Analysing the effect of soil organic matter on bacterial communities using T-RFLP fingerprinting: different methods, different stories?

Blaud A.[a], Diouf F.[b], Herrmann A.M.[c], Lerch T.Z. [b,c] *

[a] *Department of Civil and structural Engineering, University of Sheffield, S3 7HQ United Kingdom*

[b] *Institute of Ecology and Environmental Sciences of Paris (IEES-Paris), Université Paris-Est Créteil, 94010 Créteil Cedex, France*

[c] *Department of Chemistry & Biotechnology, Uppsala BioCentre, Swedish University of Agricultural Sciences, P.O. Box 7015, 750 07 Uppsala, Sweden*

*Corresponding author: (T.Z. Lerch)

Address: Faculté des Sciences et Technologies, Université Paris-Est Créteil,

61 avenue du Général de Gaulle, 94010 Créteil Cedex, France.

Tel: +33 1 45 17 16 60.

Fax: +33 1 45 17 19 99.

E-mail address: thomas.lerch@u-pec.fr

Abstract

Soil microbial ecology needs robust tools to elucidate ecological questions, such as the impact of fertilisation on soil microbial communities. However, the methods and data analysis used can directly affect the biological conclusions. In this study, the sensitivity of terminal-restriction fragment length polyphorism (T-RFLP) to four restriction enzymes (RE), six peak area thresholds (PAT) from 0 to 10 % and two matrices (presence/absence and relative abundance) was assessed on soils subjected to eight different long-term amendments. The T-RFLP profiles were analysed using a three-step multivariate analysis approach: (i) cluster analysis and non-metric multi-dimensional scaling, (ii) ANOSIM and PERMANOVA and (iii) correlations. The application of organic and mineral fertilisers over 53 years changed the bacterial community composition regardless if the RE, PAT and matrix were used. However, the clustering of the community, the strength of these differences, the correlations with environmental variables and, subsequently, the biological conclusions varied with the use of RE, PATand matrix. Hence, the bacterial community composition was found to be either highly sensitive to any changes in soil organic matter strongly correlated to C and N concentration, or only affected by large inputs of C or soil management. Different REs can reveal different bacterial populations affected by different drivers, but PATs 0.5 and 1 % should be used especially when using presence/absence matrix. This study also shows the complexity of the effect of organic and mineral amendment on bacterial community composition and stresses the importance to inform on methodological and data analysis parameters.

Keywords: Bacterial community structure, soil organic matter, T-RFLP, ANOSIM, PERMANOVA, RELATE

Introduction

Soil microbial ecology needs robust tools to elucidate ecological questions, which often require the analysis of large sample numbers, such as the link between soil fertility and microorganisms. Soil fertility is essential to maintain or increase soil productivity to feed the growing world population and is sustained by applying fertilisers to fields. Mineral and organic fertilisers are known not only to affect the bacterial community composition and abundance (Wessén et al. 2010; Hassan et al. 2013; Cederlund et al. 2014;Wu et al. 2014), but can also affect the functions that microorganisms can deliver to the ecosystems (Enwall et al. 2007; Lerch et al. 2013). Hence, it is essential to use robust tools not only to analyse the composition and the activity of microbial communities, but also to use accurate methods to analyse the results to understand the effects of agricultural management on these microbial properties.

Among DNA fingerprinting methods, terminal-restriction fragment length polymorphism (T-RFLP) has become a popular method to rapidly assess the composition of soil microbial communities (Thies 2007; Singh et al. 2009; Rousidou et al. 2013; Reardon et al. 2014). Although next-generation sequencing (NGS) now provides much more information on the composition of microbial communities, T-RFLP is able to capture the same major trends of bacterial community composition. For example, Elsayed et al. (2014) reported similar Shannon diversity index (Spearman's ρ=0.83, P=0.003) when using T-RFLP and pyrosequencing approaches on the same samples. Furthermore, van Dorst et al. (2014) showed that T-RFLP has a similar ability compared with 454 sequencing to separate bacterial community composition between sample locations and to identify correlations with environmental variables. As a cheaper method compared to NGS (van Dorst et al. 2014), T-RFLP allows us to analyse many true field replicates, and therefore, assesses the potential

variation arising from environmental heterogeneity or other sources of variability, which is a fundamental need in microbial ecology (Prosser 2010).

The interpretation of genetic profiles can be affected by several methodological issues such as the choice of restriction enzymes (RE; i.e. different REs will not have the same resolution to characterise microbial community composition), peak area threshold and matrix choices for data interpretation. The peak area threshold (PAT) involves the removal of peaks under a certain percentage in relation to their contribution to the entire data matrix, after alignment of samples, in order to remove false fragments and background noise. The PAT can be applied to data based on the area under the peaks (or peak height). Furthermore, data can be analysed using a variety of ordination methods and statistical tests based on relative abundances or presence/absence matrices.

We reviewed 159 articles using T-RFLP and published between 2002 and 2014 in the top five of soil biology journals to examine variation in the processing of T-RFLP data in the field of soil microbial ecology. We found that T-RFLP is still largely used in soil microbial ecology (Fig. S1), but the choice of RE, PAT, matrix and ordination/statistical analysis varied greatly between studies (Table S1). Most of the articles reported the use of only one RE, where the most frequently used were *HhaI*, *MspI*, *HaeIII*, *TaqI* and *AluI*. Less than half of all the publications indicated the size of the T-RFs used for analysis. A third of these studies used the baseline threshold (i.e. threshold based on peak fluorescence before alignment) and approximately the same proportion used the peak area/height threshold; many peak thresholds were reported from 0.1 to 5 %. When indicated, the relative abundance matrix was used twice more often than the presence/absence matrix. Although non-exhaustive, this review of the literature in soil biology revealed that T-RFLP methods are far from being harmonised, yet the effects of such parameters on the determination of soil microbial community composition and have been scarcely studied (Bennett et al. 2008).

The aim of this study was to test the robustness of T-RFLP method for assessing the effect of soil organic matter on the composition of bacterial communities. We used soil samples taken from the Ultuna long-term field experiment where previous studies showed that fertilisation strongly affected the composition of soil bacterial communities due to change in soil chemical parameters such as pH or C to N ratio (Enwall et al. 2007; Wessén et al. 2010). The sensitivity of T-RFLP fingerprints was evaluated by comparing (i) four different REs (*AluI*, *HaeIII*, *MspI* and *RsaI*) separately or combined together, (ii) six different thresholds (from 0%to 10 %) and (iii) two different types of matrices (relative abundance vs. presence/ absence). This study not only simultaneously evaluated the effects of these parameters on T-RFLP profiling, but also assessed the impact of these factors on the biological interpretation related to organic and mineral amendment using multivariate data analysis.

Material and methods

Soil sampling

Soil sampling was conducted in June 2009 (Lerch et al. 2013) at the Ultuna Long-Term Soil Organic Matter Experiment (Uppsala, Sweden; 60°N, 17°E). The experiment was started in 1956 on a post-glacial clay loam soil classified as an Eutric Cambisol (Witter et al. 1993). Since then, the soils have been treated with different N fertilisers or organic amendments. The soil texture was 36.5%clay, 41%silt and 22.5%sand. In this experiment, soils (2×2 m blocks) were treated with mineral N fertilizers (annual addition of 80 kg N ha−1) or organic amendments (biennial addition of 8 Mg ash-free organic matter per hectare). The different amendments resulted in a wide range of soil organic C contents ranging from 1 to 4 % (Table 1). The treatments were replicated in four blocks, but one of the four blocks did not have randomly distributed treatments and was therefore omitted from the current study. Eight sub-

samples from 0 to 7 cm depth were taken from each plot, sieved< 4 mm, bulked, mixed and stored at −20 °C before DNA extraction and T-RFLP analyses.

DNA extraction and purification

DNA extraction followed the ISO-11063 (Petric et al. 2010) procedure which is a modified version of the method described by Martin-Laurent et al. (2001). The procedure involved three main steps: (i) microbial cell lysis by chemical (SDS) and physical (bead beating) action, (ii) deproteination and (iii) alcohol precipitation and washing of the extracted nucleic acids. Two hundred and fifty milligrams (dry weight) of each soil sample were mixed with a solution of 100 mM Tris (pH 8.0), 100 mM EDTA (pH 8.0), 100 mM NaCl and 2 % (w/v) sodium dodecyl sulphate. Glass beads of different diameters were added in a bead-beater tube and the soil solution was shaken for 40 s at 6 m.s−1 in a mini bead-beater cell disruptor (Fast Prep, MP Bio) before centrifugation at 14, 000×g for 1 min. For protein precipitation, supernatants were incubated on ice for 10 min with 1/10 volume of 3 M sodium acetate and centrifuged (14,000×g, 5 min, 4 °C). In the last step, nucleic acids were precipitated from the collected supernatants by adding 1 volume of ice-cold isopropanol. The DNA pellets obtained after centrifugation (14,000×g, 5 min, 4 °C) were washed with 70 % ethanol. Soil DNA was purified as described by Petric et al. (2011). Finally, DNA was eluted in 100 µl of milli-Q water.

PCR and T-RFLP analysis

PCR was performed with 2 µl of diluted (1:10) DNA template (i.e. 1 ng of DNA per microliter) in a total volume of 20 µl (Master Mix Kit, Qiagen) and 0.05 mM of primer 63F (5′-CAGGCCTAACACATGCAAGTC-3′) and 1389R (5′-ACGGGCGGTGTGTACAAG-3′; Marchesi et al. 1998; Osborn et al. 2000). The forward primers were fluorescently labelled at

the 5′ end with FAM dye. PCR amplifications were carried out in a T100 thermocycler (BioRad) with an initial enaturation at 94 °C for 2 min, followed by 30 cycles of 94 °C for 30 s, 57 °C for 45 s and 72 °C for 90 s followed by a final extension time at 72 °C for 10 min. PCR products were purified using the QIAquick PCR purification kit (Qiagen) following the manufacturer's instructions.

The purified PCR product of each sample was split into four aliquots (10 µl), which were digested with 10 U of a single RE (either AluI, HaeIII, MspI or RsaI and 1× specific RE buffer (Fermentas) in a total volume of 15 µl at 37 °C for 3 h. Five microliters of the digests were desalted using a precipitation step with 0.25 µl of glycogen (20 mg ml−1) and 75 µl of 0.3 mM MgSO4.7H2O in 70 % ethanol. The solution was briefly vortexed and incubated at room temperature for 15 min, then centrifuged at 3,991×g for 30 min. The solution was removed by inverting, centrifuged for 1 min at 900×g and the pellet was resuspended in 5 µl of nuclease-free water (Qiagen). Desalted products (0.5 or 1 µl) were mixed with formamide containing 0.5 % LIZ500 internal size standard (Applied Biosystems) in a total volume of 10 µl. Desalted products were denatured at 94 °C for 3 min and electrophoresed for 20 min on an ABI 310 capillary DNA sequencer (Applied Biosystems) filled with the POP-7 polymer. The TRFLP profiles obtained with the sequencer were analysed using GeneMarker® V1.97 software (SoftGenetics). The terminal restriction fragments (T-RFs) were binned with a 0.5 bp interval. T-RFs between 50 and 500 bp and with a peak height>0 fluorescent units were included in the analysis.

Statistical analysis

The richness of the T-RFLP profiles was expressed as the total number of T-RFs, and the evenness of profiles was estimated using the Shannon index (H'; Shannon 1948). T-RF richness and Shannon index were first used to describe the overall TRFLP profiles, but were

not used as true indicators of the overall soil bacterial community richness or diversity (Blackwood et al. 2007). A three step statistical approach was then chosen to determine the influence of the different factors on T-RFLP results: (i) data ordination; (ii) tests of significant differences between treatments, REs or PATs; and (iii) correlation between bacterial community composition and environmental variables.

The influence of RE on T-RFLP results was investigated not only on the individual 4 REs mentioned above but also on the combination of the 4 RE results. It is often stated that multiple REs increase the resolution of T-RFLP (see Thies 2007), but it still unclear if the results obtained from combining several REs are more robust towards different data analyses (Bennett et al. 2009). Several matrices were produced using six different PATs ranging from 0 % to 10 %. Thresholds of 0 %, 0.1 %, 0.5 % and 1 % are commonly used with T-RFLP analyses; we also included thresholds of 5%and 10 % to assess when the use of threshold can negatively impact the T-RFLP results. To analyse the T-RFLP profile datasets (i.e. for each PAT and matrix types), the total peak area for each profile was normalised across peaks (i.e. the area under each peak was divided by the total peak area of each sample) to account for run-to-run variations. Data from the TRFLP matrices were then square root transformed and similarity matrices were constructed using the Bray-Curtis method (Clarke et al. 2006). Bray-Curtis distance was chosen because it is not affected by the number of null values between samples as with the Euclidean distance method (Clarke and Warwick 2001).

Similarities between samples were displayed using non-metric multi-dimensional scaling (nMDS) plots and dendrograms (Ramette 2007; Culman et al. 2008). Each nMDS plot was presented with a 2D Stress value, which indicated the mismatch between the rank similarity matrices and the nMDS 2D representation. A 2D stress value close to 0 indicates an excellent representation in 2D. Values above 0.2 indicate a weak 2D representation (i.e. data is more spread in 3D). To indicate percentage of similarity between samples on the nMDS,

196 the clusters from the dendrograms were overlaid onto the nMDS. Dendrograms were
197 produced using the group average linking method based on the Bray-Curtis similarity
198 matrices. Furthermore, the composition of dendrograms was tested using SIMPROF
199 (PRIMER software v6), to determine if the composition of the dendrogram was random or
200 not, i.e. if the different clusters were significantly different from each other or not (999
201 permutations).

202 Differences in bacterial community composition among soils studied, REs used and
203 PATs were tested using one-way and two-way ANOSIM analysis (100,000 permutations) on
204 the similarity matrices obtained using the Bray-Curtis method. One-way ANOSIM was used
205 to compare, for example, differences between soils for a specific RE and PAT, while two-way
206 ANOSIM was used, for example, to compare differences between soils and RE
207 simultaneously. The significance levels, i.e. P value, and R value, i.e. the strength of the
208 factors on samples were determined. R values close to 1 indicated high separation between
209 groups (e.g. between soil treatments), while R values close to 0 indicated no separation
210 between groups. Permutational multivariate analysis of variance (PERMANOVA) was also
211 used to test for difference in TRFLP profiles between soil treatments (999 permutations)
212 giving a P and F values.

213 The relationship between the bacterial community composition and the environmental
214 variables (C, N, C/N and pH) was tested by performing correlation analysis between the
215 similarity matrices of T-RFLP profiles obtained using the Bray-Curtis method and the
216 matrices of each environmental variables obtained using the Euclidean distance (Clarke and
217 Ainsworth 1993). The RELATE test from the PRIMER software was used to perform the
218 analysis, which is a permutation-based test (rank correlation method: Spearman, 999
219 permutations) giving the significance levels of the correlation, i.e. P value, and the correlation
220 strengths, i.e. Spearman coefficient ρ. The ρ value varies between 0 and 1; a ρ value close to 1

indicates a strong correlation between an environmental variable and the microbial community composition. Similarly, the relationship between the T-RFs richness or Shannon index and each environmental variable were performed using Spearman rank correlations. The Spearman rank correlation gives the significance level of the correlation (i.e. P value) and the strength of the correlation (i.e. Spearman rank coefficient ρ). The ρ value varies between -1 and +1, ρ value close to zero indicates no correlation, and value close to -1 or +1 indicates a strong negative or positive correlation, respectively. To display the Spearman correlations, heatmaps were generated using the gplots R package.

All data analysis from the T-RFLP was performed using the PRIMER software (v6, PRIMER-E Ltd, Plymouth, UK) and R version 3.1.0 (R Development Core Team 2014).

Results

Bacterial community richness and evenness

The restriction enzyme *AluI* showed a high T-RFs number 44 ±5 and Shannon index 2.9±0.3, while lower numbers were found for *HaeIII* (richness=36±14; H'=2.9±0.5), and *MspI* showed the lowest T-RFs number 30±10 (Fig. 1e) and H'=2.2±0.2 (Fig. 1). *RsaI* showed the highest T-RFs number (57±12) and Shannon index (3.2±0.4) at peak area threshold of 0 % across the different soil treatments. This hierarchy was consistent with PATs up to 1 %, but at 5 % and 10 % *RsaI* showed lower or the lowest richness and evenness. The T-RFs number and H' varied between soil treatments but none of the soil consistently showed the lowest/highest richness or H' across the different REs. Increasing PAT reduced richness at different rates for each RE (Fig. 1). On average, between 9 and 23 % of T-RFs were lost at 0.1 %, 28-47 % at 0.5 %, 42-61 % at 1 %, 82-90 % at 5 % and 91-99 % at 10 %. The 10 % PAT led to a reduction of richness down to 1-4, or even the complete loss of T-RFs for all replicates of green manure amended soils and 1 soil replicate for soils amended with farmyard manure

(Fig. 1g), and subsequently null Shannon index for several soil treatments (Fig. 1). The H' was less sensitive to the different PATs than richness. Hence, H' was only reduced between 7-13 % at 1 % PAT in comparison to 0 %, and it was only at 5 % and 10 % PATs that H' was reduced by 38-50 % and 60-85 %, respectively. When the results from all the Res were combined, the richness and H' at PAT 0 % were 167.0±24.1 and 4.2±0.2 respectively (Fig. 1i, j). The effect of PAT on richness and H' was similar for all RE, including the combination of RE.

Bacterial community composition

Different enzymatic digestions generated different bacterial community compositions, regardless of the PAT and the matrix used (Fig. 2, Fig. S2). For example, at PAT of 0.5 % from the relative abundance matrix, the bacterial community generated with *AluI* showed that soil amended with sewage sludge and peat had the lowest similarity percentage (~60 %) in comparison to the other soil treatments (Fig. 2). In contrast, *MspI* showed that green manure amended soils had the lowest similarity percentage (~60 %) in comparison with other soil treatments. The bare fallow soil treatment with ~70 % of similarity and the other soil treatments clustered in 2 groups with ~75 % of similarity between groups. Sewage sludge and peat showed more similarity (~75 %) than that generated with *AluI* (Fig. 2). Bacterial community generated with *HaeIII* and *RsaI* showed more variation than *AluI* and *MspI* between field replicates of sewage sludge, peat and sawdust amended soils (Fig. 2). The bacterial community generated with *HaeIII*, for the peat and sewage sludge treatments showed the lowest similarity with other soil treatments, as found with that generated with *AluI*. The restriction enzyme *RsaI* showed that the green manure amended soil and bare fallow had the lowest similarity (50-60 %) with other soil treatments, such as found with *MspI*. Peat amended soil showed high difference with all soil treatments but also high

variability between fields replicates (Fig. 2). When the results from each RE were combined, the cluster analysis showed similar clustering that when using AluIalone (Fig. 2). The use of presence/absence matrix increased the similarity percentage between clusters for all soils by 10 to 20 % regardless the RE, and could have led to non-significant differences between soil treatments as revealed by SIMPROF analysis (Fig. S2). When all the soil treatments were significantly different to each other with relative abundancematrix for each RE, with presence/absence matrix, many soil treatments grouped together and some soil treatments did not group together anymore (Fig. 2, S2).

The effect of PAT on bacterial community composition varied with the RE and matrix used. For example, increasing PAT did not have strong effect on nMDS for *MspI* based on relative abundance matrix between 0 % and 1 % (Fig. 3). The different soil treatments were separated from each other and the treatments replicates grouped together (Fig. 3). In contrast, at PATs of 5% and 10 %, the nMDS did not discriminate 6 out of 8 soil treatments, and the treatment replicates showed high variability and changes in similarity (Fig. 3). When the nMDS for *MspI* were generated based on presence/absence matrix, PAT had more effect on the nMDS (Fig. S3). At 0 % PAT, the soil treatments were not well separated by nMDS and treatment replicates showed high variability. Between 0.1 % and 1 %, the separation between soil treatments increased continuously and the treatment replicates grouped together. However, at 5 % and 10 %, the nMDS representation did not separate the soil treatments as found with relative abundance matrix.

ANOSIM and PERMANOVA analysis

Significant differences (P=10-6) between the bacterial community composition of the different soil treatments were observed for the 4 REs, regardless of PAT and type of matrix

(Fig. 4). However, the R (from ANOSIM) and F (from PERMANOVA) values differed greatly between REs and matrices. The restriction enzymes AluI and MspI showed the highest R (from 0.32 up to 0.97) and F (from 2.5 up to 54.0) values regardless of matrix used. In contrast, HaeIII and RsaI showed lower R (from 0.23 to 0.74) and F (from 2.2 to 13.3) values than AluI and MspI. HaeIII showed overall the lowest R values in comparison with other REs but similar F values than RsaI. When the REs results were combined, the R values were slightly lower than for AluI and MspI, but the F values were much lower than for AluI and MspI (up to 3.6 times lower).

The R and F values increased for all REs and both matrices with increasing PATs from 0.5 % to 1 % and then decreased (except for the F values of HaeIII that increased continuously). The R and F values reached their highest values at different PATs, but were not different for both matrices, for the different REs: for AluI at 0.5 %, for MspI and RsaI at 1 %, and for HaeII at 5 % (except for HaeIII F value: 10%). The Rand F values often decreased sharply after the highest value for the different REs and both matrices, leading sometimes to R values inferior than PAT 0%The R and F values showed more sensitivity to increase in PAT when the analyses were based on the presence/absence matrix rather than the relative abundance matrix (Fig. 4).

The bacterial community composition generated by HaeIII and RsaI showed variability between replicates of the sewage sludge amended soil for HaeIII and, peat and saw dust amended soil for RsaI (Fig. 2), which could affect the results of ANOSIM, PERMANOVA and RELATE test. However, when the ANOSIM and RELATE tests were performed for both REs without the soil treatments showing variability, and for the different PAT and matrices, HaeIII showed similar R and ρ values and RsaI showed similar ρ values but 0.2 to 0.3 times higher R values than when the variable replicates were included.

Relationship between composition of bacterial community and soil chemical properties

The bacterial community composition showed different relationships with the C and N content, C/N and soil pH for the four REs and combined REs results (Fig. 5). Hence, bacterial community composition generated with *AluI*, *HaeIII* and combination of all REs results were significantly (P=0.05) correlated to each of the soil chemical properties. However, it was not always significantly correlated for some specific PATs. The communities generated with *AluI*, *HaeIII* and the combined RE results, showed strong correlations with C content (ρ values up to 0.65), while the strength of the correlations with N content, C/N and soil pH were more variable, influenced by PAT and matrix. In contrast, the bacterial community composition generated with *MspI* and *RsaI* were not significantly (P≤0.05) correlated with all of the soil chemical properties. The bacterial community generated with *MspI* showed only ρ values that were relatively high for C content (up to 0.4 for PAT 0.1 % and presence/absence matrix) for both matrices (Fig. 5). The bacterial community composition generated with *RsaI* was only significantly correlated to C and C/N, regardless of the matrix used, with ρ values up to 0.4 and 0.56 for C and C/N, respectively (presence/absence matrix). The use of either relative abundance or presence/absence matrix can affect the strength of the correlation and the significance between bacterial community composition and the soil chemical properties.

The effect of PAT on the correlations between bacterial community composition and the soil chemical properties mainly depend on the matrix used. The strength of thecorrelations from bacterial community composition generated with relative abundance matrix were weakly affected by the PATs commonly used (i.e. 0 – 1 %), but decreased with PATs ≥ 5 %, decreasing the ρ values from 0.1 to 0.3 (Fig. 5). In contrast, correlations based on presence/absence matrix were more strongly affected by PATs, either increasing or

decreasing the strength and significance of the correlations. Increasing PATs, increased the ρ values between bacterial community and some chemical properties, but often decreased at PATs 5 % and 10 %. In contrast, increasing PAT had a negative effect on the correlations between bacterial community composition and all soil chemical properties for *MspI* and *RsaI* REs (Fig. 5).

The relationship between the T-RFLP profiles richness or evenness (Shannon index) and environmental variables showed different patterns between REs and PATs (Fig. 6). *MspI* and the combined results of all REs showed the highest number of significant correlations (16), and both were negatively correlated with C and N content. In contrast, *HaeIII* and *RsaI* showed nearly no significant correlations with the environmental variables (5 for both REs). Then, *AluI* showed a high number of significant correlations (11) with most environmental variables, but the correlations were greatly affected by PAT, as some correlations were either positive, negative and non-significant for the same variables depending on PAT (Fig. 6). In contrast, the other REs were affected in the same way by PAT, i.e. correlations became significant or stopped being significant. For each RE, PATs ≥5 % decreased the number of significant correlations from 2 to 7 times, while PATs 0.1 %, 0.5%and 1%showed high number of significant correlations (15, 11 and 10, respectively).

Discussion

Influence of soil organic matter on bacterial community composition

The application of organic materials and N fertilisers over 53 years significantly changed the amount and the quality the soil organic matter, and subsequently, the bacterial community composition. The cluster and nMDS analysis showed clear separation between all treatments, which was confirmed by the SIMPROF, ANOSIM and PERMANOVA analysis. Hence, the treatments likely affected the composition of bacterial community via direct effects, due to

changes in C and N concentration and C/N ratio, or indirect changes, for example, due to soil acidification from the treatment applications. This result agrees with previous studies investigating the bacterial community composition at the Ultuna experiment using TRFLP (Enwall et al. 2007; Hallin et al. 2009; Cederlund et al. 2014) or other fingerprinting methods such as ribosomal intergenic spacer analysis (Enwall et al. 2005) or phospholipid fatty acids (PLFA; Elfstrand et al. 2007; Börjesson et al. 2012).

Each soil treatment grouped separately, with high similarity between field replicates, highlighting sufficient resolution of T-RFLP to discriminate the communities. Hence, each treatment was affected to some extent by the organic or inorganic amendment, showing specific bacterial community composition. The sewage sludge treatment was previously found to harbour distinct bacterial community compositions in comparison to the other treatments (Enwall et al. 2005, 2007; Elfstrand et al. 2007; Hallin et al. 2009), which was attributed to the low soil pH (Enwall et al. 2007) and mainly to the high heavy metal concentration (such as Cd and Pb) in soil comingfrom the sewage sludge applied (Enwall et al. 2005; Börjesson et al. 2012). The bacterial community composition of the peat treatment was also previously found to be different from the one of straw (with or without addition of Ca(NO3)2) and unfertilised treatment. The different bacterial community composition of the peat treated soil could be related to C and N concentration and C/N ratio, which were higher for the peat treatment than any other treatments.

Different restriction enzymes reveal different stories

Despite the clear differences in bacterial community composition generated by T-RFLP between the treatments, the bacterial communities clustered differently in relation to the RE used, directly influencing the biological conclusions. When the T-RFLPs were generated using *AluI*, *HaeIII* or the combination of all the REs, the soil treated with sewage sludge and

394 peat showed clear differences in bacterial community composition compared to the other soil
395 treatments. In this case, T-RFLP profiles were strongly correlated with C and N
396 concentration, and to a lesser extent, to C/N and soil pH, confirming previous studies
397 (Elfstrand et al. 2007; Hallin et al. 2009; Cederlund et al. 2014). In contrast, when the
398 bacterial community composition was generated with *MspI*, the green manure and bare fallow
399 treatments showed the most distinct bacterial community composition from the other
400 treatments. Using *RsaI*, results showed that peat, green manure and bare fallow were the
401 treatments with distinct bacterial communities. A previous study showed that the green
402 manure treatment harbours a distinct bacterial community composition generated by PLFA,
403 but peat and sewage sludge treatment were not included in the analysis (Elfstrand et al.
404 2007).When the T-RFLP profiles were generated by *MspI* and *RsaI*, only weak correlations
405 with a few variables were found, indicating that the differences in microbial community
406 composition depended on other environmental variables (Elfstrand et al. 2007; Hallin et al.
407 2009). The specific bacterial community of the bare fallow could be due to the fact that the
408 bare fallow treatment is the only one treatment where no crops were grown and was weeded
409 manually, leaving a bare soil. In contract, the specific bacterial community of the green
410 manure treated soil is likely to be related to the nature of the amendment, which showed the
411 lowest humification coefficient in comparison to other organic amendments (Kätterer et al.
412 2011), highlighting high and rapid mineralisation that could favour fast growing bacteria.
413 Many studies have shown that the use of different Res gives different results, i.e.
414 different profiles (Burke et al. 2005; Osborne et al. 2006; Bennett et al. 2008, 2009; Kasel et
415 al. 2008; Barkovskii et al. 2009). The selection of RE is usually empirical or based on the
416 enzyme that gives the highest number of T-RFs for the gene of interest, with the expectation
417 that it will give the best representation of the community composition (Marsh 2005). This
418 selection is often obtained by *in silico* digestion of sequences. However, the RE commonly

used to digest the bacterial 16S rRNA gene (i.e. *HhaI*, *MspI*, *HaeIII*, *RsaI* and *AluI*) can give similar numbers of T-RFs and/or peak areas for the same samples (Osborne et al. 2006; Bennett et al. 2008). Here we performed in silico digestion on 51 sequences previously published from the

Ultuna experiment (Sessitch et al. 2001). We found that different REs can result in similar numbers of T-RFs with in silico digestion or in T-RFLP, and that the results between in silico digestion and T-RFLP can differ (Table S2). This study clearly shows that different REs target different bacterial populations which are affected by the treatments, but in different ways, due to their sensitivities to different chemical parameters such as pH, C and N contents in soil. Thus, selecting an appropriate enzyme is more complex than purely selecting the one that gives the highest number of T-RFs. This means that empirical selection is crucial (Schütte et al. 2008) and should be based on existing literature as it also vary with the primers pair used and the targeted gene.

Several studies have stressed the importance of using multiple REs (see Thies 2007), either by simultaneous digestion (Bastias et al. 2007; Kluber et al. 2011; Aislabie et al. 2012; Godin et al. 2012) or by combining the data (Klamer and Hedlund 2004; Kasel et al. 2008; Bennett et al. 2009; Trabelsi et al. 2012) to obtain an accurate representation of the microbial community composition (Marsh et al. 2000). In the present study, the effect of soil treatments on bacterial community composition and the relationship between community composition and environmental variables were lower when the four REs were combined, and when only *AluI* and *MspI* were combined, most of the soil treatment replicates did not group together and soil treatment effect was low (data not shown). The combination of all the RE results may represent a summary of the results showing only the strongest change in the composition of bacterial community due to the most important factors (here, soil C and N content).

444 **Variability in the effects of peak area threshold**

445 Although the selection of RE was clearly the dominant factor affecting microbiological
446 conclusions, the use of different peak area thresholds (PAT) and relative abundance or
447 presence/absence matrices also affected the interpretation. The effect of PAT on bacterial
448 community composition varied with RE, the matrix used and statistical analysis. Peak area
449 threshold had a stronger effect on bacterial community composition when generated with
450 presence/absence matrix rather than relative abundance. This was expected as presence/
451 absence give the same weight to all the T-RFs and is likely to be more sensitive to different
452 noise thresholds than relative abundance due to the loss of T-RFs (Clarke 1993; Bennett
453 et al. 2008).

454 Overall, the use of PAT improved the separation/significance between microbial
455 groups and correlations with environmental variables. The positive effect of PAT is likely
456 due to the reduction of the variability in T-RF richness, which could be related to background
457 noise (Fig. 1). The peak area threshold between 0.5 % and 1 % gave the best results for the
458 different REs. Peak area thresholds>1 % had overall a negative effect on ordination and
459 statistical analyses and should therefore not be used for most REs. Peak area threshold of 0.1
460 % did not result in high R and ρ values, indicating that it may not be a strong threshold to
461 improve the results, which was confirmed by the absence of significant differences between
462 bacterial community composition obtained with 0 % and 0.1 % PAT (data not shown). This
463 study stresses the importance of using peak threshold especially for presence/absence matrix,
464 as only 29 % of the studies surveyed (Table S1) used a threshold and about77% of the studies
465 that used presence/absence matrix did not apply any peak threshold on the data.

466

467 **Relative abundance matrix: a robust method?**

The choice of matrix between presence/absence and relative abundance depends on whether the specific aim of the study is qualitative (presence/absence) or quantitative (relative abundance). The presence/absence matrix gives the same weight to all the T-RFs, i.e. rare T-RFs have the same impact on the data than abundant T-RFs and can be highlighted in the results, while abundant peaks dominate the relative abundance matrix. Thus, the type of used matrix can have a direct effect on the results and the biological conclusions. The presence/absence matrix showed that only high input of C≥23 % and N>2 % can affect bacterial community composition. This suggests that bacteria are resilient to small changes in soil C and N content even over long periods of time. Hence, presence/absence matrix showed only the strong differences between samples, reducing the complexity of the results. Rees et al. (2004) also showed that relative abundance matrix (peak area) over presence/absence matrix, generated better nMDS representation (i.e. better separation between groups) and higher R values . In the present study, the relative abundance matrix appeared to be more robust and reliable than presence/absence method. Thus, based on the results, the use of relative abundance matrix is recommended over presence/absence matrix to investigate complex bacterial community composition and to reveal the full extent of the changes in microbial community composition.

Conclusions

Fifty three years of organic amendments and the addition of N fertilisers strongly changed the composition of bacterial community of all the treatments, with the sewage sludge and peat treated soil being the most affected. The C and N concentration (and to a lesser extent C/N ratio and soil pH) were identified as the main drivers of these differences in the composition bacterial community. However, biological conclusions found in this study were clearly affected by the methods and data analysis. The selection of RE was found to be the main

493 factor influencing T-RFLP, highlighting the importance of empirical and literature based RE
494 selection. Nevertheless, different REs can reveal different bacterial populations. PAT also
495 affect the results using the presence/absence matrix, and to a lesser extent, using relative
496 abundance matrix. Thus, PATs of 0.5 % or 1 % were found to be the most appropriate for
497 determining meaningful biological conclusions. The relative abundance matrix was found to
498 be a robust and reliable measure in comparison to presence/absence matrix, as relative
499 abundance was less sensitive to PAT, generated less variable results and revealed the full
500 complexity of the results, whilst the presence/absence matrix lost information. This study also
501 demonstrates the importance of using a variety of multivariate analysis to fully assess the
502 effect of different factors on T-RFLP and to obtain accurate biological conclusions. Here, we
503 demonstrated that cluster or nMDS analysis alone is not sufficient. As suggested by Rees et
504 al. (2004), the use of statistical test such as ANOSIM or PERMANOVA is essential for data
505 interpretation.
506
507 **Acknowledgments**
508 The authors would like to thank the Department of Soil and Environment, SLU (Uppsala,
509 Sweden) for access to soil samples and managing the Ultuna Long-Term Soil Organic Matter
510 Experiment. We would also like to thank Dr Susan Johnston for proof reading the article.
511
512 **References**
513 Aislabie JM, Ryburn J, Gutierrez-Zamora M-L, Rhodes P, Hunter D, Sarmah AK, Barker
514 GM, Farrell RL, (2012) Hexadecane mineralization activity in hydrocarbon-
515 contaminated soils of Ross Sea region Antarctica may require nutrients and
516 inoculation. Soil Biol Biochem 45:49–60

517 Barkovskii AL, Fukui H, Leisen J, Kim S-H, Marsh TL, Khijniak AI (2009) Rearrangement
518 of bacterial community structure during peat diagenesis. Soil Biol Biochem 41:135–
519 143
520 Bastias BA, Anderson IC, Xu Z, Cairney JWG (2007) RNA- and DNA-based profiling of soil
521 fungal communities in a native Australian eucalypt forest and adjacent Pinus elliotti
522 plantation. Soil Biol Biochem 39:3108–3114
523 Bennett LT, Kasel S, Tibbits J (2008) Non-parametric multivariate comparisons of soil fungal
524 composition: Sensitivity to thresholds and indications of structural redundancy in T-
525 RFLP data. Soil Biol Biochem 40:1601–1611
526 Bennett LT, Kasel S, Tibbits J (2009) Woodland trees modulate soil resources and conserve
527 fungal diversity in fragmented landscapes. Soil Biol Biochem 41:2162–2169
528 Blackwood CB, Hudleston D, Zak DR, Buyer JS (2007) Interpreting ecological diversity
529 indices Applied to terminal restriction fragment length polymorphism data: insights
530 from simulated microbial communities. Appl Environ Microbiol 73:5276–5283
531 Börjesson G, Menichetti L, Kirchmann H, Kätterer T (2012) Soil microbial community
532 structure affected by 53 years of nitrogen fertilisation and different organic
533 amendments. Biol Fertil Soils 48:245–257
534 Burke DJ, Martin KJ, Rygiewicz PT, Topa MA (2005) Ectomycorrhizal fungi identification
535 in single and pooled root samples: terminal restriction fragment length polymorphism
536 (TRFLP) and morphotyping compared. Soil Biol Biochem 37:1683–1694
537 Cederlund H, Wessén E, Enwall K, Jones CM, Juhanson J, Pell M, Philippot L, Hallin S
538 (2014) Soil carbon quality and nitrogen fertilization structure bacterial communities
539 with predictable responses of major bacterial phyla. Appl Soil Ecol 84:62–68
540 Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure.
541 Aus J Ecol 18:117–143

542 Clarke KR, Ainsworth M (1993) A method of linking multivariate community structure to
543 environmental variables. Mar Ecol Prog Ser 92:205–219.
544 Clarke KR, Somerfield PJ, Chapman MG (2006) On resemblance measures for ecological
545 studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis
546 coefficient for denuded assemblages. J Exp Mar Biol Ecol 330:55–80
547 Clarke KR, Warwick RM (2001) Change in marine communities: an approach to statistical
548 analysis and interpretation, 2nd edition. PRIMER-E: Plymouth.
549 Culman SW, Gauch HG, Blackwood CB, Thies JE (2008) Analysis of T-RFLP data using
550 analysis of variance and ordination methods: A comparative study. J Microbiol Meth
551 75:55–63
552 Elfstrand S, Hedlund K, Mårtensson A (2007) Soil enzyme activities, microbial community
553 composition and function after 47 years of continuous green manuring. Appl Soil Ecol
554 35:610–621
555 Elsayed OF, Maillard E, Vuilleumier S, Imfeld G (2014) Bacterial communities in batch and
556 continuous-flow wetlands treating the herbicide S-metolachlor. Sci Total Environ
557 499:327–335
558 Enwall K, Nyberg K, Bertilsson S, Cederlund H, Stenström J, Hallin, S (2007) Long-term
559 impact of fertilization on activity and composition of bacterial communities and
560 metabolic guilds in agricultural soil. Soil Biol Biochem 39:106–115
561 Enwall K, Philippot L, Hallin S (2005) Activity and composition of the denitrifying bacterial
562 community respond differently to long-term fertilization. Appl Environ Microbiol
563 71:8335–8343
564 Godin A, McLaughlin JW, Webster KL, Packalen M, Basiliko N (2012) Methane and
565 methanogen community dynamics across a boreal peatland nutrient gradient. Soil Biol
566 Biochem 48:96–105

Hallin S, Jones CM, Schloter M, Philippot L (2009) Relationship between N-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment. ISME J 3:597–605.

Hassan SED, Liu A, Bittman S, Forge TA, Hunt DE, Hijri M, St-Arnaud M (2013) Impact of 12-year field treatments with organic and inorganic fertilizers on crop productivity and mycorrhizal community structure. Biol Fertil Soils 49:1109–1121

Kasel S, Bennett LT, Tibbits J (2008) Land use influences soil fungal community composition across central Victoria, south-eastern Australia. Soil Biol Biochem 40:1724–1732

Kätterer T, Bolinder MA, Andrén O, Kirchmann H, Menichetti L (2011) Roots contribute more to refractory soil organic matter than above-ground crop residues, as revealed by a long-term field experiment. Ag Ecosyst Environ 141:184–192

Klamer M, Hedlund K (2004) Fungal diversity in set-aide agricultural soil investigated using terminal-restriction fragment length polymorphism. Soil Biol Biochem 36:983–988

Kluber LA, Smith JE, Myrold DD (2011) Distinctive fungal and bacterial communities are associated with mats formed by ectomycorrhizal fungi. Soil Biol Biochem 43:1042–1050

Lerch TZ, Coucheney E, Herrmann AM (2013) Sensitivity of soil microbial catabolic profiles to a gradient of carbon inputs: Does the soil organic matter matter? Soil Biol Biochem 57:911–915

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Wade WG (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol 64:795–799

Marsh TL (2005) Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. Methods Enzymol 397:308–329

592 Marsh TL, Saxman P, Cole J, Tiedje J (2000) Terminal restriction fragment length
593 polymorphism analysis program, a web-based research tool for microbial community
594 analysis. Appl Environ Microbiol 66:3616–3620
595 Martin-Laurent F, Philippot L, Hallet S, Chaussod R, Germon JC, Soulas G, Catroux G
596 (2001) DNA Extraction from soils: old bias for new microbial diversity analysis
597 methods. Appl Environ Microbiol 67:2354–2359
598 Osborn AM, Moore ERB, Timmis KN (2000) An evaluation of terminal-restriction fragment
599 length polymorphism (T-RFLP) analysis for the study of microbial community
600 structure and dynamics. Environ Microbiol 2:39–50
601 Osborne CA, Rees GN, Bernstein Y, Janssen PH (2006) New threshold and confidence
602 estimates for terminal restriction fragment length polymorphism analysis of complex
603 bacterial communities. Appl Environ Microbiol 72:1270–1278
604 Petric I, Philippot L, Abbate C, et al (2011) Inter-laboratory evaluation of the ISO standard
605 11063 “Soil quality — Method to directly extract DNA from soil samples.” J
606 Microbiol Meth 84:454–460
607 Prosser JI (2010) Replicate or lie. Environ Microbiol 12:1806–1810
608 Ramette A (2007) Multivariate analyses in microbial ecology. FEMS Microbiol Ecol 62:142–
609 160
610 R Development Core Team (2014) R: a language and environment for statistical computing.
611 Reardon CL, Gollany HT, Wuest SB (2014) Diazotroph community structure and abundance
612 in wheat–fallow and wheat–pea crop rotations. Soil Biol Biochem 69:406–412
613 Rees GN, Baldwin DS, Watson GO, Perryman S, Nielsen DL (2004) Ordination and
614 significance testing of microbial community composition derived from terminal
615 restriction fragment length polymorphisms: application of multivariate statistics.
616 Anton Leeuw 86:339–347

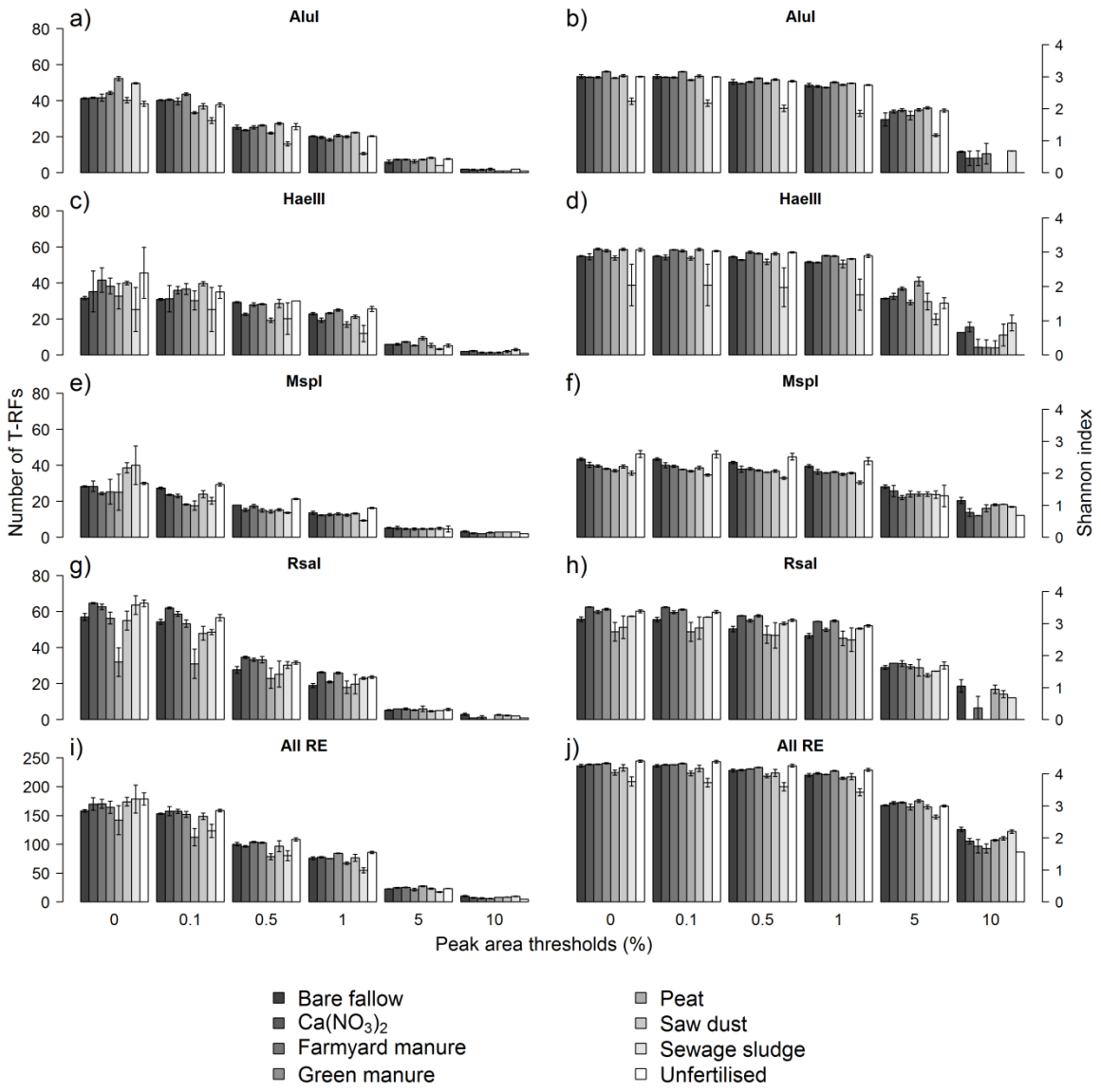
617 Rousidou C, Papadopoulou ES, Kortsinidou M, Giannakou M, Singh BK, Menkissoglu-
618 Spiroudi U, Karpouzas DG (2013) Bio-pesticides: Harmful or harmless to ammonia
619 oxidizing microorganisms? The case of a Paecilomyces lilacinus-based nematicide.
620 Soil Biol Biochem 67:98–105
621 Schütte U, Abdo Z, Bent S, Shyu C, Williams C, Pierson J, Forney L (2008) Advances in the
622 use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S
623 rRNA genes to characterize microbial communities. Appl Microbiol Biot 80:365–380
624 Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423.
625 Singh BK, Tate KR, Ross DJ, Singh DJ, Dando J, Thomas N, Milliard P (2009) Soil methane
626 oxidation and methanotroph responses to afforestation of pastures with Pinus radiata
627 stands. Soil Biol Biochem 41:2196–2205
628 Thies JE (2007) Soil microbial community analysis using terminal restriction fragment length
629 polymorphisms. Soil Sci Soc Am J 71:579–591
630 Trabelsi D, Ben Ammar H, Mengoni A, Mhamdi R (2012) Appraisal of the crop-rotation
631 effect of rhizobial inoculation on potato cropping systems in relation to soil bacterial
632 communities. Soil Biol Biochem 54:1–6
633 Van Dorst J, Bissett A, Palmer AS, et al (2014) Community fingerprinting in a sequencing
634 world. FEMS Microbiol Ecol 89:316–330
635 Wessén E, Nyberg K, Jansson JK, Hallin S (2010) Responses of bacterial and archaeal
636 ammonia oxidizers to soil organic and fertilizer amendments under long-term
637 management. Appl Soil Ecol 45:193–200
638 Witter E, Mårtensson AM, Garcia FV (1993) Size of the soil microbial biomass in a long-
639 term field experiment as affected by different n-fertilizers and organic manures. Soil
640 Biol Biochem 25:659–669

641 Wu K, Yuan S, Wang L, Shi J, Zao J, Shen B, Shen Q (2014) Effects of bio-organic fertilizer
642 plus soil amendment on the control of tobacco bacterial wilt and composition of soil
643 bacterial communities. Biol Fert Soils 50:961–971

644 **Table 1.** Chemical characteristics of the 8 different soil treatments of the Ultuna Long
645 Term Field Experiment (Uppsala, Sweden). Means values ± standard errors (n = 3) are
646 shown.

Soils treatments	Total C (mg g^{-1} soil)	Total N (mg g^{-1} soil)	C/N	pH (water)
Bare fallow	9.8 ± 0.08	1.06 ± 0.01	9.2 ± 0.01	6.1 ± 0.03
$Ca(CO_3)_2$	14.0 ± 0.12	1.45 ± 0.01	9.6 ± 0.03	6.7 ± 0.02
Farmyard manure	23.0 ± 0.08	2.27 ± 0.003	10.1 ± 0.02	6.5 ± 0.03
Green manure	16.9 ± 0.07	1.73 ± 0.01	9.8 ± 0.02	6.1 ± 0.02
Peat	38.2 ± 0.46	2.07 ± 0.01	18.5 ± 0.14	6.1 ± 0.02
Saw dust	20.9 ± 0.45	1.49 ± 0.02	14.0 ± 0.14	6.3 ± 0.04
Sewage sludge	28.6 ± 0.19	3.08 ± 0.02	9.3 ± 0.01	4.9 ± 0.02
Unfertilised	11.1 ± 0.04	1.18 ± 0.004	9.4 ± 0.03	6.2 ± 0.02

647
648



649

650 **Fig. 1.** Variation in the number of T-RFs (a, c, e, g, i) and the Shannon index (b, d, f, h, j)
651 from the bacterial community structures in the eight soil treatments studied, at six peak area
652 thresholds and for each of the four enzymatic digestions (*AluI*, *HaeIII*, *MspI* and *RsaI*) and the
653 combination of the four restriction enzymes results (All RE). Mean values ± standard errors (*n*
654 = 3) are shown. NB: the y-scale for the number of T-RFs from All RE is different than the
655 other plots.

656

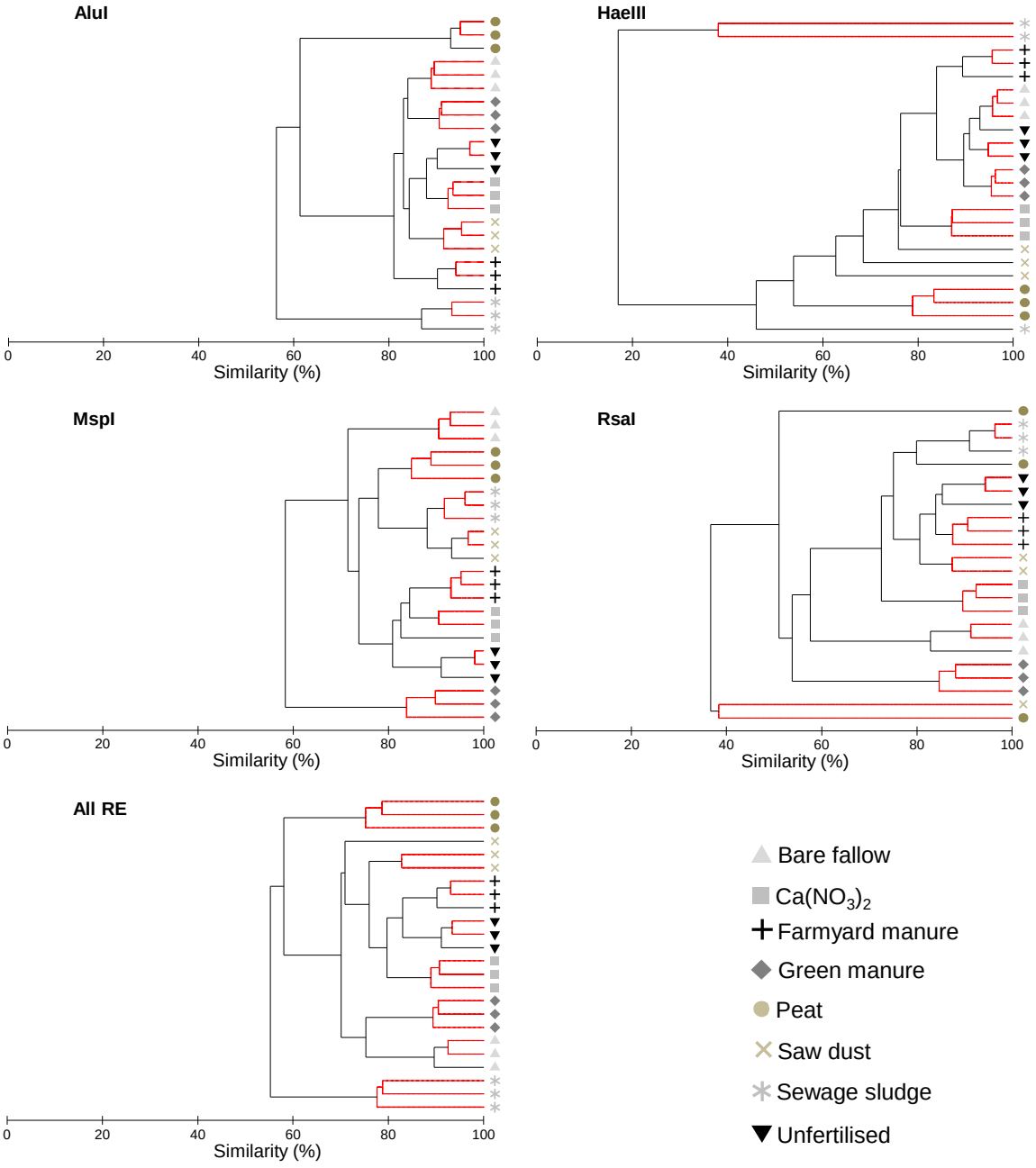
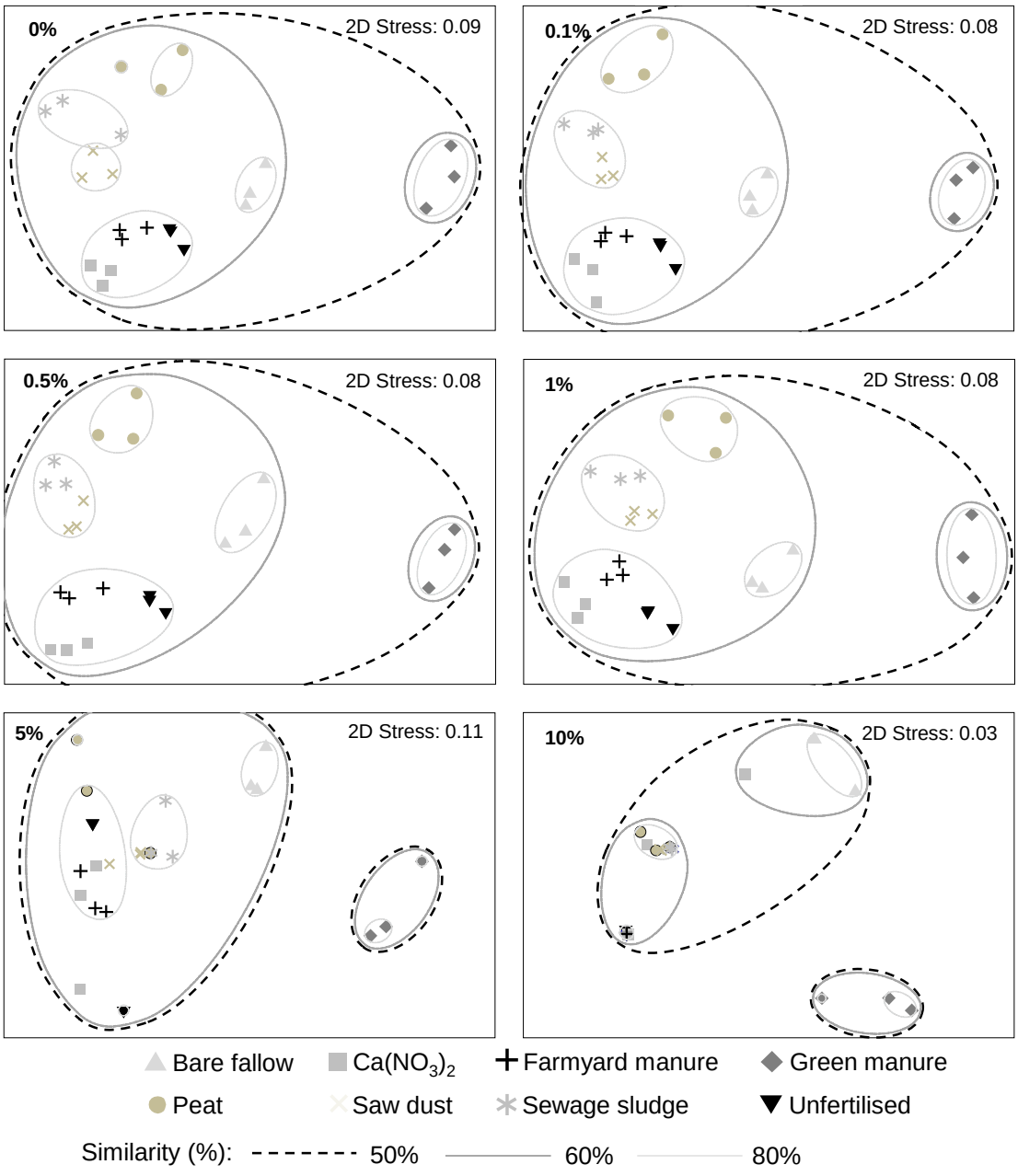


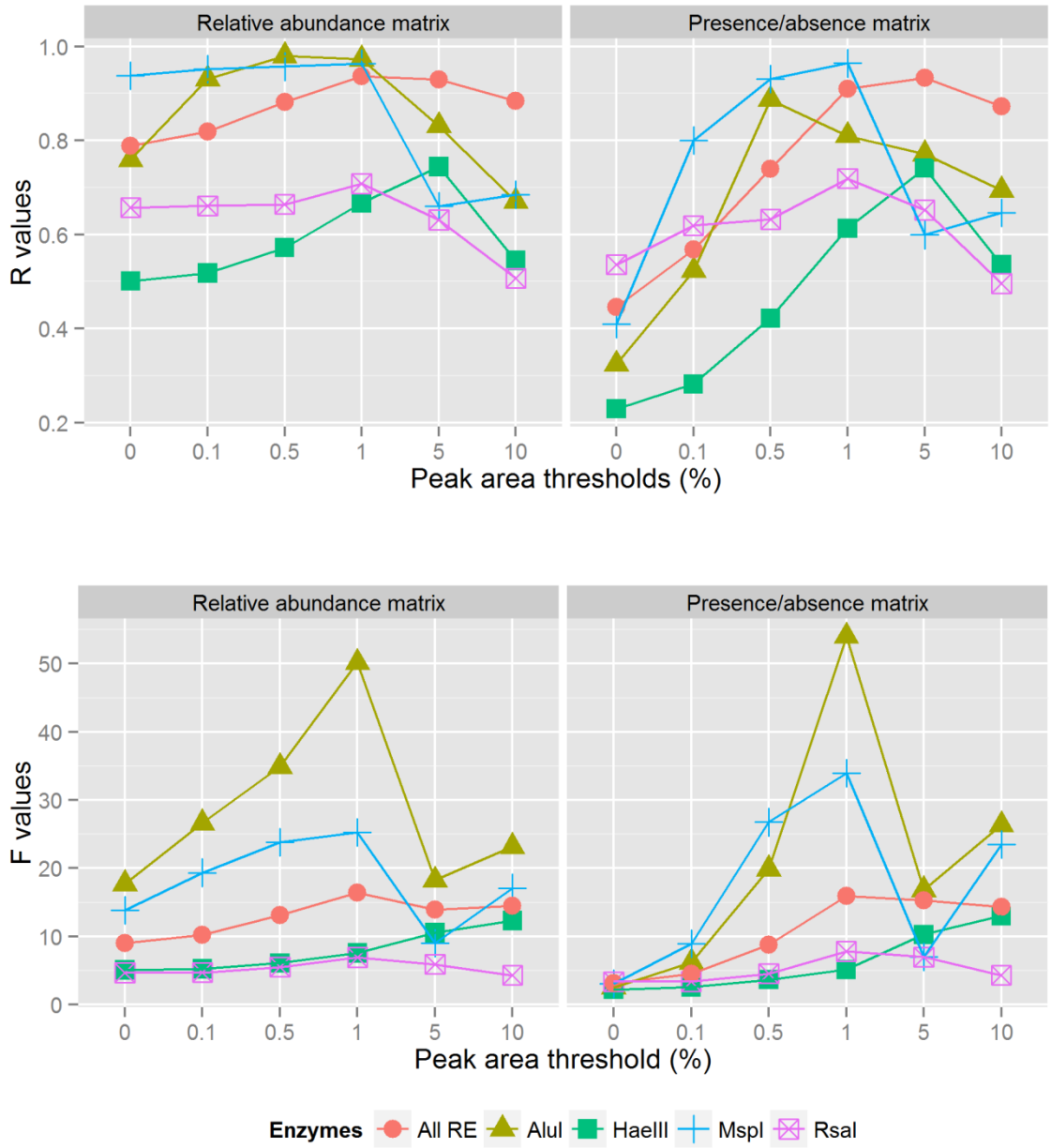
Fig. 2. Cluster analysis of the bacterial community structure of eight soil treatments studied at 0.5% peak area threshold generated from relative abundance matrix, from four different enzymatic digestions (*AluI*, *HaeIII*, *MspI* and *RsaI*) and the combination of the four restriction enzymes results (All RE). Different soil treatments are indicated in the key (3 experimental replicates). Red lines indicate clusters that are not significantly different ($P < 0.05$).



665

666 **Fig. 3.** nMDS plots of the bacterial community structure generated from the relative
667 abundance matrix, obtained by T-RFLP and digested with the restriction enzyme *MspI*, from
668 eight different soil treatments studied, at 0%, 0.1% 0.5%, 1%, 5% and 10% peak area
669 thresholds. Different soil treatments are indicated in the key (3 experimental replicates). The
670 2D stress is given for each nMDS plot. Circles indicate percentage of similarity between
671 samples based on cluster analysis.

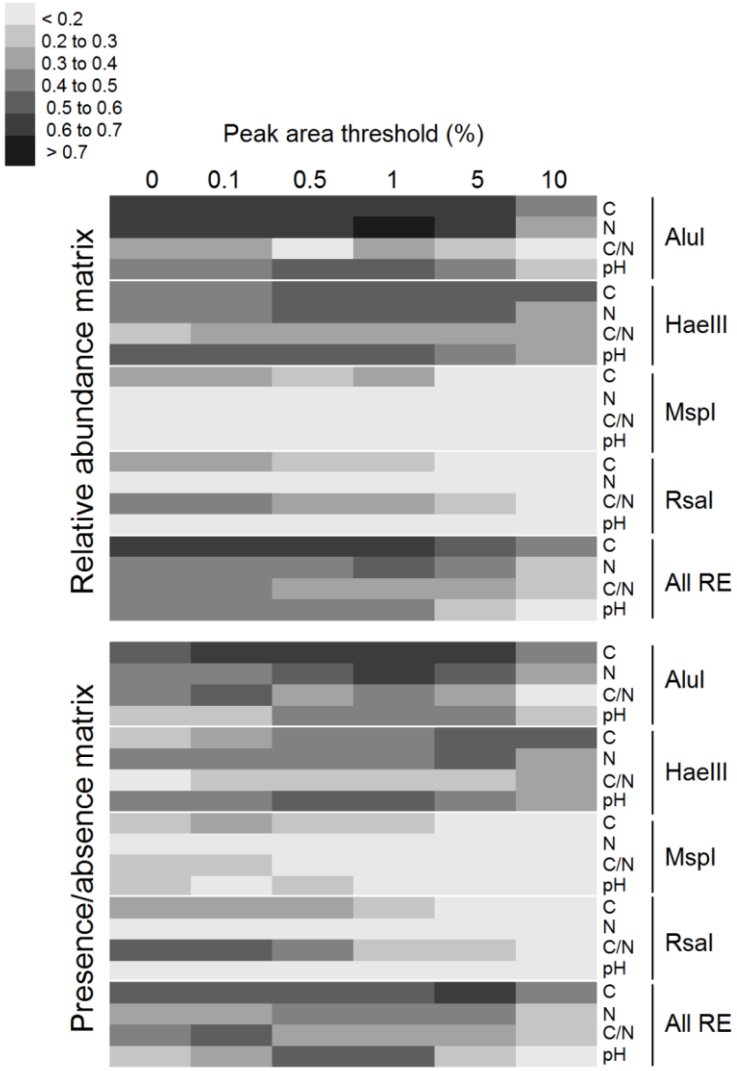
672



673

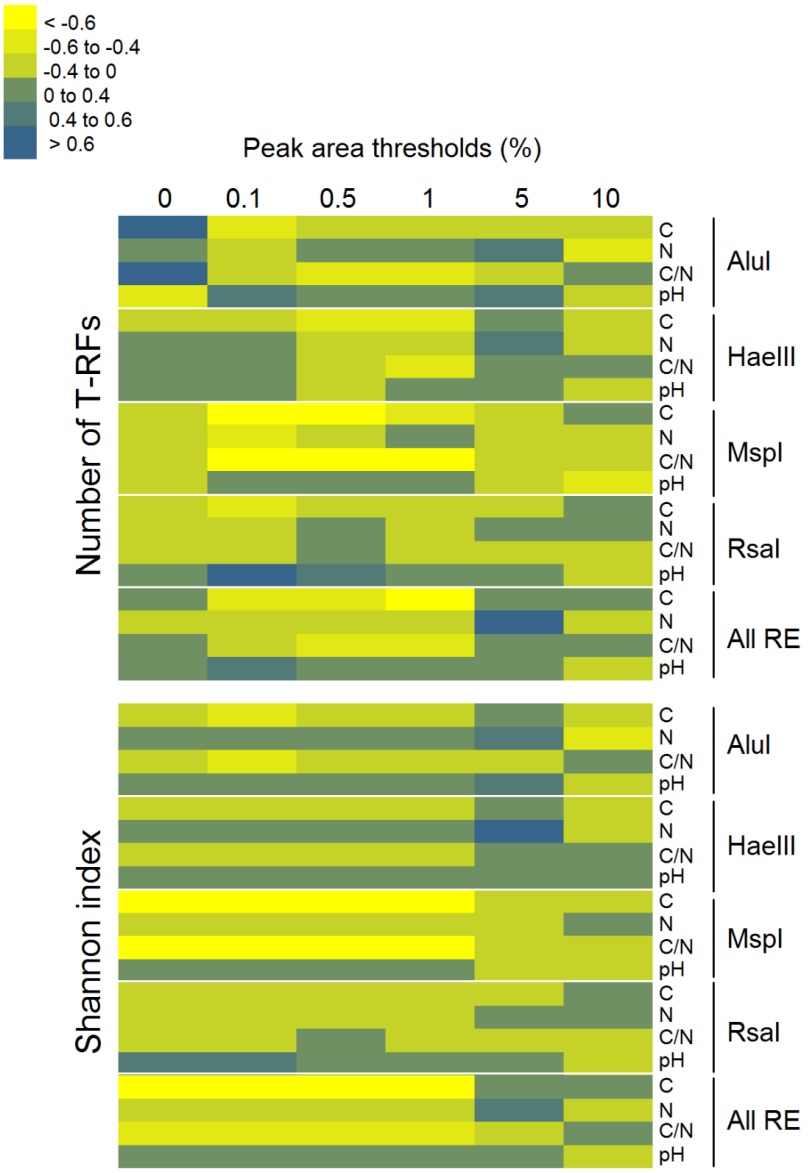
674 **Fig. 4.** Variation in the effect of soil organic and mineral amendments on bacterial community
675 structure generated from four different enzymatic digestions (*AluI*, *HaeIII*, *MspI* and *RsaI*)
676 and the combination of the four restriction enzymes results (All RE), 6 different peak area
677 thresholds (0 to 10%) and from relative abundance or presence/absence matrices. The effect
678 of soil organic and mineral amendments on bacterial community structure was expressed as R
679 (top plots) and F (bottom plots) values obtained from one-way ANOSIM and PERMANOVA,
680 respectively. All the analysis were significant at $P = 0.00001$.

681



682

683 **Fig. 5.** Heatmaps of Spearman rank correlations between each environmental variable (i.e. C
684 and N content, C/N and pH) and T-RFLP profiles from 8 different samples, generated with 4
685 different enzymatic digestions (AluI, *HaeIII*, *MspI* and *RsaI*), 6 different peak area thresholds
686 (0, 0.1, 0.5, 1, 5 and 10%) and from relative abundance or presence/absence matrices. Colours
687 represent the ρ values of Spearman rank correlations, i.e. the strength of the correlations
688 varying between 0 and +1. All RE corresponds to the combination of the 4 restriction
689 enzymes. Variations in correlation between bacterial community structure and environmental
690 variables were expressed as ρ values generated from the RELATE test from the software
691 PRIMER. Significant ($P < 0.05$) correlations were found for Spearman's rank correlation
692 superior ~0.2.



693

694 **Fig. 6.** Heatmaps of Spearman rank correlations between each environmental variable (i.e. C
695 and N content, C/N and pH) and richness or evenness of the T-RFLP profiles (See Fig. 1)
696 from 8 different samples and generated with 4 different enzymatic digestions (*AluI*, *HaeIII*,
697 *MspI* and *RsaI*) and 6 different peak area thresholds (0, 0.1, 0.5, 1, 5 and 10%). Colours
698 represent the ρ values of Spearman rank correlations, i.e. the strength of the correlations
699 varying between -1 and +1. Significant ($P < 0.05$) correlations were found for Spearman's
700 rank correlation > 0.4 and < -0.4.

701

Supplementary Information

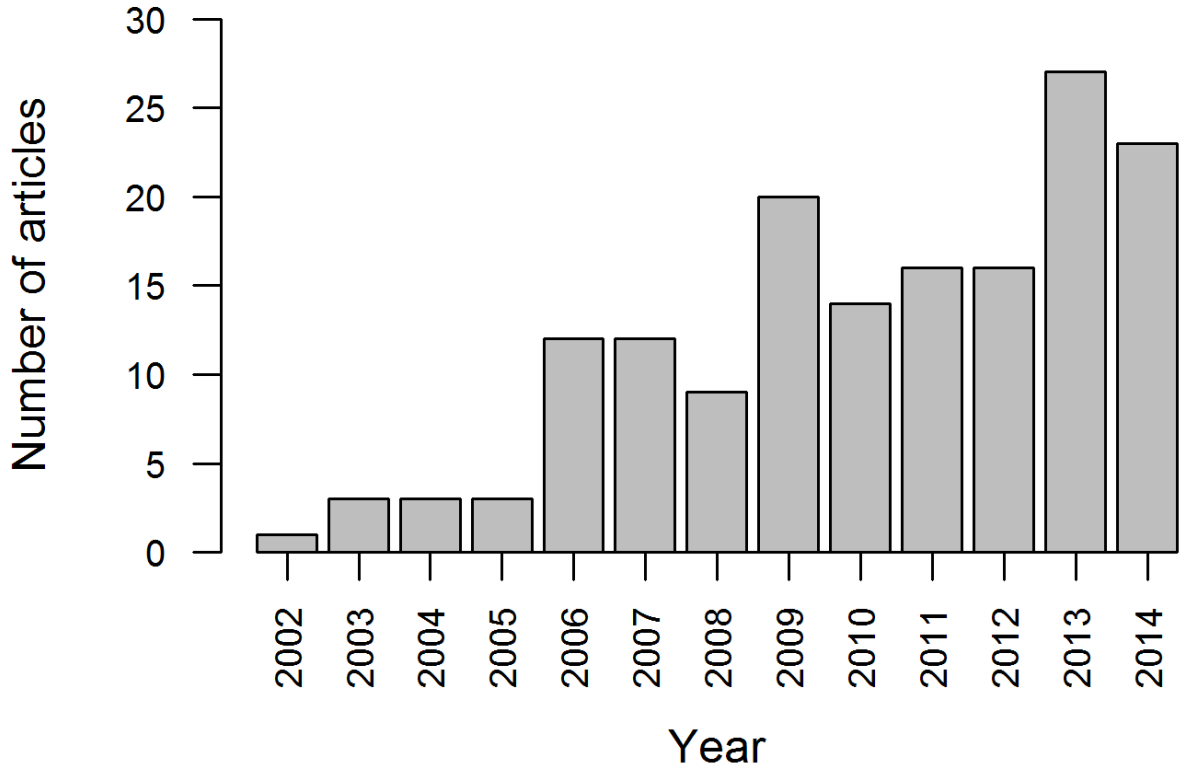
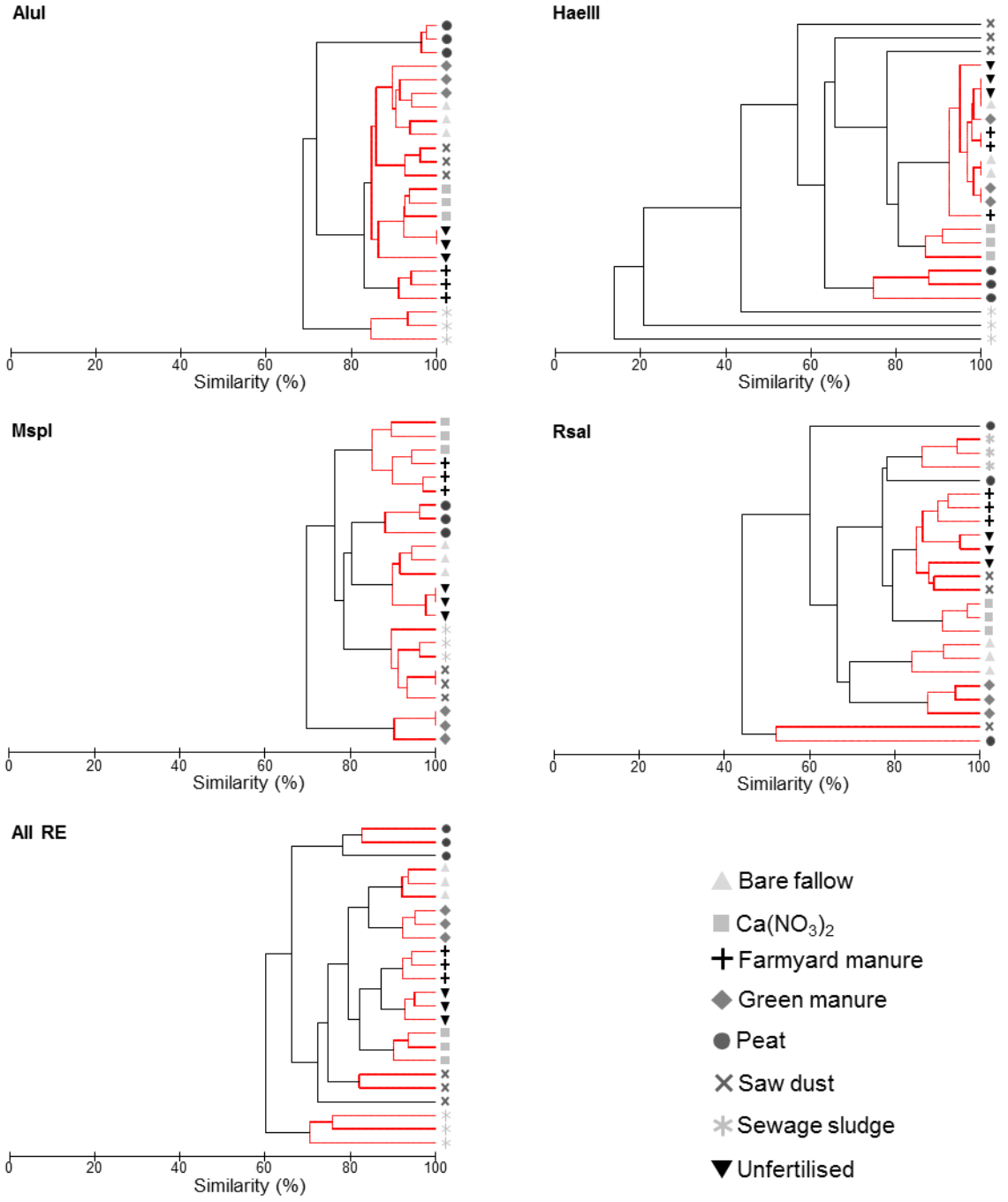


Fig. S1. Number of articles using T-RFLP and published in Biology and Fertility of Soils (17), European Journal of Soil Sciences (4), Geoderma (2), Plant and Soil (24) and Soil Biology & Biochemistry (112), between 2002 (first article published) and 2014 (See also Table S1).

716



717

718 **Fig. S2.** Cluster analysis of the bacterial community structure of eight soil treatments studied
719 at 0.5% peak area threshold generated from presence/absence matrix, from four different
720 enzymatic digestions (AluI, HaeIII, MspI and RsaI) and the combination of the four
721 restriction enzymes results (All RE). Different soils treatments are indicated in the key (3

experimental replicates). Red lines indicate clusters that are not significantly different (P < 0.05).

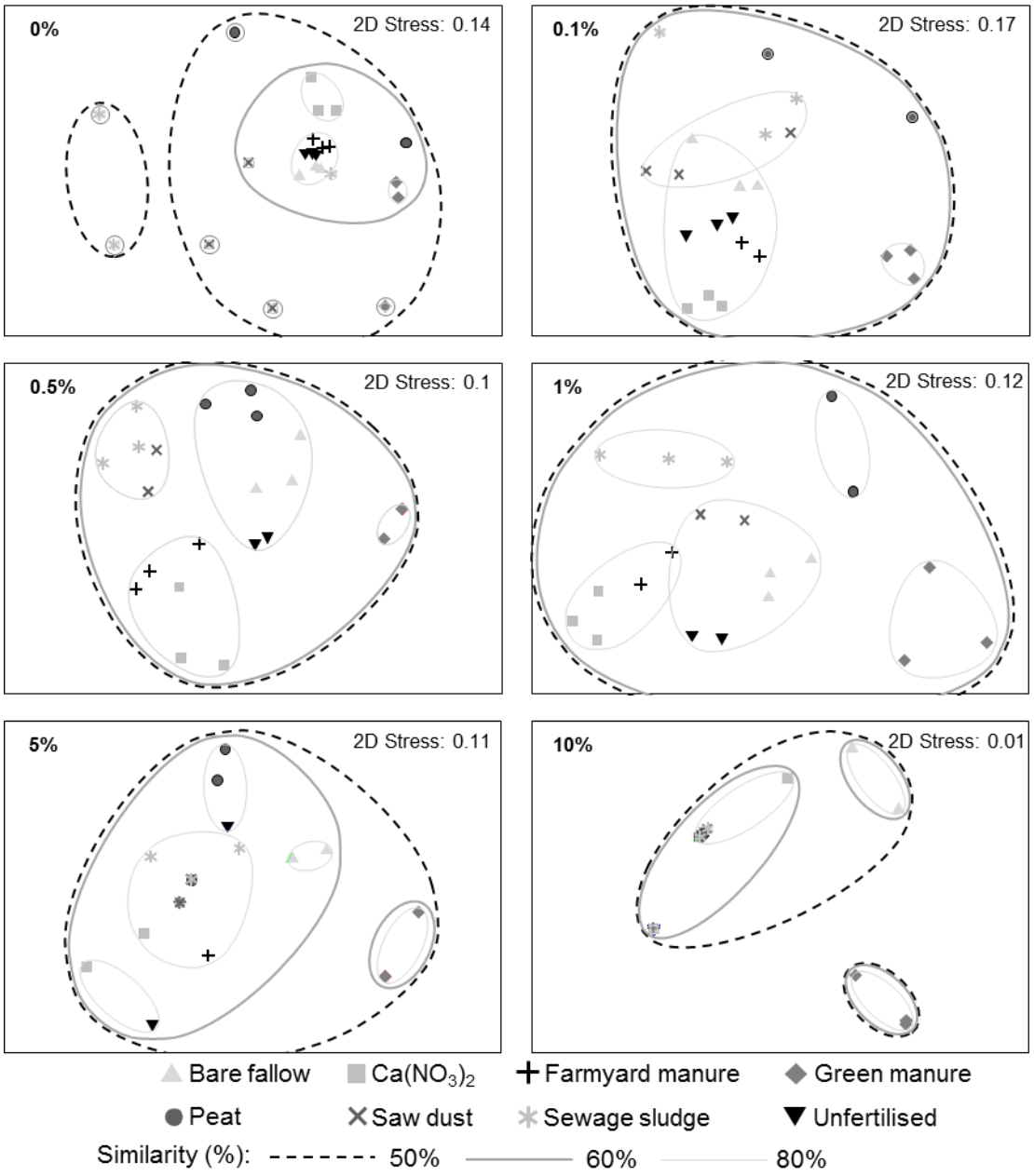


Fig. S3. nMDS plots of the bacterial community structure generated from presence/absence matrix, obtained by T-RFLP and digested with the restriction enzyme MspI, from eight soils treatments studied, at 0%, 0.1% 0.5%, 1%, 5% and 10% peak area thresholds. Different soils treatments are indicated in the key (3 experimental replicates). The 2D stress is given for each

nMDS plot. Circles indicate percentage of similarity between samples based on cluster analysis.

Table S1. Analysis of 159 studies published in *Biology and Fertility of Soils* (17), *European Journal of Soil Sciences* (4), *Geoderma* (2), *Plant and Soil* (24) and *Soil Biology & Biochemistry* (112), between 2002 (first article published) and 2014, in which T-RFLP was used. The data present the percentage of articles that indicate: 1 or >1 restriction enzymes used, mentioned the size interval of T-RFs analysed, used noise threshold (baseline or peak), normalised data, the type of matrix used for data analysis and the main ordination/statistical analysis performed.

Criteria for T-RFLP analysis	Proportion (*n* = 159)
Only one enzyme[a,b]	62%
More than one enzyme[b]	43%
T-RFs size known[c]	46%
Baseline threshold	38%
Peak area/height threshold[d]	29%
Presence/Absence matrix[e]	29%
Relative abundance matrix[e]	61%

Main ordination/statistical analysis	Proportion (*n* = 159)
Cluster	18%
nMDS	25%
PCA	25%
ANOSIM	13%
MRPP	6%
PERMANOVA	11%

[a] For 81 studies amplifying the bacterial 16S rRNA gene, the restriction enzymes the most frequently used were HhaI (37%), MspI (36%), HaeIII (30%), RsaI (12%) and AluI (10%).

[b] The total percentage is > 100% because some studies using one or more enzymes for different communities within the same study, and were subsequently counted twice.

[c] article that used or indicated the interval-size (base pair or nucleotide) of T-RFs included in their analysis.

[d] 1% peak threshold was the most used (16%) and then 0.5% (8%), 0.1% (3%), followed by 1.5%, 2%, 3% or 5% representing 1% of the studies.

[e] 9% of the studies did not indicate which matrix they used. The total percentage of studies is > 100% because some studies used presence / absence and relative abundance in their analysis and were counted twice.

Table S2. Number of T-RFs obtained by *in silico* digestion or T-RFLP. The *in silico* digestion was performed on 51 sequences previously published from the Ultuna experiment (Sessitsch et al., 2001). The number of unique T-RFs and the total number of T-RFs for all the sequences or samples are given at peak area threshold 0%.

Restriction enzyme	*In silico* digestion		T-RFLP	
	Unique T-RFs	Total number of T-RFs	Unique T-RFs	Total number of T-RFs
AluI	33	51	136	1049
HaeIII	34	49	162	872
MspI	27	41	137	720
RsaI	24	33	138	1368