Structure and function of the soil microbiome underlying N$_2$O emissions from global wetlands

Mohammad Bahram$^{1,2,9}$, Mikk Espenberg$^{1,9}$, Jaan Pärn$^{1,9}$, Laura Lehtovirta-Morley$^3$, Sten Anslan$^1$, Kuno Kasak$^1$, Urmas Köljalg$^1$, Jaan Liira$^1$, Martin Maddison$^1$, Mari Moora$^1$, Ülo Niinemets$^4$, Maarja Öpik$^1$, Meelis Pärtel$^1$, Kaido Soosaar$^1$, Martin Zobel$^1$, Falk Hildebrand$^{5,6}$, Leho Tedersoo$^{7,8,9}$ & Ülo Mander$^{1,9}$

Wetland soils are the greatest source of nitrous oxide (N$_2$O), a critical greenhouse gas and ozone depleter released by microbes. Yet, microbial players and processes underlying the N$_2$O emissions from wetland soils are poorly understood. Using in situ N$_2$O measurements and by determining the structure and potential functional of microbial communities in 645 wetland soil samples globally, we examined the potential role of archaea, bacteria, and fungi in nitrogen (N) cycling and N$_2$O emissions. We show that N$_2$O emissions are higher in drained and warm wetland soils, and are correlated with functional diversity of microbes. We further provide evidence that despite their much lower abundance compared to bacteria, nitrifying archaeal abundance is a key factor explaining N$_2$O emissions from wetland soils globally. Our data suggest that ongoing global warming and intensifying environmental change may boost archaeal nitrifiers, collectively transforming wetland soils to a greater source of N$_2$O.
Despite covering only 8% of the terrestrial Earth surface, wetland soils (including gley and peat soils) store one of the largest organic carbon (C) stocks. Microbial degradation of C and nitrogen (N) stocks can lead to substantial releases of greenhouse gases (GHGs), including nitrous oxide (N₂O). N₂O is a potent GHG with a global warming potential 265 times that of CO₂. N₂O is the most important ozone-depleting substance¹. This is particularly alarming as microbial sources of N₂O may shift with environmental changes. Wetland soils are increasingly subject to land-use changes such as afforestation and transformation to agricultural land, both preceded by drainage, with long-term consequences for N₂O emissions². To reduce N₂O emissions from wetland soils, we need a thorough understanding of biogeochemical pathways and critical environmental parameters, which shape the microbial activities underpinning the N cycle and N₂O dynamics.

Microbial processes such as classical denitrification, nitrifier denitrification, and dissimilatory nitrate reduction to ammonia (DNRA) all contribute to N₂O production mainly in anoxic conditions³. By contrast, ammonia oxidation, which is the first step in nitrification, is an aerobic process performed by three groups of ammonia oxidizing microorganisms: canonical ammonia oxidizing bacteria (AOB), ammonia oxidizing archaea (AOA), and complete ammonia oxidizers (comammox Nitrosopira), AOA not only directly produce N₂O, but also provide substrate for denitrification⁴. Yet, little is known about the environmental conditions that favor each process and thereby N₂O production and consumption. AOA may play a pivotal, underexplored role in fueling denitrification and facilitating terrestrial N₂O emissions⁵ in many soil environments.

Here we analyzed 645 wetland soils (Fig. 1a; Supplementary Data 1) to determine how the structure and function of microbial communities contribute to N₂O emissions. Our unique dataset integrated global-scale analysis of functional metagenomes (to estimate relative abundance of N-cycle genes independently of PCR biases), multi-group metabarcoding (bacterial 16S, fungal 18S-ITS rRNA genes), absolute quantification of N-cycle gene abundances, as well as in situ N₂O flux and ex situ potential N₂ production analyses. We further leveraged available genomics data to understand genetic mechanisms underlying N₂O production. We hypothesized that the high N₂O production in global wetland soils is mainly related to the diversity and abundance of nitrifying microbes, and that archaeal nitrifiers, both in terms of absolute and relative abundance to denitifiers, are the most robust and accurate explanatory factor of N₂O emissions from wetland soils globally.

Results and discussion

Global patterns of N₂O fluxes. Our analysis indicated that warmer soils and more intensive land use progressively may enhance N₂O release from wetland soils. N₂O emissions showed exponentially increasing relationships with temperature of the warmest month (Supplementary Fig. 1). In addition, the N₂O emissions were strongly explained by land-use type (r² = 0.364, p < 0.001), with greatest values in the bare soils and lowest in the forest soils (Supplementary Fig. 2). Assessment of environmental determinants of N₂O fluxes revealed that N₂O emissions decline towards higher latitudes (Fig. 1, Supplementary Fig. 1). Contrary to the N₂O emissions, potential N₂ production peaked in the temperate climate in negative correlation with land-use intensity (Supplementary Fig. 3). In agreement with our findings, a recent local warming experiment⁶ and global models⁷ predict an increase in N₂O production in response to warming across various ecosystems.

Relationships of global N₂O fluxes to microbial diversity and taxa. Our analyses of microbial communities of wetland soils revealed that, like the increasing N₂O emissions towards the equator (Fig. 1b), archaeal diversity significantly increased towards low latitudes (Supplementary Fig. 4a). By contrast, mid-latitude wetland soils harbored the highest bacterial diversity, whereas fungal diversity showed no significant relationships with latitude but peaked at mean annual temperature of 10–15 °C (Supplementary Fig. 4). Across all associations among archaea, bacteria, and fungi of the wetland soils, climate and soil variables had the greatest impact on microbial diversity (Supplementary Fig. 5). General linear models combined with machine learning techniques indicated that archaeal diversity was best explained by soil C/N ratio, which agrees with a previous study on mineral soils⁸. Soil pH was the primary determinant of bacterial diversity (Supplementary Fig. 5) and relative abundance of the most common bacterial phyla (Fig. 2; Supplementary Fig. 6), whereas fungal diversity showed a weak relationship with environmental factors (Supplementary Fig. 5). These results corroborate those from mineral soils, where bacteria show stronger environmental associations than fungi and warm temperate regions harbor the highest bacterial diversity⁹. In addition, soil pH constitutes the main determinant of bacterial diversity in the mineral soil microbiome⁹,10.

To determine the main microbial groups associated with N₂O emission in global wetland soils, we related N₂O fluxes with the relative abundance of various microbial lineages based on 16S and 18S rRNA gene metabarcoding. The microbial phyla Proteobacteria, Acidobacteria, and Chlororflexi are the most abundant globally (Fig. 2). However, these groups were not significantly associated with N₂O fluxes (p > 0.05), whereas the relative abundance of AOA from the phylum Thaumarchaeota emerged as the most strongly correlated group with N₂O emission (Fig. 3). This is in agreement with a previous study on arctic peat soils, where the contribution of ammonia oxidizing archaea to N₂O flux was confirmed by group-specific ammonia oxidation inhibitors as well as molecular approaches⁹. A previous study also reports a strong association between the thaumarchaeal 16S rRNA and amoA genes in environmental samples¹¹. We also found that among all prokaryotic and eukaryotes genera uncovered in metagenomics data, the Soil Crenarchaeotic Group (SCG) showed the strongest positive correlation with N₂O emissions (Supplementary Data 2). Furthermore, of the total 620 archaeal OTUs uncovered by a long-read sequencing technology (PacBio) occurring in >5 sites, 11 OTUs (including 5 in the order Nitrososphaerales, which are confirmed ammonia oxidizers; Supplementary Data 3) showed positive correlations (r > 0.35, q < 0.2) with N₂O emission. Of these, N₂O fluxes showed the strongest correlation with the relative abundance of OTUs most closely associated with ‘Candidatus Nitrosotenus chungbukensis MY2’ (r = 0.488, p < 0.001) and ‘Candidatus Nitrosococcus oleiphilus MY3’ (Spearman’s rank-correlation r = 0.477, p < 0.001). Both taxa produce N₂O in pure culture¹². In agreement with our study, a previous study found that in arctic peatlands N₂O emission was driven by only two OTUs of Thaumarchaeota, one of which was closely affiliated to ‘Ca. N. oleiphilus MY3’¹⁰. Ammonia oxidizing archaea play a key role in nitrification in unfertilized soils and soils with low ammonia concentrations¹³. In addition, in unfertilized soils, nitrite and nitrate may be predominantly made available for denitifiers through nitrification, making nitrification a limiting factor for denitrification.

Metagenomic analysis of pathways underlying global N₂O fluxes. To investigate functional pathways contributing to N₂O emission, we examined clusters of orthologous gene groups (OGs) using metagenomes (see the “Methods” section). Among all potential key genes involved in N₂O emission from archaea, the relative abundance of the archaeal amoA (ENOG411114F) showed
the strongest correlation with N$_2$O ($r = 0.625$, $p < 0.001$; Fig. 3), followed by an OG with unknown functions (Supplementary Data 4). To further evaluate the genetic basis facilitating N$_2$O emission, we compared the nucleotide sequences of the archaeal OTUs correlating with N$_2$O emission with those showing no such correlation. Using BlastN searches of 16S rRNA gene reads against complete archaeal genomes, we located the closest genome-sequenced relatives and obtained the corresponding genomic functional profiles. Based on these, we found that the aerobic ammonia oxidation pathway was restricted to four archaeal genera belonging to Thaumarchaeota—*Nitrososphaera*, *Nitrososphaeropsis*, *Nitrosotenuis*, and *Nitrosoarchaeum* (Supplementary Data 5). A strong association between the archaeal amoA gene abundance and N$_2$O emission occurred across both natural and disturbed sites. Soil nitrate (NO$_3^-$) content was also strongly correlated with the relative abundance of archaeal amoA ($r = 0.551$, $p < 0.001$). Comparative genomics analysis further revealed that archaea were more enriched in aerobic ammonia-oxidizing pathways compared with bacteria (5.3% vs 0.3%; Supplementary Data 6–8). Overall, our results support the potential key role of Thaumarchaeota in N$_2$O emissions from wetland soils globally.

While the pathways and enzymes involved in thaumarchaeal N$_2$O production are not fully understood, it has been suggested that AOA can produce N$_2$O through both nitrosating hybrid formation and enzymatic denitrification$^{12,14}$. Jung and colleagues proposed that ‘*Ca. N. oleophilus* MY3’ has a denitrification capacity using the putative cytochrome P450 NO reductase, homologs of which are present in other representatives of the genus *Nitrososphaeropsis*$.^{12}$ However, ammonia oxidizing archaea lacking these homologs are also able to produce N$_2$O$^{12,14}$. Further studies are needed to establish the mechanisms behind thaumarchaeal N$_2$O production.

**Functional genes driving global N$_2$O fluxes.** To validate the observations from the metagenomic analysis and determine specific microbial genes involved in N$_2$O dynamics, we related
N2O emissions to the absolute abundance of main genes involved in the N cycle using quantitative polymerase chain reaction (qPCR). The abundance of archaeal ammonia monooxygenase (archaeal amoA; $r = 0.458, p < 0.001$) and bacterial amoA (bacterial amoA; $r = 0.313, p < 0.001$) had strongest positive correlations with N2O emission (Fig. 4, Supplementary Fig. 7). The relative increase in archaeal nitrifiers compared to denitrifiers in lower latitudes coincided with the greater N2O emissions in these regions (Fig. 1). The absolute archaeal amoA abundance was slightly higher than the bacterial amoA abundance (qPCR: $F = 6.00, p = 0.015$), substantiating the importance of archaea in nitrification across wetland soils (Supplementary Fig. 8), as previously reported for grassland and agricultural soils\textsuperscript{15}. Our results also corroborate a local-scale metatranscriptomics study in mineral soils\textsuperscript{16}, suggesting that archaea predominate over bacteria for ammonia oxidation in soils.

Other major genes involved in the N cycle, including those known to be involved in N2O production, were surprisingly of limited importance in explaining N2O emission (Fig. 3, Supplementary Fig. 7). The correlation between comammox amoA and N2O emission was expectedly weak, which may be related to the apparent adaptation of comammox *Nitrospira* to low ammonia or because comammox *Nitrospira* produce relatively small quantities of N2O\textsuperscript{17,18}. Furthermore, the absolute abundance of comammox amoA was lower than that of archaeal amoA (Supplementary Fig. 7). In addition, the abundance of reads related to anaerobic ammonium oxidation (anammox) and the nitrite/nitrate-dependent anaerobic methane oxidation (n-damo) did not correlate with the N2O fluxes. The abundance of nosZ genes, which encode the nitrous oxide reductase enzyme that consumes N2O, was positively correlated with N2O emission (Supplementary Fig. 7). The denitrification genes responsible for N2O production (*nirK* and *nirS*) showed weak or no correlation with N2O emission (Supplementary Fig. 7). The abundance of nir genes was strongly correlated with that of nosZ genes (Fig. 4a) that reduce N2O into inert N2. This consumption may explain...
low N$_2$O emissions from the soils enriched with nir genes. Potential N$_2$ production, however, was not significantly correlated with nosZ abundance (Fig. 4a). In addition, the set of genes associated with denitrification may vary in different species, and not all denitrifiers possess all genes related to this process$^{19,20}$. Soil pH and organic carbon concentration may also affect the amount of N$_2$O produced from denitrification (Supplementary Fig. 5). Nevertheless, denitrifiers may be more metabolically versatile than nitrifiers and use a range of compounds for both energy and respiration, which is reflected in their weaker environmental associations (Supplementary Figs. 2, 5, 9). There have been many previous attempts to correlate denitrification genes with soil N$_2$O fluxes; whilst some studies have found a good correlation$^{21}$, others have not$^{22–25}$.

Next, we related N$_2$O emissions with the diversity of all major genes involved in the N cycle (based on their absolute abundances quantified by qPCR) and found higher N$_2$O emissions with increasing diversity of N cycle functional genes (Fig. 4c). The
greater N$_2$O emissions with increased functional gene diversity can be related to the functional complementarity of different N-related processes, in particular denitrification and nitrification, in producing N$_2$O. This can occur in drained wetland soils with a variety of anoxic and oxic conditions, where nitrifiers may contribute to the generation of nitrate required for denitrification. Variability of environmental conditions governed by water level dynamics has been shown to determine the diversity of microbes that affect N$_2$O fluxes in wetland soils. We found that the effect of soil factors (C/N ratio and pH) and temperature on functional gene diversity in our dataset (collectively explaining 57% of the variation; Supplementary Figs. 5, 9, 10) exceeded that of soil water content. Among the studied functional genes, the abundance of archaeal amoA correlated the best to temperature and C/N ratio (Supplementary Fig. 5), similarly to the N-cycle gene diversity. Contrary to our expectation, taxonomic diversity of microbes showed no significant correlation with N$_2$O emissions ($p > 0.05$). Previous studies have shown conflicting results on the relationship between microbial diversity and N$_2$O emissions, as reviewed in ref. 30. The decoupling of taxonomic and functional diversities may be due to functional redundancy in microbial taxa active in the N cycle.

**Environmental determinants of N$_2$O related microbial communities.** We explored the environmental conditions favoring microbial taxa and genes driving N$_2$O emission. The archaeal amoA displayed a unimodal relationship with mean annual air temperature peaking around 20 °C ($r_{adj}^2 = 0.255$, $p < 0.001$; Supplementary Fig. 9). This supports earlier findings of greater AOA activities in warmer seasons. The strong positive correlations of the AOA/AOB ratio, mean annual air temperature and soil temperature (Supplementary Fig. 11) is in line with the relatively high temperature optimum of AOA. This finding suggests that elevated (>15 °C) soil temperature in combination with optimal soil moisture may promote N$_2$O emissions from soils due to an increased AOA abundance. Nevertheless, how this may be offset by their altered balance with AOB remains to be determined.

**Implications for predicting global N$_2$O fluxes.** Our analyses indicate that both the structure and function of wetland soil microbiome and climatic conditions determine N$_2$O fluxes globally. Considering the combined effect of optimal soil moisture and temperature, archaea are important contributors to N cycling in drained wetland soils. Furthermore, we provide evidence that...
archaeal abundance is a key factor associated with ammonia oxidation pathway that underlies N₂O emission in wetlands globally. Our results complement previous findings on the major role of archaea in N₂O emission in alpine soils and oceans. In particular, the global distribution of AOA and their adaptation to low oxygen and ammonia concentrations may be suggestive of the substantial role of this microbial group in N cycling of wetland soils.

Taken together, our results suggest that nitifying microbes may contribute more strongly to N₂O emission than previously thought, and that the diversity of microbes involved in the N cycle may be the integral predictor of N₂O emissions. To determine the mechanisms underlying global N₂O emissions, we need to understand the relative role of nitrification and denitrification across a broad variety of habitat types as well as the effects of climate, vegetation, and land use. We predict that future drainage and warming of wetland soils will have negative consequences for regulating ecosystem services of wetlands through accelerating archaeal nitrification which, collectively, promote N₂O emission. Although we could not distinguish cause and effect, our study generates insights into nitrogen cycling and microbial drivers of N₂O emission in wetlands.

Methods

**Study sites and sampling.** We sampled gas and soil in 29 regions throughout the A (rainy tropical), C (temperate), and D (boreal) climate types of the Köppen classification from six continents during the vegetation period between August 2011 and June 2018, following a standard protocol. According to the protocol, the gas and soil samples were collected from locations in public domain or in previous agreement with the local community and/or property owner. The samples were exported from the origin countries and imported to Estonia, EU in cooperation with customs officers of the respective states, following the legal provisions of soil export and import, specifically exemptions for scientific purposes. To capture the full range of environmental conditions in each region, we established 76 wetland soil sites under different vegetation (mosses, sedges, grasses, herbs, trees, and bare soil) and land-use types (natural—raised bog, fen, and forest; agricultural—arable, hay field and pasture; and a peat extraction area) (Fig. 1a; Supplementary Data 1). We used a four-grade land-use intensity index to quantify the effect of land conversion: 0, no agricultural land use (natural mire, swamp, or bog forest); 1, moderate grazing or mowing (once a year or less); 2, intensive grazing or mowing (more than once a year); and 3, arable land (directly fertilized or unfertilized). The vegetation and land-use intensity types and the land-use intensity index were estimated from observations and contacts with site managers and local researchers. Within the sites, we established 1–4 stations 15–50 m apart to maximize the capture efficiency. Each of the 645 stations were equipped with 3–5 opaque PVC 65 L truncated conical chambers 1.5–5 m apart and an observation well (perforated, 50 mm diameter PP-HT pipe wrapped in geotextile; 1 m in length). From each of the 645 chambers, N₂O fluxes were measured following the static chamber method using PVC collars (0.5 m diameter, installed to 0.1 m depth in soil). Stabilization of 3–12 h was allowed before gas sampling to reduce the disturbance effect of inserting the collars on fluxes. The chambers were placed into water-filled rings on top of the collars. Gases were sampled from the chamber headspace into a 50 mL glass vial every 20 min during a 1-h session. The vials had been evacuated in the laboratory 2–6 days before the sampling. At least three sampling sessions per location were run within 3 days. Water-table height was recorded from the observation wells during the gas sampling at least 8 h after placement. Soil temperature was measured at the 10 and 20 cm depth.

Soil samples of 150–200 g were collected from the chambers at 0–10 cm depth after the final gas sampling, and transported to laboratories in Tartu, Estonia. The homogenized samples were divided into subsamples for physical–chemical analyses and DNA extraction. The samples for chemical analyses were stored at 4°C and microbiological samples were stored at −20°C. DNA extraction was provided at the Tartu University environmental microbiology laboratory (see details below). Using a PP-HT plastic cylinder, intact soil cores (diameter 6.8 cm, height 6 cm) for the N₂O analysis, and the He–O₂ method were collected from the topsoil (0–10 cm) inside 252 chambers at 26 sites, starting from 2014. Samples from different climates were run at respective temperatures. During transport, the soil samples were kept below the ambient soil temperature from which they were collected.

**Gas flux analyses.** The gas samples were analyzed for N₂O concentration within 2 weeks using two Shimadzu GC-2014 gas chromatographs equipped with ECD, TCD, and an autosampler. The calibration curves were determined from the regression equations obtained from consecutive N₂O concentrations taken when the chamber was closed, using p < 0.05 for the goodness of fit as a quality threshold for the linear regression. During the quality control, in cases of insignificant regression (p > 0.05 we removed one outlier. If the regression remained insignificant but the flux was below the gas-chromatography measuring accuracy (regression change of N₂O concentration δt < 10 ppb), we included it in the subsequent analyses as a zero value.

The helium atmosphere soil incubation technique was used to measure potential N₂ fluxes from soil cores. The cylinders with intact soil cores were placed into special gas-tight incubation vessels located in a climate chamber. Gases were removed by flushing with an artificial gas mixture (21.0% O₂, 358 ppm CO₂, 0.313 ppm N₂O, 1.67 ppm CH₄, 5.97 ppm N₂, and He). The new atmosphere equilibrium was established after 12–24 h by continuously flushing the vessel headspace with the artificial gas mixture at 20 mL/min. The flushing time depended on the soil moisture. Incubation temperature was kept similar with the field conditions. The gas-chromatograph (Shimadzu GC-2014) equipped with a thermal conductivity detector was used to measure N₂ concentration in the mixture of emitted gases accumulated in the headspace (start value, 40, 80, and 120 min as final value) of the cylinder after 2 h of closure. The gas concentration in the chambers increased in a near-linear fashion and linear regression was applied for calculation of the fluxes. The flux measurements with r² of 0.81 (p < 0.1) or greater were used.

**Soil physico-chemical analysis.** Plant-available phosphorus (P, NH₄-lactate extractable) was determined on a FusiStar5000 flow-injection analyzer. Plant-available potassium (K) was determined from the soil using a flame photometric method and plant-available magnesium (Mg) was determined from a 100 mL NH₄-acetate solution with a titanium-yellow reagent on the flow-injection analyzer. Plant-available calcium (Ca) was analyzed using the same solution by a flame-photometric method. Soil pH was determined using a 1 N KCl solution; soil moisture was determined on a 2 M KCl extract of soil by flow-injection analysis (APHA-AWWA-WEF, 2005). Total nitrogen and carbon contents of oven-dry samples were determined by a dry-combustion method on a varioMAX CNS elemental analyzer (Elementar Analysetechnik GmbH, Germany). Organic matter content of dry matter was determined by loss on ignition. We determined soil water content (SWC) from dry matter content and empirically established bulk densities of mineral and organic matter fractions.

**DNA extraction, DNA library preparation, and sequencing.** DNA extraction was performed from 0.2 g of frozen soil samples (homogenized) using the Qiagen DNeasy PowerSoil Kit (12888-100), following the manufacturer’s recommendations. DNA concentrations were measured with Qubit 1X dsDNA HS Assay Kit using Qubit 3 fluorometer (Invitrogen). Altogether 645 individual soil samples were selected for metabarcoding of bacteria, archaea, and eukaryotes. For bacteria, we used the primers 515F (5′-GTGYCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACNNGGGTWTCTAA-3′) to amplify the variable V4 region of the 16S rRNA gene. Although these primers co-amplify archaea to some extent, we subsequently separately amplified a portion of their genes to capture their full diversity, using the primers SSU1AIRF (5′-TCCGTGCGTATTCCGCG-3′) and SSU1AIRAR (5′-G GCCATGCMAMYWCTCCTC-3′). To amplify a broad range of eukaryotes, we used the primers ITS9mm (5′-GTACACCCCGCCG-3′) and ITS4nguUni (5′-CGCCTSCCTTTANTDATGTC-3′) that cover the V4–V8 region of the 18S rRNA gene and the full internal transcribed spacer (ITS) region. Both the forward and reverse primers were tagged with a 12-base multiplex identifier (MID), except in the case of archaea where only the forward primer was tagged with MID. All PCRs were performed in two replicates using 5 × HOT FIREPol Blend Master Mix (Solis BioDyne, Tartu, Estonia) in 25 μl volume. By default, the bacteria, archaea, and eukaryotes were amplified using 35, 35, and 30 cycles, respectively. In case of no amplification, two or five extra cycles were added, or DNA was re-extracted and re-purified. Thermal cycling included an initial denaturation at 95°C for 15 min; 25–40 cycles of denaturation for 30 s at 95°C, annealing for 30 s at 55°C, elongation at 72°C, and final elongation at 72°C for 10 min; and storage at −20°C. The two replicates of each reaction were pooled and visualized on TBE 1% agarose gel.

The bacterial amplicons were sequenced using the Illumina NovaSeq platform at 2 × 250 bp paired-end mode. Illumina amplicon libraries were generated using Truseq DNA PCR-Free High Throughput Library Prep Kit with Truseq DNA CD Indexes (Illumina). To increase identification accuracy and coverage, the bacterial and eukaryote amplicons were sequenced using a long-read sequencing technology on PacBio Sequel II platform. SMRTbell library preparation followed the Pacific Biosciences Amplicon library preparation protocol. Metabarcoding analysis was repeated for samples yielding <5000 prokaryotic reads (Illumina), <500 archeal reads (PacBio), or <1000 eukaryote reads (PacBio).

For the functional metagenome analysis, three replicate soil samples per station were pooled based on equimolar amount of DNA. Library preparation and indexing of each 196 pooled samples was performed using Nextera XT DNA Library Prep Kit in combination with Nextera XT Index kits v2 (Illumina). Illumina sequencing reads were demultiplexed from 5,000,000 reads using Illumina NovaSeq with 2 × 150 bp paired-end mode. The samples with <1,000,000 reads were subjected to resequencing.
Quantitative PCR. We used qPCR to quantify the absolute abundance of bacterial and archaeal 16S rRNA genes as well as the key genes involved in N cycle pathways, including nosZ, nirK, nirS, strk, nsoX, and nosZ (nifH), dissimilatory nitrate reduction to ammonia (DNRA: nrfA), ammonia oxidation (bacterial amoA, archaeal amoA, ammonium, and n-damo-specific 16S rRNA genes (Supplementary Fig. 7). The qPCR assays were performed using RotorGene® Q equipment (Qiagen, Valencia, CA, USA). The qPCR was performed following ref. 34. Briefly, the qPCR reactions were performed in 10 μl volume containing 3 μl Maxima SYBR Green Master Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA), an optimized concentration of forward and reverse primers, 1 μl of template DNA and sterile distilled water. The gene-specific primer sets, optimized primer concentrations and thermal cycling conditions for each target gene are shown in Supplementary Data 8. The quantification data were analyzed with RotorGene Series Software (version 2.0.2; Qiagen, Hilden, Germany) and LinRegPCR program (version 2020.0)42. The gene abundances were calculated as a mean of fold differences between a sample and each 10-fold standard dilution in respective standard as recommended by ref. 35; gene abundances were reported as gene copy numbers per gram of dry soil.

Bioinformatics

Metabarcoding. Illumina MiSeq sequences were analyzed using Lotus software13 following ref. 44. Briefly, the reads were demultiplexed and quality-filtered by trimming individual reads to 170 bp and removing reads with an accumulated error >2 or an estimated accumulated error >2.5 at a probability of q = 0.01. To pass to the next step, each unique read (reads preclustered at 100% identity) was required to present at least eight times in the same sample, four or more times in at least two samples, or three or more times in at least three samples. Chimeric OTUs were removed based on both reference-based and de novo chimera checking algorithms as implemented in uchime45. The resulting OTUs were taxonomically annotated by aligning their sequences with Lambda49 to SILVA v135 database45 and then the most abundant taxon as reference (COCA) option and QIIME (options: --derepMin 8.1,42 2.3 –sim_fastText 2 -redoSVL SLV -thr 0.03). For processing PaCBio sequencing data, PaCrafe46 was used as follows. Raw sequencing data was demultiplexed via mothur (version 1.36.1)49 module in PaCrafe by allowing one mismatch to tag region (i.e. to index sequence that was used for multiplexing); quality filtering was performed using vsearch (version 1.11.1)16 module with maximum expected error threshold of 1 (−fastq_maxx = 1) and discarding sequences with ambiguous bases (−fastq_maxx = 0); putative chimeric reads were discarded using vsearch uchime_denoovo algorithm; prior clustering, full length ITS reads without conservative regions (18S and 28S rRNA genes; i.e. primer-binding regions) were discarded using vsearch uchime_denovo algorithm; prior clustering, full length ITS reads were demultiplexed and quality filtered with custom Perl and PaCBio ITS (fungi), respectively.

Metagenomics. Analysis of metagenomic reads was done using MATAFILER pipeline43. Briefly, reads obtained from the shotgun metagenomic sequencing of PaCBio 16S rRNA genes (Supplementary Fig. 7) were filtered by removing reads shorter than 70% of the maximum expected read length (150 bp), with an observed accumulated error >2 or an estimated accumulated error >2.5 with a probability of q = 0.01, or >1 ambiguous position. Using mdm software (version 1.46)43, reads were trimmed if base quality dropped below 20 in a window of 15 bases at the 3′ end, or if the accumulated error exceeded 2. Altogether 196 samples produced sufficient quantity of reads and were retained for statistical analyses. To estimate the functional composition of each sample, we implemented a similarity search approach using DIAMOND (version 2.0.5; options -k 5 -e 1e-4 -sensitive -in blastx mode)44. Prior to that, the quality-filtered read pairs were merged using FLASH (version 1.2.10)46. The mapping scores of two unmerged query reads that mapped to the same target were combined to avoid double counting. In these cases, the hit scores were combined by averaging the percent identity of both hits. The best hit for a given query was on the highest bit score and highest percent identity to the subject sequence. Using this method, we calculated the relative abundance of (clusters of orthologous gene groups) (OG) by mapping of filtered reads against 70% of the eggnog database (version 4)46. We also calculated metagenomic relative abundances (i.e. mTag47) of different taxonomic groups based on small subunit (SSU) rRNA genes. For this, SortMeRNA (version 2.0.5)36 was used to extract and blast search rRNA genes against the SILVA SSU database (v128). Reads approximately matching this database with e < 10^-4 were further filtered with custom Perl and c + c scripts, and merged using FLASH. In case read pairs could not be merged, the reads were interleaved such that the second read pair was reverse complemented and then sequentially added to the first read. Of these preselected reads, 50,000 reads were fine-matched the Silva SSU database using Lambda and the lowest common ancestor (LCA) algorithm adapted from LotuS.

Genomic analysis. The taxonomic functional community of microbes in affecting ecosystem biogeochemistry. Thus, we followed a trait-based approach to confirm our findings. We downloaded 385 complete archaeal genomes from NCBI as of 2017/2020 (search terms: archaea[Organism] AND “complete genome”). These were used to build a reference database for a Blastn search to identify corresponding genomes and functional annotations of our archaeal OTUs. In addition, to better understand the potential functions of different archaeal lineages in N cycling, the functional annotation of all available archaeal genomes was retrieved from the Integrated Microbial Genomes and Microbiomes database (img.jgi.doe.gov) as of 15/7/2020.

Data analysis. To account for differences in sequencing depth across samples, diversity indices (Shannon diversity index) were calculated based on rarefied abundance matrices (metabarcoding datasets) in vegan package43 of R (version 2.5-6). Multivariate analyses were performed using Bray-Curtis dissimilarity on normalized taxa abundance matrices in vegan. All raw P-values of multiple tests were corrected using Benjamini–Hochberg method. Taxonomic abundance data were normalized using Hellinger transformation as implemented in vegan. To test the effect of biotic variables on N2O emissions, we used Spearman correlation analysis components to identify the bacterial and archaeal taxonomic lineages and fungal OTUs most strongly associated with N2O emissions. Functional gene (OG) composition and taxonomic community matrices were normalized by library size using Hellinger transformation. We subsequently used partial least squares (PLS) analysis to predict N2O emissions based on taxonomic groups, which allows the dimensionality of multivariate data to be reduced into PLS components using plsdof package40 of R (version 0.11.17). Prior to this, we performed a backward variable elimination procedure to remove variables with low explanatory power (VIP threshold <1), as implemented in plsVarSel package41 of R (version 0.9.6).

For univariate analysis, the best predictors of the diversity and relative abundances of taxonomic and functional groups were identified using a machine learning approach implemented in randomForest package42 of R (version 4.6.14). To further test direct and indirect effects of variables in the best model, structural equation modeling (SEM) was used as implemented in piecewiseSEM package43 of R (version 2.1.0). The prior model was constructed based on our hypothesis (see the section “Introduction”). This model was compared to a second order polynomial model for certain analyses. The best polynomial fit was determined on the basis of Akaike Information Criterion (AIC) scores using “AIC” and “poly” functions of R.

Data availability

All metabarcoding sequences and associated metadata have been deposited in the European Bioinformatics Institute Sequenced Read Archive database. https://www.ncbi.nlm.nih.gov/bioproject/PRJA718418; metagenomics sequences and associated metadata have been deposited at The European Nucleotide Archive under accession number https://www.ebi.ac.uk/ena/browser/view/PRJEB44414. Additional data generated in this study are provided in the Supplementary Information/Source Data file. SILVA database is available at https://www.arb-silva.de; UNITE database is available at https://unite.ut.ee/repository.php; Integrated Microbial Genomes is available at https://img.jgi.doe.gov; eggnog database is available at http://eggnog5.embl.de/download/eggnog_4.0/ Source data are provided with this paper.

Code availability

The pipeline to process metabarcoding samples is available under https://psbweb05.psb.ugent.be/lotus/downloads.html and https://doi.org/10.15156/bsc/587450. The pipeline to process shotgun metagenomic samples is available under https://github.com/hildebra/MATAFILER (https://doi.org/10.25211/zenodo.5831723).

Received: 7 August 2021; Accepted: 23 February 2022; Published online: 17 March 2022

References

1. Ravishankara, A. R., Daniel, J. S. & Portmann, R. W. Nitrous oxide (N2O): the dominant ozone-depleting substance emitted in the 21st century. Science 326, 123–125 (2009).
2. Tian, H. et al. A comprehensive quantification of global nitrous oxide sources and sinks. Nature 586, 248–256 (2020).
3. Hu, H.-W., Chen, D. & He, J.-Z. Microbial regulation of terrestrial nitrous oxide emissions. Nature 749 (2019).
4. Tian, H. et al. A comprehensive quantification of global nitrous oxide sources and sinks. Nature 586, 248–256 (2020).
5. Hu, H.-W., Chen, D. & He, J.-Z. Microbial regulation of terrestrial nitrous oxide emissions: understanding the biological pathways for prediction of emission rates. FEMS Microbiol. Rev. 39, 729–749 (2015).

DATA AVAILABILITY

https://doi.org/10.1038/s41467-022-29161-3

ARTICLE

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-022-29161-3

www.nature.com/naturecommunications
29. Shan, J. et al. Beyond denitrification: the role of microbial diversity in controlling nitrous oxide reduction and soil nitrous oxide emissions. *Glob. Change Biol.*, 27, 2669–2683 (2021).

30. Butterbach-Bahl, K., Baggs, E. M., Dannenmann, M., Kiese, R. & Zeichmeister-Boltenstern, S. Nitrous oxide emissions from soils: how well do we understand the processes and their controls? *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 368, 20130122 (2013).

31. Allison, S. D. & Martiny, J. B. Resistance, resilience, and redundancy in microbial communities. *Proc. Natl Acad. Sci. USA*, 105, 11512–11519 (2008).

32. Taylor, A. E., Zeglin, L. H., Wanzek, T. A., Myrold, D. D. & Bottomley, P. J. Dynamics of ammonium-oxidizing archaea and bacteria populations and contributions to soil nitration potentials. *ISME J.*, 6, 2024–2032 (2012).

33. Wu, Y. et al. Autotrophic growth of bacterial and archaeal ammonia oxidizers in freshwater sediment microcosms incubated at different temperatures. *Appl. Environ. Microbiol.*, 79, 3076–3084 (2013).

34. Espenberg, M. et al. Differences in microbial community structure and nitrogen cycling in natural and drained tropical peatland soils. *Sci. Rep.*, 8, 4772 (2018).

35. Peng, Y. et al. Soil Temperature dynamics modulate N2O flux response to multiple nitrogen amendments from Arctic peatlands after permafrost thaw. *Proc. Natl Acad. Sci. USA*, 114, 6238–6243 (2017).

36. Karimi, B. et al. Biogeography of soil bacteria and archaea across France. *Sci. Adv.*, 4, eaat1808 (2018).

37. Bahram, M. et al. Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237 (2018).

38. Dehalg-Baquero, M. et al. A global atlas of the dominant bacteria found in soil. *Science*, 359, 320–325 (2018).

39. Siljander, H. M. P. et al. Archaeal nitrification is a key driver of high nitrous oxide emissions from arctic peatlands. *Soil Biol. Biochem.*, 137, 107539 (2020).

40. Ono, E. V., Quince, C., Nicol, G. W., Prosser, J. I. & Gubry-Rangin, C. Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *ISME J.*, 10, 85–96 (2016).

41. Jung, M.-Y. et al. Indications for enzymatic denitrification to N2O at low pH in an ammonia-oxidizing archaeon. *ISME J.*, 13, 2633–2638 (2019).

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

43. Steiglmeier, M. et al. Aerobic nitrous oxide production through N-nitrating hybrid formation in ammonia-oxidizing archaea. *ISME J.*, 8, 1135–1146 (2014).

44. Leininger, S. et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, 442, 806–809 (2006).

45. Griffiths-Brown, S. Climate change induces shifts in abundance and activity pattern of bacteria and archaea catalyzing major transformation steps in nitrogen turnover in a soil from a mid-European Beech Forest. *PloS ONE*, 9, e114278 (2014).

46. Kits, K. D. et al. Low yield and abiotic origin of N2O formed by the complete oxido-reductases for nitrous oxide emissions. *Nucleic Acids Res.*, 37, e45 (2009).

47. Hildebrand, F., Tadeo, R., Voigt, A. Y., Boek, P. & Raes, J. LotuS: an efficient and user-friendly OTU processing pipeline. *Microbiome*, 2, 30 (2014).

48. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200 (2011).

49. Hauswold, H., Singer, J. & Reinert, K. Lambda: the local aligner for massive high-throughput amplicon sequencing data. *Mol. Ecol. Resour.*, 17, e234–e240 (2017).

50. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 74, 1217–1220 (2008).

51. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4, e2584 (2016).

52. Bengtsson-Palme, J. et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for metabarcoding using full-length ITS sequences. *ISME J.*, 41, D590–D596 (2017).

53. Ansland, S., Bahram, M., Hiesalu, I. & Tedersoo, L. PipeCraft: flexible open-source toolkit for bioinformatics analysis of custom high-throughput amplicon sequencing data. *Mol. Ecol. Resour.*, 17, e234–e240 (2017).

54. Schloss, P. D. et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 74, 1217–1220 (2008).

55. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, 4, e2584 (2016).

56. Bengtsson-Palme, J. et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods Ecol. Evol.*, 4, 914–919 (2013).

57. Kölijal, U. et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.*, 22, 5271–5277 (2013).

58. Hildebrand, F. et al. Dispersal strategies shape persistence and evolution of human gut bacteria. *Cell Host Microbe*, 29, 1167–1176 (2021).

59. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 12, 59–60 (2014).

60. Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27, 2957–2963 (2011).

61. Huerta-Cepas, J. et al. eggNOG 4.5: a hierarchical orthology framework with higher resolution, more species, and improved annotation. *Database*, 2017, B16 (2017).

62. Logares, R. et al. Metagenomic 16S rDNA I Illumina tags are a powerful alternative to amplicon sequencing to explore diversity and structure of microbial communities. *Environ. Microbiol.*, 16, 2659–2671 (2014).

63. Vopelova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, 28, 3211–3217 (2012).

64. Oksanen, J. et al. The vegan package. *Community Ecol. Package*, 10, 631–637 (2009).

65. Sanchez, G. ploDpetot: Partial Least Squares (PLS) Data Analysis Methods. https://CRAN.R-project.org/package=ploDpetot (2012).

66. Mather, T. T., Laland, K. H., Snipen, L. & Sabo, S. A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.*, 118, 62–69 (2012).
Acknowledgements
We thank O. Botsarova, L. Lopp, K. Kanger, R. Puusepp, E.J. Sova, H. Tamm, I. Varik for assistance in molecular laboratory analyses. We thank Lorenzo Menichetti for useful discussions. We are grateful to S. Egurov, I. Filippov, G. Gaberi, I. Gallagher, I. Gheorghe, W. Hartman, R. Iturraspe, J. Jarveoja, A. Kull, F. Lagoun-Delarge, E. Lapshina, A. Lohula, C. Luwata, S. Mander, M. Metspalu, W. Mitsch, R. Moreton, K. Oopkaup, H. Oskarsson, J. Paal, T. Pae, E. Parrodi, S. Pellerin, F. Sabater, J. Salm, F. Sgouridis, D. Silveira Batista, K. Sohar, K. Storey, M. Tenywa; S. Ullah, E. Uuemaa, G. Veber, J. Villa, L. Yang and S.S. Zaw for assistance on study-site selection and field investigation. ME, JP, LT and UM were supported by the Ministry of Education and Science of Estonia (SF0180127s08 grant), the Estonian Research Council (JUT2-16, PUTJD619, PRG-352, PRG-609, and MOBERC20), EU through the European Regional Development Fund (Centres of Excellence ENVIRON, grant number TK-107, EcoChange, grant number TK-131, and the MOBTP101 returning researcher grant by the Mobilitas Plus program); M.B. was supported by the Swedish Research Council Formas (2020-00807); J.P., K.K., and M.Ma. were supported by the European Social Fund (Doctoral School of Earth Sciences and Ecology); L.L.-M. was supported by Royal Society Dorothy Hodgkin Research Fellowship (DH150187); F.H. was supported by European Research Council (ERC) Starting Grant (UNITY 852993), the BBSRC Institute Strategic Program Gut Microbes and Health BB/S012490/1, its constituent project BB/S012490/1, the European Union’s Horizon 2020 research and innovation program (grant agreement no. 948219).

Author contributions
U.M., J.P., and M.E. set up and conducted the field experiments. L.T. supervised DNA extraction and sequencing analysis. F.H. processed metagenomics data. S.A. processed metabarcoding data; M.E. performed qPCR analysis. M.B., M.E. and J.P. analyzed the data. L.L.-M. revised taxonomic and functional annotations. M.B. wrote the first draft of the manuscript with input from M.E., J.P., L.T. and U.M. The manuscript was revised by M.B., M.E., J.P., L.L.-M., K.K., U.K., J.L., M.Ma., M.Me., U.N., M.O., M.P., K.S., M.Z., F.H., L.T., and U.M. All authors approved the submitted version.

Funding
Open access funding provided by Swedish University of Agricultural Sciences.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-29161-3.

Correspondence
and requests for materials should be addressed to Mohammad Bahram.

Peer review information Nature Communications thanks Linta Reji, Sarah Harpenslager and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022