Identification and Characterization of Microbial Key Functions in Soils of the German Biodiversity Exploratories Representing Different Land Use and Management Types

D i s s e r t a t i o n

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

vorgelegt von

Heiko Nacke

aus Bad Pyrmont

Göttingen 2011

D7

Referent:	PD Dr. Rolf Daniel
Korreferent:	Prof. Dr. Wolfgang Liebl
Tag der mündlichen Prüfung:	20.10.2011

Contents

List of publications

Will, C., **H. Nacke**, A. Thürmer, and R. Daniel. 2010. Schlaglicht Biodiversität: Charakterisierung und Nutzung der bakteriellen Diversität in Bodenmetagenomen. GenomXpress 1.10:9-11.

Will, C., A. Thürmer, A. Wollherr, **H. Nacke**, N. Herold, M. Schrumpf, J. Gutknecht, T. Wubet, F. Buscot, and R. Daniel. 2010. Horizon-specific bacterial community composition of German grassland soils as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol 76:6751-6759.

Nacke, H., A. Thürmer, A. Wollherr, C. Will, L. Hodac, N. Herold, I. Schöning, M. Schrumpf, and R. Daniel. 2011. Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils. PLoS ONE 6:e17000.

Nacke, H., C. Will, S. Herzog, B. Nowka, M. Engelhaupt, and R. Daniel. 2011. Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. FEMS Microbiol Ecol 78:188-201.

Nacke, H., M. Engelhaupt, S. Brady, C. Fischer, J. Tautzt, and R. Daniel. 2012. Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. Biotechnol Lett 34:663-675

A Introduction

1 Insights into the diversity and structure of soil microbial communities achieved by metagenomic and small subunit ribosomal RNA analyses

Soils are important ecosystems for global nutrient cycling, and exhibit a high physico-chemical complexity. Microbial communities colonizing the habitat soil are considered to be enormously diverse, with probably the highest level of prokaryotic diversity of any environment (Delmont et al. 2011). One gram of soil has been reported to contain approximately 1,000 Gbp of microbial genome sequences (Vogel et al. 2009), including an estimated 2,000 to 18,000 prokaryotic genomes (Daniel 2005). Different microorganisms such as members of the prokaryotic phyla *Proteobacteria*, *Actinobacteria*, *Acidobacteria*, and *Verrucomicrobia* have been isolated from soil using culture-based approaches (Janssen et al. 2002; Sait et al. 2002). *Acidobacteria* appear to be among the most abundant bacterial phyla in diverse soils, representing 5 to 46% of soil bacterial communities (Janssen 2006), but despite improvements of culturing techniques the available isolates poorly represent the known diversity of *Acidobacteria* (Jones et al. 2009). It has been reported that culture-based discovered taxa do not necessarily reflect the dominant taxa in an environmental habitat (Amann and Ludwig 2000; Griffiths et al. 2011). Currently, less than 1% of microbial species are considered to grow under laboratory conditions (Singh et al. 2009). Thus, culture-based approaches provide only a first glimpse into the soil microbial diversity. At the end of the 20th century, invention of culture-independent metagenomic approaches complemented traditional culture-based techniques (Handelsman et al. 1998; Rondon et al. 1999). Theoretically, metagenomic approaches permit access to the collective nucleic acids of all indigenous microorganisms present in an environmental sample, referred to as the metagenome (Handelsman et al. 1998).

Metagenomic DNA has been extracted from various soils such as Arctic and Antarctic soil (Chu et al. 2010; Teixeira et al. 2010), prairie soil (Elshahed et al. 2008), wetland soil (Hartman et al. 2008), and forest soil (Roesch et al. 2007; Uroz et al. 2010) to enable the analysis of diversity and structure of soil microbial communities. In particular, the 16S rRNA gene has proven to be a suitable marker for the taxonomic assessment of prokaryotic microorganisms in environmental samples (Rappé and Giovanni 2003; Si-

mon and Daniel 2011). To allow the analysis of 16S rRNA genes present in soil metagenomes, traditional molecular ecological approaches including fingerprinting methods (e.g., denaturing gradient gel electrophoresis and terminal fragment length polymorphism analysis) and Sanger sequencing of 16S rRNA gene clone libraries have been successfully applied (Dunbar et al. 1999; McCaig et al. 2001; Brons and Elsas 2008; Jesus et al. 2009). These methods offered first insights into the uncultured microbial majority and expanded the knowledge on microbial diversity in soil. The employment of fingerprinting methods allows rapid bacterial community comparisons of a larger number of soil samples but only provide coarse phylogenetic information. Sanger sequencing strategies targeting the 16S rRNA gene led to the identification of numerous new uncultured species in soil samples (Hackl et al. 2004). Janssen (2006) performed a meta-analysis of 32 16S rRNA gene libraries from a variety of soils. The phyla *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Bacteroidetes*, *Chloroflexi*, *Planctomycetes*, *Gemmatimonadetes*, and *Firmicutes* dominated in 21 libraries (only libraries with a clone number ≥ 90 were considered). A total of 2,763 16S rRNA gene sequences were analyzed, with the number of sequences per soil sample ranging from 56 to 396. Typically, environmental rRNA gene sequence datasets generated by Sanger sequencing incorporate about 500 sequences or less (Narang and Dunbar 2004; Ashelford et al. 2006). Only a few studies on single soil samples offering comprehensive 16S rRNA gene sequence datasets of approximately 5,000 and 13,000 sequences are available (Elshahed et al. 2008; Morales et al. 2009). Although a respectable body of phylogenetic data on soil-inhabiting bacteria has been gathered via Sanger sequencing, it is evident that the technique is not suitable for detailed comparisons of a larger number of soil samples. The time consuming, labor-intensive, and highly technical nature of Sanger DNA sequencing (Petrosino et al. 2009), considered as a "first-generation technology", demanded improved sequencing techniques to allow deeper metagenomic analyses of soil.

Next-generation sequencing technologies allow the cloning-independent production of tens to hundreds of thousands sequences in a reasonable time and at moderate costs. Of the available next-generation sequencing technologies primarily Roche-454 pyrosequencing (Roche Applied Science) has been adapted for use in 16S rRNA gene surveys focusing on highly diverse microbial communities in environments such as sediment (Hollister et al. 2010; Youssef et al. 2010; dos Santos et al. 2011), soil (Roesch et al.

2007; Acosta-Martínez et al. 2008; Lauber et al. 2009; Kolton et al. 2011), and ocean water (Sogin et al. 2006; Kirchman et al. 2010; Eloe et al. 2011). The pyrosequencing technology currently generates the longest read length of the existing next-generation sequencing platforms, which increases the likelihood of accurate annotation of genetic fragments using viable databases (Gilbert et al. 2011). In addition, the large pyrosequencing-derived datasets allow the detection of rare bacterial groups in environmental samples (Youssef et al. 2009). Fulthorpe and colleagues (2008) listed genera such as *Alcaligenes*, *Bdellovibrio*, *Rhodospirillum*, and *Flexibacter* which were represented by only 46 to 92 sequences in a sugarcane field-derived pyrosequencing dataset consisting of 28,328 partial 16S rRNA gene sequences. Furthermore, Teixeira et al. (2010) found a predominance of spore-forming and anaerobic bacterial genera such as *Bifidobacterium*, *Ruminococcus*, and *Faecalibacterium* in Antarctic rhizospheres of 10 sampling sites by applying pyrosequencing (total analyzed partial 16S rRNA genes, 27,088). In addition, Lauber et al. (2009) determined 49,944 OTUs (operational taxonomic units) at a genetic distance of 3% (species level) when evaluating 152,359 pyrosequencing-derived partial 16S rRNA gene sequences of 88 soil samples.

1.1 Factors inducing soil microbial community shifts

Deciphering the factors altering diversity, abundance, and distribution of soil microorganisms is challenging and complex. The knowledge on these factors provides important information for the prediction of ecosystem responses to environmental changes and to understand the role of different microbial taxa in soils (Jesus et al. 2009; Griffiths et al. 2011). In recent 16S and 18S rRNA gene surveys it has been emphasized that changes in soil chemistry induce microbial community shifts (Fierer and Jackson 2006; Lauber et al. 2008). The major factor driving soil bacterial community structure appears to be soil pH (Lauber et al. 2009). This impact of soil pH was shown at coarse levels of taxonomic resolution (Baker et al. 2009), but also for individual bacterial groups (Jones et al. 2009). A number of bacterial phyla such as *Acidobacteria*, *Actinobacteria*, and *Bacteroidetes* show strong correlations with soil pH (Lauber et al. 2009). Below the phylum level, the occurrence of a number of acidobacterial subgroups is highly dependent on soil pH. Jones et al. (2009) documented that the relative abundances of acidobacterial subgroups 1, 2, 3, 12, 13, and 15 decreased with pH whereas those of acido-

bacterial subgroups 4, 6, 7, 10, 11, 16, 17, 18, 22, and 25 were positively correlated with pH.

In addition to the importance of soil pH, it has been shown that other soil characteristics such as soil type (Girvan et al. 2003), soil texture (Sessitsch et al. 2001), and carbon content (Fierer et al. 2007) can influence soil microbial community structure. Along depth profiles of soil, surface and subsurface soil comprising different horizons can be distinguished. The A horizon (also referred to as topsoil), part of the surface soil, is often characterized by the accumulation of humidified organic matter. B horizons (also referred to as subsoil) belong to the subsurface soil, are characterized by the obliteration of all or much of the original rock structure, and may contain minerals and clay (FAO 2006). Most currently available surveys investigated soil bacterial communities in surface soils where the microbial community density is known to be highest (Fierer et al. 2003). In a study conducted by Hansel et al. (2008) also horizons of the subsurface soil were considered when analyzing a continuous soil profile. The composition of bacterial communities changed significantly with soil depth, more precisely, the relative abundances of *Alphaproteobacteria*, *Acidobacteria*, *Deltaproteobacteria*, and *Firmicutes* differed along the depth profile of the analyzed soil.

It has been reported that in some cases land use has long-term effects, mainly arising from varying plant species and coherent management practices, on soil characteristics such as soil texture, soil carbon, and pH (Murty et al. 2002; Lauber et al. 2008). Consequently, land use can indirectly evoke microbial community shifts by the modification of soil characteristics (Jesus et al. 2009). Wieland et al. (2001) documented that the type of plant species (clover, bean, or alfalfa) induces variations of microbial communities in soil, rhizosphere, and rhizoplane. In addition, Hackl et al. (2004) showed that soils under Austrian pine forests harbored distinct bacterial communities compared to soils under oak-hornbeam and spruce-fir-beech forests. In Austrian pine forest soils high-G+C gram-positive bacteria (49%) dominated, whereas in oak-hornbeam and spruce-fir-beech forest soils members of the *Holophaga*/*Acidobacterium* group (28% and 35%) were most abundant. In one of the few available pyrosequencing surveys about land use effects on soil bacterial communities Acosta-Martínez et al. (2008) provided a detailed list about bacterial groups classified down to the genus level in soils of two undisturbed grass systems and two agricultural systems. Soil bacteria only present in undisturbed grass systems were *Holophaga*, *Ramlibacter*, and *Streptomyces*, whereas *Alistipes*, *En-*

terococcus, *Prosthetobacter*, and *Nitrospira* were only found in agriculturally managed systems. Nevertheless, most studies on land use and management effects on soil bacterial communities do not allow statistical evaluation, as analysis of replicates is often lacking.

2 Mining soil metagenomes for novel biocatalysts

In complex evolutionary processes nature itself created a gigantic set of enzymes within billions of years. These enzymes function under various environmental conditions in often highly specific reactions. Metagenomic libraries containing directly cloned DNA from various environments such as Antarctic sediment (Zhang et al. 2011), sea sediment (Jeon et al. 2009; Hu et al 2010), glacier ice (Simon et al. 2009), sea water (Chu et al. 2008), and soil (Henne et al. 2000; Voget et al. 2006; Liu et al. 2011) have already been screened for a broad range of biocatalysts. The unprecedentedly diverse pool of microorganisms in soil represents an almost unlimited source for the discovery of genes encoding novel biocatalysts. A high number of metagenome-derived biocatalysts including amylases (Sharma et al. 2010), proteases (Waschkowitz et al. 2009), cellulases (Kim et al. 2008; Wang et al. 2009), xylanases (Hu et al. 2008), agarases (Voget al. 2003), lipases (Elend et al. 2007; Wei et al. 2009), and esterases (Elend et al. 2006; Yu et al. 2011) originate from soil. It has been reported that two esterases from soil and a drinking water biofilm exhibited characteristics, e.g. stability at high pH and unexpected substrate spectra, which could not be related to their environment (Elend et al. 2006; Steele et al. 2009). In addition, a highly halotolerant cellulase derived from a soil metagenome showed activity in the presence of metal ions, solvents, detergents, and chelating agents such as Fe^{2+}, Ni^{2+}, Co^{2+}, Methanol, Ethanol, DMSO, and EDTA, which are often components of industrial processes (Voget et al. 2006). To date, the majority of enzymes relevant for industrial processes are of microbial origin (Uchiyama and Miyazaki 2009). However, many enzymes currently used may not be “ideal” for given industrial demands. In some cases the catalyzed processes had to be adapted to suboptimal enzymes (Fernández-Arrojo et al. 2010). Thus, to optimize defined existing industrial processes and develop more cost-effective new methods, mining soil metagenomes for novel biocatalysts is one way to achieve this.

2.1 Lipases and esterases

Microorganisms express different classes of lipolytic enzymes including lipases (EC 3.1.1.3) and esterases (EC 3.1.1.1) which catalyze the hydrolysis and the synthesis of acylglycerides and other fatty acid esters. Esterases preferentially hydrolyze short-chain esters and, unlike lipases, show no activity toward water-insoluble esters (Arpigny and Jaeger 1999). Lipolytic enzymes are part of the structural superfamily of α/β-hydrolases characterized by a catalytic triad typically formed by serine, histidine, and aspartate residues (Ollis et al. 1992) (Fig. 1a). The serine residue is commonly integrated in the conserved pentapeptide sequence GXSXG and plays a key role in the hydrolysis of ester bonds (Arpigny and Jaeger 1999; Akoh et al. 2004). A nucleophilic attack of the serine residue on the carbonyl carbon-atom of an ester bond initiates the hydrolysis of a lipase/esterase substrate (Hausmann and Jaeger 2010) (Fig. 1b). The result is a tetrahedral intermediate stabilized by hydrogen bonding to amide residues of the so-called oxyanion hole. The collapse of the tetrahedral intermediate leads to the release of the corresponding alcohol by diffusion, followed by the formation of a lipase/esterase-acyl complex. Finally, the hydrolysis of this complex releases the fatty acid by diffusion.

Based on amino acid sequence similarity and some biological properties, bacterial lipolytic enzymes have been classified into eight families (I-VIII) (Arpigny and Jaeger 1999). Recently, additional families have been suggested due to the identification of novel lipolytic enzymes such as LipG (Lee et al. 2006), EstA (Chu et al. 2008), LipEH166 (Kim et al. 2009), EstZ3 and EstGK1 (Bayer et al. 2010), FLS18C and FLS18D (Hu et al. 2010), EstD2 (Lee et al. 2010), and EstPE (Park et al. 2011), which could not be classified according to Arpigny and Jaeger (1999). Lipolytic enzymes exhibit useful features for biotechnological applications such as stereoselectivity, remarkable stability in organic solvents, no cofactor requirement, positional selectivity, and broad substrate specificity (Roh and Villatte 2008, Lee et al. 2010). To date, these enzymes are well established and constantly used in fine chemistry, cosmetic production, pharmaceutical and paper industries, and food technology (Jaeger and Eggert 2002).

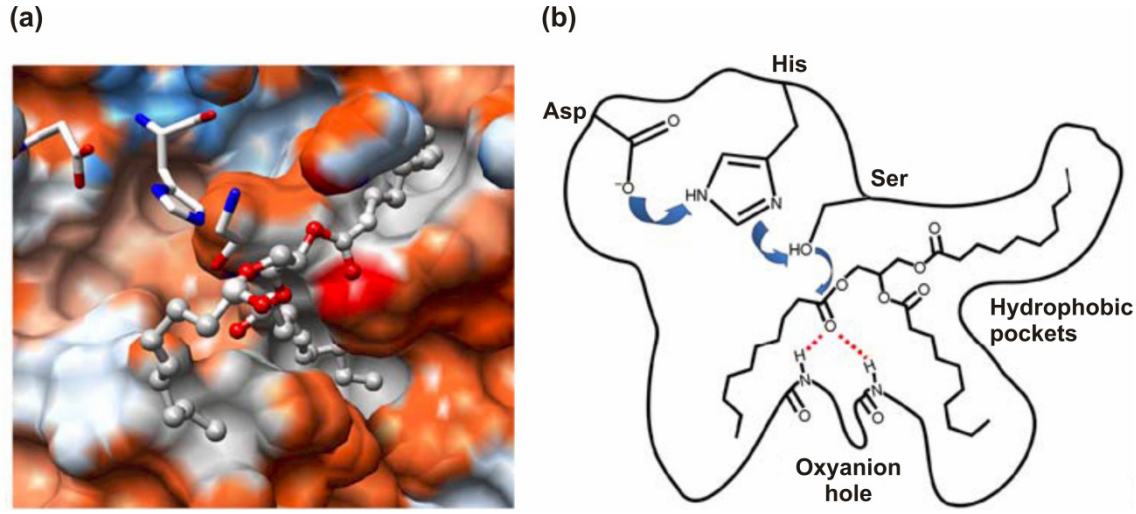


Fig. 1. (a) Active site of *Pseudomonas aeruginosa* lipase LipA (PDB Code 1EX9) (Nardini et al. 2000). The catalytic triad residues Ser^{82}, Asp^{229}, and His^{251} are shown as sticks, the substrate (a triacylglycerol analog) is highlighted in ball and stick mode. Surface electrostatic charges are indicated in blue for positive and red for negative charges, respectively (modified from Hausmann and Jaeger 2010). (b) Active site of a lipolytic enzyme. Blue arrows indicate proton transfer mediated by the catalytic triade residues Asp, His, and Ser, and the nucleophilic attack of the catalytic Ser on the carbonyl carbon-atom of the substrate ester bond. Red dashed lines indicate the interaction of backbone amides forming the oxyanion hole with the substrate (modified from Hausmann and Jaeger 2010).

2.2 Cellulases and xylanases

The degradation of the most abundant polysaccharide in nature, cellulose, is mediated by the synergistical action of three classes of cellulases, including exoglucanases (EC 3.2.1.91), endoglucanases (EC 3.2.1.4), and β-glucosidases (Lynd et al. 2002). Two different types of organization of cellulases have been discovered in microorganisms. The non-complexed system, observed in aerobic fungi and bacteria, is characterized by the secretion of cellulolytic enzymes to the environment, where cellulose substrates are hydrolyzed (Duan and Feng 2010) (Fig. 2a). Additionally, a complexed system present in anaerobic microorganisms has been described. In complexed systems, the cellulolytic enzymes are organized in cellulosomes which are attached to the microbial cell wall (Doi and Kosugi 2004). A fibrillar cellulosomal protein, the so-called scaffoldin protein, enables the concentrated action of different cellulolytic enzymes on cellulose substrates. More precisely, cohesins exposed by the scaffoldin protein allow the attachment of cellulosomal enzymes via their dockerin domains. In addition to the concentration of dif-

ferent cellulases, the scaffoldin protein binds cellulosic material with the help of carbohydrate-binding modules (CBMs) (Fig. 2b).

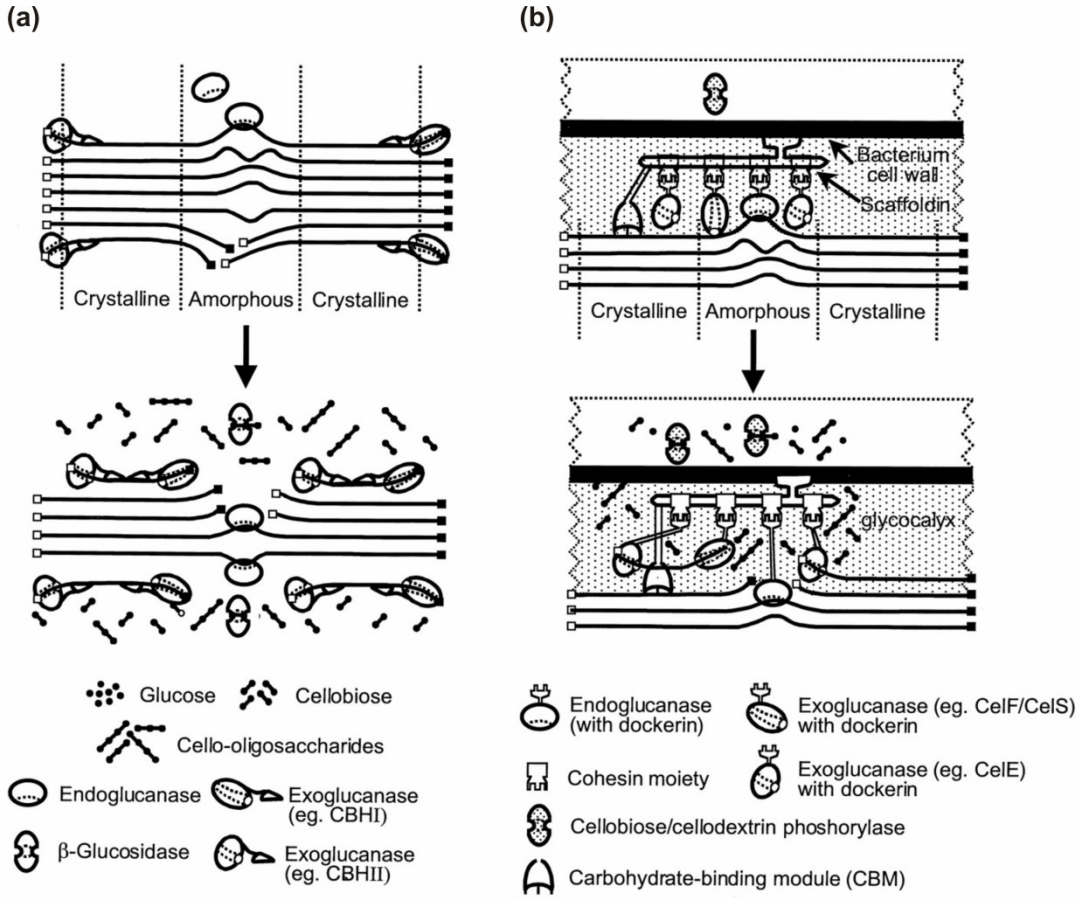


Fig. 2. Schematic representation of the degradation of amorphous and microcrystalline cellulose by non-complexed (a) and complexed (b) cellulase systems. Solid squares represent reducing ends, and open squares represent nonreducing ends. Amorphous and crystalline regions are indicated (modified from Lynd et al. 2002).

Xylanases hydrolyze the second most abundant polysaccharide in nature, xylan (Polizeli et al. 2005). Compared to cellulose, which consists of β-1,4-linked glucose molecules, xylans show a high variety of different side chains including glucuronopyranosyl, 4-O-methyl-D-glucuronopyranosyl, α-L-arabinofuranosyl, acetyl, p-coumaroyl, and feruloyl groups carried by a β-1,4-linked xylose backbone (Beg et al. 2001). To hydrolyze the complex heteropolysaccharide xylan completely, a large set of different enzymes is required. Endo-1,4-β-xylanases (EC 3.2.1.8) and β-xylosidases (EC 3.2.1.37) hydrolyze the xylan backbone. The side chains of xylans are hydrolyzed by α-D-glucuronidases (EC 3.2.1.139), α-L-arabinofuranosidases (EC 3.2.1.55), acetylxylan esterases (EC

3.1.1.72), p-coumaric acid esterases (EC 3.1.1.-), and ferulic acid esterases (EC 3.1.1.73) (Collins et al. 2005).

Both enzyme types, cellulases and xylanases, have high potential in the generation of biofuel as their released sugars can be fermented to ethanol. Currently, butanol which can also be derived from (hemi)cellulosic biomass is considered to be a promising alternative to ethanol, with respect to energy density, corrosiveness, volatility, and ease of separation (Stephanopoulos 2007). Furthermore, (hemi)cellulolytic enzymes are used in the textile, food, and paper industry (Duan and Feng 2010). However, the industrial enzymatic breakdown of (hemi)cellulolosic biomass is presently poorly developed as most available (hemi)cellulolytic enzymes show low activity and efficiency (Hess et al. 2011).

3 Sampling region and study framework

In this study, soil samples of the three German Biodiversity Exploratories Schorfheide-Chorin (northeastern Germany), Hainich-Dün (central Germany), and Schwäbische Alb (southwestern Germany) were analyzed (Fischer et al. 2010a) (Fig. 3). The Biodiversity Exploratories serve as a joint research platform to study the links between land use, biodiversity, and ecosystem processes in replicate investigation areas over a broad period of time. The term "Exploratory" highlights the complementation of comparative observation and monitoring (characteristics of an observatory) by manipulative experiments.

Approximately 1,000 study plots, designated grid plots, were selected in each Exploratory (Fischer et al. 2010a, 2010b). The grid plots incorporate the two land use types forest and grassland, and are intended to permit large-scale biodiversity data analyses and future sampling. A subsample of 100 plots comprising different management types such as fertilized meadow, spruce age class forest, and unmanaged beech forest per exploratory were selected. These plots are designated experimental plots, serving for more detailed observations and experiments such as recordings of climate data, repeated vegetation studies, and seed addition. From the 100 experimental plots per Exploratory, a subset of 18 very intensive plots representing triplicates of different management types were selected for detailed research.

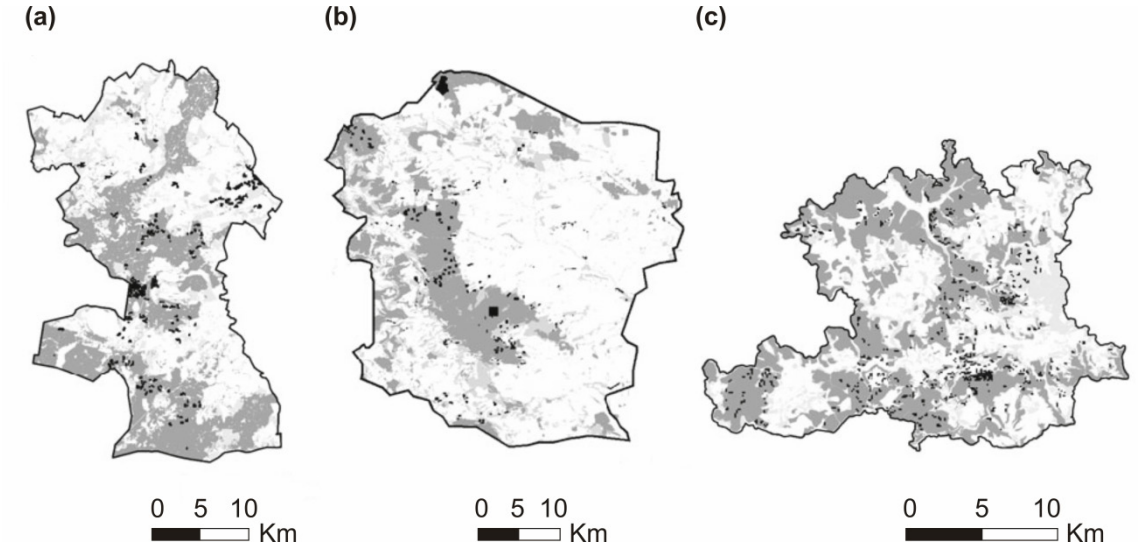


Fig. 3. Maps of the three German Biodiversity Exploratories. (a) Schorfheide-Chorin, (b) Hainich-Dün, and (c) Schwäbische Alb. Forest areas (dark grey) and the distribution of the study plots (black dots) are indicated (modified from Fischer et al. 2010a).

4 Aim of this thesis

The aim of this study was the assessment of the bacterial diversity along different management types in German forest and grassland soils. Phylogenetic analyses of topsoil microbial communities originating from the Schwäbische Alb (Baden-Württemberg, Germany) were performed. In addition, topsoil as well as subsoil samples derived from the Hainich area (Thuringia, Germany) were included in the analyses. The basis for assessment of taxonomic composition of soil bacterial communities in the studied systems was the generation of large 16S rRNA gene datasets by employing pyrosequencing-based approaches.

Microbial community DNA derived from soil samples of the Schwäbische Alb and the Schorfheide-Chorin (Brandenburg, Germany) was used for the construction of small-insert and large-insert metagenomic libraries. To identify novel lipolytic and (hemi)cellulolytic genes and gene families, function-based screens of the constructed metagenomic libraries were performed. The substrate sepcificity of clones carrying lipolytic genes was determined by the hydrolysis of triacylglycerides and *p*-nitrophenyl esters varying in chain length. In addition, biochemical characterization of (hemi)cellulolytic enzymes was carried out.

5 References

Acosta-Martínez, V., S. Dowd, Y. Sun, and V. Allen. 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol Biochem **40**:2762–2770.

Akoh, C., G. Lee, Y. Liaw, T. Huang, and J. Shaw. 2004. GDSL family of serine esterases/lipases. Prog Lipid Res **43**:534–552.

Amann, R. and W. Ludwig. 2000. Ribosomal RNA-targeted nucleic acid probes for studies in microbial ecology. FEMS Microbiol Rev **24**:555–565.

Arpigny, J. and K. Jaeger. 1999. Bacterial lipolytic enzymes: classification and properties. Biochem J **343**:177-183.

Ashelford, K., N. Chuzhanova, J. Fry, A. Jones, and A. Weightman. 2006. New screening software shows that most recent 16S rRNA gene clone libraries contain chimeras. Appl Environ Microbiol **72**:5734-5741.

Baker, K., S. Langenheder, G. Nicol, D. Ricketts, K. Killham, C. Campbell, and J. Prosser. 2009. Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. Soil Biol Biochem **41**:2292–2298.

Bayer, S., A. Kunert, M. Ballschmiter, and T. Greiner-Stoeffele. 2010. Indication for a new lipolytic enzyme family: isolation and characterization of two esterases from a metagenomic library. J Mol Microbiol Biotechnol **18**:181-187.

Beg, Q., M. Kapoor, L. Mahajan, and G. Hoondal. 2001. Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol **56**:326–338.

Brons, J. and J. Elsas. 2008. Analysis of bacterial communities in soil by use of denaturing gradient gel electrophoresis and clone libraries, as influenced by different reverse primers. Appl Environ Microbiol **74**:2717-2727.

Chu, X., H. He, C. Guo, and B. Sun. 2008. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. Appl Microbiol Biotechnol **80**:615-625.

Chu, H., N. Fierer, C. Lauber, J. Caporaso, R. Knight, and P. Grogan. 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol **12**:2998-3006.

Collins, T., C. Gerday, and G. Feller. 2005. Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev **29**:3-23.

Daniel, R. 2005. The metagenomics of soil. Nat Rev Microbiol **3**:470-478.

Delmont, T., P. Robe, S. Cecillon, I. Clark, F. Constancias, P. Simonet, P. Hirsch, and T. Vogel. 2011. Accessing the soil metagenome for studies of microbial diversity. Appl Environ Microbiol **77**:1315-1324.

Doi, R. and A. Kosugi. 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol **2**:541-551.

dos Santos, H., J. Cury, F. do Carmo, A. dos Santos, J. Tiedje, J. van Elsas, A. Rosado, and R. Peixoto. Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. 2011. PLoS One **6**:e16943

Duan, C. and J. Feng. 2010. Mining metagenomes for novel cellulase genes. Biotechnol Lett **32**:1765-1775.

Dunbar, J., S. Takala, S. Barns, J. Davis, and C. Kuske. 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. Appl Environ Microbiol **65**:1662-1669.

Elend, C., C. Schmeisser, C. Leggewie, P. Babiak, J. Carballeira, H. Steele, J. Reymond, K. Jaeger, and W. Streit. 2006. Isolation and biochemical characterization of two novel metagenome-derived esterases. Appl Environ Microbiol **72**:3637-3645.

Elend, C., C. Schmeisser, H. Hoebenreich, H. Steele, and W. Streit. 2007. Isolation and characterization of a metagenome-derived and cold-active lipase with high stereospecificity for (R)-ibuprofen esters. J Biotechnol **130**:370-377.

Eloe, E., D. Fadrosh, M. Novotny, L. Zeigler Allen, M. Kim, M. Lombardo, J. Yee-Greenbaum, S. Yooseph, E. Allen, R. Lasken, S. Williamson, and D. Bartlett. 2011. Going deeper: metagenome of a hadopelagic microbial community. PLoS One **6**:e20388.

Elshahed, M., N. Youssef, A. Spain, C. Sheik, F. Najar, L. Sukharnikov, B. Roe, J. Davis, P. Schloss, V. Bailey, and L. Krumholz. 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl Environ Microbiol **74**:5422-5428.

FAO. 2006. IUSS Working Group WRB. 2006. World reference base for soil resources 2006. 2nd edition. World Soil Resources Reports No. 103. FAO, Rome.

Fernández-Arrojo, L., M. Guazzaroni, N. López-Cortés, A. Beloqui, and M. Ferrer. 2010. Metagenomic era for biocatalyst identification. Curr Opin Biotechnol **21**:725-733.

Fierer, N., J. Schimel, and P. Holden. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol Biochem **35**:167–176.

Fierer, N. and R. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103**:626-631.

Fierer, N., M. Bradford, and R. Jackson. 2007. Toward an ecological classification of soil bacteria. Ecology **88**:1354-1364.

Fischer, M., E. Kalko, K. Linsenmair, S. Pfeiffer, D. Prati, E.-D. Schulze, and W. Weisser. 2010a. Exploratories for large-scale and long-term functional biodiversity research. In Müller F., C. Baessler, H. Schubert, and S. Klotz (eds.), Long-term ecological research – between theory and application (pp. 429-443). Dordrecht: Springer.

Fischer, M., O. Bossdorf, S. Gockel, F. Hänsel, A. Hemp, D. Hessenmöller, G. Korte, J. Nieschulze, S. Pfeiffer, D. Prati, S. Renner, I. Schöning, U. Schumacher, K. Wells, F. Buscot, E. Kalko, K. Linsenmair, E.-D. Schulze, and W. Weisser. 2010b. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl Ecol **11**:473–485.

Fulthorpe, R., L. Roesch, A. Riva, and E. Triplett. 2008. Distantly sampled soils carry few species in common. ISME J **2**:901-910.

Gilbert, J., B. Laverock, B. Temperton, S. Thomas, M. Muhling, and M. Hughes. Metagenomics. 2011. Methods Mol Biol **733**:173-183.

Girvan, M., J. Bullimore, J. Pretty, A. Osborn, and A. Ball. 2003. Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Appl Environ Microbiol **69**:1800-1809.

Griffiths, R., B. Thomson, P. James, T. Bell, M. Bailey, and A. Whiteley. 2011. The bacterial biogeography of British soils. Environ Microbiol **13**:1642-1654.

Hackl, E., S. Zechmeister-Boltenstern, L. Bodrossy, and A. Sessitsch. 2004. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Appl Environ Microbiol **70**:5057-5065.

Handelsman, J., M. Rondon, S. Brady, J. Clardy, and R. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol **5**:R245-249.

Hansel, C., S. Fendorf, P. Jardine, and C. Francis. 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74**:1620-1633.

Hartman, W., C. Richardson, R. Vilgalys, G. Bruland. 2008. Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc Natl Acad Sci U S A **105**:17842-17847.

Hausmann, S. and K. Jaeger. 2010. Lipolytic enzymes from bacteria. In Timmis, K. (ed.), Handbook of hydrocarbon and lipid microbiology (pp. 1099-1126).

Henne, A., R. Schmitz, M. Bömeke, G. Gottschalk, and R. Daniel. 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol **66**:3113-3116.

Hess, M., A. Sczyrba, R. Egan, T. Kim, H. Chokhawala, G. Schroth, S. Luo, D. Clark, F. Chen, T. Zhang, R. Mackie, L. Pennacchio, S. Tringe, A. Visel, T. Woyke, Z. Wang, and E. Rubin.

2011. Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. Science **331**:463-467.

Hollister, E., A. Engledow, A. Hammett, T. Provin, H. Wilkinson, and T. Gentry. 2010. Shifts in microbial community structure along an ecological gradient of hypersaline soils and sediments. ISME J **4**:829-838.

Hu, Y., G. Zhang, A. Li, J. Chen, and L. Ma. 2008. Cloning and enzymatic characterization of a xylanase gene from a soil-derived metagenomic library with an efficient approach. Appl Microbiol Biotechnol **80**:823-830.

Hu, Y., C. Fu, Y. Huang, Y. Yin, G. Cheng, F. Lei, N. Lu, J. Li, E. Ashforth, L. Zhang, and B. Zhu. 2010. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS Microbiol Ecol **72**:228-237.

Jaeger, K. and T. Eggert. 2002. Lipases for biotechnology. Curr Opin Biotechnol **13**:390-397.

Janssen, P., P. Yates, B. Grinton, P. Taylor, and M. Sait. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria*, *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia*. Appl Environ Microbiol **68**:2391-2396.

Janssen, P. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol **72**:1719-1728.

Jeon, J., J. Kim, Y. Kim, H. Kim, H. Lee, S. Kang, S. Kim, and J. Lee. 2009. Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. Appl Microbiol Biotechnol **81**:865-874.

Jesus, E., T. Marsh, J. Tiedje, and F. Moreira. 2009. Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J **3**:1004-1011.

Jones, R., M. Robeson, C. Lauber, M. Hamady, R. Knight, and N. Fierer. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J **3**:442-453.

Kim, S., C. Lee, B. Han, M. Kim, Y. Yeo, S. Yoon, B. Koo, and H. Jun. 2008. Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett **282**:44-51.

Kim, E., K. Oh, M. Lee, C. Kang, T. Oh, and J. Yoon. 2009. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. Appl Environ Microbiol **75**:257-260.

Kirchman, D., M. Cottrell, and C. Lovejoy. 2010. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environ Microbiol **12**:1132-1143.

Kolton, M., Y. Meller Harel, Z. Pasternak, E. Graber, Y. Elad, and E. Cytryn. 2011. Impact of biochar application to soil on the root-associated bacterial community structure of fully developed greenhouse pepper plants. Appl Environ Microbiol **77**:4924-4930.

Lauber, C., M. Strickland, M. Bradford, and N. Fierer. 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem **40**:2407–2415.

Lauber, C., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol **75**:5111-5120.

Lee, M., C. Lee, T. Oh, J. Song, and J. Yoon. 2006. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. Appl Environ Microbiol **72**:7406-7409.

Lee, M., K. Hong, S. Malhotra, J. Park, E. Hwang, H. Choi, Y. Kim, W. Tao, and S. Lee. 2010. A new esterase EstD2 isolated from plant rhizosphere soil metagenome. Appl Microbiol Biotechnol **88**:1125-1134.

Liu, J., W. Liu, X. Zhao, W. Shen, H. Cao, and Z. Cui. 2011. Cloning and functional characterization of a novel endo-β-1,4-glucanase gene from a soil-derived metagenomic library. Appl Microbiol Biotechnol **89**:1083-1092.

Lynd, L., P. Weimer, H. Willem, and I. Pretorius, 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev **66**:506-577.

McCaig, A., L. Glover, J. Prosser. 2001. Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns. Appl Environ Microbiol **67**:4554-4559.

Morales, S., T. Cosart, J. Johnson, and W. Holben. 2009. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. Appl Environ Microbiol **75**:668-675.

Murty, D., M. Kirschbaum, R. McMurtrie, and H. McGilvray. 2002. Does conversion of forest to agricultural land change soil carbon and nitrogen? A review of the literature. Global Change Biol **8**:105-123.

Narang, R. and J. Dunbar. 2004. Modeling bacterial species abundance from small community surveys. Microb Ecol **47**:396-406.

Nardini, M., D. Lang, K. Liebeton, K. Jaeger, and B. Dijkstra. 2000. Crystal structure of *Pseudomonas aeruginosa* lipase in the open conformation. The prototype for family I.1 of bacterial lipases. J Biol Chem **275**:31219-31225.

Ollis, D., E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. Franken, M. Harel, S. Remington, I. Silman, J. Schrag, J. Sussman, K. Verschueren, and A. Goldman. 1992. The alpha/beta hydrolase fold. Protein Eng **5**:197-211.

Park, S., H. Shin, and G. Kim. 2011. Screening and identification of a novel esterase EstPE from a metagenomic DNA library. J Microbiol **49**:7-14.

Petrosino, J., S. Highlander, R. Luna, R. Gibbs, and J. Versalovic. 2009. Metagenomic pyrosequencing and microbial identification. Clin Chem **55**:856-866.

Polizeli, M., A. Rizzatti, R. Monti, H. Terenzi, J. Jorge, and D. Amorim. 2005. Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol **67**:577-591.

Rappé, M. and S. Giovannoni. 2003. The uncultured microbial majority. Annu Rev Microbiol **57**:369-394.

Roesch, L., R. Fulthorpe, A. Riva, G. Casella, A. Hadwin, A. Kent, S. Daroub, F. Camargo, W. Farmerie, and E. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J **1**:283-290.

Roh, C. and F. Villatte. 2008. Isolation of a low-temperature adapted lipolytic enzyme from uncultivated micro-organism. J Appl Microbiol **105**:116-123.

Rondon, M., P. August, A. Bettermann, S. Brady, T. Grossman, M. Liles, K. Loiacono, B. Lynch, I. MacNeil, C. Minor, C. Tiong, M. Gilman, M. Osburne, J. Clardy, J. Handelsman, and R. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol **66**:2541-2547.

Sait, M., P. Hugenholtz, and P. Janssen. 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. Environ Microbiol **4**:654-666.

Sessitsch, A., A. Weilharter, M. Gerzabek, H. Kirchmann, and E. Kandeler. 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl Environ Microbiol **67**:4215-4224.

Sharma, S., F. Khan, and G. Qazi. 2010. Molecular cloning and characterization of amylase from soil metagenomic library derived from Northwestern Himalayas. Appl Microbiol Biotechnol **86**:1821-1828.

Simon, C., J. Herath, S. Rockstroh, and R. Daniel. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. Appl Environ Microbiol **75**:2964-2968.

Simon, C. and R. Daniel. 2011. Metagenomic analyses: past and future trends. Appl Environ Microbiol **77**:1153-1161.

Singh, B., C. Campbell, S. Sorenson, and J. Zhou. 2009. Soil genomics. Nat Rev Microbiol **7**:756.

Sogin, M., H. Morrison, J. Huber, D. Mark Welch, S. Huse, P. Neal, J. Arrieta, and G. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A **103**:12115-12120.

Steele, H., K. Jaeger, R. Daniel, and W. Streit. 2009. Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol **16**:25-37.

Stephanopoulos, G. 2007. Challenges in engineering microbes for biofuels production. Science **315**:801-804.

Teixeira, L., R. Peixoto, J. Cury, W. Sul, V. Pellizari, J. Tiedje, and A. Rosado. 2010. Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. ISME J **4**:989-1001.

Uchiyama, T. and K. Miyazaki. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol **20**:616-622.

Uroz, S., M. Buée, C. Murat, P. Frey-Klett, and F. Martin. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environ Microbiol Rep **2**:281-288.

Vogel, T., P. Simonet, J. Jansson, P. Hirsch, J. Tiedje, J. van Elsas, M. Bailey, R. Nalin, and L. Philippot. 2009. TerraGenome: a consortium for the sequencing of a soil metagenome. Nat Rev Microbiol **7**:252.

Voget, S., C. Leggewie, A. Uesbeck, C. Raasch, K. Jaeger, and W. Streit. 2003. Prospecting for novel biocatalysts in a soil metagenome. Appl Environ Microbiol **69**:6235-6242.

Voget, S., H. Steele, and W. Streit. 2006. Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol **126**:26-36.

Wang, F., F. Li, G. Chen, and W. Liu. 2009. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res **164**:650-657.

Waschkowitz, T., S. Rockstroh, and R. Daniel. 2009. Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. Appl Environ Microbiol **75**:2506-2516.

Wei, P., L. Bai, W. Song, and G. Hao. 2009. Characterization of two soil metagenome-derived lipases with high specificity for p-nitrophenyl palmitate. Arch Microbiol **191**:233-240.

Wieland, G., R. Neumann, and H. Backhaus. 2001. Variation of microbial communities in soil, rhizosphere, and rhizoplane in response to crop species, soil type, and crop development. Appl Environ Microbiol **67**:5849-5854.

Youssef, N., C. Sheik, L. Krumholz, F. Najar, B. Roe, and M. Elshahed. 2009. Comparison of species richness estimates obtained using nearly complete fragments and simulated pyrosequencing-generated fragments in 16S rRNA gene-based environmental surveys. Appl Environ Microbiol **75**:5227-5236.

Youssef, N., M. Couger, and M. Elshahed. 2010. Fine-scale bacterial beta diversity within a complex ecosystem (Zodletone Spring, OK, USA): the role of the rare biosphere. PLoS One **5**:e12414.

Yu, E., M. Kwon, M. Lee, J. Oh, J. Choi, J. Lee, B. Song, D. Hahm, and J. Song. 2011. Isolation and characterization of cold-active family VIII esterases from an arctic soil metagenome. Appl Microbiol Biotechnol **90**:573-581.

Zhang, Y., J. Zhao, and R. Zeng. 2011. Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment. Extremophiles. **15**:23-29.

B Publications

1

Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils

Heiko Nacke[1], Andrea Thürmer[2], Antje Wollherr[2], Christiane Will[1], Ladislav Hodac[3], Nadine Herold[4], Ingo Schöning[5], Marion Schrumpf[4], and Rolf Daniel[1,2]

PLoS ONE (2011), Vol. 6, e17000

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [3]Experimental Phycology and Culture Collection of Algae at the University of Göttingen, Germany, [4]Max Planck Institute for Biogeochemistry, Jena, Germany, [5]Institute of Ecology, Friedrich-Schiller-University, Jena, Germany

Author contributions to the work:

Performed the experiments: HN, AT, CW.

Analyzed data: HN, AT, AW, LH.

Contributed data on soil properties and analysis of these data: IS, NH, MS.

Wrote the paper: HN, RD

Conceived and designed the experiments: RD.

OPEN ACCESS Freely available online

PLoS one

Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils

Heiko Nacke[1], Andrea Thürmer[2], Antje Wollherr[2], Christiane Will[1], Ladislav Hodac[3], Nadine Herold[4], Ingo Schöning[5], Marion Schrumpf[4], Rolf Daniel[1,2]*

1 Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany, **2** Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany, **3** Experimental Phycology and Culture Collection of Algae at the University of Göttingen, Göttingen, Germany, **4** Max Planck Institute for Biogeochemistry, Jena, Germany, **5** Institute of Ecology, Friedrich-Schiller-University, Jena, Germany

Abstract

Background: Soil bacteria are important drivers for nearly all biogeochemical cycles in terrestrial ecosystems and participate in most nutrient transformations in soil. In contrast to the importance of soil bacteria for ecosystem functioning, we understand little how different management types affect the soil bacterial community composition.

Methodology/Principal Findings: We used pyrosequencing-based analysis of the V2-V3 16S rRNA gene region to identify changes in bacterial diversity and community structure in nine forest and nine grassland soils from the Schwäbische Alb that covered six different management types. The dataset comprised 598,962 sequences that were affiliated to the domain Bacteria. The number of classified sequences per sample ranged from 23,515 to 39,259. Bacterial diversity was more phylum rich in grassland soils than in forest soils. The dominant taxonomic groups across all samples (>1% of all sequences) were *Acidobacteria*, *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, and *Firmicutes*. Significant variations in relative abundances of bacterial phyla and proteobacterial classes, including *Actinobacteria*, *Firmicutes*, *Verrucomicrobia*, *Cyanobacteria*, *Gemmatimonadetes* and *Alphaproteobacteria,* between the land use types forest and grassland were observed. At the genus level, significant differences were also recorded for the dominant genera *Phenylobacter*, *Bacillus*, *Kribbella*, *Streptomyces*, *Agromyces*, and *Defluviicoccus*. In addition, soil bacterial community structure showed significant differences between beech and spruce forest soils. The relative abundances of bacterial groups at different taxonomic levels correlated with soil pH, but little or no relationships to management type and other soil properties were found.

Conclusions/Significance: Soil bacterial community composition and diversity of the six analyzed management types showed significant differences between the land use types grassland and forest. Furthermore, bacterial community structure was largely driven by tree species and soil pH.

Citation: Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, et al. (2011) Pyrosequencing-Based Assessment of Bacterial Community Structure Along Different Management Types in German Forest and Grassland Soils. PLoS ONE 6(2): e17000. doi:10.1371/journal.pone.0017000

Editor: Jack Gilbert, Argonne National Laboratory, United States of America

Received October 14, 2010; **Accepted** January 18, 2011; **Published** February 16, 2011

Funding: The work has been funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories" (DA 374/4-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: rdaniel@gwdg.de

Introduction

Soils are considered to be the most diverse microbial habitat on Earth with respect to species diversity and community size. Bacteria are the most abundant group of microorganisms in soil [1]. The calculated number of distinct bacterial genomes ranges from 2,000 to 18,000 per gram of soil [2]. Although the importance of bacteria for ecosystem functions and maintaining soil quality in agriculturally managed systems has long been recognized, the influence of land use type and management type on soil bacterial communities is poorly understood. In a recent pyrosequencing survey, bacterial diversity of forest soil was more phylum rich compared to agricultural soils, which were more species rich [3]. Furthermore, it has been described that *Bacteroidetes* were more predominant in Pullman soil in agricultural systems than in the same soil under non-disturbed conditions, whereas the opposite trend was found for *Actinobacteria* [4]. It has been reported that land use indirectly affects the bacterial community structure by modification of soil properties [5]. Other studies also indicated that soil properties are important drivers of soil bacterial community structure [6], but soil pH appears to be a major factor influencing community composition [7]. This influence of soil pH has been recognized at coarse levels of taxonomic resolution [8], but also within individual phyla [9]. In addition, it has been shown that the type of plant species [10], soil type [11], soil texture [12], and nitrogen availability [13] can affect bacterial community structure. Tree species influences on soil bacterial communities are indicated by previous studies [14], but

detailed information on the affected bacterial groups and degree of these influences is still lacking.

In most previous studies the effects of land use and soil properties on soil bacterial communities have been assessed by employing traditional molecular methods such as Sanger sequencing-based analysis of 16S rRNA gene libraries or fingerprinting methods [15]. These approaches are often limited to the analysis of a relatively small number of clones and a few different soil samples. Taking into account the large bacterial community size and the heterogeneity of soils, only a tiny fraction of the bacterial diversity was unraveled by these studies. Recently, high-throughput pyrosequencing of 16S rRNA gene fragments has been applied for in-depth analysis of soil bacterial communities [3,4]. However, most of the available pyrosequencing studies do not allow a statistical assessment of land use and management effects on soil bacterial communities, as analyses of replicates were often not performed.

In this report, we applied pyrosequencing of the V2-V3 16S rRNA gene region to analyze bacterial community structure in A horizons of forest and grassland sites, which varied in management type. A horizons are mineral soil horizons formed at the surface or below an O horizon, which is dominated by organic material consisting of undecomposed or partially decomposed litter. A horizons are often characterized by accumulation of humidified organic matter [16]. It has been shown that analysis of the V2-V3 region provides a taxonomic resolution ranging from the phylum level to the genus level [17]. Thus, it is possible to detect variations in bacterial communities at different taxonomic levels. We analyzed 18 different soil samples derived from the Schwäbische Alb, which is one of the three German Biodiversity Exploratories [18]. Schwäbische Alb is a mosaic of forest and grasslands with a higher proportion of grassland. This is due to traditional sheep herding. We determined soil bacterial community structure in A horizons of 9 forest and 9 grassland sites. The selected grassland and forest sites covered a range of 6 different management types. Triplicates of the different management types were analyzed, which is an important feature of this study, as it allows statistical analysis of management effects on soil bacterial communities. For each sample, the relative abundance and the distribution of bacterial groups were determined. Subsequently, we correlated variations in the relative abundances with land use type, management type, and soil properties.

Results and Discussion

General characteristics of the soil samples

In this study, we assessed and compared the composition of soil bacterial communities present in the A horizons of 18 soil samples derived from forest and grassland sites of the Schwäbische Alb (Germany) by large-scale pyrosequencing-based analysis of 16S rRNA gene sequences. The soil samples represented triplicates of 6 different management types, which encompassed spruce age class forest (SAF1-3), beech age class forest (BAF1-3), unmanaged beech forest (BF1-3), fertilized intensely managed grassland (FUG1-3), fertilized mown pasture grazed by horse and cattle (FMG1-3), and unfertilized pasture grazed by sheep (UPG1-3) (Tables 1 and S1). The soil groups of the forest soils and the grassland soils were Cambisols and Leptosols, respectively (Table 1). In addition, soil properties such as total nitrogen (N) content, organic carbon (OC) content, pH, and soil texture were determined. The soils had overall low sand (71 ± 64 g kg^{-1}) and highly variable clay contents with values ranging from 188 to 670 g kg^{-1} (average 412 g kg^{-1}). Similarly, OC contents showed a huge variability (68 ± 16 g kg^{-1}). Total N contents were on average lower in forest sites than in grassland sites and C/N ratios were accordingly higher (14 ± 1 forest and 11 ± 1 grassland)

Table 1. Physical and geochemical characteristics of the analyzed grassland and forest soil samples.

Management type	Sample	Soil group	pH	OC (g kg^{-1})	Total N (g kg^{-1})	C:N ratio	Gravimetric water content (%)	Particle size (g kg^{-1})		
								Sand	Silt	Clay
Spruce age class forest	SAF1	Cambisol	3.30	64.57	3.97	16.26	62.8	26	668	306
Spruce age class forest	SAF2	Cambisol	4.55	65.19	4.35	14.99	65.2	43	446	511
Spruce age class forest	SAF3	Cambisol	5.04	74.68	5.14	14.53	76.5	60	445	495
Beech age class forest	BAF1	Cambisol	6.38	78.50	6.01	13.06	75.1	70	534	396
Beech age class forest	BAF2	Cambisol	4.52	57.53	4.45	12.93	70.4	47	587	368
Beech age class forest	BAF3	Cambisol	5.36	39.05	3.15	12.40	50.8	107	575	318
Unmanaged beech forest	BF1	Cambisol	4.87	77.62	5.54	14.01	75.7	109	371	520
Unmanaged beech forest	BF2	Cambisol	5.10	105.00	6.77	15.51	96.6	34	296	670
Unmanaged beech forest	BF3	Cambisol	6.37	60.03	4.49	13.37	54.9	56	495	449
Fertilized intensely managed grassland	FUG1	Leptosol	6.71	77.09	7.58	10.17	66.2	38	543	419
Fertilized intensely managed grassland	FUG2	Leptosol	6.92	72.25	7.18	10.06	59.6	139	646	215
Fertilized intensely managed grassland	FUG3	Leptosol	6.32	53.74	5.18	10.37	57.2	25	449	526
Fertilized mown pasture, horse and cattle	FMG1	Leptosol	5.11	51.61	5.35	9.65	57.5	80	475	445
Fertilized mown pasture, horse and cattle	FMG2	Leptosol	6.36	85.16	7.87	10.82	76.4	56	694	250
Fertilized mown pasture, horse and cattle	FMG3	Leptosol	6.14	68.17	6.67	10.22	64.0	32	492	476
Unfertilized pasture, sheep	UPG1	Leptosol	7.24	40.85	3.65	11.19	46.7	282	530	188
Unfertilized pasture, sheep	UPG2	Leptosol	6.45	81.15	7.41	10.95	74.3	18	384	598
Unfertilized pasture, sheep	UPG3	Leptosol	6.65	68.89	5.82	11.84	67.6	44	684	272

doi:10.1371/journal.pone.0017000.t001

(Table 1). The forest samples showed lower pH values than the grassland soils, which were all, except FMG1, near neutral. The analysis of differences of soil properties and management types by employing one-way analysis of variance and Tukey pair-wise comparisons showed that the analyzed management types did not vary significantly in OC, total N, and soil texture (Table S2). The only significant difference between management types was observed for the pH values, which were higher in unfertilized pastures grazed by sheep (6.9±0.4) than in spruce age class forests (4.7±0.9).

General analyses of the pyrosequencing-derived dataset

Profiling of pylogenetic diversity and community composition by large-scale pyrosequencing of 16S rRNA gene sequences provides more sequence information compared to traditional Sanger sequencing of 16S rRNA gene clone libraries [19]. Although the per-base error rate of pyrosequencing of 16S rRNA genes is not higher than that of Sanger sequencing, the intrinsic error rate of pyrosequencing might lead to overestimation of the number of rare phylotypes. Since each pyrosequencing read is treated as an unique identifier of a community member and correction by assembly and sequencing depth applied during genome projects is not feasible, errors can result in overestimation of diversity [20,21]. To minimize the overestimation of rare phylotypes, we used quality filtering of the pyrosequencing-derived dataset, and clustering and diversity estimates were performed at genetic divergences of ≥3% [21]. Alpha diversity analysis was performed at the same level of surveying effort (22,000 sequences per sample). In addition, denoising of each sequence subset was performed to avoid overestimation of operational taxonomic units (OTUs) and diversity [22,23]. The pyrosequencing-based analysis of the V2-V3 region of the 16S rRNA genes resulted in recovery of 599,284 high quality sequences with a read length of ≥200 bp across all 18 samples. The average read length was 255 bp. The number of sequences per sample ranged from 23,519 to 39,273 with an average of 33,275 (Table S1). We were able to assign 598,962 sequences to the domain Bacteria and to classify 474,868 (79.3%) of these sequences below the domain level. Taking into account the number of sequences per sample and the number of analyzed sequences, the size of this study exceeded other published studies on pyrosequencing-based determination of soil bacterial community composition [3,4,7].

Bacterial diversity and richness

To determine rarefaction curves, richness, and diversity, OTUs were identified at genetic distances of 3, 5, and 20% by using 22,000 randomly selected and denoised sequences per sample. At 20% sequence divergence most rarefaction curves reached saturation, indicating that the surveying effort covered almost the full extent of taxonomic diversity at this genetic distance (Figure S1). Comparison of the rarefaction analyses with the number of OTUs determined by Chao1 and ACE richness estimators revealed that 50.0 to 100% (20% genetic distance) of the estimated taxonomic richness was covered by the surveying effort (Table S3). At 3 and 5% genetic distance, the rarefaction curves were not saturated and the richness estimators indicated that 35.5 to 89.3% and 38.9 to 84.8% of the estimated richness, respectively, were recovered by the sequencing effort (Figures 1, 2 and S1, and Table S3). Thus, we did not survey the full extent of taxonomic diversity at these genetic distances, but a substantial fraction of the bacterial diversity within individual soil samples was assessed at species and genus level by the surveying effort (Figure 1 and Table S3). The comparison of mean Chao1 richness estimates of all forest soils with all grassland soils showed similar values at genetic distances of 3% (3,219 OTUs and 2,611 OTUs, respectively) and 5% (2,331 OTUs and 2,095 OTUs, respectively) but at a genetic distance of 20% (75 OTUs and 153 OTUs, respectively) the richness was higher in grassland ($P<0.05$). The analysis of differences of richness estimates at genetic distances of 3% and 20% and the six management types by employing one-way analysis of variance showed that the analyzed management types did not vary significantly in the predicted number of OTUs ($P>0.05$ in both cases). Comparing this result to previous studies is difficult, as the number of analyzed sequences per sample has an effect on the predicted number of OTUs. In addition, denoising of amplicon sequences was not performed in other studies employing soil-derived pyrosequencing datasets [3,24]. In our study, richness estimates at 3% sequence divergence were approximately 2-fold higher in non-denoised datasets than in the corresponding denoised datasets (data not shown). In addition, in most other studies far fewer 16S rRNA fragments derived from a few soil samples have been analyzed.

The Shannon index of diversity (H') was determined for all samples (Table S3). At a genetic distance of 3%, the Shannon index ranged from 4.96 to 5.92 in the grassland samples and from 4.74 to 5.99 in the forest samples. Comparison of the mean H' of the different management types revealed that the highest bacterial diversity at a genetic distance of 3% was found in unmanaged beech forest, followed by fertilized intensely managed grassland, fertilized mown pastures grazed by horse and cattle, beech age class forest, spruce age class forest, and unfertilized pastures grazed by sheep (Table S3). In forest soils, the sample with the lowest pH (SAF1; pH 3.3) showed the lowest predicted diversity of all forest samples at all analyzed genetic distances (Figures 1, 2 and S1, and Table S3). Similar results were obtained by Fierer and Jackson [25] but a peak of diversity in soils with near-neutral pH values (BAF1 and BF3) that has been found in other studies [7] was not recorded. The spruce forest samples SAF2 and SAF3 showed higher diversity and richness estimates at phylum level but lower richness estimates at species level than the beech forest samples (Figure 2 and Table S3). Thus, an influence of the tree species on bacterial diversity is indicated. In addition, the rarefaction curves and the H' values derived from beech age class forest soils and unmanaged beech forest soils were not separated at all analyzed genetic distances (Figures 1 and S1, and Table S3), indicating that harvesting type (age class forest or unmanaged forest) has a minor or no impact on overall bacterial diversity and richness.

In grassland soils, similar values for estimated bacterial richness were obtained for the three samples derived from fertilized mown pastures grazed by horse and cattle whereas the replicated samples from the other two management types showed strong variations in estimated bacterial richness (Figure 2 and Table S3). At a genetic distance of 3%, the highest average bacterial richness according to Chao1 richness estimator was predicted for fertilized intensely managed grassland (2,887 OTUs), followed by fertilized mown pastures grazed by horse and cattle (2,720 OTUs), and unfertilized pastures grazed by sheep (2,226 OTUs). Nevertheless, the soil sample UPG3 derived from an unfertilized pasture grazed by sheep showed the second highest OTU estimate of all grassland soils (3,413 OTUs). Thus, bacterial diversity showed strong variations within management types in grassland soils.

Distribution of taxa and phylotypes across all samples

The 474,868 sequences classified below domain level were affiliated to 17 bacterial phyla and 4 proteobacterial classes (Tables S4 and S5). The dominant phyla and proteobacterial classes across all samples were *Acidobacteria*, *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, and *Firmi-*

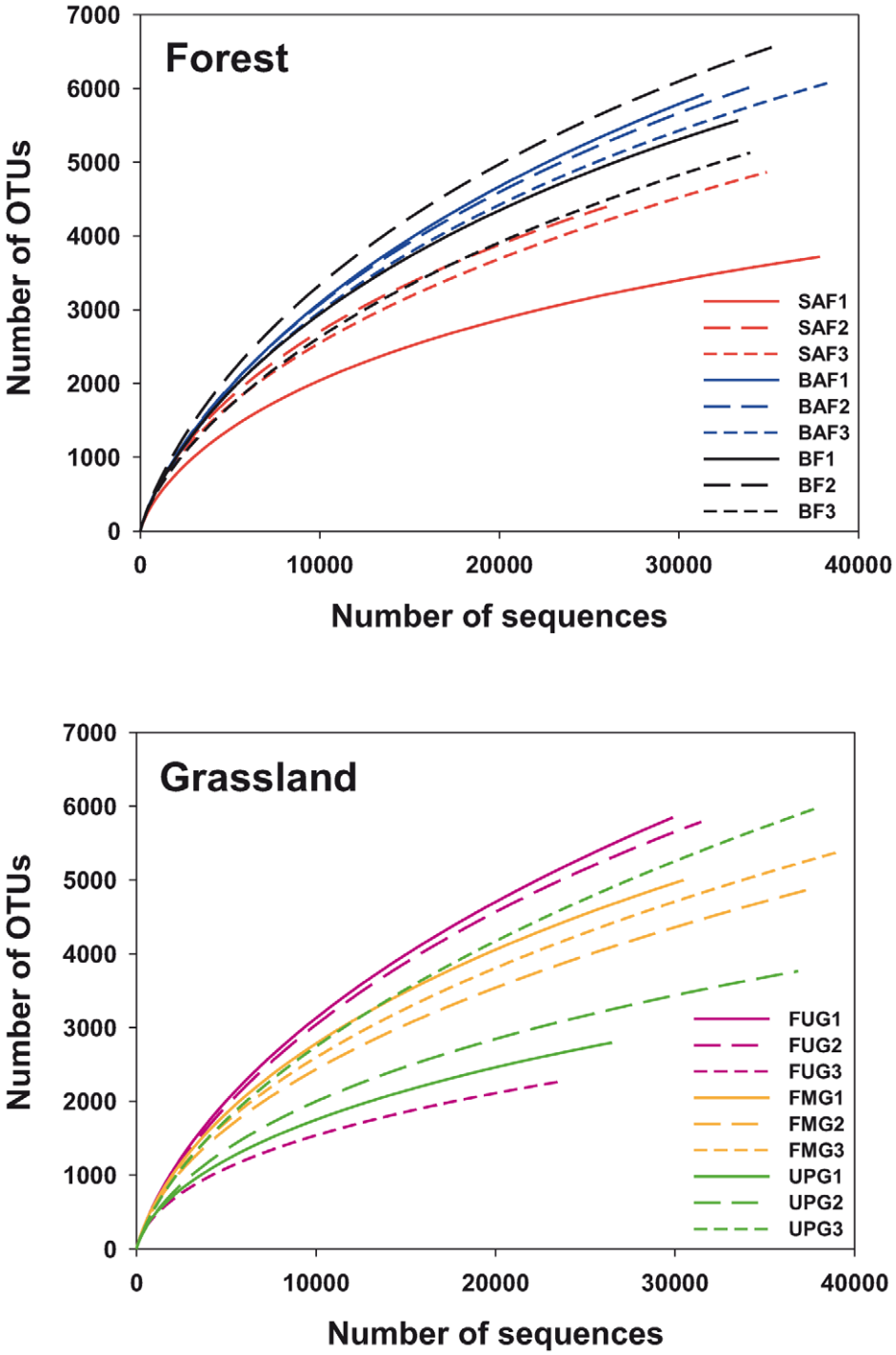


Figure 1. Rarefaction curves indicating the observed number of operational taxonomic units (OTUs) at a genetic distance of 3% in different forest and grassland soils. The spruce age class forest (SAF1-3), beech age class forest (BAF1-3), and unmanaged beech forest (BF1-3) sampling sites are marked by the red, blue, and black color, respectively. The fertilized intensely managed grassland (FUG1-3), fertilized mown pasture grazed by horse and cattle (FMG1-3), and unfertilized pasture grazed by sheep (UPG1-3) sampling sites are shown in purple, orange, and green, respectively.
doi:10.1371/journal.pone.0017000.g001

cutes, representing 19.6, 18.3, 16.1, 5.9, 3.4, 2.9, and 1.2%, respectively, of all sequences that were assigned to the domain Bacteria. The dominant taxa were present in all samples and corresponded roughly with those reported in other studies on soil bacterial community composition [26]. The members of rare phyla (<1% of all classified sequences) included WS3, *Bacteroidetes*, TM7, *Chloroflexi*, *Verrucomicrobia*, *Cyanobacteria*, *Fibrobacteres*, *Spirochaetes*, *Gemmatimonadetes*, *Planctomyces*, OP11, *Deinococcus-Thermus*, and *Fusobacteria* (Figures 3 and 4, and Tables S4 and S5). The most abundant phylotype at a genetic distance of 3% across all samples was an unclassified member of the *Alphaproteobacteria*, representing 2.9% of all sequences. The most abundant phylotype at a genetic

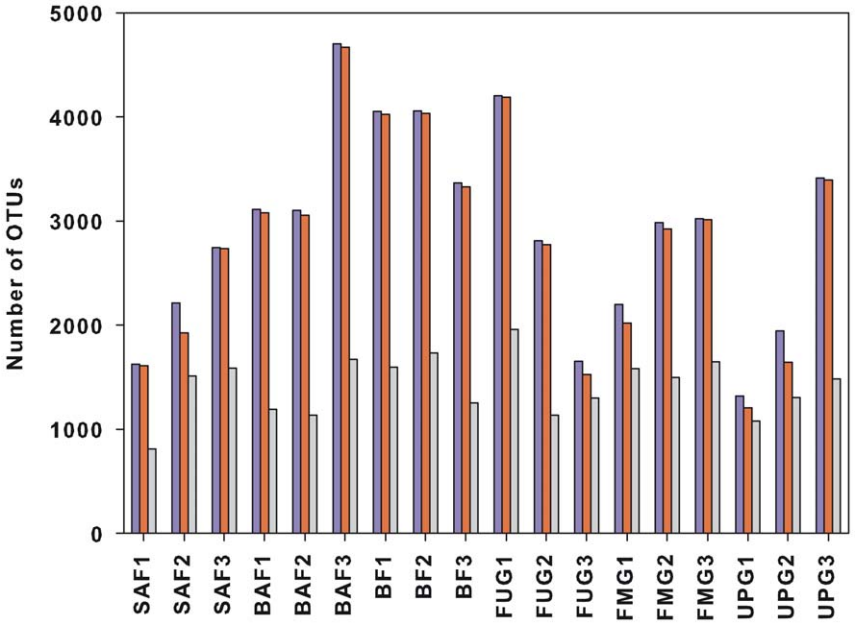


Figure 2. Bacterial richness estimates of German grassland and forest soils representing different management types at a genetic distance of 3%. Richness is expressed as number of observed unique OTUs. In addition, richness has been estimated by the abundance-based coverage estimator (ACE), which is a nonparametric richness estimator based on distribution of abundant (>10) and rare (≤10) OTUs, and the richness estimator Chao1, which is a nonparametric richness estimator based on distribution of singletons and doubletons. Richness prediction from Chao1 is colored in blue, richness prediction from ACE is colored in red, and richness observed is colored in grey. Sample numbers indicating the different management types are given below the graph. A description of the samples is shown in Table 1.
doi:10.1371/journal.pone.0017000.g002

distance of 3% within one individual forest soil sample (SAF1) was a member of the family *Caulobacteraceae*, representing 7.9% of the sequences from that soil. In grassland, an unclassified member of the *Proteobacteria* was the predominant phylotype (22.5% of all sequences) within an individual soil sample (UPG2).

Differences in community structure between forest and grassland soils

The relative abundances of dominant taxa varied between grassland and forest soils. The dominant taxa in forest soils were *Alphaproteobacteria* (25.1±8.9%), *Acidobacteria* (20.4±3.0%), *Actinobacteria* (12.7±2.1%), and *Betaproteobacteria* (6.0±2.1%), whereas in grassland soils the predominant phylogenetic group was *Actinobacteria* (19.6±6.5%) followed by *Acidobacteria* (18.7±4.4%), *Alphaproteobacteria* (11.4±4.4%), and *Betaproteobacteria* (5.9%±1.2) (Figure 3, and Tables S4 and S5). The bacterial phyla and proteobacterial classes observed in our forest and grassland soils were also present in similar relative abundances in a meta-analysis of 32 bacterial 16S rRNA gene libraries derived from a variety of different soils, including samples from pristine forest, grassland and

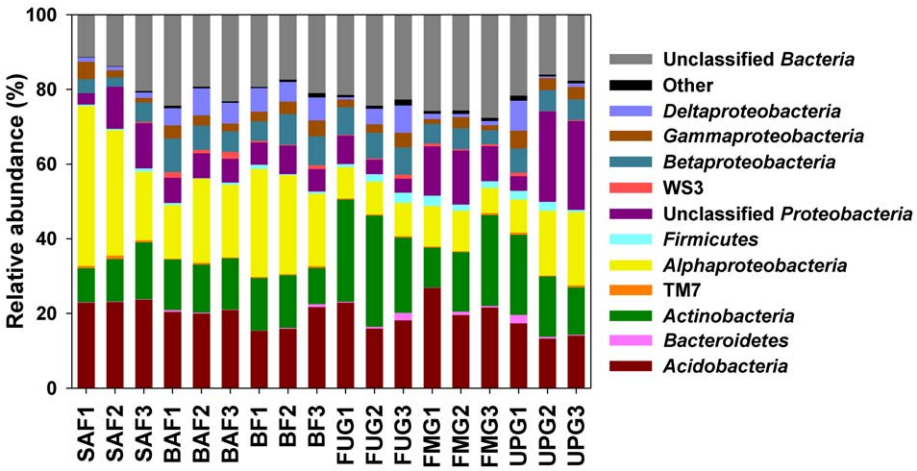


Figure 3. Relative abundances of phylogenetic groups in soils derived from the different grassland and forest sampling sites. Sample numbers indicating the different management types are given below the graph. A description of the samples is shown in Table 1. Phylogenetic groups accounting for ≤0.4% of all classified sequences are summarized in the artificial group 'others'.
doi:10.1371/journal.pone.0017000.g003

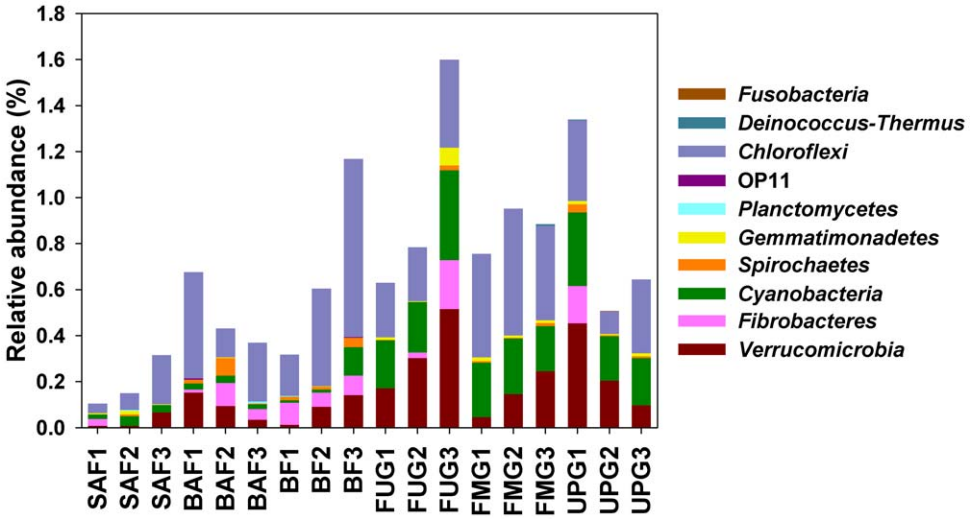


Figure 4. Relative abundances of rare phylogenetic groups of all sequences that were assigned to the domain Bacteria in soils derived from the different grassland and forest sampling sites. A description of the samples is shown in Table 1.
doi:10.1371/journal.pone.0017000.g004

agricultural soils [26]. Principal components analysis (PCA) based on the relative abundances of the different bacterial phyla and proteobacterial classes confirmed that the bacterial communities in grassland soils, except the one in sample UPG3, differed from communities in forest soils (Figure 5). We observed significant higher relative abundances of *Actinobacteria*, *Firmicutes*, *Verrucomicrobia*, *Cyanobacteria*, and *Gemmatimonadetes* in grassland soils than in forest soils whereas *Alphaproteobacteria* showed the opposite pattern ($P<0.05$ in all cases) (Figures 3 and 4). Thus, the shifts in soil bacterial community composition correlated with a change from forest to grassland. A similar trend was also found by comparison of Typic Placandept soils derived from a forest site and a pasture grazed by cattle [27]. In addition, sequences affiliated to *Alphaproteobacteria* dominated in 16S rRNA clone libraries of a spruce-fir-beech forest soil in Austria as well as in a Canadian boreal forest soil [14,28].

Differences of bacterial community structure between grassland and forest soils were also found in the phylogenetic structure within individual lineages. Members of the phylum *Acidobacteria* were predominant across all samples and the second most abundant group in forest and grassland soils, representing approximately 20% of all classified sequences. Correspondingly, members of this phylum have been reported to constitute an average of 20% in bacterial communities derived from various soils [29]. Based on their abundance and the presence in various soil types, *Acidobacteria* appear to play an important role in ecosystem functions of soils, but little is known about physiology and metabolic functions of acidobacterial species. The phylum *Acidobacteria* is divided into 26 subgroups [30] with subgroups 1, 2, 3, 4, and 6 being most abundant within a variety of diverse soils [26,31]. Here, we detected 18 and 22 of these subgroups in grassland soils and forest soils, respectively. Most abundant in the grasslands soils were subgroups 16, 6, 4, 3, and 7, which represented 6.8, 4.4, 2.8, 1.8, and 1.4%, respectively, of all sequences that were classified in grassland. In forest soils, the dominant subgroups were 3, 16, 6, 1, and 4, representing 7.0, 3.0, 2.9, 2.9, and 2.1%, respectively, of all sequences that were classified (Tables S6 and S7).

Most of the sequences belonging to the second most abundant phylum *Alphaproteobacteria* across all samples were affiliated on the order level to the *Rhodospirillales* in forest soils and to *Rhizobiales* in grassland soils. *Actinobacteridae* and *Rubrobacteridae* were the most abundant subclasses within the *Actinobacteria* in both land use types, but the actinobacterial subclass *Coriobacteridae* was only detected in grassland (Tables S8 and S9). Taking into account that members of this subclass are frequently found in gut or rumen samples [32,33] it is possible that they were introduced in the grassland sites by cattle or sheep.

At the genus level, comparison of the relative abundances revealed significant differences between grassland and forest soil bacterial communities. *Mycobacterium* was the most abundant genus across all soil samples, representing 3.7% of all classified sequences in forest soils and 5.7% in grassland soils. *Mycobacteria* are free-living saprophytes and well adapted to a variety of different environments including soils [34]. The distribution of the other dominant genera *Phenylobacter*, *Bacillus*, *Kribbella*, *Agromyces*, and *Defluviicoccus* varied significantly between forest and grassland soils ($P<0.05$). *Phenylobacter* showed a higher relative abundance in forest soils than in grassland soils whereas *Bacillus*, *Kribbella*, *Agromyces*, and *Defluviicoccus* showed the opposite pattern (Figure 6). *Rubrobacter* and *Streptomyces* were present in higher proportions in grassland soils compared to forest soils ($P<0.05$) (Figure 6). Consistently, Acosta-Martínez et al. [4] found *Rubrobacter* and *Streptomyces* among the top 20 predominant bacteria in two non-disturbed grass systems derived from Texas High Plains.

In summary, significant differences of the community structure between the two analyzed land use types forest and grassland were visible. Here, the different analyzed management types in grassland and forest were not reflected by significant changes in bacterial community structure. Thus, soils derived from an identical management type, i.e., UPG1 to UPG3 do not necessarily harbor similar bacterial communities. An exception was the significant impact of tree species (beech or spruce) on community structure in our forest soils. The comparison of relative abundances of bacterial phyla and proteobacterial classes with respect to tree species revealed significant differences between soils derived from spruce and beech forests (Figure 5). Based upon two sample t-test analyses, *Deltaproteobacteria* were less abundant in spruce forest than in beech forests ($P<0.05$) (Figure 3). At the genus level, *Methylocapsa* and *Burkholderia* were more abundant in

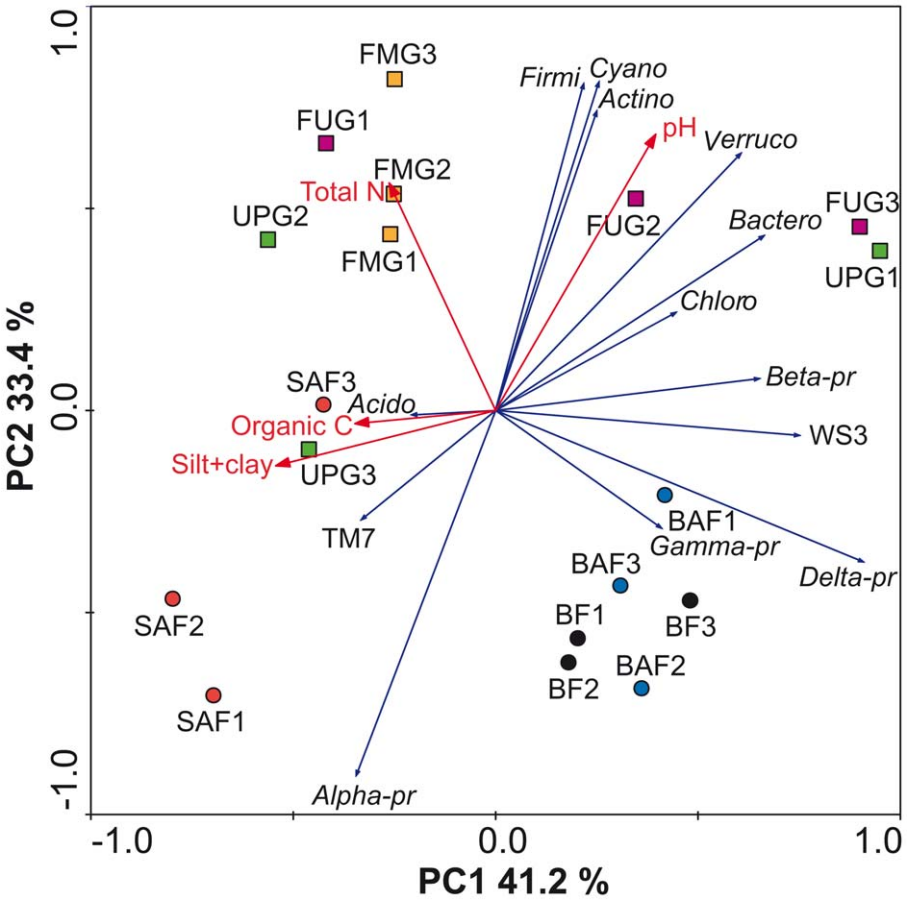


Figure 5. Principal components analysis of bacterial communities as affected by land use, based on the relative abundance of bacterial phyla and proteobacterial classes. Every vector points to the direction of increase for a given variable so that soil samples with similar bacterial communities are localized in similar positions in the diagram. The spruce age class forest (SAF1-3), beech age class forest (BAF1-3), and unmanaged beech forest (BF1-3) sampling sites are marked by the red, green, and black circles, respectively. The fertilized intensely managed grassland (FUG1-3), fertilized mown pasture grazed by horse and cattle (FMG1-3), and unfertilized pasture grazed by sheep (UPG1-3) sampling sites are depicted by red, green, and black squares, respectively. Abbreviations in figure: *Firmi*, *Firmicutes*; *Cyano*, *Cyanobacteria*; *Actino*, *Actinobacteria*; *Verruco*, *Verrucomicrobia*; *Bactero*, *Bacteroidetes*; *Chloro*, *Chloroflexi*; *Beta-pr*, *Betaproteobacteria*; *Delta-pr*, *Deltaproteobacteria*; *Gamma-pr*, *Gammaproteobacteria*; *Alpha-pr*, *Alphaproteobacteria*; *Acido*, *Acidobacteria*.
doi:10.1371/journal.pone.0017000.g005

spruce forest soil than in beech forest soil, whereas *Nocardioides*, *Leptothrix*, and *Amaricoccus* showed the opposite pattern (Figure 6). Thus, tree species appear to be an important driver of soil bacterial community structure, but the type of harvesting (age class forest or unmanaged forest) does not significantly affect bacterial community composition (Figure 5).

Impact of soil properties on the relative abundances of bacterial taxa

Previous studies indicated that soil properties such as pH value or soil texture are important drivers of bacterial community structure [12,35]. We used correlation analysis to identify relationships between the relative abundances of bacterial groups and soil properties. The relative abundances of bacterial groups at different taxonomic levels responded strongly to soil pH. This is in accordance to other surveys on soil bacterial communities derived from different management types in which pH-dependent changes in abundance and distribution of bacterial phyla were observed [36,37]. At the phylum level, relative abundances of *Bacteroidetes* and *Actinobacteria* in the analyzed soils significantly increased with higher pH values ($P<0.05$ in both cases) (Table 2).

As described for a freshwater lake [38] and diverse soils [9], we also found strong correlations of pH and relative abundances of bacterial groups below the phylum level. The relative abundances of the proteobacterial classes *Alphaproteobacteria* and *Betaproteobacteria* were significantly correlated to pH ($P<0.05$). The abundances of *Alphaproteobactia* were negatively correlated with soil pH, whereas the abundances of *Betaproteobacteria* increased with pH (Table 2). Within the *Alphaproteobacteria*, the relative abundances of the order *Caulobacterales* and the family *Acetobacteraceae* showed similar correlations to soil pH as the *Alphaproteobacteria* in general ($P<0.05$ in both cases) (Figure 7). This result corresponded to a cultivation-dependent study of Jimenez-Salgado et al. [39], in which more members of the *Acetobacteraceae* were isolated from low pH soils than from high pH soils. Although relative abundances of *Gammaproteobacteria* showed no significant correlation to soil pH at the class level, the relative abundances of the gammaproteobacterial genus *Dyella* significantly increased with lower pH values ($P<0.05$) (Figure 7). The genus *Dyella* has been recently described by Xie and Yokota [40]. So far, it includes seven species isolated from soil, but no growth of these isolates below pH 4.0 was described [41,42]. In contrast, the highest relative abundances for sequences affiliated to the genus *Dyella* (0.6% of all classified

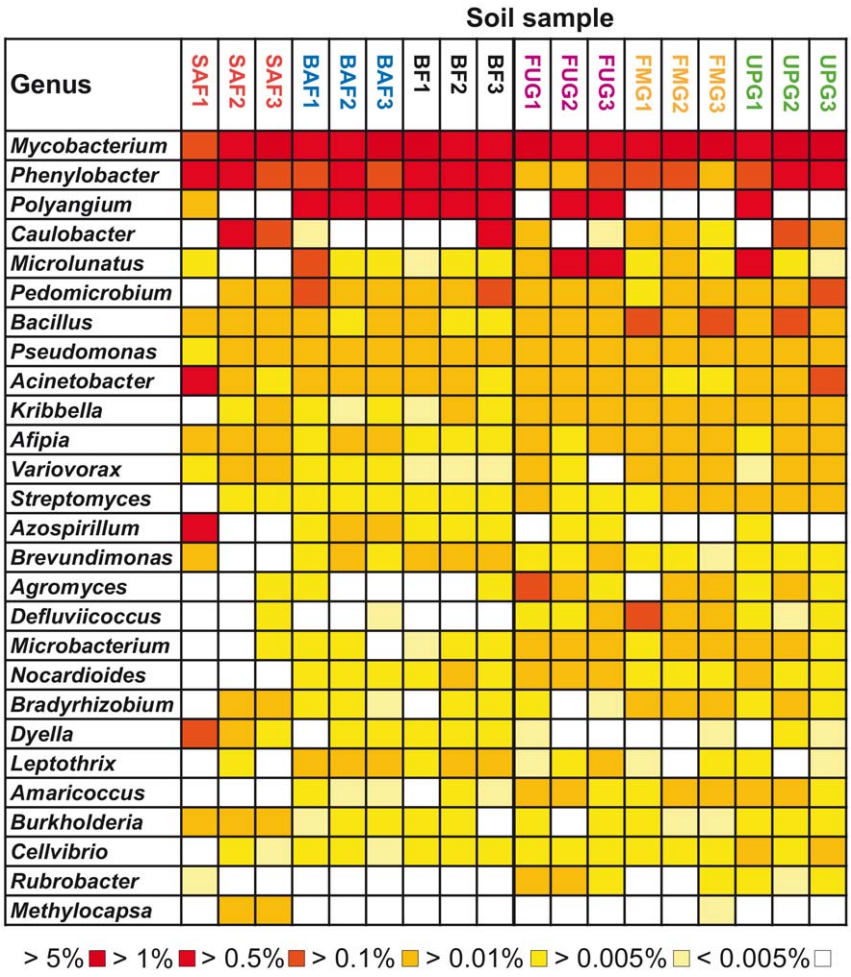


Figure 6. Relative abundances of the most abundant genera as affected by land use. Percentages below the map indicate the abundance of each genus relative to all bacterial sequences that were classified in each of the 18 soils. A description of the samples is shown in Table 1. Grassland and forest samples are separated by a bold line. Samples of different management types are colored in red (SAF1-3), blue (BAF1-3), black (BF1-3), purple (FUG1-3), orange (FMG1-3), and green (UPG1-3).
doi:10.1371/journal.pone.0017000.g006

sequences) were found in sample SAF1, which exhibited the lowest pH value of all samples (pH 3.3). Furthermore, we obtained the highest relative abundances for genera *Azospirillum* and *Acinetobacter* (each representing more than 0.5% of all classified *Bacteria*) in soil sample SAF1 (Figure 6). Thus, our results might help to identify conditions that are best suited for a targeted cultivation of members belonging to these genera.

The occurrence of several subgroups of the *Acidobacteria*, which were predominant across all samples, was also dependent on soil pH. The relative abundances of acidobacterial subgroups 1, 3, 6, 13, 17, and 18 showed strong significant correlations to soil pH ($P<0.001$ in all cases). The relative abundances of subgroups 1, 3, and 13 decreased with pH whereas those of subgroups 6, 17, and 18 were positively correlated with pH (Figure 7 and Table S10). Similar correlations of soil pH and the abundances of acidobacterial subgroups 1, 3, 6, 13, 17, and 18 have been reported by Jones et al. [9]. In addition, the inverse relationship of soil pH on the abundance of members affiliated to subgroup 1 has been reported for soils derived from rotationally grazed perennial ryegrass and white clover pasture [43].

In general, more groups at different taxonomic levels showed significant correlations to soil pH in forest soils than in grassland soils (data not shown). This might be due to the different pH range covered by the analyzed forest and grassland soils. The pH in our forest samples ranged from pH 3.30 to 6.37 (Table 1) whereas the pH values of the grassland samples were all, except sample FMG1, near neutral. Thus, a relatively small pH range was covered by our grassland samples (Table 1), so there is simply less pH range from which to determine correlations. Significant correlations of relative abundances with other soil properties were found for *Deltaproteobacteria* and *Actinobacteria*. The *Deltaproteobacteria* showed a significant correlation to OC ($P<0.05$) with higher abundances in soils with low OC content, whereas *Actinobacteria* showed a significant correlation to total N ($P<0.05$) with higher abundances in soils with high total N content (Table 2), but a connection to the observed correlations was not evident.

Conclusion

The analysis of one of the largest bacterial 16S rRNA-based datasets from soils revealed statistically significant differences in soil bacterial diversity and community structure between the two land use types forest and grassland. Additionally, the occurrence of different tree species had statistically significant effects on soil bacterial diversity, richness, and community composition in forest. The analysis of influences of soil properties on bacterial community structure revealed that pH had the strongest effect

Table 2. Spearman's rank correlations between the relative abundances of the six most abundant bacterial phyla and proteobacterial classes and the soil properties in grassland and forest soils.

Taxonomic group	Correlation			
	pH	OC	Total N	Sand/Silt/Clay
Actinobacteria	**0.58**	0.26	**0.52**	0.02/−0.08/−0.02
Bacteroidetes	**0.71**	0.14	0.33	−0.08/0.17/−0.19
Alphaproteobacteria	**−0.68**	0.05	−0.44	−0.12/−0.13/0.22
Betaproteobacteria	**0.56**	0.22	0.35	0.04/0.04/0.00
Deltaproteobacteria	−0.10	**−0.48**	−0.55	0.43/−0.15/−0.04
Gammaproteobacteria	0.27	−0.04	−0.17	−0.13/0.19/−0.19

Correlations for *Acidobacteria* are shown at higher taxonomic resolution Table S10.
Bold numbers: $P<0.05$; Bold and underlined numbers $P<0.001$.
doi:10.1371/journal.pone.0017000.t002

on bacterial community structure of the analyzed soil properties. Management type and other soil properties appear to have a minor impact on soil bacterial community structure and diversity.

In this survey, the correlations between land use type and community composition were obvious. The relative abundances of a number of taxonomic groups changed significantly between forest and grassland soils (e.g., *Actinobacteria*), but the abundances of other taxa (e.g., *Gammaproteobacteria*) were almost unaffected by land use type, indicating that the abundances of the latter groups are influenced by other factors. Specific bacterial groups such as *Amaricoccus* or *Methylocapsa* showed significantly higher abundances in beech or spruce forest soils. Finally, we cannot determine whether pH has a direct or indirect effect on community composition, as a number of soil properties (e.g., OC) are directly or indirectly related to pH [44]. Thus, the effect of a number of different factors is reflected by soil pH and these factors may also drive community composition.

Availability

The 18 pyrosequencing-derived 16S rRNA gene sequence datasets have been deposited in the GenBank short-read archive under accession number SRA022075.

Materials and Methods

Site description, sampling, DNA extraction, and soil characterization

In the frame of the German Biodiversity Exploratories, initiative soil samples were collected from 9 forest and 9 grassland plots of the German Biodiversity Exploratory Schwäbische Alb. The Schwäbische Alb covers more than 450 km×450 km in the state of Baden-Württemberg (southwestern Germany). Soil samples were collected in April 2008. The forest sampling sites included 3 spruce age class forests (SAF1-3), 3 beech age class forests (BAF1-3), and 3 unmanaged beech forests (BF1-3). Grassland sampling sites comprised 3 fertilized intensely managed grasslands (FUG1-3), 3 fertilized mown pastures grazed by horse and cattle (FMG1-3), and 3 unfertilized pastures grazed by sheep (UPG1-3) (Table S1). The dominant grasses included *Poa trivialis*, *Trisetum flavescens*, and *Arrhenaterum elatius* in sites FUG1-3, *Poa trivialis*, *Alopecurus pratensis*, *Trisetum flavescens*, *Dactylis glomerata*, *Festuca pratensis*, *Lolium perenne*, and *Arrhenaterum elatius* in sites FMG1-3, and *Brachypodium pinnatum*, *Bromus erectus*, and *Festuca guestfalica* in sites UPG1-3. A detailed description of the dominant grasses of the individual plots is provided in Table S11.

Soil samples were collected and classified at each of the grassland and forest sites as described by Will et al. [45]. Briefly, five soil cores (8.3 cm in diameter) were sampled with a motor driven soil column cylinder at each corner and in the center of each plot within a given area of 20 m×20 m. Composite samples of the five collected A horizons per plot were used for DNA extraction, after the soils were homogenized and coarse roots and stones (>5 mm) were removed. Total microbial community DNA was extracted from approximately 8 g soil derived from the A horizons of each plot by employing the MoBio PowerMax Soil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA) as recommended by the manufacturer. DNA concentrations were quantified by using a NanoDrop ND-1000 UV-Vis Spectrophotometer (NanoDrop Technologies, USA) according to the manufacturer's protocol.

OC content, total N content, soil texture, and soil pH were measured as described by Will et al. [45]. To determine the gravimetric water content, 10 g of moist soil were dried to constant weight at 105°C for 24 h. The mass of water was calculated per mass of dry soil.

Amplification of 16S rRNA genes and pyrosequencing

The V2-V3 region of the 16S rRNA gene was amplified by PCR. The PCR reaction mixture (33 µl) contained 3.3 µl 10-fold reaction buffer (Fusion GC buffer, FINNZYMES, Espoo, Finland), 800 µM of each of the four deoxynucleoside triphosphates, 3% DMSO, 1.2 µM of each of the primers, 0.5 U of Phusion hot start high-fidelity DNA Polymerase (FINNZYMES), and 20 ng of isolated DNA as template. The V2-V3 region was amplified with the following set of primers containing the Roche 454 pyrosequencing adaptors (underlined): V2for 5′-GCCTCCCTCGCGCCATCAGAGTGGCGGACGGGTGAGTAA-3′ and V3rev 5′-GCCTTGCCAGCCCGCTCAGCGTATTACCGCGGCTGCTG-3′ (modified from Schmalenberger et al. [46]). The following thermal cycling scheme was used: initial denaturation at 98°C for 5 min, 25 cycles of denaturation at 98°C for 45 s, annealing at 68°C for 45 s, and extension at 72°C for 25 s followed by a final extension period at 72°C for 5 min. All samples were amplified in triplicate, pooled in equal amounts, and purified using the peqGold gel extraction kit as recommended by the manufacturer (Peqlab Biotechnologie GmbH, Erlangen, Germany). Quantification of the PCR products was performed using the Quant-iT dsDNA BR assay kit and a Qubit fluorometer (Invitrogen GmbH, Karlsruhe, Germany) as recommended by the manufacturer. The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA genes by using a Roche GS-FLX 454 pyrosequencer (Roche, Mannheim, Germany) and the instructions of the manufacturer for amplicon sequencing.

Analysis of pyrosequencing data

Sequences that were shorter than 200 bp in length and reads containing any unresolved nucleotides were removed from the 18 pyrosequencing-derived datasets. For taxonomy-based analysis, the RDP Classifier of the Ribosomal Database Project (RDP) was used [47] at a confidence threshold of 80%. Pyrosequencing noise was removed for alpha diversity analyses by using the denoiser program [23]. For the determination of OTUs, we defined species, genus, and phylum level at 3, 5, and 20%, respectively, sequence divergence according to Schloss and Handelsman [48]. OTUs were determined for each denoised sequence dataset by using the uclust OTU picker version 1.2.21q of the QIIME software pipeline [49]. We calculated rarefaction curves as well as the

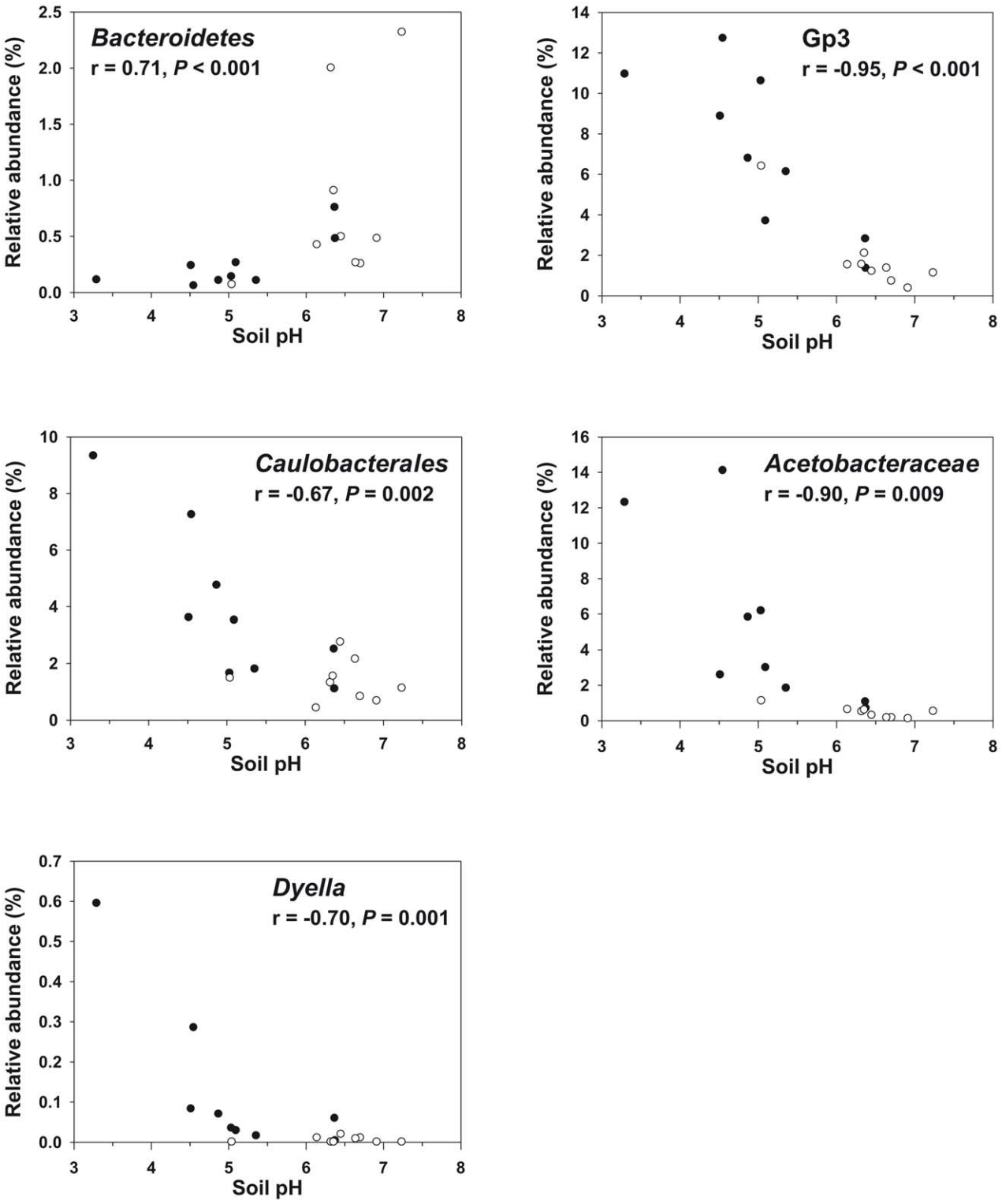


Figure 7. Correlations between relative abundances of different taxonomic groups and soil pH. Black circles represent forest sites and white circles represent grassland sites. Spearman's rank correlation coefficients (r) with the associated *P* values are shown for each taxonomic group. Abbreviation: Gp3, acidobacterial subgroup 3.
doi:10.1371/journal.pone.0017000.g007

Shannon [50] index based on OTU picker data, by employing the RDP pyrosequencing pipeline [51]. ACE and Chao1 [52] indices were calculated using the EstimateS program version 8.2.0 (http://purl.oclc.org/estimates).

Statistical analyses

Normality tests (Shapiro-Wilk) were performed with data that were used for principal component analysis (PCA), and one-way analysis of variance. Data that did not pass normality test were log transformed

and normality test was repeated. Only data that passed normality test were used for further analyses. For each soil attribute and each richness estimate at 3 and 20% genetic distance, one-way analysis of variance and Tukey pair-wise comparisons were used to determine the minimum significant difference ($P<0.05$) between management types by employing STATISTICA 8.0 (StatSoft, Inc., Tulsa, USA). To compare bacterial community structures across all samples based on the relative abundance of bacterial phyla and proteobacterial classes, PCA was performed by using CANOCO for Windows [53]. To correlate bacterial taxonomic groups with soil properties, Spearman's rank correlations were determined by using the SigmaPlot program version 11.0 (Systat Software, Inc., San Jose, CA). We used two sample t-test analyses and M-W-U-Test for non-parametric data to compare relative abundances of bacterial groups and richness estimates between grassland and forest, and on a second level between different management types using the software package PAST [54].

Supporting Information

Figure S1 Rarefaction curves indicating the observed number of OTUs at genetic distances of 5 and 20% in the different forest and grassland soils. The spruce age class forest (SAF1-3), beech age class forest (BAF1-3), and unmanaged beech forest (BF1-3) sampling sites are marked by the red, blue, and black color, respectively. The fertilized intensely managed grassland (FUG1-3), fertilized mown pasture grazed by horse and cattle (FMG1-3), and unfertilized pasture grazed by sheep (UPG1-3) sampling sites are shown in purple, orange, and green, respectively.
(DOC)

Table S1 Localization of the sampling sites and number of 16S rRNA gene sequences derived from the analyzed grassland and forest soil samples.
(DOC)

Table S2 Mean values of soil properties and standard deviation for each management type and ANOVA *P* values. Differences of soil properties between management types were analyzed by employing one-way analysis of variance and Tukey pair-wise comparisons. Significant ANOVA *P* values are shown in bold ($P<0.05$). Figures followed by different letters indicate differences among management types ($P<0.05$). Abbreviations: SAF, spruce age class forest; BAF, beech age class forest; BF, unmanaged beech forest; FUG, fertilized intensely managed grassland; FMG, fertilized mown pasture grazed by horse and cattle; UPG, unfertilized pasture grazed by sheep. Complete soil and site information for all 18 sampling sites is provided in Table 1.
(DOC)

Table S3 Bacterial diversity as assessed by Shannon index (H') and species richness estimation in all forest and grassland soils. The results from the rarefaction analyses are also depicted in Figure 1 and Figure S1.
(DOC)

Table S4 Relative abundances of bacterial phyla and proteobacterial classes in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could not be assigned to a specific phylum or a proteobacterial class.
(DOC)

Table S5 Relative abundances of bacterial phyla and proteobacterial classes in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could not be assigned to a specific phylum or a proteobacterial class.
(DOC)

Table S6 Relative abundances of acidobacterial subgroups in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could be assigned to the phylum level only.
(DOC)

Table S7 Relative abundances of acidobacterial subgroups in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could be assigned to the phylum level only.
(DOC)

Table S8 Relative abundances of taxonomic groups within the phylum *Actinobacteria* and within proteobacterial classes in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could be assigned to the phylum level only.
(DOC)

Table S9 Relative abundances of taxonomic groups within the phylum *Actinobacteria* and within proteobacterial classes in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could be assigned to the phylum level only.
(DOC)

Table S10 Spearman's rank correlations between relative abundances of *Acidobacteria* subgroups and soil properties. Only relative abundances of acidobacterial subgroups that represented $\geq$0.029% of all analyzed sequences were considered.
(DOC)

Table S11 Dominant grasses of the analyzed grassland sites.
(DOC)

Acknowledgments

Field work permits were given by the responsible state environmental office of Baden-Württemberg. We thank the local implementation team of the Schwäbische Alb exploratory, the BEO (biodiversity exploratories office), and Fabian Alt for supporting the coordinated soil sampling campaign. In addition, we thank Steffen Boch for providing vegetation data of the analyzed study sites.

Author Contributions

Conceived and designed the experiments: RD. Performed the experiments: HN AT CW. Analyzed the data: HN AT AW LH. Wrote the paper: HN RD. Contributed data on soil properties and analysis of these data: IS NH MS.

References

1. Gans J, Wolinsky M, Dunbar J (2005) Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Science 309: 1387–1390.
2. Daniel R (2005) The metagenomics of soil. Nat Rev Microbiol 3: 470–478.
3. Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, et al. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1: 283–290.

4. Acosta-Martínez V, Dowd S, Sun Y, Allen V (2008) Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol Biochem 40: 2762–2770.
5. Jesus ED, Marsh TL, Tiedje JM, Moreira FMD (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J 3: 1004–1011.
6. Wakelin SA, Macdonald LM, Rogers SL, Gregg AL, Bolger TP, et al. (2008) Habitat selective factors influencing the structural composition and functional capacity of microbial communities in agricultural soils. Soil Biol Biochem 40: 803–813.
7. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol 75: 5111–5120.
8. Bååth E, Anderson TH (2003) Comparison of soil fungal/bacterial ratios in a pH gradient using physiological and PLFA-based techniques. Soil Biol Biochem 35: 955–963.
9. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, et al. (2009) A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J 3: 442–453.
10. Wieland G, Neumann R, Backhaus H (2001) Variation of microbial communities in soil, rhizosphere, and rhizoplane in response to crop species, soil type, and crop development. Appl Environ Microbiol 67: 5849–5854.
11. Girvan MS, Bullimore J, Pretty JN, Osborn AM, Ball AS (2003) Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Appl Environ Microbiol 69: 1800–1809.
12. Sessitsch A, Weilharter A, Gerzabek MH, Kirchmann H, Kandeler E (2001) Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl Environ Microbiol 67: 4215–4224.
13. Frey SD, Knorr M, Parrent JL, Simpson RT (2004) Chronic nitrogen enrichment affects the structure and function of the soil microbial community in temperate hardwood and pine forests. Forest Ecol Manage 196: 159–171.
14. Hackl E, Zechmeister-Boltenstern S, Bodrossy L, Sessitsch A (2004) Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Appl Environ Microbiol 70: 5057–5065.
15. McCaig AE, Glover LA, Prosser JI (2001) Numerical analysis of grassland bacterial community structure under different land management regimens by using 16S ribosomal DNA sequence data and denaturing gradient gel electrophoresis banding patterns. Appl Environ Microbiol 67: 4554–4559.
16. FAO (2006) Guidelines for soil description, 4th edn. Rome, Italy: FAO.
17. Chakravorty S, Helb D, Burday M, Connell N, Alland D (2007) A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Meth 69: 330–339.
18. Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, et al. (2010) Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic Appl Ecol 11: 473–485.
19. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA 103: 12115–12120.
20. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Gen Biol 8: R143.
21. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ Microbiol 12: 118–123.
22. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Noise and the accurate determination of microbial diversity from 454 pyrosequencing data. Nat Meth 6: 639–641.
23. Reeder J, Knight R (2010) Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Meth 7: 668–669.
24. Morales S, Cosart T, Johnson J, Holben W (2009) Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. Appl Environ Microbiol 75: 668–675.
25. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103: 626–631.
26. Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol 72: 1719–1728.
27. Nüsslein K, Tiedje JM (1999) Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. Appl Environ Microbiol 65: 3622–3626.
28. Dimitriu PA, Grayston SJ (2010) Relationship between soil properties and patterns of bacterial β-diversity across reclaimed and natural boreal forest soils. Microb Ecol 59: 563–573.
29. Dunbar J, Barns SM, Ticknor LO, Kuske CR (2002) Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol 68: 3035–3045.
30. Barns SM, Cain EC, Sommerville L, Kuske CR (2007) *Acidobacteria* phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. Appl Environ Microbiol 73: 3113–3116.
31. Barns SM, Takala SL, Kuske CR (1999) Wide distribution and diversity of members of the bacterial kingdom *Acidobacterium* in the environment. Appl Environ Microbiol 65: 1731–1737.
32. Anderson RC, Rasmussen MA, Jensen NS, Allison MJ (2000) *Denitrobacterium detoxificans* gen. nov., sp. nov., a ruminal bacterium that respires on nitrocompounds Int J Syst Evol Microbiol 50: 633–638.
33. Collado MC, Sanz Y (2007) Quantification of mucosa-adhered microbiota of lambs and calves by the use of culture methods and fluorescent in situ hybridization coupled with flow cytometry techniques. Vet Microbiol 121: 299–306.
34. Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, et al. (2007) Genomics of *Actinobacteria*: tracing the evolutionary history of an ancient phylum. Microbiol Mol Biol Rev 71: 495–548.
35. Rousk J, Bååth E, Brookes P, Lauber C, Lozupone C, et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4: 1340–1351.
36. Lauber CL, Strickland MS, Bradford MA, Fierer N (2008) The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem 40: 2407–2415.
37. Hartman WH, Richardson CJ, Vilgalys R, Bruland GL (2008) Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc Natl Acad Sci USA 105: 17842–17847.
38. Newton RJ, Jones SE, Helmus MR, McMahon KD (2007) Phylogenetic ecology of the freshwater *Actinobacteria* acI lineage. Appl Environ Microbiol 73: 7169–7176.
39. Jimenez-Salgado T, Fuentes-Ramirez LE, Tapia-Hernandez A, Mascarua-Esparza MA, Martinez-Romero E, et al. (1997) *Coffea arabica* L., a new host plant for *Acetobacter diazotrophicus*, and isolation of other nitrogen-fixing acetobacteria. Appl Environ Microbiol 63: 3676–3683.
40. Xie CH, Yokota A (2005) *Dyella japonica* gen. nov., sp. nov., a gamma-proteobacterium isolated from soil. Int J Syst Evol Microbiol 55: 753–756.
41. Weon HY, Anandham R, Kim BY, Hong SB, Jeon YA, et al. (2009) *Dyella soli* sp. nov. and *Dyella terrae* sp. nov., isolated from soil. Int J Syst Evol Microbiol 59: 1685–1690.
42. Anandham R, Kwon SW, Indira Gandhi P, Kim SJ, Weon HY, et al. (2010) *Dyella thiooxydans* sp. nov., a facultative chemolithotrophic thiosulfate oxidizing bacterium isolated from rhizosphere soil of sunflower (*Helianthus annuus* L.). Int J Syst Evol Microbiol: doi:10.1099/ijs.0.022012-0.
43. Sait M, Davis KER, Janssen PH (2006) Effect of pH on isolation and distribution of members of subdivision 1 of the phylum *Acidobacteria* occurring in soil. Appl Environ Microbiol 72: 1852–1857.
44. Brady NC, Weil RR (2007) The nature and properties of soil, 13th edn. Upper Saddle River, New Jersey, USA: Prentice Hall.
45. Will C, Thürmer A, Wollherr A, Nacke H, Herold N, et al. (2010) Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol 76: 6751–6759.
46. Schmalenberger A, Schwieger F, Tebbe C (2001) Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. Appl Environ Microbiol 67: 3557–3563.
47. Wang Q, Garrity G, Tiedje J, Cole J (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73: 5261–5267.
48. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol 71: 1501–1506.
49. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Meth 7: 335–336.
50. Shannon CE, Weaver W (1949) The mathematical theory of communication. MD Comput 14: 306–317.
51. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucl Acids Res 37: D141–145.
52. Chao A, Bunge J (2002) Estimating the number of species in a stochastic abundance model. Biometrics 58: 531–539.
53. ter Braak CJF, Smilauer P (2002) CANOCO reference manual and CanoDraw for windows user's guide: software for canonical community Ordination (version 4.5). Ithaca, NY, USA: Microcomputer Power.
54. Hammer Ø, Harper DAT, Ryan PD (2001) PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4: 1–9.

1.1 Supplemental information for chapter B1

Contents

Table S10. Spearman's rank correlations between relative abundances of *Acidobacteria* subgroups and soil properties.

Table S11. Dominant grasses of the analyzed grassland sites.

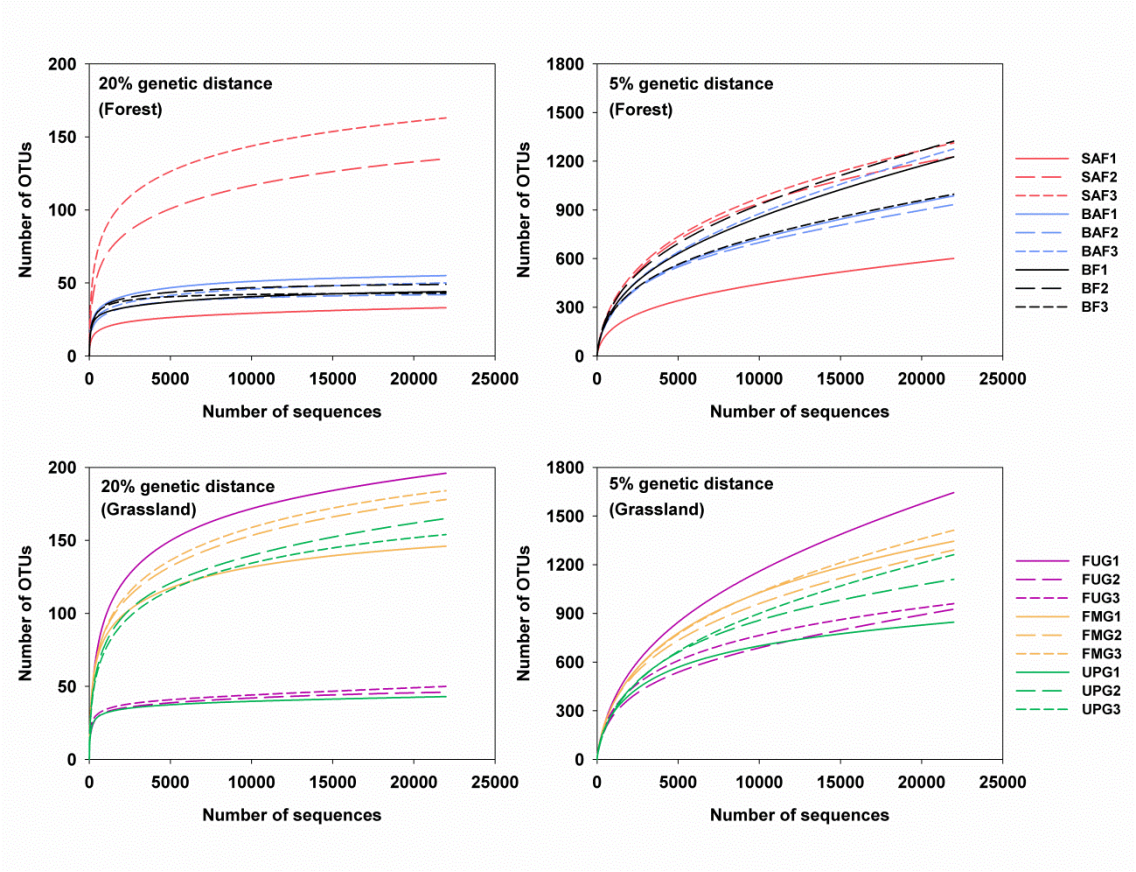


Fig. S1. Rarefaction curves indicating the observed number of operational taxonomic units (OTUs) at a genetic distance of 5% and 20% in different forest and grassland soils. The spruce age class forest (SAF1-3), beech age class forest (BAF1-3), and unmanaged beech forest (BF1-3) sampling sites are marked by the red, blue, and black color, respectively. The fertilized intensely managed grassland (FUG1-3), fertilized meadow grazed by horse and cattle (FMG1-3), and unfertilized pasture grazed by sheep (UPG1-3) sampling sites are shown in purple, orange, and green, respectively.

Table S1. Localization of the sampling sites and number of 16S rRNA gene sequences derived from the analyzed grassland and forest soil samples.

Management type	Sample	Localization of the plots (deg min sec)		No. of sequences ≥ 200 bp	No. of sequences assigned to domain Bacteria	No. of sequences classified below domain level
		Latitude (N)	Longitude (E)			
Spruce age class forest	SAF1	48 28 41.0628	9 20 3.876	37,861	37,861	33,562
Spruce age class forest	SAF2	48 22 47.9634	9 21 5.2236	25,987	25,974	22,399
Spruce age class forest	SAF3	48 24 44.1468	9 21 20.1276	34,903	34,812	27,681
Beech age class forest	BAF1	48 23 56.7558	9 14 41.3772	31,373	31,372	23,723
Beech age class forest	BAF2	48 25 10.6242	9 24 52.8552	34,095	34,092	27,518
Beech age class forest	BAF3	48 23 38.583	9 26 45.3732	38,580	38,580	29,621
Unmanaged beech forest	BF1	48 23 46.4562	9 15 40.881	33,298	33,298	26,859
Unmanaged beech forest	BF2	48 22 57.324	9 22 56.5824	35,373	35,372	29,228
Unmanaged beech forest	BF3	48 22 9.6456	9 24 54.7842	33,974	33,974	26,839
Fertilized intensely managed grassland	FUG1	48 23 52.8174	9 20 31.1526	29,890	29,865	23,444
Fertilized intensely managed grassland	FUG2	48 22 36.6852	9 28 22.0224	31,483	31,481	23,806
Fertilized intensely managed grassland	FUG3	48 24 31.9716	9 31 56.5644	23,519	23,515	18,173
Fertilized mown pasture, horse and cattle	FMG1	48 22 51.1962	9 25 8.0004	30,498	30,471	22,617
Fertilized mown pasture, horse and cattle	FMG2	48 23 45.1536	9 26 21.1236	37,798	37,745	28,060
Fertilized mown pasture, horse and cattle	FMG3	48 24 4.5432	9 26 30.0402	39,273	39,259	28,413
Unfertilized pasture, sheep	UPG1	48 23 29.1156	9 22 36.6486	26,510	26,500	20,761
Unfertilized pasture, sheep	UPG2	48 25 21.504	9 29 31.6494	36,854	36,790	30,888
Unfertilized pasture, sheep	UPG3	48 23 40.8156	9 30 10.0506	38,015	38,001	31,276

Table S2. Mean values of soil properties (± Standard deviation) for each management type and ANOVA *P* values.

Soil property	Management type						ANOVA
	SAF	BAF	BF	FUG	FMG	UPG	*P*
pH	4.7±0.9 b	5.9±0.9 ab	5.9±0.8 ab	6.7±0.3 a	6.1±0.7 ab	6.9±0.4 a	0.011
Organig C (g kg^{-1})	68.2±5.7 a	58.4±19.7 a	80.9±22.7 a	67.7±12.3 a	68.3±16.8 a	63.6±20.7 a	0.731
Total N (g kg^{-1})	4.5±0.6 a	4.5±1.4 a	5.6±1.1 a	6.7±1.3 a	6.6±1.3 a	5.6±1.9 a	0.247
Sand (g kg^{-1})	43±17.0 a	75±30.3 a	66±38.6 a	67±62.4 a	56±24 a	115±145.5 a	0.967
Silt (g kg^{-1})	520±128.5 a	565±27.8 a	387±100.5 a	546±98.5 a	554±121.8 a	533±150.0 a	0.436
Clay (g kg^{-1})	437±114.0 a	361±39.5 a	546±112.8 a	387±158.0 a	390±122.5 a	353±216.6 a	0.559

Differences of soil properties between management types were analyzed by employing one-way analysis of variance and Tukey pair-wise comparisons. Significant ANOVA *P* values are shown in bold ($P < 0.05$). Figures followed by different letters indicate differences among management types ($P < 0.05$). Abbreviations: SAF, spruce age class forest; BAF, beech age class forest; BF, unmanaged beech forest; FUG, fertilized intensely managed grassland; FMG, fertilized meadow grazed by horse and cattle; UPG: unfertilized pasture grazed by sheep. Complete soil and site information for all 18 sample sites is provided in Table 1.

Table S3. Bacterial diversity as assessed by Shannon index (H´) and species richness estimation in all forest and grassland soils. The results from the rarefaction analyses are also depicted in Figure 1 and Figure S1.

Management type	Sample	Genetic distance (%)	H´	Rarefaction	Chao1	ACE
				No. of operational taxonomic units		
Spruce age class forest	SAF1	3	4.74	810	1625	1608
		5	4.21	601	1119	1108
		20	1.62	33	37	36
Spruce age class forest	SAF2	3	5.75	1509	2214	1924
		5	5.40	1227	1786	1781
		20	2.98	135	149	154
Spruce age class forest	SAF3	3	5.81	1584	2745	2735
		5	5.55	1313	2137	2144
		20	3.66	163	193	183
Beech age class forest	BAF1	3	5.55	1192	3112	3076
		5	5.21	987	2328	2305
		20	2.50	55	60	58
Beech age class forest	BAF2	3	5.46	1134	3103	3056
		5	5.13	933	2399	2366
		20	2.06	42	42	43
Beech age class forest	BAF3	3	5.87	1669	4703	4669
		5	5.37	1276	3127	3107
		20	2.26	50	53	53
Unmanaged beech forest	BF1	3	5.68	1594	4050	4024
		5	5.23	1227	2751	2735
		20	2.00	44	45	46
Unmanaged beech forest	BF2	3	5.99	1734	4056	4033
		5	5.52	1324	3072	3054
		20	2.22	49	49	50
Unmanaged beech forest	BF3	3	5.66	1254	3366	3326
		5	5.24	997	2263	2242
		20	2.33	43	43	43

Table S3 (continued)

Management type	Sample	Genetic distance (%)	H´	Rarefaction	Chao1	ACE
				No. of operational taxonomic units		
Fertilized intensely managed grassland	FUG1	3	5.86	1960	4203	4188
		5	5.55	1645	3236	3224
		20	3.47	196	234	221
Fertilized intensely managed grassland	FUG2	3	5.46	1134	2807	2773
		5	5.10	926	2163	2139
		20	2.35	46	48	47
Fertilized intensely managed grassland	FUG3	3	5.92	1301	1652	1524
		5	5.52	960	1359	1353
		20	2.60	50	100	73
Fertilized mown pasture, horse and cattle	FMG1	3	5.79	1580	2199	2017
		5	5.54	1345	1912	1907
		20	3.67	146	156	157
Fertilized mown pasture, horse and cattle	FMG2	3	5.71	1498	2938	2923
		5	5.53	1291	2320	2309
		20	3.60	178	207	205
Fertilized mown pasture, horse and cattle	FMG3	3	5.60	1648	3022	3011
		5	5.37	1413	2561	2551
		20	3.50	184	202	209
Unfertilized pasture, sheep	UPG1	3	5.64	1078	1320	1207
		5	5.35	846	1119	998
		20	2.36	43	56	48
Unfertilized pasture, sheep	UPG2	3	4.96	1302	1945	1644
		5	4.79	1110	1681	1674
		20	3.23	165	201	198
Unfertilized pasture, sheep	UPG3	3	4.99	1482	3413	3394
		5	4.81	1262	2507	2495
		20	3.10	154	171	171

Table S4. Relative abundances of bacterial phyla and proteobacterial classes in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could not be assigned to a specific phylum or a proteobacterial class.

Phylogenetic group	Relative abundance (%)									
	Average	Spruce age class forests			Beech age class forests			Unmanaged beech forests		
		SAF1	SAF2	SAF3	BAF1	BAF2	BAF3	BF1	BF2	BF3
Proteobacteria	45.539	55.355	50.204	38.366	38.152	45.609	38.955	49.018	51.040	43.477
Alphaproteobacteria	25.072	42.759	33.603	18.272	14.338	22.480	19.453	28.957	26.515	19.388
Acidobacteria	20.391	22.852	23.081	23.661	20.404	19.981	20.910	15.325	15.868	21.725
*Bacteria**	18.964	11.355	13.764	20.484	24.382	19.283	23.222	19.337	17.370	21.001
Actinobacteria	12.655	9.165	11.408	15.253	13.550	12.871	13.735	14.052	14.104	9.690
*Proteobacteria**	7.155	3.056	11.246	12.171	6.732	6.650	6.353	5.925	7.633	5.972
Betaproteobacteria	5.991	3.735	2.476	5.191	8.992	6.482	5.534	5.220	8.187	7.738
Deltaproteobacteria	4.356	1.138	0.970	1.454	4.568	7.257	5.568	6.271	5.335	6.125
Gammaproteobacteria	2.966	4.667	1.910	1.278	3.522	2.740	2.048	2.646	3.370	4.253
WS3	0.728	0.005	0.046	0.273	1.540	0.930	1.861	0.478	0.209	1.071
Firmicutes	0.562	0.409	0.354	0.942	0.583	0.170	0.591	1.078	0.271	0.636
TM7	0.450	0.642	0.936	0.566	0.236	0.487	0.254	0.288	0.269	0.474
Chloroflexi	0.285	0.042	0.073	0.215	0.462	0.126	0.257	0.180	0.424	0.774
Bacteroidetes	0.249	0.111	0.058	0.138	0.478	0.238	0.104	0.105	0.263	0.756
Verrucomicrobia	0.067	0.008	0.008	0.066	0.153	0.094	0.034	0.012	0.090	0.141
Fibrobacteres	0.049	0.029	0.000	0.000	0.013	0.100	0.047	0.096	0.062	0.085
Cyanobacteria	0.035	0.021	0.042	0.032	0.026	0.032	0.021	0.012	0.014	0.124
Spirochaetes	0.019	0.000	0.008	0.000	0.016	0.076	0.003	0.015	0.011	0.041
Gemmatimonadetes	0.003	0.005	0.015	0.003	0.000	0.003	0.000	0.000	0.003	0.000
Planctomycetes	0.002	0.000	0.004	0.000	0.000	0.000	0.008	0.003	0.000	0.000
OP11	0.001	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.003

Table S5. Relative abundances of bacterial phyla and proteobacterial classes in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could not be assigned to a specific phylum or a proteobacterial class.

Phylogenetic group	Relative abundance (%)									
	Average	Fertilized intensely managed grassland			Fertilized mown pasture, horse and cattle			Unfertilized pasture, sheep		
		FUG1	FUG2	FUG3	FMG1	FMG2	FMG3	UPG1	UPG2	UPG3
Proteobacteria	34.863	26.044	26.028	31.218	31.906	34.699	22.318	31.989	50.886	53.356
*Bacteria**	22.542	21.500	24.380	22.717	25.775	25.659	27.627	21.657	16.042	17.697
Actinobacteria	19.625	27.420	29.796	20.174	10.722	15.949	24.374	21.460	16.167	12.684
Acidobacteria	18.710	22.873	15.946	18.150	26.835	19.571	21.572	17.306	13.226	14.034
*Proteobacteria**	12.437	7.604	3.888	3.776	13.190	14.428	9.361	3.921	24.224	23.799
Alphaproteobacteria	11.434	8.321	8.758	8.930	10.768	10.862	6.765	8.830	17.374	19.636
Betaproteobacteria	5.863	7.504	6.953	7.272	5.274	5.595	3.739	6.408	5.624	5.584
Gammaproteobacteria	2.743	2.032	2.271	3.934	1.303	2.967	1.409	4.751	3.284	3.339
Deltaproteobacteria	2.387	0.583	4.158	7.306	1.372	0.845	1.044	8.079	0.381	0.997
Firmicutes	1.845	0.837	1.992	2.760	2.822	1.674	1.752	2.366	2.324	0.603
Bacteroidetes	0.723	0.254	0.480	1.999	0.069	0.906	0.423	2.317	0.495	0.263
WS3	0.459	0.161	0.305	1.050	0.791	0.376	0.540	0.992	0.068	0.192
TM7	0.371	0.281	0.289	0.332	0.325	0.215	0.515	0.577	0.277	0.526
Chloroflexi	0.338	0.238	0.235	0.383	0.450	0.551	0.410	0.351	0.098	0.321
Cyanobacteria	0.237	0.208	0.219	0.391	0.236	0.241	0.196	0.321	0.193	0.205
Verrucomicrobia	0.226	0.171	0.302	0.515	0.046	0.146	0.245	0.453	0.204	0.097
Fibrobacteres	0.034	0.000	0.025	0.213	0.000	0.000	0.000	0.162	0.000	0.000
Gemmatimonadetes	0.016	0.010	0.003	0.077	0.016	0.011	0.013	0.015	0.005	0.013
Spirochaetes	0.010	0.003	0.000	0.021	0.007	0.003	0.013	0.034	0.005	0.008
Deinococcus-Thermus	0.001	0.000	0.000	0.000	0.000	0.000	0.008	0.003	0.000	0.000
Fusobacteria	0.0003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000

Table S6. Relative abundances of acidobacterial subgroups in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could be assigned to the phylum level only.

Phylogenetic group	Relative abundance (%)									
	Average	Spruce age class forests			Beech age class forests			Unmanaged beech forests		
		SAF1	SAF2	SAF3	BAF1	BAF2	BAF3	BF1	BF2	BF3
Acidobacteria	**20.391**	**22.852**	**23.081**	**23.661**	**20.404**	**19.981**	**20.910**	**15.325**	**15.868**	**21.725**
Gp3	7.024	10.937	12.705	10.606	1.348	8.861	6.115	6.784	3.684	2.796
Gp16	2.951	0.103	2.495	4.010	5.049	2.256	3.595	2.204	3.017	4.080
Gp6	2.942	0.127	0.574	1.703	8.068	1.596	1.903	1.697	4.388	6.676
Gp1	2.931	10.626	5.532	3.143	0.064	2.443	1.063	1.535	1.179	0.439
Gp4	2.121	0.164	0.543	2.172	2.630	1.296	4.217	1.009	1.563	5.101
Gp7	1.145	0.021	0.562	1.129	1.001	2.074	2.447	1.030	0.958	0.889
*Acidobacteria**	0.330	0.449	0.377	0.149	0.354	0.422	0.384	0.291	0.263	0.280
Gp17	0.289	0.008	0.096	0.293	0.771	0.170	0.280	0.126	0.300	0.580
Gp5	0.201	0.032	0.108	0.178	0.233	0.340	0.241	0.258	0.136	0.286
Gp11	0.168	0.016	0.015	0.118	0.434	0.150	0.308	0.120	0.102	0.238
Gp22	0.160	0.005	0.012	0.092	0.360	0.138	0.228	0.150	0.195	0.253
Gp10	0.037	0.011	0.000	0.006	0.032	0.070	0.044	0.060	0.051	0.056
Gp2	0.030	0.158	0.015	0.020	0.006	0.041	0.005	0.009	0.000	0.000
Gp13	0.026	0.114	0.012	0.011	0.000	0.041	0.013	0.024	0.008	0.000
Gp15	0.022	0.026	0.035	0.020	0.010	0.038	0.041	0.006	0.011	0.006
Gp25	0.007	0.000	0.000	0.003	0.010	0.012	0.008	0.009	0.006	0.018
Gp18	0.006	0.000	0.000	0.009	0.019	0.000	0.000	0.006	0.006	0.015
Gp14	0.006	0.045	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Gp12	0.004	0.008	0.000	0.000	0.000	0.015	0.013	0.000	0.000	0.000
Gp20	0.003	0.003	0.000	0.000	0.010	0.000	0.003	0.003	0.003	0.009
Gp8	0.002	0.000	0.000	0.000	0.000	0.018	0.000	0.003	0.000	0.000
Gp19	0.001	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.006
Gp9	0.001	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.000	0.000

Table S7. Relative abundances of acidobacterial subgroups in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could be assigned to the phylum level only.

Phylogenetic group	Relative abundance (%)									
	Average	Fertilized intensely managed grassland			Fertilized mown pasture, horse and cattle			Unfertilized pasture, sheep		
		FUG1	FUG2	FUG3	FMG1	FMG2	FMG3	UPG1	UPG2	UPG3
Acidobacteria	**18.710**	**22.873**	**15.946**	**18.150**	**26.835**	**19.571**	**21.572**	**17.306**	**13.226**	**14.034**
Gp16	6.798	9.925	4.380	3.321	9.137	8.197	11.816	3.925	4.629	4.147
Gp6	4.408	8.331	7.967	5.192	1.549	3.624	3.724	5.113	2.422	3.100
Gp4	2.798	1.771	1.372	3.593	3.669	2.567	2.331	3.638	3.025	3.505
Gp3	1.795	0.717	0.365	1.535	6.390	2.090	1.516	1.117	1.188	1.358
Gp7	1.386	1.005	0.588	2.135	3.771	1.258	0.963	1.721	0.889	0.784
Gp17	0.460	0.321	0.394	0.863	0.289	0.440	0.469	0.668	0.321	0.516
Gp11	0.304	0.321	0.337	0.315	0.269	0.400	0.252	0.238	0.307	0.287
Gp5	0.221	0.080	0.098	0.621	0.167	0.400	0.143	0.275	0.247	0.071
Gp1	0.183	0.003	0.000	0.166	1.267	0.064	0.059	0.102	0.024	0.076
Gp22	0.160	0.167	0.108	0.162	0.062	0.368	0.176	0.275	0.060	0.066
*Acidobacteria**	0.122	0.177	0.254	0.140	0.167	0.053	0.076	0.136	0.043	0.100
Gp18	0.024	0.010	0.029	0.034	0.010	0.042	0.015	0.038	0.030	0.011
Gp25	0.017	0.020	0.016	0.021	0.030	0.008	0.005	0.030	0.016	0.013
Gp20	0.016	0.010	0.019	0.000	0.026	0.048	0.010	0.008	0.019	0.000
Gp10	0.006	0.003	0.010	0.043	0.000	0.000	0.000	0.015	0.000	0.000
Gp15	0.005	0.000	0.000	0.009	0.016	0.005	0.015	0.000	0.003	0.000
Gp13	0.003	0.003	0.006	0.000	0.010	0.005	0.003	0.004	0.000	0.000
Gp9	0.002	0.007	0.003	0.000	0.000	0.000	0.000	0.004	0.003	0.000
Gp8	0.001	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000

Table S8. Relative abundances of taxonomic groups within the phylum *Actinobacteria* and within proteobacterial classes in the analyzed forest soils. Values represent percentages of all sequences assigned to the domain Bacteria for all forest soils or individual forest soils. Groups labeled with asterisks could be assigned to the phylum level only.

Phylogenetic group	Relative abundance (%)									
	Average	Spruce age class forests			Beech age class forests			Unmanaged beech forests		
		SAF1	SAF2	SAF3	BAF1	BAF2	BAF3	BF1	BF2	BF3
Actinobacteria	**12.655**	**9.165**	**11.408**	**15.253**	**13.550**	**12.871**	**13.735**	**14.052**	**14.104**	**9.690**
Actinobacteridae	10.858	6.360	11.173	14.860	11.211	11.114	11.659	12.367	11.874	7.497
*Actinobacteria**	1.658	2.697	0.100	0.221	2.180	1.590	1.947	1.589	2.069	2.060
Rubrobacteridae	0.139	0.103	0.135	0.172	0.159	0.167	0.130	0.096	0.161	0.132
Acidimicrobidae	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Alphaproteobacteria	**25.072**	**42.759**	**33.603**	**18.272**	**14.338**	**22.480**	**19.453**	**28.957**	**26.515**	**19.388**
*Alphaproteobacteria**	10.729	10.924	8.628	6.788	7.988	11.492	11.801	12.451	14.271	11.323
Rhodospirillales	7.106	20.359	14.191	6.256	1.093	4.376	2.934	8.475	4.577	2.107
Caulobacterales	3.904	9.318	7.238	1.646	1.097	3.605	1.796	4.742	3.514	2.502
Rhizobiales	3.234	2.155	3.469	3.433	4.004	2.919	2.869	3.222	3.983	3.303
Rhodobacterales	0.081	0.003	0.073	0.147	0.124	0.053	0.034	0.048	0.141	0.115
Other	0.016	0.000	0.004	0.003	0.000	0.018	0.013	0.015	0.011	0.021
Betaproteobacteria	**5.991**	**3.735**	**2.476**	**5.191**	**8.992**	**6.482**	**5.534**	**5.220**	**8.187**	**7.738**
*Betaproteobacteria**	2.981	1.706	0.805	2.407	4.906	2.957	3.258	2.505	3.729	4.271
Burkholderiales	2.886	2.013	1.602	2.775	3.777	3.394	2.182	2.568	4.266	3.300
Other	0.124	0.016	0.069	0.009	0.309	0.132	0.093	0.147	0.192	0.168
Gammaproteobacteria	**2.966**	**4.667**	**1.910**	**1.278**	**3.522**	**2.740**	**2.048**	**2.646**	**3.370**	**4.253**
*Gammaproteobacteria**	1.940	2.242	0.616	0.755	2.783	1.980	1.369	2.039	2.242	3.247
Pseudomonadales	0.569	1.144	0.501	0.169	0.555	0.525	0.568	0.294	0.670	0.612
Xanthomonadales	0.267	0.977	0.628	0.190	0.019	0.103	0.039	0.120	0.221	0.127
Other	0.190	0.304	0.166	0.164	0.166	0.132	0.073	0.192	0.237	0.268
Deltaproteobacteria	**4.356**	**1.138**	**0.970**	**1.454**	**4.568**	**7.257**	**5.568**	**6.271**	**5.335**	**6.125**
Myxococcales	3.354	0.961	0.516	1.048	3.232	5.960	4.150	4.922	4.252	4.671
*Deltaproteobacteria**	0.969	0.169	0.262	0.399	1.316	1.258	1.407	1.333	1.069	1.416
Other	0.033	0.008	0.193	0.006	0.019	0.038	0.010	0.015	0.014	0.038

Table S9. Relative abundances of taxonomic groups within the phylum *Actinobacteria* and within proteobacterial classes in the analyzed grassland soils. Values represent percentages of all sequences assigned to the domain Bacteria for all grassland soils or individual grassland soils. Groups labeled with asterisks could be assigned to the phylum level only.

Phylogenetic group	Relative abundance (%)									
	Average	Fertilized intensely managed grassland			Fertilized mown pasture, horse and cattle			Unfertilized pasture, sheep		
		FUG1	FUG2	FUG3	FMG1	FMG2	FMG3	UPG1	UPG2	UPG3
Actinobacteria	**19.625**	**27.420**	**29.796**	**20.174**	**10.722**	**15.949**	**24.374**	**21.460**	**16.167**	**12.684**
Actinobacteridae	17.595	26.322	23.719	14.497	10.374	15.451	23.668	16.106	15.515	12.281
*Actinobacteria**	1.728	0.573	5.387	5.252	0.282	0.363	0.555	4.864	0.367	0.279
Rubrobacteridae	0.281	0.465	0.686	0.425	0.062	0.125	0.135	0.491	0.234	0.095
Acidimicrobidae	0.015	0.060	0.000	0.000	0.003	0.008	0.015	0.000	0.016	0.029
Coriobacteridae	0.005	0.000	0.003	0.000	0.000	0.003	0.000	0.000	0.035	0.000
Alphaproteobacteria	**11.434**	**8.321**	**8.758**	**8.930**	**10.768**	**10.862**	**6.765**	**8.830**	**17.374**	**19.636**
*Alphaproteobacteria**	5.737	2.726	4.056	3.849	5.192	4.581	2.838	3.936	9.508	12.850
Rhizobiales	3.115	3.817	2.935	2.173	2.251	3.245	2.308	2.457	4.142	4.158
Caulobacterales	1.390	0.824	0.670	1.314	1.474	1.539	0.415	1.117	2.748	2.145
Rhodospirillales	0.816	0.378	0.756	1.301	1.684	1.062	0.909	0.940	0.359	0.232
Rhodobacterales	0.354	0.559	0.327	0.259	0.154	0.419	0.283	0.351	0.576	0.226
Other	0.022	0.017	0.012	0.034	0.013	0.016	0.013	0.031	0.041	0.026
Betaproteobacteria	**5.863**	**7.504**	**6.953**	**7.272**	**5.274**	**5.595**	**3.739**	**6.408**	**5.624**	**5.584**
Burkholderiales	4.282	6.553	4.622	4.257	3.029	4.125	2.838	3.894	4.795	4.658
*Betaproteobacteria**	1.465	0.847	2.011	2.909	1.979	1.367	0.825	2.445	0.818	0.892
Other	0.116	0.104	0.321	0.106	0.266	0.103	0.076	0.068	0.011	0.034
Gammaproteobacteria	**2.743**	**2.032**	**2.271**	**3.934**	**1.303**	**2.967**	**1.409**	**4.751**	**3.284**	**3.339**
*Gammaproteobacteria**	1.791	1.122	1.544	2.717	0.748	2.143	1.032	3.479	2.088	1.758
Pseudomonadales	0.631	0.720	0.524	0.642	0.371	0.487	0.224	0.657	0.628	1.400
Other	0.177	0.137	0.165	0.374	0.089	0.143	0.107	0.434	0.139	0.129
Xanthomonadales	0.143	0.054	0.038	0.200	0.095	0.193	0.046	0.181	0.429	0.053
Deltaproteobacteria	**2.387**	**0.583**	**4.158**	**7.306**	**1.372**	**0.845**	**1.044**	**8.079**	**0.381**	**0.997**
Myxococcales	1.321	0.285	2.935	1.761	0.545	0.419	0.644	6.268	0.188	0.389
*Deltaproteobacteria**	0.978	0.275	1.191	5.426	0.541	0.363	0.242	1.713	0.182	0.579
Other	0.088	0.023	0.032	0.119	0.286	0.064	0.158	0.098	0.011	0.029

Table S10. Spearman's rank correlations between relative abundances of *Acidobacteria* subgroups and soil properties. Only relative abundances of acidobacterial subgroups that represented ≥ 0.029% of all analyzed sequences were considered.

Acidobacteria subgroup	Correlation			
	pH	Organic C	Total N	Sand/Silt/Clay
1	**-0.87**	-0.32	**-0.71**	0.02/-0.26/0.23
2	None	None	None	None
3	**-0.95**	-0.21	**-0.56**	-0.07/-0.22/0.29
4	**0.49**	-0.40	-0.11	0.20/0.04/-0.11
5	0.01	-0.16	-0.12	0.16/-0.15/0.12
6	**0.80**	0.24	**0.47**	0.08/0.11/-0.12
7	-0.07	-0.40	-0.19	0.35/-0.06/-0.04
10	-0.23	-0.18	-0.39	0.25/-0.13/0.10
11	**0.67**	0.15	**0.54**	0.14/0.43/-0.34
13	**-0.75**	-0.20	-0.45	0.25/0.02/-0.09
16	**0.60**	0.22	**0.64**	0.01/0.18/-0.15
17	**0.77**	0.03	0.30	0.02/0.14/-0.16
18	**0.76**	0.20	**0.49**	0.05/0.12/-0.19
22	0.41	0.09	0.15	0.34/0.15/-0.22

Bold numbers: $P < 0.05$; Bold and underlined numbers: $P < 0.001$. None: subgroup 2 was not detected in grassland.

Table S11. Dominant grasses of the analyzed grassland sites.

Management type	Sample	Dominant grasses
Fertilized intensely managed grassland	FUG1	*Arrhenaterum elatius*, *Trisetum flavescens*, *Poa trivialis*
Fertilized intensely managed grassland	FUG2	*Poa trivialis*, *Trisetum flavescens*
Fertilized intensely managed grassland	FUG3	*Poa trivialis*, *Trisetum flavescens*
Fertilized mown pasture, horse and cattle	FMG1	*Poa trivialis*, *Alopecurus pratensis*, *Festuca pratensis*, *Lolium perenne*
Fertilized mown pasture, horse and cattle	FMG2	*Poa trivialis*, *Trisetum flavescens*, *Alopecurus pratensis*, *Dactylis glomerata*
Fertilized mown pasture, horse and cattle	FMG3	*Dactylis glomerata*, *Arrhenaterum elatius*, *Trisetum flavescens*, *Poa trivialis*, *Alopecurus pratensis*
Unfertilized pasture, sheep	UPG1	*Brachypodium pinnatum*, *Festuca guestfalica*
Unfertilized pasture, sheep	UPG2	*Bromus erectus*
Unfertilized pasture, sheep	UPG3	*Bromus erectus*, *Brachypodium pinnatum*

2

Horizon-specific bacterial community composition of German grassland soils as revealed by pyrosequencing-based analysis of 16S rRNA genes

Christiane Will[1], Andrea Thürmer[2], Antje Wollherr[2], Heiko Nacke[1], Nadine Herold[3], Marion Schrumpf[3], Jessica Gutknecht[4], Tesfaye Wubet[4], François Buscot[4], and Rolf Daniel[1,2]

Applied and Environmental Microbiology (2010), Vol. 76, p. 6751-6759

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [3]Max Planck Institute for Biogeochemistry, Jena, Germany, [4]UFZ-Helmholtz Centre for Environmental Research, Department of Soil Ecology, Halle/Saale, Germany

Author contributions to the work:

Performed the experiments: CW, AT, HN.

Analyzed data: CW.

Bioinformatic support: AW.

Contributed data on soil properties and analysis of these data: NH, MS.

Contributed data on microbial biomass and analysis of these data: JG, TW, FB.

Wrote the paper: CW, RD.

Conceived and designed the experiments: RD.

Applied and Environmental Microbiology, Oct. 2010, p. 6751–6759 Vol. 76, No. 20
0099-2240/10/$12.00 doi:10.1128/AEM.01063-10

Horizon-Specific Bacterial Community Composition of German Grassland Soils, as Revealed by Pyrosequencing-Based Analysis of 16S rRNA Genes▽†

Christiane Will,[1] Andrea Thürmer,[2] Antje Wollherr,[2] Heiko Nacke,[1] Nadine Herold,[3] Marion Schrumpf,[3] Jessica Gutknecht,[4] Tesfaye Wubet,[4] François Buscot,[4] and Rolf Daniel[1,2]*

Department of Genomic and Applied Microbiology[1] and Göttingen Genomics Laboratory,[2] Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany; Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, D-07745 Jena, Germany[3]; and UFZ-Helmholtz Centre for Environmental Research, Department of Soil Ecology, Theodor-Lieser-Str. 4, D-06120 Halle/Saale, Germany[4]

Received 3 May 2010/Accepted 9 August 2010

The diversity of bacteria in soil is enormous, and soil bacterial communities can vary greatly in structure. Here, we employed a pyrosequencing-based analysis of the V2-V3 16S rRNA gene region to characterize the overall and horizon-specific (A and B horizons) bacterial community compositions in nine grassland soils, which covered three different land use types. The entire data set comprised 752,838 sequences, 600,544 of which could be classified below the domain level. The average number of sequences per horizon was 41,824. The dominant taxonomic groups present in all samples and horizons were the *Acidobacteria*, *Betaproteobacteria*, *Actinobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Deltaproteobacteria*, *Chloroflexi*, *Firmicutes*, and *Bacteroidetes*. Despite these overarching dominant taxa, the abundance, diversity, and composition of bacterial communities were horizon specific. In almost all cases, the estimated bacterial diversity (H') was higher in the A horizons than in the corresponding B horizons. In addition, the H' was positively correlated with the organic carbon content, the total nitrogen content, and the C-to-N ratio, which decreased with soil depth. It appeared that lower land use intensity results in higher bacterial diversity. The majority of sequences affiliated with the *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Fibrobacteres*, *Firmicutes*, *Spirochaetes*, *Verrucomicrobia*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* were derived from A horizons, whereas the majority of the sequences related to *Acidobacteria*, *Chloroflexi*, *Gemmatimonadetes*, *Nitrospira*, TM7, and WS3 originated from B horizons. The distribution of some bacterial phylogenetic groups and subgroups in the different horizons correlated with soil properties such as organic carbon content, total nitrogen content, or microbial biomass.

Soil is probably the most complex microbial environment on Earth with respect to species richness and community size. The microbial richness in soils exceeds that of other environments (44) and is higher by orders of magnitude than the biodiversity of plants and animals. Cultivated soil or grassland soil contains an estimated 2×10^9 prokaryotic cells per gram (12). Soil microbial communities are an important factor of agriculturally managed systems, as they are responsible for most nutrient transformations in soil and influence the above-ground plant diversity and productivity (53).

To analyze the bacterial community in soils, most approaches target the 16S rRNA gene by PCR amplification and subsequent analysis employing sequencing of clone libraries (10, 24), denaturing gradient gel electrophoresis (DGGE) (38), or terminal restriction fragment length polymorphism (T-RFLP) (17, 52). Most of these approaches provided limited insights into the structure of soil bacterial communities, as the survey sizes and the number of compared sampling sites were small with respect to the enormous bacterial diversity present in different soil samples. For example, the reported clone libraries vary considerably in size, but small sample sizes (500 or fewer 16S rRNA gene sequences) are usually analyzed and employed for the theoretical estimation of species richness (39). This provides snapshots of the predominant bacterial community members, but phylogenetic groups that are present in a low abundance and which may possess important ecosystem functions are not assessed (47). In addition, it has been shown that rich sampling (several thousands of clones) of complex bacterial communities is required to perform robust measurements and estimations of community diversity parameters (37). Thus, the detection bias accompanying analyses of small sample sizes can lead to invalidated assumptions. Genetic profiling techniques such as DGGE and T-RFLP have high-throughput capability. These approaches allow researchers to unravel differences in community structure but are limited for assessing diversity (23, 40). To deeply survey the diversity and the composition of the bacterial communities within different soil samples, large-scale pyrosequencing of partial 16S rRNA

* Corresponding author. Mailing address: Department of Genomic and Applied Microbiology and Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. Phone: 0049-551-393827. Fax: 0049-551-3912181. E-mail: rdaniel@gwdg.de.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 20 August 2010.

TABLE 1. Physical and geochemical characteristics of the soil samples from two different soil horizons

Horizon	Plot	Land use type	pH[a]	Water content (%)[a]	Microbial biomass (nmol PLFA/g dry soil)	OC content (g/kg)[a]	N content (g/kg)[a]	C-to-N ratio[a]	Soil texture (%)		
									Sand	Silt	Clay
A	1	Fertilized meadow	6.63	38.50	81.85	66.20	6.24	10.61	6.8	45.8	47.4
A	2	Fertilized meadow	7.12	25.97	35.83	32.60	3.34	9.75	8.8	37.1	54.1
A	3	Fertilized meadow	7.20	25.49	53.83	26.02	2.90	8.97	5.6	37.2	57.2
A	4	Fertilized mown pasture, cattle	6.49	45.68	131.00	66.95	6.02	11.11	6.7	51.4	41.9
A	5	Fertilized mown pasture, cattle	6.91	35.96	95.19	53.41	5.09	10.50	7.0	46.5	46.5
A	6	Fertilized mown pasture, cattle	6.03	21.22	24.13	14.24	1.63	8.71	7.0	66.4	26.6
A	7	Unfertilized pasture, cattle	6.91	43.06	117.60	70.08	6.36	11.02	6.2	41.7	52.1
A	8	Unfertilized pasture, cattle	6.97	41.45	139.82	74.84	6.90	10.85	6.7	44.4	48.9
A	9	Unfertilized pasture, cattle	6.62	30.94	119.91	48.27	4.13	11.69	7.9	51.1	41.0
B	1	Fertilized meadow	7.10	23.16	27.99	17.77	2.00	8.88	7.0	46.2	46.8
B	2	Fertilized meadow	7.32	22.47	1.34	4.38	0.58	7.54	22.5	27.2	50.3
B	3	Fertilized meadow	7.40	22.14	5.82	7.19	0.99	7.25	2.9	33.0	64.1
B	4	Fertilized mown pasture, cattle	7.35	23.19	9.43	6.13	0.83	7.34	4.4	36.8	58.8
B	5	Fertilized mown pasture, cattle	7.18	22.32	12.87	10.26	1.19	8.62	8.6	53.3	38.1
B	6	Fertilized mown pasture, cattle	6.30	20.10	5.64	4.23	0.66	6.44	5.6	67.7	26.8
B	7	Unfertilized pasture, cattle	7.26	26.29	60.36	34.39	3.62	9.51	5.3	44.7	50.0
B	8	Unfertilized pasture, cattle	7.28	21.75	17.36	19.12	2.22	8.59	11.1	45.6	43.3
B	9	Unfertilized pasture, cattle	7.35	19.94	7.92	5.02	0.63	7.93	8.5	50.4	41.1

[a] Statistically significant differences between the A and B horizons ($P \leq 0.01$).

genes has been employed recently. Previous pyrosequencing-based studies of soil (1, 30, 34, 43) have generated large data sets, which comprised 39,707 (30) to 152,359 (34) 16S rRNA partial gene sequences. Those studies provided comprehensive insights into the biogeography of bacterial soil communities and taxa that were present in a low abundance. However, all those studies focused on the analysis of microbial communities present in topsoil. The subsoil is also known to harbor an important part of the soil microbial biomass (18). It has been shown that the microbial population in the shallow subsurface is impacted by agricultural production to a similar extent as that in topsoil (5).

In this study, we performed large-scale pyrosequencing-based analyses of 16S rRNA genes to assess the bacterial community composition in topsoil and the corresponding subsoil of nine different grassland sites in the Hainich region (Thuringia, Germany). To provide a high level of coverage at the species level (97% genetic distance) and minimize detection bias, we exceeded the above-described numbers of analyzed 16S rRNA gene sequences (752,838 in this study). To examine the impact of land use on bacterial diversity and community composition, the selected grassland sites covered a range of three different land use types, including samples from unfertilized pastures grazed by cattle, fertilized mown pastures grazed by cattle, and fertilized meadows. In many recent studies, surveys were focused on comprehensive analyses of a single soil or a few soil samples (1, 14, 37, 43). This allowed the determination of overall bacterial species richness and community composition, but the assessment of spatial patterns and environmental factors that drive these patterns is hampered by the limited number of examined soils. To assess spatial distribution and the impact of soil edaphic factors and land use on community structure, we used triplicate samples of each land use type from different locations. In addition, composite samples derived from five soil cores after the separation of soil horizons were employed.

MATERIALS AND METHODS

Study sites and soil sampling. Soil samples were collected from nine different grassland sites of the Hainich region in Germany, which is located in the west of Thuringia near the border to Hessen (latitude, 51.2167/N 51°13′0″; longitude, 10.45/E 10°27′0″). The Hainich region is one of the three locations investigated within the framework of the German Biodiversity Exploratories initiative (www.biodiversity-exploratories.de). The nine sampling sites encompassed the following three different land use types: fertilized meadow (plots 1 to 3), fertilized mown pasture grazed by cattle (plots 4 to 6), and unfertilized pasture grazed by cattle (plots 7 to 9) (for coordinates, see Table S1 in the supplemental material). Sampling was performed in April and May 2008. At each sampling site, five soil cores (8.3 cm in diameter) were sampled with a motor-driven soil column cylinder at each corner and in the center of the plot within a given area of 20 m by 20 m. The soil was classified using the World Reference Base of Soil Resources (27). The predominant soil type in the studied plots is Stagnosol, which is characterized by a perched water table, strong mottling, and reducing conditions (27). For each soil core, we determined soil horizons according to the *Guidelines for Soil Description* (28). The horizons were homogenized and pooled into one composite sample per plot and horizon. Coarse roots and stones (>5 mm) were removed from the samples. In the majority of the samples, the horizons Ah (topsoil) and Btg (subsoil) were detected. In plots 2 and 3 the top horizon was a transitional horizon (ABth) that was employed instead of the missing Ah horizon. In plots 7 and 8 only a transitional horizon (ABth) between topsoil and parent rock material was present. This horizon was employed instead of the missing Btg horizon. Throughout the study the topsoil horizon and the subsoil horizon were designated horizon A and horizon B, respectively.

Edaphic properties of the soil samples. For determinations of organic carbon (OC) content, total nitrogen (N) content, and soil texture, subsamples from the same composite sample were dried at 40°C and sieved to <2 mm. Total carbon and nitrogen were measured after grinding subsamples to a size of <100 μm in a ball mill. The ground samples were analyzed for total carbon and nitrogen by dry combustion with a Vario Max CN analyzer (Elementar Analysensysteme GmbH, Hanau, Germany). Inorganic carbon was quantified by measuring the total amount of carbon after the removal of organic carbon by the ignition of samples at 450°C for 16 h. To determine soil pH, the subsamples were suspended at a soil-to-liquid ratio of 1:2.5 (soil/0.01 M $CaCl_2$). Subsequently, pH was measured in the supernatant with a glass electrode. Soil texture was determined on 30 g soil according to a method described previously by Schlichting and Blume (45). The edaphic properties are depicted in Table 1.

Determination of microbial biomass. To determine microbial biomass, we performed phospholipid fatty acid analysis (PLFA) on soil samples from the A and B horizons of the sampling sites. The composite samples were kept frozen at −80°C after sampling and freeze-dried prior to PLFA extractions. PLFA extractions were performed by using a modified Bligh and Dyer (4) method. Briefly, 2 g

of freeze-dried sample was extracted twice in a chloroform-methanol-citrate buffer (1:2:0.8), followed by overnight phase separation. Fatty acids in the organic phase were then separated by using a silica-bonded phase column (silica-based solid-phase extraction [SPE-SI] Bond Elut, 3 ml, 500 mg; Varian Inc., Darmstadt, Germany) to remove glycolipids and neutral lipids. The polar lipids were then converted to fatty acid methyl esters by mild alkaline methanolysis. Methyl-esterified fatty acids were analyzed by using a Hewlett-Packard 6890 gas chromatograph equipped with a DB-5MS column (60-m length; Agilent Technologies, Böblingen, Germany) and interfaced with an Agilent 5973 mass selective detector. Peak areas of each lipid were converted to nmol/g soil using internal standards (19:0 nonadecanoic methyl ester). The total nmol lipid/g dry soil (sum of all lipids present, 20 or fewer carbons in length) was used as an index of microbial biomass (19, 25).

DNA extraction, amplification of 16S rRNA genes, and pyrosequencing. Total microbial community DNA was isolated from approximately 10 g of soil per sample. For this purpose, the MoBio Power Max soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA) was used according to the manufacturer's instruction. To analyze the taxonomic composition of the soil bacterial community, the V2-V3 region of the 16S rRNA gene (*Escherichia coli* positions 101 to 536) was chosen for the amplification and subsequent pyrosequencing of the PCR products. The V2-V3 region was amplified with the following primer set, containing the Roche 454 pyrosequencing adaptors (underlined): V2for (5′-GCCTCCCTCGCGCCATCAGAGTGGCGGACGGGTGAGTAA-3′) (modified from that described previously by Schmalenberger et al. [48]) and V3rev (5′-GCCTTGCCAGCCCGCTCAGCGTATTACCGCGGCTGCTG-3′) (7).

For each sample, three independent PCRs were performed. The PCR mixture (final volume, 50 μl) contained 5 μl 10-fold reaction buffer (MBI Fermentas GmbH, St. Leon-Rot, Germany), 30 to 70 ng of soil DNA, 0.4 μM each primer, 0.5 U *Pfu* polymerase (MBI Fermentas), and 800 μM concentration of each of the four deoxynucleoside triphosphates. In some cases, to achieve amplification of 16S rRNA genes, a different DNA polymerase was used as recommended by the manufacturer (PCR Extender system; VWR International, Hannover, Germany). The polymerase was applied to samples derived from the A horizons of plots 2, 3, 4, and 6 and from the B horizons of plots 4, 6, and 8. Negative-control reactions lacked template DNA. The following thermal cycling scheme was used: initial denaturation at 94°C for 2 min and 25 cycles of denaturation at 94°C for 1 min, annealing for 1 min using a temperature gradient ranging from 60.9°C to 68.2°C, and extension at 72°C for 1.25 min, followed by a final extension period at 72°C for 10 min. Subsequently, the three PCR products per soil sample were pooled in equal amounts and purified by employing the peqGOLD gel extraction kit (Peqlab Biotechnologie GmbH, Erlangen, Germany). Quantification of the PCR products was performed by using the Quant-iT dsDNA BR assay kit and a Qubit fluorometer (Invitrogen GmbH, Karlsruhe, Germany) as recommended by the manufacturer. The Göttingen Genomics Laboratory determined the sequences of the partial 16S rRNA genes by employing the Roche GS-FLX 454 pyrosequencer (Roche, Mannheim, Germany) and using picotiter sequencing plates subdivided into 8 parts (1 part per sample). Amplicons were sequenced as recommended in the instructions of the manufacturer for amplicon sequencing.

Analysis of pyrosequencing-derived data. Sequences that were shorter than 200 bp in length or of low quality were removed from the pyrosequencing-derived data sets. For taxonomy-based analysis, the Naïve Bayesian rRNA classifier of the Ribosomal Database Project (RDP) (55; http://rdp.cme.msu.edu/) was used. In this way, a rapid taxonomic classification of large partial and full-length rRNA gene sequence data sets according to the new Bergey's bacterial taxonomy (20) was feasible. The bootstrap value was set to ≥80%. Rarefaction curves were calculated by using the tools Aligner, Complete Linkage Clustering, and Rarefaction of the RDP pyrosequencing pipeline (11). We calculated Shannon (50) and Chao1 (8) indices based on the Complete Linkage Clustering data.

Statistical analyses of the pyrosequencing-derived data were carried out with STATISTICA 8.0 (StatSoft Inc., Tulsa, OK). *P* values of ≤0.05 were considered significant. Significant effects of soil horizon on edaphic soil properties were determined by using the Mann-Whitney U test for nonparametric data. Correlations between phylogenetic groups and soil properties were tested for significance by using Pearson's correlation coefficient. The distributions of phyla between the two soil horizons were calculated by employing the chi-square test. Microbial community compositions were compared to the land use types by using a post hoc Tukey honestly significant difference (HSD) analysis of variance. If the normality test revealed a non-Gaussian distribution, data were transformed.

Nucleotide sequence accession number. The 16S rRNA gene sequences derived from pyrosequencing have been deposited in the NCBI Sequence Read Archive under accession number SRA020168.1.

RESULTS AND DISCUSSION

General characteristics of the soil samples and the pyrosequencing-derived data set. In this study, we assessed and compared the compositions of soil bacterial communities present in the A and B horizons of nine different grassland sites of the Hainich region in Germany by a pyrosequencing-based analysis of the 16S rRNA gene sequences. The grassland sites covered a range of the following three different land use types: fertilized meadow (plots 1 to 3), fertilized mown pasture grazed by cattle (plots 4 to 6), and unfertilized pasture grazed by cattle (plots 7 to 9). The soil type of all samples was Stagnosol, except for plot 1, which was a Vertic Cambisol. In addition, further analysis of the Stagnosols revealed that plot 6 was a Luvic Stagnosol, whereas the other plots were Vertic Stagnosols (see Table S1 in the supplemental material).

The two analyzed soil horizons showed significant differences with respect to edaphic properties such as soil pH, OC content, N content, C-to-N ratio, and water content (Table 1). The pH in the A horizons ranged from 6.03 to 7.20, and the pH in the B horizons ranged from 6.30 to 7.40. In general, the pH value of the B horizon was higher than that of the corresponding A horizon, whereas the water content, the amounts of OC and N, and the C-to-N ratio showed 1.1- to 2.0-fold, 2.0- to 10.9-fold, 1.8- to 7.3-fold, and 1.2- to 1.5-fold decreases with depth, respectively.

The microbial biomass in the B horizons of all samples was lower by 48.7 to 96.3% than that in the corresponding A horizons (Table 1). A decrease in the total microbial biomass with soil depth was previously reported (5, 9, 16, 18). The total microbial biomass was positively correlated with the concentration of OC ($r = 0.88$; $P < 0.01$). This supports the assumption reported previously by Blume et al. (5), that carbon availability is closely associated with microbial biomass. In addition, significant correlations of microbial biomass with the concentration of N ($r = 0.84$; $P < 0.01$) and the C-to-N ratio ($r = 0.89$; $P < 0.01$) were detected.

The pyrosequencing-based analysis of the V2-V3 region of the 16S rRNA gene was employed for assessments of bacterial community compositions from the A and B horizons of the nine sampling sites. Short pyrosequencing reads assess the microbial diversity almost as reliably as near-full-length sequences when appropriate primers are chosen. Primers derived from V2-V3 region of the 16S rRNA gene were shown previously to be suitable for this purpose (36). In addition, this region is the most effective region for universal genus identification (7, 42). Across all 18 samples, we recovered 752,838 quality sequences with a read length of ≥200 bp. The average read length was 262 bp. The number of sequences per sample ranged from 25,851 to 61,366, with an average of 41,824 (see Table S2 in the supplemental material). We were able to classify 600,544 (79.77%) of the quality sequences below the domain level. The percentage of classified 16S rRNA gene sequences was in the range of those of other pyrosequencing-based studies (35), but the average number of sequences per sample and the total number of analyzed sequences exceeded those of other previously reported studies of pyrosequencing-based determinations of soil bacterial community composition (34, 43).

Bacterial richness and diversity indices. To determine rarefaction curves, richness, and diversity, we identified oper-

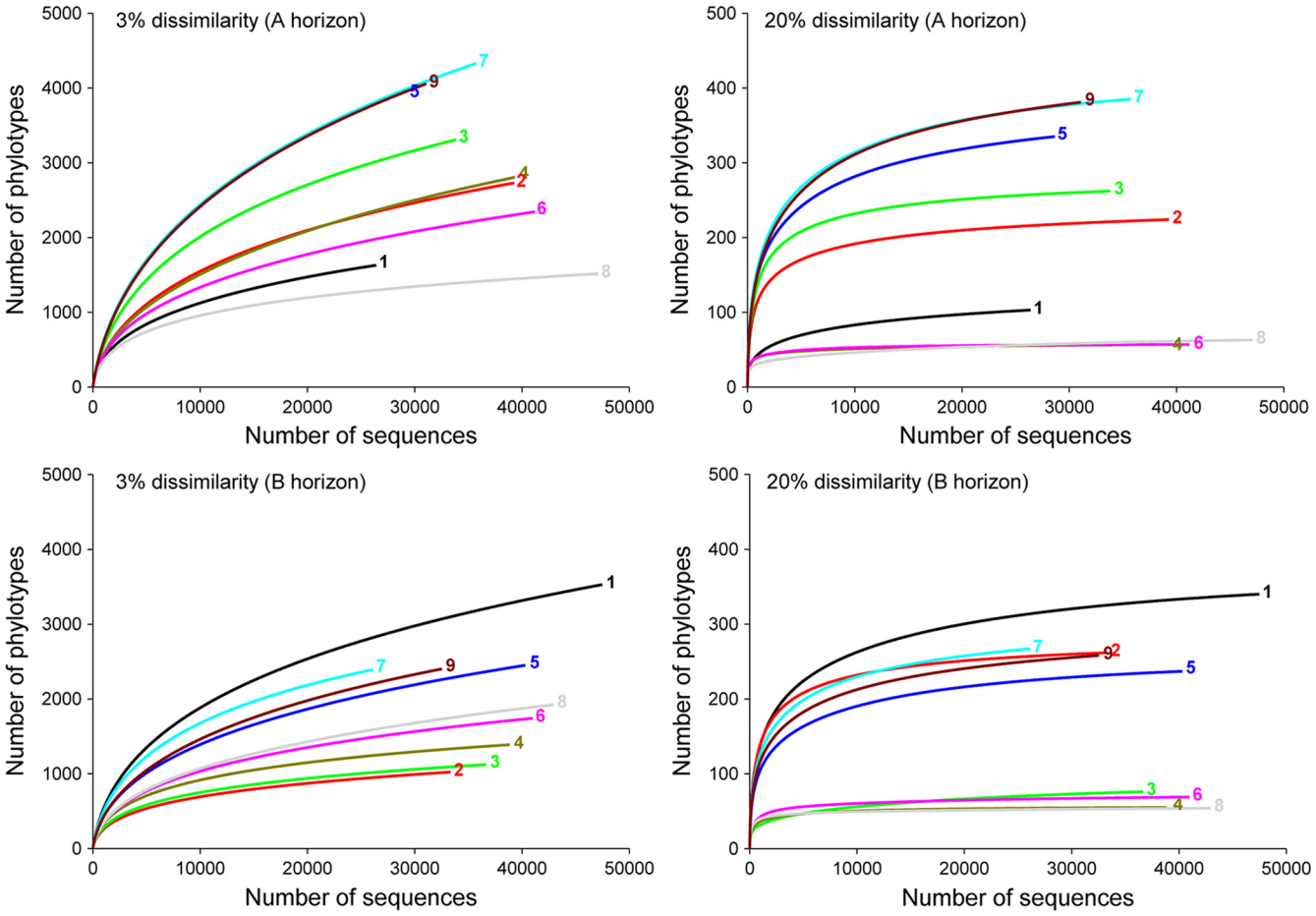


FIG. 1. Rarefaction curves indicating the observed number of OTUs within the 16S rRNA gene sequences of the A and B horizons, derived from nine German grassland sites. OTUs are shown at the 3 and 20% genetic distance levels. The rarefaction curves were calculated by employing the tools Aligner, Complete Linkage Clustering, and Rarefaction of the RDP pyrosequencing pipeline (11). The colored numbers mark the different sampling sites. A description of the sampling sites is given in Table 1.

ational taxonomic units (OTUs) at sequence divergences of 3% (species level) and 20% (phylum level). The rarefaction analysis of bacterial communities derived from the A and B horizons of the nine sampling sites is depicted in Fig. 1. At a 20% genetic distance, almost all curves showed saturation, indicating that the surveying effort covered almost the full extent of taxonomic diversity at this genetic distance. In addition, a comparison of rarefaction analyses with the number of OTUs estimated by the Chao1 richness estimator revealed that 83 to 100% of the estimated taxonomic richness was covered by the sequencing effort. At a 3% genetic distance, the observed richness was 63 to 80% of that predicted by the Chao1 richness estimator (Table 2). Thus, we did not survey the full extent of taxonomic diversity at the species level. Taking into account that at genetic distances below 5%, rarefaction analyses underestimate the bacterial richness whereas Chao1 estimators overestimate it (43), a substantial fraction of the bacterial diversity at the species level was assessed by the surveying effort. It is important that pyrosequencing provides an unprecedented sampling depth compared to that of traditional Sanger sequencing of 16S rRNA genes (51), but the intrinsic error of pyrosequencing could result in the overestimation of rare phylotypes, since each pyrosequencing read is treated as a unique identifier of a community member and correction by assembly and sequencing depth, which is typically applied during genome projects, is not feasible (26, 32). To ensure per-base error rates lower than that of conventional Sanger sequencing, we used quality filtering of the pyrosequencing-derived data set, such as the removal of reads with atypical lengths (26). In addition, to minimize the overestimation of rare phylotypes, clustering and diversity estimates were performed only at genetic divergences of ≥3% (32).

Acosta-Martínez et al. (1) postulated previously that in managed soils, the maximum number of OTUs is less than 3,400 at a genetic distance of 3%. This is in contrast to our results, as up to 4,781 and 6,231 OTUs were predicted for fertilized meadows (plot 3, A horizon) and fertilized mown pasture grazed by cattle (plot 5, A horizon), respectively (Table 2). The differences in the results might be explained by the different surveying efforts. Several studies (13, 37, 43, 46, 56) showed that the number of analyzed sequences per sample has an effect on the predicted number of OTUs. For example, Roesch et al. (43) previously plotted the number of observed OTUs against the sequencing effort using the bacterial community present in a

TABLE 2. Species richness estimates obtained at genetic distances of 3% and 20%[c]

Horizon	Plot	Shannon index (H')[a]		Rarefaction (no. of OTUs)		Chao1[b] (no. of OTUs)		Coverage (%)	
		3%	20%	3%	20%	3%	20%	3%	20%
A	1	5.92	2.67	1,629	103	2,335	124	69.8	83.1
A	2	6.11	4.03	2,730	224	4,084	236	66.9	94.8
A	3	6.76	4.27	3,307	262	4,781	272	69.2	96.2
A	4	6.30	2.93	2,805	57	4,395	59	63.8	96.6
A	5	7.07	4.49	3,937	335	6,231	366	63.2	91.6
A	6	6.15	2.91	2,344	57	3,551	57	66.0	100
A	7	7.16	4.50	4,329	385	6,487	407	66.7	94.7
A	8	5.65	2.51	1,516	63	1,924	68	78.8	93.3
A	9	7.05	4.41	4,056	381	6,232	438	65.1	87.0
B	1	6.72	4.34	3,528	340	5,168	360	68.3	94.5
B	2	5.01	2.64	1,022	84	1,399	94	73.0	89.6
B	3	5.14	2.50	1,122	76	1,509	82	74.4	92.6
B	4	5.57	2.72	1,388	55	1,745	55	79.5	99.4
B	5	6.15	3.98	2,450	237	3,635	252	67.4	94.0
B	6	5.64	2.83	1,741	69	2,420	73	72.0	94.8
B	7	6.51	4.15	2,392	267	3,293	293	72.6	91.2
B	8	5.57	2.88	1,923	54	2,854	57	67.4	94.7
B	9	6.09	4.10	2,402	258	3,606	280	66.6	92.1

[a] A higher number indicates more diversity.
[b] Nonparametric richness estimator based on the distribution of singletons and doubletons.
[c] The estimates were calculated by employing the tools Aligner, Complete Linkage Clustering, and Rarefaction of the RDP pyrosequencing pipeline (11). The results from the rarefaction analyses are also depicted in Fig. 1.

Canadian forest soil sample. The employment of the whole data set (53,632 sequences) revealed 5,500 OTUs at a genetic distance of 3%, whereas the reduction of the same data set to 30,000 sequences yielded 3,500 OTUs. In general, fewer sequences result in lower curve progression and a lower number of predicted OTUs. In addition, the comparison of richness estimates between different surveys might be hampered by the differences in sequence conservation and sequence length of the analyzed 16S rRNA gene regions. Recently, Engelbrektson et al. (15) showed that amplicon length and differences in the analyzed 16S rRNA gene regions markedly influenced estimates of richness and evenness.

The Shannon index of diversity (H') was determined for all samples (Table 2). At a genetic distance of 3%, it ranged from 5.65 to 7.16 in the A horizons and from 5.01 to 6.72 in the B horizons. The predicted diversity in the topsoil exceeded that of the corresponding subsoil, except for plot 1 (Table 2). To our knowledge, no other study assessing bacterial diversity along a soil profile was conducted with a comparable surveying effort. However, a significant decrease of bacterial diversity with soil depth was also recorded by a community analysis employing terminal restriction fragment length polymorphism (33) and phospholipid fatty acid analysis (18) of soil profiles derived from Californian grassland and soil samples of the Sedgwick Reserve (California), respectively.

The bacterial diversity at a genetic distance of 3% was strongly related to the content of OC and N as well as to the C-to-N ratio. Positive correlations between the H' and the OC content ($r = 0.60$; $P < 0.01$), the N content ($r = 0.58$; $P < 0.05$), and the C-to-N ratio ($r = 0.65$; $P < 0.01$) were observed. Similar correlations were detected by analyzing soil samples from South American grasslands, in which the H' correlated positively with the microbial biomass C and N ($r = 0.53$ to 0.58; $P = 0.02$ to 0.03) (3). Interestingly, the mean H' was lower in fertilized meadows (plots 1 to 3), with intermediate values in fertilized mown meadows (plots 4 to 6) and the highest values in unfertilized pastures (plots 7 to 9), over both horizons (Table 2). Thus, a higher bacterial diversity in samples from unfertilized plots, which represent the lowest land use intensity in this study, is indicated. Nevertheless, within the same land use types, strong variations in diversity were observed.

Distribution of taxa and phylotypes across all samples. The 600,544 classifiable sequences were affiliated with 23 phyla across the entire data set. The dominant phyla across all samples were *Acidobacteria*, *Betaproteobacteria*, *Actinobacteria*, *Gammaproteobacteria*, *Alphaproteobacteria*, *Deltaproteobacteria*, *Chloroflexi*, *Firmicutes*, and *Bacteroidetes*, representing 26.98, 15.76, 11.62, 11.10, 9.69, 5.09, 3.85, 3.22, and 1.45%, respectively, of all sequences that were classified below the domain level (Fig. 2 and see Table S3 in the supplemental material). In addition, the dominant phyla were present in all samples. These results are in accordance with results from a previously reported meta-analysis of bacterial community composition in soils (29). In addition, the abundances of the five dominant phyla, *Acidobacteria*, *Actinobacteria*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*, which represented approximately 75% of all classified sequences, corresponded roughly to those found by other studies (21, 34). Thus, despite the different surveying efforts and sampling sites used in the different studies, it is indicated that a variety of soils contain the same dominant bacterial groups.

The most abundant phylotypes in the A and B horizons were a member of the *Alphaproteobacteria* (*Acetobacteraceae*) and a member of acidobacterial subgroup 4, respectively. The corresponding sequences represented 0.74 and 2.86% of all classified sequences in each respective horizon. The most abundant phylotype within one individual sample (plot 3, B horizon) was

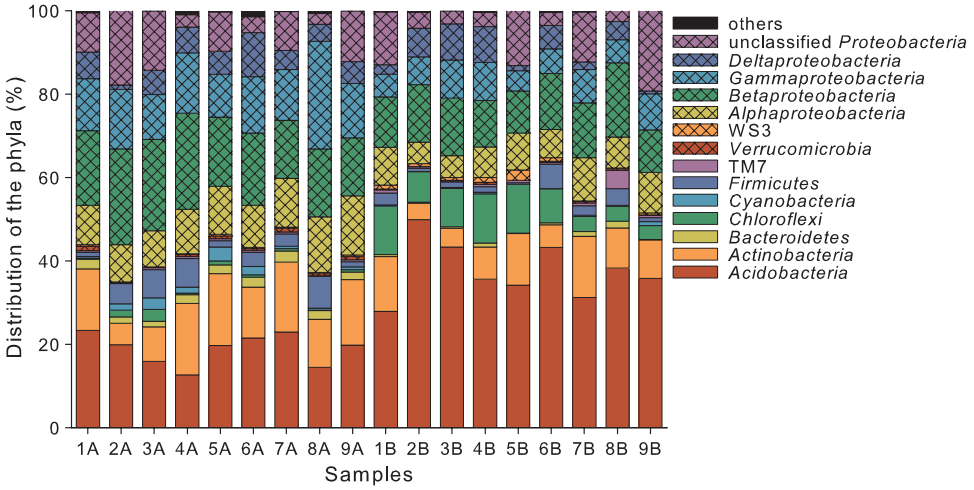


FIG. 2. Distribution of phylogenetic groups in the A and B horizons derived from the different grassland sampling sites. Plot numbers are given below the graph. A description of the plots is given in Table 1. A and B indicate the different horizons. Shown are the percentages of the classified sequences. Phylogenetic groups accounting for ≤0.25% of the classified sequences are summarized in the artificial group "others."

the above-mentioned member of acidobacterial subgroup 4, representing 8.36% of the sequences from that soil.

Distribution and abundance of the predominant phylum *Acidobacteria*. Members of the phylum *Acidobacteria* were predominant across all samples. This finding is in accordance with findings of other studies of the composition of soil-derived bacterial communities from a variety of environments, such as pristine forest, grassland, and agricultural soils (29). Here, members of the *Acidobacteria* form a significant fraction (12.68 to 49.86%) of the 16S rRNA gene sequences in both horizons from all land use types (see Tables S3 and S4 in the supplemental material). Correspondingly, members of this phylum have been reported to constitute an average of 20% and a maximum of approximately 50% of bacterial communities derived from various soils (13). Thus, based on their abundance and their presence in various soil types, the *Acidobacteria* appear to play an important role in the ecosystem function of soils.

The phylum *Acidobacteria* is divided into 26 subgroups, but only little is known with respect to the physiological and metabolic capabilities of the different subgroups (2). We detected 18 and 22 of these subgroups in the A and B horizons, respectively. Most abundant in both horizons were subgroups 6, 4, 16, and 7 (see Tables S4 and S5 in the supplemental material). In the A horizons, these subgroups were represented by 47, 21, 13, and 7% of all acidobacterial sequences, respectively, and in the B horizons, these subgroups were represented by 26, 38, 5, and 14% of all acidobacterial sequences, respectively (see Table S5 in the supplemental material). These results are in contrast to a previous study by Hansel et al. (21) of samples from a continuous watershed soil profile (Oak Ridge), which is the only other report of acidobacterial diversity with respect to soil horizon. In the A horizon, those researchers detected primarily subgroups 3 (21%), 4 (29%), and 6 (29%), whereas in our samples, these subgroups were represented by 5, 21, and 47%, respectively. In the B horizon, Hansel et al. (21) detected primarily subgroups 1 (32%) and 2 (61%), which were represented by less than 1% of all acidobacterial sequences derived from our soil samples. The predominant subgroups in the B horizons from the Hainich region were subgroups 4, 5, and 7. The major differences in the occurrences of acidobacterial subgroups in the B horizon might be due to the dissimilar pH values of the samples used in both studies. The pH in our subsoil samples ranged from 6.30 to 7.40 (Table 1) whereas the pH of the soil samples studied by Hansel et al. (21) was 4.5. It was reported previously that the abundance of the phylum *Acidobacteria* correlates with the soil pH (22, 30). Lauber et al. (34) showed previously that acidobacterial subgroups 1 and 2 were most abundant in acidic soils and decreased with the increase of the pH. Here, no significant correlations of changes in the abundance of the dominant acidobacterial subgroups and other phylogenetic groups with pH were observed. A reason for this finding might be that the sampling effort in most of the other studies was much less than that of this study. Another possibility is that almost all the pH values of our samples were near neutral. Correspondingly, a relatively small pH range was covered by our soil samples (Table 1), so there is simply a lower pH range from which to determine correlations. Nevertheless, we observed negative correlations between the abundant acidobacterial subgroup 4 and the OC content ($r = -0.84$; $P < 0.01$), N content ($r = -0.83$; $P < 0.01$), or C-to-N ratio ($r = -0.77$; $P < 0.05$) in the A horizons (see Fig. S1 in the supplemental material). In the B horizons, subgroup 4 also correlated negatively with the C-to-N ratio ($r = -0.70$; $P < 0.05$), whereas the relative abundance of subgroup 6 showed a positive correlation with the C-to-N ratio ($r = 0.70$; $P < 0.05$). Thus, the subgroup distribution varied with respect to the soil profile (horizon) and soil properties and provided some insights into the conditions that are required by the different subgroups. For example, low nutrient/OC conditions (B horizons) appear to favor subgroups 4 and 7, whereas higher nutrient/OC conditions (A horizons) favor subgroup 16. Interestingly, for members of subdivision 6, a high tolerance to nutrient/OC availability was indicated, as they constituted a substantial fraction in the A horizon and the B horizon (8.87 and 9.73% of all classified sequences, respectively).

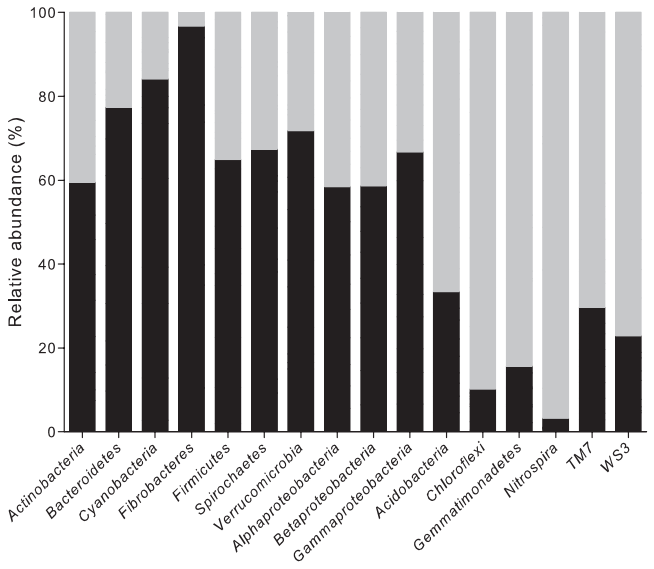


FIG. 3. Comparison of the overall distribution of selected phyla within the A and B horizons. The black bar represents the sum of all members of a phylum in the A horizon, while the gray bar represents the sum of all members in the B horizon. Box-and-whisker plots of the data are depicted in Fig. S2 in the supplemental material.

Taxonomic compositions in A horizons and the corresponding B horizons. The compositions of the bacterial community and the distributions of the phyla varied between A and B soil horizons. The most frequently present phyla in the A horizons were *Acidobacteria* (13 to 23%), *Betaproteobacteria* (14 to 23%), *Gammaproteobacteria* (10 to 26%), *Actinobacteria* (5 to 17%), and *Alphaproteobacteria* (9 to 14%). The most abundant phyla in the B horizons were *Acidobacteria* (28 to 50%), *Betaproteobacteria* (10 to 18%), *Actinobacteria* (4 to 15%), *Chloroflexi* (3 to 12%), and *Alphaproteobacteria* (5 to 10%) (Fig. 2).

We analyzed the respective abundances of the 15 most represented phyla in the A and B horizons. For almost all phyla and land use types, a significant ($P < 0.00001$) difference in distribution between the two horizons was apparent (see Table S6 in the supplemental material). The distribution of selected phyla in the two horizons is shown in Fig. 3 (see also Fig. S2 in the supplemental material). The majority of sequences affiliated with the *Actinobacteria*, *Bacteroidetes*, *Cyanobacteria*, *Fibrobacteres*, *Firmicutes*, *Spirochaetes*, *Verrucomicrobia*, *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria* were derived from A horizons, whereas the majority of the sequences related to *Acidobacteria*, *Chloroflexi*, *Gemmatimonadetes*, *Nitrospira*, TM7, and WS3 originated from B horizons. In many other reports, a pH gradient was identified as a major factor for changes in soil community structure, but as mentioned above, this trend was not observed in this study. Therefore, other factors appear to control the distribution of the phyla along the soil profile. As stated above, the total biomass decreased with soil depth (Table 1). A significant correlation between the total microbial biomass and the occurrence of several phyla was recorded. The relative abundance of the *Acidobacteria*, *Chloroflexi*, and *Nitrospira*, which increased with depth, correlated negatively with total biomass ($r = -0.53$ to -0.79; $P < 0.05$). The relative abundances of the *Actinobacteria*, *Bacteroidetes*, *Verrucomicrobia*, *Alphaproteobacteria*, and *Gammaproteobacteria*, which decreased with depth, showed a positive correlation with biomass ($r = 0.69$ to 0.84; $P < 0.01$). A positive correlation of some of the dominant acidobacterial groups (see above) and the alphaproteobacterial order *Rhizobiales* with the concentration of OC ($r = 0.77$; $P < 0.05$), the concentration of N ($r = 0.73$; $P < 0.05$), and the C-to-N ratio ($r = 0.87$; $P < 0.01$) was detectable in the B horizon (see Fig. S3 in the supplemental material). A statistically significant positive correlation of the *Rhizobiales* with the C-to-N ratio ($r = 0.80$; $P < 0.01$) was also detected in the A horizon. Taking into account that the *Rhizobiales* include the genera *Rhizobium* and *Bradyrhizobium*, which comprise members that are able to fix nitrogen and are associated with roots of legumes, a positive correlation was expected (49). In addition, land use and management regimens seem to have an impact on the *Rhizobiales*, as the fertilized plots cluster and the plots with cattle cluster (data not shown). An impact of fertilization on the structure and diversity of rhizobial populations was observed previously in other studies (6, 41). For example, rhizobial populations differed between cultivated and uncultivated Mexican soils. In addition, the affinity of host cultivars for different members of the *Rhizobiales* influenced the composition of rhizobial populations (54).

Striking is the distribution of the phyla *Fibrobacteres* and *Nitrospira*, which occurred almost exclusively in the A or the B horizon, respectively. Members of the *Fibrobacteres* are part of the microbial community in the first stomach of ruminant animals and degrade plant-based cellulose (31). Taking into account that members of the *Fibrobacteres* hardly occurred in grassland samples without the presence of cattle (i.e., plots 1 to 3) and almost exclusively in the topsoil, it can be assumed that members of this phylum were introduced into the samples by cattle. Members of the *Nitrospira* are found in interspace soils and rarely in the rhizosphere (13). In the latter environment, heterotrophic root-associated microorganisms suppress the growth of autotrophic *Nitrospira*. This might explain why in our samples, members of the *Nitrospira* occurred in the subsoil, with less rooting than in the topsoil. In addition, the concentrations of OC and N decreased with soil depth (Table 1). Thus, chemolithoautotrophic organisms adapted to darkness, like *Nitrospira*, have a selective advantage in subsoil samples.

Conclusions. Although we recovered an average of 41,824 sequences per sample, we did not survey the full extent of bacterial richness at the species level within an individual soil or horizon. Thus, an increase in surveying efforts would probably result in the identification of more bacterial taxa, which are present in a low abundance. In most cases, the B horizons showed a lower estimated bacterial diversity than the corresponding A horizons (Table 2). Correspondingly, a greater coverage of the bacterial community in the B horizons can be achieved by using the same surveying effort. In addition, the identification of bacterial taxa at the finest level of taxonomic resolution is currently not feasible by applying large-scale pyrosequencing. However, the advancement of the technology will result in an increase of the read length, and this limitation will become less relevant in the near future.

To provide a robust assessment of the impact of land use, soil factors, or soil depth on bacterial diversity, distribution,

and community composition, we used triplicate samples of each land use type from different locations. Overall, the abundance, composition, and diversity of the bacterial communities were strongly depth dependent. The Shannon index of diversity along with the nutrient content (N and OC), water content, and biomass decreased with depth (Tables 1 and 2). Based on the sharp decrease of the OC content (up to 10.9-fold) in the B horizons compared to the corresponding A horizons (Table 1) and other surveys (18, 33, 57), the concentration of OC appears to be the major driver for the diversity and structure of bacterial communities along the soil profile at near-neutral pH values. Nevertheless, we observed a variability of bacterial communities within an individual land use type, and exceptions to the above-mentioned general results were found; i.e., a slightly higher Shannon index was recorded for the B horizon of plot 1. Thus, it is advisable to survey as many soil samples as possible for the identification of general patterns and comparison of the results with those of other soil surveys. One caveat of the latter, however, is the limited comparability of different surveys, as sampling strategy, survey effort, number and type of soil factors measured, and approaches used to analyze the sequence data vary considerably (37). To take full advantage of the increasing number of data sets on soil bacterial communities, minimal requirements for sampling and the set of analyzed soil factors as well as rules for sequence analysis and phylogenetic assignment should be defined.

ACKNOWLEDGMENTS

We thank Rudolf-Josef Fischer (Department of Medical Informatics and Biomathematics, Westphalian Wilhelms University, Münster, Germany) for supporting the statistical analyses. We thank the local implementation team of the Hainich exploratory and the BEO (Biodiversity Exploratories Office) for organizing the coordinated soil sampling campaign.

The work has been funded by the DFG Priority Program 1374, Infrastructure-Biodiversity-Exploratories (DA 374/4-1).

Field work permits were given by the responsible state environmental offices of Thüringen.

REFERENCES

1. **Acosta-Martínez, V., S. Dowd, Y. Sun, and V. Allen.** 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol. Biochem. **40:**2762–2770.
2. **Barns, S., E. Cain, L. Sommerville, and C. Kuske.** 2007. *Acidobacteria* phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. Appl. Environ. Microbiol. **73:**3113–3116.
3. **Berthrong, S., C. Schadt, G. Piñeiro, and R. Jackson.** 2009. Afforestation alters the composition of functional genes in soil and biogeochemical processes in South American grasslands. Appl. Environ. Microbiol. **75:**6240–6248.
4. **Bligh, E., and W. Dyer.** 1959. A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. **37:**911–917.
5. **Blume, E., M. Bischoff, J. M. Reichert, T. Moorman, A. Konopka, and R. F. Turco.** 2002. Surface and subsurface microbial biomass, community structure and metabolic activity as a function of soil depth and season. Appl. Soil Ecol. **20:**171–181.
6. **Caballero-Mellado, J., and E. Martínez-Romero.** 1999. Soil fertilization limits the genetic diversity of *Rhizobium* in bean nodules. Symbiosis **26:**111–121.
7. **Chakravorty, S., D. Helb, M. Burday, N. Connell, and D. Alland.** 2007. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods **69:**330–339.
8. **Chao, A., and J. Bunge.** 2002. Estimating the number of species in a stochastic abundance model. Biometrics **58:**531–539.
9. **Cho, S., S. Tsai, A. Ravindran, A. Selvam, and S. Yang.** 2008. Seasonal variation of microbial populations and biomass in Tatachia grassland soils of Taiwan. Environ. Geochem. Health **30:**255–272.
10. **Chow, M., C. Radomski, J. McDermott, J. Davies, and P. Axelrood.** 2002. Molecular characterization of bacterial diversity in Lodgepole pine (*Pinus contorta*) rhizosphere soils from British Columbia forest soils differing in disturbance and geographic source. FEMS Microbiol. Ecol. **42:**347–357.
11. **Cole, J., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. Farris, A. Kulam-Syed-Mohideen, D. McGarrell, T. Marsh, G. Garrity, and J. Tiedje.** 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. **37:**D141–D145.
12. **Daniel, R.** 2005. The metagenomics of soil. Nat. Rev. Microbiol. **3:**470–478.
13. **Dunbar, J., S. Barns, L. Ticknor, and C. Kuske.** 2002. Empirical and theoretical bacterial diversity in four Arizona soils. Appl. Environ. Microbiol. **68:**3035–3045.
14. **Elshahed, M., N. Youssef, A. Spain, C. Sheik, F. Najar, L. Sukharnikov, B. Roe, J. Davis, P. Schloss, V. Bailey, and L. Krumholz.** 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl. Environ. Microbiol. **74:**5422–5428.
15. **Engelbrektson, A., V. Kunin, K. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, and P. Hugenholtz.** 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J. **4:**642–647.
16. **Federle, T., D. Dobbins, J. Thornton-Manning, and D. Jones.** 1986. Microbial biomass, activity, and community structure in subsurface soils. Ground Water **24:**365–374.
17. **Fierer, N., and R. Jackson.** 2006. The diversity and biogeography of soil bacterial communities. Proc. Natl. Acad. Sci. U. S. A. **103:**626–631.
18. **Fierer, N., J. P. Schimel, and P. A. Holden.** 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol. Biochem. **35:**167–176.
19. **Frostegård, A., and E. Bååth.** 1996. The use of fatty acid analysis to estimate bacterial and fungal biomass in soil. Biol. Fertil. Soils **22:**59–65.
20. **Garrity, G. M., J. A. Bell, and T. G. Lilburn.** 2004. Taxonomic outline of the prokaryotes, p. 4–23. *In* G. M. Garrity (ed.), Bergey's manual of systematic bacteriology, 2nd ed., vol. 2. Springer-Verlag, New York, NY.
21. **Hansel, C., S. Fendorf, P. Jardine, and C. Francis.** 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl. Environ. Microbiol. **74:**1620–1633.
22. **Hartman, W., C. Richardson, R. Vilgalys, and G. Bruland.** 2008. Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc. Natl. Acad. Sci. U. S. A. **105:**17842–17847.
23. **Hartmann, M., and F. Widmer.** 2006. Community structure analyses are more sensitive to differences in soil bacterial communities than anonymous diversity indices. Appl. Environ. Microbiol. **72:**7804–7812.
24. **He, J., Z. Xu, and J. Hughes.** 2006. Molecular bacterial diversity of a forest soil under residue management regimes in subtropical Australia. FEMS Microbiol. Ecol. **55:**38–47.
25. **Hill, T., E. McPherson, J. Harris, and P. Birch.** 1993. Microbial biomass estimated by phospholipid phosphate in soils with diverse microbial communities. Soil Biol. Biochem. **25:**1779–1786.
26. **Huse, S., J. Huber, H. Morrison, M. Sogin, and D. Welch.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. **8:**R143.
27. **IUSS Working Group.** 2006. World reference base for soil resources 2006, 2nd ed. World soil resources reports no. 103. FAO, Rome, Italy.
28. **Jahn, R., H.-P. Blume, V. B. Asio, O. Spaargaren, and P. Schad.** 2006. Guidelines for soil description, 4th ed. FAO, Rome, Italy.
29. **Janssen, P.** 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl. Environ. Microbiol. **72:**1719–1728.
30. **Jones, R., M. Robeson, C. Lauber, M. Hamady, R. Knight, and N. Fierer.** 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J. **3:**442–453.
31. **Kobayashi, Y., T. Shinkai, and S. Koike.** 2008. Ecological and physiological characterization shows that *Fibrobacter succinogenes* is important in rumen fiber digestion—review. Folia Microbiol. (Praha) **53:**195–200.
32. **Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz.** 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. Environ. Microbiol. **12:**118–123.
33. **LaMontagne, M., J. Schimel, and P. Holden.** 2003. Comparison of subsurface and surface soil bacterial communities in California grassland as assessed by terminal restriction fragment length polymorphisms of PCR-amplified 16S rRNA genes. Microb. Ecol. **46:**216–227.
34. **Lauber, C., M. Hamady, R. Knight, and N. Fierer.** 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. **75:**5111–5120.
35. **Lazarevic, V., K. Whiteson, S. Huse, D. Hernandez, L. Farinelli, M. Østerås, J. Schrenzel, and P. François.** 2009. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. J. Microbiol. Methods **79:**266–271.
36. **Liu, Z., C. Lozupone, M. Hamady, F. Bushman, and R. Knight.** 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. **35:**120–130.
37. **Morales, S., T. Cosart, J. Johnson, and W. Holben.** 2009. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA

fractionation on classification and ecological parameters. Appl. Environ. Microbiol. **75:**668–675.
38. **Muyzer, G., E. de Waal, and A. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. **59:**695–700.
39. **Narang, R., and J. Dunbar.** 2004. Modeling bacterial species abundance from small community surveys. Microb. Ecol. **47:**396–406.
40. **Osborn, A., E. Moore, and K. Timmis.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. **2:**39–50.
41. **Palmer, K., and J. Young.** 2000. Higher diversity of *Rhizobium leguminosarum* biovar viciae populations in arable soils than in grass soils. Appl. Environ. Microbiol. **66:**2445–2450.
42. **Petrosino, J., S. Highlander, R. Luna, R. Gibbs, and J. Versalovic.** 2009. Metagenomic pyrosequencing and microbial identification. Clin. Chem. **55:** 856–866.
43. **Roesch, L., R. Fulthorpe, A. Riva, G. Casella, A. Hadwin, A. Kent, S. Daroub, F. Camargo, W. Farmerie, and E. Triplett.** 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J. **1:**283–290.
44. **Rosselló-Mora, R., and R. Amann.** 2001. The species concept for prokaryotes. FEMS Microbiol. Rev. **25:**39–67.
45. **Schlichting, E., and H.-P. Blume.** 1966. Bodenkundliches Praktikum: eine Einführung in pedologisches Arbeiten für Ökologen, insbesondere Land- und Forstwirte und für Geowissenschaftler. Paul Parey, Hamburg, Germany.
46. **Schloss, P., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501–1506.
47. **Schloss, P., and J. Handelsman.** 2006. Toward a census of bacteria in soil. PLoS Comput. Biol. **2:**e92.
48. **Schmalenberger, A., F. Schwieger, and C. Tebbe.** 2001. Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. Appl. Environ. Microbiol. **67:**3557–3563.
49. **Sessitsch, A., J. G. Howieson, X. Perret, H. Antoun, and E. Martínez-Romero.** 2002. Advances in *Rhizobium* research. Crit. Rev. Plant Sci. **21:**323–378.
50. **Shannon, C.** 1997. The mathematical theory of communication. 1963. MD Comput. **14:**306–317.
51. **Sogin, M., H. Morrison, J. Huber, D. Mark Welch, S. Huse, P. Neal, J. Arrieta, and G. Herndl.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere." Proc. Natl. Acad. Sci. U. S. A. **103:**12115–12120.
52. **Tiedje, J. M., S. Asuming-Brempong, K. Nüsslein, T. L. Marsh, and S. J. Flynn.** 1999. Opening the black box of soil microbial diversity. Appl. Soil Ecol. **13:**109–122.
53. **van der Heijden, M., R. Bardgett, and N. van Straalen.** 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. Ecol. Lett. **11:**296–310.
54. **Wang, E., J. Martínez-Romero, and E. Martínez-Romero.** 1999. Genetic diversity of rhizobia from *Leucaena leucocephala* nodules in Mexican soils. Mol. Ecol. **8:**711–724.
55. **Wang, Q., G. Garrity, J. Tiedje, and J. Cole.** 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73:**5261–5267.
56. **Youssef, N., and M. Elshahed.** 2008. Species richness in soil bacterial communities: a proposed approach to overcome sample size bias. J. Microbiol. Methods **75:**86–91.
57. **Zhou, J., B. Xia, D. Treves, L. Wu, T. Marsh, R. O'Neill, A. Palumbo, and J. Tiedje.** 2002. Spatial and resource factors influencing high microbial diversity in soil. Appl. Environ. Microbiol. **68:**326–334.

2.1 Supplemental information for chapter B2

Contents

Table S1. Soil type, land use type, and coordinates of the sampling sites.

Plot	Soil type	Land use type	Latitude	Longitude
1	Vertic Cambisol (eutric)	Fertilized meadow	50°58'17.99"	10°24'19.24"
2	Vertic Stagnosol (eutric)	Fertilized meadow	51°0'2.75"	10°25'47.96"
3	Vertic Stagnosol (eutric)	Fertilized meadow	50°59'53.19"	10°25'58.54"
4	Vertic Stagnosol (eutric)	Fertilized mown pasture, cattle	51°6'48.17"	10°26'10.17"
5	Vertic Stagnosol (eutric)	Fertilized mown pasture, cattle	51°12'57.28"	10°19'21.1"
6	Luvic Stagnosol (siltic)	Fertilized mown pasture, cattle	51°12'53.83"	10°23'28.31"
7	Vertic Stagnosol (eutric)	Unfertilized pasture, cattle	51°16'24.96"	10°24'37.40"
8	Vertic Stagnosol (eutric)	Unfertilized pasture, cattle	51°16'16.59"	10°25'4.52"
9	Vertic Stagnosol (eutric)	Unfertilized pasture, cattle	51°13'26.9"	10°22'50.75"

Table S2. Number of 16S rRNA sequences derived from the A and B horizons of nine different grassland sampling sites. For a description of the sampling sites, see Table 1 and supplemental Table S1.

Horizon	Plot	No. of obtained sequences	No. of classified sequences
A	1	47,063	38,693
A	2	39,270	32,702
A	3	33,804	27,960
A	4	39,328	34,651
A	5	53,422	43,443
A	6	41,239	35,062
A	7	61,366	49,383
A	8	47,068	41,810
A	9	51,870	42,243
B	1	47,264	33,558
B	2	33,209	23,875
B	3	36,598	28,428
B	4	38,856	30,565
B	5	40,189	25,511
B	6	41,027	31,572
B	7	25,851	19,662
B	8	42,994	37,439
B	9	32,420	23,987
Sum		752,838	600,544

Table S3. Relative abundance of the phylogenetic groups. Values are given as percentage. Phylogenetic groups accounting for ≤0.25% of the classified sequences are summarized in the artificial group 'others', containing the phyla *Deinococcus-Thermus*, *Fibrobacteres*, *Fusobacteria*, *Gemmatimonadetes*, *Nitrospira*, OP11, *Planctomycetes*, *Spirochaetes*, *Tenericutes*, and *Epsilonproteobacteria*.

Horizon	Plot	*Acidobacteria*	*Actinobacteria*	*Bacteroidetes*	*Chloroflexi*	*Cyanobacteria*	*Firmicutes*	TM7	*Verrucomicrobia*	WS3	*Alphaproteobacteria*	*Betaproteobacteria*	*Gammaproteobacteria*	*Deltaproteobacteria*	unclassified *Proteobacteria*	others
A	1	23.27	14.80	2.26	0.30	0.43	1.04	0.34	1.03	0.43	9.40	17.89	12.51	6.33	9.44	0.50
A	2	19.88	5.21	1.46	1.59	1.57	4.84	0.20	0.11	0.05	9.00	22.95	14.31	1.00	17.79	0.05
A	3	15.90	8.26	1.33	2.86	2.78	6.75	0.44	0.14	0.09	8.68	21.91	10.77	5.80	14.13	0.17
A	4	12.68	17.14	2.09	0.41	1.37	6.91	0.45	0.44	0.17	10.70	23.08	14.42	6.20	2.92	1.02
A	5	19.73	17.17	2.11	0.96	3.31	1.54	0.47	0.66	0.43	11.54	16.57	10.23	5.55	9.46	0.27
A	6	21.51	12.22	2.36	0.53	2.04	3.39	0.50	0.44	0.22	10.09	17.37	13.48	10.60	3.82	1.43
A	7	22.97	16.73	2.65	0.54	0.60	3.00	0.46	0.83	0.30	11.72	13.89	12.24	4.52	9.33	0.23
A	8	14.40	11.59	2.05	0.17	0.42	7.64	0.30	0.54	0.12	13.31	16.32	25.80	4.09	2.59	0.66
A	9	19.79	15.73	1.75	0.61	0.66	1.28	0.46	0.76	0.38	14.22	13.91	13.02	5.19	12.08	0.16
B	1	27.85	13.19	0.43	11.64	0.32	2.86	0.71	0.26	0.87	9.08	12.13	5.42	2.26	12.70	0.30
B	2	49.86	3.95	0.26	7.30	0.05	0.77	0.34	0.09	0.65	5.13	13.91	6.64	6.85	3.78	0.41
B	3	43.26	4.53	0.32	9.13	0.20	1.26	0.34	0.07	0.69	5.24	13.83	9.14	8.61	3.01	0.17
B	4	35.61	7.73	0.92	11.81	0.41	1.26	0.61	0.52	1.07	7.35	11.13	9.14	8.63	3.43	0.37
B	5	34.19	12.40	0.09	11.65	0.04	0.34	0.58	0.04	2.44	8.89	10.13	4.81	1.26	13.05	0.11
B	6	43.20	5.44	0.48	8.17	0.07	5.84	0.36	0.30	0.83	6.85	13.37	5.89	5.60	3.16	0.43
B	7	31.23	14.70	1.13	3.57	0.26	2.35	0.69	0.34	0.08	10.33	13.19	7.96	1.84	12.02	0.29
B	8	38.28	9.66	1.55	3.65	0.19	4.01	4.39	0.26	0.26	7.44	17.76	5.61	4.38	2.43	0.12
B	9	35.81	9.22	0.10	3.34	0.93	0.95	0.57	0.06	0.51	9.71	10.25	8.61	0.66	19.14	0.14

Table S4. Percentages of the different acidobacterial subgroups in the A and B horizons of the German grassland soils derived from each plot. Percentages are given relative to the total number of classified sequences in that soil. Sequences of acidobacterial subgroups 14, 23, 24, and 26 were not detected.

Plot[a]	Total No.[b]	Relative abundance of acidobacterial subgroups[c] (%)																				
		Gp1	Gp2	Gp3	Gp4	Gp5	Gp6	Gp7	Gp8	Gp9	Gp10	Gp11	Gp12	Gp13	Gp15	Gp16	Gp17	Gp18	Gp19	Gp20	Gp21	Gp22	Gp25
1A	38,693	0.05	0.00	1.22	2.55	0.28	13.30	1.67	0.00	0.00	0.00	0.29	0.00	0.01	0.00	2.85	0.46	0.02	0.00	0.00	0.00	0.42	0.02
2A	32,702	0.02	0.00	0.59	7.70	0.10	6.64	1.86	0.00	0.00	0.00	0.81	0.00	0.00	0.00	1.10	0.13	0.01	0.00	0.50	0.00	0.33	0.05
3A	27,960	0.01	0.00	1.14	4.46	0.09	7.01	0.64	0.00	0.00	0.02	0.34	0.00	0.00	0.00	1.53	0.25	0.03	0.00	0.08	0.00	0.25	0.03
4A	34,651	0.01	0.00	0.37	2.12	0.11	6.11	0.44	0.01	0.00	0.02	0.23	0.00	0.00	0.00	2.59	0.34	0.01	0.00	0.02	0.00	0.16	0.03
5A	43,443	0.09	0.00	0.84	3.51	0.26	8.63	1.62	0.00	0.00	0.03	0.24	0.00	0.00	0.00	3.34	0.64	0.00	0.00	0.03	0.00	0.34	0.02
6A	35,062	0.08	0.00	1.17	7.69	0.13	7.70	1.22	0.00	0.00	0.10	0.31	0.00	0.01	0.01	2.14	0.26	0.02	0.00	0.07	0.00	0.37	0.12
7A	49,383	0.02	0.00	1.29	3.73	0.38	11.25	1.56	0.00	0.00	0.03	0.27	0.00	0.00	0.00	3.21	0.71	0.02	0.00	0.01	0.00	0.31	0.07
8A	41,810	0.01	0.00	0.29	2.36	0.11	7.74	0.81	0.00	0.00	0.03	0.18	0.00	0.00	0.00	2.39	0.12	0.01	0.00	0.02	0.00	0.16	0.01
9A	42,243	0.03	0.00	1.10	2.91	0.41	9.56	1.45	0.00	0.00	0.05	0.23	0.00	0.00	0.00	2.81	0.71	0.02	0.00	0.00	0.00	0.36	0.03
1B	33,558	0.01	0.00	1.59	6.65	0.22	10.79	2.36	0.00	0.02	0.01	0.88	0.00	0.00	0.01	2.43	1.14	0.59	0.00	0.09	0.00	0.72	0.22
2B	23,875	1.91	0.00	1.10	16.16	0.24	14.00	10.06	0.00	0.05	0.04	1.53	0.00	0.35	0.09	1.78	0.91	1.08	0.00	0.04	0.00	0.06	0.04
3B	28,428	0.10	0.00	1.38	19.45	0.14	9.98	7.29	0.00	0.02	0.06	2.05	0.00	0.04	0.04	0.82	0.70	0.66	0.00	0.04	0.00	0.11	0.21
4B	30,565	0.12	0.00	1.38	15.95	0.95	7.66	3.09	0.02	0.03	0.04	3.15	0.00	0.02	0.04	0.86	0.77	0.60	0.01	0.27	0.00	0.38	0.23
5B	25,511	0.05	0.00	1.27	8.48	0.20	7.58	7.98	0.00	0.06	0.02	1.52	0.00	0.02	0.08	3.00	2.59	0.54	0.00	0.05	0.00	0.58	0.11
6B	31,572	3.22	0.01	1.69	19.39	0.77	7.82	2.88	0.00	0.01	0.09	1.73	0.08	0.02	0.10	1.38	1.06	0.91	0.02	0.93	0.00	0.38	0.61
7B	19,662	0.01	0.00	1.29	7.36	0.35	14.67	2.31	0.00	0.00	0.08	0.43	0.00	0.00	0.03	2.85	0.74	0.29	0.00	0.19	0.00	0.38	0.14
8B	37,439	0.04	0.00	0.87	20.45	0.61	9.39	2.53	0.00	0.00	0.08	1.67	0.00	0.00	0.03	1.23	0.24	0.20	0.00	0.15	0.01	0.32	0.27
9B	23,987	0.04	0.00	1.73	9.70	0.15	7.65	11.96	0.00	0.03	0.01	0.54	0.00	0.00	0.03	2.33	0.61	0.37	0.00	0.14	0.00	0.34	0.10

[a], A and B represent samples from A and B horizon, respectively; [b], No. of all classified sequences in that soil; [c], Only subgroups that were detected are depicted.

Table S5. Relative abundance of the different acidobacterial subgroups in the A and B horizons of the German grassland soils within the *Acidobacteria*. Percentages are given relative to all acidobacterial sequences per plot. Sequences of acidobacterial subgroups 14, 23, 24, and 26 were not detected.

Plot[a]	Total No.[b]	Relative abundance of acidobacterial subgroups[c] relating to the total number of classified acidobacterial sequences in that soil (%)																				
		Gp1	Gp2	Gp3	Gp4	Gp5	Gp6	Gp7	Gp8	Gp9	Gp10	Gp11	Gp12	Gp13	Gp15	Gp16	Gp17	Gp18	Gp19	Gp20	Gp21	Gp22	Gp25
1A	8954	0.23	0.00	5.27	11.01	1.19	57.47	7.23	0.00	0.00	0.01	1.26	0.00	0.02	0.01	12.32	1.98	0.10	0.00	0.00	0.00	1.80	0.09
2A	6488	0.08	0.00	2.97	38.81	0.51	33.49	9.36	0.00	0.00	0.02	4.08	0.00	0.00	0.00	5.53	0.63	0.06	0.00	2.54	0.00	1.68	0.23
3A	4438	0.07	0.00	7.19	28.12	0.56	44.19	4.03	0.00	0.00	0.14	2.12	0.00	0.00	0.00	9.62	1.55	0.18	0.00	0.50	0.00	1.55	0.18
4A	4355	0.05	0.00	2.96	16.83	0.90	48.59	3.47	0.05	0.00	0.16	1.86	0.00	0.00	0.02	20.57	2.71	0.07	0.00	0.18	0.00	1.31	0.28
5A	8520	0.46	0.00	4.31	17.89	1.34	44.01	8.26	0.00	0.00	0.15	1.24	0.00	0.00	0.00	17.05	3.29	0.01	0.00	0.13	0.00	1.75	0.11
6A	7501	0.36	0.00	5.45	35.96	0.59	36.01	5.71	0.00	0.01	0.45	1.43	0.00	0.04	0.05	10.00	1.20	0.09	0.00	0.35	0.00	1.75	0.56
7A	1128	0.08	0.00	5.66	16.30	1.67	49.21	6.80	0.00	0.01	0.12	1.18	0.00	0.00	0.02	14.03	3.13	0.09	0.00	0.04	0.00	1.36	0.31
8A	5952	0.10	0.00	2.03	16.55	0.74	54.35	5.70	0.00	0.00	0.20	1.26	0.00	0.00	0.03	16.77	0.84	0.10	0.00	0.13	0.00	1.09	0.10
9A	8316	0.17	0.00	5.60	14.79	2.10	48.58	7.35	0.00	0.00	0.24	1.15	0.00	0.01	0.01	14.30	3.60	0.11	0.00	0.02	0.00	1.80	0.16
1B	9303	0.05	0.00	5.72	23.97	0.78	38.92	8.50	0.01	0.06	0.05	3.18	0.00	0.00	0.04	8.75	4.10	2.13	0.01	0.33	0.00	2.58	0.80
2B	1180	3.85	0.00	2.23	32.68	0.49	28.31	20.34	0.00	0.09	0.08	3.10	0.01	0.70	0.19	3.59	1.85	2.19	0.00	0.08	0.01	0.13	0.08
3B	1224	0.24	0.00	3.19	45.14	0.32	23.18	16.91	0.00	0.06	0.13	4.77	0.01	0.08	0.10	1.89	1.62	1.53	0.00	0.09	0.01	0.25	0.48
4B	1087	0.34	0.00	3.88	44.84	2.68	21.54	8.68	0.06	0.07	0.12	8.86	0.01	0.06	0.12	2.42	2.16	1.69	0.02	0.74	0.00	1.08	0.64
5B	8713	0.16	0.00	3.72	24.82	0.59	22.21	23.38	0.00	0.17	0.05	4.44	0.00	0.05	0.24	8.79	7.60	1.58	0.00	0.16	0.00	1.71	0.33
6B	1360	7.47	0.01	3.92	44.98	1.78	18.14	6.69	0.00	0.01	0.21	4.01	0.18	0.05	0.24	3.21	2.47	2.11	0.05	2.15	0.00	0.87	1.42
7B	6121	0.03	0.00	4.13	23.66	1.13	47.12	7.43	0.00	0.00	0.25	1.39	0.00	0.00	0.10	9.17	2.39	0.95	0.00	0.62	0.00	1.21	0.44
8B	1426	0.10	0.00	2.29	53.68	1.61	24.64	6.65	0.00	0.01	0.21	4.38	0.00	0.01	0.08	3.22	0.64	0.52	0.00	0.40	0.01	0.84	0.70
9B	8567	0.11	0.00	4.83	27.16	0.42	21.43	33.48	0.00	0.07	0.04	1.51	0.00	0.00	0.08	6.53	1.70	1.03	0.00	0.40	0.00	0.96	0.27

[a], A and B represent samples from A and B horizon, respectively; [b], No. of classified acidobacterial sequences in that soil; [c], Only subgroups that were detected are depicted.

Table S6. *P* values estimating the difference of the abundance of selected phyla between A and B horizon with respect to different land use types.

Phylum	Land use type		
	Fertilized meadow (plot 1 to 3)	Fertilized mown pasture, cattle (plot 4 to 6)	Unfertilized pasture, cattle (plot 7 to 9)
Acidobacteria	*P* <0.00001	*P* <0.00001	*P* <0.00001
Actinobacteria	*P* <0.00001	*P* <0.00001	*P* <0.00001
Bacteroidetes	*P* <0.00001	*P* <0.00001	*P* <0.00001
Chloroflexi	*P* <0.00001	*P* <0.00001	*P* <0.00001
Cyanobacteria	*P* <0.00001	*P* <0.00001	*P* 0.00003
Fibrobacteres	*P* <0.00001	*P* <0.00001	*P* < 0.00001
Firmicutes	*P* <0.00001	*P* <0.00001	*P* <0.00001
Nitrospira	*P* <0.00001	*P* <0.00001	*P* 0.47635
TM7	*P* <0.00001	*P* 0.21149	*P* <0.00001
Verrucomicrobia	*P* <0.00001	*P* <0.00001	*P* <0.00001
WS3	*P* <0.00001	*P* <0.00001	*P* 0.29925
Alphaproteobacteria	*P* <0.00001	*P* <0.00001	*P* <0.00001
Betaproteobacteria	*P* <0.00001	*P* <0.00001	*P* 0.15213
Gammaproteobacteria	*P* <0.00001	*P* <0.00001	*P* <0.00001
Deltaproteobacteria	*P* <0.00001	*P* <0.00001	*P* <0.00001

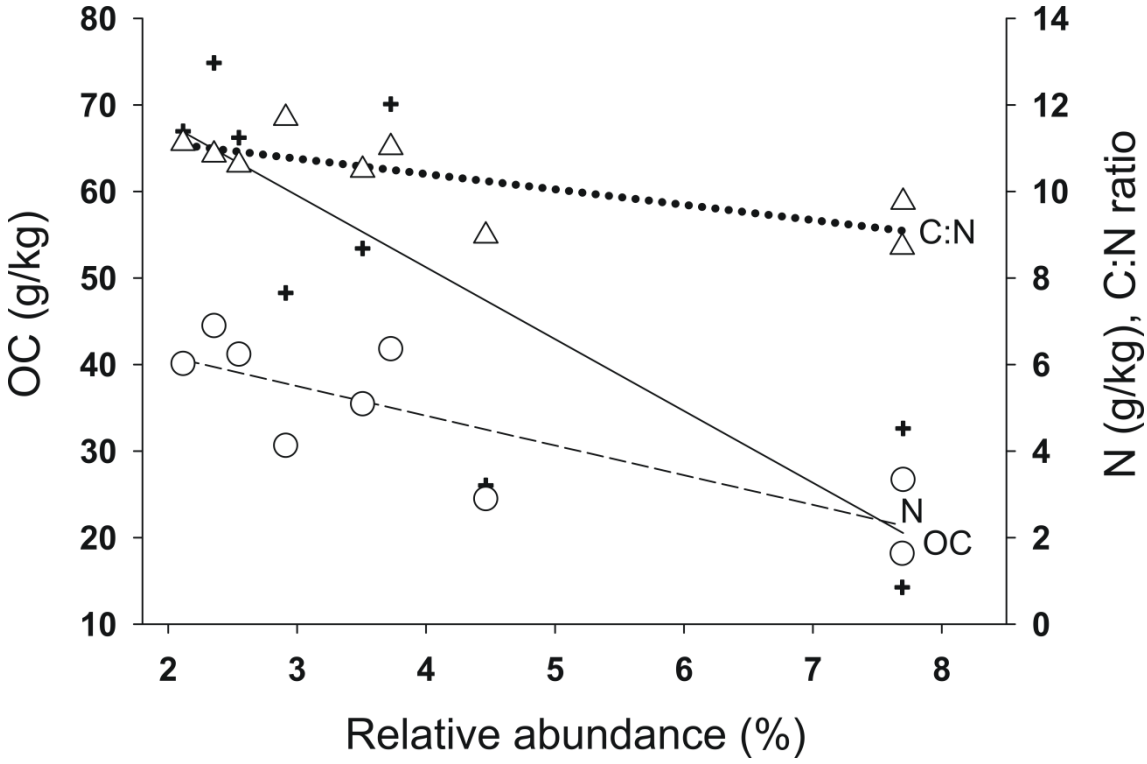


Fig. S1. Correlation of the relative abundance of the acidobacterial subgroup 4 and the edaphic soil properties OC content, N content, and C:N ratio in the A horizon. Crosses and open circles indicate the OC and N content, respectively. The C:N ratio is marked by triangles.

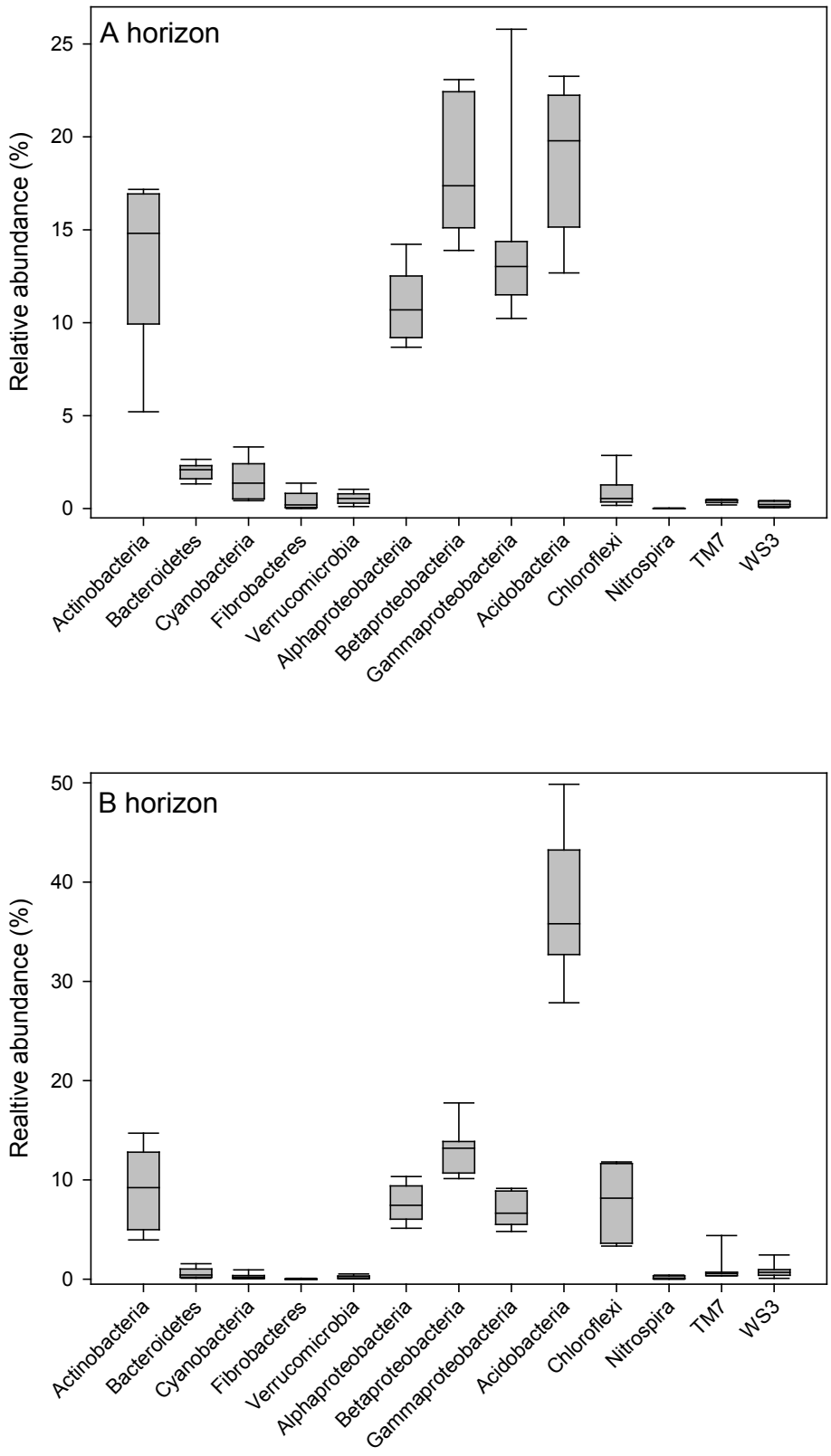


Fig. S2. Box-and-whiskers plot of relative distribution between A and B horizon for selected phyla.

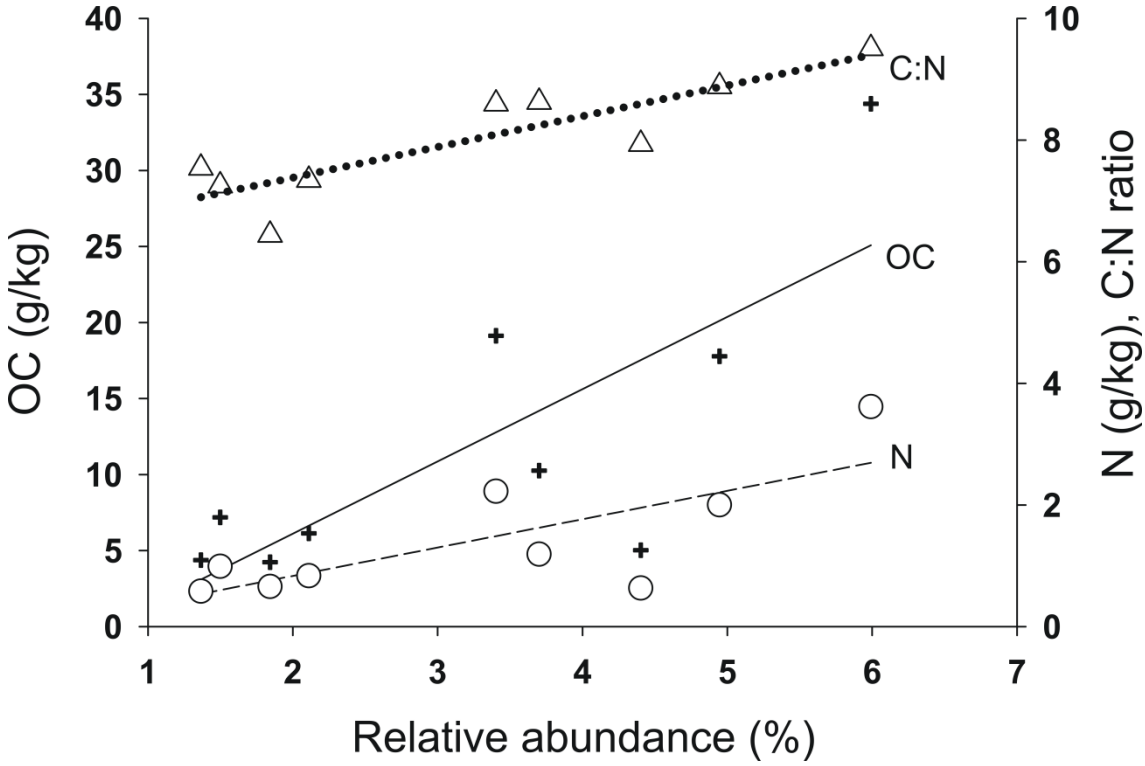


Fig. S3. Correlation of the relative abundance of *Rhizobiales* and the edaphic soil properties OC content, N content, and C:N ratio in the B horizon. Crosses and open circles indicate the OC and N content, respectively. The C:N ratio is marked by triangles.

3

Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories

Heiko Nacke[1*], Christiane Will[1*], Sarah Herzog[1], Boris Nowka[1], Martin Engelhaupt[1], and Rolf Daniel[1,2]

FEMS Microbiology Ecology (2011), Vol. 78, p. 188-201

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany

Author contributions to the work:

* H.N. and C.W. contributed equally to this work.

Performed the experiments: HN, CW, SH, BN, ME.

Analyzed data: HN, CW.

Wrote the paper: HN, CW, RD.

Conceived and designed the experiments: HN, CW, RD.

RESEARCH ARTICLE

Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories

Heiko Nacke[1], Christiane Will[1], Sarah Herzog[1], Boris Nowka[1], Martin Engelhaupt[1] & Rolf Daniel[1,2]

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany; and [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Göttingen, Germany

Correspondence: Rolf Daniel, Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany. Tel.: +49 551 393 827; fax: +49 551 391 2181; e-mail: rdaniel@gwdg.de

Present address: Boris Nowka, Universität Hamburg, Biozentrum Klein Flottbek, Ohnhorststraße 18, D-22609 Hamburg, Germany.

Received 24 December 2010; revised 2 March 2011; accepted 3 March 2011.
Final version published online 6 April 2011.

DOI:10.1111/j.1574-6941.2011.01088.x

Editor: Christoph Tebbe

Keywords
soil metagenome; metagenomic libraries; lipolytic enzymes.

Abstract

Microbial metagenomes derived from soils are rich sources for the discovery of novel genes and biocatalysts. Fourteen environmental plasmid and seven fosmid libraries obtained from 10 German forest soils (A horizons) and six grassland soils (A and B horizons) were screened for genes conferring lipolytic activity. The libraries comprised approximately 29.3 Gb of cloned soil DNA. Partial activity-based screening of the constructed libraries resulted in the identification of 37 unique lipolytic clones. The amino acid sequences of the 37 corresponding lipolytic gene products shared 29–90% identity to other lipolytic enzymes, which were mainly uncharacterized or derived from uncultured microorganisms. Multiple sequence alignments and phylogenetic tree analysis revealed that 35 of the predicted proteins were new members of known families of lipolytic enzymes. The remaining two gene products represent two putatively new families. In addition, sequence analysis indicated that two genes encode true lipases, whereas the other genes encode esterases. The determination of substrate specificity and chain-length selectivity using different triacylglycerides and *p*-nitrophenyl esters of fatty acids as substrates supported the classification of the esterases.

Introduction

Lipolytic enzymes such as lipases (EC 3.1.1.3) and esterases (EC 3.1.1.1.) are ubiquitous enzymes found in animals, plants, and microorganisms. These enzymes exhibit broad substrate specificity and catalyze both the hydrolysis and the synthesis of esters formed from glycerol and fatty acids. Lipases resemble esterases, but differ from them in their ability to act on water-insoluble esters (Arpigny & Jaeger, 1999). Lipolytic enzymes have been recognized as very useful biocatalysts because of their wide-ranging versatility in industrial applications, including food technology, detergent production, biodiesel formation, fine chemistry, and biomedical sciences (Jaeger & Eggert, 2002).

Soils harbor enormously diverse microbial communities and are a major reservoir of microbial genomic and taxonomic diversity. The microbial diversity in soils exceeds that of other environments and, by far, that of eukaryotic organisms. One gram of soil may contain up to 10 billion microorganisms of possibly thousands of different species (Rosselló-Mora & Amann, 2001). Soil microorganisms have been the major source for lipolytic enzymes and other biomolecules of industrial importance (Strohl, 2000). However, of late, the discovery rate of novel biomolecules is extremely low by applying traditional cultivation techniques, because most of the soil microorganisms cannot be cultured and only a small fraction of soil microbial diversity is assessed in this way. Culture-dependent methods have been complemented or replaced by culture-independent metagenomic approaches, which theoretically provide access to the collective nucleic acids of all indigenous microorganisms present in the studied environment (Handelsman, 2004;

FEMS MICROBIOLOGY ECOLOGY

Daniel, 2005). Functional metagenomics comprising the isolation of DNA from environmental samples without prior enrichment of individual microorganisms, construction of libraries from the recovered DNA, and function-driven screening of the generated libraries has led to the identification and characterization of a variety of novel enzymes (Ferrer *et al.*, 2005; Simon & Daniel, 2009; Steele *et al.*, 2009), including lipolytic enzymes (Roh & Villate, 2008; Rashamuse *et al.*, 2009). Lipolytic enzymes have been derived from different environmental samples such as soils (Henne *et al.*, 2000; Lee *et al.*, 2004; Elend *et al.*, 2006), sea water (Chu *et al.*, 2008), and sediments (Jeon *et al.*, 2008, 2009).

In this study, we used the soil metagenome as a source for the recovery of novel genes encoding lipolytic enzymes. We report on the construction of small-insert and large-insert metagenomic libraries from 16 different forest and grassland soil samples, which were derived from the three German Biodiversity Exploratories Hainich-Dün, Schorfheide-Chorin, and Schwäbische Alb (Fischer *et al.*, 2010). Subsequently, the constructed libraries were subjected to activity-based screening for genes encoding lipolytic enzymes. In this way, 37 novel lipolytic enzymes were identified. Lipolytic activities of the genes were confirmed by analysis of substrate specificity and chain-length selectivity. Our results demonstrate that function-driven soil-based metagenomics is a very powerful approach for the discovery of novel biomolecules and soil microorganisms can continue to play a major role as a resource for natural product discovery.

Materials and methods

Soil sampling

Soil samples used for metagenomic library construction were derived from the A horizons of 10 forest and six grassland sites of the three German Biodiversity Exploratories Hainich-Dün (samples HEG1, HEG9, HEW2, HEW5, HEW9), Schorfheide-Chorin (samples SEG2, SEG6, SEG9, SEW2, SEW5, SEW8), and Schwäbische Alb (samples AEG2, AEW1, AEW4, AEW5, AEW9). In addition, the B horizons from the Hainich-Dün samples HEG1 and HEG9 were used. Samples were collected in April and May 2008. Sampling was performed as described by Will *et al.* (2010) and Nacke *et al.* (2011). Descriptions of the sampling sites and soil characteristics are provided in Supporting Information, Table S1. Names of the metagenomic libraries refer to the designation of the samples from which the libraries were derived.

Isolation of soil DNA and construction of metagenomic DNA libraries

Total microbial community DNA was isolated from 10 g of soil per sample. For this purpose, the MoBio Power Max Soil DNA extraction kit (MoBio Laboratories, Carlsbad, CA) was used according to the instructions of the manufacturer.

Small-insert libraries were constructed using the plasmid pCR-XL-TOPO as a vector (TOPO XL PCR Cloning Kit; Invitrogen GmbH, Karlsruhe, Germany). Approximately 10 μg extracted DNA of each soil sample was separated by agarose gel electrophoresis. Subsequently, DNA fragments > 6 kb were recovered and purified from the gels using the peqGold Gel Extraction Kit (Peqlab Biotechnologie GmbH, Erlangen, Germany). The purified DNA fragments were subjected to blunt-end polishing using T4 DNA polymerase (MBI Fermentas, St Leon-Rot, Germany) as suggested by the manufacturer. Subsequently, the DNA was purified using SureClean solution (Bioline GmbH, Luckenwalde, Germany) and the resulting DNA pellet was suspended in 35 μL H_2O. Subsequently, a deoxyadenosine was added to the 3′ termini of the DNA to facilitate cloning by the TA method. For this purpose, 1 μL dATP solution (100 mM), 6 μL $MgCl_2$ solution (25 mM), 7 μL of 10-fold $(NH_4)_2SO_4$-containing *Taq* DNA polymerase buffer (MBI Fermentas), 1 μL of *Taq* DNA polymerase (5 U), and 20 μL of H_2O were mixed with the DNA solution, incubated at 72 °C for 30 min, and purified using SureClean solution (Bioline GmbH). The resulting DNA pellet was suspended in 15 μL H_2O and dephosphorylated using 5 U Antarctic Phosphatase (NEB, Ipswich, MA) as described by the manufacturer. Finally, the recovered DNA fragments were inserted into pCR-XL-TOPO using the TOPO XL PCR cloning kit (Invitrogen). To screen the small-insert metagenomic libraries for lipolytic activity, *Escherichia coli* DH5α (Ausubel *et al.*, 1987) was used as a host. Large-insert metagenomic libraries were constructed using the fosmid pCC1FOS as a vector and the Copy Control Fosmid Library Production kit (Epicentre Biotechnologies, Madison, WI) as recommended by the manufacturer. The extracted DNA (5 μg) was directly inserted into the fosmid without prior size fractionation (Simon & Daniel, 2010). Subsequently, the resulting recombinant fosmids were packaged into λ phages using MaxPlax Lambda Packaging Extracts (Epicentre Biotechnologies), and used to infect *E. coli* EPI300-T1^R cells according to the protocol of the manufacturer.

Growth condition and activity-based screening

Escherichia coli strains were routinely grown in Luria–Bertani (LB) medium at 30 °C. For activity-based screening of metagenomic libraries, recombinant *E. coli* strains were grown under aerobic conditions in LB medium, which was supplemented with 1% tributyrin and solidified with agar (15 g L^{-1}). For the determination of substrate specificity, the following compounds were added instead of tributyrin: tricaproin, tricaprylin, tricaprin, trilaurin, trimyristin, and tripalmitin. In addition, to maintain the presence of

recombinant plasmids and fosmids, the medium contained 50 mg L^{-1} kanamycin or 12.5 mg L^{-1} chloramphenicol, respectively. Clones showing lipolytic activity were identified by the formation of clear zones (halos) against the creamy background after incubation for 1–7 days at 37 °C. To avoid the isolation of false-positive clones and to confirm that the lipolytic activity of the positive clones was plasmid encoded, the recombinant plasmids were isolated and used to transform *E. coli*. The resulting *E. coli* strains were screened again on tributyrin-containing agar.

Subcloning and sequence analysis

To subclone DNA fragments containing the lipolytic genes from large-insert fosmids, the recombinant fosmids from positive clones were sheared by sonication (UP200S Sonicator, Dr Hielscher GmbH, 5 s at 30% amplitude, cycle 0.5). Subsequently, sheared DNA fragments were separated by agarose gel electrophoresis, and 2–6-kb fragment were excised and extracted using the peqGold gel extraction kit (Peqlab Biotechnologie GmbH). The resulting DNA fragments were ligated into pCR-XL-TOPO or pCR4-TOPO (Invitrogen), and used to transform *E. coli* as recommended by the manufacturer (Invitrogen). The resulting recombinant *E. coli* strains were screened on tributyrin-containing indicator agar for the presence of genes conferring lipolytic activity.

The recombinant plasmids derived from all 37 positive clones were sequenced by the Göttingen Genomics Laboratory (Göttingen, Germany). The initial prediction of ORFs located on the inserts of plasmids pLE01–pLE08 and pLE10–pLE38 was accomplished using the ORF FINDER program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) provided by the National Center for Biotechnology Information and the ARTEMIS program (Rutherford *et al.*, 2000). The results were verified and improved manually using criteria such as the presence of a ribosome-binding site, GC frame plot analysis, and similarity to known lipolytic-protein-encoding sequences. Initial annotation of the deduced proteins was performed by searching the amino acid sequences against the public GenBank database using the BLAST program (Ye *et al.*, 2006). All coding sequences were examined for similarities to protein families and domains using searches against the CDD databases (Marchler-Bauer *et al.*, 2007). Signal peptides of putative lipolytic proteins were predicted using the SIGNALP 3.0 server (Bendtsen *et al.*, 2004). To construct a phylogenetic tree of the lipolytic proteins recovered, multiple alignments of the deduced protein sequences were performed using CLUSTALW version 2.0.12 (Thompson *et al.*, 1994) and examined with the BIOEDIT program (Hall, 1999). The phylogenetic tree was constructed using the program MEGA version 4.0.2 (Tamura *et al.*, 2007) using the neighbor-joining method. Bootstrapping based on 1000 resamplings was used to estimate the robustness of the tree (Felsenstein, 1985).

The nucleotide sequences of the recombinant plasmids harboring the esterase genes *est01–est08* and *est10–est38* (pLE01–pLE08 and pLE10–pLE38) have been submitted to GenBank under accession numbers HQ156900–HQ156907 and HQ156909–HQ156937.

Lipase/esterase activity assay

To analyze the lipolytic activity of *E. coli* cells harboring the individual plasmids pLE01–pLE08 and pLE10–pLE38, *p*-nitrophenyl esters of fatty acids were used as substrates. *Escherichia coli* strains carrying the cloning vector pCR-XL-TOPO were used as controls. The *E. coli* clones were grown in LB medium to an $OD_{600\,nm}$ of 3.0–4.0 and the cell cultures were directly used for the lipolytic activity assay described by Lee *et al.* (2006), with modifications. The activity was determined by measuring *p*-nitrophenol formation from enzymatic hydrolysis of fatty acid *p*-nitrophenyl esters (C_4, *p*-nitrophenyl butyrate; C_6, *p*-nitrophenyl caproate; C_8, *p*-nitrophenyl caprylate; C_{10}, *p*-nitrophenyl caprate; C_{12}, *p*-nitrophenyl laurate; and C_{16}, *p*-nitrophenyl palmitate). Measurements were performed at 410 nm using a Cary 100 UV-visible spectrophotometer with a dual cell peltier accessory (Varian Inc., Vic., Australia). Enzyme activity was measured at 25 °C. The reaction mixture contained 890 μL 50 mM Tris-HCl (pH 7.5), 100 μL culture supernatant, and 10 μL 1 mM *p*-nitrophenyl ester as a substrate. The reaction was initiated by substrate addition. One unit (U) of enzyme activity was defined as the amount of activity required for the release of 1 μmol *p*-nitrophenol min^{-1} from *p*-nitrophenyl ester.

Results and discussion

Construction of environmental DNA libraries

DNA derived from soil samples of six grassland and 10 forest plots was used for the construction of 21 metagenomic DNA libraries (Table 1). DNA was directly isolated from the samples without previous enrichment or extraction of microbial cells. The DNA yield ranged in the A horizons from 17 to 56 μg g^{-1} soil and in the B horizons from 7 to 9 μg g^{-1} soil (Table S1). The yield of the soil DNA derived from the A horizons exceeded the 2–15 μg g^{-1} soil described in other publications (Lee *et al.*, 2004; Hong *et al.*, 2007; Lämmle *et al.*, 2007). In recently published studies, soil-derived metagenomic libraries comprised 11 000–80 000 and 2400–200 000 clones using plasmids and fosmids as vectors, respectively (Hong *et al.*, 2007; Kim *et al.*, 2007; Tirawongsaroj *et al.*, 2008; Cieśliński *et al.*, 2009; Waschkowitz *et al.*, 2009; Couto *et al.*, 2010). In this study, the 14 plasmid libraries and the seven fosmid libraries contained

Table 1. Characterization of constructed metagenomic libraries and screening for genes conferring lipolytic activity on *Escherichia coli*

Library*	Sample site	Vector	Number of clones	Average insert size (kb)†	Insert frequency (%)	Estimated library size (Gb)	Screening extend (Gb)	No. of lipolytic *E. coli* clones (designation)
SEG2	Schorfheide	pCC1FOS	86 944	24.0	100	2.09	0.22	1 (pLE01)
SEG6	Schorfheide	pCR-XL-TOPO	39 825	6.0	91	0.22	0.22	1 (pLE02)
SEG9	Schorfheide	pCR-XL-TOPO	68 770	7.3	94	0.47	0.47	1 (pLE03)
SEG9	Schorfheide	pCC1FOS	147 888	23.3	100	3.45	0.29	1 (pLE04)
SEW2	Schorfheide	pCR-XL-TOPO	135 240	5.7	95	0.73	0.45	1 (pLE05)
SEW5	Schorfheide	pCR-XL-TOPO	166 040	4.0	95	0.63	0.34	1 (pLE06)
SEW8	Schorfheide	pCR-XL-TOPO	69 984	5.5	90	0.35	0.35	1 (pLE07)
HEG1	Hainich	pCR-XL-TOPO	70 313	2.6	98	0.18	0.18	1 (pLE08)
HEG9	Hainich	pCR-XL-TOPO	161 940	6.4	69	0.72	0.72	6 (pLE10-pLE15)
HEG1‡	Hainich	pCR-XL-TOPO	510 808	5.7	97	2.80	2.80	2 (pLE16; pLE17)
HEG9‡	Hainich	pCR-XL-TOPO	150 782	9.4	96	1.36	0.54	4 (pLE18-pLE21)
HEW2	Hainich	pCR-XL-TOPO	340 990	4.9	88	1.48	1.48	2 (pLE22; pLE23)
HEW5	Hainich	pCR-XL-TOPO	181 958	6.7	92	1.13	1.13	4 (pLE24-pLE27)
HEW9	Hainich	pCC1FOS	60 000	27.8	100	1.67	0.13	1 (pLE28)
AEG2	Schwäbische Alb	pCC1FOS	299 880	26.3	100	7.89	0.64	2 (pLE29; pLE30)
AEW1	Schwäbische Alb	pCR-XL-TOPO	129 748	6.7	91	0.79	0.79	2 (pLE31; pLE32)
AEW1	Schwäbische Alb	pCC1FOS	50 952	27.8	100	1.42	0.63	2 (pLE33; pLE34)
AEW4	Schwäbische Alb	pCC1FOS	61 530	19.0	100	1.17	0.18	1 (pLE35)
AEW5	Schwäbische Alb	pCR-XL-TOPO	90 300	5.2	89	0.42	0.42	1 (pLE36)
AEW5	Schwäbische Alb	pCC1FOS	4600	30.0	100	0.14	0.14	1 (pLE37)
AEW9	Schwäbische Alb	pCR-XL-TOPO	100 950	2.6	89	0.23	0.13	1 (pLE38)

*Names of the metagenomic libraries refer to the designation of the sampling sites. Descriptions of the sampling sites are provided in Table S1.
†The average insert size was determined by analysis of 20 insert-containing recombinant plasmids or fosmids.
‡Libraries constructed from soil derived from B horizon.

approximately 40 000–341 000 clones and 4600–300 000 clones, respectively (Table 1). The quality of the 21 different environmental libraries was controlled by determination of the average insert sizes and the percentage of insert-bearing *E. coli* clones. The average insert sizes ranged from 2.6 to 9.4 kb (plasmids) and 19 to 30 kb (fosmids). These values are in accordance with those published for other soil-derived small-insert or large-insert metagenomic libraries (Henne *et al.*, 2000; Lämmle *et al.*, 2007; Waschkowitz *et al.*, 2009; Couto *et al.*, 2010). The percentage of insert-carrying clones was 69–98% (plasmids) and 100% (fosmids). Thus, approximately 11.5 and 17.8 Gb of cloned soil DNA were stored in the constructed plasmid and fosmid libraries, respectively. Assuming an average prokaryotic genome size of 5 Mb (Hårdeman & Sjöling, 2007), both types of metagenomic libraries represented approximately 5500 prokaryotic genomes.

Screening for genes conferring lipolytic activity

To isolate genes encoding lipolytic activity from the constructed libraries, a function-driven approach was chosen. As sequence information is not required before screening, this is the only strategy that bears the potential to discover entirely novel lipolytic genes (Daniel, 2005; Simon & Daniel, 2009). In addition, it is selective for full-length genes and functional gene products. The screen for genes exhibiting lipolytic activity was based on the ability of library-bearing *E. coli* clones to form halos when grown on agar medium containing tributyrin. Halo formation is caused by the hydrolysis of tributyrin. This function-based screen has been used to identify the lipolytic activity of individual microorganisms (Seo *et al.*, 2005; Hantsis-Zacharov & Halpern, 2007), and recombinant *E. coli* strains that harbor gene libraries from single microorganisms (Hotta *et al.*, 2002) or metagenomic libraries (Heath *et al.*, 2009; Hu *et al.*, 2010). In the case of metagenomic libraries, genes conferring lipolytic activity have been recovered from diverse environments such as mangrove sediment (Couto *et al.*, 2010), marine sediment (Hårdeman & Sjöling, 2007; Hu *et al.*, 2010), water samples (Ranjan *et al.*, 2005; Chu *et al.*, 2008; Wu & Sun, 2009), compost (Lämmle *et al.*, 2007), and soils (Henne *et al.*, 2000; Lee *et al.*, 2004; Kim *et al.*, 2005; Hong *et al.*, 2007).

In this study, the constructed plasmid libraries and fosmid libraries were partially screened using the above-mentioned activity-based approach. The screening effort comprised approximately 10.0 Gb (plasmids) and 2.2 Gb (fosmids) of the cloned soil DNA. Positive *E. coli* clones were collected after incubation for 1–7 days at 37 °C on tributyrin-containing indicator agar. In order to confirm that the lipolytic activity of the positive clones was vector

encoded, the recombinant plasmids or fosmids were isolated from the positive clones and used to transform *E. coli*. The resulting *E. coli* strains were screened again on indicator agar. Twenty-eight different recombinant plasmids and nine fosmids conferred a stable lipolytic phenotype (Table 1). Eighteen of these were derived from A horizons of forest soil samples, 13 from A horizons of grassland soil samples, and six from B horizons of grassland soil samples.

In this study, the average hit rate was approximately one lipolytic gene per 240 Mb (fosmids) and 360 Mb (plasmids), respectively, of screened soil DNA. For comparison, one lipolytic gene per 480 Mb of screened soil DNA, which was cloned into a high-copy plasmid vector, was identified during screening on tributyrin agar in another study (Henne *et al.*, 2000). Other soil metagenomic studies using fosmid or bacterial artificial chromosomes as vectors achieved hit rates of one lipolytic gene per 148 Mb (Lee *et al.*, 2004) or 50 Mb of screened soil DNA (Rondon *et al.*, 2000).

Molecular analysis

The inserts of all 28 recombinant plasmids (pLE02, pLE03, pLE05–pLE08, pLE10–pLE27, pLE31, pLE32, pLE36, and pLE38) recovered from the positive clones were sequenced. The insert sizes of the plasmids ranged from 1107 to 11 172 bp (Table S2). The nine fosmid-harboring lipolytic clones carried insert DNA ranging from 19 800 to 36 500 bases in size (Table S2). DNA fragments of fosmids containing the genes conferring lipolytic activity were identified by subcloning and screening for lipolytic subclones before sequencing. In this way, recombinant plasmids carrying the desired DNA fragments for all nine fosmids were recovered. The insert sizes of the nine corresponding plasmids (pLE01, pLE04, pLE28, pLE29, pLE30, pLE33, pLE34, pLE35, and pLE37) ranged from 1511 to 3568 bp (Table S2). The insert sequences of pLE01–pLE08 and pLE10–pLE38 were sequenced and analyzed. In all 37 cases, a putative gene showing similarities to known genes encoding lipases or esterases was found. The amino acid sequences deduced from the 37 identified genes (*est01–est08* and *est10–est38*) comprised 230–556 amino acids with calculated molecular masses from 25.1 to 57.9 kDa. The sequence identities to the closest similar known lipolytic protein ranged from 29% to 90% (Table 2). Fourteen of the 37 putative lipolytic proteins showed the highest similarity to esterases/lipases from uncultured bacteria and the remaining 23 to lipolytic proteins deduced from genome sequences of individual microorganisms. Interestingly, 50% of the lipolytic genes derived from forest soils (nine genes), but only 30% of those recovered from grassland soils (five genes) showed the closest similarity to esterases/lipases from uncultured bacteria. In addition, almost all of the forest soil-derived enzymes showed the closest amino acid identity (48–87%) to six putative lipolytic enzymes that have been recovered during activity-based screening of a Korean forest soil-derived metagenomic library (Lee *et al.*, 2004). In the remaining cases in which a lipolytic protein from an uncultured bacterium was the best hit, the matching lipase/esterase was detected in other recently published metagenomic surveys (Hong *et al.*, 2007; Hu *et al.*, 2010). Hu *et al.* (2010) used samples from marine sediment and the identities to our lipolytic enzyme sequences were lower than those to our lipolytic enzymes recovered from other soil metagenomic libraries (data not shown). Thus, the similarity of the habitat seems to have an impact on the degree of amino acid identity.

Classification of the lipolytic enzymes

We used the classification system of Arpigny & Jaeger (1999), who subdivided bacterial lipolytic enzymes into eight different families (I–VIII) based on the amino acid sequence similarity and some fundamental biological properties. As shown in Fig. 1, 35 of the enzyme sequences grouped into five of these families. The majority of the classified enzymes were affiliated to family IV, followed by families V, VIII, I, and VI. The remaining two enzyme sequences (Est01 and Est15) could not be assigned to the families described by Arpigny & Jaeger (1999).

Family IV

Most of the enzyme sequences (17 sequences) were affiliated to family IV. Members of this family show significant similarity to the mammalian hormone-sensitive lipase (HSL). Therefore, family IV is also known as the 'HSL family' of lipolytic enzymes (Arpigny & Jaeger, 1999; Hausmann & Jaeger, 2010). All 17 enzyme sequences contained the lipase-conserved catalytic triad residues aspartate, histidine and the nucleophile serine in the consensus pentapeptide motif GXSXG (Fig. 2). The only exceptions were the amino acid sequences of Est05, Est06, Est29, Est34, and Est38, in which the aspartate residue was replaced by a glutamate residue (Fig. 2). This substitution is common for members of family IV (Chu *et al.*, 2008; Hu *et al.*, 2010). The highly conserved motif HGGGF was present in 16 enzyme sequences. In the amino acid sequence of Est35, the motif HGGGF was replaced by PGGGF (Fig. 2). Lipolytic enzymes of the HSL family were predominant and widespread in our samples, as the 17 enzymes belonging to this family originated from forest and grassland samples and were present in all three German Biodiversity Exploratories studied. Lipolytic enzymes of family IV were also predominantly recovered in similar activity-based screens of other metagenomic libraries derived from a variety of environments such as forest soil (Lee *et al.*, 2004; Hong *et al.*, 2007), deep sea sediment (Hu *et al.*, 2010), and arctic seashore sediment

Table 2. Description of the lipolytic gene products and their observed sequence similarities

Gene (accession no.)	No. of encoded amino acids	Closest similar lipolytic protein, accession no. (no. of encoded amino acids), organism	*E* value	Amino acid homology to the closest similar lipolytic protein (% identity)
est01 (HQ156900)	397	Phospholipase/carboxylesterase, ZP_06237474 (349), *Frankia* sp. EuI1c	3e − 23	101/298 (33%)
est02 (HQ156901)	311	Lipase/esterase, ABQ11271 (310), uncultured bacterium	1e − 79	160/311 (51%)
est03 (HQ156902)	333	Putative α/β hydrolase, YP_555239 (276), *Burkholderia xenovorans* LB400	2e − 42	101/267 (37%)
est04 (HQ156903)	294	Lipase/esterase, ABQ11272 (296), uncultured bacterium	1e − 61	133/267 (49%)
est05 (HQ156904)	310	Lipase/esterase, AAS77233 (296), uncultured bacterium	3e − 140	239/284 (84%)
est06 (HQ156905)	296	Lipase/esterase, AAS77236 (296), uncultured bacterium	1e − 136	235/284 (82%)
est07 (HQ156906)	321	Putative lipase, ZP_05767946 (319), *Mycobacterium tuberculosis* T46	8e − 138	243/320 (75%)
est08 (HQ156907)	384	Esterase, ZP_04691113 (371), *Streptomyces ghanaensis* ATCC 14672	9e − 64	144/365 (39%)
est10 (HQ156909)	337	β-Lactamase, YP_001682441 (408), *Caulobacter* sp. K31	0.0	305/337 (90%)
est11 (HQ156910)	314	α/β hydrolase fold-3 domain protein, YP_002946812 (314), *Variovorax paradoxus* S110	2e − 132	239/314 (76%)
est12 (HQ156911)	408	β-Lactamase, YP_001530546 (390), *Desulfococcus oleovorans* Hxd3	1e − 89	179/391 (45%)
est13 (HQ156912)	272	Putative lipase, YP_766845 (293), *Rhizobium leguminosarum* bv. *viciae* 3841	2e − 39	89/232 (38%)
est14 (HQ156913)	377	Putative esterase, ZP_01617169 (381), marine gammaproteobacterium HTCC2143	4e − 98	193/380 (50%)
est15 (HQ156914)	448	Lipase, ZP_06703106 (440), *Xanthomonas fuscans* ssp. *aurantifolii* str. ICPB 11122	1e − 04	52/177 (29%)
est16 (HQ156915)	272	α/β fold family hydrolase, YP_628483 (314), *Myxococcus xanthus* DK 1622	2e − 24	87/262 (33%)
est17 (HQ156916)	556	Lipase class 2, YP_953514 (573), *Mycobacterium vanbaalenii* PYR-1	1e − 113	222/357 (62%)
est18 (HQ156917)	331	Lipolytic enzyme, ACL67843 (311), uncultured bacterium	2e − 73	151/308 (49%)
est19 (HQ156918)	337	Triacylglycerol lipase, ZP_06045720 (317), *Aeromicrobium marinum* DSM 15272	5e − 84	159/298 (53%)
est20 (HQ156919)	443	β-Lactamase, YP_589716 (424), *Candidatus* Koribacter versatilis Ellin345	1e − 108	218/433 (50%)
est21 (HQ156920)	230	Phospholipase/carboxylesterase, YP_413093 (227), *Nitrosospira multiformis* ATCC 25196	1e − 67	128/223 (57%)
est22 (HQ156921)	424	β-Lactamase, YP_577943 (424), *Nitrobacter hamburgensis* X14	0.0	340/424 (80%)
est23 (HQ156922)	310	Lipase/esterase, ABQ11271 (310), uncultured bacterium	1e − 158	268/310 (86%)
est24 (HQ156923)	277	α/β hydrolase fold protein, YP_003321859 (267), *Thermobaculum terrenum* ATCC BAA-798	4e − 37	86/268 (32%)
est25 (HQ156924)	276	α/β hydrolase domain-containing protein, ZP_03632535 (297), bacterium Ellin514	2e − 49	115/273 (42%)
est26 (HQ156925)	257	α/β hydrolase fold, YP_001773617 (254), *Methylobacterium* sp. 4-46	6e − 87	164/252 (65%)
est27 (HQ156926)	402	β-Lactamase, YP_484201 (395), *Rhodopseudomonas palustris* HaA2	6e − 84	155/382 (40%)
est28 (HQ156927)	245	α/β fold family hydrolase, YP_360549 (258), *Carboxydothermus hydrogenoformans* Z-2901	8e − 37	94/232 (40%)
est29 (HQ156928)	300	Lipolytic enzyme, ACL67845 (307), uncultured bacterium	1e − 78	158/299 (52%)
est30 (HQ156929)	363	Esterase, AAY45707 (362), uncultured bacterium	2e − 105	199/351 (56%)
est31 (HQ156930)	312	Lipase/esterase, AAS77247 (311), uncultured bacterium	8e − 127	214/312 (68%)
est32 (HQ156931)	266	Carboxylesterase (est-1), NP_069699 (266), *Archaeoglobus fulgidus* DSM 4304	3e − 30	91/277 (32%)
est33 (HQ156932)	297	Lipase/esterase, ABQ11272 (296), uncultured bacterium	5e − 102	178/296 (60%)
est34 (HQ156933)	296	Lipase/esterase, AAS77236 (296), uncultured bacterium	1e − 144	248/285 (87%)
est35 (HQ156934)	314	α/β hydrolase fold-3 domain protein, ZP_05908953 (314), *Vibrio parahaemolyticus* AQ4037	2e − 51	113/285 (39%)
est36 (HQ156935)	312	Lipolytic enzyme, ACL67843 (311), uncultured bacterium	2e − 99	186/311 (59%)
est37 (HQ156936)	309	Lipase/esterase, AAS77247 (311), uncultured bacterium	1e − 78	151/309 (48%)
est38 (HQ156937)	330	Lipase/esterase, AAX37296 (297), uncultured bacterium	2e − 92	169/272 (62%)

(Jeon *et al.*, 2009). Hong *et al.* (2007) identified four enzymes clustering with family IV lipolytic enzymes by screening of a forest-soil metagenomic library. In addition, two enzymes showed similarity to family V lipolytic enzymes and one to a lysophospholipase from family II. However, 20 of the 37 putative lipolytic enzymes identified in this study were not affiliated to family IV.

Family V

Eight of the lipolytic enzyme sequences grouped into family V (Est03, Est13, Est16, Est24–Est26, Est28, and Est32). The multiple sequence alignment revealed that all of them contained the catalytic triad residues (Fig. 2). The consensus motif GXSXG was present in all protein sequences of the

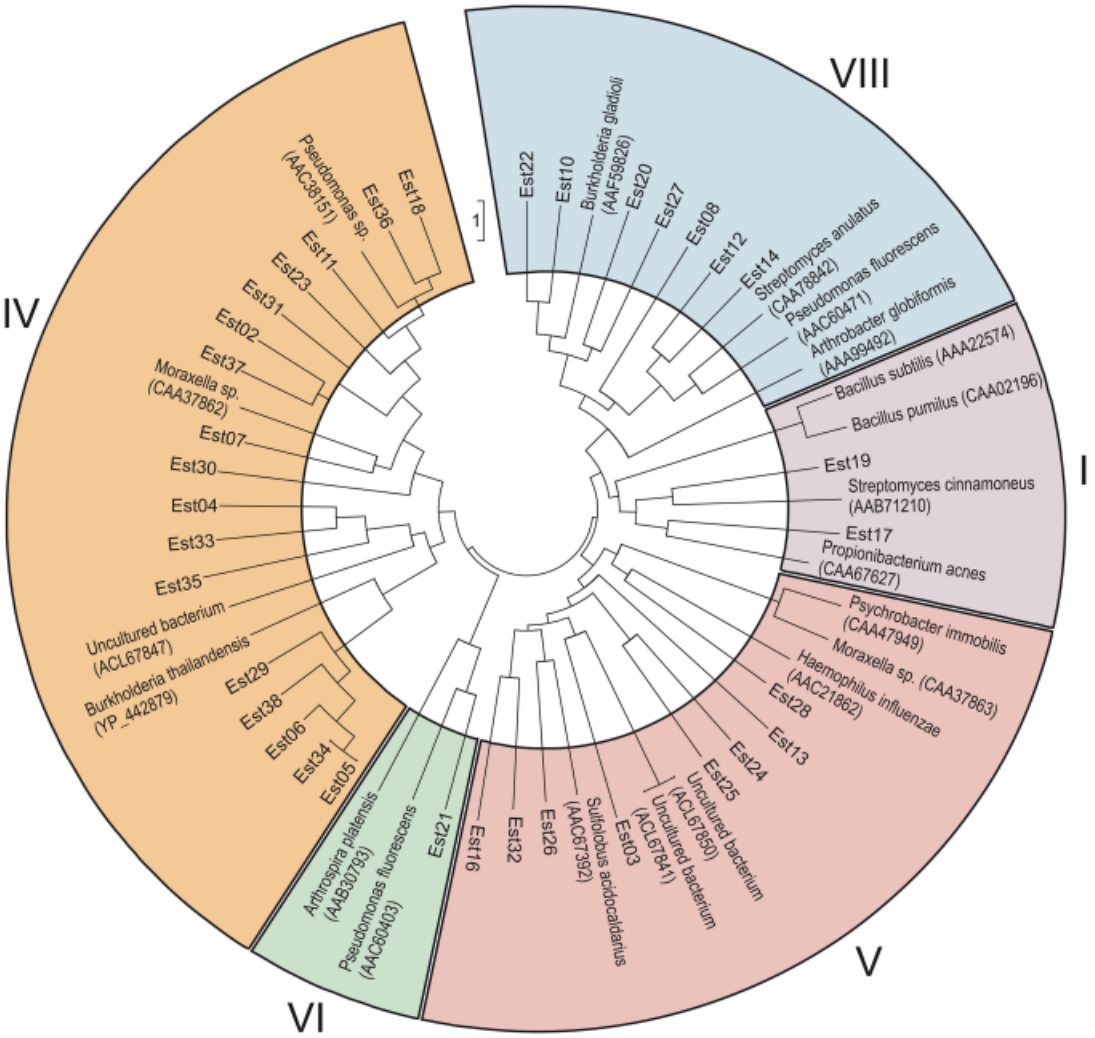


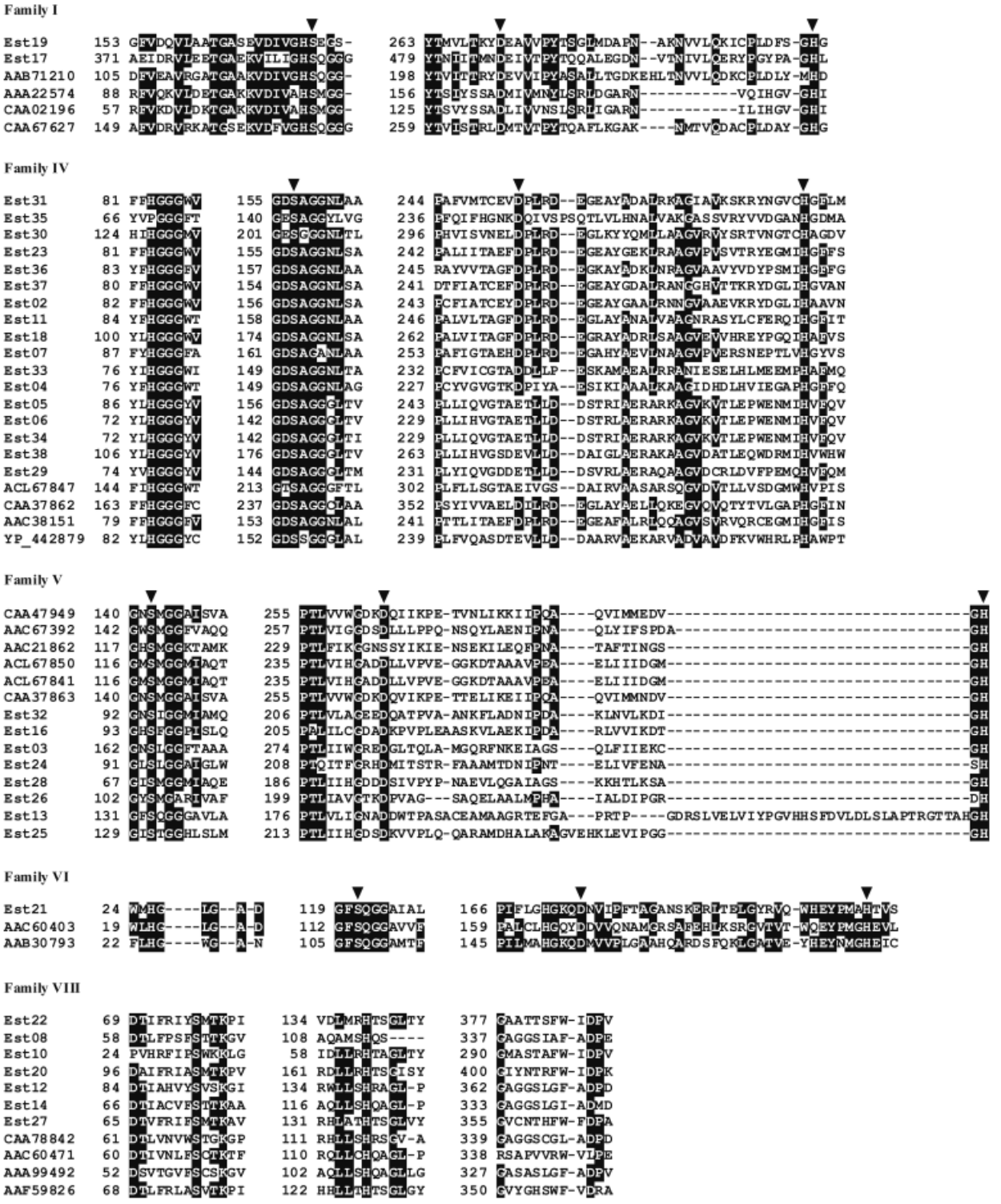
Fig. 1. Unrooted neighbor-joining tree of lipolytic enzymes obtained from forest and grassland soil metagenomes in this study and representative members of families I, IV, V, VI, and VIII. Amino acid sequences of published esterases/lipases were retrieved from GenBank. Lipolytic enzymes belonging to putative new families were analyzed in an alignment with similar proteins retrieved from GenBank (Fig. 3). The length of the branches of the phylogenetic tree indicates the difference of the protein sequences.

family V enzymes. The PTL motif, which is another common motif among family V esterases, was detected in six of the lipolytic enzymes affiliated to family V (Est03, Est13, Est25, Est26, Est28, and Est32) (Fig. 2). In addition, Est24 contained the amino acid sequence PTQ (Fig. 2), which is a known variation of the PTL motif (Arpigny & Jaeger, 1999). Est16 contained the amino acid sequence PAL instead of PTL.

Fig. 2. Multiple sequence alignment of conserved regions of lipolytic enzymes belonging to families I, IV, V, VI, and VIII. Identical amino acid residues are shown as white letters on a dark background. Triangles indicate amino acid residues belonging to the catalytic triad. References: Est02–Est14 and Est16–Est37 (this study); AAB71210, lipase LipA from *Streptomyces cinnamoneus*; AAA22574, lipase from *Bacillus subtilis*; CAA02196, lipase from *Bacillus pumilus*; CAA67627, triacylglycerol lipase from *Propionibacterium acnes*; ACL67847, lipolytic enzyme from uncultured bacterium; CAA37862, triacylglycerol lipase from *Moraxella* sp.; AAC38151, lipase from *Pseudomonas* sp. B11-1; YP_442879, lipase/esterase from *Burkholderia thailandensis* E264; CAA47949, triacylglycerol lipase from *Psychrobacter immobilis*; AAC67392, lipolytic enzyme from *Sulfolobus acidocaldarius*; AAC21862, putative esterase/lipase from *Haemophilus influenzae* Rd KW20; ACL67850, lipolytic enzyme from uncultured bacterium; ACL67841, lipolytic enzyme from uncultured bacterium; CAA37863, triacylglycerol lipase from *Moraxella* sp.; AAC60403, esterase II from *Pseudomonas fluorescens*; AAB30793, serine esterase from *Arthrospira platensis*; CAA78842, esterase A from *Streptomyces anulatus*; AAC60471, esterase III from *Pseudomonas fluorescens*; AAA99492, carboxylic ester hydrolase from *Arthrobacter globiformis*; AAF9826, esterase EstB from *Burkholderia gladioli*.

Family VIII

According to the multiple sequence alignment, seven esterases (Est08, Est10, Est12, Est14, Est20, Est22, and Est27) belonged to family VIII (Fig. 2). Esterases belonging to this family show a remarkable similarity to class C β-lactamases and penicillin-binding proteins (Bornscheuer, 2002). Members of this family comprise approximately 380 amino



residues with a molecular mass of approximately 42 kDa (Hausmann & Jaeger, 2010). The length of the amino acid sequences and the molecular mass of all seven enzymes were in the same range (337–443 amino acids and 36.9–48.6 kDa, respectively). The enzymes showed 39–90% identity to the closest similar known lipolytic enzyme, which was, in most cases, a putative β-lactamase (Table 2). The conserved pentapeptide GXSXG, which includes the active-site serine, was only present in the sequences of Est12 and Est14 (Fig. 2). Analysis of the protein sequence of the remaining enzymes revealed that the active site serine is part of the consensus motif SXXK, which is conserved in class C β-lactamases, penicillin-binding proteins, and family VIII esterases (Arpigny & Jaeger, 1999; Wagner *et al.*, 2002; Hausmann & Jaeger, 2010) (Fig. 2). Moreover, the family VIII motif LLXHXXG described by Ranjan *et al.* (2005) appeared in four of these lipolytic enzymes (Est10, Est12, Est14, and Est20) (Fig. 2). In the remaining three enzymes, the two leucine residues were replaced by alanine and methionine (Est08), leucine and methionine (Est22), or leucine and alanine residues (Est27) (Fig. 2). Thus, these variations indicated that the proposed motif is not as conserved as previously assumed. However, Est08, Est22, and Est27 did not form a separate branch within family VIII (Fig. 1). Interestingly, family VIII members were derived only from plots in the Hainich region and were absent in the other two Exploratories.

Families I and VI

Two enzyme sequences (Est17 and Est19) were affiliated to family I containing the true lipases and one to family VI (Est21). All three enzymes were derived from metagenomic libraries constructed from subsoil samples (B horizons). To our knowledge, no other studies on screening of subsoil-derived metagenomic libraries for genes conferring lipolytic enzymes have been published. True lipases (family I members) have rarely been discovered during activity-based screens of soil-derived and other metagenomic libraries for lipolytic enzymes. So far, cold-adapted and thermostable lipases (Wei *et al.*, 2009), as well as a lipase showing similarity to the lipase (Lip) of *Streptomyces albus* (Henne *et al.*, 2000) were derived from soil metagenomic libraries. Furthermore, an enzyme isolated from mangrove sediment formed a unique branch within family I of true lipases (Couto *et al.*, 2010). In this study, we detected two putative lipases. As it is typical for family I lipases, the active-site serine of Est17 and Est19 is embedded in the motif GHSXG (Arpigny & Jaeger, 1999) (Fig. 2).

Esterases belonging to family VI have also been recovered from metagenomic libraries that have been constructed from soil of a hot spring area (Kim *et al.*, 2005) and compost (Lämmle *et al.*, 2007). Arpigny & Jaeger (1999) described a molecular mass in the range of 23–26 kDa for esterases belonging to family VI. This is in accordance with the calculated molecular mass of Est21 (25.1 kDa). The multiple sequence alignment and the phylogenetic tree with two known esterases from family VI indicated that Est21 is a new member of family VI (Figs 1 and 2).

Novel families

Two of the lipolytic enzyme sequences (Est01 and Est15) could not be classified according to Arpigny & Jaeger (1999). In addition, these enzymes also did not group into other novel families of lipolytic enzymes, which have been described recently (Lee *et al.*, 2006; Chu *et al.*, 2008; Kim *et al.*, 2009; Bayer *et al.*, 2010; Hu *et al.*, 2010). Est01 has a size of 397 amino acids and a predicted molecular mass of 40.7 kDa. A phospholipase/carboxylesterase from *Frankia* sp. was the closest similar known lipolytic protein (33% amino acid identity) (Table 2). An alignment (Fig. 3a) of Est01 with similar known lipolytic proteins revealed a conserved GHSXG (amino acids 195–199) motif containing the active-site serine. The other putative residues of the catalytic triad, histidine and aspartate, are also highly conserved in the aligned sequences (Fig. 3a). The overall low similarities to members of already existing families of lipolytic enzymes indicated that Est01 is a member of a new family of lipolytic enzymes. As Est01 was derived from the German Biodiversity Exploratory Schorfheide, we propose to designate the putative new family EstGS. The protein sequence of Est01 showed the closest similarity (31% and 33%) to two hypothetical proteins from *Congregibacter litoralis* KT71 (ZP_01103967) and gammaproteobacterium NOR5-3 (ZP_05127748). These proteins are probably also members of the new EstGS family.

Est15 is one of the largest lipolytic enzymes detected within this study. It comprises 448 amino acids with a predicted molecular mass of 45.9 kDa (Table 2). The closest similar lipolytic enzyme is a lipase from *Xanthomonas fuscans* (29% identity; ZP_06703106). The deduced protein sequence of Est15 exhibited no significant similarity to conserved motifs described for true lipases or other lipolytic enzyme families. An alignment (Fig. 3b) with the four best-matching protein sequences (55–66% identity), which were all hypothetical proteins from *Mycobacteria* (YP_001073968, YP_642492, YP_890989, and YP_951301), revealed highly conserved sequence regions. The putative active-site serine is embedded in a GHSLG motif (amino acids 273–277). For the other conserved residues of the catalytic triad, histidine and aspartate, analysis of the alignment revealed several candidate histidine and aspartate residues. Thus, the results indicated that Est15 and the hypothetical proteins of *Mycobacteria* are members of a new family of lipolytic enzymes. As Est15 was derived from the German Biodiversity

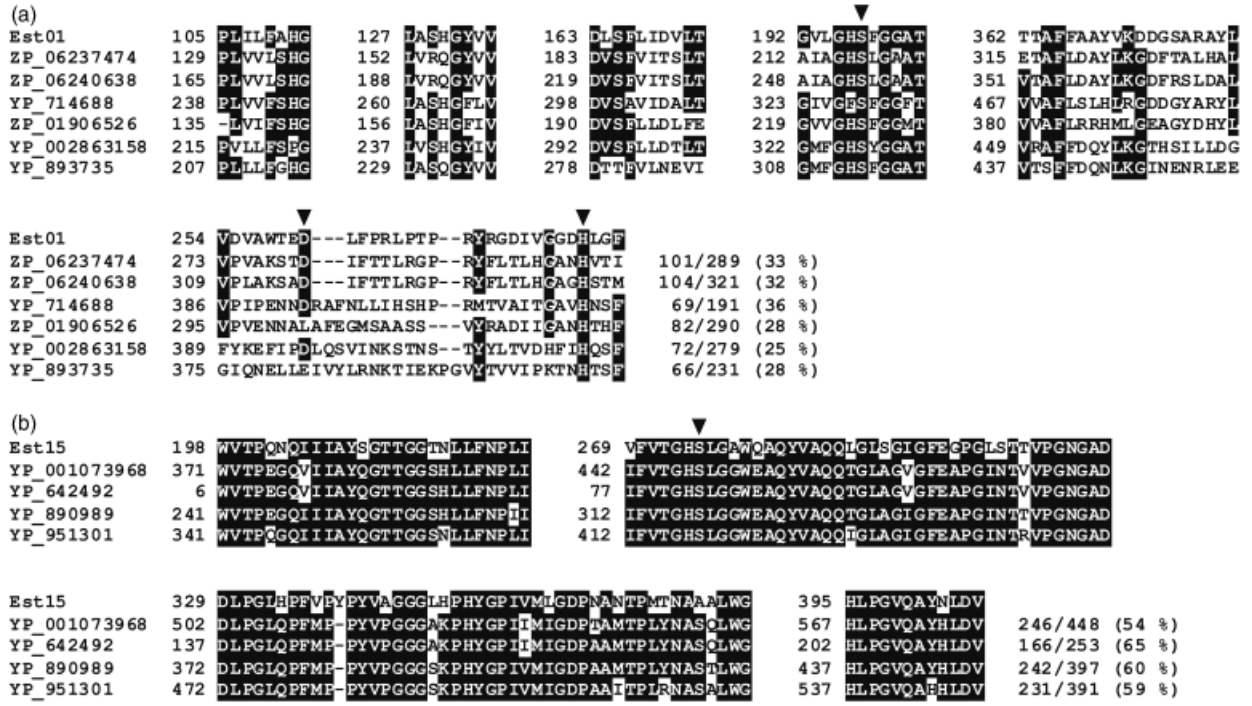


Fig. 3. Multiple sequence alignment of partial amino acid sequences harboring conserved regions of homology. Identical amino acid residues are shown as white letters on a dark background. The identity of closely related proteins to putative lipolytic enzymes identified in this study is indicated behind each sequence. Triangles indicate amino acid residues belonging to the catalytic triad. (a) Sequence alignment of Est01 with similar proteins. (b) Sequence alignment of Est15 with closely related proteins. References: Est01 and Est15 (present study); ZP_06237474, phospholipase/carboxylesterase from *Frankia* sp. EuI1c; ZP_06240638, phospholipase/carboxylesterase from *Frankia* sp. EuI1c; YP_714688, putative secreted lipase from *Frankia* ACN14a; ZP_01906526, putative secreted lipase from *Plesiocystis pacifica* SIR-1; YP_002863158, carboxylic ester hydrolase from *Clostridium botulinum* Ba4 str. 657; YP_893735, carboxylic ester hydrolase from *Bacillus thuringiensis* str. Al Hakam; YP_001073968, hypothetical protein Mjls_5714 from *Mycobacterium* sp. JLS; YP_642492, hypothetical protein Mmcs_5335 from *Mycobacterium* sp. MCS; YP_890989, hypothetical protein MSMEG_6781 from *Mycobacterium smegmatis* str. MC2 155; YP_951301, hypothetical protein Mvan_0447 from *Mycobacterium vanbaalenii* PYR-1.

Exploratory Hainich, we propose to designate the putative new family EstGH.

Putative secreted lipolytic enzymes

A potential signal peptide was predicted at the N terminus for six of the 37 deduced amino acid sequences (Est01, Est03, Est13, Est19, Est20, and Est25) using the SIGNALP 3.0 server (Bendtsen *et al.*, 2004). The number of amino acid residues of the predicted signal peptides ranged from 19 (Est25) to 32 (Est03). In addition, the amino acid sequences of all putative signal peptides showed the typical orientation of signal peptides with three distinct parts (N, H, and C domains) (Pugsley, 1993). This observation suggested that these lipolytic enzymes are secreted and function outside of the cell. Three of the putative signal peptide containing lipolytic enzymes, Est03, Est13, and Est25, belong to family V. To our knowledge, no family V lipases/esterases containing putative signal peptides have been identified in previous metagenomic studies, but signal peptide-containing family V members derived from genome sequences of individual microorganisms such as *Psychrobacter immobilis* (CAA47949) or *Moraxella* sp. (CAA37863) are known. The three remaining signal peptide-containing lipolytic enzymes grouped into families EstGS (Est01), I (Est19), and VIII (Est20). Signal peptide-harboring enzymes of the latter two families were also observed in other metagenomic studies (Meilleur *et al.*, 2009; Rashamuse *et al.*, 2009). No signal peptides were predicted for the 17 members of family IV recovered in this study. Thus, family IV lipolytic enzymes seem to play a minor role in degrading extracellular lipids in the analyzed grassland and forest soils.

Hydrolysis of triacylglycerides and *p*-nitrophenyl esters varying in chain length

The ability to hydrolyze different triacylglycerides and *p*-nitrophenyl esters was used to determine the chain-length selectivity of the 37 unique lipolytic clones and the corresponding lipolytic gene products. Each unique lipolytic clone was plated on LB agar emulsified with tributyrin (C_4), tricaproin (C_6), tricaprylin (C_8), tricaprin (C_{10}),

trilaurin (C_{12}), trimyristin (C_{14}), or tripalmitin (C_{16}). Clones with hydrolytic activity were identified after 7 days of incubation at 37 °C by halo formation. All clones showed hydrolysis activity toward the screening substrate tributyrin (Table 3), but only *E. coli* strains carrying pLE19 exhibited hydrolysis activity toward long-chain acylglycerides such as tricaprin and trilaurin (Table 3). This result supports the sequence-based classification of Est19 into the family of true lipases (family I), as true lipases show activity toward water-insoluble long-chain triacylglycerides (Arpigny & Jaeger, 1999). The other *E. coli* clone carrying a gene encoding a putative true lipase (Est17) hydrolyzed tributyrin, but none of the other tested triacylglycerides (Table 3). Besides tributyrin, tricaproin was hydrolyzed by the majority of the *E. coli* clones (21 clones) (Table 3). Thus, the substrate preference for substrates containing short-chain fatty acids ($\leq$ C10) indicated that almost all enzymes were esterases.

Hydrolysis of different *p*-nitrophenyl esters was used to further analyze the substrate specificity of the 37 unique lipolytic clones using cell culture supernatant directly. The enzyme activities determined were in the range of those previously reported by Lee *et al.* (2004), who used a similar assay to initially characterize soil-derived lipolytic enzymes. All clones, except *E. coli*/pLE12, exhibited higher activity toward *p*-nitrophenyl butyrate (C_4) than the negative control strain harboring the cloning vector (Table 3). Furthermore, the analyzed clones, except *E. coli*/pLE01, showed the highest activity toward *p*-nitrophenyl butyrate (C_4) compared with the other tested *p*-nitrophenyl esters *p*-nitrophenyl caproate (C_6), *p*-nitrophenyl caprylate (C_8), *p*-nitrophenyl caprate (C_{10}), *p*-nitrophenyl laurate (C_{12}), and *p*-nitrophenyl palmitate (C_{16}). *Escherichia coli* strains carrying pLE01–pLE07, pLE10, pLE11, pLE23, pLE24, and pLE30–pLE37 also showed activity toward *p*-nitrophenyl caproate ($\geq 0.15\,U\,mL^{-1}$) and *E. coli* carrying pLE01, pLE03, pLE10, pLE11, and pLE32 toward *p*-nitrophenyl caprylate ($\geq 0.15\,U\,mL^{-1}$) (Table 3). These results supports those derived from the plate assays using triacylglycerides in which also a preference for substrates containing short-chain fatty acids was obvious. Taking only the activities with *p*-nitrophenyl butyrate into account, *E. coli* clones harboring genes encoding family IV lipolytic enzymes showed higher activity than the clones carrying genes encoding members of other families. In summary, as expected from the sequence analysis, the determination of the substrate specificity revealed that almost all genes encode esterases. In the case of both putative lipase-encoding genes (*est17* and *est19*), further biochemical characterization is required to verify that both enzymes are true lipases. In addition, although the activity assay did not provide quantitative data to compare activity between enzymes, it is suitable to determine substrate specificity and chain-length selectivity of a large numbers of clones, which are usually recovered during metagenomic screens.

Table 3. Determination of substrate specificity and chain-length selectivity of *Escherichia coli* cells harboring pLE01 to pLE08 and pLE10 to pLE38

	Enzyme activity ($U\,mL^{-1}$)				
Plasmid	C_4	C_6	C_8	C_{10}	C_{12}
Cloning vector	0.018	0.001	0.004	ND	ND
pLE01	0.034*	0.037*	0.018	0.001	0.004
pLE02	0.169*	0.030*	0.010	0.004	0.001
pLE03	0.119*	0.039*	0.021	0.009	0.001
pLE04	0.141*	0.024*	0.011	0.005	0.002
pLE05	0.120*	0.022*	0.007	0.007	0.004
pLE06	0.151*	0.029*	0.008	0.005	0.005
pLE07	0.116*	0.017*	0.008	ND	0.004
pLE08	0.029*	0.005	0.006	0.001	0.003
pLE10	0.204*	0.083*	0.033	ND	0.004
pLE11	0.736*	0.352	0.036	0.007	0.002
pLE12	0.018*	0.003	0.004	ND	ND
pLE13	0.024*	0.001	0.005	0.003	0.003
pLE14	0.021*	ND	0.003	0.001	0.002
pLE15	0.023*	ND	0.002	0.001	ND
pLE16	0.031*	0.002	0.005	0.005	0.001
pLE17	0.021*	ND	0.003	ND	ND
pLE18	0.028*	ND	0.004	0.001	0.001
pLE19	0.063*	0.011*	0.014*	0.006*	0.005*
pLE20	0.047*	0.004*	0.003	0.001	0.003
pLE21	0.019*	0.003	0.006	ND	ND
pLE22	0.018*	0.006*	0.008	0.006	0.003
pLE23	0.057*	0.021*	0.008	0.003	0.003
pLE24	0.080*	0.028	0.003	0.003	0.003
pLE25	0.022*	0.004*	0.003	0.003	ND
pLE26	0.034*	0.006	0.008	0.003	0.003
pLE27	0.063*	0.011	0.007	0.003	0.002
pLE28	0.037*	0.006*	0.006	0.001	0.001
pLE29	0.098*	ND	0.001	0.004	ND
pLE30	0.835*	0.289*	0.006	0.037	0.007
pLE31	0.217*	0.055*	0.003	0.004	ND
pLE32	0.110*	0.016	0.017	0.003	0.003
pLE33	0.220*	0.056*	0.006	0.008	ND
pLE34	0.238*	0.031*	0.007	0.005	0.003
pLE35	0.371*	0.097*	0.009	ND	0.002
pLE36	0.142*	0.016*	ND	0.001	ND
pLE37	0.150*	0.031*	0.005	0.002	0.002
pLE38	0.115*	ND	0.002	ND	0.001

For this purpose, a plate assay using triacylglycerides and an activity assay using *p*-nitrophenyl (pNP) esters of fatty acids as substrates were used. Hydrolysis of triacylglycerides was identified by the formation of halos on agar plates. The following pNP esters of fatty acids and triacylglycerides were used as substrates: C_4, pNP-butyrate and tributyrin; C_6, pNP-caproate and tricaproin; C_8, pNP-caprylate and tricaprylin; C_{10}, pNP-caprate and tricaprin; C_{12}, pNP-laurate and trilaurin. Trimyristin (C_{14}), pNP-palmitate (C_{16}), and tripalmitin (C_{16}) were also tested, but resulted in no activity or halo formation.

*Activity toward triacylglycerides.

ND, no activity detected.

Conclusions

A total of 37 clones conferring lipolytic activity were identified by function-driven screening of soil-derived metagenomic libraries. All of the corresponding 37 lipolytic enzymes were new members of known or putatively new lipase/esterase families, and most of the enzymes were assigned to family IV. Amino acid sequence analysis and substrate specificity showed that mainly esterases that hydrolyze esters containing short-chain fatty acids were identified.

Despite the inherent limitations and biases of cloning and activity-based screens, a diverse set of genes conferring the targeted reaction was recovered using the metagenomic approach. The novelty of the lipolytic enzymes encountered arises from the largely untapped enormous genetic diversity of uncultured soil microorganisms. This study demonstrated that soils are an important source of novel lipolytic enzymes. In addition, the results presented here showed that soil-based metagenomics and the use of a simple activity-based screening system is a method for the isolation of a large number of diverse genes conferring the targeted reaction. The number of genes and gene products performing this reaction can be extended by increasing the number of screened clones. In addition, as significant differences exist in expression modes between different taxonomic groups of prokaryotes, different hosts for the constructed libraries can be used to further expand the diversity of lipolytic and other enzymes recovered during functional screening. In this way, gene banks consisting of several hundreds of genes conferring lipolytic activity or other activities can be prepared rapidly. These gene banks or the corresponding clones can serve as the starting material for the development of novel processes and products.

Acknowledgements

The work has been funded by the DFG Priority Program 1374 'Infrastructure-Biodiversity-Exploratories' (DA 374/4-1). Field work permits were given by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG). We thank the Biodiversity Exploratories Office (BEO) for support and project administration.

Authors' contribution

H.N. and C.W. contributed equally to this work.

References

Arpigny J & Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* **343**: 177–183.

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA & Struhl K (1987) *Current Protocols in Molecular Biology*. John Wiley & Sons Inc., New York, NY.

Bayer S, Kunert A, Ballschmiter M & Greiner-Stoeffele T (2010) Indication for a new lipolytic enzyme family: isolation and characterization of two esterases from a metagenomic library. *J Mol Microb Biotech* **18**: 181–187.

Bendtsen JD, Nielsen H, von Heijne G & Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**: 783–795.

Bornscheuer UT (2002) Microbial carboxyl esterases: classification, properties and application in biocatalysis. *FEMS Microbiol Rev* **26**: 73–81.

Chu X, He H, Guo C & Sun B (2008) Identification of two novel esterases from a marine metagenomic library derived from South China Sea. *Appl Microbiol Biot* **80**: 615–625.

Cieśliński H, Białkowskaa A, Tkaczuk K, Długołecka A, Kur J & Turkiewicz M (2009) Identification and molecular modeling of a novel lipase from an Antarctic soil metagenomic library. *Pol J Microbiol* **58**: 199–204.

Couto G, Glogauer A, Faoro H, Chubatsu L, Souza E & Pedrosa F (2010) Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. *Genet Mol Res* **9**: 514–523.

Daniel R (2005) The metagenomics of soil. *Nat Rev Microbiol* **3**: 470–478.

Elend C, Schmeisser C, Leggewie C, Babiak P, Carballeira JD, Steele HL, Reymond J-L, Jaeger KE & Streit WR (2006) Isolation and biochemical characterization of two novel metagenome-derived esterases. *Appl Environ Microb* **72**: 3637–3645.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Martins dos Santos VAP, Yakimov MM, Timmis KN & Golyshin PN (2005) Microbial enzymes mined from the Urania deep-sea hypersaline anoxic basin. *Chem Biol* **12**: 895–904.

Fischer M, Bossdorf O, Gockel S *et al.* (2010) Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. *Basic Appl Ecol* **11**: 473–485.

Hall A (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41**: 95–98.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol R* **68**: 669–685.

Hantsis-Zacharov E & Halpern M (2007) Culturable psychrotrophic bacterial communities in raw milk and their proteolytic and lipolytic traits. *Appl Environ Microb* **73**: 7162–7168.

Hårdeman F & Sjöling S (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiol Ecol* **59**: 524–534.

Hausmann S & Jaeger KE (2010) Lipolytic enzymes from bacteria. *Handbook of Hydrocarbon and Lipid Microbiology, Vol. 2* (Timmis KN, ed), pp. 1099–1126. Springer, Berlin.

Heath C, Hu X, Cary S & Cowan D (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. *Appl Environ Microb* **75**: 4657–4659.

Henne A, Schmitz R, 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 Microb* **66**: 3113–3116.

Hong KS, Lim HK, Chung EJ, Park EJ, Lee MH, Kim JC, Choi GJ, Cho KY & Lee SW (2007) Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes. *J Microbiol Biotechn* **17**: 1655–1660.

Hotta Y, Ezaki S, Atomi H & Imanaka T (2002) Extremely stable and versatile carboxylesterase from a hyperthermophilic archaeon. *Appl Environ Microb* **68**: 3925–3931.

Hu Y, Fu C, Huang Y *et al.* (2010) Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. *FEMS Microbiol Ecol* **72**: 228–237.

Jaeger KE & Eggert T (2002) Lipases for biotechnology. *Curr Opin Biotech* **13**: 390–397.

Jeon JH, Kim JT, Kang SG, Lee JH & Kim SG (2008) Characterization and its potential application of two esterases derived from the arctic sediment metagenome. *Mar Biotechnol* **11**: 307–316.

Jeon JH, Kim JT, Kim YJ, Kim HK, Lee HS, Kang SG, Kim SJ & Lee JH (2009) Cloning and characterization of a new cold-active lipase from a deep-sea sediment metagenome. *Appl Microbiol Biot* **81**: 865–874.

Kim BS, Kim SY, Park J, Park W, Hwang KY, Yoon YJ, Oh WK, Kim BY & Ahn JS (2007) Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library. *J Appl Microbiol* **102**: 1392–1400.

Kim E, Oh K, Lee M, Kang C, Oh T & Yoon J (2009) Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microb* **75**: 257–260.

Kim J, Seo M, Cho E, Lee S, Kim S, Cheigh C & Pyun Y (2005) Screening and characterization of an esterase from a metagenomic library. *J Microbiol Biotechn* **15**: 1067–1072.

Lämmle K, Zipper H, Breuer M, Hauer B, Buta C, Brunner H & Rupp S (2007) Identification of novel enzymes with different hydrolytic activities by metagenome expression cloning. *J Biotechnol* **127**: 575–592.

Lee M, Lee C, Oh T, Song J & Yoon J (2006) Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. *Appl Environ Microb* **72**: 7406–7409.

Lee S, Won K, Lim H, Kim J, Choi G & Cho K (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biot* **65**: 720–726.

Marchler-Bauer A, Anderson JB, Derbyshire MK *et al.* (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res* **35**: D237–D240.

Meilleur C, Hupé J, Juteau P & Shareck F (2009) Isolation and characterization of a new alkali-thermostable lipase cloned from a metagenomic library. *J Ind Microbiol Biot* **36**: 853–861.

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M & Daniel R (2011) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS One* **6**: e17000.

Pugsley AP (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**: 50–108.

Ranjan R, Grover A, Kapardar R & Sharma R (2005) Isolation of novel lipolytic genes from uncultured bacteria of pond water. *Biochem Bioph Res Co* **335**: 57–65.

Rashamuse K, Magomani V, Ronneburg T & Brady D (2009) A novel family VIII carboxylesterase derived from a leachate metagenome library exhibits promiscuous beta-lactamase activity on nitrocefin. *Appl Microbiol Biot* **83**: 491–500.

Roh C & Villate F (2008) Isolation of a low-temperature adapted lipolytic enzyme from uncultivated micro-organism. *J Appl Microbiol* **105**: 116–123.

Rondon MR, August PR, Bettermann AD *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microb* **66**: 2541–2547.

Rosselló-Mora R & Amann R (2001) The species concept for prokaryotes. *FEMS Microbiol Rev* **25**: 39–67.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA & Barrel B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Seo HJ, Bae SS, Yang SH, Lee J-H & Kim S-J (2005) *Photobacterium aplysiae* sp. nov., a lipolytic marine bacterium isolated from eggs of the sea hare *Aplysia kurodai*. *Int J Syst Evol Micr* **55**: 2293–2296.

Simon C & Daniel R (2009) Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biot* **85**: 265–276.

Simon C & Daniel R (2010) Construction of small-insert and large-insert metagenomic libraries. *Methods Mol Biol* **668**: 39–50.

Steele HL, Jaeger KE, Daniel R & Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. *J Mol Microb Biotech* **16**: 25–37.

Strohl WR (2000) The role of natural products in modern drug discovery. *Drug Discov Today* **5**: 39–41.

Tamura K, Dudley J, Nei M & Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Thompson J, Higgins D & Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.

Tirawongsaroj P, Sriprang R, Harnpicharnchai P, Thongaram T, Champreda V, Tanapongpipat S, Pootanakit K & Eurwilaichitr L (2008) Novel thermophilic and thermostable lipolytic enzymes from a Thailand hot spring metagenomic library. *J Biotechnol* **133**: 42–49.

Wagner U, Petersen E, Schwab H & Kratky C (2002) EstB from *Burkholderia gladioli*: a novel esterase with a beta-lactamase fold reveals steric factors to discriminate between esterolytic and beta-lactam cleaving activity. *Protein Sci* **11**: 467–478.

Waschkowitz T, Rockstroh S & Daniel R (2009) Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. *Appl Environ Microb* **75**: 2506–2516.

Wei P, Bai L, Song W & Hao G (2009) Characterization of two soil metagenome-derived lipases with high specificity for *p*-nitrophenyl palmitate. *Arch Microbiol* **191**: 233–240.

Will C, Thürmer A, Wollherr A, Nacke H, Herold N, Schrumpf M, Gutknecht J, Wubet T, Buscot F & Daniel R (2010) Horizon-specific bacterial community composition of German grassland soils as revealed by pyrosequencing-based analysis of 16S rRNA genes. *Appl Environ Microb* **76**: 6751–6759.

Wu C & Sun B (2009) Identification of novel esterase from metagenomic library of Yangtze river. *J Microbiol Biotechn* **19**: 187–193.

Ye J, McGinnis S & Madden TL (2006) BLAST: improvements for better sequence analysis. *Nucleic Acids Res* **34**: W6–W9.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. DNA yields, and physical and geochemical characteristics of the analyzed grassland and forest soil samples.
Table S2. Insert sizes of plasmids pLE01–pLE08 and pLE10–pLE38.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

3.1 Supplemental information for chapter B3

Contents

Table S1. DNA yields, and physical and geochemical characteristics of the analyzed grassland and forest soil samples.

Sample	DNA yield (µg/g soil)	Soil group	pH	OC[a] (g kg^{-1})	Total N (g kg^{-1})	C:N ratio	Gravimetric water content (%)
SEG2	54.3	Histosol	7.42	120.8	12.3	9.8	84.80
SEG6	47.8	Histosol	5.22	284.1	23.9	11.9	162.90
SEG9	53.6	Histosol	6.23	229.7	18.7	12.3	114.80
SEW2	35.6	Arenosol	3.46	17.0	1.0	16.7	15.70
SEW5	36.9	Arenosol	3.05	29.6	1.6	18.3	25.90
SEW8	22.4	Abeluvisol	3.09	29.2	1.8	16.1	29.30
HEG1	26.4	Cambisol	6.63	66.2	6.2	10.6	38.50
HEG9	29.6	Stagnosol	6.62	48.3	4.1	11.7	30.90
HEG1[b]	8.6	Cambisol	7.10	17.8	2.0	8.9	23.20
HEG9[b]	6.9	Stagnosol	7.35	5.0	0.6	7.9	19.90
HEW2	17.2	Luvisol	4.45	50.6	3.1	16.2	49.11
HEW5	36.5	Leptosol	4.78	61.8	4.9	12.5	77.90
HEW9	31.6	Luvisol	4.09	48.1	3.4	14.0	64.83
AEG2	37.6	Leptosol	6.92	72.3	7.2	10.1	59.60
AEW1	56.3	Cambisol	3.30	64.6	4.0	16.3	62.80
AEW4	55.9	Cambisol	6.38	78.5	6.0	13.1	75.10
AEW5	49.9	Cambisol	4.52	57.5	4.5	12.9	70.40
AEW9	37.9	Cambisol	6.37	60.0	4.5	13.4	54.90

[a], Organic carbon content

[b], Libraries constructed from soil derived from B horizon

Table S2. Insert sizes of plasmids pLE01 to pLE08 and pLE10 to pLE38. For plasmids resulting from subcloning of fosmids (pLE01, pLE04, pLE28, pLE29, pLE30, pLE33, pLE34, pLE35, and pLE37), the insert sizes of the corresponding fosmids are also depicted.

Plasmid	Insert size (bp)	Insert size of corresponding fosmid (bp)
pLE01	2,300	19,800
pLE02	3,970	Not applicable
pLE03	11,172	Not applicable
pLE04	2,290	36,000
pLE05	2,658	Not applicable
pLE06	3,608	Not applicable
pLE07	2,752	Not applicable
pLE08	6,357	Not applicable
pLE10	1,629	Not applicable
pLE11	1,245	Not applicable
pLE12	4,202	Not applicable
pLE13	3,332	Not applicable
pLE14	4,589	Not applicable
pLE15	2,439	Not applicable
pLE16	8,591	Not applicable
pLE17	6,310	Not applicable
pLE18	1,107	Not applicable
pLE19	4,995	Not applicable
pLE20	8,411	Not applicable
pLE21	7,637	Not applicable
pLE22	3,087	Not applicable
pLE23	2,568	Not applicable
pLE24	3,264	Not applicable
pLE25	4,951	Not applicable
pLE26	2,804	Not applicable
pLE27	2,176	Not applicable
pLE28	2,764	25,000
pLE29	3,568	31,000
pLE30	1,708	30,400
pLE31	7,606	Not applicable
pLE32	3,381	Not applicable
pLE33	1,583	21,500
pLE34	2,448	36,500
pLE35	1,945	34,800
pLE36	2,663	Not applicable
pLE37	1,511	28,000
pLE38	2,716	Not applicable

4

Schlaglicht Biodiversität: Charakterisierung und Nutzung der bakteriellen Diversität in Bodenmetagenomen

Christiane Will[1], Heiko Nacke[1], Andrea Thürmer[2], and Rolf Daniel[1,2]

GenomXpress (2010), Vol. 1.10, p. 9-11

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany

Author contributions to the work:

Performed the experiments: HN, CW, AT.

Analyzed data: HN, CW.

Wrote the paper: CW, HN, RD.

Conceived and designed the experiments: HN, CW, RD.

Schlaglicht Biodiversität
Charakterisierung und Nutzung der bakteriellen Diversität in Bodenmetagenomen

Seit 2006 fördert die Deutsche Forschungsgemeinschaft in einem Schwerpunktprogramm das Verbundprojekt „Exploratorien zur funktionellen Biodiversitätsforschung", kurz Biodiversitäts-Exploratorien (www.biodiversity-exploratories.de). Drei Exploratorien dienen als offene Forschungsplattform für Wissenschaftler aus ganz Deutschland: das Biosphärenreservat Schorfheide-Chorin in Brandenburg, der Nationalpark Hainich und seine Umgebung in Thüringen und das Biosphärengebiet Schwäbische Alb in Baden-Württemberg. Untersucht werden die Beziehungen zwischen der Biodiversität verschiedener Taxa und Ebenen, die Rolle von Landnutzung und Management für die Biodiversität und die Rolle der Biodiversität für Ökosystemprozesse. In den Exploratorien waren und sind über 330 Mitarbeiter aus 61 Arbeitsgruppen von ingesamt 33 Forschungseinrichtungen tätig. Im Folgenden werden erste Ergebnisse aus dem Teilprojekt Boden (Biotik) mit Schwerpunkt auf der bodenmikrobiellen Ökologie dargestellt. Geplant sind der Aufbau von metagenomischen Banken, das Screening für Targets (Organismen und Funktionen) zur Idenfizierung von *key players* und parallel die bodenökologische Charakterisierung.

Christiane Will, Heiko Nacke, Andrea Thürmer und Rolf Daniel

Die mikrobielle Diversität in Böden ist sehr viel größer als die in anderen Habitaten und übertrifft um Größenordnungen die Diversität von Pflanzen und Tieren. Bakterien bilden die häufigste Gruppe der Mikroorganismen in Böden. Es wird geschätzt, dass 2.000 bis 18.000 bakterielle Arten und bis zu 109 bakterielle Zellen ein Gramm Boden besiedeln (Daniel, 2005). Bodenbakterien sind unverzichtbar für die Funktionalität von geochemischen Stoffkreisläufen, stabilisieren die Bodenstruktur und verbessern die Speicherung von Wasser im Boden. Durch die wechselnde physische, chemische und biologische Beschaffenheit des Bodens kann die Zusammensetzung von mikrobiellen Gemeinschaften mit zunehmender Bodentiefe und in unterschiedlichen Böden variieren.

Die Metagenomik ermöglicht trotz der enormen Komplexität von Bodenhabitaten die Gewinnung tiefgehender Erkenntnisse über Struktur und Funktion von mikrobiellen Gemeinschaften im Boden. Zur Erforschung der Diversität in Bodenproben werden direkte Kultivierungsansätze und indirekte molekulare Verfahren verwendet. Da aber nur ca. 1% der Bodenbakterien mit Standardmethoden kultivierbar sind, wurden zur Erschließung der Komplexität der mikrobiellen Bodengemeinschaft indirekte molekulare Verfahren entwickelt, die auf der direkten Isolierung von Nukleinsäuren (Metagenomen) aus Bodenproben basieren. Das Metagenom umfasst die Gesamtheit der mikrobiellen genetischen Information eines Standortes. Im Rahmen der hier vorliegenden Studie wurden Bodenproben der deutschen Biodiversitäts-Exploratorien Hainich-Dün, Schorfheide-Chorin und Schwäbische Alb analysiert.

Im Verlauf der Metagenomanalyse wurde die phylogenetische Diversität der Bakterien im A-Horizont (Oberboden) und im B-Horizont (Unterboden) von Grünlandbodenproben aus dem Hainich-Dün Exploratorium untersucht. Dazu wurde die Amplikon-Sequenzierung, eine Variante der Pyrosequenzierung, etabliert und angewendet. Zusätzlich wurde eine analoge phylogenetische Analyse mit A-Horizont Wald- und Grünlandbodenproben aus der Schwäbischen Alb durchgeführt, die unterschiedliche Landnutzungstypen repräsentieren. Neben der phylogenetischen Analyse wurde das genetische Potential der Standorte erschlossen. In diesem Zusammenhang wurden Metagenom-Bibliotheken konstruiert und auf die Existenz von lipolytischen Enzymen durchmustert.

Phylogenetische Analyse von bakteriellen Gemeinschaften im Boden

Wald- und Grünlandflächen mit unterschiedlichen Landnutzungsintensitäten wurden für die Entnahme von Bodenproben ausgewählt. Im Wald umfasste dies vorwiegend Buchenforste aber auch Flächen mit Kiefern und Fichten. Die Nutzungsintensität reichte von Altersklassenwald über Plenterwald (d.h. ein sich stetig verjüngender Dauerwald, in dem Bäume aller Altersklassen kleinstflächig bis einzelstammweise vermischt sind) bis hin zu Naturwald. Im Grünland wurden sowohl gedüngte als auch ungedüngte Flächen herangezogen, welche als Mähweiden und Wie-

Abb. 1: Bohrkerne von Waldproben (oben) und Grünlandproben (unten). Der Durchmesser der Bohrkerne beträgt 8,3 cm und es wurden Ober- und Unterboden beprobt.

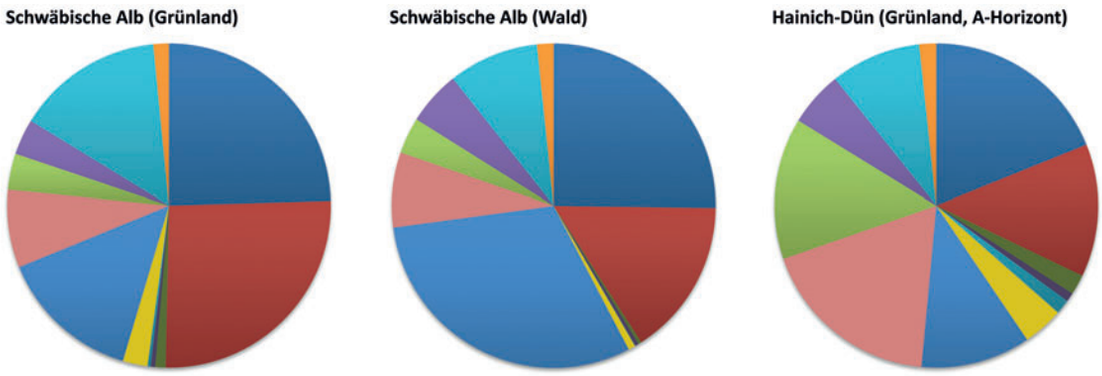


Abb. 2: Relative Verteilung von bakteriellen Phyla im Wald und Grünland sowie im A- und B- Horizont.

sen oder als Schafs-, Pferde- bzw. Rinderweiden dienten.

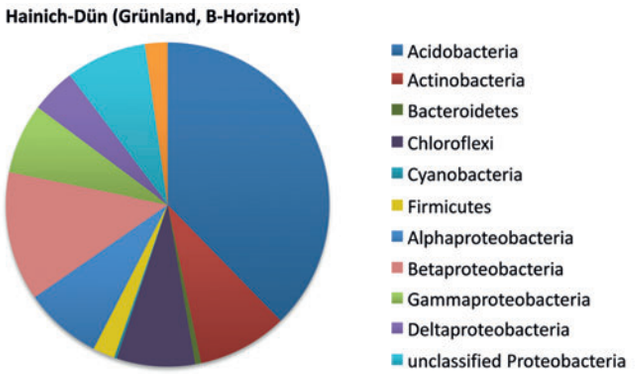
Auf jeder Entnahmestelle wurden innerhalb einer 20 x 20 m Fläche fünf Bohrkerne entnommen (Abbildung 1). Die Bohrkerne wurden in Bodenhorizonte getrennt und zu einem Feldlabor transportiert. Steine und Grobwurzeln wurden aus den Bodenproben entfernt und größere Bodenpartikel homogenisiert. Die Proben der einzelnen Entnahmestellen wurden horizontweise zu Mischproben vereint. Aus je 10 g Mischprobe des A-Horizonts wurde mikrobielle DNA isoliert. Zusätzlich wurde mikrobielle DNA aus dem B-Horizont aus den Bodenproben des Hainich-Dün-Exploratoriums isoliert. Für die phylogenetische Analyse der Bakteriengemeinschaften wurde aus der gewonnenen DNA der Bodenproben aus der Schwäbischen Alb sowie dem Grünland des Hainich-Dün eine Teilregion (V2-V3 Region) des 16S rRNA-Gens durch PCR amplifiziert und anschließend sequenziert und analysiert. Durch Vergleiche mit einer 16S rRNA-Sequenzdatenbank wurden in den 36 Proben insgesamt 1.348.962 bakterielle Sequenzen mit einer durchschnittlichen Leselänge von 259 bp identifiziert. Die dominanten bakteriellen Phyla umfassen die *Proteobacteria, Acidobacteria* und *Actinobacteria*. Es konnten signifikante Unterschiede in der Verteilung einzelner Phyla zwischen den Proben aus dem Wald und dem Grünland festgestellt werden sowie zwischen den Proben von A- und B-Horizont (Abbildung 2).

Beim Vergleich der Wald- und Grünlandproben aus der Schwäbischen Alb ist auffällig, dass *Actinobacteria* und *Firmicutes* im Grünland einen höhere Abundanz aufweisen als im Wald, wogegen die *Alphaproteobacteria* im Wald im Vergleich zum Grünland eine höhere Abundanz aufweisen. Außerdem zeigte sich, dass die relative Abundanz einiger bakterieller Phyla, Ordnungen, Familien und Gattungen innerhalb der Wald- und Grünlandproben mit dem pH-Wert des Bodens korrelierte. Im Wald nahm die relative Abundanz der *Alphaproteobacteria*, *Caulobacterales* und *Acetobacteraceae* mit zunehmendem pH-Wert ab, wogegen die der *Bacteroidetes* und *Betaproteobacteria* mit zunehmendem pH-Wert anstieg. Im Grünland erhöhte sich die relative Abundanz der *Acidobacteria* und *Acetobacteraceae* mit abnehmendem pH-Wert.

Beim Vergleich von Proben aus dem Oberboden mit denen aus dem Unterboden des Hainich-Dün Grünlands zeigt sich, dass eine größere Umverteilung der relativen Häufigkeiten stattfindet. Auffällig ist, dass die relative Abundanz der *Acidobacteria* und die der *Chloroflexi* von oben nach unten in der Bodensäule stark zunimmt. Dafür reduziert sich der Anteil der *Actinobacteria*, *Alpha-*, *Beta-* und *Gammaproteobacteria*. Insgesamt gesehen nimmt sowohl die Biomasse als auch die Anzahl der auftretenden Phyla oder Spezies im Bodenprofil von oben nach unten ab.

Lipase/ Esterase-Familie	Anzahl der gefundenen Vertreter	Proteinsequenz-ähnlichkeit zu bekannten lipolytischen Enzymen (%)	Bekanntes Protein mit der höchsten Ähnlichkeit (Zugangsnummer in GenBank-Datenbank)	Bekanntes Protein mit der geringsten Ähnlichkeit (Zugangsnummer in GenBank-Datenbank)
I	2	44 – 62	Lipase, Klasse 2 aus *Mycobacterium vanbaalenii* PYR-1 (YP_953514)	Vorhergesagte Acetyltransferase/ Hydrolase aus *Tsukamurella pauro-metabola* DSM 20162 (ZP_04028715)
IV	17	48 – 87	Lipase/Esterase aus einem unkultivierten Bakterium (AAS77236)	Lipase/Esterase aus einem unkultivierten Bakterium (AAS77247)
V	10	32 – 65	Alpha/Beta Hydrolase aus *Methylobacterium* sp. 446 (YP_001773617)	Vorhergesagte Hydrolase oder Acyltransferase aus *Lentisphaera araneosa* HTCC2155 (ZP_01875730)
VI	1	57	Phospholipase/Carboxylesterase aus *Nitrosospira multiformis* ATCC 25196 (YP_413093)	Phospholipase/Carboxylesterase aus *Nitrosospira multiformis* ATCC 25196 (YP_413093)
VIII	6	39 – 90	Beta-Lactamase aus *Caulobacter* sp. K31 (YP_001682441)	Beta-Lactamase aus *Burkholderia* sp. H160 (ZP_03264354)

Tab. 1. Einordnung der identifizierten Genprodukte in bekannte Lipase/Esterase-Familien und Sequenzähnlichkeiten.



Mit dieser umfangreichen Studie über verschiedene Bodenproben aus Deutschland konnte ein tiefgehender Einblick in die bakterielle Biodiversität und die Veränderung der Verteilung im Habitat Boden gewonnen werden. Dabei zeigte sich, dass der pH-Wert und der Bodenhorizont einen Einfluss auf die Zusammensetzung von bakteriellen Gemeinschaften haben.

Das genetische Potential des Boden-Metagenoms

In Anbetracht der Tatsache, dass ein Gramm Boden bis 2 x 10^9 prokaryotische Zellen beinhaltet, die zum überwiegenden Teil im Labor nicht kultivierbar sind, bietet dieses Habitat ein nahezu unerschöpfliches Potential an Genen für neuartige Biokatalysatoren und andere Biomoleküle.

Eine industriell bedeutende Rolle spielen dabei Lipasen (EC 3.1.1.3) und Esterasen (EC 3.1.1.1). Diese finden biotechnologische Anwendung als Katalysator bei der Synthese von Polymeren und bei der Herstellung von Biodiesel. Des Weiteren können Esterasen bei der Produktion von chemischen Reinstoffen wie zum Beispiel Arzneimitteln, Herbiziden, Kosmetika, Aroma- und Duftstoffen eingesetzt werden.

Um die Bodenproben auf neuartige Lipasen und Esterasen zu durchmustern, wurden metagenomische Genbibliotheken in Plasmiden und Fosmiden aus der isolierten DNA angelegt. Die insgesamt 32 Plasmidbanken umfassen 29 Gb klonierte DNA, die 19 Fosmidbanken 64 Gb. Davon wurden 11 Gb DNA (Plasmidbanken) bzw. 1,17 Gb DNA (Fosmidbanken) auf das Vorhandensein von Genen, die für lipolytische Aktivität kodieren, untersucht. Das Screening erfolgte mittels Plattentest, indem das Triglycerid Tributyrin als Indikatorsubstanz fungiert. Auf diesen Testplatten wurden die metagenomischen Banken in *E. coli* als Wirt ausplattiert. Klone, welche in der Lage sind, die Esterbindungen des Testsubstrats in Glycerin und Butyrat zu spalten, zeigen einen Aufklarungshof um die Kolonie (Abbildung 3). Insert-DNA von positiven Klonen wurde sequenziert und analysiert. Mit dieser Methode konnten 36 unterschiedliche Gene für potentielle Esterasen gefunden werden. Die korrespondierenden Klone wurden auch auf Agarplatten mit längerkettigen Triglyceriden getestet. Dabei zeigte sich, dass Trihexanoat (C6) überwiegend abgebaut werden konnte, während allerdings auf Agarplatten mit Trioctanoat (C8) nur einmal Aktivität zu erkennen waren. Einige der Esterase-Gene wurden in Expressionsvektoren subkloniert und exprimiert. Auch hier zeigte sich, dass Substrate mit kurzkettigen Fettsäuren in der Regel bevorzugt wurden. Es konnte für alle Substrate eine erhöhte Enzymaktivität bei steigenden Temperaturen festgestellt werden. Es wurden Esterasen gefunden die bei extremen pH-Werten (pH 3 bis 4 und pH 10 bis12) sowie über einen längeren Zeitraum bei 60 °C keinen nennenswerten Aktivitätsverlust zeigten.

Die 36 gefundenen Esterasen wurden in die acht Familien der lipolytischen Enzyme einsortiert (siehe Tabelle 1). Sie entstammen zum größten Teil den Familien IV, V und VIII, weiterhin sind auch die Familien I und VI vertreten. Familie I umfasst echte Lipasen. Diese zeigen eine höhere Aktivität gegenüber wasserunlöslichen längerkettigen Triglyceriden während Esterasen kurzkettige wasserlösliche Triglyceride bevorzugen. Die Ähnlichkeiten zu bereits bekannten Esterasen reichen von 32 bis 90 %. Die Größen der lipolytischen Genprodukte bewegen sich zwischen 111 und 556 Aminosäuren. Da Esterasen weit verbreitet sind und in allen untersuchten Bodenproben vorkamen, eignen sie sich gut, um die Diversität eines Standortes widerzuspiegeln. Einige der hier gefundenen Esterasen haben aufgrund ihrer Temperatur- und pH-Stabilität biotechnologisches Anwendungspotential.

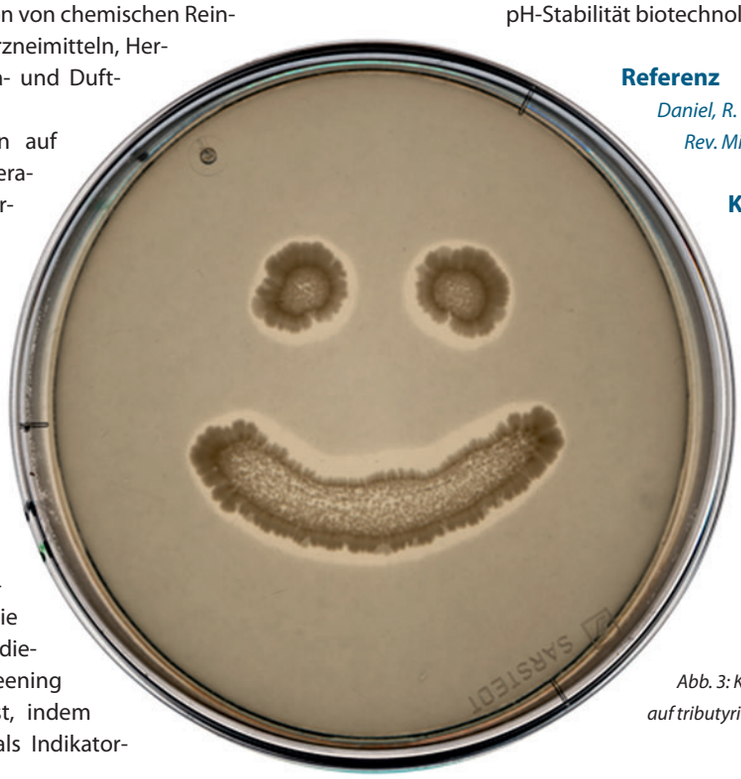


Abb. 3: Klon mit lipolytischer Aktivität auf tributyrinhaltigem Agar.

Referenz

Daniel, R. (2005) The metagenomics of soil. Nat. Rev. Microbiol. 3:470-478.

Kontakt

PD Dr. Rolf Daniel
Georg-August-Universität Göttingen
E-Mail: rdaniel@gwdg.de

5

Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes

Heiko Nacke[1], Martin Engelhaupt[1], Silja Brady[1], Christiane Fischer[1], Janine Tautzt[1], and Rolf Daniel[1,2]

Biotechnology Letters (2012), Vol. 34, p. 663-675

[1]Department of Genomic and Applied Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany, [2]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Germany

Author contributions to the work:

Performed the experiments: HN, ME, JT, SB.

Analyzed data: HN, CW.

Wrote the paper: HN, RD.

Conceived and designed the experiments: HN, RD.

Biotechnol Lett (2012) 34:663–675
DOI 10.1007/s10529-011-0830-2

ORIGINAL RESEARCH PAPER

Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes

Heiko Nacke · Martin Engelhaupt · Silja Brady · Christiane Fischer · Janine Tautzt · Rolf Daniel

Received: 9 September 2011 / Accepted: 8 December 2011 / Published online: 21 December 2011

Abstract Soil metagenomes represent an unlimited resource for the discovery of novel biocatalysts from soil microorganisms. Three large-inserts metagenomic DNA libraries were constructed from different grassland soil samples and screened for genes conferring cellulase or xylanase activity. Function-driven screening identified a novel cellulase-encoding gene (*cel01*) and two xylanase-encoding genes (*xyn01* and *xyn02*). From sequence and protein domain analyses, Cel01 (831 amino acids) belongs to glycoside hydrolase family 9 whereas Xyn01 (170 amino acids) and Xyn02 (255 amino acids) are members of glycoside hydrolase family 11. Cel01 harbors a family 9 carbohydrate-binding module, previously found only in xylanases. Both Xyn01 and Xyn02 were most active at 60°C with high activities from 4 to 10 and optimal at pH 7 (Xyn01) and pH 6 (Xyn02). The cellulase gene, *cel01*, was expressed in *E. coli* BL21 and the recombinant enzyme (91.9 kDa) was purified. Cel01 exhibited high activity with soluble cellulose substrates containing β-1,4-linkages. Activity with microcrystalline cellulose was not detected. These data, together with the analysis of the degradation profiles of carboxymethyl cellulose and barley glucan indicated that Cel01 is an endo 1,4-β-glucanase. Cel01 showed optimal activity at 50°C and pH 7 being highly active from pH range 5 to 9 and possesses remarkable halotolerance.

Keywords Soil metagenome · Cellulase · Xylanase · Metagenomic libraries · Activity-based screening

Introduction

The plant cell wall constituents cellulose and the hemicellulose xylan are the most abundant polysaccharides in nature (Beg et al. 2001; Lynd et al. 2005; Polizeli et al. 2005). The structure of cellulose is based on linear polymers of β-1,4-linked D-glucose residues. The hemicellulose xylan consists of a linear β-1,4-linked xylose backbone that can be substituted by arabinofuranosyl and/or glucuronopyranosyl side chains with the degree and nature of substitution varying between tissues and species. Xylans can be modified by additional sugar side chains, methylation, acetylation, or feruloylation (Beg et al. 2001; Kulkarni et al. 1999). Both polysaccharides have an enormous

Electronic supplementary material The online version of this article (doi:10.1007/s10529-011-0830-2) contains supplementary material, which is available to authorized users.

H. Nacke · M. Engelhaupt · S. Brady · C. Fischer · J. Tautzt · R. Daniel (✉)
Department of Genomic and Applied Microbiology, Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstr. 8, 37077 Göttingen, Germany
e-mail: rdaniel@gwdg.de

Present Address:
J. Tautzt
Biology Faculty, University of Osnabrück, Barbarastraße 11, 49069 Osnabrück, Germany

potential as renewable sources for production of biofuels (Sanchez and Cardona 2008).

Enzymes degrading (hemi)cellulose are found in microbes, plants, and the digestive tracts of animals. Three classes of cellulases, including endoglucanases (EC 3.2.1.4), exoglucanases (EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21) are involved in the degradation of cellulose (Lynd et al. 2002). Endo-1,4-β-xylanase (EC 3.2.1.8) and β-xylosidase (EC 3.2.1.37) can hydrolyze the backbone of xylan. Additionally, acetyl esterases (EC 3.1.1.6), α-D-glucuronidases (EC 3.2.1.1), and α-L-arabinofuranosidases (EC 3.2.1.55) cleave the side chains of xylan (Polizeli et al. 2005). Cellulases and hemicellulases have been recognized as very useful biocatalysts because of their wide-ranging versatility in industrial applications, including food technology, textile production, biofuel formation, and paper production (Collins et al. 2005; Lynd et al. 2002; Steele et al. 2009).

Soils are considered to be the most diverse microbial habitat on Earth with respect to species diversity and community size. One gram of soil may contain up to 10 billion microorganisms of possibly thousands of different species (Rosselló-Mora and Amann 2001). Soil microorganisms have been the major source for the isolation of novel biocatalysts and other biomolecules of industrial importance (Strohl 2000). So far, most soil-derived cellulases and xylanases and the corresponding genes have been recovered from cultured soil microorganisms such as *Cellvibrio mixtus* and *Clostridium thermocellum* (Fontes et al. 2000; Gilad et al. 2003). Since less than 1% of soil microorganisms are readily culturable, only a small fraction of soil microbial diversity is assessed by cultivation-dependent approaches. To expand the range of natural product discovery, culture-based methods have been complemented or replaced by culture-independent metagenomic approaches, which theoretically provide access to the collective nucleic acids of all indigenous microorganisms present in an environmental sample (Handelsman 2004; Simon and Daniel 2011). Functional metagenomics based on the direct isolation of DNA from environmental samples, generation of metagenomic libraries from the isolated DNA, and function-driven screening of the constructed libraries has led to the identification and characterization of a variety of novel biocatalysts (Simon and Daniel 2009), including proteases (Waschkowitz et al. 2009; Zhang et al. 2011), lipolytic enzymes (Nacke et al. 2011b; Yu et al. 2011) and (hemi)cellulolytic enzymes (Brennan et al. 2004; Duan et al. 2009; Shedova et al. 2009; Voget et al. 2006).

In this study, we report on the identification of one cellulase and two xylanases, which were derived from the soil metagenome. We constructed large-insert metagenomic libraries from three different grassland soil samples, which were collected in the German Biodiversity Exploratories Schorfheide-Chorin and Schwäbische Alb (Fischer et al. 2010). The library-containing *Escherichia coli* clones were screened for genes encoding cellulase or xylanase activity. Two novel xylanases and one cellulase were identified. Characterization of the xylanases and the purified cellulase was performed.

Materials and methods

Bacterial strains and vectors

Fosmids and plasmids used in the present study are shown in Table 1. *Escherichia coli* strain EPI300-T1^R (Epicentre Biotechnologies, Madison, WI, USA) was used as a host for the cloning of metagenomic DNA. In addition, *E. coli* strains TOP10 and BL21(DE3) (Invitrogen GmbH, Karlsruhe, Germany) were employed for subcloning and expression of the targeted genes, respectively.

Soil sampling, isolation of DNA, and construction of metagenomic DNA libraries

Soil metagenome-derived genes encoding (hemi)cellulolytic enzymes were recovered from A horizons of soil samples, which had been taken from three grassland sites of the German Biodiversity Exploratories Schorfheide-Chorin (sample SEG9) and Schwäbische Alb (samples AEG3 and AEG6). Samples were collected in April and May 2008 and sampling was performed as described by Nacke et al. (2011a). Descriptions of the soil characteristics are provided in Supplementary Table S1. Names of the metagenomic libraries refer to the designation of the samples from which the libraries were derived. To generate metagenomic libraries total microbial community DNA was isolated from 10 g soil per sample. For this purpose, the MoBio Power Max Soil DNA extraction kit

Table 1 Vectors used in this study

Type of vector	Designation	Relevant characteristics[a]	Source
Fosmid	pCC1FOS	Chl^r, T7 promoter, FI and pMB1 replicon	Epicentre
	fLC01	pCC1FOS: 16,250 bp fragment of cloned metagenomic DNA	This study
	fLX01	pCC1FOS: 36,000 bp fragment of cloned metagenomic DNA	This study
	fLX02	pCC1FOS: 30,200 bp fragment of cloned metagenomic DNA	This study
Plasmid	pCR-XL-TOPO	Kan^r, Zeo^r, *lac* promoter, pMB1 replicon	Invitrogen
	pET101/D	Ap^r, T7 promoter, pMB1 replicon	Invitrogen
	pLC01	pCR-XL-TOPO: 3,220 bp fragment of metagenomic DNA subcloned from fLC01	This study
	pLX01	pCR-XL-TOPO: 5,523 bp fragment of metagenomic DNA subcloned from fLX01	This study
	pLX02	pCR-XL-TOPO: 4,985 bp fragment of metagenomic DNA subcloned from fLX02	This study
	pCel01	pET101/D: 2,433 bp fragment containing bases 646–3,075 of *cel01* gene and a synthetic ATG start codon under the control of the T7 promoter	This study

[a] Ap^r ampicillin resistance, Chl^r chloramphenicol resistance, Kan^r kanamycin resistance, Zeo^r zeocin resistance

(MoBio Laboratories, Carlsbad, CA, USA) was used according to the instructions of the manufacturer.

The large-insert metagenomic fosmid libraries AEG3 and AEG6 were constructed by using the Copy Control Fosmid Library Production kit (Epicentre) as described by Nacke et al. (2011b). The fosmid library SEG9 has been previously generated by employing the same approach (Nacke et al. 2011b). Approx. 4,600 resulting library-containing clones per soil sample were arrayed and stored in 96-well microtiter plates. The remaining clones were collected and stored as clone pools at −80°C.

Growth conditions and activity-based screening

Escherichia coli strains were routinely grown in Luria-Bertani (LB) medium at 37°C. For activity-based screening, the arrayed library-containing *E. coli* clones were replica streaked on LB agar plates containing colored insoluble forms of hydroxyethyl cellulose (HEC_{red}) or xylan ($xylan_{green}$) as indicator substrates. The substrates were generated by employing the cross-linking reagent 1,4-butanediol diglycidyl ether and the dyes Cibacron Brilliant Red 3B-A (HEC_{red}) or Cibacron Brilliant Green T3GE ($xylan_{green}$) as described by Lee and Lee (1997) and Ten et al. (2005). In addition, to maintain the presence of recombinant fosmids and increase the copy number of the fosmids, the indicator agar contained 12.5 mg chloramphenicol l^{-1} and 0.001% arabinose, respectively. Clones showing activity with the indicator substrates were identified by the formation of clear zones (halos) after incubation for 1–14 days at 37°C under aerobic conditions.

Subcloning and sequence analysis

To subclone DNA fragments containing genes conferring (hemi)cellulolytic activity from large-insert fosmids, the recombinant fosmids from positive clones were sheared by sonication (UP200S Sonicator, Dr. Hielscher GmbH, 5 s at 30% amplitude, cycle 0.5). Subsequently, the resulting DNA fragments were separated by agarose gel electrophoresis, and fragments (2–6 kbp) were excised and extracted by using the peqGold gel extraction kit (Peqlab Biotechnologie GmbH, Erlangen, Germany). The resulting DNA fragments were ligated into pCR-XL-TOPO, and used to transform *E. coli* TOP10 as recommended by the manufacturer (Invitrogen). The resulting recombinant *E. coli* strains were screened on the corresponding indicator agar for the presence of genes conferring cellulase or xylanase activity.

The recombinant plasmids derived from positive clones were sequenced by the Göttingen Genomics Laboratory (Göttingen, Germany). The initial prediction of ORFs located on the inserts of plasmids pLC01, pLX1, and pLX2 (Table S2) was performed by using

the ORF-finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) provided by the National Center for Biotechnology Information (NCBI) and the Artemis program (Rutherford et al. 2000). The results were verified and improved manually by using criteria such as the presence of a ribosome-binding site, GC frame plot analysis, and similarity to known genes. All coding sequences were examined for similarities to protein families and domains using searches against the CDD databases (Marchler-Bauer et al. 2007). Signal peptides of proteins were predicted by using the SignalP 3.0 server (Bendtsen et al. 2004). Multiple alignments of deduced protein sequences were performed with ClustalW2, version 2.0.12 (Thompson et al. 1994) and examined with the Bioedit program (Hall 1999).

The nucleotide sequences of the recombinant plasmids harboring the xylanase genes *xyn01* and *xyn02* (pLX01 and pLX02) and the cellulase gene *cel01* (pLC01) have been submitted to GenBank under accession numbers JF799945, JF799946, and JF799947, respectively.

Cloning of the cellulase-encoding gene *cel01* into expression vector pET101/D

The cellulase gene *cel01* was amplified from plasmid pLC01 (Table 1) without the potential signal peptide sequence by PCR. The following set of primers with a synthetic site (underlined) that contained an ATG start codon and allowed directional cloning into pET101/D using the pET101/D directional TOPO expression kit (Invitrogen) was used: 5′-GCGTTCGTTGAAACGC-3′ and 5′-CACCATGCAGGAAATGCTCGCGCCC-3′. The resulting plasmid was designated pCel01 (Table 1). The PCR reaction mixture (33 µl) contained 3.3 µl tenfold reaction buffer (Finnzymes, Espoo, Finland), 800 µM of each of the four deoxynucleoside triphosphates, 3% DMSO, 1.2 µM of each of the primers, 0.5 U Phusion hot start high-fidelity DNA Polymerase (Finnzymes), and 50 ng plasmid DNA as template. The PCR reactions were initiated at 98°C (5 min), followed by 30 cycles of 98°C (45 s), a temperature gradient ranging from 60 to 68°C (45 s), 72°C (170 s) and ended with incubation at 72°C for 5 min. The PCR product was purified using the peqGold gel extraction kit (Peqlab) and cloned into pET101/D (Invitrogen) as recommended by the manufacturer. The coding region was thereby placed under the control of the IPTG-inducible T7 promoter. In addition, sequences encoding a His_6 tag and a V5 epitope provided by the vector were added to the 3′-end of the coding region.

Preparation of cell extracts

Cells from 500 ml cultures were harvested by centrifugation at 10,000×*g* and 4°C for 10 min. To prepare cell extracts of the cellulolytic strain (*E. coli* BL21/pCel01) and xylanolytic strains (*E. coli* TOP10/pLX01 and *E. coli* TOP10/pLX02) the resulting cell pellets were washed twice with LEW buffer (Macherey and Nagel, Düren, Germany) or 50 mM sodium phosphate buffer (pH 7), respectively. Subsequently, the cells were resuspended in 2 ml per g cell pellet of the same buffer. The cells were disrupted by using a French press (1.38×10^8 Pa) and the extract was cleared by centrifugation at 18,000×*g* and 4°C for 30 min. The resulting supernatant (crude cell-free extract) was used for further analyses.

Purification of cellulase

The purification of His_6-tagged cellulase from cell extracts was performed by nickel affinity chromatography using Protino-Ni-2000 prepacked columns as recommended by the manufacturer (Macherey and Nagel). Ultrafiltration employing Vivaspin concentrators (exclusion limits 100,000 and 50,000 Da; Sartorius AG, Göttingen, Germany) was used to change buffer systems and to further purify the cellulase. The purity of the resulting protein preparations was analyzed by SDS–PAGE. Detection of V5 epitope-tagged proteins by Western blot hybridization was performed as described by Waschkowitz et al. (2009).

Protein and enzyme assays

Protein was measured by the Bradford method. To analyze cellulase and xylanase activity released reducing sugars were measured as D-glucose and D-xylose equivalents, respectively (Miller 1959) and measured at 575 nm. Barley glucan (Megazyme, Bry, Ireland), carboxymethyl cellulose (Sigma–Aldrich), HEC (Sigma–Aldrich), laminarin (Sigma–Aldrich), lichenan (Megazyme), xylan from birch wood (Sigma–Aldrich), xylan from oat spelt (Sigma–Aldrich), xylan from larch wood (Sigma–Aldrich),

xylan from beech wood (Sigma–Aldrich), and microcrystalline cellulose (Serva, Heidelberg, Germany) were used as substrates. The standard assay mixture contained purified enzyme (0.4 μg) or crude cell extract and 1% barley glucan (cellulase) or 1% xylan from birch wood (xylanase) in 50 mM sodium phosphate buffer adjusted to pH 7 (final volume 0.5 ml). To determine substrate specificity barley glucan was replaced by the above-mentioned substrates at a final concentration of 1%. Enzyme activity was measured after incubation of the reaction mixture at 40°C for 20 min. All enzyme assays were performed in triplicate. One unit (U) of enzyme activity was defined as the amount of enzyme required to release 1 μmol of reducing sugar per min. Enzyme activities were determined from 20 to 70°C and pH 4–10. The pH range was determined with 50 mM sodium acetate buffer (pH 4.0–6.0), 50 mM sodium phosphate buffer (pH 6.0–8.0), 50 mM Tris/HCl buffer (pH 8.0–9.0), and 50 mM glycine/NaOH buffer (pH 9.0–10.0). The thermal stability of Cel01 was analyzed after incubation of the enzyme at 30–60°C. The halotolerance of Cel01 was determined by measuring residual activity using the standard assay after incubation of the enzyme in 3 M KCl or 4 M NaCl at 4°C for 12 h. Metal ions and chemical agents were added to the standard assay mixture to investigate their effects on enzyme activity.

The mode of action of the cellulase was determined by incubating barley glucan and CMC (each 2%) with 3 μg purified enzyme in 0.2 M Na_2HPO_4/0.1 M citric acid buffer (pH 7) at 40°C. Reaction products were separated on a silica 60 TLC plate developed in 1-propanol/ethylacetate/H_2O (6:1:3, by vol.). After spraying the plates with ethanol/sulphuric acid (9:1, v/v), sugars were visualized by heating (120°C for 20 min).

Results and discussion

Construction of metagenomic DNA libraries

Two different grassland soil samples derived from the Schwäbische Alb (AEG3 and AEG6) were used for the construction of two large-insert metagenomic DNA libraries using the fosmid pCC1FOS as vector. In addition, one previously constructed metagenomic library (SEG9) derived from a grassland soil sample of the Schorfheide-Chorin was employed for subsequent screening. In all cases, DNA was directly isolated from the soil samples without previous enrichment or extraction of microbial cells. The DNA yields ranged from 25.3 to 56.7 μg per g soil (Supplementary Table 1). These yields were in the same range as described for the isolation of DNA from other soil samples (Waschkowitz et al. 2009). The three metagenomic libraries contained 147,888 (SEG9), 121,520 (AEG3), and 95,160 (AEG6) clones. The total numbers of recovered metagenomic library-containing *E. coli* clones are higher than in most other studies in which fosmids were used as vectors for the construction of large-insert metagenomic libraries (Couto et al. 2010; Hong et al. 2007; Jeon et al. 2009; Lim et al. 2005). The quality of the three different libraries was controlled by determination of the average insert sizes and the percentage of insert-bearing *E. coli* clones. The average insert sizes were 23.3 (SEG9), 26.9 (AEG3), and 28.4 (AEG6) kbp. These values roughly corresponded to those published for other soil-derived large-insert metagenomic libraries (Donato et al. 2010; Heath et al. 2009; Nacke et al. 2011b). The libraries harbored 9.42 Gbp of cloned soil DNA. Assuming an average prokaryotic genome size of 5 Mbp (Hårdeman and Sjöling 2007), the metagenomic libraries represented approximately 1,900 prokaryotic genomes.

Activity-based screening for cellulase and xylanase genes

The screen for genes conferring cellulase or xylanase activity was based on the ability of library-bearing *E. coli* clones to form halos when grown on agar medium containing dye-labeled hydroxyethyl cellulose (HEC_{red}) or xylan ($xylan_{green}$) as indicator substrates. As function-driven screening is the only strategy that harbors the potential to discover entirely novel and functional genes from metagenomes (Simon and Daniel 2009), similar screens have been employed to identify genes encoding (hemi)cellulolytic enzymes from various environments such as soil (Kim et al. 2007, 2008), cow rumen and manure (Ferrer et al. 2005; Li et al. 2009; Zhao et al. 2010), and sludge from a biogas reactor (Jiang et al. 2010).

To perform the above-mentioned activity-based approach approximately 4,600 clones per library were arrayed and screened for the targeted activities.

Positive *E. coli* clones were collected after incubation on the indicator agar for 1–14 days at 37°C. In order to confirm that the (hemi)cellulolytic activity of the positive clones was fosmid-encoded the recombinant fosmids were isolated and used to transform *E. coli*. The resulting *E. coli* strains were screened again on indicator agar. Two different recombinant fosmids, fLX01 and fLX02, conferred a stable xylanolytic phenotype and one (fLC01) a cellulolytic phenotype. The fosmids fLX01, fLX02, and fLC01 were derived from libraries AEG3, AEG6, and SEG9, respectively (Table 1). Studies on recovery of genes encoding xylanases and cellulases from soil-derived metagenomic large-insert libraries are rare. For example, function-driven screening of 10,000 clones derived from a metagenomic phage forest soil library resulted in the recovery of one cellulolytic clone (Wang et al. 2009). In addition, one clone producing cellulase activity on carboxymethyl cellulose (CMC) was identified by screening of 70,000 clones, which harbored a metagenomic library constructed from Korean soil (Kim et al. 2008). This clone also contained a xylanase gene, which was identified by sequence analysis of the cloned DNA.

Molecular analyses

The insert sizes of the three recombinant fosmids, fLX01, fLX02 and fLC01, recovered from the positive *E. coli* clones were 36,000, 30,200, and 16,250 bp, respectively. In order to identify genes conferring xylanase or cellulase activity subcloning of the recombinant fosmids was carried out by cloning of fosmid-derived DNA fragments (2–6 kbp) into the plasmid vector pCR-XL-TOPO. Subsequently, the corresponding *E. coli* clones were screened on HEC_{red}-containing and $xylan_{green}$-containing indicator agar. In this way, recombinant plasmids carrying the desired DNA fragments for all three fosmids were recovered.

The insert sizes of the three plasmids, pLC01, pLX01, and pLX02 (Table 1) derived from the corresponding fosmids fLC01, fLX02, and fLX02, were 3,220, 5,523, and 4,985 bp, respectively. The inserts of pLC01, pLX01, and pLX02 were sequenced and analyzed. Each of the three plasmids harbored a putative gene that exhibited similarities to known genes encoding cellulases (pLC01) or xylanases (pLX01 and pLX02) (Supplementary Table 2). Cellulase or xylanase activity was detected in crude extracts of corresponding recombinant *E. coli* strains *E. coli* TOP10/pLC01, *E. coli* TOP10/pLX01, and *E. coli* TOP10/pLX02 (data not shown).

Analysis and classification of cellulase-encoding gene

Cellulases and xylanases are subdivided in different glycoside hydrolase families based on amino acid sequence similarity (Cantarel et al. 2009; Henrissat 1991). The amino acid sequence deduced from the putative cellulase gene *cel01* of pLC01 comprised 831 amino acids with a calculated molecular mass of 90.4 kDa. A potential signal peptide of 21 amino acids was predicted at the *N*-terminus of Cel01. The amino acid sequence of the putative signal peptide showed the typical orientation of signal peptides with three distinct parts (N, H, and C domains) (Pugsley 1993). Protein domain analyses suggested that Cel01 belongs to the glycoside hydrolase family 9, as it contains a catalytic domain (amino acids 350–814) that is typical for family 9 members (Fig. 1). The multiple sequence alignment revealed that Cel01 contained putative catalytic aspartate and glutamate residues (D418, D421, and E808), which act as nucleophile (aspartate) or proton donor (glutamate) during substrate hydrolysis (Parsiegla et al. 2002) (Fig. 2a). In addition to the catalytic domain of family 9 glycoside hydrolases, Cel01 harbored a family 9 carbohydrate-binding module (CBM9, amino acids 46–231) and a *N*-terminal cellulase domain (Cel-*N*-term, amino acids 255–338) (Fig. 1). To our knowledge, family 9 carbohydrate-binding modules have previously been detected only in xylanases. The protein sequence of Cel01 was most similar (50% identity) to a cellulase from *Sorangium cellulosum* 'So ce 56' (Schneiker et al. 2007), but not over the entire length. The cellulase from *Sorangium cellulosum* possesses a similar modular structure as Cel01, but lacks a family 9 carbohydrate-binding module. Most of the metagenome-derived cellulase genes that have been identified by activity-based screening belonged to glycoside hydrolase families 5 and 9 (Duan and Feng 2010). Metagenome-derived family 9 cellulases were isolated from an aquatic community (Pottkämper et al. 2009), elephant dung (Wang et al. 2009), and an enrichment culture of an alkaline lake (Grant et al. 2004) but not from other soils with the exception of

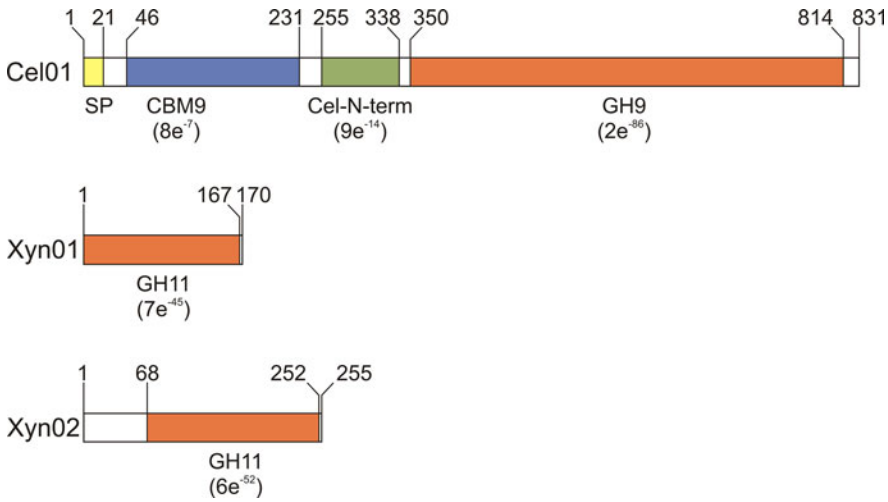


Fig. 1 Domain structure of Cel01, Xyn01, and Xyn02. The recorded E values for the family 9 carbohydrate-binding module (cd00005), the cellulase *N*-terminal domain (cd02850), the family 9 glycoside hydrolase catalytic domain (pfam00759), and the family 11 glycoside hydrolase catalytic domain (pfam00457) are given in parentheses. Abbreviations: *SP* signal peptide, *CBM9* family 9 carbohydrate-binding module, *Cel-N-term* cellulase *N*-terminal domain, *GH9* family 9 glycoside hydrolase, *GH11* family 11 glycoside hydrolase

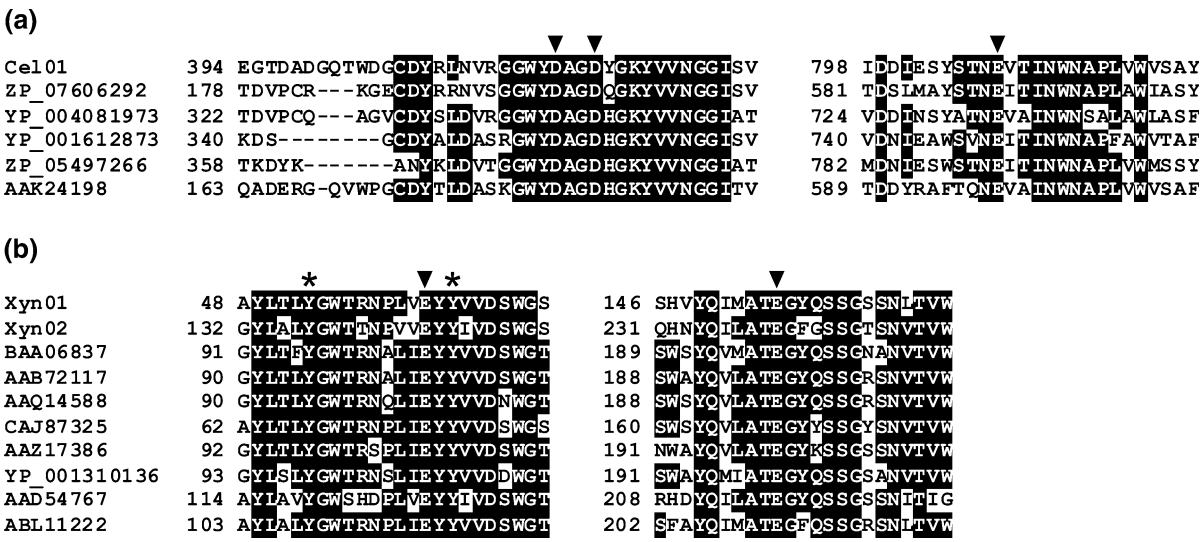


Fig. 2 Multiple sequence alignment of conserved regions in glycoside hydrolases belonging to family 9 (**a**) and family 11 (**b**). Identical amino acid residues are shown as *white letters* on a dark background. *Triangles* indicate catalytic residues and *asterisks* represent residues that are involved in substrate binding. References: Cel01, Xyn01, and Xyn02 (this study); ZP_07606292, GH9 from *Streptomyces violaceusniger* Tu 4113; YP_004081973, GH9 from *Micromonospora* sp. L5; YP_001612873, cellulase from *Sorangium cellulosum* 'So ce 56'; ZP_05497266, GH9 from *Clostridium papyrosolvens* DSM 2782; AAK24198, GH9 from *Caulobacter crescentus* CB15; BAA06837, xylanase I precursor from *Aeromonas punctata*; AAB72117, endo-beta-1,4-xylanase from *Geobacillus stearothermophilus*; AAQ14588, xylanase from *Bacillus firmus*; CAJ87325, endo-1,4-beta-xylanase B from *Thermobacillus xylanilyticus*; AAZ17386, endo-1,4-beta-xylanase from *Paenibacillus macerans*; YP_001310136, endo-1,4-beta-xylanase from *Clostridium beijerinckii* NCIMB 8052; AAD54767, endo-1,4-beta-xylanase from *Xylanimicrobium pachnodae*; ABL11222, xylanase from uncultured bacterium

compost soil (Pang et al. 2009). In summary, molecular analysis indicated that Cel01 is an extracellular cellulase belonging to glycoside hydrolase family 9. In addition, the enzyme exhibited a modular structure, which is not known from other cellulases.

Analysis and classification of xylanase-encoding genes

The gene products of the putative xylanase genes *xyn01* (170 amino acids) and *xyn02* (255 amino acids) exhibited calculated molecular masses of 18.6 and 27 kDa, respectively. Putative signal peptides were not found in the protein sequences of both gene products. Xyn01 and Xyn02 showed the highest similarity (78 and 69% identity, respectively) to a xylanase derived from a soil metagenomic library (Kim et al. 2008). The presence of a catalytic domain typical for members of the glycoside hydrolase family 11 indicated that Xyn01 and Xyn02 belong to this family (Fig. 1). In addition, the multiple sequence alignment showed that putative catalytic glutamate residues and tyrosine residues involved in substrate binding are present in the amino acid sequences of Xyn01 (E62 and E155, Y53 and Y64) and Xyn02 (E146 and E240, Y137 and Y148) (Fig. 2b). Genes encoding family 11 xylanases were also derived from metagenomic libraries of insect guts (Brennan et al. 2004; Warnecke et al. 2007) and single microorganisms such as *Dictyoglomus thermophilum* (Morris et al. 1998) and *Cellulomonas pachnodae* (Cazemier et al. 1999).

Initial characterization of the xylanases Xyn01 and Xyn02

Crude cell-free extracts of *E. coli* TOP10/pLX01 and *E. coli* TOP10/pLX02 were used to initially characterize Xyn01 and Xyn02. Both xylanases exhibited activity with different xylans as substrates, but no activity was recorded by employing barley glucan, CMC, HEC, laminarin, lichenan, or microcrystalline cellulose as substrates. Xyn01 showed the highest activity with xylan from larch wood whereas Xyn02 was most active with xylan from birch wood (Table 2). Larch wood xylan is a non-acetylated softwood xylan, which contains a higher 4-*O*-methylglucuronic acid content than the acetylated hardwood xylan from birch wood. In addition, softwood xylans are less branched than hardwood xylans (Sunna and Antranikian 1997). Thus, the observed differences in activity of both xylanases suggest that they require the presence of a particular substituent or branching extent of the xylan polymer to enhance their activity. The optimum activities of both enzymes were measured in the crude extracts from 10 to 70°C and range pH 4–10. Both enzymes were most active at 60°C and showed high activity over the entire tested pH range with optimal activities at pH 7 (Xyn01) and pH 6 (Xyn02). In addition, both xylanases retained approx. 50% activity at pH 4 and 10 (Supplementary Fig. 1). So far, the highest activities of xylanases derived from functional screens of metagenomes were found at 50°C and the retained activities at pH 4 were lower than 40% (Brennan et al. 2004; Lee et al. 2006; Li et al. 2009; Zhao et al. 2010). Only one metagenome-derived xylanase (XynA), which has been identified from a hot pool environmental DNA sample by a PCR-based approach, exhibited a higher temperature optimum than Xyn01 and Xyn02 (Sunna and Bergquist 2003).

Expression and purification of Cel01

To facilitate purification of the cellulase Cel01 the corresponding gene was amplified by PCR without the signal peptide sequence. The latter necessitated the addition of a start codon to the 5′ end of the coding

Table 2 Determination of xylanase activity in crude cell-free extracts derived from *E. coli* TOP10 cells harboring the plasmids pLX01 and pLX02

Plasmid	Specific activity (U mg^{-1})			
	Xylan (oat spelt)	Xylan (birch wood)	Xylan (larch wood)	Xylan (beech wood)
Cloning vector	0.008 ± 0.004	0.014 ± 0.006	ND	ND
pLX01	4.859 ± 0.051	4.569 ± 0.040	6.005 ± 0.158	3.362 ± 0.082
pLX02	0.168 ± 0.002	0.355 ± 0.003	0.337 ± 0.017	0.321 ± 0.013

ND no activity detected

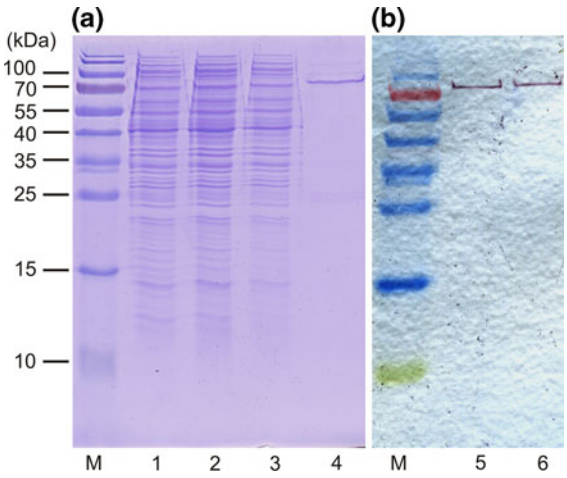


Fig. 3 Analysis of Cel01 production and purification by SDS PAGE (**a**) and Western blot analysis (**b**). His_6-tagged Cel01 was purified from cell extract of *E. coli* BL21/pCel01 by nickel affinity chromatography. **a** *Lanes*: *M*, marker proteins; 1, crude extract of *E. coli* BL21/pCel01; 2, flow through fraction; 3, wash fraction; 4, eluate (further purified by ultrafiltration); **b** *Lanes*: *M*, marker proteins 5, crude extract; 6, purified Cel01

region. The resulting PCR product was cloned into the expression vector pET101/D, thereby placing the genes under control of the IPTG-inducible T7 promoter and adding sequences encoding a His_6 tag and a V5 epitope. The resulting construct pCel01 (Table 1) was transformed into *E. coli* BL21 and production of Cel01 was induced by addition of 0.5 mM IPTG. After incubation for 12 h at 37°C, cells were harvested and crude cell-free extracts were prepared. The production of His_6-tagged Cel01 in the cell extracts was confirmed by Western-Blot analysis using antibodies against the V5 epitope (Fig. 3) and detection of cellulase activity. Subsequently, His_6-tagged Cel01 was purified from cell-free extracts by metal ion affinity chromatography and ultrafiltration. The specific cellulase activity of the final enzyme preparation was 780 ± 11.5 U mg^{-1} with barley glucan as substrate. SDS–PAGE of the purified enzyme revealed that Cel01 has a MW of approx. 90 kDa (Fig. 3). The observed molecular mass is in good agreement with the one deduced from the sequence of His_6-tagged version of Cel01 (91.9 kDa).

Properties of purified Cel01

The purified Cel01 exhibited the highest cellulolytic activity by employing soluble substrates containing β-1,4-linkages such as barley glucan (780 ± 11.5 U mg^{-1}), lichenan (516 ± 12.9 U mg^{-1}), CMC (90 ± 2.0 U mg^{-1}), and HEC (47.7 ± 2.5 U mg^{-1}). No activity with xylan from birch wood, oat spelt, larch wood or beech wood, microcrystalline cellulose, and laminarin (β-1,3/β-1,6-linkages) was detected. Thus, Cel01 is most likely an endo 1,4-β-glucan hydrolase. This was confirmed by the time courses of CMC and barley glucan hydrolysis, which were analyzed by thin layer chromatography (Fig. 4). At

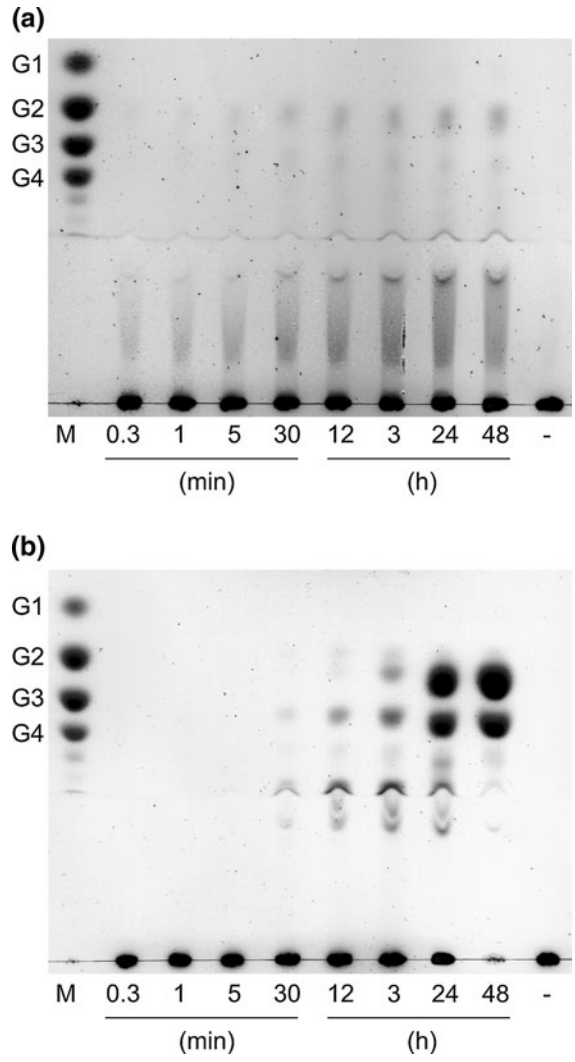


Fig. 4 Thin layer chromatography analysis of hydrolysis products released by Cel01. Degradation of CMC (**a**) and barley glucan (**b**) was analyzed at the indicated time points. *Lane M*, mixed standard sugars: glucose (G1), cellobiose (G2), cellotriose (G3), and cellotetraose (G4); *lane* -, control containing substrate without enzyme

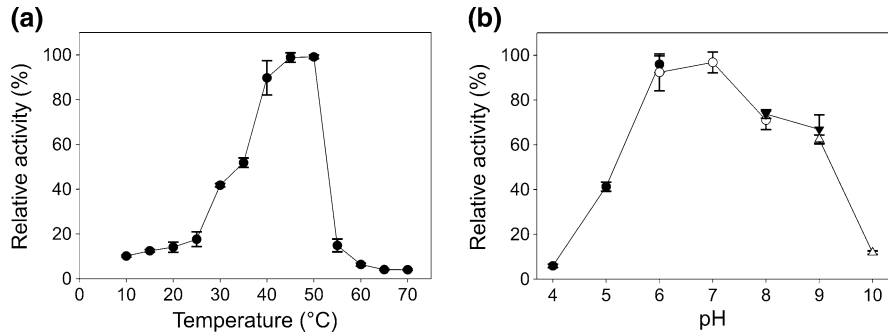


Fig. 5 Effect of temperature (**a**) and pH (**b**) on activity of Cel01. **a** Activity of Cel01 was determined using the standard assay at temperatures between 10 and 70°C. **b** Activity was measured at pH values between 4 and 10 using sodium acetate buffer (*black circles*), sodium phosphate buffer (*white circles*), Tris–HCl buffer (*black triangles*), and glycine-NaOH buffer (*white triangles*). The average of triplicate experiments is presented

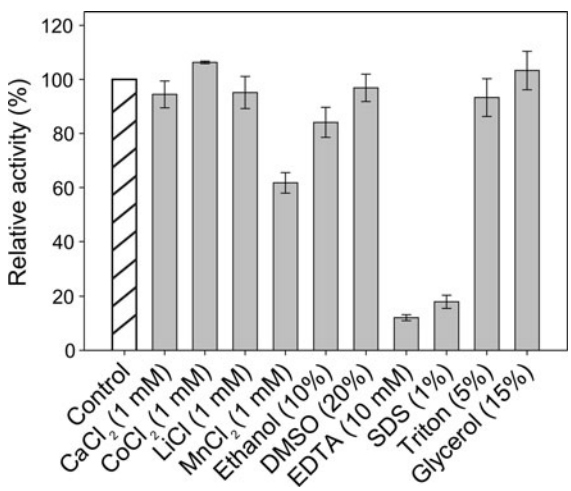


Fig. 6 Effect of metal ions and chemical agents on activity of Cel01. Activity without addition of metal ions or chemical agents was defined as 100%. The average of triplicate experiments is presented

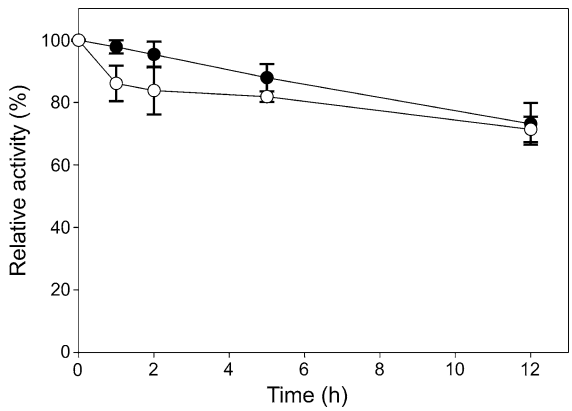


Fig. 7 Halotolerance of Cel01. Halotolerance was investigated by measuring residual activity using the standard assay after incubation of Cel01 for 12 h in 3 M KCl (*black circles*) and 4 M NaCl (*white circles*). The average of triplicate experiments is presented

first, CMC and barley glucan degradation resulted only in formation of high-molecular-mass products (Fig. 4). During prolonged incubation (CMC, 30 min to 48 h; barley glucan, 12–48 h) also small oligosaccharides were accumulated. In addition, no glucose was accumulated during the hydrolysis of CMC and barley glucan within incubation of 48 h. These results are also typical for an endo-wise action of the enzyme.

Purified Cel01 had significant activity between 30 and 50°C and exhibited optimal activity at 45 and 50°C (Fig. 5a). Cel01 was stable for 96 h below 50°C with more than 55% remaining activity (Supplementary Fig. 2). Incubation of Cel01 at 60°C caused complete inactivation of the enzyme. Interestingly, the optimal activities of other cellulases derived from soil metagenomic libraries ranged from 45 to 50°C and incubation at 60°C also resulted in rapid inactivation (Kim et al. 2008; Liu et al. 2011; Voget et al. 2006) (Table S3). Cel01 was highly active from pH 5 to pH 9 with optimal activity at pH 7 (Fig. 5b, Table S3). The cellulase activity was reduced by addition of $MnCl_2$, ethanol, EDTA, and SDS to 62 ± 3.8%, 84 ± 5.5%, 12 ± 1.1%, and 18 ± 2.4%, respectively (Fig. 6). $CoCl_2$ weakly stimulated the enzyme activity. When the enzyme was incubated in 3 M KCl or 4 M NaCl, it showed high halotolerance and retained more than 70% of its activity after 12 h incubation (Fig. 7).

Conclusions

In this study, construction and screening of large-insert soil-derived metagenomic libraries has led to the discovery and characterization of two novel xylanases and one cellulase. The novelty of these enzymes arises from the enormous genetic diversity of uncultured soil microorganisms. The recovered enzymes have some interesting properties such as high activity of the xylanases over a wide range of temperatures and pH values. Especially the metagenome-derived cellulase has potential for industrial application, as the enzyme is highly active and stable over a wide pH range, and shows high halotolerance.

Taking the relatively low number of screened library-containing *E. coli* clones screened into account the results presented here demonstrated that soil-based metagenomics in combination with use of a simple activity-based screening system is a method for the rapid isolation of diverse genes conferring the desired reactions.

Acknowledgments The work has been funded by the DFG Priority Program 1374 "Infrastructure-Biodiversity-Exploratories" (DA 374/4-1 and DA 374/6-1). Field work permits were given by the responsible state environmental offices of Baden-Württemberg and Brandenburg (according to §72 BbgNatSchG). We thank the BEO (biodiversity exploratories office) for support and project administration.

References

Beg QK, Kapoor M, Mahajan L, Hoondal GS (2001) Microbial xylanases and their industrial applications: a review. Appl Microbiol Biotechnol 56:326–338

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783–795

Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, Hernandez M, Keller M et al (2004) Unusual microbial xylanases from insect guts. Appl Environ Microbiol 70:3609–3617

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238

Cazemier AE, Verdoes JC, van Ooyen AJ, Op den Camp HJ (1999) Molecular and biochemical characterization of two xylanase-encoding genes from *Cellulomonas pachnodae*. Appl Environ Microbiol 65:4099–4107

Collins T, Gerday C, Feller G (2005) Xylanases, xylanase families and extremophilic xylanases. FEMS Microbiol Rev 29:3–23

Couto GH, Glogauer A, Faoro H, Chubatsu LS, Souza EM, Pedrosa FO (2010) Isolation of a novel lipase from a metagenomic library derived from mangrove sediment from the south Brazilian coast. Genet Mol Res 9:514–523

Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein FC, McManus PS, Handelsman J (2010) Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. Appl Environ Microbiol 76:4396–4401

Duan CJ, Feng JX (2010) Mining metagenomes for novel cellulase genes. Biotechnol Lett 32:1765–1775

Duan CJ, Xian L, Zhao GC, Feng Y, Pang H, Bai XL, Tang JL, Ma QS et al (2009) Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. J Appl Microbiol 107:245–256

Ferrer M, Golyshina OV, Chernikova TN, Khachane AN, Reyes-Duarte D, Santos VA, Strompl C, Elborough K et al (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. Environ Microbiol 7:1996–2010

Fischer M, Bossdorf O, Gockel S, Hänsel F, Hemp A, Hessenmöller D, Korte G, Nieschulze J et al (2010) Implementing large-scale and long-term functional biodiversity research: the biodiversity exploratories. Basic Appl Ecol 11:473–485

Fontes CM, Gilbert HJ, Hazlewood GP, Clarke JH, Prates JA, McKie VA, Nagy T, Fernandes TH et al (2000) A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. Microbiology 146:1959–1967

Gilad R, Rabinovich L, Yaron S, Bayer EA, Lamed R, Gilbert HJ, Shoham Y (2003) CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol 185:391–398

Grant S, Sorokin DY, Grant WD, Jones BE, Heaphy S (2004) A phylogenetic analysis of Wadi el Natrun soda lake cellulase enrichment cultures and identification of cellulase genes from these cultures. Extremophiles 8:421–429

Hall A (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Hårdeman F, Sjöling S (2007) Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. FEMS Microbiol Ecol 59:524–534

Heath C, Hu XP, Cary SC, Cowan D (2009) Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from Antarctic desert soil. Appl Environ Microbiol 75:4657–4659

Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316

Hong KS, Lim HK, Chung EJ, Park EJ, Lee MH, Kim JC, Choi GJ, Cho KY et al (2007) Selection and characterization of forest soil metagenome genes encoding lipolytic enzymes. J Microbiol Biotechnol 17:1655–1660

Jeon JH, Kim JT, Kang SG, Lee JH, Kim SJ (2009) Characterization and its potential application of two esterases derived from the arctic sediment metagenome. Mar Biotechnol 11:307–316

Jiang C, Hao ZY, Jin KE, Li SX, Che ZQ, Ma GF, Wu B (2010) Identification of a metagenome-derived β-glucosidase from bioreactor contents. J Mol Catal B Enzym 63:11–16

Kim SJ, Lee CM, Kim MY, Yeo YS, Yoon SH, Kang HC, Koo BS (2007) Screening and characterization of an enzyme with beta-glucosidase activity from environmental DNA. J Microbiol Biotechnol 17:905–912

Kim SJ, Lee CM, Han BR, Kim MY, Yeo YS, Yoon SH, Koo BS, Jun HK (2008) Characterization of a gene encoding cellulase from uncultured soil bacteria. FEMS Microbiol Lett 282:44–51

Kulkarni N, Shendye A, Rao M (1999) Molecular and biotechnological aspects of xylanases. FEMS Microbiol Rev 23:411–456

Lee ST, Lee JJ (1997) Insoluble dye substrate for screening and assay of xylan-degrading enzymes. J Microbiol Meth 29:1–5

Lee CC, Kibblewhite-Accinelli RE, Wagschal K, Robertson GH, Wong DWS (2006) Cloning and characterization of a cold-active xylanase from an environmental library. Extremophiles 10:295–300

Li R, Kibblewhite R, Orts WJ, Lee CC (2009) Molecular cloning and characterization of multidomain xylanase from manure library. World J Microbiol Biotechnol 25:2071–2078

Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW (2005) Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. Appl Environ Microbiol 71:7768–7777

Liu J, Liu WD, Zhao XL, Shen WJ, Cao H, Cui ZL (2011) Cloning and functional characterization of a novel endo-β-1, 4-glucanase gene from a soil-derived metagenomic library. Appl Microbiol Biotechnol 89:1083–1092

Lynd LR, Weimer PJ, Willem HZ, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506–577

Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577–583

Marchler-Bauer A, Anderson JB, Derbyshire MK, DeWeese-Scott C, Gonzales NR, Gwadz M, Hao L, He S et al (2007) CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Res 35:D237–D240

Miller GL (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal Chem 31:426–428

Morris DD, Gibbs MD, Chin CW, Koh MH, Wong KK, Allison RW, Nelson PJ, Bergquist PL (1998) Cloning of the *xynB* gene from *Dictyoglomus thermophilum* Rt46B.1 and action of the gene product on kraft pulp. Appl Environ Microbiol 64:1759–1765

Nacke H, Thürmer A, Wollherr A, Will C, Hodac L, Herold N, Schöning I, Schrumpf M, Daniel R (2011a) Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. PLoS ONE 6:e17000

Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R (2011b) Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil samples of the German Biodiversity Exploratories. FEMS Microbiol Ecol 78:188–201

Pang H, Zhang P, Duan CJ, Mo XC, Tang JL, Feng JX (2009) Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Curr Microbiol 58:404–408

Parsiegla G, Belaïch A, Belaïch JP, Haser R (2002) Crystal structure of the cellulase Cel9M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases. Biochemistry 41:11134–11142

Polizeli ML, Rizzatti AC, Monti R, Terenzi HF, Jorge JA, Amorim DS (2005) Xylanases from fungi: properties and industrial applications. Appl Microbiol Biotechnol 67: 577–591

Pottkämper J, Barthen P, Ilmberger N, Schwaneberg U, Schenk A, Schulte M, Ignatiev N, Streit WR (2009) Applying metagenomics for the identification of bacterial cellulases that are stable in ionic liquids. Green Chem 11:957–965

Pugsley AP (1993) The complete general secretory pathway in gram-negative bacteria. Microbiol Rev 57:50–108

Rosselló-Mora R, Amann R (2001) The species concept for prokaryotes. FEMS Microbiol Rev 25:39–67

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B (2000) Artemis: sequence visualization and annotation. Bioinformatics 16:944–945

Sanchez OJ, Cardona CA (2008) Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresour Technol 99:5270–5295

Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, Bartels D, Bekel T et al (2007) Complete genome sequence of the myxobacterium *Sorangium cellulosum*. Nat Biotechnol 25:1281–1289

Shedova EN, Berezina OV, Lunina NA, Zverlov VV, Schwarz WH, Velikodvorskaya GA (2009) Cloning and characterization of a large metagenomic DNA fragment containing glycosyl-hydrolase genes. Mol Gen Microbiol Virol 24: 12–16

Simon C, Daniel R (2009) Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol 85:265–276

Simon C, Daniel R (2011) Metagenomic analyses: past and future trends. Appl Environ Microbiol 77:1153–1161

Steele HL, Jaeger KE, Daniel R, Streit WR (2009) Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol 16:25–37

Strohl WR (2000) The role of natural products in a modern drug discovery program. Drug Discov Today 5:39–41

Sunna A, Antranikian G (1997) Xylanolytic enzymes from fungi and bacteria. Crit Rev Biotechnol 17:39–67

Sunna A, Bergquist PL (2003) A gene encoding a novel extremely thermostable 1, 4-beta-xylanase isolated directly from an environmental DNA sample. Extremophiles 7:63–70

Ten LN, Im W-T, Kim M-K, Lee S-T (2005) A plate assay for simultaneous screening of polysaccharide- and protein-degrading microorganisms. Lett Appl Microbiol 40:92–98

Thompson J, Higgins D, Gibson T (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Voget S, Steele HL, Streit WR (2006) Characterization of a metagenome-derived halotolerant cellulase. J Biotechnol 126:26–36

Wang F, Li F, Chen G, Liu W (2009) Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. Microbiol Res 164:650–657

Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, Mc Hardy AC et al (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450:560–565

Waschkowitz T, Rockstroh S, Daniel R (2009) Isolation and characterization of metalloproteases with a novel domain structure by construction and screening of metagenomic libraries. Appl Environ Microbiol 75:2506–2516

Yu EY, Kwon MA, Lee M, Oh JY, Choi JE, Lee JY, Song BK, Hahm DH, Song JK (2011) Isolation and characterization of cold-active family VIII esterases from an arctic soil metagenome. Appl Microbiol Biotechnol 90:573–581

Zhang Y, Zhao J, Zeng R (2011) Expression and characterization of a novel mesophilic protease from metagenomic library derived from Antarctic coastal sediment. Extremophiles 15:23–29

Zhao S, Wang J, Bu D, Liu K, Zhu Y, Dong Z, Yu Z (2010) Novel glycoside hydrolases identified by screening a Chinese Holstein dairy cow rumen-derived metagenome library. Appl Environ Microbiol 76:6701–6705

5.1 Supplemental information for chapter B5

Contents

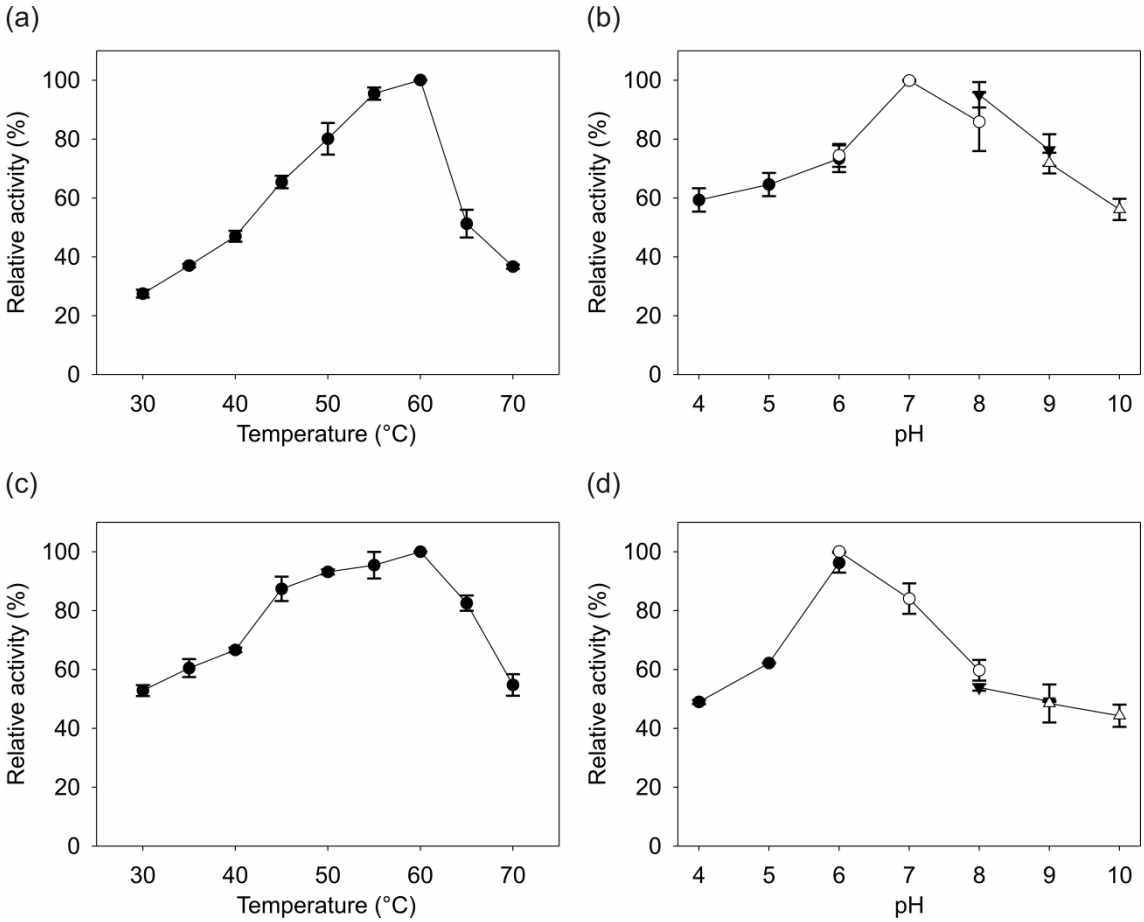


Fig. S1. Effects of temperature and pH on activity of Xyn01 and Xyn02. The effects were measured by using crude cell-free extracts prepared from *E. coli* TOP10 carrying pLX01 (Xyn01) or pLX02 (Xyn02). Activity was determined using the standard assay with xylan from birch wood as a substrate at temperatures between 30 and 70°C. The dependence of activity on pH was measured at pH values between 4 and 10 using sodium acetate buffer (black circles), sodium phosphate buffer (white circles), Tris-HCl buffer (black triangles), and glycine NaOH buffer (white triangles). (a) Effect of temperature on activity of Xyn01, (b) effect of pH on the activity of Xyn01, (c) effect of temperature on activity of Xyn02, and (d) effect of pH on the activity of Xyn02. The average of triplicate experiments is presented.

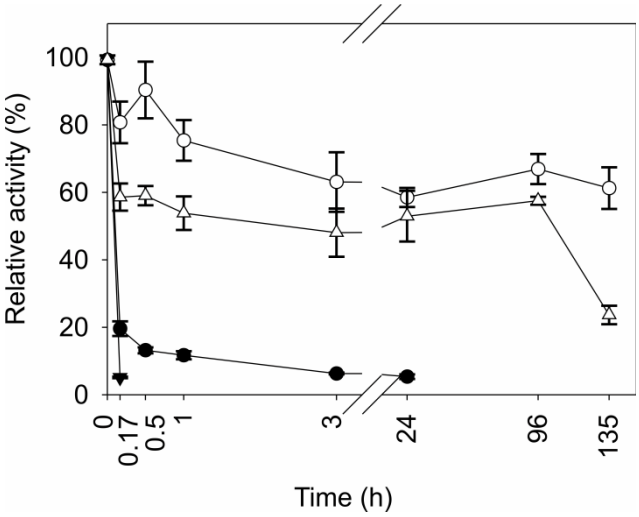


Fig. S2. Thermal stability of Cel01. The thermal stability of Cel01 was investigated after incubation of the enzyme at different temperatures (30°C, white circles; 40°C, white triangles; 50°C, black circles; 60°C, black triangles), and then measuring the residual activity with the standard assay. The average of triplicate experiments is presented.

Table S1. DNA yields and characteristics of the analyzed grassland soil samples.

Sample	DNA yield (µg/g soil)	Soil group	pH	OC[a] (g kg^{-1})	Total N (g kg^{-1})	C:N ratio	Gravimetric water content (%)
SEG9	53.6	Histosol	6.2	229.7	18.7	12.3	114.8
AEG3	25.3	Leptosol	6.3	53.7	5.2	10.4	57.2
AEG6	56.7	Leptosol	6.1	68.2	6.7	10.2	64.0

[a], Organic carbon content

Table S2. Open reading frames identified on pLC01, pLX01, and pLX02 and description of corresponding gene products and their observed sequence similarities.

Plasmid (accession no.)	ORF (Position)	No. of encoded amino acids	Closest similar protein, accession no. (no. of encoded amino acids), organism	E value	Amino acid homology to the closest similar protein (% identity)
pLC01 (JF799947)	*cel01* (3138-643)	831	Cellulase, YP_001612873 (772), *Sorangium cellulosum* 'So ce 56'	7e-150	288/581 (50%)
pLX01 (JF799945)	*xyn01* (2839-2327)	170	Xylanase, ABL11222 (226), uncultured bacterium	2e-71	132/171 (78%)
	orf01 (3226-4524)	432	Transcriptional regulator of DeoR family, YP_004119403 (322), *Pantoea* sp. At-9b	0.001	62/249 (25%)
pLX02 (JF799946)	*xyn02* (757-1524)	255	Xylanase, ABL11222 (226), uncultured bacterium	1e-69	140/203 (69%)
	orf01 (1762-2961)	398	Ser/Thr protein phosphatase family protein, ZP_05031548 (321), *Brevundimonas* sp. BAL3	2e-29	88/277 (32%)

C Discussion

The introduction and application of metagenomics and rRNA gene sequencing opened a new era of microbial diversity research. Small subunit ribosomal RNA gene-based phylogenetic analyses revealed the presence of previously unknown prokaryotic lineages in various environments on Earth (López-García and Moreira 2008). Additionally, novel biocatalysts have been recovered by function-based and sequence-based metagenomics (Steele et al. 2009). In both cases, the discovery of uncultured microorganisms as well as the identification of novel biocatalysts, environmental DNA represents the basis.

In this study, German forest and grassland soils that covered different management types were explored. The bacterial diversity and community structure of forest topsoil, and grassland topsoil and subsoil samples were analyzed by a pyrosequencing-based approach. Partial 16S rRNA genes, amplified from soil-derived DNA, represented the starting material for pyrosequencing. Bacterial taxa were classified down to the genus level, and diversity and richness estimates were performed. Subsequently, the impacts of land use type, management type, soil horizon, tree species, and soil characteristics on diversity and relative abundances of bacterial groups were investigated by statistical analyses.

In addition, metagenomic small-insert and large-insert libraries were constructed from soil-derived DNA. The metagenomic libraries were subjected to function-based screening for novel lipolytic and (hemi)cellulolytic genes. Insert DNA of positive clones was sequenced and deduced amino acid sequences of potential lipolytic and (hemi)cellulolytic ORFs were used for enzyme classification. Partial characterization of clones carrying genes encoding putative lipolytic enzymes and biochemical characterization of cellulases and xylanases were carried out. The approaches conducted in this study are briefly summarized in Figure 4.

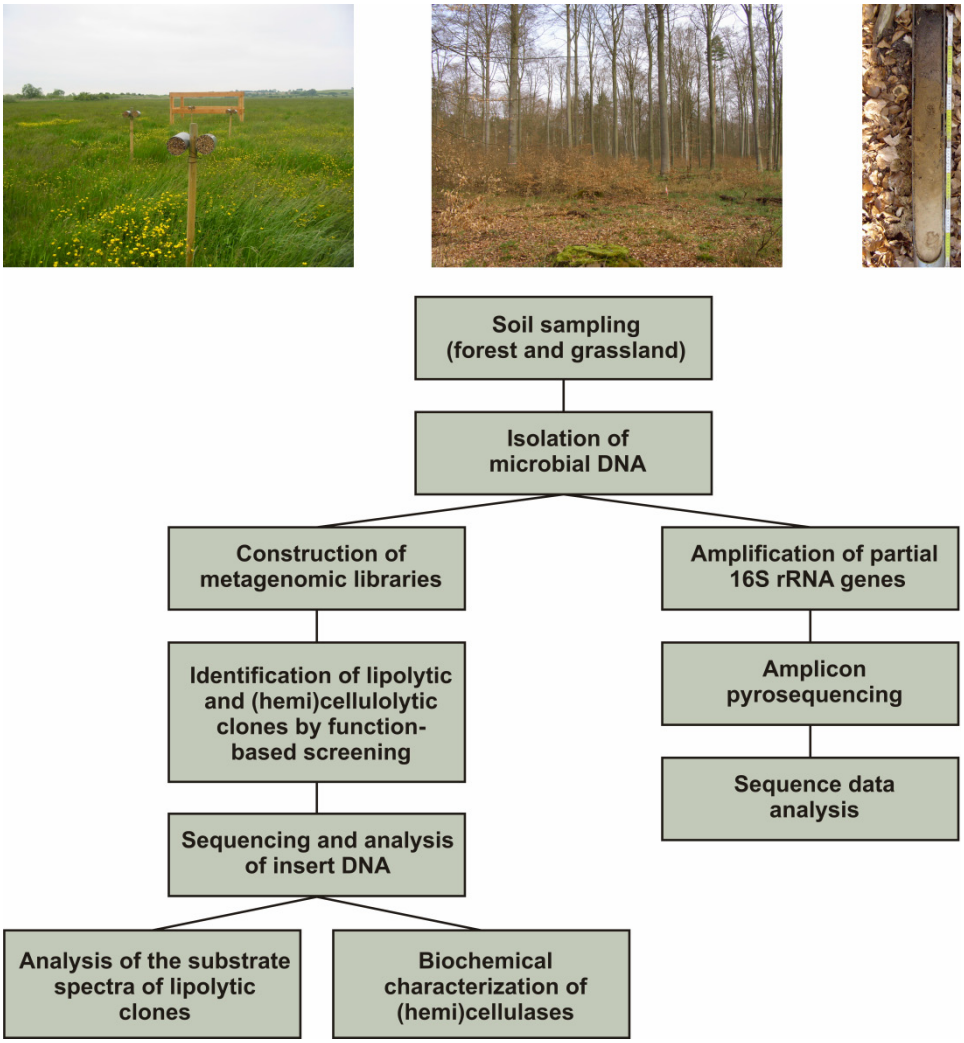


Fig. 4. Schematic representation of the strategy to analyze soil microbial communities with the help of metagenomic approaches. The pictures at the top show forest and grassland sampling sites as well as a forest soil core.

1. Phylogenetic analysis of microbial communities in soils under different management regimes

1.1 OTU-based diversity and richness estimation

Small subunit ribosomal RNA gene surveys and DNA:DNA hybridization experiments based on environmental DNA have demonstrated that microbial diversity in soil is orders of magnitude higher than the 5,000 microbial species described in Bergey's tax-

onomic outline (Torsvik et al. 1990; Guerrero 2001; Elshahed et al. 2008). Sanger sequencing as well as pyrosequencing of small ribosomal subunit genes and subsequent OTU (operational taxonomic unit) analyses have been applied for overall diversity evaluation of soil-inhabiting microorganisms (Fierer et al. 2007; Roesch et al. 2007). Especially, pyrosequencing is considered as a door opener for comprehensive phylogenetic assessment. However, the increasing number of pyrosequencing surveys focusing on soil microbial communities show strong variations in OTU numbers. Consequently, the question about factors responsible for this phenomenon arises.

Defining a phylum at 20% sequence divergence, nearly complete or full estimated taxonomic richness in individual soil samples of the Schwäbische Alb and the Hainich region was covered by the pyrosequencing effort (22,000 to 61,366 sequences per sample) (Chapter B1 and B2). Consistently, rarefaction analyses described in pyrosequencing studies by Roesch et al. (2007) (26,140-53,533 sequences per sample) and Acosta-Martínez et al. (2008) (6,000-7,000 sequences per sample) also revealed the coverage of the taxonomic diversity at phylum level in single forest and agricultural soil samples. The full extent of diversity was neither assessed in soil samples from the Schwäbische Alb nor in soil samples from the Hainich region at species level (3% genetic distance) (Chapter B1 and B2). This is in accordance with all other published pyrosequencing and Sanger sequencing surveys on soil microbial communities in which predicted richness was also higher than observed OTUs at 3% genetic distance (representative examples are depicted in Fig. 5).

The analyzed forest soils in this thesis covered a wide pH range (pH 3.30 to 6.37), with the most acidic soil sample exhibiting the lowest estimated diversity at all analyzed genetic distances (Chapter B1). Accordingly, effects of pH on diversity of soil bacterial communities have been recognized in other surveys (Fierer and Jackson 2006; Hartman et al. 2008; Lauber et al. 2009; Chu et al. 2010; Griffiths et al. 2011). By analyzing 98 and 88 soils from across North and South America, Fierer and Jackson (2006) and Lauber et al. (2009) found a peak of bacterial diversity in soils with near-neutral pH values. Despite the coverage of a similar pH range no peak of bacterial diversity at near-neutral pH in approximately 1,000 soil cores collected across Great Britain was assessed by Griffiths et al. (2011). Similarly, other recent surveys, in which correlations between soil pH and bacterial diversity in wetland soils (Hartman et al. 2008) and in arctic tundra soils (Chu et al. 2010) were analyzed, also showed no peak of bacterial diversity at

near-neutral pH. In this study too, a peak of bacterial diversity in soils with near-neutral pH values was not confirmed, suggesting locally different correlations between pH and overall bacterial diversity.

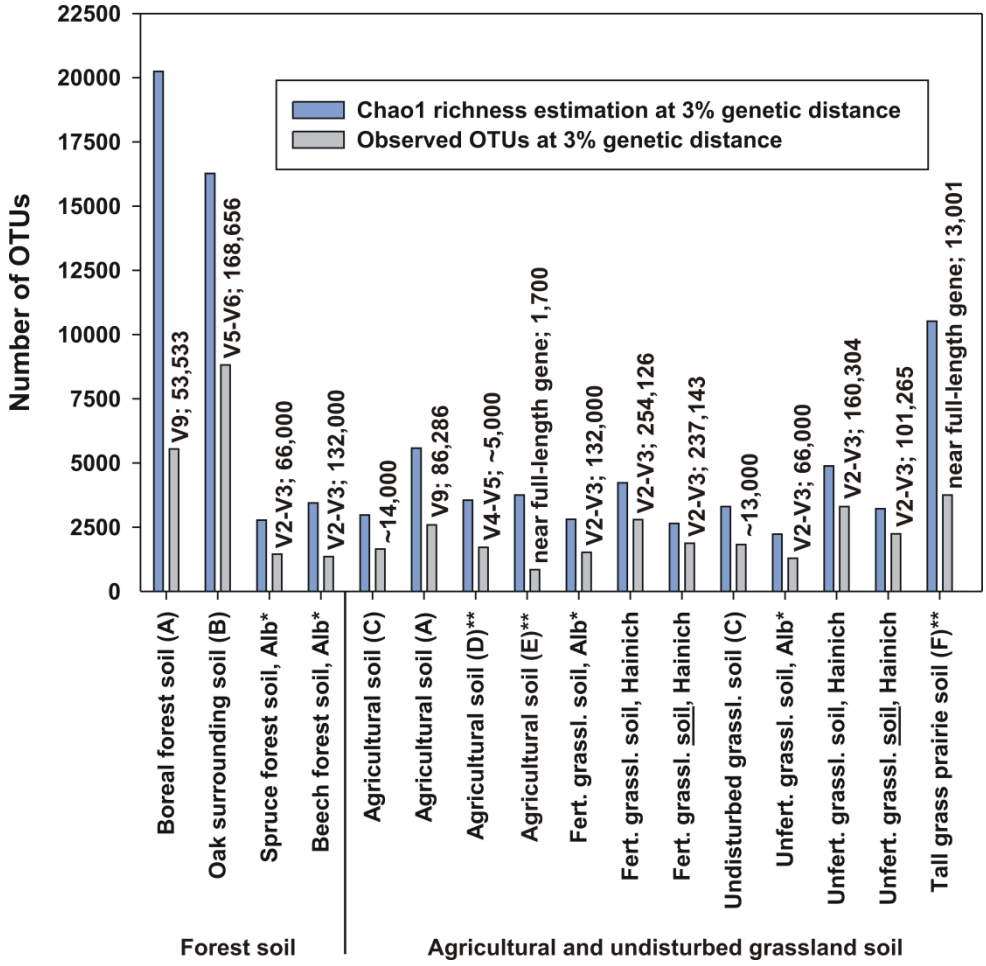


Fig. 5. Average bacterial richness estimates of soils under different management at a genetic distance of 3% within this thesis and other studies. Richness is expressed as number of observed unique OTUs. In addition, richness has been estimated by Chao1 richness estimator, which is a nonparametric richness estimator based on distribution of singletons and doubletons. (A), Roesch et al. 2007; (B), Uroz et al. 2010; (C), Acosta-Martínez et al. 2008; (D), Morales et al. 2009; (E), Tringe et al. 2005; (F), Elshahed et al. 2008. No asterisks, non-denoised pyrosequencing datasets; *, denoised pyrosequencing datasets; **, Sanger sequencing-based datasets. Sequenced regions of the 16S rRNA (if provided in the respective study) and sequence numbers are depicted next to the bars. Fert., fertilized; Unfert., unfertilized; grassl., grassland; soil, subsoil (all other analyzed soils were topsoils).

Considering the grassland soils investigated in this study, the collected samples showed strong variations in diversity, even within the same management type (Chapter B1 and B2). Nevertheless, a significant soil depth effect on bacterial diversity in the Hainich samples was recorded (Chapter B2). More precisely, Hainich grassland soil-derived

community analysis revealed that bacterial diversity was significantly higher in topsoil samples than in subsoil samples (Fig. 5). This finding is in accordance with previous PLFA (phospholipid fatty acid) and 16S rRNA gene analyses, in which a decrease of overall microbial community richness and the detected number of species with soil depth were indicated (Fierer et al. 2003; Hansel et al. 2008).

To come back to the fact that available studies on microbial diversity in soil show strong deviations of OTU numbers, driving forces other than soil characteristics and management have to be considered. First, different DNA extraction methods were employed in the available pyrosequencing surveys. These methods exhibit distinct biases that restrict the original diversity of soil metagenomic DNA (Delmont et al. 2011). Furthermore, PCR biases, amplicon length, and the selected 16S rRNA gene regions have an impact on the outcome of diversity evaluation (Engelbrektson et al. 2010). Recently, it has been reported that noise resulting from pyrosequencing can lead to overestimates of OTUs (Quince et al. 2009; Reeder and Knight 2010). Consistently, observed as well as predicted OTUs in soil samples of the Schwäbische Alb were approximately 2 to 3-fold higher in non-denoised datasets than in denoised datasets (Chapter B1). This is the first study on soil microbial communities employing denoising tools. Therefore, comparisons with currently available pyrosequencing-derived estimates of OTUs are limited. In this thesis, two hypervariable 16S rRNA gene regions were analyzed (V2-V3) and the average read length was 255 bp (Schwäbische Alb) and 262 bp (Hainich region). Roesch et al. (2007) performed one of the first pyrosequencing-based analyses of soil microbial communities. The generated sequences spanned only one hypervariable region of the 16S rRNA gene (V9; average read length, 103 bp) and denoising was not performed. Thus, despite of a similar sequence number derived from coniferous forest soil samples (Roesch et al. 2007, boreal forest, 53,533; Schwäbische Alb, spruce forest, 66,000) it is not suprising that the richness estimate by Roesch et al. (2007) (20,244 OTUs) is much higher than the one from the Schwäbische Alb (2,775 OTUs) (Fig. 5). However, continuous development of GS FLX Titanium chemistry, allowing the inclusion of more hypervariable 16S rRNA gene regions, and standardized application of denoising will increase the comparability of pyrosequencing-based surveys in the future.

1.2 Distribution of classifiable taxa

Classification of bacterial taxa using ribosomal databases represents a valuable complement to "anonymous" OTU-based analyses. Currently, the Ribosomal Database Project (RDP) (Cole et al. 2009) comprises 1,921,179 aligned and annotated 16S rRNA gene sequences. Of these sequences, 7,877 originate from species type-strains, commonly used for taxonomic classification of environmental 16S rRNA gene sequences. Classification of bacterial taxa is relatively simple and rapid, but the question about the roots for varying distributions of classifiable taxa in soil is often neglected in current studies. This is mainly attributable to lacking determination of soil characteristics or missing replicates that would allow statistical evaluation.

In general, soil bacterial communities of the Schwäbische Alb and the Hainich region exhibited phyla typically encountered in soil. The nine dominant phyla in soil according to Janssen (2006) are *Proteobacteria*, *Acidobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Bacteroidetes*, *Chloroflexi*, *Planctomycetes*, *Gemmatimonadetes*, and *Firmicutes*, representing 78.1 (Schwäbische Alb) and 77.9% (Hainich region) of the sequences that were assigned to the domain bacteria in this study. Accordingly, all or most of these phyla were also detected in other recent Sanger-based and pyrosequencing-based surveys on bacterial communities in forest and grassland soils (Acosta-Martínez et al. 2008; Dimitriu and Grayston 2010; Uroz et al. 2010). Nevertheless, the relative abundances of the mentioned phyla can differ widely between and within different soils (Janssen, 2006) (Fig. 6). Methodical differences between studies bear a side effect on the number and relative abundances of detected phyla. The *Verrucomicrobia* can be referred to as a representative example. Bergmann et al. (2011) reported that commonly used PCR primers targeting partial regions of the 16S rRNA gene result in recovery of low relative abundances of *Verrucomicrobia*. The analysis of 181 soil samples derived from across North America, South America, Europe, and Antarctica indicate that *Verrucomicrobia* relative abundances make up an average of 23% of bacterial sequences in soil (Bergmann et al. 2011). In contrast, *Verrucomicrobia* relative abundances in this study and other recent surveys on the microbial habitat soil accounted for only 0-5.1% of bacterial sequences (Fig. 6).

Apart from methodical peculiarities of different studies, the relative abundances of dominant phyla and proteobacterial classes identified in this thesis corresponded roughly

to those reported in a meta-analysis of multiple soil-derived bacterial 16S rRNA gene libraries (Janssen 2006). Fittingly, pasture, meadow, and forest soils analyzed in this thesis were also considered in the meta-analysis conducted by Janssen (2006). The most abundant phyla and proteobacterial classes (≥ 1%) across all soil samples of the Schwäbische Alb were *Acidobacteria*, *Alphaproteobacteria*, *Actinobacteria*, *Betaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, and *Firmicutes*, representing 19.6, 18.3, 16.1, 5.9, 3.4, 2.9, and 1.2%, respectively, of all bacterial sequences. Topsoil and subsoil grassland samples derived from the Hainich region were dominated by *Acidobacteria* (21.5%), *Betaproteobacteria* (12.8%), *Actinobacteria* (9.1%), *Gammaproteobacteria* (8.9%), *Alphaproteobacteria* (7.7%), *Deltaproteobacteria* (4.1%), *Chloroflexi* (3.0%), *Firmicutes* (2.7%), and *Bacteroidetes* (1.1%). Compared to the average relative abundances of *Bacteroidetes* (5%), *Planctomycetes* (2%), and *Gemmatimonadetes* (2%) reported by Janssen (2006), soil samples analyzed in this study appear to be relatively poor with respect to these phyla (Schwäbische Alb: 0.5, 0.001, and 0.01%, respectively; Hainich region, 1.1, 0.0003, and 0.03%, respectively). However, soil samples collected from an oak forest located in Breuil-Chenue (France) were also poor in *Bacteroidetes* (1.7%) (Uroz et al. 2010) whereas grassland soils derived from Texas High Plains region were rich in this phylum (12.1%) (Acosta Martínez et al. 2008) (Fig. 6). Furthermore, the *Bacteroidetes* were not detected in any of the samples analyzed in a Sanger sequencing survey on Canadian boreal forest soils (Dimitriu and Grayston 2010) (Fig. 6).

In a previous study on tropical soil, a bacterial community shift correlated with change from forest to pasture vegetation was observed (Nüsslein and Tiedje 1999). This finding corresponds to significantly varying relative abundances of different soil bacterial taxa between the land use types forest and grassland in this study. The most abundant phylogenetic group across all forest soil samples was the proteobacterial class *Alphaproteobacteria* (25.1%), which was only represented by 11.4 (Schwäbische Alb) and 9.2% (Hainich region) relative abundance in grassland topsoils (Chapter B1 and B2). Consistently, the *Alphaproteobacteria* were the dominant phylogenetic group in French oak forest (24.5%, Uroz et al. 2010) and Canadian boreal forest soils (28.1%, Dimitriu and Grayston 2010) too, whereas their relative abundance was also drastically reduced in other grassland soils (8.8%, Acosta Martínez et al. 2008; 10.3%, Elshahed et al. 2008) (Fig. 6). Besides the statistically significant land use type effect, a tree species effect

was recorded in this study: *Deltaproteobacteria* were less abundant in spruce forest than in beech forest. Significant land use type and tree species effects were also detected at genus level (Chapter B1), indicating vegetation dependency of the appearance and abundance of various bacterial groups in soil. In this respect, analysis of the distribution of single soil bacterial genera between forest and grassland sites based on hundreds of thousands sequences is currently unique. Thus, referring to the correlations obtained in this study will be valuable for future surveys.

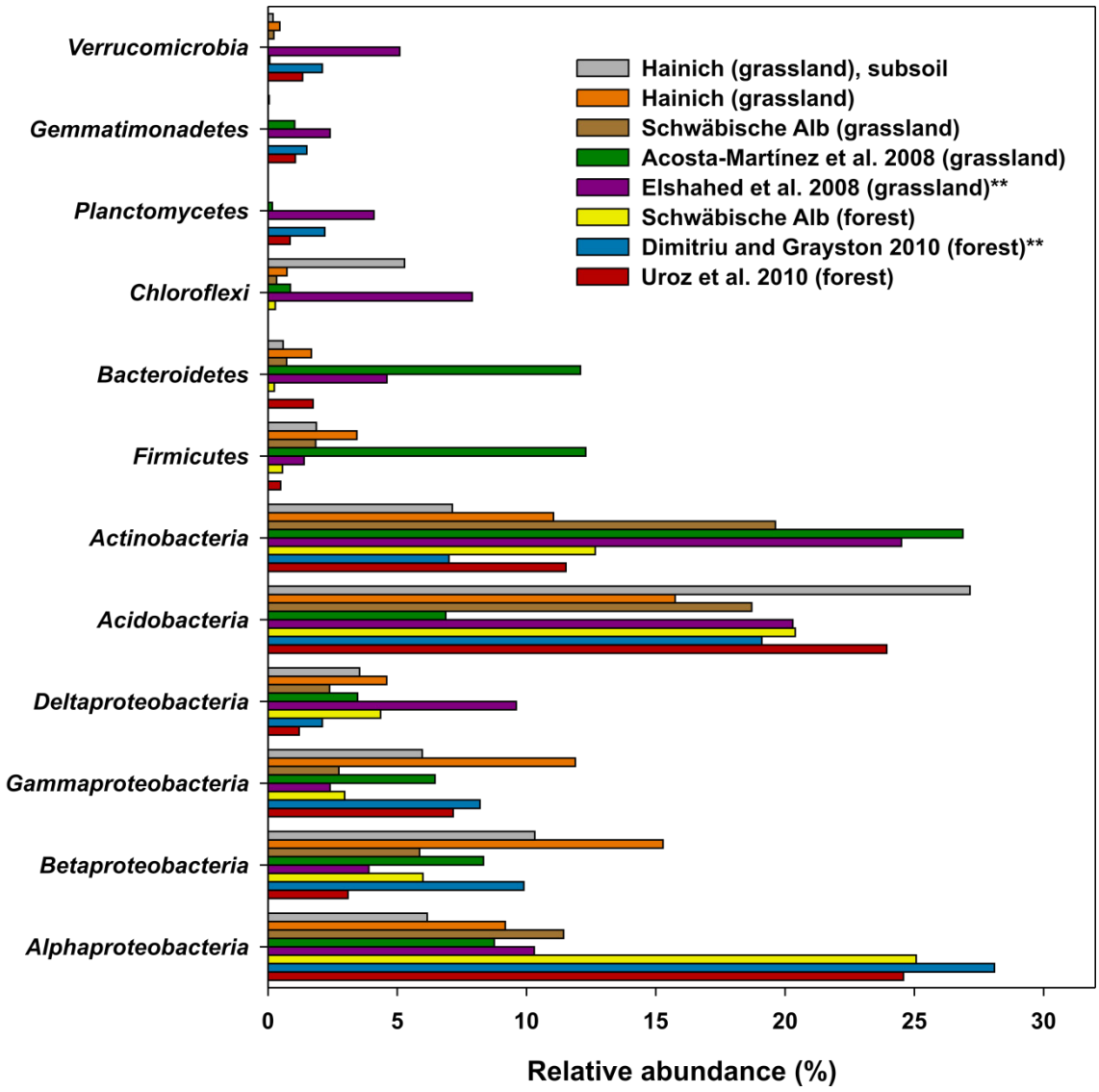


Fig. 6. Relative abundances of dominant phylogenetic groups typically encountered in soil according to Janssen (2006) within forest and grassland samples investigated in this thesis and other surveys. The total number of analyzed grassland and forest soils of each study was considered. In case of the *Proteobacteria*, dominant classes were regarded. The abundances of phylogenetic groups relative to all bacterial sequences that were classified in each study are depicted. Topsoil samples were analyzed, unless otherwise noted. No asterisks, pyrosequencing-derived datasets; **, Sanger sequencing-based datasets.

Among the determined soil characteristics such as total nitrogen content, organic carbon content, pH, and soil texture, pH exhibited the strongest impact on soil bacterial community structure within the Schwäbische Alb samples (Chapter B1). This result supports the findings of other recent surveys in which pH also induced soil microbial community shifts (Lauber et al. 2009). Similar to this study, pH effects on soil microbial communities were recorded along varying land use by Hartman et al. (2008) and Lauber et al. (2008), but also across continental and small scales (Fierer and Jackson 2006; Baker et al. 2009; Lauber et al. 2009). The relative abundances of *Actinobacteria* and *Bacteroidetes* significantly increased with higher soil pH values in this study ($P < 0.05$ in both cases). Accordingly, both phyla showed analogous correlations to pH in 29 Arctic soils (Chu et al. 2010) and 88 soils from across North and South America (Lauber et al. 2009) ($P < 0.05$ in both cases) (Fig. 7). In addition to pH effects at phylum level, correlations to soil pH were detected at the order, family, and genus level as well as for various acidobacterial subgroups in this study (Chapter B1). The soil pH effect on acidobacterial subgroups has to be emphasized, since several (subgroups 1, 3, 6, 13, 17, and 18) of the currently 26 described acidobacterial subgroups (Barns et al. 2007) showed strong correlations to soil pH ($P < 0.001$ in all cases). The detection of these correlations seems to be highly reproducible, as Jones et al. (2009) found similar soil pH-dependent distributions of subgroups 1, 3, 6, 13, 17, and 18.

Regarding the soil samples derived from the Hainich region, pH effects on bacterial taxa were not detected. This result is most likely related to the small pH range covered by the Hainich soil samples (pH 6.03 to 7.40) compared to the Schwäbische Alb soil samples (pH 3.30 to 7.24). Instead of pH, organic carbon content induced significant effects on bacterial community composition within the Hainich soil samples (Chapter B2). Furthermore, the abundances of numerous bacterial taxa differed strongly between A horizon (topsoil) and B horizon (subsoil) samples (Chapter B2) (Fig. 6). Hansel and colleagues (2008) also determined horizon specific distributions of bacterial groups, but their study was based on less than 400 bacterial 16S rRNA gene sequences. Thus, the large pyrosequencing dataset derived from Hainich topsoil and subsoil samples (18 samples, on average 41,824 sequences per sample) represents so far the most comprehensive dataset to assess soil depth effects.

In conclusion, factors altering bacterial community structure in forest and grassland soil samples were identified. The major factors shifting the distribution of bacterial taxa

were land use type, tree species, pH (when analyzed soils covered a wide pH range), organic carbon content (when pH values of analyzed soils were near-neutral), and soil depth. No management type effects on bacterial community structure were discovered. This study and previous surveys show that the distribution of bacterial taxa in soil is to some extent non-randomly and partially predictable through the knowledge of environmental parameters.

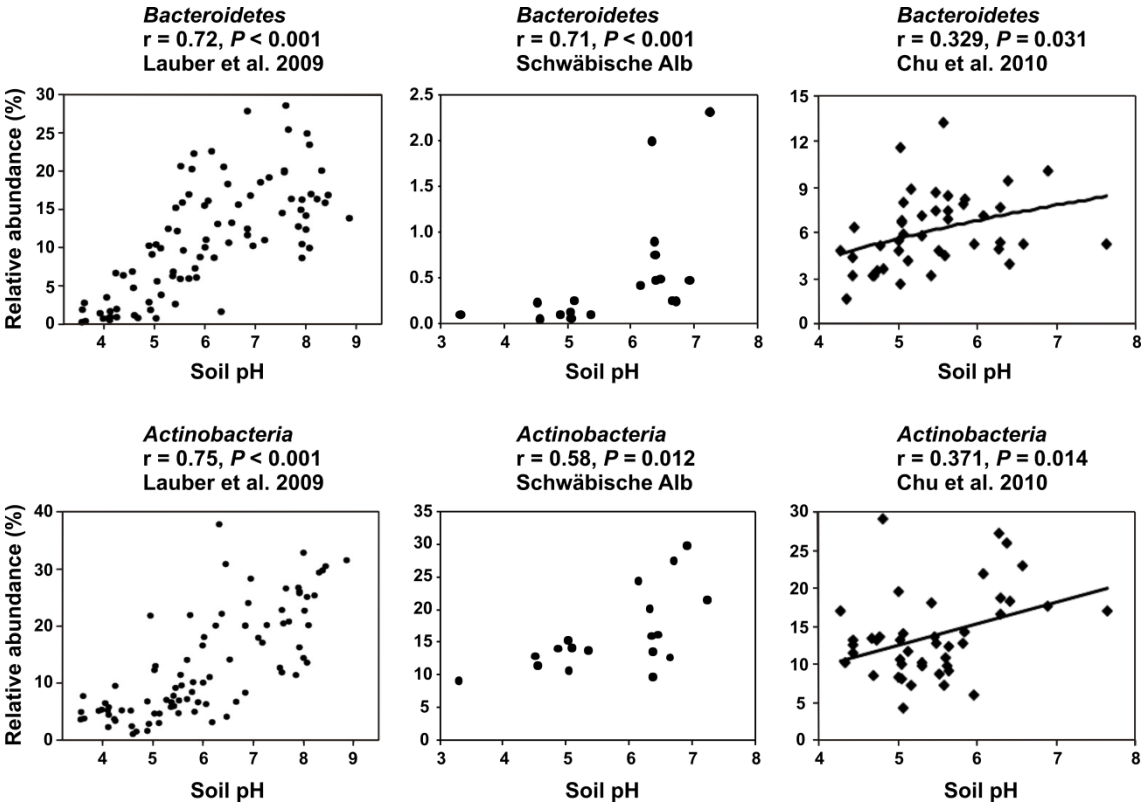


Fig. 7. Correlations of relative abundances of bacterial phyla *Bacteroidetes* and *Actinobacteria* to soil pH within this thesis and two other studies. Spearman's rank (Schwäbische Alb) and Pearson (Lauber et al. 2009; Chu et al. 2010) correlation coefficients (r) with the associated *P* values are shown for both phyla.

2 Metagenomic biocatalyst discovery

The genetic diversity present in different environments represents a rich source for the discovery of industrially important biocatalysts. Enzymes with valuable characteristics such as an alkaliphilic esterase (Heath et al. 2009), a cold-active xylanase (Lee et al. 2006a), a chloride tolerant laccase (Fang et al. 2011), and a beta-glucosidase with excellent glucose tolerance (Fang et al. 2010) have been identified by applying metagenomic

approaches. Nevertheless, in the face of environmental and economic pressure the request for efficient production solutions increases. Currently, the enormous potential for biocatalyst discovery by employing samples from diverse environments like soil is far from being exhausted. In this study, forest and grassland soils have been used for the construction of metagenomic libraries.

Small-insert and large-insert metagenomic libraries containing directly cloned environmental DNA can be screened to identify genes encoding novel biocatalysts. Plasmids are appropriate vectors for cloning of small environmental DNA fragments (≤ 15 kb), whereas fosmids, cosmids, and bacterial artificial chromosomes (BACs) can be used to clone large environmental DNA fragments (fosmids and cosmids, ≤ 40 kb; BACs, 100 - 200 kb) (Uchiyama and Miyazaki 2009; Simon and Daniel 2011). In contrast to plasmids which only bear capacity for single genes or small operons, fosmids, cosmids, and BACs can harbor whole gene clusters (Daniel 2005). Plasmids often have high copy numbers and strong promoters, enabling the identification of weakly-expressed metagenomic genes by function-based screening. To benefit from the advantages of both, small-insert and large-insert libraries, several plasmid and fosmid libraries have been constructed in this study (Chapter B3 and B5).

The function-based screening approach represents one of two routinely applied screening strategies to identify genes encoding novel biocatalysts. Library clones expressing a certain enzyme activity are usually detected when grown on agar plates supplemented with an indicator substrate. The most important advantage of this method is the potential to identify entirely novel genes. Unusual microbial xylanases with domains of unknown function and remarkably high phylogenetic distance to known xylanases have been derived from insect guts by function-based screening (Brennan et al. 2004). Disadvantages of this screening approach include incomplete expression of cloned metagenomic genes by the conventionally used host *E. coli* and insufficient sensitivity of agar plate-based screening. The second strategy, the sequence-based screening approach, is based on sequence homology. Conserved regions of available gene sequences serve to design PCR primers or probes that are then used for the discovery of novel variants of genes. Extracted environmental DNA or metagenomic libraries can be used as a target. In this way, previously unknown genes encoding various potential biocatalysts such as pullulanases (Tang et al. 2008), alkane hydroxylases (Xu et al. 2008), or nitrite reductases (Demanèche et al. 2009) have been identified. The sequence-based screening approach

is often more rapid compared to the function-based screening approach but it does not allow the isolation of so far unknown gene families (Simon and Daniel 2009). To allow the identification of entirely novel genes and gene families, a function-based screening approach has been applied in this study.

2.1 Identification and analysis of lipolytic genes and gene families

To isolate genes encoding lipolytic activity from the constructed metagenomic libraries, a simple agar plate assay using tributyrin as a substrate was chosen. Thirty-seven unique clones harboring novel lipolytic genes have been identified from forest and grassland soil-derived plasmid and fosmid libraries (Chapter B3). Other studies using plasmids or fosmids as vectors also identified unknown lipolytic genes from diverse soil environments such as Antarctic desert (Heath et al. 2009), Chinese wheat field (Sang et al. 2011), German meadow, sugarbeet field, and river valley (Henne et al. 2000). Bacterial lipolytic enzymes can be subdivided into eight different families (I-VIII) based on amino acid similarity and some fundamental biological properties (Arpigny and Jaeger 1999). Esterases and lipases belonging to putatively new families of lipolytic enzymes have been derived from sheep rumen (Bayer et al. 2010), surface sea water (Chu et al. 2008), marine sediment (Hu et al. 2010), tidal and intertidal flat sediment (Lee et al. 2006b; Kim et al. 2009), and salted shrimp (Park et al. 2011). All of the deduced proteins of available soil metagenome-derived lipolytic genes could be classified according to Arpigny and Jaeger (1999). The only exception was esterase EstD2 originating from plant rhizosphere soil (Lee et al. 2010). Considering the high number of so far isolated lipolytic enzymes from soil metagenomes, novel families of lipolytic enzymes were rarely discovered.

In this study, two of the 37 gene products of lipolytic genes could not be assigned to known families, suggesting that these proteins present new families of lipolytic enzymes. Furthermore, two genes encode true lipases, which are also rarely discovered from soil metagenomes. Only one cold-adapted, one thermostable, and a lipase highly stable in organic solvents as well as a lipase showing similarity to lipase Lip of *Streptomyces albus* were so far derived from soil metagenomic libraries by function-based screening (Henne et al. 2000; Wei et al. 2009; Glogauer et al. 2011). Substrate specificity of the clones carrying lipolytic genes and deduced amino acid sequence analysis of

these genes revealed that the remaining 35 predicted lipolytic enzymes identified in this study represent esterases. In conclusion, a high number of novel genes encoding lipolytic enzymes have been isolated by employing a simple function-driven screening approach employing indicator agar plates. The high diversity of lipolytic enzymes recovered in this study, highlights the enormous potential of soil for discovery of novel biocatalysts.

2.2 Identification and characterization of cellulolytic and hemicellulolytic genes and enzymes

Only a few metagenome-derived (hemi)cellulases have been published. This is mainly due to low probability (hit rate) of identifying (hemi)cellulases by function-based screening of soil metagenomic libraries. For example, 825 Mb of DNA had to be screened to identify one cellulase gene from compost soil (Pang et al. 2009; Duan and Feng 2010). Thus, it is not surprising that currently most known soil-derived cellulases and the corresponding genes originate from cultured microorganisms such as *Cellvibrio mixtus* or *Clostridium thermocellum* (Fontes et al. 2000; Gilad et al. 2003).

The screening of large-insert libraries, which theoretically leads to higher hit rates compared to small-insert libraries (Daniel, 2005), resulted in the identification of one cellulase-encoding gene and two xylanase-encoding genes in this study (Chapter B5). Sequence and protein domain analyses revealed that the cellulase belongs to glycoside hydrolase family 9 whereas the xylanases were new members of glycoside hydrolase family 11. Especially the predicted cellulase amino acid sequence showed low similarity to known protein sequences. This is mainly caused by a family 9 carbohydrate-binding module embedded in the amino acid sequence of the cellulase. So far, family 9 carbohydrate-binding modules were only detected in xylanases.

Biochemical characterization of the cellulase revealed remarkable salt tolerance and high activity over a wide pH range. Furthermore, the xylanases exhibited higher temperature optima (60°C) than other xylanases obtained from functional screens of metagenomes. Thus, all three enzymes possess features which are useful for industrial applications. Particularly the cellulase that can be also rapidly purified bears high potential for adaptation to industrial processes.

3 References

Acosta-Martínez, V., S. Dowd, Y. Sun, and V. Allen. 2008. Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biol Biochem **40**:2762–2770.

Arpigny, J. and K. Jaeger. 1999. Bacterial lipolytic enzymes: classification and properties. Biochem J **343**:177-183.

Baker, K., S. Langenheder, G. Nicol, D. Ricketts, K. Killham, C. Campbell, and J. Prosser. 2009. Environmental and spatial characterisation of bacterial community composition in soil to inform sampling strategies. Soil Biol Biochem **41**:2292–2298.

Barns, S., E. Cain, L. Sommerville, and C. Kuske. 2007. *Acidobacteria* phylum sequences in uranium-contaminated subsurface sediments greatly expand the known diversity within the phylum. Appl Environ Microbiol **73**:3113-3116.

Bayer, S., A. Kunert, M. Ballschmiter, and T. Greiner-Stoeffele. 2010. Indication for a new lipolytic enzyme family: isolation and characterization of two esterases from a metagenomic library. J Mol Microbiol Biotechnol **18**:181-187.

Bergmann, G., S. Bates, K. Eilers, C. Lauber, J. Caporaso, W. Walters, R. Knight, and N. Fierer. 2011. The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. Soil Biol Biochem **43**:1450–1455.

Brennan, Y., W. Callen, L. Christoffersen, P. Dupree, F. Goubet, S. Healey, M. Hernández, M. Keller, K. Li, N. Palackal, A. Sittenfeld, G. Tamayo, S. Wells, G. Hazlewood, E. Mathur, J. Short, D. Robertson, and B. Steer. 2004. Unusual microbial xylanases from insect guts. Appl Environ Microbiol **70**:3609-3617.

Chu, X., H. He, C. Guo, and B. Sun. 2008. Identification of two novel esterases from a marine metagenomic library derived from South China Sea. Appl Microbiol Biotechnol **80**:615-625.

Chu, H., N. Fierer, C. Lauber, J. Caporaso, R. Knight, and P. Grogan. 2010. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. Environ Microbiol **12**:2998-3006.

Cole, J., Q. Wang, E. Cardenas, J. Fish, B. Chai, R. Farris, A. Kulam-Syed-Mohideen, D. McGarrell, T. Marsh, G. Garrity, and J. Tiedje. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res **37**:D141-145.

Daniel, R. 2005. The metagenomics of soil. Nat Rev Microbiol **3**:470-478.

Delmont, T., P. Robe, S. Cecillon, I. Clark, F. Constancias, P. Simonet, P. Hirsch, and T. Vogel. 2011. Accessing the soil metagenome for studies fo microbial diversity. Appl Environ Microbiol **77**:1315-1324.

Demanèche, S., L. Philippot, M. David, E. Navarro, T. Vogel, and P. Simonet. 2009. Characterization of denitrification gene clusters of soil bacteria via a metagenomic approach. Appl Environ Microbiol **75**:534-537.

Dimitriu, P. and S. Grayston. 2010. Relationship between soil properties and patterns of bacterial β-diversity across reclaimed and natural boreal forest soils. Microb Ecol **59**:563-573.

Duan, C. and J. Feng. 2010. Mining metagenomes for novel cellulase genes. Biotechnol Lett **32**:1765-1775.

Elshahed, M., N. Youssef, A. Spain, C. Sheik, F. Najar, L. Sukharnikov, B. Roe, J. Davis, P. Schloss, V. Bailey, and L. Krumholz. 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl Environ Microbiol **74**:5422-5428.

Engelbrektson, A., V. Kunin, K. Wrighton, N. Zvenigorodsky, F. Chen, H. Ochman, and P. Hugenholtz. 2010. Experimental factors affecting PCR-based estimates of microbial species richness and evenness. ISME J **4**:642-647.

Fang, Z., W. Fang, J. Liu, Y. Hong, H. Peng, X. Zhang, B. Sun, and Y. Xiao. 2010. Cloning and characterization of a β-glucosidase from marine microbial metagenome with excellent glucose tolerance. J Microbiol Biotechnol **20**:1351-1358.

Fang, Z., T. Li, Q. Wang, X. Zhang, H. Peng, W. Fang, Y. Hong, H. Ge, and Y. Xiao. 2011. A bacterial laccase from marine microbial metagenome exhibiting chloride tolerance and dye decolorization ability. Appl Microbiol Biotechnol **89**:1103-1110.

Fierer, N., J. Schimel, and P. Holden. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol Biochem **35**:167–176.

Fierer, N. and R. Jackson. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103**:626-631.

Fierer, N., M. Breitbart, J. Nulton, P. Salamon, C. Lozupone, R. Jones, M. Robeson, R. Edwards, B. Felts, S. Rayhawk, R. Knight, F. Rohwer, and R. Jackson. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of Bacteria, Archaea, Fungi, and Viruses in soil. Appl Environ Microbiol **73**:7059-7066.

Fontes, C., H. Gilbert, G. Hazlewood, J. Clarke, J. Prates, V. McKie, T. Nagy, T. Fernandes, and L. Ferreira. 2000. A novel *Cellvibrio mixtus* family 10 xylanase that is both intracellular and expressed under non-inducing conditions. Microbiology **146**:1959-1967.

Gilad, R., L. Rabinovich, S. Yaron, E. Bayer, R. Lamed, H. Gilbert, and Y. Shoham. 2003. CelI, a noncellulosomal family 9 enzyme from *Clostridium thermocellum*, is a processive endoglucanase that degrades crystalline cellulose. J Bacteriol **185**:391-398.

Glogauer, A., V. Martini, H. Faoro, G. Couto, M. Müller-Santos, R. Monteiro, D. Mitchell, E. de Souza, F. Pedrosa, and N. Krieger. 2011. Identification and characterization of a new true lipase isolated through metagenomic approach. Microb Cell Fact doi:10.1186/1475-2859-10-54.

Griffiths, R., B. Thomson, P. James, T. Bell, M. Bailey, and A. Whiteley. 2011. The bacterial biogeography of British soils. Environ Microbiol **13**:1642-1654.

Guerrero, R. 2001. Bergey's manuals and the classification of prokaryotes. Int Microbiol **4**:103-109.

Hansel, C., S. Fendorf, P. Jardine, and C. Francis. 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74**:1620-1633.

Hartman, W., C. Richardson, R. Vilgalys, G. Bruland. 2008. Environmental and anthropogenic controls over bacterial communities in wetland soils. Proc Natl Acad Sci U S A **105**:17842-17847.

Heath, C., X. Hu, S. Cary, and D. Cowan. 2009. Identification of a novel alkaliphilic esterase active at low temperatures by screening a metagenomic library from antarctic desert soil. Appl Environ Microbiol **75**:4657-4659.

Henne, A., R. Schmitz, M. Bömeke, G. Gottschalk, and R. Daniel. 2000. Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on *Escherichia coli*. Appl Environ Microbiol **66**:3113-3116.

Hu, Y., C. Fu, Y. Huang, Y. Yin, G. Cheng, F. Lei, N. Lu, J. Li, E. Ashforth, L. Zhang, and B. Zhu. 2010. Novel lipolytic genes from the microbial metagenomic library of the South China Sea marine sediment. FEMS Microbiol Ecol **72**:228-237.

Janssen, P. 2006. Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol **72**:1719-1728.

Jones, R., M. Robeson, C. Lauber, M. Hamady, R. Knight, and N. Fierer. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J **3**:442-453.

Kim, E., K. Oh, M. Lee, C. Kang, T. Oh, and J. Yoon. 2009. Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. Appl Environ Microbiol **75**:257-260.

Lauber, C., M. Strickland, M. Bradford, and N. Fierer. 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biol Biochem **40**:2407–2415.

Lauber, C., M. Hamady, R. Knight, and N. Fierer. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl Environ Microbiol **75**:5111-5120.

Lee, C., R. Kibblewhite-Accinelli, K. Wagschal, G. Robertson, and D. Wong. 2006a. Cloning and characterization of a cold-active xylanase from an environmental DNA library. Extremophiles **10**:295-300.

Lee, M., C. Lee, T. Oh, J. Song, and J. Yoon. 2006b. Isolation and characterization of a novel lipase from a metagenomic library of tidal flat sediments: evidence for a new family of bacterial lipases. Appl Environ Microbiol **72**:7406-7409.

Lee, M., K. Hong, S. Malhotra, J. Park, E. Hwang, H. Choi, Y. Kim, W. Tao, and S. Lee. 2010. A new esterase EstD2 isolated from plant rhizosphere soil metagenome. Appl Microbiol Biotechnol **88**:1125-1134.

López-García, P. and D. Moreira. 2008. Tracking microbial biodiversity through molecular and genomic ecology. Res Microbiol **159**:67-73.

Morales, S., T. Cosart, J. Johnson, and W. Holben. 2009. Extensive phylogenetic analysis of a soil bacterial community illustrates extreme taxon evenness and the effects of amplicon length, degree of coverage, and DNA fractionation on classification and ecological parameters. Appl Environ Microbiol **75**:668-675.

Nüsslein, K. and J. Tiedje. 1999. Soil bacterial community shift correlated with change from forest to pasture vegetation in a tropical soil. Appl Environ Microbiol **65**:3622-3626.

Pang, H., P. Zhang, C. Duan, X. Mo, J. Tang, and J. Feng. 2009. Identification of cellulase genes from the metagenomes of compost soils and functional characterization of one novel endoglucanase. Curr Microbiol **58**:404-408.

Park, S., H. Shin, and G. Kim. 2011. Screening and identification of a novel esterase EstPE from a metagenomic DNA library. J Microbiol **49**:7-14.

Quince, C., A. Lanzén, T. Curtis, R. Davenport, N. Hall, I. Head, L. Read, and W. Sloan. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Meth **6**:639-641.

Reeder, J. and R. Knight. 2010. Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. Nat Meth **7**:668-669.

Roesch, L., R. Fulthorpe, A. Riva, G. Casella, A. Hadwin, A. Kent, S. Daroub, F. Camargo, W. Farmerie, and E. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J **1**:283-290.

Sang, S., G. Li, X. Hu, and Y. Liu. 2011. Molecular cloning, overexpression and characterization of a novel feruloyl esterase from a soil metagenomic library. J Mol Microbiol Biotechnol **20**:196-203.

Simon, C. and R. Daniel. 2009. Achievements and new knowledge unraveled by metagenomic approaches. Appl Microbiol Biotechnol **85**:265-276.

Simon, C. and R. Daniel. 2011. Metagenomic analyses: past and future trends. Appl Environ Microbiol **77**:1153-1161.

Steele, H., K. Jaeger, R. Daniel, and W. Streit. 2009. Advances in recovery of novel biocatalysts from metagenomes. J Mol Microbiol Biotechnol **16**:25-37.

Tang, K., R. Kobayashi, V. Champreda, L. Eurwilaichitr, and S. Tanapongpipat. 2008. Isolation and characterization of a novel thermostable neopullulanase-like enzyme from a hot spring in Thailand. Biosci Biotechnol Biochem **72**:1448-1456.

Torsvik, V., J. Goksoyr, and F. Daae. 1990. High diversity in DNA of soil bacteria. Appl Environ Microbiol **56**:782-787.

Tringe, S., C. von Mering, A. Kobayashi, A. Salamov, K. Chen, H. Chang, M. Podar, J. Short, E. Mathur, J. Detter, P. Bork, P. Hugenholtz, and E. Rubin. 2005. Comparative metagenomics of microbial communities. Science **308**:554-557.

Uchiyama, T. and K. Miyazaki. 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. Curr Opin Biotechnol **20**:616-622.

Uroz, S., M. Buée, C. Murat, P. Frey-Klett, and F. Martin. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environ Microbiol Rep **2**:281-288.

Wei, P., L. Bai, W. Song, and G. Hao. 2009. Characterization of two soil metagenome-derived lipases with high specificity for p-nitrophenyl palmitate. Arch Microbiol **191**:233-240.

Xu, M., X. Xiao, and F. Wang. 2008. Isolation and characterization of alkane hydroxylases from a metagenomic library of Pacific deep-sea sediment. Extremophiles **12**:255-262.

D Summary

Soil probably harbors the highest microbial species diversity of any environment on Earth and represents a major reservoir of microbial taxonomic, genomic, and metabolic diversity. Bacteria are the most abundant group within this largely unexplored habitat. The most promising approach to gain insights into diversity and structure of soil microbial communities is the application of cultivation-independent techniques. In addition, these techniques are also valuable tools for the recovery of novel natural products from soils. In this study, forest and grassland soil samples derived from the German Biodiversity Exploratories Schorfheide-Chorin, Hainich-Dün, and Schwäbische Alb were analyzed by applying metagenomic approaches.

Environmental DNA was isolated from topsoil and subsoil samples covering different management types. To assess taxonomic composition, the V2-V3 region of the 16S rRNA gene region was amplified from the isolated DNA by PCR. Subsequently, pyrosequencing and analysis of the amplicons were performed. The pyrosequencing-derived datasets from 18 Schwäbische Alb soil samples (nine forest and nine grassland topsoil samples) and 18 Hainich soil samples (nine topsoil and nine subsoil grassland samples) comprised 599,284 and 752,838 sequences, respectively. OTUs (operational taxonomic units) were determined for each sample and subsequently diversity and richness were evaluated. Bacterial diversity was higher in grassland soils than in forest soils on the phylum level. In addition, a soil depth effect was recorded; topsoil samples exhibited a higher bacterial diversity than subsoil samples. In forest soils, which covered a wide pH range, diversity was lowest in the most acidic soil sample. In addition, a tree species effect on bacterial diversity was indicated. The relative abundances of different phyla, proteobacterial classes, and genera such as *Actinobacteria*, *Alphaproteobacteria*, and *Streptomyces* showed significant variations between forest and grassland soils. Furthermore, bacterial groups at different taxonomic levels showed strong correlations to soil pH. This pH effect was not found in Hainich grassland soil samples, because of the small pH range covered by these samples. Instead of pH, organic carbon content induced statistically significant effects on bacterial community structure at the near-neutral pH of the Hainich soil samples.

Moreover, 14 small-insert and nine large-insert libraries were constructed from environmental DNA of all three German Biodiversity Exploratories. The plasmid libraries

comprised approximately 40,000 to 511,000 clones with average insert sizes of 3 to 9 kb whereas the fosmid libraries contained 4,600 to 300,000 clones with average insert sizes of 19 to 30 kb. Partial function-based screening of the metagenomic libraries for lipolytic and (hemi)cellulolytic genes resulted in the identification of 37 unique lipolytic and three individual (hemi)cellulolytic clones. Sequencing and subsequent analysis of insert DNA of these clones was carried out. Thirty-five gene products of the 37 identified lipolytic genes were new members of families I (true lipases), IV, V, VI, and VIII of lipolytic enzymes. The remaining two gene products represented putatively new families of lipolytic enzymes.

Insert DNA of two of the three (hemi)cellulolytic clones harbored xylanase-encoding genes whereas insert DNA of the remaining clone contained a cellulase-encoding gene. Amino acid sequence analysis of the gene products revealed that the putative xylanases belong to glycoside hydrolase family 11. The cellulase could be assigned to glycoside hydrolase family 9. Furthermore, the cellulase harbored a family 9 carbohydrate-binding module which was so far only detected in xylanases. Initial characterization of the two xylanases showed that both enzymes exhibit high activity over a wide range of temperatures and pH values. In addition, biochemical characterization of the purified cellulase was performed. The enzyme was highly active over a wide pH range and showed high halotolerance. Thus, a high diversity of novel biocatalysts with valuable properties for industrial applications was discovered from soil. Furthermore, the phylogenetic analysis combined with statistical analysis revealed that soil bacterial diversity and community structure is to some extent non-randomly.

Acknowledgements

First of all, I would like to express my gratitude to PD Dr. Rolf Daniel for giving me the opportunity to work on this challenging topic as well as for supervision and constant support during my thesis. Many thanks for critical reading of several manuscripts and for giving me the opportunity to attend to national and international conferences.

I am grateful to Prof. Dr. Wolfgang Liebl for his readiness to be the second examiner of this thesis.

Special thanks to Birgit, Christiane, and Silja for the thorough proofreading of this thesis. I want to thank Christiane, especially for relieving me of my workload within the Biodiversity Exploratories during the past weeks. Thanks to Sarah, Martin, Janine, and Ines who were involved in the performance of different experiments. Additionally, I thank Andrea, Steffi, and Frauke for the generation of pyrosequencing data. I am thankful to Antje and Sascha who helped me to preprocess my pyrosequencing data for subsequent analyses.

I was very lucky to be surrounded by great people in the lab. Many Thanks to all my current and former labmates, Andrea, Angel, Bernd, Birgit, Boris, Caro, Carola, Christiane, Christina, Dominik, Ines, Jenny, Jörg, Jörn, Juli, Julia, Kristin, Leo, Marco, Marie, Markus, Martin, Mechthild, Michael, Milad, Robert, Romano, Sarah, Silja, Simon, Steffi (thank you for endless running sessions!), and Tanja for creating a pleasant atmosphere in the lab and helping me throughout my work.

Many thanks to Birgit, Carola, and Prof. Dr. Gottschalk for sharing a wonderful time in South Africa. I will never forget the Kruger park safari tour and really enjoyed our well-mixed group.

I highly appreciate the cooperation with Ingo, Ladislav, and Nadine from the Biodiversity Exploratories. Thanks for providing soil parameters and assistance in statistical analyses. In addition, I thank the BEO- and the BExIS-team for support, project administration, and data management. It was a big pleasure for me to participate in the soil

sampling campaigns 2008 and 2011. Sometimes privacy was rare and sampling rough-running, but altogether I enjoyed the temporary change from lab work to sunny field work.

Thanks to Andreas, Christian, Damion, Jan, Julian, Markus, Marvin, Robert, and Thomas for friendship and diverse activities in Göttingen and Bad Pyrmont.

I am extremely grateful to my family, in particular to my parents, for their everlasting support, encouragement, and for always offering a refuge from work when necessary.

Finally, I want to thank Katharina for always being there for me at every step of the way.

Curriculum vitae

Name:	Heiko Nacke
Date of birth:	6th May, 1982
Place of birth:	Bad Pyrmont, Germany

Since 04/2008	PhD thesis at the Institute of Microbiology and Genetics, Georg-August University Göttingen
06/2007-02/2008	Diploma thesis: „Charakterisierung von transformationsdefekten Mutanten in *Thermus thermophilus*“
10/2002-05/2007	Study of Biology (Diploma) at the Georg-August University Göttingen; major subject: Microbiology, minor subjects: Organic Chemistry and Biochemistry
07/2001-03/2002	Military service at the 4./Sanitätsregiment 1, Hildesheim
06/2001	Abitur at the Städtisches Gymnasium, Barntrup