

Full Length Research Paper

Isolation of DNA from saltern soils collected in Taiwan and whole-genome amplification of minute amounts of DNA for construction of metagenomic libraries

Shie-Jea Lin, Hsuen-Chun Liao, Tzu-Jun Liu and Wen-Shen Chu*

Food Industry Research and Development Institute, Hsinchu 30062, Taiwan, Republic of China.

Accepted 30 November, 2012

Environmental DNAs from 21 samples of saltern soil in Taiwan were isolated by using the SDS-lysis method, resulting in yields ranging from 0.03 ng to 8.06 µg per gram of soil. However, sample 143 collected from saltern soil near a crystallizer had a low yield of 1.2 ng per gram of soil. Comparative analyses of the sequence data of representative clones with other 16S rRNA samples indicated that not all clones for sample 143 were closely related to the soil bacteria. A minute amount of DNA (0.15 ng) was amplified 100,000-times to 15 µg by multiple displacement amplification (MDA). The MDA method was validated by analysis of amplified bacteriorhodopsin (bR) genes. Two clone libraries were constructed from DNA samples before and after amplification and were compared. The result suggests that bR diversity was relatively conserved during whole-genome amplification (WGA). The constructed metagenome fosmid library consists of 1.7×10^6 clones with an average insert size of 26.1 kb. Taken together, WGA of metagenomic DNA from very minute microbial sources allows for construction of metagenomic libraries that are previously inaccessible.

Key words: Saltern soil, DNA extraction, metagenomic DNA, multiple displacement amplification, fosmid library construction.

INTRODUCTION

Metagenomics, the study of genetic material recovered directly from environmental samples, is a new and rapidly developing field. Metagenomic techniques have been used to analyze the complex genomes contained within microbial communities (Kowalchuk et al., 2007; Schmeisser et al., 2007) and are based on the direct isolation of DNA from environmental samples from which metagenomic libraries are generated. When clones that contain phylogenetic genes, such as the 16S rRNA gene, are retrieved, the DNA sequence information surrounding these genes provides access to the genomes of unculturable microorganisms and can provide clues to the physiology of such microorganisms (Hallam et al., 2006a). Furthermore,

function-based screening of the libraries has led to identification and characterization of a variety of novel metabolites and biocatalysts, such as lipases, amylases, nitrilases, and oxidoreductases (Handelsman et al., 2004; Ferrer et al., 2009).

Isolation of metagenomic DNA is difficult because co-extracted polyphenolic substances found in soil interferes with downstream applications (Tsai and Olson, 1992) and a few studies have attempted to quantify the efficiency of various DNA extraction protocols using environmental samples (Frostegard et al., 1999; Bertrand et al., 2005). Isolation of high molecular weight (HMW DNA) is important to reduce the risk of chimera formation during PCR

amplification (Liesack and Stackebrandt, 1992) and to allow for the construction of large-insert metagenomic libraries. Besides, DNA is simply found at low levels in many soil samples (Webster et al., 2003). However, accurate analysis of minute amounts of DNA has been a challenge for geneticists. The limitations in obtaining sufficient specimens and the difficulties in extracting high-quality DNA from environmental samples have impeded the understanding of microbial community structures (Yokouchi et al., 2006). Therefore, techniques to obtain information from small amounts of DNA are necessary. Whole genome amplification (WGA) is an increasingly common technique that can potentially be used to eliminate DNA yield as a limiting factor for genetic assays. For example, only small amounts of DNA can be extracted from low-biomass soils. Relatively low amounts of DNA (0.120-2.8 ng/g) have been retrieved from deeply buried marine sediments from the ODP Site 1229 on the Peru Margin (Webster et al., 2003).

Four primary forms of WGA have been previously described and are commonly used: multiple displacement amplification (MDA) (Dean et al., 2002; Ling et al., 2009), primer extension preamplification (PEP) (Zhang et al., 1992), degenerate oligonucleotide-primed PCR (DOP-PCR) (Telenius et al., 1992; Devries et al., 2005), and linker adapter PCR (LAP) (Klein et al., 1999; Pirker et al., 2004). These WGA methods have been compared, but the comparisons have been limited in scale. The amplification process should be highly accurate to limit the introduction of errors. Amplification should not induce a bias in the distribution of the product DNA. Large amplification factors are required, so that WGA can generate a useful amount of DNA from small starting samples. Additionally, multiple displacement amplification (MDA) was used to amplify whole-genome DNA from single bacterial cells and was highly efficient (Rodrigue et al., 2009).

One general approach of metagenomics begins with the preparation of a library of clones that contain large inserts that were obtained from microbial communities. Fosmid libraries containing inserts comprised fragments of environmental genomes approximately 35 kb in size have been constructed for many microbial communities (Grzymalski et al., 2006; Hallam et al., 2006b). Therefore, new methods are required to combine environmental WGA with library construction for metagenomic analyses of low-cell-density environments.

The aim of this study was to construct a metagenomic library with whole-genome amplification of minute amounts DNA and using molecular techniques to examine the presence and diversity of bacteriorhodopsin (bR) genes from trace amounts of metagenomic DNA acquired from environmental samples containing culturable and unculturable bacteria within a saltern soil region of the Chiku site of Tanan, Taiwan. Focus was placed on comparing the phylogeny of the bR genes before and after amplification. The results of this study indicate that WGA of minute amounts of DNA can be used for the construction

of metagenomic libraries for further study and application.

MATERIALS AND METHODS

Materials

Decylmethylammonium bromide was purchased from Acros Organics (Fisher Scientific, UK). Miracloth was obtained from Merck & Co. Polyethylene glycol, polyvinylpyrrolidone (PVPP), and SYBR[®] Green I were brought from Sigma (St. Louis, MO). Bio-spin disposable chromatography columns and random hexamers were acquired from Amersham Pharmacia Biotech. Pulse-field agarose gel BioRad electroelutor and DCODE gel electrophoresis systems were purchased from BioRad. The EPI 300[™] *E. coli* strain was acquired from Epicentre (Madison, WI). The GELase[™] Agarose Gel-Digesting Preparation kit, RepliPHI[™] Phi29 DNA Polymerase, and the CopyControl[™] fosmid library production kit were obtained from Epicentre. The plasmid miniprep purification kit was purchased from Genemark. TOPO TA Cloning kits were purchased from Invitrogen (Carlsbad, CA).

Soil sample collection and preparation

Environmental samples of saltern soil were collected from southern Taiwan, including Budai, Chiayi, Yung-An, Kaohsiung, Chiku and Tainan. The soil samples from three different sites at each location were combined and used for further experimentation. The samples were processed immediately, stored at -20°C, and simultaneously subjected to metagenomic DNA isolation. The soil samples were sieved to remove plant debris and particulates larger than 2 mm. Total DNA was isolated from the soil samples by the method described by Zhou et al. (1996) with modifications. Isolation consisted of sodium dodecyl sulfate and proteinase K treatment, which was designed to isolate total DNA from a variety of soil types. Soil samples were resuspended in Solution A [100 mM Tris-HCl, 100 mM Na₂EDTA, 100 mM sodium potassium, 1.5 M NaCl, and 1% hexadecyltrimethyl ammonium bromide (CTAB), pH 7.0] at 500 mg/ml. The solutions were mixed at 37°C for 30 min at a speed of 150 rpm. SDS was added to a final concentration of 2% and the mixture was further incubated for 2 hr and agitated by inversion every 15 min. After incubation and centrifugation, 5 M potassium acetate (pH 5.5) was added to a final concentration of 0.5 M and the mixture was gently agitated by inversion. Following a 20-min incubation on ice, the samples were centrifuged, and 0.6 volumes of isopropanol was added and samples were incubated at room temperature for 1 h (Zhou et al., 1996). After centrifugation, the supernatant fluid was removed and the precipitated DNA was washed with 70% ethanol. Afterwards, total DNA was resuspended in 50 µl of 10 mM Tris-HCl (pH 8.0) and purified by agarose gel electrophoresis or by four rounds of ultrafiltration in a Microcon-100 microconcentrator (Amicon) for PCR amplification. After electrophoresis, DNA-containing regions were cut from the unstained gel and stored overnight in 0.5 × TE buffer (Tris-EDTA). Further purification of the DNA for cloning into a fosmid vector was performed following the methods of Rondon et al. (2000).

PCR amplification and sequencing of partial 16S rDNA

Universal bacterial 16S rDNA was amplified with gene-specific primers [5'-GAGTTTGATCCTGGCTCAG-3' (sense) and 5'-AGAAAGGAGGTGATCCAGCC-3' (antisense) (Brambilla et al., 2001)]. PCR amplification was performed in 20 µl reaction mixtures containing 200 µM of each dNTP, 40 µM of each primer, 100 ng of template DNA, 1.5 mM MgCl₂, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, and 1 U AmpliTaq Gold[™] (Perkin-Elmer). The PCR reactions were amplified for 35 cycles of 95°C for 1 min, 49°C for 1 min, and

72°C for 2 min. The PCR products were separated on agarose gels, purified with a QIAquick gel extraction kit (QIAGEN). The resulting products were cloned into TOPO TA vectors (Invitrogen) according to the manufacturer's directions. Sequence information from cloned 16S rRNA genes was obtained using an ABI PRISM BigDye cycle sequencing system and was analyzed with an ABI model 3700 automated sequencer. The resulting sequence was compared with the non-redundant sequence database at the National Center for Biotechnology Information (NCBI) using BLAST.

Amplification of DNA isolated from saltern soil

Metagenomic DNA was amplified by MDA using a RepliPHI™ Phi29 Reagent Set (Epicentre). Amplification was performed according to the manufacturer's protocol with slight modifications. Briefly, 10 µl (0.15 ng) saltern soil template DNA was placed in a 1-cm polyethylene tube for incubation in a DNA thermal cycler (model 2400; Perkin-Elmer) for 3 min. DNA was cooled slowly to room temperature over 30 min. 0.8 µl RepliPHI™ Phi29 DNA polymerase (1000 U/µl), 4 µl 25 mM dNTP, 10 µl 10X reaction buffer, 2 µl 100 mM dithiothreitol, 25 µl 200 µM random hexamer primers (5'-NNNNNN-3'), and 48.2 µl H₂O were added to a final volume of 100 µl. Reactions were incubated for 18 h at 30°C, followed by heat inactivation at 65°C for 3 min. Reaction products were electrophoresed through 1.0% agarose gels (in Tris-borate-EDTA buffer) and through a pulse-field agarose gel BioRad electroelutor (Bio-Rad CHEF-DR II, 1- to 6-sec switch, 6 V/cm, 120° fixed angle, 5-h run time).

PCR amplification, sequencing, and phylogenetic analyses of bacteriorhodopsin (bR) genes

Bacteriorhodopsin fragments were PCR amplified using *Taq* polymerase (Invitrogen) from metagenomic DNA extracted from sample 143 before and after WGA. The bR-specific primers were 5'-GACTGGYTGTTCACACRCC-3' (sense) and 5'-ASGTCRAKSACCATGAA-3' (antisense) (Papke et al., 2003). Products were amplified with the following program: 94°C for 3 min, 20 cycles of 94°C for 45 s, 58°C for 45 s, and 72°C for 45 s. To further amplify the DNA, samples were subjected to 20 additional cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 45 s. The resulting products were then cloned into TOPO TA Cloning vectors. Two different libraries using DNA from before and after WGA were constructed respectively (S1 and S2), then clones positive for inserts were sequenced as above. Totally, twenty-four cloned bR gene sequences were compared with reference sequences in the NCBI Nucleotide & Protein Sequence Database by using the tBLASTX program. Homology scores were calculated using Vector NTI Advance™ 10 software.

Construction of fosmid library

To fractionate the isolated DNA by size, approximately 100 µg of post-amplification metagenomic DNA was separated by preparative pulsed-field gel electrophoresis (Bio-Rad CHEFMapper; 0.5-sec switch time, 9 V/cm, 0.5X TBE, 120° included angle, 5 h). DNA fragments greater than 40 kb (40-148 kb) was recovered by electro elution and dialyzed against 1X TE buffer. A metagenomic library for the sample was constructed by use of a CopyControl fosmid library production kit (Epicentre) according to the manufacturer's instructions. Briefly, isolated DNA was digested with Gelase (Epicentre). Both ends of the size-fractionated DNA were end-repaired to generate blunt, 5'-phosphorylated ends and then ligated directly into cloning - ready CopyControl pCC1FOS vectors (Epicentre) with Fast-Link DNA ligase at 16°C overnight. After in vitro packing into

lambda phage by use of the supplied lambda packaging extracts, the DNA fragments were transformed into an EPI300-T1R phage T1-resistant *E. coli* host. The transformed cells were plated on LB medium (containing 12.5 µg/ml chloramphenicol). White colonies were transferred onto plates gridded to be compatible with 384-well microtiter plates containing 50 µl of LB medium and 7% glycerol (v/v). The plates were incubated at 37°C for 24 h. The library was replicated into duplicate sets of 384-well microtiter plates with freezing medium and stored at -80°C.

Nucleotide sequence accession numbers

The partial 16S rRNA sequences from this study have been deposited in the GenBank nonredundant database and have accession numbers EF429664 through EF431845. All partial bacteriorhodopsin sequences from this study have been deposited in the GenBank with accession numbers HM475109 through HM475131.

RESULTS

Extraction and purification of metagenomic DNA from saltern soil

Table 1 includes the yields of 21 samples of metagenomic DNA extracted from saltern soil by the SDS-Lysis method (Zhou et al., 1996). With the exception of samples 140 and 143 with very low yields, the yields were between 0.05 to 8.06 µg metagenomic DNA per gram of sample (Table 1). The yields varied between the various ranges of general soils, but compared with soils from organic farms (5-10 µg per gram of sample; unpublished data) the yields from saltern soil were lower. Samples 140 and 143 were collected from a crystallizer and the soils near a crystallizer, respectively and contained the highest ratios of salt crystals (Figure 1). The metagenomic DNA yields from these samples were 0.03 and 1.2 ng/g, respectively. These yields were less than 1/1000 of the yields from general soil and were chosen to be the bioresources for minute amounts DNA used in these experiments.

Efficiency of preparation and 16S rRNA diversity of metagenomic DNA from sample 143

Surveying the 16S rRNA genes in soil formed a more complete census of soil bacteria, without the limitations inherent in cultivation-based studies. 16S rRNA genes from members of the domain Bacteria were isolated from saltern soil sample 143 and analyzed to gain an understanding of the general composition of saltern soil bacterial communities (Figure 2). Total DNA from soil sample 143 was isolated and PCR products of approximately 400-500 bp were successfully obtained. Sequences from the library that were less than 300 nucleotides were excluded, as phylogenetic assignment from very short sequences can be unreliable (Ludwig and Klenk, 2001).

The available library of 16S rRNA and 16S rRNA genes permitted an initial survey of the saltern soil bacterial community structure. From the mixture, a library based on more than 23 distinct 16S rRNA genes were cloned and sequenced. The sequences from the library were

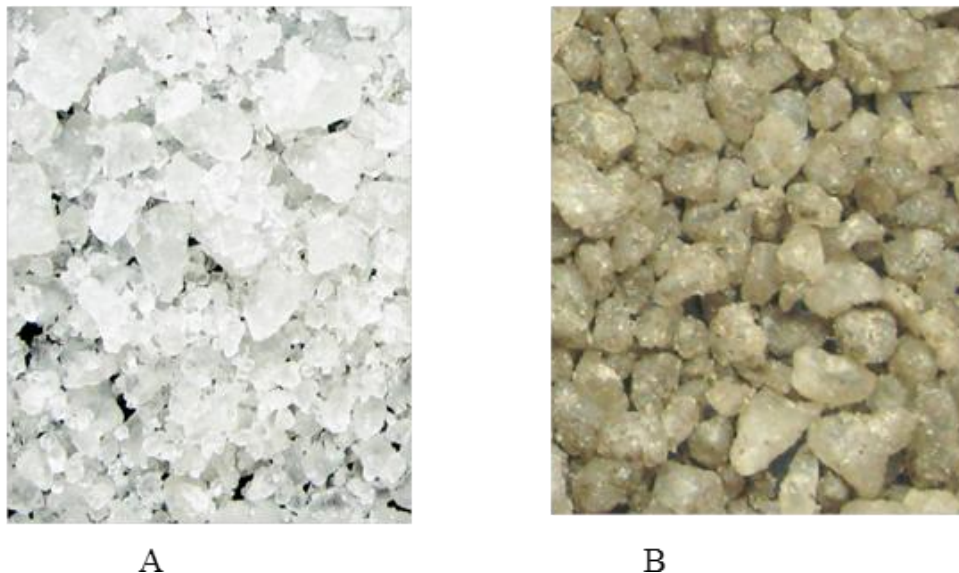


Figure 1. Photographs of samples 140 (A) and 143 (B) that were collected from crystallizer and saltern soils in Chiku, Tainan Taiwan.

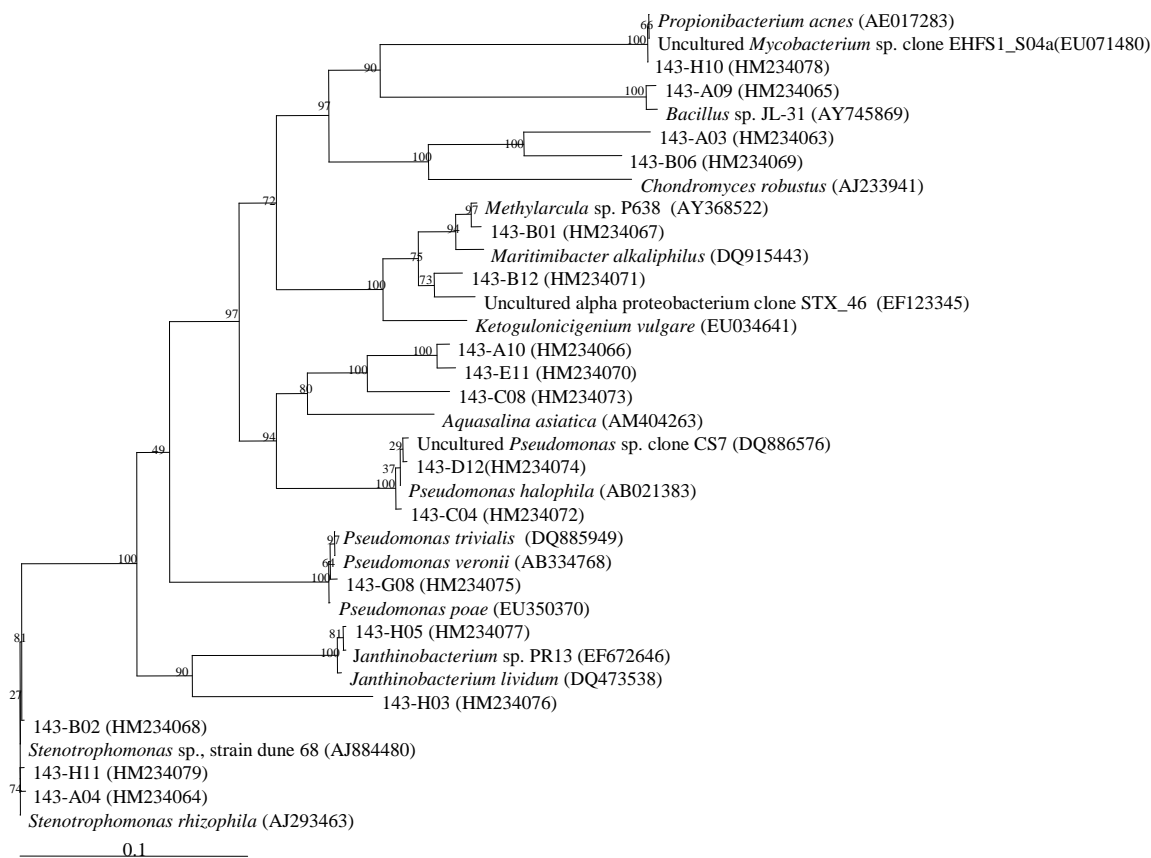


Figure 2. Phylogenetic relationships of 16S rRNA sequences from the clone library sorted by universal bacterial 16S rDNA primers and using 143 environmental DNA samples as templates. Twenty-three sequences from the represented phyla were used to construct the tree by using the neighbor-joining method with bootstrap values calculated from 1000 trees. The number at each branch point represents the bootstrap support percentage. Bar, 0.1 sequence divergence. GenBank accession numbers of nucleotide sequences are listed in parentheses.

assigned to genus-level groupings and then weighted for multiple clone assignments to one sequence type. This pooled set of clones was treated as one saltern soil set. The contribution of phylum-level groupings to soil bacterial communities was calculated from 17 clones. Altogether, the 16S rRNA sequences represented a broad spectrum of sequences, some of which fall readily within known bacterial families isolated from all over the world. However, the majority of the sequences represented unidentified bacterial families, presumably from still-uncultivated and otherwise undescribed bacterial populations. Some of the better-characterized dominant phyla identified are *Pseudomonas*, *Bacillus*, *Proteobacterium*, *Acidovorax*, *Chondromyces* (Myxobacteriales), *Propionibacterium*, *Stenotrophomonas*, and *Janthinobacterium*.

Of the 23 sequences, *Pseudomonas* spp. was the most abundant in soil bacterial communities, contributing 13.0% of the cloned sequences from the library. It is interesting that clone H03 shared the same sequences with the nitroaromatic compound-degrader *Acidovorax* sp. JS42 strain and clone H05 shared the same sequences with the *Janthinobacterium* sp. PR 13 strain. Additionally, *Propionibacterium acnes* that can cause a number of infections, including the common skin disease acne vulgaris, appeared as clone H10.

Among the 41 PCR-amplified 16S rRNA genes, 18 sequences from the library were highly similar in that they differed by only one or two bases within the 400-500 bp sequences (data not shown). These results suggest that there was a surprisingly low diversity of bacteria represented in the library. It is important to consider that libraries of PCR-amplified 16S rRNA and 16S rRNA genes may not represent a complete or accurate picture of the bacterial community.

Amplification of minute metagenomic DNA extracted from saltern soil

In this study, we used 0.15 ng of DNA extracted from saltern soil sample 143 as a template for amplification. Approximately 15 μ g of amplified DNA was acquired; a 10,000-fold amplification in only a few hours. The sample was analyzed by agarose electrophoresis, and the pattern of the DNA sample was the same as λ DNA (48.5 kb). In addition, the DNA sample was further analyzed by electrophoresis on a 1% pulse-field agarose gel. The molecular weights of the amplified DNAs ranged from 20 to 97 kb.

Phylogenetic analyses of bR genes isolated from saltern soil DNA before and after WGA

To better understand the extent of naturally occurring bR variability associated with bR-carrying genes in metagenomic DNA before and after WGA, we compared bR

genes from insert libraries (with 400-bp inserts) with DNA from sample 143 using degenerate bR primers. The PCR products from the two different sources before and after WGA were single bands in 1% agarose gels and were the expected size (400 bp) (data not shown). Even though the samples from the amplified DNA gave visible amplification products, the intensity of the specific bands was clearly higher in the amplified DNA than in the original DNA samples. This suggests that bR genes may have been distributed in each sample.

Four batches of the PCR products before and after amplification were subcloned and clones positive for inserts were selected. There were additional PCR products retrieved from the two libraries that were not bR. Twelve clones from the original DNA (S1 library) and 12 clones from the amplified DNA (S2 library) were found to contain bRs. To assess the overall diversity among the 24 incomplete bR sequences, a separate phylogeny was constructed. bR sequences were retrieved from the NCBI database and were phylogenetically compared with our clone sequences (Figure 3). In a phylogenetic tree, we identified bR sequences related to Bacteriorhodopsin (uncultured *Halobacteriales archaeon*), Rhodopsin (*Halobacterium halobium*), Archaerhodopsin (*Halorubrum xinjiangense*), and Cruxrhodopsin (*Haloarcula japonica*). Proteorhodopsin was not retrieved from the blast searches. Importantly, a similar pattern was observed for bR sequences from the two different libraries using DNA from before and after WGA. This suggests that bR diversity is relatively conserved during WGA.

The homology of bR nucleic acid sequences between the two clone libraries was also examined (Table 2). The homology scores of bR sequences from each clone were between 56 and 100% in the S1 library. For example, the bR sequence of clone a04 was the same as clone a01 and clone d07 had only two different nucleotides than clone a12. The average score was 67.7%. The homology scores of the S2 library bR sequences were 58-94%, similar to the S1 library. Clone g01 was 94% homologous with clones f02 and g05. The average score was 72.3%, only 4.6% greater than the S1 library. We also compared the partial bR sequences among the 24 clones, and found that clone g07 was 99% homologous with clone a03, and clone g02 was 98% homologous with clone d12. The average score between 144 comparisons was 69.3%. They might have been amplified because of the high bR gene sequence diversity within the samples.

We compared the amino acid sequences of the partial bR sequences of the 24 clones, and found that the average score was 64.9%, about 4.4% lower than the nucleic acid sequences (data not shown). The phylogenetic identity of a BLAST identification provided an initial analysis of the metagenome. Based on the BLASTN comparison, the library clone sequences retrieved before and after WGA showed that the two libraries had bR genes similar (66-99%) to those found in the NCBI database (data not shown).

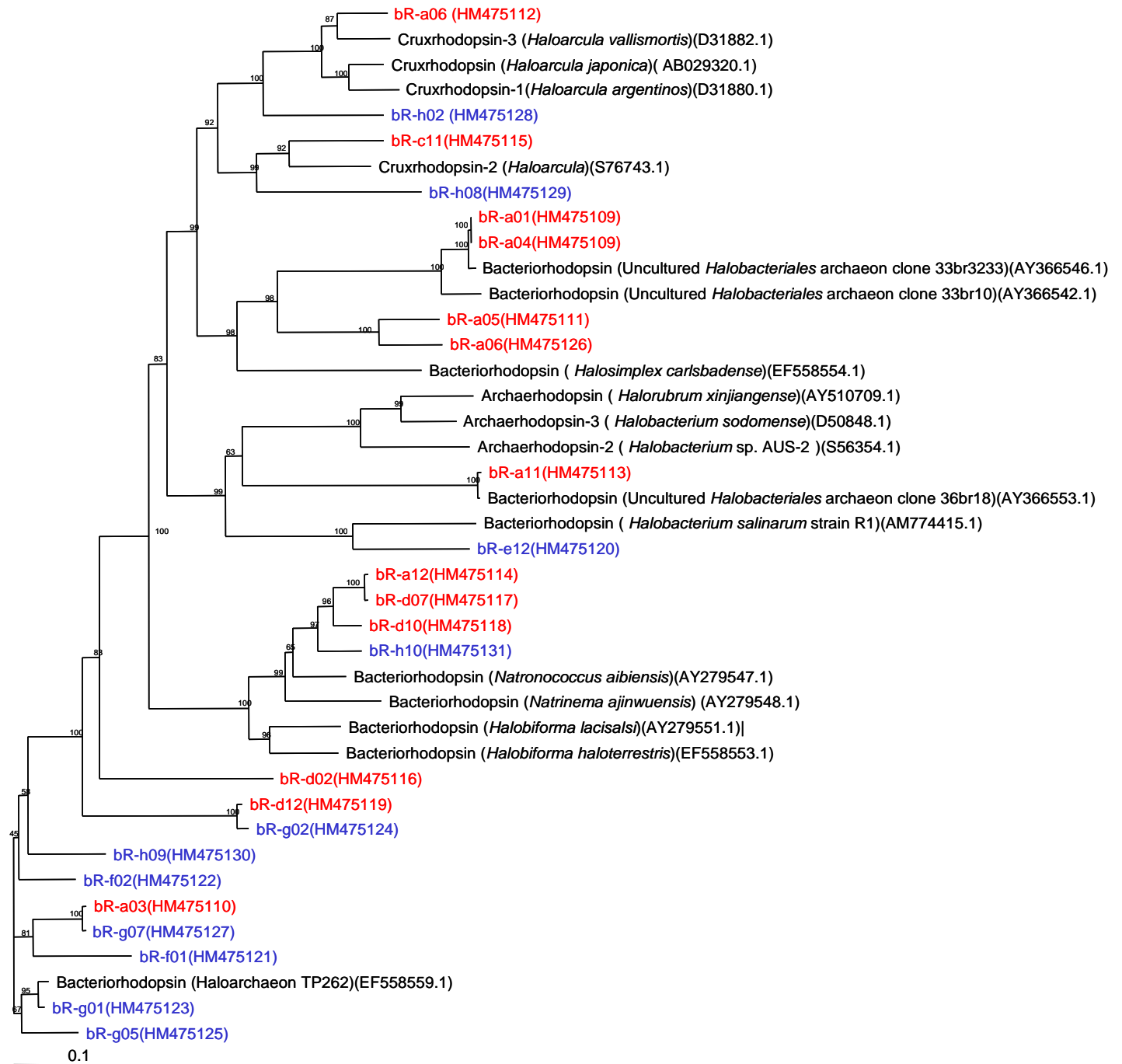


Figure 3. Phylogenetic relationships of partial bacteriorhodopsin sequences from two clone libraries (S1 and S2) sorted by the degenerate bR-specific primers. Twenty-four sequences from the represented phyla were used to construct the tree using the neighbor-joining method with bootstrap values calculated from 1000 trees. The number at each branch point represents the percentage bootstrap support. Bar, 0.1 sequence divergence. GenBank accession numbers of nucleotide sequences are listed in parentheses.

Construction of fosmid library from amplified DNA

To access genomic information from minute samples of soil microbes, including microbes that are not readily cultured, we developed methods to extract and amplify minute quantities of DNA. We then constructed a metagenomic

fosmid library containing clones of DNA fragments from the metagenomic DNA of saltern soil samples. Since the quantity of the metagenomic DNA isolated from saltern soil sample 143 was so small, its library construction was less efficient than that of DNA extracted from normal soil. However, prior to amplification, the quality of isolated

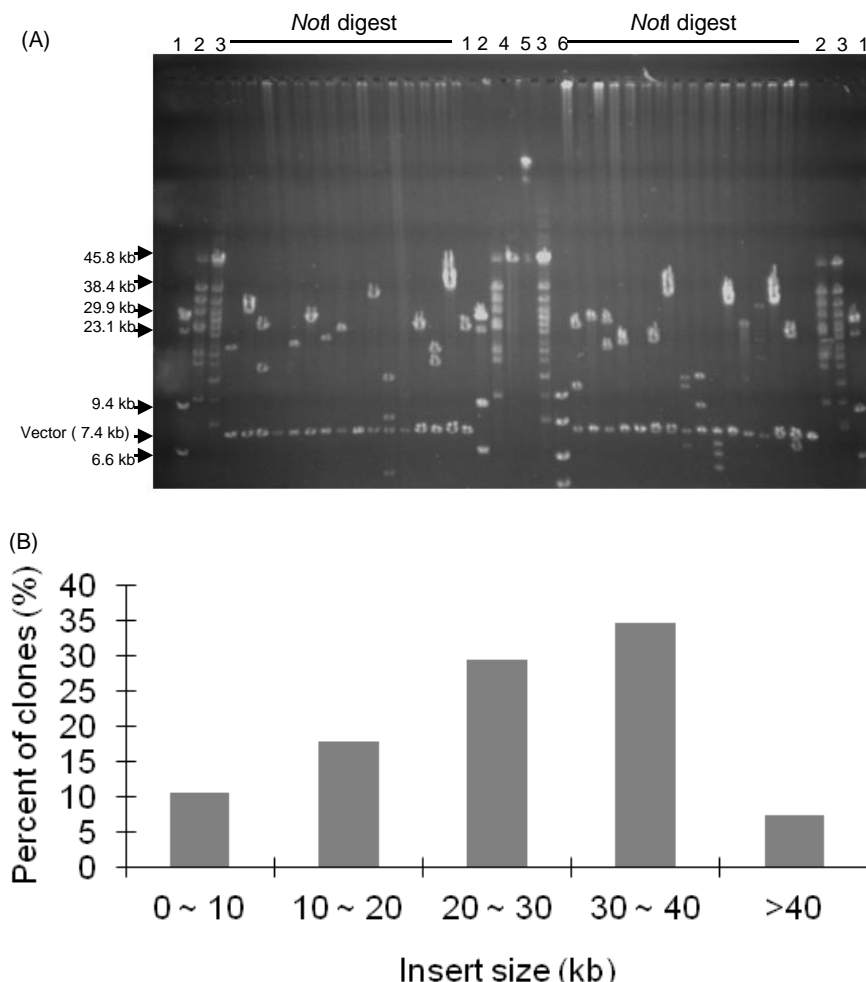


Figure 4. (A) Restriction digests of clones from the metagenomic library digested with *NotI*. The fosmid vector is indicated at 7.4 kb. Reference ladders: 1. λ DNA/*Hind* III, 2. λ DNA/Mono cut mix, 3. 8-48 kb, 4. λ DNA, 5. λ PFG marker, and 6. 1-kb ladder. (B) Representation of insert sizes in the soil metagenome libraries. Results were based on the analysis of 95 clones (2.5% of total).

DNA used to construct the metagenomic library was very similar to our previous study of farm soil (data not shown).

In general, enzymatic manipulation of the DNA for library construction was difficult due to the presence of humic substances, which were not completely removed through the purification steps. After amplification, the metagenomic DNA contained fewer humic substances, therefore making it simpler to efficiently construct the metagenomic library. Amplification of the metagenomic DNA from sample 143 increased the yield 10,000-times (15 $\mu\text{g/g}$ soil) from what was extracted from saltern soil (0.15 ng/g soil). After amplification, the metagenomic DNA (49-97 kb) was recovered from 1% pulse-field agarose gels and ligated into the Fosmid vector pCC1FOSTM. A library consisting of several thousand clones was generated by packaging and transfecting the DNA into the *E. coli* host strain. This is a prototype metagenomic library and it consists of 1.7×10^6 clones and 3744 clones

arrayed in thirty-nine 96-well microtiter plates. We examined approximately 2.5% ($n = 95$) of the clones for inserts, of which 97% contained DNA inserts with an average insert size of 26.1 kb (Figure 4). We estimated that there is approximately 98 Mbp of DNA sequence contained in the library. The libraries constructed from the saltern soil metagenomic DNA are maintained for further study and functional screening.

DISCUSSION

Many DNA extraction methods have been used to isolate DNA from soil (Bruce et al., 1992; Zhou et al., 1996; Yeates et al., 1998; Bertrand et al., 2005; Desai and Madamwar, 2007). Physical disruption methods, such as bead-beating and sonication, produce considerable DNA yields, but often cause severe DNA shearing (Leff et al., 1995; Yeates et al., 1998), which is not suitable for large-insert metagenomic library construction. In this study, we

modified the DNA extraction method described by Zhou et al. (1996) in which enzymatic and chemical lysis methods are used instead of bead-beating. We found all of metagenomic DNA from 21 samples, like sample 143 that contain extremely low cell counts, such as deeply buried marine sediments from ODP Site 1229 (0.120-2.8 ng/g), are usually inaccessible for environmental sequencing (Webster et al., 2003). WGA is an efficient approach for amplifying the small amounts of DNA extracted from microbes in low-biomass samples found in nitrate- and heavy-metal-contaminated soils. These samples typically yield small DNA quantities that have limited use for direct, native analysis and screening (Abulencia et al., 2006).

In the present study, species across several major phyla, including *Pseudomonas*, *Bacillus*, *Proteobacteria*, *Acidovorax*, *Chondromyces*, *Propionibacterium*, *Stenotrophomonas*, and *Janthinobacterium*, were found in sample 143 based on its 16S rDNA library. Comparative analyses of the sequence data of representative clones with other 16S rDNA samples indicated that not all clones were closely related to the soil bacteria. For this study, our goal was not to identify every individual microorganism found in saltern soil. Rather, our goal was to compare the phylogenetic diversity of the saltern soil microbes within sample 143. Besides *Pseudomonas* and *Bacillus*, the identified genera were different than the nine genera that were significant in soils identified by Martin Alexander: *Agrobacterium*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Flavobacterium*, *Micromonospora*, *Nocardia*, *Pseudomonas*, and *Streptomyces* (Alexander, 1977). Moreover, through the comparison of soil bacterial communities, Janssen (2006) showed that members of six genera (*Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Verrucomicrobia*) were significant in soil (Janssen, 2006). As described previously, clone library analyses based on PCR-amplified 16S rDNAs could contain biases in terms of quantitative analysis of clone distribution (Fuhrman, 2002). Such biases could particularly occur during the DNA extraction and PCR amplification steps due to different DNA extraction efficiencies for different cell types and variable primer specificities during amplification. The low diversity obtained in this study could be attributed to biases introduced during each of the multiple steps involved in this molecular approach.

Recovery of metagenomic DNA that is suitable in both quality and quantity for PCR and metagenomic library construction remains a challenge (Daniel, 2005). The current minimum amount of DNA needed to construct a library for shotgun sequencing is around 0.5 to 4 µg of DNA, which can be obtained from a minimum of 0.5 g of microbe-rich material. In the case of sediment with cell densities as low as 10^4 cells/g, 11- 88 kg of sample would be required (Tringe et al., 2005). According to recent reports, to construct fosmid libraries for further studies, such as functional gene investigation, more than 1 µg of metagenomic DNA is necessary (Couto et al., 2010; Yung

et al., 2009). In this case, 1 kg of sample 143 would be required to construct a fosmid library.

To overcome the obstacles of low DNA yields, we used WGA to construct fosmid libraries. We amplified the metagenomic DNA of sample 143 from 0.15 ng to 15 µg. Isolation of HMW DNA is important to reduce the risk of chimera formation during PCR amplification (Liesack and Stackebrandt, 1992) and to allow for the construction of large-insert metagenomic libraries and improve the possibilities of retaining the gene clusters that involve biosynthetic pathways (Bertrand et al., 2005). Construction of large-insert metagenomic libraries is currently used as a genomic approach to study the physiology of unculturable microorganisms (Liles et al., 2003).

In situ diversity studies using the 16S rRNA gene as a genetic marker are limited, as the gene is too highly conserved to provide useful genetic information and does not provide direct relevant ecological/physiological information. However, bR-based genes are diverse in saltern soils and *in situ* analysis of 16S rDNA identified diversity parallel to that identified by bR-gene analysis (Papke et al., 2003). In this study, bR was used as another gene marker for the identification of bacteria and archaea present in sample 143. The MDA method was validated by analysis of amplified bR genes. Our limited sequencing analysis of the minute metagenomic DNA obtained before and after amplification was not intended to define the soil metagenome, but rather to gain insight into the possibilities provided by MDA using Phi29 DNA polymerase. We identified bR sequences related to Bacteriorhodopsin (uncultured *Halobacteriales archaeon*), Rhodopsin (*Halobacterium halobium*), Archaeorhodopsin (*Halorubrum xinjiangense*), and Cruxrhodopsin (*Haloarcula japonica*) in sample 143 both before and after DNA amplification. We also compared the blast scores of bR sequences based on the homology of nucleic acid sequences between the two clone libraries, and within each individual clone library. A similar pattern was observed for bR sequences from the two different libraries using DNA before and after WGA, suggesting that bR diversity from the two different libraries was relatively consistent.

We examined the characteristics and quality of the constructed fosmid library by using amplified sample 143 DNA. Molecular characterization of the library showed that most of the inserts were between 30-40 kb with an average of 26 kb (Figure 4). We performed size-selected DNA insertion (49-97 kb) prior to metagenomic library construction (data not shown). This was a satisfactory insert size since the average DNA insert of metagenomic fosmid library clones from other studies were about 35 kb (Lee et al., 2006; Lim et al., 2005).

To the best of our knowledge, this study was the first to use amplified DNA to construct a large-insert metagenomic library. The MDA method was validated by analyzing amplified bR genes from DNA before and after amplification to construct a metagenomic library for environ-

mental analysis. This method was simple, rapid, and efficient at amplifying relatively low amounts of DNA that was suitable for metagenomic analysis and metagenomic library construction. The MDA method should be further validated by analyzing the quality of amplified genes, especially the gene clusters that are involved in biosynthetic pathways. Downstream analyses should involve functional screening for active clones or sequence-based screening using probes homologous to known genes.

Taken together, current results suggest that the metagenomic libraries can be readily screened for native genes or any target of interest. WGA of metagenomic DNA from very minute microbial sources, will allow access to genomic information that was not previously accessible. This method of accessing and exploiting natural biodiversity, together with high-throughput screening systems, will have a great impact on microbial biotechnology in the future (Park et al., 2008).

Conclusions

The construction of metagenomic libraries using a fosmid cloning system is rapidly becoming one method of choice for exploring environmental microbial communities, especially those that cannot be cultured. However, the challenge remains primarily in the ability to isolate large enough quantities of quality metagenomic DNA from low-abundance organisms. By amplifying the minute amounts of extracted DNA by WGA and constructing fosmid libraries from the amplified DNA, it is now possible to access genomic information from extreme environments.

ACKNOWLEDGEMENTS

This work was funded by the Ministry of Economic Affairs of the Republic of China (Grant 93-EC-17-A-17-R7-0563).

REFERENCES

- Abulencia CB, Wyborski DL, Garcia JA, Podar M, Chen W, Chang SH, Chang HW, Watson D, Brodie EL, Hazen TC, Keller M (2006). Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* 72:3291-3301.
- Alexander M (1977). *Introduction to Soil Microbiology*. 2nd Edition; John Wiley and Sons Inc. NY. pp. 1- 467
- Bertrand H, Poly F, Van VT, Lombard N, Nalin R, Vogel NT, Simonet P (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J. Microbiol. Methods*, 62:1-11.
- Bruce KD, Hiorns WD, Hobman JL, Osborn AM, Strike P, Ritchie DA (1992). Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Appl. Environ. Microbiol.* 58:3413-3416.
- Couto GH, Glogauer A, Faoro H, Chubatsu LS, Souza EM, Pedrosa FO (2010) Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genet. Mol. Res.* 9:514-523.
- Daniel R (2005) The metagenomics of soil. *Nat. Rev. Microbiol.* 3:470-478.
- Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M, Lasken RS (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA.*, 99:5261-5266.
- Desai C, Madamwar D (2007). Extraction of inhibitor-free metagenomic DNA from polluted sediments, compatible with molecular diversity analysis using adsorption and ion-exchange treatments. *Bioresour. Technol.* 98:761-768.
- Devries S, Nyante S, Korkola J, Segraves R, Nakao K, Moore D, Bae H, Wilhelm M, Hwang S, Waldman F (2005). Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J. Mol. Diagn.* 7:65-71.
- Ferrer M, Beloqui A, Timmis KN, Golyshin PN (2009). Metagenomics for mining new genetic resources of microbial communities. *J. Mol. Microbiol. Biotechnol.* 16:109-123.
- Frostegard A, Courtois S, Ramisse V, Clerc S, Bernillon D, Francoise Le Gall FL, Fuhrman JA (2002). Community structure and function in prokaryotic marine plankton. *Ant. Van Leeuwen.* 81:521-527.
- Grzymski JJ, Carter BJ, DeLong EF, Feldman RA, Ghadiri A, Murray AE (2006). Comparative genomics of DNA fragments from six Antarctic marine planktonic bacteria. *Appl. Environ. Microbiol.* 72:1532-1541.
- Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J, Preston C, de la Torre J, Richardson PM, DeLong EF (2006a). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. USA.*, 103: 18296-18301.
- Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM, DeLong EF (2006b). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Bio.* 4:e95.
- Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68:669-685.
- Janssen PH (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA Genes. *Appl. Environ. Microbiol.* 72:1719-1728.
- Klein CA, Schmidt-Kittler O, Schardt JA, Pantel K, Speicher MR, Gert Riethmüller G (1999). Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Nat. Acad. Sci., USA*, 96:4494-44949.
- Kowalchuk GA, Speksnijder AG, Zhang K, Goodman RM, van Veen JA (2007). Finding the needles in the metagenome haystack. *Microb. Ecol.* 53:475-485.
- Lee MH, Lee CH, Oh TK, Song JK, Yoon JH (2006). Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Appl. Environ. Microbiol.* 72:7406-7409.
- Leff LG, Dana JR, McArthur JV, Shimkets LJ (1995). Comparison of methods of DNA extraction from stream sediments. *Appl. Environ. Microbiol.* 61:1141-1143.
- Liesack W, Stackebrandt E (1992). Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* 174:5072-5078.
- Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM (2003). A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl. Environ. Microbiol.* 69:2684-2691.
- Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW (2005). Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl. Environ. Microbiol.* 71:7768-7777.
- Ling J, Zhuang G, Tazon-Vega B, Zhang C, Cao B, Rosenwaks BZ, Xu K (2009). Evaluation of genome coverage and fidelity of multiple displacement amplification from single cells by SNP array. *Mol. Hum. Reprod.* 15:739-747.
- Ludwig W, Klenk HP (2001). Overview: A phylogenetic backbone and taxonomic framework for prokaryotic systematics. in "Bergey's Manual of Systematic Bacteriology" (D. R. Boone, and R. W. Castenholz, Eds.). pp.49-65, Springer-Verlag, Berlin.
- Papke RT, Douady CJ, Doolittle WF, Rodríguez-Valera F (2003). Diversity of bacteriorhodopsins in different hypersaline waters from a

- single Spanish saltern. *Environ. Microbiol.* 5:1039-1045.
- Park SJ, Kang CH, Chae JC, Rhee SK (2008). Metagenome microarray for screening of fosmid clones containing specific genes. *FEMS Microbiol. Lett.* 284:28-34.
- Pirker C, Raidl M, Steiner E, Elbling L, Holzmann K, Spiegl-Kreinecker S, Aubele M, Grasl-Kraupp B, Marosi C, Micksche M, Berger W (2004). Whole genome amplification for CGH analysis: Linker-adaptor PCR as the method of choice for difficult and limited samples. *Cytometry A* 61:26-34.
- Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* 66:2541-2547.
- Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn M, Chisholm SW (2009). Whole genome amplification and de novo assembly of single bacterial. *Cells PLoS ONE*, 4:e5299.
- Schmeisser C, Steele H, Streit WR (2007). Metagenomics, biotechnology with non-culturable microbes. *Appl. Microbiol. Biotechnol.* 75:955-962.
- Telenius H, Carter NP, Bebb CE, Nordenskjöld M, Ponder BA, Tunnacliffe A (1992). Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13:718-725.
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur E J, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005). Comparative metagenomics of microbial communities. *Science* 308:554-557.
- Tsai Y, Olson BH (1992). Rapid method for separation of bacterial DNA from humic substances for polymerase chain reaction, *Appl. Environ. Microbiol.* 58:2292-2295.
- Webster G, Newberry CJ, Fry JC, Weightman AJ (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.
- Yeates C, Gillings MR, Davison AD, Altavilla N, Veal DA (1998). Methods for microbial DNA extraction from soil for PCR amplification. *Biol. Proced. Online*, 1:40-47.
- Yokouchi H, Fukuoka Y, Mukoyama D, Calugay R, Takeyama H, Tadashi MT (2006). Whole-metagenome amplification of a microbial community associated with scleractinian coral by multiple displacement amplification using ϕ 29 polymerase. *Environ. Microbiol.* 8:1155-1163.
- Yung PY, Burke C, Lewis M, Egan S, Kjelleberg S, Thomas T (2009). Phylogenetic screening of a bacterial, metagenomic library using homing endonuclease restriction and marker insertion. *Nucleic Acids Res.* 37:e144.
- Zhang L, Cui X, Schmitt K, Hubert R, Navidi W, Arnheim N (1992). Whole Genome Amplification from a Single Cell: Implications for Genetic Analysis. *Proc. Natl. Acad. Sci. USA*, 89:5847-5851.
- Zhou J, Bruns MA, Tiedje JM (1996). DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* 62:316-322.