

University of Groningen

Metagenomics-guided exploration of natural habitats for bacterial chitinases

Cretoiu, Mariana Silvia

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2013

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Cretoiu, M. S. (2013). *Metagenomics-guided exploration of natural habitats for bacterial chitinases*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter III

The metagenomics of plant pathogen-suppressive soils

Jan Dirk van Elsas, Anna Maria Kielak, Mariana Silvia Cretoiu

Chapter in “Handbook of Molecular Microbial Ecology, Volume I: Metagenomics and Complementary Approaches”, First edition. Editor Frans de Bruijn. Wiley - Blackwell, 2011.

Abstract

Soils contain a vast microbial diversity and can be considered as the biggest reservoir of genetic novelty on Earth. This novel genetic information present in the soil microbiota is an excellent genetic resource that awaits our exploitation in order to increase our understanding of soil ecosystem functioning. Only a fraction of the soil microorganisms has been cultured and the difficulties of isolating members of this microbiota hamper our ability to unlock the genetic treasures locked up in them.

Advanced DNA-based methods have become available to circumvent the cultivation dilemma by directly examining genomic DNA derived from the soil microbiota. Thus, we obtain information on the collective soil metagenome. In particular, genes that encode proteins that serve functions of key interest to soil ecology (and biotechnology) can be explored. The metagenome of plant disease-suppressive soils is of special interest given the expected prevalence of antibiotic biosynthetic or otherwise antagonistic gene clusters. In this study, we will draw on our experience on the metagenomics of disease-suppressive soils. We describe the progress achieved in developing tools that are required for metagenomic exploration of suppressive soil and report on some of the results obtained. We also examine the critical challenges that impinge on future applications such as the isolation of biopolymer-attacking enzymes.

Introduction

Soil is known to contain an often extreme microbial diversity per unit mass or volume (Gans *et al.*, 2005). By inference, the soil microbiota offers an excellent angle at novel microbial functions of ecological and industrial interest. For instance, on the basis of cultivation-based approaches, the soil microbiota has been found to harbour a wealth of antibiotic biosynthesis loci. Such functions, in particular cases, may underlie the suppressiveness of soils to plant pathogens (Steinberg *et al.*, 2006). Moreover, the soil microbiota is also known as a goldmine for novel biocatalysts involved in biodegradation processes, including those of human-made polluting compounds (Galvao *et al.*, 2005). However, the soil microbiota as-a-whole has remained largely cryptic due to a phenomenon called the “Great Plate Count Anomaly” (Janssen *et al.*, 2002; van Elsas *et al.*, 2006), which describes the lack of direct culturability of many microorganisms in soil. Our understanding of soil functioning has thus been severely hampered and many key traits of the soil microbiota that are involved in particular population regulatory processes (such as antibiotic production loci and particular enzymatic functions) have remained cryptic. In the light of the currently available high-throughput DNA-based technologies, the potential for examining and exploring the genetic treasures present in the soil microbiota is enormous. Thus, examination of the entire soil metagenome (here defined as the collective genomes of the microorganisms present in a soil sample) has been proposed as a means to address the issue (Rondon *et al.*, 2000). However, there are definite problems in this approach, being of technical as well as fundamental nature (Sjoling *et al.*, 2006). The fundamental caveats of soil metagenomics revolve around the relative ease to captivate the dominant soil microbiota versus the difficulty to access the so-called “rare” biosphere. Rank-abundance curves constructed for the soil microbiota have often demonstrated this rare biosphere to consist of an extremely long tail of ever-rarer species. It is a fact of metagenomic life that, without *a priori* measures to remediate this, soil-based metagenomes are almost always biased towards the dominant community members.

A European research project denoted Metacontrol, which was executed in the early days of soil metagenomics, i.e. between 2002 and 2007, aimed to unravel the antagonistic capacities locked up in the microbiota of disease-suppressive soils. The basic idea was to find clues with respect to the involvement of such traits in the suppression of plant pathogens as well as to explore these for application purposes. The project has yielded a wealth of methodological advances and has given glimpses of the antagonistic potential of the soils studied (van Elsas *et al.*, 2008b). However, a full understanding of the antagonistic diversity in suppressive soils against plant pathogens is still missing, and this is largely due to the astounding diversity found in the soil microbiota at this functional level. We here describe the major advances that have been achieved in metagenomic studies of disease-suppressive soils and address the major challenges that still lie ahead of us.

Disease-suppressive soils

Disease-suppressive soils are defined by their ability to restrict the activity and/or survival of plant-pathogenic microorganisms. Some soils possess a natural ability to suppress plant pathogens (Borneman & Becker, 2007; Steinberg *et al.*, 2007), whereas in other soils disease suppressiveness can be the result of soil management practices, such as monocropping. The key to plant disease suppression often lies in the soil microbiota, that is, the pathogen-suppressive microbiota of any kind or composition that is present. This microbiota may be involved in competition for essential substrates the plant pathogen grows on (leading to niche exclusion), or it may be directly antagonistic to the pathogen. In the latter case, the *in situ* production of antibiotics that inhibit or kill the pathogen or of enzymes directly affecting the pathogen may be involved. A key example of the latter mechanism is the production of fungal pathogen-attacking chitinases or of competitive proteases, in disease-suppressive soils.

Chitinolytic activity and disease suppressive soils

Chitinases produced by soil microorganisms can also be involved in the suppression of plant disease, in this case caused by fungi that have chitinous cell walls. The level of disease suppressiveness can even be raised by adding chitin to soils (Mankau & Das, 1969; Spiegel *et al.*, 1989). Several studies have reported (based on the measurement of activity of nematodes and fungi) that the induction of soil suppressiveness by chitin amendment is a biotic process (Chernin *et al.*, 1995; Kamil *et al.*, 2007). However, the mechanisms by which soils inhibit plant disease via chitin has not been completely elucidated. The exploration of the diversity of chitinases produced by the soil microbiota is a subject of current research, especially with respect to suppressiveness and to the possibility of manipulating this property (Downing & Thomson, 2000; Kobayashi *et al.*, 2002). Also, chitinolytic bacteria like *Enterobacter agglomerans*, *Serratia marcescens*, *Pseudomonas fluorescens*, *Stenotrophomonas maltophilia* and *Bacillus subtilis* have been used as biological control agents of fungal or nematodal plant disease agents (Downing & Thomson, 2000; Zhang *et al.*, 2000; Kotan *et al.*, 2009; Kobayashi *et al.*, 2002). Moreover, fungi of the genera *Gliocladium* and *Trichoderma* have also been found to produce chitinolytic enzymes with protective roles for plants (di Pietro *et al.*, 1993; Elad *et al.*, 1982). So far, the pathway of chitinase activity has been partially elucidated with respect to their protective role for plants (Cleveland *et al.*, 2004).

Current insight in disease suppressiveness of soils indicates that, in most of the cases, the phenomenon is complex. That is, various mechanisms may be involved. Suppression of a particular pathogen may include, besides the production of chitinases, efficient rhizosphere colonization leading to niche exclusion of the pathogen, and the production of one to several antibiotics, as well as of different proteases.

Exploration of disease-suppressive soils

A collaborative European project with acronym Metacontrol (2002-2007) had as its stated objective to examine selected phytopathogen-suppressive soils for their antagonistic potential. A range of assessments of the nature of different suppressive soils were obtained (Adesina *et al.*, 2007; Bertrand *et al.*, 2005; Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; Hjort *et al.*, 2007; Lefevre *et al.*, 2008; Nalin *et al.*, 2004; van Elsas *et al.*, 2008a). The assumption was that the microbiota of suppressive soils would provide reservoirs of genetic loci involved in *in situ* antibiosis or antagonism. A focus was placed on genes for phytopathogen-suppressive polyketide antibiotics and chitinases. As shown in Table 1, four soils that were suppressive to varying phytopathogens were identified in the Netherlands [W, *Rhizoctonia solani* AG3 (Garbeva *et al.*, 2004; 2006)], Sweden [U, *Plasmodiophora brassicae*], France [C, *Fusarium*] and the UK [Wy, *Fusarium*]. Metagenomic libraries were constructed from these soils plus one control soil, M (Table 1), and screened for the occurrence of antibiotic and antagonistic functions (Adesina *et al.*, 2007; Bertrand *et al.*, 2005; Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; Hjort *et al.*, 2007; Lefevre *et al.*, 2008; Nalin *et al.*, 2004; van Elsas *et al.*, 2008a,b). In addition, a range of methodologies were developed that facilitated the preparation and exploration of the resulting libraries (Bertrand *et al.*, 2005; Ginolhac *et al.*, 2004; Hjort *et al.*, 2007; Sjöling *et al.*, 2006).

The exploration of the antagonistic potential of disease-suppressive soils by using a metagenomics-centered approach appeared straightforward at the onset of the work, however it turned out to be utterly complex (van Elsas *et al.*, 2008a,b). A major issue was the prior estimation of target gene abundance, which was felt to be a strong determinant of the hit rate in the final metagenomic libraries. In the absence of a clear notion of the nature of the antagonistic compounds produced and genes involved, such an *a priori* assumption was very difficult to make. Other issues were of technical nature and revolved around the uncertainties and technicalities with respect to soil DNA extraction and cloning as well as the positive detection of the active compounds. In the following, we discuss the technology developed and the choices that had to be made prior to each analytical step with respect to: (i) the soil DNA extraction methodology, (ii) the potential to “bias” the soil community or DNA, (iii) the suitability of the vector/host system for the objectives, (iv) the optimal screening procedure, and (v) the final analysis.

Table 1. Soil metagenomic libraries constructed and their characteristics*.

Soil	Library vector / no. of clones	Screening (functional / genetic)	No. of positive clones	Remarks
W - Wildekamp grassland, suppressive to <i>Rhizoctonia solani</i> AG3	Fosmid / 15,000	Functional: antagonism against <i>Rhizoctonia solani</i> AG3 and <i>Bacillus subtilis</i> . Genetic: use of soil-generated PKS1** probe	7	Combined functional / genetic (PKS1) screening: 7 clones. Five confirmed as PKS1-positive clones. Three completely sequenced, one insert showing high similarity with <i>Acidobacterium</i> sp.
Wy - Wytham grassland & <i>Fusarium</i> -suppressive agricultural soil	Fosmid / 100,000	Functional: antagonism against <i>Fusarium</i> sp. (agar plate based dual-culture assay).	13 (grassland) 2 (agricultural soil). Each clone distinct.	Average insert size 35.6 kb. Grassland effective source of clones (high diversity). Agricultural soil low diversity and limited functional traits. End-sequencing / subcloning: mostly unidentified ORFs. Efficacy of clones lower than strains isolated from same source.
C - Chateaufrenard <i>Fusarium</i> -suppressive soil	Fosmid/ 51,000; BAC/ 60,000	Genetic / functional screening	22	Combined functional and genetic screening. Functional screening: <i>Fusarium</i> spore generation / hyphal production, <i>Aspergillus nidulans</i> growth, <i>Hebeloma cylindrosporum</i> hyphal generation. Genetic screening: sequencing PKS positive clones.
U - Uppsala <i>Plasmodiophora brassicae</i> -suppressive soil	Fosmid / 8,000	Functional: antagonism against <i>Pythium ultimum</i> Genetic: chitinase genes and 16S rRNA gene	4	Selection of <i>Streptomyces mutomycini</i> , <i>Kitosatospora</i> , <i>Lentzea</i> , <i>Oerskovia</i> revealed by fingerprinting. <i>S. mutomycini</i> and <i>S. clavifer</i> prevalent in library. Chitinase genes from soil, library and isolates. Cluster prevailing in soil not in library; library cluster not found in soil.
Montrond (control) soil	Fosmid / 60,000	Genetic / functional screening	39	Thirty-nine novel PKS1 positive clones, most with supernatants showing antimicrobial activity.

*Modified from van Elsland et al, 2008b; **PKS1: polyketide synthesis operon for type-I polyketides.

Soil DNA extraction and processing

For reliable library preparation, metagenomic soil DNA - which accurately represents the genetic make-up of the soil microbiota - is required in representative quantity (Bertrand *et al.*, 2005, Inceoglu *et al.*, 2010; van Elsas *et al.*, 2008a,b). In addition, the DNA needs to be of sufficient quality with respect to purity, integrity and fragment length in order to be suitable for cloning into a suitable vector (Bertrand *et al.*, 2005). A minimal size of 40 kb will increase the chance that entire pathways, e.g. those involved in the biosynthesis of polyketide antibiotics, can be cloned (Ginolhac *et al.*, 2004; 2005; van Elsas *et al.*, 2008a).

In several laboratories, advanced methodology that allowed to produce pure high molecular weight (HMW) DNA from soil was developed (Bertrand *et al.*, 2005; Ginolhac *et al.*, 2004; Hjort *et al.*, 2007; Lefevre *et al.*, 2008; van Elsas *et al.*, 2008a). An efficient approach consisted of the extraction of cells from soil followed by gentle DNA extraction and purification using pulsed-field gel electrophoresis (Bertrand *et al.*, 2005; van Elsas *et al.*, 2008a) Cushion (Percoll and/or Nycodenz) pre-separation of cells from soil was also tested as a pre-step for subsequent isolation of the HMW soil metagenomic DNA. Moreover, the microbial growth status in soil was assessed as an important determinant of the chemical quality of the extracted DNA. The quality could even be boosted by incubation with growth substrates such as glycerol (Bertrand *et al.*, 2005; van Elsas *et al.*, 2008a). Typically, the approach produced adequate HMW soil DNA, often with size > 60 - 100 kb (Bertrand *et al.*, 2005; van Elsas *et al.*, 2008a). It was also found that high amounts of cells, minimally ca. 10^{11} , were required to yield sufficient DNA for efficient library construction (Hjort *et al.*, 2007). As soils often contain on the order of 10^8 to 10^{10} cells per g, this finding sets a standard for the construction of soil metagenomic libraries. However, in spite of the improved soil DNA extractions and subsequent metagenomic library constructions, the hit rates of target genes were found to be low. Theoretically, assuming an incidence of target genes of 1% (that is, occurring once in every 100 bacterial genomes - average genome size of 4-5 Mbp), the constructed metagenome library would need to contain at least about 57,000 clones with 40 kb inserts to be able to find - with 99% probability - a single copy (Leveau, 2007). This phenomenon has been likened to “looking for a needle in the haystack” (Kowalchuk *et al.*, 2007) and strongly hampers the efficiency of metagenomics for bioexploration. Deliberate biasing of the habitat by applying pre-enrichment techniques has been suggested as a useful strategy that may boost hit rates (van Elsas *et al.*, 2008b).

Metagenomic libraries - production and screening

Clone libraries for four disease-suppressive soils (Table 1), each one consisting of approximately 6,000 to 60,000 clones, were constructed in *Escherichia coli* (van Elsas *et al.*, 2008b). Both large insert size vectors, such as bacterial artificial chromosomes (BACs), that allow cloning of inserts up to 200 kb, and fosmids (that allow insertion of 35 - 45 kb

fragments) were used. BAC vectors enable the cloning of complex large operons and facilitate the analysis of a gene/operon within its original genomic context. In contrast, fosmids are able to accommodate smaller inserts and thereby only allow the cloning of smaller operons. Using a fosmid vector (such as the Epicentre pCC1FOS system) allows for the positive selection of vectors that have acquired inserts (Bertrand *et al.*, 2005; Ginolhac *et al.*, 2005; Nalin *et al.*, 2004; Sjöling *et al.*, 2006; van Elsas *et al.*, 2008a). Three libraries were based on fosmid vectors, the reason being the ease of obtaining appropriately-sized libraries. One library, for the M soil, was constructed in a BAC vector (Courtois *et al.*, 2003; Ginolhac *et al.*, 2004). The latter vector also contained a replicon that was compatible with a *Streptomyces* host, which allowed shuttling between the *E. coli* and *Streptomyces* metagenomic hosts. Consequently, the probability of heterologous gene expression was enhanced for the clones obtained in this library (Courtois *et al.*, 2003; Ginolhac *et al.*, 2004).

Given the fact that soil metagenomic libraries are based on the random insertion of clonable DNA fragments into vectors, such libraries stochastically contain the genetic material of all genomes that were extracted from the soil microbiota and entered the DNA pool. Assuming that the prevalence of antagonistic functions across all microbial genomes in soil is low (ranging from 0.1 to 10%) and that these genes/operons may be 1-200 kb in size (over a 4-5 Mb average soil genome), soil DNA based metagenomes may contain only few clones that carry genes/operons of interest. Furthermore, there may be potential constraints to efficient gene expression in the metagenomic host strain. Hence, library screening is often a tedious task. For the metagenomes of the four disease-suppressive soils, functional as well as molecular screenings were employed in order to uncover antagonistic functions (Bertrand *et al.*, 2005; Ginolhac *et al.*, 2004; Nalin *et al.*, 2004; Sjöling *et al.*, 2006; van Elsas *et al.*, 2008a) and, expectedly, rather low numbers of phytopathogen-suppressive clones were found (Bertrand *et al.*, 2005; Ginolhac *et al.*, 2004; van Elsas *et al.*, 2008a).

Functional screening

Functional screenings of the libraries were performed using high-throughput dual-culture assays. These assays allow target phytopathogenic organisms to grow over metagenomic library clones arrayed on large Petri dishes. Scoring during and following growth was for irregularities / inhibitions in growth of the target organism (Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; van Elsas *et al.*, 2008a). This experimental set-up led to the detection of positive clones (up to 48 per library), amounting to < 0.05% of positives for all libraries. Such low numbers can be attributed either to a rare occurrence of target genes/operons in the clones, to a lack of expression of the genes/operons in the host used or to the required molecular machineries being significantly larger than the vector inserts. The latter fact has indeed been reported for many polyketide production loci. Other factors that potentially impede the detection of function of the target genes could relate to a lack of

adequate signals, such as found in expression systems that are controlled by quorum sensing (a cellular communication mechanism commonly found in bacteria). Shuttling from *E. coli* to *Streptomyces* as the metagenomic host facilitated the expression of an antibiotic (amphotericin) production locus (Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; van Elsas *et al.*, 2008b), as indicated by high activity in resulting clones against target fungi.

In the light of the foregoing, the low numbers of functionally-positive clones did not come as a surprise. Corroborating this result, a screening of forest soil libraries for antifungal traits also yielded a low hit rate, i.e. one positive signal among 113,700 fosmid clones examined was found (Chung *et al.*, 2008). We conclude that substantial methodological improvements are required to boost the hit rates in explorations of soils for antagonistic function. Ways forward are given below.

Genetic (molecular) screening

The libraries obtained from disease-suppressive soils were also screened using molecular tools, such as hybridization and PCR based methods (van Elsas *et al.*, 2008b). In this case, success in detecting novel operons, such as those involved in polyketide biosynthesis, was dependent on the application of deliberate degeneracy in the probes and primers used (Courtois *et al.*, 2003; Ginolhac *et al.*, 2004). The rationale was that, using this strategy, the screening would not be restricted to exactly known genes, enabling a broader range of positive hits within the metagenomic library. The method facilitated the identification of target genes that were sufficiently similar to the mixed query sequences generated from the same soil. Using the total soil community DNA as a target, we thus amplified the KS gene of the polyketide biosynthesis operon PKS1 from soil DNA with degenerated primers. The amplicons obtained were used as probes to detect PKS1 sequences in the library (Ginolhac *et al.*, 2004; Nalin *et al.*, 2004; van Elsas *et al.*, 2008a). The approach yielded a total of seven positive clones in the W soil library, of which the majority contained genes that were likely involved in the biosynthesis of novel polyketides. This was confirmed by end-sequencing of the clones (van Elsas *et al.*, 2008a). In addition, the roughly 60,000 M soil clones were divided into pools, which were subsequently used as templates for PKS-based PCR screenings. This yielded over 100 positive pools (0.22% hit rate; Ginolhac *et al.*, 2004). The amplicons were then sequenced to check for redundancies and for known PKS sequences. In total, 39 unique PKS sequences were thus found, which all represented promising novel PKS biosynthesis operons (van Elsas *et al.*, 2008b). The positive clones, identified using colony hybridization with relevant probes, were then tested, following shuttling into *Streptomyces*, for antagonistic activities. *Bacillus subtilis* 1A72, *Staphylococcus aureus* 21, *Enterococcus faecalis* 40, *Escherichia coli* 9, *Pseudomonas aeruginosa* 39, *Fusarium oxysporum* LNPV, *Aspergillus fumigatus* Gasp 4707 and *Neurospora crassa* HK were used. The clones exhibited 56% antimicrobial activity against at least *B. subtilis*, 13% against *S. aureus*, 4% against *E. faecalis* and <1% partial inhibition of growth of *Neurospora crassa* mycelium (van Elsas *et al.*, 2008b). We

conclude that the screening procedures allowed quick access to novel PKS sequences from soil. Further testing in respect of applicability is obviously required prior to any large-scale production and use in agriculture or medicine (van Elsas *et al.*, 2008b).

Hit rates of target genes via pre-enrichment of soil or samples

In the metagenomics studies performed on disease-suppressive soils, direct (unselective) approaches were used to examine the soils for anti-phytopathogen functions (Adesina *et al.*, 2007; Bertrand *et al.*, 2005; Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; Hjort *et al.*, 2007; Lefevre *et al.*, 2008; Nalin *et al.*, 2004; van Elsas *et al.*, 2008b). In these approaches, the possibility of applying positive growth selection was dismissed, the reason being that a direct unbiased assessment of the antagonistic potential of the soil microbiota was felt to be required. Moreover, most of the selected target genes/operons were assumed not to offer an *a priori* growth advantage in the host used for cloning, even in the presence of the target phytopathogens (Ginolhac *et al.*, 2004; van Elsas *et al.*, 2008a). However, targeted approaches were also developed to enhance the hit rates of specific targets in the microbial communities. For instance, the U soil was pre-treated with chitin in order to enhance the abundance of particular chitinase producers (Hjort *et al.*, 2007; Sjöling *et al.*, 2006). Starting from the premise that the most successful antagonists in soil are those that are active *in situ*, attempts were also made to enrich the metabolically-active bacterial cell fractions with the help of flow cytometric cell sorting (Hjort *et al.*, 2007; van Elsas *et al.*, 2008a). Metabolically-active cells were indeed successfully sorted from the soil (van Elsas *et al.*, 2008a,b). However, as a result of the limited flow rate of the cell sorter used, throughput was too limited (10^6 cells/hour) to yield sufficient biomass for library construction (van Elsas *et al.*, 2008a), and advanced machines with higher flow rates were deemed necessary.

The way ahead –improvement of hit rates

What are the challenges for the further metagenomics exploration of suppressive and other soils? Let us assume that most questions revolve around enhancing the efficiency of the metagenomics-based exploitation of the soil microbiota for beneficial traits. This translates into enhancing the hit rates of target genes and may imply the application of a deliberate bias to favor target organisms and genes/operons in the starting material. In addition, strong improvements should be made in all steps in soil metagenomics, i.e. in DNA extraction methodology, cloning and screening methods, all aimed at increasing the throughput of metagenomics.

Deliberate bias in sampled communities

Deliberate manipulation of microbial communities from soil offers unique possibilities to enhance metagenomics hit rates. For instance, prior growth selection can be applied, as outlined before. Here, an intelligent selection of growth conditions will guide the bias. In this approach, fluorescence-assisted cell sorting can be applied, not only to sort the metabolically-active cell fractions, but also to obtain particular interesting fractions of the community. For instance, the high-G+C% Gram-positive bacteria (in which antibiotic production loci are abundantly present) can be selected following staining with specific fluorescent probes. Another promise is offered by the use of stable isotopes. Stable isotope probing (SIP) introduces ^{13}C -labelled substrates into soil communities. Members of such communities can take up the ^{13}C and incorporate it in their cellular macromolecules including DNA. The resulting “heavy” DNA is separated from ^{12}C -DNA by ultracentrifugation and sequenced, thus identifying the organisms that captured the substrate (Radajewski *et al.*, 2003). This approach can be coupled to soil metagenomics studies (Dumont *et al.*, 2006), resulting in the identification of a complete methane monooxygenase operon, allowing insight in this process in soil. Application in the detection of beneficials in suppressive soils will depend on unique substrates that are used by particular phytopathogens and their competitors, thus allowing to identify the latter. As an example, by tracking the fate of ^{13}C -labelled CO_2 fixed by plants into the soil microbiota, key data on plant-responsive microorganisms – which often produce antibiotics as secondary metabolites - can be obtained (Ostle *et al.*, 2003). The application of SIP using other organic substrates bears great potential in future explorative metagenomic studies in which organisms with particular ecological roles are the targets.

Searching for improved metagenomic library hosts

Working with *E. coli* as the metagenomics host has clear advantages with respect to the ease of the laboratory work and the experience gained with it over many years. However, the use of *E. coli* is limited with regard to the screening of phenotypes from the soil metagenome, as *E. coli* is not a typical soil organism. The main restriction arises from the fact that particular promoters and associated factors required for the expression of inserted genes may be poorly recognized in this host. Moreover, essential post-translational processing and/or transport functions may be lacking in this host. Rondon *et al.* (1999) showed that only about 30% of *Bacillus* traits could be expressed in *E. coli*, which indicates that *E. coli* is - at best - a suboptimal host for the heterologous expression of genes from such typical soil bacteria. Bear in mind that soil microbial communities are often dominated by just five bacterial phyla: *Alpha*-, *Beta*-, *Gamma-Proteobacteria*, *Acidobacteria* and *Actinobacteria* (Fierer *et al.*, 2007). Bacteria belonging to several of these phyla are appealing hosts for use in functional metagenomic studies of soil habitats. Thus, efforts are ongoing to develop alternative hosts preferably within these bacterial groups (van Elsas *et al.*, 2008b). Recently, six novel bacterial hosts belonging to the phylum *Proteobacteria*

(*Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Escherichia coli*, *Pseudomonas putida* and *Ralstonia metallidurans*) were tested for their performance in functional metagenomic screening of a library constructed using broad-host-range cloning vectors (Craig *et al.*, 2010). This work, as well as ours, supports the premise that working with several hosts instead of just *E. coli* will allow to strongly boost the metagenomics hit rates as well as expanding functional and genetic diversity of target traits. However, the study of Craig *et al.* (2010) also revealed that not all hosts perform equally well in phenotypic screenings. It remains possible that some surrogate expression hosts are better suited for the expression of foreign genetic material than others. Furthermore, each host can be differentially sensitive to toxic compounds that are produced from inserted genes. Such host clones will disappear from the library. Obviously, these constraints are true for any host strain selected for soil metagenomics.

The metagenomic library vector

The data discussed in this chapter and elsewhere (van Elsas *et al.*, 2008b) affirm that critical evaluation of the host/vector system to be used in soil metagenomics is required (Sjoling *et al.*, 2006). For *E. coli* and for some other hosts, three types of vectors, i.e. small-, medium- and large-insert size, are available. Small-insert-size vectors, that primarily permit screening for single gene-encoded functions, are of use in shotgun sequencing approaches, allowing construction of libraries from mechanically-sheared DNA. Such an approach was used for the detection of small open reading frames (ORFs) derived from uncultured prokaryotes from sediment (Wilkinson *et al.*, 2002). On the other hand, both fosmid and BAC vectors allow incorporation of larger fragments and even intact operons within their genomic context. Although, this provides a better handle at gene expression from complex operons, the fact that pure HMW soil DNA is required in high amounts for efficient cloning into BAC vectors makes these less suitable for routine cloning efforts (such as required for high-throughput setups). The identification of novel activities requires, as mentioned in the foregoing, successful transcription and translation systems. This is obviously connected with expansion of the range of bacterial hosts that allow capturing of additional expression capabilities. Preferably, complete metagenomic libraries should be transferable to different alternative hosts, which will require the development of new shuttle vectors. The recently described vector pRS44 (Aakvik *et al.*, 2009) may serve as an example of such a broad-host-range vector system. This vector can be efficiently transferred to numerous hosts by conjugation, which is spurred by the plasmid RK2 replication origin. In *E. coli*, this plasmid replicates via its plasmid F origin.

Improved screening methods

Efficient selection of clones of interest still remains a critical point in any soil metagenomics approach. The possibility of missing a target due to problems with expression of genes in the metagenomic host plus the sizes of libraries forces the

development of novel genetic screening methods. For example, Demaneche *et al.* (2009) proposed the use of multiple probes in order to reduce the pools containing the potentially interesting clones. The simultaneous use of pooled probes increases the probability of finding clones of interest. In their study, pooled probes targeting genes associated with a range of functions (e.g. genes for antibiotic resistance, denitrification and dehalogenation) were used for the library screening. The pooled-probes approach proved to be useful for rapid library screening. Another method to enhance the detection frequencies of genes of interest lies in the use of “heavy” DNA from SIP experiments, as discussed in the foregoing. The use of stable isotope labeled substrate in enrichment experiments thus increases the chances of discovery of novel enzymes from the environment (e.g. Knietsch *et al.*, 2003; Woo *et al.*, 2009).

Direct pyrosequencing

Currently, direct pyrosequencing technology is favored as a technology that provides a quick insight into the gene repertoire of a particular soil sample. Although this method produces only short reads (currently of about 450 bp), it compensates this limitation by its speed, simplicity and especially its tremendous output. Thus, this technology, next to being useful for studying the microbial diversity and community composition based on the 16S rRNA gene (Jones *et al.*, 2009; Roesch *et al.*, 2007; Spain *et al.*, 2009), is also a good alternative for studying functional genes and assisting in probe design for recovery of the complete gene sequences. The most critical part in this approach is the design of a proper primer pair. On the one hand, it should cover the diversity of the gene family within the community. On the other hand, primers should not have a too high degree of degeneracy so as not to lose specificity. Also, the amplicon cannot be too long to not reduce the emulsion PCR efficiency.

Conclusions

For the foregoing, it is thus apparent that major challenges still lie ahead of us. Granted, several interesting novel biological functions have already been uncovered in the microbiota of the suppressive soils studied, but this may be considered to represent the tip-of-the-iceberg of the diversity that is out there. The (partial) biosynthetic machineries likely involved in the production of novel polyketide antibiotics, e.g. a leinamycin-like antibiotic, as well as other polyketides (Bertrand *et al.*, 2005; Courtois *et al.*, 2003; Ginolhac *et al.*, 2004; Nalin *et al.*, 2004; van Elsas *et al.*, 2008a) were promising, but the work is in need of a follow-up. These discoveries were also plagued by the low hit rates of promising antibiotic biosynthesis clones, even for the disease-suppressive soils. We thus assume that, in spite of the successes, a major part of the extant antibiotic biosynthesis machineries may have been missed in the metagenomic screens for reasons explained in the foregoing. Thus, a current ‘rule-of-thumb’, that the search for *nonhousekeeping* functions in soil metagenomes can be compared to looking for a needle-in-a-haystack (Kowalchuk *et al.*,

2007), was affirmed. The characteristics of the soil microbiota, exemplified by typical rank-abundance distribution of particular microbial populations, dictate this outcome and incite the application of creative tricks and tools to overcome these limitations. Such tricks and tools in all cases would need to enhance the “visibility” of target genes, allowing us to detect and get hold of these. To mention a few, soil DNA extraction could be geared towards community members or metagenome fractions that most likely contain the target genes. For instance, total plasmid DNA, the “metamobilome”, can be extracted when novel biodegradation, metal or antibiotic resistance genes (frequently found on plasmids) are targeted. Moreover, specific fractions of the chromosomal DNA pool, e.g. high G+C content DNA in the case of actinobacterial genes, may be targeted. Improvements in subsequent screens may also be cogitated, e.g. by using a gene-centered approach (Iwai *et al.*, 2010) or improving high-throughput formats of increased accuracy. Further, the screening (gene expression) data should be increasingly linked to high-throughput sequencing.

At the level of the sample, improvements may build on the invention of new positive selection strategies for desired traits, either based on growth or on overcoming resistance. Another tool that will foster explorative metagenomics is the pre-screening of habitats in respect of the incidence, abundance and expression of target genes. So-called global-scale gene mapping (GGM – analogous to the concept of environmental gene tagging) describes habitats in terms of gene abundance and/or expression (Tringe *et al.*, 2005). GGM can compare microbial gene pools across soils and provide a global perspective on target gene prevalence. For instance, PKS1-type polyketide biosynthetic loci are more prevalent in soil than in whale carcass, acid mine drainage or Sargasso sea metagenomes (Tringe *et al.*, 2005). GGM thus allows to predict hit rates of target genes.

These forthcoming advances will boost our capacities to finally come to grips with the astounding soil microbial diversity and harness it to our advantage. Guidance by GGM will be an important asset in the progress. The improved or finetuned soil metagenomics approaches will enable us to (1) mine soil for genes / pathways of interest to biotechnological applications, (2) decipher the identity and function of as-yet- uncultured microorganisms, and (3) obtain a characterization of soil with regard to antagonistic function, diversity and genetic complement. The quickly-increasing throughput of (pyro)sequencing technologies will also assist us in the rapid assessment of the prevalence of target genes, shedding increasing light on the soil genetic reservoir and potential for biotechnological exploration and application.

Acknowledgements

We thank all partners of the Metacontrol project for their collaboration: Kornelia Smalla (Julius Kuhn Institut, Braunschweig, Germany), Pascal Simonet and Tim Vogel (Ecole Superieure de Lyon, Lyon, France), Janet Jansson (LBL, Berkeley, USA), Mark Bailey (CEH, Oxford, UK), Sara Sjoling (Sodertorns Universitet, Stockholm, Sweden), Leo van Overbeek (Plant Research International, Wageningen, The Netherlands) and Renaud Nalin (LibraGen, Toulouse, France).

