

The abundance of nitrogen genes cycle and potential greenhouse gas fluxes depend on land use type and little on soil aggregate size

Aimeric Blaud^{a, 1*}, Bas van der Zaan^b, Manoj Menon^a, Georg J. Lair^{c, d}, Dayi Zhang^{a, 2}, Petra Huber^c, Jasmin Schiefer^c, Winfried E.H. Blum^c, Barbara Kitzler^e, Wei E.Huang^{a, 3}, Pauline van Gaans^b, and Steve Banwart^a

^a Department of Civil and Structural Engineering, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, United Kingdom.

^b Deltares, Subsurface and Groundwater Systems, Princetonlaan 6-8, 3508 Al Utrecht, the Netherlands.

^c University of Natural Resources and Life Sciences (BOKU), Institute of Soil Research, Vienna, Peter-Jordan-Str. 82, 1190 Vienna, Austria.

^d University of Innsbruck, Institute of Ecology, Sternwartestr. 15, 6020 Innsbruck, Austria.

^e Department of Forest Ecology and Soil, Soil Ecology, Federal Research Centre for Forests, Seckendorff-Gudent-Weg 8, 1131 Vienna Austria.

*Corresponding Author.

E-mail address: aimeric.blaud@gmail.com

¹ Current address: Sustainable Agriculture Science Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.

² Current address: Lancaster Environment Centre, Lancaster University, Lancaster LA1 2YQ, UK.

³ Current address: Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

Keywords: Quantitative-PCR; nitrogen-fixation; nitrification; denitrification; soil aggregates; land use

Abstract

Soil structure is known to influence microbial communities in soil and soil aggregates are the fundamental ecological unit of organisation that support soil functions. However, still little is known about the distribution of microbial communities and functions between soil aggregate size fractions in relation to land use. Thus, the objective of this study was to determine the gene abundance of microbial communities related to the nitrogen cycle and potential greenhouse gas (GHG) fluxes in six soil aggregate sizes (0-0.25, 0.25-0.5, 0.5-1.0, 1-2, 2-5, 5-10 mm) in four land use (i.e. grassland, cropland, forest, young forest). Quantitative-PCR (Q-PCR) was used to investigate the abundance of bacteria, archaea and fungi, and functional guilds involved in N-fixation (*nifH* gene), nitrification (bacterial and archaeal *amoA* genes) and denitrification (*narG*, *nirS*, and *nosZ* genes). Land use leads to significantly different abundances for all genes analysed, with the cropland site showing the lowest abundance for all genes except *amoA* bacteria and archaea. In contrast, not a single land use showed consistently the highest gene abundance for all the genes investigated. Variation in gene abundance between aggregate size classes was also found, but the patterns were gene specific and without common trends across land uses. However, aggregates within the size class of 0.5 – 1.0 mm showed high *nifH*, *amoA* bacteria, *narG*, *nirS* and *nosZ* gene abundance, for the two forest sites. The potential GHG fluxes were affected by land use but the effects were far less pronounced than for microbial gene abundance, inconsistent across land use and soil aggregates. However, little difference in GHG fluxes were found between soil aggregates sizes. From this study, land use emerges as the dominant factor that explains the distribution of N functional communities and potential GHG fluxes in soils, with less pronounced and less generalized effects of aggregate size.

1. Introduction

Soil is a complex and heterogeneous matrix made up of an intricate organisation of pores filled with water and gas, mineral particles, and organic matter influencing the microorganisms that live within. Soil aggregates are essential for soil fertility (Amézqueta, 1999; Bronick and Lal, 2005) and some of the fertile soils have been described as soils dominated by 0.25 – 10 mm soil crumbs (Shein, 2005). The vast variation in the size of aggregates, as well as their physico-chemical properties provides a huge diversity of microhabitats for microorganisms influencing carbon and nutrients dynamics within the soil. This study starts from the premise that soil aggregates are a fundamental ecological unit of organisation that support soil functions. These soil functions include biomass production, soil water retention and transmission, nutrient transformation, contaminant attenuation, C and N, P, K sequestration, and a major terrestrial pool of genetic diversity. The microbial community was found to vary with the size of soil aggregates, and to be linked to the specific environmental conditions in the different sizes of aggregates, such as soil organic matter (SOM) quality and quantity, pore sizes. Previous studies showed differences in microbial community structure, diversity and abundance/biomass between soil aggregates of different size, which was correlated to the quality of organic matter available (Blaud et al., 2012; Davinic et al., 2012), the size of the pores (Kravchenko et al., 2014) or tillage (Helgason et al., 2010).

Although the distribution of microbial communities in soil aggregates was studied, much less is known about the distribution of the microbial functional guilds among soil aggregates and how their sizes influence microbial functions. The size of soil aggregates in relation to their porosity (i.e. size and number of pores) was found to affect the GHG fluxes, with CO₂ emissions found to be higher in microaggregates (< 0.25 mm) than in macroaggregates (> 0.25 mm) in sandy loam soil (Sey et al., 2008; Mangalassery et al., 2013). Similar results were found for CH₄ in sandy loam and clay loam soil (Mangalassery et al., 2013), but the contrary was found in paddy rice soil (Ramakrishnan et al., 2000). Only a few studies have investigated specific

microbial functional guilds such as N fixation (Mendes and Bottomley, 1998; Poly et al., 2001; Chotte et al., 2002; Izquierdo and Nüsslein, 2006) and denitrifiers (Beauchamp and Seech, 1990; Lensi et al., 1995) in soil aggregates. The biomass and composition of diazotrophs varies with the size of soil aggregates which was correlated with total C and N, and soil texture (Poly et al., 2001; Izquierdo and Nüsslein, 2006). Aggregates within size classes 600 – 2000 μm and < 75 μm (from tundra, pasture and forest) were found to have the highest diazotroph richness (Izquierdo and Nüsslein, 2006) and microaggregates (< 250 μm) to host between 30% and 90% of the diazotrophic population (Mendes and Bottomley, 1998; Chotte et al., 2002). In contrast, denitrifiers were found to occur mainly in microaggregates (< 250 μm), where nearly 90% of the potential denitrification activity can occur (Lensi et al., 1995). Hence, the diazotroph and denitrifiers communities seem to exploit specific and different anaerobic niches within different soil aggregate size classes, although the drivers of these communities in different soil aggregate sizes remains unclear.

The type of land use and management directly influences the physico-chemical properties of soil aggregates as well as the distribution of microbial communities, their functions and resulting nutrient transformations and GHG fluxes. For example, the soil aggregates turnover rate is increased by soil tillage (Six et al., 2004), which decreases the C storage within the aggregates (Bossuyt et al., 2002), but can also decrease N_2O fluxes (Ball, 2013). Furthermore, the type of vegetation and input of organic manure influence the aggregate size distribution and the contents of organic C and N within soil aggregates (Pinheiro et al., 2004; Six et al., 2004; An et al., 2010). Subsequently, bacterial and fungal community composition was found to differ between land use types (Lauber et al., 2008) and also microbial activity such as nitrification (Hayden et al., 2010).

The above leads to the overarching hypothesis that in conjunction with land use, different microbial functions are preferentially hosted or fostered by specific size of aggregates. The specific objectives of the current study were: i) to assess the difference in microbial genes abundance between different soil aggregate size classes and bulk soil from different land uses, ii)

to assess the difference in greenhouse gases fluxes between soil aggregate sizes classes and bulk soil from different land uses, iii) to identify possible relationships between microbial gene abundances, potential GHG fluxes and the physico-chemical characteristics of the soil aggregates.

2. Material and methods

2.1 Study area

The study area is originated from the Critical Zone Observatory Marchfeld/Fuchsenbigl area (Banwart, 2011) located east of Vienna, Austria, in the National Park “Donau-Auen” on a floodplain of the Danube River (Fig. S1). The mean annual temperature in the area is ~9 °C and mean annual precipitation ~550 mm. The study sites are located along a chronosequence starting from a young river island (created <70 years; average inundation frequency: 10 d yr⁻¹) named “young forest”, and sites disconnected from the river through a flood control dike: forest, grassland and cropland. The young forest and forest sites are presently covered by the plant communities *Salicetum albae* and *Fraxino-Ulmetum* (Schubert et al., 2001), respectively. The grassland site was converted from forest to grassland (presently *Onobrychido viciifoliae-Brometum*) between 1809 and 1859 and is currently cut twice a year. The cropland site was grassland before 1781 and was converted to intensive cropland in the first half of the 20th century. According to Lair et al. (2009), the topsoil (0-10 cm) of the young forest was deposited after 1986, whereas a topsoil age of approx. 250-350 years on the forest, grassland, and cropland site can be estimated (Lair et al., 2009). The soils are classified as Epigleyic Fluvisol (young forest) and Mollic Fluvisols (forest, grassland and cropland; IUSS Working Group WRB, 2014).

2.2 Soil sampling and fractionation

The soil sampling was identical at all sites and was performed in September 2011 under dry soil moisture conditions (capillary potential pF 3.8 - 4.0). At each site, three sampling spots

(70 x 70 cm) were randomly selected within a circle of about 30 m radius. The soil layer from 5 - 10 cm soil depth was sampled below the main rooting zone. The soil samples were manually dry sieved to obtain 6 soil aggregates size classes: < 0.25, 0.25 - 0.5, 0.5 - 1, 1 - 2, 2 - 5, and 5 - 10 mm. The soil fraction > 10 mm was not included in the study as it was composed of a wide range of aggregates and large clumps (100 – 500 g per clump). During dry sieving, visible roots were removed. Sieving continued with freshly excavated soil until ~200 g of soil aggregates was obtained for each aggregate size class. Additional bulk soil samples were collected at each site and sampling spot. Soil aggregate size fractions and bulk soil samples were stored at 4 °C and samples for DNA extraction at -20°C before subsequent analysis. Dry-sieving was chosen over wet-sieving to avoid any bias due to dry/wet cycles with wet-sieving that could have direct effect on GHG emissions (Kaiser et al., 2015). Despite that the sieving method affect the gene abundance quantification, dry-sieving can reveal differences in gene abundance between soil aggregates sizes (Blaud et al., 2017).

2.3. DNA extraction and quantitative-PCR

Total nucleic acids were extracted from 0.20 to 0.55 g of fresh soil aggregates from all size classes and from bulk soil samples with PowerSoil® DNA Isolation Kit (Mo-Bio laboratories, Carlsbad, CA, USA) according to manufacturer's instruction, except for the final step where the nucleic acids were eluted in 100 µl of sterile nuclease free water instead of solution C6. Microbial abundance was investigated by Quantitative-PCR (Q-PCR) targeting specific genes or genetic regions. Bacterial and archaeal communities were targeted via the 16S rRNA genes, while the fungal community abundance was investigated by targeting the ITS region. The different communities involved in most steps of the N-cycle were investigated: the nitrogen fixing microorganisms were quantified based on the *nifH* gene; nitrification was investigated by targeting the ammonia oxidising bacteria (AOB) and archaea (AOA) via the *amoA* gene, and denitrifiers were targeted via the *narG* gene coding for the nitrate reductase, the *nirS* gene

coding for the nitrite reductase and the *nosZ* gene coding for the nitrous oxide reductase (Table S1).

Q-PCR standards for each molecular target were obtained using a 10-fold serial dilution of plasmids carrying a single cloned target gene or relevant part thereof. Standard curve template DNA and the “no template control” (NTC) were amplified in duplicate in the same plate as the environmental samples. Q-PCR amplifications were performed in 25 µl volumes containing 12.5 µl of iQ™ SYBR® Green Supermix (Bio-Rad, Hemel Hempstead, UK), 8.5 µl of nuclease-free water (Ambion, Warrington, UK), 1.25 µl of each primer (10 µM) and 1 µl of template DNA using a CFX96™ Real-Time System (Bio-Rad, Hemel Hempstead, UK). Amplification for all Q-PCR assays are given in the supplementary material and Table S1. The efficiency of the Q-PCR assays was above 90%, except for fungi and AOA (~70%). The r^2 were > 0.99, except for *nifH* and *nosZ* genes (~0.97).

2.4. Microbial respiration

Greenhouse gas fluxes from the aggregate size fractions and the bulk soil were measured from field moist bulk soil and soil aggregates (pF 3.8 -4.0; hereafter named “field moisture”) and from moistened samples (40 – 60 % of field capacity) by adding distilled water 48 hours before flux measurements started (hereafter named “elevated moisture”). Soil temperature was set to 20 °C. The soil moisture was increased because at the time of soil sampling the soil moisture content was low (pF 3.8-4.0), potentially reducing microbial activity and subsequent GHG fluxes. For full details on the GHG measurements, refer to the supplementary material.

Fluxes of CO₂ and NO were measured with a fully automated laboratory measuring system as described in detail by Schindlbacher et al., (2004) and Schaufler et al., (2010). Carbon dioxide was measured with a PP Systems WMA-2 (Amesbury, MA, USA), infrared CO₂ analyser, and NO was measured with a HORIBA APNA-360 (Kyoto, Japan) chemoluminescence NO_x analyser. Determination of N₂O and CH₄ fluxes was done manually by closed chamber technique.

The analysis was done immediately after gas sampling by gas chromatography (AGILENT 6890N) connected to an automated system sample-injection (AGILENT TECH G1888, Network HEADSPACE-SAMPLER) at an oven temperature of 40 °C. Nitrous oxide was measured by a ⁶³Ni-electron-capture detector and CH₄ by a flame ionization detector.

2.5. Physico-chemical analysis of bulk soil and aggregates

The soil moisture content, organic C, total N, N-NO₃⁻, N-NH₄⁺, P-PO₄³⁻, and carbonate concentration, C/N, and soil texture (i.e. sand, silt and clay contents) were measured for each aggregates size and bulk soil. Three different fractions of soil organic matter (SOM) were determined by simultaneous thermal analysis (STA) according to Barros et al. (2007): labile SOM stable SOM and refractory SOM. Particle size distribution in the various aggregate size classes as well as the SOM fractions (STA) were measured on one composite sample for each site (i.e. mixture of the 3 replicates at each site). For full details of the methods used, refer to the supplementary material.

2.6 Statistical analysis

To test the effects of land use and soil aggregate size on microbial gene abundance, GHG fluxes and soil aggregate characteristics, analyses of variance (ANOVA) were performed with land use and soil aggregate size as factors (3 and 6 degrees of freedom (df) respectively). The normality of the model residuals and the homoscedasticity of the variances were checked before statistical analysis. When, one or both of these conditions were not met the data were log transformed to reach the conditions. However, if log transformation did not lead to normality or homoscedasticity or could not be applied (presence of negative values for GHG), one-way ANOVA was performed to test the effect of land use within each aggregate size class separately. Similarly, to test the effect of soil moisture level on GHG fluxes for each land use, two-way ANOVA was applied with soil aggregates size and soil moisture level as main factors. To test the effect of aggregate size within each land use on microbial gene abundance, GHG fluxes and soil

aggregate characteristics, one-way ANOVA was performed with aggregates size as factor ($df = 6$) for each land use separately, insuring conditions are met as previously. When significant ($P < 0.05$) effects were found for ANOVA, the Tukey HSD (honest significant difference) test was used to reveal the significance of the differences between class pairs.

In order to get insight into the potential drivers of microbial gene abundances and GHG fluxes, Spearman's rank correlation coefficients ρ ($-1 \leq \rho \leq 1$) were calculated between microbial gene abundance, GHG and soil characteristics, across all the land uses to reveal the factors explaining the differences due to land use, or for each land use to reveal the factors explaining the differences due to soil aggregates size classes. To display the correlations, heatmaps were constructed using the library "gplots" from R software, where colours represent the direction and strength of the correlation.

All statistical analyses were performed using R v3.2.1 (R Development Core Team, 2015) and a significance level of $P < 0.05$ was used throughout.

3. Results

3.1 Variation in soil aggregates characteristics

The physico-chemical parameters of soil aggregates significantly differed between land use, and between aggregates size classes. The soil aggregate mass distribution showed the same pattern for all the land uses, with the size class 2.0 – 5.0 mm being the most abundant (20 – 40 w/w %), and size classes < 0.25 mm the lowest ($< 10\%$; Fig. S2). Young forest and forest showed significantly higher soil water content for most soil aggregates sizes in comparison to cropland and grassland (Fig. S2). The cropland soil had the lowest soil organic C (SOC) and total N concentrations (~ 25 and ~ 1.5 g kg⁻¹ soil, respectively), whereas the grassland soil showed the highest concentrations (~ 50 and ~ 3 g kg⁻¹ soil, respectively; Fig. S3). Grassland showed significantly lower N-NO₃⁻ concentration for soil aggregates > 0.5 mm (~ 10 times) than the other sites, but significantly higher N-NH₄⁺ for the bulk soil (~ 5 times) and some soil aggregates (Fig. S4). The P-PO₃⁻ in cropland was significantly lower than the other sites in aggregates 1 – 2

mm, while in young forest P-PO₃³⁻ was significantly higher for 0.5 – 1 mm in comparison to grassland and cropland.

Significant differences in physico-chemical parameters between aggregates size classes were found, mainly at the young forest and forest site, and between the classes < 0.5 mm and the other classes. The aggregates size classes < 0.5 mm at the young forest and forest sites had significantly lower SOC concentrations than bulk soil and most larger size classes, while their C/N was higher (Fig. S3). Similarly, the water content of < 0.25 mm was significantly lower than most aggregates sizes at young forest, forest and grassland sites. In contrast, soil aggregates < 0.5 mm at grassland showed significantly higher N-NO₃⁻ concentrations than other soil aggregates sizes or bulk soil (Fig. S4).

3.2. Variation in microbial gene abundance in bulk soil and soil aggregate size classes

All microbial gene abundances investigated showed significant differences between land use types for at least one soil aggregate size classes or bulk soil (Fig. 1, Fig. S7-S9, Table S2). The cropland site consistently showed lower abundance of bacterial 16S rRNA, *nifH*, *narG*, *nirS* and *nosZ* genes, while *amoA* bacteria (AOB) was lower in grassland (Fig. S8) and *amoA* archaea (AOA) in young forest (Fig. 1, S8). In contrast, the forest site tends to harbour the highest abundance for the different aggregate sizes of bacterial and archaeal 16S rRNA, AOB and AOA genes (Fig. S7, S8), while the *nifH*, *narG* and *nirS* genes showed the highest abundance in young forest site (Fig. 1, S8, S9), and *nosZ* gene in grassland site (Fig. 1, S9).

Significant effects of aggregate size within individual land uses were found (one-way ANOVA and Tukey HSD) for all microbial amplicon abundances investigated, except archaeal 16S rRNA, fungal ITS, and AOA (Fig. S7-9). However, significant pairwise differences were only found for the young forest (for bacterial 16S rRNA, *nifH*, and *narG* genes) and forest sites (for AOB, *narG*, *nirS* and *nosZ* genes). Trends at the young forest site were similar, where genes abundances were overall found relatively high in 0.5 -1.0 mm aggregates and relatively low in 2.0-5.0 mm and < 0.25 mm aggregates (Fig. 2). For the forest site a similar trend is also found,

the abundances being higher in the 0.25 – 0.5 and 0.5 – 1.0 mm aggregates than in the other aggregate size fractions (Fig. 2).

3.3. Changes in potential greenhouse gas fluxes

The types of land use and moisture levels were the main factors differentiating GHG fluxes, although differences between land uses were not as strong as for microbial abundances and consistent across land uses. Greenhouse gas fluxes were significantly different between land use types at both moisture levels for at least one soil aggregates size, except for NO at field moisture (Fig. S10, S11). The CO₂ emissions were significantly different (Tukey HSD) only for 0.5 – 1 mm and bulk soil between cropland and forest site, and also between grassland with cropland and young forest sites for the bulk soil (Fig. 3, S10). At elevated moisture, CO₂ emissions were consistently significantly lower in cropland compared to grassland sites regardless of the aggregates size classes and bulk soil (Fig. 3, S10, S11). Overall, the CO₂ emissions were significantly different between soil moisture levels, and mainly higher at the elevated moisture content than at field moisture content (Fig. S10). The other GHG fluxes showed large standard deviation (Fig. 2, 3) and overall significant differences between land use types for a few specific aggregate size classes such as < 0.25 (CH₄ elevated moisture), 0.25 – 0.5 (NO, N₂O soil moisture), 1.0 – 2.0 (CH₄ both moisture levels and N₂O field moisture), 5.0 – 10.0 mm (CH₄ and N₂O elevated moisture) (Fig. S10, S11).

Within the separate land use types, significant effects of aggregate size at field moisture were only observed for CH₄ at the forest site and for NO at the grassland site. The 0.5 – 1.0 mm aggregates acted as a sink for CH₄ at field moisture while the other aggregates classes were sources of CH₄ (Fig. 4). The aggregate size classes < 0.5 mm from grassland were found to be sources of NO, while larger size classes were sinks at field moisture (Fig. 4). At elevated moisture, the bulk soil showed significantly lower CO₂ emissions than the aggregates size classes, while it was a source of CH₄ and aggregates size classes (except 2.0 – 5.0 mm) were sinks (Fig. 4).

3.4. Relationship between microbial gene abundance, potential greenhouse gases and soil characteristics

When the correlations were performed on all the land uses, bacteria, fungi and *nosZ* gene abundances showed similar and significant positive correlations with the following soil characteristics: labile SOM, stable SOM, refractory SOM, SOC, total N, and silt for all land uses combined (Fig. 5a; Fig S3, S5, S6). The *narG*, *nirS* and *nifH* gene abundances showed significant positive correlations with silt and carbonate contents and P-PO₄³⁻ concentrations (Fig. S2, S4-S5). In contrast, AOB, AOA and archaea gene abundances showed negative correlations with silt and carbonate contents, but positive correlations with soil water content, N-NO₃⁻ concentration and sand content (Fig. 5a). The CO₂ emissions at elevated moisture for the combined land uses were strongly and positively correlated ($\rho > 0.5$) with the three SOM pools, total N, SOC, carbonate and silt, but negatively with sand content ($\rho = -0.74$; Fig. 5b). The CO₂ and CH₄ fluxes at field moisture showed significant and positive correlations with the three SOM pools, total N and SOC. The other GHG fluxes showed significant correlations with only a few specific variables (Fig. 5b). Most gene abundances were significantly and positively correlated to CO₂ emissions at elevated moisture, except AOB, archaea and AOA genes which were negatively correlated (see supplementary and Fig. S12 for details).

The heatmaps for the separate land uses did not reveal similar patterns across land use types but unique to each land use, even for young forest and forest sites where significant differences in gene abundances between soil aggregates sizes were found (Fig. 6, S13, S14). Hence, at the young forest site, the N contents and to a lesser extent SOM contents (especially the labile SOM pool) were positively correlated to bacteria, *nifH*, AOB, *narG* and *nirS* genes (Fig. 6). At the forest site, different parameters explained the differences in genes abundance between soil aggregates sizes; soil texture explained the distribution of several gene abundances, with clay content positively correlated with *nifH*, bacteria, *narG* and AOB genes and sand with fungi, while sand content was negatively correlated with *nosZ*, and *nirS* genes.

The correlations between GHG fluxes and soil properties showed no similar patterns across land uses and relatively low number of correlations (Fig. S13). At the grassland site, where most differences in GHG fluxes between soil aggregate sizes were found, the CH₄ fluxes at field moisture were positively correlated to labile, stable and refractory SOM content, but negatively correlated to these SOM fractions at elevated moisture (Fig. S13). The correlations between gene abundances and GHG fluxes for each land use are presented in supplementary material (Fig. S14)

4 Discussion

4.1 Land use is a dominant explaining factor for microbial gene abundance in soil

The type of land use was the main factor of the microbial abundance and the nitrogen cycling community in soils studied. Regardless the gene investigated, the gene abundances were always affected by the different type of land uses. The different type of land uses and management was previously found to affect the abundance of microorganisms (Enwall et al., 2010; Hallin et al., 2009; Leininger et al., 2006; Ma et al., 2008; Morales et al., 2010, 2010; Wallenstein and Vitgalys, 2005). This study present a comprehensive evaluation of the distribution of N cycling genes across land uses with similar parent material (fluvial sediments) and climate (co-located sites).

Cropping clearly had a negative effect on the abundance of microorganisms in soil and most of their N functions. The SOC and total N concentrations explained the distribution of bacteria, fungi and *nosZ* gene, highlighting that the depletion of SOC and total N in cropland (Fig. S3) due to soil management (e.g. tillage), soil erosion and plant harvest, limit the abundance of microorganisms. Soil tillage was found to directly affect negatively the biomass of bacteria and fungi (Muruganandam et al., 2009; Helgason et al., 2010), but also the *narG* gene abundance (Chèneby et al., 2009). Hence, the negative effect of cropping on microbial communities is likely due to a combination of factors limiting microbial growth. In contrast, the AOA and AOB were

abundant in cropland likely due to application of fertiliser (containing NH_4) that maintains AOA and AOB abundance and stimulates nitrification which was supported by the significant correlations of the ammonium oxidizing microorganisms with NO_3^- concentration and soil water content. However, distinct drivers to each community were also found across land uses, such as SOC/N and sand content for AOB, and total N, thermally more stable SOM and clay contents for AOA (Fig. S3, S6). Thus, it further supports the idea that despite AOA and AOB delivering the same function, the two communities live in different microhabitats with specific environments stimulating their activity separately (Prosser and Nicol, 2008).

When focusing on the community showing the highest abundance in young forest (i.e. *nifH*, *narG* and *nirS* genes), they all showed a strong and positive correlation to PO_4^{3-} concentration which was higher at the young forest (Table 1, Fig. S3). The *nifH* gene abundance was found to be higher in forest soil than in agricultural soil (Morales et al., 2010). In contrast, for the communities with higher abundance at the forest site (i.e. bacterial and archaeal 16S rRNA genes, AOB and AOA), different variables were correlated, without a common variable explaining microbial distribution. Hence, this result highlights the complexity of the variables explaining microbial distribution in forest soil (Levy-Booth et al., 2014). The fungal ITS and *nosZ* genes showed similar factors explaining their distribution (i.e. SOC, N, SOM and NO_3^-). Fungi in soils were found to produce N_2O , which in return could be reduced into N_2 , which could explain the similar factors between fungal ITS and *nosZ* gene (Maeda et al., 2015). Furthermore, *nosZ* gene distribution showed different factors than *narG* and *nirS* genes, suggesting that the different steps of the denitrification do not simultaneously occur within the same microhabitat which is expected due to the different environment required to perform the different steps of denitrification.

4.2 Soil aggregate size is secondary explaining factor for microbial gene abundance in soil

Soil aggregates size was a secondary factor in explaining nitrogen genes abundance compared to land use. The effects of soil aggregates size classes on genes abundance was

specific to the land use type and not presents for all genes studied or land use. Neuman et al. (2013) found that the size of soil aggregates were the dominant factor of the abundance of bacterial, archaeal and fungal community over soil management (i.e. fertilisation). However, they investigated microaggregates (2 - 20 μm , 20 - 63 μm , > 63 μm) and the silt and clay fractions (< 2 μm), which could physically protect organisms against environmental changes. Hence, the current study showed that the size of macroaggregates are not the main factor of the microbial distribution and N functional guilds after land use type, whereas aggregates < 63 μm could have a greater effect on the distribution of microbial communities.

The presence or absence of differences in gene abundance between soil aggregates in different land use may be related to the balance between stability and instability of the microhabitats, hindering or promoting differentiation of specific microhabitats and associated microbial communities. The low variation in gene abundance for cropland and grassland may be related to the soil aggregates and organic matter turnover, which is expected to be high due to anthropogenic activity such as tillage and plant harvest (Six et al., 2000, 2002; Tisdall and Oades, 1982), and the absence of land management. Furthermore, forest sites were likely to show a more stable temperature and soil moisture regime throughout the year than cropland and grassland because of the tree cover. Overall, specific drivers for each land use are responsible of the distribution of gene abundance in soil aggregates.

At the forest and young forest sites, the size of soil aggregates was an important factor of several microbial community abundances and functional genes with specific sizes harbouring higher gene abundances. Furthermore, a similar pattern of distribution was found between functional genes at a specific site, suggesting that these functions coexist in similar niches. Hence, the aggregate size class 0.5 – 1.0 mm showed consistently the highest gene abundance regardless of the specific microbial functions, possibly hosting a high number of active microbial functions, and is within the range of soil aggregates that characterise fertile soils as described by Shein (2005). However, some dissimilarities were present, such as the soil aggregates size class 1.0 – 2.0 mm which showed high gene abundances at the young forest while low gene

abundances were found at the forest site. Thus, differences between similar land use, such as tree cover, and soil characteristics may also play a role in gene abundance distribution within soil aggregate size classes. Although those genes colonized preferentially similar niches, which differ with their distribution across land uses. different factors were responsible for their abundances, in the young forest and forest site.

4.3 Effects of land use and soil aggregate size on potential greenhouse gas fluxes

The potential GHG fluxes were affected by land use, soil moisture levels and soil aggregate size, but the effects were far less pronounced than for microbial gene abundance, and inconsistent across land use and soil aggregates. This was partly due to the high variability in the measure of GHG fluxes, but also reveal differences compared to the microbial gene abundance. Hence, the effect of land use on the bulk soil samples were mainly found for CO₂ emissions, while for the other GHG only specific soil aggregate sizes revealed the potential effect of land use. The different effect of land use found on GHG fluxes between soil aggregate compared to the bulk soil may be linked to different porosity present for each size and how land use affect it differentially (Rabbi et al., 2016). Thus, working on bulk soil may mask some potential GHG fluxes (Kravchenko et al., 2014). However, it should be acknowledged that each soil aggregate size was in artificial conditions for the GHG measurement (e.g. air fluxes), likely leading to different behaviour than *in situ*. The CO₂ emissions were consistently lower in cropland compared to the other sites regardless of the soil water content, indicating the potential low microbial activity in cropland due to SOM depletion also supported by the low bacterial gene abundance. The other GHG fluxes showed inconsistent effect of land use depending on soil moisture and soil aggregates size, highlighting the complexity of drivers of CH₄, NO and N₂O fluxes.

Change in soil moisture had significant effects on GHG fluxes, although it varies between GHG, land use, and soil aggregates. Higher CO₂ emissions were consistently found at elevated soil moisture than field moisture across all land use, highlighting the importance of soil

moisture for microbial activity and CO₂ emissions (Sey et al., 2008). For CH₄, NO and N₂O the effect of increased soil moisture was not as consistent as for CO₂, indicating that other factors limit their fluxes. Surprisingly, increasing soil water content in the current study did not necessarily increase the CH₄ production, as it could be expected because methanogens are more active in high water content/anaerobic soils. The CH₄ was either emitted or consumed depending on the soil water content for a specific land use and soil aggregate size class. This indicates that both methanogens and methane-oxidizing bacteria were present in the same soil aggregates as previously found by Sey et al. (2008).

Overall, the GHG fluxes did not occur in a specific aggregate size class within a land use as found for microbial gene abundances in forest sites. Previous studies found higher CO₂ emissions in microaggregates (< 250 µm) while being sinks of CH₄ (Sey et al., 2008). However, it was also shown to be highly sensitive to water filled pore space (WFPS), with no difference in CO₂ emissions between aggregate size at 60% WFPS and microaggregates being sink of CH₄ at 20% WFPS but a source at higher WFPS (Ramakrishnan et al., 2000; Sey et al., 2008). However, in the current study increased soil moisture did not reveal more significant differences in GHG fluxes between soil aggregates indicating that other factors may drive differences or that the size of soil aggregate may not be an important driver for GHG fluxes.

5. Conclusions

This study demonstrates that land use is the main factor in explaining nitrogen genes abundance and greenhouse gas fluxes while soil aggregates was a secondary factor. This goes against our initial hypothesis suggesting that different microbial functions are preferentially hosted or fostered by specific size of aggregates. This is due to the stronger difference in soil physico-chemical characteristics between land use types than between soil aggregates sizes. Cropping had a clear negative effect on the abundance of most microbial communities, likely due to the depletion of SOC and total N by tillage and plant harvest. Although soil aggregate size was not a dominant factor, it affected the distribution of the N functional communities at the

semi-natural forest sites, showing that some microbial functions are probably related to specific microhabitats (i.e. the architecture and distribution of pores filled with water and air, the availability of organic matter and other nutrients) in soil, where anthropogenic activity is limited allowing differences between microhabitats to develop. However, no specific size of soil aggregates enhanced the abundance of a specific microbial function across all four land uses. Soil aggregates size had little effect on GHG fluxes, indicating that the size of soil aggregates may not be affecting much GHG fluxes but it also highlights the difficulties to measure GHG fluxes in aggregates.

This study only addresses a single point in time, limiting our understanding of the distribution of microbial functions over soil aggregates of different size. Further studies are needed, taking into consideration the dynamics of soil aggregates and its relation with microbial communities by sampling at multiple time points. A wider range of aggregate size classes (e.g. size classes < 250 μm) and land use types should be studied, combining microbiology and soil architecture (e.g. x-ray tomography) as well as nutrient availability in local and time scale. Such approach would to fully reveal the physical distribution of microhabitats, the microbial communities and functions among soil aggregates. Comparing microbial functions between soil aggregates of varying size from a specific land use (e.g. forest) but from different locations or soil types may also provide more insight into the role of soil aggregates in microbial functioning.

Acknowledgments

This work was supported by the European Commission 7th Framework Program as a Large Integrating Project, SoilTrEC (www.soiltrec.eu), Grant Agreement No. 244118.

References

Amézketa, E., 1999. Soil aggregate stability: a review. *Journal of Sustainable Agriculture* 14, 83–151. doi:10.1300/J064v14n02_08

484 An, S., Mentler, A., Mayer, H., Blum, W.E.H., 2010. Soil aggregation, aggregate stability, organic
 485 carbon and nitrogen in different soil aggregate fractions under forest and shrub
 486 vegetation on the Loess Plateau, China. CATENA 81, 226–233.
 487 doi:10.1016/j.catena.2010.04.002

488 Ball, B.C., 2013. Soil structure and greenhouse gas emissions: a synthesis of 20 years of
 489 experimentation. European Journal of Soil Science 64, 357–373. doi:10.1111/ejss.12013

490 Banwart, S., 2011. Save our soils. Nature 474, 151–152. doi:10.1038/474151a

491 Barros, N., Salgado, J., Feijóo, S., 2007. Calorimetry and soil. Thermochimica Acta, XIVth ISBC
 492 Proceedings Special Issue Fourteenth conference of the International Society for
 493 Biological Calorimetry 458, 11–17. doi:10.1016/j.tca.2007.01.010

494 Beauchamp, E.G., Seech, A.G., 1990. Denitrification with different sizes of soil aggregates
 495 obtained from dry-sieving and from sieving with water. Biology and Fertility of Soils 10,
 496 188–193. doi:10.1007/BF00336134

497 Blaud, A., Lerch, T.Z., Chevallier, T., Nunan, N., Chenu, C., Brauman, A., 2012. Dynamics of
 498 bacterial communities in relation to soil aggregate formation during the decomposition
 499 of ¹³C-labelled rice straw. Applied Soil Ecology 53, 1–9.
 500 doi:10.1016/j.apsoil.2011.11.005

501 Blaud, A., Menon, M., van der Zaan, B., Lair, G.J., Banwart, S.A., 2017. Chapter Five - Effects of Dry
 502 and Wet Sieving of Soil on Identification and Interpretation of Microbial Community
 503 Composition, in: Sparks, S.A.B. and D.L. (Ed.), Advances in Agronomy, Quantifying and
 504 Managing Soil Functions in Earth's Critical Zone Combining Experimentation and
 505 Mathematical Modelling. Academic Press, pp. 119–142.
 506 doi:10.1016/bs.agron.2016.10.006

507 Bossuyt, H., Six, J., Hendrix, P.F., 2002. Aggregate-protected carbon in no-tillage and
 508 conventional tillage agroecosystems using carbon-14 labeled plant residue. Soil Science
 509 Society of America Journal 66, 1965–1973.

510 Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. *Geoderma* 124, 3–22.
511 doi:10.1016/j.geoderma.2004.03.005

512 Chèneby, D., Brauman, A., Rabary, B., Philippot, L., 2009. Differential Responses of Nitrate
513 Reducer Community Size, Structure, and Activity to Tillage Systems. *Applied and
514 Environmental Microbiology* 75, 3180–3186. doi:10.1128/AEM.02338-08

515 Chotte, J.L., Schwartzmann, A., Bally, R., Jocteur Monrozier, L., 2002. Changes in bacterial
516 communities and *Azospirillum* diversity in soil fractions of a tropical soil under 3 or 19
517 years of natural fallow. *Soil Biology and Biochemistry* 34, 1083–1092.
518 doi:10.1016/S0038-0717(02)00041-X

519 Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak,
520 J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal
521 distinct aggregate stratification of soil bacterial communities and organic matter
522 composition. *Soil Biology and Biochemistry* 46, 63–72.
523 doi:10.1016/j.soilbio.2011.11.012

524 Dobbie, K.E., Smith, K.A., 2001. The effects of temperature, water-filled pore space and land use
525 on N₂O emissions from an imperfectly drained gleysol. *European Journal of Soil Science*
526 52, 667–673. doi:10.1046/j.1365-2389.2001.00395.x

527 Enwall, K., Throbäck, I.N., Stenberg, M., Söderström, M., Hallin, S., 2010. Soil resources influence
528 spatial patterns of denitrifying communities at scales compatible with land management.
529 *Applied and Environmental Microbiology* 76, 2243–2250. doi:10.1128/AEM.02197-09

530 Hallin, S., Jones, C.M., Schlöter, M., Philippot, L., 2009. Relationship between N-cycling
531 communities and ecosystem functioning in a 50-year-old fertilization experiment. *The
532 ISME Journal* 3, 597–605.

533 Hayden, H.L., Drake, J., Imhof, M., Oxley, A.P.A., Norng, S., Mele, P.M., 2010. The abundance of
534 nitrogen cycle genes *amoA* and *nifH* depends on land-uses and soil types in South-
535 Eastern Australia. *Soil Biology and Biochemistry* 42, 1774–1783.
536 doi:10.1016/j.soilbio.2010.06.015

537 Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial
 538 biomass and alters community profiles in soil aggregates. *Applied Soil Ecology* 46, 390–
 539 397. doi:10.1016/j.apsoil.2010.10.002
 540 IUSS Working Group WRB, 2006. World reference base for soil resources 2006, World soil
 541 resources rep 103. ed. FAO, Rome.
 542 Izquierdo, J., Nüsslein, K., 2006. Distribution of extensive nifH gene diversity across physical soil
 543 microenvironments. *Microbial Ecology* 51, 441–452. doi:10.1007/s00248-006-9044-x
 544 Kaiser, M., Kleber, M., Berhe, A.A., 2015. How air-drying and rewetting modify soil organic
 545 matter characteristics: An assessment to improve data interpretation and inference. *Soil*
 546 *Biology and Biochemistry* 80, 324–340. doi:10.1016/j.soilbio.2014.10.018
 547 Kravchenko, A.N., Negassa, W.C., Guber, A.K., Hildebrandt, B., Marsh, T.L., Rivers, M.L., 2014.
 548 Intra-aggregate pore structure influences phylogenetic composition of bacterial
 549 community in macroaggregates. *Soil Science Society of America Journal* 78, 1924.
 550 doi:10.2136/sssaj2014.07.0308
 551 Lair, G.J., Zehetner, F., Hrachowitz, M., Franz, N., Maringer, F.-J., Gerzabek, M.H., 2009. Dating of
 552 soil layers in a young floodplain using iron oxide crystallinity. *Quaternary*
 553 *Geochronology* 4, 260–266. doi:10.1016/j.quageo.2008.11.003
 554 Lauber, C.L., Strickland, M.S., Bradford, M.A., Fierer, N., 2008. The influence of soil properties on
 555 the structure of bacterial and fungal communities across land-use types. *Soil Biology and*
 556 *Biochemistry* 40, 2407–2415. doi:10.1016/j.soilbio.2008.05.021
 557 Leininger, S., Urich, T., Schlöter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C.,
 558 Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils.
 559 *Nature* 442, 806–809. doi:10.1038/nature04983
 560 Lensi, R., Clays-Josser, A., Jocteur Monrozier, L., 1995. Denitrifiers and denitrifying activity in
 561 size fractions of a mollisol under permanent pasture and continuous cultivation. *Soil*
 562 *Biology and Biochemistry* 27, 61–69. doi:10.1016/0038-0717(94)00132-K

563 Levy-Booth, D.J., Prescott, C.E., Grayston, S.J., 2014. Microbial functional genes involved in
 564 nitrogen fixation, nitrification and denitrification in forest ecosystems. *Soil Biology and*
 565 *Biochemistry* 75, 11–25. doi:10.1016/j.soilbio.2014.03.021
 566 Ma, W.K., Bedard-Haughn, A., Siciliano, S.D., Farrell, R.E., 2008. Relationship between nitrifier
 567 and denitrifier community composition and abundance in predicting nitrous oxide
 568 emissions from ephemeral wetland soils. *Soil Biology and Biochemistry* 40, 1114–1123.
 569 doi:10.1016/j.soilbio.2007.12.004
 570 Maeda, K., Spor, A., Edel-Hermann, V., Heraud, C., Breuil, M.-C., Bizouard, F., Toyoda, S., Yoshida,
 571 N., Steinberg, C., Philippot, L., 2015. N₂O production, a widespread trait in fungi.
 572 *Scientific Reports* 5. doi:10.1038/srep09697
 573 Mangalassery, S., Sjögersten, S., Sparkes, D.L., Sturrock, C.J., Mooney, S.J., 2013. The effect of soil
 574 aggregate size on pore structure and its consequence on emission of greenhouse gases.
 575 *Soil and Tillage Research* 132, 39–46. doi:10.1016/j.still.2013.05.003
 576 Mendes, I.C., Bottomley, P.J., 1998. Distribution of a population of *Rhizobium leguminosarum* bv.
 577 *trifolii* among different size classes of soil aggregates. *Applied and Environmental*
 578 *Microbiology* 64, 970–975.
 579 Morales, S.E., Cosart, T., Holben, W.E., 2010. Bacterial gene abundances as indicators of
 580 greenhouse gas emission in soils. *The ISME Journal* 4, 799–808.
 581 doi:10.1038/ismej.2010.8
 582 Muruganandam, S., Israel, D.W., Robarge, W.P., 2009. Activities of nitrogen-mineralization
 583 enzymes associated with soil aggregate size fractions of three tillage systems. *Soil*
 584 *Science Society of America Journal* 73, 751. doi:10.2136/sssaj2008.0231
 585 Neumann, D., Heuer, A., Hemkemeyer, M., Martens, R., Tebbe, C.C., 2013. Response of microbial
 586 communities to long-term fertilization depends on their microhabitat. *FEMS*
 587 *Microbiology Ecology* 86, 71–84. doi:10.1111/1574-6941.12092

588 Pinheiro, E.F.M., Pereira, M.G., Anjos, L.H.C., 2004. Aggregate distribution and soil organic matter
 589 under different tillage systems for vegetable crops in a Red Latosol from Brazil. *Soil and*
 590 *Tillage Research* 77, 79–84. doi:10.1016/j.still.2003.11.005
 591 Poly, F., Ranjard, L., Nazaret, S., Gourbiere, F., Jocteur Monrozier, L., 2001. Comparison of nifH
 592 gene pools in soils and soil microenvironments with contrasting properties. *Applied and*
 593 *Environmental Microbiology* 67, 2255–2262. doi:10.1128/AEM.67.5.2255-2262.2001
 594 Prosser, J.I., Nicol, G.W., 2008. Relative contributions of archaea and bacteria to aerobic
 595 ammonia oxidation in the environment. *Environmental Microbiology* 10, 2931–2941.
 596 doi:10.1111/j.1462-2920.2008.01775.x
 597 R Development Core Team, 2015. R: a language and environment for statistical computing.
 598 Rabbi, S.M.F., Daniel, H., Lockwood, P.V., Macdonald, C., Pereg, L., Tighe, M., Wilson, B.R., Young,
 599 I.M., 2016. Physical soil architectural traits are functionally linked to carbon
 600 decomposition and bacterial diversity. *Scientific Reports* 6, 33012.
 601 doi:10.1038/srep33012
 602 Ramakrishnan, B., Lueders, T., Conrad, R., Friedrich, M., 2000. Effect of soil aggregate size on
 603 methanogenesis and archaeal community structure in anoxic rice field soil. *FEMS*
 604 *Microbiology Ecology* 32, 261–270. doi:10.1111/j.1574-6941.2000.tb00719.x
 605 Schindlbacher, A., Zechmeister-Boltenstern, S., Butterbach-Bahl, K., 2004. Effects of soil
 606 moisture and temperature on NO, NO₂, and N₂O emissions from European forest soils.
 607 *Journal of Geophysical Research: Atmospheres* 109, 1–12. doi:10.1029/2004JD004590
 608 Sey, B.K., Manceur, A.M., Whalen, J.K., Gregorich, E.G., Rochette, P., 2008. Small-scale
 609 heterogeneity in carbon dioxide, nitrous oxide and methane production from aggregates
 610 of a cultivated sandy-loam soil. *Soil Biology and Biochemistry* 40, 2468–2473.
 611 doi:10.1016/j.soilbio.2008.05.012
 612 Shein, E.V., 2005. *Kurs fiziki pochv (A Course of Soil Physics)* [in Russian]. Moscow State Univ.
 613 Publishing.

Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between (micro)aggregates, soil biota, and soil organic matter dynamics. *Soil and Tillage Research* 79, 7–31. doi:10.1016/j.still.2004.03.008

Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant and Soil* 241, 155–176. doi:10.1023/A:1016125726789

Six, J., Elliott, E.T., Paustian, K., 2000. Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. *Soil Biology and Biochemistry* 32, 2099–2103. doi:10.1016/S0038-0717(00)00179-6

Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. *European Journal of Soil Science* 33, 141–163. doi:10.1111/j.1365-2389.1982.tb01755.x

Wallenstein, M.D., Vitgalys, R.J., 2005. Quantitative analyses of nitrogen cycling genes in soils. *Pedobiologia* 49, 665–672. doi:10.1016/j.pedobi.2005.05.005

Table 1. Soil characteristics and soil aggregate size distribution of bulk soil samples on a dry mass basis. Mean value \pm one standard deviation ($n = 3$) are shown.

Location		Cropland 48°09'N, 16°41'E	Young forest 48°07'N, 16°43'E	Forest 48°08'N, 16°39'E	Grassland 48°11'N, 16°44'E
Soil characteristics	Soil (0-10 cm) age (yr)	< 70	250-350	250-350	250-350
	Water content (%)	11.3 \pm 0.26	14.1 \pm 1.11	17.1 \pm 0.69	12.0 \pm 0.26
	Soil pH (H ₂ O)	7.7 \pm 0.14	7.5 \pm 0.07	7.4 \pm 0.17	7.4 \pm 0.09
	Organic C (%)	2.4 \pm 0.36	3.2 \pm 0.08	3.8 \pm 0.28	5.0 \pm 0.60
	Total N (%)	0.13 \pm 0.01	0.17 \pm 0.01	0.25 \pm 0.02	0.33 \pm 0.04
	C _{org} /N	18.1 \pm 1.83	18.5 \pm 1.60	15.1 \pm 1.02	15.0 \pm 0.52
	N-NH ₄ ⁺ (mg kg ⁻¹)	1.59 \pm 0.29	0.49 \pm 0.01	0.57 \pm 0.03	4.77 \pm 0.98
	N-NO ₃ ⁻ (mg kg ⁻¹)	20.3 \pm 3.07	18.6 \pm 4.00	24.3 \pm 3.13	1.5 \pm 0.66
	P-PO ₄ ³⁻ (g kg ⁻¹)	0.35 \pm 0.10	1.13 \pm 0.47	0.85 \pm 0.48	0.59 \pm 0.04
	CaCO ₃ (%)	19.0 \pm 1.90	20.6 \pm 1.11	20.4 \pm 0.62	21.1 \pm 1.41
	Sand, 63-2000 μ m (%)	32.7	20.2	22.5	8.2
	Silt, 2-63 μ m (%)	43.8	63.4	51.2	63.0
	Clay, < 2 μ m (%)	23.5	16.4	26.3	28.8
Soil aggregate size distribution (%)	> 10 mm	37.3 \pm 9.1	11.3 \pm 1.0	11.9 \pm 4.4	7.9 \pm 2.4
	5.0 - 10.0 mm	14.6 \pm 2.4	15.5 \pm 1.1	18.3 \pm 2.7	21.5 \pm 2.0
	2.0 - 5.0 mm	20.5 \pm 4.0	26.1 \pm 3.1	31.2 \pm 2.2	37.8 \pm 3.6
	1.0 - 2.0 mm	11.8 \pm 2.4	21.8 \pm 4.1	23.1 \pm 8.4	14.5 \pm 0.5
	0.5 - 1.0 mm	6.4 \pm 3.5	9.3 \pm 2.8	5.9 \pm 1.7	5.2 \pm 0.4
	0.25 - 0.5 mm	7.1 \pm 4.6	12.7 \pm 2.6	7.5 \pm 2.7	6.9 \pm 0.1
	< 0.25 mm	1.9 \pm 1.3	3.3 \pm 0.4	2.0 \pm 0.8	6.1 \pm 0.7

Figures captions

Fig. 1 Variation in gene abundance between bulk soil from four land use types. The following genes and microbial communities were targeted: bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixating (*nifH* gene), ammonia oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), nitrate reductase (*narG* gene), nitrite reductase (*nirK* gene) and nitrous oxide reductase (*nosZ* gene). All abundances are expressed on the basis of 1 g of dry soil. Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between land use.

Fig. 2. Variation in gene abundance between bulk soil and six soil aggregates sizes classes from young forest and forest. The following genes and microbial communities were targeted: bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixating (*nifH* gene), ammonia oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), nitrate reductase (*narG* gene), nitrite reductase (*nirK* gene) and nitrous oxide reductase (*nosZ* gene). All abundances are expressed on the basis of 1 g of dry mass of the bulk soil or the specific aggregate size fraction. Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

Fig. 3. Variation in GHG fluxes ($\mu\text{g kg}^{-1} \text{h}^{-1}$) between bulk soil from four land use types at field moisture or elevated moisture (40 – 60 % of field capacity). Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

Fig. 4. Variation in GHG fluxes ($\mu\text{g kg}^{-1} \text{h}^{-1}$) between bulk soil and six soil aggregates sizes classes from grassland or forest at field moisture or elevated moisture (40 – 60 % of field capacity).

Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

Fig. 5. Heatmaps of Spearman's rank correlation coefficients ρ between a) soil properties and microbial genes abundance, b) soil properties and greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25 , $0.25 - 0.5$, $0.5 - 1.0$, $1.0 - 2.0$, $2.0 - 5.0$ and $5.0 - 10.0$ mm) and four land uses. AOB: amoA bacteria; AOA: amoA archaea. The ρ values > 0.24 and < -0.24 are significant ($P < 0.05$).

Fig. 6 Heatmaps of Spearman's rank correlation coefficients ρ between soil properties and microbial genes abundance from samples across six soil aggregates sizes classes (< 0.25 , $0.25 - 0.5$, $0.5 - 1.0$, $1.0 - 2.0$, $2.0 - 5.0$ and $5.0 - 10.0$ mm) and for a) young forest and b) forest sites separately. AOB: amoA bacteria; AOA: amoA archaea. The ρ values > 0.47 and < -0.47 are significant ($P < 0.05$).

Table 1. Soil characteristics and soil aggregate size distribution of bulk soil samples on a dry mass basis. Mean value \pm one standard deviation ($n = 3$) are shown.

		Cropland	Young forest	Forest	Grassland
Location		48°09'N, 16°41'E	48°07'N, 16°43'E	48°08'N, 16°39'E	48°11'N, 16°44'E
Soil characteristics	Soil (0-10 cm) age (yr)	< 70	250-350	250-350	250-350
	Water content (%)	11.3 \pm 0.26	14.1 \pm 1.11	17.1 \pm 0.69	12.0 \pm 0.26
	Soil pH (H ₂ O)	7.7 \pm 0.14	7.5 \pm 0.07	7.4 \pm 0.17	7.4 \pm 0.09
	Organic C (%)	2.4 \pm 0.36	3.2 \pm 0.08	3.8 \pm 0.28	5.0 \pm 0.60
	Total N (%)	0.13 \pm 0.01	0.17 \pm 0.01	0.25 \pm 0.02	0.33 \pm 0.04
	C _{org} /N	18.1 \pm 1.83	18.5 \pm 1.60	15.1 \pm 1.02	15.0 \pm 0.52
	N-NH ₄ ⁺ (mg kg ⁻¹)	1.59 \pm 0.29	0.49 \pm 0.01	0.57 \pm 0.03	4.77 \pm 0.98
	N-NO ₃ ⁻ (mg kg ⁻¹)	20.3 \pm 3.07	18.6 \pm 4.00	24.3 \pm 3.13	1.5 \pm 0.66
	P-PO ₄ ³⁻ (g kg ⁻¹)	0.35 \pm 0.10	1.13 \pm 0.47	0.85 \pm 0.48	0.59 \pm 0.04
	CaCO ₃ (%)	19.0 \pm 1.90	20.6 \pm 1.11	20.4 \pm 0.62	21.1 \pm 1.41
	Sand, 63-2000 μ m (%)	32.7	20.2	22.5	8.2
	Silt, 2-63 μ m (%)	43.8	63.4	51.2	63.0
	Clay, < 2 μ m (%)	23.5	16.4	26.3	28.8
Soil aggregate size distribution (%)	> 10 mm	37.3 \pm 9.1	11.3 \pm 1.0	11.9 \pm 4.4	7.9 \pm 2.4
	5.0 - 10.0 mm	14.6 \pm 2.4	15.5 \pm 1.1	18.3 \pm 2.7	21.5 \pm 2.0
	2.0 - 5.0 mm	20.5 \pm 4.0	26.1 \pm 3.1	31.2 \pm 2.2	37.8 \pm 3.6
	1.0 - 2.0 mm	11.8 \pm 2.4	21.8 \pm 4.1	23.1 \pm 8.4	14.5 \pm 0.5
	0.5 - 1.0 mm	6.4 \pm 3.5	9.3 \pm 2.8	5.9 \pm 1.7	5.2 \pm 0.4
	0.25 - 0.5 mm	7.1 \pm 4.6	12.7 \pm 2.6	7.5 \pm 2.7	6.9 \pm 0.1
	< 0.25 mm	1.9 \pm 1.3	3.3 \pm 0.4	2.0 \pm 0.8	6.1 \pm 0.7

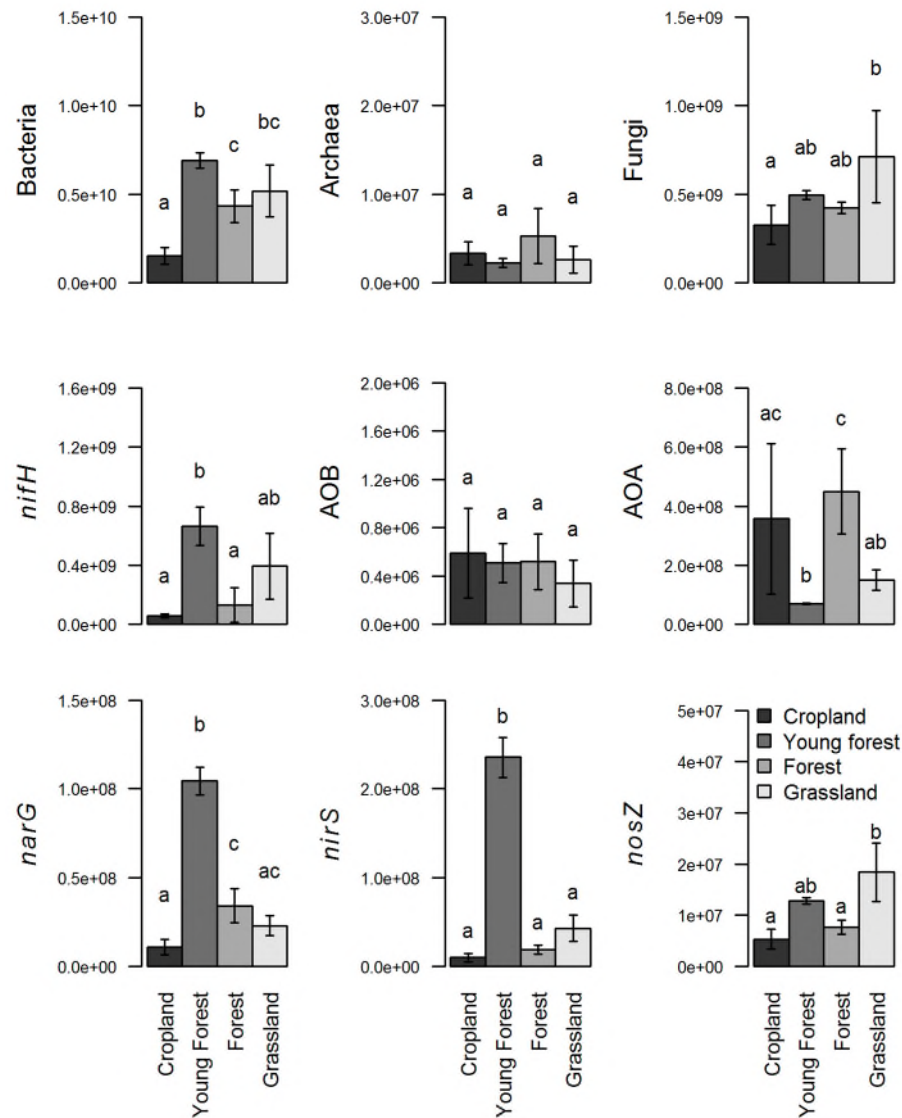


Fig. 1 Variation in gene abundance between bulk soil from four land use types. The following genes and microbial communities were targeted: bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixating (*nifH* gene), ammonia oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), nitrate reductase (*narG* gene), nitrite reductase (*nirK* gene) and nitrous oxide reductase (*nosZ* gene). All abundances are expressed on the basis of 1 g of dry soil. Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between land use.

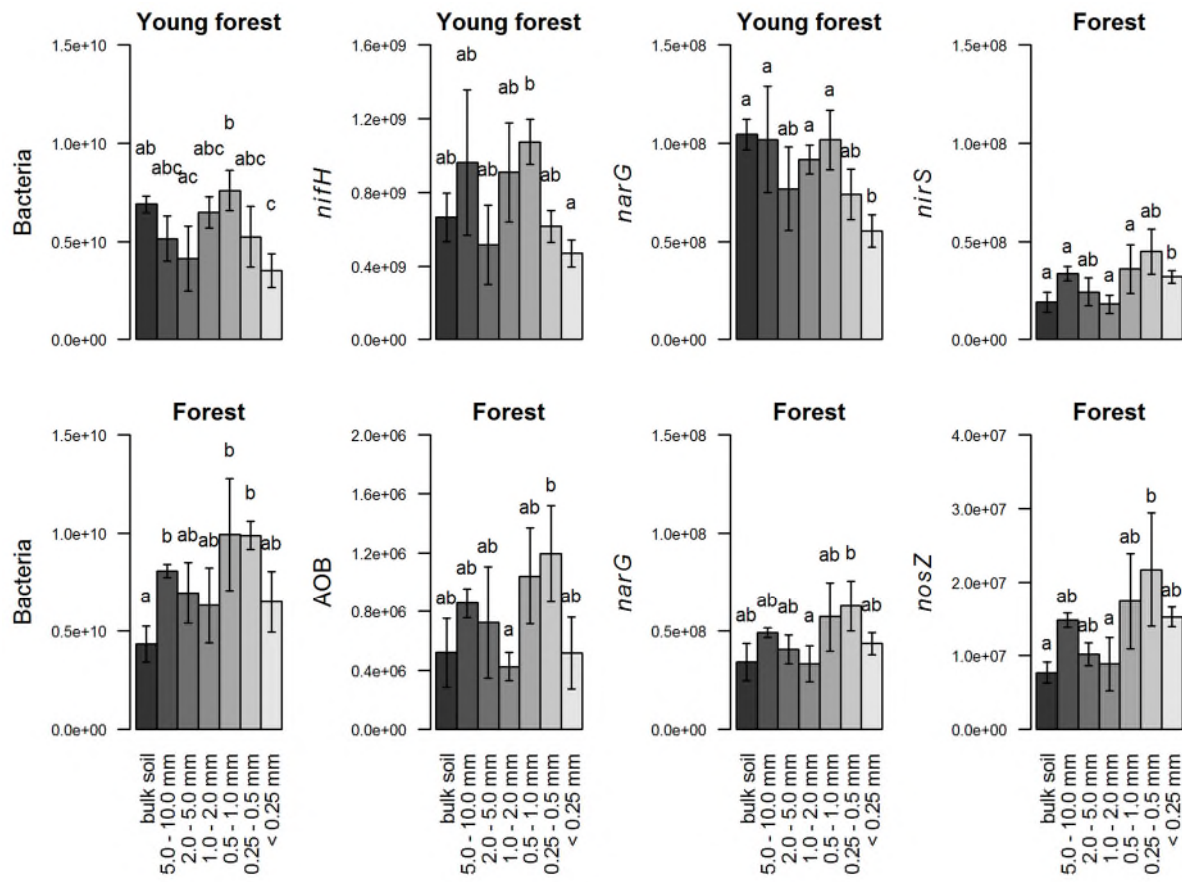


Fig. 2 Variation in gene abundance between bulk soil and six soil aggregates sizes classes from young forest and forest. The following genes and microbial communities were targeted: bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixating (*nifH* gene), ammonia oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), nitrate reductase (*narG* gene), nitrite reductase (*nirK* gene) and nitrous oxide reductase (*nosZ* gene). All abundances are expressed on the basis of 1 g of dry mass of the bulk soil or the specific aggregate size fraction. Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

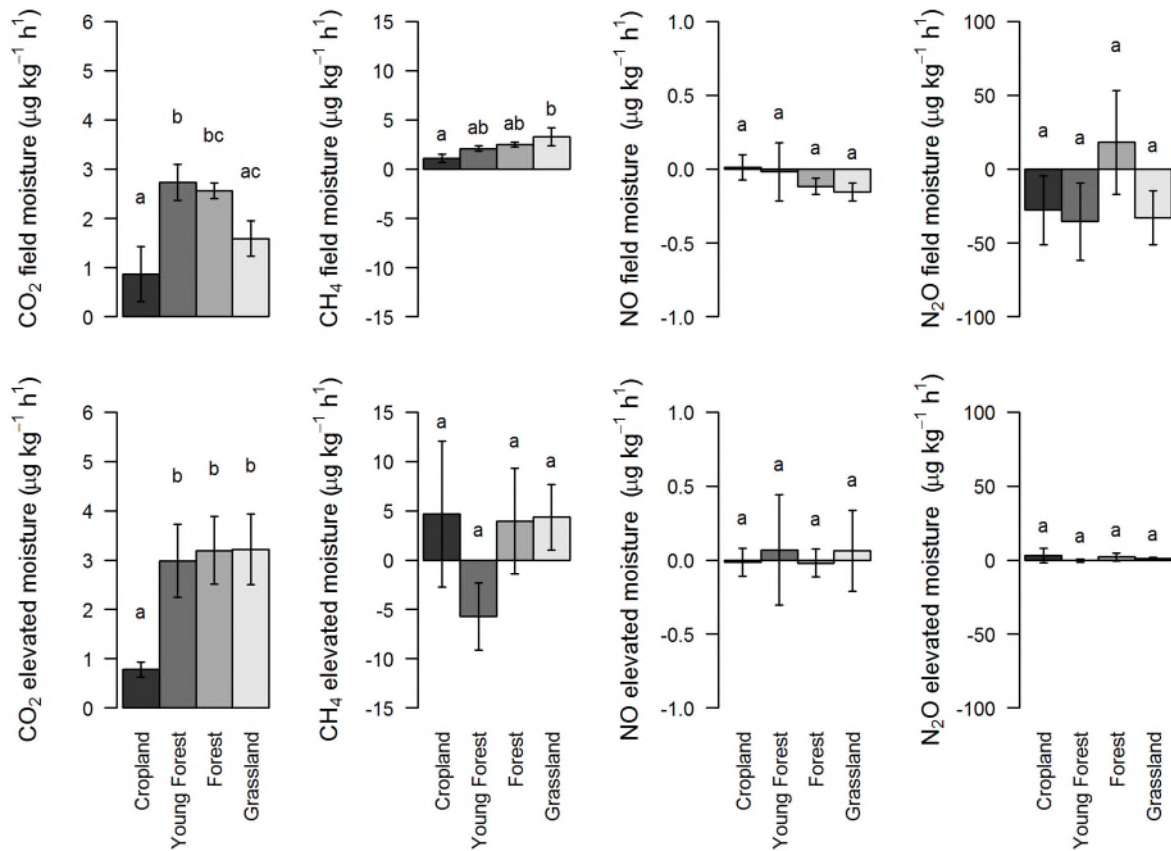


Fig. 3. Variation in GHG fluxes ($\mu\text{g kg}^{-1} \text{h}^{-1}$) between bulk soil from four land use types at field moisture or elevated moisture (40 – 60 % of field capacity). Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

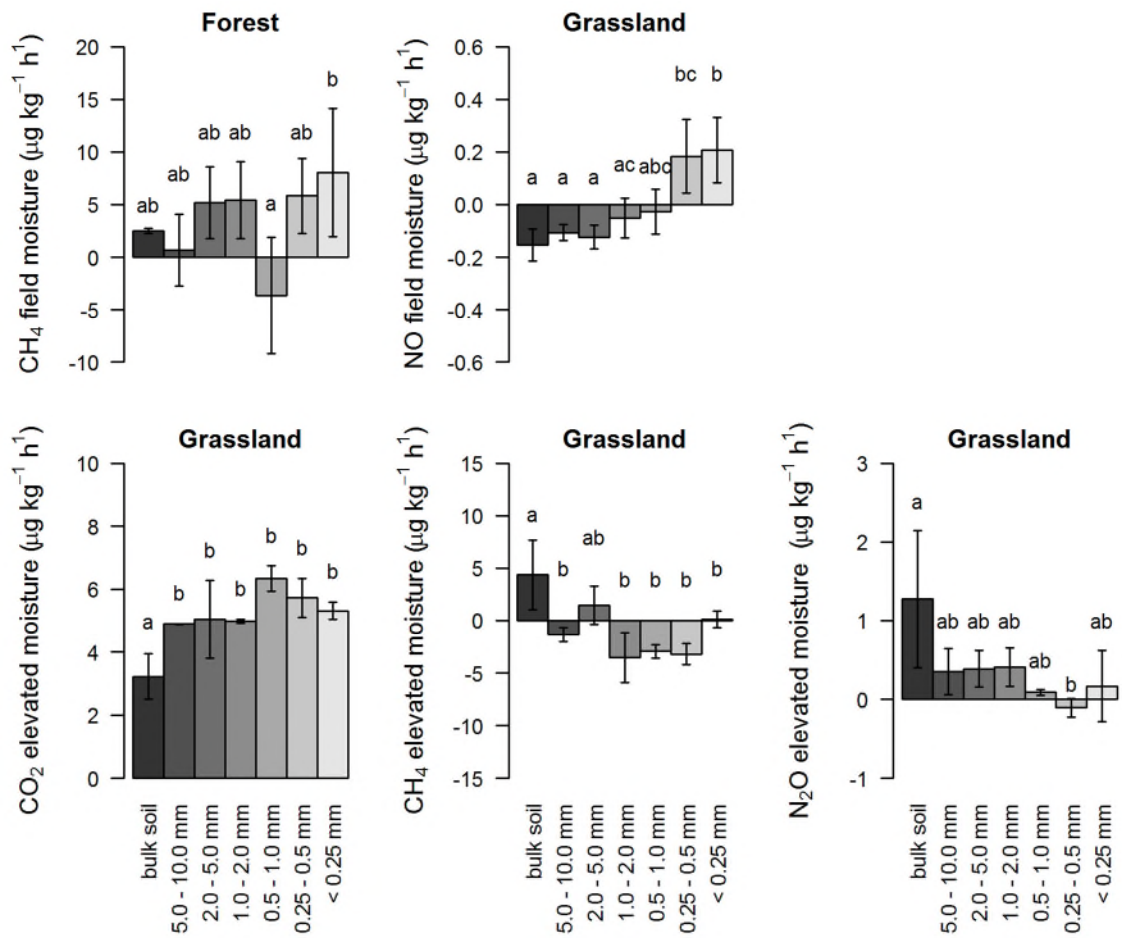


Fig. 4. Variation in GHG fluxes ($\mu\text{g kg}^{-1} \text{h}^{-1}$) between bulk soil and six soil aggregates sizes classes from grassland or forest at field moisture or elevated moisture (40 – 60 % of field capacity). Mean value \pm one standard deviation ($n = 3$) are shown. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

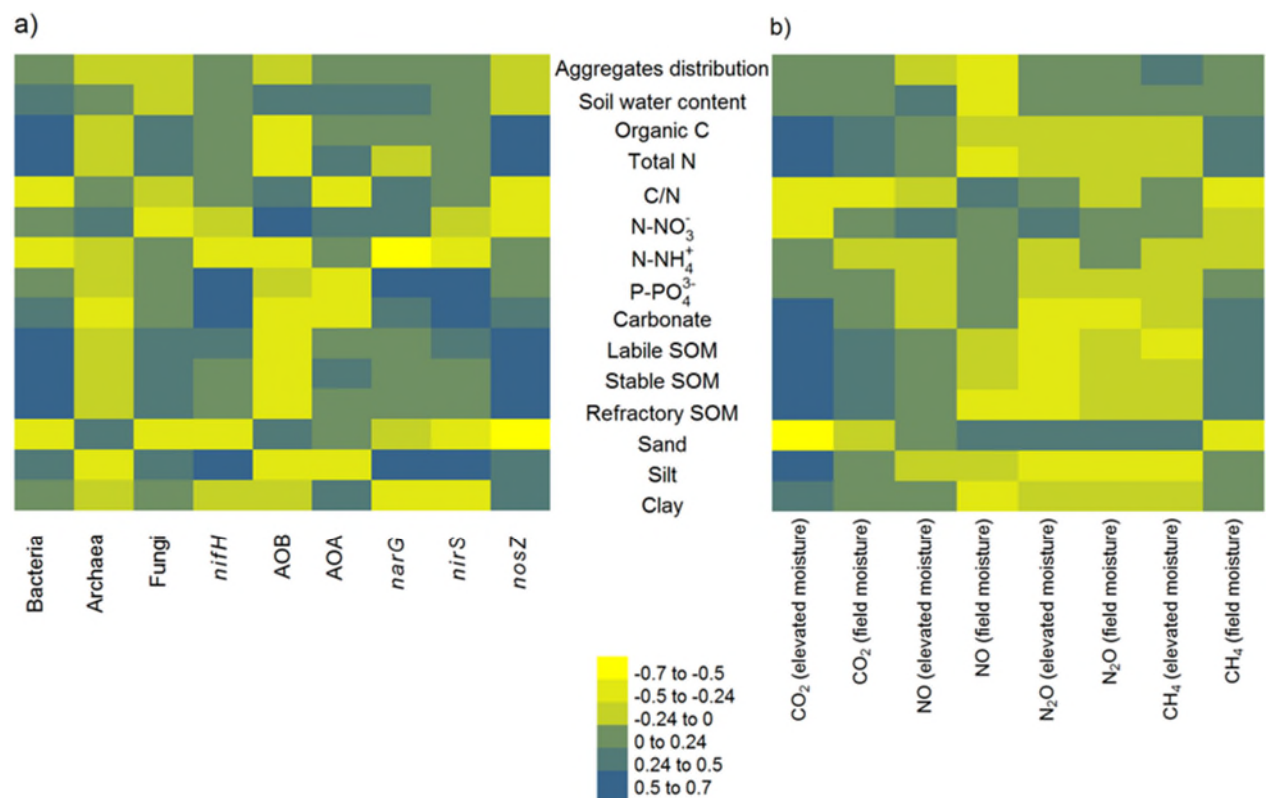


Fig. 5. Heatmaps of Spearman's rank correlation coefficients ρ between a) soil properties and microbial genes abundance, b) soil properties and greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25, 0.25 – 0.5, 0.5 – 1.0, 1.0 – 2.0, 2.0 – 5.0 and 5.0 – 10.0 mm) and four land uses. AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ values > 0.24 and < -0.24 are significant ($P < 0.05$).

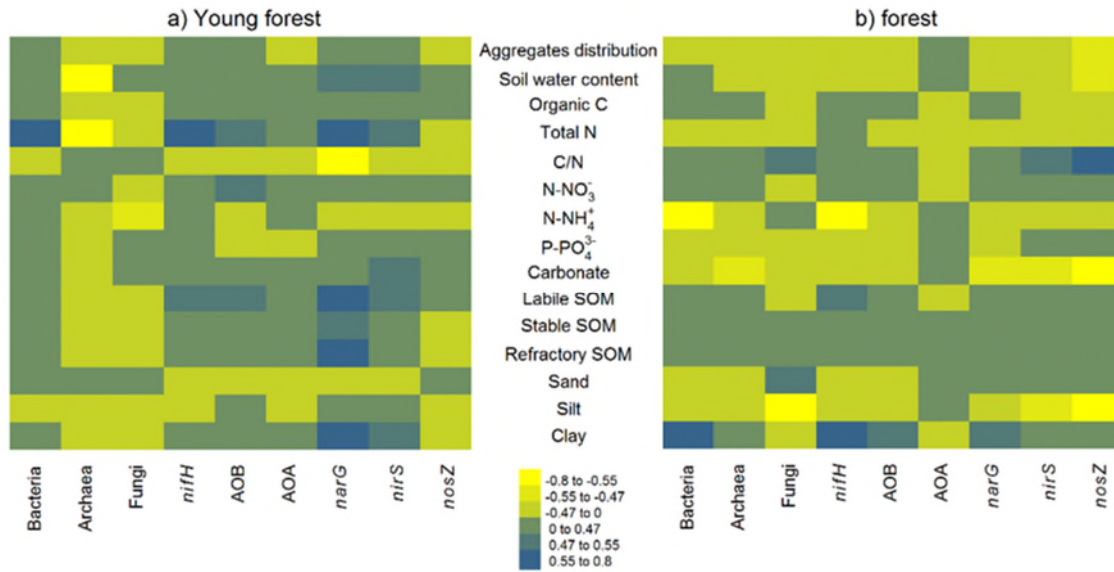


Fig. 6 Heatmaps of Spearman's rank correlation coefficients ρ between soil properties and microbial genes abundance from samples across six soil aggregates sizes classes (< 0.25 , $0.25 - 0.5$, $0.5 - 1.0$, $1.0 - 2.0$, $2.0 - 5.0$ and $5.0 - 10.0$ mm) and for a) young forest and b) forest sites separately. AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ values > 0.47 and < -0.47 are significant ($P < 0.05$).

The abundance of nitrogen genes cycle and potential greenhouse gas fluxes depend on land use type and little on soil aggregate size

Aimeric Blaud^{a, 1}, Bas van der Zaan^b, Manoj Menon^a, Georg J. Lair^{c, d}, Dayi Zhang^{a, 2}, Petra Huber^c, Jasmin Schiefer^c, Winfried E.H. Blum^c, Barbara Kitzler^e, Wei E.Huang^{a, 3}, Pauline van Gaans^b, and Steve Banwart^a*

^a Department of Civil and Structural Engineering, Kroto Research Institute, University of Sheffield, Broad Lane, Sheffield S3 7HQ, United Kingdom.

^b Deltares, Subsurface and Groundwater Systems, Princetonlaan 6-8, 3508 Al Utrecht, the Netherlands.

^c University of Natural Resources and Life Sciences (BOKU), Institute of Soil Research, Vienna, Peter-Jordan-Str. 82, 1190 Vienna, Austria.

^d University of Innsbruck, Institute of Ecology, Sternwartestr. 15, 6020 Innsbruck, Austria.

^e Department of Forest Ecology and Soil, Soil Ecology, Federal Research Centre for Forests, Seckendorff-Gudent-Weg 8, 1131 Vienna Austria.

***Corresponding Author.**

E-mail address: aimerik.blaud@gmail.com

¹ Current address: Sustainable Agriculture Science Department, Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK.

² Current address: Lancaster Environment Centre, Lancaster University, Lancaster LA1 2YQ, UK.

³ Current address: Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK.

Supplementary material and methods

Quantitative-PCR

Q-PCR standards for each molecular target were obtained using a 10-fold serial dilution of plasmids carrying a single cloned target gene or relevant part thereof. The standards were constructed by cloning the PCR product of the environmental samples of each individual PCR assay into pCR2.1 TOPO vector by using the TOPO TA cloning kit (Invitrogen, Breda, the Netherlands) according to the manufacturer's protocol. Cloned inserts were isolated using the Qiagen Plasmid mini Kit and checked for concentration and purity on a Nanodrop ND-1000 spectrophotometer (Isogen). Presence of the gene of interest was confirmed by sequence-analysis (MWG-Biotech, Germany). The total number of plasmids with cloned target genes in the Q-PCR Standard was calculated based on its total DNA concentration (Nanodrop), assuming an average molecular mass for each nucleotide pair of 660 pg/ml (Smith et al., 2006).

Standard curve template DNA and the "no template control" (NTC) were amplified in duplicate in the same plate as the environmental samples. Five tenfold dilutions were used for each Q-PCR assay. Q-PCR amplifications were performed in 25 µl volumes containing 12.5 µl of iQ™ SYBR® Green Supermix (Bio-Rad, Hemel Hempstead, UK), 8.5 µl of nuclease-free water (Ambion, Warrington, UK), 1.25 µl of each primer (10 µM) and 1 µl of template DNA using a CFX96™ Real-Time System (Bio-Rad, Hemel Hempstead, UK). Standard amplification was used for all Q-PCR assays except archaeal *amoA*, starting with an initial denaturation at 95 °C for 3 min, followed by 40 cycles of 30 s at 95 °C, 0.5 to 1 min of annealing (annealing temperature and time for each primers pairs are given in Table S1), and 30 s at 72 °C. Amplification for the archaeal *amoA* gene followed the procedure as described by (Tsiknia et al., 2013). The fluorescence was measured at the end of each synthesis step (i.e. at 81 °C for archaeal *amoA* and at 72 °C for all other genes)

Threshold cycle (Ct) values and amplicon numbers were determined automatically using the Bio-Rad CFX Manager™ software. The efficiency of the Q-PCR assays was above 90%, except for fungi and AOA (~70%). The r^2 were > 0.99, except for *nifH* and *nosZ* genes (~0.97).

Specificity of the Q-PCR was assessed via a melting curve analysis (increase of temperature from annealing temperature to 95 °C by 0.5 °C per step of 0.05 s) at the end of each Q-PCR amplification (Ririe et al., 1997). The melting curves for the bacterial and archaeal 16S rRNA, *nifH*, *amoA*, *narG*, *nirS*, and *nosZ* genes Q-PCR assays showed specificity for the amplified targeted genes (i.e. single peak). As expected, the melting curve of the Q-PCR for fungal ITS showed the amplification of products of different lengths, due to the variability in length of ITS regions among different fungal taxa (Manter and Vivanco, 2007).

Microbial respiration

Greenhouse gas fluxes from the aggregate size fractions and the bulk soil were measured from field moist bulk soil and soil aggregates (pF 3.8 -4.0; hereafter named “field moisture”) and from moistened samples (40 – 60 % of field capacity) by adding distilled water 48 hours before flux measurements started (hereafter named “elevated moisture”). Soil temperature was set to 20 °C. The soil moisture was increased because at the time of soil sampling the soil moisture content was low (pF 3.8-4.0), potentially reducing microbial activity and subsequent GHG fluxes.

Fluxes of CO₂ and NO were measured with a fully automated laboratory measuring system with 13 adapted Kilner jars serving as test chambers in a temperature-controlled incubator and connected to a CO₂ and a NO_x analyser. Twelve test chambers were used as incubation chambers. One chamber was used as a reference where no soil was incubated. The measuring system is described in detail by Schindlbacher et al., (2004) and Schaufler et al., (2010). For CO₂ and NO flux determination, air from inside the incubator was drawn through the chambers to the CO₂ and NO_x analysers with a constant flow rate of 1.0 l min⁻¹. To avoid accumulation of CO₂ and NO in the incubator, the incubator was flushed with compressed ambient air (1.0 l min⁻¹). Carbon dioxide was measured with a PP Systems WMA-2 (Amesbury, MA, USA), infrared CO₂ analyser, and NO was measured with a HORIBA APNA-360 (Kyoto,

Japan) chemoluminescence NO_x analyser. The measuring time of each chamber was 8 min according to achievement of steady state.

Determination of N₂O and CH₄ fluxes was done manually by closed chamber technique. The soil samples were put into Kilner jars and closed air-tight with a PVC lid. A glass tube, with a total volume of 685 cm³, was fitted into the PVC lid and closed air-tight with rubber septa and sealed with silicon grease. Twelve ml of the gas sample were extracted from each chamber in triplicate at intervals of 15min and injected into sealed and pre-evacuated sampling vials with a glass syringe. The analysis was done immediately by gas chromatography (AGILENT 6890N) connected to an automated system sample-injection (AGILENT TECH G1888, Network HEADSPACE-SAMPLER) at an oven temperature of 40 °C. Nitrous oxide was measured by a ⁶³Ni-electron-capture detector (ECD; detector: 350 °C) and CH₄ by a flame ionization detector (FID; detector: 250 °C). Standard gases (Inc. Linde Gas) were used as a reference and contained 0.5, 1 and 2.5 µl l⁻¹ N₂O; 1, 2 and 4µl l⁻¹ CH₄. Data were calculated as described in Kitzler et al. (2006).

Physico-chemical analysis of bulk soil and aggregates

The moisture content of each aggregate size class and the bulk soil was measured gravimetrically at 105 °C. The mass distribution over the predefined aggregate size classes was obtained by dry sieving of 100 g bulk soil from each sampling spot in triplicate (i.e. 9 replicates per site). Particle size distribution (i.e. the fractions of sand, silt and clay) for each aggregate size class and the bulk soil was determined by wet-sieving (20–2000 µm fractions) and sedimentation of the < 20 µm fraction in an X-ray sedigraph (Micromeritics Sedigraph 5000ET) after removal of organic matter with hydrogen peroxide and dispersion with sodium polyphosphate (Soil Survey Staff, 2004).

Total carbon was quantified by dry combustion (Tabatabai and Bremner, 1991) in an elemental analyser (Carlo Erba Nitrogen Analyser 500, Milano, Italy), and carbonate was measured gas-volumetrically (Soil Survey Staff, 2004). Soil organic C (SOC) was calculated as the difference of total and carbonate C. Soil and aggregate samples were extracted for N-NO₃⁻, N-

NH_4^- , and P-PO_4^- using 2 g of soil and 20 ml of KCl (1 M) shaken for 1 h. Concentration of N-NO_3^- was determined by the vanadium reduction method (Miranda et al., 2001), concentration of N-NH_4^- by the sodium salicylate-sodium nitroprusside method (Rowland, 1983), and the P-PO_4^- concentration by the ammonium molybdate–ascorbic acid method (Olsen et al., 1954).

Three different fractions of soil organic matter (SOM) were determined by simultaneous thermal analysis (STA) according to Barros et al. (2007), using 50 mg of oven dried (60°C) samples (Netzsch STA 409 PC). The samples were heated from 25 to 600°C at a rate of 5°C min^{-1} in a reaction atmosphere of synthetic air (flow rate: 50 mL min^{-1}). According to De la Rosa et al. (2008) STA allows the distinction of the amount of total SOM (decomposes between 190 and 550°C), into thermally labile SOM (decomposes between 190 and 390°C), thermally more stable SOM (decomposes between 390 and 450°C), and refractory SOM (decomposes between 450 and 550°C). In the labile fraction, SOM consists mainly of carbohydrates and proteins (De la Rosa et al., 2008), whereas in the thermally more stable SOM fraction polyphenolic and aromatic organic structures dominate (Lopes-Capel et al., 2005). Black carbon present in soil burns at higher temperatures within the refractory fraction (De la Rosa et al., 2008).

Particle size distribution in the various aggregate size classes as well as the SOM fractions (STA) were measured on one composite sample for each site (i.e. mixture of the 3 replicates/sampling spot at each site).

Table S1. Description of the primers used to target each community and the annealing temperature of each Q-PCR assays.

Target gene	Primer	Sequence 5'-3'	Annealing temp. (°C) and time (s)	References
Bacterial	519F	GCCAGCAGCCGCGGTAAT	58 (30 s)	Lane, 1991
<i>16SrRNA</i>	907R	CCGTCAATTCTTTGAGTTT		Stubner and Meuser, 2000
Archaeal	Arch 0025F	CTGGTTGATCCTGCCAG	58 (30 s)	Vetriani et al., 1999
<i>16SrRNA</i>	Arch 364R	ACGGGGCGCACGAGGCGCGA		Vetriani et al., 1999
Fungal	ITS1f	TCCGTAGGTGAACCTGCGG	50 (45 s)	Gardes and Bruns, 1993
<i>ITS</i>	5.8s	CGCTGCGTTCTTCATCG		Vilgalys and Hester, 1990
<i>nifH</i>	nifHF	AAAGGYGGWATCGGYAARTCCACCAC	62.5 (60 s)	Rösch and Bothe, 2005
	nifHRb	TGSGCYTTGTCYTCRCGGATBGGCAT		Rösch and Bothe, 2005
<i>amoA</i>	amoA_F	GGHGACTGGGAYTTCTGG	55.3 (30 s)	Holmes et al., 1995
Bacteria	amoA_R	CCTCKGSAAAGCCTTCTTC		Okano et al., 2004
<i>amoA</i>	amoAF	STAATGGTCTGGCTTAGACG	55 (35 s)	Francis et al., 2005
Archaea	amoAR	GCGGCCATCCATCTGTATGT		Francis et al., 2005
<i>narG</i>	NARG F	TCGCCSATYCCGGCSATGTC	63 (30 s)	López-Gutiérrez et al., 2004
	NARG R	GAGTTGTACCAGTCRGC SGAYTCSG		López-Gutiérrez et al., 2004
<i>nirS</i>	NIRS4Q F	G TSAACGYSAAGGARACSGG	63 (30 s)	Braker et al., 1998
	NIRS6Q R	GASTTCGGRTGSGTCTTSAYGAA		Braker et al., 1998
<i>nosZ</i>	nosZ1840_F	CGCRACGGCAASAAGGTSMSST	67 (30 s)	Henry et al., 2006
	nosZ2090_R	CAKRTGCAKSGCRTGGCAGAA		Henry et al., 2006

Table S2. Overview table of the two-way ANOVA with land use and aggregate size as factors.

Significant *P* values ($P < 0.05$) are shown in bold.

	Land use		Aggregate size		Interaction	
	F values	<i>P</i> values	F values	<i>P</i> values	F values	<i>P</i> values
Bacteria	54.458	< 2 10⁻¹⁶	4.154	0.00161	2.754	0.00197
Archaea	9.878	2.51 10⁻⁵	0.963	0.459	0.806	0.685
Fungi	9.768	2.79 10⁻⁵	1.594	0.166	0.830	0.6559
<i>nifH</i>	97.755	< 2 10⁻¹⁶	1.635	0.155	1.535	0.112
AOB	16.231	1.04 10⁻⁷	1.275	0.28353	2.473	0.00511
AOA	88.972	< 2 10⁻¹⁶	0.432	0.855	1.004	0.470
<i>narG</i>	184.079	< 2 10⁻¹⁶	2.843	0.017331	3.314	0.000305
<i>nirS</i>	246.065	< 2 10⁻¹⁶	0.768	0.5986	2.045	0.0216
<i>nosZ</i>	73.592	< 2 10⁻¹⁶	4.694	0.00062	1.889	0.03633

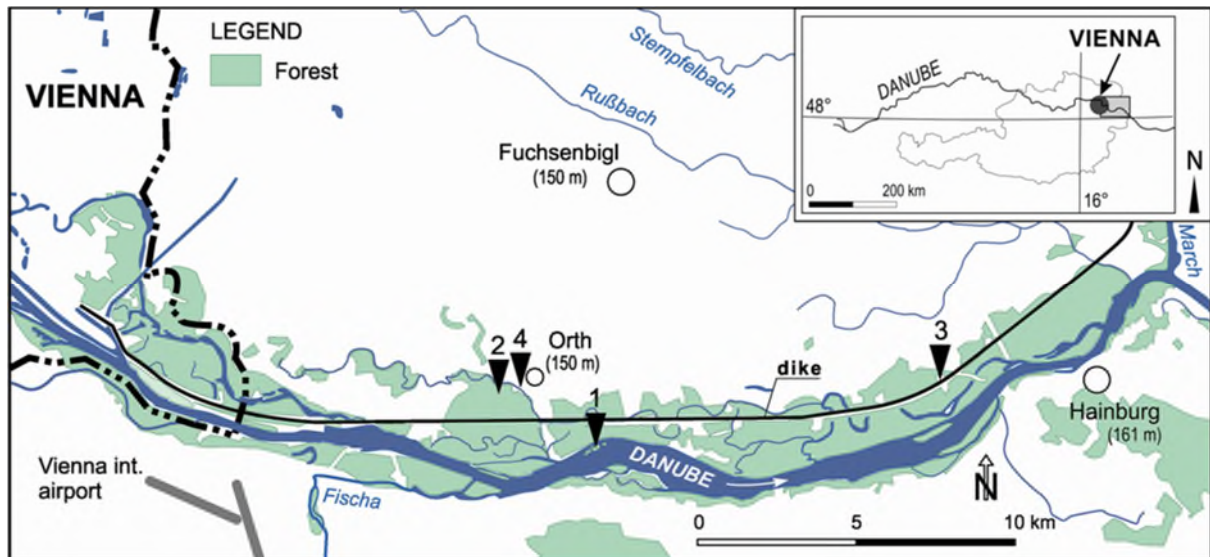


Fig. S1. Study area in the National Park “Donau-Auen” east of Vienna. The continuous black line represents a dike built from 1882 to 1905 to prevent flooding of the enclosed land. The numbers 1 to 4 indicate the 4 field sites/land uses: site 1: young forest; site 2: Forest, site 3: Grassland; site 4: Cropland.

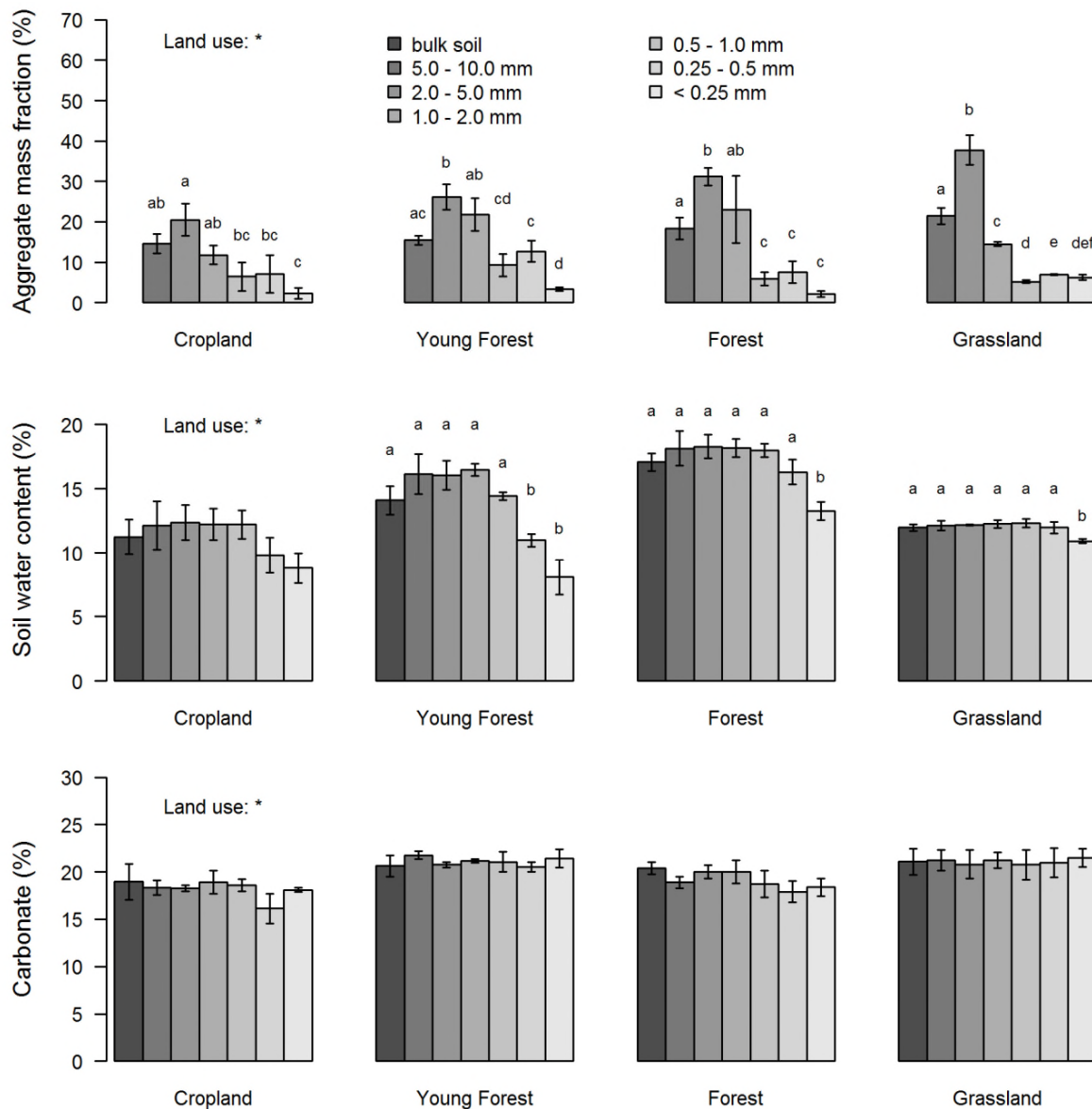


Fig. S2. Variation in mass aggregates distribution (%), soil water content (%), and carbonate concentration (%) between bulk soil and six soil aggregates sizes classes from four land use types. Mean value \pm one standard deviation ($n = 3$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

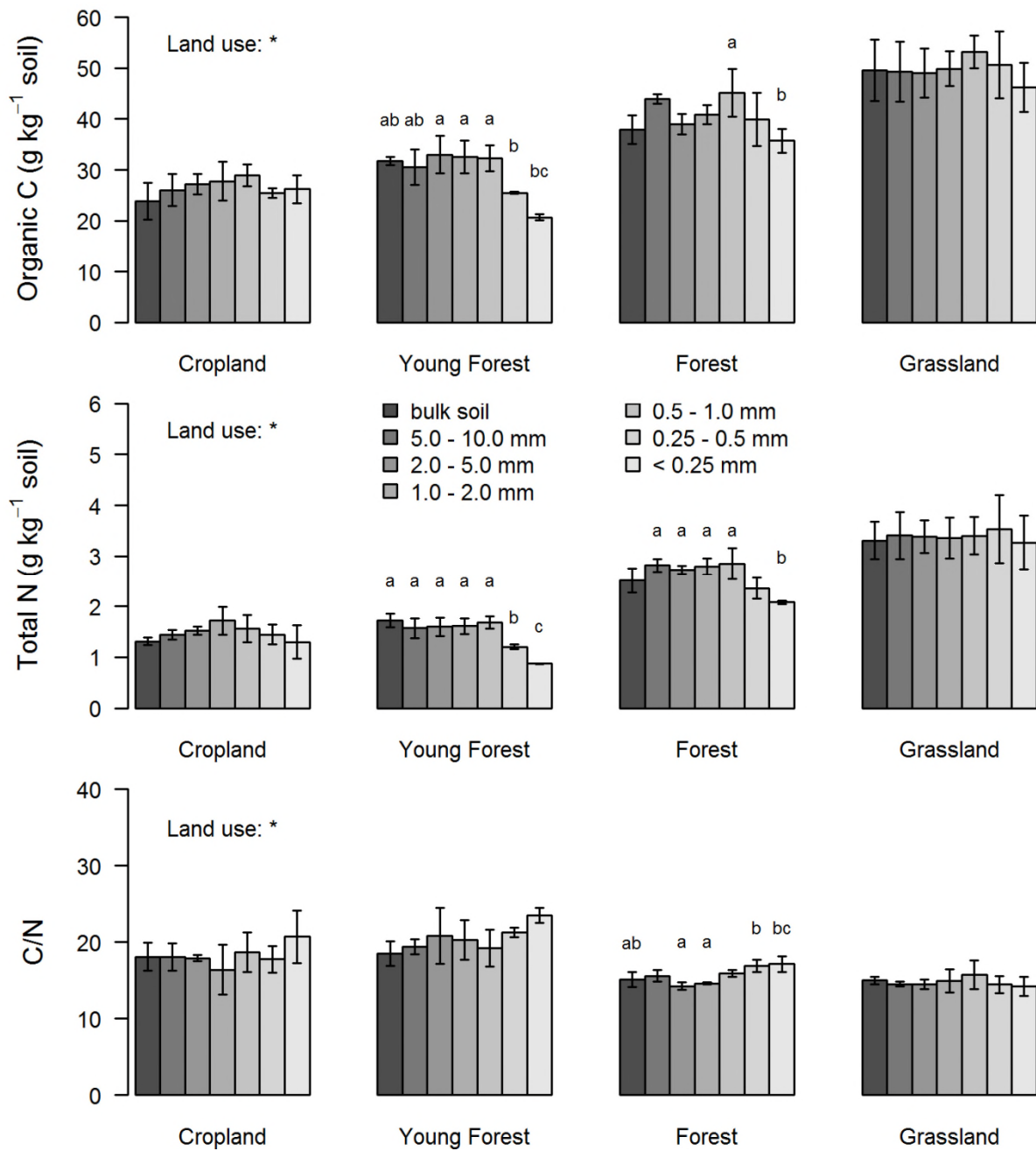


Fig. S3. Variation in organic C (g kg⁻¹ soil) and total N (g kg⁻¹ soil) concentration and C/N ratio between bulk soil and six soil aggregates sizes classes from four land use types. Mean value ± one standard deviation ($n = 3$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use.

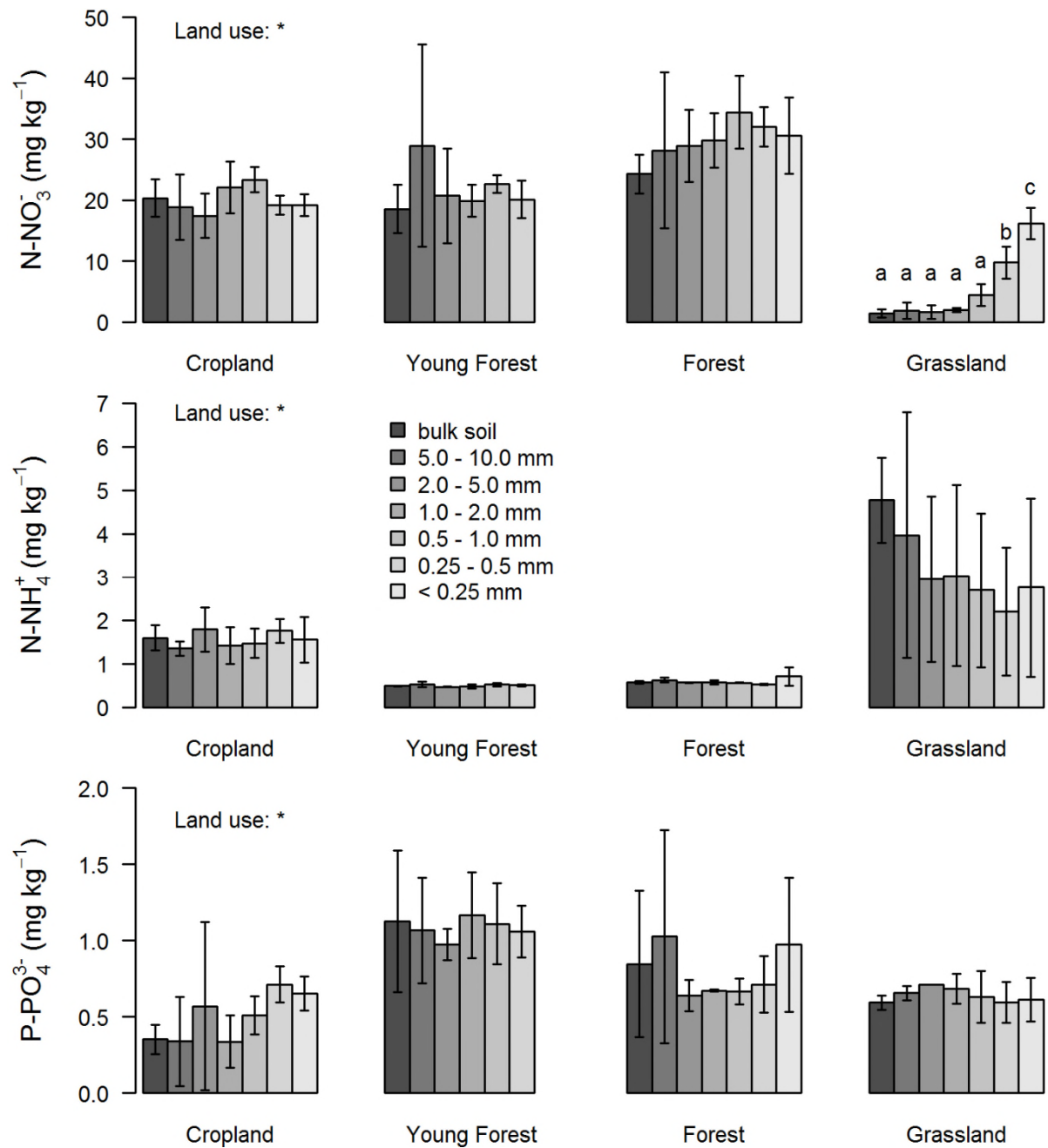


Fig. S4. Variation in N-NO₃⁻ (mg kg⁻¹ soil), N-NH₄⁺ (mg kg⁻¹ soil) and P-PO₄³⁻ (mg kg⁻¹ soil) concentrations between bulk soil and six soil aggregates sizes classes from four land use types. Mean value ± one standard deviation ($n = 3$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use. Small letters indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes within a specific land use. The N-NO₃⁻, N-NH₄⁺ and P-PO₄³⁻ concentrations were not measured on the < 0.25 mm aggregates from young forest site.

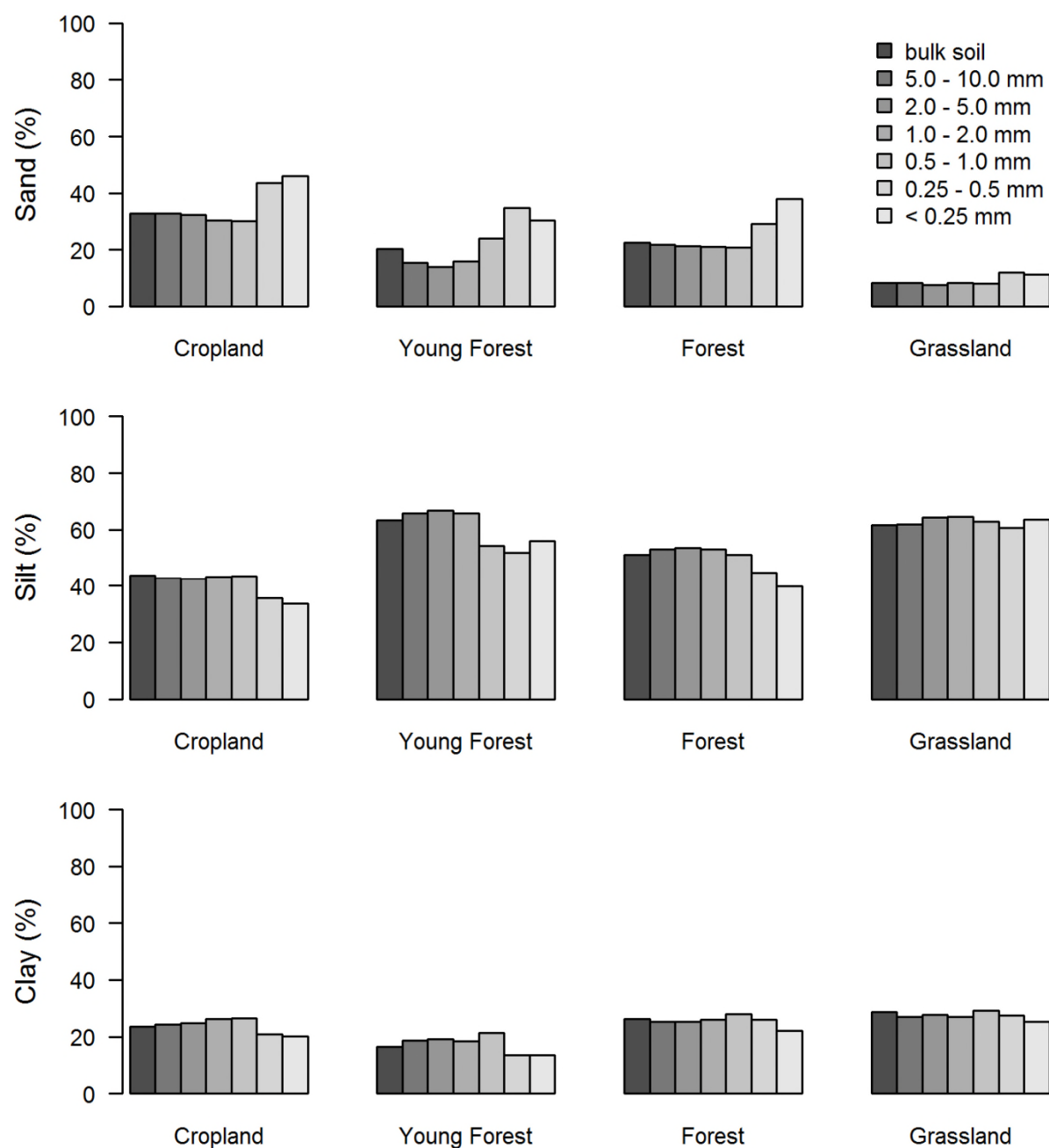


Fig. S5. Variation in sand, silt and clay contents (%) between bulk soil and six soil aggregates sizes classes from four land use types. The measurements were performed on one composite sample (mixture of 3 soil replicates).

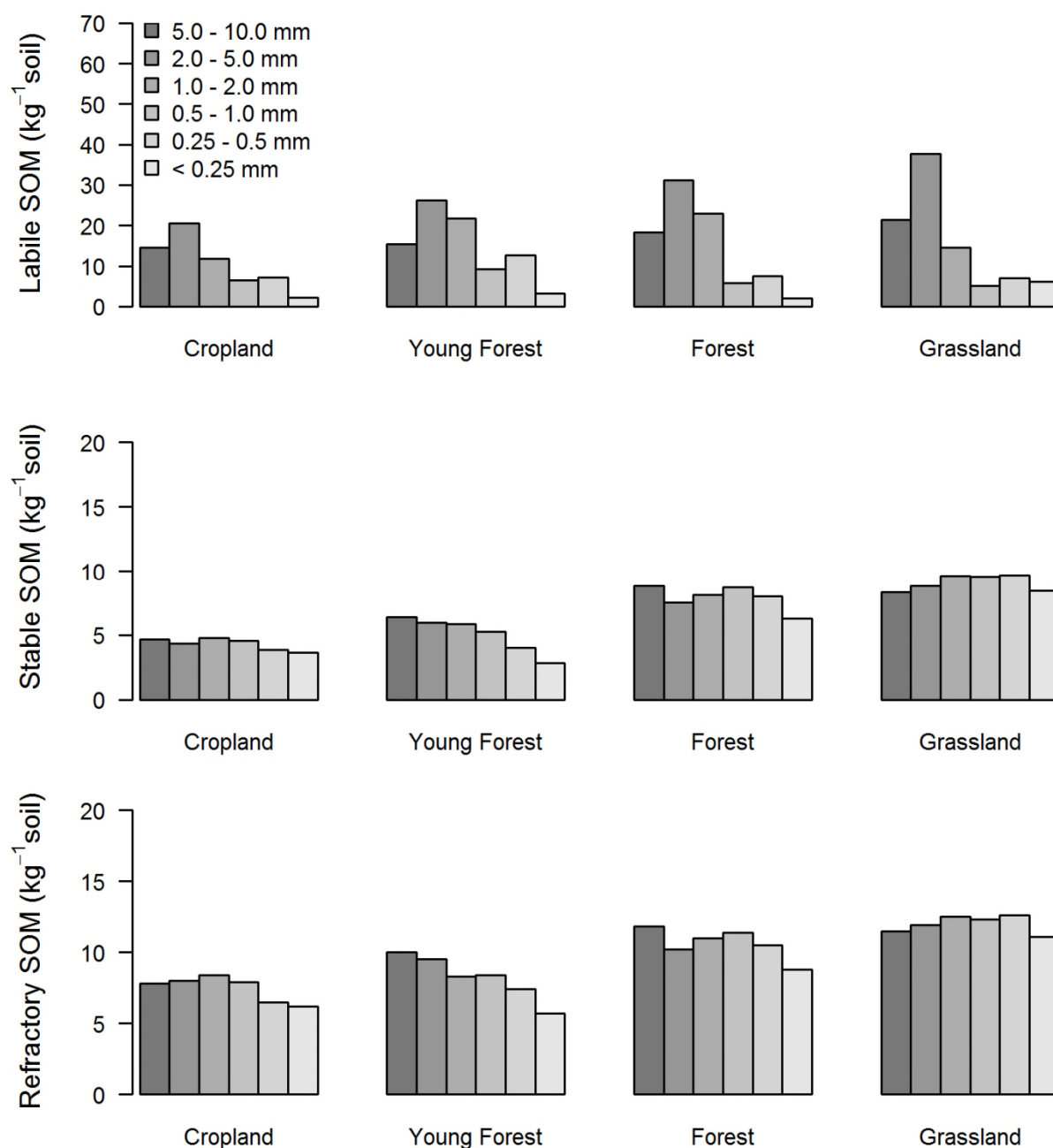


Fig. S6. Variation in labile, stable and refractory soil organic matter (SOM; g kg^{-1} soil) between bulk soil and six soil aggregates sizes classes from four land use types. The measurements were performed on one composite sample (mixture of 3 soil replicates).

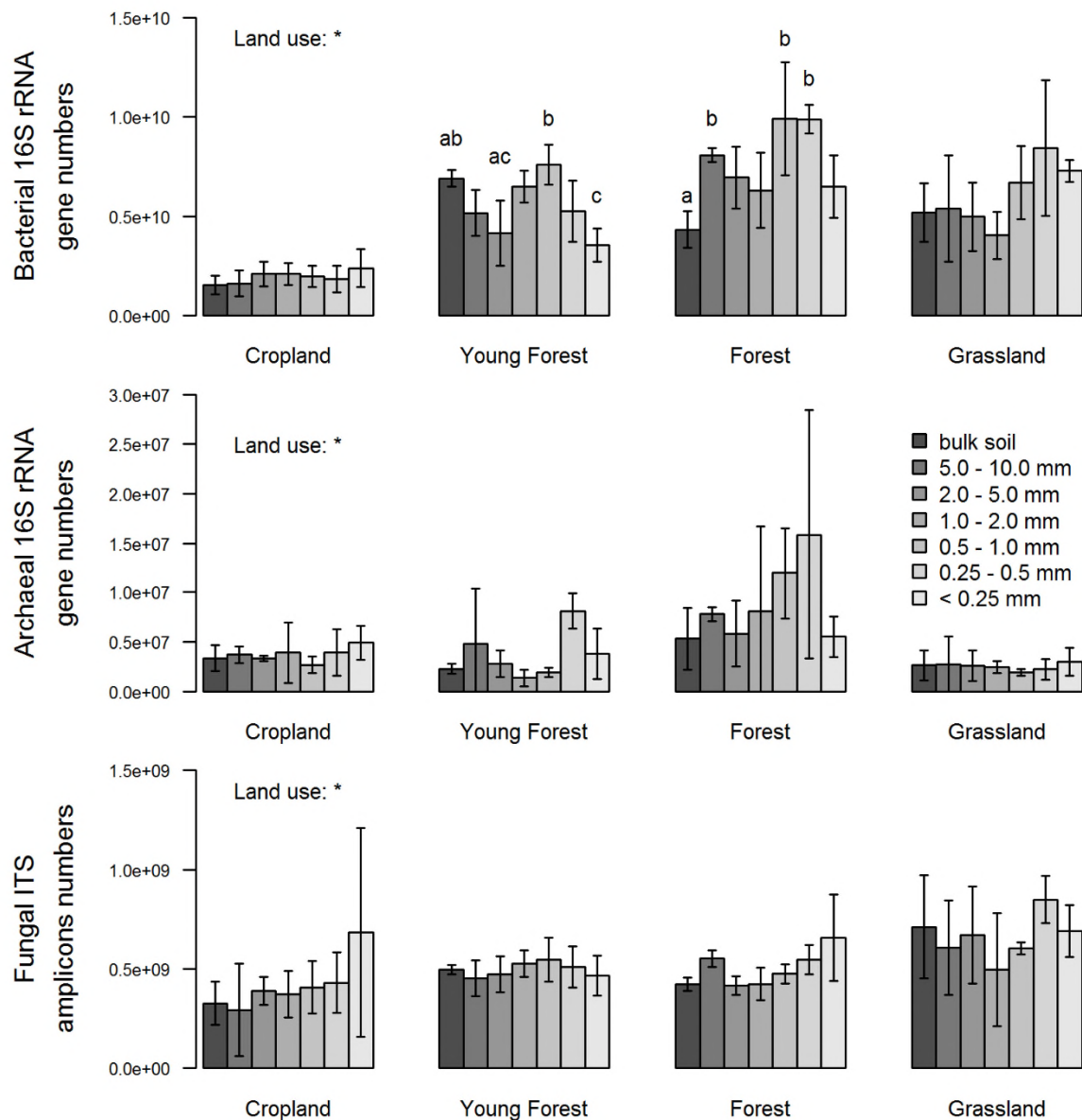


Fig. S7. Variation in gene abundance of bacteria and archaea (16S rRNA gene) and fungi (ITS amplicon) between bulk soil and 6 different soil aggregates sizes classes from 4 different land uses. The abundances of microbial communities are express by g^{-1} dry soil aggregates or by g^{-1} dry soil for the bulk soil. Means values \pm standard deviation ($n = 3$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use on microbial gene abundance. Different minusculer letter indicate significant ($P < 0.05$) differences between soil aggregates sizes classes for a specific land use.

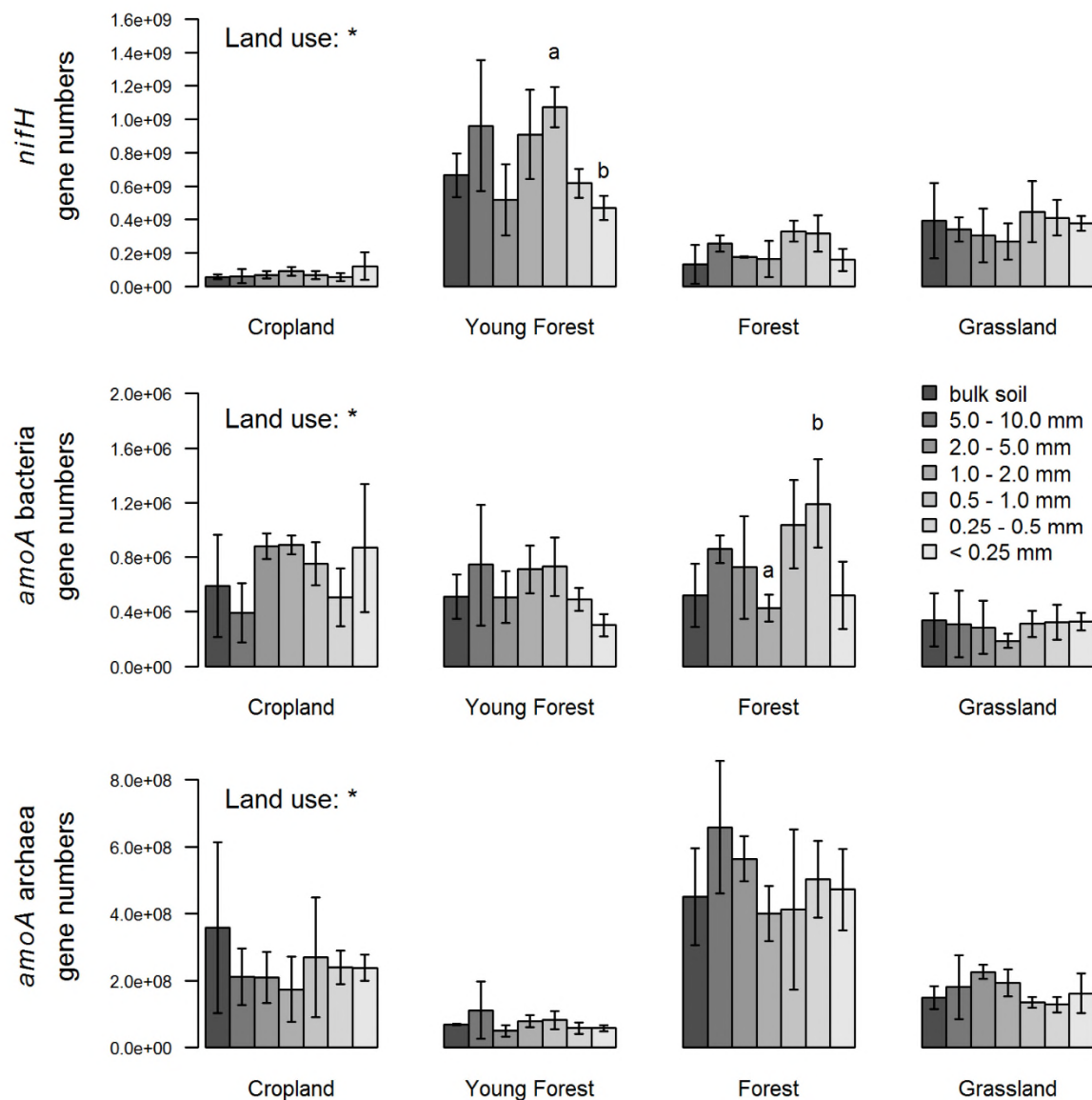


Fig. S8. Variation in gene abundance of N fixation (*nifH* gene) and ammonia oxidizing bacteria and archaea (*amoA* gene) between bulk soil and 6 different soil aggregates sizes classes from 4 different land uses. The abundances of microbial communities are express by g^{-1} dry soil aggregates or by g^{-1} dry soil for the bulk soil. Means values \pm standard deviation ($n = 3$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use on microbial gene abundance. Different minuscule letter indicate significant ($P < 0.05$) differences between soil aggregates sizes classes for a specific land use.

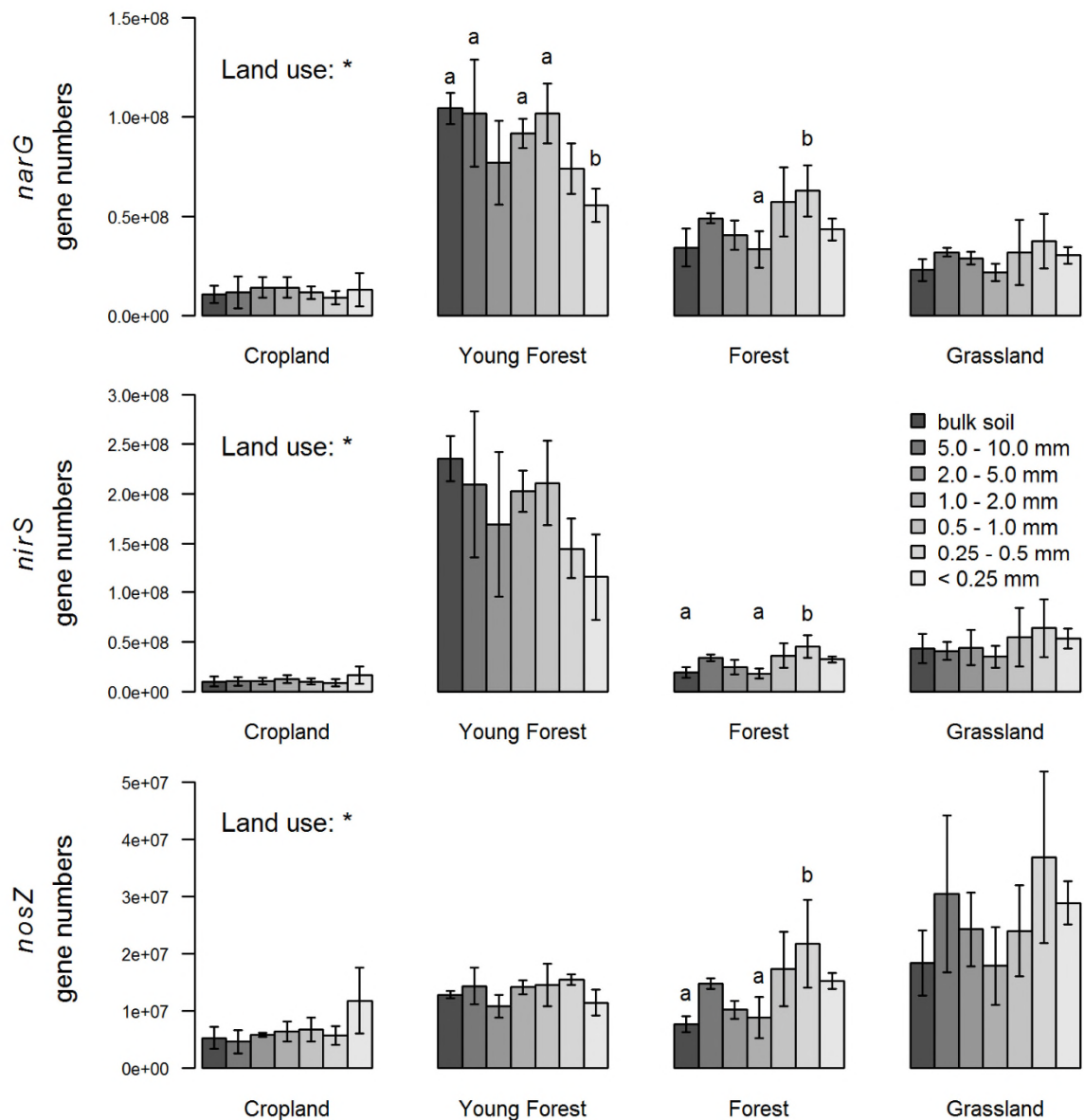


Fig. S9. Variation in gene abundance of nitrate reductase (*narG* gene), nitrite reductase (*nirK* gene) and nitrous oxide reductase (*nosZ* gene) between bulk soil and 6 different soil aggregates sizes classes from 4 different land uses. The abundances of microbial communities are expressed by g^{-1} dry soil aggregates or by g^{-1} dry soil for the bulk soil. Means values \pm standard deviation ($n = 3$; except for *nosZ* gene from cropland of the 1.0 – 2.0 mm soil aggregates, for which $n = 2$) are shown. Land use: * indicates significant ($P < 0.05$) effect of land use on microbial gene abundance. Different minuscule letter indicates significant ($P < 0.05$) differences between soil aggregates sizes classes for a specific land use.

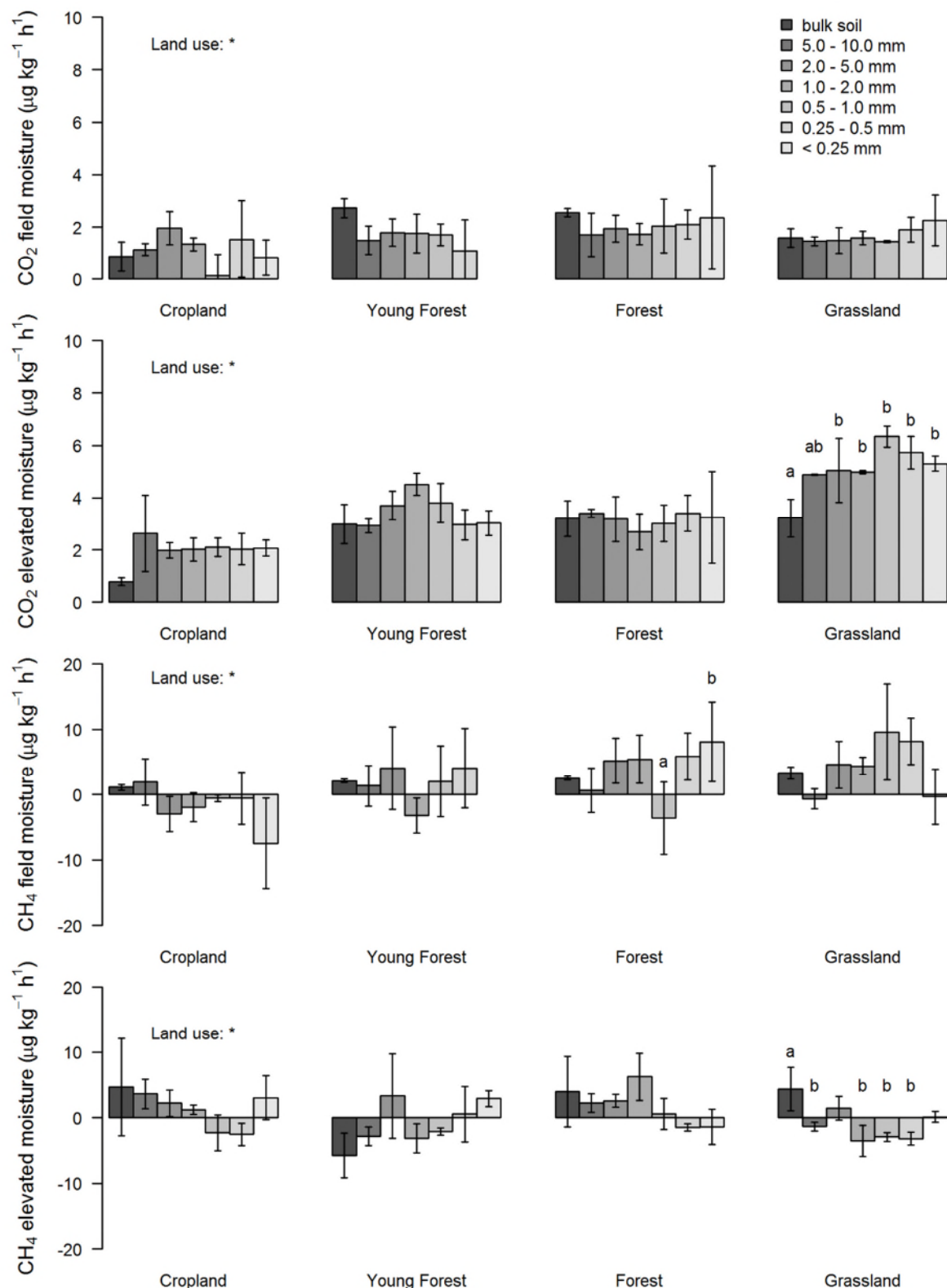


Fig. S10. Variation in CO₂ and CH₄ production (μg kg⁻¹ h⁻¹) between 6 sizes fractions and bulk soil, from 4 different land uses at the field moisture or elevated moisture (40 – 60 % of field capacity). Means values ± standard deviation (*n* = 3). Land use: * indicates significant (*P* < 0.05) effect of land use on microbial gene abundance. Different minuscule letter indicate significant (*P* < 0.05) differences between soil aggregates sizes for a specific land use. The CO₂ and CH₄ emissions were not measured for the < 0.25 mm soil fractions from young forest site at field moisture.

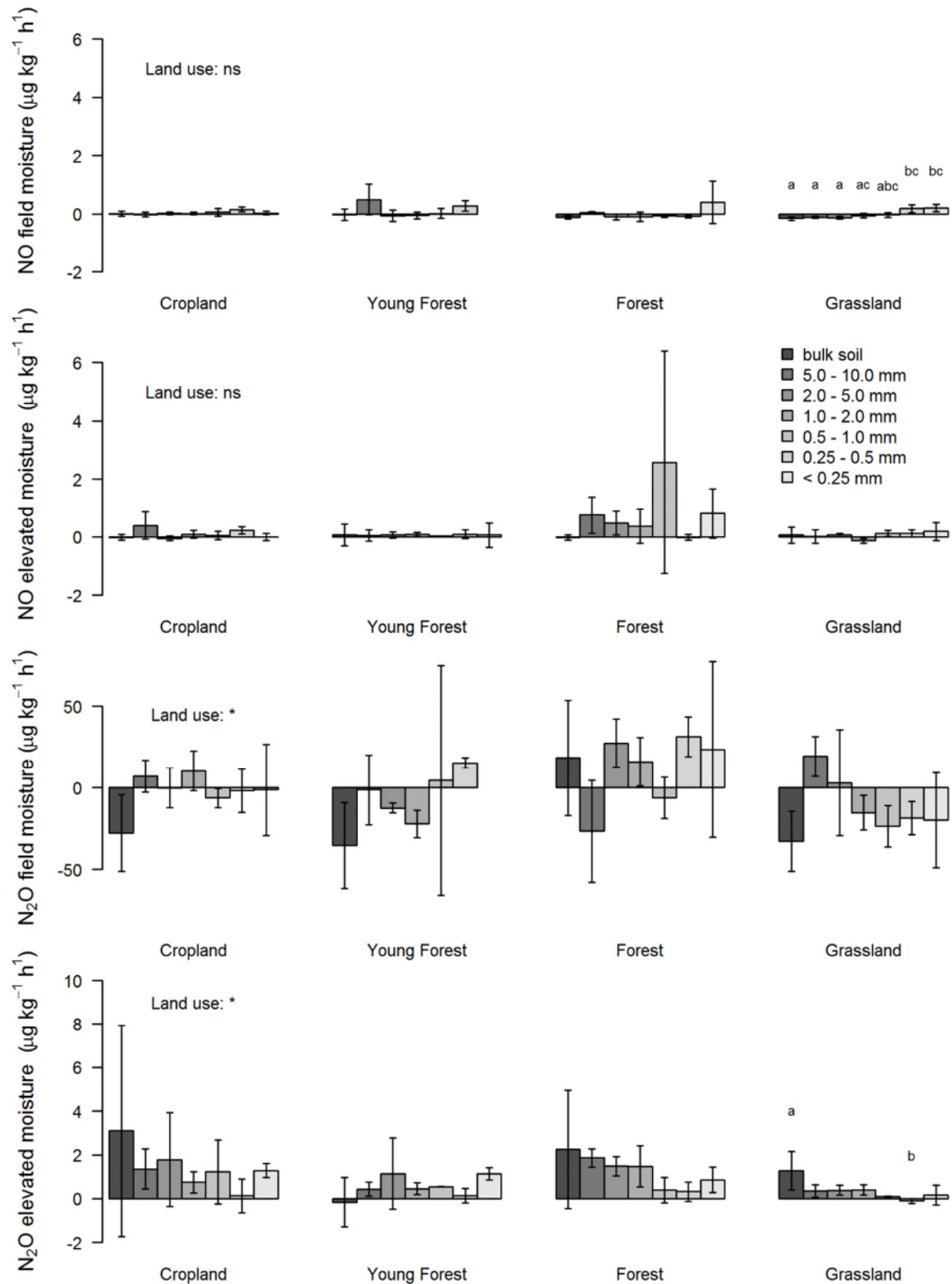


Fig. S11. Variation in NO and N₂O production (μg kg⁻¹ h⁻¹) between 6 sizes fractions and bulk soil, from 4 different land uses at the field moisture or elevated moisture (40 – 60 % of field capacity). Means values ± standard deviation (*n* = 3). Land use: indicates significant (*: *P* < 0.05) or no (ns: non-significant *P* > 0.05) effect of land use on microbial gene abundance. Different minuscule letter indicate significant (*P* < 0.05) differences between soil aggregates sizes for a specific land use. The NO and N₂O emissions were not measured for the < 0.25 mm soil fractions from young forest site at field moisture. NB: the y-scale of N₂O is different between plots based on field moisture or elevated soil moisture.

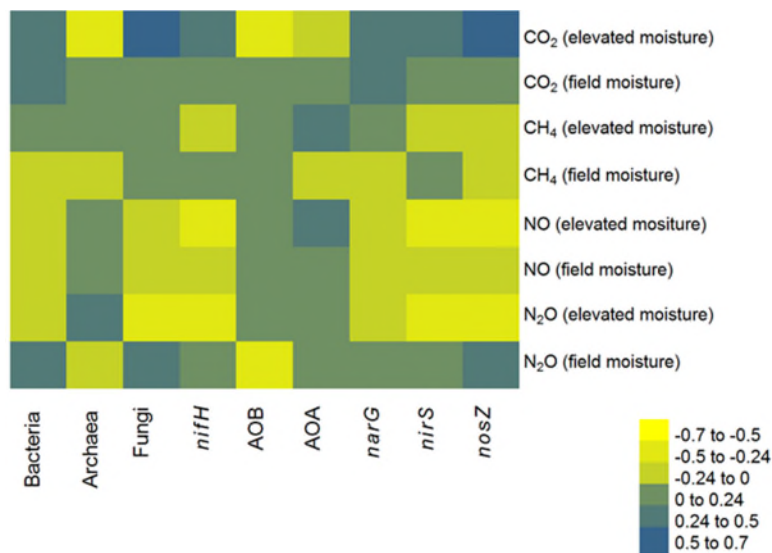


Fig. S12. Heatmaps of Spearman's rank correlation coefficients ρ between microbial genes abundance and greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25 , $0.25 - 0.5$, $0.5 - 1.0$, $1.0 - 2.0$, $2.0 - 5.0$ and $5.0 - 10.0$ mm) and four land uses. AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ values > 0.24 and < -0.24 are significant ($P < 0.05$).

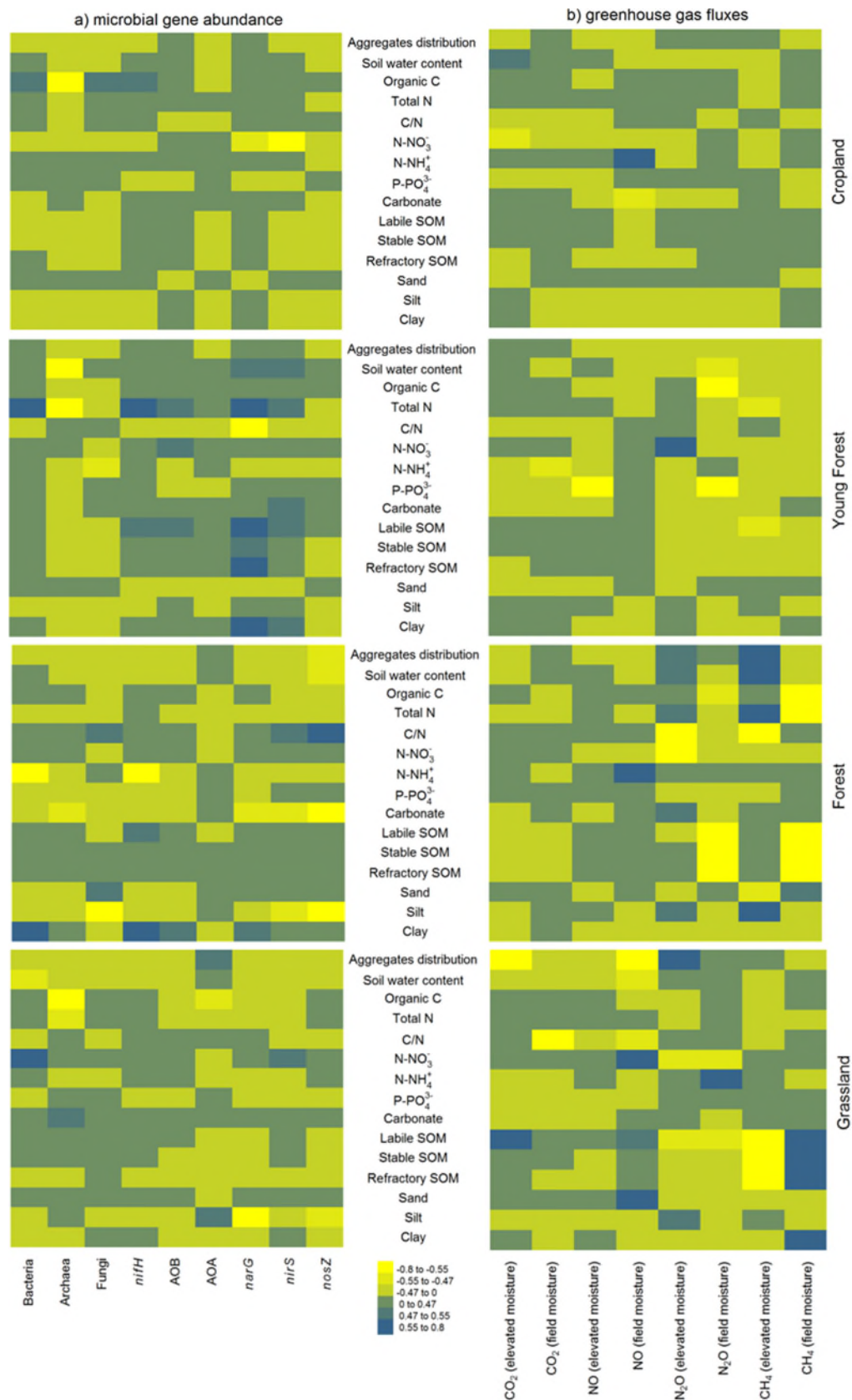


Fig. S13. Heatmaps of Spearman's rank correlation coefficients ρ between soil properties and a) microbial genes abundance or b) greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25, 0.25 – 0.5, 0.5 – 1.0, 1.0 – 2.0, 2.0 – 5.0 and 5.0 – 10.0 mm) and for four land uses separately. AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ values > 0.47 and < -0.47 are significant ($P < 0.05$).

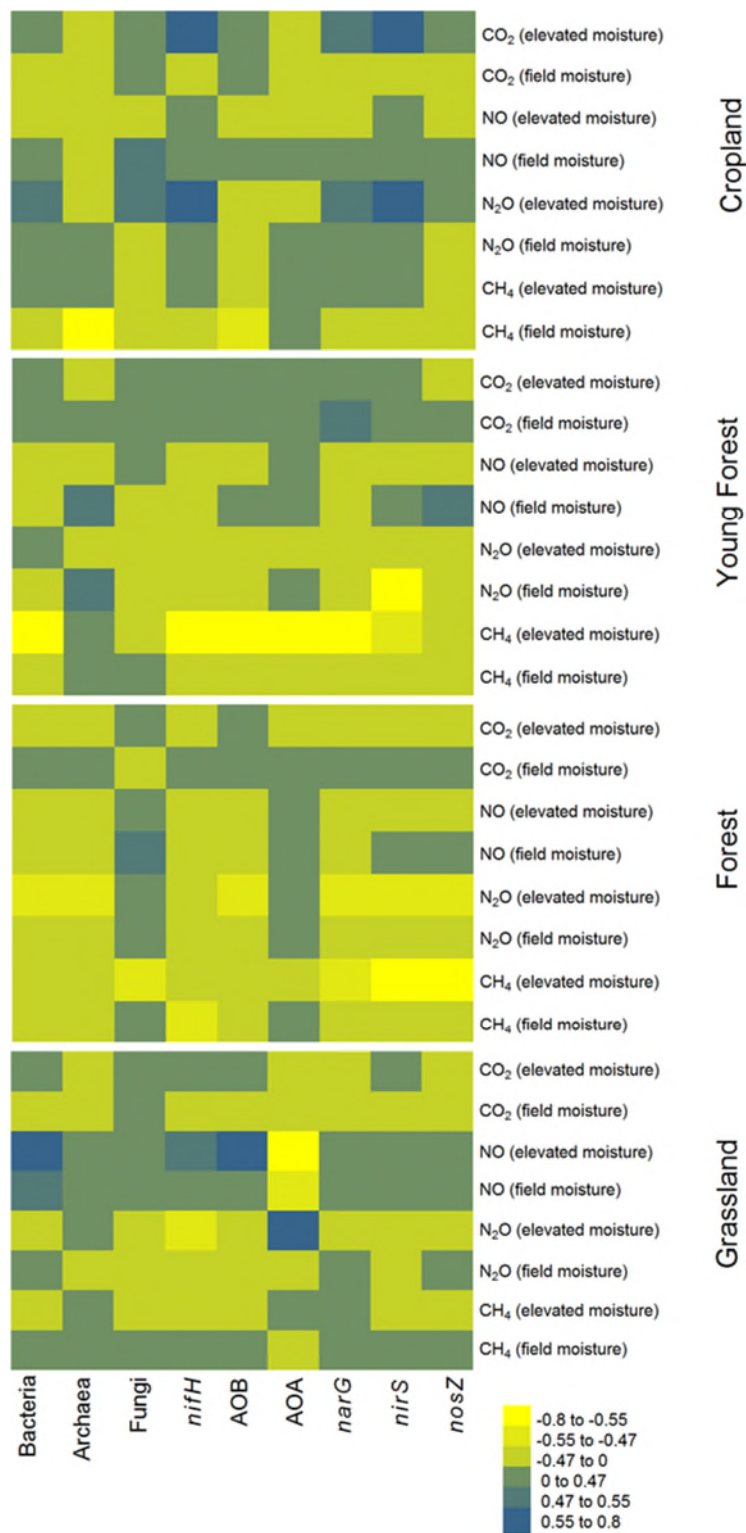


Fig. S14. Heatmaps of Spearman's rank correlation coefficients ρ between microbial genes abundance and greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25, 0.25 – 0.5, 0.5 – 1.0, 1.0 – 2.0, 2.0 – 5.0 and 5.0 – 10.0 mm) and for four land uses separately. AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ values > 0.47 and < -0.47 are significant ($P < 0.05$).

References:

- Barros, N., Salgado, J., Feijóo, S., 2007. Calorimetry and soil. *Thermochimica Acta*, XIVth ISBC Proceedings Special Issue Fourteenth conference of the International Society for Biological Calorimetry 458, 11–17.
- Braker, G., Fesefeldt, A., Witzel, K.-P., 1998. Development of PCR primer systems for amplification of nitrite reductase genes (*nirK* and *nirS*) to detect denitrifying bacteria in environmental samples. *Applied and Environmental Microbiology* 64, 3769–3775.
- De la Rosa, J.M., Knicker, H., López-Capel, E., Manning, D.A.C., González-Perez, J.A., González-Vila, F.J., 2008. Direct detection of black carbon in soils by Py-GC/MS, carbon-13 NMR spectroscopy and thermogravimetric techniques. *Soil Science Society of America Journal* 72, 258.
- Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., Oakley, B.B., 2005. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14683–14688.
- Gardes, M., Bruns, T.D., 1993. ITS primers with enhanced specificity for basidiomycetes - application to the identification of mycorrhizae and rusts. *Molecular Ecology* 2, 113–118.
- Henry, S., Bru, D., Stres, B., Hallet, S., Philippot, L., 2006. Quantitative detection of the *nosZ* Gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. *Applied and Environmental Microbiology* 72, 5181–5189.
- Holmes, A.J., Costello, A., Lidstrom, M.E., Murrell, J.C., 1995. Evidence that participate methane monooxygenase and ammonia monooxygenase may be evolutionarily related. *FEMS Microbiology Letters* 132, 203–208.
- Kitzler, B., Zechmeister-Boltenstern, S., Holtermann, C., Skiba, U.M., Butterbach-Bahl, K., 2006. Controls over N₂O, NO_x and CO₂ fluxes in a calcareous mountain forest soil. *Biogeosciences* 3, 383–395.
- Lane, D.J., 1991. *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons.

- Lopes-Capel, E., Sohi, S., Gaunt, J.L., Manning, D.A.C., 2005. Use of thermo gravimetry-differential scanning calorimetry to characterize soil organic matter fractions. *Soil Science Society of America Journal* 69, 136–140. doi:10.2136/sssaj2005.0136
- López-Gutiérrez, J.C., Henry, S., Hallet, S., Martin-Laurent, F., Catroux, G., Philippot, L., 2004. Quantification of a novel group of nitrate-reducing bacteria in the environment by real-time PCR. *Journal of Microbiological Methods* 57, 399–407.
- Manter, D.K., Vivanco, J.M., 2007. Use of the ITS primers, ITS1F and ITS4, to characterize fungal abundance and diversity in mixed-template samples by qPCR and length heterogeneity analysis. *Journal of Microbiological Methods* 71, 7–14. doi:10.1016/j.mimet.2007.06.016
- Miranda, K.M., Espey, M.G., Wink, D.A., 2001. A Rapid, simple spectrophotometric method for simultaneous detection of nitrate and nitrite. *Nitric Oxide: Biology & Chemistry* 5, 62–71.
- Okano, Y., Hristova, K., Leutenegger, C., Jackson, L.E., Denison, R.F., Gebreyesus, B., Lebauer, D., Scow, K.M., 2004. Application of real-time PCR to study effects of ammonium on population size of ammonia-oxidizing bacteria in soil. *Applied and Environmental Microbiology* 70, 1008–1016.
- Olsen, S., Cole, C.V., Dean, L.A., Watanabe, F.S., 1954. Estimation of available phosphorus in soils by extraction with sodium bicarbonate. (USDA Circular No. 939). Washington, D.C.: U.S. Government Printing Office.
- Ririe, K.M., Rasmussen, R.P., Wittwer, C.T., 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry* 245, 154–160.
- Rösch, C., Bothe, H., 2005. Improved assessment of denitrifying, N₂-fixing, and total-community bacteria by terminal restriction fragment length polymorphism analysis using multiple restriction enzymes. *Applied and Environmental Microbiology* 71, 2026–2035.
- Rowland, A.P., 1983. An automated method for the determination of ammonium-n in ecological materials. *Communications in Soil Science and Plant Analysis* 14, 49–63.

- Schindlbacher, A., Zechmeister-Boltenstern, S., Butterbach-Bahl, K., 2004. Effects of soil moisture and temperature on NO, NO₂, and N₂O emissions from European forest soils. *Journal of Geophysical Research: Atmospheres* 109, 1–12. doi:10.1029/2004JD004590
- Smith, C.J., Nedwell, D.B., Dong, L.F., Osborn, A.M., 2006. Evaluation of quantitative polymerase chain reaction-based approaches for determining gene copy and gene transcript numbers in environmental samples. *Environmental Microbiology* 8, 804–815.
- Soil Survey Staff, 2004. Keys to soil taxonomy, 10th edn, USDA – NRCS. ed. Washington, DC.
- Stubner, S., Meuser, K., 2000. Detection of *Desulfotomaculum* in an Italian rice paddy soil by 16S ribosomal nucleic acid analyses. *FEMS Microbiology Ecology* 34, 73–80.
- Tabatabai, M.A., Bremner, J.M., 1991. Automated instruments for determination of total carbon, nitrogen, and sulfur in soils by combustion techniques, in: Smith, K.A. (Ed.), *Soil Analysis*. New York, p. pp 261-286.
- Tsiknia, M., Tzanakakis, V.A., Paranychanakis, N.V., 2013. Insights on the role of vegetation on nitrogen cycling in effluent irrigated lands. *Applied Soil Ecology* 64, 104–111.
- Vetriani, C., Jannasch, H.W., MacGregor, B.J., Stahl, D.A., Reysenbach, A.-L., 1999. Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. *Applied and Environmental Microbiology* 65, 4375–4384.
- Vilgalys, R., Hester, M., 1990. Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *Journal of Bacteriology* 172, 4238–4246.