J. Mol. Microbiol. Biotechnol. (2001) 3(2): 301-308.

JMMB Research Article

Expression and Isolation of Antimicrobial Small Molecules from Soil DNA Libraries

I.A. MacNeil*, C. L. Tiong, C. Minor[1], P.R. August, T.H. Grossman[2], K.A. Loiacono[3], B.A. Lynch[4], T. Phillips[5], S. Narula, R. Sundaramoorthi, A. Tyler[6], T. Aldredge[7], H. Long, M. Gilman[8], D. Holt[9], and M.S. Osburne

Aventis Pharmaceuticals, 26 Landsdowne St, Cambridge, MA 02139, USA
[1]Current address: Millenium Pharmaceuticals Inc., Cambridge, MA 02139, USA
[2]Current address: Vertex Pharmaceuticals Inc., Cambridge, MA 02139, USA
[3]Current address: Eisai Research Institute of Boston Inc., Andover, MA 01810, USA
[4]Current address: UCB Research Inc., Cambridge, MA 02139, USA
[5]Current address: Massachucetts Biologics Labs, Jamaica Plain, MA 02130, USA
[6]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
[7]Current address: Center for Genomics Research, Harvard University, Cambridge, MA 02138, USA
[8]Current address: Biogen Inc., Cambridge, MA 02142, USA
[9]Current address: SmithKline Beecham, King of Prussia, PA, USA

Abstract

Natural products have been a critically important source of clinically relevant small molecule therapeutics. However, the discovery rate of novel structural classes of antimicrobial molecules has declined. Recently, increasing evidence has shown that the number of species cultivated from soil represents less than 1% of the total population, opening up the exciting possibility that these uncultured species may provide a large untapped pool from which novel natural products can be discovered. We have constructed and expressed in *E. coli* a BAC (bacterial artificial chromosome) library containing genomic fragments of DNA (5-120kb) isolated directly from soil organisms (S-DNA). Screening of the library resulted in the identification of several antimicrobial activities expressed by different recombinant clones. One clone (mg1.1) has been partially characterized and found to express several small molecules related to and including indirubin. These results show that genes involved in natural product synthesis can be cloned directly from S-DNA and expressed in a heterologous host, supporting the idea that this technology has the potential to provide novel natural products from the wealth of environmental microbial diversity and is a potentially important new tool for drug discovery.

Received October 1, 2000; revised December 5, 2000; accepted December 6, 2000. *For correspondence. Email ian.macneil@aventis.com; Tel. (617) 768-4084; Fax. (617) 374-8811.

Introduction

The majority of the more than 5000 known anti-infective compounds are natural product derivatives, with over 100 in clinical use (Harvey, 2000; Dixon, 1994). Furthermore, natural products are the main source of cancer chemotherapeutics (Cragg and Newman, 2000), immunomodulating compounds, and other pharmaceutical molecules(Clark, 1996; Morris *et al.,* 1990). Despite advances in high throughput screening technology and attempts to isolate and culture novel microorganisms from exotic environments, the discovery rate of novel structural classes of antimicrobial molecules has declined (Strohl, 2000; Silver and Bostian, 1990). However, evidence that the vast majority of microorganisms in environmental samples are still unknown is accumulating (Pace, 1997; Stahl, 1993; Stahl *et al.,* 1985; Suzuki *et al.,* 1997; Ward *et al.,* 1990), as many are unculturable under standard laboratory conditions (Bull *et al.,* 2000). These newly-detected bacteria belong to either known families or apparently novel groups, as determined by 16S rRNA gene analysis (Bintrim *et al.,* 1998; Hugenholtz *et al.,* 1998). Since the number of microbial species currently cultivatable from soil is thought to represent only 1% or less of the total population, these uncultured species may provide a large untapped pool of novel natural products (Griffiths *et al.,* 1996; Torsvik *et al.,* 1990; Whitman *et al.,* 1998).

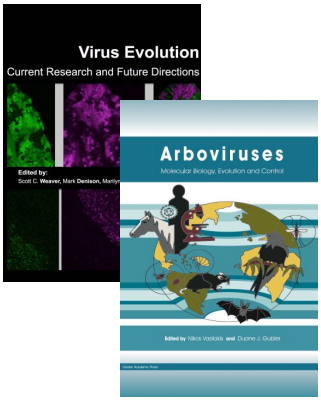
To circumvent the inability to culture and screen most soil microrganisms, microbial DNA can be extracted directly from the soil matrix and inserted into vectors propagated in bacterial strains such as *E. coli*, that are easy to grow and genetically manipulate. Theoretically, this method provides access to the entire “genomic” DNA content of the soil (the soil “metagenome” (Handelsman *et al.,* 1998; Rondon *et al.,* 2000; Rondon *et al.,* 1999)). However, the success of this technology depends upon both the ability to clone sufficiently large contiguous fragments of S-DNA into appropriate vectors and the ability to express heterologous DNA effectively (Rondon *et al.,* 2000).

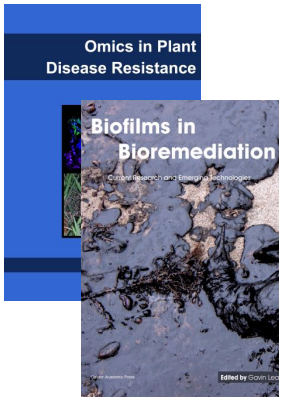
The assumption that large DNA fragments are important stems from the fact that many biosynthetic gene clusters encoding natural products range from approximately 30-100kb (Donadio *et al.,* 1991; Krause and Marahiel, 1988; Schwecke *et al.,* 1995). A number of vector systems can maintain large fragments of heterologous DNA, including BAC vectors. These vectors replicate in single copy in *E. coli* and can stably maintain heterologous DNA inserts of up to 300 kb (Cai *et al.,* 1995; Stone *et al.,* 1996; Suzuki *et al.,* 1997). Although early BAC vectors were problematic in terms of low DNA yield and lack of screens

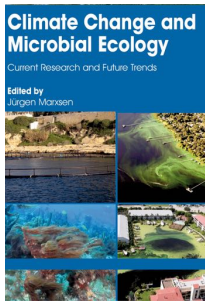
Further Reading

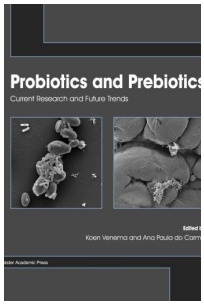
Caister Academic Press is a leading academic publisher of advanced texts in microbiology, molecular biology and medical research. Full details of all our publications at **caister.com**

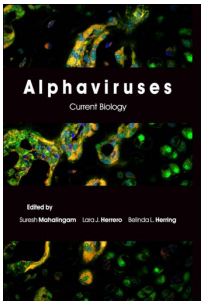
- **MALDI-TOF Mass Spectrometry in Microbiology**
  Edited by: M Kostrzewa, S Schubert (2016)
  www.caister.com/malditof
- ***Aspergillus* and *Penicillium* in the Post-genomic Era**
  Edited by: RP Vries, IB Gelber, MR Andersen (2016)
  www.caister.com/aspergillus2
- **The Bacteriocins: Current Knowledge and Future Prospects**
  Edited by: RL Dorit, SM Roy, MA Riley (2016)
  www.caister.com/bacteriocins
- **Omics in Plant Disease Resistance**
  Edited by: V Bhadauria (2016)
  www.caister.com/opdr
- **Acidophiles: Life in Extremely Acidic Environments**
  Edited by: R Quatrini, DB Johnson (2016)
  www.caister.com/acidophiles
- **Climate Change and Microbial Ecology: Current Research and Future Trends**
  Edited by: J Marxsen (2016)
  www.caister.com/climate
- **Biofilms in Bioremediation: Current Research and Emerging Technologies**
  Edited by: G Lear (2016)
  www.caister.com/biorem
- **Microalgae: Current Research and Applications**
  Edited by: MN Tsaloglou (2016)
  www.caister.com/microalgae
- **Gas Plasma Sterilization in Microbiology: Theory, Applications, Pitfalls and New Perspectives**
  Edited by: H Shintani, A Sakudo (2016)
  www.caister.com/gasplasma
- **Virus Evolution: Current Research and Future Directions**
  Edited by: SC Weaver, M Denison, M Roossinck, et al. (2016)
  www.caister.com/virusevol
- **Arboviruses: Molecular Biology, Evolution and Control**
  Edited by: N Vasilakis, DJ Gubler (2016)
  www.caister.com/arbo
- ***Shigella*: Molecular and Cellular Biology**
  Edited by: WD Picking, WL Picking (2016)
  www.caister.com/shigella
- **Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment**
  Edited by: AM Romaní, H Guasch, MD Balaguer (2016)
  www.caister.com/aquaticbiofilms
- **Alphaviruses: Current Biology**
  Edited by: S Mahalingam, L Herrero, B Herring (2016)
  www.caister.com/alpha
- **Thermophilic Microorganisms**
  Edited by: F Li (2015)
  www.caister.com/thermophile

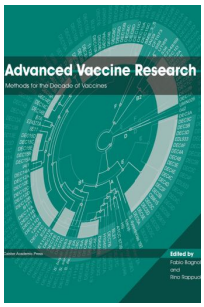


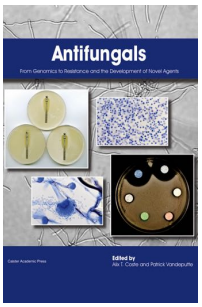












- **Flow Cytometry in Microbiology: Technology and Applications**
  Edited by: MG Wilkinson (2015)
  www.caister.com/flow
- **Probiotics and Prebiotics: Current Research and Future Trends**
  Edited by: K Venema, AP Carmo (2015)
  www.caister.com/probiotics
- **Epigenetics: Current Research and Emerging Trends**
  Edited by: BP Chadwick (2015)
  www.caister.com/epigenetics2015
- ***Corynebacterium glutamicum*: From Systems Biology to Biotechnological Applications**
  Edited by: A Burkovski (2015)
  www.caister.com/cory2
- **Advanced Vaccine Research Methods for the Decade of Vaccines**
  Edited by: F Bagnoli, R Rappuoli (2015)
  www.caister.com/vaccines
- **Antifungals: From Genomics to Resistance and the Development of Novel Agents**
  Edited by: AT Coste, P Vandeputte (2015)
  www.caister.com/antifungals
- **Bacteria-Plant Interactions: Advanced Research and Future Trends**
  Edited by: J Murillo, BA Vinatzer, RW Jackson, et al. (2015)
  www.caister.com/bacteria-plant
- ***Aeromonas***
  Edited by: J Graf (2015)
  www.caister.com/aeromonas
- **Antibiotics: Current Innovations and Future Trends**
  Edited by: S Sánchez, AL Demain (2015)
  www.caister.com/antibiotics
- ***Leishmania*: Current Biology and Control**
  Edited by: S Adak, R Datta (2015)
  www.caister.com/leish2
- ***Acanthamoeba*: Biology and Pathogenesis (2nd edition)**
  Author: NA Khan (2015)
  www.caister.com/acanthamoeba2
- **Microarrays: Current Technology, Innovations and Applications**
  Edited by: Z He (2014)
  www.caister.com/microarrays2
- **Metagenomics of the Microbial Nitrogen Cycle: Theory, Methods and Applications**
  Edited by: D Marco (2014)
  www.caister.com/n2

Order from caister.com/order

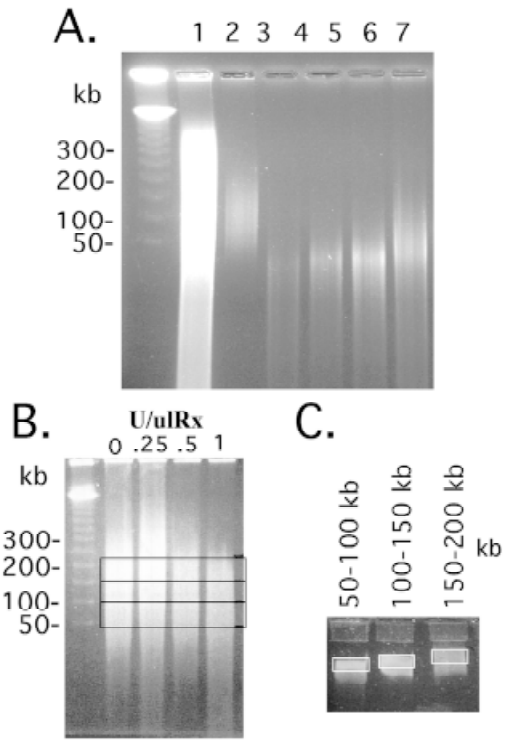


Figure 1. Panel A: Size analysis of S-DNA. DNA extracted from soil samples as described in Experimental Procedures were analysed using PFGE (1% agarose/0.5X TBE gel, 0.5 s initial switch time, 35 s final switch time, 6V/cm, 14°C, 20 h). Lane 1- molecular size standards, lane 2- partially purified S-DNA before sucrose gradient purification, lane 3- undigested DNA following sucrose gradient purification, lane 4- post-sucrose digested with *Eco*RI, lane 5- post-sucrose digested with *Hin*dIII, lane 6- post-sucrose digested with twice with *Hin*dIII, lane 7- post-sucrose digested with *Alu*I. Panel B: Sucrose gradient-purified DNA was digested with varying amounts of *Hin*dIII then subjected to pulse-field gel electrophoresis, and gel slices corresponding to various molecular weight ranges were excised. Numbers above lanes correspond to units/50μl reaction. Panel C- Excised gel slices were subjected to a second round of PFGE using low-melting point agar (1% agaraose/0.5X TBE, 0.5 s initial, 35 s final switch time, 6V/cm, 14°C, 4h). Excised gels fragments were used for the ligation reactions.

to indicate the presence of DNA inserts, improved vectors that are more amenable to generating large insert genomic libraries have been developed (Mejía and Monaco, 1997; Osoegawa *et al.,* 1998 this report).

The technology to construct random DNA libraries from environmental sources comprising many microbial species has been difficult to develop. The construction and screening of two S-DNA libraries, with average inserts of 27-45 kb were reported recently (Rondon *et al.,* 2000). These libraries were shown to encode and express many types of enzymatic activities produced by a variety of proteins. Here we report details of the construction and analysis of an *E. coli* soil library containing 12,000 clones, 4 of which expressed antimicrobial activity in the surrogate *E. coli* host strain and were identified by screening the BAC library in a high-throughput format. We also describe several new tools that facilitate library construction and analysis, and demonstrate that the S-DNA used to construct the library represents a wide spectrum of diverse and unknown organisms.

Results

Construction of S-DNA Library

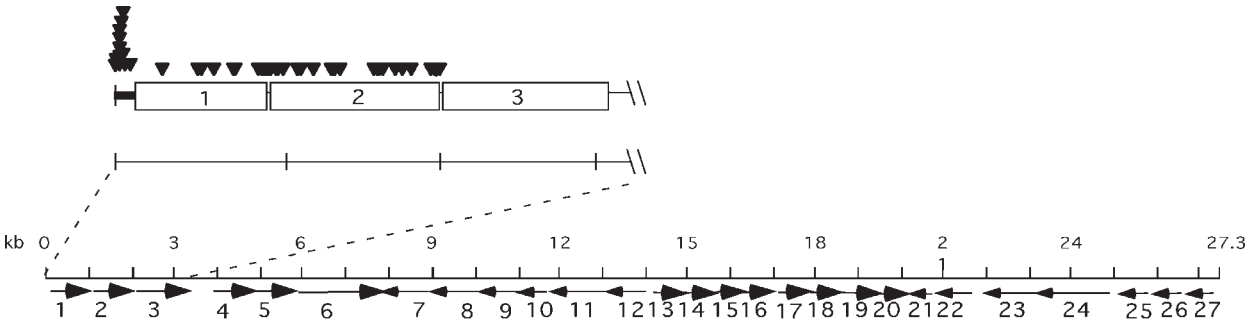
Because of their inherent fragility, the isolation of large fragments of S-DNA that are amenable to cloning is a difficult process. Therefore, we focused on developing methods that allow relatively large fragments of S-DNA to be routinely obtained in a form suitable for ligation. For this process, gentle handling of the material is crucial, although sequential purification schemes utilizing methods that separate DNA based on different physical and chemical properties are necessary. Humic acids are a major contaminant of soil samples and can inhibit PCR reactions (Porteous *et al.,* 1994; Tsai and Olson, 1992; Zhou *et al.,* 1996), restriction enzymes(Porteous *et al.,* 1994), and reduce transformation efficiency (Tebbe and Vahjen, 1993). Due to the overall charge and size of humic acids, simple electrophoresis and phenol extraction is insufficient for their removal. Figure 1 (panel A, lane 2) depicts the size range of DNA isolated from soil using gentle hot phenol extraction followed by treatment with CTAB. This latter treatment has been shown to remove humic acids as well as carbohydrates (Zhou *et al.,* 1996). However, despite an abundance of large DNA fragments, the material was refractory to digestion with restriction endonucleases (not shown) and also contained smaller (<20 kb) fragments. Therefore, we purified and sized the DNA further using a discontinuous sucrose density gradient. As shown in panel A, lane 3, DNA isolated from this gradient ranged from 50 - 200 kb, and could be efficiently digested with restriction endonucleases (panel A, lanes 4-7). Following partial digestion with *HindIII* (Figure 1, panel B) and further sizing and purification (Methods), separate ligations into modified BAC vector pBTP2 were performed with three size ranges of purified DNA (Figure 1, panel C), followed by transformation into *E. coli* strain DH10B. The insert sizes of 100 random clones representing all 3 molecular weight pools ranged from 5-120 kb, with an average size of 37 kb. Twenty seven per cent of clones had no insert.

12,000 colonies were picked and arrayed in 384-well plates, constituting the MG1 library. This number of clones represents only a small fraction of the total number obtainable from 400 ng of S-DNA, since only a small portion of the purified DNA (representing approximately 440 Mbp of DNA) was used. In total, 500 μg of purified HMW S-DNA was isolated from 400 g of soil. Ligation of the entire amount could potentially produce 1.5 X 10^7 clones, representing over 500 GB of S-DNA.

Phylogenetic Analysis of S-DNA

To evaluate the microbial diversity represented by the DNA used to construct our soil library, we carried out a phylogenetic analysis of DNA sequences encoding small subunit ribosomal RNA (SSU rRNA) (Pace, 1997). As we reported previously (Osburne *et al.,* 2000), the results indicated a wide representation of sequences that could be placed within known families isolated from all over the world. However, the majority of the 16S sequences were from previously unidentified bacterial families. A representative phylogenetic tree based on 76 sequences obtained from the S-DNA is presented in Figure 2. Also included within the tree are several known bacterial

Figure 2. Phylogenetic tree. Purified DNA obtained for the construction of the MG1 library was used as a template for PCR amplification (Methods). For the construction of the phylogenetic tree, alignment of the first 400 nucleotides (5') from 76 independent clones and known 16S rRNA sequences (obtained from Ribosomal Database Project (RDP) at the University of Illinois) was acomplished using ClustalX (freeware). Phylogenetic analysis was performed using the Phylogenetic Inference Package (PHYLIP, version 3.57, J. Felsenstein, University of Washington, Seattle). The tree was inferred by neighbor-joining analysis of 437 homologous positions. The percent bootstrap values (over 75%) from 100 resamplings that support the major topological elements for neighbor joining (above line) and parsimony (below line) are shown.



1 2-hydroxyhepta-2,4-diene-1,7-dioate Isomerase
2 indole dioxygenase
3 indole dioxygenase
4 unknown
5 unknown
6 Acetyl-CoA synthetase
7 β-ketoadipyl-CoA thiolase
8 maleylacetate reductase
9 gluconate CoA-transferase subunit B
10 gluconate CoA-transferase subunitA
11 aldehyde dehydrogenase
12 methionyl aminopeptidase
13 unknown
14 succinyl coA transferase
15 unknown
16 unknown
17 unknown
18 nod B
19 Acetyl CoA acetyl transferase
20 Acetyl CoA synthetase
21 unknown
22 unknown
23 invasion-associated protein
24 GMP synthase
25 unknown
26 amino transferase
27 unknown

Figure 3. ORF map of mg1.1. Sequence was determined as described in the methods. ORF's were identified using MapDraw (DNAStar Inc.). Homology search was done using BLAST (Basic Local Alignment Search Tool, http://www.ncbi.nlm.nih.gov/BLAST). Inset illustrates relative positions of transposon insertions which produced colorless mutants (triangles)

families. Of note is the match of clone 1906B01F to a species in the Flexibacter-Bacteroides-Cytophaga phylum (Bowman *et al.,* 1997; Osburne *et al.,* 2000) which has been found in glacial antarctic ice.

Initial Characterization of Small Molecules Produced by MG1 Clones

Four nonidentical clones containing S-DNA-encoded antibacterial activity were detected in the MG1 library. This represents a hit rate of roughly 1 anti-bacterial clone per 60 MB of soil-derived DNA. Two clones, with S-DNA insert sizes of approximately 23 and 35 kb, produced weak antibacterial activities. However, when the copy numbers of the encoding plasmids were increased by insertion of pTRANS (Methods), antibacterial activities were enhanced (data not shown). Efforts to extract these activities were unsuccessful, as were attempts to recover the activities from filtered culture supernatants. These data suggest that the activities may be cell-associated and/or possibly conferred by a protein or peptide.

A third clone (pA2), with S-DNA insert size of only 7 kb, displayed weak antimicrobial activity against *B. subtilis* and also produced a brown pigment. DNA sequence analysis of the insert revealed 4 ORFs with homologies to genes encoding long chain fatty acid coA ligase, phenylalanine-4-monooxygenase, 4-hydroxyphenyl-pyruvate dioxygenase and cylcohexadienyl dehydratase (from the Proteobacteria beta- and gamma- subdivisions). The latter 3 enzymes are known to be involved in the conversion of tyrosine and phenylalanine to homogentisic acid, which polymerizes non-catalytically to a brown-colored ochronotic pigment (Denoya *et al.,* 1994). Purified homogentisic acid (Sigma) was found to have slight antibacterial activity against *B. subtilis*, suggesting the reason for the activity of clone pA2.

The fourth antibacterial clone, mg 1.1 (insert size of 27 kb), had activity against *B. subtilis* and *S. aureus* and produced a purple pigment. To obtain genetic, functional, and DNA sequence information for the biosynthetic genes encoding these activities, transposon mutagenesis using pTRANS-*SacB* was employed. All the mutant colonies with dramatically reduced pigment production were found to have a single transposon insertion within or near one of two ORFs (Figure 3). The first ORF (ORF1) encodes a putative protein with homology to 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase, a bifunctional enzyme with both isomerase and decarboxylase activities and with homologues in many classes of microorganisms including *E. coli* (Roper and Cooper, 1993). The second ORF (ORF2) encodes a protein similar to indole dioxygenases (Figure 4), a family of enzymes that produces indole-related compounds in other organisms by catalyzing the incorporation of molecular oxygen (August *et al.,* 2000; O'Connor *et al.,* 1997; Yen *et al.,* 1991). This latter gene is followed by a second indole dioxygenase-like gene (ORF 3), although none of the insertions that visually affected pigment production mapped in this gene. Despite reduced pigment production, all of the mutants with single insertions in ORFs 1 or 2 retained antibacterial activity. To analyze these genes further, each of the 3 unmutated ORFs was then subcloned individually into pTOPO4Blunt (Invitrogen) and transformed into strain DH10B. Either of the

```
          71                                                  120

Orf 2     SREVARGCGS TAWVTALMNV CAFFVACMNE QAQDDVWGAN PDARIAGVLN
Orf 3     MSALGRGCGS TAWVAGVVHA HAWLVSHFPE QGQDDVYGNN PDAVVSAVIG
     A    VAVLGEACAS TAWYASLTAS LGRMAAYLPD EGQAELWSDG PDALIVGALM
     B    VMRMASLNPS AGWVHGIVGV HPWQLAFADP KVQQEIWGSD PDTWMASPYM
     C    MATLTEADGS FAWVTSIYNA VGHMVCAFGD QALEEFLGS. .EIPRSAGVF
     D    VAVVGEGCTS AAWAASLTAS LGRMAAYLPE AGRRRIWAGG PDTLIVGALM
     E    LVTVGTACAA SSWSASIVAG VGRMAGFLPA EGRAEVWKDG PDAVVVGSLA
     F    VRTVARYSVA AGWLTYFYSM HEVWAAYLPP KGREEIFGQG ..GLLADVVA
     G    IERISEADGS AGWVASFGDG .ARYLAALPD DTLRKVYANG PDVVLAGALF
     H    AEAVARGDAS AGWCVSIAIT SALLVAYLPA RSREEMFGGG .RGVAAGVWA

          121                                                 170
Orf 2     P.TATTRKVD GGIVVTGAWP WASGSYHADW SYVGVPIMN. DEGEFMYP..
Orf 3     P.RGKAVRTG DGYRLEGVWP FGSGSERADW LLLGAVVFD. EDGNEIDQ..
     A    P.LGRAEKTP GGWHVSGTWP FVSVVDHSDW ALICAKVG.. .EEP......
     B    P.GGMCIPTD GGYKFSGRWQ FSSGTDHCDW AFLGAMACDK DGNMEMPPRM
     C    AVTGRSAAVD GGYRVSGKWG FSSGQHHAGW IIVPGIPE.. .EGG..AP..
     D    P.FGRARREE GGWRIGGTWP YVSVVDHADW ALVCAMTT.. ..EERPVV..
     E    P.LGRAKAVP GGWRLSGTWP SISVVDFSDW ALVRAVVA.. D.SEGQAL..
     F    P.VGRVEKDG DGYRLYGQWN FCSGVLHSDW IGLGAMMELP DGNS...PEY
     G    P.LQPARRCA GGFVVNGVWP FASGSPGADL IGVGIKLED. D.PAGDLP..
     H    P.RGTARSVD GGVVVSGRWP FCSGINHADI MFAGCFVD.. DR...QVP..

          171                                                 220
Orf 2     AMALIPNSDI T.IEDTWFVA GMRGTGSNTL HADEVFVPDH RLHWVPGLLN
Orf 3     GDFMVPQSEV T.IKDDWYVS GLAGTGSCTM VVTGTDVPAH RFLSFPGLIM
     A    WFFAVPRQEY G.IVDSWYPM GMRGTGSNTL VLDGVFVPDA RACTRAAIAA
     B    LHVIIPRTDY EIIEDSWDVM GLRGTGSKDL VVKDAYVPDY RVMDCDEVID
     C    IAFLVPKAEF E.VKDDWFVT GFIGTGSNTV VLDDVFVPEY RAIPFMDIVT
     D    RFFAVPRGSW R.SEDTWSSV GMRGTGSNTL HVEDVFVSDE LTFTRDAVAT
     E    RVFVVPRAGY E.IQDTWSNV GMRATGSNTL VVDDVFVPDA RTFEGDDLFQ
     F    CLLVLPKSDV Q.IVENWDTM GLRASGSNGV LVEGAYVPLH RIFPAGRVMA
     G    RVAVIPAEKV T.IRPNWDVI GLRGTGSHDV VVSDVIVPEQ WTLIRGGPAT
     H    SVVALNKDEL Q.VLDTWHTL GLRGTGSHDC VADDVFVPAD RVFSVFDGPI
```

Figure 4. Partial comparison of the translated products of ORF 2 and 3 from mg1.1 with other bacterial dioxygenase proteins. Identical and similar amino acids are underlined and bolded. Numbering is with respect to ORF2 and 3.

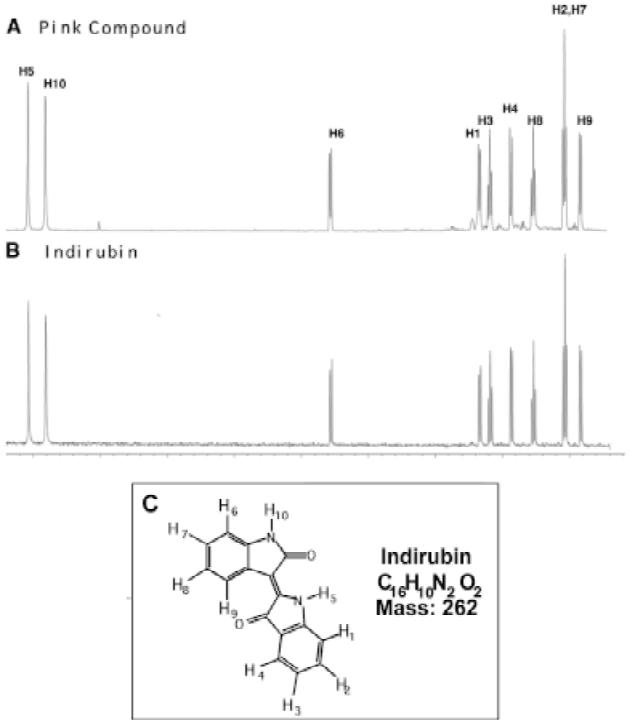


Figure 5. One-dimensional proton(^{1}H) NMR spectrum of pink compound (A) is compared to authentic sample of Indirubin (B) (Sigma). Both spectra are recorded in d_6-DMSO solvent at 27°C.The peaks are assigned and labelled according to the numbering scheme shown in the insert. The intensity of the triplet peak at 7.03 ppm integrated to two protons resulting in a total of 10 observable protons. The chemical shift value of all the protons indicate that these protons are either aromatic or bonded to heteroatoms such as oxygen or nitrogen. The two-dimensional (2D) ^{1}H-^{13}C heteronuclear (44) experiments, 2D ^{1}H-^{13}C multiple-bond correlation (45) and 2D ^{1}H-^{1}H correlation (46) experiments (data not shown) confirmed the proposed structure. The proton chemical shifts are referenced to TMS as standard at 0.0ppm. Panel C is the structure, determined by NMR, of indirubin isolated from mg1.1.

dioxygenase genes alone (ORF 2 or 3) produced both pigment and antibacterial activity (data not shown). This result explains why none of the mg1.1 single insertion mutants lacked antibacterial activity. The isomerase gene (ORF1) alone did not produce either activity, although addition of this gene to either ORF2 or ORF3 resulted in increased pigment and antibacterial activity.

TLC analysis of organic extracts of mg1.1 yielded both pink and blue pigments. Both pigments had a MW= 262, as determined by MS, and weak antibacterial activity. The genetic and DNA sequence data indicating that indole-dioxygenase genes were involved in pigment production (Figures 3 and 4), combined with the MW determination, suggested that the pigments could be indirubin and indigo blue, structural isomers previously shown to be co-produced in several microorganisms (Eaton and Chapman, 1995; Hart *et al.,* 1992). Authentic samples of indirubin, an antileukemic drug known to inhibit tyrosine kinases (Han, 1994; Hoessel *et al.,* 1999) and indigo blue were obtained (Apin Chemicals, UK) and shown to co-migrate with the unknown pigments purified from mg1.1 in TLC analysis. NMR analysis of the pink compound as compared with indirubin (Figure 5) was carried out using several methods (legend to Figure 5). The NMR results confirmed the indentity of the pink pigment produced by mg1.1 as indirubin.

Interestingly, TLC and HPLC fractionation of the extract and assay of the fractions showed that the majority of antibacterial activity was actually unpigmented (*B. subtilis* was sensitive to 1-4 μg/ml in LB liquid culture using HPLC-purified unpigmented material). Initial MS and NMR analysis of the unpigmented activity indicated that it is also related to indole, suggesting that *E. coli* clone mg1.1 produces a family of structurally related antimicrobial molecules encoded by the S-DNA insert. Further structural and genetic analysis of this family is in progress.

Discussion

Natural products have been a major source of pharmaceuticals used to treat human disease. Classically, the discovery process involves the isolation of novel producing organisms (usually microbes or plants), screening for activities, and extensive purification and testing of active compounds which may then form the starting material for new drugs. As the number of novel structures isolated from readily culturable microorganisms diminishes, the ability to tap into the biological diversity of other microorganisms, as described here, becomes paramount to the discovery of novel natural product-derived pharmaceuticals. Previously we screened random S-DNA libraries expressed in *E. coli* and found novel genes expressing novel enzymatic activities (Rondon *et al.,* 2000). Here we demonstrated that, using a related approach, biologically active small molecules from unknown organisms can be expressed and detected in laboratory strains of *E. coli.*

Our 12,000 clone library represents only a tiny fraction of the microbial genomes present in the soil. By scaling up the procedure, it should be relatively easy to produce a S-DNA library containing 10 million clones or more, that could theoretically represent multiple entire genomes from different organisms. Improvements to the current procedure (e.g., removing empty or very small inserts after ligation of the genomic DNA into the vector, eliminating restriction enzyme digestion by ligating the fragmented DNA directly into a suitable vector) should greatly enhance our ability to capture large genomic fragments consistently. However, as shown here, potentially interesting antibacterial activities can be encoded on small multigene clusters. Furthermore, heterologous genes might be expressed more readily from small clusters that reside on small fragments, since their expression might be driven from host promoters. Thus a comprehensive effort to discover novel natural products might include producing and screening both small-insert and large-insert environmental DNA libraries, especially if "producing" clones could be culled from the many non-producing clones by the use of engineered reporter strains.

The use of *E. coli* as a heterologous expression host may limit the ability to express DNA from many soil microbes. Ideally, libraries using heterologous hosts with different expression properties, such as *Bacillus*, *Streptomyces*, or fungi, could complement *E. coli* libraries (Sosio, M. *et al.,* 2000) Thus, expression of a particular library in several hosts should enhance the ability to detect novel activities.

Using more conventional methods of natural products discovery involving fermentation of native strains, it is often difficult to duplicate production conditions in re-fermentation attempts. A significant advantage of the approach described here is the potential to manipulate the already-cloned producing genes to increase production. Further advantages of this method include the use of DNA sequence information to help identify unknown natural products, as clearly demonstrated here for the genes encoding monooxygenases. An additional benefit stems from the potential for enzymes produced by heterologous genes to interact with endogenous proteins, thereby producing novel natural products.

In summary, using a simple "shotgun" approach of cloning DNA derived from the soil into BAC vectors and expressing the DNA in *E. coli*, we were able to screen for and discover small molecule antibacterial activities. Our efforts were aided substantially by improvements to existing BAC and transposon vectors, and by methods for isolating and preparing DNA from the soil. Our results clearly support the potential of this technology to provide novel natural products from the wealth of environmental microbial diversity, and to contribute to our understanding of the vast reservoir of unknown terrestrial microbes.

Experimental Procedures

Isolation of S-DNA

Soil was obtained during the month of March from a local New England source that has remained uncultivated for at least 50 years. Approximately 75 g of sieved soil was gently mixed with 100 ml Buffer I (25 mM Tris 8.0, 150 mM NaCl, 25 mM EDTA) and the suspension was extracted with 50 ml phenol (pre-equilibrated to 65°C) and then centrifuged at 4000 rpm for 20 min. Crude DNA in the aqueous layer was precipitated with 0.7 vol of isopropanol. Six ml of Buffer I + 0.5% SDS and 100 μg/ml Proteinase K were added to the DNA pellet and the sample was incubated at 50°C for 2h. Six hundred μl of 5M NaCl and 6 mls of 2% CTAB (Cetyltrimethyl-ammonium bromide) in 2 M NaCl were added and the sample was mixed for 15 min. Proteins and carbohydrates precipitated by CTAB were removed by chloroform extraction. The DNA was precipitated with 0.7 vol isopropanol, washed with 70% ethanol and resuspended without mixing in 1 ml TE. The supernatant was layered onto a 32-ml sucrose gradient (four 8-ml steps:

50%, 40%, 30%, and 20% sterile sucrose dissolved in TE) and centrifuged for 21 h at 28,000 rpm (SW-40 rotor). Fractions containing high molecular weight (HMW) DNA were collected from the bottom of the tube, which was punctured with a 16 gauge needle. The DNA of each fraction was precipitated with 0.7 vol of isopropanol and resuspended in TE. HMW fractions were identified by gel electrophoresis, pooled, and used for library construction.

Construction of BAC Library

S-DNA was partially digested with restriction enzyme *HindIII* (New England Biolabs) and then purified by pulsed-field gel electrophoresis (PFGE). Gel slices containing HMW DNA were excised and digested with GELase (Epicentre) for 1 h at 40°C. S-DNA was ligated to *HindIII*-digested BAC vector pBTP2 at a ratio of 10:1 (vector to insert). Vector pBTP2 is a modified version of pBeloBAC11 (Shizuya *et al.,* 1992) that contains additional cloning sites as well as a pUC origin of replication inserted into the polylinker, allowing for high-copy replication of the empty vector to facilitate purification. The pUC sequence is removed by restriction endonuclease digestion before ligation to insert DNA.

Ligations (16°C for 10 h, T4 DNA ligase, New England Biolabs), were used to transform *E. coli* strain DH10B electrocompetent cells (Gibco/BRL). Recombinants were selected at 37°C overnight on LB agar plates containing 12.5 μg/ml chloramphenicol, 50 μg/ml X-gal, and 25 μg/ml IPTG. Clones were arrayed into 384-well plates using a Q-Bot (Genetix) robot and grown overnight in LB at 37°C. Glycerol was added to a final concentration of 20% and plates were stored at -80°C.

Diversity Analysis

Purified DNA used in the construction of the MG1 library was the template for PCR amplification of 16S rRNA genes. The oligonucleotide primers chosen have been shown to amplify most eubacterial 16S rRNA sequences (27f: 5'-AGAGTTTGATCMTGGCTCAG-3' and 1492r: 5'TACGGYTAC CTTGTTACGACTT-3', where M=C/A, and Y=C/T; Weisburg *et al.,* 1991). Amplified sequences were cloned into pZero-Blunt (Invitrogen) and sequenced on an ABI 377 sequencer. 16S rRNA gene sequences were identified using BLAST (Basic Local Alignment Search Tool, http://www.ncbi.nlm.nih.gov/BLAST).

Assays

For antibacterial assays, clones were grown on LB+chloramphenicol (10 μg/ml) for 1 day at 37ºC and 6 days at room temperature (~25ºC). Plates were then overlaid with top agar containing exponentially growing *Bacillus subtilis* strain BR151(pPL608) (Bacillus Genetic Stock Center, Columbus, OH) and incubated overnight at 37ºC, followed by further incubation at room temperature for several more days. Clones producing antibacterial activities were identified by a zone of inhibition in the lawn surrounding the clone. In some cases*, Staphylococcus aureus* (smith strain, ATCC #13709) was used for screening.

Transposon Mutagenesis and DNA Sequencing

Clones expressing heterologous activities were characterized by transposon mutagenesis and DNA sequencing using plasmid pTRANS-*sacB*. Plasmid pTRANS-*sacB* encodes the T7-based transposon TRANS (from plasmid pGPS1, New England BioLabs, Beverly, MA), a colE1 origin of replication, and a kanamycin resistance gene. *In vitro* transposition of TRANS allows for random insertion of the colE1 *ori* into a target BAC plasmid, increasing its copy number and thereby facilitating plasmid DNA isolation and sequencing, and expression. Plasmid pTRANS-*sacB* also encodes the *Bacillus subtilis sacB* gene in the vector portion of the plasmid, allowing its counterselection in the presence of 5% sucrose. Transposition reactions were performed essentially according to the published protocol for pGPS1, followed by transformation of electrocompetent *E. coli* strain DH10B with 5 μl of the transposon reaction and selection of transformants on LB plates containing kanamycin (50 μg/ml), chloramphenicol (10 μg/ml), and sucrose (5%). The resulting transformants contained multicopy BAC plasmids with TRANS insertions. For sequencing, plasmid DNA was isolated using the Qiagen Biorobot 9600 according to the manufacturer's instructions, and sequenced using ABI Big Dye sequencing kit and run on an ABI 377 DNA sequencer. Bases were assigned using the Unix program Phred, the data was assembled using Phrap, and edited using Consed (University of Washington, Seattle).

Isolation and Analysis of Small Molecules from MG1 Clones

E. coli clones were grown overnight in LB medium in 20 L batches at 37° in a 30L fermentor. The cell pellet (~170g) was resuspended in 500 ml PBS and extracted with 1L of CH_2Cl_2. The organic phase was concentrated by rotary evaporation and pigments were isolated by flash chromatography on silca in 1:1 Ethyl Acetate:Hexane. Fractions were assayed for antibacterial activity by spotting directly onto LB plates overlayed with *Bacillus subtilis*. Compounds were further purified on RP-HPLC using acetonitrile/water/0.1% TFA gradients. Thin-layer chromatography (TLC) analysis was performed in 1:1 Ethyl Acetate:Hexane.

Atmospheric Pressure Chemical Ionization (ApCI) mass spectra were obtained using a Platform II mass spectrometer (Micromass Inc, Beverly MA) with 50:50 water : acetonitrile: 0.1 formic acid as the mobile phase. The NMR data was acquired on Bruker DRX600 MHz NMR instrument equipped with triple resonance probe and triple axis gradient system.

Acknowledgements

We wish to thank Lynn Silver, David Rothstein, David Berstein, Sharon Hausdorff, Karen Krupen and Mark Nelligan for helpful discussions and suggestions. We thank Jon Clardy, Robert Goodman, Jo Handelsman, and Michelle Rondon for developing the concept of metagenomics and sharing the idea, their methods, and data prior to publication. We thank Robert Goodman and Brian Manske for assistance with the phylogenetic analysis and Jon Clardy and Sean Brady for assistance with structural chemistry.

References

August, P. R., Grossman, T. H., Minor, C., Draper, M. P., MacNeil, I. A., Pemberton, J. M., Call, K. M., Holt, D., and Osburne, M. S. 2000. Sequence Analysis and Functional Characterization of the Violacein Biosynthetic Pathway from *Chromobacterium violaceum*. J. Mol. Microbiol. Biotechnol. 2: 1-7.

Bintrim, S. B., Donohue, T. J., Handelsman, J., Roberts, G. P., and Goodman, R. M. 1998. Molecular phylogeny of Archaea from soil. Proc. Natl. Acad. Sci. USA. 94: 277-282.

Bowman, J. P., McCammon, S. A., Brown, M. V., Nichols, D. S., and McMeekin, T. A. 1997. Diversity and association of psychrophilic bacteria in Antartic sea ice. Appl. Environ. Microbiol. 63: 3068-3078.

Cai, L., Taylor, J. F., Wing, R. A., Gallagher, D. S., Woo, S. S., and Davis, S. K. 1995. Construction and characterization of a bovine bacterial artificial chromosome library. Genomics. 29: 413-425.

Clark, A. M. 1996. Natural products as a resource for new drugs. Pharm. Res. 13: 1133-44.

Bull, A.T., Ward, A.C., and Goodfellow, M. 2000. Search and discovery strategies for biotechnology: the paradigm shift. Microbiol. Mol. Biol. Rev. 64:573-606.

Cragg, G. M., and Newman, D. J. 2000. Antineoplastic agents from natural sources: achievements and future directions. Expert Opin. Investig. Drugs. 9:2783-2797.

Denoya, C. D., Skinner, D. D., and Morgenstern, M. R. 1994. A *Streptomyces avermitilis* gene encoding a 4-hydroxyphenylpyruvic acid dioxygenase-like protein that directs the production of homogentisic acid and an ochronotic pigment in *Escherichia coli*. J. Bacteriol. 176: 5312-5319.

Dixon, B. 1994. Power unseen: how microbes rule the world, W.H. Freeman, New York.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. 1991. Modular organization of genes required for complex polyketide biosynthesis. Science. 252: 675-679.

Eaton, R. W., and Chapman, P. J. 1995. Formation of indigo and related compounds from indolecarboxylic acids by aromatic acid-degrading bacteria: chromogenic reactions for cloning genes encoding dioxygenases that act on aromatic acids. J. Bacteriol. 177: 6983-6988.

Griffiths, B. S., Ritz, K., and Glover, L. A. 1996. Broad-scale approaches to the determination of soil microbial community structure:application of the community DNA hybridization technique. Microbial. Ecol. 31: 269-280.

Han, R. 1994. Highlight on the studies of anticancer drugs derived from plants in China. Stem Cells (Dayt). 12: 53-63.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem. Biol. 5: 245-249.

Hart, S., Koch, K. R., and Woods, D. R. 1992. Identification of indigo-related pigments produced by *Escherichia coli* containing a cloned *Rhodococcus* gene. J. Gen. Microbiol. 138: 211-216.

Harvey, A. 2000. Strategies for discovering drugs from previously unexplored natural products. Drug Discovery Today. 5: 294-300.

Hoessel, R., Leclerc, S., Endicott, J. A., Nobel, M. E. M., Lawrie, A., Tunnah, P., Leost, M., Damiens, E., Marie, D., Marko, D., Niederberger, E., Tang, W., Eisenbrand, G., and Meijer, L. 1999. Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. Nature Cell Biol. 1: 60-67.

Hugenholtz, P., Goebel, B. M., and Pace, N. R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180: 4765-4774.

Krause, M., and Marahiel, M. A. 1988. Organization of the biosynthesis genes for the peptide antibiotic gramicidin S. J. Bacteriol. 170: 4669-4674.

Mejía, J. E., and Monaco, A. P. 1997. Retrofitting vectors for *Escherichia*

coli-based artificial chromosomes (PACs and BACs) with markers for transfection studies. Genome Res. 7: 179-186.
Morris, R. E., Wu, J., and Shorthouse, R. 1990. A study of the contrasting effects of cyclosporine, FK 506, and rapamycin on the suppression of allograft rejection. Transplant Proc. 22: 1638-41.
O'Connor, K., Dobson, A. D., and Hartmans, S. 1997. Indigo formation by microorganisms expressing styrene monooxygenase activity. Appl. Environ. Microbiol. 63: 4287-4291.
Osburne, M. S., Grossman, T. H., August, P. R., and MacNeil, I. A. 2000. Tapping into microbial diversity for natural products drug discovery. ASM News. 66: 411-417.
Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., and de Jong, P. J. 1998. An improved approach for construction of bacterial artificial chromosome libraries. Genomics. 52: 1-8.
Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science. 276: 734-740.
Porteous, L. A., Armstrong, J. L., Seidler, R. J., and Watrud, L. S. 1994. An effective method to extract DNA from environmental samples for polymerase chain reaction amplification and DNA fingerprint analysis. Curr. Microbiol. 29: 301-307.
Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., and Goodman, R. M. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl. Environ. Microbiol. 66: 2541-2547.
Rondon, M. R., Goodman, R. M., and Handelsman, J. 1999. The Earth's bounty: assessing and accessing soil microbial diversity. Trends Biotechnol. 17: 403-409.
Roper, D. I., and Cooper, R. A. 1993. Purification, nucleotide sequence and some properties of a bifunctional isomerase/decarboxylase from the homoprotocatechuate degradative pathway of E. coli. Eur. J. Biochem. 217: 575-580.
Schwecke, T., Aparicio, J. F., Molnár, I., König, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortés, J., and Lester, J. B. 1995. The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. Proc. Natl. Acad. Sci. USA. 92: 7839-7843.
Shizuya, H., Birren, B., Kim, U. J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA. 89: 8794-8797.
Silver, L., and Bostian, K. 1990. Screening of natural products for antimicrobial agents. Eur. J. Clin. Microbiol. Infect. Dis. 9: 455-461.
Sosio, M., Giusin,o F., Cappellano, C., Bossi, E., Puglia, A.M., and Donadio S. 2000 Artificial chromosomes for antibiotic-producing actinomycetes. Nat. Biotechnol. 18: 343-345.
Stahl, D. A. 1993. The natural history of microorganisms. ASM News. 59: 609-613.
Stahl, D. A., Lane, D. J., Olsen, G. J., and Pace, N. R. 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. Appl. Environ. Microbiol. 49: 1379-1384.
Stone, N. E., Fan, J. B., Willour, V., Pennacchio, L. A., Warrington, J. A., Hu, A., de la Chapelle, A., Lehesjoki, A. E., Cox, D. R., and Myers, R. M. 1996. Construction of a 750-kb bacterial clone contig and restriction map in the region of human chromosome 21 containing the progressive myoclonus epilepsy gene. Genome Res. 6: 218-225.
Strohl, W.R. 2000. The role of natural products in a modern drug discovery program. Drug Discovery Today. 5:39-41.
Suzuki, M. T., Rappe, M. S., Haimberger, Z. W., Winfield, H., Adair, N., Strobel, J., and Giovannoni, S. J. 1997. Bacterial diversity among small-subunit rRNA gene clones and cellular isolates from the same seawater sample. Environ. Microbiol. 63: 983-989.
Tebbe, C. C., and Vahjen, W. 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl. Envir. Microbiol. 59: 2657-2665.
Torsvik, V., Salte, K., Sørheim, R., and Goksøyr, J. 1990. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl. Environ. Microbiol. 56: 776-781.
Tsai, Y. L., and Olson, B. H. 1992. Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. Appl. Envir. Microbiol. 58: 2292-2295.
Ward, D. M., Weller, R., and Bateson, M. M. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature. 345: 63-65.
Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. 173: 697-703.
Whitman, W. B., Coleman, D. C., and Wiebe, W. J. 1998. Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. USA. 95: 6578-6583.
Yen, K. M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., Lu, H. S., Harcourt, A. A., and Chen, K. K. 1991. Cloning and characterization of a *Pseudomonas mendocina* KR1 gene cluster encoding toluene-4-monooxygenase. J. Bacteriol. 173: 5315-5327.
Zhou, J., Bruns, M. A., and Tiedje, J. M. 1996. DNA recovery from soils of diverse composition. Appl. Envir. Microbiol. 62: 316-322.