View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Helsingin yliopiston digitaalinen arkisto

1 Characterization of successional changes in bacterial community composition during
2 bioremediation of used motor oil-contaminated soil in a boreal climate

3 Lijuan Yan[1], Hanna-Mari Sinkko[2], Petri Penttinen[1], Kristina Lindström[1]

4 [1] Department of Environmental Sciences, PO Box 65 (Viikinkaari 2a), 00014 University of
5 Helsinki, Finland

6 [2] Department of Food and Environmental Sciences, PO Box 27 (Latokartanonkaari 11), 00014
7 University of Helsinki, Finland

8 Corresponding author: Lijuan Yan, Department of Environmental Sciences, PO Box 65, 00014
9 University of Helsinki, Finland, +358-45-8747988, lijuan.yan@helsinki.fi

10 Keywords: LH-PCR, Bioremediation, Bacterial community, Oil contamination, Soil, Fodder galega,
11 Smooth brome

12 Running title: Bacterial community succession during bio-remediation

13

14 Abstract

15 The widespread use of motor oil makes it a notable risk factor to cause scattered contamination in
16 soil. The monitoring of microbial community dynamics can serve as a comprehensive tool to assess
17 the ecological impact of contaminants and their disappearance in the ecosystem. Hence, a field
18 study was conducted to monitor the ecological impact of used motor oil under different perennial
19 cropping systems (fodder galega, brome grass, galega-brome grass mixture and bare fallow) in a
20 boreal climate zone. Length heterogeneity PCR characterized a successional pattern in bacterial
21 community following oil contamination over a four-year bioremediation period. Soil pH and
22 electrical conductivity were associated with the shifts in bacterial community composition. Crops
23 had no detectable effect on bacterial community composition or complexity. However, the legume
24 fodder galega increased soil microbial biomass, expressed as soil total DNA. Oil contamination
25 induced an abrupt change in bacterial community composition at the early stage, yet the effect did
26 not last as long as the oil in soil. The successional variation in bacterial community composition can
27 serve as a sensitive ecological indicator of oil contamination and remediation *in situ*.

1. INTRODUCTION

Petroleum hydrocarbons (PHCs) originating from crude oil or refined petroleum products are detrimental to environmental health as soil contaminants. Used motor oil or crankcase oil is lubricating oil that is removed from the crankcase of internal combustion engines of vehicles (Irwin et al. 1997). The widespread handling of small volumes of used motor oil by enterprises, farms and private persons makes it a notable risk factor to cause scattered contamination. Besides physical removal (leaching and volatilization), PHCs are subjected to biodegradation, the metabolic ability of microorganisms to transform or mineralize organic contaminants to less harmful, non-hazardous substances (Margesin and Schinner 1997, Margesin and Schinner 2001, Namkoong et al. 2002, Chaîneau et al. 2003). Hydrocarbon fractions differ in their susceptibility to microbial attack (Leahy and Colwell 1990). In used motor oil, the concentrations of long-chain aliphatics, benzene-, and naphthalene-based compounds, polycyclic aromatic hydrocarbons (PAHs) and heavy metals are high; once released, these carcinogenic compounds can result in long lasting contamination due to their high resistance to microbial degradation (Irwin et al. 1997, Dominguez-Rosado et al. 2004).

Nitrogen is often a limiting factor in biodegradation of hydrocarbon-contaminated soils. Leguminous plants that are resistant to hydrocarbon pollutants assist bioremediation of oil-polluted sites effectively and sustainably as substitutes of N-fertilizers (Dominguez-Rosado et al. 2004, Kamath et al. 2004, Chiapusio et al. 2007). The perennial legume fodder galega (*Galega orientalis*) and smooth brome grass (*Bromus inermis*) are both suitable to grow in a boreal climate and have great potential to enhance bioremediation of oil-contaminated soil in microcosm and mesocosm studies (Suominen et al. 2000, Kulakow et al. 2000, Lindstrom et al. 2003, Kaksonen et al. 2006, Muratova et al. 2008, Jasinskas et al. 2008, Kryževičienė et al. 2008, Mikkonen et al. 2011a). Further assistance to the bioremediation process may be provided by plant growth promoting bacteria (PGPB) that have potential to mitigate plant stress response and increase the bioavailability of soil contaminants,

therefore enhancing the degradation of contaminants (Gurska et al. 2009, Hong et al. 2011, Pajuelo et al. 2011, Bhattacharyya and Jha 2012).

Effectiveness and completeness are ultimate goals in a successful remediation project (White et al. 1998). Complete removal of contaminants in the environment is not always easy to achieve. White et al. (1998) proposed an ecologically based test of “how clean is clean” using assessment of microbial community dynamics as a comprehensive tool to estimate contaminant disappearance. Hence, understanding the successional dynamics of bacterial communities on contaminated sites is an important aspect of risk assessment needed for the planning of following remediation actions. Due to the operational simplicity and high reproducibility in analyzing large sample series, length heterogeneity analysis of polymerase chain reaction products (LH-PCR, Suzuki et al. 1998) was widely used to monitor the succession of microbial communities in response to oil pollution (Mills et al. 2003, Mills et al. 2006, Mikkonen et al. 2011b, Mikkonen et al. 2012). The possibility to compare the sizes of the amplicons against 16S rRNA gene sequences *in silico* enables preliminary identification of bacterial groups in the community (Mills et al. 2003, Tiirola et al. 2003).

To date, bacterial community succession in used motor oil-polluted soil in a boreal climate zone has received little experimental attention. The studies on bacterial community succession in oil-polluted vegetated soil have been limited to short-term microcosm and mesocosm experiments (Mikkonen et al. 2011b, Mukherjee et al. 2013, Simarro et al. 2013). The successional patterns of soil microbial community following oil contamination in a boreal field are plausibly different from those in short-term controlled conditions. Hence, a systematic field bioremediation study was established with the main aim to monitor the impact of used motor oil, different perennial cropping systems (fodder galega, brome grass, galega-brome grass mixture and bare fallow), plant growth promoting bacteria and soil parameters on bacterial community composition over a four-year period (2009-2012) in a boreal region, using LH-PCR microbial community fingerprinting analysis.

76 2. MATERIALS AND METHODS

77 2.1 Experimental design, samplings and chemical analysis of soil

78 The multi-year bioremediation field experiment was established in a split-plot design at Viikki
79 experimental farm, Helsinki, Finland (60°14'N, 25°01'E, 8 m AMSL). Crop treatments of
80 monocultures of brome grass and fodder galega, their mixture and bare fallow were the main plots
81 in four replicated blocks. Used motor treatments (oil+/-) and plant growth promoting bacteria
82 treatments (PGPB+/-) were the sub-plot factors. About 6 kg of used motor oil (Teboil Lubricants
83 Classic Mineral Motor oil, SAE 10W-30, API SF/CD, Finland) was mixed with 10 kg of coarse
84 sand (0.5-1.2 mm), spread and spiked onto the top 20 cm of each designated-to-be oil-contaminated
85 plot with a rotary tiller on 17 June 2009, making the target contamination approximately to 7000
86 ppm (7 g kg^{-1} dry soil). The non-contaminated control plots received pure sand on the top 20 cm
87 soil. Before sowing, seeds of *G. orientalis* cv. 'Gale' (Naturcom Oy, Ruukki, Finland) were all
88 inoculated with *Neorhizobium galegae* strain HAMBI 540 (University of Helsinki, Helsinki,
89 Finland). The seeds of *Neorhizobium galegae*-inoculated *G. orientalis* and *B. inermis* cv. 'Lehis'
90 (Jõgeva Plant Breeding Institute, Estonia) were inoculated with two PGPB strains, *Pseudomonas*
91 *trivialis* 3Re27 (Graz University of Technology, Graz, Austria) and *Pseudomonas extremorientalis*
92 TSAU20 (National University of Uzbekistan) according to Egamberdieva et al. (2010), as the co-
93 inoculation of these two PGPB strains with *Neorhizobium galegae* were found to improve growth
94 and symbiotic performance of fodder galega in a greenhouse experiment (Egamberdieva et al.
95 2010). PGPB-free seeds were used as controls. The seeds were manually sown and lightly covered
96 by raking. Crops were harvested twice a year from 2010 on. Weeds were controlled manually. Soil
97 samples were taken from the top 20 cm layer in the field at six time points (July 2009, May 2010,
98 November 2010, May 2011, May 2012 and October 2012) and stored at -20°C until the analysis.
99 Soil chemical properties of three sample sets (July 2009, November 2010 and May 2012) were
100 measured. Electrical conductivity (EC) and soil pH were measured in a 1:2.5 (v:v) soil-water

101 suspension with MeterLab™ CDM210 (Radiometer Analytical) and SCHOTT CG842 pH-meter (SI
102 Analytics), respectively. Soil dry matter content was determined by drying to constant mass at 105
103 ºC. Soil total C and N contents were analysed using the VarioMax CN-analyzer (Elementar
104 Analysensysteme GmbH, Hanau, Germany) and corrected to the dry-weight basis. The oil
105 concentration in each oil-spiked plot was determined as the difference of total solvent extractable
106 material (TSEM) concentration between the plot and the average of 4 to 5 randomly selected
107 control plots at each sampling time. Detailed information on the field design, oil spike, soil
108 sampling, measurements of soil chemical properties and TSEM determination are described in Yan
109 et al. (2015).

110 2.2 DNA extraction and LH-PCR

111 Soil DNA was directly extracted from 0.50 g moist soil samples with FastDNA SPIN kit for Soil
112 (Qbiogene, USA) according to the manufacturer's instructions. The final elution volume was 75-125
113 µL. The DNA yield of the first four sample sets was measured fluorometrically on a 96-well plate
114 according to the manufacturer's instructions (PicoGreen dsDNA Quantification Reagent Kit;
115 Molecular Probes).

116 Soil DNA extract was diluted 1/50 with sterile deionized water to avoid PCR inhibition by co-
117 extracted humic substances in soil. Length heterogeneity PCR (LH-PCR) with 0.5-5 ng of DNA as a
118 template was performed as described by Mikkonen et al. (2011b). The amplified fragments were
119 separated with polyacrylamide capillary electrophoresis using ABI PRISM 310 Genetic Analyzer
120 (Applied Biosystems).

121 2.3 LH-PCR data processing

122 The fingerprint electropherograms were imported from the GeneScan v. 3.7 (Applied Biosystems) as
123 12-bit densitometric curves with Curve Converter into an artificial gel in BioNumerics v. 6.6 (Applied
124 Maths, Sint-Martens-Latem, Belgium). The bands (peaks) of each sample profile (FAM-labeled)

125 were manually assigned to avoid background noise. The bands were aligned and normalized with the
126 internal HEX-labelled size standards. The active area of each profile was set to the expected amplicon
127 size of 460-565 base pairs (bp) with normalized position ranging between 18.11% and 64.92%
128 (resolution = 1942 points). The densitometric curve of each bacterial community profile was directly
129 exported from BioNumerics as curve-based raw data. The relative fluorescence ratio of each band
130 point was calculated as its contribution of the fluorescence intensity to the summed fluorescence
131 intensity of the 1942 band points within the size range of 460-565 base pairs.

132 The fluorescence intensity, area and size (bp) of each peak and the number of peaks present in each
133 LH-PCR profile were exported directly from the BioNumerics LH-PCR fingerprint report for peak-
134 based analysis. Each LH-PCR peak differentiated by BioNumerics software was considered an
135 operational taxonomic unit (OTU), identified by its LH-PCR amplicon size (bp). The number of peaks
136 (OTUs) was used as proxy of the species richness (S) of the bacterial community. The relative area
137 of each OTU was calculated as its proportion in the summed area of all the peaks in that profile within
138 the size range. Peak-based Shannon diversity index (H) of each bacterial community profile was
139 calculated according to the formula: $H= - \Sigma p_i \ln p_i$, where p_i is the relative fluorescence intensity of
140 the peak of the *i*th operational taxonomic unit (OTU).

141 2.4 Statistical analyses

142 LH-PCR curve-based fingerprinting data, which represented soil bacterial communities, were non-
143 normally distributed and included high numbers of zeroes. Therefore the LH-PCR and soil chemical
144 data were subjected to non-parametric distance-based multivariate methods. Bray-Curtis distance was
145 calculated between observations for all the following distance-based nonparametric multivariate
146 analyses. Variation in the entire LH-PCR curve-based data was first visualized by the distance-based
147 principal coordinates (PCoA), which was performed in the R environment (R Development Core
148 Team 2014), using the function "cmdscale" in package Vegan (Oksanen et al. 2015).

149 The effects of crops (legume, grass, legume-grass mixture and bare fallow), oil and PGPB treatments,
150 sampling time, and replicated blocks as well as their interactions on soil bacterial community
151 composition were analysed using permutational multivariate analysis of variance (PERMANOVA)
152 (Anderson 2001, McArdle and Anderson 2001) in PRIMER v.6 software (Clarke and Gorley 2006)
153 with add-on package PERMANOVA+ (Anderson et al. 2008). We used 9999 permutations to
154 calculate the significance of the treatment effects.

155 To test differences of bacterial communities based on the *a priori* groups (e.g. crops, oil+/-, PGPB+/-,
156 sampling times, experimental blocks, times in a growing season), we performed non-parametric
157 distance-based discriminant analysis (db-DA, Anderson and Robinson 2003) using the function
158 "CAPdiscrim" of R package BiodiversityR (Kindt and Coe 2005). Discriminant analysis also
159 calculated the proportion of observations that were correctly classified based on the above tested *a*
160 *priori* groups. The significance of the classification was calculated using 9999 permutations. The
161 multivariate homogeneity of group variances (dispersions) (Anderson 2006) was tested using the
162 function "betadisper" in the package Vegan (Oksanen et al. 2015). The function
163 "permutest.betadisper" with 9999 permutations was used to calculate significance for the pairwise
164 comparisons of the multivariate dispersions of the groups (Supplementary Figure S1), the null
165 hypothesis being that there were no differences in dispersion between groups.

166 To study the variation in bacterial community composition as a function of soil physiochemical
167 variables, a constrained analysis of principal coordinates (CAP), also called as distance-based
168 redundancy analysis (Legendre and Anderson 1999) was performed. The CAP, which used soil
169 physicochemical variables and LH-PCR curve-based data of PGPB-untreated samples from three
170 sampling times (July 2009, Nov. 2010 and May 2012), was executed in the R package Vegan
171 (Oksanen et al. 2015) using the function "capscale". We used 9999 permutations of LH-PCR data
172 with the function "permutest" to test significance. Insignificant and collinear soil chemical properties
173 were excluded from the final CAP model. The idea behind CAP analysis is to apply multivariate

174 linear regression to represent the bacterial community assemblages as a function of explanatory
175 variables such as our soil physiochemical variables. Subsequently, the principal coordinates of fitted
176 values (Legendre and Anderson 1999) can be used to visualize the significant differences among the
177 community assemblages. To be able to visualize each fragment as base pairs in the CAP ordinations,
178 we combined all band points produced by LH-PCR within 1 bp by summarizing the relative
179 fluorescence of these band points (summarized LH-PCR fragment) and calculating the average
180 proportion of each summarized fragment. Thus, CAP analyses were based on the relative abundance
181 of summarized LH-PCR fragments. The scores of individual components of the bacterial community
182 assemblages (LH-PCR fragments) were calculated using the function "scores.rda" of the package
183 vegan (Oksanen et al. 2015).

184 Repeated measures split-plot analysis of variance (RM ANOVA) with the sampling time as the
185 repeated factor (within-subject factor) was used to test the overall between- and within-subjects
186 effects (sphericity assumed) on soil total DNA concentration and peak-based ecological indices (H
187 and S) in SPSS (version 22, IBM Inc., Armonk, NY, USA). Crop and oil treatments were input as
188 fixed factors and block (replicate) as a random factor. Crop was tested against the interaction term
189 crop × block to take out the effect of the main plot from the residual variance so it does not skew the
190 error variance of the subplot stratum. Oil treatment and its remaining interaction with crop treatment
191 were tested against the subplot error mean square. For each sampling time, the dependent variables,
192 e.g. H, S and soil DNA concentration were roughly normally distributed, checked with Normal Q-Q
193 plots and Shapiro-Wilk normality test in SPSS, prior to parametric analysis. The population variances
194 were assumed equal for treatment groups as the sample sizes were equal. Bonferroni multiple pairwise
195 test was applied to compare the means, when treatment effect was significant. When the effects of
196 interactions between sampling times and other treatment factors were significant, the split-plot
197 univariate analysis of variance (UV ANOVA) was applied to further test the between-subjects effects

198 (oil, crop and oil × crop) on soil bacterial diversity at separate sampling times. In all statistical
199 analysis, differences were concluded significant at p<0.05.

200 3. Results

201 In our multi-year bioremediation field experiment, soil total DNA concentration was monitored for
202 three years as a proxy for total soil microbial biomass (Table 1). Soil DNA concentration was different
203 in different cropping systems (RM ANOVA, $p<0.05$). The soil planted with the legume fodder galega
204 gained 3.2 ng g^{-1} fresh soil (18%) more soil microbial DNA than bare fallow on average (Table 1).
205 The difference between legume and bare fallow on soil DNA concentration was most significant in
206 May 2011 (Table 1). The impact of oil and PGPB treatment on soil total DNA was insignificant.

207 3.1 Evaluation of treatment effect on bacterial community composition using curve-based measures

208 We used LH-PCR to assess the effect of crop, oil and PGPB treatment on microbial diversity patterns
209 in a four-year field experiment. In a principal coordinate analysis (PCoA) the 192 curve-based
210 bacterial LH-PCR community-fingerprinting profiles showed a clear time-dependent shifting pattern
211 (Supplementary Figure S2). The bacterial community profiles in contaminated and non-contaminated
212 soil samples were different at the first and second sampling times. Crop and PGPB treatments had no
213 detectable effect on the communities.

214 The effect of treatments (crops, oil and PGPB) and their interactions on bacterial community structure
215 was further evaluated using split-plot PERMANOVA. The effect of oil contamination on bacterial
216 community composition was statistically significant at the beginning of the growing seasons of 2009,
217 2010 and 2011, but insignificant at other sampling times (Supplementary Table S1). Experimental
218 blocks significantly affected the variation in bacterial community composition (Supplementary Table
219 S1). Crop treatment (brome grass, galega and their mixture) and plant growth promoting bacteria
220 showed no significant impact on bacterial community composition (Supplementary Table S1).

Discriminant analysis was applied as a follow-up procedure to the PERMANOVA to confirm and visualize how the composition of microbial communities differs between *a priori* groups. The separation of the *a priori* groups based on sampling times (Figure 1a), oil treatment (Figure 1b) and times of a growing season (Figure 1d) was clear. However, the dispersions of observations at different sampling times (Supplementary Figure S1a) and time of a growing season (Supplementary Figure S1d) were not equal, which may have affected the differentiation of groups. The averaged LH-PCR profiles in the end of the growing seasons were also clearly similar to each other (Supplementary Figure S3c and S3f). The crop treatments showed no effect on bacterial community compositions (Figure 1c).

3.2 Evaluation of treatment effect on bacterial community composition using peak-based measures

The effects of oil, crops, sampling time and their interactions on the Shannon diversity and richness of bacterial community compositions were assessed with RM ANOVA (Table 2). The 192 LH-PCR profiles showed 10 to 38 peaks that were regarded as proxies for operational taxonomic units (OTUs). The peak-based ecological indices showed a strong time-dependent pattern. The species richness (number of OTUs) was not influenced by oil, but by time (RM ANOVA, $p<0.05$); the highest values were observed in spring 2010 and the lowest in both autumns (Table 2). The Shannon diversity indices differed with time and time × oil treatment (RM ANOVA, $p<0.05$). Bacterial diversity in the contaminated plots was significantly different from diversity in the non-contaminated plots at the first and fourth sampling times in July 2009 ($p<0.05$) and in May 2011 ($p=0.051$), respectively (Table 2). The Shannon diversities were highest in spring 2010 and lowest in November 2010 in both the contaminated and the non-contaminated plots (Table 2). Crop treatments did not affect soil bacterial diversity and species richness. The variation of bacterial diversity between blocks was high (RM ANOVA, $p<0.05$).

244 There was a time-dependent pattern in the abundance of the dominant OTUs. In non-contaminated
245 plots, the OTUs with amplicons 469-470 bp, 495-496 bp, 520-521 bp and 535-536 bp were the most
246 dominant bacterial groups with the relative peak area of over 10% in the averaged LH-PCR
247 fingerprint profiles. The abundance of the OTUs 535 and 536 bp increased over time. In autumn 2012,
248 the relative peak area of the OTU 536 bp reached 33%.

249 3.3 Compositional changes in microbial community in response to changing soil variables

250 Constrained analysis of principal coordinates (CAP) revealed a time-dependent pattern and a strong
251 dependence between microbial community composition and soil variables (oil concentration, total C,
252 total N, C:N ratio electrical conductivity and pH) in all PGPB-untreated plots from three sampling
253 times (July 2009, November 2010 and May 2012). Among the measured soil variables, oil
254 concentration (p=0.001), pH (p=0.002) and EC (p=0.011) were factors that accounted for 13.5% of
255 the total variation in the bacterial community composition (Figure 2, p<0.0001). The first two CAP
256 axes accounted for 97.6% of the variance. Both soil total C and C:N ratio were associated with the
257 changes of bacterial community composition (data not shown); however, they were removed from
258 the final CAP model due to the collinearity (high correlation) with oil concentration.

259 The influence of sampling time was mostly shown along the first CAP axis, as the bacterial profiles
260 shifted from right to left in the ordination space over time (Figure 2). Oil effect was better shown on
261 the second CAP axis than on the first CAP axis, as 53.8% of the total variation originating from the
262 oil concentration was loaded onto the second axis. Bacterial communities of the first sampling time
263 from the non-contaminated samples correlated positively with EC and pH whereas those from oil-
264 contaminated soils correlated positively with oil concentration and pH. At later sampling times,
265 bacterial communities of both oil-contaminated and control soils showed negative correlation with
266 all tested parameters (Figure 2).

267 The bacterial taxa that primarily responded to the quantitative changes in soil variables were also
268 defined by CAP analysis. The LH-PCR fragments 495-497 bp correlated positively with oil
269 concentration (Figure 2). Specifically, the relative fluorescence ratio of the LH-PCR fragment 497 bp
270 was much higher in oil-contaminated than in non-contaminated plots (Supplementary Figure S3a),
271 accounting for 19% of the total peak area in July 2009, when the oil concentration was above 4.00 g
272 kg^{-1}. Although the oil concentration at the second sampling time was almost unchanged (3.85 g kg^{-1}),
273 the 497 bp peak disappeared (Supplementary Figure S3b). The LH-PCR fragments 471-472 bp, 517
274 bp and 535-536 bp correlated negatively with oil concentration (Figure 2). The LH-PCR fragment
275 469 bp correlated positively with soil EC in non-contaminated samples in the first sampling time. The
276 OTUs 465-466 bp, 492 bp, 501 bp and 531-532 bp that were abundant in the first sampling time
277 correlated positively with soil pH (Figure 2). LH-PCR fragments 537-539 correlated negatively with
278 soil pH and EC.

279 In addition, CAP revealed a strong negative correlation between soil electrical conductivity (EC) and
280 oil concentration, and a strong positive correlation between electrical conductivity and pH (Figure 2).
281 Oil concentration was weakly linked with pH (Figure 2).

282 4. Discussion

283 Microbial communities can be considered as functional units that are characterized by the sum of the
284 metabolic properties of the microbial taxa involved (Wünsche et al. 1995). The field results showed
285 that oil contamination had no effect on soil microbial biomass. It disagrees with the greenhouse
286 experiment where soil total DNA concentration increased in the presence of oil during the first 15
287 weeks' time (Mikkonen et al. 2011a), due to the more complex environmental condition in the field
288 than in the controlled greenhouse.

289 The response of the microbial community to changes in the environment can be monitored with
290 community fingerprinting methods, e.g. with length heterogeneity analysis of polymerase chain

291 reaction (LH-PCR) that is based on the natural length variation of 16S rRNA gene (Tiirola et al. 2003,
292 Mikkonen et al. 2011b). The analysis of peak-based ecological indices (diversity and richness) is a
293 simplified measure routinely used in most LH-PCR data analysis (Mills et al. 2003, Mills et al. 2006,
294 Mikkonen et al. 2011b, Mikkonen et al. 2012, Wu et al. 2015, Zou et al. 2015). Nevertheless, these
295 traditional ecological indices are not as sensitive as the distance-based nonparametric multivariate
296 measures to observe the potential treatment effect on microbial community, due to the inherently
297 lower resolution because of limited number of peaks resolved (Mills et al. 2006, Mikkonen et al.
298 2011b). However, Mills et al. (2003) found that LH-PCR is operationally simpler and has better
299 reproducibility than T-RFLP, another popular community fingerprint technique in profiling diverse
300 microbial communities during bioremediation of petroleum-contaminated soils. The LH-PCR
301 technique was also successfully implemented to study the changes in the bacterial community
302 composition in the multi-year bioremediation field experiment. In this study, both curve-based
303 multivariate analysis and peak-based univariate analysis were used to assess the succession of
304 bacterial community during bioremediation. An LH-PCR densitometric curve-based profile was used
305 to pattern the whole soil bacterial community structure, whereas LH-PCR peaks were taken as proxies
306 for operational taxonomic units (OTUs) that approximate bacterial species or species groups (Zou et
307 al. 2015). Every LH-PCR fingerprint included at least ten peaks, indicating a fine resolution of 16S
308 rRNA gene fragments among diverse bacterial populations, appropriate for monitoring the response
309 of microbial community to oil contamination. The number of LH-PCR peaks did not change
310 significantly in the presence of oil contamination, suggesting resilience of the dominant bacterial
311 populations towards oil contamination in the agricultural soil. In the community profiles, the relative
312 peak areas of the most abundant OTUs were less than 40% and those of several OTUs were over 10%,
313 indicative of inherent compositional complexity of bacterial populations in the community.

314 The discriminant analysis based on Bray-Curtis dissimilarity, produced an effective discrimination
315 of bacterial LH-PCR profiles based on *a priori* groups including different sampling times, oil

316 contamination and growing seasons in the field study. Time and oil contamination lead to
317 successional changes in the bacterial community composition. These findings agreed with the
318 previous greenhouse experiment where the oil effect on bacterial communities was constantly
319 significant during the whole 20-week experiment using Pearson dissimilarity (distance)-based
320 discriminant analysis of LH-PCR curve data (Mikkonen et al. 2011b). The db-DA method, and
321 particularly PERMANOVA, assumes equal dispersion (variance) of observations between *a priori*
322 groups to be analyzed. However, the *a priori* groups of sampling times and seasons showed unequal
323 dispersion, which may have effect on the discrimination based on the different sampling times and
324 seasons. As the oil-contaminated samples separated clearly from non-contaminated ones in db-DA,
325 particularly in the first two sampling times when the oil concentration was highest, one could assume
326 that oil contamination level resulted in the unequal dispersion of observations between sampling times.
327 The unequal dispersion of observations in the beginning and at the end of growing season was
328 probably caused by the unequal sample sizes (beginning: n=224 and end: n= 96). Despite the unequal
329 dispersions of times and seasons, the p-values of db-DA were highly significant, which indicate that
330 the effect of time and season was considerable. In addition, db-DA (Anderson and Robinson 2003) is
331 quite robust to violations of the assumptions.

332 The successional shifts of microbial community reflected the changes of soil condition, such as motor
333 oil addition and degradation in our experiment. One month after oil spike, oil diminished bacterial
334 diversity significantly. The CAP analysis elucidated a significant difference in LH-PCR profiles
335 between oil-contaminated and control plots. The structural change of microbial community in the
336 presence of oil, especially the increased abundance of the bacterial group (OTU 497 bp), was
337 associated with the rapid loss of oil (approximately 42% on average, Yan et al. 2015) observed in the
338 first month following oil spike (June-July 2009). As indigenous microbial populations of differing
339 taxonomic microbial groups capable of degrading hydrocarbons exist widely in natural environment
340 (Atlas 1981), the presence of oil hydrocarbons may increase the absolute and relative abundance of

341 hydrocarbon-utilizing bacteria in the community. This structural difference between oil-contaminated
342 and non-contaminated soil was associated with rapid oil reduction, suggesting that the difference in
343 community composition was related to a functional difference, such as biodegradation of
344 hydrocarbons in the oil-contaminated plots. Interestingly, the dominance of the OTU 497 bp
345 disappeared on the later sampling occasions. This OTU likely represented a group of r-strategic
346 hydrocarbon-utilizing bacteria that responded to certain easy-degradable hydrocarbon substrates,
347 responsible for the initial biodegradation of hydrocarbons. According to Tiirola et al. (2003), the OTU
348 497 bp likely belongs to Epsilon-*Proteobacteria*, *Thermus*/*Deinococus*, Alpha-*Proteobacteria* and
349 Gram positives. This prediction agrees with the studies that demonstrated the dominance of Alpha-
350 *Proteobacteria* during the whole process of biodegradation, especially at the early stages of
351 biodegradation (Mills et al. 2003, Vinas et al. 2005). *Thermus* sp. was also reported to effectively
352 degrade hexadecane/pyrene mixture as the sole carbon and energy source at high temperature (70 °C)
353 in bioreactor (Feitkenhauer et al. 2003). Gram-positive bacteria were also suggested to adapt to
354 hydrocarbon biodegradation in soil in cold climate due to its high resistance to low temperature
355 (Eriksson et al. 2001).

356 Bacterial populations can adapt to and recover from oil contamination owing to their unique
357 biological features including fast reproduction rates, high degree of physiological flexibility and rapid
358 evolution through mutations or horizontal gene transfer (Winding et al. 2005, Allison and Martiny
359 2008). From July 2009 to May 2010, the oil concentration remained almost unchanged (Yan et al.
360 2015), yet the impact of oil on the composition of bacterial community decreased significantly,
361 indicating an intensive adaptation of microbial populations in oil-contaminated soil. Bacterial
362 diversity and species richness were significantly higher in both contaminated and non-contaminated
363 soils in May 2010 than those in July 2009, reflecting the development of indigenous microbial
364 populations over time, regardless of oil contamination.

365 The overall rate of biodegradation of the component fractions is affected by the compositional
366 heterogeneity of crude oil products (Leahy and Colwell 1990). A second rapid oil reduction occurred
367 during the second growing season in 2010 (Yan et al. 2015). Together with the optimum
368 environmental conditions (especially high summer temperature, Yan et al. 2015), the biodegradation
369 of available easy-degradable hydrocarbons was accelerated. The composition of the oil hydrocarbons
370 remaining in the soil matrix was plausibly changed after this period of rapid biodegradation. After
371 that, oil reduction slowed down, likely due to the exhaustion of easy-degradable hydrocarbons in soil.
372 The low oil reduction rate was coupled with the reduced dissimilarity between bacterial community
373 fingerprinting profiles in contaminated and non-contaminated plots over years. The return of a
374 baseline community indicates that the risk associated with contamination is significantly decreased
375 (White et al. 1998). However, since microbial communities change in time, the return of the pre-
376 contamination community composition may be impossible. Although oil hydrocarbons were not
377 completely removed in the fourth growing season 2012 (Yan et al. 2015), the microbial communities
378 in contaminated and non-contaminated soil were similar. The similarity of the communities can be
379 considered as an indication of significantly decreased risk. If suitable clean control soil is available,
380 the successional variation in bacterial community composition can serve as a sensitive ecological
381 indicator of oil contamination and remediation *in situ*.

382 Microbial community composition is strongly dependent on the presence of vegetation (Habekost et
383 al. 2008), likely due to an input of nutrients as the vegetation cover is decomposed (Hobbie 2015). In
384 a greenhouse bioremediation experiment, fodder galega increased the total microbial biomass and
385 induced dissimilarity in the microbial community but did not affect the bacterial species diversity
386 (Mikkonen et al. 2011a, 2011b). In field-lysimeters galega increased the diversity of bacteria in the
387 rhizosphere (Kaksonen et al. 2006). We observed that under fodder galega the soil total DNA
388 concentration was higher compared to the bare fallow, especially when the legume reached stable
389 growth in 2011. None of the crop treatments (galega, brome grass or their mixture) showed detectable

390 effects on bacterial community composition and diversity, not even at the later phase of
391 bioremediation when crops had fully established their roots in soil. There were no significant
392 differences in soil total N content between crop treatments (Yan et al. 2015), disagreeing our
393 hypothesis that crops, especially the legume inoculated with rhizobia, would increase soil N content
394 and thus result in a change in bacterial community composition compared to bare fallow plots. The
395 inherent complexity of the microbial populations in soil and the resolution limit of the LH-PCR
396 technique might together make the effect of crop treatment on bacterial community composition hard
397 to detect. In the range of 30:1 to 10:1, the C:N ratio is not considered to limit bioremediation (Alkoaik
398 and Ghaly 2006). Most aerobic heterotrophic bacteria, which are associated with hydrocarbon
399 degradation (Wrenn and Venosa 1996, Zhuang et al. 2003, Saul et al. 2005), favor a neutral pH (Leahy
400 and Colwell 1990). As the C:N ratio and pH in our field were favorable for bioremediation (Yan et
401 al. 2015), the lack of effect of the plants and PGPB on bacterial community structure and on oil
402 reduction rate may be attributed to the optimum soil conditions. Thus, when soil is nutrient-rich and
403 satisfies the metabolic requirements of the soil organisms, the effect of vegetation and PGPB on
404 microbial communities is negligible.

405 Seasonal changes in the bacterial community structure were distinct. It is consistent with the finding
406 that seasonality was the most influential factor influencing microbial community structure, provided
407 that the experimental plots share the same soil type (Schutter et al. 2001). At the end of each growing
408 season, microbial communities were less diverse than in the beginning. In contrast to our results, a
409 phospholipid fatty acid (PLFA)-based analysis of microbial community structure in grassland in Jena,
410 Germany, revealed a more diverse pattern of microbial populations in October than in May (Habekost
411 et al. 2008). The authors conferred the seasonal changes to the higher availability and quality of
412 organic input by vegetation in the autumn. In our experiment, the seasonal succession was not driven
413 by organic input since the succession was similar in both bare fallow and vegetated plots. An earlier
414 study revealed that the seasonal difference in PFLA-based microbial community patterns in temperate

415 grassland systems was related to soil mineral nitrogen and soil moisture contents (Bardgett et al.
416 1999). Environmental conditions, especially the difference in soil temperature and moisture between
417 spring and autumn, were likely to be the major factor behind the seasonal variation. The lowest values
418 of ecological indices, regardless of oil contamination, were observed in autumn 2010, when the soil
419 was covered with snow. The similarity of LH-PCR profiles of the autumn samples between
420 contaminated and non-contaminated plots suggested that the bacterial populations exhibited a similar
421 development pattern in both soils in autumn. Thus, seasonal variation in bacterial community was
422 stronger than variation caused by oil contamination. As biodegradation rates increase with
423 temperature increase (Leahy and Colwell 1990) and low temperature limits the oxidation of
424 hydrocarbons in motor oil-contaminated soil (Alkoaik and Ghaly 2006), the similarities in the LH-
425 PCR profiles between oil-contaminated and non-contaminated soils might be associated with the low
426 oil reduction rate in autumn.

427 Taken together, the LH-PCR community fingerprinting technique and the following data analysis
428 demonstrated a dynamic succession of the bacterial community in field soil. The microbial
429 communities responded quickly to oil contamination, yet the effect of oil on community composition
430 did not last as long as the oil in soil. Besides oil concentration, the changes in soil chemical properties
431 such as soil pH and electrical conductivity significantly influenced the structural changes in bacterial
432 community. Linking the oil degradation to the changes in community structure more strongly would
433 require additional studies on functional genes. LH-PCR accompanied with multivariate data analysis
434 was an effective method for monitoring microbial succession. However, if the goal is taxon
435 identification it needs to be complemented with sequencing-based methods.

436 Acknowledgements

437 This work was funded by Legume Futures, an international research project funded by the European
438 Union through the Framework 7 Programme (FP7) under grant agreement number 245216 (FP7-

439 KBBE-2009-3), and by MUTKU ry for oil analysis. Asko Simojoki and Frederick L. Stoddard were
440 acknowledged for their supervision in soil and crop analysis. We thank Markku Tykkyläinen for his
441 technical assistance in the field management and plant harvest. The authors also acknowledge Anni-
442 Mari Pulkkinen and Kati P. Pulkkinen for their technical assistance with oil analysis.

443 References

444 Alkoaik, F. & Ghaly, A. 2006. Thermobioremediation of soil contaminated with used motor oil.
445 Energy Sources, Part A 28: 487-500.

446 Allison, S.D. & Martiny, J.B. 2008. Colloquium paper: resistance, resilience, and redundancy in
447 microbial communities. Proceedings of the National Academy of Sciences of the United States of
448 America 105 Suppl 1: 11512-11519.

449 Anderson, M.J., Gorley, R.N. & Clarke, K.R. 2008. PERMANOVA+ for PRIMER: Guide to
450 Software and Statistical Methods. Plymouth, UK: Primer-E Ltd.

451 Anderson, M.J. 2006. Distance-Based Tests for Homogeneity of Multivariate Dispersions.
452 Biometrics 62: 245-253.

453 Anderson, M.J. & Robinson, J. 2003. Generalized discriminant analysis based on distances.
454 Australian & New Zealand Journal of Statistics 45: 301-318.

455 Atlas, R.M. 1981. Microbial degradation of petroleum hydrocarbons: an environmental perspective.
456 Microbiological Reviews 45: 180-209.

457 Bardgett, R.D., Lovell, R.D., Hobbs, P.J. & Jarvis, S.C. 1999. Seasonal changes in soil microbial
458 communities along a fertility gradient of temperate grasslands. Soil Biology and Biochemistry 31:
459 1021-1030.

460 Bhattacharyya, P.N. & Jha, D.K. 2012. Plant growth-promoting rhizobacteria (PGPR): emergence in
461 agriculture. World Journal of Microbiology & Biotechnology 28: 1327-1350.

462 Chaîneau, C., Yepremian, C., Vidalie, J., Ducreux, J. & Ballerini, D. 2003. Bioremediation of a crude
463 oil-polluted soil: biodegradation, leaching and toxicity assessments. Water, Air, and Soil Pollution
464 144: 419-440.

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

468 Clarke, K.R. & Gorley, R.N. 2006. PRIMER v6: User Manual/tutorial. Plymouth, UK: Primer-E
469 Ltd.

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

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

476 Eriksson, M., Ka, J.O. & Mohn, W.W. 2001. Effects of low temperature and freeze-thaw cycles on
477 hydrocarbon biodegradation in Arctic tundra soil. Applied and Environmental Microbiology 67:
478 5107-5112.

479 Feitkenhauer, H., Müller, R. & MAuml, H. 2003. Degradation of polycyclic aromatic hydrocarbons
480 and long chain alkanes at 6070 C by Thermus and Bacillus spp. Biodegradation 14: 367-372.

481 Gurska, J., Wang, W., Gerhardt, K.E., Khalid, A.M., Isherwood, D.M., Huang, X., Glick, B.R. &
482 Greenberg, B.M. 2009. Three year field test of a plant growth promoting rhizobacteria enhanced
483 phytoremediation system at a land farm for treatment of hydrocarbon waste. Environmental Science
484 & Technology 43: 4472-4479.

485 Habekost, M., Eisenhauer, N., Scheu, S., Steinbeiss, S., Weigelt, A. & Gleixner, G. 2008. Seasonal
486 changes in the soil microbial community in a grassland plant diversity gradient four years after
487 establishment. Soil Biology and Biochemistry 40: 2588-2595.

488 Hobbie, S.E. 2015. Plant species effects on nutrient cycling: revisiting litter feedbacks. Trends in
489 Ecology & Evolution .

490 Hong, S.H., Ryu, H., Kim, J. & Cho, K. 2011. Rhizoremediation of diesel-contaminated soil using
491 the plant growth-promoting rhizobacterium Gordonia sp. S2RP-17. Biodegradation 22: 593-601.

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

496 Jasinskas, A., Zaltauskas, A. & Kryzeviciene, A. 2008. The investigation of growing and using of tall
497 perennial grasses as energy crops. Biomass and Bioenergy 32: 981-987.

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

500 Kamath, R., Rentz, J., Schnoor, J.L. & Alvarez, P. 2004. Phytoremediation of hydrocarbon-
501 contaminated soils: principles and applications. Studies in Surface Science and Catalysis 151: 447-
502 478.

503 Kindt, R. & Coe, R. 2005. Tree Diversity Analysis: A Manual and Software for Common Statistical
504 Methods for Ecological and Biodiversity Studies. Nairobi, Kenya: World Agroforestry Centre.

505 Kryževičienė, A., Jasinskas, A. & Gulbinas, A. 2008. Perennial grasses as a source of bioenergy in
506 Lithuania. Agron.Res 6: 229-239.

507 Kulakow, P.A., Schwab, A. & Banks, M. 2000. Screening plant species for growth on weathered,
508 petroleum hydrocarbon-contaminated sediments. International Journal of Phytoremediation 2: 297-
509 317.

510 Leahy, J.G. & Colwell, R.R. 1990. Microbial degradation of hydrocarbons in the environment.
511 Microbiological Reviews 54: 305-315.

512 Legendre, P. & Anderson, M.J. 1999. Distance-based redundancy analysis: testing multispecies
513 responses in multifactorial ecological experiments. Ecological Monographs 69: 1-24.

514 Lindstrom, K., Jussila, M.M., Hintsa, H., Kaksonen, A., Mokelke, L., Makelainen, K., Pitkajarvi, J.
515 & Suominen, L. 2003. Potential of the Galega-Rhizobium galegae system for bioremediation of oil-
516 contaminated soil. Food Technology and Biotechnology 41: 11-16.

517 Margesin, R. & Schinner, F. 2001. Biodegradation and bioremediation of hydrocarbons in extreme
518 environments. Applied Microbiology and Biotechnology 56: 650-663.

519 Margesin, R. & Schinner, F. 1997. Laboratory bioremediation experiments with soil from a diesel-
520 oil contaminated siteation experiments with so-adapted microorganisms and fertilizers. Journal of
521 Chemical Technology and Biotechnology 70: 92-98.

522 McArdle, B.H. and Anderson M.J. Fitting multivariate models to community data: A comment on
523 distance-based redundancy analysis. *Ecol* 2001; **82**: 290-297.

524 Mikkonen, A., Hakala, K.P., Lappi, K., Kondo, E., Vaalama, A. & Suominen, L. 2012. Changes in
525 hydrocarbon groups, soil ecotoxicity and microbiology along horizontal and vertical contamination
526 gradients in an old landfarming field for oil refinery waste. Environmental Pollution 162: 374-380.

527 Mikkonen, A., Kondo, E., Lappi, K., Wallenius, K., Lindström, K., Hartikainen, H. & Suominen, L.
528 2011a. Contaminant and plant-derived changes in soil chemical and microbiological indicators during
529 fuel oil rhizoremediation with< i> Galega orientalis. Geoderma 160: 336-346.

530 Mikkonen, A., Lappi, K., Wallenius, K., Lindström, K. & Suominen, L. 2011b. Ecological inference
531 on bacterial succession using curve-based community fingerprint data analysis, demonstrated with
532 rhizoremediation experiment. FEMS Microbiology Ecology 78: 604-616.

533 Mills, D.K., Entry, J.A., Voss, J.D., Gillevet, P.M. & Mathee, K. 2006. An assessment of the
534 hypervariable domains of the 16S rRNA genes for their value in determining microbial community
535 diversity: the paradox of traditional ecological indices. FEMS Microbiology Ecology 57: 496-503.

536 Mills, D.K., Fitzgerald, K., Litchfield, C.D. & Gillevet, P.M. 2003. A comparison of DNA profiling
537 techniques for monitoring nutrient impact on microbial community composition during
538 bioremediation of petroleum-contaminated soils. Journal of Microbiological Methods 54: 57-74.

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

542 Muratova, A.Y., Dmitrieva, T., Panchenko, L. & Turkovskaya, O. 2008. Phytoremediation of Oil-
543 Sludge–Contaminated Soil. International Journal of Phytoremediation 10: 486-502.

544 Namkoong, W., Hwang, E., Park, J. & Choi, J. 2002. Bioremediation of diesel-contaminated soil with
545 composting. Environmental Pollution 119: 23-31.

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. http://CRAN.R-project.org/package=vegan.

Pajuelo, E., Rodríguez-Llorente, I.D., Lafuente, A. & Caviedes, M.Á 2011. Legume–rhizobium symbioses as a tool for bioremediation of heavy metal polluted soils. Legume–rhizobium symbioses as a tool for bioremediation of heavy metal polluted soils. Biomanagement of metal-contaminated soils. Springer. p. 95-123.

R Development Core Team. 2014. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available online at http://www.R-project.org/.

Saul, D.J., Aislabie, J.M., Brown, C.E., Harris, L. & Foght, J.M. 2005. Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. FEMS Microbiology Ecology 53: 141-155.

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.

Simarro, R., González, N., Bautista, L. & Molina, M. 2013. Assessment of the efficiency of in situ bioremediation techniques in a creosote polluted soil: change in bacterial community. Journal of Hazardous Materials 262: 158-167.

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.

568 Suzuki, M., Rappé, M.S., & Giovannoni, S.J. 1998. Kinetic bias in estimates of coastal picoplankton
569 community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length
570 heterogeneity. Applied and environmental microbiology, 64(11), 4522-4529.

571 Tiirola, M.A., Suvilampi, J.E., Kulomaa, M.S. & Rintala, J.A. 2003. Microbial diversity in a
572 thermophilic aerobic biofilm process: analysis by length heterogeneity PCR (LH-PCR). Water
573 Research 37: 2259-2268.

574 Vinas, M., Sabate, J., Espuny, M.J. & Solanas, A.M. 2005. Bacterial community dynamics and
575 polycyclic aromatic hydrocarbon degradation during bioremediation of heavily creosote-
576 contaminated soil. Applied and Environmental Microbiology 71: 7008-7018.

577 White, D.C., Flemming, C.A., Leung, K.T. & Macnaughton, S.J. 1998. In situ microbial ecology for
578 quantitative appraisal, monitoring, and risk assessment of pollution remediation in soils, the
579 subsurface, the rhizosphere and in biofilms. Journal of Microbiological Methods 32: 93-105.

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

582 Wrenn, B.A. & Venosa, A.D. 1996. Selective enumeration of aromatic and aliphatic hydrocarbon
583 degrading bacteria by a most-probable-number procedure. Canadian Journal of Microbiology 42:
584 252-258.

585 Wu, T., Milner, H., Díaz-Pérez, J.C. & Ji, P. 2015. Effects of soil management practices on soil
586 microbial communities and development of southern blight in vegetable production. Applied Soil
587 Ecology 91: 58-67.

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

591 Yan, L., Penttinen, P., Simojoki, A., Stoddard, F.L. & Lindström, K. 2015. Perennial crop growth in
592 oil-contaminated soil in a boreal climate. Science of the Total Environment 532: 752-761.

593 Zhuang, W., Tay, J., Maszenan, A., Krumholz, L. & Tay, S. 2003. Importance of Gram-positive
594 naphthalene-degrading bacteria in oil-contaminated tropical marine sediments. Letters in Applied
595 Microbiology 36: 251-257.

596 Zou, L., Tuulos, A., Mikkonen, A., Stoddard, F.L., Lindström, K., Kontro, M.H., Koponen, H. &
597 Mäkelä, P.S. 2015. Fusarium-suppressive effects of green manure of turnip rape. European Journal
598 of Soil Biology 69: 41-51.

599

600 Tables

601 Table1. Development of total soil microbial DNA concentration (ng g^{-1} fresh soil) between July

602 2009 and May 2011.

Crop treatment		mean	July 2009	May 2010	Nov. 2010	May 2011
bare fallow		18.2^{b}	15	18.7	18.6	20.4^{b}
brome grass		20.0^{ab}	15.5	21.9	20.2	22.5^{ab}
galega		21.4^{a}	16.8	22.6	21.5	24.9^{a}
mixture		20.2^{ab}	14.8	21.1	20.7	24.2^{ab}
SEM		0.7	0.8	1.8	0.6	1.1
		significance level				
Source	df	RM	UV	UV	UV	UV
crop	3	$p<0.05$	ns	ns	ns	$p<0.05$

603 *mean* the average value from the four sampling times under each crop treatment regardless of oil

604 treatment, because oil had no effect on soil DNA concentration, *SEM* standard errors of mean (SEM),

605 *ns* not significant, *RM* repeated measures ANOVA, *UV* univariate ANOVA based on each sampling

606 time. Different superscript letters (a and b) indicate significant differences ($p < 0.05$) between the

607 means of crop treatment, based on Boferroni post-hoc pairwise comparisons. The tests were based on

608 split-plot-based repeated measures (RM) ANOVA model: Y_1 = residue (error) + crop + crop ×

609 replicate + oil + PGPB + oil × PGPB + oil × crop + crop × PGPB + oil × crop × PGPB, with time as

610 the repeated factor. The factors or interactions that had no significant effects on soil physiological

611 parameters are not presented in this table.

612

613 Table 2. The development of peak-based ecological indices of bacterial communities in the PGPB-

614 untreated plots.

Sampling time	Shannon diversity (H)				Species richness (S)			
	mean	oil+	oil-	SLO	mean	oil+	oil-	SLO
July 2009	2.56^{b}(0.03)	2.48(0.05)	2.64(0.05)	$p<0.05$	19^{b}(1)	17(1)	18(1)	ns
May 2010	2.85^{a}(0.04)	2.87(0.06)	2.83(0.05)	ns	25^{a}(1)	23(2)	22(1)	ns
Nov. 2010	2.28^{d}(0.02)	2.24(0.03)	2.32(0.04)	ns	15^{c}(0.5)	14(0.5)	15(1)	ns
May 2011	2.59^{b}(0.03)	2.52(0.04)	2.66(0.04)	$p=0.051$	20^{b}(1)	18(1)	20(1)	ns
May 2012	2.35^{cd}(0.02)	2.37(0.03)	2.33(0.04)	ns	16^{c}(0.3)	15(1)	15(1)	ns
Oct. 2012	2.44^{c}(0.02)	2.42(0.03)	2.45(0.02)	ns	15^{c}(0.4)	14(0.3)	15(1)	ns

615 *SLO* significance level of oil effect analysed using univariate analysis of variance based on a split-

616 plot experimental design (model: Y_2 = residue (error) + crop + crop × replicate + oil + oil × crop), *ns*

617 not significant, *mean* the average value from all plots at each sampling time regardless of oil treatment,

618 *oil+* oil-contaminated plots and *oil-* control plots. Standard errors of mean were indicated in brackets.

619 Species richness was estimated as the number of peaks in each sample. Different superscript letters

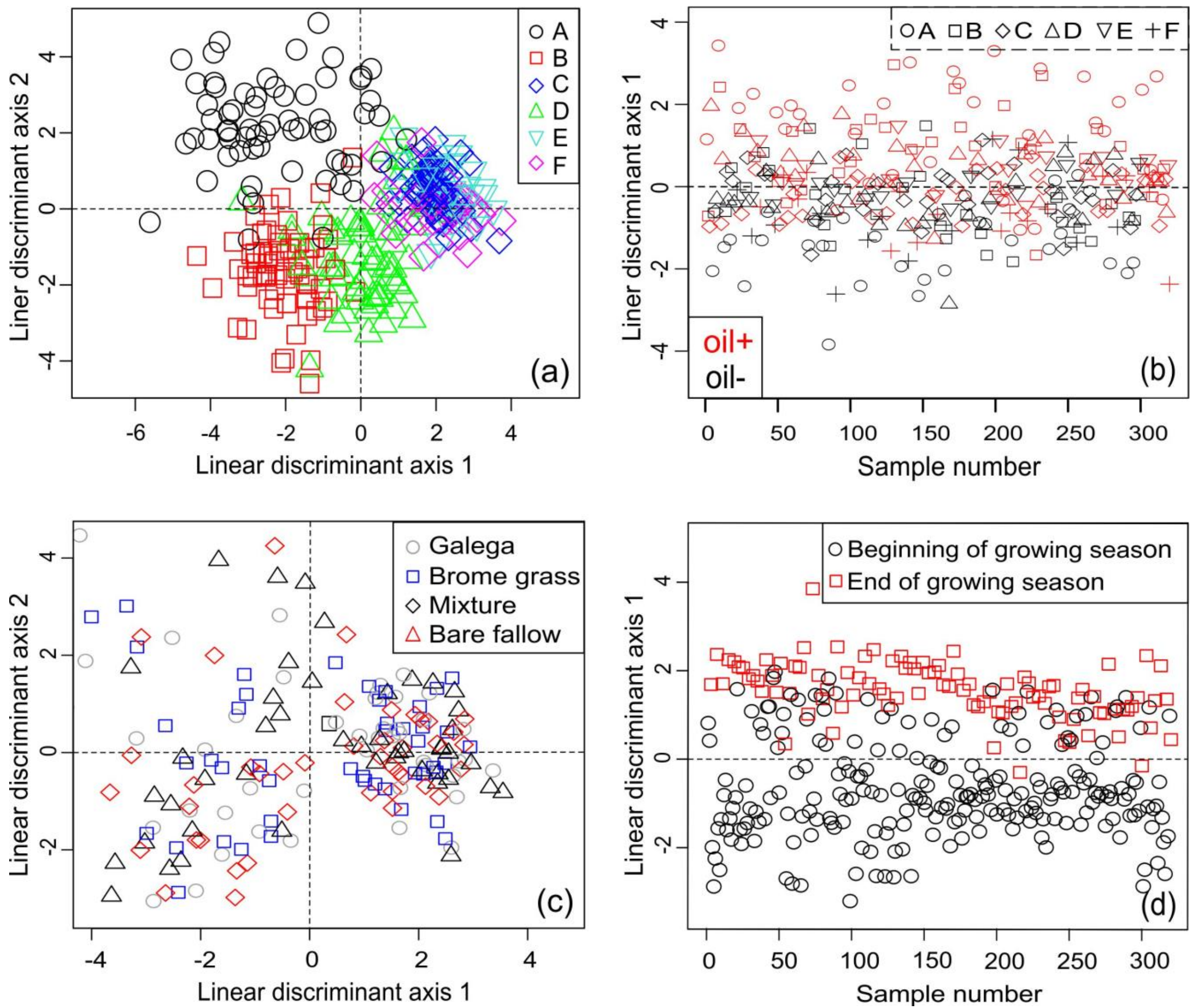
620 (a, b, c and d) indicate significant differences between the means of sampling times regardless of oil

621 treatment, based on Bonferroni post-hoc pairwise comparisons.

622

623

624 Figures



625

626 Figure 1. Differences in soil bacterial communities of *a priori* groups using distance-based

627 discriminant analysis (db-DA) of the LH-PCR curve-based data. *A priori* groups were (a) sampling

628 times: *A*: July 2009, *B*: May 2010, *C*: November 2010, *D*: May 2011, *E*: May 2012 and *F*: October

629 2012; (b) oil treatment; (c) crop treatment: galega, brome grass, galega-brome grass mixture and

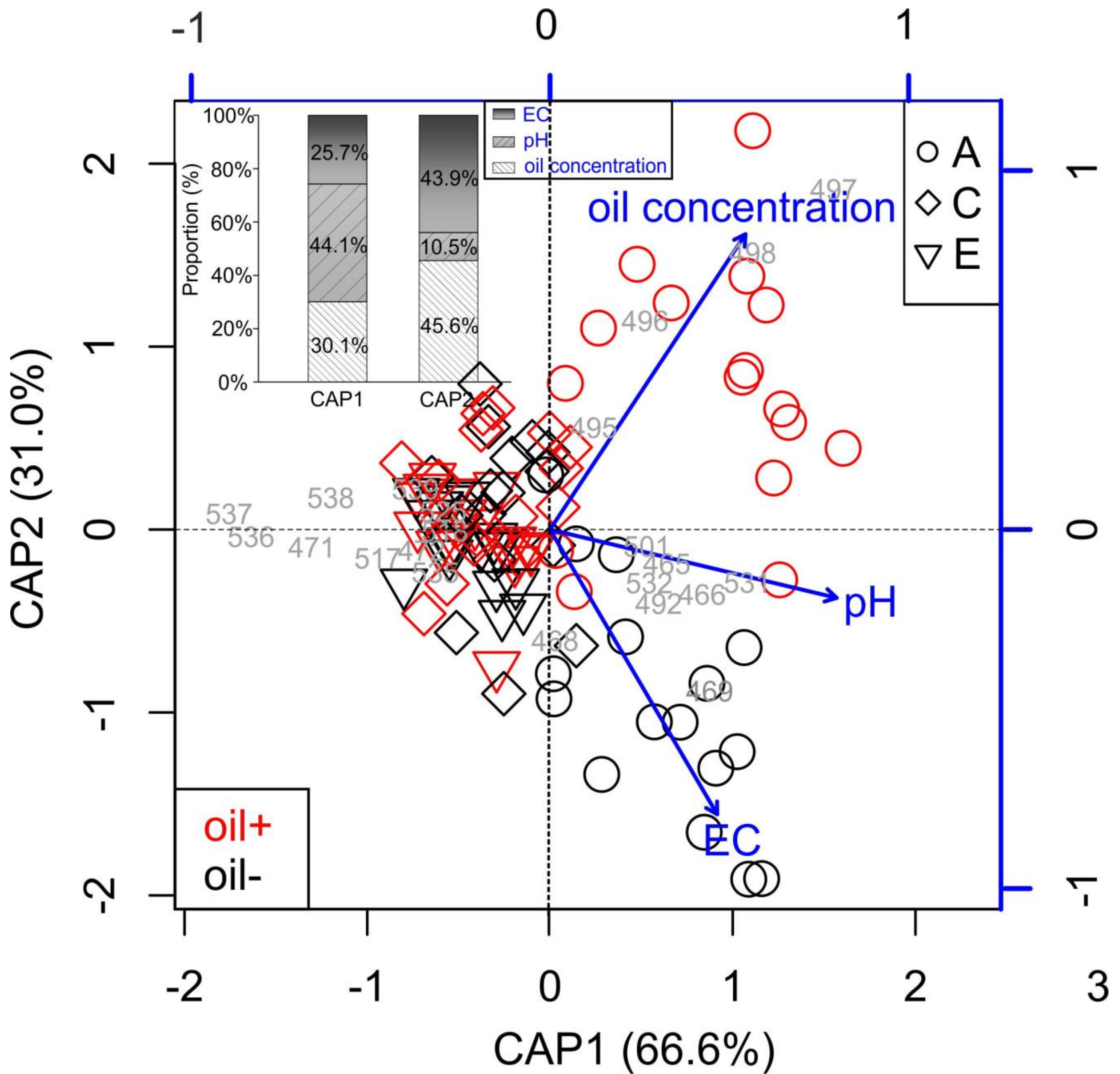
630 bare fallow and (d) the stages of a growing season: *beginning*: May-July and *end*: October-

631 November. In addition, sampling times were shown in the figure b with labels in different colors.

632 “As there was only one dimension in the figure (b) and (d) to discriminate the oil and seasonal

633 effect, sample numbers (n=320) were plotted on the x-axis. The percentage of the observations

634 which were correctly classified based on a priori hypotheses were (a) 81% (p<0.001), (b) 68%
635 (p<0.001), (c) 29% (p=0.4978) and (d) 85% (p<0.001). Altogether, 20 (a), 14 (b), 19 (c) and 19 (d)
636 principal coordinates used in the discriminant analyses explained 99.3 % (a), 93.1% (b), 98.5% (c),
637 and 98.5% (d) of the variation in LH-PCR data, respectively.



638

639 Figure 2. Relationships between bacterial community composition, bacterial taxa (LH-PCR
640 fragments) and soil parameters. Constrained analysis of principal coordinates (CAP) was performed
641 using soil chemical parameters (oil concentration, pH and electrical conductivity) as explanatory
642 variables (blue arrows) and bacterial LH-PCR curve-based community profiles (n=96) as response
643 variables. The observations were bacterial community profiles of three sampling times (*A*: July 2009,
644 *C*: November 2010, and *E*: May 2012). In order to fit the LH-PCR fragments, microbial community
645 observations and soil chemical parameters in the same figure, the scores of each LH-PCR fragment
646 on the first two CAP axes were scaled by 15 times to the eigenvalues and labeled by its amplicon

647 size. Bacterial LH-PCR fragments that fell close to the origin with scores between -0.5 and 0.5 on
648 both axes were removed. All results presented were from PGPB un-inoculated plots. Proportions of
649 soil constraining variables loaded on the first two CAP axes were calculated from their absolute
650 values of the biplot scores, illustrated in the upper-left stacked column plot.

651

652

653

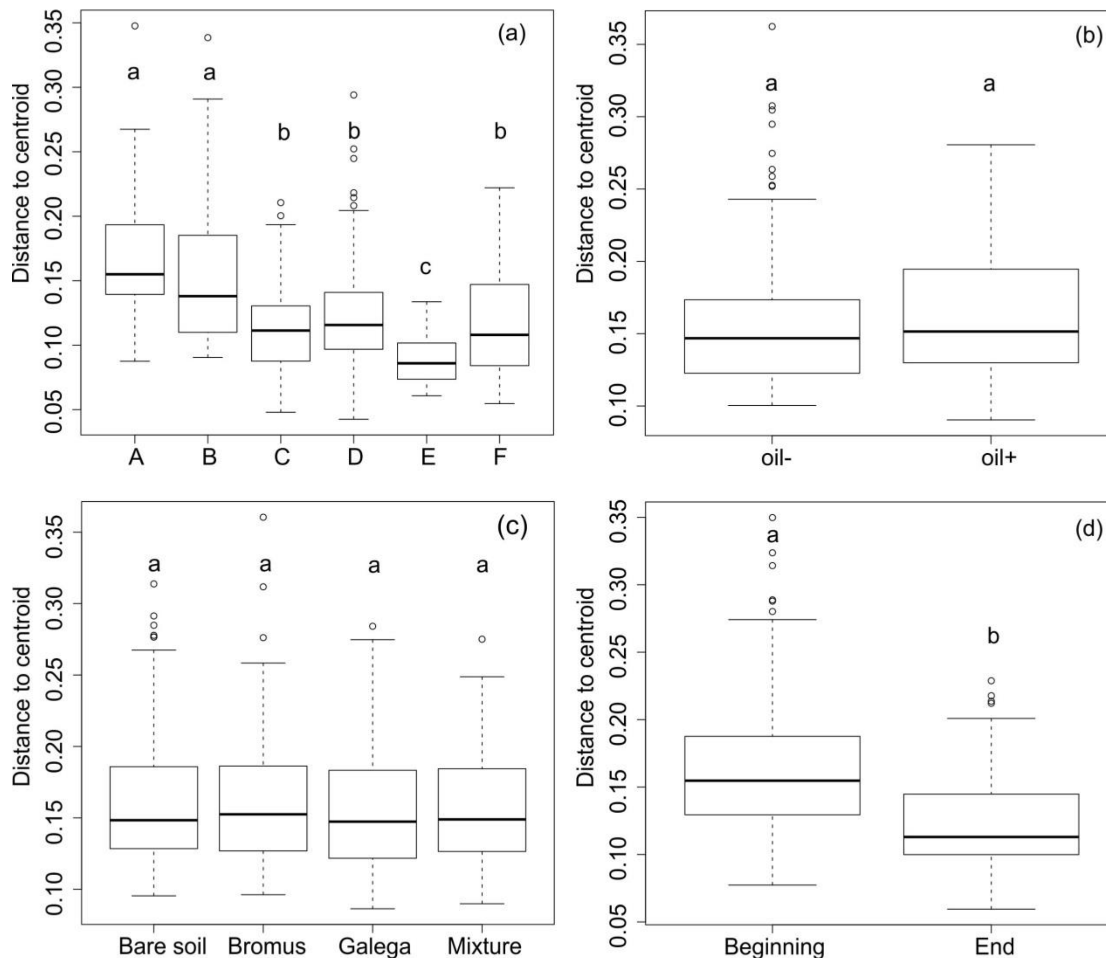
654 Supplementary materials

655 Table S1. The significance of the main effects (crop, oil and PGPB) and their interactions on bacterial
656 community composition

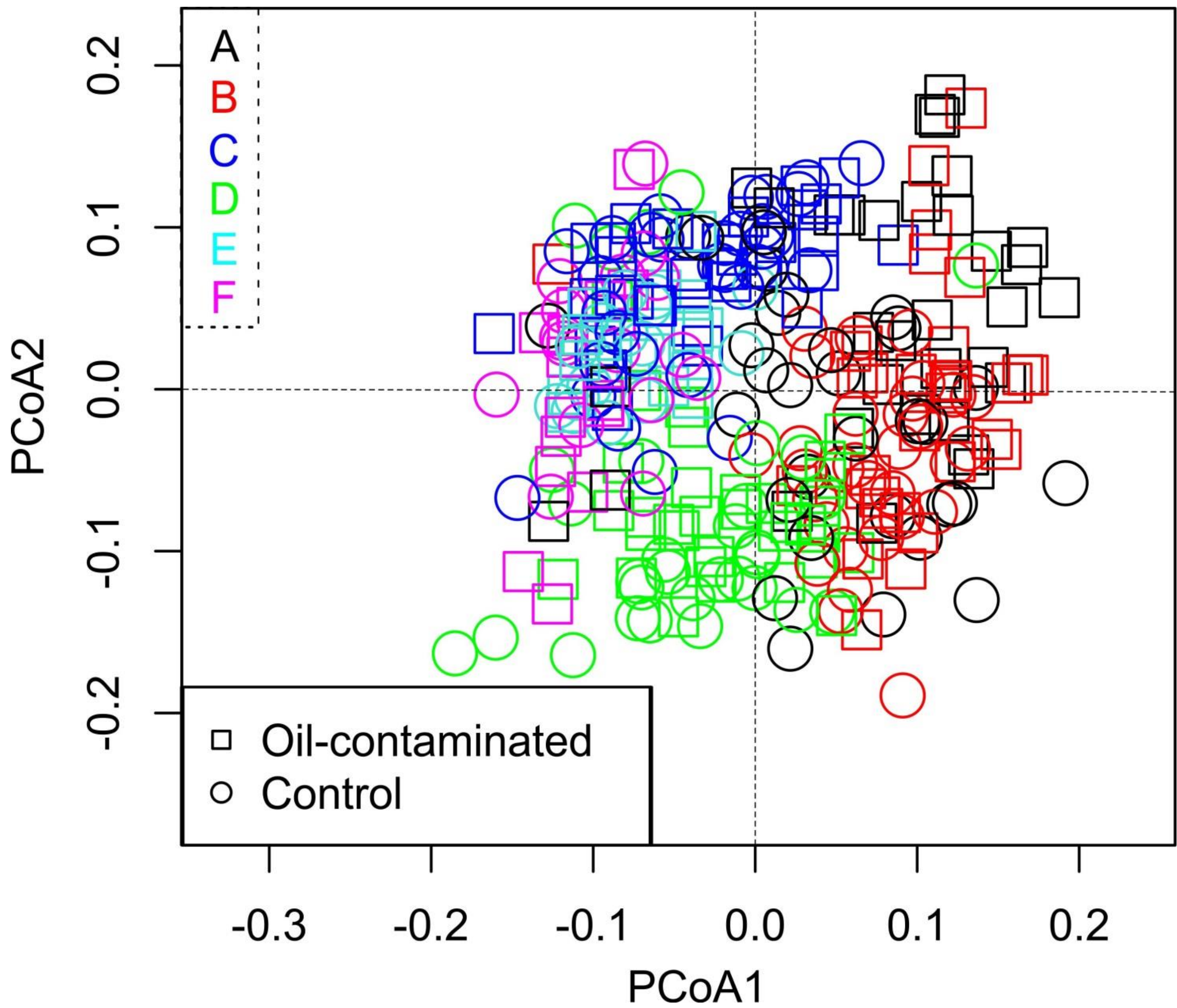
Source	df	July 2009	May 2010	Nov. 2010	May 2011	May 2012	Oct 2012
B	3	***	***	***	***	***	***
C	3	ns	ns	ns	ns	ns	ns
O	1	***	***	ns	*	ns	ns
P	1	ns	ns	ns	ns	-	-
C×O	3	ns	ns	ns	ns	ns	ns
C×P	3	ns	ns	ns	ns	-	-
O×P	1	ns	ns	ns	ns	-	-
C×O×P	3	ns	ns	ns	ns	-	-

657 *B* experimental block (replicate), *C* crop treatment, *O* oil treatment, *P* PGPB treatment, *df* degrees
658 of freedom, *ns* not significant, * $p < 0.05$ and *** $p < 0.001$. The tests were based on Bray-Curtis
659 distance-based PERMANOVA model on the basis of a split-plot design with 9999 permutations.
660 Samples taken in 2012 were all from PGPB-untreated plots, so the PGPB effect was excluded in the
661 analysis model for these two sampling sets.

662



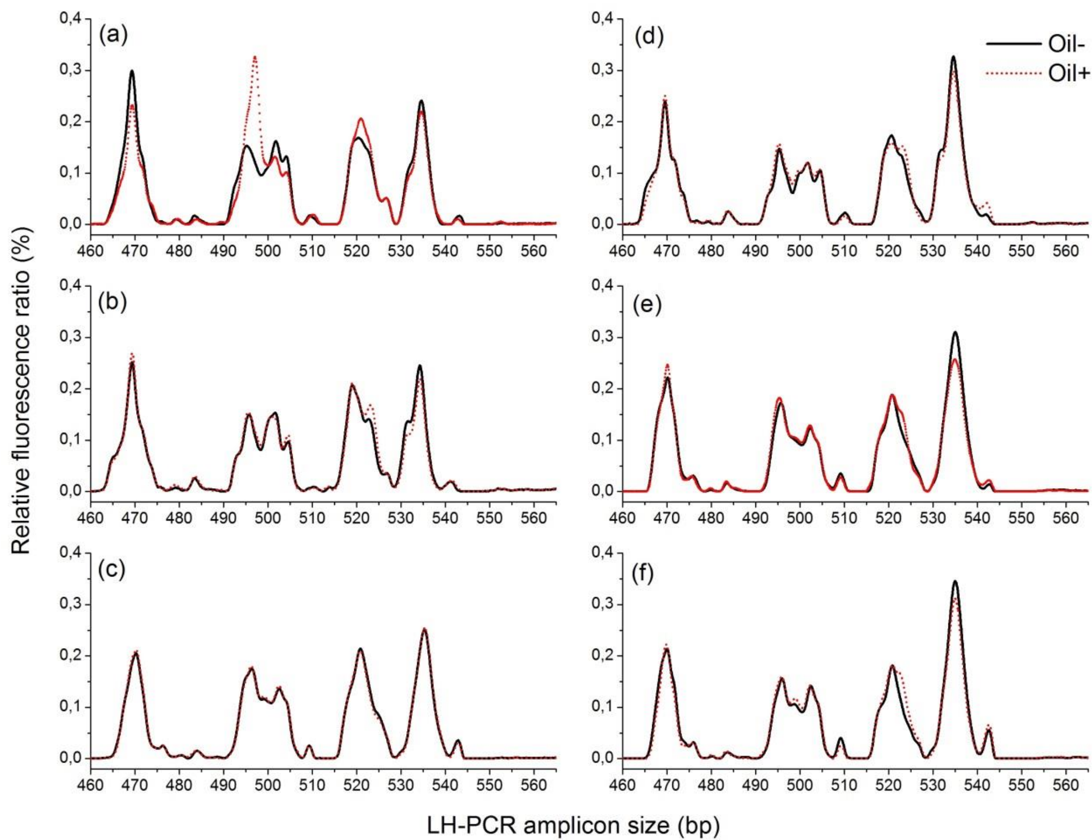
663 Figure S1. Homogeneity of variances between *a prior* groups used in the distance-based
664 discriminant analysis on bacteria LH-PCR data. A priori groups were (a) sampling times: *A*: July
665 2009 (n=64), *B*: May 2010 (n=64), *C*: November 2010 (n=64), *D*: May 2011 (n=64), *E*: May
666 2012(n=32) and *F*: October 2012 (n=32); (b) oil treatment: *oil-* control, *oil+* oil-contaminated; (c)
667 crop treatment and (d) the stages of a growing season: *beginning*: May-July and *end*: October-
668 November. The significance of homogeneity of multivariate dispersions were (a) p=0.0001, (b)
669 p=0.6112, (c) p=0.1840 and (d) p=0.0001, based on 9999 permutations. Different letters (a, b and c)
670 in the figures indicate unequal dispersions between *a prior* groups based on pairwise comparisons
671 (permuted p<0.01).



672

673 Figure S2. Principal Coordinate Analysis of 192 LH-PCR profiles obtained from PGPB-untreated
674 plots. The first two principal coordinates explain 25.3% and 21.4% of the total variation, respectively.
675 Oil-contaminated and control samples are labelled with different symbols and sampling times (*A*:
676 July 2009, *B*: May 2010, *C*: November 2010, *D*: May 2011, *E*: May 2012 and *F*: October 2012) in
677 different colors.

678



679

680 Figure S3. Averaged curve-based LH-PCR profiles between the oil-treated and control plots over
681 time: (a) July 2009, (b) May 2010, (c) November 2010, (d) May 2011, (e) May 2012 and (f) October
682 2012. Each averaged oil-treated bacterial profile was created by 16 profiles (4 crop treatments × 4
683 blocks) and each averaged control profile was created by 16 control profiles (4 crop treatments × 4
684 blocks) in PGPB-untreated plots at each sampling time using the Create Averaged Fingerprint script
685 in BioNumerics software.

686

687

688