Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Diversity of archaea and niche preferences among putative ammonia-oxidizing Nitrososphaeria dominating across European arable soils

Aurélien Saghaï
Swedish University of Agricultural Sciences: Sveriges lantbruksuniversitet
Samiran Banerjee
Agroscope
Florine Degrune
Freie Universitat Berlin
Anna Edlinger
Agroscope
Pablo García-Palacios
Instituto de Ciencias Agrarias
Gina Garland
Agroscope
Marcel GA van der Heijden
Agroscope
Chantal Herzog
Agroscope
Fernando T Maestre
Universidad de Alicante: Universitat d'Alacant
David S Pescador
Universidad Rey Juan Carlos
Laurent Philippot
INRA UMR Agroécologie: Agroecologie
Matthias Rillig
Freie Universitat Berlin
Sana Romdhane
INRA UMR Agroécologie: Agroecologie
Sara Hallin (sara.hallin@slu.se)
Sveriges lantbruksuniversitet https://orcid.org/0000-0002-9069-9024

Research

Keywords: archaea, phylogeny, niche differentiation, community assembly, Thermoplasmata

Posted Date: May 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-429240/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at Environmental Microbiology on November 18th, 2021. See the published version at https://doi.org/10.1111/1462-2920.15830.

Abstract

Background: Archaeal communities in arable soils are dominated by Nitrososphaeria, a class within Thaumarchaeota comprising all known ammonia-oxidizing archaea (AOA). AOA are key players in the nitrogen cycle and defining their niche specialization can help predicting effects of environmental change on these communities. However, hierarchical effects of environmental filters on AOA and the delineation of niche preferences of nitrososphaerial lineages remain poorly understood. Here we combined multiple environmental gradients with fine-scale phylogenetic analyses and machine learning for new insights into ecological preferences within Nitrososphaeria. With this approach, we could identify climatic, edaphic and geomorphological drivers of Nitrososphaeria and other archaea in arable soils along a 3 000 km European gradient.

Results: Mean annual temperature, C:N ratio and pH were the best predictors of diversity, evenness and distribution of Nitrososphaeria and thresholds in the predictions could be defined for C:N ratio and cation exchange capacity. Recent adaptations to soil pH were indicated in the Nitrososphaeria phylogeny. Our analyses further suggest the coexistence of widespread ecophysiological differences between closely related soil Nitrososphaeria. Only limited insights into the ecology of the low-abundant, but highly diverse classes Thermoplasmata and Woesearchaeia could be inferred. However, for explaining the overall archaeal community composition at the continental scale the contribution of distance (reflecting stochastic assembly processes) and environmental factors (reflecting deterministic processes) were comparable.

Conclusions: Our findings underline the multifactorial nature of niche differentiation in soil Nitrososphaeria and the ecophysiological differences between closely related members observed suggest that the phylogeny does not reflect their ecology. Our results also imply multiple, independent specializations to low pH. The study highlights that the ecology of Nitrososphaeria is best studied at fine phylogenetic scale. Finally, the identification of thresholds in the responses show that simple correlations are not sufficient when determining environmental drivers of archaeal diversity.

Background

Archaea are pivotal for the functioning of all major biomes, as they play a critical role in both carbon (C) and nitrogen (N) cycles [1, 2]. In terrestrial ecosystems, archaeal communities tend to be phylogenetically clustered and are commonly dominated by Thaumarchaeota [3–5]. This phylum harbors the globally important ammonia-oxidizing archaea (AOA), restricted to the class Nitrososphaeria [6]. All AOA characterized so far use the ammonia monooxygenase (encoded by *amoABC* genes) to catalyze the first step of nitrification, the oxidation of ammonia to nitrite [7]. Globally, nitrification contributes to the circulation of N [8], but locally this process causes N losses, directly through nitrate leaching and production of the greenhouse gas nitrous oxide and indirectly by fueling denitrification leading to gaseous N losses as dinitrogen gas or nitrous oxide. Altogether, this corresponds to an average loss of 50 % of the N added to arable soils [9]. Thus, nitrification affects N use efficiency in cropping systems,

causes eutrophication of watersheds and contributes to global warming. In arable soils, AOA are typically abundant and important contributors to nitrification [10–13] and, therefore, there is great interest in understanding their ecology and evolution (see [14] for a recent review).

Soil pH has previously been proposed as the main driver of thaumarchaeotal and AOA diversity [5, 15, 16] and evolution at relatively broad phylogenetic scales [17–19]. However, this view has recently been challenged by Alves et al., who suggested more recent adaptations to low pH from cosmopolitan clades [6]. In addition to soil pH, experimental work using isolates has linked niche differenciation in AOA to differences in ammonia affinity/tolerance and organic C preferences [20–22]. Yet, the relative importance of other environmental factors, including C:N ratio [4], moisture [23, 24], temperature [25, 26] and soil organic carbon content [27], for niche differentiation across AOA lineages remains poorly understood since there are few reports on hierarchical effects of environmental filters in the delineation of ecological preferences in Thaumarchaeota or, more specifically, AOA [28]. Defining niche specialization is important for understanding and predicting effects of environmental change on AOA diversity and composition, with implications for soil functioning.

The aim of this study was to identify ecological niches among Nitrososphaeria (putative AOA) at fine phylogenetic scale by taking advantage of the congruence between 16S rRNA and *amoA* phylogenies [6, 27]. We further aimed at identifying the best environmental (climatic, edaphic and geomorphological) predictors of the overall archaeal diversity and determining the factors governing their community assembly processes in soil. To this end, we sampled arable soils along a 3 000 km European gradient, spanning from northern Sweden to southern Spain [29]. Such continental surveys focusing on archaea are rare [30] and represent a significant opportunity to gain insight into their ecology by capturing broad environmental gradients. By only including arable fields under cereal cultivation and conventional tillage [29], management effects were minimized. For the AOA, machine learning in terms of random forest modeling [31] was used to determine the environmental drivers of Nitrososphaeria diversity and AOA abundances. The extent of niche differentiation in Nitrososphaeria was assessed by examining how multifactorial changes in environmental conditions were reflected in the phylogeny, using multivariate regression trees [32], as opposed to previous studies that have assessed the importance of each variable individually (e.g. [6, 18, 27]). We hypothesized that the combination of multiple environmental gradients with fine-scale phylogenetic analyses would reveal if ecological preferences were conserved within lineages of Nitrososphaeria. Random forest modeling was also used to determine the environmental drivers of α-diversity of other archaeal classes to gain insight into the ecology of archaea in arable soils. By considering the spatial distance (*i.e.* dispersal limitation) and edaphic factors (*i.e.* environmental filtering), we could also evaluate the relative importance of stochastic versus deterministic processes in the assembly of the overall archaeal communities since this is crucial to predict how current and future environmental change will affect the structure of these communities and ultimately the ecosystem functions they support [33].

Material And Methods

Sampling and measurement of edaphic parameters

Soil samples were collected across a north-south gradient in Europe (Sweden, Germany, Switzerland, France and Spain) in a total of 151 agricultural fields (Additional file 1: **Figure S1**) [29]. To homogenize variation in plant development stages and associated farming practices, the sampling was performed around flowering time (*i.e.* anthesis) between May and August 2017 depending on country and location. Only fields under cereal cultivation (barley, oat or wheat) and conventional tillage were surveyed. In each field, eight soil cores (Ø5 x 20 cm) were taken within a 10 m radius. Five cores were sieved (2 mm) and homogenized into a composite sample, which was air-dried before measuring soil parameters (Table 1) using the Swiss standard protocols [34]. A fresh subsample of the composite soil was taken before drying for DNA extraction and stored at -20°C until DNA extraction. The three remaining cores were kept intact and used to measure bulk density of the fine soil (< 8 mm). Mean annual atmospheric temperature (MAT) data (1987–2017) were obtained for each sampling site using their GPS coordinates and the closest weather station in the NOAA database (https://www.noaa.gov/), through the R package 'rnoaa' (v. 0.8.4; [35]). The same GPS coordinates were also used to gather elevation data using the R package 'elevatr' (v. 0.2.0; [36]).

Table 1
Environmental variables used in this study.

Category	Variable	Range
Climatic	Mean annual temperature (MAT; ºC)	3.7–19.5
Geomorphological	Elevation (m)	7.8-1 022.0
Edaphic	Bulk density (g cm^{-3})	0.7–1.7
	C:N ratio	7.8–63.9
	C:P ratio	7.3-1 042.9
	Calcium (cmol kg^{-1})	0.2–36.1
	Cation Exchange Capacity (CEC; cmol kg^{-1})	5.6–49.6
	Clay (%)	8.6–54.9
	Magnesium (cmol kg^{-1})	0.3–5.6
	Moisture (%)	0.03–0.35
	N:P ratio	0.1–19.1
	pH	5.4–8.3
	Silt (%)	12.4–63.1
	SOC (g kg^{-1})	3.2–36.1
	Total C (g kg^{-1})	6.6–90.2
	Total N (g kg^{-1})	0.5–4.8
	Total P (g kg^{-1})	0.1–1.7

DNA extraction, amplification and sequencing

DNA was extracted on the homogenized soil samples using the DNeasy PowerSoil-htp 96 well DNA isolation kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. Archaeal 16S rRNA gene fragments, encompassing the V3-V4 hypervariable regions, were amplified using the primer pair S-D-Arch-0349-a-S-17 [37] and S-D-Bact-0785-a-A-21 [38] to capture both Thaumarchaeota and low-abundant and under-studied groups typically present in soils (e.g. different classes of methanogens, Thermoplasmata; [30, 39, 40]). The PCRs were run in duplicate 15 µl reactions under the following conditions: 3 minutes at 98°C, followed by 30 cycles of 98°C for 30 s, 65°C for 30 s and 72°C for 30 sec and a final extension step of 10 minutes at 72°C. The PCR products were then pooled and inspected by gel electrophoresis. For the second (indexing) PCR, a single 30 µl reaction was performed using 0.2 µM of primers with Nextera adaptor and index sequences, and 3 µl of the pooled PCR product from the first PCR

as template. Conditions were the same as in the first PCR, except an annealing temperature of 55°C, an extension time of 45 sec, and 8 cycles. The final PCR products were purified using AMPure XP PCR purification beads (Beckman Coulter, Indianapolis, IN, USA) following the manufacturer's protocol. The amplicons were checked by gel electrophoresis and using a 2100 BioAnalyzer (Agilent, Santa Clara, CA, USA). After quantification using a Qubit™ fluorometer (Thermo Fischer Scientific, Waltham, MA, USA), a single library was created by pooling equal amounts of purified amplicons from all the 151 samples. Sequencing was performed by SciLifeLab (Uppsala, Sweden) on the Illumina MiSeq (2 x 250 bp) platform.

Quantitative PCR analysis

The abundance of ammonia-oxidizing archaea was determined by quantitative real-time PCR (qPCR) based on SYBR green detection and the archaeal *amoA* gene (encoding the ammonia monooxygenase subunit A). The qPCR reactions were carried out in duplicate runs on a ViiA7 (Life Technologies, Carlsbad, CA, USA) and a 15 µl reaction volume containing 7.5 µL of Takyon Master Mix (Eurogentec, Liège, Belgium), 1 µM of each primer (CrenamoA23f and CrenamoA616r; [26]), 250 ng of T4 gene 32 (QBiogene, Carlsbad, CA, USA) and 1 ng of DNA). Cycling conditions were 15 min at 95°C, 35 cycles of 15 s at 95°C, 30 s at 55°C, 30 s at 72°C and a plate read of 15 s at 80°C (efficiency: 88 %). Standard curves were obtained by serial dilutions of linearized plasmids with cloned fragments of the specific gene. The amplifications were validated by melting curve analyses. Potential inhibition of PCR reactions was checked by amplifying a known amount of the pGEM-T plasmid (Promega, Madison, WI, USA) with the plasmid specific T7 and SP6 primers when added to the DNA extracts or non-template controls. No inhibition was detected with the amount of DNA used.

Sequence processing and phylogenetic reconstruction

All sequence analyses were performed using the R software (v. 3.6.4, R Core Team, 2019). The archaeal 16S rRNA gene amplicons were processed with the 'dada2' package (v. 1.6.0; [42]) to infer amplicon sequence variants (ASVs), which allowed detection of ecological preferences at the finest phylogenetic scale [43, 44]. Briefly, primer sequences were removed and the reads merged using default parameters. Chimeras were discarded using a *denovo* approach with the removeBimeraDenovo function ('consensus' method). The resulting ASVs were aligned to the SILVA reference database (SSU132 Ref NR) using the SINA aligner (v. 1.6.0; [45]) and classified using SINA's least common ancestor algorithm. After elimination of the bacterial ASVs, 2 042 archaeal ASVs remained based on a total of 6 316 715 high-quality 16S rRNA gene amplicons (corresponding to ca. 70 % of the original dataset). The reads were also clustered into OTUs at a similarity cut-off of 97 % (Additional file 1: **Supplementary Methods**) to compare the diversity of nitrosophaerial taxa based on ASVs and OTUs.

Rarefaction curves of species richness were generated from the raw ASV table using the rarcurve function in 'vegan' (v. 2.5.5; [46]) (Additional file 1: **Figure S2**) and a rarefied table (n = 1 079 ASVs) was obtained by averaging the ASV counts over 1 000 computations using the rrarefy function in 'vegan'. A phylogeny was built with the rarefied ASVs and a broad taxonomic selection of reference sequences extracted from

SILVA. Sequences were aligned using the SINA aligner and the phylogeny generated using FastTreeMP (v. 2.1.10; [47]) with the GTR + CAT model of nucleotide evolution.

Partitioning and transformation of sequence data

The ASV abundance distributions were examined to further partition the rarefied table between frequent and rare community members. An index of dispersion corresponding to the ratio of the variance to the mean abundance multiplied by the occurrence was calculated [48] to split the dataset according to the frequency of occurrence of each ASV [49]. This index was then used to model whether ASVs followed a stochastic (Poisson) distribution and those falling below the 2.5 % confidence limit of the χ^2 distribution were discarded [50] (Additional file 1: **Figure S3**). By focusing on the frequent community members, we minimized the risk of sampling artefacts that would bias the distribution of the ASVs and thus increase the likelihood to detect relevant ecological patterns. Such partitioning remains rarely used in microbial ecology (but see [51–53]), despite being more statistically robust than traditional approaches using arbitrary cut-offs of local and regional abundances (e.g. [54]). The resulting community included 718 ASVs representing 99.6 % of the reads in the rarefied dataset. Zero count ASVs were replaced by an imputed value using the Bayesian-multiplicative replacement method available in the 'zCompositions' package (v. 1.2.0; [55]). An isometric log-ratio transformation [56], as implemented in the 'philr' package (default parameters, v. 1.10.0; [57]), was then applied to the zero replaced dataset using the phylogenetic tree of the ASVs as the sequential binary partition. The output of this transformation consists of a matrix of sites x nodes containing the balances calculated on the internal nodes of the phylogeny. This approach accounts for the compositional nature of amplicon datasets [58] by inferring changes between phylogeny-based subcommunities (i.e. the two set of branches stemming from any given node) rather than changes of individual taxa (i.e. branch tips) [57, 59].

Statistical analyses

Alpha-diversity indices and taxonomic composition were calculated using the full rarefied table. Faith's phylogenetic diversity (PD; [60]) was calculated at both domain and class levels using the phylogenetic tree of the ASVs and the pd function in the 'picante' package (v. 1.8; [61]), and Pielou's evenness [62] with the diversity function in 'vegan'. Random forest (RF) based variable selection was performed on the entire set of variables (Table 1) using the 'VSURF' package (v. 1.1.0; [63]) to identify the best predictors for PD and evenness of archaeal groups and the absolute abundance of AOA (response variables). Random forests are well suited to model non-linear relationships between predictors and response variables and can deal with non-normality and high collinearity among predictors [31]. Briefly, variables were first ranked according to a variable importance score, averaged across 50 RFs. The set of variables leading to the model with the smallest out-of-bag error, averaged across a nested collection of 25 RFs starting from one with only the most important variable, was selected. To account for the random nature of RFs, the algorithm was run with default parameters 100 times and only the variables selected in the second step in > 95 % of the runs were retained. Random forest analyses, as implemented in the 'randomForest' package (v. 4.6–14; [64]), were then used to study the relationship between the selected environmental variables and the response variables. A grid search was first conducted to find the optimal combination

of tuning parameters (with ntree = 500): the number of variables to randomly sample as candidates at each split (mtry; range 1–10, step = 1), the minimal number of samples within the terminal nodes (node_size; range 2–10, step = 1) and the fraction of samples to train the model on (samp_size; 55 %, 63.25 % (default), 70 % and 80 %). The search was run 100 times and the combination of parameters corresponding to the best model fit (or lowest out-of-bag root-mean-square error) was selected (Additional file 1: **Table S1**). The relationship between each environmental variable and PD, evenness and the absolute abundance of AOA was visualized using accumulated local effects (ALE) plots (grid.size = 30) implemented in the 'iml' package (v. 0.9.0; [65]). These plots show how the prediction of a response variable in a given RF model (PD, evenness or abundance of AOA) changes on average over the range of each individual environmental variable, while accounting for potential correlations amongst explanatory variables [66].

All statistical analyses on β-diversity were conducted on the philr-transformed [57] ASV table. Differences in community composition and structure were visualized with a principal component analysis (PCA) using the rda function in 'vegan'. Distance-decay curves were calculated as the linear regression relationship between geographical distance and Euclidean distance-based community similarity. The relative influence of climatic, edaphic and spatial factors on the patterns of β-diversity was estimated by variation partitioning analysis (VPA; varpart function in 'vegan'). To this end, cartesian coordinates of each sampling site were obtained from the GPS data (geoXY function, package 'SoDA' v. 1.0.6) and used to construct a matrix of distance-based Moran's eigenvectors maps (dbMEM). The edaphic factors were selected following a procedure similar to Power et al. [67]. First, individual permutational multivariate analyses of variance (PERMANOVA) and Mantel tests were conducted between overall β-diversity and each variable using the adonis and mantel functions implemented in 'vegan' (number of permutations = 9 999), respectively (Additional file 1: **Table S2**). Collinearity among edaphic factors was assessed by pair-wise Spearman correlations; only the variable with the highest mantel statistic was retained in each collinear group ($|r| > 0.7$). Thereby, the selected edaphic factors were calcium, clay, C:N and C:P ratios, magnesium, moisture, pH, silt, SOC and total C. Finally, a forward selection step ($p < 0.05$) was applied to select the final set of variables before running the variation partitioning analysis (forward.sel function, package 'adespatial' v. 0.3.4). The significance of each component of the VPA was estimated by a permutation test (adonis function). Elevation, although significant, only improved the explanatory power by 0.4 % and was not included in the final analysis.

Since Nitrososphaeria represent putative AOA [6], their ecological preferences were predicted at a finer taxonomic scale by recursive partitioning of the corresponding philr-transformed ASV table with the entire set of metadata (Table 1). First, a subset of the full phlir-transformed ASV table was generated by extracting the nodes corresponding to Nitrososphaeria. Then, multivariate regression trees (MRT) were computed using the 'mvpart' package (v. 1.6.2; [32]). to predict the relationships between the set of variables (Table 1) and the community composition of Nitrososphaeria. This approach has the advantage to allow for the examination of the sequential effect of several environmental filters (contrary to [6, 17, 18, 27] who assessed effects of each variable individually) and delineate clusters of samples in which the variation in environmental conditions is minimized (Additional file 1: **Table S3**). The selected

tree represented the most parsimonious solution within one standard error above the minimal cross-validated relative error (Additional file 1: **Figure S4**; n = 10 000 trees), following [68], and explained 42 % of the variation. Indicator nodes were searched for in each partition (i.e. clusters of samples) and corresponded to balances that significantly differed from the mean across all partitions according to Tukey's HSD ($P < 0.01$). The balances within each MRT group were plotted on the nitrososphaerial phylogeny with ggtree (v. 1.16.0; [69]), with blue and red branches indicating an increase and a decrease, respectively. Sequences of Group 1.1c thaumarchaeota were used to root the phylogeny.

Results

Archaeal taxa, diversity and community structure

Thaumarchaeota dominated the archaeal communities in arable soils across Europe, both in terms of the relative abundance of ASVs and reads, followed by the phylum Euryarchaeota and Woesearchaeota within the DPANN superphylum [70] (Fig. 1). The 19 ASVs with a frequency > 1 % represented ca. 70 % of the reads and, consequently, the archaeal communities were uneven (J = 0.61 ± 0.08) and displayed low phylogenetic diversity (PD = 5.09 ± 1.51). At the class level, Nitrososphaeria alone represented more than 90 % of the reads. We obtained 322 and 25 nitrososphaerial ASVs and OTUs at a similarity cut-off of 97 %, respectively. Most of the diversity was found within the order Nitrososphaerales (322 ASVs or ~ 91 % of AO ASVs and ~ 98 % of the reads), whereas Nitrosopumilales and Nitrosotaleales represented a minor fraction of the nitrososphaerial communities (~ 1 and 8 % of ASVs; < 1 % and < 2 % of the reads, respectively). Thermoplasmata and Woesearchaeia were diverse, respectively representing ca. 27 and 25 % of the ASVs, but also relatively rare (< 3 % of the reads).

The combination of PCA and PERMANOVA showed that the overall archaeal communities were structured following a spatial gradient of pH and temperature along PC1, with a shift from alkaline pH and warm temperatures to acidic soils and colder climate (Fig. 2a). This is reflected by the increasing dissimilarity between communities with increasing spatial distance between sampling sites (Fig. 2b). Besides pH and MAT (R^2 = 0.11 and 0.10, respectively), soil C:N ratio (R^2 = 0.07) and soil moisture (R^2 = 0.06) were also important contributors to β-diversity ($p < 0.001$; Additional file 1: **Table S2**). Climatic, edaphic and spatial factors collectively explained 36 % of the variation in the archaeal community composition and all three groups of factors were significant (ANOVA, $p < 0.01$; Fig. 2c). When partitioning the variance, spatial distance (*i.e.* dispersal limitation) explained nearly as much variation in community composition as the edaphic factors (*i.e.* environmental filtering; ca. 10 %), whereas the climatic component defined solely by MAT explained only 1 %.

Environmental predictors of diversity of the individual archaeal classes and abundance of AOA

Random forest-based variable selection analyses revealed that different environmental variables (Table 1) contributed to the PD and evenness of the taxa-specific archaeal communities, although MAT

was important for diversity of nearly all taxonomic groups (Figs. 3a and **b**, Additional file 1: **Figures S5** and **S6**). Across the archaeal domain, edaphic factors were more important for PD than for evenness. For the methanogens, PD and evenness of Methanobacteria and Methanomicrobia displayed the same relationship to calcium, elevation, soil organic carbon (SOC) and, to a lesser degree, MAT. All three categories of environmental factors significantly influenced the PD of Woesearchaeia, while soil texture and MAT were major predictors of the diversity and evenness of Thermoplasmata.

For the Nitrososphaeria, elevation, cation exchange capacity (CEC), C:N ratio and moisture had strong effects on the diversity, whereas evenness was driven by multiple climatic, edaphic and geomorphological (*i.e.* elevation) variables. Here, diversity and pH exhibited a u-shaped relationship, with an increase associated with pH below 7 and above 7.5. More acidic soils were associated with an increase in both PD and evenness in Group 1.1c, i.e. a defined lineage within the Thaumarchaeota that likely do not oxidize ammonia [71]. The abundance of AOA, measured as the copy number of the archaeal *amoA* gene (Additional file 1: **Figure S7a**), was positively influenced by elevation, total N and total P, while AOA abundances tended to decrease with increasing silt content and bulk density (Fig. 3c and Additional file 1: **Figure S7b**).

Ecological preferences of nitrososphaerial taxa

Each of the eight clusters obtained using the MRT analysis (labelled A to H) contained 10 to 31 samples (Fig. 4), with the exception of the four samples from northern Sweden that formed a separate cluster (cluster A in Fig. 4). Eight variables, among edaphic (bulk density, calcium, CEC, C:N ratio, pH, SOC and total C) and climatic (MAT) factors, were selected in the regression analyses. The MRT identified two to four levels of environmental filtering, with MAT being the most important driver. The importance of geographic distance was also evidenced by the origin of the samples present in each cluster, with a clear North-South gradient. Several variables contributed equally to the split between clusters A and B (CEC, C:N ratio, MAT, SOC and total C) and clusters G and H (calcium and SOC), indicating that soil C content also plays a role in defining ecological preferences. It should be noted that this eight-cluster partition does not imply within-cluster homogeneity for the rest of the measured variables (Additional file 1: **Table S3**)

The balances, which depict relative changes in abundance between two neighboring clades relative to each other, revealed extensive niche differentiation throughout the phylogeny (Fig. 4). Within the less abundant orders Nitrosopumilales and Nitrosotaleales, balances at both deep (depicting ancient evolutionary events conserved in the phylogeny) and shallow (depicting more recent adaptations) nodes were only significant in the northernmost samples (Fig. 4 clusters A and B). Within the dominating order Nitrososphaerales, the depth of the phylogenetic signals differed depending on the combinations of environmental factors. All clusters displayed significant shallow nodes, whereas clusters A, B, C, F and H also exhibited significant deeper nodes. Of particular interest was the presence of multiple shallow nodes associated with more acidic soils in clusters A (pH range: 5.4–5.9), B (5.8–6.9) and E (6.1-7.0) (Fig. 4; Additional file 1: **Table S3**).

Discussion

The congruence between the phylogeny of Nitrososphaeria and their ecological preferences was relatively limited, suggesting the existence of a considerable ecotypical intra-diversity within this class [6] due to high diversification rates [17]. These observations corroborate work conducted on marine systems, which shows that closely related AOA isolates could differ in a range of physiological traits [72, 73]. Extensive niche specialization could be due to autotrophic *vs.* mixotrophic/heterotrophic growth and preferences for C compounds [20] in addition to differences in affinity for ammonia [21, 22]. Observed differences in the balances between lineages at fine phylogenetic scale also reflect a low level of intra-genomic heterogeneity in the 16S rRNA gene [74], and highlight the relevance of using ASVs to study the environmental determinants of soil Nitrososphaeria (more than 10 times as many nitrososphaerial ASVs than OTUs were detected).

By using MRT, we could evaluate the hierarchical effects of environmental filters in the delineation of ecological preferences of nitrososphaerial lineages. The machine learning approach used also allowed us to go beyond simple correlations when exploring drivers of archaeal diversity. Across our broad geographical and environmental gradients, MAT was the most prominent variable in the MRT analysis and ranked among the best predictors of evenness. Temperature affects composition and nitrification activity of AOA communities [25, 26, 75], but since MAT and latitude correlated (Spearman's r = 0.9), MAT could also reflect the importance of spatial distance. Soil pH was another significant predictor of phylogenetic diversity and evenness of Nitrososphaeria across the European gradient. This agrees with the dominant idea that soil pH drives diversification of terrestrial Thaumarchaeota at broad phylogenetic scales [17–19, 27]. However, the balance shifts in the shallow nodes in our study suggest recent adaptations in relation to soil pH. This would imply multiple independent specializations to acidic pH in Nitrososphaeria, which support the conclusions of a recent study based on analyses of the *amoA* gene [6]. The limited genomic information available on Nitrososphaerales prevents us using the congruence between 16S rRNA and *amoA* phylogenies [6, 27] to match the clades observed in this study for identifying *amoA*-based low-pH AOA lineages (within clades NS-α, -β, -γ and -ζ in [6]). Nevertheless, since pH controls the equilibrium between ammonia and ammonium in soils, these adaptations could reflect differences in substrate affinity [21, 22], possibly through distinct molecular adaptations of the ammonia monooxygenase [76]. Different N-related variables such as CEC, C:N ratio were also important predictors of the diversity and/or evenness of Nitrososphaeria and for the abundance of AOA, it was Total N. These results illustrate the expected selective role played by soil N, and are consistent with demonstrated effects of soil C:N ratio [4, 77] and ammonium supply [78, 79] on the abundance and structure of soil AOA communities. Random forest modeling revealed that nearly of the above-mentioned relationships were non-linear, and some ALE curves suggested the existence of thresholds (TH). The increase of nitrososphaerial PD in the low range of the C:N ratio (TH ~ 10) could be due to preferences for mineralized N from organic matter [80]. This aligns with the observed sharp increase of archaeal nitrification rates at C:N ratios below 20 [81], indicating a coupling between nitrososphaerial diversity levels and nitrification activity. However, the high number of significant shallow nodes in both high and low C:N ratio clusters in Southern Europe shows that niche specialization in Nitrososphaeria extends past

preferences in soil C:N. For example, total N was a better predictor for AOA abundances than C:N, showing a strong positive effect across the measured range, although it appeared weaker between 2-3.5 g N kg^{-1} soil. The capacity for cation exchange was also important, likely by influencing the retention of ammonium. We identified a threshold above which CEC had a strong negative effect on nitrososphaerial PD (TH ~ 15 cmol kg^{-1}). Soil CEC is linked to soil texture and clay content has previously been reported to negatively affect the abundance of AOA [13].

For the overall soil archaeal communities, mean annual temperature (MAT) was, similarly as for the Nitrososphaeria, an important factor and the main driver of *β*-diversity. However, the importance of MAT also reflects the effect of distance on the changes in community composition, which was further supported by the high fitness value of the distance-decay relationship (R^2_{adj} = 0.22) along the large gradient studied. This observation contrasts with a recent survey where substantially weaker decays of community similarity were found in archaeal communities in maize and rice fields across Eastern China [82]. At large spatial scales, distance-decay relationships are typically influenced by dispersal limitation (stochastic process) and species sorting (deterministic process), *i.e.* the combined effect of environmental filtering and biotic interactions [83]. In the present study, dispersal limitation and environmental filtering had comparable effects on the overall *β*-diversity of archaea, whereas the few other studies of archaeal communities across broad spatial scales in arable soils have reported larger sorting:dispersal effect ratios [40, 82]. The relative importance of dispersal limitation in the present study could have been overestimated, since some of the variation explained by spatial factors alone likely encompass unmeasured environmental variables and we could have missed environmental variables that are relevant for archaea. Nevertheless, the use of ASVs instead of OTUs and the removal of the rare taxa combined with broader environmental and geographic gradients, as we did, should lead to a more accurate assessment of the relative importance of environmental filtering versus dispersal limitation for the assembly of archaeal communities when compared to other studies [40, 82]. Our results thus indicate that the structure of archaeal communities associated to fields under cereal cultivation could be less sensitive to changes in environmental conditions than previously thought. Balanced effects of stochastic versus deterministic processes were recently found to promote diverse, yet uneven ecosystem functions in a study comparing three types of agroecosystems [84]. This interpretation of effects of balanced stochastic and deterministic processes fits with the observed low evenness of the archaeal community and the presence of a high diversity of low abundant groups harboring potentially diverse functional capabilities at the continental scale.

Archaeal classes present at low abundance exhibited different responses to the environmental factors evaluated. Soil pH displayed a negative relationship with both PD and evenness in Group 1.1c thaumarchaeota. Accordingly, these microorganisms have mainly been detected in acidic environments [85, 86]. The few known representatives of this group do not have *amoA* homologs and do not produce nitrite or nitrate in culture. It is therefore hypothesized that Group 1.1c is a non-ammonia oxidizing lineage within the Thaumarchaeota [71, 87], which is supported by the lack of significant effects of N-related variables observed in this study. Woesearchaeia and the two classes of methanogens shared only a few

environmental preferences (*i.e.* elevation and clay content) with regards to *α*-diversity, despite recent findings that they that are potential metabolic partners [53]. This partnership or their importance in arable soils thus remain elusive. Finally, diversity and evenness of the Thermoplasmata, which was the second most abundant class in terms of both ASVs and relative abundance, appeared to be predominantly driven by MAT and soil texture along the gradient. Their ecology in agricultural soils is largely unknown as Thermoplasmata have mainly been studied in acid mine drainage [88] and hot environments [89]. Members of this class however account for ca. 5 % of archaeal sequences in global soil samples [3] and have been detected at high abundances in deeper soil layers in both boreal [90] and temperate, acidic deciduous forests [91]. There are few cultivated representatives, but more than 400 genomes available that suggest a versatile metabolic potential for this class, including methanogenesis, sulfur cycling and even dinitrogen fixation [92]. Sulfur cycling was also detected in a genome obtained from peat soil [87], while methanogenic Thermoplasmata appear to be widespread in wetlands [93]. Several low-abundant archaeal lineages could thus be involved in C and S cycling and play a more important role than previously thought in arable soils.

Conclusions

This study provides novel insights into the ecology of archaea, particularly putative ammonia-oxidizing Nitrososphaeria, in arable soils across Europe. Our results suggest extensive ecophysiological differences between closely related Nitrososphaeria, and underline the multifactorial nature of niche differentiation in this class. Both MAT and soil C:N ratio were unexpectedly better predictors of their diversity and distribution than soil pH, for which multiple, independent adaptations were inferred. Thresholds for soil C:N and CEC were also identified. Overall, we show that future studies aiming at deciphering the ecology of Nitrososphaeria should be performed at fine phylogenetic scale, using methods accounting for non-linear relationships between environmental drivers and the diversity of these functionally important archaea in arable soils.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Sequencing data has been deposited at the European Nucleotide Archive (ENA) under the accession number PRJEB35080. The datasets and code generated during the current study are available in the

Zenodo repository (http://doi.org/10.5281/zenodo.4095504). They include metadata, R code, *amoA* gene abundances and the phylogeny in newick format.

Competing interests

The authors declare that they have no competing interests.

Funding

The Digging Deeper project was funded through the 2015-2016 BiodivERsA call, with national funding from the Swiss National Science Foundation (grant 31BD30-172466), the Deutsche Forschungsgemeinschaft (grant 317895346), the Swedish Research Council Formas (grant 2016-0194), the Spanish Ministerio de Economía y Competitividad (grant PCIN-2016-028) and the Agence Nationale de la Recherche (grant ANR-16-EBI3-0004-01). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

Authors' contributions

A.S., S.H., M.G.A.v.d.H, F.T.M., L.P. and M.C.R. initiated the study and planned the field work. A.S., S.B., F.D., A.E., P.G-P., G.G., C.H., D.S.P., and S.R. contributed to data collection. A.S. performed the analyses and drafted the manuscript together with S.H. All authors commented on and approved the final manuscript.

Acknowledgements

We thank Christopher Jones for help with plotting the balances on the phylogeny.

References

1. Offre P, Spang A, Schleper C. Archaea in biogeochemical cycles. Annu Rev Microbiol. 2013;67:437–57.
2. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. Science. 2008;320:1034–9.
3. Auguet JC, Barberan A, Casamayor EO. Global ecological patterns in uncultured Archaea. ISME J. 2010;4:182–90.
4. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. ISME J. 2011;5:908–17.
5. Tripathi BM, Kim M, Tateno R, Kim W, Wang J, Lai-Hoe A, et al. Soil pH and biome are both key determinants of soil archaeal community structure. Soil Biol Biochem. 2015;88:1–8.
6. Alves RJE, Minh BQ, Urich T, Von Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. Nat Commun.

2018;9:1517.

7. Stahl DA, de la Torre JR. Physiology and diversity of ammonia-oxidizing archaea. Annu Rev Microbiol. 2012;66:83–101.
8. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nat Rev Microbiol. 2018;16:263–76.
9. Lassaletta L, Billen G, Grizzetti B, Anglade J, Garnier J. 50 year trends in nitrogen use efficiency of world cropping systems: The relationship between yield and nitrogen input to cropland. Environ Res Lett. 2014;9:105011.
10. Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature. 2006;442:806–9.
11. Prosser JI, Nicol GW. Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. Environ Microbiol. 2008;10:2931–41.
12. Schauss K, Focks A, Leininger S, Kotzerke A, Heuer H, Thiele-Bruhn S, et al. Dynamics and functional relevance of ammonia-oxidizing archaea in two agricultural soils. Environ Microbiol. 2009;11:446–56.
13. Wessén E, Söderström M, Stenberg M, Bru D, Hellman M, Welsh A, et al. Spatial distribution of ammonia-oxidizing bacteria and archaea across a 44-hectare farm related to ecosystem functioning. ISME J. 2011;5:1213–25.
14. Gubry-Rangin C, Williams W, Prosser JI. Approaches to understanding the ecology and evolution of understudied terrestrial archaeal ammonia-oxidisers. Emerg Top Life Sci. 2018;2:619–28.
15. Bru D, Ramette A, Saby NPA, Dequiedt S, Ranjard L, Jolivet C, et al. Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. ISME J. 2011;5:532–42.
16. Hu HW, Zhang LM, Dai Y, Di HJ, He JZ. pH-dependent distribution of soil ammonia oxidizers across a large geographical scale as revealed by high-throughput pyrosequencing. J Soils Sediments. 2013;13:1439–49.
17. Gubry-Rangin C, Kratsch C, Williams TA, McHardy AC, Embley TM, Prosser JI, et al. Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. Proc Natl Acad Sci USA. 2015;112:9370–5.
18. Gubry-Rangin C, Hai B, Quince C, Engel M, Thomson BC, James P, et al. Niche specialization of terrestrial archaeal ammonia oxidizers. Proc Natl Acad Sci USA. 2011;108:21206–11.
19. Nicol GW, Leininger S, Schleper C, Prosser JI. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. Environ Microbiol. 2008;10:2966–78.
20. Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. Trends Microbiol. 2012;20:523–31.
21. Hink L, Lycus P, Gubry-Rangin C, Frostegård Å, Nicol GW, Prosser JI, et al. Kinetics of NH_3-oxidation, NO-turnover, N_2O-production and electron flow during oxygen depletion in model bacterial and

archaeal ammonia oxidisers. Environ Microbiol. 2017;19:4882–96.

22. Lehtovirta-Morley LE, Ross J, Hink L, Weber EB, Gubry-Rangin C, Thion C, et al. Isolation of "*Candidatus* Nitrosocosmicus franklandus", a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. FEMS Microbiol Ecol. 2016;92:fiw057.
23. Placella SA, Firestone MK. Transcriptional response of nitrifying communities to wetting of dry soil. Appl Environ Microbiol. 2013;79:3294–302.
24. Thion C, Prosser JI. Differential response of nonadapted ammonia-oxidising archaea and bacteria to drying-rewetting stress. FEMS Microbiol Ecol. 2014;90:380–9.
25. Gubry-Rangin C, Novotnik B, Mandič-Mulec I, Nicol GW, Prosser JI. Temperature responses of soil ammonia-oxidising archaea depend on pH. Soil Biol Biochem. 2017;106:61–8.
26. Tourna M, Freitag TE, Nicol GW, Prosser JI. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environ Microbiol. 2008;10:1357–64.
27. Oton EV, Quince C, Nicol GW, Prosser JI, Gubry-Rangin C. Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. ISME J. 2016;10:85–96.
28. Aigle A, Gubry-Rangin C, Thion C, Estera-Molina KY, Richmond H, Pett-Ridge J, et al. Experimental testing of hypotheses for temperature- and pH-based niche specialisation of ammonia oxidising archaea and bacteria. Environ Microbiol. 2020;22:4032–45.
29. Garland G, Edlinger A, Banerjee S, Degrune F, García-Palacios P, Pescador DS, et al. Crop cover is more important than rotational diversity for soil multifunctionality and cereal yields in European cropping systems. Nat Food. 2021;2:28–37.
30. Jiao S, Xu Y, Zhang J, Lu Y. Environmental filtering drives distinct continental atlases of soil archaea between dryland and wetland agricultural ecosystems. Microbiome. 2019;7:15.
31. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
32. De'ath G. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology. 2002;83:1105–17.
33. Dini-Andreote F, Stegen JC, Van Elsas JD, Salles JF. Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. Proc Natl Acad Sci USA. 2015;112:E1326–32.
34. FAL FAW. RAC. Referenzmethoden der Eidg. landwirtschaftlichen Forschungsanstalten. 1. Bodenuntersuchung zur Düngeberatung. Zürich- Reckenholz; 1996.
35. Chamberlain S rnoaa: "NOAA" weather data from R. 2019.
36. Hollister J, Shah T. elevatr: access elevation data from various APIs. 2017.
37. Takai K, Horikoshi K. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl Environ Microbiol. 2000;66:5066–72.
38. Herlemann DPR, Labrenz M, Jürgens K, Bertilsson S, Waniek JJ, Andersson AF. Transitions in bacterial communities along the 2 000 km salinity gradient of the Baltic Sea. ISME J. 2011;5:1571–9.

39. George PBL, Lallias D, Creer S, Seaton FM, Kenny JG, Eccles RM, et al. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. Nat Commun. 2019;10:1107.
40. Liu J, Yu Z, Yao Q, Sui Y, Shi Y, Chu H, et al. Biogeographic distribution patterns of the archaeal communities across the black soil zone of Northeast China. Front Microbiol. 2019;10:23.
41. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. 2019.
42. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.
43. Hunt D, David L, Gevers D, Preheim S, Alm E, Polz M. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science. 2008;320:1081–5.
44. Larkin AA, Martiny AC. Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. Environ Microbiol Rep. 2017;9:55–70.
45. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.
46. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: community ecology package. 2018.
47. Price MN, Dehal PS, Arkin AP. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.
48. Hubbell S. The unified neutral theory of biodiversity and biogeography. Priceton: Princeton University Press; 2001.
49. Magurran AE. Species abundance distributions over time. Ecol Lett. 2007;10:347–54.
50. Krebs CJ. Ecological Methodology. 2nd ed. New York: Addison-Wesley Educational Publishers;: Inc; 1999.
51. Fillol M, Auguet JC, Casamayor EO, Borrego CM. Insights in the ecology and evolutionary history of the Miscellaneous Crenarchaeotic Group lineage. ISME J. 2016;10:665–77.
52. Jeanbille M, Gury J, Duran R, Tronczynski J, Agogué H, Saïd O, Ben, et al. Response of core microbial consortia to chronic hydrocarbon contaminations in coastal sediment habitats. Front Microbiol. 2016;7:1637.
53. Liu X, Li M, Castelle CJ, Probst AJ, Zhou Z, Pan J, et al. Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. Microbiome. 2018;6:102.
54. Mo Y, Zhang W, Yang J, Lin Y, Yu Z, Lin S. Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. ISME J. 2018;12:2198–210.
55. Martín-Fernández JA, Hron K, Templ M, Filzmoser P, Palarea-Albaladejo J. Bayesian-multiplicative treatment of count zeros in compositional data sets. Stat Modelling. 2015;15:134–58.

56. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. Math Geol. 2003;35:279–300.
57. Silverman JD, Washburne AD, Mukherjee S, David LA. A phylogenetic transform enhances analysis of compositional microbiota data. Elife. 2017;6:e21887.
58. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol. 2017;8:2224.
59. Morton JT, Sanders J, Quinn RA, Mcdonald D, Gonzalez A, Vázquez-Baeza Y, et al. Balance trees reveal microbial niche differentiation. mSystems. 2017;2:e00162-16.
60. Faith DP. Conservation evaluation and phylogenetic diversity. Biol Conserv. 1992;61:1–10.
61. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26:1463–4.
62. Pielou EC. The measurement of diversity in different types of biological collections. J Theor Biol. 1966;13:131–44.
63. Genuer R, Poggi JM, Tuleau-Malot C. VSURF: an R package for variable selection using random forests. R J. 2015;7:19–33.
64. Liaw A, Wiener M. Classification and regression by randomForest. R News. 2002;2:18–22.
65. Molnar C, Bischl B, Casalicchio G. iml: an R package for interpretable machine learning. J Open Source Softw. 2018;3:786.
66. Apley DW, Zhu J. Visualizing the effects of predictor variables in black box supervised learning models. J R Stat Soc Ser B Stat Methodol. 2020;82:1059–86.
67. Power JF, Carere CR, Lee CK, Wakerley GLJ, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018;9:2876.
68. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. California: Wadsworth Publishing Company;: Belmont; 1984.
69. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8:28–36.
70. Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr Biol. 2015;25:690–701.
71. Weber EB, Lehtovirta-Morley LE, Prosser JI, Gubry-Rangin C. Ammonia oxidation is not required for growth of group 1.1c soil Thaumarchaeota. FEMS Microbiol Ecol. 2015;91:fiv001.
72. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. Proc Natl Acad Sci USA. 2014;111:12504–9.
73. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JAL, et al. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. ISME J.

2016;10:1051–63.

74. Sun DL, Jiang X, Wu QL, Zhou NY. Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. Appl Environ Microbiol. 2013;79:5962–9.
75. Alves RJE, Wanek W, Zappe A, Richter A, Svenning MM, Schleper C, et al. Nitrification rates in Arctic soils are associated with functionally distinct populations of ammonia-oxidizing archaea. ISME J. 2013;7:1620–31.
76. Macqueen DJ, Gubry-Rangin C. Molecular adaptation of ammonia monooxygenase during independent pH specialization in Thaumarchaeota. Mol Ecol. 2016;25:1986–99.
77. Jiang H, Huang L, Deng Y, Wang S, Zhou Y, Liu L, et al. Latitudinal distribution of ammonia-oxidizing bacteria and archaea in the agricultural soils of Eastern China. Appl Environ Microbiol. 2014;80:5593–602.
78. Verhamme DT, Prosser JI, Nicol GW. Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. ISME J. 2011;5:1067–71.
79. Hink L, Gubry-Rangin C, Nicol GW, Prosser JI. The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. ISME J. 2018;12:1084–93.
80. Levičnik-Höfferle Š, Nicol GW, Ausec L, Mandić-Mulec I, Prosser JI. Stimulation of thaumarchaeal ammonia oxidation by ammonia derived from organic nitrogen but not added inorganic nitrogen. FEMS Microbiol Ecol. 2012;80:114–23.
81. Lu X, Bottomley PJ, Myrold DD. Contributions of ammonia-oxidizing archaea and bacteria to nitrification in Oregon forest soils. Soil Biol Biochem. 2015;85:54–62.
82. Jiao S, Yang Y, Xu Y, Zhang J, Lu Y. Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. ISME J. 2019;14:202–16.
83. Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol. 2012;10:497–506.
84. Liu W, Graham EB, Dong Y, Zhong L, Zhang J, Qiu C, et al. Balanced stochastic versus deterministic assembly processes benefit diverse yet uneven ecosystem functions in representative agroecosystems. Environ Microbiol. 2021;23:391–404.
85. Tripathi BM, Kim M, Lai-Hoe A, Shukor NAA, Rahim RA, Go R, et al. pH dominates variation in tropical soil archaeal diversity and community structure. FEMS Microbiol Ecol. 2013;86:303–11.
86. Lehtovirta LE, Prosser JI, Nicol GW. Soil pH regulates the abundance and diversity of Group 1.1c Crenarchaeota. FEMS Microbiol Ecol. 2009;70:367–76.
87. Lin X, Handley KM, Gilbert JA, Kostka JE. Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. ISME J. 2015;9:2740–4.
88. Baker BJ, Banfield JF. Microbial communities in acid mine drainage. FEMS Microbiol Ecol. 2003;44:139–52.

89. Massello FL, Chan CS, Chan KG, Goh KM, Donati E, Urbieta MS. Meta-analysis of microbial communities in hot springs: recurrent taxa and complex shaping factors beyond ph and temperature. Microorganisms. 2020;8:906.
90. Kemnitz D, Kolb S, Conrad R. High abundance of Crenarchaeota in a temperate acidic forest soil. FEMS Microbiol Ecol. 2007;60:442–8.
91. Isoda R, Hara S, Tahvanainen T, Hashidoko Y. Comparison of archaeal communities in mineral soils at a boreal forest in Finland and a cold-temperate forest in Japan. Microbes Environ. 2017;32:390–3.
92. Baker BJ, De Anda V, Seitz KW, Dombrowski N, Santoro AE, Lloyd KG. Diversity, ecology and evolution of Archaea. Nat Microbiol. 2020;5:887–900.
93. Söllinger A, Schwab C, Weinmaier T, Loy A, Tveit AT, Schleper C, et al. Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat. FEMS Microbiol Ecol. 2016;92:fiv149.

Figures

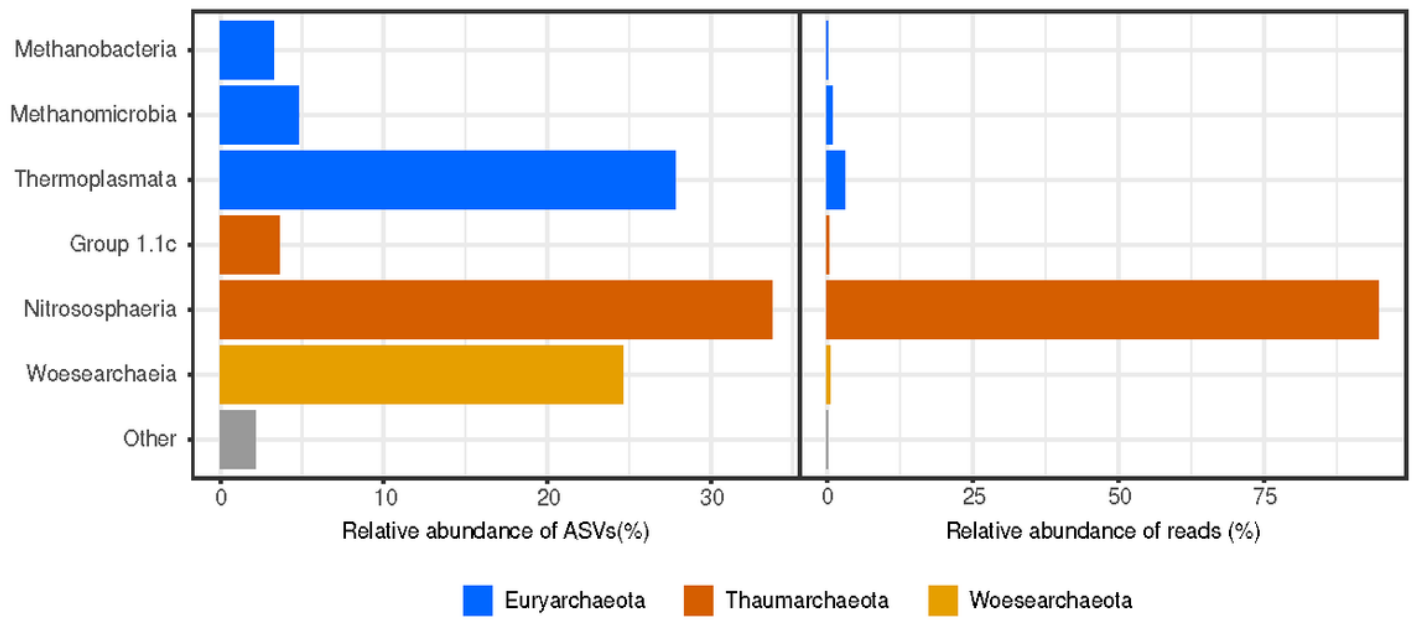

Figure 1. Saghaï et al.

Figure 1

Relative abundances of high-rank archaeal taxa in the rarefied dataset of 16S rRNA gene sequences. Abundances are shown in terms of ASVs (left panel) and reads (right panel).

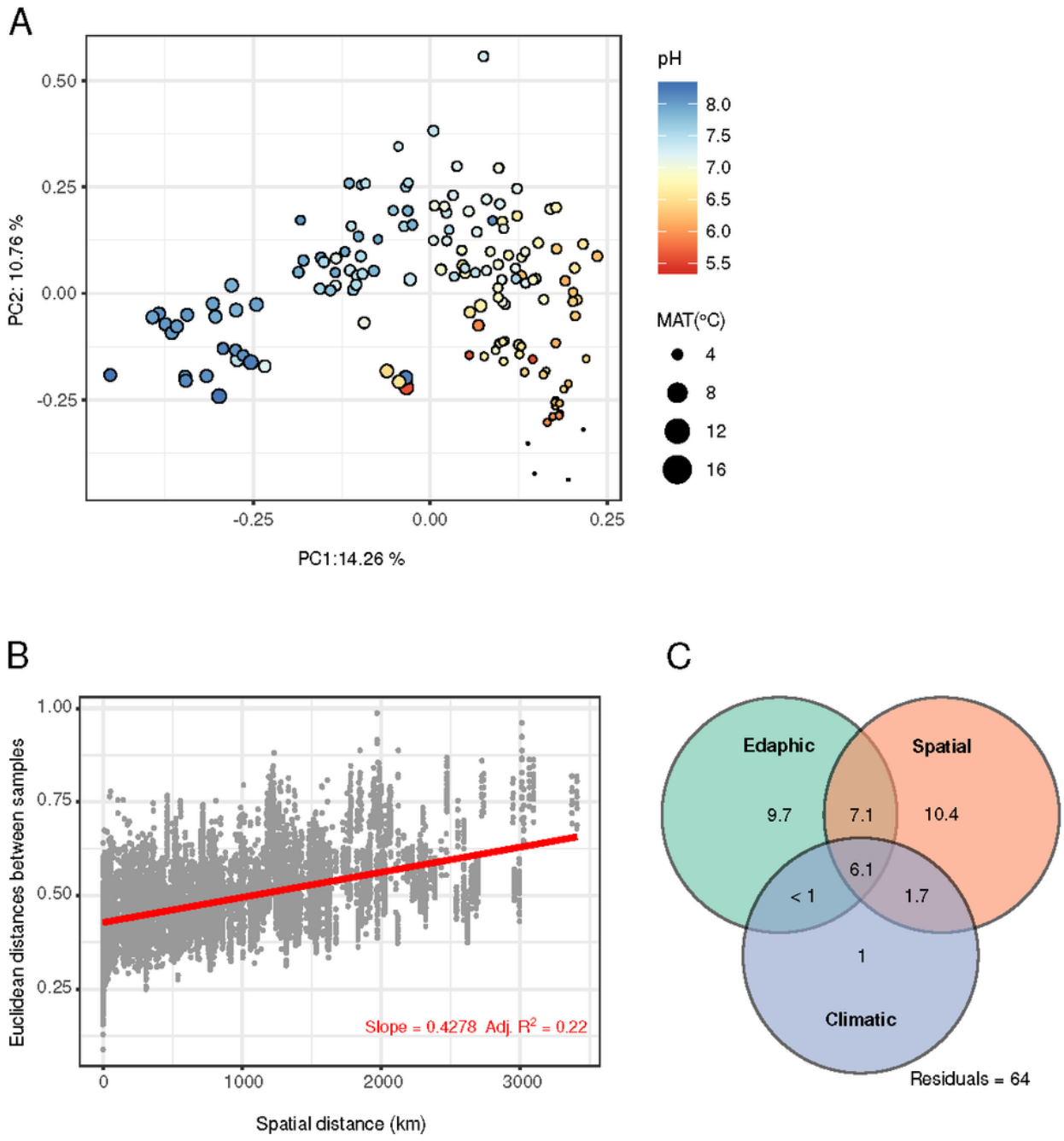

Figure 2. Saghaï et al.

Figure 2

Factors driving the variation in archaeal community composition and structure across the European gradient. a Principal Component Analysis (PCA) showing differences in archaeal communities between all samples and the associated changes in pH and MAT, identified as the two best explanatory variables (PERMANOVA, $p < 0.001$). b Distance-decay relationship between geographic distance and community similarity. The red line indicates the ordinary least squares linear regression ($p < 0.001$). c Variation

partitioning analysis (VPA) between climatic, edaphic and spatial components. All fractions were significant ($p < 0.01$) and the variance explained is indicated (%). Both PCA and VPA were performed on the philr-transformed data.

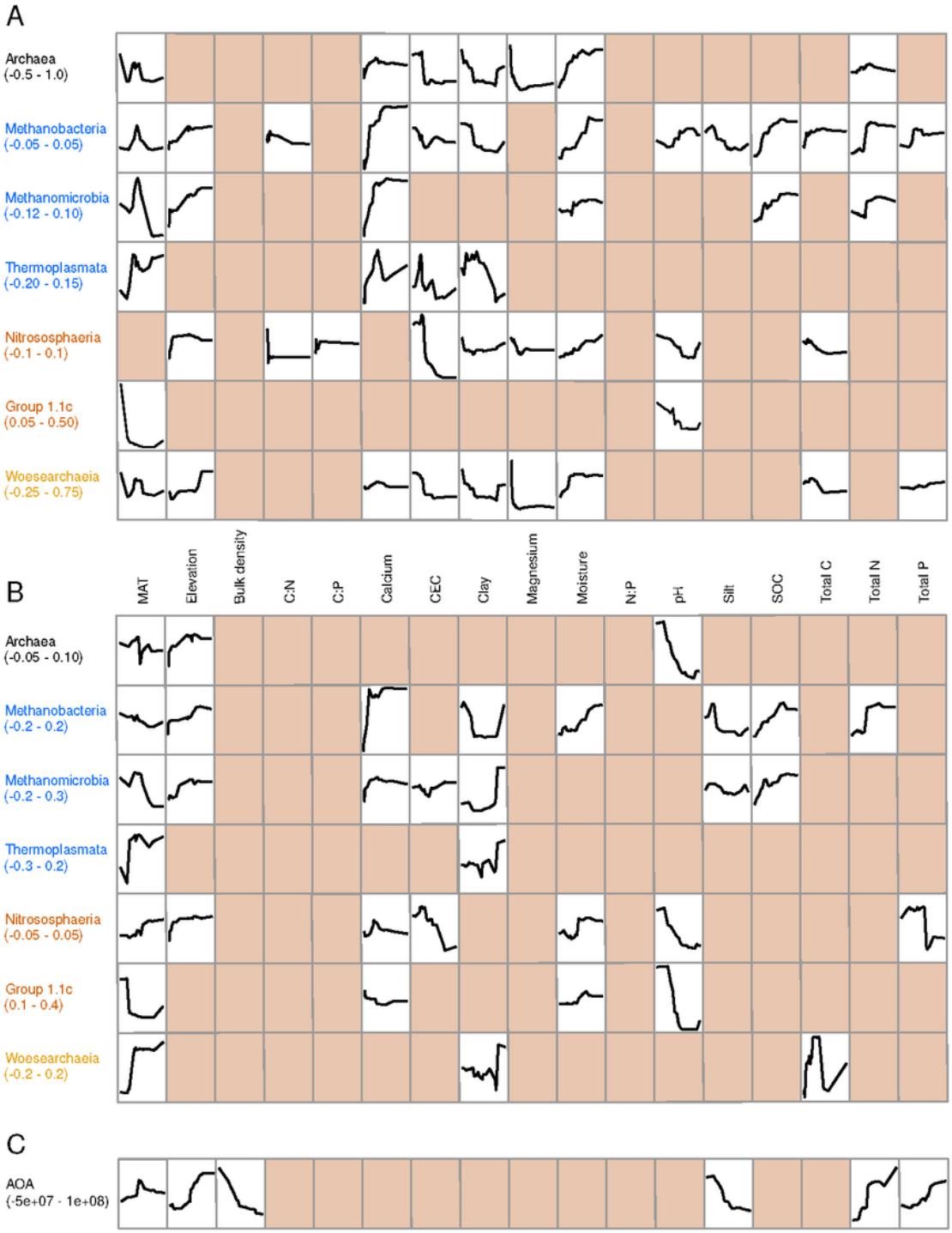

Figure 3. Saghaï et al.

Figure 3

Relationship between environmental variables and (a) phylogenetic diversity (PD), (b) evenness of archaeal communities, and (c) the abundance of AOA, based on random forest (RF) analyses. Variables

selected by VSURF (x-axis; see Table 1 for units and range) were used to generate accumulated local effects plots, which show how the prediction of the response variables (PD, evenness or abundance of AOA) changes along the range of each environmental variable in each of the RF models (y-axis; range indicated in brackets). Model parameters and fit are indicated in Additional file 1: Table S1. Full plots are available as Additional file 1: Figures S5, S6 and S7b.

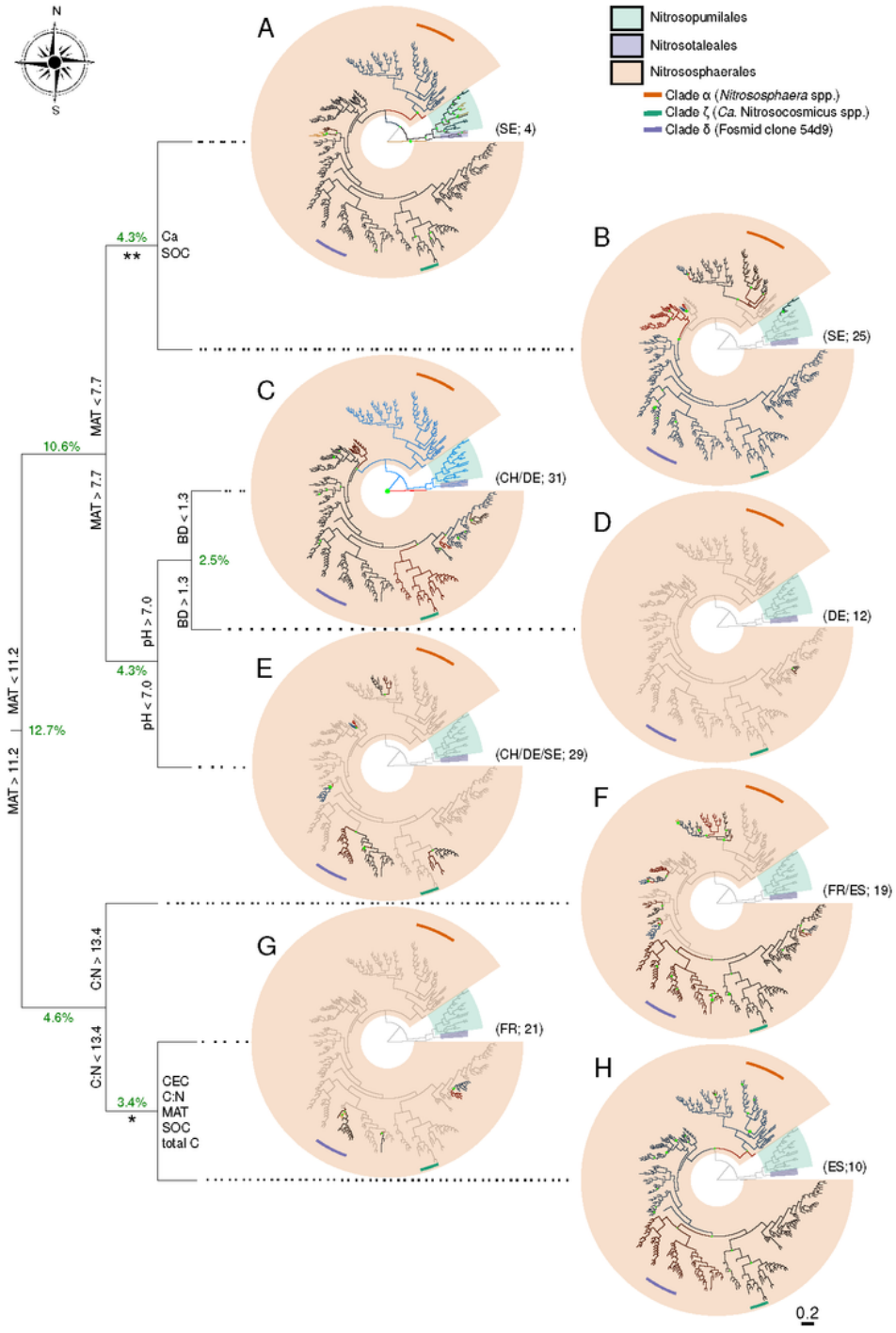

Figure 4. Saghaï et al.

Figure 4

Effect of environmental parameters on the balances within the phylogeny of Nitrososphaeria. The variables contributing the most to explain the variation in community composition and structure were selected by multivariate regression trees and are displayed on the left (R2 = 0.42; see Table 1 for the units). In each environmental cluster (A-H), only the significant nodes (green dots; $p < 0.01$) and their associated balances were plotted (country of origin and number of samples are indicated in brackets; CH: Switzerland, DE: Germany, ES: Spain, FR: France and SE: Sweden). The branch color and its intensity depict the direction and magnitude of change between two neighboring clades, relatively to each other (blue: increasing clade, red: decreasing clade). The scale bar represents the average substitutions per site in the phylogeny. The clades within Nitrososphaerales correspond to Nitrososphaera spp. (α), Nitrosocosmicus sp. (ζ) and fosmid clone 54d9 (δ) from Alves et al., 2018. *Variables with equal predictive power: cation exchange capacity (< 17.0 to the right), C:N ratio (< 13.7 to the left), mean annual temperature (< 5.0 to the right), SOC (< 22.1 to the left) and total C (< 26.1 to the left); **Variables with equal predictive power: calcium (< 1.8 to the right) and soil organic carbon (SOC; < 6.6 to the left).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf