Discovery and Characterization of Novel Antibiotic Synthesis and Resistance Determinants from Diverse Microbial Metagenomes

by

Larissa Cherie Parsley

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 9, 2010

Keywotds: metagenomics, polyketide, antibiotic resistance,
activated sludge, soil, phylogeny

Approved by

Mark R. Liles, Chair, Assistant Professor of Biological Sciences
Paul A. Cobine, Assistant Professor of Biological Sciences
Willie F. Harper, Jr., Associate Professor of Civil and Environmental Engineering
Omar A. Oyarzabal, Associate Professor of Biological Sciences

Abstract

The application of metagenomics to investigate the genetic and functional

diversity of as-yet-uncultured microorganisms from natural environments has been one of

the most significant breakthroughs in molecular microbiology. Using a variety of

metagenomic approaches, we conducted several studies of wastewater and soil

communities, which are known to harbor a dynamic and complex assemblage of

microorganisms expressing an array of metabolic activities.

In the first study, we explored the bacterial and viral diversity in activated sludge

(AS), its influent (IN, and a laboratory-scale sequencing batch reactor (SBR) originally

seeded from the AS aeration basin using both culture-based and metagenomic

approaches. A total of 91 unique cultured bacterial isolates, 28 cultured bacteriophages,

1,103 bacterial 16S rDNA clones, and 1,779 viral metagenomic clones were subjected to

various phylogenetic and functional analyses, including BLAST comparisons,

microscopic observations, metabolic profiling, and biochemical characterization. These

surveys represent the most comprehensive census of wastewater microbial diversity than

ever performed.

The second study focused on discovery of antibiotic resistance determinants

encoded on bacterial, plasmid, and viral metagenomes. All three genetic sources

were screened for resistance to any of 12 antibiotics, and clones or plasmids of interest

were subjected to bioinformatic and biochemical analyses. We identified several resistance genes of interest, including six clones that exhibited high levels of chloramphenicol (Cm) resistance but that share no homology with any known CmR genes. The potential for the lateral transfer of antibiotic resistance genes was also addressed, and several putative mobile genetic elements were identified from all three metagenomes, and possible *in vivo* transposition of kanamycin and ampicillin resistance from the multi-drug resistant plasmid pAS1 to *E. coli* was observed.

The third study describes the discovery of novel pathways from a soil metagenomic library involved in the biosynthesis of polyketides, a structurally diverse group of secondary metabolites that often exhibits antimicrobial activity. Although many of the cloned pathways shared homology with known genes, others appear to be distantly related to genes associated with polyketide synthesis. More specifically, many of the pathways may have originated from the *Cyanobacterial* lineage, and one clone exhibited significant similarity with *Solibacter usitatus*, an as-yet-uncultured species from the poorly-characterized bacterial division *Acidobacteria.*

Because metagenomics is an evolving field that depends on the emergence of new approaches and technologies for success, we also developed two protocols for the characterization of diverse bacterial metagenomes. The first protocol describes a novel process for the recovery, purification, and cloning of pure, high-molecular-weight metagenomic DNA from soil and is a valuable contribution toward overcoming the challenges of constructing large-insert libraries from complex environments. The second protocol enables the fluorescence *in situ* hybridization (FISH) of multiple probes to

bacteria in pure or mixed cultures and environmental samples using an in-solution

approach. This method provides a more cost- and time-effective alternative to traditional

FISH analyses, and the probed sample can be used directly in downstream applications

such as flow cytometry or fluorometry.

Acknowledgments

I wish to thank my mentor, Dr. Mark Liles, for his constant encouragement, good humor in the midst of negative results, and his willingness to place his students' well-being above his own. I could not have done this without him. I'm also indebted to my advisory committee, Drs. Willie Harper, Al Brown, Evert Duin, Paul Cobine, and Omar Oyzarbal for sharing their wisdom, advice, and expertise throughout the years. I would like to acknowledge Nadia Bhuiyan, Nancy Capps, Erin Consuegra, Ann Marie Goode, Kavita Kakirde, Andrew Land, and Stephen Thomas for their valuable contributions to these research projects.   I owe much of this accomplishment to my friends Kauser Hussain, Trafina Jadhav, and Molli Newman, and to Drs. Randall Harris and Ann Rushing, who first taught me what it means to be a scientist. I'm also thankful for the support of Jack and Kathy Feminella, which has been invaluable during my time at Auburn. Finally, I am forever grateful to my loving parents and best friends Ralph and Robyn Parsley, who have inspired me to push beyond all boundaries by their perfect example and countless sacrifices. Thank you for teaching me the importance of hard work, integrity, and personal responsibility, and the eternal rewards of persevering in a world that doesn't always value these qualities. I would like to dedicate this dissertation to my Savior Jesus Christ who, by His infinite grace, allows us to study His creation. May this work in some way bring glory to His name.

Table of Contents

List of Tables

List of Figures

CHAPTER I

LITERATURE REVIEW

A. METAGENOMICS

Metagenomic analyses provide extensive information on the structure, composition, and predicted gene functions of diverse environmental microbial communities. Each environment presents its own unique challenges to metagenomic investigation and requires a specifically designed approach to fully characterize the functional and genetic diversity of their indigenous microbial communities. Here, the advantages, disadvantages, technical hurdles, and significant discoveries associated with metagenomic-based approaches are discussed.

1. Accessing As-Yet-Uncultured Microorganisms in Diverse Environments

Previous cultivation-based studies have already proven natural environments to be an excellent resource for the discovery of novel genes, biological activities, and microbial species. The extensive genetic diversity of microorganisms makes them a rich reservoir of new bioactive compounds and metabolic processes, but the inability to successfully cultivate the majority of these microbes has been a deterrent to exploiting microbial communities to their full potential. Overcoming the "great plate count anomaly" (252) in order to access both greater numbers and a greater diversity of bacteria has become one of the most significant breakthroughs in the field of molecular microbial ecology.

Many studies have demonstrated that the true phylogenetic and functional diversity of microbes in various habitats, including soil, remains untapped (10, 36, 108, 218, 284). Fortunately, the recent development of metagenomic-based approaches now makes it possible to further investigate this diversity without the inherent biases of cultivation.

Metagenomics can be defined as the genomic analysis of the collective microbial assemblage found in an environment and often involves cloning the total DNA extracted from the community in an appropriate vector followed by expression in a heterologous host (94). However, many metagenomic approaches do not depend on cloning, such as 454 pyrosequencing and community-wide fingerprinting methods (e.g., T-RFLP). One of the main advantages of metagenomics is that it is a powerful technique for identifying genes from as-yet-uncultured microbes which have very low or no similarity to known genes and may encode novel bioactive compounds or metabolic pathways. Many studies have used metagenomics to characterize the microbial diversity in a number of different environments, such as soils (218, 271, 277), the complex microbiome of the rumen (64), planktonic marine microbial assemblages (23, 37), deep sea microbiota (135), acid mine drains (273), arctic sediments (114) and the Sargasso Sea (276).

The extent of known microbial diversity is much less than that of the actual biodiversity. Cultivated bacteria constitute only a small fraction of the total bacterial diversity found in natural environments (108), but the development of metagenomic technologies has enabled better characterization of uncultured microbes both *in situ* and *in vitro*. The use of 16S rRNA genes as powerful markers in phylogenic studies due to their universal distribution, the presence of extremely conserved and variable regions, and the ability to yield an RNA product that serves as a target for hybridization has

2

revolutionized the field of microbial ecology. PCR amplification of 16S rRNA genes is

commonly utilized for studying the phylogenetic diversity of microbial consortia using

community fingerprinting methods such as ribosomal intergenic spacer and terminal

restriction fragment length polymorphism analyses. 16S rRNA sequences are also used to

design universal or taxa-specific probes and primers for identifying large genomic

fragments in environmental DNA libraries (141, 203, 255), as well as fluorescence *in situ*

hybridization using 16S rDNA oligonucleotide probes (148, 183, 222, 299). In addition,

heterogenous 16S rDNA amplicons obtained from PCR with metagenomic DNA as a

template can be used to construct clone libraries that can be sequenced and analyzed with

bioinformatics tools. In any PCR-based metagenomic survey, it is critical to design

appropriate primers for successful annealing that allows for the detection of diverse

bacterial species. As the number of metagenomic projects increases, however,

information on more rare 16S rRNA genes will also increase.

Although culture-independent methods of DNA isolation enable the recovery of

DNA from a great diversity of microorganisms within environmental samples, they may

still fail to identify certain members of the microbial community due to biases associated

with DNA extraction and PCR amplification (16). An alternative to PCR-based

phylogenetic analysis lies in large-scale shotgun sequencing (276), which can predict the

level of diversity within a community but may not necessarily provide information on

novel genes discovered.

2. Challenges to Obtaining High-Quality Metagenomic DNA

Environmental metagenomes remain both promising yet problematic microbial

communities to be explored. Often, environmental DNA is contaminated with associated

organic compounds that can inhibit downstream applications such as PCR and metagenomic library construction. Depending on the physical and chemical composition of the environment, these contaminants may include humic acids, polyphenols, polysaccharides, and nucleases, which can also degrade DNA (79, 260, 267, 304). The removal of these co-isolated contaminants is critical to successful DNA manipulation, and extraction and purification methods should be selected to yield DNA suitable for the ultimate metagenomic application.

2.1. General Issues Concerning DNA Extraction Methods

When obtaining metagenomic DNA for any application, it is critical to isolate DNA from diverse microorganisms that are truly representative of the entire community. Otherwise, downstream analyses may be biased against or in favor of a particular group of microbes, which leads to an incomplete characterization of the metagenome of interest (74, 141). Thus, one of the major issues to consider when extracting metagenomic DNA relates to selecting a method that will effectively yield DNA from as many different microbes as possible (e.g., both Gram-positive and –negative bacteria).

Two general approaches exist for environmental metagenomic DNA extraction: 1) DNA is directly extracted from the environmental sample; or 2) bacterial cells are recovered from the environmental sample prior to lysis and DNA purification (i.e., "indirect extraction"). Direct extraction of metagenomic DNA has many advantages, including its decreased processing time and that it provides a greater DNA yield compared to other methods (182). Unfortunately, this method often results in the significant co-isolation of contaminants from the environmental sample and a higher percentage of non-bacterial DNA (182, 267, 272). Indirect DNA extraction overcomes

4

some limitations of the direct extraction method because it results in less non-bacterial DNA (184) and can yield DNA from phylogenetically diverse origins (81). However, indirect extraction methods are more time-consuming, in general provide lower DNA yields, and may bias against microorganisms that are not easily dissociated from the environmental matrix. Selecting which extraction method to adopt depends greatly on the desired downstream application. For example, the direct extraction method is most commonly used for PCR-based or pyrosequencing studies. The DNA template for PCR or pyrosequencing must be extensively purified in order to remove environmental contaminants, and the purification schemes used will result in genomic DNA fragmentation. The decrease in genomic DNA fragment size resulting from harsh direct extraction and purification methods is typically not a problem in PCR-based or pyrosequencing studies since the targeted genetic loci are of relatively small size (e.g., less than a few kilobase pairs). Conversely, the indirect extraction method is generally used when the size of extracted DNA fragments must be maintained for use in constructing large-insert metagenomic libraries and/or a high proportion of bacterial DNA template is desired prior to the molecular application.

Selecting a cell lysis method is another critical issue, as this will determine if DNA from a particular microbial group is preferentially extracted. In most cases, the successful cell lysis method should yield genomic DNA from both Gram-positive and – negative bacteria. This is generally achieved by using a combination of mechanical (e.g., sonication, bead-beating) and chemical (e.g. lysozyme, detergents) means, which provides conditions harsh enough to lyse Gram-positive cells. Many studies have evaluated the efficiency of this "combined lysis" approach and found that although this

method yields DNA from diverse bacteria, it also results in sheared DNA that may not be suitable for some large-insert applications and can cause PCR artifacts (67, 172, 245).

More recently, commercial kits have provided an alternative to these extraction methods. These kits often include one or more purification steps that remove environmental contaminants without the need for organic extraction with phenol and chloroform. Another advantage of using commercial kits is that the procedure has been standardized, which leads to more consistent DNA yields and decreased processing time. However, two major limitations of commercial kits involve the scale of the extraction and the size of DNA obtained. If DNA will be directly extracted from the environmental sample, a large quantity of the sample may be required to isolate sufficient DNA for downstream applications and a macro-scale commercial kit should be adopted (287). Also, column-based methods generally will retain DNA greater than 20 kb and are not suitable for large-insert library construction.

Although certain DNA extraction methods are appropriate for specific metagenomic applications, DNA yield and size are not the only factors to consider when choosing an extraction approach; equally important are the purity and integrity of the resulting DNA. Some methods, such as direct extraction, require significant purification steps in addition to the extraction protocol, and most molecular applications, such as PCR-based studies, depend on high-purity DNA for success. Therefore, it is essential that an effective purification strategy be selected with respect to both the extraction method used and the desired downstream application.

2.2. Purification of DNA for PCR-based Applications and Pyrosequencing

The use of PCR in various metagenomic approaches has become invaluable in determining the phylogenetic diversity of environmental microorganisms, as well as identifying genes of interest from as-yet-uncultured microbes. Methods such as denaturing gradient gel electrophoresis, analyses of clone libraries of 16S/18S rDNA or specific functional genes, and pyrosequencing-based metagenomic surveys allow for the characterization of microbial metagenomes from a variety of habitats. However, all of these techniques rely on a sufficient yield of high-purity DNA from both Gram-positive and Gram-negative bacteria, which can be quite challenging with respect to complex microbial communities.

When selecting which DNA extraction and purification methods to use with respect to PCR, there are three major goals to consider: 1) isolating genomic DNA from a diverse group of microorganisms that is representative of the entire community; 2) obtaining a sufficient amount of DNA as template for PCR; and 3) removing potential PCR inhibitors (e.g., humic acids, tannins) that are co-isolated with the DNA. Regardless of which DNA extraction method is used, it is likely that further purification will be required for consistent PCR amplification, although the degree of purification necessary may be reduced depending on which extraction method is selected.

Many DNA purification methods have been developed to remove PCR inhibitors, such as the addition of phenol and chloroform, hexadecyltrimethylammonium bromide (CTAB), and polyvinylpolypyrrolidone (PVPP), as well as CsCl density centrifugation and hydroxyapatite column chromatographic purification (104, 124, 137, 219, 231).

However, it has been shown that many of these methods (PVPP addition, CsCl density centrifugation, and hydroxyapatite column chromatographic purification) resulted in a decreased DNA yield (254). Some studies have found that a washing step prior to cell lysis is useful for the removal of soluble inhibitors and extracellular DNA (97, 297). The precipitation of DNA following cell lysis with ethanol, isopropanol, or polyethyleneglycol (PEG) is a common purification step, but it has been found that ethanol precipitation often results in a greater co-precipitation of humic acids and a lower yield of DNA than PEG precipitation. However, PEG must be removed by phenol extraction because it can interfere with PCR (202). Unfortunately, many environmental samples require a combination of these purification steps, which significantly increases processing time and can lead to an even greater loss of DNA. For example, one study compared DNA extracted from five different soils with various organic matter content and found that the high organic matter sample required five purification steps to yield DNA for PCR (275). Conversely, if the amount of contaminants is low enough, alternative methods may be sufficient to permit PCR amplification, such as diluting the DNA template (and thus the contaminants) (4), adding bovine serum albumin to the reaction mixture (129), performing sucrose gradient centrifugation (199) or simple dialysis (216), and embedding the DNA in agarose prior to the PCR (173, 303). The latter method is most useful when humic acids must be removed and it is desired to use the recovered DNA directly for PCR. Because humic acids migrate through agarose more quickly than genomic DNA, electrophoresing the DNA through low-melting-point agarose and then extracting it from the gel allows for its direct use in PCR (101, 137, 302). Recently, a protocol was developed which allows for the parallel processing of

8

multiple samples with two steps of purification, embedding the DNA in an agarose plug and incubating the plug in a formamide and NaCl solution to facilitate the separation of contaminants from the DNA while maintaining its integrity (177).

Pyrosequencing, a fairly recent alternative to Sanger sequencing, is often referred to as "sequencing by synthesis" because it involves DNA polymerase synthesizing a complementary strand of DNA in the presence of a single-stranded template (217). Because pyrosequencing relies on a previously PCR-amplified target sequence and the activity of DNA polymerase during the sequencing reaction, the same environmental contamination challenges that apply to other PCR-based applications also apply to pyrosequencing. However, since pyrosequencing currently generates reads only 300-500 bp in length, obtaining intact, larger DNA is not as critical as it may be in other PCR-based studies (163).

2.3. Preparation of DNA for Small-insert Cloning

The construction of small-insert metagenomic libraries is a useful method for the identification of specific target genes or for screening metagenomic libraries for a functional activity that is encoded by a relatively small genetic locus. Ideally, these libraries should contain DNA reflecting the composition of the microbial community from which the library is constructed, but some prokaryotic taxa may be selectively omitted given potential biases in cell lysis and cloning. Therefore, it is important to select a DNA extraction method that will successfully lyse both Gram-positive and – negative cells (if the library is from a bacterial metagenome) and also yield a high proportion of DNA from the microorganisms of interest (e.g., bacteria vs. fungi).

The preparation of DNA for small-insert libraries is a somewhat different process from that for the preparation of DNA for PCR-based applications, although many of the extraction and purification issues are similar. For example, a significant yield of DNA is necessary for successful library construction, and environmental contaminants co-isolated with the DNA such as humic acids can interfere with efficient cloning. DNA extraction and purification should be harsh enough to lyse a variety of microbes and remove the majority of contaminants but not so stringent that the resulting DNA is significantly sheared. If the desired average insert size of the DNA is < 20 kb, a commercial kit (e.g., MoBio Laboratories, Qiagen) may provide a useful method for obtaining DNA of sufficient size, purity, and yield for small-insert cloning. Although many kits do not yield DNA pure enough for direct PCR amplification, it is usually unnecessary to perform further purification for small-insert cloning. However, one study using Antarctic top soil used two separate commercial kits to further purify the DNA after cell lysis for construction of small-insert libraries (46).   In cases when commercial kits are not suitable, such as soils with a high clay content, other purification methods such as sucrose/Percoll density gradient centrifugation and Nycodenz treatment have been shown to generate DNA appropriate for small-insert cloning (18). Regardless of the extraction and purification method chosen, replicate samples may be processed in parallel, pooled after the extraction and/or purification, and precipitated in order to concentrate the DNA and achieve sufficient yield prior to cloning. Following extraction and purification of the DNA, it may be physically sheared or restriction digested and then size-selected by extracting the DNA in the desired size range from an agarose gel. Because the size-selected DNA will likely be < 20 kb, it can be column-purified with any number of

commercial kits without the need for manual purification methods. Alternatively, the gel slices may be treated with GELase or the DNA may be electroeluted from the gel; both of these methods will be discussed below.

2.4. Preparation of DNA for Large-insert Cloning

Large-insert metagenomic libraries are the most challenging to construct but also provide significant advantages since they enable identification and characterization of intact functional pathways encoded on large, contiguous DNA fragments (54, 218). All of the considerations discussed previously regarding the selection of DNA extraction and purification methods apply to large-insert cloning, along with an additional critical issue: the construction of large-insert metagenomic libraries typically depends on obtaining DNA in excess of 100 kb. However, most extraction and purification methods result in DNA significantly smaller than this size (130, 164, 269). Although a few methods can yield DNA from soil greater than 1 Mbp in size (24, 142), it has been demonstrated that these indirect extraction methods also result in inefficient cloning due to contaminating nucleases that are co-isolated with the metagenomic DNA and require further purification prior to library construction.

The successful recovery of high molecular weight (HMW) metagenomic DNA from environmental bacteria presents many extraction and purification challenges. Since the goal of many large-insert libraries is to capture functional pathways from diverse bacterial hosts, obtaining DNA from an assemblage of Gram-positive and -negative cells that contains little or no contaminating DNA is critical to constructing a metagenomic library with representative community DNA. However, the harsh extraction methods necessary to achieve more complete lysis of diverse bacteria also result in sheared DNA

that is much too small for large-insert cloning. The use of indirect DNA methods can somewhat alleviate this dilemma by first separating the cells from the environmental sample, embedding them in an agarose plug, and then carefully lysing the cells and purifying the resulting DNA rather than performing the extraction *in situ*. Repeated homogenization and differential centrifugation are often sufficient to separate the cells from the environmental sample (73, 105), although other dispersion methods include the use of cation-exchange resin (113, 149) and incubating the sample with sodium deoxycholate or polyethylene glycol (143).

Several purification methods can also lead to lower yields and decreased size of DNA. Commercial kits are generally not suitable for isolating HMW DNA from environmental samples because many of them have DNA size limitations and do not recover enough HMW DNA for efficient cloning. Often, a combination of phenol:chloroform extraction, CTAB treatment to remove polysaccharides (57), and incubation with PVPP to remove polyphenols (56) may be needed to achieve high-purity DNA, but with each purification step there is a risk of losing DNA and thus lowering the resulting yield.

The choice of extraction and purification method also depends on which large-insert cloning strategy will be employed, such as fosmid-based or bacterial artificial chromosome (BAC)-based. Metagenomic libraries constructed in a fosmid vector are introduced into their heterologous host using a λ phage-based packaging system, which limits the clone insert size to 40-50 kb. Although DNA isolated for fosmid libraries must be treated carefully to prevent excessive shearing of DNA, using a fosmid vector does allow the use of harsher extraction and purification methods than those that may be used

for BAC cloning. Also, during fosmid library construction, the DNA is typically size-selected by physically shearing the DNA into fragments of a desired length rather than by restriction digestion. This "direct size-selection" method eliminates the need for gel extraction (which can lead to DNA loss) and the possibility of DNA degradation due to over-digestion. An alternative to the physical shearing method was proposed by Quaiser and colleagues, who constructed fosmid libraries containing metagenomic DNA contaminated with humic and fulvic acids by embedding the DNA in agarose, incubating it with polyvinylpyrrolidone (PVP), and then combining the subsequent removal of the PVP with the size-selection step. This method resulted in purified, "clonable" DNA in the 30-100 kb size range (203).   In combination with other purification steps, the inclusion of a formamide plus NaCl treatment was shown to significantly reduce the level of nuclease contamination and allow for the efficient cloning of large DNA fragments into fosmid or bacterial artificial chromosome (BAC) vectors (143). Factors that have been demonstrated to affect the size of DNA recovered include not only the DNA extraction method used but also the microbial growth status and chemical composition of the environment (25). In general, DNA extracted from bacterial cells is significantly larger than DNA directly extracted from the environmental sample but is also found in lower yields (143).

Cloning HMW DNA into a BAC vector presents a unique challenge because the DNA must be of sufficient purity, size, and yield to achieve metagenomic libraries with a sufficient number of large-insert transformants. As discussed previously, many of the extraction methods available either result in DNA that fits one or two of the above criteria but not all three. Nevertheless, several studies have successfully constructed and

screened metagenomic libraries in a BAC vector with DNA from various habitats (68, 188, 201, 218), emphasizing the necessity of selecting the appropriate combination of DNA extraction and purification methods.

Unlike fosmid cloning, standard BAC library construction methods require that DNA be partially restriction digested and size-selected on an agarose gel prior to ligation. The efficiency of the restriction digest depends on the purity of the DNA, as co-extracted contaminants can inhibit the restriction enzyme and result in insufficient digestion (202, 213). While these extra steps serve to further purify the DNA, they also provide an opportunity for DNA loss. After restriction digestion and electrophoresis, gel slices containing DNA of the desired size can either be treated with GELase (Epicentre) or electroeluted from the agarose (185). DNA loss can be minimized by running multiple size-selection gels, using wide-bore pipet tips to reduce shearing, and keeping gel slices as small as possible during size-selection(143)

3. Various Metagenomic-Based Approaches

A variety of approaches may be employed for analyzing environmental metagenomes, depending on the specific aims of the study (e.g., phylogenetic survey, isolation of functional activity). The ultimate downstream application should dictate the methods used for DNA extraction and purification, library construction, and screening (if necessary), as well as the types of analyses that are applied to the environmental sample (e.g, *in situ* methods, PCR). Selecting the appropriate combination of methods and analyses will ensure that a comprehensive and accurate study is performed which is truly representative of the microbial community present within the environmental sample. When working with environmental samples that have not been well-characterized, it may

be advisable to utilize a variety of different methods for DNA extraction and purification to empirically determine the ideal combination that will yield high-quality metagenomic DNA. Many of the metagenomic-based approaches used to study environmental communities, as well as the approach-specific factors to consider when performing such analyses, are discussed below.

3.1. *In situ* methods using fluorescence *in situ* hybridization (FISH)

FISH is a powerful tool for the visual identification and enumeration of specific bacteria from environmental samples which combines microscopy, phylogenetic analysis, and often other methods such as flow cytometry (268). Fluorescently-labeled oligonucleotide probes are designed to target DNA in individual cells within a sample without the need for prior cultivation. With the application of 16S rDNA-based phylogenetic typing, the use of this gene target in FISH methods has become a valuable means to identify and quantify viable cells with specific taxanomic affiliations (6, 92, 289) or metabolic abilities (40, 61, 301). In addition to 16S rDNA genes, FISH has been used to target a variety of functional genes (111, 194), investigate bacterial endosymbiosis (106, 223), and observe single-cell phylogeny and function (107, 146). FISH methods yield best results with samples in which cells are actively growing, as this often leads to greater numbers of gene targets and therefore higher levels of fluorescence. As an alternative, laser confocal scanning microscopy is an excellent technique for the visualization of low fluorescence signals and allows for the observation of microbes immobilized in a complex matrix (6).

One challenge related to FISH methods, especially with highly heterogeneous environmental samples such soils, involves the sufficient interaction of the probe with its

intended gene target. Particulates and aggregates can prevent this interaction because microbial cells may be sequestered within these complexes, thus the probe cannot successfully hybridize to its target. Secondly, autofluorescence of environmental particles and non-specific binding of the probe due to aggregation can adversely affect the integrity of the fluorescent signal and make cell enumeration difficult. To circumvent these issues, cell extraction coupled with density gradient centrifugation has been shown to yield a high level of cell recovery from environmental samples for clear visualization of fluorescence (21), as well as the entrapment of bacteria of interest in cation-exchange membranes that can then be hybridized with appropriate probes (133). However, the successful hybridization and subsequent detection of representative microorganisms from complex environments remains a significant challenge that must be considered prior to employing FISH-based investigations.

3.2. *In situ* Methods using Stable Isotope Probing (SIP)

SIP is a pioneering molecular technique that links phylogeny with function by studying active microorganisms in environmental samples and can be used to characterize microbes involved in specific metabolic processes (68). The underlying principle is to use stable isotopes (e.g. $[^{13}C]$) to label specific substrates provided to a microbial community and monitor their subsequent assimilation and/or degradation. Although stable isotopes of any element found in the DNA molecule can be used for DNA-SIP, $[^{13}C]$-labeled substrates are often preferred because carbon is the most abundant element in DNA and may yield higher levels of isotope detection. The isolated DNA from the treated environmental sample is then subjected to ultracentrifugation or isopycnic density gradient centrifugation to separate the labeled DNA for analysis. Only

organisms that utilized the labeled substrate will incorporate the stable isotope into their DNA, thus differentiating the active target population from the inactive portion of the microbial community.

The stable isotope labeled-DNA, also referred to as "heavy" DNA, can be used in constructing bacterial 16S rRNA gene clone libraries. Subsequent sequencing of the 16S rRNA gene helps to identify the microorganism involved in the utilization of the specific substrate. Although 16S rRNA has been widely used for gaining insights into the phylogeny of microbial communities, sequence analysis alone gives limited information on the functional role of a microorganism in the environment (292). Metagenomic libraries may sometimes fail to capture low-abundance species but, in conjunction with DNA-SIP, can recognize a multitude of functional genes because the labeled DNA represents the whole genomes of active populations (24) involved in a specific metabolic process of interest. DNA-SIP has also been used to retrieve genomic fragments of an active population by cloning the [^{13}C]-incorporated DNA without initial PCR amplification (68).

Thus, SIP coupled with metagenomic techniques is an excellent strategy to explore functional genes from uncultivated organisms that are actively involved in the utilization of a variety of compounds or in degradation of pollutants in the environment but are relatively low in abundance. Like FISH, SIP analysis of complex microbial communities has its limits and is unsuitable for use under certain conditions. For example, unintentional dilution of the labeled substrate with the unlabeled substrate before assimilation decreases the relative amount of DNA that is labeled, thus reducing the chances of identifying the target microorganisms. Also of concern is the potential

incorporation of [^{13}C] from the labeled substrate into metabolic intermediates that might be metabolized by other non-primary organisms (68). When using DNA-SIP for metagenomic analyses, the small amount of heavy DNA available can also be a hurdle to successful library construction. To overcome this challenge, methods such as multiple-displacement amplification (43, 68) and community growth enrichment by sediment slurries (117) have been used to increase the amount of heavy DNA available for analysis.

3.3. PCR and Pyrosequencing

The use of PCR in any metagenomic study is critical in the determination of phylogenic diversity of microorganisms and identification of target genes from the as-yet uncultured microorganisms. A combination of approaches such as community-wide fingerprinting, analysis of 16S/18S rDNA library clones, detection of specific functional genes, and pyrosequencing is often used in characterizing the metagenomic diversity of the selected environment. For successful amplification in this context, it is important to have a sufficient concentration of the DNA template that is free from potential PCR inhibitors as described previously.

The use of universal primers that anneal to conserved regions of phylogenetically divergent loci enables the detection of diverse genes but may lack sufficient sensitivity when a template of low concentration and/or high degree of heterogeneity is used. Specific primers greatly increase the chances of amplifying and detecting a relatively minor microbial constituent that may be significant for study but may exclude a divergent gene product. Though PCR-based methods can be used for the rapid identification of target sequences from extracted environmental DNA, the quality of the amplicon must be

high in order to make accurate predictions as to its encoded product and phylogenetic origin. Thus, designing primers that encompass the entirety of the target sequence present the best chance of obtaining high-quality amplicons for downstream (274).

Pyrosequencing is also known as "sequencing by synthesis," wherein DNA polymerase is used to synthesize a complimentary DNA strand from a single-stranded DNA template (217). PCR-amplification of the target sequence is sometimes used prior to pyrosequencing and hence it is prone to similar problems related to purity and integrity of the isolated DNA. Pyrosequencing has the capacity to generate thousands of relatively short sequences in a single run, and the sequence information obtained has been shown to be adequate in differentiating samples based on physiological functions such as carbon utilization, nitrogen assimilation, and respiration (10). This technology has also been used to characterize a wide variety of unique sequences from deep sea samples (135). Using pyrosequecning for metagenomic analysis eliminates the need for cloning and is thus free from the associated biases, which will be discussed later in more detail. One potential drawback to pyrosequencing lies in the difficulty of data interpretation due to the sheer volume of resulting sequence information. Adequate bioinformatics pipelines should be in place to enable comprehensive analyses of the resulting sequence data.

3.4. Small-Insert Library Construction

Small-insert metagenomic libraries provide a means of discovering specific genes or functional activities of interest that are encoded within a relatively small region of DNA. Because these clone inserts do not exceeed 20 kb, harsher DNA extraction and purification methods may be used, which may result in the collection of more diverse metagenomic DNA and increased cloning efficiency due to the use of high-purity DNA.

These considerations have been discussed in detail previously; however, the selection of an appropriate cloning vector and host also determines the success of small-inset library construction and screening.

Vectors used in construction of small-insert libraries often possess a promoter for transcription of the cloned gene inserts and should be compatible with the host selected for screening. A vector with two promoter sites flanking the multi-cloning site facilitates gene expression that is independent of gene orientation and the promoters associated with inserts (134). With the possibility of the expressed gene products having toxic effects on the host organism, it is important to regulate the expression levels of the cloned genes, which can be achieved by using vectors with inducible control over gene expression of the insert or plasmid copy number. Additional issues to consider when selecting a vector include its ability to replicate in multiple hosts if heterologous expression is of interest, as well as the insert size limitations of the vector.

The selection of an appropriate host organism is dependent on the ultimate goal of the study. Although the utility of using *E. coli* as a heterologous host for metagenomic library construction has been well-established (87, 94, 195, 218), other bacterial hosts may be more suitable for some applications, such as demonstrating horizontal gene transfer or reproducing a specific functional activity in a host that represents a dominant phylogenetic taxa from the environmental sample. Regardless of the host species that is chosen, its genotype should be considered (e.g., *rec*A⁻) as well as its suitability for high-efficiency cloning (e.g, highly electrocompetent).

Refer to Figure 1.1 for a schematic representation of the overall steps involved in metagenomic library construction.

20

3.5. Large-Insert Library Construction

Large-insert metagenomic libraries contain large, contiguous DNA fragments that have a high probability of expressing novel compounds or metabolic activities from pathways encoded by the cloned genes (54, 218). This strategy increases the likelihood of cloning intact genetic loci involved in the synthesis of antimicrobial compounds, multiple enzymes with catabolic activity, or operons encoding other complex metabolic functions.

The challenges of obtaining sufficiently pure HMW DNA has been described above, but additional factors can influence the successful construction of large-insert libraries, such as host and vector selection. Because the success of many large-insert metagenomic libraries depends on their ability to capture large, intact genetic pathways, the selection of an appropriate vector is critical to the maintenance and expression of the cloned pathways. Several vector options exist for cloning HMW DNA from environmental samples, such as cosmids, fosmids, and BACs. The cosmid, a hybrid plasmid that contains *cos* sequences from the λ phage genome, was one of the first vectors used for cloning (50). While typical plasmids can maintain inserts of 1-20 kb, cosmids are capable of containing DNA inserts up to 40 kb. They can replicate like plasmids in presence of a suitable origin of replication and frequently possess selective genes such as antibiotic resistance to facilitate transformant screening. Fosmid vectors, which are similar to cosmids but are based on the *E. coli* F-factor replicon, were developed for constructing stable libraries from complex genomes (123). The low copy number of fosmid vectors offers higher stability than comparable high-copy number

21

cosmids. Fosmid copy number is tightly regulated in *E. coli* to 1-2 copies per cell, and

fosmids can typically accommodate cloned inserts between 40 and 50 kb. BAC vectors

are based on the same F-factor replicon but have the capacity to maintain large inserts in

excess of 100 kb (238). Along with the long-term stability conferred by the F-factor for

maintenance, BAC vectors also possess the property of conditional amplification that

allows for alternating between single-copy and high-copy BAC maintenance (291).

Fosmid vectors have a greater insert size limitation (40-50 kb) when compared to

BAC vectors, but they result in higher numbers of transformants with cloned inserts.

Conversely, BAC vectors can accommodate insert sizes in excess of 100 kb, but their

cloning efficiency is lower than that of fosmid vectors. As mentioned previously, HMW

DNA for fosmid-based cloning may be treated with harsher extraction and purification

methods, which could yield a higher concentration of DNA from more diverse

microorganisms than that of DNA isolated for BAC-based cloning. However, because

BAC vectors can stably maintain cloned inserts hundreds of kilobases in size, they offer a

greater chance of isolating intact pathways that encode complex functional activities.

Therefore, the predicted size of the pathway of interest, its native level of activity, and its

relative abundance within the community must be considered when choosing a suitable

large-construct vector.

Similarly with small-insert libraries, *E. coli* is the preferred host for the

construction of large-insert metagenomic libraries because it has been well-characterized

genetically and provides high-efficiency cloning of DNA. This host has been

successfully used as an expression host for the detection of novel compounds from a rich

diversity of metagenomic DNA by function-based screening methods (95, 99). In

addition, *Bacillus subtilis* and *Streptomyces lividans* have been used as successful heterologous hosts for library construction, but they have more stringent promoter recognition and regulation properties when compared to *E. coli* (54). Because large-insert libraries likely harbor some biologically active pathways that may encode gene products that are toxic to the library host, it is important to consider the use of multiple hosts to increase the probability of identifying and characterizing the function of interest. It has been shown that clones positive for a specific activity detected using one host may not be detected in a different host and vice-versa (55, 139). A range of Gram-positive and -negative bacteria can be used as hosts for heterologous expression, and the corresponding vectors selected should be compatible with those hosts. In addition, the promoter and ribosome binding sites should be compatible with the host machinery. Several other factors are necessary for successful expression of the cloned pathways (e.g., co-factors, post-translation modification enzymes, inducers, chaperones etc.), which may be provided by the vector or the host organism.

3.6. Screening of Metagenomic Libraries

The analysis of metagenomic libraries involves two main strategies, function-based or sequence-based screening. The choice of screening method depends on many factors, including the type of library constructed, the genetic loci or functional activity of interest, and the time and resources available to characterize the library. Both approaches offer advantages and disadvantages, which will be discussed here.

Function-based methods involve screening a metagenomic library to detect the expression of a particular phenotype conferred on the host by cloned DNA (100). Because the frequency of discovering active pathways from metagenomic libraries is

often low, high-throughput screening of library clones is the most efficient approach for function-based detection of activity, such as biodegradation or the synthesis of antimicrobial compounds. Screening for clones that have acquired resistance to antibiotics or heavy metals due to the expression of the cloned genes can be performed by excluding microorganisms which are unable to grow in the presence of these selective compounds (166, 210).

Another approach for the functional screening metagenomic libraries is to use host strains or mutants of host strains that require heterologous complementation for growth under selective conditions (240). Growth is exclusively observed in the case of recombinant clones that possess the gene of interest and produce an active product. This strategy has been applied for the detection of enzymes involved in poly-3-hydroxybutyrate metabolism (282), DNA polymerase I (241), operons for biotin biosynthesis (72), lysine racemases (42), glycerol dehydratases (125) and naphthalene dioxygenase (33).

Screening can also be performed by detecting a specific phenotypic characteristic, in which individual clones are assayed for a particular trait. Incorporation of specific substrates in the growth medium helps to identify the corresponding enzymatic activities encoded by metagenomic clones. Examples include the identification of esterases (45, 71) by formation of a clear halo around a colony on the indicator medium and the identification of extradiol dioxygenases by the production of a *meta*-cleavage yellow compound (256). Production of compounds with antimicrobial activity is detected by growth inhibition assays of a suitable tester organism using soft agar overlays over the clone colonies or a microtiter plate assay using the supernatant extracts from the clone

cultures (35, 54, 55, 218). As discussed previously, substrate-induced gene expression

screening may be a suitable approach to enhance the discovery of functional pathways

from microorganisms that exist in the community at a relatively low abundance.

Although function-based screening is a powerful tool to identify novel natural

products or metabolic activities from as-yet-uncultured organisms, it is often limited by a

number of obstacles which are difficult to overcome. For example, identifying clones of

interest depends on a series of successful gene expression steps, such as transcription,

translation, protein folding, and secretion from the host organism. As discussed above,

the gene product may be toxic to the host and is therefore not identified in function-based

assays, although it may be expressed in a different library host. In addition, the isolation

of intact, functionally active pathways relies on the cloning of high-purity HMW DNA

that is likely to harbor large, contiguous regions of DNA.

Sequence-based screening involves using conserved nucleotide sequences of

known genes to design hybridization probes or PCR primers for detection, amplification,

and identification of specific target genes. A second approach is the direct sequencing of

the insert DNA and subjecting the sequences to bioinformatic analyses (83, 244).

Sequence-based methods often have a limited scope because they rely on sequences of

known genes for probe or primer design, although universal oligonucleotides targeting

conserved regions within genes of interest may increase the likelihood of identifying

novel genes. Other disadvantages include the lack of functional data from the clone of

interest, as well as the limited functional information that can be inferred from sequence-

based analyses alone. However, successful sequence-based screening does not depend on

the heterologous expression of gene products in the library host, and sequence data can

provide insight into the phylogenetic origin of the genetic locus. Additionally, screening

a library by hybridization or PCR requires significantly less time and resources than

function-based screening, and multiple genes of interest within the library can be

investigated simultaneously.

3.7 High-Throughput Sequencing

With advances in automated DNA sequencing, the development of high-

performance computational platforms has provided additional methods for analyzing

large subsets of metagenomic DNA, such as whole genome shotgun sequencing,

microarray analysis, and sequencing of large-insert library clones.

High-throughput sequencing technologies are very useful in analyzing gene

functions and have been valuable for identifying genes involved in the synthesis of

potentially novel antimicrobial compounds.   Random shotgun sequencing of

metagenomic DNA from natural environments is used to characterize microbial

communities following the assembly and analysis of large, contiguous sequences. The

use of microarrays is another high throughput strategy for sequence-based screening of

metagenomic libraries in which vector and insert DNA from the library are spotted onto a

slide and hybridized with highly specific labeled probes (45, 230).

Whole genome shotgun (WGS) sequencing is a rapid and cost-effective method

for sequencing genomes from microorganisms that are recalcitrant to cultivation. Large

fragments of DNA are extracted directly from an environmental sample, clone sequences

are assembled into contiguous fragments, and gaps are closed by primer walking or

similar method. WGS sequencing enables the capture of the entire genetic complement

within a bacterial community (271). Novel genes that are unrelated to known genes and

cannot be detected by PCR or heterologously expressed in library host bacteria have a greater chance of being captured using WGS sequencing. Though increased availability of high-throughput sequencing technologies has made it possible for scientists to gain access to the genetic diversity within environmental communities (135), a tremendous amount of current high-throughput efforts are underway to improve the sequencing and analysis methodologies that can be applied to new sets of bulk metagenomic DNA.

Massively parallel "sequencing by synthesis" has been used to collate a number of sequences in exploring the microbial diversity in various environments. The next generation 454 Life Sciences sequencing platform, the GS FLX Titanium, provides up to 400 Mb of data per run with an average read length of 400 bp. This high-throughput sequencing technology is superior to Sanger sequencing, as it is more cost- and time-effective per sequenced nucleotide (109) and has a higher capacity of sequencing in each run. Pyrosequencing (454-based sequencing) has been widely used in the investigations of microbial diversity in soil (214), deep sea ecosystems (135) and phage populations from various environments (64). It has revolutionized metagenomic analyses with its ability to capture the entire microbial diversity in a sample by eliminating the biases associated with PCR and cloning.

In addition to the sequence-based screening of metagenomic libraries, the entire sequence of large-insert clones may be obtained using a high-throughput approach. Using 454 pyrosequencing technology, cloned DNA is bar-coded and resulting sequences are assembled into contiguous regions that can be annotated for genes of interest. Site-directed mutagenesis can then be used to investigate the genes' activity by analyzing loss-of-function mutants, or other bioassays may be employed to assign functionality to

predicted open reading frames. The effective removal of contaminating host genomic DNA is absolutely critical to the successful sequencing of large-insert, which can be achieved by digesting the cloned DNA with plasmid-safe nucleases and/or extracting the cloned DNA from an agarose gel prior to sequencing.

4. Natural Product Discovery

Natural products from cultured soil microorganisms have been an integral part of the drug discovery process for many decades. However, as-yet-uncultured microoragnisms are an untapped reservoir for the discovery of novel natural products such as antibiotics (87). Though these microorganisms themselves essentially remain inaccessible in the laboratory, the natural products they produce can be accessed via metagenomics methods. The genetic and functional diversity of these uncultured microbes can be captured by cloning large fragments of contiguous metagenomic DNA into heterologous hosts that are easier to manipulate *in vitro*, such as *E. coli* (87, 94, 218). Many low molecular weight molecules obtained from natural environments are secondary metabolites produced under specific conditions such as during developmental stages, starvation, or cell signaling (47) and exhibit bioactive properties. For many years, screening environmental microbial communities for natural products has led to the discovery of different bioactive molecules such as polyketides (89, 167, 226, 286) and proteolytic systems (23) from a variety of environmental metageomes.

Complex microbial assemblages are a rich source of antibiotics due to the biotic and abiotic stresses that affect them, which often prompts the production of secondary metabolites with antimicrobial activity. Genes encoding the biosynthesis of antibiotics are often found in clusters; therefore, the isolation of large contiguous fragments of DNA

is necessary to obtain intact biosynthetic pathways. The construction and screening of large-insert (>20 kb) metagenomic libraries remains the most efficient approach for harnessing active biosynthetic pathways from as-yet-uncultured microorganisms. However, PCR-based approaches and small-insert libraries are often adequate for the sequence-based detection of genes of interest.

Streptomyces species have been used as hosts for soil DNA libraries because of their ability to express polyketide pathways and ease of genetic manipulation due to the absence of methylation dependent restriction systems (54). In another study, a PCR-based screening approach was used to analyze DNA extracted from desert soil for identifying sequences related to OxyC, which is an oxidative coupling enzyme involved in the synthesis of glycopeptide antibiotics (19). The same group also discovered eDNA clones producing long-chain N-acyltyrosine antibiotics after screening seven libraries constructed from different environmental samples that were geographically distinct (35). Additional studies have investigated other metagenomes and have identified pathways involved in the biosynthesis of various antimicrobial compounds such as beta-lactamases (206, 292) and antifungal agents (45).

5. Xenobiotic Degradation Pathways

The term xenobiotic is derived from the Greek words *xenos* (foreigner) and *bios* (life) and can be defined as any substance foreign to an entire biological system. Xenobiotics often include compounds such as antibiotics, pesticides, hormones, and other biological or chemical contaminants that affect a microbial community. Other examples of xenobiotics include aromatic compounds and their derivatives and polychlorinated biphenyls (PCBs), which are anthropogenic chemical pollutants that persist in the

environment long-term and are recalcitrant to complete removal. Xenobiotic degradation can be achieved by biotic and abiotic reactions or a combination of the two, but many techniques used for this purpose are very expensive and have the potential to further pollute and adversely affect the contaminated sites. Harnessing the natural abilities of microbial communities in the degradation of xenobiotic compounds has emerged as a promising alternative to traditional bioremediation methods and has previously been applied as biostimulation, bioventing, bioaugmentation and composting to remove pollutants and restore environmental quality of the ecosystem. More recently, the application of metagenomics may aid in the isolation of novel catabolic pathways that degrade a wide range of xenobiotic compounds and prove to be a more "natural" alternative to mechanical and chemical bioremediative methods.

A combined approach using metagenomics and other molecular techniques is commonly used to deduce genetic information from microorganisms that are involved in degradation of xenobiotics in contaminated environments. Labeled substrates have been used to target and recover genes from populations involved in the degradation process (74). Sul et al. used [^{13}C]-labeled biphenyl to isolate biphenyl dioxygenase genes from bacteria capable of growing in PCB-contaminated river sediments, which are known to be involved in PCB oxidation. Other metagenomic studies have identified catabolic pathways that encode nitrilases, which play an important role in both biosynthetic and catabolic reactions (212) and enzymes with catalytic properties that degrade organic contaminants (122).

High throughput screening of environmental metagenomic DNA libraries has also led to the discovery of many novel enzymes that are of great use in industrial

applications. Microbes that can digest cellulose and hemicellulose produce special enzymes that target specific bonds in these molecules for their breakdown (e.g., glucanases and glycosyl hydrolases). A multifunctional glycosyl hydrolase was discovered from the diverse microflora of the rumen gut using metagenomic analyses (190). Moreover, a low pH, thermostable α-amylase with high efficiency for starch liquefaction was discovered by screening microbial DNA libraries constructed from various deep sea and acidic soil environments (208). Other examples of metagenomic analyses involved in the discovery of novel enzymes and biocatalysts include the identification of genes encoding pectinolytic lyases (249) and lipolytic enzymes such as esterases and lipases (136).

Extraction of total metagenomic DNA and cloning to construct libraries requires extensive effort and resources. The number of positive clones obtained from screening these libraries for the presence or expression of a specific gene or function is often very low because the target pathways comprise a small percentage of the total cloned DNA. There are various strategies that can be employed prior to library construction and/or screening which can improve the frequency of biosynthetic or catabolic pathway isolation. Although these methods may result in a loss of considerable diversity from the environmental sample, they also have the power to select for a particular population or function of interest. The loss of diversity can be reduced by altering the degree of the selective pressure criteria used.

Substrate-induced gene expression screening, also known as SIGEX, was proposed by Uchiyama et al., in which catabolic gene expression can be induced in the presence of relevant substrates and is under the control of regulatory elements proximal

to the catabolic genes (274). Another strategy is to enrich the environmental sample so as to select for genes of interest and involves exposing the microbial population to different physico-chemical factors prior to library construction. Use of a selective medium will result in favorable growth and enrichment of the targeted population due to specific substrate utilization.

Direct cloning from enriched cultures enables studying metabolic activities of microbial communities. Enrichment cultures are used to select specific microbes that produce a compound of interest. Instead of direct DNA extraction from the environmental sample, the sample is inoculated in a minimal or oligotrophic medium that is enriched with a particular nutrient known to favor the growth of the target microbes. Cells from the culture are then lysed, the isolated DNA is ligated to a suitable vector, and is then used in the construction of libraries followed by DNA sequencing from harvested clones. This approach has been used in the identification of biotin synthesis genes by isolation of clones carrying the biotin biosynthesis operon (72).

6. Summary

The development of metagenomic approaches has provided an unprecedented level of access to microbial genomes from many different environments, making it possible to characterize the phylogenetic and functional diversity of as-yet-uncultured microorganisms from various biomes of interest. Despite the potential power of these techniques, the success of metagenomic analyses depends upon the isolation of nucleic acids that are of the appropriate size, composition, quantity, and purity for the specific metagenomic application to be employed. Because of inherent physical and chemical heterogeneity, environmental samples present many unique and complex challenges for

the isolation of metagenomic DNA of sufficient quantity and purity for downstream

applications. In addition, the dynamic and competitive nature of many microbial

populations influence the type of metagenomic approach that is most appropriate for the

ultimate objective of the study (e.g, PCR, library screening, *in situ* methods). Selecting

the most suitable combination of soil sampling, DNA extraction and purification, cloning,

and sequencing methods and the type of metagenomic approach that is most appropriate

for the ultimate objective of the study is vital to the successful characterization of a

microbial metagenome.

The use of cutting-edge metagenomic-based technologies to access as-yet-

uncultured microbial communities has led to a remarkable increase in the discovery of

pathways that encode a range of important gene products, such as antimicrobial

compounds and enzymes involved in the catabolism of xenobiotics. Community-based

phylogenetic surveys have provided an unparalleled glimpse into the true level of

biodiversity that exists in the soil biome that was previously poorly characterized by

cultivation-based methods alone. Although environmental metagenomes present many

technical hurdles that must be overcome when performing metagenomic analyses, these

environments promise to be a rich resource of novel genetic and functional pathways

from diverse microorganisms that simply await to be discovered.

**B. ANTIBIOTIC RESISTANCE AND SYNTHESIS RELEVANT TO SOIL AND
     SLUDGE BACTERIA**

1. Mechanisms and Horizontal Transfer of Antibiotic Resistance

The rise of antibiotic resistance in microorganisms is one of the most significant

public health concerns within the past 50 years. In this "post-antibiotic era," genetic

manipulation, synthetic chemistry, and an enhanced search for natural antibiotic products

have become our best strategies for combating this developing clinical crisis. Many

studies focus on the ecology and evolution of resistance determinants both in the

environment and in the clinical setting, while others aim to characterize the molecular

mechanisms of resistance in hopes of integrating this knowledge with the development of

novel antibiotic strategies. Current case studies demonstrate the widespread distribution

of antibiotic resistance as well as the emergence of the "multi-resistance phenomenon,"

which describes bacteria that are resistant to a host of antibiotic therapies (257).

1.1. Origins and Evolution of Antibiotic Resistance Genes

Although past horizontal gene transfer of resistance between taxonomically

distant bacteria is largely responsible for generating a reservoir of resistance genes, this

interaction is difficult to reproduce *in vitro* for many reasons (8). Instead, recent studies

have focused on a "retrospective analysis" in which phylogenetic tools are used to verify

this hypothesis and track the origin of various resistances. A metagenomic survey of

deep-sea sediments near Papua New Guinea identified a TEM-type β-lactamase from a

sample estimated to be nearly 10,000 years old (250). Other studies have discovered the

presence of resistant bacteria in culture collections that were assembled before the

appearance of modern-day antibiotics: Smith identified resistance to sulfadiazine,

spectinomycin, and tetracycline in *E. coli* isolates collected before 1950 (246), and

Barlow demonstrated that *amp*C β-lactamse genes recovered from *Citrobacter freundii*

strains collected in the 1920s were as effective in conferring β-lactam resistance as were

plasmid-borne resistance genes from more recent clinical isolates (20). In addition,

plasmids found in Gram-negative bacteria that were isolated before the use of modern antibiotics are very similar to currently described plasmids (110).

A series of experiments by Connell and colleagues (51, 52) described the evolution of ribosome protection proteins (RPPs), which confer resistance to tetracycline. Their phylogenetic analyses revealed that clusters of *tet* genes have a monophyletic origin and underwent an early branching event that separated them from other classes of RPPs. Further studies confirmed the branching and diversification of eight clusters of RPPs well before the "antibiotic era" (7). With respect to macrolide resistance, the *erm*X gene found within one clade of bacteria appears to have evolved prior to similar resistance genes in antibiotic-producing microorganisms (8). One of the most characterized and clinically relevant phylogenetic reconstructions pertains to the genes that confer resistance to vancomycin. Most likely, the gene cluster evolved in a soil bacterium but was then acquired by *Bacillus circulans*, a bacterium that inhabits both environmental and clinical settings (140). Further transfer of the *van* genes into the enterococci was probably mediated by the transposon *Tn*1546, which contributed to the appearance of the gene cluster in clinical isolates such as *Staphylococcus aureus* and *Enterococcus faecalis* (98).

The phylogenetic reconstruction of resistance evolution, especially those with significant clinical implications, is key in determining the origin of various resistance genes. The next step in this process lies in identifying reservoirs of resistance and their distribution throughout natural and clinical environments. In addition, this knowledge may allow us to predict the transmission of resistance determinants, not only between bacterial species themselves but also between animals and humans.

Although antibiotic resistance in the clinical setting has gained much attention in recent years, the occurrence of resistance in natural environments (*i.e.*, soil, marine systems) may provide clues to the transmission of resistance genes from native, commensal bacteria to pathogenic bacteria across ecological gradients (296). A review by Nwosu (180) identifies four main ecosystems with regard to the transmission of antibiotic resistance: soil, water, animals, and humans. The various mechanisms of horizontal gene transfer both within and between these environments will be described in a later section, but the *modes* of resistance will be briefly discussed here.

Antibiotic resistance genes found on bacterial chromosomes are generally associated with a transposon, integron, or other insertion sequence the bacterium has acquired over time (165). Although the resistance gene(s) may have integrated into the host's chromosome, it is still possible (if not likely) that they may be transferred to other bacterial hosts. For example, a recent study identified a novel transposon that conferred erythromycin on various pneumococci that were previously sensitive to tetracycline (48). After a series of mating experiments, it was determined that this transposon could be successfully transferred from *Streptococcus pyogenes* and expressed in other pneumococcal species. Several additional studies have described integrons that confer resistance to a wide range of antibiotics to bacterial species of *Campylobacter, Salmonella, Acinetobacter,* and *Shigella,* all of which represent significant human pathogens (84, 169, 204, 224, 250).

A second well-characterized mode of antibiotic resistance is through plasmids, which have the ability to replicate and move between different bacteria. Plasmids often carry genes for resistance to antibiotics and other environmental stresses that may aid the

36

bacterial host in surviving adverse conditions (196). Many plasmids conferring antibiotic

resistance in activated sludge communities have been characterized. Droge and

colleagues (66) identified 12 unique plasmids from an activated sludge basin that carried

genes for amoxicillin, spectinomycin, chloramphenicol, and gentamicin resistance,

among others . All but one of these plasmids was multi-resistant, meaning they conferred

resistance to more than one antibiotic. In another study (227), a plasmid containing a

novel erythromycin resistance gene was isolated from activated sludge and was

determined to carry a transposon-like element that could potentially move the resistance

gene from one host to another. An additional study by the same group (262) also

characterized the multi-resistance plasmid pRSB101 (also isolated from activated

sludge), which harbors a 20 kb region of contiguous resistance genes and mobilizable

elements. This plasmid is of interest because it encodes a host of diverse resistance

mechanisms, including a novel multi-drug resistant transport system.

A third reservoir for antibiotic resistance lies in bacteriophage and their ability to

enter bacterial cells and integrate into the host genome (78). Like plasmids, the phage act

as vehicles by transferring resistance genes from one host to another but with the

additional ability to establish themselves within the bacterial chromosome in a process

known as transduction. Although there are reports of phage-mediated antibiotic

resistance between diverse bacteria (44, 60, 115, 156), there appears to be no discussion

of the transduction of antibiotic resistance in activated sludge microbial communities.

1.2. Mechanisms of Antibiotic Resistance Relevant to Activated Sludge Bacteria

Antibiotics are a structurally and functionally diverse group of natural and

synthetic molecules used to prevent and eliminate the spread of bacteria in various hosts.

They are often classified by their origin, mode of action, or chemical structure. This discussion focuses on the mechanisms of chloramphenicol, ampicillin, and kanamycin resistance, which differ in many ways with respect to biochemistry and genetics. These particular antibiotics were chosen for review because we have isolated genes that encode their resistances from activated sludge microbial communities.

Chloramphenicol was originally isolated from *Streptomyces venezuelae* and has been used for decades to treat infections caused by *Salmonella, Neisseria,* and *Streptococcus* (120). This antibiotic inhibits translation by binding to the 50S subunit of the bacterial ribosome, which inhibits the peptidyltransferase reaction and therefore prevents the formation of the polypeptide chain (76). Because of its mode of action, chloramphenicol has been useful for determining when proteins for cell division and chromosome replication are synthesized in the bacterial cell (280).

The most common mechanism of resistance depends on chloramphenicol acetyltransferase (CAT), a monomeric enzyme that chemically modifies the antibiotic and prevents its interaction with 23S rRNA at the 50S subunit (33). An –OH group at the C-3 position of chloramphenicol is necessary for this interaction, during which hydrogen bonds form between the antibiotic and a branched loop in the V domain of the 23S rRNA region. The CAT protein, with its cofactor acetyl-CoA, acetylates the C-3 –OH group. In a subsequent step, there is a non-enzymatic transfer of the acetyl group from C-3 to C-1, resulting in a mixture of 1- and 3-O-acetylchloramphenicol. The 1-O-acetylechloramphenicol is then available for another round of enzymatic acetylation at C-3. Both mono- and diacetylated chloramphenicol are inactive forms of the antibiotic (158).

Another mechanism of chloramphenicol resistance was discovered, ironically, in the bacterium from which the antibiotic was first isolated, *S. venezuelae* (112). Although enzymatic modification of the antibiotic took place, it was mediated by a 3-O-phosphotransferase, not an acetyltransferase. In this situation, the C-3 –OH group was phosphorylated, which prevented the formation of hydrogen bonds between chloramphenicol and the 23S rRNA. This discovery suggests that other enzymatic modification systems may exist which mediate resistance to chloramphenicol.

Because chloramphenicol resistance arises due to the expression of *cat* genes (resulting in CAT proteins) or other enzymatic modification systems and not a spontaneous mutation, its potential for horizontal gene transfer between bacteria is great. The *cat* genes are often associated with integrons and transposons, DNA fragments that have the ability to move from one bacterium to another and integrate themselves into the host genome (157). Another common vehicle for the spread of chloramphenicol resistance are plasmids, which are extra-chromosomal DNA elements known to harbor multiple resistance genes (76). These plasmids are easily transferred between bacteria, thus transferring chloramphenicol resistance with them.

The fact that resistance to chloramphenicol in a particular bacterium is dependent on only one *cat* gene makes it ideal for its detection in the laboratory. Previous studies describe various classes of *cat* genes, noting that there appears to be little homology between *cat* genes from Gram positive and Gram negative bacteria (76). Many sequence-based techniques have been used to isolate chloramphenicol resistance determinants from microbial communities, such as PCR (270) and DNA hybridization (16). In our study,

we have identified 7 unique chloramphenicol resistance determinants from activated sludge using a functional screening approach.

Ampicillin is a member of the β-lactam group, mostly broad-spectrum antibiotics that inhibit cell wall synthesis in both Gram positive and Gram negative bacteria (120). However, because of the extra layers of peptidoglycan cross-links that exist in Gram positive bacteria that offer a more promising target to β-lactams, these antibiotics are mainly used to treat infections caused by species of *Staphylococcus, Streptococcus,* and *Enterococcus*. Unfortunately, bacteria have developed widespread resistance to multiple β-lactams, including amoxicillin, penicillin, ampicillin, and most recently, methicillin (280).

During bacterial cell wall synthesis, a transpeptidase enzyme mediates the cross-linking of peptide chains that provide covalent connectivity and mechanical strength to peptidoglycan, thus protecting the cell from osmotic pressure. The transpeptidase belongs to a family of enzymes called active-site serine hydrolases, which carry a serine residue as a necessary nucleophile during the cross-linking reaction (158). When ampicillin interacts with the transpeptidase, the enzyme mistakes the antibiotic for another substrate to be incorporated into the peptidoglycan cross-link. Instead, the serine active site of the transpeptidase interacts with the β-lactam ring, generating an acylated transpeptidase intermediate that cannot continue the cross-linking cycle. The enzyme intermediate is resistant to hydrolysis (and therefore has a long half-life), so incorporation of the antibiotic into the cross-link results in eventual cell death (280). In addition, the excess of peptidoglycan precursors triggers the release of autolytic hydrolases, which digest existing peptidoglycan without the generation of new cross-links (140).

Many of the enzymes involved in the cross-linking reaction, including the transpeptidase, are also known as penicillin binding proteins (PBPs) because of their binding affinity for penicillin and its derivatives (220). PBPs are classified by molecular weight and are the most common mediators of β-lactam resistance in Gram positive bacteria. More specifically, Class 4 PBPs show high affinity for ampicillin and have become that antibiotic's primary target (158). Resistant bacteria possess altered PBPs that do not bind effectively to the antibiotic, thus limiting its incorporation into the peptidoglycan cross-link. However, this mechanism of resistance is most notably associated with methicillin-resistant *Staphylococcus aureus* and penicillin-resistant *Streptococcus pneumoniae*, among other Gram positive bacteria (280).

A second mechanism more applicable to ampicillin resistance is the bacterial expression of β-lactamases, enzymes that chemically destroy the antibiotic by hydrolyzing its β-lactam ring (158). To date, over 250 unique β-lactamases have been identified and organized into four classes (A through D) based on their chemical structure, molecular weight, and precise mechanism of action (159). The AmpC protein, which mediates ampicillin resistance in *E. coli,* is a Class C β-lactamase, as are all previously described β-lactamases that confer ampicillin resistance (158). Most β-lactamases are outer membrane or periplasmic proteins that inactivate their antibiotic targets before they can interact with PBPs, but there is evidence for a functional relationship between β-lactamases and PBPs that protects the cell with a combination of these resistance mechanisms (33).

The acquisition of PBP-mediated resistance is often through natural transformation or recombination events between neighboring bacteria (33). Sequence

and structural data support the hypothesis that PBPs gave rise to β-lactamases, which are more efficient in hydrolyzing and destroying β-lactam antibiotics (161). These enzymes are generally associated with the bacterial genome but are easily moved from one host to another by way of integrons, transposons, and other insertion elements (62, 140). Recent studies have also characterized ampicillin resistance mediated by plasmids, with the AmpC-encoded β-lactamase often part of a multiple resistance drug cassette (147, 151, 191). Because of the extensive use of ampicillin and other β-lactam antibiotics to treat a variety of bacterial infections, the emergence of β-lactam resistance is not surprising considering the widespread transfer of β-lactamase genes between bacterial species via transformation and conjugation (161). In addition, the fact that only one enzyme is necessary for conferring ampicillin resistance makes it an ideal candidate for horizontal transfer. The widespread use of ampicillin and other β-lactams make the horizontal transfer of their resistance genes a logical choice for study in the activated sludge environment.

Kanamycin belongs to the aminoglycoside class of antibiotics and is effective against both Gram-positive and -negative bacteria (120). Its modes of action involve the inhibition of translocation during translation and the mistranslation of proteins due to its interaction with the 30S ribosomal subunit (280). Because the antibiotic has low-level activity against plant cells, it is often used to select for transgenic plants that carry bacterial genes for resistance. The resistance gene is also a popular marker in molecular cloning experiments, as spontaneous mutations resulting in kanamycin resistance are very rare (248).

Kanamycin, like its close relative spectinomycin, has two modes of action. Both cause aberrant translation in bacterial cells, but two possible targets exist for the antibiotic. In the first mode of action, kanamycin prevents the translocation of the aminoacyl-tRNA from the A site to the P site of the ribosome by inhibiting the activity of translation elongation factor G (EF-G). Although the chemical nature of this interaction is still unknown, the result is a truncated polypeptide chain at the P site of the ribosome and a blocked A site that can no longer accept aminoacyl-tRNA molecules (138, 280).

In the second mode of action, kanamycin irreversibly binds to certain domains of the 16S rRNA region at the 30S ribosomal subunit, preventing translation of downstream mRNA (280). Various hydroxyl and amino groups within the antibiotic's cyclitol rings provide a high-affinity binding site for the interaction between the antibiotic and the ribosomal subunit. The presence of kanamycin physically blocks translation initiation because necessary initiation factors (especially IF-3) cannot mediate the successful binding of mRNA to the 30S subunit (158, 280).

The primary mechanism of resistance to aminoglycosides occurs through enzymatic inactivation in one of three ways: phosphorylation, acetylation, or nucleotidylation, although the first two modifications are most common with kanamycin resistance (295). In phosphorylation, an O-phosphotransferase moves a phosphate group from ADP to the antibiotic, disrupting the site necessary for binding to the 16S rRNA region. During acetylation, an O-acetyltransferase removes an acetyl group from its coenzyme, acetyl-CoA, and disrupts the binding site in a similar way. Nucloetidylation occurs in the same manner, except an O-nucleotidyltransferase transfers a nucleotide from an NTP molecule to the antibiotic binding site (138, 158). Of the three types of

modification, phosphorylation is the most energetically favorable, and therefore the most common (33).

However, other resistance mechanisms have been identified within specific groups of bacteria. For example, an efflux system that could transport kanamycin out of the cell was characterized in *Burkholderia cepaciae* and *B. pseudomallei* (171). In another study, resistance occurred via target modification, in which mutations in certain ribosomal proteins conferred protection from interactions with kanamycin (77). Decreased cell permeability is sometimes attributed to aminoglycoside resistance in *Pseudomonas aeruginosa*, which can alter the number of porin channels in its cell membrane to prevent kanamycin from entering the cell (200).

Although kanamycin resistance was first identified in association with the transposon *Tn*5 (248), many studies have demonstrated the movement of resistance genes via plasmids, integrons, and other conjugal elements. Analysis of various bacterial genomes, including *Mycobacterium tuberculosis, Bacillus subtilis,* and *Pseudomonas aeruginosa*, has revealed the presence of resistance genes, suggesting the widespread distribution of these genes among very diverse bacteria (235).

Phosphotransferase genes have been identified on plasmids and transposons in *Klebsiella pneumoniae* (250)*, Vibrio cholerae* (103), and *Proteus vulgaris* (175), as well as within the composite transposon *Tn*903 in *Corynebacterium striatum*, which supports the idea of horizontal transfer of these genes between both Gram-positive and -negative bacteria (265). Nucleotidyltransferase genes are often carried on small plasmids but have also been found on larger conjugative plasmids, the transposon *Tn*5405, and within the bacterial chromosome itself (76). Acetyltransferase genes are equally ubiquitous,

appearing on plasmids, class I integrons, and transposons (157). It has been hypothesized that these "resistance enzymes" evolved from proteins active in normal biosynthetic pathways, which could explain their presence in diverse groups of bacteria in a variety of genetic contexts (280).

Kanamycin resistance genes are ideal targets for our study because of their genetic characteristics. Like ampicillin, only one enzyme is needed for resistance, so the isolation of an intact pathway of multiple resistance determinants is not necessary to obtain functional clones. Again, our small-insert libraries should be sufficient to identify most genes that encode kanamycin resistance, although other mechanisms such as the efflux system may be overlooked because of their potentially large genetic loci. Another advantage to screening for kanamycin resistance is that rarely do spontaneous mutations give rise to resistance. Therefore, any clones exhibiting kanamycin resistance likely carry the actual resistance determinant and are not harboring mutations.

Because only one enzyme (and therefore only one or two genes) is usually necessary to confer chloramphenicol, ampicillin, or kanamycin resistance, we are confident that we have successfully surveyed the most likely resistance determinants in our activated sludge small-insert metagenomic libraries. Also, given the evidence in other studies as well as our own, the horizontal transfer of these resistance genes is of great interest, both clinically and ecologically.

1.3. Horizontal Transfer of Antibiotic Resistance Genes in Activated Sludge

Horizontal gene transfer (HGT) is any event which involves the movement of genetic material from one organism to another organism that is not its offspring. In recent years, HGT has become notable in microbiology, as it mediates the acquisition of genes in bacteria that do not naturally express them. Although any genetic material can

45

be exchanged by HGT, genes that encode antibiotic resistance are of great public health interest because of their ability to confer resistance in bacteria that were previously sensitive to clinical antibiotics. Several mechanisms of HGT exist, which are illustrated in Figure 1.2.

Transformation was the first mechanism of bacterial HGT to be discovered (261). In this process, DNA is exchanged between closely-related bacteria that are naturally "transformable." Although "transformation" has become a generic term for any situation in which a bacterium receives foreign DNA, it specifically describes the uptake of DNA by a bacterial cell without the aid of transfer elements that are often necessary in other HGT events. The most common transformation event *in vivo* occurs when a bacterial cell absorbs exogenous DNA directly through its cell membrane (271), as first observed in the classic experiment by Frederick Griffith in which an avirulent strain of *Streptococcus pneumoniae* gained the ability to produce a virulence capsule after exchanging DNA with dead, encapsulated *S. pneumoniae* cells (150). Since Griffith's discovery, many studies have described the role of transformation in the dissemination of antibiotic resistance genes between various bacteria. A recent study characterized the transfer of multiple antibiotic resistance genes between species of *Streptococcus* and *Gemella* without the help of mobilizable or conjugal elements (41). Another study examined the spread of antibiotic resistance in hospital-associated bacteria, again mediated by nothing more than physical contact between bacterial cells (69).

A second, more complex HGT mechanism is conjugation, in which a plasmid, transposon, integron, or other conjugal element is transferred from one bacterium to another by means of a sex pilus (261). In this situation, there is a defined donor (male)

46

and recipient (female) cell, as well as a possible helper cell that contributes mobilizable elements to aid in the stable transfer of DNA. Unlike transformation, conjugation is a common mechanism for genetic exchange between bacteria that may not be closely related. Because of this significant difference, conjugation as a means to transfer antibiotic resistance genes is of great interest. Conjugal plasmids are often a reservoir of resistance genes and have been isolated from various environments, including activated sludge (59, 66, 227, 257). In addition, the prevalence of other conjugative elements such as transposons and integrons has prompted studies into their role in the spread of resistance genes. Schubert and colleagues (229) identified a novel conjugative element in *E. coli* that encodes multiple resistance genes, an entire Type IV secretion pathway, mobilizable elements, and an integrase. The authors hypothesized that this element is the potential precursor for a pathogenicity island found in *Yersinia* species. Several studies have also described the occurrence of the HGT of antibiotic resistance by transposons, segments of DNA that have the ability to "jump" from host to host with the aid of the enzyme transposase (48, 140, 281).

One aspect of conjugation that contributes to its complexity is the variety of mechanisms that related to this type of HGT. For example, the conjugation system in Gram-negative bacteria resembles the anatomy of filamentous phage and requires the presence of a Type IV secretion system to assemble the sex pilus. In Gram-positive bacteria, conjugation is generally mediated by cell surface proteins that initiate mating-pair formation (78). Another level of conjugal complexity lies in the unpredictable need for a helper cell and exogenous mobilizable elements. As mentioned previously, some conjugation events occur independently of a "third party" bacterial cell, while others

require a complicated system of pili and mobilizable elements *in trans* (261). It has also been found that through conjugation, large sections of a bacterial chromosome (in addition to smaller elements) can also be transferred to the recipient cell (78).

Transduction, a third HGT mechanism, refers to the transfer of DNA by bacteriophage and was first described in 1952 by Zinder and Lederberg (305). Phage, which are the most abundant life forms on Earth (39), possess an enormous range of genomic diversity and an impressive repertoire of strategies for invading bacterial cells (38). Lytic phage replicate inside the bacterium and eventually lyse the cell, releasing phage, chromosomal, and plasmid (if present) DNA which can be absorbed by neighboring cells during transformation. Lysogenic phage (also called temperate phage) integrate into the host's genome or replicate as plasmids instead of immediately lysing the cell. Environmental stimuli can induce a switch from lysogeny to lysis, during which host cell DNA can be packaged and later injected into a new host (55). During these events, a variety of genes may be transferred from one host to another, including those for antibiotic resistance. However, because the transduced chromosomal DNA must be able to recombine with the chromosome of the recipient in order to survive, transduction is often restricted to bacteria in the same species (78). Literature describing the transduction of antibiotic resistance genes is limited, although a few studies describe this phenomenon with respect to gentamicin, ceftazidime, and erythromycin resistance (30, 31, 34, 198, 276). To our knowledge, transduction of antibiotic resistance genes in activated sludge communities has not been previously characterized.

2. Polyketide Synthase Pathways

Microbial metabolites have long been a primary source of therapeutically important drugs such as polyketides, .a diverse family of natural products found in bacteria, fungi, and plants (237). Examples of polyketides include the antibiotics erythromycin, tylosin, rifamycin, tetracyclines; immunosuppressants FK506 and rapamycin; and the antitumor compounds doxorubicin and mithramycin. (64). See Figure 1.3 for examples of common polyketides.

Polyketides were discovered serendipitously in 1893 by James Collie, who was attempting to determine the structure of dedydroacetic by boiling it with barium hydroxide and acid. He was surprised to find oricinol, an aromatic compound, as one of the products.   He proposed the mechanism of oricinol formation, suggesting that a polyketone intermediate could be formed from the γ-pyrone of dehydroacetic acid by the addition of water and ring opening. Collie further hypothesized that these polyketone intermediates might be generated and produced by living cells, but his peers did not readily accept this hypothesis. In 1917, Robert Robinson's findings re-affirmed Collie's original belief that polyphenols are produced from polyketones. Another scientist, Arthur Birch, revived the scientific community's interest in polyketides in the 1950s by demonstrating that polyketones could be generated from acetate units by repeated condensation reactions, which later became known as the Collie-Birch polyketide hypothesis (29), which proposes that polyketide biosynthesis is similar to fatty acid synthesis. Refer to Figure 1.4., which illustrates the complete biochemical mechanism of polyketide synthsis (253).

2.1. Biochemical Synthesis

Both fatty acids and polyketides are assembled from acetyl-coenzyme A or related acyl-coenzymes in a series of repeated head-to-tail linkages until the desired chain length is reached (226, 253). Experiments with isotope labeling further confirmed the Collie-Birch theory, verifying that the mechanisms of polyketide synthesis fall within the four types of reactions used to make fatty acids (181). Although the biosynthetic pathways are similar, there is one critical difference between fatty acid synthesis and polyketide synthase synthesis: fatty acid biosynthetic pathways reduce and dehydrate each resulting β-keto carbon, producing an inert hydrocarbon. Polyketide systems modify or omit some reactions, preserving some chemical reactivity along portions of the polyketide chain. Still other polyketide enzymes selectively promote internal cyclization and π-bond rearrangements to produce a number of structurally diverse products (15).

Polyketide synthases (PKSs), the pathways responsible for polyketide synthesis, can be divided into three groups: Type I, Type II, and Type III. Type I PKSs contain multifunctional proteins that are each responsible for a different active site for catalytic reactions in polyketide chain assembly (170). The actinorhodins, or Type II PKSs, are composed of individual proteins that perform one enzymatic activity to catalyze the formation of aromatic polyketides (170, 237). Type III PKSs, previously thought to be found only in plants, participate in the assembly of small aromatic compounds (170).

Type I PKSs are large multifunctional proteins organized into modules, each of which contains an individual enzyme responsible for the catalysis of polyketide chain elongation. Modular PKSs are exemplified by 6-deoxyerythromycin B synthases, which are responsible for the biosynthesis of reduced polyketides such as erythromycin A (237),

seen in Figure 1.3. Three open reading frames (ORFs), *eryAI*, *eryAII*, *eryAIII*, comprise the genes responsible for the synthesis of multienzyme polypeptides 1, 2, and 3, respectively. Each ORF is further organized into two modules, which each contain three domains to catalyze one cycle of chain extension: a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). Some modules carry variable domains, such as a ketoreductase (KR), a dehydratase (DH) or an enoyl reductase (ER), which are responsible for keto-group modification (26, 65).

The manipulation and development of PKS pathways offers vast opportunities for the isolation of novel compounds. The modular organization of Type I PKSs present diverse possibilities to determine if new metabolites can be discovered from various modular arrangements using a combination of genes from individual or multiple organisms. If these rearrangements are successful, this method would provide another source of chemical diversity in the quest for new antibiotics (226).

The structure of Type II PKS is comprised of enzymes similar to those of Type I PKSs: α-KS, ACP, chain length factor (CLF) and cyclases (64). These four subunits constitute the "minimal PKS" found in all Type II PKSs (15). The CLF controls the chain length of the polyketide, while the ACP assembles the polyketone chain from one unit of acetyl-CoA and seven units of chain-extending malonyl-CoA. The cyclases act in cyclization and aromatization of the resulting compounds (64, 253). More specifically, after decarboxylation of malonyl-ACP, the acetyl group is transferred to the α-KS active site where the first condensation step will occur, resulting in acetoacetyl-ACP. The ketoester is then transferred from the ACP back to α-KS where another cycle of chain extension can occur (253).

Type I and Type II PKSs differ in the formation of cyclic aromatic compounds that do not require the dehydration and reduction steps required in modular synthesis. Type II PKSs were actually discovered several years before Type I PKSs; however, a complete set of Type II PKS enzymes has not been found (226). The first genes encoding Type II PKSs were discovered by studying the genetics of actinorhodin biosynthesis in *Streptomyces coelicolor* and tetracenomycin biosynthesis in *Streptomyces glaucescens* (152, 174). *S. coelicolor* produces a blue-pigmented polyketide, actinorhodin (279), from which Type II PKSs get their name. For actinorhodin biosynthesis, an aromatic polyketide gene cluster encodes the PKS genes involved in cyclization, aromatization, and tailoring enzymes. Further sequence analysis of the PKS-encoding genes from the granaticin, tetracenomycin, and actinorhodin synthesis pathways in *Streptomyces violaceoruber* was used to confirm the basic design of Type II PKS systems (258, 259). Previous studies involving Type II PKS pathways have revealed novel products from new combinations of heterologous bacterial genes. The exact functions of many Type II enzymes have yet to be determined and could provide more insight into the potential for novel PKS metabolites from these pathways..

In 1999, a third type of PKS was identified: Type III PKSs are small homodimeric proteins that are similar to chalcone synthase (CHS) and stilbene synthase (STS) families that were thought to be found only in plants (228, 237). Their sequence homology and vague evolutionary history led some to believe that a number of bacteria had acquired the CHS-like genes via horizontal gene transfer from plants. However, the recent increase in complete genome sequence data has revealed only ~ 25% amino acid homology with CHS, confirming that Type III PKSs are indeed a novel system found in

bacteria (15). Type III PKSs differ from CHS and STS families in their starter enzymes, the number of acetyl additions they catalyze, and their mechanism of chain termination (15). This type of PKS generally uses a malonyl-CoA starter unit, continues through malonyl-CoA decarboxylation, polyketide chain elongation, and cyclization, yielding a chalcone or stilbene. There are currently 12 known Type III PKS pathways attributed to plants and three from bacteria. (15).

The first demonstration of the enzymatic production of an aromatic PKS product by a Type III polyketide synthase was demonstrated in the Chinese club moss, *Huperzia serrata*. This study determined that Type III PKS enzymes accept a larger starter substrate and possibly contain a larger starter-substrate binding site at the active site than chalcone synthase pathways found in other plants (283). Furthermore, the Type III enzyme RppA offers broad substrate specificity and allows for increased diversity in final end products, leading some to believe that RppA-like enzymes are involved in the biosynthesis of a wide variety of metabolites in various different hosts (80).

Currently, two-thirds of known bioactive polyketide natural products originate from actinomycetes; other microbial sources of polyketides include myxobacteria, cyanobacteria, and fungi (195). While the exact function of many polyketides is unknown, it is thought that in many cases they result as secondary metabolites to aid their hosts during competition for nutrients and other resources.   Polyketide synthases have several distinguishing characteristics that set them apart from other biosynthetic systems. They are quite large, usually 100 to 10,000 kDa, and are soluble cytosolic multienzyme systems that require no intracellular substructure or organelle to maintain activity. The possibilities of joining various purified protein components *in vitro* to yield novel PKS

activity makes these pathways ideal for expression in heterologus hosts (195). Heterologus expression can be employed to produce a desired protein in specific quantities and can be used to overcome limitations associated with manipulating the larger modular arrangement of PKS in native microorganisms. *Streptomyces coelicolor* and *E. coli* have been engineered in such a way to be able to express PKS pathways that might otherwise not be characterized. The macrocyclic core of the antibiotic erythromycin, 6-dEB), is a complex natural product produced by the soil bacterium *Saccharopolyspora erythraea* through a PKS reaction (26). A derivative of *Escherichia coli* has been genetically engineered to convert exogenous propionate into 6-dEB (211). The gene encoding dEB- 3, an enzyme proposed to catalyze the fifth and sixth condensentation cycle sin the assmelby of 6-dEB, has been engineered into a pT7-based expression system for over-expression in *Escherichia coli*. (211).

2.2. Distribution of Polyketide Synthases

Although the synthesis of polyketides is most often associated with the actinomycetes (195), PKS genes have also been identified in a diverse group of bacteria, including species of *Pseudomonas* (154, 205), *Mycobacterium* (126), *Stigmatella* (54), *Sorganium* (85, 86), and *Acidobacteria* (285). The widespread distribution of PKS genes is of great interest because additional reservoirs of polyketides likely exist but remain uncharacterized due to cultivation-based limitations and the targeting of specific phylogenetic groups.

2.3. Discovery of Novel Polyketides

The isolation of secondary metabolites produced by soil microorganisms has historically been an effective means of identifying novel compounds with antimicrobial

activity. Soil environments contain an impressive 10^{16} microbial cells per ton, while only an estimated 1% of these have been successfully cultivated in the laboratory (58, 108). While this small percentage of cultured soil microorganisms has already illustrated the phylogenetic diversity of these communities and have provided many important antibiotics to date, a potentially richer resource for novel antibiotic discovery lies with as-yet-uncultured bacteria (87, 94, 218). The application of a metagenomic approach to isolate large, contiguous regions of DNA from uncultured microorganisms remains a good strategy for accessing intact biosynthetic pathways that are involved in antimicrobial compound production. Previous studies have demonstrated the utility of screening environmental metagenomic libraries to identify bioactive molecules (54), proteolytic systems (23), and polyketide synthases (89, 167, 226, 286) .

3. Summary

The close physical interaction of bacteria within activated sludge provides an environment ideal for genetic exchange, more specifically the HGT of antibiotic resistance genes. Because this community contains a diverse array of bacteria, plasmids, and phage, it is logical to explore all three mechanisms of HGT with respect to antibiotic resistance. We hypothesize that the prokaryotic community structure of activated sludge can serve as a model for various mechanisms of HGT that mediate the spread of antibiotic resistance within environmental microbial communities.

As microorganisms undergo HGT and become more resistant to known antibiotics, the need for innovative pharmaceutical treatment will increase significantly. Evolutionary processes have resulted in an incredible diversity of chemical moieties with antibiotic activity that remains undiscovered. To access a greater degree of the

biosynthetic capacity of environmental microorganisms, we applied a metagenomic approach for the discovery of novel polyketides, combining recent advances in the isolation and cloning of HMW DNA from diverse soil microorganisms with both sequence-based and function-based screening of soil metagenomic libraries to identify recombinant clones either encoding or expressing the biosynthetic machinery for natural products with potential antimicrobial activity.

Soil   Symbionts   Plant   Marine

Environmental Sample

Recover microbial cells &
extract genomic DNA

DNA purification & size selection

Clone into a suitable vector

Metagenomic Library-construction

Large-insert cloning   Small-insert cloning

Transform into a suitable host & select for transformants

Metagenomic Library Screening

Function-based

- Gene expression
- Identify small molecules, proteins

Sequence-based

- Identify DNA sequences
- PCR, hybridization

Figure 1.1. Schematic of microbial metagenomic library construction and screening.

57

Donor cell

Recipient cell

Chromosome

Chromosome

Prophage

Specialized

Transduction

Generalized

Mobilizable plasmid

incX

Transposon

incY

Integron

Pilus

Conjugative plasmid

incY

Mobile gene cassettes

incZ
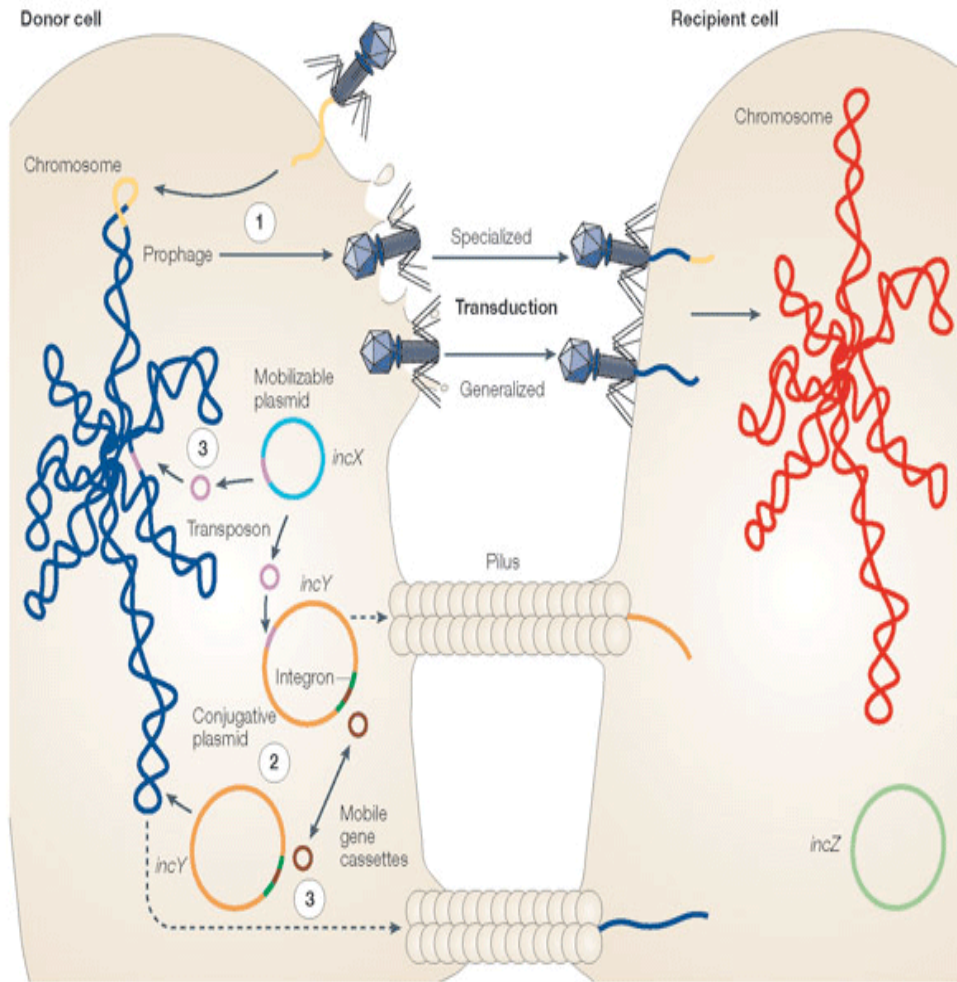
Copyright © 2005 Nature Publishing Group

Nature Reviews | Microbiology

Figure 1.2. Mechanisms of horizontal gene transfer among microorganisms (78).
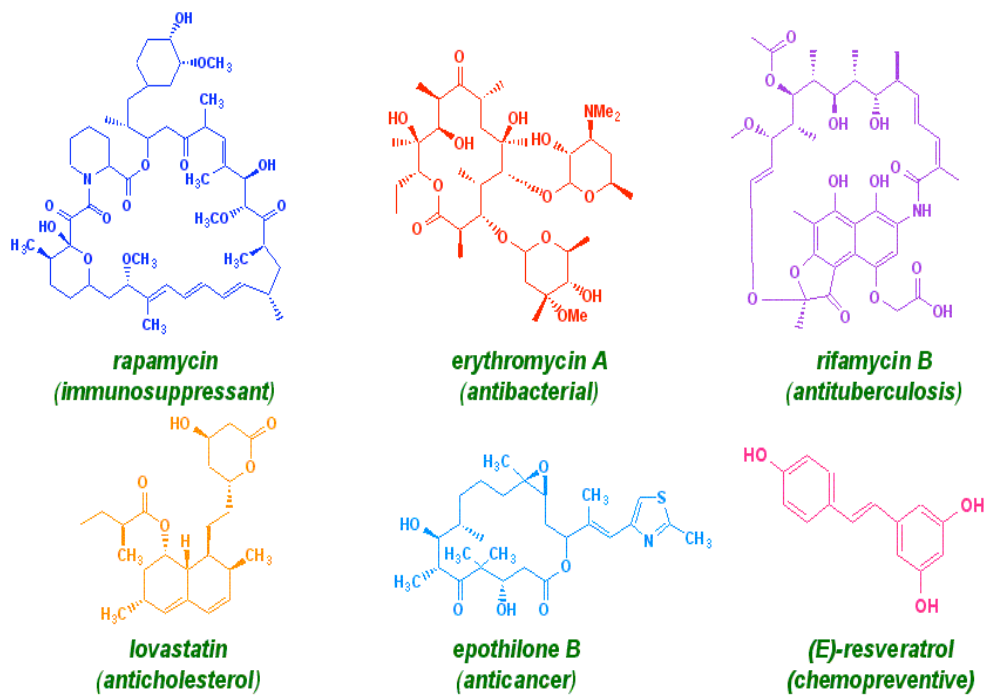
Figure 1.3. Structural diversity of polyketide compounds.
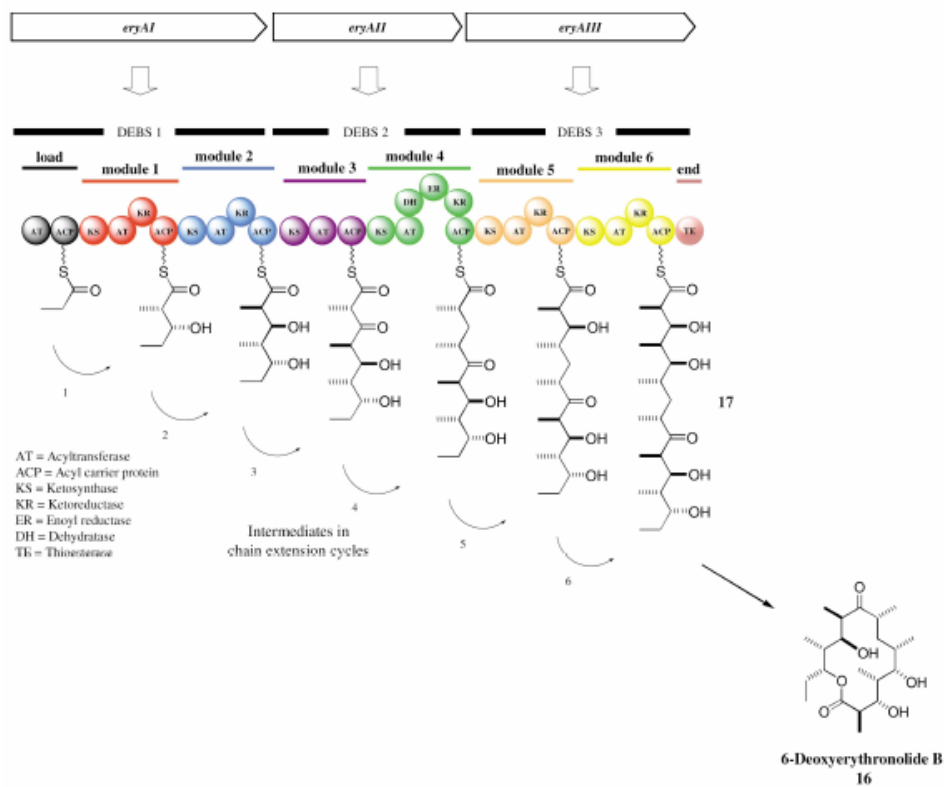
Figure 1.4. Biochemical mechanism of polyketide synthesis (253).

CHAPTER II

CHARACTERIZATION OF BACTERIAL AND VIRAL PHYLOGENETIC
DIVERSITY IN WASTEWATER COMMUNITIES USING BOTH CULTURE-BASED
AND METAGENOMIC APPROACHES

A. ABSTRACT

We evaluated the bacterial and viral diversity within an activated sludge (AS)
aeration basin, its influent (IN), and a laboratory-scale sequencing batch reactor (SBR)
using culture-dependent and culture-independent methods.   A total of 91 unique cultured
bacterial isolates, 28 cultured bacteriophages, 1,103 bacterial 16S rDNA clones, and
1,779 viral metagenomic clones were subjected to phylogenetic analyses.   The culture-
dependent survey of bacterial diversity indicated that the AS and SBR microbial
assemblages were dominated by *γ-Proteobacteria*, whereas the culture-independent
survey revealed the dominance of members from the *Bacteroidetes* division that
comprised more than half of the rDNA sequences obtained.   Among the cultured
bacterial isolates, a high frequency (19.7%) of bacterial co-isolation was observed, which
may be attributable to the intimate bacterial associations present within floc microbial
assemblages in AS and the SBR.   The bacterial culture collection was also used to
identify 28 unique bacteriophages that infected members of the *Actinobacteria*,

Firmicute, and γ-*Proteobacteria* divisions.   The most abundant viral metagenomic

sequences were used to query the cultured bacteriophages by PCR, revealing two genetic

loci that had representatives within both the viral metagenome and a cultured

bacteriophage and very low percent matches within the GenBank nr/nt database. These

observations support the premise that the vast majority of microorganisms in AS, IN, and

SBR communities are not readily cultivated under laboratory conditions and that culture-

dependent and –independent approaches have a limited overlap. Together, these

techniques can provide a more complete characterization of the phylogenetic and

functional diversity of bacteria and viruses in these environments.

B. INTRODUCTION

The activated sludge system in wastewater treatment plants is an important

biological process that has a significant impact on both environmental and public health.

As activated sludge and its influent harbor a functionally diverse microbial population, an

extensive phylogenetic survey of these communities is important to gain an

understanding of the complex microbial interactions at a community level. In addition,

the use of laboratory-scale bioreactors allows for more controlled studies of phylogeny

and metabolic activity over time and under various environmental conditions.

Many studies have explored the abundance of specific bacterial taxa within

activated sludge and bioreactors, revealing the ubiquity of members from the

Bacteroidetes and *β-Proteobacterial* divisions (53, 127, 128, 232, 288), as well as

bacteria with certain metabolic abilities (32, 107, 145, 183, 234, 250, 299, 302).

Phylogenetic surveys have targeted a wide range of genetic and physiological markers

including quinone profiles (102), polyamine patterns (14), functional genes such as *amoA*

(178, 186, 221, 250) and stable isotope probing of precursors utilized during various

metabolic cycles (88, 153, 179) in activated sludge and bioreactors. Perhaps the most

widely-used phylogenetic approach involves sequencing and classifying 16S rRNA genes

from both cultured isolates (119) and community rDNA clone libraries (145, 160, 242,

247, 250) and by detecting 16S rRNA genes using various *in situ* labeling methods (118,

155, 215). However, because each of these methods has potential advantages and biases,

a combination of these approaches remains the best approach to assess the phylogenetic

diversity of bacteria in activated sludge and bioreactor communities.

One aspect of activated sludge microbiology which remains relatively

uncharacterized is the occurrence of bacterial co-isolation, also referred to as co-

aggregation, in wastewater systems. Inducing co-isolation between diverse bacterial

species has been used to produce a specific metabolic activity (63) or to isolate obligate

intracellular bacteria (49), but to our knowledge, only one study has previously

characterized bacterial co-isolation in the activated sludge environment (115). However,

that study focused on the specific interaction of *Actinobacteria* and *Burkholderia* bacteria

within aerobic granules that could degrade phenol, rather than the frequency and

phylogenetic diversity of co-isolation. Because co-isolate associations may have

important implications for bacterial community structure and function, their

characterization could reveal potentially novel bacterial interactions.

In addition, bacteriophages play an active role in the ecology of natural

environments, such as influencing population dynamics (71) and transducing horizontal

gene transfer between diverse bacterial species. Historically, the focus of wastewater

viral studies has been on host/viral interactions and the use of phages as indicators of the

presence of specific bacterial hosts. Total viral counts in AS have varied based on isolation techniques and counting methods, but the highest counts were observed by Otawa *et al.* who observed $10^8 – 10^9$ virus-like particles (VLPs) per ml of AS sample (187). From culture-dependent studies on AS, phages have been isolated from 30%-60% of the cultured bacterial isolates from the same source (96, 243, 250). It has been observed in several studies that coliphages are seen in high titers in influent, but seem to decline in activated sludge (90, 298). Otawa *et al.* also used pulsed-field gel electrophoresis to examine viral communities from 14 full-scale wastewater treatment plants. They noted the range of size of viral genomes in AS to be predominately between 40 and 200 kb, with 40 to 70 kb-sized genomes the most frequent, and similar banding patterns were observed between samples, suggesting that the viral communities shared many viruses in common among the treatment facilities. In addition, varying temporal patterns in viral populations in a laboratory-scale reactor were observed with two distinct bands emerging and disappearing within a short time period (187).

Recently, there has been discussion of using phages to improve the efficiency of the wastewater treatment process, including pathogen control and foam control in the AS community (293). With the idea of using phages as a means of bacterial population control in AS, it is even more important to understand the viral community that exists naturally in AS. Although it has been established that phages exist in high numbers in AS (187), to our knowledge a phylogenetic survey of the AS viral community has yet to be conducted.

In this study, we present comparative phylogenetic analyses of bacterial and viral diversity in activated sludge, its influent, and a sequencing batch reactor using both

culture-dependent and –independent methods. The rationale for these surveys is three-fold: 1) to obtain biological materials and knowledge for future studies, especially in describing the genetic determinants of antibiotic resistance borne on plasmid, phage, and bacterial genomes; 2) to enable comparisons between culture-dependent and -independent phylogenetic data; and 3) to provide the foundation for additional surveys of bacteria, plasmids, and bacteriophages that impact the functional diversity of microbial communities in activated sludge, influent, and bioreactors. This study also demonstrates the value of combining both culture-based methods, which yield a resource for the characterization of biological activities and interactions, and culture-independent methods, which allow for the direct querying of microbial functional and phylogenetic diversity without the substantial bias of laboratory cultivation. A summary and comparison of the eight phylogenetic surveys conducted in this study can be found in Table 2.1.

C. MATERIALS AND METHODS

1. Sampling site and bioreactor operation. Samples were collected from the activated sludge (AS) aeration basin at the H.C. Morgan Water Pollution Control Facility (HCMWPCF) in Auburn, AL, and a laboratory-scale sequencing batch reactor (SBR) originally seeded from the aeration basin. The HCMWPCF is a dual train, conventional activated sludge system that includes grit removal, aeration, secondary clarification, and effluent disinfection. The SBR had a working volume of 4 liters and was operated on a continuous 6 h cycle consisting of 4 phases: effluent withdrawal (12 min), nutrient feed and deionized water addition (3 min), aeration (280 min), and settling (30 min). Aerobic conditions were maintained by cycling ambient air through a porous diffuser and were

verified by measuring dissolved oxygen levels (YSI Model 57 Oxygen Meter, YSI

Incorporated, Yellow Springs, OH) and redox potential (Model 200, Eutech Instruments,

Singapore) at regular intervals. The pH of the SBR was kept between 6.8 and 7.3 by

adding 0.1M HCl or 0.1M NaOH when appropriate. The SBR was maintained at 25°C.

The composition of the synthetic feed was (mg per COD/L): acetate (360), casamino

acids (20), yeast extract (< 1). The inorganic salts content was (as mg/L total influent

concentration): KCl (210), $MgCl_2$-$6H_2O$ (394), $MgSO_4$-$7H_2O$ (26), $CaCl_2$ (80), H_3BO_3

(0.11), $ZnSO_4$-$7H_2O$ (0.0.50), KI (0.027), $CuSO_4$-$5H_2O$ (0.11), $Co(NO3)_2$-$6H_2O$ (0.135),

$NaMoO_4$-$2H_2O$ (0.056), $MnSO_4$-H_2O (0.62), and $FeSO_4$-$7H_2O$ (0.55). The influent

phosphate (P) concentration was supplied as NaH_2PO_4-$2H_2O$ and was maintained at 8.0

mg P/L. The influent nitrogen (N) was supplied as NH_4Cl and was maintained at 40.3 mg

N/L.

2. Bacterial culture collection. Bacterial cultures were obtained by plating serial

dilutions (10^{-3} to 10^{-6} in 1X PBS, pH 7.2) of AS and SBR samples onto LB agar and a

synthetic wastewater (SWW) medium made from SBR nutrient and mineral stocks (207).

The plates were incubated for two weeks at 25°C, and colony forming units per ml were

determined for AS and SBR on both media at various time points. Colonies exhibiting a

unique morphotype were re-streaked for isolation onto their respective medium, and these

secondary cultures were incubated for 24 or 48 h at 25°C. Isolated colonies were

subcultured into LB broth and grown overnight at 37°C with shaking to generate glycerol

stocks for the bacterial cultured isolate collection.

3. Genomic DNA isolation and ribotyping of cultured isolates. Genomic DNA was

extracted from a 2 ml overnight culture of each bacterial isolate using the Promega

Wizard Genomic DNA extraction protocol for Gram positive bacteria (Promega, Madison, WI). The 16S rRNA gene from each isolate was amplified by PCR in a 50 µl reaction containing 1X GoGreen master mix (Promega, Madison, WI) , 20 nM each of universal bacterial primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-GYTACCTTGTTACGACTT-3') (135) and approximately 1µg of DNA template. The thermalcycling conditions included an initial denaturing step at 95°C for 1 min, followed by 40 cycles of denaturing (95°C for 30 sec), annealing (55°C for 15 sec), and extension (72°C for 2 min) with a final extension step at 72°C for 5 min. Each amplicon was digested with the restriction enzyme RsaI and electrophoresed through a 1X TAE 1.2% (w/v) agarose gel for restriction fragment length polymorphism (RFLP) analysis, which identified unique cultured isolates from the total culture collection. Unique 16S rDNA fragments were purified using the Wizard PCR Clean-up system (Promega) and sequenced at the Lucigen Corporation (Middleton, WI) using dideoxy chain termination sequencing reactions.

4. Resolution and identification of bacterial co-isolates. Mixed colony morphologies, RFLP analysis, and 16S rDNA sequence chromatograms indicated co-isolation of some bacteria in the culture collection. Gram staining, negative staining, and endospore staining confirmed the presence of more than one cell type in the co-aggregates.

To resolve bacterial co-isolates and identify the separate bacterial constituents, the mixed bacterial cultures were sequentially re-streaked multiple times on LB agar to obtain separate, isolated colonies. In several cases, the co-isolate pairs could not be physically resolved using this method, so their respective 16S rRNA genes were PCR amplified and then separated using denaturing gradient gel electrophoresis (DGGE)

67

according to the following methods: DNA was used as template in 50 μl PCR reactions containing 1X GoGreen master mix (Promega), 20 nM each of the universal primer set 338F-GC (5'ACTCCTACGGGAGGCAGCAG-3' plus a GC-clamp at the 5' end) and 518R (5'-ATTACCGCGGCTGCTGG-3') (176) to amplify 16S rDNA for DGGE under the following thermalcycling conditions: 95°C for 1 min (initial denaturing), 95°C for 30 sec, 55°C for 45 sec, 72°C for 1 min (30 cycles), and 72°C for 5 min (final extension). The amplicons from each time point were electrophoresed through an 8% (v/v) denaturing polyacrylamide gel at a 45%-65% denaturant gradient range for 18 h at 100V using the phorU DGGE system (Ingeny, Goes, The Netherlands). The DNA was visualized by low-wavelength UV light excitation after SYBR Gold staining, and individual rDNA amplicon bands were extracted from the gel and served as template for a secondary PCR using the 338F (without a GC clamp) and 518R primer set. PCR amplicons were purified as described above and sequenced using BigDye sequencing reactions at the Auburn University Genetic Analysis Laboratory.

5. Characterization of co-isolate AS 6. For one putative co-isolate (AS 6), DGGE data repeatedly yielded only one 16S rRNA gene sequence, although observations on LB agar and multiple stains suggested the presence of more than one bacterial morphotype. A panel of biochemical tests (indole, motility, methyl red, Voges-Proskauer, multiple sugar fermentations, H_2S production, phenylalanine deaminase, and lysine, arginine, and ornithine decarboxylase) was performed on AS 6 and the ATCC reference strain (*Aeromonas caviae,* ATCC 14486) of its top hit from the nr/nt database of GenBank.

In addition, fluorescence *in situ* hybridization was performed on AS 6 using the *Aeromonas*-specific probe AER66 (5'-CTACTTCCCGCTGCCGC-3') (118) labeled with

FITC and the universal bacterial probe EUB338 (5'-GCTGCCTCCCGTAGGAGT-3')
(5) labeled with Texas Red. An overnight broth culture of AS 6 was centrifuged at
13,000 x g for 5 min, the cell pellet was resuspended in 4% formaldehyde in 1X PBS, and
the suspension was incubated at room temperature for 3 h. The suspension was then
centrifuged again at 13,000 x g for 5 min and the fixative was discarded. The pellet was
sequentially washed at room temperature for 5 min in 50%, 80%, and 95% EtOH and
dried for 10 min in a SpeedVac (Eppendorf, Westbury, NY). Hybridization buffer
(20mM Tris-HCl, pH 8.0, 0.9M NaCl, 0.01% SDS, 40% formamide) was added to the
dried cell pellet, and the suspension was pre-hybridized at 65°C for 30 min. The
appropriate probe was then added to the suspension at a final concentration of 50 ng/μl,
and the hybridization was carried out at 65°C for 3 h. After hybridization, the cells were
washed 2X in 0.1X SSC at 37°C for 30 min, and the final cell pellet was resuspended in
0.1X SSC. To prevent photobleaching, SlowFade reagent (Invitrogen, Carlsbad, CA)
was added to each suspension, which was then viewed using a Zeiss Axiovert 200
inverted fluorescence microscope (Zeiss, Thornwood, NY) with a Nikon camera
workstation and Elements imaging software (Nikon, Melville, NY).

6. Construction of 16S rDNA clone libraries. 50 ml samples were obtained from the
AS aeration basin and the SBR in June 2006. Microbial community genomic DNA was
isolated using a modified MoBio UltraClean Genomic DNA extraction protocol (MoBio
Laboratories, Carlsbad, CA) and further purified by CTAB extraction and ethanol
precipitation. Briefly, 5M NaCl was added to the genomic DNA to a final concentration
of 0.7M NaCl. Then, a solution containing 10% hexadecyltrimethylammonium bromide
in 0.7M NaCl was added, and the mixture was vortexed and incubated at 65°C for 10

min. Genomic DNA was then extracted using isoamyl alchohol:chloroform (1:1 v/v),

precipitated with 100% ethanol, and resuspended in TE buffer (10mM Tris-HCl, 1mM

EDTA).   Samples were run in triplicate and stored at -80°C.   Purified genomic DNA

from AS and SBR was used as template in separate 50 μl PCR reactions to amplify

community 16S rDNA under the following conditions: Each 50 μL reaction contained

1X Go Green master mix (Promega), 20 nM each of primers 27F and 1492R, and

approximately 1 μg of DNA template. The thermalcycling conditions included an initial

denaturing step at 95°C for 1 min, followed by 40 cycles of denaturing (95°C for 30 sec),

annealing (55°C for 15 sec), and extension (72°C for 2 min) with a final extension step at

72°C for 5 min.   Amplicons were purified using the Wizard PCR Clean-up kit (Promega)

and ligated into the TOPO TA pCR2.1 vector (Invitrogen, Carlsbad, CA) for cloning of

PCR products. Ligations were transformed into electrocompetent *E. coli* DH5α TOP10

cells (Invitrogen) according to the manufacturer's instructions. Transformants were

robotically picked into 96-well format using the QPix2 colony-picker (Genetix,

Hampshire, United Kingdom) and submitted for sequencing at the SymBio Corporation

(Menlo Park, CA).

7. Phylogenetic analyses of bacterial 16S rDNA. Resulting 16S rDNA sequences

(≥600bp) from the 16S rDNA library clones and bacterial cultured isolates were trimmed

for quality using ChromasPro software (Technelysium Pty Ltd, Queensland, Australia)

and compared to the GenBank nr/nt database using the BLASTn algorithm. Top hits

from the database with ≥90% similarity to the submitted 16S rRNA gene sequences were

used to identify a division-level affiliation for each rDNA library clone and cultured

isolate.

16S rDNA sequences from the library clones were then aligned using ClustalX to produce a genetic distance matrix.   The resulting matrix was further analyzed by the DOTUR program (292) to generate rarefaction curves to estimate taxa richness. The clone library sequences from AS and SBR were submitted to DOTUR, which assigned the sequences to operational taxonomic units at genetic distances of 0.03 (species-level) and 0.15 (division-level).

8. Analysis of bacterial community structure over time. Samples were taken from the AS aeration basin and the SBR at five time points between June 2006 and January 2007. Microbial community genomic DNA was isolated and purified as described above. The resulting DNA was used as template to PCR amplify community 16S rDNA for DGGE analysis as described above.   After SYBR Gold staining, the DGGE gel was visualized and analyzed using Bionumerics 5.0 (Applied Maths, Austin, TX). The similarity between time points was calculated using the Pearson product-moment correlation coefficient, and clustering was performed with the unweighted pair-group method using arithmetic averages (UPGMA).

9. Identification of cultured bacteriophages. To determine the best method to extract viruses from AS, varying extraction methods were evaluated prior to sampling for viral DNA extraction. Previous studies demonstrated that a beef extract buffer was an effective elluent of viruses from activated sludge and solid sludge samples (1, 2, 168). Viral extraction trials were performed using varying ratios of beef extract buffer to AS to determine the best viral extraction method. Following the combination of AS and beef extract buffer, the samples were agitated at room temperature for 30 min.   Cellular debris was spun down by differential centrifugation at 10,000 x g for 30 min. The VLPs in the

resulting supernatant were filtered through a 0.45 μm PVDF filter, stained with SYBR Gold, and quantified using a Zeiss Axiovert 200 inverted fluorescence microscope with a Nikon 2 Camera with Elements software. Based on microscopic counts, it was determined that 10% beef extract buffer at equal volume or greater ratio with AS provided the highest VLP yield (data not shown). Based on the results of these trials, virions were extracted from AS by incubating 200 ml of AS sample with 200 ml of 10% beef extract buffer for 18 h at room temperature with continual mixing in a hybridization bottle.

Subsequently, to prepare viral enrichments for each of the bacterial isolates from the culture collection were grown in both LB broth and LB broth supplemented with 1 mM $CaCl_2$ and 1 mM $MgCl_2$ at room temperature with shaking until log phase growth was achieved. The cultures were then incubated with approximately 1.09×10^6 VLPs extracted from AS, and the incubation was continued under the same conditions overnight. Following incubation, the culture solutions were evenly divided, and half the culture was chloroform-lysed, while the other half of the culture was stored in a 7% DMSO solution at -80°C. Spot tests with both chloroform-lysed and non-lysed enrichment cultures were performed on all of the cultures to observe plaque formation. Plaque-positive samples were then purified by serial soft agar isolation, and phages that were plaque-purified were collected for DNA extraction.

To extract viral DNA, the isolated plaques were treated with 100 μL of benzonase (250 U/μl final concentration) at 37˚C overnight to degrade residual bacterial chromosomal DNA. Following benzonase inactivation (addition of 10 mM EDTA and heating at 70°C for 10 min), proteinase K (1mg/ml final concentration) and 1% sodium

dodecyl sulfate were added, and the mixture was incubated at 37˚C for 2 h to degrade viral protein coats. Proteins were removed by phenol:cholorfom extraction and DNA was recovered by isopropanol precipitation.

10. Viral metagenomic library construction and sequencing. IN and AS samples for metagenomic library construction were combined with an equal volume of 10% beef extract and agitated for 18 h at room temperature, then spun at 10,000 x g for 30 min. The supernatant was filtered through a 0.45 μm PVDF filter and precipitated with polyethylene glycol (PEG) at 4˚C overnight. The overnight PEG solution was spun at 7,000 x g for 45 min, and the resulting viral pellet was resuspended in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl (pH 7.5)). The resuspended solution was spun again (12,000 x g for 10 min) and the supernatant was loaded onto a cesium chloride (CsCl) step gradient and subjected to ultracentrifugation at 100,000 x g for 2 h. The viral band evident at the 1.3 mg/ml zone was collected and used for DNA extraction and direct counts. Viral DNA was extracted and purified from this fraction as described above.

Random shotgun metagenomic libraries of the viral communities from IN and AS were prepared at the Lucigen Corporation (37, 288). Viral DNA isolated from IN and AS was fragmented to approximately 1 to 3 kb fragments and end-repaired to blunt-end each dsDNA molecule. The dsDNA was then ligated to an adaptor (20 bp dsDNA sequence), and then primers specific to the adaptor sequence were used to PCR amplify the dsDNA fragments, using only 15 rounds of amplification. The viral amplicons were then purified over a silica column and ligated into the pSMART-LCKan vector, a low-copy cloning vector containing transcriptional terminators flanking the cloning site to prevent

transcription of viral genes (and potential toxicity to the *E. coli* host cell) resulting from vector promoters (37, 288). The ligation was transformed into electrocompetent *E. coli* strain DH10B, and a glycerol stock was prepared of the transformation mixture for storage at -80°C. The transformation mix was shipped to Auburn University, where the number of transformants per ml was determined. Transformants were plated onto LB agar containing kanamycin (30 μg/ml) and were picked into 96-well format using the QPix2 colony picking robot (Genetix, Hampshire, United Kingdom). After incubation at 37°C overnight with shaking, sterile 50% glycerol was added to each well (final concentration, 15%) and plates were frozen at -80°C. Seven 96-well plates (672 clones) from IN and 13 96-well plates (1,248 clones) from AS were sent to the Lucigen Corp. (Middlteon, WI) for sequencing.

11. Sequence analysis of viral metagenomic library clones. DNA sequences were screened for removal of vector and linker sequences, and minimum base quality using ChromasPro and Sequencher software packages (Technelysium Pty Ltd; Gene Codes). Following removal of nucleotide bases with poor quality reads, there were 1161 quality sequences from AS (0.76 Mbp) and 618 quality sequences from IN (0.41Mbp). For taxonomy assignments, sequences were compared against the GenBank nr/nt and env databases using the tBLASTx algorithm. The top five hits were used for analysis, and only matches with an E-value of ≤0.001 were considered significant. Significant GenBank database hits were categorized as either viral, bacterial, archeal, eukaryotic, or a mobile genetic element. In cases where multiple significant hits were observed for a single query sequence, the sequence was preferentially classified as viral if the hit occurred within the top five hits (37). Mobile elements consisted of plasmids,

transposons, transposases, pathogenicity islands, and insertion sequences.   Significant

hits to phages were further classified into phage families according to The International

Committee on Taxonomy of Viruses classification system. Hits to bacterial genomes

were further compared to a list of known prophages (http://phage-

finder.sourceforge.net/). Sequences predicted to be located on the genome of a known

prophage were categorized in the viral category.

12. Estimates of viral diversity and community structure. The online PHACCS tool

(http://biome.sdsu.edu/phaccs/) was used for assessing viral community diversity as

described by Angly *et al.* (9) The exponential rank abundance form was predicted to be

the best fit for the AS viral community, and the lognormal rank abundance form was

predicted by PHACCS to be the best fit for the IN viral community.

13. PCR amplification of abundant metagenomic loci. Contig assembly of viral

metagenomic sequences was performed using assembly criteria of 80% identity with at

least a 20 bp overlap in order to assemble sequences from the most abundant viral types

(37). These contigs were then analyzed by BLASTx comparisons to GenBank nr/nt and

env databases, and contigs with significant hits to bacterial taxa that were also present

within the AS bacterial culture collection were selected. Primers were designed to target

regions of the metagenomic contigs that were 100% identical with homologs in

GenBank. Touchdown PCR was performed using each metagenomic contig-specifc

primer set on all cultured phage DNAs under the following conditions: 95°C for 1 min

(initial denaturing), 95°C for 1 min, step-wise annealing cycles of 65°C to 50°C for 30

sec each, and 72°C for 1 min (16 cycles) followed by 30 cycles of 95°C for 1 min, 50°C

for 30 sec, and 72°C for 1 min , followed by a final extension step of 72°C for 5 min.

PCR amplicons were analyzed by gel electrophoresis, and amplicons was column purified using the Wizard PCR Clean-up kit (Promega). PCR products were then sequenced at the Auburn University Genetic Analysis Laboratory using both forward and reverse primer reactions to obtain a consensus sequence.

14. Metabolic profiling of viral metagenomic DNA. Sequences from the IN and AS viral metagenomic library clones were analyzed by the MetaGene annotator, which identified 933 ORFs from the IN library and 1815 ORFs from the AS library. The ORF sequences were then compared to those found in the SEED platform (http://www.theseed.org) and organized into predicted functional categories based on the subsystems approach described by Overbeek *et al*. (189).

15. Nucleotide accession numbers. All 16S rDNA sequences from the cultured isolates and bacterial library clones were submitted to GenBank under the accession numbers GU002706 to GU003868. The viral metagenomic DNA sequences were registered as part of the Wastewater Viral Metagenome project at GenBank (Trace Archives) and are under TI reference numbers 2251203077 to 2251204802.

D. RESULTS

1. Phylogenetic classification of cultured bacterial isolates. The bacterial culture collection contained 58 unique cultures from AS and 18 from SBR. Resolution of bacterial co-isolate pairs (14 from AS, one from SBR, see below) yielded a total of 91 bacterial isolates from which at least partial 16S rRNA gene sequences ($\geq$600bp) were obtained. Based on sequence homology with other 16S rRNA genes in the GenBank nr/nt database, the AS community was dominated by representatives of the subdivision *γ-Proteobacteria*, followed by members of the *Firmicute* division, whereas the SBR

76

community showed a nearly even representation of members from these divisions (Fig. 2.1A). The *β-Proteobacteria* and *Bacteroidetes* divisions were present in AS but were not represented in the SBR community in these culture-dependent surveys.

2. Resolution of bacterial co-isolates. Sequential re-streaking and DGGE were used to identify the separate bacterial taxa that were co-isolated in 14 bacterial cultures from AS (24%) and one from SBR (5%). Sequences of the 16S rDNA amplicons separated by DGGE were added to the culture-dependent survey so that each distinct bacterial taxa was included in the respective AS and SBR surveys. The phylogeny of each bacterial co-isolate pair is given in Table 2.2, which shows that all but three pairs contain bacteria from two different bacterial phyla. The most common bacterial co-isolates involved taxa from the *Firmicute* and γ-*Proteobacteria* divisions, which comprised 60% of the total co-isolate pairs identified.

3. Identification of a pleiomorphic *Aeromonas* isolate. One cultured isolate, AS 6, first appeared to be a co-isolate pair due to its morphological characteristics observed on LB agar and after Gram staining and negative staining (Fig. 2.2). However, separated 16S rDNA sequences obtained from the DGGE analysis repeatedly showed significant homology (99%) with only one bacterial species, *Aeromonas caviae*. Subsequent biochemical tests of AS 6 did not provide a diagnostic profile consistent with that of *A. caviae*, although FISH probing confirmed that only *Aeromonas* 16S rDNA is present in AS 6 (data not shown). The biochemical profile of AS 6 changed dramatically over time (24 h vs. 2 weeks), suggesting that in addition to morphological changes, this culture may also be experiencing metabolic variations that have been previously uncharacterized. The morphological and biochemical evidence suggests that AS 6 represents a pleiomorphic

Aeromonas species that is structurally and functionally disparate from its top hit in the GenBank nr/nt database.

4. Culture-independent phylogeny of bacterial 16S rDNA library clones. Partial 16S rRNA genes (≥600bp) from 829 (AS) and 274 (SBR) library clones were analyzed using the BLASTn algorithm and assigned to bacterial divisions based on sequence homology with 16S rRNA genes in the GenBank nr/nt database. Figure 2.3 shows the relative abundance of representatives from over 10 bacterial divisions in AS and 5 bacterial divisions in SBR. Although very low numbers of AS sequences were affiliated with bacterial divisions such as *Nitrospira, Acidobacteria*, and *Planctomycetes*, the presence of bacteria from these divisions, along with the *α-* and *β-Proteobacteria*, account for a two-fold increase in the number of divisions represented in AS when compared to SBR. Not surprisingly, the SBR community appears to be a subset of the AS community, as the bioreactor was originally seeded from the AS aeration basin. Members of the *Bacteroidetes* division dominated both communities, comprising 51.3% of the AS library and 67% of the SBR library, followed by nearly equal representation from members of the *γ-Proteobacteria* division in both communities (25.6% in AS, 26.9% in SBR).

The relationship between the total number of sequences and the number of unique operational taxonomic units (OTU, defined as "species-level" at 97% identity and "division-level" at 85% identity) for the culture-independent 16S rRNA survey are shown in Figure 2.4. In the AS community, the species-level curve continues to increase in a nearly linear fashion, even after the analysis of nearly 900 clone sequences, but the division-level curve appears to level off slightly (Figure 4A). Although the number of sequences analyzed from the SBR was three-fold less than those analyzed from AS, the

slope of its division-level curve has also begun to approach zero (Figure 4B). Like the AS community, the species-level richness of the SBR has not yet been fully sampled.

5. Assessment of AS and SBR bacterial community diversity changes over time.

The dendrogram in Figure 2.5 compares the ribotype richness patterns between AS and SBR bacterial communities at five time points. The greatest ribotype richness appeared in AS at the second and third time point (June 27, 2006 and July 25, 2006, respectively), and the least richness appeared in SBR at the final time point (January 22, 2007). At all but one time point (June 20, 2006), a greater level of taxa richness was observed for AS when compared to SBR. The UPGMA analysis of the banding patterns at each time divided the samples into two distinct groups, one containing AS and SBR from the July, October, and January time points, and the other containing AS and SBR from the June time points. This separation indicates a temporal shift in the level of ribotype richness and composition of the bacterial communities in both AS and SBR. In addition, many of the AS and SBR samples at various time points shared common bands, suggesting that a set of bacterial taxa persist in both communities, often at similar abundances (indicated by the intensity of the bands seen in Fig. 2.5). However, other samples such AS (10/12/06) appear to harbor several unique taxa at that time point, even though its ribotype richness estimate is one of the lowest among all of the samples.

6. Isolation of cultured bacteriophages. Viral enrichments performed on all of the bacterial isolates identified 28 bacterial taxa from which bacteriophages were isolated, 24 from AS and 4 from SBR. Of these 28 bacteriophage isolates, only 13 bacteriophages were capable of being serially purified from isolated plaques, which suggests the other 15 bacteriophages may be lysogenic or thermosensitive, as observed by Hantula *et al.* (96).

Seventy-five percent of the AS and SBR bacteriophages were isolated from *γ-Proteobacterial* hosts, while the remainder of the phages were isolated from *Firmicute* (AS and SBR) and *Actinobacterial* (AS only) hosts (data not shown). Interestingly, it was not necessary to use a viral enrichment to identify one of the cultured phages, which was determined to be a pseudolysogenic phage from a *Brevibacterium* spp. bacterial isolate.

7. Phylogenetic analysis of viral metagenomic library clones. Following library construction and sequencing, the quality sequence reads from AS (n=1161) and IN (n=618) were compared to multiple GenBank databases. Sequences were categorized as: 1) known, 2) unknown, or 3) novel based on comparison to the non-redundant and environmental nucleotide databases by tBLASTx analysis as described previously (294). A sequence with a significant hit to the nr/nt database was considered known. A sequence with no significant hit to the nr/nt database and a significant hit to the env/nt database was considered as unknown. A sequence with no significant hit to either database was considered novel. For the AS library, 55% of the sequences were known (n=694), 13% were unknown (n=97), and 32% were novel (n=370). For the IN library, 74% of the sequences were known (n=457), 8% were unknown (n=47), and 18% were novel (n=114).

The known hits were further categorized into taxonomic groups. Taxonomic assignments were made based on significant hits to the nr/nt database using a tBLASTx comparison. Of the 694 known hits for the AS library, 389 (61%) were similar to eubacterial genomes, 191 (22%) were similar to viral genomes, 57 (9%) were similar to mobile genetic elements, 55 (8%) were similar to eukaryotic genomes and 2 (<1%) were

similar to archaeal genomes (Fig. 2.6A). Of the 457 known hits for the IN library, 309 (68%) were similar to bacterial genomes, 76 (17%) were similar to mobile genetic elements, 66 (14%) were similar to viral genomes, 6 (1%) were similar to eukaryotic genomes. Of the 191 known sequences with homology to viral genomes, 95% were homologous to bacteriophage sequences, within the viral families Myoviridae (40.3%), Siphoviridae (31.9%), Podoviridae (25.6%), and to unclassified phage (2.2%), which is illustrated in Figure 2.6B. Ninety-five percent of all viral genome hits found in both libraries were similar to DNA sequences annotated as bacteriophage in a GenBank database.

Eleven different bacterial divisions were represented in the 389 AS eubacterial sequences. The phylogenetic breakdown of the AS bacterial hits is illustrated in Figure 2.6C. The *γ-Proteobacteria* were the most dominant bacterial division represented, comprising 26% of the total bacterial sequences, while the α- and *β-Proteobacteria* and *Firmicute* divisions are the second- and third-most dominant taxa present, respectively. Interestingly, the phylogenetic distribution of these eubacterial hits does not conform to either our culture-dependent or -independent surveys. For example, only two sequences (0.5%) from the AS bacteriophage metagenomic library were highly similar to *Bacteroidetes* DNA found in the nr/nt and env databases; in contrast, 51.3% of the sequences from the AS bacterial rDNA library were highly similar to the *Bacteroidetes* division. In addition, three of the bacterial phyla predicted to be present within the AS phage library (*Deinococcus, Cyanobacteria*, and *Spirochetes*) were totally absent from each of the bacterial phylogenetic surveys.

Although the IN library contained a slightly higher percentage of hits to bacterial genomes, only 10 different bacterial divisions were represented (data not shown). Similarly to the AS library, the *γ-Proteobacteria* were the most dominant division

present, representing 58% of the bacterial hits. The *α-* and *β-Proteobacteria* were again

the second- and third-most dominant divisions in this library. We observed the same

disparity between the bacterial division relative abundance patterns of the IN phage

library and the bacterial phylogenetic surveys.

8. Comparison of cultured phage and phage metagenomic DNA sequences. To

identify cultured phage genomes containing gene homologs identified from the viral

metagenomic library, assembled contig data from the AS metagenomic library was

analyzed. Contig assembly at 80% mismatch and 20 bp overlap resulted in 152 different

contigs with at least two sequences per contig. Contigs were then compared to the

GenBank nr/nt database using the BLASTx algorithm and sorted based on similarities to

the cultured bacterial isolate phylogeny. Four viral metagenomic contigs showed high

similarity matches to the following bacterial genera that were also represented within the

bacterial culture collection: *Citrobacter* spp., *Aeromonas* spp., *Acinetobacter* spp., and

Klebsiella spp. Primers were designed to PCR amplify genes identified from viral

metagenomic sequences in order to identify any homologs within cultured phage

genomes. Using this strategy three gene sequences were PCR-amplified, one from

cultured phage 47 (contig 30) and two from cultured phages 5 and 53 (contig 110), and

the PCR products were sequenced in both the forward and reverse directions. A

phylogenetic analysis was conducted to determine the phylogenetic relationships between

the cultured phage sequences, the metagenomic sequences, and related sequences in

GenBank (Fig. 2.7). In every case, the viral metagenomic sequences were more closely

related to the cultured phage sequences than to any other sequence in the GenBank

database. The phylogenetic affiliation between the culture-independent viral sequences

and the culture-dependent viral sequences had high bootstrap support, with >50% values for each clade.

9. Viral metagenomic library diversity estimates. Since there is no universally conserved set of genetic loci among viruses, assessing the diversity and viral species abundance based on metagenomic library sequence data is challenging. One method of estimating viral genotype diversity is by contig assembly, which can be influenced by a number of parameters. Breitbart *et al.*(37) determined that using the contig assembly parameters of 98% identity with a 20 bp overlap enabled differentiation of T3 and T7 coliphages. Contig assembly using these parameters was performed on both the AS and IN phage metagenomic library sequences. The contig spectra obtained from both libraries (AS: 892 113 32 5 2 0 0 0 0 0 0 and IN: 522 47 13 2 1 0 0 0 0 0 0 0) were then processed with the PHACCS tool (http://biome.sdsu.edu/phaccs/) (9). The exponential and the niche-preemption rank-abundance forms were the best fit for the AS library, while the lognormal rank-abundance form was the best fit for the IN library (data not shown). Based on the exponential form, the AS community contained 10,001 unique predicted viral genotypes with the most abundant virus representing 2.27% of the community. Based on the lognormal form, the IN community contained 160 unique predicted viral genotypes with the most abundant virus representing 7% of the community. For direct comparison purposes, the exponential form applied to the IN library contig spectra yielded a nine point increase in error over the lognormal form but predicted 10,001 different genotypes, with the most abundant virus representing 3.4% of the community.

10. Metabolic profiling of viral metagenomic DNA sequences. Predicted metabolic profiles for the IN and AS viral metagenomic libraries using the SEED platform are shown in Figure 2.8. In both libraries, the most dominant functional categories were those involved with carbohydrate, amino acid, and protein metabolism. Although the overall profiles for IN and AS are similar, there are some notable differences between the two libraries. For example, ORFs predicted to encode carbohydrate metabolism activities (which was the most dominant category for both libraries) accounted for over 40% of ORFs in the IN library but only 25% of the AS library, indicating different relative abundance patterns may exist for some metabolic activities in these two environments. Also, no ORFs with putative involvement in potassium metabolism were found in the IN library, although a small number (<5%) were present in the AS library. Other dominant categories found in both libraries included aromatic compound metabolism, virulence, and fatty acid metabolism. Considering the lack of an exhaustive survey of AS viral metagenomes, no conclusions may be made regarding changes in the relative dominance of each subsystem, but overall there is a striking similarity in the dominant metabolic systems distribution in an AS viral metagenome in comparison with viral metagenomes from other environments (64).

11. Comparison of bacterial and viral phylogenetic affiliations. Because bacterial community structure is expected to be the primary determinant of viral diversity within an environment, the relative dominance of bacterial divisions represented in both the bacterial and viral phylogenetic surveys is compared in Figure 2.9. The culture-independent 16S rDNA clone library revealed more than a three-fold increase in the number of represented bacterial divisions (n=10) compared to that observed from the

culture collection (n=3), and a significant shift in the relative dominance of each phylum was observed between the two surveys. Although *Proteobacteria* taxa dominated the culture-dependent survey (58.6% of analyzed sequences), the percentage of sequences representing this division in the culture-independent survey dropped to 25.6%. Similarly, *Firmicute* representation decreased by almost 10-fold in the culture-independent survey (35.7% to 3.6%) when compared to the culture-dependent study. In contrast, the *Bacteroidetes* taxa, which were only minimally represented in the culture-dependent survey, comprised over half (51.3%) of the bacterial taxa identified in 16S rDNA metagenomic library. The minimal overlap between data obtained from these culture-dependent and –independent analyses reflects the strong cultivation bias that is common to virtually all environmental studies.

Viral enrichments performed on all of the bacterial isolates identified 24 bacterial hosts from which phage could be isolated. Seventy-five percent of the phages were isolated from *γ-Proteobacterial* hosts, while the remainder of the phages was isolated from *Firmicute* (8.3%) or *Actinobacteria* (16.7%) hosts. The 389 sequences from the viral metagenomic library that were most similar to eubacterial genomes were also included in this taxonomic classification, as well as the predicted hosts of the 187 viral sequences with significant homology to bacteriophage DNA.

Although both viral surveys demonstrate similar phylogenetic profiles, the distribution of the bacterial hits from the viral metagenomic library does not conform to that of either bacterial survey. For example, only two sequences (0.5%) from the viral metagenomic library (0.5%, bacterial hits; 3.1%, phage host hits) were highly similar to *Bacteroidetes* DNA found in the nr/nt and env databases; in contrast, 51.3% of the

sequences from the bacterial rDNA library were highly similar to the *Bacteroidetes* division. In addition, four of the bacterial phyla predicted to be present within the viral metagenomic library (*Deinococcus, Cyanobacteria, Chloroflexi*, and *Spirochetes*) were absent from both of the bacterial phylogenetic surveys. The *Proteobacteria* dominated every survey except for the bacterial culture-independent study, in which the *Bacteroidetes* division was most represented. The phylogenetic distribution of the predicted phage host sequences most closely resembles that of the bacterial hits from the viral metagenomic library, with the *Proteobacteria, Firmicute,* and *Actinobacteria* taxa dominating both surveys. Other bacterial divisions with < 2% representation include the *Chlamydiae* (0.1%, Bacterial, CI), Nitrospira (0.2%, Bacterial, CI), *Planctomycetes* (0.7%, Bacterial, CI), *Cyanobacteria* (1.6%, Viral, CI), *Spirochetes* (0.7%, Viral, CI), *Chloroflexi* (0.5%, Viral CI) and *Deinococcus* (0.5%, Viral, CI).

E. DISCUSSION

The combined use of culture-dependent and –independent methods to characterize bacterial and viral community structure in activated sludge allowed a more extensive phylogenetic census of this environment than has been previously reported. Observations made from the culture-dependent bacterial survey (e.g., co-isolation, psuedolysogeny) would have gone unnoticed in a culture-independent survey and would not have permitted the isolation of cultured bacteriophages. While many biologically interesting phenomena and interactions may be determined from such culture collections, it is also clear that the culture-independent surveys in this study revealed substantially higher phylogenetic diversity for the bacterial and viral communities compared to culture-based methods.

Not surprisingly, both culture-dependent and -independent bacterial phylogenetic surveys showed that the SBR community (which was originally seeded from the AS aeration basin) was a subset of the AS community (Figs. 2.1, 2.3). In both surveys, the number of bacterial divisions represented in AS was much greater than those represented in SBR. A comparison of the culture-dependent and –independent surveys of both communities revealed both the appearance of additional bacterial divisions in the culture-independent study, as well as a significant shift in the relative abundance of bacterial divisions represented in both AS and SBR. For example, the culture-independent AS survey revealed representatives from 10 different bacterial divisions, whereas the culture-dependent AS survey identified bacteria from only five divisions. In addition, while the *γ-Proteobacteria* dominated both AS and SBR bacterial culture collections, the *Bacteroidetes* division comprised 51.3% of sequences from the AS rDNA library and 67% of sequences from the SBR rDNA library.

An unexpected observation from the cultivation of bacteria from both AS and SBR was a high frequency of co-isolated bacterial taxa. Over 20% of the cultured bacteria exhibited this phenomenon in which multiple bacterial taxa were present within seemingly isolated colonies, and in some cases required multiple rounds of sequential restreaking for isolation to achieve pure cultures. The high frequency of co-isolation is likely explained by the presence of bacterial assemblages within floc particles. Bacteria become associated with these flocs via glycocalyx formation and/or fimbrial attachment, making them difficult to physically separate. Alternatively, their association may be symbiotic or commensal in nature, associations which have evolved over time due to their close physical interactions within aggregates and metabolic co-dependence. No

obligate associations were observed among the bacterial co-isolates under the laboratory growth conditions employed, although the difficulty in separation of some co-isolate pairs suggest active mechanisms of cellular attachment.

The bacterial culture collection also provided a resource for the enrichment and identification of cultured bacteriophages isolated from AS and SBR. Of the 91 bacterial isolates, 28 exhibited phage plaques when grown on solid media. Only 13 of the isolated phage could be purified by serial plaquing on their respective bacterial host, suggesting that the remaining phage may be lysogenic or thermosensitive. Phages were most commonly isolated from bacterial cultures in the *γ-Proteobacteria*, *Firmicute,* and *Actinobacteria* divisions, which were also the most commonly represented divisions in AS and the only divisions found in SBR. One pseudolysogenic phage was isolated from a bacterial culture with high similarity to *Brevibacterium sanguinis*, which, to our knowledge, is the first time a pseudolysogenic phage has been identified from this bacterial species.

A metagenomic survey of the viral communities in AS and IN revealed a very high percentage of novel predicted gene products. Approximately 37% of sequences obtained from the AS viral metagenomic library and 18% of sequences from the IN library were considered novel, illustrating a relative scarcity of AS phage DNA sequences in the GenBank nr/nt or env databases. The viral metagenomic surveys also indicated that sequences in the AS viral library were less similar to known sequences and more phylogenetically diverse than those in the IN sample, suggesting that the AS viral community has a greater level of genomic heterogeneity compared to the IN viral community. Of the known sequences in each library, over half of these were most similar

to eubacterial DNA sequences in GenBank. The remaining sequences were classified as eukaryotic, mobile genetic elements, archaeal, or viral; the latter comprised only 22% and 14% of the AS and IN libraries, respectively. The dominance of viral metagenomic sequences with significant homology to eubacterial DNA is likely the consequence of the high frequency of prophage sequences within bacterial genomes deposited in GenBank databases (9). The taxonomic heterogeneity of sequences obtained in the viral metagenomic library underscores the diversity of the AS system, which receives an influx of microbiota from both human and environmental sources. In addition, these surveys revealed a high percentage of novel predicted gene products. Approximately 30% of the total sequences obtained from the metagenomic libraries were considered novel, illustrating a relative scarcity of wastewater community phage DNA sequences in the GenBank nr/nt or env databases.

Metabolic profiles of the viral metagenomic libraries were generated by the SEED platform, which classified viral ORFs into 27 predicted metabolic subsystems. Many of the more dominant processes included various macromolecule metabolisms, consistent with the most highly represented categories observed by Dinsdale *et al.* (13). Despite the lack of exhaustiveness of the viral metagenomic libraries in covering the functional genomic diversity present in either AS or IN viral communities, there is a striking overall similarity in the metabolic category distribution observed with viral metagenomes from other environments.

Our data support the use of both culture-dependent and –independent approaches to gain a more comprehensive understanding of the phylogenetic and functional diversity of microbial assemblages within wastewater communities. There was minimal overlap

between data obtained from culture-dependent and –independent analyses, reflecting the

strong cultivation bias that is common to virtually all environmental studies. These

surveys will also serve as a valuable resource for further investigations into other aspects

of bacterial and viral functional diversity within these wastewater communities, as well

as biological activities (e.g., antibiotic resistance) with ecological and clinical

implications.

Survey	Sequences Analyzed	No. of Divisions Represented
Culture-dependent		
AS, Bacterial	72	5
SBR, Bacterial	19	3
AS, Phage	24	3
SBR, Phage	4	2
Culture-independent		
AS, Bacterial	829	10
SBR, Bacterial	274	5
AS, Phage	1161	11
IN, Phage	618	10

Table 2.1. Summary and comparison of the eight phylogenetic surveys performed in this study. (AS, activated sludge, SBR, sequencing batch reactor; IN, influent)
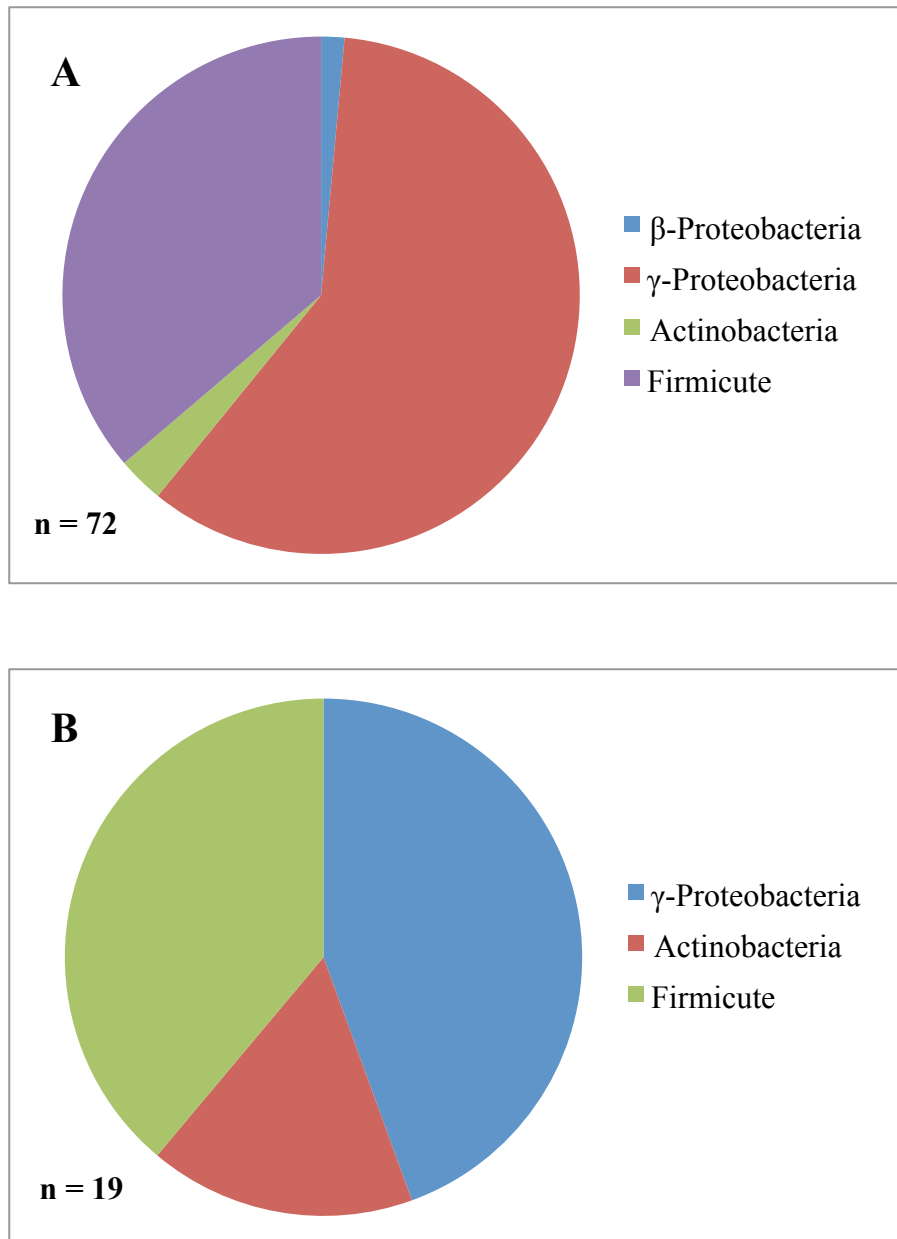
Figure 2.1. Phylogenetic breakdown of bacterial cultured isolates from activated sludge (Panel A) and a sequencing batch reactor (Panel B).
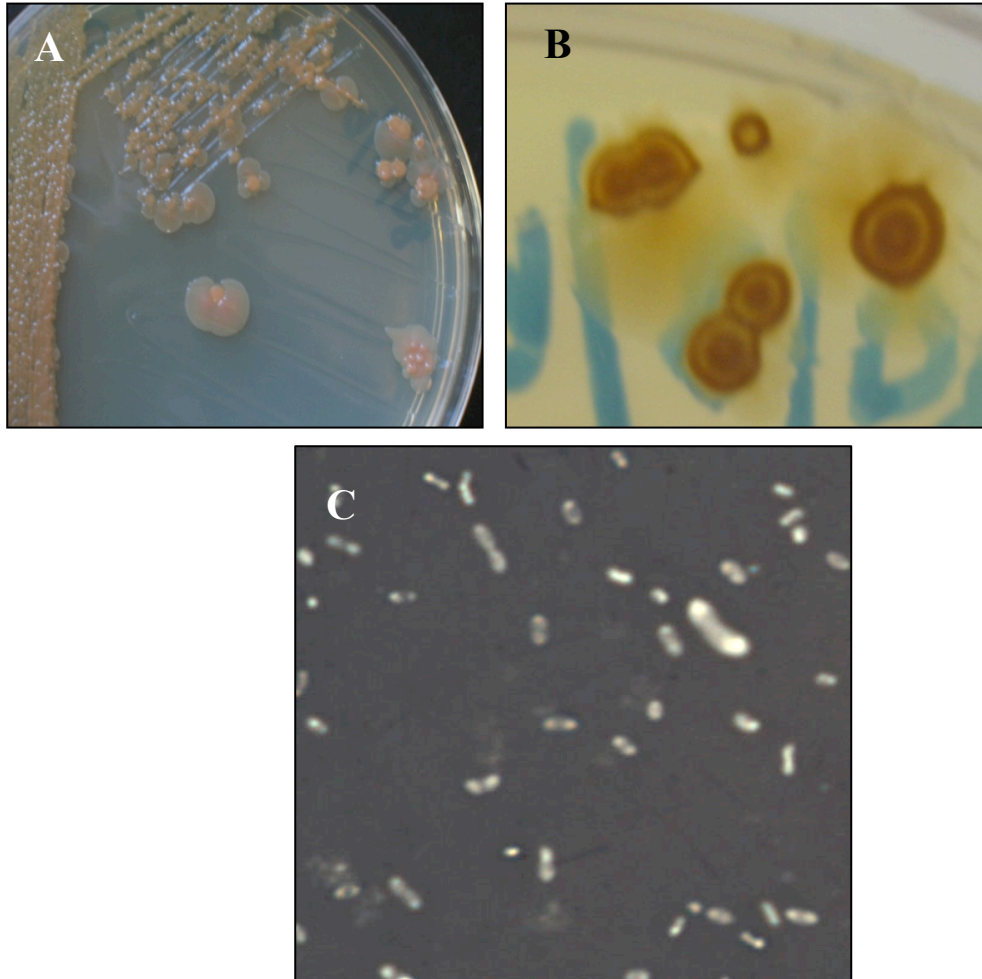
Figure 2.2. Visualization of culture AS 6, a potentially novel *Aeromonas* species. Panels A and B illustrate the production of a brown pigment and "bull's-eye" morphology when grown on solid medium. Panel C, showing a negative stain of AS 6, supports the hypothesis that this is a pleiomorphic isolate.

Culture	Isolate A	Isolate B
AS 4	Actinobacteria (*Micrococcus*)	Firmicute (*Staphylococcus*)
AS 7	γ-Proteobacteria (*Acinetobacter*)	γ-Proteobacteria (*Aeromonas)*
AS 13	Firmicute (*Bacillus*)	γ-Proteobacteria (*Klebsiella*)
AS 32	Firmicute (*Exiguobacterium*)	γ-Proteobacteria (*Pantoea*)
AS 37	Firmicute (*Staphylococcus*)	γ-Proteobacteria (*Aeromonas*)
AS 38	Firmicute (*Bacillus*)	γ-Proteobacteria (*Pantoea*)
AS 47	Firmicute (*Bacillus*)	γ-Proteobacteria (*Aeromonas*)
AS 52	Firmicute (*Staphylococcus*)	γ-Proteobacteria (*Acinetobacter*)
AS 57	Actinobacteria (*Arthrobacter)*	γ-Proteobacteria (*Klebsiella*)
AS 61	Bacteroidetes (*Chryseobacterium*)	γ-Proteobacteria (*Citrobacter*)
AS 71	Firmicute (*Staphylococcus*)	γ-Proteobacteria (*Acinetobacter*)
AS 74	γ-Proteobacteria (*Enterobacter*)	γ-Proteobacteria (*Aeromonas)*
AS 75	Firmicute (*Bacillus*)	γ-Proteobacteria (*Acinetobacter*)
AS 78	Firmicute (*Staphylococcus*)	Firmicute (*Bacillus*)
SBR 10	Firmicute (*Bacillus*)	γ-Proteobacteria (*Morganella*)

Table 2.2. Phylogenetic affiliations of bacterial co-isolates identified from activated sludge and a sequencing batch reactor. The bacterial division with which each bacterial isolate is affiliated is indicated followed by its respective genus in parentheses.
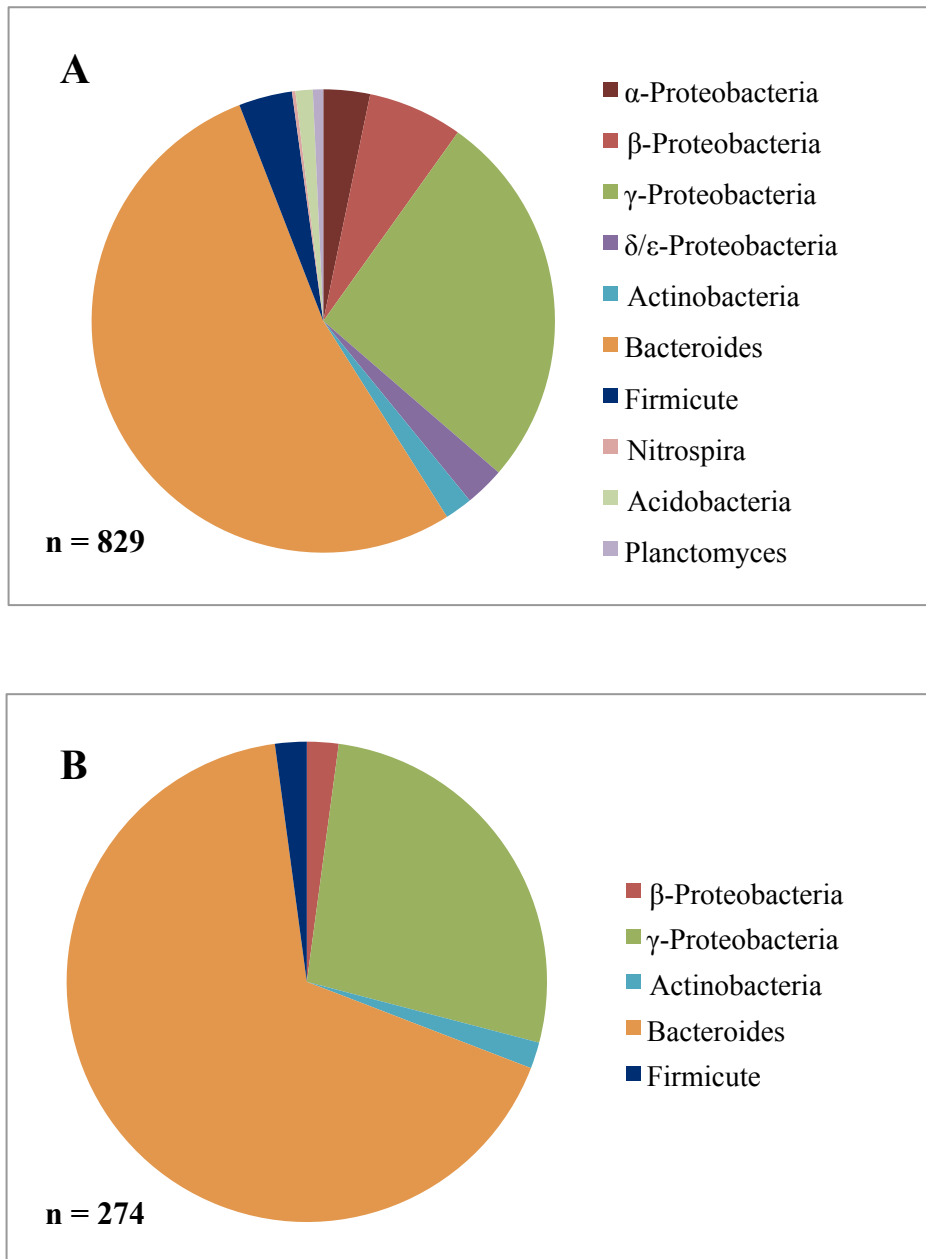
Figure 2.3. Division-level phylogenetic affiliations of sequences from bacterial 16S rDNA clone libraries (Panel A, activated sludge; Panel B, sequencing batch reactor). Gene sequences were compared to known sequences in the GenBank nr/nt database.
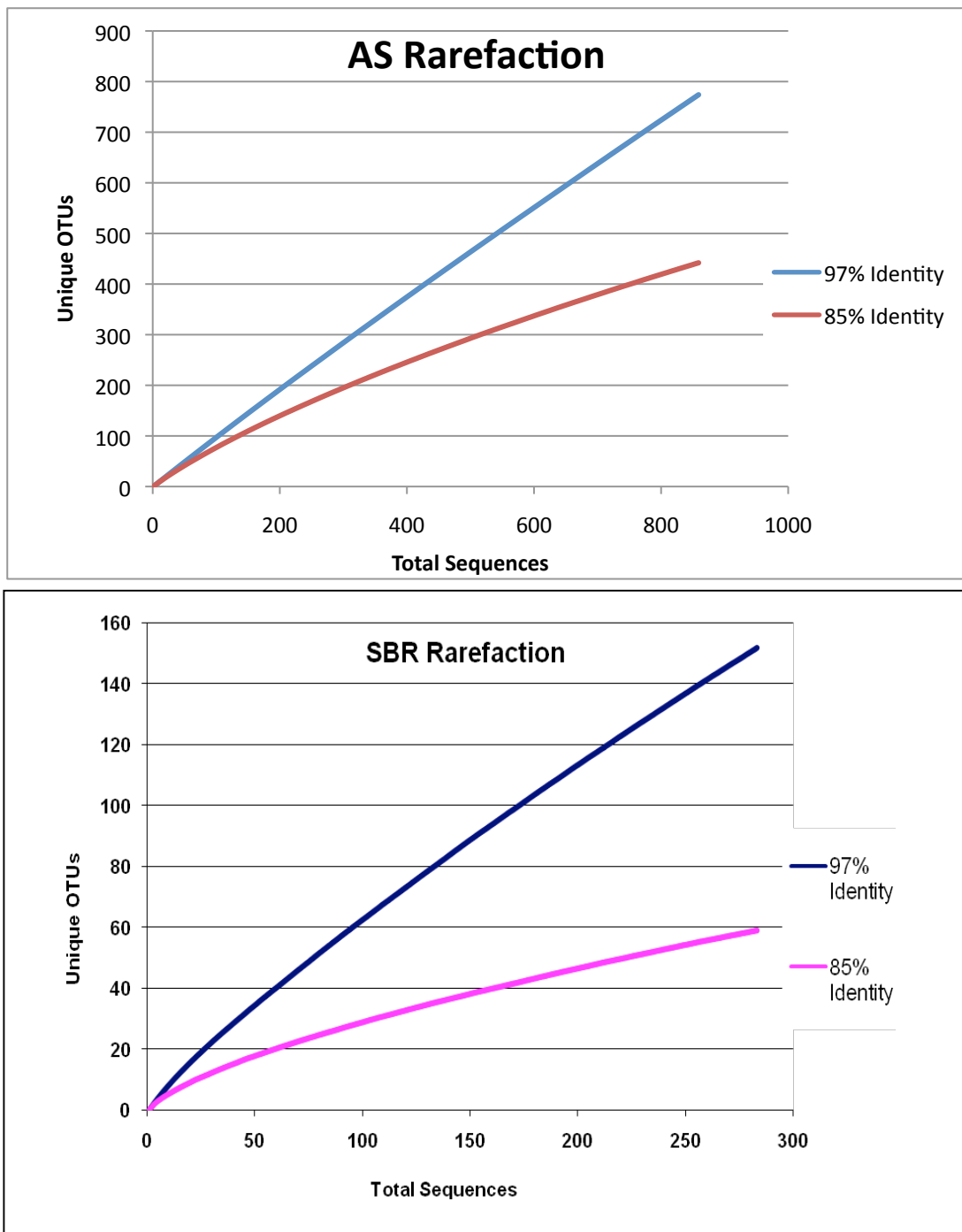
Figure 2.4. Rarefaction analyses of 16S rDNA library sequences from activated sludge (AS) and the sequencing batch reactor (SBR).

96
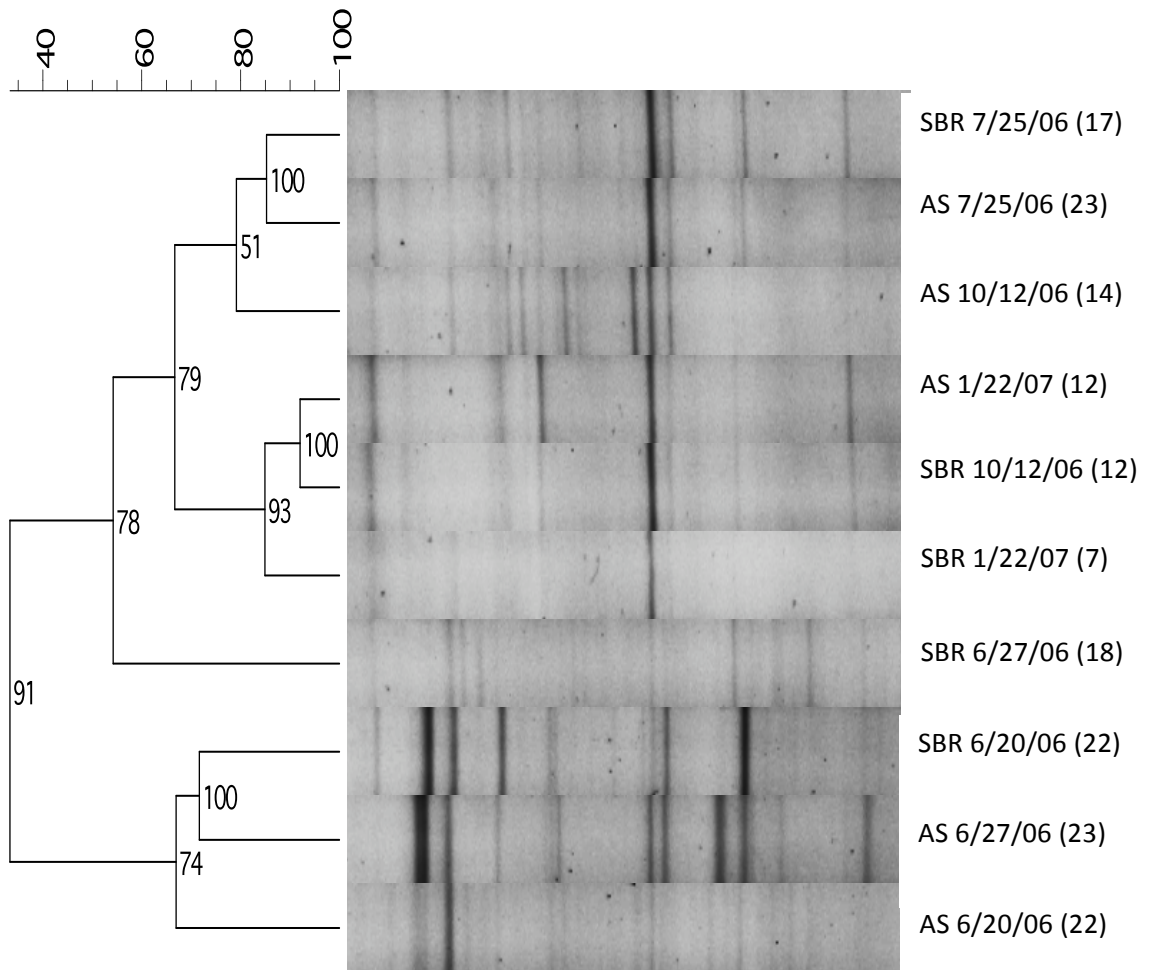
Figure 2.5. UPGMA dendrogram based on the banding patterns of community 16S rDNA from each time point. Cophenetic correlation values are shown at each node. The number given in parenthesis behind each sample name represents the taxa richness estimate of that sample.
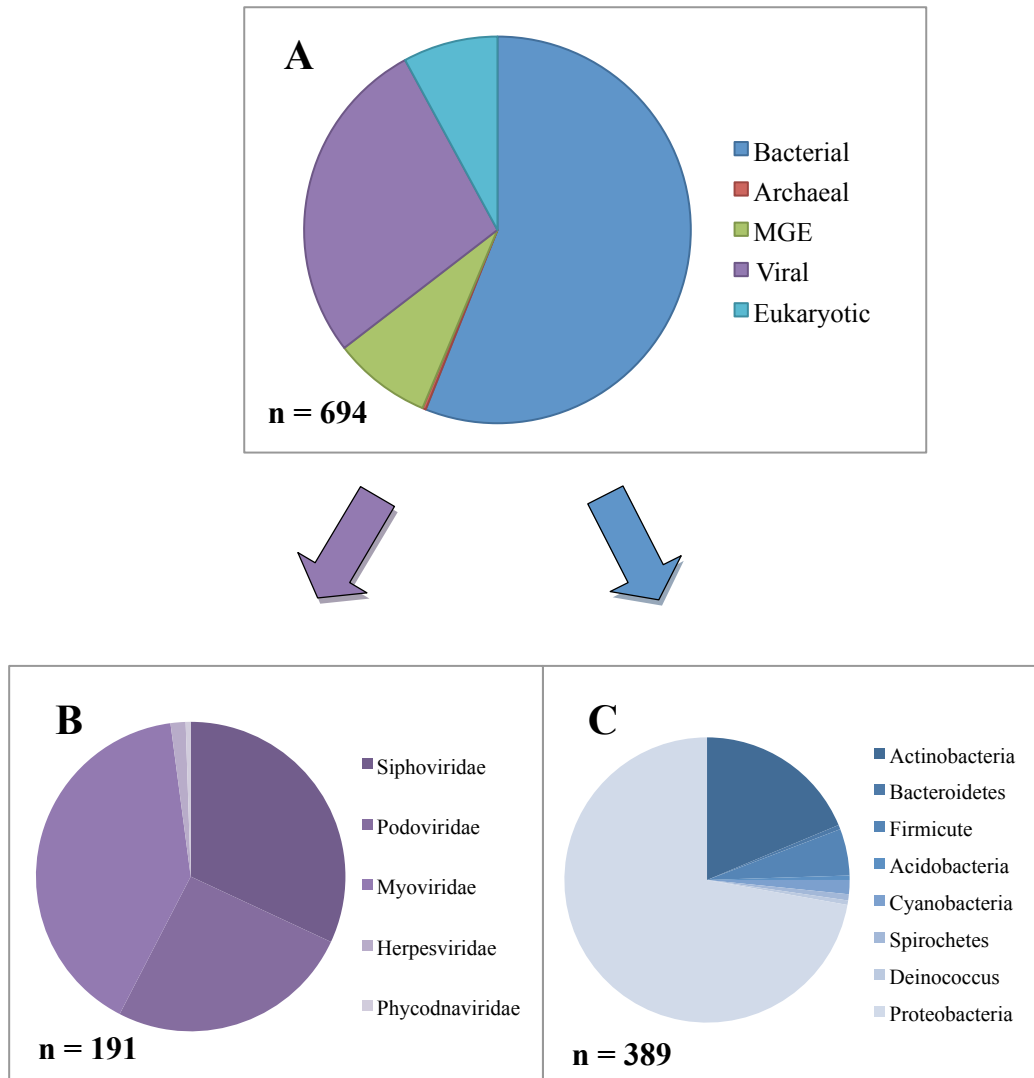
Figure 2.6. Breakdown of eubactrial and viral hits from the AS viral metagenomic library. Panel A illustrates the overall functional affiliations of the library clones, while Panel B and Panel C further separate the the hits by viral family affiliation and eubacterial division affiliation, respectively.
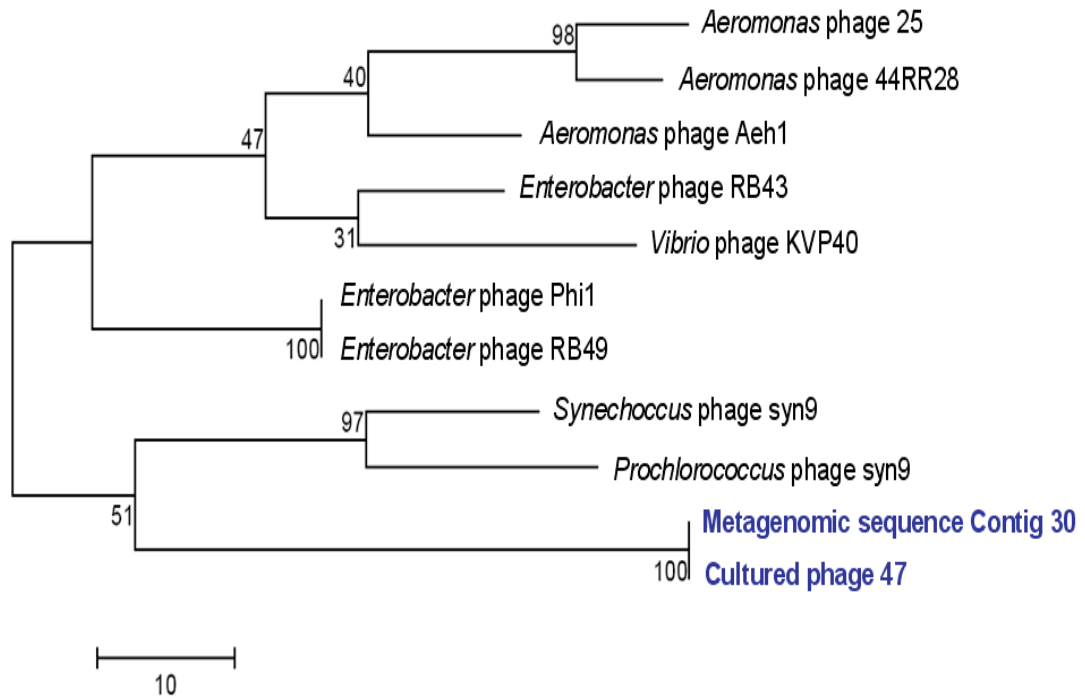
Figure 2.7. Maximum parsimony tree of viral gene sequences derived from a cultured

bacteriophage, a metagenomic sequence, and related gp4 head completion proteins. The

number at each node represents the bootstrap value after 100 replicates.
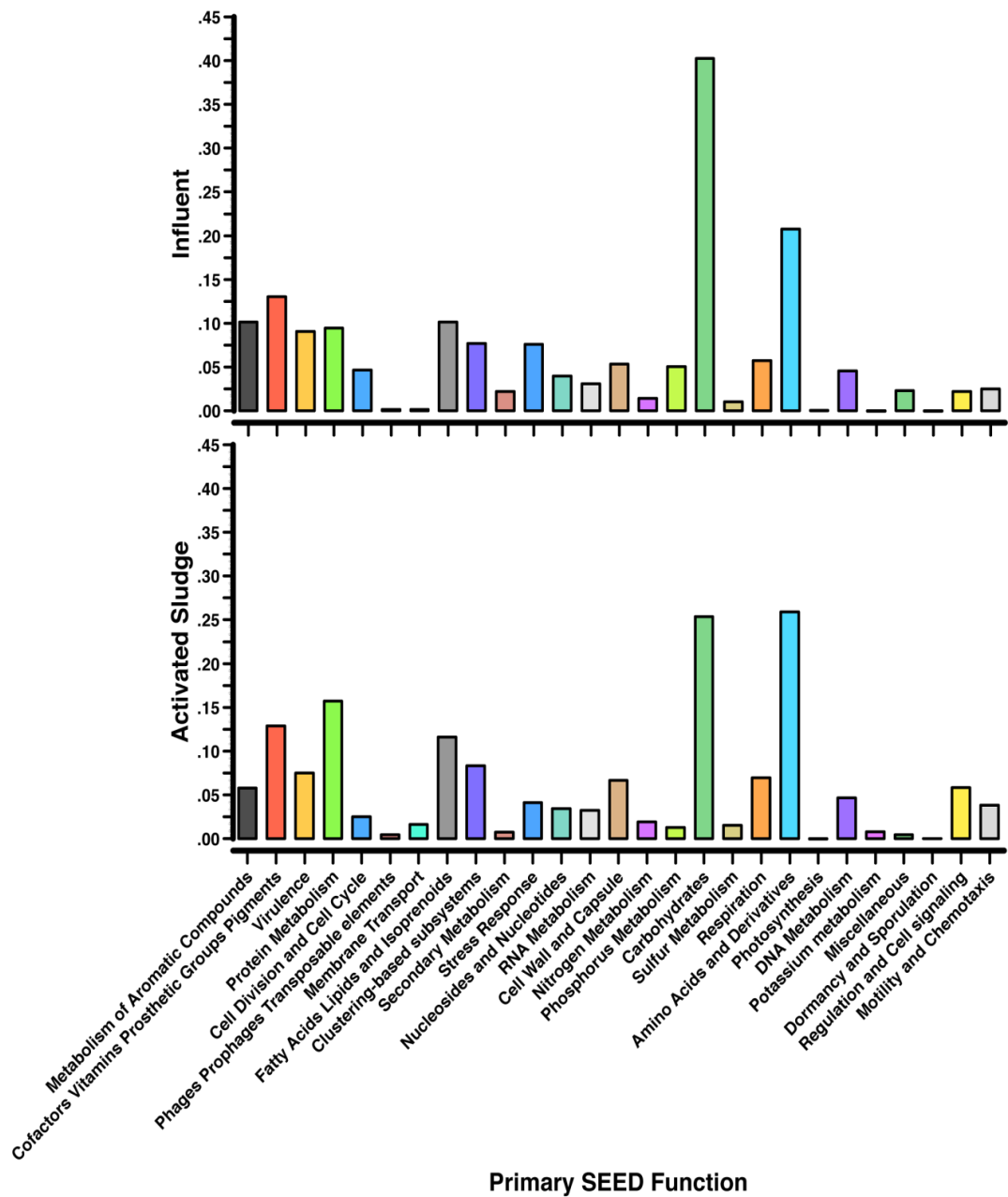
Figure 2.8. Comparison of influent and activated sludge metabolic profiles obtained

from SEED analyses of viral metagenomic DNA. The y-axis indicates the percentage of

ORF sequences classified into each metabolic category.
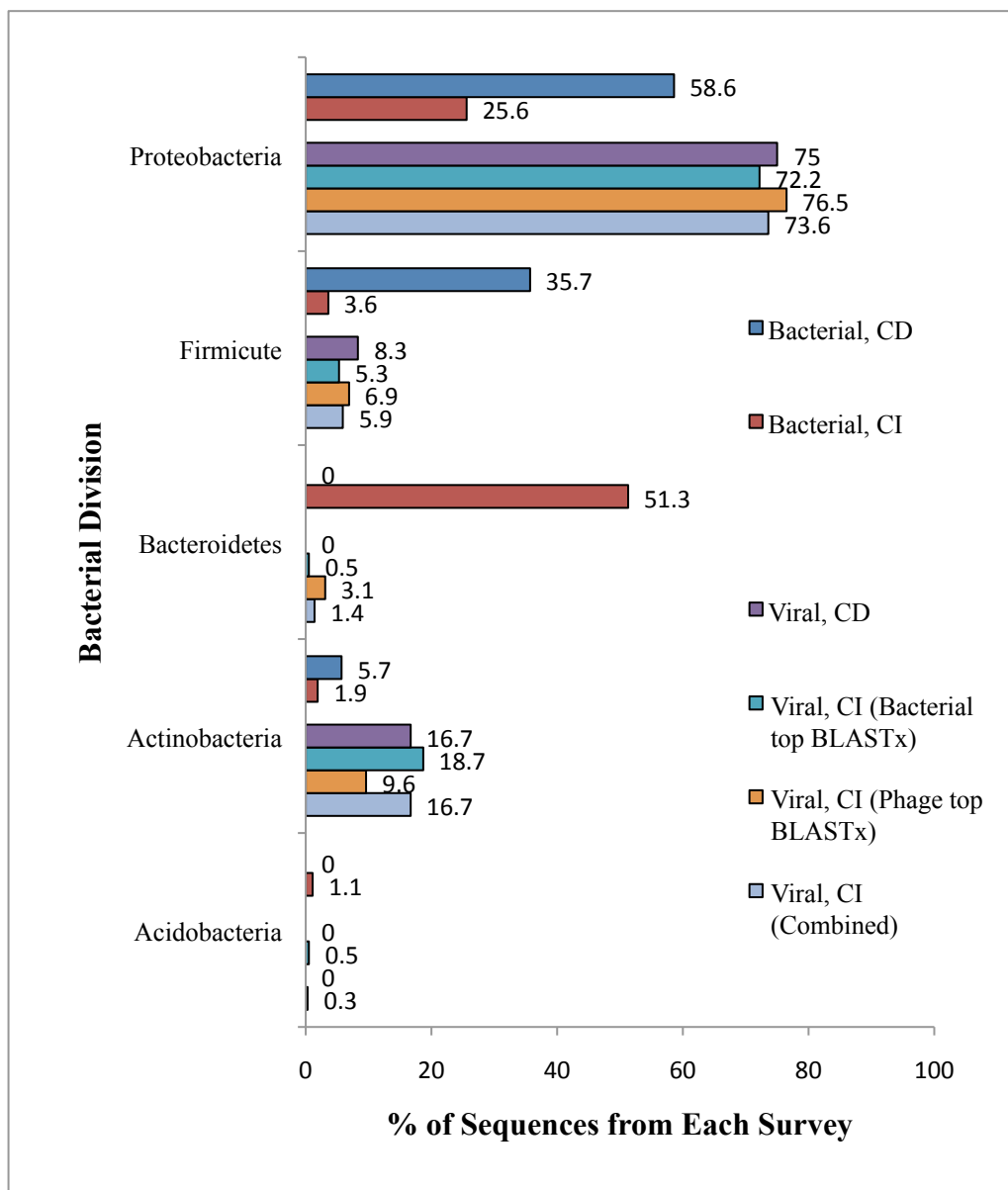
Figure 2.9. Relative dominance of bacterial divisions represented by sequences from the bacterial culture collection, 16S rDNA clone library, cultured bacteriophage, and viral metagenomic library (eubacterial and phage sequences). The percentage of sequences affiliated with each bacterial division is compared among the four phylogenetic surveys. (CD, culture-dependent; CI, culture-independent).

CHAPTER III

DIVERSE ANTIMICROBIAL RESISTANCE DETERMINANTS ISOLATED
FROM BACTERIAL, PLASMID, AND BACTERIOPHAGE METAGENOMES
IN ACTIVATED SLUDGE

A. ABSTRACT

Using both sequence- and function-based approaches, metagenomic libraries were constructed from bacterial chromosomes, plasmids, and bacteriophages from an activated sludge microbial assemblage, and screened for the presence of genes conferring resistance to any of 12 antibiotics to an *Escherichia coli* heterologous host. Sampling from three genetic sources (i.e., genome, plasmid, and phage) was performed to access the extent of antibiotic resistance in this environment amongst each potential reservoir of antibiotic resistance determinants. Our surveys resulted in the identification of nine bacterial metagenomic clones encoding resistance to chloramphenicol (n=7), ampicillin (n=1), and kanamycin (n=1), a multi-drug resistant plasmid (pAS1), and six phage metagenomic clones with sequences similar to those in the GenBank nr/nt database that encode resistance to tetracycline, bleomycin, acriflavin, and ampicillin. Although many of the analyzed clone sequences have homology with genes generally associated with antibiotic resistance, six of the recombinant clones conferring chloramphenicol resistance lacked any significant homology with previously characterized chloramphenicol

resistance genes. In addition to the antibiotic resistance determinants, the metagenomic

surveys also identified mobile genetic elements (MGEs) from each genetic source,

indicating the potential for the lateral transfer of antibiotic resistance genes among

various bacterial hosts. The antibiotic resistance gene and MGE sequences from all

surveys represent bacteria from seven distinct taxa, including the *Bacteroidetes,*

Proteobacteria, Firmicutes, and *Chlamydiae*, further demonstrating the benefit of using a

metagenomic approach to access genomes from phylogenetically diverse bacteria to

explore the extant diversity of potentially mobile antibiotic resistance determinants.

B. INTRODUCTION

Biological wastewater treatment systems are engineered to treat polluted

wastewater before the effluent is discharged into the environment. These systems are

commonly referred to as the activated sludge process, and the associated microbial

community is now known to be both dynamic and functionally and phylogenetically

diverse (32, 107, 145, 183, 192, 234, 250). Microorganisms in activated sludge play a

significant role in the wastewater treatment process, as they remove nutrients and degrade

harmful compounds in wastewater before it is released into the environment (83, 117).

More specifically, the horizontal genetic transfer (HGT) of antibiotic resistance

determinants within these microbial communities has become of great interest because

activated sludge may serve as a reservoir for diverse antibiotic resistance mechanisms

that can potentially be dispersed throughout the environment, with both public and

environmental health implications.

HGT among microbiota occurs through transformation, conjugation, and

transduction (22, 187, 225, 266). Since the activated sludge process promotes close

physical interactions among diverse microorganisms, there is great potential for HGT of antibiotic resistance genes between microbes in activated sludge and in downstream environments. Several studies have previously identified antibiotic resistance determinants from wastewater communities that are encoded on bacterial chromosomes (13, 82, 278) and plasmids (227, 262, 263), but to our knowledge, a survey of bacteriophage-encoded resistance genes in this environment has not been performed. To achieve a more comprehensive assessment of antibiotic resistance genes in the activated sludge microbial community, it is crucial to consider all possible origins of these genes, as well as the potential for their lateral transfer among diverse microorganisms.

The application of metagenomics to characterize the phylogenetic and functional diversity of a microbial community is a powerful tool that can be used to identify gene products with specific functions, without *a priori* knowledge of their genetic determinants. Previous studies have used a metagenomic approach to identify novel genetic pathways and specific microbial genomes in wastewater microorganisms, such as those encoding aromatic degradation or phosphorus removal (83, 256). In addition, the use of both functional- and sequence-based screening to identify genetic loci encoding antibiotic resistance has revealed the presence of novel and diverse resistance determinants that can be transferred between bacterial hosts (82, 83, 132). In this study, we have used both functional- and sequence-based metagenomic approaches to identify diverse and potentially novel antibiotic resistance determinants encoded on bacterial chromosomes, plasmids, or bacteriophages within an activated sludge microbial assemblage.

C. MATERIALS AND METHODS

1. Isolation of bacterial, plasmid, and bacteriophage metagenomic DNA. At two time points in 2007, samples of activated sludge (AS) from the aeration basin of the H.C. Morgan Water Pollution Control Facility in Auburn, AL were collected as previously described (192). Bacterial genomic DNA was extracted from 2 ml aliquots (x10 replicates) of AS using the MoBio UltraClean Microbial DNA extraction protocol (MoBio Laboratories, Carlsbad, CA). DNA from each aliquot was eluted in 30 μl of $T_{10}E_1$ buffer (10 mM Tris-HCl, 1 mM EDTA), pooled, and stored at -20°C. The pooled DNA was then purified by ethanol precipitation and CTAB extraction. Briefly, 5 M NaCl was added to the genomic DNA to a final concentration of 0.7 M NaCl. Then, a solution containing 10% hexadecyltrimethylammonium bromide in 0.7 M NaCl was added, and the mixture was vortexed and incubated at 65°C for 10 min. Genomic DNA was then extracted using isoamyl alchohol:chloroform (1:1 v/v), precipitated with 100% ethanol, and resuspended in $T_{10}E_1$ buffer (10 mM Tris-HCl, 1mM EDTA). Purified DNA was stored at -20°C.

Community plasmid DNA was isolated from a 50 ml sample of AS using a manual plasmid extraction method. Briefly, the AS sample was centrifuged for 10 minutes at 3600 x g. The cell pellet was resuspended in Solution P1 (15 mM Tris-HCl, 10 mM EDTA, 100 μg/mL RNase A) and lysed with Solution P2 (0.2% NaOH, 1% SDS). Cellular proteins were precipitated using Solution P3 (3 M KOAc, pH 5.5), and the mixture was centrifuged as before. Supernatant was aspirated into a fresh tube and mixed with 0.6x total volume of 100% isopropanol. DNA was then precipitated, and the

resulting DNA pellet was vacuum-dried and resuspended in 50 μL of $T_{10}E_1$ buffer. Plasmid DNA was stored at -20°C.

Activated sludge samples for viral DNA isolation were combined with an equal volume of 10% beef extract and agitated for 18 h at room temperature, then subjected to centrifugation at 10,000 x g for 30 min. The supernatant was filtered through a 0.45 μm PVDF filter and precipitated with polyethylene glycol (PEG) at 4°C overnight. The overnight PEG solution was centrifuged at 7,000 x g for 45 min, and the resulting viral pellet was resuspended in SM buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl, pH 7.5). The resuspended solution was spun again (12,000 x g for 10 min) and the supernatant was loaded onto a cesium chloride step gradient and subjected to ultracentrifugation at 100,000 x g for 2 h. The viral band evident at the 1.3 mg/mL zone was collected and used for DNA extraction. The resulting DNA was treated with 100 μl of benzonase (250 U/μl final concentration) at 37°C overnight to degrade extra-viral DNA. Following benzonase inactivation (addition of 10 mM EDTA and heating at 70°C for 10 min), proteinase K (1mg/ml final concentration) and 1% sodium dodecyl sulfate were added, and the mixture was incubated at 37°C for 2 h to degrade viral protein coats. Proteins were removed by phenol:cholorfom extraction and DNA was recovered by isopropanol precipitation. The purified phage DNA was stored at -20°C.

2. Small-insert bacterial metagenomic library construction. Purified microbial metagenomic DNA from AS at two time points was sheared to generate random fragments using a HydroShear machine (Gene Machines, Singapore) at a setting of 7, 8, or 9 to produce 3-5 kb fragments. The DNATerminator kit (Lucigen, Middleton, WI) was then used to blunt end 3'and 5' overhangs for efficient cloning. To size-select

repaired fragments, DNA was electrophoresed on a 1X TAE 1% (w/v) HiRes agarose gel (ISC BioExpress, Kaysville, UT), stained with SYBR Gold, and visualized on a low-wavelength UV transilluminator. DNA in the 1-5 kb range was isolated using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA). Because the amount of DNA in this size range was greater than the capacity of a single DNA binding column, multiple columns were used, and the resulting eluted DNA was pooled.   The DNA was concentrated by ethanol precipitation and resuspended in 20 µL of $T_{10}E_1$ buffer. The purified DNA was then ligated into the pSMART LCKan and LCAmp vectors (Lucigen, Middleton, WI), which were chosen because this would allow for "cross-screening" of the resulting libraries (i.e., for kanamycin and ampicillin resistance, respectively) and because they transcriptional terminators flanking the cloning site to direction by the host machinery only. All ligations were transformed into electrocompetent *E. coli* DH10B Supreme cells (Lucigen, Middleton, WI) at 1800V/cm, plated onto 2xYT agar with 30 µg/ml kanamycin (LCKan ligation) or 100 µg/ml ampicillin (LCAmp ligation), and incubated overnight at 37°C.

Ten clones from the kanamycin and ampicillin plates were subcultured into 2xYT broth with appropriate antibiotic (Ab), and plasmid DNA from each clone was extracted as described above. Plasmid DNA was then restriction digested with EcoRI and electorphoresed through a 1x TAE agarose gel (1.2% w/v) to estimate the average insert size and the percentage of clones with insert. The total amount of cloned metagenomic DNA was determined based on estimates of the transformation efficiency, average insert size, and the percentage of clones with an insert. Library statistics are shown in Table 3.1.

3. Screening bacterial metagenomic libraries for antibiotic resistance. Each

metagenomic library transformation mix was plated onto 2xYT agar containing one of 12

antibiotics (Table 3.2) and incubated overnight at 37°C. All *E. coli* clones exhibiting

antibiotic resistance were picked onto a fresh plate of 2xYT agar with appropriate

antibiotic (Ab). Viable clones were subcultured into 2xYT broth containing respective

Ab, incubated overnight at 37°C, and frozen as glycerol stocks. Plasmid DNA from each

clone was then extracted as described above and re-transformed into chemically

competent *E. coli* DH10B to verify the original phenotype.

Clones that exhibited antibiotic resistance upon re-transformation of naïve *E. coli*

were subcultured into 2x YT broth and plasmid DNA was extracted as described above.

To identify unique clones, plasmid DNA was restriction digested with Sau3AI, and

restriction fragment patterns for each clone were visualized on a 1x TAE agarose gel

(1.2% w/v). A representative collection of unique clones is shown in Figure 3.1.

4. Sequence analysis of unique chromosomal-encoded antibiotic resistance

determinants. Unique clones were grown overnight in 2xYT containing appropriate Abs

at 37°C and transferred into a 96-well plate. Glycerol was added to each at well at a final

concentration of 10% (v/v), and the clones were submitted for sequencing to the Lucigen

Corp. (Middleton, WI) using vector-specific primers. Sequences were then trimmed for

quality using ChromasPro (Technelysium Pty, Queensland, Australia), and the resulting

sequences were analyzed by the BLASTx algorithm and compared to the non-redundant

nucleotide database (GenBank, NCBI). In addition, predicted open reading frames

(ORFs) were identified using ChromasPro software and the NCBI ORFinder tool. Each

ORF was compared to the GenBank non-redundant protein database by the BLASTp

algorithm. Primer walking and additional sequencing reactions were performed to obtain longer sequence reads for each insert (≥600 bp) in order to obtain a complete contiguous sequence for each cloned insert.

For six of the chloramphenicol resistant (Cm^R) clones with no apparent homology to known Cm ^{R}-associated determinants, transposon mutagenesis was performed to identify the genetic loci necessary for antibiotic resistance. The EZ-Tn5<TET-1> transposon (Tn) system (Epicentre, Madison, WI) was used according to the manufacturer's instructions. Loss-of-function mutants were identified by transformation of each mutagenized clone onto 2xYT with selection for the vector Ab^R (Kan or Amp), and then re-plating each transformant onto 2xYT containing Cm to identify Cm^S clones. The insertion site for each loss-of-function transposon mutant was identified by sequencing using mutagenized plasmid template with Tn5-specific primers and mapping the Tn insertion site against the previously sequenced clone insert DNA sequence. In the two cases where a single ORF was indicated as necessary and sufficient to confer Cm^R (i.e., clones A7 and A10) and significant additional cloned DNA was present within the insert, the ORFs identified as responsible for Cm^R were subcloned with their native promoters into the TOPO TA vector (Invitrogen, Carlsbad, CA), transformed into *E. coli*, and screened for Cm^R.

The clone sequences were also analyzed by comparing them to their significant homologs in the nr/nt database, and to all known genetic loci that can confer Cm^R (e.g., Type I, II, and III chloramphenicol acetyltransferases) (236). A molecular phylogenetic analysis was conducted for each of the ORFs implicated in Cm^R (total of 10 ORFs within

six clones) using ClustalX for multiple alignments and the maximum parsimony method
with bootstrap support (1000 replicates) on the MEGA 4.1 software package (131, 264).

5. Functional analyses of Cm^R clones. Minimum inhibitory concentration (MIC)
values were obtained from each of the six Cm^R clones that lacked homology with known
Cm^R determinants using the macrodilution susceptibility method according to the CLSI
standardized protocol. Each clone and *E. coli* ATCC 25922 (reference strain) were tested
in triplicate. The following chloramphenicol concentrations (μg/ml) were tested: 0, 1, 2,
4, 8, 16, 32, 64, 128, 256, and 512.

Chloramphenicol from clone supernatants after overnight incubation detected by
HPLC (Hewlett-Packard, HP 1200). The six clones, pSU2719 (positive control plasmid
carrying the CAT gene), and chloramphenicol alone (negative control) were incubated
overnight at 37ºC with shaking at 150 rpm. The supernatants from each culture were
harvested by subjecting the samples to centrifugation at 13,000 x g for 5 min. Each
supernatant was then filtered through a 0.22 μm syringe filter to remove media
components that may interfere with chloramphenicol detection. The HPLC system
consisted of a degasser (G1322A), a quaternary pump (G1311A), an ALS auto-sampler
(G1313A), a colcomp column oven (G1316A), and variable wavelength UV-VIS detector
(G1314A). A Hypersil ODS C18 (125x46 mm, 5μm) column was used. HPLC operating
conditions were as follows: UV detector wavelength, 197 nm; mobile phase, acetonitrile
and water (40:60) with solvent delivered at a constant flow rate of 1mL/min; and a run
time of 15 min. The profiles for each clone and control were obtained and analyzed
using the ChemStation/LC software (Agilent Technologies, Santa Clara, CA).

To further explore the mechanism of the Cm^R clones, each clone was tested for susceptibility to florfenicol, a fluorinated synthetic analog of chloramphenicol. The clones were tested against florfenicol (30 µg/ml) using the Kirby-Bauer disk diffusion method. Clones were considered susceptible if they exhibited a zone of inhibition >21 mm.

6. Identification of antibiotic resistance plasmids. Community plasmid DNA was treated with a plasmid-safe exonuclease (Epicentre, Madison, WI) to degrade contaminating chromosomal DNA. A portion of the community plasmids was mutagenized with the <R6Kγori/KAN-2> transposon (Epicentre, Madison, WI) to rescue plasmids that may be unable to replicate in an *E. coli* host (116). The naïve and mutagenized plasmids were then transformed into electrocompetent *E. coli* pir+ Transformax EC100 cells (Epicentre, Madison, WI) and plated onto 2xYT agar containing one of 12 antibiotics (Table 3.2) and incubated overnight at 37°C. Plasmid DNA was extracted from 20 (10 naïve and 10 mutagenized) plasmids, digested with *Eco*RI, and analyzed for unique restriction patterns. Unique plasmids were re-transformed into chemically competent *E. coli* to verify the Ab^R phenotype.

7. Shotgun subcloning of plasmid pAS1. The multi-drug resistant plasmid pAS1, isolated from both naïve and transposon-rescue transformations, was subcloned into pSU2719 to obtain additional sequence data. *E. coli* strain DH10B harboring pAS1 was grown in 500 ml culture in 2xYT medium containing 50 µg/ml kanamycin overnight with shaking at 37°C for large-scale plasmid preparation of pAS1 as described above. The pAS1 DNA was digested with plasmid-safe exonuclease (Epicentre, Madison, WI) for 2 h at 37°C to remove chromosomal DNA, followed by heat inactivation for 10 min at

70°C. Plasmid DNA was then sheared, end-repaired, and size-selected using the same methods described above for bacterial metagenomic library construction. The pSU2719 vector was prepared by restriction digestion with SmaI to generate linearized vector and treatment with alkaline phosphatase to prevent vector self-ligation. Purified pAS1 fragments were then ligated into pSU2719 at 15°C overnight using T4 DNA ligase, and the ligation mixture was transformed into electrocompetent *E. coli* DH10B. The transformation was plated onto 2xYT medium containing 12.5 µg/ml Cm and 30 µg/m X-gal for transformant selection and blue-white screening. White transformants were picked into 96-well plates with each well containing 1 ml of 2xYT broth containing 12.5 µg/ml Cm and 25% glycerol. After overnight incubation at 37°C, the plates were sent to the Lucigen Corp. (Middleton, WI) for sequencing.

8. Construction of small-insert viral metagenomic libraries. Random shotgun metagenomic libraries of the viral communities in AS were prepared at the Lucigen Corporation (Middleton, WI) (193, 288). Purified phage DNA isolated from AS was fragmented to approximately 1-3 kb fragments using a HydroShear machine (Gene Machines, Singapore) and end-repaired to blunt-end each dsDNA molecule. The dsDNA was then ligated to an adaptor (20 bp dsDNA sequence), and then primers specific to the adaptor sequence were used to PCR amplify the dsDNA fragments using only 15 rounds of amplification. The viral amplicons were then purified over a silica column and ligated into the pSMART-LCKan vector, a low-copy cloning vector that has transcriptional terminators flanking the cloning site to prevent transcription of viral genes. The ligation was transformed into electrocompetent *E. coli* strain DH10B, and a glycerol stock was prepared of the transformation mixture for storage at -80°C. The transformation mix was

shipped to Auburn University, where the number of transformants per mL was determined. Transformants were plated onto LB agar containing kanamycin (30 µg/mL) and were picked into 96-well format using the QPix2 colony picking robot (Genetix, Hampshire, United Kingdom). After incubation at 37°C overnight with shaking, sterile 50% glycerol was added to each well (final concentration, 15%) and plates were frozen at -80°C. Thirteen 96-well plates (1,248 clones) from AS were sent to the Lucigen Corp. (Middlteon, WI) for dideoxy sequencing.

9. Identification of antibiotic resistance determinants from viral metagenomic libraries. Resulting clone sequences were screened for vector sequence, linker sequence, and minimum base quality using ChromasPro software. Sequences were then manually analyzed for sequence quality. Following removal of poor sequence, there were 1,161 quality sequences from AS (approx. 0.76 Mbp). For taxonomy assignments, sequences were compared against the GenBank nr/nt and env databases using the tBLASTx algorithm. The top five hits were used for analysis, and only matches with an E-value of ≤ 0.001 were considered significant. Significant GenBank database hits were categorized as viral, bacterial, archaeal, eukaryotic, or a mobile genetic element.   In cases where multiple significant hits were observed for a single query sequence, the sequence was preferentially classified as viral if the hit occurred within the top five hits (37). Mobile elements consisted of plasmids, transposons, transposases, pathogenecity islands, and insertion sequences. Each clone sequence was observed for significant homology with known antibiotic resistance determinants in the GenBank databases.

The viral metagenomic libraries were also functionally screened for antibiotic resistance as described above (bacterial metagenomic library screening), by plating the

viral metagenomic library transformation mix onto 2x YT medium containing each of the 12 respective antibiotics in this study (Table 3.2).

10. Nucleotide accession numbers. Novel chromosomal- and plasmid-encoded resistance genes and predicted MGEs are deposited in GenBank under the accession numbers GU720994 to GU721005. The viral metagenomic DNA sequences were registered as part of the Wastewater Viral Metagenome project at GenBank (Trace Archives) and are under TI reference numbers 2251203077 to 2251204802.

D. RESULTS

1. Bacterial chromosomal-encoded antibiotic resistance determinants. The small-insert bacterial metagenomic libraries contained approximately 1.85 Gbp of cloned DNA, or approximately 400 *E. coli* genome equivalents. Function-based screening of the libraries identified nine unique clones exhibiting resistance to chloramphenicol (n=7), ampicillin (n=1), and kanamycin (n=1) whose predicted gene products shared significant homology with both known and potentially novel antibiotic resistance determinants in the GenBank nr/nt database (Table 3.3). The clones that conferred resistant to ampicillin and kanamycin contain inserts with sequences similar to a Class C β-lactamase and an O-methyltransferase, respectively. However, six of the seven clones showing resistance to chloramphenicol appeared to encode gene products that are not generally associated with chloramphenicol resistance (i.e., chloramphenicol acetyltransferase (CAT)). Clone A5 shared no significant homology to any predicted gene products in either the nr/nt or env GenBank database, and three of the Cm^R clones (A9, A10, and B4) showed top BLASTx hits to bacterial strains that are typically susceptible to chloramphenicol (28, 121, 251, 298).

114

To identify the genetic loci necessary for Cm^R in these six clones, full-length sequences of the clones were obtained, and the clones were mutagenized with a *Tn*5 transposon containing a tetracycline resistance cassette to generate loss-of-function mutants (Figure 3.2). In two cases, the interrupted ORFs did not show homology with any sequences in the GenBank nr/nt or env databases. Three clones contained interrupted ORFs whose top BLASTx hits indicated hypothetical proteins, including clone B4, which contained an additional interrupted ORF with similarity to an ABC transporter protein that appeared to be responsible for chloramphenicol resistance.

The predicted phylogeny of the Ab^R determinants from all nine clones is shown in Table 3.3. The cloned inserts showed homology with representatives of the *Bacteroidetes* phylum, as well as all four sub-divisions of the *Proteobacteria*. In addition, BLAST comparison and multiple alignments of five of the six Cm^R clones with known Cm^R determinants (Type I, II, and III CATs) revealed that these genes do not share any significant similarity with CATs and likely originate from diverse bacterial lineages. Figure 3.3 shows a representative maximum parsimony analysis of clone A4 and its predicted evolutionary relationship to its nearest significant neighbors in the GenBnk nr/nt database. Predicted gene products from several bacterial taxa are represented, including the *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*. The observed similarity between clone A4 and these homologs suggest the gene either be conserved among diverse bacterial genomes or that the gene has a monophyletic origin and has been laterally transferred among multiple hosts.

To further investigate the resistance mechanism(s) encoded by these six Cm^R clones, clone supernatants were analyzed by HPLC, and the resulting profiles were

compared to that of pSU2719, a plasmid that encodes the CAT enzyme. All six of the

clones showed comparable peak profiles to that of pSU2719 (representative clone shown

in Figure 3.4), suggesting that in each of the respective clones, the Cm may be

enzymatically modified and inactivated. The presence of a CAT-like mechanism was

further supported by the results of the florfenicol susceptibility testing, in which all six

Cm^R clones tested exhibited zones of inhibition measuring ≥25 mm and were considered

susceptible. Because florfenicol resistance is typically conferred by efflux pumps which

also provide resistance to chloramphenicol (290), the observation of the clones'

resistance to chloramphenicol but not florfenicol indicates the presence of a Cm-specific

resistance mechanism. Interestingly, these functional data do not corroborate the

sequence data obtained, which did not indicate CAT-associated chloramphenicol

resistance. To our knowledge, none of their significant homologs in GenBank have been

previously linked to chloramphenicol resistance of any kind (e.g., CAT, efflux pumps).

MIC data obtained for each Cm^R clone indicates a higher level of resistance than is often

seen with this particular antibiotic (Table 3.4). While many of the clones showed similar

MIC thresholds (64 μg/m), two of the clones exhibited much higher levels of

chloramphenicol resistance (256 μg/ml).

Because a CAT-like mechanism of Cm^R was indicated by the functional analyses

of the six Cm^R clones, the clone inserts were compared to known CAT genes from

diverse bacteria. Hydrophobicity plots of predicted ORFs conferring Cm^R on clone A4

and the CAT gene found on pSU2719 are compared in Figure 3.5. The variable profiles

suggest the proteins display very different primary, secondary, and tertiary structures and

that the gene product from clone A4 likely contains more hydrophobic domains than the

CAT enzyme encoded by pSU2719. In addition, the predicted ORFs from all six Cm^R clones were compared to known CAT genes from several bacterial hosts by multiple alignment (Figure 3.6). The ORFs did not align with the known CAT genes, although several conserved regions were observed among the CAT genes themselves. Because these genes are known to have a high degree of conservation even among diverse bacteria, we expected to observe some alignment between the clone ORFs and known genes. However, we observed very little conservation among the sequences analyzed, supporting our hypothesis that these clones harbor novel genes encoding CAT-like activity that are significantly divergent from known genes involved with Cm resistance.

2. **Plasmid-encoded antibiotic resistant determinants.** Functional screening of community plasmid DNA identified a 4.2 kb multi-drug resistant plasmid, pAS1, which was recovered from both direct and transposon-rescued transformations into *E. coli*. This plasmid conferred both kanamycin and ampicillin resistance on naïve *E. coli* and showed evidence of mediating the transposition of both resistance genes into the host genome (Figure 3.7). A 900 bp fragment present in plasmid preparations from primary transformations was absent in all plasmid preparations from downstream transformations, yet the antibiotic resistances persisted. Secondary transformations performed with pAS1 lacking the 900 bp fragment did not result in any clones resistant to either antibiotic, which supports the hypothesis that these resistance genes may be easily transposed from plasmid to host genome.

Shotgun subcloning and sequence analysis of pAS1 identified two predicted gene products most likely responsible for the antibiotic resistance phenotypes, an aminoglycoside phosphotransferase from the genus *Rhodopseudomonas* (α-

Proteobacteria) and a TEM β-lactamase from the genus *Bacillus* (*Firmicutes*) (Table 3.3). Further evidence for the transposition of kanamycin and ampicillin resistance into the host genome was obtained from the sequencing of pAS1 shotgun subclones, which identified two gene fragments sharing 100% homology with two unique *γ-Proteobacterial* transposases (Table 3.5).

3. Viral-encoded antibiotic resistance determinants. Functional screening of the viral metagenomic libraries did not yield any antibiotic resistant clones. The small average insert size of these clones (662 bp) likely contributed to the inability to isolate any clones with intact resistance determinants. Therefore, a sequence-based approach was employed to identify phage-encoded resistance genes from the 1,161 phage metagenomic clones. Six unique clone inserts were identified that had homology with antibiotic resistance genes in the GenBank nr/nt database, representing bacteria from four distinct taxa, including the *Bacteroidetes* and *Chlamydiae* divisions and two sub-divisions of the *Proteobacteria* (Table 3.3). These clones contain putative antibiotic resistance determinants which vary in both function and phylogeny when compared to those encoded by bacterial chromosomes or plasmids. Based on BLASTx data, these clones appear to carry partial genes that may be responsible for resistance to several antibiotics, including tetracycline, ampicillin, acriflavin, and bleomycin, as well as efflux systems that may mediate resistance to additional antibiotics. None of the clones identified by sequence analysis conferred their predicted Ab^R to *E. coli* when plated onto media containing the respective Ab.

4. Isolation of mobile genetic elements (MGEs) from bacterial, plasmid, and phage metagenomes. Sequence analysis of DNA from all three metagenomic surveys was

performed to identify gene products potentially involved in the lateral transfer of antibiotic resistance genes (Table 3.5). Clone E5, isolated from the bacterial metagenomic library, contains an insert sequence with 90% homology to a transposase from *Acidovorax avenae*, a bacterium classified in the *β-Proteobacteria* sub-division. This clone also carries a functional O-methyltransferase that confers kanamycin resistance on *E. coli* but shares significant homology with a gene product from the genus *Microscilla* of the *Bacteroidetes* division. The close proximity of the transposase and O-methyltransferase genes within the E5 clone insert suggests that the O-methyltransferase gene, and therefore kanamycin resistance, may be transposed among bacterial hosts.

As mentioned previously, the multi-drug resistance plasmid pAS1 also carried genes with 100% homology to *γ-Proteobacteria* transposases, indicating its ability to shuttle genes between bacterial hosts not only by plasmid transfer but also transposition of specific regions of DNA into the host genome. The functionality of at least one of these transposases was demonstrated by the *in vitro* transposition of kanamycin and ampicillin resistance genes into the *E. coli* host genome. Interestingly, the antibiotic resistance determinants encoded by pAS1 show strong homology with those from bacteria in the *α-Proteobacteria* and *Firmicute* divisions, neither of which was indicated by sequence analysis of the MGEs from pAS1.

Bacteriophages, which in themselves mediate the transduction of genes between bacterial hosts, also harbored a phylogenetically diverse group of putative transposases identified from the viral metagenomic libraries (Table 3.5). These transposases may be responsible for the transfer of various genetic loci between diverse bacterial hosts, as

evidenced by the number of bacterial divisions represented by the antibiotic resistance gene and MGE sequences from phage-derived clones.

E. DISCUSSION

Because activated sludge in wastewater treatment plants is a microbial system that constantly receives an influx of organic material and subsequently releases microbiota and genetic material into the environment, the occurrence of antibiotic resistance determinants in this assemblage and their potential for horizontal gene transfer are of great public and environmental health concern. The phylogenetic diversity of activated sludge has been shown to be very complex, consisting of many bacteria that are recalcitrant to cultivation in the laboratory, such as members of the *Bacteroidetes*, *Acidobacteria*, and deeply-branching *Proteobacteria* clades (75, 119, 193). Previous studies have demonstrated the limitations of culture-based methods for the identification and characterization of the microbial assemblages in activated sludge (6, 247); therefore, a metagenomic approach was used to identify antibiotic resistance genes that may not have been detected by culture-based methods alone. To access multiple reservoirs of antibiotic resistance genes, functional screening and sequence analysis were applied to metagenomic libraries constructed from bacterial chromosomal, plasmid, and bacteriophage community DNA.

Functional screening of bacterial metagenomic libraries yielded nine unique clones that conferred to *E. coli* resistance to one of three antibiotics: chloramphenicol, ampicillin, or kanamycin. Although BLASTx analysis indicated that the ampicillin and kanamycin resistance determinants were similar to previously known resistance genes in the GenBank nr/nt database, only one of the chloramphenicol resistant clones shared

significant homology with a predicted gene product that is generally associated with Cm^R. Subsequent transposon mutagenesis to identify the genetic loci responsible for Cm^R in the remaining clones confirmed that none of the interrupted ORFs showed similarity to chloramphenicol resistance genes in GenBank, which likely originated from diverse bacterial lineages. However, preliminary functional studies to determine the mechanism of Cm^R suggest that these clones harbor resistance determinants that have Cm-modifying enzymatic activity. Further investigations into the specific enzymatic modification and resultant structural changes to chloramphenicol by these clones using MS/NMR and various *in vitro* assays will be useful in identifying the mechanism(s) of resistance and perhaps attributing a novel phenotype to predicted proteins in the GenBank databases.

The multi-drug resistant plasmid pAS1, encoding kanamycin and ampicillin resistance, was isolated from both directly transformed plasmid DNA as well as plasmid DNA that was previously mutagenized with the <R6Kγ*ori*/KAN-2> transposon to mediate their replication in *E. coli*. These results indicate that pAS1 contains an origin of replication that allows for its maintenance in *E. coli*, and presumably other Gram-negative hosts. Interestingly, the TEM β-lactamase gene found on pAS1 shares 95% homology with a similar gene from the Gram positive genus *Bacillus*, suggesting the occurrence of previous horizontal gene transfer events between diverse bacterial hosts. The significant degree of homology observed for the Ab^R genes identified on pAS1 is in contrast to the poor or nonexistent homology observed for the bacterial chromosome-derived Ab^R clones. This is presumably the result of the much more frequent transmission of plasmid-borne Ab^R within wastewater, other natural environments, and

bacterial pathogens resulting in the frequent deposition of plasmid-acquired Ab^R genes in GenBank databases. The lower degree of homology for Ab^R genes present within bacterial genomes also reflects the latent functional genetic potential resident within wastewater metagenomes (e.g., novel Cm^R determinants) that have the potential for future lateral transfer. In addition, the greater degree of homology observed between the plasmid-encoded resistance genes (as opposed to the chromosomal-encoded genes) and those found in GenBank supports the hypothesis that plasmids may play a more significant role in the lateral transfer of antibiotic resistance genes among prokaryotes (93).

Although no antibiotic resistance clones were isolated from functional screening of the viral metagenomic libraries, sequence analysis identified six unique putative resistance determinants that may encode direct resistance to tetracycline, bleomycin, acriflavin, and ampicillin, as well as drug resistance transport systems that may confer antibiotic resistance(s) on their hosts.

The phylogenetic diversity of antibiotic resistance determinants identified in this survey and the presence of multiple MGEs identified from all three metagenomes indicate that activated sludge does serve as a reservoir of Ab^R genetic determinants that may be laterally transferred among diverse bacterial hosts. Seven distinct bacterial taxa were represented in our metagenomic surveys, including the *Bacteroidetes*, *Chlamydiae*, and *Firmicute* divisions and all four sub-divisions of the *Proteobacteria*. Although all of the MGEs identified from the surveys are likely derived from *Proteobacteria* taxa, the observation that they were often linked with antibiotic resistance loci with a high degree of homology to Ab^R genes from other phyla suggests that genes from phylogenetically

122

diverse sources have been transferred through multiple mechanisms (e.g., conjugation, transduction, transformation). Along with the *in vitro* transposition of kanamycin and ampicillin resistance genes into *E. coli* observed with pAS1, these data reflect the great potential for the mobilization, transfer, and persistence of diverse antibiotic resistance determinants among various bacterial hosts.

The application of sequence- and function-based screening of bacterial, plasmid, and bacteriophage metagenomic libraries allowed a more inclusive study into the occurrence, phylogenetic origins, and potential transfer of Ab^R genes in activated sludge than has been previously described. The discovery of these resistance genes illustrates an advantage of function-based screening methods, as many of these Ab^R genes likely would not have been identified by sequence-based screening alone. It should be noted that this survey was far from exhaustive nor necessarily representative of all of the major phyla present within activated sludge microbial assemblages. While the selection of *E. coli* as a heterologous host afforded advantages in terms of cloning efficiency and phylogenetic relationship to many of the bacterial taxa present in activated sludge, future studies using shuttle vectors may expand the diversity of MGEs and Ab^R loci identified. Future work will also identify the mechanism of action for each of the novel Cm^R genetic determinants. The identification of these functionally and phylogenetically diverse antibiotic resistance determinants and MGEs from all three metagenomes demonstrates the widespread occurrence and potential movement of antibiotic resistance that are a driving force in prokaryotic evolution within activated sludge microbial communities.

Library	# of clones	Avg. insert size	Cloned DNA
LCKan-1	312,800	2.1 kbp	~656 Mbp
LCAmp-1	380,000	1.67 kbp	~634 Mbp
LCKan-2	156,000	1.66 kbp	~259 Mbp
LCAmp-2	174,400	1.74 kbp	~303 Mbp
TOTAL	**1,023,200**	**1.79 kbp**	**~1.85 Gbp**

Table 3.1. Cloning statistics of the small-insert bacterial metagenomic libraries.

Antibiotic	Class	Concentration
Ampicillin	Pencillins	100 µg/mL
Azithromycin	Macrolides	10 µg/mL
Ceftriaxone	Cephalosporins	25 µg/mL
Chloramphenicol	50S Inhibitors	12.5 µg/mL
Ciprofloxacin	Fluoroquinolones	10 µg/mL
Kanamycin	Aminoglycosides	50 µg/mL
Nalidixic Acid	Quinolones	15 µg/mL
Polymyxin B	Polypeptides	8 µg/mL
Rifampicin	Rifamycils	150 µg/mL
Spectinomycin	Aminocyclitols	100 µg/mL
Tetracycline	Tetracyclines	25 µg/mL
Trimethoprim	DHFR Inhibitors	1 µg/mL

Table 3.2. Antibiotics used in the functional screening of metagenomic libraries and plasmids.
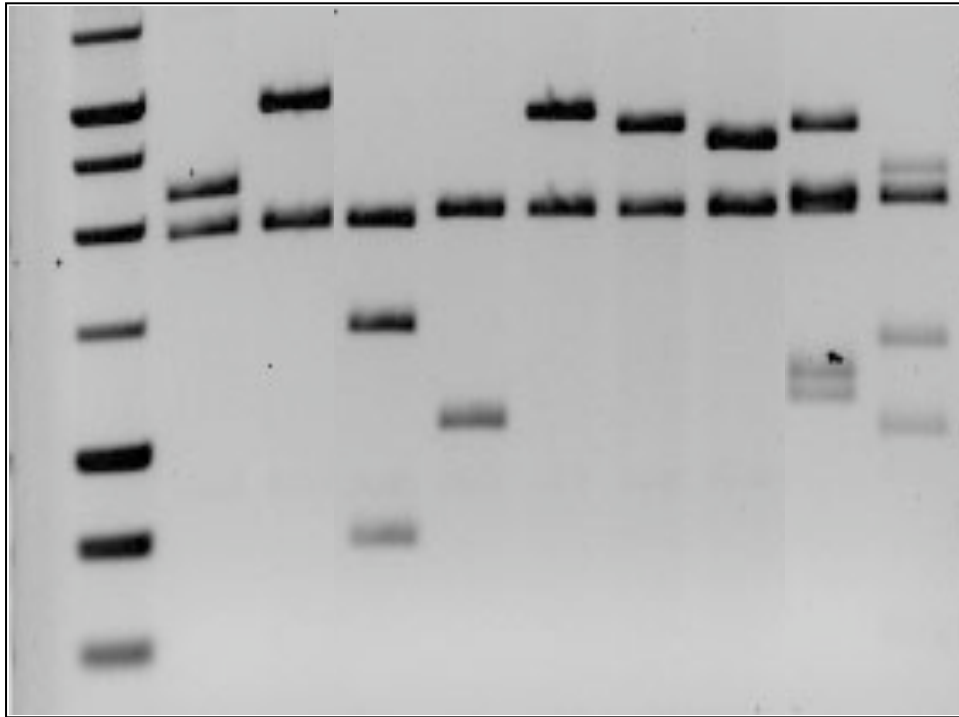
Figure 3.1. Representative RFLP patterns of unique bacterial metagenomic clones conferring an antibiotic resistance phenotype on *E. coli.*

Clone	Source	Ab^R	Top BLASTx	Nearest Neighbor	% ID
A4	Bacterial genome	Cm	PKD-repeating protein	*Cytophaga hutchinsoni*	31%
A5	Bacterial genome	Cm	No significant hits found	N/A	N/A
A7	Bacterial genome	Cm	Hypothetical protein FB2170	Flavobacteriales bacterium HTCC2170	59%
A9	Bacterial genome	Cm	Hypothetical protein sce4763	*Sorangium cellulosum*	57%
A10	Bacterial genome	Cm	Large subunit of N,N-dimethylformidase	*Hoeflea phototrophica* DFL-43	29%
A12	Bacterial genome	Cm	Class V aminotransferase	*Hyphomonas neptunium*	33%
B4	Bacterial genome	Cm	Hypothetical protein Mpe_A2882	*Methylibium petroleiphilum*	67%
D10	Bacterial genome	Amp	PBP, Class C β-lactamase	*Spirosoma linguale*	66%
E5	Bacterial genome	Km	O-methyltransferase	*Microscilla* sp.	55%
pAS1	Plasmid	Km	Aminoglycoside phosphotransferase	*Rhodopseudomonas palustris*	70%
pAS1	Plasmid	Amp	TEM β-lactamase	*Bacillus* sp. BT-192	95%
Ph-A8	Phage genome	N/A	Acriflavin resistance protein	*Xanthobacter autotrophicus*	53%
Ph-C2	Phage genome	N/A	Drug resistance transporter, Bcr/CflA subfamily	*Serratia proteamaculans*	96%
Ph-C3	Phage genome	N/A	putative transmembrane multidrug-efflux system lipoprotein	*Ralstonia pickettii*	58%
Ph-D8	Phage genome	N/A	Class A β-lactamase	*Capnocytophaga ochracea*	93%
Ph-E10	Phage genome	N/A	TetC protein	*Chlamydia suis*	89%

Table 3.3. Predicted gene products and phylogeny of putative antibiotic resistance determinants from metagenomic clones and plasmid pAS1 based on BLASTx output.
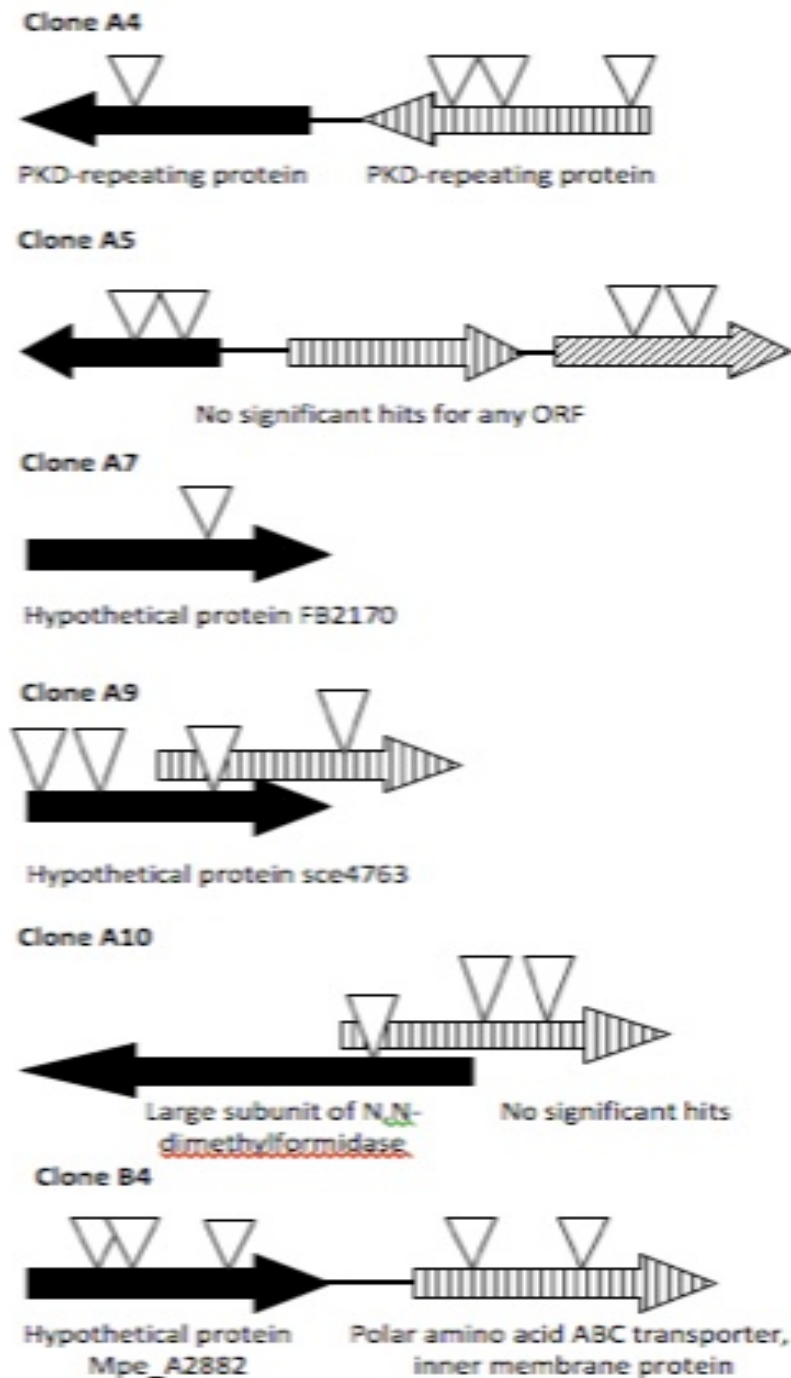
Figure 3.2. Transposon insertion maps of metagenomic clones that confer Cm^R to *E. coli* but have no homology with previously characterized Cm^R genes. Triangles indicate the transposon insertion site which resulted in loss of the antibiotic resistance phenotype.
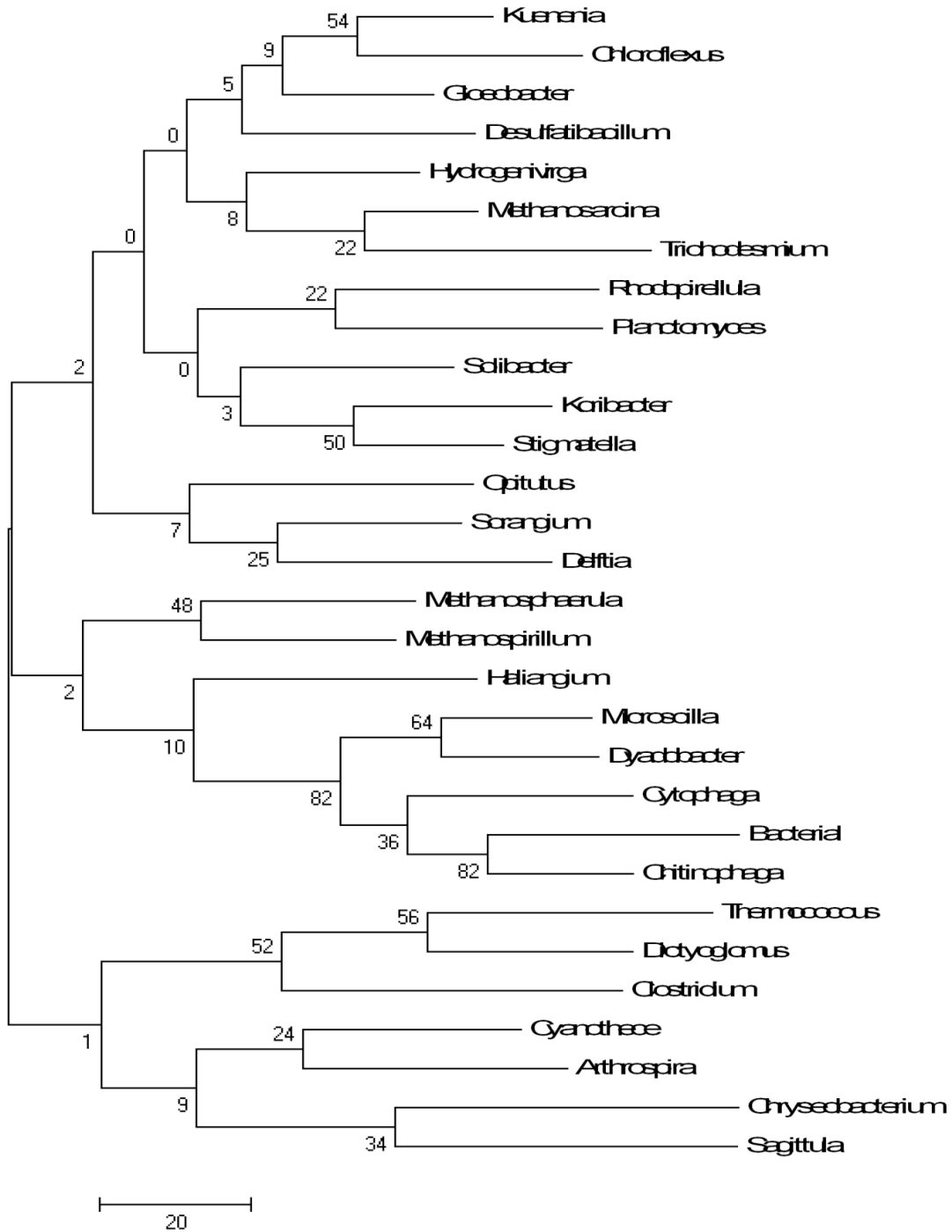
Figure 3.3. Phylogenetic relationships between Cm^R clone A4 and significant homologs in the nr/nt GenBank database were predicted using maximum parsimony analysis. Bootstrap values are shown on the nodes of the trees, and each tree was rooted with the sequence of a distantly related but non-significant homolog.
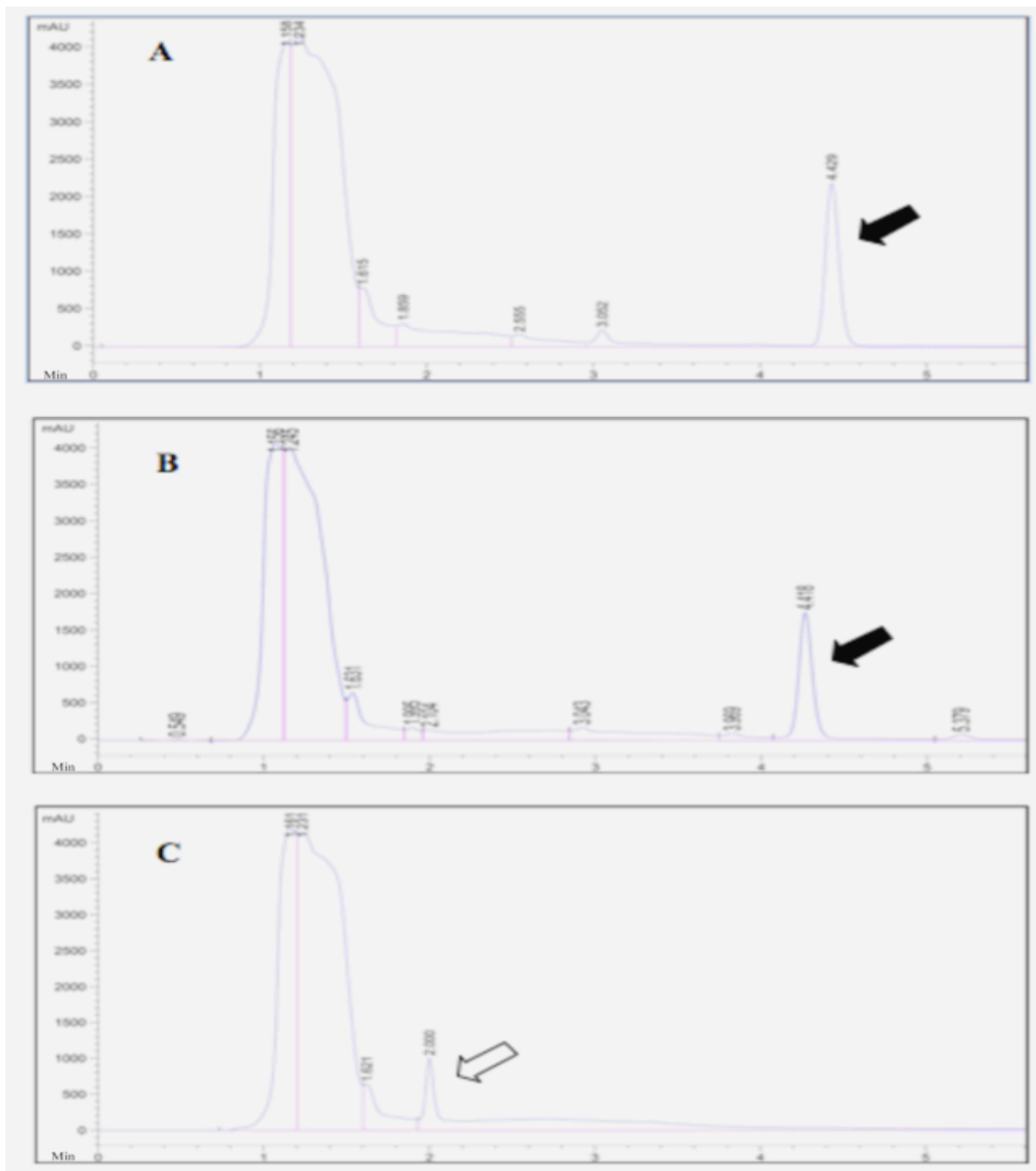
Figure 3.4. HPLC profiles of Clone A4 (Panel A) and pSU2719 (Panel B) supernatants from cultures grown overnight in the presence of Cm. Panel C represents the negative control in which chloramphenicol only was incubated overnight. Black arrows indicate the appearance of a byproduct with a peak at 4.4 minutes, and the white arrow indicates the native chloramphenicol peak at 2.0 minutes.

Clone	MIC (μg/ml)
A4	64
A5	64
A7	256
A9	64
A10	64
B4	256
E. coli ATCC 25922	4

Table 3.4. Mininum inhibitory concentration (MIC thresholds for the six Cm[R] clones with no significant homology to known resistance determinants and *E. coli* ATCC 25922 (reference strain).
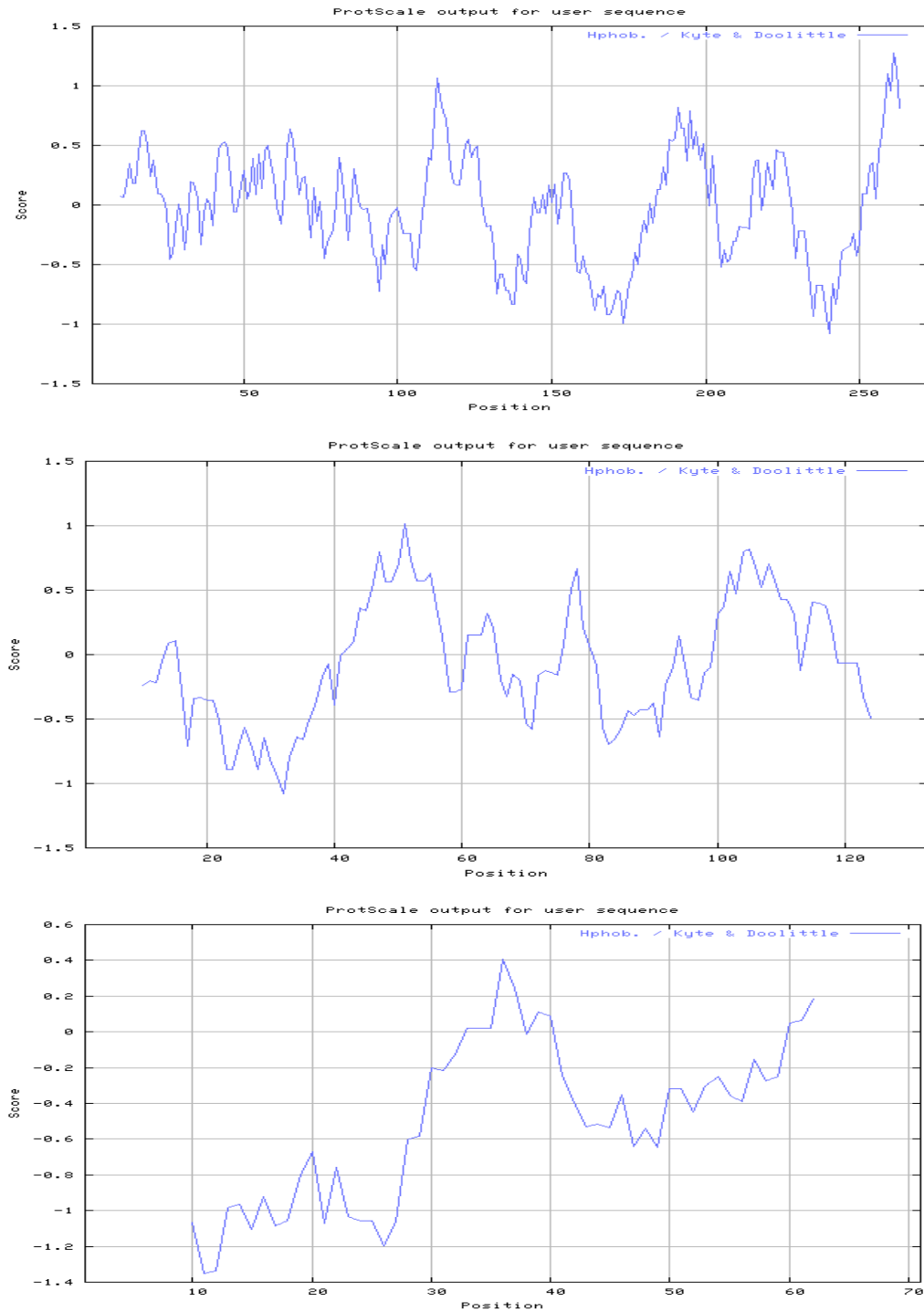
Figure 3.5. Hydrophobicity plots of predicted ORFs conferring Cm^R on clone A4 (Panels A and B) and the CAT gene on plasmid pSU2719 (Panel C).

132
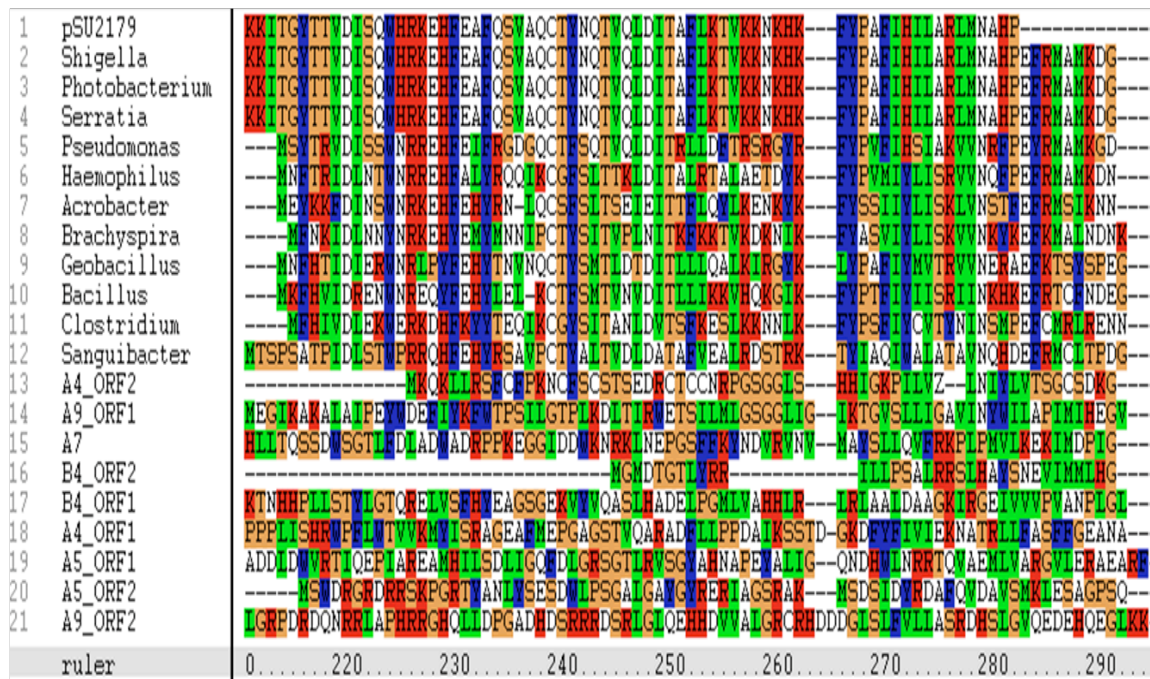
Figure 3.6. Multiple alignment of ORFs conferring Cm^R on the bacterial metagenomic clones, the CAT enzyme of pSU2719, and phylogenetically diverse CAT gene products from the GenBank nr/nt database. The alignment was performed in ClustalX.
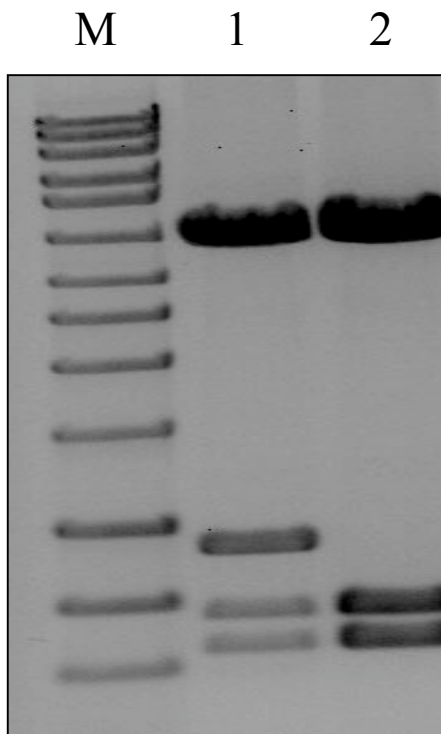
Figure 3.7. Predicted transposition of a 900 bp element from pAS1 into the host

genome. Shown are EcoRI restriction digests of pAS1 following primary transformation

(Lane 1) and secondary transformation (Lane 2). The 900 bp fragment seen in Lane 1

was absent in all downstream plasmid preparation, but a loss of antibiotic resistance was

not observed during the screening of transformants, indicating the transposition of

antibiotic resistance gene(s) into the *E. coli* host genome. (M, molecular marker)

Sample	Source	Top BLASTx	Nearest Neighbor	% Identity
E5	Bacterial genome	Transposase	*Acidovorax avenae*	90%
pAS1	Plasmid	IS10 transposase	*Salmonella enteric*	100%
pAS1	Plasmid	ISPsy10 transposase	*Vibrio sp*. Ex 25	100%
Ph-A11	Phage genome	transposase protein A	*Methylococcus capsulatus*	50%
Ph-C11	Phage genome	transposase	*Nitrobacter* sp. Nb-311A	70%
Ph-C12	Phage genome	transposase	*Gluconobacter oxydans*	75%
Ph-E06	Phage genome	transposase, IS605 OrfB family	*Burkholderia cenocepacia* MC03	33%
Ph-F11	Phage genome	transposase of ISAba1, IS4 family	*Acinetobacter baumannii*	99%
Ph-G04	Phage genome	TniA transposase	*Pseudomonas stutzeri*	34%

Table 3.5. Predicted gene products and phylogeny of putative mobile genetic elements identified from metagenomic clones and plasmid pAS1.

CHAPTER IV

DISCOVERY OF TYPE I POLYKETIDE SYNTHASE PATHWAYS FROM AN

ARRAYED SOIL METAGENOMIC LIBRARY

A. ABSTRACT

Polyketides are structurally diverse bacterial secondary metabolites, many of which have

antibiotic or anti-cancer activity. Modular polyketide synthase (PKS) enzymatic

complexes contain conserved ketoacyl synthase (KS) domains, and most PKS

biosynthetic pathways exceed 30 kb in size. A fosmid metagenomic library constructed

from soil at the Hancock Agricultural Research Station in Hancock, WI, (9,216 clones

with an average insert of 42 kb) was spotted onto a nylon membrane. The macroarray

was then screened for the presence of KS domains using a degenerate DNA probe

targeting a conserved region of this domain. Thirty-four clones containing KS domains

were identified by Southern hybridization; however, only 21 of the 34 KS-positive clones

produced a PCR product using KS-specific primers, indicating that these domains may be

related yet phylogenetically divergent from known KS genes. Many of the KS amplicons

shared significant homology with PKS or non-ribosomal peptide synthesis genes from the

Cyanobacteria or *Proteobacteria* phyla. Sequence analysis from a KS-containing clone

that was consistently PCR-negative revealed PKS-like domains on the cloned fragment,

which is hypothesized to have an origin from the bacterial division *Acidobacteria*, which is prevalent within soils yet has few cultured representatives.

B. INTRODUCTION

The isolation of secondary metabolites produced by soil microorganisms has historically been an effective means of identifying novel compounds with antimicrobial activity. Soil contains an impressive 10^{16} microbial cells per ton, while only an estimated 1% of these have been cultivated in the laboratory (58, 108). While this small percentage of cultured soil microorganisms has already illustrated the phylogenetic diversity of these communities and have provided many important antibiotics to date, a potentially richer resource for novel antibiotic discovery resides within as-yet-uncultured bacteria (87, 94, 218). The application of a metagenomic approach to isolate large, contiguous regions of DNA from uncultured microorganisms is a viable strategy to access intact genetic pathways that are involved in natural product synthesis. Previous studies have demonstrated the utility of screening environmental metagenomic libraries to identify bioactive molecules (54), proteolytic systems (23), and polyketide synthases (89, 167, 226, 286) .

Polyketides are a diverse group of bioactive secondary metabolites produced by bacteria, fungi, and plants and include several therapeutically important drugs such as erythromycin, tetracyclines, immunosuppressants FK506 and rapamycin, and antitumor compounds doxorubicin and mithramycin (226, 253). Polyketide biosynthetic genes are organized into polyketide synthase (PKS) pathways and are generally classified into three major groups: Type I, Type II, or Type III (209). Type I PKSs contain multi-functional proteins that are each responsible for a different active site for catalytic

reactions in polyketide chain assembly (170). The actinorhodins, or Type II PKSs, are composed of individual proteins that perform one enzymatic activity to catalyze the formation of aromatic polyketides (170, 237). Type III PKSs, previously thought to be found only in plants, participate in the assembly of small aromatic compounds (170). This study focuses on the identification of Type I PKSs, which are large, multi-domain enzymes organized into modules, each of which contains an individual enzyme responsible for the catalysis of polyketide chain elongation. Type I PKSs are characterized by the presence of a 6-deoxyerythromycin B synthase for the biosynthesis of reduced polyketides such as erythromycin A (237), and contain loading and extender modules. The extender module is further organized into three domains that catalyze one cycle of chain extension: ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP). Some modules also carry variable domains, such as ketoreductase, dehydratase, thioesterase, and enoylreductase , which are responsible for keto-group modification and lead to the synthesis of a wide range of combinatorial compounds (26, 65).

Although the synthesis of polyketides is often associated with bacteria within the division *Actinobacteria* (195), PKS genes have also been identified in a diverse group of bacteria, including species of *Pseudomonas* (154, 205), *Mycobacterium* (126), *Stigmatella* (54), *Sorganium* (85, 86), and *Acidobacteria* (285). The widespread distribution of PKS pathways is of interest because additional reservoirs of polyketides in soil microorganisms likely exist but have remained uncharacterized due to cultivation bias. Therefore, we used degenerate probes to identify potentially novel PKS pathways from a large-insert soil metagenomic library via macroarray hybridization. The isolation

and manipulation of PKS pathways offers opportunities for the production of novel compounds via manipulation of the modular organization of Type I PKSs for synthesis of novel metabolites. These PKS pathways can therefore provide another resource for chemical diversity in the search for novel antimicrobial compounds (226).

C. MATERIALS AND METHODS

1. Soil collection and DNA isolation. Soil samples were collected from soil cores (10 cm to 50 cm depth) from an agricultural soil in the Central Sands area of Wisconsin at the University of Wisconsin-Madison's Hancock Agricultural Research Station (HARS) in Hancock, WI. The soil was a Plainfield loamy sand taken from the plow zone at HARS. Bacterial cells were collected by soil homogenization and differential centrifugation and then embedded and lysed within an agarose plug. High molecular weight metagenomic DNA was recovered and purified as described previously (143, 144).

2. Fosmid library construction. The metagenomic library was constructed in the pCC1FOS vector with the CopyControl™ fosmid library system according to the manufacturer's instructions (Epicentre, Madison, WI). Briefly, purified metagenomic DNA was randomly sheared and size-selected from a pulsed-field agarose gel prior to ligation into the pCC1FOS vector. Ligated DNA was then packaged and plated onto the EPI100™-T1 plating strain (Epicentre) and plated onto LB + 12.5 μg/ml chloramphenicol for selection of recombinant clones. Arabinose was used to control the copy number of pCC1FOS by activating the P_{BAD} promoter and *trf*A gene, which in turn induces expression of *ori*V; therefore, the addition of arabinose (0.01% v/v) to media will induce a high copy number of the fosmid. The recombinant *E. coli* clones were robotically picked in duplicate onto nylon membranes at the Clemson University

Genomics Institute, resulting in duplicate clone macroarrays. Macroarrays were generated with and without arabinose-induction of fosmid copy number and contained 9,216 fosmid clones with an average insert size of 42 kb (ca. 100 *E. coli* genome equivalents).

3. Screening of fosmid macroarrays. Pooled fosmid DNA served as template in PCR reactions to generate a PKS probe for downstream library screening. The β-ketosynthase (KS) domain was targeted with the degenerate primer set 5LL (5'-GGRTCNCCIARYTGIGTICCIGTICCRTGIGC) and 4UU (5'-MGIGARGCIYTICARATGGAYCCICARCARMG) (David Sherman, University of Michigan, personal communication). Seven other Type I and Type II PKS-specific and non-ribosomal peptide synthetase-specific primer sets were also evaluated for probe synthesis (Table 1) but were not used because they either yielded PCR products from control reactions containing *E. coli* genomic DNA as template or did not yield the expected PCR product based on sequences of representative amplicons. The resulting KS domain amplicon was labeled with digoxigenin (DIG) using the DIG PCR Synthesis system (Roche, Indianapolis, IN) and used as a probe in macroarray screening by colony blot hybridization at 42°C with medium hybridization washing stringency as per manufacturer's instructions. The hybridization was performed using the DIG detection system (Roche) with detection mediated by the CSPD chemiluminescent substrate. Imaging of hybridizations was accomplished using autoradiography with multiple exposure times for each blot. Clones that exhibited chemiluminescence in duplicate (n=34) were considered KS-positive and subjected to further analyses.

4. Verification and sequencing of KS domains from positive clones. To confirm that
the 34 hybridization-positive clones were unique and to estimate insert size, restriction
fragment length polymorphism (RFLP) profiles were generated for each clone using the
restriction enzyme BamHI. Resulting digest reactions were electrophoresed on a 0.8%
(w/v) 1x TAE pulsed-field agarose gel at 5 V/cm with a 1 to 15 second switch time for 16
hours.

Twenty-one of the KS-positive clones produced a KS amplicon following PCR
with the 5LL/4UU primer set to verify the presence of a KS domain. To obtain sequence
data from these clones, the KS primers were modified with T3 (5LL) and T7 (4UU)
primer recognition sequences at the 5'-end of each respective primer and used in a
secondary PCR to generate KS amplicons for sequencing. The KS genes were then
sequenced using primers T3 and T7, which resulted in higher quality sequence data than
using either the degenerate 5LL or 4UU primers. All 34 KS-positive clones were end-
sequenced using vector primers EpiFOSF and EpiFOSR.

5. KS domain phylogenetic analysis. KS gene sequences from the 21 PCR-positive
clones and known KS gene sequences from bacterial and fungal sources were aligned
using ClustalX software. A maximum parsimony phylogenetic tree was generated using
MEGA4 (131) with 1000 iterations of the maximum parsimony analysis performed for
the calculation of bootstrap support.

6. Southern hybridization of putative KS-positive clones. To investigate the 13
clones that were KS-positive after colony blot hybridization but failed to yield a KS PCR
product, a Southern hybridization was performed using the same DIG-labeled
heterogeneous KS probe used in previous experiments. Fosmid DNA from each of the

13 clones was restriction digested with BamHI, electrophoresed on a 1% (w/v) 1x TAE agarose gel, and transferred to a nylon membrane using the Whatman TurboBlotter system (Kent, UK) and hybridized with the KS probe. Identification of DNA fragments hybridized to the KS gene probe was visualized by the NBT/BCIP colorimetric detection method (Roche).

8. Isolation and 454 sequencing of fosmid DNA. Ten clones were selected for complete insert sequencing by 454 pyrosequencing based on their predicted phylogenetic and functional domains. Fosmid DNA was extracted from clones A2, A3, A11, A12, B1, B8, B10, B11, C1, and C5 using the Qiagen Large-Construct DNA Isolation kit (Qiagen, Valencia, CA) with an additional plasmid-safe exonuclease digestion step for 4 h at 37ºC to remove contaminating genomic DNA. Purified DNA was sent to the Lucigen Corporation (Middleton, WI) for bar-coded shotgun subclone library construction and then sequenced at the Engencore Center at the University of South Carolina (Columbia, SC) using a Genome Sequencer FLX system (Roche, Nutley, New Jersey) according to manufacturer's instructions.

9. Bioinformatic analysis of fosmid-derived open reading frames. Fosmid DNA sequences generated by 454 sequencing were assembled into contiguous fragments (contigs) using the CLC Genomics Workbench (Cambridge, MA) assembly algorithm and contigs were analyzed for the presence of putative PKS-related genes. Open reading frames were identified using a GeneMark heuristic approach for gene prediction in prokaryotes (http://opal.biology.gatech.edu/GeneMark/). Additionally, GLIMMER 3.02, and NCBI's ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) were utilized to corroborate the predicted ORFs obtained from GenMark analysis. Prediction of gene

function was done by comparing each phage ORF sequence against the GenBank nr/nt database using the BLASTp and BLASTn algorithms (3). Any predicted ORFs lacking any significant homology to other gene products in GenBank (E value > 0.001) were further investigated for secondary structures (profiles, patterns, blocks, motifs, and protein families) using a web server (http://motif.genome.jp/).

10. Analysis of clone A12. Preliminary sequence data from clone A12 indicated that this clone may have significant homology to the genome of *Solibacter usitatus* from the division *Acidobacteria*. Because the *mtaD* homologous gene sequence did not assemble with other contigs from this clone during the first sequencing analysis, multiple approaches were utilized to verify and characterize its presence on the clone.

Primers specific to the putative PKS gene were designed and used to generate a DIG-incorporated PCR probe for Southern hybridization. Purified DNA from A12 was restriction digested with the enzymes BamHI, ClaI, EcoRI, EcoRV, Hind III, NotI, PstI, SalI, SmaI, SphI, and XbaI and electrophoresed on a 0.8% (w/v) 1x TAE agarose gel. The digested DNA was transferred to a nylon membrane and hybridized with the *mtaD* homolog probe as described above for the KS domain probe.

To obtain additional data on the regions flanking the *mtaD* homologous gene, shotgun subclone libraries of the A12 insert were constructed. Purified A12 DNA was partially digested with Sau3AI for 1 h, electrophoresed on a 1% (w/v) 1x TAE agarose gel, and size-selected. Fragments approximately 5-8 kb were ligated into the BamHI-cut CopyRight BAC vector (Lucigen Corp., Middleton, WI) overnight at 15ºC. The ligation mix was transformed into BAC-optimized electrocompetent *E. coli* cells, and transformants were selected on 2x YT agar supplemented with 12.5 µg/ml of

chloramphenicol (Cm). Ten random subclones were chosen for library validation. BAC

DNA was extracted from 10 clones by alkaline lysis and digested with EcoRI at 37ºC for

1 h. The percentage of clones with an insert and the average insert size was determined

by electrophoresing the restriction digests on a 1% (w/v) 1x TAE agarose gel.

The remaining transformation mix was plated onto additional 2x YT+Cm medium

and incubated overnight at 37ºC. Colonies were transferred to nitrocellulose membranes

and hybridized with the DIG-labeled *mtaD* homolog probe as described above.

Subclones demonstrating a color change after overnight detection were sequenced and

included in additional contig assemblies.

11. Detection of *mtaD* homologs from cultured *Acidobacteria*. Cultures representing

multiple subdivisions of *Acidobacteria* (George *et al*., manuscript in preparation) were

screened for *mtaD* homologous sequences by PCR using *mtaD*-specific primers (Table

1). The following touchdown thermalcycling program was used: initial denaturing at

94ºC for 2 min, followed by 14 cycles of denaturing at 94ºC for 30 sec, stepwise

annealing from 63ºC to 50ºC for 30 sec, and extension at 72ºC for 2 min, and an

additional 30 cycles of the above with a 50ºC annealing temperature. Amplicons were

subcloned into the TOPO-TA pCR 2.1 vector (Invitrogen, Carlsbad, CA) according to the

manufacturer's instructions, and inserts were amplified in a colony PCR using vector

primers M13F and M13R. The resulting PCR reactions were purified using the Promega

Wizard PCR Clean-up System (Madison, WI) and submitted to the Auburn University

Genetic Analysis Laboratory for sequencing. The sequences were trimmed for quality

and compared to their nearest significant neighbors in GenBank by multiple alignment

(ClustalX) and maximum parsimony analysis (MEGA4) (131, 264).

D. RESULTS

1. Generation of a KS-specific probe. Multiple primer sets specific to PKS or NRPS pathways were used in this study to prepare a probe for macroarray hybridization (Table 4.1).   In each experiment, a pooled fosmid library DNA template was used as a template with each primer set. Every primer set that generated a product from library template (but not *E. coli* DNA template) was cloned and a representative number of clones (n=8) was sequenced to validate that the amplicon was the intended gene target. The only primer set that produced the desired PCR amplicon using a metagenomic library template and not with host genomic DNA template was primer set 4UU/5LL, which is specific to the ketosynthase (KS) domain, the most conserved region of Type I PKS pathways (11, 93).   This universal primer set was selected for DIG-labeling and subsequent macroarray hybridization.

2. Identification of Type I PKS-containing clones from the fosmid library. A total of 34 clones from the arabinose-induced macroarray exhibited chemiluminescence after hybridization with the DIG-labeled Type I KS probe (in duplicate), indicating the presence of a KS domain and therefore a possible PKS pathway (Fig. 4.1). Only 19 KS-containing clones were identified from the macroarray prepared without arabinose induction, which were also detected in the arabinose-induced macroarray in addition to 15 other clones. The position of the duplicate clones relative to one another enabled determination of the specific fosmid clone identity, and RFLP profiles of the 34 clones confirmed they contained unique inserts (Fig. 4.2).   Of the 34 fosmid clones that were identified by hybridization, only 21 of these clones yielded a PCR product after a

secondary PCR with primers 4UU and 5LL, suggesting the remaining 13 clones may contain KS domains which are phylogenetically divergent from known genes.

3. Phylogenetic analysis of KS-positive clones. DNA sequences derived from KS amplicons obtained from the 21 hybridization-positive clones were compared to the GenBank nr/nt database and were found to have homology with known KS genes (identities ranging from 51-73%). As the highest percent homology observed was 73%, it is likely that these KS domains are phylogenetically distinct from the previously characterized KS domains present in GenBank. Maximum parsimony analysis of library-derived KS domains and known KS domains indicates that the KS domains identified from the soil metagenomic library represent diverse lineages but are still related to known KS domains (Fig. 4.3). Interestingly, the majority of the library-derived KS domains had nearest neighbors among the phylum *Cyanobacteria*. Even though this bacterial division has been previously associated with polyketide production, we did not expect to find a high frequency of *Cyanobacteria*-derived sequences because the fosmid library was not constructed from surface soil. The KS domains encoded by these fosmid clones showed similarity to those found in pathways responsible for a variety of polyketide compounds, such as the myxobacterial compounds epothilone and stigmatellin (27), the cyanobacterial jamaicamide (70), and the actinobacterial polyketide antibiotics erythromycin (65), and pikromycin (300). Although many of these library-derived KS domains have similarity to known KS domains, the resultant chemical moieties produced by these pathways in their native host may be structurally distinct from known polyketides, as the other modules present in these PKS pathways may not be similar in organization to known PKS pathways.

4. **Southern hybridization of PCR-negative fosmid clones.** Of the 34 fosmid clones identified by the original library macroarray hybridization, 13 clones did not yield a KS domain PCR product despite multiple attempts at PCR amplification using a variety of reaction conditions. To determine if each of these fosmid clones actually contained a KS domain, a Southern hybridization was performed using a DIG-labeled KS domain probe generated by PCR amplification of the KS domains from a pool of the 21 KS domain-positive fosmid DNAs. Eight of the 13 PCR-negative clones gave KS-positive Southern blot results, thus verifying the presence of a putative KS domain on these clones. The five fosmid clones which were originally considered KS-positive by macroarray hybridization but have not yielded a KS PCR product or shown a positive result in subsequent hybridizations may represent false positive macroarray hybridizations, misidentification of the fosmid clone from the metagenomic library stored at -80°C, or may require lower stringency hybridization conditions to be detected.

5. 454 sequencing of fosmid clones. Ten PCR-negative, hybridization-positive clones were selected for whole-insert sequencing by 454 pyrosequencing based on their predicted phylogeny and functional domains. To date, three of the ten clone inserts have been completely assembled and contain many open reading frames (ORFs) with similarity to gene products associated with polyketide synthesis, such as acyl-carrier proteins and beta-ketoacyl synthases, as well as cell transporters and translocation systems from various bacterial lineages (Figure 4.4). Preliminary annotation of the remaining seven clones has also revealed the presence of similar ORFs but also many sequences with no significant neighbors in GenBank or which have currently un-assigned functions.

147

6. Bioinformatic analysis of clone A12. 454 sequence data from clone A12 generated two points of interest. First, this clone appears to contain a gene with 52% identity and 70% similarity to a gene that is homologous to *mtaD*, known to encode an enzyme from the myxobacterium *Stigmatella aurantiaca* that is involved in chain extension during the synthesis of the polyketide myxothiazol (239). Secondly, the majority of DNA carried by A12 shares strong homology with the genome of *Solibacter usitatus*, a representative of the division *Acidobacteria* (Fig. 4.5).

However, the MtaD-like domain did not assemble with other contigs from A12, so Southern hybridization was performed to verify its presence on this clone, which indicated this may be a repeating domain within the clone (data not shown). Although the *mtaD* homologous gene sequence can be readily obtained from A12 by PCR, we have been unable to detect it in any additional 454 sequencing reactions or contig assemblies using a variety of bioinformatics pipelines.

7. Identification of *mtaD* homologous sequences from cultured *Acidobacteria*. After PCR-based screening of cultured *Acidobacteria*, we identified three *mtaD* homologs from *Acidobacteria* representing subdivisions 3, 4, and 6 (Fig. 4.6). Maximum parsimony analysis of these sequences, the *mtaD* homolog from A12, and other significant neighboring sequences from GenBank suggests that this gene (and possibly other PKS determinants) is highly conserved among the *Acidobacteria* but do not group significantly with any other specific bacterial phylum. *Solibacter usitatus* (nearest neighbor of clone A12) and cultured *Acidobacteria* J13 both represent subdivision 3, yet the *mtaD* homologous sequence from culture H6 (subdivision 4) appears most similar to

A12, indicating a widespread distribution of diverse *mtaD*-like genes among

Acidobacteria subdivisions.

E. DISCUSSION

We constructed and screened a large-insert fosmid metagenomic library from soil

to detect the presence of Type I polyketide synthase pathways using a ketosynthase-

specific probe. A sequence-based screening approach was selected in order to identify

PKS pathways derived from phylogenetically diverse prokaryotes that may not be

expressed in an *E. coli* host. Probing of the fosmid library, which is estimated to contain

over 400 Mbp of cloned metagenomic DNA, revealed the presence of 34 clones that

appeared to contain a KS domain. With the metagenomic library representing

approximately 100 *E. coli* genome equivalents and assuming equal relative abundance of

prokaryotic genomes within the library, then at this hit frequency approximately 40% of

the genomes represented within the library showed a KS domain positive hybridization.

Compared to previous estimates of KS domain-positive cultured soil bacteria of 15-25%

(209), this is a higher frequency, suggesting at the very least given the assumptions

necessary in this comparison that a culture-independent survey of PKS pathways is

capable of identifying a significant yield of PKS-containing clones. The other

acknowledged constraint in screening a fosmid library is the limited insert size possible

for any given clone. Future soil metagenomic libraries available and being constructed

exceed the average insert size of this fosmid library, but this study does represent a proof-

of-concept for the detection of KS-positive clones within a macroarrayed metagenomic

library and permit description of the phylogenetic origin and PKS pathway architecture

among these clones.

The majority of the 21 PCR-positive clones encoded KS domains most similar to those from the *Cyanobacteria* and *Myxobacteria* taxa and were phylogenetically and functionally diverse, including some expected to be involved in fatty acid synthesis and polyketide chain elongation. Additional bioinformatic analysis of complete insert sequences from the PCR-positive clones indicated many predicted ORFs with significant homology to sequences encoding β-ketosynthase domains, acyl carrier proteins, and enzymes involved in the post-synthesis modification of various polyketides. The predicted bacterial lineages of these sequences included the *Acidobacteria, Cyanobacteria,* and multiple subdivisions of the *Proteobacteria.* These data clearly demonstrate that many soil-derived PKS pathways are present in diverse Gram negative bacteria, and that the current public databases may not adequately reflect the extant diversity of soil bacterial PKS pathways.

A significant percentage of the 34 fosmid clones identified by library macroarray hybridization never yielded a KS domain-specific PCR product, leading to the initial conclusion that these may be false positive identifications or that the incorrect fosmid clones had been selected from the 384-well formatted library. Subsequent Southern hybridization analysis indicated that 8 of these 13 clones did contain a KS-homologous domain(s). It is likely possible that base pair mismatches between the KS domain PCR primers and fosmid clone DNA prevented successful PCR amplification of KS genes, yet they shared homology for successful hybridization. This is strong supportive data for the use of macroarray hybridization rather than PCR-based strategies for identification of metagenomic clones representing diverse phyla.

Sequence analysis of the PCR-negative clones revealed the presence of PKS-related ORFs that appear to originate from diverse bacterial lineages, such as the *Planctomycetes, Proteobacteria, Acidobacteria,* and *Cyanobacteria* phyla. Many of the predicted gene products appear to be involved in polyketide biosynthesis, post-synthesis modification, or the transport of such compounds outside of the cell. Due to the phylogenetically diverse nature of these sequences, it is unlikely that all of the pathways would have been functionally active in *E. coli*, thus underscoring the advantage of a sequence-based screening approach for the initial identification, to be followed by research to express PKS domains or pathways within a heterologous host(s).

Sequence analysis of clone A12 revealed the presence of an MtaD homologous domain, which in the myxobacterium *Stigmatella aurantiaca* is involved in the hybrid PKS/NRPS biosynthetic pathway of myxothiazol. The presence of such a gene on this clone suggests that the host organism from which the A12 clone was derived may encode a pathway for the synthesis of a myxothiazol-like compound. Subsequent investigation revealed that clone A12 contained ORFS with homology to the genome of *Solibacter usitatus* from the *Acidobacteria* division. To our knowledge, no polyketide product has been identified from a *Acidobacteria* spp., although the results of a genome sequencing study indicate that members of this division may harbor PKS-related genes (285). These findings reflect the scarcity of known *Acidobacteria* sequences and that these bacteria, like many other soil bacteria, may represent an unexplored reservoir of novel PKS pathways.

1	A3	GCSTACSYSATSTACACSTCSGG
	A7R	SASGTCVCCSGTSCGGTAS
2	K1	TSAAGTCSAACATCGGBCA
	M6R	CGCAGGTTSCSGTACCAGTA
3	KSLF	CCSCAGSAGCGCSTSYTSCTSGA
	KSLR	GTSCCSGTSCCGTGSGYSTCSA
4	KSDPQQF	MGNGARGCNNWNSMNATGGAYCCNCARCANMG
	KSHGTGR	GGRTCNCCNARNSWNGTNCCNGTNCCRTG
5	5LL	GGRTCNCCIARYTGIGTICCIGTICCRTGIGC
	4UU	MGIGARGCIYTICARATGGAYCCICARCARMG
6	F	TSGCSTGCTTGGAYGCSATC
	R	TGGAANCCGCCGAABCCGCT
7	540F	GGITGCACSTCIGGIMTSGAC
	1100R	CCGATSGCICCSAGIGAGTG
8	kSALPHA	TTCGGCGGXTTCCAGTCXGCCATG
	ACP	CCXATGCTCAGCXACCGCGACGACCT
9	A12-TS-F2	GCGTCGACTTGTCCT
	A12-TS-R2	GCGCATATTCTGACC

Table 4.1. Primers used in the generation of ketosynthase-specific probes (17, 89, 162, 197, 233, 286). Primer set 5, specific to Type I KS domains, was the only primer set that yielded a PCR product with pooled fosmid KS domains while not giving a similar product with *E.coli* template. Primer set 9 was used to amplify the *mtaD* homologous sequence from clone A12 and the cultured *Acidobacteria.*
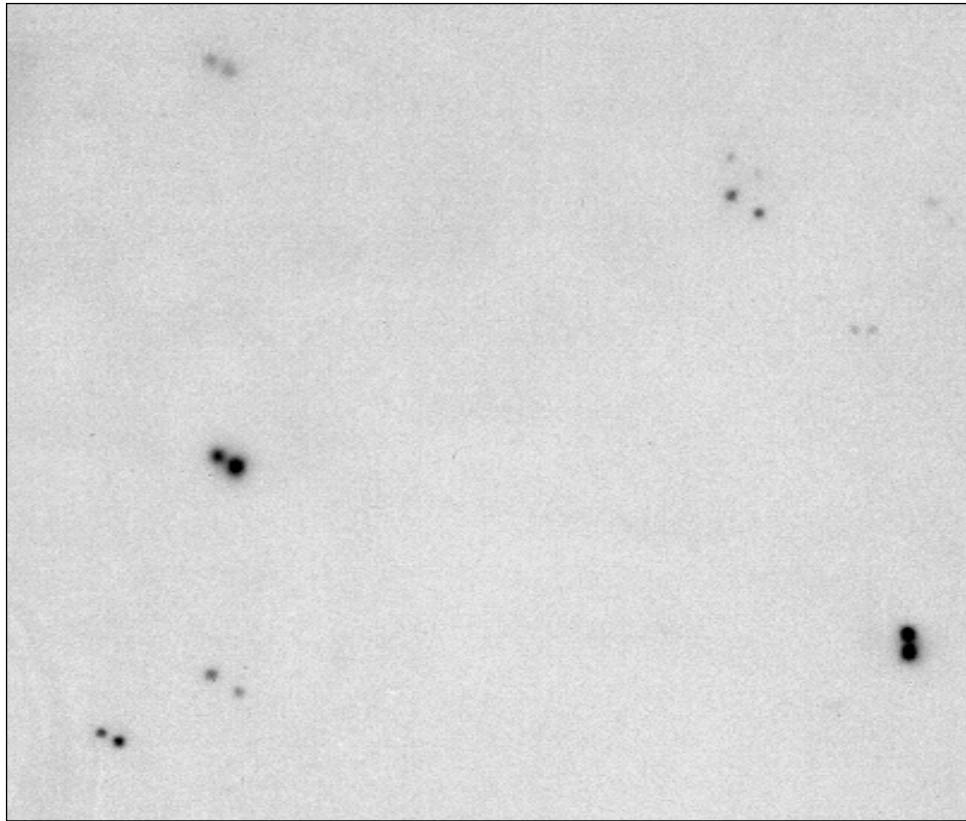
Figure 4.1. Clones from a soil metagenomic library exhibit chemilumniscence after

hybridization with a KS-specific, DIG-labeled probe.
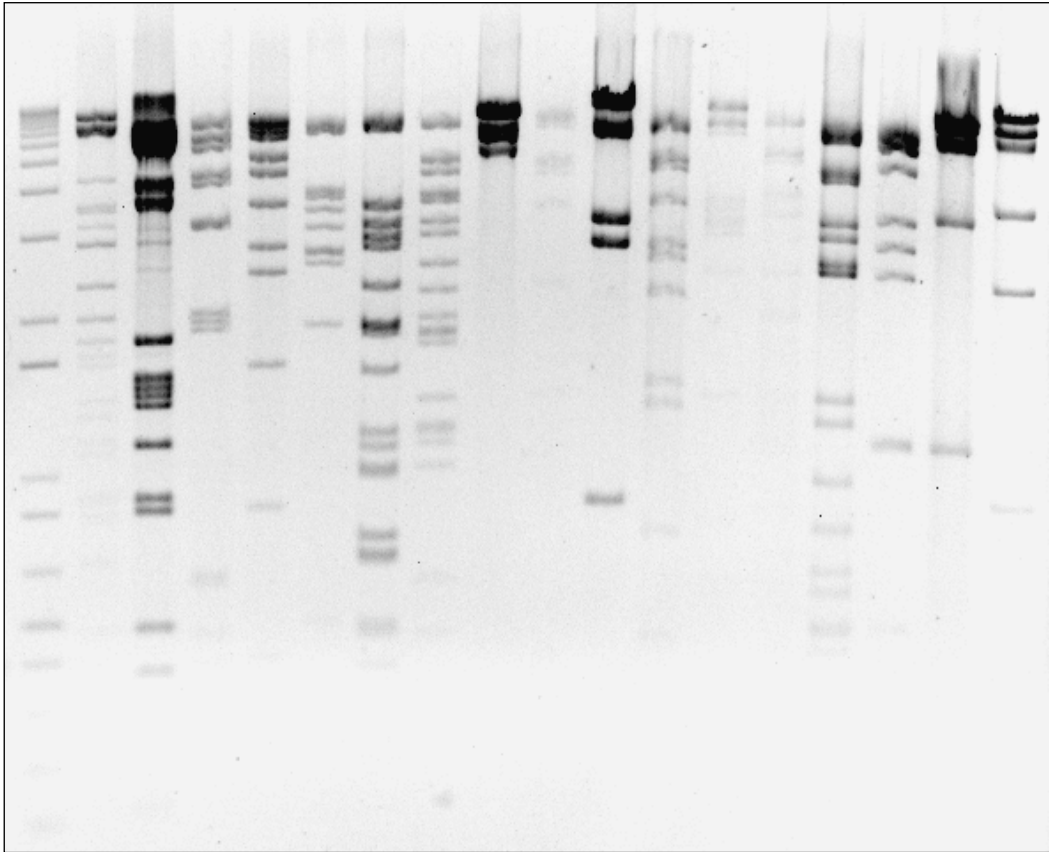
Figure 4.2. BamHI restriction digests of unique KS-containing clones. RFLP patterns were also used to estimate the insert size of each clone.
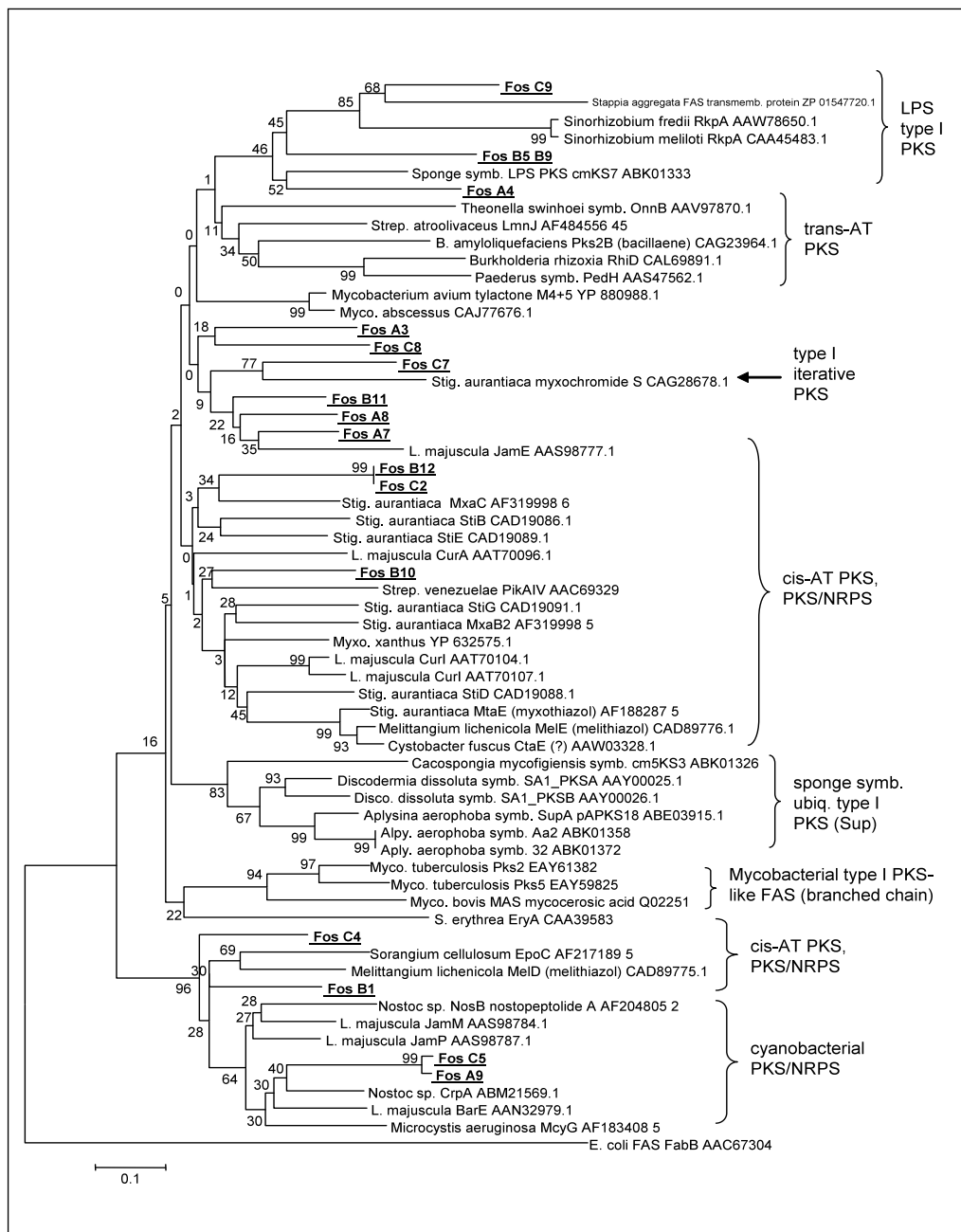
Figure 4.3. Maximum parsimony analysis of cloned KS genes and known KS domains

from GenBank. KS domain sequences recovered from the soil metagenomic library are

underlined, and in bold. Numbers at tree nodes represent bootstrap values, and the tree is

rooted using the *E. coli* FabB (KS-ACP) sequence.
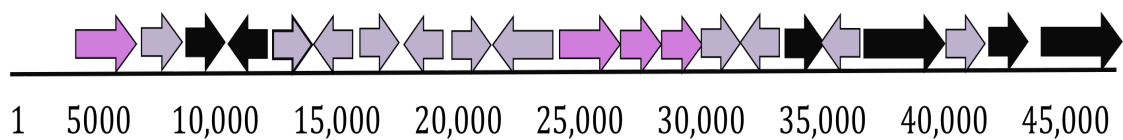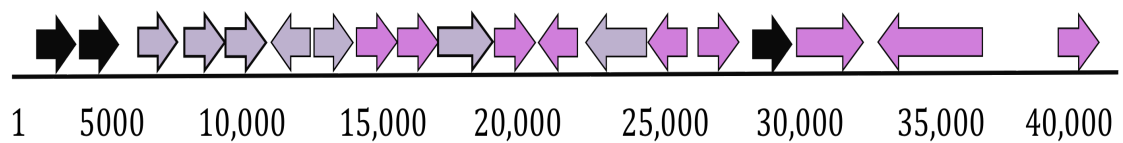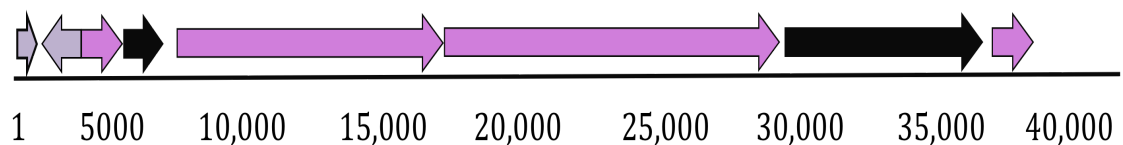
Figure 4.4. Predicted ORFs from Clones A2, A11, and C5 after complete insert sequencing. ORFs were categorized based on their predicted functions after comparison to the GenBank nr/nt database.

156
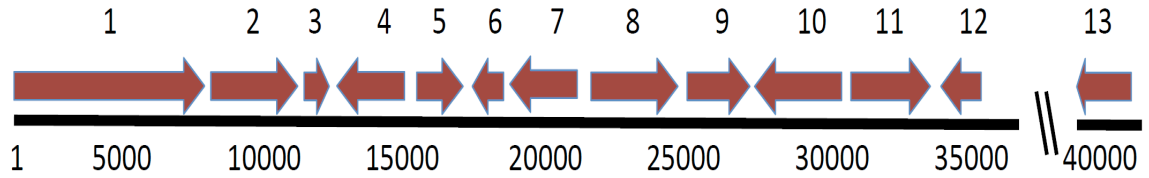
ORF	Top BLASTp	Nearest neighbor	E-value	% Similarity
1	ATPase	*Shewanella amazonensis*	7e-168	76%
2	ATPase	*Shewanella amazonensis*	2e-166	76%
3	Aspartate aminotransferase	*Solibacter usitatus*	2e-48	64%
4	Tetratricopeptide TPR_4	*Methylobacterium* sp. 4-46	2e-124	58%
5	Amidohydrolase	*Arthrobacter* sp. FB24	5e-47	56%
6	Fe-S protein, radical SAM family	*Koribacter versatilis* Ellin345	0.0	86%
7	Hypothetical protein	*Solibacter usitatus* Ellin6076	2e-28	41%
8	Permease	*Opitutus terrae* PB90-1	5e-121	55%
9	Hypothetical protein	*Solibacter usitatus* Ellin6076	4e-156	66%
10	TonB-dependent receptor	*Acidobacteria* sp. Ellin345	2e-72	42%
11	SSS family sodium symporter	*Nostoc punctiforme*	0.0	85%
12	RTX toxin Ca2+-binding domain	*Bradyrhizobium* sp. BTAi1	1e-52	62%
13	MtaD, myxothiazol synthesis	*Stigmatella aurantiaca*	1e-25	71%

Figure 4.5. Predicted ORFs from Clone A12, shown with their corresponding predicted functions and nearest neighbors in the GenBank nr/nt database.
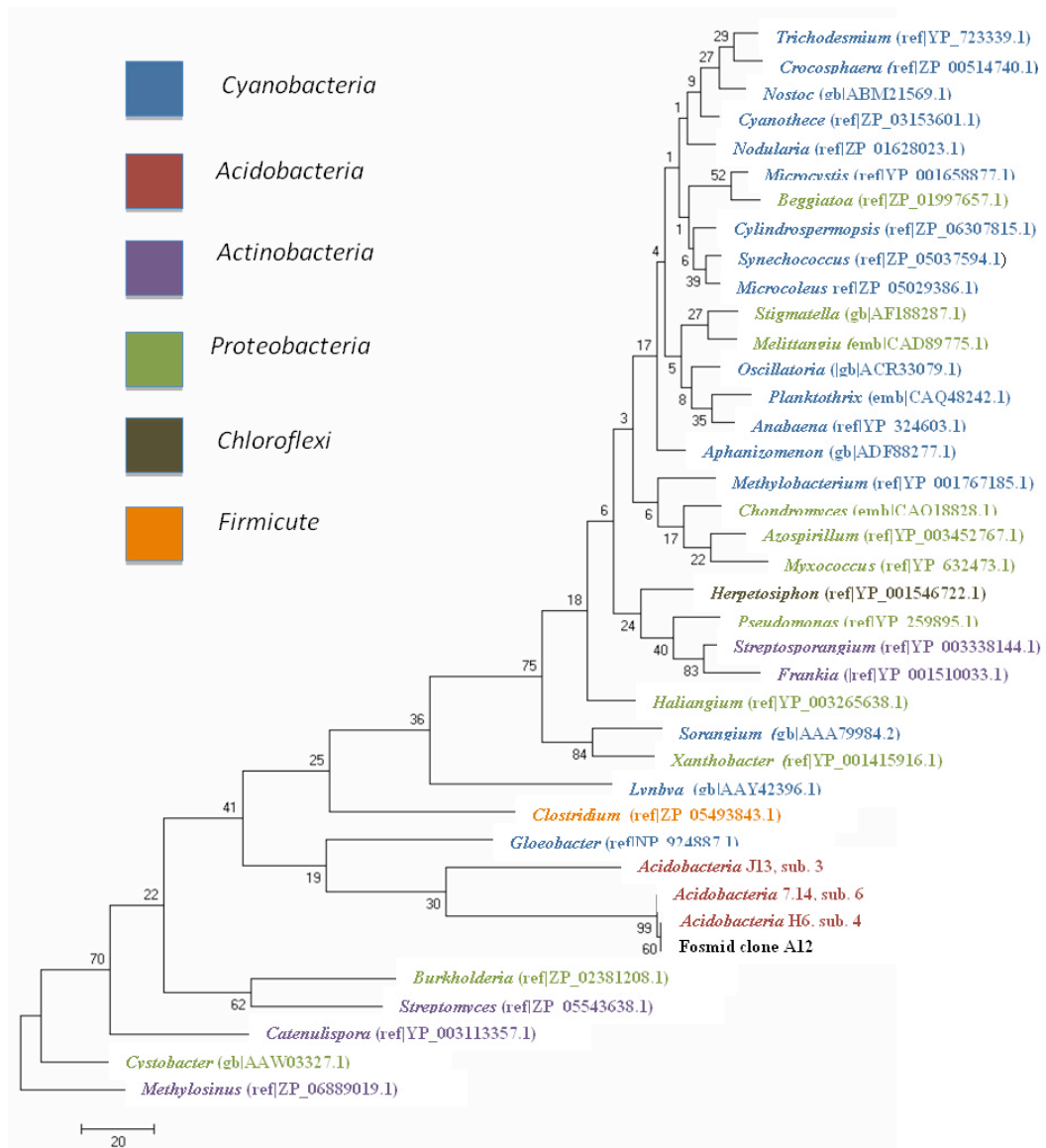
Figure 4.6. Maximum parsimony analysis of *Acidobacteria*-derived sequences and
known MtaD homologs from the GenBank nr/nt database. Numbers at tree node
represent bootstrap values (n=1000 repetitions), and the tree was rooted with the MtaD
homolog from a *Methylosinus* sp.

CHAPTER V

DEVELOPMENT AND EVALUATION OF METHODS FOR THE

CHARACTERIZATION OF BACTERIAL METAGENOMES FROM

ENVIRONMENTAL SAMPLES

A. ISOLATION AND CLONING OF HIGH MOLECULAR WEIGHT METAGENOMIC DNA FROM SOIL MICROORGANISMS

1. Introduction

The successful construction of large-insert metagenomic libraries from natural

environments is dependent on several parameters, including effective cell lysis, DNA

purity, and a high transformation efficiency. One challenge associated with constructing

metagenomic libraries from soil microbes involves the co-isolation of contaminants that

leads to the degradation of DNA due to nuclease activity. As the isolation of intact

genetic pathways from soil microbes is necessary to characterize their genetic and

functional diversity, obtaining high-purity, high molecular weight (HMW) DNA for

library construction is absolutely critical. The following protocol describes the steps for

the indirect extraction of bacterial DNA from soil, embedding the DNA in an agarose

matrix, using a formamide and high salt treatment to eliminate nucleases, size-selecting

DNA by restriction digestion and pulsed-field gel electrophoresis, and cloning the HMW

DNA into a large-insert vector. The resulting metagenomic libraries contain high-purity,

stable, HMW DNA that can be screened for various genetic loci (sequence-based) or phenotypic traits (function-based).

2. Protocol

2.1. Materials

Reagents:

Antibiotic(s) for transformant selection

Agarose, low-melting-point (Promega) in 1x TAE

Agarose, CleanCut (Bio-Rad Laboratories)

Agarose

Bovine serum albumin

Centrifuge bottles, 250 or 500 mL

Cloning vector(s), commercially prepared for large-insert cloning

Conical centrifuge tubes, 15 mL and 50 mL (sterile)

Crombach buffer

Double-distilled H_2O

Electrocompetent *E. coli* cells

Electroporation cuvettes

ESP solution

Formamide, deionized

GELase (Epicentre)

Isopropanol

Ligase, T4 DNA

Ligase buffer, 10X

Lysis buffer [R]

Membrane, Type VS (0.025 μm) (Millipore)

Microcentrifuge tubes, 2mL

Polyvinylpolypyrrolidone, acid-washed (optional)

PMSF solution

Restriction enzyme Sau3A1 and HindIII

Restriction enzyme buffer, 10X

Sodium hexametaphosphate $((NaPO_3)_n, n=6)$, 2% (w/v) (adjusted to pH 8.5 with 0.2% (w/v) Na_2CO_3)

ST buffer

Storage buffer

Syringes, 1 cc

TAE buffer stock, 50X

$T_{10}E_1$ buffer

Equipment:

Cold room (4°C) or similar environment where gel electrophoresis can be performed

Electroporator

Equipment for running agarose gels

Equipment for running pulsed-field gels

Hybridization oven with rotating cylinders (or equivalent)

Magnetic stir plate

Soil sieve (mesh size 2 mm)

Super-speed centrifuge

Tissue homogenizer

Waring blendor

2.2. Method

Homogenization of soil and extraction of bacterial cells

1. Weigh out a 30 g soil sample that has been passed through a soil sieve.

Do not use dessicated soil, as this will result in a significant decrease in DNA yield. Storage of soil at -20℃ or lower is preferable to storage at 4℃. The number of 30 g soil samples used is optional, but the parallel processing of multiple samples may be desired depending upon the yield of HMW genomic DNA.

2. Add 100 ml of ice-cold sterile distilled H_2O to each portion and homogenize in a

    Waring blendor at low speed 3 times for 1 min, with 5 min cooling on ice between

    each run.

 If the soil humic acid content is high, 10 g of acid-washed polyvinvylpolypyrrolidone (PVPP) can be added to the soil suspension before homogenization. PVPP will remove humic acid from the suspension, but the bacterial cell recovery may be reduced.

3. Transfer the soil suspensions to 500 ml centrifuge bottles, using 50 ml sterile distilled

    H_2O to wash out blendor bucket. Centrifuge at 1000 x *g* for 15 min at 5°C.

4. Combine the resulting supernatants and store at 4°C for later steps.

5. Transfer the soil pellets back into the Waring blendor with 100 ml fresh sterile

    distilled H_2O and homogenize for 1 min.

6. Repeat centrifugation in step 3.

7. Pool the resulting supernatant with the previous supernatants at 4°C.

8. Repeat the homogenization and centrifugation steps once more.

The supernatants can be filtered through an 8-ply cheesecloth (or other coarse filter) to remove any particulates prior to collection of bacterial cells.

9. Centrifuge the combined supernatants from the three low-speed spins at 15,000 x

g for 30 min at 5°C. Discard supernatants but store the bacterial pellets at 4°C.

10. Combine the bacterial pellets and resuspend them in 200 ml cold 2% sodium

    hexametaphosphate solution. Homogenize in the Waring blendor 3 times for 1 min

    at low speed with 5 min cooling on ice between each run.

11. Centrifuge the homogenate at 15,000 x *g* for 30 min at 5ºC. Discard the supernatant.

12. Resuspend the pellet in 200 ml Crombach buffer and centrifuge as in step 11.

It may be necessary to repeat these wash steps with Crombach buffer until the supernatant becomes relatively clear.

13. Transfer the pellet to a new centrifuge bottle containing 3 ml Crombach buffer and

    centrifuge as in step 11. Discard the supernatant.

Generation of agarose plugs and cell lysis

14. Prepare a 1.4% low-melting-point agarose solution in 1x TAE and let cool to 45°C.

15. Quickly mix 500 μl of the agarose solution with 500 μl of extracted bacterial cells in

    a 2 ml microcentrifuge tube.

16. Draw up the entire solution into a 1 cc syringe. Place the cap on the syringe and

    incubate at 4°C for approximately 30 minutes until the plug has set.

The syringes containing agarose plugs may be stored at 4°C for several weeks, but ideally cell lysis is performed immediately after plug preparation.

17. Extrude the agarose plug from the syringe by removing the plunger, and allow the

    plug to slide from the syringe into a sterile 50 ml conical tube containing 10 ml of

    lysis buffer.

Multiple plugs can be incubated in the same tube if the volume of lysis buffer is increased accordingly.

18. Incubate the plugs for 3 h at 37°C with gentle agitation in a hybridization oven or

    equivalent.

If needed, the lysis step can be extended to 24 h. This step and subsequent incubations are ideally performed in a hybridization oven with rotating cylinders.

19. Carefully transfer the plugs into a new sterile 50 ml conical tube containing 40 ml of ESP solution and incubate the plugs for 16 h at 55°C with gentle agitation.

20. Repeat incubation with 40 ml of fresh ESP solution for an additional 1 to 24 h.

Up to four 1-cc agarose plugs can be incubated in the same tube containing 40 ml ESP. Especially with multiple plugs per tube, this step is repeated with fresh ESP for an additional 24 hr.

21. Remove the ESP solution and wash 3 times with 10 ml $T_{10}E_1$ per 1 cc agarose plug, letting sit at room temp 10 min each time. Decant and replace with 50 ml of $T_{10}E_1$ and 1 mM PMSF solution. Incubate the plugs for 1 h at room temperature with gentle agitation. Decant solution and repeat PMSF treatment with fresh solution.

22. Remove $T_{10}E_1$ buffer and PMSF solution and replace with 10 ml of storage buffer per 1-cc agarose plug. Incubate at 4°C overnight.

At this stage, the agarose plugs are ready for subsequent steps. However, plugs can also be stored at 4°C in storage buffer for several weeks.

Electrophoresis and formamide treatment of agarose plugs

23. Remove the storage buffer from the plugs and wash with 50 ml of $T_{10}E_1$ buffer three times. After washing, incubate for 1 h at room temperature with gentle agitation.

These washing steps are critical for the removal of excess EDTA present in the storage buffer.

24. Prepare a 1% agarose gel in 1x TAE. Cast the gel without wells and reserve 10 ml of the agarose solution for the next step.

25. Place the agarose plug along the top of the gel and pipet 10 ml of molten agarose over the plugs to secure them in the gel.

26. Electrophorese the DNA from the plugs for 4 to 5 h at 70 V at 4°C (ideally, in a cold

room if available).

27. Remove a narrow strip of agarose to stain and visualize the location of the desired DNA.

28. Excise the corresponding unstained agarose containing the compressed DNA. Notch the bottom of the agarose strip to mark its orientation, and place the new plug in a 15 ml conical tube containing 8 ml of ST buffer.

When excising the DNA, try to minimize the amount of agarose transferred to the conical tube to prevent a decrease in DNA concentration.

29. Incubate the plug overnight at 15°C.

30. Transfer the agarose plug directly into 1 L of $T_{10}E_1$ at 4°C and gently stir the solution on a magnetic plate for 24 to 48 h, with one exchange of $T_{10}E_1$.

31. Place and seal the purified agarose plug along the top of a 1% agarose gel. Align the agarose strip in the same orientation as the prior gel, with the HMW DNA toward the top of the gel.

For improved restriction digestion within an agarose plug, it is recommended to electrophorese the HMW DNA into CleanCut® agarose. Immediately below the HMW plug, and along the entire length of the agarose gel, remove a 3 cm-deep section of normal agarose and pour a 1% Clean Cut agarose solution (in 1x TAE) into this trough so that the height of the CleanCut agarose is equal to the rest of the gel.

32. Electrophorese the DNA from the plug for 3 h at 70 V at 4°C.

33. Visualize and excise the high molecular weight DNA as described in steps 28-29.

34. Immediately slice five 1 mm to 3 mm pieces from the resulting agarose plug, and notch the bottom of each piece to mark the orientation.

These agarose sections will be used for a test digest. Store the remaining agarose plug in storage buffer at 4°C for up to 24 h until results of the test digest are known.

35. Place one of the 1 mm to 3 mm pieces in $T_{10}E_1$ on ice as a control to check for

nuclease activity. Use the remaining 4 pieces to test different levels of restriction enzyme.

Size-selection of DNA by restriction digestion and PFGE

The following protocol is designed for production of BAC vector metagenomic libraries. There are alternatives to partial restriction digestion to construct libraries, but this is the most commonly used protocol. For production of a fosmid library, which has an advantage in terms of numbers of transformants that may be obtained but is limited in insert size, blunt-ended fragments of approximately 40 kb should be isolated and then ligated into a suitable fosmid vector (e.g., Epicentre's EpiFOS).

36. Prepare test digest reaction mixtures (5 total) as follows:

DNA (1 mm slice for concentrated DNA, 3 mm slice for more dilute DNA)

0 to 5 µl (10U/µl) of HindIII restriction enzyme

30 µl of 10X reaction buffer

30 µl of 10X bovine serum albumin

240 µl of sterile H_2O

Alternatively, the restriction enzyme Sau3A1 may be used instead of HindIII.

37. Incubate the test digests at 37°C for 1 h, then replace with new reagents and continue digest for an additional hour at 37°C.

38. Immediately load the digested DNA plugs into a 1% agarose gel for pulsed-field gel electrophoresis.

39. Electrophorese the digested DNA under the following conditions: 0.3- to 3-s switch time -1.5 ramping factor, 120° angle, and 6 V/cm for 10 to 12 h at 14°C. Stain the gel and visualize the size of the restricted DNA.

40. After the optimal digest conditions are determined, repeat digests with the remaining agarose plug sections.

41. Perform pulsed-field gel electrophoresis (same conditions as step 40) of the partially

166

restriction digested DNA.

42. Recover DNA of desired size (e.g., greater than 50 kb) from the gel and place gel slice in a microcentrifuge tube.

43. Extract partially restriction digested, and size-selected genomic DNA from the agarose gel slice, either using GELase treatment (follow manufacturer's instructions) or electroelution (185).

Both of these methods have been used successfully to recover DNA from agarose gel slices. In our labs, electroelution is the method of choice, although it does require some experience to achieve reproducible DNA recovery. For electroelution, the agarose gel slice is placed within a SpectraPor 12-13 kDa dialysis membrane (previously boiled for 10 min, and can be stored in 70% EtOH) and a minimal volume of 1x TAE buffer is added into the dialysis bag (less than 100 μl). The HMW DNA is then electroeluted within an agarose gel electrophoresis apparatus at 70V for 3-4 h at 4^oC. The dialysis bag is placed within ~500 ml of $T_{10}E_1$ buffer for 2-3 h at 4^oC to exchange the buffer. The agarose fragment is then carefully removed from the dialysis bag and the volume of buffer containing DNA at the bottom of the dialysis bag is then removed with a wide-bore pipet tip and placed immediately into a ligation reaction.

Cloning HMW DNA into a BAC (or fosmid) vector

44. Ligate the size-selected DNA to a commercially prepared fosmid or bacterial artificial chromosome vector at a 10:1 molar ratio of insert DNA to vector, respectively.

     1 μl vector DNA (typically 25 ng/μl)

     5 to 16 μl of insert DNA (typically 100 ng or greater)

     1 μl (3 U/μl) of T4 DNA ligase

     2 μl of 10X reaction buffer

     sterile H_2O to a total volume of 20 μl

Incubate the ligation mixture overnight at 15°C.

It is critical to determine the optimum insert:vector ratio for each reaction empirically. In addition, dephosphorylation of the cloning vector to prevent self-ligation is critical to obtain recombinant clones containing genomic DNA inserts.

45. Heat inactivate the ligation mixture at 65°C for 15 min.

46. Dialyze the ligation mixture to remove salts that will interfere with electroporation.

     A drop dialysis method can be used (spot 10 μl of the ligation onto a Millipore Type

     VS, 0.025 μm membrane), or alternatively an agarose cone dialysis may be used (12).

47. Electroporate the ligation mixture into highly electrocompetent *E. coli* cells. Keep

     the cells on ice for 30 min prior to adding 1 μl of dialyzed ligation mixture. Use

     settings according to manufacturer's instructions for *E. coli* electroporation.

     Immediately after electroporation, add 1 ml of 37°C preheated recovery medium (e.g.,

     SOB) and incubate at 37°C with shaking for 30 min.

48. Select for transformants using the appropriate antibiotic on LB media for 15 h at

     37°C.

At this stage, LB plates may be sealed and stored at 4°C formany days. The transformants can then be picked into 96- or 384-well format and stored in 10% glycerol at -80°C for further analysis by sequence-based or functional screening.

49. Isolate clone DNA from a representative number of transformants, restriction digest

     the DNA, and evaluate the average insert size for the metagenomic library. Based on

     the estimate of average insert size, and the total number of transformants obtained,

     the total size (in Gbp, hopefully) of the metagenomic library can be determined.

2.3. Troubleshooting

Problem: Insufficient yield of bacterial cells [Step 12]

Solution: The prior step to remove eukaryotic cells and soil particles may have also

resulted in centrifugation of bacterial cells adherent to particulates. Increasing the degree

of homogenization of the sample may improve bacterial cell recovery. Alternatively, using a different extraction buffer may serve to release bacterial cells that are adsorbed to soil surfaces (e.g., metaphosphate or beef extract buffers), thereby improving cell yield. Also, increasing the size of the sample used and the number of batches of samples that are processed may be helpful in obtaining sufficient biomass for subsequent steps.

Problem: Insufficient yield of DNA [Step 27]

Solution: The efficiency of bacterial cell lysis within a plug will depend upon many factors, including cell concentration, the presence of humic acids and other soil derived impurities, and the composition of the microbial community. Efficient Gram-positive cell lysis may require additional treatments (e.g., phage-derived lysins that target peptidoglycan), as previous work has shown that soil metagenomic libraries produced via similar methods have a bias in favor of Gram-negative bacteria (141). The most significant factor affecting DNA yield is the origin of the sample, as the largest differences in DNA yield are typically a function of the sampling site, and perhaps time of year that the sample is taken. However, abundant HMW DNA has been obtained from a Wisconsin soil sampled in February, so even frozen soil can yield sufficient DNA. The temporal and sampling variability of the site should be evaluated if possible.

Problem: Loss of DNA after buffer control experiment [step 35]

Solution: This is an important control experiment to determine if DNA degradation occurs from a preparation of HMW DNA in the absence of restriction enzyme. The formamide and denaturation buffer has been observed to remove DNA degradative activity from multiple soil types; however, it is possible that soils with exceedingly high organic/humic acid contents may require extended incubation in the formamide buffer

and/or repeated treatments to remove all DNA degradation activity. Performing more bacterial cell washes prior to embedding cells within an agarose plug [steps 10 and 11] may also improve the purity of the genomic DNA.

Problem: Inefficient restriction digestion [Step 36]

Solution: Each environmental sample may require a different amount of restriction enzyme to achieve a similar degree of restriction digestion. It has been observed that two different soil samples required consistently different amounts of restriction enzyme to achieve the same degree of restriction, which could reflect the different yields of DNA and associated impurities from each sample. Some restriction enzymes (e.g., HindIII and Sau3AI) appear to be more tolerant of impurities in the genomic DNA preparation. Also, restriction enzymes from different commercial sources can give radically different degrees of restriction digestion, so evaluate each DNA preparation empirically to determine the optimal partial digests conditions.

Problem: Poor recovery of DNA from size-selection gel [Step 44]

Solution: The extraction of DNA from a gel slice can be problematic, as the HMW genomic DNA is susceptible to shearing forces present once removed from an agarose gel. A test electroelution (or GELase treatment) of DNA embedded within agarose is advisable when beginning this work, to evaluate the time required to electroelute a good yield of DNA from a gel slice. Keeping gel slices as small as possible, while still recovering a good fraction of HMW DNA, is the objective. Be certain to use wide-bore pipet tips at all times once DNA has been recovered from a gel slice to prevent DNA shearing.

Problem: Low efficiency of transformation [Step 48]

Solution: A low concentration of DNA, or *E. coli* cells that have a low transformation efficiency, can produce disappointing results for transformations. Rather than use "lab-prepared" *E. coli* cells for transformation, use a commercial supplier with a strain (i.e., *E. coli* strain DH10B) developed for large-insert library generation with excellent electroporation efficiency (e.g., greater than 10^{10} transformants per mg pUC19 DNA).

Problem: Poor percentage of insert-containing clones [step 50]

Solution: There can be many reasons why insert size is poor. The first question should be whether small inserts are present. If no inserts are observed, the likely problem is with vector self-ligation, resulting in a very large number of transformants but without observable inserts. The solution for this result would be to either 1) switch to a different phosphatase to achieve better dephosphorylation of the vector prior to ligation with insert DNA, or 2) switch to a commercial supplier of ready-to-ligate vector that has been quality tested for low cloning background due to vector self-ligation. If small inserts are present, and it does not appear that vector self-ligation is the problem, then removal of small DNAs present in the ligation may be the solution. Various methods have been shown to help remove small DNA fragments that have become trapped within HMW DNA, particularly when high concentrations of HMW DNA are used. Running the size-selection PFGE [step 42] the wrong direction toward the top of the gel by switching the electrodes for 20-30 min, prior to electrophoresis toward the bottom of the gel, may help remove smaller DNAs that may preferentially ligate into the vector and produce transformants. Running a second size-selection gel may also improve removal of small DNAs, but should only be considered when the concentration of HMW DNA is not limiting.

3. Evaluation of Protocol

Genomic DNA isolated from soil microorganisms using harsh extraction methods is often less than 100 kb (130, 164, 254). Unfortunately, this can limit the amount of contiguous environmental DNA isolated for metagenomic library construction and therefore limit the downstream analyses (*e.g.,* functional screening, phylogenetic analysis) performed on the libraries (54). Various methods for the initial extraction and purification of bacterial DNA exist, including direct DNA extraction (267), ion exchange for dissociating bacterial cells from soil particles (149), Nycodenz density gradient centrifugation (142), and homogenization (24). Depending on the composition of the soil used (*i.e.,* high clay or silt content), a combination of these approaches may be used to obtain HMW DNA, and the inclusion of the formamide and high salt treatment may be critical in increasing cloning efficiency.

Liles *et al.* (143) evaluated the protocol described here by using it to isolate HMW metagenomic DNA from multiple soil samples, comparing the yield and purity to that of DNA isolated by various methods (e.g., Nycodenz treatment), and constructing BAC and fosmid metagenomic libraries from the resulting DNA. The success of the formamide treatment is shown in Figure 5.2, and a summary of library statistics is shown in Table 5.1.

4. Discussion

The prevention of DNA degradation due to nuclease activity is a major hurdle in the construction of large-insert metagenomic libraries from natural environments. Although high molecular weight (HMW) DNA greater than 1 Mbp in size can be isolated from soil using the indirect DNA extraction method presented here, the DNA isolated

from some samples cannot be readily cloned due to nucleases initially co-isolated with the bacterial cells. To overcome this hurdle, a formamide and high salt treatment of agarose plugs containing HMW metagenomic DNA can help eliminate nuclease activity and other co-isolated impurities while maintaining integrity of the DNA for subsequent cloning.

The extraction and purification methods presented here provide efficient solutions for recovering metagenomic DNA from soil microorganisms. However, due to soil's complex nature and the need for DNA of high purity, yield, and sometimes size for various metagenomic approaches, using only a single method is often insufficient for success. In addition, choosing a DNA extraction and/or purification method that is not suited for the ultimate application will likely result in reduced efficiency of the overall experimental process. For example, using a commercial kit for DNA extraction may be sufficient for PCR-based studies but not large-insert cloning. Similarly, performing HMW DNA purification techniques (e.g., formamide treatment) is usually not necessary for PCR or small-insert cloning. Fortunately, new methods for metagenomic DNA extraction and purification have been developed to allow investigations into the functional and phylogenetic diversity of various microbial habitats. Selecting the appropriate combination of DNA extraction and purification methods for the desired application remains the best chance for achieving success in any metagenomic approach.

Soil sample

Homogenize soil & isolate cells
• Use differential centrifugation
  to recover microbial cells

Microbial cell plug
• Extract cells from soil
• Embed in low-melt agarose
• Lyse cells within the plug

Well

Compressed
DNA

DNA Isolation
• Electrophorese DNA from plug
• Excise non-stained DNA
  within agarose

DNA Purification
•Treat DNA (still in agarose)
  with formamide solution
• Electrophorese DNA into
  low-melting point agarose

Restriction and size-selection
• Restrict DNA (partial digestion)
• Electrophorese DNA
• Extract HMW DNA in agarose
Increasing [enzyme] • Electroelute DNA from agarose

Cloning environmental DNA
• Ligate insert with cloning vector
• Electroporate into *E. coli*
• Select for *E. coli* transformants
E. coli transformants • Characterize library

Figure 5.1. Diagram outlining the steps in the the recovery, purification, and cloning of

HMW metagenomic DNA from soil microorganisms.

174

Lane 1, 1kb DNA Ladder (Promega);

Lane 2, MidRange II PFGE Marker (New England BioLabs);

Lane 3, BCEF soil DNA **+ formamide and salt treatment**;

Lane 4, BCEF soil DNA **+ formamide and salt treatment** + Sau3AI (10U);

Lane 5, BCEF soil DNA;

Lane 6, BCEF soil DNA + Sau3AI (10 U).
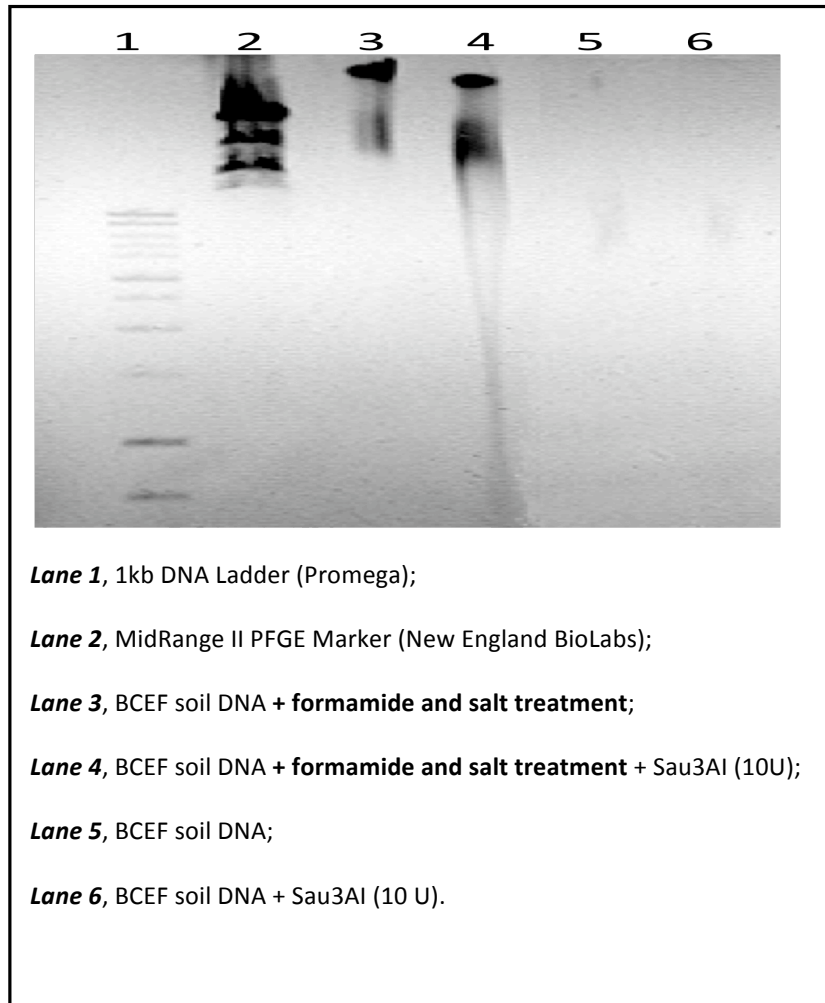
Figure 5.2. Formamide treatment of soil metagenomic DNA prevents its degradation by nucleases. (BCEF, Bonanza Creek Experimental Forest near Fairbanks, Alaska)

LIBRARY NAME	LIBRARY TYPE	NYCODENZ ISOLATION	FORMAMIDE TREATMENT	NUMBER CLONES	APPROX % INSERT	AVG. INSERT SIZE (kb)
AK 5	BAC	YES	NO	2,200	70	33
AK 7	BAC	YES	NO	2,300	60	47
AK 8	BAC	YES	NO	2,700	70	9
AK 9	BAC	YES	NO	2,400	25	20
AK 13	BAC	NO	NO	8,400	10	29
AK 17	BAC	NO	NO	1,000	20	29
AK 15	BAC	NO	YES	32,100	60	20
AK 19	FOSMID	NO	YES	3,300	98	31
AK 20	BAC	NO	YES	48,100	80	8
AK 21	FOSMID	NO	YES	333,700	100	30
AK 22	FOSMID	NO	YES	50,500	100	30
AK 23	FOSMID	NO	YES	8,200	100	30

Table 5.1. Soil metagenomic libraries constructed with and without formamide treatment.

B. FLUORESCENCE *IN SITU* HYBRIDIZATION OF BACTERIAL CELL SUSPENSIONS AND ENVIRONMENTAL SAMPLES

1. Abstract

The use of fluorescence *in situ* hybridization (FISH) to identify and enumerate specific bacteria within a mixed culture or environmental sample has become a powerful tool in combining microscopy with molecular phylogenetic analysis. However, processing a large number of samples in parallel can be difficult because the bacterial cells are typically fixed and hybridized on microscope slides rather than processed in solution. In addition, Gram-positive cells and certain environmental samples present a unique challenge to achieve adequate cell fixation and uniform hybridization for optimum FISH analysis. Here, we describe a protocol for FISH in solution that can be performed entirely in microcentrifuge tubes prior to microscopy and can be applied to both Gram-positive and -negative cell and a homogenized environmental sample. This protocol utilized genus-specific oligonucleotide 16S rDNA FISH probes and was able to differentiate among three bacterial genera from the γ-Proteobacteria division, as well as identify *Bacillus* spp. from pure and mixed cultures and from catfish gut microbiota. Many other phylogenetically informative probes could be used in this protocol. This method is a rapid protocol for performing multiple hybridizations simultaneously, which may be used to qualitatively assess the presence of specific phylogenetic groups in bacterial cultures or environmental samples and/or directly quantify fluorescence by fluorometry or flow cytometry.

2. Protocol

2.1. Materials

Reagents:

Anti-fade reagent (SlowFade; Invitrogen)

Deionized H_2O

Ethanol

Fixative solution

Hybridization solution

Medium for bacterial broth cultures

Oligonucleotide probe, labeled with appropriate fluorophore

PBS, 1X, pH 7.4

SSC, 0.1X, pH 7.0

SYBR Gold (optional, for the detection of non-hybridized cells)

Equipment:

Aluminum foil

Cover slips

Culture tubes

Fluorescence microscope with appropriate filters

Microcentrifuge

Microcentrifuge tubes

Micropipettor with tips

Microscope slides, cleaned with ethanol

Nail polish, clear (or other alternative to tack slides)

Spectrophotometer

Speed vacuum dryer or aspirator (to completely dry fixed cells)

Water bath, variable temperature

2.2. Method

Bacterial cell fixation

The following steps describe the procedure for fixation of Gram-positive bacteria, Gram-negative bacteria, or an environmental sample.

A. Gram-positive cell fixation

A1. For each bacterial culture, inoculate a 2 ml broth culture and incubate at the bacterium's optimum growth temperature, medium, and incubation conditions.

After incubation, it may be beneficial to perform cell counts (or OD_{600} spectrophotometer readings) so that the desired cell density may be achieved prior to hybridization.

A2. Harvest bacterial cells during the logarithmic growth phase (OD_{600} of 0.4-1.0) by centrifugation at 13,000 x g for 10 min.

Harvesting the cells during log phase allows for optimum hybridization of vegetative cells only, which is an important consideration for spore-forming bacteria such as Bacillus *spp. Because the cell count at this growth stage may be low, multiple 2 ml cultures can be combined and harvested in the same microcentrifuge tube to increase cell yield.*

A3. Discard supernatant and thoroughly resuspend cells in 1 ml of 1X PBS, pH 7.4.

A4. Centrifuge cells at 13,000 x g for 10 min and discard supernatant.

A5. Resuspend cells in 500 μl of 1X PBS, pH 7.4.

A6. Add 500 μl of cold 100% ethanol and store tubes at -20ºC.

Fixed cells can be stored at -20ºC for several months.

A7. Prior to hybridization, pellet the cells by centrifugation at 13,000 x g for 5 min and discard supernatant.

B. Gram-negative cell dixation

B1. For each bacterial culture, inoculate a 2 ml broth culture and incubate overnight at the bacterium's optimum growth temperature, medium, and incubation conditions.

After incubation, it may be beneficial to perform cell counts (or OD_{600} spectrophotometer readings so that the desired cell density may be achieved prior

to hybridization.

B2. Transfer the culture into a 2 ml microcentrifuge tube and centrifuge at 13,000 x g for 5 min. Discard supernatant.

B3. Resuspend the cell pellet in 1 ml fixative solution and incubate at room temperature for 3 h.

B4. Centrifuge the cell mixture at 13,000 x g for 5 min and discard supernatant.

B5. Wash the cell pellet by resuspending in 50% ethanol and incubating at room temperature for 5 min.

B6. Centrifuge tube at 13,000 x g for 2 min and discard supernatant.

B7. Repeat wash steps with 80% and 95% ethanol.

B8. Dry cell pellet by aspiration or by placing tubes in a speed vacuum dryer for 10 min.

It is critical that all residual ethanol be removed from the cell pellets. Fixed cells may be stored at 4°C for up to 1 month, or they can be immediately processed.

C. Environmental sample fixation

With environmental samples especially, it is necessary to include a sample that is not incubated with a labeled probe to assess the degree of autofluorescence present in the sample.

C1. If possible, homogenize the environmental sample to reduce the adverse effects of aggregation on obtaining a high-integrity fluorescent signal. It may also be advantageous to wash the environmental sample or separate the cells by centrifugation prior to hybridization.

C2. Proceed with one of the above-described protocols for cell fixation. If the environmental sample likely contains a mixture of both Gram-positive and –negative bacteria, samples may be processed in parallel and then images may be overlayed after viewing.

Pre-hybridization and hybridization

1. Resuspend fixed cells in 500 μl of hybridization buffer and incubate at 37°C for 30 min.

After the incubation, it is recommended that part of the pre-hybridization mixture

(e.g., 50 µl) be removed for hybridization. This allows for multiple hybridizations per sample but also decreases the amount of probe and other reagents necessary in each reaction.

2. Add fluorescently-labeled probe(s) to the pre-hybridization mixture.

   The optimum amount of probe to be used should be determined empirically. Several factors can affect the amount of probe needed, such as sensitivity of the probe, cell density, and the length of hybridization. In our experiments, we found that adding 5 µl of a 50 ng/µl probe stock to 50 µl of pre-hybridization mixture (5 ng/µl final concentration) yielded consistent and uniform fluorescence. This protocol is optimized for a fluorescently-labeled oligonucleotide probe that does not require denaturation prior to hybridization.

3. Incubate the hybridization mixture at 37°C to 65°C for 2-3 h in the dark or by completely covering tubes with aluminum foil.

   Because multiple probes were used in our experimental protocol, a high stringency hybridization at 65°C was necessary for specific probe binding. Depending on other conditions used, a lower temperature may be adequate for specific probe hybridization. It is recommended to test a range of hybridization temperatures between 37°C and 65°C to determine the optimum hybridization temperature. Many protocols adopt 42ºC and manipulate the formamide concentration (40% in this protocol) to achieve desired stringency.

4. Centrifuge mixture at 13,000 x g for 5 min and discard supernatant.

5. Wash cells by resuspending cell pellet in 20 µl 0.1X SSC and incubate at 37°C for 15 min in the dark as in Step 10.

6. Centrifuge mixture at 13,000 x g for 5 minutes and discard supernatant.

7. Repeat 0.1X SSC wash steps twice.

8. Resuspend washed cell pellet in 20 µl 0.1X SSC.

9. Transfer 10 µl of the sample to a microscope slide that has been cleaned with ethanol. Add 5 µl of SlowFade reagent (or other reagent to reduce photobleaching) and place a cover slip over the mixture. If necessary, tack a cover slip to the microscope slide (i.e, if an inverted microscope will be used). Immediately store slides in the dark.

   Clear acrylic fingernail polish is an inexpensive and effective means to tack the cover slip to the microscope slide. The remaining hybridization mixture and the prepared slides may be stored in the dark at 4°C for up to 1 week before viewing.

10. View sample using appropriate filters using a fluorescent microscope.

It is helpful to focus the microscope on a desired field of cells before the fluorescence filters are applied (i.e., in "light only" mode), as this will minimize photobleaching. Only cells containing hybridized probe will fluoresce, although using a counterstain such as SYBR Gold will allow for the visualization of non-hybridized cells. In our experiments, we viewed with a Zeiss Axiovert 200 inverted fluorescence microscope with a Nikon camera workstation and Elements imaging software.

2.3. Troubleshooting

Problem: There is a poor fluorescent signal observed during viewing of slides.

[Steps 2, 3, 4, 7, 9]

Solution: Several factors can impact the level of signal observed during microscopy, either due to a lack of probe hybridization or washing conditions which are too stringent. Using a sufficient amount of labeled probe is critical in achieving efficient hybridization signal. In addition, performing the hybridization and subsequent washing steps at a lower stringency may produce conditions that are more favorable for hybridization. The following factors may be interfering with obtaining an optimum fluorescent signal and should be determined empirically for each sample and probe combination:

 i)   Hybridization temperature - A lower temperature will result in increased probe hybridization but will also decrease probe specificity.

 ii)   Formamide concentration - A lower concentration of formamide may be needed. Generally, a range of 0%-40% formamide concentrations is tested during pilot experiments.

 iii)   Probe concentration - Depending on which probe(s) is used, more probe may be required for adequate hybridization.

 iv)   Concentration of SSC used in washing steps - Using 1X SSC instead of 0.1X SSC

182

will decrease the stringency of the washing steps.

v)   Washing temperature - Although 37°C is generally considered to be the optimum washing temperature, decreasing the temperature in 5ºC increments correlates to less stringent washing conditions.

vi)   The number of wash steps - Decreasing the number of wash steps decreases the stringency of washing. However, it is not recommended to perform less than 2 washes.

Also, degradation of fluorescence may occur if the hybridization mixture or prepared slides are not stored in the dark or are stored longer than the recommended time prior to viewing.

Problem: The probe(s) did not specifically hybridize with the desired bacterial cells. [Steps 2, 3, 5, 7]

Solution: Alternatively, the above conditions may be adjusted to produce a more stringent hybridization reaction. Increasing the hybridization and washing temperatures, as well as the number of wash steps, will increase the specificity of probe hybridization and the stringency of the washes. The amount of formamide in the hybridization buffer can also be increased, but the recommended maximum (40%) is given in this method. It is recommended to address each of these conditions individually because increasing the stringency of the reaction too much will interfere with the hybridization signal achieved.

Problem: The fluorescent signal is too great or there is photobleaching. [Steps A1, B1, C1, 8]

Solution: The following factors that should be considered if an excess of fluorescence is observed or there is photobleaching:

i) Cell density - Using too many cells in the hybridization reaction will result in a "crowded" viewing field, which can make it difficult to observe individual cells.

ii) Amount of SlowFade reagent - Many anti-fade reagents are available, and we highly recommend the use of these products. If more or less hybridization mixture is applied to the slide than what is described in the protocol, the amount of SlowFade reagent should be adjusted accordingly. In addition, it may be necessary to use different amounts of alternate anti-fade reagents according to the manufacturer's instructions.

iii) Prepared slide storage - Prepared slides should be immediately stored in the dark until they are ready to view. It is recommended they be kept in the dark until they are transferred to the microscope and that images be captured as soon as possible during viewing.

3. Discussion

There are many applications for fluorescent *in situ* hybridization (FISH) in environmental microbiology, such as the detection of phylogenetically or functionally specific bacteria in a heterogeneous sample (91, 148, 183, 194), bacterial endsymbiosis (106, 223), and studies of single-cell phylogeny and function (107, 146) Various genetic markers have been utilized for bacterial FISH studies, including the 16S rRNA gene (148, 183, 222, 299) and genes encoding certain metabolic activities (111, 194)   In the specific experiments being investigated in our laboratory, a bacterial culture obtained from an activated sludge microbial assemblage was isolated that exhibited multiple colony and cell morphologies, a single ribotype by 16S rRNA gene analysis, and varying enzymatic profiles depending upon the age of the bacterial culture. To resolve whether

the differing cell morphologies were two distinct bacterial taxa or was actually a pure bacterial culture with pleiomorphic characteristics, the preceding protocol was developed to enable FISH in solution prior to applying the hybridized sample onto a slide for fluorescent microscopy. More recently, this protocol was also used to detect the prevalence and persistence of *Bacillus* spp. from the gut microbiota of catfish that had been given *Bacillus*-infused feed for biocontrol purposes. For many different applications, the use of probes specific to different bacterial taxa in addition to a universal bacterial probe can be helpful in making a preliminary identification of microorganisms within a mixed culture, co-aggregate, or environmental sample. The large database of 16S rDNA gene sequences and available probe sets (e.g., http://rdp.cme.msu.edu/) may be utilized for phylogenetic discrimination by using both a universal bacterial probe (e.g., EUB338) combined with other taxa-specific probes. Since in this protocol the hybridization reactions are carried out in solution rather than on microscope slides, it is also easier to process a larger number of samples in parallel and manipulate various conditions (e.g., different temperatures, probe concentrations, etc.) when optimizing the hybridization reaction and proceed to quantitative measures of fluorescence such as fluorometry or flow cytometry.
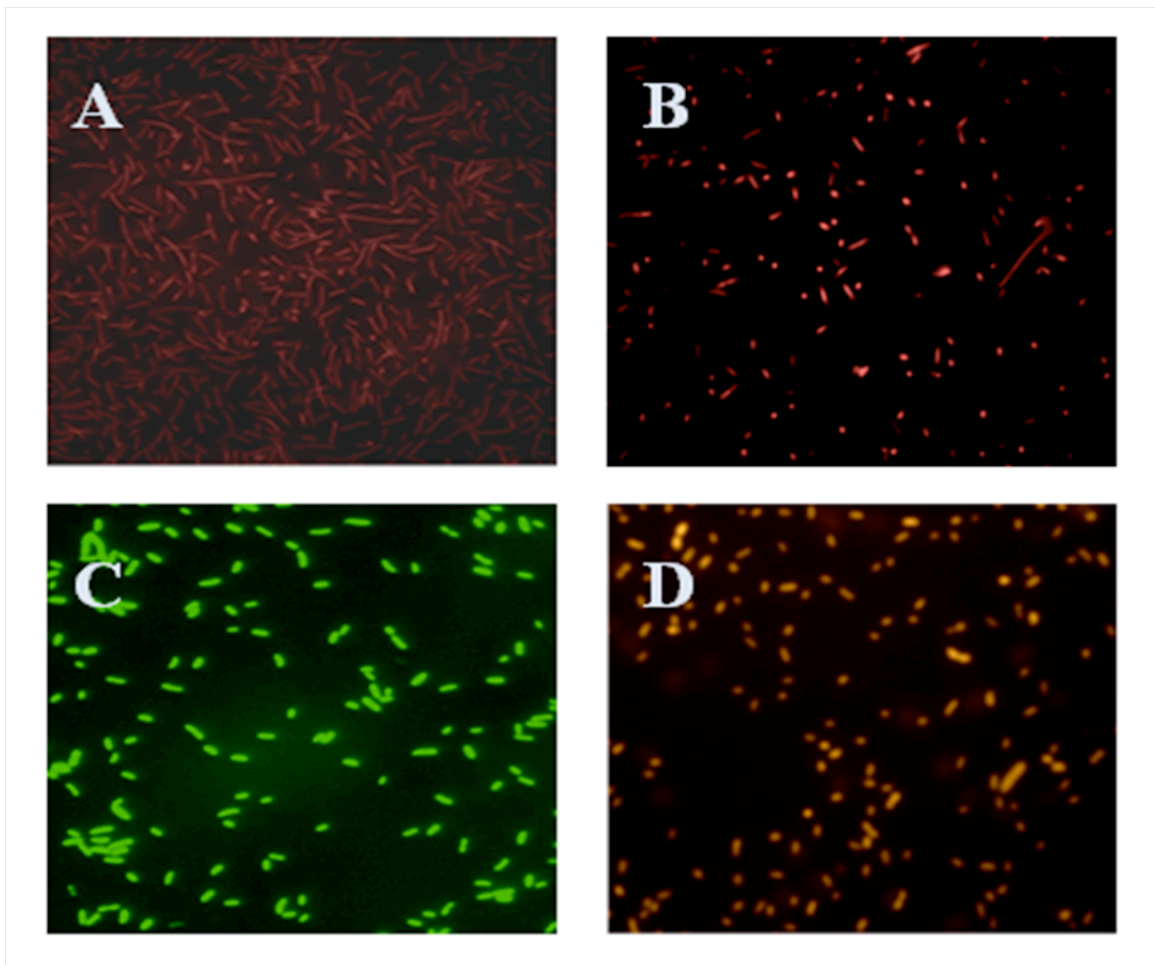
Figure 5.3. Fluorescence micrographs of bacterial cultures analyzed by in-solution FISH probing with Texas Red-labeled EUB338 (A, B, D) or FITC-labeled AER66 (C,D).

COMPREHENSIVE BIBLIOGRAPHY

1. **Ahmed, A. U., and D. L. Sorensen.** 1995. Kinetics of pathogen destruction during storage of dewatered biosolids. Water Environment Research **67:**143-150.

2. **Alouini, S., and M. D. Sobsey.** 1995. Evaluation of an extraction-precipitation method for recovering Hepatitis A virus and poliovirus from hardshell clams(Mercenaria mercenaria). Water Science & Technology **31:**465-469.

3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. Journal of molecular biology **215:**403-410.

4. **Altshuler, M. L.** 2006. PCR troubleshooting: the essential guide, 1st ed. Caister Academic Press, Wymondham, Norfolk (UK).

5. **Amann, R. I., L. Krumholz, and D. A. Stahl.** 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J Bacteriol **172:**762-70.

6. **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev **59:**143-69.

7. **Aminov, R. I., N. Garrigues-Jeanjean, and R. I. Mackie.** 2001. Molecular ecology of tetracycline resistance: development and validation of primers for detection of tetracycline resistance genes encoding ribosomal protection proteins. Appl Environ Microbiol **67:**22-32.

8. **Aminov, R. I., and R. I. Mackie.** 2007. Evolution and ecology of antibiotic resistance genes. FEMS Microbiol Lett **271:**147-61.

9. **Angly, F., B. Rodriguez-Brito, D. Bangor, P. McNairnie, M. Breitbart, P. Salamon, B. Felts, J. Nulton, J. Mahaffy, and F. Rohwer.** 2005. PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information. BMC Bioinformatics **6:**41.

10. **Angly, F. E., B. Felts, M. Breitbart, P. Salamon, R. A. Edwards, C. Carlson, A. M. Chan, M. Haynes, S. Kelley, H. Liu, J. M. Mahaffy, J. E. Mueller, J. Nulton, R. Olson, R. Parsons, S. Rayhawk, C. A. Suttle, and F. Rohwer.** 2006. The marine viromes of four oceanic regions. PLoS Biol **4:**e368.

11. **Aparicio, J. F., R. Fouces, M. V. Mendes, N. Olivera, and J. F. Martin.** 2000. A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in *Streptomyces natalensis*. Chem Biol **7:**895-905.

12. **Atrazhev, A. M., and J. F. Elliott.** 1996. Simplified desalting of ligation reactions immediately prior to electroporation into E. coli. Biotechniques **21:**1024.

13. **Auerbach, E. A., E. E. Seyfried, and K. D. McMahon.** 2007. Tetracycline resistance genes in activated sludge wastewater treatment plants. Water Res **41:**1143-51.

14. **Auling, G., F. Pilz, H. J. Busse, S. Karrasch, M. Streichan, and G. Schon.** 1991. Analysis of the polyphosphate-accumulating microflora in phosphorus-eliminating, anaerobic-aerobic activated sludge systems by using diaminopropane as a biomarker for rapid estimation of Acinetobacter spp. Appl Environ Microbiol **57:**3585-92.

15. **Austin, M. B., and J. P. Noel.** 2003. The chalcone synthase superfamily of type III polyketide synthases. Nat Prod Rep **20:**79-110.

16. **Axelrood, P. E., M. L. Chow, C. C. Radomski, J. M. McDermott, and J. Davies.** 2002. Molecular characterization of bacterial diversity from British Columbia forest soils subjected to disturbance. Can J Microbiol **48:**655-74.

17. **Ayuso-Sacido, A., and O. Genilloud.** 2005. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. Microb Ecol **49:**10-24.

18. **Bakken, L. R., and V. Lindahl.** 1995. Recovery of bacterial cells from soil, p. 13-27. *In* J. T. Trevors and J. D. Van Elsas (ed.), Nucleic Acids in the Environment. Springer-Verlag, Berlin.

19. **Banik, J. J., and S. F. Brady.** 2008. Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci U S A **105:**17273-7.

20. **Barlow, M., and B. G. Hall.** 2002. Origin and evolution of the AmpC beta-lactamases of Citrobacter freundii. Antimicrob Agents Chemother **46:**1190-8.

21. **Barra-Caracciolo, A., P. Grenni, C. Cupo, and S. Rossetti.** 2005. In situ analysis of native microbial communities in complex samples with high particulate loads. FEMS Microbiol Lett **253:**55-8.

22. **Bathe, S.** 2004. Conjugal transfer of plasmid pNB2 to activated sludge bacteria leads to 3-chloroaniline degradation in enrichment cultures. Lett Appl Microbiol **38:**527-31.

23. **Beja, O., M. T. Suzuki, E. V. Koonin, L. Aravind, A. Hadd, L. P. Nguyen, R. Villacorta, M. Amjadi, C. Garrigues, S. B. Jovanovich, R. A. Feldman, and E. F. DeLong.** 2000. Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. Environ Microbiol **2:**516-29.

24. **Berry, A. E., C. Chiocchini, T. Selby, M. Sosio, and E. M. Wellington.** 2003. Isolation of high molecular weight DNA from soil for cloning into BAC vectors. FEMS Microbiol Lett **223:**15-20.

25. **Bertrand, H., F. Poly, V. T. Van, N. Lombard, R. Nalin, T. M. Vogel, and P. Simonet.** 2005. High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. Journal of Microbiological Methods **62:**1-11.

26. **Bevitt, D. J., J. Cortes, S. F. Haydock, and P. F. Leadlay.** 1992. 6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. Eur J Biochem **204:**39-49.

27. **Beyer, S., B. Kunze, B. Silakowski, and R. Muller.** 1999. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of Stigmatella aurantiaca Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain Sorangium cellulosum So ce90. Biochim Biophys Acta **1445:**185-95.

28. **Biebl, H., B. J. Tindall, R. Pukall, H. Lunsdorf, M. Allgaier, and I. Wagner-Dobler.** 2006. Hoeflea phototrophica sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll a. Int J Syst Evol Microbiol **56:**821-6.

29. **Birch, A. J., P. A. Massy-Westropp, and C. J. Moye.** 1955. Studies in relation to biosynthesis III. 2-Hydroxy-6-methylbenzoic acid in Penicillium griseofuluum Dierekx. Australian Journal of Chemistry **8:**539-544.

30. **Blahova, J., K. Kralikova, V. Krcmery, and P. Jezek.** 2000. Low-Frequency transduction of imipenem resistance and high-frequency transduction of ceftazidime and aztreonam resistance by the bacteriophage AP-151 isolated from a Pseudomonas aeruginosa strain. J Chemother **12:**482-6.

31. **Blahova, J., K. Kralikova, V. Krcmery, Sr., and N. Bartonikova.** 1999. High-frequency transduction of antibiotic resistance in Pseudomonas aeruginosa by a wild-type bacteriophage with restricted specificity for recipient strains. Eur J Clin Microbiol Infect Dis **18:**152-4.

32. **Bond, P. L., P. Hugenholtz, J. Keller, and L. L. Blackall.** 1995. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. Appl Environ Microbiol **61:**1910-6.

33. **Bonomo, R. A., and M. Tolmasky.** 2007. Enzyme-mediated resistance to antibiotics : mechanisms, dissemination, and prospects for inhibition. ASM Press, Washington, D.C.

34. **Brabban, A. D., E. Hite, and T. R. Callaway.** 2005. Evolution of foodborne pathogens via temperate bacteriophage-mediated gene transfer. Foodborne Pathog Dis **2:**287-303.

35. **Brady, S. F., C. J. Chao, and J. Clardy.** 2004. Long-chain N-acyltyrosine synthases from environmental DNA. Appl Environ Microbiol **70:**6865-70.

36. **Breitbart, M., I. Hewson, B. Felts, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer.** 2003. Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol **185:**6220-3.

37. **Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer.** 2002. Genomic analysis of uncultured marine viral communities. Proceedings of the National Academy of Sciences of the United States of America **99:**14250-14255.

38. **Brussow, H., C. Canchaya, and W. D. Hardt.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev **68:**560-602.

39. **Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brussow.** 2003. Phage as agents of lateral gene transfer. Curr Opin Microbiol **6:**417-24.

40. **Caracciolo, A. B., P. Grenni, R. Ciccoli, G. Di Landa, and C. Cremisini.** 2005. Simazine biodegradation in soil: analysis of bacterial community structure by in situ hybridization. Pest Manag Sci **61:**863-9.

41. **Cerda Zolezzi, P., L. M. Laplana, C. R. Calvo, P. G. Cepero, M. C. Erazo, and R. Gomez-Lus.** 2004. Molecular basis of resistance to macrolides and other antibiotics in commensal viridans group streptococci and Gemella spp. and transfer of resistance genes to Streptococcus pneumoniae. Antimicrob Agents Chemother **48:**3462-7.

42. **Chen, I. C., W. D. Lin, S. K. Hsu, V. Thiruvengadam, and W. H. Hsu.** 2009. Isolation and characterization of a novel lysine racemase from a soil metagenomic library. Appl Environ Microbiol **75:**5161-6.

43. **Chen, Y., M. G. Dumont, J. D. Neufeld, L. Bodrossy, N. Stralis-Pavese, N. P. McNamara, N. Ostle, M. J. Briones, and J. C. Murrell.** 2008. Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated Methylocystis in acidic peatlands. Environ Microbiol **10:**2609-22.

44. **Chopra, I., and M. Roberts.** 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev **65:**232-60 ; second page, table of contents.

45. **Chung, E. J., H. K. Lim, J. C. Kim, G. J. Choi, E. J. Park, M. H. Lee, Y. R. Chung, and S. W. Lee.** 2008. Forest soil metagenome gene cluster involved in antifungal activity expression in Escherichia coli. Appl Environ Microbiol **74:**723-30.

46. **Cieslinski, H., A. Dlugolecka, J. Kur, and M. Turkiewicz.** 2009. An MTA phosphorylase gene discovered in the metagenomic library derived from Antarctic top soil during screening for lipolytic active clones confers strong pink fluorescence in the presence of rhodamine B. FEMS Microbiol Lett **299:**232-40.

47. **Clardy, J., and C. Walsh.** 2004. Lessons from natural molecules. Nature **432:**829-37.

48. **Cochetti, I., E. Tili, M. Vecchi, A. Manzin, M. Mingoia, P. E. Varaldo, and M. P. Montanari.** 2007. New Tn916-related elements causing erm(B)-mediated erythromycin resistance in tetracycline-susceptible pneumococci. J Antimicrob Chemother **60:**127-31.

49. **Collingro, A., S. Poppert, E. Heinz, S. Schmitz-Esser, A. Essig, M. Schweikert, M. Wagner, and M. Horn.** 2005. Recovery of an environmental Chlamydia strain from activated sludge by co-cultivation with Acanthamoeba sp. Microbiology **151:**301-9.

50. **Collins, J., and B. Hohn.** 1978. Cosmids: a type of plasmid gene-cloning vector that is packageable in vitro in bacteriophage lambda heads. Proc Natl Acad Sci U S A **75:**4242-6.

51. **Connell, S. R., D. M. Tracz, K. H. Nierhaus, and D. E. Taylor.** 2003. Ribosomal protection proteins and their mechanism of tetracycline resistance. Antimicrob Agents Chemother **47:**3675-81.

52. **Connell, S. R., C. A. Trieber, G. P. Dinos, E. Einfeldt, D. E. Taylor, and K. H. Nierhaus.** 2003. Mechanism of Tet(O)-mediated tetracycline resistance. Embo J **22:**945-53.

53. **Cook, K. L., A. C. Layton, H. M. Dionisi, J. T. Fleming, and G. S. Sayler.** 2004. Evaluation of a plasmid-based 16S-23S rDNA intergenic spacer region array for analysis of microbial diversity in industrial wastewater. J Microbiol Methods **57:**79-93.

54. **Courtois, S., C. M. Cappellano, M. Ball, F. X. Francou, P. Normand, G. Helynck, A. Martinez, S. J. Kolvek, J. Hopke, M. S. Osburne, P. R. August, R. Nalin, M. Guerineau, P. Jeannin, P. Simonet, and J. L. Pernodet.** 2003.

Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol **69:**49-55.

55. **Craig, J. W., F. Y. Chang, and S. F. Brady.** 2009. Natural products from environmental DNA hosted in Ralstonia metallidurans. ACS Chem Biol **4:**23-8.

56. **Créach, V., Lucas, F., Deleu, C., Bertru, G., Mariotti, A. .** 1999. Combination of biomolecular and stable isotope techniques to determine the origin of organic matter used by bacterial communities: application to sediment. J Microbiol Methods **38:**43-52.

57. **Cullings, K. W.** 1992. Design and testing of a plant-specific PCR primer for ecological and evolutionary studies. Molecular Ecology **1:**233-240.

58. **Curtis, T. P., and W. T. Sloan.** 2005. Microbiology. Exploring microbial diversity--a vast below. Science **309:**1331-3.

59. **De Gelder, L., F. P. Vandecasteele, C. J. Brown, L. J. Forney, and E. M. Top.** 2005. Plasmid donor affects host range of promiscuous IncP-1beta plasmid pB10 in an activated-sludge microbial community. Appl Environ Microbiol **71:**5309-17.

60. **Debartolomeis, J., and V. J. Cabelli.** 1991. Evaluation of an Escherichia coli host strain for enumeration of F male-specific bacteriophages. Appl Environ Microbiol **57:**1301-5.

61. **Dedysh, S. N., M. Derakshani, and W. Liesack.** 2001. Detection and enumeration of methanotrophs in acidic Sphagnum peat by 16S rRNA fluorescence in situ hybridization, including the use of newly developed oligonucleotide probes for Methylocella palustris. Appl Environ Microbiol **67:**4850-7.

62. **Depardieu, F., I. Podglajen, R. Leclercq, E. Collatz, and P. Courvalin.** 2007. Modes and modulations of antibiotic resistance gene expression. Clin Microbiol Rev **20:**79-114.

63. **Di Gioia, D., L. Fambrini, E. Coppini, F. Fava, and C. Barberio.** 2004. Aggregation-based cooperation during bacterial aerobic degradation of polyethoxylated nonylphenols. Research in Microbiology **155:**761-769.

64. **Dinsdale, E. A., R. A. Edwards, D. Hall, F. Angly, M. Breitbart, J. M. Brulc, M. Furlan, C. Desnues, M. Haynes, L. L. Li, L. McDaniel, M. A. Moran, K. E. Nelson, C. Nilsson, R. Olson, J. Paul, B. R. Brito, Y. J. Ruan, B. K. Swan, R. Stevens, D. L. Valentine, R. V. Thurber, L. Wegley, B. A. White, and F. Rohwer.** 2008. Functional metagenomic profiling of nine biomes (vol 452, pg 629, 2008). Nature **455:**830-830.

65. **Donadio, S., M. J. Staver, J. B. McAlpine, S. J. Swanson, and L. Katz.** 1991. Modular organization of genes required for complex polyketide biosynthesis. Science **252:**675-9.

66. **Droge, M., A. Puhler, and W. Selbitschka.** 2000. Phenotypic and molecular characterization of conjugative antibiotic resistance plasmids isolated from bacterial communities of activated sludge. Mol Gen Genet **263:**471-82.

67. **Duarte, G. F., A. S. Rosado, L. Seldin, A. C. Keijzer-Wolters, and J. D. van Elsas.** 1998. Extraction of ribosomal RNA and genomic DNA from soil for studying the diversity of the indigenous bacterial community. Journal of Microbiological Methods **32:**21-29.

68. **Dumont, M. G., S. M. Radajewski, C. B. Miguez, I. R. McDonald, and J. C. Murrell.** 2006. Identification of a complete methane monooxygenase operon from soil by combining stable isotope probing and metagenomic analysis. Environ Microbiol **8:**1240-50.

69. **Dzidic, S., and V. Bedekovic.** 2003. Horizontal gene transfer-emerging multidrug resistance in hospital bacteria. Acta Pharmacol Sin **24:**519-26.

70. **Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick.** 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. Chem Biol **11:**817-33.

71. **Elend, C., C. Schmeisser, C. Leggewie, P. Babiak, J. D. Carballeira, H. L. Steele, J. L. Reymond, K. E. Jaeger, and W. R. Streit.** 2006. Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl Environ Microbiol **72:**3637-45.

72. **Entcheva, P., W. Liebl, A. Johann, T. Hartsch, and W. R. Streit.** 2001. Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia. Appl Environ Microbiol **67:**89-99.

73. **Faegri, A., V. Torsvik, and J. Goksoyr.** 1977. Bacterial and fungal activities in soil: separation of bacteria and fungi by a rapid fractionated centrifugation technique. Soil Biol Biochem **9:**105-112.

74. **Feinstein, L. M., W. J. Sul, and C. B. Blackwood.** 2009. Assessment of bias associated with incomplete extraction of microbial DNA from soil. Appl Environ Microbiol **75:**5428-33.

75. **Figuerola, E. L., and L. Erijman.** 2007. Bacterial taxa abundance pattern in an industrial wastewater treatment system determined by the full rRNA cycle approach. Environ Microbiol **9:**1780-9.

76. **Fluit, A. C., M. R. Visser, and F. J. Schmitz.** 2001. Molecular detection of antimicrobial resistance. Clin Microbiol Rev **14:**836-71, table of contents.

77. **Fong, D. H., and A. M. Berghuis.** 2002. Substrate promiscuity of an aminoglycoside antibiotic resistance enzyme via target mimicry. Embo J **21:**2323-31.

78. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol **3:**722-32.

79. **Frostegard, A., S. Courtois, V. Ramisse, S. Clerc, D. Bernillon, F. Le Gall, P. Jeannin, X. Nesme, and P. Simonet.** 1999. Quantification of bias related to the extraction of DNA directly from soils. Appl Environ Microbiol **65:**5409-20.

80. **Funa, N., Y. Ohnishi, Y. Ebizuka, and S. Horinouchi.** 2002. Properties and substrate specificity of RppA, a chalcone synthase-related polyketide synthase in Streptomyces griseus. J Biol Chem **277:**4628-35.

81. **Gabor, E. M., E. J. Vries, and D. B. Janssen.** 2003. Efficient recovery of environmental DNA for expression cloning by indirect extraction methods. FEMS Microbiol Ecol **44:**153-63.

82. **Gallert, C., K. Fund, and J. Winter.** 2005. Antibiotic resistance of bacteria in raw and biologically treated sewage and in groundwater below leaking sewers. Appl Microbiol Biotechnol **69:**106-12.

83. **Garcia Martin, H., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, E. Szeto, E. Dalin, N. H. Putnam, H. J. Shapiro, J. L. Pangilinan, I. Rigoutsos, N. C. Kyrpides, L. L. Blackall, K. D. McMahon, and P. Hugenholtz.** 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat Biotechnol **24:**1263-9.

84. **Gassama, A., A. Aidara-Kane, D. Chainier, F. Denis, and M. C. Ploy.** 2004. Integron-associated antibiotic resistance in enteroaggregative and enteroinvasive Escherichia coli. Microb Drug Resist **10:**27-30.

85. **Gerth, K., N. Bedorf, G. Hofle, H. Irschik, and H. Reichenbach.** 1996. Epothilons A and B: antifungal and cytotoxic compounds from Sorangium cellulosum (Myxobacteria). Production, physico-chemical and biological properties. J Antibiot (Tokyo) **49:**560-3.

86. **Gerth, K., N. Bedorf, H. Irschik, G. Hofle, and H. Reichenbach.** 1994. The soraphens: a family of novel antifungal compounds from Sorangium cellulosum (Myxobacteria). I. Soraphen A1 alpha: fermentation, isolation, biological properties. J Antibiot (Tokyo) **47:**23-31.

87. **Gillespie, D. E., S. F. Brady, A. D. Bettermann, N. P. Cianciotto, M. R. Liles, M. R. Rondon, J. Clardy, R. M. Goodman, and J. Handelsman.** 2002. Isolation of antibiotics turbomycin a and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol **68:**4301-6.

88. **Ginige, M. P., P. Hugenholtz, H. Daims, M. Wagner, J. Keller, and L. L. Blackall.** 2004. Use of stable-isotope probing, full-cycle rRNA analysis, and fluorescence in situ hybridization-microautoradiography to study a methanol-fed denitrifying microbial community. Appl Environ Microbiol **70:**588-96.

89. **Ginolhac, A., C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T. M. Vogel, G. Perriere, P. Simonet, and R. Nalin.** 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol **70:**5522-7.

90. **Goyal, S. M., C. P. Gerba, and G. Bitton.** 1987. Phage ecology. Wiley, New York.

91. **Guardabassi, L., S. Schwarz, and D. H. Lloyd.** 2004. Pet animals as reservoirs of antimicrobial-resistant bacteria. J Antimicrob Chemother **54:**321-32.

92. **Hahn, D., R. I. Amann, and J. Zeyer.** 1993. Whole-cell hybridization of Frankia strains with fluorescence- or digoxigenin-Labeled, 16S rRNA-targeted oligonucleotide probes. Appl Environ Microbiol **59:**1709-1716.

93. **Halary, S., J. Leigh, B. Cheaib, P. Lopez, and E. Bapteste.** 2009. Network analyses structure genetic diversity in independent genetic worlds. Proc Natl Acad Sci U S A **http://www.pnas.org/cgi/doi/10.1073/pnas.0908978107**.

94. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev **68:**669-85.

95. **Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman.** 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol **5:**R245-9.

96. **Hantula, J., A. Kurki, P. Vuoriranta, and D. H. Bamford.** 1991. Ecology of bacteriophages infecting activated sludge bacteria. Appl Environ Microbiol **57:**2147-51.

97. **Harry, M., B. Gambier, Y. Bourezgui, and E. Garnier-Sillam.** 1999. Evaluation of purification procedures for DNA extracted from organic rich samples: interference with humic substances. Analysis **27:**439-442.

98. **Hasman, H., F. M. Aarestrup, A. Dalsgaard, and L. Guardabassi.** 2006. Heterologous expression of glycopeptide resistance vanHAX gene clusters from soil bacteria in Enterococcus faecalis. J Antimicrob Chemother **57:**648-53.

99. **Heath, C., X. P. Hu, S. C. Cary, and D. Cowan.** 2009. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. Appl Environ Microbiol **75:**4657-9.

100. **Henne, A., R. Daniel, R. A. Schmitz, and G. Gottschalk.** 1999. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4-hydroxybutyrate. Appl Environ Microbiol **65:**3901-7.

101. **Herrick, J. B., E. L. Madsen, C. A. Batt, and W. C. Ghiorse.** 1993. Polymerase chain reaction amplification of naphthalene-catabolic and 16S rRNA gene sequences from indigenous sediment bacteria. Appl Environ Microbiol **59:**687-94.

102. **Hiraishi, A., M. Iwasaki, and H. Shinjo.** 2000. Terminal restriction pattern analysis of 16S rRNA genes for the characterization of bacterial communities of activated sludge. J Biosci Bioeng **90:**148-56.

103. **Hochhut, B., Y. Lotfi, D. Mazel, S. M. Faruque, R. Woodgate, and M. K. Waldor.** 2001. Molecular analysis of antibiotic resistance gene clusters in vibrio cholerae O139 and O1 SXT constins. Antimicrob Agents Chemother **45:**2991-3000.

104. **Holben, W. E., J. K. Jansson, B. K. Chelm, and J. M. Tiedje.** 1988. DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. Appl Environ Microbiol **54:**703-711.

105. **Hopkins, D. W., S. J. Macnaughton, and A. G. O'Donnell.** 1991. A dispersion and differential centrifugation technique for representatively sampling microorganisms from soil. Soil Biol Biochem **23:**217-225.

106. **Horn, M., and M. Wagner.** 2004. Bacterial endosymbionts of free-living amoebae. J Eukaryot Microbiol **51:**509-14.

107. **Huang, W. E., K. Stoecker, R. Griffiths, L. Newbold, H. Daims, A. S. Whiteley, and M. Wagner.** 2007. Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ Microbiol **9:**1878-89.

108. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180:**4765-74.

109. **Hugenholtz, P., and G. W. Tyson.** 2008. Microbiology: metagenomics. Nature **455:**481-3.

110. **Hughes, V. M., and N. Datta.** 1983. Conjugative plasmids in bacteria of the 'pre-antibiotic' era. Nature **302:**725-6.

111. **Imachi, H., Y. Sekiguchi, Y. Kamagata, A. Loy, Y. L. Qiu, P. Hugenholtz, N. Kimura, M. Wagner, A. Ohashi, and H. Harada.** 2006. Non-sulfate-reducing, syntrophic bacteria affiliated with desulfotomaculum cluster I are widely distributed in methanogenic environments. Appl Environ Microbiol **72:**2080-91.

112. **Izard, T., and J. Ellis.** 2000. The crystal structures of chloramphenicol phosphotransferase reveal a novel inactivation mechanism. Embo J **19:**2690-700.

113. **Jacobsen, C. S., and O. F. Rasmussen.** 1992. Development and application of a new method to extract bacterial DNA from soil based on separation of bacteria from soil with cation-exchange resin. Appl Environ Microbiol **58:**2458-2462.

114. **Jeon, J. H., J. T. Kim, S. G. Kang, J. H. Lee, and S. J. Kim.** 2009. Characterization and its potential application of two esterases derived from the arctic sediment metagenome. Mar Biotechnol (NY) **11:**307-16.

115. **Jiang, H. L., A. M. Maszenan, and J. H. Tay.** 2007. Bioaugmentation and coexistence of two functionally similar bacterial strains in aerobic granules. Appl Microbiol Biotechnol **75:**1191-200.

116. **Jones, B. V., and J. R. Marchesi.** 2007. Transposon-aided capture (TRACA) of plasmids resident in the human gut mobile metagenome. Nat Methods **4:**55-61.

117. **Kalyuzhnaya, M. G., A. Lapidus, N. Ivanova, A. C. Copeland, A. C. McHardy, E. Szeto, A. Salamov, I. V. Grigoriev, D. Suciu, S. R. Levine, V. M. Markowitz, I. Rigoutsos, S. G. Tringe, D. C. Bruce, P. M. Richardson, M. E. Lidstrom, and L. Chistoserdova.** 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. Nat Biotechnol **26:**1029-34.

118. **Kampfer, P., R. Erhart, C. Beimfohr, J. Bohringer, M. Wagner, and R. Amann.** 1996. Characterization of Bacterial Communities from Activated Sludge: Culture-Dependent Numerical Identification Versus In Situ Identification Using Group- and Genus-Specific rRNA-Targeted Oligonucleotide Probes. Microb Ecol **32:**101-21.

119. **Kapley, A., T. De Baere, and H. J. Purohit.** 2007. Eubacterial diversity of activated biomass from a common effluent treatment plant. Res Microbiol **158:**494-500.

120. **Kayser, F. H.** 2005. Medical microbiology. Georg Thieme Verlag, Stuttgart ; New York, NY.

121. **Kelch, W. J., and J. S. Lee.** 1978. Antibiotic resistance patterns of gram-negative bacteria isolated from environmental sources. Appl Environ Microbiol **36:**450-6.

122. **Kim, B. S., S. Y. Kim, J. Park, W. Park, K. Y. Hwang, Y. J. Yoon, W. K. Oh, B. Y. Kim, and J. S. Ahn.** 2007. Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. J Appl Microbiol **102:**1392-400.

123. **Kim, U. J., H. Shizuya, P. J. de Jong, B. Birren, and M. I. Simon.** 1992. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucleic Acids Res **20:**1083-5.

124. **Knaebel, D. B., and R. L. Crawford.** 1995. Extraction and purification of microbial DNA from petroleum-contaminated soils and detection of low numbers of toluene, octane and pesticide degraders by multiplex polymerase chain reaction and Southern analysis. Mol Ecol **4:**579-91.

125. **Knietsch, A., S. Bowien, G. Whited, G. Gottschalk, and R. Daniel.** 2003. Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. Appl Environ Microbiol **69:**3048-60.

126. **Kowalchuk, G. A.** 2004. Molecular microbial ecology manual, 2nd ed. Kluwer Academic, Dordrecht ; London.

127. **Kragelund, C., C. Levantesi, A. Borger, K. Thelen, D. Eikelboom, V. Tandoi, Y. Kong, J. Krooneman, P. Larsen, T. R. Thomsen, and P. H. Nielsen.** 2008. Identity, abundance and ecophysiology of filamentous bacteria belonging to the Bacteroidetes present in activated sludge plants. Microbiology **154:**886-94.

128. **Kragelund, C., C. Levantesi, A. Borger, K. Thelen, D. Eikelboom, V. Tandoi, Y. Kong, J. van der Waarde, J. Krooneman, S. Rossetti, T. R. Thomsen, and P. H. Nielsen.** 2007. Identity, abundance and ecophysiology of filamentous Chloroflexi species present in activated sludge treatment plants. FEMS Microbiol Ecol **59:**671-82.

129. **Kreader, C. A.** 1996. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. Appl Environ Microbiol **62:**1102-6.

130. **Krsek, M., and E. M. Wellington.** 1999. Comparison of different methods for the isolation and purification of total community DNA from soil. J Microbiol Methods **39:**1-16.

131. **Kumar, S., M. Nei, J. Dudley, and K. Tamura.** 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform **9:**299-306.

132. **Kurokawa, K., T. Itoh, T. Kuwahara, K. Oshima, H. Toh, A. Toyoda, H. Takami, H. Morita, V. K. Sharma, T. P. Srivastava, T. D. Taylor, H. Noguchi, H. Mori, Y. Ogura, D. S. Ehrlich, K. Itoh, T. Takagi, Y. Sakaki, T. Hayashi, and M. Hattori.** 2007. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. DNA Res **14:**169-81.

133. **Kurola, J., C. Wittmann, M. Salkinoja-Salonen, T. Aarnio, and M. Romantschuk.** 2005. Application of cation-exchange membranes for characterisation and imaging ammonia-oxidising bacteria in soils. FEMS Microbiol Ecol **53:**463-72.

134. **Lammle, K., H. Zipper, M. Breuer, B. Hauer, C. Buta, H. Brunner, and S. Rupp.** 2007. Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. J Biotechnol **127:**575-92.

135. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A **82:**6955-9.

136. **Lee, S. W., K. Won, H. K. Lim, J. C. Kim, G. J. Choi, and K. Y. Cho.** 2004. Screening for novel lipolytic enzymes from uncultured soil microorganisms. Appl Microbiol Biotechnol **65:**720-6.

137. **Lee, S. Y., J. Bollinger, D. Bezdicek, and A. Ogram.** 1996. Estimation of the abundance of an uncultured soil bacterial strain by a competitive quantitative PCR method. Appl Environ Microbiol **62:**3787-93.

138. **Lewis, K.** 2002. Bacterial resistance to antimicrobials. Marcel Dekker, New York.

139. **Li, X., and L. Qin.** 2005. Metagenomics-based drug discovery and marine microbial diversity. Trends Biotechnol **23:**539-43.

140. **Ligozzi, M., G. Lo Cascio, and R. Fontana.** 1998. vanA gene cluster in a vancomycin-resistant clinical isolate of Bacillus circulans. Antimicrob Agents Chemother **42:**2055-9.

141. **Liles, M. R., B. F. Manske, S. B. Bintrim, J. Handelsman, and R. M. Goodman.** 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl Environ Microbiol **69:**2684-91.

142. **Liles, M. R., L. L. Williamson, R. M. Goodman, and J. Handelsman.** 2004. Isolation of high molecular weight genomic DNA from soil bacteria for genomic library construction, p. 839-852. *In* G. A. Kowalchuk, F. J. de Bruijn, M. Head, A. D. Akkermans, and J. D. Van Elsas (ed.), Molecular microbial ecology manual, 2nd ed. Kluwer Academic Publishers, Dordrecht, The Netherlands.

143. **Liles, M. R., L. L. Williamson, J. Rodbumrer, V. Torsvik, R. M. Goodman, and J. Handelsman.** 2008. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. Appl Environ Microbiol **74:**3302-5.

144. **Liles, M. R., L. L. Williamson, J. Rodbumrer, V. Torsvik, L. C. Parsley, R. M. Goodman, and J. Handelsman.** 2009. Isolation and cloning of high molecular weight metagenomic DNA from soil microorganisms. Cold Spring Harbor Press Protocols **Online: http://cshprotocols.cshlp.org/cgi/content/full/2009/8/pdb.prot5271**

145. **Limpiyakorn, T., Y. Shinohara, F. Kurisu, and O. Yagi.** 2004. Distribution of ammonia-oxidizing bacteria in sewage activated sludge: analysis based on 16S rDNA sequence. Water Sci Technol **50:**9-14.

146. **Liu, W., H. J. Kim, E. M. Lucchetta, W. Du, and R. F. Ismagilov.** 2009. Isolation, incubation, and parallel functional testing and identification by FISH of rare microbial single-copy cells from multi-species mixtures using the combination of chemistrode and stochastic confinement. Lab Chip **9:**2153-62.

147. **Lu, W. X., Z. Q. Wei, Y. S. Yu, Y. G. Chen, and L. J. Li.** 2005. Study on the genetic structure and transmission mechanism of a plasmid-mediated AmpC beta-lactamase. Zhonghua Yi Xue Za Zhi **85:**3043-7.

148. **Lucker, S., D. Steger, K. U. Kjeldsen, B. J. MacGregor, M. Wagner, and A. Loy.** 2007. Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence in situ hybridization. J Microbiol Methods **69:**523-8.

149. **Macdonald, R. M.** 1986. Sampling soil microflora-dispersion of soil by ion-exvhange and extraction of specific microorganisms from suspension by elutriation. Soil Biol Biochem **18:**399-406.

150. **Madigan, M. T., J. M. Martinko, P. V. Dunlap, and D. P. Clark.** 2005. Brock's biology of microorganisms, 11th ed. Benjamin Cummings.

151. **Malik, R., P. Pristas, and P. Javorsky.** 2004. Occurrence of plasmid-mediated ampicillin resistance among enterobacteria from the ovine rumen. Folia Microbiol (Praha) **49:**187-90.

152. **Malpartida, F., and D. A. Hopwood.** 1984. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. Nature **309:**462-4.

153. **Manefield, M., R. Griffiths, N. P. McNamara, D. Sleep, N. Ostle, and A. Whiteley.** 2007. Insights into the fate of a 13C labelled phenol pulse for stable isotope probing (SIP) experiments. J Microbiol Methods **69:**340-4.

154. **Mantle, P. G., and K. M. Macgeorge.** 1990. The role of monic acid A in pseudomonic acid A biosynthesis in Pseudomonas fluorescens. Appl Microbiol Biotechnol **33:**709-11.

155. **Manz, W., M. Wagner, R. Amann, and K.-H. Scleifer.** 1994. In situ characterization of the microbial consortia active in two wastewater treatment plants. Water Res **28:**1715-1723.

156. **Markogiannakis, A., P. T. Tassios, M. Lambiri, L. R. Ward, J. Kourea-Kremastinou, N. J. Legakis, and A. C. Vatopoulos.** 2000. Multiple clones within multidrug-resistant Salmonella enterica serotype Typhimurium phage type DT104. The Greek Nontyphoidal Salmonella Study Group. J Clin Microbiol **38:**1269-71.

157. **Martinez, J. L., and F. Baquero.** 2002. Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance. Clin Microbiol Rev **15:**647-79.

158. **Mascaretti, O. A.** 2003. Bacteria versus antibacterial agents : an integrated approach. ASM Press, Washington, D.C.

159. **Massova, I., and S. Mobashery.** 1998. Kinship and diversification of bacterial penicillin-binding proteins and beta-lactamases. Antimicrob Agents Chemother **42:**1-17.

160. **McDermott, P. F., S. Zhao, D. D. Wagner, S. Simjee, R. D. Walker, and D. G. White.** 2002. The food safety perspective of antibiotic resistance. Anim Biotechnol **13:**71-84.

161. **Medeiros, A. A.** 1997. Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. Clin Infect Dis **24 Suppl 1:**S19-45.

162. **Metsa-Ketela, M., L. Halo, E. Munukka, J. Hakala, P. Mantsala, and K. Ylihonko.** 2002. Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. Appl Environ Microbiol **68:**4472-9.

163. **Metzker, M. L.** 2005. Emerging technologies in DNA sequencing. Genome Res **15:**1767-76.

164. **Miller, D. N., J. E. Bryant, E. L. Madsen, and W. C. Ghiorse.** 1999. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. Appl Environ Microbiol **65:**4715-24.

165. **Miller, R. V., and M. J. Day.** 2004. Microbial evolution : gene establishment, survival, and exchange. ASM Press, Washington, D.C.

166. **Mirete, S., C. G. de Figueras, and J. E. Gonzalez-Pastor.** 2007. Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. Appl Environ Microbiol **73:**6001-11.

167. **Moffitt, M. C., and B. A. Neilan.** 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. J Mol Evol **56:**446-57.

168. **Monpoeho, S., A. Maul, B. Mignotte-Cadiergues, L. Schwartzbrod, S. Billaudel, and V. Ferre.** 2001. Best Viral Elution Method Available for Quantification of Enteroviruses in Sludge by Both Cell Culture and Reverse Transcription-PCR. Applied and Environmental Microbiology **67:**2484-2488.

169. **Mooij, M. J., I. Schouten, G. Vos, A. Van Belkum, C. M. Vandenbroucke-Grauls, P. H. Savelkoul, and C. Schultsz.** 2005. Class 1 integrons in ciprofloxacin-resistant Escherichia coli strains from two Dutch hospitals. Clin Microbiol Infect **11:**898-902.

170. **Moore, B. S., and J. N. Hopke.** 2001. Discovery of a new bacterial polyketide biosynthetic pathway. Chembiochem **2:**35-8.

171. **Moore, R. A., D. DeShazer, S. Reckseidler, A. Weissman, and D. E. Woods.** 1999. Efflux-mediated aminoglycoside and macrolide resistance in Burkholderia pseudomallei. Antimicrob Agents Chemother **43:**465-70.

172. **More, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl Environ Microbiol **60:**1572-80.

173. **Moreira, D.** 1998. Efficient removal of PCR inhibitors using agarose-embedded DNA preparations. Nucleic Acids Res **26:**3309-10.

174. **Motamedi, H., and C. R. Hutchinson.** 1987. Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumor antibiotic of Streptomyces glaucescens. Proc Natl Acad Sci U S A **84:**4445-9.

175. **Murata, T., M. Ohnishi, T. Ara, J. Kaneko, C. G. Han, Y. F. Li, K. Takashima, H. Nojima, K. Nakayama, A. Kaji, Y. Kamio, T. Miki, H. Mori, E. Ohtsubo, Y. Terawaki, and T. Hayashi.** 2002. Complete nucleotide sequence of plasmid Rts1: implications for evolution of large plasmid genomes. J Bacteriol **184:**3194-202.

176. **Muyzer, G., E. C. de Waal, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl Environ Microbiol **59:**695-700.

177. **Newman, M. M., J. W. Feminella, and M. R. Liles.** In press. Purification of genomic DNA from environmental sources for use in a polymerase chain reaction. Cold Spring Harbor Press Protocols.

178. **Nielsen, J. L., S. Juretschko, M. Wagner, and P. H. Nielsen.** 2002. Abundance and phylogenetic affiliation of iron reducers in activated sludge as assessed by fluorescence in situ hybridization and microautoradiography. Appl Environ Microbiol **68:**4629-36.

179. **Nishimura, K., T. Hosaka, S. Tokuyama, S. Okamoto, and K. Ochi.** 2007. Mutations in rsmG, encoding a 16S rRNA methyltransferase, result in low-level streptomycin resistance and antibiotic overproduction in Streptomyces coelicolor A3(2). J Bacteriol **189:**3876-83.

180. **Nwosu, V. C.** 2001. Antibiotic resistance with particular reference to soil microorganisms. Res Microbiol **152:**421-30.

181. **O'Hagan, D.** 1991. The Polyketide Metabolites. Ellis Horwood, New York.

182. **Ogram, A., G. S. Sayler, and T. J. Barkay.** 1987. DNA extraction and purification from sediments. J Microbiol Methods **7:**57-66.

183. **Olsen, K. N., M. Henriksen, M. Bisgaard, O. L. Nielsen, and H. Christensen.** 2008. Investigation of chicken intestinal bacterial communities by 16S rRNA targeted fluorescence in situ hybridization. Antonie Van Leeuwenhoek **94:**423-37.

184. **Osborn, A. M., and C. J. Smith.** 2005. Molecular microbial ecology. Taylor & Francis, New York ; Abingdon [England].

185. **Osoegawa, K., P. Y. Woon, B. Zhao, E. Frengen, M. Tateno, J. J. Catanese, and P. J. de Jong.** 1998. An improved approach for construction of bacterial artificial chromosome libraries. Genomics **52:**1-8.

186. **Otawa, K., R. Asano, Y. Ohba, T. Sasaki, E. Kawamura, F. Koyama, S. Nakamura, and Y. Nakai.** 2006. Molecular analysis of ammonia-oxidizing bacteria community in intermittent aeration sequencing batch reactors used for animal wastewater treatment. Environ Microbiol **8:**1985-96.

187. **Otawa, K., S. H. Lee, A. Yamazoe, M. Onuki, H. Satoh, and T. Mino.** 2007. Abundance, diversity, and dynamics of viruses on microorganisms in activated sludge processes. Microb Ecol **53:**143-52.

188. **Ouyang, Y., S. Dai, L. Xie, M. S. Ravi Kumar, W. Sun, H. Sun, D. Tang, and X. Li.** 2009. Isolation of High Molecular Weight DNA from Marine Sponge Bacteria for BAC Library Construction. Mar Biotechnol (NY).

189. **Overbeek, R., T. Begley, R. M. Butler, J. V. Choudhuri, H. Y. Chuang, M. Cohoon, V. de Crecy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. D. Frank, S. Gerdes, E. M. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. C. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. D. Pusch, D. A. Rodionov, C. Ruckert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. Nucleic Acids Res **33:**5691-702.

190. **Palackal, N., C. S. Lyon, S. Zaidi, P. Luginbuhl, P. Dupree, F. Goubet, J. L. Macomber, J. M. Short, G. P. Hazlewood, D. E. Robertson, and B. A. Steer.** 2007. A multifunctional hybrid glycosyl hydrolase discovered in an uncultured microbial consortium from ruminant gut. Appl Microbiol Biotechnol **74:**113-24.

191. **Pallecchi, L., C. Lucchetti, A. Bartoloni, F. Bartalesi, A. Mantella, H. Gamboa, A. Carattoli, F. Paradisi, and G. M. Rossolini.** 2007. Population structure and resistance genes in antibiotic-resistant bacteria from a remote community with minimal antibiotic exposure. Antimicrob Agents Chemother **51:**1179-84.

192. **Parsley, L. C., E. J. Consuegra, S. J. Thomas, J. Bhavsar, A. M. Land, N. N. Bhuiyan, M. A. Mazher, R. J. Waters, K. E. Wommack, W. F. Harper, Jr., and M. R. Liles.** 2010. Census of the viral metagenome within an activated sludge microbial assemblage. Appl Environ Microbiol **76:**2673-7.

193. **Parsley, L. C., E. J. Consuegra, S. J. Thomas, J. Bhavsar, A. M. Land, N. N. Bhuiyan, M. A. Mazher, R. J. Waters, K. E. Wommack, W. F. Harper, Jr., and M. R. Liles.** 2010. A census of the viral metagenome within an activated sludge microbial assemblage. Appl Environ Microbiol **Published ahead of print:**doi:10.1128/AEM.02520-09.

194. **Pavlekovic, M., M. C. Schmid, N. Schmider-Poignee, S. Spring, M. Pilhofer, T. Gaul, M. Fiandaca, F. E. Loffler, M. Jetten, K. H. Schleifer, and N. M. Lee.** 2009. Optimization of three FISH procedures for in situ detection of anaerobic ammonium oxidizing bacteria in biological wastewater treatment. J Microbiol Methods **78:**119-26.

195. **Pfeifer, B. A., and C. Khosla.** 2001. Biosynthesis of polyketides in heterologous hosts. Microbiol Mol Biol Rev **65:**106-18.

196. **Phillips, G., and B. E. Funnell.** 2004. Plasmid biology. ASM Press, Washington, D.C.

197. **Piel, J.** 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. Proc Natl Acad Sci U S A **99:**14002-7.

198. **Piepersberg, W., B. Brau, and J. Davies.** 1983. High frequency transduction of R-factor encoded gentamicin resistance by bacteriophage P1Cm. J Antibiot (Tokyo) **36:**588-94.

199. **Pillai, S. D., K. L. Josephson, R. L. Bailey, C. P. Gerba, and I. L. Pepper.** 1991. Rapid method for processing soil samples for polymerase chain reaction amplification of specific gene sequences. Appl Environ Microbiol **57:**2283-6.

200. **Poole, K.** 2001. Overcoming antimicrobial resistance by targeting resistance mechanisms. J Pharm Pharmacol **53:**283-94.

201. **Pope, P. B., and B. K. Patel.** 2008. Metagenomic analysis of a freshwater toxic cyanobacteria bloom. FEMS Microbiol Ecol **64:**9-27.

202. **Porteous, L. A., R. J. Seidler, and L. S. Watrud.** 1997. An improved method for purifying DNA from soil for polymerase chain reaction amplification and molecular ecology applications. Mol Ecol **6:**787-791.

203. **Quaiser, A., T. Ochsenreiter, H. P. Klenk, A. Kletzin, A. H. Treusch, G. Meurer, J. Eck, C. W. Sensen, and C. Schleper.** 2002. First insight into the genome of an uncultivated crenarchaeote from soil. Environ Microbiol **4:**603-11.

204. **Randall, L. P., S. W. Cooles, M. K. Osborn, L. J. Piddock, and M. J. Woodward.** 2004. Antibiotic resistance genes, integrons and multiple antibiotic resistance in thirty-five serotypes of Salmonella enterica isolated from humans and animals in the UK. J Antimicrob Chemother **53:**208-16.

205. **Rangaswamy, V., S. Jiralerspong, R. Parry, and C. L. Bender.** 1998. Biosynthesis of the Pseudomonas polyketide coronafacic acid requires monofunctional and multifunctional polyketide synthase proteins. Proc Natl Acad Sci U S A **95:**15469-74.

206. **Rashamuse, K., V. Magomani, T. Ronneburg, and D. Brady.** 2009. A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous beta-lactamase activity on nitrocefin. Appl Microbiol Biotechnol **83:**491-500.

207. **Rezania, B., J. A. Oleszkiewicz, N. Cicek, and H. Mo.** 2005. Hydrogen-dependent denitrification in an alternating anoxic-aerobic SBR membrane bioreactor. Water Sci Technol **51:**403-9.

208. **Richardson, T. H., X. Tan, G. Frey, W. Callen, M. Cabell, D. Lam, J. Macomber, J. M. Short, D. E. Robertson, and C. Miller.** 2002. A novel, high performance enzyme for starch liquefaction. Discovery and optimization of a low pH, thermostable alpha-amylase. J Biol Chem **277:**26501-7.

209. **Ridley, C. P., H. Y. Lee, and C. Khosla.** 2008. Evolution of polyketide synthases in bacteria. Proc Natl Acad Sci U S A **105:**4595-600.

210. **Riesenfeld, C. S., R. M. Goodman, and J. Handelsman.** 2004. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. Environ Microbiol **6:**981-9.

211. **Roberts, G. A., J. Staunton, and P. F. Leadlay.** 1993. Heterologous expression in Escherichia coli of an intact multienzyme component of the erythromycin-producing polyketide synthase. Eur J Biochem **214:**305-11.

212. **Robertson, D. E., J. A. Chaplin, G. DeSantis, M. Podar, M. Madden, E. Chi, T. Richardson, A. Milan, M. Miller, D. P. Weiner, K. Wong, J. McQuaid, B. Farwell, L. A. Preston, X. Tan, M. A. Snead, M. Keller, E. Mathur, P. L. Kretz, M. J. Burk, and J. M. Short.** 2004. Exploring nitrilase sequence space for enantioselective catalysis. Appl Environ Microbiol **70:**2429-36.

213. **Rochelle, P. A., J. C. Fry, R. J. Parkes, and A. J. Weightman.** 1992. DNA extraction for 16S rRNA gene analysis to determine genetic diversity in deep sediment communities. FEMS Microbiol Lett **100:**59-66.

214. **Roesch LF, F. R., Riva A, Casella G, Hadwin AKM, et al.** 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. **1:**ISME J. .

215. **Roller, C., M. Wagner, R. Amann, W. Ludwig, and K. H. Schleifer.** 1994. In situ probing of gram-positive bacteria with high DNA G + C content using 23S rRNA-targeted oligonucleotides. Microbiology **140 (Pt 10):**2849-58.

216. **Romanowski, G., M. G. Lorenz, and W. Wackernagel.** 1993. Use of polymerase chain reaction and electroporation of Escherichia coli to monitor the persistence of extracellular plasmid DNA introduced into natural soils. Appl Environ Microbiol **59:**3438-46.

217. **Ronaghi, M.** 2001. Pyrosequencing sheds light on DNA sequencing. Genome Res **11:**3-11.

218. **Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman.** 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol **66:**2541-7.

219. **Roose-Amsaleg, C. L., E. Garnier-Sillam, and M. Harry.** 2001. Extraction and purification of microbial DNA from soil and sediment samples. Applied Soil Ecology **18:**47-60.

220. **Rosenberg, E.** 1999. Microbial ecology and infectious disease. ASM Press, Washington, DC.

221. **Rotthauwe, J. H., K. P. Witzel, and W. Liesack.** 1997. The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. Appl Environ Microbiol **63:**4704-12.

222. **Rowe, A. R., B. J. Lazar, R. M. Morris, and R. E. Richardson.** 2008. Characterization of the community structure of a dechlorinating mixed culture and comparisons of gene expression in planktonic and biofloc-associated "Dehalococcoides" and Methanospirillum species. Appl Environ Microbiol **74:**6709-19.

223. **Ruehland, C., A. Blazejak, C. Lott, A. Loy, C. Erseus, and N. Dubilier.** 2008. Multiple bacterial symbionts in two species of co-occurring gutless oligochaete worms from Mediterranean sea grass sediments. Environ Microbiol **10:**3404-16.

224. **Ruiz, J., M. M. Navia, C. Casals, J. M. Sierra, M. T. Jimenez De Anta, and J. Vila.** 2003. Integron-mediated antibiotic multiresistance in Acinetobacter baumannii clinical isolates from Spain. Clin Microbiol Infect **9:**907-11.

225. **Salyers, A., and N. B. Shoemaker.** 2006. Reservoirs of antibiotic resistance genes. Animal Biotechnology **17:**137-46.

226. **Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, and C. R. Hutchinson.** 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge Discodermia dissoluta. Appl Environ Microbiol **71:**4840-9.

227. **Schluter, A., R. Szczepanowski, N. Kurz, S. Schneiker, I. Krahn, and A. Puhler.** 2007. Erythromycin resistance-conferring plasmid pRSB105, isolated from a sewage treatment plant, harbors a new macrolide resistance determinant, an integron-containing Tn402-like element, and a large region of unknown function. Appl Environ Microbiol **73:**1952-60.

228. **Schroder, J.** 1999. Comprehensive Natural Products in Chemistry.749-771.

229. **Schubert, S., S. Dufke, J. Sorsa, and J. Heesemann.** 2004. A novel integrative and conjugative element (ICE) of Escherichia coli: the putative progenitor of the Yersinia high-pathogenicity island. Mol Microbiol **51:**837-48.

230. **Sebat, J. L., F. S. Colwell, and R. L. Crawford.** 2003. Metagenomic profiling: microarray analysis of an environmental genomic library. Appl Environ Microbiol **69:**4927-34.

231. **Selenska, S., and W. Klingmuller.** 1991. Direct detection of nif-gene sequences of Enterobacter agglomerans in soil. FEMS Microbiol Lett **80:**243-246.

232. **Selvakumaran, S., A. Kapley, V. C. Kalia, and H. J. Purohit.** 2008. Phenotypic and phylogenic groups to evaluate the diversity of Citrobacter isolates from activated biomass of effluent treatment plants. Bioresour Technol **99:**1189-95.

233. **Seow, K. T., G. Meurer, M. Gerlitz, E. Wendt-Pienkowski, C. R. Hutchinson, and J. Davies.** 1997. A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. J Bacteriol **179:**7360-8.

234. **Seviour, R. J., T. Mino, and M. Onuki.** 2003. The microbiology of biological phosphorus removal in activated sludge systems. FEMS Microbiol Rev **27:**99-127.

235. **Shakil, S., R. Khan, R. Zarrilli, and A. U. Khan.** 2008. Aminoglycosides versus bacteria--a description of the action, resistance mechanism, and nosocomial battleground. J Biomed Sci **15:**5-14.

236. **Shaw, W. V., and A. G. W. Leslie.** 1991. Chloramphenicol acetyltransferase. Annu Rev Biophys Biophys Chem **20:**363-386.

237. **Shen, B.** 2003. Polyketide biosynthesis beyond the type I, II, and III polyketide synthase paradigms. Current Opinion in Chemical Biology **7:**285-295.

238. **Shizuya, H., B. Birren, U. J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon.** 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:**8794-7.

239. **Silakowski, B., H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blocker, G. Hofle, S. Beyer, and R. Muller.** 1999. New lessons for combinatorial biosynthesis from myxobacteria. The myxothiazol biosynthetic gene cluster of Stigmatella aurantiaca DW4/3-1. J Biol Chem **274:**37391-9.

240. **Simon, C., and R. Daniel.** 2009. Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol.

241. **Simon, C., J. Herath, S. Rockstroh, and R. Daniel.** 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. Appl Environ Microbiol **75:**2964-8.

242. **Simpson, J. M., P. G. Stroot, S. Gelman, I. Beydilli, S. Dudley, and D. B. Oerther.** 2006. 16S ribosomal RNA tools identify an unexpected predominance of Paenibacillus-like bacteria in an industrial activated sludge system suffering from poor biosolids separation. Water Environ Res **78:**864-71.

243. **Singh, V., V. Praveen, F. Khan, and C. K. Tripathi.** 2009. Phylogenetics of an antibiotic producing Streptomyces strain isolated from soil. Bioinformation **4:**53-8.

244. **Sleator, R. D., C. Shortall, and C. Hill.** 2008. Metagenomics. Lett Appl Microbiol **47:**361-6.

245. **Smalla, K., N. Cresswell, L. C. Mendoncahagler, A. Wolters, and J. D. Vanelsas.** 1993. Rapid DNA Extraction Protocol from Soil for Polymerase Chain Reaction-Mediated Amplification. Journal of Applied Bacteriology **74:**78-85.

246. **Smith, D. H.** 1967. R factor infection of Escherichia coli lyophilized in 1946. J Bacteriol **94:**2071-2.

247. **Snaidr, J., R. Amann, I. Huber, W. Ludwig, and K. H. Schleifer.** 1997. Phylogenetic analysis and in situ identification of bacteria in activated sludge. Appl Environ Microbiol **63:**2884-96.

248. **Snyder, L., and W. Champness.** 2007. Molecular genetics of bacteria, 3rd ed. ASM Press, Washington, D.C.

249. **Solbak, A. I., T. H. Richardson, R. T. McCann, K. A. Kline, F. Bartnek, G. Tomlinson, X. Tan, L. Parra-Gessert, G. J. Frey, M. Podar, P. Luginbuhl, K. A. Gray, E. J. Mathur, D. E. Robertson, M. J. Burk, G. P. Hazlewood, J. M. Short, and J. Kerovuo.** 2005. Discovery of pectin-degrading enzymes and directed evolution of a novel pectate lyase for processing cotton fabric. J Biol Chem **280:**9431-8.

250. **Song, J. S., J. H. Jeon, J. H. Lee, S. H. Jeong, B. C. Jeong, S. J. Kim, J. H. Lee, and S. H. Lee.** 2005. Molecular characterization of TEM-type beta-lactamases identified in cold-seep sediments of Edison Seamount (south of Lihir Island, Papua New Guinea). J Microbiol **43:**172-8.

251. **Stackebrandt, E., A. Fruhling, S. Cousin, E. Brambilla, H. Lunsdorf, and S. Verbarg.** 2008. Methylibium subsaxonicum spec. nov., a betaproteobacterium Isolated from a hardwater rivulet. Curr Microbiol **56:**298-305.

252. **Staley, J. T., and A. Konopka.** 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annual Review of Microbiology **39:**321-346.

253. **Staunton, J., and K. J. Weissman.** 2001. Polyketide biosynthesis: a millennium review. Nat Prod Rep **18:**380-416.

254. **Steffan, R. J., J. Goksoyr, A. K. Bej, and R. M. Atlas.** 1988. Recovery of DNA from soils and sediments. Appl Environ Microbiol **54:**2908-15.

255. **Stein, J. L., T. L. Marsh, K. Y. Wu, H. Shizuya, and E. F. DeLong.** 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol **178:**591-9.

256. **Suenaga, H., T. Ohnuki, and K. Miyazaki.** 2007. Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. Environ Microbiol **9:**2289-97.

257. **Summers, A. O.** 2006. Genetic linkage and horizontal gene transfer, the roots of the antibiotic multi-resistance problem. Anim Biotechnol **17:**125-35.

258. **Summers, R. G., E. Wendt-Pienkowski, H. Motamedi, and C. R. Hutchinson.** 1992. Nucleotide sequence of the tcmII-tcmIV region of the tetracenomycin C biosynthetic gene cluster of Streptomyces glaucescens and evidence that the tcmN gene encodes a multifunctional cyclase-dehydratase-O-methyl transferase. J Bacteriol **174:**1810-20.

259. **Summers, R. G., E. Wendt-Pienkowski, H. Motamedi, and C. R. Hutchinson.** 1993. The tcmVI region of the tetracenomycin C biosynthetic gene cluster of Streptomyces glaucescens encodes the tetracenomycin F1 monooxygenase, tetracenomycin F2 cyclase, and, most likely, a second cyclase. J Bacteriol **175:**7571-80.

260. **Sylvia, D. M.** 2005. Principles and applications of soil microbiology, 2nd ed. Pearson Prentice Hall, Upper Saddle River, N.J.

261. **Syvanen, M., and C. I. Kado.** 2002. Horizontal gene transfer, 2nd ed. Academic Press, San Diego.

262. **Szczepanowski, R., I. Krahn, B. Linke, A. Goesmann, A. Puhler, and A. Schluter.** 2004. Antibiotic multiresistance plasmid pRSB101 isolated from a wastewater treatment plant is related to plasmids residing in phytopathogenic bacteria and carries eight different resistance determinants including a multidrug transport system. Microbiology **150:**3613-30.

263. **Szczepanowski, R., B. Linke, I. Krahn, K. H. Gartemann, T. Gutzkow, W. Eichler, A. Puhler, and A. Schluter.** 2009. Detection of 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to selected antibiotics. Microbiology **155:**2306-19.

264. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **24:**1596-9.

265. **Tauch, A., S. Krieft, J. Kalinowski, and A. Puhler.** 2000. The 51,409-bp R-plasmid pTP10 from the multiresistant clinical isolate Corynebacterium striatum M82B is composed of DNA segments initially identified in soil bacteria and in plant, animal, and human pathogens. Mol Gen Genet **263:**1-11.

266. **Tauch, A., A. Schluter, N. Bischoff, A. Goesmann, F. Meyer, and A. Puhler.** 2003. The 79,370-bp conjugative plasmid pB4 consists of an IncP-1beta backbone loaded with a chromate resistance transposon, the strA-strB streptomycin resistance gene pair, the oxacillinase gene bla(NPS-1), and a tripartite antibiotic efflux system of the resistance-nodulation-division family. Mol Genet Genomics **268:**570-84.

267. **Tebbe, C. C., and W. Vahjen.** 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl Environ Microbiol **59:**2657-65.

268. **Thomas, J. C., M. Desrosiers, Y. St-Pierre, P. Lirette, J. G. Bisaillon, R. Beaudet, and R. Villemur.** 1997. Quantitative flow cytometric detection of specific microorganisms in soil samples using rRNA targeted fluorescent probes and ethidium bromide. Cytometry **27:**224-32.

269. **Tien, C. C., C. C. Chao, and W. L. Chao.** 1999. Methods for DNA extraction from various soils: a comparison. Journal of Applied Microbiology **86:**937-943.

270. **Trieu-Cuot, P., G. de Cespedes, F. Bentorcha, F. Delbos, E. Gaspar, and T. Horaud.** 1993. Study of heterogeneity of chloramphenicol acetyltransferase (CAT) genes in streptococci and enterococci by polymerase chain reaction: characterization of a new CAT determinant. Antimicrob Agents Chemother **37:**2593-8.

271. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P.**

Hugenholtz, and E. M. Rubin. 2005. Comparative metagenomics of microbial communities. Science **308:**554-7.

272. **Tsai, Y. L., and B. H. Olson.** 1991. Rapid method for direct extraction of DNA from soil and sediments. Appl Environ Microbiol **57:**1070-4.

273. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:**37-43.

274. **Uchiyama, T., T. Abe, T. Ikemura, and K. Watanabe.** 2005. Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol **23:**88-93.

275. **Van Elsas, J. D., V. Mantynen, and A. C. Wolters.** 1997. Soil DNA extraction and assessment of the fate of Mycobacterium chlorophenicolum strain PCP-1 in different soils by 16S ribosomal RNA gene sequence based most-probable-number PCR and immunofluorescence. Biol Fertil Soils **24:**188-195.

276. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.

277. **Voget, S., C. Leggewie, A. Uesbeck, C. Raasch, K. E. Jaeger, and W. R. Streit.** 2003. Prospecting for novel biocatalysts in a soil metagenome. Appl Environ Microbiol **69:**6235-42.

278. **Volkmann, H., T. Schwartz, P. Bischoff, S. Kirchen, and U. Obst.** 2004. Detection of clinically relevant antibiotic-resistance genes in municipal wastewater using real-time PCR (TaqMan). J Microbiol Methods **56:**277-86.

279. **Wakil, S. J.** 1989. Fatty acid synthase, a proficient multifunctional enzyme. Biochemistry **28:**4523-30.

280. **Walsh, C.** 2003. Antibiotics : actions, origins, resistance. ASM Press, Washington, D.C.

281. **Walsh, T. R.** 2006. Combinatorial genetic evolution of multiresistance. Curr Opin Microbiol **9:**476-82.

282. **Wang, C., D. J. Meek, P. Panchal, N. Boruvka, F. S. Archibald, B. T. Driscoll, and T. C. Charles.** 2006. Isolation of poly-3-hydroxybutyrate metabolism genes from complex microbial communities by phenotypic complementation of bacterial mutants. Appl Environ Microbiol **72:**384-91.

283. **Wanibuchi, K., P. Zhang, T. Abe, H. Morita, T. Kohno, G. Chen, H. Noguchi, and I. Abe.** 2007. An acridone-producing novel multifunctional type III polyketide synthase from Huperzia serrata. FEBS J **274:**1073-82.

284. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature **345:**63-5.

285. **Ward, N. L., J. F. Challacombe, P. H. Janssen, B. Henrissat, P. M. Coutinho, M. Wu, G. Xie, D. H. Haft, M. Sait, J. Badger, R. D. Barabote, B. Bradley, T. S. Brettin, L. M. Brinkac, D. Bruce, T. Creasy, S. C. Daugherty, T. M. Davidsen, R. T. DeBoy, J. C. Detter, R. J. Dodson, A. S. Durkin, A. Ganapathy, M. Gwinn-Giglio, C. S. Han, H. Khouri, H. Kiss, S. P. Kothari, R. Madupu, K. E. Nelson, W. C. Nelson, I. Paulsen, K. Penn, Q. Ren, M. J. Rosovitz, J. D. Selengut, S. Shrivastava, S. A. Sullivan, R. Tapia, L. S. Thompson, K. L. Watkins, Q. Yang, C. Yu, N. Zafar, L. Zhou, and C. R. Kuske.** 2009. Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol **75:**2046-56.

286. **Wawrik, B., L. Kerkhof, G. J. Zylstra, and J. J. Kukor.** 2005. Identification of unique type II polyketide synthase genes in soil. Appl Environ Microbiol **71:**2232-8.

287. **Webster, G., C. J. Newbury, J. C. Fry, and A. J. Weightman.** 2003. Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. J Microbiol Methods **55:**155-164.

288. **Wegley, L., R. Edwards, B. Rodriguez-Brito, H. Liu, and F. Rohwer.** 2007. Metagenomic analysis of the microbial community associated with the coral Porites astreoides. Environ Microbiol **9:**2707-19.

289. **Weller, R., F. O. Glockner, and R. Amann.** 2000. 16S rRNA-targeted oligonucleotide probes for the in situ detection of members of the phylum Cytophaga-Flavobacterium-Bacteroides. Syst Appl Microbiol **23:**107-14.

290. **White, D. G., C. Hudson, J. J. Maurer, S. Ayers, S. Zhao, M. D. Lee, L. Bolton, T. Foley, and J. Sherwood.** 2000. Characterization of chloramphenicol and florfenicol resistance in Escherichia coli associated with bovine diarrhea. J Clin Microbiol **38:**4593-8.

291. **Wild, J., Z. Hradecna, and W. Szybalski.** 2002. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res **12:**1434-44.

292. **Williamson, L. L., B. R. Borlee, P. D. Schloss, C. Guan, H. K. Allen, and J. Handelsman.** 2005. Intracellular screen to identify metagenomic clones that induce or inhibit a quorum-sensing biosensor. Appl Environ Microbiol **71:**6335-44.

293. **Withey, S., E. Cartmell, L. M. Avery, and T. Stephenson.** 2005. Bacteriophages--potential for application in wastewater treatment processes. Sci Total Environ **339:**1-18.

294. **Wommack, K. E., S. R. Bench, J. Bhavsar, D. Mead, and T. Hanson.** 2009. Isolation independent methods of characterizing phage communities 2: characterizing a metagenome. Methods Mol Biol **502:**279-89.

295. **Wright, G. D.** 1999. Aminoglycoside-modifying enzymes. Curr Opin Microbiol **2:**499-503.

296. **Wright, G. D.** 2007. The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol **5:**175-86.

297. **Xia, X., J. Bollinger, and A. Ogram.** 1995. Molecular genetic analysis of the response of three soil microbial communities to the application of 2,4-D. Molecular Ecology **4:**17-28.

298. **Xia, Z. J., J. Wang, W. Hu, H. Liu, X. Z. Gao, Z. H. Wu, P. Y. Zhang, and Y. Z. Li.** 2008. Improving conjugation efficacy of Sorangium cellulosum by the addition of dual selection antibiotics. J Ind Microbiol Biotechnol **35:**1157-63.

299. **Xu, M., F. Wang, J. Meng, and X. Xiao.** 2007. Construction and preliminary analysis of a metagenomic library from a deep-sea sediment of east Pacific Nodule Province. FEMS Microbiol Ecol **62:**233-41.

300. **Xue, Y., L. Zhao, H. W. Liu, and D. H. Sherman.** 1998. A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity. Proc Natl Acad Sci U S A **95:**12111-6.

301. **Yang, Y., and J. Zeyer.** 2003. Specific detection of Dehalococcoides species by fluorescence in situ hybridization with 16S rRNA-targeted oligonucleotide probes. Appl Environ Microbiol **69:**2879-83.

302. **Young, C. C., R. L. Burghoff, L. G. Keim, V. Minak-Bernero, J. R. Lute, and S. M. Hinton.** 1993. Polyvinylpyrrolidone-Agarose Gel Electrophoresis Purification of Polymerase Chain Reaction-Amplifiable DNA from Soils. Appl Environ Microbiol **59:**1972-1974.

303. **Zhen, L., and R. T. Swank.** 1993. A simple and high yield method for recovering DNA from agarose gels. Biotechniques **14:**894-8.

304. **Zhou, J., M. A. Bruns, and J. M. Tiedje.** 1996. DNA recovery from soils of diverse composition. Appl Environ Microbiol **62:**316-22.

305. **Zinder, N. D., and J. Lederberg.** 1952. Genetic exchange in Salmonella. J Bacteriol **64:**679-99.