

Genome-encoded metabolic potential of the
***Nitrosocosmicus* genus and related ammonia-**
oxidizing archaea

by

Calvin Cornell

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2024

© Calvin Cornell 2024

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Since the discovery of ammonia-oxidizing archaea (AOA), genomic and experimental evidence suggests that mixotrophic growth may complement chemolithotrophic ammonia oxidation for some AOA representatives. Although members of the *Candidatus Nitrosocosmicus* genus have been implicated in the use of alternate sources of energy and carbon, stronger genomic evidence is needed to support testing of hypotheses with robust experimental design. This study involved analysis of 58 genomes spanning all four known AOA groups (i.e., I.1a, I.1b, 1.1a-associated, thermophilic AOA; ThAOA). The analyses focused primarily on transporters and enzymes, linking the former to potential substrates from the transporter classification database (TCDB) and the latter to KEGG, Carbohydrate-Active enZymes (CAZymes), and pathways from GapSeq. Correlated with genome size, the results demonstrate that genomes of the I.1b group, including those from the *Ca. Nitrosocosmicus* genus, showed the highest abundances of exclusive (only found within this group) and preferred (primarily associated with this group) proteins for alternative metabolism, followed by ThAOA, I.1a, and 1.1a-associated. There was extensive functional heterogeneity among representative genomes, particularly for *Ca. Nitrosocosmicus* and *Nitrososphaera* genera. The group I.1b genomes encoded for the potential to use nitrogenous substrates such as urea, alanine, glycine, asparagine, and glutamine. Although *Ca. Nitrosocosmicus* genomes were more commonly associated with use of TCA cycle intermediates, such as citrate, malate, oxaloacetate, and succinate, this potential was likewise found in ThAOA and I.1a representatives. Several I.1b representative genomes encoded enzymes for metabolism of ureidoacrylates and carbamate, CAZymes associated with glycoside hydrolases, genes implying oxidation of sulfur and manganese, and polyamine metabolism. However, genes for synthesis and degradation of polyamines were common to many AOA genomes analyzed. Overall, the results suggest that I.1b-associated representatives encode for mixotrophic metabolism, and future research should verify associated substrate predictions experimentally. Given broad distributions of AOA within terrestrial and aquatic environments, these findings have implications for biogeochemical cycling of carbon and nitrogen on a global scale.

Acknowledgements

I want to express my deepest thanks to my supervisor Dr. Josh Neufeld for help and guidance during this project and opening my eyes to the beautiful world of bioinformatics. Also, a huge thanks to him and my committee, Drs. Andrew Doxey and Laura Hug, for many insightful opinions and suggestions. Thanks to Alex Umbach of the Neufeld Lab for helping me learn about the server and Dr. Cheryl Duxbury for being a fantastic TA supervisor and for helping and motivating me when times got tough.

Heartfelt thanks to my parents, best friend Farhan Tasin, and girlfriend Allison Geyer, for unwavering support, help, and patience during these last two years. Your belief in me and encouragement during times of difficulty were crucial for the completion of this project.

Finally, a thank you to McDonald's for always being there for me in the good and the bad times and my beautiful kitten Somersault looking over me and stealing my food. And to John Steinbeck and Cormac McCarthy, your literature and stories filled me with confidence, inspiration, and encouraged my love for writing.

Table of contents

Author's Declaration	ii
Abstract	iii
Acknowledgements	iv
List of figures.....	vi
List of abbreviations	viii
<i>Chapter 1</i>	<i>1</i>
<i>Introduction and literature review</i>	<i>1</i>
1.1 Nitrification	1
1.2 AOA background	1
1.2.1 Phylogenetics and ecology	1
1.2.2 Evolutionary history	4
1.3 AOA metabolism.....	5
1.3.1 Chemolithoautotrophy.....	5
1.3.3.1 Nitrogen substrates	7
1.3.3.2 Carbon substrates	9
1.4 Mixotrophy.....	10
1.5 Limitations of previous work and direct experimentation	11
1.6 Implications of AOA metabolism	12
1.7 Bioinformatics	13
<i>Chapter 2</i>	<i>16</i>
<i>Materials and Methods</i>	<i>16</i>
2.1 Genome selection	16
2.2 Genome quality	16
2.3 Annotations	16
2.4 Phylogenetic tree	17
2.5 Functional annotation	18
2.5.1 KEGG annotations	19
2.5.2 Categorization of KEGG IDs	20
2.5.3 Reactions	21
2.5.3.1 Functional redundancy removal	21
2.5.3.1 Reaction sorting.....	21
2.5.4 Transporters.....	22

2.5.5 Pathways	24
2.5.6.1 Exclusive analysis.....	25
2.5.6.2 Lineage specificity analysis (LSP)	25
2.6 Visualizations.....	27
2.6.1 Principal coordinate analysis (PCoA)...	27
2.6.2 Phylogenetic trees figure, heatmaps, and Venn diagrams	28
<i>Chapter 3</i>	<i>29</i>
<i>Results & Discussion</i>	<i>29</i>
3.1 Overview.....	29
3.1.1 Genomic and protein sizes.....	29
3.1.2 Broad metabolic differences	30
3.1.3 Averages and exclusivity.....	32
3.1.4 Evolutionary explanations and heterogeneity.....	34
3.2 Differences in metabolic potentials	35
3.2.1 Reactive oxygen species (ROS) enzymes	35
3.2.2 Transporters	37
3.2.3 Cofactors.....	41
3.2.3.1.1 Complex 1.....	43
3.2.3.1.2 lipoquinones.....	45
3.2.3.1.3 Cytochrome C	46
3.2.3.1.4 Pyrroloquinoline quinone (PQQ)....	47
3.2.4 Energy and fermentation.....	48
3.2.5 Anabolism	51
3.3 Substrate differences.....	53
3.3.1 Nitrogen	53
3.3.1.1 Urea.....	53
3.3.1.2 Peptidases.....	55
3.3.1.3 Polyamines.....	60
3.3.1.4 Other nitrogenous substrates.....	63
3.3.2 Carbon substrate	65
3.3.2.1 Glycolysis and TCA intermediates	65
3.3.2.2 Single carbon substrate	72
3.3.2.3 Sugars.....	73
3.3.2.3.1 Simple sugars.....	73
3.3.2.3.2 Polysaccharides.....	75
3.3.2.4 Alcohols	76
3.3.3 Inorganic substrates	77
3.3.3.1 Manganese	78
3.3.3.2 Sulfite.....	78
3.3.3.3 Bacterioferritin.....	79
3.3.3.4 Mercury.....	80
3.4 Mixotrophy	80
3.5 Alternatives to mixotrophy	80
3.6 Benefits of mixotrophy	82

<i>Chapter 4</i>	84
<i>Conclusion</i>	84
4.1 Concluding remarks	84
4.2 Dark matter and future research	85
4.3 Importance.....	89
<i>References</i>	91

List of figures

Figure 1.1 Phylogenetic tree of the 58 species of ammonia-oxidizing archaea (AOA).....	4
Figure 2.1 A subset of the phylogenetic tree from Figure 1.1, showing clades for lineage specificity analysis (LSP)	27
Figure 3.1 Scatter plot showing the relationship between genome size and number of proteins.....	30
Figure 3.2 PCoA plot showing overall metabolic repertoire differences between the 58 AOA genomes..	31
Figure 3.3 Boxplots and Venn diagrams showing the abundance and spread of enzymes and pathways associated with metabolism.....	33
Figure 3.4 Presence and absence heatmap for the detoxification and removal of reactive oxygen species (ROS).....	36
Figure 3.5 Presence and absence heatmap for I.1b specific proteins associated with CAZymes (A), peptidases (B) and transporters (C).....	39
Figure 3.6 Heatmap showing the presence of I.1b specific pathways as predicted to be complete by GapSeq.....	44
Figure 3.7 Presence and absence heatmap showing I.1b specific reactions associated with production of energetic products.....	50
Figure 3.8 Heatmap showing the presence and absence of reactions associated with producing biosynthetic products.....	52
Figure 3.9 Heatmap showing the genomes of AOA that are predicted to both encode for substrates that yield free ammonia and the transporter to transport said substrate across.....	54
Figure 3.10 Proportional heatmap showing the degree of polyamine pathway completion as predicted by GapSeq.....	61
Figure 3.11 Heatmap showing reactions associated with the production or usage of anaplerosis (i.e., intermediates of TCA and 3-HP/4-HB cycle).....	68

List of abbreviations

3HP/4HB: 3-hydroxypropionate/4-

hydroxybutyrate

AA: auxiliary activities

Amt: ammonia transporter

Anammox: anaerobic ammonia oxidizing
bacteria

AOA: ammonia-oxidizing archaea

AOB: ammonia-oxidizing bacteria

AOM: ammonia-oxidizing microorganism

ATP: adenosine triphosphate

Ca: Candidatus

CAZyme: carbohydrate-active enzymes

CBM: carbohydrate-binding modules

CE: carbohydrate esterases

CH₄: methane

CO₂: carbon dioxide

CO₃²⁻: carbonate

CoA: coenzyme A

COG: clusters of orthologous genes

Comammox: complete ammonia oxidation

CSA: coding sequences

CTP: cytidine triphosphate

DRAM: Distilled and Refined Annotation of
Metabolism

EC: enzyme commission number

ETC: electron transport chain

FADH₂: flavin adenine dinucleotide + 2

hydrogen (H)

Fe²⁺: iron(II)

Fe³⁺: iron(III)

FLOW-SIP: flow-through stable-isotope probing

GC content: guanine-cytosine content

GDGT: glycerol dialkylglycerol tetraether lipid

GH: glycoside hydrolases

GH: glycoside hydrolases

GT: glycosyl transferases

GTP: guanosine-5'-triphosphate

H₂O₂: hydrogen peroxide

HAO: hydroxylamine oxidoreductase

HAOB: heterotrophic ammonia-oxidizing
bacteria

HCO₃⁻: bicarbonate

HGT: horizontal gene transfer

KAAS: KEGG Automatic Annotation Server

KEGG: Kyoto Encyclopedia of Genes and
Genomes

Mya: millions of years ago

MAG: metagenome-assembled genome

N₂: dinitrogen gas

N₂O: nitrous oxide

NADH: nicotinamide adenine dinucleotide
(NAD) + hydrogen (H)

NADPH: nicotinamide adenine dinucleotide
phosphate + hydrogen (H)

NCBI: National Center for Biotechnology
Information

NH₂OH: hydroxylamine

NH₃: ammonia

NH₄⁺: ammonium

NO: nitric oxide

NO₂⁻: nitrite

NO₃⁻: nitrate

NOB: nitrite-oxidizing bacteria

ORF: open reading frame

PCoA: principal coordinates analysis

PCR: polymerase chain reaction

PL: polysaccharide lyases

PLFA: phospholipid fatty acid analysis

PPS: pseudo-periplasmic space

PQQ: pyrroloquinoline quinone

qPCR: quantitative polymerase chain reaction

RASTtk: Rapid Annotation using Subsystem
Technology

ROS: reactive oxygen species

S-layer: surface layer

SIP: stable-isotope probing

sp: species

TCA: tricarboxylic acid

TCDB: Transporter Classification Database

TTP: thymidine triphosphate

V-type ATPase: vacuolar-type ATPase

WWTP: wastewater treatment plant

Chapter 1

Introduction and literature review

1.1 Nitrification

Nitrogen is a macronutrient required by all organisms, and its concentration, species, and distribution influences productivity, abundance, and diversity of communities in all environments [1, 2]. The most biologically useful form of nitrogen, which depending on pH is ammonium (NH_4^+) or ammonia (NH_3), can enter the environment by nitrogen-fixing microorganisms, ammonification, degradation of organic nitrogen-containing compounds (e.g., proteins or DNA), or from non-biological sources such as lightning and anthropogenic industries such as farming via fertilizers or untreated human waste from sewage. Ammonia can be used by cells either in anabolic metabolism, such as synthesizing proteins and nucleic acids [3], or in catabolic metabolism, such as nitrification. Nitrification is the aerobic oxidation of ammonium (NH_4^+) to nitrite (NO_2^-) and then to nitrate (NO_3^-) [4] with the rate-limiting step of nitrification being the oxidation of NH_3 to NO_2^- [5]. Two groups of microorganisms mediate this first step: ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AOA). The second step, nitrite (NO_2^-) oxidation to nitrate (NO_3^-), is accomplished by nitrite-oxidizing bacteria (NOB; 6). In addition, a recently discovered third group of aerobic ammonia oxidizers belongs to the *Nitrospira*, with representatives capable of complete ammonia oxidation (comammox) to nitrate [6].

1.2 AOA background

1.2.1 Phylogenetics and ecology

For over a century, AOB were considered to be the only aerobic ammonia oxidizers [7, 8]. In 2004, an archaeal metagenome assembly scaffold was found containing a homolog to the enzyme that catalyzes the first step in ammonia oxidation (ammonia monooxygenase; *amoA*), associated with a Sargasso Sea sample [9]. A subsequent soil metagenomic survey found an archaeal fosmid clone that

encoded homologs to bacterial *Amo* subunits [10]. Both studies proposed that some archaea could catalyze ammonia oxidation. These observations were confirmed by the isolation and characterization of *Nitrosopumilus maritimus* SCM1, which was cultivated from marine aquarium gravel, demonstrating chemolithoautotrophic growth with ammonia as energy and electron source, with bicarbonate as a source of carbon [11].

Since the discovery of *N. maritimus* SCM1, all known AOA are associated with the phylum *Thermoproteota* (formally known as *Thaumarchaeota*) and the class *Nitrososphaerota* [12]. Species of AOA are typically classified into four groups (families; Figure 1.1). Most AOA species are associated with Group I.1a (*Ca. Nitrosopumilales*) and Group I.1b (*Ca. Nitrososphaerales*), commonly referred to as the marine and soil groups, respectively [5, 13–15] on account of being found predominantly within these environments. Other AOA are grouped within the I.1a-associated (*Ca. Nitrosotaleales*) [16, 17] and thermophilic AOA (ThAOA; *Ca. Nitrosocaldales*) [18], with representatives commonly found in acidic and hot environments, respectively, although less is known about the physiology and metabolism of these two groups [19].

Metagenomic surveys have detected AOA in nearly every niche on Earth [19] and, for many habitats, AOA and the phylum *Thermoproteota* are among the most abundant phyla of archaea present [20, 21]. Such high abundance has been documented for mesophilic soils [21–23] and oceans [11, 23], as well as extreme environments including hot (up to 97°C) and extremely acid (pH 2.5) soils [24, 25]. Such high AOA abundances might be related to their ability to adapt to new conditions and environments [26], differences in membrane stability [27, 28], and some species possessing ammonia affinities higher than for AOB [29–31]. Some species' ammonia affinities are higher (i.e., lower K_m values) than AOB which might provide them with an advantage over AOB within oligotrophic environments. However, several members of the family *Nitrososphaeraceae*, in particular species of *Ca. Nitrosocosmicus*, are adapted to relatively high ammonia concentrations [32–34]. Such ammonia tolerance has led some members of the *Nitrososphaeraceae* to outnumber AOB in several hypoxic reactors [35] and eutrophic WWTPs [21, 36, 37].

Most known AOA show a preference for lower oxygen concentrations than AOB [30, 38] due to many representatives lacking a catalase [39]. Likewise, many AOA are associated with optimum pH values ranging from acidic to slightly basic [20, 40, 41]. However, these preferred pH ranges pose a problem for the concentration of ammonia, the preferred substrate of AOA [42], which below pKa 9.25 becomes increasingly dominated by the less preferred NH_4^+ . Which might not pose a problem to certain group of genera of AOA who demonstrate a higher affinity of AOA for ammonia [20, 43], due to the presence of ammonia concentrating methods, such as the S-layer [44], the high-affinity amt-type ammonium transporter [32, 34], or a high surface area to volume ratio (SA/V) [45]. Yet several genera of AOA, in particular those within *Ca. Nitrosocosmicus* are predicted to lack the S-layer, high-affinity amt-type ammonium transporter, and likewise have significantly smaller SA/V due to their larger diameters [32–34] and thus might have a problem within low ammonia concentrations.

Environments having more accessible copper over iron is consistent with the higher abundance of copper-dependent enzymes in AOA that are associated with crucial functions such as electron transfer and ammonia oxidation [46, 47]. Therefore, it is predicted if oxygen, ammonia, or copper drops below required concentrations certain species, in particular those that are non-motile such as many *Ca. Nitrosocosmicus* [32–34], could either enter a stationary phase or could resort to using alternative substrates to support their growth, nutrient, or energy needs.

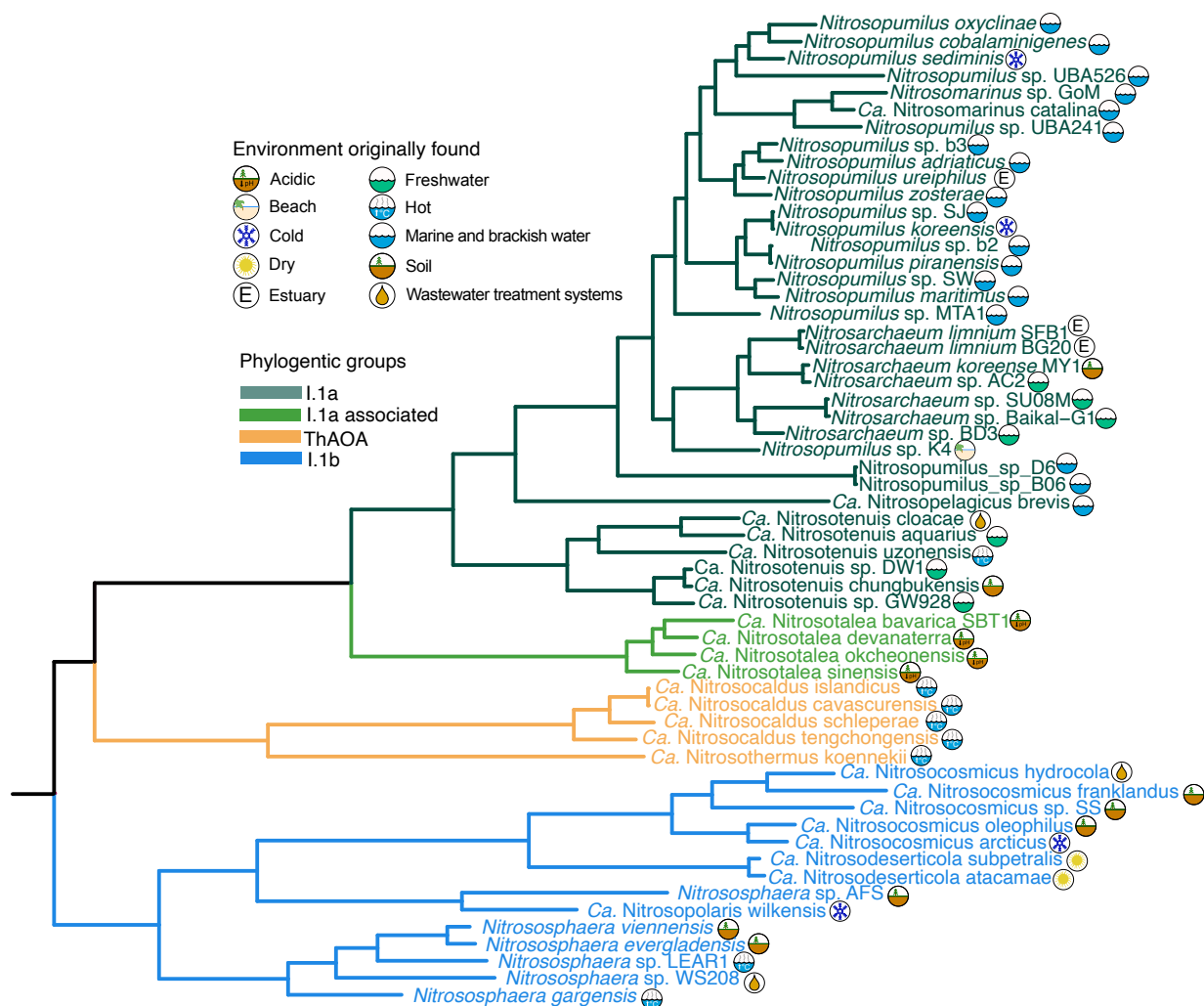


Figure 1.1 Phylogenetic tree of the 58 species of ammonia-oxidizing archaea (AOA) used in this thesis research. The phylogenetic tree was generated using GToTree (v1.6.31) and the option to use the 76 single-copy genes for archaea as marker genes [48]. Visualization of the tree and the colorization of the lines was achieved using ggtree [49]. The four colours of the lines represent the four groups (i.e., families) associated with the 58 species. Whereas the symbols following the species names represent the type of habitat where the species or the metagenomically assembled genomes (MAGs) were first discovered or detected. This figure and symbols were adapted from a previous publication [50].

1.2.2 Evolutionary history

Despite their recent discovery, phylogenetic research predicts that AOA originated from a terrestrial non-AOA ancestor [51] between 1,165 to 2,300 million years ago (Mya) [52], in association with geothermal environments [53, 54]. This basal species was likely thermophilic, autotrophic, aerobic,

and capable of ammonia oxidation [19]. The transition to other aquatic and terrestrial habitats is predicted to have occurred in a stepwise manner from geothermal locations, to mesophilic terrestrial environments, then to freshwater and estuaries, to upper oceans, and finally to deep oceans and sediments [52, 55]. Such a progression has been linked to major geological events, such as glaciation and oxygenation, in the same order in which AOA expanded and likely occurred within the Neoproterozoic era between 1000–543 Mya [51, 52]. Adaptation to these new environments was likely facilitated by high levels of horizontal gene transfer (HGT) that have occurred from bacterial, archaeal, and eukaryotic donors, which has allowed many AOA representatives to obtain genes for metabolism, substrate transport, membrane biogenesis, nutrient acquisition, and energy conservation [52, 56–58]. For example, HGT of vacuolar-type ATPase (V-type ATPase), potentially from a member of *Thermoplasmatales* likely enabled some AOA to survive in high-pressure and acidic environments on account of its presumptive function in pumping protons out at low pH [59]. Under high pressures, the pKa of many weak acids decreases, particularly those of biological systems or those located within the cytosol [60]. This releases more free protons within the cell, decreases pH, and threatens a cell's metabolism. To combat this, the V-type ATPase pumps these extra protons out of the cell. While the marine species of AOA diversified within the oceans many lineages, particularly within Group I.1b (*Ca. Nitrososphaerales*), also underwent extensive gene duplication and loss, suggesting further niche-specific adaptation and specialization, with genes encoding for coenzyme metabolism and carbohydrate transport commonly duplicated [61, 62].

1.3 AOA metabolism

1.3.1 Chemolithoautotrophy

Known AOA are recognized as being chemolithoautotrophic, oxidizing NH_3 to NO_2^- for energy and fixing bicarbonate (HCO_3^-) into biomass via a modified version of the 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle [11, 19, 20, 32, 33]. Evidence to support this metabolic strategy includes AOA genomes containing required genes for carbon fixation and coding for the three subunits of *Amo* [19, 32–34]. In addition, all known cultured species grow in minimal media without organic carbon,

with cell numbers closely following NO_2^- production, demonstrating that AOA can derive energy solely from autotrophic nitrification [19, 32–34].

1.3.2 Evidence for alternative metabolic strategies

Several studies challenge the strictness of a chemolithoautotrophic metabolism for AOA representatives. For example, Mußmann *et al.* showed that the abundance of *Ca. Nitrosocosmicus* was two orders of magnitude higher than what could be sustained by autotrophic ammonia oxidation alone, based on comparing predicted abundances via metabolic modelling and actual abundances via quantitative PCR (qPCR) and fluorescence *in situ* hybridization (FISH) images [21]. Other studies have shown that, despite *amoA* gene transcripts being present, ammonia oxidation was not detected within agricultural microcosm soil experiments [63]. Likewise, adding inorganic ammonia to soil failed to stimulate archaeal oxidation of ammonia [24, 64, 65]. Additionally, the abundances of soil AOA increased despite the presence of acetylene, an inhibitor of nitrification, suggesting that the AOA were relying on another form of metabolism instead of ammonia oxidation [63, 66]. Because terrestrial environments are typically dominated by members of the I.1b group [63], these observations based on terrestrial systems suggest alternate metabolic strategies may be involved. Other studies have failed to show bicarbonate uptake to support autotrophic growth for I.1b archaea. For example, both Mußmann *et al.* and Sauder *et al.* demonstrated using microautoradiography fluorescence *in situ* hybridization (MAR-FISH) that cells of the *Ca. Nitrosocosmicus* genus within WWTP biofilm samples failed to assimilate bicarbonate, even though they did so within enrichment cultures [21, 33].

1.3.3 Substrates that may support alternative metabolic strategies

That some AOA may not be strict chemolithoautotrophs has been demonstrated both genomically, via the discovery of genes for the use and transport of alternative substrates, and experimentally, with exposure to different substrates resulting in enhanced activity and/or growth. These substrates typically are nitrogen-based, such as urea and amino acids, or those that are commonly

considered carbon sources, such as tricarboxylic acid (TCA) cycle intermediates and single carbon compounds, each of which will be examined below.

1.3.3.1 Nitrogen substrates

Members of group I.1b commonly possess genes encoding proteins involved in the transport and catabolism of urea, amino acids, and peptides, and these are encoded by all members of the *Ca. Nitrosocosmicus* genus [19, 32, 33, 67], multiple species from *Nitrososphaera* [51, 57, 61, 68], and *Candidatus Nitrosopolaris wilkensis* [69]. Homologs of these genes have also been reported in other groups from the I.1a, including *Nitrosopumilus* [26, 70, 71], *Nitrosopelagicus* [72], and ThAOA members *Ca. Nitrosothermus* and *Ca. Nitrosocaldus* [73]. However, these genes are rarely present in species from these latter groups [74], yet this might also be due to relatively few studies testing their distributions. Experimentally, many group I.1b species have been observed to use urea to stimulate growth, such as those from the genera *Ca. Nitrosocosmicus* [19, 32–34] and *Nitrososphaera*, *Nitrososphaera gargensis* Ga 9.2 [75], and *Nitrososphaera viennensis* EN76 [68]. Several species from other groups also use urea, including group I.1a representatives *Nitrosopumilus piranensis* [76] and *Nitrosopumilus theophilus* [77], yet others within this group do not [76–78]. Similarly, growth stimulation by urea was shown for *Ca. Nitrosotalea devanattera*, a member of group I.1a-associated AOA [20]. This spectrum of effects, within and between groups, suggests that this metabolism depends on location and ecotype [79].

Several studies indicate that AOA-associated archaea can uptake and assimilate amino acids from both marine [80–82] and soil [65] habitats, suggesting that, at a minimum, certain species can reuse amino acids for protein synthesis. However, these studies did not identify the types of AOA responsible for this uptake. The only group I.1b species directly tested for amino acid use was *Nitrososphaera viennensis*, which failed to show any stimulation [83]. Therefore, whether group I.1b members can use amino acids is still unknown. Other groups of AOA have been sparsely tested for their ability to use amino acids, yet none showed growth stimulation, including species from the genus of *Nitrosopumilus*

[76], *Candidatus Nitrosopelagicus brevis* [84], and *Nitrosarchaeum koreense* [85] from group I.1a and *Candidatus Nitrosotalea devanaterrea* of group I.1a-associated [67].

Both *Nitrososphaera gargensis* [86] and *N. maritimus* SCM1 [87] of groups I.1b and I.1a, respectively, showed an ability to use cyanate. The presence of cyanate lyase in the *N. gargensis* genome reportedly came from an HGT event [86]. However, the finding of cyanate use by *N. maritimus* SCM1 is unexpected because the genome does not encode a known cyanate lyase homologue [87].

Other nitrogenous substrates relevant for AOA include polyamines, which are organic polycations with a carbon backbone and at least two amines attached. Polyamines are widespread in both eukaryotic and prokaryotic cells and serve many functions from assisting translation and transcription to nucleic acid synthesis and stabilization. However, despite being present at millimolar concentrations within cells [88], aquatic environments show average concentrations in the nanomolar or picomolar ranges [88, 89]. These differences suggest that they are rapidly broken down within the environment, biotically or abiotically. Similar to how some bacteria use polyamines as carbon, nitrogen, and energy sources [90–93], there is preliminary evidence to suggest that certain AOA can be involved in polyamine transformations. For example, *Ca. N. agrestis* sp. SS, a member of group I.1b, was the first AOA species shown to actively synthesize polyamines and secrete them into the extracellular space, a strategy that may permit them to survive in high ammonia environments [94]. However, whether AOA contain pathways for polyamine catabolism or whether the polyamine synthesis enzymes present within *Ca. N. agrestis* sp. SS can operate in reverse, are unknown but could be advantageous when ammonia is limiting. Although the ability to produce or use polyamines has not yet been identified for other AOA groups, this could be due to a lack of experiments as opposed to a lack of ability. However, Damashek *et al.* [95] showed that group I.1a members *Nitrosopumilus piranensis* D3C and *Nitrosopumilus adriaticus* NF5 could only oxidize putrescine in the presence of heterotrophic bacteria, which suggests syntrophy (i.e. co-metabolism), where the co-occurring bacterial species actively or passively degrade putrescine and release waste products that would then be available for the two AOA species.

1.3.3.2 Carbon substrates

When enriched in vitro I.1b cultures of *Ca. N. hydrocola* [33] and *N. viennensis* EN76 [68] were supplemented with TCA cycle intermediates succinate and malate, in addition to ammonia, they both showed higher NO_2^- production rates compared to controls. Similar experiments have not yet been done for the remaining three *Ca. Nitrosocosmicus* representatives, leaving it unknown whether organic carbon substrates can also promote their growth. However, *Ca. Nitrosocosmicus* species encode many of the same genes and pathways as *Ca. N. hydrocola*, suggesting that these other representatives may use these substrates as well [19, 32, 34]. Whether TCA cycle intermediates can be used by AOA as alternative carbon or energy sources is yet to be determined. In contrast, both *Ca. Nitrosotenuis chungbukensis* [96] and *Nitrosarchaeum koreense* MY1 [85], species within Group I.1a, were not stimulated by organic carbon substrates. Despite *N. viennensis* being stimulated by an extensive range of organic substrates, such as pyruvate, further analysis showed that less than 10% of cellular carbon came from pyruvate [68]. This indicated that even though pyruvate can be used by *N. viennensis*, it is not the preferred carbon source. However, caution is necessary when measuring AOA activity when exposed to certain alpha-keto acids as some can detoxify reactive oxygen species (ROS), particularly hydrogen peroxide, which can likewise trigger higher activity and growth, as documented for other AOA [39]. Yet, it is possible given that all known *Ca. Nitrosocosmicus* species encode genes for reactive oxygen species (ROS) detoxification, such as superoxide dismutase, peroxidase, and catalase [33]. Thus, these I.1b species may use these enzymes instead of relying on alpha-keto acids; whether other AOA encodes catalase or other homologs of it is not known.

Other carbon substrates that may serve as carbon or energy sources for AOA are single carbon molecules, such as formate, given that *Ca. N. hydrocola* encodes for formate dehydrogenase [33], and potentially methane given that the AmoA enzyme is a member of the copper-containing membrane monooxygenase (CuMMO) protein superfamily that includes methane monooxygenases [97]. Finally, members of I.1b can be positively correlated to total or organic carbon abundances, whereas other AOA

groups show anticorrelations [98, 99]. Such positive correlations invite experiments to test whether these substrates may be metabolized directly.

1.4 Mixotrophy

Many microorganisms possess more complex metabolic strategies than strict autotrophy or heterotrophy, with many being capable of using a mixture of organic and inorganic carbon substrates for assimilation and as energy sources. Typically, this ability to vary substrates used to support growth is termed mixotrophy [100]. However, mixotrophy can be divided into three subcategories. The first is necrotrophic (i.e., predatory) mixotrophy, a method often implied by planktonic protists' [101] or carnivorous plants (e.g., venus fly trap) and involves the uptake of organics/substrates by means of predation of living organisms and hence often resulting in the preys death [101]. The second is biotrophic (i.e., symbiotic) mixotrophy in which interactions (from parasitic to mutualistic) occur between the species of interest and other living organisms. Examples of such include hemiparasitic land plants and micro-eukaryotes such as metazoans or protists that live and form obligatory symbiosis with photosynthetic algae [100]. The third subcategory is absorbotrophic (i.e., osmotrophic) mixotrophy, which is the form in which cells actively or passively uptake organic matter found in the environment [101]. Although low molecular weight molecules are often preferred/used, some enzymatic pathways, such as those involved in proteinase activity, can occur outside the cell [100]. From preliminary evidence, specific AOA genomes, particularly those from *I.1b* and *Ca. Nitrosocosmicus* show the potential for absorbotrophic mixotrophy, which this thesis will endeavour to answer whether they encode for the potential for and whether there are differences between the groups and genera of AOA. In addition, there is an ongoing debate about what defines a species as mixotrophic. Some suggest that this applies only to species with the ability to switch from an inorganic to an organic carbon source, from autotrophy to heterotrophy (Muñoz-Marín, 2022). Others define mixotrophy as the capability to switch both the carbon source and the energy source and to source all major nutrients (e.g., nitrogen, sulfur) from multiple

different sources and compounds [102], even if the breakdown of substrates supplements, chemolithoautotrophy. This latter definition will be adopted for this thesis.

1.5 Limitations of previous work and direct experimentation

Although comparative genomic studies have the potential to reveal genome-encoded potential for substrate usage among characterized AOA genomes, there are limitations with prior studies that have used such approaches. Several prior studies did not analyze potential metabolism as the main focus of the analysis or used a relatively small collection of genomes. For example, one study found that marine and terrestrial AOA share well-conserved nitrogen and carbon metabolic pathways, and that terrestrial AOA possess extensive environmental adaptations that include detoxification and carbohydrate conversions [103]. However, this study used only seven genomes and did not include representatives from the *Ca. Nitrosocaldus*, *Ca. Nitrosomarinus*, *Ca. Nitrosopolaris*, *Ca. Nitrosotenuis*, *Ca. Nitrosocosmicus*, or *Ca. Nitrosothermus* genera. Another study analyzed 44 AOA genomes from representative ocean regions and showed that nutrients (e.g., ammonia, phosphorus) and resistance against hydrostatic pressure were selective pressures on marine AOA [104]. The potential to biosynthesize methylphosphonate was found in a large variety of species, even from those that showed prevalence in distinct regions of the ocean. However, this study focused on marine AOA and their adaptations to different marine habitats and did not include terrestrial AOA for comparison. Therefore, there is a need for a comprehensive study of AOA metabolic potential that includes all known AOA genomes available.

Another limitation of existing studies exploring the metabolic potential of known AOA relates to the analysis of growth stimulation, without also testing whether the taxa possess the ability to use the substrates directly. For example, *Nitrosopumilus* sp. HCA1 and sp. PS0 were initially thought capable of alternative metabolism given a requirement for alpha-keto acids, such as pyruvate, to grow in culture [105]. However, it was later found that instead of pyruvate being used as a carbon or energy source, it was required for scavenging ROS, such as hydrogen peroxide [39]. Likewise, many amino acids can be used by marine prokaryotes as compatible solutes, which can serve to stabilize cells and macromolecular

structures against high pressures and temperature-associated stress [106]. In addition, some substrates hypothesized to be alternate substrates, such as cyanate [86, 87] and urea [86, 87], can be rapidly and abiotically decomposed into both ammonia and bicarbonate [86, 87], both of which can then be used to support AOA growth. How much of these compounds is abiotically degraded is not yet clear [107]. Therefore, experimentally determining whether AOA members possess the ability to use such substrates for growth may be at risk of false positive results.

Many AOA species either have not been cultured by researchers in the lab or are only available as enrichment cultures alongside other non-AOA microorganisms. Such “satellite” microorganisms share metabolites with AOA, which may in turn stimulate growth through a process known as “cross-feeding” [5]. The presence of satellite microorganisms in AOA cultures makes directly demonstrating substrate usage by AOA technically challenging. Flow-through stable-isotope probing (FLOW-SIP) may help to determine whether substrates are assimilated by specific populations in mixed cultures [108]. This method retains cells on a stationary surface as media flows through, minimizing cell-cell nutrient exchanges. Even though this method may help in some circumstances, experimental logistics related to the use of multiple labelled substrates, low biomass, and relatively long generation times of AOA may limit use of this approach for AOA enrichment cultures. Other potential methods include assessing transcriptome [94] or metabolome [109] after incubations with different substrates.

1.6 Implications of AOA metabolism

Better understanding the metabolic diversity of AOA will help provide insight into their physiology, ecology, and evolutionary history [32]. Comparative genomics can provide insight into AOA physiology, better identifying their contributions to biogeochemical cycling of nitrogen and carbon [23, 53, 104]. Likewise, the presence of alternative metabolisms may provide an advantage for starvation tolerance, adaptation to extreme environments [110], or overcoming mismatches between required nutrients/resources and available supplies [104]. Identifying previously undescribed metabolic roles may help explain why specific AOA are enriched, compared to AOB and other microorganisms in various

environments [25, 111, 112], with implications for plant nutrition [23], eutrophication [33], human industry [113], and aquatic life [114]. Understanding metabolism can also help to form effective strategies for increasing ecosystem usage of nitrogen and reducing its losses, such as leaching into groundwater, which can affect downstream water bodies [115].

The presence of alternative metabolisms for AOA may also have implications for bioremediation. Previous research shows that several AOA can help remove anthropogenic pollutants, such as polycyclic aromatic hydrocarbons, organic fertilizers, and pesticides [26]. Some evidence suggests that AOA can transform xenobiotics compounds in co-metabolism with ammonia oxidation [116, 117], including sulfonamide antibiotics [118], micropollutants [119], tertiary amines, mianserin, and ranitidine [120]. However, whether all these compounds are metabolized by AOA to detoxify or use as a source of energy or carbon is currently unknown. Additionally, as previously mentioned, due to wet-lab experimental logistics, it is advantageous to find methods that minimize these drawbacks and identify potential substrates.

1.7 Bioinformatics

Bioinformatics, particularly functional annotations and analyses through comparative genomics, can serve as an important first step in identifying the metabolic potential of microorganisms based on predicted pathway presence and completion. One way to identify alternative metabolisms is to search for specific classes of proteins. These classes, which include transporters, carbohydrate-active enzymes (CAZymes), and peptidases, can each reveal different aspects of metabolism and potential sources of nutrients and energy for a cell.

An important protein class for predicting function is transporters. Whether species can utilize substrates can often be governed by whether they encode for the corresponding transportation proteins [121, 122]. This is due to transporters being the interface with the environment [123] and due to the inability of many substrates to diffuse freely through the membrane. Therefore, by assessing the AOA presence and absence of transporters and the substrates they can hypothetically bring across, it becomes

possible to predict substrate usage and help mitigate false positives, such as compounds that remain external to cells to offer benefits such as detoxifying H_2O_2 [124].

Another group of enzymes that might be used to support alternative metabolism are those associated with carbohydrate degradation. Such enzymes, classified as carbohydrate-active enzymes (CAZymes), typically serve as the first metabolic steps for modification of broad-spectrum carbohydrates such as polysaccharides, glycolipids, and glycoproteins [125]. Thus, CAZymes are involved in carbon metabolism and their presence suggests an ability to use more complex carbon sources for energy or assimilation [125]. Each of the five classes of CAZymes play a different role in carbohydrate metabolism, but this project will focus primarily on glycoside hydrolases (GHs) because they are enzymatically involved in carbohydrate degradation. The GHs are a widespread group targeting glycosidic bonds of substrates including monosaccharides, polysaccharides, and complex sugars such as glucans, fucans, mannans, or cellulose [126]. The third class of proteins are peptidases, which catalyze peptide hydrolysis, typically by breaking peptide bonds [127]. The presence of peptidase provides cells with an ability to obtain nitrogen or amino acids from proteins/peptides that can make up a substantial proportion of organic nitrogen within the environment. Likewise, peptidases can be especially valuable for environments with limited bioavailable nitrogen [128]. Previous research has shown that AOA within deep ocean floor sediments conduct proteolysis and deamination to regenerate or obtain ammonia [129]. Assessing CAZymes and peptidases within AOA genomes can help predict whether individual species can obtain necessary substrates from more complex environmental compounds.

Another strategy for predicting or comparing AOA metabolic potential is linking as many enzymes as possible to their reaction(s) and sorting them for essential compounds/products of interest. The purpose of such a process is multifold. Enzyme names often do not indicate their metabolic role or reactions. Additionally, enzymes can have multiple functions that are not apparent from their names. There may also be functional redundancy, where different enzymes perform the same reactions, which names alone cannot reveal. Briefly, each species' protein sequences can be converted to Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology identifiers (KEGG IDs) using a combination of

four tools. Once duplicates are removed, each KEGG ID can be linked to their reaction(s) using the KEGG API. However, reactions by themselves might not easily reveal differences between groups or potential to use different substrates therefore by sorting reactions based on their specific compounds/products of interest can help overcome this difficulty. These compounds were selected based on what is known about AOA metabolism, such as using ammonia and bicarbonate as an energy source and carbon source, respectively [11, 32–34]. In addition, popular or ubiquitous electron carriers such as NADPH, NADH, and FADH₂ were selected based on many independent studies have shown that NADPH is the main electron carrier for most enzyme-driven anabolic reactions and is responsible for the biosynthesis of all major cellular components from DNA to membranes [130], whereas NADH and FADH₂ are primarily used for oxidation or catabolic metabolism [131]. Likewise, many electron transport chain-related proteins or intermediates, such as quinones and cytochromes, were also included. This is because a previous study showed that all marine and terrestrial AOA genomes tested encoded quinone reductase, the conventional complex III and IV, ferredoxin, and complex I (NADH: ubiquinone oxidoreductase) [104]. Finally, to aid and potentially reduce false positive alternative metabolisms, which can arise when α -keto acids are supplied to AOA on account of their ability to detoxify H₂O₂ [39], enzymes associated with the removal or destruction of ROS will also be assessed, in particular those with H₂O₂.

The objective of this thesis research was to test for potential metabolic diversity among AOA representatives, with a specific emphasis on group I.1b representatives and *Ca. Nitrosocosmicus* spp. more specifically. This research used complete genomes to ensure a comprehensive comparison of putative functions. The comparative genomic pipeline used to assess potential AOA metabolisms is outlined in Chapter 2 and a detailed analysis of the results, testing for clade-specific distributions of potential substrate usage and a discussion of how the results relate to previous AOA research is found in Chapter 3. Finally, Chapter 4 contextualizes this work in a broader ecological context, emphasizing how metabolic diversity among AOA may have implications for the environment and industry.

Chapter 2

Materials and Methods

2.1 Genome selection

Representative complete genomes of AOA were downloaded from the National Center for Biotechnology Information (NCBI) between May and June 2023. Genomes were identified by searching the NCBI phylogenetic tree of *Nitrososphaerota* and by a literature search with a recent five-year period to retrieve recent genomes that had not yet been given a formal name or NCBI repository page. At the end of this survey there were 91 genomes collected to run CheckM [132] upon (Supplementary Table 1).

2.2 Genome quality

Each genome fasta file was analyzed by CheckM [132], a tool designed to assess the quality of genomes and their assemblies by using sets of ubiquitous genes that are present as single copies within a predefined lineage. The two measures used to gauge quality were completeness, the proportion of genes that are expected to be found, and contamination, which checks for single-copy genes that occur more than once in a genome sequence [132]. To be considered for this study, all genomes were required to be $\geq 90\%$ complete and $\leq 5\%$ contaminated [132–134]. Removing low-quality genomes helps minimize false negatives related to missing functional annotations [132] and resulted in 58 genomes.

2.3 Annotations

The genome fasta files for all species that passed CheckM were annotated first using Rapid Annotation using Subsystem Technology (RASTtk) [135] to detect open reading frames (ORFs), which uses Prodigal [136] followed by Glimmer3 [137]. The addition of Prodigal in 2015 to RAST [135] has the benefit of increasing accuracy with predicating short ORFs and their start positions, being more robust to differences in G+C content between genes and sequences, and being well suited to handle MAGs

compared to other ORF detection software [138]. The resulting Genbank file for each species was then used as input for Distilled and Refined Annotation of Metabolism (DRAM) [139] located on Kbase [140] under default setting. By using DRAM on Kbase servers, it was possible to overcome the high storage requirements and dependencies noted for DRAM usage [140]. DRAM independently annotates the ORFs predicted by RAST using a combination of KEGG [141], UniRef90 [142], MEROPS [143], and MMseqs2 [144], with the database having the highest bit score being used for the annotation of each putative protein. The decision to first detect ORFs using RAST and sequentially annotate them using DRAM was threefold. First, RAST uses a combination of Prodigal [136], followed by Glimmer3 [137], whereas DRAM only uses Prodigal [139]. Second, RAST relies only on the FIGfam databases, which are collected and maintained within the SEED project [133] for annotating ORF. In contrast, DRAM has a higher number of databases for ORF annotation as mentioned above. Third, there is value in using multiple annotation tools. For example, a previous study used a combination of both RAST and DRAM in the annotation of *Escherichia coli* strain K-12 and *Paceibacter normanii* AAA255-P19 and demonstrated that there were 1.5× and 1.4× more reactions identified than when using DRAM alone and 1.6× and 2.3× more functions annotated than when using RAST alone for *E. coli* and *P. normanii*, respectively [140].

2.4 Phylogenetic tree

Amino acid sequences for all 58 AOA genomes were used to generate a phylogenetic tree using GToTree (v1.6.31) and the prepackaged 76 single-copy genes for archaea [48] were utilized. GToTree had the advantage of automatically and rapidly running all steps required for phylogenetic analysis. In brief, GToTree first searched each species' protein sequences for the 76 target genes with HMMER3 [145] using predefined and default model cutoffs. If a protein target was detected more than once per genome, the target protein for that species was considered absent and was not used for alignment. Once each genome had its target proteins identified, each amino acid sequence was individually aligned with muscle v5.1 [146] and automatically trimmed with trimal v1.3.rev15 [147]. All alignments were then concatenated, and the resulting Newick file was visualized and customized with the addition of bootstrap

values, thickness of lines, and rotation of the nodules to cluster groups together using FigTree

(<http://tree.bio.ed.ac.uk/software/figtree/>)."

2.5 Functional annotation

In order to classify putative proteins into specific metabolisms and pathways, as well as to assign putative reactions, it was necessary to assign proteins annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology identifiers (KEGG IDs), which is one of the largest databases DRAM pulls from to provide KEGG IDs for its output files [140]. To maximize the number of proteins given KEGG IDs, each genome's protein fasta files were also annotated against three additional assignment/annotation tools (see section 2.5.1 below). The addition of each tool increased the total number of KEGG IDs per genome by roughly 20 to 30 new KEGG IDs (Supplementary Table 2). Duplicates were removed and all KEGG IDs were maintained even if the same protein was given different KEGG IDs from different annotation software. The rationale for keeping these IDs were as follows:

- All annotation software have stringent thresholds (i.e., above a specific bit-score or below a certain e-value). Therefore, any proteins given a KEGG ID can be considered accurate for a specific tool.
- The same protein can have multiple functions, domains, and reactions. Thus, different KEGG IDs can be detected by each database.
- Different KEGG tools leverage unique assignment algorithms and databases. Thus, multiple databases can enhance the likelihood of capturing a broader spectrum/range of KEGG ID annotations.
- Although KEGG databases are regularly updated, annotation tools may incorporate updates at different times. Using multiple tools can help reduce the risk of missing new annotations or features introduced in updated database versions.

- This study focused on identifying functions for archaea that are less well known compared to bacteria. Using multiple tools can increase the chances of annotating hypothetical proteins and their predicted functions.

2.5.1 KEGG annotations

The three KEGG ID assignment tools used for this thesis research were the KEGG Automatic Annotation Server (KAAS) [148], GhostKOALA [149], and KofamKOALA [150]. In brief, KAAS uses sequence similarities between reference proteins with assigned KEGG IDs and genome query sequences to determine if the degree of overlap (i.e., similarity between sequences) is sufficient to assign the query sequence with the corresponding KEGG ID. If multiple reference proteins hit a single query sequence, KAAS program automatically selects the one with the highest assignment score. Thus, each of the 58 genomes was independently run on the KAAS server using GHOSTZ, which is a homology search tool that is roughly 200 times more efficient in finding and detecting remote homologues than BLAST [151] under default settings and bi-directional best hit (BBH) [148]. The prokaryotic dataset that serves as the reference list for KAAS was used and amended with representative AOA genomes in addition to representatives from other archaeal and bacterial species (i.e., nmr, csy, nga, kcr, neq, ffo, asc, tpe, vdi, tuz, aho, mse, sai, sso, sto, pfm, hbu, thg, iag, tag, dka, iho, smr, ape, abi, pto, tac, nou, hru, nge, nat, hxa, nmg, htu, hah, mpd, mpl, mbn, mem, and mla).

For GhostKOALA, query sequences were assigned KEGG IDs based on sequence similarity. However, reference sequences and their corresponding KEGG IDs are pulled from a non-redundant GENES database and not individual species like that of KAAS. These sequence comparisons were achieved using the GHOSTX [151] program. If multiple hits were made against the query sequence, GhostKOALA automatically selects the top-scoring hit [141]. GhostKOALA was selected instead of BlastKOALA because recent updates resulted in the GhostKOALA GENES database having an order of magnitude more reference sequences and KEGG IDs than that of BlastKOALA [149].

The KofamKOALA tool assigns KEGG IDs based on homology search of each genome's amino acid sequences against a database of hidden Markov models with precomputed adaptive score thresholds [150]. Thus, if the protein sequence is equal to or greater than the score threshold, it was assigned a KEGG ID [150].

All Python and bash scripts used in the sequent sections are saved to the Neufeld Lab's GitHub (<https://github.com/neufeld/AOA-genomics>).

2.5.2 Categorization of KEGG IDs

In order to assign each genome's KEGG IDs to specific metabolic categories, reference KEGG ID linked to metabolism were obtained from the KEGG Brite Hierarchy (https://www.genome.jp/kegg-bin/get_htext), including those relating to carbohydrate metabolism, amino acid metabolism (the module associated with the metabolism of the 20 common amino acids), metabolism of other amino acids (the metabolism of less common or modified amino acids), energy metabolism, oxidation-reduction, sulfur, transporters, and peptidases (Supplementary Table 3) were selected as they include the KEGG IDs for central metabolism other modules from the KEGG Brite Hierarchy are associated with immune system, disease or structural proteins, which are outside the scope of this project. Brite hierarchies were collected separately and used as reference lists to check for presence and absence against the 58 AOA genomes. More specifically, the KEGG IDs for each species' genome were searched against metabolic reference KEGG ID lists '2_presence_absence/presence_absence.py'. By doing this, it was possible to obtain binary matrices (presence 1, absence 0) for specific metabolisms. To condense the number of assignments, many classes of metabolisms that shared a dominant element were joined together and duplicates were removed. For example, amino acid metabolism and metabolism of other amino acids classes were combined due to their common element of nitrogen. This resulted in a binary (presence/absence) matrices for specific categories of metabolism and included carbon, nitrogen,

peptidase/proteinases (Supplementary Table 3), oxidation-reduction, sulfur, and transporters for all 58 species of AOA.

2.5.3 Reactions

2.5.3.1 Functional redundancy removal

Using the KEGG REST API (<https://www.kegg.jp/kegg/rest/keggapi.html>), a high portion of the KEGG IDs could be connected or linked to their corresponding reactions using '3_Compare/compare_two_lists.py'. The KEGG IDs that were not connected to reactions were those involved with cellular regulation, transport, or signalling. Linking KEGG IDs and reactions resulted in a binary matrix with species names as row headers and KEGG IDs and their respective reactions in columns. However, due to functional redundancy, different enzymes (KEGG IDs) can conduct identical reactions, using the same substrates to yield the same products but with different names and KEGG IDs. To account for this functional redundancy, all instances of enzyme names were removed from the identical reactions were then combined into a single row again using '4_Duplicate remove/duplicates_combine.py'. The resulting .xlsx file had single instances of each reaction in one column, and all the KEGG IDs predicted to drive it in another, with species of AOA being row headers and 1 symbolizing the species could conduct the reaction and 0 they could not.

2.5.3.1 Reaction sorting

To identify reactions related to energy (Supplementary Table 5) and biosynthesis (Supplementary Table 6), '5_Mol_interest/molecules_interest.py' was used to search for specific compounds associated with metabolism. This operates on the idea that any potential reaction should produce or use specific or critical compounds for downstream metabolism. Therefore, reactions were identified as energy reactions if they produced reduced energy electron carriers, FADH₂ and NADH, reduced components of the ETC such as quinones and ferricytochrome, and/or released ammonia or hydrogen. Ammonia was considered

due to the ability of AOA to oxidize ammonia, whereas hydrogen was considered due to it being produced or oxidized in many energy-driven reactions.

Biosynthesis reactions were identified as those that produced/used reduced electron biosynthesis carriers such as NADPH, and/or released CO₂ or ammonia. CO₂ was included because all known AOA can grow autotrophically, whereas ammonia was included because it can be assimilated to produce amino acids and nucleotides. Finally, all reactions producing or using ATP, CTP, GTP, and TTP were considered as biosynthetic pathways because they are often associated with phosphorylation, which is a critical step in many biosynthetic reactions, contrary to being considered as energy reactions on account that most are produced via ATP synthase.

The final group of potential substrates includes those generating anaplerosis and cataplerosis metabolites, particularly those associated with the TCA cycle and the AOA-specific 3-hydroxypropionate/4-hydroxybutyrate cycle (Supplementary Table 7). These intermediates play an important role in being the initial substrates used to produce new macromolecules (cataplerosis) as well as being resupplied to the cycles (anaplerosis), to replace those that were used for cataplerosis or biosynthesis. Thus, all intermediates for these two cycles were collected and searched against all genome reactions and any reaction that used or produced these intermediates was retained.

2.5.4 Transporters

Because named transporters often do not specify potential substrates, and the promiscuity of many transporters, transporter KEGG IDs were linked to potential substrates based on previously published literature. In order to do this, the transporter matrix which consists of a binary file containing all the predicted transporters of AOA with the species as columns names and transporter's KEGG ID as row names was converted to the Transporter Classification Database (TCDB) ID [152] using '3_Compare/compare_two_lists.py'. Once converted, the TCDB IDs were linked to all their potential substrates using files found within the TCDB download files section

(<https://www.tcdb.org/download.php>). If the TCDB lacked a specific ID, a literature search was conducted to find potential substrates. This process resulted in a binary (presence and absence) matrix file with the TCDB ID, and KEGG IDs in one column each and another column with the potential substrate transported with species names as row names.

Duplicate substrates (i.e., the same substrate transported by different transporters) were consolidated into one row using '6_search_combine_TC/search_combine_TC.py', leading to a more condensed binary matrix with one instance of each substrate and all the KEGG IDs potentially responsible for transport (Supplementary Table 8) in separate columns. This led to quickly assessing what substrates can be carried across AOA membranes while accounting for transport redundancy, the lack of information from transporter names, and avoiding manually searching the literature. Nevertheless, relying on homology to predict substrates can introduce false positives and false negatives, especially when comparing transporters and their potential substrates across domains and kingdoms so while this pipeline is used caution is suggested. However, using the pipeline developed for this project makes it possible to assess the metabolic potential of AOA to a greater extent than if names were used as a sole approach for functional inference.

2.5.5 CAZymes

Carbohydrate-active enzymes (CAZymes) include both enzymes and carbohydrate-binding modules/molecules involved in complex carbohydrate decomposition and synthesis [153]. Potential CAZymes were identified by uploading amino acid fasta files for each genome into the automated Carbohydrate-active enzyme ANnotation (dbCAN3) server [154], using the dbCAN3 default and recommend databases and parameters of HMMER: dbCAN (E-Value < 1e-15, coverage > 0.35), DIAMOND: CAZy (E-Value < 1e-102) and HAMMER: dbCAN-sub (E-Value < 1e-15, coverage > 0.35). CAZymes are then assigned based on homology and automatically separated into five categories: carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), auxiliary activities (AA), glycosyl transferases (GT), and polysaccharide lyases (PL). Positive hits and

identification of proteins as CAZymes were considered when two or more previously mentioned databases agreed on their assignment according to previously published recommendations [153]. All the AOA genomes CAZymes were then compiled and converted to a presence and absence binary matrix (Supplementary Table 9; '2_presence_absence/presence_absence.py').

2.5.5 Pathways

The presence and completeness of metabolic pathways were achieved by running annotated protein sequences on GapSeq (v.1.2) using the “find” function [155]. GapSeq operates via homology, where query protein sequences are searched against reference sequences for a given enzyme. If significant homology is detected (i.e., bit score ≥ 100 [155]), then the query protein is predicted to function as that reference protein [155]. More specifically, for each pathway, the names of enzyme are retrieved from the user-selected MetaCyc, KEGG, and ModelSEED databases. Once enzyme names are obtained from the three databases, GapSeq retrieved the enzymes sequences as references from the UniProt Consortium, 2016 (Uniport) taking homologs of enzymes from only the archaeal domain, using unreviewed protein sequences only when reviewed sequences were unavailable [155]. Next, GapSeq compared the homology of query sequences against those obtained from Uniport and positive hits were recorded for bit scores ≥ 100 and coverages of $\geq 75\%$ [155]. This process was done for each enzyme within all pathways searched and if the overall pathway showed $\geq 75\%$ of it enzymes present, or $\geq 66\%$ if key pathway enzymes were present, the pathway was recorded as complete [155]. Key pathway enzymes are enzymes that are often regulated and, as a result, their concentration within a cell can affect the rate of an entire metabolic pathway, these enzymes are classified as key according to the databases MetaCyc, KEGG, and ModelSEED. Using these parameters, it was possible to manually select classes of pathways/substrates for analysis in particular selecting those involved with central metabolism and included biosynthesis and degradation of amino acids, amide, carbohydrates, fermentation, polyamine, as well as prokaryotic respiration/energy pathways, such as the TCA cycle, and those pertaining to inorganic substrate metabolism (e.g., sulfur). Manually selecting the pathways was important on account of the large number

of pathways that GapSeq can pull from that have little or no immediate connections to microbial metabolism such as immune systems or virulence factors or are outside the scope of this project including signalling pathways. In addition, as of writing (mid-July 2024) GapSeq cannot run all pathways in one run and thus requires the user to manually select the pathways of interest.

The resulting GapSeq pathway files were then run through two self-made bash file sets, which served the overall purpose of condensing and consolidating all genome outputs into one matrix (Supplementary Table 10). The first served to collect the proportion (percentage) of pathway completion ('7_path/Codes_2/Together_1.sh'). The second set collected was whether the pathway was present (complete) and absent (incomplete) (1 or 0) ('7_path/FT/collect_FT.sh'), the completeness of a pathway is based on the previously mentioned parameters.

2.5.6 Exclusive and lineage specificity analysis

2.5.6.1 Exclusive analysis

Traits are defined as any pathway, reaction, or protein that was exclusively found in each or only one group or genus of AOA. All the binary matrices generated above were used for this analysis and were organized into files in which row headers were traits, the second column was species names, while the first column was a taxonomic level. By using '9_Only_LSP/Group_only.py' it was possible to pull the columns that show presence in one group or genus, but none outside, or in other words, columns that showed complete exclusivity for a specific taxonomic level.

2.5.6.2 Lineage specificity analysis (LSP)

To identify traits that showed lineage specificity, or the traits that were found or predicted within species of an in-group at a higher proportion compared to those in the out-group(s) relied on the method outlined previously [156], which calculates a trait's lineage-specificity. The two variables of lineage-specificity are precision, the degree/proportion of which the trait is conserved within the in-group, and sensitivity, which is how exclusive the trait is to the in-group.

- Precision = [number of trait-containing species in the in-group] ÷ [number of species in the in-group]
- Sensitivity = [number of trait-containing species in the in-group] ÷ [number of trait-containing species across all AOA].

Both precision and sensitivity are entered within the formula:

$$F1 = 2 [\text{precision} * \text{sensitivity}] [\text{precision} + \text{sensitivity}]$$

This resulting F1 score shows the degree to which the trait is exclusive to the in-groups or the in-group ability to predict the presence of the trait [156]. In order to calculate lineage specificity values, ‘9_Only_LSP/LSP.py’ was used on the same binary matrices used for exclusive analysis as input, but with the taxonomic level/column changed to calculate lineage specificity for different in and out groups of AOA genomes. The output of the script only kept the reactions and proteins with precision and sensitivity scores ≥ 0.8 (80%), a threshold set by previous lineage specificity studies [157] setting the threshold at a high value minimized the chances of false positives. However, such a high threshold can potentially risk false negatives, which are traits that should show lineage specificity but will not be classified as such. To account for this, all traits below these thresholds were manually assessed for potential AOA relevance, including nitrogenous substrates such as amino acids, carbon substrates such as TCA intermediates, and sugars.

To test lineage-specificity with different AOA groups, it was necessary to run these scripts multiple times, each time setting the in-groups to clusters or nodes of species identified as potential groups or subgroups, either from previously published work, phylogeny, and the metabolic PCoA plot. Examples include phylogenetic groups (e.g., I.1b, I.1a, I.1a-associated, and ThAOA), the two sub-clades of I.1b, one subclade *Ca. Nitrosocosmicus*, *Ca. Nitrosodeserticola*, *Ca. Nitrosopolaris*, and *Nitrososphaera* sp. AFS and the second subclade with the five other species of *Nitrososphaera* (Figure 2.1). These two subclades of I.1b were identified based on phylogenetic (Figure 1.1) and metabolic PCoA similarities between the species. The reasoning to include *Nitrososphaera* sp. AFS within the *Ca. Nitrosocosmicus* clade was due to it being significantly closer to *Ca. Nitrosocosmicus* in terms of

phylogeny and metabolic potential, then to the other *Nitrososphaera* species. The next analysis included the node after sub-clades but before individual genera and included *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola* species as one group, *Ca. Nitrosopolaris wilkensis*, and *Nitrososphaera* sp. AFS as a second, and the other five species of *Nitrososphaera* as a third (Figure 2.1). The final analysis was done on all the individual genera of AOA separately (Figure 2.1). If two or more in-groups or nodes, in which the larger in-group contained all the species from a smaller in-group, contained precision and sensitivity greater than ≥ 0.8 , lineage specificity was assigned to the node with the highest F1 score [156]. The same narrowing of clades was not done for the other groups of AOA because below the node of group there were very few or a complete absence of traits that showed lineage specify.

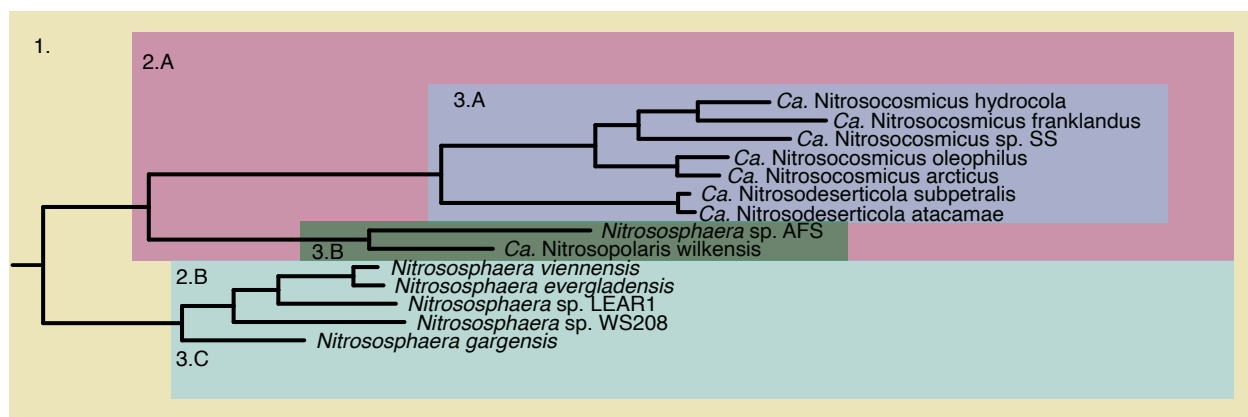


Figure 2.1 A subset of the phylogenetic tree from Figure 1.1, only showing the species within group I.1b. The different coloured rectangles and letters represent the different in-groups that were run for lineage specificity analysis (using LSP.py). More specifically, to determine if different clades (groups of species) showed differences in their abundance or variety of lineage-specific traits, it was necessary to alter these species' taxonomic levels within the binary matrix or to assign them to separate in-groups from one another. Species within different coloured rectangles and letters represent different in-groups, whereas the number is the times LSP.py was run. The clades were determined by the phylogenetic tree presented here and the PCoA metabolic plot.

2.6 Visualizations

2.6.1 Principal coordinate analysis (PCoA)

To generate principal coordinate analysis (PCoA) ordinations, presence/absence matrices were consolidated from the different classes of metabolism including transporters, peptidases, CAZymes,

carbon and nitrogen metabolism, and oxidative and reductive proteins with duplicates removed (Supplementary Table 4). This was done to compare the species based on their metabolism and not based on “housekeeping” proteins, such as those involved in DNA replication, transcription, or translation. The binary annotation file (Supplementary Table 4) was run using `vegdist()` from the `vegan` package [158], with Jaccard as the distance metric. The resulting distance matrix was analyzed using `cmdscale()` from the `stats v. 3.6.2` package [159], with the maximum dimension set to two and with the addition of a minimal additive constant (`add=TRUE`). The obtained eigenvalues were then used for preparing a PCoA using `ggplot2` [160] and `ggrepel` was used to show the species names in greater clarity (<https://ggrepel.slowkow.com/>).

2.6.2 Phylogenetic trees figure, heatmaps, and Venn diagrams

The phylogenetic tree was generated using `GToTree (v1.6.31)` [48]. Proteins predicted by KEGG to be associated with metabolism were standardized with `standardize=TRUE`. Heat maps, box plots, and Venn diagrams were prepared using `R studio (RStudio Team (2023). 2023.06.1)` and the packages `ComplexHeatmap` [161], `ggplot2` [160], and `ggvenn` [162] packages, respectively.

All Python and bash scripts are saved to the Neufeld Lab’s GitHub (<https://github.com/neufeld/AOA-genomics>). In addition, all Supplementary Tables referenced within this thesis are located at https://docs.google.com/spreadsheets/d/12B9z61tmRBdU44GhA3viL5k_qs_6c7aNcG4uGOgMVTU/edit?gid=633529713#gid=633529713

Chapter 3

Results & Discussion

3.1 Overview

3.1.1 Genomic and protein sizes

Genomes within the I.1b clade showed a wide diversity of genome sizes, with most I.1b genomes being larger and possessing higher numbers of encoded proteins compared to genomes associated with the other three AOA groups (Figure 3.1). Yet there is a high degree of heterogeneity between I.1b genomes, with the smallest I.1b genomes being *Nitrososphaera* sp. WS208 and *Nitrososphaera* sp. L.E.AR.1, which were similar in size to those of other AOA groups. The largest AOA genome was associated with *Ca. Nitrosopolaris wilkensis*, with many more predicted proteins compared to those of other I.1b species. The species from the other AOA groups possessed smaller genomes and did not separate clearly by clade in terms of genome size and the number of predicted proteins (Figure 3.1). The differences in both genome sizes and protein numbers between groups of AOA parallels previous results showing that there are substantial size and protein abundance differences between AOA species/group genomes, in particular between I.1a and I.1b [104].

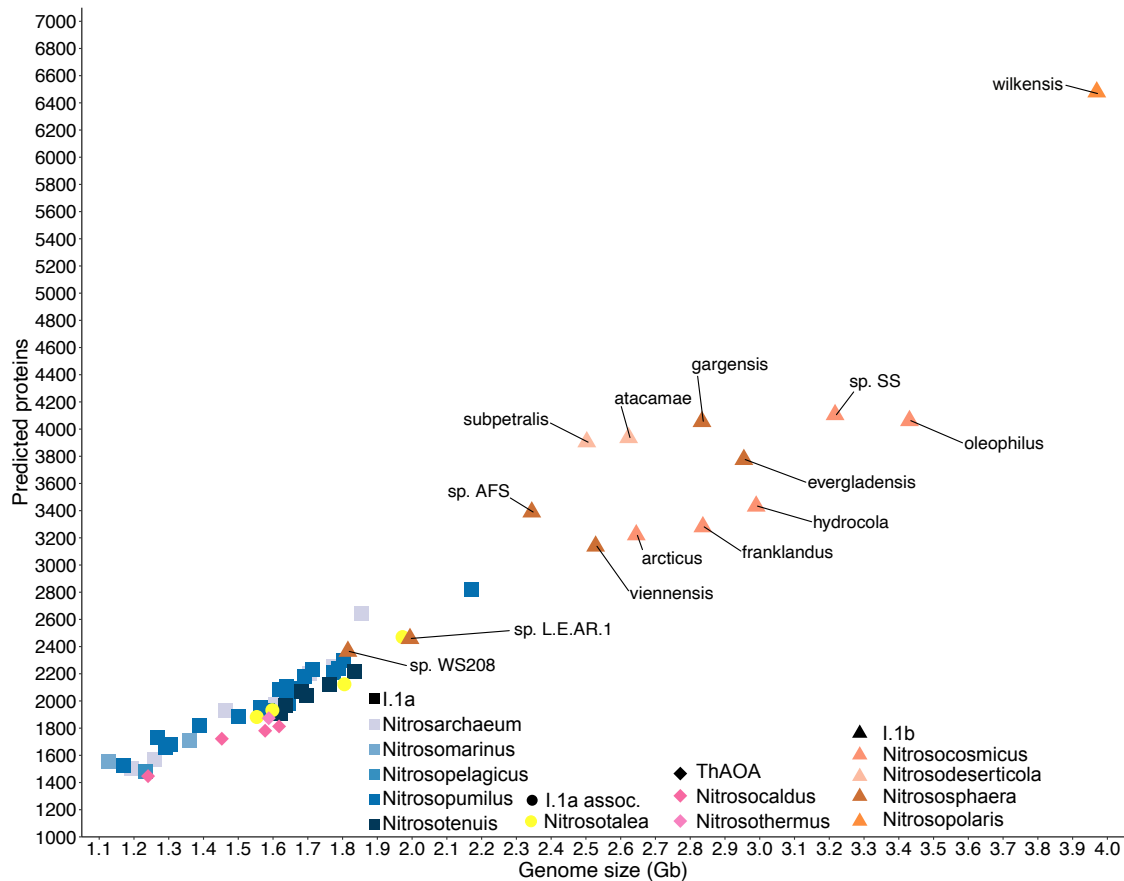


Figure 3.1 A scatter plot generated using ggplot2 [160] shows a strong positive relationship between genome size (gigabase (Gb) $\times 10^9$) versus the number of proteins for the 58 genomes of AOA. Genome sizes were obtained from the NCBI, and predicted protein counts were obtained from the output of DRAM (including both proteins with complete identifications (i.e., name) and those that were unknown (i.e., hypothetical protein)). Genomes belonging to the same group were given the same shape and base colour, with different hues representing different genera (shape, colour): I.1b (triangles, orange), ThAOA (diamonds, pink), I.1a (squares, blue), and I.1a-associated (circles, yellow).

3.1.2 Broad metabolic differences

This differentiation between the genomes of AOA genera and groups, in particular those of I.1b, was also within the metabolic PCoA plot (Figure 3.2), which shows distinctive groupings of the four groups of AOA, suggesting that metabolic potential is linked to phylogenetic relatedness. However, the degree of difference between groups and genera varies, with I.1a and I.1a-associated being much closer and hence similar to each other compared to ThAOA and I.1b, which both show sizable distances from

the former two groups and each other (Figure 3.2). This wide variety of functions and metabolic potentials was previously cited for the number of ecotypes and intra-diversity divergences of *Nitrososphaeria* [163]. Finally, within the ordination (Figure 3.2) and phylogenetic tree (Figure 1.1), *Nitrososphaeria* sp. AFS is more closely associated with the *Ca. Nitrosocosmicus* clade over the *Nitrososphaeria* clade. Perhaps this is related to annotation errors, the lower completeness and higher contamination compared to other *Nitrososphaeria* at 93.2% and 2.91% respectively, or potentially could suggest that *Nitrososphaeria* sp. AFS taxonomic name and phylogenetic relations need to be changed to reflect a phylogeny or genus within the *Ca. Nitrosocosmicus* clade.

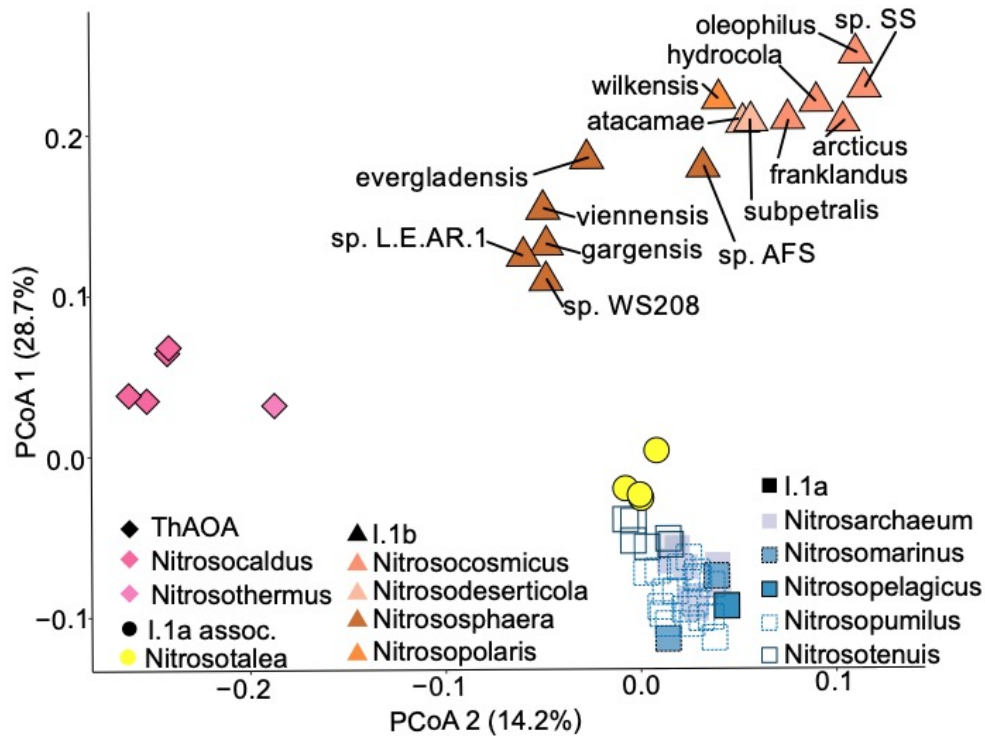


Figure 3.2 A PCoA plot showing overall metabolic repertoire differences between the 58 AOA genomes. Genomes belonging to the same group are given the same shape and base colour with different hues representing different genera (shape, colour): I.1b (triangles, red), ThAOA (diamonds, pink), I.1a (squares, blue), and I.1a-associated (circles, yellow). The metabolic repertoires considered were all the proteins associated with carbon, nitrogen, peptidase/proteinases, OxRed, non-inorganic transporters, and CAZymes, concatenated within a single presence and absence matrix with duplicates removed (Supplementary Table 4), which was subsequently run using the vegan package [158], with Jaccard similarity as the distance matrix, followed by cmdscale() from the stats v. 3.6.2 package [159], with the maximum dimension set to two and with the addition of a minimal additive constant (add=TRUE). The obtained eigenvalues were then used to generate the PCoA using ggplot2 [160].

3.1.3 Averages and exclusivity

When comparing the total number and range of proteins, reactions, and pathways, the results demonstrate that the larger genomes of the I.1b AOA encoded for a greater metabolic potential than for the other groups (Figure 3.3A-F), and thus suggests that these AOA may supplement chemolithoautotrophic growth with alternative substrates for carbon and energy than found in the other three groups (I.1a, I.1a-associated, and ThAOA). For example, species within I.1b show a higher abundance both in terms of total number and the number of exclusive proteins, or those only found within I.1b, for all metabolic protein categories included in this thesis, such as CADzy, peptidase, and transporters (Figure 3.3A-C) and reactions predicted to be involved with specific elements of metabolism, including energetic, biosynthesis, and anaplerosis (Figure 3.3D-F). The one exception is for predicted complete pathways (Figure 3.3G), in which I.1a-associated and I.1b are predicted to have similar averages. However, I.1b genomes had a higher diversity or number of exclusive pathways than I.1a-associated (Figure 3.3G), which suggest that I.1b has a greater metabolic potential/versatility than the other three groups. However, additional statistical support and/or verifying these findings with *in vitro* experiments would be required to increase confidence in the claim that I.1b representatives have a greater ability to use alternative substrates over the other groups. This higher metabolic potential predicted for I.1b genomes might be directly related to them possessing larger genomes and higher number of proteins on account that species with larger genomes did likewise show higher counts of enzymes and transporters corresponding with alternative metabolisms. The finding that larger genomes of AOA were also associated with increased metabolic potential is likewise seen across many bacterial and archaea domains [164, 165].

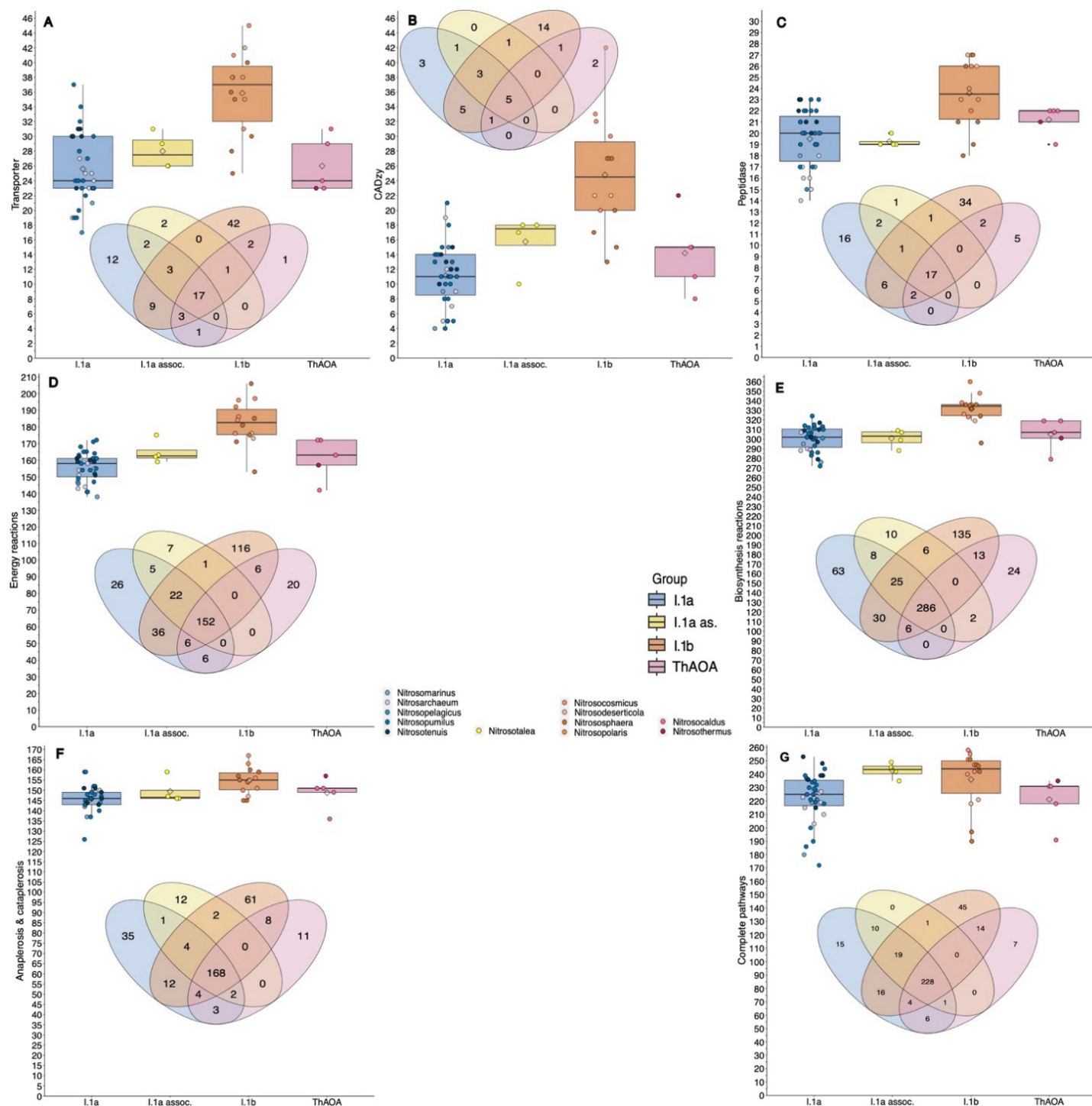


Figure 3.3 Boxplots and Venn diagrams generated using ggplot2 and ggvenn, respectively, to show the total abundance and group average (clear diamond) of genomes for each metabolic category, including: Transporters (A), CAZymes (B), peptidase (C), energy reactions (D), biosynthetic reactions (E), anaplerosis & cataplerosis reactions (F), and complete pathways (G) as predicted by GapSeq [155]. All genomes belonging to the same group were given the same base with different shades for each genus. In addition, each has a corresponding Venn diagram. Groups are colour-coded, and numbers show the abundance of proteins or reactions each group present in at least one genome.

3.1.4 Evolutionary explanations and heterogeneity

The distinctions between the groups in terms of metabolism (Figure 3.3A-G), genome size, and protein counts (Figure 3.1), particularly for I.1b and I.1a, likely arose due to their ancestors establishing in different environments, terrestrial and aquatic, respectively and evolutionary (i.e., I.1b species arose first within terrestrial environments, followed by I.1a aquatic representatives), which led the different groups to acquire differing traits and strategies [55, 104], with I.1b obtaining both a higher number of proteins and at a greater rate of change than that of members of I.1a [55, 61]. Moreover, despite I.1a acquiring genes that deal with higher osmotic stresses and lower temperatures, both of which are critical stresses of the oceans [55], some evidence indicates that at the same time I.1b genomes and total number of proteins were increasing in size, I.1a was undergoing genome streamlining [55, 104], likely as a result of many species existing within marine environments, where the oligotrophic conditions (e.g., lower inorganic nitrogen supply), has been predictive of invoking genome streamlining in free-living bacterioplankton [166]. However, within and between species of the same group and genus, in particularly for I.1b, there is a large spread of metabolic potentials, in terms of the number of proteins (Figure 3.3A-C), reactions (Figure 3.3D-F), and pathways (Figure 3.3G) with metabolism (Figure 3.2, Figure 3.3); with *Ca. Nitrosocosmicus*, followed by *Nitrososphaera*, on average, showing the highest average diversity and number. The metabolic PCoA (Figure 3.2) shows this high heterogeneity between *Nitrososphaera* and *Ca. Nitrosocosmicus*, whereas the other genera of I.1b such as *Ca. Nitrosodeserticola* and *Ca. Nitrosopolaris* are closer to *Ca. Nitrosocosmicus* than *Nitrososphaera*.

The clustering of I.1b genera into two clusters leads to the prediction that there are two sub-groups (i.e., clades) of I.1b, *Ca. Nitrosodeserticola*, *Ca. Nitrosopolaris*, *Ca. Nitrosocosmicus* and *Nitrososphaera* sp. AFS within one informally named here the *Ca. Nitrosocosmicus* clade, and the remaining five *Nitrososphaera* within another again informally called the *Nitrososphaera* clade. The high differentiation between species belonging to the *Ca. Nitrosocosmicus* and *Nitrososphaera* clades is backed up by previous work suggesting that the last common ancestor of I.1b diverged relatively early

into two clades, with each clade gaining a high number of metabolic genes yet distinctive from one another in terms of function and activity. The genes present within different species and clades likely resulted from differences in HGT [55] and gene duplications or loss leading them to encode for distinctive genomic repertoires and abilities [61, 62]. Wide intra-group variability of metabolic potentials is not limited to I.1b but can be seen to a lesser extent for all AOA groups (Figure 3.2 and Figure 3.3), nor is it limited to just metabolism as previous work has noted significant differences in I.1a motility and pH tolerances [71, 105] and I.1a-associated in regards to their optimal growth temperature, response to organic acids, and growth rates [67].

3.2 Differences in metabolic potentials

3.2.1 Reactive oxygen species (ROS) enzymes

Reactive oxygen species (ROS) found within the oxic rhizosphere can be formed from lignin polymerization within plant cell walls [167], heterotrophs actively metabolizing, and other species converting ROS through superoxide dismutase and peroxidase [168, 169]. When the spectrum of ways ROS can be produced is coupled with their highly destructive nature, it becomes crucial for aerobic species to control or eliminate them. With all AOA genomes encoding superoxide dismutase (K04564 SOD2) and peroxiredoxin (K03564) (Figure 3.4), it can be predicted that all AOA are capable of dealing with superoxide radicals, yet these reactions leads to the formation of hydrogen peroxide (H_2O_2), which is likewise produced as by-product of aerobic metabolism [170]. The presence of H_2O_2 within cells can lead to DNA damage, lipid peroxidation, and protein disintegration [171, 172]. Because H_2O_2 can inhibit ammonia oxidation at nanomolar concentrations [173], AOA require methods for its removal. Of the ROS detoxifying enzymes found within AOA, four are predicted to be capable of detoxifying H_2O_2 : non-heme chloroperoxidase (K00433), putative iron-dependent peroxidase (K07223), Mn-containing catalase (K07217), and catalase (K03781) (Figure 3.4). These four enzymes show the highest prevalence within I.1b and, with the exceptions of *Ca. Nitrosocosmicus franklandianus*, *Nitrososphaera* sp. AFS, and *Nitrososphaera* sp. LEAR1, all I.1b can potentially detoxify H_2O_2 (Figure 3.4). In terms of other groups,

possible H₂O₂ detoxifying enzymes include non-heme chloroperoxidase (K00433), but it is only predicted within *Ca. Nitrosotenuis* sp. DW1 and *Ca. Nitrosotenuis chungbukensis* (Figure 3.4), which suggests that only these two species can enzymatically remove H₂O₂.

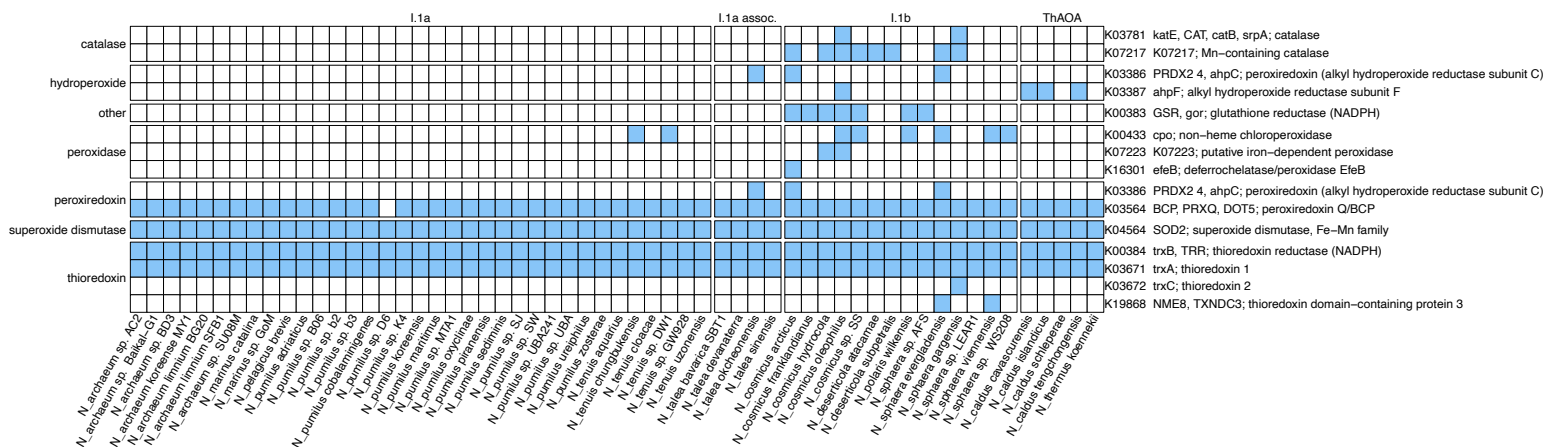


Figure 3.4 Heatmap produced using ComplexHeatmap [161] and a binary colour scheme to show the presence (blue) and absence (white) of KEGG IDs and enzymes predicted to be associated with the detoxification and removal of reactive oxygen species (ROS). The rows of heatmap are organized based on the class of ROS, while the columns are the groups (i.e., families) of AOA.

Of the enzymes that can remove and detoxify H₂O₂, catalase is among the most efficient enzymes [174], which is an enzyme prevalent only within I.1b and the *Ca. Nitrosocosmicus* clade in particular (Figure 3.4). Despite *N. viennensis* containing non-heme chloroperoxidase (K00433), growth of this AOA, was still increased when supplied with H₂O₂ scavengers (i.e., catalase, alpha-keto acids) [39]. The presence of catalase within only a few I.1b genomes, and its absence in others, suggests that it was acquired by HGT, and this was also demonstrated by previous research showing that HGT occurred between an ancestor of the *Nitrososphaerales* and *Terrabacteria*, which is a massive *bacteria* clade that includes *Actinomycetota* and *Bacillota* [55].

Caution is necessary when interpreting the ROS-encoding enzyme results. Even though the results show that *N. gargensis* contains Mn-containing catalase (Figure 3.4), previous work has suggested that it is truncated and hence not functional [39]. However, recent research has suggested that the *Ca. Nitrosocosmicus* catalase enzymes are operational for species of *Ca. Nitrosocosmicus* that showed

dominance among archaea and AOA within the root systems of pepper and ginseng, which was attributed to them expressing a manganese catalase [175]. In addition, when members of *Ca. Nitrosocosmicus* were inoculated by researchers in H₂O₂-amended bulk soils, they became highly enriched [175]. This, coupled with the ability of *Ca. Nitrosocosmicus* to grow in the absence of a H₂O₂ scavenger [32, 33], suggests that these taxa and their catalase genes can persist in the presence of H₂O₂. The rationale for discussing enzymes capable of detoxifying H₂O₂ is also important given that their absence forces species to likely rely on other tactics for its detoxification, such as using α -keto acids (pyruvate and α -ketoglutarate), which was previously suggested as being used by AOA for non-chemolithoautotrophic mixotrophic growth [105] or by depending on catalase-encoding bacteria that can be recruited to grow alongside AOA via the release of labile organic matter [76]. Together, these tactics pose more significant challenges in predicting and experimentally verifying alternative metabolic capabilities because the presence of any transporters or enzymes related to the production or transport of α -keto acids might be related to this removal of H₂O₂ and not to metabolism. Likewise, this ability to release organic matter to recruit H₂O₂-detoxifying bacteria might partially explain why I.1a species maintain transporters associated with these substrates. Therefore, AOA encoding catalase enzymes may not have to rely on external detoxification methods, either being provided with H₂O₂ scavengers or recruitment of catalase-containing microorganisms [76] and potentially removing this as a rationale for the presence of these transporters.

3.2.2 Transporters

One of the most significant indicators of whether species can use different carbon or energy sources is genome-encoded transporters [122]. Because transporters are an organism's interface with the environment, and many substrates cannot diffuse freely across their membranes [176], transporters can provide evidence for a species' ecological preferences and the substrates they use [176]. Others have suggested that the presence of transporters for organic substrates is more indicative of using alternative substrates than specific pathways, such as the TCA cycle [177]. Some of the core potential substrates, or those predicted to be transported across by all genomes of AOA within this study, include carbon

substrates such as maltose, mannose, melibiose, rhamnose, and nitrogenous substrates including methylamine, oligopeptide/peptide, spermidine, and ammonia/ammonium (Supplementary Table 8). Outside of these core substrates, the four groups of AOA possess widely different transporter abundances and diversity, even for members of the same genus, which is consistent with previous work [176]. These differences might reflect differences in habitat, differences in evolutionary histories, or differences in genome sizes, which could be reflective of HGT and gene duplications [55].

Of the four AOA groups, I.1b genomes showed the most extensive ranges of potential transporters and transported substrates. This observation is consistent with high levels of HGT and gene duplications that correspond to substrates that include carbohydrates [62]. But outside of the potential substrate of hyaluronate which is predicted to be transported by all I.1b genomes except *Nitrososphaera* sp. WS208 (Figure 3.5C), genomes of I.1b show a large heterogeneity between one another even genomes from the same genera. The potential to transport hyaluronate, a linear polymer composed of units of β -1, 4-glucuronic acid and β -1, 3-N-acetyl glucosamine [178], might be to export it to the extracellular space, as opposed to its import. The export of hyaluronate is based on hyaluronate having been detected within the extracellular space of certain bacteria, where it serves to help avoid immune system detection [179]. Although hyaluronate presence within I.1b is unlikely to be used for immune system invasion, it might help provide protection from terrestrial stressors such as desiccation.

The genomes of I.1b showed a wide diversity of potential substrates that, while exclusive to this group, were only predicted within a small number of genomes. Examples include L-aspartate and formate both being primarily found in *Nitrososphaera* and tetraethylammonium/tetramethylammonium found within limited and diverse genomes of the *Ca. Nitrosocosmicus* clade (Figure 3.5C). The genomes of *Ca. Nitrosodeserticola* showed a prevalence of genes for transporting nucleotide sugars such as GDP-D-mannose, UDP-alpha-D-glucuronic acid, and UDP-alpha-D-xylose (Figure 3.5C). The transport of these nucleotide sugars might be used to support the building of the extracellular matrix as documented in *Cryptococcus neoformans* [180].

Figure 3.5 Heatmap produced using ComplexHeatmap [161] and a binary colour scheme to show the presence (blue) and absence (white) of CAZymes (A), peptidases (B) and transporters (C) that are lineage-specificity for certain in-groups of I.1b including related clades and genera. For proteins to be considered lineage-specific, precision and specificity must be ≥ 0.8 , and the highest F1 if multiple in-groups are past this 0.8 threshold. The order of the heatmap follows a decrease in the size of the in-group, starting first with the group (i.e., all I.1b genomes) and clade *Ca. Nitrosocosmicus* and *Nitrososphaeria* (as based on phylogenetic tree and PCOA plot), node before single genera (*Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*, *Ca. N. wilkensis* and *Nitrososphaera* sp. AF2 and the remaining four *Nitrososphaera*, finally, individual genera. The only I.1b group represents those below the 0.8 threshold but shows complete exclusivity within the I.1b group.

It is unlikely that UDP-alpha-D-glucuronic acid, and UDP-alpha-D-xylose are used in AOA to evade the immune system [44], but instead might serve to protect the *Ca. Nitrosodeserticola* from the high temperature and risks of desiccation prevalent with the Atacama Desert where these two species originate. Although the prediction to transport substrates such as nitrogenous and carbon substrates were prevalent for I.1b genomes, despite being predicted within other groups and genera (Supplementary Table 8), these genes were not considered to be I.1b specific, likely due to the strict and high precision and sensitivity thresholds used for the analysis. Each of these classes of substrates will be discussed within later sections.

One substrate class within I.1b genomes that shows a wide variability both between kind and species potentially capable of transporting them are those related to the transport of structural complex (large, or high diversity of functional groups) molecules or drugs (pharmaceuticals or antibiotics) (Figure 3.5C). Although the precise substrates for these potential transporters are uncertain, previous work with bacteria, has shown that some multi-drug transporters, have a higher flexibility or range in transporting a variety of structurally different and unrelated substrates [181, 182]. For example, multi-drug transporters have been shown to transport spermidine [183], fatty acids [184], and gastrointestinal bile salts [185]. Another unknown in regards to these transporters is when or if they are actively transcribed, yet some *E.coli* multi-drug transporters are constitutively expressed and thus must serve biological roles outside of just transporting these drugs [186].

Although the presence of organic transporters may be predictive of metabolism, their presence may also be explained by export of organic molecules to recruit catalase-containing microorganisms, as demonstrated using three catalase negative *Nitrosopumilus* species, which were shown to actively export

organic molecules to promote growth of catalase-encoding *Oceanicaulis alexandrii* [76]. Another explanation relates to detoxification, where transporters are involved in detoxification processes, removing harmful organic compounds from the cell to maintain cellular health, such as interfering with the proton motive force (PMF) as can occur for many organic molecules. Some organic molecules might be transported into or out of microbial cells to produce secondary metabolites, which can serve various functions, including defence mechanisms against other microbes, predators, or environmental stressors. Alternatively, microorganisms might use organic transporters to sense the presence of specific organic molecules in their environment, which can serve as signals for altering gene expression or metabolic pathways in response to environmental changes. Overall, the TCDB and KEGG can provide only a hypothetical substrate that can be transported across the membrane. To ascertain the true range of substrates used by each transporter direct experimental validation is required [31, 176]. This low confidence related to substrate transport may also be related to homology [187, 188], which can pose problems and errors, especially when homology is low [189].

3.2.3 Cofactors

Cellular activities and functions, in particular those related to catabolism and anabolism, are driven in part by specific cofactors, such as NAD^+ , FAD^+ , coenzyme A (CoA), NADP^+ , and ATP [190, 191]. From assessing the predicted pathways from GapSeq, the genome annotation data show that all AOA can biosynthesize FAD^+ , NADP^+ , CoA, and ATP, whereas most species can synthesize NAD^+ (Supplementary Table 10). Those genomes that do not encode for the potential for NAD^+ synthesis either have lower CheckM completeness scores or this pathway is controlled by unknown homologs (Supplementary Table 10). This is contrary to other archaea species, such as *Woesearchaeota*, which lack NAD^+ synthesis potential [192]. Thus, the potential synthesis of these cofactors makes the alternative metabolic reactions that require them more likely than if they were absent.

All AOA genomes are predicted to be capable of 3PG-factor 420 biosynthesis and factor 420 biosynthesis II (mycobacteria) (Supplementary Table 10), a finding documented in previous AOA research [52, 68, 193]. Although once thought only present within methanogens, F₄₂₀ has been shown to be actively produced by AOA [75, 103]. In addition to methanogenesis, F₄₂₀ has been associated with biosynthesis pathways of tetracycline antibiotics [191], biodegradation of picrate and aflatoxins [194, 195], and increasing the metabolic flexibility of mycobacteria for the catalysis of endogenous and exogenous organic compounds [196, 197]. Thus, it may be that AOA likewise use F₄₂₀ for similar metabolic purposes. This cofactor is also protective against reactive nitrogen species, as was also observed for mycobacteria [198]. Although these research results cannot predict activation or abundances of such cofactors, it will be necessary for future research to note that the alterations or changes in the distribution and abundance of these cofactors can alter particular metabolisms in different ways [199]. For instance, intracellular NADH and NADPH can facilitate over 1,500 enzymatic reactions in some microbial species [98]. Thus, when species activate or regulate non-metabolic coupled expression, major pathways that rely on NADH and NADPH can be shut down, leading to effects on the synthesis of metabolites and overall microbial growth [98]. More specifically, when intracellular NADH concentrations were decreased within *Klebsiella pneumonia*, there was increased production of acetoin and acetate and decreased production of lactate acid and ethanol [200]. Likewise, when NADH and, to a lesser extent, NADPH were reduced within *Saccharomyces cerevisiae*, there was a significant decrease in the cell's ability to produce volatile compounds due to decreases in the production of α -keto acids and acetyl coenzyme A [201]. The usage of CoA, NAD⁺, ATP, and NADP⁺ within the 3-HP/4-HB cycle [202] does suggest that AOA can produce/synthesize these cofactors, at least during autotrophic growth on account that all known AOA both encode and utilize the 3-HP/4-HB cycle for autotrophic fixation of CO₂/HCO₃⁻ [11, 33].

3.2.3.1 Energy and electron transport chain

3.2.3.1.1 Complex 1

Nicotinamide adenine dinucleotide (NAD^+) is one of the universal coenzyme/electron carriers distributed across all domains of life and its subsequent ability to accept hydride ion [203], and then donate this ion to components of the ETC, makes it highly valuable for catabolic metabolisms [203]. Thus, NAD^+ thus serves as a predictor of whether microorganisms can utilize alternative substrates for energy [204]. In this research, the GapSeq analysis indicated that both *Ca. N. wilkensis* and *Ca. Nitrosocosmicus* species, except for *Ca. N. franklandianus*, encodes for the predicted complete respiratory complex I (Figure 3.6A) responsible for the oxidation of NADH to the reduction of a quinone, with the subsequent energy released able to pump protons across the membrane [205, 206]. This absence within *Nitrososphaera* agrees with a previous study showing that 10 MAGs resolved to the *Nitrososphaera* lacked complex 1 [207]. However, caution might be warranted for conclusions made for MAGs that may be incomplete. The high threshold for contamination and completeness in this project reinforces that *Nitrososphaera* members lack this complex.

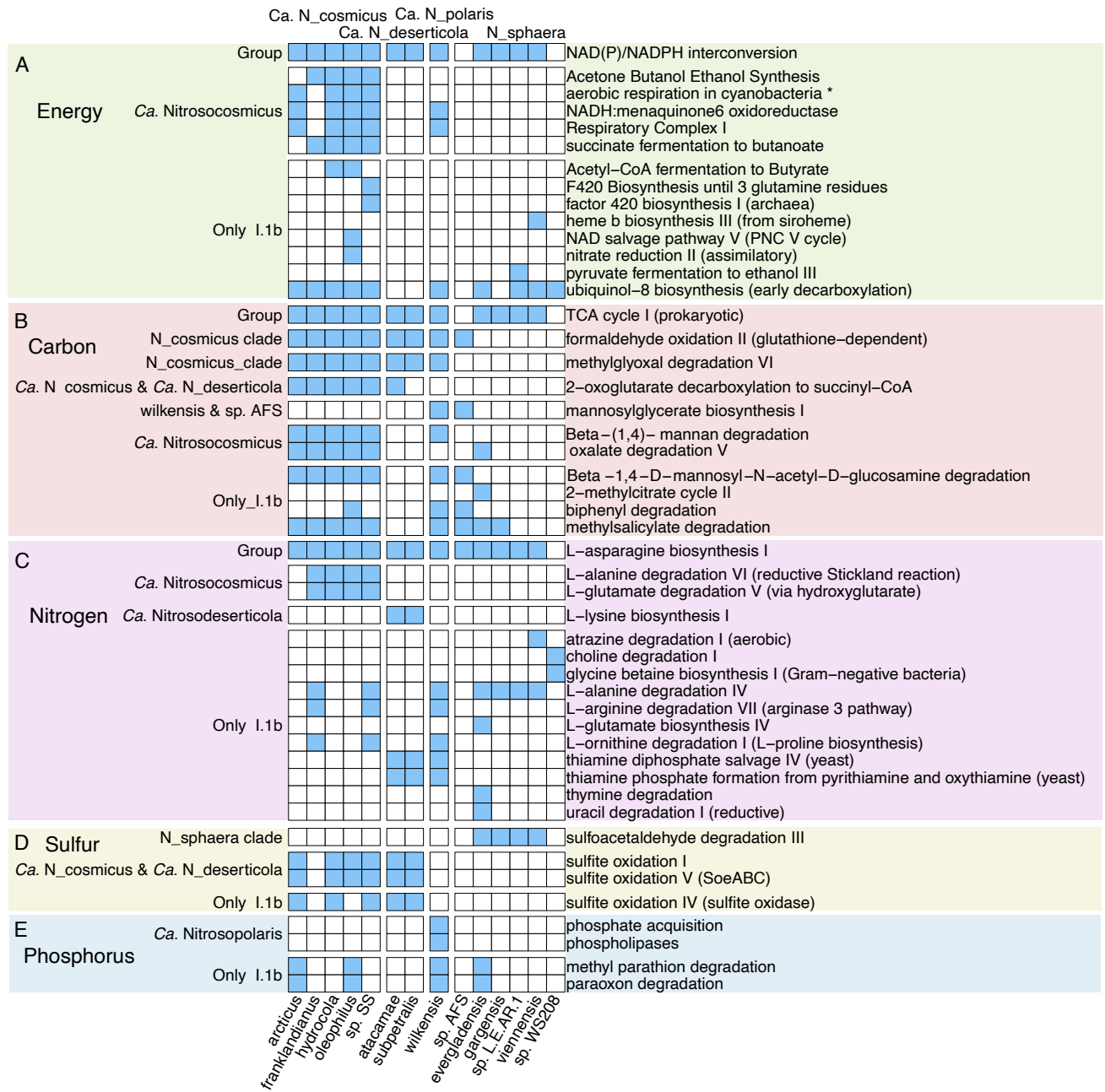


Figure 3.6 Heatmap produced using ComplexHeatmap [161] and a binary colour scheme to show the presence (blue) and absence (white) of specific and complete metabolic pathways predicted to be lineage specific for certain in-groups of I.1b including related clades and genera and organized to specific classes of metabolism (based on initial substrate or general purpose for the cell, such as produce energy) thus classes include Energy (A), Carbon (B), Nitrogen (C), Sulfur (D), and Phosphate (E). For pathways to be considered lineage-specific, precision and specificity must be ≥ 0.8 , and the highest F1 if multiple in-groups are past this 0.8 threshold. The order of the heatmap follows a decrease in the size of the in-group, starting first with the group (i.e., all I.1b genomes) and clade *Ca. Nitrosocosmicus* and *Nitrososphaeria* (as based on phylogenetic tree and PCoA plot), node before single genera (*Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*, *Ca. N. wilkensis* and *Nitrososphaeria* sp. AF2 and the remaining four *Nitrososphaeria*, finally, individual genera. The only I.1b group represents those below the 0.8 threshold but shows complete exclusivity within the I.1b group. Complete pathways are according to the default GapSeq parameters: enzymes that makeup pathways are $\geq 75\%$ are present or if all key pathway enzymes $\geq 66\%$).

Although the results from KEGG show that all AOA possess most complex I subunits, all genomes were missing genes for NADH-quinone oxidoreductase subunits E-G (Supplementary Table 5), and the absence of these subunits is in agreement with both previous AOA and archaeal research [208–210]. However, what is unclear is whether these two complexes are separate from one another or are linked to database bias is currently unknown.

If *Ca. Nitrosocosmicus* is verified to have respiratory complex I, this poses an entry point for electrons into the ETC, which means that any reaction that reduces NAD^+ could potentially be used as an energy source. Alternatively, other researchers have proposed that due to the having the *Amo* complex, respiratory complex I might operate in reverse and reduce NAD^+ , which is used in carbon fixation and biomass accumulation [57, 211]. Therefore, future experiments should test complex 1 directionality because this will have major impacts on what substrates can be used as alternative energy sources.

3.2.3.1.2 lipoquinones

The electrons from complex 1 are typically passed to lipoquinones, which are lipid molecules capable of shuttling electrons between membrane-bound complexes and typically fall into two major groups: ubiquinones and menaquinones [212]. The difference between lipoquinones typically relies on isoprenyl sidechains of varying length, degree of saturation, and presence in different amounts/diversity across the phyla. For example, Gram-negative *E. coli* mainly uses ubiquinone-8 and menaquinone-8 in oxic and anoxic conditions, respectively [213], whereas Gram-positive *Mycobacterium tuberculosis* utilizes only menaquinones-9 [214]. According to GapSeq, all AOA show the potential to biosynthesize menaquinone-4,6,8,9 (Supplementary Table 10), whereas most I.1b species show the potential to produce ubiquinol-8 as well (Figure 3.6A). Other species of archaea, such as *Acidianus ambivalens* [215, 216] and methanoperedens-like members, have likewise been shown to encode for menaquinones within their ETC [217]. In addition to ferrying electrons between complexes, specific metabolic reactions can directly reduce quinones and thus provide an entry point into the ETC [218]. For example, within *Ca.*

Nitrosocosmicus there is the Type-II NADH: quinone oxidoreductase (Figure 3.6A), which are membrane-bound complexes found in diverse phyla of bacteria, fungi, and protozoa, but not mammals [219–221]. This complex can oxidize NADH to reduce phyloquinone and menaquinone, thus posing another possible way for these electron carriers to be oxidized, allowing electrons to enter the ETC [218] and avoiding complex 1.

3.2.3.1.3 Cytochrome C

All species of AOA, except for *Ca. Nitrosotenuis* sp. GW928, contain the predicted ubiquinone menaquinone-cytochrome C reductase complex (Supplementary Table 5). This complex is responsible for oxidizing quinol to reduce cytochrome C and expels between two and four protons across the membrane. Contrary to previously published work, GapSeq predicted the presence of a pathway that can produce c-type cytochromes, and with very few genome exceptions, all species of AOA included in this project have this pathway complete (Supplementary Table 10), with bit scores greater than 100 for all the enzymes responsible for its synthesis. The species that are exceptions are roughly 75% complete for c-type cytochrome pathway, which might be more related to errors in annotations, unknown homologs, or enzymes sequences below the homology thresholds, but is also unlikely due to phylogeny because there is no phylogenetic relationship between the species that lack or show incomplete c-type cytochrome biosynthesis. The predictions of complete c-type cytochrome biosynthesis are in contrast to previous studies that predicted the absence c-type cytochromes for AOA such as *N. maritimus* [222], yet when searched for its presence within the RAST database, which uses protein sequences independent of the ones used here, *N. maritimus* appears to encode the c-type cytochrome biogenesis pathways, suggesting that the finding presented here is not an artifact or false positive. GapSeq also show that AOA are potentially capable of producing complexes that both reduce (complex III) and oxidize (complex IV) cytochrome C (Supplementary Table 10). If the presence of c-type cytochromes is indeed verified within AOA, this can indicate a way for electrons to continue down the ETC. Likewise, this would also increase

the number and diversity of potential substrates that could be used. For example, *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola* genomes show the potential to oxidize certain alcohols to reduce ferricytochrome, such as ethanol, 2-propyn-1-ol, and 2-chloroethanol (Figure 3.7).

3.2.3.1.4 Pyrroloquinoline quinone (PQQ)

One category of molecules that can assist with the reduction of cytochrome C is periplasmic or membrane-bound pyrroloquinoline quinone (PQQ)-dependent dehydrogenases, to which certain AOA species, mainly *Ca. Nitrosocosmicus*, show the presence of associated reactions [19, 33]. Their presence and high degree of expression have been shown in heterotrophic non-AOA marine Thaumarchaeota and thus suggest they are essential for energy metabolism [223]. One such example of PQQ-dependent dehydrogenases is the previously mentioned oxidation of certain alcohols to reduce ferricytochrome, which has been shown to occur within the Gram-negative bacterium *Pseudomonas aeruginosa* that can grow on ethanol as its sole carbon and energy source [224]. However, according to GapSeq, all AOA lack the predicted ability to biosynthesize pyrroloquinoline quinone (PQQ). However, the number of enzymes detected in the 58 genomes, and in other previous work [33], implies that AOA, and in particular I.1b representatives, can make use of them. Therefore, this failure to detect PQQ sequences might result from a lack of homology between GapSeq protein database, a novel pathway for its biosynthesis, or suggest that it is obtained from surrounding non-AOA community members. Finally, contrary to Qin *et al.* [104] who showed that all 44 tested species of AOA encode for plastocyanin-like proteins [104], integral to respiration and electron transfer, the analysis done here found that 10 out of 14 I.1b genomes are associated with the presence of plastocyanin (K02638). Those that are absent include both *Ca. Nitrosodeserticola* species, *Ca. N. oleophilus*, and *Nitrososphaera* sp. WS208. Outside of I.1b, only three genomes are predicted to encode plastocyanin and include *Ca. Nitrosotalea devanattera*, *Ca. Nitrosotalea okcheonensis*, and *Ca. Nitrosotenuis* sp. GW928 (Supplementary Table 4). Plastocyanins are blue copper proteins that contain copper in the active site [225] and are typically found within species capable of

oxygenic photosynthesis in that they transfer electrons between the photosystems II and I [226]. Although their function is currently unknown for AOA [222], they are predicted to have roles in shuttling electrons between complexes and intermediates in the ETC [227].

3.2.4 Energy and fermentation

An important purpose of fermentation is oxidation of NADH for regeneration of NAD^+ and production of ATP, both of which can be subsequently used for other pathways. Therefore, the initial substrates used in fermentation can be seen as electron acceptors that can circumvent the ETC [228]. This may be useful when conditions do not permit oxygenic ammonia oxidation or usage of the ETC, such as in low oxygen concentrations, or impairments of the *Amo* complex. Such impairments were shown from low ammonia concentrations, which has been shown to reduce the translation of the *Amo* complex [46, 229, 230], inhibited by organic matter [231] or limitations on the elements needed for the complexes such as copper and iron, which has been shown to down regulate genes associated with ammonia oxidation and electron transport [232]. Thereby the potential to switch to fermentation might be useful for non-motile AOA such as those within *Ca. Nitrosocosmicus*. In terms of specific pathways, all tested AOA encode the potential to ferment pyruvate to either ethanol (pyruvate fermentation to ethanol) or iso-butanol (pyruvate fermentation to iso-butanol) (Supplementary Table 10). However, only selected genomes, including those of I.1b as well as I.1a-associated and *Ca. Nitrosotenuis*, encoded for pyruvate transporters (Supplementary Table 8). Thus, the widespread presence of transporters predictive of transporting pyruvate might partly be explained by pyruvate being used to fuel fermentation. In addition, there is a wide diversity of other pyruvate-related fermentation pathways which are predicted to be complete within most AOA genomes (Supplementary Table 10). Yet the high presence of these fermentation pathways leads to the hypothesis that pyruvate is transported and used by many AOA species either sourcing it extracellularly, from the catabolism of another substrate/intermediate resourced either extracellularly or endogenously. This aligns with the concept that transporters often have lower positive annotations than

other classes of proteins, and their full range of substrates is unknown [187, 188]. In terms of producing ethanol as an end product, as is common for many fermentation pathways, previous research has shown that archaea are not known to produce ethanol as a primary fermentation end-product [233]. However, a genetic analysis of *Micrarchaeota* and *Parvarchaeota* [234] suggests that the potential exists for this process. One reason ethanol might not be detected or released within the extracellular media is that for some species, such as *Ca. Nitrosocosmicus*, it can be used as a possible alternative energy source in that its oxidation reduces two ferricytochrome c proteins (K00114) (Figure 3.7). Likewise, previous work using *Ca. N. franklandianus* showed that certain alcohols can inhibit the *Amo* complex [231], which therefore warrants alcohol's immediate breakdown.

In terms of pyruvate fermentation to iso-butanol, it has been reported within several yeast species such as *Saccharomyces cerevisiae*, *Pichia pastoris*, *Kluyveromyces lactis*, and *Magnusiomyces magnusii* [235]. The production of iso-butanol is of bio-industrial and biofuel interest due to it being used as a precursor in a wide range of polymers, paints, and plastics [236], and its advantages over ethanol including lower volatility, and higher energy density [236]. Compared to other groups, I.1b genomes show a higher proportion of pathways associated with fermentation, particularly for *Ca. Nitrosocosmicus* species, and examples include succinate fermentation to butanoate and acetone butanol ethanol synthesis. Both pathways are found in four out of five *Ca. Nitrosocosmicus* genomes but not that of *Ca. N. arcticus* (Figure 3.6A). Both *Ca. N. hydrocola* and *Ca. N. oleophilus* genomes show potentially complete pathways for acetyl-CoA fermentation to butanoate (Figure 3.6A), which is a pathway also found in several bacterial species such as *Clostridium butyricum* [237], *E. coli* [238], and *C. tyrobutyricum* [239]. However, without purpose-built experiments, the presence of these enzymes might be false positive of fermentation, as noted by Meier *et al.* [121] who put forth that most enzymes involved in fermentation are bidirectional and part of other metabolic pathways. In addition, the low yield of oxidized electron carriers and ATP and the high cost of producing such enzymes and transporters might mean that fermentation is rarely used or only under extraordinary circumstances [121] such as under low ammonia or copper concentrations, when the *Amo* complex is inhibited, or under anoxic conditions.



Figure 3.7 Heatmap produced using ComplexHeatmap (152) and a binary colour scheme to show the presence (blue) and absence (white) of reactions that can produce or reduce metabolites for energy/catabolic purposes including NADH, quinones and ammonia. The reactions presented and organized based on those that are predicted to be lineage-specificity for certain clades of I.1b. For proteins to be considered lineage-specific, precision and specificity must be ≥ 0.8 , and the highest F1 if multiple clade are past this 0.8 threshold. The order of the heatmap follows a decrease in the size of the in-group, starting first with the all I.1b species, then clade *Ca. Nitrosocosmicus* (*Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*), *Ca. N. wilkensis* and *Nitrososphaeria* sp. AF2) and *Nitrososphaeria* (*N. evergladensis*, *N. gargensis*, *Nitrososphaeria* sp. LEAR.1, *N. viennensis*, and *Nitrososphaeria* sp. WS208), as determined by phylogenetic tree and PCOA plot, individual genera. Finally, the ‘only I.1b’ group represents those below the 0.8 threshold but showed complete exclusivity within the I.1b. To save space for the figure any reactions that had excessively long reactions were assigned their KEGG based reaction codes (e.g., R04844) for these please refer to Supplementary Table 10.

3.2.5 Anabolism

Many independent studies have shown that NADPH is the primary electron carrier for most enzyme-driven anabolic reactions and aids in the biosynthesis cellular components [130]. Furthermore, upon inspection of the modified 3-HP/4-HB cycle of AOA, five reactions require the oxidation of NADPH [202]. Therefore, any reactions that involve reduction of NADP^+ by oxidizing another substrate could potentially be used as a substrate for anabolism. In addition, all I.1b genomes, with the exceptions of *Nitrososphaera* sp. WS208 and *Nitrososphaera* sp. AFS, contain NAD(P)/NADPH interconversion (Figure 3.6A), which suggests that these species can generate reduced NADH or NADPH as needed if the other exists in suitable concentrations. Group I.1b genomes, in particular those of *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*, and three out of six sampled *Nitrososphaera* genomes, also contain carbonic anhydrase (Figure 3.8), which is capable of reversible hydration of CO_2 to HCO_3^- [240, 241]. Although some species such as *Ca. Nitrosocosmicus* sp. SS can have multiple copies of carbonic anhydrase [19] at least one is predicted to reside within the cytosol because of the absence of any signal peptides [19]. Therefore, any CO_2 produced by decarboxylation reactions that is not turned into HCO_3^- due to the slightly alkaline pH of the cytosol could be converted into such by carbonic anhydrase. The importance of obtaining HCO_3^- is that it is the inorganic carbon used for the 3HP/4HB pathway [202]. Any remaining CO_2 could be subsequently used by carboxylation reactions, or lost via degassing, either actively or passively.

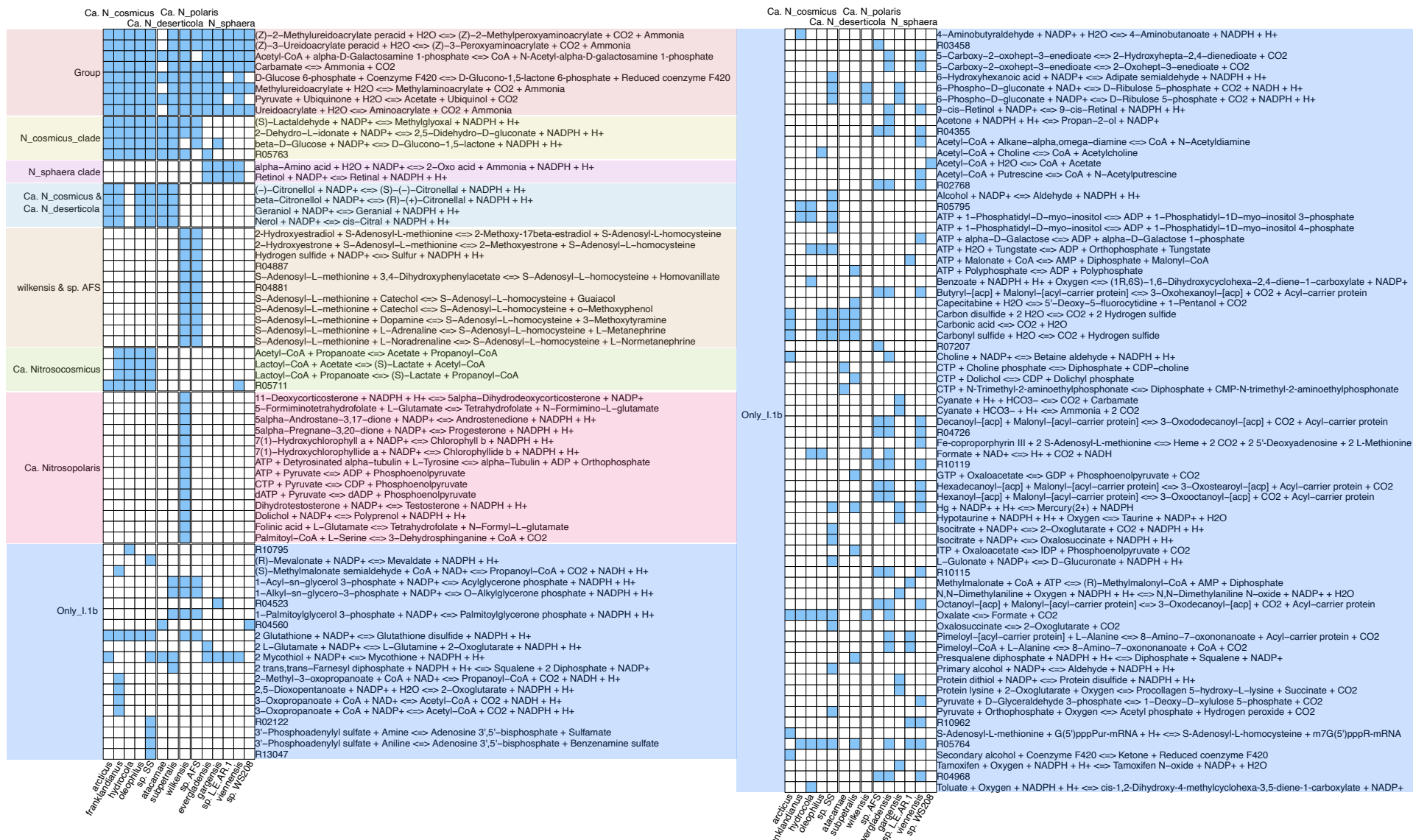


Figure 3.8 Heatmap produced using ComplexHeatmap (152) and a binary colour scheme to show the presence (blue) and absence (white) of reactions that can produce or reduce metabolites for biosynthesis metabolites that can support anabolic or autotrophic metabolism of AOA including NAPH and CO₂. The reactions presented are organized based on those that are predicted to be lineage-specificity for certain clades of I.1b. For proteins to be considered lineage-specific, precision and specificity must be ≥ 0.8 , and the highest F1 if multiple clade are past this 0.8 threshold. The order of the heatmap follows a decrease in the size of the in-group, starting first with the all I.1b species, then clade *Ca. Nitrosocosmicus* (*Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*, *Ca. N. wilkensis* and *Nitrososphaeria* sp. AF2) and *Nitrososphaeria* (*N. evergladensis*, *N. gargensis*, *Nitrososphaeria* sp. LEAR.1, *N. viennensis*, and *Nitrososphaeria* sp. WS208), as determined by phylogenetic tree and PCOA plot, individual genera. Finally, the ‘only I.1b’ group represents those below the 0.8 threshold but showed complete exclusivity within the I.1b. Those that have KEGG based reaction codes (e.g., R02122) held reactions that were excessively long and complex for these please refer to Supplementary Table 11.

3.3 Substrate differences

3.3.1 Nitrogen

3.3.1.1 Urea

Despite nearly all genomes, excluding those of several ThAOA, showing the potential to transport urea, and many encoding for urease, those within I.1b are unique in that it is predicted that all genomes included in this study encode for urease and at least one transporter (Figure 3.9). This finding agrees with previous comparative genomic analyses that show ureases primarily associated with terrestrial AOA (I.1b) rather than their marine (I.1a) counterparts [242]. The finding that the potential to use urea is not universal across AOA likewise agrees with previous research suggesting that urea usage depends on the location and ecotype [79]. Although it is uncertain whether all AOA species can use urea, multiple studies of both AOA [20, 32, 33, 71, 163] and other species [243] have shown that microorganisms with encoded ureases are also capable of degrading urea. Such use of urea may offer advantages over ammonia, for unlike ammonia that decreases as the pH drops below its pKa of 9.25 into the less preferred ammonium [42], urea is pH-independent and thus may be an especially valuable substrate for AOA to obtain sufficient ammonia under limiting conditions, such as in acidic environments [243, 244]. Urea can likewise comprise a high proportion of dissolved organic nitrogen [245] given the myriad ways it is derived from natural sources, such as by degradation of organic matter, via macro and micro animal excretions [246], atmospheric deposition, or anthropogenic sources such as fertilizers [247]. The fast rate at which urea undergoes intracellular hydrolysis, typically via urease, provides an easy source of two ammonia ions and one CO₂ and does not require cofactors such as electron carriers or ATP to be accomplished [248]. Therefore, urea degradation could provide AOA with ammonia to fuel ammonia oxidation and carbon dioxide for carbon fixation [75].

In-vitro experiments have provided direct evidence that members of the genus *Ca.* Nitrosocosmicus can use urea as a sole nitrogen and energy source [33, 34]. There is evidence there is more complexity to the story for example, *Ca.* Nitrosocosmicus sp. SS can use urea as its sole nitrogen and energy source, yet the associated ammonia oxidation rate was only a tenth of that when ammonia was

the energy source [94]. This suggests that for some species, urea can be used for metabolism but is not as metabolically favourable. In addition, a recent study showed that tested AOA species preferentially used ammonia over urea, and the authors attributed these results to higher affinities for ammonia [249].

However, to reach these conclusions, they tested only two AOA species: *N. viennensis* and

Nitrosopumilus piranensis. Although these AOA encode for urea transporters and urease (Figure 3.9), it is possible that the cells could not afford to expel the energy and resources to activate this pathway

considering the relatively low energy production from ammonia oxidation. Yet the results presented here can potentially explain the lack of activity from urea that was demonstrated within *Nitrosopumilus*

adriaticus [76], *Nitrosopumilus cobalaminigenes* [77], and *Nitrosopumilus koreensis* sp. AR1 [70] might

be because they lack a urease (Supplementary Table 5). For *Ca. N. cloacae* specifically, the results

presented here predict that it can use urea, yet those who characterized this species from a WWTP showed

that it could not on account of a truncated transporter [250]. This thus highlights a strong limitation of

relying on metabolic predictions based on genomes or subsets of relevant genome annotations.

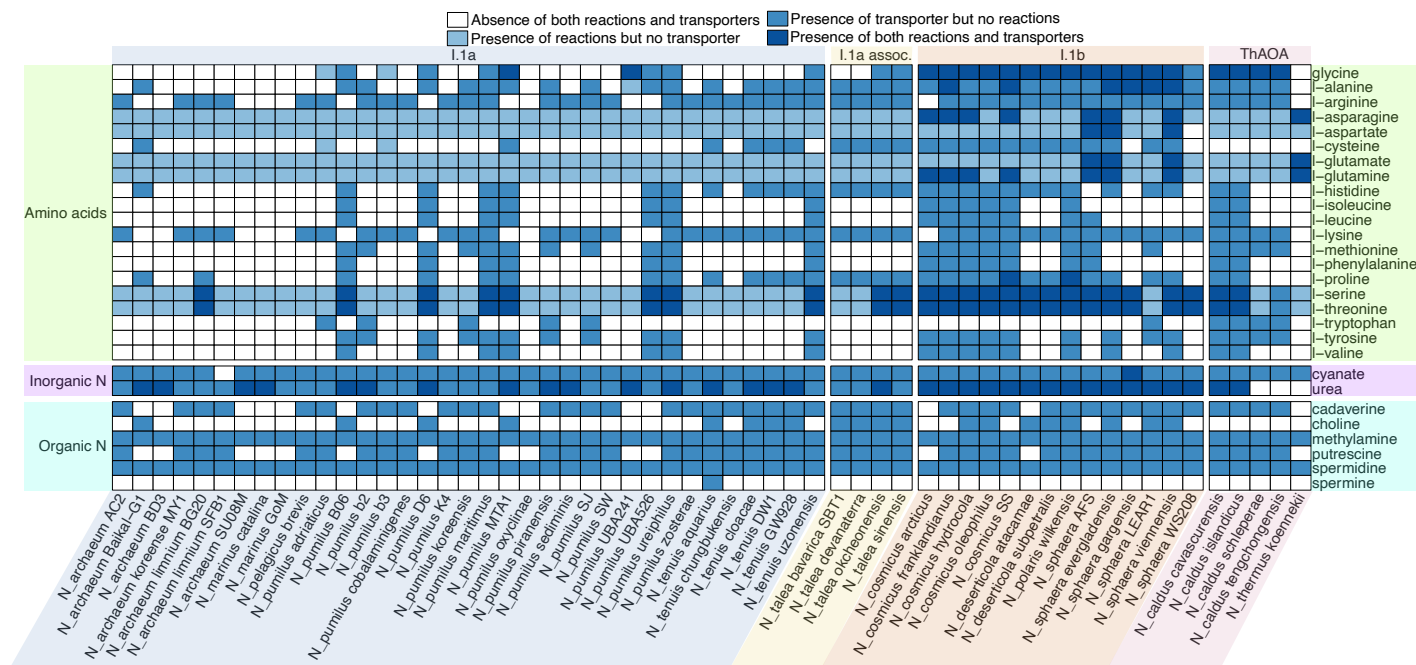


Figure 3.9 Heatmap produced using ComplexHeatmap [161] and shows which genomes of AOA encode reactions that can deamination nitrogenous substrates predicted by KEGG, but lack transporters as predicted by TCDB [152] (Presence of reactions but no transporter), encode for transporters but no reactions (Presence of transporter but no reactions), both reactions and transporters (Presence of both reactions and transporters) or neither reactions and transporters (Absence of both reactions and transporters). The nitrogenous substrates are separated based on whether they are amino acids, inorganic nitrogenous compounds (Inorganic N; not deemed entirely organic in nature), or organic nitrogenous compounds (organic N; considered organic but not amino acids).

3.3.1.2 Peptidases

Group I.1b genomes show both a higher diversity and total number of peptidases compared to other groups, and more specifically the species belonging to the *Ca. Nitrosocosmicus* clade show a higher presence over those within the *Nitrososphaera* clade (Figure 3.5B). Peptidases, also known as proteinases, encompass a comprehensive and unique class of enzymes capable of catalyzing protein or peptide hydrolysis [127], typically via the breaking of the peptide bonds. Some of the peptidases predicted within all AOA within this thesis include amidophosphoribosyltransferase, Xaa-Pro dipeptidase, asparagine synthase (glutamine-hydrolysing), and signal peptidase I (Supplementary Table 3) the last of which is important for the extracellular release of proteins from the cytoplasm [251] and might suggest that AOA is actively exporting proteins into the extracellular space. In terms of I.1b specificity, two peptidases were predicted sigma E protease and specificity stage IV sporulation protein, which is predicted to be absent within *Ca. N. hydrocola* and *Ca. N. arcticus* and present within I.1a associated, *Ca. Nitrosotalea bavarica* SBT1 (Supplementary Table 3). At other taxonomic levels, such as the *Ca. Nitrosocosmicus* clade, there is a predicted prevalence of carboxypeptidase Taq, but it was absent within *Ca. N. wilkensis* and present within *N. evergladensis* and *N. gargensis* (Figure 3.5B). In terms of genus specificity, all *Ca. Nitrosocosmicus* species showed the presence of bacterial leucyl aminopeptidase, with the exception being *Ca. N. franklandianus*, whereas both genomes of *Ca. Nitrosodeserticola* were predicted to encode ubiquitin carboxyl-terminal hydrolase 47 and epoxide hydrolase 4 (Figure 3.5B).

Peptidases are ubiquitous among microorganisms and between domains [127, 252, 253] and their function and use in the cell depends partly on their location. Peptidases that remain inside the cell are

typically associated with protein turnover or regulatory functions. In contrast, the peptidases that are exported, completely or membrane-bound, are typically involved with degrading proteins/peptides within the environment for the acquisition of carbon and nitrogen by the cell, despite the high resource and investment in resources [254, 255]. Unfortunately, this thesis research cannot currently state whether any of the 58 AOA genomes encode for the ability to export peptidases and thus, future work should rely on software such as SignalP to predict their location [256]. However, the widespread presence of peptide transporters within AOA, both shown within the results (Supplemental table 6) and with previous findings [30, 71] suggests that peptides could be uptaken and degraded by intracellular peptidase. Location notwithstanding, the cleavage of peptides, which can make up a sizable proportion of organic nitrogen within the environment [257], could provide AOA with options for obtaining amino acids. Thus, it could be especially useful for environments with limited bioavailable nitrogen [258] or energy-limiting conditions. Several other archaeal groups, such as five from marine sediments including those from Lokiarchaeota subgroup 2c, *Ca. Proteinoplasmatales*, and *Bathyarchaeota* subgroup were identified as active protein degraders based on SIP analysis [259], which showed the formation of heavy CO₂ as well as labeled RNA and lipids the latter suggests that the degraded proteins, which were likely cleaved to amino acids, entered central carbon metabolism of the catabolizing archaea.

3.3.1.3 Amino acids

On average, members of I.1b, particularly genomes belonging to the *Ca. Nitrosocosmicus* clade show a relatively high prevalence of transporters predicted to facilitate the transport of amino acids (Figure 3.9). Examples of potential transported amino acids include L-glutamine, L-leucine, L-phenylalanine, and L-asparagine (Figure 3.9). Outside of I.1b genomes, species within the genus of *Ca. Nitrosotenuis* and the ThAOA group likewise showed a high abundance of amino acid transporters (Figure 3.9), and which is consistent with previous work [260]. The presence of other species outside of AOA having shown the presence of amino acid transporters, including Group-3.unk, a sub-non-ammonia

oxidizing archaea group [261], many cyanobacteria [262, 263], Lokiarchaeon [264], and Caldiarchaeum [265] suggests that this is not a process strictly associated with AOA species.

Once taken up by the cell, the question that arises is what happens to the amino acids. One possibility is that the amino acids are reused for peptide synthesis, thus reducing costs of protein synthesis [75, 266, 267]. Such reuse of exogenous amino acids to support protein synthesis has been shown in many microorganisms surviving under higher pressures [266–268] and from isotopic analysis of AOA within marine sediment and oceanic crust communities [268–270]. Others have postulated that the amino acids are broken down, which has been demonstrated by a wide diversity of microorganisms that can use amino acids and peptides across many ecosystems [271–273]. For example, marine cyanobacteria can incorporate amino acids into new cellular macromolecules in the ocean's mesotrophic and oligotrophic regions [274]. For all AOA genomes studied within this thesis, there was a wide selection of enzymes predicted for either the deamination or transamination of amino acids, which governs the release of the amino function group or transfer to another substrate respectively (Supplementary Table 5). For example, all AOA show the presence of enzymes for the deamination of specific amino acids such as D/L-glutamine, L-glutamate, and L-asparagine, yet only I.1b species show the presence of a predicted transporter for these amino acids (Figure 3.9). This does not exclude the possibility that other groups have transporters, and which remain unclassified due to low homology. But at this time the differences between species metabolic potentials are related more to the presence or absence of a transporter than having the enzyme for the amino acid metabolism. As a result, I.1b genomes are associated with the highest dominance and exclusivity of annotation predictions for the use of exogenous amino acids, in particular for glycine, L-alanine, L-asparagine, L-aspartate L-glutamate, and L-glutamine, and the potential to use these amino acids were infrequent outside of I.1b genomes (Figure 3.9). Although predicted genes for L-serine and L-threonine use were primarily found in I.1b genomes, this potential was also found in other groups (Figure 3.9). Thus, it is possible that AOA genomes that lack transporters use their deamination or transamination enzymes more to support the general anabolic metabolism of endogenous amino acids rather than those from the environment.

Most reactions involved with the amino acid breakdown involve transferring the amino group to 2-oxoglutarate, which results in L-glutamate, which itself can be broken down or deaminated (Supplementary Table 5). If the ammonia group is freed, it could be used to satisfy nitrogen requirements in terms of ammonia oxidation and anabolism of nitrogenous macromolecules. However, there is the potential for these reactions to run in reverse, where the glutamate could transfer its amino group to the carbon skeleton of a pre-amino acid, leading to its generation of both the amino acid and 2-oxoglutarate. Hence, the directionality of these reactions might depend on the ratio of 2-oxoglutarate/L-glutamate and suggests that this ratio and L-glutamate is an important biological intermediate. The resulting carbon skeleton could be further degraded, enter a key fundamental cycle (anaplerosis), or be exported (presence of organic transporters). More specifically, the carbon skeleton could replenish intermediates of the TCA or 3-HP/4-HB cycles [75] where they could either be pulled away at a particular step for anabolism of new macromolecules (i.e., undergo cataplerosis) or oxidized to CO₂ coupled with the reduction of electron carriers, which in turn could fuel autotrophic anabolism or deliver the electron carriers to the ETC respectively. The majority of I.1b and ThAOA also show the potential to degrade glycine to yield 5,10-methylenetetrahydrofolate (Supplementary Table 6), which can be subsequently used for biosynthesis pathways, including those of purines and regeneration of methionine [275], this reaction also leads to the formation CO₂ and NH₄⁺.

Group I.1b genomes show a higher prevalence of pathways corresponding to amino acid degradation on account of having a predicted transporter. These pathways might produce intermediates or end products that can be subsequently used for central carbon cycles. For example, *Ca. Nitrosocosmicus* species genomes, except for that of *Ca. N. arcticus*, show the presence of L-glutamate degradation V (via hydroxyglutarate) (Figure 3.6C). This pathway can form intermediates that can be used to resupply both the 3-HP/4-HB cycle, with intermediates such as (S)-3-hydroxy butanol-CoA and crotonyl-CoA, and those for the TCA cycle, such as 2-oxoglutarate. Whereas the end product of acetyl-CoA could be used for the anabolism of lipids or other amino acids, fermentation, or entrance within the TCA and 3-HP/4-

HB cycle. I.1b genomes likewise show dominance for L-serine degradation I and L-alanine degradation IV both resulting in pyruvate and ammonia (Supplementary Table 7 & 8).

The breakdown of some amino acids can involve multi-pathway processes. For example, *Nitrososphaera* genomes shows higher prevalence for both L-asparagine transport and L-asparagine degradation I (Table 6 & 8), which yields ammonia and L-aspartate. The L-aspartate could then enter the L-aspartate degradation I pathway to yield L-glutamate and oxaloacetate, a pathway again predicted within *Nitrososphaera* genomes (Supplementary Table 7 & 8). The following products could be deaminated to enter the TCA or 3-HP/4-HB cycles or used for the generation of new amino acids. However, reactions that deliver the carbon skeleton of the amino acid to the TCA or 3-HP/4-HB cycles might be reversible and may also have a role in removing excessive ammonia, serving as ammonia storage [75], or keeping the concentration of ammonia low to continue the transport of ammonia into the cell.

Some evidence suggests that certain AOA use only the ammonia from the amino acids and not the carbon skeleton [259, 270]. For example, AOA within the sediments of the deep oceans have shown the ability to conduct proteolysis and deamination, mainly to obtain ammonia and export deaminated amino acids back to the external environment [129]. However, to date no studies have assessed whether terrestrial or I.1b shows preference for one atom or another.

The results presented here can also be used to contextualize previous studies. For example, both *Nitrosopumilus adriaticus* and *Nitrosopumilus piranensis* failed to be stimulated by leucine in the absence of ammonia [270], which is consistent with the results presented here for both corresponding species lack annotated transporters for leucine (or any other amino acid) (Figure 3.9). In addition, the failure of *Ca. Nitrosopelagicus brevis* [84] and *Nitrosarchaeum koreense* [85] to be stimulated by amino acids makes sense given that, other than for L-arginine (that was not tested in those studies), the genomes of these species do not contain any predicted amino acid transporters (Figure 3.9).

3.3.1.3 Polyamines

The potential to transport polyamines show differences between both species and the type of polyamine potentially brought across (Supplementary Table 8). For example, spermidine transport is predicted for nearly all AOA genomes, yet putrescine and cadaverine transporters were lacking for I.1a and *Nitrosopumilus* genomes specifically, and only the genome of *Ca. Nitrosotenuis aquarius* was predicted to encode for spermine transport (Supplementary Table 8). However, the corresponding transporters are not direct polyamine transporters but are instead predicted as possible substrates of amino acid, drugs, or antibiotics transporters. All AOA genomes included in this thesis lack polyamine transporter genes (potABCD and potFGHI), which is in agreement with the conclusion of previous research [276].

In terms of polyamine anabolism predicted by GapSeq, there is a high number of AOA genomes predicted to encode for the complete putrescine biosynthesis pathway I, but is absent in four I.1b members, all ThAOA, and two I.1a (Figure 3.10). The putrescine biosynthesis pathway I produces putrescine from the starting reagent of L-arginine, which shows widespread potential to be transported (Supplementary Table 8). In addition, all genomes except that of *Nitrososphaera* sp. AFS are predicted to encode for the complete homospermidine biosynthesis I (Figure 3.10), which combines spermidine and putrescine to yield syn-homospermidine and propane-1,3-diamine. The findings that AOA are potentially capable of synthesizing polyamines intracellularly are in agreement with previous studies that showed that *Ca. Nitrosocosmicus* sp. SS was capable of producing them as well [94].

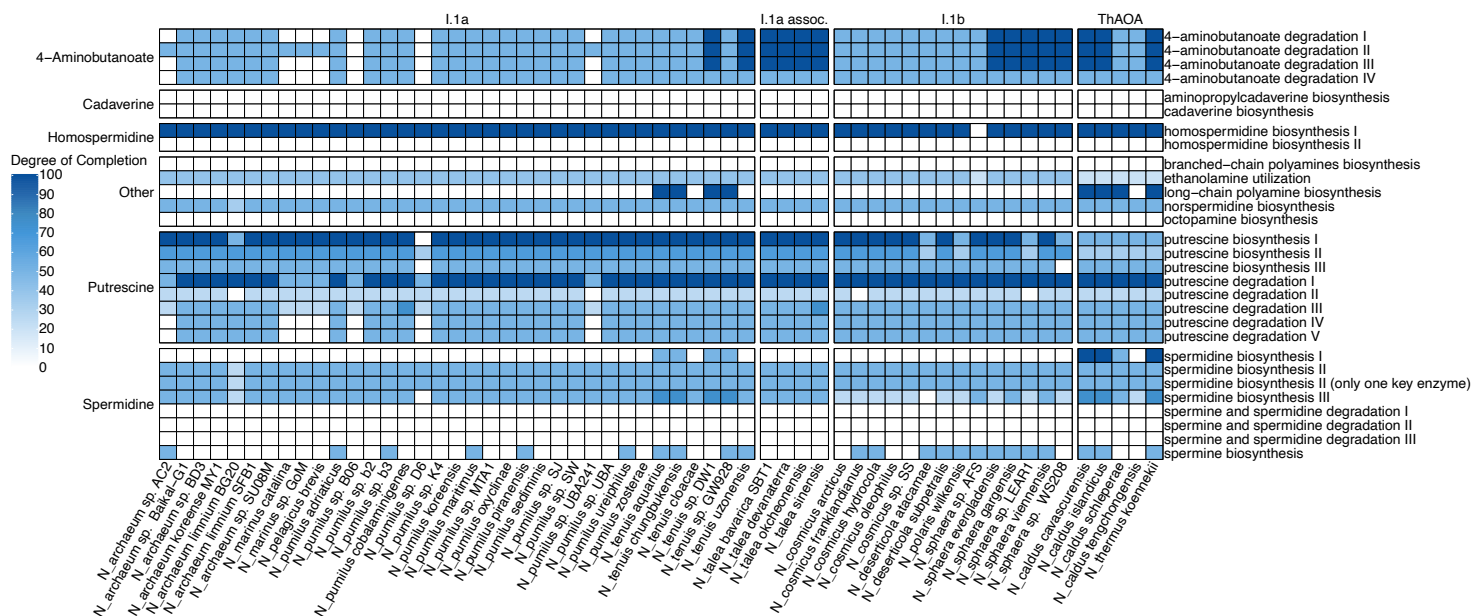


Figure 3.10 Heatmap produced using ComplexHeatmap [161], and the proportion of pathways completed for polyamine biosynthesis and degradation pathways. The pathways were obtained from KEGG, RAST, and Metacyc, and GapSeq was used to analyze them [155]. White indicates the complete absence of any enzymes associated with said pathway, while the darkest blue indicates the 100% or complete presence of all associated enzymes. The heatmap rows are organized based on the polyamine, whereas the columns are organized based on the phylogenetic groups the 58 genomes of AOA are associated with.

Although the biological role of polyamines within AOA remains unknown, one proposal is that they serve to remove excessive ammonia [94] and hence enable some species to survive in higher concentrations. However, the costs, both in terms of energy and resources (carbon atoms), of synthesizing and exporting these substrates might pose a limit or restriction. Other potential functions include the involvement of polyamines in processes such as cell division, neglecting negative charges [277], intercellular signals to trigger biofilms, assisting with the synthesis of proteins [278] or DNA [279]. In addition, some species of bacteria use polyamines as ROS scavengers, which, due to their positive charges, can effectively bind and detoxify alkyl, hydroxyl, and peroxy radicals [280].

Although I.1b, I.1a-associated, and I.1a genomes do not show any preference/group-specific pathways for the synthesis or degradation of polyamines, ThAOA genomes do (Figure 3.10). For example, the synthesis of long-chain polyamines is associated with four of five ThAOA genomes, except for *Ca. Nitrosocaldus tengchongensis* and spermidine biosynthesis III is associated with three out of five ThAOA species genomes, excluding *Ca. Nitrosocaldus schleperae* and *Ca. Nitrosocaldus tengchongensis*

(Figure 3.6). The ThAOA preference and potential to biosynthesize long-chain polyamines as well as spermidine has previously been found in other thermophilic bacteria and archaea and is thought to aid in thermostability [208, 281]. In particular, polyamines aid in stabilizing and maintaining tRNAs, rRNAs, 70S ribosomes, and DNA [208]. This protection against stresses is strengthened by polyamines binding to negatively charged phospholipid head groups or proteins in membranes, which can provide defence against osmotic or acidic stress [282]. These potential roles can also explain the presence of polyamine transporters (Supplementary Table 8), as mentioned previously.

The results presented here show that a wide selection of AOA genomes have putrescine degradation IV, which is present in all I.1a-associated, I.1b, and ThAOA, but absent in seven I.1a species (Figure 3.10). Putrescine degradation IV produces 4-aminobutanoate as one of its major products (Figure 3.10) yet the ability to break down this molecule is limited to members of I.1a-associated taxa, certain members of ThAOA, and *Nitrososphaera*. However, within this pathway, ammonia is released which could in turn undergo ammonia oxidation.

Other studies hypothesize the possibility that AOA can use polyamines as an alternative energy source without them being transported by the cell or undergoing these degradation pathways. These hypotheses are based on the idea that the active site of *Amo* faces out and, thus, the subsequent oxidation of ammonia occurs within the (pseudo)periplasmic space (PPS), the region between the cytoplasmic membrane and the S-layer [211]. The first of two potential hypotheses involves nitric oxide (NO) which AOA has been shown to produce [283], reacting with either O₂ or superoxide to form reactive nitrogen oxide species (RNOS; NO₂ or N₂O₃) [283], RNOS can then react with primary amines such as polyamines to produce nitrosamines and subsequently deaminated to NH₄⁺ which can then be oxidized by the *Amo* active site [284]. The second hypothesis involves ammonia and primary amines, such as polyamines, showing close structural and chemical similarities, which means they may be oxidized by *Amo* directly without even requiring RNOS [285]. Evidence to support these predictions includes polyamine-N oxidation rates being higher than their associated uptake rates [89, 285]. However, when *N. piranensis* sp. D3C and *N. adriaticus* sp. NF5 were supplied with putrescine as their sole energy

source, both failed to grow [285], yet this experiment should be repeated with additional species. Moreover, the lack of polyamine response might be due to the reduction in the presence of the *Amo* complex, which has been documented to occur within *Nitrosopumilus maritimus* in times of ammonia limitations [230], and the absence of the complex may result in putrescine being unable to be oxidized. Other prokaryotic species have shown the ability to use polyamines as a source of energy or its components for anabolic metabolism. For example, bacterioplankton have been shown to use polyamines as a source of nitrogen [94] and *Veillonella alcalescens* requires the presence of either cadaverine or putrescine to grow [286].

3.3.1.4 Other nitrogenous substrates

The findings of this project support that *N. gargensis* is the only known AOA species that encodes cyanase (Figure 3.7), which degrades cyanate to ammonia and CO₂ [86]. However, the major setback of relying on homology is that it is based only on what is currently established and not what is unknown. For example, the enzyme and mechanism responsible for *N. maritimus*' ability to produce ammonia from cyanate despite lacking cyanase is still unknown [287]; however, *N. maritimus* could only use cyanate when it was already growing on ammonia [87] suggestive that this breakdown is more complicated and warranting closer assessment. Nonetheless, *N. maritimus* and many other AOA, particularly those belonging to I.1b, have a transporter predicted to transport cyanate (Supplemental table 6). Thus, the potential for cyanate utilization might be more widespread than initially considered. However, future experiments must account for rapid abiotic breakdown of cyanate [86], despite that cyanate may be another preferred substrate, like urea, as its pH is stable and yields ammonia and CO₂, potentially fueling chemolithoautotrophic metabolism once broken down.

Almost all I.1b genomes encode for the potential to hydrolytically cleaved ureidoacrylate (K09020) (Figure 3.7), an acrylamide that often arises from the degradation of pyrimidine nucleotides [288, 289]. The resulting acrylamide can then hydrolyze spontaneously which leads to the formation of two ammonia, carbon dioxide and malonic semialdehyde [289]. The first two products (ammonia and

carbon dioxide) from spontaneously hydrolyze could be subsequently used to fuel chemolithoautotrophic growth; whereas, malonic semialdehyde (3-oxopropanoate) can potentially enter the 3-HP/4-HB cycle [202]. The presence of ureidoacrylate degrading enzymes has likewise been shown within *Rhizobium leguminosorum* [288] and *E. coli* sp. K-12 [289]. However, no known AOA have predicted transporters for ureidoacrylate.

Another potential substrate that shows a higher prevalence in *Nitrososphaera* and *Ca. Nitrosodeserticola* is biuret, which can be cleaved to urea and ammonia by biuret amidohydrolase (K23359) (Figure 3.7). Biuret is typically an impurity of urea fertilizers [290] and often added as a non-protein nitrogen supplement for ruminants [291]. Even though there is no known biological source for biuret [292], its addition to the environment might nonetheless provide a possible substrate for I.1b members that comprise a high proportion of AOA populations in soils, in particular those in agricultural soils [34, 293]. Given that several studies indicate that biuret might be phytotoxic [294], it may be possible to use species encoding for biuret amidohydrolase in bioremediation. The presence of biuret amidohydrolase is likewise detected in many soil species, including Actinobacteria, Alpha- and Beta-proteobacteria, fungi, archaea, and green algae [295]. In addition, *Pseudomonas* sp. BH-2 was shown to encode biuret amidohydrolase and could grow on this substrate as its nitrogen source [295]. However, the presence of biuret amidohydrolase within I.1b does not appear to be linked entirely to phylogeny (Figure 3.7), as genera not immediately related show the presence of it, such as both *Ca. Nitrosodeserticola* species and four out of six *Nitrososphaera*. This suggests that either biuret amidohydrolase was obtained multiple times via HGT or was more prevalent within basal species but was lost over time.

Both *Ca. N. arcticus* and *Ca. N. oleophilus* show a predicted transporter and reaction for the deamination of methylamine to facilitate its attachment to L-glutamate (K22081/K22083) (Figure 3.7), the presence and activity of this enzyme likewise shown within the methylotrophic beta-proteobacterium *Methyloversatilis universalis* sp. FAM5 [296]. Methylamines often form as by-products of the decomposition of proteins [297] and thus might pose an opportunity for these two *Ca. Nitrosocosmicus* species to further degrade proteins with their variety of peptidases. Moreover, methylamine levels

increase within soils when exposed to animal husbandry [298], biomass-burning [297], and certain pesticides [299]. Due to its neurotoxic and oxidative stresses, it causes to animals [300], there is increased interest to find degradation pathways. If this reaction is verified for these *Ca. Nitrosocosmicus* species, it might be possible to thus use these species (*Ca. N. arcticus* and *Ca. N. oleophilus*) in the bioremediation of methylamine.

The *N. evergladensis* genome shows the presence of both a transporter predicted to transport choline and the reaction corresponding to its oxidation to betaine aldehyde via the reduction of NAD⁺ (K00001) (Figure 3.7). A homolog of such a reaction controlled by choline dehydrogenase has likewise been found within *E. coli* [301] and *Pseudomonas aeruginosa* [302], and the resulting betaine aldehyde might subsequently be used in the synthesis of glycine or serine. Finally, *Ca. Nitrosocosmicus* genomes, *Ca. N. wilkensis*, and *Nitrososphaera*, as well as two ThAOA and several I.1a genomes show predictions in transporting hydroxyurea, thiourea, and gamma-aminobutyric acid (Supplementary Table 8) however from our analysis there are no reactions that directly use these compounds.

3.3.2 Carbon substrate

3.3.2.1 Glycolysis and TCA intermediates

Previous research has suggested that AOA can conduct glycolysis [32] yet when assessing the metabolic pathways from GapSeq, only the five ThAOA genomes and *N. gargensis* appear to encode for a complete glycolysis pathway (Supplementary Table 10). Theoretically, these species can use fructose 6-phosphate and glucose 6-phosphate as substrates, resulting in pyruvate and production of ATP and NADH.

Glycolysis ends at pyruvate, a substrate that is also associated with widespread transporters within AOA genomes, particularly those belonging to I.1b, but is likewise predicted within other groups (Supplementary Table 4). Pyruvate can also be produced by either the decarboxylation of oxaloacetate or from malate (Figure 3.8). Both processes release CO₂ and the latter releases NADH (K01003). Once within the cell, pyruvate can lead to the anabolism of specific amino acids, such as alanine [303] or

undergo fermentation which many AOA genomes show complete pathways for (Figure 3.6A).

Fermentation serves to help oxidize electron carriers and provide small amounts of ATP. Group I.1b, in particular those belonging to the *Ca. Nitrosocosmicus* clade, show the potential for oxidizing pyruvate to acetate and carbon dioxide, using ubiquinone as an electron acceptor (Figure 3.7), which is typically found associated with the ETC and thus may help with formation of the PMF [304, 305]. A homolog of this enzyme is likewise found in *E. coli*, which could support cell growth when supplied with pyruvate as sole carbon source [306]. In addition, all genomes of AOA presented here showed the potential to decarboxylate (S)-2-acetolactate to yield two pyruvates (Supplementary Table 7).

All I.1b except for *Ca. N. franklandianus* and *Nitrososphaera* sp. AFS also show the prediction of pyruvate produced by the methylglyoxal degradation VI pathway (Figure 3.6B), which results in two reduced c-type cytochromes and one pyruvate, its absence in both *Ca. N. franklandianus* and *Nitrososphaera* sp. AFS was due to the absence of the completed pathway and a predicted transporter, respectively. However, because of the toxic nature of methylglyoxal on cellular metabolism [307], there is also the potential that this pathway is intended for detoxification instead of dissimilatory metabolism. All AOA can potentially convert pyruvate into acetyl-CoA by reducing two oxidized ferredoxins (K00174) (Supplementary Table 6).

Despite the high presence of these reactions and the pyruvate transporter, caution is needed, especially for future experiments, because previous studies showing pyruvate stimulation of AOA growth [53, 105] were more related to pyruvate's ability to detoxify H_2O_2 over it being used to support alternative metabolisms [39]. Therefore, experiments to test its metabolism should potentially rely on isotopically labelled pyruvate and subsequently check for incorporation into biomolecules.

Acetate can either be obtained from the breakdown of pyruvate or acetate can be sourced extracellularly, which is predicted to be transported in I.1b, I.1a associated, and several I.1a genomes (Supplementary Table 8). The same species potentially capable of transporting acetate also shows the potential to transport propionate, phenylacetate, butyrate, and 2-hydroxybutyrate (Supplementary Table 8). Once obtained acetate can be converted into acetyl-CoA, via AMP-dependent acetyl-CoA synthetase

(K01895) or ADP-dependent acetate-CoA ligase (K24012) (Supplementary Table 7). These two enzymes are found in all AOA genomes (Supplementary Table 6) and have likewise been found in certain anammox bacteria [308]. In addition, all species hold the potential to convert citrate to acetyl-CoA (K01644) (Supplementary Table 7). However, only species encoding for the sodium-dependent dicarboxylate transporter [309] can transport citrate. Both acetate and acetyl-CoA are considered keystone substrates for microbial metabolism, in particular for AOA because it is the main product produced by the 3-HP/4-HB cycle [202]. Moreover, these compounds are used for the biosynthesis of macromolecules such as carbohydrates, amino acids [303], and nucleotides [310], as well as being important in several fermentation pathways.

The *Ca. Nitrosocosmicus* clade, again except for *Ca. N. arcticus*, shows the potential to transfer the CoA subunit between acetate and propionate (K01026) (Figure 3.11). The formation of either of these intermediates can serve as a link between the TCA cycle and the 3-HP/4-HB cycle. Acetyl-CoA can also enter the TCA cycle, which GapSeq predicts that I.1b genomes have as complete (Figure 3.6B). A complete TCA cycle is also found among members of the phylum Loki- and Heimdallarchaeia [311] and therefore is not an AOA-specific cycle among archaea. Although the genomes of the other three groups of AOA are predicted to encode for many enzymes of the TCA cycle and hence might be completed from unknown homologs. What distinguishes I.1b genomes from the other groups is the higher number and diversity of group-specific transporters (Supplementary Table 8), reactions (Figure 3.7-10), and pathways (Figure 3.6C) that can convert organic substrates, such as amino acids, into TCA cycle intermediates.

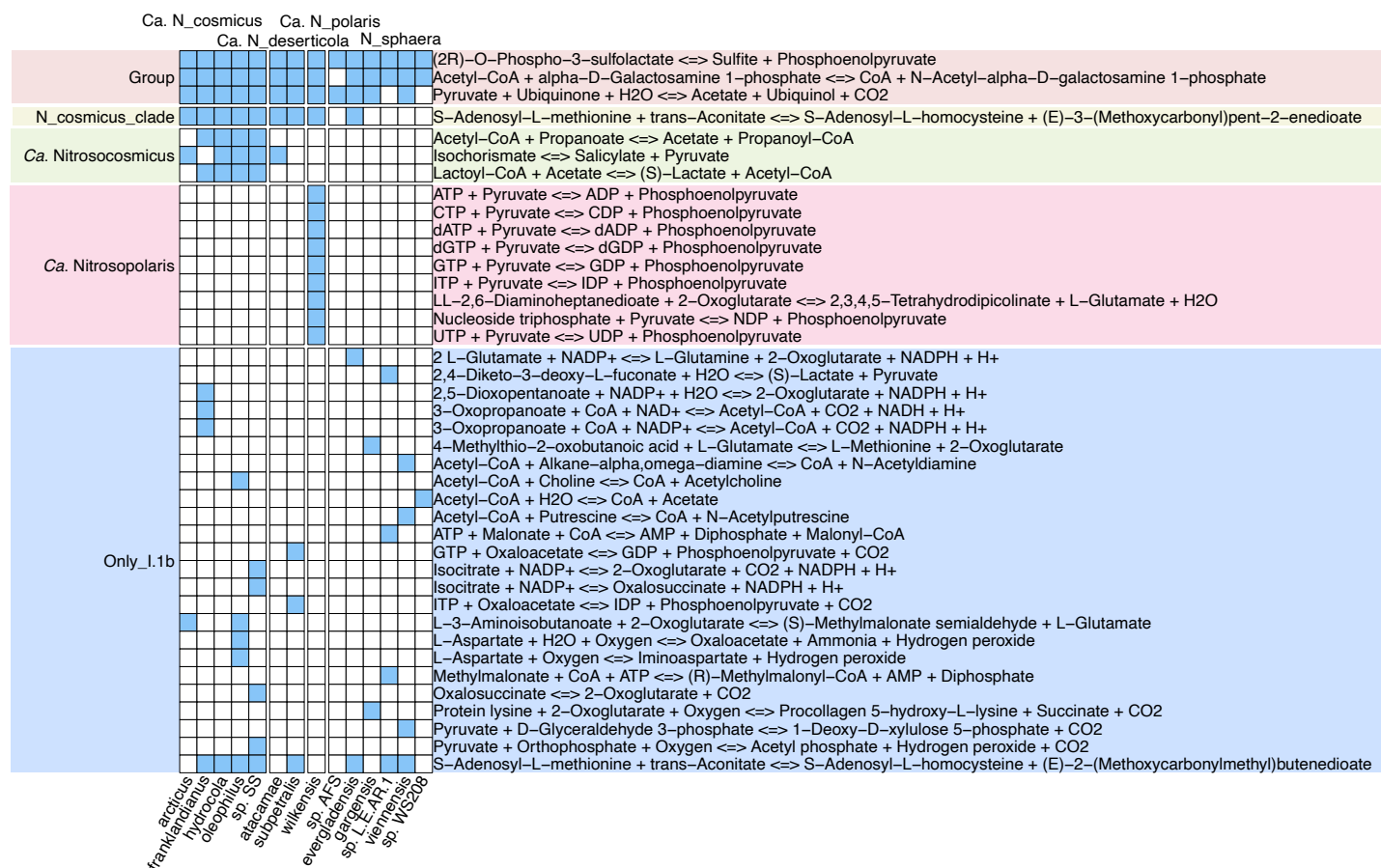


Figure 3.11 Heatmap produced using ComplexHeatmap [161] and a binary colour scheme to show the presence (blue) and absence (white) of reactions that are predicted to produce metabolites of interest including anaplerosis (A) intermediates of the TCA and 3-HP/4-HB cycle. The reactions presented are organized based on those that are predicted to be lineage-specificity for certain clades of I.1b. For proteins to be considered lineage-specific, precision and specificity must be ≥ 0.8 , and the highest F1 if multiple clade are past this 0.8 threshold. The order of the heatmap follows a decrease in the size of the in-group, starting first with all I.1b species, then clade *Ca. Nitrosocosmicus* (*Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola*, *Ca. N. wilkensis* and *Nitrososphaeria* sp. AF2) and *Nitrososphaeria* (*N. evergladensis*, *N. gargensis*, *Nitrososphaeria* sp. LEAR.1, *N. viennensis*, and *Nitrososphaeria* sp. WS208), as determined by phylogenetic tree and PCOA plot, individual genera. Finally, the ‘only I.1b’ group represents those below the 0.8 threshold but showed complete exclusivity within the I.1b. Those that have KEGG based reaction codes (e.g., R02122) held reactions that were excessively long and complex for these please refer to appendix.

There is a high degree of heterogeneity between genera of I.1b with *Ca. Nitrosocosmicus* species, except for *Ca. N. arcticus*, showing a greater potential to transport specific TCA intermediates and substrates than *Nitrososphaeria* species genomes. Some examples include succinate, oxoglutarate, malate, oxaloacetate, citrate, and fumarate (Supplementary Table 8). This potential is associated with the higher presence of sodium-dependent dicarboxylate transporter, which is well documented in the transport of

these substrates [309]. Once within the TCA cycle, four fates can be predicted for the carbon atoms that make up these intermediates. The first is that the carbon atoms are processed through the cycle to completion, back to the stage they entered. Second, the intermediate is broken down completely to CO₂ and electron carriers, at which point it could be used for autotrophic growth and delivering of electrons to the ETC, respectively. Third, the cycle forms an intermediate that can serve as a linkage between the TCA cycle and the 3-HP/4-HB cycle. Fourth, they could be subsequently extracted from the cycle for anabolic reactions, which help to synthesize new subunits or macromolecules of the cells, such as amino acids, lipids, and nucleotides [303].

Typically, four major metabolites or intermediates serve as major branching points for anabolism: phosphoenolpyruvate, pyruvate (already discussed above), oxaloacetate, and malate [312], with the proportions of each along with the availability of specific enzymes driving the biosynthesis or cataplerosis (i.e., reactions that involve removing the intermediates of a pathway or cycle, often for the synthesis of new cellular building blocks such as amino acids, lipids, or nucleotides.) of products emanating from the TCA cycle [313]. Phosphoenolpyruvate is expected to have a significant role in gluconeogenesis, a pathway [312] predicted by GapSeq to have a high presence within I.1b (Figure 3.6B). The production of phosphoenolpyruvate might occur via two reactions. The first reaction, the desulfation of (2R)-O-phospho-3-sulfolactate to generate phosphoenolpyruvate (K08097) is completely present within I.1b genomes and is largely absent from other AOA groups (Figure 3.11). Although this reaction is predicted to proceed towards the reverse reaction [314], if it is capable of operating in the forward direction, phosphoenolpyruvate will become available. The second reaction, is the phosphorylation and decarboxylation of oxaloacetate to phosphoenolpyruvate, which despite being prevalent in all genomes (Supplementary Table 6) is a more likely substrate for I.1b on account of the transportation potential of oxaloacetate (Supplementary Table 8), shown to be potentially done by the sodium-dependent dicarboxylate transporter [309]. Oxaloacetate can also lead to the synthesis of aspartate and asparagine [303] or converted into pyruvate. Moreover, if the reaction generating oxaloacetate from succinyl-CoA can operate in reverse, it could provide a route for oxaloacetate to enter the 3-HP/4-HB cycle.

Malate could be converted to pyruvate (K00027) and oxaloacetate (K00024), both resulting in the reduction of NAD^+ . But while the potential to conduct these reactions is predicted in 34 and 57 genomes respectively (Supplementary Table 5), only the species carrying the sodium-dependent dicarboxylate transporter (Supplementary Table 8) might be able to transport malate, suggestive that those with this transporter might be able to rely on exogenous malate to support their metabolism. Whereas those without a transporter might use endogenous malate produced from upstream substrates. The usage of malate has precedent and is regarded as one of the preferred carbon sources for several microbial species [293, 315].

For AOA, it is also important to consider succinate as a central metabolic intermediate given its potential use in multiple roles within AOA. For example, all AOA contain the subunits that comprise ETC complex II, which oxidizes succinate to reduce quinone (Supplementary Table 5), providing energy to the ETC [316]. The TCA cycle and 3-HB/4-HP cycle both independently produce succinyl-CoA. For the TCA cycle, succinyl-CoA can be converted to methionine, isoleucine, and valine [303], whereas within the 3-HB/4-HP cycle, succinyl-CoA leads to the formation of many precursor metabolites, such as pyruvate, phosphoenolpyruvate, oxaloacetate, 2-oxoglutarate, [202, 317]. The importance of succinyl-CoA was identified for *Metallosphaera medulla*, which also contains the 3-HP/4-HP cycle [318], and which showed that the majority of anabolic precursors for macromolecules are derived by succinyl-CoA [319]. With both TCA cycle and 3-HB/4-HP cycle using and forming succinyl-CoA, it may be that this intermediate can be shared between these cycles and used differently depending on cellular factors. Another crucial TCA intermediate is 2-oxoglutarate (i.e., α -ketoglutarate), which can lead to the synthesis of glutamine and either proline, histidine, arginine [303]. In addition to their roles in protein synthesis amino acids can also serve to store ammonia or to encourage the continued diffusion of ammonia into the cell.

The presence of the TCA cycle, or its partial presence, has also been theorized to be used for reductive or autotrophic fixation over oxidative purposes, as shown for Lokiarchaeota members [311]. Despite requiring higher oxygen concentrations, restricted under lower, the reduced TCA cycle shows a higher efficiency in fixing CO_2 , with four ATP being used per one molecule of pyruvate, compared to the

oxygen-insensitive 3-HP/4-HB pathway, which can use between five to nine ATPs per one molecule of pyruvate [317]. Thus, hypothetically, species could utilize the oxygen-sensitive TCA cycle under higher oxygen concentrations, whereas they could switch to the oxygen-insensitive 3-HP/4-HB pathway under lower oxygen concentrations. However, currently, all AOA lack the subunits that make up ATP-citric lyase and fumarate reductase key enzymes to drive the reductive TCA cycle. Unless unknown homologs encode these, it can be predicted that the TCA cycle might only be able to run in the oxidative direction.

The metabolic predictions in this thesis can potentially explain several previous findings. For example, both *N. koreense* sp. MY1 [85] and *Ca. Nitrosopelagicus brevis* [84] failed to be stimulated by substrates such as succinate, citrate, and malate, which is now understood given that these species lack a predicted transporter for these substrates (Supplementary Table 8). However, both species were stimulated by pyruvate [84, 85], despite lacking the presence of a corresponding transporter (Supplementary Table 8). This can be potentially explained given that both species lack any predicted enzymes for the detoxification of H_2O_2 (Figure 3.4) and thus pyruvate was used for H_2O_2 detoxification instead, thus giving the false positive of alternative metabolism [39]. Previous work has also showed that *Ca. N. hydrocola* was stimulated by pyruvate, malate, succinate, and citrate in the presence of ammonia [33], which can now be related to the possibility of *Ca. N. hydrocola* using these as an alternative substrates due to both bearing the sodium-dependent dicarboxylate transporter [309] (Supplementary Table 8) and encoding for enzymes and or pathways, such as a complete TCA cycle (Figure 3.6B), for the catabolism or incorporation of these exogenous substrates into new cellular metabolites. In addition, as *Ca. N. hydrocola* encodes for catalase (Figure 3.4) it is thus unlikely that these substrates were used for the detoxification of H_2O_2 . However, a replication of the experiment, in which *Ca. N. hydrocola* was supplemented with pyruvate, failed to show pyruvate being able to stimulate *Ca. N. hydrocola* but was reasoned to be because of a lower ratio of *Ca. N. hydrocola* in the culture, which led to inconsistent or possible false negative stimulation [320]. More specifically, it may have been that non-AOA heterotrophic bacteria were present at a higher abundance than *Ca. N. hydrocola* and they then consumed these substrates preferentially and thus leaving none available for *Ca. N. hydrocola*.

These results presented within this thesis cannot explain why *N. viennensis* sp. EN76 was stimulated by a range of TCA intermediates such as succinate, malate, and pyruvate in a media containing them and ammonia [83], despite lacking known transporters (Figure 3.5C), and thus it is possible that these substrates were detoxifying the H_2O_2 [39] or was transported by unknown transport homologs. Additionally, despite showing both potential transporters, enzymes, and catalase, another study demonstrated that TCA cycle intermediates pyruvate, oxaloacetate, and α -ketoglutarate failed to stimulate the growth of strain *Ca. N. oleophilus* [32]. It is possible that being within an enrichment culture led non-AOA species to utilize it first, or maybe the cellular expense to turn on these pathways was not sustained by available energy or resources.

3.3.2.2 Single carbon substrate

Despite the *Ca. N. hydrocola* genome encoding enzymes for formate breakdown (Figure 3.8), which agrees with previous work [33], *Ca. N. hydrocola* lacks a predicted transporter for formate which is predicted to be present in the genomes of *N. gargensis* and *Nitrososphaera* sp. WS208 (Supplementary Table 8), as well as several basal and non-AOA members of Thaumarchaeota [321, 322]. In addition, two nitrite-oxidizing bacterial strains of *Nitrobacter winogradskii* and *Nitrospira moscoviensis* showed the potential to use formate both bearing a transporter and the enzymes necessary for its degradation [323, 324]. While all AOA genomes showed the potential to breakdown formate to CO_2 (K21636) at the cost of one NTP (e.g., ATP, UTP, CTP, GTP), although this breakdown does not appear coupled to the reduction of NAD^+ (Supplementary Table 6) it might not be catabolically favourable due to the energetic cost.

All members of the *Ca. Nitrosocosmicus* clade likewise show the complete formaldehyde oxidation II (glutathione-dependent) pathway (Figure 3.6B), which can yield one reduced NADPH and formate per molecule of formaldehyde. In addition, three species of *Nitrososphaera* and *Ca. N. arcticus* show the potential to oxidize formaldehyde using glutathione-independent formaldehyde dehydrogenase (K00148) to formate to the reduction of NAD^+ (Figure 3.7). However, no relevant transporters for formaldehyde were predicted within the genomes of *Ca. Nitrosocosmicus* (Figure 3.5C). There is the

possibility that due to formaldehyde being highly toxic to cells in which it acts as a strong electrophile, stealing electrons from crucial macromolecules such as proteins [325, 326] and DNA [327] this pathway has evolved and been maintained within the *Ca. Nitrosocosmicus* clade on account of their non-motility. In addition, formaldehyde can lead to the formation of ROS [325], which, due to the *Ca. Nitrosocosmicus* clade bearing the catalase enzyme (Figure 3.4), poses the possibility they can deal with the consequences of formaldehyde at a higher degree than other AOA and thus its maintenance within their genomes. One possible route of formaldehyde metabolism is its spontaneous combination with tetrahydrofolate to form 5,10-methylenetetrahydrofolate (Figure 3.8), which assists with a variety of biosynthesis pathways, including that of purines and regeneration of methionine [275]. In addition, 5,10-methylenetetrahydrofolate can combine with glycine via glycine hydroxymethyltransferase (K00600), an enzyme that shows universal presence with AOA to form L-serine (Supplementary Table 7). In addition, three species of *Nitrososphaera* and *Ca. N. arcticus* show the potential to oxidize formaldehyde using glutathione-independent formaldehyde dehydrogenase (K00148) to formate to the reduction of NAD⁺ (Supplementary Table 5).

3.3.2.3 Sugars

3.3.2.3.1 Simple sugars

All tested AOA genomes encode for the potential to oxidize certain sugar alcohols such as xylitol, D-sorbitol, and iditol with the reduction of NAD⁺ a reaction conducted by the enzyme L-iditol 2-dehydrogenase (K00008) (Supplementary Table 5) and has likewise been found within species such as *Gluconobacter frateurii* and *Rhodobacter sphaeroides* [328]. However, some evidence suggests that L-iditol 2-dehydrogenase exhibits low substrate specificities, substrate promiscuous, and thus might be able to oxidize a greater number and variety of sugars than predicted here [328]. Some evidence suggests that in addition to oxidizing sugars L-iditol 2-dehydrogenase can also have a role in regulating the balance between NADH/NAD⁺ [329]. However, the absence of predicted transporter for sorbitol or iditol, and the very limited xylitol transporter (Supplementary Table 8), might suggest that this reaction may operate

more on endogenous substrate than exogenous ones. Genomes within I.1b, in particular *Ca.*

Nitrosocosmicus species, except for *Ca. N. hydrocola*, showed a higher presence for the potential transport of simple and complex sugars including xylobiose, stachyose, isomaltotriose, hexanol, alpha-lactose, aldehyde-D-xylose, D-mannitol, and isomaltose over other groups and genera (Supplementary Table 8). The finding of sugar/carbohydrate-related transporters is likewise documented for *Lokiarchaeota* [311], *Prochlorococcus*, and *Synechococcus* [330], yet there is a predicted absence within all AOA genomes for predicted reactions to use these sugars.

The prediction to transport glucose is prevalent for almost all genomes of AOA (Supplementary Table 8), yet I.1b genomes were distinct in that all I.1b species, with the exceptions of *Nitrososphaera* sp. LEAR1 and *Nitrososphaera* sp. WS208, have the potential to oxidize D-glucose to either the reduction of NAD^+ (Figure 3.7) or NADP^+ (K00034) (Figure 3.8). The majority of *Nitrososphaera* species are also predicted to encode for the oxidation of beta-D-glucose by glucose 1-dehydrogenase (Figure 3.8), which is likewise encoded and expressed by species such as *B. subtilis* [331] and *Bacillus megaterium* [332], and has been noted for its high rate of cofactor regeneration due to both its high activity and coenzyme promiscuity [333]. The, genomes of I.1b do show a higher prevalence of a complete pathway for gluconeogenesis (Supplementary Table 10), which means that they can potentially biosynthesize glucose-6P from acetyl-CoA and pyruvate, which can subsequently be used to assist with biosynthesis of various cellular components, including nucleic acids, storage molecules, or into components for the extracellular matrix [334]. This potential is likewise found in many *Ca. Angelarchaeales* and *Ca. Gimiplasmatales* MAGs [335, 336]. The resulting glucose-6P could also be used in the reduction of coenzyme F_{420} driven by coenzyme F_{420} -dependent glucose-6-phosphate dehydrogenase (K15510; F_{420} -G6PDH), an enzyme that I.1b genomes show the exclusive potential for among AOA (Figure 3.8). Yet, F_{420} -G6PDH is found in many diverse and metabolically distinct microorganisms, including *Sulfolobus solfataricus* [337] and *Halobacterium salinarum* [338] where it may play a role as a reduced electron cofactor regenerator [339, 340].

3.3.2.3.2 Polysaccharides

In terms of breaking down more complex polysaccharides, only five CAZymes were shared among all four AOA groups, with only GT2 being found in all genomes (Supplementary Table 9). As mentioned, genomes within I.1b had a higher average and diversity of group-specific CAZymes among the four AOA groups (Figure 3.3B), in particular those belonging to glycoside hydrolases (GH), which can be seen as important for carbohydrate degradation given their ability to cleave glycosidic bonds [341]. The GHs that I.1b are predicted to encode can be characterized by the carbohydrates they can degrade. For example, GH1 and GH63 (Figure 3.5A) are predicted to be α -glucosidases that can aid with the breakdown of starches and complex sugars [126]. Another potential carbohydrate substrate is chitin, which is acted on by GH19 (Figure 3.5A), which is an endo-acting hydrolase enzyme [342, 343]. The potential to cleave chitin, a major component of fungal cell walls that are common within soils [344], might mean I.1b can contribute to the breakdown of recalcitrant organic matter. Other CAZymes found specifically within I.1b genomes include GH3 (Figure 3.5A), coding for β -xylosidases [345], which are found in a diversity of bacteria, fungi, and actinomycetes, and can act upon xylosylated phytochemicals to remove xylose groups [346]. One such xylosylated phytochemical is xylan which is a common component of hemicellulose of land plants and widely found in the cell walls of seaweeds [346], yet can also serve alternative roles such as antimicrobial and antioxidant functions [347]. In addition, α -mannosidase are predicted by GH38 (Figure 3.5A) [348], which showed the potential to hydrolyze all linkage types of α -glycosidically linked mannose disaccharides [349] and is present both in eukaryotic species [350] and bacterial representatives as well [351, 352]. There is the presence of GH133 (Figure 3.5A), which is predicted to encode for amylo- α -1,6-glucosidase (EC 3.2.1.33), which can show hydrolytic activity towards α -1,6-glycosidic bonds of branched substrates [353], or act as a debranching enzyme that can help to remove the branches on amylopectin and in doing so generates maltotriose, maltose, and other linear oligosaccharides [354].

The I.1b genomes were also associated with a wide diversity and heterogeneity among species for GHs predicted to operate upon mannose/mannans and that is potentially metabolized by GH113, which links to mannanase and mannosidase [355]. The GH26 enzyme connect to β -mannanases and are responsible for the cleavage of residues on β -linked mannopyranose whereas GH130 enzyme are exo-acting β -mannosidase phosphorylases [356]. GH130 is also capable of cleaving the β -1,2- and β -1,4-mannosyl linkages between a range of polysaccharides [357, 358]. Gapseq also predicted beta-(1,4)-mannan degradation predominantly within *Ca. Nitrosocosmicus* species (Supplementary Table 10). The presence of β -mannan polysaccharides are typical components of cell walls and storage molecules for plants [359] and has been shown to be degraded by a large diversity of microorganisms and fungi [360, 361]. In terms of transportation potentials, all I.1b species contain a predicted transporter for xylan and mannose, yet all lack transporters for substrates like chitin (Figure 3.5C), yet the absence of these transporters, or in cases where the substrate is too large, might suggest that the certain GH are exported extracellularly, and the resulting products can be transported into the cell. However, in contrast with catabolic metabolism, the prevalence of GH and other CAZymes might instead be used to assist and modify the synthesis of polysaccharides on the cell surface or help with biofilm formation. Synthesizing such biofilms could provide AOA species, especially those that are non-motile, with protection and increased survival from changing or adverse environmental conditions or even help to facilitate cooperation with bacteria [94].

3.3.2.4 Alcohols

Many *Ca. Nitrosocosmicus* species encode for periplasmic or membrane-bound pyrroloquinoline quinone (PQQ)-dependent dehydrogenases, which can oxidize organic substrates such as sugars and alcohols, which in turn reduce electron acceptors [362]. The presence of such (PQQ)-dependent dehydrogenases is hinted at by specific substrate reactions. For example, there is the potential for *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola* species to oxidize particular alcohols (e.g., ethanol, 2-

chloroethanol, 2-propyn-1-ol) while reducing two ferricytochrome (K00114) (Figure 3.7). This usage of ethanol was likewise seen in *Pseudomonas aeruginosa* ATCC17933, which was shown to be capable of growing using ethanol as its sole source of carbon and energy [363]. In addition, a newly discovered heterotrophic marine Thaumarchaeota showed that its PPQ-dependent dehydrogenases were among the highest expressed genes and are thus suggestive to be essential for energy production, in particular since these Thaumarchaeota lack the ability to oxidize ammonia [223]. However, *Ca. Nitrosocosmicus* does not contain a predicted transporter for ethanol transportation (Supplementary Table 8). Thus, one possibility is that ethanol is produced by fermentation within those AOA representatives.

Species belonging to the *Ca. Nitrosocosmicus* clade also show the potential for the oxidation of both simple and complex alcohols to the reduction of NADP^+ , and thus include glycerol and ethanol (K00002/K12957/K13979) and nerol, geraniol, beta-citronellol, and (-)-citronellol (K12957) (Figure 3.8). This latter enzyme, which produces geraniol dehydrogenase (NADP^+), has shown a higher affinity for allyl alcohols such as geraniol and nerol over nonallylic citronellol [364] and is likewise present and active within many members of the genus of *Pedobacter* [365] and for *Castellaniella defragrans* [364]. Moreover, *Pseudomonas aeruginosa* can metabolize citronellol and geraniol as the only carbon and energy sources using a series of seven proteins [366] in which geraniol dehydrogenase (NADP^+) plays a predicted important role.

3.3.3 Inorganic substrates

All AOA show the potential to oxidize organochlorine compounds, trichloroethane, cis-3-chloro-2-propene-1-ol, and trans-3-chloro-2-propene-1-ol (K13953) all of which upon oxidation reduce NAD^+ (Supplementary Table 5). All genomes likewise showed the potential to encode oxidation of inorganic substrates such as hydrogen selenide and hydrogen sulfide to either the reduction of a general acceptor, or to an oxidized ferredoxin (K00392) (Supplementary Table 5). However, none of the AOA genomes within this study showed potential transporters for any of these organochlorine alcohols or for hydrogen selenide or hydrogen sulfide (Supplementary Table 8).

3.3.3.1 Manganese

All *Nitrosocosmicus*-associated genomes were associated with enzyme sequences for manganese oxidation in the presence of oxygen (Figure 3.7). The ability to oxidize manganese is found across both fungal [367, 368] and bacteria [369, 370] phyla. For example, *Leptothrix* is a dominant manganese-oxidizing bacteria in certain biofilters [371] and full-scale water supply plant biofilters [372]. However, the function and benefits of manganese oxidation for AOA remains speculative [373]. Some postulate that this oxidation enhances resistance to ROS, particularly H₂O₂ [374], and others have demonstrated that some species can derive energy from its oxidation [375]. For example, a species from the *Nitrospira* phylum could solely use manganese as an energy source and could likewise use the resulting energy for autotrophic growth [376]. In many species that oxidize manganese, its activity is on the outer surface of the cells, leading to oxidized manganese to coat their surfaces [373]. Therefore, a potential experiment to test *Ca. Nitrosocosmicus* usage is the detection of oxidized manganese nodules on their surfaces or within the surrounding media.

3.3.3.2 Sulfite

Many *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola* species also show the presence of sulfite oxidation pathways. Examples include sulfite oxidation V (SoeABC) and sulfite oxidation I found in all *Ca. Nitrosocosmicus* and *Ca. Nitrosodeserticola* but not *Ca. N. franklandianus*, and sulfite oxidation IV found in both *Ca. Nitrosodeserticola* species and three out of five *Ca. Nitrosocosmicus* species excluding *Ca. N. franlandus* and *N. oleophilus* (Figure 3.6D). Each sulfite oxidation pathway is slightly different in terms of the compound it reduces. For example, sulfite oxidation I reduces c-type cytochrome, which this thesis proposes that AOA can potentially produce and which is the central metabolism of *Starkeya novella* and *Thiobacillus novellus* [377]. Whereas sulfite oxidation V (SoeABC) (Figure 3.6D) involves the reduction of menaquinone, to which AOA show a wide diversity of biosynthetic pathways for synthesizing many different kinds/types (Supplemental table 8). Sulfite oxidation V is used by the microaerophilic bacterium *Aquifex aeolicus* as its primary source of electrons [378] and is also shown

within the genomes of a wide variety of Thioalkalivibrio species [379]. Sulfite oxidation IV couples the oxidation of sulfite to sulphate with the combination of H₂O and oxygen to form hydrogen peroxide and thus is unlikely used for energy production on account it not reducing or oxidizing any cofactors or electron carriers. However, due to sulfite being both highly reactive and toxic [211]. Potentially rising to toxic levels within cells from general metabolisms such as recycling sulfur-containing metabolites/amino acids [211] or being transported by a generalist transporter there is the possibility that these pathways of sulfite oxidation may be instead a method to remove/detoxify it [211]. The detoxification of sulfite might be useful or required for *Ca. Nitrosocosmicus* species due to their lack of motility [19, 32–34], which if present might allow them to get away from potential toxic levels of sulfite.

3.3.3.3 Bacterioferritin

All I.1b genomes show the presence of Fe²⁺:oxygen oxidoreductase (i.e., bacterioferritin) (Supplementary Table 5), which is likewise present within a variety of bacteria, such as nitrogen fixation in *Azorhizobium caulinodans* [380], and pathogens such as *Agrobacterium tumefaciens* [381]. However, this enzyme and the associated reaction are predicted to be useful for iron storage instead of catabolic or energy metabolism. Bacterioferritin are spherically shell-like iron storage proteins that catalyze iron oxidation and hydrolysis inside the protein shell, resulting in a hydrated ferric oxide core within the protein cavity [382]. More specifically, bacterioferritin stores free iron in the form of Fe³⁺ and can release it upon need in the form of Fe²⁺ when it is required by the cell for metabolic activities [383]. The purpose of such a reaction is unknown because I.1b species, like other AOA, use copper over iron for many cofactors and enzymes [222]. One possibility is that this enzyme can be used to remove the high reactivity Fe²⁺, which can, in turn, create the toxic ROS species [384]. However, Gapseq does predict that all AOA can biosynthesize heme a (Supplementary Table 10), which among many functions can have roles in oxygen transport and activation, DNA synthesis, gene regulation, and protease/peptidase [384, 385]. This does lead to the possibility that iron has a greater role within AOA than previously found but requires verification and in-vitro experimentation to verify.

3.3.3.4 Mercury

Ca. Nitrosocosmicus sp. SS, *Ca. N. subpetralis*, *N. gargensis*, and *N. viennensis* genomes encode the potential to oxidize elemental mercury to Hg^{2+} to reduce NADP^+ via mercuric reductase (K00520) (Figure 3.8). The potential to transport elemental mercury is prevalent within 53 of the 58 genomes (Supplementary Table 8). However, it is likely that instead of operating in the forward direction, which would result in the production of highly toxic Hg^{2+} , mercuric reductase, which is part of the mercury resistance system, catalyzes the reduction of Hg^{2+} to elemental mercury [386], which is both less toxic but also capable of leaving the cell by diffusion or by the previously mentioned transporters.

3.4 Mixotrophy

Based on the definitions in the introduction, it may be that AOA species within the I.1b group and *Ca. Nitrosocosmicus* genus specifically are absorbotrophic mixotrophs. This form of mixotrophy involves cells actively or passively up-taking organic matter from the environment [101]. This is due to the genomes, particularly I.1b and *Ca. Nitrosocosmicus*, showing an abundance and diversity of predicted enzymes, for example, those involved with pyruvate or alcohol oxidation (Supplementary Table 5), transporters in particular amino acids and TCA intermediates (Supplementary Table 8), and pathways such as the TCA cycle and a number of amino acid catabolic pathways (Supplementary table 10), compared to other AOA groups to obtain and support their anabolic and catabolic demands.

3.5 Alternatives to mixotrophy

While the presence of these enzymes and transporters might suggest and support mixotrophy, other potential explanations for their presence does exist. For example, certain organic substrates can sequester copper ions or become toxic at lower pH [47, 67, 222]. Many AOA depend on copper for critical enzymes and elements of their ETC [47, 55] and even organic acids can become toxic under acidic environmental conditions, as small protonated acids can diffuse easier through the cell membrane [387]. Upon entering the more neutral cytosol, these acids undergo dissociation, which can serve to acidify the cytoplasm and disrupt the proton motive force [388]. In addition, certain organic substrates (e.g., aliphatic

alcohols) can inhibit the *Amo* complex and interrupt ammonia oxidation [389]. Thus, the presence of pathways or transporters to either remove or actively metabolize these organics to detoxify can be a potential reason for their presence.

Organic matter can also be significantly higher within soils than in marine environments [390]. Given that known *Ca. Nitrosocosmicus* clade representatives are non-motile, there might be greater evolutionary pressures to preserve these proteins over species that are either motile or exist in more oligotrophic aquatic environments. Another potential explanation for the presence of these transporters and enzymes is the removal of resources, substrates, and nutrients needed by competitors, and thus obtained an advantage [391]. This hypothesis has been applied to certain species of *Picocyanobacteria*, which might use this tactic to become the dominant species within their habitat [392], and if verified within AOA it might explain their high dominance within certain environments [21]. Another hypothesis postulates that these proteins are evolutionary relics from AOA basal ancestors. More specifically, evolutionary analysis showed that as AOA switched to ammonia oxidation, there was a corresponding minimization of their ability to grow heterotrophically, losing pathways such as glycolysis and families involved in amino acid catabolism [55]. Such a hypothesis of genomic relics might explain why all substrates tested for *Ca. N. oleophilus* failed to encourage growth [393] but also potentially explains why there is a high degree of heterogeneity even between phylogenetically related species as each loses proteins at different rates as there is little or no selective pressures to keep these genes.

Despite showing predictions and prevalence of both transporters and reactions, some species fail to show stimulation or increase growth or activity for organic substrates. For example, *Ca. Nitrosocosmicus oleophilus* [32], and *Ca. N. franklandianus* [34], or in the case of *Ca. N. hydrocola*, showed stimulation in one study [33] but not another [320]. Two possible explanations are that ammonia is a poor energy source or signalling limitations. It is well established that the oxidation of ammonia is the primary method by which AOA generates ATP and energy [11, 394, 395]. Yet, archaeal ammonia monooxygenase preferred substrate has been documented to be NH_3 [42]. As the pH drops from its equilibrium pKa of 9.25, the environment becomes increasingly populated by NH_4^+ [151]. This decrease

in concentrations of available substrate for species within neutral or acidic environments [34, 40, 41] might not pose a problem to oligotrophic AOA due to the presence of ammonia concentrating methods, such as the S-layer [44], the high-affinity amt-type ammonium transporter [32, 34], or a high surface to volume ratio (S/V) [42]. However, low ammonia concentrations might be problematic for *Ca.* Nitrosocosmicus species that have the lowest affinity for ammonia [33, 45]. When *Ca.* Nitrosocosmicus lowered affinity is coupled with low concentrations of their preferred NH₃ substrate, which is a relatively poor energy-yielding substrate [396]. It might force cells to rely solely on basal metabolism or more specifically lead to cells being able to metabolically afford the high energy costs (i.e., the alternative metabolism energy threshold) required to synthesize the transporters and enzymes to support alternative metabolism. It may be that only when *Ca.* Nitrosocosmicus spp. obtain additional energy, such as from substrates with transporters that are continuously expressed or from a sudden spike in ammonia concentrations, that cells overcome this threshold. Likewise, it is possible that a specific stimulus or condition must be met to trigger the signalling pathways for the synthesis of this alternative metabolism, such as low ammonia concentrations or quorum sensing molecules from neighbouring bacteria.

3.6 Benefits of mixotrophy

The classification of microorganisms conducting mixotrophy is increasingly being applied and discovered as an alternative trophic strategy for a broad spectrum of microorganisms and is thus emerging as a more widespread phenomenon, likely as more refined and in-depth experimental work and genome mining occurs. Examples of this include many microeukaryotes in a variety of aquatic habitats [397], nitrite-oxidizing bacteria [6], and even some AOB such as *Nitrosomonas europaea* can assimilate low molecular weight organic compounds [398]. The benefits of being mixotrophic include having the ability to overcome mismatches between crucial resources/nutrients and the environmental supply [399] or within environments that are oligotrophic or have fluctuations in crucial nutrients [400, 401] on account of being able to mineralize, supplement, or replace a broader range of substrates and thus increase and satisfy their needs in limiting conditions [402]. This may be especially useful when competition is high or

when the environment changes due to stress or disturbance [403]. Using exogenous substrates from the environment can also help to offset costs in terms of the amount of energy and number of reduced electron carriers that autotrophic carbon fixation can provide. In other words, the higher number of substrates available for supplementing crucial intermediates within a cycle would give species a growth and competitive advantage over those that are strict in their trophic style [404]. This view of not being strictly autotrophic or heterotrophic is strengthened by the idea that, within their natural habitats, and in particular terrestrial ones, microorganisms are likely to encounter a wide diversity of potential energy and carbon substrates, some of which might not be in high concentrations. However, if microorganisms contain transporters and enzymes for those substrates, or strategies for concentrating substrates such as phosphorylation of glucose [405], or trapping them within the extracellular matrix or S-layer as proposed for *Nitrosopumilus maritimus* for ammonia [42], then even substrates at low concentrations can be used.

Chapter 4

Conclusion

4.1 Concluding remarks

The discovery that some archaea can conduct ammonia oxidation was revolutionary in unlocking the metabolic potential of an entirely new branch of the tree of life, with global biogeochemical implications. However, a longstanding unanswered question relates to whether all AOA rely solely on autotrophic ammonia oxidation or whether clues from their genomes [33, 222] and *in vitro* work [21, 61] implicate several representatives in additional metabolic potential. The primary objective of this thesis research was to perform a comparative genome analysis of AOA representatives, with an emphasis on the genomes of I.1b species given prior evidence for metabolic diversity in this group [33]. The results here show that I.1b genomes are larger and possess correspondingly greater metabolic diversity, particularly with respect to transformations of inorganic compounds, organic compounds, including fermentation. The results also demonstrate extensive heterogeneity among AOA genomes, and future work should test whether these variations are driven by factors related to the environment or phylogenetics.

Overall, the results presented here suggests that AOA can “switch” to alternative substrates to support carbon, nitrogen, and energy demands when provided suitable conditions and substrates. Such a potential for “mixotrophy” is compatible with a primarily chemolithoautotrophic metabolism. Indeed, two studies previously used MAR-FISH to demonstrate that biofilms dominated by a *Ca. Nitrosocosmicus* I.1b species within different WWTP systems failed to assimilate labelled bicarbonate from the environment despite doing so *in vitro* when supplied only with ammonia and bicarbonate [21, 33]. Such a possibility to use alternative carbon and energy sources may help to explain why I.1b species can outnumber AOB by many orders of magnitude, especially in soils and relatively substrate-rich environments [21]. Alternative substrate usage can also potentially explain why two clades of AOA that are typically found in terrestrial environments have never been documented as being autotrophic based on SIP incubations [163].

Although this thesis research profiled genome-encoded enzymes for direct use of substrates, an emphasis of this work was using the presence of transporters as a primary predictor of alternative metabolism. This is particularly important because our results demonstrate that most AOA genomes share a common core of metabolic reactions. However, one limitation of this work is that the substrates of predicted transporters are not always known, as is also true for the directionality of reactions encoded by predicted enzymes. In addition, gene regulation stimuli and related signalling pathways typically cannot be inferred using sequences and homology alone. Likewise, caution is needed when making conclusions based on a limited number of isolates and genomes to explain ecological or metabolic phenomena, as well as make generalizations about the four groups of AOA species. Instead, this work provides a collection of potential AOA substrates that are worth testing with subsequent experimental work with additional methodologies. The genomes analyzed in this research together demonstrate that I.1b and *Ca. Nitrosocosmicus* representatives possess extensive additional genes that imply a strong potential for conducting alternative metabolisms beyond strictly relying on ammonia oxidation and bicarbonate fixation.

4.2 Dark matter and future research

This project endeavored to include all AOA genomes that passed the CheckM requirements, and this study now represents the most extensive metabolic comparative analysis of AOA. However, there are still unknowns and weaknesses. For example, Lehtovirta-Morley *et al.* expressed the need for caution when inferring differences between physiological characteristics of *Ca. Nitrosotalea* on a limited number of isolates [67]. Thus, the same caution can be expanded to apply to all AOA [67]. In particular, I.1a-associated and ThAOA have a lower number of representative genomes such that it is not possible to make generalizations about their groups' metabolic capabilities as can be better done with I.1a and I.1b genomes. Likewise, the large proportion of genomic “dark matter” within each genome leads to a large number of proteins identified by annotation software but without an established name, function, or reaction [406], with some estimates predicting that between 30-50% of genes within each genome lacking

any functional annotations [407]. Because at least 30% of these unannotated genes are likely to contribute metabolic functions [408], this leaves open a significant gap in understanding of the metabolic potential of any given AOA species. This problem is becoming well documented for Nitrososphaerota members because many of their genes lack homologs in databases and thus cannot be identified and used for functional analysis [202]. For instance, on average, the 58 genomes of AOA assessed in this study had 42.4% of their genomes uncharacterized or not given a protein name. Genomes from I.1b showed above average abundances and a wider range of proteins without names from slightly above *Nitrososphaera* sp. WS208 at 42.6% to *Nitrosopolaris wilkensis* at 68.4%. Without such identification, it is possible to state that a protein is present, but nothing can be said about its use by the cell or possible alternative metabolism. Therefore, failing to detect pathways/proteins within databases does not definitively indicate their absence as they might be controlled by yet-to-be-established homologs [406]. Likewise, there is great uncertainty in transporters specificity or discrimination for substrates [31]. While the TCDB can provide predictions about potential substrates that transporters can translocate, this is based on homology and might not contain the complete range of substrates. As a result, *in vitro* experiments are required to determine what, if any, substrates predicted here can be transported [31].

One major drawback of relying on the KEGG and TCDB databases is the lack of information about directionality. More specifically, there is a lack of rules that govern the direction in which direction reactions and transporters proceed. Without this information, it is hard to determine a species' metabolic potential and the purpose of reactions in a cell. Future work should test substrates predicted *in silico* using *in vitro* or *in situ* experimentation. For example, *Candidatus* Kuenenia stuttgartiensis, an anaerobic ammonium-oxidizing (anammox) bacteria, revealed several instances where genomic predictions were not in alignment with *in vitro* experiments [409]. For example, conducting alpha-ketoglutarate biosynthesis for the TCA cycle despite not having an annotated citrate synthase and synthesizing methionine despite lacking an annotated pathway for its biosynthesis and, contrary to genomic data, they were unable to use acetate as a carbon or energy source [409].

Certain enzymes also show promiscuity in the pathways or the functions they facilitate. Therefore, enzymes used in autotrophic pathways can be similar or the same as those used in heterotrophic pathways and hence give a false positive for function [177]. Likewise, many copper-based enzymes show promiscuity, suggesting they can use other substrates for energy-yielding reactions [227]. Although the *in silico* comparative analysis done here revealed potential substrates and metabolic capabilities, these predictions alone cannot determine actual functional physiology or metabolism [410]. Thus, the next step is confirmation using techniques that involve either *in vitro* or *in situ* techniques to measure and detect metabolisms predicted here. For example, stable-isotope probing (SIP) [410] can be used to measure anabolic metabolism. Alternatively, tetrazolium dyes can reveal catabolic metabolism given that in their oxidized state they are colourless, but become coloured when reduced by cellular oxidoreductases (e.g., NADH dehydrogenase), and this change can be measured either qualitatively or quantitatively using a range of colorimetric metabolic assays [411, 412]. The use of these dyes has been used to determine the metabolic activity of both bacteria [413] and eukaryotes [414] and is especially appealing because catabolic activity can be measured even without cell division [415]. Methods such as SIP and tetrazolium dyes can trigger false positives, especially if the culture is enriched. AOA that are flanked by non-AOA species can result in cross-feeding, which is when microorganisms share metabolites either actively or passively between themselves [5, 416].

Three potential experimental strategies exist to address the cross-feeding problem. The first approach obtain pure cultures or metabolically inhibit the contaminating bacteria via the selective additions of antibiotics, thereby suppressing any competing bacteria [259]. However, caution is required because the antibiotics could potentially harm the AOA of interest or cause the susceptible bacteria to spill nutrient-rich substrates, which could either dilute or cause a false negative/positive for the substrate being tested.

In addition, unpublished work by Laura Sauder showed that as the culture of *Ca. N. hydrocola* became increasingly purified, growth rates were reduced. This resistance to purification has been noted for many other AOA species [417], as well as decrease in growth rates [33, 68]. All of which suggests

that either the antibiotics were having an indirect effect on AOA or that the surrounding bacteria were necessary for growth. Another option is to conduct SIP on a macromolecule that shows more rapid or continuous synthesis, such as RNA, which can also bring the benefit of reflecting what proteins or pathways are active under specific substrate exposure [418]. Because cross-feeding arises due to stable media which accumulate waste products, one solution is to have media continuously flowing through the culture, or more specifically utilize a technique known as flow-through stable isotope probing (FLOW-SIP) [108]. This, like other SIP methods, would use stably labelled substrates, which would determine whether anabolism metabolism is occurring. However, unlike traditional methods in which the media is stationary, which increases the chances of cross-feeding, FLOW-SIP instead relies on a biofilm or culture growing on a stationary surface, whereas new media is dripped onto the culture, and waste media is removed simultaneously. Other methods to elucidate metabolism include transcriptomics, proteomics [94], and metabolomics [109]. The combination of these methods has been used and provided evidence of mixotrophy for multiple species of bacteria [419, 420]. However, as noted for urea in soils, the oxidation rate of specific substrates cannot be directly deduced from meta-transcriptomic data [421] and therefore care and attention is required to ensure accurate results and interpretation of them. In addition, metabolomics has shown to be well adapted to detect both the functional state of a wide selection of organisms and any changes within the organism's metabolite, and hence metabolic patterns [422]. However, all these techniques increase in difficulty, time, and cost as the culture becomes increasingly complex or populated by species that are not the target.

Another method that can measure a microorganism's metabolic activity is the ratio of reduced to oxidized electron carriers, particularly NADH/NAD⁺ [423]. In particular, a lowered ratio might indicate active and efficient electron transport and turnover of NADH to NAD⁺ [259]. Likewise, using the ratio of ATP/ADP, precisely a decreased proportion can be indicative of higher energy consumption and more active metabolism [424] and has been used to predict possible mixotrophic usage of acetate within *in vitro* anammox cultures [259]. However, these techniques face the same problems of cross-feeding as those relying on SIP.

Future studies assessing possible growth of AOA species should be cautious when using the *amoA* gene and its transcripts, especially when providing the species in question with alternative substrates because multiple studies have demonstrated that growth of AOA can become uncoupled or uncorrelated to the *amoA* transcripts [19, 21, 63]. Finally, assessing the links between AOA genomic diversity, habitat preference, molecular adaptations, and biogeography will ultimately rely on technological and methodological developments that will enable large-scale AOA population genomics studies, high-throughput characterization of AOA physiology, gene knockouts, and characterization of their microhabitat *in situ* [176]. In addition, metabolic predictions based on metagenomic studies can also support the design of media and consequently facilitate the cultivation of uncharacterized microorganisms [425]. To understand their ecological impact, it is important to identify and characterize in detail the metabolic functions and physiological versatility of more representatives of these archaea [83]. However, many genes encoding enzymes in *Thermoproteota* do not have homologs in databases and cannot be identified using the bioinformatics approach, including a malonyl-CoA reductase in *N. maritimus* [72]. Each of these techniques have setbacks and limitations, but if applied appropriately and accurately can help in answering whether the potential substrates outlined here can support the metabolisms of AOA.

4.3 Importance

The widespread distribution and relatively high abundance of AOA in aquatic [81] and terrestrial [426] environments implies longstanding adaptations to myriad global niches. Indeed, the AOA dominate over their bacterial counterparts in a variety of terrestrial environments, such as Arctic [111], ammonium-depleted [23], and acidic [24, 25] soils. This is also true for marine environments, both for open water and sediments [427, 428], geothermal habitats [429], as well as artificial or engineered environments, such as aquarium biofilters and aquaculture systems [112, 430, 431]. Together, these studies suggest that AOA play important ecological roles in the structure and organization of microbial communities. More broadly, the activity of such globally relevant AOA has profound impacts on biogeochemical cycling of carbon and nitrogen in most environments [11, 432]. Given such relevance, better understanding the genomes of

AOA is an important way to develop strategies to utilize, maintain, and promote their use and importance both in industry and the environment.

This thesis research presents genomic evidence that AOA possess previously unrecognized metabolic diversity, which greatly expands potential substrates that can be used by AOA representatives. The results largely suggest expanded roles for AOA in carbon and nitrogen cycling than initially understood. These roles can include mineralization and decomposition of organic nitrogen and carbon sources to compounds that can be redirected to ammonia oxidation or to better support autotrophic growth. In addition, this thesis research predicts that AOA can play a role in the cycling of sulfur, and metals such as iron and manganese.

Future experimental work should test the predictions generated from this comparative genome analysis. If alternative metabolisms are indeed verified, this additional knowledge could better explain the distributions, presence, and rate of metabolism of organic substrates within the environment. This information can help improve biogeochemical models and predictions, which will only become more important in relation to climate change. With AOA also being implicated as a source of N_2O , a potent greenhouse gas and known ozone destroyer [84, 433], the need to understand the proper role and metabolism of AOA becomes even more urgent. The results presented here may also offer industrial relevance, particularly given the vast diversity of peptidases and CAZymes predicted to be encoded by I.1b genomes, which could help the breakdown and metabolism of recalcitrant organic matter. In addition, many GHs originating from archaea show high thermostability and thus can be potentially used for high temperature processes, such as biofuel generation and biomass conversion [348].

Some evidence also suggests that AOA can transform complex pharmaceutical compounds in co-metabolism with ammonia oxidation [116, 117], including sulfonamide antibiotics [118]. As well as biodegrading micropollutants [119] such as tertiary amines, mianserin and ranitidine [120] and with the vast diversity of potential drug or antibiotic transporters predicted within I.1b the potential to breakdown and use xenobiotics could be higher than what is currently considered.

Although previous AOA comparative genome studies have explored the physiology and adaptations of AOA within the environment, one limitation is that previous studies did not make the metabolism of AOA the primary focus. For example, a previous study of 44 AOA genomes focused solely on marine AOA and their adaptations to different aquatic habitats [434]. In some, while alternative metabolic potentials are reviewed, a very narrow range of substrates was assessed, such as urea, cyanate, and glutamate [396, 435]. In some comparative studies that assessed metabolism, their limitations lay within a very narrow range of species, which likely biased the conclusions. For example, Zou et al.'s [436] findings that AOA in estuaries has higher abundances of genes supporting alternative metabolisms (i.e., amino acid, carbohydrates, and transport) were associated with small sample sizes of each group and included no I.1b members. Another study showed that marine and terrestrial AOA share well-conserved nitrogen and carbon metabolic pathways, and that terrestrial AOA possess extensive environmental adaptations that include detoxification and carbohydrate conversions [103]. However, this study used only seven genomes and did not include representatives from *Ca. Nitrosocaldus*, *Ca. Nitrosomarinus*, *Ca. Nitrosopolaris*, *Ca. Nitrosotenuis*, *Ca. Nitrosocosmicus*, or *Ca. Nitrosothermus* genera [103]. Finally, almost all comparative studies suffer from manually selecting the metabolic pathways or enzymes of interest. Although this allows for rapid determination of traits of interest, it can lead to false negatives or miss additional metabolic potentials of AOA representatives.

Even though the comparative study done here was limited by homology (i.e., protein sequences known to associate with specific enzymes or transporters), currently established or known species, and a high degree of unknown or unidentified proteins, it is an inclusive AOA comparative genomics study that includes all species both known and having a high degree of completeness (as performed by CheckM). In addition, this study uniquely assessed fermentation,

anaplerosis, polyamine metabolism (both degradation and biosynthesis), inorganics, and finding the total metabolic profiles using a pipeline to link KEGG IDs to reactions, which minimized objectivity and manual searches. Finally, this study assessed metabolic traits that are lineage-specific by using lineage-specificity analysis. Thus, despite limitations related to the current states of knowledge, this project addresses limitations of previous studies and showed that AOA, and I.1b in particular, can potentially use alternative substrates to support growth and metabolism.

References

1. Mills MM, Ridame C, Davey M, La Roche J, Geider RJ. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* 2004; **429**: 292–294.
2. Canfield DE, Glazer AN, Falkowski PG. The evolution and future of Earth's nitrogen cycle. *Science* 2010; **330**: 192–196.
3. Holmes DE, Dang Y, Smith JA. Chapter Four - Nitrogen cycling during wastewater treatment. In: Gadd GM, Sariaslani S (eds). *Adv. Appl. Microbiol.* 2019. Academic Press, pp 113–192.
4. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. *Nat Rev Microbiol* 2018; **16**: 263–276.
5. French E, Kozłowski JA, Mukherjee M, Bullerjahn G, Bollmann A. Ecophysiological characterization of ammonia-oxidizing archaea and bacteria from freshwater. *Appl Environ Microbiol* 2012; **78**: 5773–5780.
6. Daims H, Lebedeva EV, Pjevac P, Han P, Herbold C, Albertsen M, et al. Complete nitrification by *Nitrospira* bacteria. *Nature* 2015; **528**: 504–509.
7. Purkhold U, Pommerening-Röser A, Juretschko S, Schmid MC, Koops H-P, Wagner M. Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* 2000; **66**: 5368–5382.
8. Klotz MG, Arp DJ, Chain PSG, El-Sheikh AF, Hauser LJ, Hommes NG, et al. Complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceanus* ATCC 19707. *Appl Environ Microbiol* 2006; **72**: 6299–6315.
9. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004; **304**: 66–74.
10. Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk H-P, Schleper C. Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic Crenarchaeota in nitrogen cycling. *Environ Microbiol* 2005; **7**: 1985–1995.
11. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 2005; **437**: 543–546.
12. Brochier-Armanet C, Gribaldo S, Forterre P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of Archaea and Eucarya. *Biol Direct* 2008; **3**.
13. Schleper C, Jurgens G, Jönuscheit M. Genomic studies of uncultivated archaea. *Nat Rev Microbiol* 2005; **3**: 479–488.
14. Nicol GW, Schleper C. Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? *Trends Microbiol* 2006; **14**: 207–212.
15. Pester M, Schleper C, Wagner M. The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* 2011; **14**: 300–306.
16. Daebeler A, Abell G, Bodelier P, Bodrossy L, Frampton D, Hefting M, et al. Archaeal dominated ammonia-oxidizing communities in Icelandic grassland soils are moderately affected by long-term N fertilization and geothermal heating. *Front Microbiol* 2012; **3**: 352–366.
17. Zhang L-M, Hu H-W, Shen J-P, He J-Z. Ammonia-oxidizing archaea have more important role than ammonia-oxidizing bacteria in ammonia oxidation of strongly acidic soils. *ISME J* 2012; **6**: 1032–1045.
18. De La Torre JR, Walker CB, Ingalls AE, Könneke M, Stahl DA. Cultivation of a thermophilic ammonia oxidizing archaeon synthesizing crenarchaeol. *Environ Microbiol* 2008; **10**: 810–818.
19. Alves RJE, Kerou M, Zappe A, Bittner R, Abby SS, Schmidt HA, et al. Ammonia oxidation by the arctic terrestrial Thaumarchaeote *Candidatus Nitrosocosmicus arcticus* is stimulated by increasing temperatures. *Front Microbiol* 2019; **10**: 1571–1587.

20. Lehtovirta-Morley LE, Stoecker K, Vilcinskas A, Prosser JI, Nicol GW. Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *Proc Natl Acad Sci USA* 2011; **108**: 15892–15897.
21. Mußmann M, Brito I, Pitcher A, Sinninghe Damsté JS, Hatzenpichler R, Richter A, et al. Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci USA* 2011; **108**: 16771–16776.
22. Ochsenreiter T, Selezi D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol* 2003; **5**: 787–797.
23. Leininger S, Urich T, Schlöter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 2006; **442**: 806–809.
24. Stopnisek N, Gubry-Rangin C, Höfferle S, Nicol GW, Mandic-Mulec I, Prosser JI. Thaumarchaeal ammonia oxidation in an acidic forest peat soil is not influenced by ammonium amendment. *Appl Environ Microbiol* 2010; **76**: 7626–7634.
25. Yao H, Gao Y, Nicol GW, Campbell CD, Prosser JI, Zhang L, et al. Links between ammonia oxidizer community structure, abundance, and nitrification potential in acidic soils. *Appl Environ Microbiol* 2011; **77**: 4618–4625.
26. Zou D, Wan R, Han L, Xu MN, Liu Y, Liu H, et al. Genomic characteristics of a novel species of ammonia-oxidizing archaea from the Jiulong river estuary. *J Appl Environ Microbiol* 2020; **86**: 20–35.
27. Koga Y. Thermal adaptation of the Archaeal and Bacterial lipid membranes. *Archaea* 2012; **2012**: e789652.
28. Siliakus MF, van der Oost J, Kengen SWM. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles* 2017; **21**: 651–670.
29. Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature* 2009; **461**: 976–979.
30. Limpiyakorn T, Fühacker M, Haberl R, Chodanont T, Srithep P, Sonthiphand P. amoA-encoding archaea in wastewater treatment plants: a review. *Appl Microbiol Biotechnol* 2013; **97**: 1425–1439.
31. Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol* 2012; **20**: 523–531.
32. Jung M-Y, Kim J-G, Sinninghe Damsté JS, Rijpstra WIC, Madsen EL, Kim S-J, et al. A hydrophobic ammonia-oxidizing archaeon of the *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment. *Environ Microbiol Rep* 2016; **8**: 983–992.
33. Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of *Candidatus Nitrosocosmicus exaquare*, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J* 2017; **11**: 1142–1157.
34. Lehtovirta-Morley LE, Ross J, Hink L, Weber EB, Gubry-Rangin C, Thion C, et al. Isolation of ‘*Candidatus Nitrosocosmicus franklandus*’, a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol Ecol* 2016; **92**: 57–67.
35. Park S-J, Ghai R, Martín-Cuadrado A-B, Rodríguez-Valera F, Chung W-H, Kwon K, et al. Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments. *PLOS ONE* 2014; **9**: 331–441.
36. Bai Y, Sun Q, Wen D, Tang X. Abundance of ammonia-oxidizing bacteria and archaea in industrial and domestic wastewater treatment systems. *FEMS Microbiol Ecol* 2012; **80**: 323–330.
37. Pan K-L, Gao J-F, Li H-Y, Fan X-Y, Li D-C, Jiang H. Ammonia-oxidizing bacteria dominate ammonia oxidation in a full-scale wastewater treatment plant revealed by DNA-based stable isotope probing. *Bioresour Technol* 2018; **256**: 152–159.
38. Han Y, Zhang M, Chen X, Zhai W, Tan E, Tang K. Transcriptomic evidences for microbial carbon and nitrogen cycles in the deoxygenated seawaters of Bohai Sea. *Environ Int* 2022; **158**: 106889.

39. Kim J-G, Park S-J, Damsté JSS, Schouten S, Rijpstra WIC, Jung M-Y, et al. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci USA* 2016; **113**: 7888–7893.
40. Shen J-P, Zhang L-M, Di HJ, He J-Z. A review of ammonia-oxidizing bacteria and archaea in Chinese soils. *Front Microbiol* 2012; **3**.
41. Gubry-Rangin C, Nicol GW, Prosser JI. Archaea rather than bacteria control nitrification in two agricultural acidic soils. *FEMS Microbiol Ecol* 2010; **74**: 566–574.
42. Jung M-Y, Sedlacek CJ, Kits KD, Mueller AJ, Rhee S-K, Hink L, et al. Ammonia-oxidizing archaea possess a wide range of cellular ammonia affinities. *ISME J* 2021; **16**: 272–283.
43. Liu X-Y, Tao K-Y, Sun J, He C-Q, Cui J, Chen X-P. The introduction of woody plants for freshwater wetland restoration alters the Archaeal community structure in soil. *Land Degrad Dev* 2017; **28**: 1933–1942.
44. Li P-N, Herrmann J, Tolar BB, Poitevin F, Ramdasi R, Bargar JR, et al. Nutrient transport suggests an evolutionary basis for charged archaeal surface layer proteins. *ISME J* 2018; **12**: 2389–2402.
45. Jung M-Y, Sedlacek CJ, Kits KD, Mueller AJ, Rhee S-K, Hink L, et al. Ammonia-oxidizing archaea possess a wide range of cellular ammonia affinities. *ISME J* 2022; **16**: 272–283.
46. Carini P, Dupont CL, Santoro AE. Patterns of thaumarchaeal gene expression in culture and diverse marine environments. *Environ Microbiol* 2018; **20**: 2112–2124.
47. Gwak J-H, Jung M-Y, Hong H, Kim J-G, Quan Z-X, Reinfelder JR, et al. Archaeal nitrification is constrained by copper complexation with organic matter in municipal wastewater treatment plants. *ISME J* 2020; **14**: 335–346.
48. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinform* 2019; **35**: 4162–4164.
49. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017; **8**: 28–36.
50. Al-Ajeel S, Spasov E, Sauder LA, McKnight MM, Neufeld JD. Ammonia-oxidizing archaea and complete ammonia-oxidizing *Nitrospira* in water treatment systems. *Water Res X* 2022; **15**: 100131.
51. Yang Y, Zhang C, Lenton TM, Yan X, Zhu M, Zhou M, et al. The evolution pathway of ammonia-oxidizing archaea shaped by major geological events. *Mol Biol Evol* 2021; **38**: 3637–3648.
52. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J* 2019; **13**: 2150–2161.
53. Stahl DA, de la Torre JR. Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol* 2012; **66**: 83–101.
54. Beam JP, Jay ZJ, Kozubal MA, Inskeep WP. Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J* 2014; **8**: 938–951.
55. Abby SS, Kerou M, Schleper C. Ancestral reconstructions decipher major adaptations of ammonia-oxidizing archaea upon radiation into moderate terrestrial and marine environments. *mBio* 2020; **11**.
56. Brochier-Armanet C, Deschamps P, López-García P, Zivanovic Y, Rodríguez-Valera F, Moreira D. Complete-fosmid and fosmid-end sequences reveal frequent horizontal gene transfers in marine uncultured planktonic archaea. *ISME J* 2011; **5**: 1291–1302.
57. Spang A, Poehlein A, Offre P, Zumbärgel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* 2012; **14**: 3122–3145.
58. Deschamps P, Zivanovic Y, Moreira D, Rodríguez-Valera F, López-García P. Pangenome evidence for extensive interdomain horizontal transfer affecting lineage core and shell genes in uncultured planktonic Thaumarchaeota and Euryarchaeota. *Genome Biol Evol* 2014; **6**: 1549–1563.
59. Wang B, Qin W, Ren Y, Zhou X, Jung M-Y, Han P, et al. Expansion of Thaumarchaeota habitat range is correlated with horizontal transfer of ATPase operons. *ISME J* 2019; **13**: 3067–3079.

60. Weber G, Drickamer HG. The effect of high pressure upon proteins and other biomolecules. *Q Rev Biophys* 1983; **16**: 89–112.
61. Zhalnina KV, Dias R, Leonard MT, Quadros PD de, Camargo FAO, Drew JC, et al. 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* 2014; **9**: 648–663.
62. Sheridan PO, Raguideau S, Quince C, Holden J, Zhang L, Williams TA, et al. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nat Commun* 2020; **11**: 5494.
63. Jia Z, Conrad R. Bacteria rather than Archaea dominate microbial ammonia oxidation in an agricultural soil. *Environ Microbiol Rep* 2009; **11**: 1658–1671.
64. Verhamme DT, Prosser JI, Nicol GW. Ammonia concentration determines differential growth of ammonia-oxidising archaea and bacteria in soil microcosms. *ISME J* 2011; **5**: 1067–1071.
65. Levičnik-Höfferle Š, Nicol GW, Ausec L, Mandić-Mulec I, Prosser JI. Stimulation of thaumarchaeal ammonia oxidation by ammonia derived from organic nitrogen but not added inorganic nitrogen. *FEMS Microbiol Ecol* 2012; **80**: 114–123.
66. Prosser JI, Nicol GW. Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environ Microbiol* 2008; **10**: 2931–2941.
67. Lehtovirta-Morley LE, Ge C, Ross J, Yao H, Nicol GW, Prosser JI. Characterisation of terrestrial acidophilic archaeal ammonia oxidisers and their inhibition and stimulation by organic compounds. *FEMS Microbiol Ecol* 2014; **89**: 542–552.
68. Tourna M, Stieglmeier M, Spang A, Könneke M, Schintlmeister A, Urich T, et al. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. *Proc Natl Acad Sci USA* 2011; **108**: 8420–8425.
69. Pessi IS, Rutanen A, Hultman J. *Candidatus Nitrosopolaris*, a genus of putative ammonia-oxidizing archaea with a polar/alpine distribution. *FEMS Microbes* 2022; **3**: 626–637.
70. Park S-J, Ghai R, Martín-Cuadrado A-B, Rodríguez-Valera F, Chung W-H, Kwon K, et al. Genomes of two new ammonia-oxidizing archaea enriched from deep marine sediments. *PLOS ONE* 2014; **9**: 331–441.
71. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JA, et al. Physiological and genomic characterization of two novel marine Thaumarchaeal strains indicates niche differentiation. *ISME J* 2016; **10**: 1051–1063.
72. Pinto OHB, Silva TF, Vizzotto CS, Santana RH, Lopes FAC, Silva BS, et al. Genome-resolved metagenomics analysis provides insights into the ecological role of Thaumarchaeota in the Amazon River and its plume. *BMC Microbiol* 2020; **20**: 2–11.
73. Abby SS, Melcher M, Kerou M, Krupovic M, Stieglmeier M, Rossel C, et al. *Candidatus Nitrosocaldus cavascurensis*, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front Microbiol* 2018; **9**: 29434576.
74. Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, et al. Ammonia-oxidising archaea living at low pH: Insights from comparative genomics. *Environ Microbiol* 2017; **19**: 4939–4952.
75. Spang A, Poehlein A, Offre P, Zumbärgel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing *Candidatus Nitrososphaera gargensis*: insights into metabolic versatility and environmental adaptations. *Environ Microbiol* 2012; **14**: 3122–3145.
76. Bayer B, Hansman RL, Bittner MJ, Noriega-Ortega BE, Niggemann J, Dittmar T, et al. Ammonia-oxidizing archaea release a suite of organic compounds potentially fueling prokaryotic heterotrophy in the ocean. *Environ Microbiol* 2019; **21**: 4062–4075.
77. Qin W, Heal KR, Ramdasi R, Kobelt JN, Martens-Habbena W, Bertagnolli AD, et al. *Nitrosopumilus maritimus* gen. nov., sp. nov., *Nitrosopumilus cobalaminigenes* sp. nov., *Nitrosopumilus oxyclinae* sp. nov., and *Nitrosopumilus ureiphilus* sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. *Int J Syst Evol Microbiol* 2017; **67**: 5067–5079.

78. Ahlgren NA, Chen Y, Needham DM, Parada AE, Sachdeva R, Trinh V, et al. Genome and epigenome of a novel marine Thaumarchaeota strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ Microbiol* 2017; **19**: 2434–2452.
79. Tolar BB, Wallsgrove NJ, Popp BN, Hollibaugh JT. Oxidation of urea-derived nitrogen by Thaumarchaeota-dominated marine nitrifying communities. *Environ Microbiol* 2017; **19**: 4838–4850.
80. Ouverney CC, Fuhrman JA. Marine planktonic Archaea take up amino acids. *Appl Environ Microbiol* 2000; **66**: 4829–4833.
81. Herndl GJ, Reinthaler T, Teira E, van Aken H, Veth C, Pernthaler A, et al. Contribution of Archaea to total prokaryotic production in the deep Atlantic Ocean. *Appl Environ Microbiol* 2005; **71**: 2303–2309.
82. Teira E, van Aken H, Veth C, Herndl GJ. Archaeal uptake of enantiomeric amino acids in the meso- and bathypelagic waters of the North Atlantic. *Limnol Oceanogr* 2006; **51**: 60–69.
83. Stieglmeier M, Klingl A, Alves RJE, Rittmann SK-MR, Melcher M, Leisch N, et al. Nitrososphaera viennensis gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. *Int J Syst Evol Microbiol* 2014; **64**: 2738–2752.
84. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of “*Candidatus Nitrosopelagicus brevis*”: An ammonia-oxidizing archaeon from the open ocean. *Proc Natl Acad Sci U S A* 2015; **112**: 1173–1178.
85. Jung M-Y, Islam MdA, Gwak J-H, Kim J-G, Rhee S-K. Nitrosarchaeum koreense gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon member of the phylum Thaumarchaeota isolated from agricultural soil. *Int J Syst Evol Microbiol* 2018; **68**: 3084–3095.
86. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, von Bergen M, et al. Cyanate as an energy source for nitrifiers. *Nature* 2015; **524**: 105–108.
87. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol* 2019; **4**: 234–243.
88. Liu Q, Nishibori N, Imai I, Hollibaugh JT. Response of polyamine pools in marine phytoplankton to nutrient limitation and variation in temperature and salinity. *Mar Ecol Prog Ser* 2016; **544**: 93–105.
89. Lu X, Zou L, Clevinger C, Liu Q, Hollibaugh JT, Mou X. Temporal dynamics and depth variations of dissolved free amino acids and polyamines in coastal seawater determined by high-performance liquid chromatography. *Mar Chem* 2014; **163**: 36–44.
90. Höfle MG. Degradation of uprescine and cadaverine in seawater cultures by marine bacteria. *Appl Environ Microbiol* 1984; **47**: 843–849.
91. Mou X, Vila-Costa M, Sun S, Zhao W, Sharma S, Moran MA. Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environ Microbiol Rep* 2011; **3**: 798–806.
92. Liu Q, Lu X, Tolar BB, Mou X, Hollibaugh JT. Concentrations, turnover rates and fluxes of polyamines in coastal waters of the South Atlantic Bight. *Biogeochemistry* 2015; **123**: 117–133.
93. Noell SE, Barrell GE, Suffridge C, Morré J, Gable KP, Graff JR, et al. SAR11 cells rely on enzyme multifunctionality to metabolize a range of polyamine compounds. *mBio* 2021; **12**: e01091-21.
94. Liu L, Liu M, Jiang Y, Lin W, Luo J. Production and excretion of polyamines to tolerate high ammonia, a case study on soil ammonia-oxidizing archaeon “*Candidatus Nitrosocosmicus agrestis*”. *Environ Microbiol* 2021; **6**: 3–20.
95. Damashek J, Bayer B, Herndl GJ, Wallsgrove NJ, Allen T, Popp BN, et al. Limited accessibility of nitrogen supplied as amino acids, amides, and amines as energy sources for marine Thaumarchaeota. 2021. , 2021.07.22.453390

96. Jung M-Y, Park S-J, Kim S-J, Kim J-G, Sinninghe Damsté JS, Jeon CO, et al. A mesophilic, autotrophic, ammonia-oxidizing Archaeon of Thaumarchaeal Group I.1a cultivated from a deep oligotrophic soil horizon. *Appl Environ Microbiol* 2014; **80**: 3645–3655.
97. Hodgskiss LH, Melcher M, Kerou M, Chen W, Ponce-Toledo RI, Savvides SN, et al. Unexpected complexity of the ammonia monooxygenase in Archaea. 2022. bioRxiv. , 2022.04.06.487334
98. Sun W, Xia C, Xu M, Guo J, Wang A, Sun G. Distribution and abundance of archaeal and bacterial ammonia oxidizers in the sediments of the Dongjiang River, a drinking water supply for Hong Kong. *Microbes Environ* 2013; **28**: 457–465.
99. Wei D, Zeng S, Hou D, Zhou R, Xing C, Deng X, et al. Community diversity and abundance of ammonia-oxidizing archaea and bacteria in shrimp pond sediment at different culture stages. *J Appl Microbiol* 2021; **130**: 1442–1455.
100. Selosse M-A, Charpin M, Not F. Mixotrophy everywhere on land and in water: the grand écart hypothesis. *Ecol Lett* 2017; **20**: 246–263.
101. Schmidt S, Raven JA, Paungfoo-Lonhienne C. The mixotrophic nature of photosynthetic plants. *Functional Plant Biol* 2013; **40**: 425–438.
102. Stoecker DK, Hansen PJ, Caron DA, Mitra A. Mixotrophy in the Marine Plankton. *Annu Rev Mar Science* 2017; **9**: 311–335.
103. Kerou M, Offre P, Valledor L, Abby SS, Melcher M, Nagler M, et al. Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* 2016; **113**: 7937–7946.
104. Qin W, Zheng Y, Zhao F, Wang Y, Urakawa H, Martens-Habbena W, et al. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *ISME J* 2020; **14**: 2595–2609.
105. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Natl Acad Sci USA* 2014; **111**: 12504–12509.
106. Empadinhas N, da Costa MS. Osmoadaptation mechanisms in prokaryotes: distribution of compatible solutes. *Int Microbiol* 2008; **11**: 151–161.
107. Widner B, Mulholland MR, Mopper K. Distribution, sources, and sinks of cyanate in the coastal North Atlantic Ocean. *Environ Sci Technol Lett* 2016; **3**: 297–302.
108. Mooshammer M, Kitzinger K, Schintlmeister A, Ahmerkamp S, Nielsen JL, Nielsen PH, et al. Flow-through stable isotope probing (Flow-SIP) minimizes cross-feeding in complex microbial communities. *ISME J* 2021; **15**: 348–353.
109. Tautenhahn R, Cho K, Uritboonthai W, Zhu Z, Patti GJ, Siuzdak G. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat Biotechnol* 2012; **30**: 826–828.
110. Agogué H, Brink M, Dinasquet J, Herndl GJ. Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic. *Nature* 2008; **456**: 788–791.
111. Alves RJE, Wanek W, Zappe A, Richter A, Svenning MM, Schleper C, et al. Nitrification rates in Arctic soils are associated with functionally distinct populations of ammonia-oxidizing archaea. *ISME J* 2013; **7**: 1620–1631.
112. Bagchi S, Vlaeminck SE, Sauder LA, Mosquera M, Neufeld JD, Boon N. Temporal and spatial stability of ammonia-oxidizing archaea and bacteria in aquarium biofilters. *PLOS ONE* 2014; **9**: e113515.
113. Sauder LA, Peterse F, Schouten S, Neufeld JD. Low-ammonia niche of ammonia-oxidizing archaea in rotating biological contactors of a municipal wastewater treatment plant. *Environ Microbiol* 2012; **14**: 2589–2600.
114. Wu J, Hong Y, He X, Liu X, Ye J, Jiao L, et al. Niche differentiation of ammonia-oxidizing archaea and related autotrophic carbon fixation potential in the water column of the South China Sea. *iScience* 2022; **25**: 104333.
115. van Groenigen JW, Huygens D, Boeckx P, Kuyper TW, Lubbers IM, Rütting T, et al. The soil N cycle: new insights and key challenges. *SOIL* 2015; **1**: 235–256.

116. Han P, Yu Y, Zhou L, Tian Z, Li Z, Hou L, et al. Specific micropollutant biotransformation pattern by the comammox bacterium *Nitrospira inopinata*. *Environ Sci Technol* 2019; **53**: 8695–8705.
117. Li S, Peng L, Yang C, Song S, Xu Y. Cometabolic biodegradation of antibiotics by ammonia oxidizing microorganisms during wastewater treatment processes. *J Environ Manage* 2022; **305**: 114336–114349.
118. Zhou L-J, Han P, Yu Y, Wang B, Men Y, Wagner M, et al. Cometabolic biotransformation and microbial-mediated abiotic transformation of sulfonamides by three ammonia oxidizers. *Water Res* 2019; **159**: 444–453.
119. Helbling DE, Johnson DR, Honti M, Fenner K. Micropollutant biotransformation kinetics associate with WWTP process parameters and microbial community characteristics. *Environ Sci Technol* 2012; **46**: 10579–10588.
120. Men Y, Han P, Helbling DE, Jehmlich N, Herbold C, Gulde R, et al. Biotransformation of two pharmaceuticals by the ammonia-oxidizing archaeon *Nitrososphaera gargensis*. *Environ Sci Technol* 2016; **50**: 4682–4692.
121. Meier DV, Imminger S, Gillor O, Woebken D. Distribution of mixotrophy and desiccation survival mechanisms across microbial genomes in an arid biological soil crust community. *mSystems* 2021; **6**: e00786-20.
122. Russum S, Lam KJK, Wong NA, Iddamsetty V, Hendargo KJ, Wang J, et al. Comparative population genomic analyses of transporters within the Asgard archaeal superphylum. *PLOS ONE* 2021; **16**: e0247806.
123. Ferrer-González FX, Widner B, Holderman NR, Glushka J, Edison AS, Kujawinski EB, et al. Resource partitioning of phytoplankton metabolites that support bacterial heterotrophy. *ISME J* 2021; **15**: 762–773.
124. Kim J-G, Park S-J, Damsté JSS, Schouten S, Rijpstra WIC, Jung M-Y, et al. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci USA* 2016; **113**: 7888–7893.
125. Deng F, Li Y, Peng Y, Wei X, Wang X, Howe S, et al. The diversity, composition, and metabolic pathways of archaea in pigs. *Animals* 2021; **11**: 2139.
126. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 2009; **37**: 233–238.
127. Rao MB, Tanksale AM, Ghatge MS, Deshpande VV. Molecular and biotechnological aspects of microbial proteases. *Microbiol Mol Biol Rev* 1998; **62**: 597–635.
128. Geisseler D, Horwath WR, Joergensen RG, Ludwig B. Pathways of nitrogen utilization by soil microorganisms – A review. *Soil Biol Biochem* 2010; **42**: 2058–2067.
129. Vuillemin A, Wankel SD, Coskun ÖK, Magritsch T, Vargas S, Estes ER, et al. Archaea dominate oxic subseafloor communities over multimillion-year time scales. *Sci Adv* 2019; **5**: 4108–4119.
130. Spaans SK, Weusthuis RA, Van Der Oost J, Kengen SW. NADPH-generating systems in bacteria and archaea. *Front Microbiol* 2015; **6**: 742.
131. Blank LM, Ebert BE, Buehler K, Bühler B. Redox biocatalysis and metabolism: molecular mechanisms and metabolic network analysis. *Antioxid Redox Signaling* 2010; **13**: 349–394.
132. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.
133. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017; **35**: 725–731.
134. Zoccarato L, Sher D, Miki T, Segrè D, Grossart H-P. A comparative whole-genome approach identifies bacterial traits for marine microbial interactions. *Commun Biol* 2022; **5**: 276.

135. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 2015; **5**: 8365–8371.
136. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf* 2010; **11**: 119.
137. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinform* 2007; **23**: 673–679.
138. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinform* 2014; **30**: 2068–2069.
139. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 2020; **48**: 8883–8900.
140. Shaffer M, Borton MA, Bolduc B, Faria JP, Flynn RM, Ghadermazi P, et al. kb_DRAM: annotation and metabolic profiling of genomes with DRAM in KBase. *Bioinform* 2023; **39**: 110.
141. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017; **45**: D353–D361.
142. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, the UniProt Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinform* 2015; **31**: 926–932.
143. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the peptidase database. *Nucleic Acids Res* 2010; **38**: 227–233.
144. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017; **35**: 1026–1028.
145. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 2013; **41**: e121.
146. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32**: 1792–1797.
147. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinform* 2009; **25**: 1972–1973.
148. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007; **35**: 182–185.
149. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016; **428**: 726–731.
150. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinform* 2020; **36**: 2251–2252.
151. Suzuki S, Kakuta M, Ishida T, Akiyama Y. Faster sequence homology searches by clustering subsequences. *Bioinform* 2015; **31**: 1183–1190.
152. Saier MH, Reddy VS, Moreno-Hagelsieb G, Hendargo KJ, Zhang Y, Iddamsetty V, et al. The transporter classification database (TCDB): 2021 update. *Nucleic Acids Res* 2021; **49**: D461–D467.
153. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014; **42**: 490–495.
154. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. *Nucleic Acids Res* 2023; 115–121.
155. Zimmermann J, Kaleta C, Waschina S. gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biol* 2021; **22**: 81–117.
156. Mendler K, Chen H, Parks DH, Lobb B, Hug LA, Doxey AC. AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* 2019; **47**: 4442–4448.

157. Potter RF, Zhang K, Reimler B, Marino J, Muenks CE, Alvarado K, et al. Uncharacterized and lineage-specific accessory genes within the *Proteus mirabilis* pan-genome landscape. *mSystems* 2023; **8**: e00159-23.
158. Oksanen J, Kindt R, Legendre P, Hara B, Simpson G, Solymos P, et al. The vegan Package. 2009.
159. R Core Team. R: A language and environment for statistical computing. 2013. R Foundation for Statistical Computing, Vienna, Austria.
160. Wickham H. ggplot2: elegant graphics for data analysis. 2016. Springer-Verlag New York.
161. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinform* 2016; **32**: 2847–2849.
162. Gao C-H, Yu G, Cai P. ggVennDiagram: An intuitive, easy-to-use, and highly customizable R package to generate Venn diagram. *Front Genet* 2021; **12**: 706907.
163. Alves RJE, Minh BQ, Urich T, von Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun* 2018; **9**: 1517.
164. Rodriguez J, Chakrabarti S, Choi E, Shehadeh N, Sierra-Martinez S, Zhao J, et al. Nutrient-limited enrichments of nitrifiers from soil yield consortia of nitrosocosmicus-affiliated aoa and nitrospira-affiliated nob. *Front Microbiol* 2021; **12**: 1962.
165. Guieysse B, Wuertz S. Metabolically versatile large-genome prokaryotes. *Curr Opin Biotechnol* 2012; **23**: 467–473.
166. Grzymski JJ, Dussaq AM. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J* 2012; **6**: 71–80.
167. Dunand C, Crèvecoeur M, Penel C. Distribution of superoxide and hydrogen peroxide in Arabidopsis root and their influence on root development: possible interaction with peroxidases. *New Phytol* 2007; **174**: 332–341.
168. Becana M, Aparicio-Tejo P, Irigoyen JJ, Sanchez-Diaz M. Some enzymes of hydrogen peroxide metabolism in leaves and root nodules of *Medicago sativa*. *Plant Physiol* 1986; **82**: 1169–1171.
169. Kontchou CY, Blondeau R. Isolation and characterization of hydrogen peroxide producing *Aerococcus* sp. from soil samples. *FEMS Microbiol Lett* 1990; **68**: 323–327.
170. Veal EA, Day AM, Morgan BA. Hydrogen peroxide sensing and signaling. *Mol Cell* 2007; **26**: 1–14.
171. Park S, Imlay JA. High levels of intracellular cysteine promote oxidative DNA damage by driving the fenton reaction. *J Bacteriol* 2003; **185**: 1942–1950.
172. Valko M, Morris H, Cronin MTD. Metals, toxicity and oxidative stress. *Curr Med Chem* ; **12**: 1161–1208.
173. Tolar BB, Powers LC, Miller WL, Wallsgrove NJ, Popp BN, Hollibaugh JT. Ammonia oxidation in the ocean can be inhibited by nanomolar concentrations of hydrogen peroxide. *Front Mar Sci* 2016; **3**: 237.
174. Imlay JA. Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem* 2008; **77**: 755–776.
175. Lee U-J, Gwak J-H, Choi S, Jung M-Y, Lee TK, Ryu H, et al. *Ca*. Nitrosocosmicus members are the dominant archaea associated with pepper (*Capsicum annuum* L.) and ginseng (*Panax ginseng* C.A. Mey.) plants' rhizospheres. 2024. bioRxiv. , 2024.01.08.574571
176. Offre P, Kerou M, Spang A, Schleper C. Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol* 2014; **22**: 665–675.
177. Parada AE, Mayali X, Weber PK, Wollard J, Santoro AE, Fuhrman JA, et al. Constraining the composition and quantity of organic matter used by abundant marine Thaumarchaeota. *Environ Microbiol* 2023; **25**: 689–704.
178. Zheng X, Wang B, Tang X, Mao B, Zhang Q, Zhang T, et al. Absorption, metabolism, and functions of hyaluronic acid and its therapeutic prospects in combination with microorganisms: A review. *Carbohydr Polym* 2023; **299**: 120153.
179. Fehervari Z. Coinfection enhances inflammation. *Nat Immunol* 2016; **17**: 121–121.

180. Li LX, Rautengarten C, Heazlewood JL, Doering TL. UDP-glucuronic acid transport is required for virulence of *Cryptococcus neoformans*. *mBio* 2018; **9**: e02319-17.
181. Nikaido H. Multidrug efflux pumps of gram-negative bacteria. *J Bacteriol* 1996; **178**: 5853–5859.
182. Paulsen IT, Brown MH, Skurray RA. Proton-dependent multidrug efflux systems. *Microbiol Rev* 1996; **60**: 575–608.
183. Woolridge DP, Vazquez-Laslop N, Markham PN, Chevalier MS, Gerner EW, Neyfakh AA. Efflux of the natural polyamine spermidine facilitated by the *Bacillus subtilis* multidrug transporter Blt. *J Biol Chem* 1997; **272**: 8864–8866.
184. Lennen RM, Politz MG, Kruziki MA, Pfeleger BF. Identification of transport proteins involved in free fatty acid efflux in *Escherichia coli*. *J Bacteriol* 2013; **195**: 135–144.
185. Rosenberg EY, Bertenthal D, Nilles ML, Bertrand KP, Nikaido H. Bile salts and fatty acids induce the expression of *Escherichia coli* AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Mol Microbiol* 2003; **48**: 1609–1619.
186. Piddock LJV. Multidrug-resistance efflux pumps - not just for resistance. *Nat Rev Microbiol* 2006; **4**: 629–636.
187. Chang Y-C, Hu Z, Rachlin J, Anton BP, Kasif S, Roberts RJ, et al. COMBREX-DB: an experiment centered database of protein function: knowledge, predictions and knowledge gaps. *Nucleic Acids Res* 2016; **44**: 330–335.
188. Erbilgin O, Rübel O, Louie KB, Trinh M, Raad M de, Wildish T, et al. MAGI: A method for metabolite annotation and gene integration. *ACS Chem Biol* 2019; **14**: 704–714.
189. Schnoes AM, Brown SD, Dodevski I, Babbitt PC. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* 2009; **5**: e1000605.
190. Liu Y, Zhou Z, Pan J, Baker BJ, Gu J-D, Li M. Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota. *ISME J* 2018; **12**: 1021–1031.
191. Wang Y, San K-Y, Bennett GN. Cofactor engineering for advancing chemical biotechnology. *Curr Opin Biotechnol* 2013; **24**: 994–999.
192. Huang L, Xing X, Zhou P, Li Puma G. Mixotrophic bacteria for environmental detoxification of contaminated waste and wastewater. *Appl Microbiol Biotechnol* 2021; **105**: 6627–6648.
193. Kozłowski JA, Price J, Stein LY. Revision of N₂O-producing pathways in the ammonia-oxidizing bacterium *Nitrosomonas europaea* ATCC 19718. *Appl Environ Microbiol* 2014; **80**: 4930–4935.
194. Ebert S, Fischer P, Knackmuss HJ. Converging catabolism of 2,4,6-trinitrophenol (picric acid) and 2,4-dinitrophenol by *Nocardioides simplex* FJ2-1A. *Biodegradation* 2001; **12**: 367–376.
195. Taylor MC, Jackson CJ, Tattersall DB, French N, Peat TS, Newman J, et al. Identification and characterization of two families of F420 H₂-dependent reductases from *Mycobacteria* that catalyse aflatoxin degradation. *Mol Microbiol* 2010; **78**: 561–575.
196. Ahmed FH, Carr PD, Lee BM, Afriat-Jurnou L, Mohamed AE, Hong N-S, et al. Sequence–structure–function classification of a catalytically diverse oxidoreductase superfamily in *Mycobacteria*. *J Mol Biol* 2015; **427**: 3554–3571.
197. Greening C, Ahmed FH, Mohamed AE, Lee BM, Pandey G, Warden AC, et al. Physiology, biochemistry, and applications of F420- and Fo-dependent redox reactions. *Microbiol Mol Biol Rev* 2016; **80**: 451–493.
198. Purwantini E, Mukhopadhyay B. Conversion of NO₂ to NO by reduced coenzyme F420 protects *mycobacteria* from nitrosative damage. *Proc Natl Acad Sci USA* 2009; **106**: 6333–6338.
199. Sellés Vidal L, Kelly CL, Mordaka PM, Heap JT. Review of NAD(P)H-dependent oxidoreductases: properties, engineering and application. *BBA* 2018; **1866**: 327–347.
200. Ji X-J, Xia Z-F, Fu N-H, Nie Z-K, Shen M-Q, Tian Q-Q, et al. Cofactor engineering through heterologous expression of an NADH oxidase and its impact on metabolic flux redistribution in *Klebsiella pneumoniae*. *Biotechnol Biofuels* 2013; **6**: 7.

201. Bloem A, Sanchez I, Dequin S, Camarasa C. Metabolic impact of redox cofactor perturbations on the formation of aroma compounds in *Saccharomyces cerevisiae*. *Appl Environ Microbiol* 2016; **82**: 174–183.
202. Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *Proc Natl Acad Sci USA* 2014; **111**: 8239–8244.
203. Covarrubias AJ, Perrone R, Grozio A, Verdin E. NAD⁺ metabolism and its roles in cellular processes during ageing. *Nat Rev Mol Cell Biol* 2021; **22**: 119–141.
204. Xie N, Zhang L, Gao W, Huang C, Huber PE, Zhou X, et al. NAD⁺ metabolism: pathophysiologic mechanisms and therapeutic potential. *Sig Transduct Target Ther* 2020; **5**: 227.
205. Sharma LK, Lu J, Bai Y. Mitochondrial respiratory complex I: structure, function and implication in human diseases. *Curr Med Chem* 2009; **16**: 1266–1277.
206. Galemou Yoga E, Angerer H, Parey K, Zickermann V. Respiratory complex I – mechanistic insights and advances in structure determination. *BBA* 2020; **1861**: 148153.
207. Sarkar S, Kazarina A, Hansen PM, Ward K, Hargreaves C, Reese N, et al. Ammonia-oxidizing archaea and bacteria differentially contribute to ammonia oxidation in soil under precipitation gradients and land legacy. *bioRxiv* 2023; 2023.11.08.566028.
208. Nakashima M, Yamagami R, Tomikawa C, Ochi Y, Moriya T, Asahara H, et al. Long and branched polyamines are required for maintenance of the ribosome, tRNA^{His} and tRNA^{Tyr} in *Thermus thermophilus* cells at high temperatures. *Genes to Cells* 2017; **22**: 628–645.
209. Pereira MM, Bandejas TM, Fernandes AS, Lemos RS, Melo AM, Teixeira M. Respiratory chains from aerobic thermophilic prokaryotes. *J Bioenerg Biomembr* 2004; **36**: 93–105.
210. Patridge EV, Ferry JG. WrbA from *Escherichia coli* and *Archaeoglobus fulgidus* is an NAD(P)H:quinone oxidoreductase. *J Bacteriol* 2006; **188**: 3498–3506.
211. Simon J, Klotz MG. Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. *BBA* 2013; **1827**: 114–135.
212. Braasch-Turi MM, Koehn JT, Kostenkova K, Van Cleave C, Ives JW, Murakami HA, et al. Electron transport lipids fold within membrane-like interfaces. *Front Chem* 2022; **10**.
213. Kurosu M, Begari E. Vitamin K2 in electron transport system: are enzymes involved in vitamin K2 biosynthesis promising drug targets? *Molecules* 2010; **15**: 1531–1553.
214. Truglio JJ, Theis K, Feng Y, Gajda R, Machutta C, Tonge PJ, et al. Crystal structure of *Mycobacterium tuberculosis* MenB, a key enzyme in vitamin K2 biosynthesis. *J Biol Chem* 2003; **278**: 42352–42360.
215. Gomes CM, Bandejas TM, Teixeira M. A new type-II NADH dehydrogenase from the archaeon *Acidianus ambivalens*: characterization and in vitro reconstitution of the respiratory chain. *J Bioenerg Biomembr* 2001; **33**: 1–8.
216. Bandejas TM, Salgueiro C, Kletzin A, Gomes CM, Teixeira M. *Acidianus ambivalens* type-II NADH dehydrogenase: genetic characterisation and identification of the flavin moiety as FMN. *FEBS Lett* 2002; **531**: 273–277.
217. Abken H-J, Tietze M, Brodersen J, Bäumer S, Beifuss U, Deppenmeier U. Isolation and characterization of methanophenazine and function of phenazines in membrane-bound electron transport of *Methanosarcina mazei* Gö1. *J Bacteriol* 1998; **180**: 2027–2032.
218. Kerscher S, Dröse S, Zickermann V, Brandt U. The three families of respiratory NADH dehydrogenases. In: Schäfer G, Penefsky HS (eds). *Bioenergetics: Energy Conservation and Conversion*. 2008. Springer, Berlin, Heidelberg, pp 185–222.
219. Jaworowski A, Mayo G, Shaw DC, Campbell HD, Young IG. Characterization of the respiratory NADH dehydrogenase of *Escherichia coli* and reconstitution of NADH oxidase in *ndh* mutant membrane vesicles. *Biochem* 1981; **20**: 3621–3628.
220. Rasmusson AG, Svensson ÅS, Knoop V, Grohmann L, Brennicke A. Homologues of yeast and bacterial rotenone-insensitive NADH dehydrogenases in higher eukaryotes: two enzymes are present in potato mitochondria. *TPJ* 1999; **20**: 79–87.

221. Melo AMP, Duarte M, Videira A. Primary structure and characterisation of a 64 kDa NADH dehydrogenase from the inner membrane of *Neurospora crassa* mitochondrial. *BBA* 1999; **1412**: 282–287.
222. Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proc Natl Acad Sci U S A* 2010; **107**: 8818–8823.
223. Aylward FO, Santoro AE. Heterotrophic Thaumarchaea with small genomes are widespread in the dark ocean. *mSystems* 2020; **5**: e00415–e00420.
224. Reichmann P, Görisch H. Cytochrome c550 from *Pseudomonas aeruginosa*. *Biochem J* 1993; **289**: 173–178.
225. Bertini I, Pierattelli R. Plastocyanin. In: Kretsinger RH, Uversky VN, Permyakov EA (eds). *Encyclopedia of metalloproteins*. 2013. Springer, New York, NY, pp 1694–1698.
226. Höhner R, Pribil M, Herbstová M, Lopez LS, Kunz H-H, Li M, et al. Plastocyanin is the long-range electron carrier between photosystem II and photosystem I in plants. *Proc Natl Acad Sci USA* 2020; **117**: 15354–15362.
227. Hatzenpichler R. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol* 2012; **78**: 7501–7510.
228. Spormann AM. Fermentative Metabolism. In: Spormann AM (ed). *Principles of Microbial Metabolism and Metabolic Ecology*. 2023. Springer International Publishing, Cham, pp 215–269.
229. Nakagawa T, Stahl DA. Transcriptional response of the archaeal ammonia oxidizer *Nitrosopumilus maritimus* to low and environmentally relevant ammonia concentrations. *Appl Environ Microbiol* 2013; **79**: 6911–6916.
230. Qin W, Amin SA, Lundeen RA, Heal KR, Martens-Habbenha W, Turkarslan S, et al. Stress response of a marine ammonia-oxidizing archaeon informs physiological status of environmental populations. *ISME J* 2018; **12**: 508–519.
231. Oudova-Rivera B, Crombie AT, Murrell JC, Lehtovirta-Morley LE. Alcohols as inhibitors of ammonia oxidizing archaea and bacteria. *FEMS Microbiology Letters* 2023; **370**: fnad093.
232. Reyes C, Hodgskiss LH, Kerou M, Pribasniig T, Abby SS, Bayer B, et al. Genome wide transcriptomic analysis of the soil ammonia oxidizing archaeon *Nitrososphaera viennensis* upon exposure to copper limitation. *ISME J* 2020; **14**: 2659–2674.
233. Basen M, Schut GJ, Nguyen DM, Lipscomb GL, Benn RA, Prybol CJ, et al. Single gene insertion drives bioalcohol production by a thermophilic archaeon. *Proc Natl Acad Sci USA* 2014; **111**: 17618–17623.
234. Chen H, Yue Y, Jin W, Zhou X, Wang Q, Gao S, et al. Enrichment and characteristics of ammonia-oxidizing archaea in wastewater treatment process. *Chem Eng J* 2017; **323**: 465–472.
235. Siripong W, Wolf P, Kusumoputri TP, Downes JJ, Kocharin K, Tanapongpipat S, et al. Metabolic engineering of *Pichia pastoris* for production of isobutanol and isobutyl acetate. *Biotechnol Biofuels* 2018; **11**: 1.
236. Milne N, Wahl SA, van Maris AJA, Pronk JT, Daran JM. Excessive by-product formation: A key contributor to low isobutanol yields of engineered *Saccharomyces cerevisiae* strains. *Metab Eng Commun* 2016; **3**: 39–51.
237. Sushkova VI, Zhukovskii SV, Berezina OV, Yarotskii SV. Biosynthesis of butyric acid from cabbage stem and molasses by the strain *Clostridium butyricum* VKPM B-9619. *Russ J Bioorg Chem* 2013; **39**: 771–776.
238. Kataoka N, Vangnai AS, Pongtharangkul T, Yakushi T, Matsushita K. Butyrate production under aerobic growth conditions by engineered *Escherichia coli*. *J Biosci Bioeng* 2017; **123**: 562–568.
239. Jiang L, Wu Q, Xu Q, Zhu L, Huang H. Fermentative hydrogen production from Jerusalem artichoke by *Clostridium tyrobutyricum* expressing exo-inulinase gene. *Sci Rep* 2017; **7**: 7940.
240. Smith KS, Ferry JG. Prokaryotic carbonic anhydrases. *FEMS Microbiol Rev* 2000; **24**: 335–366.
241. Ferry JG. The γ class of carbonic anhydrases. *BBA* 2010; **1804**: 374–381.

242. Zhao J, Huang L, Chakrabarti S, Cooper J, Choi E, Ganan C, et al. Nitrogen and phosphorous acquisition strategies drive coexistence patterns among archaeal lineages in soil. *ISME J* 2023; **17**: 1839–1850.
243. Burton SAQ, Prosser JI. Autotrophic ammonia oxidation at low pH through urea hydrolysis. *Environ Microbiol* 2001; **67**: 2952–2957.
244. Lu L, Han W, Zhang J, Wu Y, Wang B, Lin X, et al. Nitrification of archaeal ammonia oxidizers in acid soils is supported by hydrolysis of urea. *ISME J* 2012; **6**: 1978–1984.
245. Bogard MJ, Donald DB, Finlay K, Leavitt PR. Distribution and regulation of urea in lakes of central North America. *Freshwater Biol* 2012; **57**: 1277–1292.
246. Kearns KD, Hunter MD. Green algal extracellular products regulate antialgal toxin production in a cyanobacterium. *Environ Microbiol* 2000; **2**: 291–297.
247. Solomon CM, Collier JL, Berg GM, Glibert PM. Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat Microb Ecol* 2010; **59**: 67–88.
248. Mobley HL, Island MD, Hausinger RP. Molecular biology of microbial ureases. *Microbiol Rev* 1995; **59**: 451–480.
249. Qin W, Wei SP, Zheng Y, Choi E, Li X, Johnston J, et al. Ammonia-oxidizing bacteria and archaea exhibit differential nitrogen source preferences. *Nat Microbiol* 2024; **9**: 524–536.
250. Li Y, Ding K, Wen X, Zhang B, Shen B, Yang Y. A novel ammonia-oxidizing archaeon from wastewater treatment plant: Its enrichment, physiological and genomic characteristics. *Sci Rep* 2016; **6**: 23747.
251. Auclair SM, Bhanu MK, Kendall DA. Signal peptidase I: Cleaving the way to mature proteins. *Protein Sci* 2012; **21**: 13–25.
252. Nguyen TTH, Myrold DD, Mueller RS. Distributions of extracellular peptidases across prokaryotic genomes reflect phylogeny and habitat. *Front Microbiol* 2019; **10**: 413.
253. Patra AK, Yu Z. Genomic insights into the distribution of peptidases and proteolytic capacity among *Prevotella* and *Paraprevotella* Species. *Microbiol Spectrum* 2022; **10**: e02185-21.
254. Allison SD, Gartner TB, Mack MC, McGuire K, Treseder K. Nitrogen alters carbon dynamics during early succession in boreal forest. *Soil Biol Biochem* 2010; **42**: 1157–1164.
255. Geisseler D, Horwath WR. Regulation of extracellular protease activity in soil in response to different sources and concentrations of nitrogen and carbon. *Soil Biol Biochem* 2008; **40**: 3040–3048.
256. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsigirgos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 2022; **40**: 1023–1025.
257. Jämtgård S, Näsholm T, Huss-Danell K. Nitrogen compounds in soil solutions of agricultural land. *Soil Biol Biochem* 2010; **42**: 2325–2330.
258. Schimel JP, Bennett J. Nitrogen mineralization: challenges of a changing paradigm. *Ecology* 2004; **85**: 591–602.
259. Yin X, Zhou G, Cai M, Zhu Q-Z, Richter-Heitmann T, Aromokeye DA, et al. Catabolic protein degradation in marine sediments confined to distinct archaea. *ISME J* 2022; **16**: 1617–1626.
260. Kerou M, Ponce-Toledo RI, Zhao R, Abby SS, Hirai M, Nomaki H, et al. Genomes of Thaumarchaeota from deep sea sediments reveal specific adaptations of three independently evolved lineages. *ISME J* 2021; **15**: 2792–2808.
261. Zhang R-Y, Wang Y-R, Liu R-L, Rhee S-K, Zhao G-P, Quan Z-X. Metagenomic characterization of a novel non-ammonia-oxidizing Thaumarchaeota from hadal sediment. *Microbiome* 2024; **12**: 7.
262. Montesinos ML, Herrero A, Flores E. Amino acid transport in taxonomically diverse cyanobacteria and identification of two genes encoding elements of a neutral amino acid permease putatively involved in recapture of leaked hydrophobic amino acids. *J Bacteriol* 1997; **179**: 853–862.
263. Zotina T, Köster O, Jüttner F. Photoheterotrophy and light-dependent uptake of organic and organic nitrogenous compounds by *Planktothrix rubescens* under low irradiance. *Freshwater Biol* 2003; **48**: 1859–1872.

264. Orsi WD, Vuillemin A, Rodriguez P, Coskun ÖK, Gomez-Saez GV, Lavik G, et al. Metabolic activity analyses demonstrate that Lokiarchaeon exhibits homoacetogenesis in sulfidic marine sediments. *Nat Microbiol* 2020; **5**: 248–255.
265. Takami H, Arai W, Takemoto K, Uchiyama I, Taniguchi T. Functional classification of uncultured *Candidatus* Caldiarchaeum subterraneum using the maple system. *PLOS ONE* 2015; **10**: e0132994.
266. Cario A, Lormières F, Xiang X, Oger P. High hydrostatic pressure increases amino acid requirements in the piezo-hyperthermophilic archaeon *Thermococcus barophilus*. *Res Microbiol* 2015; **166**: 710–716.
267. Li J, Mara P, Schubotz F, Sylvan JB, Burgaud G, Klein F, et al. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature* 2020; **579**: 250–255.
268. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, et al. Carbon and nitrogen assimilation in deep seafloor microbial cells. *Proc Natl Acad Sci USA* 2011; **108**: 18295–18300.
269. Molari M, Manini E, Dell’Anno A. Dark inorganic carbon fixation sustains the functioning of benthic deep-sea ecosystems. *Global Biogeochem Cycles* 2013; **27**: 212–221.
270. Dekas AE, Parada AE, Mayali X, Fuhrman JA, Wollard J, Weber PK, et al. Characterizing chemoautotrophy and heterotrophy in marine Archaea and Bacteria with single-cell multi-isotope nanoSIP. *Front Microbiol* 2019; **10**: 2682–2697.
271. Jones DL, Kielland K, Sinclair FL, Dahlgren RA, Newsham KK, Farrar JF, et al. Soil organic nitrogen mineralization across a global latitudinal gradient. *Global Biogeochem Cycles* 2009; **23**: GB1016.
272. Farrell M, Hill PW, Farrar J, DeLuca TH, Roberts P, Kielland K, et al. Oligopeptides represent a preferred source of organic N uptake: a global phenomenon? *Ecosystems* 2013; **16**: 133–145.
273. Chen T-H, Chen T-L, Hung L-M, Huang T-C. Circadian rhythm in amino acid uptake by *Synechococcus* RF-1 1. *Plant Physiol* 1991; **97**: 55–59.
274. Aldunate M, Henríquez-Castillo C, Ji Q, Lueders-Dumont J, Mulholland MR, Ward BB, et al. Nitrogen assimilation in picocyanobacteria inhabiting the oxygen-deficient waters of the eastern tropical North and South Pacific. *Limnol Oceanogr* 2020; **65**: 437–453.
275. Lazar CS, Baker BJ, Seitz KW, Teske AP. Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISME J* 2017; **11**: 1118–1129.
276. Mou X, Sun S, Rayapati P, Moran MA. Genes for transport and metabolism of spermidine in *Ruegeria pomeroyi* DSS-3 and other marine bacteria. *Aquat Microb Ecol* 2010; **58**: 311–321.
277. Igarashi K, Kashiwagi K. Characteristics of cellular polyamine transport in prokaryotes and eukaryotes. *J Plant Biochem Physiol* 2010; **48**: 506–512.
278. Friedman SM, Oshima T. Polyamines of sulfur-dependent Archaeobacteria and their role in protein synthesis. *J Biochem* 1989; **105**: 1030–1033.
279. Terui Y, Ohnuma M, Hiraga K, Kawashima E, Oshima T. Stabilization of nucleic acids by unusual polyamines produced by an extreme thermophile, *Thermus thermophilus*. *Biochem J* 2005; **388**: 427–433.
280. Valdés-Santiago L, Ruiz-Herrera J. Stress and polyamine metabolism in fungi. *Front Chem* 2013; **1**: 42.
281. Michael AJ. Polyamine function in archaea and bacteria. *J Biol Chem* 2018; **293**: 18693–18701.
282. Marton LJ, Morris DR. 4 - Molecular and Cellular Functions of the Polyamines. In: McCann PP, Pegg AE, Sjoerdsma A (eds). *Inhibition of Polyamine Metabolism*. 1987. Academic Press, pp 79–105.
283. Wink DA, Mitchell JB. Chemical biology of nitric oxide: insights into regulatory, cytotoxic, and cytoprotective mechanisms of nitric oxide. *Free Radical Biol Med* 1998; **25**: 434–456.

284. Ridnour LA, Thomas DD, Mancardi D, Espey MG, Miranda KM, Paolocci N, et al. The chemistry of nitrosative stress induced by nitric oxide and reactive nitrogen oxide species. Putting perspective on stressful biological situations. *Biol Chem* 2004; **385**: 1–10.
285. Damashek J, Bayer B, Herndl GJ, Wallsgrove NJ, Allen T, Popp BN, et al. Limited accessibility of nitrogen supplied as amino acids, amides, and amines as energy sources for marine Thaumarchaeota. 2021. , 2021.07.22.453390
286. Kamio Y, Nakamura K. Putrescine and cadaverine are constituents of peptidoglycan in *Veillonella alcalescens* and *Veillonella parvula*. *Journal of Bacteriology* 1987; **169**: 2881–2884.
287. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol* 2019; **4**: 234–243.
288. Esquirol L, Peat TS, Wilding M, Lucent D, French NG, Hartley CJ, et al. Structural and biochemical characterization of the biuret hydrolase (BiuH) from the cyanuric acid catabolism pathway of *Rhizobium leguminosorum* bv. *viciae* 3841. *PLoS One* 2018; **13**: e0192736.
289. Kim K-S, Pelton JG, Inwood WB, Andersen U, Kustu S, Wemmer DE. The *Rut* pathway for pyrimidine degradation: novel chemistry and toxicity problems. *J Bacteriol* 2010; **192**: 4089–4102.
290. Murray TP, Austin ER, Howard RG, Horn RC. Simultaneous determination of biuret, triuret, and methylenediurea in urea by reversed-phase liquid chromatography. *Anal Chem* 1982; **54**: 1504–1507.
291. MacMahon S, Begley TH, Diachenko GW, Stromgren SA. A liquid chromatography-tandem mass spectrometry method for the detection of economically motivated adulteration in protein-containing foods. *J Chromatogr A* 2012; **1220**: 101–107.
292. Robinson SL, Badalamenti JP, Dodge AG, Tassoulas LJ, Wackett LP. Microbial biodegradation of biuret: defining biuret hydrolases within the isochorismatase superfamily. *Environ Microbiol* 2018; **20**: 2099–2111.
293. Wang W, Paschalidis K, Feng J-C, Song J, Liu J-H. Polyamine catabolism in plants: A universal process with diverse functions. *Front Plant Sci* 2019; **10**: 561.
294. Achor DS, Albrigo LG. Biuret toxicity symptoms in citrus leaves mimics cell senescence rather than nutritional deficiency chlorosis. *JASHS* 2005; **130**: 667–673.
295. Robinson SL, Badalamenti JP, Dodge AG, Tassoulas LJ, Wackett LP. Microbial biodegradation of biuret: defining biuret hydrolases within the isochorismatase superfamily. *Environ Microbiol* 2018; **20**: 2099–2111.
296. Latypova E, Yang S, Wang Y-S, Wang T, Chavkin TA, Hackett M, et al. Genetics of the glutamate-mediated methylamine utilization pathway in the facultative methylotrophic beta-proteobacterium *Methyloversatilis universalis* FAM5. *Mol Microbiol* 2010; **75**: 426–439.
297. Neff JC, Holland EA, Dentener FJ, McDowell WH, Russell KM. The origin, composition and rates of organic nitrogen deposition: A missing piece of the nitrogen cycle? *Biogeochemistry* 2002; **57**: 99–136.
298. Schade GW, Crutzen PJ. Emission of aliphatic amines from animal husbandry and their reactions: Potential source of N₂O and HCN. *J Atmos Chem* 1995; **22**: 319–346.
299. Kiene RP, Capone DG. Stimulation of methanogenesis by aldicarb and several other N-methyl carbamate pesticides. *Appl Environ Microbiol* 1986; **51**: 1247–1251.
300. Yu PH, Dyck RF. Impairment of methylamine clearance in uremic patients and its nephropathological implications. *Clin Nephrol* 1998; **49**: 299–302.
301. Landfald B, Strøm AR. Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. *J Bacteriol* 1986; **165**: 849–855.
302. Russell R, Scopes RK. Use of hydrophobic chromatography for purification of the membrane-located choline dehydrogenase from a *Pseudomonas* strain. *Biomed J* 1994; **4**: 279–284.
303. Owen OE, Kalhan SC, Hanson RW. The key role of anaplerosis and cataplerosis for citric acid cycle function. *J Biol Chem* 2002; **277**: 30409–30412.

304. Mather MW, Gennis RB. Spectroscopic studies of pyruvate oxidase flavoprotein from *Escherichia coli* trapped in the lipid-activated form by cross-linking. *J Biol Chem* 1985; **260**: 10395–10397.
305. Schreiner ME, Eikmanns BJ. Pyruvate:quinone oxidoreductase from *Corynebacterium glutamicum*: purification and biochemical characterization. *J Bacteriol* 2005; **187**: 862–871.
306. Abdel-Hamid AM, Attwood MM, Guest JR. Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*. *Microbiology* 2001; **147**: 1483–1498.
307. Lee C, Park C. Bacterial responses to glyoxal and methylglyoxal: reactive electrophilic species. *Int J Mol Sci* 2017; **18**: 169.
308. Gori F, Tringe SG, Kartal B, Marchiori E, Jetten MSM. The metagenomic basis of anammox metabolism in *Candidatus Brocadia fulgida*. *Biochem Soc Trans* 2011; **39**: 1799–1804.
309. Hall JA, Pajor AM. Functional characterization of a Na⁽⁺⁾-coupled dicarboxylate carrier protein from *Staphylococcus aureus*. *J Bacteriol* 2005; **187**: 5189–5194.
310. Carr SA, Jungbluth SP, Eloë-Fadrosch EA, Stepanauskas R, Woyke T, Rappé MS, et al. Carboxydolithotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. *ISME J* 2019; **13**: 1457–1468.
311. Bulzu P-A, Andrei A-Ş, Salcher MM, Mehrshad M, Inoue K, Kandori H, et al. Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche. *Nat Microbiol* 2019; **4**: 1129–1137.
312. Schink SJ, Christodoulou D, Mukherjee A, Athaide E, Brunner V, Fuhrer T, et al. Glycolysis/gluconeogenesis specialization in microbes is driven by biochemical constraints of flux sensing. *Mol Syst Biol* 2022; **18**: e10704.
313. Kappelmann J, Wiechert W, Noack S. Cutting the Gordian Knot: Identifiability of anaplerotic reactions in *Corynebacterium glutamicum* by means of ¹³C-metabolic flux analysis. *Biotechnol Bioeng* 2016; **113**: 661–674.
314. Graham DE, Xu H, White RH. Identification of coenzyme M biosynthetic phosphosulfolactate synthase: a new family of sulfonate-biosynthesizing enzymes. *J Biol Chem* 2002; **277**: 13421–13429.
315. Meyer FM, Stülke J. Malate metabolism in *Bacillus subtilis*: distinct roles for three classes of malate-oxidizing enzymes. *FEMS Microbiol Lett* 2013; **339**: 17–22.
316. Bose HS, Marshall B, Debnath DK, Perry EW, Whittall RM. Electron transport chain complex II regulates steroid metabolism. *iScience* 2020; **23**: 101295.
317. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. *Nat Rev Microbiol* 2010; **8**: 447–460.
318. Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 2007; **318**: 1782–1786.
319. Estelmann S, Hügler M, Eisenreich W, Werner K, Berg IA, Ramos-Vera WH, et al. Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J Bacteriol* 2011; **193**: 1191–1200.
320. Spasov E. Ammonia-oxidizing archaea and complete ammonia-oxidizing bacteria in a municipal wastewater treatment plant. 2018. Master's thesis, University of Waterloo.
321. Kemnitz D, Kolb S, Conrad R. High abundance of Crenarchaeota in a temperate acidic forest soil. *FEMS Microbiol Ecol* 2007; **60**: 442–448.
322. Bombar D, Heller P, Sanchez-Baracaldo P, Carter BJ, Zehr JP. Comparative genomics reveals surprising divergence of two closely related strains of uncultivated UCYN-A cyanobacteria. *ISME J* 2014; **8**: 2530–2542.
323. Gruber-Dorninger C, Pester M, Kitzinger K, Savio DF, Loy A, Rattei T, et al. Functionally relevant diversity of closely related *Nitrospira* in activated sludge. *ISME J* 2015; **9**: 643–655.
324. Koch H, Lückner S, Albertsen M, Kitzinger K, Herbold C, Spieck E, et al. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus *Nitrospira*. *Proc Natl Acad Sci USA* 2015; **112**: 11371–11376.
325. Feldman MY. Reactions of nucleic acids and nucleoproteins with formaldehyde. *Prog Nucleic Acid Res Mol Biol* 1973; **13**: 1–49.

326. Conaway CC, Whysner J, Verna LK, Williams GM. Formaldehyde mechanistic data and risk assessment: endogenous protection from DNA adduct formation. *Pharmacol Ther* 1996; **71**: 29–55.
327. Chen NH, Djoko KY, Veyrier FJ, McEwan AG. Formaldehyde stress responses in bacterial pathogens. *Front Microbiol* 2016; **7**: 257.
328. Kubota K, Nagata K, Okai M, Miyazono K-I, Soemphol W, Ohtsuka J, et al. The crystal structure of l-sorbose reductase from *Gluconobacter frateurii* complexed with NADPH and l-sorbose. *J Mol Biol* 2011; **407**: 543–555.
329. Zhang Y, Li Z, Kholodkevich S, Sharov A, Feng Y, Ren N, et al. Cadmium-induced oxidative stress, histopathology, and transcriptome changes in the hepatopancreas of freshwater crayfish (*Procambarus clarkii*). *Sci Total Environ* 2019; **666**: 944–955.
330. Yelton AP, Acinas SG, Sunagawa S, Bork P, Pedrós-Alió C, Chisholm SW. Global genetic capacity for mixotrophy in marine picocyanobacteria. *ISME J* 2016; **10**: 2946–2957.
331. Zhang X, Rao Z, Zhang L, Xu M, Yang T. Efficient 9 α -hydroxy-4-androstene-3,17-dione production by engineered *Bacillus subtilis* co-expressing *Mycobacterium neoaurum* 3-ketosteroid 9 α -hydroxylase and *B. subtilis* glucose 1-dehydrogenase with NADH regeneration. *SpringerPlus* 2016; **5**: 2–8.
332. Pumirat P, Boonyuen U, Vanaporn M, Pinweha P, Tandhavanant S, Korbsrisate S, et al. The role of short-chain dehydrogenase/oxidoreductase, induced by salt stress, on host interaction of *B. pseudomallei*. *BMC Microbiol* 2014; **14**: 1.
333. Liu W, Wang P. Cofactor regeneration for sustainable enzymatic biosynthesis. *Biotechnol Adv* 2007; **25**: 369–384.
334. Tanadchangsang N, Yu J. Microbial synthesis of polyhydroxybutyrate from glycerol: gluconeogenesis, molecular weight and material properties of biopolyester. *Biotechnol Bioeng* 2012; **109**: 2808–2818.
335. Diamond S, Lavy A, Crits-Christoph A, Matheus Carnevali PB, Sharrar A, Williams KH, et al. Soils and sediments host Thermoplasmata archaea encoding novel copper membrane monooxygenases (CuMMOs). *ISME J* 2022; **16**: 1348–1362.
336. Hu W, Pan J, Wang B, Guo J, Li M, Xu M. Metagenomic insights into the metabolism and evolution of a new Thermoplasmata order (*Candidatus Gimiplasmatales*). *Environ Microbiol* 2021; **23**: 3695–3709.
337. Lamble HJ, Heyer NI, Bull SD, Hough DW, Danson MJ. Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. *J Biol Chem* 2003; **278**: 34066–34072.
338. Bhaumik SR, Sonawat HM. Kinetic mechanism of glucose dehydrogenase from *Halobacterium salinarum*. *Indian J Biochem Biophys* 1999; **36**: 143–149.
339. Wichmann R, Vasic-Racki D. Cofactor Regeneration at the Lab Scale. In: Kragl U (ed). *Technology transfer in biotechnology: from lab to industry to production*. 2005. Springer, Berlin, Heidelberg, pp 225–260.
340. Zhu Y, Chen X, Chen T, Shi S, Zhao X. Over-expression of glucose dehydrogenase improves cell growth and riboflavin production in *Bacillus subtilis*. *Biotechnol Lett* 2006; **28**: 1667–1672.
341. Phitsuwan P, Laohakunjit N, Kerdchoechuen O, Kyu KL, Ratanakhanokchai K. Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. *Folia Microbiol (Praha)* 2013; **58**: 163–176.
342. Davies G, Henrissat B. Structures and mechanisms of glycosyl hydrolases. *Structure* 1995; **3**: 853–859.
343. Oyeleye A, Normi YM. Chitinase: diversity, limitations, and trends in engineering for suitable applications. *Biosci Rep* 2018; **38**: BSR2018032300.
344. Muzzarelli RAA, Boudrant J, Meyer D, Manno N, DeMarchis M, Paoletti MG. Current views on fungal chitin/chitosan, human chitinases, food preservation, glucans, pectins and inulin: A tribute

- to Henri Braconnot, precursor of the carbohydrate polymers science, on the chitin bicentennial. *Carbohydr Polym* 2012; **87**: 995–1012.
345. Naas AE, Solden LM, Norbeck AD, Brewer H, Hagen LH, Heggenes IM, et al. *Candidatus* Paraporphyromonas polyenzymogenes encodes multi-modular cellulases linked to the type IX secretion system. *Microbiome* 2018; **6**: 44.
 346. Li N, Zhang R, Zhou J, Huang Z. Structures, biochemical characteristics, and functions of β -xylosidases. *J Agric Food Chem* 2023; **71**: 7961–7976.
 347. Brusa C, Muzard M, Rémond C, Plantier-Royon R. β -Xylopyranosides: synthesis and applications. *RSC Adv* 2015; **5**: 91026–91055.
 348. Strazzulli A, Cobucci-Ponzano B, Iacono R, Giglio R, Maurelli L, Curci N, et al. Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *The FEBS Journal* 2020; **287**: 1116–1137.
 349. Angelov A, Putyrski M, Liebl W. Molecular and biochemical characterization of α -glucosidase and α -mannosidase and their clustered genes from the thermoacidophilic Archaeon *Picrophilus torridus*. *J Bacteriol* 2006; **188**: 7123–7131.
 350. Daniel PF, Winchester B, Warren CD. Mammalian α -mannosidases—multiple forms but a common purpose? *J Glycobiol* 1994; **4**: 551–566.
 351. Nakajima M, Imamura H, Shoun H, Wakagi T. Unique metal dependency of cytosolic alpha-mannosidase from *Thermotoga maritima*, a hyperthermophilic bacterium. *Arch Biochem Biophys* 2003; **415**: 87–93.
 352. Maruyama Y, Nakajima T, Ichishima E. A 1,2- α -D-mannosidase from a *Bacillus* sp.: purification, characterization, and mode of action. *Carbohydr Res* 1994; **251**: 89–98.
 353. Wang S, Ping Q, Li Y. Comprehensively understanding metabolic pathways of protein during the anaerobic digestion of waste activated sludge. *Chemosphere* 2022; **297**: 134117.
 354. Hii SL, Tan JS, Ling TC, Ariff AB. Pullulanase: role in starch hydrolysis and potential industrial applications. *Enzyme Res* 2012; **2012**: 921362.
 355. Couturier M, Touvrey-Loiodice M, Terrapon N, Drula E, Buon L, Chirat C, et al. Functional exploration of the glycoside hydrolase family GH113. *PLOS ONE* 2022; **17**: e0267509.
 356. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res* 2022; **50**: 571–577.
 357. Chiku K, Nihira T, Suzuki E, Nishimoto M, Kitaoka M, Ohtsubo K, et al. Discovery of two β -1,2-mannoside phosphorylases showing different chain-length specificities from *Thermoanaerobacter* sp. X-514. *PLOS ONE* 2014; **9**: e114882.
 358. Kawahara R, Saburi W, Odaka R, Taguchi H, Ito S, Mori H, et al. Metabolic mechanism of mannan in a ruminal bacterium, *Ruminococcus albus*, involving two mannoside phosphorylases and cellobiose 2-epimerase: discovery of a new carbohydrate phosphorylase, β -1,4-mannooligosaccharide phosphorylase. *J Biol Chem* 2012; **287**: 42389–42399.
 359. Pérez Recalde M, Carlucci MJ, Nosedá MD, Matulewicz MC. Chemical modifications of algal mannans and xylomannans: Effects on antiviral activity. *Phytochemistry* 2012; **73**: 57–64.
 360. Wang X, Wang C, Bao L, Xie S. Impact of carbon source amendment on ammonia-oxidizing microorganisms in reservoir riparian soil. *Ann Microbiol* 2015; **65**: 1411–1418.
 361. Jana UK, Suryawanshi RK, Prajapati BP, Soni H, Kango N. Production optimization and characterization of mannoooligosaccharide generating β -mannanase from *Aspergillus oryzae*. *Bioresour Technol* 2018; **268**: 308–314.
 362. Kulys J, Tetianec L, Bratkovskaja I. Pyrroloquinoline quinone-dependent carbohydrate dehydrogenase: activity enhancement and the role of artificial electron acceptors. *Biotechnol J* 2010; **5**: 822–828.
 363. Matsushita K, Shinagawa E, Adachi O, Ameyama M. o-Type cytochrome oxidase in the membrane of aerobically grown *Pseudomonas aeruginosa*. *FEBS Lett* 1982; **139**: 255–258.

364. Lüddecke F, Wülfing A, Timke M, Germer F, Weber J, Dikfidan A, et al. Geraniol and geranial dehydrogenases induced in anaerobic monoterpene degradation by *Castellaniella defragrans*. *Appl Environ Microbiol* 2012; **78**: 2128–2136.
365. Camargo FP, Sakamoto IK, Delforno TP, Midoux C, Duarte ICS, Silva EL, et al. Microbial and functional characterization of granulated sludge from full-scale UASB thermophilic reactor applied to sugarcane vinasse treatment. *Environ Technol* 2023; **44**: 3141–3160.
366. Chen L-X, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes from metagenomes. *Genome Res* 2020; **30**: 315–333.
367. Glenn JK, Akileswaran L, Gold MH. Mn(II) oxidation is the principal function of the extracellular Mn-peroxidase from *Phanerochaete chrysosporium*. *Arch Biochem Biophys* 1986; **251**: 688–696.
368. Santelli CM, Webb SM, Dohnalkova AC, Hansel CM. Diversity of Mn oxides produced by Mn(II)-oxidizing fungi. *Geochimica et Cosmochimica Acta* 2011; **75**: 2762–2776.
369. Tang W, Liu Y, Gong J, Chen S, Zeng X. Analysis of manganese oxidase and its encoding gene in *Lysinibacillus* strain MK-1. *Process Saf Environ Prot* 2019; **127**: 299–305.
370. Therdkiattikul N, Ratpukdi T, Kidkhunthod P, Chanlek N, Siripattanakul-Ratpukdi S. Manganese-contaminated groundwater treatment by novel bacterial isolates: kinetic study and mechanism analysis using synchrotron-based techniques. *Sci Rep* 2020; **10**: 13391.
371. Hope CK, Bott TR. Laboratory modelling of manganese biofiltration using biofilms of *Leptothrix discophora*. *Water Res* 2004; **38**: 1853–1861.
372. Burger MS, Krentz CA, Mercer SS, Gagnon GA. Manganese removal and occurrence of manganese oxidizing bacteria in full-scale biofilters. *JWR* 2008; **57**: 351–359.
373. Tebo BM, Johnson HA, McCarthy JK, Templeton AS. Geomicrobiology of manganese(II) oxidation. *Trends Microbiol* 2005; **13**: 421–428.
374. Banh A, Chavez V, Doi J, Nguyen A, Hernandez S, Ha V, et al. Manganese (Mn) oxidation increases intracellular Mn in *Pseudomonas putida* GB-1. *PLOS ONE* 2013; **8**: e77835.
375. Ehrich HL, Salerno JC. Energy coupling in Mn²⁺ oxidation by a marine bacterium. *Arch Microbiol* 1990; **154**: 12–17.
376. Yu H, Leadbetter JR. Bacterial chemolithoautotrophy via manganese oxidation. *Nature* 2020; **583**: 453–458.
377. Kappler U, Bennett B, Rethmeier J, Schwarz G, Deutzmann R, McEwan AG, et al. Sulfite: Cytochrome c oxidoreductase from *Thiobacillus novellus*. Purification, characterization, and molecular biology of a heterodimeric member of the sulfite oxidase family. *J Biol Chem* 2000; **275**: 13202–13212.
378. Boughanemi S, Infossi P, Giudici-Orticoni M-T, Schoepp-Cothenet B, Guiral M. Sulfite oxidation by the quinone-reducing molybdenum sulfite dehydrogenase SoeABC from the bacterium *Aquifex aeolicus*. *BBA* 2020; **1861**: 148279.
379. Berben T, Overmars L, Sorokin DY, Muyzer G. Diversity and distribution of sulfur oxidation-related genes in *Thioalkalivibrio*, a genus of chemolithoautotrophic and haloalkaliphilic sulfur-oxidizing Bacteria. *Front Microbiol* 2019; **10**: 160.
380. Liu L, Liu M, Jiang Y, Lin W, Luo J. Physiological and genomic analysis of *Candidatus Nitrosocosmicus agrestis*, an ammonia tolerant ammonia-oxidizing archaeon from vegetable soil. *bioRxiv* 2019.
381. Yang Y, Daims H, Liu Y, Herbold CW, Pjevac P, Lin J-G, et al. Activity and metabolic versatility of complete ammonia oxidizers in full-scale wastewater treatment systems. *mBio* 2020; **11**: e03175-19.
382. Yang X, Le Brun NE, Thomson AJ, Moore GR, Chasteen ND. The iron oxidation and hydrolysis chemistry of *Escherichia coli* bacterioferritin. *Biochemistry* 2000; **39**: 4915–4923.
383. Soldano A, Yao H, Chandler JR, Rivera M. Inhibiting iron mobilization from bacterioferritin in *Pseudomonas aeruginosa* impairs biofilm formation irrespective of environmental iron availability. *ACS Infect Dis* 2020; **6**: 447–458.

384. Rivera M. Bacterioferritin: structure, dynamics, and protein–protein interactions at play in iron storage and mobilization. *Acc Chem Res* 2017; **50**: 331–340.
385. Andrews S, Norton I, Salunkhe AS, Goodluck H, Aly WSM, Mourad-Agha H, et al. Control of iron metabolism in Bacteria. In: Banci L (ed). *Metallomics and the Cell*. 2013. Springer Netherlands, Dordrecht, pp 203–239.
386. Barkay T, Kritee K, Boyd E, Geesey G. A thermophilic bacterial origin and subsequent constraints by redox, light and salinity on the evolution of the microbial mercuric reductase. *Environ Microbiol* 2010; **12**: 2904–2917.
387. Alam AKMN, Mollah MFA, Islam MA, Haq MS, Haque MM. Status of phytoplankton in newly constructed ponds and their relation to some meteorological and limnological factors. *Bangladesh J Fish Res* 1987; **10**: 75–81.
388. Austin JA, Colman SM. Lake Superior summer water temperatures are increasing more rapidly than regional air temperatures: A positive ice-albedo feedback. *Geophys Res Lett* 2007; **34**.
389. Oudova-Rivera B, Wright CL, Crombie AT, Murrell JC, Lehtovirta-Morley LE. The effect of methane and methanol on the terrestrial ammonia-oxidizing archaeon ‘*Candidatus Nitrosocosmicus franklandus* C13’. *Environ Microbiol* ; **25**: 948–961.
390. Verrone V, Gupta A, Laloo AE, Dubey RK, Hamid NAA, Swarup S. Organic matter stability and lability in terrestrial and aquatic ecosystems: A chemical and microbial perspective. *Sci Total Environ* 2024; **906**: 167757.
391. Ghoul M, Mitri S. The ecology and evolution of microbial competition. *Trends Microbiol* 2016; **24**: 833–845.
392. Muñoz-Marín MC, Gómez-Baena G, López-Lozano A, Moreno-Cabezuelo JA, Díez J, García-Fernández JM. Mixotrophy in marine picocyanobacteria: use of organic compounds by *Prochlorococcus* and *Synechococcus*. *ISME J* 2020; **14**: 1065–1073.
393. Jung M-Y, Kim J-G, Sinninghe Damsté JS, Rijpstra WIC, Madsen EL, Kim S-J, et al. A hydrophobic ammonia-oxidizing archaeon of the *Nitrosocosmicus* clade isolated from coal tar-contaminated sediment. *Environ Microbiol Rep* 2016; **8**: 983–992.
394. Arp DJ, Stein LY. Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. *Crit Rev Biochem Mol Biol* 2003; **38**: 471–495.
395. Junier P, Molina V, Dorador C, Hadas O, Kim O-S, Junier T, et al. Phylogenetic and functional marker genes to study ammonia-oxidizing microorganisms (AOM) in the environment. *Appl Microbiol Biotechnol* 2010; **85**: 425–440.
396. Wright CL, Lehtovirta-Morley LE. Nitrification and beyond: metabolic versatility of ammonia oxidising archaea. *ISME J* 2023; **17**: 1358–1368.
397. Stoecker DK. Conceptual models of mixotrophy in planktonic protists and some ecological and evolutionary implications. *Eur J Protistol* 1998; **34**: 281–290.
398. Hommes NG, Sayavedra-Soto LA, Arp DJ. Chemolithoorganotrophic growth of *Nitrosomonas europaea* on fructose. *J Bacteriol* 2003; **185**: 6809–6814.
399. Ward BA. Mixotroph ecology: More than the sum of its parts. *Proc Natl Acad Sci USA* 2019; **116**: 5846–5848.
400. Litchman E. Chapter 16 - resource competition and the ecological success of phytoplankton. In: Falkowski PG, Knoll AH (eds). *Evolution of Primary Producers in the Sea*. 2007. Academic Press, Burlington, pp 351–375.
401. Waibel A, Peter H, Sommaruga R. Importance of mixotrophic flagellates during the ice-free season in lakes located along an elevational gradient. *Aquat Sci* 2019; **81**: 45.
402. Crane KW, Grover JP. Coexistence of mixotrophs, autotrophs, and heterotrophs in planktonic microbial communities. *J Theor Biol* 2010; **262**: 517–527.
403. Biggs CR, Yeager LA, Bolser DG, Bonsell C, Dichiera AM, Hou Z, et al. Does functional redundancy affect ecological stability and resilience? A review and meta-analysis. *Ecosphere* 2020; **11**: e03184.

404. Eiler A. Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences. *Appl Environ Microbiol* 2006; **72**: 7431–7437.
405. Carreón-Rodríguez OE, Gosset G, Escalante A, Bolívar F. Glucose transport in *Escherichia coli*: from basics to transport engineering. *Microorganisms* 2023; **11**: 1588.
406. Reji L, Francis CA. Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J* 2020; **14**: 2105–2115.
407. Hanson AD, Pribat A, Waller JC, de Crécy-Lagard V. ‘Unknown’ proteins and ‘orphan’ enzymes: the missing half of the engineering parts list--and how to find it. *Biochem J* 2009; **425**: 1–11.
408. Ellens KW, Christian N, Singh C, Satagopam VP, May P, Linster CL. Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic Acids Res* 2017; **45**: 11495–11514.
409. Lawson CE, Nuijten GHL, de Graaf RM, Jacobson TB, Pabst M, Stevenson DM, et al. Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo ¹³C and ²H metabolic network mapping. *ISME J* 2021; **15**: 673–687.
410. Hwang Y, Schulze-Makuch D, Arens FL, Saenz JS, Adam PS, Sager C, et al. Leave no stone unturned: individually adapted xerotolerant Thaumarchaeota sheltered below the boulders of the Atacama Desert hyperarid core. *Microbiome* 2021; **9**: 234.
411. Berridge MV, Herst PM, Tan AS. Tetrazolium dyes as tools in cell biology: New insights into their cellular reduction. *Biotechnol Annu Rev* 2005; **11**: 127–152.
412. Vistica DT, Skehan P, Scudiero D, Monks A, Pittman A, Boyd MR. Tetrazolium-based assays for cellular viability: a critical examination of selected parameters affecting formazan production. *Cancer Res* 1991; **51**: 2515–2520.
413. Hatzinger PB, Palmer P, Smith RL, Peñarrieta CT, Yoshinari T. Applicability of tetrazolium salts for the measurement of respiratory activity and viability of groundwater bacteria. *J Microbiol Methods* 2003; **52**: 47–58.
414. Bochner BR, Siri M, Huang RH, Noble S, Lei X-H, Clemons PA, et al. Assay of the multiple energy-producing pathways of mammalian cells. *PLOS ONE* 2011; **6**: e18147.
415. Bailey JV, Flood BE, Ricci E, Delherbe N. Imaging of cellular oxidoreductase activity suggests mixotrophic metabolisms in *Thiomargarita* spp. *mBio* 2017; **8**: e01263-17.
416. Zhao R, Wang H, Yang H, Yun Y, Barton HA. Ammonia-oxidizing archaea dominate ammonia-oxidizing communities within alkaline cave sediments. *Geomicrobiol J* 2017; **34**: 511–523.
417. Wang Y, Qin W, Jiang X, Ju F, Mao Y, Zhang A, et al. Seasonal prevalence of ammonia-oxidizing archaea in a full-scale municipal wastewater treatment plant treating saline wastewater revealed by a 6-year time-series analysis. *Environ Sci Technol* 2021; **55**: 2662–2673.
418. Aoyagi T, Hanada S, Itoh H, Sato Y, Ogata A, Friedrich MW, et al. Ultra-high-sensitivity stable-isotope probing of rRNA by high-throughput sequencing of isopycnic centrifugation gradients. *Environ Microbiol Rep* 2015; **7**: 282–287.
419. Yang C, Hua Q, Shimizu K. Integration of the information from gene expression and metabolic fluxes for the analysis of the regulatory mechanisms in *Synechocystis*. *Appl Microbiol Biotechnol* 2002; **58**: 813–822.
420. Narayan KD, Sabat SC, Das SK. Mechanism of electron transport during thiosulfate oxidation in an obligately mixotrophic bacterium *Thiomonas bhubaneswarensis* strain S10 (DSM 18181T). *Appl Microbiol Biotechnol* 2017; **101**: 1239–1252.
421. Zhao J, Wang B, Jia Z. Phylogenetically distinct phylotypes modulate nitrification in a paddy soil. *Appl Environ Microbiol* 2015; **81**: 3218–3227.
422. Marcobal A, Kashyap PC, Nelson TA, Aronov PA, Donia MS, Spormann A, et al. A metabolomic view of how the human gut microbiota impacts the host metabolome using humanized and gnotobiotic mice. *ISME J* 2013; **7**: 1933–1943.
423. Wos ML, Pollard PC. Cellular nicotinamide adenine dinucleotide (NADH) as an indicator of bacterial metabolic activity dynamics in activated sludge. *Water Sci Technol* 2009; **60**: 783–791.
424. Kilstrup M, Hammer K, Ruhdal Jensen P, Martinussen J. Nucleotide metabolism and its control in lactic acid bacteria. *FEMS Microbiol Rev* 2005; **29**: 555–590.

425. Tyson GW, Banfield JF. Cultivating the uncultivated: a community genomics perspective. *Trends Microbiol* 2005; **13**: 411–415.
426. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. *ISME J* 2011; **5**: 908–917.
427. Beman JM, Popp BN, Francis CA. Molecular and biogeochemical evidence for ammonia oxidation by marine Crenarchaeota in the Gulf of California. *ISME J* 2008; **2**: 429–441.
428. Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. *Proc Natl Acad Sci USA* 2006; **103**: 12317–12322.
429. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. *Nat Commun* 2015; **6**: 8933–8939.
430. Brown MN, Briones A, Diana J, Raskin L. Ammonia-oxidizing archaea and nitrite-oxidizing *Nitrospiras* in the biofilter of a shrimp recirculating aquaculture system. *FEMS Microbiol Ecol* 2013; **83**: 17–25.
431. Sauder LA, Engel K, Stearns JC, Masella AP, Pawliszyn R, Neufeld JD. Aquarium nitrification revisited: Thaumarchaeota are the dominant ammonia oxidizers in freshwater aquarium biofilters. *PLoS One* 2011; **6**: e23281.
432. Doxey AC, Kurtz DA, Lynch MDJ, Sauder LA, Neufeld JD. Aquatic metagenomes implicate Thaumarchaeota in global cobalamin production. *ISME J* 2015; **9**: 461–471.
433. Jung M-Y, Park S-J, Min D, Kim J-S, Rijpstra WIC, Sinninghe Damsté JS, et al. Enrichment and characterization of an autotrophic ammonia-oxidizing Archaeon of mesophilic Crenarchaeal Group I.1a from an agricultural soil. *Appl Environ Microbiol* 2011; **77**: 8635–8647.
434. Qin W, Zheng Y, Zhao F, Wang Y, Urakawa H, Martens-Habbena W, et al. Alternative strategies of nutrient acquisition and energy conservation map to the biogeography of marine ammonia-oxidizing archaea. *ISME J* 2020; **14**: 2595–2609.
435. Stein LY. Insights into the physiology of ammonia-oxidizing microorganisms. *Curr Opin Chem Biol* 2019; **49**: 9–15.
436. Zou D, Chen J, Zhang C, Kao S-J, Liu H, Li M. Diversity and salinity adaptations of ammonia oxidizing archaea in three estuaries of China. *Appl Microbiol Biotechnol* 2023; **107**: 6897–6909.