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

Aimeric Blaud[a, 1]*, Bas van der Zaan[b], Manoj Menon[a, 2], Georg J. Lair[c, d], Dayi Zhang[a, 3], Petra Huber[c], Jasmin Schiefer[c], Winfried E.H. Blum[c], Barbara Kitzler[e], Wei E.Huang[a ,4], 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: Department of Geography, University of Sheffield, Winter street, Sheffield S10 2TN, United Kingdom.

[3] Current address: School of Environment, Tsinghua University, Beijing 200084, PR China.

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

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 uses (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 consistently showed 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 bacterial 16S, *nifH*, *amoA* bacteria, *narG*, *nirS* and *nosZ* gene abundance for the two forest sites but not for fungal ITS and archaeal 16S. 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, few differences in GHG fluxes were found between soil aggregate 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.

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

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 (Amézketa, 1999; Bronick and Lal, 2005) and some 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 has been found to vary with the size of soil aggregates, and to be linked to the specific environmental conditions in the different sizes of aggregates. 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 has been 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 CO_2 emissions found to be higher in microaggregates (< 0.25 mm) than in macroaggregates (> 0.25 mm) in cropland sandy loam soil (Sey et al., 2008; Mangalassery et al., 2013). Similar results were found for CH_4 in cropland 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

82 studies have investigated specific microbial functional guilds such as N fixation (Mendes and
83 Bottomley, 1998; Poly et al., 2001; Chotte et al., 2002; Izquierdo and Nüsslein, 2006) and
84 denitrifiers (Beauchamp and Seech, 1990; Lensi et al., 1995) in soil aggregates. The biomass and
85 composition of diazotrophs varies with the size of soil aggregates which was correlated with
86 total C and N, and soil texture (Poly et al., 2001; Izquierdo and Nüsslein, 2006). Aggregates
87 within size classes 0.6 – 2.0 mm and < 0.075 mm (from tundra, pasture and forest) were found
88 to have the highest diazotroph richness (Izquierdo and Nüsslein, 2006) and microaggregates (<
89 0.25 mm) to host between 30% and 90% of the diazotrophic population (Mendes and
90 Bottomley, 1998; Chotte et al., 2002). In contrast, denitrifiers were found to occur mainly in
91 microaggregates, where nearly 90% of the potential denitrification activity can occur (Lensi et
92 al., 1995). Hence, the diazotroph and denitrifier communities seem to exploit specific and
93 different anaerobic niches within different soil aggregate size classes, although the drivers of
94 these communities in different soil aggregate sizes remains unclear.
95 The type of land use and management directly influences the physico-chemical
96 properties of soil aggregates as well as the distribution of microbial communities, their
97 functions and resulting nutrient transformations and GHG fluxes. For example, the soil
98 aggregates turnover rate is increased by soil tillage (Six et al., 2004), which decreases the C
99 storage within the aggregates (Bossuyt et al., 2002), but can also decrease N_2O fluxes (Ball,
100 2013). Furthermore, the type of vegetation and input of organic manure influence the aggregate
101 size distribution and the contents of organic C and N within soil aggregates (Pinheiro et al.,
102 2004; Six et al., 2004; An et al., 2010). Subsequently, bacterial and fungal community
103 composition was found to differ between land use types (Lauber et al., 2008) and also microbial
104 activity such as nitrification (Hayden et al., 2010).
105 The above leads to the overarching hypothesis that in conjunction with land use,
106 different microbial functions are preferentially hosted or fostered by specific size classes of
107 aggregates. The specific objectives of the current study were: i) to assess the difference in
108 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 day yr^{-1}) named "young forest", and sites disconnected from the river through a flood control dike: forest, grassland and cropland. The young forest is impacted by flood events, and covered by "soft-wood" dominated by *Salicetum* albae, while the forest site is covered by "hard-wood" dominated by *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. Cropland site was conventionally managed, with annual tillage and NPK mineral fertilisers. The field is under crop rotation (maize, sugar beet, barley and wheat), with summer wheat the year of the sampling which was shortly harvested before the soil sampling. 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 . The soils are classified as Epigleyic Fluvisol (young forest) and Mollic Fluvisols (forest, grassland and cropland; (IUSS Working Group WRB, 2014). The Epigleyic Fluvisol is at least one time of the year impacted by

groundwater and is located close to the Danube River. In contrast, the Mollic Fluvisols have no impact of groundwater and are characterized by a fast OC accumulation in the topsoil. In our study area Mollic Fluvisols develop towards a Chernozem.

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 to avoid the main rooting zone in grassland and the litter layer in forest sites, focusing on the similar mineral soil layer across sites. The soil samples were manually dry sieved to obtain 6 soil aggregate 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 knowing that the sieving method affects the gene abundance quantification, dry-sieving can nonetheless reveal differences in gene abundance between soil aggregate 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 conditions 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 CO_2 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 CO_2 analyser, and NO was measured with a HORIBA APNA-360 (Kyoto, Japan) chemoluminescence NO_x analyser. Determination of N_2O and CH_4 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 ^{63}Ni-electron-capture detector and CH_4 by a flame ionization detector.

2.5. Physico-chemical analysis of bulk soil and aggregates

The soil moisture content, organic C, total N, $N\text{-}NO_3^-$, $N\text{-}NH_4^+$, $P\text{-}PO^{3-}_4$, and carbonate concentration, C/N, and soil texture (i.e. sand, silt and clay contents) were measured for each aggregate size class 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 comply with 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 aggregate 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 a factor (df = 6) for each land use separately, insuring conditions were met as described 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 aggregate size classes. To display the correlations, heatmaps were constructed using the library "gplots" from R software, were 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 least (< 10%; Fig. S2). Young forest and forest showed significantly higher soil water content for most soil aggregate 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-NO_3^- concentration for soil aggregates > 0.5 mm (~10 times) than the other sites, but significantly higher N-NH_4^+ for the bulk soil (~5 times) and some soil aggregates (Fig. S4). The P-PO^{3-}_4 in cropland was significantly lower than the other sites in aggregates 1 – 2 mm, while in young forest P-PO^{3-}_4 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-NO_3^- concentrations than other soil aggregate sizes or bulk soil (Fig. S4). The sand content was higher in cropland and lower in grassland and was higher in aggregate size classes < 0.5 mm regardless of the land use (Fig. S5). In contrast, the silt content was lower in cropland and higher in grassland, while clay content was lower in young forest. Both silt and clay contents tend to decrease in aggregate size classes < 0.5 mm. The different fractions of SOM were lower in cropland and higher in grassland, while labile SOM was higher in aggregate size classes 2 -5 and 1 -2 mm and stable and refractory SOM both tend to decrease in aggregate size classes < 0.5 mm (Fig. S6).

3.2. Variation in microbial gene abundance between land uses and soil aggregate size classes

All microbial gene abundances investigated showed significant differences between land use types for at least one soil aggregate size class or bulk soil (Fig. 1, Fig. S7-S9, Table S2). The cropland site consistently (i.e. across bulk soil and soil aggregates) 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-S9). 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 between land uses and soil aggregate size classes

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 aggregate size, except for NO at field

moisture (Fig. S10, S11). The CO_2 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, CO_2 emissions were consistently significantly lower in cropland compared to grassland sites regardless of the aggregates size classes and bulk soil (Fig. 3, S10). Overall, the CO_2 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. 3) and overall significant differences between land use types for a few specific aggregate size classes such as < 0.25 (CH_4 elevated moisture), 0.25 – 0.5 (NO, N_2O soil moisture), 1.0 – 2.0 (CH_4 both moisture levels and N_2O field moisture), 5.0 – 10.0 mm (CH_4 and N_2O elevated moisture) (Fig. S10, S11).

Within the separate land use types, significant effects of aggregate size at field moisture were only observed for CH_4 at the forest site and for NO at the grassland site. The 0.5 – 1.0 mm aggregates acted as a sink for CH_4 at field moisture while the other aggregates classes were sources of CH_4 (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 CO_2 emissions than the aggregates size classes, while it was a source of CH_4 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). The *narG, nirS* and *nifH* gene abundances showed significant positive correlations with silt and carbonate contents and P-PO^{3-}_4 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, $N\text{-}NO^{3-}$ concentration and sand content (Fig. 5a). The CO_2 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 CO_2 and CH_4 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 CO_2 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 aggregate 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 aggregate 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 CH_4 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 of the gene investigated, gene abundances were always affected by the different types of land use. The different types of land use and management were previously found to affect the abundance of microorganisms (Enwall et al., 2010; Hallin et al., 2009; Lauber et al., 2008; Leininger et al., 2006; Ma et al., 2008; Morales et al., 2010; Wallenstein and Vilgalys, 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 have a direct and negative effect on the biomass of bacteria and fungi (Muruganandam et al., 2009; Helgason et al., 2010), and also on n*arG* 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 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 of 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. 5a). Thus, it further supports the idea that despite AOA and AOB delivering the same function, the two communities live in different niches/microhabitats with specific environments stimulating their activity separately (Prosser and Nicol, 2008). Low soil pH and low NH_4^+

concentration were found to be important conditions favouring *amoA* archaea abundance while the contrary was found for *amoA* bacteria (Leininger et al., 2006; Verhamme et al., 2011). However, in the current study the soil pH was above 7 and both bacterial and archaeal *amoA* showed strong positive correlation with NO_3^- and NH_4^+ for archaea, showing that these drivers are not the only ones responsible for niche differentiation of *amoA.* Hence, the quantity and quality of SOM might play an important role in the studied soil, as organic C can differently inhibit or stimulate ammonia oxidizer (Erguder et al., 2009).

The community showing the highest abundance in young forest (i.e. *nifH*, *narG* and *nirS* genes) showed a strong and positive correlation to phosphate concentration which was higher in the young forest and could be a limiting factor in the other land use (Table 1, Fig. S3). Their high abundance could also be related to the location of the site, with a slightly different soil type (Epigleyic Fluvisol for young forest and Mollic Fluvisols for the other sites) which is also younger (70 yr against 250-350 yr). Furthermore, the site is located along the Danube River, subjected to flood (~10 days yr^{-1}), creating anaerobic conditions over long period of time that would favour the denitrification and N fixation processes. In contrast, the other sites are protected from flood by a dike. 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 by bacteria, 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 existence of *nosZ* in bacteria lacking other genes for denitrification and the different

environment required to perform the different steps of denitrification. Thus, there is a niche differentiation of the different steps of the denitrification, with SOM quantity and quality (directly related to the plant residues input and root exudates) playing a key role for *nosZ* gene abundance, while *narG* and *nirS* genes were both regulated by the P, carbonate and silt concentration.

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

Soil aggregate size was a minor factor in explaining nitrogen genes abundance, compared to land use. The effects of soil aggregate size classes on gene abundances was specific to the land use type and not present for all genes or land uses studied. Neuman et al. (2013) found that the size of soil aggregates was the dominant factor in the abundance of bacterial, archaeal and fungal community, over soil management (i.e. fertilisation). However, they investigated microaggregates (0.002 – 0.020 mm, 0.020 – 0.063 mm, > 0.063 mm) and the silt and clay fractions (< 0.002 mm), which could physically protect organisms against environmental changes. Hence, the current study shows that the sizes of macroaggregates are not the main factor determining microbial distribution and N functional guilds after land use type, whereas aggregates < 0.063 mm 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 higher due to anthropogenic activity such as tillage and plant harvest (Blaud et al., 2014; Six et al., 2002, 2000; Tisdall and Oades, 1982). The lower variation in microbial abundance between soil aggregate size fractions in grassland in comparison to young forest and forest, might be explained by a high organic matter input due to fine grass root system and root exudates, resulting in the

highest SOC and total N concentration in comparison to the other land uses, and no significant difference in their concentrations between grassland aggregate sizes classes (Fig. S3). 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, as well as a different quantity and quality of plant input that affected SOM concentration in soil aggregate size classes (Fig. S6). Overall, specific drivers for each land use are responsible of the distribution of gene abundance in soil aggregates, such as total N and labile SOM that explained bacteria, *nifH*, AOB, *narG* and *nirS* genes distribution for young forest, while soil texture, especially clay content, was explaining most gene distribution in forest. In contrast, for cropland and grassland organic C and silt content respectively, explained few genes distribution.

At the forest and young forest sites, the size of soil aggregates was an important factor in the abundance of several microbial communities 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 consistently showed 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 aggregate 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 preferentially colonised similar niches, which differ in 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 to a lesser extent 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 revealed differences compared to the microbial gene abundance. Hence, the effect of land use on the bulk soil samples were mainly found for CO_2 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 size classes compared to the bulk soil may be linked to different porosity present for each size and how land use affects 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 CO_2 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, but also strong correlations with most genes abundance. The other GHG fluxes showed inconsistent effect of land use depending on soil moisture and soil aggregate size, highlighting the complexity of drivers of CH_4, NO and N_2O fluxes. Only few correlations were found between CH_4, NO and N_2O fluxes and genes abundance, showing the difficulty to relate gene abundance and GHG fluxes, due to the high variability of GHG fluxes and possible dissimilarity between genes and activity.

Change in soil moisture had significant effects on GHG fluxes, although it varies between GHG, land use, and soil aggregate size classes. Higher CO_2 emissions were consistently found at elevated soil moisture compared to field moisture across all land use, highlighting the importance of soil moisture for microbial activity and CO_2 emissions (Sey et al., 2008). For CH_4, NO and N_2O the effect of increased soil moisture was not as consistent as for CO_2, indicating that other factors limit their fluxes. Surprisingly, increasing soil water content in the current study

483 did not necessarily increase the CH_4 production, as might be expected because methanogens are
484 more active in high water content/anaerobic soils. The CH_4 was either emitted or consumed
485 depending on the soil water content for a specific land use and soil aggregate size class. This
486 indicates that both methanogens and methane-oxidizing bacteria were present in the same soil
487 aggregates as previously found by Sey et al. (2008) and can co-exist in the same niche. Similarly,
488 increasing soil water content did not increase the anaerobic process of denitrification
489 responsible for NO and N_2O fluxes, indicating that other factors are regulating these fluxes and
490 the microorganisms responsible, or the increase in soil water content was not enough to reach
491 anaerobic conditions.

492 Overall, the GHG fluxes did not occur in a specific aggregate size class within a land use
493 as found for microbial gene abundances in forest sites. Previous studies found higher CO_2
494 emissions in microaggregates whilst acting as sinks of CH_4 (Sey et al., 2008). However, CO_2
495 emissions were also shown to be highly sensitive to water filled pore space (WFPS), with no
496 difference in emissions between aggregate size at 60% WFPS; microaggregates acted as sinks of
497 CH_4 at 20% WFPS but a source at higher WFPS (Ramakrishnan et al., 2000; Sey et al., 2008).
498 However, in the current study, elevated soil moisture did not reveal more significant differences
499 than at soil moisture in GHG fluxes between soil aggregates, indicating that other factors may
500 drive differences or that the size of soil aggregate may not be an important driver for GHG fluxes.
501

502 **5. Conclusions**

503 This study demonstrates that land use is the main factor in explaining abundance of
504 nitrogen genes and greenhouse gas fluxes, while soil aggregate size class was a minor factor.
505 This goes against our initial hypothesis suggesting that different microbial functions are
506 preferentially hosted or fostered by specific size of aggregates. This is due to the stronger
507 difference in soil physico-chemical characteristics between land use types than between soil
508 aggregate sizes. Cropping had a clear negative effect on the abundance of most microbial
509 communities, likely due to the depletion of SOC and total N by tillage, plant harvest, and soil

510 erosion. Although soil aggregate size was not a dominant factor, it affected the distribution of the
511 N functional communities at the semi-natural forest sites, showing that some microbial
512 functions are probably related to specific microhabitats (i.e. the architecture and distribution of
513 pores filled with water and air, the availability of organic matter and other nutrients) in soil,
514 where anthropogenic activity is limited, allowing differences between microhabitats to develop.
515 However, no specific size of soil aggregates enhanced the abundance of any specific microbial
516 function across all four land uses. Soil aggregate size had little effect on GHG fluxes, indicating
517 that the size of soil aggregates may not have much effect on GHG fluxes but it also highlights the
518 difficulties of measuring GHG fluxes in aggregates.
519 This study only addresses a single point in time, limiting our understanding of the
520 distribution of microbial functions over soil aggregates of different size. Further studies are
521 needed, taking into consideration the dynamics of soil aggregates and its relation with microbial
522 communities by sampling at multiple time points, work on a wider range of aggregate size
523 classes (e.g. size classes < 0.25 mm) and land use types. Furthermore, combining microbiology
524 and soil architecture (e.g. x-ray tomography) as well as nutrient availability in local and time
525 scale, would fully reveal the physical distribution of microhabitats, the microbial communities
526 and functions among soil aggregates. Comparing microbial functions between soil aggregates of
527 varying size from a specific land use (e.g. forest) but from different locations or soil types may
528 also provide more insight into the role of soil aggregates in microbial functioning.
529

530 ## Acknowledgements

531 This work was supported by the European Commission 7th Framework Program as a Large
532 Integrating Project, SoilTrEC (www.soiltrec.eu), Grant Agreement No. 244118.
533

534 ## References

535 Amézketa, E., 1999. Soil aggregate stability: a review. J. Sustain. Agr. 14, 83–151.

536 An, S., Mentler, A., Mayer, H., Blum, W.E.H., 2010. Soil aggregation, aggregate stability, organic
537 carbon and nitrogen in different soil aggregate fractions under forest and shrub
538 vegetation on the Loess Plateau, China. CATENA 81, 226–233.
539 Ball, B.C., 2013. Soil structure and greenhouse gas emissions: a synthesis of 20 years of
540 experimentation. Eur. J. Soil Sci. 64, 357–373.
541 Banwart, S., 2011. Save our soils. Nature 474, 151–152.
542 Barros, N., Salgado, J., Feijóo, S., 2007. Calorimetry and soil. Thermochimica Acta, XIVth ISBC
543 Proceedings Special Issue Fourteenth conference of the International Society for
544 Biological Calorimetry 458, 11–17.
545 Beauchamp, E.G., Seech, A.G., 1990. Denitrification with different sizes of soil aggregates
546 obtained from dry-sieving and from sieving with water. Biol. Fertil. Soils 10, 188–193.
547 Blaud, A., Lerch, T.Z., Chevallier, T., Nunan, N., Chenu, C., Brauman, A., 2012. Dynamics of bacterial
548 communities in relation to soil aggregate formation during the decomposition of ^{13}C-
549 labelled rice straw. Appl. Soil Ecol. 53, 1–9.
550 Blaud, A., Chevallier, T., Virto, I., Pablo, A.-L., Chenu, C., Brauman, A., 2014. Bacterial community
551 structure in soil microaggregates and on particulate organic matter fractions located
552 outside or inside soil macroaggregates. Pedobiologia 57, 191–194.
553 Blaud, A., Menon, M., van der Zaan, B., Lair, G.J., Banwart, S.A., 2017. Chapter Five - Effects of Dry
554 and Wet Sieving of Soil on Identification and Interpretation of Microbial Community
555 Composition, in: Sparks, S.A.B. and D.L. (Ed.), Advances in Agronomy, Quantifying and
556 Managing Soil Functions in Earth's Critical Zone Combining Experimentation and
557 Mathematical Modelling. Academic Press, pp. 119–142.
558 Bossuyt, H., Six, J., Hendrix, P.F., 2002. Aggregate-protected carbon in no-tillage and conventional
559 tillage agroecosystems using carbon-14 labeled plant residue. Soil Sci. Soc. Am. J. 66,
560 1965–1973.
561 Bronick, C.J., Lal, R., 2005. Soil structure and management: a review. Geoderma 124, 3–22.

562 Chèneby, D., Brauman, A., Rabary, B., Philippot, L., 2009. Differential responses of nitrate reducer
563 community size, structure, and activity to tillage systems. Appl. Environ. Microbiol. 75,
564 3180–3186.
565 Chotte, J.L., Schwartzmann, A., Bally, R., Jocteur Monrozier, L., 2002. Changes in bacterial
566 communities and Azospirillum diversity in soil fractions of a tropical soil under 3 or 19
567 years of natural fallow. Soil Biol. Biochem. 34, 1083–1092.
568 Davinic, M., Fultz, L.M., Acosta-Martinez, V., Calderón, F.J., Cox, S.B., Dowd, S.E., Allen, V.G., Zak,
569 J.C., Moore-Kucera, J., 2012. Pyrosequencing and mid-infrared spectroscopy reveal
570 distinct aggregate stratification of soil bacterial communities and organic matter
571 composition. Soil Biol. Biochem. 46, 63–72.
572 Enwall, K., Throbäck, I.N., Stenberg, M., Söderström, M., Hallin, S., 2010. Soil resources influence
573 spatial patterns of denitrifying communities at scales compatible with land
574 management. Environ. Microbiol.Appl. Environ. Microbiol. 76, 2243–2250.
Erguder, T.H., Boon, N., Wittebolle, L., Marzorati, M., Verstraete, W., 2009. Environmental factors
shaping the ecological niches of ammonia-oxidizing archaea. FEMS Microbiol. Rev. 33,
855–869.
575 Hallin, S., Jones, C.M., Schloter, M., Philippot, L., 2009. Relationship between N-cycling
576 communities and ecosystem functioning in a 50-year-old fertilization experiment. ISME
577 J. 3, 597–605.
578 Hayden, H.L., Drake, J., Imhof, M., Oxley, A.P.A., Norng, S., Mele, P.M., 2010. The abundance of
579 nitrogen cycle genes *amoA* and *nifH* depends on land-uses and soil types in South-
580 Eastern Australia. Soil Biol. Biochem. 42, 1774–1783.
581 Helgason, B.L., Walley, F.L., Germida, J.J., 2010. No-till soil management increases microbial
582 biomass and alters community profiles in soil aggregates. Appl. Soil Ecol. 46, 390–397.
583 IUSS Working Group WRB, 2014. World reference base for soil resources 2006, World soil
584 resources rep 103. ed. FAO, Rome.

585 Izquierdo, J., Nüsslein, K., 2006. Distribution of extensive *nifH* gene diversity across physical soil
586 microenvironments. Microbial Ecol. 51, 441–452.
587 Kaiser, M., Kleber, M., Berhe, A.A., 2015. How air-drying and rewetting modify soil organic matter
588 characteristics: An assessment to improve data interpretation and inference. Soil Biol.
589 Biochem. 80, 324–340.
590 Kravchenko, A.N., Negassa, W.C., Guber, A.K., Hildebrandt, B., Marsh, T.L., Rivers, M.L., 2014. Intra-
591 aggregate pore structure influences phylogenetic composition of bacterial community in
592 macroaggregates. Soil Sci. Soc. Am. J. 78, 1924.
593 Lair, G.J., Zehetner, F., Hrachowitz, M., Franz, N., Maringer, F.-J., Gerzabek, M.H., 2009. Dating of
594 soil layers in a young floodplain using iron oxide crystallinity. Quatern Geochronol 4,
595 260–266.
596 Lauber, C.L., Strickland, M.S., Bradford, M.A., Fierer, N., 2008. The influence of soil properties on
597 the structure of bacterial and fungal communities across land-use types. Soil Biol.
598 Biochem. 40, 2407–2415.
599 Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C.,
600 Schleper, C., 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils.
601 Nature 442, 806–809.
602 Lensi, R., Clays-Josser, A., Jocteur Monrozier, L., 1995. Denitrifiers and denitrifying activity in size
603 fractions of a mollisol under permanent pasture and continuous cultivation. Soil Biol.
604 Biochem. 27, 61–69.
605 Levy-Booth, D.J., Prescott, C.E., Grayston, S.J., 2014. Microbial functional genes involved in
606 nitrogen fixation, nitrification and denitrification in forest ecosystems. Soil Biol.
607 Biochem. 75, 11–25.
608 Ma, W.K., Bedard-Haughn, A., Siciliano, S.D., Farrell, R.E., 2008. Relationship between nitrifier
609 and denitrifier community composition and abundance in predicting nitrous oxide
610 emissions from ephemeral wetland soils. Soil Biol. Biochem. 40, 1114–1123.

611 Maeda, K., Spor, A., Edel-Hermann, V., Heraud, C., Breuil, M.-C., Bizouard, F., Toyoda, S., Yoshida,
612 N., Steinberg, C., Philippot, L., 2015. N_2O production, a widespread trait in fungi. Sci. Rep.
613 5.
614 Mangalassery, S., Sjögersten, S., Sparkes, D.L., Sturrock, C.J., Mooney, S.J., 2013. The effect of soil
615 aggregate size on pore structure and its consequence on emission of greenhouse gases.
616 Soil Till. Res. 132, 39–46.
617 Mendes, I.C., Bottomley, P.J., 1998. Distribution of a population of *Rhizobium leguminosarum bv.*
618 *trifolii* among different size classes of soil aggregates. Appl. Environ. Microbiol. 64, 970–
619 975.
620 Morales, S.E., Cosart, T., Holben, W.E., 2010. Bacterial gene abundances as indicators of
621 greenhouse gas emission in soils. ISME J 4, 799–808.
622 Muruganandam, S., Israel, D.W., Robarge, W.P., 2009. Activities of nitrogen-mineralization
623 enzymes associated with soil aggregate size fractions of three tillage systems. Soil Sci.
624 Soc. Am. J. 73, 751.
625 Neumann, D., Heuer, A., Hemkemeyer, M., Martens, R., Tebbe, C.C., 2013. Response of microbial
626 communities to long-term fertilization depends on their microhabitat. FEMS Microbiol.
627 Ecol. 86, 71–84.
628 Pinheiro, E.F.M., Pereira, M.G., Anjos, L.H.C., 2004. Aggregate distribution and soil organic matter
629 under different tillage systems for vegetable crops in a Red Latosol from Brazil. Soil Till.
630 Res. 77, 79–84.
631 Poly, F., Ranjard, L., Nazaret, S., Gourbiere, F., Jocteur Monrozier, L., 2001. Comparison of *nifH*
632 gene pools in soils and soil microenvironments with contrasting properties. Environ.
633 Microbiol.Appl. Environ. Microbiol. 67, 2255–2262.
634 Prosser, J.I., Nicol, G.W., 2008. Relative contributions of archaea and bacteria to aerobic ammonia
635 oxidation in the environment. Environ. Microbiol. 10, 2931–2941.
636 R Development Core Team, 2015. R: a language and environment for statistical computing.

637 Rabbi, S.M.F., Daniel, H., Lockwood, P.V., Macdonald, C., Pereg, L., Tighe, M., Wilson, B.R., Young,
638 I.M., 2016. Physical soil architectural traits are functionally linked to carbon
639 decomposition and bacterial diversity. Sci. Rep. 6, 33012.
640 Ramakrishnan, B., Lueders, T., Conrad, R., Friedrich, M., 2000. Effect of soil aggregate size on
641 methanogenesis and archaeal community structure in anoxic rice field soil. FEMS
642 Microbiol. Ecol. 32, 261–270.
643 Schaufler, G., Kitzler, B., Schindlbacher, A., Skiba, U., Sutton, M.A., Zechmeister-Boltenstern, S.,
644 2010. Greenhouse gas emissions from European soils under different land use: effects of
645 soil moisture and temperature. Eur. J. Soil Sci. 61, 683–696.
646 Schindlbacher, A., Zechmeister-Boltenstern, S., Butterbach-Bahl, K., 2004. Effects of soil moisture
647 and temperature on NO, NO_2, and N_2O emissions from European forest soils. J. Geophys.
648 Res. 109, 1–12.
649 Sey, B.K., Manceur, A.M., Whalen, J.K., Gregorich, E.G., Rochette, P., 2008. Small-scale
650 heterogeneity in carbon dioxide, nitrous oxide and methane production from aggregates
651 of a cultivated sandy-loam soil. Soil Biol. Biochem. 40, 2468–2473.
652 Shein, E.V., 2005. Kurs fiziki pochv (A Course of Soil Physics) [in Russian]. Moscow State Univ.
653 Publishing.
654 Six, J., Bossuyt, H., Degryze, S., Denef, K., 2004. A history of research on the link between
655 (micro)aggregates, soil biota, and soil organic matter dynamics. Soil Till. Res. 79, 7–31.
656 Six, J., Conant, R.T., Paul, E.A., Paustian, K., 2002. Stabilization mechanisms of soil organic matter:
657 Implications for C-saturation of soils. Plant Soil 241, 155–176.
658 Six, J., Elliott, E.T., Paustian, K., 2000. Soil macroaggregate turnover and microaggregate
659 formation: a mechanism for C sequestration under no-tillage agriculture. Soil Biol.
660 Biochem. 32, 2099–2103.
661 Tisdall, J.M., Oades, J.M., 1982. Organic matter and water-stable aggregates in soils. Eur. J. Soil Sci.
662 33, 141–163.

Verhamme, D.T., Prosser, J.I., Nicol, G.W., 2011. Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. ISME J. 5, 1067–1071.

Wallenstein, M.D., Vilgalys, R.J., 2005. Quantitative analyses of nitrogen cycling genes in soils. Pedobiologia 49, 665–672.

Table

Table 1. Soil characteristics and soil aggregate size distribution of bulk soil samples on a dry mass basis. Mean value ± 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 ± 0.26	14.1 ± 1.11	17.1 ± 0.69	12.0 ± 0.26
	Soil pH (H_2O)	7.7 ± 0.14	7.5 ± 0.07	7.4 ± 0.17	7.4 ± 0.09
	Organic C (%)	2.4 ± 0.36	3.2 ± 0.08	3.8 ± 0.28	5.0 ± 0.60
	Total N (%)	0.13 ± 0.01	0.17 ± 0.01	0.25 ± 0.02	0.33 ± 0.04
	C_{org}/N	18.1 ± 1.83	18.5 ± 1.60	15.1 ± 1.02	15.0 ± 0.52
	$N\text{-}NH_4^+$ (mg kg^{-1})	1.59 ± 0.29	0.49 ± 0.01	0.57 ± 0.03	4.77 ± 0.98
	$N\text{-}NO_3^-$ (mg kg^{-1})	20.3 ± 3.07	18.6 ± 4.00	24.3 ± 3.13	1.5 ± 0.66
	$P\text{-}PO_4^{3-}$ (g kg^{-1})	0.35 ± 0.10	1.13 ± 0.47	0.85 ± 0.48	0.59 ±0.04
	$CaCO_3$ (%)	19.0 ± 1.90	20.6 ± 1.11	20.4 ± 0.62	21.1 ± 1.41
	Sand, 63-2000 µm (%)	32.7	20.2	22.5	8.2
	Silt, 2-63 µm (%)	43.8	63.4	51.2	63.0
	Clay, < 2 µm (%)	23.5	16.4	26.3	28.8
	Soil texture	loam	silt loam	silt loam	silt loam
Soil aggregate size distribution (%)	> 10 mm	37.3 ± 9.1	11.3 ± 1.0	11.9 ± 4.4	7.9 ± 2.4
	5.0 - 10.0 mm	14.6 ± 2.4	15.5 ± 1.1	18.3 ± 2.7	21.5 ± 2.0
	2.0 - 5.0 mm	20.5 ± 4.0	26.1 ± 3.1	31.2 ± 2.2	37.8 ± 3.6
	1.0 - 2.0 mm	11.8 ± 2.4	21.8 ± 4.1	23.1 ± 8.4	14.5 ± 0.5
	0.5 - 1.0 mm	6.4 ± 3.5	9.3 ± 2.8	5.9 ± 1.7	5.2 ± 0.4
	0.25 - 0.5 mm	7.1 ± 4.6	12.7 ± 2.6	7.5 ± 2.7	6.9 ± 0.1
	< 0.25 mm	1.9 ± 1.3	3.3 ± 0.4	2.0 ± 0.8	6.1 ± 0.7

Figures captions

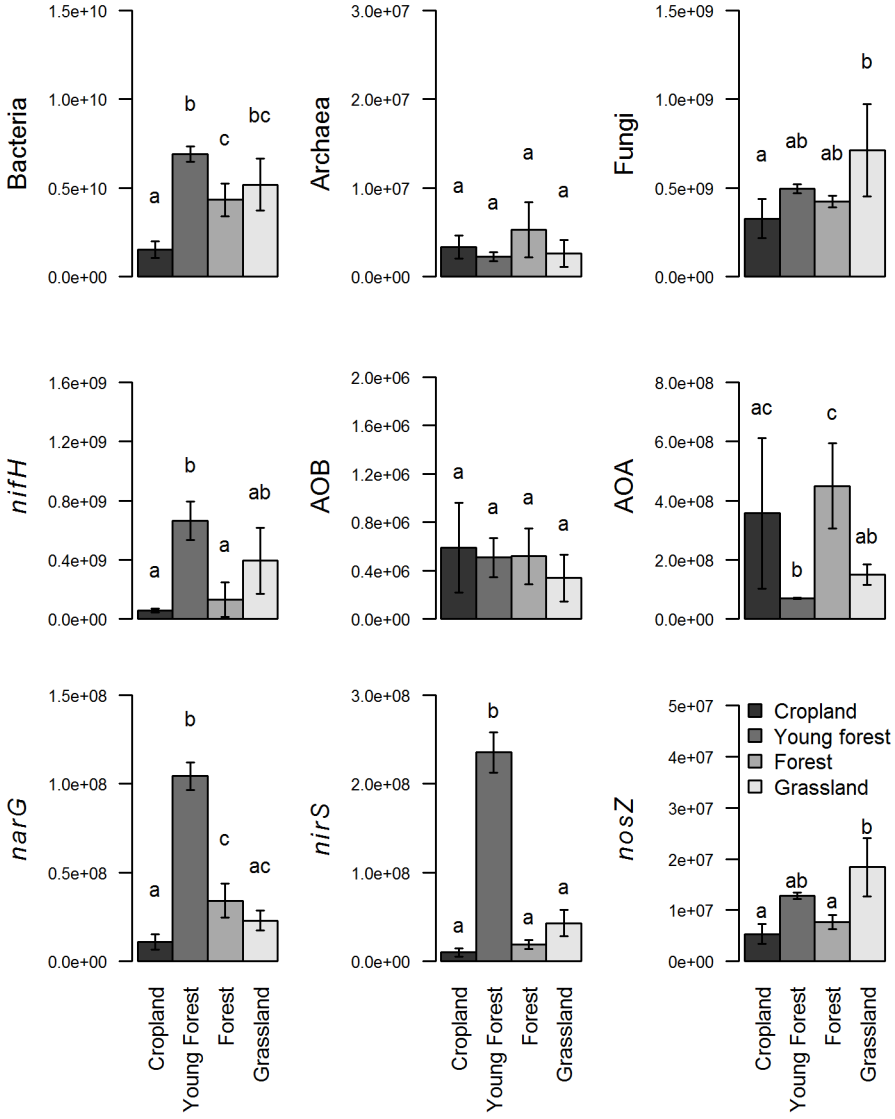
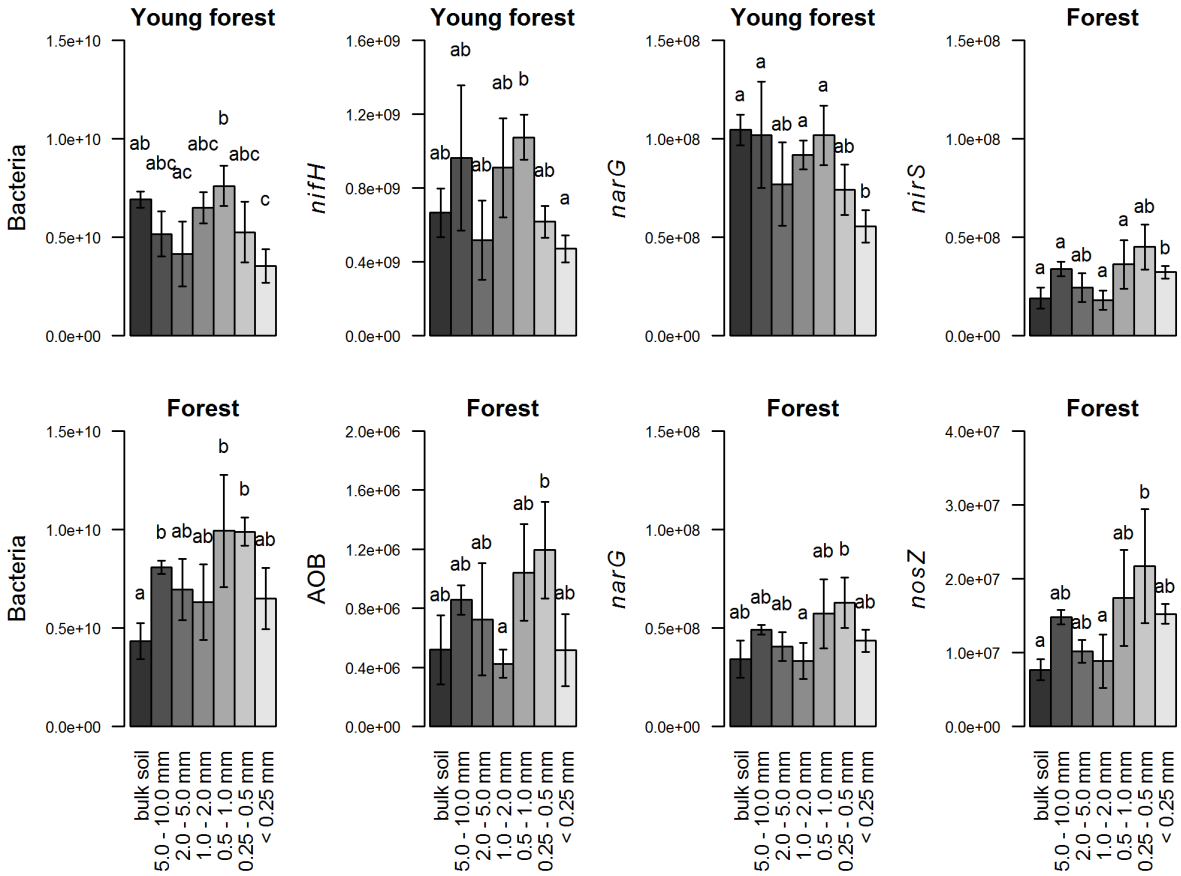


Fig. 1 Variation in gene abundance between bulk soil from four land use types. The following genes and microbial communities were targeted: bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixation (*nifH* gene), ammonia oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), 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 ± one standard deviation (n = 3) are shown. Small letters indicate significance (P < 0.05) of pairwise differences between land use.



704 **Fig. 2.** Variation in gene abundance between bulk soil and six soil aggregates sizes classes from
705 young forest and forest. The following genes and microbial communities were targeted:
706 bacterial and archaea (16S rRNA gene), fungi (ITS region), N fixation (*nifH* gene), ammonia
707 oxidizing bacteria and archaea (*amoA* gene, named AOB and AOA, respectively), nitrate
708 reductase (*narG* gene), nitrite reductase *(nirK* gene) and nitrous oxide reductase (*nosZ* gene). All
709 abundances are expressed on the basis of 1 g of dry mass of the bulk soil or the specific
710 aggregate size fraction. Mean value ± one standard deviation (n = 3) are shown. Small letters
711 indicate significance ($P < 0.05$) of pairwise differences between soil aggregate size classes
712 within a specific land use.
713

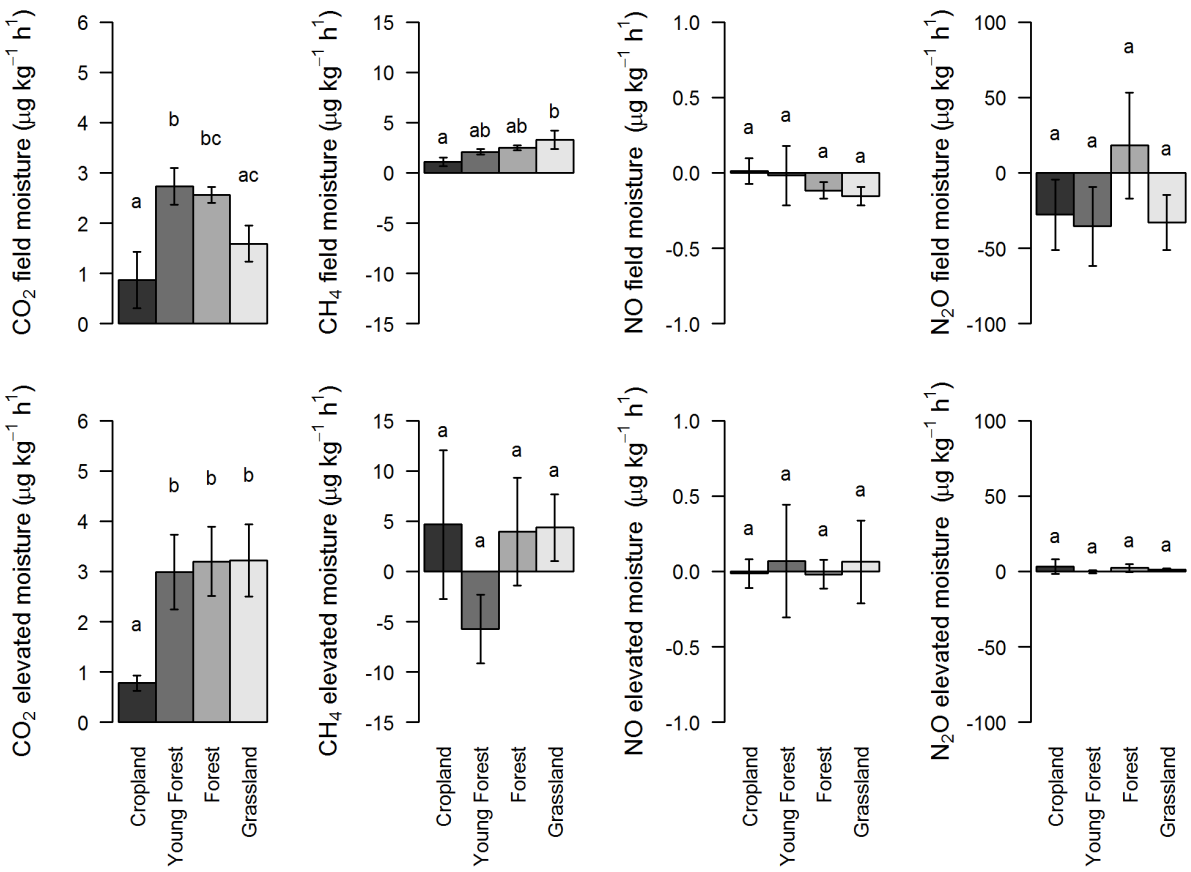
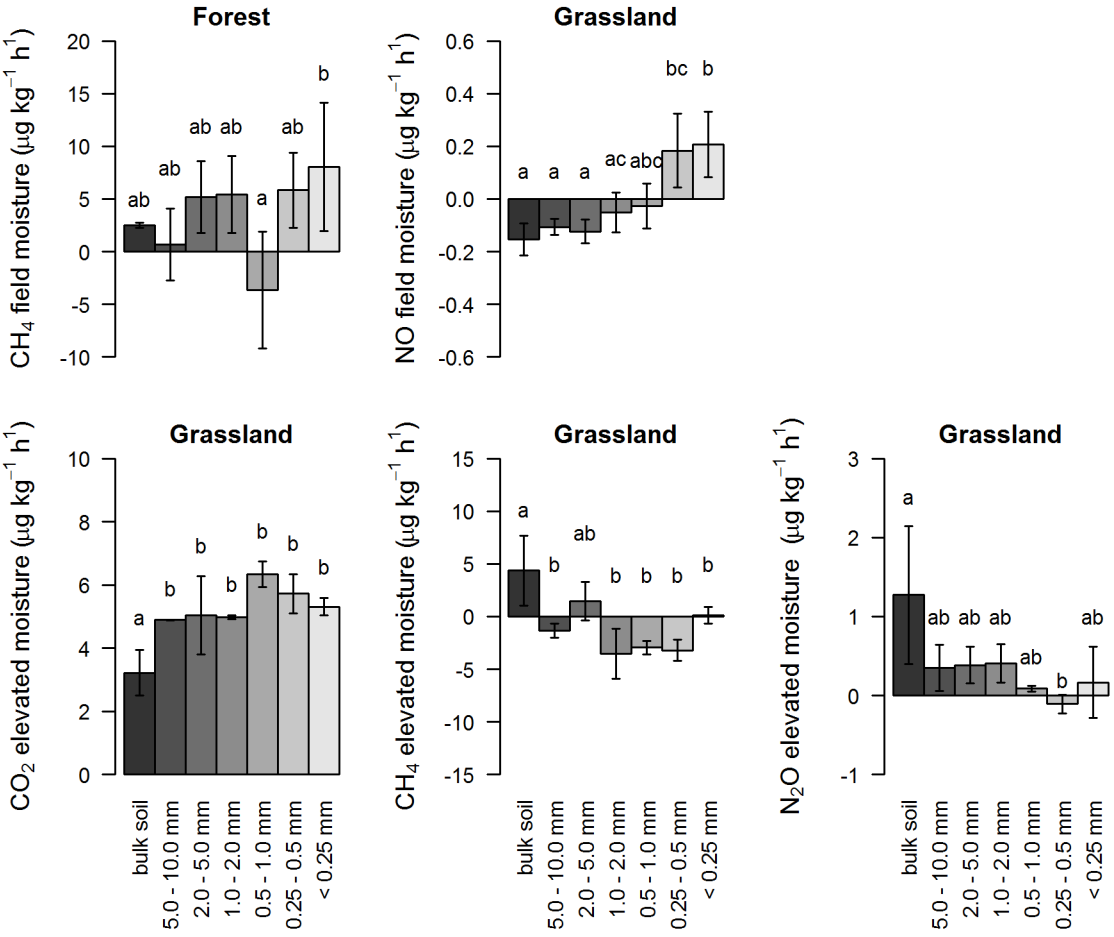


Fig. 3. Variation in GHG fluxes (µg kg^{-1} h^{-1}) between bulk soil from four land use types at field moisture or elevated moisture (40 – 60 % of field capacity). Mean value ± one standard deviation (n = 3) are shown. Small letters indicate significance (P < 0.05) of pairwise differences between soil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



721 **Fig. 4.** Variation in GHG fluxes (µg kg^{-1} h^{-1}) between bulk soil and six soil aggregates sizes classes
722 from grassland or forest at field moisture or elevated moisture (40 – 60 % of field capacity).
723 Mean value ± one standard deviation (n = 3) are shown. Small letters indicate significance (P <
724 0.05) of pairwise differences between soil aggregate size classes within a specific land use.
725

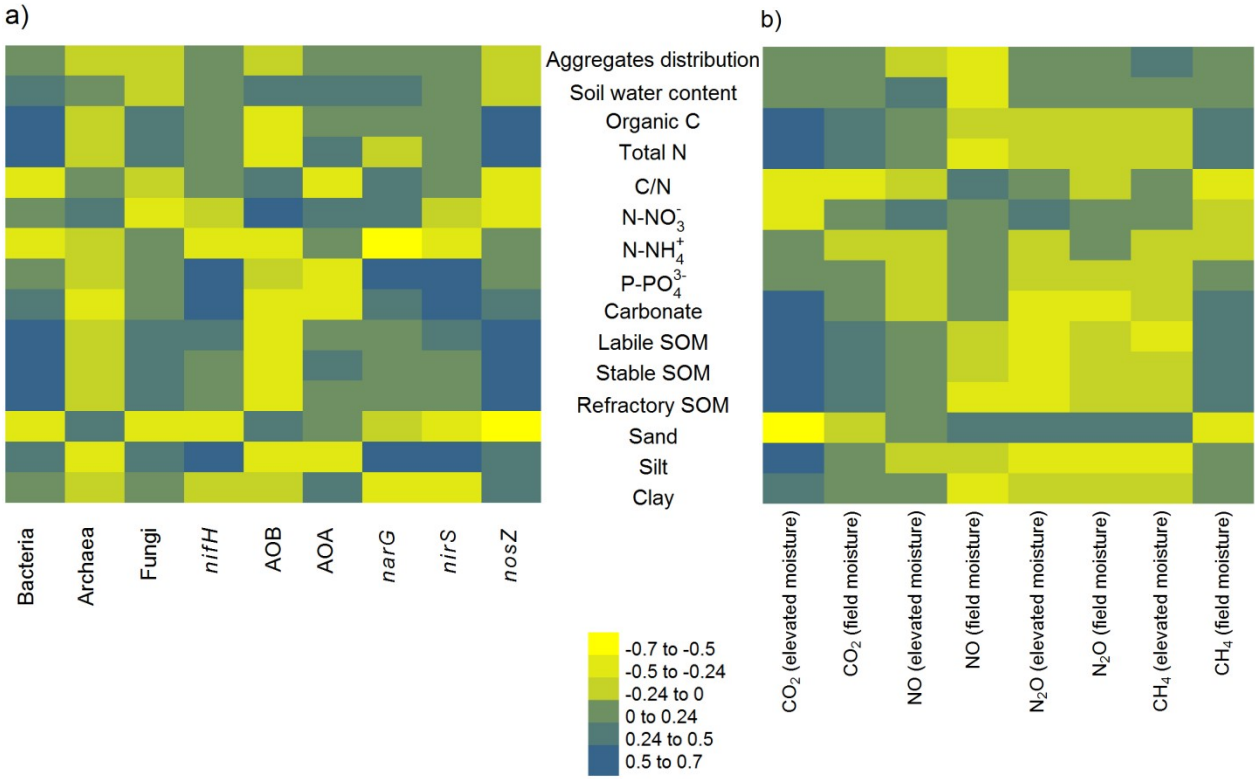
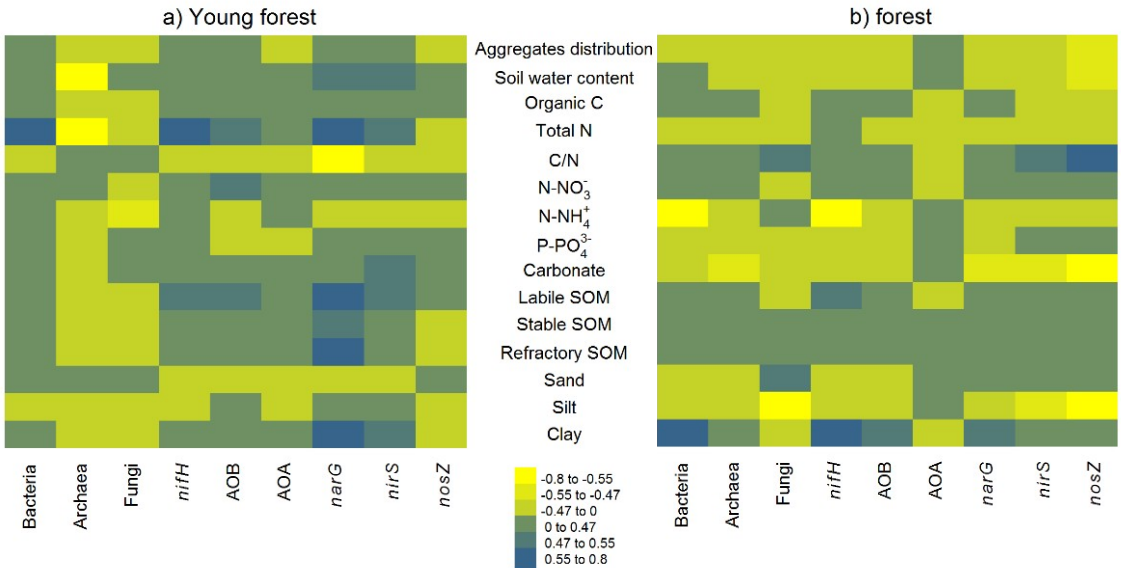


Fig. 5. Heatmaps of Spearman's rank correlation coefficients ρ between a) soil properties and microbial genes abundance, b) soil properties and greenhouse gas fluxes from samples across six soil aggregates sizes classes (< 0.25, 0.25 – 0.5, 0.5 – 1.0, 1.0 – 2.0, 2.0 – 5.0 and 5.0 – 10.0 mm) and four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



734 **Fig. 6** Heatmaps of Spearman's rank correlation coefficients ρ between soil properties and
735 microbial genes abundance from samples across six soil aggregates sizes classes (< 0.25, 0.25 –
736 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
737 separately, which showed significant variation in gene abundance with aggregates size classes
738 (refers to figure S13 for the other land uses). AOB: *amoA* bacteria; AOA: *amoA* archaea. The ρ
739 values > 0.47 and < -0.47 are significant ($P < 0.05$).