

Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

Winter 3-13-2019

Symbiosis in Archaea: Functional and Phylogenetic Diversity of Marine and Terrestrial Nanoarchaeota and their Hosts

Emily Joyce St. John
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Bacteriology Commons](#), and the [Biology Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

St. John, Emily Joyce, "Symbiosis in Archaea: Functional and Phylogenetic Diversity of Marine and Terrestrial Nanoarchaeota and their Hosts" (2019). *Dissertations and Theses*. Paper 4939.
<https://doi.org/10.15760/etd.6815>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Symbiosis in *Archaea*: Functional and Phylogenetic Diversity of Marine and Terrestrial
Nanoarchaeota and their Hosts

by

Emily Joyce St. John

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biology

Thesis Committee:
Anna-Louise Reysenbach, Chair
Anne W. Thompson
Rahul Raghavan

Portland State University
2019

Abstract

The *Nanoarchaeota* are an enigmatic lineage of *Archaea* found in deep-sea hydrothermal vents and geothermal springs across the globe. These small (~100-400 nm) hyperthermophiles live ectosymbiotically with diverse hosts from the *Crenarchaeota*. Despite their broad distribution in high-temperature environments, very few *Nanoarchaeota* have been successfully isolated in co-culture with their hosts and nanoarchaeote genomes are poorly represented in public databases. However, the *Nanoarchaeota* provide unique insights into the structure and function of symbiosis in the archaeal domain. This study describes novel nanoarchaeotes from multiple geothermal habitats, using a combination of direct cultivation techniques and genomic analysis. A new nanoarchaeote from a New Zealand hot spring, *Candidatus* Nanoclepta minutus, was isolated in co-culture with its host. Like other terrestrial *Nanoarchaeota*, *Cand.* Ncl. minutus harbors genes for gluconeogenesis and archaeal flagella. *Zestosphaera tikiterensis*, the New Zealand host, was also isolated in pure culture and characterized. Phylogenetic analysis showed that both *Cand.* Ncl. minutus and *Z. tikiterensis* are new genera in the *Nanoarchaeota* and *Crenarchaeota*, respectively. Metagenome-assembled genomes (MAGs) from the *Nanoarchaeota* were also recovered from deep-sea hydrothermal vent sites. These MAGs capture a wide range of diversity in the *Nanoarchaeota*, representing three new species and two novel genera. Key nanoarchaeotal features were identified in the MAGs, including marker genes for archaeal flagella, gluconeogenesis and CRISPR-Cas regions. These studies greatly contribute to our understanding of nanoarchaeotal ecophysiology and provide key insights into the coding potential and diversity of *Nanoarchaeota* and their hosts.

Dedication

I dedicate this thesis to Brady St. John, my best friend, partner and husband, for his unending faith in me.

Acknowledgements

First of all, I would like to thank Dr. Anna-Louise Reysenbach. Thank-you for your endless support and patience, and for inspiring me to explore the beautiful and strange world of extremophiles. Without you, none of this would be possible. I would also like to thank my committee members, Dr. Anne Thompson and Dr. Rahul Raghavan, for their insight and support throughout this process. Thanks to Dr. Mircea Podar for many thoughtful email discussions and to Dr. Annie Lindgren for always being available to talk. I would also like to thank the members of the Reysenbach Lab, especially Jennifer Meneghin and Yitai Liu - working with both of you has been a privilege and a joy. Thank-you to my each one of my friends and family, especially my husband Brady – thanks for walking with me through this process and always seeing the best in me. Finally, I would like to thank Leila for her exuberance and zest for life and her phenomenal hair. Research was funded by grants to Anna-Louise Reysenbach from the National Science Foundation (OCE1235432, OCE1558795, OCE0728391 and DEB1134877) and the National Aeronautics and Space Administration (NNX16AJ66G).

Table of Contents

Abstract.....	i
Dedication.....	ii
Acknowledgements	iii
List of Tables	vii
List of Figures	viii
Chapter 1: Introduction.....	1
<i>Diversity and exploration of the archaeal domain</i>	1
<i>Global distribution of nanoarchaeotes</i>	4
<i>Diversity within nanoarchaeote-host systems</i>	5
<i>Nanoarchaeote-host interaction dynamics</i>	8
<i>Shared genomic features of cultivated nanoarchaeotes</i>	11
<i>Unique features of various nanoarchaeotes</i>	14
<i>Conclusions</i>	17
<i>Scope of thesis</i>	17
Chapter 2: A New Symbiotic Nanoarchaeote (<i>Candidatus Nanoclepta minutus</i>) and its Host (<i>Zestosphaera tikiterensis</i> gen. nov., sp. nov.) from a New Zealand Hot Spring.....	31
<i>Abstract</i>	31
<i>Introduction</i>	32
<i>Materials and methods</i>	34
<i>Sample site and culturing conditions</i>	34
<i>Host isolation and 16S rRNA gene sequencing</i>	34
<i>Physiological characterization of the host</i>	35

	v
<i>Core and intact lipid preparation and analysis</i>	36
<i>Electron microscopy</i>	36
<i>Cell sorting and metagenomic sequencing</i>	37
<i>Binning and assembly of metagenome</i>	37
<i>Accessibility of data and biological material</i>	38
<i>Functional annotation</i>	38
<i>Phylogenetic analysis</i>	39
<i>Analysis of predicted rRNA and tRNA intron sequences</i>	40
<i>Results and discussion</i>	40
<i>Establishment of a stable nanoarchaeotal-host enrichment culture</i>	40
<i>Isolation of the nanoarchaeote in co-culture and the host in pure culture</i>	42
<i>Characterization of the host, strain NZ3^T</i>	46
<i>The genome of strain Ncl-1 more closely resembles its terrestrial relatives than marine Nanoarchaeum equitans</i>	51
<i>Metabolic insights from the draft genome of the host, NZ3^T</i>	59
<i>Proposal of a novel Candidatus taxon in the Nanoarchaeota and a novel genus and species in the Desulfurococcaceae</i>	65
<i>Conclusions</i>	66
<i>Acknowledgements</i>	68
<i>Contributions</i>	69
Chapter 3: Deep-sea Hydrothermal Vent Metagenome-Assembled Genomes Provide	
<i>Insight into the Phylum Nanoarchaeota</i>	89
<i>Summary</i>	89

	vi
<i>Introduction</i>	90
<i>Materials and methods</i>	91
<i>Sample collection and processing</i>	91
<i>Metagenomic sequencing, assembly and binning</i>	91
<i>Genomic analysis</i>	92
<i>Phylogenetic analysis</i>	93
<i>Data availability</i>	94
<i>Results & Discussion</i>	94
<i>Phylogenetic analysis of nanoarchaeotal MAGs</i>	94
<i>Basic genomic features</i>	98
<i>Comparative analysis of nanoarchaeotal coding potential</i>	100
<i>ATP generation in the Nanoarchaeota</i>	103
<i>Ecophysiology of deep-sea Nanoarchaeota and their hosts</i>	106
<i>Acknowledgements</i>	107
<i>Contributions</i>	107
Chapter 4: Conclusion	117
Appendix A. Supplementary methods.....	127
Appendix B. List of supplementary files.....	129

List of Tables

Table 1.1	47
Characteristics distinguishing strain NZ3 ^T from other representative species in the <i>Desulfurococcaceae</i> .	
Table 1.2	49
NZ3 ^T IPLs identification.	
Table 1.3	50
Features of the nanoarchaeote strain Ncl-1 and the host strain NZ3 ^T and their closest cultivated relatives, based on our re-analysis of reference genomes.	
Table 1.4	60
Comparison of genomic features in strain Ncl-1 and other <i>Nanoarchaeota</i> .	
Table 2.1	95
Geological features and chemical composition of end-member fluids from deep-sea hydrothermal vent sites.	

List of Figures

Figure 1.1	43
Negatively stained transmission electron micrographs (TEM) of the New Zealand enrichment culture.	
Figure 1.2	44
Maximum likelihood phylogenetic tree of <i>Nanoarchaeota</i> based on 16S rRNA gene sequences.	
Figure 1.3	45
Predicted secondary structure and placement of intervening sequences in the 16S rRNA gene of strain NZ3 ^T .	
Figure 1.4	45
16S rRNA gene maximum likelihood phylogenetic tree of select <i>Desulfurococcaceae</i> .	
Figure 1.5	48
MS-spectra of intact polar lipid peaks identified in Table 1.2.	
Figure 1.6	49
HPLC-MS traces of the core lipids of NZ3 ^T .	
Figure 1.7	50
LC-MS analysis of the NZ3 ^T intact polar lipid extract.	
Figure 1.8	52
Phylogenetic analysis of 16 concatenated ribosomal proteins (rpL2, 3, 4, 5, 6, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) using RAxML.	
Figure 1.9	53
Circos-based genome alignments comparing overall protein homology between strain Ncl-1 and (A) <i>Cand. Nps. acidilobi</i> , (B) <i>N. equitans</i> and (C) the NZ13 draft genome, and between (D) strain NZ3 ^T and <i>Ignisphaera aggregans</i> .	
Figure 1.10	56
Comparison of the archaeal flagellar gene arrangement in (A) strain Ncl-1, (B) <i>Cand. Nps. acidilobi</i> and (C) <i>Pyrococcus furiosus</i> .	
Figure 1.11	59
Gene arrangement of the CRISPR-Cas locus in (A) strain Ncl-1 and (B) <i>N. equitans</i> .	
Figure 1.12	63
Maximum likelihood phylogenetic tree of archaeal <i>cooS</i> genes.	

Figure 1.13	64
Maximum likelihood phylogenetic analysis of coenzyme F ₄₂₀ -reducing hydrogenases.	
Figure 2.1	96
Maximum-likelihood phylogenetic tree based on eight concatenated ribosomal proteins.	
Figure 2.2	97
Average amino acid identity analysis of novel deep-sea <i>Nanoarchaeota</i> MAGs and their cultivated relatives.	
Figure 2.3	98
Maximum-likelihood phylogenetic analysis of nanoarchaeotal 16S rRNA genes (710 nucleotide positions).	
Figure 2.4	101
Pan-genome analysis of cultivated <i>Nanoarchaeota</i> and deep-sea nanoarchaeotal MAGs produced in Anvi'o [14,15].	
Figure 2.5	104
Reconstruction of the potential metabolic pathway for ATP generation in MC-1.	
Figure 3.1	120
Current understanding of metabolism in the <i>Nanoarchaeota</i> and nanoarchaeote-host interactions.	

Chapter 1: Introduction

Diversity and exploration of the archaeal domain

Karl Woese's 1977 discovery of the archaeal domain dramatically shifted and expanded our view of the universal tree of life. Prior to this discovery, the *Archaea* or "Archaeobacteria" were thought to represent divergent members of the *Bacteria* [66,77]. However, Woese's 16S ribosomal RNA (rRNA) gene trees revealed that the *Archaea* were very unique on the molecular level, often more similar to their eukaryotic relatives than to the *Bacteria* [73,74]. Because of the small number of cultivated archaeal lineages, the *Archaea* were thought to be exclusively extremophilic or methanogenic, limited to a few specific niches across the planet. The archaeal domain was initially divided into two main kingdoms, which are now called phyla: the thermophilic *Crenarchaeota* and the *Euryarchaeota*, comprised of methanogens, halophiles, thermophiles and sulfate reducers [74].

Exploration into the third domain of life quickly began to reveal the true extent of archaeal diversity, both on the phylogenetic and physiological level. A third archaeal phylum, the *Korarchaeota*, was discovered in the early 1990's based on 16S rRNA gene sequences. Like the *Crenarchaeota*, the *Korarchaeota* are a thermophilic lineage, and were detected in a geothermal spring in Yellowstone National Park (YNP) [5]. However, korarchaeotal 16S rRNA genes sequences branched deep within the archaeal tree of life, prior to the split between *Euryarchaeota* and *Crenarchaeota* [4,5]. The *Korarchaeota* represented the first phylum-level division in the *Archaea* to be proposed without a single cultivated representative, setting a precedent for many later-described lineages. Around the same time period, archaeal 16S rRNA gene sequences from marine environments

were also amplified and sequenced [16,22] and a cold-loving archaeal lineage was found associated with a marine sponge [54], revealing that the *Archaea* were not solely limited to methanogenic and extremophilic lifestyles. Hints regarding the physiology of these moderate archaeal lineages came from a fosmid clone obtained from a soil sample, which contained both archaeal rRNA and genes typically involved in ammonia oxidation [69]. Shortly thereafter, ammonia oxidation was demonstrated in a cultivated isolate [37]. Although initially identified as cold *Crenarchaeota*, these temperate taxa were found to be widely distributed in marine and terrestrial environments [62] and were later proposed to represent a novel mesophilic phylum, the *Thaumarchaeota* [6]. The unique ability of some *Thaumarchaeota* to oxidize ammonia suggested that *Archaea* may play major roles in global biogeochemical cycling that were previously overlooked.

Yet another archaeal phylum, the *Nanoarchaeota*, was proposed in the early 2000's, once again expanding our understanding of archaeal diversity and physiology. These ultra-small *Archaea* were first identified in anaerobic enrichment cultures, where they were seen attached to the outer surface of hyperthermophilic *Crenarchaeota* [33]. Sequencing revealed that *Nanoarchaeum equitans*, the first identified nanoarchaeote, harbored a highly unique 16S rRNA gene. This gene could not be amplified using traditional archaeal 16S rRNA gene primers [33], explaining why previous diversity studies failed to detect the lineage. Direct cultivation studies showed that *N. equitans* required direct cell-to-cell contact with its crenarchaeotal host for survival, although the host did not need any nanoarchaeotes [33]. Whole genome sequencing also revealed that *N. equitans* contained a small, reduced genome with few biosynthetic genes, consistent with its host-dependent lifestyle [71]. The *Nanoarchaeota*-host system became the first

described interspecies symbiosis between two members of the *Archaea*, once again demonstrating the diversity and functional flexibility of the archaeal domain.

Since the identification of these first archaeal phyla, metagenomic and single-cell sequencing techniques have revolutionized the study of microbiology, leading to the proposal of many new archaeal phyla and higher-order divisions [3,36,60,61,63,76]. Today, the *Archaea* are understood to be ubiquitous, thriving in both extreme and temperate environments across the globe. Key lineages of abundant, widely distributed *Archaea* have been discovered in subsurface and aquatic environments, contributing significantly to biogeochemical cycling [3,41,60], while other lineages of divergent *Archaea* have provided insights into the relationship between the archaeal and eukaryotic domains [63,76]. As genetic material has continued to increase, the archaeal tree of life has also undergone significant restructuring, with several higher-order groupings proposed as a means to categorize major archaeal phyla [27,61,76].

In the time since the discovery of the *Nanoarchaeota*, several other archaeal lineages with reduced genomes have been identified, largely by metagenomics and single-cell genome sequencing [2,10,46,61]. These taxa have been grouped with the *Nanoarchaeota* to form the DPANN superphylum, based on their tendency to cluster together in phylogenetic analyses [61]. The DPANN superphylum was originally named for five key lineages (the *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanohaloarchaeota*, and *Nanoarchaeota*) but now includes several additional phylum-level designations [10,61]. Although they harbor relatively small genomes and often have small cell sizes, members of the DPANN are highly diverse, and DPANN sequences have been recovered from a wide range of environments, including hypersaline habitats,

groundwater, alpine lakes and acid mine drainage [2,10,46,48]. Although some *Archaea* in the DPANN superphylum are thought to be free-living [46], no organism from the group has been successfully isolated in axenic culture. Genomic evidence also suggests that particular DPANN lineages may rely on a host or partner for survival, like the *Nanoarchaeota* [10]. *In situ* imaging and enrichment cultivation have revealed intimate associations between two different members of the DPANN and *Euryarchaeota* [2,15,26], suggesting that interspecies dependence may be a hallmark of some taxa in this group.

Despite the recovery of many DPANN genome sequences, the placement of the entire superphylum in the archaeal tree of life, and the *Nanoarchaeota* in particular, remains unclear [7,10,51]. It also continues to be debated whether or not the DPANN superphylum truly represents a monophyletic group [10,51,61]. Nonetheless, the discovery of diverse DPANN taxa suggests that the host-dependent lifestyle of the *Nanoarchaeota* may be widely distributed across multiple archaeal lineages. As the first characterized symbiosis between two *Archaea*, the *Nanoarchaeota* and their hosts greatly expand our understanding of the nature and evolution of interspecies interaction in the third domain of life, and they may provide valuable insights into the adaptations that allow diverse, small archaeal lineages to thrive across the globe.

Global distribution of nanoarchaeotes

The first cultivated nanoarchaeote, *N. equitans*, was discovered in enrichment cultures from Kolbeinsey Ridge, a shallow marine hydrothermal vent site near the coast of Iceland [33]. However, with the development of nanoarchaeotal-specific primers,

sequences related to *N. equitans* were soon identified in geothermal sites around the world, including hot springs in YNP, New Zealand, Chile, Kamchatka and central Asia [9,12,13,30]. 16S rRNA gene diversity studies also identified nanoarchaeotal-like sequences in lower temperature environments, including hypersaline ponds and the photic zone of Yellowstone Lake [9,13]. While it is possible that *Nanoarchaeota* survive in lower temperature environments, the sequences identified in the photic zone of Yellowstone Lake may also represent cells that have drifted away from geothermal vent sites on the lake floor [35]. In either case, however, it appears that *Nanoarchaeota* predominantly inhabit high-temperature habitats. Nanoarchaeotal genomic signatures have also been identified at deep-sea hydrothermal vent sites along the Eastern Lau Spreading Center, Mid-Atlantic Ridge, Guaymas Basin and East Pacific Rise [20,21,30,39]. A 16S rRNA gene sequence study at East Pacific Rise demonstrated that *Nanoarchaeota* and their hosts may be among the first organisms to colonize young deep-sea hydrothermal vent deposits [39], suggesting that *Nanoarchaeota*-host pairings may play a key role in the ecology of high-temperature hydrothermal systems. This study also recovered nanoarchaeotal sequences both in the presence and absence of *Ignicoccus*-like sequences, providing the first hint that nanoarchaeotes may associate with diverse hosts [39].

Diversity within nanoarchaeote-host systems

Despite their broad distribution across geothermal environments, only three *Nanoarchaeota*-host systems have been successfully cultivated in the laboratory setting, one of which is described in Chapter 2. A fourth nanoarchaeote-host system has also

been extensively described by single-cell genome sequencing. Cultivated nanoarchaeotes vary in size, but they have consistently small cell diameters (~100-400 nm), and each nanoarchaeote relies on a distinct host from the *Crenarchaeota* [33,64,75]. The three cultivated nanoarchaeote-host pairings all appear to be specific, although the crenarchaeotal hosts are highly diverse.

N. equitans, the first described nanoarchaeote, associates with *Ignicoccus hospitalis*, a hyperthermophilic *Crenarchaeota* with a highly streamlined genome [52]. *I. hospitalis* fixes carbon using the efficient dicarboxylate/4-hydroxybutyrate cycle [32,52] and obtains energy via sulfur reduction [49]. Members of the *Ignicoccus* are relatively unique among the *Archaea* due to the extensive double membrane surrounding their cells [31,49]. *I. hospitalis* has a large periplasmic space between its inner and outer membrane, which often contains vesicles and/or protrusions [29,49]. Energy generation in *I. hospitalis* takes place across the energized outer membrane, which houses an ATP synthase complex and enzymatic machinery needed to generate a proton gradient [38]. This extensive endomembrane system likely has played a key role in the evolution of the *N. equitans*-*I. hospitalis* interaction.

To explore the diversity of nanoarchaeotes in high-temperature terrestrial settings, a second and third cultivated nanoarchaeote-host system were isolated from geothermal springs in YNP and New Zealand. The YNP nanoarchaeote, *Candidatus Nanopusillus acidilobi*, was obtained from Cistern Spring, an acidic geothermal pool [75]. *Cand. Nps. acidilobi* associates with the acidophilic lineage *Acidilobus* sp. 7A [75]. Unlike autotrophic *I. hospitalis*, *Acidilobus* sp. 7A is a heterotroph that ferments protein- and carbohydrate-rich substrates [75]. Chapter 2 describes the cultivation of the New

Zealand nanoarchaeote, *Candidatus Nanoclepta minutus*, from a circum-neutral hot spring in New Zealand [64]. *Cand. Ncl. minutus* lives on the surface of *Zestosphaera tikiterensis*, the first representative of a novel genus in the Crenarchaeota [64]. Direct characterization in Chapter 2 showed that *Z. tikiterensis* is also a heterotroph that degrades proteinaceous substrates and requires thiosulfate for growth [64]. In contrast to *I. hospitalis*, neither of the cultivated terrestrial hosts have a double membrane system. Instead, both *Acidilobus* sp. 7A and *Z. tikiterensis* encode for an S-layer, a protein-based lattice that envelops the cell [64,75].

Several additional nanoarchaeote-host pairings have also been identified using single-cell genome sequences from hot springs in YNP. Because of the close relationship between the *Nanoarchaeota* and their hosts, single-cell genome sequencing techniques often recover both nanoarchaeotal and host DNA, providing a method of assessing host diversity. Putative nanoarchaeote hosts from YNP span several genera and include lineages related to the *Sulfolobales*, *Metallosphaera*, *Caldivirga*, *Thermocladium* and *Vulcanisaeta* [35,44,53]. Despite this diverse array of hosts, however, single-cell amplified nanoarchaeotes show high 16S rRNA gene sequence similarity (>95%) to the cultivated lineage *Cand. Nps. acidilobi* [35], suggesting that YNP *Nanoarchaeota* may undergo frequent host-switching events. It is currently unclear what may be driving this rapid host diversification. However, both nanoarchaeote and host cell envelope structure may play a crucial role. An analysis of synonymous versus non-synonymous mutations in single-cell amplified YNP nanoarchaeotal genomes suggested that hypothetical cell-surface proteins in the *Nanoarchaeota* undergo rapid evolution [35], which may provide the novel mutations needed to facilitate new host associations. Additionally, several of

the putative host lineages likely have S-layers [53,70], a similarity that may allow for more frequent host-switching events.

Although no *Nanoarchaeota* have been cultivated from deep submarine hydrothermal vent environments, three nanoarchaeotal metagenome-assembled genomes (MAGs) have been obtained from deep-sea hydrothermal vents sites across the world [65], as described in Chapter 3. However, without cultivation or single-cell sequencing approaches, it is nearly impossible to determine what host organism(s) associate with these *Nanoarchaeota*. No *Ignicoccus*-like bins were recovered from any of these deep-sea metagenomes, but one data set contained a bin closely related to the New Zealand host lineage *Zestosphaera* [65]. This suggests that marine nanoarchaeotes may associate with multiple hosts, not just *I. hospitalis*, and some of these host organisms may resemble their terrestrial counterparts on the phylogenetic level.

Nanoarchaeote-host interaction dynamics

Although all three cultivated nanoarchaeotes have an ectosymbiotic lifestyle, most of our understanding of *Nanoarchaeota*/host interactions comes from the *N. equitans*-*I. hospitalis* system, which has been studied for over 15 years and can be cultivated in large volumes using an anaerobic fermenter [33]. Like all cultivated nanoarchaeotes, *N. equitans* largely relies on its host for the production of metabolic building blocks [64,71,75]. Labelling studies have demonstrated that *I. hospitalis*-synthesized lipids and amino acids are transferred to *N. equitans* [34,49]. Likely, this exchange takes place via a pore-like structure that connects *I. hospitalis* and *N. equitans*, where the cytoplasm of host and nanoarchaeote appear to meet [29]. While studies of the *N. equitans*-

I. hospitalis system are much more extensive, electron microscopy has also revealed a similar pore-type formation between terrestrial nanoarchaeote *Cand. Ncl. minutus* and its host, shown in Chapter 2 [64], and membrane stretching at the attachment site of *Cand. Nps. acidilobi* and *Acidilobus* sp. 7A [75]. Due to the differences in host cell envelope, the mechanism of association in these terrestrial nanoarchaeote-host systems may be very different compared to that of *N. equitans* and *I. hospitalis* [53]. Nevertheless, imaging studies suggest an equally intimate attachment between all cultivated *Nanoarchaeota*, which may involve cytoplasmic bridging.

In the laboratory setting, different nanoarchaeotes appear to have variable effects on their respective hosts. Although *N. equitans* does not significantly affect the growth rate or final population of *I. hospitalis* cells, the growth curves of the two organisms are not synchronized during cultivation [25,49]. *N. equitans* cells continue dividing exponentially after *I. hospitalis* enters stationary phase, and nanoarchaeotes may eventually outnumber *I. hospitalis* cells by a factor of ten [25,49]. Also, infection with nanoarchaeotes can hamper or even halt the ability of *I. hospitalis* to divide, indicating a high cost for individual host cells [49]. In contrast, the growth of the YNP nanoarchaeote *Cand. Nps. acidilobi* is synchronized closely with that of its host, *Acidilobus* sp. 7A, with *Acidilobus* cells consistently outnumbering nanoarchaeotes [75]. Growth profiles of pure culture and co-cultured *Acidilobus* sp. 7A also show no appreciable difference [75]. Overall, this may suggest that *Cand. Nps. acidilobi* is less damaging to host growth and survival than *N. equitans*. In the *Cand. Ncl. minutus*-*Z. tikerensis* system, discussed in Chapter 2, direct counts of host cells and qPCR-derived counts of nanoarchaeotes also appear to be synchronous, pointing to a system more similar to the YNP

nanoarchaeote/host symbiosis [64]. Notably, Chapter 2 also reveals that *Cand. Ncl. minutus* is able to grow to much higher densities in complex enrichment cultures compared to co-culture with *Z. tikiaterensis* [64]. Although *Z. tikiaterensis* functions as the obligate and primary host for *Cand. Ncl. minutus*, it is possible that other lineages provide nutrients or small molecules that encourage nanoarchaeote growth. To date, all cultivated nanoarchaeote-host pairings appear to be specific, and most cross-infection studies have been largely unsuccessful [49,75]. Although rare host-switching events have been observed in the laboratory, none of these novel *Nanoarchaeota*-pairings have been maintained (Mircea Podar, personal communication). However, this does not rule out the possibility that nanoarchaeotes engage in a variety of interactions with diverse organisms in their natural geothermal habitats, all while maintaining their obligate host associations.

Proteomic data comparing pure *I. hospitalis* cultures and *N. equitans*-*I. hospitalis* co-cultures has also revealed some of the effects of *Nanoarchaeota* association on the host. *I. hospitalis* cells show differential changes in transcriptional regulator expression when cultivated with *N. equitans*, suggesting that nanoarchaeote association causes highly controlled, specific changes within the host [24,25]. Over time, co-cultivation with *N. equitans* also causes increases in expression of host carbon fixation and energy generation genes, likely in response to the increased metabolic demands from the nanoarchaeotes [25]. Stress proteins are shown to be highly expressed in the *I. hospitalis* proteome, although the organism does not appear to mount a significant defensive response to nanoarchaeotal infection [24,25]. High levels of stress proteins were also detected in the proteome of YNP host *Acidilobus* sp. 7A when co-cultivated with its

Nanoarchaeota [75]. However, because no proteomic data from pure *Acidilobus* sp. 7A cultures is available, it is unclear whether or not these proteins are expressed in reaction to nanoarchaeotal infection, or if they represent a constitutive response to environmental stressors. In the future, comparative proteomic studies such as those performed on the *N. equitans*-*I. hospitalis* system will be pivotal in determining whether or not terrestrial hosts react to their respective nanoarchaeotes in a similar manner.

Shared genomic features of cultivated nanoarchaeotes

Like many obligate symbionts and parasites, cultivated *Nanoarchaeota* have reduced genomes, ranging in size from 0.491 (*N. equitans*) to 0.606 Mbp (*Cand. Nps. acidilobi*) [71,75]. Although described *Nanoarchaeota* have a fairly extensive set of genes used in cell replication, RNA transcription and protein translation, they have lost most biosynthetic capabilities [64,71,75]. Without these core functions, *Nanoarchaeota* rely on host-produced cofactors, nucleotides, lipids and amino acids [71]. As shown in Chapter 2, each of the three genomes also encode split protein-coding genes, which are in multiple open reading frames spread across the genome [64,71,75]. At least some of these split genes likely maintain their functionality, as demonstrated for the split alanyl-tRNA synthase in *N. equitans* [71]. Although the three cultivated *Nanoarchaeota* share a common set of split protein-coding genes, several unique split genes are also scattered across the three genomes. Some of these split genes have also been recovered in single-cell nanoarchaeotal genomes from YNP [44,53] and *Nanoarchaeota* MAGs from deep-sea hydrothermal vents [65], as described in Chapter 3. These patterns of distribution suggest that the last common ancestor of described nanoarchaeotal lineages already had

split protein-coding genes and that genomic fragmentation has continued to occur after these taxa diverged [53,64]. Because of the fast evolving nature of the reduced and fragmented *Nanoarchaeota* genomes, the positioning of the nanoarchaeotal lineage and the DPANN group in the archaeal tree remains unclear [8,17,61,71,72]. However, split protein-coding genes in the *Nanoarchaeota* are typically seen as a derived feature, resulting from genome degradation due to host association [53].

In addition to their lack of biosynthetic genes, cultivated *Nanoarchaeota* do not have any apparent means of ATP production. Although *N. equitans* encodes a partial ATP synthase, the complex is quite divergent and cannot undergo conformational changes, which are necessary for enzyme functionality [42]. Additionally, as discussed in Chapter 2, no ATP synthase complex genes could be identified in the genomes of *Cand. Nps. acidilobi* or *Cand. Ncl. minutus* [64,75], suggesting this method of ATP generation is not functional across nanoarchaeotal lineages. The glycolysis pathway, which can produce minimal ATP through substrate-level phosphorylation, is also not complete in any of the three cultivated nanoarchaeotal genomes [64,71,75]. One deep-sea *Nanoarchaeota* MAG reported in Chapter 3 contains genes for the pentose biphosphate pathway, which may provide a way for the nanoarchaeote to produce minimal ATP after circumventing the early steps of glycolysis [65]. However, this pathway is thus far restricted to one nanoarchaeotal MAG, and it is unknown whether or not these enzymes are functional *in vivo*. Without a widely-distributed means of synthesizing ATP, most *Nanoarchaeota* likely rely on their hosts for ATP production. Co-cultures of *N. equitans* and *I. hospitalis* have been shown to contain less ATP than pure *I. hospitalis* cultures, supporting this hypothesis and suggesting that nanoarchaeotes

are utilizing host-produced ATP to fuel their own metabolic processes [28]. Complete and partial dependence on host ATP has been identified previously in obligate intracellular symbionts and parasites, including pathogenic *Chlamydia* and the insect symbiont *Cardinium* [14,50]. Both of these intracellular organisms contain ATP transporters to import host-produced ATP across the membrane [50,67], which have not been identified in the *Nanoarchaeota*. However, nanoarchaeotes may siphon ATP directly away from their hosts through the pore-like bridge connecting the two organisms.

Despite their small size and highly reduced gene content, the three cultivated *Nanoarchaeota* have efficient genomes with high coding density. Compared to obligate intracellular *Bacteria*, the nanoarchaeotes have maintained a higher level of genomic integrity and avoided extensive genome collapse [71]. As shown in Chapter 3, the three cultivated *Nanoarchaeota* have also retained their DNA repair mechanisms [65,71], which are often lost in endosymbiotic lineages [43], allowing them to avoid a buildup of deleterious mutations. Also, *Nanoarchaeota* have the added advantage of exposure to the outside environment, which provides them with avenues for horizontal gene transfer (HGT) from viruses and neighboring cells [44,47]. Based on genomic analysis of the *N. equitans*-*I. hospitalis* system, HGT also appears to occur between the host and nanoarchaeotal symbiont [71], which may help *Nanoarchaeota* compensate for gene loss events. The differences between nanoarchaeotal genomes and those of obligate endosymbiotic *Bacteria* may point to the different selection pressures affecting genome size and structure in these diverse environments [47]. In the case of endosymbionts, isolation within the host typically leads to massive gene loss and degradation, with few mechanisms for recovering lost functions [33,35]. However, genome reduction in the

Nanoarchaeota may be a product of both host dependence and the high-temperature geothermal habitat, which tends to select for smaller cell volumes, streamlined genomes and shorter gene variants, all while avoiding the extreme isolation of the intercellular environment [47].

Unique features of various nanoarchaeotes

Although all cultivated *Nanoarchaeota* share a similar ectosymbiotic lifestyle, their genomes harbor some key differences. *N. equitans* has the smallest genome of any cultivated *Nanoarchaeota* and shows more extensive gene loss and reduction compared to its terrestrial relatives, which is especially evident in its RNA systems [71]. *N. equitans* is lacking the RNase P complex, previously thought to be ubiquitous across all life [19]. RNase P performs a crucial step in the generation of mature RNA molecules by trimming the 5' leader sequence from precursor tRNAs [19]. However, *N. equitans* can function without RNase P due to its shifted tRNA promoter regions, which allow precursor tRNAs to be transcribed without a leader sequence [58]. Both cultivated terrestrial *Nanoarchaeota* appear to retain the enzyme (Chapter 2), and RNase P protein subunits have been identified in multiple nanoarchaeotal MAGs from marine hydrothermal vents (Chapter 3) [64,65,75], suggesting that RNase P is typically maintained across nanoarchaeotal lineages. The loss of RNase P in *N. equitans* likely reflects the high level of genomic reduction found in this particular *Nanoarchaeota* [53,58]. *Trans*-encoded tRNA genes in the *N. equitans* genome also support a scenario of extensive genome re-arrangement [53]. These split tRNA genes are encoded in multiple pieces, which are fused into functional tRNAs after transcription using a

nanoarchaeotal splicing endonuclease [55–57]. Since their discovery in *N. equitans*, *trans*-spliced tRNA genes have also been identified in several lineages of free-living *Crenarchaeota* [11,23]. Potentially, these split tRNA genes represent a defense mechanism used to avoid integration of viral DNA, which targets conserved tRNA gene sequences, or they may be the product of previous insertion events [57,59]. Alternately, it has also been proposed that these *trans*-encoded tRNA genes represent an ancestral gene form, which carries implications for the evolutionary history of extant tRNA genes [18]. To date, no split tRNA genes have been identified in any nanoarchaeotal lineage outside of *N. equitans* [44,53,64,75].

Genomic evidence suggests that diverse *Nanoarchaeota* interact with the abundant viral populations found in geothermal environments. CRISPR-Cas cassettes have been identified in *N. equitans* [55,71] and *Cand. Ncl. minutus* [64], and *cas* genes have been discovered in deep-sea nanoarchaeotal MAGs [65], as discussed in Chapters 2 and 3, respectively. Integrated prophage DNA has also been found in single-cell amplified nanoarchaeotal genomes from YNP [44,45]. However, no CRISPR-Cas regions were detected in YNP *Nanoarchaeota* genomes [44,53,75], suggesting that defense mechanisms against viral infection may not be distributed across all nanoarchaeotal lineages. While viruses may pose a threat to nanoarchaeotal cells, they may also provide a reservoir of genetic material which could serve to replace lost gene functions through HGT [67].

Compared to *N. equitans*, both terrestrial nanoarchaeotes have somewhat expanded genomes, and both contain genes that are lacking their marine relative, which is shown in Chapter 2. Most notably, *Cand. Nps. acidilobi* and *Cand. Ncl. minutus* both

encode genes for the archaeal flagellum, which are expressed in the proteome of *Cand. Nps. acidilobi* [75] and are visible by electron microscopy in *Cand. Ncl. minutus* [64]. Unlike most *Archaea*, however, the archaeal flagellar genes are not encoded in an operon [1], once again pointing to patterns of genome reshuffling in the *Nanoarchaeota*.

Currently, the function of the archaeal flagella is unclear, especially given the apparent inability of cultivated nanoarchaeotes to produce ATP needed to fuel motility. However, as discussed further in Chapter 2, it is possible that the *Nanoarchaeota* may use their archaeal flagella only when attached to their hosts, fueling motility with host-produced ATP [64]. Alternately, the archaeal flagella may have a function in the association between *Nanoarchaeota* and their hosts [53,64,75]. Both terrestrial nanoarchaeotes also encode genes for the gluconeogenic pathway, which are highly expressed in the *Cand. Nps. acidilobi* proteome [75]. Likely, these enzymes aid in the production of sugars for cell-surface glycosylation [64,75]. This process may play a key role in the *Nanoarchaeota*-host association [75], potentially facilitating the wide diversity of nanoarchaeote-host pairings identified in YNP hot springs [35]. Although archaeal flagella and gluconeogenic genes were not detected in the genome of *N. equitans* [71], their distribution does not appear to be governed by habitat type alone. In Chapter 3, marker genes for gluconeogenesis and archaeal flagella were also identified in nanoarchaeotal MAGs from geographically distant deep-sea hydrothermal vents, suggesting these features may be widely distributed across the *Nanoarchaeota* [65].

Once again, this pattern suggests that *N. equitans* represents a more highly reduced lineage compared to other described nanoarchaeotes. However, as more *Nanoarchaeota*-

host pairings are identified, additional ecophysiological patterns may emerge that are not yet apparent.

Conclusions

The *Nanoarchaeota* represent a diverse, widely distributed lineage thriving in geothermal environments around the world. The highly specialized host-dependent lifestyle of the nanoarchaeotes provides unique insights into interspecies relationships across the *Archaea*. However, the three described *Nanoarchaeota*-host systems defy easy classification. Although the nanoarchaeotes initially resemble parasites, relying on their hosts for the majority of metabolic products, they cause fewer detrimental effects to the host than expected. Marine and terrestrial *Nanoarchaeota* also appear to interact with their hosts in different ways, as evidenced by unique growth patterns during cultivation. To date, no positive benefits of *Nanoarchaeota* infection have been demonstrated in the laboratory setting, although nanoarchaeotal association may provide yet unknown benefits to hosts in their diverse natural communities. Further cultivation and genome sequencing will be crucial to understanding the intricacies of the *Nanoarchaeota*-host interaction, and the role this symbiotic system plays in the ecology of hot springs and deep-sea hydrothermal vents.

Scope of thesis

Although our understanding of nanoarchaeotal diversity and physiology has advanced significantly since the discovery of *N. equitans* in the early 2000's, many questions regarding nanoarchaeotal biology remain unanswered. Despite the prevalence

of nanoarchaeotal 16S rRNA gene sequences in high-temperature environments, few *Nanoarchaeota* have been cultivated with their hosts, and only a small number of nanoarchaeote genomes have been sequenced. To assess the functional and phylogenetic diversity of nanoarchaeotes from geographically distant hot spring environments, a new *Nanoarchaeota*-host system from New Zealand was cultivated and described on the genomic level, as reported in Chapter 2. The similarities between the New Zealand and YNP nanoarchaeotes and their divergence from marine *N. equitans* suggested that terrestrial nanoarchaeotes may harbor a unique set of features lacking in their marine relatives. To further investigate this question, nanoarchaeote MAGs were recovered from deep submarine hydrothermal vent sites (Chapter 3). However, genomic analysis of these MAGs revealed no clear pattern between marine and terrestrial lineages in terms of coding potential. Nonetheless, the studies presented in Chapters 2 and 3 display the phylogenetic diversity of the *Nanoarchaeota* and provide fresh insights into this enigmatic phylum.

References

- [1] Albers, S.-V., Jarrell, K.F. (2015) The archaellum: how *Archaea* swim. *Front. Microbiol.* 6, 23, <https://doi.org/10.3389/fmicb.2015.00023>.
- [2] Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., Verberkmoes, N.C., Hettich, R.L., Banfield, J.F. (2010) Enigmatic, ultrasmall, uncultivated *Archaea*. *Proc. Natl. Acad. Sci. U. S. A.* 107(19), 8806–8811.
- [3] Baker, B.J., Saw, J.H., Lind, A.E., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., Ettema, T.J.G. (2016) Genomic inference of the metabolism of cosmopolitan subsurface *Archaea*, *Hadesarchaea*. *Nat Microbiol* 1, 16002, <https://doi.org/10.1038/nmicrobiol.2016.2>.
- [4] Barns, S.M., Delwiche, C.F., Palmer, J.D., Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* 93(17), 9188–9193.
- [5] Barns, S.M., Fundyga, R.E., Jeffries, M.W., Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc. Natl. Acad. Sci. U. S. A.* 91(5), 1609–1613.
- [6] Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P. (2008) Mesophilic *Crenarchaeota*: proposal for a third archaeal phylum, the *Thaumarchaeota*. *Nat. Rev. Microbiol.* 6, 245–252.
- [7] Brochier-Armanet, C., Forterre, P., Gribaldo, S. (2011) Phylogeny and evolution of the *Archaea*: one hundred genomes later. *Curr. Opin. Microbiol.* 14, 274–281.
- [8] Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., Forterre, P. (2005)

- Nanoarchaea*: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to *Thermococcales*? *Genome Biol.* 6(5), R42, <https://doi.org/10.1186/gb-2005-6-5-r42>.
- [9] Casanueva, A., Galada, N., Baker, G.C., Grant, W.D., Heaphy, S., Jones, B., Yanhe, M., Ventosa, A., Blamey, J., Cowan, D.A. (2008) Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* 12, 651–656.
- [10] Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., Taylor, R.C., Williams, K.H., Banfield, J.F. (2015) Genomic expansion of domain *Archaea* highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* 25(6), P690–P701.
- [11] Chan, P.P., Cozen, A.E., Lowe, T.M. (2011) Discovery of permuted and recently split transfer RNAs in *Archaea*. *Genome Biol.* 12, R38, <https://doi.org/10.1186/gb-2011-12-4-r38>.
- [12] Chernyh, N.A., Mardanov, A. V., Gumerov, V.M., Miroshnichenko, M.L., Lebedinsky, A. V., Merkel, A.Y., Crowe, D., Pimenov, N. V., Rusanov, I.I., Ravin, N. V., Moran, M.A., Bonch-Osmolovskaya, E.A. (2015) Microbial life in Bourlyashchy, the hottest thermal pool of Uzon Caldera, Kamchatka. *Extremophiles* 19(6), 1157–1171.
- [13] Clingenpeel, S., Kan, J., Macur, R.E., Woyke, T., Lovalvo, D., Varley, J., Inskeep, W.P., Nealson, K., McDermott, T.R. (2013) Yellowstone Lake *Nanoarchaeota*. *Front. Microbiol.* 4, 274, <https://doi.org/10.3389/fmicb.2013.00274>.

- [14] Collingro, A., Köstlbacher, S., Musmann, M., Stepanauskas, R., Hallam, S.J., Horn, M. (2017) Unexpected genomic features in widespread intracellular *Bacteria*: evidence for motility of marine *Chlamydiae*. ISME J. 11(10), 2334–2344.
- [15] Comolli, L.R., Banfield, J.F. (2014) Inter-species interconnections in acid mine drainage microbial communities. Front. Microbiol. 5, 367, <https://doi.org/10.3389/fmicb.2014.00367>.
- [16] DeLong, E.F. (1992) *Archaea* in coastal marine environments. Proc. Natl. Acad. Sci. U. S. A. 89(12), 5685–5689.
- [17] Di Giulio, M. (2014) The split genes of *Nanoarchaeum equitans* have not originated in its lineage and have been merged in another *Nanoarchaeota*: A reply to Podar et al. J. Theor. Biol. 349, 167–169.
- [18] Di Giulio, M. (2008) Transfer RNA genes in pieces are an ancestral character. EMBO Rep. 9(9), 820.
- [19] Evans, D., Marquez, S.M., Pace, N.R. (2006) RNase P: interface of the RNA and protein worlds. Trends Biochem. Sci. 31(6), 333–341.
- [20] Flores, G.E., Campbell, J.H., Kirshtein, J.D., Meneghin, J., Podar, M., Steinberg, J.I., Seewald, J.S., Tivey, M.K., Voytek, M.A., Yang, Z.K., Reysenbach, A.L. (2011) Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. Environ. Microbiol. 13(8), 2158–2171.
- [21] Flores, G.E., Shakya, M., Meneghin, J., Yang, Z.K., Seewald, J.S., Geoff Wheat, C., Podar, M., Reysenbach, A.-L. (2012) Inter-field variability in the microbial

- communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10(4), 333–346.
- [22] Fuhrman, J.A., McCallum, K. (1992) Novel major archaeobacterial group from marine plankton. *Nature* 356(6365), 148–149.
- [23] Fujishima, K., Sugahara, J., Kikuta, K., Hirano, R., Sato, A., Tomita, M., Kanai, A. (2009) Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in *Archaea*. *Proc. Natl. Acad. Sci.* 106(8), 2683–2687.
- [24] Giannone, R.J., Huber, H., Karpinets, T., Heimerl, T., Küper, U., Rachel, R., Keller, M., Hettich, R.L., Podar, M. (2011) Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans*-*Ignicoccus hospitalis* relationship. *PLoS One* 6(8), e22942, <https://doi.org/10.1371/journal.pone.0022942>.
- [25] Giannone, R.J., Wurch, L.L., Heimerl, T., Martin, S., Yang, Z., Huber, H., Rachel, R., Hettich, R.L., Podar, M. (2015) Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. *ISME J.* 9(1), 101–114.
- [26] Golyshina, O. V., Toshchakov, S. V., Makarova, K.S., Gavrillov, S.N., Korzhenkov, A.A., La Cono, V., Arcadi, E., Nechitaylo, T.Y., Ferrer, M., Kublanov, I. V., Wolf, Y.I., Yakimov, M.M., Golyshin, P.N. (2017) “ARMAN” *Archaea* depend on association with euryarchaeal host in culture and in situ. *Nat. Commun.* 8, 60, <https://doi.org/10.1038/s41467-017-00104-7>.
- [27] Guy, L., Ettema, T.J.G. (2011) The archaeal “TACK” superphylum and the origin

of eukaryotes. Trends Microbiol. 19(12), 580–587.

- [28] Hamerly, T., Tripet, B.P., Tigges, M., Giannone, R.J., Wurch, L., Hettich, R.L., Podar, M., Copié, V., Bothner, B. (2015) Untargeted metabolomics studies employing NMR and LC–MS reveal metabolic coupling between *Nanoarchaeum equitans* and its archaeal host *Ignicoccus hospitalis*. Metabolomics 11, 895–907.
- [29] Heimerl, T., Flechsler, J., Pickl, C., Heinz, V., Salecker, B., Zweck, J., Wanner, G., Geimer, S., Samson, R.Y., Bell, S.D., Huber, H., Wirth, R., Wurch, L., Podar, M., Rachel, R. (2017) A complex endomembrane system in the archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. Front. Microbiol. 8, 1072, <https://doi.org/10.3389/fmicb.2017.01072>.
- [30] Hohn, M.J., Hedlund, B.P., Huber, H. (2002) Detection of 16S rDNA sequences representing the novel phylum “*Nanoarchaeota*”: Indication for a wide distribution in high temperature biotopes. Syst. Appl. Microbiol. 25, 551–554.
- [31] Huber, H., Burggraf, S., Mayer, T., Wyschkony, I., Rachel, R., Stetter, K.O. (2000) *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic *Archaea*, represented by two new species, *Ignicoccus islandicus* sp. nov. and *Ignicoccus pacificus* sp. nov. Int. J. Syst. Evol. Microbiol. 50(6), 2093–2100.
- [32] Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I.A., Kockelkorn, D., Eisenreich, W., Fuchs, G. (2008) A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic archaeum *Ignicoccus hospitalis*. Proc. Natl. Acad. Sci. 105(22), 7851–7856.
- [33] Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002)

A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* 417(6884), 63–67.

- [34] Jahn, U., Summons, R., Sturt, H., Grosjean, E., Huber, H. (2004) Composition of the lipids of *Nanoarchaeum equitans* and their origin from its host *Ignicoccus* sp. strain KIN4/I. *Arch. Microbiol.* 190(5), 1743–1750.
- [35] Jarett, J.K., Nayfach, S., Podar, M., Inskeep, W., Ivanova, N.N., Munson-McGee, J., Schulz, F., Young, M., Jay, Z.J., Beam, J.P., Kyrpides, N.C., Malmstrom, R.R. (2018) Single-cell genomics of co-sorted *Nanoarchaeota* suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* 6(161), <https://doi.org/10.1186/s40168-018-0539-8>.
- [36] Jungbluth, S.P., Amend, J.P., Rappé, M.S. (2017) Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci. Data* 4, 170037, <https://doi.org/10.1038/sdata.2017.37>.
- [37] Könneke, M., Bernhard, A.E., De La Torre, J.R., Walker, C.B., Waterbury, J.B., Stahl, D.A. (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* 437(22), 543–546.
- [38] Kuper, U., Meyer, C., Muller, V., Rachel, R., Huber, H. (2010) Energized outer membrane and spatial separation of metabolic processes in the hyperthermophilic archaeon *Ignicoccus hospitalis*. *Proc. Natl. Acad. Sci.* 107(7), 3152–3156.
- [39] McCliment, E.A., Voglesonger, K.M., O’Day, P.A., Dunn, E.E., Holloway, J.R., Cary, S.C. (2006) Colonization of nascent, deep-sea hydrothermal vents by a novel archaeal and nanoarchaeal assemblage. *Environ. Microbiol.* 8(1), 114–125.
- [40] McCutcheon, J.P., Moran, N.A. (2012) Extreme genome reduction in symbiotic

- Bacteria*. Nat. Rev. Microbiol. 10, 13–26.
- [41] Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., Wang, F. (2014) Genetic and functional properties of uncultivated MCG *Archaea* assessed by metagenome and gene expression analyses. ISME J. 8, 650–659.
 - [42] Mohanty, S., Jobichen, C., Chichili, V.P.R., Velázquez-Campoy, A., Low, B.C., Hogue, C.W.V., Sivaraman, J. (2015) Structural basis for a unique ATP synthase core complex from *Nanoarchaeum equitans*. J. Biol. Chem. 290(45), 27280–27296.
 - [43] Moran, N.A., Plague, G.R. (2004) Genomic changes following host restriction in *Bacteria*. Curr. Opin. Genet. Dev. 14, 627–633.
 - [44] Munson-McGee, J.H., Field, E.K., Bateson, M., Rooney, C., Stepanauskas, R., Young, M.J. (2015) *Nanoarchaeota*, their *Sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone National Park hot springs. Appl. Environ. Microbiol. 81(22), 7860–7868.
 - [45] Munson-McGee, J.H., Peng, S., Dewerff, S., Stepanauskas, R., Whitaker, R.J., Weitz, J.S., Young, M.J. (2018) A virus or more in (nearly) every cell: ubiquitous networks of virus-host interactions in extreme environments. ISME J. 12, 1706–1714.
 - [46] Narasingarao, P., Podell, S., Ugalde, J.A., Brochier-Armanet, C., Emerson, J.B., Brocks, J.J., Heidelberg, K.B., Banfield, J.F., Allen, E.E. (2012) *De novo* metagenomic assembly reveals abundant novel major lineage of *Archaea* in hypersaline microbial communities. ISME J. 6, 81–93.
 - [47] Nicks, T., Rahn-Lee, L. (2017) Inside out: archaeal ectosymbionts suggest a

second model of reduced-genome evolution. *Front. Microbiol.* 8, 384,

<https://doi.org/10.3389/fmicb.2017.00384>.

- [48] Ortiz-Alvarez, R., Casamayor, E.O. (2016) High occurrence of *Pacearchaeota* and *Woesearchaeota* (*Archaea* superphylum DPANN) in the surface waters of oligotrophic high-altitude lakes. *Environ. Microbiol. Rep.* 8(2), 210–217.
- [49] Paper, W., Jahn, U., Hohn, M.J., Kronner, M., Näther, D.J., Burghardt, T., Rachel, R., Stetter, K.O., Huber, H. (2007) *Ignicoccus hospitalis* sp. nov., the host of “*Nanoarchaeum equitans*.” *Int. J. Syst. Evol. Microbiol.* 57(Pt 4), 803–808.
- [50] Penz, T., Schmitz-Esser, S., Kelly, S.E., Cass, B.N., Müller, A., Woyke, T., Malfatti, S.A., Hunter, M.S., Horn, M. (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in *Cardinium hertigii*. *PLoS Genet.* 8(10), e1003012, <https://doi.org/10.1371/journal.pgen.1003012>.
- [51] Petitjean, C., Deschamps, P., López-García, P., Moreira, D. (2014) Rooting the domain *Archaea* by phylogenomic analysis supports the foundation of the new kingdom *Proteoarchaeota*. *Genome Biol. Evol.* 7(1), 191–204.
- [52] Podar, M., Anderson, I., Makarova, K.S., Elkins, J.G., Ivanova, N., Wall, M.A., Lykidis, A., Mavromatis, K., Sun, H., Hudson, M.E., Chen, W., Deciu, C., Hutchison, D., Eads, J.R., Anderson, A., Fernandes, F., Szeto, E., Lapidus, A., Kyrpides, N.C., Saier Jr, M.H., Richardson, P.M., Rachel, R., Huber, H., Eisen, J.A., Koonin, E. V., Keller, M., Stetter, K.O. (2008) A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*. *Genome Biol.* 9(11), R158, <https://doi.org/10.1186/gb-2008-9-11-r158>.
- [53] Podar, M., Makarova, K.S., Graham, D.E., Wolf, Y.I., Koonin, E. V., Reysenbach,

- A.-L. (2013) Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct* 8, 9, <https://doi.org/10.1186/1745-6150-8-9>.
- [54] Preston, C.M., Wu, K.Y., Molinski, T.F., DeLong, E.F. (1996) A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl. Acad. Sci. U. S. A.* 93(13), 6241–6246.
- [55] Randau, L. (2012) RNA processing in the minimal organism *Nanoarchaeum equitans*. *Genome Biol.* 13(7), R63, <https://doi.org/10.1186/gb-2012-13-7-r63>.
- [56] Randau, L., Calvin, K., Hall, M., Yuan, J., Podar, M., Li, H., Söll, D. (2005) The heteromeric *Nanoarchaeum equitans* splicing endonuclease cleaves noncanonical bulge-helix-bulge motifs of joined tRNA halves. *Proc. Natl. Acad. Sci. U. S. A.* 102(50), 17934–17939.
- [57] Randau, L., Pearson, M., Söll, D. (2005) The complete set of tRNA species in *Nanoarchaeum equitans*. *FEBS Lett.* 579(13), 2945–2947.
- [58] Randau, L., Schröder, I., Söll, D. (2008) Life without RNase P. *Nature* 453(7191), 120–123.
- [59] Randau, L., Söll, D. (2008) Transfer RNA genes in pieces. *EMBO Rep.* 9(7), 623–628.
- [60] Rinke, C., Rubino, F., Messer, L.F., Youssef, N., Parks, D.H., Chuvochina, M., Brown, M., Jeffries, T., Tyson, G.W., Seymour, J.R., Hugenholtz, P. (2018) A phylogenomic and ecological analysis of the globally abundant Marine Group II *Archaea* (Ca. *Poseidoniales* ord. nov.). *ISME J.*, <https://doi.org/10.1038/s41396-018-0282-y>.

- [61] Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.-T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459), 431–437.
- [62] Schleper, C., Jurgens, G., Jonscheit, M. (2005) Genomic studies of uncultivated *Archaea*. *Nat. Rev. Microbiol.* 3(6), 479–488.
- [63] Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G. (2015) Complex *Archaea* that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551), 173–179.
- [64] St. John, E., Liu, Y., Podar, M., Stott, M.B., Meneghin, J., Chen, Z., Lagutin, K., Mitchell, K., Reysenbach, A.-L. (In press) A new symbiotic nanoarchaeote (*Candidatus* Nanoclepta minutus) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.*, <https://doi.org/10.1016/j.syapm.2018.08.005>.
- [65] St. John, E., Meneghin, J., Flores, G.E., Reysenbach, A.-L. (In review) Deep-sea hydrothermal vent metagenome-assembled genomes provide insight into the phylum *Nanoarchaeota*. *Environ. Microbiol. Rep.*
- [66] Taylor, G.T., John Pirt, S. (1977) Nutrition and factors limiting the growth of a methanogenic bacterium (*Methanobacterium thermoautotrophicum*). *Arch. Microbiol.* 113(1–2), 17–22.

- [67] Tjaden, J., Winkler, H.H., Schwöppe, C., Van Der Laan, M., Möhlmann, T., Neuhaus, H.E. (1999) Two nucleotide transport proteins in *Chlamydia trachomatis*, one for net nucleoside triphosphate uptake and the other for transport of energy. J. Bacteriol. 181(4), 1196–1202.
- [68] Touchon, M., Moura De Sousa, J.A., PC Rocha, E. (2017) Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Curr. Opin. Microbiol. 38, 66–73.
- [69] Treusch, A.H., Leininger, S., Kietzin, A., Schuster, S.C., Klenk, H.P., Schleper, C. (2005) Novel genes for nitrite reductase and Amo-related proteins indicate a role of uncultivated mesophilic *Crenarchaeota* in nitrogen cycling. Environ. Microbiol. 7(12), 1985–1995.
- [70] Veith, A., Klingl, A., Zolghadr, B., Lauber, K., Mentele, R., Lottspeich, F., Rachel, R., Albers, S.V., Kletzin, A. (2009) *Acidianus*, *Sulfolobus* and *Metallosphaera* surface layers: structure, composition and gene expression. Mol. Microbiol. 73(1), 58–72.
- [71] Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. Proc. Natl. Acad. Sci. U. S. A. 100(22), 12984–12988.
- [72] Williams, T.A., Embley, T.M. (2014) Archaeal “dark matter” and the origin of eukaryotes. Genome Biol. Evol. 6(3), 474–481.

- [73] Woese, C.R., Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74(11), 5088–5090.
- [74] Woese, C.R., Kandler, O., Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc. Natl. Acad. Sci. U. S. A.* 87(12), 4576–4579.
- [75] Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L., Podar, M. (2016) Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. *Nat. Commun.* 7, 12115, <https://www.http://doi.org/10.1038/ncomms12115>.
- [76] Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., Stott, M.B., Nunoura, T., Banfield, J.F., Schramm, A., Baker, B.J., Spang, A., Ettema, T.J.G. (2017) Asgard *Archaea* illuminate the origin of eukaryotic cellular complexity. *Nature* 541(7637), 353–358.
- [77] Zeikus, J.G. (1977) The biology of methanogenic *Bacteria*. *Bacteriol. Rev.* 41(2), 514–541.

Chapter 2: A New Symbiotic Nanoarchaeote (*Candidatus Nanoclepta minutus*) and its Host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand Hot Spring

St. John, E, Liu, Y, Podar, M, Stott, MB, Meneghin, J, Chen, Z, Lagutin, K, Mitchell, K, Reysenbach, A-L (In press). Systematic and Applied Microbiology.
<https://doi.org/10.1016/j.syapm.2018.08.005>.

Abstract

Three thermophilic *Nanoarchaeota*-*Crenarchaeota* symbiotic systems have been described. We obtained another stable anaerobic enrichment culture at 80°C, pH 6.0 from a New Zealand hot spring. The nanoarchaeote (Ncl-1) and its host (NZ3^T) were isolated in co-culture and their genomes assembled. The small (~200 nm) flagellated cocci were often attached to larger cocci. Based on 16S rRNA gene similarity (88.4%) and average amino acid identity (52%), Ncl-1 is closely related to *Candidatus Nanopusillus acidilobi*. Their genomes both encode for archaeal flagella and partial glycolysis and gluconeogenesis pathways, but lack ATP synthase genes. Like *Nanoarchaeum equitans*, Ncl-1 has a CRISPR-Cas system. Ncl-1 also likely relies on its crenarchaeotal host for most of its biosynthetic needs. The host NZ3^T was isolated and grows on proteinaceous substrates but not on sugars, alcohols, or fatty acids. NZ3^T requires thiosulfate and grows best at 82°C, pH 6.0. NZ3^T is most closely related to the *Desulfurococcaceae*, *Ignisphaera aggregans* (~92% 16S rRNA gene sequence similarity, 45% AAI). Based on phylogenetic, physiological and genomic data, Ncl-1 and NZ3^T

represent novel genera in the *Nanoarchaeota* and *Desulfurococcaceae*, respectively, with the proposed names *Candidatus* Nanoclepta minutus and *Zestosphaera tikiterensis* gen. nov., sp. nov., type strain NZ3^T (=DSMZ 107634^T =OCM 1213^T =JCM 32895^T).

Introduction

The *Archaea* still remain the most unexplored of life's domains, despite a ubiquitous distribution and a diversity that may parallel that of the *Bacteria*. Many new lineages without cultivated representatives have been discovered through metagenomic and single cell genomic approaches, giving rise to a significant increase in new archaeal phyla and divisions [4,7,89,95,105]. One exception is the *Nanoarchaeota*, a phylum with only two representatives in culture [36,113]. The first isolate, *Nanoarchaeum equitans*, was obtained from a shallow marine hydrothermal site off Iceland and forms an obligate parasitic relationship with its autotrophic archaeal host, *Ignicoccus hospitalis* [36]. With its highly reduced genome, *N. equitans* lacks most primary biosynthetic functions and has to acquire many metabolic precursors from its host [111]. The second nanoarchaeote in culture was obtained from an acidic hot spring in Yellowstone National Park (YNP) [113]. This terrestrial representative, *Candidatus* Nanopusillus acidilobi, is also an obligate symbiont, and its host is an acidophile and heterotroph, *Acidilobus* sp. 7A. Like its marine relative, *Cand. Nps. acidilobi* shows extensive genome reduction and requires its host for most primary biosynthetic and energetic needs, yet it has retained some potential for glycolysis and gluconeogenesis, expresses archaeal flagella genes and lacks a membrane ATP synthase complex [113]. The differences between the two cultivated nanoarchaeotes point to how diverse symbiotic associations may be within the

Nanoarchaeota. Additional terrestrial *Nanoarchaeota* symbiotic systems were characterized based on single cell sequencing from geothermal acidic pools in YNP [72,89]. Even though these environmental nanoarchaeote genomes were very similar to *Cand. Nps. acidilobi*, their implied hosts were quite dissimilar and more closely related to *Sulfolobus* than *Acidilobus* [72,89].

Based on 16S rRNA gene surveys, *Nanoarchaeota* have been detected globally in terrestrial and marine hot springs. These include many sites in YNP [16,34,71,72], Kamchatka [15,34], Chile [14], Central Asia [14], China [14], New Zealand [14], and deep-sea vents in the Atlantic [24] and Pacific oceans [26,34,67,73]. These different nanoarchaeotes form phylogenetic clades that can be separated somewhat by geographical region. It is therefore likely that as more members of this group and their hosts are characterized, a much better understanding of their biology and evolutionary history will emerge.

Here, we report on the enrichment and isolation of a third nanoarchaeote in stable co-culture with its host. This nanoarchaeote was obtained from a terrestrial hot spring in New Zealand and more closely resembles its relatives from YNP than *N. equitans*. We also describe the isolation and characterization of the host, a novel genus in the *Desulfurococcaceae*. The *Nanoarchaeota*-host system highlights the need for formal recognition and classification for organisms that can be cultivated, but only in the presence of a symbiotic partner.

Materials and methods

Sample site and culturing conditions

Samples from “Cooking Pots” hot springs at Hell’s Gate, Tikitere, New Zealand (38.0654°S, 176.3595°E) were inoculated into anaerobic enrichment media: Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) media #88 with (g/L): $\text{Na}_2\text{O}_3\text{S}_2 \cdot 5\text{H}_2\text{O}$, 2.0; NaOH, 0.16; yeast extract, 0.5; tryptone, 0.5; resazurin, 0.001; $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 0.004; $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 0.00036; and 10 mL additional trace mineral solution DSMZ #141 and dispensed under N_2/CO_2 (80/20, v/v) and pH adjusted to 6.0-6.1. The enrichments were incubated at 80°C for four days and checked by phase contrast microscopy. DNA was extracted using the Qiagen DNeasy Blood and Tissue DNA kit and the presence of *Nanoarchaeota* was checked using nanoarchaeotal-specific 16S rRNA gene primers [34]. Nanoarchaeotal-positive enrichment cultures were monitored over many transfers and maintained stably for over two years. A nanoarchaeote-host co-culture was isolated after several dilution-to-extinction transfers. Subsequently, the co-culture was transferred into, and routinely grown in, the above media with the following modifications: 0.5 g/L casein hydrolysate, 1.0 g/L yeast extract, 1.0 g/L tryptone, 0.0625 g/L L-cysteine-HCl and 10 mL/L glycerol.

Host isolation and 16S rRNA gene sequencing

The host was isolated by several series of dilution-to-extinction transfers in the described modified enrichment media. Genomic DNA was extracted as above, and the purity of the isolate was verified by PCR and sequencing of the 16S rRNA gene. Due to the presence of introns in the host 16S rRNA gene, two overlapping primer sets were

utilized for sequencing: 4F (5'-TCCGGTTGATCCTGCCRG-3'), 907RN (5'-CCGYCAATTCMTTTRAGTTT-3'), 341FA (5'-CCTAYGGGGYGCASCAGGCG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). PCR conditions consisted of initial denaturation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min 30 s, followed by a final elongation cycle at 72°C for 7 min.

Physiological characterization of the host

Unless otherwise noted, all characterization studies were carried out in triplicate at 80-85°C in sealed anaerobic Balch tubes. Cell density was monitored by phase contrast microscopy with a Petroff-Hausser counting chamber. Use of carbon sources was determined at a 0.1% (w/v) final concentration using a modified enrichment media with reduced yeast extract (0.01% w/v) and no casein hydrolysate, tryptone, or glycerol. Cultures were transferred at least once to minimize carry-over. The carbon sources that were tested included yeast extract, tryptone, casamino acids, gelatin, acetate, butyrate, cellulose, fructose, glucose, glycerol, mannose, propionate, sorbitol, ribose, starch, sucrose and xylose.

Alternate electron acceptor use was determined by replacing L-cysteine-HCl and thiosulfate in the enrichment media with a variety of electron acceptors at both 2.5 and 10 mM final concentrations. To prevent sublimation, elemental sulfur (2.5 and 10 mM) was added directly to media in individual Balch tubes prior to autoclaving at 105°C for 1 hour [25]. Potential electron acceptors included elemental sulfur, nitrate, thiosulfate, sulfite and L-cysteine-HCl. A positive growth response was only recorded after three consecutive transfers in the same medium.

To assess the pH range of isolate growth, enrichment media was supplemented with 25 mM MES (pH 5.5-6.0) and 25 mM MOPS (pH 6.3-6.8) [78]. The pH was adjusted to the desired range prior to autoclaving. Oxygen tolerance was tested by adding defined volumes (0-10%, v/v) of filtered oxygen directly to anaerobic culture tubes containing media with 10 mM thiosulfate but lacking L-cysteine HCl. Requirement for NaCl was determined using 0-1% NaCl (w/v).

Core and intact lipid preparation and analysis

Lipids were recovered using a modified Bligh and Dyer method [9], and polar head groups were removed by acidic methanolysis [79]. Core and intact lipids were analyzed using a modified version of the protocol described by Elling et al, 2016 [22]. Liquid chromatography-mass spectrometry (LCMS) was performed using a Shimadzu 8040-LCMS system equipped with a photodiode array detector and a triple quadrupole low resolution MS utilizing an ESI source for ionization and Q3 scanning from 300 to 2000 m/z . Lipids were identified based on comparison with published mass spectrometry data [69,115]. Additional methodologies for core and intact polar lipid analysis are available in Appendix A.

Electron microscopy

Negatively stained cells were prepared for transmission electron microscopy (TEM) by spotting 5 μ L of culture onto formvar-carbon copper grids (Ted Pella, 400 mesh), and stained with 2% (w/v) uranyl acetate for 15 s and air dried. The micrographs were taken at acceleration voltage of 200 kV using a FEI TECNAI F-20 TEM.

qPCR assay

qPCR was performed using the QuantiTect® SYBR® Green PCR kit (Qiagen). Standards were developed with double-stranded 16S rRNA gene blocks (500bp, gBlocks®, Integrated DNA Technologies) containing 165 bp regions flanking the target qPCR product. The qPCR primers were specific for the New Zealand nanoarchaeote: 515F-NZ (5'-GTGGGCAAGAGAGGTGG-3') and 650R-NZ (5'-CGGCTCCTTCCGTC CCCTAG-3') and were used at 0.2 µM final concentration. qPCR reaction conditions were as follows: 95°C for 15 min, followed by 35 cycles of 94°C for 15 s, 55°C for 30 s, 72°C for 30 s, followed by melt curve analysis from 55 to 95°C.

Cell sorting and metagenomic sequencing

To obtain low diversity metagenomes, enrichment cultures containing nanoarchaeotes were sorted using flow cytometry as described previously [89]. To ensure recovery of nanoarchaeotes, multiple cells (5-10) were sorted into each well. After multiple displacement amplification (MDA), wells were screened using nanoarchaeotal-specific PCR primers [34,89]. Two wells with positive nanoarchaeotal amplification were selected for sequencing using the Illumina MiSeq. Adaptor sequences and low-quality regions were trimmed using Trim Galore [52]. After trimming, the two datasets yielded approximately 4.0 and 4.5 million paired-end reads, respectively.

Binning and assembly of metagenome

To maximize recovery of *Nanoarchaeota* and host genomes, raw reads from the two low diversity metagenomes were merged into one dataset. Reads were initially

assembled using IDBA-UD v.1.1.0 [85,86]. Contigs ≥ 5 kb were binned using MaxBin v.1.4.5 [112]. To enhance data recovery, the combined dataset was re-assembled using the SPAdes genome assembler v.3.9.1-Linux [8]. Supervised binning of contigs ≥ 2.5 kb was performed using emergent self-organizing mapping (ESOM) based on tetranucleotide frequency [109]. Both the IDBA-UD and SPAdes assemblies were then merged and viewed in Geneious v.10.0.2 [46] for manual contig extension, curation and identification of misassembled regions. Completeness and contamination of final draft genomes was assessed using CheckM [84].

Accessibility of data and biological material

Draft genomes of the New Zealand nanoarchaeote and its host are available in Genbank under accession numbers MWM100000000 and NBVN00000000, respectively. 16S rRNA genes of the above strains have also been deposited under accession numbers MH298645 and MH252993. Project details are available in the NCBI BioProject database (PRJNA363059), and sampling metadata can be found in the NCBI BioSample database for the New Zealand nanoarchaeote (SAMN06250747) and its host (SAMN06648824). The nanoarchaeote-host co-culture has also been deposited in the Oregon Culture Collection of Methanogens (OCM 1214), and the host has been deposited in the OCM (OCM 1213^T), the DSMZ (DSM 107634^T) and the JCM (JCM 32895^T).

Functional annotation

After assembly, open reading frames were predicted and annotated using the Rapid Annotation and Subsystem Technology (RAST) server [5,11,82]. Additional

annotation (Appendix B, File S1, Table A1-A2) was performed using the updated Archaeal Clusters of Orthologous Genes (arCOG) database [64,65]. As needed, annotations were cross-referenced using the NCBI Basic Local Alignment Search Tool (BLAST) [3] and the NCBI Conserved Domain (CD) Database [66,77]. Ribosomal RNAs were predicted with Infernal v.1.1.2 [75] and Rfam v.12.0 [74] and tRNAs were predicted with the tRNAscan-SE 2.0 web server [60]. Hydrogenases were predicted using the HydDB web server [104]. Circos ribbon plots [53] were used to compare overall genomic synteny between draft genomes and their closest described relatives. Pairwise average amino acid identity matrices were generated using the ANI/AAI Matrix Calculator [96].

Phylogenetic analysis

To assess the phylogenetic placement of the nanoarchaeote and its host, their 16S rRNA genes were aligned using the Geneious Aligner [46] and hypervariable regions were manually masked. Maximum-likelihood analysis was inferred with RAxML v.8.2.8 [106] under the GTRGAMMA model with 1000 replicate bootstrap trees, and trees were viewed using the Interactive Tree of Life [58]. 16S rRNA gene similarity was also assessed using EzBioCloud [48,114] and pairwise Geneious alignments. Additional phylogenetic trees were built using a set of 16 ribosomal proteins [4] extracted with PhyloSift v.1.0.1 [19]. After each set of ribosomal proteins was aligned in Muscle [20], alignments were concatenated in Geneious and maximum-likelihood analysis was performed using the PROTGAMMALG model as described. This model was also used to build single gene trees.

Analysis of predicted rRNA and tRNA intron sequences

Transfer RNA introns in strain Ncl-1 were predicted using tRNAscan-SE [60]. To search for the presence of Group I and Group II introns, Ncl-1 tRNAs with predicted introns were screened against custom Group I intron covariance models (CMs) generated by Nawrocki et al. [76] and the Rfam v.14.0 database [44,45,74] using Infernal [75]. As previously described [76], all hits with an E-value ≤ 0.001 were considered in these analyses. The same approach was also used to classify introns in the strain NZ3^T 16S rRNA gene sequence. In both cases, intron sequences were also translated into all six potential translation frames using Geneious [46] and screened against the CD Database [66,77]. A high E-value threshold (0.1) was set to ensure the detection of low-homology matches. The Mfold web server [118] was used to predict the secondary structure of introns in the NZ3^T 16S rRNA gene sequence.

Results and discussion

Establishment of a stable nanoarchaeotal-host enrichment culture

Initial enrichments containing the New Zealand nanoarchaeote were complex (Fig. 1.1A) and often dominated by *Bacteria* belonging to the *Thermodesulfobacteriaceae* (97.47% similar in 16S rRNA gene sequence to *Thermodesulfobacterium geofontis* [32,48,114]). Other members of these enrichments included the *Korarchaeota*, *Thermoanaerobacteraceae* related to the *Caldanaerobacter*, *Dictyoglomus* sp., *Fervidobacterium* sp., *Thermofilum* sp., *Pyrobaculum* sp. and some unclassified *Thermoprotei* related to *Desulfurococcus* and *Ignisphaera*. Notably, some enrichments had up to 20% korarchaeotal sequences. Tracking the initial enrichment

cultures by barcoded 16S rRNA gene sequencing (data not shown) revealed that *Thermodesulfobacteriaceae* reached stationary phase after 120 h. *Thermodesulfobacteriaceae* were also in highest relative abundance when the *Korarchaeota* were prevalent. Potentially, *Thermodesulfobacteriaceae* stimulated korarchaeotal growth by scavenging molecular hydrogen, as proposed by Elkins et al. [21]. The *Caldanaerobacter* sp. and *Fervidobacterium* sp. grew rapidly and reached stationary phase of growth after 24 h, while the *Dictyoglomus* sp. peaked after 72 h and the *Korarchaeota* after 144 h. The *Thermoprotei* and nanoarchaeotes were in very low relative sequence abundance and reached stationary growth phase around 72 h of incubation. From these enrichments, we successfully isolated pure cultures of strains related to *Dictyoglomus*, *Thermofilum*, *Thermodesulfobacterium*, *Caldanaerobacter*, *Desulfurococcus* and *Ignisphaera*.

Based on qPCR, the abundance of nanoarchaeotes in the enrichments increased with additional transfers, but then plateaued around $4\text{--}8 \times 10^6$ nanoarchaeotes/mL. Several early enrichments showed nanoarchaeotes attached to some, but not all, cocci (Fig. 1.1B). Nanoarchaeotal qPCR numbers closely tracked direct counts of the cocci in the cultures. Addition of sodium sulfide (0.005% w/v) did not stimulate nanoarchaeotal growth. Starch (0.5% w/v) and the presence of hydrogen in the headspace inhibited growth. Highest nanoarchaeotal numbers were obtained between 80–85°C although they were still detected at 60°C and at 90°C.

Early dilution-to-extinction series were not successful in isolating the New Zealand nanoarchaeote and its host. No nanoarchaeotes could be detected in initial serial dilutions greater than 10^{-3} . Possibly the dilutions were outgrown by some other members

of the enrichment, particularly the fastidious *Thermodesulfobacteriaceae*. Furthermore, attempts to obtain colonies with nanoarchaeotes on solid media roll tubes (Gelrite™, Research Products International) were unsuccessful. As repeated attempts to isolate pure nanoarchaeotal-host cultures by these methods were not successful, additional approaches to stimulate nanoarchaeotal growth and obtain less complex enrichments were used. A series of defined enrichments were reconstructed by combining several of the isolate cultures described above. Cultures were grown for four days and then inoculated with nanoarchaeotes by adding 0.22 µm-filtered spent nanoarchaeote-containing media (10% v/v). These mixed cultures were then monitored by qPCR after four, five and six days of incubation. Very low nanoarchaeotal numbers were detected in one reconstructed culture after six days of incubation. This culture was then reinoculated several times into fresh media until a stable culture was established. The resulting mixed culture had $2\text{--}5 \times 10^7$ nanoarchaeotes/mL. All further isolation attempts and metagenomic analyses of the nanoarchaeote-host system were conducted using this reconstructed culture.

Isolation of the nanoarchaeote in co-culture and the host in pure culture

Subsequently, we successfully isolated the New Zealand nanoarchaeote and its crenarchaeotal host in co-culture by screening several 10^{-3} dilutions of the reconstructed culture by PCR. The purity of the co-cultures was confirmed by PCR of the *Nanoarchaeota* and host 16S rRNA genes. On the basis of 16S rRNA gene sequence, the nanoarchaeote (strain Ncl-1) was most closely related (96.6%) to a clone sequence (Genbank EF562624.1, [14]) from Tokaanu hot springs in New Zealand and only 88.% and 84.4% similar to *Cand. Nps. acidilobi* and *N. equitans*, respectively. Phylogenetic

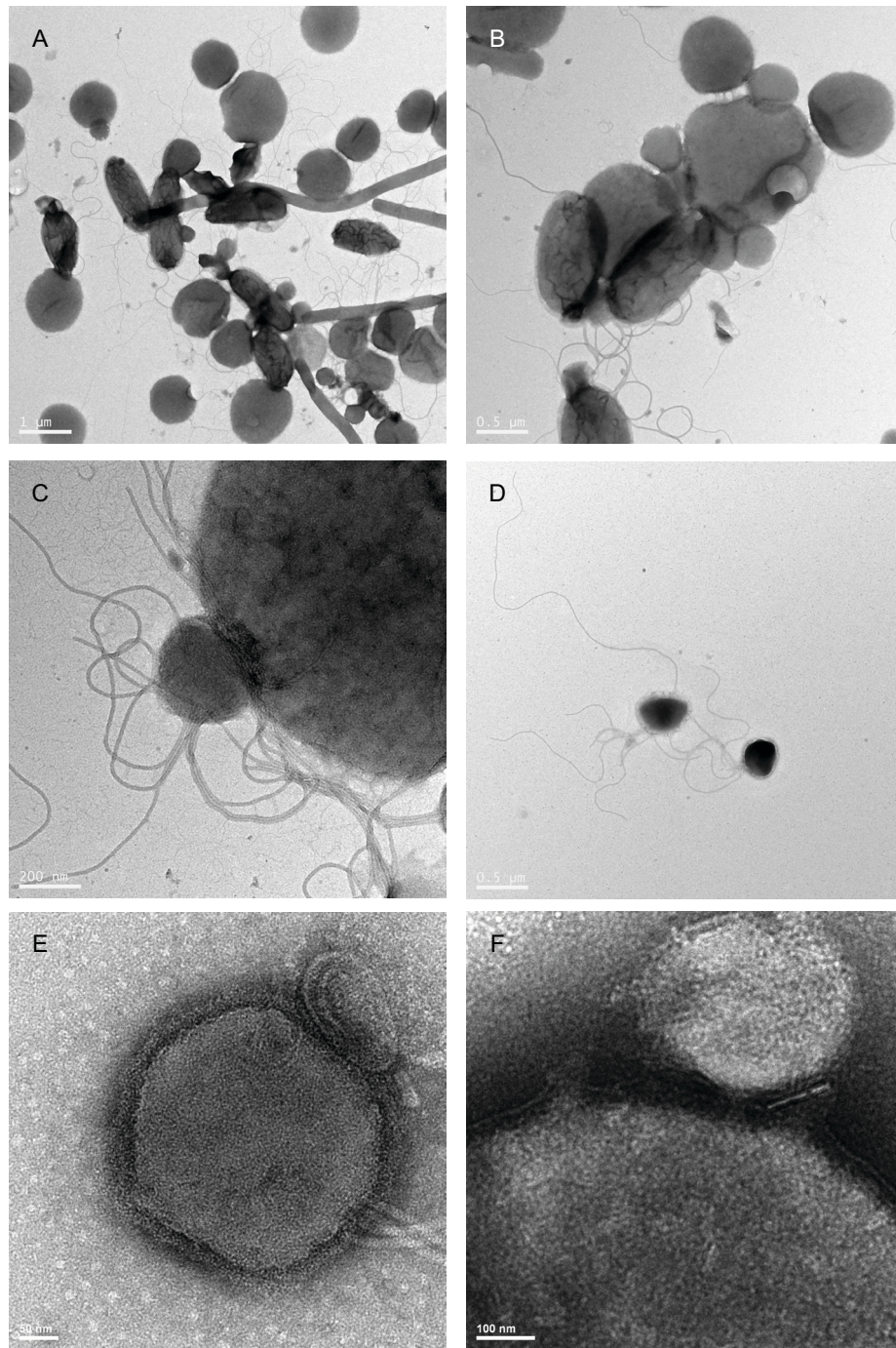


Figure 1.1. Negatively stained transmission electron micrographs (TEM) of the New Zealand enrichment culture. (A) Enrichment culture with nanoarchaeotes, (B) several nanoarchaeotes attached to host coccus, (C) nanoarchaeote attached to host with lophotrichous archaeal flagella, (D) unattached nanoarchaeotes with lophotrichous flagella, (E) detailed view of archaeal flagella attachment, and (F) cell connection between host and nanoarchaeote.

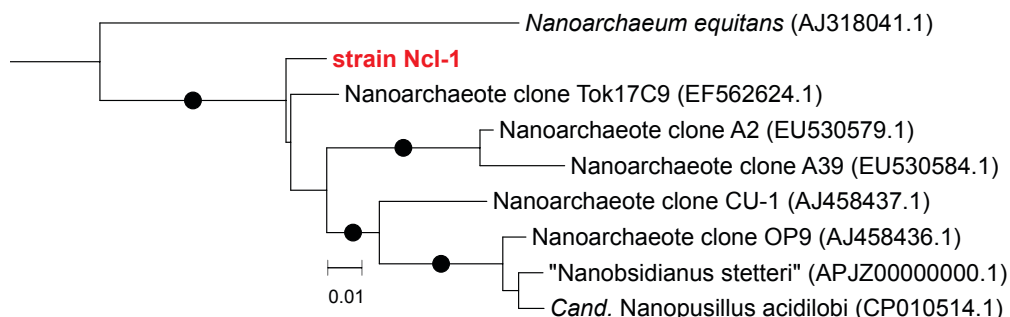


Figure 1.2. Maximum likelihood phylogenetic tree of *Nanoarchaeota* based on 16S rRNA gene sequences. Bootstrap support of $\geq 70\%$ is shown by closed circles. *Desulfurococcaceae* were used as the outgroup. The scale bar indicates 0.01 nucleotide substitutions per site.

analysis of the 16S rRNA gene also confirmed the position of Ncl-1 in the *Nanoarchaeota* (Fig. 1.2).

The New Zealand host (strain NZ3^T) was also isolated after several rounds of dilution-to-extinction transfers of the co-culture. The purity of the isolate was verified by amplification and sequencing of the 16S rRNA gene, which was found to contain four introns (2273 bp total gene length; Fig. 1.3). Introns in archaeal 16S rRNA genes have been widely reported in the *Crenarchaeota*, particularly in the *Desulfurococcales* and *Thermoproteales* [13,39,41,61,80,108], and often encode homing endonucleases which may facilitate their spread throughout microbial communities [1,41,61]. The largest of the four predicted introns in the NZ3^T 16S rRNA gene has extensive secondary structure and contains a predicted homing endonuclease, while the other three introns form small hairpin structures [118]. However, none of the intron sequences showed significant

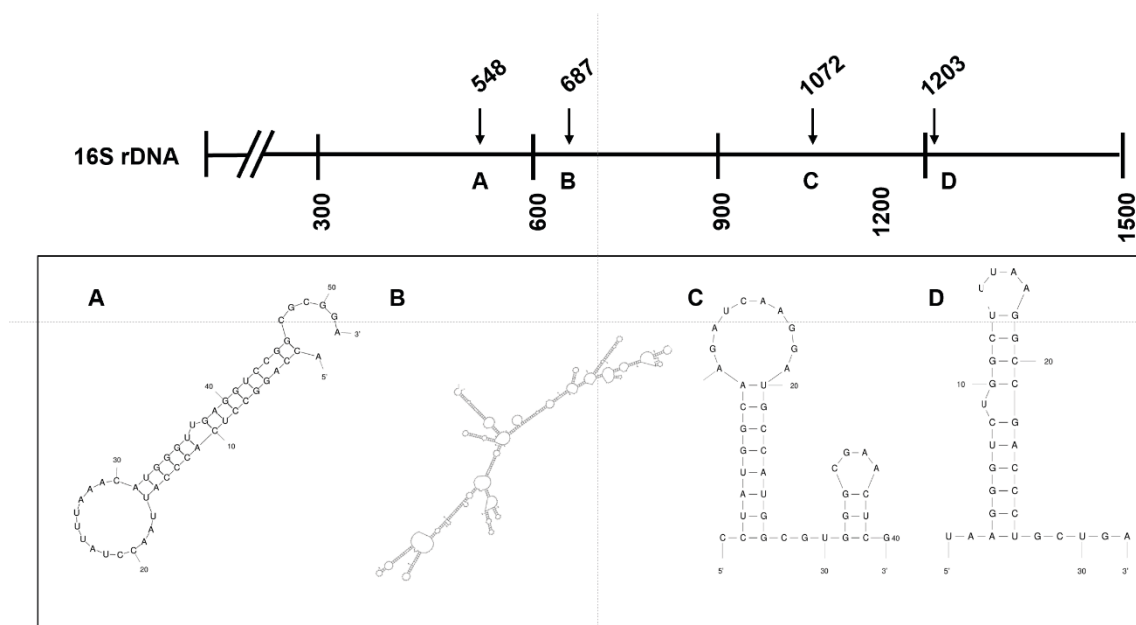


Figure 1.3. Predicted secondary structure and placement of intervening sequences in the 16S rRNA gene of strain NZ3^T. Structures were predicted and drawn in Mfold. Only the most energetically favorable conformations are presented.

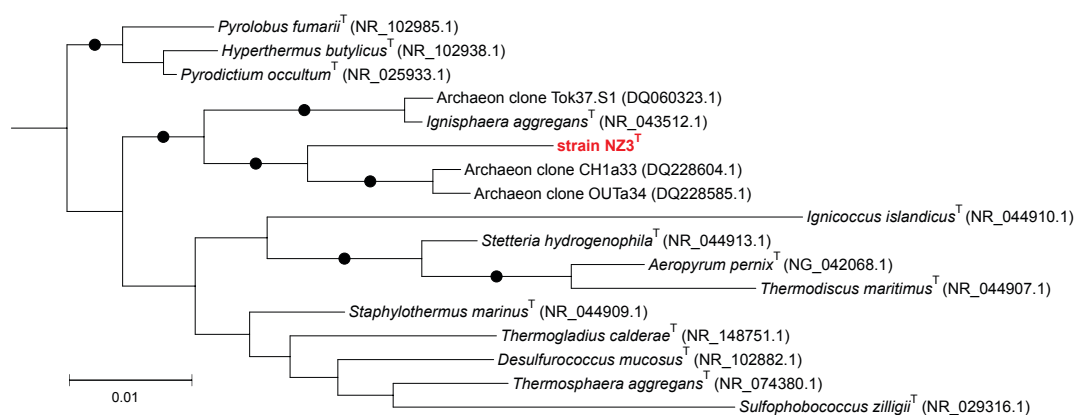


Figure 1.4. 16S rRNA gene maximum likelihood phylogenetic tree of select *Desulfurococcaceae*. *Pyrodictiaceae* were used as the outgroup, and filled circles indicate $\geq 70\%$ bootstrap support. The scale bar represents 0.1 substitutions per nucleotide.

homology (E-value ≤ 0.001) to Group I or Group II introns found in the *Archaea*. After trimming the introns from the sequence, the NZ3^T 16S rRNA gene showed highest sequence similarity (~93%) to clones from deep-sea hydrothermal vents (Genbank DQ228585.1, DQ228604.1, unpublished) and 92% to *Ignisphaera aggregans* [78]. Based on maximum likelihood phylogenetic analysis of the 16S rRNA gene, strain NZ3^T formed a deeply-branching clade in the *Desulfurococcaceae* with *I. aggregans* (Fig. 1.4).

Characterization of the host, strain NZ3^T

The morphology of strain NZ3^T was analyzed using phase contrast microscopy and TEM. Cells were coccoid, approximately 1 μm in diameter, and no archaeal flagella were observed. NZ3^T grew between 60 and 90°C, growing best between 82-83°C (Table 1.1). No growth was observed at 55 or 95°C. The isolate grew from pH 4.9 and 7.2, with optimal growth at pH 6.0, but no growth observed at pH 4.6 or 7.2.

Strain NZ3^T grew well on peptide-rich substrates (0.1% w/v), including yeast extract, tryptone, casamino acids and gelatin. However, no growth was observed on sugars, fatty acids, or alcohols, including fructose, glucose, mannose, ribose, sucrose, xylose, acetate, butyrate, propionate, sorbitol, glycerol, cellulose or starch. Thiosulfate (2.5-10 mM) was required as a terminal electron acceptor and could not be replaced by 2.5 or 10 mM L-cysteine-HCl, sulfite, nitrate or elemental sulfur.

Strain NZ3^T grew in the presence of 0-0.5% NaCl (w/v), but no growth was observed at 1% NaCl. NZ3^T is an obligate anaerobe, as trace amounts of oxygen in the headspace (3% v/v) inhibited growth.

Table 1.1. Characteristics distinguishing strain NZ3^T from other representative species in the *Desulfurococcaceae*.

Strains: 1, NZ3^T (this study); 2, *Ignisphaera aggregans* AQ1.S1^T [78]; 3, *Desulfurococcus mucosus* DSM2162^T [87,116]; 4, *Stetteria hydrogenophila* DSM1227^T [43]; 5, *Ignicoccus islandicus* Kol8^T [35].

Characteristic	1	2	3	4	5
Temperature for growth (°C)					
Range	60-90	85-98	up to 97	70-102	70-98
(optimal)	(82-83)	(92-95)	(85)	(95)	(90)
pH for growth					
Range	4.9-7.0	5.4-7.0	2.2-6.5*	4.5-7.0	3.8-6.5
(optimal)	(6.0)	(6.4)	(5.5-6.0)	(6.0)	(5.8)
NaCl for growth (% w/v)	0 – 0.5	0 – 0.5	x	0.5-6.0	0.3-5.5
Doubling time (hr)	8	7.5	4	3.7	1.4
DNA G+C content (mol%)	41.9**	52.9	53	65	41
Electron acceptors	thiosulfate	none	S ^o	S ^o , thiosulfate	S ^o
Substrate utilization					
Casamino acids	+	-	-	+	-
Gelatin	+	ND	ND	ND	+
Tryptone	+	ND	+	+	+
Yeast extract	+	-	+	+	-
Acetate	-	-	ND	-	-
Butyrate	-	ND	ND	ND	ND
Cellulose	-	ND	ND	ND	ND
Fructose	-	ND	ND	ND	ND
Glucose	-	+	ND	-	-
Glycerol	-	ND	ND	ND	ND
Mannose	-	+	ND	ND	ND
Propionate	-	ND	ND	ND	ND
Ribose	-	ND	ND	-	ND
Sorbitol	-	ND	ND	ND	ND
Starch	-	+	-	-	-
Sucrose	-	+/-	ND	ND	ND
Xylose	-	-	ND	-	ND

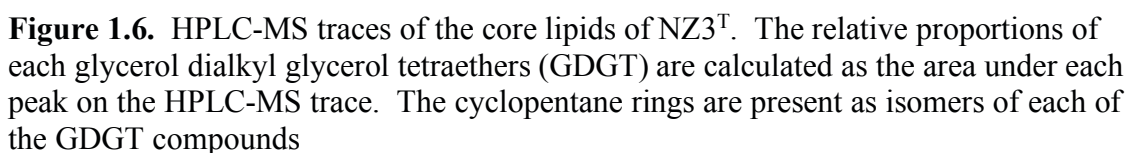
+/-, weakly supported or enhanced growth; +*, stimulates growth; ND, no data available.

*The pH range for *Desulfurococcus mucosus* indicates the ecological conditions in which the strain is found.

**Based on weighted average of % G+C across the draft genome.



* Number in parentheses for GDGTs denotes the number of cyclopentane moieties.
 ** Dihexose was detected in trace amounts and was identified only by molecular ion.



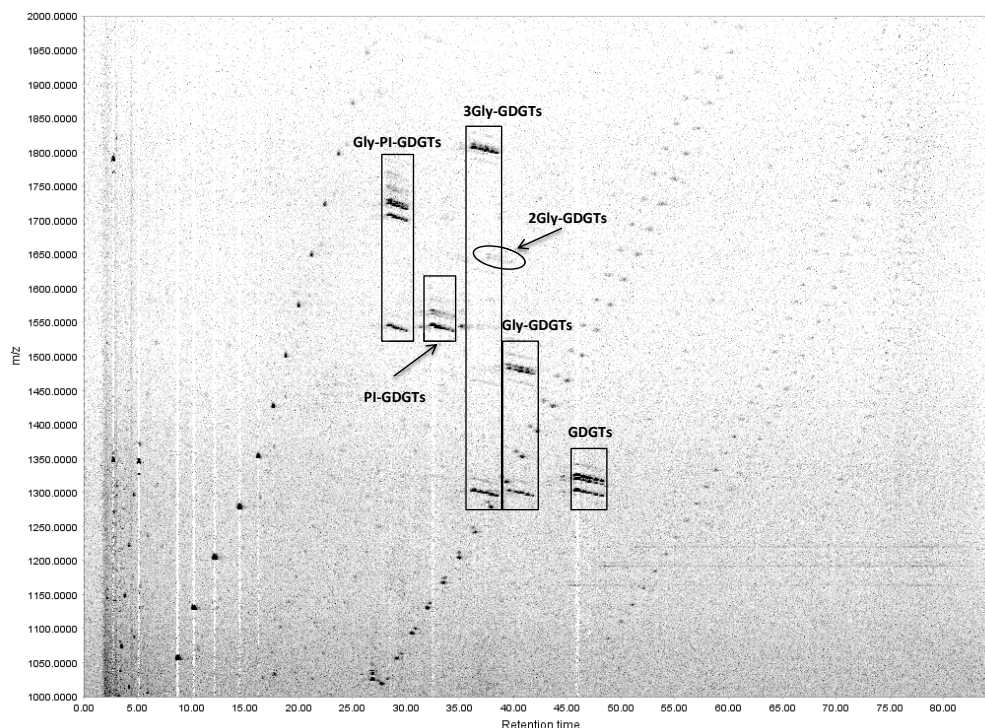


Figure 1.7. LC-MS analysis of the NZ3^T intact polar lipid extract. GDGT; glycerol dialkyl glycerol tetraethers, Gly; glycoside (monohexose), 2Gly; diglycoside (dihexose), 3Gly; triglycoside (trihexose), PI; phosphoinositol, Gly-PI; phosphoinositol.

Table 1.3. Features of the nanoarchaeote strain Ncl-1 and the host strain NZ3^T and their closest cultivated relatives, based on our re-analysis of reference genomes.

	Strain Ncl-1	<i>Cand. Nps. acidilobi</i>	<i>N. equitans</i>	Strain NZ3 ^T	<i>I. aggregans</i>
Contigs	9	1	1	19	1
Assembly size (bp)	575637	605887	490885	1808184	1875953
% G+C	32.2	24.2	31.6	41.9	35.7
Protein coding genes	653	657	570	1961	2055
5S, 16S, 23S rRNA	+	+	+	+	+
tRNA genes	41	43	41	49	47
CRISPR regions	1	0	2	3	8
% arCOG classification (E-value <=0.00001)	73.51	75.19	71.75	80.32	81.51
Marker genes*	121 of 149	119 of 149	118 of 149	241 of 245	245 of 245
Location	NZ hot spring	YNP hot spring	Shallow marine vent	NZ hot spring	NZ hot spring

NZ = New Zealand; YNP = Yellowstone National Park.

*Determined using CheckM [84].

The genome of strain Ncl-1 more closely resembles its terrestrial relatives than marine

Nanoarchaeum equitans

In order to gain insight into the relationship between the nanoarchaeote and its host, their draft genomes were obtained from single-cell sorted metagenomes of the enrichment. Using the 16S rRNA genes, the nanoarchaeote and host bins were identified. The nanoarchaeote strain Ncl-1 draft genome consists of nine contigs, totaling 0.576 Mbp, with a 32% G+C content (Table 1.3). Using CheckM, the draft genome contains 121 of the 149 marker genes used to estimate completeness, while the closed genomes of *N. equitans* and *Cand. Nps. acidilobi* have 118 and 119 of the marker genes, respectively. Ncl-1 is 52% similar to *Cand. Nps. acidilobi* and 43% similar to *N. equitans* based on average amino acid identity (AAI) [96], indicating genus-level divergence [51]. Concatenated ribosomal protein phylogeny also supports the position of Ncl-1 in the *Nanoarchaeota* (Fig. 1.8).

On a genomic level, Ncl-1 most closely resembles its terrestrial relatives, “*Nanobsidianus stetteri*” Nst1 and *Cand. Nps. acidilobi* (Table 1.3), and like other nanoarchaeotes, it lacks the genes encoding for the ability to synthesize most metabolic precursors such as nucleotides, amino acids, cofactors and lipids. Ncl-1 shows modest synteny with the genomes of *Cand. Nps. acidilobi* and *N. equitans* (Fig. 1.9A, B), and it is also highly syntenic with a nanoarchaeote draft genome obtained from a New Zealand hot spring enrichment metagenome collected in 2013 [47] (Fig. 1.9C). As “*N. stetteri*” Nst1 and *Cand. Nps. acidilobi* are very similar at the genomic level, the majority of the comparative analyses in this study were done with the closed genome of *Cand. Nps. acidilobi*. Like *Cand. Nps. acidilobi*, the Ncl-1 draft genome is about 100 kb larger than

N. equitans, has partial glycolytic and gluconeogenic pathways, lacks an apparent ATP synthase complex and contains a full suite of archaeal flagellar genes [113].

In contrast to *N. equitans*, which only has one of the enzymes in the Embden-Meyerhof-Parnas (EMP) glycolysis pathway [10], Ncl-1 encodes several of the EMP pathway genes. Like *Cand. Nps. acidilobi*, however, glucokinase, phosphofructokinase (PFK) and fructose 1,6-bisphosphate aldolase (FBPA) were not detected, suggesting the EMP pathway may not be functional for ATP production in terrestrial nanoarchaeotes. The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) and pyruvate kinase (PK) genes found in *Cand. Nps. acidilobi* were not found in the Ncl-1

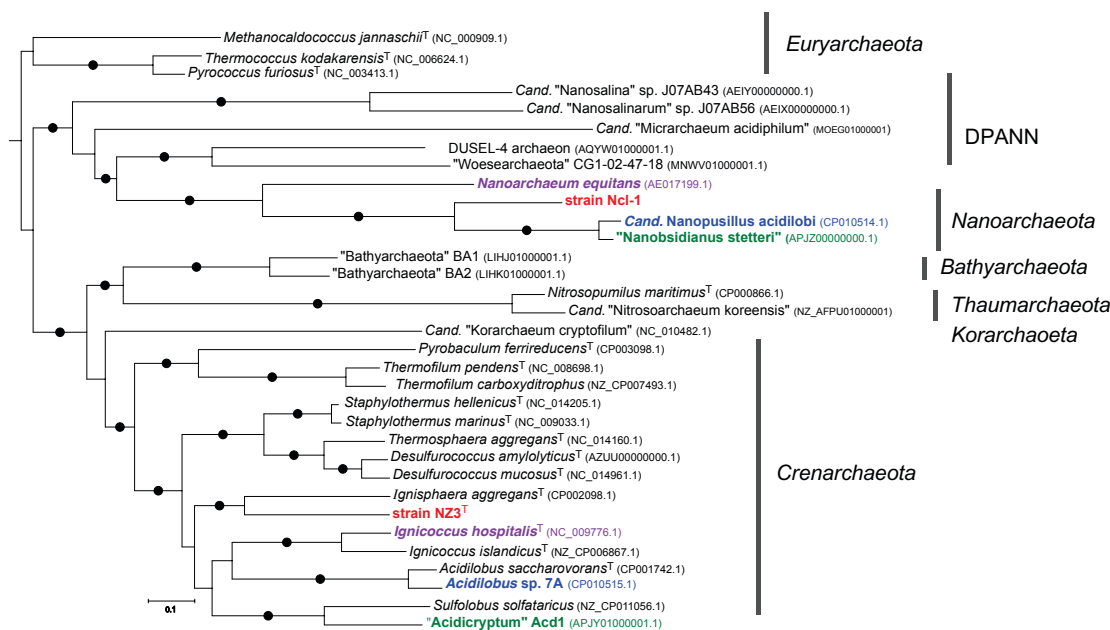


Figure 1.8. Phylogenetic analysis of 16 concatenated ribosomal proteins (rpL2, 3, 4, 5, 6, 15, 16, 18, 22, 24 and rpS3, 8, 10, 17, 19) using RAXML. The tree was rooted in the *Euryarchaeota*. Closed circles represent bootstrap support of $\geq 70\%$. The four different *Nanoarchaeota*-host systems are in red, blue, purple and green. The scale bar represents 0.1 amino acid substitutions per site.

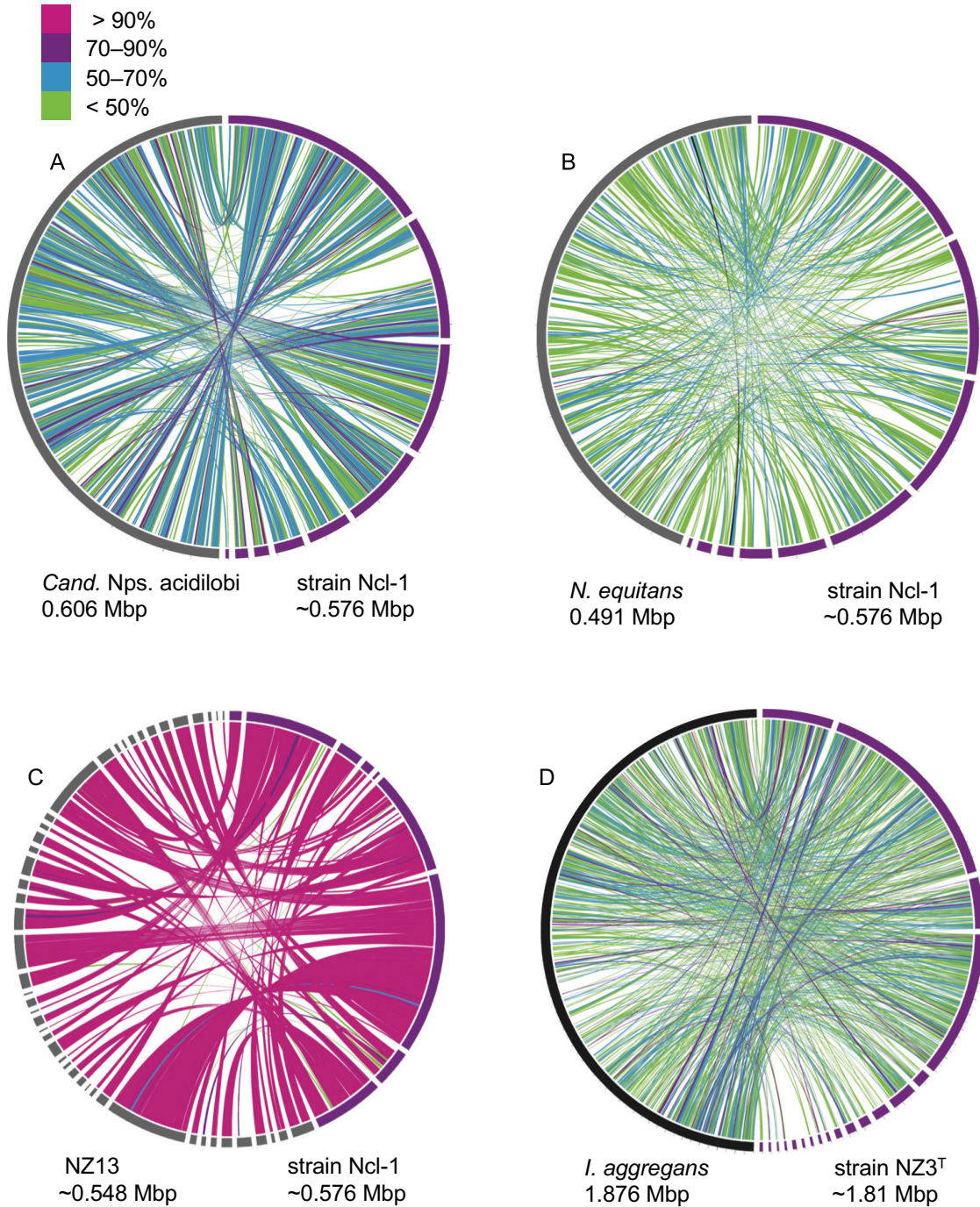


Figure 1.9. Circos-based genome alignments comparing overall protein homology between strain Ncl-1 and (A) *Cand. Nps. acidilobi*, (B) *N. equitans* and (C) the NZ13 draft genome, and between (D) strain NZ3^T and *Ignisphaera aggregans*. Colored lines link pairs of orthologs between genomes, with the color scaled to % amino acid similarity levels.

draft genome, although the PK gene also was not expressed in the proteome of *Cand.*

Nps. acidilobi [113]. As an alternative, the PK step may be catalyzed in both *Cand. Nps. acidilobi* and Ncl-1 by phosphoenolpyruvate synthase (PEPS, Ncl-1 451, 487), as reported in *Thermococcus kodakarensis* [10,38,113]. Nonetheless, the EMP glycolytic pathway is likely nonfunctional in either *Cand. Nps. acidilobi* or Ncl-1.

Most of the genes for gluconeogenesis, several which are bidirectional, are present in both terrestrial nanoarchaeotes (absent in *N. equitans*) and are expressed in the proteome of *Cand. Nps. acidilobi* [113]. It is therefore likely that gluconeogenesis can proceed to some extent in these two nanoarchaeotes and its primary role may be to synthesize activated sugars for protein and lipid glycosylation, as *Cand. Nps. acidilobi* has a glycosylated, lectin-reactive surface [113]. Although the glyceraldehyde-3-phosphate dehydrogenase/phosphoglycerate kinase (GAPDH/PGK) couple was not detected in either *Cand. Nps. acidilobi* or Ncl-1, Wurch et al, 2016 [113] proposed that its function may be replaced by a reversible ferredoxin-dependent aldehyde oxidoreductase (Ncl-1 615). Furthermore, while both terrestrial nanoarchaeotes have an acetyl-CoA synthase (Ncl-1 481, 606), it was not expressed in the *Cand. Nps. acidilobi* proteome under tested culture conditions [113], leaving it unclear whether *Cand. Nps. acidilobi* and Ncl-1 can convert acetate to acetyl-CoA for gluconeogenesis, or if acetyl-CoA is obtained from their hosts or other sources.

Strain Ncl-1 has two ABC-type multidrug transporter subunits (Ncl-1 256-257), two ABC-type antimicrobial peptide transporter subunits (Ncl-1 473, 476), and two gene copies of a cation transport ATPase (Ncl-1 290, 573) which all show strong homology to transporters reported for *Cand. Nps. acidilobi*. However, the major facilitator

superfamily permeases were not detected in the draft genome. In contrast, Ncl-1 has two Trk-system proteins involved in potassium uptake (Ncl-1 249-250) that have not previously been reported in the *Nanoarchaeota*. Both *Cand. Nps. acidilobi* and Ncl-1 have a highly reduced set of transporter genes, however, suggesting that transport proteins may not represent the primary mode of metabolite transfer between *Nanoarchaeota* and their hosts [113].

Strain Ncl-1 and *Cand. Nps. acidilobi* encode archaeal flagella using very similar gene arrangements (Fig. 1.10A-B), both containing the essential *fla* genes required for assembly and function [2], but lacking the traditional operon structure found in other *Archaea* (Fig. 1.10C). Like *Cand. Nps. acidilobi*, Ncl-1 has a *flaD/E* gene (Ncl-1 218) typical of *Euryarchaeota* [2]. TEM micrographs revealed that archaeal flagella genes are functional in strain Ncl-1 and expressed under cultivation conditions (Fig. 1.1C-E). Strain Ncl-1 is often seen with several (~5) lophotrichous archaeal flagella, both when associated (Fig. 1.1C) and detached from its host (Fig. 1.1D-E). If motile, the nanoarchaeotal flagella may play a role in host-seeking behavior, as proposed for marine obligate intracellular parasitic *Chlamydiae* [17], or they may allow the nanoarchaeote to detach and search for a new host [113]. Alternatively, the nanoarchaeotal flagella may mediate adhesion to surfaces [40,117] or attachments between the symbiont and its host [99]. However, the presence of archaeal flagella presents a conundrum for both Ncl-1 and *Cand. Nps. acidilobi*, as neither possesses any apparent means of producing ATP to fuel motility.

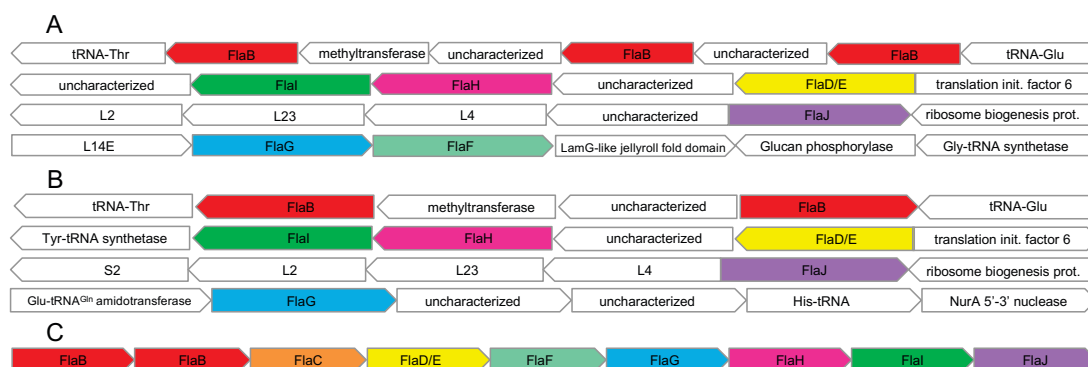


Figure 1.10. Comparison of the archaeal flagellar gene arrangement in (A) strain Ncl-1, (B) *Cand. Nps. acidilobi* and (C) *Pyrococcus furiosus*.

Strain Ncl-1 shares additional similarities with the central metabolic genes of *N. equitans* and *Cand. Nps. acidilobi*. Like all characterized nanoarchaeotes, Ncl-1 has a non-discriminating Glu-tRNA synthetase (Ncl-1 520) and both of the amidotransferase subunits required for transamidation of Glu-tRNA^{Gln} to Gln-tRNA^{Gln} (Ncl-1 159-160). Both *Cand. Nps. acidilobi* and Ncl-1 appear to synthesize Asn-tRNA^{Asn} directly via an asparaginyl-tRNA synthetase (Ncl-1 446), instead of using the transamidation reaction found in *N. equitans*. However, Ncl-1 encodes a class I lysyl-tRNA synthetase (Ncl-1 78) similar to marine *N. equitans*, thought to be the ancestral gene form [89], while *Cand. Nps. acidilobi* harbors the class II enzyme. Similar to *Cand. Nps. acidilobi*, Ncl-1 can also convert L-cysteine to L-alanine via a cysteine desulfurase (Ncl-1 421). Furthermore, Ncl-1 possesses some unique nucleotide biotransformation capabilities not found in either *Cand. Nps. acidilobi* or *N. equitans*, including a purine-nucleoside phosphorylase (Ncl-1 611), a CTP synthase (Ncl-1 590), and a nucleoside-triphosphatase THEP1 (Ncl-1 633).

The Ncl-1 draft genome contains 41 tRNAs, encoding for all 20 standard amino acids. Like all described nanoarchaeotes, Ncl-1 contains several *cis*-spliced tRNAs which are common throughout the *Archaea* [27,49,107,111,113]. Although some of the *cis*-spliced tRNA genes found in Ncl-1 are shared by other *Nanoarchaeota*, the Ncl-1 intron sequences are unique (Appendix B, File S1, Table A3). In contrast, two tRNA intron sequences identified in *Cand. Nps. acidilobi* and “N. stetteri” Nst1 are identical, suggesting the intron sequence were already present in the last common ancestor of these two closely-related nanoarchaeotes. Although the host strain NZ3^T also contains *cis*-spliced tRNA genes, none of the intron sequences found in the host match the corresponding intron sequences in the Ncl-1 nanoarchaeote, suggesting these introns were not transferred horizontally from host to *Nanoarchaeota*. The Ncl-1 tRNA intron sequences did not show any significant homology (E-value ≤ 0.001) to Group II introns in the Rfam database [44] or to the custom set of Group I introns provided by Nawrocki et al. [76]. However, one large tRNA intron sequence in the Ncl-1 draft genome (tRNA-Asp; 486 bp) contains a putative homing endonuclease domain (E-value 0.003) and complex secondary structure. Notably, this intronic sequence does not show significant homology (E-value ≤ 0.00001) to the putative homing endonuclease found in the host 16S rRNA gene intron, suggesting it was not acquired from the host.

The *trans*-spliced tRNAs identified in *N. equitans* [91,92] were not found in either Ncl-1 or *Cand. Nps. acidilobi*. Similarly, both the genomes of *Cand. Nps. acidilobi* and Ncl-1 encode for RNase P (inferred in Ncl-1 on the basis of RNase P protein subunits) while it was not identified in the genome of *N. equitans* [59,93]. The Ncl-1 draft genome has a common set of split protein coding genes, shared by all described nanoarchaeotes

(Table 1.4). Fragmented protein coding genes have also been described in some members of the *Euryarchaeota* [111]. Split genes provide further evidence that genome reduction in the *Nanoarchaeota* resulted in stochastic fragmentation, with genes often split in the same location [89]. Some unique split genes also occur across the nanoarchaeote genomes, which is not surprising given the phylogenetic diversity of the three cultivated nanoarchaeotes. Genes with conserved split regions likely represent ancestral fragmentation sites, present in the last common ancestor of extant nanoarchaeotes, while proteins split in some but not all *Nanoarchaeota* may be more recent splitting events.

Similar to its marine relative *N. equitans*, Ncl-1 has a putative CRISPR-Cas system, which is the first described CRISPR-Cas system for a terrestrial nanoarchaeote [72,89,113]. The system contains 43 spacer regions and a suite of Type 1 *cas* genes. Although several of the *cas* genes show strong homology to their counterparts in *N. equitans*, the gene arrangement is quite different (Fig. 1.11) and the system contains an additional *cas6* gene (Ncl-1 284) not identified in *N. equitans*. The target of the Ncl-1 CRISPR-Cas system is unclear as the spacer regions do not show significant homology (E-value ≤ 0.00001) to viral sequences in the NCBI nucleotide database or to the genome sequence of the host, strain NZ3^T. Although no CRISPR-Cas system has been previously reported in terrestrial nanoarchaeotes, a nanoarchaeotal draft genome from Nymph Lake, YNP, showed evidence of integrated prophage DNA [72], demonstrating that geographically distant nanoarchaeotes are vulnerable to viral infection.

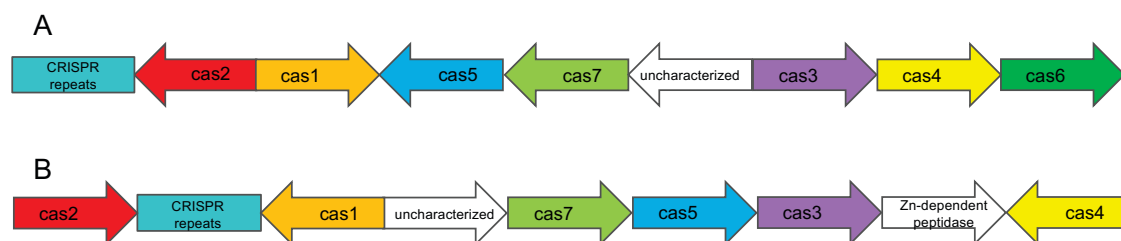


Figure 1.11. Gene arrangement of the CRISPR-Cas locus in (A) strain Ncl-1 and (B) *N. equitans*.

Metabolic insights from the draft genome of the host, NZ3^T

To explore the metabolic capabilities of NZ3^T further, the draft genome was reconstructed from low-diversity metagenomic data as previously described (Materials and Methods, pg. 37). The NZ3^T draft genome consists of 19 contigs, with a 41.9% G+C content and 97.5% completion based on CheckM (Table 1.3). Phylogenetic reconstruction using a set of 16 ribosomal proteins (Fig. 1.8) confirmed the placement of NZ3^T in the *Desulfurococcaceae*, in a deeply-branching clade with *I. aggregans*. The genome is 45% similar to *I. aggregans* based on AAI, which is consistent with the level of 16S rRNA gene sequence divergence between the two organisms [51]. The strain NZ3^T genome also shows modest levels of synteny to *I. aggregans* (Fig. 1.9D), and the two strains are similar in size (~1.808 Mbp and 1.876 Mbp, respectively) [29]. Forty-nine tRNA genes (19 of 20 standard amino acids), three rRNA genes and RNase P were also identified in the NZ3^T genome. Similar to *Fervidicoccus fontis*, another member of the *Thermoprotei*, NZ3^T appears to be auxotrophic for a variety of amino acids and coenzymes, such as biotin and coenzyme A [55].

Table 1.4. Comparison of genomic features in strain Ncl-1 and other *Nanoarchaeota*. *

Characteristic	Strain Ncl-1	<i>N. equitans</i>	“ <i>N. stetteri</i> ” Nst1	<i>Cand. Nps.</i> <i>acidilobi</i>	Split site
Split protein coding genes					
Reverse gyrase	Ncl-1 419-492	NEQ 318-434	Nst 337-402	Nps 2030-2705	Same
Glu-tRNA ^{Gln} amidotransferase subunit E	<i>Ncl-1 160</i>	NEQ 245-396	Nst 197-449	Nps 2280-3415	<i>Not split in Ncl-1</i>
Predicted RNA-binding protein	Ncl-1 37-631	NEQ 438-506	Nst 176-251	Nps 30-3140	Same
Archaeosine tRNA-guanine transglycosylase	Ncl-1 339-649	NEQ 124-305	Nst 096-232	Nps 1135-3235	Same
RNA polymerase subunit B	<i>Ncl-1 121</i>	NEQ 156-173	Nst 632-633	Nps 545-550	<i>Not split in Ncl-1</i>
Large helicase-related protein	Ncl-1 158-643	NEQ 003-409	Nst 172-239	Nps 745-3195	Different**
DNA polymerase I	Ncl-1 402	NEQ 068-528	Nst 417	Nps 2105	Not split in Ncl-1
Topoisomerase I	Ncl-1 39	NEQ 045-324	Nst 174	Nps 45	Not split in Ncl-1
P-loop ATPase acetyltransferase fusion prot.	Ncl-1 420	NEQ 096-495	Nst 401	Nps 2025	Not split in Ncl-1
Alanyl-tRNA synthetase	Ncl-1 77	NEQ 211-547	Nst 054	Nps 290	Not split in Ncl-1
Diphthamide synthase sub. DPH2	Ncl-1 5-382	Absent	Nst 222-440	Nps 3285-2235	Same
Uncharacterized prot. (arCOG04253)	Ncl-1 582	Absent	Nst 474-80	Nps 2410	Not split in Ncl-1
Glucosamine F6P aminotransferase	<i>Ncl-1 169-171</i>	Absent	Absent	Absent	<i>Unique to Ncl-1</i>
<i>Cis</i>-spliced tRNA genes	Ile Tyr, Met, <i>Leu</i> , <i>Asp</i>	Ile, Tyr, Met, Trp	Ile, Tyr	Ile, Tyr	
<i>Trans</i>-spliced tRNA genes	None	iMet, His, Lys, Gln, Glu (2)	None	None	
RNase P	Present	Absent	Present	Present	
Gluconeogenesis/glycolysis	Present (partial)	Absent	Present (partial)	Present (partial)	
Polyamine biosynthesis	Present	Absent	Present	Present	
ATP synthase complex	Absent	Incomplete	Absent	Absent	
Glutamate dehydrogenase	Present	Present	Absent	Absent	

Bold-italics indicate where the new Ncl-1 differs from other described *Nanoarchaeota*. F6P, fructose 6-phosphate.

*Modified from Podar et al, 2013 [83].

** *N. equitans* harbors a distinct split site; all three other *Nanoarchaeota* share a split site.

Consistent with its growth on peptide-rich substrates, the NZ3^T draft genome contains secreted and cytosolic peptidases and proteases, and a suite of transporters to import dipeptides, oligopeptides and amino acids into the cell. Once inside the cell, amino acid degradation likely proceeds using pyruvate:ferredoxin oxidoreductases (NZ3 1006-1009;1927-1930) and indolepyruvate:ferredoxin oxidoreductases (NZ3 279-280; 1116-1117) as described for members of the *Thermococcaceae* [33,62,63]. Breakdown of amino acids provides NZ3^T with limited ATP and a pool of reduced ferredoxin molecules.

The NZ3^T genome has incomplete gluconeogenesis and EMP glycolysis pathways, and the Entner-Doudoroff glycolysis pathway and pentose phosphate pathway were not detected. Possibly, NZ3^T may utilize an archaeal type III Rubisco (NZ3 708, 1082) to generate 3-phosphoglycerate (3-PGA) as proposed by Sato et al, 2007 [97]. 3-PGA could then be converted to acetate using enzymes involved in glycolysis and amino acid fermentation, generating ATP for the cell.

Despite its dependence on thiosulfate as a terminal electron acceptor, no thiosulfate reductase gene was identified in the NZ3^T draft genome. However, it is possible that the gene may be encoded in one of the unsequenced portions of the genome. Alternately, NZ3^T may utilize a divergent gene for thiosulfate reduction that has yet to be identified. Consistent with its lifestyle as an anaerobe, no cytochrome c oxidase or cytochrome *bd* ubiquinol oxidase genes were found in the draft genome of NZ3^T.

Strain NZ3^T has several different hydrogenases, including two membrane-bound [NiFe] group 4b hydrogenase clusters which evolve molecular hydrogen [31]. One of these clusters contains a putative catalytic carbon monoxide dehydrogenase subunit *cooS*

(NZ3 969), suggesting that NZ3^T may be able to conserve energy through carboxydrotrophic hydrogenogenesis (coupling CO oxidation and proton reduction) in a similar manner to that of *Thermococcus onnurineus* and *Thermococcus* sp. AM4 [56,102,103]. The NZ3^T *cooS* gene shows strong phylogenetic relatedness to *cooS* genes found in known carboxydrotrophs such as *Thermococcus barophilus* (Genbank NC_014804.1, [110]) and *Thermococcus* sp. AM4 (Genbank CP002952.1, [81]) (Fig. 1.12). However, direct cultivation studies under a CO atmosphere will be required to determine whether or not CO oxidation is functional in strain NZ3^T. The second [NiFe] group 4b hydrogenase cluster (NZ3 314-329) in the draft genome encodes a multi-subunit Na⁺/H⁺ antiporter which may facilitate the formation of a sodium ion gradient across the membrane [31]. This sodium gradient may be used to generate ATP using an archaeal-type ATP synthase (NZ3 1842-1845, 1847-1849), as reported for *Desulfurococcus kamchatkensis*, *Pyrococcus furiosus*, and *Methanosarcina acetivorans* [88,94,98]. The NZ3^T genome also encodes a [NiFe] group 3A hydrogenase cluster containing a coenzyme F₄₂₀-reducing hydrogenase (NZ3 1861) similar to those found in methanogens and *T. onnurineus* [31,42,56,57]. Phylogenetic analysis clusters the NZ3^T gene with coenzyme F₄₂₀-reducing hydrogenases from *T. onnurineus* and other non-methanogenic lineages with high confidence (Fig. 1.13). To our knowledge, this represents the first example of a coenzyme F₄₂₀-reducing hydrogenase reported in the *Crenarchaeota*. As suggested for *T. onnurineus*, this hydrogenase likely does not utilize coenzyme F₄₂₀, since there is no genomic evidence that either organism is capable of synthesizing the coenzyme [42,57]. While its function in non-methanogens remains unclear, transcriptomic data from *T. onnurineus* indicates that the coenzyme F₄₂₀-reducing

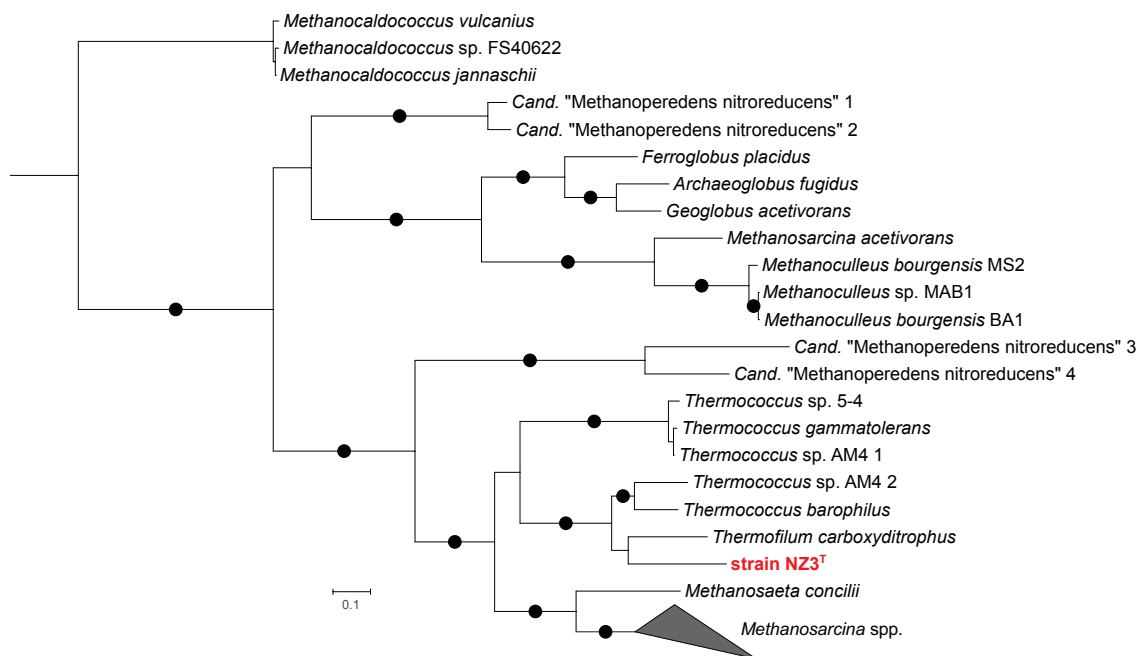


Figure 1.12. Maximum likelihood phylogenetic tree of archaeal *cooS* genes. Filled circles represent $\geq 70\%$ bootstrap support. Scale bar indicates 0.1 amino acid substitutions per site.

hydrogenase impacts the expression of CO oxidation genes, potentially pointing to a larger role in central metabolism [57].

Genomic evidence suggests that NZ3^T encodes genes for synthesis of an S-layer (NZ3 717), similar to *Acidilobus* sp. 7A, the host of terrestrial *Cand. Nps. acidilobi* [113]. This is in contrast with *I. hospitalis*, the host of *N. equitans*, which has a distinctive double membrane system and lacks an S-layer [83,90]. Based on a recent study using electron tomography, the double membrane structure in *I. hospitalis* appears to play a crucial role in its association with *N. equitans*, as the host and symbiont form a cytoplasmic bridge at their attachment site [90]. A similar pore-like structure was detected between NZ3^T and its nanoarchaeotal symbiont at their attachment site, with no membrane clearly visible between the cells (Fig. 1.1F) and suggests an equally intimate

association between the terrestrial nanoarchaeote and its host. However, given the extreme differences in host outer membrane structure, it is likely that terrestrial nanoarchaeotes have evolved a unique mechanism of recognizing, attaching, and interacting with their hosts.

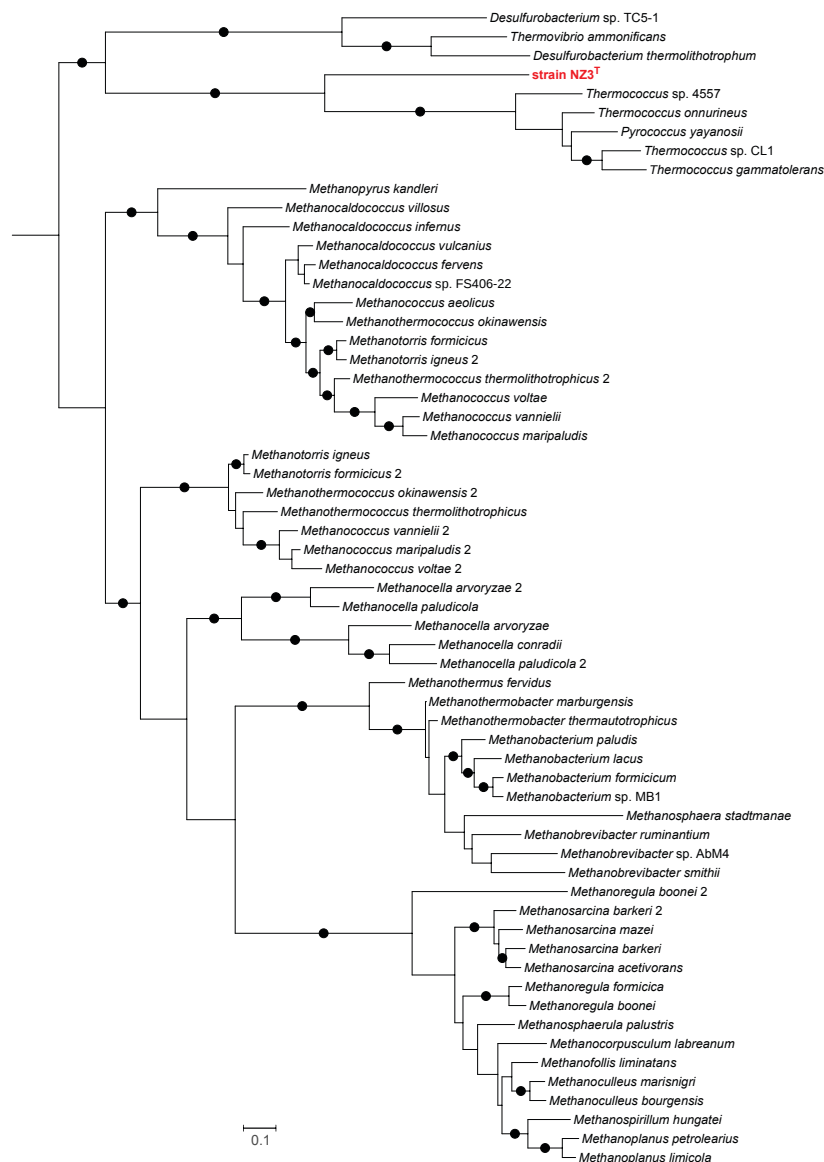


Figure 1.13. Maximum likelihood phylogenetic analysis of coenzyme F₄₂₀-reducing hydrogenases. Enzyme sequences were obtained from the HydDB web server [104]. The scale shows 0.1 substitutions per amino acid, and filled circles represent ≥70% bootstrap support.

Proposal of a novel Candidatus taxon in the Nanoarchaeota and a novel genus and species in the Desulfurococcaceae

The New Zealand nanoarchaeote strain Ncl-1 is distinct from its described relatives at both a phylogenetic and a genomic level. While 16S rRNA gene and concatenated ribosomal protein phylogenetic trees place Ncl-1 in the *Nanoarchaeota*, the 16S rRNA gene of the strain is divergent from its described relatives (~84-88% gene sequence similarity). AAI analysis also suggests genus-level divergence between Ncl-1 and the cultivated taxa *Cand. Nps. acidilobi* and *N. equitans* (45-52%). Although the genome of Ncl-1 shows stronger homology to its terrestrial relatives, it contains several genes not previously found in the *Nanoarchaeota*. Ncl-1 is also the first cultivated *Nanoarchaeota* from a New Zealand geothermal spring, and its host is a new crenarchaeotal genus. The clear phylogenetic divergence between Ncl-1 and other described nanoarchaeotes suggests that it represents a novel genus and species in the *Nanoarchaeota*. The formal proposal of a new *Candidatus* taxon in the *Nanoarchaeota*, *Candidatus* Nanoclepta minutus, is given in Appendix B, File S1, Table A4, with the Taxonumber CA00029.

Strain NZ3^T forms a clade in the *Desulfurococcaceae* with *Ignisphaera aggregans*. However, NZ3^T is only ~92% similar to *I. aggregans* by 16S rRNA gene similarity and 45% similar by AAI. Unlike *I. aggregans*, NZ3^T requires thiosulfate as a terminal electron acceptor and grows well on several proteinaceous substrates but cannot use any sugars for growth. Moreover, NZ3^T grows optimally between 82-83°C, which is much lower than the optimal temperature (92-95°C) of *I. aggregans*. Given these physiological and phylogenetic differences, we propose that strain NZ3^T represents a new

species of a novel genus in the *Desulfurococcaceae*. The formal proposal of the novel genus *Zestosphaera* gen. nov. and the species *Zestosphaera tikiterensis* sp. nov. is provided in Appendix B, File S1, Table A4, with the Taxonumbers GA00064 and TA00512, respectively.

Conclusions

Here we expand on nanoarchaeote-host symbiotic systems by describing a fourth association. A stable co-culture containing a novel nanoarchaeote and its host, a new genus in the *Desulfurococcaceae*, was cultivated from a geothermal hot spring in New Zealand. The New Zealand nanoarchaeote closely resembles other terrestrial nanoarchaeotes, with the presence of archaeal flagella and no apparent ATP synthase genes, but it also contains a CRISPR-Cas system like its marine relative *N. equitans*. The fragmentation of archaeal flagellar genes and the presence of shared split protein-coding genes distributed among marine and terrestrial nanoarchaeotes provide further evidence that genome reduction and fragmentation in this lineage began before the marine and terrestrial taxa diverged. The mechanisms by which phylogenetically different hosts interact with their nanoarchaeotal symbionts remain elusive. However, as more nanoarchaeote-host systems are characterized, additional insights will emerge about the cell-cell interactions in these symbioses, and whether host metabolism and ultrastructure have played roles in the evolutionary history of the symbiosis.

The New Zealand *Nanoarchaeota*-host system also points to the challenges associated with describing a cultivated organism that is only viable in the presence of its host. Under the current system of nomenclature, regulated by the International Code of

Nomenclature of *Bacteria*, a pure culture is required for the valid naming and recognition of a novel taxon [50,54]. However, no provision is made for obligate symbionts that are cultivated in the minimum consortium needed for growth (co-culture). For some intracellular pathogens, such as *Lawsonia* [68], *Diplorickettsia* [70], and members of the *Chlamydiae* [28,30], valid names may be assigned if the strain is cultivable in cell culture. However, this option is not available for many environmental taxa requiring a specific host for cultivation, such as ectosymbiotic *Nanoarchaeota*, or the diverse epibiotic *Archaea* and *Bacteria* that colonize ciliates [12,23,101]. While the *Candidatus* designation provides a means of identifying these novel lineages, it is a provisional status and implies that the organism will be isolated in the future, which is not feasible for obligate symbionts, parasites, or some epibionts [50]. Ultra-small genomes recovered from the archaeal DPANN superphylum [6,95] and bacterial Candidate Phyla Radiation [18] suggest a large diversity of yet-uncultivated microbial taxa that may rely on hosts for survival. Looking ahead, as more of these lineages are discovered and cultivated, the coherent classification of symbionts will require a greater acceptance and standardization of the *Candidatus* designation, or a novel system of symbiont nomenclature. Further complicating matters are names that have become adopted as validly described through repeated use such as *Nanoarchaeum equitans*.

Acknowledgements

We would like to thank Dr. Aharon Oren for providing correct etymology for our novel taxa, Dr. Kenneth Stedman for initial help with TEM, Dr. Richard Davis, John Donaho and Nicole Wagner for assistance in the bioinformatics, Kristen Brileya for developing the qPCR primers, Kira Makarova for guidance with protein annotation, and Stefan L'Haridon for many helpful discussions. We also thank Steve Allman, Zamin Yang and Dawn Klingeman from Oak Ridge National Laboratory (ORNL) for assistance with cell sorting, molecular biology techniques and MiSeq sequencing. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. This research was funded through grants from the National Science Foundation (DEB1134877), the National Aeronautics and Space Administration (NNX16AJ66G), and the Geothermal Resources of New Zealand (GRN) Program at GNS Science. We acknowledge and thank Tikitere Trust for their enthusiasm for our research and assistance in access and sampling of the Hell's Gate geothermal features. TEM data acquisition was performed at the Portland State University Center for Electron Microscopy and Nanofabrication (CEMN), which is generously supported by the NSF, Murdock Charitable Trust, ONAMI, and FEI (currently Thermo Fisher Scientific).

Contributions

Sampling was performed by ALR and MBS, and initial enrichment cultures were grown by ALR and YL. Negative staining for TEM was done by ALR, and ZC obtained TEM images. The NZ *Nanoarchaeota*-host system was isolated by YL. *Z. tikiterensis* was isolated by ESJ, who also identified the optimal pH and temperature, doubling time and electron acceptor use. YL determined carbon source usage, NaCl sensitivity, and temperature and pH minima and maxima for *Z. tikiterensis*. ESJ routinely extracted genomic DNA, monitored *Nanoarchaeota* numbers by qPCR, validated the purity of the *Nanoarchaeota*-host co-culture and the *Z. tikiterensis* isolate culture and validated the presence of 16S rRNA introns in *Z. tikiterensis*. KL and KM extracted and identified the polar and intact lipids of *Z. tikiterensis*. iTAG sequencing and analysis and sorting, screening and amplification of single cells was done by ALR and MP. Assembly, trimming, and binning was performed by JM. ESJ performed manual curation and final binning in ESOM, followed by analysis of the coding potential of the *Nanoarchaeota* and its host. ESJ generated ribosomal protein and 16S rRNA gene trees, average amino acid identity analyses, Circos synteny plots and intron and gene arrangement diagrams. The manuscript was prepared by ALR and ESJ.

References

- [1] Aagaard, C., Dalgaard, J.Z., Garrett, R.A. (1995) Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron- cells of *Sulfolobus acidocaldarius*. Proc. Natl. Acad. Sci. U. S. A. 92(26), 12285–12289.
- [2] Albers, S.-V., Jarrell, K.F. (2015) The archaeellum: how *Archaea* swim. Front. Microbiol. 6, 23, <https://doi.org/10.3389/fmicb.2015.00023>.
- [3] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J. Mol. Biol. 215(3), 403–410.
- [4] Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S., Banfield, J.F. (2016) Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219, <https://doi.org/10.1038/ncomms13219>.
- [5] Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75, <https://doi.org/10.1186/1471-2164-9-75>.
- [6] Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., Verberkmoes, N.C., Hettich, R.L., Banfield, J.F. (2010) Enigmatic,

- ultrasmall, uncultivated *Archaea*. Proc. Natl. Acad. Sci. U. S. A. 107(19), 8806–8811.
- [7] Baker, B.J., Saw, J.H., Lind, A.E., Lazar, C.S., Hinrichs, K.-U., Teske, A.P., Ettema, T.J.G. (2016) Genomic inference of the metabolism of cosmopolitan subsurface *Archaea*, *Hadesarchaea*. Nat Microbiol 1, 16002, <https://doi.org/10.1038/nmicrobiol.2016.2>.
- [8] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19(5), 455–477.
- [9] Bligh, E.G., Dyer, W.J. (1959) A rapid method of total lipid extraction and purification. Can. J. Biochem. Physiol. 37(8), 911–917.
- [10] Bräsen, C., Esser, D., Rauch, B., Siebers, B., Brasen, C., Esser, D., Rauch, B., Siebers, B. (2014) Carbohydrate metabolism in *Archaea*: current insights into unusual enzymes and pathways and their regulation. Microbiol. Mol. Biol. Rev. 78(1), 89–175.
- [11] Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., Olson, R., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomason, J.A., Stevens, R., Vonstein, V., Wattam, A.R., Xia, F. (2015) RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, <https://doi.org/10.1038/srep08365>.
- [12] Bright, M., Espada-Hinojosa, S., Lagkouvardos, I., Volland, J.-M. (2014) The

giant ciliate *Zoothamnium niveum* and its thiotrophic epibiont *Candidatus*

Thiobios zoothamnicoli: a model system to study interspecies cooperation. *Front. Microbiol.* 5, 145, <https://doi.org/10.3389/fmicb.2014.00145>.

- [13] Burggraf, S., Larsen, N., Woese, C.R., Stetter, K.O. (1993) An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci. U. S. A.* 90(6), 2547–2550.
- [14] Casanueva, A., Galada, N., Baker, G.C., Grant, W.D., Heaphy, S., Jones, B., Yanhe, M., Ventosa, A., Blamey, J., Cowan, D.A. (2008) Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* 12, 651–656.
- [15] Chernyh, N.A., Mardanov, A. V., Gumerov, V.M., Miroshnichenko, M.L., Lebedinsky, A. V., Merkel, A.Y., Crowe, D., Pimenov, N. V., Rusanov, I.I., Ravin, N. V., Moran, M.A., Bonch-Osmolovskaya, E.A. (2015) Microbial life in Bourlyashchy, the hottest thermal pool of Uzon Caldera, Kamchatka. *Extremophiles* 19(6), 1157–1171.
- [16] Clingenpeel, S., Kan, J., Macur, R.E., Woyke, T., Lovalvo, D., Varley, J., Inskeep, W.P., Nealson, K., McDermott, T.R. (2013) Yellowstone Lake *Nanoarchaeota*. *Front. Microbiol.* 4, 274, <https://doi.org/10.3389/fmicb.2013.00274>.
- [17] Collingro, A., Köstlbacher, S., Musmann, M., Stepanauskas, R., Hallam, S.J., Horn, M. (2017) Unexpected genomic features in widespread intracellular *Bacteria*: evidence for motility of marine *Chlamydiae*. *ISME J.* 11(10), 2334–2344.
- [18] Danczak, R.E., Johnston, M.D., Kenah, C., Slattery, M., Wrighton, K.C., Wilkins,

- M.J. (2017) Members of the Candidate Phyla Radiation are functionally differentiated by carbon- and nitrogen-cycling capabilities. *Microbiome* 5(1), 112, <https://doi.org/10.1186/s40168-017-0331-1>.
- [19] Darling, A.E., Jospin, G., Lowe, E., Matsen 4th, F.A., Bik, H.M., Eisen, J.A. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2, e243, <https://doi.org/10.7717/peerj.243>.
- [20] Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797.
- [21] Elkins, J.G., Podar, M., Graham, D.E., Makarova, K.S., Wolf, Y., Randau, L., Hedlund, B.P., Brochier-Armanet, C., Kunin, V., Anderson, I., Lapidus, A., Goltsman, E., Barry, K., Koonin, E. V., Hugenholtz, P., Kyrpides, N., Wanner, G., Richardson, P., Keller, M., Stetter, K.O. (2008) A korarchaeal genome reveals insights into the evolution of the *Archaea*. *Proc. Natl. Acad. Sci. U. S. A.* 105(23), 8102–8107.
- [22] Elling, F.J., Becker, K.W., Könneke, M., Schröder, J.M., Kellermann, M.Y., Thomm, M., Hinrichs, K.-U. (2016) Respiratory quinones in *Archaea*: phylogenetic distribution and application as biomarkers in the marine environment. *Environ. Microbiol.* 18(2), 692–707.
- [23] Filker, S., Kaiser, M., Rosselló-Móra, R., Dunthorn, M., Lax, G., Stoeck, T. (2014) "*Candidatus* Haloectosymbiotes riaformosensis" (*Halobacteriaceae*), an archaeal ectosymbiont of the hypersaline ciliate *Platynematum salinarum*. *Syst. Appl. Microbiol.* 37(4), 244–251.
- [24] Flores, G.E., Campbell, J.H., Kirshtein, J.D., Meneghin, J., Podar, M., Steinberg,

- J.I., Seewald, J.S., Tivey, M.K., Voytek, M.A., Yang, Z.K., Reysenbach, A.L. (2011) Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environ. Microbiol.* 13(8), 2158–2171.
- [25] Flores, G.E., Hunter, R.C., Liu, Y., Mets, A., Schouten, S., Reysenbach, A.-L. (2012) *Hipaea jasoniae* sp. nov. and *Hipaea alviniae* sp. nov., thermoacidophilic members of the class *Deltaproteobacteria* isolated from deep-sea hydrothermal vent deposits. *Int. J. Syst. Evol. Microbiol.* 62(Pt 6), 1252–1258.
- [26] Flores, G.E., Shakya, M., Meneghin, J., Yang, Z.K., Seewald, J.S., Geoff Wheat, C., Podar, M., Reysenbach, A.-L. (2012) Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10(4), 333–346.
- [27] Fujishima, K., Sugahara, J., Tomita, M., Kanai, A. (2010) Large-scale tRNA intron transposition in the archaeal order *Thermoproteales* represents a novel mechanism of intron gain. *Mol. Biol. Evol.* 27(10), 2233–2243.
- [28] Fukushi, H., Hirai, K. (1992) Proposal of *Chlamydia pecorum* sp. nov. for *Chlamydia* strains derived from ruminants. *Int. J. Syst. Bacteriol.* 42(2), 306–308.
- [29] Göker, M., Held, B., Lapidus, A., Nolan, M., Spring, S., Yasawong, M., Lucas, S., Glavina Del Rio, T., Tice, H., Cheng, J.-F., Goodwin, L., Tapia, R., Pitluck, S., Liolios, K., Ivanova, N., Mavromatis, K., Mikhailova, N., Pati, A., Chen, A., Palaniappan, K., Brambilla, E., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C.D., Brettin, T., Detter, J.C., Han, C., Rohde, M., Sikorski, J., Woyke, T., Bristow, J., Eisen, J.A., Markowitz, V., Hugenholtz, P., Kyrpides, N.C., Klenk, H.-P. (2010)

Complete genome sequence of *Ignisphaera aggregans* type strain (AQ1.S1^T).

Stand. Genomic Sci. 3(1), 66–75.

- [30] Grayston, J.T., Kuo, C.-C., Campbell, L.A., Wang, S.-P. (1989) *Chlamydia pneumoniae* sp. nov. for *Chlamydia* sp. strain TWAR. Int. J. Syst. Evol. Microbiol. 39(1), 88–90.
- [31] Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., Cook, G.M., Morales, S.E. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. ISME J. 10(3), 761–777.
- [32] Hamilton-Brehm, S.D., Gibson, R.A., Green, S.J., Hopmans, E.C., Schouten, S., van der Meer, M.T.J., Shields, J.P., Damsté, J.S.S., Elkins, J.G. (2013) *Thermodesulfobacterium geofontis* sp. nov., a hyperthermophilic, sulfate-reducing bacterium isolated from Obsidian Pool, Yellowstone National Park. Extremophiles 17(2), 251–263.
- [33] Heider, J., Ma, K., Adams, M.W. (1995) Purification, characterization, and metabolic function of tungsten-containing aldehyde ferredoxin oxidoreductase from the hyperthermophilic and proteolytic archaeon *Thermococcus* strain ES-1. J. Bacteriol. 177(16), 4757–4764.
- [34] Hohn, M.J., Hedlund, B.P., Huber, H. (2002) Detection of 16S rDNA sequences representing the novel phylum “*Nanoarchaeota*”: Indication for a wide distribution in high temperature biotopes. Syst. Appl. Microbiol. 25, 551–554.
- [35] Huber, H., Burggraf, S., Mayer, T., Wyschkony, I., Rachel, R., Stetter, K.O. (2000) *Ignicoccus* gen. nov., a novel genus of hyperthermophilic,

- chemolithoautotrophic *Archaea*, represented by two new species, *Ignicoccus islandicus* sp. nov. and *Ignicoccus pacificus* sp. nov. *Int. J. Syst. Evol. Microbiol.* 50(6), 2093–2100.
- [36] Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002) A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* 417(6884), 63–67.
- [37] Huber, R., Dyba, D., Huber, H., Burggraf, S., Rachel, R. (1998) Sulfur-inhibited *Thermosphaera aggregans* sp. nov., a new genus of hyperthermophilic archaea isolated after its prediction from environmentally derived 16S rRNA sequences. *Int. J. Syst. Evol. Microbiol.* 48(1), 31–38.
- [38] Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H., Imanaka, T. (2006) Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.* 61(4), 898–909.
- [39] Itoh, T., Suzuki, K., Nakase, T. (1998) Occurrence of introns in the 16S rRNA genes of members of the genus *Thermoproteus*. *Arch. Microbiol.* 170(3), 155–161.
- [40] Jarrell, K.F., Stark, M., Nair, D.B., Chong, J.P.J. (2011) Flagella and pili are both necessary for efficient attachment of *Methanococcus maripaludis* to surfaces. *FEMS Microbiol. Lett.* 319(1), 44–50.
- [41] Jay, Z.J., Inskeep, W.P. (2015) The distribution, diversity, and importance of 16S rRNA gene introns in the order *Thermoproteales*. *Biol. Direct* 10, 35, <https://doi.org/10.1186/s13062-015-0065-6>.
- [42] Jeon, J.H., Lim, J.K., Kim, M.-S., Yang, T.-J., Lee, S.-H., Bae, S.S., Kim, Y.J.,

- Lee, S.H., Lee, J.-H., Kang, S.G., Lee, H.S. (2015) Characterization of the frhAGB-encoding hydrogenase from a non-methanogenic hyperthermophilic archaeon. *Extremophiles* 19(1), 109–118.
- [43] Jochimsen, B., Peinemann-Simon, S., Völker, H., Stüben, D., Botz, R., Stoffers, P., Dando, P.R., Thomm, M. (1997) *Stetteria hydrogenophila*, gen. nov. and sp. nov., a novel mixotrophic sulfur-dependent crenarchaeote isolated from Milos, Greece. *Extremophiles* 1(2), 67–73.
- [44] Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., Petrov, A.I. (2018) Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 46(D1), D335–D342.
- [45] Kalvari, I., Nawrocki, E.P., Argasinska, J., Quinones-Olvera, N., Finn, R.D., Bateman, A., Petrov, A.I. (2018) Non-coding RNA analysis using the Rfam database. *Curr. Protoc. Bioinforma.* 62(1), e51, <https://doi.org/10.1002/cpbi.51>.
- [46] Kears, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12), 1647–1649.
- [47] Kelley, J. 2017. Expanding Metabolic Diversity of Two Archaeal Phyla: *Nanoarchaeota* and *Korarchaeota*. M. Sc. Thesis, Biology Department, Portland State University, Portland, OR, USA.
- [48] Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon,

- Y.S., Lee, J.-H., Yi, H., Won, S., Chun, J. (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int. J. Syst. Evol. Microbiol.* 62(Pt 3), 716–721.
- [49] Kjems, J., Leffers, H., Olesen, T., Garrett, R.A. (1989) A unique tRNA intron in the variable loop of the extreme thermophile *Thermofilum pendens* and its possible evolutionary implications. *J. Biol. Chem.* 264(30), 17834–17837.
- [50] Konstantinidis, K.T., Rosselló-Móra, R. (2015) Classifying the uncultivated microbial majority: a place for metagenomic data in the *Candidatus* proposal. *Syst. Appl. Microbiol.* 38(4), 223–230.
- [51] Konstantinidis, K.T., Rosselló-Móra, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J.* 11(11), 2399–2406.
- [52] Krueger, F. Trim Galore! Babraham Bioinformatics.
https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. (Accessed 2016).
- [53] Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.* 19(9), 1639–1645.
- [54] Lapage, S.P., Lessel, E.F., Skerman, V.B.D., Clark, W.A., Seeliger, H.P.R. 1992. *International Code of Nomenclature of Bacteria. 1990 Revision.* ASM Press, Washington, D.C.
- [55] Lebedinsky, A. V., Mardanov, A. V., Kublanov, I. V., Gumerov, V.M., Beletsky, A. V., Perevalova, A.A., Bidzhieva, S.K., Bonch-Osmolovskaya, E.A., Skryabin, K.G., Ravin, N. V. (2014) Analysis of the complete genome of *Fervidococcus*

fontis confirms the distinct phylogenetic position of the order *Fervidicoccales* and suggests its environmental function. *Extremophiles* 18(2), 295–309.

- [56] Lee, H.S., Kang, S.G., Bae, S.S., Lim, J.K., Cho, Y., Kim, Y.J., Jeon, J.H., Cha, S.-S., Kwon, K.K., Kim, H.-T., Park, C.-J., Lee, H.-W., Kim, S. Il., Chun, J., Colwell, R.R., Kim, S.-J., Lee, J.-H. (2008) The complete genome sequence of *Thermococcus onnurineus* NA1 reveals a mixed heterotrophic and carboxydrotrophic metabolism. *J. Bacteriol.* 190(22), 7491–7499.
- [57] Lee, S.H., Kim, M.-S., Kim, Y.J., Kim, T.W., Kang, S.G., Lee, H.S. (2017) Transcriptomic profiling and its implications for the H₂ production of a non-methanogen deficient in the *frh*AGB-encoding hydrogenase. *Appl. Microbiol. Biotechnol.* 101(12), 5081–5088.
- [58] Letunic, I., Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44(W1), W242–W245.
- [59] Li, Y., Altman, S. (2004) In search of RNase P RNA from microbial genomes. *RNA* 10(10), 1533–1540.
- [60] Lowe, T.M., Chan, P.P. (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44(W1), W54–W57.
- [61] Lykke-Andersen, J., Aagaard, C., Semionenkov, M., Garrett, R.A. (1997) Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci.* 22(9), 326–331.
- [62] Ma, K., Hutchins, A., Sung, S.J., Adams, M.W. (1997) Pyruvate ferredoxin

- oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. Proc. Natl. Acad. Sci. U. S. A. 94(18), 9608–9613.
- [63] Mai, X., Adams, M.W. (1996) Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. J. Bacteriol. 178(20), 5897–5903.
- [64] Makarova, K., Wolf, Y., Koonin, E. (2015) Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. Life 5(1), 818–840.
- [65] Makarova, K.S., Sorokin, A. V., Novichkov, P.S., Wolf, Y.I., Koonin, E. V. (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. Biol. Direct, 33, <https://doi.org/10.1186/1745-6150-2-33>.
- [66] Marchler-Bauer, A., Bryant, S.H. (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32(Web Server issue), W327–W331.
- [67] McCliment, E.A., Voglesonger, K.M., O'Day, P.A., Dunn, E.E., Holloway, J.R., Cary, S.C. (2006) Colonization of nascent, deep-sea hydrothermal vents by a novel archaeal and nanoarchaeal assemblage. Environ. Microbiol. 8(1), 114–125.
- [68] McOrist, S., Gebhart, C.J., Boid, R., Barns, S.M. (1995) Characterization of *Lawsonia intracellularis* gen. nov., sp. nov., the obligately intracellular bacterium of porcine proliferative enteropathy. Int. J. Syst. Bacteriol. 45(4), 820–825.
- [69] Meador, T.B., Gagen, E.J., Loscar, M.E., Goldhammer, T., Yoshinaga, M.Y., Wendt, J., Thomm, M., Hinrichs, K.-U. (2014) *Thermococcus kodakarensis*

modulates its polar membrane lipids and elemental composition according to growth stage and phosphate availability. *Front. Microbiol.* 5, 10, <https://doi.org/10.3389/fmicb.2014.00010>.

- [70] Mediannikov, O., Sekeyová, Z., Birg, M.-L., Raoult, D. (2010) A novel obligate intracellular gamma-proteobacterium associated with ixodid ticks, *Diplorickettsia massiliensis*, gen. nov., sp. nov. *PLoS One* 5(7), e11478, <https://doi.org/10.1371/journal.pone.0011478>.
- [71] Menzel, P., Gudbergisdóttir, S.R., Rike, A.G., Lin, L., Zhang, Q., Contursi, P., Moracci, M., Kristjansson, J.K., Bolduc, B., Gavrilov, S., Ravin, N., Mardanov, A., Bonch-Osmolovskaya, E., Young, M., Krogh, A., Peng, X. (2015) Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70(2), 411–424.
- [72] Munson-McGee, J.H., Field, E.K., Bateson, M., Rooney, C., Stepanauskas, R., Young, M.J. (2015) *Nanoarchaeota*, their *Sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone National Park hot springs. *Appl. Environ. Microbiol.* 81(22), 7860–7868.
- [73] Nakagawa, T., Ishibashi, J.-I., Maruyama, A., Yamanaka, T., Morimoto, Y., Kimura, H., Urabe, T., Fukui, M. (2004) Analysis of dissimilatory sulfite reductase and 16S rRNA gene fragments from deep-sea hydrothermal sites of the Suiyo Seamount, Izu-Bonin Arc, western Pacific. *Appl. Environ. Microbiol.* 70(1), 393–403.
- [74] Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., Finn, R.D. (2015) Rfam 12.0:

- updates to the RNA families database. *Nucleic Acids Res.* 43(D1), D130–D137.
- [75] Nawrocki, E.P., Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22), 2933–2935.
- [76] Nawrocki, E.P., Jones, T.A., Eddy, S.R. (2018) Group I introns are widespread in *Archaea*. *Nucleic Acids Res.* 46(15), 7970–7976.
- [77] NCBI Resource Coordinators. (2017) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 45(D1), D12–D17.
- [78] Niederberger, T.D., Götz, D.K., McDonald, I.R., Ronimus, R.S., Morgan, H.W. (2006) *Ignisphaera aggregans* gen. nov., sp. nov., a novel hyperthermophilic crenarchaeote isolated from hot springs in Rotorua and Tokaanu, New Zealand. *Int. J. Syst. Evol. Microbiol.* 56(Pt 5), 965–971.
- [79] Nishihara, M., Koga, Y. (1987) Extraction and composition of polar lipids from the archaebacterium, *Methanobacterium thermoautotrophicum*: effective extraction of tetraether lipids by an acidified solvent. *J. Biochem.* 101(4), 997–1005.
- [80] Nomura, N., Sako, Y., Uchida, A. (1998) Molecular characterization and postsplicing fate of three introns within the single rRNA operon of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *J. Bacteriol.* 180(14), 3635–3643.
- [81] Oger, P., Sokolova, T.G., Kozhevnikova, D.A., Chernyh, N.A., Bartlett, D.H., Bonch-Osmolovskaya, E.A., Lebedinsky, A. V. (2011) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus* sp. strain AM4, capable of organotrophic growth and growth at the expense of hydrogenogenic or

sulfidogenic oxidation of carbon monoxide. *J. Bacteriol.* 193(24), 7019–7020.

- [82] Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R. (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42(Database issue), D206–D214.
- [83] Paper, W., Jahn, U., Hohn, M.J., Kronner, M., Näther, D.J., Burghardt, T., Rachel, R., Stetter, K.O., Huber, H. (2007) *Ignicoccus hospitalis* sp. nov., the host of “*Nanoarchaeum equitans*.” *Int. J. Syst. Evol. Microbiol.* 57(Pt 4), 803–808.
- [84] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25(7), 1043–1055.
- [85] Peng, Y., Leung, H.C.M., Yiu, S.-M., Chin, F.Y.L. (2010) IDBA—a practical iterative de Bruijn graph de novo assembler. In: *Annual International Conference on Research in Computational Molecular Biology*, Springer, pp. 426–440.
- [86] Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L. (2012) IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28(11), 1420–1428.
- [87] Perevalova, A.A., Svetlichny, V.A., Kublanov, I. V., Chernyh, N.A., Kostrikina, N.A., Tourova, T.P., Kuznetsov, B.B., Bonch-Osmolovskaya, E.A. (2005) *Desulfurococcus fermentans* sp. nov., a novel hyperthermophilic archaeon from a Kamchatka hot spring, and emended description of the genus *Desulfurococcus*. *Int. J. Syst. Evol. Microbiol.* 55(Pt 3), 995–999.

- [88] Pisa, K.Y., Huber, H., Thomm, M., Müller, V. (2007) A sodium ion-dependent A1AO ATP synthase from the hyperthermophilic archaeon *Pyrococcus furiosus*. FEBS J. 274(15), 3928–3938.
- [89] Podar, M., Makarova, K.S., Graham, D.E., Wolf, Y.I., Koonin, E. V., Reysenbach, A.-L. (2013) Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. Biol. Direct 8, 9, <https://doi.org/10.1186/1745-6150-8-9>.
- [90] Rachel, R., Wyszchony, I., Riehl, S., Huber, H. (2002) The ultrastructure of *Ignicoccus*: evidence for a novel outer membrane and for intracellular vesicle budding in an archaeon. Archaea 1(1), 9–18.
- [91] Randau, L., Münch, R., Hohn, M.J., Jahn, D., Söll, D. (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'-and 3'-halves. Nature 433(7025), 537–541.
- [92] Randau, L., Pearson, M., Söll, D. (2005) The complete set of tRNA species in *Nanoarchaeum equitans*. FEBS Lett. 579(13), 2945–2947.
- [93] Randau, L., Schröder, I., Söll, D. (2008) Life without RNase P. Nature 453(7191), 120–123.
- [94] Ravin, N. V., Mardanov, A. V., Beletsky, A. V., Kublanov, I. V., Kolganova, T. V., Lebedinsky, A. V., Chernyh, N.A., Bonch-Osmolovskaya, E.A., Skryabin, K.G. (2009) Complete genome sequence of the anaerobic, protein-degrading hyperthermophilic crenarchaeon *Desulfurococcus kamchatkensis*. J. Bacteriol. 191(7), 2371–2379.
- [95] Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-

- F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.-T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499(7459), 431–437.
- [96] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Prepr.* (4), e1900v1, <https://doi.org/10.7287/peerj.preprints.1900v1>.
- [97] Sato, T., Atomi, H., Imanaka, T. (2007) Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315(5814), 1003–1006.
- [98] Schlegel, K., Leone, V., Faraldo-Gómez, J.D., Müller, V. (2012) Promiscuous archaeal ATP synthase concurrently coupled to Na⁺ and H⁺ translocation. *Proc. Natl. Acad. Sci. U. S. A.* 109(3), 947–952.
- [99] Schopf, S., Wanner, G., Rachel, R., Wirth, R. (2008) An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri*. *Arch. Microbiol.* 190(3), 371–377.
- [100] Schouten, S., van der Meer, M.T.J., Hopmans, E.C., Rijpstra, W.I.C., Reysenbach, A.-L., Ward, D.M., Sinninghe Damsté, J.S. (2007) Archaeal and bacterial glycerol dialkyl glycerol tetraether lipids in hot springs of Yellowstone National Park. *Appl. Environ. Microbiol.* 73(19), 6181–6191.
- [101] Schuster, L., Bright, M. (2016) A novel colonial ciliate *Zoothamnium ignavum* sp. nov. (*Ciliophora*, *Oligohymenophorea*) and its ectosymbiont *Candidatus Navis piranensis* gen. nov., sp. nov. from shallow-water Wood Falls. *PLoS One* 11(9),

e0162834, <https://doi.org/10.1371/journal.pone.0162834>.

- [102] Sokolova, T.G., Henstra, A.-M., Sipma, J., Parshina, S.N., Stams, A.J.M., Lebedinsky, A. V. (2009) Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol. Ecol.* 68(2), 131–141.
- [103] Sokolova, T.G., Jeanthon, C., Kostrikina, N.A., Chernyh, N.A., Lebedinsky, A. V., Stackebrandt, E., Bonch-Osmolovskaya, E.A. (2004) The first evidence of anaerobic CO oxidation coupled with H₂ production by a hyperthermophilic archaeon isolated from a deep-sea hydrothermal vent. *Extremophiles* 8(4), 317–323.
- [104] Søndergaard, D., Pedersen, C.N.S., Greening, C. (2016) HydDB: A web tool for hydrogenase classification and analysis. *Sci. Rep.* 6, 34212, <https://doi.org/10.1038/srep34212>.
- [105] Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G. (2015) Complex *Archaea* that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551), 173–179.
- [106] Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9), 1312–1313.
- [107] Sugahara, J., Kikuta, K., Fujishima, K., Yachie, N., Tomita, M., Kanai, A. (2008) Comprehensive analysis of archaeal tRNA genes reveals rapid increase of tRNA introns in the order *Thermoproteales*. *Mol. Biol. Evol.* 25(12), 2709–2716.
- [108] Takai, K., Horikoshi, K. (1999) Molecular phylogenetic analysis of archaeal intron-containing genes coding for rRNA obtained from a deep-subsurface

geothermal water pool. Appl. Environ. Microbiol. 65(12), 5586–5589.

- [109] Ultsch, A., Mörchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM, Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany, no. 46.
- [110] Vannier, P., Marteinsson, V.T., Fridjonsson, O.H., Oger, P., Jebbar, M. (2011) Complete genome sequence of the hyperthermophilic, piezophilic, heterotrophic, and carboxydrotrophic archaeon *Thermococcus barophilus* MP. J. Bacteriol. 193(6), 1481–1482.
- [111] Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. Proc. Natl. Acad. Sci. U. S. A. 100(22), 12984–12988.
- [112] Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., Singer, S.W. (2014) MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2(1), 26, <https://doi.org/10.1186/2049-2618-2-26>.
- [113] Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L., Podar, M. (2016) Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. Nat. Commun. 7, 12115, <http://doi.org/10.1038/ncomms12115>.

- [114] Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J. (2017) Introducing EzBioCloud: A taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67(5), 1613–1617.
- [115] Yoshinaga, M.Y., Kellermann, M.Y., Rossel, P.E., Schubotz, F., Lipp, J.S., Hinrichs, K.-U. (2011) Systematic fragmentation patterns of archaeal intact polar lipids by high-performance liquid chromatography/electrospray ionization ion-trap mass spectrometry. *Rapid Commun. Mass Spectrom.* 25(23), 3563–3574.
- [116] Zillig, W., Stetter, K.O., Prangishvilli, D., Schäfer, W., Wunderl, S., Janekovic, D., Holz, I., Palm, P. (1982) *Desulfurococcaceae*, the second family of the extremely thermophilic, anaerobic, sulfur-respiring *Thermoproteales*. *Zentralblatt Für Bakteriologie, Mikrobiologie, Und Hygiene, I. Abteilung, Originalien, Angewandte Und Ökologische Mikrobiologie* 3(2), 304–317.
- [117] Zolghadr, B., Klingl, A., Koerdt, A., Driessen, A.J.M., Rachel, R., Albers, S.-V. (2010) Appendage-mediated surface adherence of *Sulfolobus solfataricus*. *J. Bacteriology* 192(1), 104–110.
- [118] Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31(13), 3406–3415.

Chapter 3: Deep-sea Hydrothermal Vent Metagenome-Assembled Genomes Provide

Insight into the Phylum *Nanoarchaeota*

St. John E, Flores G, Meneghin J, Reysenbach AL (In review). Environmental
Microbiology Reports

Summary

Ectosymbiotic *Nanoarchaeota* live on the surface of diverse archaeal hosts. Despite a broad distribution in global geothermal systems, only three *Nanoarchaeota* have been successfully co-cultivated with their hosts, and until now no nanoarchaeotal cultures or genomes have been described from deep-sea hydrothermal vents. We recovered three nanoarchaeotal metagenome-assembled genomes (MAGs) from deep-sea hydrothermal vent sites at the Eastern Lau Spreading Center (M10-121), Guaymas Basin (Gua-46) and the Mid-Cayman Rise (MC-1). Based on average amino acid identity analysis, M10-121 is a novel species in the candidate genus *Nanoclepta*, while the other two MAGs represent novel genera in the *Nanoarchaeota*. Like previously sequenced *Nanoarchaeota*, each MAG encodes at least one split protein-coding gene. The MAGs also contain a mosaic of key nanoarchaeotal features, including CRISPR repeat regions and marker genes for the gluconeogenic pathway and archaeal flagella. MC-1 also encodes the pentose biphosphate pathway, which may allow the nanoarchaeote to bypass several steps in glycolysis and produce ATP.

Introduction

The *Nanoarchaeota* are ultra-small hyperthermophilic *Archaea* that have been detected at marine hydrothermal vents [3,16,17,34] and terrestrial geothermal springs across the globe [9,10,12,19,36]. All cultivated *Nanoarchaeota* are obligate ectosymbionts that associate with unique hosts from the *Crenarchaeota* [20,47,53]. With their hosts, they may play a key role in colonizing newly-formed deep-sea hydrothermal vent deposits [34]. Despite their broad distribution and ecological importance however, no *Nanoarchaeota* genomes from deep-sea hydrothermal systems have been characterized.

To date, only three *Nanoarchaeota* have been successfully co-cultivated with their hosts. The marine lineage, *Nanoarchaeum equitans*, was obtained from a shallow hydrothermal system near Iceland [20], while the terrestrial taxa *Candidatus* *Nanopusillus acidilobi* [53] and *Candidatus* *Nanoclepta minutus* [47] were cultivated from geothermal springs in Yellowstone National Park (YNP), USA and Tikitere, New Zealand, respectively. The small genomes of cultivated *Nanoarchaeota* lack nearly all biosynthetic pathways and have split protein-coding genes [47,52,53], pointing to a pattern of genomic reduction and fragmentation. Although no functional ATP synthase has been identified in the genome of any cultivated *Nanoarchaeota* [35,47,53], the genomes retain a core set of genes involved in transcription, translation and DNA replication. Both *Cand. Nps. acidilobi* and *Cand. Ncl. minutus* have slightly expanded genomes compared to *N. equitans*, encoding genes for gluconeogenesis and archaeal flagella [47,53]. Several putative *Nanoarchaeota*-host pairings from YNP also have been identified through single-cell genomics, due to the tendency of *Nanoarchaeota* to co-sort

with their hosts [22,40]. These co-sorted genomes suggest that a wide diversity of crenarchaeotal hosts may associate with *Nanoarchaeota* across small geographic and spatial scales [22].

Here we report the assembly of three novel *Nanoarchaeota* metagenome-assembled genomes (MAGs) from three deep-sea hydrothermal vent systems in the Eastern Lau Spreading Center (ELSC), Guaymas Basin and the Mid-Cayman Rise. These MAGs broaden our understanding of the phylogenetic and genomic diversity of the *Nanoarchaeota*, while highlighting the need for continued sequencing and cultivation efforts in this lineage.

Materials and methods

Sample collection and processing

Hydrothermal vent deposits were collected in December 2009 from a thermocouple array placed on a hydrothermal vent in Guaymas Basin (27° 00.637' N, 111° 24.378' W) 14.8 days prior. Samples were collected using the human-occupied vehicle HOV *Alvin*. Additional hydrothermal deposits from Mariner vent field, ELSC (22° 10.821' S, 176° 36.100' W) were sampled in 2015 using the remotely operated vehicle ROV *Jason*. Sample collection, subsampling, storage and DNA extraction were performed as previously described [17,45].

Metagenomic sequencing, assembly and binning

Metagenomic libraries from Guaymas Basin and Eastern Lau Spreading Center (ELSC) were prepared and sequenced on the Illumina HiSeq at the Department of

Energy, Joint Genome Institute (JGI) and Oregon State University, respectively.

Metagenomic reads from Guaymas Basin also underwent quality control and assembly with SPAdes v. 3.11.1 [6] at the JGI. Poor quality sequences and adaptors were removed from ELSC metagenomic sequences using Trimmomatic [7] and trimmed reads were assembled with MegaHit v.1.1.1 [28,29]. Both metagenomes were binned based on tetranucleotide frequency and coverage using MetaBat v.0.32.4 [23]. *Nanoarchaeota* bins were identified with PhyloSift [13] and verified by concatenated ribosomal protein phylogenetic analysis.

Metagenomic reads from a diffuse-flow hydrothermal vent fluid sample (Von Damm vent field, Mid-Cayman Rise [3]) were downloaded from the NCBI Sequence Read Archive (SRA accession ERR1662245-ERR1662254 [27]). Initial binning with MetaBat produced three low-quality *Nanoarchaeota* bins. Contigs from these poor-quality bins were extracted and manually re-binned using ESOM [51], generating a single higher quality bin. To further curate the bin, we calculated tetranucleotide counts of the contigs using an in-house Perl script. These counts were used to calculate a resemblance matrix using the Bray-Curtis similarity measure in Primer v.6.0 [11]. Hierarchical cluster analysis was then performed on the resemblance matrix, and multidimensional scaling plots were created to visualize the results. One cluster (>90% similarity) contained the majority of the high-quality contigs (>10000 bp) and was extracted as the final bin.

Genomic analysis

Completeness and contamination of the *Nanoarchaeota* MAGs, along with identification and annotation of tRNAs, rRNAs, CRISPR repeats, open reading frames

and tRNA intron sequences were performed as previously described [47]. The completeness of the MAGs was also calculated by dividing the number of marker genes found in each MAG by the average marker gene count from all three cultivated *Nanoarchaeota*. Average amino acid identity comparisons were generated using the ANI/AAI Matrix Calculator tool [46]. Anvi'o was used to analyze and visualize pan and core nanoarchaeotal genomes [14,15]. The pangenome was generated using the 'anvi-pan-genome' command with the following parameters: --minbit 0.5, --mcl-inflation 6 and --use-ncbi-blast. COG and PFAM annotations in Anvi'o were also determined using BlastP [2]. Pairwise best-hit BlastP comparisons [2] were also used to identify genes shared by all cultivated *Nanoarchaeota* (E-value ≤ 0.00001). These shared genes were then placed into functional categories using the Archaeal Clusters of Orthologous Genes (arCOG) database [31,32]. Using the same method, the MAGs were each compared against the list of shared genes to identify homologues.

Phylogenetic analysis

Concatenated protein phylogenetic trees were constructed as described [47], using a set of eight ribosomal proteins (rpL2, 15, 18, 24, rpS3, 8, 10, 19). To generate small subunit ribosomal RNA gene trees, nanoarchaeotal 16S rRNA genes were aligned in Geneious v.10.0.2 [24] using the Clustal W aligner [26], and short clone sequences were added to the original alignment using the "Consensus Align" option. Curation and masking of hypervariable regions were performed manually. Maximum-likelihood phylogenetic trees were constructed and visualized as described [47]. Pairwise 16S rRNA gene comparisons were also performed using the Geneious aligner [24].

Data availability

The MAGs used in this study are available at <https://bitbucket.org/alrlab/deep-sea-hydrothermal-vent-nanoarchaeota-mags/src/master/>. Metadata and sampling information have been deposited in Genbank for the ELSC (BioProject PRJNA495050, BioSample SAMN10217875) and Guaymas Basin samples (BioProject PRJNA469236, BioSample SAMN09093796).

Results & Discussion

Phylogenetic analysis of nanoarchaeotal MAGs

To explore the diversity and coding potential of deep-sea *Nanoarchaeota*, we assembled metagenomes from three geographically distinct hydrothermal vent sites located in diverse geological settings (Table 2.1). Two metagenomic read sets were obtained by *de novo* sequencing, while the third was downloaded from the NCBI Sequence Read Archive (SRA). One nanoarchaeotal MAG was recovered from each of the three deep-sea hydrothermal vent metagenomes (Appendix B, File S2, B1-B3), and the position of all three MAGs in the *Nanoarchaeota* was confirmed using a concatenated ribosomal protein phylogenetic tree (Fig. 2.1). Although none of the MAGs were closed, each was compared to cultivated *Nanoarchaeota* using average amino acid identity (AAI) analysis (Fig. 2.2) and 16S rRNA gene sequence alignments, when applicable. Based on AAI comparisons, the ELSC nanoarchaeote (M10-121) is most closely related to the New Zealand taxon, *Cand. Nanoclepta minutus* (69% AAI), and it likely represents a novel

Table 2.1. Geological features and chemical composition of end-member fluids from deep-sea hydrothermal vent sites. Values less than 0.001 μM or 0.001 μm are reported as zero.

Vent field, site	Geographical feature	Host rock	Depth (m)	Temp. ($^{\circ}\text{C}$)	pH	H_2 (mM)	CH_4 (mM)	CO_2 (mm)	H_2S (mM)	Ref.
Mariner, ELSC*	Back-arc basin	Basaltic andesite	1877-1951	338-359	2.2-2.4	0.033-0.179	0.009-0.020	40-91	6.1-19	[17]
Von Damm, MCR	Slow-spreading ridge	Ultramafic	2300	226	5.56	19.2	2.84	2.78	3.2	[3,44]
Guaymas Basin	Marginal rift basin	Basalt sediment	2000	172-299	4.48-6.09	0.52-3.3	44.2-63.5	35.0-50.8	6.8-8.0	[44,49]
Seawater	--	--	--	2	7.6	0	0	2.3	<0.0001	[17]

ELSC, Eastern Lau Spreading Center; MCR, Mid-Cayman Rise; mM, mmol/L; mm, mmol/kg; μM , $\mu\text{mol/L}$; μm , $\mu\text{mol/kg}$; Ref, reference(s)
 * Reported data gathered in 2009 except for H_2 (2005)

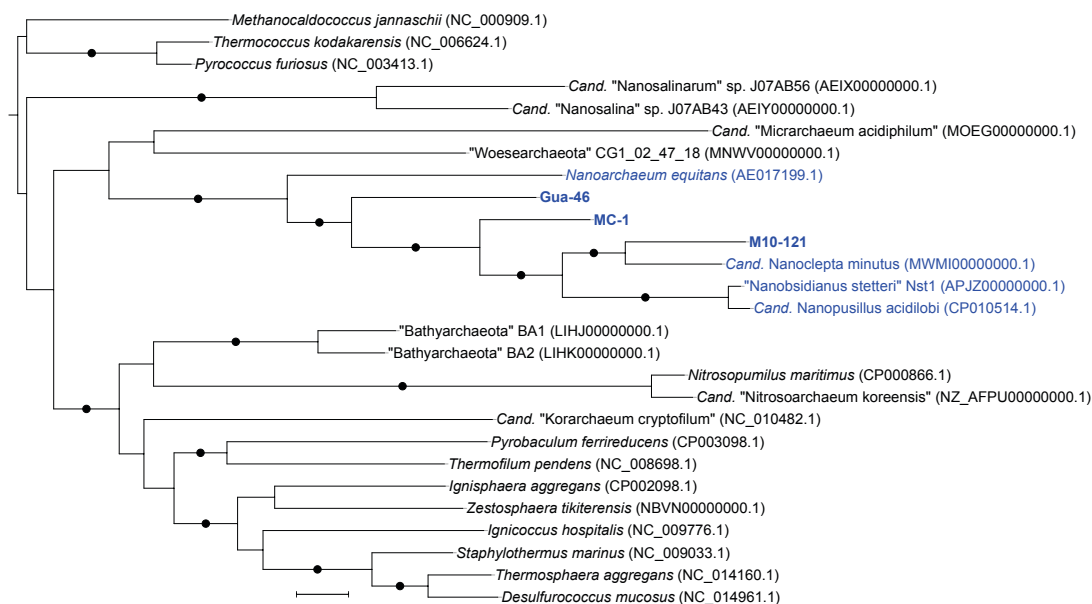


Figure 2.1. Maximum-likelihood phylogenetic tree based on eight concatenated ribosomal proteins. Members of the *Nanoarchaeota* are indicated in blue. The scale bar indicates 0.1 amino acid substitutions per site. Bootstrap support of 80-100% is indicated, based on 1000 replicate bootstrap trees. Members of the *Euryarchaeota* were used as the outgroups.

species in this genus [25]. The Mid-Cayman Rise MAG (MC-1) also shows the strongest similarity to *Cand. Ncl. minutus* by AAI (52%) and 16S rRNA gene sequence similarity (87.5%, Fig. 2.3), and it may represent the first species of a new genus in the *Nanoarchaeota*. AAI analysis also suggests that the Guaymas Basin MAG (Gua-46) represents a novel species in a second new genus, distinct from both *Cand. Ncl. minutus* and marine *N. equitans* (46% AAI). Although Gua-46 is equally similar to these cultivated taxa based on AAI, its partial 16S rRNA gene (779 bp) shows much stronger

gene sequence similarity to *N. equitans* (91.3%) than to *Cand. Ncl. minutus* (86.5%; Figure 2.3). While this discrepancy may be due to the partial length of the 16S rRNA gene, it may also point to the inherent limitations of AAI in assessing the similarity of incomplete genomes. AAI (51%) and 16S rRNA gene sequence similarity (92.8%) also suggest that Gua-46 is more closely related to MC-1 than to any of its cultivated relatives, although the two MAGs represent two distinct, novel genera in the *Nanoarchaeota*.

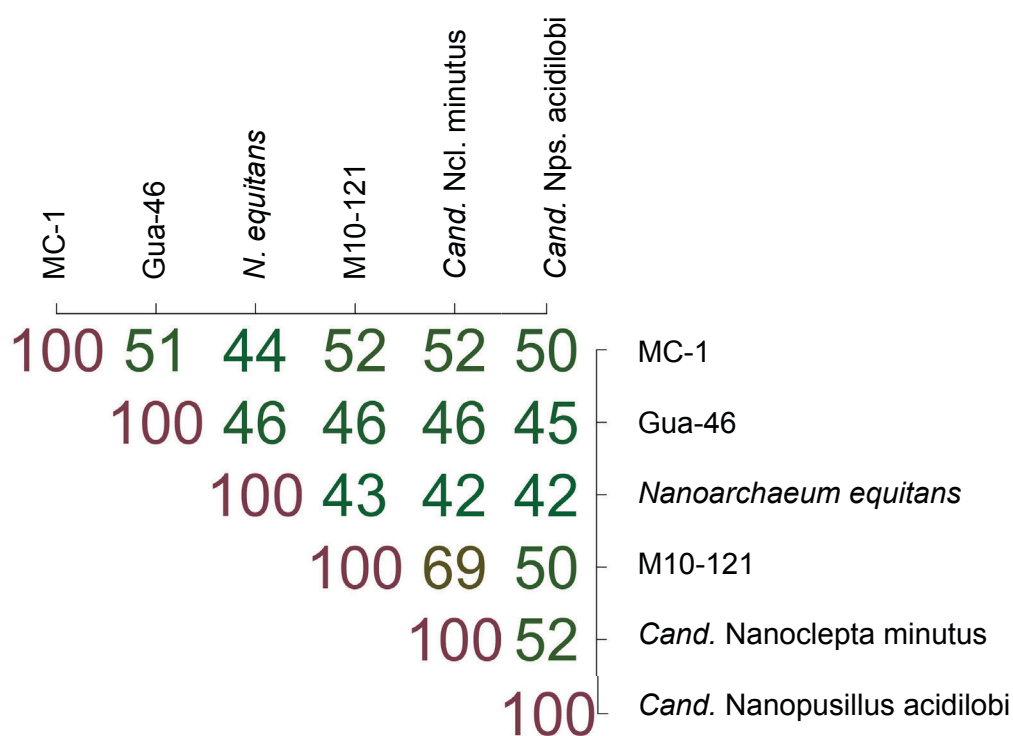


Figure 2.2. Average amino acid identity analysis of novel deep-sea *Nanoarchaeota* MAGs and their cultivated relatives

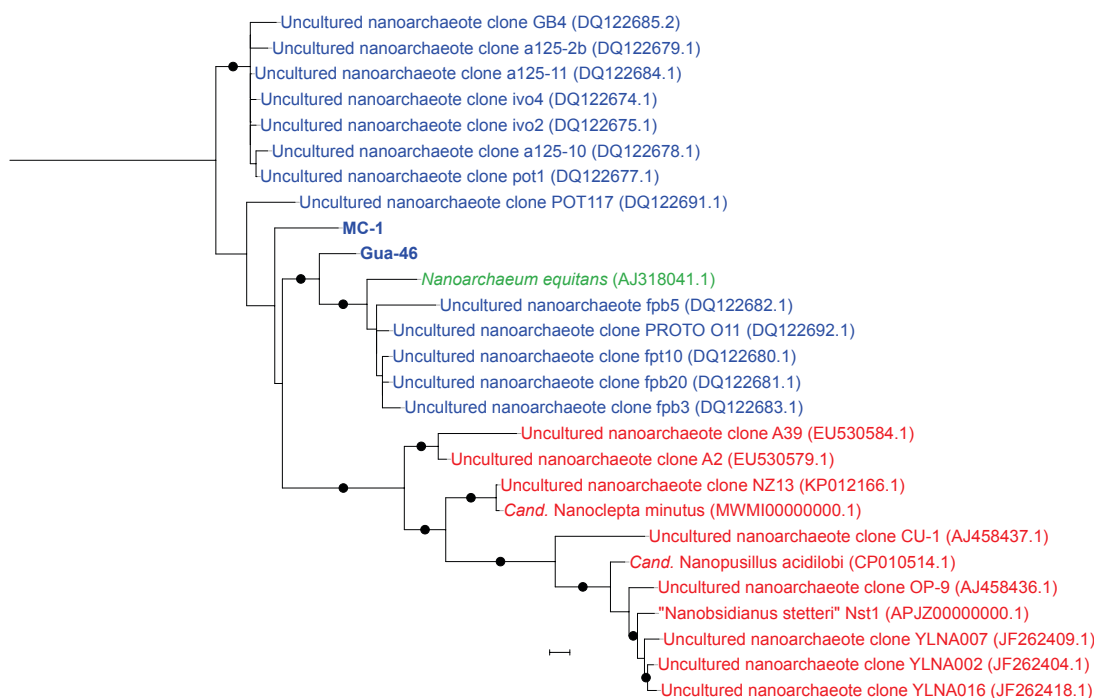


Figure 2.3. Maximum-likelihood phylogenetic analysis of nanoarchaeotal 16S rRNA genes (710 nucleotide positions). Because no 16S rRNA gene sequence was identified in M10-121, the MAG is not represented in this analysis. Bootstrapping was performed using 1000 replicate trees, and bootstrap support of 80% and greater is shown. The scale bar indicates 0.01 substitutions per nucleotide. The outgroup consists of 16S rRNA sequences from *Ignisphaera aggregans* AQ1.S1^T, *Desulfurococcus mucosus* O7/1^T and *Staphylothermus marinus* F1^T. Sequence color indicates environment type (red, terrestrial; green, shallow marine hydrothermal vent; blue, deep-sea hydrothermal vent).

Basic genomic features

The three *Nanoarchaeota* MAGs all have low contamination and are variable in length (Appendix B, File S2, Table B4). M10-121 is the most incomplete based on CheckM analysis (37.8%; [40]), containing 65 of the archaeal marker genes used to estimate completion. Gua-46 and MC-1 are more complete (48.92% and 67.99%), with 85 and 108 marker genes, respectively. CheckM analysis likely underestimates the true completeness of symbionts and parasites like the *Nanoarchaeota*, whose genomes are

highly reduced and do not contain the full set of archaeal marker genes. For example, the closed genomes of *N. equitans* and *Cand. Nps. acidilobi* are only estimated to be 73-74% complete by CheckM. Due to this limitation, directly comparing the number of marker genes found in each MAG to those identified in their cultivated relatives may provide a more accurate assessment of their relative completion. Based on comparison with the average marker genes recovered from *N. equitans*, *Cand. Nps. acidilobi* and *Cand. Ncl. minutus*, the MAGs are estimated to be 54.47% (M10-121), 71.23% (Gua-46) and 90.50% complete (MC-1).

Like previously sequenced *Nanoarchaeota* [40,47,52,53], the three MAGs each encode at least one tRNA gene containing an intron sequence (Appendix B, File S2, Table B4-B5). None of the predicted intronic sequences found in the MAGs were identified as Group I or Group II introns (E-value ≤ 0.001), and none of the introns had identical nucleotide sequences to other intronic regions in described *Nanoarchaeota* tRNA genes. None of the *trans*-encoded tRNA genes found in *N. equitans* [41] were identified in any of the MAGs. In our re-analysis of cultivated *Nanoarchaeota* genomes, we only recovered 41 of the tRNA genes found in *N. equitans*. However, we were able to identify additional intron-containing tRNAs in the *Cand. Nps. acidilobi* genome, which have thus far been unreported (Appendix B, File S2, Table B4-B5). Unlike *N. equitans* [42], both Gua-46 and MC-1 contain multiple RNase P protein subunits, suggesting RNase P may be functional in these lineages. Although no RNase P was identified in M10-121, it may be encoded in the unsequenced portion of the genome.

Several split protein-coding genes were found in the MAGs, similar to described *Nanoarchaeota* (Appendix B, File S2, Table B4, B6). Each MAG encodes a split gene

for the large helicase-related (LHR) protein, which is split in all cultivated *Nanoarchaeota* genomes [47,52,53]. Notably, the LHR protein split site is conserved in all nanoarchaeotes in our analysis, with the exception of *N. equitans*. Several truncated genes were also identified, predominantly in M10-121. Due to the partial nature of the MAGs, we were unable to determine conclusively whether these truncated genes represent halves of split protein coding genes, or if they merely encode for shortened gene variants. However, several of the truncated genes align with the 5' or 3' regions of described split protein coding genes in other *Nanoarchaeota*, suggesting they may be true split genes. Some truncated genes are also the terminal open reading frames encoded on particular contigs, and thus they may be artificially shortened. However, the detection of split protein-coding genes across sequenced *Nanoarchaeota* genomes points to a shared history of genomic fragmentation [47].

Comparative analysis of nanoarchaeotal coding potential

All cultivated *Nanoarchaeota* genomes retain a common set of protein coding genes, predominantly involved in DNA repair and replication, ribosomal structure, transcription, translation and post-translational modification (Archaeal Clusters of Orthologous Genes (arCOG) clusters J, K, L and O; Appendix B, File S2, Table B7-B10). Jarett et al. [22] recently proposed that *Nanoarchaeota* are under strong selective pressure to maintain the genes in these key functional groups. We identified several of these shared genes in the deep-sea nanoarchaeotal MAGs (Appendix B, File S2, Table B11-B14), including ribosomal proteins, genes involved in DNA replication and several subunits of the DNA-directed RNA polymerase. Pan-genomic analysis using Anvi'o

[14,15] also confirmed the presence of shared functional gene clusters across the MAGs and their cultivated relatives (Fig. 2.4). The recovery of these genes suggests that vital information-processing and replicative functions are maintained across nanoarchaeotal lineages.

Geographically distinct *Nanoarchaeota* may interact with diverse viral populations, as evidenced by CRISPR-Cas regions identified in *N. equitans* and *Cand.*

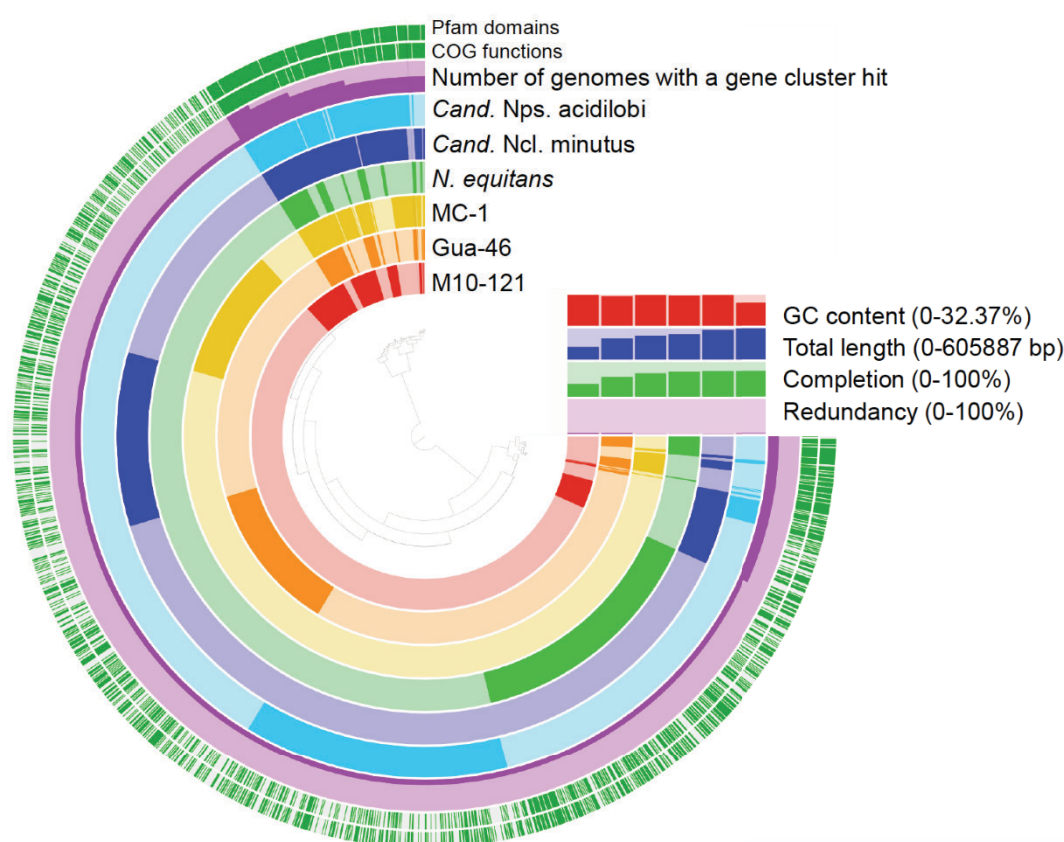


Figure 2.4. Pan-genome analysis of cultivated *Nanoarchaeota* and deep-sea nanoarchaeotal MAGs produced in Anvi'o [14,15]. Each MAG and cultivated *Nanoarchaeota* genome is shown as a separate layer, with genes ordered by functional categories. Shaded regions in each organism layer show the presence of genes in functional gene clusters.

Ncl. minutus [47,52] and integrated prophage DNA found in a single-cell nanoarchaeotal genome from YNP [36]. Although no CRISPR repeat regions were identified in Gua-46, the MAG contains an operon with three *cas* genes typically associated with a type III CRISPR-Cas cassette [33]. We also identified a putative CRISPR repeat region in MC-1, along with a separate *cas6* gene that could be part of a type I or type III CRISPR-Cas system [33]. None of the MC-1 CRISPR repeat sequences showed significant similarity (E-value ≤ 0.00001) to viral sequence in the NCBI non-redundant nucleotide database [37]. However, given that deep-sea hydrothermal vents harbor large viral communities [38], it is likely these CRISPR repeats may correspond to viruses that have not yet been described.

Unlike their marine relative *N. equitans*, cultivated terrestrial *Nanoarchaeota* contain genes for the gluconeogenic pathway and express archaeal flagella [47,53]. Analysis of the deep-sea nanoarchaeotal MAGs suggests these traits may not be limited to terrestrial *Nanoarchaeota*. Although we were unable to identify a complete gluconeogenic pathway in any of the MAGs, M10-121 contains a bifunctional fructose-1,6-bisphosphate aldolase/phosphatase (FBPA/ase), one of the marker genes for gluconeogenesis [53]. Both M10-121 and Gua-46 also contain one to three *flaB* genes, which encode for the long protein stalk of the archaeal flagella [1,50]. It is unclear whether or not nanoarchaeotal flagella are functional for motility, or if they perform an alternate role in host attachment and interaction [47,53]. However, the maintenance of archaeal flagellar genes in several of the *Nanoarchaeota* genomes may suggest they have a crucial function in these *Nanoarchaeota*.

ATP generation in the Nanoarchaeota

Currently, it is unknown if the *Nanoarchaeota* have the ability to produce any ATP, or if they rely on their hosts for all ATP synthesis. It was previously suggested that *N. equitans* may utilize NADH from its host to generate a proton motive force using a sulfide:quinone oxidoreductase, which would be used to generate ATP, presumably via a partial ATP synthase [43]. However, structural analysis has shown that the *N. equitans* ATP synthase complex is not capable of conformational changes and is therefore non-functional [35]. Recently it was hypothesized that *Nanoarchaeota* from YNP might also tap into the host's electron transport chain as suggested for *N. equitans* [22]. However, instead of utilizing a sulfide:quinone oxidoreductase, these *Nanoarchaeota* could potentially use a cytochrome *bd* ubiquinol oxidase, fueled by quinols from the host [22]. Although cytochrome *bd* ubiquinol oxidase could contribute to the generation of a membrane potential [8], it is unclear what role this enzyme would play in nanoarchaeotal bioenergetics, since no ATP synthase has been identified in any terrestrial *Nanoarchaeota* genome to date [40,47,53]. Even if *Nanoarchaeota* were capable of generating a membrane potential, they would still be reliant on their host's ATP synthase complex to harness that potential for ATP production.

Without an ATP synthase, *Nanoarchaeota* may still be able to produce small amounts of ATP through substrate-level phosphorylation. Glycolysis was suggested as a means of ATP production in terrestrial *Nanoarchaeota*, based on analysis of the single-cell genome “*Nanobsidianus stetteri*” Nst1 from YNP [40]. However, the glycolytic pathway is incomplete both in “*N. stetteri*” Nst1 [40] and in the genomes of *Cand. Nps. acidilobi* and *Cand. Ncl. minutus* [47,53] and is therefore likely non-functional in these

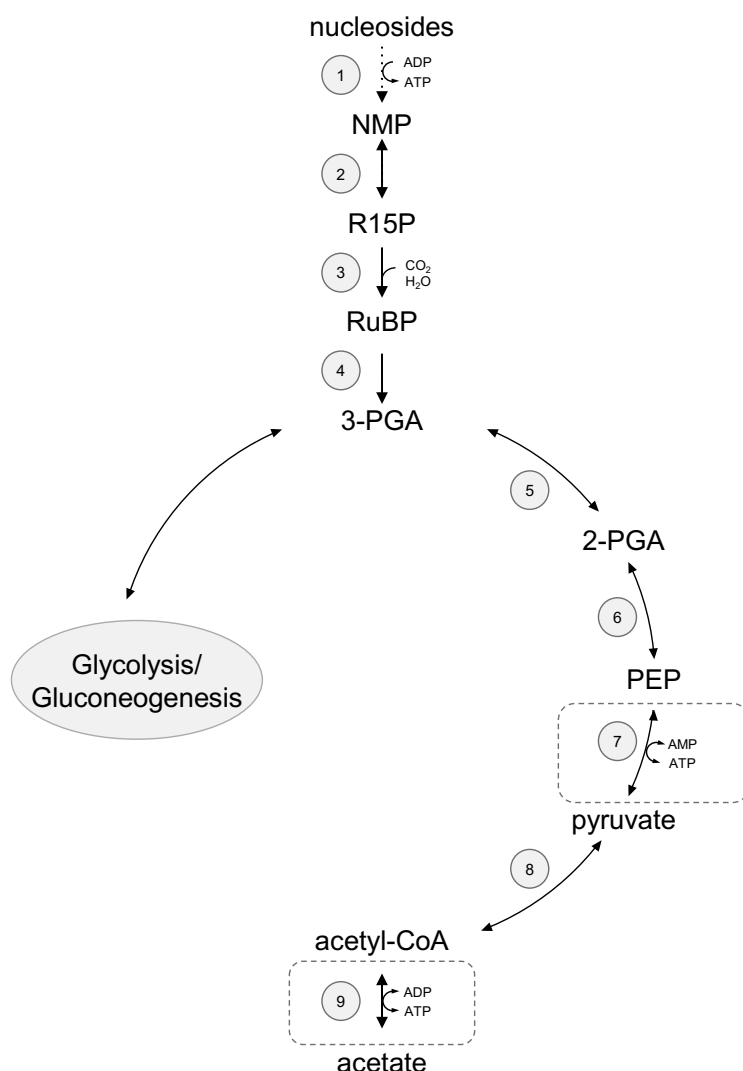


Figure 2.5. Reconstruction of the potential metabolic pathway for ATP generation in MC-1. Predicted genes: 1, putative nucleoside kinase (MC_288); 2, AMP phosphorylase (MC_480); 3, ribose-1,5-bisphosphate isomerase (MC_481); 4, type III Rubisco (MC_479); 5, phosphoglycerate mutase (MC_357); 6, enolase (MC_111); 7, phosphoenolpyruvate synthase (MC_225); 8, pyruvate:ferredoxin oxidoreductase (MC_146, 304); 9, acetyl-CoA synthetase (MC_147-149). Gene products: NMP, nucleoside monophosphate; R15P, ribose 1,5-bisphosphate; RuBP, ribulose 1,5-bisphosphate; PGA, phosphoglycerate; PEP, phosphoenolpyruvate.

lineages. Consistent with described *Nanoarchaeota*, none of the MAGs contained an evident, complete glycolytic pathway, and no ATP synthase or electron transport chain genes were detected. However, MC-1 does encode genes for the pentose bisphosphate pathway (Fig. 2.5), which would allow nanoarchaeotal cells to produce 3-phosphoglycerate and bypass the missing steps in glycolysis [4,5]. The pentose bisphosphate pathway typically begins with nucleoside monophosphates, which may be produced by a putative nucleoside kinase (MC_288) that shows 35.1-37.4% amino acid similarity to nucleoside kinases in *Methanocaldococcus jannaschii* and *Aeropyrum pernix* [4,18]. The gene also shows 25.6-28.5% similarity to a cytidine kinase and an ADP-dependent ribose 1-phosphate kinase found in *T. kodakarensis*, which feed into the pentose bisphosphate pathway at two different enzymatic steps [4]. Three-phosphoglycerate produced by the pentose bisphosphate pathway could be shuttled into glycolysis, where ATP synthesis could be carried out using a phosphoenolpyruvate synthase, as shown in *T. kodakarensis* [21] and suggested for *Cand. Nps. acidilobi* and *Cand. Ncl. minutus* [47,53]. Acetyl coenzyme A synthase may then catalyze a second ATP-producing reaction [30]. If functional, this metabolic network would allow some deep-sea *Nanoarchaeota* to circumvent the early steps of glycolysis and produce limited ATP apart from their hosts. With a means of independent ATP production, *Nanoarchaeota* may be able to detach from their hosts for short periods of time, possibly facilitating infection of new host cells or rare host-switching events. However, cultivation of deep-sea *Nanoarchaeota*-host systems will be essential to determine if this pathway is functional, and what role it may play in host/symbiont interactions.

Previous studies of cultivated *Nanoarchaeota* have shown a clear physiological distinction between marine and terrestrial lineages, with marine *N. equitans* lacking genes for archaeal flagella and gluconeogenesis that are found in terrestrial taxa [40,47,52,53]. However, our study suggests this apparent dichotomy does not exist. Genomic evidence of these key features is scattered across deep-sea nanoarchaeotal MAGs, with no clear pattern of distribution.

Although we do not propose any putative nanoarchaeotal-host pairings, the apparent absence of ATP synthase genes suggests that deep-sea *Nanoarchaeota* rely on intimate associations with host organisms, similar to their described relatives. Studies based on 16S rRNA gene sequencing have shown that deep-sea hydrothermal vents support diverse communities of *Crenarchaeota*, including lineages related to the marine *N. equitans* host, *Ignicoccus*, and many taxa that have not yet been described [16,17,34,48]. The ELSC metagenome also contained sequences most closely related to the New Zealand nanoarchaeote host organism, *Zestosphaera*. It is likely that deep-sea nanoarchaeotes associate with some, or many, of these crenarchaeotal hosts, and that the relationship between symbiont and host may influence which traits are retained or lost in particular nanoarchaeotal lineages. Host type may also vary significantly between different hydrothermal vent sites due to the diverse geochemical settings of the marine vent sites, which can strongly affect community structure [16,17]. Cultivation and single-cell genomic studies of deep-sea *Nanoarchaeota* will be key to understanding the scope of nanoarchaeotal diversity and the complexities of this symbiont-host interaction.

Acknowledgements

This work was supported through grants from the National Aeronautics and Space Administration (NNX16AJ66G), the US-National Science Foundation (OCE1235432, OCE1558795, OCE0728391 and DEB1134877) and the Department of Energy Joint Genome Institute (Community Science Program award 339) to ALR and the Forbes-Lea Endowment Fund for Student Research award to ESJ. We thank the captain and crew members of the RV *Thomas Thompson* and the RV *Atlantis*, and the ROV *Jason* and HOV *Alvin* teams for assistance in collecting samples.

Contributions

ESJ performed genomic analysis of the MAGs and BlastP-based core-genome analysis, 16S rRNA gene and ribosomal protein phylogenetic analyses and average amino acid identity comparisons. JM performed iterative binning of the MC-1 MAG. Anvi'o-based pan-genomic analysis was done by GF. ESJ and ALR prepared and edited the manuscript.

References

- [1] Albers, S.-V., Jarrell, K.F. (2015) The archaellum: how *Archaea* swim. *Front. Microbiol.* 6, 23, <https://doi.org/10.3389/fmicb.2015.00023>.
- [2] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215(3), 403–410.
- [3] Anderson, R.E., Reveillaud, J., Reddington, E., Delmont, T.O., Eren, A.M., McDermott, J.M., Seewald, J.S., Huber, J.A. (2017) Genomic variation in microbial populations inhabiting the marine seafloor at deep-sea hydrothermal vents. *Nat. Commun.* 8(1), 1114, <https://doi.org/10.1038/s41467-017-01228-6>.
- [4] Aono, R., Sato, T., Imanaka, T., Atomi, H. (2015) A pentose bisphosphate pathway for nucleoside degradation in *Archaea*. *Nat. Chem. Biol.* 11(5), 355–360.
- [5] Aono, R., Sato, T., Yano, A., Yoshida, S., Nishitani, Y., Miki, K., Imanaka, T., Atomi, H. (2012) Enzymatic characterization of AMP phosphorylase and ribose-1,5-bisphosphate isomerase functioning in an archaeal AMP metabolic pathway. *J. Bacteriol.* 194(24), 6847–6855.
- [6] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19(5), 455–477.
- [7] Bolger, A.M., Lohse, M., Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114–2120.
- [8] Borisov, V.B., Gennis, R.B., Hemp, J., Verkhovsky, M.I. (2011) The cytochrome

- bd* respiratory oxygen reductases. *Biochim. Biophys. Acta - Bioenerg.* 1807(11), 1398–1413.
- [9] Casanueva, A., Galada, N., Baker, G.C., Grant, W.D., Heaphy, S., Jones, B., Yanhe, M., Ventosa, A., Blamey, J., Cowan, D.A. (2008) Nanoarchaeal 16S rRNA gene sequences are widely dispersed in hyperthermophilic and mesophilic halophilic environments. *Extremophiles* 12, 651–656.
- [10] Chernyh, N.A., Mardanov, A. V., Gumerov, V.M., Miroshnichenko, M.L., Lebedinsky, A. V., Merkel, A.Y., Crowe, D., Pimenov, N. V., Rusanov, I.I., Ravin, N. V., Moran, M.A., Bonch-Osmolovskaya, E.A. (2015) Microbial life in Bourlyashchy, the hottest thermal pool of Uzon Caldera, Kamchatka. *Extremophiles* 19(6), 1157–1171.
- [11] Clarke, K.R., Gorley, R.N. (2006) Primer V6: User Manual - Tutorial. Plymouth Marine Laboratory, Plymouth, England.
- [12] Clingenpeel, S., Kan, J., Macur, R.E., Woyke, T., Lovalvo, D., Varley, J., Inskeep, W.P., Nealson, K., McDermott, T.R. (2013) Yellowstone Lake *Nanoarchaeota*. *Front. Microbiol.* 4, 274, <https://doi.org/10.3389/fmicb.2013.00274>.
- [13] Darling, A.E., Jospin, G., Lowe, E., Matsen 4th, F.A., Bik, H.M., Eisen, J.A. (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2, e243, <https://doi.org/10.7717/peerj.243>.
- [14] Delmont, T.O., Eren, A.M. (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. *PeerJ* 6, e4320, <https://doi.org/10.7717/peerj.4320>.
- [15] Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L.,

- Delmont, T.O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 3, e1319, <https://doi.org/10.7717/peerj.1319>.
- [16] Flores, G.E., Campbell, J.H., Kirshtein, J.D., Meneghin, J., Podar, M., Steinberg, J.I., Seewald, J.S., Tivey, M.K., Voytek, M.A., Yang, Z.K., Reysenbach, A.L. (2011) Microbial community structure of hydrothermal deposits from geochemically different vent fields along the Mid-Atlantic Ridge. *Environ. Microbiol.* 13(8), 2158–2171.
- [17] Flores, G.E., Shakya, M., Meneghin, J., Yang, Z.K., Seewald, J.S., Geoff Wheat, C., Podar, M., Reysenbach, A.-L. (2012) Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* 10(4), 333–346.
- [18] Hansen, T., Arnfors, L., Ladenstein, R., Schönheit, P. (2007) The phosphofructokinase-B (MJ0406) from *Methanocaldococcus jannaschii* represents a nucleoside kinase with a broad substrate specificity. *Extremophiles* 11(1), 105–114.
- [19] Hohn, M.J., Hedlund, B.P., Huber, H. (2002) Detection of 16S rDNA sequences representing the novel phylum “*Nanoarchaeota*”: Indication for a wide distribution in high temperature biotopes. *Syst. Appl. Microbiol.* 25, 551–554.
- [20] Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002) A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* 417(6884), 63–67.
- [21] Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H., Imanaka, T. (2006) Phosphoenolpyruvate synthase plays an essential role for glycolysis in the

modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. Mol.

Microbiol. 61(4), 898–909.

- [22] Jarett, J.K., Nayfach, S., Podar, M., Inskeep, W., Ivanova, N.N., Munson-McGee, J., Schulz, F., Young, M., Jay, Z.J., Beam, J.P., Kyrpides, N.C., Malmstrom, R.R. (2018) Single-cell genomics of co-sorted *Nanoarchaeota* suggests novel putative host associations and diversification of proteins involved in symbiosis. Microbiome 6(161), <https://doi.org/10.1186/s40168-018-0539-8>.
- [23] Kang, D.D., Froula, J., Egan, R., Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165, <https://doi.org/10.7717/peerj.1165>.
- [24] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A. (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12), 1647–1649.
- [25] Konstantinidis, K.T., Rosselló-Móra, R., Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. ISME J. 11(11), 2399–2406.
- [26] Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., Mcgettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21), 2947–2948.
- [27] Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration. (2010) The sequence read archive. Nucleic Acids Res.

39(Database issue), D19–D21.

- [28] Li, D., Liu, C.-M., Luo, R., Sadakane, K., Lam, T.-W. (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31(10), 1674–1676.
- [29] Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H., Lam, T.-W. (2016) MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11.
- [30] Mai, X., Adams, M.W. (1996) Purification and characterization of two reversible and ADP-dependent acetyl coenzyme A synthetases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* 178(20), 5897–5903.
- [31] Makarova, K., Wolf, Y., Koonin, E. (2015) Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5(1), 818–840.
- [32] Makarova, K.S., Sorokin, A. V., Novichkov, P.S., Wolf, Y.I., Koonin, E. V. (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct*, 33, <https://doi.org/10.1186/1745-6150-2-33>.
- [33] Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., Koonin, E. V. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.*

13(11), 722–736.

- [34] McCliment, E.A., Voglesonger, K.M., O'Day, P.A., Dunn, E.E., Holloway, J.R., Cary, S.C. (2006) Colonization of nascent, deep-sea hydrothermal vents by a novel archaeal and nanoarchaeal assemblage. *Environ. Microbiol.* 8(1), 114–125.
- [35] Mohanty, S., Jobichen, C., Chichili, V.P.R., Velázquez-Campoy, A., Low, B.C., Hogue, C.W.V., Sivaraman, J. (2015) Structural basis for a unique ATP synthase core complex from *Nanoarchaeum equitans*. *J. Biol. Chem.* 290(45), 27280–27296.
- [36] Munson-McGee, J.H., Field, E.K., Bateson, M., Rooney, C., Stepanauskas, R., Young, M.J. (2015) *Nanoarchaeota*, their *Sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone National Park hot springs. *Appl. Environ. Microbiol.* 81(22), 7860–7868.
- [37] NCBI Resource Coordinators. (2017) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 45(D1), D12–D17.
- [38] Ortmann, A.C., Suttle, C.A. (2005) High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. *Deep. Res. Part I Oceanogr. Res. Pap.* 52(8), 1515–1527.
- [39] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25(7), 1043–1055.
- [40] Podar, M., Makarova, K.S., Graham, D.E., Wolf, Y.I., Koonin, E. V., Reysenbach, A.-L. (2013) Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone

National Park. Biol. Direct 8, 9, <https://doi.org/10.1186/1745-6150-8-9>.

- [41] Randau, L., Münch, R., Hohn, M.J., Jahn, D., Söll, D. (2005) *Nanoarchaeum equitans* creates functional tRNAs from separate genes for their 5'-and 3'-halves. Nature 433(7025), 537–541.
- [42] Randau, L., Schröder, I., Söll, D. (2008) Life without RNase P. Nature 453(7191), 120–123.
- [43] Rawle, R.A., Hamerly, T., Tripet, B.P., Giannone, R.J., Wurch, L., Hettich, R.L., Podar, M., Copié, V., Bothner, B. (2017) Multi-omics analysis provides insight to the *Ignicoccus hospitalis*-*Nanoarchaeum equitans* association. Biochim. Biophys. Acta - Gen. Subj. 1861(9), 2218–2227.
- [44] Reeves, E.P., McDermott, J.M., Seewald, J.S. (2014) The origin of methanethiol in midocean ridge hydrothermal fluids. Proc. Natl. Acad. Sci. 111(15), 5474–5479.
- [45] Reysenbach, A.-L., Liu, Y., Banta, A.B., Beveridge, T.J., Kirshtein, J.D., Schouten, S., Tivey, M.K., Von Damm, K.L., Voytek, M.A. (2006) A ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature 442(7101), 444–447.
- [46] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr., 4, e1900v1, <https://doi.org/10.7287/peerj.preprints.1900v1>.
- [47] St. John, E., Liu, Y., Podar, M., Stott, M.B., Meneghin, J., Chen, Z., Lagutin, K., Mitchell, K., Reysenbach, A.-L. (In press) A new symbiotic nanoarchaeote (*Candidatus Nanoclepta minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. Syst. Appl. Microbiol.,

<https://doi.org/10.1016/j.syapm.2018.08.005>.

- [48] Takai, K., Horikoshi, K. (1999) Genetic diversity of *Archaea* in deep-sea hydrothermal vent environments. *Genetics* 152(4), 1285–1297.
- [49] Teske, A., De Beer, D., McKay, L.J., Tivey, M.K., Biddle, J.F., Hoer, D., Lloyd, K.G., Lever, M.A., Røy, H., Albert, D.B., Mendlovitz, H.P., MacGregor, B.J. (2016) The Guaymas Basin hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front. Microbiol.* 7(FEB), 1–23.
- [50] Thomas, N.A., Bardy, S.L., Jarrell, K.F. (2001) The archaeal flagellum: a different kind of prokaryotic motility structure. *Microbiol. Rev.* 25(2), 147-174.
- [51] Ultsch, A., Mörchen, F. (2005) ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM, Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany, no. 46.
- [52] Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.* 100(22), 12984–12988.
- [53] Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L., Podar, M. (2016) Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. *Nat. Commun.* 7, 12115,

<http://doi.org/10.1038/ncomms12115>.

Chapter 4: Conclusion

The research presented in this thesis provides key insights into the *Nanoarchaeota*, one of several novel phyla discovered over the past four decades of archaeal research [1–3,13,17]. The cultivation of the first *Nanoarchaeota*-host system revolutionized archaeal biology, revealing that *Archaea* engage in close, interspecies relationships with other *Archaea*. Today, this lifestyle is thought to be widespread across several archaeal lineages [1,5,6,9]. However, many aspects of nanoarchaeotal ecophysiology remain enigmatic, due to the limited amount of genetic material available. In this thesis, direct cultivation, phylogenetic analysis and comparative genomics are used to reveal previously unknown diversity in the *Nanoarchaeota*, both on the functional and taxonomic level.

In Chapter 2, anaerobic cultivation was used to isolate a novel *Nanoarchaeota* and its host from a geothermal spring in Tikitere, New Zealand. The New Zealand nanoarchaeote (*Candidatus* Nanoclepta minutus) was compared to its cultivated relatives on the genomic level. Despite their large geographical separation, *Cand. Ncl. minutus* showed distinct similarities to terrestrial lineages from Yellowstone National Park (YNP), encoding genes for gluconeogenesis and archaeal flagella which are not found in the shallow marine nanoarchaeote, *Nanoarchaeum equitans* [12,14,15]. However, *Cand. Ncl. minutus* also encoded for a CRISPR-Cas system like *N. equitans* [14]. The phylogenetic placement of *Cand. Ncl. minutus* and its host *Zestosphaera tikitensis* was determined using a multi-faceted approach, which showed that *Cand. Ncl. minutus* is a new genus in the *Nanoarchaeota*, while *Z. tikitensis* is a novel genus in the *Crenarchaeota*. *Z. tikitensis* was also purified in axenic culture, described and

compared to its closest relative. Although *Z. tikiterensis* represents a novel genus, it shared some key features with the distantly-related nanoarchaeote host *Acidilobus* sp. 7A from YNP, including genes associated with a heterotrophic lifestyle and an outer S-layer [22].

In light of the similarities between the New Zealand and YNP *Nanoarchaeota* and their divergence from *N. equitans*, the studies presented in Chapter 3 focused on the relationship between cultivated terrestrial and shallow marine *Nanoarchaeota* and their undescribed deep-sea relatives. Phylogenetic and genomic analysis were used to assess if marine and terrestrial nanoarchaeotes from various geothermal settings show consistent, observable differences, either on the phylogenetic or genomic scale. Three metagenome-assembled genomes (MAGs) from diverse deep-sea hydrothermal vent environments were assembled and described. Based on phylogenetic reconstruction, the three MAGs represented two new genera and an additional novel species of *Nanoarchaeota*. However, terrestrial and marine *Nanoarchaeota* did not resolve into distinct monophyletic lineages. Comparative genomic analysis of the MAGs also revealed a patchy distribution of functional genes previously thought to be limited to terrestrial *Nanoarchaeota*. The phylogenetic relatedness of these *Nanoarchaeota* and the distribution of key marker genes suggests a complex evolutionary history between marine and terrestrial nanoarchaeotal lineages, with no clear pattern to delineate the two. One of the deep-sea *Nanoarchaeota* MAGs from the Mid-Cayman Rise also contained a novel pathway for potential ATP production. If functional, the pathway may allow certain *Nanoarchaeota* to produce low amounts of ATP independent from their hosts.

In light of the studies presented in this thesis, an integrated view of nanoarchaeotal biology and ecophysiology has begun to emerge (Fig. 3.1A-B). To date, all cultivated *Nanoarchaeota* share an ectosymbiotic lifestyle with hosts from the *Crenarchaeota* [12,20,22]. Metabolic building blocks such as lipids, amino acids, cofactors and nucleotides are likely shuttled from hosts to their *Nanoarchaeota*, due to the reduced biosynthetic capabilities of nanoarchaeotes [18,20–22]. Likely, this transfer of nutrients is shuttled through a pore-type structure bridging the *Nanoarchaeota* and host. Although this direct connection has only been verified in the *N. equitans*-*I. hospitalis* system [11], electron microscopy suggests it may be further distributed across nanoarchaeotal lineages [20,22], as discussed in Chapter 2.

The method by which ATP synthesis and energy generation occur in the *Nanoarchaeota* still remains unclear. However, the apparent absence of a complete ATP synthase complex or glycolytic pathway in described *Nanoarchaeota* [14,16–18,20,22] indicates that most nanoarchaeotes likely rely on ATP provided by their hosts. The discovery of the pentose biphosphate pathway in one nanoarchaeotal MAG suggests some limited ATP production may be possible in deep-sea *Nanoarchaeota*. However, direct cultivation will be necessary to determine whether or not this pathway is functional.

The means by which nanoarchaeotes and their hosts recognize one another, initiate attachment and maintain their association is not yet fully understood. Thus far, *I. hospitalis*, the host of *N. equitans* is unique among nanoarchaeote hosts due to its double membrane system (Fig. 3.1B), which may be crucial in the attachment of *N. equitans* [11]. In contrast, terrestrial hosts from New Zealand and YNP encode for

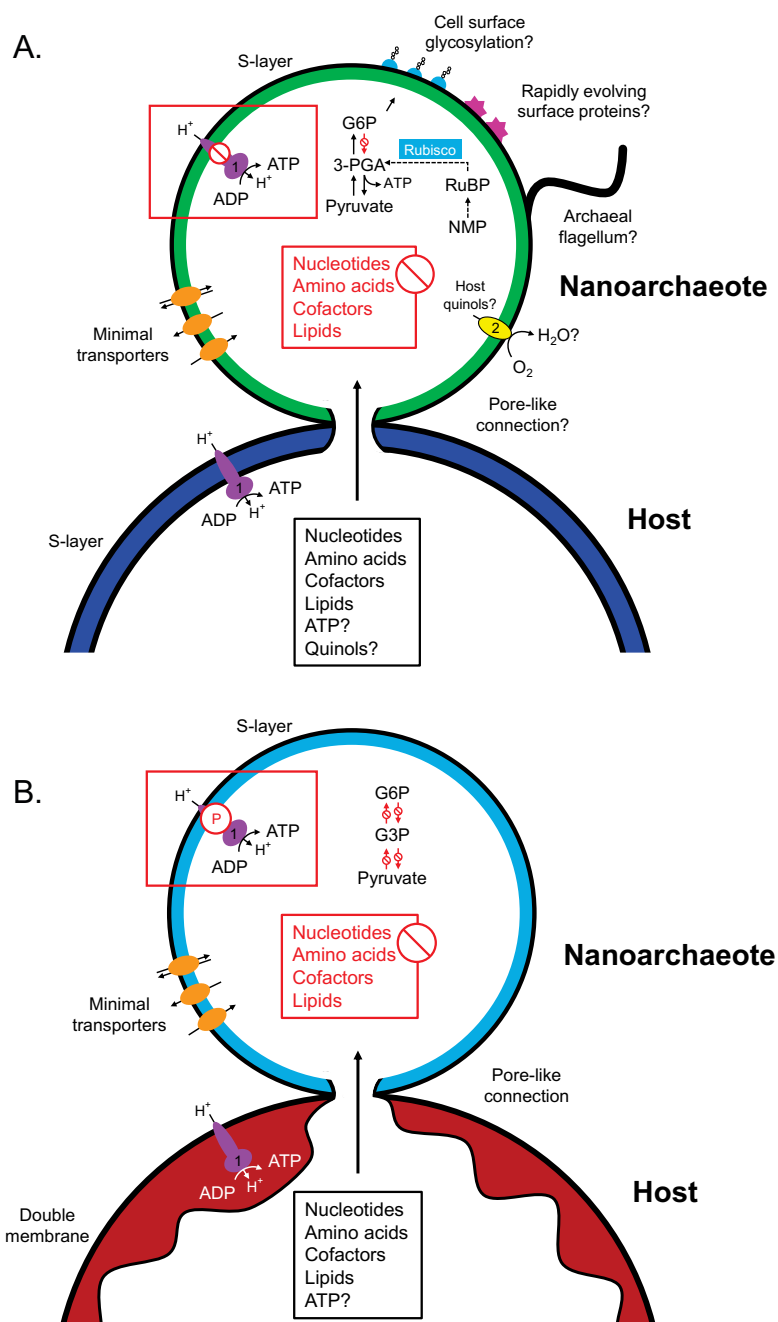


Figure 3.1. Current understanding of metabolism in the *Nanoarchaeota* and nanoarchaeote-host interactions. (A) Features from terrestrial nanoarchaeotes and deep-sea nanoarchaeotal MAGs, shown with a prototypical non-*Ignicoccus* host; (B) features of *N. equitans* and its host *I. hospitalis*. G6P, glucose 6-phosphate; NMP, nucleoside monophosphate; 3-PGA, 3-phosphoglycerate; RuBP, ribulose 1,5-bisphosphate; 1, ATP synthase complex; 2, cytochrome *bd* ubiquinol oxidase; P, partial; Ø, absent/not detected. Genes from the pentose bisphosphate pathway are indicated with dashed lines.

S-layers [18,20,22]. The presence of archaeal flagella, glycosylated cell surfaces and rapidly-evolving surface proteins (Fig. 3.1A) have all been implicated in the attachment between terrestrial nanoarchaeotes and their hosts [14,18,22], and may also play a role in deep-sea high-temperature *Nanoarchaeota*-host systems, as discussed in Chapter 3. However, the exact role of these various systems still remains unclear.

Looking forward, the next steps in nanoarchaeotal research can build on the analyses in this thesis by further exploring the interaction between the *Nanoarchaeota* and their hosts. In the laboratory setting, cultivated terrestrial nanoarchaeotes do not reach the high concentrations seen in the *N. equitans*-*I. hospitalis* co-culture [13,20,22]. Nonetheless, comparative transcriptomics and proteomics may be a valuable tool to further understand these systems. By comparing gene transcription and protein expression in terrestrial *Nanoarchaeota*-host co-cultures and isolated host cultures, the effects of nanoarchaeote infection on diverse hosts may be elucidated, as shown in the *N. equitans*-*I. hospitalis* pairing [7,8]. Differences in host response to nanoarchaeote infection may also clarify whether different *Nanoarchaeota*-host systems more strongly resemble parasitic or commensal relationships. As previously demonstrated in the *N. equitans*/*I. hospitalis* system [10], metabolomics may also be used to assess the effects of symbiont infection on host physiology, particularly in terms of ATP consumption.

The New Zealand nanoarchaeote, *Cand. Ncl. minutus*, also provides a unique opportunity for studying the effects of other community members on interspecies interaction. *Cand. Ncl. minutus* grows to higher densities in the mixed enrichment culture setting as compared to pure co-cultures, suggesting that other lineages in the enrichment community may positively affect the host/symbiont pairing. To assess the

potential effects of other microbial species on *Cand. Ncl. minutus* and its host, individual microbes isolated from the enrichment cultures (described in Chapter 2) could be added to pure nanoarchaeote-host co-cultures. By tracking growth response and applying differential transcriptomic, proteomic or metabolomic approaches to these cultures, the various effects of non-host lineages from the natural environment could be observed.

Additionally, nanoarchaeotal research would greatly benefit from the continued cultivation of *Nanoarchaeota*-host systems from diverse environments. With their putative novel pathway for ATP generation, *Nanoarchaeota* from the Mid-Cayman Rise represent an especially vital lineage for cultivation. Assessing the functionality of this pathway and its apparent effect on the nanoarchaeote-host symbiosis may provide crucial insights into the evolution of host dependency in this phylum. Additional sites for targeted nanoarchaeotal cultivation would include geographically distinct deep-sea hydrothermal vent sites and hot springs with diverse biogeochemical conditions. As more nanoarchaeotes and their hosts are described from these diverse environments, patterns may become clear regarding the mechanism of host/symbiont interaction and the role that environment type, host taxonomy and nanoarchaeote genomics play in this complex association.

References

- [1] Baker, B.J., Comolli, L.R., Dick, G.J., Hauser, L.J., Hyatt, D., Dill, B.D., Land, M.L., Verberkmoes, N.C., Hettich, R.L., Banfield, J.F. (2010) Enigmatic, ultrasmall, uncultivated *Archaea*. Proc. Natl. Acad. Sci. U. S. A. 107(19), 8806–8811.
- [2] Barns, S.M., Delwiche, C.F., Palmer, J.D., Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc. Natl. Acad. Sci. U. S. A. 93(17), 9188–9193.
- [3] Barns, S.M., Fundyga, R.E., Jeffries, M.W., Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. U. S. A. 91(5), 1609–1613.
- [4] Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P. (2008) Mesophilic *Crenarchaeota*: Proposal for a third archaeal phylum, the *Thaumarchaeota*. Nat. Rev. Microbiol. 6, 245–252.
- [5] Castelle, C.J., Brown, C.T., Anantharaman, K., Probst, A.J., Huang, R.H., Banfield, J.F. (2018) Biosynthetic capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat. Rev. Microbiol. 16(10), 629–645.
- [6] Castelle, C.J., Wrighton, K.C., Thomas, B.C., Hug, L.A., Brown, C.T., Wilkins, M.J., Frischkorn, K.R., Tringe, S.G., Singh, A., Markillie, L.M., Taylor, R.C., Williams, K.H., Banfield, J.F. (2015) Genomic expansion of domain *Archaea* highlights roles for organisms from new phyla in anaerobic carbon cycling. Curr. Biol. 25(6), P690–P701.
- [7] Giannone, R.J., Huber, H., Karpinets, T., Heimerl, T., Küper, U., Rachel, R.,

- Keller, M., Hettich, R.L., Podar, M. (2011) Proteomic characterization of cellular and molecular processes that enable the *Nanoarchaeum equitans*-*Ignicoccus hospitalis* relationship. PLoS One 6(8), e22942, <https://doi.org/10.1371/journal.pone.0022942>.
- [8] Giannone, R.J., Wurch, L.L., Heimerl, T., Martin, S., Yang, Z., Huber, H., Rachel, R., Hettich, R.L., Podar, M. (2015) Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. ISME J. 9(1), 101–114.
- [9] Golyshina, O. V., Toshchakov, S. V., Makarova, K.S., Gavrilov, S.N., Korzhenkov, A.A., La Cono, V., Arcadi, E., Nechitaylo, T.Y., Ferrer, M., Kublanov, I. V., Wolf, Y.I., Yakimov, M.M., Golyshin, P.N. (2017) “ARMAN” *Archaea* depend on association with euryarchaeal host in culture and in situ. Nat. Commun. 8, 60, <https://doi.org/10.1038/s41467-017-00104-7>.
- [10] Hamerly, T., Tripet, B.P., Tigges, M., Giannone, R.J., Wurch, L., Hettich, R.L., Podar, M., Copié, V., Bothner, B. (2015) Untargeted metabolomics studies employing NMR and LC–MS reveal metabolic coupling between *Nanoarchaeum equitans* and its archaeal host *Ignicoccus hospitalis*. Metabolomics 11, 895–907.
- [11] Heimerl, T., Flechsler, J., Pickl, C., Heinz, V., Salecker, B., Zweck, J., Wanner, G., Geimer, S., Samson, R.Y., Bell, S.D., Huber, H., Wirth, R., Wurch, L., Podar, M., Rachel, R. (2017) A complex endomembrane system in the archaeon *Ignicoccus hospitalis* tapped by *Nanoarchaeum equitans*. Front. Microbiol. 8, 1072, <https://doi.org/10.3389/fmicb.2017.01072>.
- [12] Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., Stetter, K.O. (2002)

A new phylum of *Archaea* represented by a nanosized hyperthermophilic symbiont. *Nature* 417(6884), 63–67.

- [13] Jahn, U., Gallenberger, M., Paper, W., Junglas, B., Eisenreich, W., Stetter, K.O., Rachel, R., Huber, H. (2008) *Nanoarchaeum equitans* and *Ignicoccus hospitalis*: new insights into a unique, intimate association of two *Archaea*. *J. Bacteriol.* 190(5), 1743–1750.
- [14] Jarett, J.K., Nayfach, S., Podar, M., Inskeep, W., Ivanova, N.N., Munson-McGee, J., Schulz, F., Young, M., Jay, Z.J., Beam, J.P., Kyrpides, N.C., Malmstrom, R.R. (2018) Single-cell genomics of co-sorted *Nanoarchaeota* suggests novel putative host associations and diversification of proteins involved in symbiosis. *Microbiome* 6(161), <https://doi.org/10.1186/s40168-018-0539-8>.
- [15] Jungbluth, S.P., Amend, J.P., Rappé, M.S. (2017) Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. *Sci. Data* 4, 170037, <https://doi.org/10.1038/sdata.2017.37>.
- [16] Mohanty, S., Jobichen, C., Chichili, V.P.R., Velázquez-Campoy, A., Low, B.C., Hogue, C.W.V., Sivaraman, J. (2015) Structural basis for a unique ATP synthase core complex from *Nanoarchaeum equitans*. *J. Biol. Chem.* 290(45), 27280–27296.
- [17] Munson-McGee, J.H., Field, E.K., Bateson, M., Rooney, C., Stepanauskas, R., Young, M.J. (2015) *Nanoarchaeota*, their *Sulfolobales* host, and *Nanoarchaeota* virus distribution across Yellowstone National Park hot springs. *Appl. Environ. Microbiol.* 81(22), 7860–7868.
- [18] Podar, M., Makarova, K.S., Graham, D.E., Wolf, Y.I., Koonin, E. V., Reysenbach,

- A.-L. (2013) Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct* 8, 9, <https://doi.org/10.1186/1745-6150-8-9>.
- [19] Spang, A., Saw, J.H., Jørgensen, S.L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A.E., van Eijk, R., Schleper, C., Guy, L., Ettema, T.J.G. (2015) Complex *Archaea* that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551), 173–179.
- [20] St. John, E., Liu, Y., Podar, M., Stott, M.B., Meneghin, J., Chen, Z., Lagutin, K., Mitchell, K., Reysenbach, A.-L. (In press) A new symbiotic nanoarchaeote (*Candidatus Nanoclepta minutus*) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.*, <https://doi.org/10.1016/j.syapm.2018.08.005>.
- [21] Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M., Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. U. S. A.* 100(22), 12984–12988.
- [22] Wurch, L., Giannone, R.J., Belisle, B.S., Swift, C., Utturkar, S., Hettich, R.L., Reysenbach, A.-L., Podar, M. (2016) Genomics-informed isolation and characterization of a symbiotic *Nanoarchaeota* system from a terrestrial geothermal environment. *Nat. Commun.* 7, 12115, <http://doi.org/10.1038/ncomms12115>.

Appendix A. Supplementary methods

Lipid extraction and analysis

Wet NZ3^T biomass (approx. 0.1 g) was thawed and extracted to recover lipids using a modified Bligh and Dyer method [9]. The cells were suspended in methanol (MeOH; 1 mL) and chloroform (1 mL). Following further tip sonication (5 min) and centrifugation (5 min; 13000 rpm), the organic layer was collected. The residual pellet was reextracted and organic layers were combined. Water (1.35 mL) and 10% aqueous NaCl (0.15 mL) were then added, the resultant biphasic mixture was centrifuged (5 min, 3500 rpm). The organic layer was transferred and dried under a stream of N₂.

The lipid extract was subjected to acidic methanolysis by heating with 5% methanolic HCl at 100°C for 3 h in order to remove polar head groups as described [79]. Resulting core lipids were analyzed by LCMS.

Analysis of core lipids was carried out using a modified version of the method described by Elling et al [22]. Samples dissolved in chloroform/methanol (1:1) were analyzed by UHPLC-ESI-MS on a system comprising of a Shimadzu 8040-LCMS system equipped with photodiode array detector and triple quadrupole low resolution MS utilizing an ESI source for ionization and Q3 scanning from 300 to 2000 *m/z*. LCMS chromatography was conducted using a Waters Acquity BEH C18 1.7 µm, 2x150 mm column at 45°C. A linear gradient elution profile was employed, with solvents comprised of solvent A (MeOH/NH₃/formic acid 100:0.1:0.4) and solvent B (2-propanol/NH₃/formic acid 100:0.1:0.4). Initial elution conditions were 100% solvent A, which was held for 10 min. The percentage of solvent B was then increased to 24% (v/v) at 15 min from sample injection, then increased linearly to 65% at 70 min, to 90% at 75

min. and held at this composition for 5 min before being reduced to initial conditions over 1 min. The column was equilibrated for 9 min prior to a subsequent injection being made. A flow rate of 0.2 ml min⁻¹ was employed.

Lipids were identified by comparison of MS data with published information [69,115].

Appendix B. List of supplementary files

File S1.....

This file contains tables that correspond to data in Chapter 2, as detailed below. The document is approximately 5 megabytes in size and is saved in the extensible markup language (XML) format. Contents can be viewed in Microsoft Office XML Handler.

Table A1. Annotations for the New Zealand nanoarchaeote, strain Ncl-1.

Table A2. Annotations for the New Zealand host organism, strain NZ3^T.

Table A3. Characteristics of intron sequences predicted in nanoarchaeote tRNA genes.

Table A4. Descriptions of *Candidatus* Nanoclepta minutus, *Zestosphaera* gen. nov. and *Zestosphaera tikiterensis* sp. nov., based on the digital protologues CA00029, GA00064 and TA00512, respectively.

File S2.....

This file holds supplementary tables for Chapter 3, as listed below. The XML file is readable in Microsoft Office XML Handler and is approximately 418 megabytes in size.

Table B1. M10-121 gene annotations, produced using RAST and the arCOG database.

Table B2. Annotations for the Guaymas Basin *Nanoarchaeota*, Gua-46, predicted by the RAST server and arCOG database.

Table B3. MC-1 gene annotations, produced using RAST and significant arCOG ($E \leq 0.00001$) gene predictions.

Table B4. Key genomic features of novel *Nanoarchaeota* MAGs and their cultivated relatives.

Table B5. Comparison of predicted transfer RNA introns found in previously described *Nanoarchaeota* genomes and MAGs.

Table B6. Split protein-coding genes identified in previously described *Nanoarchaeota* and MAGs.

Table B7. Genes shared in cultivated *Nanoarchaeota* based on pairwise best-hit BlastP comparisons between each genome.

Table B8. *Cand.* Nanoclepta minutus genes used in core genome analysis.

Table B9. Genes from *Nanoarchaeum equitans* used in the core genome analysis, identified per protocols described in Materials and Methods.

Table B10. *Cand.* Nanopusillus acidilobi genes, predicted as described in Materials and Methods.

Table B11. M10-121 core gene homologues, identified by pairwise best-hit BlastP analysis ($E \leq 0.00001$) comparing M10-121 to each cultivated *Nanoarchaeota* genome.

Table B12. Gua-46 core gene homologues, identified as described in Table B11.

Table B13. MC-1 core gene homologues, identified using methods described in Table B11.

Table B14. Core genes identified in all cultivated *Nanoarchaeota* and all deep-sea vent MAGs.

File S3.....

Concatenated ribosomal protein alignment used to construct Figure 1.8. This text file is 71 kilobytes in size and can be opened with a text editor.