LIJUAN YAN

Bacterial Community Dynamics and Perennial Crop Growth in Motor Oil-Contaminated Soil in a Boreal Climate

DEPARTMENT OF ENVIRONMENTAL SCIENCES
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES
DOCTORAL PROGRAMME IN INTERDISCIPLINARY ENVIRONMENTAL SCIENCES
UNIVERSITY OF HELSINKI

Department of Environmental Sciences

Faculty of Biological and Environmental Sciences

University of Helsinki

Finland

Bacterial community dynamics and perennial crop growth in motor oil-contaminated soil in a boreal climate

Lijuan Yan

Academic dissertation

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in auditorium 1041, Biocenter 2, Viikinkaari 5, on June 10th at 12 noon.

Helsinki 2016

Supervisor:	Professor Kristina Lindström
	Department of Environmental Sciences, University of Helsinki
Co-supervisors:	Professor Frederick L. Stoddard
	Department of Food and Environmental Sciences, University of Helsinki
	Department of Agricultural Sciences, University of Helsinki
	Doctor Petri Penttinen
	Department of Environmental Sciences, University of Helsinki
	Doctor Anu Mikkonen
	Department of Biological and Environmental Science, University of Jyväskylä
Advisory committee:	Docent Leena Suominen
	Department of Food and Environmental Sciences, University of Helsinki
	Professor Martin Romantschuk
	Department of Environmental Sciences, University of Helsinki
Pre-examiners:	Associate Professor Ole Nybroe
	Department of Plant and Environmental Sciences, University of Copenhagen
	Doctor Göran Bergkvist
	Department of Crop Production Ecology, Swedish University of Agricultural Sciences
Opponent:	Docent Kirsten S. Jørgensen
	Finnish Environment Institute
Custos:	Professor Kristina Lindström

To my family

TABLE OF CONTENTS

A LIST OF ORIGINAL PUBLICATIONS

I. Lijuan Yan, Petri Penttinen, Asko Simojoki, Frederick L. Stoddard, & Kristina Lindström (2015). Perennial crop growth in oil-contaminated soil in a boreal climate. Science of the Total Environment, 532, 752-761.

II. Lijuan Yan, Hanna Sinkko, Petri Penttinen, & Kristina Lindström (2016). Characterization of successional changes in bacterial community composition during bioremediation of used motor oil-contaminated soil in a boreal climate. Science of the Total Environment, 542, 817-825.

III. Lijuan Yan, Petri Penttinen, Anu Mikkonen & Kristina Lindström. Bacterial community dynamics during multi-year bioremediation of oil-contaminated soil in a boreal climate. Manuscript, Submitted.

Papers I-II are reproduced with the permission of the original publisher.

The contribution of the author to the publications:

Paper I   Lijuan Yan took the soil and crop samples, performed the chemical analysis of soil and crop samples, did the statistical analysis, interpreted the results, wrote the manuscript and was the corresponding author.

Paper II   Lijuan Yan performed the genomic DNA extraction and LH-PCR fingerprinting analyses. Lijuan Yan performed the statistical analysis with Hanna Sinkko, interpreted the results, wrote the manuscript and was the corresponding author.

Paper III   Lijuan Yan performed the PCR, statistical analysis, interpreted the results, wrote the manuscript and was the corresponding author.

ABSTRACT

Soil pollution by petroleum hydrocarbons (PHCs) as a result of anthropogenic activities poses significant threats in the environment. In particular, used motor oil that contains high concentrations of aliphatics, polycyclic aromatic hydrocarbons (PAHs) and heavy metals (e.g. lead, zinc, chromium, barium and arsenic) contribute to chronic hazards including carcinogenicity. Microorganisms are able to degrade and utilize many recalcitrant compounds as carbon and energy sources in a natural attenuation process. However, in boreal regions this process is limited by the cool climate.

The main goal of most bioremediation designs should be an optimization of environmental conditions for microbial growth and metabolic activities. Plant growth can stimulate the activities of soil microflora in the rhizosphere, thus enhancing the bioremediation of oil-polluted soil. Nitrogen deficiency is a frequent limiting factor of biomediation in oil-contaminated soils. Legumes that form symbiotic association with N-fixing bacteria are able to assist the biodegradation of PHCs. The planting of oil-tolerant perennial crops, especially legumes, in oil-contaminated soil holds promise for great economic benefits for bioenergy production while accelerating the oil degradation process. Fodder galega (*Galega orientalis* Lam.), a perennial forage legume, and smooth brome (*Bromus inermis* L.), a cool-season perennial sod-forming grass, are both persistent in boreal zones and have been shown to grow well together in crop mixtures without N-fertilizer supply. The oil tolerance and oil-rhizoremediation potential of *G. orientalis* and its microsymbiont *Neorhizobium galegae* have been demonstrated at microcosm and mesocosm scales. Plant growth promoting bacteria (PGPB) have potential to increase nodulation of galega, mitigate plant stress response and increase the bioavailability of soil contaminants, therefore enhancing the degradation of contaminants. These components can form a powerful combination to be used for bioremediation of oil-contaminated soil.

However, the competitiveness and effectiveness of the crop- and PGPB-assisted bioremediation system need to be evaluated in field conditions. To date, there were no systematically described studies on bioremediation of oil-contaminated soil combined with crop biomass production in boreal regions. This multidisciplinary research project was conducted to fill this knowledge gap by evaluating the sustainability of the legume-cropping bioremediation system economically in terms of crop yield, and environmentally in terms of oil degradation rate and the dynamics of bacterial communities. To reach these aims, we established a multi-year bioremediation field experiment at the Viikki Experimental Farm, University of Helsinki, Finland (60°14'N, 25°01'E, 8 m AMSL) with crop treatments (brome grass, fodder galega, their mixture and bare fallow) as the main plots in four replicated blocks, and used motor oil treatments (7000 ppm +/-) and PGPB (+/-) treatments as the sub-plot factors.

Soil samples were taken from the top 20 cm layer at six time points (July 2009, May 2010, November 2010, May 2011, May 2012 and October 2012). Soil chemical properties e.g. pH, electrical conductivity (EC), total C, total N and C:N ratio of three sample sets (July 2009, November 2010 and May 2012) were measured. Oil concentration was determined based on the difference of total solvent extractable material (TSEM) concentration between the oil-spiked plot and the average of control plots at each sampling time using the gravimetrical method. Crop physiological properties e.g. annual DM yield, total C, total N, C: N ratio, chlorophyll and BNF of the legume were measured. Soil-borne bacterial communities were investigated using i) LH-PCR community fingerprinting technique (all samples) and ii) Illumina's MiSeq sequencing (spring-summer samples).

Oil contamination had a significant impact on soil chemical properties, e.g. pH, EC, total C and C:N ratio. The oil degradation was incomplete 40 months after the oil spike, with a dissipation of 73% - 92% of oil concentration (Paper I). As the field soil condition was good for oil degradation, the advantage of using crops to assist oil degradation was not evident. The oil degradation followed first-order kinetics with the reduction rates decreasing as follows: bare fallow > galega-brome grass mixture > brome grass > galega. Oil, surprisingly, increased crop dry matter and nitrogen yield, particularly in the fourth year (Paper I). The legume-grass mixture produced significantly higher crop dry biomass than the pure stands. For instance, the unfertilized galega-brome grass mixture out-yielded the N-fertilized pure grass swards over years by an average of 32% (Paper I), suggesting that the inoculated galega could fully replace N-fertilizer for brome grass. PGPB enhanced the efficiency of biological nitrogen fixation of the legume, especially in legume-grass mixture plots (Paper I). The LH-PCR community fingerprinting technique produced similar results as the 16S rRNA gene amplicon sequencing. Both time and oil contamination were the main drivers of bacterial community dynamics (Papers II and III). The effect of oil was initially negative on overall bacterial diversity (Papers II and III), but variable on the diversity of bacterial sub-communities (Paper III). The bacterial communities responded quickly to oil contamination, but the effect of oil on community composition was recoverable over time (Papers II and III). Crop cultivation had a small impact on the composition of bacterial community (Paper III). The oil-favored taxa that discriminated bacterial communities between oil-contaminated and non-contaminated soils were mainly assigned to the two prevalent phyla *Actinobacteria* and *Proteobacteria* (Paper III). The operational taxonomic units (OTUs) with significantly different oil-specific abundance changes over time were all favored by oil; therefore, these oil-specific taxa were suggested as suitable bio-indicators to monitor the ecological impact of oil contamination (Paper III). Besides oil concentration, the changes in soil chemical properties, e.g. soil pH and EC, significantly affected bacterial community structure (Paper II and III).

To summarize, oil contamination affected soil chemical and biological properties (e.g. crop growth and bacterial community), but the impact of oil decreased with time. The cultivation of oil-tolerant perennial crops, especially galega-brome grass mixture, in oil-contaminated soil can hopefully produce considerable biomass for bioenergy industry. Bacterial communities underwent a significant time- and season-dependent succession, regardless of oil contamination. Therefore, studies restricted to a single snapshot of time without any non-contaminated samples as reference cannot reveal oil contamination-related changes in the dynamic patterns of bacterial communities in the field soil. With the development and decreasing cost of high-throughput sequencing (NGS), NGS-based metagenomics analysis has become the mainstream method in microbial ecology research, providing in-depth view on bacterial populations at different taxonomic levels in the community. However, the LH-PCR technique is still suggested as a cost-effective method to monitor microbial community dynamics for assessing the ecological impact of oil contamination. Oil degradation was rather slow in the boreal climate. Long-term stimulation and monitoring of soil chemical properties, oil concentration, crop growth and microbial community are still needed for risk control. All these suggestions can be applied to soil contaminated by PHC contaminants other than used motor oil, despite hydrocarbon compositional differences.

ABBREVIATIONS

ANOVA analysis of variance

BNF biological nitrogen fixation

CAP canonical analysis of principle coordinates

db-DA distance-based discriminant analysis

db-RDA distance-based redundancy analysis

DISTLM distance-based linear models

DM dry matter

DNA deoxyribonucleic acid

LH-PCR length heterogeneity polymerase chain reaction

nMDS non-metric multidimensional scaling

OTU operational taxonomic unit

PAH polycyclic aromatic hydrocarbon

PCA principal component analysis

PCO principle coordinate ordination

PCR polymerase chain reaction

PERMANOVA permutational univariate or multivariate analysis of variance

PERMDISP test of homogeneity of dispersions by permutations

PGPB plant growth promoting bacteria

RM ANOVA repeated measures analysis of variance

SOM soil organic matter

TPH total petroleum hydrocarbon

T-RFLP terminal restriction fragment length polymorphism

TSEM total solvent extractable material

UCM unresolved complex mixture

UV ANOVA univariate analysis of variance

%Ndfa proportion of legume shoot N derived from atmospheric N_2

1. INTRODUCTION

1.1. Soil contamination by petroleum hydrocarbons in Europe

"*The coming 50 years are likely to be the final period of rapidly expanding, global human environmental impacts*" (Tilman *et al.* 2002).

Soil pollution by petroleum hydrocarbons (PHCs) has been increasing considerably as a result of anthropogenic activities. These contaminants induce serious ecotoxicological impacts to the environment, depending on the types and bioavailability of these hydrocarbon compounds. PHC mixtures, including crude oil, diesel fuel and creosote, comprise different concentrations of *n-* and branched alkanes, cycloalkanes, phenolics, aromatics and polycyclic aromatic hydrocarbons (PAHs) (Hamamura *et al.* 2006). They are composed of two categories: 1) gasoline range organics refers to small chain alkanes (<C10), e.g. methane, ethane, propane, volatile aromatic compounds (e.g. benzene, toluene, ethylbenzene, and xylenes BTEX) and common oxygenates (e.g. Methyl Tertiary Butyl Ether MTBE and ethanol); and 2) diesel range organics are the alkanes with longer chains (C10-C40) or non-halogenated semi-volatile organic compounds and hydrophobic chemicals such as PAHs (e.g. naphthalene, phenanthrene, anthracene, benzo[a]pyrene) (Kamath *et al.* 2004). In Europe, petroleum hydrocarbons and hydrocarbon-related organic materials, e.g. mineral oil, BTEX, PAHs, chlorinated hydrocarbons (CHCs) and phenols, account for 54.5% of the total soil contaminants (Fig. 1). In Europe, industrial and commercial activities and waste disposal and treatment are the main sources of oil pollution (Fig. 1).
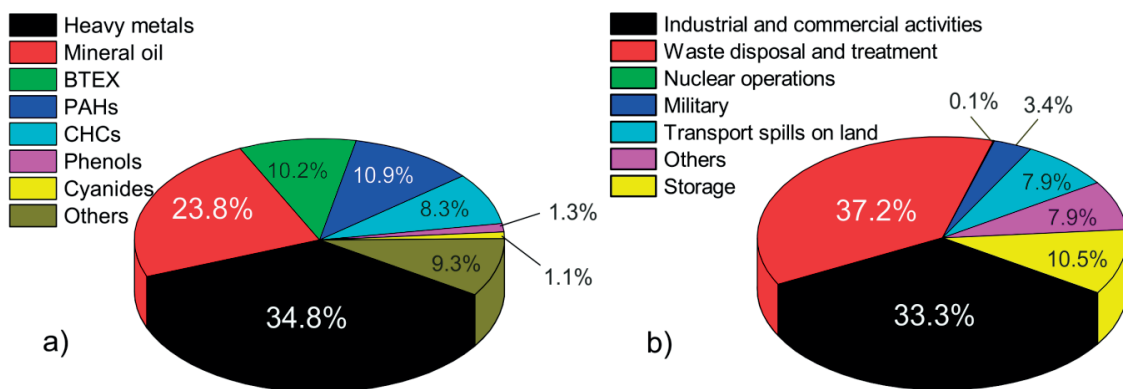
Fig. 1. Soil contamination in Europe (summarized as average over 22 EEA countries/regions): a) contaminants affecting soil matrix (soil, sludge, sediment), as reported in 2011 and b) a breakdown of the main sources causing soil contamination in Europe as % of sources over the total number of sources identified. Both figures are redrawn from EEA (2014). Abbreviations: "*CHCs*" chlorinated hydrocarbons, "*PAHs*" polycyclic aromatic hydrocarbons and "*BTEX*" monoaromatic hydrocarbons (benzene, toluene, ethyl benzene and xylene).

In Finland, soil contamination has been investigated since the 1980s (Pyy *et al.* 2013). The service sector and the production sector contributed 74.1% and 25.5% to soil contamination, respectively (EEA 2014). The number of identified contaminated sites was 19.9 per 1000 capita, whereas potentially contaminated sites were estimated to 35.5 per 1000 capita (EEA 2014).

1.2. Ecological impacts of hydrocarbon contamination

Prior to the assessment of the ecological impacts, it is important to understand the fate of hydrocarbon contaminants in soil. Once released to soil, the PHC mixture may separate into individual compounds, depending on their chemical properties. Hydrocarbons of lower molecular weight, e.g. monoaromatic BTEX compounds, are highly mobile in the environment and tend to volatilize, according to the type of hydrocarbon, temperature and wind conditions; those of higher molecular weight tend to leach to the groundwater (Kamath *et al.* 2004; Gawel & North Dakota Industrial Commission 2006). Hydrocarbons that were removed through evaporation were generally assumed to account for 30% of the total hydrocarbon content upon spills (Bragg *et al.* 1994). Soil texture and moisture content together determine how fast and how deep the hydrocarbons leach into the soil profile (Gawel & North Dakota Industrial Commission 2006). The volume of the oil spills also has an impact on the migration of the PHCs in soil and further to the ground water (Kamath *et al.* 2004). These hazards pose significant threats to the environment, resulting in serious acute or chronic ecotoxicological effects, depending on the accessibility and bioavailability of these hydrocarbon compounds. Extensive studies have revealed that hydrocarbons, especially PAHs, undergo metabolic activation that induces toxicity, mutagenicity or carcinogenicity in mammalian systems (Chaudhry 1994). Due to the mutagenic and carcinogenic properties that pose a risk to human health, 16 PAHs were listed as priority pollutants by the United States Environmental Protection Agency (Chaudhry 1994).

1.2.1. Soil quality

Soil quality is dependent on the interactions between soil physical, chemical and biological properties (Dexter 2004). Oil contamination has direct impact on soil quality in terms of soil physical and chemical properties. Oil spills impact soil physical structure by coating soil aggregates, affecting soil water holding capacity, reducing and diverting water infiltration into the soil and obstructing air and water movement in the soil matrix (Gawel & North Dakota Industrial Commission 2006). Oil contamination affects soil chemical properties by reducing ion exchange on soil aggregates (Gawel & North Dakota Industrial Commission 2006). The changed soil chemical properties following oil contamination therefore lead to indirect changes in soil biological properties in terms of the growth and activities of soil organisms and plants.

1.2.2. Plants

Plants represent the basal component of most ecosystems as primary producers (Loreau *et al.* 2001). Soil contamination by petroleum hydrocarbons results in indirect and direct impacts on plants. However, the phytotoxicity of hydrocarbons is dependent on the specific contaminant composition, soil quality and the plant species (Gawel & North Dakota Industrial Commission 2006; Kamath *et al.* 2004). At low concentrations, the impact of hydrocarbons on plants is due to the physical impact on soil structure (Gawel & North Dakota Industrial Commission 2006). Certain annual plants were able to survive in weakly to moderately contaminated sites with oil content below 10% by weight in soil (Radwan *et al.* 1995). Nonetheless, hydrocarbon concentrations over 4% are directly toxic to sensitive plant species, and those over 7% cause direct phytotoxicity to most plants (Gawel & North Dakota Industrial Commission 2006). Major adverse effects of hydrocarbons are typically shown as decreased germination and growth, if contaminant concentrations are high enough (Kamath *et al.* 2004). Particularly, PAHs can impact plant morphologies. For instance, *Arabidopsis* exhibited a series of morphological symptoms of PAH stress, including reduced growth of the root and shoot, deformed

trichomes, reduced root hairs, chlorosis, late flowering, and the appearance of white spots, which later developed into necrotic lesions; at the tissue and cellular levels, the detoxification of PAHs in plant cells resulted in oxidative stress mechanisms, including the production of hydrogen peroxide (H_2O_2) and cell death (Alkio *et al.* 2005).

1.2.3. Microorganisms

Microorganisms are the key players in ecosystem processes by driving the biogeochemical cycling of elements (Zhang *et al.* 2007). Microorganisms, due to their small size, are in direct communication with their environment (McArthur 2006). The cytoplasmic membrane is fundamental to the survival and reproductive success of organisms by controlling the exchange of materials between the living environment and the organisms. Factors that affect the membrane are crucial for the fitness of the organisms. Petroleum hydrocarbon mixture contains a large proportion of lipophilic compounds, such as cycloalkanes, alkanes and phenols; the accumulation of these lipophilic compounds have direct toxic effects on microbial cells by affecting the lipid part of the membrane and the proteins embedded in the membrane (Sikkema *et al.* 1995). However, microorganisms that can degrade a variety of hydrocarbon compounds widely exist in soil. The abundance and diversity of genes involved in hydrocarbon degradation were commonly found in pristine soils, providing a basis for fast adaptation of the indigenous microbial communities following oil spills (Kloos *et al.* 2006). Microbial populations adapt sensitively to changing environmental conditions owing to features such as varying individual activity, increasing reproduction of species with favorable abilities, and spreading new capabilities via horizontal gene transfer (Winding *et al.* 2005). Microorganisms (bacteria and fungi) generally show higher sensitivity in response to environmental changes, due to their high surface/volume ratio (Winding *et al.* 2005). Among them, bacteria are the most abundant microbial group in oil-contaminated soil (Dominguez-Rosado *et al.* 2004).

Microorganisms exhibit different intrinsic growth rates in response to oil disturbances. The responses of a microorganism to environmental disturbances are related to its survival strategies. The *r*- and *K*-traits of microorganisms and their associated environment are important to be considered in population biology. The *r*-*K* scheme assumes that evolutionary selection favors adaptation either to rapid (*r* strategists) or slower development (*K* strategists), in terms of resource availability (Pianka 1970). The concept of generalists and specialists is similar, yet not the same as that of the *r*- and *K*- strategists, and is also used to describe the traits of bacterial taxa in response to oil contamination. Generalists are described as organisms that can use many different resources at somewhat similar efficiencies and specialists are restricted to a few resources (McArthur 2006). In certain situations, specialists will outcompete generalists because of their ability to use the newly available resource more efficiently; outside those situations, generalists out-compete specialists because of their ability to use more than one resources (McArthur 2006). Increased catabolic capacity and a change from specialist to generalist strategy of bacterial communities were reported during the course of secondary succession following oil pollution (Mukherjee *et al.* 2015), where secondary succession is defined as a pattern of alternations in the microbial community composition after a radical disturbance in the physical environment (Mukherjee *et al.* 2013).

According to the review by McArthur (2006), soil was an important determinant, defining the specific microorganisms that can respond to hydrocarbon contaminants. Of all the communities and habitat types, the soil environment is perhaps the most complex and difficult to observe, measure, and

understand. Soils are extremely heterogeneous matrices that vary over very small to very large scales. The size of particles, amount of vegetation, types of vegetation, root mass, and amount of water all affect the ecology of soil. Therefore, oil contamination can also indirectly influence the activities of microorganisms via changes in the physiochemical properties of their habitat in soil.

Changes in bacterial diversity

Biodiversity is defined as "the variety of life at all levels, from the level of genetic variation within and among species to the level of variation within and among ecosystems and biomes" (Tilman 1997). The ability to measure bacterial diversity is a prerequisite for the systematic study of bacterial biogeography and community assembly (Curtis *et al.* 2002). However, the absolute diversity of a naturally occurring prokaryotic community is widely recognized as unknown and unknowable at any scale in any environment (Curtis *et al.* 2002; McArthur 2006). The simplest estimate of species diversity is "species richness", the number of species per unit area (McArthur 2006). The order of magnitude for the total number of bacterial and archaeal species on Earth was estimated to range from 10^4 to 10^{12} (Yarza *et al.* 2014). Nevertheless, species richness fails to convey information on the abundance of each species in an area (McArthur 2006). The estimation of species diversity (e.g. Shannon-Weiner index, Simpson's index) incorporates the abundance of each species (evenness) and assumes that all species from a large community are known (McCune *et al.* 2002; McArthur 2006). One gram of soil contains more than 10^{10} bacterial cells using direct numeration with fluorescence microscopy (Torsvik *et al.* 1996). It is suggested not necessary to count every single cell or taxon to estimate the extent of microbial diversity in a sample (Curtis *et al.* 2002). The diversity estimation is challenged by shortage of reliable data sets on the relative abundance of microbial communities at any level of scale (Curtis *et al.* 2002).

Currently, the relative abundance of gene sequences provides the most comprehensive information available to estimate relative abundance of bacterial species in one sample using next generation sequencing approaches. Diversity estimates made on higher-level taxa are usually meaningless (McArthur 2006). Bacterial diversity, as estimated by phylotype richness and diversity indices, varies across different ecosystem types (Fierer & Jackson 2006). As soil bacteria are highly diverse, the composition and diversity of bacterial community vary in different geographic locations, and depend on oil contamination, soil type and physicochemical properties, as well as plant species and land management in terrestrial environments (Torsvik & Øvreås 2002; da C Jesus *et al.* 2009; Liang *et al.* 2011; Liao *et al.* 2015).

In a historical view, increasing levels of environmental stress result in a decrease in diversity, species richness and evenness, i.e. an increase in dominance (Clarke & Warwick 2001). Oil spills lead to a decrease in microbial diversity and to changes in structure and functions of bacterial communities in oil-contaminated environments (Liang *et al.* 2011; Ros *et al.* 2014). Certain ecosystem functions are maintained through more than one species, though, resulting in functional redundancy (McArthur 2006). Although alleviated by functional redundancy, the changes (decrease or increase) in microbial diversity following habitat disturbances are widely accepted to affect the ecosystem functions such as bioremediation and nutrient cycling to varying degrees.

Successional dynamics of bacterial community

 "*Organisms modify their environments, and over time, the environments modify the organisms*" (McArthur 2006).

Microbial community structures are distinctive in varied ecosystems and sensitive to disturbances or environmental alternations. Differences in microbial community composition due to disturbances may lead to functional dissimilarities. The composition and structure of hydrocarbons in oil contaminants affect the microbial functional gene composition (Liang *et al.* 2011). Allison & Martiny (2008) proposed a simple schematic of how disturbance can change microbial composition and thereby affect ecosystem processes in terms of microbial resistance, resilience and functional redundancy. There are three potential scenarios of a disturbance on microbial composition and/or ecosystem processes: i) if the microbial community is resistant to the disturbance, its composition stays the same; ii) if the microbial community composition is sensitive and does change, it can be resilient and quickly return to the original composition; and iii) if the community is sensitive and not resilient, it might either produce ecosystem process rates similar to the original community as a result of functional redundancy or perform differently from the original community.

The term **'Community dynamics'**, defined as "changes in community size and composition that result from various forces that control and regulate communities over time" (Head *et al.* 2006), is used to describe community development in community ecology. Community development is dependent on the physical nature of the habitat, the amount and duration of disturbances, the potential and realized species pool, the types of species interactions, and other large- and small-scale processes and events (McArthur 2006). **Succession**, as one type of community development, expresses temporal changes in communities (McArthur 2006). According to the succession theory, all activities, whether abiotic or biological, that remove or add microbial species from an area would result in micro-scale succession occurring at that site (McArthur 2006).

Microbial responses to contaminant exposure reveal a natural attenuation process as a shift or return of the *in situ* microbial communities to the baseline level if the contaminants are successfully removed or transformed to a non-toxic status by a comparison with the community structure of a similar safe or uncontaminated environment (White *et al.* 1998). Significant shifts in microbial community following oil-contamination were widely characterized based on phylogenetic composition in diverse environmental conditions (Labbe *et al.* 2007; Yang *et al.* 2014), suggesting that the overall community response to oil contamination was, to some extent, predictable. The successional dynamics of microbial community has been suggested as a bio-indicator of oil contamination and recovery in terms of its sensitivity (White *et al.* 1998; Mikkonen *et al.* 2011b; Mukherjee *et al.* 2014).

1.3. Remediation of hydrocarbon-contaminated soils

1.3.1. The aim of remediation

The aim of remediation is to recover the contaminated site to a state consistent with original land-use (Gawel & North Dakota Industrial Commission 2006). Effectiveness and completeness are ultimate goals in a successful remediation project (White *et al.* 1998). The remediation approach should be case-specific, as many environmental variables (e.g. nature of pollutants, environmental condition and the microorganisms present) must be evaluated (Riser-Roberts 1998). Excavation is practical at sites contaminated by high concentration of pollutants that pose an environmental hazard (Riser-Roberts

1998). The popular *ex situ* physical approaches (e.g. thermal treatment, incineration and soil washing) and chemical approaches (e.g. peroxide spraying, solvent extraction and supercritical fluid oxidation) are effective to remove hydrocarbon-contaminated soil (Riser-Roberts 1998). In Finland, remediation of contaminated soil relies mostly on *ex situ* treatment of contaminated soil transferred from the initial contaminated site (Fig. 2). *Ex situ* treatment using physical and chemical approaches is more widely applied at contaminated sites than biological approaches. These techniques based on soil excavation are quick and effective by removing contaminants from the initial contaminated sites but are often costly and laborious, and can destroy the biological, and potentially also the chemical and physical properties of soil.
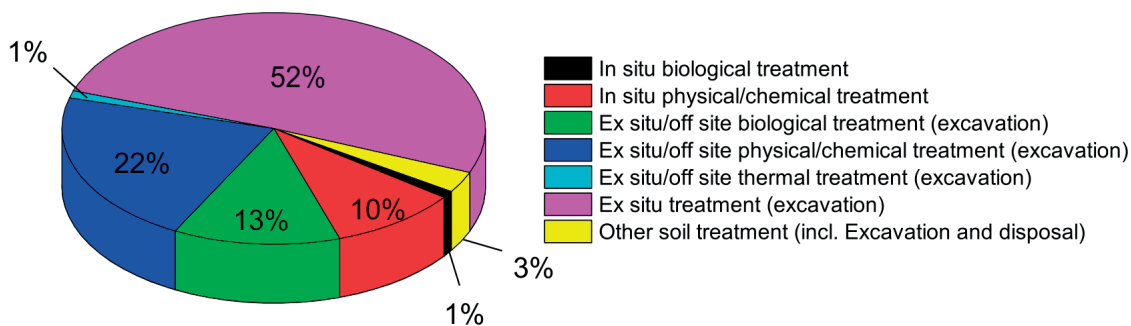

Fig. 2. Most frequently applied remediation techniques for contaminated soil in Finland, reproduced from EEA (2014).

1.3.2. Bioremediation
"*Remediation requires patience*" (Gawel & North Dakota Industrial Commission 2006)

Microbes (mainly fungi and bacteria) are able to degrade a wide variety of hydrocarbons by various pathways and mechanisms. Because hydrocarbons derived from fatty acid metabolites of plants, insects and microorganisms widely exist in natural environments (Nie *et al.* 2014), the hydrocarbon-utilizing microorganisms are ecologically widespread in diverse environments, even in pristine soil (Van Hamme *et al.* 2003; Schulz *et al.* 2012). Hydrocarbon-degrading taxa as well as the key hydrocarbon degradation genes become prevalent in oil-impacted environments, owing to natural selection resulting from the stress and newly available substrates of the oil contaminants (Head *et al.* 2006; Liang *et al.* 2011).

Generally, hydrocarbons with straight and short chains are degraded more readily than those with highly condensed ring structures. Hydrocarbon biodegradation correlates strongly with the molecular ring structure, carbon number (molecular weight) and treatment conditions with a sequential order of biodegradability of petroleum components as follows: *n*-alkanes > branched-chain alkanes > branched alkenes > low-molecular-weight n-alkyl aromatics > monoaromatics > cyclic alkanes > polynuclear aromatics >> asphaltines (Huesemann 1995; Van Hamme *et al.* 2003). Head *et al.* (2006) suggested that it was the lower degradation rates of the high-molecular-weight hydrocarbons that caused the illusion of sequential phenomenon of hydrocarbon degradation. In an aged oil-contaminated site, the remaining oil was composed mainly of recalcitrant aromatic compounds with high resistance to biodegradation (Hatzinger & Alexander 1995). Certain bacteria (e.g. *Pseudomonas aeruginosa*) can

produce biosurfactants to increase bioavailability of persistent organic pollutants and therefore largely enhance the biodegradation of the recalcitrant compounds (van Elsas *et al.* 2006).

In situ bioremediation using indigenous microbes is thought to be an effective and low-cost strategy to take the components of degraded contaminants back to their natural cycles in moderately or slightly contaminated sites. However, this natural attenuation process, relying only on biological degradation, is relatively slow and limited by the composition of the autochthonous microbial community, bioavailability and characteristics of contaminants and other environmental factors such as soil texture, temperature, moisture, oxygen content, pH and nutrient conditions (Leahy & Colwell 1990; Balba *et al.* 1998; Boopathy 2000). Bioremediation is therefore used as a common follow-up procedure after extensive physical and chemical treatment at severely contaminated sites. In Finland, only 1% and 13% of the contaminated soils were directly subjected to biological treatment *in situ* and *ex situ*, respectively (Fig. 2), as the efficiency of bioremediation is limited by the cool climate (Eriksson *et al.* 2001).

Temperature affects hydrocarbon biodegradation by its effect on the physical nature and chemical composition of the oil contaminants, the rate of hydrocarbon degradation and the composition of the microbial community (Leahy & Colwell 1990). Cold-adapted psychrophilic and psychrotrophic microorganisms, which have adapted their metabolism to function optimally at low temperatures, play an important role in the *in situ* bioremediation of hydrocarbons in cold environments, where ambient summer temperatures often coincide with the optimal temperature range of their growth (Margesin *et al.* 2003). A large number of hydrocarbon-utilizing bacteria from oil-contaminated soils in boreal and other cold regions (e.g. alpine, arctic and antarctic) have been identified, mostly belonging to the phyla *Actinobacteria* and *Proteobacteria* (Whyte *et al.* 2002a; Margesin *et al.* 2003; Mukherjee *et al.* 2014; Mukherjee *et al.* 2015). Gram-positive bacteria were suggested to adapt to hydrocarbon biodegradation in soil in cold climate due to their high resistance to low temperature (Eriksson *et al.* 2001). Identified isolates with the best degradation ability of used engine oil have the optimum temperature for degradation at 30-37°C (Mandri & Lin 2007).

1.3.3. Phytoremediation
The main goal of a successful bioremediation design should be the creation of optimal conditions for microbial growth and metabolic activities (Balba *et al.* 1998). Rhizoremediation is an emerging technology that uses plants and their associated rhizosphere microorganisms to remove, transform, or extract toxic chemicals in various environments such as soils, sediments, ground water, surface water, and even the atmosphere (Susarla *et al.* 2002). Plants can well adapt to different environmental conditions and can also modify conditions of the environment to some extent (Susarla *et al.* 2002). Plants are capable of changing the composition or the amount of root exudation to stimulate bioremediation in the way of altering the microbial community structure, stimulating the growth of microorganisms, or increasing microbial catabolic activities (Radwan et al. 1995, Suominen et al. 2000, Kawasaki *et al.* 2012; Acharya et al. 2014). However, interactions between plants, soil and contaminants are rather complex. The efficiency of phytoremediation relies on the establishment of healthy plants with sufficient shoot and root biomass growth to support the activities of a flourishing microbial consortium in the rhizosphere (Wenzel 2009). Dense cultivation of suitable crops in polluted sites was thus suggested as a promising approach for bioremediation (Radwan *et al.* 1995).

Legume-cropping bioremediation system

The deficiency of nutrients, particularly nitrogen, is frequently a limiting factor that affects the efficiency of bioremediation (Wenzel 2009; Romantschuk *et al.* 2000). Fertilizer stimulates microbial growth and increases the rate of hydrocarbon biodegradation in a process called biostimulation (Gawel & North Dakota Industrial Commission 2006). However, the excessive N-fertilizers that cannot be taken up by crops tend to be lost through mechanisms such as ammonia volatilization, denitrification, and leaching, thus resulting in serious environmental problems such as polluting the atmosphere, aquatic systems, and groundwater (Choudhury & Kennedy 2005). In addition, the application of commercial N-fertilizers is laborious.

The cultivation of suitable oil-tolerant legumes can replace N-fertilizers in the bioremediation sites. According to the review by Franche *et al.* (2009), nitrogen fixation by legumes in their root nodules was firstly reported by the German scientists Hellriegel and Wilfarth in 1886. Two years later, a Dutch microbiologist Beijerinck successfully isolated a N-fixing bacterial strain (*Rhizobium leguminosarum*) from root nodules for the first time. Leguminous plants such as peas, beans and soybeans with the symbiotic association with various species of N-fixing bacteria, notably *Rhizobium*, can be used to increase soil fertility. During the symbiosis, rhizobia inhabit the legume-formed nodules on roots and stems of legumes, where they reduce atmospheric N_2, making it available to the host legumes. In the last 30 years, an increasing number of N-fixing bacteria have been identified in most of the phyla of Bacteria, e.g. *Proteobacteria*, *Chlorobi*, *Firmicutes*, *Cyanobacteria* and also in methanogenic Archaea. The worldwide distribution of these N-fixing organisms suggests that the N-fixing way of life is robust and successful, and that even if all other components of the N cycle were disturbed, atmospheric nitrogen would continue to be fixed (McArthur 2006). This capacity of biological nitrogen fixation (BNF) is substantial, often exceeding 100 kg ha^{-1} y^{-1}, more than enough to maintain N pools in ecosystems and to replenish N losses (Vitousek *et al.* 2002). Legumes, due to their capacity for BNF, are thus promising to assist bioremediation of soils contaminated with petrochemical waste as substitutes of fertilizers sustainably (Kamath *et al.* 2004; Chiapusio *et al.* 2007). In addition, legumes showed higher resistance to used motor oil contamination than other plants, in terms of the germination rate and overall growth (Dominguez-Rosado *et al.* 2004). Legumes such as alfalfa and clover have been successfully used to remediate PHC-contaminated soils due to their substantial benefit of BNF (Kamath *et al.* 2004; Chiapusio *et al.* 2007). Provided with proper management strategies for irrigation, fertilization, weed control (mowing, mulching, or spraying) and pest control, a cropping system assisted by N-fixing bacteria therefore holds promise for the successful bioremediation of contaminated soils.

Grass species are effective in binding and transforming hydrophobic contaminants such as BTEX and PAHs in contaminated sites, due to their large fine root biomass that can hold a higher microbial population than other species of a comparable size (Kamath *et al.* 2004; Chiapusio *et al.* 2007). Fast-growing grasses that are highly adaptive to new environments are therefore widely used as a primary remediation species in oil-contaminated sites. Once harvested, the grasses can be disposed as compost (bio-fertilizer) or for biogas production (Kamath *et al.* 2004). As a popular agricultural practice, the use of crop mixture of legumes and grasses increases the overall economic and environmental output by improving the soil nutrient condition, raising the yields of grasses, breaking the disease cycle and increasing soil biodiversity (Sleugh *et al.* 2000; Sanderson *et al.* 2004).

Plant growth-promoting bacteria (PGPB)

Certain plant-associated bacteria enhance plant growth. They can defend plants against pathogens, promote beneficial plant-microbe symbioses, increase nutrient uptake by solubilizing phosphate, fix nitrogen, stimulate plant growth by secreting phytohormones, exhibit antifungal activity, and induce systemic resistance (Pajuelo *et al.* 2011; Bhattacharyya & Jha 2012). The most studied group of plant growth-promoting bacteria (PGPB) are plant growth promoting rhizobacteria (PGPR) that colonize the root surfaces and the closely adhering soil, the rhizosphere (Compant *et al.* 2005). These PGPR, including *Pseudomonas, Acinetobacter, Achromobacter, Flavobacterium, Bacillus, Nocardia* and *Rhizobium* species have been shown to increase plant yields and soil organic matter (SOM) contents (Pajuelo *et al.* 2011). Several studies have reported that the inoculation of stress-tolerant plant species with the appropriate PGPR is able to effectively enhance the rhizoremediation of both organic and heavy metal polluted soils (Gerhardt *et al.* 2009; Hong *et al.* 2011; Pajuelo *et al.* 2011; Bhattacharyya & Jha 2012).

1.4. Methods used to study the environmental impacts of contaminants

1.4.1. Molecular methods

Classically, investigating the effectiveness and completeness of bioremediation requires a combination of different assessment tools such as culture-based isolation of oil-tolerant microbes, toxicity testing with a single-species biosensor, chemical quantification of contaminants, enumeration by direct microscopic cell counting or most probable number (MPN), enzymatic activities, and other biochemical methods (White *et al.* 1998; Dominguez-Rosado *et al.* 2004). However, these traditional methods and their combination only partially reflect the fate and environmental impacts of the contaminants with ignorance of the interactions between microbes, contaminants and the environment, especially for environmental samples (White et al. 1998).

Statistical analysis of DNA sequences from natural microbial communities can be used to accurately evaluate the ecological impacts of various environmental contaminants including uranium, nitrate and oil hydrocarbons (Smith *et al.* 2015). The development of molecular methods provides reliable tools and databases to explore the microbial diversity and community composition of environmental samples without the bias imposed by traditional cultural-based approaches (Torsvik *et al.* 1996), as less than 1% of microorganisms can be cultivated, characterized and identified using cultural based methods (Kellenberger 2001). It is possible to show the degree of relatedness by examining the similarity of the sequences of a certain type of DNA (e.g. 16S rRNA gene for prokaryotes and 18S rRNA for eukaryotes) from one organism to another using molecular biological approaches. To discriminate among bacterial species, the use of the amplication of 16S rRNA genes is based on at least two major assumptions: 1) that there has been no horizontal transfer of 16S rRNA genes and 2) that the amount of evolution or dissimilarity between sequences is representative of the entire genomes of bacteria (Goodfellow *et al.* 1997). Various other molecular markers such as the gene sequences of 23S rRNA, ATPase sub-units, elongation factors, and RNA polymerase generally characterize a high degree of correspondence with the 16S rRNA results to discriminate among bacteria (McArthur 2006). As 16S rRNA genes are highly conserved, species-specific dissimilarity in 16S rRNA genes are primarily found in certain hypervariable regions of the gene. To describe bacterial diversity in comparative studies, molecular techniques are mainly designed to amplify the hypervariable regions of the gene or the entire gene, if needed.

DNA profiling techniques using 16S rRNA genes as phylogenetic markers are rapid and economical methods to produce an overall pattern of the microbial community with considerable resolution rather than absolute identification of each individual species or genus in that community (Torsvik *et al.* 1996; Mills *et al.* 2003). Automated fluorescence-based detection techniques such as terminal restriction fragment length polymorphism (T-RFLP, Liu *et al.* 1997), length heterogeneity analysis of polymerase chain reaction (LH-PCR, Suzuki *et al.* 1998) and automated method of ribosomal intergenic spacer analysis (ARISA, Fisher & Triplett 1999) gained their popularity in microbial community studies due to the higher reproducibility of PCR and faster processing of large sample series. These automated microbial community fingerprinting methods provide both qualitative and semi-quantitative means to assess the remediation processes in terms of both efficiency and completeness, so they gradually replaced the laborious polyacrylamide gel electrophoresis-based methods such as denaturing gradient gel electrophoresis (DGGE) and rRNA intergenic spacer analysis (RISA, analogous to LH-PCR) (Borneman & Triplett 1997).

The LH-PCR method utilizes inherent variation in sequence lengths of the first two or three hypervariable regions of the 16S rRNA genes (Suzuki *et al.* 1998; Tiirola *et al.* 2003). A comparative study found that LH-PCR profiled a more diverse microbial community with higher operational simplicity and better reproducibility during bioremediation of petroleum contaminated soils than the T-RFLP method that is based on the enzymatic digestion of the 16S rRNA genes (Mills *et al.* 2003). The possibility to compare the amplicons against 16S rDNA *in silico* databases in size makes LH-PCR more practical for further specifying bacterial groups in the community (Tiirola *et al.* 2003).

Sequencing of 16S rRNA gene amplicons is now the most common approach for investigating environmental taxonomic diversity and structure of prokaryotic communities, despite the known biases introduced during PCR (Logares *et al.* 2014). The advent of next generation sequencing (NGS) technologies has enabled the quick and in-depth exploration of microbial genetic diversity at an unprecedented scale (Logares *et al.* 2014), even in populations in which little or no genetic information is available (Davey *et al.* 2011). Sequencing technologies are comprised of various methods from the processes of template preparation, sequencing and imaging, to data analysis; the unique combination of specific protocols differentiates one technology from another and determines the type of data produced from each platform (Metzker 2010). Several cross-platform studies have indicated the advantages and biases associated with different high-throughput platforms, e.g. Roche/454 and Illumina's HiSeq/MiSeq and SOLiD, serving as guidelines for platform selection to address biological questions of interest (Harismendy *et al.* 2009; Metzker 2010; Luo *et al.* 2012; Logares *et al.* 2014). In metagenomics studies, given the monetary savings and small errors, the short-read sequencing technique based on Illumina's MiSeq platform, has gained increasing popularity, compared to the Roche 454 technology (Luo *et al.* 2012). The development of multivariate statistical analysis methods facilitates the exploration of microbial community dynamics from multivariate data produced by both community fingerprinting and NGS sequencing techniques.

1.4.2. Statistical analysis of multivariate microbial community data

Data pretreatment and choice of distance measures
Traditional multivariate analysis of variance (MANOVA) is too strict in its assumptions for most multivariate ecological data sets, because the microbial community data (being counts or other

measures of abundances of species) i) tend to be highly aggregated or skewed (non-normally distributed); ii) contain high numbers of rare species that contribute lots of zeroes to the data set and iii) have numbers of operational taxonomic units (OTUs) that far exceed the number of observations per group (Anderson *et al.* 2008; Anderson 2001a).

Non-parametric multivariate methods that make no explicit assumptions regarding the distribution of multivariate variables are appropriate and more powerful in the analysis of multivariate ecological data (Clarke *et al.* 2014). These methods compare two (or more) samples that share particular species at comparable levels of abundance, either explicitly or implicitly, based on the similarity coefficients calculated between every pair of samples (Clarke *et al.* 2014). A distance/dissimilarity (1-similarity) measure is used as the basis of the multivariate data analysis as it affects the nature of the results, not just statistically but for interpretation as well (Anderson & Robinson 2003). There are two types of distance measures commonly used in the analysis of ecological data, depending on how they treat "zero" values in the multivariate data set. Some distance measures such as Euclidean (metric) distance treat "zero" like any other value on the number line; others, such as Chi-squared distance and Bray-Curtis distance, treat "zero" as a lack of information. According to Anderson & Robinson (2003), among these popular distance measures, Euclidean distance strongly emphasizes changes in relative abundance, rather than changes in composition (identities) of species; Chi-squared distance strongly emphasizes changes in composition rather than changes in abundance; Bray-Curtis distance (Bray & Curtis 1957), a semi-metric distance, reflects differences not only in relative abundance but also in compositional change. Therefore, to choose a suitable distance measure is important for better revealing the ecological patterns and largely depends on the study question. Nevertheless, distances calculated from the original data can often be over-dominated by a few highly abundant species. These dominant species cannot reflect distances of overall community composition in the following data analysis, so standardization (if any) and/or an appropriate transformation (e.g. square root or logarithmic transformation, reduction of the sample information to presence/absence for each species) can be applied to the data to reduce the contribution of the dominant species, prior to computing the similarities (Clarke *et al.* 2014). Bray-Curtis distance computed after suitable transformation is recommended as a satisfactory coefficient for the analysis of biological data of community structure (Clarke *et al.* 2014).

The work flow of multivariate community data analysis
The workflow and the most common multivariate methods of community data analysis are summarized in Table 1, based on Anderson *et al.* (2008) and Clarke *et al.* (2014).

At the first stage, a robust unconstrained ordination is used to view the overall multivariate patterns of observation dispersion based on the community composition. Representation of the ecological data by an ordination is more appropriate where the community pattern responds to a continuous abiotic gradient (Clarke *et al.* 2014). Among the popular ordination methods, principle component analysis (PCA, Chatfield & Collins 1980), based on Euclidean distance, is not a suitable ordination method for species-by-samples matrices of high dimensions, but it is suitable for normalized environmental data as environmental variables are always low in numbers (Clarke *et al.* 2014). Non-metric multidimensional scale ordination (nMDS, Kruskal & Wish 1978) and principle coordinate ordination (PCO, Gower 1966; Torgerson 1958) are the two routinely used ordination methods to visualize the overall pattern of multivariate ecological data. PCO (like PCA) is a projection of points

(perpendicularly) onto axes that minimize residual variation in the multivariate space, yet using a distance measure other than the Euclidean distance, whereas nMDS emphasizes preservation of only the rank order of dissimilarities for a chosen number of ordination dimensions (Anderson *et al.* 2008). However, the patterns of observation dispersion can be masked in unconstrained ordination if the direction of greatest total variation is completely different from the direction of group differences in multivariate space (Anderson & Willis 2003). Another community-representing method, the clustering technique, which uses a dendrogram to link the samples in hierarchical groups based on the similarity values between each cluster, is more appropriately used where the samples are expected to be divided into well-defined groups, particularly structured by some clear-cut environmental distinctions (Clarke *et al.* 2014).

At the second stage, a rigorous statistical test of the hypothesis about differences in the ecological assemblages induced by space, time and treatment factors can be performed using non-parametric multivariate methods such as analysis of similarity (ANOSIM, Clarke 1990; Clarke & Green 1988), permutational ANOVA and MANOVA (PERMANOVA, Anderson 2001a,b; McArdle & Anderson 2001; Anderson & Braak 2003) and generalized discriminant analysis (Anderson & Robinson 2003; Anderson & Willis 2003). All of these non-parametric methods use permutations to obtain *p*-values, so they allow rigorous probabilistic statements to be made for multivariate ecological data (Anderson 2001a). ANOSIM, based on the relationship (ranks) of dissimilarities, lacks the ability to partition the multivariate variation of overall effects into "main" and "interaction" components, so it is not suitable to be used in studies with complex experimental designs (Anderson *et al.* 2008). Unlike ANOSIM, PERMANOVA, which focuses on the dissimilarity values themselves, allows variance partitioning on the basis of any resemblance measure, and thus is suitable for testing the simultaneous response of one or more variables to multi-factorial ANOVA experimental design (Anderson *et al.* 2008). In addition, distance-based discriminant analysis (db-DA) or so-called generalized discriminant analysis or canonical analysis of principle coordinates (CAP), is used to both test for and visualize significant differences among *a priori* groups in multivariate space (Anderson & Robinson 2003; Anderson & Willis 2003). It is a constrained ordination method that shows maximized patterns of group differences through the multivariate cloud of points by finding axes that are best at discriminating among *a priori* groups, in regard to unconstrained ordinations (Anderson & Robinson 2003; Anderson & Willis 2003). Compared to PERMANOVA, db-DA is more powerful to detect effects if they are small in size and appear in a different direction to the axis of greatest total variation (Anderson *et al.* 2008). In many applications, PERMANOVA is done first, and if significant differences are found, db-DA is used as a following rigorous measure to characterize the difference/distinctiveness of the groups in the multivariate space (Anderson *et al.* 2008).

At the third stage in the determination of stress levels (Table 1), we intend to characterize the directional changes (e.g. adding or deleterious consequence) in community composition induced by the treatments. Are those changes "good" or "bad" and to what extent do they affect the community and even the whole ecosystem? This can be achieved either by referring to historical data for the similar sites or by investigating the changes (increase/decrease) in diversity indices, indicator species, opportunists and taxonomic distinctness (Clarke *et al.* 2014). In order to reduce the multivariate complexity of microbial community data into a small number of indices, single-variable ecological indices (e.g. diversity, richness and evenness) are calculated, before they can then be subject to

traditional univariate statistical analyses in many studies (Clarke & Warwick 2001). The analysis of the specific response of one single taxon (based on *a priori* selection) as an **indicator species** to a particular environment gradient is another popular alternative univariate method in interpreting ecological data (Clarke *et al.* 2014). However, the direct extraction of single variables from a multivariate data matrix seems to lower the sensitivity of the detection of the treatment effects. Hence, a combination of multivariate and univariate methods can be used to select the indicator species posteriorly to infer the impacts of the treatment factors in a case-specific manner. For instance, firstly, similarity percentages (SIMPER, Clarke 1993) can be used to characterize the species responsible for multivariate patterns or effects and to calculate the proportion of the total differences that each species accounts for between the groups of interest. Secondly, a robust univariate method (e.g. t-test, ANOVA and Beta binomial analysis) can be used as a follow-up procedure to test the significance of each representative species that accounts for large differences between *a priori* groups. Based on the above results, treatment association networks can then be drawn to show the holistic response patterns of the species to treatment. In some studies, temporal effects can interfere with the treatment effects and hence reduce the sensitivity of the above methods to select the suitable representative species that behaved differentially between treatments over time. In such circumstances, the indicator species can be further assigned appropriately in a time-series analysis, e.g. using the DESeq2 package (differential analysis of count data, Love et al. 2014). However, there are conditions when the popular multivariate techniques are sensitive. Alternative methods such as meta-analysis, multivariate dispersion and seriation can be used (Clarke *et al.* 2014).

At the last stage, we are interested to characterize the relationship between the changes in the community composition and the environmental variables (quantitative data). A bubble plot, an ordination plot (e.g. MDS) with bubbles of different sizes as vectors, can be used to visualize the patterns of communities and the effect of a single or several environmental variables in one ordination plot at one time (Clarke *et al.* 2014). However, a bubble plot based on unconstrained MDS ordination is a poor approximation to the community similarity matrix (Clarke *et al.* 2014). The non-parametric BEST/BIOENV (best subset of environmental variables with maximum rank correlation with community dissimilarities) procedure is a heuristic approach designed to display the multivariate pattern of the environmental data and to find the combination of environmental variables that reveals the extent to which the community pattern can be explained by knowledge of the full set, or a subset, of the environmental variables, using the rank-based approach in parallel with MDS and ANOSIM (Clarke *et al.* 2014). Distance-based linear models (DISTLM, Legendre & Anderson 1999; Anderson 2001b; McArdle & Anderson 2001), like PERMANOVA, partition variation according to a (multiple) regression model in which the predictor variables are quantitative and continuous; nonetheless, it is different from PERMANOVA where the variance portioning was based on an ANOVA model in which the predictor variables are categorical (Anderson *et al.* 2008). The purpose of DISTLM is to find the predictor (environmental) variables that are best at explaining variation in the response community data. Nevertheless, based on the concept of parsimony, the model can be reduced by removing the non-significant and the multi-collinear environmental variables using a suitable selection criterion, e.g. R^2 (the proportion of explained variation for the model), adjusted R^2, AIC (Akaike information criterion) or BIC (Bayesian information criterion) (Anderson *et al.* 2008). Noticeably, the predictor variables that were selected in a model should not be interpreted as being necessarily causative of the change in the community composition, whereas they may act as proxies for some other

important variables that were either unmeasured or omitted from the model for the sake of parsimony (Anderson *et al.* 2008). DISTLM has two main advantages over BEST/BIOENV: i) it can directly provide a quantitative measure of the relative importance of the selected individual environmental variables and ii) most tellingly, it can provide sequential (partial) tests for the statistical significance of adding or deleting an explanatory variable from the current model (Anderson *et al.* 2008). Distance-based redundancy analysis (db-RDA) is used to visualize the fitted model that is produced from DISTLM in an ordination. The environmental variables are overlaid as vectors individually on the ordination diagram to indicate the strength and directions of their effects on the community data, based on the calculation of the multiple partial correlations between each environmental variable and the db-RDA axis scores (Anderson *et al.* 2008). An alternative ordination method, canonical correlation analysis (CCorA, Anderson & Robinson 2003; Anderson & Willis 2003), is routinely used to find the axes in the multivariate space (matrix of community variables) that have the strongest correlation with another set of variables (matrix of environment variables). A constrained binary divisive clustering method, linkage trees (LINKTREE, multivariate regression trees), which gives the maximized separation of community at one site into two groups, is another good candidate for characterizing the relationship between the environmental variables and the multivariate community data (Clarke *et al.* 2014).

Most of these multivariate methods, e.g. ANOSIM, PERMANOVA, db-DA and DISTLM, are sensitive to differences in location, dispersion or correlation structure among groups (Anderson & Willis 2003). Hence, tests of homogeneity of observation dispersions (PERMDISP, Anderson 2006) across treatments should be performed to allow that the unequal dispersion and the correlation structure among samples, if any, are ignored under permutation in multivariate data analysis.

Table 1. Common methods of multivariate community data analysis, according to Clarke *et al.* (2014) and Anderson *et al.* (2008)

Stage	Purpose	Approach	Methods
1	Representing communities	By low-dimensional graphic description of the relationship between the samples	Clustering (dendrogram) or ordination (e.g. MDS, PCA or PCO)
2	Discriminating sites/conditions (between *a priori* groups)	By hypothesis testing on the basis of their community composition	ANOSIM, PERMANOVA, db-DA and SIMPER
3	Determining levels of stress or disturbance (to infer directional changes)	By constructing biological measures from a range of community data which are indicative of disturbed conditions	Meta-analysis, multivariate dispersion and seriation, univariate analysis of diversity indices, indicator species, opportunists and taxonomic distinctness
4	Linking to environmental variables and examining the issues of causality of any changes	By regression techniques	MDS with bubble plots, LINKTREE, BEST/BIOENV, DISTLM/db-RDA, CCorA (CAP)

Abbreviations: *MDS* multidimensional scaling, *PCA* principal component analysis, *PCO* principle coordinate ordination, *ANOSIM* analysis of similarity, *PERMANOVA* permutational univariate or multivariate analysis of variance, *db-DA* distance-based discriminant analysis, *SIMPER* Similarity percentages, *LINKTREE* multivariate regression trees, *BEST/BIOENV* (best subset of environmental variables with maximum rank correlation with community dissimilarities), *DISTLM* distance-based linear models, *db-RDA* distance-based redundancy analysis, *CCorA* canonical correlation analysis, *CAP* canonical analysis of principle coordinates.

2. Significance of the empirical bioremediation project in a boreal climate

2.1. Used motor oil

Soil contamination by used motor oil is not well characterized in field studies. Used motor oil or used crankcase oil is used lubricating oil that is removed from the crankcase of internal combustion engines of vehicles (Irwin *et al.* 1997). The increasing use of automobiles and machinery vehicles inevitably increases the spillage of used motor oil into the surrounding environment. Motor oil is made up of mainly a base lubricating oil (80%-90% by volume) and performance-enhancing additives (Irwin *et al.* 1997). In addition to the contamination with heavy metals and other impurities during the circulation in the engines, the concentrations of long-chain aliphatics, benzene-, and naphthalene-based compounds, polycyclic aromatic hydrocarbons (PAHs) and heavy metals are high in used motor oil; once released, these contaminants can result in chronic hazards to the environment, due to their high resistance to microbial degradation (Michelsen & Boyce 1993; Irwin *et al.* 1997; Dominguez-Rosado *et al.* 2004).

2.2. Selection of crops and PGPB

The pursuit of bioremediation and agriculture in the aspect of social, economic and environmental sustainability will require substantial knowledge of the scientifically sound technologies/approaches that can enhance the decision-making at the field level according to the local climatic and environmental conditions. Plant-assisted bioremediation of oil-contaminated soil in the Nordic countries is limited by the shortness of the growing season, so the selection of suitable crops in the boreal climate is the key to a successful bioremediation project.

As bioremediation is a slow process that does not allow many disturbances of the contaminated soil, the use of perennial crops with proper field management holds promise for accelerating oil degradation with reduced yearly inputs. The other advantage to cultivate perennial crops is that we can harvest the biomass for bioenergy production, without replanting them from year-to-year. Fodder galega (*Galega orientalis* Lam.), a perennial, fast-growing forage legume, and smooth brome (*Bromus inermis* L.), a cool-season perennial, sod-forming grass, are two good candidates to remediate oil contaminated soils in the low temperature and acid soil conditions of northern regions. *G. orientalis* has a well-developed root system (Varis 1986) that improves the aeration of the penetrated below-ground soil and its exudates enhance the growth and metabolic activities of beneficial soil organisms at the rhizosphere. The cultivation practice of this legume has also been optimized for ecological cropping systems (Lindström et al. 2003). Several microcosm and mesocosm studies have demonstrated the oil-tolerance capacity and rhizoremediation potential of *G. orientalis* and its microsymbiont *Neorhizobium galegae* to remediate soils contaminated with BTEX) (Suominen *et al.* 2000; Lindstrom *et al.* 2003; Jussila *et al.* 2006; Kaksonen *et al.* 2006; Mikkonen *et al.* 2011a). The long and strong stem of fodder galega is suitable for direct combustion, and young green plants are well suited to digest to biofuels (Võsa & Meripõld 2008). Adamovich *et al.* (2007) elucidated that fodder galega in mixture with grasses is able to economically provide continuous and high forage production during summer without additional N fertilization. Field conditions and management practices such as sowing time, N fertilizer use, and the frequency of cutting significantly affected the productivity of galega-grass swards (Zolotarev 2010). Therefore, the combination of these two

perennial crops was expected to alleviate soil contamination in boreal soils while providing us with crop biomass for bioenergy production in the field.

A previous greenhouse experiment demonstrated that the co-inoculation of fodder galega seeds with its specific rhizobia and PGPB strains *Pseudomonas trivialis* 3Re27 and *P. extremorientalis* TSAU20, improved plant growth, nodulation and BNF efficiency (Egamberdieva *et al.* 2010). However, the competitiveness and effectiveness of these PGPB strains were not evaluated in uncontrolled field conditions.

The biodiversity and composition of bacterial communities that underlie the metabolic processes in oil-contaminated soils have been studied using 16S rRNA gene taxonomy. As the environmental DNA pool is highly complex with many genomes that are unknown and normally present in very low abundances, kinetics and amplification, PCR biases may behave differently than in controlled studies (Logares *et al.* 2014). However, most of the previous studies were performed in either a small-scale microcosm experiments or a short-term field experiment where there were no non-contaminated control samples for comparative study. The impact of crop cultivation on structural and functional bacterial community may vary, e.g. from crop type and age. Hence, it is also important to empirically evaluate the impact of different perennial crops and their combination on soil bacterial communities. Before the establishment of this field experiment in 2009, there was no systematically described study on bioremediation of oil-contaminated agricultural soil combined with crop biomass production in boreal regions, with both non-contaminated and non-vegetated plots as controls. This study was conducted to fill this knowledge gap by monitoring the changes in diversity and composition of the soil-borne microbial community in a Finnish legume-cropping system.

3. Objectives of the empirical study

We established a multi-year bioremediation field experiment on Viikki Experimental Farm (60°14'N, 25°01'E), Helsinki, with crop treatments of monocrops, their mixture and bare fallow, oil+/- treatments (7000 ppm), and plant growth promoting bacteria (PGPB+/PGPB-) treatments.

The overall aims of this project was to i) assess the sustainability of the legume-cropping bioremediation system economically in terms of crop yield, and environmentally in terms of oil degradation rate; and ii) reveal the dynamics of bacterial communities and certain oil-responding bacterial OTUs over time to provide ecological inferences for a remediation process in oil-contaminated soil. Our ultimate goal was to develop a systematic and comprehensive plant-microbe-soil monitoring tool to evaluate bioremediation as well as agricultural soil management practices in Finland. The multidisciplinary research framework of the whole bioremediation project was split into four parts reported in three papers (I-III), with specific focuses summarized in Fig. 3 and objectives in Table 2.
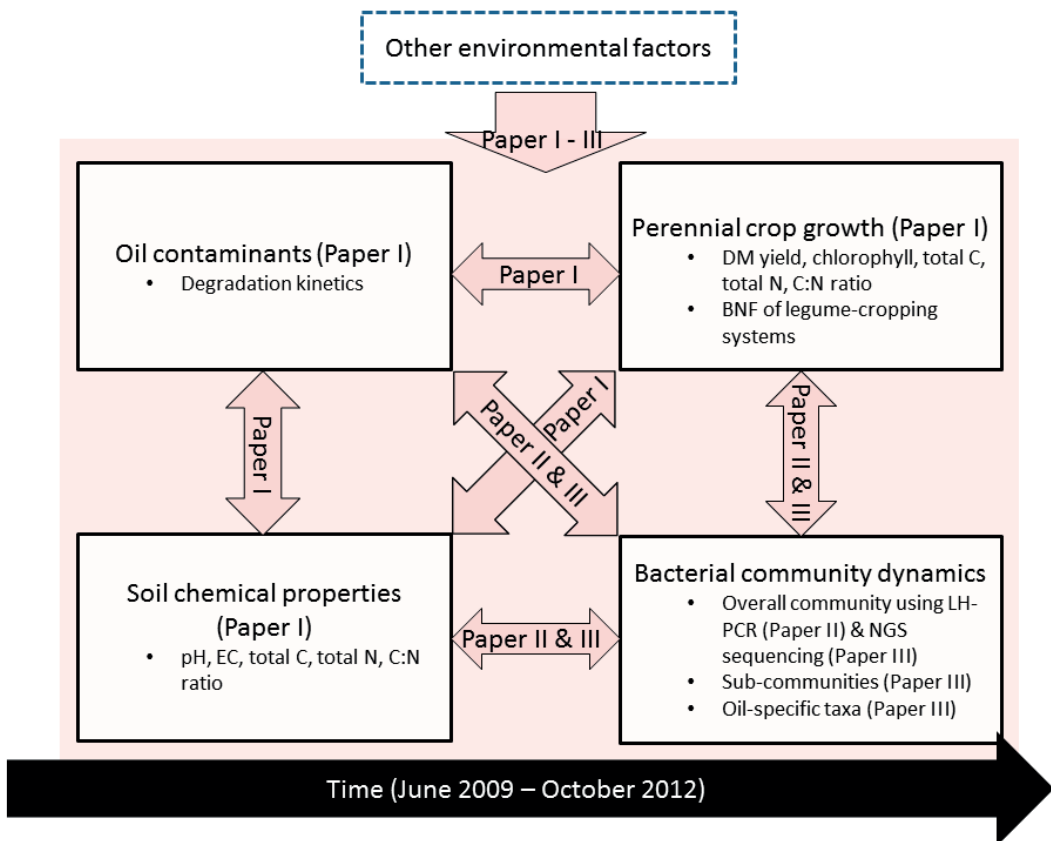
Fig. 3. The research framework of the multi-year bioremediation field experiment under different perennial cropping systems in a boreal climate. Arrows denote interactions between the studied abiotic and biotic factors, summarized accordingly from Paper I-III.

Table 2. Research objectives of the bioremediation field experiment, specified in Paper I-III

Paper	Objectives
I	i) to evaluate the ecological suitability and potential economic benefits of perennial cropping systems (pure fodder galega , pure brome grass and their mixture) during bioremediation of oil-contaminated soil in a boreal climate
	ii) to study the oil degradation patterns and soil agrochemical properties in different perennial cropping systems in a boreal climate
II	i) to monitor the impact of used motor oil, perennial cropping systems, PGPB and soil parameters on diversity and structure of bacterial community
	ii) to pre-screen soil samples for high-throughput sequencing for further identification of bacterial taxa of interest
III	i) to provide general phylogenetic information on the bacterial communities present in a Finnish agricultural soil
	ii) to analyze the successional patterns displayed by the whole bacterial assemblage and sub-communities in response to oil
	iii) to identify the taxa of oil-favored bacteria that were responsible for the dissimilarity between oil-contaminated and non-contaminated soil at different phases of bioremediation
	iv) to link the changes in the bacterial community with environmental variables (soil chemical and crop physiochemical properties)

4. Materials and methods

4.1. Geographic location of the bioremediation field

The multi-year bioremediation field experiment was established in Viikki, Helsinki in June 2009 (Fig. 4). Detailed information of the experimental site (e.g. soil texture, cropping history and meteorological data) is given in Paper I.

Fig. 4. Geographic location of the bioremediation experimental field (marked in red) at Viikki Experimental Farm, Helsinki, Finland (60°14'N, 25°01'E, 8 m AMSL)

4.2. Experimental design and field management

The experiment was designed in split-plots with crop treatments (brome grass, fodder galega, their mixture and bare fallow) as the main plots in four replicated blocks (Fig. 5). The sub-plot factor was factorial combinations of oil (+/-) and PGPB (+/-) treatments. About 6 kg of used motor oil (Teboil Lubricants Classic Mineral Motor oil, SAE 10W-30, API SF/CD, Finland) was mixed with 10 kg of coarse sand (0.5-1.2 mm), spread and spiked into the top 20 cm of oil-contaminated plot with a rotary tiller on 17 June 2009, making the target contamination approximately 7000 ppm (7 g kg^{-1} dry soil). The non-contaminated control plots received pure sand into the top 20 cm soil. Before sowing, seeds of *G. orientalis* (Naturcom Oy, Ruukki, Finland) and *B. inermis* cv. 'Lehis' (Jõgeva Plant Breeding Institute, Estonia) were surface-sterilized. *G. orientalis* seeds were inoculated with *Neorhizobium galegae* strain HAMBI 540 (University of Helsinki, Helsinki, Finland). The seeds were manually sown

24

and lightly covered by raking. Weeds were controlled manually in the growing season. The establishment (e.g. oil spike, bacterial inoculation and sowing of seeds) and management (e.g. weeding and harvesting) of the experimental field are described in detail in Paper I.
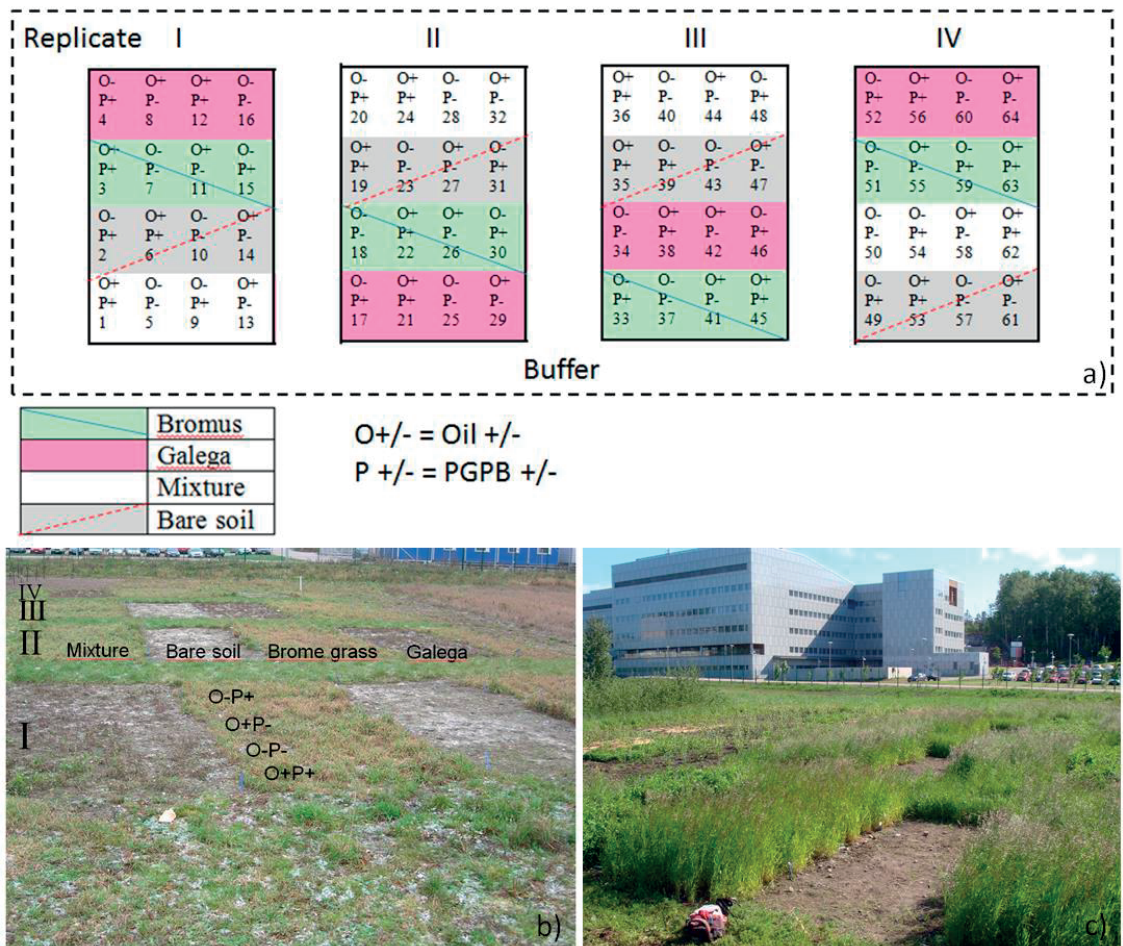

Fig. 5. a) Layout of the field experiment; b) a view of the field in winter, photographed by Fred Stoddard in November 2010; c) a view of the field in summer (on the first-cut day), photographed by Lijuan Yan in June 2010. There are four replicates (I-IV) of four main plots (galega, brome grass, brome grass-galega mixture and bare fallow), each of which has 4 sub-plots of ±oil and ±PGPB in factorial combinations, making a total of 64 plots, each 2.5 m ×1.5 m.

4.3. Data collection

Soil samples were taken from the top 20 cm layer in the field at six times (July 2009, May 2010, November 2010, May 2011, May 2012 and October 2012). Soil chemical properties (oil concentration, pH, EC, total C, total N and C:N ratio) of three sample sets (July 2009, November 2010 and May 2012) were measured. Crops were harvested in midsummer and autumn each year. Crop physiological properties (annual DM yield, total C, total N, C: N ratio, chlorophyll and BNF of the legume) were measured. Detailed information concerning the measurements and calculations of soil and crop properties were described in Paper I. Soil-borne bacterial communities were characterized by using the LH-PCR technique for both spring and autumn samples in Paper II. Illumina's MiSeq sequencing

was performed only on spring samples in Paper III. The target amplicon region of the bacterial 16S rDNA was from V1 to V3 for both LH-PCR and MiSeq sequencing techniques with primers and thermal cycler programs described in Papers II and III, respectively.

To profile the fingerprints of the hydrocarbon composition of used motor oil, we followed the gas chromatographic method according to the international standard method (ISO 16703:2004). The extraction procedures of oil hydrocarbons in soil contaminated with used motor-oil were described in Paper I. Florisil (magnesium silicate, 150-250 µM, 60-100 mesh, activated by heating for at least 16 h at 140° C before analysis) was used to trap the interfering polar compounds from the oil extract in a clean-up column, prior to GC-FID (gas chromatograph with flame ionization detector) separation of the eluate. The GC-FID (Agilent Technologies 6890N Network GC System) was equipped with a splitless injector (Agilent 7683 Series), an auto-sampler and a RTX silica capillary column (length 30 m, I.D. 0.32 mm, 5% diphenyl-95% dimethylepolysiloxane, film thickness 0.25 µm, maximum temperature 350°C, with 5 m long precolumn). The injection volume was 1.0 µL. Helium was used as a carrier gas (flow: 1.4 ml /min). The temperature of the detector was adjusted to 320°C. The temperature program (OLJY) was used for driving oil samples and heptane (60°C 2 min, 10°C/min to 320°C, 320°C 25 min, 50°C/min to 340°C, 7 min 340°C). Cross-contamination was controlled by column cleaning after each drive of oil samples with the PAAHTO program (60°C 1 min, 10°C/min to 340°C, 320°C 40 min) and injector washing (pre-injection 2×A; post injection 2×A + 3×B; A=acetone (HPLC), B=heptane (HPLC)).

4.4. Statistical analysis
The statistical methods used in Paper I-III are shown in Table 3. Repeated measures (RM) and univariate (UV) analysis of variance (ANOVA) were used to test the significance of the treatment effects on the univariate variables based on a split-plot design. Bray-Curtis distance-based nonparametric analyses were used to test and visualize the treatment effects on bacterial multivariate community data. All nonparametric analyses were performed with 9999 permutations to determine the significance of the treatment effects. In all statistical analysis, differences were concluded significant at $p < 0.05$.

Table 3. Statistical methods used in this study

Variables	Times	Methods	Software/Package	Important references
Univariate data				
Soil chemical properties	A, C and E	RM/UV ANOVA	SPSS v. 20, IBM Inc., Armonk, NY, USA	
Oil concentration	All A to F	RM/UV ANOVA and Kinetic modeling		Snoeyink & Jenkins 1980
Crop physiological properties		RM/UV ANOVA		
Ecological indices (diversity, richness and evenness)	LH-PCR: A to F	RM/UV ANOVA		
	MiSeq: A, B, D and E			
Multivariate data (based on Bray-Curtis distance)				
LH-PCR community fingerprinting data	A to F	PCO	R environment / "cmdscale" in package Vegen	Torgerson 1958; Gower 1966; Oksanen *et al.* 2015
		PERMANOVA	PRIMER V.6 software / in PERMANOVA+ package	Anderson 2001a,b; McArdle & Anderson 2001; Anderson & Braak 2003; Anderson *et al.* 2008
		db-DA	R environment / "CAPDISCRIM" in package BiodiversityR	Anderson & Robinson 2003; Kindt & Coe 2005
		PERMDISP	R environment / "betadisper"and "permutest.betadisper" in package Vegen	Anderson 2006
Overall community based on MiSeq OTU data	A, B, D and E	nMDS	PRIMER V.7 software	Clarke *et al.* 2014
		PERMANOVA and RM PERMANOVA	PRIMER V.7 software / in PERMANOVA+ package	Anderson 2001a,b; McArdle & Anderson 2001; Anderson *et al.* 2008
		PERMDISP		Anderson 2006
		db-DA		Anderson & Robinson 2003

		SIMPER	PRIMER V.7 software	Clarke 1993; Clarke *et al.* 2014
		Beta binomial analysis	R environment / ibb package	Pham *et al.* 2010
		Time series analysis	R environment / DESeq2	Love et al. 2014
Sub-communities (major phyla and classes) based on MiSeq OTU data	A, B, D and E	db-DA	PRIMER V.7 software / in PERMANOVA+ package	Anderson & Robinson 2003; Anderson & Willis 2003

Linking bacterial community to environmental variables

Between LH-PCR fingerprinting data and soil chemical properties	A and E	Constrained analysis of principle coordinates (CAP/db-RDA)	R environment / "capscale", "permutest" and "scores.rda" in package Vegen	Legendre & Anderson 1999
Between MiSeq OTU data and soil and crop parameters	A and E	DISTLM and db-RDA	PRIMER V.7 software / in PERMANOVA+ package	Legendre & Anderson 1999; McArdle & Anderson 2001; Anderson 2001b

Sampling times: *A*: July 2009, *B*: May 2010, *C*: November 2010, *D*: May 2011, *E*: May 2012 and *F*: October 2012. Abbreviations: *RM* repeated mesures, *UV* univariate, *ANOVA* analysis of variance, *PCO* principle coordinate ordination, *PERMANOVA* permutational univariate or multivariate analysis of variance, *db-DA* distance-based discriminant analysis, *PERMDISP* test of homogeneity of dispersions by permutations, *nMDS* non-metric multidimensional scaling, *SIMPER* Similarity percentages, *DISTLM* distance-based linear models, *CAP* canonical analysis of principle coordinates, *db-RDA* distance-based redundancy analysis.

5. Results and discussion

5.1. Evaluation of oil degradation

The monitoring of oil concentration over time provided a direct evaluation of oil dissipation during bioremediation in the soil. The oil concentration decreased over time (Paper I, Fig. 1), following a typical first-order kinetic pattern (Paper I, Table 9), agreeing with the majority of oil-degradation kinetic studies (Jørgensen *et al.* 2000; Nocentini *et al.* 2000; Van Gestel *et al.* 2003). Oil degradation was most intense during the first month, as shown by a reduction of oil concentration by an average of 42% (Paper I, Fig. 1). As the initial oil concentrations estimated in the first-order kinetic models were lower than the designed oil input value by an average of 29% (Fig. 6), the initial rapid loss of oil was attributed to a substantial abiotic loss of volatile hydrocarbon compounds. High summer temperatures tend to enhance volatilization, especially when the soil begins to dry out (ATSDR 1999). The second significant loss of oil (Paper I, Fig. 1) during the growing season in 2010 was probably caused by the biodegradation of easily available hydrocarbons at optimal temperature in summer, since biodegradation rates increase as temperature increases (Leahy & Colwell 1990). The degradation was rather slow in the cold seasons (Paper I, Fig.1), in line with the finding that the relatively low temperature limited the efficiency of hydrocarbon removal in cold regions (Eriksson *et al.* 2001). After the summer 2010, oil concentration remained relatively stable. At the end of the experiment, the oil degradation was incomplete, with 8% remaining in bare soil, 25% in pure grass, 27% in pure legume and 23% in mixture plots. This phenomenon of a stationary oil-degradation phase at the end of a bioremediation project prevails (Nocentini *et al.* 2000), explained by the poor bioavailability and high resistance of aged hydrocarbons to microbes (Huesemann 1997).
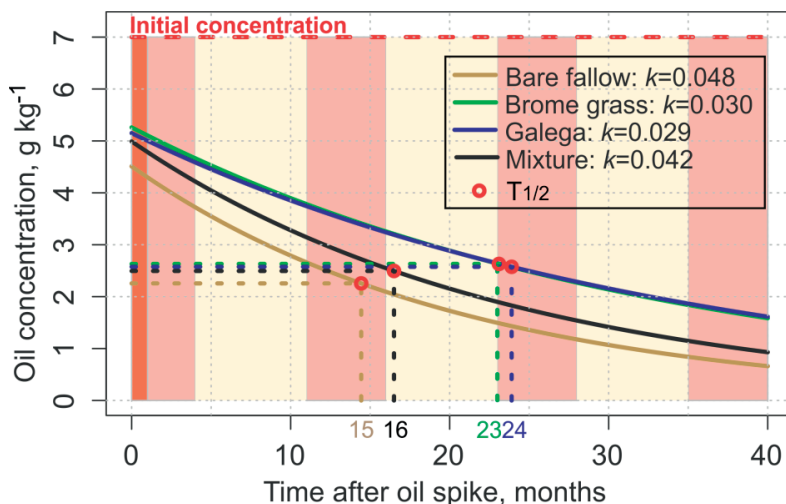
Fig. 6. Oil degradation kinetics under different cropping systems, reproduced from Table 10 in Paper I. The rectangles in different colors denote different periods: first month after oil spike (in orange), the growing season (May- October) (in pink) and the cold season (November-April) in each year. Abbreviations: "k" rate constant and "$T_{1/2}$" the time (half-life) when the oil concentration was halved in the soil, according to the first-order kinetics.

In the first-order kinetic model, the oil degradation rate was proportional to the oil concentration. The oil concentration at the starting point of biodegradation, fitted by the first-order kinetic model, was higher in vegetated plots than in bare fallow (Fig. 6). From 15 to 24 months were required to halve the hydrocarbon concentration in this region, depending on the crop type (Fig. 6). The degradation rate constants (k) in different cropping systems decreased in the following order: bare fallow > galega-brome grass mixture > pure brome grass > pure galega (Fig. 6). There are three possible reasons to explain the negative effect of crops on oil reduction rate. Firstly and most likely, as the bare soil was directly exposed to the sun, the soil-oil complex at its surface layer tends to absorb more radiation, thus contributing to the greater volatilization of the light hydrocarbons than in the vegetated plots (ATSDR 1999). Particularly, a sharper (abiotic) decrease of oil concentration occurred in bare fallow plots one month after the oil spike, in regard to vegetated plots (Paper I, Fig. 1). Secondly, plant-derived organic materials were more favorable substrates for microorganisms (Chaudhary *et al.* 2011), thus affecting the distribution and bioavailability of PHCs. Thirdly, some plant-derived oil-like compounds such as waxes and chlorophyll could have been co-extracted as TSEM, contributing to an overestimation of oil concentration in the vegetated plots. This explanation is supported by the consistency between the high oil concentration and the high crop chlorophyll content and dry matter yield in the vegetated plots (especially in pure galega plots) in later years (2011-2012).
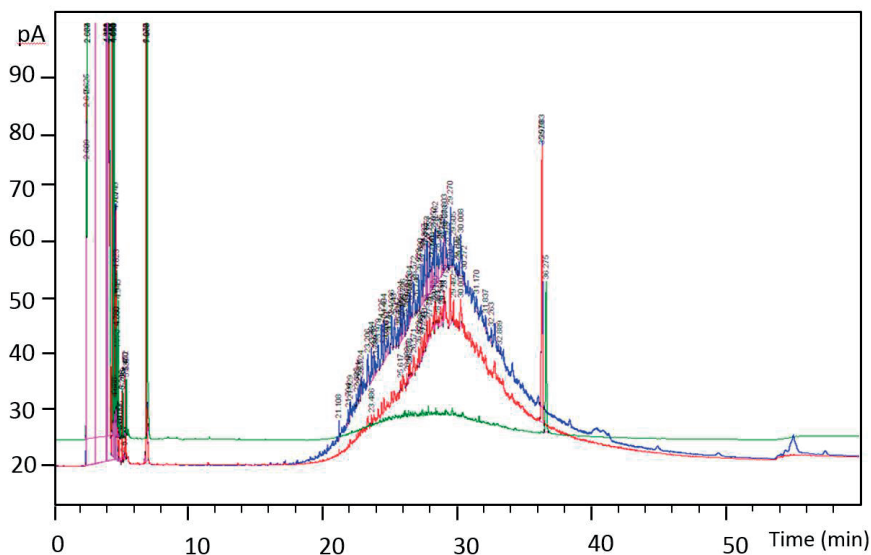
Fig. 7. An example of GC-FID fingerprinting of hydrocarbon composition from a motor oil-contaminated soil (Plot No. 11, pure brome grass, Oil+, PGPB-) at three sampling times (blue: July 2009; red: November 2010; green: May 2012). The peaks of the interior hydrocarbon standards *n*-decane (C_{10}) and *n*-tetracontane (C_{40}) appeared approximately at 7 min and at 36 min, respectively.

Although different types of PHC mixtures have similar constituents, the relative abundance, toxicity and bioavailability of these components vary considerably, and these variations are potentially associated with the observed different structures of oil-degrading microbial consortia at different contaminated sites (Hamamura *et al.* 2006). The fingerprint of the extracted PHC mixture from the contaminated soil, as revealed by GC-FID chromatography (Fig. 7), showed a "bell shape" that was similar to the GC-MS spectrum reported by Dominguez-Rosado *et al.* (2004), characterized by a huge

unresolved complex mixture (UCM) hump. The non-GC-FID-resolvable compounds are polar organic compounds with high molecular weight originating from the polar hydrocarbon metabolites (Robertson *et al.* 2007). This UCM-dominated profile of motor oil was highly different from the peak-dominated GC-FID fingerprint of fuel oil shown by Mikkonen *et al.* (2011a). The UCM hump decreased with time. However, heavier distillates that are more recalcitrant to degradation are expected to remain at the end of a bioremediation experiment (Sarkar *et al.* 2005), as observed in our experiment.

5.2. Perennial crop growth

5.2.1. Crop growth in non-contaminated soil

We studied the ecological suitability and agronomic productivity of the two perennial crops, fodder galega and brome grass, in terms of their adaptation ability, biological nitrogen fixation (BNF) efficiency and dry matter (DM) production in a boreal region. Both fodder galega and brome grass are suitable for cultivation in boreal regions as potential bioenergy swards, despite the slow initial growth of the galega in the first year. Galega tend to exhibit a slow initial growth in the first year after sowing, but a stable production in the second or third year was reported in studies, e.g. Halling *et al.* (2002), Adamovich (2007), Kryževičienė *et al.* (2008) and Zolotarev (2010). Our measurements of dry matter yield, chlorophyll, total C, total N and BNF on galega over years agreed with this finding. The unfertilized galega-brome grass mixture out-yielded the N-fertilized pure grass swards over years by an average of 34%. This finding is in accordance with previous field studies (Halling *et al.* 2002; Adamovich *et al.* 2007; Kryževičienė *et al.* 2008), in which the better performance of the grass in legume-grass mixture was attributed to the substantial BNF capacity of the legume fodder galega.

Our data showed that BNF was more efficient in the galega-brome grass mixture than in pure legume stands, confirming the advantage of legume-grass intercropping. The observed decreasing $\delta^{15}N$ values and higher yield of the brome grass in the mixture plots over time indicated N translocation from the legume to the associated grass, probably via litter decomposition and/or root exudation, as suggested by Fujita *et al.* (1992) and Tilman *et al.* (2002). Some studies have suggested that free-living N-fixing bacteria have an important role in satisfying the nitrogen needs of perennial crop growth (Mao *et al.* 2013), but the impact of free-living N-fixing bacteria on the overall BNF yield and crop growth was negligible in our study.

5.2.2. Comparison of crop growth between oil-contaminated and non-contaminated soil

Our field data showed that used motor oil (7000 ppm) was insufficient to depress the growth of both crops, despite a minor decrease of crop total C in brome grass in 2010 (data not shown) and a decline of %Ndfa (proportion of legume shoot N derived from atmospheric N_2) in galega and mixture plots in 2011 (Paper I, Table 6). The normal growth and symbiotic functions of fodder galega under oil stress in the field validated the findings of previous microcosm and mesocosm studies (Suominen *et al.* 2000; Jussila *et al.* 2006; Mikkonen *et al.* 2011a). Surprisingly, oil contamination significantly increased the overall crop production in terms of DM and N yield (Paper I, Table 5). The enhancement of oil on crop growth was more evident in 2012, with a marked increase of DM yield, %Ndfa, BNF and N yield, especially in the pure galega plots (Table 5 and Table 6). The oil-degradation activities of the heterotrophic microbes might have provided other soil organisms with extra C and energy source in oil-contaminates soil. The subsequent recovery of soil microbial activities, especially those of the rhizobia, from oil inhibition could therefore have

contributed to the increased BNF, DM yield and N yield of the legume. Therefore, the galega-brome grass mixture is a suitable cropping system to grow in oil-contaminated soil for bioenergy production in boreal regions, in terms of its remarkable biomass yield and high BNF efficiency.

5.3. Soil properties

"*Fertile soils with good physical properties to support root growth are essential for sustainable agriculture*" (Tilman *et al.* 2002).

Oil contamination has a substantial impact on soil physiochemical properties. The PCA plot (Fig. 8) showed the internal structure of the soil samples in a way that best explains the variance in the data. The oil-contaminated and non-contaminated samples were clearly separated by the first principle component that explained 38.3% of the variance. Oil addition was, as expected, a significant source of soil total carbon (Paper I, Table 8). Oil concentration was positively correlated with soil total C (Pearson correlation 0.53) and C:N ratio (Pearson correlation 0.60). Soil electrical conductivity (EC), which is influenced by properties of the pore-filling contaminants at the interface region of a soil (Börner *et al.* 1993), was lower in oil-contaminated plots than in non-contaminated ones at the start (Paper I, Table 8), but the difference of EC values between oil-contaminated and non-contaminated soils decreased with time. The lowered EC values did not result in a decline in soil function, as the crop dry matter yield increased significantly in lower-EC (oil-contaminated) plots. The field soil was a well buffered system. We observed no significant difference in soil pH and total N concentration between oil-treated and clean plots in our field experiment (Paper I, Table 8).
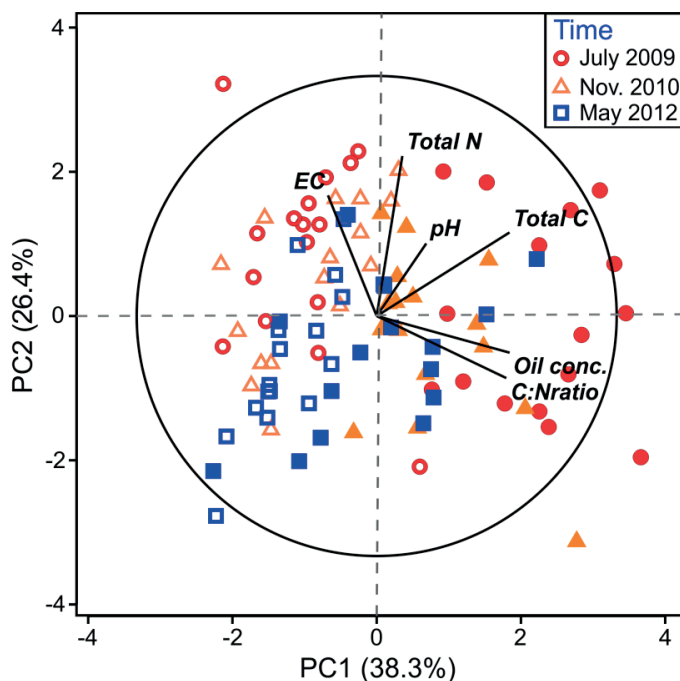
Fig. 8. Principal component analysis (PCA) showing the distribution of soil samples over three time points based on the measured soil chemical variables. Vectors of soil chemical variables were overlaid onto the plot with lengths denoting Pearson correlation values between the vector and the PCA coordinate. The radius of the blue circle equals 1 (Pearson correlation r = 1). Oil-contaminated samples are in solid symbols and non-contaminated in open symbols. See Table 8 (Paper I) for

details of these variables. Abbreviations: "*EC*" electrical conductivity (µS/cm), "*Oil conc.*" oil concentration (g/kg dry soil), "*Total C*" total carbon content (g/kg dry soil), "*Total N*" total nitrogen (g/kg dry soil) and "*C:Nratio*" carbon:nitrogen ratio.

Soil quality is important to assess as it directly affects the effectiveness and efficiency of bioremediation. Several soil properties (e.g. soil texture, SOM, C:N ratio, and pH) have been shown to be greatly involved in hydrocarbon degradation in non-vegetated soils. Soil pH influences plant growth, cation solubility, microbial activity and clay dispersion (Haynes & Naidu 1998). The ideal pH to promote biodegradation of soil contamination ranges between 6 and 8, depending on the species of the degraders (ATSDR 1999). In our study, soil pH was optimum for oil degradation, even though it decreased slightly over time (Paper I, Table 8). A C:N ratio less than 25:1 leads to mineralization and a ratio greater than 30:1 leads to immobilization of soil organic materials (Xu & Johnson 1997). The mean C:N ratio in all plots was less than 12.1:1 (Paper I, Table 8), indicating that there was sufficient N available for crop growth and for soil microorganisms to rapidly break down soil organic materials (including PHCs) as energy sources. Crops had a small impact on these measured soil properties (Paper I, Table 8).

5.4. Microbial community dynamics

5.4.1. Total genomic DNA
Soil total genomic DNA, as a proxy for total soil microbial biomass, was monitored for three years. The fodder galega significantly increased soil microbial DNA content over values in bare fallow, particularly after its growth stabilized (Paper II, Table 1), as found in the greenhouse experiment (Mikkonen *et al.* 2011a). However, the impact of oil on soil total DNA was not significant (Paper II, Table 1), in contrast with the greenhouse experiment where soil total DNA concentration increased in the presence of oil during the first 15 weeks (Mikkonen *et al.* 2011a), probably owing to two reasons. Firstly, our field soil was already "more saturated" with microbes, whereas a major part of the soil used for the greenhouse experiment was silt that had been taken from a deeper layer and stored for years before use, so probably without a very high microbial biomass; and secondly, the environmental conditions in the field are more complex than in the controlled conditions in the greenhouse.

5.4.2. Taxonomic composition of the bacterial community in the field
The 16S rRNA gene sequences revealed the co-existence of 32 phyla in the field soil, the most abundant being *Actinobacteria* (average abundance: 38.4%), *Proteobacteria* (29.3%), *Chloroflexi* (8.6%), *Acidobacteria* (8.3%), *Gematimonadetes* (5.8%), *Bacteroidetes* (2.3%), *Firmicutes* (2.0%), *Planctomycetes* (1.7%) and *Nitrospirae* (0.9%), despite of some differences in abundance of these phyla between oil-contaminated and control soils (Paper III, Fig. S1).

The majority of metabolic processes in a habitat are driven by the clearly dominant organisms (McArthur 2006). We focused on the two most dominant and diverse bacterial groups *Actinobacteria* and *Proteobacteria*. Among the phylum *Actinobacteria*, *Actinobacteria* class accounted for 53.5%, *Thermoleophilia* for 32.1%, *Acidimicrobiia* for 5.4%, *MB-A2-108* for 3.4% and unclassified for 5.6% of the total abundance. Among the phylum *Proteobacteria*, *Alphaproteobacteria* accounted for 63.3%, *Betaproteobacteria* for 18.1%, *Deltaproteobacteria* for 10.1%, *Gammaproteobacteria* for 8.0% and unclassified for 0.4%.

5.4.3. Bacterial diversity

Microbial diversity, which encompasses genetic variability of different taxa, the number (richness) and relative abundance (evenness) of taxa and the functional groups (guilds) in communities, characterizes the complexity and variability of biological organization at different levels (Torsvik & Øvreås 2002). The classes *Actinobacteria*, *Acidobacteria*, *Thermoleophilia*, *Alphaproteobacteria* and *Deltaproteobacteria* showed high diversity (average Shannon-Wiener index H' > 4.0, Paper III, Supplementary Table 3) in control soil. The effects of oil, crop and sampling times on bacterial diversity were assessed using both LH-PCR fingerprinting (Paper II, Table 2) and high-throughput sequencing approaches (Paper III, Fig. 1).
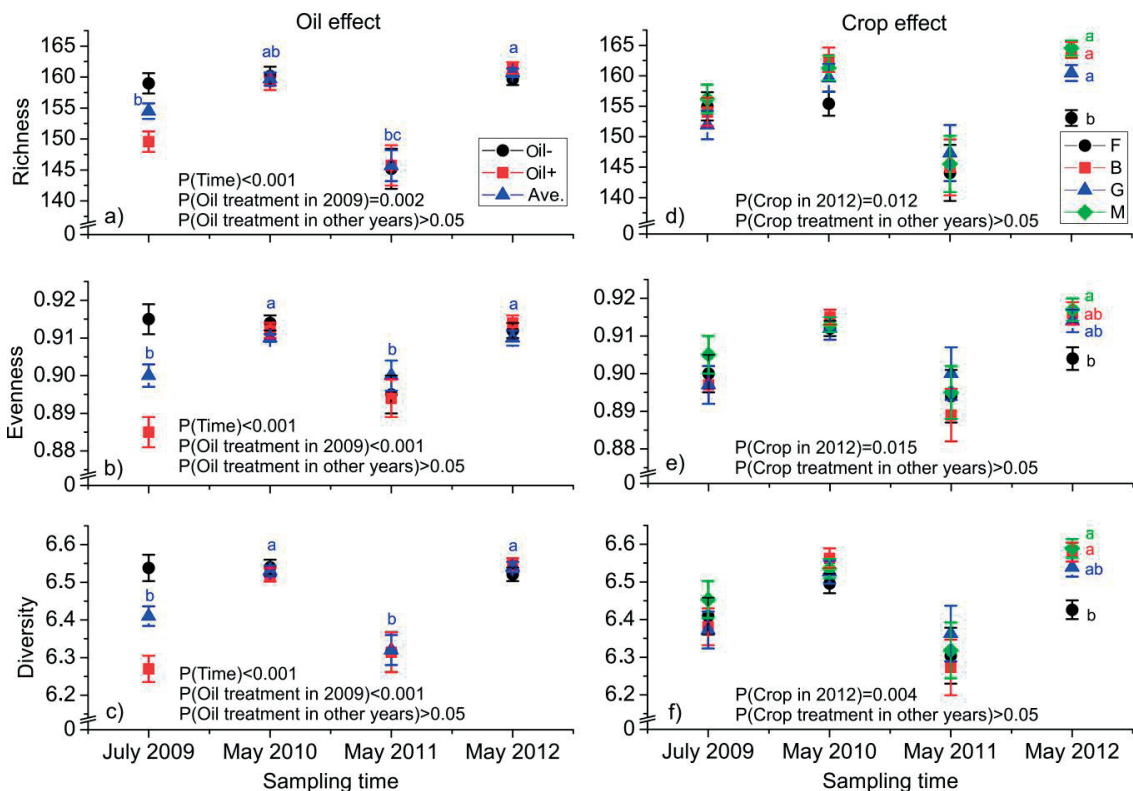
Fig. 9. Effects of oil and perennial crops on species richness, evenness and diversity over time based on bacterial16S rRNA gene amplicon sequencing data. Diversity: Shannon-Wiener index $H'= - \Sigma\, p_i \ln p_i$, where p_i is the proportion of the total count arising from the ith species; Richness: Margelef's index $d = (S-1)/\log N$, where S denotes total number of OTUs observed in each sample and N presents the total number of individuals (N=2977/normalized sample); Evenness: Pielou's index $J'=H'/H'\mathrm{max}=H'/\log S$. *P*-values were produced from univariate ANOVA based on a split-plot design. Different letters denote significant differences between the means based on Bonferroni post-hoc pairwise comparisons ($p < 0.05$). Abbreviations: "*Oil+*" oil-contaminated plots; "*Oil-*" control plots; "*Ave.*" mean value averaged from *oil+* and *oil-* plots; "*F*" bare fallow; "*B*" brome grass; "*G*" fodder galega; "*M*" mixture of brome grass and fodder galega.

The estimated evenness (Fig. 9b and e) and richness indices (Fig. 9e and f) showed the same oil-, time- and vegetation-dependent patterns as the Shannon diversity index (Fig. 9a and b). Oil significantly decreased the overall diversity, evenness and richness of the soil-borne bacterial community in the first year (Fig. 9). The reduced complexity in bacterial community following oil spike at the contamination

level of 7000 ppm likely implied an immediate ecotoxicity of used motor oil. However, the negative effect of oil on the overall bacterial diversity was not detected thereafter. Oil had contrasting effects on the diversity of the major bacterial classes (Paper III, Supplementary Table 3). The addition of oil consistently increased the diversity of *Gammaproteobacteria,* indicating the high competence of this class in the presence of oil. In contrast, the diversity of *Actinobacteria*, *Alphaproteobacteria*, *Acidobacteria* and *Deltaproteobacteria* decreased immediately following oil addition, suggesting their high sensitivity or lower relative competitivity of these classes in response to oil contamination. The negative effect of oil on these classes disappeared over time, indicating gradual recovery of these bacterial groups from contamination.

High-throughput sequencing identified the positive effect of vegetation (particularly, brome grass) on the complexity (diversity, evenness and richness) of bacterial community in 2012 (Paper III, Fig. 1b), but LH-PCR did not (Paper II, Table 2), demonstrating the higher sensitivity of high-throughput sequencing in the study of microbial ecology. The average value of bacterial diversity in pure galega plots was higher than that in the bare fallow, but not statistically significantly different, in agreement with finding of the recent greenhouse bioremediation experiment (Mikkonen *et al.* 2011a). However, the diversity of none of the major bacterial classes was significantly affected by crop treatment (UV ANOVA, $p > 0.05$).

Bacterial diversity of the overall assemblage and the sub-communities of major classes all exhibited a strong time-dependent pattern (Fig. 9 and Paper II, Table 2). The overall diversity was highest in the May 2010. The biodegradation efficiency of PHCs was found positively related to bacterial biodiversity. The observed sharp decrease in oil concentration was thus likely directly or indirectly associated with the observed high bacterial diversity in May 2010.

5.4.4. Bacterial communities
Microbial communities can be considered as functional units that are characterized by the sum of the metabolic properties of the microbial taxa involved (Wünsche *et al.* 1995). The multi-year bioremediation experiment based on a split-plot design enabled the comparative investigation of successional dynamics in bacterial communities over time i) between oil-contaminated and non-contaminated soil and ii) between vegetated soil and bare fallow.

Impact of time and oil contamination
The overall soil bacterial community went through a strong time- and oil contamination-dependent succession, as revealed by both LH-PCR fingerprinting (Paper II, Fig. 1) and high-throughput sequencing (Paper III, Fig. 2). In non-contaminated soil, the temporal dynamics in bacterial communities reflected the changing environmental conditions over time. The compositions of bacterial communities were distinctive between oil-contaminated soil and non-contaminated soils. The major phyla *Actinobacteria* contributed to an average proportion of 9.3%, *Acidobacteria* of 9.2%, *Protebacteria* of 7.9%, *Bacterioidetes* of 7.7%, *Gemmatimonadetes* of 6.2% and *Chloroflexi* of 5.6% of the total dissimilarity of bacterial assemblage between oil-contaminated and uncontaminated control samples over years at the phylum level (Paper III, Fig. S1). The effect of oil on bacterial overall community (Paper III, Fig. 2c) was most significant in 2009. This rapid and extensive alternation in the composition of bacterial communities induced by oil addition supports the notion that most bacterial groups in the assemblage were sensitive and not immediately resilient in response to

disturbance (Allison & Martiny 2008). Oil contamination also increased the multivariate dispersion among replicates, in comparison to controls (Paper III, supplementary Fig. S4b), agreeing with the demonstration that stress increases variability of multivariate ecological data (Warwick & Clarke 1993). Among all the sub-communities, determined at the class level, the succession of *Actinobacteria*, *Gammaproteobacteria* and *Betaproteobacteria* was highly oil-specific (Paper III, Fig. 3), indicating the high sensitivity of these taxonomic groups in response to changing oil contamination level over time.

Although the effect of oil on bacterial diversity was not detected after 2009, it was still significant on the composition of bacterial community at the end of the experiment. Since oil degradation was incomplete at the end of the experiment (Paper I), the multivariate community analysis is thus suggested to be more robust and reliable to reveal the persistent ecological impact of oil contamination than the diversity analysis.

Impact of crop cultivation
The presence of crops shifted bacterial community structure, particularly after the legume reached stable growth (Paper III, supplementary Fig. 3), in comparison with bare fallow, highly agreeing with the greenhouse experiment results (Mikkonen *et al.* 2011b). Plant biomass, plant species combination and diversity do not necessarily affect microbial biomass, colony-forming units of major microbial groups and microbial processes (Malý *et al.* 2000). Similarly, the patterns of bacterial community succession were identical in vegetated plots, independent of crop type. Specifically, crop physiochemical properties such as crop dry matter, chlorophyll content, total C and BNF yield explained small but significant proportions of bacterial community composition after the perennial crops had established their stable growth (Paper III, Supplementary Table 5), suggesting the importance of agricultural cultivation in influencing the structure of soil-borne bacterial communities. Changes in microbial community in the presence of vegetation are attributable to an input of nutrients from the decomposition of vegetation cover (Hobbie 2015). The crop physiological properties (e.g. biomass yield and BNF yield) were significantly associated with certain bacterial groups such as *Bacteroidetes*, *Armatimonadetes*, and *Fibrobacteres* (Paper III, Fig. 3). These crop-associated bacterial groups may be involved in the exploitation of bioavailable OM, colonizing aggregates and the decomposition of plant materials, polysaccharide-based substances, or photosynthetic biomass (Acosta-Martinez *et al.* 2008; Ransom-Jones *et al.* 2012; Lee *et al.* 2014). However, each of these crop-associated bacterial taxa accounted for only a comparatively low proportion of relative abundance in the overall bacterial community, thus resulting in the failure of detection using LH-PCR. According to this evidence, the ecological effect of different crops on soil bacterial communities would be expected to be more significant in a longer period of cultivation.

Impact of soil chemical properties
In this study, the measured soil chemical properties (pH, EC, total C and C:N ratio) explained a small but significant proportion of variations in bacterial communities monitored by both LH-PCR fingerprinting (Paper II, Fig. 2) and high-throughput sequencing (Paper III, Table 5). The importance of these soil parameters in shaping bacterial community and its succession was previously characterized by DNA fingerprinting and pyrosequencing analysis (Johnson *et al.* 2003; Lauber *et al.* 2009; Kuramae *et al.* 2010). As discussed earlier, however, the values of these soil chemical properties changed significantly in the presence of oil contamination. Since total C and C:N ratio were highly

correlated with oil concentration, they were removed from the multivariate regression model in sequential DISTLM analysis. The pH of the environment is one factor that directly affects the transport of materials across the cytoplasmic membrane (Sikkema *et al.* 1995). Most bacteria can grow in a range of pH from 4 to 9 with the optimum range from 6.5 to 8.5 (McArthur 2006). Changes in pH can lead to changes in functions and morphology of some bacteria. Although the average value of soil pH in this field soil was close to neutral in both oil-contaminated and control soils, it declined slightly with time, associated with alternations in the bacterial assemblage over time. Soil electrical conductivity is an indicator of nutrient availability, and since it decreased with oil contamination and time (Paper I, Table 8), it was found to correlate positively with bacterial community samples taken in 2009 (Paper III, Fig. 6).

Impact of seasonality
Seasonal changes in the bacterial community structure were clearly characterized by LH-PCR fingerprinting (Paper II, Fig. 1), agreeing with the finding that seasonality was the most influential factor influencing microbial community structure, provided that the experimental plots share the same soil type (Schutter *et al.* 2001). The seasonal effect was significant with higher diversity and richness exhibited in spring and lower diversity in autumn (Paper II, Table 2). The similarity of LH-PCR profiles of the autumn samples between contaminated and non-contaminated plots (Paper II, supplementary Fig. 3) suggested that the bacterial populations exhibited a similar development pattern in both treatments in autumn. Environmental conditions, especially the difference in soil temperature and moisture between spring and autumn, were likely to be the major factor behind the seasonal variation. Temperature plays an overriding role on all biological transformations (ATSDR 1999). As low temperature limits the oxidation of hydrocarbons in motor oil-contaminated soil (Ludzack & Kinkead 1956; Alkoaik & Ghaly 2006), the similarities in the LH-PCR profiles between oil-contaminated and non-contaminated soils might be associated with the low oil reduction rate in autumn. In addition to climatic factors (e.g. temperature, precipitation and wind) freezing and thawing cycles of the soil influence the availability of liquid water, which is essential to support the growth and metabolism of bacteria (Vázquez *et al.* 2009), resulting in seasonal effects.

However, there was tremendous unexplained variation within soil microbial communities in the field. Hence, unmeasured environmental factors such as soil organic matter, nutrient status, water and oxygen content might also play an important role in affecting bacterial communities as well as bioremediation efficiency. In a broader ecological view, organisms experience different competitors, different predators, and different parasites as seasons change and the structure of their communities change (McArthur 2006).

5.4.5. Oil-specific bacterial taxonomic groups
Different bacterial taxonomic groups behaved differently in response to oil contamination. In this study, the observed prevalence of certain taxonomic groups in oil-contaminated and uncontaminated control soils may be explained by the *r-K* scheme.

The majority of OTUs assigned by the LH-PCR community fingerprinting method were *K* strategists that are more stable and permanent members of the community. Nevertheless, there was one bacterial group (OTU 497 bp) that showed an immediate and sharp increase in relative abundance, accounting for 19% of the total peak area in July 2009 after oil addition (Fig. 10a). The dominance of the OTU

497 bp disappeared at the later sampling times. This OTU likely represented a group of *r* strategists that grew rapidly and took advantage of the added substrate (motor oil) quickly before other microorganisms adapted to the changed conditions. However, the exact composition of this oil-specific bacterial group was not identified. When hydrocarbons were less available, *r*-strategists grow more slowly, so *K*-strategists become more competitive in the community.
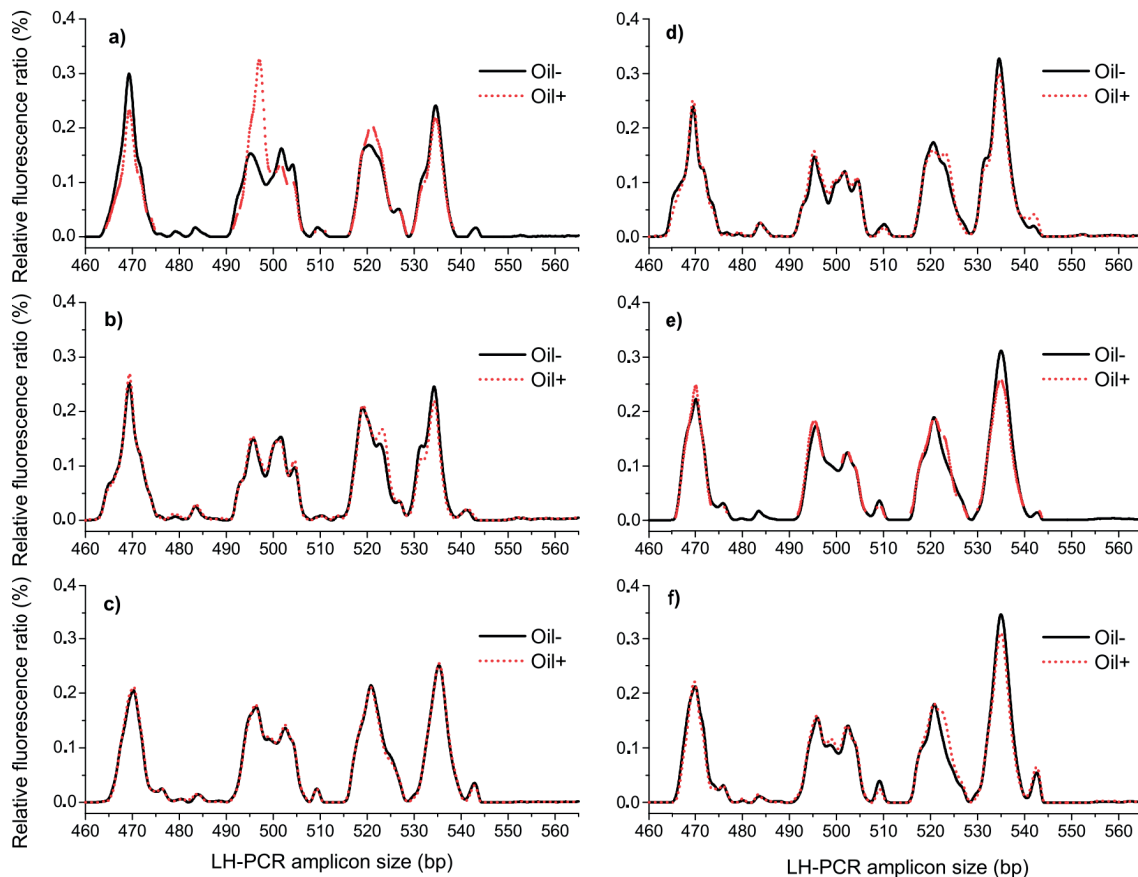

Fig. 10. Averaged curve-based LH-PCR profiles of the oil-treated and control plots over time: (a) July 2009, (b) May 2010, (c) November 2010, (d) May 2011, (e) May 2012 and (f) October 2012. Each bacterial profile was created from 16 profiles (4 crop treatments × 4 blocks) in PGPB-untreated plots at each sampling time using the Create Averaged Fingerprint script in BioNumerics software.

One important attribute of high-throughput sequencing is its potential to identify bacterial taxa that are responsible for shifts in community structure induced by the treatment factors at a fine resolution. The taxonomic treatment network provides an integrated view of the response patterns (positive, negative or neutral in terms of changes in the relative abundance) of bacterial taxa from phylum to OTU level in response to oil over time. The OTUs that responded positively to oil contamination were assigned mostly to the most diverse groups of *Actinobacteria* and *Proteobacteria*, in accordance with previous reports.

Among *Actinobacteria*, the oil-favored family *Nocardiaceae* (e.g. genera *Nocardia*, *Rhodococcus* and *Williamsia*) contributed an average proportion of 1.6%, *Nocardioidaceae* (e.g. *Nocardioides* and *Marmoricola*) of 0.5%, *Mycobacteriaceae* (e.g. *Mycobacterium*) of 0.5%, *Microbacteriaceae* (e.g. *Lysinimonas*) of 0.4% and *Geodermatophilaceae* (e.g. *Blastococcus*) of 0.4% to bacterial communities discriminating between oil-contaminated and non-contaminated control soil over years. The hydrocarbon-degradation capacity and mechanisms of these bacterial groups have been widely characterized in previous studies, e.g. (Nie *et al.* 2014; Tsuboi *et al.* 2015; Saul *et al.* 2005; Rodríguez-Mejía *et al.* 2008). For instance, the two alkane-catabolic genes (*alkB* and *CYP153*) were both detected in genomes of *Rhodococcus*, *Nocardia*, *Marmoricola and Mycobacteria* (Nie *et al.* 2014; Tsuboi *et al.* 2015). The increase in the relative abundance of these OTUs in those families coincided with the high rates of hydrocarbon biodegradation especially in the early-middle phase of bioremediation in the summer of 2009-2010 (Paper I), further confirming their oil-tolerance and competence in oil-contaminated soil. Specifically, *Rhodococcus* was the most abundant genus found in oil-contaminated soil one month after the oil spike. Particularly, the relative abundance of *Rhodococcus* OTU8 was 6.71% in oil-contaminated soil and 0.42% in non-contaminated soil in 2009. These early-phase oil-favored taxa are likely *r* strategists. Similarly, a study on catabolic bacterial communities in oil-polluted soil in the same boreal climate reported that the early-phase *alkB* OTUs detected displayed a high sequence similarity to *alkB* homologs of *Rhodococcus* (Mukherjee *et al.* 2015), as it carries multiple alkane catabolic genes (*alkB*) and the catechol 1,2-dioxygenase gene (C12O) for degrading both alkane and aromatic compounds, as well as three- to five-ring high-molecular-weight PAHs (Song *et al.* 2011; Shen *et al.* 2009; Andreoni *et al.* 2000; Whyte *et al.* 2002b).

The 16S rRNA gene amplicon length of the MiSeq OTU8 was only 459 bp, which was not long enough for the *in silico* LH-PCR analysis; however, OTU 8 showed 100% identity to *Rhodococcus erythropolis* using Basic Local Alignment Search Tool (BLAST, http://blast.ncbi.nlm.nih.gov/Blast.cgi). The LH-PCR primers of fD1 (5'-AGAGTTTGATCCTGGCT-CAG-3') and PRUN518r (5'-ATTACCGCGGCTGCTGG-3') were used to cut the partial 16S rRNA gene sequences of three *Rhodococcus erythropolis* strains 3C-9 (GenBank: DQ000156.2), K14 (GenBank: KF976878.1) and DSM 25948 (GenBank: KM047507.1) *in silico* using the online program Sequence Extractor (http://www.bioinformatics.org/seqext/). The *in silico* LH-PCR analysis produced the same amplicon length of 499 bp for all of the three strains. Technically, 1-3 bp horizontal shifts of the sequences sometimes occur in practice using the LH-PCR method, because the G+C% of the internal size standards can affect the accuracy (Anu Mikkonen, personal communication, January 11, 2016). Hence, the LH-PCR OTU 497 bp that had an immediate abundance increase in the presence of oil most likely represents MiSeq OTU8 (*Rhodococcus erythropolis)*, although different bacterial taxa might have the same LH-PCR amplicon length. In addition, we observed increasingly higher relative abundance of OTU29 and OTU132, both of which belong to the family *Microbacteriaceae,* in oil-contaminated soil over time, reflecting their increasing competence in the oil-impacted community.

Among *Proteobacteria*, *Alpha-*, *Beta-*, *Delta-* and *Gamma-* subdivisions accounted for 4.2%, 2.7%, 1.5% and 3.1% of the total dissimilarity between oil-contaminated and control plots at the class level, respectively. The comparatively less abundant subdivisions of *Gammaproteobacteria* and *Betaproteobacteria* displayed clear successional changes, elucidating the sensitivity in response to

changing oil concentration and composition during bioremediation. Similarly, the importance of the two classes *Gammaproteobacteria* and *Betaproteobacteria* in PAH-degradation was addressed by a stable isotope probing-based study (Singleton *et al.* 2006). Among *Gammaproteobacteria*, the oil-favored family *Xanthomonadaceae* (e.g. genera *Arenimonas* and *Lysobacter*) contributed 0.3% and *KCM-B-112* 0.9% to dissimilarity of bacterial communities between oil-contaminated and control soil. *Arenimonas* and *Lysobacter* are closely related to *Pseudomonas*, *Xanthomonas* and *Thermomonas*, which are well-known aerobic hydrocarbon degraders (Young *et al.* 2007; Akbari & Ghoshal 2015). The family *KCM-B-112* that includes a number of uncultured soil bacteria positively responded to oil contamination throughout the bioremediation process, suggesting its high resistance and competence in oil-contaminated soil. Among *Betaproteobacteria*, the oil-selected family *Comamonadaceae* (e.g. genera *Aquabacterium*, *Variovorax* and *Rhizobacter*) contributed an average proportion of 0.5% and *Oxalobacteriaceae* (e.g. genera *Noviherbaspirillum* and *Massila)* 0.4% to the difference in the bacterial communities between oil-contaminated and control soil over years. In particular, *Aquabacterium* spp. that harbor *alkB* genes and are able to utilize both liquid and solid alkanes (Aburto-Medina *et al.* 2012; Giebler *et al.* 2013; Masuda *et al.* 2014), showed significant oil-specific trajectories over time, indicating its high competence in bacterial community in the presence of oil, especially at the early stage.

This study showed that *Alphaproteobacteria* were more abundant in control soil than in oil-contaminated soil, in accordance with the comparative phylogenetic study (Labbe *et al.* 2007), but inconsistent with the studies of Mills *et al.* (2003) and Vinas *et al.* (2005). Since the soil condition, pollutant characteristics and original composition of alphaproteobacterial communities varied between studies, it is reasonable to uncover contradictory phenomena. Among *Alphaproteobacteria*, the families such as *Xanthobacteraceae* (contribution 0.5%, *esp.* genus *Pseudolabrys*) and *DA111* (0.2%) responded negatively, whereas families such as *Bradyrhizobiaceae* (contribution 0.5%) and *Caulobacteraceae* (0.3%, e.g. *Phenylobacterium*) responded positively to oil contamination through time. *Phenylobacterium* was the only alphaproteobacterial genus that consistently showed higher abundance in the presence of oil, according with the finding that alkane- and PAH-degrading *Phenylobacterium* strains were detected even in cold environments (Giebler *et al.* 2013; Yang *et al.* 2014). The alkane- and anthracene-degrading potential of *Rhizobiales* was widely reported (Jones *et al.* 2011; Dunlevy *et al.* 2013; Giebler *et al.* 2013). We observed the prevalence of *Rhizobiales* (e.g. family *Bradyrhizobiaceae*) in oil-contaminated soil, especially at the middle-late phase.

Earlier culture-based studies proposed to use *Gammaproteobacteria*, particularly cultivable genera such as *Pseudomonas*, *Acinetobacter* and *Enterobacteria*, as bio-indicators to detect the impact on the bacterial community in PAH-contaminated agricultural soils due to their PAH-degrading capacity and better resistance to PAH toxicity (Labbe *et al.* 2007; Vázquez *et al.* 2009; Lors *et al.* 2010; Niepceron *et al.* 2013). However, the importance of these genera in hydrocarbon degradation was overestimated in the culture-dependent enrichment studies, as they are rare organisms in microbial communities (Shade *et al.* 2012), accounting for less than 0.1% in relative abundance in the overall bacterial community in our soil. Therefore, to choose suitable bio-indicators for evaluating oil contamination and remediation is case-specific. The time-series analysis demonstrated that the OTUs with significantly different oil-specific trajectories over time (*Lysinimonas* OTU132, *Microbacteriaceae* OTU29, *Marmoricola* OTU541 and *Aquabacterium* OTU188) were all favored by oil (Paper III, Fig.

5), revealing the high sensitivity of these oil-favored OTUs to the changing hydrocarbon composition in different phases of bioremediation. In contrast, the OTUs that were negatively or not affected by the initial oil input (e.g. *Pseudolabrys* OTU2) all followed similar patterns in both oil-contaminated and non-contaminated plots over time. This result agrees with the recent finding of Smith *et al.* (2015) that the bacterial strains that are most useful for detecting soil contaminants such as oil and uranium are known to interact with these substrates. Hence, the genera of these four oil-specific OTUs with different temporal patterns show potential as bio-indicators to monitor the ecological impact of oil contamination and the following long-term recovery process in boreal soil.

5.5. PGPB effect

The effect of plant growth promoting bacteria on crop growth, soil quality, oil reduction, soil total genomic DNA and bacterial community composition was evaluated. In the present study, PGPB strains enhanced BNF efficiency of fodder galega in terms of %Ndfa and BNF yield (Paper I, Table 7), in agreement with the corresponding pot experiment (Egamberdieva *et al.* 2010). The promotion of BNF by PGPB was more significant in the legume-grass mixture plots than in pure galega plots, suggesting that it was triggered mainly by the nutrient requirement of the legume. When there was nutrient competition from the grass in the mixture plots, the nitrogen fixation efficiency of the legume was improved by the PGPB strains. However, effect of PGPB on crop DM yield was not significant.

PGPB inoculation showed no impact on soil properties (data not shown), oil concentration (data not shown) or bacterial communities (Paper II, Table S1) in our field, in contrast to other studies where PGPB enhanced the rhizoremediation of polluted soils (Pajuelo *et al.* 2011; Bhattacharyya & Jha 2012; Vershinina *et al.* 2012). These results were not surprising, as PGPB strains that were supposed to exist mostly in the rhizosphere would have minor impact on those parameters measured from the bulk soil. In addition, the PGPB strains used, their oil tolerance ability, the local field factors and the hydrocarbon composition were all different from those in other studies.

6. CONCLUSIONS AND FUTURE PROSPECTS

In this study, the ecological impacts of motor oil contamination (7000 ppm) were evaluated in the field, considering four aspects: 1) the oil concentration in soil, 2) soil physicochemical properties, 3) crop physiological properties and 4) bacterial community dynamics during a four-year period. The results suggested that the effect of oil was initially severe, but diminished with time.

It seemed that physical removal (volatilization and leaching) was the main processes of oil reduction during the first month following the oil spike, accounting for 29% of oil reduction, according to the first-order kinetic model. Afterwards, biodegradation was likely the main process. The oil reduction rate was higher in the first two summers, after which it became very slow. Soil conditions in our field were optimal for biodegradation of hydrocarbons, with neutral pH and an optimal C:N ratio. Soil physiochemical properties were buffered very well under oil treatment, except for the changes of soil electrical conductivity and the total carbon content in the first year. Both crops (*Galega orientalis* and *Bromus inermis*) were highly tolerant of an oil contamination of 7000 ppm and, surprisingly, the oil contamination triggered higher dry matter production of both crops. Brome grass in the mixture plots (without synthetic N fertilizer) had better yield than in pure plots (with synthetic N fertilizer), attributed to biological nitrogen fixation of fodder galega with its associated rhizobia. Therefore, oil-tolerant perennial galega-brome grass mixture, without any additional N-fertilizer can be cultivated and managed as bioenergy crops in oil-contaminated soil.

Bacterial overall community and sub-communities of *Actinobacteria*, *Gammaproteobacteria* and *Betaproteobacteria* underwent successional changes in oil-contaminated soil, indicating their sensitivity to changing oil concentration over time. The time-series analysis demonstrated that the four oil-favored OTUs (*Lysinimonas* OTU132, *Microbacteriaceae* OTU29, *Marmoricola* OTU541 and *Aquabacterium* OTU188) had significantly different temporal patterns. Therefore, together with the bacterial overall community and sub-communities of *Actinobacteria*, *Gammaproteobacteria* and *Betaproteobacteria*, these four oil-specific OTUs can be used as sensitive bio-indicators of oil contamination and remediation over time, with regard to the comparable controls. Crops (fodder galega, brome grass and their mixture) had no significant effect on microbial community composition until they fully established their roots in soil. Other unmeasured field variables (e.g. climatic factors) might have had an overriding influence on the removal of oil and might explain high variations in bacterial communities. The temporal effect was overriding in altering bacterial communities, regardless of oil contamination. Therefore, studies restricted to a single snapshot of time, as well as studies of succession without any non-contaminated reference, can hardly effectively give insight into the bacterial populations in the communities, if the goal is to characterize the long-term impacts of oil contamination.

LH-PCR revealed highly consistent patterns in bacterial community dynamics, as did the high-throughput sequencing, despite the inherent resolution limit of the former for the identification of bacterial groups. Specifically, by comparing the initially oil-favored OTUs between the LH-PCR and MiSeq techniques, it is further suggested that the LH-PCR fingerprinting of amplified bacterial 16S rRNA fragments is a reliable and cost-effective technique to profile the structure of bacterial communities for pre-screening of soil DNA samples for further in-depth sequencing analysis.

After the monitoring period of 40 months, oil was not completely removed. Nevertheless, the bioremediation process can be considered satisfactory, since all the measured parameters, e.g. crop growth, soil chemical properties and bacterial community, gradually recovered over time. Taking all together, the combination of these two perennial crops is a good candidate for vegetation of abandoned oil-contaminated boreal sites or contaminated soil dumping grounds for simultaneous bioremediation and bioenergy production.

The co-existence and responses of bacterial communities to oil contamination were variable as part of an ecological network that involved diverse direct and indirect interactions with each other and the environment. Although some bacterial groups showed negative or neutral response to oil contamination, it cannot be ruled out that these bacteria were functionally important for other ecosystem processes, such as biogeochemical cycling, energy flow and water cycling. In the future, NGS-based transcriptome and gene expression analysis can be incorporated to elucidate the functional significance of the observed communities. Furthermore, in a broader aspect of sustainability, we can perform a meta-analysis by combining similar case studies to identify the threshold of "a safe operating space", below which the risk of a change in ecosystem services is likely to remain low, to provide guidance for policy making in pollutant-remediation activities in boreal regions.

ACKNOWLEDGEMENTS

REFERENCES

Aburto-Medina A, Adetutu EM, Aleer S, et al (2012) Comparison of indigenous and exogenous microbial populations during slurry phase biodegradation of long-term hydrocarbon-contaminated soil. *Biodegradation,* **23**, 813-822.

Acosta-Martinez 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 Biology and Biochemistry,* **40**, 2762-2770.

Adamovich A, Adamovicha O, Vliegher Ad, Carlier L (2007) Forage quality dynamics and productivity of fodder galega-grass swards. Permanent and temporary grassland: plant, environment and economy. Proceedings of the 14th Symposium of the European Grassland Federation, Ghent, Belgium, 3–5 September 2007. *Belgian Society for Grassland and Forage Crop*s, pp. 181–184.

Agency for Toxic Substances and Disease Registry (ATSDR) (1999). Toxicological profile for total petroleum hydrocarbons (PHC). Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service. Available at: http://www.atsdr.cdc.gov/ToxProfiles/tp123.pdf. Last visited: March 20, 2012.

Akbari A, Ghoshal S (2015) Effects of diurnal temperature variation on microbial community and petroleum hydrocarbon biodegradation in contaminated soils from a sub-Arctic site. *Environmental Microbiology*, **89**(1), 62–69.

Alkio M, Tabuchi TM, Wang X, Colon-Carmona A (2005) Stress responses to polycyclic aromatic hydrocarbons in Arabidopsis include growth inhibition and hypersensitive response-like symptoms. *Journal of Experimental Botany,* **56**, 2983-2994.

Alkoaik F, Ghaly A (2006) Thermobioremediation of soil contaminated with used motor oil. *Energy Sources, Part A,* **28**, 487-500.

Allison SD, Martiny JB (2008) Colloquium paper: resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America,* **105 Suppl 1**, 11512-11519.

Anderson MJ (2006) Distance- Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics,* **62**, 245-253.

Anderson MJ (2001a) A new method for non- parametric multivariate analysis of variance. *Austral Ecology,* **26**, 32-46.

Anderson MJ, Robinson J (2003) Generalized discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics,* **45**, 301-318.

Anderson MJ, Willis TJ (2003) Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. *Ecology,* **84**, 511-525.

Anderson M, Braak CT (2003) Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation,* **73**, 85-113.

Anderson M, Gorley RN, Clarke RK (2008) *Permanova for Primer: Guide to Software and Statistical Methods*. PRIMER-E: Plymouth, UK, pp. 87–104.

Anderson MJ (2001b) Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences,* **58**, 626-639.

Andreoni V, Bernasconi S, Colombo M, Van Beilen JB, Cavalca L (2000) Detection of genes for alkane and naphthalene catabolism in Rhodococcus sp. strain 1BN. *Environmental Microbiology,* **2**, 572-577.

Balba M, Al-Awadhi N, Al-Daher R (1998) Bioremediation of oil-contaminated soil: microbiological methods for feasibility assessment and field evaluation. *Journal of Microbiological Methods,* **32**, 155-164.

Bhattacharyya PN, Jha DK (2012) Plant growth-promoting rhizobacteria (PGPR): emergence in agriculture. *World Journal of Microbiology & Biotechnology,* **28**, 1327-1350.

Boopathy R (2000) Factors limiting bioremediation technologies. *Bioresource Technology,* **74**, 63-67.

Borneman J, Triplett EW (1997) Molecular microbial diversity in soils from eastern Amazonia: evidence for unusual microorganisms and microbial population shifts associated with deforestation. *Applied and Environmental Microbiology,* **63**, 2647-2653.

Börner F, Gruhne M, Schön J (1993) Contamination indications derived from electrical properties in the low frequency RANGE1. *Geophysical Prospecting,* **41**, 83-98.

Bragg JR, Prince RC, Harner EJ, Atlas RM (1994) Effectiveness of bioremediation for the Exxon Valdez oil spill. *Nature,* **368**, 413-418.

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs,* **27**, 325-349.

Chatfield C, Collins AJ (1980) Principal component analysis. In: *Introduction to multivariate analysis*, Springer, pp. 57-81.

Chaudhary P, Sharma R, Singh SB, Nain L (2011) Bioremediation of PAH by Streptomyces sp. *Bulletin of environmental contamination and toxicology,* **86**, 268-271.

Chaudhry GR (1994) *Biological degradation and bioremediation of toxic chemicals*, Springer, pp. 204-205.

Chiapusio G, Pujol S, Toussaint M, Badot P, Binet P (2007) Phenanthrene toxicity and dissipation in rhizosphere of grassland plants (Lolium perenne L. and Trifolium pratense L.) in three spiked soils. *Plant and Soil,* **294**, 103-112.

Choudhury ATMA and Kennedy IR (2005) Nitrogen Fertilizer Losses from Rice Soils and Control of Environmental Pollution Problems. *Communications in Soil Science and Plant Analysis*, **36**, 1625–1639.

Clarke KR (1990) Comparisons of dominance curves. *Journal of experimental marine biology and ecology,* **138**, 143-157.

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology,* **18**, 117-117.

Clarke K, Green R (1988) Statistical design and analysis for a" biological effects" study. *Marine Ecology Progress Series,* **46**, 213-226.

Clarke K, Warwick R (2001) An approach to statistical analysis and interpretation. *Change in Marine Communities,* **2**.

Clarke KR, Gorley RN, Somerfield PJ, Warwick RM (2014). *Change in marine communities*: *an approach to statistical analysis and interpretation*, *3rd edition*. PRIMER-E, Plymouth.

Compant S, Duffy B, Nowak J, Clement C, Barka EA (2005) Use of plant growth-promoting bacteria for biocontrol of plant diseases: principles, mechanisms of action, and future prospects. *Applied and Environmental Microbiology,* **71**, 4951-4959.

Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America,* **99**, 10494-10499.

da C Jesus E, Marsh TL, Tiedje JM, de S Moreira, Fatima M (2009) Changes in land use alter the structure of bacterial communities in Western Amazon soils. *The ISME journal,* **3**, 1004-1011.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics,* **12**, 499-510.

Dexter A (2004) Soil physical quality: Part I. Theory, effects of soil texture, density, and organic matter, and effects on root growth. *Geoderma,* **120**, 201-214.

Dominguez-Rosado E, Pichtel J, Coughlin M (2004) Phytoremediation of soil contaminated with used motor oil: I. Enhanced microbial activities from laboratory and growth chamber studies. *Environmental Engineering Science,* **21**, 157-168.

Dunlevy SR, Singleton DR, Aitken MD (2013) Biostimulation reveals functional redundancy of anthracene-degrading bacteria in polycyclic aromatic hydrocarbon-contaminated soil. *Environmental Engineering Science,* **30**, 697-705.

Egamberdieva D, Berg G, Lindström K, Räsänen L (2010) Co-inoculation of< i> Pseudomonas spp. with< i> Rhizobium improves growth and symbiotic performance of fodder galega (< i> Galega orientalis Lam.). *European Journal of Soil Biology,* **46**, 269-272.

European Environmental Agency (EEA) (2014). Progress in management of contaminated sites. Available at: http://www.eea.europa.eu/data-and-maps/indicators/progress-in-management-of-contaminated-sites-3/assessment. Last visited: November 20, 2015.

Eriksson M, Ka JO, Mohn WW (2001) Effects of low temperature and freeze-thaw cycles on hydrocarbon biodegradation in Arctic tundra soil. *Applied and Environmental Microbiology,* **67**, 5107-5112.

Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America,* **103**, 626-631.

Fisher MM, Triplett EW (1999) Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology,* **65**, 4630-4636.

Franche C, Lindström K, Elmerich C (2009) Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant Soil*, **321**, 35–59.

Fujita K, Ofosu-Budu K, Ogata S (1992) Biological nitrogen fixation in mixed legume-cereal cropping systems. In: *Biological Nitrogen Fixation for Sustainable Agriculture*, Springer, pp. 155-175.

Gawel LJ, North Dakota Industrial Commission (2006) *A guide for remediation of salt/hydrocarbon impacted soil* North Dakota Industrial Commission, Department of Mineral Resources.

Gerhardt KE, Huang X, Glick BR, Greenberg BM (2009) Phytoremediation and rhizoremediation of organic soil contaminants: potential and challenges. *Plant Science,* **176**, 20-30.

Giebler J, Wick LY, Chatzinotas A, Harms H (2013) Alkane-degrading bacteria at the soil-litter interface: comparing isolates with T-RFLP-based community profiles. *FEMS microbiology ecology,* **86**, 45-58.

Goodfellow M, Manfio G, Chun J (1997) Towards a practical species concept for cultivable bacteria.

Gower JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika,* **53**, 325-338.

Halling MA, Hopkins A, Nissinen O, Paul C, Tuori M, Soelter U (2002) Forage legumes– productivity and composition. *Legume Silages for Animal Production-LEGSIL*, 5-15.

Hamamura N, Olson SH, Ward DM, Inskeep WP (2006) Microbial population dynamics associated with crude-oil biodegradation in diverse soils. *Applied and Environmental Microbiology,* **72**, 6316-6324.

Harismendy O, Ng PC, Strausberg RL, et al (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome biol,* **10**, R32.

Hatzinger PB, Alexander M (1995) Effect of aging of chemicals in soil on their biodegradability and extractability. *Environmental science & technology,* **29**, 537-545.

Head IM, Jones DM, Röling WF (2006) Marine microorganisms make a meal of oil. *Nature Reviews Microbiology,* **4**, 173-182.

Hobbie SE (2015) Plant species effects on nutrient cycling: revisiting litter feedbacks. *Trends in ecology & evolution*.

Hong SH, Ryu H, Kim J, Cho K (2011) Rhizoremediation of diesel-contaminated soil using the plant growth-promoting rhizobacterium Gordonia sp. S2RP-17. *Biodegradation,* **22**, 593-601.

Huesemann MH (1997) Incomplete hydrocarbon biodegradation in contaminated soils: limitations in bioavailability or inherent recalcitrance? *Bioremediation Journal,* **1**, 27-39.

Huesemann MH (1995) Predictive model for estimating the extent of petroleum hydrocarbon biodegradation in contaminated soils. *Environmental science & technology,* **29**, 7-18.

Irwin, R.J., M. VanMouwerik, L. Stevens, M.D. Seese, and W. Basham. (1997). Environmental Contaminants Encyclopedia. National Park Service, Water Resources Division, Fort Collins, Colorado. Distributed within the Federal Government as an Electronic Document (Projected public availability on the internet or NTIS: 1998).

ISO 16703:2004. Soil Quality –Determination of content of hydrocarbon in the range C10 to C40 by gas chromatography. ISO 2004.

Johnson M, Lee K, Scow K (2003) DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. *Geoderma,* **114**, 279-303.

Jones MD, Crandell DW, Singleton DR, Aitken MD (2011) Stable- isotope probing of the polycyclic aromatic hydrocarbon- degrading bacterial guild in a contaminated soil. *Environmental microbiology,* **13**, 2623-2632.

Jørgensen K, Puustinen J, Suortti A (2000) Bioremediation of petroleum hydrocarbon-contaminated soil by composting in biopiles. *Environmental pollution,* **107**, 245-254.

Jussila MM, Jurgens G, Lindström K, Suominen L (2006) Genetic diversity of culturable bacteria in oil-contaminated rhizosphere of< i> Galega orientalis. *Environmental pollution,* **139**, 244-257.

Kaksonen A, Jussila M, Lindström K, Suominen L (2006) Rhizosphere effect of Galega orientalis in oil-contaminated soil. *Soil Biology and Biochemistry,* **38**, 817-827.

Kamath R, Rentz J, Schnoor JL, Alvarez P (2004) Phytoremediation of hydrocarbon-contaminated soils: principles and applications. *Studies in surface science and catalysis,* **151**, 447-478.

Kawasaki A, Watson ER, Kertesz MA (2012) Indirect effects of polycyclic aromatic hydrocarbon contamination on microbial communities in legume and grass rhizospheres. *Plant and Soil,* **358**, 169-182.

Kellenberger E (2001) Exploring the unknown: The silent revolution of microbiology. *EMBO Reports*, **2**, 5–7.

Kindt R, Coe R (2005) *Tree diversity analysis: a manual and software for common statistical methods for ecological and biodiversity studies* World Agroforestry Centre.

Kloos K, Munch JC, Schloter M (2006) A new method for the detection of alkane-monooxygenase homologous genes (alkB) in soils based on PCR-hybridization. *Journal of microbiological methods,* **66**, 486-496.

Kruskal JB, Wish M (1978) *Multidimensional scaling* Sage.

Kryževičienė A, Jasinskas A, Gulbinas A (2008) Perennial grasses as a source of bioenergy in Lithuania. *Agron.Res,* **6**, 229-239.

Kuramae EE, Gamper HA, Yergeau E, et al (2010) Microbial secondary succession in a chronosequence of chalk grasslands. *The ISME journal,* **4**, 711-715.

Labbe D, Margesin R, Schinner F, Whyte LG, Greer CW (2007) Comparative phylogenetic analysis of microbial communities in pristine and hydrocarbon-contaminated Alpine soils. *FEMS microbiology ecology,* **59**, 466-475.

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. *Applied and Environmental Microbiology,* **75**, 5111-5120.

Leahy JG, Colwell RR (1990) Microbial degradation of hydrocarbons in the environment. *Microbiological reviews,* **54**, 305-315.

Lee KC, Dunfield PF, Stott MB (2014) The Phylum Armatimonadetes. In: *The Prokaryotes*, Springer, pp. 447-458.

Legendre P, Anderson MJ (1999) Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecological Monographs,* **69**, 1-24.

Liang Y, Van Nostrand JD, Deng Y, et al (2011) Functional gene diversity of soil microbial communities from five oil-contaminated fields in China. *The ISME journal,* **5**, 403-413.

Liao J, Wang J, Huang Y (2015) Bacterial Community Features Are Shaped by Geographic Location, Physicochemical Properties, and Oil Contamination of Soil in Main Oil Fields of China. *Microbial ecology*, 1-10.

Lindstrom K, Jussila MM, Hintsa H, et al (2003) Potential of the Galega-Rhizobium galegae system for bioremediation of oil-contaminated soil. *Food Technology and Biotechnology,* **41**, 11-16.

Liu WT, Marsh TL, Cheng H, Forney LJ (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology,* **63**, 4516-4522.

Logares R, Sunagawa S, Salazar G, et al (2014) Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environmental microbiology,* **16**, 2659-2671.

Loreau M, Naeem S, Inchausti P, et al (2001) Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science (New York, N.Y.),* **294**, 804-808.

Lors C, Ryngaert A, Périé F, Diels L, Damidot D (2010) Evolution of bacterial community during bioremediation of PAHs in a coal tar contaminated soil. *Chemosphere,* **81**, 1263-1271.

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* **15**, 550.

Ludzack F, Kinkead D (1956) Persistence of oily wastes in polluted water under aerobic conditions. *Industrial & Engineering Chemistry,* **48**, 263-267.

Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PloS one,* **7**, e30087.

Malý S, Korthals G, Van Dijk C, Van der Putten W, De Boer W (2000) Effect of vegetation manipulation of abandoned arable land on soil microbial properties. *Biology and Fertility of Soils,* **31**, 121-127.

Mandri T, Lin J (2007) Isolation and characterization of engine oil degrading indigenous microorganisms in Kwazulu-Natal, South Africa. *African journal of Biotechnology,* **6**.

Mao Y, Yannarell AC, Davis SC, Mackie RI (2013) Impact of different bioenergy crops on N-cycling bacterial and archaeal communities in soil. *Environmental microbiology,* **15**, 928-942.

Margesin R, Labbe D, Schinner F, Greer CW, Whyte LG (2003) Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine Alpine soils. *Applied and Environmental Microbiology,* **69**, 3085-3092.

Masuda H, Shiwa Y, Yoshikawa H, Zylstra GJ (2014) Draft Genome Sequence of the Versatile Alkane-Degrading Bacterium Aquabacterium sp. Strain NJ1. *Genome announcements,* **2**, 10.1128/genomeA.01271-14.

McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology,* **82**, 290-297.

McArthur JV (2006) *Microbial ecology: an evolutionary approach* Academic press, 432 pages.

McCune B, Grace JB, Urban DL (2002) *Analysis of ecological communities* MjM software design Gleneden Beach, OR.

Metzker ML (2010) Sequencing technologies—the next generation. *Nature reviews genetics,* **11**, 31-46.

Michelsen TC, Boyce CP (1993) Cleanup standards for petroleum hydrocarbons. Part 1. Review of methods and recent developments. *Soil and Sediment Contamination,* **2**, 109-124.

Mikkonen A, Kondo E, Lappi K, et al (2011a) Contaminant and plant-derived changes in soil chemical and microbiological indicators during fuel oil rhizoremediation with< i> Galega orientalis. *Geoderma,* **160**, 336-346.

Mikkonen A, Lappi K, Wallenius K, Lindström K, Suominen L (2011b) Ecological inference on bacterial succession using curve- based community fingerprint data analysis, demonstrated with rhizoremediation experiment. *FEMS microbiology ecology,* **78**, 604-616.

Mills DK, Fitzgerald K, Litchfield CD, Gillevet PM (2003) A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils. *Journal of microbiological methods,* **54**, 57-74.

Mukherjee S, Heinonen M, Dequvire M, Sipilä T, Pulkkinen P, Yrjälä K (2013) Secondary succession of bacterial communities and co-occurrence of phylotypes in oil-polluted Populus rhizosphere. *Soil Biology and Biochemistry,* **58**, 188-197.

Mukherjee S, Juottonen H, Siivonen P, et al (2014) Spatial patterns of microbial diversity and activity in an aged creosote-contaminated site. *The ISME journal,* **8**, 2131-2142.

Mukherjee S, Sipilä T, Pulkkinen P, Yrjälä K (2015) Secondary successional trajectories of structural and catabolic bacterial communities in oil- polluted soil planted with hybrid poplar. *Molecular ecology,* **24**, 628-642.

Nie Y, Chi C, Fang H, et al (2014) Diverse alkane hydroxylase genes in microorganisms and environments. *Scientific reports,* **4**.

Niepceron M, Martin-Laurent F, Crampon M, et al (2013) GammaProteobacteria as a potential bioindicator of a multiple contamination by polycyclic aromatic hydrocarbons (PAHs) in agricultural soils. *Environmental Pollution,* **180**, 199-205.

Nocentini M, Pinelli D, Fava F (2000) Bioremediation of a soil contaminated by hydrocarbon mixtures: the residual concentration problem. *Chemosphere,* **41**, 1115-1123.

Pajuelo E, Rodríguez-Llorente ID, Lafuente A, Caviedes MÁ (2011) Legume–rhizobium symbioses as a tool for bioremediation of heavy metal polluted soils. *Biomanagement of Metal-contaminated Soils*, Springer, pp. 95-123.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara R et al (2015). Vegan: community ecology package [online]. R package version 2.2-1. Available at: http://CRAN.R-project.org/package=vegan. Last visited: November 20, 2015.

Pham TV, Piersma SR, Warmoes M, Jimenez CR (2010) On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics (Oxford, England),* **26**, 363-369.

Pianka ER (1970) On r-and K-selection. *American naturalist,* **104**(940), 592-597.

Pyy, O., Haavisto, T., Niskala, K., & Silvola, M. (2013). Contaminated land in Finland- Report 2013, Finnish Environmental Institute. Available at http://ymparistonc.edelkey.net/en-US/Consumption_and_production/Contaminated_soil_sites. Last visited: November 20, 2015.

Radwan S, Sorkhoh N, Elnemr I (1995) Oil Biodegradation Around Roots. *Nature,* **376**, 302-302.

Ransom-Jones E, Jones DL, McCarthy AJ, McDonald JE (2012) The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microbial Ecology,* **63**, 267-281.

Riser-Roberts E (1998) Remediation of petroleum contaminated soils: biological, physical, and chemical processes CRC press, 576 pages.

Robertson SJ, McGill WB, Massicotte HB, Rutherford PM (2007) Petroleum hydrocarbon contamination in boreal forest soils: a mycorrhizal ecosystems perspective. *Biological Reviews,* **82**, 213-240.

Rodríguez- Mejía JL, Martínez- Anaya C, Folch- Mallol JL, Dantán- González E (2008) A two-step electrodialysis method for DNA purification from polluted metallic environmental samples. *Electrophoresis,* **29**, 3239-3244.

Romantschuk M, Sarand I, Petänen T, et al (2000) Means to improve the effect of in situ bioremediation of contaminated soil: an overview of novel approaches. *Environmental Pollution,* **107**, 179-185.

Ros M, Rodríguez I, García C, Hernández MT (2014) Bacterial community in semiarid hydrocarbon contaminated soils treated by aeration and organic amendments. *International Biodeterioration & Biodegradation,* **94**, 200-206.

Sanderson M, Skinner R, Barker D, Edwards G, Tracy B, Wedin D (2004) Plant species diversity and management of temperate forage and grazing land ecosystems. *Crop Science,* **44**, 1132-1144.

Sarkar D, Ferguson M, Datta R, Birnbaum S (2005) Bioremediation of petroleum hydrocarbons in contaminated soils: comparison of biosolids addition, carbon supplementation, and monitored natural attenuation. *Environmental Pollution,* **136**, 187-195.

Saul DJ, Aislabie JM, Brown CE, Harris L, Foght JM (2005) Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. *FEMS Microbiology Ecology,* **53**, 141-155.

Schulz S, Giebler J, Chatzinotas A, et al (2012) Plant litter and soil type drive abundance, activity and community structure of alkB harbouring microbes in different soil compartments. *The ISME journal,* **6**, 1763-1774.

Schutter M, Sandeno J, Dick R (2001) Seasonal, soil type, and alternative management influences on microbial communities of vegetable cropping systems. *Biology and Fertility of Soils,* **34**, 397-410.

Shade A, Hogan CS, Klimowicz AK, Linske M, McManus PS, Handelsman J (2012) Culturing captures members of the soil rare biosphere. *Environmental Microbiology,* **14**, 2247-2252.

Shen F, Lin J, Huang C, et al (2009) Molecular detection and phylogenetic analysis of the catechol 1, 2-dioxygenase gene from Gordonia spp. *Systematic and Applied Microbiology,* **32**, 291-300.

Sikkema J, de Bont JA, Poolman B (1995) Mechanisms of membrane toxicity of hydrocarbons. *Microbiological Reviews,* **59**, 201-222.

Singleton DR, Sangaiah R, Gold A, Ball LM, Aitken MD (2006) Identification and quantification of uncultivated Proteobacteria associated with pyrene degradation in a bioreactor treating PAH-contaminated soil. *Environmental Microbiology,* **8**, 1736-1745.

Sleugh B, Moore KJ, George JR, Brummer EC (2000) Binary legume–grass mixtures improve forage yield, quality, and seasonal distribution. *Agronomy Journal,* **92**, 24-29.

Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, Campbell JH, Fortney JL, Mehlhorn TL, Lowe KA, Earles JE, Phillips J, Techtmann SM, Joyner DC, Elias DA, Bailey KL, Hurt RA, Jr, Preheim SP, Sanders MS, Yang J, Mueller MA, Brooks S, Watson DB, Zhang P, He Z, Dubinsky EA, Adams PD, Arkin AP, Fields MW, Zhou J, Alm EJ, Hazen TC (2015). Natural bacterial communities serve as quantitative geochemical biosensors. *mBio*, **6**, e00326-15.

Snoeyink VL, Jenkins D (1980) *Water chemistry* John Wiley.

Song X, Xu Y, Li G, Zhang Y, Huang T, Hu Z (2011) Isolation, characterization of Rhodococcus sp. P14 capable of degrading high-molecular-weight polycyclic aromatic hydrocarbons and aliphatic hydrocarbons. *Marine Pollution Bulletin,* **62**, 2122-2128.

Suominen L, Jussila M, Mäkeläinen K, Romantschuk M, Lindström K (2000) Evaluation of the< i> Galega–Rhizobium< i> galegae system for the bioremediation of oil-contaminated soil. *Environmental Pollution,* **107**, 239-244.

Susarla S, Medina VF, McCutcheon SC (2002) Phytoremediation: an ecological solution to organic chemical contamination. *Ecological Engineering,* **18**, 647-658.

Suzuki M, Rappe MS, Giovannoni SJ (1998) Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Applied and Environmental Microbiology,* **64**, 4522-4529.

Tiirola MA, Suvilampi JE, Kulomaa MS, Rintala JA (2003) Microbial diversity in a thermophilic aerobic biofilm process: analysis by length heterogeneity PCR (LH-PCR). *Water Research,* **37**, 2259-2268.

Tilman D (1997) Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology,* **78**, 81-92.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature,* **418**, 671-677.

Torgerson WS (1958) Theory and methods of scaling. *Biometrika*.

Torsvik V, Sørheim R, Goksøyr J (1996) Total bacterial diversity in soil and sediment communities—a review. *Journal of Industrial Microbiology,* **17**, 170-178.

Torsvik V, Øvreås L (2002) Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology,* **5**, 240-245.

Tsuboi S, Yamamura S, Nakajima-Kambe T, Iwasaki K (2015) Diversity of alkane hydroxylase genes on the rhizoplane of grasses planted in petroleum-contaminated soils. *SpringerPlus,* **4**, 1-10.

van Elsas JD, Jansson JK, Trevors JT (2006) *Modern soil microbiology* CRC Press, 672 papes.

Van Gestel K, Mergaert J, Swings J, Coosemans J, Ryckeboer J (2003) Bioremediation of diesel oil-contaminated soil by composting with biowaste. *Environmental Pollution,* **125**, 361-368.

Van Hamme JD, Singh A, Ward OP (2003) Recent advances in petroleum microbiology. *Microbiology and molecular biology reviews: MMBR,* **67**, 503-549.

Varis E (1986) Goat's rue (*Galega orientalis* Lam.), a potential pasture legume for temperate conditions. *Journal of Agricultural Science in Finland*, **58**(2), 83-101.

Vázquez S, Nogales B, Ruberto L, et al (2009) Bacterial community dynamics during bioremediation of diesel oil-contaminated Antarctic soil. *Microbial Ecology,* **57**, 598-610.

Vershinina ZR, Baymiev AK, Blagova DK, Chubukova OV, Baymiev AK, Chemeris AV (2012) Artificial colonization of non-symbiotic plants roots with the use of lectins. *Symbiosis,* **56**, 25-33.

Vinas M, Sabate J, Espuny MJ, Solanas AM (2005) Bacterial community dynamics and polycyclic aromatic hydrocarbon degradation during bioremediation of heavily creosote-contaminated soil. *Applied and Environmental Microbiology,* **71**, 7008-7018.

Vitousek PM, Cassman K, Cleveland C, et al (2002) Towards an ecological understanding of biological nitrogen fixation. *Biogeochemistry,* **57**, 1-45.

Võsa T, Meripõld H (2008) Growing technology and production costs for dry mass for direct burning and green mass for biogas of Galega Orientalis. *Agronomy Research*, **6**, 415-421.

Warwick R, Clarke K (1993) Increased variability as a symptom of stress in marine communities. *Journal of Experimental Marine Biology and Ecology,* **172**, 215-226.

Wenzel WW (2009) Rhizosphere processes and management in plant-assisted bioremediation (phytoremediation) of soils. *Plant and Soil,* **321**, 385-408.

White DC, Flemming CA, Leung KT, Macnaughton SJ (1998) In situ microbial ecology for quantitative appraisal, monitoring, and risk assessment of pollution remediation in soils, the subsurface, the rhizosphere and in biofilms. *Journal of Microbiological Methods,* **32**, 93-105.

Whyte LG, Schultz A, Beilen JB, et al (2002a) Prevalence of alkane monooxygenase genes in Arctic and Antarctic hydrocarbon-contaminated and pristine soils. *FEMS Microbiology Ecology,* **41**, 141-150.

Whyte LG, Smits TH, Labbe D, Witholt B, Greer CW, van Beilen JB (2002b) Gene cloning and characterization of multiple alkane hydroxylase systems in Rhodococcus strains Q15 and NRRL B-16531. *Applied and Environmental Microbiology,* **68**, 5933-5942.

Winding A, Hund-Rinke K, Rutgers M (2005) The use of microorganisms in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety,* **62**, 230-248.

Wünsche L, Brüggemann L, Babel W (1995) Determination of substrate utilization patterns of soil microbial communities: an approach to assess population changes after hydrocarbon pollution. *FEMS Microbiology Ecology,* **17**, 295-305.

Yang S, Wen X, Zhao L, Shi Y, Jin H (2014) Crude oil treatment leads to shift of bacterial communities in soils from the deep active layer and upper permafrost along the China-Russia Crude Oil Pipeline route. *PloS one,* **9**, e96552.

Yarza P, Yilmaz P, Pruesse E, Glöckner FO, Ludwig W, Schleifer K-H, Whitman WB, Euzéby J, Amann R and Rosselló-Móra R (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nature Reviews Microbiology* **12**, 635-45.

Young CC, Kampfer P, Ho MJ, et al (2007) Arenimonas malthae sp. nov., a gammaproteobacterium isolated from an oil-contaminated site. *International Journal of Systematic and Evolutionary Microbiology,* **57**, 2790-2793.

Zhang W, Ricketts TH, Kremen C, Carney K, Swinton SM (2007) Ecosystem services and dis-services to agriculture. *Ecological Economics,* **64**, 253-260.

Zolotarev VN (2010). Agroecological estimate of sowing times of fodder galega for seeds in the central region of Russia. *Russian Agricultural Sciences*, *36*(6), 433-436.

Recent Publications in this Series