

Introduction and major aim

All living cells can be assigned to one of three major domains of life: Bacteria, Eukarya and Archaea. Eukaryotes comprise both unicellular and multicellular organisms, the latter of which include Animals, Plants and Fungi that make up the visible biosphere. Even though bacteria are unicellular and can only be observed under the microscope, they have been studied extensively during the past century and comprise a wide variety of different major lineages. In contrast, unicellular archaea are much less investigated and have long been thought to represent the least diverse domain of life. Just three years ago, only five archaeal phyla were known, the Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota (collectively referred to as **TACK** superphylum) (Gut and Ettema, 2011) and Euryarchaeota. However, there was accumulating evidence that various additional archaeal lineages may inhabit anoxic environments such as deep-sea sediments as well as aquatic habitats (Figure 1). In addition, pioneering studies have shown that archaea are ecologically important and have played a major role in the origin of eukaryotes. For instance, there was increasing support for the evolution of the eukaryotic cell from an archaeal ancestor that engulfed an alpha-proteobacterial endosymbiont eventually evolving into mitochondria.

The major aim of my proposal was to make use of the most recently developed and powerful culture independent sequencing techniques such as single cell genomics and metagenomics, to obtain and analyse genome sequences of potentially novel archaeal phyla. This should increase our understanding of archaeal phylogenetic diversity, ecological importance, geographical distribution as well as their evolution and relationship to eukaryotes.

Description of the work carried out to achieve the project's objectives

In collaboration with various research groups, we have obtained sediment samples from diverse regions all around the world, including for instance hot spring sediments from Yellowstone National park, marine sediments from Århus bay and hydrothermal vent fields (e.g. at Loki's Castle and off the coast of Japan), as well as river and aquifer sediments from the USA. From these sediments, community DNA was extracted and sequenced and subsequently, the microbial community was analyzed. Using sophisticated bioinformatics approaches, genomic reads could be assembled and resulting contigs were assigned to individual genomes. Using this approach, we were able to obtain various genome sequences from uncultivated and novel members of the Archaea. Furthermore, we also obtained three novel archaeal genomes using a single cell genomics approach, in which single cells are sorted, sequenced and assembled individually. A major focus of my work was then to analyze those genomes using diverse bioinformatics approaches in order to get insights into the biology and evolution of the respective organisms. For instance, protein coding genes were identified and carefully annotated to determine the functional potential of these novel archaea. For this, proteins of interest (e.g. proteins involved in studied pathways) were studied by assigning clusters of orthologous groups, analyzing the presence and arrangement of protein domains (e.g. PFAM and IPR domains), performing phylogenetic analyses, investigating operon and gene cluster arrangements and by homology modeling to predict protein structures.

In addition, extensive and careful phylogenomic approaches were used to place the novel archaeal genomes in an updated tree of life and determine their relationship to eukaryotes. Our analyses are based on maximum likelihood and Bayesian phylogenetic analyses using sophisticated models of sequence evolution, that minimize various phylogenetic artifacts. For instance, determining the relationships of members of the three domains of life relative to each other is a challenging problem, because evolutionary distances are extremely large while reliable datasets are small and difficult to identify.

The main results

Our metagenomics and single cell genomics approaches allowed us to obtain various novel archaeal genomes that represent previously unknown archaeal lineages of high taxonomic rank. First of all, using single cell genomics combined with metagenomics, we could extend the TACK superphylum by one relative of Korarchaeum cryptophilum as well as a two members of the Bathyarchaeota. The analysis of these genomes was published in Saw et al., 2015.

1) Discovery of a novel archaeal phylum *Lokiarchaeum* – the closest relative of Eukaryotes known so far. The major focus of my work was however the obtainment of genomes from members of the DSAG and AAG archaeal lineages as well as of more deeply branching Korarchaeota using a purely metagenomics approach (Figure 1). From 16SrRNA taq sequencing, we knew that DSAG archaea are abundant in Loki's Castle hydrothermal vent field sediments comprising up to 10% of the microbial community. Therefore, we decided to obtain our first metagenomics sequencing data from these sediment samples from which we could reconstruct a composite genome of the first member of the DSAG named *Lokiarchaeum* (Spang et al., 2015). The lokiarchaeal genome hold various surprises: First of all, *Lokiarchaeota* appeared to represent the closest archaeal sister lineage of Eukaryotes (Figure 2). Furthermore, its genome encodes various proteins that are most closely related to eukaryotic homologs and are absent from previously sequenced archaeal and bacterial genomes. These proteins, which we referred to as eukaryotic signature proteins (ESPs) include for instance *bona fide* actins, components of all ESCRT complexes (Figure 2) and a large amount of small Ras and Arf superfamily GTPases (Klinger et al., 2016). Notably, in eukaryotic cells, these proteins are involved in fundamental cellular processes

such as endomembrane and vesicular trafficking machineries or are essential for actin cytoskeletons and ubiquitin systems. The identification of these ESPs in an archaeal lineage, which appeared to represent the most closely known relative of the archaeal ancestor of Eukaryotes, strongly indicated that the eukaryotic cell was much more complex than previously anticipated. Thus, the discovery of *Lokiarchaeum* (Spang et al., 2015) has fundamentally changed our perception of the evolution of eukaryotes and has attracted unprecedented media attention from all around the world. Our findings have been publicized in major news outlets such as New York Times, BBC, and National Geographic and the paper was ranked as the number one paper among all papers published from Uppsala University and number two among all papers published from Sweden in 2015 based on Altmetric scores.

2) Extension of the Lokiarchaeota by the description of a novel archaeal superphylum comprise of four archaeal phyla. Building on our acquired expertise in metagenomics, we were subsequently able to obtain 3 novel genomes of deep-branching Koarchaeota and 8 novel genomes of Lokiarchaeota-related lineages from various sediment samples all over the world. Most importantly, this has allowed us to uncover the ASGARD superphylum, which is comprised of Lokiarchaeota, Thorarchaeota, Odinarchaeota and Heimdallarchaeota (Figure 3a) (Zaremba-Niedzwiedzka *et al., under revision in Nature*). ASGARDs group together with eukaryotes in phylogenetic analyses and encode a large amount of ESPs (Figure 3b). Notably, some of these ESPs represent novel key components of the eukaryotic trafficking machinery and were not found in the genome of *Lokiarchaeum*. Indeed, ASGARD archaea contain all fundamental building blocks assumed to have been key for the evolution of eukaryotic complexity. In addition, initial analyses of the central carbon and energy metabolism in these lineages revealed interesting novel insights into the metabolic diversity of a novel archaeal superphylum (Spang et al., in prep.). Therefore, our most recent work on ASGARD archaea further extends and corroborates our initial findings and is expected to represent an additional major discovery of global significance.

Conclusions

The known archaeal diversity has increased tremendously during the past two years. Today, we know that archaea comprise at least four superphyla (DPANN, ASGARD, TACK and Euryarchaeota) each of which is comprised of various phyla. With the discovery of the ASGARD superphylum and its four phyla, our work made an important contribution to this increased knowledge of archaeal phylogenetic diversity. In addition, the ASGARD superphylum has proven to be key for our understanding of the evolution of eukaryotes as this archaeal phylum represents the closest related clade of the previously elusive archaeal ancestor of eukaryotes.

Certainly, prospective analyses of the metabolic potential of this novel archaeal superphylum will yield additional insights into the evolution of metabolism in archaea and help to disentangle the ecological role of ASGARD lineages whose members are abundant in anoxic sediments all over the world. For instance, first indications suggest that members of this group encode proteins that may be of economic and ecological significance as they may degrade substrates that are major pollutants in our ecosystems. Future studies of these archaea are therefore not only of interest for basic but also applied natural sciences and may have wider societal applications than anticipated.

Figure 1. Phylogeny of uncultivated diversity in the archaeal TACK superphylum based on environmental 16S rDNA sequences (from Guy and Ettema, 2011, Trends in Microbiology). The deeply diverging Korarchaeota, DSAG and AAG lineages, which we targeted in our research are highlighted with purple arrows.

Figure 2: A schematic tree depicting the placement of *Lokiarchaeum* - our first representative of DSAG archaea – in the tree of life and an overview of ESPs detected in the different archaeal phyla and Eukaryotes (from Spang et al., 2015). *Lokiarchaeum* contains a large amount of novel ESPs, previously not found in other archaeal lineages.

Figure 3: a) Schematic tree depicting the relationship of the new phyla comprising the ASGARD superphylum. **b)** Schematic tree showing an updated version of the tree of life with ASGARD and Eukaryotes as sister groups.

Contact details and address of the project Website

Anja Spang

Department of Cell- and Molecular Biology, Science for Life Laboratory
Uppsala University, Husargatan 3, 75237 Uppsala

Tel.: 0046-18-471 4006

E-mail: anja.spang@icm.uu.se

<http://www.ettemalab.org/>

References

Guy L, Ettema T.J.G. (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends in Microbiology* 19: 580-587.

Saw, J.H., Spang, A., Zaremba-Niedzwiedzka, K., Juzokaite, L., Dodsworth, J.A., Murugapiran, S.K., Colman, D.R., Takacs-Vesbach, C., Hedlund, B.P., Guy, L. and Ettema, T.J. (2015). Exploring microbial dark matter to resolve the deep archaeal ancestry of eukaryotes. *Philos Trans R Soc Lond B Biol Sci.* 370(1678):20140328.

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*, and Ettema, T.J. ASGARD archaea illuminate the origin of eukaryotic cellular complexity (*in revision at Nature*).

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. (2015). Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature*: 521(7551):173-9.

Klinger, C.; Spang, A., Ettema, T. and Dacks, J. (submitted 2015) Tracing the archaeal origins of eukaryotic membrane-trafficking system building blocks". *Submitted to Molecular Biology and Evolution*