Cancer Chemother Pharmacol (2004) 54: 1–6
DOI 10.1007/s00280-004-0771-8

REVIEW

Robin K. Pettit

Soil DNA libraries for anticancer drug discovery

Received: 10 October 2003 / Accepted: 8 December 2003 / Published online: 8 April 2004

Abstract Soil has the largest population of microbes of any habitat, but only about 0.3% of soil microbes are cultivable with current techniques. Cultured soil microbes have been an incredibly productive source of drugs, for example the cancer chemotherapeutics doxorubicin hydrochloride, bleomycin, daunorubicin and mitomycin. Unfortunately, the current yield of new drugs from soil microbes is low due to repeated cultivation of the same small fraction of cultivable microbes. Uncultured soil species represent a tremendous untapped resource of new antineoplastic agents. Methods have recently been developed to access the diversity of secondary metabolites from uncultured soil microbes. Briefly, total DNA is extracted from soil samples, purified, partially digested, and fragments inserted into vectors for expression in readily fermented microbes such as *Escherichia coli*. Clones expressing enzymatic and antibiotic activities that are encoded by novel sequences have been reported.

Keywords Bacterial artificial chromosome · Soil DNA libraries · Uncultured soil microbes · Anticancer agents

Introduction

The development of clinically useful drugs continues to rely heavily on isolation from natural sources. Even with major efforts in rational drug design, combinatorial chemistry, peptide libraries and antibody libraries, natural products are still unsurpassed in novelty and complexity [7]. One area where natural products have had a major impact on longevity and quality of life is cancer chemotherapy. Plant and microbial natural products have been the main source of anticancer drugs [12, 13].

With billions of years of evolution and rapid generation times, microbes are an obvious choice for drug discovery, and they have certainly fulfilled their potential. Microbes isolated from soil, in particular, have been an excellent source of diverse anticancer drugs (Table 1). Soil microbes, notably *Bacillus*, the actinomycetes and filamentous fungi, have an amazing capacity to produce biologically active secondary metabolites. Two of the most potent anticancer drugs, doxorubicin hydrochloride and actinomycin D, and a variety of other FDA-approved anticancer drugs, were isolated from soil actinomycetes (Table 1). As summarized in Table 1, most of these compounds were isolated over 30 years ago. The rate at which clinically useful compounds are being discovered in soil microbes by traditional culturing methods has diminished significantly. Fortunately, though, this resource has not been tapped out, far from it in fact. While soil contains the largest population of microbes of any habitat [50], only about 0.3% of soil microbes are cultivable with current techniques [47, 48]. Exhaustive studies of soil microbes repeatedly yield this same small fraction of cultivable microbes, which in turn produce a large number of previously described compounds. The enormous number of as-yet-uncultured microbes in soil is probably the most promising source of new chemical structures on earth.

Diversity of soil microbes

Hammond [18] estimated that there are 1.9×10^6 species of microorganisms. Fewer than 100,000 species of bacteria and fungi are described [6], because, in the case of bacteria at least, such a small percent form visible colonies on laboratory media. For fungi, inadequate sampling is believed to be a contributing factor [21]. Even with media and growth conditions designed to mimic the natural environment, noncultivability remains

R. K. Pettit
Cancer Research Institute and Department of Microbiology, Arizona State University, Tempe, AZ 85287, USA
E-mail: pettitr@asu.edu
Tel.: +1-480-9654907
Fax: +1-480-9658558

Table 1 FDA-approved anticancer drugs from microbes (from reference 9 and the FDA Center for Drug Evaluation and Research). Not listed are the various isomers, derivatives and synthetic analogs

Anticancer drug	Chemical type	Example of producing microorganism	Isolation date
Bleomycin	Glycopeptides	*Streptomyces verticillus*	1975
Daunorubicin (daunomycin)	Anthracycline	*Streptomyces peucitius*	1964
Doxorubicin hydrochloride (adriamycin)	Anthracycline	*Streptomyces peucitius*	1969
Pentostatin	Nucleoside	*Streptomyces antibioticus* (and the fungus *Aspergillus nidulans*)	1974
Streptozocin	Aminoglycoside	*Streptomyces achromogenes*	1959
Mithracin	Anthracycline	*Streptomyces argillaceus*	1953
Mitomycin C	Benzoquinone	*Streptomyces verticillatus*	1964
Actinomycin D	Depsipeptide	*Actinomyces* spp.	1957

a major challenge. Evidence was recently provided [28] that a major reason for bacterial noncultivability is the requirement for cellular signals from organisms in coculture.

Using reassociation kinetics, Torsvik et al. [47] estimated that soil contains more than 4000 species per gram. The genetic diversity of cultivable microbes from soil is about 200-fold less than this [47]. Procaryotic genomic diversity is much higher in soils than in aquatic environments. Relatively nutrient-rich freshwater and estuarine waters have a microbial diversity equivalent to about 160 species/10 liters [38]. In genetic diversity studies of uncultured microbes from marine, hot springs and terrestrial environments, most of the sequences analyzed belong to organisms with no close relatedness to sequenced, cultured representatives [2, 27, 44, 52]. In the last decade, analysis of rRNA gene sequences has resulted in a more than tripling of the number of identifiable bacterial divisions [26, 36]. As of 1998 [26], more than a third of the divisions were characterized only by environmental sequences; members had not been cultured. Culture-independent studies reflect what was shown with culture-dependent studies, that representatives of some bacterial divisions are cosmopolitan in the environment, while others appear restricted to certain habitats [26, 27].

The metagenomics approach to drug discovery

The fact that the discovery rate of new structural classes of antineoplastic agents has significantly declined, and the mounting evidence that the vast majority of soil microbes are unknown, has inspired a new approach to natural product drug discovery. We now have the ability to tap into the diverse chemistry of uncultured microbes by preparing DNA libraries from the soil metagenome (collective genomes of soil organisms) [19], and cloning functional genes in laboratory workhorses such as *Escherichia coli*. In this new approach, DNA fragments are purified from soil and cloned in vectors that can stably maintain large DNA inserts.

The need for high molecular weight DNA

In all cases where the genes encoding biosynthesis of bacterial (and in some cases, fungal) secondary metabolites have been cloned and sequenced, the biosynthetic, regulatory and resistance regions are tightly linked [1, 10, 14, 15, 25, 31, 34, 35]. Clusters of antibiotic synthesis genes in actinomycetes range from approximately 20 kb or less for a relatively simple aromatic polyketide such as actinorhodin, to at least 90–100 kb for complex polyketides such as rifamycin or rapamycin [25]. These linked genes enable cloning of entire secondary metabolite pathways into vectors on contiguous pieces of DNA. Another benefit of the linked biosynthesis and resistance genes for drug discovery is that metabolites that could be toxic to the cloning host would likely be coexpressed with the resistance mechanism.

Isolation and purification of metagenomic DNA from soil

Cloning high molecular weight DNA from soil is particularly difficult because of polyphenolic compounds (e.g. humic acid) that copurify with DNA. Soils high in clay or organic matter pose particular challenges in obtaining pure, high molecular weight DNA [51]. Commonly, polyvinylpyrrolidone or hexadecyl methyl ammonium bromide are incorporated to reduce humic contamination. DNA isolated from soil needs to reflect the existing genetic diversity. Previous work suggests that the direct lysis technique recovers a more representative fraction of the genetic diversity than the cell extraction technique [45]. This may be explained in part by the inability of the cell extraction technique to capture the metagenome, which includes extracellular DNA. The direct lysis method yields at least tenfold greater total DNA recovery than the cell extraction method [45]. The direct lysis method of DNA isolation from soil is based on the initial extraction of extracellular DNA with alkaline buffer, followed by the direct lysis of cells in the soil by chemical or mechanical means, and then quantitative extraction of released DNA.

Chemical disruption uses compounds and conditions such as sodium dodecyl sulfate plus heat, lysozyme, osmotic shock, sodium hydroxide, proteinase K or peptidase. Physical disruption methods include bead mill homogenization, French press, boiling, microwave treatment, freeze-thaw cycles and sonication. DNA extraction is a critical first step, and the choice of more gentle chemical means or harsher mechanical means depends on whether there is a requirement for small (less than about 10 kb) or larger DNA pieces. As discussed above, large inserts are required to capture operons encoding biosynthetic pathways. Furthermore, large inserts reduce the number of clones that need to be screened.

Whether one is attempting to clone small or large fragments, soil DNA extraction always results in coextraction of interfering humic substances [24]; so following DNA extraction with DNA purification is essential [51]. For DNA purification, multiple electrophoresis runs are commonly employed [4, 40]. An optimized purification procedure has recently been reported [37] in which a two-phase agarose gel with polyvinylpyrrolidone in the first half and no polyvinylpyrrolidone in the second half is used. Polyphenolic compounds are removed in the first phase, and the polyvinylpyrrolidone is eliminated in the second phase. In this one-step electrophoresis procedure, DNA shearing effects that occur from repeated electrophoresis are minimized.

Size selection

After purification from soil, metagenomic DNA is partially digested with restriction enzymes. Size selection is typically accomplished with pulsed-field gel electrophoresis [32, 37, 40]. Restriction digestion requires starting with DNA minimally three times longer than the desired insert size [17]. If large inserts (≥100 kb) are desired, DNA of several hundred kilobase is needed. DNA of this length is quite vulnerable to shearing. Thus, modifications are continually being explored that will decrease the number of steps in the lysis procedure and eliminate the need for restriction digestion. Toward this end, Quaiser et al. [37] prepared a soil library in the fosmid vector pEpiFOS (Epicentre, Madison, Wis.) using blunt end cloning. Their library has 2.5×10^4 clones with insert sizes of 30 randomly chosen clones between 32.5 and 43.5 kb. Eliminating restriction digestion also avoids the bias introduced by restriction digests, and naturally decreases the amount of DNA handling.

Choice of vector and host

The choice of vectors for cloning soil metagenomic DNA includes fosmids, cosmids and bacterial artificial chromosomes (BACs). Fosmids and cosmids are suitable for libraries with intermediate size inserts (38–52 kb) [3, 4, 46]. These vectors have much greater cloning efficiencies than BACs. However, the vector of choice for cloning large fragments of DNA is the BAC. BACs are based on the *E. coli* F factor, and can carry DNA inserts >300 kb [42]. BAC vectors containing >100 kb of DNA are stably maintained at one or two copies per cell [42]. Because BACs are low copy number, there is less potential for recombination between DNA fragments and for lethal overexpression of cloned genes [5]. BACs exist as supercoiled plasmids in *E. coli*, permitting easy isolation and manipulation with minimal breakage [42], and allowing transformation into *E. coli* via electroporation [42]. A BAC vector that has been used to create soil metagenomic libraries is pBeloBAC11 [40]. This BAC has two selective markers for cloning purposes, the lacZ gene for color-based recombinant selection and chloramphenicol resistance for transformant selection. In addition, pBeloBAC11 has three unique cloning sites flanked by the T7 and SP6 promoters. NotI cuts on each side of the cloning site, excising the inserts and permitting, for example, size determination by pulsed-field gel electrophoresis. Regulatory genes of pBeloBAC11 include oriS and repE, which mediate the unidirectional replication of the F factor, and parA and parB, which maintain the copy number at one or two per cell.

Evidence that BACs are useful for heterologous gene expression in procaryotes was provided when a BAC library in *E. coli* was constructed from genomic DNA of the Gram-positive bacterium *Bacillus cereus* [39]. Of ten *B. cereus* activities tested, six were found in the *E. coli* library, including lipase, ampicillin resistance, starch hydrolysis and secretion of an orange pigment. The gene expression and protein export machinery in Gram-negative and Gram-positive bacteria differ, so the demonstration of some extracellular biological activities is remarkable.

A widely used *E. coli*strain for BAC cloning is DH10B. This strain contains mutations that block restriction of foreign DNA by endogenous restriction endonucleases, restriction of DNA containing methylated DNA and recombination. Only deoR mutants such as DH10B take up large DNA sufficient for >100 kb insert library construction [41]. Other benefits of cloning in *E. coli* include high transformation efficiency [41], the ability to express heterologous genes from inserted sequences [39, 40], the presence of simple primary metabolites for the production of heterologous secondary metabolites and ease of fermentation scaleup. Regarding anticancer drug discovery, *E. coli* lacks endogenous metabolites that would be active in cancer cell line screens.

Expanding the host machinery and metabolic background would certainly increase the probability of drug discovery from soil libraries. Two obvious host choices are *Streptomyces* and *Bacillus*. Several groups are developing shuttle vectors that allow conjugation between *E. coli* and a second, unrelated, host. Sosio et al. [43] have developed BACs that shuttle between *E. coli*,

where they replicate autonomously, and *Streptomyces*, where they integrate site-specifically into the chromosome. Handelsman's group [20] has developed the vector superBAC1 that shuttles between *E. coli* and *Bacillus*. We are not aware of any reports of soil libraries in eucaryotic hosts. This is surprising, given the even greater chemical versatility, and thus potential for drug discovery, of eucaryotes. Eucaryotic fungi share much of the same core molecular biology, and gene exchange is possible between many of the most common fungal species [49]. A candidate eucaryotic host is the yeast *Schizosaccharomyces pombe*, whose molecular biology is well described. Importantly, *S. pombe* is capable of properly splicing and expressing genes of other species of fungi and plants. As for vectors, yeast artificial chromosomes, like BACs, are capable of propagating several hundred kilobases of DNA, have cloning sites that allow color selection of clones with inserts and have regulatory regions that maintain the copy number at one per cell.

Biological activities detected in soil metagenomic libraries

To date, two categories of biological activity have been detected in soil metagenomic libraries, enzymatic and antibiotic (Table 2). In each of the studies summarized in Table 2, the DNA was extracted by the direct lysis method and expressed inserts were heterologous to the *E. coli* host. In some cases the sequences were completely novel [33, 40]. The following heterologous gene products were isolated: indirubin [32], violacein and deoxyviolacein [4], tyrosine derivatives [3] and turbomycins A and B [16]. Where reported, hit rates for enzymatic activities range from 0.0001% to 0.2% [23, 33, 40]. Hit rates for antibiotic activities range from 0.009% to 0.03% [3, 33, 40]. The low success rates are probably due in large part to barriers in heterologous expression [20]. These barriers could include differences in required transcription and translation machinery and the absence of required substrates for secondary metabolism. The low hit rates for the BAC studies summarized in Table 2 could also be due to low levels of expression from single-copy BACs, making detection of heterologous functions difficult. BACs with inducible copy number, such as pCC1BAC (Epicentre, Madison, Wis.) and SuperBAC1 [20], permit library construction and maintenance at low copy number and detection of heterologous gene expression at high copy number.

Should there be a focus on high-probability taxa?

Cultured actinomycetes produce more than half of all known biologically active microbial products [29]. However, molecular studies suggest that the actual proportion of actinomycetes in soil is not nearly as great as cultivability studies have indicated [30, 44, 52]. While this could be an artifact of the metagenomic library protocols, Liles et al. [30], using a harsh bead-beating method, which yields DNA broadly representative of the soil microbial community [8], had 0–4% actinomycete abundance in a Wisconsin soil sample. A subtropical Australian soil library, prepared using glass beads and analyzed with a streptomycete-specific primer, did not yield any actinomycete sequences [44]. A tundra soil DNA sample extracted with a high-temperature/salt/sodium dodecyl sulfate-based procedure had fewer than 12% of the clones in the Gram-positive group, and none were in the genus *Streptomyces* [52]. It will be very

Table 2 Biological activities detected in soil metagenomic libraries

Reference	Vector	Host	Library size	Biological activity	Detection method
33	Plasmid pBluescript SK^+	*E. coli* DH5α	1.5×10^6	Na^+/H^+ transporters	Complementation of antiporter-deficient strain
23	Plasmid pBluescript SK^+	*E. coli* DH5α	7.3×10^5	Lipase	Orange halos on LB agar + triolein + rhodamine B
			2.9×10^5	Esterase	Clear halos on LB agar + tributyrin
22	Plasmid pBluescript SK^+	*E. coli* DH5α	9.3×10^5	4-Hydroxybutyrate dehydrogenase	Tetrazolium indicator plates with 4-hydroxybutyrate as sole carbon source
40	BAC pBeloBAC11	*E. coli* DH10B	3.6×10^3	Lipase	Clear halos on LB agar + Bacto Lipid
				Amylase	Orange halos on iodine-flooded starch agar
				DNase	Orange halos on methyl green agar
16, 40	BAC pBeloBAC11	*E. coli* DH10B	2.4×10^4	Hemolysin	Clear halos on blood agar overlay
				Antibacterial	Hemolysis and brown pigment on LB agar-purified gene products were antibacterial
32	BAC pBTP2	*E. coli* DH10B	1.2×10^4	Antibacterial	Zones of inhibition on *Bacillus subtilis* agar overlay
4	Cosmid SuperCos I	*E. coli* (no strain info.)	ND	Antibacterial	Blue pigment-purified gene products were antibacterial
3	Cosmid SuperCos I	*E. coli* XL1	7×10^5	Antibacterial	Zones of inhibition on *Bacillus subtilis* agar overlay

interesting to watch the actinomycete story unfold as more soil libraries are created. Results so far suggest that relative to cultured members, the uncultured component of the actinomycetes may make a smaller contribution to human medicine.

Bacillus are also quite biologically active, and appear numerically abundant with standard soil plating methods. When bead-beating was compared to DNA extraction for library preparation, using identical primers and polymerase chain reaction conditions, *Bacillus* made up about 6–14% of the bead-beating library, and $<1\%$ of the BAC library [30]. Thus, the chemical extraction method may result in incomplete lysis of *Bacillus* spores [30].

While industry has focused on actinomycetes, completely unrelated bacteria, the myxobacteria, chromobacteria and cyanobacteria, produce antineoplastic agents, some of which are in clinical development [11, 13]. The FDA-approved anticancer drug pentostatin is produced by both *Streptomyces* and *Aspergillus* [9]. Clearly, advances in metagenomic methods should emphasize the capture and expression of DNA from diverse microbes.

The promise of metagenomic libraries for anticancer drug discovery

Secondary metabolites from cultured microbes have been a major source of clinically useful chemotherapeutics. Assuming that useful natural products will be as frequent in the remaining $>99\%$ of uncultured microbes, soil metagenomic DNA should be a prolific source of novel anticancer agents. The enormous libraries that are being created should prove useful in high-throughput screens for identification of compounds that interact specifically with cancer-specific molecular targets. In addition to their therapeutic promise, the libraries of chemically diverse structures with potentially new mechanisms may lead to the discovery of novel molecular targets. Arguably the biggest technological hurdle that must be overcome before metagenomic libraries become productive sources of new anticancer agents and other pharmaceuticals is the successful cloning of large inserts ($\geq$100 kb). The average insert sizes in BAC soil libraries that have resulted in heterologous gene expression are 37–44.5 kb [32, 40]. The average insert sizes in plasmid soil libraries that have resulted in heterologous gene expression are 5–8 kb [23, 33]. While the only screened biological activities in the soil libraries published to date are antibiotic and enzymatic, one can be sure that efforts are being made in various academic and private laboratories to discover anticancer agents using this approach, and that publications will be forthcoming. The application of technologies developed to express genes from uncultured soil microbes to other areas where microbial culture conditions are extremely difficult to define, for example, symbionts of anticancer-active marine organisms and terrestrial plants, is an exciting prospect.

Acknowledgements This paper is dedicated to Professor G.R. Pettit, who pioneered another productive source of antineoplastic agents, marine natural products.

References

1. Aharonowitz Y, Cohen G, Martin JF (1992) Penicillin and cephalosporin biosynthetic genes: structure, organization, regulation, and evolution. Annu Rev Microbiol 46:461
2. Barns SM, Fundyga RE, Jeffries MW, Pace NR (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc Natl Acad Sci USA 91:1609
3. Brady SF, Clardy J (2000) Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. J Am Chem Soc 122:12903
4. Brady SF, Chao CJ, Handelsman J, Clardy J (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org Lett 3:1981
5. Brosch R, Gordon SV, Billault A, Garnier T, Eiglmeier K, Soravito C, Barrell BG, Cole ST (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect Immun 66:2221
6. Bull AT, Goodfellow M, Slater JH (1992) Biodiversity as a source of innovation in biotechnology. Annu Rev Microbiol 46:219
7. Bull AT, Ward AC, Goodfellow M (2000) Search and discovery strategies for biotechnology: the paradigm shift. Microbiol Mol Biol Rev 64:573
8. Burgmann H, Pesaro M, Widmer F, Zeyer J (2001) A strategy for optimizing quality and quantity of DNA extracted from soil. J Microbiol Methods 45:7
9. Bycroft BW (1988) Dictionary of antibiotics and related substances. Chapman and Hall, University Press, Cambridge, UK
10. Chater KF, Bruton CJ (1985) Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. EMBO J 4:1893
11. Christian MC, Pluda JM, Ho PTC, Arbuck SG, Murgo AJ, Sausville EA (1997) Promising new agents under development by the Division of Cancer Treatment, Diagnosis, and Centers of the National Cancer Institute. Semin Oncol 24:219
12. Clark AM (1996) Natural products as a resource for new drugs. Pharm Res 13:1133
13. Cragg GM, Newman DJ (2000) Antineoplastic agents from natural sources: achievements and future directions. Exp Opin Invest Drugs 9:2783
14. Distler J, Braun C, Ebert A, Piepersberg W (1987) Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: analysis of a central region including the major resistance gene. Mol Gen Genet 208:204
15. Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. Science 252:675
16. Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68:1
17. Gillespie DE, Rondon MR, Handelsman J (2004) Metagenomic libraries from uncultured microorganisms. In: Osborn AM (ed) Molecular microbial ecology. BIOS Scientific Publishers, Oxford (in press)
18. Hammond PM (1995) Described and estimated species numbers: an objective assessment of current knowledge. In: Microbial

diversity and ecosystem function: proceedings of the IUBS/IUMS workshop, Egham, UK. University Press, Cambridge, UK, p 29
19. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245
20. Handelsman J, Liles M, Mann D, Riesenfeld C, Goodman RM (2002) Cloning the metagenome: culture-independent access to the diversity and functions of the uncultivated microbial world. Methods Microbiol 33:241
21. Hawksworth DL (1991) The fungal dimension of biodiversity: magnitude, significance, and conservation. Mycol Res 95:641
22. Henne A, Daniel R, Schmitz RA, Gottschalk G (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
23. Henne A, Schmitz RA, Bömeke M, Gottschalk G, Daniel R (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol 66:3113
24. Holben WE, Jansson JK, Chelm BK, Tiedje JM (1988) DNA probe method for the detection of specific microorganisms in the soil bacterial community. Appl Environ Microbiol 54:703
25. Hopwood DA (1999) Forty years of genetics with *Streptomyces*: from in vivo through in vitro to in silico. Microbiology 145:2183
26. Hugenholtz P, Goebel BM, Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765
27. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998) Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 180:366
28. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating ''uncultivable'' microorganisms in pure culture in a simulated natural environment. Science 296:1127
29. Khosla C (2000) Natural product biosynthesis: a new interface between enzymology and medicine. J Org Chem 65:8127
30. 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
31. Lomovskaya N, Doi-Katayama Y, Filippini S, Nastro C, Fonstein L, Gallo M, Colombo AL, Hutchinson CR (1998) The *Streptomyces peucetius dps*Y and *dnr*X genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. J Bacteriol 180:2379
32. MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3:301
33. Majernik A, Gottschalk G, Daniel R (2001) Screening of environmental DNA libraries for the presence of genes conferring $Na^+(Li^+)/H^+$ antiporter activity on *Escherichia coli*: characterization of the recovered genes and the corresponding gene products. J Bacteriol 183:6645
34. Martín JF (1992) Clusters of genes for the biosynthesis of antibiotics: regulatory genes and overproduction of pharmaceuticals. J Ind Microbiol 9:73
35. Martín JF, Liras P (1989) Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. Annu Rev Microbiol 43:173
36. Pace NR (1997) A molecular view of microbial diversity and the biosphere. Science 276:734
37. Quaiser A, Ochsenreiter T, Klenk H-P, Kletzin A, Treusch AH, Meurer G, Eck J, Sensen CW, Schleper C (2002) First insight into the genome of an uncultivated crenarchaeote from soil. Environ Microbiol 4:603
38. Ritz K, Griffiths BS, Torsvik VL, Hendriksen NB (1997) Analysis of soil and bacterioplankton community DNA by melting profiles and reassociation kinetics. FEMS Microbiol Lett 149:151
39. Rondon MR, Raffel SJ, Goodman RM, Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of *Bacillus cereus*. Proc Natl Acad Sci U S A 96:6451
40. 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
41. Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Res 23:1990
42. Shizuya H, Birren B, Kim U-J, Mancino V, Slepak T, Tachiiri Y, 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 U S A 89:8794
43. Sosio M, Giusino FA, Cappellano C, Bossi E, Puglia AM, Donadio S (2000) Artificial chromosomes for antibiotic-producing actinomycetes. Nat Biotechnol 18:343
44. Stackebrandt E, Liesack W, Goebel BM (1993) Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J 7:232
45. Steffan RJ, Goksøyr J, Bej AK, Atlas RM (1988) Recovery of DNA from soils and sediments. Appl Environ Microbiol 54:2908
46. Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996) Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J Bacteriol 178:591
47. Torsvik V, Goksøyr J, Daae FL (1990) High diversity in DNA of soil bacteria. Appl Environ Microbiol 56:782
48. Torsvik V, Sørheim R, Goksøyr J (1996) Total bacterial diversity in soil and sediment communities—a review. J Ind Microbiol 17:170
49. Van den Hondel, CAMJJ, Punt PJ, van Gorcom RFM (1991) Heterologous gene expression in filamentous fungi. In: Bennett JW, Lasure LL (eds) More gene manipulations in fungi. Academic Press, San Diego, p 396
50. Whitman WB, Coleman DC, Wiebe WJ (1998) Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95:6578
51. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. Appl Environ Microbiol 62:316
52. Zhou J, Davey ME, Figueras JB, Rivkina E, Gillichinsky D, Tiedje JM (1997) Phylogenetic diversity of a bacterial community determined from Siberian tundra soil DNA. Microbiology 143:3913