

**BABEȘ-BOLYAI UNIVERSITY
FACULTY OF BIOLOGY AND GEOLOGY
DOCTORAL SCHOOL OF INTEGRATIVE BIOLOGY**

**ASGARD ARCHAEA FROM SALINE SAPROPELS PROVIDE
NEW INSIGHTS INTO THE ARCHAEOAL ANCESTRY OF
EUKARYOTES**

– SUMMARY OF DOCTORAL THESIS –

PhD Student

Paul-Adrian BULZU

Scientific Supervisor

Prof. Dr. Horia Leonard BANCIU

**CLUJ-NAPOCA
2019**

TABLE OF CONTENTS

Preface.....	3
List of abbreviations.....	4
Chapter I: General introduction.....	5
1.1 Eukaryogenesis.....	6
1.2 Emergence of bacteria, archaea and eukaryotes.....	8
1.3 Phagocytosis as a trigger of eukaryogenesis.....	10
1.4 An inventory of available Asgardarchaeota genomes.....	12
1.5 Geographical distribution and environment of Asgardarchaeota.....	13
1.6 Asgardarchaeota phylogenomics.....	14
1.7 Eukaryote Signature Proteins (ESP) are abundant in Asgard archaea.....	15
1.8 Metabolically versatile microbes.....	16
1.9 Aims of the thesis.....	19
Chapter II: Illuminating microbial diversity and function in saline sapropels.....	20
1. Introduction.....	20
2. Materials and methods.....	21
2.1 Sampling sites and methodology.....	21
2.2 Sediment chemical analyses.....	23
2.3 DNA extraction, purification and concentration.....	23
2.4 Sequencing and data preprocessing.....	24
2.5 Abundance estimation for prokaryotes.....	25
2.6 Metagenomes assembly.....	25
2.7 Phylogenetic analyses of SSU rRNA genes.....	25
3. Results and discussion.....	26
3.1 Bacteria diversity and inferred functionality.....	28
3.2 Archaea diversity and inferred functionality.....	35
3.3 Eukaryote diversity and inferred functionality.....	39
4. Conclusions.....	43
5. Acknowledgements.....	44
Chapter III: Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche.....	45
1. Introduction.....	45
2. Materials and methods.....	45
2.1 Sampling.....	45
2.2 Sediment chemical analyses.....	46
2.3 DNA extraction and sequencing.....	47
2.4 Metagenome assembly and binning.....	47

2.5 Genome annotation.....	48
2.6 Phylogenetic trees.....	49
2.7 Multiple sequence alignment of rhodopsins.....	50
2.8 RuBisCO tree reconstruction.....	50
2.9 Primase tree construction	50
2.10 Phylogenies of Heimdallarchaeia glucokinases and kynurenine pathway	51
2.11 Abundance estimation for Loki- and Heimdallarchaeia.....	51
2.12 Environmental distribution of Heimdallarchaeia	51
3. Results and discussion.....	52
3.1 Sediment chemical analyses	52
3.2 Asgardarchaeota phylogenomics.....	52
3.3 Identification of a novel class of rhodopsins in Asgardarchaeota.....	57
3.4 Eukaryotic signature proteins identified in Asgardarchaeota.....	62
3.5 Eukaryotic-specific primase in Heimdallarchaeia.....	63
3.6 Metabolic reconstructions	65
3.7 Environmental distribution of Heimdallarchaeia	75
4. Conclusions	75
5. Acknowledgements	76
Chapter IV: Visualization of Loki- and Heimdallarchaeia (Asgardarchaeota) by fluorescence <i>in situ</i> hybridization and catalyzed reporter deposition (CARD-FISH).....	77
1. Introduction	77
2. Materials and methods.....	77
3. Results and discussion.....	80
4. Conclusions	83
5. Acknowledgements	84
Chapter V: General conclusions and perspectives	85
1. General conclusions	85
2. Perspectives.....	86
References	87
Appendices	109
List of publications included in the thesis as chapters	132
List of publication not included in the thesis.....	132
List of conference participations	133
List of participations as member in research projects	134
General acknowledgements.....	135

List of abbreviations

AAI	–	average amino acid identity
ATP	–	adenosine triphosphate
BLAST	–	basic local alignment search tool
bp	–	base pairs
CARD	–	catalyzed reporter deposition
CoA	–	coenzyme A
CPR	–	candidate phyla radiation
ESP	–	eukaryotic signature protein
FECA	–	First Eukaryotic Common Ancestor
FISH	–	fluorescence <i>in situ</i> hybridization
Ga	–	billion years ago
ISO	–	International Organization for Standardization
kbp	–	thousand base pairs
LACA	–	Last Archaeal Common Ancestor
LECA	–	Last Eukaryotic Common Ancestor
LGT	–	lateral gene transfer
LUCA	–	Last Universal Common Ancestor
Ma	–	million years ago
MAG	–	metagenome-assembled genome
Mbp	–	million base pairs
NAD	–	nicotinamide adenine dinucleotide
NGS	–	next next-generation sequencing
nt	–	nucleotides
ORF	–	open reading frame
PCR	–	polymerase chain reaction
RT	–	room temperature
SAG	–	single-cell amplified genome
SAR	–	Stramenopiles, Alveolata, Rhizaria (supergroup)
SMTZ	–	sulfate-methane transition zone
SSU	–	small subunit (ribosomal RNA)
TCA	–	tricarboxylic acid cycle
THF	–	tetrahydrofuran
WGS	–	Whole Genome Shotgun (sequencing)
WL	–	Wood-Ljungdahl (pathway)

Keywords: Asgardarchaeota, Asgard CARD-FISH, eukaryogenesis, kynurenine pathway, protoeukaryote, schizorhodopsin, saline sediments

Chapter I: General introduction

[Parts of Chapter I were published as: Bulzu, P.-A., Cristea, A., Buda, D.M., Banciu, H.L. (2019). 'Asgardarchaeota – a novel prokaryotic group discovered in aquatic sediments that might shed light on the origin and early evolution of eukaryotes'. *Aerul si Apa: Componente ale Mediului/ Air and Water: Components of the Environment*, ISSN 2067-743X. pp. 435-446. doi: 10.24193/AWC2019_43]

The emergence of eukaryotes followed by their rapid diversification is widely regarded as one of the major events in the history of life (Adl *et al.*, 2012; López-García and Moreira, 2015). Although recent insights place eukaryotes at the top of cellular evolution, stemming from an early interplay between an archaeal host (Eme *et al.*, 2017) and an alphaproteobacterial endosymbiont (Martijn *et al.*, 2018), questions concerning the precise identity and nature of the archaeal partner still persist (Zaremba-Niedzwiedzka *et al.*, 2017; Da Cunha *et al.*, 2018; Spang *et al.*, 2018).

The metagenomics-based discovery of Lokiarchaeia and their phylogenomic placement as direct ancestors of the eukaryotes – a position previously held by members of the TACK superphylum (comprised of the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota phyla) (Guy and Ettema, 2011) – has fueled the already reignited debate about the topology of the Tree of Life (Zaremba-Niedzwiedzka *et al.*, 2017; Da Cunha *et al.*, 2018; Spang *et al.*, 2018). This discovery was soon followed by three other novel candidate archaeal phyla (Odin-, Thor- and Heimdallarchaeia) that together with Lokiarchaeia comprise the recently defined Asgardarchaeota superphylum, or “Asgard archaea”, named after gods from Norse mythology (Zaremba-Niedzwiedzka *et al.*, 2017).

Geographical distribution Asgardarchaeota: At the time of writing (January 2019), the inventory of Asgardarchaeota members consisted of 27 metagenome-assembled genomes (MAGs) deposited in the NCBI Genome database, 3 in the MG-RAST database, and 35 unpublished genomes reported in a preprint (Bulzu *et al.*, 2018). The sixty-five MAGs include 10 Heimdallarchaeia, 29 Lokiarchaeia, 25 Thorarchaeia and one Odinararchaeia genome. Following the discovery of Lokiarchaeia (Spang *et al.*, 2015) in marine sediments from the Arctic Mid-Ocean Ridge, Asgard archaea were identified in a wide range of environments with markedly different biological and chemical parameters (Spang *et al.*, 2015; Seitz *et al.*, 2016; Dombrowski *et al.*, 2017; Zaremba-Niedzwiedzka *et al.*, 2017; Y. Liu *et al.*, 2018; Tully, Graham and Heidelberg, 2018). Although evidence of Asgardarchaeota presence is frequently derived solely based on 16S rRNA genes or individual contigs (Zaremba-Niedzwiedzka *et al.*, 2017; Cai *et al.*, 2018), the list of geographical sites that yielded more or less complete genomes is significantly shorter (Figure 1).

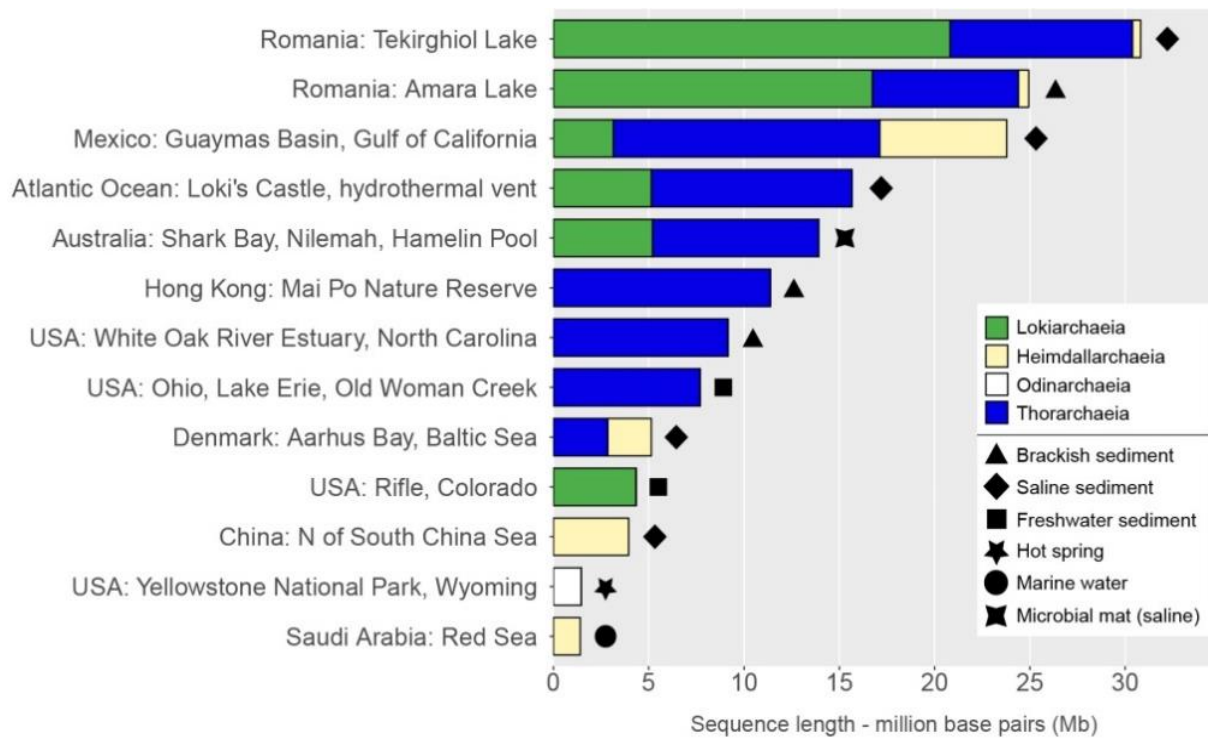


Figure 1. The main geographical sites that have yielded Asgardarchaeota MAGs. Summed contig length is represented by horizontal bars, partitioned according to taxonomic affiliation. Taxonomy and environments are indicated in the legend.

Asgardarchaeota phylogenomics: The few available studies tackling the phylogenetic positioning of Asgardarchaeota lineages relative to eukaryotes as well as to other archaea derive their conclusions from analyses based on: 1) individual or concatenated 16S and 23S rRNA genes (Spang *et al.*, 2015; Seitz *et al.*, 2016; Zaremba-Niedzwiedzka *et al.*, 2017; Y. Liu *et al.*, 2018), 2) concatenated ribosomal proteins conserved between archaea and eukaryotes (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017; this study), 3) concatenated universal marker genes (Da Cunha *et al.*, 2017, 2018; Zaremba-Niedzwiedzka *et al.*, 2017) and 4) individual marker genes - eukaryote signature proteins (ESPs) (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017; Akıl and Robinson, 2018; this study) or other relevant markers such as elongation factor 2 (EF-2) (Narrowe *et al.*, 2018).

Eukaryote Signature Proteins (ESP) are abundant in Asgard archaea: Study of Asgardarchaeota genomes revealed the largest number of genes coding for “eukaryote signature proteins” (ESPs) (Hartman and Fedorov, 2002) detected in any known archaeal lineage (Spang *et al.*, 2015, 2018; Zaremba-Niedzwiedzka *et al.*, 2017). Identified ESP genes encode homologues of eukaryotic proteins acting in membrane maintenance and trafficking, endosomal sorting, ribosome structure, N-glycosylation, protein ubiquitination as well as cytoskeleton structure and dynamics (e.g. profilin, actin, gelsolin, folliculin) (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017; Y. Liu *et al.*, 2018).

Metabolically versatile microbes: It is assumed that knowledge regarding the fine cellular structure and metabolic capabilities of Asgardarchaeota will aid in testing the competing hypotheses on the origin of eukaryotes (Koonin, 2015; Sousa *et al.*, 2016; Bernard *et al.*, 2018) thus ultimately providing valuable insights into the *Archaea-to-Eukaryota* transition. However, in lack of cultivable representatives, this scientific endeavor has relied exclusively upon state-of-the-art cultivation-independent metagenomics approaches.

- **Aims of the thesis**

In the light of recent research, eukaryotes emerged from the interplay between an archaeal host sharing close ancestry with extant Asgardarchaeota members and an alphaproteobacterial endosymbiont that eventually gave rise to mitochondria. Since their original discovery in the vicinity of hot hydrothermal vents in the Northern Atlantic, a number of partial or complete Asgard archaea genomes have been recovered in a wide range of environments by next-generation sequencing approaches. However, the paucity of recovered genomes and incomplete nature of the few available ones had precluded previous attempts to perform in-depth phylogenetic or metabolic characterizations for this clade. Noteworthy, Asgard archaea became the focus point of this thesis after their discovery in organic-rich sediments (i.e. sapropels) of Amara and Tekirghiol lakes, during the exploratory phase of our studies.

The main objectives of this thesis were:

1. To describe and compare the microbial diversity and their inferred involvement in the biogeochemical processes possibly occurring in the sapropels of three Romanian lakes with markedly different salinities (**Chapter II**).
2. To resolve the evolutionary history of Asgardarchaeota with respect to eukaryotes, other archaea clades and between different lineages comprising this superphylum (**Chapter III**).
3. To unravel the metabolic networks and putative eukaryotic-like cellular features of individual Asgardarchaeota lineages (**Chapter III**).
4. To provide the first images of Asgardarchaeota cells by using the CARD-FISH technique (**Chapter IV**).

Chapter II: Illuminating microbial diversity and function in saline sapropels

[This chapter is currently under the form of a manuscript in preparation; parts of the Methods section from this chapter were published as: Bulzu, P.-A., Andrei, A.-Ş., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R. and Banciu, H. L. (2019) 'Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche', *Nature Microbiology*. doi: 10.1038/s41564-019-0404-y.]

1. Introduction

Although widely disputed among different branches of environmental scientists, the term “sapropel” (gr., *sapros* = rotten, *pelos* = soil) (Stein, 2014) is broadly used to describe organic carbon-enriched (>2% of dry weight), dark-colored, fine-grained sediments formed in stratified stagnant water bodies under reducing conditions (Kidd, Cita and Ryan, 1978; de Lange and ten Haven, 1983; Stein, 2014). Sapropels typically accumulate in water bodies where detrital remains of aquatic and/or terrestrial macrophytes, phytoplankton and zooplankton are subjected to complex anaerobic microbe-mediated transformations and mixed with fine mineral particles (e.g. carbonates, silicates) (Stankevica *et al.*, 2015; Yermolaeva *et al.*, 2016). While studies focusing on ancient marine sapropels have pictured microbial assemblages typical of the subsurface biosphere (Sass and Parkes, 2011; Parkes *et al.*, 2014), little is known about microbial communities dwelling in recently formed saline sapropels. In this regard, state-of-the-art culture-independent techniques, combined with detailed physico-chemical analyses, hold the potential to unravel the diversity, structure, and biogeochemical roles of such organisms. In this study we aim to investigate the microbial diversity and its role in the genesis of organic rich-sediments from three Romanian lakes with markedly different characteristics.

2. Materials and methods

Amara, Tekirghiol and Ursu are naturally-formed Romanian lakes defined by markedly different salinities, hydric regimes, trophic state, local climate and genesis mechanisms (Alexe, 2010; Gheorghievici *et al.*, 2012). Notably, all three lakes harbor large deposits of organic-rich sediments (i.e. sapropels) pinpointing them as suitable models for deciphering microbiological diversity driving the genesis of saline sapropels.

Sediment sampling: Was performed for chemical and metagenomics analyses in Romanian lakes Tekirghiol, Amara and Ursu. In Tekirghiol Lake, sampling was performed in the shallow shore area using a custom sediment corer while in Amara and Ursu lakes, increased water depth impeded the use of the hand-held sediment corer and therefore sampling was performed using a Petite Ponar dredge (Wildco, Saginaw, MI, USA).

Environmental DNA extraction: Was performed from sapropel samples using a dedicate isolation kit following instructions provided by the manufacturer.

DNA whole genome shotgun (WGS) sequencing: Was performed by a commercial company. The amount of total raw sequence data generated for each metagenome was: 64.5 Gbp - Amara, 57.6 Gbp - Tekirghiol and 59.6 - Ursu.

Microbial diversity and relative abundance estimation: Preprocessed Illumina WGS sequencing sets were randomly subsampled and putative small subunit ribosomal RNA gene sequences were queried in each subset by scanning with UBLAST (Edgar, 2010) against the latest non-redundant SILVA database (v. 132, clustered at 85% identity) (Pruesse *et al.*, 2007). Resulting bona fide 16S and 18S rRNA gene sequences were compared by blastn (Altschul *et al.*, 1990) (e-value <1e-5) against the curated SILVA database (v. 132, clustered at 99%). Sequences assigned to major prokaryotic phyla were used to calculate their abundances (expressed as percent “%” of total SSU rRNA gene reads) in their originating environments.

Assembling and annotating full-length small subunit rRNA gene sequences: Preprocessed Illumina paired-end sequencing sets generated for all three samples were individually *de novo* assembled using Megahit (Li *et al.*, 2016) v.1.1.1 with default parameters and the following list of k-mers: 39, 49, 69, 89, 109, 129, 149. Sequences encoding ribosomal RNAs were retrieved from contigs with SSU-ALIGN (Nawrocki, 2009), length filtered and uploaded to the SILVA SINA online alignment and classification service (Pruesse, Peplies and Glockner, 2012). Classification files with assigned taxonomy were retrieved and multiple sequence alignment files including original query sequences and inferred neighbours were further used to construct maximum-likelihood phylogenies with IQ-TREE (Nguyen *et al.*, 2015).

3. Results and conclusions

Taxonomic classifications based on 16S rRNA gene fragments from unassembled metagenomics reads revealed the presence of 34 distinct prokaryotic phyla (28 from *Bacteria* and 6 from *Archaea*) with estimated phylum-level relative abundances $\geq 0.5\%$ in at least one of the three sampled sites (Figure 2).

The structure and composition of sapropel microbial communities reflected the distinctive limnological properties inherent to each sampled lake. Prokaryotic fractions in all sapropels were dominated by bacteria, judging by overall abundance as well as diversity of phyla. Nevertheless, an increasing tendency for Archaea presence was noticeable in the direction of decreasing salinity: 7.5% in Ursu, 12.8% in Tekirghiol and 14.3% in Amara sapropel. Multiple prokaryotic phyla lacking isolated representatives were identified in all three samples.

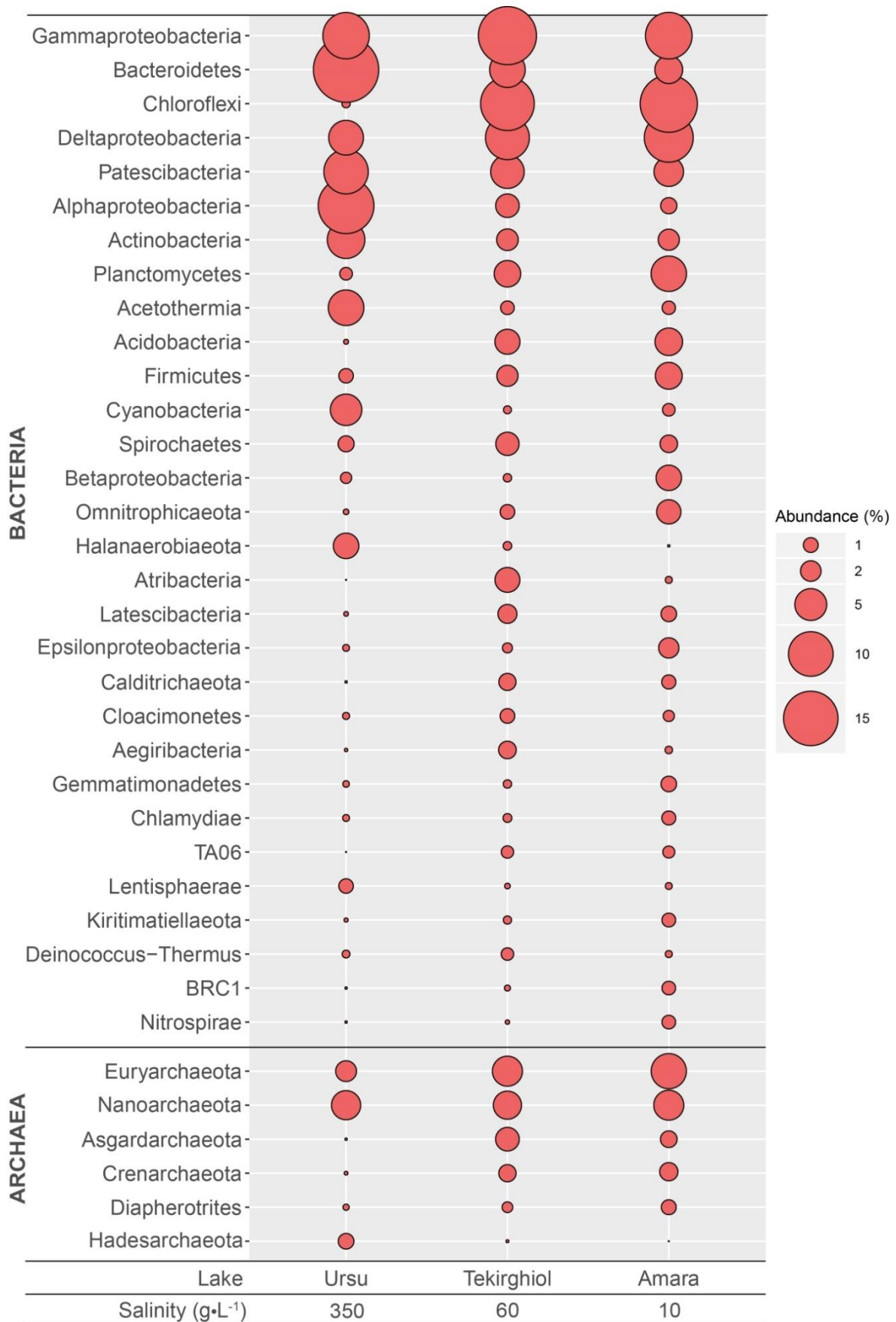


Figure 2. Relative abundances of phyla (and classes within Proteobacteria) comprising $\geq 0.5\%$ of the microbial community in at least one dataset, estimated using SSU rRNA unassembled gene reads.

Particularly, 16S rDNA-level evidence for the presence of elusive Asgard archaea in Amara and Tekirghiol sapropels prompted further genome-level investigations, now published as a separate study (Bulzu *et al.*, 2019). Regardless of varying salinities, eco-physiological inferences showed that sapropel dwelling microbial communities were mainly involved in organic matter degradation and sulfur cycling.

Eukaryotes were classified based on 18S rRNA gene sequences detected in sapropel samples. Identified taxa covered all trophic levels, ranging from green algae and macrophytes to macroinvertebrates. The allochthonous input of organic matter from circumlacustrine plants and from autochthonous algae and/or macroinvertebrates to the process of sapropel formation was also confirmed by this approach. Notably, most eukaryotic taxa were certainly light and/or oxygen dependent water column dwellers, thus owing their detection to favorable DNA preservation conditions. Potential sapropel dwellers included fungi and benthic detritivorous invertebrates that likely interact with the local prokaryotic communities and facilitate organic matter decomposition.

Chapter III: Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche

[Part of this chapter was published as: Bulzu, P.-A., Andrei, A.-Ş., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R. and Banciu, H. L. (2019) 'Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche', Nature Microbiology. ISSN: 2058-5276, doi: 10.1038/s41564-019-0404-y.]

1. Introduction

The discovery of Asgard archaea, the closest extant relative of eukaryotes to date, has reignited the two-domain of life debate. While it is apparent that Asgard archaea encode multiple eukaryotic-specific proteins, the lack of genomic information and metabolic characterization has precluded inferences about the closest eukaryotic ancestor and the metabolic landscape that laid the grounds for the emergence of the hallmark eukaryotic subcellular architecture. Here, we employ state-of-the-art metagenomics, sensitive phylogenetic analyses and genome-scale metabolic reconstructions with the aim of shedding light upon the deep archaeal ancestry of eukaryotes.

2. Materials and methods

Sampling: Amara and Tekirghiol are naturally-formed shallow lakes in South-Eastern Romania that harbor large deposits of organic-rich sediments (or 'sapropels'). Sediment for exploratory chemical and metagenomics analyses was collected in Tekirghiol and Amara lakes in October 2017. The successful recovery of Asgardarchaeota genomes prompted a second sampling

campaign for fine chemical profiling in April 2018, in Tekirghiol Lake, on the site of the previous sampling.

Sediment chemical analyses: Were performed on both mixed samples (0 – 40 cm Tekirghiol and 0 – 10 cm Amara) taken in 2017 and the vertical profile obtained from Tekirghiol Lake in 2018.

DNA extraction and sequencing: DNA was extracted from approximately 10 g of wet mixed sediment samples (0 – 40 cm Tekirghiol and 0 – 10 cm Amara) collected in 2017, by using the DNeasy PowerMax Soil Kit (Qiagen, Hilden, Germany).

Metagenome assembly and binning: *De novo* assembly of preprocessed paired-end Illumina reads was done by Megahit (Li *et al.*, 2015) v.1.1.1 with default parameters. Assembled contigs were binned by a combination of taxonomy-dependent and -independent methods. Bin completeness and contamination were estimated using CheckM (Parks *et al.*, 2015). Final curated bins with estimated completeness above 10% and contamination below 3% were denominated as metagenome assembled genomes (MAGs).

Genome annotation: Protein coding sequences were predicted *de novo* with Prokka (Seemann, 2014) for all available Asgard MAGs (35 from this study, 14 from NCBI/Genome). Metabolic inferences were guided by KEGG functional annotations. InterProScan (v.5.25) (Jones *et al.*, 2014) was used to annotate protein domains. Potential eukaryote signature proteins (ESP) were identified based on previously published lists (Zaremba-Niedzwiedzka *et al.*, 2017).

Phylogenetic trees: A total of 131 taxa were considered for concatenated small subunit (SSU) and larger subunit (LSU) ribosomal RNA phylogenetic analyses. A total of 93 taxa were considered for concatenated ribosomal protein phylogenomic analyses. rRNA as well as protein sequences were aligned independently by PRANK (Loytynoja, 2014), trimmed using BMGE (Criscuolo and Gribaldo, 2010) and concatenated within each type of sequence (rRNA or protein). Maximum-likelihood phylogeny was inferred using IQ-TREE with ultrafast bootstrapping (-bb 1000) and significance testing (-alrt 1000) (Nguyen *et al.*, 2015; Hoang *et al.*, 2018). Bayesian phylogenies were constructed using PhyloBayes MPI 1.8 (Lartillot *et al.*, 2013), with the CAT-Poisson model.

Multiple sequence alignment of rhodopsins: The three groups of rhodopsins (type-1, schizorhodopsins and heliorhodopsins), were first aligned independently using T_Coffee (Notredame, Higgins and Heringa, 2000) (<http://tcoffee.crg.cat/>) in accurate mode, that employs protein structure information, wherever available, or sequence comparisons with homologues in

databases to improve accuracy. These individual alignment were aligned to each other using the profile alignment mode in T_Coffee.

RuBisCO tree reconstruction: MUSCLE (Edgar, 2004) was used for aligning the sequences (n=146) of the large subunit of RuBisCO (types I-III) and RuBisCO-like (type IV) (rbcL, K01601) proteins. Sequences not generated in this study were recovered from previous studies (Tabita *et al.*, 2007; Wrighton *et al.*, 2016). Maximum-likelihood trees were constructed with FastTree2 (Price, Dehal and Arkin, 2010) using a JTT model, a gamma approximation, and 100 bootstrap replicates.

Primase tree construction: All publicly available Archaea genomes (n=3242; 1st of March 2019) were downloaded from the NCBI Genome section. Additionally, 22 eukaryotic protein sequences for small (PriS) and large (PriL) subunits of DNA primase were retrieved from the UniProt database (<https://www.uniprot.org/>). InterProScan was used locally to annotate predicted proteins. Sequences containing domains appertaining to primase subunits were retrieved from predicted proteomes. The two subsets were aligned using PRANK (-protein +F). Single subunit maximum-likelihood trees were generated by IQ-TREE (-bb 1000, -alrt 1000) with ultrafast bootstrapping.

Phylogenies of Heimdallarchaeia glucokinases and kynurenine pathway: ADP-dependent phosphofructokinase/glucokinase protein sequences were identified by their assigned KO number (K00918) in 3 MAGs (AMARA_4, Heimdall_AB_125, Heimdall_LC_3). Retrieved sequences were used along with 49 other sugar kinases published in a previous study (Castro-Fernandez *et al.*, 2017). Aminoacid sequences of enzymes within the kynurenine pathway - tryptophan 2,3-dioxygenase (TDO), kynurenine 3-monooxygenase (KMO) and 3-hydroxyanthranilate 3,4-dioxygenase (HAAO) – that were identified only in Heimdallarchaeia MAGs, were used along with sequences of corresponding enzymes from 12 Eukaryotes and 15 Bacteria retrieved from NCBI RefSeq. MAFFT-L-INS-i (Katoh, 2002) (default parameters) and PRANK (parameters: -PROTEIN +F) were used for aligning enzyme sequences, followed by trimming using BMGE (-m BLOSUM30 -t AA -g 0.5 -b 3). Single marker maximum-likelihood trees were constructed with FastTree2, using 100 standard bootstrap replicates.

3. Results and discussion

Sediment chemical analyses: The analysis of leachable salt contents of Tekirghiol (0-40cm) and Amara (0-10cm) sediments indicated that the dominant cations and anions ($\text{g} \cdot \text{Kg}^{-1}$) were: Na^+ (16.5 and 7.0), K^+ (1.0 and 0.22), Mg^{2+} (1.1 and 4.0), Cl^- (27.7 and 11.2), and SO_4^{2-} (0.25 and 13.2).

Asgardarchaeota phylogenomics: Homology-based searches were employed to recover Asgardarchaeota-related contigs from *de novo* metagenomic assemblies of two deep-sequenced, shallow, brackish lake sediment samples (sediment pore-water salinities: 5.7% in Tekirghiol Lake and 3.9% in Amara Lake). By utilizing a hybrid binning strategy and performing manual inspection and data curation, we obtained 11 high- and medium-quality (>50% completeness; <2% contamination) and 24 low completeness (<50% completeness; <3% contamination) MAGs, spanning 3 (out of four) evolutionary lineages within the superphylum. The maximum likelihood phylogenetic tree, based on concatenation of small (SSU) and large (LSU) ribosomal RNA genes, pictured a topology in which eukaryotes branched with high-support from within Asgardarchaeota (Archaea). Even more remarkably, in addition to recreating a previously described Asgardarchaeota/*Eukaryota* branching pattern (Zaremba-Niedzwiedzka *et al.*, 2017), we provide support for a close evolutionary linkage between Heimdallarchaeia and eukaryotes.

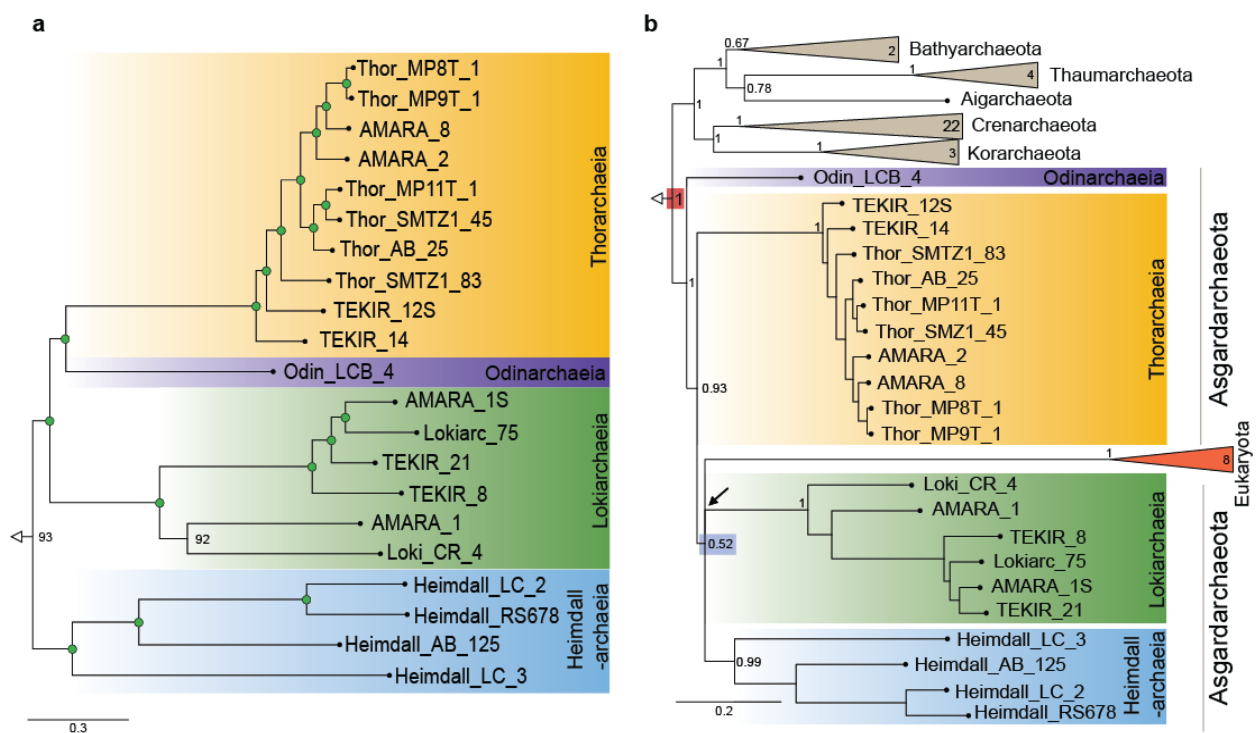


Figure 3. Asgardarchaeota phylogenomics. **a)** Maximum likelihood (LG+C60, general matrix and 60-profile protein models of amino acid substitution) phylogeny of the Asgardarchaeota superphylum. The green circles highlight UFBoot values higher than 95. **b)** Asgardarchaeota phylogeny generated through Bayesian inference (CAT-Poisson, CAT model of amino acid substitution with uniform global exchange rates). The posterior probability values are shown above the internal nodes. The black arrow indicates the unresolved position of Lokiarchaeia.

The genome-focused phylogeny of Asgardarchaeota revealed a pattern of ancestry, divergence, and descent, in which Heimdallarchaeia comprise the basal branch of the superphylum and Thor-/Odinararchaeia the youngest one (Figure 3a). Although dissimilar in branching pattern with the SSU + LSU tree, the phylogenomic one was found to be robust

(Figure 3a) and to support a topology brought into attention by an earlier study (Zaremba-Niedzwiedzka *et al.*, 2017). The SR4-recoded (Susko and Roger, 2007) Bayesian tree (maxdiff=0.1) resolved with high support the monophyly of Asgardarchaeota/*Eukaryota*, but failed to confidently resolve the internal topology of the superphylum and the branching point of eukaryotes (Figure 3b).

Regarding Thorarchaeia, the phylogenomic trees (Figure 3a, b) showed that the basal branches of this phylum were represented by MAGs recovered from the Tekirghiol hypersaline sediment (i.e. TEKIR_14 and TEKIR_12S). The other ones form a compact cluster (n=8), which appeared to be the outcome of a more recent diversification event in brackish environments. Noteworthy, this reduced phylogenomic diversity within Thorarchaeia contrasts with the highly divergent MAGs of Loki- and Heimdallarchaeia (Figure 3a, b).

Identification of a novel class of rhodopsins in Asgardarchaeota: Recent findings reporting the presence of a previously uncharacterized family of rhodopsins (Pushkarev *et al.*, 2018) (i.e. heliorhodopsins; abbreviated as HeR) in monoderms (Flores-Uribe *et al.*, 2019) prompted a dedicated screening in all available Asgardarchaeota MAGs. The results indicated that the Heimdallarchaeota RS678 MAG encoded two HeR and what appears to be a type-1 proton-pumping rhodopsin (Figure 4), suggestive of light sensitivity. Remarkably, we found that the Asgardarchaeota MAGs recovered during this study encoded rhodopsin sequences similar in membrane orientation to type-1 rhodopsins, and which organized during phylogenetic analysis in a monophyletic clade placed in-between HeR and type-1 rhodopsins (Figure 4).

Given their phylogenetically intermediate position, as type-1 rhodopsins closest to HeR, and presence of features found in both type-1 and HeR, we denote them as **schizorhodopsins** (schizo, ‘split’, plus ‘rhodopsin’, abbreviated as SzR). The very recent discovery of HeR and their inconclusive functional role (Pushkarev *et al.*, 2018; Flores-Uribe *et al.*, 2019) precludes tentative functional assertions for SzR capacity in Asgardarchaeota. However, the plethora of rhodopsins that we identified in Heimdallarchaeia (putative type-1 proton pumps, HeR and SzR), together with the SzR found in Loki- and Thorarchaeia suggests that, during their evolutionary history, Asgardarchaeota were present in light-exposed habitats.

Eukaryotic signature proteins identified in Asgardarchaeota: To further substantiate the phylogenetic connections between Asgardarchaeota members and eukaryotes, we screened all recovered MAGs and the publicly available ones (14 in July 2018) for the presence of potential eukaryotic signature proteins (ESP).

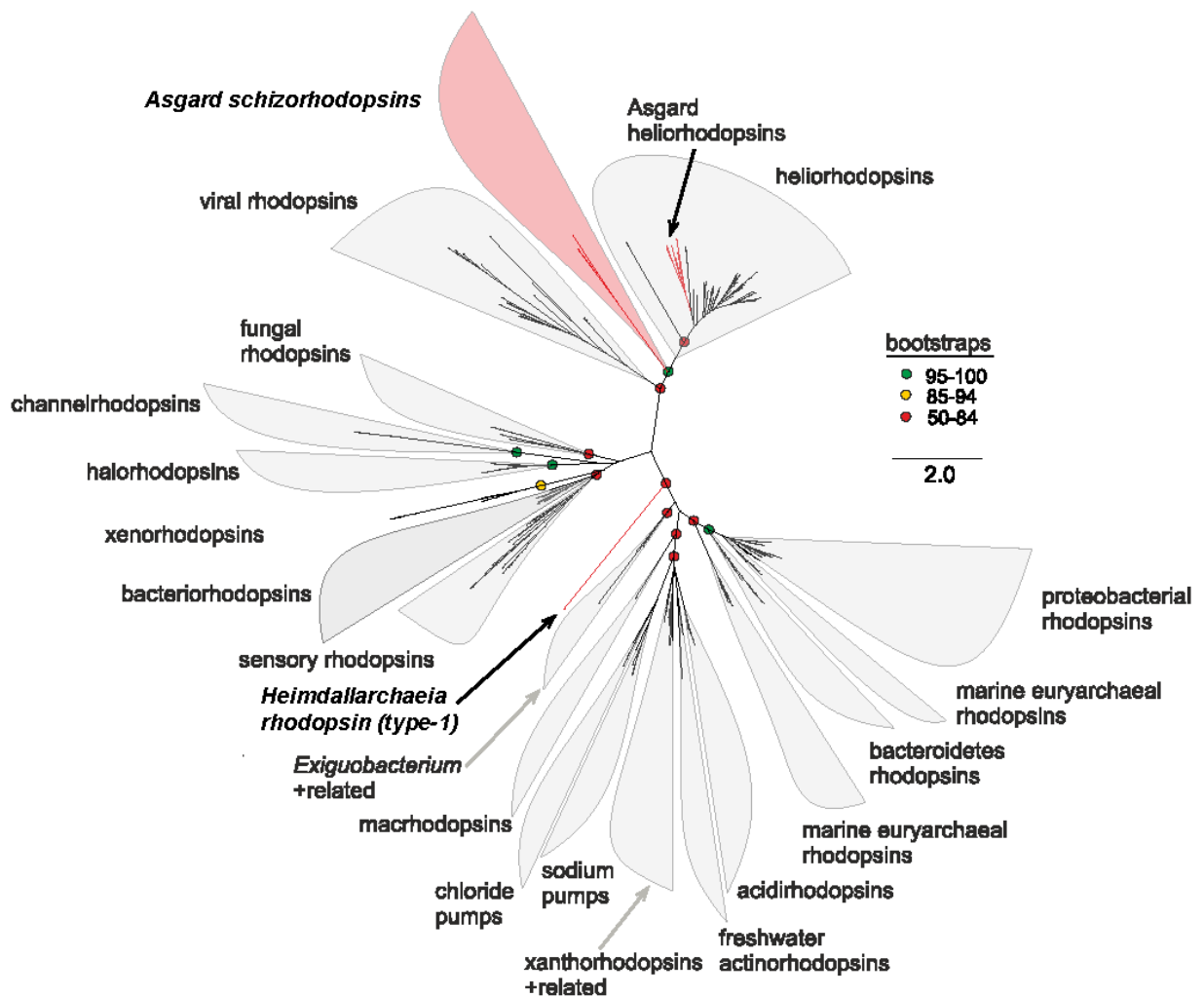


Figure 4. Phylogenetic analysis of rhodopsins. An unrooted maximum likelihood tree of all Asgardarchaeota schizorhodopsins (n=6) identified in this work, heliorhodopsins and representative known type-1 rhodopsins, is shown. The branches colored red are sequences from the Asgardarchaeota. Bootstrap values on nodes are indicated by colored circles (see color key at the right). A total of 392 sequences, spanning known rhodopsin families and including schizorhodopsins retrieved in this study, were used for phylogenetic inference.

Similar to previous reports (Spang *et al.*, 2015; Zaremba-Niedzwiedzka *et al.*, 2017; Liu *et al.*, 2018), the MAGs were found to be highly enriched in ESP, which further reinforced their ancestral linkage to eukaryotes. In addition to the reported ESP (Zaremba-Niedzwiedzka *et al.*, 2017), we identified a potential subunit of the COPII vesicle coat complex (associated with intracellular vesicle traffic and secretion) in Thorarchaeia and proteins that harbor the N-terminal domain of folliculin - a eukaryote-specific protein in Lokiarchaeia. Furthermore, we retrieved conclusive hits for the ESP-related domains Ezrin/radixin/moesin C-terminal domain and ELKS in Lokiarchaeia.

Eukaryotic-specific primase in Heimdallarchaeia Following protein annotation, it became apparent that Loki- and Thorarchaeia encode a typical archaeal DNA primase while a distinct,

eukaryote-specific, homologue was particular to Heimdallarchaeia. This intriguing observation prompted further phylogenetic analyses taking into account DNA primases from all archaeal orders along with representatives of major eukaryotic lineages and all available Asgardarchaeotes.

Metabolic reconstructions: The genome-scale metabolic reconstruction placed special emphasis on Heimdallarchaeia (Figure 5), since it was suggested by phylogenetic analyses to encompass the most probable candidates for the archaeal protoeukaryote ancestor. While the anaerobic lifestyles inferred for Loki- (Sousa *et al.*, 2016) and Thorarchaeia (Liu *et al.*, 2018) were considered to be accompanied by autotrophy (Sousa *et al.*, 2016) and mixotrophy (Liu *et al.*, 2018), respectively, no consistent metabolic reconstructions existed for Heimdallarchaeia. The physiology inferred here pointed towards mixotrophic lifestyles (for Asgardarchaeota), simultaneously showing the presence of transporters for the uptake of exogenous organic matter and the metabolic circuitry responsible for its catabolism.

While all analyzed Asgardarchaeota phyla encode components of the glycolytic pathway, three Heimdallarchaeia MAGs (LC_3, AB_125 and AMARA_4) were found to employ non-canonical ADP-dependent kinases and, as previously noted (Liu *et al.*, 2018), no glucokinase homologue could be identified in Thorarchaeia. We reason that the well represented non-oxidative PPP in this group could either represent an alternative point of entry for sugars in the EMP, or that the function of canonical glucokinase is achieved by yet unidentified archaea-specific sugar kinases (Brasen *et al.*, 2014).

The identified homologues for ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes were found to appertain to types III (Loki- and Heimdallarchaeia) and IV (Loki- and Thorarchaeia). We consider that the MAGs encoding type III-like RuBisCO utilize the nucleotide monophosphate degradation pathway (Kono *et al.*, 2017), performing CO₂ fixation by linking nucleoside catabolism to glycolysis/gluconeogenesis. This conclusion is supported by the co-occurrence of genes encoding for: RuBisCO type III, AMP phosphorylases, ribose 1,5-bisphosphate isomerases, and carbonic anhydrases.

As nicotinamide adenine dinucleotide (NAD⁺) is an essential cofactor in redox biochemistry and energetics (Ternes and Schönknecht, 2014), we investigated its *de novo* synthesis mechanisms. As expected, all Asgardarchaeota phyla harbored the ubiquitous aspartate pathway - a set of metabolic transformations that can occur both in presence or absence of oxygen (Gazzaniga *et al.*, 2009). Additionally, Heimdallarchaeia also encoded the exclusively aerobic kynurenine pathway of NAD⁺ biosynthesis, present in few bacteria and eukaryotes (Ternes and Schönknecht, 2014). The phylogenetic inferences showed that this pathway,

considered to be present in the protoeukaryote ancestor (Ternes and Schönknecht, 2014), was probably acquired by the ancestor of Heimdallarchaeia through LGT from bacteria. As far as the authors are aware, Heimdallarchaeia are the first reported archaeal organisms harboring the aerobic kynurenine pathway.

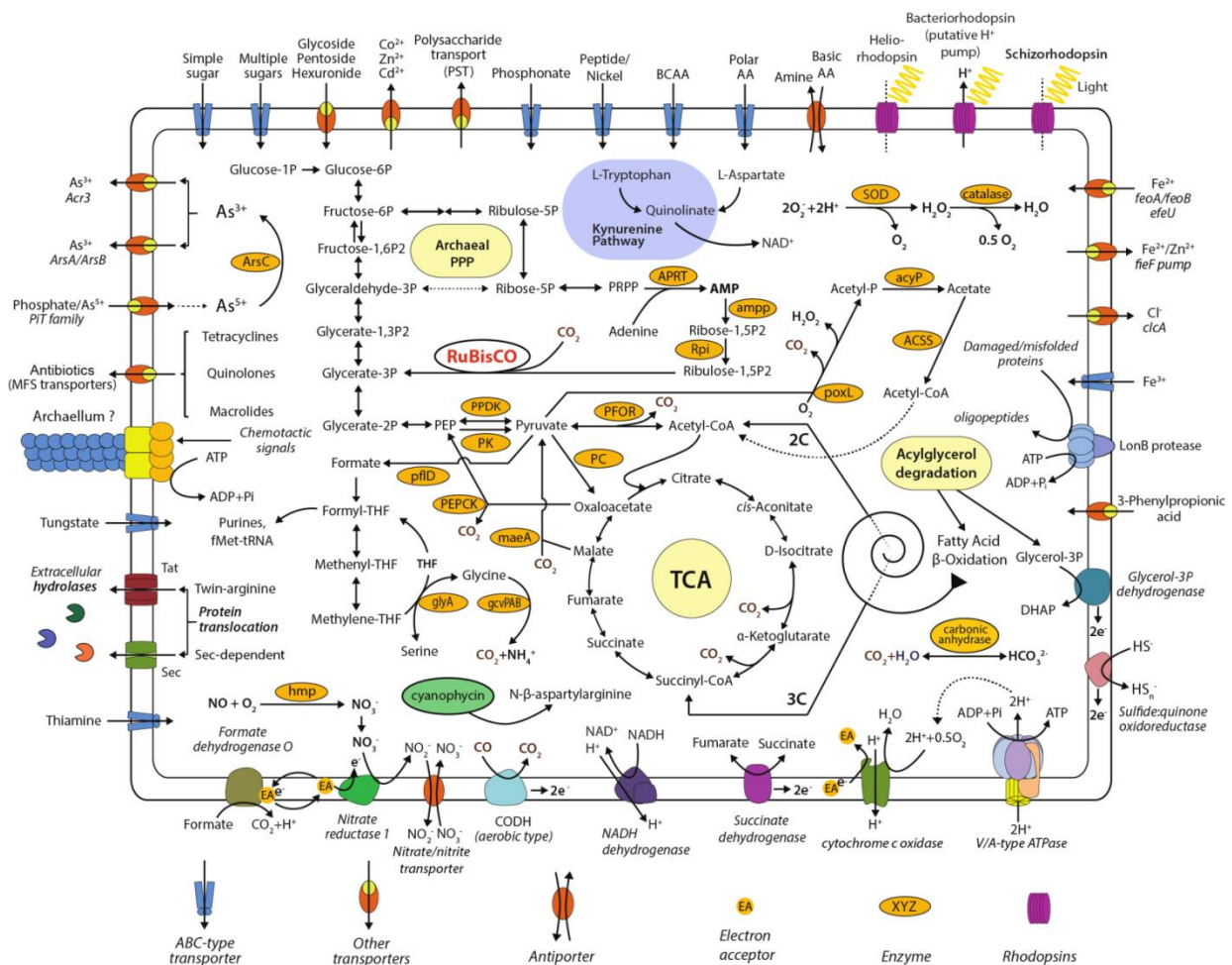


Figure 5. Metabolic reconstruction of Heimdallarchaeia. The text in the yellow panels depicts names of pathways and metabolic processes. Abbreviations: ACSS - acetyl-CoA synthetase and carbonic anhydrase; acyP - acylphosphatase; ampp - AMP phosphorylase; APRT – AMP pyrophosphorylase; ArsC - arsenate reductase (glutaredoxin); BCAA - branched-chain amino acid; CODH - carbon monoxide dehydrogenase; gcvPAB – glycine dehydrogenase; glyA – glycine hydroxymethyltransferase; hmp – nitric oxide dioxygenase; maeA - malate dehydrogenase (decarboxylating); PC - pyruvate carboxylase; PEPCK - phosphoenolpyruvate carboxykinase; pflD - formate C-acetyltransferase; PFOR - pyruvate ferredoxin oxidoreductase; PK - pyruvate kinase; poxL – pyruvate oxidase; PPK - pyruvate, phosphate dikinase; PPP - pentose phosphate pathway; Rpi - ribose-5-phosphate isomerase; RuBisCO - Ribulose-1,5-bisphosphate carboxylase/oxygenase; SOD - superoxide dismutase; TCA - tricarboxylic acid cycle.

Among all analyzed phyla, the complete TCA cycle was identified in Loki- and Heimdallarchaeia. Additionally, Lokiarchaeia contain key genes (isocitrate dehydrogenase, 2-oxoglutarate-ferredoxin oxidoreductase, ATP-citrate lyase) that are indicative of a reductive TCA cycle, involved in the autotrophic fixation of CO₂. Regarding oxidative phosphorylation, while V/A-type ATPase appears mostly complete in Loki- and Thorarchaeia, the other

components involved in oxidative phosphorylation, are missing or incomplete, emphasizing anaerobic lifestyles. For Heimdallarchaeia we could identify complete V/A-type ATPase, succinate dehydrogenase, almost complete NADH:quinone oxidoreductase and importantly – cytochrome c oxidase – another hallmark of aerobiosis.

4. Conclusions

Here, we propose that Heimdallarchaeia (phylum within Asgardarchaeota) are the closest extant relatives to all eukaryotes and shed light on their facultative aerobic lifestyle, characterized by the capacity to use Sun's energy and metabolic pathways unique among archaea.

Chapter IV: Visualization of Loki- and Heimdallarchaeia (Asgardarchaeota) by fluorescence *in situ* hybridization and catalyzed reporter deposition (CARD-FISH)

[Part of this chapter was published as a preprint article (unreviewed): Salcher, M. M., Andrei, A.-Ş., Bulzu, P.-A., Keresztes, Z. G., Banciu, H. L. and Ghai, R. (2019) 'Visualization of Loki- and Heimdallarchaeia (Asgardarchaeota) by fluorescence in situ hybridization and catalyzed reporter deposition (CARD-FISH)', bioRxiv, p. 580431. doi: 10.1101/580431.]

1. Introduction

Metagenome-assembled genomes (MAGs) of Asgardarchaeota are starting to be recovered from a variety of habitats, broadening their environmental distribution and providing access to the genetic makeup of this archaeal lineage. Despite their singular phylogenetic position at the base of the eukaryotic tree of life, the morphology of these bewildering organisms remains a mystery. In order to visualize this elusive group, we applied a combination of CARD-FISH and epifluorescence microscopy on coastal hypersaline sediment samples, using specifically designed CARD-FISH probes for Heimdallarchaeia and Lokiarchaeia lineages and provide the first visual evidence for both these groups. Here, we show that while Heimdallarchaeia are characterized by a uniform cellular morphology typified by central DNA localization, Lokiarchaeia display a plethora of shapes and sizes that likely reflect their broad phylogenetic diversity and ecological distribution.

2. Materials and methods

Oligonucleotide probe design: We manually optimized the alignment of all 16S sequences classified as Asgardarchaeota in ARB (Ludwig *et al.*, 2004) using the latest SILVA database (v.132, clustered at 99% identity) (Pruesse *et al.*, 2007) and constructed a RAxML tree (GTR-GAMMA model, 100 bootstraps (Stamatakis, Ludwig and Meier, 2005)) for all high-quality near

full-length sequences. All attempts to construct a general probe targeting all Asgardarchaeota sequences (n=935) failed, likely due to the high diversity of this superphylum. We designed three distinct competitor probes for Heimdallarchaeia and two for Lokiarchaeia. Each competitor was used in the same concentrations as the CARD-FISH probes in order to prevent non-specific binding.

Testing of probes: Probes were tested *in silico* (Yilmaz, Parnerkar and Noguera, 2011) and in the laboratory with different formamide concentrations in the hybridization buffer until stringent conditions were achieved.

CARD-FISH and cell visualization: We tested these probes in sediment samples from two sites from where recently several Asgardarchaeota genomes were recovered by metagenomics (Lakes Tekirghiol and Amara, Romania) (Bulzu *et al.*, 2018). Seven sediment layers (0-70 cm, in 10-cm ranges) were sampled in Lake Tekirghiol, and the top 10 cm was sampled in Lake Amara. Samples were fixed with formaldehyde, treated by sonication, vortexing and centrifugation to detach cells from sediment particles and aliquots were filtered onto white polycarbonate filters (0.2 μm pore size, Millipore). CARD-FISH was conducted as previously described (Ishii *et al.*, 2004). Filters were counterstained with DAPI and inspected by epifluorescence microscopy.

3. Results and discussion

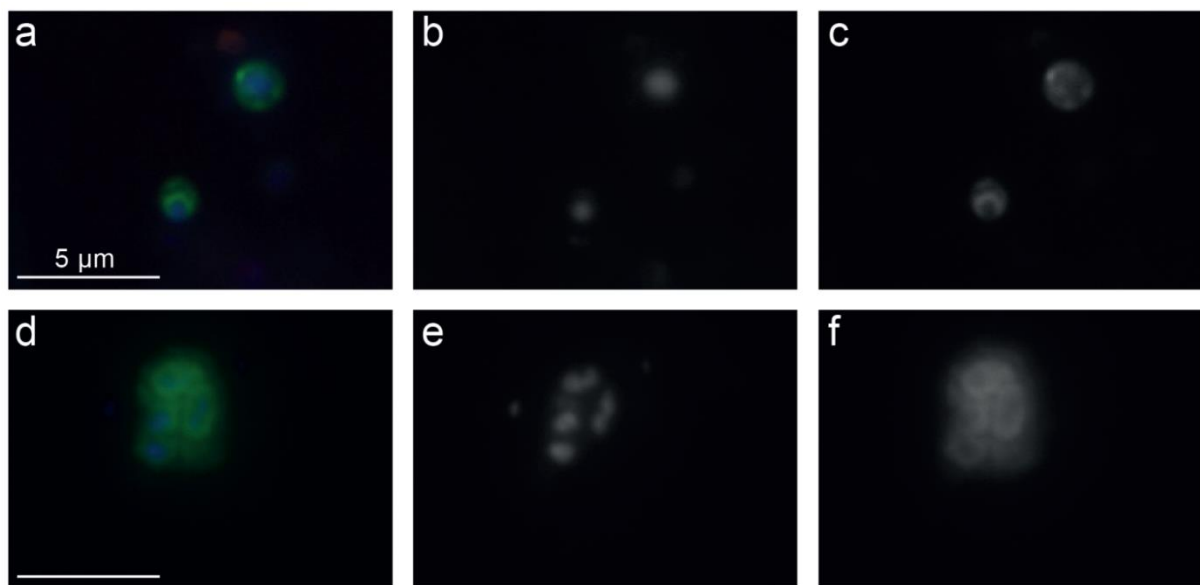


Figure 6. CARD-FISH imaging of Heimdallarchaeia hybridized with probe heim-526. The left panels (**a**, **d**) display overlay images of probe signal (green), DAPI staining (blue) and autofluorescence (red), the middle panels (**b**, **e**) DAPI staining of DNA, the right panels (**c**, **f**) CARD-FISH staining of proteins. The scale bar (5 μm) in the left images applies to all microphotographs.

Both phyla were rare and appeared completely absent below depths of 40 cm. All observed Heimdallarchaeia were similar in cell size (2.0 ± 0.5 μm length x 1.4 ± 0.4 μm width,

n=23) and of conspicuous shape with DNA condensed ($0.8\pm0.2 \times 0.5\pm0.2 \mu\text{m}$) at the center of the cells (Figure 6). In contrast, Lokiarchaeia presented diverse shapes and sizes and we could distinguish at least three distinct morphotypes: Most common were small-medium sized ovoid cells ($2.0\pm0.5 \times 1.4\pm0.3 \mu\text{m}$, n=30, Figure 7 a-c) that were found in different depth layers in Lake Tekirghiol (0-10 cm, 10-20 cm, 20-30 cm) and in the top 10 cm sample from Lake Amara. A single large round cell ($3.8 \times 3.6 \mu\text{m}$, Figure 7 d-f) with bright fluorescence signal and condensed DNA at the center was detected in Lake Amara, and large rods/filaments ($12.0\pm4.3 \times 1.4\pm0.5 \mu\text{m}$, n=6, Figure 7 g-i) with filamentous, condensed DNA ($10.2\pm4.8 \times 0.6\pm0.1 \mu\text{m}$) were present at 30-40 cm sediment depth in Lake Tekirghiol and in 0-10 cm depth in Lake Amara.

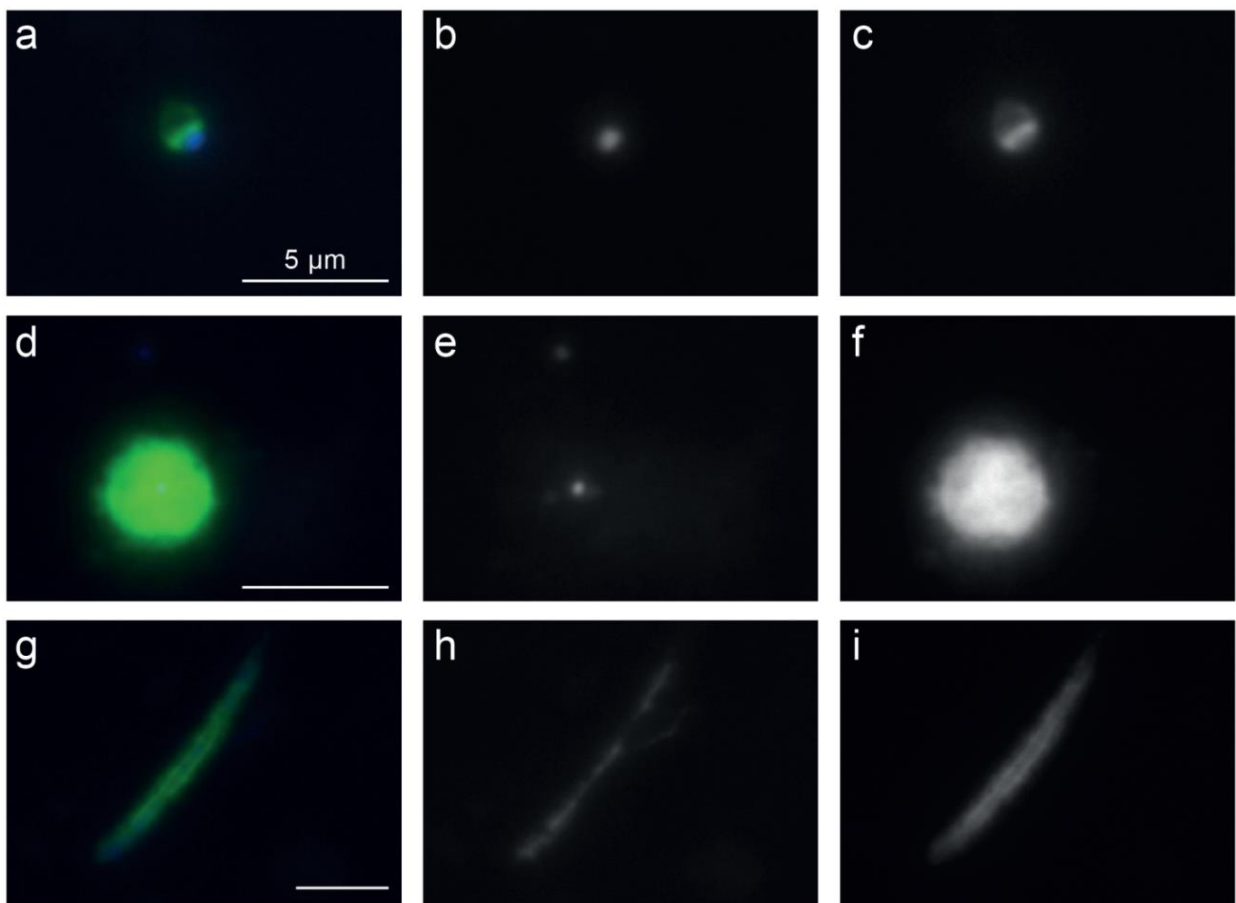


Figure 7. CARD-FISH imaging of Lokiarchaeia hybridized with probe loki1-1184. Three different morphotypes are displayed: **(a-d)** small-medium sized ovoid cells, **(d-f)** large round cell, and **(g-i)** large filamentous cells. The left panels **(a, d, g)** display overlay images of probe signal (green), DAPI staining (blue) and autofluorescence (red), the middle panels **(b, e, h)** DAPI staining of DNA, the right panels **(c, f, i)** CARD-FISH staining of proteins. The scale bar (5 μm) in the left images applies to all microphotographs.

The variety of Lokiarchaeia morphologies most likely reflects the higher sampling of the phylogenetic diversity within this phylum. During microscopic inspections, we carefully checked for potential non-specific or autofluorescent signals at wavelengths not interfering with the probe signal and found no overlap for any of the inspected cells. A set of negative controls were

conducted to rule out false-positive signals due to unspecific binding of dye or nucleic acid components of probes by using a nonspecific probe (Wallner, Amann and Beisker, 1993).

To avoid false-positive signals from cellular peroxidases, we performed additional control experiments including the CARD reaction only. All these control treatments resulted in low, unspecific background signals, but no obvious staining of cells. Further evidence of specificity was seen in all cells hybridized with the Heimdallarchaeia probe, both the shapes and staining patterns coupled to DAPI were remarkably consistent.

4. Conclusions

Examining these first images of Asgardarchaeota, we could not avoid to note the unusual happenstance of their naming with shapes and ecology. Inspired by their first detection in the deep, steaming hot vents of ‘Loki’s castle’ off the Norwegian coast (Spang *et al.*, 2015), their initial baptism after mythological characters from Norse mythology (Odin, Thor, Loki and Heimdall (Zaremba-Niedzwiedzka *et al.*, 2017)) has been unusually prescient.

Chapter V: General conclusions and perspectives

The research detailed in this PhD thesis combines state-of-the-art high-throughput environmental DNA (eDNA) sequencing with bioinformatics and physicochemical analyses to describe the biodiversity in saline, organic-rich sediments (i.e. sapropels) and to unravel the lifestyles and evolutionary history of Asgard archaea.

The first study (Chapter II) illustrated the biodiversity and inferred functionality of organisms contributing to sapropel formation in three Romanian lakes with different physicochemical parameters. This is the first attempt to compare microbial groups dwelling in recently formed lacustrine sapropels by using culture-independent methods. Results pinpointed multiple input sources for detrital organic matter in the process of sapropel genesis. Bacteria dominated the prokaryotic communities in all tree sampled sapropels (from Amara, Tekirghiol and Ursu lakes). Rare and/or uncultivated taxa comprising the so-called “microbial dark matter” were detected from both prokaryotic domains of life. The presence of enigmatic Asgard archaea prompted further investigations focused on this clade.

The second study (Chapter III) employed genome-resolved metagenomics to reconstruct and characterize Asgard archaea from Amara and Tekirghiol sapropels. By using *de novo* assembly and hybrid binning strategies, thirty-five partial genomes were reconstructed, thus effectively doubling the amount of genomic data available for this clade. Phylogenomic analyses

confirmed the intriguing eocyte topology noted in previous studies, depicting eukaryotes emerging from within Asgardarchaeota. A detailed metabolic reconstruction was for the first time performed for Heimdallarchaeia revealing their potential for mixotrophy and capacity for aerobiosis. We identified a previously unknown family of rhodopsins within Asgard archaea which we denominated as “schizorhodopsins”. By corroborating phylogenomic and metabolic data we show that Heimdallarchaeia are the closest archaeal ancestors of eukaryotes and propose the “aerobic-protokaryote”.

The third study (Chapter IV) focused on obtaining the first images of Loki- and Heimdallarchaeia cells. CARD-FISH combined with an epifluorescence microscopy approach showed that all imaged Heimdallarchaeia were similar with respect to their cell size and shape while Lokiarchaeia cells could be assigned to at least three distinct morphotypes. The observed morphological diversity was attributed to low sequence sampling available for Asgardarchaeota at the time experiments were conducted.

Selective references

- Adl, S. M., Simpson, A. G. B., Lane, C. E., Lukes, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A. D., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A. and Spiegel, F. W. (2012) 'The revised classification of eukaryotes.', *J. Eukaryot. Microbiol.*, 59(5), pp. 429–493.
- Akıl, C. and Robinson, R. C. (2018) 'Genomes of Asgard archaea encode profilins that regulate actin', *Nature*, 562(7727), pp. 439–443.
- Alexe, M. (2010) 'Studiul lacurilor sărate din Depresiunea Transilvaniei', Ed. Presa Universitară Clujeană: Cluj-Napoca, Romania, (book in Romanian).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool.', *J. Mol. Biol.*, 215(3), pp. 403–410.
- Bernard, G., Pathmanathan, J. S., Lannes, R., Lopez, P. and Baptiste, E. (2018) 'Microbial Dark Matter Investigations: How Microbial Studies Transform Biological Knowledge and Empirically Sketch a Logic of Scientific Discovery', *Genome Biol. Evol.*, 10(3), pp. 707–715.
- Brasen, C., Esser, D., Rauch, B. and Siebers, B. (2014) 'Carbohydrate metabolism in *Archaea*: current insights into unusual enzymes and pathways and their regulation.', *Microbiol. Mol. Biol. Rev.*, 78(1), pp. 89–175.
- Bulzu, P.-A., Andrei, A.-S., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R. and Banciu, H. (2018) 'The sunlit microoxic niche of the archaeal eukaryotic ancestor comes to light', *bioRxiv*, p. 385732, doi: <https://doi.org/10.1101/385732>
- Bulzu, P.-A., Andrei, A.-Ş., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R. and Banciu, H. L. (2019) 'Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche', *Nat. Microbiol.*, <https://doi.org/10.1038/s41564-019-0404-y>.
- Cai, M., Liu, Y., Zhou, Z., Yang, Y., Pan, J., Gu, J.-D. and Li, M. (2018) 'Asgard archaea are diverse, ubiquitous, and transcriptionally active microbes', *bioRxiv*, p. 374165, doi: <https://doi.org/10.1101/374165>
- Castro-Fernandez, V., Herrera-Morande, A., Zamora, R., Merino, F., Gonzalez-Ordenes, F., Padilla-Salinas, F., Pereira, H. M., Brandao-Neto, J., Garratt, R. C. and Guixé, V. (2017) 'Reconstructed ancestral enzymes reveal that negative selection drove the evolution of substrate specificity in ADP-dependent kinases.', *J. Biol. Chem.*, 292(51), p. 21218.
- Criscuolo, A. and Gribaldo, S. (2010) 'BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments', *BMC Evol. Biol.*, 10(1), p. 210.
- Da Cunha, V., Gaia, M., Gadelle, D., Nasir, A. and Forterre, P. (2017) 'Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes', *PLOS Genet.*, 13(6), p. e1006810.
- Da Cunha, V., Gaia, M., Nasir, A. and Forterre, P. (2018) 'Asgard archaea do not close the debate about the universal tree of life topology', *PLOS Genet.*, 14(3), p. e1007215.

- Dombrowski, N., Seitz, K. W., Teske, A. P. and Baker, B. J. (2017) ‘Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments.’, *Microbiome.*, 5(1), p. 106.
- Edgar, R. C. (2004) ‘MUSCLE: multiple sequence alignment with high accuracy and high throughput’, *Nucleic Acids Res.*, 32(5), pp. 1792–1797.
- Edgar, R. C. (2010) ‘Search and clustering orders of magnitude faster than BLAST.’, *Bioinformatics.*, 26(19), pp. 2460–2461.
- Eme, L., Spang, A., Lombard, J., Stairs, C. W. and Ettema, T. J. G. (2017) ‘Archaea and the origin of eukaryotes.’, *Nat. Rev. Microbiol.*, 15(12), pp. 711–723.
- Flores-Uribe, J., Hevroni, G., Ghai, R., Pushkarev, A., Inoue, K., Kandori, H. and Béjà, O. (2019) ‘Heliorhodopsins are absent in diderm (Gram-negative) bacteria: Some thoughts and possible implications for activity’, *Environ. Microbiol. Rep.*, <https://doi.org/10.1111/1758-2229.12730>.
- Gazzaniga, F., Stebbins, R., Chang, S. Z., McPeck, M. A. and Brenner, C. (2009) ‘Microbial NAD Metabolism: Lessons from Comparative Genomics’, *Microbiol. Mol. Biol. Rev.*, 73(3), pp. 529–541.
- Gheorghievici, L. M., Pompei, I., Gheorghievici, G. and Tănase, I. (2012) ‘The influence of abiotic factors on suppliers of organic matter in the peloidogenesis process from Lake Techirghiol, Romania’, *AAEL Bioflux.*, 5(2), pp. 401–405.
- Guy, L. and Ettema, T. J. G. (2011) ‘The archaeal “TACK” superphylum and the origin of eukaryotes.’, *Trends Microbiol.*, 19(12), pp. 580–587.
- Hartman, H. and Fedorov, A. (2002) ‘The origin of the eukaryotic cell: A genomic investigation’, *Proc. Natl. Acad. Sci. U.S.A.*, 99(3), p. 1420 LP-1425.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. and Vinh, L. S. (2018) ‘UFBoot2: Improving the ultrafast bootstrap approximation’, *Mol. Biol. Evol.*, 35(2), pp. 518–522.
- Ishii, K., Mußmann, M., MacGregor, B. J. and Amann, R. (2004) ‘An improved fluorescence *in situ* hybridization protocol for the identification of bacteria and archaea in marine sediments.’, *FEMS Microbiol. Ecol.*, 50(3), pp. 203–213.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R. and Hunter, S. (2014) ‘InterProScan 5: genome-scale protein function classification.’, *Bioinformatics.*, 30(9), pp. 1236–1240.
- Katoh, K. (2002) ‘MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform’, *Nucleic Acids Res.*, 30(14), pp. 3059–3066.
- Kidd, R. B., Cita, M. B. and Ryan, W. B. F. (1978) ‘Stratigraphy of eastern Mediterranean sapropel sequences recovered during DSDP Leg 42A and their paleoenvironmental significance’, *Initial Reports DSDP.*, 42A, pp. 421–443.
- Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., Mizohata, E., Inoue, T., Hasunuma, T., Yokota, A., Matsumura, H. and Ashida, H. (2017) ‘A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea’, *Nat. Commun.*, 8, p. 14007.

- Koonin, E. V (2015) ‘Archaeal ancestors of eukaryotes: not so elusive any more.’, *BMC Biol.*, 13, p. 84.
- de Lange, G. J. and ten Haven, H. L. (1983) ‘Recent sapropel formation in the eastern Mediterranean’, *Nature*, 305(5937), pp. 797–798.
- Lartillot, N., Rodrigue, N., Stubbs, D. and Richer, J. (2013) ‘PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment.’, *Syst. Biol.*, 62(4), pp. 611–615.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K. and 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), pp. 1674–1676.
- Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., Yamashita, H. and Lam, T.-W. (2016) ‘MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices.’, *Methods.*, 102, pp. 3–11.
- Liu, Y., Zhou, Z., Pan, J., Baker, B. J., Gu, J.-D. and Li, M. (2018) ‘Comparative genomic inference suggests mixotrophic lifestyle for Thorarchaeota’, *ISME J.*, 12(4), pp. 1021–1031.
- López-García, P. and Moreira, D. (2015) ‘Open questions on the origin of eukaryotes’, *Trends Ecol. Evol.*, 30(11), pp. 697–708.
- Loytynoja, A. (2014) ‘Phylogeny-aware alignment with PRANK.’, *Methods Mol. Biol.*, 1079, pp. 155–170.
- Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A. and Schleifer, K.-H. (2004) ‘ARB: a software environment for sequence data.’, *Nucleic Acids Res.*, 32(4), pp. 1363–1371.
- Martijn, J., Vosseberg, J., Guy, L., Offre, P. and Ettema, T. J. G. (2018) ‘Deep mitochondrial origin outside the sampled alphaproteobacteria.’, *Nature.*, 557(7703), pp. 101–105.
- Narrowe, A. B., Spang, A., Stairs, C. W., Caceres, E. F., Baker, B. J., Miller, C. S. and Ettema, T. J. G. (2018) ‘Complex evolutionary history of translation Elongation Factor 2 and diphthamide biosynthesis in *Archaea* and parabasalids’, *Genome Biol. Evol.*, 10(9), pp. 2380–2393.
- Nawrocki, E. P. (2009) *Structural RNA Homology Search and Alignment using Covariance Models*. Washington University School of Medicine. PhD Thesis, <http://eddylab.org/publications/Nawrocki09b/Nawrocki09b-phdthesis.pdf>.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. and Minh, B. Q. (2015) ‘IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies’, *Mol. Biol. Evol.*, 32(1), pp. 268–274.
- Notredame, C., Higgins, D. G. and Heringa, J. (2000) ‘T-Coffee: A novel method for fast and accurate multiple sequence alignment.’, *J. Mol. Biol.*, 302(1), pp. 205–217.

- Parkes, R. J., Cragg, B., Roussel, E., Webster, G., Weightman, A. and Sass, H. (2014) ‘A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere:geosphere interactions’, *Mar. Geol.*, 352, pp. 409–425.
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. and Tyson, G. W. (2015) ‘CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.’, *Genome Res.*, 25(7), pp. 1043–1055.
- Price, M. N., Dehal, P. S. and Arkin, A. P. (2010) ‘FastTree 2--approximately maximum-likelihood trees for large alignments.’, *PLoS One.*, 5(3), p. e9490.
- Pruesse, E., Peplies, J. and Glockner, F. O. (2012) ‘SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes.’, *Bioinformatics.*, 28(14), pp. 1823–1829.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J. and Glockner, F. O. (2007) ‘SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.’, *Nucleic Acids Res.*, 35(21), pp. 7188–7196.
- Pushkarev, A., Inoue, K., Larom, S., Flores-Uribe, J., Singh, M., Konno, M., Tomida, S., Ito, S., Nakamura, R., Tsunoda, S. P., Philosof, A., Sharon, I., Yutin, N., Koonin, E. V, Kandori, H. and Beja, O. (2018) ‘A distinct abundant group of microbial rhodopsins discovered using functional metagenomics.’, *Nature.*, 558(7711), pp. 595–599.
- Sass, H. and Parkes, R. J. (2011) ‘Sub-seafloor sediments: An extreme but globally significant prokaryotic habitat (Taxonomy, Diversity, Ecology) BT - Extremophiles Handbook’, in Horikoshi, K. (ed.). Tokyo: Springer Japan, pp. 1015–1041.
- Seemann, T. (2014) ‘Prokka: rapid prokaryotic genome annotation.’, *Bioinformatics.*, 30(14), pp. 2068–2069.
- Seitz, K. W., Lazar, C. S., Hinrichs, K.-U., Teske, A. P. and Baker, B. J. (2016) ‘Genomic reconstruction of a novel, deeply branched sediment archaeal phylum with pathways for acetogenesis and sulfur reduction.’, *ISME J.*, 10(7), pp. 1696–1705.
- Sousa, F. L., Neukirchen, S., Allen, J. F., Lane, N. and Martin, W. F. (2016) ‘Lokiarchaeon is hydrogen dependent.’, *Nat. Microbiol.*, 1, p. 16034.
- Spang, A., Eme, L., Saw, J. H., Caceres, E. F., Zaremba-Niedzwiedzka, K., Lombard, J., Guy, L. and Ettema, T. J. G. (2018) ‘Asgard archaea are the closest prokaryotic relatives of eukaryotes’, *PLOS Genet.*, 14(3), p. e1007080.
- Spang, A., Saw, J. H., Jørgensen, S. L., Zaremba-Niedzwiedzka, K., Martijn, J., Lind, A. E., van Eijk, R., Schleper, C., Guy, L. and Ettema, T. J. G. (2015) ‘Complex archaea that bridge the gap between prokaryotes and eukaryotes’, *Nature.*, 521, p. 173.
- Stamatakis, A., Ludwig, T. and Meier, H. (2005) ‘RAxML-II: a program for sequential, parallel and distributed inference of large phylogenetic trees’, *Concurr. Comput. Pract. Exp.*, 17(14), pp. 1705–1723.
- Stankevica, K., Kalnina, L., Klavins, M., Cerina, A., Ustupe, L. and Kaup, E. (2015) ‘Reconstruction of the Holocene palaeoenvironmental conditions accordingly to the multiproxy sedimentary records from Lake Pilvelis, Latvia’, *Quat. Int.*, 386, pp. 102–115.
- Stein, R. (2014) ‘Sapropel’, in Harff, J., Meschede, M., Petersen, S., and Thiede, J. (eds) *Encyclopedia of Marine Geosciences*, Springer Heidelberg, pp. 746–747.

- Susko, E. and Roger, A. J. (2007) 'On reduced amino acid alphabets for phylogenetic inference.', *Mol. Biol. Evol.*, 24(9), pp. 2139–2150.
- Tabita, F. R., Hanson, T. E., Li, H., Satagopan, S., Singh, J. and Chan, S. (2007) 'Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs.', *Microbiol. Mol. Biol. Rev.*, 71(4), pp. 576–599.
- Ternes, C. M. and Schönknecht, G. (2014) 'Gene Transfers Shaped the Evolution of De Novo NAD⁽⁺⁾ Biosynthesis in Eukaryotes', *Genome Biol. Evol.*, 6(9), pp. 2335–2349.
- Tully, B. J., Graham, E. D. and Heidelberg, J. F. (2018) 'The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans', *Sci. Data.*, p. 170203.
- Wallner, G., Amann, R. and Beisker, W. (1993) 'Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms.', *Cytometry.*, 14(2), pp. 136–143.
- Wrighton, K. C., Castelle, C. J., Varaljay, V. A., Satagopan, S., Brown, C. T., Wilkins, M. J., Thomas, B. C., Sharon, I., Williams, K. H., Tabita, F. R. and Banfield, J. F. (2016) 'RubisCO of a nucleoside pathway known from *Archaea* is found in diverse uncultivated phyla in bacteria.', *ISME J.*, 10(11), pp. 2702–2714.
- Yermolaeva, N. I., Zarubina, E. Y., Romanov, R. E., Leonova, G. A. and Puzanov, A. V (2016) 'Hydrobiological conditions of sapropel formation in lakes in the south of Western Siberia', *Water Resour.*, 43(1), pp. 129–140.
- Yilmaz, L. S., Parnerkar, S. and Noguera, D. R. (2011) 'mathFISH, a web tool that uses thermodynamics-based mathematical models for *in silico* evaluation of oligonucleotide probes for fluorescence *in situ* hybridization.', *Appl. Environ. Microbiol.*, 77(3), pp. 1118–1122.
- Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Backstrom, 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. and Ettema, T. J. G. (2017) 'Asgard archaea illuminate the origin of eukaryotic cellular complexity.', *Nature.*, 541(7637), pp. 353–358.

List of publications included in the thesis as chapters

Chapter I

Bulzu, P.-A., Cristea, A., Buda, D.M., Banciu, H.L. (2019). ‘Asgardarchaeota – a novel prokaryotic group discovered in aquatic sediments that might shed light on the origin and early evolution of eukaryotes’. *Aerul si Apa: Componente ale Mediului/ Air and Water: Components of the Environment*, ISSN 2067-743X. pp. 435-446. doi: 10.24193/AWC2019_43

Chapter III

Bulzu, P.-A., Andrei, A.-Ş., Salcher, M. M., Mehrshad, M., Inoue, K., Kandori, H., Beja, O., Ghai, R. and Banciu, H. L. (2019) ‘Casting light on Asgardarchaeota metabolism in a sunlit microoxic niche’, *Nature Microbiology*. ISSN: 2058-5276, doi: 10.1038/s41564-019-0404-y. (AIS₂₀₁₇=7.515; IF₂₀₁₇=14.174)

Chapter IV

Salcher, M. M., Andrei, A.-Ş., Bulzu, P.-A., Keresztes, Z. G., Banciu, H. L. and Ghai, R. (2019) ‘Visualization of Loki- and Heimdallarchaeia (Asgardarchaeota) by fluorescence *in situ* hybridization and catalyzed reporter deposition (CARD-FISH)’, *bioRxiv*, p. 580431, doi: 10.1101/580431. (preprint article – not peer reviewed)