Impact of nitrogen fertilization on the soil microbiome and nitrous oxide emissions

Cassman, N.A.

Impact of nitrogen fertilization on the soil microbiome and nitrous oxide emissions

Noriko A. Cassman

Impact of nitrogen fertilization on the soil microbiome and nitrous oxide emissions

The research described in this thesis was carried out in the Department of Microbial Ecology at the Netherlands Institute for Ecology (NIOO-KNAW). Financial support was provided by NWO Grant number 729.004.003.

Cover photo by Noriko A. Cassman and the nitrous oxide space-filling illustration from Wikimedia Commons.

Cover design and thesis layout by Noriko A. Cassman

Printed by UFB / Grafimedia

Impact of nitrogen fertilization on the soil microbiome and nitrous oxide emissions

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties

te verdedigen op 17 april 2019
klokke 13:45 uur

door

Noriko Alexandra Cassman

geboren 1 juli 1986

in Santiago, Chile

Promotor:	Prof. Dr. J.A. van Veen Netherlands Institute for Ecology Leiden University
Co-promotor:	Dr. E.E. Kuramae Netherlands Institute for Ecology
Overige:	Prof. dr. G.P. van Wezel Leiden University
	Prof. Dr. J.M. Raaijmakers Netherlands Institute for Ecology Leiden University
	Prof. Dr. H.J. Laanbroek Netherlands Institute for Ecology Utrecht University
	Dr. M.H. Medema Wageningen University
	Dr. M.M. Hefting Utrecht University

Table of Contents

Chapter 1

General Introduction

Introduction

Nitrogen is an essential component of living systems. Nitrogen is abundant in the atmosphere as inert di-nitrogen gas and its transformation to bioavailable forms is limited by biological nitrogen fixation ($N_2 \rightarrow NH_3$; Canfield et al 2010). This latter process is carried out by a few genera of microbes and as the rate of conversion is slow, natural ecosystems are often N-limited (Kuypers et al 2018). The invention of synthetic nitrogen fixation by Fritz Haber (now called the Haber-Bosch process) in 1909 led to the mass production and widespread use of nitrogen fertilizers and corresponding high yields in agriculture in the next century (Ellis 2011). The agricultural boom of the past century has substantially attributed to a seven-fold increase in the human population, which is now over seven billion (Galloway et al 2008). About half of the world's population relies on food grown using synthetic N (Erisman et al 2008). This increased input of N fertilizers into agricultural systems has had serious impacts beyond increasing food productivity, including long-term decreases in biodiversity, in soil quality, waterway eutrophication and acidification, and greenhouse gas emissions (Foley et al 2011, Fowler et al 2015, Smith 2017).

As only about 50% of the input N to agricultural soils is used by plants, the excess nitrogen is leached out of the soil matrix and into the air and surrounding water sources, resulting in an imbalance of nitrogen in surrounding ecosystems which contributes to the degradation of surface and groundwater quality (Schlesinger et al 2009, Erisman et al 2013). Moreover, N transformations in the soil matrix include processes resulting in the greenhouse gases NO and N_2O. Before 2050, global food production is expected to double to feed the projected human population of 9 billion people (Godfray et al 2010, Tilman et al 2011). Updating nitrogen fertilizer management strategies toward long-term sustainability without decreasing crop productivity is therefore of global importance. This requires deep knowledge of the soil system, especially regarding the effects of nitrogen fertilizers on the soil microbes, which are the main players in nutrient cycling, litter decomposition and energy flows in terrestrial and agroecosystems (Baggs 2011; Hu et al 2014b). While the astronomical diversity of soil microbes has hampered detailed study of the soil microbiome, the recent advances in sequencing technology have allowed for an unprecedented glimpse into the "black box" of the microbial role in soil functioning (Torsvik et al 1990, Fierer et al 2012). Here, the overall research aim was to apply next-generation sequencing technology and associated advanced data analyses to gain detailed insight into the responses of soil microbial communities to various nitrogen fertilizer regimes,

including long term fertilization, with a focus on the potentially N_2O-producing microbial community.

1.1 Nitrous oxide emissions as a function of N fertilizer input

Reactive nitrogen generally is supplied to agricultural soils in the form of ammonium-based fertilizers, such as urea ($CO(NH_2)_2$), ammonium nitrate (NH_4NO_3), ammonium sulphate ($(NH_4)_2SO_4$) and synthetic ammonia (NH_3; Mosier 1994). About 220 Tg N yr^{-1} of nitrogen fertilizers are applied to agricultural soils globally, of which about half is lost into groundwater as soluble NO_x or as gaseous NO_x species (Gruber & Galloway 2008, Fowler et al 2015). This follows the conceptual "hole-in-the-pipe" model, also known as the nitrogen cascade, which describes soil nitrogen transformations as limited by the availability of reactive nitrogen, which then "leak" through a cascade of reactions (Galloway et al 2003). Roughly 18.8 Tg of $N\text{-}N_2O$ are emitted per year, with agricultural soils directly contributing to 16% of these emissions (Syakila & Kroeze 2011, Smith 2017). A general rule is to consider that 1% of applied fertilizer N is emitted as N_2O based on a rough estimation by IPCC (2007). However, recent studies show that this value may fluctuate from 0.2 to 4% depending on many factors, including site, soil type and management (Carmo et al 2013, Filoso et al 2015). Nitrous oxide emissions threaten the global climate because N_2O has a global warming potential 298 times that of CO_2 due to its radiative forcing and long presence (114 years/molecule) in the atmosphere (Robertson & Vitousek 2009, Snyder et al 2009). Further, once in the atmosphere it is converted to NO which reacts with tropospheric ozone; this implicates N_2O as a major ozone-depleting substance (Ravishankara et al 2009). Efforts to develop N_2O mitigation strategies focus on efficiency in N fertilizer utilization and more recently on identifying the controls and mechanisms of N_2O emissions, including the microbial role (Signor and Cerri 2013, Butterbach-Bahl et al 2013, Soares et al 2016, Pitombo et al 2016, Galloway et al 2017, Bakken & Frostegård 2017, Lourenco et al 2018, Kuypers et al 2018).

Nitrous oxide emissions from agricultural soils are mainly attributed to the cumulative effects of the biotic pathways nitrification and denitrification (Butterbach-Bahl 2013). Nitrification is the two-step oxidation of NH_4^+ to NO_2^- and NO_2^- to NO_3^-, in which N_2O is an intermediate, while denitrification is the sequential reduction of NO_3^-, NO_2^-, NO, N_2O and N_2 in which N_2O is a product of NO reduction and a reactant of N_2O reduction to N_2 (Baggs et al 2011). In ammonia oxidation, the rate-limiting step is ammonia oxidation to hydroxylamine, which is generally catalyzed by ammonia monooxygenase and encoded by the gene amoA. The other main biotic pathway leading to N_2O, denitrification ($NO_3^- \rightarrow NO_2^- \rightarrow$

$NO \rightarrow N_2O \rightarrow N_2$), is catalyzed by a series of enzymes which are encoded by different genes (Hu et al 2015). The first step ($NO_3^- \rightarrow NO_2^-$) is carried out by the enzyme nitrate reductase, which is encoded by the *narG* or *napA* gene; the second step ($NO_2^- \rightarrow NO$) can be catalyzed by two types of nitrite reductases encoded respectively by the the nirK or nirS genes. The third step ($NO \rightarrow N_2O$) is carried out by the genes cnorB or qnorB; last, the enzyme nitric oxide reductase catalyzes the reduction of N_2O ($N_2O \rightarrow N_2$) and is encoded by the nosZ gene, which exists in two forms (nosZ I and II). The relative contributions of nitrification and denitrification to overall N_2O production are challenging to untangle due to the many interrelated reactions and microbes with overlapping function (Zhu et al 2013; Shcherbak, Millar and Robertson 2014). Other sources of N_2O emissions are denitrification by nitrifiers (nitrifier denitrification), anaerobic ammonium oxidation (anammox), complete nitrification (comammox) and dissimilatory nitrate reduction to ammonium (DNRA, or nitrate ammonification (Hu et al 2015, Kuypers et al 2018). However, due to the main contributions of nitrification and denitrification to N_2O emissions in agriculture, in the current research the focus was on nitrification and denitrification.

1.2 Nitrification and denitrification

Nitrification and denitrification are mediated by microbes (archaea, bacteria and fungi) which use these pathways to gain energy or assimilate N. Nitrifiers encompass a narrow phylogenetic range of a few bacterial and archaeal genera. Ammonia oxidation is mediated by the ammonia-oxidizing archaea (AOA), such as the Thaumarchaeota *Nitrososphaera*, and the ammonia-oxidizing bacteria (AOB), such as the Betaproteobacteria *Nitrosomonas* and *Nitrosospira*; Upon ammonium oxidation, nitrite can be formed which can be further oxidized by nitrite oxidizing bacteria (NOB), including the Nitrospirae *Nitrospira* and the Alphaproteobacteria *Nitrobacter*. Further, the process of complete nitrification by the recently discovered comammox bacteria, which have so far been found in the NOB *Nitrospira* genus, might also contribute to N_2O emissions (Liu et al 2017). Comammox bacterial genomes have revealed the full set of nitrification genes, that is for ammonia oxidation ($NH_3 \rightarrow NH_2OH$, *amoA*) and hydroxylamine oxidation (NH_2OH, *hao*), as well as the genes for nitrite oxidation ($NO_2^- \rightarrow NO_3^-$, *nxrB*; Daims et al 2015; van Kessel et al 2015; Camejo et al 2017). Both ammonia and nitrite oxidation is an obligately aerobic process, with nitrifiers being chemolitho-heterotrophic and -autotrophic.

Denitrification is a facultative anaerobic process carried out by microorganisms widely dispersed over the bacterial, archaeal and fungal domains, and deni-

trification genes can also be carried by nitrifiers in what is termed nitrifier denitrification. Some denitrifiers contain the full suite of denitrification genes and are able to reduce NO_3^- to N_2; these are known as full denitrifiers. Others contain a truncated set of denitrification pathway genes and may produce one of the intermediates, such as NO (which is rapidly converted to N_2O) or N_2O. The genetic potential of the denitrification community for full or incomplete denitrification is directly linked to the N_2O or N_2 output of the soil. A community with a higher proportion of n*osZ* to *norB* or n*irS* + n*irK* (full denitrifiers) may present a sink for N_2O (Jones, Graf et al 2013). In support, Philippot et al (2011) found increased N_2O emissions from soils when increasing dilutions of bacteria lacking *NosZ* were added to microcosms. Further, recent studies provided evidence for this as well (Domeignoz−Horta 2015 and 2018); for example, as the addition of non-denitrifier *nosZII*-containing bacteria in microcosms was linked to lower N_2O emissions (Domeingoz−Horta 2016). Thus, the overall genetic potential of a nitrifying or denitrifying community, along with environmental controls, impacts the amount of N_2O emitted.

1.3 Management factors influencing soil microbial nitrifiers and denitrifiers

The proximal, or immediate and short-term, factors influencing nitrifier and denitrifiers are carbon availability, NO_3^- concentrations, moisture levels and oxygen availability, while distal, or indirect, and long-term factors are plant growth, micronutrient availability, and pH (Hénault 2012 and Saggar, Jha et al 2013). When N fertilizers are applied, microbial decomposition can be increased or decreased, depending on recalcitrance of the organic substrate and N availability. Application of plant residues with low C:N ratios often result in high rates of N mineralization, or the conversion of organic N to plant-available NH_3 (usually by microbial death), while residues with higher C:N ratios stimulate N immobilization into microbial biomass (reviewed in Chen et al 2014). Soil organic matter directly affects N_2O production because it provides a diverse suite of substrates for heterotrophic denitrifier activity (Schmidt & Torn et al 2011). For instance, soluble sugars, or labile carbon, can easily dissolve into the water-filled spaces in soil and become available for microbial or plant uptake. In contrast, the insoluble compound lignin requires specialized microbial enzymes for degradation and otherwise remains in the soil as soil organic matter (SOM) (Swift et al 1979). Additionally, rapid decomposition can drive down oxygen levels faster than the rate of oxygen diffusion, establishing anaerobic conditions for denitrification. Parkin (1987) showed that the frequencies of N_2O emissions correlate with predictions based on the spatially heterogeneous distribution of organic compounds that are

found in soils. Nitrite levels control the nitrification and denitrification processes as it is a reaction intermediate and reactant, respectively. Therefore, N and organic matter additions -- such as in agricultural management practices of fertilization with N and plant residues -- can either promote or reduce N_2O production by their effect on the factors controlling the activity and growth of nitrifiers and denitrifiers.

1.4 Sugarcane agriculture

The N_2O emissions of sugarcane production cycles has recently drawn attention due to the use of sugarcane bioethanol as a sustainable biofuel (Crutzen et al 2008, Lisboa et al 2011, Seabra et al 2011). The largest producer of sugarcane, *Saccharum sp.*, is Brazil, which devotes almost 7.5 million hectares to sugarcane production mainly for its use as a biofuel (Christofoletti et al 2013). Sustainability of sugarcane production stems partly from the crop characteristics and partly because of efficiency in its production. After the sugarcane stalk is cut during a harvest, the regrowth yields another crop, known as the ratoon crop, during the following harvest season. The growth from the ratoon crop decreases each year, which warrants replanting of the plant crop every three to eight years, without tilling the soil in the intervening years, and this promotes SOM formation. Historically, sugarcane leaves were burned to remove the plant leaves from the sugar-containing stalks prior to harvest. Now, most Brazilian sugarcane is harvested using a 'green harvest' method in which the stalks are stripped of leaves and this so-called "straw" is left on the field (Carvalho et al 2017). The amount of dry sugarcane straw on fields in Brazil ranges between 8–30 Mg ha^{-1} dry mass of straw (Carvalho et al 2017). The green harvest method has several advantages over the burning method, namely, that application of the residues increases moisture retention and provides a long-term source of nutrients (Carvalho et al 2017) and contributes to overall lower greenhouse gas emissions (Capaz 2013). Depending on the cultivar and the conditions in which it was grown, sugarcane leaves have a C:N of roughly 125:1, which is relatively high (Carvalho et al 2017). Decomposition of plant residues with C:N of above 30 generally promotes N immobilization, or the uptake of available soil-borne N into microbial biomass. This immobilized N can turn into soil organic N following microbial death, which serves as a long-term source of N to subsequent crops (Otto et al 2013). Application of crop residues with high C:N content, such as sugarcane leaves, may lead to microbial decomposers using soil organic N for their N needs, ultimately lowering the soil N pool, unless combined with an N fertilization regime (Trivelin et al 2013, Ferreira et al 2015).

1.5 Sugarcane bioethanol and vinasse production

In the bioethanol production cycle, the sugarcane stalk is crushed, and the sugarcane juice is separated from the pulpy stalk residue. Sugarcane juice is heated, clarified with lime and cooled to crystallize sugar and molasses. The molasses is further fermented and heated to produce bioethanol and the waste product, vinasse. Up to 13 L of vinasse per liter of bioethanol may be generated (Boddey et al 2008). Essentially all of the vinasse is recycled onto the sugarcane fields as a K fertilizer source according to Brazilian agricultural practices (Moran-Salazar et al 2016). Vinasse is comprised of about 93% water and organic acids, solids and nutrients such as magnesium, calcium and potassium (Christofoletti et al 2013). It is effective as a K and P fertilizer (Moran-Salazar et al 2016) and is also used in animal feed and as a source of biogas (Christofoletti et al 2013). Benefits of using vinasse as fertilizer include improved soil quality due to the addition of moisture and micronutrients (Jiang et al 2012) and improved crop production and crop quality (Yi-Ding et al 2006, Zani et al 2018). However, when vinasse is used in conjunction with an N fertilizer, potentially detrimental effects on long-term soil fertility and greenhouse gas emissions have been observed, especially the emission of N_2O and reduction of soil C stocks due to the addition of labile C from vinasse (Fuess et al 2017, Pitombo et al 2016, do Carmo et al 2013). These negative consequences might outweigh the benefits of sugarcane bioethanol as an energy source (Lapola et al 2010, Erisman et al 2010). Further, microbial contaminants of the bioethanol process are thought to be present in vinasse (Costa et al 2015) with unknown effects on the soil microbiome upon fertilization.

1.6 Insight into microbial communities through sequencing

The soil matrix contains an astronomical number and diversity of microorganisms, which can reach up to 10^{13} cells and contain between 10^4-10^9 genotypes in one gram of soil (Torsvik and Øvreås 2002). This great diversity is a challenge to study, not least because the majority of soil microbes are unculturable. Recent advances in high-throughput sequencing technologies and computational methods, largely driven by the less diverse microbial communities of the marine and human gut environments, have enabled scientists to begin tackling the soil ecosystem (Zhou et al 2015). Briefly, a comparative metagenomics study encompasses experimental design, DNA or RNA extraction from environmental samples, sequencing, quality control of the reads, followed by taxonomic and/or functional potential identification of the reads and statistical analysis to address hypotheses.

The data subjected to the statistical analyses generally come in the form of taxonomic or functional profiles. Multivariate statistics are then applied, e.g. to identify taxa differing between groups of samples, or to find the most represented metabolic pathways in a metagenome. Several molecular methods are used to generate this data, including amplicon of phylogenetic markers or functional genes and shotgun metagenomics (Luo et al 2014, Orellana et al 2017).

The PCR of phylogenetic markers from microbial DNA in soil samples allows for the surveying of the taxonomic composition and diversity of soil microbial communities (Pace 1997, Huse et al 2008). Generally, the 16S rRNA gene is used to profile the bacterial and archaeal community while the 18S rRNA gene and/or ITS region are used for eukaryotes, including fungi. Advantages to using this method are lower cost per sample and the availability of large databases of marker genes representing sequences from millions of species. However, this strategy, so-called amplicon metagenomics, is limited by the conservation of the primers used, which can miss highly novel, divergent sequences as well as viruses; further, only taxonomic information is obtained (Logares et al 2014). Regarding the latter, several bioinformatic analysis methods have tackled gaining functional information by matching 16S taxonomy information to the functional potential of similar genomes, for example Picrust and Tax4Fun (Langille et al 2013 and Aßhauer et al 2015). These tools depend on prior knowledge of full genomes in the reference databases, which might limit the accounting of the true functional diversity of the sample. Further, the precision of reference-based methods depend on which lineages are represented in the databases.

Similar to the information derived from amplicon metagenomics, PCR of functional markers can reveal taxonomic and diversity information about a functional subgroup of the soil microbial community, e.g. the amplification and sequencing of the *amoA* gene gives insight into the ammonia-oxidizing bacterial community (Ouyang et al 2016). As functional amplicon metagenome techniques are limited to revealing relative abundances of taxa in the sample, these surveys can be supplemented by alternatives to measuring microbial biomass, such as real-time PCR, which is a quantitative method for measuring the number of copies of a gene, as a proxy for the number of cells, in a sample. The FUNGENE database is one such tool that provides a platform for functional amplicon metagenomic analysis and includes databases and Hidden Markov Models (HMMs) of a range of functional genes, including the main genes involved in nitrification (*amoA*) and denitrification (*nirS*, *nirK*, *nosZ*; Fish et al 2013). Further, the database dbCAN provides a stand-alone database for the analysis of genes encoding for enzymes involved in carbohydrate metabolism (Zhang et al 2018).

Functional potential as well as taxonomic information can be derived from shotgun metagenomes, which are genomic sequences derived from all the cells in a sample (Thomas et al 2012). Function is inferred by translating the sequences through a gene predictor followed by homology searching against a protein sequence or protein family database. Common databases for functional potential analysis include the Kyoto Encyclopedia of Genes and Genomes (KEGG), in which the genes are cross-referenced into metabolic pathways, and the protein family database (Pfam), in which protein domains are represented as HMMs (Kanehisa et al 2014, Finn et al 2016). The model organism E. coli, humans and the human gut microbiome only have 90%, 82% and 75% functionally annotated genes, respectively. In a complex, less-studied environment such as soil, the percentage of functionally annotated genes may further drop to 55% (Prakash & Taylor 2012). There are several widely used platforms for metagenomic analysis, including the MG-RAST and EBI platforms which allow users to upload and store data and to run their samples through automated pipelines. In addition to the application of amplicon and shotgun metagenomics to DNA, these analyses have also been applied to RNA transcripts (metatranscriptomics) and protein sequences (proteomics), which allow for gene expression and protein sequence levels to quantify soil microbial activity (Urich et al 2008, Hirsch et al 2010). This is useful in studies linking the activity of microbes with a potential function, e.g. the abundance of *amoA* gene transcripts, to responses, e.g. N_2O emissions (Theodorakopoulos et al 2017). Further, the sheer volume of sequencing coupled with high-throughput analytical techniques have enabled the binning of draft genomes, or metagenome-assembled genomes, from environments with low and medium diversity, with soil on the horizon (Sharon & Banfield 2013, Orellana et al 2018). Further goals are the linking of metabolomes, or all the proteins in a sample, with the metatranscriptome, metagenome and genomic information.

1.7 Research aims and thesis outline

The purpose of this dissertation was to investigate the connected system of the soil microbial community, nitrogen and organic fertilizers, and N_2O emissions. This will help to devise strategies targeting the microbes specifically affected by nitrogen fertilization. To do this, I analyzed long- and short-term studies of the effects of different N fertilizer treatments on the microbial soil communities in Dutch pasture soils and in Brazilian sugarcane fields. This was to identify the microbial taxa that responded to the treatments, with a focus on the microbial taxa that were directly involved in N_2O emissions.

In **Chapter 2** I describe potential direct and indirect effects of long-term fertilization with N, P and K on the plant and soil bacterial and fungal communities. To this end I applied co-variation analysis to the taxonomic compositions of each community across the treatments and to a suite of soil physicochemical measurements. In **Chapter 3** I focus on the effects of long-term inorganic fertilization on soil physicochemical characteristics and the soil microbial taxa in Dutch pasture soils. This was done by combining shotgun metagenomic analysis with soil physicochemical measurements using multivariate statistics.

In **Chapter 4** I investigated the effect of different urea fertilization treatments with or without nitrification inhibitors on nitrous oxide fluxes, soil physicochemical characteristics and the soil microbial community in a field experiment. Using 16S rDNA amplicon metagenomes, I evaluated the effect of these treatments on the overall bacterial community composition and diversity, and on functional subgroups using qPCR of nitrification and denitrification genes. In **Chapter 5** I describe further the effect of these urea and nitrification inhibitor treatments on the abundance of ammonia-oxidizing bacteria and other nitrifying species by the analysis of an *amoA* amplicon sequences combined with data mining of the previously published 16S rDNA dataset. Further, I identify the likely species directly responsible for the N_2O emissions in a tropical soil.

In **Chapter 6** I focused on vinasse, which contains a previously uncharacterized microbial assemblage. I obtained metagenome assembled genomes from vinasse samples taken over 1.5 years from a bioethanol factory in Brazil. Based on the functional potential described in these genomes, I describe potential effects of these vinasse bacteria on N_2O emissions in the field when used in fertirrigation.

Last, in **Chapter 7** I provide a general discussion of the research chapters, and present conclusions as well as some thoughts on future directions. This thesis showcases several advanced statistical and bioinformatic methods applied to metagenomic data. Further, the results of this thesis will contribute to the literature serving as a reference for farmers and policy-makers to steer the soil microbiome in agriculture toward long-term sustainability.

1.8 References

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31, 2882–2884.

Baggs, E.M., 2011. Soil microbial sources of nitrous oxide: recent advances in knowledge, emerging challenges and future direction. Current Opinion in Environmental Sustainability, Carbon and nitrogen cycles 3, 321–327.

Bakken, L.R., Frostegård, Å., 2017. Sources and sinks for N_2O, can microbiologist help to mitigate N_2O emissions? Environmental Microbiology 19, 4801–4805.

Boddey, R.M., Soares, L.H. de B., Alves, B.J.R., Urquiaga, S., 2008. Bio-Ethanol Production in Brazil, in: Pimentel, D. (Ed.), Biofuels, Solar and Wind as Renewable Energy Systems: Benefits and Risks. Springer Netherlands, Dordrecht, pp. 321–356.

Butterbach-Bahl, K., Baggs, E.M., Dannenmann, M., Kiese, R., Zechmeister-Boltenstern, S., 2013. Nitrous oxide emissions from soils: how well do we understand the processes and their controls? Phil. Trans. R. Soc. B 368, 20130122.

Camejo, P.Y., Domingo, J.S., McMahon, K.D., Noguera, D.R., 2017. Genome-Enabled Insights into the Ecophysiology of the Comammox Bacterium "Candidatus Nitrospira nitrosa." mSystems 2, e00059-17.

Canfield, D.E., Glazer, A.N., Falkowski, P.G., 2010. The Evolution and Future of Earth's Nitrogen Cycle. Science 330, 192–196.

Capaz, R.S., Carvalho, V.S.B., Nogueira, L.A.H., 2013. Impact of mechanization and previous burning reduction on GHG emissions of sugarcane harvesting operations in Brazil. Applied Energy, Special Issue on Advances in sustainable biofuel production and use - XIX International Symposium on Alcohol Fuels - ISAF 102, 220–228.

Carmo, J.B. do, Filoso, S., Zotelli, L.C., Neto, E.R. de S., Pitombo, L.M., Duarte−Neto, P.J., Vargas, V.P., Andrade, C.A., Gava, G.J.C., Rossetto, R., Cantarella, H., Neto, A.E., Martinelli, L.A., 2013. Infield greenhouse gas emissions from sugarcane soils in Brazil: effects from synthetic and organic fertilizer application and crop trash accumulation. GCB Bioenergy 5, 267–280.

Carvalho, J.L.N., Nogueirol, R.C., Menandro, L.M.S., Bordonal, R. de O., Borges, C.D., Cantarella, H., Franco, H.C.J., 2017. Agronomic and environmental implications of sugarcane straw removal: a major review. GCB Bioenergy 9, 1181–1195.

Chen, B., Liu, E., Tian, Q., Yan, C., Zhang, Y., 2014. Soil nitrogen dynamics and crop residues. A review. Agron. Sustain. Dev. 34, 429–442.

Christofoletti, C.A., Escher, J.P., Correia, J.E., Marinho, J.F.U., Fontanetti, C.S., 2013. Sugarcane vinasse: Environmental implications of its use. Waste Management 33, 2752–2761.

Costa, O.Y.A., Souto, B.M., Tupinambá, D.D., Bergmann, J.C., Kyaw, C.M., Kruger, R.H., Barreto, C.C., Quirino, B.F., 2015. Microbial diversity in sugarcane ethanol production in a Brazilian distillery using a culture-independent method. J Ind Microbiol Biotechnol 42, 73–84.

Crutzen, P.J., Mosier, A.R., Smith, K.A., Winiwarter, W., 2007. N_2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. Atmospheric Chemistry and Physics Discussions 7, 11191–11205.

Daims, H., Lebedeva, E.V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R.H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P.H., Wagner, M., 2015. Complete nitrification by *Nitrospira* bacteria. Nature 528, 504–509.

Domeignoz−Horta, L.A., Philippot, L., Peyrard, C., Bru, D., Breuil, M.-C., Bizouard, F., Justes, E., Mary, B., Léonard, J., Spor, A., 2018. Peaks of in situ N_2O emissions are influenced by N_2O-producing and reducing microbial communities across arable soils. Global Change Biology 24, 360–370.

Domeignoz-Horta, L.A., Putz, M., Spor, A., Bru, D., Breuil, M.C., Hallin, S., Philippot, L., 2016. Non-denitrifying nitrous oxide-reducing bacteria - An effective N_2O sink in soil. Soil Biology and Biochemistry 103, 376–379.

Domeignoz-Horta1, L., Spor, A., Bru, D., Breuil, M.-C., Bizouard, F., Leonard, J., Philippot, L., 2015. The diversity of the N_2O reducers matters for the N_2O:N2 denitrification end-product ratio across an annual and a perennial cropping system. Front. Microbiol. 6.

Ellis, E.C., 2011. Anthropogenic transformation of the terrestrial biosphere. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 369, 1010–1035.

Erisman, J.W., Galloway, J.N., Seitzinger, S., Bleeker, A., Dise, N.B., Petrescu, A.M.R., Leach, A.M., Vries, W. de, 2013. Consequences of human modification of the global nitrogen cycle. Phil. Trans. R. Soc. B 368, 20130116.

Erisman, J.W., Sutton, M.A., Galloway, J., Klimont, Z., Winiwarter, W., 2008. How a century of ammonia synthesis changed the world. Nature Geoscience 1, 636–639.

Erisman, J.W., van Grinsven, H., Leip, A., Mosier, A., Bleeker, A., 2010. Nitrogen and biofuels; an overview of the current state of knowledge. Nutr Cycl Agroecosyst 86, 211–223.

Ferreira, D.A., Franco, H.C.J., Otto, R., Vitti, A.C., Fortes, C., Faroni, C.E., Garside, A.L., Trivelin, P.C.O., 2016. Contribution of N from green harvest residues for sugarcane nutrition in Brazil. GCB Bioenergy 8, 859–866.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., Caporaso, J.G., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. PNAS 109, 21390–21395.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44, D279–D285.

Fish, J.A., Chai, B., Wang, Q., Sun, Y., Brown, C.T., Tiedje, J.M., Cole, J.R., 2013. FunGene: the functional gene pipeline and repository. Front. Microbiol. 4.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockström, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478, 337–342.

Fowler, D., Steadman, C.E., Stevenson, D., Coyle, M., Rees, R.M., Skiba, U.M., Sutton, M.A., Cape, J.N., Dore, A.J., Vieno, M., Simpson, D., Zaehle, S., Stocker, B.D., Rinaldi, M., Facchini, M.C., Flechard, C.R., Nemitz, E., Twigg, M., Erisman, J.W., Butterbach-Bahl, K., Galloway, J.N., 2015. Effects of global change during the 21st century on the nitrogen cycle. Atmospheric Chemistry and Physics 15, 13849–13893.

Fuess, L.T., Garcia, M.L., 2014. Implications of stillage land disposal: A critical review on the impacts of fertigation. Journal of Environmental Management 145, 210–229.

Galloway, J.N., Aber, J.D., Erisman, J.W., Seitzinger, S.P., Howarth, R.W., Cowling, E.B., Cosby, B.J., 2003. The Nitrogen Cascade. BioScience 53, 341–356.

Galloway, J.N., Leach, A.M., Erisman, J.W., Bleeker, A., 2017. Nitrogen: the historical progression from ignorance to knowledge, with a view to future solutions. Soil Res. 55, 417–424.

Galloway, J.N., Townsend, A.R., Erisman, J.W., Bekunda, M., Cai, Z., Freney, J.R., Martinelli, L.A., Seitzinger, S.P., Sutton, M.A., 2008. Transformation of the Nitrogen Cycle: Recent Trends, Questions, and Potential Solutions. Science 320, 889–892.

Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food Security: The Challenge of Feeding 9 Billion People. Science 327, 812–818.

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451, 293–296.

Hénault, C., Grossel, A., Mary, B., Roussel, M., Léonard, J., 2012. Nitrous Oxide Emission by Agricultural Soils: A Review of Spatial and Temporal Variability for Mitigation. Pedosphere, Special Issue on Bioremediation of Contaminated Soil and Water 22, 426–433.

Hirsch, P.R., Mauchline, T.H., Clark, I.M., 2010. Culture-independent molecular techniques for soil microbial ecology. Soil Biology and Biochemistry 42, 878–887.

Hu, H.-W., Chen, D., He, J.-Z., 2015. Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates. FEMS Microbiol Rev 39, 729–749.

Huse, S.M., Dethlefsen, L., Huber, J.A., Welch, D.M., Relman, D.A., Sogin, M.L., 2008. Exploring Microbial Diversity and Taxonomy Using SSU rRNA Hypervariable Tag Sequencing. PLOS Genetics 4, e1000255.

Jiang, Z.-P., Li, Y.-R., Wei, G.-P., Liao, Q., Su, T.-M., Meng, Y.-C., Zhang, H.-Y., Lu, C.-Y., 2012. Effect of Long-Term Vinasse Application on Physico-chemical Properties of Sugarcane Field Soils. Sugar Tech 14, 412–417.

Jones, C.M., Graf, D.R., Bru, D., Philippot, L., Hallin, S., 2013. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink. The ISME Journal 7, 417–426.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42, D199–D205.

Kuypers, M.M.M., Marchant, H.K., Kartal, B., 2018. The microbial nitrogen-cycling network. Nature Reviews Microbiology 16, 263–276.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31, 814–821.

Lapola, D.M., Schaldach, R., Alcamo, J., Bondeau, A., Koch, J., Koelking, C., Priess, J.A., 2010. Indirect land-use changes can overcome carbon savings from biofuels in Brazil. PNAS 107, 3388–3393.

Lisboa, C.C., Butterbach−Bahl, K., Mauder, M., Kiese, R., 2011. Bioethanol production from sugarcane and emissions of greenhouse gases – known and unknowns. GCB Bioenergy 3, 277–292.

Liu, S., Han, P., Hink, L., Prosser, J.I., Wagner, M., Brüggemann, N., 2017. Abiotic Conversion of Extracellular NH2OH Contributes to N_2O Emission during Ammonia Oxidation.

Logares, R., Sunagawa, S., Salazar, G., Cornejo−Castillo, F.M., Ferrera, I., Sarmento, H., Hingamp, P., Ogata, H., Vargas, C. de, Lima−Mendez, G., Raes, J., Poulain, J., Jaillon, O., Wincker, P., Kandels−Lewis, S., Karsenti, E., Bork, P., Acinas, S.G., 2014. Metagenomic 16S rDNA Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. Environmental Microbiology 16, 2659–2671.

Lourenço, K.S., Cassman, N.A., Pijl, A.S., van Veen, J.A., Cantarella, H., Kuramae, E.E., 2018. Nitrosospira sp. Govern Nitrous Oxide Emissions in a Tropical Soil Amended With Residues of Bioenergy Crop. Front. Microbiol. 9.

Luo, C., Rodriguez-R, L.M., Johnston, E.R., Wu, L., Cheng, L., Xue, K., Tu, Q., Deng, Y., He, Z., Shi, J.Z., Yuan, M.M., Sherry, R.A., Li, D., Luo, Y., Schuur, E.A.G., Chain, P., Tiedje, J.M., Zhou, J., Konstantinidis, K.T., 2014. Soil Microbial Community Responses to a Decade of Warming as Revealed by Comparative Metagenomics. Appl. Environ. Microbiol. 80, 1777–1786.

Mosier, A.R., 1994. Nitrous oxide emissions from agricultural soils. Fertilizer Research 37, 191–200.

Orellana, L.H., Chee-Sanford, J.C., Sanford, R.A., Löffler, F.E., Konstantinidis, K.T., 2018. Year-Round Shotgun Metagenomes Reveal Stable Microbial Communities in Agricultural Soils and Novel Ammonia Oxidizers Responding to Fertilization. Appl. Environ. Microbiol. 84, e01646-17.

Otto, R., Mulvaney, R.L., Khan, S.A., Trivelin, P.C.O., 2013. Quantifying soil nitrogen mineralization to improve fertilizer nitrogen management of sugarcane. Biol Fertil Soils 49, 893–904.

Ouyang, Y., Norton, J.M., Stark, J.M., Reeve, J.R., Habteselassie, M.Y., 2016. Ammonia-oxidizing bacteria are more responsive than archaea to nitrogen source in an agricultural soil. Soil Biology and Biochemistry 96, 4–15.
Pace, N.R., 1997. A Molecular View of Microbial Diversity and the Biosphere. Science 276, 734–740.
Parkin, T.B., 1987. Soil Microsites as a Source of Denitrification Variability 1. Soil Science Society of America Journal 51, 1194–1199.
Philippot, L., Andert, J., Jones, C.M., Bru, D., Hallin, S., 2011. Importance of denitrifiers lacking the genes encoding the nitrous oxide reductase for N_2O emissions from soil. Global Change Biology 17, 1497–1504.
Pitombo, L.M., Carmo, J.B. do, Hollander, M. de, Rossetto, R., López, M.V., Cantarella, H., Kuramae, E.E., 2016. Exploring soil microbial 16S rRNA sequence data to increase carbon yield and nitrogen efficiency of a bioenergy crop. GCB Bioenergy 8, 867–879.
Prakash, T., Taylor, T.D., 2012. Functional assignment of metagenomic data: challenges and applications. Brief Bioinform 13, 711–727.
Ravishankara, A.R., Daniel, J.S., Portmann, R.W., 2009. Nitrous Oxide (N_2O): The Dominant Ozone-Depleting Substance Emitted in the 21st Century. Science 326, 123–125.
Robertson, G.P., Vitousek, P.M., 2009. Nitrogen in Agriculture: Balancing the Cost of an Essential Resource. Annual Review of Environment and Resources 34, 97–125.
Saggar, S., Jha, N., Deslippe, J., Bolan, N.S., Luo, J., Giltrap, D.L., Kim, D.-G., Zaman, M., Tillman, R.W., 2013. Denitrification and N_2O:N2 production in temperate grasslands: Processes, measurements, modelling and mitigating negative impacts. Science of The Total Environment, Soil as a Source & Sink for Greenhouse Gases 465, 173–195.
Schlesinger, W.H., 2009. On the fate of anthropogenic nitrogen. PNAS 106, 203–208.
Schmidt, M.W.I., Torn, M.S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I.A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D.A.C., Nannipieri, P., Rasse, D.P., Weiner, S., Trumbore, S.E., 2011. Persistence of soil organic matter as an ecosystem property. Nature 478, 49–56.
Seabra, J.E.A., Macedo, I.C., Chum, H.L., Faroni, C.E., Sarto, C.A., 2011. Life cycle assessment of Brazilian sugarcane products: GHG emissions and energy use. Biofuels, Bioproducts and Biorefining 5, 519–532.
Sharon, I., Banfield, J.F., 2013. Genomes from Metagenomics. Science 342, 1057–1058.
Shcherbak, I., Millar, N., Robertson, G.P., 2014. Global metaanalysis of the nonlinear response of soil nitrous oxide (N_2O) emissions to fertilizer nitrogen. PNAS 201322434.
Signor, D., Cerri, C.E.P., 2013. Nitrous oxide emissions in agricultural soils: a review. Pesquisa Agropecuária Tropical 43, 322–338.
Smith, K.A., 2017. Changing views of nitrous oxide emissions from agricultural soil: key controlling processes and assessment at different spatial scales. European Journal of Soil Science 68, 137–155.
Snyder, C.S., Bruulsema, T.W., Jensen, T.L., Fixen, P.E., 2009. Review of greenhouse gas emissions from crop production systems and fertilizer management effects. Agriculture, Ecosystems & Environment, Reactive nitrogen in agroecosystems: Integration with greenhouse gas interactions 133, 247–266.
Soares, J.R., Cassman, N.A., Kielak, A.M., Pijl, A., Carmo, J.B., Lourenço, K.S., Laanbroek, H.J., Cantarella, H., Kuramae, E.E., 2016. Nitrous oxide emission related to ammonia-oxidizing bacteria and mitigation options from N fertilization in a tropical soil. Scientific Reports 6, 30349.
Swift, M.J., Heal, O.W., Anderson, Jonathan Michael, Anderson, J. M., 1979. Decomposition in Terrestrial Ecosystems. University of California Press.
Syakila, A., Kroeze, C., 2011. The global nitrous oxide budget revisited. Greenhouse Gas Measurement and Management 1, 17–26.
Theodorakopoulos, N., Lognoul, M., Degrune, F., Broux, F., Regaert, D., Muys, C., Heinesch, B., Bodson, B., Aubinet, M., Vandenbol, M., 2017. Increased expression of bacterial amoA

during an N_2O emission peak in an agricultural field. Agriculture, Ecosystems & Environment 236, 212–220.
Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. Microbial Informatics and Experimentation 2, 3.
Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. PNAS 108, 20260–20264.
Torsvik, V., Goksøyr, J., Daae, F.L., 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. 56, 782–787.
Torsvik, V., Øvreås, L., 2002. Microbial diversity and function in soil: from genes to ecosystems. Current Opinion in Microbiology 5, 240–245.
Trivelin, P.C.O., Franco, H.C.J., Otto, R., Ferreira, D.A., Vitti, A.C., Fortes, C., Faroni, C.E., Oliveira, E.C.A., Cantarella, H., 2013. Impact of sugarcane trash on fertilizer requirements for São Paulo, Brazil. Scientia Agricola 70, 345–352.
Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C., 2008. Simultaneous Assessment of Soil Microbial Community Structure and Function through Analysis of the Meta-Transcriptome. PLOS ONE 3, e2527.
van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op den Camp, H.J.M., Kartal, B., Jetten, M.S.M., Lücker, S., 2015. Complete nitrification by a single microorganism. Nature 528, 555–559.
Yi-Ding, W., Yun-Chuan, M., Wei-Hao, W., Yang-Rui, L., Yan-Ping, Y., 2006. Effect of vinasse irrigation on the activity of three enzymes and agronomic characters at seedling stage of sugarcane. Sugar Tech 8, 264–267.
Zani, C.F., Barneze, A.S., Robertson, A.D., Keith, A.M., Cerri, C.E.P., McNamara, N.P., Cerri, C.C., 2018. Vinasse application and cessation of burning in sugarcane management can have positive impact on soil carbon stocks. PeerJ 6, e5398.
Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., Yin, Y., 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46, W95–W101.
Zhou, J., He, Z., Yang, Y., Deng, Y., Tringe, S.G., Alvarez-Cohen, L., 2015. High-Throughput Metagenomic Technologies for Complex Microbial Community Analysis: Open and Closed Formats. mBio 6, e02288-14.
Zhu, X., Burger, M., Doane, T.A., Horwath, W.R., 2013. Ammonia oxidation pathways and nitrifier denitrification are significant sources of N_2O and NO under low oxygen availability. PNAS 201219993.

Chapter 2

Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland

Noriko A. Cassman, Marcio F.A. Leite, Yao Pan, Mattias de Hollander, Johannes A. van Veen and Eiko E. Kuramae

Published as:

Cassman NA, Leite MFA, Pan Y, de Hollander M, van Veen JA & Kuramae EE, 2016. "Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland." *Scientific Reports*. 6: 23680.

Abstract

Inorganic fertilization and mowing alter soil factors with subsequent effects –direct and indirect - on above- and below-ground communities. We explored direct and indirect effects of long-term fertilization (N, P, NPK, Liming) and twice yearly mowing on the plant, bacterial and fungal communities and soil factors. We analyzed co-variation using 16S and 18S rRNA genes surveys, and plant frequency and edaphic factors across treatments. The plant and fungal communities were distinct in the NPK and L treatments, while the bacterial communities and soil factors were distinct in the N and L treatments. Plant community diversity and evenness had low diversity in the NPK and high diversity in the liming treatment, while the diversity and evenness of the bacterial and fungal communities did not differ across treatments, except of higher diversity and evenness in the liming treatment for the bacteria. We found significant co-structures between communities based on plant and fungal comparisons but not between plant and bacterial nor bacterial and fungal comparisons. Our results suggest that plant and fungal communities are more tightly linked than either community with the bacterial community in fertilized soils. We found co-varying plant, bacterial and fungal taxa in different treatments that may indicate ecological interactions.

2.1 Introduction

The plant-soil feedback drives primary productivity and is fundamental to terrestrial ecosystem functioning[1]. Plant-derived litter and rhizo-deposits present C resources to soil microbes, and microbial decomposition of plant litter and soil organic matter (SOM) releases available nutrients to plants[2]. Furthermore, the plant-soil feedback underlies plant and soil microbial (bacteria and fungi) species interactions, many mediated by nutrient quality and availability. For example, plant species select a subset of the soil microbial (bacterial and fungal) community around the roots and in the topsoil through the biochemical composition of rhizo-deposits and leaves[3-5]. Nutrient competition, especially for N[6], plays an important role in structuring bacterial communities over time in cooperation with the overlying plant species[7, 8]. Moreover, plant and microbial species may be linked through negative pathogenic interactions or positive symbiotic associations, for example those of plants with Arbuscular Mycorrhizal Fungi (AMF)[9] or N-fixing bacteria[1] to enhance plant nutrient uptake. Through micro-climatic change, such as modulation of soil pH and subsequent effects on soil nutrient status, plant species may also indirectly alter the composition of microbial communities[2]. Thus, above- and below-ground communities may have significant associations at the community level. While links between soil microbial community diversity and biomass, and plant productivity[1, 2, 10] or soil C storage[8] are known, links between above- and below-ground community compositions have not received as much attention.

Altered plant, bacterial and fungal species interactions may have consequences for ecosystem function. Long-term grassland management practices of adding nitrogen (N), potassium (K) and phosphate (P) intended to increase vegetation productivity affect the composition and diversity of the plant community and the composition of the soil bacterial community[11, 12]; this management practice can be considered a disturbance on above- and below-ground community compositions[13]. Nutrient additions can affect community compositions i) directly, e.g. by favoring species functionally adapted to the nutrient inputs and over time selecting those species so that they increase in abundance, or indirectly; these indirect effects include iia) the indirect effect of nutrient additions on soil physicochemical status, which can drive community composition changes (e.g. nutrient additions alter micronutrient availability which can structure microbial community compositions), iib) an indirect effect of nutrient inputs on the biomass of other communities which alters the composition of the original community; for example, nutrient additions increase plant productivity, which provides more C resources, accelerating soil microbial growth, and iic) an indirect effect of nutrient

inputs on the compositions of other communities inducing shifts in the original community, e.g. nutrient additions favor some plant species, which may also favor microbial symbionts of these plants.

Here, we explored the related effects of long-term nutrient addition on the composition, evenness and diversity of three grassland communities –plant and soil bacterial and fungal assemblages as measured by 16S and 18S rRNA gene sequence surveys. We examined the direct effect of nutrient addition on the plant, bacterial and fungal communities across the treatments. We hypothesized H1) that each treatment would affect plant, bacterial and fungal communities differently. To look at the indirect effect of nutrient additions through altering soil factors, we considered the co-variation in soil factor profiles and plant, bacterial and fungal communities. In addition, we included a treatment in which lime (L) was added to raise the pH of the soil to near-neutral values, consequently altering macro and micronutrients availability. We hypothesized H2) that each community – plant, bacterial and fungal – would be affected by the L treatment compared to the unfertilized treatment, confirming the indirect effect of nutrient additions through altering soil factors on each community. We examined the indirect effect of nutrients through changes in a different community, in which altering the composition or diversity of another communities alters the composition or diversity of the original community, by examining the co-variation in communities across the treatments. Because of the importance of the plant-soil feedback, we hypothesized H3) that each community would co-vary across the treatments. The third indirect effect, of plant litter structuring soil microbial community composition, was minimized by the management of removing the plant biomass. Soil bacterial (16S rRNA) and soil fungal (18S rRNA) OTU abundances, plant species frequencies and soil factor chemical profiles were analyzed using multivariate statistics. Co-variation, or co-structures, between soil factor profiles and community compositions were determined using co-inertia analysis combining all four data sets. To our knowledge, this is the first study to apply co-inertia analysis to examine above- and below-ground community links in grasslands under long-term nutrient additions. Moreover, this is the first report on fungal community compositions sampled with 18S rRNA gene marker sequencing over these long-term fertilizer treatments.

2.2 Material and Methods

2.2.1 Site description

The Ossenkampen Grassland Experiment fields were established in 1958 in species-rich meadows on heavy-clay soil in Wageningen, The Netherlands with coordinates 51 degrees 58'15"N; 5 degrees 38'18"E. Prior to the experiment, the land was grazed and had been used in alternate years for haymaking. Fertilizer treatments of chalk (L; 1000 kg CaO ha^{-1} yr^{-1}), nitrogen (N; ammonium nitrate, 100 kg N ha^{-1} yr^{-1}), phosphate (P; superphosphate, 22 kg P ha^{-1} yr^{-1}) and NP-potassium (NPK; ammonium nitrate, superphosphate and potassium sulfate, 160 kg N ha^{-1} yr^{-1}, 33 kg P ha^{-1} yr^{-1} and 311 kg K ha^{-1} yr^{-1}) were applied to duplicate 40 m^2 (16 m X 2.5 m) plots annually since 1958[34]. The control treatment fields were established without nutrient amendment. The five treatment fields were mown twice a year, in July and in October, and the biomass was removed toward the 2.5 m ends of the plots to prevent seeds from draining into the other treatment plots. Treatment and control plots were separated by unfertilized 2.5 m buffer strips, which were similarly maintained by mowing.

2.2.2 Plant sampling regime

The botanical composition was measured in September 2011, after peak plant growth[14]. Briefly, fifty samples of 25 cm^2 were clipped from each plot (2 plots X 5 treatments = 10 plant samples). The presence of each species was recorded to determine its frequency percentage (i.e. the proportion of 50 samples in which the species was present). In addition, the first, second and third most abundant species were recorded in each sample by visual estimation. The Dry Weight Rank (DWR) method was developed to estimate the species composition of grassland swards on a dry weight basis[35]. The DWR method calculates for each species its dry weight proportion (DW percentage for species A) from the percentages of cases the species takes the first (A1%), second (A2%) and third (A3%) rank. These proportions are multiplied by 0.702, 0.211 and 0.87, respectively, culminating in the following equation: DWA% = 0.702 (A1%) + 0.211 (A2%) + 0.087 (A3%). This method was tested with different sampling methods, including small samples, and it was concluded that the DWR method is well suited for studying vegetation changes in old, floristically diverse grasslands[36], such as ours. For each plant species, the functional classification was noted (grass, herb, legume or forb).

2.2.3 Soil sampling and total DNA extraction

In September 2011, nine random soil core samples (10 cm depth, two cm diameter) were taken from three independent areas within each duplicate plot of the five treatment fields. The nine soil cores were homogenized per area, sieved through 5 mm pores stored at -80 degrees C for molecular analyses (see below) or stored at -20^{o}C for one night before being sent for physico-chemical analysis as follows. The soil samples were combined per plot for physicochemical analyses for a total of 10 soil factor samples (3 areas homogenized to 1 sample X 2 plots X 5 treatment fields). These samples were dried at 60^{o}C to measure moisture content. Soil pH, extractable N, total carbon and nitrogen concentrations, available potassium, phosphate and sulfur, total organic matter and available trace elements such as Al, As, Cd, Cr, Cu, Fe, Mg, Zn, Mn, Na, Ni and Pb were analyzed at the Soil Science Department of Wageningen University (Supplementary Table S4) [37]. For the molecular investigations, total DNA was extracted from 0.3 g fractions of each soil sample (3 areas X 2 plots X 5 treatments = total 30 soil samples). The Power Soil kit (MolBio, Carlsbad, CA) protocol was followed with the modification of 5.5 m s^{-1} for 10 min. bead beating. Total DNA concentrations were measured using an ND-1000 spectrophotometer (Nanodrop, Wilmington, DE).

2.2.4 16S rRNA amplicon library preparation

The V4 region of the 16S rRNA gene marker was amplified from each sample, for a total of 30 bacterial samples (3 areas X 2 plots X 5 treatments). Amplicons for barcoded pyrosequencing were obtained using the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 806R (5'-GGACTACVSGGGTATCTAAT-3'). The 515F primer included the Roche 454-A adapter, a 10-bp barcode and a GT linker, and the reverse primer included the Roche 454-B adapter, the same 10-bp barcode as the 515F primer, and a GG linker. Amplification reactions were performed using 5 micromolar of each forward and reverse primer, 5 mM dNTPs (Invitrogen, Carlsbad, CA), 1 unit of *Taq* polymerase (Roche, Indianapolis, IN), and 1 microliter of sample DNA as the template in a total volume of 25 microliters. The PCR was conducted with an initial incubation of 5 min. at 94 ^{0}C, followed by 25 cycles of 1 min. at 94 ^{0}C, 1 min. at the annealing temperature of 53 ^{0}C and 1 min. at 72^{0}C, followed by a final incubation of 10 min. at 72 ^{0}C. Each sample was amplified in four reactions, and resulting amplification products were pooled to achieve equal mass concentrations in the final mixture.

2.2.5 18S rRNA amplicon library preparation

The 18S rRNA gene marker was amplified from two of the three soil samples per plot, for a total of 20 fungal samples (2 areas x 2 plots x 5 treatments). The 18S rRNA gene marker was amplified from the total community DNA samples using the primers FR1 (5'AICCATTCAATCGGTAIT-3') and FF390.1 (5'-CGWTAACGAACGAGACCT-3') based on published methods[38]. The Roche MID tag IDs 24 to 26 and 61 to 69 were added to barcode the samples. Amplification reactions were performed using 5 micromolar of each primer, 2 mM dNTPS (Invitrogen, Carlsbad, CA), 0.5 microliters of BSA, 10 PCR buffer, 0.56 units of Fast Start Exp-Polymerase and 1 microliter of sample DNA template in a total reaction volume of 25 microliters. The PCR was conducted with initial incubation of 5 min at 95 ^{0}C C followed by 25 cycles of 30 sec. at 95 ^{0}C, 1 min. at the annealing temperature of 57 ^{0}C, 1 min. at the extension temperature of 72 ^{0}C, followed by a final extension for 10 min. at 72 ^{0}C. Each sample was amplified in two reactions, and resulting amplification products were pooled to achieve equal mass concentrations in the final mixture.

2.2.6 16S and 18S rRNA amplicon library sequencing and processing

Amplification products were cleaned using the QIAquick PCR Purification Kit following the manufacturer's instructions (Qiagen, Valencia, CA). The purified 16S and 18S rRNA amplicon products were sequenced on the Roche 454 FLX Titanium platform (Macrogen Inc, South Korea). The 18S rRNA sequence data was processed using QIIME v1.3.0-dev on a local installation of Galaxy[39]. The 16S rRNA sequence data was processed using MOTHUR on a local 64-node server running Ubuntu (Ubuntu-14.04-trusty). Low-quality sequences that were less than 150 bp in length or that had an average quality score of less than 25 were removed. Denoising and chimera checking were accomplished using USEARCH and UCHIME[40, 41]. Operational Taxonomic Units (OTUs) were identified using USEARCH with a phylotype defined at 97% sequence similarity level. For the 18S rRNA dataset, representative phylotype sequences were taxonomically assigned against the SILVA 104 database[42] using BLAST at an e-value of 0.001. For the 16S rRNA dataset, representative phylotype sequences were assigned to the SILVA database using the RDP classifier v10.

2.2.7 Statistical analyses

Statistical analyses were conducted in R v3.1.1 (Team, R. C. R: A language and environment for statistical computing). The sequenced datasets (16S and 18S rRNA) were handled in R with the "phyloseq" package (McMurdie, P. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data). Further methods can be found in the Supplementary Methods.

2.2.7.1 Treatment effects on each component

To determine the effect of treatment on the plant (n=10), bacterial (n=27) and fungal (n=20) community compositions and the soil factor profiles (n=10), the "ade4" R package was used[43]. Between-class analysis (BCA) was applied to explore the dissimilarity between treatments within each community (plant, bacterial, fungal) or soil factors (environmental). Correspondence Analysis (CA) was applied to each community dataset while correlation-based Principal Component Analysis (PCA) was used for the soil factors[44]. A Monte-Carlo test of the treatment groups was conducted using 999 random permutations of sample rows. Ordination was used to visualize the BCA results. To determine the effect of treatment on the soil factors or community diversities (for calculations, see Supplementary Methods), multiple group comparisons were conducted in R. Of the 27 measured soil factors, the variable Cu was removed for having the same value for all treatments; similarly, the variables PO_4 and Fe were removed for containing values that fell below the detection limit for four or more treatments (Table S4). The remaining 24 soil factors were standardized and checked for normality with the Shapiro-Wilk test (normal variables: total N, Al, Mg, Mn, Na, Pb, Zn, NH_4, extractable N, C.1, OM, C.2; non-normal variables: total K, total P, As, Cd, Cr, K, Ni, P, S, NO_3, pH, moisture). The Kruskal-Wallis H or ANOVA statistical test was applied to compare the non-normal and normal variable median and means, respectively, between treatments. Post-hoc comparisons (two-tailed) were conducted using Tukey's HSD test (alpha level = 0.05) or Dunn's method (alpha level = 0.10), respectively, and boxplots were constructed.

2.2.7.2 Between-component analyses

To examine the common structure in plant, bacterial and fungal community compositions and soil factor profiles across the long-term treatments, co-inertia analysis was conducted. The "RV.test" R function was used to perform a Monte-Carlo test on the sum of eigenvalues in the co-inertia[45]. For the co-inertia analyses, the community and soil factor datasets were required to have the same number of samples per treatment. Because we had two plant samples and two soil fac-

tor samples per treatment, for these datasets we duplicated the values to result in four plant samples and four soil factor samples per treatment. First, co-inertia was performed for paired combinations of community CA ordinations and the soil factor PCA ordination, and visualizations were constructed. Additional co-inertia analyses were carried out between the community CA ordinations and PCA ordinations of three subsets of the soil factor dataset: macro-nutrients (total K, total P, extractable K, extractable P, NO_3^-, Nt, NH_4^+, extractable N, C and OM), micro-nutrients (As, Cd, Cr, Ni, Al, Mg, Mn, Na, Pb and Zn) and structural components (pH, moisture); the community-soil factor subset co-inertia analyses were visualized. To determine whether diversity of another community as well as Treatment had an effect on the composition of the original community, we tested these hypotheses using PERMANOVA ("vegan" R package).

2.3 Results

Overall, the long-term fertilization treatments resulted in considerable shifts in the soil factor profiles and plant, bacterial and fungal community compositions, but generally not in community diversity or evenness indexes, with the exception of the plant community. We found different plant and soil fungal communities under the nitrogen-phosphate-potassium (NPK) and liming (L) treatments compared to the unfertilized control (C) treatment; furthermore, we found different soil bacterial communities and soil factor profiles under nitrogen (N) and L treatments compared to the unfertilized treatment. Interestingly, the plant-fungal phylum co-structure was significant, while that of the plant-bacterial phylum-level and fungal (phylum) and bacterial (phylum) comparisons were not. From pairwise co-inertia comparisons of communities, we identified taxonomic groups that co-varied in the N, NPK or L treatments.

2.3.1Treatment effects on plant and fungal community compositions and diversity

Between-Class Analysis (BCA) revealed similar structures in the plant and fungal communities across treatments. For these communities, the unfertilized control (C), nitrogen (N) and phosphorous (P) treatments clustered together while the liming (L) and N-P-potassium (NPK) treatments clustered separately (**Figure 1**). From the BCAs of plant and fungal community compositions, 86% and 72% of total variation, respectively, could be attributed to treatment (Monte-Carlo test of groups, both $P = 0.001$).

Of the fungal samples across all treatments, 65 to 96% of the sequences could be classified at the phylum level. The five most abundant fungal groups represented an average of 96% of the proportion of classified sequences across the five treatments, and these groups were *Agaricomycotina* (average over all treatments 64%), *Saccharomyceta* (24%), "mitosporic" (3%), *Ascomycota* (3%) and *Glomerales* (2%). Fungal groups that differed significantly between the treatments were "mitosporic", *Ascomycota*, *Glomerales*, *Paraglomerales*, *Chytridiales*, *Archaeosporales*, *Lobulomycetales* and *Rhizophydiales* (Tukey-Kramer, corrected p less than 0.05; **Supplementary Table S1**).

Fungal community evenness and diversity did not differ by treatment compared to the control treatment; similarly, plant community evenness did not differ by treatment compared to the control treatment (**Supplementary Figure S1 and S2**). However, the plant communities showed the lowest diversity in the NPK treatment while the highest diversity was present in the L treatment (**Supplementary Figure S2a**). In addition to treatment and interaction with treatment, plant richness, bacterial diversity indices were significant effects on plant community compositions. For fungal community compositions, treatment and interaction with treatment as well as fungal diversity indices had significant effects (**Supplementary Table S2**).

2.3.2Treatment effects on bacterial community composition and diversity, and soil factor profiles

In contrast to the structures of plant and fungal communities across treatments that was revealed by BCA, the bacterial communities and the soil factors in the C, P and NPK treatments were grouped together while the L and N treatments each formed separate clusters (**Figure 1**). The treatments could explain 55% and 72% of total variation in the BCAs of bacterial community compositions and soil factors, respectively (Monte-Carlo tests, P = 0.001). Of the 24 soil factors, 21 differed significantly between treatments; only the soil factors As, Cr and Na did not differ between treatments (**Supplementary Figure S3**). When the soil factors were divided into subsets of micronutrients, macronutrients and structural factors BCA revealed that treatment could explain 71% (P = 0.001), 83% (P = 0.001) and 74% (P = 0.002) of the variation in each soil factor profile subset, respectively (**Supplementary Figure S4).** Furthermore, while all treatments were distinct in the macronutrient profiles, only the L treatment was distinct when looking at micronutrient and structural factor profiles.

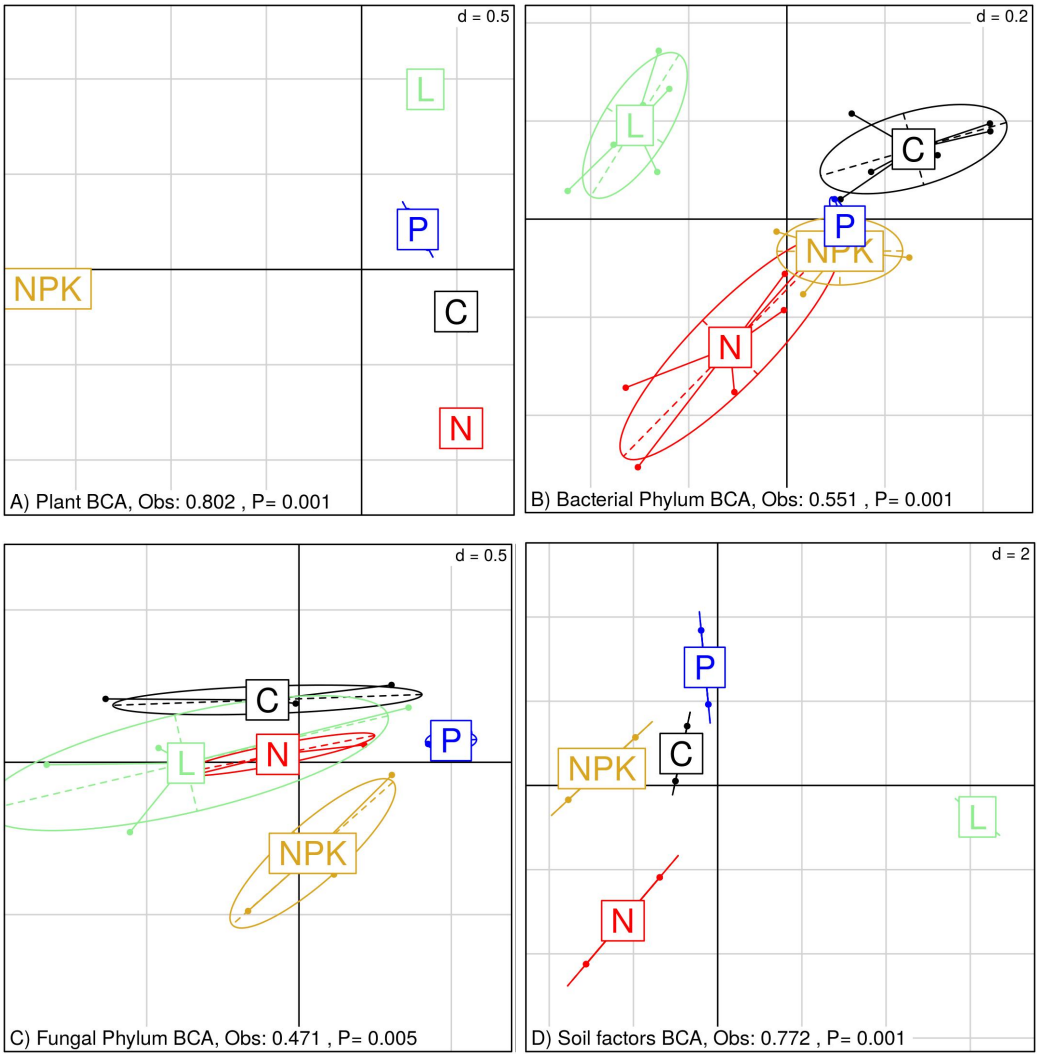


Figure 1. Between-Class Analysis (BCA) of the (A) plant, B) bacterial, C) fungal community compositions and D) soil factor factors based on correspondence analysis (A, B, C) or principal components analysis (D) over the long-term control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) treatments of the Ossenkampen experiment are presented. Group significances were assessed by Monte-Carlo tests.

Between 79 and 91% of the bacterial sequences across all treatments could be classified at the Phylum level. Of the proportion of sequences classified within bacterial phyla, 97% were distributed across the treatments within the six most abundant Bacterial phyla, *Proteobacteria* (average across treatments 40%), *Acidobacteria* (22%), *Verrucomicrobia* (15%), *Actinobacteria* (13%), *Bacteroidetes* (3%) and *Planctomycetes* (1%). Significantly different bacterial phyla between treatments were *Proteobacteria*, *Acidobacteria*, *Verrucomicrobia*, *Actinobacteria*,

Bacteroidetes, *Planctomycetes* and *Nitrospira* (Tukey-Kramer, corrected p less than 0.05, see **Supplementary Table S3**).

Bacterial community diversity and evenness was higher in the L treatment compared to the control treatment at all Renyi alpha levels but were the same among the other treatments (**Supplementary Figure S1 and S2**). In addition to treatment and interaction with treatment, bacterial diversity (Shannon and Inverse Simpson indexes) and plant richness had significant effects on bacterial community compositions (**Supplementary Table S3**).

2.3.3 Plant, bacterial and fungal community co-structures with soil factors

Co-inertia analysis allowed us to find the global similarity of the structures imposed by the long-term fertilization treatments on the grassland components. Co-structures resulting from the community-soil factor comparisons were each significant (**Table 1**). Further, we sub-categorized the soil factors into macronutrients (total K, total P, extractable K, P, NO3, total N, NH_4, extractable N, C and OM), micronutrients (As, Cd, Cr, Ni, Al, Mg, Mn, Na, Pb and Zn) and structural components (pH, moisture), and conducted co-inertia analysis between communities and the soil factor subsets. Each of the community comparisons with soil factors subsets (**Table 1**).

Table 1. Co-inertia analysis results from community-soil factors and soil factor subset comparisons from the Ossenkampen experiment (n=20 per dataset). Significance of co-structures was assessed by Monte-Carlo tests.

Co-inertia comparison		Cumulative projected inertia (%)	Observed Rv	P
Plant species	All soil factors	85	0.763	0.001
	Micronutrients	90	0.646	0.001
	Macronutrients	85	0.687	0.001
	Structural	100	0.650	0.001
Bacterial phyla	All soil factors	97	0.346	0.004
	Micronutrients	97	0.251	0.029
	Macronutrients	96	0.400	0.001
	Structural	100	0.221	0.087
Fungal phyla	All soil factors	87	0.437	0.001
	Micronutrients	89	0.300	0.046
	Macronutrients	93	0.496	0.001
	Structural	100	0.308	0.025

2.3.4 Plant, bacterial and fungal co-structures and co-variates

Co-inertia analysis identified significant co-structures between the plant and fungal communities (**Figure 2**) but not the plant and bacterial (**Figure 3**) nor the bacterial and fungal (**Figure 4**) communities. The first two co-inertia axes captured 86, 88, and 92% of the total variance in the plant-fungal, plant-bacterial and bacterial-fungal comparisons, respectively. The co-inertia analysis additionally identified taxonomic groups that most contributed to each co-inertia axis, i.e. groups that contributed the most to the total co-variance between samples (summarized in **Table 2**).

In the plant-fungal co-inertia (**Figure 2**), the second co-inertial axis separated the NPK treatment from the other treatments, and this distinction was associated with Heracsph (herb), Anthrsyl (herb), Alopepra (dominant grass) and Elymurep (grass) plant species and *Monoblepharidales*, *Basidiomycetales*, *Ascomycetales* and *Paraglomerales* fungal groups. The P treatment was separated by the first co-inertia axis, and this was driven by Cerasfon (grass), Holculan (grass), Dactyglo (grass) and Luzulcam (sedge) plant species but not clearly any fungal groups. Last, Festupra (grass), Galiumol (herb), Bromuh H (grass), Glechhed (herb), Galiuuli (herb) and Ajugarep (herb) plant species and *Pucciniomycotina*, *Cladochytriales*, *Chytridiales* and *Spizellomycetales* fungal groups co-varied in the L treatment.

In the plant-bacterial co-inertia (**Figure 3**), the NPK treatment was separated by the first co-inertia axis, and Anthrsyl (herb), Heracsph (herb), Alopepra (grass) and Elymurep (grass) plant species co-varied with *Nitrospira* bacterial phyla. The second co-inertia axis separated the N treatment from the other treatments, and this distinction was related to Carexhir (sedge), Carexdit (sedge), Juncuart (sedge) and Filipulm (herb) plant species and the *Tenericutes* bacterial phylum. Last, the L treatment grouping in the co-inertia between plant and bacterial ordinations was driven mainly by Bromuh H (grass), Galiumol (herb), Galiuuli (herb) and Glechhed (herb) plant species and WS3 and *Bacteroidetes* bacterial Phyla.

In the bacterial-fungal comparison (**Figure 4**), treatments were not distinctly clustered. However, the L treatment was weakly separated by the first co-inertial axis, and the co-variation was driven by *Pucciniomycotina*, *Cladochytriales*, *Spizellomycetales* and *Chytridiales* fungal groups and the OD1 bacterial group. Furthermore, the N treatment was weakly separated by the second co-inertial axis, and this separation was driven by variation in the *Tenericutes* bacterial Phylum and the fungal groups *Ustilaginomycotina* and *Rhizophydiales*.

Table 2. List of co-variate taxonomic groups from pairwise co-inertia analysis of plant, bacterial and fungal community composition from the Ossenkampen experiment. Taxonomic groups that contributed to the co-inertia axes clearly separating the nitrogen (N), N-phosphate-potassium (NPK), P and chalk (L) treatments are listed. Plant functional group information is included (s=sedge, h=herb, g=grass).

Community	Co-variates within treatment			
	N	NPK	P	L
Plant	Filipulm (h) Carexhir (s) Carexdit (s) Juncuart (s)	Heracsph (h) Anthrsyl (h) Alopepra (g) Elymurep (g)	Cerasfon (g) Holculan (g) Dactyglo (g) Luzulcam (s)	Galiuuli (h) Glechhed (h) Ajugarep (h) Galiumol (h) Bromuh H (g) Festupra (g) Poa..tri (g)
Bacteria	Tenericutes	Nitrospira	--	Bacterioidetes WS3 OD1
Fungus	Ustilaginomycotina Rhizophydiales	Paraglomerales Basidiomycetales Ascomycetales Monoblepharidales	--	Pucciniomycotina Cladochytriales Chytridiales Spizellomycetales

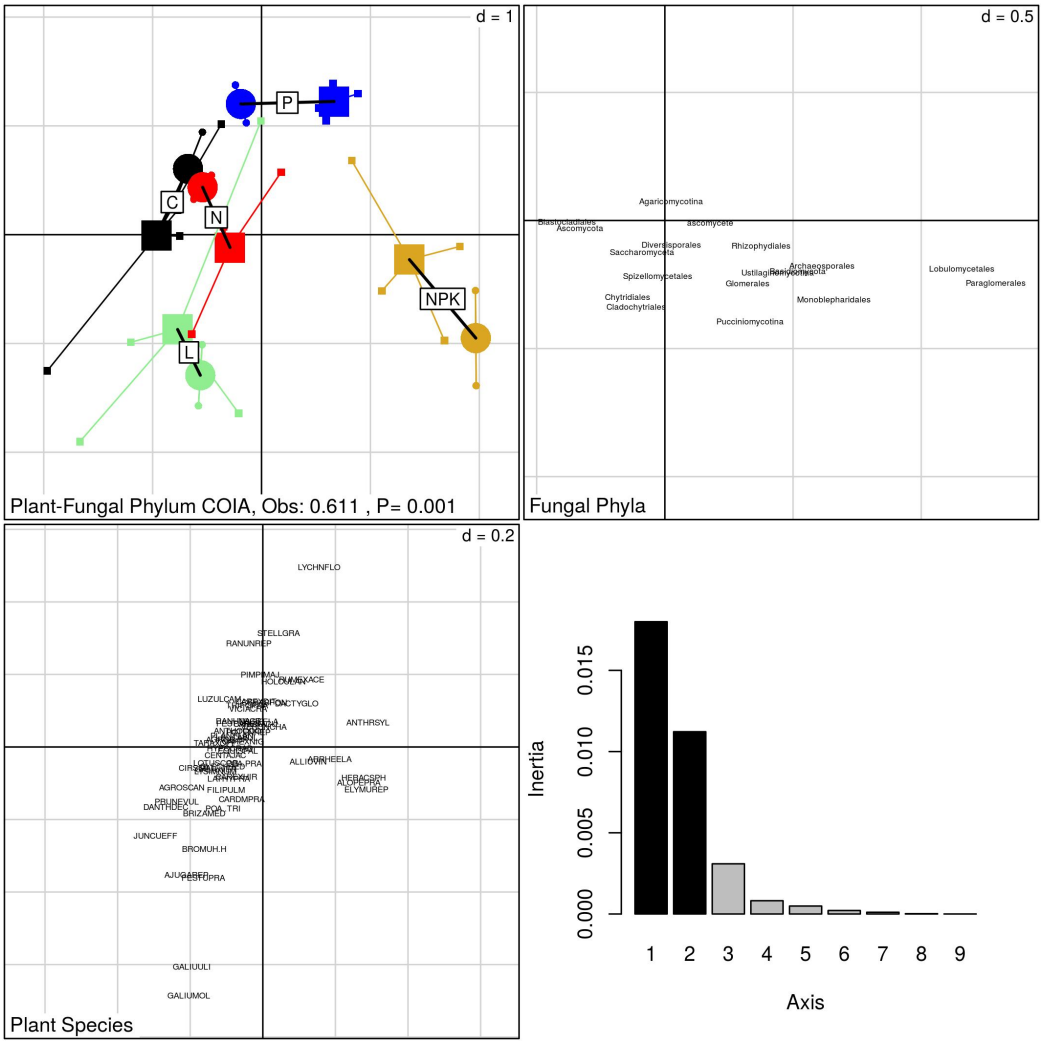


Figure 2. Co-inertia analysis (COIA) between correspondence analysis of plant and fungal community composition across the long-term control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) treatments of the Ossenkampen experiment (cumulative projected inertia = 86%). Significance of co-structure was assessed by a Monte-Carlo test. Circle = Plant; Square = Fungi.

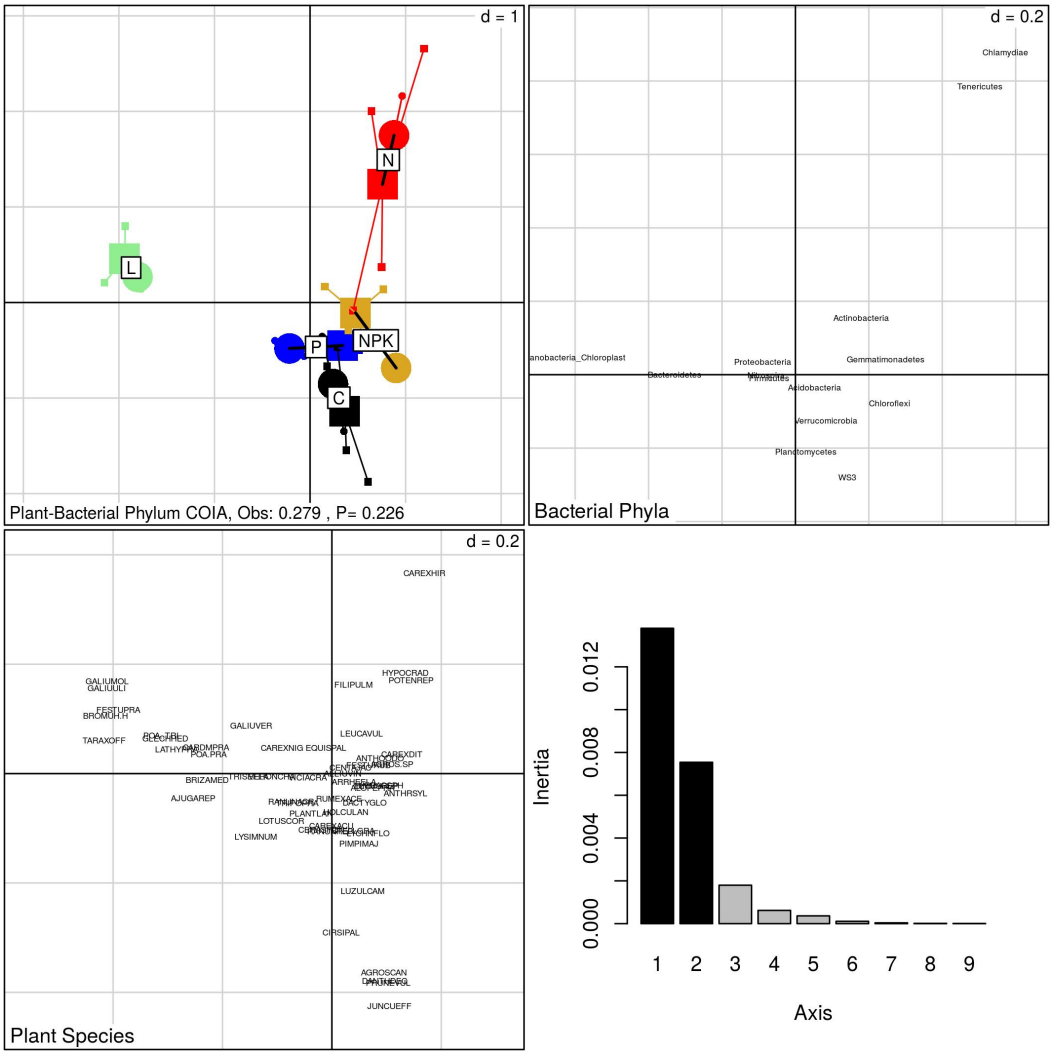


Figure 3. Co-inertia analysis (COIA) between correspondence analysis of plant and bacterial community composition across the long-term control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) treatments of the Ossenkampen experiment (cumulative projected inertia = 88%). Significance of the co-structure was assessed by a Monte-Carlo test. Circle = Plant; Square = Bacteria.

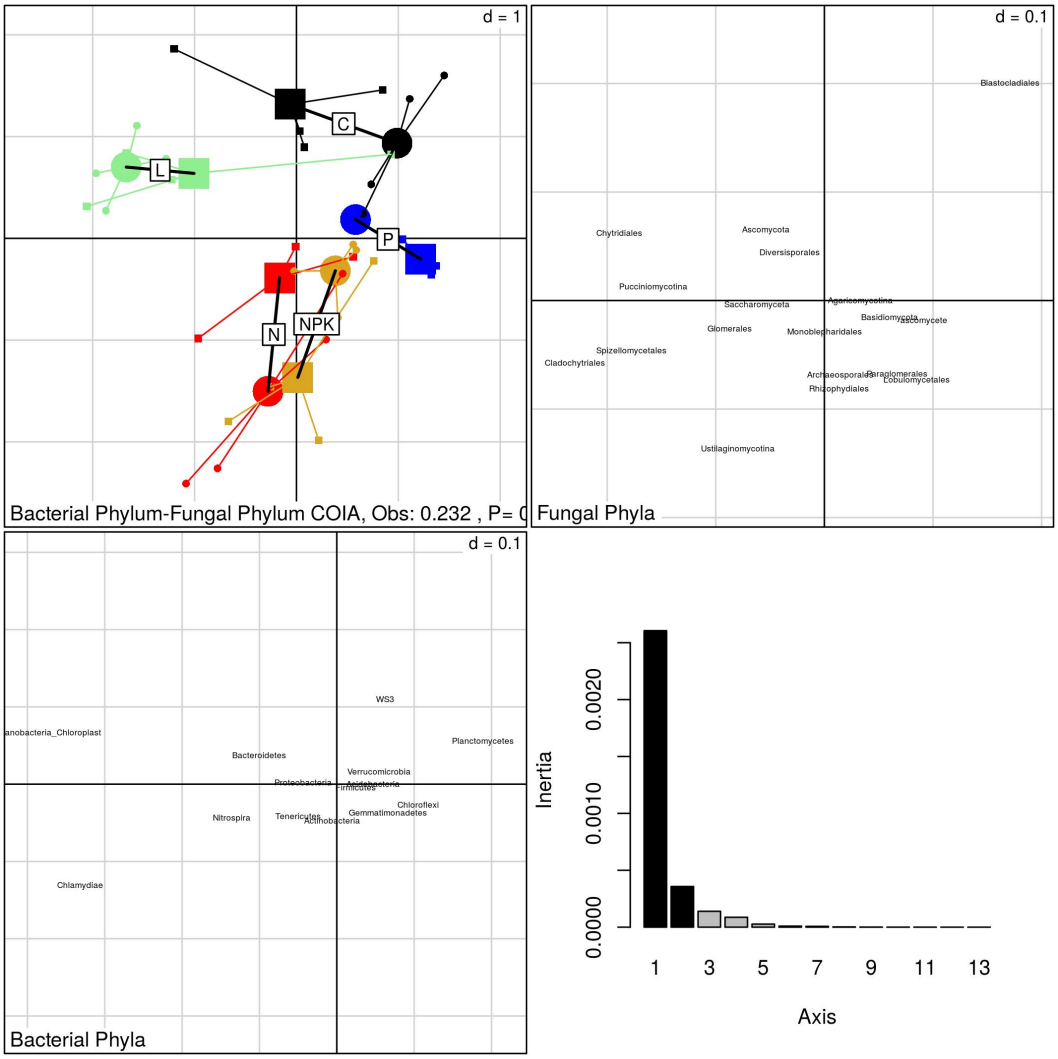


Figure 4. Co-inertia analysis (COIA) between correspondence analysis of bacterial and fungal community composition across the long-term control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) treatments of the Ossenkampen experiment (cumulative projected inertia = 92%). Significance of the co-structure was assessed by a Monte-Carlo test. Circle = Bacteria; Square = Fungi.

2.4 Discussion

Here we explored indirect effects (pH and composition of other communities but not biomass of the plant community) and the direct effect of nutrient additions on the composition of three communities (plant, soil bacterial and soil fungal) of improved grassland. Changes in the relative abundances of many taxonomic groups within one community resulted in community-level composition changes in some treatments. Our first hypothesis was supported, in that the bacterial communities were altered in the nitrogen (N) treatment but not phosphorus (P) nor NP-potassium(K) treatments, and the plant and fungal communities were altered in the NPK but not the P nor N treatments, compared to the unfertilized treatment. Combining the treatment groupings of each grassland community and the links between each community and macronutrients, we confirmed the direct effect of the nutrient additions on the communities.

The nutrient addition and mowing management practice provided selective pressures that drove changes in the relative abundances of some plant, bacterial and fungal groups. The treatments represented new habitats that allowed some functional types to succeed; for instance, in the NPK treatment, faster-growing individual plants better able to compete for light and uptake nutrients were most successful[14]. This trait selection led to changes in the relative abundances of grasses in the NPK treatment. We found a direct effect of the long-term N fertilizer treatment on the bacterial communities, compared to the unfertilized treatment, while no direct effect was found in the NPK treatment despite an equivalent rate of N application. In our previous work and here, we found so-called copiotrophic bacterial groups that were functionally adapted to the N additions to increase in relative abundance in the N treatment, which has been found at other grassland sites under long-term N additions[15, 16]. Furthermore, the NPK treatment, but not the N treatment, had a direct effect on plant and fungal community compositions compared to the unfertilized treatment.

Community compositions can be altered indirectly by nutrient additions through the effect on pH. The liming (L) treatment changed the soil pH, which alters other soil factors, such as Al concentration and micronutrient availability[17]. Thus, any changes in L community composition were due to the increase in pH and mowing management practice. Our second hypothesis, that each community would be affected by the L treatment when compared to the unfertilized control treatment, and that each community would show a link to soil factors, was supported. Each community was altered in the L treatment compared to the unfertilized treatment, supporting pH changes and related edaphic factor changes as a major driver of bacterial and fungal composition changes[18-20]. As

shown previously, the near-neutral pH of the L treatment soils increased nutrient availability resulting in native plant species recolonizing the L plots and overall increased plant richness[14]. The L treatment was characterized by high plant diversity and evenness and additionally by non-pathogenic nematode and earthworm biomass, which characterize fertile (e.g. productive) grasslands[21].

Here we considered that some of the changes in the relative abundances of taxonomic groups reflected not only functional adaptation to the treatment but a link with taxonomic groups from other communities. Therefore, we explored the indirect effect of nutrient additions on the composition of a community through the composition of another community. In other words, these were community links apart from the direct effects of the nutrient additions. Most interestingly, our third hypothesis, that each community would be significantly linked to each other, was not supported. Community-level composition links were not found for plant and bacterial, and fungal and bacterial communities. Thus, the indirect effect of nutrient additions altering community compositions through changes in other community compositions was supported between plant and fungal (phylum) communities, and between each community when the soil microbial communities were compared at a lower taxonomic level. Recent work has found that the beta diversity, or site differences, of grassland soil microbial communities are predicted by the beta diversity of the overlying plant communities[12]. The authors also found that the beta diversities of plants and soil microbes were driven mutually by mean annual temperature and the C:N ratio. We speculate that our NPK bacterial communities may have shifted in association with the plant community if the plant biomass had remained on the fields. In support, Millard and Singh[22] suggested that while plant community composition drives fungal community composition, the bacterial community structure is more influenced by SOM quality, hence driven indirectly by plant community composition via plant biomass.

The acquisition-conservation trade-off hypothesis states that under soil resource limitation, the plant community is more dependent on the below-ground community for nutrients, resulting in more ecological links between plant and microbial species[23]. Conversely, during high nutrient availability, fewer links should be found between the plant and soil microbial communities, with plants less dependent on nutrients from fungal or bacterial sources. Here, plant community diversity decreased in the NPK treatment compared to the unfertilized treatment and grasses were better adapted to the high nutrient conditions. Consequently, fungal groups associated with the grasses might have shifted under the high nutrient conditions, thereby changing fungal community composition in the NPK treatment. Selection of the grasses on fungal groups that increase grass root nutri-

ent uptake may have occurred, although whether the plant-fungal link describes positive or negative interactions is beyond the scope of the current study.

One interesting NPK fungal co-variate was *Paraglomerales*, which are a fungal group that includes arbuscular mycorrhizal fungi (AMF); these fungi form hyphal networks in association with plant roots to aid plant absorption of P. There is evidence that under conditions of high P availability, plants form fewer AMF associations because available P is directly absorbed from soil[24, 25], and this was supported in our results as the P treatment fungal communities had the lowest proportions of AMF compared to the other treatments. However, in the NPK treatment with N and K added, there was a high soil Al-content, which may block P uptake by plant by fixing it in the soil[26]. Thus, the NPK plants may be investing in AMF associations in order to improve soil P uptake despite the high Al content. Alternately, plant-AMF interactions can become competitive under high-nutrient conditions, and these results may indicate a negative interaction[27, 28].

An interesting bacterial NPK co-variate was the *Nitrospira* bacterial phyla, which are a group of autotrophic, nitrite-oxidizing bacteria[29]. This group was abundant in the NPK treatment but not in the N treatment, compared to the control treatment. *Nitrospira* are classified as K-strategists, that is, slow-growing, oligotrophic and with low affinity for N substrates[30]. Similarly, *Nitrospira* relative abundances decreased in grasslands fertilized for 27 years and agricultural fields under 8 years of N fertilization consistent with the rate in our plots[15]. However, under high nutrient fertilization, plant competition for macro-nutrients with bacteria may keep the soil nutrient status poor; thus, we hypothesize that the *Nitrospira* co-varied with NPK plants due to the relatively low-nutrient conditions available in these soils for the bacterial community, in contrast to the high N conditions available in the N treatment.

In the N treatment comparisons, microbial co-variates included potential pathogenic taxa. For instance, the N treatment co-variate *Rhizophydiales* are a zoosporic fungal taxa that are found in soils as pathogens and decomposers[31]. The *Tenericutes* bacterial Phyla encompasses the phytoplasmas, which are regarded as plant pathogens, infecting up to 98 plant families[32]. The *Ustilaginomycotina* are a group of plant parasitic fungi, also known as smut fungi[33]. These results suggest that plant health may be negatively impacted under N fertilization, with consequences on the long-term stability of plant communities regularly fertilized with N.

The current study presents a way to investigate concurrently the plant, bacterial and fungal communities using a not-widely used statistical analysis. However, we acknowledge that our results are limited by the low statistical power in-

duced by having only two plots per treatment for the plant and soil factor data. We sampled soil samples from independent sampling areas within each of the two plots, resulting in two samples within each plot for the fungal data (n=20), and three samples within each plot for the bacterial data (n=27).

In summary, we explored direct (long-term fertilizer treatments) and indirect (pH and composition of another community) effects of nutrient additions on plant, soil microbial community compositions from grassland. This is, to our knowledge, the first study of the topsoil fungal community composition to long-term inorganic fertilization treatments in grasslands using the 18S rRNA gene marker. In addition, this is the first study to examine concurrently the above- and below-ground community compositions in an improved grassland using co-inertia analysis. Nitrogen treatment had a direct effect on bacterial community compositions and soil factors while NPK treatment had a direct effect on plant and fungal community compositions. Co-inertia results highlighted the link between plant and fungal community compositions, suggesting that indirect effect of nutrient additions on plant community compositions are observed due to fungal community compositions, or vice versa, while the same is not true for plant and bacterial communities, nor for fungal and bacterial communities. However, there was also an effect between plant and bacterial community diversities impacting the plant and bacterial compositions, respectively, suggesting that these communities also are linked. To examine the indirect effects of nutrient additions on the grassland communities, we necessarily used symmetric co-inertia analyses; therefore, it should be emphasized that these results imply correlations and not causation. In addition to community-level links, we found potential association between plant, bacterial and fungal taxonomic groups in the N, NPK and L treatments that can be explored in future studies.

2.5 Declarations

Acknowledgments

The authors thank Robert HEM Geerts and H Korevaar for assistance with plant and soil sampling and analysis. This work was supported by The Netherlands Organization for Scientific Research (NWO) and FAPESP grant number 729.004.003. Publication number 6044 of the NIOO-KNAW, Netherlands Institute of Ecology.

Author contribution statement

NAC, JAV and EEK designed the study. YP and EEK collected the data. NAC, MH and EEK processed the data. NAC and MFAL performed data analysis. NAC wrote the paper with contributions from MFAL, EEK and JAV. All authors discussed the results and commented on the manuscript.

Competing interests statement

The authors declare no competing financial interests.

Accession codes

European Nucleotide Archive study accession number PRJEB11582.

2.6 Supplementary Material

Supplementary Table S1. Fungal community compositions in the long-term unfertilized control (C), liming (L), nitrogen (N), N-phosphate-potassium (NPK) and P treatments of the Ossenkampen experiment. Fungal phyla were included if the average proportion of classified sequences was above 1% in at least one treatment.

Fungal Phyla	Effect size	Average proportion of classified sequences in treatments (%)				
		C	L	N	NPK	P
Agaricomycotina	0.506	61±17	46±21	62±10	63±9	89±3
Saccharomyceta	0.405	28±15	32±16	30±10	22±7	7±2
"mitosporic"+	0.679	2±2 a*	8±3 b	1±0 a	5±2 ab	1±0 a
Ascomycota	0.740	6±2 a	5±0 a	3±1 ab	1±1 b	1±1 b
Glomerales	0.682	1±0 a	3±2 b	2±1 ab	4±1 b	1±0 a
Paraglomerales	0.768	a	a	a	1±1 b	a
Chrytidiales	0.608	a	2±1 b	a	a	a
Pucciniomycotina	0.208		1±2		1±1	

+bolded phyla indicate significant difference among treatments (ANOVA, corrected $p<0.05$ from STAMP analysis)
*Similar letters represent no significant differences between treatments (Tukey-Kramer, 95% CI, Benjamini-Hochberg FDR multiple test correction)

Supplementary Table S2. PERMANOVA test results of treatment and diversity effects on plant, bacterial and fungal community compositions in the Ossenkampen experiment.

Community	Source of Variation	Pseudo-F	p-value
Plant species	**Treatment**	**3.1444**	**0.045**
	Plant Richness	**14.9652**	**0.001**
	Interaction	**3.4186**	**0.043**
	Treatment	**60.824**	**0.001**
	Bacterial Shannon	**3.975**	**0.014**
	Interaction	**3.371**	**0.002**
	Treatment	**51.359**	**0.001**
	Bacterial Inverse Simpson	**2.901**	**0.053**
	Interaction	**2.571**	**0.016**
	Treatment	**42.879**	**0.001**
	Fungal Shannon	1.048	0.351
	Interaction	**2.077**	**0.041**
	Treatment	**35.710**	**0.001**
	Fungal Inverse Simpson	0.862	0.450
	Interaction	1.314	0.245
Bacterial phyla	**Treatment**	**33.772**	**0.001**
	Plant Richness	**15.379**	**0.001**
	Interaction	**6.753**	**0.001**
	Treatment	**20.4142**	**0.001**
	Bacterial Shannon	**8.8261**	**0.001**
	Interaction	**3.2105**	**0.009**
	Treatment	**13.0175**	**0.001**
	Bacterial Inverse Simpson	**6.1752**	**0.006**
	Interaction	1.0046	0.482
	Treatment	**9.1368**	**0.001**
	Fungal Shannon	0.9903	0.395
	Interaction	0.7958	0.643
	Treatment	**8.5066**	**0.001**
	Fungal Inverse Simpson	0.7469	0.473
	Interaction	0.6123	0.784
Fungal phyla	**Treatment**	**5.3954**	**0.007**
	Plant Richness	0.6262	0.458
	Interaction	**3.1178**	**0.048**
	Treatment	**3.3005**	**0.033**
	Bacterial Shannon	0.6617	0.464
	Interaction	0.8668	0.548
	Treatment	**3.2082**	**0.057**
	Bacterial Inverse Simpson	0.4438	0.583
	Interaction	0.8226	0.582
	Treatment	**9.3895**	**0.002**
	Fungal Shannon	3.6953	0.067
	Interaction	**6.6252**	**0.005**
	Treatment	**9.0983**	**0.001**
	Fungal Inverse Simpson	**5.4546**	**0.023**
	Interaction	**5.8736**	**0.004**

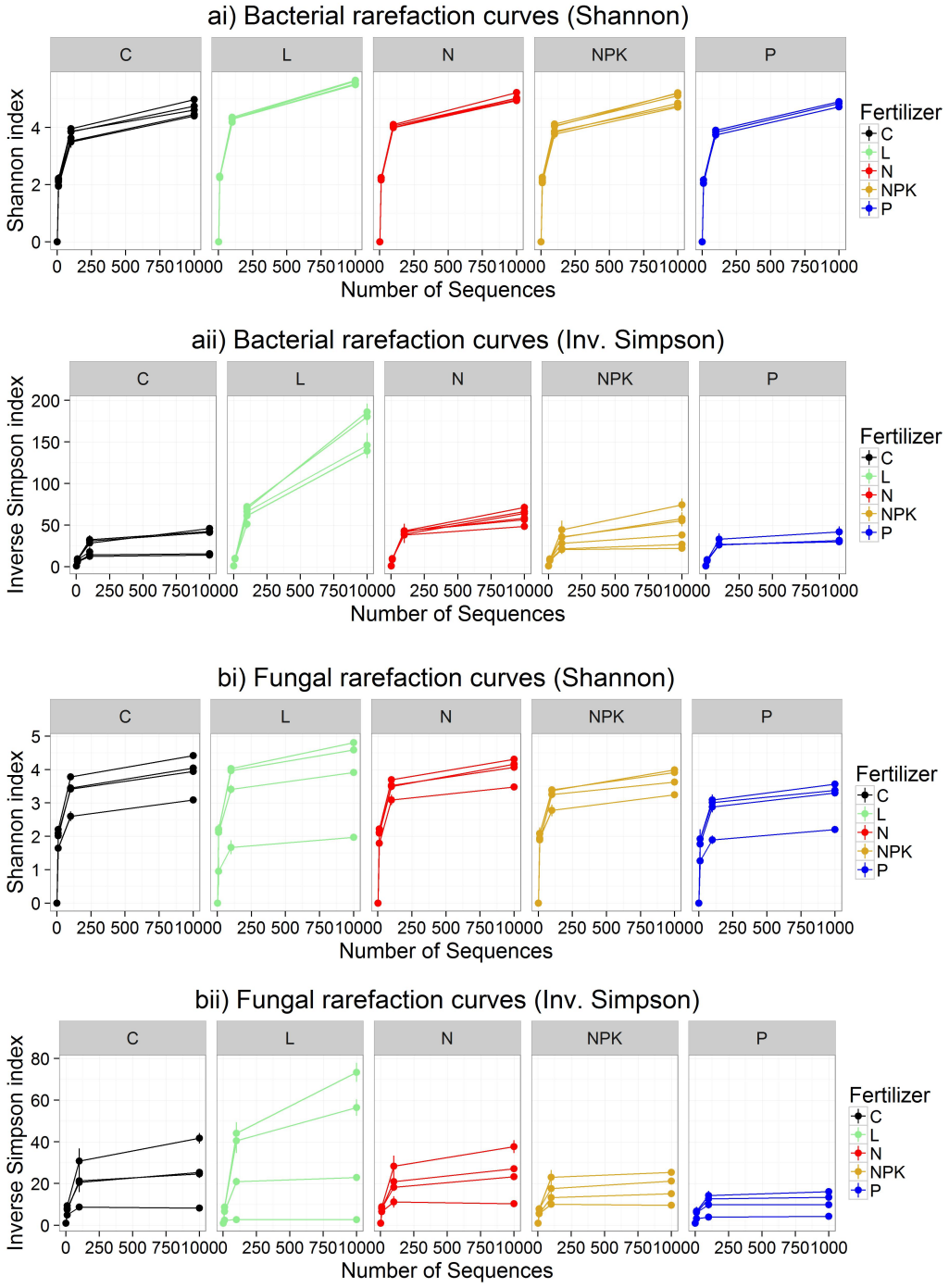
Supplementary Table S3. Bacterial community compositions in the long-term unfertilized control (C), liming (L), nitrogen (N), N-phosphate-potassium (NPK) and P treatments of the Ossenkampen experiment. Phyla were included if the average proportion of classified sequences was above 1% in one or more treatments.

Bacterial Phyla	Effect size	Average proportion of classified sequences (%)									
		C		L		N		NPK		P	
Proteobacteria+	0.836	35±2	a*	51±1	c	41±4	b	38±3	ab	37±2	ab
Acidobacteria	0.501	28±4	a	19±4	b	20±3	b	23±2	ab	23±2	ab
Verrucomicrobia	0.642	24±7	ac	8±2	b	13±4	bc	19±4	a	20±2	a
Actinobacteria	0.675	5±2	a	9±3	ab	18±5	c	12±2	b	11±1	ab
Bacteroidetes	0.900	3±0	a	8±1	b	2±1	a	2±1	a	3±0	a
Firmicutes	0.285	2±0		3±1		2±1		2±1		3±0	
Planctomycetes	0.709	2±0	abc	1±0	ab		ac	1±0	abc	1±0	ab
Gemmatimonadetes	0.527	1±0				1±0		1±1		1±0	

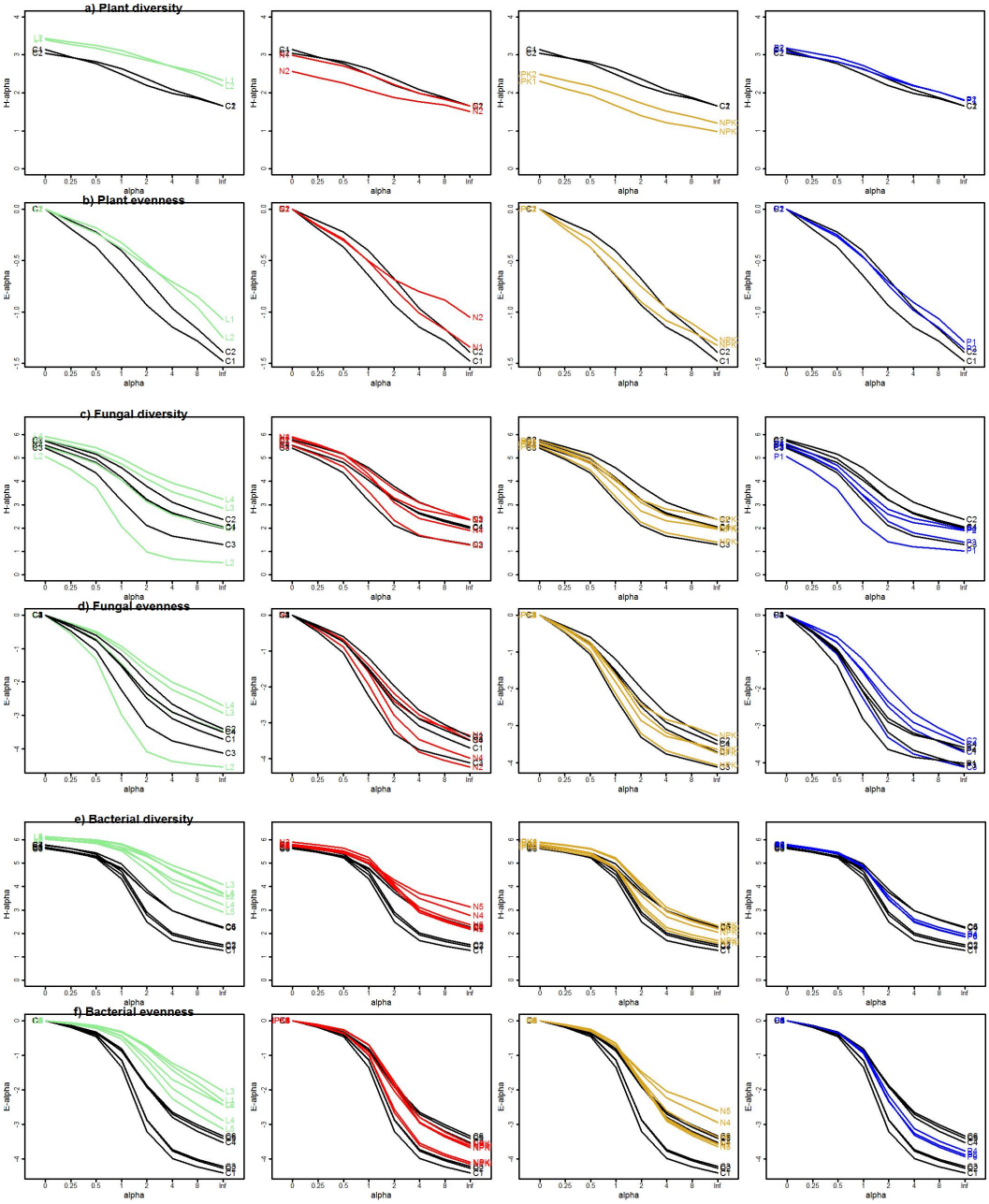
+bolded phyla indicate significant difference among treatments (ANOVA, corrected p<0.**05** from STAMP analysis)
*Similar letters represent no significant differences between treatments (Tukey-Kramer, 95% CI, Benjamini-Hochberg FDR multiple test correction)

Supplementary Table S4. Additional information regarding the soil factors measured in the Ossenkampen experiment. Horizontal lines group measurements from the same method.

Soil Factor	Factor ID	Units	Method	Notes
Total K	Kt	mg/kg	F-AES	Aqua Regia digestion
Total N	Nt	g/kg	SFA-Nt/Pt	H_2SO_4-H_2O_2-Se digestion
Total P	Pt	mg/kg		
Al	Al	mg/kg	ICP-AES Thermo	0,01M $CaCl_2$ extraction
As	As	mg/kg	ICP-AES Thermo	
Cd	Cd	mg/kg	ICP-AES Thermo	
Cr	Cr	mg/kg	ICP-AES Thermo	
Cu	Cu	mg/kg	ICP-AES Thermo	Factor removed; values all the same
Fe	Fe	mg/kg	ICP-AES Thermo	Factor removed; values below detection limi
Extractable K	K	mg/kg	ICP-AES Thermo	
Mg	Mg	mg/kg	ICP-AES Thermo	
Mn	Mn	mg/kg	ICP-AES Thermo	
Na	Na	mg/kg	ICP-AES Thermo	
Ni	Ni	mg/kg	ICP-AES Thermo	
Extractable P	P	mg/kg	ICP-AES Thermo	
Pb	Pb	mg/kg	ICP-AES Thermo	
S	S	mg/kg	ICP-AES Thermo	
Zn	Zn	mg/kg	ICP-AES Thermo	
NH_4^+	NH_4^+	mg/kg	SFA-$CaCl_2$	
NO_3^-	NO_3^-	mg/kg	SFA-$CaCl_2$	
Extractable N	Nts	mg/kg	SFA-$CaCl_2$	
PO_4	--	mg/kg	SFA-$CaCl_2$	Factor removed; values below detection limi
C	C.1	mg/kg	SFA-TOC	
Organic matter	OM	%	baking oven	Loss on ignition (105-550°C)
C	C.2	g/kg	spectrophotometer	Kurmies
pH	pH	at 20±1˚C	pH-meter	pH-H_2O
Moisture	Moisture	dry-matt	dry matter	moisture determination

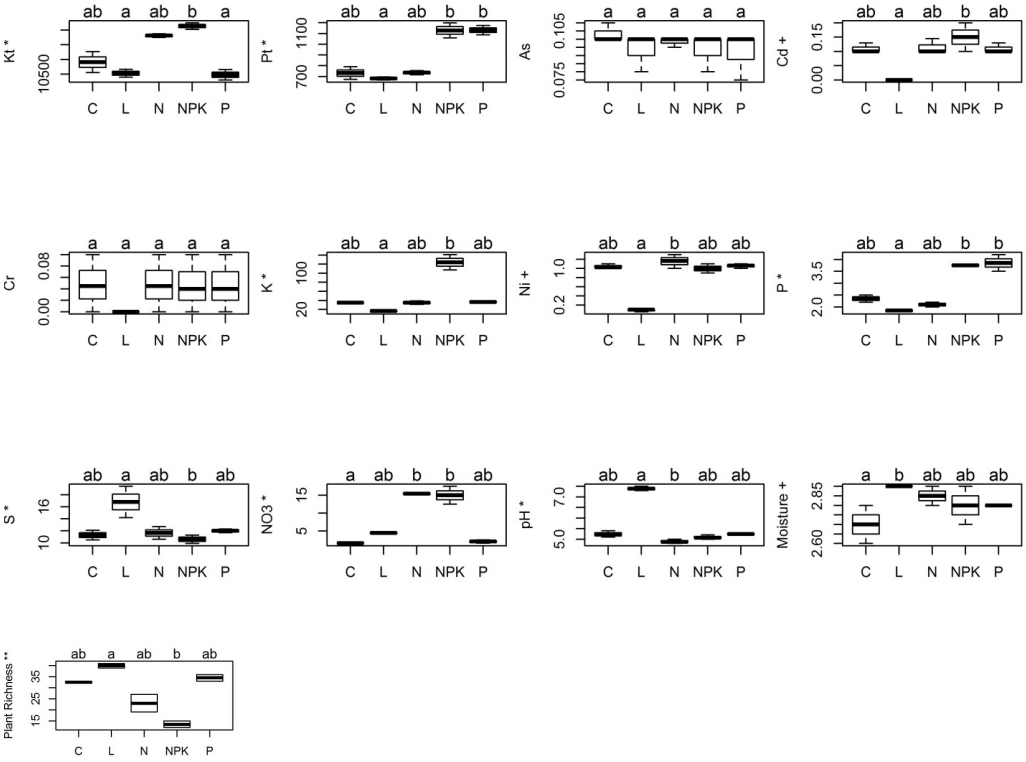


Supplementary Figure S1. Rarefaction curves of the sequenced communities from the Ossenkampen experiment, including the a) bacterial and b) fungal communities, are presented using **(i)** Shannon and **(ii)** Inverse Simpson diversity indices. Legend: control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK and phosphorus (P) fertilizer treatments.



Supplementary Figure S2. Renyi diversities and evenness of a-b) plant (n=10), c-d) fungal (n=20) and e-f) bacterial (n=27) communities in the Ossenkampen experiment. Legend: control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) fertilizer treatments.

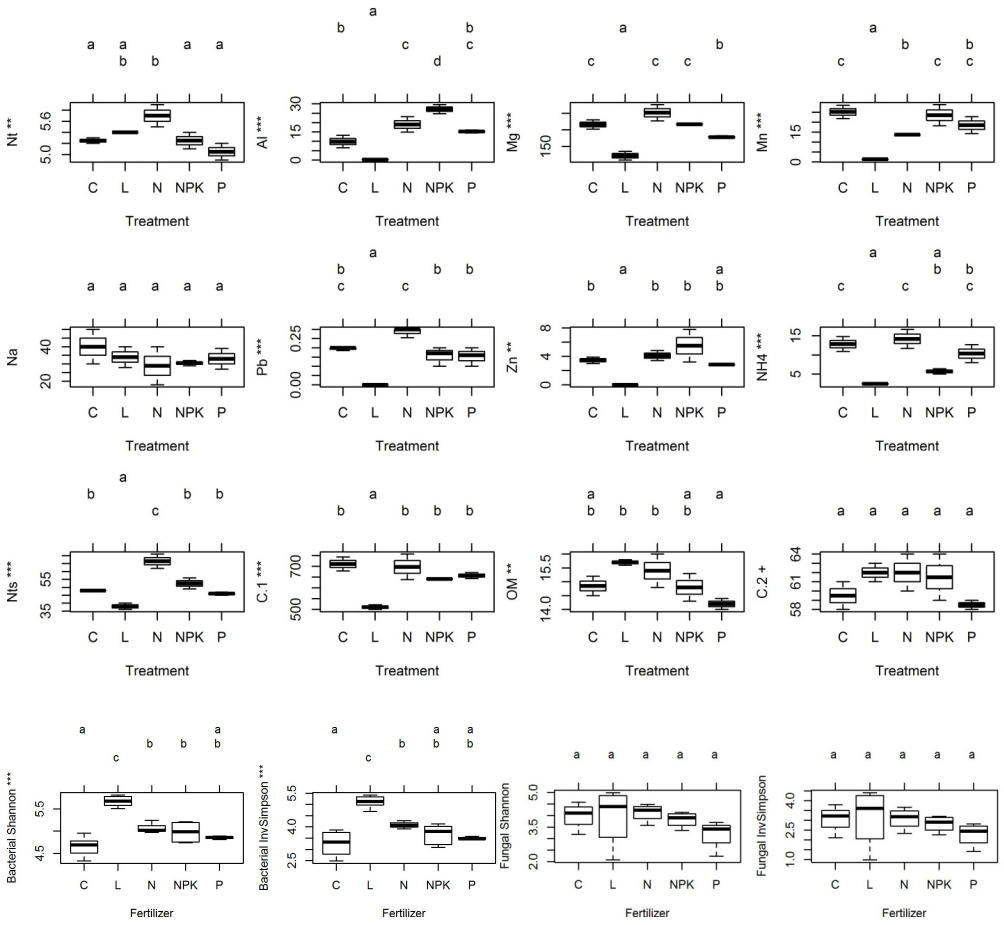
a) Non-normal soil factors and diversity indices



Kruskal-Wallis H test significance: ** p<0.01, * p < 0.05, + p < 0.1
Dunn's post-hoc test significance: similar letters mean no difference at alpha level 0.10

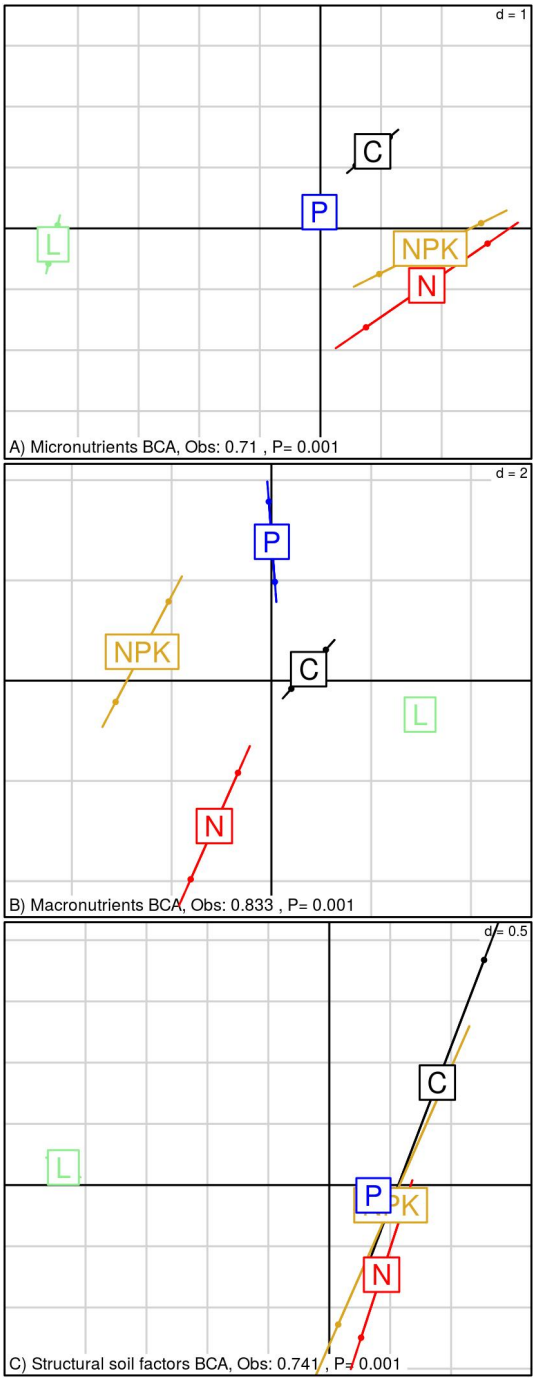
Supplementary Figure S3. Boxplots of the a) non-normal soil physicochemical parameters and community diversity indices in the Ossenkampen experiment. Legend: control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) fertilizer treatments. Asterisks by variable names indicate significantly different mean or median values across all treatments from Kruskal-Wallis or ANOVA tests, respectively. Soil factor identifiers (IDs) are listed in **Supplementary Table S4**.

b) Normal soil factors and diversity indices



ANOVA test significance: *** p < 0.001, ** p = 0.001, * p = 0.01, + p = 0.1
Tukey-Kramer post-hoc test significance: similar letters represent no difference at alpha < 0.05

Supplementary Figure S3 cont'd. Boxplots of the b) normal soil physicochemical parameters and community diversity indices in the Ossenkampen experiment. Legend: control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) fertilizer treatments. Asterisks by variable names indicate significantly different mean or median values across all treatments from Kruskal-Wallis or ANOVA tests, respectively. Soil factor identifiers (IDs) are listed in **Supplementary Table S4**.



Supplementary Figure S4. Between-Class Analysis (BCA) of the soil factor subsets (A) micronutrients, B) macronutrients, C) structural) over the long-term control (C), liming (L), nitrogen (N), nitrogen-potassium-phosphorus (NPK) and phosphorus (P) treatments of the Ossenkampen experiment are presented. Significance of groups was assessed by Monte-Carlo tests.

2.7 Supplementary Methods

2.6.1 Diversity and evenness calculations. Renyi diversities and evenness were calculated and visualized from the fungal (n=20), bacterial (n=27) datasets and the plant species (n=10) frequency dataset using the "BiodiversityR" R package. Plant richness was measured during sampling as the total number of species present. To obtain alpha diversity indices for the sequenced datasets, the bacterial and fungal OTU tables were rarified to the size of the smallest sample in each dataset. For the bacterial and the fungal samples, Renyi diversity indices at alpha = 1 (Shannon) and alpha = 2 (Inverse Simpson) were kept for group testing.

2.6.2 STAMP analysis. The fungal (n=20) and bacterial (n=27) datasets were agglomerated at the taxonomic rank of Phylum for the STAMP[1] analysis. Missing taxonomy information in the 18S rRNA dataset was resolved as follows: 1) if an entry was blank with classified entries before and after, then the blank was replaced with a copy of the entry after (eg. "Eukaryota", "", "Fungi", became "Eukaryota", "Fungi", "Fungi"), 2) if an entry was blank and the previous or following entry was "unclassified", it was replaced with "unclassified" (eg. "Eukaryota", "Fungi", "", "unclassified" became "Eukaryota", "Fungi", "unclassified", "unclassified") and 3) if an entry was blank and the previous entry was "Fungi" and the next entry was "fungal", the blank space was replaced with "Fungi." In addition, for the 18S rRNA dataset, ambiguous classifications (e.g. "environmental" and "unknown") were replaced with "unclassified". Unclassified reads were removed; then, the ANOVA statistical test was selected with a Tukey-Kramer post-hoc test (CI= 95%). A Benjamini-Hochberg FDR multiple test correction was applied.

2.6.3 Treatment effects on plant, bacterial and fungal communities and soil factor profiles. Between-Class Analysis (BCA) selects the orthogonal axis that maximizes between-group variance and measures the amount of variance restricted to the grouping factor as a percentage of the inertia captured through the new axis over the total inertia[2]. This allowed us to assess the amount of variability that could be explained by treatment for each community. The bacterial and fungal OTU abundances were summarized at the Phylum and Genus, or Phylum and Class levels, respectively. The plant species frequencies and the bacterial and fungal abundances were converted to relative abundances (contingency tables), and the soil factors were normalized and scaled to unit variance.

2.6.4 Between-component analyses. Co-inertia analysis is a multivariate method that identifies the common structure in two tables related by the same samples[3]. We chose this analysis because it can tolerate a high variable-to-sample ratio. Furthermore, we leveraged the imposed structure of the long-term fertilizer treatments. The co-inertia of two ordinations is the sum of squares of the co-variances of the variables in each table; thus, co-varying variables can be identified through correlation with the co-inertia axis[3]. Here we could simultaneously identify, for example, the taxonomic groups from two communities that contributed to treatment groupings in the factor map. The degree of multivariate co-variance between two ordinations is summarized within an array correlation and co-variance coefficient, the RV value, which is a measure of the global similarity.

2.6.5 Supplementary Methods References

1. Parks, D. H., Tyson, G. W., Hugenholtz, P. & Beiko, R. G. Stamp: statistical analysis of taxonomic and functional profiles. *Bioinformatics* **30**, 3123–3124 (2014).
2. Dray, S. & Dufour, A. B. The ade4 package: implementing the duality diagram for ecologists. *J. Statist. Software*, ***22***:1-20 (2007).
3. Dray, S., Chessel, D. & Thioulouse, J. Co-inertia analysis and the linking of ecological tables. *Ecology* **84**: 3078-3089 (2003).

2.8 References

1. van der Heijden, M.G.A., R.D. Bardgett, and N.M. van Straalen, *The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems.* Ecology Letters, 2008. **11**(3): p. 296-310.
2. Wagg, C., et al., *Soil biodiversity and soil community composition determine ecosystem multifunctionality.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(14): p. 5266-5270.
3. Cline, L.C. and D.R. Zak, *Soil microbial communities are shaped by plant-driven changes in resource availability during secondary succession.* Ecology, 2015. **96**(12): p. 3374-85.
4. Paterson, E., et al., *Rhizodeposition shapes rhizosphere microbial community structure in organic soil.* New Phytol, 2007. **173**(3): p. 600-10.
5. Mendes, L.W., et al., *Taxonomical and functional microbial community selection in soybean rhizosphere.* Isme Journal, 2014. **8**(8): p. 1577-1587.
6. Bardgett, R.D., T.C. Streeter, and R. Bol, *Soil microbes compete effectively with plants for organic-nitrogen inputs to temperate grasslands.* Ecology, 2003. **84**(5): p. 1277-1287.
7. Reynolds, H.L., et al., *Grassroots ecology: Plant–microbe–soil interactions as drivers of plant community structure and dynamics.* Ecology, 2003. **84**(9): p. 2281-2291.
8. Kuzyakov, Y. and X. Xu, *Competition between roots and microorganisms for nitrogen: mechanisms and ecological relevance.* New Phytol, 2013. **198**(3): p. 656-69.
9. Bever, J.D., et al., *Host-Dependent Sporulation and Species Diversity of Arbuscular Mycorrhizal Fungi in a Mown Grassland.* Journal of Ecology, 1996. **84**(1): p. 71-82.
10. Schnitzer, S.A., et al., *Soil microbes drive the classic plant diversity-productivity pattern.* Ecology, 2011. **92**(2): p. 296-303.
11. Honsová, D., et al., *Species composition of an alluvial meadow after 40 years of applying nitrogen, phosphorus and potassium fertilizer.* Preslia 2007. **79**: p. 245-258.
12. Leff, J.W., et al., *Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe.* Proc Natl Acad Sci U S A, 2015. **112**(35): p. 10967-72.
13. Pickett, S.T.A., et al., *The Ecological Concept of Disturbance and Its Expression at Various Hierarchical Levels.* Oikos, 1989. **54**(2): p. 129-136.
14. Pierik, M., et al., *Recovery of plant species richness during long-term fertilization of a species-rich grassland.* Ecology, 2011. **92**(7): p. 1393-1398.
15. Ramirez, K.S., et al., *Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems.* Ecology, 2010. **91**(12): p. 3463-3470.
16. Coolon, J.D., et al., *Long-Term Nitrogen Amendment Alters the Diversity and Assemblage of Soil Bacterial Communities in Tallgrass Prairie.* PLoS ONE, 2013. **8**(6): p. e67884.
17. McBride, M.B., *in Environ. Chem. Soils, New York : Oxford University Press.* 1994.
18. Lauber, C.L., et al., *Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale.* Applied and Environmental Microbiology, 2009. **75**(15): p. 5111-5120.
19. Rousk, J., et al., *Soil bacterial and fungal communities across a pH gradient in an arable soil.* Isme Journal, 2010. **4**(10): p. 1340-1351.
20. de Vries, F.T., et al., *Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities.* Ecol Lett, 2012. **15**(11): p. 1230-9.
21. van der Wal, A., et al., *Dissimilar response of plant and soil biota communities to long-term nutrient addition in grasslands.* Biology and Fertility of Soils, 2009. **45**(6): p. 663-667.
22. Millard, P. and B.K. Singh, *Does grassland vegetation drive soil microbial diversity?* Nutrient Cycling in Agroecosystems, 2010. **88**(2): p. 147-158.
23. Pérez-Ramos, I.M., et al., *Evidence for a 'plant community economics spectrum' driven by nutrient and water limitations in a Mediterranean rangeland of southern France.* Journal of Ecology, 2012. **100**(6): p. 1315-1327.
24. Bolan, N.S., *A critical review on the role of mycorrhizal fungi in the uptake of phosphorus by plants.* Plant and Soil, 1991. **134**(2): p. 189-207.

25. Mosse, B., *Plant Growth Responses to Vesicular-Arbuscular Mycorrhiza. IV. In Soil Given Additional Phosphate.* The New Phytologist, 1973. **72**(1): p. 127-136.
26. Larsen, J.E., G.F. Warren, and R. Langston, *Effect of Iron, Aluminum and Humic Acid on Phosphorus Fixation by Organic Soils1.* Soil Science Society of America Journal, 1959. **23**(6).
27. Schreiner, R.P., et al., *Mycorrhizal fungi influence plant and soil functions and interactions.* Plant and Soil, 1997. **188**(2): p. 199-209.
28. van der Heijden, M.G.A., et al., *Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity.* Nature, 1998. **396**(6706): p. 69-72.
29. Xia, W., et al., *Autotrophic growth of nitrifying community in an agricultural soil.* Isme j, 2011. **5**(7): p. 1226-36.
30. Attard, E., et al., *Shifts between Nitrospira- and Nitrobacter-like nitrite oxidizers underlie the response of soil potential nitrite oxidation to changes in tillage practices.* Environ Microbiol, 2010. **12**(2): p. 315-26.
31. Letcher, P.M., et al., *Ultrastructural and molecular phylogenetic delineation of a new order, the Rhizophydiales (Chytridiomycota).* Mycol Res, 2006. **110**(Pt 8): p. 898-915.
32. Lee, I.M., R.E. Davis, and D.E. Gundersen-Rindal, *Phytoplasma: phytopathogenic mollicutes.* Annu Rev Microbiol, 2000. **54**: p. 221-55.
33. Begerow, D., M. Stoll, and R. Bauer, *A phylogenetic hypothesis of Ustilaginomycotina based on multiple gene analyses and morphological data.* Mycologia, 2006. **98**(6): p. 906-16.
34. Elberse, W.T., J.P. van der Berg, and J.G.P. Dirven, *Effect of use and mineral supply on the botanical composition and yield of old grassland on heavy-clay soil.* Netherlands Agriculture Science, 1983. **31**: p. 63-88.
35. Mannetje, L. and K.P. Haydock†, *The dry-weight-rank method for the botanical analysis of pasture.* Grass and Forage Science, 1963. **18**(4): p. 268-275.
36. Neuteboom, J.H., E.A. Lantinga, and P.C. Struik, *Evaluation of the dry weight rank method for botanical analysis of grassland by means of simulation.* Netherlands Journal of Agricultural Science, 1998. **46**(3/4): p. 285-304.
37. Pan, Y., et al., *Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil.* Fems Microbiology Ecology, 2014. **90**(1): p. 195-205.
38. Verbruggen, E., et al., *Testing Potential Effects of Maize Expressing the Bacillus thuringiensis Cry1Ab Endotoxin (Bt Maize) on Mycorrhizal Fungal Communities via DNA- and RNA-Based Pyrosequencing and Molecular Fingerprinting.* Applied and Environmental Microbiology, 2012. **78**(20): p. 7384-7392.
39. Goecks, J., et al., *Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.* Genome Biology, 2010. **11**(8).
40. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.
41. Edgar, R.C., et al., *UCHIME improves sensitivity and speed of chimera detection.* Bioinformatics, 2011. **27**(16): p. 2194-2200.
42. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic Acids Research, 2013. **41**(Database issue): p. D590-D596.
43. Dray, S. and A.-B. Dufour, *The ade4 Package: Implementing the Duality Diagram for Ecologists.* 2007, 2007. **22**(4): p. 20.
44. Kenkel, N.C., *On selecting an appropriate multivariate analysis.* Canadian Journal of Plant Science, 2006. **86**(3): p. 663-676.
45. Moonseong, H. and K. Ruben Gabriel, *A permutation test of association between configurations by means of the rv coefficient.* Communications in Statistics - Simulation and Computation, 1998. **27**(3): p. 843-856.

Chapter 3

Impact of long-term N, P, K and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil

Noriko A. Cassman, Yao Pan, Mattias de Hollander, Lucas W. Mendes, Hein Korevaar, Rob H.E.M. Geerts, Johannes A. van Veen and Eiko E. Kuramae

Published as:

Pan Y*, **Cassman NA*,** de Hollander M, Mendes LW, Korevaar H, Geerts RHEM, van Veen JA & Kuramae EE, 2014. "Impact of long-term N, P, K and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil." *FEMS Microbiol Ecol.* 90 (1): 195-205.

* indicates co-first authorship

Abstract

Soil abiotic and biotic interactions govern important ecosystem processes. However, the mechanisms behind these interactions are complex and the links between specific environmental factors, microbial community structures and functions are not well understood. Here, we applied DNA shotgun metagenomic techniques to investigate the effect of inorganic fertilizers N, P, K and NPK on the bacterial community composition and potential functions in grassland soils in a 54-year experiment. Differences in total and available nutrients were found in the treatment soils; interestingly, Al, As, Mg and Mn contents were variable in N, P, K and NPK treatments. Bacterial community compositions shifted and *Actinobacteria* were overrepresented under the four fertilization treatments compared to the control. Redundancy analysis of the soil parameters and the bacterial community profiles showed that Mg, total N, Cd and Al were linked to community variation. Using correlation analysis, *Acidobacteria*, *Bacteroidetes* and *Verrucomicrobia* were linked similarly to soil parameters, and *Actinobacteria* and *Proteobacteria* were linked separately to different suites of parameters. Surprisingly, we found no fertilizers effect on microbial functional profiles which supports functional redundancy as a mechanism for stabilization of functions during changes in microbial composition. We suggest that functional profiles are more resistant to environmental changes than community compositions in the grassland ecosystem.

Keywords: metagenomics/microbial ecology/microbial function redundancy/environmental factors

3.1 Introduction

Soil harbors a huge variety of organisms, including microorganisms which are central to terrestrial ecosystem processes such as C and N flows[1]. The recent development of culture-independent techniques and high-throughput sequencing technologies allows detailed study of bacterial communities and the factors that cause shifts in bacterial community structure[2-4]. Many studies apply 16S rRNA amplicon sequence profiling to address not only changes in community composition and function[5, 6] but also the relationship between community components and external drivers (to what extent community compositional shifts are influenced by external factors)[7]. Changes in the soil bacterial communities due to disturbances or altered resource availability may influence ecosystem processes by altering bacterial functions, compositions or interactions. However, the influence of specific environmental factors on changes in bacterial community composition and function due to nutrient inputs alone is not clear.

Studies of bacterial structural drivers at the regional scale have shown that bacterial community structure is driven by soil type and general chemical characteristics, including pH. Soil pH is shown as a key environmental factor that influences bacterial community composition and can explain the distributions of bacterial phyla at local scales[8, 9]. However, there is some indication that pH only indirectly drives shifts in community composition. For instance, with pH change, the soil moisture, cations availability and C:N ratios often co-vary as well[10]. Linking bacterial community shifts only to pH does not take into account the individual influence of soil parameters, e.g., NO_3, NH_4 and Ca. One study in Amazon soils that measured a wide range of soil parameters found a relationship between a specific group of *Acidobacteria* and the available Ca, Mg and Mn[11]. Because of the range of abiotic and biotic interactions between bacteria and their chemical environment, it is important to investigate a suite of soil paramenters in studies of the drivers of bacterial community changes, which we have applied here.

Nutrient amendments in managed grasslands are historically used to improve plant productivity. However, this management practice impacts all components of the grassland ecosystem, including soil functioning. The effect of chronic, long-term (more than five years) N fertilization is intensively studied, e.g. for the relationship between nutrients and plant productivities[12]; or for evaluating anthropogenic effects on bacterial and fungal biomass, community structure, composition and activities[8]. Several studies find long-term N fertilization effects on bacterial community compositional shifts and subsequent alteration of ecosystem functions[13-15]. Long-term application of N fertilizer affects the abundance of

specific groups that are associated with N cycling, as shown in studies targeting nitrifiers and denitrifiers[15, 16], ammonium oxidizing Archaea[17, 18] and methanotrophs[19]. In general, the long-term addition of N appears to select for copiotrophic taxa though these taxa may respond to C dynamics associated with the N additions[13]. Whether a similar mechanism affects bacterial community composition in long-term P and K fertilization is unknown.

In this study, we explored the shifts in abiotic and biotic factors between long-term N-, P-, K- and NPK-fertilization. First, we determined the soil chemical parameters in each long-term fertilization regime. Next, we tracked the compositional and functional changes in the bacterial communities across fertilization regimes. Last, we identified links between soil chemical parameters and bacterial community shifts. We hypothesized that soil factors other than pH were linked to bacterial community compositional and functional alterations (H_1). In addition, we hypothesized that each fertilization regime would alter bacterial community composition and function (H_2). To our knowledge, our study is the first to apply shotgun metagenome techniques to the soil bacterial communities in clay-soil grassland with yearly addition of inorganic fertilizers for such duration (54 years). Advantages over the widely used 16S rRNA amplicon method include the ability to examine community functional potential and to exclude PCR biases. In addition, this study is one of the first to include a suite of soil measurements to explore the environmental factors driving simultaneously soil bacterial community composition and functions under long-term nutrient input.

3.2 Material and methods

3.2.1 Site description

The Ossenkampen Grassland Experiment fields were established in Wageningen, The Netherlands (51°58'15"N; 5°38'18"E) on heavy-clay soil in 1958 to track plant species shifts under long-term application of inorganic fertilizers[20]. Fertilizer amendments, including N (ammonium nitrate, 160 kg N ha^{-1} yr^{-1}), P (superphosphate, 22 kg P ha^{-1} yr^{-1}), K (potassium sulfate, 108 kg K ha^{-1} yr^{-1}) and NPK (ammonium nitrate, superphosphate and potassium sulfate, 160 kg N ha^{-1} yr^{-1}, 33 kg P ha^{-1} yr^{-1} and 311 kg K ha^{-1} yr^{-1}) were applied to the fields annually since 1959. The fields were mown twice a year: once in July and once in October. The mown matter was left on the fields.

3.2.2 Sampling regime and soil parameters

Soil samples were collected on September 20, 2011 from the four treatment fields and one control field (five treatment fields). Each treatment field included 5 m × 2.5 m triplicate plots. Twenty-seven soil cores (10 cm depth and 2 cm diameter) were sampled from each plot and then pooled to give one bulk sample for each plot (Supplementary Figure 1). Five bulk soil samples (3 replicates/sample) were collected and homogenized through a 5 mm sieve. Fractions from two of the three soil samples from each field were sent for physical and chemical measurements at the Soil Science Department of Wageningen University. The remaining soil fractions were stored at -80^{o}C for molecular analyses. For the physical and chemical measurements, levels of soil moisture content, soil pH, extractable N, total carbon and nitrogen concentrations, available potassium, phosphate and sulfur, total organic matter, and available trace elements Al, As, Cd, Cr, Cu, Fe, Mg, Zn, Mn, Na, Ni and Pb were determined.

3.2.3 Shotgun metagenome preparation and processing

DNA from soil fractions (0.3 g) was extracted using the Power Soil kit (MolBio, Carlsbad, CA) with 5.5 m·s^{-1} for 10-min bead beating. DNA concentrations were measured using an ND-1000 spectrophotometer (Nanodrop, Wilmington, DE) to ensure at least 500 ng for library preparation.

The libraries were pyrosequenced on the Roche 454 FLX platform with titanium chemistry (Roche, USA) at Macrogen (South Korea). After demultiplexing and barcode removal, the fragments were trimmed to 300 bp with a quality score cutoff of 30. Artificial duplicates were removed (http://microbiomes.msu.edu/replicates/) from further analysis[21].

3.2.4 Sequence annotation and abundance normalization

The metagenomes were uploaded to the Metagenome Rapid Annotation with Subsystem Technology (MG-RAST; http://metagenomics.anl.gov) web server, version 3.1.2[22]. The sequences were deposited in MG-RAST under ID numbers 4485389.3 to 4485403.3. The metagenomes were compared by BLASTx to the Clusters of Orthologous Groups (COGs) database for the genome size normalization calculation, to the SEED database for functional annotation and the RefSeq microbial database for taxonomic annotation. The sequences with similarity to database entries with the default parameters (E-value of 1x10^{-5}, minimum sequence nucleotide identity of 60% and minimum alignment length of 15 bp)

were counted. The count of metagenome sequences with similarity to the 35 COG marker genes described by Raes *et al.* (2007) [23] [2323] were filtered from the COG tables and used to calculate the effective genome size (EGS) for each metagenome. The annotation abundances were normalized by multiplying each value by the weight of the sample EGS to the average EGS of all sample metagenomes (for detailed method see references[24, 25]. The taxonomic and functional annotation profiles were created by calculating the relative abundances of annotations within each taxonomic or functional category out of the total number of annotations for the sample.

3.2.5 Statistical analyses

Soil parameter values were compared between the control and treatment fields using permutation t-tests (n = 2) and among the five treatments using Analysis of Variance (ANOVA). The t-tests and ANOVA were performed using PAST software version 2.15[26]. The similarity of treatment groups was calculated based on the Euclidean distance matrices between the taxonomic and functional profiles with the Analysis of Similarity (ANOSIM) test. PCA[27] was applied to visualize the samples based on the taxonomic and functional profiles.

Overrepresentation of bacterial phyla or Subsystem Level 1 and 2 categories between the treatment groups and the control group were investigated using the Statistical Analysis of Metagenomic Profiles software (STAMP, version 2.0.0) [28]. Unclassified annotations were not included in the analyses. The two-sided Welch's t-test was applied with Storey's FDR (taxonomic analysis) or Benjamini-Hochberg FDR (functional analysis) multiple test correction and significance was determined based on a 95% confidence interval.

Redundancy analysis (RDA) was conducted to combine the soil parameter and the taxonomic or functional and data. For this analysis, the average of the two values for each soil variable was used as the value for the soil variable of the third replicate. The soil variables were tested for normality and standardized and in two cases were log (Nts) or inverse (Zn) transformed to achieve normality. Pearson correlations were calculated between the soil variables to identify highly correlated variables. The rda function in the "vegan" package in R (version 3.0.2) was applied. The rda function conducts an unweighted linear regression of categories (e.g. relative abundances) against variables (e.g. soil parameters) and performs unweighted singular value decomposition of the influencing variables.

3.3 Results

3.3.1 Soil parameters along the fertilizations

The autocorrelation between soil factors did not result in highly correlated variables across treatments ($p<0.05$), except organic matter correlated with total N (0.92) and with Mg (0.90). Therefore, no variables were removed for further analysis. Seventeen of the 24 analyzed soil parameters differed across the five treatment fields (ANOVA, $p < 0.05$; **Table 1**). The only soil parameter that differed in all of the four treatment fields compared to the control was total extractable N. Several soil parameters differed within the N, P or K and the NPK treatments. Expected differences due to the associated fertilizations were found. For instance, NO_3-NO_2-N was higher in the N and NPK treatments; available K was higher in the K and NPK treatments; and P was higher in NPK treatments. In both the P and NPK treatments, available C was lower than in the control treatment. Al was higher in the N and NPK treatments compared to the control. NH_4-N was lower in the K and NPK treatments compared to the control treatment. Most interestingly, the differences in Mg and Mn values that were found in the N and P comparisons were not encountered in the NPK comparison.

3.3.2 Shotgun metagenome characteristics

The 15 metagenomes comprised a total of 603 million bp contained in 1.6 million reads (**Supplementary Table 1**). Mean lengths of the sequences ranged from 362 to 398 bp and the quality scores were 36 and above. MG-RAST annotation to the M5NR database identified 53% to 57% of the predicted protein coding regions for the 15 sample metagenomes. Rarefaction curve analysis revealed that most of the diversity was sampled in the metagenomes based on number of sequences against number of species from the M5NR comparison (**Supplementary Figure 2**). Alpha diversity of the metagenomes averaged ~640 ± 34 species. The functional diversity, given by the number of functional categories divided by the metagenome size, ranged from 39% to 51%. The samples were dominated by bacteria domain sequences (average relative abundance 94.30%) followed by eukaryota (1.19%) and archaea (1.18%) sequences based on hits to the RefSeq database.

Table 1. Soil physicochemical properties from the non-fertilized (Control) field and the nitrogen (N), phosphorous (P), potassium (K) and nitrogen+phosphorous+potassium (NPK) treatment fields. Significant treatment values are in bold from the control to treatment comparison and the parameters that differed significantly among the five treatments are in bold.

Soil Properties		Values within Treatment Fields				
		Control	**N**	**P**	**K**	**NPK**
Total	**Ct (g/kg)****	59±1.00	62.7±1.33	58.7±0.33	61±0.01	60.7±1.67
	Nt (g/kg)**	5.2±0.03	5.8±0.13	5.1±0.1	5.2±0.01	5.2±0.1
	Nt (mg/kg)**	48±0.01	**68±0.01*****	**46.3±0.01****	**42±0.01****	**51.3±0.01****
	Kt (g/kg)***	10.8±0.24	11.8±0.04	10.5±0.11	12.6±0.06	12.1±0.07
	Pt (g/kg)***	0.7±0.04	0.7±0.01	1.1±0.03	0.07±0.03	1.1±0.04
	OM (%)*	14.7±0.23	15.6±0.4	14.3±0.13	14.7±0.03	14.6±0.33
	pH*	5.27±0.1	4.85±0.06	5.22±0.04	5.05±0.09	5.04±0.05
	moisture (%)	47.2±1.76	**37.5±1.09***	43.3±1.14	41.8±3.04	41.9±1.89
	C:N	11.3±0.12	10.9±0.02	11.5±0.16	11.7±0.01	11.7±0.09
Available (mg/kg)	C	699.7±21.7	716.7±39.3	**652±9.0***	702±3.0	**642±1.7***
	NH_4^+-N**	13.5±1.3	15.0±1.67	11.1±1.57	**7.3±1.17***	**5.5±0.47***
	NO_3^-+NO_2—N	1.6±0.07	**15.3±0.23****	1.9±0.3	1.2±0.13	**14.1±1.67***
	P***	2.3±0.1	2.1±0.07	**3.7±0.23**	2.1±0.1	**3.8±0.03***
	S	11.0±0.53	12.0±0.70	11.9±0.20	10.2±0.90	10.4±0.47
	Al***	8.7±2.23	**20.4±2.73***	15.5±0.53	13.4±1.54	**27.9±1.63****
	As**	0.11±0.01	0.09±0.01	0.08±0.01	0.09±0.01	0.08±0.01
	Cd	0.12±0.01	0.15±0.01	0.13±0.01	0.12±0.01	0.15±0.01
	Cu	0.1±0.00	0.1±0.00	0.1±0.00	0.1±0.00	0.1±0.00
	K***	35±0.01	36.3±2.67	37.3±1.67	**135.7±11.7***	**118.3±11.3***
	Mg**	211.3±9.33	**259.7±16.33****	**179.3±2.67***	222.7±7.33	215.7±1.67
	Mn*	26.3±2.23	**13.7±0.07***	**17.0±2.83***	24.87±4.40	21.7±3.53
	Na	36.7±6.67	32.7±7.33	35.0±4.00	34.7±2.33	31±1.00
	Ni	1.0±0.02	1.2±0.09	1.00±0.04	1.2±0.04	1.0±0.05
	Zn	3.3±0.3	4.3±0.47	2.8.0±0.03	3.5±0.0	4.7±1.53

*p < 0.05, **p < 0.01, ***p < 0.001

3.3.3 Taxonomic comparison of the sampled communities

The taxonomic profiles of the treatment metagenomes were similar to those of the control metagenomes. *Proteobacteria*, *Acidobacteria, Actinobacteria, Firmicutes, Verrucomicrobia* and *Bacterioidetes* were the most abundant phyla in the metagenomes. Fifteen additional bacterial phyla with more than 0.8% of representation in the metagenomes were identified by annotation against the RefSeq database (**Supplementary Table 2**). Principal Component Analysis (PCA) was used to visualize the samples based on relative abundances of annotations within bacterial phyla. Different inorganic fertilizations significantly changed mean relative

abundances of *Actinobacteria, Aquificae Chlorobi*, *Chloroflexi*, *Cyanobacteria, Deferribacteres, Dictyoglomi, Firmicutes, Lentisphaerae, Nitrospirae, Planctomycetes, Proteobacteria, Spirochaetes, Synergistetes* and *Verrucomicrobia.* (ANOVA, corrected $p < 0.05$). The samples were found to form distinct treatment groups by the PCA which explained ~96% of the variance between samples. This was confirmed by the ANOSIM test ($R = 0.835$, $p < 0.0001$; **Supplementary Figure 3a**).

Two-group comparisons revealed over- and under-represented phyla between the control and treatment groups, with the exception of the P treatment (**Figure 1**). The N, K and NPK treatments showed an overrepresentation in *Actinobacteria* compared to the control (corrected $p < 0.03$). In addition, the N treatment showed an underrepresentation in *Aquificae*, *Chlorobi*, *Cyanobacteria*, *Firmicutes*, *Nitrospirae*, *Planctomycetes*, and *Verrucomicrobia* (corrected $p < 0.03$), the K treatment an underrepresentation in *Firmicutes* and *Verrucomicrobia* (corrected $p < 0.04$) and the NPK treatment an underrepresentation in *Planctomycetes* (corrected $p < 0.05$).

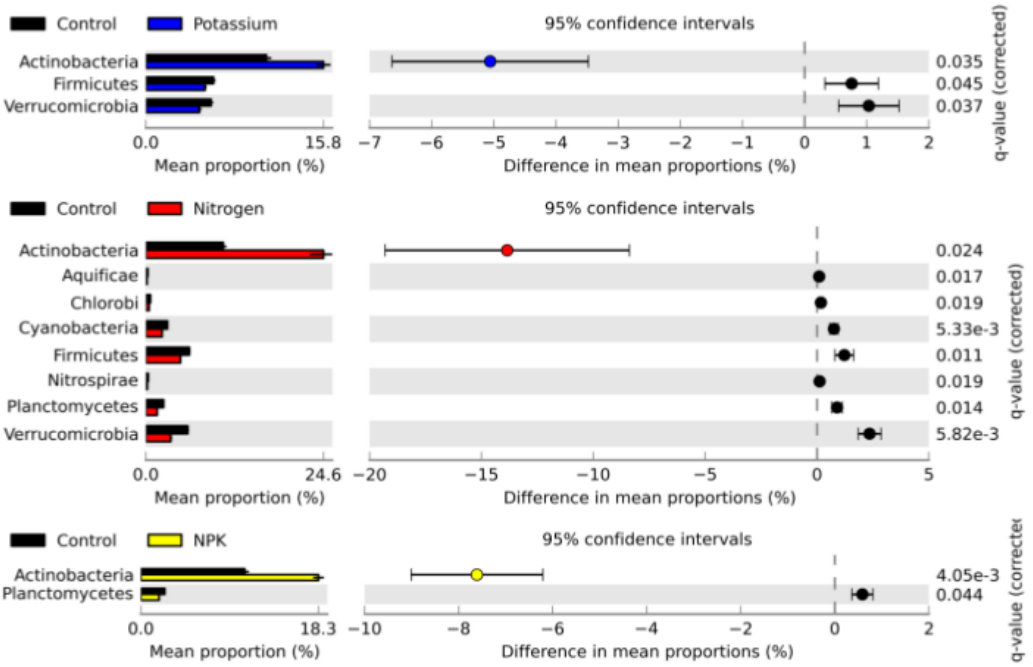


Figure 1. The bacterial phyla that were significantly overrepresented in comparisons between the treatments (Potassium, Nitrogen and NPK) and the control are shown (Welch's t-test, $p < 0.05$). Control=black, Potassium=blue, Nitrogen=red, NPK=yellow. In the Control to Phosphorus treatments no overrepresented phyla were found.

3.3.4 Functional comparison of the sampled communities

Functional profiles of the metagenomes were created from relative abundances of annotations to hierarchical subsystems through MG-RAST. The 15 sample metagenomes exhibited similar functional profiles across the 28 Subsystem Level 1 categories. The top 10 subsystems represented in the metagenomes were Carbohydrates (average 15.6%), Clustering-based subsystems (14.0%), Amino Acids and Derivatives (9.9%), Miscellaneous (7.6%), Protein Metabolism (6.0%), Cofactors, etc. (6.0%), RNA Metabolism (5.0%), Cell Wall and Capsule (4.0%), Fatty Acids, etc. (4.1%) and Virulence, etc. (3.3%). Principal Component Analysis (PCA) was used to visualize the samples based on relative abundances of annotations within Subsystems Level 1 and 2 categories. In the Level 1 comparison, the samples did not form distinct treatment groups in the PCA (~60% of the variance) which was reflected in the ANOSIM test ($R = 0.2889$, $p < 0.0246$; **Supplementary Figure 4**). In the Level 2 comparison, the treatment effects were more distinctly displayed in the PCA (~48% of variance) and this was reflected in the ANOSIM test result ($R = 0.6193$, $p < 0.0003$; **Supplementary Figure 3b**). Of the Level 2 categories, no category differed between treatments based on the multiple group comparison. Two-group comparisons revealed no over- or under-represented functional categories between the control and treatment groups based on Subsystems Level 1 and 2.

3.3.5 Linking environmental variables with bacterial taxa and functional categories

The links between the measured soil parameters and the relative abundances within bacterial phyla or functional Level 2 categories were investigated through RDA. Mn, As, Mg, Cd, Al, Pt, K, Kt, and Nt were found to be related to the samples based on the taxonomic (RDA, adjusted R-squared = 0.82) and functional Level 2 (RDA, adjusted R-squared = 0.34) profiles (**Figure 2a and Figure 2b**).

In order to examine the correlation between environmental factors and bacterial taxa, Spearman correlations were calculated between the relative abundance values of different bacterial taxa and the values for soil parameters for each control versus treatment comparison (**Table 2**). Of the soil parameters that were significantly different in the control to treatment comparisons, Al, K, Nt, $N\text{-}NH_4$, $NO_3\text{-}NO_2\text{-}N$, P and Mg were strongly correlated ($r > |0.70|$) with the relative abundances of the six major bacterial phyla. Three groups could be made of the bacterial phyla with similar correlations. First, the *Acidobacteria, Bacteroidetes* and *Verrucomicrobia* were correlated to similar soil parameters in at least two

comparisons: negatively correlated to Kt, Al, K and Ni and positively correlated with N-NH_4, NO_3-NO_2-N and Nt. The *Actinobacteria* were positively correlated to Kt, Al, K, and Ni and negatively correlated to N-NH_4 and Nt in at least two comparisons. Last, the *Firmicutes* and *Proteobacteria* were positively correlated with P, N-NH_4, NO_3-NO_2-N, Nt, Ct, OM and moisture in at least one comparison (**Supplementary Table 3**).

For investigating the correlation between soil factors and functional categories, Spearman correlations were also calculated between the relative abundance values within functional categories and the values of soil parameters. The relative abundances of certain functional groups were correlated with specific environmental factors. For example, available Cd, Al and Ni together with total K concentrations were highly positively correlated with stress responses related genes abundance (**Table 3**), which had a negative correlation with soil pH and moisture content. Available Cd, Ni and total C were negatively correlated with virulence related gene abundances whereas soil pH and moisture content were positively correlated with this function category. Total extractable N had positive correlation with cell division and cell cycle and DNA metabolism related functional profiles abundance.

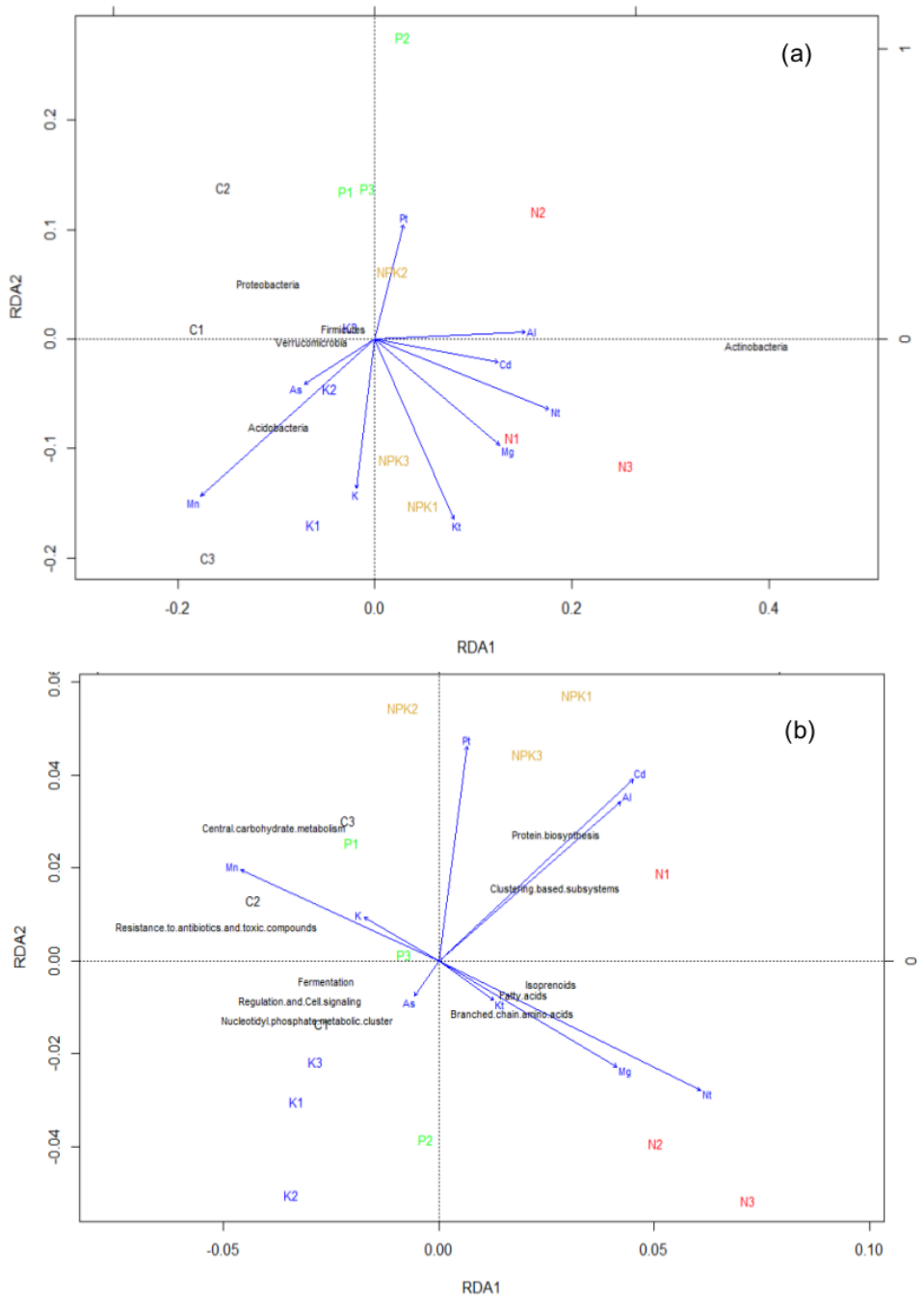
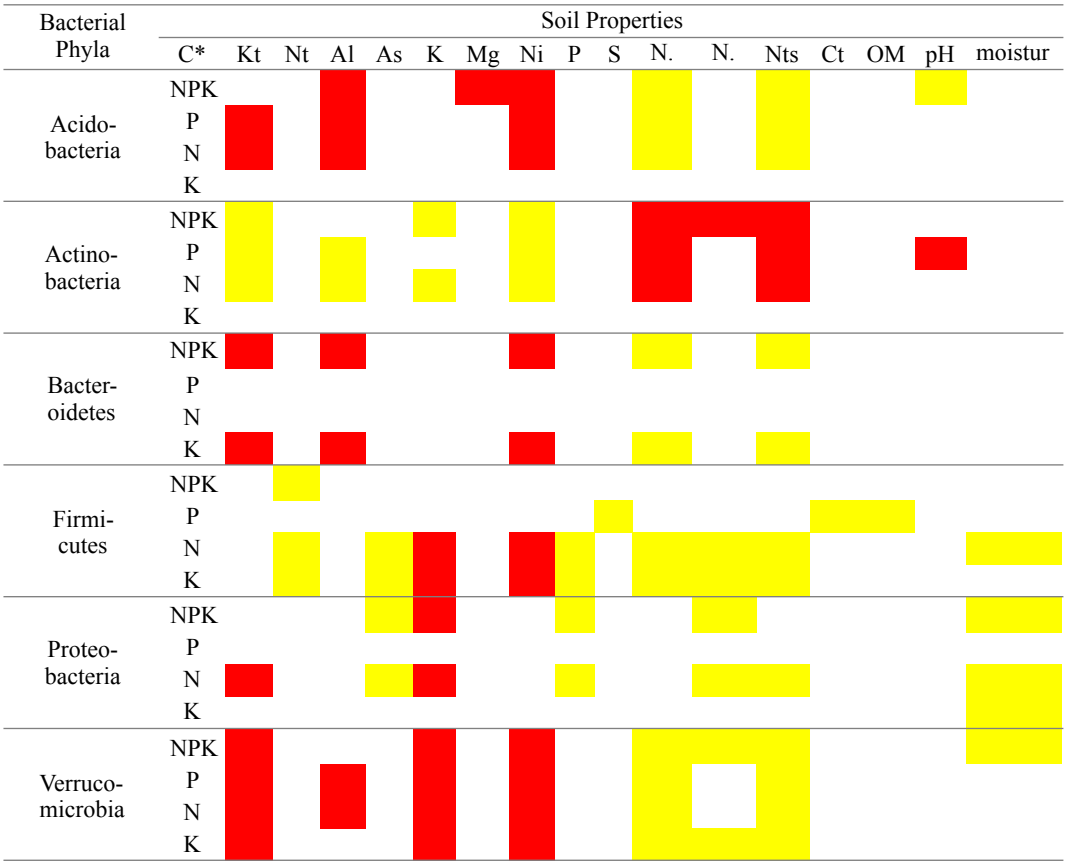


Figure 2. Redundancy analysis (RDA) of the samples based on (a) relative abundances within RefSeq Bacterial phyla or (b) relative abundances within Subsystem Level 2 categories and values of environmental variables. C=Control, N=N fertilized samples, P=P fertilized samples, K=K fertilized samples, NPK=NPK fertilized samples. 1, 2, 3 = replicates. Only the environmental variables (blue lines with arrow heads) that explained a significant amount of the variations (Monte Carlo test, p value < 0.05) are included in the figure.

Table 2. Spearman correlation values between the relative abundances of annotations within the six major bacterial phyla and soil parameters. Only the correlations values greater than |0.7| are shown. Positive correlations=yellow and negative correlations=red.



C* = control to treatment comparison; soil parameters

Table 3. Spearman correlation between soil parameters and functional profiles based on subsystems hierarchy 1 (MG-RAST). Only the correlations values greater than |0.7| are shown.

Function category	Soil Properties													
	pH	Moisture	C:N	Ct[a]	Nt[a]	Kt[a]	Pt[a]	(NO_3+NO2)-N[b]	Al[b]	Cd[b]	Mn[b]	Na[b]	Ni[b]	Pb[b]
Amino Acids and Derivatives														
Cell Division and Cell Cycle					0.7									
Cofactors, Vitamins, Prosthetic Groups, Pigments		-0.7		0.7						0.7	-0.8	0.7	0.7	0.7
DNA Metabolism					0.7			0.8	0.7	0.8				0.8
Motility and Chemotaxis														
Protein Metabolism														
Regulation and Cell signaling														-0.7
Respiration			0.7				0.7							
Secondary Metabolism														
Stress Response	-0.7	-0.7				0.7			0.8	0.8			0.7	
Virulence, Disease and Defense	0.8	0.8		-0.7						-0.8			-0.8	

a = g/kg; b = mg/kg

3.4 Discussion

Here we explored the correlations between abiotic and biotic factors in a long-term N-, P-, K- and NPK-fertilized grassland. We hypothesized that each inorganic fertilization treatment would alter bacterial community composition and functions (H_1). Bacterial community compositional and functional alterations were hypothesized to be linked to soil parameters other than pH (H_2). Our results indicated that soil chemical profiles and bacterial community composition but not functions shifted with inorganic fertilization. Several soil parameters were correlated to compositional and functional shifts in the bacterial community.

Chronic deposition of N, P and K fertilizers resulted in distinct soil profiles, i.e. with saturated levels of extractable N and NO_3-NO_2-N, P and K in the respective treatments (deduced by similar levels of these soil parameters in the NPK treatment). Thus, the treatments can be conceptualized as resource manipulations in which bacterial community assembly was driven by the four different treatment "habitats." Hereafter we will refer to each treatment as saturated with the associated fertilizer, e.g. N-saturated for the N fertilizer treatment, to facilitate our discussion on the drivers of compositional shifts that were observed within the habitats.

Microbial mining hypothesis links phyla and soil properties in N- and K-saturated fields

N-saturations of 160 kg N ha^{-1} yr^{-1} resulted in the largest differences in bacterial community compositions compared to the control field. Shen, Zhang [29] also showed significant changes in the microbial community composition in a 20-year inorganic N-fertilized field (135 kg N ha^{-1} yr^{-1}) in China. Fierer, Lauber [14] detected significant differences in microbial community profiles only in fields with the highest inorganic N-input of 136 to 145 kg N ha^{-1} yr^{-1}, which is in the range of our N treatment, and not in fields that received intermediate N input of 17 to 101 kg N ha^{-1} yr^{-1}. However, in that study, the duration of fertilization of the fields was 27 and 8 years while ours was 54 years. These studies and ours results confirm that N-saturation has variable effects on microbial community composition that depend on the duration of the experiment [13]. Total extractable N availability was significantly higher in our N-saturated soil compared to the control treatment while NH_4^+ levels were the same. Coupled with the taxonomic differences observed in the N- and K-saturated communities, the microbial mining theory offers a possible explanation. N stored in organic matter is the main source of N for microbial growth and maintenance [30] and microbes usually mineralize organic N into ammonium-N in order to access this nitrogen. In N-amended soil,

there is no need for microbes to mine N and to compete with other microbes to provide N sources. Instead, bacteria that utilize high levels of available N may be favored. We observed this prediction as a shift to copiotrophic taxa under N and K saturation. The copiotroph-oligotroph tradeoff [31] seems to explain our result of increased abundances of *Actinobacteria* and decreased abundances of *Acidobacteria*/*Verrucomicrobia/Firmicutes* in the N and K saturated fields. *Actinobacteria* are regarded as copiotrophs, whereas *Acidobacteria, Firmicutes* and *Verrucomicrobia* are regarded as oligotrophs. In support, studies have shown that N addition may decrease the decomposition of recalcitrant carbon [32], which may affect members of the phylum *Actinobacteria* since they are important decomposers and play a vital role in the carbon cycle [33]. The generally oligotrophic phyla *Verrucomicrobia* may highly depend on C availability due to a slow-growing life strategy [34, 35].

P mobility in N- and P-saturated fields may explain the chemical profiles

Our finding of no compositional differences in the P-saturated field compared to that of the control field was unexpected. In contrast, in a long-term (42-year) phosphorus fertilization field, the phosphorus addition resulted in a decrease of the relative abundance of *Acidobactera* and *Pseudomonas* in pasture soils [36]. Long-term (8-year) P-amendment also shifted the bacterial community in an alfalfa field [37]. In the P-saturated fields we observed decreased levels of available Mg and Mn compared to the control field. Because no compositional differences in the bacterial community between control and the P-saturated fields were found, a hypothesis based on P mobility may explain the differences in the chemical profiles that were observed. In acidic soils, phosphorous conversion to insoluble forms may occur by precipitation of phosphate ions with Ca, Mg, Al and Fe (precipitation); alternatively, phosphate might react to form plant-available soluble forms by reacting with Mg or Mn (reactive processes). We hypothesize that both precipitation and reactive processes are on-going in the P-saturated fields due to the excess availability of P. In support of this hypothesis, Mg and Mn levels were decreased in the P-saturated but not the NPK-saturated fields compared to the control fields; strikingly, available C levels were decreased in both fields. Our results suggest that in the P-saturated fields, phosphate reactions involving Mn and Mg have removed excess P while in the NPK-saturated fields the additions of N and K fertilizers diverted the P to alternate reactions.

Under fertilization, total community functions did not shift

Surprisingly, we found no fertilizer effect on the Level 1 and Level 2 functional profiles. This supports functional redundancy as a mechanism for stabilization of functions during changes in microbial composition [7, 38]. Functional redundancy refers to the hypothesis that each member of the community has the same functional capability, such that if the composition of the community changes, the metabolic output does not. This hypothesis suggests that functional profiles are more resistant to environmental changes than community compositions. Our result is in contrast to the observations of Fierer *et al.* [14]. They found distinct different functional profiles in the soils with the highest N-saturation at the highest hierarchical level of annotations. However, our observation was supported by a meta-analysis according to Allison and Martiny [39], who showed that microbial community composition is sensitive to environmental changes. Therefore, it is still difficult to make solid conclusions whether we could predict functional redundancy using whole community data. Further studies are needed to compose mock communities to test whether the link between community functions and metagenomic characteristics generally exists or whether the link only can be observed within certain taxa.

Abiotic factors and soil bacterial group correlations

In many long-term fertilization studies, community composition or functional changes are most often correlated with soil pH and C/N ratio [29, 40]. In addition to pH and C and N links, we report strong correlations of bacterial group abundances to several other measured parameters as well as functional categories. For example, we found that *Acidobacteria*, *Actinobacteria* and *Verrucomicrobia* were correlated with Kt, K, Al, Ni, N-NH_4 and Nt in at least 2 comparisons. Navarrete *et al.* Navarrete, Kuramae [11] has reported a correlation between Al with *Acidobacteria* in the Amazon area. *Verrucomicrobia* are positively correlated with soil moisture [41] and negatively correlated with elevated CO_2. Faoro, Alves [42] demonstrated the importance of Ca^{2+}/Mg^{2+} ratio, Al^{3+} and phosphorus content in shaping soil microbial community composition in the Southern Brazilian Atlantic Forest. The studies suggest that besides pH and C/N ratios other abiotic parameters may also influence soil microbial community structures. Cd, Al and Ni are regarded as toxic (stress) compounds to microbes [43]. In order to survive, the microbes need to have the ability to detoxify, which can explain why in our study available Cd, Al, Ni and total K concentrations had a high positive correlation with stress response-related gene relative abundances. The positive correlations

between cell division and DNA metabolism genes and total extractable N suggest that the increased N nutrient level may lead to enhanced microbial reproduction [44].

Conclusions

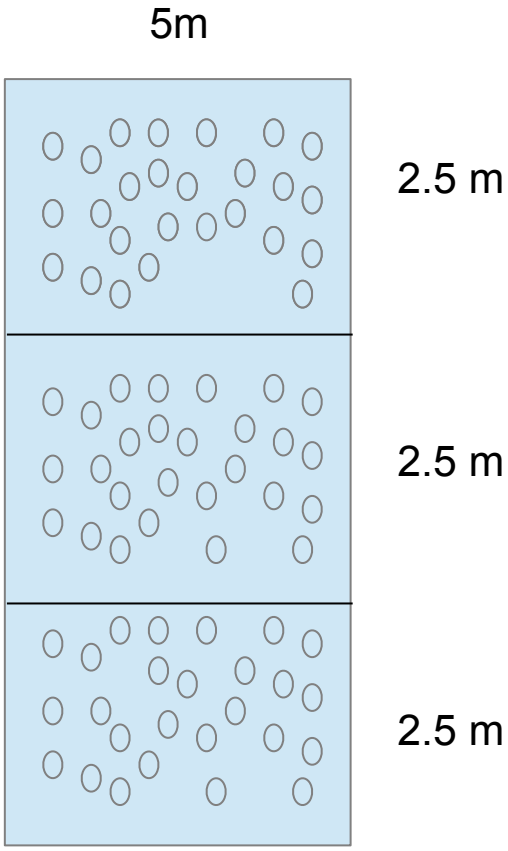
Of the numerous studies investigating the influence of long-term fertilizers on microbial community compositions, many generalize the different soil sources e.g., grassland, pasture or forest, and use 16S rRNA amplicon profiling. Here we focused on the effects of only inorganic fertilization on long-term bacterial community changes in clay soil grassland. We applied shotgun sequencing which allowed us to circumvent PCR biases as occurs 16S rRNA amplicon studies and also to evaluate the effects of fertilization not only on bacterial community composition but also on potential functions. During our analysis, we took a coarse-grained approach by examining the changes in the bacterial communities at the domain, phyla or Subsystems category levels. This was done to overcome database biases and spurious annotations from a finer resolution. Last, we demonstrated the importance of measuring a suite of soil parameters in future studies that aim to find links between abiotic and biotic ecosystem components. Whether our findings of the correlations between different soil factors (especially Fe, Al, Mg, Mn) and specific bacterial groups and functional categories have biological meanings still needs to be verified in both laboratory and field conditions.

3.5 Declarations

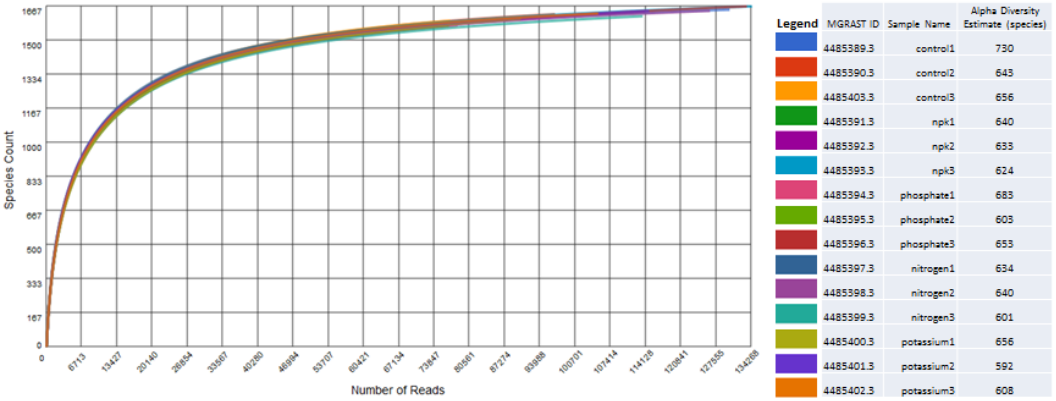
Acknowledgements

We would like to thank AS Pijl and GA Kowalchuk for laboratory assistance and support. This work was supported by NWO-FAPESP 729.004.003 program. Publication number 5636 of the NIOO-KNAW, Netherlands Institute of Ecology.

3.6 Supplementary Material

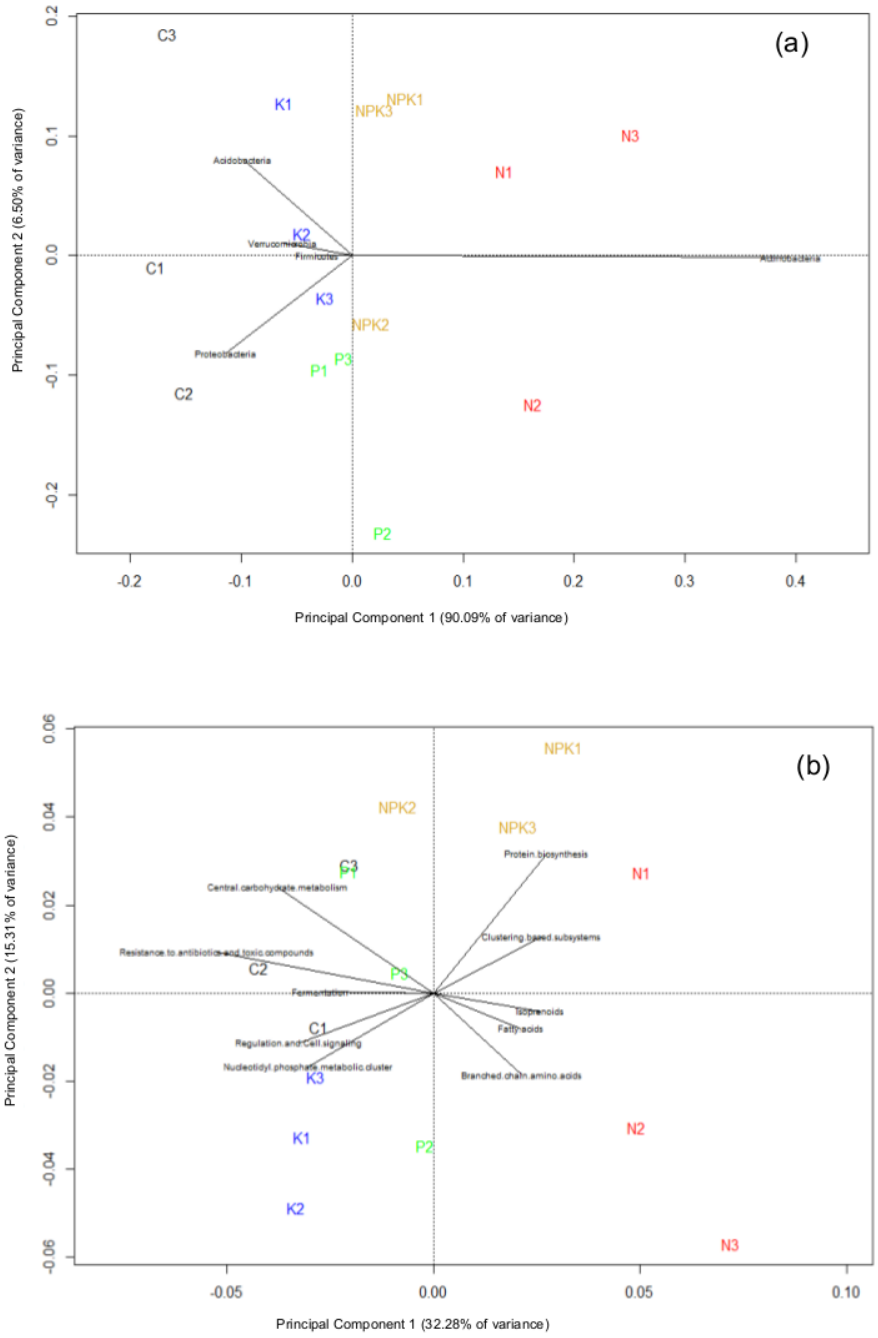


Supplementary Figure 1. One field is divided to 3 plots, 27 soil cores from each plot were sampled and pooled as one replicate for the field, and each field has 3 replicates.

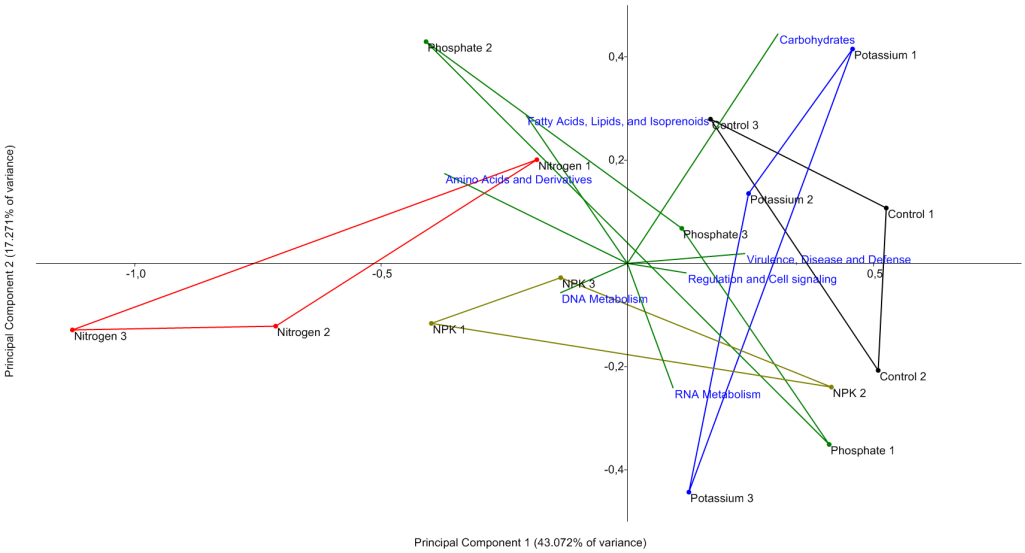


Legend	MGRAST ID	Sample Name	Alpha Diversity Estimate (species)
	4485389.3	control1	730
	4485390.3	control2	643
	4485403.3	control3	656
	4485391.3	npk1	640
	4485392.3	npk2	633
	4485393.3	npk3	624
	4485394.3	phosphate1	683
	4485395.3	phosphate2	603
	4485396.3	phosphate3	653
	4485397.3	nitrogen1	634
	4485398.3	nitrogen2	640
	4485399.3	nitrogen3	601
	4485400.3	potassium1	656
	4485401.3	potassium2	592
	4485402.3	potassium3	608

Supplementary Figure 2. Rarefaction curve analysis of the 15 sample metagenomes and estimated alpha diversity of the communities. Figure taken from MG-RAST based on subsamples plotted against M5NR species annotations.



Supplementary Figure 3. Principal Component Analysis (PCA) of the samples based on relative abundances of metagenome sequences annotated within (a) RefSeq bacterial phyla and (b) Subsystem Level 2 categories. Only the phyla that contributed to more than |0.1| units of loading and the Subsystems categories that contributed to more than |0.2| units of loading are included in the plots. C=Control, N=N fertilized samples, P=P fertilized samples, K=K fertilized samples, NPK=NPK fertilized samples. 1, 2, 3 = replicates.



Supplementary Figure 4. Principal Component Analysis (PCA) of the samples based on relative abundances of metagenome sequences annotated within Subsystem Level 1 categories. Only the phyla or Subsystem categories that contributed to more than |0.1| units of loading are included in the plots. Control=black, Phosphorus=green, Potassium=blue, Nitrogen=red, NPK=yellow; 1, 2, 3 = replicates.

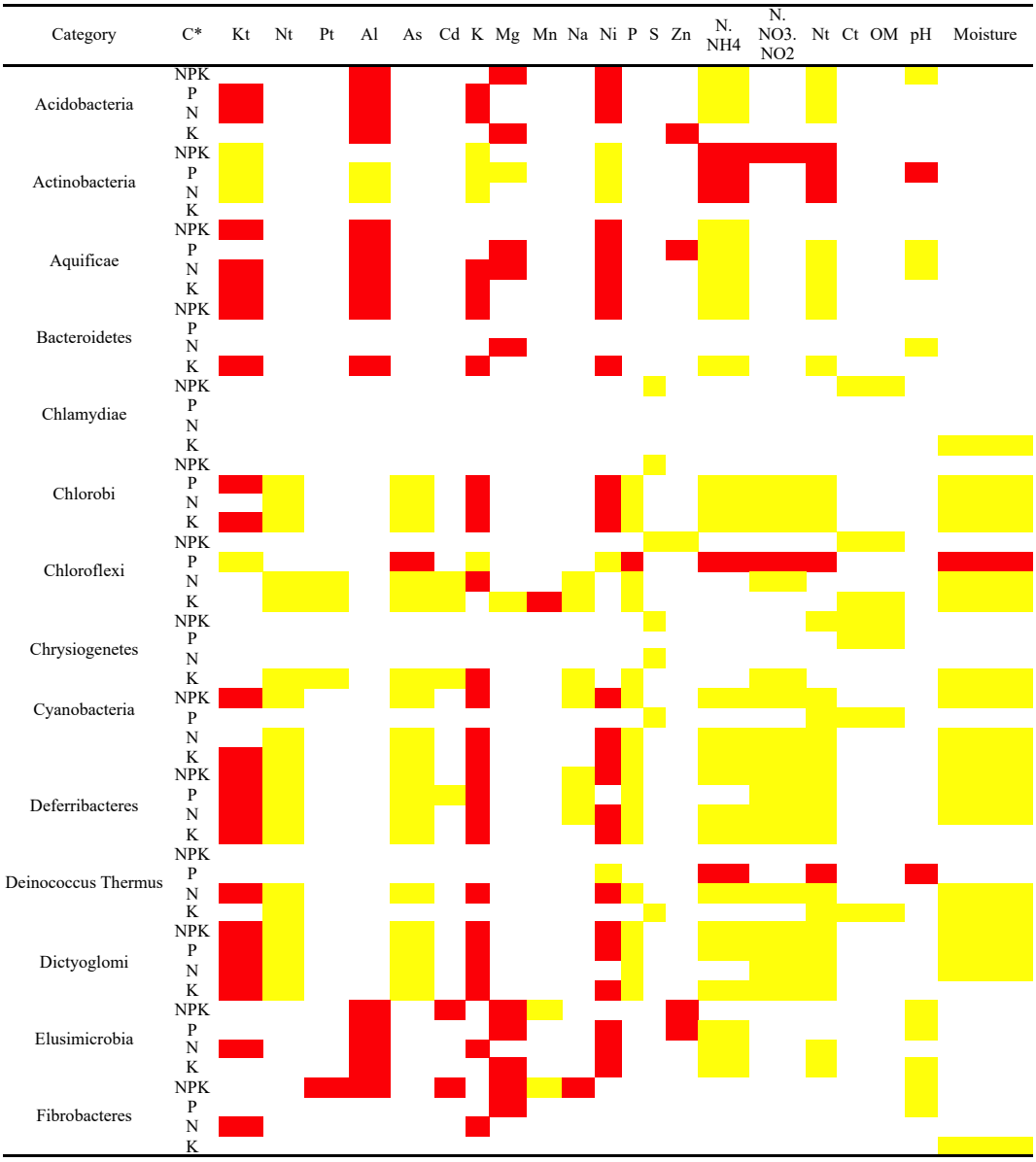
Supplementary Table 1. Sequencing information for the fifteen sample metagenomes.

Sample Name	MGRAST ID	# Sequences	# bp	Mean Sequence Length (bp)
control1	4485389.3	105010	40423549	384 ± 131
control2	4485390.3	114709	44905782	391 ± 127
control3	4485403.3	126331	49728531	393 ± 120
npk1	4485391.3	134389	52721323	392 ± 123
npk2	4485392.3	113883	42073456	369 ± 125
npk3	4485393.3	96793	38425397	396 ± 120
phosphate1	4485394.3	84053	33521299	398 ± 119
phosphate2	4485395.3	76677	29433082	383 ± 121
phosphate3	4485396.3	83399	32356773	387 ± 119
nitrogen1	4485397.3	76859	28573338	371 ± 124
nitrogen2	4485398.3	78248	28396437	362 ± 126
nitrogen3	4485399.3	90340	34738730	384 ± 123
potassium1	4485400.3	130016	51665672	397 ± 123
potassium2	4485401.3	133309	52226427	391 ± 120
potassium3	4485402.3	113565	43656772	384 ± 125

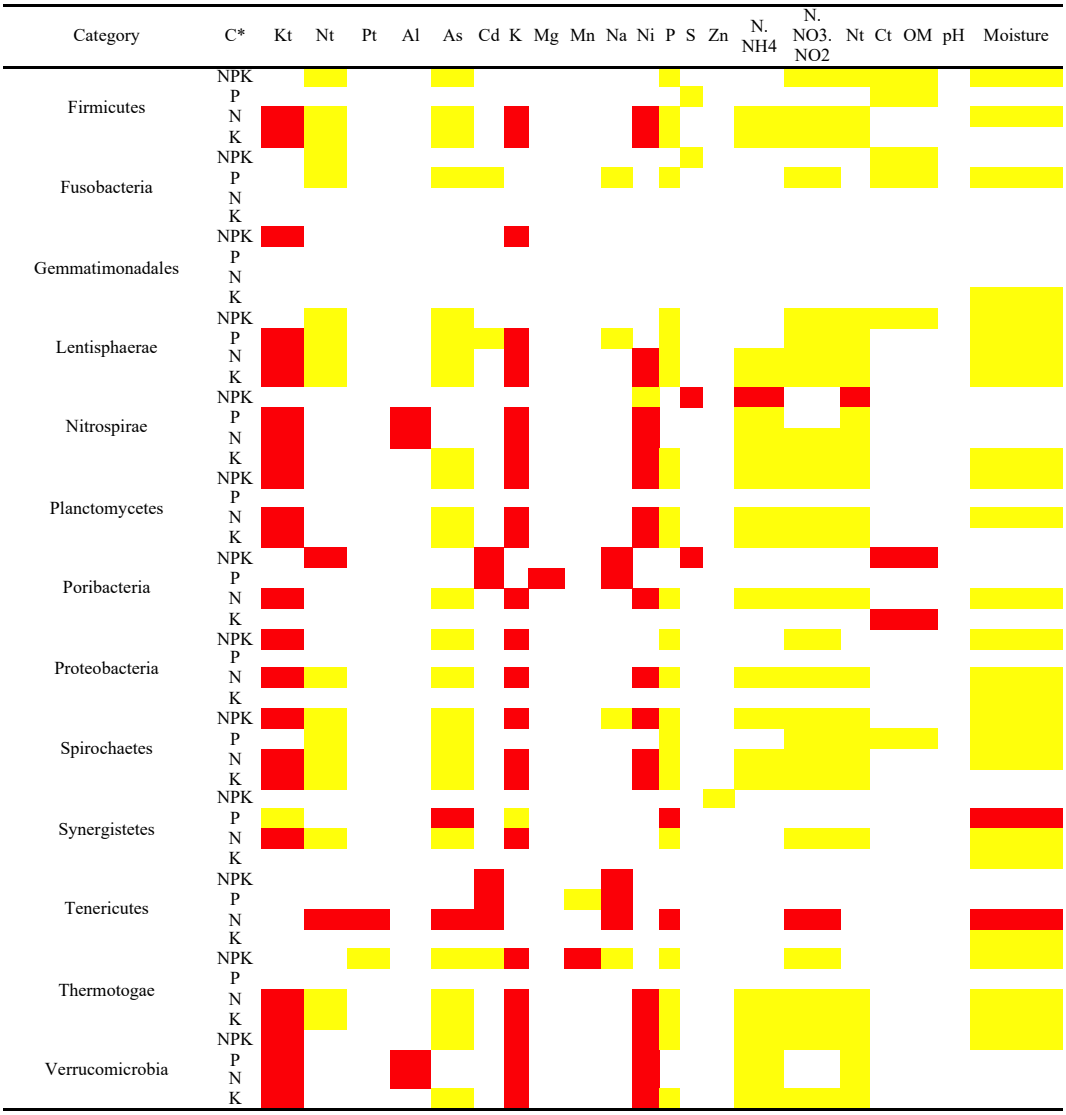
Supplementary Table 2. Relative abundances of metagenome read annotations within taxonomic categories. Values above a cutoff of 0.08% are presented and those in bold were significantly different between treatments (ANOVA, corrected $p < 0.05$).

		Average Relative Abundance of Reads (%)				
Domain	Phylum	Control	Nitrogen	Phosphorous	Potassium	NPK
Archaea	Crenarchaeota	0,14	0,12	0,12	0,12	0,17
	Euryarchaeota	0,90	0,64	0,82	0,72	0,79
Bacteria	Acidobacteria	13,78	10,59	10,44	13,36	12,20
	Actinobacteria	**10,35**	**23,70**	**16,18**	**15,19**	**17,78**
	Aquificae	**0,26**	**0,16**	**0,22**	**0,19**	**0,23**
	Bacteroidetes	4,28	3,82	4,05	3,50	3,57
	Chlamydiae	0,15	0,14	0,14	0,15	0,14
	Chlorobi	**0,61**	**0,44**	**0,54**	**0,53**	**0,58**
	Chloroflexi	**2,19**	**1,93**	**2,39**	**2,09**	**2,25**
	Cyanobacteria	**2,90**	**2,16**	**2,71**	**2,56**	**2,54**
	Deferribacteres	**0,13**	**0,08**	**0,09**	**0,08**	**0,10**
	Deinococcus-Thermus	0,58	0,54	0,60	0,53	0,60
	Firmicutes	**5,81**	**4,62**	**5,63**	**5,06**	**5,51**
	Fusobacteria	0,10	0,10	0,08	0,12	0,09
	Gemmatimonadetes	0,35	0,31	0,35	0,33	0,34
	Lentisphaerae	**0,16**	**0,11**	**0,12**	**0,11**	**0,14**
	Nitrospirae	**0,32**	**0,20**	**0,26**	**0,24**	**0,36**
	Planctomycetes	**2,38**	**1,51**	**2,37**	**2,11**	**1,82**
	Proteobacteria	**45,90**	**42,19**	**44,49**	**44,88**	**43,46**
	Spirochaetes	**0,26**	**0,17**	**0,25**	**0,21**	**0,22**
	Synergistetes	**0,14**	**0,10**	**0,15**	**0,13**	**0,15**
	Thermotogae	0,21	0,13	0,20	0,17	0,19
	Verrucomicrobia	**5,60**	**3,32**	**4,36**	**4,58**	**4,58**
Eukaryota	Arthropoda	0,12	0,17	0,16	0,09	0,11
	Ascomycota	0,79	1,23	1,10	1,30	0,55
	Basidiomycota	0,21	0,15	0,44	0,32	0,15
	Chordata	0,21	0,21	0,30	0,24	0,25
	Streptophyta	0,34	0,43	0,65	0,40	0,41

Supplementary Table 3. Spearman correlation values between the relative abundances of annotations within bacterial phyla and soil parameters using data from all five sample treatments. Only the phyla and parameters with correlations values greater than |0.6| are shown. Positive correlations=yellow and negative correlations=red. Continued on next page.



Supplementary Table 3 con't.



3.6 References

1. Prosser, J.I., *Ecosystem processes and interactions in a morass of diversity.* Fems Microbiology Ecology, 2012. **81**(3): p. 507-519.
2. Dinsdale, E.A., et al., *Functional metagenomic profiling of nine biomes.* Nature, 2008. **452**(7187): p. 629-632.
3. Torsvik, V. and L. Øvreås, *Microbial diversity and function in soil: from genes to ecosystems.* Current Opinion in Microbiology, 2002. **5**(3): p. 240-245.
4. Spor, A., O. Koren, and R. Ley, *Unravelling the effects of the environment and host genotype on the gut microbiome.* Nature Reviews: Microbiology, 2011. **9**(4): p. 279-290.
5. Frostegard, A., A. Tunlid, and E. Baath, *Use and misuse of PLFA measurements in soils.* Soil Biology & Biochemistry, 2011. **43**(8): p. 1621-1625.
6. Nemergut, D.R., et al., *Global patterns in the biogeography of bacterial taxa.* Environmental Microbiology, 2011. **13**(1): p. 135-144.
7. Langenheder, S., et al., *Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity.* PLoS ONE, 2010. **5**(5): p. e10834.
8. Berg, G. and K. Smalla, *Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere.* Fems Microbiology Ecology, 2009. **68**(1): p. 1-13.
9. Jones, R.T., et al., *A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses.* ISME J, 2009. **3**(4): p. 442-453.
10. Rousk, *Soil bacterial and fungal communities across a pH gradient in an arable soil.* ISME J, 2010. **4**(10): p. 12.
11. Navarrete, A.A., et al., *Acidobacterial community responses to agricultural management of soybean in Amazon forest soils.* FEMS Microbiology Ecology, 2013. **83**(3): p. 607-621.
12. Pierik, M., et al., *Recovery of plant species richness during long-term fertilization of a species-rich grassland.* Ecology, 2011. **92**(7): p. 1393-1398.
13. Ramirez, K.S., et al., *Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems.* Ecology, 2010. **91**(12): p. 3463-3470.
14. Fierer, N., et al., *Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients.* ISME J, 2012. **6**(5): p. 1007-1017.
15. Hallin, S., et al., *Relationship between N-cycling communities and ecosystem functioning in a 50-year-old fertilization experiment.* ISME J, 2009. **3**(5): p. 597-605.
16. Shen, J.P., et al., *Abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea communities of an alkaline sandy loam.* Environmental Microbiology, 2008. **10**(6): p. 1601-1611.
17. He, J.Z., et al., *Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices (vol 9, pg 2364, 2007).* Environmental Microbiology, 2007. **9**(12): p. 3152-3152.
18. Chu, H.Y., et al., *Community structure of ammonia-oxidizing bacteria under long-term application of mineral fertilizer and organic manure in a sandy loam soil.* Applied and Environmental Microbiology, 2007. **73**(2): p. 485-491.
19. Seghers, D., et al., *Long-term effects of mineral versus organic fertilizers on activity and structure of the methanotrophic community in agricultural soils.* Environmental Microbiology, 2003. **5**(10): p. 867-877.
20. Elberse, W.T., J.P. van der Berg, and J.G.P. Dirven, *Effect of use and mineral supply on the botanical composition and yield of old grassland on heavy-clay soil.* Netherlands Agriculture Science, 1983. **31**: p. 63-88.
21. Gomez-Alvarez, V., T.K. Teal, and T.M. Schmidt, *Systematic artifacts in metagenomes from complex microbial communities.* ISME J, 2009. **3**(11): p. 1314-1317.
22. Meyer, F., et al., *The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.* BMC Bioinformatics, 2008. **9**(1): p. 386.

23. Raes, J., et al., *Prediction of effective genome size in metagenomic samples.* Genome Biology, 2007. **8**(1).
24. Beszteri, B., et al., *Average genome size: a potential source of bias in comparative metagenomics.* ISME J, 2010. **4**(8): p. 1075-1077.
25. Plewniak, F., et al., *Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments.* Molecular Ecology, 2013. **22**(19): p. 4870-4883.
26. Hammer, Ø., D.A.T. Harper, and P.D. Ryan, *PAST: Paleontological Statistics Software Package for Education and Data Analysis.* Palaeontologia Electronica, 2001. **4**(1): p. 9.
27. Clarke, K.R., *Non-parametric multivariate analyses of changes in community structure.* Australian Journal of Ecology, 1993. **18**(1): p. 117-143.
28. Parks, D.H. and R.G. Beiko, *Identifying biologically relevant differences between metagenomic communities.* Bioinformatics, 2010. **26**(6): p. 715-721.
29. Shen, J.P., et al., *Impact of long-term fertilization practices on the abundance and composition of soil bacterial communities in Northeast China.* Applied Soil Ecology, 2010. **46**(1): p. 119-124.
30. Parfitt, R.L., et al., *Relationships between soil biota, nitrogen and phosphorus availability, and pasture growth under organic and conventional management.* Applied Soil Ecology, 2005. **28**(1): p. 1-13.
31. Ramirez, K.S., J.M. Craine, and N. Fierer, *Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes.* Global Change Biology, 2012. **18**(6): p. 1918-1927.
32. Craine, J.M., C. Morrow, and N. Fierer, *Microbial nitrogen limitation increases decomposition.* Ecology, 2007. **88**(8): p. 2105-2113.
33. Ventura, M., et al., *Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum.* Microbiology and Molecular Biology Reviews, 2007. **71**: p. 495-548.
34. Bergmann, G.T., et al., *The under-recognized dominance of Verrucomicrobia in soil bacterial communities.* Soil Biology & Biochemistry, 2011. **43**(7): p. 1450-1455.
35. He, Z.L., et al., *The phylogenetic composition and structure of soil microbial communities shifts in response to elevated carbon dioxide.* ISME J, 2012. **6**(2): p. 259-272.
36. Tan, H., et al., *Long-term phosphorus fertilisation increased the diversity of the total bacterial community and the phoD phosphorus mineraliser group in pasture soils.* Biology and Fertility of Soils, 2013. **49**(6): p. 661-672.
37. Beauregard, M.S., et al., *Long-term phosphorus fertilization impacts soil fungal and bacterial diversity but not AM fungal community in Alfalfa.* Microbial Ecology, 2010. **59**(2): p. 379-389.
38. Frossard, A., et al., *Disconnect of microbial structure and function: enzyme activities and bacterial communities in nascent stream corridors.* ISME J, 2012. **6**(3): p. 680-691.
39. Allison, S.D. and J.B.H. Martiny, *Resistance, resilience, and redundancy in microbial communities.* Proceedings of the National Academy of Sciences, 2008. **105**(Supplement 1): p. 11512-11519.
40. Wessen, E., S. Hallin, and L. Philippot, *Differential responses of bacterial and archaeal groups at high taxonomical ranks to soil management.* Soil Biology & Biochemistry, 2010. **42**(10): p. 1759-1765.
41. Buckley, D.H. and T.M. Schmidt, *Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil.* FEMS Microbiology Ecology, 2001. **35**(1): p. 105-112.
42. Faoro, H., et al., *Influence of soil characteristics on the diversity of bacteria in the southern Brazilian Atlantic forest.* Applied and Environmental Microbiology, 2010. **76**(14): p. 4744-4749.
43. Bååth, E., *Effects of heavy metals in soil on microbial processes and populations (a review).* Water, Air, and Soil Pollution, 1989. **47**(3-4): p. 335-379.
44. Smith, R.J., et al., *Metagenomic comparison of microbial communities inhabiting confined and unconfined aquifer ecosystems.* Environmental Microbiology, 2012. **14**(1): p. 240-253.

Chapter 4

Nitrous oxide emission related to ammonia-oxidizing bacteria and mitigation options from N fertilization in a tropical soil

Noriko A. Cassman, Johnny R. Soares, Anna M. Kielak, Agata Pijl, Janaina B. do Carmo, Kesia S. Lourenco, Hendrikus J. Laanbroek, Heitor Cantarella and Eiko E. Kuramae

Published as:

Soares JR*, Cassman NA*, Kielak AM, Pijl A, do Carmo JB, Lourenco KS, Laanbroek HJ, Cantarella H & Kuramae EE, 2016. "Nitrous oxide emission related to ammonia-oxidizing bacteria and mitigation options from N fertilization in a tropical soil." *Scientific Reports*. 6, 30349.

* indicates co-first-authorship

Abstract

Nitrous oxide (N_2O) from nitrogen fertilizers applied to sugarcane has high environmental impact on ethanol production. This study aimed to determine the main microbial processes responsible for the N_2O emissions from soil fertilized with different N sources, to identify options to mitigate N_2O emissions, and to determine the impacts of the N sources on the soil microbiome. In a field experiment, nitrogen was applied as calcium nitrate, urea, urea with dicyandiamide or 3,4 dimethylpyrazone phosphate nitrification inhibitors (NIs), and urea coated with polymer and sulfur (PSCU). Urea caused the highest N_2O emissions (1.7% of N applied) and PSCU did not reduce cumulative N_2O emissions compared to urea. NIs reduced N_2O emissions (95%) compared to urea and had emissions comparable to those of the control (no N). Similarly, calcium nitrate resulted in very low N_2O emissions. Interestingly, N_2O emissions were significantly correlated only with bacterial *amoA*, but not with denitrification gene (*nirK*, *nirS*, *nosZ*) abundances, suggesting that ammonia-oxidizing bacteria, via the nitrification pathway, were the main contributors to N_2O emissions. Moreover, the treatments had little effect on microbial composition or diversity. We suggest nitrate-based fertilizers or the addition of NIs in NH_4^+-N based fertilizers as viable options for reducing N_2O emissions in tropical soils and lessening the environmental impact of biofuel produced from sugarcane.

4.1 Introduction

Agriculture is the main anthropogenic source of N_2O emissions, which are predicted to increase as nitrogen fertilizer use increases worldwide to meet the global food demand[1]. Currently, N_2O emissions derived from N fertilizers account for up to 40% of total greenhouse gases (GHG) emissions in ethanol production from sugarcane[2]. High N_2O emissions can negate the benefits of GHG reduction of biofuels used to replace fossil fuels[3].

Emissions of N_2O from soils occur mainly through nitrification and denitrification processes. These processes are carried out by autotrophic and heterotrophic microorganisms belonging to Bacteria, Archaea and Fungi divisions[4-6]. Other N transformations such as nitrifier denitrification, dissimilatory reduction of NO_3^- to NH_4^+, chemo-denitrification and co-denitrification may also produce N_2O. Despite considerable knowledge of the processes evolving N_2O, the prevalence of these processes in tropical soils has only begun to be addressed.

The denitrification process has been demonstrated to contribute more to N_2O emissions than nitrification at soil moisture levels above 75% of the water-filled pore space (WFPS); however, nitrification has been observed to be more prevalent in soil at 60% WFPS[7]. High correlation between N_2O emissions and bacterial *amoA* and *nirK* abundances are observed[8], suggesting that both nitrification and denitrification and/or nitrifier denitrification processes are responsible for N_2O emissions when cattle urine is applied to soils with 100 and 130% of water-holding capacity.

In the central-west and southeast regions of Brazil, about 80% of the land area is cultivated with sugarcane[9]. The dominant soils in these regions are Red Latosols (Hapludox), which are highly weathered, deep and well-drained soils[10]. Here, we expected that denitrification would be low because the optimal conditions are at least 60% WFPS. Though high levels of rainfall and anaerobic conditions in soil micropores may increase the contribution of denitrification to N_2O emissions[11], we predicted based on the high soil drainage that nitrification would be the major pathway contributing to N_2O emissions. In this case NH_4^+-based fertilizer would result in higher N_2O emissions than those from NO_3^--based fertilizers in these soils. Up to date, this process has not been shown for these types of soils grown with sugarcane.

The Intergovernmental Panel on Climate Change (IPCC) estimates that 1% of N applied is emitted as N_2O as default value[12-14]. However, in practice, different amounts of N_2O are emitted depending on N fertilizers and soil types, and environmental conditions[12-14]. Therefore, experiment-based nitrogen manage-

ment is an important tool to decrease N_2O emissions and to reduce the environmental impact of agricultural practices[14].

Urea is the most widely used fertilizer in the world, and generally has been linked to higher N_2O emissions compared with other N sources[13]. One way to reduce N_2O emissions is the addition of specific nitrification inhibitors (NIs) such as dicyandiamide (DCD), 3,4 dimethylpyrazone phosphate (DMPP), nitrapyrin, and others with urea fertilization[14, 15]. These nitrification inhibitors block the enzyme ammonia monooxygenase in the first step of nitrification[16]. The gene encoding this enzyme is *amoA*, present in ammonia-oxidizing bacteria (AOB) and archaea (AOA). Several studies indicate that DCD and DMPP reduce AOB or AOA gene abundances, depending on which microorganism was prevalent[8, 17-19]. DCD has been reported to reduce also the abundance of *nirK*, probably because AOB abundances are correlated with *nirK* abundances, implying reduced nitrifier denitrification abundances[8].

To our knowledge, there are no studies identifying the main microbial processes, the effect of different fertilizers on N_2O emissions, and the impact of different N fertilizers on the microbial community in tropical soils grown with sugarcane. Therefore, the goals of this study were to (i) determine the main microbial process responsible for the N_2O emissions, (ii) evaluate the efficacy of enhanced-efficiency fertilizers, including nitrification inhibitors, in reducing N_2O emissions, and (iii) determine the short-term effects of the fertilizer treatments on bacterial community composition and diversity.

4.2 Methods

4.2.1 Experimental set up

The present experiment was carried out in the 2013/14 season, corresponding to the third ratoon cycle of sugarcane, the variety SP791011, in the experimental area of the Agronomic Institute in Campinas, Brazil (22º52’15” S, 47º04’57” W). The soil in the area was classified as Typic Hapludox or Red Latosol[10, 21]. The same experiment was carried out during the seasons of 2011/12 and 2012/13[15]. However, in the 2013/14 season an extra treatment with calcium nitrate was included to consider N_2O emissions due to nitrification or denitrification processes. Here, soil samples were collected in order to associate greenhouse gases (GHG) emissions with the microbial processes that were involved. The treatments were: 1) Control plot without N fertilization (control); 2) urea (UR); 3) UR + DCD; 4) UR + DMPP; 5) Polymer and Sulphur Coated Urea (PSCU); 6) UR + DCD-R; 7) UR + DMPP-R; 8) Calcium Nitrate. R stands for reapplication of in-

hibitors in the same plots during the previous two cycles of the experiment. The fertilizers were applied on 19 December 2013, 20 days after the harvest of the previous cycle. Phosphorus and potassium were concurrently applied to all plots at rates of 20 and 100 kg ha^{-1} of P and K, respectively.

Nitrogen was applied at a rate of 120 kg ha^{-1}; the nitrification inhibitor DCD (Sigma Aldrich) was added in a dose of 5% DCD-N in relation to urea-N whereas DMPP (powder form) was added as 1% DMPP (w/w) to urea-N; PSCU was produced by Produquímica (Produquímica Ltda, Brazil) and calcium nitrate by Yara (Yara International ASA). Fertilizers were incorporated at a 5 cm soil depth to avoid NH_3 volatilization from urea and the effect of NIs on this N loss[35]. The fertilizers were applied on either side of the plant row, 10 cm away from the recently harvested sugarcane plants. On one side of the plant row the greenhouse gases were measured; on the opposite side of the same plant, soil for chemical and molecular microbial analyses was collected.

Sugarcane yields were not measured in this study because the amount of N lost as N_2O is generally much too low to affect yields. Furthermore, the plot size necessary to evaluate yields usually exceeds 100 m^2. Because our focus was on GHGs emissions, which are dependent on localized soil conditions, small plots were chosen. In our study, large plots were not only unnecessary but would contribute to noise in the gas flux data.

4.2.2 Greenhouse gases analysis

Greenhouse gases were collected using static chambers[15]. Chambers were fixed in the soil 5 cm deep along two 25-m long rows of sugarcane. In total, 32 chambers were used, with four replicates per treatment, in a completely randomised design. Gases were sampled in the morning and three times per week during the first three months after fertilizer application, then biweekly as previously done[15]. In each sampling date, gas samples were taken at three time intervals: 1, 15, and 30 minutes.

After sampling, the gases were immediately stored in pre-evacuated Extainers vials (Labco Limited, Ceredigion, United Kingdom) and analysed in a Shimadzu gas chromatograph (GC-2014). Cumulated gas emissions were calculated by linear interpolation between gas samplings periods. Details of the procedures used for gas analysis and calculations are described elsewhere[14, 20].

4.2.3 Soil chemical analysis

Soil samples (0-10 cm depth) were collected more intensively in the first two months after fertilizer application, a period corresponding to higher N_2O emissions. Using an auger, three subsamples were collected as a composite sample per experimental plot. In total, eight soil sampling campaigns were collected at 7, 16, 18, 27, 35, 42, 82 and 158 days after fertilizer application. The soil samples were stored in plastic bags at -20ºC. Gravimetric moisture after constant weight was attained at 105 ºC. The water-filled pore space (WFPS) was calculated considering soil bulk density and porosity determined at the beginning of the experiment. Soil pH was measured in $CaCl_2$ (0.0125mol L^{-1}) and NH_4^+-N and NO_3^--N contents were determined by steam distillation after soil extraction in 2 mol L^{-1} KCl solution[36].

4.2.4 Real-time PCR analysis

Soil subsamples (20 g) were stored at -80ºC for molecular analyses. Total soil DNA was extracted from 0.25 g of soil using the Power Soil kit (Mobio, Carlsbad, CA USA) following the manufacturer's instructions. The quantity and quality of DNA was measured by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Montchanin, USA). The DNA samples were diluted in water free of DNase and RNase (Sigma Aldrich) and the abundance of the genes encoding for nitrification and for denitrification processes were quantified by quantitative real-time PCR with a Qiagen Rotor-Gene Q6000 cycler (RO212226). A reaction was performed in total volume of 12 µl, containing 6 µl Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.5 µl of each primer (5 pmol) and 5 µl of DNA (3 ng). Exceptions were the reaction for the *nirK* amplification, for which the Sybrgreen Qiagen Rotor-Gene SYBR Green PCR Kit was used, and the *nosZ* amplification, for which the starting DNA concentration was 30 ng. Reactions were performed by a QIAgility robot (003516).

The thermal conditions of each gene amplification are listed in Supplementary Table S5. Acquisition was done at 72°C (cycle A) or 82-86°C (cycle B) to avoid primer dimers. Melt curve analysis was done at 55-99°C to confirm specificity; the qPCR products were checked by agarose gel electrophoresis to confirm the desired amplicon size. Plasmid DNA from microorganisms containing the gene of interest or from environmental samples were used for the standard curve and then cloned into vectors as described in Table S5. Normal PCR reactions were carried out with similar thermal conditions as qPCR to confirm the fragment size of interest, then cloned and transformed into JM109 High Efficiency Competent Cells (Promega, In Vitro Technologies, Auckland, New Zealand). After overnight bacterial growth in LB medium with ampicillin at 37ºC, plasmids were extracted

using the PureLin Quick Plasmid Miniprep Kit (Life Technologies, Auckland, New Zealand). The quantity and quality of plasmid DNA were checked by spectrophotometer (NanoDrop ND-1000 Technologies, Montchanin, USA). Standard dilutions were obtained from 10 to 10^8 copies/µl of each gene. Each run included a DNA template, standard, and a no-DNA control – water free of DNase and RNase (Sigma Aldrich) – done in duplicate. Reaction efficiency was 89-105% and R^2 values ranged from 0.94 to 0.99.

4.2.5 16S rRNA partial gene sequencing

To assess the impact of the treatments on the bacterial community, we sequenced the 16S rRNA gene marker from total DNA extracted from the soil samples. The V4 region of the 16S rRNA gene was amplified by using archaeal/bacterial primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACVSGGGTATCTAAT-3'). The samples were PCR-amplified using barcoded primers linked with the Ion adapter "A" sequence (5′-CCATCTCATCCCTGCGTGTCTCCGACTCAG-3′) and Ion adapter "P1" sequence (5′-CCTCTCTATGGGCAGTCGGTGAT-3′) to obtain a sequence of primer composed for A-barcode-806R and P1-515F adapter and primers. The 16S rRNA gene amplifications for library preparation were performed on the C1000 thermocycler (Biorad, Hercules, CA, USA) with thermal conditions of 95°C-5 min.; 35x 95°C-30s, 53°C-30s, 72°C-60s; 72°C-10 min. A reaction of 25 µl in total was done, including 2.5 µl of 10X PCR Buffer, 2.5 µl dNTPs (200 µM), 0.25 µl of each primer (0.1 pmol/µl), 0.2 µl of fast startExp-Polymerase (0.056 U) and 1 µl of DNA (0.6 ng). The reactions were carried out in duplicate and included a negative control. The amplicons were checked by gel electrophoresis. The PCR products were purified by Agencourt AMPURE XP to remove primer dimers, then quantified by Quant-iT PicoGreen and equimolar mixed for sequencing using the PGM Ion Torrent (Life Technologies).

4.2.6 16S rRNA amplicon sequences processing

MOTHUR Version 1.34.2 was used to process the 16S rRNA partial genes sequences, implemented using a Snakemake workflow on a 32-node server running Linux Ubuntu 14.4[37]. Forward and reverse primer sequences were removed from each sample FASTQ file using Flexbar version 2.5[38]. Reads were filtered based on sequence quality by running the Sickle tool (minimum quality score 25, minimum length 150). Filtered reads were converted to FASTA format and concatenated into a single file, then clustered into OTUs using the UPARSE

strategy of dereplication, sorting by abundance with at least two sequences and clustering using the UCLUST smallmem algorithm[39]. These steps were performed with VSEARCH version 1.0.10, which is an open-source and 64-bit multithreaded compatible alternative to USEARCH. Chimeric sequences were detected using the UCHIME algorithm[40] implemented in VSEARCH. All reads before the dereplication step were mapped to OTUs using the USEARCH_global method implemented in VSEARCH to create an OTU table and then converted to the BIOM-Format 1.3.1[41]. Last, taxonomic information for each OTU was added to the BIOM file using the RDP Classifier version 2.10[42].

4.2.7 Statistical analysis – gas fluxes and gene abundances

Daily GHG fluxes, cumulated emissions of N_2O, CO_2, CH_4 and gene abundance values were checked for normal distribution of residues by Shapiro-Wilk test, and then submitted to variance analysis (ANOVA) and the means compared by Tukey's test at $P \leq 0.05$. Soil pH was transformed to H^+: 10^{-pH} before statistical analysis. Linear correlations between N_2O fluxes and environmental variables were evaluated at the 5% level of significance. Multiple linear regressions, which were selected by the stepwise process at $p \leq 0.05$, were fitted between N_2O fluxes and environmental variables. When necessary, the N_2O flux values were log(x) transformed and rechecked to obtain a normal distribution of residues and variance stability[43]. The calculations were performed with the SISVAR statistical software[43] and graphics plotted using Sigma Plot[43].

Cumulative N_2O emissions as a function of time were fitted by sigmoidal or exponential equations, for which the sigmoid equation was:

$$N_2O = \frac{a}{1 + e^{-(\frac{t-t_0}{b})}} \quad (1)$$

where N_2O is the cumulative N_2O-N emission, g ha-1, t is the time in days after fertilizer application, and a, t0 and b are equation parameters, where a is the maximum loss and t0 is the time in which 50% of maximum loss occurs.

The exponential rise to the maximum model had the following equation:

$$N_2O = a(1 - e^{-bt}) \quad (2)$$

where N_2O is the cumulative N_2O-N emission, g ha^{-1}, t is the time in days after fertilizer application, and a and b are equation parameters, where a is the maximum loss and b is the rate of rise.

4.2.8 Statistical analysis – 16S rRNA amplicon sequence data

The 16S rRNA samples were analysed to compare bacterial community alpha diversity and composition across treatments and time point (days 7, 16, 18, 27, 35, 42, 82 and 158). The BIOM files were handled with the "phyloseq" package[44] in R[50]. Rarefaction curves were generated to ensure adequate sequencing depth across samples. Discarding undersequenced samples, the minimum sample size was 2000. For alpha diversity analyses, the 16S rRNA samples were rarefied to 2000 sequences using the "vegan" package[51]. Renyi diversities at alpha level 1, corresponding to the Shannon diversity index were kept ("BiodiversityR" R package). The Shannon diversity data was furthermore subjected to Kruskal-Wallis tests among treatments and the Kruskal-Wallis multiple comparison test between treatments using the "pgirmess" R package.

Comparisons of bacterial community compositions were evaluated using the Statistical Analysis of Metagenomic Profiles (STAMP) software[52]. The top nine Bacterial phyla based on relative abundances across all samples were compared among and between treatments for each time point. The unclassified sequences were removed prior to analysis. The ANOVA statistical and Tukey-Kramer post-hoc tests (CI 95%) were applied using the Benjamini Hochberg multiple test correction. To explore beta diversity (treatment differences) of the bacterial communities, Between-Class Analysis (BCA) of the non-rarefied 16S rRNA samples grouped by treatment was performed using the "ade4" R package[53]. First, with unclassified sequences removed, correspondence analyses of the compositional data agglomerated at the rank of Phylum and Genus were conducted, followed by BCA. Further, the BCA groups for the phyla and genera analyses were tested using the Monte-Carlo permutation method with 999 repetitions.

4.3 Results

4.3.1 Greenhouse gases emissions

Emissions of N_2O were low in the first 10 days after fertilizer application, with less than 10 g ha^{-1} day^{-1} of N emitted (**Figure 1**). High N_2O emission followed rain events coupled with high soil inorganic N availability (**Figure 2**). The UR treatment had the highest peak of N_2O emission, on the 14th day, which corresponded to a loss rate of more than 200 g ha^{-1} day^{-1} of N. On the 29th day, another high emission peak of about 170 g ha^{-1} day^{-1} occurred. Between these peaks, N_2O emissions were still relatively high ranging from 15 to 70 g ha^{-1} day^{-1} in the UR treatment. The treatments UR+DCD, UR+DMPP, and calcium nitrate had smaller N_2O fluxes than those of UR, showing emission levels similar to the control

treatment (around 5 g ha^{-1} day^{-1} of N). Urea containing NIs (UR+DCD-R, UR+DMPP-R) had been reapplied in the same plots in the previous two years[15]. Emissions in the plots with repeated application of NIs were also low at ≤ 5 g ha^{-1} day-1 of N (**Figure 1**). The controlled release fertilizer PSCU treatment showed lower N_2O emission (80 g ha^{-1} day^{-1}) compared to the UR treatment on the 14th and 29th days, but was similar to UR treatment levels afterwards, until the 50th day. Between 70 and 120 days after fertilizer application, N_2O emissions were greater in the PSCU treatment (between 10 – 20 g ha^{-1} day^{-1}) compared to the other treatments (2 g ha^{-1} day^{-1}) (**Figure 1**).

Cumulated N_2O emissions in the control treatment were equivalent to 0.3 kg ha^{-1} after 278 days. The UR treatment emitted more than 2.3 kg ha^{-1} of N_2O-N, which corresponded to 1.7% of total N applied. The UR+DCD, UR+DMPP treatments resulted in considerable reductions in cumulated N_2O emissions compared to UR, with emissions that did not differ from those of the control (**Table 1**). The reduction of N_2O emissions by addition of NIs to UR varied from 88 to 97% (95% in average). The PSCU treatment resulted in cumulative emissions similar to those of UR. Calcium nitrate resulted in low N_2O emissions that did not differ from those of the UR+DCD and UR+DMPP treatments or the control (**Table 1**).

Cumulative N_2O emissions data were fit with sigmoidal or exponential equations for the control, UR and PSCU treatments. These models were not applied to the UR+DCD, UR+DMPP, and calcium nitrate treatments because in these treatments the N_2O emissions did not differ from the control. The N_2O emissions in the control treatment were low and whether included or excluded did not affect the model equations. The UR treatment achieved 90% of maximum N_2O loss 40 days after fertilizer application. The N_2O emission lasted longer with the PSCU treatment than with UR; 90% of total N_2O emission was achieved 187 days after fertilization with PSCU (**Figure 3**).

The total CO_2 and CH_4 emissions were around 6 t ha^{-1} and -600 g ha^{-1}, respectively. For both CO_2 emissions and CH_4 consumption, no differences between the treatments were observed (**Table 1**).

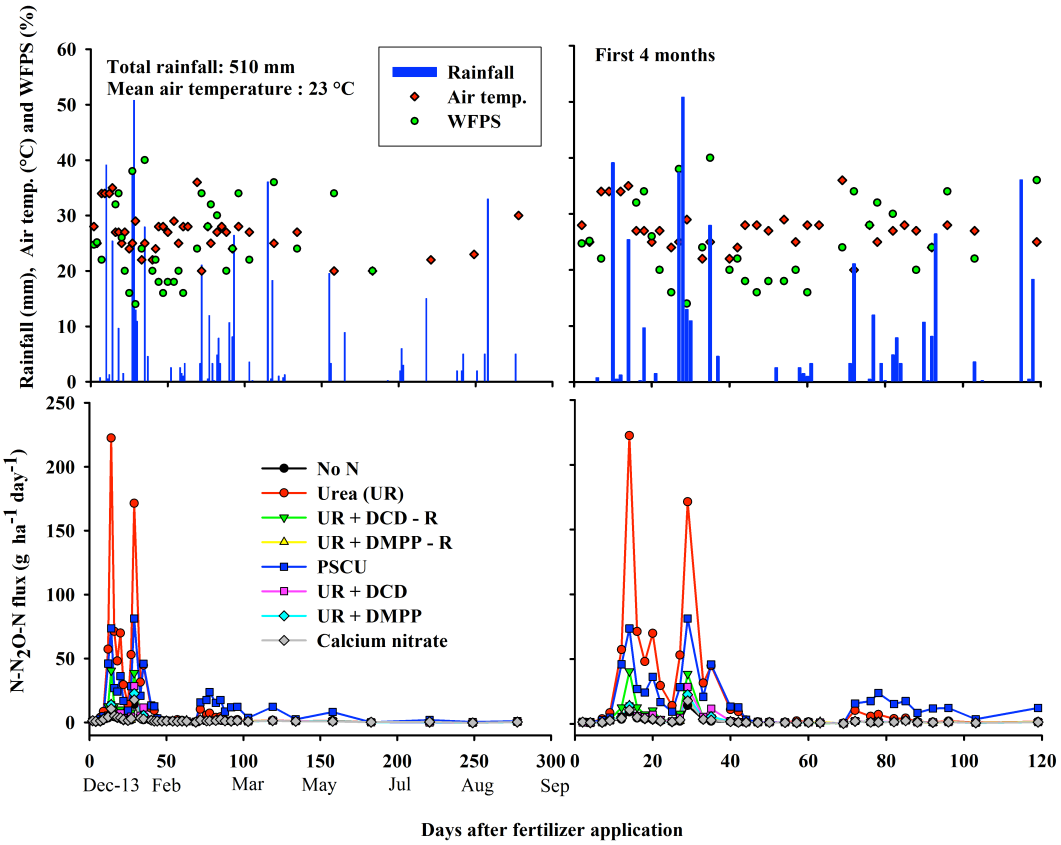


Figure 1. Rainfall, air temperature, water-filled pore space (WFPS) and nitrous oxide fluxes from Control (No N), urea (UR) with or without nitrification inhibitors (DCD and DMPP), polymer sulphur coated urea (PSCU) and calcium nitrate applied to sugarcane. R: reapplication of inhibitors in the same plots during the previous two cycles of the experiment. N fertilizers were applied on 13 December 2013.

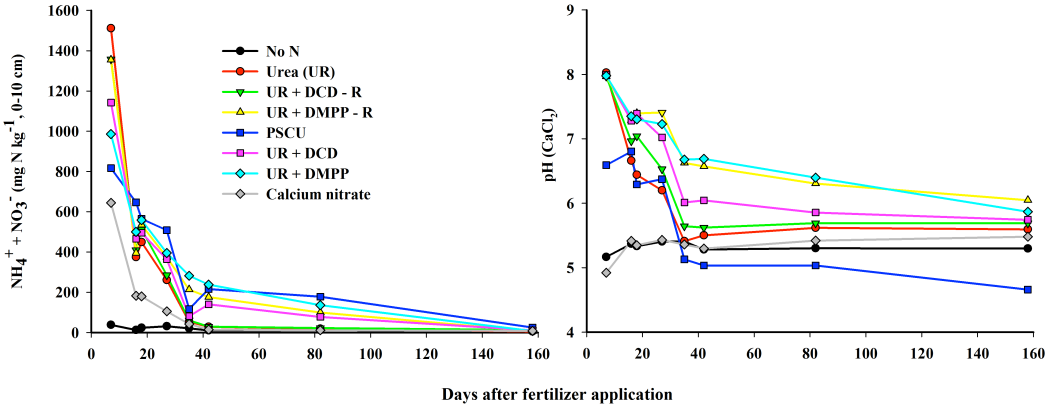


Figure 2. Soil concentration of NH4+-N + NO3--N and pH from Control (No N), urea (UR) with or without nitrification inhibitors (DCD and DMPP), polymer sulphur coated urea (PSCU) and calcium nitrate applied to sugarcane. R: reapplication of inhibitors in the same plots during the previous two cycles of the experiment.

Table 1. Cumulated nitrous oxide, carbon dioxide and methane emissions from Red Latosol soil during 278 days after application of urea with or without nitrification inhibitors (DCD and DMPP), polymer sulphur coated urea (PSCU) and calcium nitrate applied to sugarcane.

Treatments	N_2O-N					CO_2-C		CH_4-C	
	g ha^{-1}	log*		% N applied†	Reduction (%)	kg ha^{-1}*		g ha^{-1}*	
Control	286	2.4	C	-	-	5835	ns	-598	ns
UR	2301	3.4	A	1.68	-	5933		-633	
UR+DCD-R	531	2.7	B	0.20	88	5883		-612	
UR+DMPP-R	350	2.5	C	0.05	97	5871		-532	
PSCU	2165	3.3	A	1.57	7	5912		-648	
UR+DCD	410	2.6	Bc	0.10	94	5859		-633	
UR+DMPP	353	2.5	C	0.06	97	5897		-656	
Calcium nitrate	329	2.5	C	0.04	98	5973		-600	
P value		<0.00001				0.9769		0.9328	

* Tukey test, $p \leq 0.05$; ns: no significant; N_2O-N: g ha^{-1} transformed in log(X) † Results from treatment without N were subtracted for this calculation. - R means reapplication of inhibitors in the same plot in the two preceding years. Different characteristics in the column of N_2O mean significant differences ($p<0.05$) between the values.

4.3.2 Soil analysis

In the first soil sampling seven days after fertilizer application, the calcium nitrate treatment showed N inorganic content ($NH_4^+ + NO_3^-$) around 600 mg kg^{-1}, which was lower than the 1000 mg kg^{-1} of N in the 10 cm soil layer found in the UR treatment (**Figure 2**). Afterwards the N content in soil decreased exponentially; during this time, nitrification inhibitors in the UR+DCD and UR+DMPP treatments maintained soil N mostly in the NH_4^+ form (**Supplementary Figure S1**). At the 82th day, soil treated with PSCU showed N content higher than other treatments, near 200 mg kg^{-1} of N, as opposed to 100 mg kg^{-1} from the UR treatment.

The original soil pH (control treatment) was around 5.1 but increased to 8 in the UR, UR+DCD and UR+DMPP treatments by seven days after fertilizer application, because of urea hydrolysis. PSCU showed pH value 1.4 units lower (pH 6.6) but treatments with calcium nitrate did not affect soil pH. In the 42th day, the soil pH of the UR+DCD and UR+DMPP treatments had dropped to values around 7 (**Figure 2**).

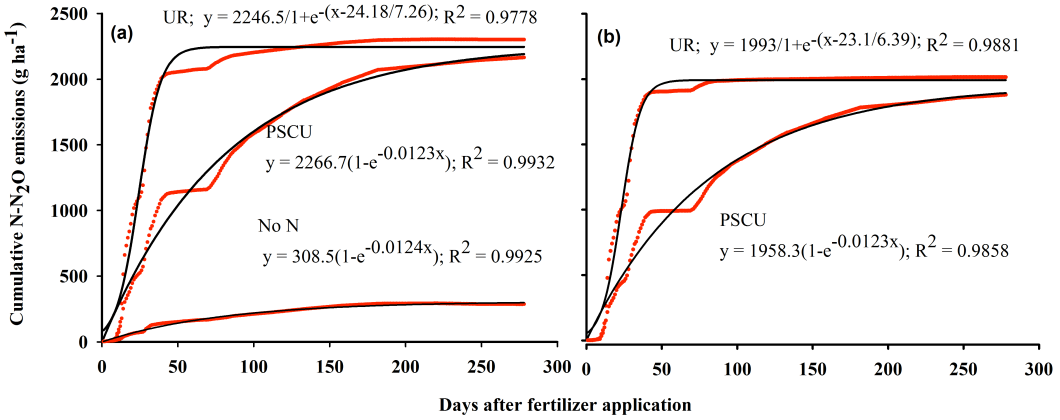


Figure 3. Cumulative N_2O emission (red dots) and sigmoidal or exponential equations fitted (black lines) to data of urea (UR) and polymer sulphur coated urea (PSCU) applied to sugarcane. (b) net UR and PSCU N_2O emissions calculated by subtracting data of the control treatment.

4.3.3 Nitrogen cycle gene abundances

The abundances of N cycling genes related to N_2O emissions are depicted in **Figure 4** for one timepoint sampling that featured high N_2O emissions: 16 days after fertilizer application, corresponding to the second soil sampling. The qPCR results from all data sampling timepoints are available in **Supplementary Table S1**. The abundance of *amoA* belonging to ammonia-oxidizing archaea (AOA) was lower in treatments with N sources than in the control plot and did not show significant differences among the N sources across nearly all data sampling points. The correlation between AOA *amoA* abundance and N_2O emissions was negative (**Supplementary Table S2**). The gene abundance representing total archaea showed a similar pattern as the AOA *amoA* abundance (**Figure 4** and **Supplementary Table S1**). On the other hand, the *amoA* abundance of ammonia-oxidizing bacteria (AOB) was best correlated with N_2O emissions, showing a coefficient (R^2) of 0.18 ($p \leq 0.05$) (**Supplementary Table S2**). Over almost all data sampling points, AOB *amoA* abundances were higher in the UR treatment than in other treatments, following the data from N_2O emissions (**Figure 4** and **Supplementary Table S1**). For example, concurrent to high N_2O emissions at day 16 after fertilizer application, the coefficient of correlation (R^2) between N_2O and AOB *amoA* was 0.53 (**Supplementary Figure S2**).

The denitrification genes *nirS* and *nosZ* as well as the 16S rRNA gene of total bacteria did not show differences in abundance between treatments over almost all data samplings. The *nirK* occurring in both ammonia-oxidizing and denitrification microorganisms had a negative correlation with N_2O emissions, while

the total bacteria abundance resulted in a positive correlation with this emission (**Supplementary Table S2**). In some data samplings, the abundance of the *nosZ* was higher in treatment with N than the control treatment, but no differences in *nosZ* abundance were observed between the N sources treatments (**Figure 4** and **Supplementary Table S1**).

A good fit, with an R^2 of 0.47, was obtained by stepwise regression model relating N_2O emissions to environmental variables, including the AOB *amoA* abundance, rain amount accumulated one week before GHGs measurement, NH_4^+-N and NO_3-N contents, total bacteria abundance, pH, and CO_2 emission (**Supplementary Table S3**). Removing the treatments without nitrification inhibitors, NH_4^+-N content in soil was correlated with N_2O, while NO_3^--N was not (**Supplementary Tables S2 and S3**).

4.3.4 Bacterial community composition and diversity

Because the AOB *AmoA* abundance was correlated with N_2O emissions, we sequenced the 16S rRNA genes from our samples to ascertain the effect of the treatments on the entire microbial (bacterial and archaeal) community. After processing the 16S rRNA amplicon sequences, the 177 samples (8 treatments x 8 timepoints x 3 replicates, excluding undersampled samples and outliers) contained between 2,000 and 56,638 sequences, with a total of 3,607,143 sequences distributed into 9,267 Operational Taxonomic Units (OTUs). Rarefaction curves indicated that most of the community diversity was captured with our sequencing depth (**Supplementary Figure S3**). The top nine bacterial phyla across the samples were *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, *Gemmatimonadetes*, *Proteobacteria* and *Verrucomicrobia* (**Supplementary Table S4**). The top phyla that differed between the treatments within at least one timepoint were *Firmicutes*, *Bacteroidetes*, *Nitrospira*, *Proteobacteria*, *Verrucomicrobia* and *Acidobacteria*. Shannon diversity indices of the bacterial communities ranged between 5.3 and 6.3 in treatments over all time points and were significantly different between treatments only for days 27 (Control *versus* UR) and 82 (PSCU *versus* UR+DMPP; **Table 2**).

Based on phylum-level relative abundances, the samples were significantly grouped by treatment on days 7, 27, 35, 42, 85 and 158 (permutation test, $P < 0.100$; **Table 2**) with a range of observed values between 0.83 and 0.44. At day 7, three separate clusters were formed respectively with the Control and calcium nitrate treatments, the PSCU treatment and the other treatments (**Supplementary Table S5**). However, this clustering pattern was not observed during the remain-

ing other days, indicating no differences in the bacterial communities between treatments. When bacterial community compositions were compared at the taxonomical level of genera, the samples were significantly grouped by treatment and the grouping was characterized by low observation values for all days except day 16 (P=0.119; **Table 2** and **Supplementary Table S5**). The sample treatment grouping pattern from day 7 was seen in the genus-level comparisons (**Supplementary Table S5**).

Table 2. Shannon indices and Between-Class Analysis (BCA) ordinations of the bacterial communities present in the Red Latosol soil under treatments with urea with or without nitrification inhibitors (DCD and DMPP); polymer sulphur coated urea (PSCU) or calcium nitrate applied to sugarcane.

Treatments	16S rRNA gene diversity (Shannon index)													
	7 DAF*		16		18		27		35		42		85		158	
Control	6.2	ns	6.1	ns	6.0	ns	6.2	a	6.1	ns	6.2	ns	6.0	ab	6.1	ns
UR	5.8		6.0		6.0		5.3	b	5.7		5.8		5.6	ab	6.0	
UR+DCD-R†	5.9		6.2		5.8		6.0	ab	6.0		6.1		5.8	ab	6.2	
UR+DMPP-R	5.6		6.1		5.9		5.8	ab	6.1		6.2		6.1	ab	6.2	
PSCU	5.9		5.9		5.8		5.8	ab	5.1		5.2		5.4	a	5.8	
UR+DCD	5.7		6.1		5.7		5.8	ab	5.9		6.0		6.0	ab	6.2	
UR+DMPP	5.7		6.0		5.9		5.8	ab	5.5		6.1		6.1	b	6.1	
Calcium nitrate	6.3		6.0		5.7		5.9	ab	5.9		6.2		6.0	ab	6.2	

*DAF: Days after fertilizer application. Means in column followed by same letter did not differ; ns: no significant. † R means reapplication of inhibitors in same plots.

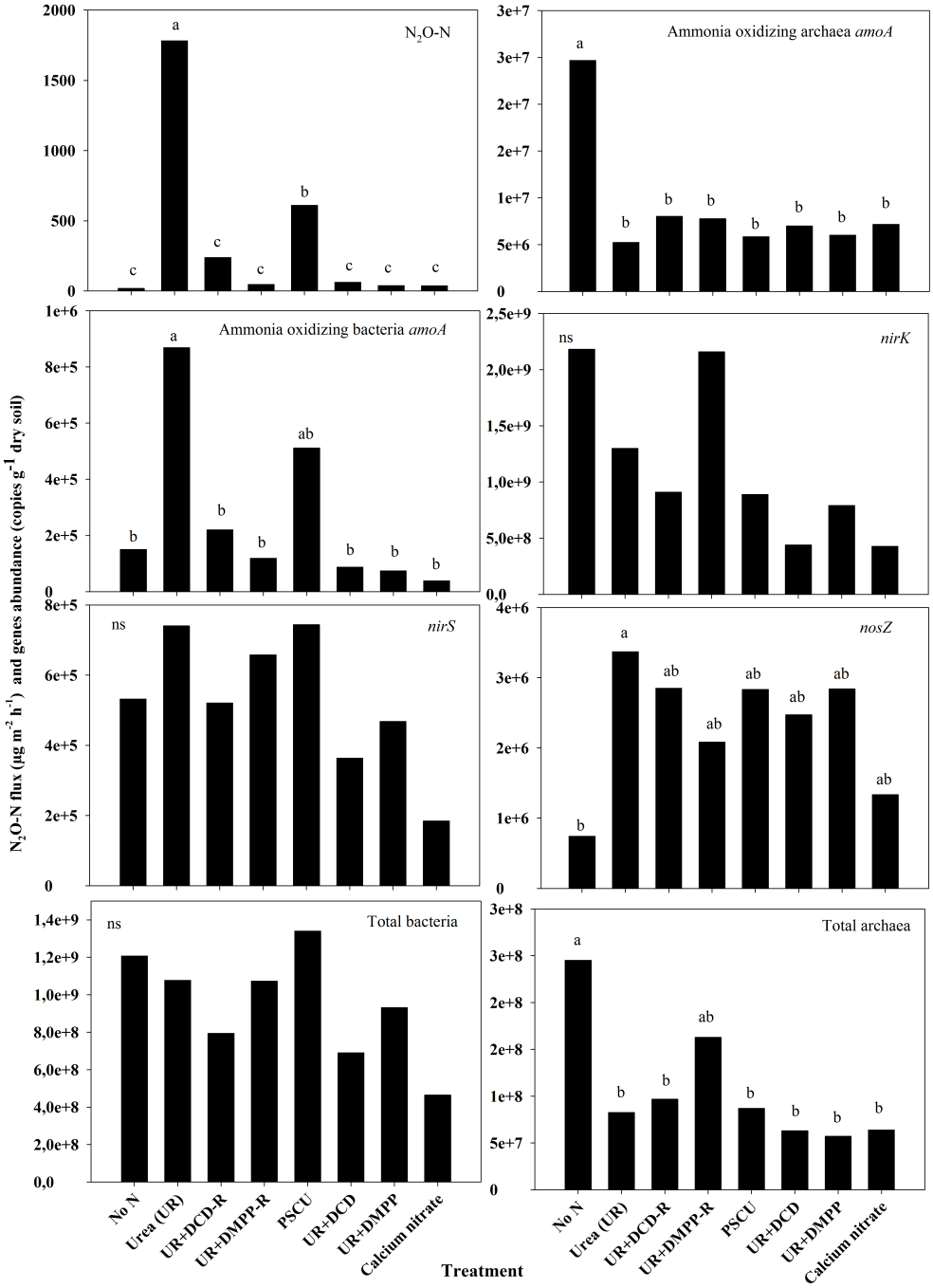


Figure 4. Nitrous oxide fluxes, nitrogen cycle genes (*amoA* bacteria, *amoA* archaea, *nirK*, *nirS*, *nosZ*) abundances and total bacteria and total archaea abundances in the Red Latosol soil 16 days after fertilizer application of urea with or without nitrification inhibitors (DCD and DMPP); polymer sulphur coated urea (PSCU) or calcium nitrate applied to sugarcane. R: reapplication of inhibitors in the same plots during the previous two cycles of the experiment.

4.4 Discussion

Nitrous oxide emissions from the urea treatment were higher than the emissions found in two previous sugarcane cycles in this experimental area[15]. Here, the emission factor was 1.7% of N applied, which is greater than the emission factors found in other studies of sugarcane soils in Brazil, around 0.7 – 1% of N applied as urea[12, 15, 20]. The sugarcane plant phenology may give insight into higher emissions. The N fertilizer treatments were applied 20 days after the previous sugarcane ratoon was harvested. At the time of fertilization, the soil was relatively dry – below 30% of the WFPS (Figure 1) and sugarcane plants were still beginning to sprout. At this stage the root system was being reformed and nutrient uptake was slow, as the high soil N concentration indicated (Figure 2). This probably led to the greater N_2O emissions than expected. The high peaks of N_2O emission occurred after two high rain events in the first 35 days (total 65 mm and 90 mm) on a mostly dry soil (15 – 20% WFPS) but with high air temperatures as shown in Figure 1. Moreover, high correlation between N_2O emission and accumulated rain in the week was found (Supplementary Table S2). Thus, the climatic conditions in this season contributed to the high N_2O emission values.

A strong reduction in N_2O emissions due to the addition of nitrification inhibitors to urea was found, as well as a lack of beneficial effects of the controlled release fertilizer PSCU, supporting similar observations in the same area[15]. The N_2O emissions from PSCU were lower than those from UR in the first 30 days after N application, as expected from a slow-release fertilizer (Figure 2). There was a dry spell from mid-January to the end of March (Figure 1), which may have slowed down the release of N from PSCU. Subsequent release of N from the PSCU pellets likely led to the observed increase in N_2O emissions (Figure 1). In this way, the N_2O emission from PSCU had lower peaks than those of UR but lasted longer (Figure 3). Thus, in the end of the experiment cumulated N_2O of UR and PSCU emissions were similar, suggesting that PSCU is not an environmentally friendly N source during one cycle of sugarcane.

In the present study, the calcium nitrate treatment showed very small N_2O emissions that were similar to those of the control plots or plots with urea and nitrification inhibitors (NIs). In the present study intensive GHG measurements under field condition were performed over a whole yearly cycle of sugarcane. We maintain that this is the first field study demonstrating much lower N_2O emissions of a nitrate-N source in comparison to high emissions with urea-N or NH_4^+-N sources; the reduction in N_2O emissions were 98% when compared to urea. This reduction in emissions might be attributed to the high drainage capacity of the soil of the present study, which was classified as Typic Hapludox[21] or Red

Latosol[10]. Water accumulation does not tend to occur in these soil profiles, and consequently, favourable conditions for denitrifiers are avoided.

Under controlled conditions with ^{15}N-labeled sources, denitrification was more important at the high soil moisture (75% WFPS), while N_2O emissions with NH_4^+ fertilizers were two times higher than with NO_3^- fertilization at 60% WFPS[7]. In our study WFPS reached a maximum of 40% (Figure 1), which is more favourable for nitrification. The O_2 concentrations were likely not low enough to favour the denitrification process in relation to nitrification.

An alternate explanation for the low N_2O emissions in the calcium nitrate treatment could be NO_3^- leaching. Indeed, the N concentration in the 0-10 cm soil layer of the calcium nitrate treatment was lower than that observed with the other N sources (Figure 2). However, with only 70 mm of rain on a dry soil in 15 days, nitrate is unlikely to have moved beyond 30 cm. Moreover studies with ^{15}N labelling showed little NO_3^- leaching in sugarcane fields in Brazil[22, 23]. Therefore, NO_3^- leaching was not expected to explain the small amount of N_2O emission found with calcium nitrate in the present study. However, further studies should include NO_3^- leaching measurements to confirm the present data.

Another aspect that may have contributed to the small N_2O emission in the calcium nitrate treatment was the relatively low organic carbon content in the soil, approximately 1%[15]. Sugarcane trash and vinasse have been reported to increase N_2O emissions[24], especially under high soil moisture conditions[24]. Here, we did not include in our treatments C sources such as vinasse, filter cake or sugarcane trash, common sugarcane residues or by-products. These residues can favour not only denitrifiers but also nitrifiers and other microorganisms related to the N cycle[25]. Application of exclusively NO_3^--N sources with the addition of C sources, as commonly applied during sugarcane production, may result in N_2O emissions different from those observed here and deserves further attention.

Smaller N_2O emissions from calcium nitrate as compared to UR or NH_4-based fertilizer have been previously reported[26, 27]. However, in one study, N_2O emissions observed with all the studied N sources were low (around 0.5% of the N applied), which makes it difficult to compare the treatments[27]. In a field grown with maize in Brazil, no differences were reported between UR and calcium nitrate in N_2O emissions, but the emission factor was 0.2% of the N applied[11]. In our study, the N_2O emissions from UR were high at 1.7 % of N applied; the low N_2O emission under calcium nitrate occurred concurrently to high N_2O emissions from UR treatment, which highlights the relevance of the present study.

If soil moisture conditions are favourable to denitrification, nitrate-based N fertilizers may produce higher N_2O emissions than urea or ammonium fertilizers. That is the case of the study conducted in a Gleysol soil in which the WFPS was above 60% during most of the experimental period[28].

Based on Between-Class ordinations of the 16S rRNA compositional data as well as the total 16S rRNA gene copy numbers, the bacterial community appeared to be more affected by sampling day than by treatment. This suggests overall a minimal impact of the treatments on bacterial community composition and diversity. Though further work should examine the long-term impacts, there appears to be a low short-term impact of NIs on the bacterial community. Culturing or shotgun metagenomic and metatranscriptomic techniques may provide future avenues to illuminate the activity of specific nitrifiers under the environmental conditions in this study and to enhance predictions of N_2O emissions due to nitrification in tropical soils.

Archaeal *amoA* abundances were highest in the control treatment compared to the treatments with any N sources. Elevated ammonia concentrations and higher soil pH are suggested to favour bacteria compared to archaea[8, 17, 29]. Interestingly, the plots with calcium nitrate also showed a reduction in archaeal *amoA* abundance compared to the control. This may reflect the accumulated effect of ammonium nitrate applied in the two previous cycles as this was the N source previously used in a separate study[15].

Significant correlations between N_2O emissions and *amoA* abundances were found for ammonia-oxidizing bacteria (AOB), indicating that in our study, N_2O emissions occurred via nitrification. During the first month after UR application, a peak in the AOB *amoA* abundances was observed. Concurrently, the nitrate content in soil increased whereas ammonium decreased, soil pH decreased from 8 to around 6.3, and the soil temperature was 25ºC (Supplementary Figure S5). Thus, the soil and climatic conditions were favourable to AOB and N_2O emissions from nitrification.

Apart from identifying nitrification as the likely source of N_2O emissions here, the data also suggest that denitrification was very low. No differences among treatments nor significant correlations between N_2O emissions and the genes encoding denitrification process as *nirK*, *nirS* and *nosZ* abundances were observed. Further, the model that best estimated N_2O emissions included bacterial *amoA* abundances and N present in the NH_4^+ form but not in NO_3^-. Thus, our gene abundance data supported the results of the low N_2O emission data obtained from application of nitrate as the N source.

Significant correlation between AOB *amoA* abundance and N_2O emissions was also shown. Under controlled conditions, Venterea et al.[30] found a high correlation of N_2O emissions from urea and NO_2^--N content in soil resulting from increased bacterial *amoA* abundance with no increase in the abundance of the *nxr* gene, which encodes for nitrite oxidation. The authors discussed that N_2O emissions occurred more during nitrification than denitrification, similar to the results found here. Dicyandiamide (DCD) application with cattle urine effectively inhibited the growth of AOB and reduced N_2O emissions as well as the numbers of the *nirK* gene, which encodes for a nitrate reductase[8]. Since DCD did not affect the abundance of other denitrification genes, the authors concluded that AOB, including nitrifier denitrifiers containing *nirK*, were the main contributors to N_2O emissions[8]. In the present study no evidences relating *nirK* and N_2O emissions was found, but the nitrifier denitrification process could have great contribution to N_2O released due the presence of the gene *norB* in AOB [4,5,32]. Besides nitrifier denitrification, nitrous oxide could be emitted during oxidation of hydroxylamine by ammonia-oxidizing bacteria[31], heterotrophic nitrifiers[32, 33], and/or abiotic chemodenitrification[34]. Abiotic N_2O emissions also occur due nitrite reduction by organic and inorganic compounds as amine, copper and iron[4,37]. Others processes that could be involved in N_2O emissions are abiotic or biotic co-denitrification, by archaea, bacteria or fungi. In co-denitrification, a reducing compound as NO^-, NO_2^- or NO_3^- combined with organic N, hydroxylamine or ammonium generates N_2O emissions in oxic and anoxic conditions[6, 34]. More studies targeting these reactions can pin down the relative contribution of factors explaining N_2O fluxes from nitrification. The present study showed high N_2O losses from urea, but very small from a nitrate fertilizer source and nitrification was the most relevant microbial process associated with such losses, which has not been reported in soil with sugarcane. The relationship between N_2O emissions and bacterial *amoA* abundances may, therefore be a useful indicator for N management strategies to mitigate N_2O emissions in tropical soils. Other classes of soils and N sources are necessary to confirm our data.

4.5 Declarations

Acknowledgments

The authors thank Dr. Eoin L. Brodie (Lawrence Berkeley National Lab) and Dr. Adriana P. D. Silveira (IAC) for scientific discussions, and Rafael M. Sousa (IAC) for technical assistance. This research was supported by grants from The Netherlands Organization for Scientific Research (NWO) and FAPESP grant number 729.004.003, CAPES/NUFFIC 037/12-13910/13-2, FAPESP/BIOEN 2008/56147-1, FAPESP/BE-BASIC 2013/50365-5 and CNPq 471886/2012-2. Publication 6124 of the Netherlands Institute of Ecology (NIOO/KNAW).

Author contribution statement

J.R.S., J.B.C., H.C. and E.E.K. designed research; J.R.S. and K.S.L. conducted the experiment; J.R.S., N.A.C., A.M.K., A.P., H.J.L. and E.E.K. conducted the qPCR and the sequencing analyses; J.R.S. and N.A.C. performed the statistical analyses; J.R.S., N.A.C., H.J.L., H.C. and E.E.K. wrote the paper. All authors reviewed the manuscript.

Competing interest statement

The authors declare no conflict of interest.

Accession codes

European Nucleotide Archive study accession number PRJEB13027.

4.6 Supplementary Material

Starts on next page.

Table S1. Nitrous oxide fluxes, nitrogen cycling gene abundances and total bacterial and total archaea abundance in a Red Latosol soil as affected by treatments with urea with or without nitrification inhibitors (DCD and DMPP); polymer sulfur coated urea (PSCU) or calcium nitrate applied to sugarcane. The R represents the second application of the nitrification inhibitors.

Treatment	N_2O-N	AOA *amoA*	AOB *amoA*	*nirK*	*nirS*	*nosZ*	Total archaea	Toral bacteria
7 DAF†	$\mu g\ m^{-2}\ h^{-1}$	10^6 copies g^{-1}	10^5 copies g^{-1}	10^8 copies g^{-1}	10^5 copies g^{-1}	10^6 copies g^{-1}	10^7 copies g^{-1}	10^9 copies g^{-1}
No N	3.8 c	7.5 a	2.3 b	7.8 ns	4.2 ns	15.9 ns	9.7 ns	2.6 abc
UR	82.0 a	2.0 b	12.2 a	6.8	4.6	21.3	4.5	5.9 a
UR+DCD - R‡	12.9 c	3.5 b	2.8 b	7.1	3.9	13.9	5.1	4.9 abc
UR+DMPP - R	5.8 c	3.2 b	3.0 b	8.5	4.8	18.7	8.6	5.6 ab
PSCU	55.5 b	1.9 b	2.6 b	4.0	1.4	6.7	2.9	1.3 bc
UR+DCD	8.9 c	3.5 b	0.6 b	2.0	0.8	5.3	2.3	1.2 c
UR+DMPP	8.3 c	2.5 b	1.4 b	3.7	2.7	10.2	3.1	2.0 abc
Calcium Nitrate	12.8 c	1.6 b	0.5 b	1.1	1.7	3.7	1.5	0.8 c
16 DAF								
No N	18.7 c	41.1 a	1.5 b	2.2 ns	5.3 ns	0.7 b	24.5 a	1.2 ns
UR	1781.3 a	8.7 b	8.7 a	1.3	7.4	3.4 a	8.3 b	1.1
UR+DCD - R	238.4 c	13.4 b	2.2 b	0.9	5.2	2.9 ab	9.7 b	0.8
UR+DMPP - R	47.7 c	12.9 b	1.2 b	2.2	6.6	2.1 ab	16.3 ab	1.1
PSCU	609.8 b	9.8 b	5.1 ab	0.9	7.4	2.8 ab	8.7 b	1.3
UR+DCD	62.2 c	11.7 b	0.9 b	0.4	3.6	2.5 ab	6.3 b	0.7
UR+DMPP	39.3 c	10.0 b	0.7 b	0.8	4.7	2.8 ab	5.7 b	0.9
Calcium Nitrate	36.9 c	11.9 b	0.4 b	0.4	1.8	1.3 ab	6.4 b	0.5
18 DAF								
No N	15.3 c	23.6 a	2.8 b	14.5 ns	35.2 ns	0.7 b	18.0 a	1.0 ns
UR	1187.0 a	8.5 b	10.7 ab	7.9	14.2	1.6 a	3.2 b	0.7
UR+DCD - R	110.6 c	4.3 b	16.7 a	10.0	20.0	1.0 ab	5.1 b	1.1
UR+DMPP - R	27.6 c	6.4 b	1.7 b	9.3	27.7	1.2 ab	5.2 b	1.0
PSCU	558.0 b	2.2 b	6.6 ab	2.8	6.0	0.8 b	2.0 b	0.3
UR+DCD	53.8 c	7.4 b	4.4 b	5.8	9.6	1.1 ab	7.3 b	0.9
UR+DMPP	20.9 c	3.5 b	1.9 b	5.9	6.4	1.0 ab	4.7 b	0.5
Calcium Nitrate	27.4 c	9.0 b	1.1 b	4.1	10.7	0.5 b	9.9 ab	0.5

Table S1 cont'd.

Treatment	N_2O-N	AOA *amoA*	AOB *amoA*	*nirK*	*nirS*	*nosZ*	Total archaea	Total bacteria
27 DAF	µg m^{-2} h^{-1}	10^6 copies g^{-1}	10^5 copies g^{-1}	10^8 copies g^{-1}	10^5 copies g^{-1}	10^6 copies g^{-1}	10^7 copies g^{-1}	10^9 copies g^{-1}
No N	8.1 c	27.8 a	6.8 bc	15.4 a	5.3 a	1.8 abc	14.4 a	1.3 ns
UR	1356.3 a	2.1 bc	53.8 a	7.3 ab	4.1 ab	2.2 a	2.0 c	0.9
UR+DCD - R	163.8 c	6.0 bc	31.3 ab	7.3 ab	4.9 ab	1.7 abc	4.1 bc	0.7
UR+DMPP - R	10.7 c	8.3 b	3.7 c	11.1 ab	6.5 a	2.1 ab	8.3 ab	1.4
PSCU	704.8 b	1.5 c	17.9 bc	3.9 b	2.1 b	1.3 bc	2.4 bc	0.8
UR+DCD	77.3 c	5.0 bc	10.2 bc	5.2 ab	2.2 b	1.4 abc	3.2 bc	0.8
UR+DMPP	39.3 c	6.0 bc	4.4 bc	7.1 ab	2.6 b	1.7 abc	3.1 bc	0.9
Calcium Nitrate	36.9 c	6.8 bc	1.8 c	3.6 b	0.8 b	0.9 c	3.0 bc	0.4
35 DAF								
No N	8.4 b	12.4 a	1.2 b	5.5 ns	1.2 ns	2.1 ns	3.7 ns	0.3 ns
UR	1137.0 a	4.0 ab	11.8 a	6.2	1.3	2.0	2.4	0.5
UR+DCD - R	83.9 b	4.4 ab	5.7 b	2.6	0.9	2.2	2.0	0.3
UR+DMPP - R	42.4 b	3.8 ab	1.6 b	5.6	1.3	1.3	3,0	0.5
PSCU	1167.5 a	1.4 b	2.8 b	2.5	0.3	1.2	1,0	0.2
UR+DCD	264.5 b	4.8 ab	3.4 b	3.0	0.6	1.9	1,7	0.3
UR+DMPP	105.4 b	9.1 ab	1.6 b	4.3	1.4	3.1	2,1	0.4
Calcium Nitrate	28.7 b	9.4 ab	1.0 b	4.7	0.7	2.0	2,8	0.3
42 DAF								
No N	4.0 b	28.1 a	1.6 ns	13.5 ns	3.4 ns	2.5 ns	7,9 a	0.7 ns
UR	225.3 ab	1.4 b	8.5	4.5	2.0	2.4	0,8 b	0.4
UR+DCD - R	10.8 b	3.3 b	8.0	10.2	2.8	1.7	3,6 ab	0.6
UR+DMPP - R	22.1 b	5.7 b	3.0	11.9	4.2	2.1	3,5 ab	0.7
PSCU	312.9 a	1.8 b	6.5	4.0	1.6	1.7	1,0 b	0.3
UR+DCD	15.3 b	6.1 b	9.7	6.0	1.4	1.7	3,3 ab	0.5
UR+DMPP	10.7 b	5.5 b	1.6	7.7	2.4	1.6	3,3 ab	0.6
Calcium Nitrate	6.5 b	7.0 b	1.2	5.0	1.0	1.1	2,7 ab	0.4

Table S1 cont'd.

Treatment	N_2O-N	AOA *amoA*	AOB *amoA*	*nirK*	*nirS*	*nosZ*	Total archaea	Total bacteria
82 DAF	µg m^{-2} h^{-1}	10^6 copies g^{-1}	10^5 copies g^{-1}	10^8 copies g^{-1}	10^5 copies g^{-1}	10^6 copies g^{-1}	10^7 copies g^{-1}	10^9 copies g^{-1}
No N	0.2 b	40.7 a	1.7 b	22.1 ns	4.6 ab	0.6 c	13.8 ns	1.3 ns
UR	1.6 b	1.0 b	10.7 a	2.9	1.1 b	1.7 abc	0.8	0.2
UR+DCD - R	0.1 b	1.9 b	2.8 b	13.4	9.2 ab	1.1 bc	6.2	0.8
UR+DMPP - R	0.2 b	6.2 b	0.8 b	3.0	3.1 ab	0.8 bc	4.0	0.1
PSCU	8.8 a	1.4 b	4.8 ab	1.9	1.4 b	0.7 c	1.0	0.1
UR+DCD	0.2 b	22.4 ab	3.7 b	19.5	8.8 ab	2.2 ab	9.8	2.2
UR+DMPP	0.1 b	23.2 ab	2.4 b	29.6	13.0 ab	3.1 a	12.6	2.4
Calcium Nitrate	0.1 b	51.3 a	2.4 b	19.7	17.4 a	2.0 abc	20.8	2.6
158 DAF								
No N	0.1 b	35.3 a	2.9 ab	23.9 ns	14.7 ns	1.0 ns	9.2 a	1.1 ns
UR	0.3 b	2.8 b	4.2 ab	6.3	5.2	1.0	2.4 ab	0.5
UR+DCD - R	0.2 b	10.1 b	5.7 a	13.8	16.5	0.9	6.2 ab	2.0
UR+DMPP - R	0.1 b	5.2 b	1.3 b	7.0	5.1	0.8	3.7 ab	0.5
PSCU	4.6 a	7.7 b	4.1 ab	6.6	4.0	0.8	2.4 ab	0.7
UR+DCD	0.1 b	6.5 b	1.7 b	5.5	4.2	0.7	0.5 b	0.5
UR+DMPP	0.2 b	10.0 b	0.9 b	5.5	4.7	0.9	0.3 b	0.4
Calcium Nitrate	0.0 b	11.9 b	1.3 b	13.1	4.4	0.7	0.5 b	0.6

Means followed by same letter in column per DAF did not differ by Tukey test 5%. ns: no significant. †Days after fertilizer application. ‡ R means reapplication of inhibitors in same plot.

Table S2. Linear regression and coefficients (R^2) relating daily N_2O flux to environmental variables from Red Latosol soil after application of urea with or without nitrification inhibitors (DCD and DMPP), polymer sulfur coated urea (PSCU) and calcium nitrate applied to sugarcane. (n = 256).

	All treatments		Without nitrification inhibitors	
Variable (x)†	Regression*	R^2	Regression	R^2
NH_4^+-N	ns	-	y = 1.87 + 0.00073x	0.0879
NO_3^--N	y = 1.64 + 0.0016x	0.0389	ns	-
WFPS	y = 0.74 + 0.031x	0.1021	y = 0.82 + 0.034x	0.0864
Temp. air	ns	-	ns	-
Temp. soil	ns	-	ns	-
pH	y = 1.70 + 2.4e4x	0.0214	ns	-
Rain day	y = 1.58 + 0.013x	0.0818	y = 1.77 + 0.013x	0.0609
Rain week	y = 1.40 + 0.0081x	0.1589	y = 1.60 + 0.0078x	0.1017
CO_2-C	ns	-	ns	-
CH_4-C	y = 1.78 + 0.0073x	0.0243	y = 2.02 + 0.011x	0.0422
AOA *amoA*	y = 1.88 - 2.83e-8x	0.0990	y = 2.18 - 3.83e-8x	0.2107
AOB *amoA*	y = 1.56 + 2.94e-7x	0.1786	y = 1.71 + 3.02e-7x	0.1944
nirK	y = 1.82 - 1.32e-11x	0.0405	y = 2.03 - 1.55e-11x	0.0485
nirS	ns	-	ns	-
nosZ	ns	-	ns	-
Total bacteria	y = 1.80 + 7.83e-11x	0.0297	ns	-
Total archaea	y = 1.84 - 2.28e-9x	0.0532	y = 2.10 - 3.24e-9x	0.1076

* $p \leq 0.05$. ns: no significant. † N_2O-N: μg m^{-2} h^{-1} transformed in log(X+10); CO_2-C: mg m^{-2} h^{-1}; CH_4-C: μg m^{-2} h^{-1}; NH_4^+-N and NO_3^--N: mg kg^{-1} 0-10 cm soil; Rainweek and Rainday: mm, Temp. air and soil: ºC; WFPS: Water-filled pore space, %; AOB *amoA*, AOA *amoA*, *nir*K, *nir*S, *nos*Z, total bacteria and total archaea: copies gene g^{-1} dry soil; pH-$CaCl_2$: transformed in H^+, 10^{-pH}.

Table S3. Multiple linear regression parameters (β) and coefficients (R^2) relating daily N_2O flux to environmental variables after fertilizer application to sugarcane (n = 256).

Model*	Parameters	R^2
All treatments†		
$N_2O\text{-}N = \beta0 + \beta AOB + \beta Rainweek + \beta pH + \beta NH_4^+\text{-}N + \beta Bacteria + \beta CO_2 + \beta NO_3^-\text{-}N$ ‡	$0.84 + 2.96e\text{-}7AOB + 0.0097Rainweek + 3.3e4pH + 0.0050NH_4^+\text{-}N - 0.9e\text{-}10Bacteria + 0.0019NO_3^-\text{-}N + 0.00093CO_2$	0.4741
Without nitrification inhibitors		
$N_2O\text{-}N = \beta0 + \beta AOB + \beta Rainweek + \beta NH_4^+\text{-}N + \beta Bacteria + \beta pH + \beta CO_2$	$0.96 + 2.93e\text{-}7AOB + 0.0096Rainweek + 0.0012NH_4^+\text{-}N - 1.8e\text{-}10Bacteria + 3.0e4pH + 0.0023CO_2$	0.5267

* Stepwise selection, $p \leq 0.05$. † Treatments: No N – control, Urea (UR), UR+DCD, UR+DMPP, Polymer Sulfur Coated Urea and calcium nitrate. ‡ N_2O-N: μg m^{-2} h^{-1} transformed in log(X+10); AOB and Bacteria: *amoA* from ammonia oxidizing bacteria and all bacteria, copies gene g^{-1} dry soil; NH_4^+-N and NO_3^--N: mg kg^{-1} 0-10 cm soil; Rainweek: mm; pH-$CaCl_2$: transformed in H^+,

Table S4. Top-nine bacterial phyla presented in the microbial communities in the Red Latosol soil under treatments with urea with or without nitrification inhibitors (DCD and DMPP); polymer sulfur coated urea (PSCU) or calcium nitrate applied to sugarcane.

Treatment	Mean relative abundance (%) of Bacterial Phyla, or effect size(corrected p-value)								
	Proteo	Firm	Actino	Acido	Bacter	Verruco	Plancto	Gemmati	Nitro
7 DAF†	ns	0.73 (0.07)	ns	ns	ns	ns	ns	ns	ns
No N	24	4	20	21	3	5	2	1	1
UR	27	24	23	5	4	2	1	1	0
UR+DCD - R‡	26	26	14	12	4	3	2	1	1
UR+DMPP – R	25	28	12	12	7	2	2	1	0
PSCU	44	6	24	7	3	2	2	1	0
UR+DCD	34	20	13	8	5	3	1	1	0
UR+DMPP	23	32	16	7	4	2	1	1	0
Calcium Nitrate	30	9	22	13	4	4	2	1	1
16 DAF†	ns	ns	ns	ns	0.65(0.09)	ns	ns	ns	0.63(0.09)
No N	28	10	31	11	3	3	1	1	1
UR	36	15	23	4	8	1	1	1	0
UR+DCD - R‡	30	12	19	12	6	4	3	1	0
UR+DMPP – R	36	13	20	11	4	2	2	1	0
PSCU	35	18	23	5	7	1	1	1	0
UR+DCD	32	15	22	7	7	3	2	1	0
UR+DMPP	30	11	21	10	4	5	3	1	0
Calcium Nitrate	22	21	24	12	2	3	2	1	0
18 DAF†	ns	ns	ns	ns	ns	ns	ns	ns	ns
No N	25	9	24	15	3	5	2	1	1
UR	34	14	24	6	6	2	2	1	0
UR+DCD - R‡	41	13	17	2	12	1	3	1	0
UR+DMPP – R	43	10	17	5	10	2	1	2	0
PSCU	38	13	23	6	6	1	1	1	0
UR+DCD	31	20	17	5	11	2	2	1	0
UR+DMPP	34	17	24	5	5	2	1	1	0
Calcium Nitrate	35	18	23	6	5	2	1	1	0
27 DAF†	ns	ns	ns	ns	ns	ns	ns	ns	ns
No N	42	3	23	8	6	2	0	2	1
UR	49	7	13	3	17	1	1	2	0
UR+DCD - R‡	49	5	16	5	7	3	1	2	0
UR+DMPP – R	48	4	12	6	13	3	1	2	0
PSCU	49	6	20	3	10	1	1	2	0
UR+DCD	48	5	16	3	14	2	1	2	0
UR+DMPP	53	6	18	3	9	1	0	1	0
Calcium Nitrate	45	7	19	6	6	2	1	2	0

Significance from Tukey-Kramer post-hoc test based on Benjamini-Hochberg corrected p-values above 0.1 from STAMP analysis. ns: not significant. †Days after fertilizer application. ‡ R means reapplication of inhibitors in same plot. Proteo=Proteobacteria, Firm=Firmicutes, Actino=Actinobacteria, Bacter=Bacteroidetes, Verruco=Verrucomicrobia, Planct=Planctomycetes, Gemmati=Gemmatimonadetes, Nitro=Nitrospira.

Table S4 cont'd.

Treatment	Mean relative abundance (%) of Bacterial Phyla, or effect size(corrected p-value)								
	Proteo	Firm	Actino	Acido	Bacter	Verruco	Planct	Gemmati	Nitro
35 DAF†	0.71(0.02)	ns	ns	ns	ns	0.67(0.04)	ns	ns	ns
No N	32	2	15	18	6	5	1	2	1
UR	48	3	15	5	12	1	1	2	0
UR+DCD - R‡	46	5	14	8	12	2	1	2	0
UR+DMPP – R	45	6	14	7	12	3	1	2	0
PSCU	57	3	12	5	13	1	1	1	0
UR+DCD	42	5	13	8	12	2	1	3	0
UR+DMPP	43	3	11	8	18	2	1	2	0
Calcium Nitrate	42	10	20	5	8	2	1	3	0
42 DAF†	0.60(0.07)	ns	ns	0.60(0.09)	0.64(0.08)	ns	ns	ns	0.74(0.02)
No N	31	2	19	17	8	6	1	1	1
UR	39	5	17	7	16	3	1	2	0
UR+DCD - R‡	43	4	15	6	15	3	1	2	1
UR+DMPP – R	37	6	19	10	11	2	2	2	0
PSCU	53	2	17	4	11	1	1	1	0
UR+DCD	40	4	18	7	13	5	1	2	0
UR+DMPP	38	4	18	11	9	4	2	2	0
Calcium Nitrate	37	4	19	8	13	4	1	2	0
82 DAF†	ns	0.80(0.003)	ns	ns	0.65(0.03)	ns	ns	0.72(0.01)	0.77(0.004)
No N	37	4	13	14	9	5	1	2	1
UR	47	3	15	5	16	1	1	3	0
UR+DCD - R‡	37	22	22	2	9	1	0	1	0
UR+DMPP – R	48	15	18	5	7	1	0	1	0
PSCU	50	4	20	3	14	0	0	1	0
UR+DCD	45	3	16	8	11	3	1	3	0
UR+DMPP	42	6	18	9	10	2	1	2	1
Calcium Nitrate	30	20	30	4	4	2	1	1	0
158DAF†	ns	ns	ns	ns	ns	ns	ns	ns	0.70(0.06)
No N	39	3	21	9	11	3	1	2	0
UR	47	3	15	6	12	3	1	3	0
UR+DCD - R‡	42	6	22	5	11	1	0	3	0
UR+DMPP – R	43	5	23	7	9	2	1	2	0
PSCU	47	3	20	5	9	2	1	2	0
UR+DCD	43	5	18	8	10	3	1	2	0
UR+DMPP	46	3	17	8	10	2	1	3	0
Calcium Nitrate	40	5	24	5	9	2	1	3	0

Significance from Tukey-Kramer post-hoc test based on Benjamini-Hochberg corrected p-values above 0.1 from STAMP analysis. ns: not significant. †Days after fertilizer application. ‡ R means reapplication of inhibitors in same plot. Proteo=Proteobacteria, Firm=Firmicutes, Actino=Actinobacteria, Bacter=Bacteroidetes, Verruco=Verrucomicrobia, Planct=Planctomycetes, Gemmati=Gemmatimonadetes, Nitro=Nitrospira.

Table S5. Primers and thermocycler conditions used in gene abundance analysis by real time qPCR

Target gene	Primer	Primer Sequence	Size (bp)	Thermal profile	Reference	Source of Standard plasmid	-Vector
AOA *amoA*	Arch-amoAF	5'-STAATGGTCTG GCTTAGACG-3'	635	95°C-5 min.; 40x 95°C-30s, 55°C-45s, 72°C-45s, 82°C-15s	Francis et al.[54]	Environmental Archaea	pGEM®-T Systems Promega
	Arch-amoAR	5'-GCGGCCATCC ATCTGTATGT-3'					
AOB *amoA*	amoA1F	5'-GGGGTTTCT ACTGGTGGT-3'	491	95°C-5 min.; 40x 95°C-30s, 56°C-45s, 72°C-45s, 82°C-15s	Rotthauwe et al.[55]	*Nitrosomonas europaea*	pGEM®-T Systems Promega
	amoA2R	5'-CCCCTCKGSA AAGCCTTCTTC-3'					
*nos*Z	nosZ2F	5'-CGCRACGGCAA SAAGGTSMSSGT-3'	267	95°C-5 min.; 40x 95°C-15s, 60°C-15s, 72°C-30s, 82°C-15s	Henry et al.[56]	*Pseudomonas stutzeri* (M13R/F)	Dh5alpha p PCR4-topo
	nosZ2R	5'-CAKRTGCAKSG CRTGGCAGAA-3'					
*nir*K	NirK876	5'-ATYGGCGG VAYGGCGA-3'	165	95°C-5 min.; 40x 95°C-15s, 63°C-30s, 72°C-30s, 82°C-15s	Henry et al.[57]	*Paracoccus denitrificans* (DSM 413)	Dh5alpha p PCR4-topo
	NirK1040	5'-GCCTCGATCA GRTTRTGGTT-3'					
*nir*S	nirScd3aF	5'-GTSAACGTSA AGGARACSGG-3'	425	95°C-5 min.; 40x 95°C-10s, 60°C-10s, 72°C-20s, 86°C-5s	Throbäck et al.[58]	*Pseudomonas stutzeri* (M13R/F)	PCR produ
	nirSR3cd	5'-GASTTCGGRT GSGTCTTGA-3'					
Total bacteria	Eub338	5'-ACTCCTACGG GAGGCAGCAG-3'	200	95°C-5 min.; 40x 95°C-5s, 53°C-10s, 72°C-20s	Fierer et al.[59]	Firmicutes	Dh5alpha p PCR4-topo
	Eub518	5'-ATTACCGC GGCTGCTGG-3'					
Total archaea	Arch1017R	5'-AGGAATTGGC GGGGGAGCAC-3'	112	95°C-10 min.; 40x 95°C-10s, 60°C-10s, 72°C-20s	Klindworth et al.[60]	Environmental Archaea	Dh5alpha p PCR4-topo
	Arch915F	5'-GGCCATGCA CCWCCTCTC-3'					

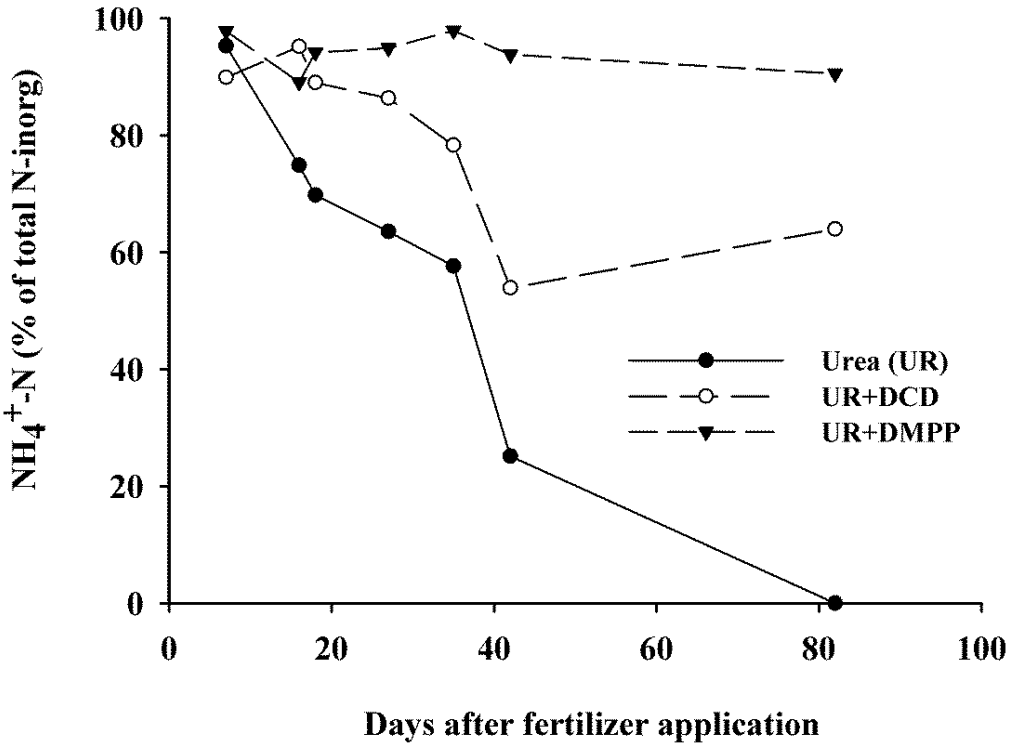


Figure S1. Percentage of NH4+ in relation to total inorganic N (NH_4+ + NO_3-) at 0 - 10 cm soil depth from application of urea with or without nitrification inhibitors (DCD and DMPP) to sugarcane.

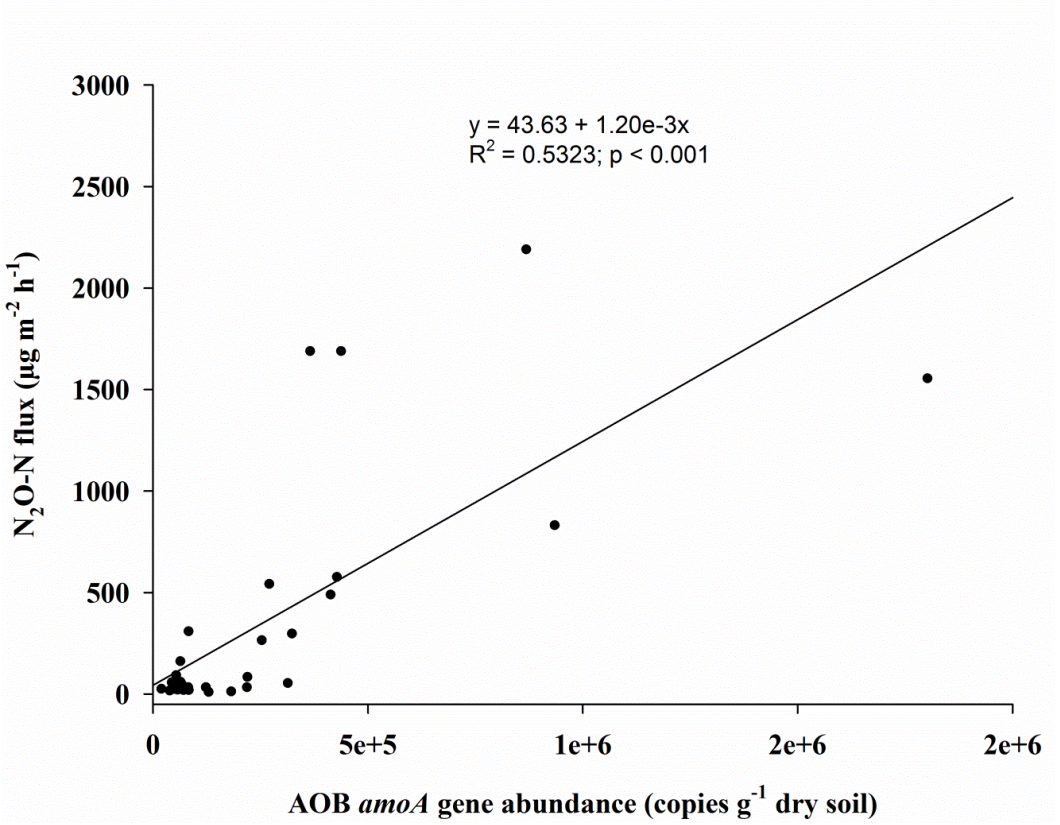


Figure S2. Correlation between ammonia oxidizing bacteria (AOB) *amoA* gene abundance and nitrous oxide emission from Red Latosol soil 16 days after application of urea with or without nitrification inhibitors (DCD and DMPP), polymer sulfur coated urea (PSCU) and calcium nitrate applied to sugarcane (n = 32).

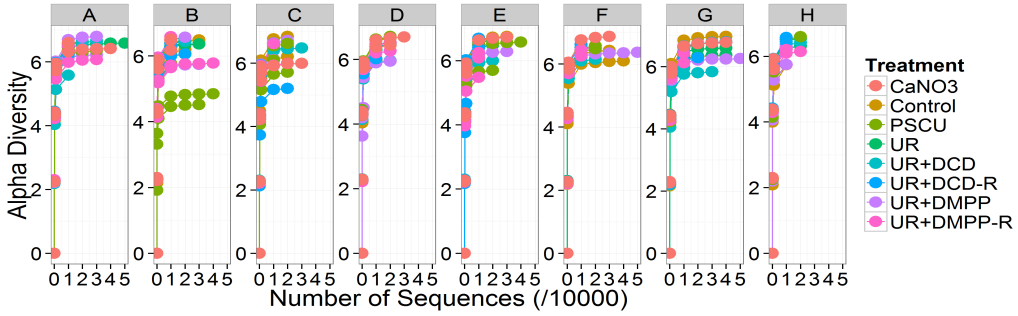


Figure S3. Rarefaction curves of the microbial communities present in the Red Latosol under treatments with urea, incorporated, with or without nitrification inhibitors (DCD and DMPP); polymer sulfur coated urea (PSCU) and calcium nitrate applied to sugarcane. Time points are separated such that A, B, C, D, E, F, G and H correspond to days 7, 16, 18, 27, 35, 42, 82 and 158 of the experiment.

Table S4. Shannon indices and Between-Class Analysis (BCA) ordinations of the bacterial communities under treatments of urea, incorporated in the Red Latosol, with or without nitrification inhibitors (DCD and DMPP); polymer sulfur coated urea (PSCU) or calcium nitrate applied to ratoon sugarcane. Continued on next four pages. DAF = days after fertilization.

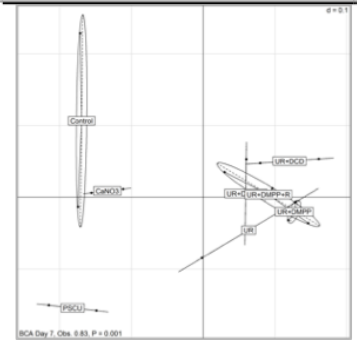
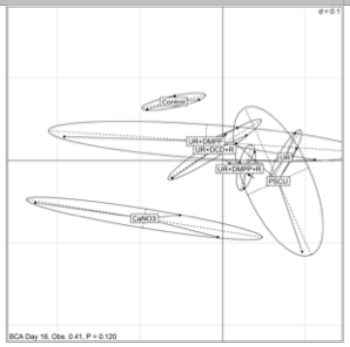
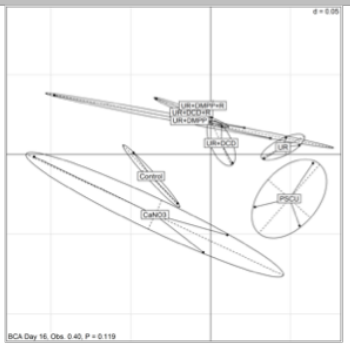
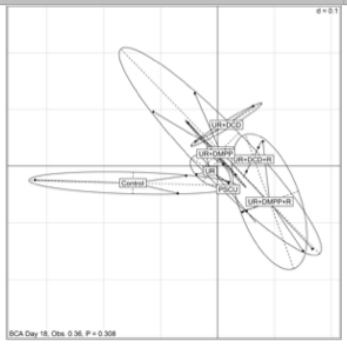
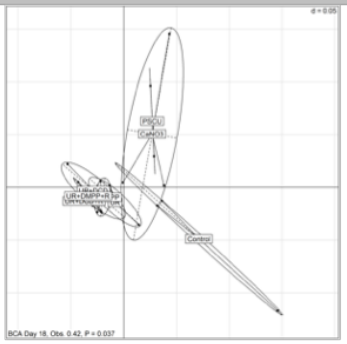
Treatment	16s rRNA gene diversity		BCA ordination based on 16S gene abundances within Phyla	BCA ordination based on 16S gene abundances within Genus
7 DAF†				
No N	6.2	ns		
UR	5.8			
UR+DCD - R	5.9			
UR+DMPP – R	5.6			
PSCU	5.9			
UR+DCD	5.7			
UR+DMPP	5.7			
Calcium Nitrate	6.3			
16 DAF				
No N	6.1	ns		
UR	6.0			
UR+DCD - R	6.2			
UR+DMPP – R	6.1			
PSCU	5.9			
UR+DCD	6.1			
UR+DMPP	6.0			
Calcium Nitrate	6.0			
18 DAF				
No N	6.0	ns		
UR	6.0			
UR+DCD - R	5.8			
UR+DMPP – R	5.9			
PSCU	5.8			
UR+DCD	5.7			
UR+DMPP	5.9			
Calcium Nitrate	5.7			

Table S4 con't.

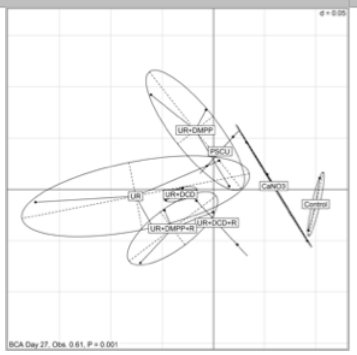
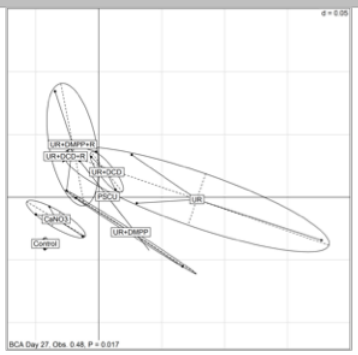
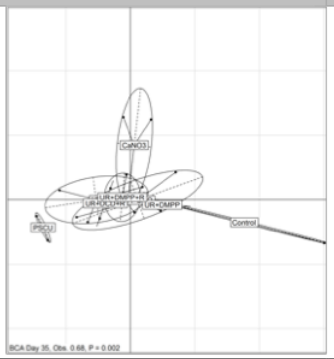
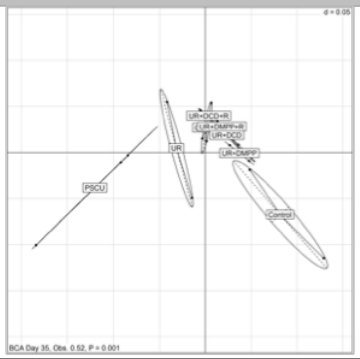
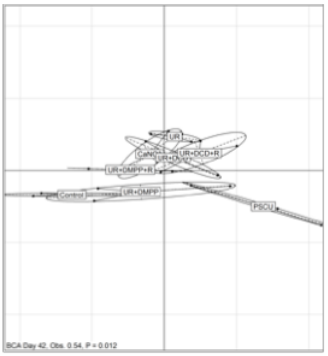
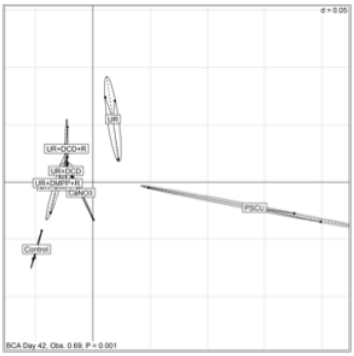
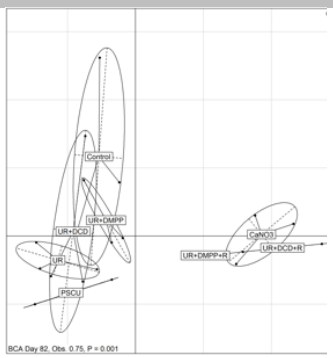
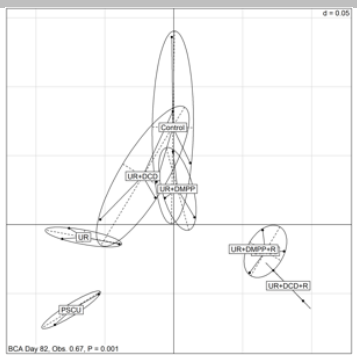
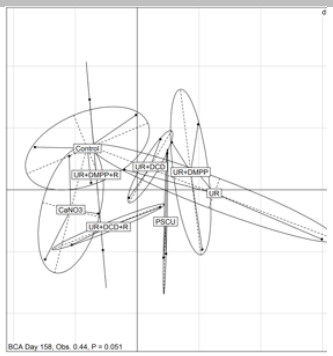
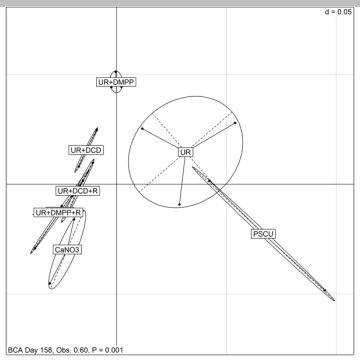
Treatment	16s rRNA gene diversity		BCA ordination based on 16S gene abundances within Phyla	BCA ordination based on 16S gene abundances within Genus
27 DAF				
No N	6.2	a		
UR	5.3	b		
UR+DCD - R	6.0	ab		
UR+DMPP – R	5.8	ab		
PSCU	5.8	ab		
UR+DCD	5.8	ab		
UR+DMPP	5.8	ab		
Calcium Nitrate		ab		
	5.9			
35 DAF				
No N	6.1	*ns*		
UR	5.7			
UR+DCD – R	6.0			
UR+DMPP – R	6.1			
PSCU	5.1			
UR+DCD	5.9			
UR+DMPP	5.5			
Calcium Nitrate	5.9			
42 DAF				
No N	6.2	ns		
UR	5.8			
UR+DCD - R	6.1			
UR+DMPP – R	6.2			
PSCU	5.2			
UR+DCD	6.0			
UR+DMPP	6.1			
Calcium Nitrate	6.2			

Table S4 con

Treatment	16S rRNA gene diversity		BCA ordination based on 16S gene abundances within Phyla	BCA ordination based on 16S gene abundances within Genus
85 DAF				
No N	6.0	ab		
UR	5.6	ab		
UR+DCD – R	5.8	ab		
UR+DMPP – R	6.1	ab		
PSCU	5.4	a		
UR+DCD	6.0	ab		
UR+DMPP	6.1	b		
Calcium Nitrate	6.0	ab		
158DAF				
No N	6.1	ns		
UR	6.0			
UR+DCD – R	6.2			
UR+DMPP – R	6.2			
PSCU	5.8			
UR+DCD	6.2			
UR+DMPP	6.1			
Calcium Nitrate	6.2			

4.7 References

1. Erisman, J.W., et al., *Reactive nitrogen in the environment and its effect on climate change.* Current Opinion in Environmental Sustainability, 2011. **3**(5): p. 281-290.
2. Egeskog, A., et al., *Greenhouse gas balances and land use changes associated with the planned expansion (to 2020) of the sugarcane ethanol industry in Sao Paulo, Brazil.* Biomass & Bioenergy, 2014. **63**: p. 280-290.
3. Crutzen, P.J., et al., N_2O *release from agro-biofuel production negates global warming reduction by replacing fossil fuels.* Atmospheric Chemistry and Physics, 2008. **8**(2): p. 389-395.
4. Braker, G. and R. Conrad, *Diversity, Structure, and Size of* N_2O*-Producing Microbial Communities in Soils-What Matters for Their Functioning?*, in *Advances in Applied Microbiology, Vol 75*, A.I. Laskin, S. Sariaslani, and G.M. Gadd, Editors. 2011. p. 33-70.
5. Hayatsu, M., K. Tago, and M. Saito, *Various players in the nitrogen cycle: Diversity and functions of the microorganisms involved in nitrification and denitrification.* Soil Science and Plant Nutrition, 2008. **54**(1): p. 33-45.

6. Muller, C., et al., *Quantification of* N_2O *emission pathways via a N-15 tracing model.* Soil Biology & Biochemistry, 2014. **72**: p. 44-54.
7. Liu, X.J.J., et al., *Dinitrogen and* N_2O *emissions in arable soils: Effect of tillage, N source and soil moisture.* Soil Biology & Biochemistry, 2007. **39**(9): p. 2362-2370.
8. Di, H.J., et al., *Effect of soil moisture status and a nitrification inhibitor, dicyandiamide, on ammonia oxidizer and denitrifier growth and nitrous oxide emissions in a grassland soil.* Soil Biology & Biochemistry, 2014. **73**: p. 59-68.
9. Companhia Nacional de Abastecimento (CONAB). Quarto levantamento de safra de cana-de-açúcar 2014/2015, p.I.P.A.o.h.w.c.g.b.O.
10. EMBRAPA, E.B.d.P.A., *Brazilian soil classification system. Empresa Brasileira de Pesquisa Agropecuária, Rio de Janeiro, Brazil.* . 2006.
11. Martins, M.R., et al., *Nitrous oxide and ammonia emissions from N fertilization of maize crop under no-till in a Cerrado soil.* Soil & Tillage Research, 2015. **151**: p. 75-81.
12. Filoso, S., et al., *Reassessing the environmental impacts of sugarcane ethanol production in Brazil to help meet sustainability goals.* Renewable and Sustainable Energy Reviews, 2015. **52**: p. 1847-1856.
13. Snyder, C.S., et al., *Review of greenhouse gas emissions from crop production systems and fertilizer management effects.* Agriculture Ecosystems & Environment, 2009. **133**(3-4): p. 247-266.
14. Snyder, C.S., et al., *Agriculture: sustainable crop and animal production to help mitigate nitrous oxide emissions.* Current Opinion in Environmental Sustainability, 2014. **9-10**: p. 46-54.
15. Soares, J.R., et al., *Enhanced-Efficiency Fertilizers in Nitrous Oxide Emissions from Urea Applied to Sugarcane.* Journal of Environmental Quality, 2015. **44**(2): p. 423-430.
16. Ruser, R. and R. Schulz, *The effect of nitrification inhibitors on the nitrous oxide (*N_2O*) release from agricultural soils-a review.* Journal of Plant Nutrition and Soil Science, 2015. **178**(2): p. 171-188.
17. O'Callaghan, M., et al., *Effect of the nitrification inhibitor dicyandiamide (DCD) on microbial communities in a pasture soil amended with bovine urine.* Soil Biology & Biochemistry, 2010. **42**(9): p. 1425-1436.
18. Yang, J.B., et al., *Influence of the nitrification inhibitor DMPP on the community composition of ammonia-oxidizing bacteria at microsites with increasing distance from the fertilizer zone.* Biology and Fertility of Soils, 2013. **49**(1): p. 23-30.
19. Zhang, L.M., et al., *Ammonia-oxidizing archaea have more important role than ammonia-oxidizing bacteria in ammonia oxidation of strongly acidic soils.* Isme Journal, 2012. **6**(5): p. 1032-1045.
20. do Carmo JB, et al., *Infield greenhouse gas emissions from sugarcane soils in Brazil: effects from synthetic and organic fertilizer application and crop trash accumulation.* Global Change Biology Bioenergy, 2013. **5**: p. 267-280.
21. Staff, S.S., *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys.* . 2nd edition. Natural Resources Conservation Service. U.S. Department of Agriculture Handbook 436., 1999.
22. Cantarella, H., *Nitrogênio. In: Fertilidade do Solo (eds Novais RF, Alvarez VVH, Barros NF, Fontes RLF, Cantarutti RB, Neves JCL). Sociedade Brasileira de Ciência do Solo, Viçosa, pp. 375-470* 2007.
23. Ghiberto, P.J., et al., *Nitrogen fertilizer leaching in an Oxisol cultivated with sugarcane.* Scientia Agricola, 2011. **68**(1): p. 86-93.
24. Vargas, V.P., et al., *Sugarcane Crop Residue Increases* N_2O *and CO2 Emissions Under High Soil Moisture Conditions.* Sugar Tech, 2014. **16**(2): p. 174-179.
25. Pitombo, L.M., et al., *Exploring soil microbial 16S rRNA sequence data to increase carbon yield and nitrogen efficiency of a bioenergy crop.* GCB Bioenergy, 2015: p. n/a-n/a.

26. Abalos, D., et al., *Management of irrigation frequency and nitrogen fertilization to mitigate GHG and NO emissions from drip-fertigated crops.* Science of the Total Environment, 2014. **490**: p. 880-888.
27. Tenuta, M. and E.G. Beauchamp, *Nitrous oxide production from granular nitrogen fertilizers applied to a silt loam.* Canadian Journal of Soil Science, 2003. **83**(5): p. 521-532.
28. Zanatta, J.A., et al., *Nitrous oxide and methane fluxes in south brazilian gleysol as affected by nitrogen fertilizers.* Revista Brasileira De Ciencia Do Solo, 2010. **34**(5): p. 1653-1665.
29. Nicol, G.W., et al., *The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria.* Environmental Microbiology, 2008. **10**(11): p. 2966-2978.
30. Venterea, R.T., et al., *Ammonium sorption and ammonia inhibition of nitrite-oxidizing bacteria explain contrasting soil* N_2O *production.* Scientific Reports, 2015. **5**.
31. Frame, C.H. and K.L. Casciotti, *Biogeochemical controls and isotopic signatures of nitrous oxide production by a marine ammonia-oxidizing bacterium.* Biogeosciences, 2010. **7**(9): p. 2695-2709.
32. Joo, H.S., M. Hirai, and M. Shoda, *Characteristics of ammonium removal by heterotrophic nitrification-aerobic denitrification by Alcaligenes faecalis no. 4.* Journal of Bioscience and Bioengineering, 2005. **100**(2): p. 184-191.
33. Zhao, B., et al., N_2O *and N-2 production during heterotrophic nitrification by Alcaligenes faecalis strain NR.* Bioresource Technology, 2012. **116**: p. 379-385.
34. Spott, O., R. Russow, and C.F. Stange, *Formation of hybrid* N_2O *and hybrid N-2 due to codenitrification: First review of a barely considered process of microbially mediated N-nitrosation.* Soil Biology & Biochemistry, 2011. **43**(10): p. 1995-2011.
35. Soares, J.R., H. Cantarella, and M.C.L. Menegale, *Ammonia volatilization losses of surface-applied urea with urease and nitrification inhibitors.* Soil Biology & Biochemistry, 2012. **52**: p. 82-89.
36. Cantarella, H. and P.C.O. Trivelin, *Determinação de nitrogênio inorgânico em solo pelo método da destilação a vapor. In: Análise química para avaliação da fertilidade de solos tropicais (eds Raij B van, Andrade JC, Cantarella H, Quaggio JA), Instituto Agronômico, Campinas, pp. 271–276.* 2001.
37. Koster, J. and S. Rahmann, *Snakemake-a scalable bioinformatics workflow engine.* Bioinformatics, 2012. **28**(19): p. 2520-2522.
38. Dodt, M., et al., *FLEXBAR—Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms.* Biology, 2012. **1**(3): p. 895.
39. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.
40. Edgar, R.C., et al., *UCHIME improves sensitivity and speed of chimera detection.* Bioinformatics, 2011. **27**(16): p. 2194-2200.
41. McDonald, D., et al., *The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome.* GigaScience, 2012. **1**: p. 7.
42. Cole, J.R., et al., *Ribosomal Database Project: data and tools for high throughput rRNA analysis.* Nucleic Acids Research, 2014. **42**(D1): p. D633-D642.
43. Rawlings, J.O., S.G. Pantula, and D.A. Dickey, *Applied Regression Analysis: A Research Tool. Second Edition, pp.656, Springer-Verlag New York, Inc.* 1998.
44. McMurdie, P.J. and S. Holmes, *phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data.* Plos One, 2013. **8**(4).

Chapter 5

Nitrification inhibitors effectively target N_2O-producing Nitrosospira spp. in tropical soil

Noriko A. Cassman, Johnny R. Soares, Agata Pijl, Késia S. Lourenço, Johannes A. van Veen, Heitor Cantarella and Eiko E. Kuramae

A version of this chapter was accepted as:

Cassman NA, Soares JR, Pijl A, Lourenco KS, van Veen JA, Cantarella H and Kuramae EE, February 2019. "Nitrification inhibitors effectively target N_2O-producing Nitrosospira spp. in tropical soil." *Environmental Microbiology*.

Abstract

The nitrification inhibitors (NIs) 3,4-dimethylpyrazole (DMPP) and dicyandiamide (DCD) effectively reduce N_2O emissions; however, which species are targeted and the effect on the nitrifying community is still unclear. Here we characterized the ammonia oxidizing bacteria (AOB) species linked to N_2O emissions and evaluated the effects of urea and urea with DCD and DMPP on the nitrifying community in a 258-day field experiment under sugarcane. Using an *amoA* AOB amplicon sequencing approach and mining a previous dataset of *16S rRNA* sequences, we characterized the most likely N_2O-producing AOB as a *Nitrosospira* spp. and identified *Nitrosospira* (AOB), *Nitrososphaera* (archaeal ammonia oxidizer) and *Nitrospira* (nitrite-oxidizer) as the main nitrifiers. The fertilizer treatments had no effect on the alpha and beta diversities of the AOB communities. Interestingly, we found three clusters of co-varying variables with nitrifier OTUs: the N_2O-producing AOB *Nitrosospira* with N_2O, NO_3^-, NH_4^+, WFPS and pH; AOA *Nitrososphaera* with NO_3^-, NH_4^+ and pH; and AOA *Nitrososphaera* and NOB *Nitrospira* with NH_4^+. These results support the co-occurrence of non-N_2O-producing *Nitrososphaera* and *Nitrospira* in the unfertilized soils and the promotion of N_2O-producing *Nitrosospira* under urea fertilization. Further, we suggest that DMPP is a more effective NI than DCD in tropical soil under sugarcane.

5.1 Introduction

Anthropogenic inputs of N fertilizers to agriculture have stimulated agricultural soils to contribute up to 59% of anthropogenic N_2O emissions [1-4]. Because N_2O has a global warming potential 298 times that of CO_2 [5] and diverts N that would otherwise be used by the crop, reducing N_2O emissions is a major target for sustainable management practices [6]. The N_2O emitted from a soil is the cumulative result of abiotic and biotic N_2O-generating pathways [7, 8]. The two main biotic processes contributing to N_2O in agricultural soils are nitrification (oxidation of NH_4^+ to NO_2^- to NO_3^-) and denitrification (anaerobic reduction of NO_3^- to NO_2^- to N_2O to N_2; reviewed in [9, 10]. Nitrification is carried out by a few bacterial and archaeal genera: ammonia oxidation is mediated by the ammonia-oxidizing archaea (AOA), such as the Thaumarchaeota *Nitrososphaera*, and the ammonia-oxidizing bacteria (AOB), such as the Betaproteobacteria *Nitrosomonas* and *Nitrosospira*; nitrite oxidation is carried out by nitrite oxidizing bacteria (NOB), including the Nitrospirae *Nitrospira* and the Alphaproteobacteria *Nitrobacter*. Denitrification is carried out by microorganisms widely dispersed over the bacterial, archaeal and fungal domains, and denitrification genes may also be carried by nitrifiers in what is termed nitrifier denitrification. Further, the process of complete nitrification by the recently discovered comammox bacteria, which have so far been found in the NOB *Nitrospira* genus, might also contribute to N_2O emissions [11].

Nitrification and denitrification processes are regulated by the abiotic factors temperature, oxygen availability, moisture, ammonia and nitrate availability, carbon availability and pH [12, 13]. These factors also affect the distribution and niche differentiation of nitrifiers; for example, the AOB numerically dominate in neutral soils with high NH_4^+ concentrations while the AOA dominate in acidic soils with low NH_4^+ concentrations [14, 15]. However, there are also exceptions to this general rule, for example the Gammaproteobacteria AOB Candidatus *Nitrosoglobus* recently isolated from acidic soils with survival in conditions down to pH 2 [16]. Within the domains there are also niche specializations, as the AOB *Nitrosomonas* is generally isolated from neutral pH soils while the AOB *Nitrosospira* is found in acid soils [17, 18]. Further, the nitrite oxidizer bacteria *Nitrobacter* and *Nitrospira* have optimal growth under higher and lower nitrite supplies, respectively, which is linked to their ecological niches [19, 20].

Nitrification is doubly implicated in N_2O production, either directly or indirectly by producing NO_3^- as the basis for denitrification, and has been shown to be the main process involved in N_2O emissions in sugarcane soils [21-27]. Thus, the addition of nitrification inhibitors with nitrogen fertilizers is currently being ex-

plored as a sustainable management practice in Brazilian sugarcane [24, 28, 29]. Nitrification inhibitors include dicyandiamide (DCD) and 3,4-dimethylpyrazole phosphate (DMPP), which are thought to be Cu-chelating agents acting on the ammonia monooxygenase enzyme [30]. The inhibitors have been shown to effectively reduce N_2O emissions by 40-95% in temperate and tropical soils [24, 31, 32]. This effect is generally restricted to the ammonia oxidizing bacteria, not affecting ammonia oxidizing archaea or the rest of the microbial community at coarse-grained levels [24, 33]. Evidence for the interdependence of ammonia and nitrite oxidizers as determined in unfertilized grassland soil suggests that the nitrifying community may be negatively affected under nitrogen fertilization with nitrification inhibitors [34]. It is yet unknown how the nitrification inhibitors DCD and DMPP might affect the nitrifying community in tropical soil under sugarcane.

Here our objectives were to identify the AOB species linked to N_2O emissions in a previous experiment and to compare the effects of urea fertilization with or without nitrification inhibitors on nitrifier abundances, with a focus on the ammonia-oxidizing bacterial community [24]. We sequenced *amoA* AOB amplicons from a 258-day field experiment encompassing treatments with urea and two nitrification inhibitors, DCD and DMPP, on soils growing ratoon sugarcane. We combined the *amoA* dataset with the 16S rRNA gene, nitrification and denitrification gene copy numbers and soil environmental variable datasets previously generated to test our hypotheses [24]. We hypothesized that the nitrification inhibitors would decrease the *amoA* AOB community diversity. Further, we hypothesized that the two nitrification inhibitors would have similar effects on the abundances of nitrifiers, including ammonia oxidizers (AOB and AOA) and nitrite oxidizers (NOB). To our knowledge, no studies to date have investigated the effect of nitrification inhibitors in urea fertilized soils on the nitrifier community growing in tropical soil.

5.2 Materials and Methods

5.2.1 Experimental design and sampling summary

A field experiment on Typic Hapludox soil (also known as Red Latosol) was set up at the Agronomic Institute in Campinas, Brazil at 22°52′15″ S, 47°04′57″ W, as described previously [24, 35]. Briefly, four treatments containing four replicate plots each were established in the 2013/2014 season on a third cycle of ratoon sugarcane (cultivar SP791011). The treatments were 1) no N fertilizer (control), 2) urea (UR), 3) urea with dicyandiamide (UR+DCD), 4) urea with 3,4-dimethylpyrazone phosphate (UR+DMPP). Urea was incorporated into the first 5

cm of soil and applied at a rate of 120 kg N ha^{-1}. The DCD (Sigma Aldritch) was added at 5% DCD-N per unit N from urea (v/v) while powdered DMPP (Sigma Aldritch) was added as 1% DMPP (w/w). Gas emission rates of CO_2, CH_4 and N_2O were measured daily to monthly using static chambers. Soil samples were taken of the top 10 cm of soil such that three subsamples were combined per plot. The soil samples were collected at eight time points: 7, 16, 18, 27, 35, 42, 82 and 158 days following fertilizer application and stored at -80 °C. Total DNA was extracted from the soil samples using a Power Soil kit from Mobio without modifications (Carlsbad, CA, USA). Further, pH, NO_3-N and NH_4-N were measured from the soil samples and water-filled pore space (WFPS) and temperature was previously calculated [24].

5.2.2 *amoA* AOB amplification and sequencing

Amplification of the partial *amoA* bacterial gene (491 bp) was performed using a two-step barcoding approach. The first PCR from the total DNA samples was carried out using forward primer H-AmoA1F-mod (5'-GCTATGCGC-GAGCTGCGGGGHTTYTACTGGTGGT-3') and reverse primer H-amoA2R (5'-GCTATGCGCGAGCTGCCCCCTCKGSAAAGCCTTCTTC -3') [36, 37]. In the second PCR, the *amoA* amplification products were amplified with primers that consisted of a 16 bp head sequence and included at the 5' end a library-specific 8 bp barcode [38]. Each PCR reaction (20 µl in first step, 50 µl in second step) consisted of 0.025 units of FastStart Taq DNA Polymerase (Roche), 1x reaction buffer with $MgCl_2$ (Roche), 0.5 mM dNTPmix (Fermentas), 0.125 µM of the forward and reverse primers, 0.1 mg/ml bovine serum albumin and 1 µl of DNA template. Thermocycler (C1000 Touch Thermal cycler, Biorad) conditions were as follows: 1) 5 minutes at 95 °C; 35 times 30 seconds at 95 °C, 30 seconds at 53 °C, 30 seconds at 72 °C; and 7 minutes at 72 °C and 2) 5 minutes at 95 °C; 10 times 30 seconds at 95 °C, 30 seconds at 53 °C, 1 minute at 72 °C; and 10 minutes at 72 °C. The first PCR reaction was performed in duplicate, screened by gel electrophoresis and pooled for use as a template in the second step, which used one primer (5'-BARCODE-HEAD-3'). Second step PCR products were checked by agarose gel electrophoresis and the concentration and quality determined using a fragment analyzer (Advanced Analytical). The bar-coded PCR products from all samples were normalized in equimolar amounts before sequencing. The *amoA* amplicon pool was sequenced using MiSeq V3 (2x300bp) technology (LGC, Germany). To complement the analysis of the *amoA* amplicon sequences, we mined the previously published dataset of 16S rRNA partial gene amplicons [24]. The *amoA* AOB

amplicons were thus, obtained from the same total DNA samples as the *16S rRNA* amplicons.

5.2.3 *amoA* AOB amplicon sequence processing

Bioinformatics steps were performed on a multi core server with 64 threads running Linux Ubuntu 16.04. Processing was accomplished through a Snakemake pipeline and bash and perl scripts. The *amoA* AOB sequences were clipped of primers and barcodes using bbduk (bbmap version 35.82) and the paired-ends were merged with the "join_paired_ends.py" script from ea_utils version 1.1.2-537. The AOB merged sequences were dereplicated and clustered into 97% AOB OTUs with minimum size of 2 using USEARCH version 9.2.64 (commands: derep_fulllength, and cluster_otus; Edgar, 2010). These parameters were chosen based on the recommendation found in the USEARCH manual (see also Figure S1). To confirm the functional potential of the OTUs as *amoA* (KEGG pathway K10944), the centroids were compared to the KEGG database (2014-03-17 version) using uproc-dna (UPROC v1.2.0; [39]). The table of OTU abundances across samples was created with the usearch global command based on 97% identity of sequences to the OTUs. Taxonomy was assigned to OTU centroids by diamond blastx v0.8.20 against the 2016-10-04 NCBI-nr database [40]. When this step yielded only classifications in the category "environmental samples," taxonomy was assigned instead by best blastn (e-value cutoff 0.02; blast v2.6.0) comparison against the custom *amoA* database described below.

To support the taxonomic classification results, a phylogenetic tree was created to depict the relationships between the 54 *amoA* OTUs and their closest matches in the custom *amoA* database. The latter was constructed as follows. High-quality *amoA* AOB sequences were downloaded from the FUNGENE RDP database (v9.4.1) with score above 350, HMM coverage above 80% and a minimum amino acid size of 270. Duplicates were reduced to one entry. The *amoA* OTU centroids and reference *amoA* AOB sequences along with an outgroup *amoA* sequence from *Nitrosococcus oceani* C-27 were aligned using ClustalW and used as input to make a phylogenetic tree in MEGA7 [41, 42]. The Maximum Composite Likelihood method was used to calculate phylogenetic distances, and bootstrap tests with 1000 replicates were performed [43]. The iTOL was used to create the final tree with bootstrap values of at least 90% depicted on the branches [44].

5.2.4 *amoA* AOB OTU processing and beta and alpha diversity analyses

Statistical analyses were carried out in R version 3.3.1 using R-Studio version 1.0.136. The R package phyloseq was used to handle the amplicon datasets. To remove outliers, the *amoA* AOB samples with less than 120 sequences were filtered out. To evaluate the sequencing coverage of the AOB communities, Good's coverage was calculated (package jfq3/QsRutils) and rarefaction curves were produced. Relative abundances of the *amoA* AOB OTUs were converted to absolute abundances by multiplying by sample the relative abundances by the relevant gene copy numbers previously obtained [24].

To ascertain the effect of treatment on the AOB community structure, we ordinated the *amoA* AOB samples using Bray-Curtis distances based on OTU relative abundance profiles. Multivariate homogeneity of dispersion was checked with function "betadisp" in the vegan R package. If dispersions were homogeneous, the effects of time point, treatment within time point, and time point within treatment were assessed through PERMANOVA analyses ("vegan" R package). Post-hoc tests of different pairwise group means were carried out using the "pairwiseAdonis" R package [45].

To determine the effect of treatment and time point, treatment within time point and time point within treatment on the AOB community alpha diversity, the data was first rarified to 120 sequences across samples using random seed 42. After confirming that all the data were not normal using the Shapiro-Wilk test and visual check of quantile plots, two-way crossed analyses of treatment and time point, and one-way analyses of treatment within time point and time point within treatment were evaluated using Kruskal-Wallis tests. These were supplemented with Dunn's post-hoc tests.

5.2.5 *16S rRNA* OTU processing and differential abundance and indicator species analyses

We supplemented the analyses of the *amoA* AOB dataset using the previously published 16S rRNA gene sequence dataset [24]. Good's coverage was calculated and rarefaction curves were produced as described for the *amoA* AOB OTU dataset. The *16S rRNA* OTU abundance dataset was processed as follows. Samples with less than 3000 sequences and *16S rRNA* OTUs with less than 23 sequences across all samples were filtered out. To determine significantly different nitrifier *16S rRNA* OTUs between treatments, differential abundance analysis was applied between treatment pairs considering all time points. The DeSeq2 package, which applies a negative binomial transformation of the filtered abundance data to

stabilize variances, was used for the differential abundance testing [46]. The Wald test with local model fit was applied to the *16S rRNA* data; orthogonal contrasts of the control and all other treatments, and of the urea against the treatments with a nitrification inhibitor, were carried out using Bonferroni-Hochberg correction for multiple tests. Significantly different *16S rRNA* OTUs with Bonferroni-adjusted p-values of less than 0.05 were identified.

The *16S rRNA* relative abundances were converted to absolute abundances using the *16S rRNA* copy numbers previously obtained by real-time PCR [24]. To examine the *16S rRNA* OTUs that were indicators of combinations of up to three treatments, we used the *multipatt* function from the "indicspec" R package to apply Legendre's indicator species analysis on the *16S rRNA* absolute abundances. Multiple comparison p-values were adjusted using the Benjamini-Hochberg correction.

5.2.6 Spearman correlations of *amoA* AOB and nitrifier *16S rRNA* OTUs with environmental variables

To reveal correlations between nitrifier OTU abundances and environmental variables, a subset of the previously published environmental data was employed [24]. Log transformations of the gene copy numbers obtained by qPCR (*nirS*, *nirK*, *amoA* AOB, *amoA* AOA, total Archaeal, *16S rRNA*) were carried out leaving the other variables (CO_2, N_2O, CH_4, soil NH_4–N, soil NO_3–N, soil pH and WFPS) untransformed (Figure S2). The nitrifier *16S rRNA* and *amoA* AOB OTU relative and absolute abundances, and the nitrifier *16S rRNA* normalized abundances, were correlated with the environmental variables using Spearman correlations. Significant correlations ($p<0.01$) were kept; for visualization the correlations were clustered using complete linkage clustering through the "corrplot" package.

5.3 Results

5.3.1 *amoA* AOB community sequencing coverage and composition

Processing of the *amoA* AOB amplicon data resulted in 68,211 sequences, which were clustered into 54 OTUs. The number of sequences ranged between 121 and 3,019 across the 127 samples (4 treatments X 8 time points X 4 replicates with one outlier sample removed). The samples had average Good's coverage of at least 94% (**Supplementary Table S1**), which was supported in the rarefaction curve results, with more sequences not adding more species in the samples with more sequences (**Supplementary Figure S3**). At the genus level, the AOB com-

munity was composed of unclassified Betaproteobacteria, *Nitrosomonas* and *Nitrosospira* (which included the *Nitrosovibrio* classification; **Figure 1A**). The phylogenetic tree of the *amoA* AOB OTUs with reference sequences indicated that these aligned with *Nitrosospira* (52/54 *amoA* OTUs) and *Nitrosomonas* (2/54 *amoA* OTUs) (**Figure 2**). In support of the low diversity of the *amoA* AOB communities, the 16S rRNA gene dataset revealed only two *Nitrosospira* OTUs (abundant OTU 30 and rare *16S rRNA* OTU 1102) and one *Nitrosomonas* OTU (rare *16S rRNA* OTU 2875). Further, the *Nitrosospira 16S rRNA* OTUs had similar absolute abundances as the *Nitrosospira amoA* AOB OTUs across the treatments (**Figure 1B and 1D**).

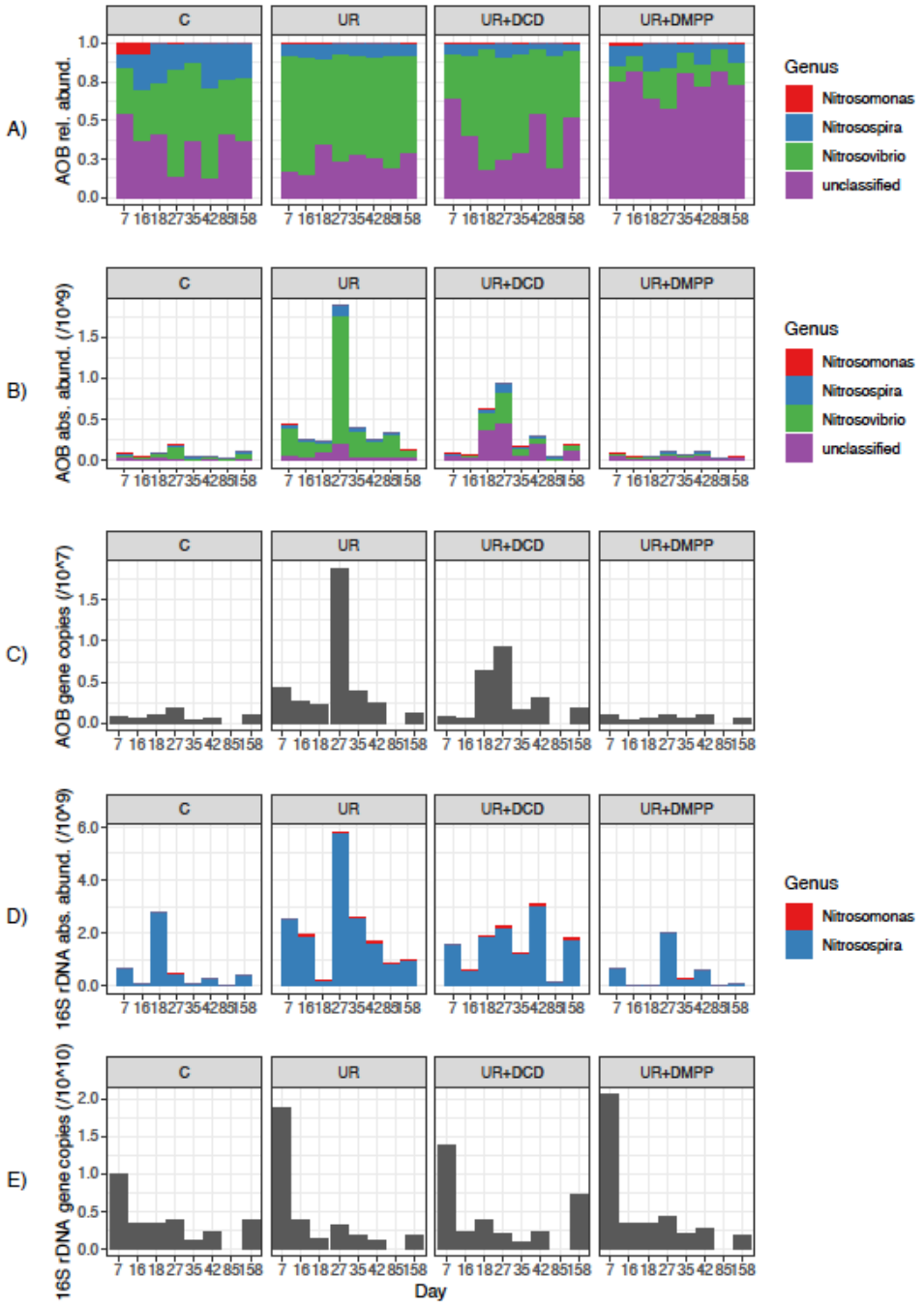


Figure 1. Taxonomic distributions of the *amoA* AOB amplicon samples by A) relative abundances or B) absolute abundances within genus, and C) the amoA AOB gene copy numbers. Also included are the D) taxonomic distributions of the *16S rRNA* amplicon samples by absolute abundances within the *Nitrosomonadaceae* family and E) the gene copy numbers of 16S rRNA gene sequences. Mean values within treatments and time points are shown. Treatments were the unfertilized control (C), urea (UR), urea with dicyanimide (UR+DCD) and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). Day = days after fertilization.

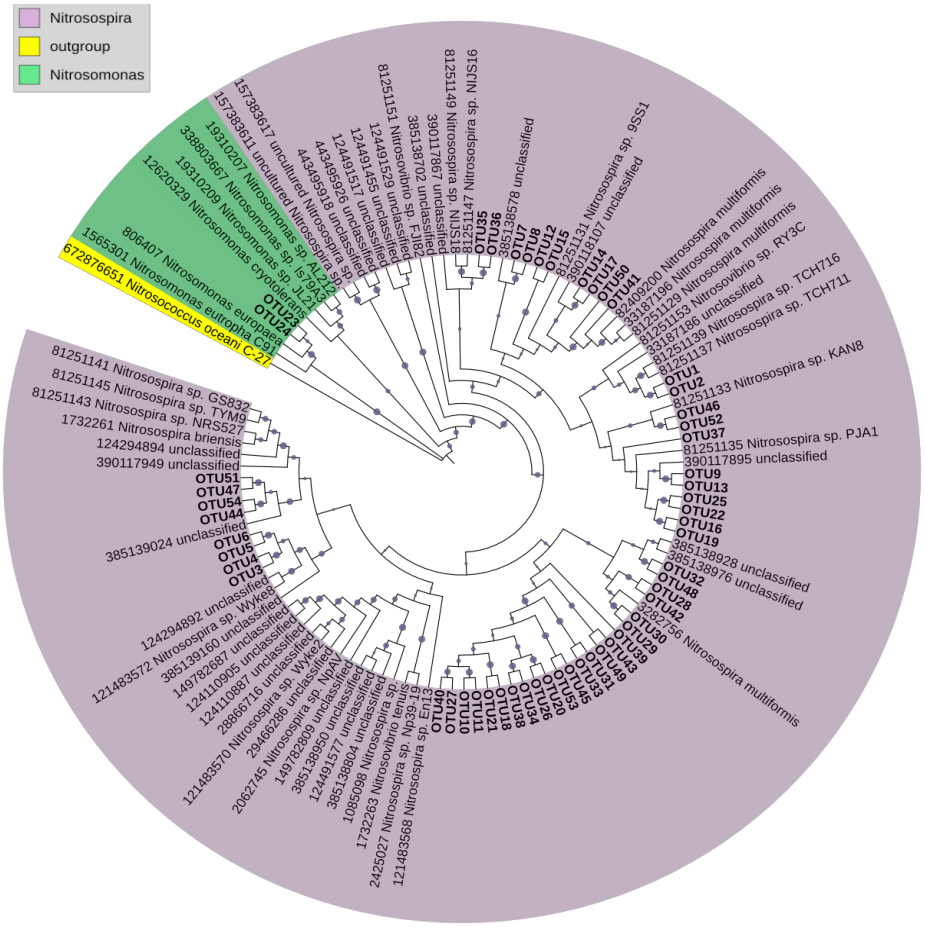


Figure 2. Phylogenetic analysis of *amoA* AOB OTUs and reference *amoA* sequences from the FUNGENE database based on the Maximum Likelihood distance method. Bootstrap values (1000 replicates) of less than 90% are depicted by the blue dots on the branches. The *Nitrosospira* are depicted with purple color bars, *Nitrosomonas* with green color bars, and the outgroup *Nitrososoccus* with the yellow color bar.

5.3.2 Treatment effects on *amoA* AOB community beta diversity

Beta dispersion analysis on all the samples revealed that treatment, but not time point, had a significant effect on the AOB community dispersions ($F=3.6529$, $P<0.05$). Subsequent beta diversity analysis revealed that time point, considering all treatments, had no effect on the AOB community structures (**Supplementary Table S2**). Ordination plots showed that the *amoA* AOB communities overlapped between treatments, considering all time points, according to 95% confidence intervals (**Figure 3A**). Within time points and treatments, the beta dispersions of the

amoA AOB communities were unaffected by treatment and time point, respectively. Treatment had a significant effect on the *amoA* AOB community structures only within days 7 and 16 (PERMANOVA; P<0.1; **Supplementary Table S2**). However, pairwise comparisons revealed that no *amoA* AOB community structures were significantly different between treatments within these time points. Time point had no effect on *amoA* AOB community structures within any treatment. Ordination plots within time point revealed that the *amoA* AOB communities did not cluster separately for treatments nor time points at 95% confidence intervals (**Figures 3C and 3D**).

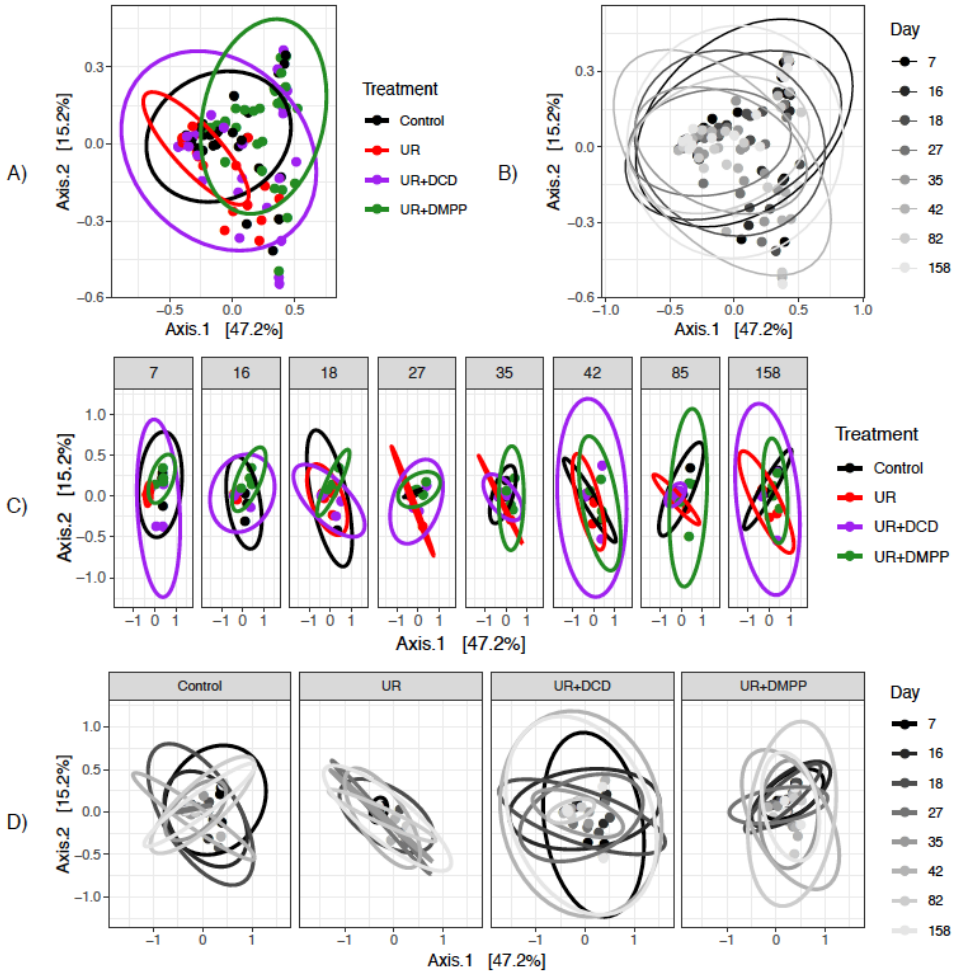


Figure 3. Ordination plots of the ammonia-oxidizing bacterial communities using PCoA on Bray-Curtis sample distances based on *amoA* AOB OTU relative abundances (n=127) A) across all time points (n=127) and B) across all treatments, C) by treatment within each time point (n=16) and D) by time point within each treatment (n=24). Time points were 7, 16, 18, 27, 35, 42, 82, and 158 days after fertilization. Treatments were unfertilized (Control, black), urea (UR, red), urea with dicyanimide (UR+DCD, purple), urea with 3,4-dimethylpyrazole phosphate (UR+DMPP, green). Confidence intervals of 0.95 are drawn around the treatments or days as ellipses.

5.3.3 Treatment and time point effects on *amoA* AOB community alpha diversity

The alpha diversities of the *amoA* AOB communities ranged from 1 to 3 based on Shannon index (**Figure 4**). Considering all time points, treatment had an effect on the alpha diversity of the *amoA* AOB communities (chi-squared value 33.884, p-value = $2.096e^{-07}$), but time point had no effect on the alpha diversities when considering all treatments. Post-hoc testing over all time points found that the *amoA* AOB communities in the DMPP treatment had higher alpha diversity compared to the other treatments (Dunn's test, $p < 0.05$; **Figure 4**). Within time point, treatment had an effect on the *amoA* AOB alpha diversities for days 7, 18, 27, with chi-squared values of 7.6103 (p-value 0.05479), 4.7792 (p-value 0.1887) and 6.7721 (p-value 0.07953), respectively. However, *post hoc* testing revealed no different pairs. Within treatment, time point had an effect on the *amoA* AOB community alpha diversities only for the Control treatment (chi-squared 12.534, p-value=0.08431); further, pairwise post hoc tests revealed no difference in alpha diversity between treatments.

5.3.4 Differential abundance of nitrifier *16S rRNA* OTUs and treatment group indicators

From the 16S rRNA gene sequence data, four genera of nitrifiers were represented: *Nitrosomonas* (1 OTU), *Nitrososphaera* (37 OTUs), *Nitrosospira* (2 OTUs) and *Nitrospira* (11 OTUs). The variance-stabilized trajectories of *Nitrosospira*, *Nitrososphaera* and *Nitrospira 16S rRNA* OTUs across the four treatments can be seen in Figure S4. The two *16S rRNA Nitrosospira* OTUs showed a similar trend across the treatments, with higher abundances in the urea and urea with DCD treatments compared to the control and the urea with DMPP treatments. The *16S rRNA Nitrososphaera* OTUs showed three trends, with OTUs 11 and 429 having lowest abundances in the control treatment and higher abundances in the treatments with urea, with the highest abundances in the urea with DMPP treatment; OTUs 40 and 45 having highest abundances in the control treatment, lower abundances in the treatments with urea, and the lowest abundance in the urea treatment; and OTUs 112 and 39 having highest abundances in the control and urea with nitrification inhibitor treatments and the lowest abundance in the urea treatment. The *16S rRNA Nitrospira* OTU followed the last trend with the highest abundances in the control and urea with nitrification inhibitor treatments and the lowest abundance in the urea treatment.

The differential abundance and indicator species analyses generally supported the abundance trajectories of the *16S rRNA* nitrifier OTUs. Differential abundance analysis revealed the nitrifier *16S rRNA* OTUs that were significantly over- and under-represented between pairwise comparisons of treatments based on variance-stabilized abundances (**Supplementary Table S3**). Of the *Nitrosospira 16S rRNA* OTUs, OTU 30 was an indicator of the control, urea and urea with DCD treatments, while OTU 1102 was an indicator of only the urea and urea with DCD treatments (adjusted p-value < 0.1; **Supplementary Table S3**). Of the *Nitrososphaera 16S rRNA* OTUs, OTU 45, OTU 112, OTU 40, OTU 39 and OTU 11 were indicators of the control, urea with DCD and urea with DMPP treatments. Of the *Nitrospira 16S rRNA* OTUs, OTU 79 was an indicator of the control, urea with DCD and urea with DMPP treatments.

5.3.5 Nitrifier *amoA* and *16S rRNA* OTU and environmental correlations

The correlations of the environmental variables with the gene copy numbers of AOB, AOA, *nirK*, *nirS*, *16S rRNA* total bacteria and total Archaea (**Supplementary Figure S5**) depict the positive links between AOB, N_2O, NO_3^-, NH_4^+, WFPS and pH, and AOA, *nirS*, *nirK*, total archaea and total bacteria; and the negative links between CO_2, CH_4^+ and WFPS, and AOA, total Archaea, NH_4^+, N_2O and NO_3^- (**Figure 5A**). As can be seen in Figure 5B which depicts correlations including the normalized abundances of *16S rRNA* OTUs, N_2O emissions were correlated with *amoA* AOB copy numbers, water-filled pore space (WFPS), NO_3^-, NH_4^+ and pH. Interestingly, the *16S rRNA* and *amoA* AOB OTU correlations clustered with the previous variables with the exception of NH_4^+ and pH, which nevertheless suggests that *Nitrosospira* (OTU 30 and OTU 1102) were the N_2O-producing AOB in these soils. Other interesting clusters were the *16S rRNA Nitrososphaera* OTUs 429 and 11 with NO_3^-, NH_4^+ and pH; the *nirS*, *nirK*, total archaeal and 16S rRNA gene copy numbers; and the *amoA* AOA, *16S rRNA Nitrospira* OTU 79, the *16S rRNA Nitrososphaera* OTUs 45, 112, 40 and 39. These clusters were found in all the correlations with absolute and relative abundances of the *amoA* AOB and the 16S rRNA gene sequence data (**Supplementary Figure S5**).

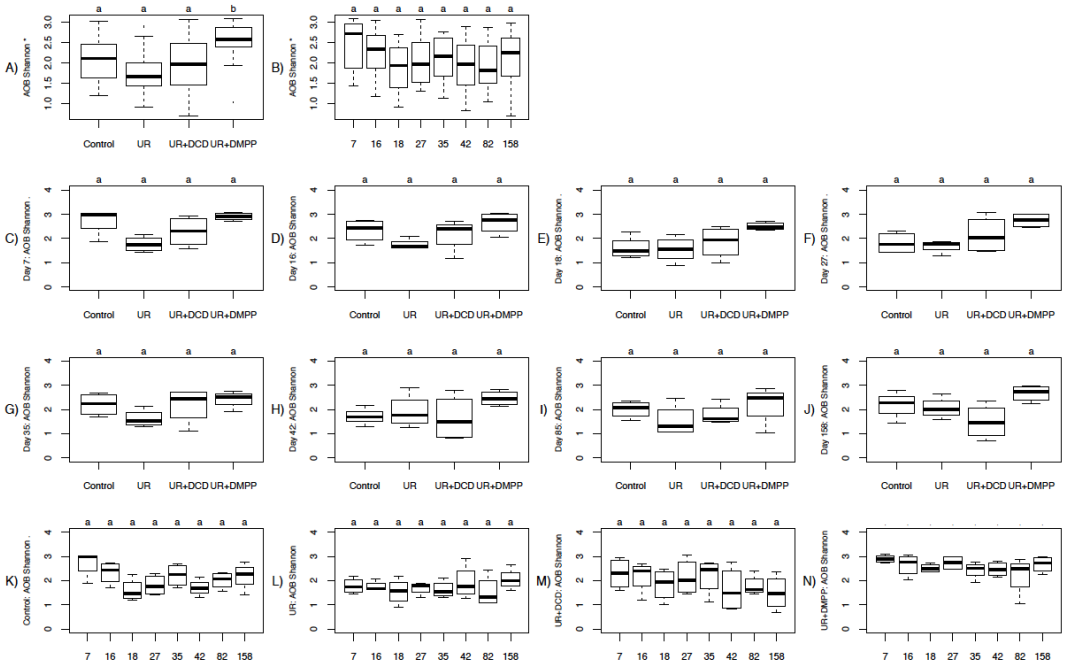


Figure 4. Alpha diversity of the amoA AOB communities as affected by A) treatment, for all time points, B) time point, for all treatments, C-J) treatment, within each time point, and K-N) time point, within each treatment. Treatments were unfertilized (C), urea (UR), urea with dicyanimide (UR+DCD) and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP); time points were 7, 16, 18, 27, 35, 42, 82,158 days after fertilization. The y-axis label includes the result of a Kruskal-Wallis chi-squared test ("*" for $p<0.05$, "." for $p<0.10$); the letters above the plots represent the results of Dunn's post hoc tests at alpha < 0.05 in which similar letters denote no difference between groups.

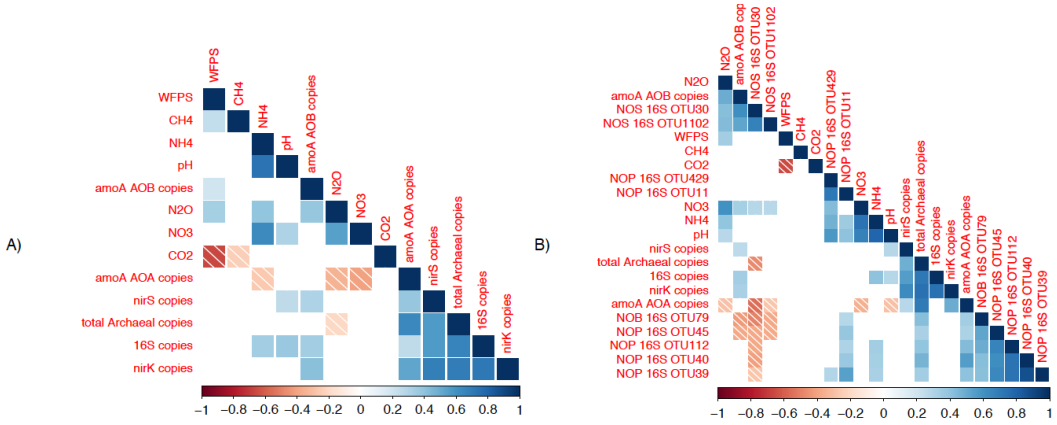


Figure 5. Cluster plots visualizing Spearman's correlations A) between environmental variables and gene copy numbers, and B) between environmental variables, gene copy numbers and the normalized abundances of the *16S rRNA* nitrifier OTUs. Normalization was carried out using DeSeq2. Only significant correlations are shown ($p<0.01$). Clusters were determined using complete linkage clustering. NOS = *Nitrosospira*, NOP = *Nitrososphaera*, NOB = *Nitrospira*.

5.4 Discussion

From our previous work, we found that bacterial *amoA* (AOB) but not archaeal *amoA* (AOA) nor denitrification gene copy numbers (*nirK*, *nirS*) were correlated with nitrous oxide emissions from tropical soil growing sugarcane [24]. Here we found evidence that the AOB responsible for the N_2O emissions was most similar to the *Nitrosospira* spp. (*Nitrosovibrio* RY3C), based on the decrease in abundance of these OTUs in soils with the nitrification inhibitors in comparison with the urea treatment and the correlation of these OTUs with N_2O emissions. The *Nitrosovibrio* RY3C species was originally isolated from avocado rhizosphere and its nitrifying activity was susceptible to DCD [47]. To our knowledge, just one other study has identified *Nitrosospira* spp. as the N_2O-generating AOB in tropical soil under sugarcane, and that study applied NH_4NO_3 as the N source [25]. The *Nitrosospira* in general are widespread spiral soil bacteria with generally low specificity for ammonia and thus found in soils under high levels of ammonia [15, 48, 49]. The only other AOB identified was *Nitrosomonas*, which was present in low abundance in the soils and was not linked to N_2O emissions. The *Nitrosomonas* are generally found in soils with high N inputs; moreover, *Nitrosomonas europaea* has a 3.5-fold higher Vmax compared to *Nitrosospira* sp., suggesting that these AOB do better in soils with consistently higher N [50]. We suggest that the conditions these soils encounter (generally low N with occasional high N inputs from fertilization) selects for the *Nitrosospira*, but that perhaps a *Nitrosomonas* species adapted to these conditions is present in low abundance.

The AOB are widely implicated in N_2O emissions under conditions favoring nitrification in tropical and temperate soils, in contrast to the AOA [51-55]. This is thought to be linked to the enzymatic capabilities of different AOB and AOA species, with the former generating higher amounts of N_2O through both abiotic (nitric oxide oxidation by O_2) and biotic (incomplete hydroxylamine oxidation and nitrifier denitrification) mechanisms, while the latter likely emits lower N_2O using only an abiotic (nitric oxidation by O_2) mechanism [56, 57]. While the AOB *Nitrosospira* was abundant in the soils under urea and urea with DCD treatments, we found that in the unfertilized and in the urea with DMPP treatment, the AOA *Nitrososphaera* were more abundant. More than 5 AOA *Nitrososphaera 16S rRNA* OTUs were identified compared to the two AOB *Nitrosospira 16S rRNA* OTUs; this supports the idea that the conditions in these unfertilized soils normally support the AOA *Nitrososphaera* rather than the AOB *Nitrosospira* or *Nitrosomonas* as the main ammonia oxidizers. Moreover, these native *Nitrososphaera* appeared to be non-N_2O-producing AOA. These results support observations that the AOA

Nitrososphaera is associated with low concentrations of ammonia linked to the stronger affinity of the archaeal ammonia monooxygenase for ammonia [48].

Interestingly, we identified two types of *Nitrososphaera* (AOA): one cluster of *Nitrososphaera* OTUs was more abundant in the soils with urea and DMPP, while the other cluster was more abundant in the unfertilized soils and co-varied with the NOB *Nitrospira*. The *Nitrospira* was the only nitrite-oxidizer found in our soils according to the 16S rRNA gene sequence data; interestingly, this was most abundant in the unfertilized soils and co-varied with AOA *Nitrososphaera* OTUs. The *Nitrospira* are thought to be adapted to low NO_2^- availability [20], which might explain their presence in our soils instead of *Nitrobacter* [19, 58]. Further, perhaps the *Nitrososphaera* and *Nitrospira* naturally interact in these unfertilized soils, as has been suggested for unfertilized grassland soils [34]. Future work could focus on this hypothesized interaction between non-N_2O-generating *Nitrososphaera* and *Nitrobacter*, which appears to be selected for by low levels of available substrate and might be enhanced by adding organic residues with high C:N [51, 59].

The inhibitors DCD and DMPP are both thought to inhibit ammonia monooxygenase by chelating the Cu co-factor in the enzyme [9].The limitation of *Nitrosospira* but not *Nitrososphaera* by DCD has been shown also in a paddy field soil and in microcosms of *Nitrosospira multiformis* but not *Nitrososphaera viennensis* [60, 61]. Based on gene copy numbers, the AOB but not the AOA were inhibited by DMPP in a sandy soil [62]; and the AOB but not the AOA were inhibited by DCD in a grazed grassland system [54]. In a Chinese vegetable soil, DMPP rather than DCD was revealed to be the more effective inhibitor of N_2O-producing AOB rather than AOA, although the N source urea was also amended with manure [63]. In studies of nitrification in agricultural soils, DMPP inhibited AOB expression under neutral pH conditions [64, 65]. The different success of the nitrification inhibitors appears to be a function of temperature, Cu-levels, and variation in abundance, genetic potential and/or expression levels of the targeted nitrifiers [9]. The different effects of DCD and DMPP on the abundance of the AOB *Nitrosospira* and the AOA *Nitrososphaera* found here suggests that evaluating the nitrification dynamics of these species in culture would be interesting for future work.

In contrast to our hypothesis that the nitrification inhibitor treatments would decrease the *amoA* AOB community alpha diversity, this diversity remained largely unchanged across treatments. There overall was low alpha diversity of the *amoA* AOB community, which was supported in both the *amoA* AOB and *16S rRNA* sequence results. Nitrifiers occupy a specific functional niche in the soil

environment, and the nitrifying functions are restricted to a handful of genera; new AOB are not likely to appear at least over the relatively short duration of this experiment (in total 258 days, subset presented here was 158 days). Moreover, the sugarcane plant competes with microbes for NH_4^+ and NO_3^- and these substrates are not likely to remain immobile long in this soil [66]. The highly weathered soils have high soil drainage capacity and have been under more than 20 years of sugarcane cultivation. Due to the long time of cultivation by sugarcane, likely the nitrifiers found in this soil are those that are adapted to the natural unfertilized conditions, to the brief high inputs of ammonia through urea fertilization, and to the competition with the sugarcane plant for ammonia. We speculate that the overall low nitrifier diversity and the selection of the nitrifiers that are present in these soils are driven by the generally low N levels.

While we inferred our results from the analysis of two different amplicon gene datasets, making our results more robust, some caveats to our methods should be acknowledged. The precision of the OTU classification was dependent upon the coverage of the databases used; for example, for our *16S rRNA* dataset we were only able to confidently classify to the genus level. This prevented us from directly comparing the classification results between the *amoA* AOB and *16S rRNA* datasets at the species level. However, we were reassured by the congruence of the *amoA* and *16S rRNA* sequence data relative to the absolute abundances of the *amoA* AOB at genus level. Further, the low diversity of the *amoA* bacterial communities was echoed in the *16S rRNA* data with just a few OTUs identified as *Nitrosospira* and only one as *Nitrosomonas*. Last, though the *16S rRNA* samples had high Good's coverage values between 85% and 99%, there is a possibility that the nitrifying subset of the community did not have such high coverage values. However, the focus of this paper was the *amoA* AOB community, although future studies could target in more depth and with more specificity the nitrifying network in these soils.

In summary, the nitrification inhibitors in our experiment were revealed to target the N_2O-producing bacterial ammonia-oxidizer *Nitrosospira* spp. in the soils. The low N availability appeared to drive the nitrifier community found in these soils. Treatment with urea and DMPP appeared to favor one functional type of AOA *Nitrososphaera* while the unfertilized soils revealed potentially interdependent AOA *Nitrososphaera* and NOB *Nitrospira*; it seems these species do not greatly contribute to N_2O emissions. Our results support the use of DMPP and especially DCD as inhibitors of N_2O-producing *Nitrosospira* spp. in tropical soils under sugarcane. The DMPP treatment may also increase the amount of NH_4^+ in the soil, allowing the sugarcane crop to uptake this N source while blocking the

N_2O from *Nitrosospira*. Furthermore, we provide evidence that the nitrification process in these soils is controlled by a few bacterial and archaeal species, driven mainly by the overall low N levels, and which have contrasting functional potentials for N_2O emission rates.

5.5 Declarations

Acknowledgements

This work was supported by NWO-FAPESP (The Netherlands Organization for Scientific Research NWO-729.004.003- Sao Paulo State foundation FAPESP-2013/50365-5). Publication number 6678 of the NIOO-KNAW, Netherlands Institute of Ecology.

Conflict of interest

The authors declare no conflict of interest.

Author's contributions

EEK and NAC designed the current study. JRS and HC designed the original experiment. JRS and KSL carried out the field work and DNA extractions. JRS and AP ran the qPCRs. AP prepared the *amoA* amplicon library for sequencing. NAC performed all bioinformatics and NAC and JRS performed analysis steps. NAC wrote the paper. JAV, HC and EEK contribute in the interpretation of the results and discussion of the paper. All authors read and approved the final version of the manuscript.

5.6 Supplementary Material

Table S1. Good's coverage of the *16S rRNA* (n=93) and *amoA* AOB (n=127) datasets. Treatments were unfertilized (C), urea (UR), urea with dicyanimide (UR+DCD) and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). Day: days after fertilization.

Treatment	Day	*16S rRNA* (avg; %)	*16S rRNA* (sd)	*amoA* (avg; %)	*amoA* (sd)
C	7	92.0	5.8	98.0	0.6
	16	96.4	1.9	96.7	1.1
	18	95.5	2.0	98.2	0.5
	27	93.9	3.2	98.1	0.3
	35	93.7	1.7	98.4	0.7
	42	91.3	8.0	99.2	0.6
	82	94.4	0.8	98.4	0.3
	158	90.9	4.6	98.4	0.8
UR	7	91.6	4.0	98.9	0.8
	16	96.7	1.0	97.9	1.0
	18	95.6	0.8	97.9	0.5
	27	95.2	0.1	98.1	1.2
	35	93.7	0.9	99.1	0.5
	42	94.7	1.3	99.0	0.7
	82	97.9	0.2	98.2	0.5
	158	91.9	2.1	96.6	2.2
UR+DCD	7	95.6	0.9	98.2	0.8
	16	94.4	2.5	98.0	1.1
	18	95.8	1.7	98.7	0.7
	27	96.4	1.9	98.0	1.9
	35	94.6	0.3	98.3	1.2
	42	90.2	2.3	99.0	0.5
	82	90.4	3.6	98.0	0.8
	158	88.7	0.9	98.5	0.8
UR+DMPP	7	94.7	0.4	95.6	1.9
	16	95.9	1.2	96.0	1.3
	18	96.4	0.6	98.6	1.0
	27	95.6	0.9	97.1	1.0
	35	96.2	0.5	97.6	1.3
	42	91.0	1.9	98.8	0.6
	82	93.4	1.0	97.9	0.9
	158	95.4	1.5	97.3	1.2

Table S2. Beta diversity results from PERMANOVA tests of the effects of treatment and time point on the AOB communities based on the Bray-Curtis distances between *amoA* OTU relative abundance profiles. Treatments were unfertilized (C), urea (UR), urea with dicyanimide (UR+DCD) and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). Days were 7, 16, 18, 27, 35, 42, 85 and 158 days after fertilization.

All treatments	df	F statistic	R^2	P-value
Day	7	0.8253	0.0463	0.711
Residuals	119		0.9537	
Within Day 7				
Treatment	**3**	**2.0553**	**0.33942**	**0.005 ****
Residuals	**12**		**0.66058**	
Within Day 16				
Treatment	**3**	**1.5286**	**0.29422**	**0.085 .**
Residuals	**12**		**0.70578**	
Within Day 18				
Treatment	3	0.9178	0.18663	0.623
Residuals	12		0.81337	
Within Day 27				
Treatment	3	1.2305	0.23525	0.207
Residuals	12		0.76475	
Within Day 35				
Treatment	3	1.1248	0.21948	0.272
Residuals	12		0.78052	
Within Day 42				
Treatment	3	1.1308	0.2204	0.271
Residuals	12		0.7796	
Within Day 82				
Treatment	3	1.4292	0.26324	0.134
Residuals	12		0.73676	
Within Day 158				
Treatment	3	1.0193	0.20308	0.410
Residuals	12		0.79692	
Within Control				
Day	7	1.1906	0.25775	0.157
Residuals	24		0.74225	
Within UR				
Day	7	0.95792	0.22573	0.584
Residuals	23		0.77427	
Within UR+DCD				
Day	7	0.81333	0.19174	0.821
Residuals	24		0.80826	
Within UR+DMPP				
Day	7	0.78015	0.18536	0.858
Residuals	24		0.81464	

Table S3. Differentially abundant nitrifier *16S rRNA* OTUs based on pairwise comparisons of *16S rRNA* OTU abundances between treatments. Treatments were unfertilized (C), urea (UR), urea with dicyanimide (UR+DCD), and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). The *16S rRNA* OTUs were included if the mean normalized abundance across all samples was greater than 4 and had log2 fold changes of more or less than 1. Values are significant log2 fold changes from DeSeq2 analysis (p adjusted < 0.05).

			Log_2 fold Change					
Genus	*16S rRNA* OTU Id	Mean Norm. Abund.	UR vs C	DCD vs C	DMPP vs C	DCD vs UR	DMPP vs UR	DMPP vs DCD
Nitrosospira	30	102	2.7	2.0			-2.5	-1.8
	1102	5	3.2	2.9			-2.1	
Nitrososphaera	11	97		2.0	1.7		1.5	1.0
	39	49	-1.1					
	40	44	-1.2					
	45	22	-3.2			2.0	2.1	
	112	15	-2.4				1.9	
	429	9	2.6	2.9	4.0		1.3	
Nitrospira	79	36	-1.3			1.1	1.6	

Table S4. Results of the indicator species analysis depicting indicator nitrifiers 16S rRNA gene sequence for groups of treatments (asterisk indicates adjusted p-value < 0.1) based on absolute abundances of *16S rRNA* OTUs. The OTUs were included if the mean normalized abundances were at least 4. Treatments were unfertilized (C), urea (UR), urea with dicyanimide (UR+DCD), and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). A "1" is in place if the *16S rRNA* OTU was an indicator of the treatment or, taken together, the group of treatments.

Max Order 3							
Genus	*16S rRNA* OTU Id	C	UR	UR+DCD	UR+DMPP	statistic	adj. p-value
Nitrosospira	30	1	1	1		0.95	0.025 *
	1102		1	1		0.82	0.015 *
Nitrososphaera	11		1	1	1	0.95	0.553
	39	1		1	1	0.92	0.459
	40	1		1	1	0.90	0.760
	45	1		1	1	0.94	0.033 *
	112	1		1	1	0.91	0.015 *
	429		1	1	1	0.93	0.009 *
Nitrospira	79	1		1	1	0.94	0.232
Max Order 2							
Genus	*16S rRNA* OTU Id	C	UR	UR+DCD	UR+DMPP	statistic	adj. p-value
Nitrosospira	30		1	1		0.88	0.010 *
	1102		1	1		0.82	0.010 *
Nitrososphaera	11			1	1	0.88	0.218
	39	1		1		0.77	0.825
	40	1		1		0.78	0.752
	45	1			1	0.78	0.854
	112	1			1	0.77	0.340
	429			1	1	0.87	0.022 *
Nitrospira	79	1			1	0.82	0.164
Max Order 1							
Genus	*16S rRNA* OTU Id	C	UR	UR+DCD	UR+DMPP	statistic	adj. p-value
Nitrosospira	30		1			0.66	0.058 *
	1102		1			0.63	0.096 *
Nitrososphaera	11				1	0.73	0.176
	39			1		0.59	0.836
	40	1				0.58	0.843
	45				1	0.55	0.971
	112	1				0.60	0.375
	429				1	0.76	0.036 *
Nitrospira	79				1	0.64	0.188

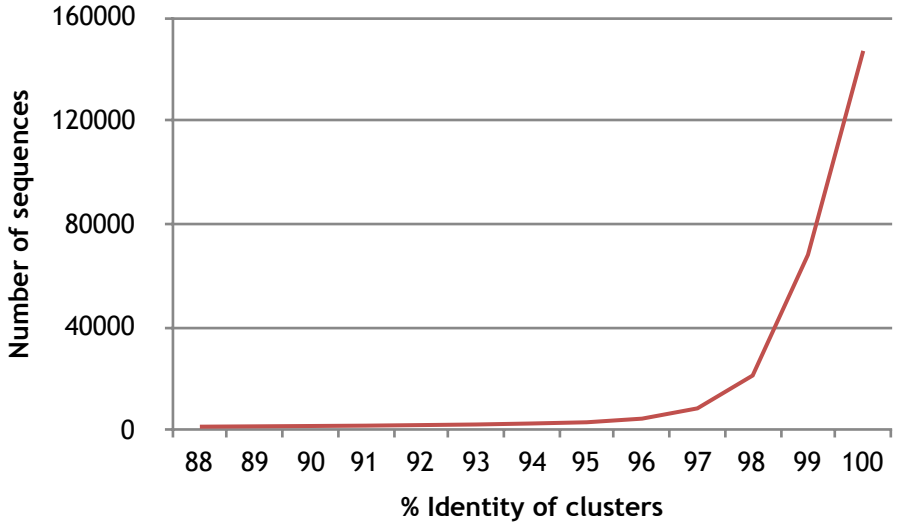


Figure S1. Line graph depicting the number of *amoA* AOB clusters (OTUs) at various percent identity levels from the USEARCH clustering algorithm. The percentage of 97% was chosen as per the recommendation of the USEARCH manual.

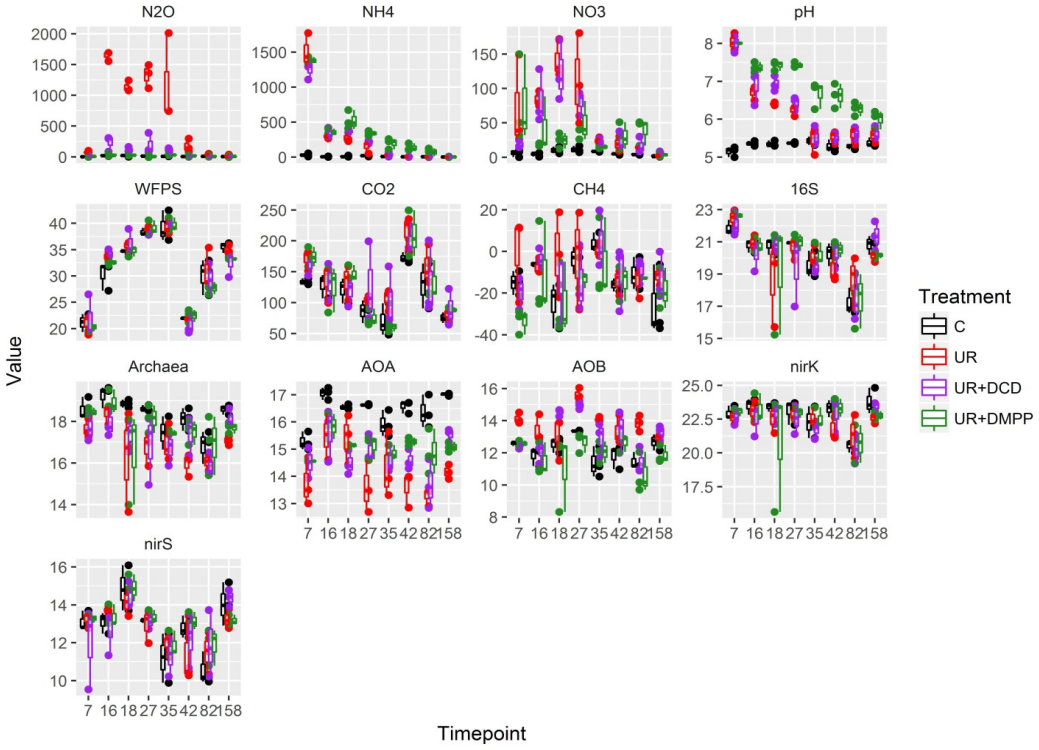
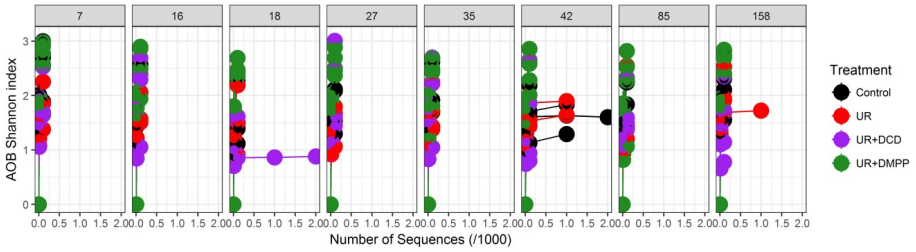


Figure S2. The previously published environmental variables and the gene copy numbers which were used in the current paper (Soares et al., 2016). Treatments were control (C; black), urea (UR; red), urea with dicyanimide (UR+DCD) and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP). Time points were 7, 16, 18, 27, 35, 42, 82 and 158 days after fertilization. The gene copy numbers (qPCR) are presented as log values.

A)



B)

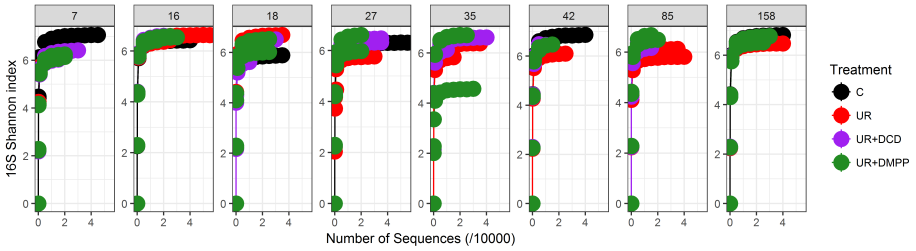


Figure S3. Rarefaction curves of the A) *amoA* AOB (n=127) and B) *16S rRNA* (n=93) datasets across time point. Treatments were unfertilized (C, black), urea (UR, red), urea with dicyanimide (UR+DCD, purple), and urea with 3,4-dimethylpyrazole phosphate (UR+DMPP, green). Days are 7, 16, 18, 27, 35, 42, 85 and 158 days after fertilization.

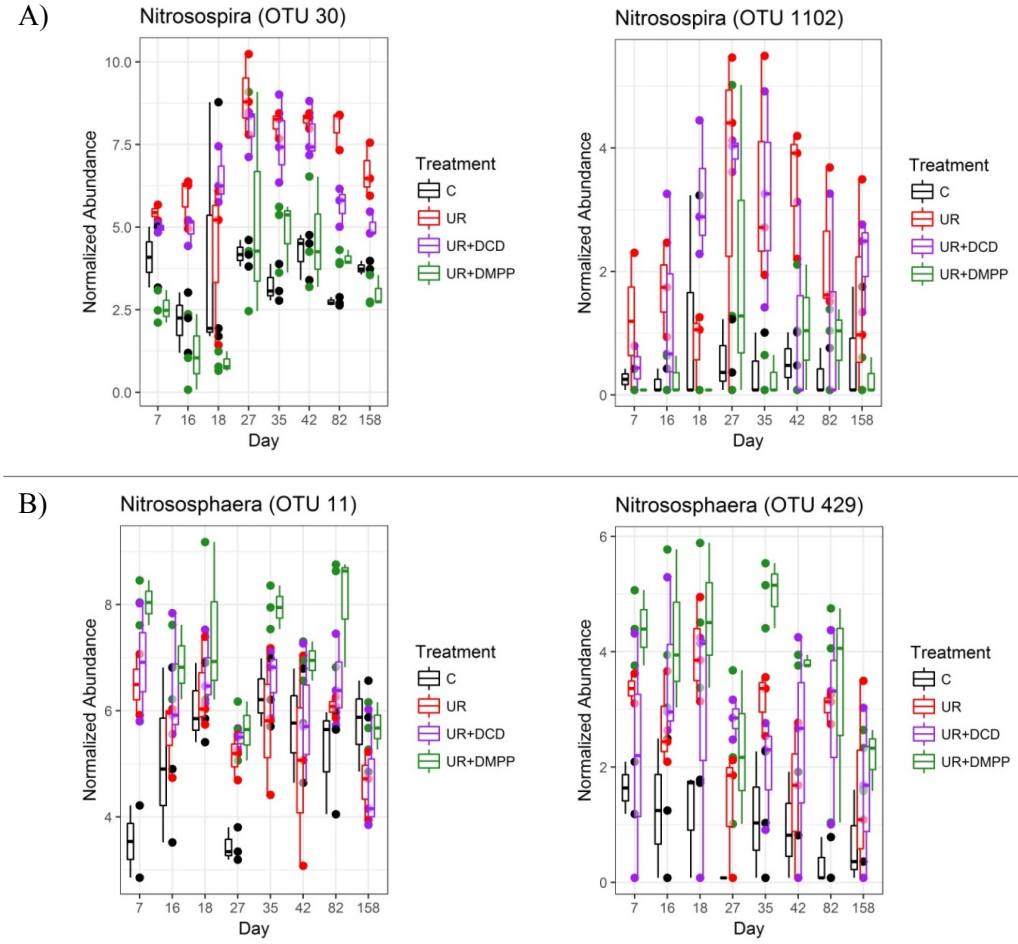


Figure S4. Normalized abundances of the A) AOB *Nitrosospira*, B) AOA *Nitrososphaera* and C) NOB *Nitrospira 16S rRNA* OTUs that were differentially abundant across the treatments based on DESeq2 analysis. Treatments were control (C), urea (UR), urea with DCD (UR+DCD) and urea with DMPP (UR+DMPP). Days are 7, 16, 18, 27, 35, 42, 85 and 158 days after fertilization. Note the different limits on the y-axes. Normalization was carried out with DeSeq2. Note the differing Y-axis limits.

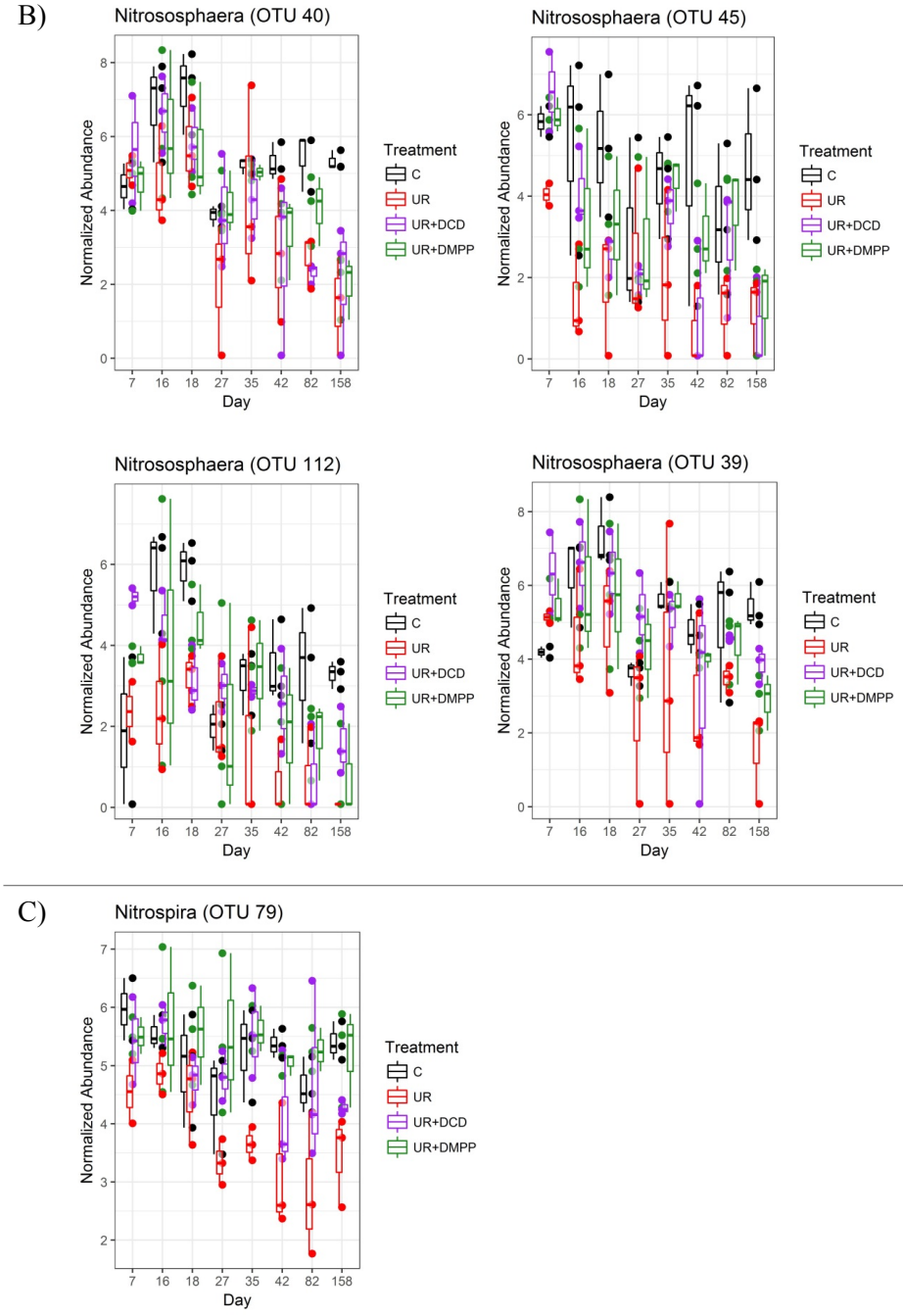
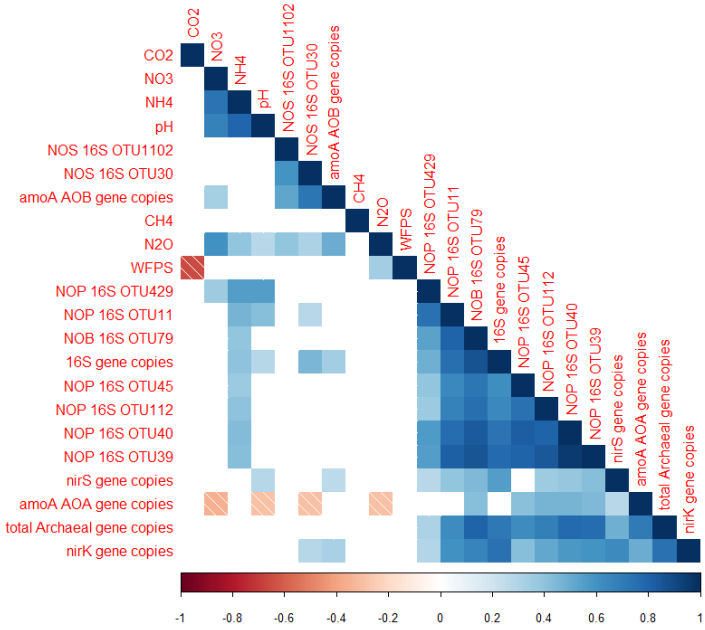
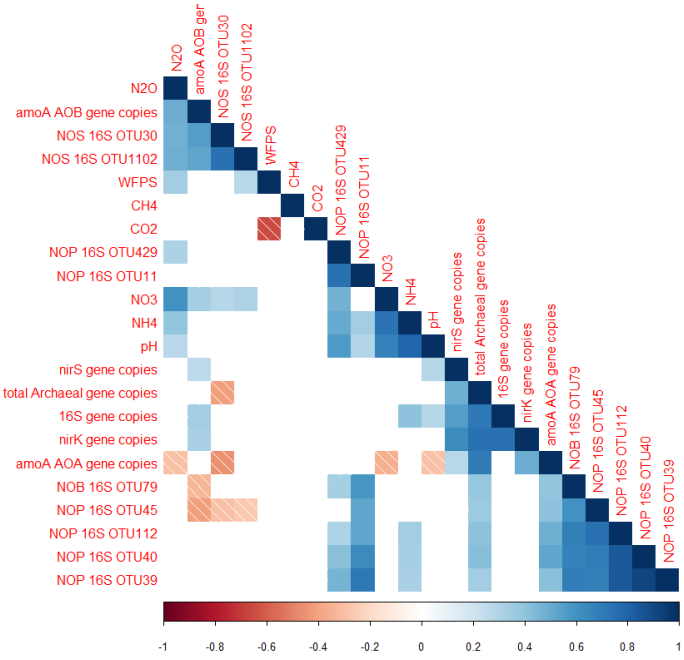


Figure S4 continued. Normalized abundances of the A) AOB *Nitrosospira*, B) AOA *Nitrososphaera* and C) NOB *Nitrospira 16S rRNA* OTUs that were differentially abundant across the treatments based on DESeq2 analysis. Treatments were control (C), urea (UR), urea with DCD (UR+DCD) and urea with DMPP (UR+DMPP). Days are 7, 16, 18, 27, 35, 42, 85 and 158 days after fertilization. Note the different limits on the y-axes. Normalization was carried out with DeSeq2. Note the differing Y-axis limits.

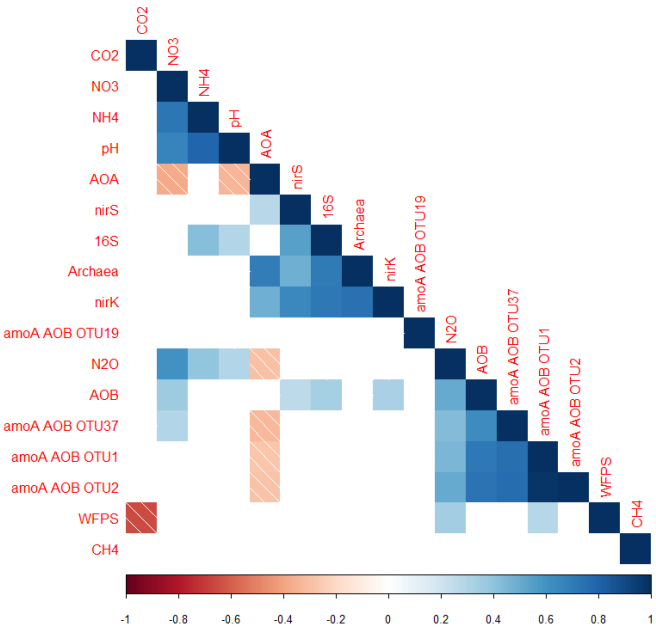


A)

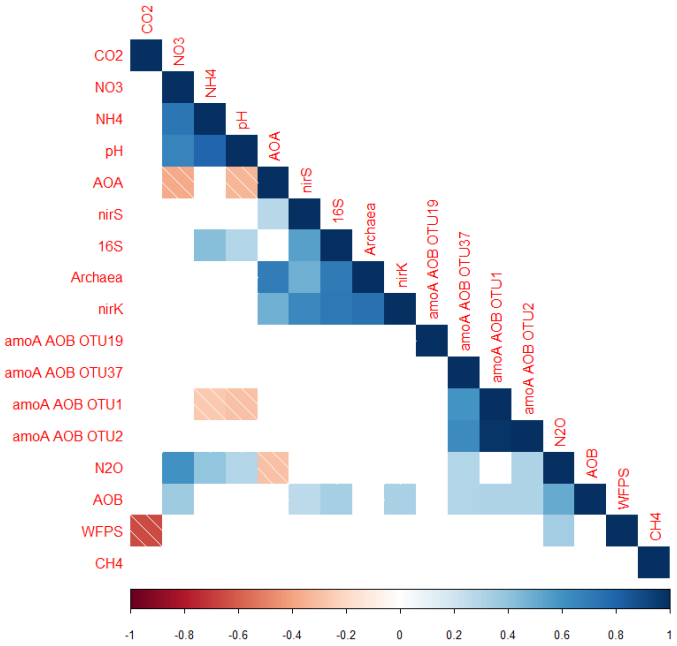


B)

Figure S5 cont'd. Cluster plots visualizing Spearman's correlations between environmental variables and the A) absolute abundances and B) relative abundances of the *16S rRNA* nitrifier OTUs. Only significant correlations are shown ($p<0.01$). Clusters were determined using complete linkage clustering.



C)



D)

Figure S5 cont'd. Cluster plots visualizing Spearman's correlations between environmental variables and the C) absolute abundances and D) relative abundances of the *amoA* AOB OTUs classified as the *Nitrosospira* spp. Nitrosovibrio RY3C. Only significant correlations are shown ($p<0.01$). Clusters were determined using complete linkage clustering.

5.7 References

1. Fields, S., *Global nitrogen: cycling out of control.* Environmental Health Perspectives, 2004. **112**(10): p. A556.
2. Robertson, G.P. and P.M. Vitousek, *Nitrogen in agriculture: balancing the cost of an essential resource.* Annual Review of Environment and Resources, 2009. **34**: p. 97-125.
3. Ciais, P., *Carbon and other biogeochemical cycles Climate Change 2013: The Physical Science Basis ed TF Stocker et al*. 2013, Cambridge: Cambridge University Press.
4. Signor, D. and C.E.P. Cerri, *Nitrous oxide emissions in agricultural soils: a review.* Pesquisa Agropecuária Tropical, 2013. **43**(3): p. 322-338.
5. Ravishankara, A., J.S. Daniel, and R.W. Portmann, *Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century.* science, 2009. **326**(5949): p. 123-125.
6. Venterea, R.T., et al., *Challenges and opportunities for mitigating nitrous oxide emissions from fertilized cropping systems.* Frontiers in Ecology and the Environment, 2012. **10**(10): p. 562-570.
7. Graham, E.B., et al., *Do we need to understand microbial communities to predict ecosystem function? A comparison of statistical models of nitrogen cycling processes.* Soil Biology and Biochemistry, 2014. **68**: p. 279-282.
8. Hu, H.-W., D. Chen, and J.-Z. He, *Microbial regulation of terrestrial nitrous oxide formation: understanding the biological pathways for prediction of emission rates.* FEMS microbiology reviews, 2015: p. fuv021.
9. Ruser, R. and R. Schulz, *The effect of nitrification inhibitors on the nitrous oxide (N_2O) release from agricultural soils—a review.* Journal of Plant Nutrition and Soil Science, 2015. **178**(2): p. 171-188.
10. Guo, J., et al., *Pathways and organisms involved in ammonia oxidation and nitrous oxide emission.* Critical reviews in environmental science and technology, 2013. **43**(21): p. 2213-2296.
11. Liu, S., et al., *Abiotic conversion of extracellular NH2OH contributes to N_2O emission during ammonia oxidation.* Environmental science & technology, 2017. **51**(22): p. 13122-13132.
12. Butterbach-Bahl, K., et al., *Nitrous oxide emissions from soils: how well do we understand the processes and their controls?* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1621): p. 20130122.
13. Wallenstein, M.D., et al., *Environmental controls on denitrifying communities and denitrification rates: insights from molecular methods.* Ecological applications, 2006. **16**(6): p. 2143-2152.
14. Di, H., et al., *Nitrification driven by bacteria and not archaea in nitrogen-rich grassland soils.* Nature Geoscience, 2009. **2**(9): p. 621.
15. Di, H.J., et al., *Ammonia-oxidizing bacteria and archaea grow under contrasting soil nitrogen conditions.* FEMS Microbiology Ecology, 2010. **72**(3): p. 386-394.
16. Hayatsu, M., et al., *An acid-tolerant ammonia-oxidizing γ-proteobacterium from soil.* The ISME journal, 2017. **11**(5): p. 1130.
17. Song, H., et al., *Changing roles of ammonia-oxidizing bacteria and archaea in a continuously acidifying soil caused by over-fertilization with nitrogen.* Environmental Science and Pollution Research, 2016. **23**(12): p. 11964-11974.
18. Li, Y., et al., *Nitrification and nitrifiers in acidic soils.* Soil Biology and Biochemistry, 2018. **116**: p. 290-301.
19. Attard, E., et al., *Shifts between Nitrospira−and Nitrobacter−like nitrite oxidizers underlie the response of soil potential nitrite oxidation to changes in tillage practices.* Environmental Microbiology, 2010. **12**(2): p. 315-326.

20. Nowka, B., H. Daims, and E. Spieck, *Comparison of oxidation kinetics of nitrite-oxidizing bacteria: nitrite availability as a key factor in niche differentiation.* Applied and environmental microbiology, 2015. **81**(2): p. 745-753.
21. Wu, D., et al., *The effect of nitrification inhibitor on N_2O, NO and N2 emissions under different soil moisture levels in a permanent grassland soil.* Soil Biology and Biochemistry, 2017. **113**: p. 153-160.
22. Wu, D., et al., *Nitrification inhibitors mitigate N_2O emissions more effectively under straw-induced conditions favoring denitrification.* Soil Biology and Biochemistry, 2017. **104**: p. 197-207.
23. Liu, R., et al., *Nitrification is a primary driver of nitrous oxide production in laboratory microcosms from different land-use soils.* Frontiers in microbiology, 2016. **7**: p. 1373.
24. Soares, J.R., et al., *Nitrous oxide emission related to ammonia-oxidizing bacteria and mitigation options from N fertilization in a tropical soil.* Scientific reports, 2016. **6**: p. 30349.
25. Lourenco, K.S., et al., *Nitrosospira sp. govern nitrous oxide emissions in a tropical soil amended with residues of bioenergy crop.* Frontiers in Microbiology, 2018. **9**: p. 674.
26. S., L.K., et al., *Dominance of bacterial ammonium oxidizers and fungal denitrifiers in the complex nitrogen cycle pathways related to nitrous oxide emission.* GCB Bioenergy. **0**(0).
27. Lourenço, K.S., et al., *Nitrosospira sp. Govern Nitrous Oxide Emissions in a Tropical Soil Amended With Residues of Bioenergy Crop.* Frontiers in Microbiology, 2018. **9**(674).
28. Soares, J.R., et al., *Enhanced-efficiency fertilizers in nitrous oxide emissions from urea applied to sugarcane.* J Environ Qual, 2015. **44**(2): p. 423-30.
29. Signor, D., C.E.P. Cerri, and R. Conant, *N_2O emissions due to nitrogen fertilizer applications in two regions of sugarcane cultivation in Brazil.* Environmental Research Letters, 2013. **8**(1): p. 015013.
30. Morales, S.E., N. Jha, and S. Saggar, *Biogeography and biophysicochemical traits link N_2O emissions, N_2O emission potential and microbial communities across New Zealand pasture soils.* Soil Biology and Biochemistry, 2015. **82**: p. 87-98.
31. Gilsanz, C., et al., *Development of emission factors and efficiency of two nitrification inhibitors, DCD and DMPP.* Agriculture, Ecosystems & Environment, 2016. **216**: p. 1-8.
32. Misselbrook, T., et al., *An assessment of nitrification inhibitors to reduce nitrous oxide emissions from UK agriculture.* Environmental Research Letters, 2014. **9**(11): p. 115006.
33. Morales, S.E., N. Jha, and S. Saggar, *Impact of urine and the application of the nitrification inhibitor DCD on microbial communities in dairy-grazed pasture soils.* Soil Biology and Biochemistry, 2015. **88**: p. 344-353.
34. Stempfhuber, B., et al., *Spatial interaction of archaeal ammonia-oxidizers and nitrite-oxidizing bacteria in an unfertilized grassland soil.* Frontiers in microbiology, 2016. **6**: p. 1567.
35. Soares, J.R., et al., *Enhanced-efficiency fertilizers in nitrous oxide emissions from urea applied to sugarcane.* Journal of Environmental Quality, 2015. **44**(2): p. 423-430.
36. Rotthauwe, J.-H., K.-P. Witzel, and W. Liesack, *The ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations.* Applied and environmental microbiology, 1997. **63**(12): p. 4704-4712.
37. Herbold, C.W., et al., *A flexible and economical barcoding approach for highly multiplexed amplicon sequencing of diverse target genes.* Frontiers in microbiology, 2015. **6**: p. 731.
38. Hamady, M., et al., *Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex.* Nature methods, 2008. **5**(3): p. 235.
39. Meinicke, P., *UProC: tools for ultra-fast protein domain classification.* Bioinformatics, 2015. **31**(9): p. 1382-1388.
40. Buchfink, B., C. Xie, and D.H. Huson, *Fast and sensitive protein alignment using DIAMOND.* Nature methods, 2015. **12**(1): p. 59-60.
41. Saitou, N. and M. Nei, *The neighbor-joining method: a new method for reconstructing phylogenetic trees.* Molecular biology and evolution, 1987. **4**(4): p. 406-425.

42. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets.* Molecular biology and evolution, 2016: p. msw054.
43. Felsenstein, J., *Confidence limits on phylogenies: an approach using the bootstrap.* Evolution, 1985: p. 783-791.
44. Letunic, I. and P. Bork, *Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.* Nucleic acids research, 2016. **44**(W1): p. W242-W245.
45. Martinez Arbizu, P., *pairwiseAdonis: Pairwise multilevel comparison using adonis.* R package version 0.0.1, 2017.
46. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
47. Matsuba, D., et al., *Susceptibility of ammonia-oxidizing bacteria to nitrification inhibitors.* Zeitschrift für Naturforschung C, 2003. **58**(3-4): p. 282-287.
48. Sterngren, A.E., S. Hallin, and P. Bengtson, *Archaeal ammonia oxidizers dominate in numbers, but bacteria drive gross nitrification in N-amended grassland soil.* Frontiers in microbiology, 2015. **6**: p. 1350.
49. Jia, Z. and R. Conrad, *Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil.* Environmental microbiology, 2009. **11**(7): p. 1658-1671.
50. Taylor, A.E. and P.J. Bottomley, *Nitrite production by Nitrosomonas europaea and Nitrosospira sp. AV in soils at different solution concentrations of ammonium.* Soil Biology and Biochemistry, 2006. **38**(4): p. 828-836.
51. Hink, L., G.W. Nicol, and J.I. Prosser, *Archaea produce lower yields of* N_2O *than bacteria during aerobic ammonia oxidation in soil.* Environmental microbiology, 2017. **19**(12): p. 4829-4837.
52. Theodorakopoulos, N., et al., *Increased expression of bacterial amoA during an N 2 O emission peak in an agricultural field.* Agriculture, Ecosystems & Environment, 2017. **236**: p. 212-220.
53. Liu, R., et al., *Nitrification is a primary driver of nitrous oxide production in laboratory microcosms from different land-use soils.* Frontiers in microbiology, 2016. **7**.
54. Di, H.J., et al., *Nitrous oxide emissions from grazed grassland as affected by a nitrification inhibitor, dicyandiamide, and relationships with ammonia-oxidizing bacteria and archaea.* Journal of Soils and Sediments, 2010. **10**(5): p. 943-954.
55. Meinhardt, K.A., et al., *Ammonia−oxidizing bacteria are the primary* N_2O *producers in an ammonia−oxidizing archaea dominated alkaline agricultural soil.* Environmental microbiology, 2018.
56. Kozlowski, J.A., et al., *Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota.* The ISME journal, 2016. **10**(8): p. 1836.
57. Harper, W.F., et al., *Novel abiotic reactions increase nitrous oxide production during partial nitrification: modeling and experiments.* Chemical Engineering Journal, 2015. **281**: p. 1017-1023.
58. Gruber-Dorninger, C., et al., *Functionally relevant diversity of closely related Nitrospira in activated sludge.* The ISME journal, 2015. **9**(3): p. 643.
59. Levičnik-Höfferle, Š., et al., *Stimulation of thaumarchaeal ammonia oxidation by ammonia derived from organic nitrogen but not added inorganic nitrogen.* FEMS microbiology ecology, 2012. **80**(1): p. 114-123.
60. Fu, Q., et al., *The short-term effects of nitrification inhibitors on the abundance and expression of ammonia and nitrite oxidizers in a long-term field experiment comparing land management.* Biology and Fertility of Soils, 2018. **54**(1): p. 163-172.
61. Shen, T., et al., *Responses of the terrestrial ammonia-oxidizing archaeon Ca. Nitrososphaera viennensis and the ammonia-oxidizing bacterium Nitrosospira multiformis to nitrification inhibitors.* FEMS microbiology letters, 2013. **344**(2): p. 121-129.

62. Duncan, E.G., et al., *The nitrification inhibitor 3, 4,-dimethylpyrazole phosphate strongly inhibits nitrification in coarse-grained soils containing a low abundance of nitrifying microbiota.* Soil Research, 2017. **55**(1): p. 28-37.
63. Kou, Y., et al., *Effects of 3, 4-dimethylpyrazole phosphate and dicyandiamide on nitrous oxide emission in a greenhouse vegetable soil.* Plant, Soil and Environment, 2015. **61**(1): p. 29-35.
64. Shi, X., et al., *Nitrifier−induced denitrification is an important source of soil nitrous oxide and can be inhibited by a nitrification inhibitor 3, 4−dimethylpyrazole phosphate (DMPP).* Environmental microbiology, 2017.
65. Shi, X., et al., *Effects of the nitrification inhibitor 3, 4-dimethylpyrazole phosphate on nitrification and nitrifiers in two contrasting agricultural soils.* Applied and environmental microbiology, 2016. **82**(17): p. 5236-5248.
66. Hajari, E., S.J. Snyman, and M.P. Watt, *Inorganic nitrogen uptake kinetics of sugarcane (Saccharum spp.) varieties under in vitro conditions with varying N supply.* Plant Cell, Tissue and Organ Culture (PCTOC), 2014. **117**(3): p. 361-371.

Chapter 6

Genome-resolved metagenomics of sugarcane vinasse bacteria

Noriko A. Cassman, Kesia S. Lourenco, Janaina B. do Carmo, Heitor Cantarella and Eiko E. Kuramae

Published as:

Cassman NA, Lourenco KS, do Carmo JB, Cantarella H & Kuramae EE, 2018. "Genome-resolved metagenomics of sugarcane vinasse bacteria." *Biotech for Biofuels*. 11(1): 48.

Abstract

Background: The production of 1L of ethanol from sugarcane generates up to 12 L of vinasse, which is a liquid waste containing an as-yet uncharacterized microbial assemblage. Most vinasse is destined for use as a fertilizer on the sugarcane fields because of the high organic and K content; however, increased N_2O emissions have been observed when vinasse is co-applied with inorganic N fertilizers. Here we aimed to characterize the microbial assemblage of vinasse to determine the gene potential of vinasse microbes for contributing to negative environmental effects during fertirrigation and/or to the obstruction of bioethanol fermentation.

Results: We measured chemical characteristics and extracted total DNA from six vinasse batches taken over 1.5 years from a bioethanol and sugar mill in Sao Paulo State. The vinasse microbial assemblage was characterized by low alpha diversity with 5 to 15 species across the six vinasses. The core genus was *Lactobacillus*. The top six represented bacterial genera across the samples were *Lactobacillus*, *Megasphaera* and *Mitsuokella* (Phylum Firmicutes, 35 – 97% of sample reads); *Arcobacter* and *Alcaligenes* (Phylum Proteobacteria, 0 – 40%); *Dysgonomonas* (Phylum Bacteroidetes, 0 – 53%); and *Bifidobacterium* (Phylum Actinobacteria, 0 – 18%). Potential genes for denitrification but not nitrification were identified in the vinasse metagenomes, with putative *nir*K and *nos*Z genes the most represented. Binning resulted in 38 large bins with between 36.0 and 99.3% completeness, and five small mobile element bins. Of the large bins, 53% could be classified at the phylum level as Firmicutes, 15% as Proteobacteria, 13% as unknown phyla, 13% as Bacteroidetes and 6% as Actinobacteria. The large bins spanned a range of potential denitrifiers; moreover, the genetic repertoires of all the large bins included the presence of genes involved in acetate, CO_2, ethanol, H_2O_2, and lactose metabolism; for many of the large bins, genes related to the metabolism of mannitol, xylose, butyric acid, cellulose, sucrose, “3-hydroxy fatty acids and antibiotic resistance genes were present. In total, 21 vinasse bacterial draft genomes were submitted to the genome repository.

Conclusions: Identification of the gene repertoires of vinasse bacteria and assemblages supported the idea that microbiological variation of vinasse might lead to varying patterns of N_2O emissions during fertirrigation. Furthermore, we uncovered draft genomes of novel strains of known bioethanol contaminants, as well as draft genomes unknown at the phylum level. This study will aid efforts to improve bioethanol production efficiency and sugarcane agriculture sustainability.

6.1 Introduction

Sao Paulo State contains a total of 5.7 million hectares of land planted with sugarcane. These fields supply the input for Brazil's large bioethanol industry, which is the second largest producer of bioethanol worldwide (UNICA). Brazil has more than 300 sugarcane processing plants, including sugar mills (producing only sugar), mills with distillery plants (sugar and ethanol production), and independent distilleries (only ethanol production). In the 2013/2014 season, the total ethanol production was 13.9 thousand m^3 (UNICA, 2013/2014 harvest). The major by-product of sugarcane ethanol production is vinasse; up to 12 L of vinasse is generated per liter of ethanol [1]. Sugarcane vinasse consists of water (about 93%) and organic compounds, and contains K, Ca and Mg, though the amount of these components depends on the characteristics of the input sugarcane and subsequent processing steps [2]. The major organic components of sugarcane vinasse are low molecular-weight organic compounds, mainly glycerol, lactic acid, ethanol, and acetic acid [3]. In general, vinasse has a low pH of around 4 and high chemical oxygen demand of 100 - 500 g per L.

The large volumes of vinasse and its chemical properties of high organic and K content have led to its widespread reuse as a fertilizer supplement for sugarcane crops. Most often the vinasse is sprayed onto the fields, which is a process termed fertirrigation. This method is low-cost and contributes to net energy savings in sugarcane bioethanol production cycles because the vinasse is locally transported and applied [4]. Benefits of using vinasse as fertilizer include improved short-term soil quality, crop production and crop quality [5-8]. However, negative effects include decreasing long-term soil fertility (lead leaching, N immobilization) and increasing greenhouse gas emissions, especially the emission of N_2O when used in conjunction with an N fertilizer [2, 9-12]. These effects depend on the soil and environmental characteristics as well as vinasse-specific nutrient contents (reviewed in [12]). The increased N_2O emissions from vinasse fertirrigation may be due to the stimulation of soil microbes by vinasse-derived organic material (i.e. a form of priming) or the activity of vinasse-derived cells containing genes in N_2O-producing pathways[8].

Nitrous oxide emissions are produced through two main microbially-mediated processes in soil: nitrification (NH_4^+ to NH_2OH to NO_3^-) and denitrification (NO_3^- to NO_2^- to NO to N_2O to N_2). Nitrification is carried out by microbes containing the ammonia monooxygenase enzyme, which is encoded by the gene *amo*A, and generally used as a functional marker of nitrifiers. Denitrifier bacteria may contain the nitrite reductase genes *nir*S and *nir*K, the nitric oxide reductase gene *nor*B and/or the nitrous oxide reductase gene *nor*B, which each encode for

the enzymes involved in the respiration of nitrite to nitric oxide to nitrous oxide to dinitrogen gas, respectively. The abundance of the different microbes containing denitrification genes, and the abundance of these genes when measured as functional markers, is known to correlate with the actual N_2O emission rates from soils [8]. While much is known regarding the chemical characteristics of vinasse, there are only a few indirect studies of its biotic components despite recent attention to the environmental effects of its use in fertirrigation.

The microbiota present in vinasse likely encompasses the microorganisms present in the bioethanol production process. The most common raw material for ethanol production in Brazil is the mixture of diluted molasses and cane juice, used in the distilleries annexed to sugar producing mills. The ethanol pipeline starts with crushing the unwashed sugarcane stalk to separate the sugarcane juice from the pulpy stalk residue known as bagasse. The sugarcane juice is heated and clarified with lime; the clarified juice is concentrated in an evaporator at 115 degrees C followed by vacuum boiling pan, at which point sugar and molasses crystallize. By centrifugation, the sugar crystals are separated from the mother liquor. This liquor is again crystallized in vacuum pans and then passed through continuous sugar centrifuges. The last residual solution is called molasses, which has high sucrose content suitable for ethanol production. The raw material for ethanol production is a mixture of unsterilized sugarcane juice, molasses and water [13]. The fermented material is then distilled at temperatures of at least 78 ^{0}C to separate the ethanol from the remaining waste vinasse. This vinasse is then transported via open channels or trucks to the sugarcane site for fertirrigation. The mixed sugarcane juice is fermented using proprietary *Saccharomyces cerevisiae* strains through two methods: batch (85% of distilleries as of 2011) or continuous fermentation (15%). In batch processing, the fermentation occurs in parallel, while in continuous fermentation the process occurs in series (reviewed in [14]). In either method, the yeast cells are treated with sulfuric acid, antibiotics, hop products and/or chemical biocides to reduce bacterial contamination, recovered by centrifugation, and reapplied to the fermentation tanks. This recycling step occurs between 400 and 600 times in a harvest season and despite the antibacterial treatment, bacteria remain the major contaminants.

The main bacterial contaminants of the bioethanol pipeline are lactic acid bacteria, which tend to dominate the samples taken from the ethanol pipeline in the steps prior to vinasse [15, 16]. These bacteria, in particular Lactobacillus species, compete with the commercial yeast strains for sugar or form exopolyssacharides that floccule yeast cells [17-19]. Contamination by bacteria – through sucrose competition, flocculation of the commercial yeast strain or fer-

mentation inhibition – can lower the efficiency of the bioethanol process by up to 30% [16, 20]. Furthermore, because of the antibiotic treatment of the yeast cells during the recycling step, contaminant bacteria may be a source of antibiotic resistance genes, as has been recently reported in a field study [21]. Other sources of contamination are wild yeast strains from the input sugarcane stalks, which are not sterilized prior to the production pipeline [22]. To date, no studies have investigated the presence of bioethanol pipeline contaminants in vinasse.

Here we investigated concurrently the chemical and microbial properties of vinasse in order to characterize the vinasse assemblage. We explored metagenomic data taken from vinasse samples over 1.5 years of production from a bioethanol mill in Piracicaba, SP, Brazil. The mill processes sugarcane produced in the region within a rough 40 km radius. Vinasse is distributed by trucks for fertirrigation during the harvest season (April to November). To characterize the microbial assemblage of this vinasse, we sequenced total DNA from six vinasse samples. We analyzed the resulting 18 shotgun metagenomes through metagenomics and differential abundance binning. To investigate the potential for N_2O emissions from fertirrigation with vinasse, special attention was given to sequences and reconstructed genomes annotated as genes involved in N_2O-related metabolism. Furthermore, we also identified genes relating to bioethanol production concerns to identify future research directions. To date this is the first culture-independent study of the vinasse microbial assemblage. Our main questions were (1) what are the overall and sample-wise taxonomic and functional characteristics of the vinasse microbial assemblages? and (2) what is the potential of the vinasse microbes for N_2O emissions, obstruction of fermentation and/or antibiotic resistance?

6.2 Materials and Methods

6.2.1 Sampling description

The bioethanol mill from which we sampled is in the region of Piracicaba in SP, Brazil. The mill takes in sugarcane from the region and produces sugar and ethanol. We obtained six time points of vinasse taken from transport trucks prior to their departure to the fields for chemical and molecular analyses. The trucks hold about 10,000 L of vinasse. Prior to sampling, the vinasse was held in the trucks for 24 hours. Of the vinasse, 0.5 L sampled from the truck was immediately kept at 4 degrees C. The six sampling dates were 13/11/2013 (A, Nov. 2013), 13/12/2013 (B, Dec. 2013), 15/07/2014 (C, July 2014), 15/08/2014 (D, Aug. 2014), 14/10/2014 (E, Oct. 2014) and 10/11/2014 (F, Nov. 2014). The dates of the vinasse sampling corresponded to summer (October, November and December) or

winter (July and August) sugarcane harvests. Because each vinasse was a random assemblage of contaminants from the bioethanol process, we considered each time point an independent measure for statistical analysis.

6.2.2 Chemical analyses, DNA extraction, and qPCR quantification and sequencing

For each vinasse sample, 500 ml was used for chemical analyses. The remaining three subsamples of 100 ml per time point were used for DNA extraction. First, the samples were centrifuged at 10,621 x g (Sigma 2-16P) for 10 min to separate cells from the liquid. Total DNA was extracted from the pellets with the MoBio PowerSoil kit according to the manufacturer's instructions. Between 553 and 5310 ng was sent for sequencing (**Additional file 1**). The DNA was prepared as a MiSeq Illumina paired-end library and sequenced (3 replicates x 6 samples = 18 metagenomes) or used for quantitative PCR of genes that encode for the enzymes involved in the sequential biochemical steps leading to N_2O production (*amo*A, *nir*K, *nir*S, *nor*B) or removal (*nos*Z). The qPCR reactions were performed in a 96-well plate (Bio-Rad) using CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). The qPCR reaction, primers combinations and thermal cycler conditions of each gene amplification are listed in **Additional file 2.** The qPCR data was acquired at 72 °C and melting curve analysis was performed to confirm specificity. Amplicon sizes were checked by agarose gel electrophoresis. Samples were analyzed with two technical replicates. Reaction efficiency varied from 80 to105% and R^2 values ranged from 0.94 to 0.99.

6.2.3 Metagenome processing and read-based sample comparisons

Bioinformatics processing was performed on a Linux server (Linux-3.13.0-76-generic-x86_64-with-Ubuntu-14.04-trusty) with 64 nodes and 250 GB RAM. Processing was performed in a Snakemake v3.7.1 workflow or with bash or perl scripts (available upon request). The 18 shotgun metagenomes were checked for tag sequences and evaluated for statistics using FastQC v0.10.1 (Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and PRINSEQ-lite version 0.20.4 [23]. Raw reads were filtered out using PRINSEQ if they had more than 1% of ambiguous (N) characters, had a mean quality score of less than 25 or were exact duplicates. Reads were trimmed at the 3'end if the mean quality score was less than 20 within a sliding window size of 10 (clean reads). The clean paired-end reads were used in further analyses unless otherwise noted. The raw paired-end reads were merged using PEAR v0.9.5; these merged

read ends were trimmed by quality and filtered out if the merged read had more than 1% ambiguous characters (parameters: -q 20 –n 0.01) with PEAR (merged reads) [24]. For downstream normalization of annotation counts, calculations of average genome size per sample were carried out using MicrobeCensus [25]. To compare the metagenomes directly, sample distances were determined from the partial de Bruijn assembly of the clean forward reads using MetaFAST 0.1.0 (revision 57253a1) [26].

6.2.4 Taxonomy, phylogeny and alpha diversity

To characterize the taxonomic composition, functional potential and diversity of the microbial assemblages in the vinasse samples, we profiled the metagenomes using different databases. First, the merged reads were uploaded to the metagenome analysis platform MG-RAST version 3.6 [27]. The metagenomes were compared using the default presets to the RefSeq or Subsystems databases to obtain taxonomic or functional profiles, respectively. Refseq annotations, including eukaryota, bacteria, archaea and viruses, were determined using the last common ancestor approach. The MG-RAST taxonomic (phylum-level) and functional (Subsystems Level 1) profiles were analyzed with the Statistical Analysis of Metagenome Profiles (STAMP) software [28]. Taxonomic or functional level abundances significantly different among vinasse samples were evaluated using ANOVA. The Tukey-Kramer post-hoc test with a 95% confidence interval and the Benjamini-Hochberg correction was used to identify differing phyla or Subsystems Level 1 category abundances between the vinasse metagenomes with significance determined at corrected $p<0.001$ or 0.05, respectively. The taxonomic profiles at genus level were kept to visualize the relative abundance of genera across samples.

Because the metagenomes were well-represented in the MG-RAST databases, we further characterized the taxonomy and functional potential of the metagenomes using metaphlan2 version 2.6.0 and humann2 version 0.9.9 pipelines [29,30]. For metaphlan2 analysis, we used the “relab” analysis with the “--ignore_eukaryotes” flags to obtain taxonomic profiles. To gain an overall view of the taxonomy present in the vinasse samples and the phylogenetic relationships between the species in the samples, the average taxonomic distributions of the vinasse samples from metaphlan2 were visualized as a cladogram using Graphlan [31]. To examine the taxonomic profiles of vinasse across samples, these were visualized through heatmaps with average linkage clustering on Euclidean distances using hclust2. For the humann2 analysis, we annotated the forward clean reads against the UniRef90 database [32]. Pathway abundances were visualized exclud-

ing the “UNMAPPED” and “UNKNOWN” categories using hclust2 heat maps with average linkage clustering on Euclidean distances. To obtain a measure of alpha diversity, we ran metaphlan2 with previously mentioned flags on samples rarified to the smallest library size (280,161 reads).

To infer the phylogenetic relationships between the organisms present in the vinasse samples, full-length 16S rRNA genes were recruited from the vinasse metagenome reads using REAGO version 1.1 on forward clean reads truncated to 201 bp [33]. The resulting full-length 16S rRNA vinasse sequences were aligned and taxonomically classified against the SSU 128 SILVA reference database using SINA [34,35]. The five nearest neighbors for each full-length 16S rRNA sequence were downloaded in addition to two *Verrucomicrobia* outgroup sequences. The 16S rRNA sequences were aligned without gaps using ClustalW in MEGA7 (121 sequences in total)[36]. A neighbor-joining tree was created with evolutionary distances computed using the Maximum Composite Likelihood method [37,38]. Phylogenetic distances were evaluated with bootstrap tests (1000 replicates) [39]. To obtain a measure of alpha diversity we recruited full-length 16S rRNA genes using REAGO as above on the rarified metagenomes. Further, we evaluated a measure of genus-level relative abundance across samples by mapping the metagenome reads to the extracted 16S sequences grouped by taxonomic affiliation using bowtie2. These abundances were calculated as percentages of the number of aligned pairs from the total number of metagenome reads per sample.

6.2.5 Putative denitrification and nitrification gene abundances

To investigate the potential for N_2O emissions from the vinasse samples, we used two approaches: 1) metagenome read matching to profile HMMs of denitrification and nitrification genes and 2) recruitment of denitrifying and nitrifying genes from the reads. Profile HMMs for the *amo*A_*AOA, amo*A_*AOB, nir*K, *nir*S, *nor*B, *nosZ, nosZ_atypical_1* and *nosZ_atypical_2* genes were downloaded from the Functional Gene Repository (FUNgene version 8.3; available at http://fungene.cme.msu.edu/). Reads were translated to protein sequences with the “meta” setting using Prodigal version 2.6.2. The ORFs were queried for HMM matches using HMMsearch (command: hmmsearch --noali -o <filename.fasta> <gene.hmm> <filename.fasta>; available at https://www.ebi.ac.uk/Tools/hmmer/search/hmmsearch). The HMM matches were normalized to reads per kilobase per genome equivalent (RPKG = (# mapped reads / HMM gene length (KB)) / genome equivalents). The RPKG normalization accounts for genome size, library size and gene length biases, allowing for gene and sample comparisons.

In parallel, the gene-targeted assembler pipeline megagta version 0.1_alpha was used to recruit full-length genes from the metagenomes [33,40]. Gene-targeted assemblies (i.e. recruitments) were carried out on *amo*A_AOA, *amo*A_AOB, *nir*S, *nir*K, *nor*B_*cNor*, *nor*B_*qNor*, *nos*Z and *nos*Z_a2 genes using megagta. Further, to infer alpha diversity, the ribosomal *rpl*B gene was recruited from the rarified metagenomes.

6.2.6 Cross-assembly and binning

We evaluated the performance of three assemblers (Ray-meta [41], Megahit [42] and metaSpades[43]) in cross-assembling the 18 vinasse metagenomes; the best cross-assembly was that from the metaSPADES assembler version 3.8.2 based on assembly characteristics evaluated using MetaQUAST (QUAST Version 3.0, build 07.07.2015 05:57 [44]). The 18 metagenomes were cross-assembled with metaSpades using kmer sizes 77, 99 and 127. The sample reads were mapped to the cross-contigs using bowtie2 to obtain cross contig abundances [45] . The final metaSPADES cross-assembly was binned using three tools for comparison: CONCOCT (with anvio version 2.3.2), Metabat [46] and MaxBin2 version 2.1.1 [47]. The contig annotation tool (CAT version 2) was used to determine the taxonomic affiliation of all ORFs identified in each bin using prodigal to find ORFs and diamond blastp against the NCBI-nr database [48]. CAT taxonomy results were formatted using custom Perl scripts and visualized with TreeMap to aid with the taxonomic characterization of the bins. Because more genomes with >90% completeness and coherent taxonomies were found from the MaxBin2 binning, these were selected for downstream analysis. CheckM was used to check the original MaxBin2 bins [49]. These bins were manually refined using anvi'o version 2.3.2 based on cross-contig taxonomy (from CAT), hierarchical clustering of the cross-contigs and sample coverage information [50]. The relative sample abundances of the bins were noted as the percent of sample reads recruited to the bin out of the total sample reads recruited to all the bins (i.e. percent recruitment anvi'o results).

The "good bins" were identified as having >90% completeness and <10% redundancy. Further "interesting bins" were further identified as those with >20% completeness and <10% redundancy and/or coherent contig taxonomies. Functional annotation of the "good and interesting bins" were carried out using prokka with the "kingdom" flag set to bacteria or viruses depending on the taxonomic classification [51]. To characterize the bins by their potential functional type, prokka annotation results were mined for lines matching EC numbers of KEGG enzymes of compounds related to bioethanol production interests and N_2O emis-

sions. These KEGG compounds were acetate (C00033), cellulose (C00760), xylose (C00181), lactose (C00242), caprylic acid (C01585, carbon dioxide (C00011), diacetyl (C00741), hydrogen peroxide (CC00027), lactaldehyde (C05999) and phenyllactate (C05607). The lists of EC numbers were obtained by querying the KEGG REST API on each compound ID. Keyword searches of "3-hydroxy" fatty acids, "cyclic dipeptide," antibiotic "resistance," and nitrification and denitrification genes were additionally used to identify the potential presence of these functions in the bins.

In parallel, to confirm potential denitrification and nitrification gene presence, bin sequences were compared to HMMs of nitrification and denitrification genes from FunGene as described previously but with the prodigal setting "single." The HMM matches were normalized by bin size (number of ORFs and total number of bp in ORFs) and HMM length in bp.

6.3 Results

6.3.1 Vinasse chemical characteristics and metagenome overview

The chemical characteristics of the vinasse samples are listed in **Table 1**. Average pH was low at 4.4±0.4, ranging between 3.9 (D) and 4.8 (C). Total organic carbon averaged 29±1.8 g L^{-1} and ranged between 25.7 (B) and 31.4 g L^{-1} (D). Total N averaged 0.64±0.15 g L^{-1}, while that of P and K was 0.16±0.07 and 3.43±1.02, respectively. The C/N ratio averaged 42±13 and ranged between 19 (F) and 57 (C). After processing, the 18 vinasse metagenomes contained a total of 2,126 Mbp distributed into 7.82 million reads. The number of reads ranged between 280,161 and 542,208 sequences per sample with between 77 and 150 Mbp (**Additional file 1**). When the metagenome distances were compared using partial de Bruijn assembly, A and C were most similar, followed by F, followed by E; least similar were B and last D (**Additional file 3**).

Table 1. Chemical characteristics of the six vinasse samples.

Group Name	Sampling Date	pH	C org	N tot	$N\text{-}NH_4^+$	$N\text{-}NO_3^-$	P	K	C:N
			g L^{-1}	g L^{-1}	mg L^{-1}	mg L^{-1}	g kg^{-1}	g kg^{-1}	
A	Nov. 2013	4.7	28.2	0.53	65.8	17.6	0.08	2.9	53
B	Dec. 2013	4.1	25.7	0.53	63.4	10.8	0.17	2.6	49
C	July 2014	4.8	28.8	0.51	45.7	8.8	0.11	3.5	57
D	Aug. 2014	3.9	31.4	0.89	41.6	4.1	0.23	4.7	35
E	Oct. 2014	4.2	29.6	0.74	37.7	6.8	0.10	2.1	40
F	Nov. 2014	4.7	30.3	1.57	75.9	6.6	0.25	4.8	19

6.3.2 Taxonomic characterization

When compared to the M5NR database containing eukaryota, bacteria, archaea and viruses on MG-RAST (**Additional file 4**), 21 to 55% of the merged reads could be classified. Of the classified reads, 96 to 100% were annotated as bacteria. The top phyla present in the vinasse samples with relative abundances greater than 1% and/or that significantly co-varied among the samples (ANOVA at $p<0.001$ and Kruskal-Wallis post-hoc test) were Firmicutes (35 to 97% of merged reads), Bacteroidetes (0.8 to 53%), Actinobacteria (0.4 to 17.5%) and Proteobacteria (0.3 to 39.4%; **Additional file 5**). The "core" phylum observed in all vinasse samples was Firmicutes. Similarly, when compared to the metaphlan2 marker gene database containing bacteria, archaea and viruses (excluding eukaryotes), between 68 and 100% of classified reads were identified as bacteria and 0 to 32% as viruses (**Figure 1**). The previous four main bacterial phyla again dominated the vinasse samples: Firmicutes (48 to 100% of classified reads), Actinobacteria (0 to 19%) and Proteobacteria (0 to 18%), as well as viruses (0 to 32%; **Figure 2**). The most abundant bacterial genera were Lactobacillus (Phylum Firmicutes), Megasphaera (Firmicutes), Mitsuokella (Firmicutes) and Bifidobacterium (Actinobacteria). Further supporting these taxonomic results, the full-length 16S rRNA genes recruited from the vinasse metagenomes were classified as Bifidobacterium (Phylum Actinobacteria), Olsenella (Phylum Actinobacteria), Prevotella (Phylum Bacteroidetes), Lactobacillus (Phylum Firmicutes), Megasphaera (Phylum Firmicutes), Mitsuokella (Phylum Firmicutes) and *Comamonas* (Phylum Proteobacteria) genera (**Additional file 6 and 13**).

When the samples were clustered based on the MG-RAST taxonomic profiles at phylum level, E and C formed a cluster while A, F, and D were separated based on the first principal component and B was separated based on the second (**Additional file 7**). When the metaphlan2 profiles were clustered at the level of class, order, family and genus, samples A, C and F formed a cluster while B, D and E formed a separate cluster (**Figure 2**).

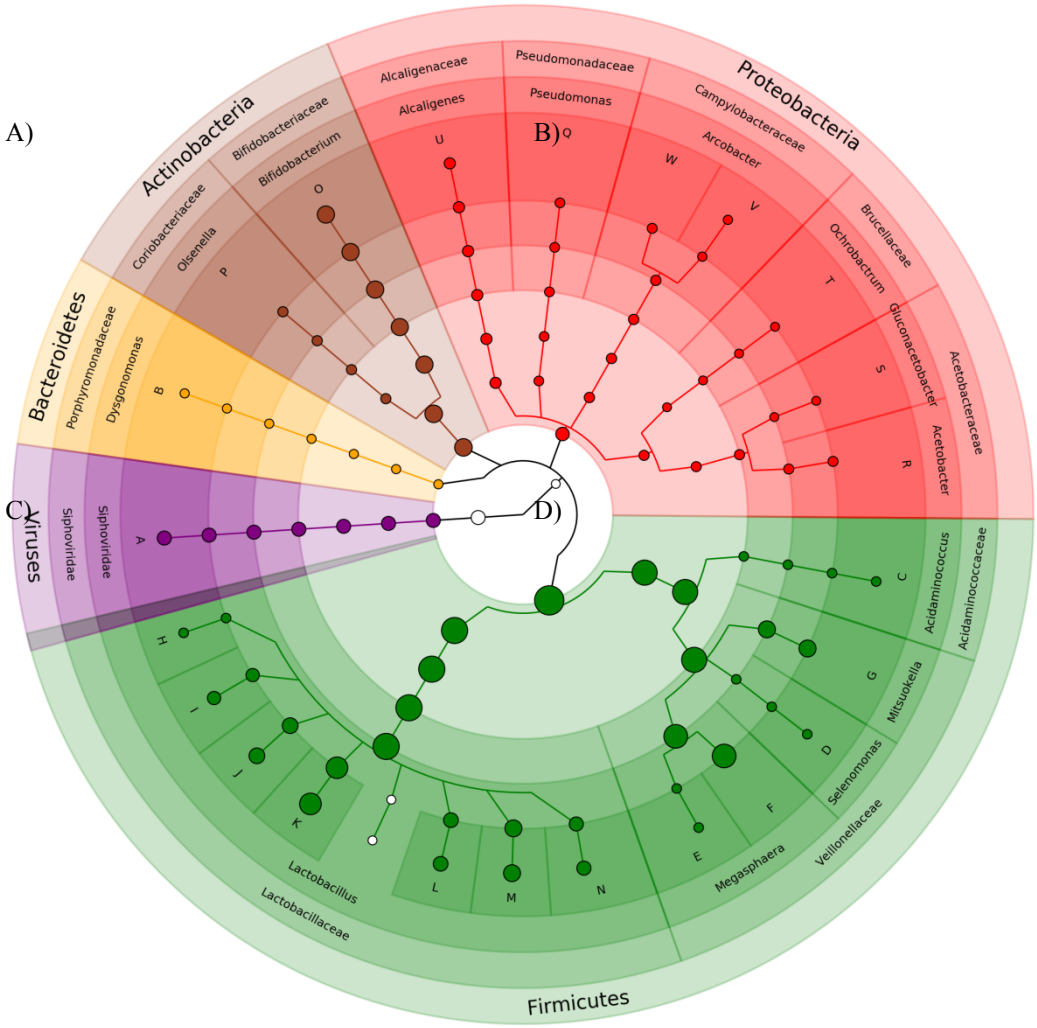


Figure 1. Average abundance of taxa in the vinasse samples. The metagenomes were analyzed using metaphlan2 and visualized with GraPhlan. Node sizes correspond to average relative abundance across the vinasse metagenomes while colors correspond to phylum. Species are noted with letters: A=*Lactobacillus* phage Lc Nu, B=*D. mossii*, C=*A. intestini*, D=*S. bovis*, E=*M. elsdenii*, F=*Megasphera* unclassified, G=*Mitsuokella* unclassified, H=*L. salivarius*, I=*L. equicursoris*, J=*L. delbrueckii*, K=*L. amylovorus*, L=*L. mucosae*, M=*L. fermentum*, N=*L. vini*, O=*B. thermophilum*, P=*Olsenella* unclassified, Q=*Pseudomonas* unclassified, R=*Acetobacter* unclassified, S=*Gluconacetobacter* unclassified, T=*Ochrobactrum* unclassified, U=*A. faecalis*, V=*A. butzleri* and W=*Arcobacter* unclassified.

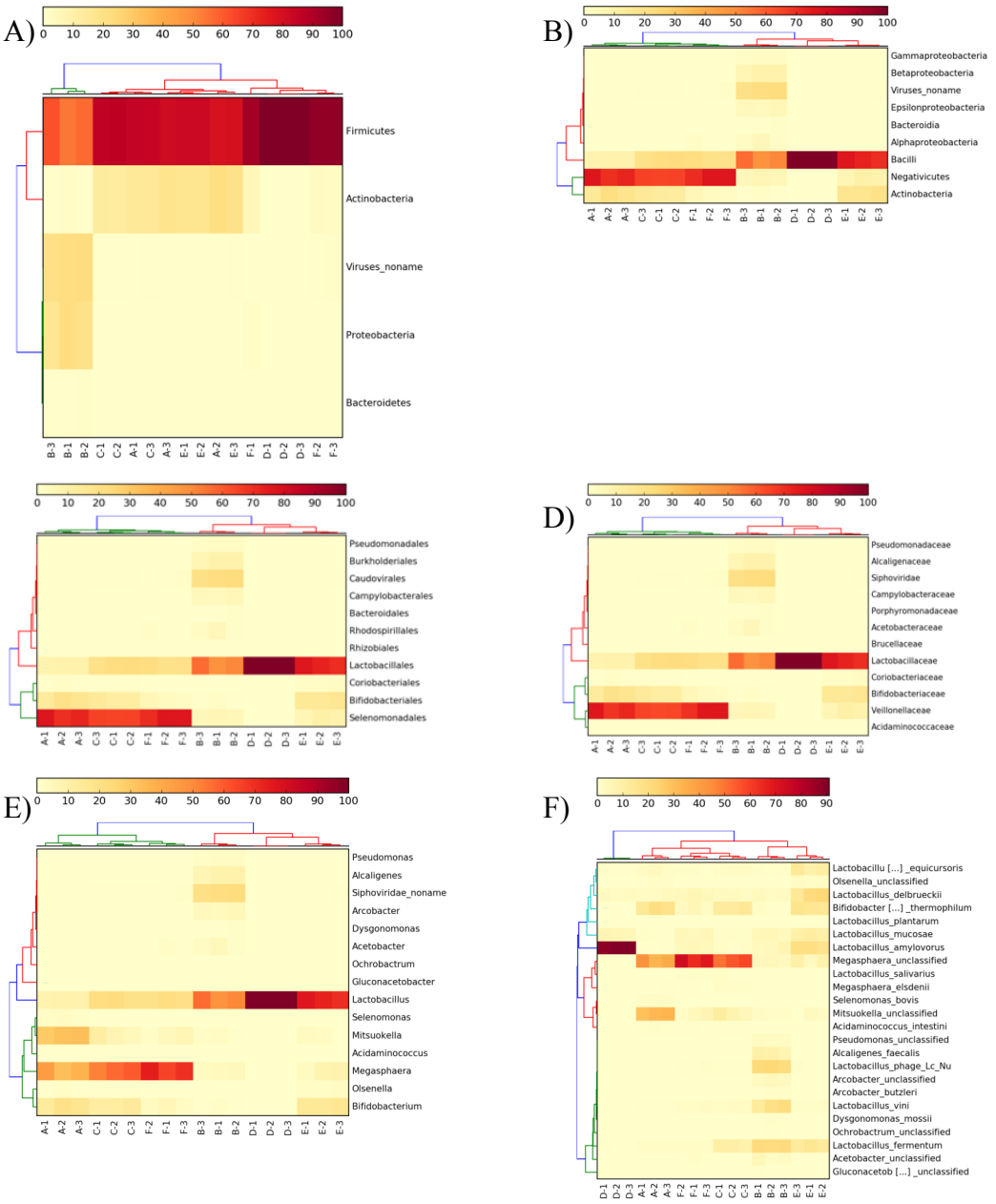


Figure 2. Taxonomic distributions across the vinasse samples at the level of A) Phylum, B) Class, C) Order, D) Family, E) Genus and F) Species. The taxonomic group and sample profiles were clustered using hclust2 from metaphlan2 results.

6.3.3 Functional potential characterization

When compared to the M5NR databases through MG-RAST, the percentage of sequences with ORFs that could be classified into functional categories ranged between 16 and 42% (**Additional file 8**). At Subsystems Level 1, the top significantly different categories were Carbohydrate metabolism, Clustering-based subsystems, Amino Acids and Derivatives, Miscellaneous, Protein metabolism, DNA metabolism, RNA metabolism, Cofactors/vitamins, Cell wall/capsule, Phages/ prophages and Nucleosides and Nucleotides. When sample distances were determined using the functional profiles at Subsystems Level 1, C, A, B and F formed a cluster while E was separated based on the first principal component and D was separated based on the second (**Additional file 9**). When the vinasse metagenomes were analyzed using the humann2 framework, abundant pathways were found in sample D, which was dominated by one *Lactobacillus* – the top abundant pathways included PWY-5100: pyruvate fermentation to acetate and lactate II and PWY-7219: adenosine ribonucleotides de novo biosynthesis (**Additional file 10**). Combining the real-time PCR, gene recruitment and gene mapping results, the vinasse metagenomes had few to no genes matching nitrification genes; in contrast, a range of denitrification genes was found (**Figure 3**). Sample B presented the most diversity of denitrification genes, with *nir*K, *nir*S, *nor*B and *nos*Z present based on the recruitment and mapped results. The presence of putative *nos*Z was supported in all samples except D. In addition, putative *nir*K was found in all samples except F.

6.3.4 Alpha diversity of the vinasse samples

Several methods were employed to obtain estimates of the alpha diversity of the vinasse samples (**Table 2**). The normalized effective number of species from MG-RAST averaged 29±14 and ranged between 3 (D) and 53 (B) species. When metaphlan2 was applied to the rarified vinasse samples, the number of classified species averaged 11±3 and ranged between 5 (D) and 14 (B) species. Partial 16S rRNA fragments recruited from the rarified samples using REAGO averaged 10±4 and ranged between 4 (D) and 17 (E). Further, when the *rpl*B gene was recruited from the rarified samples with megaGTA analysis, and average of 17±3 fragments could be found across the vinasse samples with between 13 (E) and 22 (C) *rpl*B fragments identified. When the 16S rRNA gene was amplified using real time PCR of the vinasse samples, the number of genes per kg of dry matter averaged 12e12±9e12 and ranged between 0.8 (E) and 25.7 (A); the gene abundance of 18S rRNA gene averaged 100e3±71e3 and ranged between 17e3 (D) and 208e3 (B).

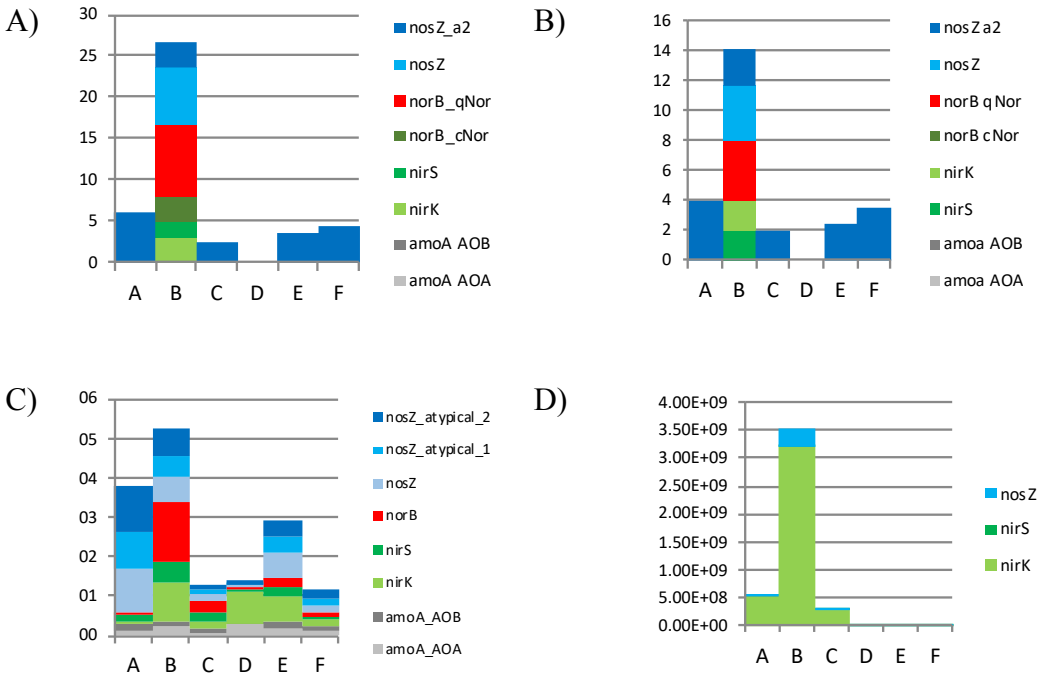


Figure 3. Putative gene abundances in the vinasse metagenomes**.** Partial gene fragments were recruited from the vinasse metagenomes using megagta on A) all reads and B) rarified reads. In parallel, vinasse metagenomes were compared to profile HMMs and the number of matches was normalized to C) reads per kilobase per genome equivalent (RPKG). In D) the gene copy numbers from real time PCR of the *nos*Z, *nir*S and *nir*K genes are depicted. Note that no qPCR of the *norB* gene was made.

Table 2. Alpha diversity estimates of the vinasse samples. Diversity was quantified by the number of partial genes recruited (REAGO and megaGTA), or the estimated number of species (metaphlan2 and MG-RAST) from the vinasse metagenomes; results from real-time PCR of the 16S gene was also included. Rarified forward reads were used as input for metaphlan2, reago and megagta analysis; merged reads were used in the MGRAST analysis and these results were normalized by library size.

	REAGO	megaGTA	metaphlan2	MGRAST	qPCR
Sample Name	# recruited 16S rRNA genes	# recruited rplB genes	# Species	Effective # species	# 16S rRNA copies (/ 1000000) kg dry matter^{-1}
A	13 ± 2	21 ± 2	10 ± 1	37 ± 1	25750 ± 13900
B	10 ± 3	16 ± 3	14 ± 1	47 ± 4	16839 ± 11664
C	12 ± 1	22 ± 2	13 ± 0	38 ± 1	16281 ± 1104
D	4 ± 0	15 ± 2	5 ± 0	3 ± 0	10749 ± 3336
E	17 ± 2	13 ± 1	10 ± 0	20 ± 1	839 ± 840
F	6 ± 2	17 ± 1	12 ± 1	29 ± 3	1135 ± 1142

6.3.5 Bin characteristics, taxonomy and functional types

The cross-assembly resulted in 221,975 cross-contigs totaling 216 Mbp. Of the cross-contigs, 40,815 were longer than 1Kbp, and 40,186 of these could be binned. After refining the bins, 20,825 cross-contigs remained distributed within the 36 good or interesting large bins (0.6 to 3.9 Mbp; hereafter referred to as the large bins). The large bins represented 39 to 68% of the sample reads. Fifty-eight percent of the large bins were classified at the phylum level as Firmicutes, 8% as Bacteroidetes, 17% as Proteobacteria, 11% as Unknown and 6% as Actinobacteria (**Table 4**). Overall, the GC percent of these bins ranged between 28 and 66%. Of the large bins, 24 were potential denitrifiers and three potential nitrifiers. The presence of genes related to acetate, CO_2, ethanol, H_2O_2 and lactose metabolism were found in all large bins while the potential presence of genes related to Lactaldehyde, mannitol, xylose, butyric acid, cellulose, diacetyl, phenyllactate, sucrose and "3-hydroxy" was variable across the large bins (**Table 5**). Last, when multidrug resistance was identified in the bin annotations, all large bins but Unknown-19 and Lactobacillus-30 contained these genes. In addition to the large bins, eight small bins (0.03 to 0.20 Mbp) lacking bacterial marker gene presence were found (**Table 3** and **Additional files 11 and 12**). The largest of the small bins, 4.2 and 8.1 were most abundant in samples E and D, respectively.

Table 3. The "good and interesting" vinasse bin characteristics and relative sample abundances (indicated by heatmap per sample).

Bin Ids	A	B	C	D	E	F	Length (Mbp)	# Contigs	N50	GC (%)	Completeness	Redundancy
1	6	1	2	0	0	1	2.37	209	23171	53	92	2
2	14	2	1	0	4	3	3.45	547	10519	49	94	4
3	8	1	18	0	2	13	2.19	352	10534	53	97	2
4.1	0	0	0	1	8	0	0.05	19	3352	37	0	0
4.2	0	0	0	0	5	0	0.14	14	32202	29	0	0
5	3	0	4	0	2	1	1.91	298	11078	60	94	3
6	4	1	2	0	1	12	2.42	539	6347	44	91	2
8.1	0	0	0	1	5	0	0.20	39	52952	36	0	0
8.2	0	1	0	9	1	0	0.03	17	1553	40	0	0
8.3	0	0	1	0	3	0	0.07	27	2920	47	0	0
8.4	0	1	0	0	0	0	0.02	10	2442	41	0	0
8.5	0	0	0	0	1	0	0.03	13	3198	39	0	0
8.6	0	0	0	1	1	0	0.03	16	3273	45	1	0
9	2	0	2	0	1	0	1.96	407	6384	60	90	6
10	1	1	1	1	3	0	2.07	183	27583	47	96	1
12	3	0	2	0	5	1	2.34	485	7489	66	91	5
13	2	0	2	0	0	0	1.48	605	2710	63	71	12
14	2	0	1	0	0	1	1.88	947	2190	52	74	15
15	2	0	2	0	1	1	2.16	1017	2297	53	76	9
16	1	1	1	0	0	13	3.02	307	22370	42	99	1
18	2	0	1	0	2	0	1.48	893	1665	63	69	22
19	1	0	1	0	3	0	2.01	917	2351	62	60	12
20	0	0	1	0	1	1	1.16	387	3766	54	64	2
21	0	0	1	1	2	1	1.78	298	9992	50	88	1
23	0	1	1	3	2	1	1.90	373	7041	47	95	4
24	0	2	1	0	1	0	1.78	220	13204	53	98	4
25	1	1	1	1	0	2	1.75	729	2822	40	91	9
26	1	1	2	0	1	1	2.12	1236	1733	41	66	15
27	0	1	1	45	3	1	1.94	262	11858	38	99	1
28	0	3	0	0	0	0	2.11	340	8670	38	96	5
29	0	1	0	2	0	0	1.02	439	2658	50	88	9
30	0	3	1	3	2	3	3.94	2021	1897	31	47	16
31	0	0	8	0	0	0	3.29	1595	2241	54	79	37
32	0	0	0	6	0	0	2.00	104	208993	36	99	1
33	0	0	0	1	3	0	1.71	447	4850	48	96	9
34	0	3	0	0	0	0	2.68	343	12289	60	92	1
35	0	4	0	0	0	0	2.72	259	13750	43	96	6
36	0	5	0	0	0	0	1.70	647	2989	49	67	5
37.1	0	10	0	0	0	0	3.08	1326	2603	57	77	24
37.2	0	4	0	0	0	0	1.22	784	1533	58	34	6
38	0	9	0	0	0	0	2.99	488	9382	60	89	3
39	0	0	0	4	0	0	1.90	205	15510	47	99	4
40.1	0	4	0	0	0	0	1.61	190	11919	28	97	2
40.2	0	2	0	0	0	0	0.57	271	2044	27	15	1

Table 4. Taxonomy of the "good and interesting" vinasse bins based on CAT classification.

Bin Id	K	Phylum	Class	Order	Family	Genus	Species
1	B	F	Nega	Selenomonadales	Veillonellaceae	Mitsuokella	U
2	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	P. mutisaccharivorax/Unclassified
3	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera	U/M. elsdenii
4.1	V				Caudovirales	Siphovirales	Lactobacillus phage Ldl1
4.2	B	U/B	U/B	U/B	U	U	U
5	B	A	Acti	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	U
6	B	B	Bact	Bacteroidales	Prevotellaceae/U	Prevotella/U	U
8.1	B/V	U	U	U	U/Caudovirales	U/Siphoviridae	U/Lactobacillus phage LdL1
8.2	B	F	B	Lactobacillales	Lactobacillaceae	Lactobacillus	U
8.3	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	P. histicola/U
8.4	B	F	Bac	Lactobacillales	Lactobacillaceae	Lactobacillus	U
8.5	B	B	U/Bact				
8.6	B	F	Bac	Lactobacillales	Lactobacillaceae	Lactobacillus	U
9	B	U/A					
10	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. equicursoris/Unclassified
12	B	A/U					
13	B	U/A					
14	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera	U/sp. DJF_B143
15	B	F	Nega	Selenomonadales	Veillonellaceae	Dialister	U/D. succinatiphilus
16	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	U
18	U/B						
19	B	U/A					
20	B	F	Clos	Clostridiales	Eubacteriaceae/U	Pseudoramibacter/U	U/P. alactolyticus
21	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. delbrueckii/U
23	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. mucosae/U
24	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. fermentum/U
25	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U/L. vini
26	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U/L. agilis
27	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U
28	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. vini
29	B	F	Clos	Clostridiales	Clostridiaceae	Clostridium	sp. CAG:568/U
30	B/A	F/E	Bacilli/Metha	Bacteroidales/Methanobacteriaceae	Lactobacillaceae/Methanobactericeae	Lactobactillus/Methanobrevibacter	U
31	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera/U	U
32	B	F	Nega	Lactobacillales	L	L	U
33	B	F	B	Lactobacillales	L	L	L. secaliphilus/U
34	B	F	B	Lactobacillales	L	L	U/L. manihotivorans
35	B	F	B	Lactobacillales	L	L	L. bifermentans/U
36	B	P	Beta	Burkholderiales	Alcaligenaceae	U	U
37.1	B	P	B	Burkholderiales	Alcaligenaceae	Alcaligenes	A. faecalis/U
37.2	B	P	B	B	A	Alcaligenes	U/A. faecalis
38	B	P	B	B	Comamonadaceae	U/Comamonas	U/C. kerstersii

K=Kingdom, F=*Firmicutes*, B=*Bacteroidetes*, A=*Actinobacteria*, P=*Proteobacteria*, E=*Euryarchaeota*, U=Unknown, Bact=*Bacteroidia*, Nega=*Negativicutes*, Clos=*Clostridia*, Mega=*Megasphaera*, Lact=*Lactobacillales*, Metha=*Methanobacteria*

Table 5. Putative gene repertoires of the large vinasse bins. Keyword searches of prokka annotation results (grey, "Y") were supplemented in the case of the N_2O metabolism-related genes with hmm profile search results (colors). Substrates for the genes related to N_2O metabolism are included.

																NO_2					
Bins	Acetate	CO_2	Ethanol	Lactaldehyde	Mannitol	Xylose	Butyric acid	Cellulose	Diacetyl	H_2O_2	Lactose	Phenyllactate	Sucrose	3-hydroxy	ABR	*nirK*	*nirS*	*norB*	*nosZ*	*nosZ_1*	*nosZ_2*
1	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
2	Y	Y	Y		Y	Y	Y	Y		Y	Y		Y	Y Y						Y	Y
3	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y Y							
5	Y	Y	Y					Y		Y	Y		Y	Y							
6	Y	Y	Y					Y		Y	Y		Y	Y Y							
9	Y	Y	Y		Y	Y		Y		Y	Y		Y	Y Y							
10	Y	Y	Y	Y	Y	Y		Y		Y	Y		Y	Y Y							
12	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
13	Y	Y	Y		Y			Y		Y	Y		Y	Y							
14	Y	Y	Y		Y			Y		Y	Y			Y Y							
15	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
16	Y	Y	Y		Y	Y	Y	Y		Y	Y		Y	Y Y							
18	Y	Y	Y		Y					Y	Y			Y							
19	Y	Y	Y					Y		Y	Y	Y	Y								
20	Y	Y	Y	Y						Y	Y			Y Y							
21	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
23	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y		Y	Y Y							
24	Y	Y	Y			Y	Y	Y		Y	Y		Y	Y Y							
25	Y	Y	Y	Y	Y	Y	Y	Y		Y	Y		Y	Y Y							
26	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y Y							
27	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
28	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y Y							
29	Y	Y	Y					Y		Y	Y			Y							
30	Y	Y	Y					Y		Y	Y	Y	Y								
31	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y							
32	Y	Y	Y		Y			Y		Y	Y	Y	Y	Y Y							
33	Y	Y	Y		Y	Y		Y	Y	Y	Y		Y	Y Y							
34	Y	Y	Y		Y		Y	Y		Y	Y	Y	Y	Y Y							
35	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y Y							
36	Y	Y	Y		Y			Y		Y	Y		Y	Y Y							
37.1	Y	Y	Y		Y		Y	Y		Y	Y		Y	Y Y							
37.2	Y	Y	Y					Y		Y	Y		Y	Y Y		Y				Y	
38	Y	Y	Y		Y			Y		Y	Y	Y	Y	Y Y							
39	Y	Y	Y		Y	Y	Y	Y		Y	Y		Y	Y Y							
40.1	Y	Y	Y		Y			Y		Y	Y			Y Y							
40.2	Y	Y	Y					Y		Y	Y			Y							

ABR = antibiotic resistance; Notes: no genes related to the metabolism of caproic acid were found in the bin annotations. No *amo*ABC, *hao*, *nxr nor*, *nir*S genes were found in the bin annotations, but the *amo*A AOA gene was identified in Bin 23 and 40.1 and the *amo*A AOB gene was identified in Bin 33 by HMM matches.

6.4 Discussion

Here, we explored concurrently the chemical and microbiological characteristics of vinasse produced over 1.5 years from one bioethanol mill in Sao Paulo State. The aims were to characterize, for the first time, the taxonomy and potential functions of the microbial assemblage in vinasse; we further recovered draft genomes from vinasse bacteria. We combined metagenomic analyses with binning techniques to characterize the vinasse assemblages and bacteria, respectively. We discuss below both potential ethanol pipeline contamination traits of vinasse bacteria and the potential ecology of vinasse fertirrigation. The vinasse chemical characteristics fell within the range of other sugarcane vinasses [52, 53]. Different vinasse inputs are known to contribute different nutrition; this is taken into account in that vinasse fertirrigation is applied depending on the amount of K present in the input vinasse [54]; however, that different vinasse inputs contribute different bacteria was not known until now. The different nutrient contents of vinasse originate from the differences in the input of sugarcane stalks to the bioethanol production process; this might also be the source of the vinasse bacteria.

The vinasse draft genomes most likely represented the bacteria that survived the selective bottleneck of the bioethanol production pipeline. The potential for bacteria found in vinasse originating from later steps in the bioethanol pipeline, such as the truck from which we sampled the vinasse, was considered a minor source of bacteria due to the large capacity (10,000 L), making this a negligible source of bacteria. The core genus found in the vinasse samples was Lactobacillus (Phylum Firmicutes), which is a previously known ubiquitous ethanol pipeline contaminant due to its tolerance of low pH [15]. Other known contaminants found prior to the distillation stage that we observed in our vinasse samples included representatives of the Acetobacter, Bacillus, Bifidobacterium, Clostridium, Gluconacetobacter, Lactobacillus and Pseudomonas genera [16, 55, 56]. Strikingly, we identified members of the genera Megasphaera and Mitsuokella that have not previously been reported as bioethanol pipeline contaminants. Members of the genus Megasphaera and Mitsuokella are Gram negative ruminant fermenters that have been found in pig hindguts, cow rumen and human dental plaque and feces; gram-positive Bifidobacterium have also been used as probiotics in humans and are found in the gut, vagina and mouth of mammals and bovine rumens. Whether these bacteria interact with each other within each vinasse sample – e.g. Megasphaera and Mitsuokella utilizing lactose provided by Lactobacillus – is unknown, as is the direction of the interactions.

Uncovering the physiological mechanisms by which these particular bacteria survive the selection bottlenecks of the bioethanol process was outside the scope of the current research since our goals were to characterize fully the metagenomic data. However, we speculated that plausible protective mechanisms are biofilm formation [16, 57], strain-dependent temperature tolerance, and unknown pipeline management considerations. For the latter, the distillation material might not homogenized, thus creating pockets of lower temperatures where the bacteria can remain. Other management considerations that might affect the viability of bacterial cells are length of time exposed to the distillation temperature and the highest temperature reached. Evaluating the physiology of cultured isolates from vinasse, which can be done building upon the work described here, is an interesting topic for further research.

Here, using differential abundance binning, we successfully obtained 21 draft genomes from vinasse bacteria likely representing bioethanol contaminants. We confirmed that roughly half of the vinasse bins were of the genus Lactobacillus (Phylum Firmicutes), which is the most ubiquitous bacterial bioethanol pipeline contaminant [16]. We also uncovered contaminants with up to 70% of sample coverage from the Prevotella (Phylum Bacteroidetes), Megasphaera (Phylum Firmicutes), and Mitsuokella (Phylum Firmicutes) genera, which have not been as well-studied. Five of the draft genomes were from bacteria unknown at the phylum level. Furthermore, most of the bins recovered here were partly uncharacterized at the species level, supporting the idea that we obtained genomes from novel strains of bioethanol contaminants. Studies of bioethanol contaminants to date have used culture-based methods, which do not capture the entire microbial diversity; or profiling of 16S rRNA genes, which does not capture the functional potential of the contaminants [13, 14]. Bacterial contaminants in general are known to compete with commercial yeast strains, lowering ethanol yield; contaminants may also flocculate with the yeast or produce compounds such as acetate, butyric acid or lactose which might inhibit yeast fermentation [16]. Many bins contained sucrose metabolism-related genes, suggesting that these might compete with the commercial yeast strain for sugarcane sucrose. Annotation of the bins revealed the potential presence of bioethanol contaminant genes related to the metabolism of acetate, ethanol, mannitol, cellulose, hydrogen peroxide, lactose, sucrose and 3-hydroxy fatty acids. These results support the idea that vinasse bacteria are an additional source in identifying likely bioethanol process contaminants.

Interesting bins included Lactobacillus/Methanobrevibacter-bin30 and Arcobacter/Methanobrevibacter-bin40.2, which contained cross-contigs annotated as both bacterial and archaeal. Methanobrevibacter is an archaeal genus whose

methanogenic members are often found in vertebrate guts consuming the end products of bacterial fermentation. Finding them here suggests that this interaction might also be present in vinasse. In addition, we binned potential phage genomes, which suggest that phages are present in the fermentation tanks along with the host contaminants. The large phage genome bin 8.2 was most abundant in vinasse sample D, corresponding to a low diversity assemblage with a dominant bin, suggesting that the host of this phage was L. amylovorous-bin27. The phage bins 4.1, 4.2 and 8.1 were all most abundant in vinasse sample D, corresponding to a more diverse assemblage of bacterial hosts across the phyla Firmicutes and Bacteroidetes. These associations suggest that phage lysis may be a factor controlling bacterial population sizes in the fermentation tanks. Attention has recently been paid to using phage therapy to control bacterial contamination in bioethanol pipelines [58-59].

In addition to investigating the potential for vinasse bacteria to be contaminants in the production of bioethanol, we evaluated the potential for vinasse bacteria to contribute to N_2O emissions during fertirrigation. Vinasse fertirrigation can be considered a disturbance on the soil microbial community; the success of the vinasse assemblage in the soil likely depends on the connectivity (e.g. strength and direction of the vinasse species interactions). Pitombo et al.[11] identified significantly abundant bacterial genera under treatments of vinasse compared to unfertilized control plots using 16S rRNA marker abundance, and the significantly differentially abundant genera in the plots amended with vinasse included the vinasse bacteria (as identified here) Lactobacillus, Bacillus, Prevotella, Gluconacetobacter, Megasphaera, Mitsuokella and Acetobacter [11]. Further, unpublished research suggests that vinasse bacteria on a field experiment may persist at low abundances. These results together suggest that vinasse bacteria may successfully invade the soil microbial community. Furthermore, the vinasse bacteria may transfer to the sugarcane stalks during plant growth and at harvest time become the contaminants that are inputted with the sugarcane to the ethanol processing pipeline. In support, a survey of the bacteria associated with the sugarcane plant found the vinasse taxa Bacillus, Acetobacter and Gluconacetobacter as part of the “core” sugarcane microbiome [59]. While this is interesting speculation, we note that caution should be taken because the referenced studies were few and based on gene marker surveys at higher taxonomic levels, which hinders robust and precise interpretation. We recommend further research into the ecological interactions of vinasse bacteria with the soil bacterial community at the species or strain level during fertirrigation with vinasse.

Actual N_2O emissions from a soil are the result of the sequential biochemical processes nitrification and denitrification carried out collectively by the microbial communities in a soil. The total rate of N_2O emissions through nitrification or denitrification is controlled by carbon availability, moisture, oxygen availability, pH, temperature, and nitrate concentrations. These factors limit enzyme activity, gene transcription levels and microbial cell growth [61]. Furthermore, the abundance of the genes involved in the production (amoA, nirK, nirS, norB) or removal (nosZ) of N_2O is correlated with the actual N_2O emissions [62]. In the case of vinasse fertirrigation, if many denitrifiers invade a soil conducive to denitrification, we would expect more denitrification to occur. Whether net N_2O or N2 (the end product of denitrification) emissions would occur would depend on the number and genetic repertoires of the invading bacteria. Therefore, vinasse containing an assemblage with a higher partial (containing nirK, nirS and/or norB) to full (containing at least nosZ) denitrifier ratio may lead to higher N_2O emissions during fertirrigation.

Four phyla (Firmicutes, Actinobacteria, Proteobacteria and Bacteroidetes) were represented across the vinasse samples, although at the genus level the diversity of each assemblage fluctuated. The samples could generally be classified as dominated by Megasphaera (A, C, F) or Lactobacillus (B, D, E) at the genus level. The second assemblage (B) was the most diverse; it was dominated by Lactobacillus and containing, uniquely compared to the other time points, Proteobacteria such as Alcaligenes, as well as phage (Lactobacillus phage Lc Nu). The least diverse assemblage was D, containing mostly Lactobacillus and phage. Of the 22 potential vinasse denitrifiers, two were potential complete denitrifiers (containing nirK or nirS, norB and nosZ) and eight were potential incomplete denitrifiers (containing nirK or nirS and norB). The abundances of these potential denitrifiers varied across timepoint, suggesting varied effects on N_2O during vinasse fertirrigation with different vinasses. For example, the Lactobacillus-bin 27 dominated to 97% of the sample D abundance, and this contained a putative nirK gene; when this vinasse is sprayed onto the fields, one would expect nitrate degradation and an increase in N_2O or N2 depending on the gene content of the endogenous microbial community. Another abundant potential denitrifier present in sample A (Prevotella-Bin 2) contained only potential nosZ, suggesting that if the vinasse A were to be used in fertirrigation, the actual emission of N_2O may be reduced due to the further reduction of N_2O into N2. Furthermore, vinasse denitrifiers might directly contribute to the N_2O emissions observed when vinasse is added in conjunction with a nitrate fertilizer [63]. This suggests that vinasse application in conjunction with a reduced nitrogen source such as ammonium sulfate may be a feasible management practice to reduce N_2O production. Further research investigat-

ing the microbes involved in N_2O emissions during fertirrigation with vinasse would greatly aid in steering future vinasse management strategies.

Vinasse fertirrigation has raised human health concerns that vinasse bacteria may carry antibiotic resistance genes (ARGs) [21]. These genes can enter the soil resistome and can be transferred using horizontal gene transfer to other soil bacteria, with potential spreading of antibiotic resistance genes to soil-derived human pathogens. Here a search of the annotation results of the recovered vinasse bins found multidrug resistance genes in 34 of the 36 large bins. Surprisingly, no drug resistance genes were found in the phage bins; this may indicate, that the phages from which these genomes were not prophage that confer auxiliary metabolic genes in the form of antibiotic resistance to the vinasse bacteria. These results warrant further study of the fate of ARG's from vinasse during fertirrigation.

While significant progress has been made in metagenome assembly and binning, some caveats should be noted to the bins we recovered here. Misassembly and misbinning can occur and bias the final results, in our case identifying relevant genes present in the bins. We addressed these issues by comparing three assemblers and three binning tools and choosing the best of each. Further, we used large kmer sizes for the final cross-assembly, and this successfully allowed MaxBin2 to bin to the level of species. We additionally used the manual refinement feature of anvi'o to improve the bins. Because bins with low completeness as determined by the presence of universal marker genes can still contain useful information regarding potential gene content, we used all useful bins to characterize the vinasse assemblage. Eight bins could not be refined, and these represent unbinned vinasse bacterial genome content; however, the information from this genomic material was characterized in the metagenomic analyses. We included several different methods for each analysis to supplement each other as database coverage and read length can bias results based on sequence alignment. Moreover, we used the metagenomic analysis to complement the bin results. Interestingly, comparing the qPCR, putative gene abundances and gene recruitment results suggested that the qPCR primers we used do not cover the entire diversity of vinasse bacteria or alternately that the HMM results may be biased toward false positives.

Here we used metagenomic analysis and genome binning to characterize in depth the assemblage of six vinasse samples from one bioethanol mill. We identified previously unknown vinasse taxa compared to taxa identified through culture- or 16S rRNA survey-based studies of the ethanol processing pipeline steps prior to vinasse. Furthermore, we obtained 21 draft genomes and 8 phage or mobile element genomes from vinasse, which to our knowledge is the first study to do so. Vinasse bacteria included mainly putative denitrifiers, which may directly affect

soil N_2O or N_2 emissions when applied during fertirrigation, although more research is needed into the ecological interactions during this event. In the vinasse bins we found the putative presence of antibiotic resistance genes and genes affecting yeast fermentation, which potentially implicate vinasse bacteria in negative impacts on human health and bioethanol production, respectively. We suggest that monitoring the vinasse assemblage is a promising option to screen both for bioethanol production contaminants and to identify vinasse batches which, when added to the fields during fertirrigation, may lead to higher N_2O emissions. Because of the decreasing costs of high-throughput sequencing, we suggest that monitoring of vinasse assemblages can be widely implemented to improve sugarcane bioethanol production sustainability.

6.5 Declarations

Availability of data and material

The datasets generated and analyzed during the current study are available in the MG-RAST repository with identification numbers described in **Additional file 8**. The metagenomes are additionally found under NCBI BioProject id: PRJNA435511. Draft genomes are available on Zenodo (DOI: 10.5281/zenodo.1194340).

Competing interests

The authors declare that they have no competing interests.

Abbreviations

MGRAST metagenomics analysis server
***nir*K, *nir*S** nitrite reductase genes
***nor*B** nitric oxide reductase gene
***nos*Z** nitrous oxide reductase gene
***amo*A** ammonium monooxygenase gene
ABR antibiotic resistance gene

Authors' contributions

NAC, EEK, HVV designed the study. KSL, JDC, HC collected the data. NAC processed and analyzed the data, and wrote the paper. All the authors contributed with the discussion, read and approved the final manuscript.

Acknowledgements

The authors thank André C. Vitti for his assistance in collecting the vinasse samples and for collecting information on the sugar and ethanol mill pipeline. This work was supported by The Netherlands Organization for Scientific Research (NWO-729.004.003), Sao Paulo State foundation (FAPESP-2013/50365-5, FAPESP BEPE-2014/24141-5, FAPESP-2013/12716-0). Publication number 6473 of the NIOO-KNAW, Netherlands Institute of Ecology.

6.6 Additional files

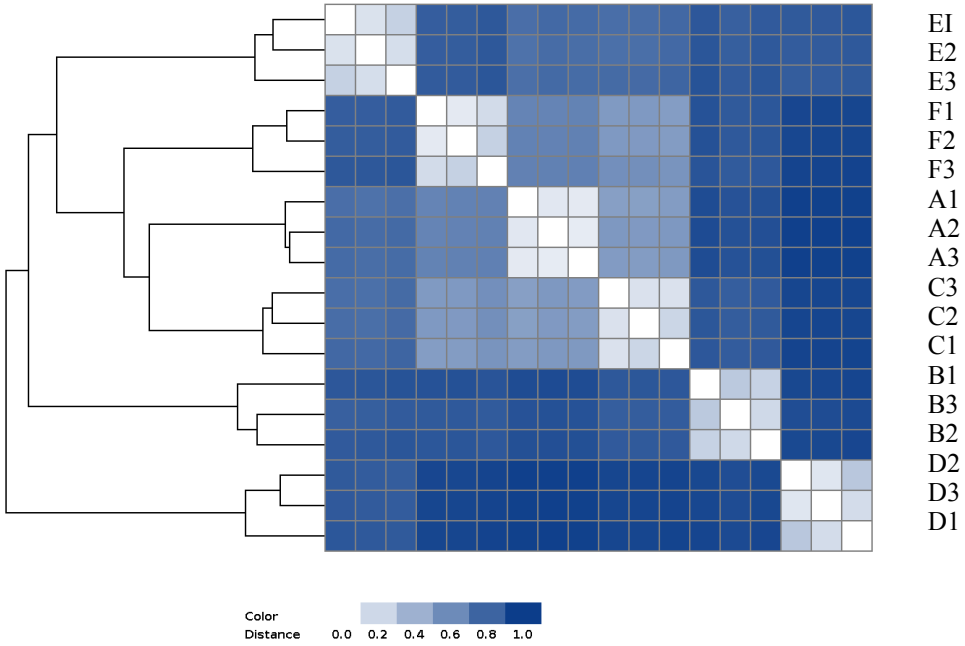
Additional file 1. Data description of the 18 vinasse metagenomes.

Date Sampled	Sample Name	Sample Id	DNA Weight (g)	DNA Conc. (ng/µl)	# reads	Forward # bases (Mbp)	Reverse # bases (Mbp)	Reads mapped to cross-contigs (%)
Nov. 2013	A-1	1V1-1	0.297	32.5	461,784	131	117	92.77
	A-2	1V1-2	0.250	39.6	454,721	130	115	92.69
	A-3	1V1-3	0.295	20.4	469,527	130	115	92.19
Dec. 2013	B-1	1V2-2	0.300	39.2	417,625	110	94	73.20
	B-2	1V2-3	0.295	53.1	517,039	142	131	73.67
	B-3	1V2-4	0.292	38.5	542,208	150	139	74.67
July 2014	C-1	2V1-1	0.301	13.9	362,499	100	89	92.38
	C-2	2V1-2	0.267	14.2	501,511	138	123	92.53
	C-3	2V1-3	0.291	14.4	432,207	119	107	92.12
Aug. 2014	D-1	2V2-1	0.295	17.2	489,336	138	132	95.12
	D-2	2V2-2	0.291	18.6	280,161	77	74	95.33
	D-3	2V2-3	0.294	21.0	351,407	971	935	95.3
Oct. 2014	E-1	3V1-1	0.294	6.19	363,382	981	914	91.30
	E-2	3V1-2	0.280	5.57	434,111	117	108	91.18
	E-3	3V1-3	0.290	7.22	472,732	135	124	91.47
Nov. 2014	F-1	3V2-1	0.294	7.29	444,056	122	114	91.14
	F-2	3V2-2	0.285	6.64	500,636	136	128	91.29
	F-3	3V2-3	0.292	5.53	323,376	883	794	91.66

Additional file 2. Primers and thermocycler conditions used in gene abundance analysis by real time qPCR of the vinasse samples.

Target gene	Primer	Primer Sequence	Size (bp)	12 µL of reaction	Thermal profile
AOA *amoA*	Arch-amoAF	5'-STAATGGTCTGGCTTAGACG-3'	635 491	6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.125 µL of each primer (10 pmol), 1.75 µL of BSA and 4 µL of DNA (3 ng).	95°C-5 min.; 40x 95°C-10s, 64°C-10s, 72°C-20s
	Arch-amoAR	5'-GCGGCCATCCATCTGTATGT-3'		6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.125 µL of each primer (10 pmol) and 4 µL of DNA (3 ng).	95°C-10min.; 40x 95°C-10s, 65°C-25s,
AOB *amoA*	amoA1F	5'-GGGGTTTCTACTGGTGGT-3'			
	amoA2R	5'-CCCCTCKGSAAAGCCTTCTTC-3'			
AOA *amoA*	Arch-amoAF	5'-STAATGGTCTGGCTTAGACG-3'	635	6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.125 µL of each primer (10 pmol), 1.75 µL of BSA and 4 µL of DNA (3 ng).	95°C-5 min.; 40x 95°C-10s, 64°C-10s, 72°C-20s
	Arch-amoAR	5'-GCGGCCATCCATCTGTATGT-3'			
	amoA2R	5'-CCCCTCKGSAAAGCCTTCTTC-3'			
*Nos*Z [3]	nosZ2F	5'-CGCRACGGCAASAAGGTSMSSGT-3'	267	6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.250 µL of each primer (10 pmol), 1.20 µL of BSA and 4 µL of DNA (1.25 ng).	95°C-5 min.; 40x 95°C-10s, 64°C-10s, 72°C-20s
	nosZ2R	5'-CAKRTGCAKSGCRTGGCAGAA-3'			
*nir*K [4]	NirK876	5'-ATYGGCGGVAYGGCGA-3'	165	6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.250 µL of each primer (10 pmol), 1.50 µL of BSA and 4 µL of DNA (1.25 ng).	95°C-5 min.; 40x 95°C-15s, 62°C-15s, 72°C-20s
	NirK1040	5'-GCCTCGATCAGRTTRTGGTT-3'			
*nir*S [5]	nirScd3aF	5'-GTSAACGTSAAGGARACSGG-3'	425	6 µL of Sybrgreen Bioline SensiFAST SYBR non-rox mix, 0.250 µL of each primer (10 pmol), 1.20 µL of BSA and 4 µL of DNA (1.25 ng).	95°C-5 min.; 40x 95°C-10s, 63°C-10s, 72°C-20s
	nirSR3cd	5'-GASTTCGGRTGSGTCTTGA-3'			

1. Francis CA, Roberts KJ, Beman JM, Santoro AE & Oakley BB. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. Proc Natl Acad Sci USA. 2005;102:14683–88.
2. Rotthauwe JH, Witzel KP & Liesack W. The Ammonia monooxygenase structural gene amoA as a functional marker: molecular fine-Scale analysis of natural ammonia-oxidizing populations. Appl Environ Microbiol. 1997;63: 4704 12.
3. Henry S, Bru D, Stres B, Hallet S & Philippot L. Quantitative detection of the nosZ gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *nar*G, *nir*K, and *nos*Z genes in soils. App. Environ Microbiol. 2006;72: 5181–89.
4. Henry S, Baudoin E, López-Gutiérrez JC, Martin-Laurent F, Brauman A & Philippot L. Quantification of denitrifying bacteria in soils by *nir*K gene targeted real-time PCR. J Microbiol Methods. 2004;59:327–35.
5. Throbäck IN, Enwall K, Jarvis A & Hallin S. Reassessing PCR primers targeting *nir*S, *nir*K and *nos*Z genes for community surveys of denitrifying bacteria with DGGE. FEMS Microbiol Ecol. 2004;49:401–17.



Additional file 3. Hierarchical clustering of the vinasse metagenomes based on partial de Bruijn graph overlap from Metafast analysis. Replicates were most similar to each other.

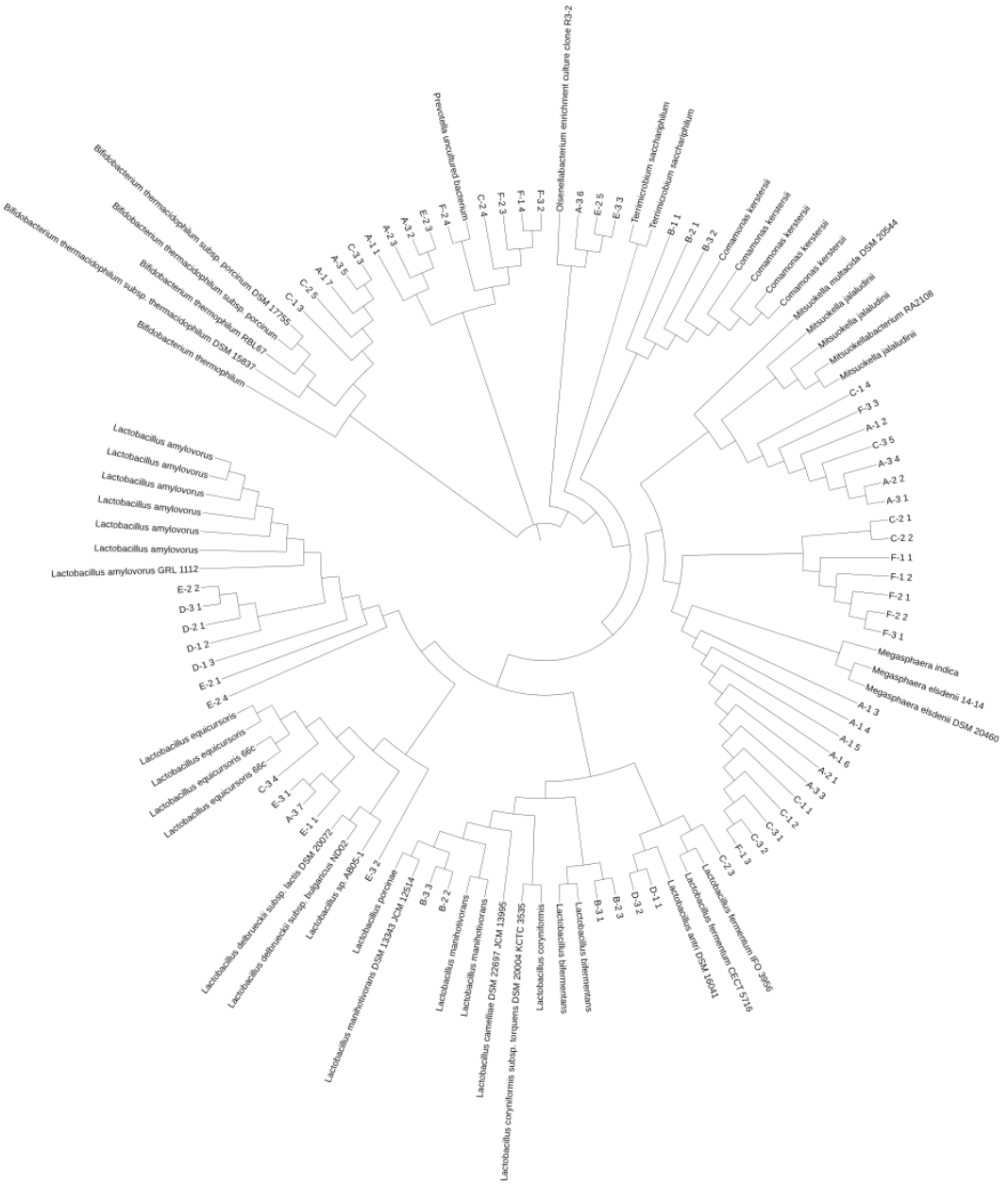
Additional file 4. Data description of the merged vinasse metagenomes uploaded to MG-RAST.

MG-RAST ID	Sample Date	Sample Name	Sample ID	Percent merged	# merged reads	Avg. merged read length	# merged bases (Mbp)
4678764.3	Nov. 2013	A-1	1V1-1	87.98%	236896	592,17	85
4678762.3		A-2	1V1-2	88.72%	230245	592,38	83
4678758.3		A-3	1V1-3	87.40%	245213	590,59	83
4678765.3	Dec. 2013	B-1	1V2-2	93.07%	263198	582,17	86
4678752.3		B-2	1V2-3	95.62%	359330	580,67	116
4678749.3		B-3	1V2-4	95.15%	376765	581,71	124
4678755.3	July 2014	C-1	2V1-1	88.74%	193318	590,16	64
4678760.3		C-2	2V1-2	91.90%	275242	589,16	91
4678754.3		C-3	2V1-3	91.96%	252621	588,15	83
4678766.3	Aug. 2014	D-1	2V2-1	95.97%	402629	581,89	139
4678761.3		D-2	2V2-2	95.88%	235413	577,50	79
4678753.3		D-3	2V2-3	96.49%	300148	576,57	100
4678756.3	Oct. 2014	E-1	3V1-1	94.29%	258649	584,54	84
4678751.3		E-2	3V1-2	92.39%	293786	585,17	96
4678759.3		E-3	3V1-3	91.44%	305881	589,63	111
4678757.3	Nov. 2014	F-1	3V2-1	93.26%	320201	586,67	109
4678763.3		F-2	3V2-2	95.37%	383266	583,43	126
4678750.3		F-3	3V2-3	84.59%	269124	575,92	94

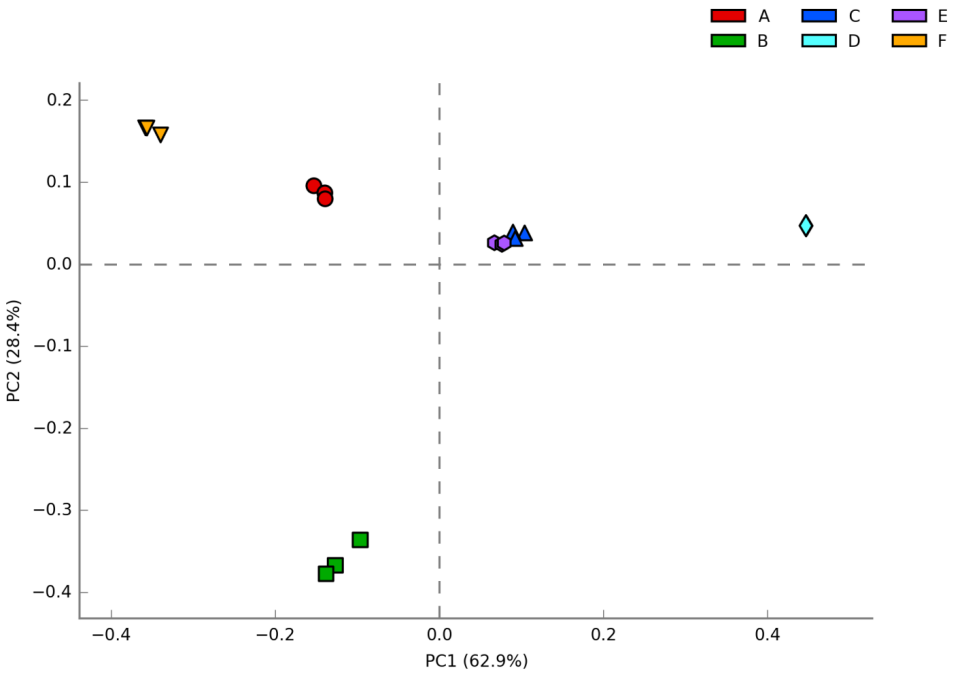
Additional file 5. Taxonomic distribution of the merged vinasse metagenomes from MG-RAST annotation against RefSeq database. Phyla with average relative abundance greater than 1% across all samples were included. Samples with significantly different phyla between groups (Tukey-Kramer post-hoc test, 95% confidence interval, $p < 0.001$) are indicated by different letters.

			Average proportion of sample					
Phylum	p-values (corrected)	Effect size	A	B	C	D	E	F
Firm.	1.94E-14	0.998	44.5±0.4 a	39.2±1.7 b	61.2±0.9 c	97.0±0.0 d	58.8±0.5 e	35.4±0.3 f
Bacter.	4.09E-14	0.998	29.4±1.5 a	9.5±0.2 b	11.3±0.7 c	0.8±0.0 d	11.5±0.3 e	52.8±1.1 f
Actino.	1.88E-09	0.985	15.9±1.5 a	2.1±0.1 b	17.5±1.4 a	0.4±0.0 b	17.8±0.6 a	3.3±0.8 b
Proteo.	7.19E-14	0.997	0.8±0.0 a	39.4±1.8 b	0.8±0.0 a	0.3±0.0 a	1.7±0.0 a	1.4±0.0 a

Effect sizes and corrected p-values were calculated using ANOVA on mean relative abundance of phyla in sample groups using the Benjamini-Hochberg multiple test correction in STAMP.



Additional file 6. Phylogenetic relationships between full-length 16S rRNA genes reconstructed from the vinasse metagenomes using REAGO. The 16S rRNA sequences from vinasse and reference sequences were aligned using ClustalW. The neighbor-joining tree was created with MEGA and visualized using iTol ignoring branch lengths.

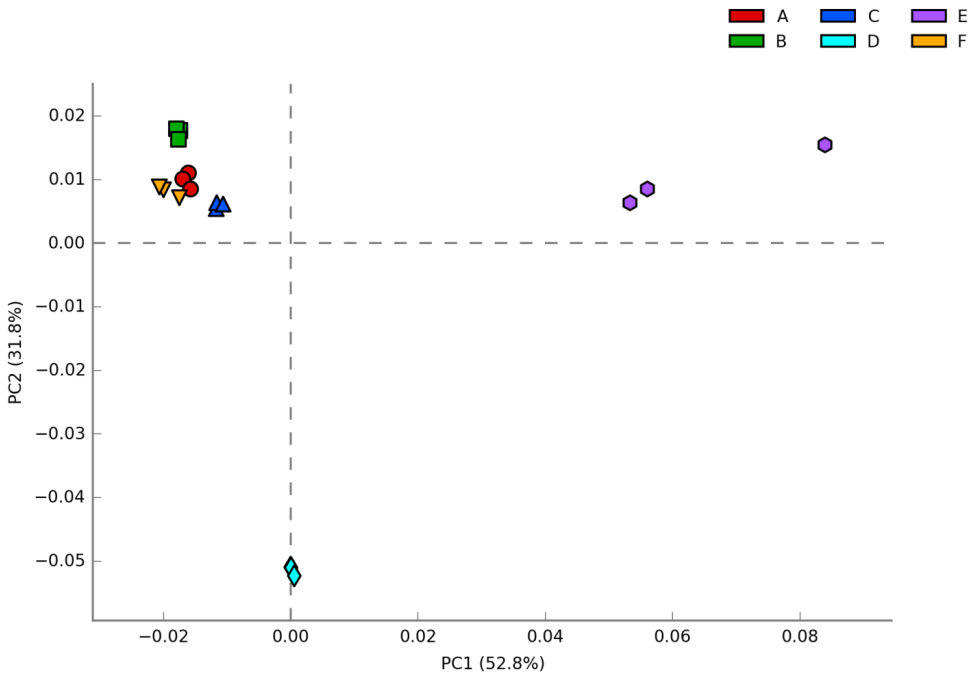


Additional file 7. Principal component analysis of the phylum-level abundance distributions of the vinasse metagenomes. Relative abundance profiles were determined using MG-RAST against the RefSeq database and phyla membership was determined using the Last Common Ancestor algorithm.

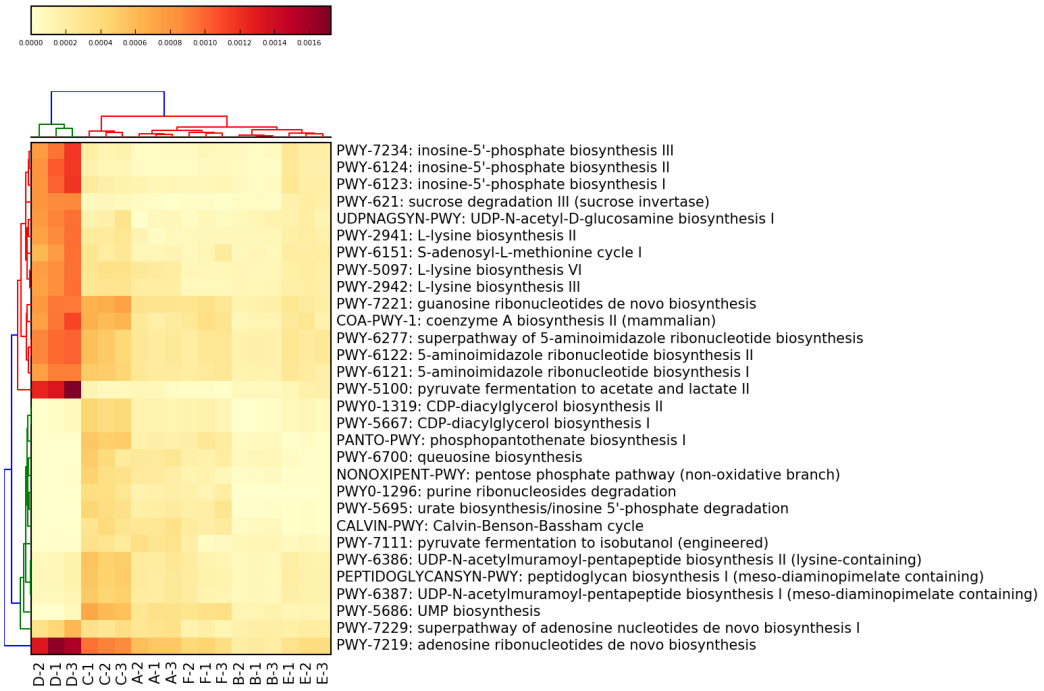
Additional file 8. Functional potential characterization of the vinasse metagenomes from MG-RAST annotation against the Subsystems database**.** Only the subsystems at level 1 with average relative abundance greater than 2% across all samples were included. Significantly different subsystems at level 1 between sample groups (Tukey-Kramer post-hoc test, 95% confidence interval, $p < 0.05$) are indicated by different letters.

			Relative sample abundance					
Subsystems Level 1 Category	p-values (corrected)	Effect size	A	B	C	D	E	F
Carbohydrates	2.84E-08	0.967	16.3±0.3 ac	12.9±0.2 b	15.5±0.1 a	15.5±0.1 a	13.7±0.3 b	17.0±0.5 c
Clustering-based subsystems	4.37E-09	0.977	14.5±0.2 a	14.6±0.0 a	15.2±0.1 b	17.2±0.0 c	14.8±0.2 ab	14.2±0.2 a
Amino Acids and Derivatives	6.35E-13	0.997	9.2±0.1 ac	9.4±0.0 a	9.3±0.1 a	4.7±0.1 b	7.0±0.2 d	8.9±0.0 c
Miscellaneous	8.50E-05	0.863	7.6±0.1 a	7.5±0.0 a	7.6±0.1 a	7.3±0.0 b	7.0±0.2 c	7.3±0.1 bc
Protein Metabolism	2.25E-08	0.969	7.1±0.1 a	6.9±0.1 a	7.8±0.1 c	8.3±0.1 d	7.0±0.1 ab	6.7±0.2 b
DNA Metabolism	3.86E-11	0.991	5.5±0.1 a	5.4±0.0 a	5.4±0.1 a	7.3±0.1 c	5.8±0.0 b	5.7±0.0 b
RNA Metabolism	3.43E-07	0.949	5.7±0.1 a	5.7±0.1 a	5.9±0.1 a	7.0±0.1 c	5.7±0.2 a	5.1±0.1 b
Cofactors, Vitamins	3.71E-11	0.991	5.6±0.0 a	5.8±0.1 a	5.1±0.1 b	3.5±0.0 c	4.0±0.2 d	5.2±0.0 b
Cell Wall and Capsule	3.04E-06	0.923	4.7±0.1 ac	4.4±0.1 bc	4.6±0.1 c	4.9±0.1 a	4.2±0.1 b	4.9±0.0 a
Phages and Prophages	8.82E-08	0.960	2.9±0.1 a	2.3±0.0 a	3.1±0.0 a	2.4±0.0 a	10.0±1.4 b	2.6±0.1 a
Nucleosides and Nucleotides	4.42E-10	0.984	3.2±0.0 a	2.7±0.0 b	3.3±0.0 a	3.7±0.1 c	4.1±0.1 d	2.9±0.1 b

Effect sizes and corrected p-values were calculated using ANOVA on mean relative abundance of Subsystems level 1 in sample groups using the Benjamini-Hochberg multiple test correction in STAMP.



Additional file 9. Principal component analysis of the Subsystems Level 1 category abundances of the vinasse metagenomes. The colors correspond to time point. Profiles were determined against the Subsystems database using MG-RAST and relative abundances of phyla were calculated out of the total number of sample reads.



Additional file 10. Functional potential profiles of the top 30 pathways across the vinasse samples, excluding “unmapped” and “uncategorized” results. The functional group and sample profiles were clustered using hclust2 from humann2 analysis against the UniRef90 database.

Additional file 11. All vinasse bin characteristics and relative sample abundances (indicated by heatmap per sample). Bin id's highlighted in green indicate "good" bins; yellow id's indicate "interesting" bins, and red id's indicate "bad" bins. Continued on next page.

Bin Id	A	B	C	D	E	F	Length (Mbp)	# Contigs	N50	GC (%)	Completeness (%)	Redundancy (%)
1	6	1	2	0	0	1	2.4	209	23171	53	92	2
2	14	2	1	0	4	3	3.5	547	10519	49	94	4
3	8	1	18	0	2	13	2.2	352	10534	53	97	2
5	3	0	4	0	2	1	1.9	298	11078	60	94	3
6	4	1	2	0	1	12	2.4	539	6347	44	91	2
7	9	2	3	0	4	2	4.9	2521	1944	51	63	40
9	2	0	2	0	1	0	2.0	407	6384	60	90	6
10	1	1	1	1	3	0	2.1	183	27583	47	96	1
11	7	2	4	0	2	10	5.8	2606	2396	44	36	9
12	3	0	2	0	5	1	2.3	485	7489	66	91	5
13	2	0	2	0	0	0	1.5	605	2710	63	71	12
14	2	0	1	0	0	1	1.9	947	2190	52	74	15
15	2	0	2	0	1	1	2.2	1017	2297	53	76	9
16	1	1	1	0	0	13	3.0	307	22370	42	99	1
17	2	0	1	0	1	0	1.4	928	1434	60	43	14
18	2	0	1	0	2	0	1.5	893	1665	63	69	22
19	1	0	1	0	3	0	2.0	917	2351	62	60	12
20	0	0	1	0	1	1	1.2	387	3766	54	64	2
21	0	0	1	1	2	1	1.8	298	9992	50	88	1
22	2	1	3	0	1	12	5.3	2375	2491	36	39	7
23	0	1	1	3	2	1	1.9	373	7041	47	95	4
24	0	2	1	0	1	0	1.8	220	13204	53	98	4
25	1	1	1	1	0	2	1.8	729	2822	40	91	9
26	1	1	2	0	1	1	2.1	1236	1733	41	66	15
27	0	1	1	45	3	1	1.9	262	11858	38	99	1
28	0	3	0	0	0	0	2.1	340	8670	38	96	5
29	0	1	0	2	0	0	1.0	439	2658	50	88	9
30	0	3	1	3	2	3	3.9	2021	1897	31	47	16
31	0	0	8	0	0	0	3.3	1595	2241	54	79	37
32	0	0	0	6	0	0	2.0	104	208993	36	99	1
33	0	0	0	1	3	0	1.7	447	4850	48	96	9
34	0	3	0	0	0	0	2.7	343	12289	60	92	1
35	0	4	0	0	0	0	2.7	259	13750	43	96	6
36	0	5	0	0	0	0	1.7	647	2989	49	67	5
36.2	0	4	0	0	0	0	1.4	868	1558	48	38	5
36.3	0	1	0	0	0	0	0.5	245	2503	48	10	0
37.1	0	10	0	0	0	0	3.1	1326	2603	57	77	24
37.2	0	4	0	0	0	0	1.2	784	1533	58	34	6
37.3	0	2	0	0	0	0	0.7	506	1236	49	16	1

Bin Id	A	B	C	D	E	F	Length (Mbp)	# Contigs	N50	GC (%)	Completeness (%)	Redundancy (%)
38	0	9	0	0	0	0	3.0	488	9382	60	89	3
39	0	0	0	4	0	0	1.9	205	15510	47	99	4
39.2	0	0	0	0	0	0	0.1	76	1387	47	5	1
40.1	0	4	0	0	0	0	1.6	190	11919	28	97	2
40.2	0	2	0	0	0	0	0.6	271	2044	27	15	1

Additional file 12. All vinasse bin taxonomic affiliations based on CAT. Bin id's highlighted in green indicate "good" bins; yellow id's indicate "interesting" bins, and red indicate "bad" bins. Continued on next page.

Bin Id	Kingdom	Phylum	Class	Order	Family	Genus	Species
1	B	F	Nega	Selenomonadales	Veillonellaceae	Mitsuokella	Unclassified
2	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	P. mutisaccharivorax/Unclassified
3	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera	Unclassified/M. elsdenii
4.1	V				Caudovirales	Siphovirales	Lactobacillus phage Ldl1
4.2	B	U/B	U/B	U/B	U	U	U
5	B	A	Acti	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	Unclassified
6	B	B	Bact	Bacteroidales	Prevotellaceae/Unclassified	Prevotella/Unclassified	Unclassified
7	B	A/B/F					
8.1	B/V	U	U	U	U/Caudovirales	U/Siphoviridae	U/Lactobacillus phage LdL1
8.2	B	F	B	Lactobacillales	Lactobacillaceae	Lactobacillus	U
8.3	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	P. histicola/U
8.4	B	F	Bac	Lactobacillales	Lactobacillaceae	Lactobacillus	U
8.5	B	B	U/Bact				
8.6	B	F	Bac	Lactobacillales	Lactobacillaceae	Lactobacillus	U
9	B	U/A					
10	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. equicursoris/Unclassified
11	B	B/F/U					
12	B	A/U					
13	B	U/A					
14	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera	U/sp. DJF_B143
15	B	F	Nega	Selenomonadales	Veillonellaceae	Dialister	U/D. succinatiphilus
16	B	B	Bact	Bacteroidales	Prevotellaceae	Prevotella	U
17	B	A/F					
18	U/B						
19	B	U/A					
20	B	F	Clos	Clostridiales	Eubacteriaceae/U	Pseudoramibacter/U	U/P. alactolyticus
21	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. delbrueckii/U
22	B	B/F					

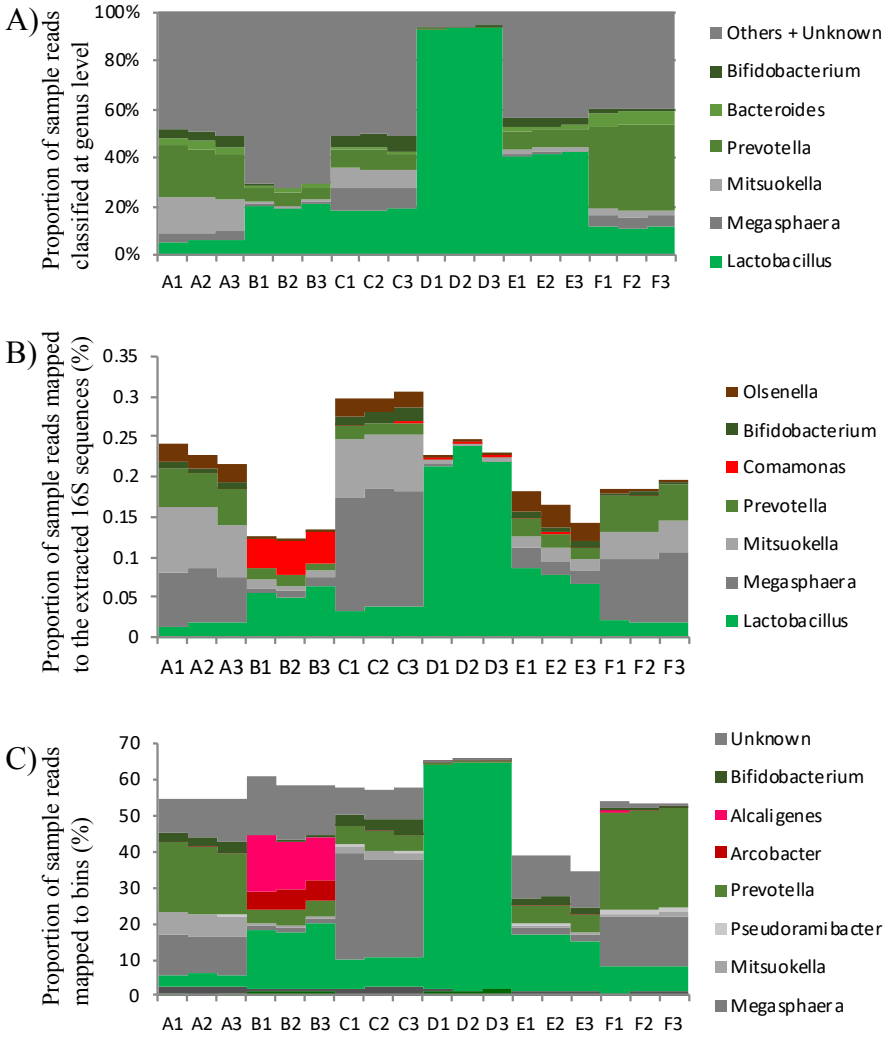
F = *Firmicutes*, B = *Bacteroidetes*, A = *Actinobacteria*, P = *Proteobacteria*, E = Euryarchaeota, U = Unknown, Bact = *Bacteroidia*, Nega = *Negativicutes*, Clos = *Clostridia*, Mega = *Megasphaera*, Lact = *Lactobacillales*

Additional file 12 cont'd.

Bin Id	Kingdom	Phylum	Class	Order	Family	Genus	Species
23	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. mucosae/U
24	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. fermentum/U
25	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U/L. vini
26	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U/L. agilis
27	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	U
28	B	F	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	L. vini
29	B	F	Clos	Clostridiales	Clostridiaceae	Clostridium	sp. CAG:568/U
30	B/A	F/E	Bacilli/Metha	Bacteroidales/Methanobacteriaceae	Lactobacillaceae/Methanobactericeae	Lactobactillus/Methanobrevibacter	U
31	B	F	Nega	Selenomonadales	Veillonellaceae	Megasphaera/U	U
32	B	F	Nega	Lactobacillales	L	L	U
33	B	F	B	Lactobacillales	L	L	L. secaliphilus/U
34	B	F	B	Lactobacillales	L	L	U/L. manihotivorans
35	B	F	B	Lactobacillales	L	L	L. bifermentans/U
36	B	P	Beta	Burkholderiales	Alcaligenaceae	U	U
36.2	B	P/F					
36.3	B	F	B	Lactobacillales	L	L	U/L. manihotivorans
37.1	B	P	B	Burkholderiales	Alcaligenaceae	Alcaligenes	A. faecalis/U
37.2	B	P	B	B	A	Alcaligenes	U/A. faecalis
37.3	B/A	P/E					
38	B	P	B	B	Comamonadaceae	U/Comamonas	U/C. kerstersii
39	B	F	B	Lactobacillales	L	L	U/L. panis
39.2	B	F	B	Lactobacillales	L	L	L. panis/L. ponis/U
40.1	B	P	E	Campylobacterales	Campylobacteraceae	Arcobacter	A. skirrowii/U
40.2	B/A	P/E	E/Metha	Campylobacterales/Methanobacteriales	Campylobacteraceae/Methanobacteriaceae	Arcobacter/Methanobrevibacter	U

F = *Firmicutes*, B = *Bacteroidetes*, A = *Actinobacteria*, P = *Proteobacteria*, E = Euryarchaeota, U = Unknown, Bact = *Bacteroidia*, Nega = *Negativicutes*, Clos = *Clostridia*, Mega = *Megasphaera*, Lact = *Lactobacillales*

Additional file 13. Comparison of genus abundances across samples from the A) MGRAST, B) extracted 16S and the C) bin taxonomy results. The colors correspond to genus and phyla as Phylum Firmicutes (green), Proteobacteria (red), Actinobacteria (brown) and Bacteroidetes (orange).



6.6 References

1. Amorim,HV, Lopes ML, de Castro Oliveira JV, Buckeridge MS, Goldman GH. Scientific challenges of bioethanol production in Brazil. App Microbiol Biotech. 2011;91:1267-1275.
2. Christofoletti CA, Escher JP, Correia JE, Marinho JFU, Fontanetti CS. Sugarcane vinasse: environmental implications of its use. Waste Manage. 2013;33:2752-2761.
3. Parnaudeau V, Condom N, Oliver R, Cazevieille P, Recous S. Vinasse organic matter quality and mineralization potential, as influenced by raw material, fermentation and concentration processes. Biores Tech. 2008;99:1553-1562.
4. Moore CCS, Nogueira AR, Kulay L. Environmental and energy assessment of the substitution of chemical fertilizers for industrial wastes of ethanol production in sugarcane cultivation in Brazil. Intl J Life Cycle Ass. 2017;22:628-643.
5. Jiang ZP, et al. Effect of long-term vinasse application on physico-chemical properties of sugarcane field soils. Sugar Tech. 2012;14:412-417.
6. Zhou M, Luo Y, Zhou Z, Gong D, Zhou X. The effect of alcohol waste liquid (as a top dressing) on growth and yield of sugarcane. Guizhou Ag Sci. 2008;36:102-103.
7. YunChuan M, YanPing Y, Qiang L, YangRui L. Effects of vinasse on the quality of sugarcane and key enzymes in sucrose synthesis. SW China J Agric Sci. 2009;22:55-59.
8. Yang SD, Liu JX, Wu J, Tan HW, Li YR. Effects of vinasse and press mud application on the biological properties of soils and productivity of sugarcane. Sugar Tech. 2013;15:152-158.
9. Navarrete AA, et al. Multi-analytical approach reveals potential microbial indicators in soil for sugarcane model systems. PloS One. 2015;10:e0129765.
10. do Carmo JB et al. Infield greenhouse gas emissions from sugarcane soils in Brazil: effects from synthetic and organic fertilizer application and crop trash accumulation. GCB Bioenergy. 2013;5:267-280.
11. Pitombo LM et al. Exploring soil microbial 16S rRNA sequence data to increase carbon yield and nitrogen efficiency of a bioenergy crop. GCB Bioenergy. 2015.
12. Moran-Salazar R, et al. Utilization of vinasses as soil amendment: consequences and perspectives. SpringerPlus. 2016;5:1-11.
13. Rein P. Proc S African Sugar Tech Ass. 196-200.
14. Lopes ML, et al. Ethanol production in Brazil: a bridge between science and industry. Brazilian J Microbiol. 2016;47:64-76.
15. Costa OY, et al. Microbial diversity in sugarcane ethanol production in a Brazilian distillery using a culture-independent method. J Ind Microbiol Biotechnol. 2015;42:73-84.
16. Brexó RP, Santana AS. Impact and significance of microbial contamination during fermentation for bioethanol production. Ren Sust Energy Rev. 2017a;73:423-434.
17. Cabrini KT, Gallo CR. Yeast identification in alcoholic fermentation process in a sugar cane industry unit of São Paulo state, Brazil. Scientia Agricola. 1999;56:207-216.
18. Lucena BT, et al. Diversity of lactic acid bacteria of the bioethanol process. BMC Microbiol. 2010;10:298.
19. de Souza RB, et al. The consequences of *Lactobacillus* vini and *Dekkera bruxellensis* as contaminants of the sugarcane-based ethanol fermentation. J Ind Microbiol & Biotech. 2012;39:1645-1650.
20. Alcarde V, Yokoya F. Effect of the bacterial population on flocculation of yeasts isolated from industrial processes of alcoholic fermentation. STAB-Açúcar, Álcool e Subprodutos. 2003;21:40-42.
21. Braga LP, et al. Vinasse fertirrigation alters soil resistome dynamics: an analysis based on metagenomic profiles. BioData Mining. 2017;10;17.
22. Antonangelo ATB, Alonso DP, Ribolla PE, Colombi, D. Microsatellite marker−based assessment of the biodiversity of native bioethanol yeast strains. Yeast. 2013;30:307-317.
23. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011;27:863-864.

24. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30:614-620.
25. Nayfach S, Pollard KS. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. Genome Biology. 2015;16:51.
26. Ulyantsev VI, Kazakov SV, Dubinkina VB, Tyakht AV, Alexeev DG. MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data. Bioinformatics. 2016;32:2760-2770..
27. Meyer F, et al. The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics. 2008;9:386.
28. Parks DH, Beiko RG. Identifying biologically relevant differences between metagenomic communities. Bioinformatics. 2010;26:715-721.
29. Segata N, et al. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 2012;9:811-814.
30. Abubucker S, et al. Metabolic reconstruction for metagenomic data and its application to the human microbiome. PLoS Comp Biol. 2012;8:e1002358.
31. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015;3:e1029.
32. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007;23:1282-1288.
33. Yuan C, Lei J, Cole J, Sun Y. Reconstructing 16S rRNA genes in metagenomic data. Bioinformatics. 2015;31:35-43.
34. Quast C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:590-596.
35. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823-1829.
36. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol & Evol. 2016;33:1870-1874.
37. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molec Biol & Evol. 1987;4:406-425.
38. Tamura K, Nei M, Kumar S. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Nat Aca Sci USA. 2004;101:11030-11035.
39. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evol. 1985:783-791.
40. Li D, et al. ISBRA 2016, Minsk, Belarus. Proceedings June 5-8, 2016. 309 (Springer).
41. Boisvert S, Raymond F, Godzaridis É, Laviolette F, Corbeil J. Ray Meta: scalable de novo metagenome assembly and profiling. Genome Biol. 2012;13:R122.
42. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674-1676.
43. Bankevich A, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comp Biol. 2012;19:455-477.
44. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072-1075.
45. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9:357-359.
46. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015;3:e1165.
47. Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016;32:605-607.
48. Cambuy DD, Coutinho FH, Dutilh BE. Contig annotation tool CAT robustly classifies assembled metagenomic contigs and long sequences. bioRxiv. 2016:072868.

49. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043-1055.
50. Eren AM, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.
51. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068-2069.
52. Moraes BS, Zaiat M, Bonomi A. Anaerobic digestion of vinasse from sugarcane ethanol production in Brazil: Challenges and perspectives. Ren & Sust Energy Rev. 2015;44:888-903.
53. Reis RCE, Hu B. Vinasse from sugarcane ethanol production: better treatment or better utilization? Front Energy Res. 2017;5:7.
54. Botellho RG, Christofoletti CA, Correira JE, Tornisielo VL. Environmental implications of using waste from sugarcane industry in agriculture. Prod Cons & Agr Manage. 2014;91.
55. Beckner, M., Ivey, M.L., Phister, T.G. (2011) Microbial contamination of fuel ethanol fermentations. Letters in applied microbiology 53, 387-394.
56. Bonatelli, M.L., Quecine, M.C., Silva, M.S., Labate, C.A. (2017) Characterization of the contaminant bacterial communities in sugarcane first-generation industrial ethanol production. FEMS microbiology letters 364.
57. Brexó, R.P., Sant'Ana, A.d.S. (2017b) Microbial interactions during sugar cane must fermentation for bioethanol production: does quorum sensing play a role? Critical Reviews in Biotechnology, 1-14.
58. Solomon E B, Okull D. (Google Patents, 2008).
59. Roach DR, Khatibi PA, Bischoff KM, Hughes SR, Donovan DM. Bacteriophage-encoded lytic enzymes control growth of contaminating Lactobacillus found in fuel ethanol fermentations. *Biotechnology for biofuels* **6**, 20 (2013).
60. De Souza RSC, et al. Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. Sci Rep. 2016;6:28774.
61. Wallenstein MD, Myrold DD, Firestone M, Voytek M. Environmental controls on denitrifying communities and denitrification rates: insights from molecular methods. Ecol Appl. 2006;16:2143-2152.
62. Philippot L, Andert J, Jones CM, Bru D, Hallin S. Importance of denitrifiers lacking the genes encoding the nitrous oxide reductase for N_2O emissions from soil. Glob Chang Biol. 2011;17:1497-1504.
63. Shade A, et al. Fundamentals of microbial community resistance and resilience. Front Microbiol. 2012;3:417.

Chapter 7

General Discussion

Discussion

In this thesis I presented the results of my research regarding the effects of nitrogen fertilizers on soil microbial communities in agriculture. Fertilization in agricultural soils creates different soil "habitats." These habitats present different levels of soil physicochemical properties from the different fertilizer applications, which can lead to differences in the resident communities. Nitrogen fertilization appears greatly to affect the composition of the soil bacterial community, while the effect of nitrogen fertilization on the plant and soil fungal communities depends on the availability of other macronutrients, in particular P and K. Further, I identified specific microbial taxa linked to N_2O emissions under different nitrogen fertilization regimes. The choice of N fertilizer will affect the actual N_2O emissions from a soil, which is also determined by the genetic potential for N_2O emissions, e.g. the activity of resident N_2O-metabolizing microbes. Elucidating the drivers of particular microbial groups, e.g. non-N_2O producers vs N_2O producers, given the soil physicochemical levels, will enable targeted management of N_2O reduction. Here I further discuss the link between advances in sequencing technology and soil microbial ecology research as demonstrated here.

7.1 General effects of N fertilization on the soil microbial and plant communities

Under N limitation, as in natural ecosystems, plant growth is dependent upon decomposition by soil microbes of N and other nutrients bound in plant litter or soil organic matter to bioavailable forms for plant uptake (LeBauer & Treseder 2008, Aislabie & Deslippe 2013). Moreover, soil microbial communities are structured in part by plant litter and plant root deposits (Berg & Smalla 2009). This dependence between plants and soil microbes, the so-called plant-soil feedback, encompasses the ecological relationships between plants and soil microbes that can lead to species co-evolution (van der Heijden et al 2008, van der Putten et al 2013, van Nuland et al 2016, ter Horst & Zee 2016). Fertilization can upset nutrient-mediated plant-microbe symbioses by removing this nutrient limitation (Wall et al 2015). Under N fertilization, the resulting excess N availability is thought by some to decrease the plant dependence on nutrients made bioavailable by soil microbial decomposition, furthermore reducing soil microbial diversity; however, this is still under investigation by the field (Thiele-Bruhn et al 2012, Bommarco et al 2013).

In practice, plant productivity is limited by N, P and K as these macronutrients are required for growth; subsequently, these macronutrients are often added

as fertilizers for crops, which can affect existing beneficial plant-soil microbe dependencies. In Chapter 2, fertilization with NPK but not N alone affected plant community composition and diversity. The plant community in the NPK plots differed from that of the control plots in the dominance of fast-growing plant species able to cope with the high NPK inputs without growth of the medium- and slow-growing species, which led to an overall lower plant diversity. Because there was no concomitant change in the bacterial community composition in the NPK plots, we concluded that indeed, the plant and bacterial communities were disassociated due to the N input. In other words, the plant communities in the N treatment were likely limited by the lack of P and K, which in turn limited their proliferation. Moreover, the N treatment resulted in a large effect in on the soil bacterial but not the plant community, suggesting that the bulk of the added N was used by the bacterial community, resulting in their proliferation and differentiation from the bacterial communities in the control plots. In contrast to the soil bacterial community, it appeared that the soil fungal community composition co-varied with that of the plant community, with a difference in the NPK but not the N treatment (Chapter 2). This suggested that there was co-dependency between the plant species and fungal phyla in these plots, which was not affected by the long-term nutrient additions. Fungi and plant co-dependences are well-known in that fungi (e.g. arbuscular mycorrhizal fungi) form mutualistic relationships with some plant species (Vályi et al 2015). It is possible that the co-dependencies between the grasses and fungi in these plots were more robust than those of the plant and bacteria.

The microbial mining hypothesis offers a framework to interpret how the increased N to the bacterial community might lead to changes in the bacterial community composition, as I observed in Chapter 3. Organic N, or N stored in organic matter, is thought to be the main source of N for microbial growth and maintenance (Parfitt et al., 2005). Thus, microbial decomposers will usually mineralize organic N into ammonium-N to access this nitrogen. However, the N fertilization provides available N so that microbes best equipped to handle the high N levels succeed and increase their abundance in the community. We observed this in the bacterial community in the N fertilized plots in Chapter 3 as a shift to so-called copiotrophic phyla under N saturation. The copiotroph– oligotroph trade-off (Fierer et al 2007) seemed to explain our result of increased abundances of Actinobacteria and decreased abundances of Acidobacteria, Verrucomicrobia and Firmicutes in the N-saturated fields. Oligotrophs are described as those that grow slower but more efficiently and succeed during resource limitation, while copiotrophs have fast growth rates and inefficiently transform resources, hence doing better during higher resource availability (Roller & Schmidt 2015). While Acidobacteria, Firmicutes, and Verrucomicrobia are regarded as oligotrophs, Acti-

nobacteria are regarded as copiotrophs (Fierer et al 2012, Leff et al 2015). While this hypothesis is a good framework to interpret results, it is clear that this is a very simplistic approach, as within a bacterial phylum there can be copiotrophic and oligotrophic populations, perhaps even in the same genus, as evidenced by the large range of organic substrate degradation by different soil microbial species (Goldfarb et al 2011). Further, in Chapter 3 I observed that based on the metagenomic analysis, the functional potential of the soil bacterial communities did not differ based on treatment, in contrast to the taxonomic composition. This lack of difference of functional profiles, at least at a coarse-grained level, seems to be widely applicable to bacterial communities in soil (Fierer et al 2012), and suggests functional redundancy provided by different taxa. However, fine-grained analyses might reveal the specific differences in the functional potential of bacterial communities in N- and NPK- saturated plots.

7.2 Microbial populations involved in N_2O metabolism under N fertilization

Here, I add to the agricultural and microbial ecology literature cementing the idea that nitrogen addition greatly influences the soil microbial community, and further describe specific microbial taxa influenced by different N sources. In Chapter 3, the effect of the long-term inorganic fertilizations on the soil bacterial community were quite large, and the differences in the nitrogen plots clearly visible even at the phylum level. Going to Chapter 5, this chapter showcased differences at the OTU-level between the nitrogen source and control plots. The effect of nitrogen input occurs first on individual OTUs by promoting certain species, which then allows these taxa to succeed over time. Fertilization with urea appeared to select for certain groups as per the microbial mining hypothesis described previously. For example, the ammonia-oxidizing bacterial *Nitrosospira*-like population that was correlated with N_2O emissions also appeared to respond to the nitrification inhibitors co-applied with the urea. Further, also in Chapter 5, the native population appeared to be an ammonia-oxidizing archaeal *Nitrososphaera*-like population. Sugarcane agriculture practices in Brazil aim to improve nitrogen use efficiency, although this is a challenge due to the tropical climate, which provides high volumes of rainfall which contribute to nitrogen loss through NO_3^- leaching and N_2O gas production (Otto et al 2016). As seen in Chapters 4 and 5, no N application in sugarcane soils were found to result in poor N availability; the resident community of nitrifiers in the N unfertilized plots was adapted to low N conditions, being ammonia-oxidizing archaea and nitrite-oxidizing bacteria. Many studies also identified ammonia-oxidizing bacteria rather than archaea as responsible for N_2O emissions in agricultural soils (Wang et al 2016,

Meinhardt et al 2018). Recent evidence suggests that this trend of archaea producing less N_2O than bacteria is generally found, which boils down to the differences in their biochemical pathways (Stieglmeier et al 2014, Hink et al 2017, Jia & Conrad 2009). This is thought to be due to the ammonia preferences of ammonia-oxidizing archaea vs bacteria, with the former having higher affinity to ammonia and therefore preferring low ammonia availability, and the latter having lower affinity, thus preferring higher ammonia availability (Hink et al 2018).

Interestingly, as found in Chapter 6 mainly putative denitrifiers were found in the metagenome-assembled genomes of vinasse bacteria. Yang et al (2013) found that Actinobacteria were stimulated under vinasse fertilization, suggesting that copiotrophs might be stimulated by mainly the organic compounds found in vinasse. Measuring the precise nutrient conditions is vital for further research on the dynamics of N_2O emissions in soils under sugarcane. Another important point from the system of vinasse fertirrigation was our finding of potential antibiotic-resistance genes in the vinasse bacterial genomes. This is inferred to be due to the antibacterial procedure used during bioethanol distillation, but which can have serious environmental effects once used in fertirrigation (Braga et al 2017). Namely, there is a potential for the spread of antibiotic-resistance genes in the soil microbial community in sugarcane soils as occurs when antibiotic-treated livestock waste is used as fertilizer for crops (Thiele Bruhn et al 2003, Tasho et al 2016). This is an important task for future research into the environmental and health effects of vinasse fertirrigation.

Here I suggest that teasing out the drivers at a finer-grained scale can help us to further describe the ecology of all microbes involved in N_2O emissions in a system. We demonstrated in Chapter 2 that measuring micronutrients as well as macronutrients pointed to a correlation between the soil bacterial community compositions in the N fertilized plots with Fe, Al, Mg and Mn. This represents a link between abiotic and biotic ecosystem components. While the cost of metabolomics and proteomics of soil samples remains prohibitive, measuring full suites of micronutrients along with macronutrients might allow us to connect microbial populations with their drivers. Specifically, this could aid in teasing out the ecological niches of different N_2O-producing microbes and lead to designing fertilizer schemes that minimize nitrogen loss and increase nitrogen fertilizer efficiency.

7.3 Sequencing technologies paved the way for soil microbial research

Advances in next-generation sequencing technology led to decreasing costs of sequencing, widespread use and increased representation of soil microbes in sequence databases. While the cost of sequencing a bacterial genome of about 5 Mbp at 100X coverage was about EU 100 in 2012, that cost dwindled to about EU 2 in the present day (Köser et al 2012, Deurenberg et al 2017). These reductions in cost have been largely driven by clinical microbiology, but also studies of plant-associated microbes. This reduction in cost has also led to overall improvement in the field of bioinformatics, as more comprehensive studies are possible when more genomic data is deposited to the database. Further, the application of metagenomics to improving culturing conditions of microbes has led to increasing the "unculturable" fraction of represented microbes, although this reduction has been mainly in human-related microbes (Lagier et al 2012). Soil microbial communities, which reach up to 10^9 cells and between 10^4-10^6 species in one gram of soil, initially presented a challenge to fully sequence (Roesch et al 2007, Schloss & Handelsman 2006). In studies of soil microbial communities, this means that the comprehensive sequencing of the full diversity of these communities can be realized. Further, more soil genomic information represented in the databases strengthens future research of soil microbial communities by providing reference sequences with which to compare unknown sequences. In this way, researchers can move from coarse- to finer-grained analyses, and to investigate the dynamics of microbial populations.

In the present thesis, my chapters span a course of about five years, in which the advances in bioinformatics tools can be seen. The soil microbial communities in long-term fertilized grassland or sugarcane soil were evaluated at the phylum-level (coarse-grained analysis) using *16S* and *18S rDNA* amplicon (Chapter 2 & 4) and shotgun metagenomics (Chapter 3). In Chapter 3, 454 pyrosequencing was applied, while in Chapters 4 and 5 Ion Torrent was used and in Chapter 6 Illumina MiSeq sequencing was used. A fine-grained analysis (species or OTU-level) was used to investigate the ammonia-oxidizing microbial community in sugarcane soils (Chapter 5) as well as the vinasse assemblage using shotgun metagenomics and metagenome assembled genomes (Chapter 6). Bioinformatic analyses are especially useful in generating testable hypotheses that future research can address. Thus, this consideration highlights the need to study microbially-mediated events using relevant molecules at relevant time scales. For example, RNA sequencing is a good tool to examine the link between expression of N_2O-producing genes from microbes and N_2O emission peaks in the field (Theodorakopoulos et al 2017). Here I showed that the coarse-grained soil bacterial community structure was sus-

ceptible to long-term (Chapter 3) but not short-term nitrogen addition (Chapter 4) in the forms of ammonium nitrate and urea, respectively. While these studies were regarding the bacterial communities of different soils, these are considered to be comparable due to the inclusion of a control treatment in each experiment. This aspect brings up the point that experiments involving microbial community research should be carefully designed, with the inclusion of a control treatment and further, enough biological replicates to provide good statistical power, eg. a recent meta-analysis of global bacterial and fungal diversity using 189 sites and 7,560 sub-samples (Bahram et al 2018). Here, changes in OTU abundances were visible over one season in the field (Chapters 4 and 5), while differences in phylum abundances were visible after a 60+ year experiment (Chapter 3).

Because there are many pathways leading to N_2O in agricultural and natural systems it is a challenge to uncover mechanistic details regarding the microbes in control of these emissions. However, the combination of improved sequencing technologies, sequencing effort, informed experimental design and improved representation of soil microbes in public databases is closing this knowledge gap. Further, more detailed measurements of soil and environmental factors combined with gases emissions will enable us to get fuller pictures of the complex processes studied. While here I focused on nitrification and denitrification as the main sources of nitrogen fertilizer-derived N_2O emissions, it is important in the context of research into N_2O-emission reduction that all sources be considered when designing future projects. As pointed out in a recent review, microbes mediating N transformations are both metabolically and taxonomically diverse; thus, research taking microbes into account when detailing the nitrogen cycle should incorporate this diversity (Kuypers et al 2018). For investigating the role of microbes in N transformations in agricultural soils, this means zooming into the details, asking questions such as, “which functional pathway is dominant?” and further, “what are the drivers of the microbes involved in these pathways?”

7.4 References

Aislabie, J., Deslippe, J.R. and Dymond, J., 2013. Soil microbes and their contribution to soil services. Ecosystem services in New Zealand–conditions and trends. Manaaki Whenua Press, Lincoln, New Zealand, pp.143-161.

Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., Bengtsson-Palme, J., Anslan, S., Coelho, L.P., Harend, H., Huerta-Cepas, J., Medema, M.H., Maltz, M.R., Mundra, S., Olsson, P.A., Pent, M., Põlme, S., Sunagawa, S., Ryberg, M., Tedersoo, L., Bork, P., 2018. Structure and function of the global topsoil microbiome. Nature 560, 233.

Berg, G., Smalla, K., 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. FEMS Microbiol Ecol 68, 1–13.

Braga, L.P.P., Alves, R.F., Dellias, M.T.F., Navarrete, A.A., Basso, T.O., Tsai, S.M., 2017. Vinasse fertirrigation alters soil resistome dynamics: an analysis based on metagenomic profiles. BioData Mining 10, 17.

Deurenberg, R.H., Bathoorn, E., Chlebowicz, M.A., Couto, N., Ferdous, M., García-Cobos, S., Kooistra-Smid, A.M.D., Raangs, E.C., Rosema, S., Veloo, A.C.M., Zhou, K., Friedrich, A.W., Rossen, J.W.A., 2017. Application of next generation sequencing in clinical microbiology and infection prevention. Journal of Biotechnology 243, 16–24.

Fierer, N., Bradford, M.A., Jackson, R.B., 2007. Toward an Ecological Classification of Soil Bacteria. Ecology 88, 1354–1364.

Fierer, N., Lauber, C.L., Ramirez, K.S., Zaneveld, J., Bradford, M.A., Knight, R., 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. The ISME Journal 6, 1007–1017.

Goldfarb, K.C., Karaoz, U., Hanson, C.A., Santee, C.A., Bradford, M.A., Treseder, K.K., Wallenstein, M.D., Brodie, E.L., 2011. Differential Growth Responses of Soil Bacterial Taxa to Carbon Substrates of Varying Chemical Recalcitrance. Front. Microbiol. 2.

Hink, L., Gubry-Rangin, C., Nicol, G.W., Prosser, J.I., 2018. The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. The ISME Journal 12, 1084.

Hink, L., Nicol, G.W., Prosser, J.I., 2017. Archaea produce lower yields of N_2O than bacteria during aerobic ammonia oxidation in soil. Environmental Microbiology 19, 4829–4837.

Jia, Z., Conrad, R., 2009. Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. Environmental Microbiology 11, 1658–1671.

Köser, C.U., Ellington, M.J., Cartwright, E.J.P., Gillespie, S.H., Brown, N.M., Farrington, M., Holden, M.T.G., Dougan, G., Bentley, S.D., Parkhill, J., Peacock, S.J., 2012. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. PLOS Pathogens 8, e1002824.

Kuypers, M.M.M., Marchant, H.K., Kartal, B., 2018. The microbial nitrogen-cycling network. Nature Reviews Microbiology 16, 263–276.

Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., Bittar, F., Fournous, G., Gimenez, G., Maraninchi, M., Trape, J.-F., Koonin, E.V., Scola, B.L., Raoult, D., 2012. Microbial culturomics: paradigm shift in the human gut microbiome study. Clinical Microbiology and Infection 18, 1185–1193.

LeBauer, D.S., Treseder, K.K., 2008. Nitrogen Limitation of Net Primary Productivity in Terrestrial Ecosystems Is Globally Distributed. Ecology 89, 371–379.

Leff, J.W., Jones, S.E., Prober, S.M., Barberán, A., Borer, E.T., Firn, J.L., Harpole, W.S., Hobbie, S.E., Hofmockel, K.S., Knops, J.M.H., McCulley, R.L., Pierre, K.L., Risch, A.C., Seabloom, E.W., Schütz, M., Steenbock, C., Stevens, C.J., Fierer, N., 2015. Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. PNAS 112, 10967–10972.

Meinhardt, K.A., Stopnisek, N., Pannu, M.W., Strand, S.E., Fransen, S.C., Casciotti, K.L., Stahl, D.A., 2018. Ammonia-oxidizing bacteria are the primary N_2O producers in an ammonia-

oxidizing archaea dominated alkaline agricultural soil. Environmental Microbiology 20, 2195–2206.
Otto, R., Castro, S.A.Q., Mariano, E., Castro, S.G.Q., Franco, H.C.J., Trivelin, P.C.O., 2016. Nitrogen Use Efficiency for Sugarcane-Biofuel Production: What Is Next? Bioenerg. Res. 9, 1272–1289.
Parfitt, R.L., Yeates, G.W., Ross, D.J., Mackay, A.D., Budding, P.J., 2005. Relationships between soil biota, nitrogen and phosphorus availability, and pasture growth under organic and conventional management. Applied Soil Ecology 28, 1–13.
Putten, W.H. van der, Bardgett, R.D., Bever, J.D., Bezemer, T.M., Casper, B.B., Fukami, T., Kardol, P., Klironomos, J.N., Kulmatiski, A., Schweitzer, J.A., Suding, K.N., Voorde, T.F.J.V. de, Wardle, D.A., 2013. Plant–soil feedbacks: the past, the present and future challenges. Journal of Ecology 101, 265–276.
Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., Triplett, E.W., 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME Journal 1, 283–290.
Roller, B.R., Schmidt, T.M., 2015. The physiology and ecological implications of efficient growth. The ISME Journal 9, 1481–1487.
Schloss, P.D., Handelsman, J., 2006. Toward a Census of Bacteria in Soil. PLOS Computational Biology 2, e92.
Stieglmeier, M., Mooshammer, M., Kitzler, B., Wanek, W., Zechmeister-Boltenstern, S., Richter, A., Schleper, C., 2014. Aerobic nitrous oxide production through N-nitrosating hybrid formation in ammonia-oxidizing archaea. The ISME Journal 8, 1135–1146.
Tasho, R.P., Cho, J.Y., 2016. Veterinary antibiotics in animal waste, its distribution in soil and uptake by plants: A review. Science of The Total Environment 563–564, 366–376.
terHorst, C.P., Zee, P.C., 2016. Eco-evolutionary dynamics in plant–soil feedbacks. Functional Ecology 30, 1062–1072.
Theodorakopoulos, N., Lognoul, M., Degrune, F., Broux, F., Regaert, D., Muys, C., Heinesch, B., Bodson, B., Aubinet, M., Vandenbol, M., 2017. Increased expression of bacterial amoA during an N_2O emission peak in an agricultural field. Agriculture, Ecosystems & Environment 236, 212–220.
Thiele-Bruhn, S., Bloem, J., de Vries, F.T., Kalbitz, K., Wagg, C., 2012. Linking soil biodiversity and agricultural soil management. Current Opinion in Environmental Sustainability, Terrestrial systems 4, 523–528.
Vályi, K., Rillig, M.C., Hempel, S., 2015. Land-use intensity and host plant identity interactively shape communities of arbuscular mycorrhizal fungi in roots of grassland plants. New Phytologist 205, 1577–1586.
Wang, Q., Zhang, L.-M., Shen, J.-P., Du, S., Han, L.-L., He, J.-Z., 2016. Nitrogen fertiliser-induced changes in N_2O emissions are attributed more to ammonia-oxidising bacteria rather than archaea as revealed using 1-octyne and acetylene inhibitors in two arable soils. Biol Fertil Soils 52, 1163–1171.
Yang, S.-D., Liu, J.-X., Wu, J., Tan, H.-W., Li, Y.-R., 2013. Effects of Vinasse and Press Mud Application on the Biological Properties of Soils and Productivity of Sugarcane. Sugar Tech 15, 152–158.

Summary

Summary

The use of N fertilizers has increased worldwide in the past century. While this increased input of N has increased food productivity, it has also contributed to decreases in biodiversity, soil quality and environmental health, including increases in greenhouse gas emissions. These emissions in agricultural soils are largely carried out by the soil microbiome, or the microorganisms living in the soil and transforming N fertilizers to different forms. Here, the overall research aim was to gain detailed insight into the effects of nitrogen fertilizer schemes, including long term fertilization, on soil microbial communities. To do this, I applied next-generation sequencing technology and associated bioinformatics analyses to field experiments in the Netherlands and in Brazil.

In Chapter 2 I found that long-term fertilization with lime, ammonium nitrate (N), super phosphate (P) and NPK resulted in "habitats" with different soil conditions in Dutch hay meadows. In these different fertilizer habitats, the taxonomic makeup of the soil bacterial community, or the community composition and diversity differed from the control plots by the varying abundance of the different taxa; however, potential functions did not differ between treatments. This suggests that each treatment affected the soil bacterial community as a whole by being associated with altered abundances of taxa, but not at a broad functional level, with the gene potential remaining the same across treatments. Nitrogen fertilization alone resulted in the most significant changes in the soil bacterial community, in which the abundances of the Actinobacteria phyla were higher compared to the other treatments. This is in line with the results from other studies, which identifies the Actinobacteria as a copiotrophic phyla, or a taxa well able to succeed under high-nutrient conditions. This suggests differences in decomposition rates in each habitat, with decomposition and associations by the bacterial community affected especially under N fertilization.

In Chapter 3, I extended the investigation from Chapter 2 to include the plant and soil fungal communities in the long-term fertilized fields. I found that the plant and soil fungal but not the plant and soil bacterial and soil fungal and bacterial communities varied similarly across the plots. The plant community was less diverse in the NPK compared to the other treatments, which was thought to be due to the success of high-nutrient-adapted grasses in the plant communities in this habitat. In addition, the liming treatment was associated with higher diversity in all three communities, which was likely due to the increased availability of nutrients from the higher pH. Regarding the plant and fungal community compositions, these were different in the NPK plots compared to the control plots, and our co-variation analysis supported their interdependence, with different potential co-

varying taxonomic groups identified. This suggested that the plant and soil fungal communities are closely interconnected in these fields, suggesting more ecological connections between these communities compared to either with the soil bacterial community. Further, the bacterial community appeared not to co-vary with either the plant nor the fungal communities, even though the macroorganisms are known to have specific associated bacterial communities. These results point to interesting questions in the differences in the ecological communities in each habitat due to interconnected effects of the fertilizers and nutrient availability on the plant and soil microbial communities.

In Chapter 4 and 5 I investigated the short-term effects of nitrogen fertilizers on the soil microbial community and N_2O emissions. Urea and urea with nitrification inhibitor treatments were evaluated over 256 days for the effect on the bacterial communities based on *16S rDNA* sequencing (Chapter 4), and on the ammonia-oxidizing subset of the bacterial community based on *amoA* sequencing (Chapter 5). Overall bacterial community compositions were unaffected by treatment, at least at phylum-level taxonomic resolution based on DNA. Furthermore, the bacterial community diversity was not affected by treatment. Moreover, we concluded that the nitrification inhibitors successfully inhibited the N_2O emissions. When looking specifically at the *amoA*-containing bacterial (AOB) OTUs, however, the abundances of different ammonia-oxidizing bacterial species indeed differed between treatments. Further, we identified a Nitrosospira-like AOB as the likely N_2O-emitter in the plots under urea fertilization, and also showed that the abundances of this bacteria decreased in the treatments with the nitrification inhibitors DMPP and DCD. These chapters indicated that short-term nitrogen fertilization with urea did not affect the soil bacterial community composition at a high taxonomic resolution, but did lead to differences at the OTU-level. Further, the ammonia-oxidizing bacterial community had low diversity in these soils, which might be due to the low levels of ammonia that are normally present. Most interestingly, we identified cohorts of ammonia- and nitrite-oxidizing OTUs that were associated with different soil factors, suggesting a picture of the nitrifying bacterial community in these soils.

In Chapter 6 I detailed for the first time the bacterial assemblage of sugarcane vinasse, which is widely used as a potassium fertilizer in conjunction with nitrogen fertilizers during sugarcane management. Targeting concerns over greenhouse gas emissions during the fertirrigation practice, I evaluated the potential presence of genes from the main N_2O-producing microbial pathways from 21 metagenome-assembled vinasse bacterial genomes (MAGs). The main genera I uncovered from the vinasse MAGs were Lactobacillus, Megasphaera and Mit-

suokella, and these had mainly denitrification gene potential. Interestingly, these had varying presence of denitrification genes, suggesting that the different vinasses used as fertilizer can be a source of bacteria with different N_2O-producing gene potential, potentially influencing the actual emissions of N_2O. Further, the potential presence of antibiotic-resistance genes was found across almost all the MAGs; this reinforced the idea that the bacterial component of vinasse is mainly the contaminants of the bioethanol production cycle, which often includes an antibacterial sterilizing step. Further, this result raises the visibility of the potential for horizontal gene transfer potential in sugarcane soil and subsequent risks for public health and crop productivity. This work also highlighted the necessity to include measures of the varying biotic component of vinasse in research of greenhouse gases from this system.

In conclusion, metagenomic and bioinformatic analyses were applied to data from field experiments investigating the effects of nitrogen deposition on the soil microbiota and related soil physicochemical factors. This thesis sheds light on the complex and interconnected changes within the soil microbial community upon nitrogen fertilization. Two separate studies in the Netherlands and Brazil demonstrated the effect of long-term nitrogen deposition on the soil bacterial community at coarse taxonomic levels, as well as the effect of short-term fertilization on soil microbes at the OTU level. This research underscores the idea that nitrogen fertilization in fields affects not only crop production, but the soil microbial community composition and the ecology of the communities in these fields, which can have long-term effects on crop productivity. Last, genome binning from bacterial DNA sequences extracted from sugarcane vinasse revealed 21 potential bacterial contaminants of the bioethanol production process. Since vinasse is widely used as a fertilizer, especially in conjunction with nitrogen fertilizers, for sugarcane in Brazil, this research paved the way for future studies linking the genetic potential of vinasse bacteria, vinasse and nitrogen fertilization, and field emissions of N_2O.

Samenvatting

Samenvatting

Het gebruik van N-meststoffen is de afgelopen eeuw wereldwijd toegenomen. Hoewel deze verhoogde input van N de voedselproductiviteit heeft verhoogd, heeft deze ook bijgedragen aan een afname van de biodiversiteit, de bodemkwaliteit en de milieukwaliteit, waaronder een toename van de uitstoot van broeikasgassen. In landbouwgronden is het microbioom in de bodem grotendeels verantwoordelijk voor deze emissies, met name de micro-organismen die in de bodem leven en N-meststoffen omzetten. Het algemene onderzoeksdoel van mijn studie was om gedetailleerd inzicht te krijgen in de effecten van stikstofkunstmest, waaronder langdurige bemesting, op microbiële gemeenschappen in de bodem. Om dit te doen, paste ik next generation sequencing technieken en bijbehorende bioinformatica-analyses toe op veldexperimenten in Nederland en Brazilië.

In hoofdstuk 2 ontdekte ik dat langdurige bemesting met kalk, ammoniumnitraat (N), superfosfaat (P) en NPK resulteerde in "habitats" met verschillende bodemgesteldheden in Nederlandse hooilanden. In deze verschillende meststofhabitats verschilden de taxonomische samenstelling van de bodembacteriegemeenschap, of de gemeenschapssamenstelling en diversiteit van de controle behandelingdoor de variërende abundantie van de verschillende taxa; de potentiële functies verschilden echter niet. Dit suggereert dat in elke behandeling de bodembacteriegemeenschap als geheel beïnvloed word door de associatie met veranderde taxa abundanties, maar dat de behandeling niet op een breed functioneel niveau invloed had, waarbij het genpotentieel hetzelfde bleef in alle behandelingen. De stikstofbemesting alleen resulteerde in de meest significante veranderingen in de bacteriële bodemgemeenschap, waarin de abundanties van de Actinobacteria phyla hoger waren in vergelijking met de andere behandelingen. Dit is overeenstemmend met de resultaten van andere studies, die de Actinobacteriën identificeren als een copiotroof phylum, of taxa die goed kunnen slagen onder nutrientenrijke omstandigheden. Dit suggereert dat er verschillen zijn tussen de afbraaksnelheden in elke habitat en dat deze afbraaksnelheid en de samenstelling van de bacteriële gemeenschap beïnvloed worden door de N-bemesting.

In hoofdstuk 3 heb ik het onderzoek van hoofdstuk 2 uitgebreid naar de plant- en bodemschimmelgemeenschappen in lange termijn bemestingsexperimenten. Ik ontdekte dat de planten- en bodem schimmelgemeenschap, maar niet de plant en bodem bacteriegemeenschap en de bodemschimmel- en bodembacteriegemeenschap op vergelijkbare wijze over de percelen varieerden. De plantengemeenschap was minder divers in de NPK in vergelijking met de andere behandelingen, wat mogelijk het gevolg was van het succes van grassen met een hoge voedingsstof-aanpassing in de plantengemeenschappen in dit leefgebied. Bovendien ging de kalkbehandeling gepaard met een hogere diversiteit in alle drie de gemeenschappen, waarschijnlijk als gevolg van de verhoogde beschikbaarheid van voedingsstoffen door de hogere pH. Met betrekking tot de plant- en schimmelgemeenschapsamenstellingen waren deze verschillend in de NPK-plots in vergelijking met de controleplots en onze co-variatieanalyse onderschreef onderlinge afhankelijkheid, waarbij verschillende potentiële co-variabele taxonomische groepen werden geïdentificeerd. Dit suggereerde dat de plant- en bodemschimmelgemeenschappen nauw met elkaar verbonden zijn in deze velden, wat duidt op meer ecologische verbanden tussen deze gemeenschappen in vergelijking met de bacteriële bodemgemeenschap en beide gemeenschappen. Verder bleek de bacteriegemeenschap niet te co-variëren met ofwel de plant noch de

schimmelgemeenschappen, hoewel het bekend is dat de macro-organismen specifieke geassocieerde bacteriële gemeenschappen hebben. Deze resultaten wijzen op interessante vragen in de verschillen in de ecologische gemeenschappen in elk leefgebied als gevolg van onderling verbonden effecten van de meststoffen en beschikbaarheid van voedingsstoffen op de microbiële gemeenschappen van planten en de bodem.

In hoofdstuk 4 en 5 heb ik de korte termijn effecten van stikstofkunstmeststoffen op de microbiële bodemgemeenschap en N_2O-emissies onderzocht. Ureum en ureum met nitrificatie-inhibitorbehandelingen werden gedurende 256 dagen beoordeeld op het effect op de bacteriële gemeenschappen op basis van 16S rDNA-sequencing (Hoofdstuk 4) en op de ammoniak-oxiderende subset van de bacteriële gemeenschap op basis van *amoA*-sequencing (Hoofdstuk 5). De algemene bacteriële samenstelling van de gemeenschap werd niet beïnvloed door de behandeling, tenminste niet op taxonomische resolutie op basis van DNA op phylum-niveau. Bovendien werd de diversiteit van de bacteriële gemeenschap niet beïnvloed door de behandeling. Bovendien concludeerden we dat de nitrificatie-inhibitoren met succes de N_2O-emissies remden. Specifiek kijkend naar de *amoA*-bevattende bacteriële (AOB) OTU's, verschilden de abundanties van verschillende ammoniak-oxiderende bacteriesoorten inderdaad tussen de behandelingen. Verder identificeerden we een Nitrosospira-achtige AOB als de waarschijnlijke N_2O-emitter in de plots met ureumbemesting, en toonden ook aan dat de abundanties van deze bacterie afnamen in de behandelingen met de nitrificatie-inhibitoren DMPP en DCD. Deze hoofdstukken gaven aan dat korte termijn stikstofbemesting met ureum de samenstelling van de bodembacterie met een hoge taxonomische resolutie niet beïnvloedde, maar wel leidde tot verschillen op OTU-niveau. Verder had de ammoniak-oxiderende bacteriële bodemgemeenschap een lage diversiteit in deze bodems, wat mogelijk te wijten is aan de lage niveaus van ammoniak die normaal aanwezig zijn. Interessant genoeg identificeerden we cohorten van ammoniak- en nitriet-oxiderende OTU's die geassocieerd waren met verschillende bodemfactoren, wat een beeld geeft van de nitrificerende bacteriële gemeenschap in deze bodems.

In hoofdstuk 6 heb ik voor het eerst de bacteriële assemblage van suikerriet vinasse beschreven, die veel wordt gebruikt als kaliummeststof in combinatie met stikstofmeststoffen tijdens het beheer van suikerriet. Ik richtte mijn aandacht op zorgen over de uitstoot van broeikasgassen tijdens de fertirrigatiepraktijk en evalueerde de mogelijke aanwezigheid van genen van de belangrijkste N_2O-producerende microbiële routes van 21 metagenoom-geassembleerde vinasse-bacteriegenen (MAG's). De belangrijkste geslachten die ik heb geidentificeerd uit de vinasse MAGs waren Lactobacillus, Megasphaera en Mitsuokella en deze hadden voornamelijk denitrificatie-genpotentiaal. Interessant is dat de aanwezigheid van denitrificerings-genen varieerde, wat suggereert dat de verschillende vinasses die als kunstmest worden gebruikt een bron van bacteriën met verschillende N_2O-producerende genpotentiëlen kunnen zijn, die mogelijk de feitelijke emissies van N_2O beïnvloeden. Verder werd de potentiële aanwezigheid van antibioticaresistentiegenen gevonden in bijna alle MAG's; dit versterkte het idee dat de bacteriële component van vinasse hoofdzakelijk de bacteriele verontreinigingen van de productiecyclus van bio-ethanol is, die vaak een antibacteriële sterilisatiestap omvat. Verder verhoogt dit resultaat de zichtbaarheid van het potentieel voor horizontale genoverdracht in de suikerriet bodem en de daaruit voortvloeiende risico's voor de volksgezondheid en de gewasproductiviteit. Dit werk benadrukte ook de noodzaak om maatregelen van de variërende biotische component van vinasse op te nemen in het onderzoek naar broeikasgassen uit dit systeem.

Concluderend werden metagenomische en bioinformatische analyses toegepast op gegevens uit veldexperimenten die de effecten van stikstofdepositie op de bodemmicrobiota en gerelateerde bodemfysisch-chemische factoren onderzoeken. Dit proefschrift belicht de complexe en onderling verbonden veranderingen in de microbiële bodemgemeenschap na stikstofbemesting. Twee afzonderlijke studies in Nederland en Brazilië toonden het effect aan van langdurige stikstofdepositie op de bacteriële bodemgemeenschap op hogere taxonomische niveaus, evenals het effect van kortdurende bemesting op bodemmicroben op OTU-niveau. Dit onderzoek onderstreept het idee dat stikstofbemesting op het veld niet alleen de productie van gewassen beïnvloedt, maar ook de samenstelling van de microbiële gemeenschap in de bodem en de ecologie van de gemeenschappen in deze gebieden, wat op lange termijn gevolgen kan hebben voor de gewasproductiviteit. Tenslotte onthulde genome-binning van bacteriële DNA-sequenties geëxtraheerd uit suikerrietvinasse 21 potentiële bacteriële verontreinigingen van het productieproces van bioethanol. Aangezien vinasse veel wordt gebruikt als meststof, vooral in combinatie met stikstofhoudende meststoffen, voor suikerrietteeltin Brazilië heeft dit onderzoek de weg geëffend voor toekomstige studies die het genetische potentieel van vinasse-bacteriën, vinasse en stikstofbemesting en veldemissies van N_2O met elkaar verbinden.

Publications

Publications in the thesis:

Pan Y*, Cassman N*, de Hollander M, Mendes LW, Korevaar H, Geerts RH, van Veen JA, Kuramae EE. Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS microbiology ecology*. 2014 Oct 1. 90(1):195-205.

Cassman NA, Leite MF, Pan Y, De Hollander M, Van Veen JA, Kuramae EE. Plant and soil fungal but not soil bacterial communities are linked in long-term fertilized grassland. *Scientific reports*. 2016 Mar 29. 6:23680.

Soares JR*, Cassman NA*, Kielak AM, Pijl A, Carmo JB, Lourenço KS, Laanbroek HJ, Cantarella H, Kuramae EE. Nitrous oxide emission related to ammonia-oxidizing bacteria and mitigation options from N fertilization in a tropical soil. *Scientific reports*. 2016 Jul 27. 6:30349.

Cassman NA, Lourenço KS, Carmo JB, Cantarella H, Kuramae EE. Genome-resolved metagenomics of sugarcane vinasse bacteria. *Biotechnology for biofuels*. 2018 Dec. 11(1):48.

Cassman NA, Soares JR, Pijl A, Lourenço KS, van Veen JA, Cantarella H, Kuramae EE. Nitrification inhibitors effectively target N_2O−producing Nitrosospira spp. in tropical soil. *Environmental microbiology*. 2019 Feb 8.

* indicates shared first authorship

Other publications (reverse chronological order):

Lourenco KS, **Cassman NA**, Pijl A, van Veen JA, Cantarella H, Kuramae EE. Nitrosospira sp. govern nitrous oxide emissions in a tropical soil amended with residues of bioenergy crop. *Frontiers in microbiology*. 2018;9:674.

Navarrete AA, Tsai SM, Mendes LW, Faust K, de Hollander M, **Cassman NA**, Raes J, van Veen JA, Kuramae EE. Soil microbiome responses to the short−term effects of Amazonian deforestation. *Molecular ecology.* 2015 May;24(10): 2433-48.

Matthews TD, Schmieder R, Silva GG, Busch J, **Cassman N**, Dutilh BE, Green D, Matlock B, Heffernan B, Olsen GJ, Hanna LF. Genomic comparison of the closely-related Salmonella enterica serovars Enteritidis, Dublin and Gallinarum. *PLoS One*. 2015 Jun 3;10(6):e0126883.

Edwards RA, Haggerty JM, **Cassman N**, Busch JC, Aguinaldo K, Chinta S, Vaughn MH, Morey R, Harkins TT, Teiling C, Fredrikson K. Microbes, metagenomes and marine mammals: enabling the next generation of scientist to enter the genomic era. *BMC genomics*. 2013 Dec;14(1):600.

Dutilh BE, **Cassman N**, McNair K, Sanchez SE, Silva GG, Boling L, Barr JJ, Speth DR, Seguritan V, Aziz RK, Felts B. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature communications.* 2014 Jul 24;5:4498.

Cassman N*, Prieto−Davó A*, Walsh K, Silva GG, Angly F, Akhter S, Barott K, Busch J, McDole T, Haggerty JM, Willner D. Oxygen minimum zones harbour novel viral communities with low diversity. *Environmental microbiology*. 2012 Nov;14(11):3043-65.

***** indicates shared first authorship

Acknowledgements

Acknowledgements

To the collaborators in Brazil – **Heitor Cantarella, Janaina Braga do Carmo, Juliana Ramos**, **Leonardo, Acacio Navarette** and **Helio Danilo** – thank you very much for your support during my trips to Brazil and your contributions to our various publications. Muita obrigada!

Big thanks go to my co-authors. **Riks**, your body of work admirably presents the path of curiosity-driven science and I look forward to finishing our collaborations. **Jos**, I really appreciated your insightful questions during presentations – yours is the most valuable skill for a scientist to have.

Johnny, Késia and **Leonardo**, without you I would not have a thesis. Thank you for sharing the data from your experiments. **Johnny**, thank you also for sharing your family's home and culture in Ouro Fino, I will always remember the beautiful fazenda and the live sertañejo. **Késia**, you know so much about soil and agriculture and I could always chat with you about N_2O and vinasse at any time. **Leo**, I enjoyed getting to know your intelligent and adventurous nature, thanks for the fun times in Brazil and the Netherlands.

To **Márcio**, fellow nerd, thanks for your statistical expertise and all the nice chats. **Agaat** and **Késia**, thank you for helping me to learn the "real meaning of qPCR," and **Agaat** thanks also for your vast and efficient work on our various sequencing projects. **Roos**, in addition to your work on the chemostat project, thank you for sharing swimming and moss-ing adventures with me.

To **Mattias**, you were there from my very first interview! For the innumerable times I poked my head around the monitor to say, "Hey Mattias…" with an accompanying question, thanks for always giving your attention and a helpful suggestion, not to mention the innumerable lunches and chats about sustainable travel and hobbies.

To my office mates past and present: **Victor de Jager**, thanks for your help with bioinformatics questions and for sharing your photography interest with me. **Fleur**, though you only joined us halfway through my PhD I feel like you have always been there. Thank you for all your friendly words and support. **Kay**, your

great scientific and communication abilities astound me and I am sure you will be a very successful scientist or any career of your choosing.

To the NIOO postdocs, especially **Victor Carrion**, **Irene**, **Mauricio**, **Viviane, Ben, Olaf**, **Annelies**, **Desalegn**, **Adrian**, **Sainur, Chunxu, Nurmi, Anna Kielak, Sang Yoon, Natalia, Lara, Emilia** and **Max,** thank you for your inspiring curiosity, scientific drive and interesting conversations about work and life during lunches and elsewhere. **Jenny Ouyang**, it was great to have you as a neighbor in Utrecht.

To the **Kuramae group**, past and present, and other ME students: **Thiago**, thank you for your infectious enthusiasm and friendship, best of luck back in Brazil. **Afnan**, you always have a calming presence and I see you becoming a great professor, thank you for being there to listen to my complaints. And of course, we are always going to be the Taquaral Survivors! **Ohana**, we share a lot of common interests and it is always nice to converse with you, I wish you lots of luck with the science! Thank you **Je Seun** for your kindness, and I will never forget your amazing artistic talent. Thank you to **Adam, Juan, Marcelo, Adriano, Manoeli, Raul, Valeria, Sarah vdB., Kristin, Anna C., Sytske, Femke, Ruth G.** and **Yani** for the snacks and all the nice conversations.

To **Sabine**, for sharing an interest in English literature; and to **Sarash**, for great conversations about everything and anything. Thanks to the **NIOO and its denizens**, especially to **Eke** for knowing how to save an injured koot, and to **Elly** and **Gerda** for your myriad support with paperwork. For the relief of freeing myself sometimes from the constraints of scientific writing, I thank **Froukje** and the rest of the **NIOOScoop** team. To my student **Stan**, thanks for your inspiring work ethic.

To the Wageningen friends, thanks for your friendship, dinners and the fun times: **Ruth S.**, **Paolo**, **Maaike, Marta**, **Julia, Antonella, Julie, Nico, Kadri** and **Kim.**

Maaike, your friendship has meant a lot to me! Thanks for sharing your creativity and sharing with me your family's farm. May you keep having memorable adventures. **Ruth S.**, my thanks for your friendship and for the fun times in Utrecht and beyond.

To people in other departments, thank you for great conversations and for contributing to the great working environment at NIOO: **Stijn, Jeff, Minghui, Jasper, Kelly, Dedmer, Antica, Peiyu, Wei** and **Tanya.**

To the Utrecht friends – **Frédèrique, Joost, Frerik, Gwen, Robert,** and everyone else – thank you for including me in your lives, your friendship has meant a lot as I integrate as much as I can into Dutch life.

To the friends around the globe, especially **Hannah, Sonny, Sally, Peter, Ben and Carol**, thank you for helping me to develop my voice, which is the crucial instrument not only for communicating science but also for self-expression. To my extended family of **Cassmans**, **Hamamotos** and **Friedmans**, and to the newest family members – to **Pravin** and the **Patel**s, to the **Dutilhs**, **Bloems** and **Ruijses** – thank you for your love and support over the past few years.

Deep thanks go to my family, who have always loved and believed in me. **Sarah**, I am so proud of you as you reach for the sky and beyond. **Eva**, you already adult incredibly well and I can't wait to see what the future holds for you. **Mom**, your work ethic, intelligence and zest for life have always inspired me. **Dad**, it is your love of knowledge and curiosity that I have brought to science.

Dear **Bas**, thanks for sailing your ship next to mine on the vast ocean of life.

Curriculum vitae

Nori was born on July 1, 1986 in Santiago, Chile, as the first of three daughters to a diplomatic family. From 2004 to 2008, she attended the Florida Institute of Technology (Melbourne, Florida, USA), during which time she contributed to research in microbial genetics and colloidal enzymes at Florida Tech and bone density loss at the cellular level at the Kennedy Space Science Center (Cape Canaveral, Florida, USA). She graduated in 2008 with a BSc. in Biochemistry and a minor in Psychology. Following an interest in computer programming, Nori completed her Masters in Bioinformatics and Medical Informatics from San Diego State University (San Diego, California, USA), which included a Masters thesis titled "The functional contributions of integrated phages to marine bacterial communities." Nori moved to the Netherlands in 2013 to work as a Junior Researcher at the Netherlands Institute of Ecology (NIOO-KNAW) in Wageningen, and continued as a PhD Researcher from 2014 to 2018 under the supervision of Prof. dr. Johannes A van Veen and Dr. Eiko Kuramae. The results of Nori's research at the NIOO-KNAW are detailed in this thesis. She lives in Utrecht as of 2013, which at five years and counting is the longest she has lived in one place. She looks forward to professional opportunities in improving environmental and human health, and currently freelances as a bioinformatician. Nori can be contacted at noriko.cassman@gmail.com.