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

Aurélien Saghaï [iD],[1] Samiran Banjeree,[2†]
Florine Degrune,[3,4†] Anna Edlinger,[2,5†]
Pablo García-Palacios,[6†] Gina Garland,[2,7,8†]
Marcel G. A. van der Heijden,[2,5†] Chantal Herzog,[2,5†]
Fernando T. Maestre,[9,10†] David S. Pescador,[11†]
Laurent Philippot,[12†] Matthias C. Rillig,[3,4†]
Sana Romdhane[12†] and Sara Hallin [iD][1*]

[1]*Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden.*

[2]*Plant-Soil Interactions Group, Agroscope, Zurich, Switzerland.*

[3]*Institute of Biology, Freie Universität Berlin, Berlin, Germany.*

[4]*Berlin-Brandenburg Institute of Advanced Biodiversity Research, Berlin, Germany.*

[5]*Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland.*

[6]*Instituto de Ciencias Agrarias, Consejo Superior de Investigaciones Científicas, Madrid, Spain.*

[7]*Soil Quality and Use Group, Agroscope, Zurich, Switzerland.*

[8]*Department of Environmental System Sciences, Soil Resources Group, ETH Zurich, Zurich, Switzerland.*

[9]*Instituto Multidisciplinar para el Estudio del Medio "Ramón Margalef", Universidad de Alicante, Alicante, Spain.*

[10]*Departamento de Ecología, Universidad de Alicante, Alicante, Spain.*

[11]*Departamento de Biología y Geología, Física y Química Inorgánica, Universidad Rey Juan Carlos, Escuela Superior de Ciencias Experimentales y Tecnología, Móstoles, Spain.*

[12]*Department of Agroecology, University of Bourgogne Franche-Comté, INRAE, AgroSup Dijon, Dijon, France.*

Summary

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. We used phylogenetic information at fine scale and machine learning approaches to identify climatic, edaphic and geomorphological drivers of Nitrososphaeria and other archaea along a 3000 km European gradient. Only limited insights into the ecology of the low-abundant archaeal classes could be inferred, but our analyses underlined the multifactorial nature of niche differentiation within Nitrososphaeria. Mean annual temperature, C:N ratio and pH were the best predictors of their diversity, evenness and distribution. Thresholds in the predictions could be defined for C:N ratio and cation exchange capacity. Furthermore, multiple, independent and recent specializations to soil pH were detected in the Nitrososphaeria phylogeny. The coexistence of widespread ecophysiological differences between closely related soil Nitrososphaeria highlights that their ecology is best studied at fine phylogenetic scale.

Introduction

Archaea are pivotal for the functioning of all major biomes, as they play a critical role in both carbon (C) and nitrogen (N) cycles (Falkowski *et al*., 2008; Offre *et al*., 2013). In terrestrial ecosystems, archaeal communities tend to be phylogenetically clustered and are commonly dominated by Thaumarchaeota (Auguet *et al*., 2010; Bates *et al*., 2011; Tripathi *et al*., 2015). This phylum harbours the globally important ammonia-oxidizing archaea (AOA), restricted to the class Nitrososphaeria (Alves *et al*., 2018). All AOA characterized so far use the

ammonia monooxygenase enzyme (encoded by *amoABC* genes) to catalyze the first step of nitrification, the oxidation of ammonia to hydroxylamine (Vajrala *et al.*, 2013). Globally, nitrification contributes to the circulation of N (Kuypers *et al.*, 2018), but locally this process causes N losses, directly through nitrate leaching and production of the greenhouse gas nitrous oxide and indirectly by fuelling 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 (Lassaletta *et al.*, 2014). 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 (Leininger *et al.*, 2006; Prosser and Nicol, 2008; Schauss *et al.*, 2009; Wessén *et al.*, 2011) and, therefore, there is great interest in understanding their ecology and evolution (see Gubry-Rangin *et al.*, 2018 for a recent review).

Soil pH has previously been proposed as the main driver of thaumarchaeotal and AOA diversity (Bru *et al.*, 2011; Hu *et al.*, 2013; Tripathi *et al.*, 2015) and evolution at relatively broad phylogenetic scales (Nicol *et al.*, 2008; Gubry-Rangin *et al.*, 2011, 2015). However, this view has recently been challenged by Alves *et al.*, who suggested more recent adaptations to low pH from cosmopolitan clades (Alves *et al.*, 2018). In addition to soil pH, experimental work using isolates has linked niche differentiation in AOA to differences in ammonia affinity/tolerance and organic C preferences (Prosser and Nicol, 2012; Lehtovirta-Morley *et al.*, 2016; Hink *et al.*, 2017). Yet, the relative importance of other environmental factors, including C:N ratio (Bates *et al.*, 2011), moisture (Placella and Firestone, 2013; Thion and Prosser, 2014), temperature (Tourna *et al.*, 2008; Gubry-Rangin *et al.*, 2017) and soil organic carbon content (Oton *et al.*, 2016), 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 (Aigle *et al.*, 2020). 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 a fine phylogenetic scale by taking advantage of the congruence between 16S rRNA and *amoA* phylogenies (Oton *et al.*, 2016; Alves *et al.*, 2018). 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 3000 km European gradient, spanning

from northern Sweden to southern Spain (Garland *et al.*, 2021). Such continental surveys focusing on archaea are rare (Jiao, Xu, *et al.*, 2019a) 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 (Garland *et al.*, 2021), management effects were minimized. For the AOA, machine learning in terms of random forest modelling (Breiman, 2001) was used to predict 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 (De'ath, 2002), as opposed to previous studies that have assessed the importance of each variable individually (e.g. Gubry-Rangin *et al.*, 2011; Oton *et al.*, 2016; Alves *et al.*, 2018). 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. Accordingly, analyses were conducted on amplicon sequence variants (ASVs), instead of operational taxonomic units (OTUs) that may group similar species into a single OTU (Callahan *et al.*, 2017). Random forest modelling 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 (Dini-Andreote *et al.*, 2015).

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 (Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaea; Castelle *et al.*, 2015) (Fig. 1). The 19 ASVs with a frequency > 1% represented approximately 70% of the reads and, consequently, the archaeal communities were uneven ($J = 0.61 \pm 0.08$) 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
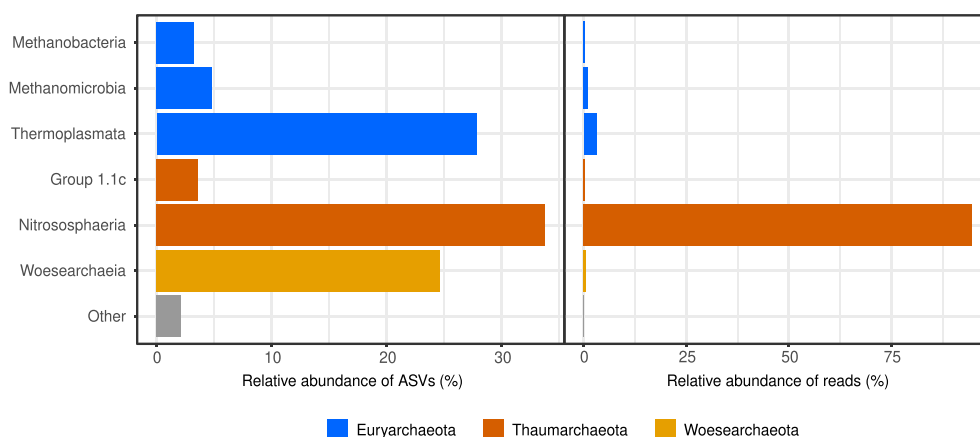
Fig. 1. Relative abundances of high-rank archaeal taxa in the rarefied dataset of 16S rRNA gene sequences across the transect represented by 151 soil samples. Abundances are shown in terms of ASVs (left panel) and reads (right panel). [Color figure can be viewed at wileyonlinelibrary.com]

of 97%, respectively. Most of the diversity was found within the order Nitrososphaerales (322 ASVs or approximately 91% of putative AO ASVs and approximately 98% of the reads), whereas Nitrosopumilales and Nitrosotaleales represented a minor fraction of the nitrososphaerial communities (approximately 1% and 8% of ASVs; < 1% and < 2% of the reads, respectively). Thermoplasmata and Woesearchaeia were diverse, respectively representing approximately 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 mean annual temperature (MAT) and pH ($R^2 = 0.11$ 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 ($p < 0.001$; Table S1). Climatic, edaphic and spatial factors collectively explained 36% of the variation in the archaeal community composition and all three groups of factors were significant ($p < 0.01$; Fig. 2C). When partitioning the variation, spatial distance (i.e. dispersal limitation) explained nearly as much variation in community composition as the edaphic factors (i.e. environmental filtering; approximately 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, S1 and S2). 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 does not oxidize ammonia (Weber *et al.*, 2015). The abundance of AOA, measured as the copy number of the archaeal *amoA* gene (Fig. S3A), was positively influenced by elevation, total N and total P, but tended to decrease with increasing silt content and bulk density (Figs. 3C and S3B).

Ecological preferences of nitrososphaerial taxa

Spatial distance explained a larger proportion of the variation in the structure of the nitrososphaerial communities than edaphic factors (14.6% and 11.2%, respectively), whereas MAT explained 1.4% (Fig. S4). The abundance of archaeal *amoA* genes, being a biotic factor predicting functionality, explained only 0.8% to the variation.

The sequential effect of several environmental filters was determined using multivariate regression trees
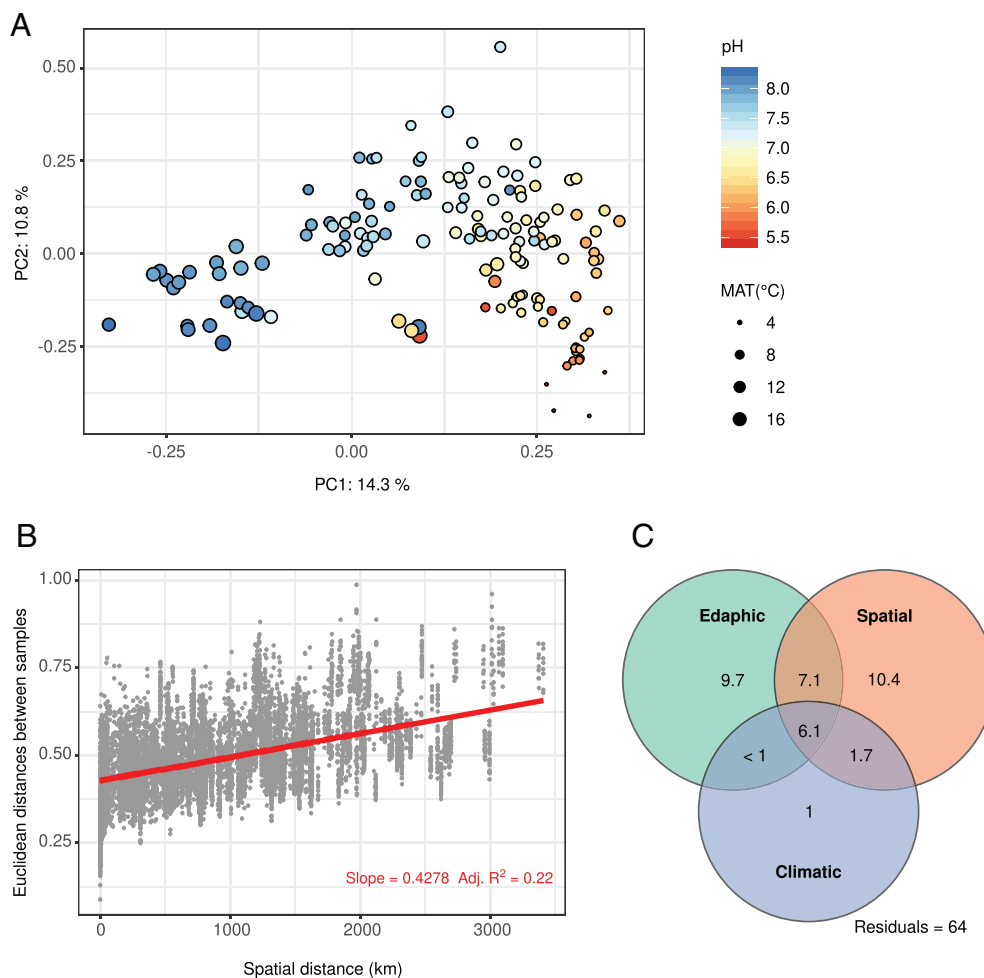
Fig. 2. Factors driving the variation in archaeal community composition and structure across the European gradient.
A. Principal component analysis (PCA) showing differences in archaeal communities between all samples and the associated changes in pH and MAT, identified as the two best explanatory variables (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 variation explained is indicated (%).
Both PCA and VPA were performed on the philr-transformed data. [Color figure can be viewed at wileyonlinelibrary.com]

(MRT), which delineate clusters of samples where the variation in environmental conditions is minimized. Each of the eight clusters (labelled A–H) contained 10–31 soil 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, with MAT also reflecting spatial distance. The MRT identified two to four levels of environmental filtering, with MAT being the most important driver. The importance of geographic distance was 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 (Table S2).

Balances, which depict relative changes in ASV abundance between two neighbouring clades relative to each other, were calculated for all inner nodes of the phylogenetic tree in each MRT cluster. They 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 in the

Table 1. Environmental variables used in this study.

Category	Variable	Range
Climatic	Mean annual temperature (MAT; °C)	3.7–19.5
Geomorphological	Elevation (m)	7.8–1022.0
Edaphic	Bulk density (g cm^{-3})	0.7–1.7
	C:N ratio	7.8–63.9
	C:P ratio	7.3–1042.9
	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

phylogenetic tree (indicated in green) 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; Table S2).

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 (Alves *et al*., 2018) due to high diversification rates (Gubry-Rangin *et al*., 2015). These observations corroborate work conducted on marine systems, where genomic and physiological data indicate that closely related AOA isolates could differ in a range of traits, including motility, pH preferences and substrate utilization (Qin *et al*., 2014; Bayer *et al*., 2016). Extensive niche specialization due to differences in affinity for ammonia (Lehtovirta-Morley *et al*., 2016; Hink *et al*., 2017; Jung *et al*., 2021) and for different organic C compounds in mixotrophic/heterotrophic growth have been discussed (Prosser and Nicol, 2012). However, organic acids have been shown to be used to detoxify the hydrogen peroxide produced during ammonia oxidation in several AOA isolates rather than used as an energy source (Kim *et al*., 2016) and the question of whether there are AOA able to grow

heterotrophically remains open. 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 (Sun *et al*., 2013), and highlight the relevance of using ASVs to study the environmental determinants of soil Nitrososphaeria (i.e. 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, whereas the random forest modelling allowed us to go beyond linear correlations when exploring drivers of archaeal diversity. Across our broad geographical and environmental gradients, MAT was the most prominent variable in the MRT analysis, as indicated by the major grouping between clusters A–B, C–E and F–H, and ranked among the best predictors of evenness. Temperature affects composition and nitrification activity of AOA communities (Tourna *et al*., 2008; Alves *et al*., 2013; Gubry-Rangin *et al*., 2017), but since MAT and latitude correlated (Spearman's $r = 0.9$, $p < 0.001$), MAT likely also reflects the importance of spatial distance. Results of the variation partitioning analysis, where spatial distance explained approximately 15% of the variation in the structure of the nitrososphaerial communities and MAT only approximately 1%, further support this interpretation. 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 (Nicol *et al*., 2008; Gubry-Rangin *et al*., 2011, 2015; Oton *et al*., 2016). However, we detected multiple balance shifts in the shallow nodes in the phylogenies corresponding to low-pH MRT clusters (i.e. A, B and E), which suggest that closely related nitrososphaerial taxa display distinct preferences with regards to soil pH. This would imply multiple independent and recent specializations to acidic pH in Nitrososphaeria, which aligns with the conclusions of a study based on analyses of the *amoA* gene (Alves *et al*., 2018). The limited genomic information available on Nitrososphaerales prevents us from using the congruence between 16S rRNA and *amoA* phylogenies (Oton *et al*., 2016; Alves *et al*., 2018) 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 Alves *et al*., 2018). Nevertheless, since pH controls the equilibrium between ammonia and ammonium in soils, these adaptations could reflect differences in substrate affinity (Lehtovirta-Morley *et al*., 2016; Hink *et al*., 2017), possibly through distinct molecular adaptations of the ammonia monooxygenase (Macqueen and Gubry-Rangin, 2016). Different N-related variables

such as CEC and C:N ratio were important predictors of the diversity and/or evenness of Nitrososphaeria; and for the abundance of archaeal *amoA* genes, it was total N. MAT (or distance) and elevation were thus the only variables linking α-diversity and predicted function (archaeal *amoA* genes). The abundance of archaeal
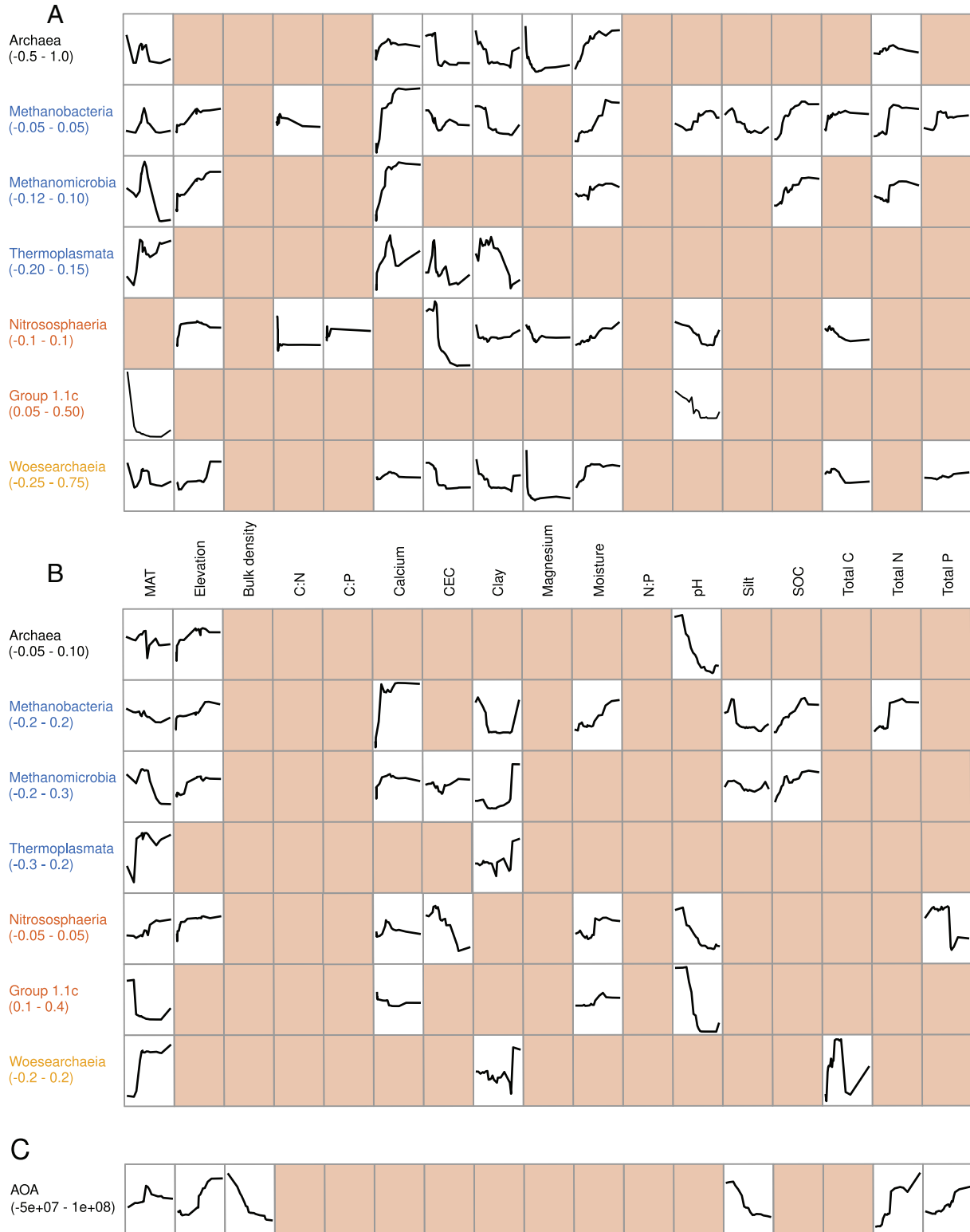

Fig. 3. Legend on next page.

amoA genes was also a poor predictor of nitrososphaerial β-diversity, suggesting a weak link between diversity of this class and the genetic potential for ammonia oxidation at broad geographic scale. Our results also illustrate the expected selective role played by soil N, and are consistent with demonstrated effects of soil C:N ratio (Bates *et al*., 2011; Jiang *et al*., 2014) and ammonium supply (Verhamme *et al*., 2011; Hink *et al*., 2018) on the abundance and structure of soil AOA communities. Random forest modelling revealed that nearly all of the above-mentioned relationships were nonlinear. Furthermore, some accumulated local effect curves, showing how the predictions change on average over the range of each individual environmental variable, suggested the existence of thresholds (TH). The increase of nitrososphaerial PD in the low range of the C:N ratio (TH $\sim$ 10) could be due to preferences for mineralized N from organic matter (Levičnik-Höfferle *et al*., 2012). This aligns with the observed sharp increase of archaeal nitrification rates at C:N ratios below 20 (Lu *et al*., 2015), indicating a coupling between nitrososphaerial diversity levels and nitrification activity. However, the high number of significant shallow nodes in the phylogeny of 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 and 3.5 g N kg^{-1} soil. The soils' CEC 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 $\sim$ 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 at the local scale (Wessén *et al*., 2011). We did not observe any effect of neither clay content or CEC on AOA abundance, which suggests that soil texture and its implications on CEC has a stronger effect on diversity than on abundance of AOA at the continental scale.

For the overall soil archaeal communities, MAT was an important factor and the main driver of β-diversity. As for the Nitrososphaeria, the importance of MAT also reflects the effect of distance on the changes in community composition, as shown by the higher explanatory power of spatial distance compared to MAT in the variation partitioning analysis and the high fitness value of the distance-decay relationship ($R^2_{adj} = 0.22$) along the geographical gradient. This latter 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 (Jiao *et al*., 2019b). 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 (Hanson *et al*., 2012). 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 (Liu *et al*., 2019; Jiao *et al*., 2019b). 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 (Liu *et al*., 2019; Jiao *et al*., 2019b). Our results thus indicate that the structure of archaeal communities associated to fields under cereal cultivation could be less influenced by changes in environmental conditions than previously suggested; an important finding in the context of ongoing climate change. Moreover, 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 (Liu *et al*., 2021). 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 harbouring potentially diverse functional capabilities at the continental scale.

Fig. 3. Relationship between environmental variables and α-diversity indices or 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 (phylogenetic diversity, 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 Table S3. Full plots are available as Figs. S1, S2 and S3B.
A. Phylogenetic diversity of archaeal communities.
B. Evenness of archaeal communities.
C. Abundance of AOA, measured as the copy number of the archaeal *amoA* gene (g^{-1} dw soil). [Color figure can be viewed at wileyonlinelibrary.com]
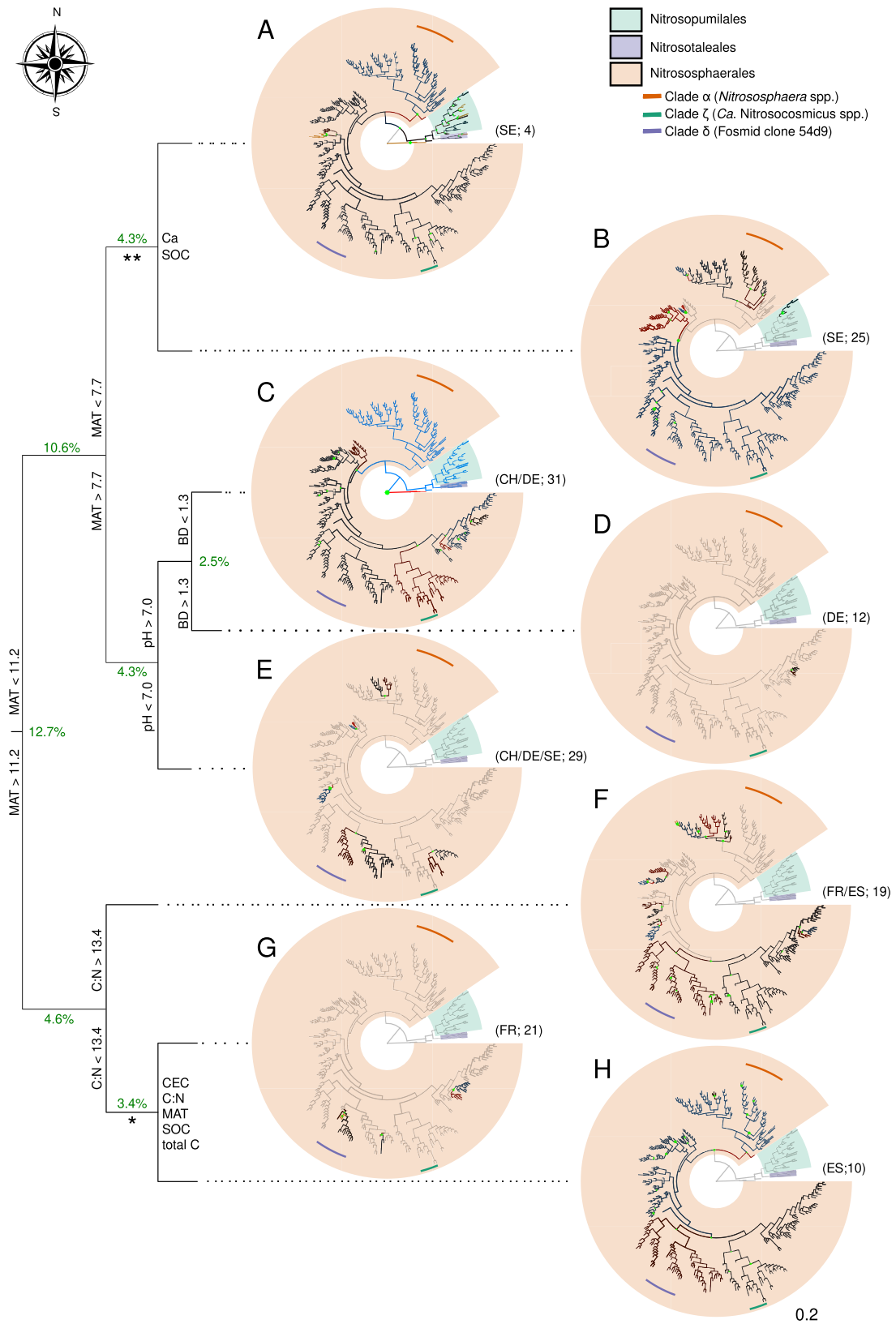
Fig. 4. Legend on next page.

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 (Lehtovirta *et al*., 2009; Tripathi *et al*., 2013). The few known representatives of this group do not have *amoA* homologues 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 (Lin *et al*., 2015; Weber *et al*., 2015), 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 (Liu *et al*., 2018). 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 (Baker and Banfield, 2003) and hot environments (Massello *et al*., 2020). Members of this class however account for approximately 5% of archaeal sequences in global soil samples (Auguet *et al*., 2010) and have been detected at high abundances in deeper soil layers in both boreal (Kemnitz *et al*., 2007) and temperate, acidic deciduous forests (Isoda *et al*., 2017). There are few cultivated representatives, but more than 400 genomes available that suggest a versatile metabolic potential for this class, including methanogenesis, sulfur (S) cycling and even dinitrogen fixation (Baker *et al*., 2020). Sulfur cycling was also detected in a genome obtained from peat soil (Lin *et al*., 2015), while methanogenic Thermoplasmata appear to be widespread in wetlands (Söllinger *et al*., 2016). 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.

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 (i.e. distance) 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.

Experimental procedures

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 (Garland *et al*., 2021) (Fig. S5). 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 × 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 (FAL *et al*., 1996). A fresh subsample of the composite soil was taken before drying for DNA extraction and stored at −20°C until DNA extraction (one extraction per sample). 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

Fig. 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 ($R^2 = 0.42$; 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 colour and its intensity depict the direction and magnitude of change between two neighbouring 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 (CEC; < 17.0 to the right), C:N ratio (< 13.7 to the left), mean annual temperature (MAT; < 5.0 to the right), soil organic carbon (SOC; < 22.1 to the left) and total C (< 26.1 to the left); **Variables with equal predictive power: calcium (Ca; < 1.8 to the right) and SOC (< 6.6 to the left). [Color figure can be viewed at wileyonlinelibrary.com]

package 'rnoaa' (v. 0.8.4; Chamberlain, 2019). The same GPS coordinates were also used to gather elevation data using the R package 'elevatr' (v. 0.2.0; Hollister and Shah, 2017).

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 (Takai and Horikoshi, 2000) and S-D-Bact-0785-a-A-21 (Herlemann *et al.*, 2011) to capture both Thaumarchaeota and low-abundant and under-studied groups typically present in soils (e.g. different classes of methanogens, Thermoplasmata; George *et al.*, 2019; Jiao *et al.*, 2019a; Liu *et al.*, 2019). The PCRs were run in duplicate 15 μl reactions under the following conditions: 3 min at 98°C, followed by 30 cycles of 98°C for 30 s, 65°C for 30 s and 72°C for 30 s and a final extension step of 10 min 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 s, 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) to verify the correct size and that there were no additional amplicons. All samples passed the control. 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 × 250 bp) platform.

Quantitative PCR analysis

The abundance of AOA 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; Tourna *et al.*, 2008), 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; Callahan *et al.*, 2016) to infer amplicon sequence variants (ASVs), which allow detection of ecological preferences at the finest phylogenetic scale (Hunt *et al.*, 2008; Larkin and Martiny, 2017). 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; Pruesse *et al.*, 2012) and classified using SINA's least common ancestor algorithm. After elimination of the bacterial ASVs, 2042 archaeal ASVs remained based on a total of 63 16 715 high-quality 16S rRNA gene amplicons (corresponding to approximately 70% of the original dataset). The reads were also clustered into OTUs at a similarity cut-off of 97% (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; Oksanen *et al.*, 2018) (Fig. S6) and a rarefied table ($n = 1079$ ASVs) was obtained by averaging the ASV counts over 1000 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; Price *et al.*, 2010) 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 (Hubbell, 2001) to split the dataset according to the frequency of occurrence of each ASV (Magurran, 2007). 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 (Krebs, 1999) (Fig. S7). By focusing on the frequent community members, we minimized the risk of sampling artefacts that would bias the distribution of the ASVs and thus increased the likelihood to detect relevant ecological patterns. Such partitioning remains rarely used in microbial ecology (but see Fillol *et al*., 2016; Jeanbille *et al*., 2016; Liu *et al*., 2018), despite being more statistically robust than traditional approaches using arbitrary cut-offs of local and regional abundances (e.g. Mo *et al*., 2018). 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-multiplicatives replacement method available in the 'zCompositions' package (v. 1.2.0; Martín-Fernández *et al*., 2015). An isometric log-ratio transformation (Egozcue *et al*., 2003), as implemented in the 'philr' package (default parameters, v. 1.10.0; Silverman *et al*., 2017), 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 inner nodes of the phylogeny. Balances were computed for each node and depict relative changes in ASV abundance between two neighbouring clades relative to each other, while ignoring abundances in the remainder of the tree. This approach accounts for the compositional nature of amplicon datasets (Gloor *et al*., 2017) 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) (Morton *et al*., 2017; Silverman *et al*., 2017).

Statistical analyses

Alpha-diversity indices and taxonomic composition were calculated using the full rarefied table. The alpha diversity metric Faith's phylogenetic diversity (PD), which measures the shared phylogenetic history among taxa occurring in a sample (Faith, 1992), 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; Kembel *et al*., 2010). Pielou's evenness (Pielou, 1966) was computed 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; Genuer *et al*., 2015) to identify the best predictors for PD and evenness of archaeal groups and the absolute abundance of AOA (response variables). Random forests represent an ensemble machine learning algorithm that is well suited to model non-linear relationships between predictors and response variables and can deal with non-normality and high collinearity among predictors (Breiman, 2001). 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; Liaw and Wiener, 2002), 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 (Table S3). The relationship between each environmental variable and PD, evenness and the absolute abundance of AOA was visualized using accumulated local effects plots (grid.size = 30) implemented in the 'iml' package (v. 0.9.0; Molnar *et al*., 2018). 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 among explanatory variables (Apley and Zhu, 2020).

All statistical analyses on β-diversity were conducted on the philr-transformed (Silverman *et al*., 2017) 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*. (2018). 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 = 9999), respectively (Table S1). 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 VPA (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. The VPA was also applied to the nitrososphaerial communities and a subset of the full phlir-transformed ASV table was generated by extracting the nodes corresponding to Nitrososphaeria. The abundance of the archaeal *amoA* gene was fitted onto the nitrososphaerial ordination using the envfit function in 'vegan' (permutations = 9999). Since the correlation was significant ($p < 0.001$), the abundance of *amoA* genes was included in the VPA as a biotic component (functional predictor), along edaphic, climatic and spatial factors.

Since Nitrososphaeria represent putative AOA (Alves *et al*., 2018), 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). Multivariate regression trees (MRT) were computed using the 'mvpart' package (v. 1.6.2; De'ath, 2002). to predict the relationships between the set of variables (Table 1) and the community structure of Nitrososphaeria. This approach has the advantage to allow for the examination of the sequential effect of several environmental filters (contrary to Gubry-Rangin *et al*., 2011, 2015; Oton *et al*., 2016; Alves *et al*., 2018 who assessed effects of each variable individually) and delineate clusters of samples in which the variation in environmental conditions is minimized (Table S2). The selected tree represented the most parsimonious solution within one standard error above the minimal cross-validated relative error (Fig. S8; $n = 10,000$ trees), following (Breiman *et al*., 1984), 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 cluster were plotted on the nitrososphaerial phylogeny with ggtree (v. 1.16.0; Yu *et al*., 2017), with blue and red branches indicating an increase and a decrease, respectively. Sequences of Group 1.1c thaumarchaeota were used to root the phylogeny.

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.

Author contribution

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.

References

Aigle, A., Gubry-Rangin, C., Thion, C., Estera-Molina, K.Y., Richmond, H., Pett-Ridge, J., *et al*. (2020) Experimental testing of hypotheses for temperature- and pH-based niche specialisation of ammonia oxidising archaea and bacteria. *Environ Microbiol* **22**: 4032–4045.

Alves, R.J.E., Minh, B.Q., Urich, T., Von Haeseler, A., and Schleper, C. (2018) Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. *Nat Commun* **9**: 1517.

Alves, R.J.E., Wanek, W., Zappe, A., Richter, A., Svenning, M.M., Schleper, C., and Urich, T. (2013) Nitrification rates in Arctic soils are associated with functionally distinct populations of ammonia-oxidizing archaea. *ISME J* **7**: 1620–1631.

Apley, D.W., and Zhu, J. (2020) Visualizing the effects of predictor variables in black box supervised learning models. *J R Stat Soc Ser B Stat Methodol* **82**: 1059–1086.

Auguet, J.C., Barberan, A., and Casamayor, E.O. (2010) Global ecological patterns in uncultured Archaea. *ISME J* **4**: 182–190.

Baker, B.J., De Anda, V., Seitz, K.W., Dombrowski, N., Santoro, A.E., and Lloyd, K.G. (2020) Diversity, ecology and evolution of Archaea. *Nat Microbiol* **5**: 887–900.

Baker, B.J., and Banfield, J.F. (2003) Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* **44**: 139–152.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., and Fierer, N. (2011) Examining the global distribution of dominant archaeal populations in soil. *ISME J* **5**: 908–917.

Bayer, B., Vojvoda, J., Offre, P., Alves, R.J.E., Elisabeth, N. H., Garcia, J.A.L., *et al*. (2016) Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J* **10**: 1051–1063.

Breiman, L. (2001) Random forests. *Mach Learn* **45**: 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984) *Classification and Regression Trees*. Belmont, California, U.S.A.: Wadsworth Publishing Company.

Bru, D., Ramette, A., Saby, N.P.A., Dequiedt, S., Ranjard, L., Jolivet, C., *et al*. (2011) Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J* **5**: 532–542.

Callahan, B.J., McMurdie, P.J., and Holmes, S.P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J* **11**: 2639–2643.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016) DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**: 581–583.

Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., *et al*. (2015) Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr Biol* **25**: 690–701.

Chamberlain, S. (2019) rnoaa: "NOAA" weather data from R.

De'ath, G. (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* **83**: 1105–1117.

Dini-Andreote, F., Stegen, J.C., Van Elsas, J.D., and Salles, J.F. (2015) Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci USA* **112**: E1326–E1332.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-Vidal, C. (2003) Isometric logratio transformations for compositional data analysis. *Math Geol* **35**: 279–300.

Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.

FAL, FAW, and RAC. (1996) Referenzmethoden der Eidg. landwirtschaftlichen Forschungsanstalten. 1. Bodenuntersuchung zur Düngeberatung. *Zürich-Reckenholz*.

Falkowski, P.G., Fenchel, T., and Delong, E.F. (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.

Fillol, M., Auguet, J.C., Casamayor, E.O., and Borrego, C.M. (2016) Insights in the ecology and evolutionary history of the miscellaneous Crenarchaeotic group lineage. *ISME J* **10**: 665–677.

Garland, G., Edlinger, A., Banerjee, S., Degrune, F., García-Palacios, P., Pescador, D.S., *et al*. (2021) Crop cover is more important than rotational diversity for soil multi-functionality and cereal yields in European cropping systems. *Nat Food* **2**: 28–37.

Genuer, R., Poggi, J.M., and Tuleau-Malot, C. (2015) VSURF: an R package for variable selection using random forests. *R J* **7**: 19–33.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., *et al*. (2019) Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nat Commun* **10**: 1107.

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., and Egozcue, J.J. (2017) Microbiome datasets are compositional: and this is not optional. *Front Microbiol* **8**: 2224.

Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P., *et al*. (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* **108**: 21206–21211.

Gubry-Rangin, C., Kratsch, C., Williams, T.A., McHardy, A. C., Embley, T.M., Prosser, J.I., and Macqueen, D.J. (2015) Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proc Natl Acad Sci USA* **112**: 9370–9375.

Gubry-Rangin, C., Novotnik, B., Mandič-Mulec, I., Nicol, G. W., and Prosser, J.I. (2017) Temperature responses of soil ammonia-oxidising archaea depend on pH. *Soil Biol Biochem* **106**: 61–68.

Gubry-Rangin, C., Williams, W., and Prosser, J.I. (2018) Approaches to understanding the ecology and evolution of understudied terrestrial archaeal ammonia-oxidisers. *Emerg Top Life Sci* **2**: 619–628.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497–506.

Herlemann, D.P.R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J.J., and Andersson, A.F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J* **5**: 1571–1579.

Hink, L., Gubry-Rangin, C., Nicol, G.W., and Prosser, J.I. (2018) The consequences of niche and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. *ISME J* **12**: 1084–1093.

Hink, L., Lycus, P., Gubry-Rangin, C., Frostegård, Å., Nicol, G.W., Prosser, J.I., and Bakken, L.R. (2017) 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* **19**: 4882–4896.

Hollister, J. and Shah, T. (2017) elevatr: access elevation data from various APIs.

Hu, H.W., Zhang, L.M., Dai, Y., Di, H.J., and He, J.Z. (2013) pH-dependent distribution of soil ammonia oxidizers across a large geographical scale as revealed by high-throughput pyrosequencing. *J Soil Sediment* **13**: 1439–1449.

Hubbell, S. (2001) *The Unified Neutral Theory of Biodiversity and Biogeography*. Priceton, NJ, USA: Princeton University Press.

Hunt, D., David, L., Gevers, D., Preheim, S., Alm, E., and Polz, M. (2008) Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science* **320**: 1081–1085.

Isoda, R., Hara, S., Tahvanainen, T., and Hashidoko, Y. (2017) Comparison of archaeal communities in mineral soils at a boreal forest in Finland and a cold-temperate forest in Japan. *Microbes Environ* **32**: 390–393.

Jeanbille, M., Gury, J., Duran, R., Tronczynski, J., Agogué, H., Saïd, O.B., *et al*. (2016) Response of core microbial consortia to chronic hydrocarbon contaminations in coastal sediment habitats. *Front Microbiol* **7**: 1637.

Jiang, H., Huang, L., Deng, Y., Wang, S., Zhou, Y., Liu, L., and Dong, H. (2014) Latitudinal distribution of ammonia-oxidizing bacteria and archaea in the agricultural soils of eastern China. *Appl Environ Microbiol* **80**: 5593–5602.

Jiao, S., Xu, Y., Zhang, J., and Lu, Y. (2019a) Environmental filtering drives distinct continental atlases of soil archaea between dryland and wetland agricultural ecosystems. *Microbiome* **7**: 15.

Jiao, S., Yang, Y., Xu, Y., Zhang, J., and Lu, Y. (2019b) Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J* **14**: 202–216.

Jung, M.-Y., Sedlacek, C.J., Kits, K.D., Mueller, A.J., Rhee, S.-K., Hink, L., *et al*. (2021) Ammonia-oxidizing archaea possess a wide range of cellular ammonia affinities. *ISME J*. https://doi.org/10.1038/s41396-021-01064-z

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., *et al*. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

Kemnitz, D., Kolb, S., and Conrad, R. (2007) High abundance of Crenarchaeota in a temperate acidic forest soil. *FEMS Microbiol Ecol* **60**: 442–448.

Kim, J.G., Park, S.J., Sinninghe Damsté, J.S., Schouten, S., Rijpstra, W.I.C., Jung, M.Y., *et al*. (2016) Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci USA* **113**: 7888–7893.

Krebs, C.J. (1999) *Ecological Methodology*, 2nd ed. New York, NY, USA: Addison-Wesley Educational Publishers.

Kuypers, M.M.M., Marchant, H.K., and Kartal, B. (2018) The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**: 263–276.

Larkin, A.A., and Martiny, A.C. (2017) Microdiversity shapes the traits, niche space, and biogeography of microbial taxa. *Environ Microbiol Rep* **9**: 55–70.

Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J., and Garnier, J. (2014) 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environ Res Lett* **9**: 105011.

Lehtovirta-Morley, L.E., Ross, J., Hink, L., Weber, E.B., Gubry-Rangin, C., Thion, C., *et al*. (2016) Isolation of "*Candidatus* Nitrosocosmicus franklandus", a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol Ecol* **92**: fiw057.

Lehtovirta, L.E., Prosser, J.I., and Nicol, G.W. (2009) Soil pH regulates the abundance and diversity of group 1.1c Crenarchaeota. *FEMS Microbiol Ecol* **70**: 367–376.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., *et al*. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806–809.

Levičnik-Höfferle, Š., Nicol, G.W., Ausec, L., Mandić-Mulec, I., and Prosser, J.I. (2012) Stimulation of thaumarchaeal ammonia oxidation by ammonia derived from organic nitrogen but not added inorganic nitrogen. *FEMS Microbiol Ecol* **80**: 114–123.

Liaw, A., and Wiener, M. (2002) Classification and regression by randomForest. *R News* **2**: 18–22.

Lin, X., Handley, K.M., Gilbert, J.A., and Kostka, J.E. (2015) Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *ISME J* **9**: 2740–2744.

Liu, J., Yu, Z., Yao, Q., Sui, Y., Shi, Y., Chu, H., and Tang, C. (2019) Biogeographic distribution patterns of the archaeal communities across the black soil zone of Northeast China. *Front Microbiol* **10**: 23.

Liu, W., Graham, E.B., Dong, Y., Zhong, L., Zhang, J., Qiu, C., *et al*. (2021) Balanced stochastic versus deterministic assembly processes benefit diverse yet uneven ecosystem functions in representative agroecosystems. *Environ Microbiol* **23**: 391–404.

Liu, X., Li, M., Castelle, C.J., Probst, A.J., Zhou, Z., Pan, J., *et al*. (2018) Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. *Microbiome* **6**: 102.

Lu, X., Bottomley, P.J., and Myrold, D.D. (2015) Contributions of ammonia-oxidizing archaea and bacteria to nitrification in Oregon forest soils. *Soil Biol Biochem* **85**: 54–62.

Macqueen, D.J., and Gubry-Rangin, C. (2016) Molecular adaptation of ammonia monooxygenase during independent pH specialization in Thaumarchaeota. *Mol Ecol* **25**: 1986–1999.

Magurran, A.E. (2007) Species abundance distributions over time. *Ecol Lett* **10**: 347–354.

Martín-Fernández, J.A., Hron, K., Templ, M., Filzmoser, P., and Palarea-Albaladejo, J. (2015) Bayesian-multiplicative treatment of count zeros in compositional data sets. *Stat Modelling* **15**: 134–158.

Massello, F.L., Chan, C.S., Chan, K.G., Goh, K.M., Donati, E., and Urbieta, M.S. (2020) Meta-analysis of microbial communities in hot springs: recurrent taxa and complex shaping factors beyond ph and temperature. *Microorganisms* **8**: 906.

Mo, Y., Zhang, W., Yang, J., Lin, Y., Yu, Z., and Lin, S. (2018) Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME J* **12**: 2198–2210.

Molnar, C., Bischl, B., and Casalicchio, G. (2018) iml: an R package for interpretable machine learning. *J Open Source Softw* **3**: 786.

Morton, J.T., Sanders, J., Quinn, R.A., Mcdonald, D., Gonzalez, A., Vázquez-Baeza, Y., *et al*. (2017) Balance trees reveal microbial niche differentiation. *mSystems* **2**: e00162–e00116.

Nicol, G.W., Leininger, S., Schleper, C., and Prosser, J.I. (2008) The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environ Microbiol* **10**: 2966–2978.

Offre, P., Spang, A., and Schleper, C. (2013) Archaea in biogeochemical cycles. *Annu Rev Microbiol* **67**: 437–457.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., *et al*. (2018) vegan: community ecology package.

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

Pielou, E.C. (1966) The measurement of diversity in different types of biological collections. *J Theor Biol* **13**: 131–144.

Placella, S.A., and Firestone, M.K. (2013) Transcriptional response of nitrifying communities to wetting of dry soil. *Appl Environ Microbiol* **79**: 3294–3302.

Power, J.F., Carere, C.R., Lee, C.K., Wakerley, G.L.J., Evans, D.W., Button, M., *et al*. (2018) Microbial biogeography of 925 geothermal springs in New Zealand. *Nat Commun* **9**: 2876.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Prosser, J.I., and Nicol, G.W. (2012) Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol* **20**: 523–531.

Prosser, J.I., and Nicol, G.W. (2008) Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* **10**: 2931–2941.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Qin, W., Amin, S.A., Martens-Habbena, W., Walker, C.B., Urakawa, H., Devol, A.H., *et al*. (2014) Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci USA* **111**: 12504–12509.

R Core Team. (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schauss, K., Focks, A., Leininger, S., Kotzerke, A., Heuer, H., Thiele-Bruhn, S., *et al*. (2009) Dynamics and functional relevance of ammonia-oxidizing archaea in two agricultural soils. *Environ Microbiol* **11**: 446–456.

Silverman, J.D., Washburne, A.D., Mukherjee, S., and David, L.A. (2017) A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* **6**: e21887.

Söllinger, A., Schwab, C., Weinmaier, T., Loy, A., Tveit, A.T., Schleper, C., and Urich, T. (2016) Phylogenetic and genomic analysis of Methanomassiliicoccales in wetlands and animal intestinal tracts reveals clade-specific habitat. *FEMS Microbiol Ecol* **92**: fiv149.

Sun, D.L., Jiang, X., Wu, Q.L., and Zhou, N.Y. (2013) Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl Environ Microbiol* **79**: 5962–5969.

Takai, K., and Horikoshi, K. (2000) Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. *Appl Environ Microbiol* **66**: 5066–5072.

Thion, C., and Prosser, J.I. (2014) Differential response of nonadapted ammonia-oxidising archaea and bacteria to drying-rewetting stress. *FEMS Microbiol Ecol* **90**: 380–389.

Tourna, M., Freitag, T.E., Nicol, G.W., and Prosser, J.I. (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* **10**: 1357–1364.

Tripathi, B.M., Kim, M., Lai-Hoe, A., Shukor, N.A.A., Rahim, R.A., Go, R., and Adams, J.M. (2013) pH dominates variation in tropical soil archaeal diversity and community structure. *FEMS Microbiol Ecol* **86**: 303–311.

Tripathi, B.M., Kim, M., Tateno, R., Kim, W., Wang, J., Lai-Hoe, A., *et al*. (2015) Soil pH and biome are both key determinants of soil archaeal community structure. *Soil Biol Biochem* **88**: 1–8.

Vajrala, N., Martens-Habbena, W., Sayavedra-Soto, L.A., Schauer, A., Bottomley, P.J., Stahl, D.A., and Arp, D.J. (2013) Hydroxylamine as an intermediate in ammonia oxidation by globally abundant marine archaea. *Proc Natl Acad Sci USA* **110**: 1006–1011.

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

Weber, E.B., Lehtovirta-Morley, L.E., Prosser, J.I., and Gubry-Rangin, C. (2015) Ammonia oxidation is not required for growth of group 1.1c soil Thaumarchaeota. *FEMS Microbiol Ecol* **91**: fiv001.

Wessén, E., Söderström, M., Stenberg, M., Bru, D., Hellman, M., Welsh, A., *et al*. (2011) Spatial distribution of ammonia-oxidizing bacteria and archaea across a 44-hectare farm related to ecosystem functioning. *ISME J* **5**: 1213–1225.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017) ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Results of Mantel tests and PERMANOVAs between the archaeal communities and the set of environmental variables (***$p < 0.001$, **$p < 0.01$; *$p < 0.05$).

Table S2. Range of environmental variables in the MRT clusters presented in Fig. 4 (see Table 1 for units).

Table S3. Model parameters and fit (residual mean squared error—RMSE) of the random forest analyses.

Fig. S1. Relationship between environmental variables and the phylogenetic diversity of archaeal communities, based on random forest analyses. Predictor variables selected by VSURF were used to generate accumulated local effects plots, which show the relationship between the predictor (*x*-axis) and response variables in the model (*y*-axis) while accounting for potential correlations among predictor values. N.S. Non-significant variable. See Table 1 for units. Model parameters and fit are indicated in Table S3.

Fig. S2. Relationship between environmental variables and the evenness of archaeal communities, based on random forest analyses. Predictor variables selected by VSURF were used to generate accumulated local effects plots, which show the relationship between the predictor (*x*-axis) and response variables in the model (*y*-axis) while accounting for potential correlations among predictor values. N.S. Non-significant variable. See Table 1 for units. Model parameters and fit are indicated in Table S3.

Fig. S3. Relationship between environmental variables and the abundance of AOA (number of gene copies g^{-1} dw soil), based on random forest analyses. (A) Abundances of AOA in the 151 samples. (B) Predictor variables selected by VSURF were used to generate accumulated local effects plots, which show the relationship between the predictor (*x*-axis) and response variables in the model (*y*-axis) while accounting for potential correlations among predictor values. See Table 1 for units. Model parameters and fit are indicated in Table S3.

Fig. S4. Factors driving the variation in nitrososphaerial community composition and structure across the European gradient according to variation partitioning analysis (VPA). The biotic component corresponds to the abundance of the archaeal *amoA* gene, which correlated to the nitrososphaerial community structure ($p < 0.001$). All fractions were significant ($p < 0.001$) and the variation explained is indicated (%). Values < 0 are not shown. The VPA was performed on the philr-transformed data corresponding to the Nitrososphaeria.

Fig. S5. Sampling sites across Europe.

Fig. S6. Rarefaction curves of archaeal communities.

Fig. S7. Occurence of each ASV plotted against its dispersion index. The dotted line represents the 2.5% confidence limit of the χ^2 distribution, with frequent ASVs depicted in orange and rare ASVs in blue.

Fig. S8. Relationship between the size of trees and the relative error (green) and the cross-validated relative error (blue) computed in the multivariate regression tree analysis. The best predictive tree is indicated by a red dot. The orange dot corresponds to the smallest tree one standard error above the best tree. The green vertical bars indicate the number of times that a given solution was selected as the best one during the cross-validation iterations.