universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Exploring the diversity of ammonia oxidizing archaea in pristine arctic and agricultural soils"

verfasst von / submitted by

Grda. Marina Monserrat Díez

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 830
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molecular Microbiology, Microbial Ecology and Immunobiology.
Betreut von / Supervisor:	Univ.-Prof. Dipl.-Biol. Dr. Christa Schleper
Mitbetreut von / Co-Supervisor:	Melina Kerou, PhD

ACKNOWLEDGMENTS

In the first place, I want to thank Melina for her serene and efficient supervision. Thank you for inspiring me confidence in myself and my work and motivation to continue pursuing the scientific pathway.

I want to thank Logan for the efforts made collecting the farm soil samples from Montana, as well as gathering the metadata from both ranches. It was fun to get to work with them!

I want to thank all the people who took part in CryoCarb project, particularly to Ricardo J. Eloy Alves, for collecting the samples and the metadata that we have analysed here. It has been challenging and exciting to get to work with these precious samples!

Also, I want to thank Rafael for all the help with the phylogenies and for his rapid answers to my bioinformatic related questions.

I also want to thank Erika, Max, Thomas and Ülku and the rest of the AOA people for the support in the lab and for cheering me during my eternal fights with PCRs.

I want to thank my Master buddies in the department Karin, Nicole and Kat for the shared time and their always kind words.

Also I want to thank all the people in the Archaeal Biology and Ecogenomics Unit, for creating a nice working environment and offering their help when I needed it.

I would like to express my deep gratitude to my family for their essential support throughout my studies. Also, I want to thank my little viennese family for their emotional support during this time.

Finally, I want to thank Christa for her contagious passion and enthusiasm for AOA and science. Thank you for introducing me into this field and for giving me the opportunity to work in this project. It has been a very inspiring journey!

ABSTRACT

Nitrogen is an essential element for all forms of life but its availability is limited in most environments. As a result of human activities, large amounts of reactive N-species are introduced into the system, disrupting the N-cycle and causing multiple issues, such as the emission of the strong greenhouse gas N2O. Thus, this unbalance has direct consequences for the actual global climatic crisis. Ammonia-oxidizing archaea (AOA) play a crucial role in the microbial N-transformations by performing the first step of the nitrification process and thus, their activity impacts directly N-cycling processes. The ecology of AOA has been often studied in relation to their nitrifying bacterial counterparts (ammonia oxidizing bacteria, AOB), and few studies focus on the diversity and environmental distribution within AOA, although it represents the most widespread archaeal group in the environment. A recent study generated a highly resolved phylogeny for AOA based on the alpha subunit of ammonia monooxygenase (amoA), which revealed distinct environmental patterns among the different AOA clades. Indeed, AOA diversity studies based on the amoA gene can provide information about the ecophysiologies and potential functions in the ecosystems, particularly valuable for this group of organisms since their cultivation is highly challenging.

Here we characterized the AOA communities from farm soils and arctic tundra soils, through 16S rRNA gene amplicon sequencing analysis. Subsequently, the variant sequences were assigned into AOA clades, using a recently developed method that links the information of the 16S rRNA gene sequences with their corresponding amoA gene sequences. In the case of the farm soils, we also quantified the nitrifier community by qPCR. Finally, we identified the main environmental drivers behind the AOA clade distribution.

Our results indicate distinct patterns of distribution of the different members of AOA in both studied ecosystems. In the farm soils, we saw that the nitrifier community is dominated by AOA, which in some samples reaches 19% of the total prokaryotic community. The main clades detected in the farm soils belonged to the NS-Zeta and NS-Alpha clades. Our observations suggest that NS-Alpha seems to prefer sites less disturbed by agricultural activities, that tend to be more oligotrophic, while NS-Zeta tends to dominate more in sites agriculturally exploited. Furthermore, pH and the soil water content appeared to have a significant effect on the AOA community structure. In the arctic tundra soils, the AOA community has a low diversity along the circumpolar area and it is dominated by two phylotypes, belonging to the NS-Gamma and the NS-Zeta clade. Correlation analysis to the environmental data revealed the oligotrophic character of the overall AOA community, with NS-Gamma appearing to have a stronger preference for nutrient poor environments than

NS-Zeta. Additionally, we validated alternative methods for the detection and classification of the amoA clades, by using the broadly used 16S rRNAmarker gene. This method broadens the options to obtain information about the ecophysiologies of the different clades in the AOA diversity studies to better understand their ecology and distribution.

ZUSAMMENFASSUNG

Stickstoff ist ein wesentliches Element für alle Lebensformen, aber seine Verfügbarkeit ist in den meisten Umgebungen begrenzt. Infolge menschlicher Aktivitäten werden große Mengen reaktiver Stickstoffarten in das System eingebracht, wodurch der Stickstoffkreislauf gestört wird und zahlreiche Probleme entstehen, wie die Emission des starken Treibhausgases N2O. Dieses Ungleichgewicht hat somit direkte Auswirkungen auf die aktuelle globale Klimakrise. Ammoniak-oxidierende Archaeen (AOA) spielen eine entscheidende Rolle bei der mikrobiellen N-Transformation, da sie den ersten Schritt des Nitrifikationsprozesses durchführen und ihre Aktivität somit direkte Auswirkungen auf die N-Kreislaufprozesse hat. Die Ökologie der AOA wurde häufig im Zusammenhang mit ihren nitrifizierenden bakteriellen Gegenstücken (Ammoniak oxidierende Bakterien, AOB) untersucht, und nur wenige Studien konzentrieren sich auf die Vielfalt und die Verteilung der AOA in der Umwelt, obwohl sie die am weitesten verbreitete archäische Gruppe in der Umwelt darstellen. In einer kürzlich durchgeführten Studie wurde eine hochaufgelöste Phylogenie für AOA auf der Grundlage der Alpha-Untereinheit der Ammoniak-Monooxygenase (amoA) erstellt, die unterschiedliche Umweltmuster unter den verschiedenen AOA-Kladen erkennen ließ. Untersuchungen der AOA-Diversität auf der Grundlage des amoA-Gens können Informationen über die Ökophysiologie und die potenziellen Funktionen in den Ökosystemen liefern, was für diese Gruppe von Organismen besonders wertvoll ist, da ihre Kultivierung sehr schwierig ist.

Hier haben wir die AOA-Gemeinschaften aus landwirtschaftlich genutzten Böden und Böden der arktischen Tundra durch 16S rRNA-Gen-Amplikon-Sequenzanalyse charakterisiert. Anschließend wurden die Sequenzvarianten mit Hilfe einer kürzlich entwickelten Methode, die die Informationen der 16S rRNA-Gensequenzen mit den entsprechenden amoA-Gensequenzen verknüpft, in AOA-Kladen eingeteilt. Im Falle der landwirtschaftlichen Böden haben wir auch die Nitrifikantengemeinschaft mittels qPCR quantifiziert. Schließlich identifizierten wir die wichtigsten umweltbedingten Faktoren für die Verteilung der AOA-Klone.

Unsere Ergebnisse weisen auf unterschiedliche Verteilungsmuster der verschiedenen AOA-Mitglieder in den beiden untersuchten Ökosystemen hin. In den landwirtschaftlich genutzten Böden wurde festgestellt, dass die Nitrifikantengemeinschaft von AOA dominiert wird, die in einigen Proben 19 % der gesamten prokaryotischen Gemeinschaft ausmachen. Die wichtigsten Kladen, die in den landwirtschaftlichen Böden nachgewiesen wurden, gehörten zu den NS-Zeta- und NS-Alpha-Kladen. Unsere Beobachtungen deuten darauf hin, dass NS-Alpha anscheinend Standorte bevorzugt, die weniger durch landwirtschaftliche Aktivitäten gestört werden und eher oligotroph sind, während NS-Zeta eher an landwirtschaftlich genutzten Standorten vorherrscht. Darüber hinaus hatten der pH-Wert und der Wassergehalt des Bodens offenbar einen erheblichen Einfluss auf die Struktur der AOA-Gemeinschaft. In den Böden der arktischen Tundra weist die AOA-Gemeinschaft in der zirkumpolaren Region eine geringe Diversität auf und wird von zwei Phylotypen dominiert, die zur NS-Gamma- und zur NS-Zeta-Klade gehören. Die Korrelationsanalyse zu den Umweltdaten zeigte den oligotrophen Charakter der gesamten AOA-Gemeinschaft, wobei NS-Gamma eine stärkere Präferenz für nährstoffarme Umgebungen zu haben scheint als NS-Zeta. Darüber hinaus haben wir alternative Methoden für den Nachweis und die Klassifizierung der amoA-Kladen validiert, indem wir das weit verbreitete 16S rRNAmarker-Gen verwendeten. Diese Methode erweitert die Möglichkeiten, Informationen über die Ökophysiologie der verschiedenen Kladen in den AOA-Diversitätsstudien zu erhalten, um ihre Ökologie und Verbreitung besser zu verstehen.

Contents

1. INTRODUCTION

1.1. The nitrogen cycle and the risks for the planet of having an imbalanced N cycle

Nitrogen is an element essential for life. However, its bioavailability is limited and conditioned to the activity of microorganisms. The largest pool of N in the planet is in the form of molecular N_2 and it is unassimilable by most living beings, with the exception of a small group of microorganisms: the diazotrophs or nitrogen fixators. These organisms (e.g. cyanobacteria or rhizobia) transform N_2 to NH_3 through a highly energetically costly process, which is the first step to supply ecosystems with reactive nitrogen species (Nr, usually in the forms of NH_3 and NO_x). Thus, N is a limiting factor in ecosystems and it plays a key role in determining their habitability. As a consequence of this N-limitation, terrestrial ecosystems harbor a high biodiversity of organisms that are adapted to different levels and forms of N availability (Dise et al., 2011).

Humans overcame this bottleneck with the invention of the Haber-Bosch reaction, an artificial process for the fixation of nitrogen. This represented a true revolution because it allowed the production of ammonia-based fertilizers at an industrial level. Since the appearance of chemical fertilizers, the food productivity exponentially increased, and consequently, so did the human population (Figure 1). Thanks to the use of fertilizers nowadays, we are feeding 50% of Earth's population, according to estimations (Erisman et al., 2015). The abundance of reactive nitrogen species on the planet has been doubled due to human activity. Therefore, the release of massive amounts of an element that naturally is a limited resource, is highly disruptive for the ecosystems' equilibrium and biodiversity.

Although we currently strongly rely on them, the prolonged and intensive use of fertilizers is giving rise to pressing environmental, social and economical issues. The initial cause of the problem is that from the N pool present in soil, plants end up taking little amounts from it. This is due to the existing competition between plants and the natural soil microbiome for the scarce N resources in soil (Coskun et al., 2017), with the additional difficulty of the low efficiency for the assimilation of the nitrogen in fertilizers by the plants. To compensate for the low N uptake, farmers apply excess amounts of fertilizer into the fields. As a result, there is a large input of ammonia in the soil, which gets rapidly oxidized by nitrifying microorganisms into nitrate. The retention of anions by soils is much lower than cations;

thus, as nitrate is a negatively charged molecule, it is easily leached out of the soil to groundwater and marine systems. When it does, it causes acidification and eutrophication of water systems, and can finally get converted and released into the atmosphere as N_2O, a gas with a 300 times stronger greenhouse effect than CO_2. The consequences of this phenomenon are known as the nitrogen cascade (Erisman et al., 2015; Galloway et al., 2003), because during its journey, the same N atom can have multiple effects in the atmosphere, in terrestrial ecosystems and in marine and fresh water systems.

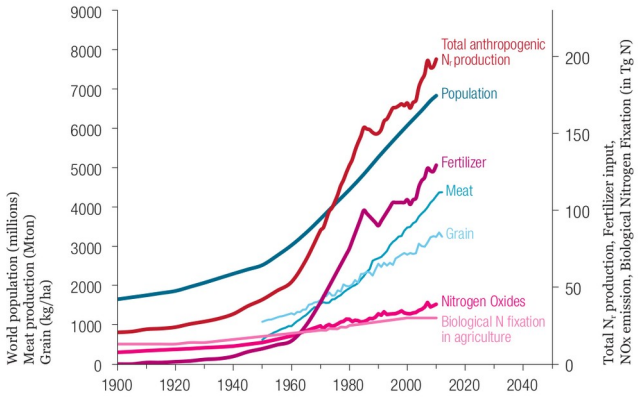


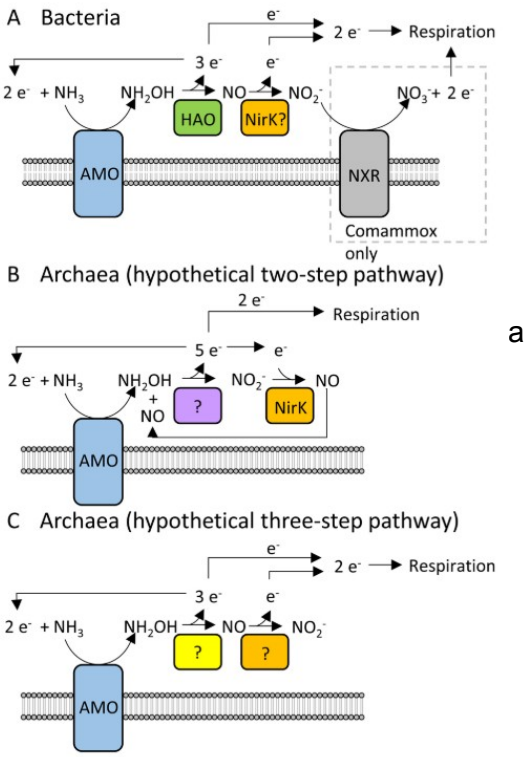
Figure 1: Global trends in human population, release of N reactive species, use of fertilizers, meat and grain production between the years 1900 and 2012. From Erisman *et. al.,* 2015.

1.2. Ammonia oxidizing microorganisms and their role in the N cycle

Nitrification is the aerobic process of converting ammonia to nitrate, performed by nitrifying organisms. It is a reaction that happens in two stages: the oxidation of ammonia to nitrite, performed by ammonia oxidizing microorganisms (AOM), and the further oxidation of nitrite to nitrate, performed by nitrite oxidizing bacteria (NOB). The former group is comprised by ammonia oxidizing bacteria (AOB), ammonia oxidizing archaea (AOA) and the complete ammonia oxidizing bacteria (comammox). AOB, already described by Winogradsky in 1890 (Winogradsky, Sergei, 1890), were considered the sole ammonia oxidizing microorganisms, until the discovery of archaea capable of ammonia oxidation through metagenomics in 2005 (Treusch et al., 2005). Since then, and thanks to the expansion of metagenomics and cultivation techniques, AOA emerged as a main group within the nitrifier community in both terrestrial and marine environments, and so far constitute the only lineage of aerobic mesophilic archaea (M.-Y. Jung et al., 2016; Kim et al., 2012; Schleper & Nicol, 2010; Stieglmeier et al., 2014). The most recently discovered ammonia oxidizing group was

comammox (Daims et al., 2015; Pinto et al., 2016; van Kessel et al., 2015), able to perform the complete oxidation of ammonia to nitrate.

The oxidation of ammonia is a process that happens in several steps, which some remain still unresolved (Figure 2). The first step, common to AOA, AOB and comammox, is the oxidation of ammonia to hydroxylamine, catalyzed by the ammonia monooxygenase (AMO), an enzyme belonging to the CuMMO superfamily with distantly related homologs in AOA and AOB (Alves *et al*. 2018). Subsequently, in the case of AOB, the process continues through the oxidation to NO, by the hydroxylamine dehydrogenase enzyme (HAO) (Caranto & Lancaster, 2017), followed by the oxidation to nitrite, possibly catalyzed by nitrosocyanin (NcyA) or another uncharacterized enzyme. However, in AOA the oxidation of hydroxylamine to nitrite is still uncharacterized and remains under debate whether it happens in one oxidation step directly to nitrite (Kozlowski et al., 2016), or in two steps with NO as an intermediate product (Carini et al., 2018).



a

Figure 2: Pathways of ammonia oxidation in bacteria and archaea. (A) AOB and comammox Nitrospira (based on Caranto and Lancaster 2017). (B) Hypothetical two-step model in AOA (Kozlowski et al. (2016). (C) Hypothetical three-step model in AOA (Lehtovirta *et. al.,* 2018.). Figure from Lehtovirta *et. al.,* 2018.

One of the pressing incentives for the study of nitrifiers in the context of the actual environmental crisis is the production of the strong pollutant gas N_2O as a product of their activity. In AOB, the reaction happens enzymatically in low oxygen conditions, through a

process called nitrifier denitrification where nitrite is reduced to NO and subsequently to N_2O (Goreau et al., 1980; Kozlowski et al., 2014). Conversely, AOA and comammox cannot perform nitrifier denitrification because they do not possess the genomic repertoire for NO reduction (Daims et al., 2015; Spang et al., 2012; Tourna et al., 2011). Still, in these organisms N_2O can be produced through the abiotic reaction of intermediate molecules of the ammonia oxidation process (i.e. hydroxylamine, NO) (Kozlowski et al., 2016), becoming a relevant source of N_2O in those environments where they are abundant and active (Siljanen et al., 2019). Additionally, nitrifier activity can indirectly fuel N_2O production by enhancing denitrifier activity, which often do not complete the total reduction of NO_2^- or NO_3^- to N_2 and instead produce N_2O.

With the identification of all the different groups performing ammonia oxidation in the environment, efforts have focused in exploring their environmental distribution and the nature of their competition for the same N resource. Both AOA and AOB are ubiquitous and highly abundant in the environment. In most soil ecosystems, AOA have been found to outnumber AOB, in some cases by some orders of magnitude, and they can represent from 1% to 5% of the total prokaryotic community. Different responses of AOA and AOB to environmental parameters as well as different distribution patterns were reported by multiple studies (reviewed by (Lehtovirta-Morley, 2018; Prosser & Nicol, 2012), resulting in the identification of certain environmental drivers that influence their relative abundance and provide insights to their respective ecophysiology.

The ammonia, as the first source of energy for both groups in aerobic environments, is thought to be one of the main environmental factors influencing the differentiation between AOA and AOB, reflected in different affinities for ammonia and utilization of different ammonia sources. As a general trend, terrestrial AOA isolates (*Nitrososphaeraceae*) possess similar ammonia affinities as AOB (M. Y. Jung et al., 2022), although other AOA lineages (*Nitrosopumilaceae*) seem to possess higher apparent and specific substrate affinities than AOB. Additionally, microcosm experiments showed higher growth and activity of AOB than AOA in heavily fertilized soils with inorganic ammonium (Di et al., 2009; Jia & Conrad, 2009) while upon addition of organic sources of ammonia, growth and nitrification activity is dominated by AOA over AOB (Hink et al., 2017; Levičnik-Höfferle et al., 2012; Prosser et al., 2020).

Another relevant environmental driving force in AOA and AOB is pH. In culture, several acidophilic AOA strains have been isolated, which led to the discovery of an acidophilic

lineage within AOA, the Nitrosotaleales (Lehtovirta-Morley et al., 2011, 2014), whereas no acidophilic AOB has been found yet. Also in the environment, AOA outnumber AOB in acidic ecosystems (Gubry-Rangin et al., 2010; Yao et al., 2011; L.-M. Zhang et al., 2012), in some cases becoming the only nitrifiers in the community (Siljanen et al., 2019). pH has also an influence on substrate availability, because under acidic pH values the natural balance between NH_3/NH_4^+ shifts towards NH_4^+, lowering the ammonia concentration which is supposed to be the main N source by AOA and AOB (M. Y. Jung et al., 2022). Thus, it has been proposed that AOA are better adapted to habitats with lower pH thanks to their higher affinity to ammonia, among other specific adaptations to low pH.

Other factors have been proposed to influence the distribution and activity of AOA and AOB such as soil water content, with observations indicating that AOB could be more resistant to drought (Placella & Firestone, 2013; Thion & Prosser, 2014) and AOA activity and abundance could be enhanced in soils frequently irrigated (Yang et al., 2017). Also temperature was suggested to drive the nitrifiers community, with higher temperatures increasing AOA nitrification activity over AOB (Taylor et al., 2017). All these studies support the hypothesis that the different groups of nitrifiers might be occupying different ecological niches, although the identification of the main environmental drivers behind this niche specialization is still debated.

1.3. AOA have diverse ecophysiologies

AOA are one of the most widespread ammonia oxidizer groups, with presence in practically all sampled environments (reviewed by Hatzenpichler, 2012; Schleper & Nicol, 2010; Stahl & Torre, 2012; Stieglmeier et al., 2014). They are found in marine environments, such as open ocean waters, marine sediments, in association with eukaryotes, and in coastal and estuarine ecosystems; in most terrestrial ecosystems; in freshwater systems; in wastewater treatment plants and in hot springs.

Currently, the AOA are classified within the phylum Thermoproteota, as a monophyletic group within the Nitrososphaeria class (Rinke et al., 2021), although previously they were classified in a separate phylum called Thaumarchaeota (Pester et al., 2011). This change in the taxonomy happened recently and the old nomenclature continues to appear in new publications. Here, we will use the new taxonomy proposed by GTDB.

All characterized AOA so far possess a chemoautolithotrophic metabolism, sustained by the oxidation of ammonia to nitrite. All AOA share the 3-hydroxypropionate-4-hydroxybutyrate pathway used for CO_2 fixation, this is so far the most energy efficient aerobic fixation pathway (Könneke et al., 2014). Additionally, from observations made *in situ* in wastewater treatment plants it was suggested that AOA could harbor alternative metabolisms to chemoautolithotrophy (Mußmann et al., 2011). For instance, potential metabolisms related to complex organic carbon molecules degradation were predicted from genomic information obtained from an enrichment culture of *Candidatus* Nitrosocosmicus arcticus (Alves et al., 2019). This led to the hypothesis that some AOA might not be obligate chemolithoautotrophs. Their cultivation is highly challenging because they are highly sensitive to environmental factors (e.g. pH, substrate concentration, temperature, light, oxygen, reactive oxygen species), they have long generation times and they could be establishing partnerships with other members of the microbial community, as discussed in Erguder et al. 2009. Therefore, we have to rely on environmental surveys and *in situ* activity experiments to understand their distribution and function in the ecosystem.

The diversity of AOA has been usually assessed using the *amoA* gene, which encodes the subunit A of the AMO enzyme. This gene has been used as a phylogenetic marker because it is exclusively found in this lineage, it is conserved and at the same time contains enough phylogenetic information so that phylogenetic relationships can be inferred (Purkhold et al., 2000; Rotthauwe et al., 1997). Furthermore, the *amoA* and the 16S rRNA genes result in congruent phylogenies, meaning that the phylogenetic relationships established at each tree agree with each other (Nicol et al., 2008; Oton et al., 2016). Indeed, *amoA* has become the second most utilized marker gene in microbial ecology after the 16S rRNA gene.

In 2018 Alves et al. reconstructed the AOA phylogeny based on 33,378 *amoA* sequences retrieved from public databases (Alves et al., 2018). The outcome was a highly resolved and supported phylogeny that allowed the definition of clearly supported and distinct clades and subclades within every main lineage (Figure 3). The main lineages known in AOA are classified here in the following families: Nitrosocaldaceae (NC), Nitrosopumilaceae (NP) and Nitrososphaeraceae (NS). Typically, although not exclusively, NS are found in terrestrial and freshwater habitats, NP in marine ecosystems and NC in hot springs. Every main lineage can be divided further into subclades, distinguished by greek letters. The Nitrososphaerales, which are the main terrestrial AOA group, are composed by the clades NS-Alpha, NS-Gamma, NS-Beta, NS-Delta, NS-Epsilon and NS-Zeta. So far, only cultivated representatives of NS-Alpha and NS-Zeta have been successfully isolated from various

types of soils. Nevertheless, the vast majority of NS diversity, including the most abundantly detected clades in soils (i.e. NS-Delta and NS-Gamma), remains so far uncultured. There are a few number of isolates in cultures representing all the main AOA lineages when we compare them to the vast diversity of the group that it is found in the environment (Figure 3). However, these cultures have been crucial for our understanding of the metabolism and physiology of AOA. From the Nitrososphaeraceae lineage there are available in pure culture representatives of clades NS-Alpha and NS-Zeta, namely *Ca.* Nitrososphaera gargensis, *Ca.* Nitrososphaera viennensis, *Ca.* Nitrosocosmicus franklandanus and *Ca.* Nitrosocosmicus oleophilus (M.-Y. Jung et al., 2016; Lehtovirta-Morley et al., 2016; Spang et al., 2012; Tourna et al., 2011).

What Alves and colleagues also observed is that there are distinct patterns of AOA distribution along different environments and environmental factors, such as pH. In the case of the clades most commonly found in soils, it was observed that NS-Delta is more frequent in pH neutral soils (6.5<pH<7.5) and NS-Gamma in acidic soils, NS-Alpha is found in soils with a wide pH range, being the most cosmopolitan soil AOA group, while the clade NS-Zeta is more common in alkaline soils, with the exception of a particular subclade NS-Zeta-2 that occurs in acidic soils. The substrate affinity of the enzyme AMO also differs between clades, with cultivated representatives from NP-Gamma having one of the highest substrate affinity and from NS-Zeta having the lowest substrate affinity of any characterized AOA (M. Y. Jung et al., 2022). This corresponds with the habitats where members from the NS-Zeta have been isolated: *Ca.* Nitrosocosmicus exaquare in wastewater treatment plants (Sauder et al., 2017) and *Ca.* Nitrosocosmicus franklandianus (Lehtovirta-Morley et al., 2016) in highly fertilized soils, both systems with high ammonia concentrations. This clade is opposed to the acidophilic and oligotrophic character that generally is associated with the AOA.

The habitat diversity and specificity in AOA denotes different ecophysiological adaptations on the clade and/or subclade level (ecotypes), the study of which can give very valuable information about the potential function of AOA in a particular habitat. The cultivation of new AOA strains is key to unraveling new physiologies, and although it has greatly advanced, cultivate representatives of the most abundant clades detected in soils by environmental surveys, such as NS-Delta and NS-Gamma, are still missing.

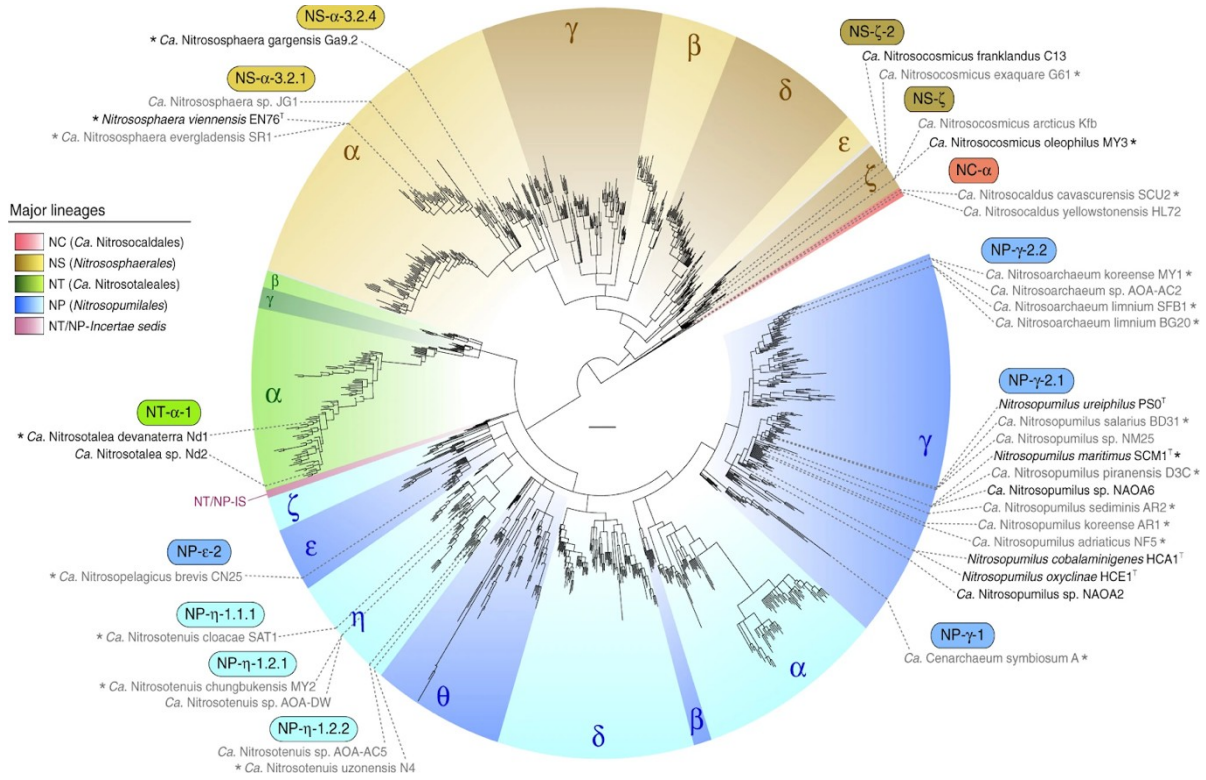


Figure 3: Global phylogeny of archaeal *amoA* genes from cultivated and environmental AOA, adapted by Alves et al. 2018. The major lineages are indicated by five main different colours and the different clades and subclades are indicated by Greek letters. Strains in pure culture are indicated in black letters, and in enrichment cultures, in grey letters. The asterisks indicate that there is available the sequenced genome of the organism. Details on the phylogenetic tree reconstruction can be found in Alves *et al.,* 2018.

1.4. Our site studies and environmental surveys of microbial communities

In the present work we carried out an environmental survey based on the marker gene 16S rRNA to assess the AOA community on a varied range of soil samples. We included on one hand, farmland and grassland soils and on the other, arctic tundra soils. There have been many studies analyzing and comparing the community structure and diversity of AOA, very often in relation with the other groups of nitrifiers (refs). It has been seen that different land management uses have an effect on the community structure and composition of AOA. Changes in the fertilization regimes through the amendments of organic material has been seen to increase diversity and abundance of AOB, as well as to influence the community

composition (Zheng et al., 2022; Zhou et al., 2015). Also, it has been seen that certain crop types can positively influence the AOA abundance (Momesso et al., 2022). Nevertheless, most of these studies do not go deeper into the diversity within the main AOA genera inhabiting soils, namely Nitrososphaeria and Nitrosothaleales. By not looking at the AOA clades composition, valuable information about the ecophysiology and potential function of the AOA in a particular habitat could be disregarded. The study of AOA communities in agricultural and grassland soils, together with the other nitrifier groups, has been in the center of attention of many researchers to help find mitigation strategies against the environmental consequences derived from their activity.

The arctic tundra, although being a completely different ecosystem, it is also a habitat in which AOA play an especially relevant role, particularly in light of climate change. Arctic ecosystems have been facing the consequences of the rising temperatures for decades already and it is predicted that they will continue to do so exponentially (Pörtner & Roberts, n.d.). The arctic soils are estimated to store in their permafrost layer over 1500 Pg of carbon (Koven et al., 2011) and 41 Pg of nitrogen (Strauss et al., 2022), which due to global warming are at risk of becoming available to the microbial communities decomposing activity. Microbial activity results in emissions of CO2, CH4 and N2O, generating a positive feed-back effect to the temperature increase.

In the Arctic, N is a major limiting element (Nordin et al., 2004; Sistla et al., 2012) (Nordin et al., 2004) and nitrification has been detected (Alves et al., 2013) and has been associated mainly to the activity of AOA, while AOB are almost undetectable in arctic soils (Alves et al., 2013; Siljanen et al., 2019). Two AOA subclades have been identified as the main members of the ammonia oxidizer community in arctic soils: NS-γ−2.3.2 and NS-ζ-1.2 (Alves et al., 2013; Siljanen et al., 2019). Only one isolated representative from the arctic NS-Zeta clade has been described, *Ca.* Nitrosocosmicus arcticus, (Alves *et al.,* 2019). A new genus, *Ca.* Nitrosopolaris, affiliated with the arctic NS-Gamma clade was recently proposed by Pessi *et al.* (Pessi et al., 2022) based on several metagenome-assembled genomes (MAG) identified in tundra soils, permafrost soils and polar desert soils from different sites in Norway, Finland, Canada and Antarctica. In the present study we aimed to study whether this apparent low diversity in some AOA arctic communities was a generalized trend to circumpolar arctic area and which could be the environmental drivers behind the community structure. The activity of AOA in the arctic is impactful for the whole microbial community, as they control the bioavailability of nitrogen in the system; thus, the study of their ecology and physiology is key to understand the biogeochemical cycles in the arctic and anticipate potential changes derived from climate change.

Environmental surveys based on amplicon sequencing of key conserved marker genes are a great tool for the exploration of new habitats where there is no previous knowledge about the microbial community structure and composition. It is a technique based on the high-throughput sequencing of a microbial marker gene such as the universally conserved 16S rRNA gene, from which the taxonomic placement of the microbial community members can be inferred and the relative abundances of the taxa present can be calculated. In AOA diversity studies, the most commonly used marker gene is the *amoA* gene, due to its group specificity and conserved nature.

In a recent publication, Wang et al built a database of co-occuring *amoA* and 16S rRNA gene sequences from the available closed genomes and MAGs (Wang et al., 2021), so that is possible to obtain information on AOA *amoA* clade assignment (Alves et al., 2018) through the sequencing of 16S rRNA gene, the most widely used marker gene in microbiome studies. The linkage between the 16S rRNA taxonomy and the *amoA* taxonomy is sustained by the phylogenetic congruency in AOA of these two marker genes (Alves et al., 2018; Oton et al., 2016). In this project, we used this database to extract the *amoA* clades information from 16S rRNA gene reads. This approach is convenient because on one hand, it offers the advantages of using a general prokaryotic marker gene to have a picture of the whole microbial community composition, and on the other, it provides the fine resolution of the *amoA* taxonomy for AOA.

The analysis of the microbial community based on amplicon-sequencing surveys have limitations, for instance, amplification biases caused by the exclusion of certain groups not detected by the primers or the lack of information in the activity and physiology of the groups detected. Nevertheless, they are a powerful exploratory tool that offers a preview of the microbial community in its environmental context and provides links between the certain environmental factors and the community composition. In combination with other techniques, such as metagenomic, metatranscriptomics and cultivation, they are a useful toolkit for the study of the ecology and evolution of AOA.

2. AIMS OF THE STUDY

Although there are multiple studies on the ecology and distribution of the AOA community, very often these do not consider the ecophysiological diversity within the group. The broad distribution of AOA across virtually all sampled ecosystems, the physiological diversity inferred from the limited available cultivated representatives and finally the phylogenetic diversity inferred from the *amoA* gene subunit tell us that AOA should not be treated as an homogeneous group of organisms. The *amoA* gene has been shown to give more information about the different ecophysiological properties from each group, yet studies tackling specifically the AOA ecology using the *amoA* taxonomy are still scarce. Thus, the main purposes of this master thesis were:

- Firstly, to assess the AOA community diversity in two very different soil ecosystems through the analysis of 16S rRNA amplicon sequencing data, with a special focus on identifying the main *amoA* clades using the dataset previously described.
- Secondly, to identify possible patterns of distribution specific for each *amoA* clade, and the environmental drivers behind them based on the available environmental data for each location hereby studied.
- Thirdly, to explore whether the low diverse AOA communities dominated by a few clades observed in other arctic ecosystems occur similarly across the arctic circumpolar area.
- Additionally, we wanted to test the newly developed method by Wang et al. 2021 to obtain the *amoA* clade information based on 16S rRNA gene amplicon sequences.

The results will give insights in the better understanding of AOA diversity and distribution across the environment.

3. MATERIALS & METHODS

3.1. Description of the study sites and sampling

3.1.1 Montana Ranch soils

The farm samples were taken in January 2020 at the west of Choteau (Montana) (Figure 4). The sampling was performed in duplicates about 1 m apart for each site, breaking the soil with a pick and axe, and collecting soil from 5 cm below the surface. Samples were stored at 4ºC until further processing. Land use, fertilization status, herbicide use and irrigation information were provided by the ranch owners. Soil type information was obtained through the Web Soil Survey, from the Natural Resources Conservation Service of the U.S. Department of Agriculture (https://websoilsurvey.nrcs.usda.gov/).

Sample id	Land use	Fertilized	Herbicide	Irrigated	Soil type	pH	SW
BKBC.A	grazed	not fertilized	none	none	loam gravelly sandy loam	8.02	0.58
BKBC.B	grazed	not fertilized	none	none	loam gravelly sandy loam	8.36	0.23
BKCB.A	grazed	not fertilized	spot sprayed	none	loam	8.2	0.3
BKCB.B	grazed	not fertilized	spot sprayed	none	loam	8.35	0.26
BKFS.A	grazed	not fertilized	none	none	mucky peat water	7.45	0.85
BKFS.B	grazed	not fertilized	none	none	mucky peat water	7.56	0.85
BKSP.A	hay field	feritilized	starane flex glyphosate	center pivot	loam	7.96	0.25
BKSP.B	hay field	feritilized	starane flex glyphosate	center pivot	loam	7.96	0.33
BKWB.A	hayed grazed	not fertilized	spot sprayed	flood irrigated	clay loam	7.67	0.65
BKWB.B	hayed grazed	not fertilized	spot sprayed	flood irrigated	clay loam	7.89	0.59
DWL.1A	grazed	not fertilized	spot spraying	flood irrigated intermittent	clay loam	8.66	0.6
DWL.1B	grazed	not fertilized	spot spraying	flood irrigated intermittent	clay loam	8.41	0.53
DWL.2A	grazed	not fertilized	spot spraying	flood irrigated intermittent	clay loam	10.24	0.08
DWL.2B	grazed	not fertilized	spot spraying	flood irrigated intermittent	clay loam	10.31	0.15
DWL.3A	CRP not grazed	not fertilized	spot spraying	none	clay loam	8.3	0.32
DWL.3B	CRP not grazed	not fertilized	spot spraying	none	clay loam	8.44	0.11
DWL.4A	grazed	not fertilized	spot spraying	none	silty clay loam loam	8.1	0.08
DWL.4B	grazed	not fertilized	spot spraying	none	silty clay loam loam	7.8	0.37
DWL.5A	grazed	not fertilized	spot spraying	none	cobbly loam	8.12	0.46
DWL.5B	grazed	not fertilized	spot spraying	none	cobbly loam	8.3	0.14
DWL.6A	grazed	not fertilized	spot spraying	none	loam	7.06	0.33
DWL.6B	grazed	not fertilized	spot spraying	none	loam	7.21	0.26
DWL.9A	crop production alfalfa	fertilized intermittment	spot spraying	wheelline irrigated	loam	8.4	0.05
DWL.9B	crop production alfalfa	fertilized intermittment	spot spraying	wheelline irrigated	loam	8.37	0.03
DWL.10A	none	not fertilized	none	none	clay loam	8.57	0.39
DWL.10B	none	not fertilized	none	none	clay loam	8.28	0.354
DWL.11	none	not fertilized	none	none	loam	9.07	0.43

Table 1: Overview on the ranch soil samples, with the environmental context data.

3.1.2 Arctic soil samples

The arctic soil samples were taken between 2007 and 2012 during different sampling expeditions, in the frame of the Cryocarb project (https://www.univie.ac.at/cryocarb/the-cryocarb-project/), an international project aiming to provide accurate estimates of organic carbon storage in cryoturbated arctic soils in Eurasia and its dynamics in a changing climate. The samples used in the present study were taken by the project research associates from 4 different locations: Cherskii (Republic of Sakha, Russia) (Gittel et al., 2014a), Zackenberg (Greenland) (Gittel et al., 2014b), Taymyr Peninsula (Central Siberia), and Tazovsky (Western Siberia) (Gentsch et al., 2015). These four locations, depicted in Figure 4, are found along the arctic circumpolar area and the dominant vegetation type is tundra. Soil data available for the different locations was collected from the original studies, and additionally from Wild *et al.,* 2013 and Gentsch *et al.,* 2018 (Supplementary Table 1 and 2).

On each location, several sites were selected based on their dominant vegetation subtypes. These subtypes are listed in Table 2. On each site, two to three peats were excavated down to the permafrost layer, approximately down to 1 m deep. Subsequently, soil profiles were classified into different soil horizons defined in the Keys to Soil Taxonomy (Soil Survey Staff, 2010). In SI Figure 1, there is a schematic representation of an arctic soil profile with its corresponding descriptions and diagnosis. Replicates were taken along the length of each horizon and were pooled together to obtain a single representative sample per horizon. In the present work, we pooled together all the organic topsoil (i.e. O horizon) samples belonging to the same site chosen based on the vegetation subtype. In Table 2, we summarize which samples were grouped together and the naming of the new pools. In addition, we also included in our study the samples belonging to the buried topsoil fractions (i.e. Ajj horizon). Site descriptions and vegetation types were taken from the publications Gittel *et al.,* 2014a, Gittel *et al.,* 2014b and Gentsch *et al.,* 2015.

Pool	Location	Description site	Vegetation type	Horizon layer	Original samples
CHP0	Cherskii	shrubby grass tundra	*Betula exilis, Salix phenophylla, Carex lugens, Calamagrostis holmii, Aulacomnium turgidum*	Ajj	CHA2
CHP1	Cherskii	shrubby grass tundra	*Betula exilis, Salix phenophylla, Carex lugens, Calamagrostis holmii, Aulacomnium turgidum*	O	CHA3, CHB7, CHC16
CHP2	Cherskii	shrubby tussock tundra	*Eriophorum vaginatum, Carex lugens, Betula exilis, Salix pulchra, Aulacomnium turgidum*	O	CHD9, CHD10, CHE7, CHE10, CHF5, CHF6,
CHP3	Cherskii	shrubby tussock tundra	*Eriophorum vaginatum, Carex lugens, Betula exilis, Salix pulchra, Aulacomnium turgidum*	O	CHG11, CHH6, CHI12
CHP4	Cherskii	Shrubby lichen tundra	*Betula exilis, Vaccinium uligonosum, Flavocetraria nivalis, Flavocetraria cucullata*	A	CHH7
CHP5	Cherskii	Shrubby lichen tundra	*Betula exilis, Vaccinium uligonosum, Flavocetraria nivalis, Flavocetraria cucullata*	Ajj	CHI4
ZKP6	Zackenberg	Tundra vegetation with old frost boils, (already re-vegetated by lichens and some higher plants, e.g. Dryas octopetala) and some frost cracks	*Vaccinium uliginosum, Salix arctica, Carex sp., Polygonum viviparum, Dryas octopetala,mosses, lichens on bare soil patches*	O	ZKA1, ZKB1, ZKC1
ZKP7	Zackenberg	Tundra vegetation with old frost boils, (already re-vegetated by lichens and some higher plants, e.g. Dryas octopetala) and some frost cracks	*Vaccinium uliginosum, Salix arctica, Carex sp., Polygonum viviparum, Dryas octopetala,mosses, lichens on bare soil patches*	Ajj "young"	ZKB4
ZKP8	Zackenberg	Wet fen with ~ 5 % mossy hummocks on flat surface	Carex species, Eriophorum angustifolium (~5%), mosses, several grasses	O	ZKD1, ZKE1, ZKF1
ZKP9	Zackenberg	Almost flat site with active frost boils and earth hummocks (differences in relief up to 30 cm)	Cassiope tetragona, Salix arctica, Vaccinium uliginosum, Dryas octopetala, grasses, Carex sp., mosses; some lichens on bare soil patches	O	ZKG1, ZKH1
AMP10	Taymyr (Ari-Mas)	shrubby grass tundra	*Betula nana, Dryas punctata, Vaccinium uliginosum, Carex arctisibirica, Aulacomnium turgidum*	O	AMA10, AMB10, AMC11
AMP11	Taymyr (Ari-Mas)	shrubby tussock tundra	*Cassiope tetragona, Carex arctisibirica, Aulacomnium turgidum*	O	AMD14, AME12, AMF11
AMP12	Taymyr (Logata)	Dryas tundra	*Dryas punctata, Rhytidium rugosum, Hylocomium splendends*	O	LGB9, LGC17
AMP13	Taymyr (Logata)	Grassy moss tundra	*Betula nana. Carex arctisibirica, Hylocomium splendens, Tomentypnum nitens*	O	LGD17, LGE16
TZP14	Tazowvsky	Shrubby lichen tundra	*Empetrum nigrum, Ledum palustre, Betula nana, Cladonia rangiferina, C. stellaris*	O	TZA11, TZB11, TZC12
TZP15	Tazowvsky	larch woodland with shrubby lichen understory (forest-tundra zone)	*Larix sibirica, Ledum palustre, Betula nana, Vaccinium uliginosum, Cladonia rangiferina, C. stellaris*	O	TZD6, TZE7, TZF6

Table 2: Overview on the arctic soil pools, with a description from the sampling sites and the original sample names that were used on each pool.

3.2. pH and soil water content measurements

Soil water content (SW) and pH were measured on the Montana farm soils. For each sample to be extracted, 6 mL of MQ-water were transferred in a 15-mL centrifuge tube, into which 2 g of fresh soil were added. The tube was shaken heavily by hand for a few seconds to make sure that all soil aggregates were homogenized. Then the tubes were incubated on an orbital shaker for 60 minutes at 300 rpm. The pH measurement was done in the supernatant, using a calibrated pH. For the SW, 3 g of fresh soil samples were weighed on an aluminum weighing dish, and again the next day after drying them overnight at 150ºC. The SW was calculated with the following formula:

$$SWC = \frac{W_{f+t} - W_{d+t}}{W_{f+t} - W_t}$$

3.3. Nucleic acid extraction

DNA for qualitative PCR and quantitative PCR (qPCR) was extracted from 300 mg of soil sample (wet weight). Soils were weighed and transferred into a Lysing tube E (MP Biomedicals, USA), with a sterilized spatula. DNA was extracted with the DNAEasy PowerSoil Pro Kit (Qiagen). The extraction was done following the steps described by the manufacturer, using the bead beating method for the cell lysis with the FastPrep-24™ Classic (MP), during 45 s at 5.5 m/s. To improve the extraction yield, the MB Spin Columns were incubated for 2' with the elution buffer, preheated at 55ºC, before proceeding to the elution of the DNA. The DNA extracts were then purified with the OneStep PCR Inhibitor Removal Kit (ZymoResearch) and kept at -20 ºC until analysis. The DNA quality was verified with NanoDrop™ (ThermoFisher) and the concentration was quantified with the Qubit™ 2.0 Fluorometer (Invitrogen) using the dsDNA BR Assay Kit.

DNA extractions were performed in the current study for the farm samples. The arctic samples had been previously extracted by Dr. Ricardo Alves during the course of the studies previously cited, and original extractions and dilutions were stored at -80 ºC and -20ºC since then.

3.4. PCR amplification, library preparation and next generation sequencing

PCR amplification of the 16S v4 rRNA gene was done for the farm samples using the universal primers 519F (5'-CAGCMGCCGCGGTAA-3') (Burggraf et al., 1997) and 805R (5'-GACTACNVGGGTWTCTAAT-3') (Apprill et al., 2015), which were previously used in (Alves et al., 2019). DNA amplicons were sequenced with Illumina MiSeq Paired-End sequencing (2x300 bp). The PCR step and the sequencing were done by Novogene (UK).

3.4.1 Nested PCR amplification

The 16S rRNA gene amplicons for the arctic samples were produced following a nested PCR approach described in (Pausan et al., 2019), which aims to enrich the AOA pool, which are expected to be in very low abundance in arctic soils. The primer pair combinations used were the following:

- First PCR round, performed in house: 344f (5'-ACGGGGYGCAGCAGGCGCGA-3') and 1041R (5'-GGCCATGCACCWCCTCTC-3'). Reactions were carried out in triplicates of 50 ul reaction volume containing: 0.4 U Phire Hot Start II DNA Polymerase (ThermoFisher), 1x Phire Reaction Buffer, 0.2 mM dNTPs, 3% DMSO, 0.5 uM of each primer. The cycling conditions were the following: 3 min initial denaturation step at 98ºC, followed by 25 cycles of 10 s denaturing at 98ºC, 30 s annealing at 56◦C and 30 s extension at 72◦C, with a final extension step of 5 min at 72◦C. Pooled triplicate PCR products were column-purified with the Monarch® PCR & DNA Cleanup Kit (New England BioLabs, US), prior to send them to the sequencing facility.
- Second PCR round: Arch-519F (5'-CAGCCGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3'), performed and sequenced by Novogene (UK).

3.5. Quantitative PCR (qPCR)

Archaeal and bacterial *amoA* genes were quantified in the farm samples using qPCR. For the archaeal *amoA* the primer pair C*amoA*-19F (5′-ATGGTCTGGYTWAGACG-3′) (Pester et al., 2012; Tourna et al., 2008) and T*amoA*-629R (5′-TGGCANTAYMGATGGATGGC-3′) (Arce et al., 2018) were used, and for the bacterial *amoA* we used the *amoA*-1F* (5'-GGGGHTTYTACTGGTGGT-3') (Stephen et al., 1999) and the *amoA*-2R (5'-CCCCTCKGSAAAGCCTTCTTC-3') (Rotthauwe et al., 1997). The quantification for each sample was performed in technical triplicates, in volumes of 20 ul, which contained: 10 ul of GoTaq® qPCR Master Mix 2x (Promega), 0.1 mg/ml BSA, 0.5 uM of each primer and 5 ml template DNA. The cycling conditions were the same for each primer pair and were as it follows: 95ºC for 2 min, followed by 40 cycles of 15 s denaturing at 95ºC, 60 s joint annealing-extension at 60ºC, and fluorescence measurement at 60ºC for 10 s. Standard dilutions were prepared in triplicates ranging from 10 to 10^8 gene copies μL^{-1} using for archaeal *amoA* standards long gene fragments from *N. viennensis* EN76 amplified with primers NV-L*amoA*-F (5′-CGCATGATCGGCCGCAGAGT-3′) and NV-L*amoA*-R (5′-GCCTAGTAGCGACCCGCCCT-3′) (Alves et al., 2019) (Alves et al. 2019). For the bacterial *amoA* standards *amoA Nitrosospira multiformis* were generated with NM-*amoA*B-F (5'-GAGGGAGGGTGCGATGAGCAGA-3') and NM-*amoA*B-R (5'-ACGTTGAAGAAGGCGCGTCCAG-3'). The specificity of the amplification products was confirmed with a melting curve analysis, performed at the end of the amplification by increasing the temperature 65.0°C to 95.0°C in increments of 0.5°C every 5 s. The quantitative PCRs were performed in CFX Real Time PCR System (Bio-rad). The efficiency of qPCR assays ranged between 74 and 90% with R2 ≥ 0.95.

3.6. Sequence processing

Firstly, the adaptor and primer sequences from the amplicons were trimmed with CUTADAPT (Martin, Marcel, 2011). Then, amplicons were processed using the open source package DADA2 (Divisive Amplicon Denoising Algorithm) (Callahan et al., 2016) in R v. 4.2.1. This pipeline manipulates the paired-end fastq files with several tools obtaining as an end product the Amplicon Sequencing Variants (ASVs). ASVs are the analogues of traditional OTUs, but they are considered to have higher resolution because ASVs are distinguished at a nucleotide level. The key steps performed by DADA2 are: filtering and the trimming of low quality sequences (forward and reverse reads were truncated at position 170 and 200, for the farm samples and arctic samples, respectively); construction of an error model based on the specific error signatures among our data; dereplication of the exact same sequences; finally, the inference of the ASVs. Next, the forward and reverse reads are merged and the chimeras are removed. In addition, DADA2 performs the taxonomy assignment from the output sequences, which for this study we used the GTDB v. 202 with the RDP classification method.

3.7. Microbial diversity and statistical analysis

Analysis was performed with the packages ampvis, phyloseq and vegan, rstatix in R. Alpha diversity measures (Observed OTUs, Shannon, Simpson, InvSimpson, Chao1 and ACE) were calculated with the amp_alphadiv function from the ampvis package. We tested for significant differences in the indexes among sites and among the available environmental parameters with ANOVA test followed by a TukeyHSD test, with the rstatix R package. With the ampvis function amp_ordinate, a PCoA was generated to represent the differences in the community structure, from a distance matrix automatically calculated by the r function, using the Bray-Curtis method. To statistically assess the effect of the environmental parameters in the community structure, we used the PERMANOVA test with the adonis function in the vegan r package. Additionally, we performed an analysis of the homogeneity of group dispersions with the betadisper function from the vegan package. This test was done to explore the variance within the groups of each environmental category.

3.8. Phylogenetic analysis

The ASVs assigned to Nitrososphaeria were extracted and aligned against a reference alignment containing 220 AOA reference sequences from most known *amoA* clades (provided by Dr. Rafael Ponce). For alignment between the ASVs sequences and the reference alignment we used the online tool MAFFT v. 7.490 (Katoh & Toh, 2008), using as

a scoring matrix the 1PAM / k=2. This alignment was then used for the construction of a maximum likelihood (ML) phylogenetic tree in IQTREE v. 2.2.0 (Trifinopoulos et al., 2016) under the TIM3e+I+G4 model with 1000 ultrafast bootstrap replicates (Minh et al., 2013).

3.9. Identification of the *amoA* clades using a 16S-*amoA* database

To identify to which known *amoA* clades the Nitrososphaeria ASVs belong to, we used the pipeline designed by (Wang et al., 2021). In their work, a database was compiled comprising 16S rRNA genes retrieved from publicly available genomes, metagenomes and contigs on which the correspondent *amoA* gene was also encoded. This resulted in a database comprising 124 16S rRNA gene sequences linked to their corresponding 124 *amoA* gene. In this way, this database (from now on referred to as the 16S-*amoA* database) allows the use of the most commonly used marker gene in microbiome studies with the addition of the valuable high-resolution taxonomic information of the *amoA*. The Nitrososphaeria ASVs were clustered with the 16S-*amoA* database using UCLUST (Edgar, 2010), with a > 90% sequence similarity setting. The allocation of an ASV into an *amoA* clade occurred when at least two out of the three best hits matched a specific assignment. In order to enrich the 16S-*amoA* database from Wang *et al.,* we added the AOA reference sequences used to build the phylogenetic tree. The enriched 16S-*amoA* database was used to assign the soil ASVs.

Additionally, we wanted to define the nucleotide identity percentage threshold that the database establishes on the *amoA* clade and the subclade level. In other words, which is the minimum identity percentage between the sequences belonging to a same clade or subclade. To do that, we performed a multiple sequence alignment with the online tool MAFFT v. 7.4 with the 16S rRNA gene sequences from the original 16S-*amoA* database and from the phylogenetic tree. Next, a nucleotide identity matrix was calculated using the R package Bios2cor, from the Comprehensive R Archive Network (cran.r-project.org), which was then represented as a heatmap using the R package pheatmap v. 1.0.12.

3.10 - Statistical analysis of the *amoA* clades distribution and the environmental parameters involved

Once the classification into *amoA* clades was defined, the relative abundance for each clade was calculated and visually represented with barplots. In the case of the farm samples, the *amoA* clade relative abundances were used to estimate the absolute abundance of each

clade, by multiplying the *amoA* copy number measured by the qPCR with the proportion of the different *amoA* clades on each sample. These were then compared between the different environmental categories with boxplots, and statistically with One way ANOVA tests followed by TukeyHSD tests. In the case of the continuous variables (i.e. pH and SW), a Spearman's rank correlation was calculated for each *amoA* clade.

To explore the potential relationships between the environmental parameters and the *amoA* clades in the arctic tundra soils, Canonical Correspondence Analysis (CCA) models were built, using the vegan package in R. For that we used the environmental data matrix with the soil parameters or the enzymatic activity that had been previously log transformed, and the counts matrix including all the Nitrososphaeria counts on each sample. The models were then tested for significance with the function anova.cca from the vegan package in R. Furthermore, we tested the specific correlation between the environmental parameters and the relative abundance each AOA clade separately with the Spearman's rank correlations.

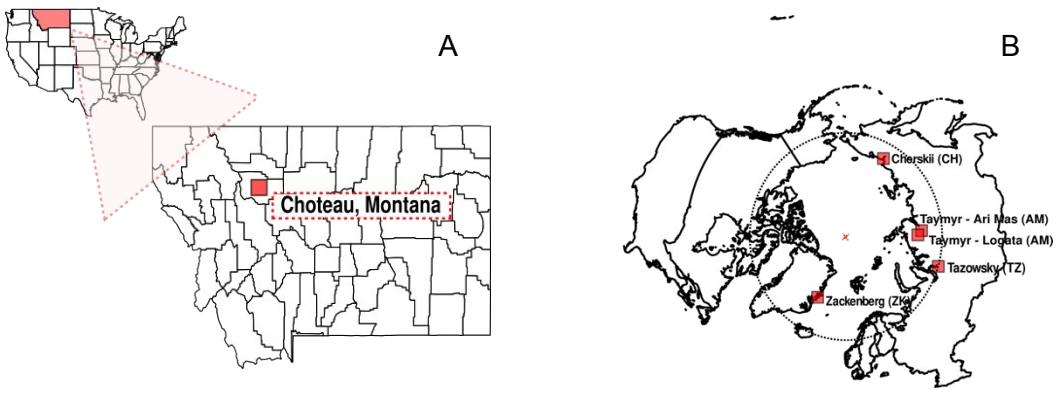


Figure 4: Maps with the samples locations explored in the present work, marked in red squares. Choteau, Montana (USA) (A), and the circumpolar Arctic region, marked with the dashed circle (B).

4. RESULTS

4.1 AOA communities of two Ranch sites in Montana, USA

4.1.1. Sites description

The area from where the Montana ranch samples were obtained is located 30 km east from the Rocky Mountains. It has a semi-arid climate with a mean annual temperature of 6.5 ºC, and a mean annual precipitation of 268mm. The warmest months (May-July) have the highest frequency of precipitation, eventually causing floods, while the driest season occurs during the coldest months. The sampling was performed in two different ranches: Ben&Karli Ranch (BK) and the Diamond Lazy W Ranch (DWL). The main activity of both ranches is the raising of grazing cattle. Additionally, they also grow hay and other grass-like plants, such as fodder for the livestock and other crop types. Both properties cover an extensive area that consists of various types of terrestrial ecosystems. Thus, the selection of the studied sites (Table 1, Figure 5) within each location aimed to cover all the different types of ecosystems present on both ranches.

In the BK (Figure 5 A-D) ranch dataset there are sites that are almost all year round saturated with water, such as BKFS and BKBC. Sites BKSP and BKCB are adjacent to each other, thus, they share the soil type, but the first is agriculturally exploited, while the latter is undisturbed. Site BKWB is an old hay field that has been untouched for the last three years. Most of the sites are frequently grazed by cattle, except the only one which is used agriculturally (BKSP).

The DWL ranch included more varied ecosystems (Figure 5 E-L). Site DWL1 is a typical range grass field used for cattle grazing. In distinct spots from this area, one finds distinct patches with highly alkaline soil (i.e. DWL2). Sites DWL3 and 4 belong to the same pasture area and are adjacent to each other, but the first is under a Crop Restoration Program, a conservation program for the recovery of the native grasses, while the latter is actively used for cattle grazing. Sites DWL5 and DWL6 are found on the top and on the slope of a rocky hill, respectively. Site DWL9 is found in an active hayfield and site DWL10 is typically always saturated with water, as it is found on the bank of a creek. Finally, site DWL11 is situated on a corral, where livestock is kept.

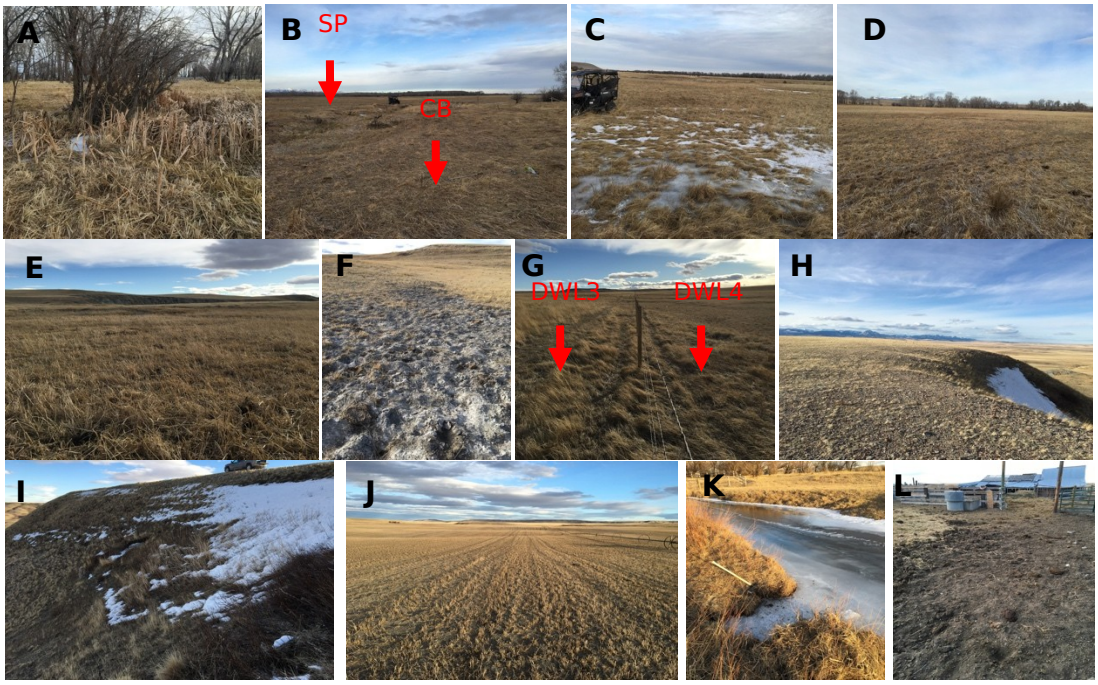


Figure 5: Images from the sampling sites in the Montana ranches. BKBC (A), BKCB and BKSP (B), BKFS (C), BKWB (D), DWL1 (E), DWL2 (F), DWL3 and DWL4 (G), DWL5 (H), DWL6 (I), DWL9 (J), DWL10 (K), DWL11 (L).

4.1.2. Sequencing results and taxonomic assignment

After the initial filtering and adaptors trimming process, the 27 samples accounted for a total of 2,956,801 reads paired-end reads, with a median per sample of 108,058 reads. The average length per sample was 253 nucleotides. The processing with DADA2 resulted in 2,350,936 amplicon reads, with a median of 86,071 reads per sample, from which 20,349 ASVs were recovered. The rarefaction curves (SI Figure 2A), show high sequencing coverage in all samples. From the obtained ASVs, 212 were successfully taxonomically assigned to Archaea and 19,984 to Bacteria, that is 1% and 98.2% from the total ASVs respectively.

4.1.3. Composition and dynamics of the general prokaryotic community

In general, the most abundant phyla are observed in both locations, although their proportions vary between sites (Figure 6). The two most abundant phyla are Proteobacteria and Actinobacteriota, with a trend for a higher prevalence of Actinobacteria over Proteobacteria observed in the DWL farm. The Proteobacteria group comprises mainly Burkholderiales and Rhizobiales (Figure 7), the first ones being more predominant in the BK farm. The Rhizobiales seem to be present across all sites in a homogeneous way representing between a 5 and a 10% of the relative abundance from the total microbial community.

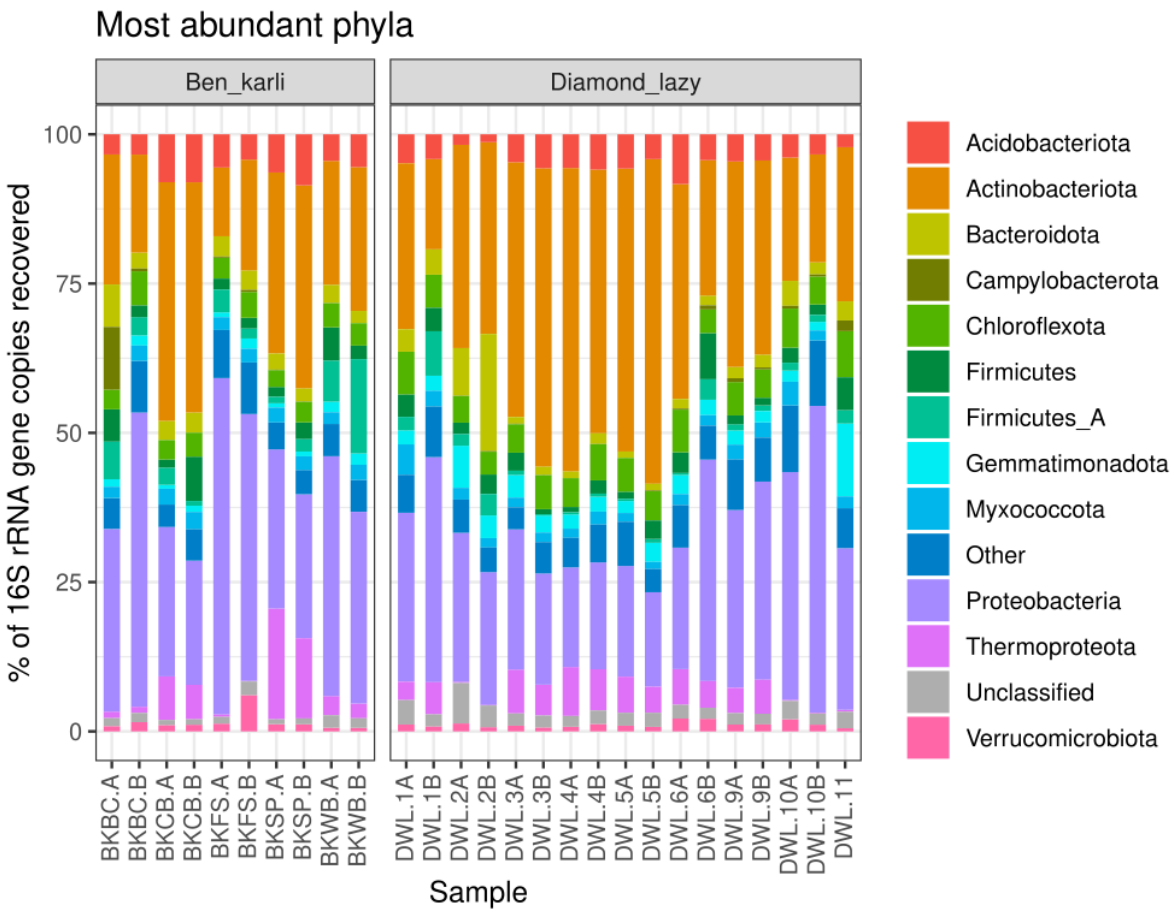


Figure 6: Barplot with relative abundances of 16S rRNA gene amplicons of the general prokaryotic community.

The Actinobacteriota group is mainly composed of Solirubrobacterales and Rubrobacterales (Figure 7). Exceptionally, in samples DWL2A and B, which came from the alkaline patches sampled in the grass field, the most abundant order is Nitriliruptorales. This group is known to inhabit super alkaline and hypersaline samples (Sorokin et al., 2009).

129 ASVs belonging to Nitrososphaerales (AOA) were detected in all samples, with a relative abundance ranging between <0.5% up to almost 19% from the total prokaryotic community. In the sites BKBC, BKFS and DWL10, where the abundance of AOA is below 1%, we find the highest proportion of Burkholderiales. The proportion of organisms from the other groups of nitrifiers (i.e. AOB and Comammox) was lower than 1% of the total prokaryotic community (SI Figure 3).

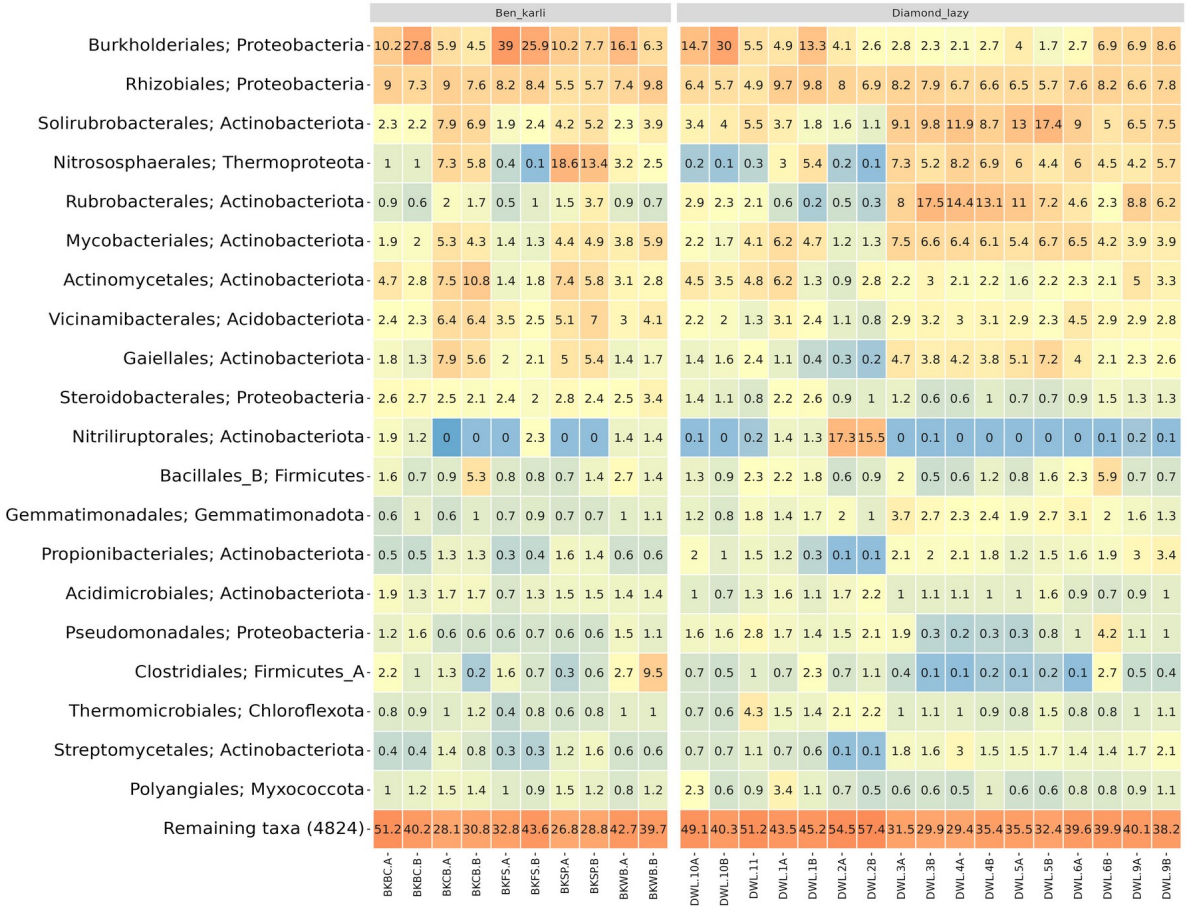


Figure 7: Heatmap showing the most abundant orders in the prokaryotic community, ordered by abundance.

4.1.4. Alpha diversity and Beta diversity

As it is commonly observed in soil ecosystems, both locations are highly diverse, as indicated by the alpha diversity estimators plotted in Figure 8. The alpha diversity indexes range in values usually observed in soil microbial communities. There are no significant differences in the alpha diversity indexes between the two locations. However, in several sampling sites the alpha diversity varies between the two biological replicates. In addition, we grouped the samples into the different environmental variables shown in the metadata table (Table 1) and compared the alpha diversity between categories by testing them with One-way ANOVA. We found that the irrigation regime and the pH give significant differences in the Chao1, with p-values of 0.04 and 0.02, respectively.

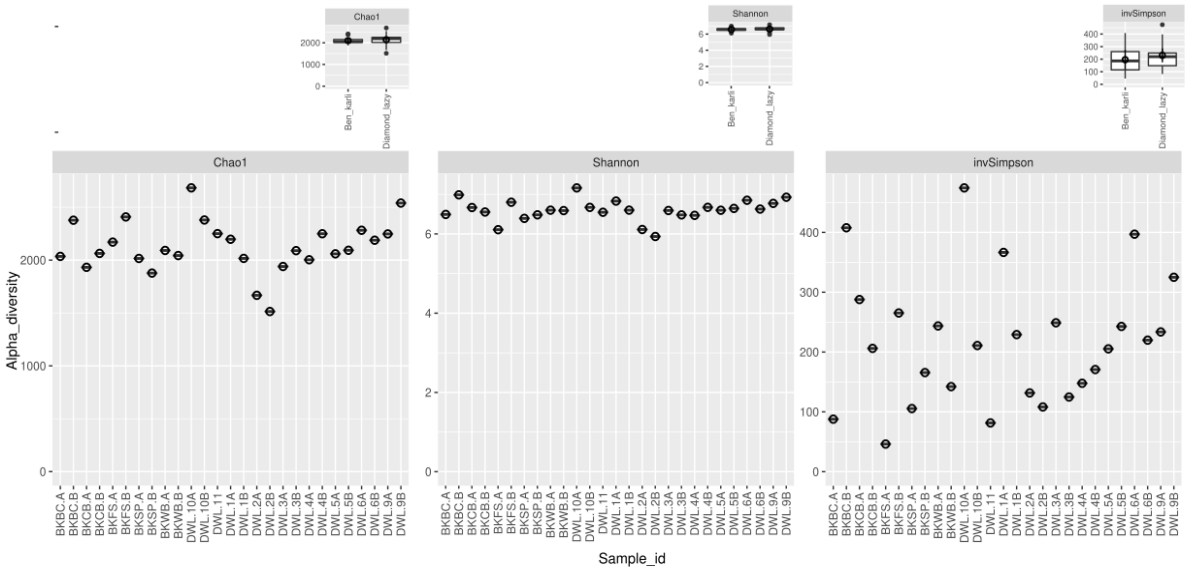


Figure 8: Plots with the **a**lpha diversity measures: Chao1, Shannon, and inverted Simpson. The averages of every alpha diversity index at each location are represented in the small plots above, and the index values for every individual sample are represented in the main panels below.

The overall microbial community structure for the 26 samples was significantly different between the two farms, based on PERMANOVA with a p-value = 0.0003332. Thus, we analyzed the beta diversity in relationship with the different environmental variables for the samples of each location separately. In Table 3 the results from this analysis are summarized, which show that the community structure is affected significantly by all the environmental variables, except fertilization. The soil type and the irrigation regime seem to have a stronger effect on the beta diversity.

Table 3: P-value from the PERMANOVA tests, used to analyse the effect of the environmental factors on the community composition in the two farms.

	Ben&Karli	**Diamond Lazy**
Soil type	PERMANOVA p-value = 0.0006	PERMANOVA p-value = 0.02
Fertilization	PERMANOVA p-value = 0.127	PERMANOVA p-value = 0.05831
Herbicides	PERMANOVA p-value = 0.014	PERMANOVA p-value = 0.04099
Irrigation	PERMANOVA p-value = 0.04465	PERMANOVA p-value = 0.0003
Land use	PERMANOVA p-value = 0.04399	PERMANOVA p-value = 0.01666

The analysis for the homogeneity of the variances within samples from the same category was tested with Betadisper and resulted in significant differences between the sample variances for all environmental variables (p-values < 0.05). This means that the differences observed in the structure of the community could be influenced by the existent heterogeneity

of the data within each group. Nevertheless, the PERMANOVA tests for the irrigation and the soil type result in high levels of significance and in the PCoA (Figure 9) there is a clear visual separation between samples belonging to some categories from these two variables.

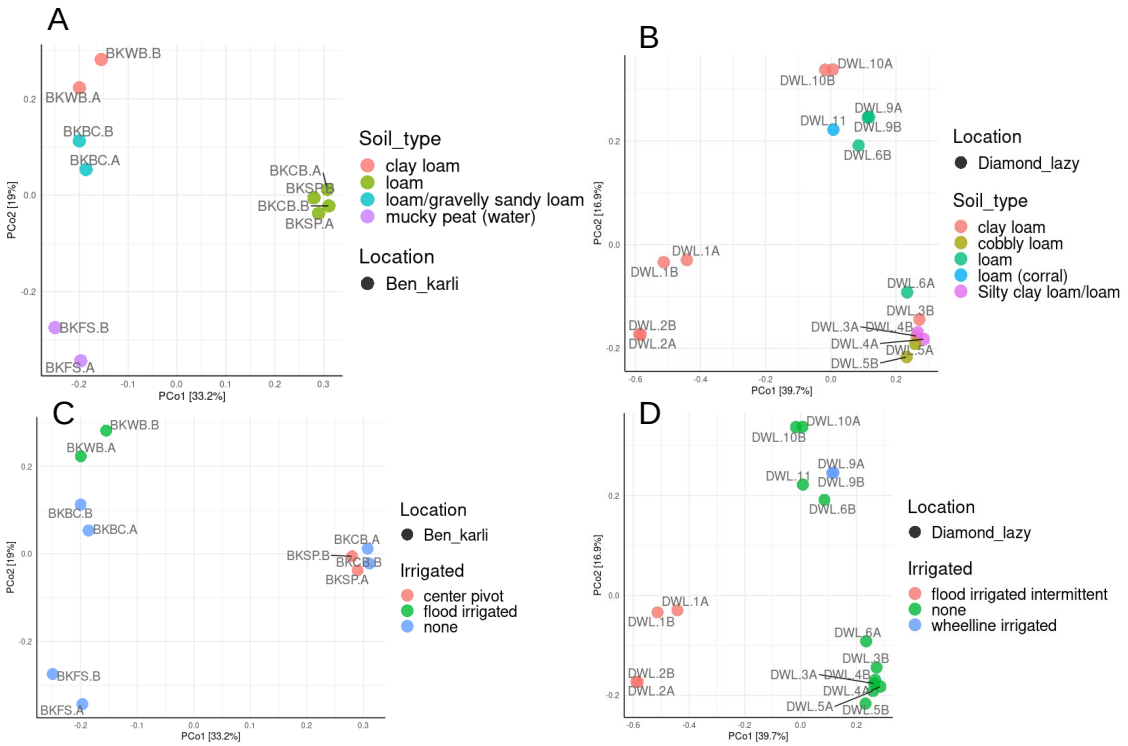


Figure 9: Principal coordinates analysis plots based on Bray-Curtis distance matrix for read counts in BK and DWL. The soil type is presented in the following color code for BK (A): clay loam (red), loam (green-olive), loam/gravelly sand (blue), mucky peat (purple); and in DWL (B): clay loam (red), cobbly loam (green-olive), loam-corral (turquoise), silty clay loam (purple). The irrigation regime is color coded in BK (C) as: center pivot (red), flood irrigated (green), none (blue); and in DWL (D): flood irrigated intermittent (red), wheeline irrigated (blue), none (green).

To study the relationship between pH and SW with the community structure, we performed a Canonical Correspondence Analysis (Figure 10). The CCA captured 29% (BK) and 22% (DWL) of the total variability among the samples as a function of the pH and SW. The CCA models for each farm were statistically supported by ANOVA CCA test, which resulted in p-values of 0.041 (BK) and 0.004 (DWL). Soil water content seems to have a negative influence in the beta-diversity of both farms, particularly in the case of samples BKFSA-B and BKWBA, in Ben&Karli farm, and in DWL10A-B and DWL11. These sites indeed have the highest values of SW, ranging from 0.4 to 0.85. In the case of the pH, only the farm DWL showed a negative impact in the beta diversity, with the site DWL2 being the most affected by it. This site has the highest pH values among all samples (i.e. pH 10).

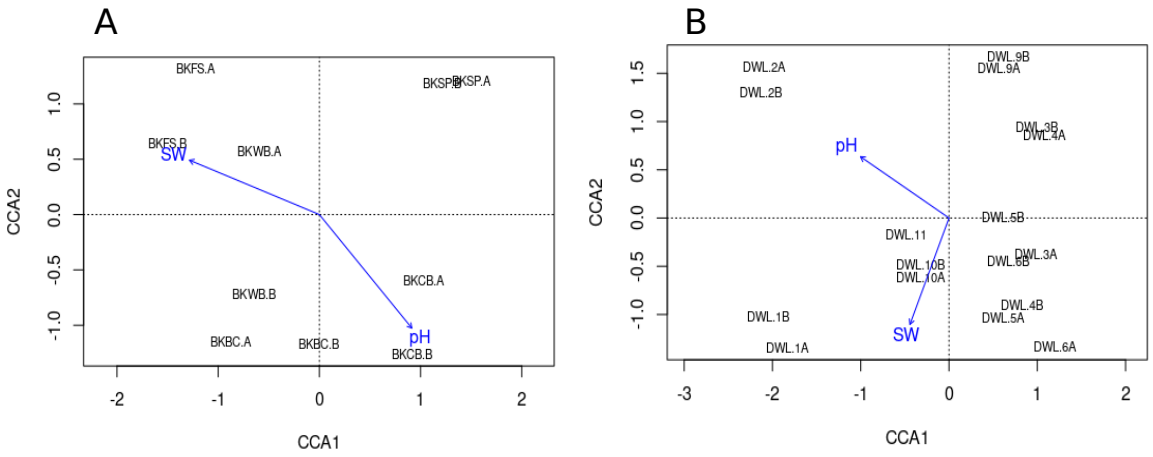


Figure 10: Canonical Correspondence Analysis biplot showing the relationship between pH, SW and the microbial community structure for BK (A) and DWL (B)

4.1.5. AOA amoA clade classification and phylogenetic tree reconstruction

In order to classify the AOA ASVs into the *amoA* clades established by Alves et al 2018, we used the 16S-*amoA* database constructed by Wang et al 2021 (Wang et al., 2021). First, we extracted the 129 ASVs that were classified within the Nitrososphaeria class by the taxonomic assignment with the GTDB database and run them against the 16S-*amoA,* using the clustering script from the original publication. From these, 125 were assigned in 6 *amoA* clades, while 4 were unclassified.

To support the accuracy of this assignment, we built a phylogenetic tree based on a reference alignment of full length 16S rRNA gene sequences from AOA, each of them with their corresponding *amoA* clade information, and the Nitrososphaeria ASVs subset from our samples (Figure 11). We expected the ASVs to mainly cluster with the reference sequences belonging to the different Nitrososphaerales clades, also known as the terrestrial *amoA* clades, and support the assignment we obtained from the database.

The placement of the ASVs on the phylogenetic tree shows us that over half of the Nitrosophaeria ASVs form independent branches which do not include any of the reference sequences used for the construction of the tree. These ASVs form two separate branches: NS-Unclassified-1 (UltraFast bootstrap 92%) and Putative NS-Zeta (UltraFast bootstrap 79%). The reason for these names is justified in the following section. Unexpectedly, NS-

unclassified-1 forms a not very well supported sister branch to the Nitrosopumilales/Nitrosotaleales lineage (UltraFast bootstrap 79%). The Nitrososphaeria ASVs that do cluster together with tree reference sequences are distributed mainly within the NS-Alpha and the NS-Zeta branches. Also, a few ASVs cluster within the NS-Delta and the NS-unclassified-2 branches, the latter formed by the reference sequences coming from partial 16S rRNA gene sequences, from which we do not have *amoA* clade information.

Subsequently, we compared the clade classification obtained by the 16S-*amoA* database with the phylogenetic tree placement (Figure 11). From this, we can see that there is a mismatch between the *amoA* clades assigned by the 16S-*amoA* database and the clades inferred from the phylogenetic tree for 32 ASVs. For instance, the database classifies over 10 ASVs as NS-Gamma, but in the tree none of these ASVs cluster within the NS-Gamma reference sequences. On the contrary, these ASVs are scattered across the tree. In the rest of ASVs (32) the results from the clade assignment in both methods were coincident.

4.1.6. Improvements on the method for assigning 16S rRNA gene amplicons into *amoA* clades

In order to improve the method accuracy and obtain reliable assignments, we followed different approaches. First, we included in the 16S-*amoA* database the 16S rRNA gene sequences used to reconstruct the reference tree that provided the *amoA* clade information. Thus, all reference sequences were included with the exception of those forming the branch NS-unclassified-2. Secondly, with this “enriched” database, we performed a multiple sequence alignment with only the database sequences. From the alignment, we calculated the nucleotide sequence identity between all sequences, that was visualized via a heatmap and hierarchical clustering (Fig. 12). In this way, we were able to see clear clustering of the sequences belonging to the same clade, which allowed us to define a threshold of sequence identity between different clades. All sequences belonging to the same clade shared ≥ 97% of their sequence, while all sequences from other clades were <97% similar.

With this threshold established, we reconstructed the alignment and sequence identity matrix but now also including the Nitrososphaeria ASV sequences from our samples. In the SI Figure 4, the sequence identities of our ASVs compared to the sequences from the “enriched” database are shown in a heatmap. From this we evaluated the sequence identity values of each ASV, taking into account the defined clade threshold. Some ASVs had a sequence identity ≥ 97% with the database sequences from a specific clade; in this case they were validated as adequately classified. On the contrary, other ASVs always had

sequence identities under 97%; thus, these ones were reclassified as NS-Unclassified. There were cases where the assignment of some ASVs with the new enriched 16S-*amoA* placed these into a different *amoA* clade. This is the case for all ASVs classified as NS-Gamma in the first classification, which after revising them they were reclassified mostly into NS-Alpha or NS-Unclassified. The second column of Figure 11 shows the improved *amoA* clade classification.

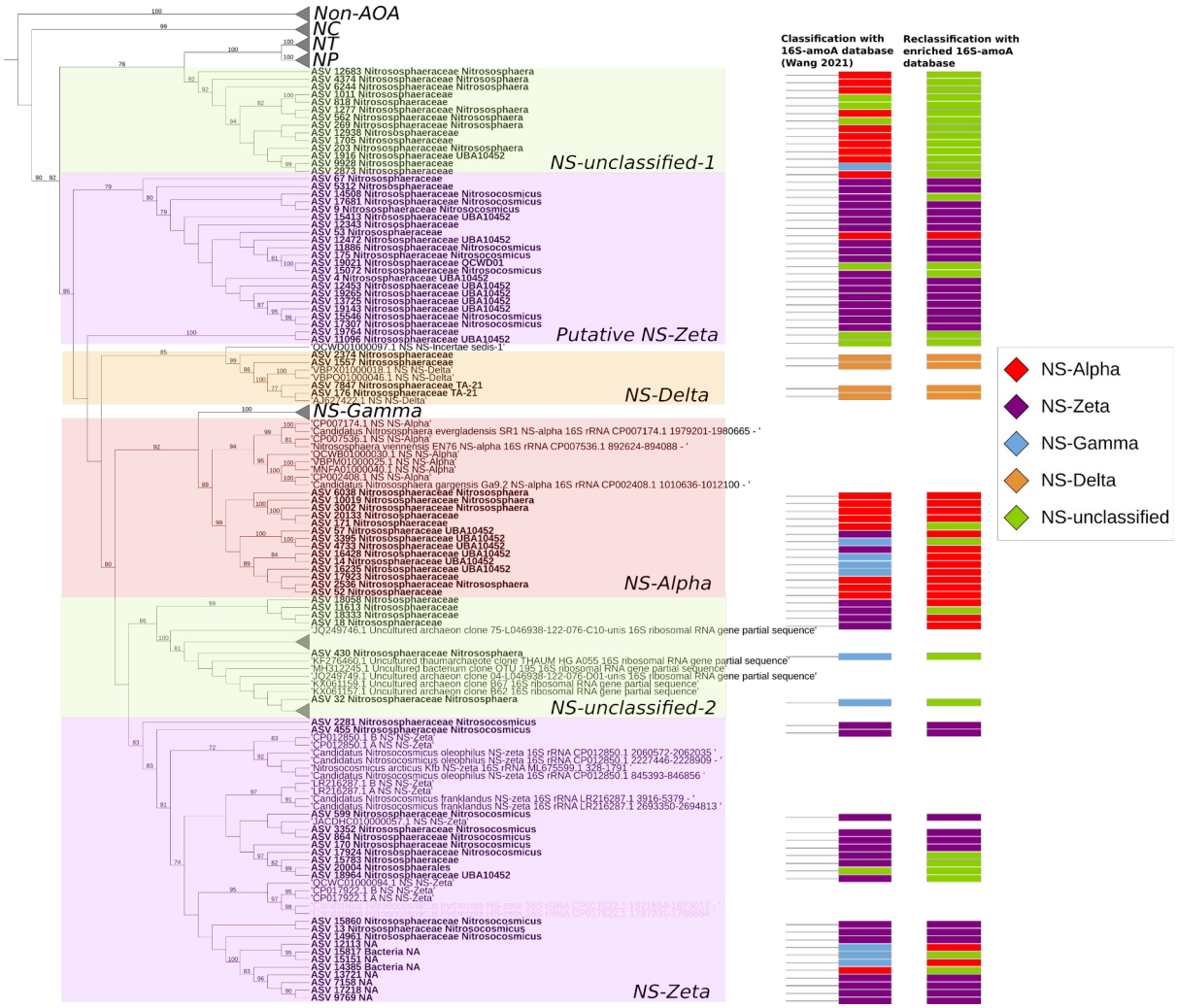


Figure 11: Phyolgenetic tree reconstruction based on a 16S rRNA full gene sequences backbone tree and 16S rRNA amplicons. The coloured areas in the tree represent the correspondent *amoA* clade based on the reference sequence from the backbone tree. The *amoA* clades assignment by the database is shown in two annotation columns on the right side of the tree: the first column corresponds to the first classification made with the original database found in Wang et al., 2021, and the second corresponds to the reclassification performed here with the “enriched” database and applying the minimum nucleotide sequence identity threshold described in the text.

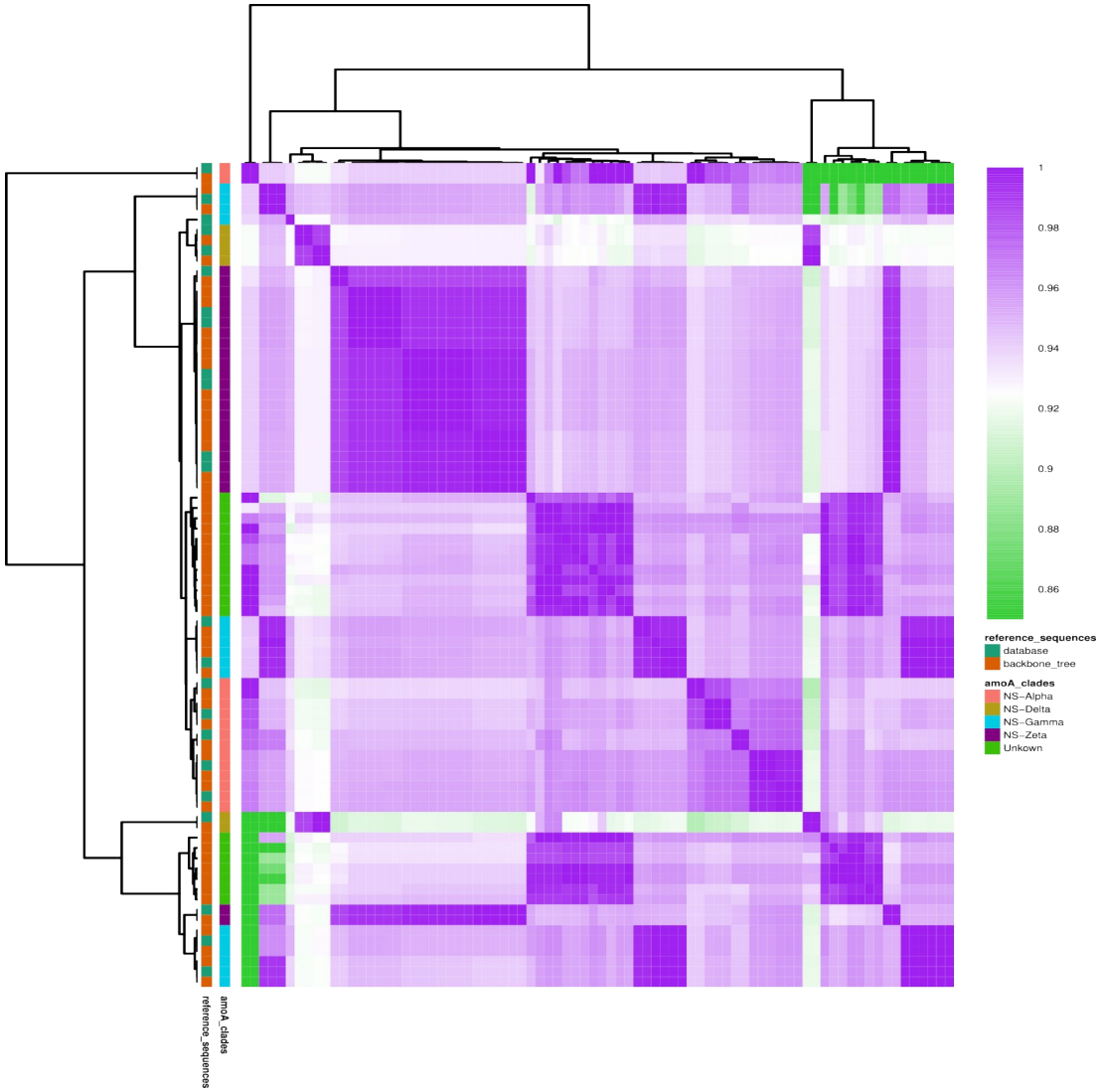


Figure 12: Heatmap with hierarchical clustering depicting the nucleotide sequence identity between the aligned sequences from the enriched database. Each sequence is colour coded with their *amoA* clade classification and their origin (i.e. the original 16S-*amoA* database or the backbone tree).

With the new classification we can observe that the *amoA* clade assignment and the phylogenetic tree information are more coherent. Most of the conflicting ASVs were reclassified as NS-Unclassified, with a few exceptions. We named the two branches containing only ASVs after the most commonly assigned *amoA* clade in each branch. Thus, the first branch is called NS-Unclassified-1, because all ASVs fall into the NS-unclassified category, while the second is Putative-NS-Zeta because most of their leaves were assigned to the NS-Zeta clade.

4.1.7. Composition and dynamics of the AOA community

The analysis on the AOA community was done based on the *amoA* clade assignment resulting from the process of re-classification described in the previous section. In Figure 13, the relative abundances from each clade are shown as a barplot, maintaining the proportions with respect to the total 16S rRNA amplicon reads. The most significant observation is that the structure and composition of the AOA community differs considerably between the two farms. As already noticed in the analysis from the general prokaryotic community, the proportion of AOA varies substantially in the BK farm, with some samples reaching almost a 20% of AOA in the total community (i.e. BKSPA-B) and some others barely reaching 1% (i.e. BKFSA-B). The NS-Zeta is the dominant clade in most samples from BK, but there is also a variable presence of NS-Alpha and NS-Unclassified. On the other hand, in the DWL the proportion of AOA is more stable between samples, with the exception of DWL2A-B, DWL10A-B and DWL11. In DWL, NS-Alpha is the dominant clade in a third of all DWL samples, while NS-Zeta dominates in only a couple of sites. It is noteworthy that in site DWL6, the clade NS-Delta represents 16-19% of the AOA community while it is virtually not found in any of the other sites.

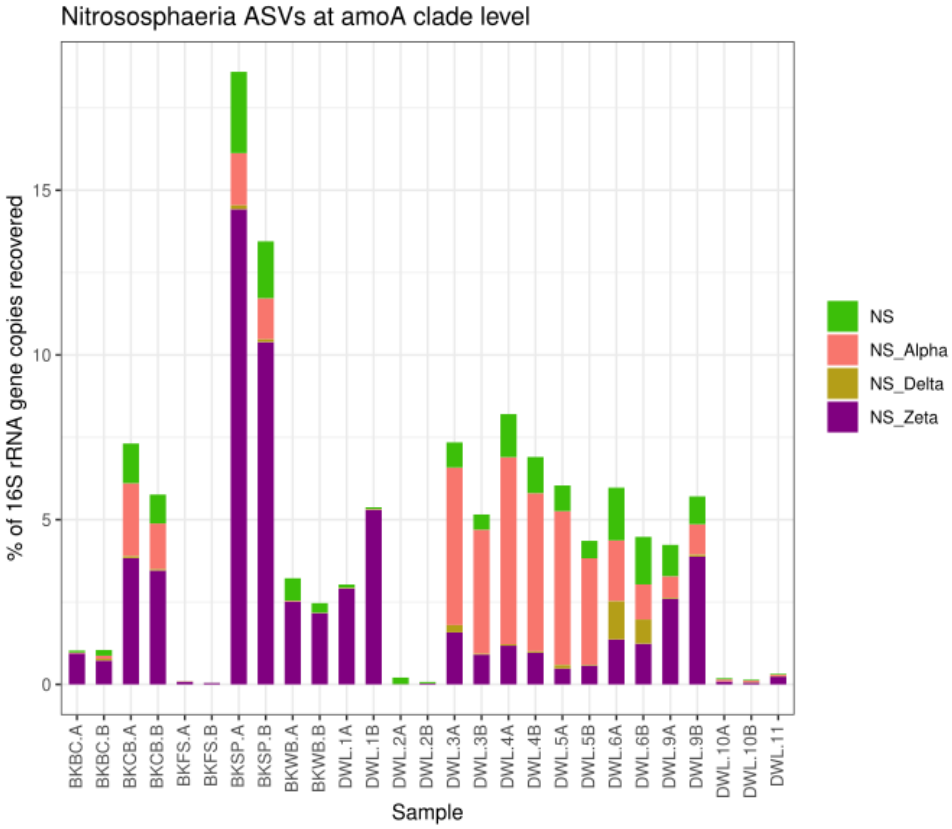


Figure 13: Barplot with the relative abundances of the 16S rRNA amplicons of the AOA members and their corresponding AOA clades, maintaining the original proportions relative to the total prokaryotic community.

If we break down the relative abundances of AOA into the ASV level, we see that the AOA community is composed by a few highly abundant ASVs and by a group of low abundant but more diverse ASVs, the latter forming about 25-30% of the AOA community (Figure 14). These more abundant members of the community are often common between sites and even in between the two farms (e.g. ASV 4). It is also interesting to see how the clades are not always dominated by the same ASV. For example, in the case of NS-zeta, the most abundant ASVs are ASV 4 and 9 in many samples, either in BK or in DWL. However, in samples BKBCA and DWL1A-B the NS-Zeta clade is almost totally populated by ASV 13.

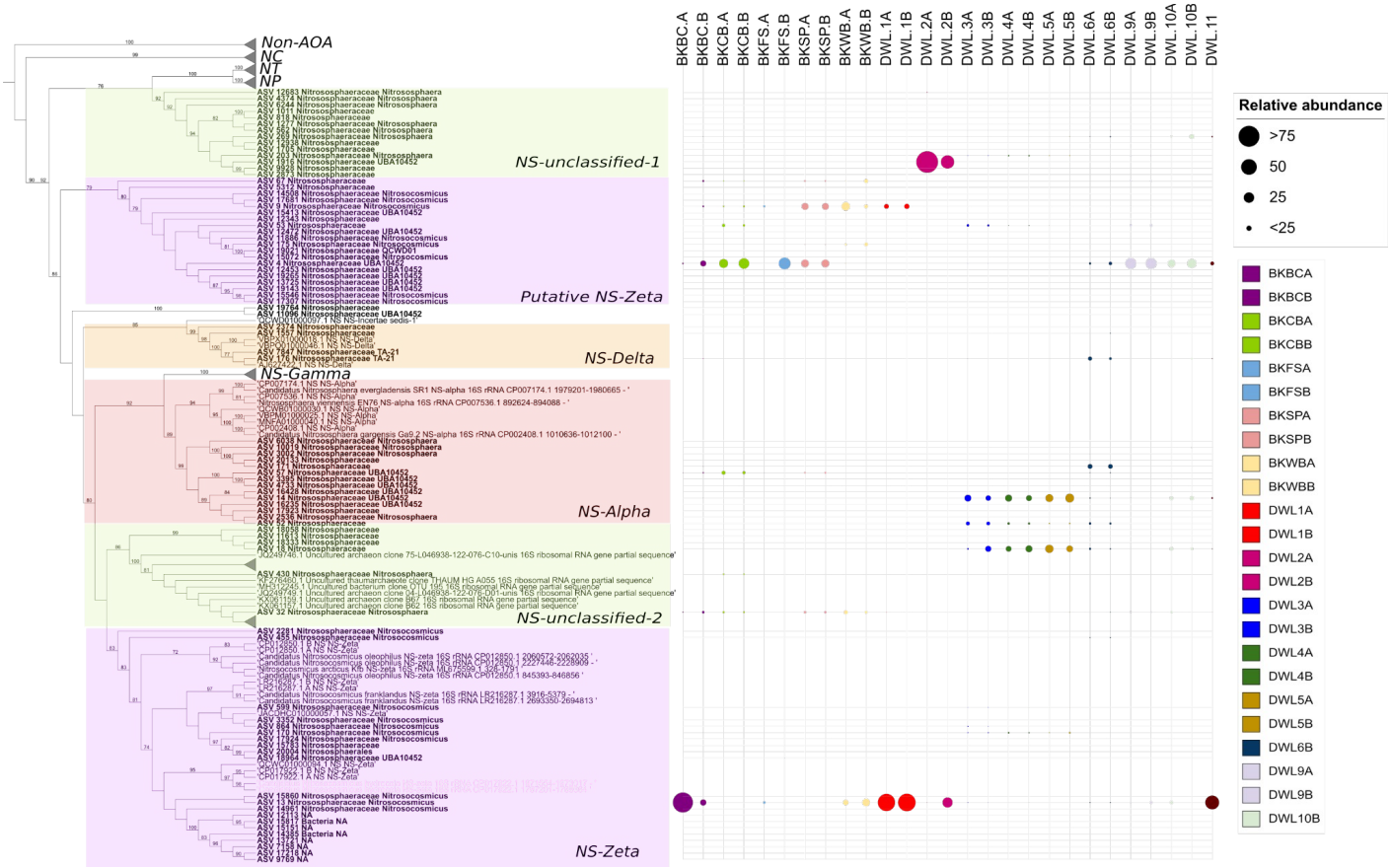


Figure 14: Phyolgenetic tree reconstruction based on a 16S rRNA full gene sequences backbone tree and the AOA 16S rRNA amplicons. The coloured areas in the tree represent the correspondent *amoA* clade based on the reference sequence from the backbone tree. The bubble plot represents the relative abundances for each *Nitrososphaeraceae* ASV. The color code indicates the sites.

4.1.8. Absolute abundances of the *amoA* clades:

Quantification of the archaeal and bacterial *amoA* genes revealed that the AOA community is more abundant in all samples (Figure 15). AOA numbers ranged from 4.89 x 10^6 to 2.92 x 10^9 *amoA* genes g−1 dry soil, with the highest abundances found in samples BKWBA and DWL4B (2.92 x 10^9 and 1.34 x 10^9 *amoA* genes g−1 dry soil, respectively), and the lowest DWL2B and BKFSB (4.89 x 10^6 and 7.11 x 10^6 *amoA* genes g−1 dry soil, respectively). On the other hand, AOB numbers ranged from 2.53 x 10^4 to 1.25 x 10^7 *amoA* genes g−1 dry soil.

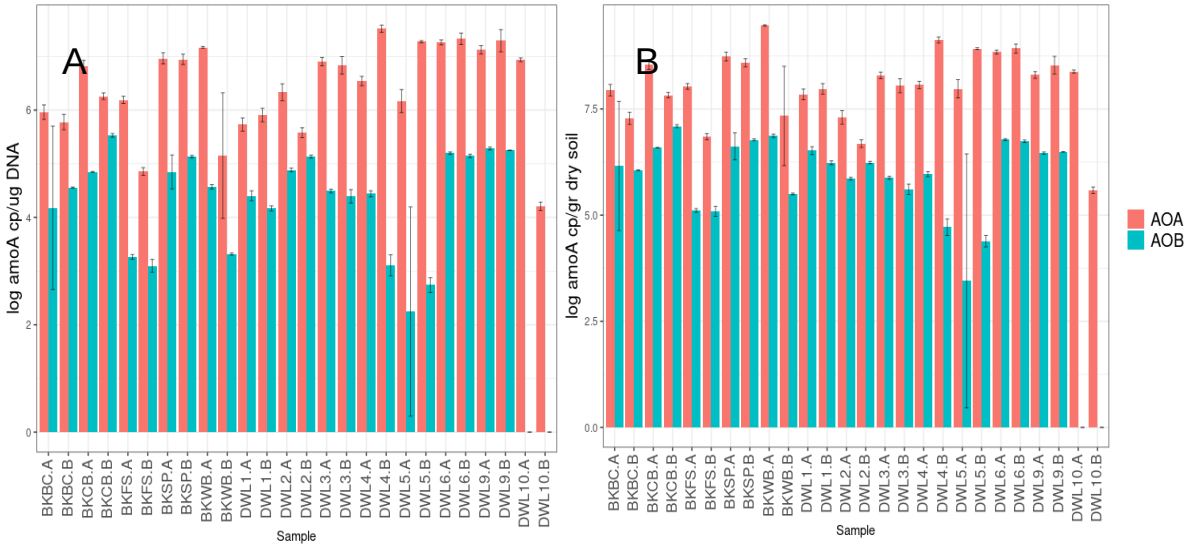


Figure 15: Barplots representing the log transformed copy number of archaeal (red) and bacterial (blue) *amoA* per µg DNA $^{-1}$ (A), and per g dry soil $^{-1}$ (B).

With the absolute abundances for AOA, we calculated the absolute abundances for each *amoA* clade by multiplying the clade relative abundances with the *amoA* counts obtained by quantitative PCR (Figure 16). With this data, we aimed to get a more realistic picture from the AOA community structure in the soil and to explore possible distribution patterns based on the available environmental data. From the overlay of the absolute abundance and the pH and SW values (Figure 16), we see that on some specific samples there seems to be a strong negative influence on the *amoA* gene abundance. This is the case for DWL2A and B, which are the most alkaline samples of the dataset, or for BKFSA-B that have the highest SW content. Nevertheless, we cannot talk about an overall and clear trend on the way these two factors influence the total AOA abundance. Correlation analysis with Spearman's rank between pH, SW and each separate *amoA* clade did show significant influences on certain clades (Table 5). NS-Alpha, NS-Delta and NS-unassigned appeared to be the most negatively affected by pH and SW content, particularly in samples from the DWL farm.

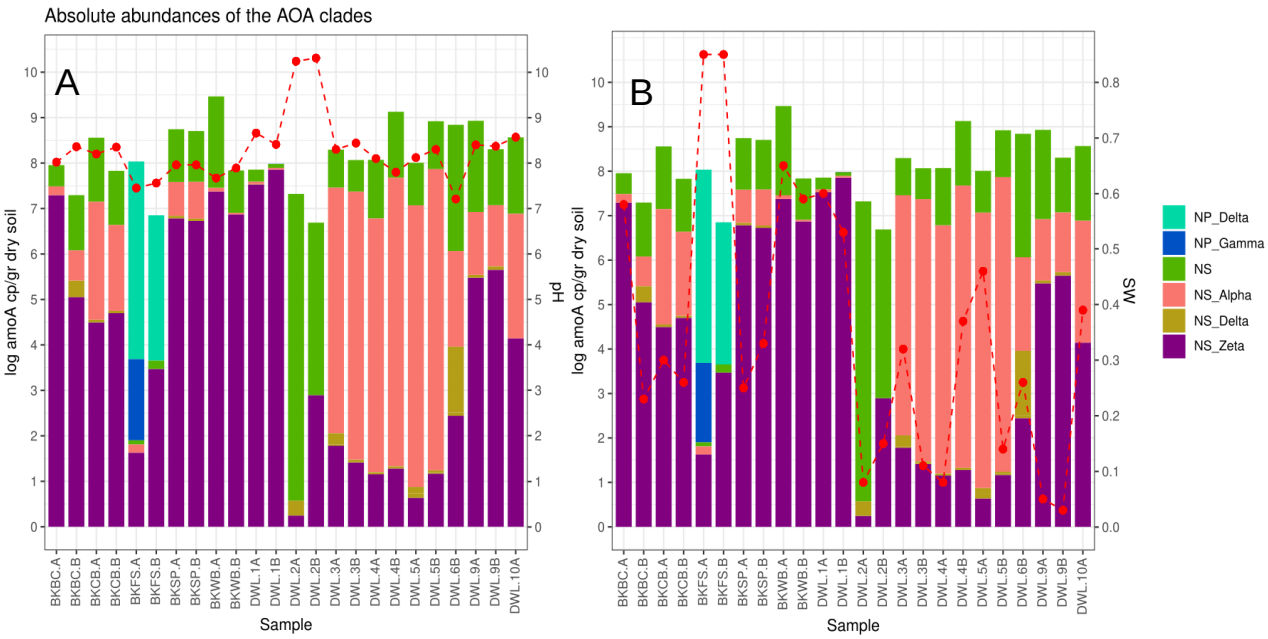
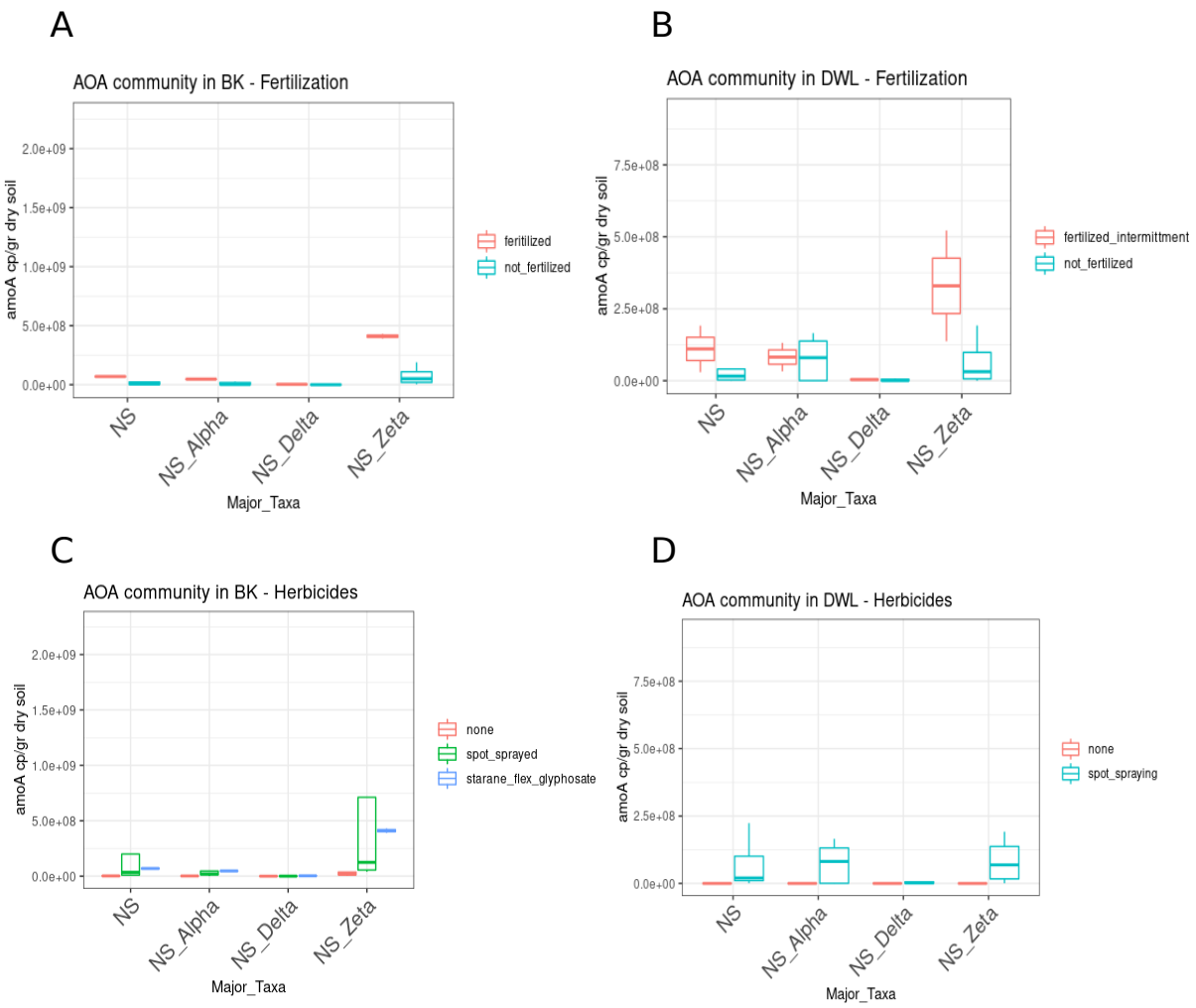


Figure 16: Barplots representing the log transformed copy number of archaeal *amoA* per µg per g dry soil $^{-1}$ broken down in clades with the pH (A) and SW (B) values on each sample.

Regarding the other environmental factors, we can observe trends of specific distributions in some cases (Figure 17). NS-Zeta is more abundant than other *amoA* clades in the land use categories describing some agricultural activity from both farms (Figure 17 G-H). It also appears to be more numerous than other clades with different types of irrigation regimes (Figure 17 E-F), and in fertilized samples compared to non fertilized samples (Figure 17 A-B). Regarding herbicides use, NS-Zeta reaches the highest abundance in the BK samples exposed to strong chemicals such as glyphosate and StaraneFlex (Figure 17 C-D). NS-Alpha exhibits a different trend to the NS-Zeta for some variables. The land use shows that NS-Alpha has not a clear preference for one of the categories in both farms, as NS-Zeta does (Figure 17 G-H). In the case of the irrigation, it seems to not thrive in heavily irrigated samples; on the contrary, it seems to prefer non irrigated environments in the DWL farm (Figure 17 E-F). For the fertilization regime, NS-Alpha does not thrive in fertilized environments as NS-Zeta does (Figure 17 A-B), in the same way that the herbicide presence does not enhance the NS-Alpha abundance as it seems to do for NS-Zeta (Figure C-D). The NS-unassigned, seems to follow the pattern from the NS-Zeta for most variables but with the average abundances lower than most NS-Zeta groups. Finally, in the case of NS-Delta it is difficult to observe any pattern, due to its low abundance across the dataset.

Spearman's rank correlations between pH/SW and amoA absolute abundances of each clade				
	pH		SW	
	Ben&Karli	Diamond_Lazy	Ben&Karli	Diamond_Lazy
NS-Zeta	0 (p-value = 1)	-0.31 (p-value = 0.28)	-0.16 (p-value = 0.65)	-0.12 (p-value = 0.69)
NS-Alpha	0.23 (p-value= 0.52)	-0.65 **(p-value = 0.01)**	-0.40 (p-value = 0.26)	-0.08 (p-value = 0.79)
NS-Delta	0.53 (p-value= 0.11)	-0.70 **(p-value = 0.005)**	-0.78 **(p-value = 0.07)**	-0.14 (p-value = 0.62)
NS-Unassigned	0.16 (p-value = 0.6628)	-0.48 **(p-value = 0.08)**	-0.34 (p-value = 0.34)	-0.42 (p-value = 0.13)

Table 4: Spearman's rank correlations between the pH and SW values and the absolute abundances of each AOA clade found in the two farm locations.



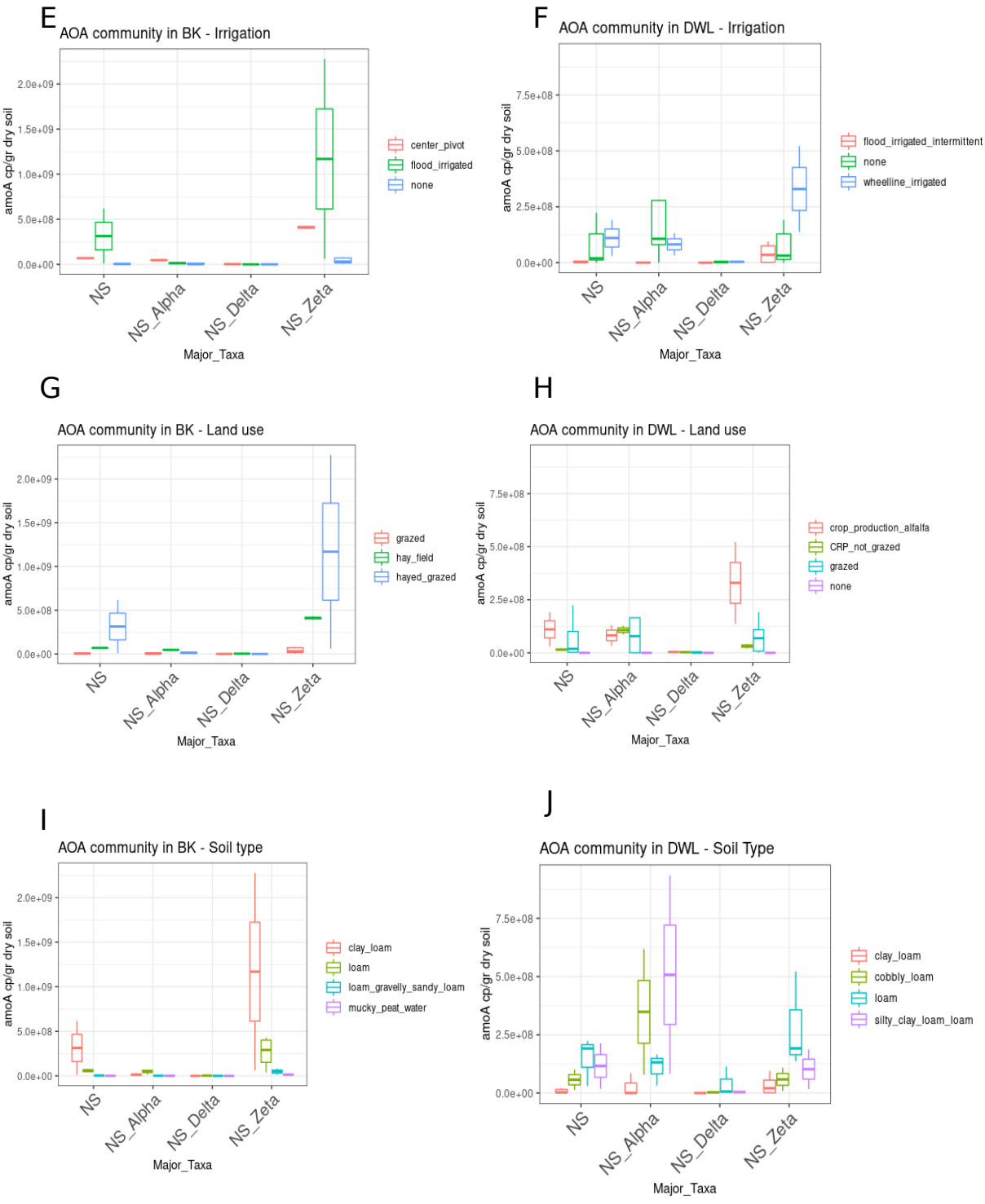


Figure 17: Boxplots comparing the absolute abundance of each AOA clade at each location between the different environmental factors: Fertilization (A, B), Herbicides (C, D), Irrigation (E, F), Land use (G, H) and Soil type (I, J).

4.2. AOA communities in the circumpolar arctic

4.2.1. Sites description

The typical climate from the areas sampled in the diverse arctic tundra locations included in the present study is characterized by a mean annual temperature below zero and a marked continentality (i.e. increased range of temperature variation and low mean precipitation values). The area of Tazovsky differs slightly to the others due to its exposure to the Arctic Sea, which influences the climate by softening the mean annual temperature and increasing precipitation. The soils in the tundra ecosystems are generally slightly acidic to neutral (pH range 5.0 - 7.0), with the pH increasing with depth. The soil water content is relatively high due to the low water evaporation rate caused by low temperatures, and due to the water accumulation thanks to the presence of the permafrost layer that stops the water drainage.

4.2.2. Attempt to generate *amoA* gene amplicons from arctic tundra soils

The archaeal *amoA* gene sequence is widely used to study the diversity and ecology of AOA, due to its lineage specificity and its highly conserved nature among the group. Initially, our goal was to realize an *amoA*-based study. Thus, prior to generating 16S rRNA gene amplicons, we attempted to generate *amoA* gene amplicons from the arctic tundra soils from all the locations and sites. Unfortunately, we did not manage to produce *amoA* gene amplicons via PCR. Despite all the efforts made to optimize the protocol, most of the resulting PCR products appeared to be smears, as depicted in the agarose gel image in Figure 18A. After multiple PCR trials, only three samples succeeded in the amplification of the *amoA* gene. Consequently, we decided to carry out the diversity analysis using a nested PCR approach based on the 16S-rRNA gene (see Materials and Methods), a method designed to detect very low abundant groups, previously described in Pausan et al., 2019. The end PCR products are shown in Figure 18B.

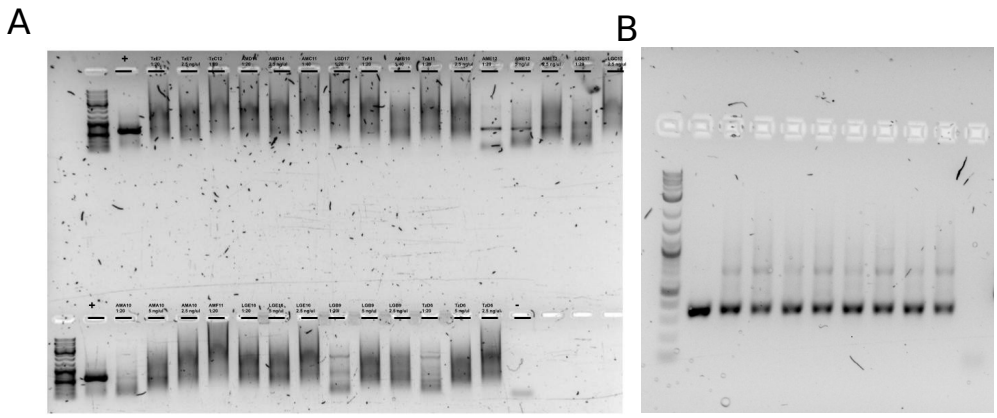


Figure 18: Agarose gel electrophoresis images with the PCR products obtained with the *amoA* primer pair C*amoA*-19F/T*amoA*-629R (A); second step of the nested-PCR with the primer pair Arch-519F/806R (B).

4.2.3. Sequencing results and taxonomical assignment

After the initial filtering and adaptors trimming process, the samples accounted for a total of 1,701,785 reads paired-end reads, with a median per sample of 115,485 reads. The median length per sample was 253 bp. The processing with DADA2 resulted in 950,188 amplicon reads, with a median of 62,403 reads per sample, from which 4,563 ASVs were recovered. The rarefaction curves (SI Fig. 2A), show high sequencing coverage in all samples. Moreover, a higher diversity level can be observed in samples AMP11 and ZKP7, with a number of observed ASVs around the 1500, while the rest of sample pools are under 500. From the obtained ASVs, 803 were successfully taxonomically assigned to Archaea and 3,744 to Bacteria, that is 17.6% and 82.1% from the total ASVs respectively.

4.2.4. AOA community analysis composition

The microbial community structure in the majority of the samples reveal an enrichment effect towards archaea, as a result of the nested PCR with archaeal 16S rRNA gene primers used to generate the amplicon sequences (Figure 19A). 78.52% is the median of counts corresponding to archaeal phyla, while 18.37% correspond to Bacteria phyla. It is interesting that the great majority of the diversity captured belongs to Bacteria and not to Archaea, but in terms of relative abundance the 16S rRNA gene community is dominated by archaeal reads. For most samples, the proportion of archaeal reads ranges between 60 and 98% from all classified reads, with the exception of samples AMP11 and ZKP7 that have 6.9 and 11.7% respectively, where most of the community consists of bacterial reads and additionally, harbor the highest number of observed ASVs (SI Figure 2).

The archaeal community is composed mainly by Thermoproteota and Nanoarchaeota, although the proportion of each group was variable between sites (Figure 19A). Thermoplasmatota and Methanobacteriota were the next most prevalent phyla, the first one being very abundant in samples ZKP6 and CHP5, and the latest in sample ZKP8. The bacterial fraction was dominated by Actinobacteria and Acidobacteria. In samples AMP11 and ZKP7, Proteobacteria and Chloroflexota are also two abundant phyla. The taxonomic distribution between and within locations appears to vary at phylum level. Zackenberg sites show the most dissimilar composition between them and also compared to the other locations. Conversely, the Cherskii sites seem to have a more homogeneous community. The Taymyr and Tazowvsky sites composition resembles the Cherskii sites, with the exception of sample AMP11 and sample TZP15.

The ASVs classified in class Nitrososphaeria were subselected and run against the 16S-*amoA* database to obtain the AOA clade classification. 64 Nitrososphaeria ASVs were successfully classified into one of the known AOA clades from a total of 169 ASVs in the subset, while the rest resulted to be unclassifiable by the database. The detected AOA clades were NS-Zeta, NS-Gamma, NS-Delta and NS-Alpha (Figure 19B). The 16S-*amoA* database classification method offers theoretically a resolution down to the *amoA* subclade level. In our case, NS-Zeta ASVs were all classified as NS-Zeta-2, the NS-Gamma as NS-Gamma-2.3.2-OTU-2, the NS-Delta as NS-Delta-1 and the NS-Alpha as NS-Alpha-3.2.1.1.1.

Additionally, we reconstructed a 16S rRNA gene-based maximum likelihood phylogenetic tree using full 16S rRNA gene sequences from known AOA and the 30 most abundant Nitrososphaeria ASVs in our dataset (Figure 20). The unclassified sequences clustered within the sequences of non ammonia oxidizing archaea (non-AOA), a sister group to the family *Nitrososphaeraceae* which had been included in the tree as an outgroup (Figure 20). Thus, we deduced that the unclassified amplicon sequences by the 16S-*amoA* database corresponded to the non AOA group.

The barplot in Figure 19B shows the composition from the Nitrososphaeria community broken down into AOA clades; the relative abundances shown are based on the total reads. Firstly, one can see that the Thermoproteota proportion in Figure 19A, is equivalent to the Nitrososphaeria proportion depicted in Figure 19B; thus, the majority of Thermoproteota sequences belong to the class Nitrososphaeria. The AOA community is dominated at each location by the terrestrial AOA clades NS-Zeta and NS-Gamma. When co-occurring, one clade is always more abundant than the other. Additionally, NS-Delta, although much less

abundant, is also prevalent in all samples. Non-AOA are also present in most locations and in some cases are the most abundant group of the Nitrososphaeria community. The community structure changes between locations (i.e. the proportion of the AOA clades varies between locations). For example, in the Cherskii sites the main AOA group is NS-Gamma, while in Taymyr NS-Zeta is the dominant. Furthermore, the structure is not homogeneous within each location. On the contrary, we see differences in the AOA communities between sites from the same location. In fact, in the case of Tazowvky and Zackenberg every site appears to be differently composed.

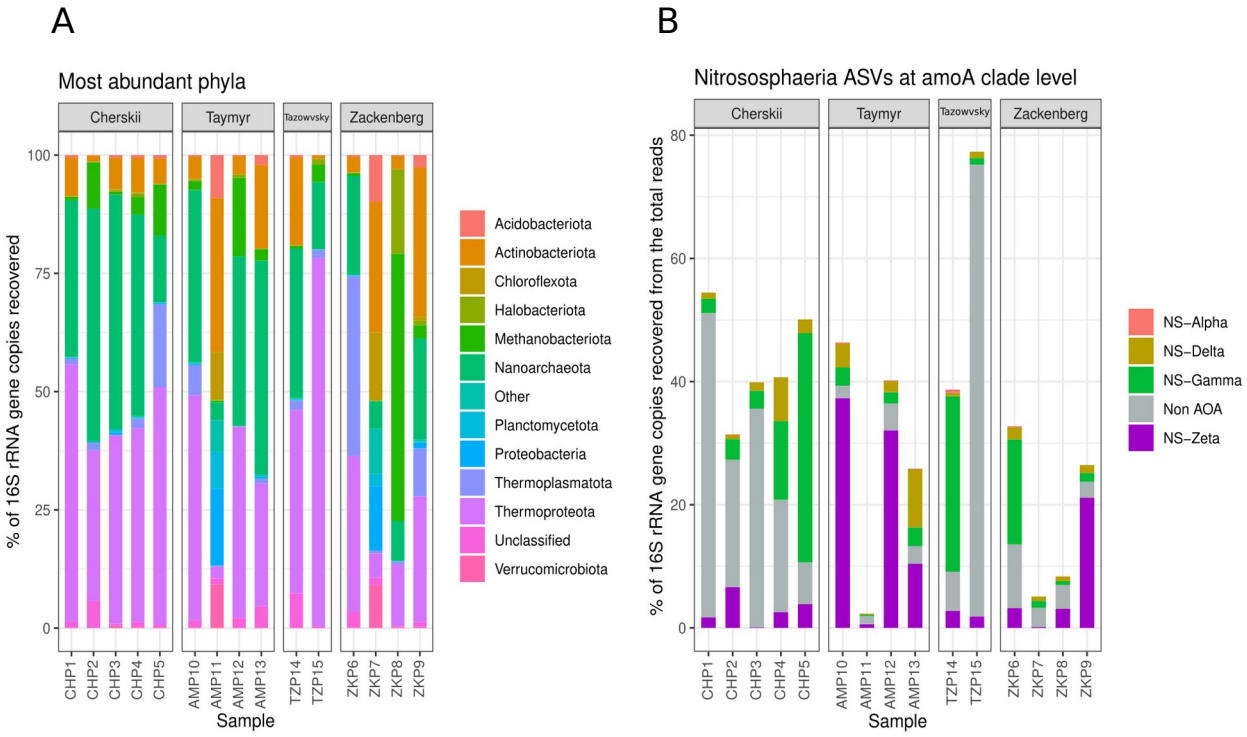


Figure 19: Barplots (A) with relative abundances of the taxonomical classification at a phylum level of the total 16S rRNA genes reads, and (B), with the relative abundance of Nitrososphaeria ASVs with the AOA clade classification, maintaining the proportions with respect to the total reads obtained from the amplicon sequencing (on average 113000 reads/sample). The amplicons were obtained with a nested PCR: firstly with a 16S rRNA primer pair with high archaeal coverage (344F-1041R) and secondly with a universal 16S rRNA (519F-806R).

In order to take a closer look into the composition and structure of the AOA community, we selected the same 30 most abundant Nitrososphaeria ASVs (representing >90% of the total Nitrososphaeria counts) used to build the phylogenetic tree, and produced a bubble plot with their relative abundances. Figure 20 depicts the merged output of both analyses. Additionally, we added the AOA clade classification obtained by the 16S-*amoA* database to show the correspondence between this result and the AOA clades inferred from the phylogenetic tree.

The AOA community has a low diversity within sites: for most locations, the AOA community is composed of a few very abundant ASVs (Figure 20). One of the interesting findings in this study is that the most abundant ASVs are often common between locations, regardless of the geographical distance between them. This is noticeable in the case of NS-Gamma ASV-1, as it is present across sites from the four locations, being at times the main AOA community member (e.g. CHP5, TZP14 or ZKP6). This trend is also observed in the case of ASV-9 and ASV-5, both in the NS-Zeta clade, although they are not as abundant as ASV-1. In the non-AOA community, some of the most abundant ASVs are also common in between locations. Nevertheless, it appears that the non-AOA populations on each location are more diverse than the AOA populations, consisting often of several ASVs.

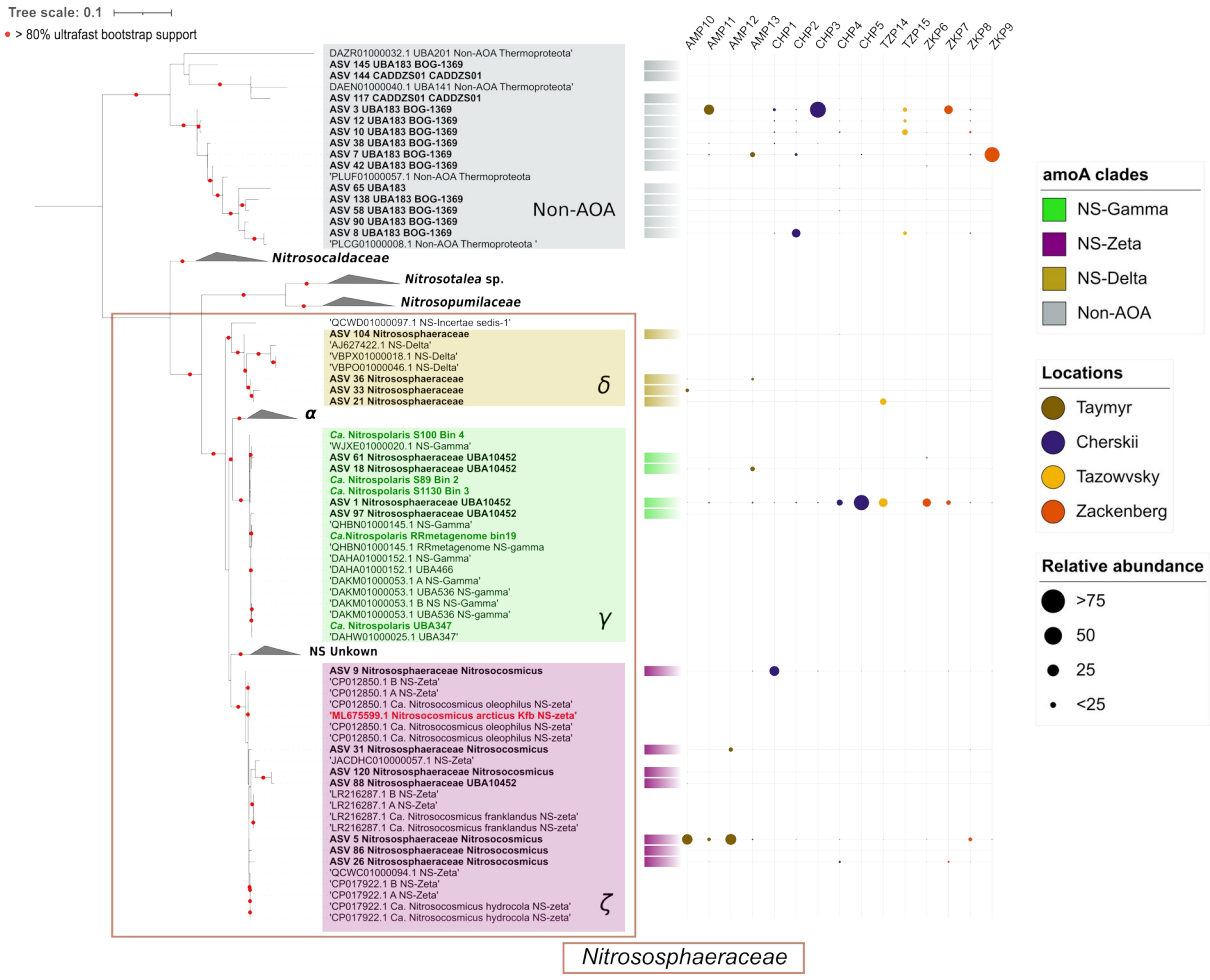


Figure 20: Phylogenetic tree reconstruction of Nitrososphaeria based on full 16S rRNA gene sequences as backbone and with the most abundant ASVs in the arctic dataset added (labeled in bold, >90% of total read counts). The classification of the ASVs into the *amoA* clades was additionally done with the 16S-*amoA* database (Wang et al. 2021). The relative abundance of the most abundant ASVs included in the tree are represented in the bubble plot, color coded by locations. The ASVs labels are marked in bold letters; the cultivated representative *Ca.* Nitrosocosmicus arcticus is marked in red; and the sequences from the putative new genus *Ca.* Nitrosopolaris are marked in green.

4.2.5. Analysis of the distribution of the AOA clades in the arctic tundra ecosystem

Despite the AOA communities in the arctic tundra being composed of only a few members, there seem to be different distribution patterns among these dominant groups. We wanted to see if the differences in AOA clade distribution could be due to environmental particularities found at each location, different vegetation types or soil horizons. To shed some light into which environmental factors could potentially be driving community structure on each site, we gathered metadata collected during the CryoCarb project summarized in SI Table 1 and 2, and published in several papers (listed here in the Material & Methods section). As described in the M&M section, we selected a set of 15 relevant environmental variables from all samples that were used within each pool. All grouped data was averaged and the result was used for further analysis.

To examine the potential relationship between these environmental variables and the AOA relative abundances, we built 2 different CCA models: one with the soil physico-chemical parameters (CCA-1, Figure 21A) and one with the enzymatic activity data (CCA-2, Figure 21B). Both explain a significant proportion of the variability present in the AOA community among the different sites (58.8% for model 1 and 68.13% in model 2, ANOVA p-value < 0.05). CCA-1 distinguishes that the three major clades in arctic soils show clearly distinct patterns based on the soil physico-chemical parameters included in the model. The non-AOA seem to be more abundant in sites richer in organic carbon and nitrogen, with higher soil water and lower pH values. The two main AOA clades tend to correlate with higher nitrate (NS-Gamma) and higher pH values (NS-Zeta). Furthermore, it seems that certain clades have a preference for certain vegetation types, as NS-Gamma is the main AOA clade in most lichen tundra sites and non-AOA are dominant in tussock tundra sites.

CCA-2 indicates distinctly that non-AOA dominated sites correspond with higher enzymatic activities of organic matter degradation, while the main two AOA clades seem to be less influenced, in the case of NS-Zeta, or even negatively impacted by the effect of catabolic enzyme activities, in the case of NS-Gamma. It is interesting the evident separation between NS-Gamma and NS-Zeta, suggesting contrasting reactions to the environmental variables by the different AOA clades.

To further explore possible environmental drivers of the AOA distribution in our dataset, we performed Spearman's rank correlations between the soil parameters or the enzymatic

activity data and the relative abundance of each AOA clade. The results, represented in Table 5, generally agreed on the trend observed in the CCA models regarding the AOA vs. non-AOA. Among soil parameters, Non-AOA correlate positively with SW, OC content, NH4+ and total N content, although statistically non-significantly, and negatively with pH, significantly, and with NH3-, non-significantly. Regarding enzymatic activities, non-AOA correlate positively with all the ones included in model 2, with the correlation with peroxidase being the only statistically significant. In the AOA groups, pH correlates significantly positively with the NS-Zeta and NS-Delta, while it does not seem to have an effect on the NS-Gamma abundance. No significant correlations were found with the other soil parameters, and the correlation coefficient values seem to be more negative and/or neutral than in the non-AOA. The correlations between the enzymatic activities and the AOA relative abundances seem to be in general more negative than correlations with the non-AOA, with some exceptions (not significant though): NS-Zeta exhibits a positive correlation with the endochitinase and actual protease activities and NS-Gamma with the peroxidase activity. The only significant correlation with an enzymatic activity was an observed negative correlation between NS-Gamma and endochitinase values. There is no clear pattern distinguishable between AOA groups in the correlations calculated here.

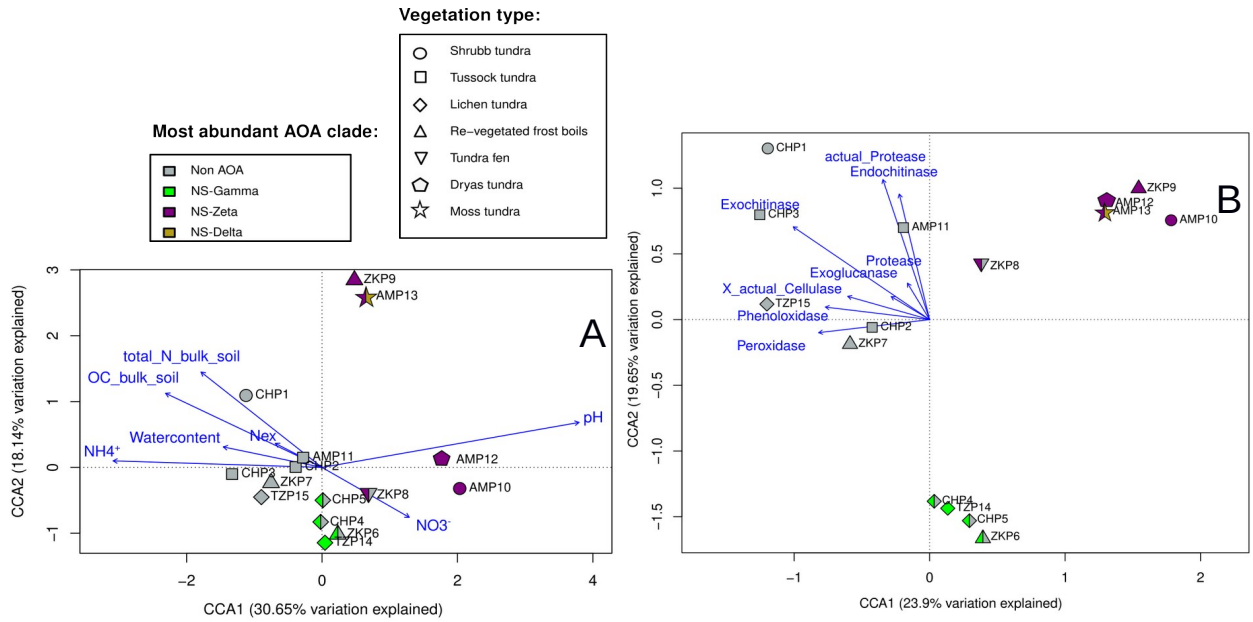


Figure 21: Relationship between AOA distribution and environmental parameters. CCA biplot of Nitrososphaeria relative abundances on each sampling site and selection of soil physico-chemical properties in model 1 (A), and enzymatic activities in model 2 (B). The environmental parameters are represented by the blue arrows. Symbols represent the main vegetation type on each site. Colors represent the most abundant *amoA* clade found on each site.

	NS-Zeta	NS-Gamma	NS-Delta	Non-AOA
pH	0.63 *	0.10	0.68 *	-0.55 *
SW	-0.13	-0.16	-0.24	0.39
OC_bulk_soil	-0.16	-0.18	-0.40	0.40
total_N_bulk	0.03	-0.16	-0.23	0.29
NH4	-0.36	0.06	-0.35	0.40
NO3	0.15	-0.22	-0.09	-0.24
Nex	0.16	0.11	-0.28	-0.03
Exoglucanase	0.00	-0.17	-0.23	0.06
Endochitinase	0.16	-0.67 *	-0.29	0.13
Exochitinase	-0.24	-0.35	-0.39	0.43
Protease	-0.03	0.01	0.04	0.28
Peroxidase	-0.45	0.19	-0.21	0.57 *
actual_cellulase	-0.03	-0.41	-0.30	0.27
actual_protease	0.21	-0.42	-0.19	0.11

Table 5: Spearman's rank correlations between different soil parameters and enzymatic activities and relative abundance of the main Nitrososphaeria groups in arctic tundra soils. Asterisks indicate the significance of the correlations (*, P-value < 0.05). The color code indicates whether the coefficient is positive (green) or negative (red).

5. DISCUSSION

5.1 - The AOA community composition varies in different ecosystems

5.1.1 - The farms soils harbor rich and abundant AOA communities

The general prokaryotic community from both ranches was composed of phyla typically found in terrestrial environments, such as Proteobacteria, Actinobacteria, Acidobacteria or Firmicutes (Banerjee et al., 2018). However, the community structure differed between the samples from the two locations, as we saw with the beta diversity analysis carried out with all the farm samples together. The BK dataset has more sites which are dominated by Proteobacteria while in the DWL ranch there are more sites dominated by Actinobacteria (Fig 2). Although this trend could be due to the different physicochemical properties found at each location, we observed that regardless of the location, the sites with higher water content (Table 1) often had a higher proportion of Proteobacteria over Actinobacteria. The most common Proteobacteria order in the samples with high water content was Burkholderiales (Fig 3). This group was rather diverse, but some of the most abundant genus were *Sideroxydans* sp., a microaerophilic iron oxidizing bacteria (Cooper et al., 2020) and *Rhodoferax* sp., a genus commonly found in freshwater systems (Finneran et al., 2003; Li et al., 2020) that has been associated with the degradation of several contaminants (Jeon et al., 2003; Kasanke et al., 2019). The Actinobacteria community was composed mainly by *Rubrobacter* sp., a genus described to be adapted to extreme environments, specially to high temperatures and high γ-radiation (e.g. volcanic soils (Norman et al., 2017), Atacama desert (Castro et al., 2019) but that also has been isolated from agricultural and pasture soils (Sait et al., 2002; Singleton et al., 2003); and *Solirubrobacter* sp., a ubiquitous group in agroecosystems that has been seen to be more abundant in soils with low organic carbon and under biochar amendments, which is associated with the capacity of degrading more complex organic molecules (Nielsen et al., 2014; Shange et al., 2012; Zhang et al., 2019). Acidobacteria can be very abundant members of soils communities, particularly in native soils over agricultural soils, and are seen as a soil health indicator group (Trivedi et al., 2016). However, here they did not reach 10% of the relative abundance in any of the samples, possibly because the pH values in all samples are neutral to alkaline.

Both Betapreobacteria and Actinobacteria are often associated with copiotrophic lifestyles

(i.e. they thrive in nutrient-rich environments) (Fierer et al., 2012;), however they can be also found in oligotrophic environments (Zhang et al., 2019). Indeed, there is a high variation within each of these phyla, which underlines the importance of analyzing their ecology at lower taxonomic levels. The oligotrophic or copiotrophic character of a microbial group can be defined by its preference for utilize labile organic C and have rapid growth rates under high nutrient availability, in the case of copiotrophs (i.e. *r* strategists), or by having lower growth rates and thriving under lower nutrient availability, in the case of oligotrophs (i.e. *K*-strategists) (Fierer et al., 2007). Still, as discussed in (Ho et al., 2017), the relative abundances of the microbial groups are influenced by multiple factors besides their microbial life strategies, such as the soil type or the availability of micronutrients. In addition, the physiological traits of a microbial group are inferred based on the few available strains in cultures, which is not representative of the vast majority of microorganisms present in the environment. Thus, the interpretation of the potential function of a microbial group based on its presence and relative abundance has to be done carefully.

Some of the limitations of the Montana soils dataset are the lack of control samples for each factor, which do not allow for the observation of the effect of a single factor, or the fact that the sample distribution across all treatments/categories is imbalanced (e.g. 23 unfertilized samples vs. 4 fertilized samples). Furthermore, the sampling was done in duplicates instead of triplicates. This is particularly relevant in terrestrial ecosystems, where homogenization of the soil is limited, as compared to other environments such as the ocean, and where nearby spots belonging to the same site can be significantly different from each other. This can be seen in the fluctuation of the alpha diversity results between duplicates (Figure 4). Taking into consideration all the aforementioned limitations, our analysis points at the pH and the SW content as relevant factors shaping the microbial diversity and structure (Figure 4, Figure 6), as it has been shown in previous studies (Drenovsky et al., 2004; Lauber et al., 2009). The irrigation regime, although not directly linked to the SW content, and the soil type seem also to be influencing the community structure (Figure 5, Table 3).

The relative abundance of the total AOA community varied substantially across the dataset (Figure 9). In terrestrial environments, AOA range between 1 and 5% (Leininger et al., 2006), which is the proportion detected in most samples. Surprisingly, in samples BKSPA-B up to 20% of the total community were AOA. In this site traditional agricultural activities are practiced, with the application of chemical fertilizers and herbicides, such as glyphosate. In BKCBA-B, which has the same soil characteristics as BKSP but the site has been undisturbed, the AOA relative abundance was about less than half that in BKSPA-B.

Between the two farms, the AOA proportion seems to be more stable in the DWL soils, than in the BK soils. Samples with the lowest AOA relative abundance are those which also have the highest proportion in Burkholderiales (BKBC, BKFS, DWL10) (Figure 3), while on the other hand abundant AOA population seem to coexist with higher proportions of the Actinobacteria community. As mentioned earlier, these samples seem to differ by their soil water content. AOA are aerobic organisms, as they need oxygen for the ammonia oxidation (Stein, 2019). Thus, we would expect that in water saturated soils their frequency is lower, due to less oxygen availability.

The *amoA* absolute abundance detected by qPCR (Figure 11) also is within the expected range for terrestrial environments (Leininger et al., 2006). Comparing the amplicon sequencing relative abundance with the absolute abundances we see that often they do not correspond with each other. For instance, AOA relative abundance in samples BKFSA-B and DWL 10A-B was almost undetectable but the absolute abundance reached up to 10^7 *amoA* cp/g dry soil. Still, the difference in abundance between BKSPA-B and BKCBA-B was roughly maintained, with BKSP being one of the most abundant sites in AOA. These discrepancies between the relative and absolute abundance data could be explained for instance if that particular sample is super densely populated by a diversity of microbial groups (AOA included), so the absolute count on AOA is high, but that the proportion of AOA respect to the rest of the community is low. The difference in the primer efficiencies should be also considered when comparing these two outputs, since two different primer pairs were used for each method. However, the use of both the amplicon sequencing and the quantification by qPCR offers a more complete and realistic picture of a microbial community.

Most of the diversity studies on AOA are based on the *amoA* gene. Despite all the advantages that come with using *amoA* sequences, they are missing information on the AOA community in the context of the total prokaryotic community. However, by using universal 16S rRNA primers and the *amoA* clade classification (Wang et al., 2021), the best of both worlds can be obtained. As we have seen in this study, AOA can represent an unusually high proportion of the total community. Further analysis could entail co-occurrence microbial network analysis to gain more insight into the possible interactions of AOA clades with other microbial groups.

The quantification of the AOB results agree with the generalized trend of AOA outnumbering AOB in most terrestrial environments (Leininger et al., 2006; Nicol et al., 2008). Again here, we found a disparity between the relative abundance data, which shows AOB to be under

1% of the prokaryotic community in all of the samples, and the absolute abundance data, which detects AOB to be on average around 10^5 *amoA* cp/g dry soil. Besides a primer bias effect of the EMP unfavorable to the AOB, another possible explanation could be that AOB belonging to the Betaproteobacteria usually have 2 to 3 copies of *amoA* per cell (Norton et al., 2002), while AOA are known to have one copy of the *amoA* per genome (L. Liu et al., 2021; Spang et al., 2012; K. V. Zhalnina et al., 2014). From our data, we can conclude that the community of nitrifiers is indeed dominated by AOA. However, since no measurements on the nitrification activity of the soils in this study were taken, we do not know how active the nitrifiers are and if there are differences in the activity among each group. There is no clear relationship between the abundance of the nitrification genes (i.e. *amoA* gene) and their expression (Prosser & Nicol, 2008), so the *amoA* abundance should not be taken as a proxy for nitrification activity. Several studies have seen that the archaeal community can be responsible for a large part of the total nitrification activity (Nicol et al., 2008; Tourna et al., 2008, Huang et al., 2021). Generally, it has been observed that AOA drive nitrification under lower ammonia concentrations and with organic fertilizers amendments (Guo et al., 2017; Hink et al., 2018), while AOB activity increases upon the addition of high amounts of chemical N-fertilizers and in overfertilized agricultural soils (Meinhardt et al., 2018; Sterngren et al., 2015).

As expected most of the AOA lineages that we detected belong to the so-called terrestrial AOA clades, grouped under the Nitrososphaeraceae family. However, NS-Delta and NS-Gamma, which are the most abundant terrestrial clades detected in environmental surveys and yet unculturable (Alves et al., 2018; Huang et al., 2021; Treusch et al., 2005), were either present in low abundance, in the case of NS-Delta, or undetectable, in the case of NS-Gamma. The AOA community varied its composition and structure between locations. The high proportion of NS-Zeta in the BK farm sites was particularly surprising, as it was seen by (Alves et al., 2018) that its *amoA* sequences represent around 1% in the databases, while NS-Delta and NS-Gamma *amoA* sequences represent a 23% and 13%, respectively. Still, NS-Zeta strains have been isolated from agricultural soils (Lehtovirta-Morley et al., 2016; L. Liu et al., 2021). This lineage is characterized by its tolerance for high ammonia concentrations and higher pH values than its NS sister lineages (Alves et al., 2018; M. Y. Jung et al., 2022; L. Liu et al., 2021). These findings align with our results, as we found the highest NS-Zeta proportion in fertilized site BKSP. On the other hand, NS-Alpha was more common in sites from DWL ranch. This clade is known to inhabit soils with a wide range of pH values, to be more sensitive to ammonia concentration due to the high substrate affinity of the AMO complex (M. Y. Jung et al., 2022; Kim et al., 2012; Stieglmeier et al., 2014) and to be adaptable to disturbances, such as high salinity concentrations (Sun et al., 2021). In a

more detailed analysis on how each clade population was structured on an ASV level (Figure 10), we observed that sometimes the most abundant members of the clade differed between sites. Unfortunately, the 16S-*amoA* database did not allow us to distinguish between subclades; to obtain a stronger taxonomic and phylogenetic resolution we would need to obtain the *amoA* gene sequences. We hypothesize that the clade populations consist of different subclades that have preference for certain sites, possibly due to adaptations to the specific environmental factors from the place. It has been reported that subclades can have specific adaptations to environmental parameters, such as pH (Alves et al., 2018; Zhao et al., 2022). In the following sections, we will elaborate on our predictions of the environmental distribution for each clade based on our available environmental and context data.

5.1.2 - The AOA community is similar across the circumpolar tundra soils

The approach that we used to analyze the AOA community in the arctic tundra soils was chosen for numerous reasons. Firstly, we failed in our attempt of generating *amoA* gene amplicons with the tundra soils, probably due to a combination of factors (Figure 14). One of the main reasons was the natural presence of PCR inhibitors in arctic soils (e.g. humic and fulvic acids) that are co-extracted with the DNA. To counteract the inhibitors effect, typically there are two methods: using specific purification kits for the PCR-inhibitors removal and using dilutions of the DNA extract in the PCR (Siljanen et al., 2019). The latter method is usually effective, but in our case we were also handling samples with low DNA concentrations that additionally, had been reported to have very low relative abundance of archaea (Gittel et al., 2014a; Gittel et al., 2014b). For all these reasons, we considered the nested-PCR method designed by (Pausan et al., 2019), together with the pooling of the samples from the same site and vegetation type, to be the most appropriate strategy to explore the AOA diversity in the arctic tundra. Evidently, this approach only enabled us to obtain a qualitative diversity assessment of the archaeal community. Still, for the purpose of this study, which was the identification of the main AOA clades in the circumpolar tundra soil ecosystems, this method was adequate.

Most of the pools belonged to topsoil horizons, which despite the observation of specific site composition patterns (Gittel et al., 2014a), they are mainly dominated by Proteobacteria (Alpha, Delta and Gamma), Acidobacteria, and Actinobacteria. The topsoils are in direct interaction with the vegetation cover and the fungal community associated with them, which implies that sources rich in C and N are more available to the prokaryotic community (Talbot al., 2008). In their studies, Gittel and colleagues saw that the archaeal community

represented less than 1% of the total prokaryotic reads in all sites and was mostly dominated by AOA and methanogens. Members of the phyla Halobacteria and Thermoplasmata were also detected. Generally, in our results we also detected these archaeal groups, although with different relative abundances, because after all, our community results are a magnified and most likely deformed picture of the actual archaeal community composition. We also detected in most samples the phylum Nanoarchaeota that is known for its very small size and their symbiotic lifestyle (Huber et al., 2003) and has been found in arctic permafrost soils (Shcherbakova et al., 2016). Tveit and colleagues 2013 (Tveit et al., 2013) performed a study in which they looked at the genetic potential of the microbial communities in peat arctic soils for the degradation of organic matter. They found that arctic communities encoded in their genomes an enzymatic battery for the degradation of plant material comparable to those found in temperate and tropical soil prokaryotic communities. Being the only ammonia-oxidizing organisms detected in many arctic soils (Siljanen et al., 2019), AOA become a key group in N-cycling. Since N is an elemental nutrient for all organisms and its availability is usually limited, the activity of AOA has a direct impact on controlling the availability of N for the rest of community members and therefore, can influence the organic carbon transformations occurring in the soils. Furthermore, with the rising temperatures causing the thawing of the permafrost, there is an increasing risk of ammonia being released to the soils inducing the activity of AOA, which would cause higher N_2O and CO_2 emissions.

The AOA community was mainly composed of the clades NS-Zeta and NS-Gamma, as was previously reported in (Alves et al., 2013; Siljanen et al., 2019). The alternating presence of one or the other clade in the different sites, led us to hypothesize about a different distribution pattern for each clade. As it has been already discussed, there is enough evidence that supports diverse specific adaptations to each clade to certain environmental contexts. Recently, (Pessi et al., 2022) have proposed a new genus, named as *Ca.* Nitrosopolaris, found in polar soils based on the recovery of several MAGs along the polar climate zone. This new genus classifies within the NS-Gamma clade and it appears to be adapted to highly oligotrophic environments. Contrary to our hypothesis, (Siljanen et al., 2019) found that NS-Zeta and NS-Gamma co-occurred in the same samples with similar relative abundances. We also detected the presence in most locations of non ammonia oxidizing Nitrososphaeria (non-AOA), sister groups to the family Nitrososphaeraceae. This group has been seen to grow independently from ammonia oxidation (Weber et al., 2015), and genomic data indicates that could have a chemoorgoheterotrophic lifestyle (Aylward & Santoro, 2020). Moreover, it has been suggested that some members could be sustained by anaerobic respiration, based on reconstructed genomes and the hypoxic and anoxic character of the soils in which they have been detected (Lin et al, 2015). However, genomic

repertoire for aerobic respiration has been also identified in some other members from the group (Beam et al., 2014; Aylward & Santoro, 2020). Still, due to the lack of cultivated representative there is not much known about its specific ecophysiological characteristics. Finally, the analysis at the ASV level confirmed us that the AOA community appears to exhibit limited diversity in the circumpolar arctic tundra soils, as most ASVs are shared among distant locations.

5.2 - The *amoA* clade classification method based on the 16S rRNA gene sequence

The classification of 16S rRNA gene amplicons into the *amoA* clades using the 16S-*amoA* database (Wang et al., 2021) was in general a success. In most of the cases, it could assign in an accurate manner the sequences in the main AOA clades. However, we came across some inconsistencies between the database output and the 16S rRNA gene phylogeny (Figure 7). Firstly, the database, which was built from 124 MAGs and contigs that have both the 16S-rRNA and *amoA* gene in their sequence, has an uneven representation of all AOA. For example, NS-Delta and NS-Gamma are represented by 3 and 5 16S-rRNA gene sequences, respectively, while NP-Gamma is represented by 49 sequences. Thus, we argue that one of the possible reasons for the incongruences could be the lack of more sequences especially from the clades that are underrepresented.

Secondly, we are also aware of the limitations present in the phylogenetic tree that we used to compare to the classification of the database. The use of 16S rRNA gene amplicon sequences to build phylogenetic models can give rise to problems because, due to the short sequences and the highly conservation of the gene, often it is challenging to capture the phylogenetic signal. After identifying the sequence identity threshold between different *amoA* clades at the 16S rRNA gene level to be at ≥ 97%, we saw that some of the ASVs were being classified to a certain clade to nucleotide sequence identities of 90%. Indeed, most of these sequences were the reason for the incongruencies between the 16S-*amoA* database output and the phylogenetic tree information. Additionally, the threshold that we found is similar to the one identified by Pessi et al 2022, were they also explore the ability of using 16S rRNA gene sequences to distinguish between AOA species. What they found was that the 16S had a limited resolution to decipher between species, due to its highly conserved nature. The classification method proposed by Wang et al 2021 uses the tool UCLUST (Edgar, 2010), with the default parameter for nucleotide sequence identity set at 0.9. We propose to increase this minimum sequence identity threshold from the UCLUST tool to

0.97, in order to get more accurate classifications.

5.3 - Identifying the main environmental drivers influencing AOA clade distribution

5.3.1 - Distinct patterns of AOA distribution in farmland soils

We managed to identify different trends of distribution of the AOA clades from the farm soils, and we tried to relate them with the environmental context data available. The patterns were more observable in the DWL farm, because the locations in BK were in general all dominated by NS-Zeta, introducing a bias towards this clade for all categories. What we observed is that NS-Alpha tends to be more abundant in non-fertilized, not highly irrigated, non agricultural soils and that is negatively impacted by high pH values and SW content. We predict that NS-Alpha tends to thrive better in oligotrophic environments. However, we do not have quantitative data on organic carbon and N-species, we cannot truly estimate which are the most oligotrophic sites. On the contrary, NS-Zeta appears to be higher in fertilized agriculturally used soils and to withstand higher pH values (Table 4). This aligns with the physiologies observed in the cultures available from both clades: the more oligotrophic character of NS-Alpha and the higher tolerance for ammonium concentration and pH of NS-Zeta. Something to take into consideration when looking at the possible effects of environmental parameters to the distribution of AOA clades is the interaction that the parameters can have between each other. For instance, these interactions have been detected between irrigation regimes and the pH (Yang et al., 2017), temperature and pH (Gubry-Rangin et al., 2017), and substrate affinity and pH (M. Y. Jung et al., 2022). As mentioned earlier, due to the experimental design in here we could not have identified possible interaction effects between the different environmental factors, due to the lack of negative controls.

The highest proportion of NS-Zeta and overall, of AOA among the whole dataset was found in the site BKSP. One of the particularities of the soil found at this site is the exposure to strong generalist herbicides, such as glyphosate. We wondered if NS-Zeta could be favored by the presence of glyphosate. This is a broad-spectrum herbicide that is used to kill weeds as a non-tillage method alternative. Nitrifiers communities, as well as general microbial communities, have been studied under the application of glyphosate, and no significant shifts on the community and activity have been found (Dennis et al., 2018; Zabaloy et al., 2017). Still, there could be a potential relationship between the effect of glyphosate and

nitrifiers, through the decrease of naturally occurring biological nitrification inhibitors (BNIs) exuded by the weeds targeted with the herbicide, which would have a positive impact on the nitrifiers community. Another possibility is that NS-Zeta could be somehow using up the herbicide as an alternative source of C. Genomic data collected from cultivated representatives from the NS-Zeta revealed potential metabolisms outside the chemolithoautotrophy that are generally associated with AOA (Alves et al., 2019). Specifically, they found genes encoding for enzymes belonging to the CAZy family, known to catalyze the degradation of organic carbon molecules. Alternative energy generation strategies were already suggested for NS-Delta, based on in situ nitrification measurements that detected a lower nitrification rates in sites dominated by this clade and based on ^{13}C-CO_2 stable isotope labeling (SIP) experiments, where NS-Delta and NS-Gamma almost did not show autotrophic activity (Alves et al., 2018). Another particularity from the BKSP samples that could potentially influence the AOA abundance is that they have a high copper concentration in the soil. Information about the other sites regarding the copper level was unfortunately not available. Copper is a key-element in AOA for the ammonia oxidation and for the respiration and its concentration has been seen to be a main driver within AOA abundance (Amin et al., 2013; Qin et al., 2018).

5.3.2 - NS-Gamma and NS-Zeta occupy different niches in arctic tundra soils

The correlation study that we carried out to relate the relative abundances of the three main Nitrososphaeria groups detected in the arctic soils to the available soil physicochemical parameters and enzymatic activity data showed distinct patterns of distribution. Some of the soil environmental factors that we identified as relevant drivers for the AOA community distribution, (e.g. pH, NO_3^-, water content) had been found previously to have a significant role in arctic tundra AOA communities (Alves et al., 2013). The patterns detected here denote the autotrophic oligotrophic lifestyle of AOA, as we see them being more abundant in sites with lower concentrations of organic carbon, N-species, and in sites with lower activity of catabolic enzymes. An opposing trend can be observed with the non-AOA, which are found to be more abundant in the presence of higher concentrations organic compounds, higher water content and higher enzymatic activities, aligning with the proposed heterotrophic lifestyle and potential anaerobic respiration (Beam et al., 2014; Lin et al, 2015). Based on these correlation analyses, we also saw distinct distribution patterns between the two main AOA clades. Although the identification of different responses to specific environmental drivers was less evident in this case, pH seemed to be having a significant impact on their distribution (Table 5). The soils samples used in this study fell into the acidic

range of the pH scale (4.5-6.5) (Suppl. Table), as it is common in arctic soils. Therefore, aligning with the findings in Siljanen and colleagues 2019, the two AOA clades detected here would be adapted to acidic soils. However, it seems that NS-Zeta would be favored by higher pH values within this acidic range. pH has been identified as a main driver for the evolution and diversification of terrestrial AOA (Gubry-Rangin et al., 2015).

In the case of the subclade NS-Gamma-2.3.2, it seems to be a clade specialized in acidic soils (Alves et al., 2018). Vegetation covers can have a strong impact on the microbial community (Malard & Pearce, 2018; Wild et al., 2016) and give us information about the physico-chemical characteristics from the soils. For instance, soils dominated by lichens are usually poor in nutrients and have higher C/N ratios than soils inhabited by vascular plants (Biasi et al., 2008; Gentsch et al., 2015). We observed that there could be specific patterns of distribution based on the dominant vegetation type present at each site. The sites in which NS-Gamma was the dominant AOA clade, had lichens as the main vegetation type in most of the cases. The NS-Gamma identified in this study appear to be very closely related to sequences belonging to the newly proposed genus *Ca.* Nitrospolaris (Figure 15), which has been postulated to have an oligotrophic lifestyle. Therefore, we hypothesize that possibly NS-Gamma might appear in higher abundance in sites dominated by lichens due to the soils harboring a lower nutrient level. In contrast, NS-Zeta in the arctic tundra soils could be occupỳing richer soils. In arctic soils, pH and vegetation cover type have been seen to be the main drivers shaping microbial communities (Malard & Pearce, 2018). Altogether, these results support the proposed hypothesis that NS-Gamma and NS-Zeta could be occupying different niches in the arctic tundra soils. As it has been previously suggested, when an AOA clade substitutes another it could be because they are functionally redundant (Allison & Martiny, 2008; Zhao et al., 2022).

5.4 - Psychrophilic AOA clades adapted to live in the arctic soils

Specific biogeographical patterns in the microbial communities across the polar/alpine climate zone (i.e. cryosphere) have been studied (Malard et al., 2019; Malard & Pearce, 2018), however the archaeal community in these areas remains understudied. The exploration of such extensive areas require large-scale studies with common standardized methodologies to be able to compare the results. Here, we were able to identify that only a few sub-lineages of the NS-Gamma and NS-Zeta clades compose the AOA community in arctic tundra soils, as opposed to temperate soils where the AOA community is highly diverse. Pessi et al., 2021 proposed that *Ca.* Nitrosopolaris is distributed along the

polar/alpine climatic zone based on the origins of their MAGs (i.e. Canada, Norway, Antarctica) and based on the high nucleotide sequence identity similarity to sequences in databases belonging to other cold areas, such as the Tibetan Plateau. The occurrence of same or very closely related archaeal lineages, on an ASV level, in geographically distinct and remote locations is also observed for other extreme environments: for instance, similar *Sulfolobus* or *Nitrosocaldus* species in hot springs around the world (Mao & Grogan 2012). Currently, there is only one cultivated arctic soil representative available of the NS-Zeta lineage: *Ca.* Nitrosocosmicus arcticus Kfb (Alves et al., 2019), while cultivated representatives from the NS-Gamma clade are still missing. The cultivation of this model organism and other potential future isolates can inform about mitigation strategies for N2O emissions, particularly relevant in the rapidly warming environment of the arctic.

6. CONCLUSIONS

The aim of this study was to analyze the AOA communities and the main drivers behind their distribution in two different soil ecosystems such as farmland soils and the arctic tundra soils. Often, in these two ecosystems there is a high emission of N_2O. Even though the causes behind these emissions are different, there is a pressing interest for the study and understanding of the N-cycling to try to foresee and mitigate the consequences of the release of N_2O in these environments. Therefore, the study of the nitrifier communities, and in particular of the AOA as the most abundant nitrifiers in most soil ecosystems, is highly relevant in farmlands and arctic soils.

Our results from the two very different geographical locations studied in this work reveal distinct patterns of distribution of the main AOA clades/lineages across the different ecosystems and how they relate to the different ecophysiologies of each clade. We also managed to identify trends regarding the influence of certain environmental parameters on the distribution of the AOA clades.

In the case of the farm soils in Montana we detected that AOA dominate the nitrifiers' community, with relative abundance values ranging from 0.5% to 19%, and exhibit a diversity of clades. The clade composition showed an unexpectedly high relative abundance of NS-Zeta, particularly in the ranch Ben&Karli, and NS-Alpha. According to our results, NS-Alpha tends to be more abundant in non-agricultural soils that are not exposed to fertilizers and not irrigated, and it was negatively impacted by higher pH and soil water content values. Conversely, NS-Zeta was more abundant in the sites actively used for agricultural practices, exposed to fertilizers, frequent irrigation and strong herbicides, and could withstand higher values of pH and SW. These observations agree with the observed physiologies of some of the available isolates from these clades.

In contrast, the AOA community in the arctic tundra soils exhibited low diversity and was composed mainly by two sublineages belonging to the clades NS-Gamma and NS-Zeta, in agreement with previous studies of arctic environments, as well as a high relative abundance of non-ammonia oxidizing Nitrososphaeria. Furthermore, we could observe that the detected clades had a different pattern of distribution across the four circumpolar locations included in the study. The AOA community in the arctic appears to be more abundant in sites with lower content of organic carbon and N-species, as well as lower

activity of enzymes of carbon catabolism, agreeing with the autotrophic and oligotrophic lifestyle generally associated with AOA in literature. Specific environmental drivers behind the different distribution of NS-Zeta and NS-Gamma could include pH, with NS-Zeta more tolerant to higher pH values, and the tundra vegetation subtype.

Additionally, the NS-Gamma ASVs detected in this study cluster closely together with 16S rRNA gene sequences from the newly proposed genus *Nitrosopolaris* sp., which comprises of species adapted to the extremely cold environments. Our findings in the arctic tundra suggest the existence of AOA psychrophilic or psychrotolerant clades that have specific adaptations to extreme cold environments.

The former phylum level classification of the lineage was Thaumarchaeota, which derives from the greek word *thaûma,* meaning "miracle". Indeed, this name was appropiate to describe the astonishing varieties of ecosystems that AOA inhabit. Overall, our results underline the importance of studying the AOA communities in depth and constructing a framework for their analysis and detection of their diverse lineages, in order to understand their diverse ecosystem functions.

7. REFERENCES

Allison, S. D., & Martiny, J. B. H. (2008). Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(SUPPL. 1), 11512–11519. https://doi.org/10.1073/pnas.0801925105

Alves, R. J. E., Kerou, M., Zappe, A., Bittner, R., Abby, S. S., Schmidt, H. A., Pfeifer, K., & Schleper, C. (2019). Ammonia oxidation by the arctic terrestrial thaumarchaeote candidatus nitrosocosmicus arcticus is stimulated by increasing temperatures. *Frontiers in Microbiology*, *10*(JULY). https://doi.org/10.3389/fmicb.2019.01571

Alves, R. J. E., Minh, B. Q., Urich, T., Haeseler, A. V., & Schleper, C. (2018). Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on *amoA* genes. *Nature Communications*, *9*(1), 1–17. https://doi.org/10.1038/s41467-018-03861-1

Alves, R. J. E., Wanek, W., Zappe, A., Richter, A., Svenning, M. M., Schleper, C., & Urich, T. (2013). Nitrification rates in Arctic soils are associated with functionally distinct populations of ammonia-oxidizing archaea. *Isme J*, *7*(8), 1620–1631.

Amin, S. A., Moffett, J. W., Martens-Habbena, W., Jacquot, J. E., Han, Y., Devol, A., Ingalls, A. E., Stahl, D. A., & Armbrust, E. V. (2013). Copper requirements of the ammonia-oxidizing archaeon Nitrosopumilus maritimus SCM1 and implications for nitrification in the marine environment. *Limnology and Oceanography*, *58*(6), 2037–2045. https://doi.org/10.4319/lo.2013.58.6.2037

Apprill, A., Mcnally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, *75*(2), 129–137. https://doi.org/10.3354/ame01753

Arce, M. I., von Schiller, D., Bengtsson, M. M., Hinze, C., Jung, H., Alves, R. J. E., Urich, T., & Singer, G. (2018). Drying and Rainfall Shape the Structure and Functioning of Nitrifying Microbial Communities in Riverbed Sediments. *Frontiers in Microbiology*, *9*, 2794. https://doi.org/10.3389/fmicb.2018.02794

Aylward, F. O., & Santoro, A. E. (2020). Heterotrophic Thaumarchaea with Small Genomes Are Widespread in the Dark Ocean. *MSystems*, *5*(3), 1–20. https://doi.org/10.1128/msystems.00415-20

Banerjee, S., Schlaeppi, K., & Heijden, M. G. A. van der. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology*, *16*(9), 567–576. https://doi.org/10.1038/s41579-018-0024-1

Beam, J. P., Jay, Z. J., Kozubal, M. A., & Inskeep, W. P. (2014). Niche specialization of

novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *The ISME journal*, *8*(4), 938-951.

Biasi, C., Meyer, H., Rusalimova, O., Hämmerle, R., Kaiser, C., Baranyi, C., Daims, H., Lashchinsky, N., Barsukov, P., & Richter, A. (2008). Initial effects of experimental warming on carbon exchange rates, plant growth and microbial dynamics of a lichen-rich dwarf shrub tundra in Siberia. *Plant and Soil*, *307*(1–2), 191–205. https://doi.org/10.1007/s11104-008-9596-2

Burggraf, S., Huber, H., & Regensburg, U. (1997). Reclassification of the Crenarchaeal Accordance with 16S rRNA Orders and Families in Sequence Data. *Order A Journal On The Theory Of Ordered Sets And Its Applications*, *47*(July), 657–660.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caranto, J. D., & Lancaster, K. M. (2017). Nitric oxide is an obligate bacterial nitrification intermediate produced by hydroxylamine oxidoreductase. *Proceedings of the National Academy of Sciences*, *114*(31), 8217–8222. https://doi.org/10.1073/pnas.1704504114

Carini, P., Dupont, C. L., & Santoro, A. E. (2018). Patterns of thaumarchaeal gene expression in culture and diverse marine environments. *Environmental Microbiology*, *20*(6), 2112–2124. https://doi.org/10.1111/1462-2920.14107

Castro, J. F., Nouioui, I., Asenjo, J. A., Andrews, B., Bull, A. T., & Goodfellow, M. (2019). New genus-specific primers for PCR identification of Rubrobacter strains. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *112*(12), 1863–1874. https://doi.org/10.1007/s10482-019-01314-3

Cooper, R. E., Wegner, C.-E., McAllister, S. M., Shevchenko, O., Chan, C. S., & Küsel, K. (2020). Draft Genome Sequence of Sideroxydans sp. Strain CL21, an Fe(II)-Oxidizing Bacterium. *Microbiology Resource Announcements*, *9*(2), e01444-19. https://doi.org/10.1128/MRA.01444-19

Coskun, D., Britto, D. T., Shi, W., & Kronzucker, H. J. (2017). Nitrogen transformations in modern agriculture and the role of biological nitrification inhibition. *Nature Plants*, *3*(June). https://doi.org/10.1038/nplants.2017.74

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R. H., Bergen, M. V., Rattei, T., Bendinger, B., Nielsen, P. H., & Wagner, M. (2015). Complete nitrification by Nitrospira bacteria. *Nature*, *528*(7583), 504–509. https://doi.org/10.1038/nature16461

Dennis, P. G., Kukulies, T., Forstner, C., Orton, T. G., & Pattison, A. B. (2018). The effects of glyphosate, glufosinate, paraquat and paraquat-diquat on soil microbial activity and

bacterial, archaeal and nematode diversity. *Scientific Reports*, *8*(1), 1–9. https://doi.org/10.1038/s41598-018-20589-6

Dise, N. B., Ashmore, M., Belyazid, S., Bleeker, A., Bobbink, R., de Vries, W., Erisman, J. W., Spranger, T., Stevens, C. J., & van den Berg, L. (2011). Nitrogen as a threat to European terrestrial biodiversity. In M. A. Sutton, C. M. Howard, J. W. Erisman, G. Billen, A. Bleeker, P. Grennfelt, H. van Grinsven, & B. Grizzetti (Eds.), *The European Nitrogen Assessment* (1st ed., pp. 463–494). Cambridge University Press. https://doi.org/10.1017/CBO9780511976988.023

Drenovsky, R. E., Vo, D., Graham, K. J., & Scow, K. M. (2004). Soil water content and organic carbon availability are major determinants of soil microbial community composition. *Microbial Ecology*, *48*(3), 424–430. https://doi.org/10.1007/s00248-003-1063-2

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Erguder, T., Boon, N., Wittebolle, L., Marzorati, M., & Verstraete, W. (2009). Environmental factors shaping the ecological niches of ammonia-oxidizing archaea. *FEMS Microbiology Reviews*, *33*(5), 855–869.

Erisman, J. W., Galloway, J. N., Dise, N. B., Sutton, M. A., Bleeker, A., Grizzetti, B., Leach, A. M., & Vries, W. de. (2015). *Nitrogen: Too much of a vital resource*. WWF Nederland.

Fierer, N., Bradford, M. A., & Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology*, *88*(6), 1354–1364. https://doi.org/10.1890/05-1839

Fierer, N., Lauber, C. L., Ramirez, K. S., Zaneveld, J., Bradford, M. A., & Knight, R. (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *The ISME journal*, *6*(5), 1007-1017.

Finneran, K. T., Johnsen, C. V., & Lovley, D. R. (2003). Rhodoferax ferrireducens sp. Nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). *International Journal of Systematic and Evolutionary Microbiology*, *53*(3), 669–673. https://doi.org/10.1099/ijs.0.02298-0

Fowler, S. J., Palomo, A., Dechesne, A., Mines, P. D., & Smets, B. F. (2018). Comammox Nitrospira are abundant ammonia oxidizers in diverse groundwater-fed rapid sand filter communities. *Environmental Microbiology*, *20*(3), 1002–1015. https://doi.org/10.1111/1462-2920.14033

Galloway, J. N., Aber, J. D., Erisman, J. W., Seitzinger, S. P., Howarth, R. W., Cowling, E. B., & Cosby, B. J. (2003). The Nitrogen Cascade. *BioScience*, *53*(4), 341. https://doi.org/10.1641/0006-3568(2003)053[0341:TNC]2.0.CO;2

Gentsch, N., Mikutta, R., Alves, R. J. E., Barta, J., Čapek, P., Gittel, A., Hugelius, G., Kuhry,

P., Lashchinskiy, N., Palmtag, J., Richter, A., Šantrůčková, H., Schnecker, J., Shibistova, O., Urich, T., Wild, B., & Guggenberger, G. (2015). Storage and transformation of organic matter fractions in cryoturbated permafrost soils across the Siberian Arctic. *Biogeosciences*, *12*(14), 4525–4542. https://doi.org/10.5194/bg-12-4525-2015

Gittel, A., Bárta, J., Kohoutová, I., Mikutta, R., Owens, S., Gilbert, J., Schnecker, J., Wild, B., Hannisdal, B., Maerz, J., Lashchinskiy, N., Čapek, P., Šantrůčková, H., Gentsch, N., Shibistova, O., Guggenberger, G., Richter, A., Torsvik, V. L., Schleper, C., & Urich, T. (2014). Distinct microbial communities associated with buried soils in the siberian tundra. *ISME Journal*, *8*(4), 841–853. https://doi.org/10.1038/ismej.2013.219

Gittel, A., Bárta, J., Kohoutová, I., Schnecker, J., Wild, B., Čapek, P., Kaiser, C., Torsvik, V. L., Richter, A., Schleper, C., & Urich, T. (2014). Site- and horizon-specific patterns of microbial community structure and enzyme activities in permafrost-affected soils of Greenland. *Frontiers in Microbiology*, *5*(OCT), 1–14. https://doi.org/10.3389/fmicb.2014.00541

Goreau, T. J., Kaplan, W. A., Wofsy, S. C., McElroy, M. B., Valois, F. W., & Watson, S. W. (1980). Production of NO_2^- and N_2O by Nitrifying Bacteria at Reduced Concentrations of Oxygen. *Applied and Environmental Microbiology*, *40*(3), 526–532. https://doi.org/10.1128/aem.40.3.526-532.1980

Gubry-Rangin, C., Kratsch, C., Williams, T. A., Mchardy, A. C., Embley, T. M., Prosser, J. I., & Macqueen, D. J. (2015). Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *Proceedings of the National Academy of Sciences of the United States of America*. file:///Users/melinaki/Documents/Papersoldmac/Papers/2015/Gubry-Rangin/Proceedings of the National Academy Sciences of.../Proceedings United States America 2015 Gubry-Rangin.pdf papers:

Gubry-Rangin, C., Novotnik, B., Mandič-Mulec, I., Nicol, G. W., & Prosser, J. I. (2017). Temperature responses of soil ammonia-oxidising archaea depend on pH. *Soil Biology and Biochemistry*, *106*(April), 61–68. https://doi.org/10.1016/j.soilbio.2016.12.007

Guo, J., Ling, N., Chen, H., Zhu, C., Kong, Y., Wang, M., Shen, Q., & Guo, S. (2017). Distinct drivers of activity, abundance, diversity and composition of ammonia-oxidizers: Evidence from a long-term field experiment. *Soil Biology and Biochemistry*, *115*, 403–414. https://doi.org/10.1016/j.soilbio.2017.09.007

Hatzenpichler, R. (2012). Diversity, Physiology, and Niche Differentiation of Ammonia-Oxidizing Archaea. *Applied and Environmental Microbiology*, *78*(21), 7501–7510.

Hink, L., Gubry-Rangin, C., Nicol, G. W., & Prosser, J. I. (2018). The consequences of niche

and physiological differentiation of archaeal and bacterial ammonia oxidisers for nitrous oxide emissions. *The ISME Journal*, *12*(4), 1084. https://doi.org/10.1038/s41396-017-0025-5

Hink, L., Lycus, P., Gubry-Rangin, C., Frostegård, Å., Nicol, G. W., Prosser, J. I., & Bakken, L. R. (2017). Kinetics of NH3-oxidation, NO-turnover, N2O-production and electron flow during oxygen depletion in model bacterial and archaeal ammonia oxidisers. *Environmental Microbiology*, *September*. https://doi.org/10.1111/1462-2920.13914

Hink, L., Nicol, G. W., & Prosser, J. I. (2017). Archaea produce lower yields of N2O than bacteria during aerobic ammonia oxidation in soil. *Environmental Microbiology*, *19*(12), 4829–4837. https://doi.org/10.1111/1462-2920.13282

Ho, A., Lonardo, D. P. D., & Bodelier, P. L. E. (2017). Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiology Ecology*, *93*(3), 1–14. https://doi.org/10.1093/femsec/fix006

Huang, L., Chakrabarti, S., Cooper, J., Perez, A., John, S. M., Daroub, S. H., & Martens-Habbena, W. (2021). Ammonia-oxidizing archaea are integral to nitrogen cycling in a highly fertile agricultural soil. *ISME Communications*, *1*(1), 1–12. https://doi.org/10.1038/s43705-021-00020-4

Huber, H., Hohn, M. J., Stetter, K. O., & Rachel, R. (2003). The phylum Nanoarchaeota: Present knowledge and future perspectives of a unique form of life. *Research in Microbiology*, *154*(3), 165–171. https://doi.org/10.1016/S0923-2508(03)00035-4

Jeon, C. O., Park, W., Padmanabhan, P., DeRito, C., Snape, J. R., & Madsen, E. L. (2003). Discovery of a bacterium, with distinctive dioxygenase, that is responsible for in situ biodegradation in contaminated sediment. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23), 13591–13596. https://doi.org/10.1073/pnas.1735529100

Jung, M. Y., Sedlacek, C. J., Kits, K. D., Mueller, A. J., Rhee, S. K., Hink, L., Nicol, G. W., Bayer, B., Lehtovirta-Morley, L., Wright, C., Torre, J. R. de la, Herbold, C. W., Pjevac, P., Daims, H., & Wagner, M. (2022). Ammonia-oxidizing archaea possess a wide range of cellular ammonia affinities. *ISME Journal*, *16*(1), 272–283. https://doi.org/10.1038/s41396-021-01064-z

Jung, M.-Y., Kim, J.-G., Damsté, J. S. S., Rijpstra, W. I. C., Madsen, E. L., Kim, S.-J., Hong, H., Si, O.-J., Kerou, M., Schleper, C., & Rhee, S.-K. (2016). A hydrophobic ammonia-oxidizing archaeon of the Nitrosocosmicus clade isolated from coal tar-contaminated sediment. *Environmental Microbiology Reports*, *11*(5), E7937–E7946. https://doi.org/10.1111/1758-2229.12477

Kasanke, C. P., Collins, R. E., & Leigh, M. B. (2019). Identification and Characterization of a Dominant Sulfolane-Degrading Rhodoferax sp. Via Stable Isotope Probing Combined

with Metagenomics. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-40000-2

Katoh, K., & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, *9*(4), 286–298. https://doi.org/10.1093/bib/bbn013

Kim, J., Jung, M., Park, S., Rijpstra, W. I., Damsté, J. S. S., Madsen, E. L., Min, D., Kim, J. S., Kim, G. J., & Rhee, S. (2012). Cultivation of a highly enriched ammonia-oxidizing archaeon of thaumarchaeotal group I.1b from an agricultural soil. *Environmental Microbiology*, *14*(6), 1528–1543.

Kits, K. D., Sedlacek, C. J., Lebedeva, E. V., Han, P., Bulaev, A., Pjevac, P., Daebeler, A., Romano, S., Albertsen, M., Stein, L. Y., Daims, H., & Wagner, M. (2017). Kinetic analysis of a complete nitrifier reveals an oligotrophic lifestyle. *Nature*, *549*, pages 269-272. https://doi.org/10.1038/nature23679

Könneke, M., Bernhard, A. E., Torre, J. R. D. L., Walker, C. B., Waterbury, J. B., & Stahl, D. A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature*, *437*(7058), 543–546. https://doi.org/10.1038/nature03911

Könneke, M., Schubert, D. M., Brown, P. C., Hügler, M., Standfest, S., Schwander, T., Borzyskowski, L. S. V., Erb, T. J., Stahl, D. A., & Berg, I. A. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(22), 8239–8244. https://doi.org/10.1073/pnas.1402028111

Koven, C. D., Ringeval, B., Friedlingstein, P., Ciais, P., Cadule, P., Khvorostyanov, D., Krinner, G., & Tarnocai, C. (2011). Permafrost carbon-climate feedbacks accelerate global warming. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 14769–14774. https://doi.org/10.1073/pnas.1103910108

Kozlowski, J. A., Price, J., & Stein, L. Y. (2014). Revision of N $_2$ O-Producing Pathways in the Ammonia-Oxidizing Bacterium Nitrosomonas europaea ATCC 19718. *Applied and Environmental Microbiology*, *80*(16), 4930–4935. https://doi.org/10.1128/AEM.01061-14

Kozlowski, J. A., Stieglmeier, M., Schleper, C., Klotz, M. G., & Stein, L. Y. (2016). Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME Journal*, *10*(8), 1836–1845. https://doi.org/10.1038/ismej.2016.2

Lauber, C. L., Hamady, M., Knight, R., & Fierer, N. (2009). Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology*, *75*(15), 5111–5120. https://doi.org/10.1128/AEM.00335-09

Lehtovirta-Morley, L. E., Ge, C., Ross, J., Yao, H., Nicol, G. W., & Prosser, J. I. (2014). Characterisation of terrestrial acidophilic archaeal ammonia oxidisers and their inhibition and stimulation by organic compounds. *FEMS Microbiology Ecology*, *89*(3), 542–552.

Lehtovirta-Morley, L. E., Ross, J., Hink, L., Weber, E. B., Gubry-Rangin, C., Thion, C., Prosser, J. I., & Nicol, G. W. (2016). Isolation of "Candidatus Nitrosocosmicus franklandus", a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiology Ecology*, *92*(5), fiw057. https://doi.org/10.1093/femsec/fiw057

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., Prosser, J. I., Schuster, S. C., & Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*, *442*(7104). http://www.nature.com/nature/journal/v442/n7104/full/nature04983.html file:///Users/melinaki/Documents/Papersoldmac/Papers/2006/Leininger/Nature/Nature 2006 Leininger.pdf papers://dfa503e6-d258-4223-9ef8-d69bedc66f2d/Paper/p2969

Li, T., Zhuo, Y., Jin, C. Z., Wu, X., Ko, S. R., Jin, F. J., Ahn, C. Y., Oh, H. M., Lee, H. G., & Jin, L. (2020). Genomic insights into a novel species Rhodoferax aquaticus sp. Nov., isolated from freshwater. *International Journal of Systematic and Evolutionary Microbiology*, *70*(8), 4653–4660. https://doi.org/10.1099/ijsem.0.004325

Lin, X., Handley, K. M., Gilbert, J. A., & Kostka, J. E. (2015). Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *The ISME journal*, *9*(12), 2740-2744.

Liu, L., Liu, M., Jiang, Y., Lin, W., & Luo, J. (2021). Production and Excretion of Polyamines To Tolerate High Ammonia, a Case Study on Soil Ammonia-Oxidizing Archaeon " Candidatus Nitrosocosmicus agrestis." *MSystems*, *6*(1). https://doi.org/10.1128/msystems.01003-20

Liu, S., Wang, H., Chen, L., Wang, J., Zheng, M., Liu, S., Chen, Q., & Ni, J. (2020). Comammox Nitrospira within the Yangtze River continuum: Community, biogeography, and ecological drivers. *ISME Journal*, *14*(10), 2488–2504. https://doi.org/10.1038/s41396-020-0701-8

Lu, X., Nicol, G. W., & Neufeld, J. D. (2018). Differential responses of soil ammonia-oxidizing archaea and bacteria to temperature and depth under two different land uses. *Soil Biology and Biochemistry*, *120*, 272–282. https://doi.org/10.1016/j.soilbio.2018.02.017

Malard, L. A., Anwar, M. Z., Jacobsen, C. S., & Pearce, D. A. (2019). Biogeographical patterns in soil bacterial communities across the Arctic region. *FEMS Microbiology*

Ecology, *95*(9), 1–13. https://doi.org/10.1093/femsec/fiz128

Malard, L. A., & Pearce, D. A. (2018). Microbial diversity and biogeography in Arctic soils. *Environmental Microbiology Reports*, *10*(6), 611–625. https://doi.org/10.1111/1758-2229.12680

Martin, Marcel, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*, 10–12.

Meinhardt, K. A., Stopnisek, N., Pannu, M. W., Strand, S. E., Fransen, S. C., Casciotti, K. L., & Stahl, D. A. (2018). Ammonia-oxidizing bacteria are the primary N2O producers in an ammonia-oxidizing archaea dominated alkaline agricultural soil. *Environmental Microbiology*, *20*(6), 2195–2206. https://doi.org/10.1111/1462-2920.14246

Minh, B. Q., Nguyen, M. A. T., & Haeseler, A. V. (2013). Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

Momesso, L., Crusciol, C. A. C., Leite, M. F. A., Bossolani, J. W., & Kuramae, E. E. (2022). Forage Grasses Steer Soil Nitrogen Processes, Microbial Populations, and Microbiome Composition in A Long-term Tropical Agriculture System. *Agriculture, Ecosystems and Environment*, *323*(January 2021), 107688. https://doi.org/10.1016/j.agee.2021.107688

Mußmann, M., Brito, I., Pitcher, A., Damsté, J. S. S., Hatzenpichler, R., Richter, A., Nielsen, J. L., Nielsen, P. H., Müller, A., Daims, H., Wagner, M., & Head, I. M. (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16771–16776. https://doi.org/10.1073/pnas.1106427108

Nicol, G. W., Leininger, S., Schleper, C., & Prosser, J. I. (2008). The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environmental Microbiology*, *10*(11), 2966–2978. https://doi.org/10.1111/j.1462-2920.2008.01701.x

Nielsen, S., Minchin, T., Kimber, S., Zwieten, L. van, Gilbert, J., Munroe, P., Joseph, S., & Thomas, T. (2014). Comparative analysis of the microbial communities in agricultural soil amended with enhanced biochars or traditional fertilisers. *Agriculture, Ecosystems and Environment*, *191*, 73–82. https://doi.org/10.1016/j.agee.2014.04.006

Nordin, A., Schmidt, I. K., & Shaver, G. R. (2004). Nitrogen uptake by arctic soil microbes and plants in relation to soil nitrogen supply. *Ecology*, *85*(4), 955–962. https://doi.org/10.1890/03-0084

Norman, J. S., King, G. M., & Friesen, M. L. (2017). Rubrobacter spartanus sp. Nov., a

moderately thermophilic oligotrophic bacterium isolated from volcanic soil. *International Journal of Systematic and Evolutionary Microbiology*, *67*(9), 3597–3602. https://doi.org/10.1099/ijsem.0.002175

Norton, J. M., Alzerreca, J. J., Suwa, Y., & Klotz, M. G. (2002). Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Archives of Microbiology*, *177*(2), 139–149. https://doi.org/10.1007/s00203-001-0369-z

Orellana, L. H., Chee-Sanford, J. C., Sanford, R. A., Löffler, F. E., & Konstantinidis, K. T. (2018). Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Applied and Environmental Microbiology*, *84*(2). https://doi.org/10.1128/AEM.01646-17

Oton, E. V., Quince, C., Nicol, G. W., Prosser, J. I., & Gubry-Rangin, C. (2016). Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *The ISME Journal*, *10*(1), 85–96. https://doi.org/10.1038/ismej.2015.101

Pausan, M. R., Csorba, C., Singer, G., Till, H., Schöpf, V., Santigli, E., Klug, B., Högenauer, C., Blohs, M., & Moissl-Eichinger, C. (2019). Exploring the Archaeome: Detection of Archaeal Signatures in the Human Body. *Frontiers in Microbiology*, *10*(December), 1–13. https://doi.org/10.3389/fmicb.2019.02796

Pessi, I. S., Rutanen, A., & Hultman, J. (2022). Candidatus Nitrosopolaris, a genus of putative ammonia-oxidizing archaea with a polar/alpine distribution. *FEMS Microbes*. https://doi.org/10.1093/femsmc/xtac019

Pester, M., Rattei, T., Flechl, S., Gröngröft, A., Richter, A., Overmann, J., Reinhold-Hurek, B., Loy, A., & Wagner, M. (2012). *amoA* -based consensus phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions. *Environmental Microbiology*, *14*(2), 525–539. https://doi.org/10.1111/j.1462-2920.2011.02666.x

Pester, M., Schleper, C., & Wagner, M. (2011). The Thaumarchaeota: An emerging view of their phylogeny and ecophysiology. *Current Opinion in Microbiology*, *14*(3), 300–306. https://doi.org/10.1016/j.mib.2011.04.007

Pinto, A. J., Marcus, D. N., Ijaz, U. Z., Bautista-de lose Santos, Q. M., Dick, G. J., & Raskin, L. (2016). Metagenomic Evidence for the Presence of Comammox *Nitrospira* -Like Bacteria in a Drinking Water System. *MSphere*, *1*(1), e00054-15. https://doi.org/10.1128/mSphere.00054-15

Pjevac, P., Schauberger, C., Poghosyan, L., Herbold, C. W., Kessel, M. A. H. J. van, Daebeler, A., Steinberger, M., Jetten, M. S. M., Lücker, S., Wagner, M., & Daims, H. (2017). *amoA*-Targeted Polymerase Chain Reaction Primers for the Specific Detection and Quantification of Comammox Nitrospira in the Environment. *Frontiers in Microbiology*, *8*, 1508. https://doi.org/10.3389/fmicb.2017.01508

Pörtner, H.-O., & Roberts, D. C. (n.d.). *Climate Change 2022: Impacts, Adaptation and Vulnerability*. 3068.

Prosser, J. I., & Nicol, G. W. (2008). Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environmental Microbiology*, *10*(11), 2931–2941. https://doi.org/10.1111/j.1462-2920.2008.01775.x

Purkhold, U., Pommerening-Röser, A., Juretschko, S., Schmid, M. C., Koops, H.-P., & Wagner, M. (2000). Phylogeny of All Recognized Species of Ammonia Oxidizers Based on Comparative 16S rRNA and *amoA* Sequence Analysis: Implications for Molecular Diversity Surveys. *Applied and Environmental Microbiology*, *66*(12), 5368–5382. https://doi.org/10.1128/AEM.66.12.5368-5382.2000

Qin, W., Amin, S. A., Lundeen, R. A., Heal, K. R., Martens-Habbena, W., Turkarslan, S., Urakawa, H., Costa, K. C., Hendrickson, E. L., Wang, T., Beck, D. A., Tiquia-Arashiro, S. M., Taub, F., Holmes, A. D., Vajrala, N., Berube, P. M., Lowe, T. M., Moffett, J. W., Devol, A. H., … Stahl, D. A. (2018). Stress response of a marine ammonia-oxidizing archaeon informs physiological status of environmental populations. *The ISME Journal*, *12*(2), 508–519. https://doi.org/10.1038/ismej.2017.186

Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P. A., Davín, A. A., Waite, D. W., Whitman, W. B., Parks, D. H., & Hugenholtz, P. (2021). A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nature Microbiology*, *6*(7), 946–959. https://doi.org/10.1038/s41564-021-00918-8

Rotthauwe, J. H., Witzel, K. P., & Liesack, W. (1997). The ammonia monooxygenase structural gene *amoA* as a functional marker: Molecular fine-scale analysis of natural ammonia-oxidizing populations. *Applied and Environmental Microbiology*, *63*(12), 4704–4712. https://doi.org/10.1128/aem.63.12.4704-4712.1997

Sait, M., Hugenholtz, P., & Janssen, P. H. (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environmental Microbiology*, *4*(11), 654–666. https://doi.org/10.1046/j.1462-2920.2002.00352.x

Sauder, L. A., Albertsen, M., Engel, K., Schwarz, J., Nielsen, P. H., Wagner, M., & Neufeld, J. D. (2017). Cultivation and characterization of Candidatus Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment plant. *Isme J*, 1–16. https://doi.org/10.1038/ismej.2016.192

Schleper, C. (2010). Ammonia oxidation: Different niches for bacteria and archaea? *The ISME Journal*, *4*(9), 1092–1094.

Schleper, C., & Nicol, G. W. (2010). Ammonia-oxidising archaea–physiology, ecology and evolution. *Advances in Microbial Physiology*, *57*, 1–41.

Shange, R. S., Ankumah, R. O., Ibekwe, A. M., Zabawa, R., & Dowd, S. E. (2012). Distinct soil bacterial communities revealed under a diversely managed agroecosystem. *PLoS ONE*, *7*(7). https://doi.org/10.1371/journal.pone.0040338

Shcherbakova, V., Yoshimura, Y., Ryzhmanova, Y., Taguchi, Y., Segawa, T., Oshurkova, V., & Rivkina, E. (2016). Archaeal communities of Arctic methane-containing permafrost. *FEMS Microbiology Ecology*, *92*(10), 1–11. https://doi.org/10.1093/femsec/fiw135

Siljanen, H. M. P., Alves, R. J. E., Ronkainen, J. G., Lamprecht, R. E., Bhattarai, H. R., Bagnoud, A., Marushchak, M. E., Martikainen, P. J., Schleper, C., & Biasi, C. (2019). Archaeal nitrification is a key driver of high nitrous oxide emissions from arctic peatlands. *Soil Biology and Biochemistry*, *137*(March), 107539. https://doi.org/10.1016/j.soilbio.2019.107539

Singleton, D. R., Furlong, M. A., Peacock, A. D., White, D. C., Coleman, D. C., & Whitman, W. B. (2003). Solirubrobacter pauli gen. Nov., sp. Nov., a mesophilic bacterium within the Rubrobacteridae related to common soil clones. *International Journal of Systematic and Evolutionary Microbiology*, *53*(2), 485–490. https://doi.org/10.1099/ijs.0.02438-0

Sorokin, D. Y., Pelt, S. van, Tourova, T. P., & Evtushenko, L. I. (2009). Nitriliruptor alkaliphilus gen. Nov., sp. Nov., a deeplineage haloalkaliphilic actinobacterium from soda lakes capable of growth on aliphatic nitriles, and proposal of Nitriliruptoraceae fam. Nov. And Nitriliruptorales ord. Nov. *International Journal of Systematic and Evolutionary Microbiology*, *59*(2), 248–253. https://doi.org/10.1099/ijs.0.002204-0

Spang, A., Poehlein, A., Offre, P., Zumbrägel, S., Haider, S., Rychlik, N., Nowka, B., Schmeisser, C., Lebedeva, E., Rattei, T., Böhm, C., Schmid, M., Galushko, A., Hatzenpichler, R., Weinmaier, T., Daniel, R., Schleper, C., Spieck, E., Streit, W., & Wagner, M. (2012). The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: Insights into metabolic versatility and environmental adaptations. *Environmental Microbiology*, *14*(12), 3122–3145.

Stahl, D. A., & Torre, J. R. de la. (2012). Physiology and diversity of ammonia-oxidizing archaea. *Annual Review of Microbiology*, *66*, 83–101.

Stein, L. Y. (2019). Insights into the physiology of ammonia-oxidizing microorganisms. *Current Opinion in Chemical Biology*, *49*, 9–15. https://doi.org/10.1016/j.cbpa.2018.09.003

Stephen, J. R., Chang, Y.-J., Macnaughton, S. J., Kowalchuk, G. A., Leung, K. T., Flemming, C. A., & White, D. C. (1999). Effect of Toxic Metals on Indigenous Soil β-Subgroup Proteobacterium Ammonia Oxidizer Community Structure and Protection against Toxicity by Inoculated Metal-Resistant Bacteria. *Applied and Environmental Microbiology*, *65*(1), 95–101. https://doi.org/10.1128/AEM.65.1.95-101.1999

Sterngren, A. E., Hallin, S., & Bengtson, P. (2015). Archaeal ammonia oxidizers dominate in numbers, but bacteria drive gross nitrification in n-amended grassland soil. *Frontiers in Microbiology*, *6*(NOV), 1–8. https://doi.org/10.3389/fmicb.2015.01350

Stieglmeier, M., Klingl, A., Alves, R. J. E., Rittmann, S. K. M. R., Melcher, M., Leisch, N., & Schleper, C. (2014). Nitrososphaera viennensis gen. Nov., sp. Nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. *International Journal of Systematic and Evolutionary Microbiology*, *64*(PART 8), 2738–2752. https://doi.org/10.1099/ijs.0.063172-0

Sun, X., Zhao, J., Zhou, X., Bei, Q., Xia, W., Zhao, B., Zhang, J., & Jia, Z. (2021). Salt tolerance-based niche differentiation of soil ammonia oxidizers. *The ISME Journal*, 1–11. https://doi.org/10.1038/s41396-021-01079-6

Talbot, J. M., Allison, S. D., & Treseder, K. K. (2008). Decomposers in disguise: Mycorrhizal fungi as regulators of soil C dynamics in ecosystems under global change. *Functional Ecology*, *22*(6), 955–963. https://doi.org/10.1111/j.1365-2435.2008.01402.x

Tourna, M., Freitag, T. E., Nicol, G. W., & Prosser, J. I. (2008). Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environmental Microbiology*, *10*(5), 1357–1364. https://doi.org/10.1111/j.1462-2920.2007.01563.x

Tourna, M., Stieglmeier, M., Spang, A., Könneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., Wagner, M., Richter, A., & Schleper, C. (2011). Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(20), 8420–8425. https://doi.org/10.1073/pnas.1013488108

Treusch, A. H., Leininger, S., Kietzin, A., Schuster, S. C., Klenk, H. P., & Schleper, C. (2005). Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environmental Microbiology*, *7*(12), 1985–1995. https://doi.org/10.1111/j.1462-2920.2005.00906.x

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., & Minh, B. Q. (2016). W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*, *44*(W1), W232–W235. https://doi.org/10.1093/nar/gkw256

Trivedi, P., Delgado-Baquerizo, M., Anderson, I. C., & Singh, B. K. (2016). Response of soil properties and microbial communities to agriculture: Implications for primary productivity and soil health indicators. *Frontiers in Plant Science*, *7*(JULY2016), 1–13. https://doi.org/10.3389/fpls.2016.00990

Tveit, A., Schwacke, R., Svenning, M. M., & Urich, T. (2013). Organic carbon transformations in high-Arctic peat soils: Key functions and microorganisms. *ISME*

Journal, *7*(2), 299–311. https://doi.org/10.1038/ismej.2012.99

van Kessel, M. A. H. J., Speth, D. R., Albertsen, M., Nielsen, P. H., Op den Camp, H. J. M., Kartal, B., Jetten, M. S. M., & Lücker, S. (2015). Complete nitrification by a single microorganism. *Nature*, *528*(7583), 555–559. https://doi.org/10.1038/nature16459

Wang, H., Bagnoud, A., Ponce-Toledo, R. I., Kerou, M., Weil, M., Schleper, C., & Urich, T. (2021). Linking 16S rRNA Gene Classification to *amoA* Gene Taxonomy Reveals Environmental Distribution of Ammonia-Oxidizing Archaeal Clades in Peatland Soils. *MSystems*, *6*(4). https://doi.org/10.1128/msystems.00546-21

Weber, E. B., Lehtovirta-Morley, L. E., Prosser, J. I., & Gubry-Rangin, C. (2015). Ammonia oxidation is not required for growth of group 1.1c soil Thaumarchaeota. *FEMS Microbiology Ecology*, *91*(3), 1–7. https://doi.org/10.1093/femsec/fiv001

Wild, B., Gentsch, N., Capek, P., Diáková, K., Alves, R. J. E., Bárta, J., Gittel, A., Hugelius, G., Knoltsch, A., Kuhry, P., Lashchinskiy, N., Mikutta, R., Palmtag, J., Schleper, C., Schnecker, J., Shibistova, O., Takriti, M., Torsvik, V. L., Urich, T., … Richter, A. (2016). Plant-derived compounds stimulate the decomposition of organic matter in arctic permafrost soils. *Scientific Reports*, *6*(February), 1–11. https://doi.org/10.1038/srep25607

Winogradsky, Sergei. (1890). Recherches sur les organisms de la nitrification. *Ann. Inst. Pasteur*, *4*, 213--231.

Xia, F., Wang, J.-G., Zhu, T., Zou, B., Rhee, S.-K., & Quan, Z.-X. (2018). Ubiquity and Diversity of Complete Ammonia Oxidizers (Comammox). *Applied and Environmental Microbiology*, *84*(24), AEM.01390-18. https://doi.org/10.1128/AEM.01390-18

Yang, Y., Wang, Z., Hu, Y., & Zeng, Z. (2017). Irrigation frequency alters the abundance and community structure of ammonia-oxidizing archaea and bacteria in a northern Chinese upland soil. *European Journal of Soil Biology*, *83*, 34–42. https://doi.org/10.1016/j.ejsobi.2017.09.005

Zabaloy, M. C., Allegrini, M., Tebbe, D. A., Schuster, K., & Gomez, E. del V. (2017). Nitrifying bacteria and archaea withstanding glyphosate in fertilized soil microcosms. *Applied Soil Ecology*, *117–118*(March), 88–95. https://doi.org/10.1016/j.apsoil.2017.04.012

Zhalnina, K., Quadros, P. D. D., Gano, K. A., Davis-Richardson, A., Fagen, J. R., Brown, C. T., Giongo, A., Drew, J. C., Sayavedra-Soto, L. A., Arp, D. J., Camargo, F. A. O., Daroub, S. H., Clark, I. M., McGrath, S. P., Hirsch, P. R., & Triplett, E. W. (2013). Ca. Nitrososphaera and bradyrhizobium are inversely correlated and related to agricultural practices in long-term field experiments. *Frontiers in Microbiology*, *4*(MAY), 1–13. https://doi.org/10.3389/fmicb.2013.00104

Zhalnina, K. V., Dias, R., Leonard, M. T., Quadros, P. D. de, Camargo, F. A., Drew, J. C.,

Farmerie, W. G., Daroub, S. H., & Triplett, E. W. (2014). Genome Sequence of Candidatus Nitrososphaera evergladensis from Group I.1b Enriched from Everglades Soil Reveals Novel Genomic Features of the Ammonia-Oxidizing Archaea. *PloS One*, *9*(7), e101648.

Zhang, B., Wu, X., Tai, X., Sun, L., Wu, M., Zhang, W., Chen, X., Zhang, G., Chen, T., Liu, G., & Dyson, P. (2019). Variation in Actinobacterial Community Composition and Potential Function in Different Soil Ecosystems Belonging to the Arid Heihe River Basin of Northwest China. *Frontiers in Microbiology*, *10*(September), 1–11. https://doi.org/10.3389/fmicb.2019.02209

Zhao, J., Wang, B., Zhou, X., Alam, M. S., Fan, J., Guo, Z., Zhang, H., Gubry-Rangin, C., & Zhongjun, J. (2022). Long-Term Adaptation of Acidophilic Archaeal Ammonia Oxidisers Following Different Soil Fertilisation Histories. *Microbial Ecology*, *83*(2), 424–435. https://doi.org/10.1007/s00248-021-01763-2

Zheng, J., Tao, L., Dini-Andreote, F., Luan, L., Kong, P., Xue, J., Zhu, G., Xu, Q., & Jiang, Y. (2022). Dynamic Responses of Ammonia-Oxidizing Archaea and Bacteria Populations to Organic Material Amendments Affect Soil Nitrification and Nitrogen Use Efficiency. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.911799

Zhou, X., Fornara, D., Wasson, E. A., Wang, D., Ren, G., Christie, P., & Jia, Z. (2015). Effects of 44 years of chronic nitrogen fertilization on the soil nitrifying community of permanent grassland. *Soil Biology and Biochemistry*, *91*, 76–83. https://doi.org/10.1016/j.soilbio.2015.08.031

7. SUPPLEMENTARY MATERIAL

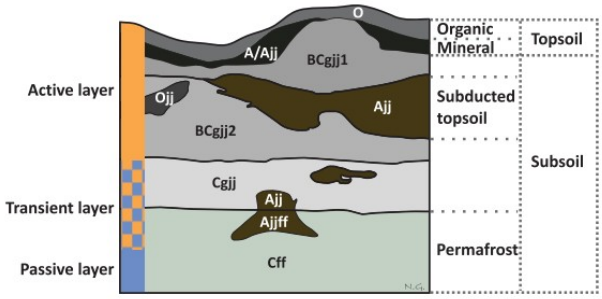


Figure 1: Diagram with the soil horizons in the arctic tundra soils (from Gentsch et al., 2015).

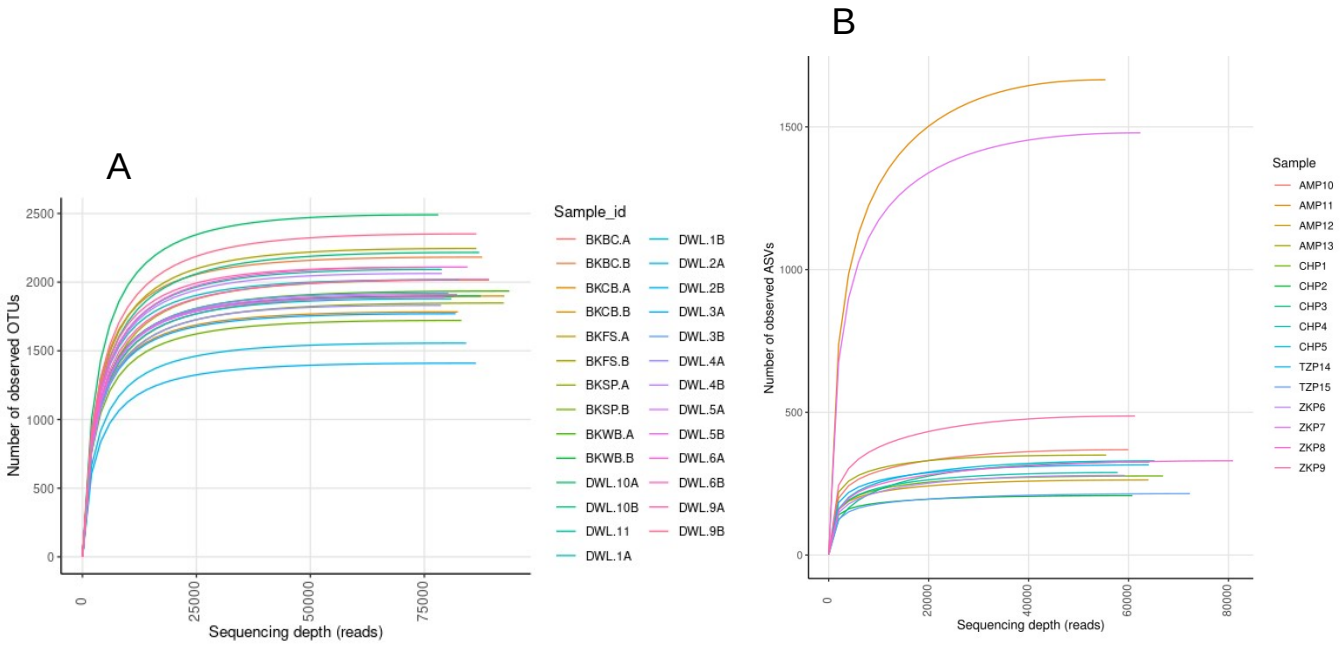


Figure 2: Rarefaction curves from the ranch soil samples (A) and the arctic tundra samples (B).

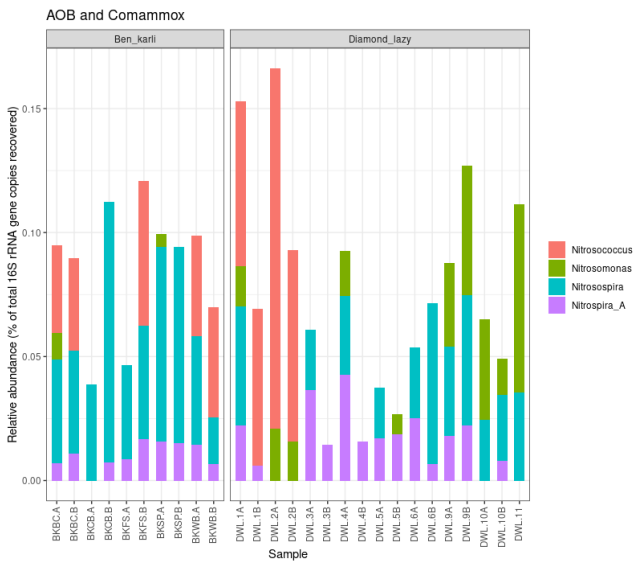


Figure 3: Barplot with the relative abundances of AOB and Comammox in the ranch soil samples.

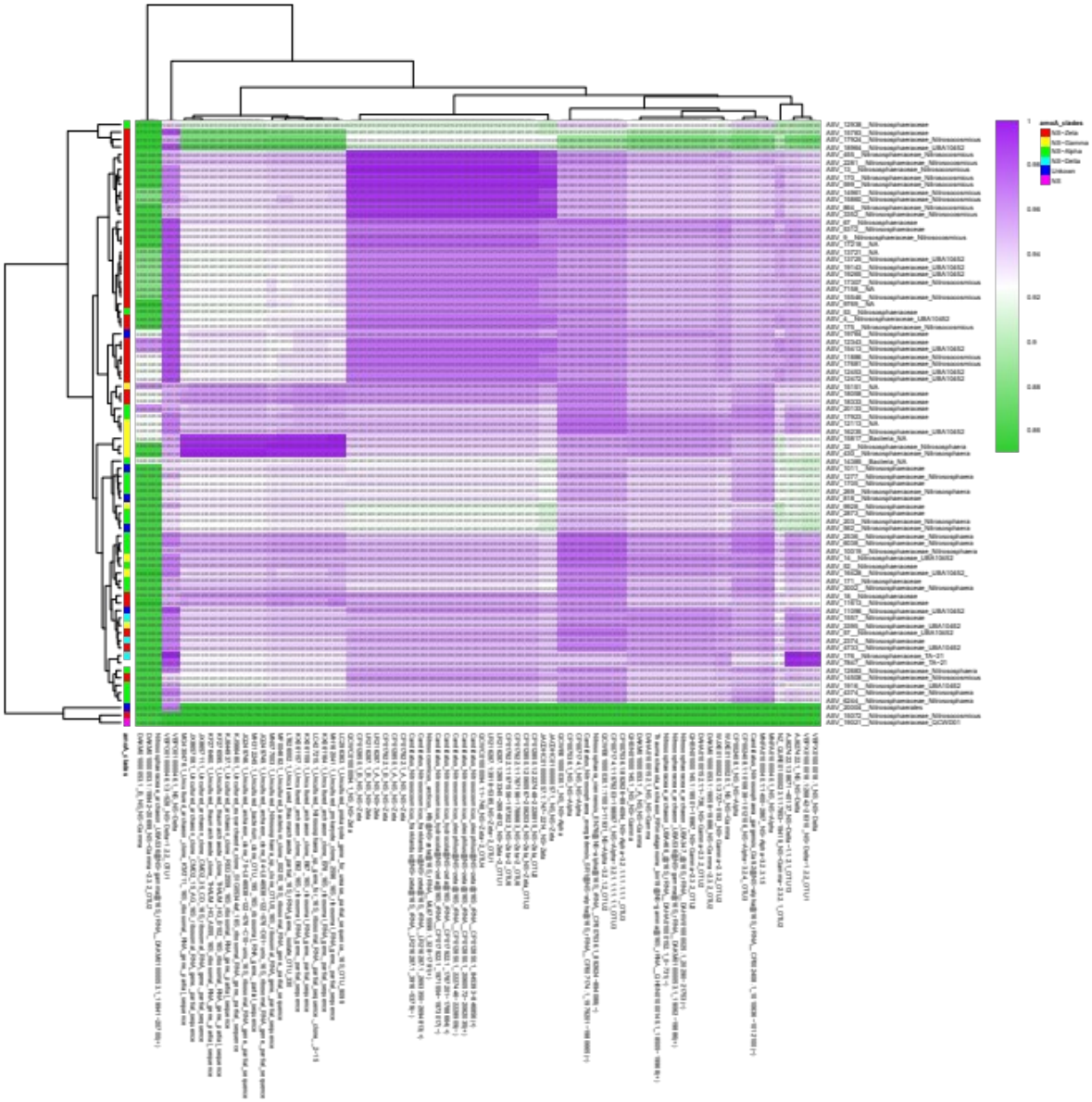


Figure 4: Heatmap with hierarchical clustering depicting the nucleotide sequence identity between the aligned sequences from the enriched database (X-axis) and the *Nitrososphaeraceae* ASVs (Y-axis). Each sequence is colour coded with their *amoA* clade classification.

Pool	Location	Water content	OC bulk soil	total N bulk soil	NH4	NO3	Cex	Nex	pH
P01	Cherskii	10.42	9.17	4.61	6.35	1.58	9.01	4.69	5.17
P02	Cherskii	10.70	9.22	4.95	7.83	1.36	9.22	6.10	5.25
P03	Cherskii	10.87	9.38	4.58	8.05	5.42	9.74	5.48	4.64
P04	Cherskii	9.54	6.60	2.32	6.08	4.09	7.90	3.52	5.26
P05	Cherskii	10.24	7.37	3.39	7.43	4.13	8.26	5.00	5.72
P06	Zackenberg	10.33	8.57	4.63	7.32	2.12	8.54	6.19	5.29
P07	Zackenberg	10.10	7.98	3.88	4.60	1.00	5.26	2.91	6.02
P08	Zackenberg	10.92	9.50	5.16	5.73	4.81	7.87	4.50	5.29
P09	Zackenberg	10.07	8.77	4.37	6.22	3.70	8.48	5.83	5.29
P10	Taymyr	10.09	6.92	2.58	3.64	5.33	7.91	3.99	6.55
P11	Taymyr	10.49	8.66	4.01	5.80	5.26	11.92	8.77	6.20
P12	Taymyr	10.50	8.42	4.26	5.67	5.78	9.52	5.98	5.93
P13	Taymyr	10.79	9.16	4.55	6.55	6.29	10.07	6.29	5.98
P14	Tazowvsky	10.61	9.50	4.30	5.52	5.65	10.10	6.47	4.86
P15	Tazowvsky	10.65	9.10	4.13	6.65	5.84	9.42	6.05	4.92

Table 1: Soil physico-chemical parameters used for the correlation analysis in the arctic tundra soils. Data was collected from the references cited in section 3.1.2.

Pools	Location	Exoglucanase	Endochitinase	Exochitinase	Protease	Phenol oxidase	Peroxidase	Actual Protease	Actual Cellulase
P01	Cherskii	208.04	84.34	1001.08	360.01	2786.49	8280.59	3.09	81.68
P02	Cherskii	224.31	48.38	857.49	728.09	3135.89	11163.95	3.35	78.10
P03	Cherskii	344.81	116.42	1247.60	670.49	2143.50	6242.05	2.88	95.55
P04	Cherskii	33.54	14.96	257.53	117.04	826.33	3135.75	0.45	10.45
P05	Cherskii	75.80	19.43	97.15	123.95	2206.80	7542.88	0.12	11.78
P06	Zackenberg	472.33	44.24	692.55	691.42	2245.00	7622.08	1.20	42.83
P07	Zackenberg	266.70	17.39	378.48	603.38	1888.15	6102.38	0.68	27.54
P08	Zackenberg	625.52	270.94	1273.47	1710.19	7290.48	20467.00	2.55	390.81
P09	Zackenberg	325.81	114.69	443.23	452.78	688.06	1959.86	3.44	48.64
P10	Taymyr	66.59	26.05	355.44	89.50	834.95	1533.65	1.54	12.39
P11	Taymyr	123.27	78.81	608.66	259.12	507.62	1324.27	3.97	62.07
P12	Taymyr	474.03	130.06	902.07	282.59	678.26	1065.53	2.56	101.97
P13	Taymyr	233.24	69.65	979.08	390.02	1197.04	2679.10	6.06	274.91
P14	Tazowvsky	465.68	60.91	1213.90	267.13	1217.65	2829.12	3.26	272.78
P15	Tazowvsky	396.78	89.68	1161.23	187.68	2072.07	5671.77	3.41	289.91

Table 2: Enzymatic activity data used for the correlation analysis in the arctic tundra soils. Data was collected from the references cited in section 3.1.2.