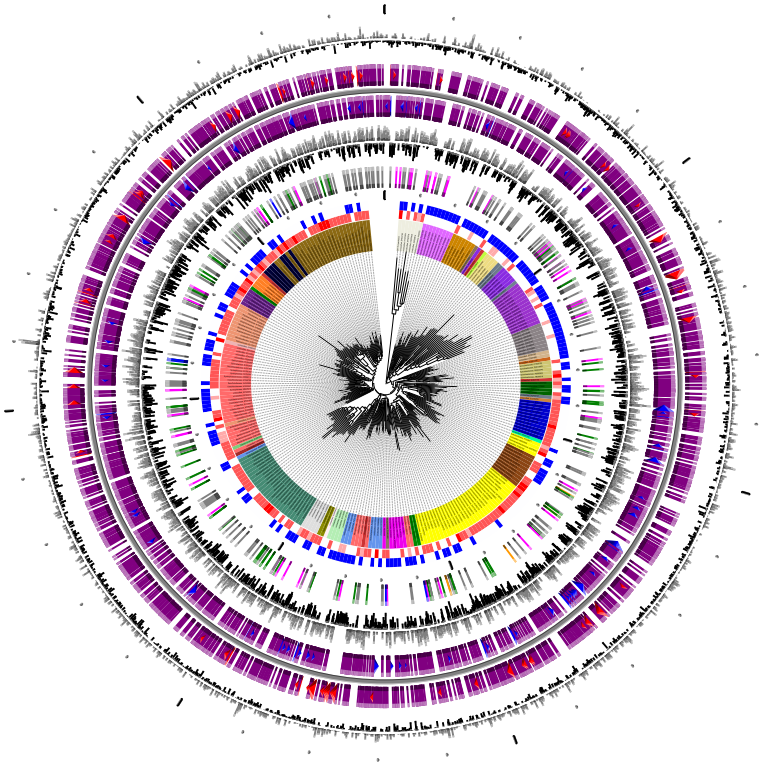

Uncultivated Archaea and Associated Bacteria in Untapped Biotopes

DISSERTATION

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der Fakultät für Biologie und Vorklinische Medizin der

Univeristät Regensburg

vorgelegt von

Alexander Josef Probst

aus Rotthalmünster

2014

Image on title page: From microbial diversity to coding potential of a genome
Circular, phylogenetic tree based on 16S rRNA gene sequences of the microbial diversity present in the Mühlbacher Schwefelquelle (Burgweinting, Germany). Each leaf represents one prokaryotic family of which signatures were either detected in SM1 Euryarchaeal biofilms (red circle) or spring water (blue circle) which are colored by higher taxonomic levels. Outer eight rings represent the genome structure of the SM1 Euryarchaeon metagenome recovered from the same spring.

Das Promotionsgesuch wurde eingereicht am:		31.03.2014
Diese Arbeit wurde angeleitet von:		Prof. Dr. Reinhard Wirth
Prüfungsausschuss:	Vorsitzender:	Prof. Dr. Jürgen Heinze
	1. Gutachter:	Prof. Dr. Reinhard Wirth
	2. Gutachter:	Prof. Dr. Reinhard Sterner
	3. Prüfer:	Prof. Dr. Robert Huber

Unterschrift: Alexander J. Probst

Uncultivated Archaea and Associated Bacteria in Untapped Biotopes

DISSERTATION

for achieving the doctoral degree of natural sciences (Dr. rer. nat.)

at the faculty of Biology and Preclinical Medicine at the University of Regensburg

by

Alexander Josef Probst

from Rotthalmuenster

2014

Running title:
Uncultivated Archaea in untapped biotopes

Key words:
microbial ecology, archaeome, subsurface microbial life, cleanroom microbiome, human skin microbiome, SM1 Euryarchaeon, sulfate-reducing bacteria, biofilm, string-of-pearls community, astrobiology, planetary protection,16S rRNA, microarray, PhyloChip, SR-FTIR, metagenomics

This dissertation was performed under the auspices of

Prof. Dr. Reinhard Wirth (PhD supervisor)

Prof. Dr. Robert Huber (1st mentor)

Dr. Tamas Torok (2nd mentor)

Dr. Christine Moissl-Eichinger (project supervisor)

at the Naturwissenschaftlichen Fakultaet III, Biology and Preclinical Medicine under fulfillment of all requirements of the Regensburger Graduate School of Life Science (RiGeL).

‘It is good to have an end to journey toward; but it is the journey that matters, in the end.’

— Ernest Hemingway

‘If you’re going through hell, keep going.’

— Winston Chruchill

‘Uuuuuuuuur Ahhhhrrrrrr Uhrrrr Ahhhhrrrrrr Aaaarhg...’

— Chewbacca

Acknowledgements

First and foremost, special thanks go to *Dr. Christine Moissl-Eichinger* for supervising my projects and giving me freedom to unfold my mind. I am very grateful to *Prof. Dr. Reinhard Wirth* for being my doctoral father, for the endorsement of every idea I had, and all the support in my career. I thank *Prof. Dr. Robert Huber* for mentoring my PhD thesis and the many advices in these three years. I am indebted to *Dr. Tamas Torok* for his great generosity to mentor my PhD project from abroad, to spending the many hours listening to my research, and to giving me valuable advices. I am also grateful to *Prof. Dr. Reinhard Sterner* as my second assessor.

A great "thank you" go to the group of *Prof. Dr. Thomas Rattei* in Vienna. Thomas, I am very happy for all your support and for all the things you thought me. *Thomas Weinmaier*, thank you for sharing your bioinformatics expertise with me and for having become an awesome friend.

Many thanks go to *Todd Z DeSantis* for teaching me microbiome research, bioinformatics and biostatistics.

I thankfully acknowledge *Prof. Dr. Gary Andersen* for introducing me to the wide field of environmental microbiology. I also thank Gary's group, particularly *Dr. Yvette Piceno* and *Lauren Tom* for introducing my to microarray data acquisition.

A fantastic and very interdisciplinary cooperation with *Dr. Hoi-Ying Holman*'s BSISB group is much appreciated. I thank Hoi-Ying for introducing me to infrared spectromicrosopy and *Dr. Giovanni Birarda* for amazing analysis times and a great friendship. *Dr. Hans Bechtel* is acknowledged for continuous support of the project via beamline maintenance.

I am grateful to *Dr. Simonetta Gribaldo* for phylogenomic analysis and hosting me at the Pasteur Institute. *Kasie Raymann*, I thank you for great times in Paris, great science and a great friendship.

Great thanks go to *Prof. Dr. Jillian Banfield* for a joint collaboration in metagenomics, and to *Dr. Ivan Berg* for the continuous discussions on metabolism.

I deeply acknowledge people from the Biotechnology and Planetary Protection group for continuous support and many joint projects. *Dr. Dr. Kasthuri Venkateswaran* for the many articles we published together and the introduction to science in general and *Dr. Parag Vaishampayan* for great collaborations and our weekly meetings. Last but not least, I thank *Myron LaDuc* for scientific discussions, proofreading, and a great friendship.

I thank the entire Department for Microbiology and Archaea Center for continuous support and encouragement during the last three years. *Prof. Dr. Michael Thomm*, thank you for hosting me at your department and the great parties, *Dr. Annett Bellack* for proofreading and friendship, *Anna Auerbach* for all the amazing lab work and the good times and the many other people, with whom I worked in the lab: *Sandra Meck*, *Alexandra Perras*, *Alexander Mahnert*, *Maximilian Mora*, and *Charlotte Völkel*.

I am grateful to *Stephanie Weber* and *Christiane Absbacher* for proofreading my thesis.

Research funds and travel support from diverse institutions is acknowledged: Studienstiftung des deutschen Volkes (German National Academic Foundation), DFG, ESA, NASA, DOE, UR, BaCaTec, EANA, ASM, ISME, and Graduate Research Academy RNA Biology.

In general, I would like to thank all people who contributed to the publications that made this cumulative dissertation possible.

I would like to thank my family and particularly my parents for unreserved support throughout the entire thesis and my whole life.

List of Publications

Publications listed here have appeared or will appear in peer-reviewed journals or are publications by the National Aeronautics and Space Administration (NASA). The 22 publications have been cited >550 times, 11 are first author publications by Alexander J. Probst and all accepted publications have a cumulative impact factor of 90 (March 2014). Publications 01-05 appeared prior to entering the Ph.D program. Copies of all publications are available on the supporting DVD (./List_of_publications).

Publications

01. Kasthuri Venkateswaran, **Alexander Probst**, Parag A Vaishampayan, and Sudeshan Ghosh (2010): NASA New Technology Report: Isolation of Resistance-bearing Microorganisms. ***NASA Tech Briefs***, March 2010

02. Parag Vaishampayan, **Alexander Probst**, Srinivasan Krishnamurthi, Sudeshna Ghosh, Shariff Osman, Alasdair McDowall, Arunachalam Ruckmani, Shanmugam Mayilraj, and Kasthuri Venkateswaran (2010): *Bacillus horneckiae* sp. nov., isolated from a spacecraft assembly clean room. ***Int J Syst Evol Microbiol***, pp. 1031-1037, Vol. 60

03. **Alexander Probst**, Parag Vaishampayan, Shariff Osman, Christine Moissl-Eichinger, Gary Andersen, and Kasthuri Venkteswaran (2010): Diversity of Anaerobic Microbes in Spacecraft Assembly Clean Rooms. ***Appl Environ Microbiol***, pp. 2837-2845, Vol. 76, No. 9

04. **Alexander Probst**, Rainer Facius, Reinhard Wirth, and Christine Moissl-Eichinger (2010): Validation of the Nylon-Flocked Swab for Efficient Recovery of Bacterial Spores from Smooth and Rough Surfaces. ***Appl Environ Microbiol***, pp. 5148-5158, Vol. 76, No. 15

05. Terry C. Hazen, Eric A. Dubinsky, Todd Z. DeSantis, Gary L. Andersen, Yvette M. Piceno, Navjeet Singh, Janet R. Jansson, **Alexander Probst**, Sharon E. Borglin, Julian L. Fortney, William T. Stringfellow, Markus Bill, Mark S. Conrad, Lauren M. Tom, Krystle L. Chavarria, Thana R. Alusi, Regina Lamendella, Dominique C. Joyner, Chelsea Spier, Jacob Baelum, Manfred Auer, Marcin L. Zemla, Romy Chakraborty, Eric L. Sonnenthal, Patrik D'haeseleer, Hoi-Ying N. Holman, Shariff Osman, Zhenmei Lu, Joy D. Van Nostrand, Ye Deng, Jizhong Zhou, and Olivia U. Mason (2010): Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria. ***Science***, pp. 204-208, Vol. 330, No. 6001

06. Reinhard Wirth, Annett Bellack, Markus Bertl, Yvonne Bilek, Thomas Heimerl, Bastian Herzog, Madeleine Leisner, **Alexander Probst**, Reinhard Rachel, Christina Sarbu, Simone Schopf, and Gerhard Wanner (2011): The Mode of Cell Wall Growth in Selected Archaea Follows the General Mode of Cell Wall Growth in Bacteria — An Analysis using Fluorescent Dyes. ***Appl Environ Microbiol***, pp. 1556-1562, Vol. 77, No. 5

07. **Alexander Probst**, Rainer Facius, Reinhard Wirth, Marco Wolf, and Christine Moissl-Eichinger (2011): Recovery of Bacillus spores from rough surfaces: A challenge to cleanliness control of space missions. ***Appl Environ Microbiol***, pp. 1628-37, Vol. 77, No. 5

08. Moogega Cooper, Myron T. La Duc, **Alexander Probst**, Parag Vaishampayan, Christina Stam, James N. Benardini, Yvette M. Piceno, Gary L. Andersen, and Kasthuri Venkateswaran (2011). Comparison of Innovative Molecular Approaches and Standard Spore Assays for Assessment of Surface Cleanliness. ***Appl Environ Microbiol***, pp. 5438-5444, Vol. 77, No. 15

09. Daniel McDonald, Morgan N. Price, Julia Goodrich, Eric P. Nawrocki, Todd Z. DeSantis, **Alexander Probst**, Gary L. Andersen, Sean R. Eddy, Adam Arkin, Rob Knight, and Philip Hugenholtz (2011). An Improved Greengenes Taxonomy for Bacteria and Archaea with explicit ranks. ***ISME J***, pp. 610- 618, Vol., No. 3

10. Agnes Weiner, Simone Schopf, Gerhard Wanner, **Alexander Probst**, and Reinhard Wirth (2012). Positive, neutral and negative interactions in cocultures between *Pyrococcus furiosus* and different methanogenic Archaea. ***Microbiol Insights***, pp. 1-10, Vol. 4

11. **Alexander Probst**, Alexander Mahnert, Christina Weber, Klaus Haberer, and Christine Moissl-Eichinger (2012). Detecting inactivated endospores in fluorescence microscopy using propidium monoazide. ***Int J Astrobiol***, p. 117-123, Vol. 11, No. 2

12. Parag Vaishampayan*, **Alexander J. Probst***, Myron T La Duc*, Emilee Bargoma, James N. Bernadini, Gary L. Andersen, Kasthuri Venkateswaran (2012). New perspectives on viable microbial communities in low-biomass cleanroom environments. ***ISME J***, pp. 312-324, Vol. 7 - *authors contributed equally

13. Kasthuri Venkateswaran*, Myron T. La Duc*, Parag Vaishampayan*, Shariff Osman, Kelly Kwan, Emilee Bargoma, **Alexander Probst**, Moogega Cooper, James N. Benardini, and James A. Spry (2012): Genetic Inventory Task: Final Report, Volume 2, ***JPL Publication*** 12-12, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA - *main contributors

14. **Alexander J. Probst**, Hoi-Ying N. Holman, Todd Z. DeSantis, Gary L. Andersen, Hans A. Bechtel, Maria Sonnleitner, Kasthuri Venkateswaran, and Christine Moissl-Eichinger (2012). Tackling the minority: Sulfate-reducing bacteria in an archaea-dominated subsurface biofilm. ***ISME J***, pp. 635-651, Vol. 7

15. **Alexander J. Probst**, Anna K. Auerbach, and Christine Moissl-Eichinger (2013). Archaea on human skin. ***PLoS One***, e65388, Vol. 8

16. Christine Moissl-Eichinger, Rüdiger Pukall, **Alexander J. Probst**, Michaela Stieglmeier, Petra Schwendner, Maximilian Mora, Simon Barczyk, Maria Bohmeier, and Petra Rettberg (2013). Lessons Learned from the Microbial Analysis of the Herschel Spacecraft during Assembly, Integration, and Test Operations. ***Astrobiology***, pp. 1125-1139, Vol. 13, No. 12.

17. Jordan E. Krebs*, Parag Vaishampayan*, **Alexander J. Probst***, Lauren M. Tom, Viggó Thór Marteinsson, Gary L. Andersen, and Kasthuri Venkateswaran (2014). Microbial community structures of novel icelandic hot spring systems revealed by PhyloChip G3 analysis. ***Astrobiology***, epub ahead of print, Vol. 14, No. 3 - *authors contributed equally

18. **Alexander J. Probst**, Pek Yee Lum, Bettina John, Eric A Dubinsky, Yvette M Piceno, Lauren M Tom, Gary L Andersen, and Todd Z DeSantis (2014). Microarray of 16S rRNA gene probes for quantifying population differences across microbiome samples. In: Microarrays: Current Technology, Innovations and Applications. Editor: Zhili He, ***Horizon Scientific Press and Caister Academic Press***, accepted

15. Anja Bauermeister, Alexander Mahnert, Anna Auerbach, Alexander Böker, Niwin Flier, Christina Weber, **Alexander J. Probst**, Christine Moissl-Eichinger, Klaus Haberer (2014). Quantification of encapsulated bioburden in spacecraft polymer materials by cultivation-dependent and molecular methods. ***PLoS One***, accepted

20. Matthew Miezeiewski, Todd Schnaufer, Michele Muravsky, Su Wang, Ivette Caro-Aguilar, Susan Secore, David S. Thiriot, Charlie Hsu, Irene Rogers, Todd Z. DeSantis, Justin Kuczynski, **Alexander J. Probst**, Christel Chehoud, Rachel Steger, Janet Warrington, Jean-Luc Bodmer, Jon H. Hinrichs (2014). An in virto culture model to study the dynamics of colonic microbiota in syrian golden hamster and their susceptibility to infection with *Clostridium difficile*. ***ISME J***, submitted

21. **Alexander J. Probst**, Giovanni Birarda, Hoi-Ying N. Holman, Todd Z. DeSantis, Gerhard Wanner, Gary L. Andersen, Alexandra K. Perras, Sandra Meck, Jörg Völkel, Hans A. Bechtel, Reinhard Wirth, and Christine Moissl-Eichinger (2014). Coupling genetic and chemical microbiome profiling reveals Heterogeneity of Archaeome and Bacteriome in Subsurface Biofilms that are Dominated by the Same Archaeal Species. ***PLoS One***, submitted

22. **Alexander J. Probst**, Thomas Weinmaier, Kasie Raymann, Alexandra Perras, Joanne B. Emerson, Thomas Rattei, Gerhard Wanner, Andreas Klingl, Ivan Berg, Bernhard Viehweger, Marcos Yoshinaga, Kai-Uwe Hinrichs, Simonetta Gribaldo, Brian C. Thomas, Sandra Meck, Anna Auerbach, Matthias Heise, Jillian F. Banfield, and Christine Moissl-Eichinger (2014). Grappling with dark matter: Biology of an uncultivated subsurface archaeon. ***Nat Comm***, in preparation.

Patent

Giovanni Birarda, **Alexander J Probst**, Hoi-Ying N Holman (2014). “Rapid and label-free infrared imaging for microbial community screening and profiling”. ***U.S. patent application*** serial number 61/908,014 filed on 22 November 2013, submitted

Conference Proceedings

Contributions at international conferences included >30 posters and talks (date: March 2014) held at e.g. International Society for Microbial Ecology (ISME), American Society for Microbiology (ASM). A full list of all contributions can be requested from the author (alexander.j.probst@gmail.com).

Table of Contents

I. Abstract

Despite the almost ubiquitous nature of Archaea about the planet, and an ever-increasing appreciation for the contribution they make to Earth's energy cycles, the distribution and metabolic potential of these microorganisms remain poorly understood due to the lack of cultivated representatives. The overarching objective of this study was to significantly bolster the current understanding of the occurrence and function of archaea, and co-existing bacteria, in two so-far poorly explored biotopes. To date, archaea have yet to be isolated and cultivated under laboratory conditions from either of these environments, which necessitated the use of metagenomic and other emerging molecular approaches available in the field of microbial ecology.

The first biotope investigated, spacecraft assembly cleanrooms, are strongly influenced by rigorous cleaning regimens, controlled humidity, and scarce nutrients. While the occurrence of Thaumarchaeota had already been reported in these environments, their origin and mode of ingress remained unknown and unexplored. Over the course of this investigation, Thaumarchaeota were identified to persist on human skin, and human activity was shown to be a source of the observed archaeal presence in spacecraft assembly facilities. In a parallel vein, a viability-dependent assay coupled to molecular microbiome profiling enabled the differentiation of living fraction of bacteria from the entire community profile. The results of these analyses represent a milestone in cleanroom monitoring, as they suggest that the viable fraction of the spacecraft assembly cleanroom microbiome accounts for a mere tenth of its total abundance and diversity.

The second biotope investigated in this study were subsurface environments that are accessible through anaerobic, sulfidic springs. It had previously been demonstrated that, with regard to biodiversity and metabolic potential, the subsurface remains poorly characterized and thus represents a great reservoir for mining novel microbial lineages. More specifically, the upwelling of sulfidic springs near Regensburg has been shown to source extensive quantities of nearly mono-species biofilms of the recently discovered SM1 Euryarchaeon (95% purity). The findings of this investigation, for the first time ever, demonstrated that bacterial contingent of this primarily archaeal biome was actively taking part in sulfate-reduction. Comparative microbiome analysis of biofilms from two geographically distinct spring locations enabled the observation of extensive variation between communities without substantial change in their function. To resolve the metabolic capability and phylogenomic position of the SM1 Euryarchaeon, metagenomic approaches were utilized in concert with other state-of-the-art techniques. The results of these efforts evidenced a distinct pyhlogenetic position of the SM1 Euryarchaea, for which the name "*Candidatus* Altiarchaeum hamiconnexum" (candidatus family "*Altiarchaeaceae*", candidatus order "*Altiarchaeales*") is proposed. The genome of this novel taxon was shown to encode the enzymes responsible for a novel, archaeal acetyl-CoA pathway free of $Factor_{420}$. Such a pathway affords these double-membraned archaea to grow anaerobically, most likely via acetate oxidation. Genes for the genesis of the biofilms mediated by nano-grappling hooks and diverse extracellular polysaccharides were identified in the genome. It is hypothesized that the biofilm provides a competitive advantage with respect to fitness, as they help to outcompete other subsurface microorganisms by filtering and capturing nutrients that seep into the groundwater.

The findings of the many investigations that make up this thesis significantly contributed to today's much enhanced understanding of archaea and the roles they play alongside the co-occurring bacteriome in two distinct biotopes. At the same time, the results represent a remarkable example of how well the ecophysiology of novel archaeal lineages can be examined with current state-of-the-art technologies.

II. General Introduction

1. Exploring the uncultivated biodiversity: breaking the three-domain mold?

1.1. Carl R. Woese and George E. Fox: proposal of the three domains of life

In 1977, Carl Woese[1] and George Fox posited that each and every form of life on Earth could be binned into one of the three primary domains, the Eukaryota, the Bacteria, and the Archaea (Woese and Fox, 1977; Woese et al., 1990). The latter two domains were termed the 'Prokaryotes'. This descriptor is not to be confused with the broader term 'microorganisms', which also encompasses Eukaryotic clades like fungi, protists, and mites. Prior to Woese's and Fox's three-domain hypotheses, all Prokaryotes were termed Bacteria reflected by the misleading nomenclature of some archaea like *Methanobacterium* (Kluyver and Van Niel, 1936) and *Halobacterium* (Harrison and Kennedy, 1922; Ventosa and Oren, 1996). Renaming these archaea subsequently, i.e. *Halobacterium* to *Haloarchaeum* (DasSarma and DasSarma, 2008), was not possible since it would stand in violation of the 'General Considerations, Principles and Rules of the International Code of Nomenclature of Bacteria' (Oren, 2008).

Woese and Fox concluded on their three-domain hypothesis based primarily on the sequence analysis of 18S ribosomal RNA (rRNA; eukaryotic) and 16S rRNA (prokaryotic) isolated from 13 different organisms. The rRNA gene is universally present in all forms of life on this planet. Since Sanger sequencing technology (Sanger et al., 1977) had not yet been available, Woese and Fox refined molecular methods of their time to find differences in sequential nucleotide compositions of rRNA gene transcripts. They digested 18S/16S rRNA with T1 RNase individually for each organism, the end product of which was used to generate oligonucleotide fingerprints via 2D-electrophoresis followed by sequencing the short oligonucleotides. The association coefficient S_{AB} was calculated in pairwise manner for each of the sequences generated, corresponding to a similarity value between transcripts (Woese and Fox, 1977). *Figure II.1-1* depicts a heatmap generated from these S_{AB} coefficients. A clear trend towards separation of three domains of life is resolved due to the low S_{AB} coefficients between them. Intra-domain similarity, however, also varied greatly.

[1] In December 2012, Carl Woese passed away. For his lifework the reader is referred to the many reviews covering this topic, e.g. Fox, G.E. (2013). Carl R. Woese, 1928-2012. Astrobiology *13*, 1201-1202.

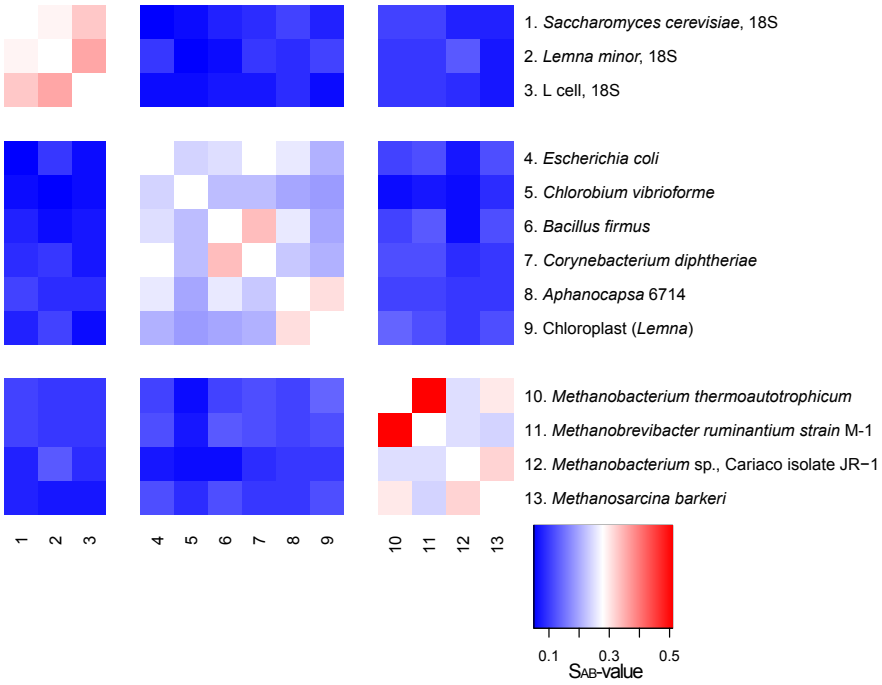

Figure II.1-1 *Heatmap based on S_{AB} coefficients provided in Woese and Fox, 1977. This particular image was generated using R (R-Development-Core-Team, 2005). Please note that Methanobacterium thermoautotrophicum is now referred to as Methanothermobacter thermoautotrophicus. Deep red is indicative of a high S_{AB} value (similarity score), whereas dark blue reflects a low similarity value.*

1.2. 16S rRNA gene based phylogenetics

Unlike Woese's and Fox's, by way of Sanger sequencing modern day scientists have easy access to the entire ca. 1500 base pair 16S rRNA gene of Prokaryotes (Sanger et al., 1977). Accurately assessing the entire rRNA gene sequence similarity shared between disparate organisms not only facilitates the generation of much more meaningful dissimilarity values, it also allows to infer phylogenetic relationships based on a single gene (Fox et al., 1977). Particularly the comparison of color intensities of the heatmap presented in *Figure II.1-1* and the heatmap presented in *Figure II.1-2A* verifies the superiority of information that can be gathered from entire 16S/18S rRNA gene compared to the S_{AB} values used by Woese and Fox, 1977.

Today, the widespread application of 16S rRNA gene sequencing and analysis has enabled many researchers to identify and infer the phylogenetic position of microbial isolates as depicted in *Figure II.1-2B* (Fox et al., 1977). However, Stackebrandt and Goebel have cautioned that while 16S rRNA gene phylogeny can be used to infer relationships for Bacteria and Archaea from kingdom down to species level, it is limiting with respect to strain specificity (Stackebrandt and Goebel, 1994). This circumstance was supported by for instance Ash et al., who reported the inability in resolving the *Bacillus cereus* group to the species level via 16S rRNA gene sequence analysis (Ash et al., 1991).

Nevertheless, 16S rRNA gene sequence analysis has revolutionized environmental microbiology. As it is currently estimated that fewer than 1 in every 100 prokaryotic lineages on Earth can be cultivated under laboratory conditions (Amann et al., 1995; Colwell, 1997), the value of methods to collect, process, purify, and analyze 16S rRNA from the environment, and thereby circumventing the

need to culture cells, is tremendous. Such techniques have, in many instances, provided the only source of understanding the unseen diversity of bacteria and archaea contributing to Earth's elemental cycles (Pace et al., 1986; Ward et al., 1990).

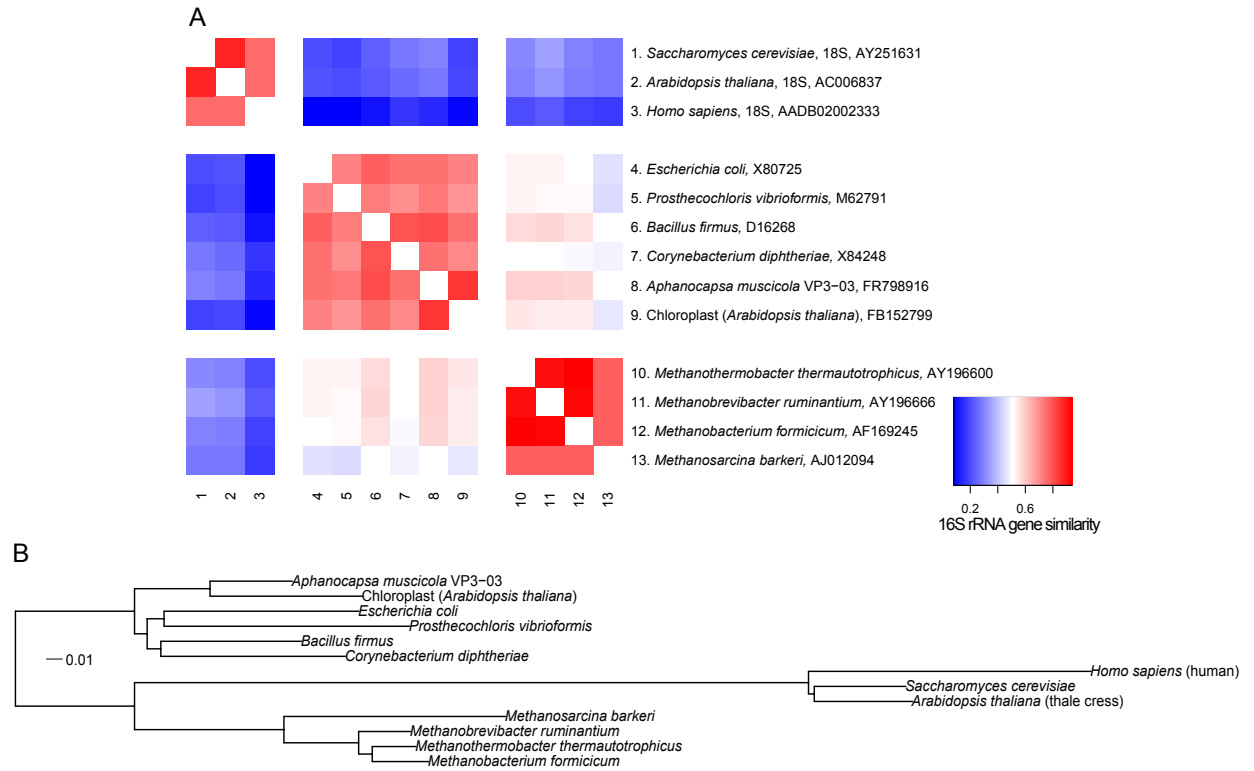

Figure II.1-2 **A**) Heatmap based on rRNA gene similarity. Gene sequences were downloaded from the SILVA database and clustered using the average neighbor algorithm in mothur (Pruesse et al., 2007; Schloss et al., 2009). Dark red is indicative of high similarity, whereas dark blue reflects low similarity. **B**) Phylogenetic tree consisting of the same rRNA genes used for A. While the intra-domain phylogeny is lacking information due to under-sampling of the phylogenetic clades available, the three-domain hypothesis easily resolved. The phylogenetic tree was generated using FastTree V2.0 (maximum likelihood approximation) and visualized in iTOL (Letunic and Bork, 2007; Price et al., 2010).

As technologies continue to evolve, 16S rRNA gene sequence analysis is nowadays barely used to infer the phylogenetic position of individual microorganisms unless scientist describe novel species. Instead, this technique has proven important for deciphering the entire microbial communities housed within a given environmental sample. To accomplish this, researchers extract total environmental 'metagenomic' DNA and amplify the 16S rRNA genes via polymerase chain reaction (PCR) with archaea- or bacteria-directed primers [a detailed evaluation and review of frequently used 16S rRNA gene primers was recently published by Klindworth and co-workers (Klindworth et al., 2013)]. The resulting 16S rRNA gene amplicons are then either cloned into competent *E. coli* cells and afterwards sequenced via Sanger methods (Sanger et al., 1977) or directly subjected to next generation sequencing [NGS, (Sogin et al., 2006); for details on NGS please see II.2.3 and a review by Hutchison (Hutchison, 2007)]. The acquisition and analysis of 16S rRNA gene NGS data has been thoroughly standardized through existing software pipelines (Caporaso et al., 2010; Edgar, 2013; Schloss et al., 2009). Consequently, the NGS approach is no longer used to describe the alpha-diversity of samples, but rather community relationships between samples as well as changes in microbial abundance of

single taxa as a function of time and/or geographic location [for a summary and comparison of the various community relationship calculations that exist in the literature, the reader is referred to the study by Kuczynski and coworkers (Kuczynski et al., 2010)]. NGS-generated beta-diversity and gamma-diversity analyses contributed significantly to the Human Microbiome Project and the Earth Microbiome Project, whose objectives were to elucidate the entire spectrum of microorganisms living on the human body and on our home planet, respectively (Gilbert et al., 2010a; Gilbert et al., 2010b; Wortman et al., 2010). Ergo, it has become commonplace in the literature to refer to the total genetic information pertaining to all of microorganisms (bacteria, archaea, fungi, and some lower Eukaryota) present in a given environmental sample as its **'microbiome'**. When such information is restricted to the domain Bacteria, the term **'bacteriome'** found great acceptance in the literature. Along these lines of reasoning, the term **'archaeome'** — referring to the total genetic information pertaining to all of the archaea detected in an environmental sample — is used in this study, although it has not appeared in the literature so far. For any given sample, its bacteriome and archaeome can determined to the exclusion of the other via 16S rRNA gene PCR/amplicon analysis with domain-directed primers, as mentioned above.

1.3. Phylogenomics and the two-domain hypothesis

Analysis and interpretation of 16S rRNA gene phylogenetics have their limitations, not only in deciphering differences between strains but also at higher taxonomic levels. Ancient evolutionary relationships can barely be recovered by molecular phylogenies in general (Gribaldo and Philippe, 2002), since successive substitutions can disguise early speciation events, particularly when computing phylogenetic relationships from single marker genes like 16S rRNA gene (Gribaldo et al., 2010). As a result, the inferred evolutionary relationship between Archaea, Bacteria, and Eukaryota can consequently lead to different results depending on the gene analyzed (Harris et al., 2003). Consequently, the 'conditioned reconstruction' of the entire genomic information (Lake and Rivera, 2004) and the concatenation of orthologous genes to large alignments (Delsuc et al., 2005; Snel et al., 2005) have each been applied for to better resolve dissimilarities between taxonomic lineages. The latter methodology, in particular, has gained considerable popularity. By calculating dissimilarities between concatenated ribosomal protein sequences, this approach has become the standard for inferring phylogenetic positions of newly sequenced bacterial and archaeal genomes (Castelle et al., 2013; Gribaldo et al., 2010; Kozubal et al., 2013; Rinke et al., 2013; Spang et al., 2010; Williams et al., 2013; Wrighton et al., 2012), although sequences of proteins involved in replication also seemed to appropriately reflect phylogeny (Raymann et al., 2014).

The associated methodologies, pitfalls, artifacts, and major encompasses of evolutionary reconstruction of the tree of life have recently been reviewed (Gribaldo et al., 2010; Williams et al., 2013), and the authors agree upon one major conclusion. Currently, there are two phylogenetic scenarios plausible for the relationship between Archaea, Bacteria, and Eukaryota: A) the 'three primary domains' (3D) and B) the 'two primary domains' (2D) scenario (*Figure II.1-3*). While the former is similar to what had been proposed by Woese et al. 1977 and 1990 (Woese and Fox, 1977; Woese et al., 1990), the latter suggests a sister relationship between Eukaryota and the Crenarchaeota (also

known as 'eocytes'), where Archaea are the direct ancestor of Eukaryota (Guy and Ettema, 2011; Lake et al., 1984).

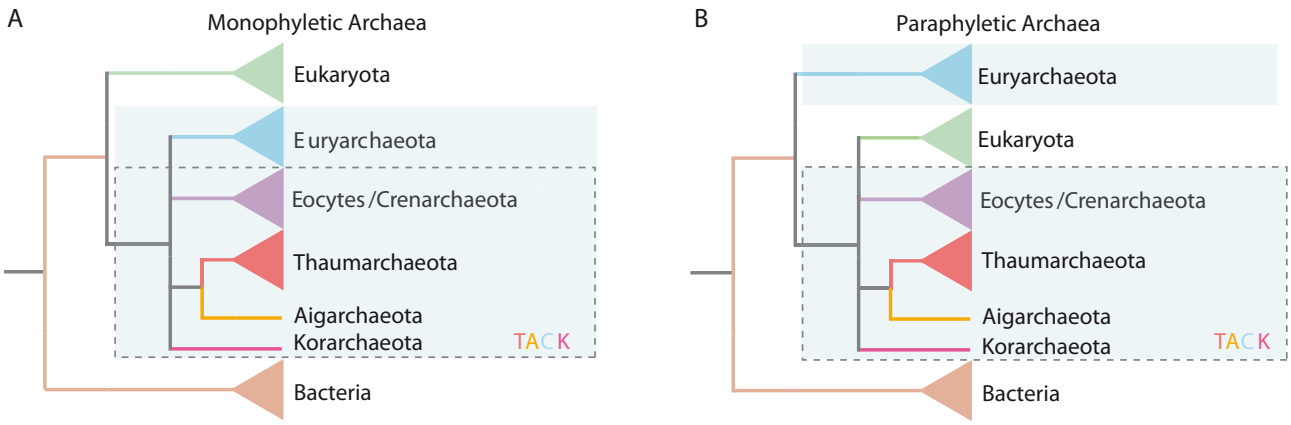

Figure II.1-3 *Phylogenetic trees taken from Williams et al., which reflect the three-domain scenario.* ***A****) and the two domain scenario.* ***B****) (Williams et al., 2013). In the 2D scenario, the Archaea form a paraphyletic group. [TACK = Thaumarchaeota-Aigarchaeota-Crenarchaeota-Korarchaeota superphylum, (Guy and Ettema, 2011)]*

Although Gribaldo et al. argued that the final scenario cannot yet be determined (with adequate confidence), Williams et al. maintain that the 2D scenario presents the most plausible solution as it is well resistant to eukaryotic-archaeal horizontal gene transfer [HGT; (Gribaldo et al., 2010; Williams et al., 2013)]. Nevertheless, many questions remain unanswered — perhaps none greater than the question for the different lipid architecture of Archaea and Eukaryota. This is, however, possibly also explainable by HGT as bacteria may have transferred their lipid biosynthesis machinery to Eukaryota (Gribaldo et al., 2010; Pereto et al., 2004). In any scenario, whether the Archaea are a sister domain of the Eukaryota or their direct ancestors, it is necessary to A) study the biology of Archaea in depth and B) taxonomically sample Earth's life sufficiently in order to be able to reconstruct proper phylogeny and understand the processes behind Eukaryogenesis (Gribaldo et al., 2010).

1.4. Refining the tree of life: Metagenomics and single cell sequencing

Neither the 2D nor the 3D scenario is supported by adequate empirical evidence, however, as the 3D paradigm has been widely accepted in the literature of the past 20 years (Woese et al., 1990), the current thesis refers to Archaea as the third domain of life. Historically, this domain comprised two major phyla, the Euryarchaeota and the Crenarchaeota (Woese et al., 1990). Nevertheless, in recent years, many more phyla were proposed, whereas the Korarchaeota (Barns et al., 1996), the Aigarchaeota (Nunoura et al., 2011), and the Thaumarchaeota (Brochier-Armanet et al., 2008; Pester et al., 2011; Spang et al., 2010) have found great acceptance in the literature so far. The phylum Nanoarchaeota is currently represented by a few genomes and only one cultured organism, *Nanoarchaeum equitans* (Huber et al., 2002; Podar et al., 2013; Waters et al., 2003). The Nanoarchaeota have been discussed to be a 'fast-evolving euryarchaeal lineage', and their status as a separate phylum is currently uncertain (Brochier et al., 2005). Additional phyla, like the Geoarchaeota, have been proposed (Kozubal et al., 2013), but recent re-analysis of the phylogenomic positioning of the Geoarchaeota revealed this archaeal lineage as Thermoproteales-related (Guy et al., 2014).

The aforementioned Thaumarchaeota are a remarkable case where single gene phylogenetics (16S rRNA gene) fail to reconstruct evolutionary relationships. Previously, members of this phylum were described as widespread crenarchaeota contributing substantially to nitrogen cycling via ammonia-oxidation in the ocean and soil biotopes (DeLong, 1992; Fuhrman et al., 1992; Karner et al., 2001; Konneke et al., 2005; Leininger et al., 2006). Using phylogenomics, these crenarchaea were separated from others and reclassified as Thaumarchaeota, whose distribution on Earth still necessitates exploration (Brochier-Armanet et al., 2008; Pester et al., 2011; Spang et al., 2010).

At least two of the aforementioned phyla of Archaea (Korarchaeota and Aigarchaeota) have been designated based on genomic data not generated from pure cultures (Elkins et al., 2008; Nunoura et al., 2011). Such a reconstruction of genomes from enrichment cultures or directly from environmental samples is termed 'metagenomics' and was launched by two independent groups in 2004 (Tyson et al., 2004; Venter et al., 2004). Venter's group generated a huge set of environmental sequences never seen before, which continues to reveal interesting hypotheses as scientists further explore and reanalyze these data. For instance, a recent publication based solely on these sequencing data claimed indications for a 'fourth domain of life' (Wu et al., 2011). Nevertheless, these sequences may also belong to unknown viruses (Wu et al., 2011), like giant viruses that even contain ribosomes (Philippe et al., 2013), and their positioning as the fourth domain of life continues in the literature (Colson et al., 2012; Colson et al., 2011; Legendre et al., 2012; Williams et al., 2011). In contrast to generating huge amounts of shotgun sequencing data, Tyson et al. focused on the reconstruction of entire genomes or draft genomes. Since then, multiple genomes of different bacteria and archaea have been reconstructed from the environment, which has lead to the proposal of many Candidate phyla (Castelle et al., 2013; Di Rienzi et al., 2013; Nunoura et al., 2011; Sharon and Banfield, 2013; Wrighton et al., 2012).

Another approach to get an insight into the genomes of the uncultivated majority of microorganisms in the environment is the amplification of DNA from a single-cell (Zhang et al., 1992) followed by sequencing, also termed single-cell genomics (Hutchison and Venter, 2006). This method has recently been performed on large scale (Rinke et al., 2013). Rinke et al. reconstructed 201 different microbial genomes, which had, however, only a completeness of 40% on average (ranging from 10% to 90%). The lack of completeness is one downside of single-cell genomics compared to metagenomics, where multiple genomes of closely related or identical organisms are assembled (Sharon and Banfield, 2013). Rinke et al. ended up proposing a superphylum called DPANN (representative for the lineages Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota) at the base of the Archaea, which unified many environmental genomes gathered from nano-sized archaea, including the Nanoarchaeota (Rinke et al., 2013). However, this incorrect placement of the Nanoarchaeota in the phylogenetic tree of Archaea has been noted earlier (Brochier et al., 2005), and Castelle et al. recently reported that not all nano-sized archaea cluster together when using appropriate phylogenomics coupled with increased taxon sampling (Castelle et al., 2013). These results were confirmed by a recent study that used DNA replication protein sequences instead of ribosomal protein sequences for phylogenetic inference (Raymann et al., 2014). Thus, the enigmatic

clustering of nano-sized archaea can be attributed to an insufficient taxon sampling of these lineages resulting in a long branch attraction artifact (Felsenstein, 1978; Lartillot et al., 2007).

Many metagenomic studies have not only amplified our knowledge on the phylogeny of Archaea and Bacteria, they have also shed light on the potential metabolic capabilities of the uncultivated majority in our ecosystems (Albertsen et al., 2013; Baker et al., 2010; Baker et al., 2006; Castelle et al., 2013; Erkel et al., 2006; Kantor et al., 2013; Nunoura et al., 2011; Tyson et al., 2004; Wrighton et al., 2012). Particularly, novel archaeal lineages were shown to carry a great proportion of dark matter concerning their genomic coding potential (Baker et al., 2010; Baker et al., 2006; Nunoura et al., 2011). These unexplored genes may encode for novel enzymatic pathways and metabolic capabilities that go way beyond the knowledge gathered from cultivated microorganisms. At the same time, new archaeal lineages also harbor great novelty concerning their cell ultrastructure (Comolli et al., 2009; Moissl et al., 2005b). For instance, more than 37% of the genes in ARMAN (Archaeal Richmond Mine Acidophilic Nanoorganisms) showed no matches to clusters of orthologous groups in Archaea, and these novel lineages were also described to be one of the first archaea to possess a double-membrane system (Baker et al., 2010; Comolli et al., 2009). These findings highlight the importance of exploring novel, untapped biotopes to understand diversity and function of uncultivated archaea. Such novel biotopes are mainly found in the subsurface, as it accommodates a great proportion of microbial life that has been barely explored due to the lack of accessibility (Whitman et al., 1998). Indeed, Castelle et al. and Nunoura et al. showed that subsurface samples can reveal many unexplored archaeal phyla and represent a great reservoir for mining novel archaeal taxa (Castelle et al., 2013; Nunoura et al., 2011). Such a novel, uncultivated archaeal taxon called SM1 Euryarchaeon is found near Regensburg, Germany, in a cold sulfidic spring, where it is constantly washed up from the subsurface (Henneberger et al., 2006). Although this archaeon appeared to be a novel lineage within the Euryarchaeota based on 16S rRNA gene sequence analysis (Rudolph et al., 2001), it remains completely unexplored with regard to its genomic information, metabolic capabilities or phylogenomic placement. Related 16S rRNA gene sequences of this novel archaeal taxon were described to be wide-spread (Rudolph et al., 2004) and also found to be dominant in hydrogen-enriched, anoxic subsurface hydrothermal waters (Chapelle et al., 2002). The latter biotope was intensely discussed to be an analogous biotope of foreign celestial bodies in our solar system like Mars and Europa [moon of Jupiter; (Chapelle et al., 2002)]. Consequently, life thriving in these subsurface environments could potentially resemble extraterrestrial life that mankind still desires to discover — if it exists.

2. Astrobiology and the search for (extra)terrestrial life

Astrobiology is an interdisciplinary field unifying numerous science disciplines and space technologies to answer mankind's fundamental questions, such as 'How did life originate?', 'Are there other habitable planets?', and 'Does life exist anywhere else in the universe?'. Many subdisciplines have evolved in recent decades, all of them shaping the picture of the relatively new field of Astrobiology (Morrison, 2001).

2.1. Panspermia as the origin of life?

Approximately 100 years ago, Arrhenius formed the hypothesis that life may not have originated on Earth but on another celestial body in the galaxy and may have been transported to Earth as microbial life (Arrhenius, 1908). This theory was named 'Panspermia', and numerous incarnations have been formed, like Lithopanspermia (transmission of microorganisms via meteorites (Tobias and Todd, 1974)) or Radiopanspermia [the distribution of life via starlight (Secker et al., 1994)]. It was concluded that life may be ubiquitous in the galaxy, if a theory like Radiopanspermia was real (Parsons, 1996). A more exotic theory is the Directed Panspermia, which hypothesizes that extraterrestrial intelligent life forms have sent microbial life to Earth (Crick and Orgel, 1973). However, it has been calculated that even within our solar system, terrestrial life could have been spread to other celestial bodies via Lithopanspermia. These celestial bodies should not be considered as isolated, and planets like Mars could therefore harbor microbial life that originated on Earth and vice versa (Melosh, 1988). Nevertheless, the search for extraterrestrial life is a great desire of mankind and continues with many ongoing and future life detection space missions.

2.2. Planetary protection a.k.a. applied astrobiology

As mentioned above, astrobiology holds many subdisciplines, one of them being the field of 'planetary protection'. In this applied research area, scientists work to prevent the inadvertent contamination of otherwise pristine celestial bodies with terrestrial life forms while conducting various space exploration initiatives (Rummel, 1989). For instance, exceedingly large numbers of terrestrial life forms must not be transported to an extraterrestrial setting (e.g. Mars) via spacecraft, as their presence could confound present and future life detection missions. In addition, they could even colonize the environment and interfere with the natural progression and evolution of indigenous life forms (termed 'forward contamination'). Conversely, 'reverse contamination' refers to a scenario in which extraterrestrial life forms are inadvertently returned to Earth. Any such instance could prove harmful to the planet's biosphere. Missions to different celestial bodies are categorized based on many attributes pertaining to the scientific interest in the selected destination, and different types of missions have different requirements with regard to cleanliness of spacecraft hardware [for details please see (COSPAR, 2002)].

As extraterrestrial life has yet to be discovered, there are those that have criticized planetary protection as it may be unnecessary but expensive (Fairén and Schulze-Makuch, 2013); others, however, contend that it represents a very important scope of application in astrobiology (Conley and Rummel, 2013). Particularly, safeguarding life detection missions to avoid false positives is of high interest — as recently demonstrated by Mars Science Laboratory's drill bit debacle[1].

As was the case for the Mars Science Laboratory (also known as 'Curiosity' rover), all spacecraft hardware outbound Earth needs to be assembled in a controlled facility, i.e. a cleanroom, in order to minimize contamination risk. This helps to ensure the integrity (with respect to cleanliness) of the hardware surfaces and minimizes the likelihood of stowaway contaminant microbes inadvertently being transferred to extraterrestrial settings of scientific interest. Such cleanroom facilities

[1] Los Angeles Times/McClatchy / September 10, 2012;
http://www.csmonitor.com/Science/2012/0910/Scientists-fear-Curiosity-rover-drill-bits-could-contaminate-Mars, last access March 23, 2014.

typically maintain a constant humidity, are equipped with redundant air-filtering systems, and are subjected to rigorous cleaning regimens. Consequently, these cleanrooms harbor low nutrients for microorganisms. This is ideal, since microorganisms that enter cleanrooms on equipment and humans can be very problematic and even threaten forward contamination (Venkateswaran et al., 2001). The molecular signatures of microorganisms that have been detected in cleanroom facilities belonged to all three domains of life, Bacteria, Archaea, and Eukaryota (La Duc et al., 2012; Moissl et al., 2008; Venkateswaran et al., 2001).

Since the Viking era, scientists have labored fervently to catalog the entire microbial population associated with spacecraft hardware and the cleanroom facilities in which it is assembled (Puleo et al., 1977). As a result, the techniques by which to assess the cleanroom microbiome have evolved significantly, coinciding with a great many other methodological advancements in the field of microbial ecology.

2.3. Deciphering microbiomes: The transfer of technologies from microbial ecology to planetary protection

The field of microbial ecology has experienced two major milestones in microbial community analysis. First and foremost, the ability to directly analyze 16S rRNA (genes) retrieved from an environmental sample has tremendously improved the current understanding of the extent of — and role of — the uncultivated portion of a given environmental microbial community (Pace et al., 1986; Ward et al., 1990). Historically, the 16S rRNA gene has been proven a suitable marker gene due to its length that can be covered by two (to several) Sanger sequencing reads and due to its 9 (hyper-)variable regions flanked by conserved regions allowing the inference of phylogenetic relationships between microorganisms (Fox et al., 1980; Lane et al., 1985; Sanger et al., 1977; Ward et al., 1990; Woese and Fox, 1977). The second milestone was the advent of so-called ‘metaomics’ and their application to microbial community samples. This discipline covers the analysis of an environmental sample to reconstruct microbial genomes [metagenomics, (Tyson et al., 2004)], microbial transcriptomes [metatranscriptomics; (Poretsky et al., 2005)] and microbial proteomes [metaproteomics; (Wilmes and Bond, 2004)] of an entire community. Other ‘omics’ approaches like metabolomics and metalipidomics have been reviewed by Zhang and coworkers (Zhang et al., 2010).

16S rRNA gene analysis, metagenomics, and metatranscriptomics have all benefited immensely from NGS technologies, such as illumina/solexa (Bennett, 2004; Bennett et al., 2005), 454 pyrosequencing (Nyrén et al., 1993; Ronaghi et al., 1996), and ion torrent (Rothberg et al., 2011). NGS facilitates a greater sampling depth of molecules that can be assayed in comparison to the outdated Sanger sequencing technology (Sanger et al., 1977). However, with respect to intrinsic fidelity, these technologies suffer from short read length and/or homopolymer errors (Huse et al., 2007; Margulies et al., 2005), which need to be quelled via appropriate bioinformatic tools (Gilles et al., 2011; Hamady et al., 2008; Kircher et al., 2009; Preheim et al., 2013; Quince et al., 2009; Schiex et al., 2003). Another technology that facilitated the high throughput of 16S rRNA gene amplicon analysis is a hybridization based technology called PhyloChip™ G3 (DeSantis et al., 2007; Hazen et al., 2010). Here, 500 ng of DNA (PCR amplicons), corresponding to approximately 10^{11} 16S rRNA gene molecules, are fragmented, labeled, and hybridized onto an array, which carries ca. 1.1 million

probes (25-mers) designed from the entire Greengenes 16S rRNA gene database (DeSantis et al., 2006; Hazen et al., 2010; Probst et al., 2014). Although 10 to 100 fold more molecules of a single sample compared to an entire illumina HiSeq 2000 lane can be analyzed [www.illumina.com; (Probst et al., 2014)], the disadvantage of the PhyloChip technology was believed to be the limitation to already *a priori* 16S rRNA genes not allowing the discovery of novel phylogenetic taxa (Brodie et al., 2007; La Duc et al., 2009). In a recent publication we introduced a novel, empirical, and unsupervised method for taxon discovery from PhyloChip data (Probst et al., 2014). The new software termed 'sinfonietta' correlates probes across different arrays and across taxonomic affiliation to reconstruct 16S rRNA genes from the existing probe collection. Our data previously published in *Science Magazine* (Hazen et al., 2010) was re-analyzed, and an outlier sample was discovered not seen by the reference-based microarray analysis applied those days (Probst et al., 2014). Ergo, this technology appears to be a very attractive alternative to NGS-based 16S rRNA gene analysis enabling a deeper assay of microbial communities.

Each and every molecular analysis technique has its advantages and pitfalls, and as such, no single technique can be used to explain composition, dynamics, and function(s) of microbial communities. For instance, 16S rRNA gene analysis can be very useful to discover microbiome changes between different samples, but has great limitations in establishing evidence for microbiome function (Langille et al., 2013). Meanwhile, metagenomics enable the reconstruction of genomes and metabolic pathways from the environment but give no information on gene expression, enzyme activity or even the viability of a detected organism. Consequently, these and other high throughput technologies need to be combined with other (molecular) tools to unveil the countless facets at play in a typical environmental microbial community.

With a delay of approximately 10 years, planetary protection practitioners have at least partially applied the aforementioned novel methodologies of microbial community analysis to generate a comprehensive microbial inventory associated with spacecraft hardware and its assembly facilities. This delay in application is largely a consequence of the extremely low-biomass nature of such environments – many emerging technologies simply require more biomaterial than these ultra-clean surfaces typically harbor (Kwan et al., 2011; Probst et al., 2010a; Probst et al., 2011). Standard cultivation-dependent assays for heat-shock resistant microbes performed by both NASA and ESA are used regularly to estimate the total spore burden of spacecraft. This approach is well established and enables comparisons between recently gathered and previously acquired datasets [e.g., Viking era spacecraft (Puleo et al., 1977)]. In contrast, molecular assays that provide insight beyond the cultivable microbiome are rather fast developing and can become outdated quickly but provide unique insights into the microbiome beyond cultivability (Ward et al., 1990). In 2001, Venkateswaran and co-workers were the first to apply bacteria-directed 16S rRNA gene cloning on spacecraft assembly facility samples, which gave raise to a more enhanced understanding of microbial cleanroom diversity (Venkateswaran et al., 2001). The utilization of PhyloChip technology by LaDuc et al. and Cooper et al. facilitated the discovery of an even greater microbial diversity, discovering parts of the 'rare biosphere' present in cleanrooms (Cooper et al., 2011; La Duc et al., 2009). Results of studies focusing on the bacteriome in spacecraft assembly cleanrooms have led to the conclusion that A)

humans are the predominant source of microbial contaminants, and B) the cleanroom bacteriome very closely resembles the human bacteriome (Moissl et al., 2007).

Using 454 pyrotagsequencing, LaDuc et al. ascertained not only the bacterial diversity but also the archaeal and fungal diversity in samples taken from a rover of a recent Mars mission (La Duc et al., 2012). This study represented the first ever assessment of microbial diversity to encompass all three domains of life from one sample set. Archaea had, however, previously been detected in spacecraft assembly facilities in two independent reports (Moissl et al., 2008; Moissl-Eichinger, 2011). Current investigations at NASA's Jet Propulsion Laboratory comprise also a metagenomic survey of a cleanroom microbiome, which may provide great information on microbiome function (pers. com. Parag, Vaishampayan, Jet Propulsion Laboratory, CA, US, funded by NASA-ROSES 2011).

2.4. Obstacles in cleanroom microbiome research

Hitherto, all of the molecular-based studies of the cleanroom environment published have focused on alpha diversity and microbial community structure and development (Cooper et al., 2011; La Duc et al., 2003; La Duc et al., 2009; La Duc et al., 2012; Moissl et al., 2008; Moissl et al., 2007; Moissl-Eichinger, 2011; Probst et al., 2010b; Vaishampayan et al., 2010; Venkateswaran et al., 2001). While these reports were appreciatively descriptive, they failed to overcome two significant obstacles, namely fractional viability of the population and the contribution of its archaeal members. In these studies, it is unclear whether the entire set of molecular data gathered originated solely from viable microbial cells or a mixture of living cells and the remnants of dead microorganisms. Many cultivation-based studies have purified bacterial isolates from cleanroom samples, and the proportion of cultivable species in such samples has been reported to approach 35% (Moissl-Eichinger et al., 2013). This finding should not be taken lightly, as this index is 35 times greater than the 1% estimated for the entirety of Earth’s cultivable microbial species (Amann et al., 1995; Colwell, 1997). Ultimately, however, the living proportion and diversity of bacteria in cleanrooms has yet to be thoroughly assessed.

The second obstacle that remains unresolved is the presence and function of archaea in cleanroom facilities. Moissl et al. and Moissl-Eichinger have described the occurrence of archaea in all spacecraft assembly cleanrooms that they studied and described their diversity to be dominated by Thaumarchaeota (Moissl et al., 2008; Moissl-Eichinger, 2011), a phylum previously classified crenarchaeotic and comprised of putative ammnonia-oxidizers (Brochier-Armanet et al., 2008; Pester et al., 2011; Spang et al., 2010). With respect to the bacterial diversity, as mentioned above, biodiversity resembles the human skin microbiome since humans are thought to be the major source of microbial contaminants in spacecraft assembly cleanrooms (Moissl et al., 2007; Moissl-Eichinger et al., 2013). However, the human skin has been considered almost Archaea-free (Grice and Segre, 2011; Hulcr et al., 2012), and Thaumarchaeota have only been reported on two human palms as transient contaminants (Caporaso et al., 2011). Thus, the current knowledge of the human microbiome and its impact on cleanroom diversity do not explain the occurrence of archaea in these facilities.

3. The SM1 Euryarchaeon — facts, heuristics, hypotheses

3.1. The discovery of the SM1 Euryarchaeon and its life styles

In 2001, Rudolph and coworkers discovered streamers with string-of-peals like morphology floating in the streamlet of a marsh environment of the Sippenauer Moor, Germany. The water was reportedly strong with sulfidic odor and was thus identified to be sulfide-rich and oxygen-limited (Rudolph et al., 2001). Detailed diverse microscopic analyses of these string-of-pearls structures revealed that the interior of each pearl consisted primarily of archaeal cocci, while the outside was formed by filamentous bacteria. These bacteria, later characterized as *Thiothrix* sp. (Moissl et al., 2002), also formed the string spanning the various pearls. This spatially coordinated association of novel archaea and (potentially) sulfide-oxidizing bacteria was appropriately termed 'string-of-pearls community' (Rudolph et al., 2001). This coordination was illustrated in many electron microscopy (EM) and fluorescence *in situ* hybridization (FISH) micrographs (Moissl et al., 2002; Moissl et al., 2003; Rudolph et al., 2004; Rudolph et al., 2001); an example is given in *Figure II.3-1*. The cells of the archaeon were ca. 0.5 µm in diameter and, while embedded in this matrix, displayed hundreds of surface-bound appendages called *hami* [plural; singular: *hamus*; (Moissl et al., 2005b), please see also chapter II.3.2.].

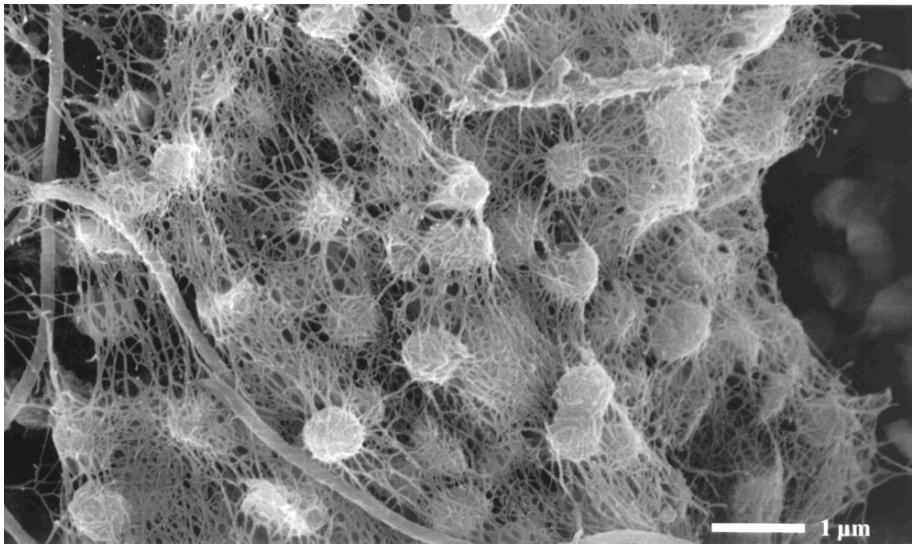

Figure II.3-1 *Scanning electron micrograph of the inner of a pearl. Small cocci (archaea) are embedded in a dense network of cell surface appendages and surrounded by filamentous bacteria (Rudolph et al., 2001).*

Using 16S rRNA gene amplicon cloning, Rudolph et al. also characterized the novel archaeon phylogenetically and demonstrated that it branched deeply within the Euryarchaeota but without any cultivated representative in the entire clade of environmental sequences. Due to its novelty, there was no prior knowledge on the environmental function of these archaea nor on their potential metabolism. The archaeon was refractory to cultivation under laboratory conditions, and was thus termed 'SM1 Euryarchaeon', whereas SM stands for Sippenauer Moor, the first biotope where the archaeon was discovered (Rudolph et al., 2001).

The string-of-pearls community was grown *in situ* on polyethylene nets in the streamlets of the Sippenauer Moor. This method facilitated the collection of appreciable amounts of SM1 Euryarchaeon

and associated community biomass with the help of state-of-the art purification techniques (Moissl et al., 2003). Although this community was referred to as the net population (Moissl et al., 2003), it was very similar to the original string-of-pearls community and is henceforth not differentiated herein. Growing the string-of-pearls community in its actual biotope and harvesting high amounts of biomass sparked many subsequent studies that investigated the archaea and the community at molecular level [(Moissl et al., 2005b; Moissl et al., 2003), please see also chapter III.3. and III.4.].

In another study of the group around Robert Huber the distribution of the SM1 Euryarchaeon was analyzed, and many other environmental 16S rRNA genes of the so-called SM1 group were retrieved (Rudolph et al., 2004). The study focused on sulfidic springs in Bavaria but also included samples from Turkey. All springs had representative sequences of the SM1 clade indicative of a certain environmental success of these microorganisms. Importantly, identical 16S rRNA gene sequences of the SM1 Euryarchaeon were retrieved from the Mühlbacher Schwefelquelle, Germany (*Figure II.3-2*). This sulfur spring is an artificial bore-hole and emits thousands of liters of water per hour; the chemical composition and temperature of its water was very similar to that found at Sippenauer Moor (Rudolph et al., 2004). In the streamlet a similar string-of-pearls community was found, also being comprised of the SM1 Euryarchaeon in the inner of the pearl but surrounded by the IMB1 Epslionproteobacterium; the environmental clade of these microoganisms was later classified *Sulfuricurvum* (genus) due to the cultivation of a representative (Kodama and Watanabe, 2004). *Sulfuricurva* are filamentous bacteria capable of sulfide-oxidation and may therefore fulfill a similar symbiotic role for SM1 Euryarchaea at the Mühlbacher Schwefelquelle as *Thiothrix* sp. at the Sippenauer Moor (Rudolph et al., 2004).

Figure II.3-2 Mühlbacher Schwefelquelle (Regensburg/Burgweinting, Germany). The bore-hole has an approximate diameter of 25 cm and is man-made (Probst et al., 2013b).

Two years after the description of the Mühlbacher Schwefelquelle as the SM1 Euryarchaeon's second biotope, an additional life style of the SM1 Euryarchaeon was discovered. Henneberger and coworkers exposed polyethylene nets to the subsurface of the spring with an approximately 1 m

depth at the Mühlbacher Schwefelquelle (Henneberger et al., 2006). Slimy, milky biofilm droplets settled within minutes onto the nets allowing the conclusion that they were washed up from deeper Earth layers. These biofilms were almost completely comprised of SM1 Euryarchaea with only a few bacterial cells (purity ca. 95%). The 16S rRNA gene sequence was identical to the SM1 Euryarchaea found in the string-of-pearls communities, which led scientists to conclude that the archaea are the same species. Analysis of the intergenic spacer region [the genomic region between 16S rRNA gene and 23S rRNA gene, which is supposedly highly variable between strains (Jensen et al., 1993)] via PCR amplification and restriction fragment length polymorphism demonstrated that these archaea were even the same strain (Henneberger et al., 2006). Furthermore, analysis of the genome size demonstrated a similar size of ca. 3 Mbps of the SM1 Euryarchaeon at both sampling sites (Henneberger et al., 2006; Moissl et al., 2003). However, the appearance of the biofilm samples were greatly different from the string-of-pearls community. Within the biofilm, the archaea maintained a consistent proximity to one another, likely due to their cell surface appendages (please see also II.3.2). The cells were embedded in a polysaccharide matrix, and most of the bacterial cells reacted positively to sulfate-reducer directed probes in hybridization studies. Consequently, the biofilm represents a new life-style of the SM1 Euryarchaeon (Henneberger et al., 2006), and the bore-hole at the Mühlbacher Schwefelquelle is a window to a third biotope of the SM1 Euryarchaeon.

Table II.3-1 Data and characteristics of biotopes of the SM1 Euryarchaeon (data publicly available in January 2011).

Biotope	**Sippenauer Moor**		**Mühlbacher Schwefelquelle**	
Life style	**Biofilm**	**String-of-pearls community**	**Biofilm**	**String-of-pearls community**
Dominant archaeon	SM1 Euryarchaeon	SM1 Euryarchaeon	SM1 Euryarchaeon	SM1 Euryarchaeon
Potential bacterial partner	unknown	*Thiothrix* sp.	possibly sulfate-reducers, details unknown	*Sulfuricurvum* sp.
Percentage of archaea (approx.)	90%	50%	95%	50%
Accessibility of biomass[1]	+	++	++	(+)
Characteristics	similar to biofilm from Mühlbacher Schwefelquelle, details unknown	microcolony of SM1 surrounded by *Thiothrix* sp. as streamers	biofilm mainly comprised of SM1 Euryarchaeon with diverse bacteria	microcolony surrounded by *Sulfuricurvum* sp. as streamer
Origin	subsurface, depth unknown	streamlet	subsurface, depth unknown	streamlet
Reference	Henneberger et al., 2006	Rudolph et al., 2001 / Moissl et al., 2002 / Moissl et al., 2003	Henneberger et al., 2006	Rudolph et al., 2004
Reference for water composition	Rudolph et al., 2001 / Rudolph et al., 2004	Rudolph et al., 2001 / Rudolph et al., 2004	Rudolph et al., 2004	Rudolph et al., 2004

[1] (+) little, + medium, ++ high amount of biomass can be harvested throughout the entire year.

Over the course of the aforementioned study, minor investigations were also performed on biofilm droplets that could be harvested at the Sippenauer Moor springs. These biofilm droplets were

also comprised of SM1 Euryarchaea, but to date, no analysis beyond 16S rRNA gene analysis (sequencing and FISH) has been performed. Consequently, there are four reported biotopes of the SM1 Euryarchaeon at two different sampling sites that have a distance of approximately 20 km. A summary of the four biotopes and the different life styles of the SM1 Euryarchaeon is provided in *Table II.3-1*.

3.2. The extraordinary traits of the SM1 Euryarchaeon

To date, 116 different archaeal genera have been validly described and made available in public databases (Euzeby, 1997; Parte, 2014). This provides many opportunities to study the biology of Archaea via for instance genetic systems (Contursi et al., 2003; Metcalf et al., 1997). However, the uncultivated SM1 Euryarchaeon features several specialties that demand studying the biology of particular this archaeon in order to advance the field of Archaea research.

The SM1 Euryarchaeon forms the only **natural archaeal biofilm** that has been described in the literature. Archaeal biofilms have sparsely been reported in the literature, and the necessity to study them has intensively been reviewed (Fröls, 2013; Orell et al., 2013). Unfortunately, none of the aforementioned reviews considered — nor made the slightest mention of — the SM1 Euryarchaeon biofilm (Henneberger et al., 2006).

Examples of biotopes that are **dominated by one archaeal species** are scarce in the literature (Jackson et al., 2013) and represent therefore the exception rather than the rule. The SM1 Euryarchaeon biofilms appear to dominate the subsurface biotope that can be accessed through the Mühlbacher Schwefelquelle bore-hole [and also through various springs at the Sippenauer Moor (Henneberger et al., 2006)].

Due to the **constancy of the biotope**, samples of the SM1 Euryarchaeon biofilm and string-of-pearls community can be continuously harvested in high biomass independently of the seasonal fluctuations (Henneberger et al., 2006; Moissl et al., 2003). The constancy and the dominance of the SM1 Euryarchaeon in its biotopes enables researchers to study the biology of an uncultivated, cold-loving Archaeon sampled in its natural environment even down to the ultrastructural and genetic level (Moissl et al., 2005b).

These ultrastructural and genetic analyses of the SM1 Euryarchaeon revealed one of the most fascinating features of microbial life [*Figure II.3-3*, (Moissl et al., 2005b)]: Each individual cell of SM1 Euryarchaea carries hundreds of **cell surface appendages**, called *hami* (singular *hamus*). These resemble a barbwire structure, carry a distal hook, enable cells to attach to (biotic and abiotic) surfaces and to connect to each other (Moissl et al., 2005b). Hitherto, these cell surface appendages have not been observed for any other life form known to man.

Biofilm samples of the SM1 Euryarchaeon originate from deeper sediments and thus represent **subsurface microbial life** (Henneberger et al., 2006; Probst et al., 2013b). In general, subsurface microbial life represents a great proportion of the Earth's life forms but remains poorly understood due to the lack of access to samples (Whitman et al., 1998). Nevertheless, the contribution of subsurface microbial life to Earth's carbon, nitrogen, and sulfur cycling has already been demonstrated as has its potential to harbor completely novel microbial phyla (Castelle et al., 2013; Wrighton et al., 2012).

For all of these reasons, the SM1 Euryarchaeon represents an ideal **model system** for studying archaeal biofilms that have originated in the subsurface (Probst et al., 2013b).

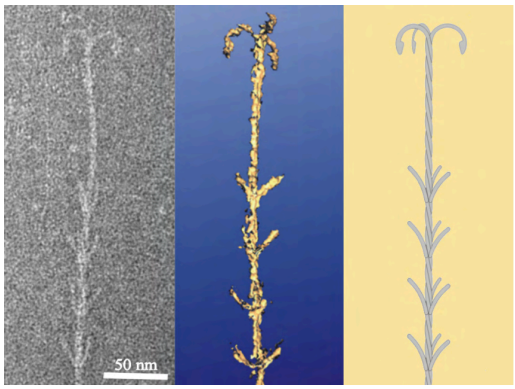

Figure II.3-3 *Cell surface appendages (hami) of the SM1 Euryarchaeon. Left panel: Transmission electron microscopy of a hamus; middle panel: cryo-electron tomography of a hamus (same scale as left panel); right panel: idealized model of a hamus (Moissl, 2004), which was, however, revised in 2005 (Moissl et al., 2005a).*

3.3. Knowledge gaps in SM1 Euryarchaeon research

Although many studies about the SM1 Euryarchaeon have appeared in the literature (Henneberger et al., 2006; Moissl et al., 2005b; Moissl et al., 2002; Moissl et al., 2003; Moissl-Eichinger et al., 2012a; Rudolph et al., 2004; Rudolph et al., 2001), there are several inconsistencies and knowledge gaps in understanding the biology of this extraordinary, uncultivated archaeon.

Henneberger and coworkers described the same strain of SM1 Euryarchaea to be found at the two sampling sites, Mühlbacher Schwefelquelle and Sippenauer Moor, since the 16S-23S intergenic spacer region was identical based on restriction fragment length polymorphism (Henneberger et al., 2006). However, restriction fragment length polymorphism does not provide the sensitivity necessary to look into single base pair mutations that may occur between closely related strains. In addition, different 16S-23S intergenic spacer regions lead to the conclusion of different microbial strains, but identical regions do not allow the conclusion of identical strains. Moreover, the intergenic spacer region can even vary within genomes if multiple rRNA operons are present [for details on intergenic spacer regions between strains and within one genome the reader is referred to the work by Nagpal and co-workers (Nagpal et al., 1998)]. Southern blot analysis performed with *hamus*-directed probes and metagenomic DNA from both sampling sites revealed different restriction patterns, indicative of different SM1 Euryarchaeon genomes present at the Mühlbacher Schwefelquelle and Sippenauer Moor sites (Moissl, 2004). Consequently, the hypothesized theory that one SM1 Euryarchaeon strain forms both biofilms at the Mühlbacher Schwefelquelle and Sippenauer Moor may not reflect the truth (Henneberger et al., 2006) and necessitates further investigation.

The bacterial counterpart(s) of the biofilms comprising the SM1 Euryarchaeon has/have not been studied in nearly enough detail. Although Henneberger et al. used a sulfate-reducer directed FISH probe (Henneberger et al., 2006), the sensitivity of the probe has been demonstrated to be insufficient (Loy et al., 2002; Loy et al., 2005). However, the function of the bacteriome as primarily sulfate-reducers has not been proven, nor has its phylogenetic diversity been determined. The bacteriome of the biofilm at the Sippenauer Moor has not been described at all. A comparative

analysis of the bacteriomes of the two biofilms could contribute to a better understanding of the two biofilms' relatedness.

The SM1 Euryarchaeon was found to form the string-of-pearls community and biofilms at two sampling sites, yet the relationship between the two life styles has not been explored. It can be hypothesized that the biofilm is a precursor of the string-of-pearls community, but an independent growth of different SM1 Euryarchaea at the surface and in the subsurface could also be possible. Preliminary indications that the former hypothesis may be correct were identified by the above-mentioned southern blot analysis with *hamus*-directed probes. Restriction patterns of the biofilm at the Sippenauer Moor were identical to the restriction patters of the string-of-pearls community. However, additional investigations are necessary to further support this hypothesis.

The phylogenomic and taxonomic position of the SM1 Euryarchaeon is currently unclear, for several reasons. Although 16S rRNA gene based phylogenetic relationships were determined (Rudolph et al., 2001), only approximately 1.0 kbps of the entire 1.5 kbps gene could be used. Mismatches of archaea-directed forward primers in the front region of the gene did not allow an amplification of the SM1 Euryarchaeon 16S rRNA gene (Rudolph et al., 2001). As described in chapter II.1, the phylogenetic information retrieved from 16S rRNA genes does not always correctly display the evolutionary relatedness of microorganisms. Consequently, it is necessary to obtain amino acid sequence data of conserved genes for performing a phylogenomic placement of the SM1 Euryarchaeon within the tree of Archaea.

As mentioned above, the 16S rRNA gene based phylogenetic clade of Euryarchaeota closely related to the SM1 Euryarchaeon remains widely unexplored. Consequently, no information is available with regard to metabolic capabilities of close relatives that could lead to a potential hypothesis; however, inferring metabolic functions from 16S rRNA gene phylogeny has anyways only partially been accepted by the scientific community (Langille et al., 2013). This may solely be applicable for certain phylogenetic clades (e.g. methanogens) or well-studied taxa. Nevertheless, Moissl et al. hypothesized a metabolic function of the SM1 Euryarchaeon (Moissl et al., 2002). Due to its intimate association with *Thiothrix*, whose main metabolic activity was described as sulfide oxidation (Harold and Stanier, 1955), the SM1 Euryarchaeon could supposedly be a sulfate-reducer living with *Thiothrix* in a syntrophic relationship based on an interspecies sulfur cycle (Moissl et al., 2002; Morris et al., 2013). This hypothesis is supported by the observation that the only cultivated representative of the genus *Sulfuricurvum* — closely related to the partner of SM1 Euryarchaeon at the Mühlbacher Schwefelquelle — has also a sulfide-based metabolism (Kodama and Watanabe, 2004). Although there is a hypothesis about the terminal electron acceptor of the SM1 Euryarchaeon (sulfate), there is now clue about the potential electron donors, carbon, nitrogen, phosphorous sources that this archaeon can utilize. The current heuristics need to be replaced with direct evidence and knowledge gaps need to be filled by applying appropriate methodologies. Such methodologies may include (meta)genomics, which may provide stronger evidence for the understanding of the SM1 Euryarchaeon's opaque metabolism and may in turn lead to its cultivation under laboratory conditions.

4. Scope of this thesis and publication guide

The scope of this thesis was to investigate the **archaeome** of two independent biotopes, namely spacecraft assembly facilities and sulfidic groundwater aquifers near Regensburg, Germany. Concurrently, the cohabiting **bacteriome** and potential bacterial partners, respectively, were analyzed to better understand the types of microbiota contributing to their corresponding biotopes. The missing pieces to these puzzles were discussed previously in chapter II.2.4. ('Obstacles in cleanroom microbiome research') and in chapter II.3.3. ('Disagreements and knowledge gaps in SM1 Euryarchaeon research'). These discussions provided the knowledge base for the investigations presented in detail henceforth in this thesis. The following two paragraphs serve as a guide for chapter III ('Publications') and summarize the major findings.

4.1. Expanding the current knowledge base on cleanroom microbiomes

Cleanroom microbiomes, bacteriomes in particular, have been investigated extensively, and have been shown to coincide significantly with the human (skin) microbiome. However, the proportion of analyzed biosignatures belonging to viable microorganisms remained cryptic. The manuscript entitled '**New perspectives on viable microbial communities in cleanroom environments**' (III.1.), ponders the viable fraction of the bacteriome associated with spacecraft assembly cleanrooms. This article presents the findings of the first investigation of the viable versus expired nature of the cleanroom microbiome.

The **archaeome** in spacecraft assembly cleanrooms has been investigated deeply by Moissl-Eichinger (Moissl et al., 2008; Moissl-Eichinger, 2011), yet the origin of these microorganisms has not been determined. In the publication '**Archaea on human skin**' (III.2.) the origin of archaeal cleanroom signatures was revealed by comparing samples from cleanroom facilities, intensive care units, and human skin. This investigation yielded the first report in the scientific addressing and defining the human skin archaeome.

4.2. Investigating the SM1 Euryarchaeon's biofilm and its potential bacterial partners

The **bacteriome** of the SM1 Euryarchaeon biofilm found at the Mühlbacher Schwefelquelle has been analyzed solely with FISH using a sulfate-reducing bacteria-directed probe (Henneberger et al., 2006). The manuscript entitled '**Tackling the minority: Sulfate-reducing bacteria in an archaea dominated subsurface biofilm**' (III.3.) contains compelling evidence for the presence of active sulfate-reducing bacteria in the SM1 Euryarchaeon biofilm and thoroughly deciphered the bacteriome composition of the SM1 Euryarchaeon biofilm from the Mühlbacher Schwefelquelle.

A comparative study of the two SM1 Euryarchaeal biofilms that can be found at Mühlbacher Schwefelquelle and Sippenauer Moor, respectively, was conducted in the manuscript '**Coupling genetic and chemical microbiome profiling reveals heterogeneity of archaeome and bacteriome in subsurface biofilms that are dominated by the same archaeal species**' (III.4.). The composition of and differences between the two biofilms were analyzed at various levels of resolution from entire community structure and microbiome relatedness down to the strain level. All these levels of resolution showed a divergence of the two biofilms and shed light onto the variability of subsurface microbial life in general.

A metagenomic survey to thoroughly decipher the genetic information of the SM1 Euryarchaeon was performed in the manuscript '**Grappling with dark matter: Biology of an uncultivated subsurface archaeon**' (III.5.). The genome of the SM1 Euryarchaeon was successfully assembled and binned from metagenomic reads. Phylogenomics revealed that the SM1 Euryarchaeon is a representative of a novel euryarchaeal lineage, and metabolic pathway reconstructions enabled the prediction of carbon and nitrogen sources of these archaea.

III. Publications

Overview

This cumulative dissertation comprises five articles. Three manuscripts have been published in, one has been submitted to and one is in preparation for submission to peer-reviewed journals. In any case, the PhD student, Alexander J. Probst, has authored every manuscript printed herein as a first author. Supplementary information and additional data can be found on the supporting DVD. Concerning the data structure on the supporting DVD the reader is referred to Chapter VII.

The PhD student's contributions to the manuscripts are as follows:

1. Vaishampayan, Probst, LaDuc et al., 2013: Alexander performed bioinformatics and biostatistical analysis and wrote the paper.
2. Probst et al., 2013a: Alexander performed experiments, bioinformatics, analyzed data and wrote the paper.
3. Probst et al., 2013b: Alexander performed experiments, bioinformatics and biostatistical analysis and wrote the paper.
4. Probst, Birarda et al., submitted: Alexander performed experiments, bioinformatics and biostatistical analysis and wrote the paper.
5. Probst et al., in preparation: Alexander performed experiments, bioinformatics and biostatistical analysis and wrote the paper.

A full list of publications authored by the PhD student are listed on page 7 to 9 and the articles are individually provided on the supplementary DVD.

Prof. Dr. Reinhard Wirth

1. New perspectives on viable microbial communities in low-biomass cleanroom environments

Parag Vaishampayan[1,4], Alexander J Probst[2,4], Myron T La Duc[1,4], Emilee Bargoma[1], James N Benardini[1], Gary L Andersen[3] and Kasthuri Venkateswaran[1]

[1]Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA; [2]Department for Microbiology and Archaea Center, University of Regensburg, Regensburg, Germany and [3]Earth Sciences Division, Ecology Department, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[4]These authors contributed equally to this work.

Correspondence: K Venkateswaran, California Institute of Technology, Jet Propulsion Laboratory Biotechnology and Planetary Protection Group; M/S 89-108, 4800 Oak Grove Drive, Pasadena, CA 91109, USA. E-mail: kjvenkat@jpl.nasa.gov

Publication information:
The ISME Journal (2013) 7, 312–324; doi:10.1038/ismej.2012.114; published online 11 October 2012
Received 3 April 2012; revised 8 August 2012; accepted 14 August 2012
Link: http://www.nature.com/ismej/journal/v7/n2/full/ismej2012114a.html

Abstract

The advent of phylogenetic DNA microarrays and high-throughput pyrosequencing technologies has dramatically increased the resolution and accuracy of detection of distinct microbial lineages in mixed microbial assemblages. Despite an expanding array of approaches for detecting microbes in a given sample, rapid and robust means of assessing the differential viability of these cells, as a function of phylogenetic lineage, remain elusive. In this study, pre-PCR propidium monoazide (PMA) treatment was coupled with downstream pyrosequencing and PhyloChip DNA microarray analyses to better understand the frequency, diversity and distribution of viable bacteria in spacecraft assembly cleanrooms. Sample fractions not treated with PMA, which were indicative of the presence of both live and dead cells, yielded a great abundance of highly diverse bacterial pyrosequences. In contrast, only 1% to 10% of all of the pyrosequencing reads, arising from a few robust bacterial lineages, originated from sample fractions that had been pre-treated with PMA. The results of PhyloChip analyses of PMA-treated and -untreated sample fractions were in agreement with those of pyrosequencing. The viable bacterial population detected in cleanrooms devoid of spacecraft hardware was far more diverse than that observed in cleanrooms that housed mission-critical spacecraft hardware. The latter was dominated by hardy, robust organisms previously reported to survive in oligotrophic cleanroom environments. Presented here are the findings of the first ever comprehensive effort to assess the viability of cells in low-biomass environmental samples, and correlate differential viability with phylogenetic affiliation.

Introduction

Microbial cells are traditionally classified as either viable (maintaining active metabolism and membrane integrity), viable but dormant (because of external pressures) or non-viable [dead; (Kaprelyants et al., 1993; Keer and Birch, 2003)]. The vast majority of microorganisms cannot be cultivated and many that can require long cultivation times (Amann et al., 1995). To minimize the time spent in determining viability and bias associated with such analyses, advanced molecular approaches for assessing cellular viability have been recently developed, including live–dead staining (Boulos et al., 1999) and flow cytometry-based techniques (Ben-Amor et al., 2005). In such cases, however, there is inherent risk of overestimating the total number of viable cells in samples because of variation in binding affinities of the dyes used. Methods probing RNA as an alternative to DNA have also been developed for assessing viability (DeAngelis et al., 2011), but these too come with complications as RNA is difficult to purify, less stable, prone to degradation and often isolated in quantities insufficient for analysis (Andorra et al., 2010; Hierro et al., 2006).

The development of high-throughput pyrosequencing and phylogenetic microarray techniques has dramatically increased the resolution and detectable spectrum of diverse microbial lineages from environmental samples (La Duc et al., 2009; Mendes et al., 2011; Sogin et al., 2006). However, DNA-based molecular technologies alone have yet to be validated for assessing the differential viability of cells across varying phylogenetic lineages in a complex microbial assemblage. DNA-derived signals, which originate from both living and dead cell types, often lead to interpretations that overestimate the viable microbial population present (Pointing et al., 2009; Rogers et al., 2008). Quantitative PCR (qPCR)-based analyses increase the speed, sensitivity and specificity of quantitative microbial detection (Hierro et al., 2006; Yanez et al., 2011). Unfortunately, qPCR techniques alone cannot differentiate live from dead cells, as the latter contribute template DNA to the overall PCR amplification (Cawthorn and Witthuhn, 2008). The persistence of nucleic acids post-cell death renders cogent estimations of live–dead cellular ratios in a given sample practically impossible, as naked DNA will ultimately result in the overestimation of total cells (Nocker et al., 2006).

The treatment of microbial cell suspensions with propidium monoazide (PMA), which first intercalates and on photo-activation, covalently binds DNA, before DNA extraction followed by downstream PCR and related molecular analyses has become an increasingly popular technique for the selective detection and enumeration of viable microbes (Bae and Wuertz, 2009; Hein et al., 2006; Nocker et al., 2006; Wagner et al., 2008). The use of PMA for discriminating live from dead cells has been studied across various applications and research testbeds, including food (Cawthorn and Witthuhn, 2008), biosolids (van Frankenhuyzen et al., 2011), infectious enteric viruses in water samples (Parshionikar et al., 2010), fungi (Vesper et al., 2008), bacteriophage T4 (Fittipaldi et al., 2011), infectious parasitic protozoa (Brescia et al., 2009) and wastewater treatment plants (Lin et al., 2011). Recently, PMA treatment has been used in combination with crude microarray analysis for assessing cell viability (Nocker et al., 2009). In addition, pyrosequencing profiles of water samples exposed to high temperature and treated with and without PMA have been comparatively analyzed (Nocker et al., 2010).

The development and validation of molecular methods to selectively detect and enumerate the living fraction of the microbial population resident on critical surfaces (for example, hospital operating rooms, pharmaceutical manufacturing and packaging facilities, semiconductor fabrication, and spacecraft assembly cleanrooms) is of immense importance to good manufacturing practices aimed at minimizing contaminant bioburden levels. Reported here for the first time are the results of pioneering efforts coupling PMA-based viability discrimination with innovative PhyloChip DNA microarray and bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) methodologies to assess the viable bacterial population present in a typical low-biomass environmental sample.

Material and Methods

Sampling location. The two cleanroom facilities examined in this study were both certified as ISO 8. Spacecraft hardware and componentry was housed and assembled in the first cleanroom site, Jet Propulsion Laboratory's (JPL's) Spacecraft Assembly Facility (SAF; sample #GI-36; *Table III.1-1*), whereas the second cleanroom site, JPL cleanroom Building (Bldg)144 (sample #GI-42; *Table III.1-1*), did not house spacecraft hardware at the time of sample collection. Both of these cleanroom facilities operated at a positive pressure, with temperatures in the range of 20±4°C, and relative humidity ranging from 30% to 50%. Ground support equipment (GSE) consisted of all non-flight hardware items used during spacecraft hardware

receipt, assembly, integration, test, storage, shipment and pre-launch activities. All GSE materials used inside the cleanrooms were inspected for compliance to visible cleanliness.

Table III.1-1 Physical, chemical and bacterial characteristics of samples collected during this study

Sample ID	Sample type[a]	123-bp qPCR (16S rRNA copies m^{-2})	Total number of PhloChip-detected genera[b]	Total number of bTEFAP-derived MOTU[c]
SAF (during a spacecraft assembly):				
GI-36-4	Clean room floor	6.70×10^5	94	122
GI-36-4(p)	PMA-treated clean room floor	4.93×10^4	9	4
GI-36-3	GSE	1.85×10^6	411	425
GI-36-3(p)	PMA-treated GSE	7.50×10^4	3	17
Bldg 144 (no mission operation):				
GI-42-1	Clean room floor	4.46×10^7	199	447
GI-42-1(p)	PMA-treated clean room floor	9.25×10^6	106	108
GI-42-2	GSE	2.68×10^7	236	571
GI-42-2(p)	PMA-treated GSE	1.86×10^6	75	42

Abbreviations: Bldg, building; bTEFAP, bacterial tag-encoded FLX amplicon pyrosequencing; GSE, ground support equipment; JPL, Jet Propulsion Laboratory; MOTU, molecular operational taxonomic unit; PMA, propidium monoazide; qPCR, quantitative PCR; rRNA, ribosomal RNA; SAF, spacecraft assembly facility.

[a]Nine individual samples (each 1 m^2) were collected using Biological Sampling Kit and pooled. All filtered samples were divided into two separate aliquots (500 µl each; equivalent to 4.5 m^2), one to be subjected to PMA pre-treatment (viability assessment), and the other to serve as a null environmental sample (viable +non-viable; total DNA). The JPL–SAF is the most frequently utilized cleanroom facility, as spacecraft assembly was underway at the time of sampling. In contrast, Bldg 144 was inactive and not in use when samples were collected. The only human traffic in the Bldg 144 facility before sampling was the janitorial servicing, which occurred once a week, or to address any other miscellaneous maintenance issues. Other metadata such as usage rate, and so on, are not in place. The dimensions of the JPL-SAF cleanroom were larger and surface area of the GSE materials were more in JPL–SAF than in Bldg 144.

[b]The amount of total 16S rRNA PCR product subjected to hybridization on PhyloChips was normalized across samples (~400ng) whenever possible.

[c]The total volume of initial PCR product used for subsequent emulsion PCR was 2 ml for strong positives (>10 ng $µl^{-1}$; all non-PMA samples), 5 µl for weak positives [5 to 10 ng $µl^{-1}$; GI-42-1(p)] and 20 µl for samples that failed to yield PCR products [<5 ng $µl^{-1}$; all PMA-treated samples except GI-42-1(p)]. This normalization step enhanced retrieval of maximum number of sequences.

An all-purpose cleaning and degreasing agent (Kleenol 30, Accurate Industrial Supply, Inc., Cerritos, CA, USA, Cat #: J-CC-00040) was used to maintain cleanliness of the floor. Surface cleaning procedures were performed twice a day in the cleanroom during periods when spacecraft componentry was actively undergoing assembly (SAF; sample #GI-36). In contrast, the quiescent Bldg 144 environmental test facility certified cleanroom was cleaned only once a week since spacecraft hardware was not present in this facility at the time of sample collection. Both of the cleanroom facilities examined were maintained with stringent protocols pertaining to the replacement of tacky mats, vacuuming and mopping of floors, and wiping down of GSE surfaces with alcohol. In addition, before entering, staff were required to don cleanroom garments and comply with appropriate practices to minimize the influx of particulate matter.

Sample collection. Samples were collected from the cleanroom floors and GSE housed in SAF and Bldg 144. Wet-surface sampling of the cleanroom floors and GSE (each 1 m^2) was performed using biological sampling kit (BiSKit; QuickSilver Analytics, Abingdon, MD, USA) as previously described (Kwan et al., 2011). To measure indigenous DNA associated with the sampling device (negative control), 15 ml of sterile phosphate-buffered saline (PBS, pH 7.0, Mo Bio Laboratories Inc., Carlsbad, CA, USA) was first processed through each BiSKit sampling module, with the expelled liquid pooled before using the very same kit for sample collection. The collection bottle corresponding to each BiSKit sampling module was replenished with 15 ml of sterile PBS and the entire assembly was inverted, saturating the macrofoam sponge. The sampling module of the BiSKit assembly was then used to sample floors and GSE sites in the ternary, unidirectional manner described elsewhere (Kwan et al., 2011). Overall, 36 individual samples (each 1 m^2) were collected from two distinct cleanroom floors and two distinct GSE locations. A detailed description of the sample types is provided in *Table III.1-1*. Sterile water, PBS, DNA extraction reagent blanks and PCR reagent blanks used in sample collection, processing and analysis, respectively, served as negative controls in all molecular analyses.

Sample processing. Sample volumes were extracted from BiSKit devices as instructed by the manufacturer, a total of three times with 15 ml of PBS each. Previous studies have demonstrated that spacecraft-associated surfaces house extremely low-biomass and seldom yield detectable PCR amplification products (Moissl et al., 2007; Vaishampayan et al., 2010).

Hence, all samples collected from the same general location or site were pooled together. The biological materials resulting from each pooled sample (~400 ml each) were further concentrated using Amicon Ultra-50 Ultracel centrifugal filter tubes (Millipore, Billerica, MA, USA). Each filter unit has a molecular mass cutoff of 50 kDa, which facilitates the concentration of bacterial cells, spores and exogenous nucleic acid fragments >100bp in a final volume of 1 ml. All filtered samples were divided into two separate aliquots (500 µl each), one to be subjected to PMA pre-treatment (viability assessment), and the other to serve as a null environmental sample (viable+non-viable; total DNA). All samples, both with and without PMA pre-treatment were subjected to DNA extraction via the Maxwell 16 automated system (Promega, Madison, WI, USA) according to the manufacturer's instructions, and resulting DNA suspensions (100 µl each) were stored at -20 °C.

PMA treatment. A 500 µl aliquot of filter-concentrated sample suspension was treated with 12.5 µl of PMA (2 mM; Biotium, Inc., Hayward, CA, USA) to a final concentration of 50 µM (Nocker et al., 2010; Rawsthorne et al., 2009), followed by thorough mixing and incubation in the dark for 50 min at room temperature. The tubes were inverted manually 5–6 times over a 10-min incubation interval to promote homogeneous PMA exposure. Sample tubes were then placed horizontally atop a bed of ice and exposed to a 500-W halogen lamp (Osram 64553 C318; Danvers, MA, USA) at a distance of 20 cm for 3 min. Samples without PMA treatment were also subjected to incubation in dark for 50 min and exposure to 500-W halogen lamp on ice along with PMA-treated samples.

Quantitative PCR. Bacteria-directed primers (1369F and 1492R) targeting the 16S ribosomal RNA (rRNA) gene were used for qPCR analysis (Suzuki et al., 2000). Each 25 µl reaction mixture consisted of 12.5 µl of Bio-Rad 2X iQ SYBR Green Supermix (Hercules, CA, USA), 10.5 µl of UltraPure water (Gibco; Grand Island, NY, USA), 0.5 µl of forward primer 1369F (10 µM), 0.5 µl of reverse primer 1492R (10 µM) and 1 µl of DNA template. Purified standards and UltraPure Gibco water no-template controls were included in all qPCR runs. Thermal cycling parameters for universal 16S rRNA gene qPCR were as follows: hold at 95°C for 3 min to achieve initial denaturation, followed by 40 cycles of: 10-s hold at 95°C to denature, ramp-down to 55°C for primer annealing and extension occurring through a 35-s ramp-up to 95°C. In this study, all samples were analyzed in triplicate.

Taxonomy. For convenience and differentiation in this communication, bTEFAP-based pyrosequence discrimination results are binned hierarchically into what are referred to as molecular operational taxonomic unit(s) [MOTU; (Blaxter, 2003; Blaxter et al., 2005)]. Variation in DNA sequence among microbes can arise via naturally occurring evolutionary events and/or methodological errors (for example, homopolymer repetition in pyrosequencing). It is the goal of the MOTU-based classification and clustering system presented here, and in detail elsewhere (Blaxter and Floyd, 2003), to separate these two sources of sequence variation based on known error rates in sequencing and measured levels of difference across various taxonomical schemes. The accuracy and specificity of a MOTU-based system can be derived from measured levels of between-taxa and within-group variation from well-defined populations, and of observational error obtained by re-sequencing (Blaxter and Floyd, 2003). In a similar vein, PhyloChip DNA microarrays use multiple ~25-bp probes, which collectively represent the full-length 1.5-kb 16S rRNA gene of each taxon. In this study, PhyloChip-derived taxonomic units (PTUs) were delineated in accordance with the hybridization scores of a given set of 25-mer probes, which have been previously designed based on the prevalence of members of a given PTU, and dissimilarity in DNA sequences outside of the given PTU. Ultimately, a microorganism can be assigned to only one given MOTU/PTU, either via similarity within a homologous sequenced DNA fragment (MOTU) or hybridization score (PTU), but neither MOTU nor PTU need be congruent with other taxonomic schemes.

PhyloChip G3. Bacterial 16S rRNA genes were amplified from DNA preparations from each sample using the primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). PCR conditions were as follows: 1 cycle of initial melting for 3 min at 95°C, followed by 35 cycles of 30-s melting at 95°C, 30-s annealing over a 48–58°C gradient (48°C, 48.8°C, 50.1°C, 51.9°C, 54.4°C, 56.3°C, 57.5°C and 58°C), and 2-min extension at 72°C, with a final 10-min incubation at 72°C. The amount of total 16S rRNA gene PCR product subjected to hybridization on PhyloChips was normalized across samples (~400 ng) whenever possible. A detailed explanation of the processing of the PhyloChip assay

has been described elsewhere (Hazen et al., 2010). Stage 1 and stage 2 analysis were performed and the cross-hybridization response score was adjusted as previous described (Cooper et al., 2011).

PhyloChip G3 data analysis. Following stage 2 analysis, hybridization intensities were transformed (log_2*1000) and were henceforth referred to as transformed hybridization intensities. Representative PTU sequences were then compared against the taxonomic architecture of the SILVA database (Pruesse et al., 2007) and PTU were grouped at the genus level. In order to be identified as having been enriched in the PMA-treated samples, PTU had to meet the following two criteria: (a) the PTU must be deemed present based on standard PhyloChip analysis in PMA-treated samples and (b) the corresponding PTU must be present in greater transformed hybridization intensity in the PMA-treated sample fraction than in the non-PMA-treated sample fraction.
One representative 16S rRNA gene sequence within each PTU was selected, a multiple sequence alignment was generated with the SINA aligner (Pruesse et al., 2012), and a neighbor-joining phylogenetic tree was compiled at the family level with MEGA 4 (Tamura et al., 2007). This tree was rendered in a circular manner with the iTOL tree viewing program (Letunic and Bork, 2007).

Tag-encoded FLX amplicon pyrosequencing. Bacterial-biased primers 28F (5'-GAGTTTGATCNTGGCTCAG-3') and 519R (5'-GTNTTACNGCGGCKGCTG-3') were used to amplify ~500-bp fragments spanning the V1–V3 hypervariable regions of the bacterial 16S rRNA gene. This primer pair was tailored for bTEFAP by adding a fusion linker and a proprietary 12-bp barcode sequence at the 5' end of the forward primer, and a biotin and fusion linker sequence at the 5' end of the reverse primer (Dowd et al., 2008). A HotStarTaq Plus master mix kit (Qiagen, Valencia, CA, USA) was used to catalyze the PCR under the following thermal cycling conditions: initial denaturing at 95°C for 5 min, followed by 35 cycles of denaturing at 95°C for 30 s, annealing at 54°C for 40 s, and extension at 72°C for 1 min, finalized by a 10-min elongation at 72°C. Resulting PCR products were purified via Rapid Tip (Diffinity Genomics, Inc., West Henrietta, NY, USA) chemistry, and were then pooled accordingly. Small fragments (<100 bp) were removed with Agencourt Ampure Beads in accordance with manufacturer's instructions (Beckman Coulter, Brea, CA, USA).
In preparation for FLX-Titanium sequencing (Roche, Nutley, NJ, USA), resulting PCR amplicon fragment size and concentration were accurately measured with DNA 1000 chips using a Bioanalyzer 2100 automated electrophoresis station (Agilent, Santa Clara, CA, USA) and a TBS-380 Fluorometer (Turner Biosystems, Sunnyvale, CA, USA). The total volume of initial PCR product used for subsequent emulsion PCR was 2 µl for strong positives (>10 ng $µl^{-1}$), 5 µl for weak positives (5 to 10 ng $µl^{-1}$) and 20 µl for samples that failed to yield PCR products (<5 ng $µl^{-1}$). This normalization step helped to ensure minimal bias favoring downstream amplification from initially strong PCR products. Approximately 9.6 x 10^6 molecules of ~600-bp double-stranded DNA were combined with 9.6 x 10^6 DNA capture beads, and then subjected to emulsion PCR conditions. Following recovery and enrichment, bead-attached DNA molecules were denatured with NaOH and sequencing primers were annealed. A 454 pyrosequencing run was performed on a GS PicoTiterPlate using the Genome Sequencer FLX System in accordance with manufacturer's instructions (Roche). In all, 24 to 30 tagged samples were applied to each quarter region of the PicoTiterPlate. All bTEFAP procedures were performed at the Research and Testing Laboratory (Lubbock, TX, USA) in accordance with well-established protocols (Dowd et al., 2008).

bTEFAP-derived bacterial diversity and data analysis. Bacterial TEFAP sequences were processed and analyzed using the MOTHUR software package (Schloss et al., 2009), with the AmpliconNoise algorithm implemented. Raw pyrosequencing data for the PMA-untreated samples was derived from previously published work and was re-analyzed alongside PMA-treated samples (La Duc et al., 2012). Previously described standard operating procedures were followed for the analysis of sequence data in this study (Schloss et al., 2011). Sequences were removed from consideration if they (a) did not contain the primer sequence, (b) contained an uncorrectable barcode, (c) were <200 nt in length, (d) had homopolymers longer than 8 nt or (e) had a quality score of <25. Unique sequences were aligned using the Greengenes reference alignment (McDonald et al., 2012; Schloss et al., 2009) and trimmed such that all the sequences overlap in the same alignment space. Filtered sequences were assigned to samples according to their 12-nt barcode. After removing chimeras, sequences were classified in accordance with the new Greengenes training set and taxonomy (McDonald et al., 2012; Werner et al., 2012), and clustered into MOTU at the 0.03 level [that is, at 97% similarity; (Schloss et al., 2011)].

Negative controls. During this study, a negative control whereby a BiSKit was only pre-moistened with PBS, and a handling control, in which a BiSKit was pre-moistened with PBS and exposed to the sampling environment, were also prepared. These negative and handling controls were processed and analyzed both with and without PMA treatment before DNA extraction. All of the resulting qPCR indices for these controls were below detection limit. Bacterial bTEFAP sequencing was not performed on any of these controls because PCR amplification did not yield any quantifiable product. Although no detectable PCR amplification products were available, all negative and handling controls were run on a PhyloChip in order to detect possible contaminants. The resulting 447 PTU (of 8943 PTU in total) detected were omitted from the entire analysis. Only one PTU was detected in one handling control after PMA treatment. No PTU were detected in any of the negative or sampling controls after PMA treatment, supporting the conclusion that the detected PTU originated from extraneous DNA, and not from viable microbes associated with sampling materials or reagents.

Statistical analysis of community data. Multiple statistical analyses were performed to study the differences between the PMA-treated and non-PMA samples, all of which were based on (a) the abundance of sequences of each MOTU and (b) the transformed hybridization intensities of each PTU. This included principal coordinate analysis (PCoA), multi-response permutation procedures and Adonis testing (999 permutations), all of which were based on Bray–Curtis distance measures. Dendrogram clustering was based on Euclidean distance. Diversity indices (Shannon–Wiener) were calculated for MOTU only. All statistical analyses, including heatmaps, were performed using the R programming environment (Vegan, MASS and ape packages [(R-Development-Core-Team, 2005)].

Results

Quantitative PCR. Total bacterial burden, as assessed by bacteria-directed qPCR, is given in *Table III.1-1*. When PMA treatment was omitted before molecular processing, the total bacterial burden (viable+non-viable) of the mission-critical SAF cleanroom floor and GSE samples was ~10^5 to 10^6 rRNA gene copies m^{-2}. Samples collected from the inactive Bldg 144 facility floor and GSE yielded 1 to 2-logs higher rRNA gene copy numbers than SAF samples both with and without PMA treatment. Following treatment with PMA, a mere 7% (floors) and 4% (GSE) of the total bacterial population encountered about the mission- critical SAF was determined to be viable. However, the bacterial population present in PMA-treated Bldg 144 samples was 21% (floors) and 7% (GSE) viable. On comparative analysis, Bldg 144 floors housed a 188-fold greater viable bacterial burden than their mission-critical SAF counterparts. Similarly, Bldg 144 GSE samples contained 25-fold more viable bacteria than their SAF GSE equivalents.

bTEFAP analysis. A breakdown of the number of MOTU observed in the various samples examined over the course of this study is provided in *Table III.1-1*. Overall, Bldg 144 cleanroom floor samples housed 3.6-fold more bacterial MOTU than the mission-critical SAF cleanroom floor, whereas the GSE surfaces from each of these cleanrooms yielded a roughly equivalent number of MOTU. The effect of PMA treatment was significantly higher in both SAF (97% MOTU reduction) and Bldg 144 (75% MOTU reduction) floor samples, which indicated the presence of a large number of dead cells or extraneous DNA on these floors. A similar trend was observed in PMA-treated GSE samples from SAF (reduced from 425 to 17 MOTU) and Bldg 144 (reduced from 571 to 42 MOTU).

A breakdown of MOTU, at the level of bacterial phyla or class, observed in the various samples examined is detailed in *Table III.1-2*. The mission-critical SAF floor retained MOTU affiliated with physiologically recalcitrant bacteria (Actinobacteria, Acidobacteria and Firmicutes), whereas the Bldg 144 cleanroom floor harbored predominantly proteobacterial MOTU. It was particularly apparent that a few acidobacterial types (4 MOTU) were present in great abundance (112 sequences) in the SAF floor samples. However, the GSE samples exhibited no such correlation between MOTU numbers and sequence occurrence. A closer examination of the pyrosequence reads resulting from the SAF cleanroom floor and GSE samples indicated a predominance of members of the genera: *Acidobacteria*, *Actinobacteria*, *Arsenicicoccus*, *Arthrobacter*, *Corynebacterium*, *Kineococcus*, *Propionibacterium*, *Nocardioides*, *Streptomyces*, *Bacillus*, *Clostridium*, *Lactobacillus*, *Deinococcus* and *Staphylococcus*. Many of the sequences arising from the physiologically recalcitrant bacteria observed in cleanroom floor samples sans PMA treatment were absent or in very low number in the PMA-treated fractions of the very same sample.

Table III.1-2 Bacterial taxa present in various cleanroom samples as determined by pyrosequencing method

Taxa	Number of MOTUs from:								Number of pyrosequences from:							
	SAF cleanroom floor (GI-36-4)		SAF-GSE (GI-36-3)		Bldg 144 cleanroom floor (GI-42-1)		Bldg 144 GSE (GI-42-2)		SAF cleanroom floor (GI-36-4)		SAF-GSE (GI-36-3)		Bldg 144 cleanroom floor (GI-42-1)		Bldg 144 GSE (GI-42-2)	
	As is	PMA	As is	PMA	As is	PMA	As is	PMA	As is	PMA	As is	PMA	As is	PMA	As is	PMA
Actinobacteria	25	1	161		48	8	155	3	350	2	1685		84	106	578	15
Armatimonadetes					3		1	1					6		1	1
Bacteroidetes	8		36		39	5	66	3	108		208		2483	60	403	8
Verrucomicrobia					2								5			
Chloroflexi	1		6				9		4		44				24	
Deinococcus-Thermus			7		5		13				26		7		40	
Acidobacteria	4		1		1		3	2	112		55		1		30	67
Firmicutes	11		24	1	11	5	24	2	157		250	2	16	30	109	10
Fusobacteria	1		1				2		1		1				2	
Gemmatimonadetes					2								2			
Nitrospirae					1		1						1		1	
Planctomycetes			1		2		2				1		2		4	
Proteobacteria																
Alpha	41	2	100	8	186	41	168	16	606	11	1154	52	7335	1058	2478	121
Beta	9		22		39	14	46	2	100		464		879	296	525	42
Delta	1		1		3		2		36		5		6			5
Gamma	19	2	26	2	67	23	35	5	357	3	231	6	4784	479	1094	239
Unidentified					9	2	4							155	2	31
Spirochaetes	1								1							
Unidentified division																
SC4				3			4				23				6	
TM7			1		1						2		1			
WPS-2					2								2			
Unclassified bacteria	4		32	5	26	10	36	8	8		112	27	145	165	103	25
Total	**122**	**4**	**425**	**17**	**447**	**108**	**571**	**42**	**1783**	**14**	**4318**	**89**	**15914**	**2196**	**5434**	**528**

Abbreviations: Bldg, building; GSE, ground support equipment; MOTU, molecular operational taxonomic unit; PMA, propidium monoazide; SAF, spacecraft assembly facility.

Although equivalent surface areas were sampled from the floors and GSE of the two cleanrooms studied (9 m^2 each), the Bldg 144 samples gave rise to many more pyrosequence reads than the SAF floor samples. Regardless of sample type, PMA-treated sample fractions consistently yielded considerably fewer pyrosequences than their untreated counterparts. Anywhere from 14 to 2196 high-quality pyrosequences (>250 bp) were obtained from samples that had been pre-treated with PMA, whereas 1783 to 15914 high-quality pyrosequences were recovered from untreated samples. Even when PMA treatment was omitted, the mission-critical SAF cleanroom floor sample (GI-36-4) yielded far fewer pyrosequences (1783 reads) than the Bldg 144 cleanroom floor sample (GI-42-1; 15914 reads). The relative abundance of pyrosequences retrieved from the PMA-untreated cleanroom floor samples is plotted as a Venn diagram in *Figure III.1-1a*. Approximately 65% of the pyrosequences retrieved from these two distinct facility floor samples were detected in both cleanrooms, although this shared fraction represented only 8% of the total observed MOTU (46 out of 569; *Figure III.1-1a*). The relative abundance of pyrosequences retrieved from the untreated SAF samples is plotted as a Venn diagram in *Figure III.1-1b*. Between the SAF floor and GSE samples, ~38% of the total number of detected pyrosequence reads were shared, although this constituted a mere 7.8% of the total MOTU (43 out of 547).

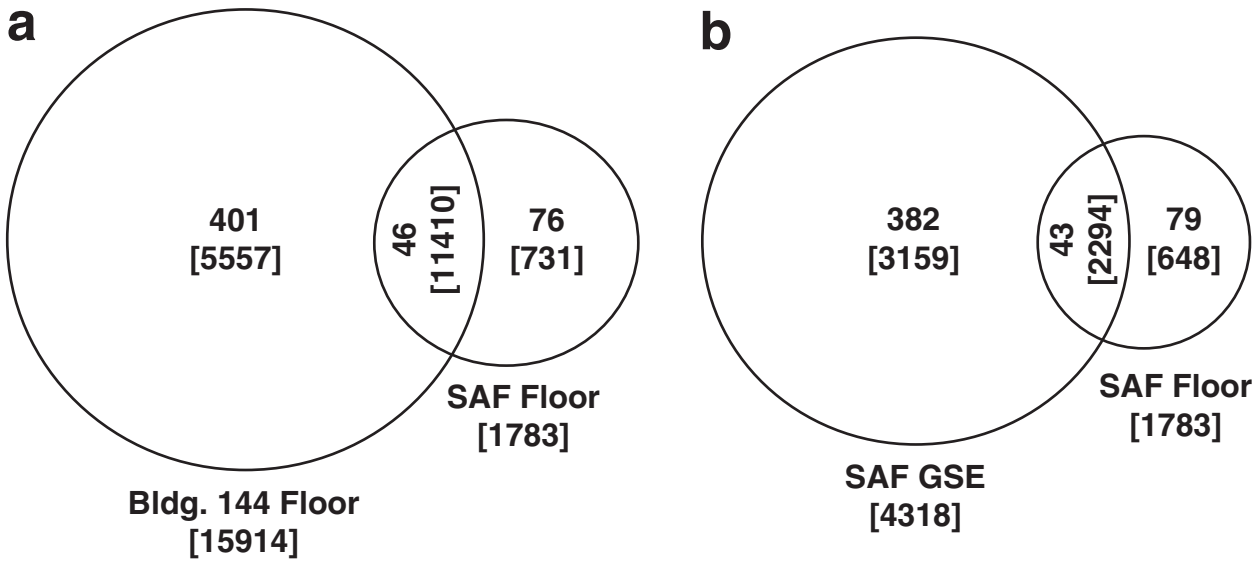

Figure III.1-1 *Venn diagram showing the MOTU detected in various samples of the cleanrooms. Comparison* ***a****) between SAF and Bldg 144 cleanroom floor samples and* ***b****) between floor and GSE samples of SAF cleanroom. Parentheses denote total number of pyrosequences generated and the numerals without parentheses are total number of MOTU present in that sample.*

PhyloChip analysis. A drastic decrease in the total number of bacterial genera was observed in all samples on pre-treatment with PMA (*Table III.1-1*). After having been treated with PMA, the SAF cleanroom floor and GSE samples exhibited very simple bacterial community structure, housing very few genera (9 and 3, respectively) of Firmicute and Proteobacteria lineage. Without PMA treatment, these very same floor and GSE samples yielded many more genera, which represented a diverse assemblage of taxa (*Table III.1-3*) dominated by Actinobacteria, Firmicutes and Proteobacteria. The Bldg 144 cleanroom floor and GSE samples also presented many more genera without PMA pre-treatment (199 and 236, respectively) than the very same samples having been treated with PMA (106 and 75).

A heatmap was generated based on the transformed hybridization intensities of detected PTU (Supplementary Figure 1). Such analysis clearly illustrates a marked decrease in the hybridization scores of all samples treated with PMA, as compared with the very same samples not treated with PMA. For instance, the SAF GSE sample (GI-36-3) exhibited very high transformed hybridization intensities for detected PTU *sans* PMA treatment. However, when treated with PMA, this very same sample showed a dramatic reduction in all resulting hybridization intensity scores, which suggested that the majority of the detected 16S rRNA genes arose from deceased members of the community. In order to identify the viable members that exhibited increased hybridization intensities after PMA treatment, ratios of non-PMA-treated and PMA-treated hybridization scores were calculated for each PTU, and are presented as a heatmap (*Figure III.1-2*). An observed increase in the transformed hybridization intensity of a given PTU following PMA treatment would likely stem from PCR bias in non-PMA-treated samples, where high levels of DNA originating from non-viable cells are co-amplified. This amplification of template DNA arising from nonviable cell types masks the presence of smaller levels of DNA template arising from viable cells. Hence, transformed hybridization intensities are elevated on removing the template DNA arising from non-viable cells from the equation (that is, PMA-treated sample fraction). The SAF cleanroom floor and GSE samples gave rise to fewer PTU with higher transformed hybridization ratios. At the same time, PMA-treated Bldg 144 cleanroom samples were richer in PTU having increased transformed hybridization intensity ratios, which was indicative of a relatively greater viable population. These PTU belonged to the Firmicutes, Cyanobacteria, Actinobacteria and Proteobacteria. A phylogenetic tree of PhyloChip PTU grouped at the family level is provided as Supplementary Figure 2, which shows the presence of the detected families in each sample with and without PMA treatment.

Table III.1-3 Bacterial taxa of various cleanroom samples as determined by PhyloChip analysis

Taxa	Number of PhyloChip-detected genera from:							
	SAF cleanroom floor (GI-36-4)		SAF-GSE (GI-36-3)		Bldg 144 cleanroom floor (GI-42-1)		Bldg 144 GSE (GI-42-2)	
	As is	PMA	As is	PMA	As is	PMA	As is	PMA
Actinobacteria	19		131		16	11	73	18
Armatimonadetes							1	
Bacteroidetes	1		16		4	2	3	2
Verrucomicrobia					1			
Chloroflexi			2					
Deinococcus-Thermus	1		2				2	1
Acidobacteria	2		4				1	
Firmicutes	12	4	67	2	6	14	28	13
Fusobacteria			1				2	
Gemmatimonadetes			1					1
Nitrospirae	1							
Planctomycetes			4			2	2	1
Proteobacteria								
Alpha	18	3	59	1	55	24	39	9
Beta	19		61		53	28	41	18
Delta			2		2			
Gamma	18	2	54		56	20	38	4
Epsilon			1					
Fibrobacteres					1			
Cyanobacteria	2		4		4	4	6	5
Unidentified division								
BRC1			1					
OP11					1			
OP3								1
TM7		1	1					1
WS3							1	1
Total number of genera:	94	9	411	3	199	106	236	75

Abbreviations: Bldg, building; GSE, ground support equipment; PMA, propidium monoazide; SAF, spacecraft assembly facility.

Statistical analysis of microbial community profiles. The various samples examined by bTEFAP analysis consistently showed lower Shannon–Wiener diversity indices when treated with PMA compared with their corresponding non-treated samples (Supplementary Table 1). PCoA was performed to study the environmental clustering and relatedness of community profiles derived from bTEFAP and PhyloChip analyses (*Figure III.1-3*). All of the samples analyzed without PMA treatment clustered together, indicative of their relatively similar community structure, compared with their coinciding PMA-treated sample fractions. The clustering observed in the PCoA plot was congruent with the dendrogram clustering presented in Supplementary Figure 1, which implied a close association among all non-PMA samples. Multi-response permutation procedure analysis was performed to assess the difference in community structure between PMA-treated sample fractions and non-PMA-treated sample fractions, derived by both bTEFAP and PhyloChip analysis. The null hypothesis (no difference in PMA- treated and non-PMA-treated samples) was rejected based on the significance of the delta for both bTEFAP and PhyloChip analyses (0.028 and 0.037, respectively). The chance-corrected within-group agreements were fairly low for bTEFAP sequence data (A = 0.0411) but high (A = 0.2605) for PhyloChip data, which reflect the observed grouping in PCoA. In addition, Adonis testing clearly showed a significant change in the detected community profiles after PMA treatment (P-value bTEFAP: 0.04, P-value PhyloChip: 0.02). Based on the multiple statistical approaches used, such as Adonis, multi-response permutation procedure, a Euclidean distance-based dendrogram (Supplementary Figure 1), and ordination (PCoA)

analyses (*Figure III.1-3*), it was clear that PMA-treated samples and non-PMA samples were significantly dissimilar with respect to diversity in bacterial community profiles.

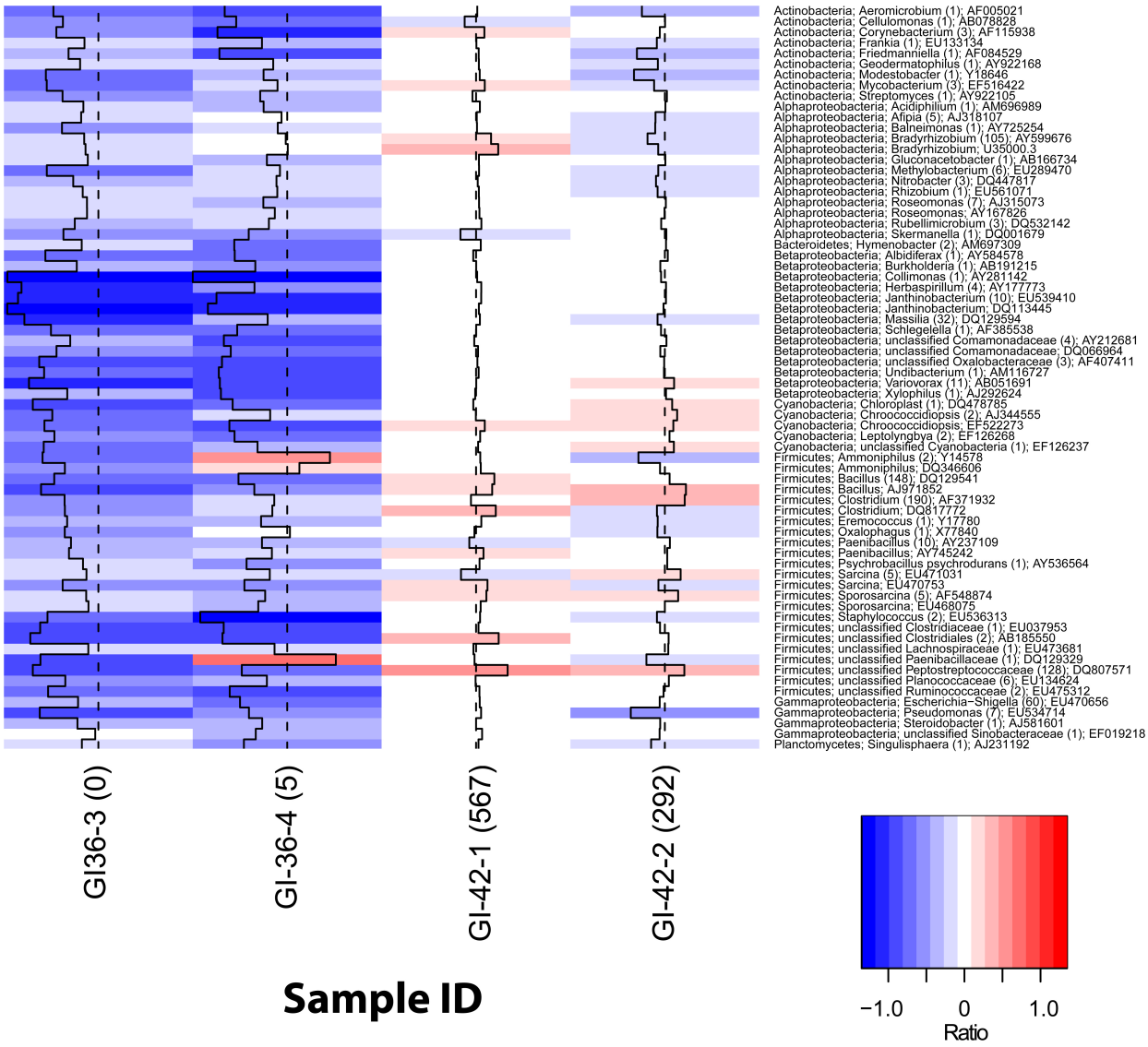

Figure III.1-2 *Heatmap of PTU that increased in transformed hybridization intensities in PMA-treated samples compared with non-PMA-treated samples and were called present in the PMA-treated sample. An increase in transformed hybridization intensities in PMA-treated sample is reflected as a positive ratio. In total, 801 PTU were identified that fulfilled this requirement, which were grouped into 70 genera. Displayed are representatives of all genera with the most drastic changes for each sample pair. Numbers in parentheses are the total number of PTU.*

Discussion

Over the past 25 years, sequence analysis of PCR-amplified rRNA genes has become the 'gold standard' for assessing species richness in mixed microbial communities, and as a result, total resolvable microbial diversity is now estimated to be threefold greater than that based solely on cultivation (Pace, 1997). As sequences from organisms in greatest abundance are far more likely to be represented in clone libraries than those from singleton and low-abundance taxa, it is advantageous to use high-throughput technologies, such as bTEFAP and PhyloChip, which have been shown to render a far superior representation of community structure (Brodie et al., 2006; DeSantis et al., 2007; La Duc et al., 2009; Sogin et al., 2006).

The application of PMA to assess the differential viability of microbial cells is increasing in popularity within the scientific community (Nocker et al., 2009; van Frankenhuyzen et al., 2011). When used to pre-treat samples, the PMA concentration applied in this study effectively precluded ~90% of the total DNA template molecules from downstream manipulation, resulting in the generation of only 2827 pyrosequences, whereas 27449 pyrosequences were obtained from the very same

samples without PMA pre-treatment. With respect to diversity, all of the pyrosequences generated after PMA treatment represented a mere 171 MOTU, which corresponded to ~12% of the total number of MOTU resulting from the very same samples without PMA treatment (*Table III.1-2*). Furthermore, 2- to 3-logs fewer pyrosequences arose from PMA-treated than untreated mission-critical cleanroom samples (0.8% for floor; 2.1% for GSE).

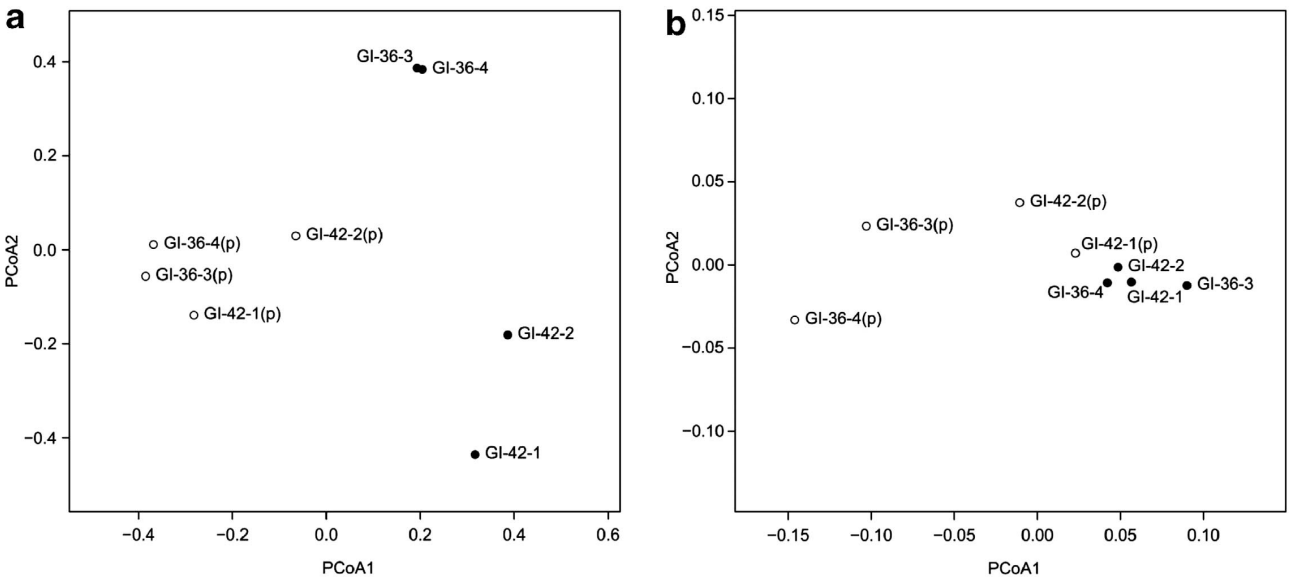

Figure III.1-3 *PCoA based on:* ***a****) number of pyrosequences per MOTU (PCoA1, percentage of explained variance: 22%; PCoA2, percentage of explained variance: 17%) and* ***b****) PhyloChip-derived transformed hybridization scores of each PTU (PCoA1, percentage of explained variance: 87%; PCoA2, percentage of explained variance: 6%). Open and closed dots represent PMA-treated and PMA-non-treated samples, respectively.*

The total number of MOTU observed in the operational mission-critical SAF cleanroom floor sample (122 MOTU) was considerably less than that associated with the quiescent Bldg 144 cleanroom floor (447 MOTU). Bioinformatic analyses of pyrosequence data demonstrated that both of these contaminant microbial populations comprised only Actinobacteria, Firmicutes and Proteobacteria. The vast majority of these bacteria, if not all, were present in the SAF mission-critical cleanroom floor samples in a non-viable state. However, about 25% of the detected MOTU (108 out of 447) were observed to be viable in the Bldg 144 facility, suggesting that certain taxa are able to withstand the desiccated and nutrient-deprived conditions of these cleanroom floors (sample #GI-42-1). Actinobacterial genera such as *Kineococcus*, *Kocuria*, *Modestobacter* and *Propionibacterium* were present in high abundance in the JPL-SAF floor samples. In contrast, the floors of the Bldg 144 facility housed predominantly Proteobacterial genera, as *Brevendimonas* and *Acinetobacter* pyrosequences were generated in great numbers. Previously, members of these and other closely related genera have been isolated from spacecraft assembly environments (Ghosh et al., 2010; La Duc et al., 2009; Osman et al., 2008; Vaishampayan et al., 2012). The molecular biological detection and isolation of these robust microbial lineages from spacecraft-associated environments is of particular consequence to National Aeronautics and Space Administration (NASA) planetary protection practices, not to mention routine validation of these cleanroom facilities. Unlike cleanroom floors, which were treated with Kleenol 30 detergent, the GSE materials housed in either cleanroom were subjected only to alcohol wiping, and yet gave rise to similar MOTU and pyrosequence occurrence, with GSE materials kept at the Bldg 144 facility marginally enriched (~1.3-fold increase in pyrosequences and MOTU). From these results, it is apparent that GSE need to be subjected to more rigorous cleaning regimens, as GSE-associated richness was two to four times greater than that of the floor surfaces.

The most frequently encountered bacterial MOTU from the stringently maintained and frequently cleaned floors of the SAF cleanroom were members of the genera: *Bacillus*, *Clostridium* and *Nocardia*. These bacteria are known to survive oligotrophic conditions for extended periods of time, tolerate alkaline and oxidative stress, and avoid death by ultraviolet radiation by morphing into highly resilient, dormant endospores (Ghosh et al., 2010; La Duc et al., 2007; Probst et al., 2010b). Although routine cleaning and maintenance regimens limit the number of bacterial taxa capable of persisting in cleanrooms, hardy spore-forming microorganisms like *Bacillus* spp., *Clostridium* spp. and *Nocardia* spp. capitalize on their selective advantage and superior fitness and survive — much to the chagrin of those challenged with bioreduction and

sterilization of these environments. The presence of these organisms is detected as a result of their (a) viability, or (b) inability to be penetrated by PMA molecules while in a non-viable state. Owing to subtle nuances inherent in the PMA-chemistry-coupled techniques described herein, endospores and non-viable cells having intact cell walls and/or outer membranes will escape PMA treatment (Nocker et al., 2009; Probst et al., 2012; Rawsthorne et al., 2009), and thus be observed as false-positive viable entities. Similarly, sampling and sample processing steps (for example, the composition of the solution to collect and store microorganisms, method for cell concentration and PMA treatment) might affect the viability of the cells but in a recent study such adverse effect of sample handling procedures was not noticed for sea and canal water samples (Kort et al., 2010; Nocker et al., 2010).

As was observed via bTEFAP procedures, PhyloChip-based analyses discerned noticeable differences in the bacterial diversity profiles resulting from PMA- and non-PMA-treated samples. One such observation was the reduction in diversity of proteobacteria detected in PMA-treated samples and prevalence of proteobacteria in samples not treated with PMA. Another consistent result was the observed prevalence of Firmicutes in the PMA-treated samples. This suggests that Gram-positive bacteria are more tolerable of the inhospitable conditions of the cleanroom environment that their Gram-negative kin. Numerous studies have reported varying accounts of the microbial diversity typical of spacecraft-associated cleanrooms, and what the presence of such communities might portend for life detection endeavors in extraterrestrial settings (La Duc et al., 2007; La Duc et al., 2004; Moissl et al., 2007; Probst et al., 2010b; Vaishampayan et al., 2010; Venkateswaran et al., 2001). Indeed, the introduction of contaminant microbes to extraterrestrial environments could have profound repercussions on (a) the scientific integrity of in situ and sample-return based life detection experiments, and (b) the uncompromised nature of such settings. The worst-case scenario for life-detection experimentation would be the inadvertent transfer of viable contaminant microbiota to an otherwise pristine location of interest. The results of this study are encouraging, as they suggest that hitherto, the breadth of diversity enveloped within the viable fraction(s) of typical spacecraft-associated microbial communities have been overestimated. At the same time, these findings enabled the first ever statistically significant differentiation between the total and viable-only portion of microbial communities in cleanroom environments. Significant differences were shown between these two populations using two independent profiling methods, namely bTEFAP and PhyloChip G3. Consequently, these methodologies are an attractive means of discerning viable phylotypes in low-biomass environments. Such a capability is of crucial importance and benefit to numerous industries (for example, healthcare, pharma, semiconductor fabrication), not least of all the NASA, whose planetary protection program is tasked with ensuring spacecraft-borne microorganisms do not result in harmful contamination of extraterrestrial environments.

Acknowledgements

Part of the research described in this study was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. AJP's contribution was supported by the German National Academic Foundation (Studienstiftung des deutschen Volkes). We are grateful to T DeSantis, L Tom, for PhyloChip analyses, S Westcott and P Schloss, for pyrosequence analysis, J Andy Spry and K Buxbaum for valuable advice and guidance. We thank M Cooper and C Stam for assistance with sample collection and processing, and acknowledge Y Sun at Research and Technology Laboratory for all next-generation sequencing and assistance with TEFAP analyses.

Supplementary Information

Supplementary information can be found online http://www.nature.com/ismej/journal/v7/n2/extref/ismej2012114x1.pdf or on the supporting DVD.

2. Archaea on human skin

Alexander J. Probst[1], Anna K. Auerbach[1] and Christine Moissl-Eichinger[1]

[1]Department for Microbiology and Archaea Center, University of Regensburg, Regensburg, Germany

Correspondence: C Moissl-Eichinger, University of Regensburg, Department for Microbiology and Archaea Center, Universitaetsstrasse 31, 93053 Regensburg, Germany. E-mail: christine.moissl-eichinger@ur.de

Publication information:
PLOS ONE, Published: June 12, 2013, DOI: 10.1371/journal.pone.0065388
Received: February 11, 2013; Accepted: April 29, 2013
Link: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0065388

Abstract

The recent era of exploring the human microbiome has provided valuable information on microbial inhabitants, beneficials and pathogens. Screening efforts based on DNA sequencing identified thousands of bacterial lineages associated with human skin but provided only incomplete and crude information on Archaea. Here, we report for the first time the quantification and visualization of Archaea from human skin. Based on 16S rRNA gene copies Archaea comprised up to 4.2% of the prokaryotic skin microbiome. Most of the gene signatures analyzed belonged to the Thaumarchaeota, a group of Archaea we also found in hospitals and clean room facilities. The metabolic potential for ammonia oxidation of the skin-associated Archaea was supported by the successful detection of thaumarchaeal amoA genes in human skin samples. However, the activity and possible interaction with human epithelial cells of these associated Archaea remains an open question. Nevertheless, in this study we provide evidence that Archaea are part of the human skin microbiome and discuss their potential for ammonia turnover on human skin.

Introduction

Archaea have long been thought of as an ancient form of microorganisms, restricted to extreme environments. However, the picture of Archaea changed within the last decade, when these organisms were found in high abundance in cold and moderate environments all around the world (Karner et al., 2001).

Archaea might also play an important role in the human body, as methanogenic archaea can contribute up to 12% of total anaerobes in the human gut (Conway de Macario and Macario, 2009). In the oral cavity, methanogens have been associated with some periodontal diseases (Conway de Macario and Macario, 2009), although pathogenesis of an archaeon is yet to be confirmed.

The Human Microbiome Project, founded to decipher the entire set of microorganisms associated with the human body, continues to provide valuable information on how microbial diversity correlates with the health status of humans (Human Microbiome Project Consortium, 2012). So far, the bacterial dynamics of the largest human organ, the skin, have been studied in detail (Grice and Segre, 2011), while only two studies report the detection of Archaea on human skin (Caporaso et al., 2011; Hulcr et al., 2012). Hulcr and co-workers studied 60 navels and found three different phylotypes of Archaea appearing marginally in a large subset of bacterial sequences. The archaeal phylotypes identified belonged to the Euryarchaeota and were retrieved from 6 samples only. The authors detected halophiles (*Halobacteriaceae*) in two samples and methanogens (*Methanobrevibacter*) in five different samples. Two of the three archaeal phylotypes were retrieved from a human subject that had not showered for years implying that Archaea are only a minor fraction of the navel and skin microbiome. Moreover, the detected archaeal taxa have previously been found associated with human oral cavity and the human gastrointestinal tract and are thus likely to be oral or fecal contaminants (Lepp et al., 2004; Oxley et al., 2010). Although Hulcr and co-workers claimed to be the first to report on Archaea in the human skin microbiome, Caporaso et al. (Caporaso et al., 2011) had already reported signatures of Archaea, in particular Thaumarchaeota, in samples taken from palms of two individuals. However, they identified these microorganisms as a minor, transient part of the human-associated microbiome and assumed an insignificant role.

Nevertheless, both studies detected Archaea via a co-amplification of their 16S rRNA genes along with Bacteria. This, and the fact that the primer pairs used did not perfectly match (thaum-)archaeal 16S rRNA genes, does not allow a conclusion about the role, abundance or diversity of Archaea on human skin. For instance, primer pair F515/R806 [used in (Caporaso et al., 2011)] hits only 50% of all Archaea and in particular, 8% of all Thaumarchaeota without mismatch. No perfect match was revealed within the soil crenarchaeotic group (I.1b), which includes *Candidatus* Nitrososphaera [SILVA TestPrime (Klindworth et al., 2013)].

The explorations of microbiomes in man-made environments such as clean room facilities, which are strongly influenced by the human (skin) microbiome, have revealed Archaea to be continuously present (Moissl et al., 2008; Moissl-Eichinger, 2011). Most of these Archaea were Thaumarchaeota, a recently proposed phylum including designated ammonia oxidizers (Pester et al., 2011). Due to the recent re-classification of the thaumarchaeal phylum and therefore an assignment of certain crenarchaeal groups to the thaumarchaeal clade, a recent study by LaDuc and co-workers (La Duc et al., 2012) wrongly claimed first evidence of Thaumarchaota in clean room environments, although this group had been identified earlier (Moissl et al., 2008; Moissl-Eichinger, 2011). So far, these clean room Archaea belonged mainly to the I.1b thaumarchaeal clade, whose representatives are commonly found in the soil microbiome, where they likely contribute to the global nitrogen cycle (Leininger et al., 2006). A probable association with humans was discussed by Moissl-Eichinger (Moissl-Eichinger, 2011) but the question if these Archaea originate from human skin remained unanswered.

In the present study we tackle the question, whether the human skin can be carrier or even habitat for Archaea. We show, that Archaea, and in particular Thaumarchaeota represent a detectable part of the human skin microbiome and their signatures are closely related to those found in hospitals and clean room facilities. Moreover, we provide insight into cell morphology and functional genes for Archaea on human skin.

Materials and Methods

Human skin samples were taken and handled with approval by and in accordance with the Ethic Commission at the University of Regensburg. The Ethics Commission stated that no ethical concerns are raised by the methods applied and approved the following procedures. Verbal informed consent was obtained from all study participants, which was in

agreement with the Ethic Commission's statement. Each participant handed over the sample right after self-sampling and verbal consent was documented manually along with receiving the samples. Samples were treated anonymously. Human material was not subject of this study. Microbial samples or data derived cannot be attributed to a certain person.
Samples from the entire front torso were taken using DNA-free wipes by the volunteers themselves. The human subjects were instructed to thoroughly wipe their torso (holding the DNA-free wipe with a sterile glove) before taking a regular shower. The volunteers did not apply cosmetics before sampling. The wipes were immediately stored on ice or frozen before processing. An overview of all human skin wipe-samples is given in Supplementary Table 1.

Sampling of indoor environments was performed with either Biological Sampling Kits (BiSKit, QuickSilver Analytics, Abingdon, MD, USA, according to manufacturer's instructions) or with a pre-moistened, DNA-free wipe attached to a DNA-free sampling tool made of steel (Supplementary Table 1). One clean room complex (EADS, Friedrichshafen, Germany) with an ISO 5 and an ISO 8 clean room was sampled. The clean rooms were under certified, fully operating conditions. Additional sampling locations were two intensive care units, one in Regensburg (Germany) and one in Graz (Austria), both maintained fully operating (Supplementary Table 1).

Sample extraction from wipes (including vortexing and sonication) was performed in 40 ml of PCR grade water (for molecular analyses) or phosphate buffered saline (PBS-buffer, for fluorescence *in situ* hybridization) as described elsewhere (Moissl-Eichinger, 2011). Liquids were concentrated to 200–500 µl using Amicon 50 filter tubes (Millipore, Billerica, MA, USA) before processing.

The recovery efficiency of sampling devices is strongly dependent on the sampling tool and the porosity of the surface (Probst et al., 2010a; Probst et al., 2011). Wipe sampling of non-porous surfaces has been proven to have generally low recovery efficiencies of 8–20% (Kwan et al., 2011; Probst et al., 2011). The fact that human skin is a very porous surface and the sampling of skin was performed by non-specialists (self-sampling) combined with the low recovery efficiency of sampling in general, allows the conclusion that only a small part of the skin microbiome was recovered. However, the ratio of Bacteria and Archaea retrieved is expected to be independent from sampling efficiency.

Propidium monoazide treatment on selected samples from the intensive care unit in Regensburg was performed as described elsewhere (Nocker et al., 2007) prior to DNA extraction. Propidium monoazide (PMA) is a chemical that intercalates to accessible DNA molecules in a given solution and forms a covalent bond after photoactivation of the azide. After PMA-binding the DNA is masked and no longer available for PCR amplification. Cells with intact cell membranes are not penetrated by PMA, their DNA remains unlabeled for PCR amplification and can therefore be detected. This assay allows distinction between the membrane-compromised and the viable microbial community.

DNA extraction was performed by a combination of bead-beating and the XS-buffer method described in Moissl-Eichinger (Moissl-Eichinger, 2011), a protocol adapted for low-biomass environments. Bead-beating was included before the application of the XS-buffer to ensure that also hardy microbes were lysed (bead-beating tubes were taken from the MO BIO Power Biofilm™ DNA Isolation Kit, MO BIO, Carlsbad, CA, USA). After bead-beating, beads were washed with 400 µl pure PCR-grade water to decrease sample loss.

Quantitative PCR of bacterial and archaeal 16S rRNA gene sequences was performed in triplicate as described elsewhere with primer pairs 338 bf/517 ur and 344 af/517 ur, respectively (Moissl-Eichinger, 2011; Probst et al., 2013b) (final primer concentration: 300 nM). 16S rRNA genes from genomic DNA of the archaeal and bacterial reference strains *Methanosarcina barkeri* and *Bacillus safensis* were amplified with the primer sets 8 af/1406 ur and 9 bf/1406 ur (Burggraf et al., 1992; Lane, 1991). Quantification of standards was performed with the Qubit Quantitation Platform 2.0 (High Sensitivity Kit, Invitrogen, Carlsbad, CA, USA). Forty cycles of qPCR (Quantitect SYBR Green PCR Mix, Qiagen, Hilden, Germany) were run (RotorGene 6000 Real-Time PCR system, Corbett Life Science, Concorde, NWS, Australia) with an initial denaturation at 95°C for 15 min and a cycling protocol as follows: denaturation at 94°C for 15 sec, annealing at 60°C for 30 sec and elongation at 72°C for 30 sec. Melting curve was performed at 72–95°C. The qPCR efficiencies ranged from 0.87 to 0.92 and R^2 values of standard curves were in the range of 0.98 to 1.00. Detection limits were defined as 105 copies/µl for

archaeal qPCRs (threshold for non-template controls) and 334 copies/µl for bacterial qPCRs. These thresholds were generated as follows: archaeal negative control qPCR revealed minor signals from primer-dimerization, which was confirmed via gel electrophoresis. To exclude such background, qPCR negative controls were averaged and used as a threshold. No archaeal signal was obtained from non-template and extraction controls. Since the qPCR reagents were not free of bacterial DNA as reported elsewhere (Mühl et al., 2010), the detection limit for bacterial qPCR was increased accordingly (334 copies/µl, averaged bacterial qPCR no template controls).

Amplification of archaeal 16S rRNA genes from samples was performed either directly [primer pair 344 af/915 ar; (Raskin et al., 1994; Stahl and Amann, 1991)] or using nested PCR (primer pairs 8 af/UA1406 R (Baker et al., 2003; Huber et al., 2002) and 340 af/915 ar (Gantner et al., 2011; Raskin et al., 1994); 2×35 cycles). Primer coverages were evaluated using TestProbe, the SILVA probe match and evaluation tool against the entire SILVA SSU Ref database 108 (Pruesse et al., 2012), which revealed those primer combinations with the highest coverage over other combinations studied. The tested combinations included the following primers. Forward: 8af (Burggraf et al., 1992), 27FLP (Burggraf et al., 1991), 21af (DeLong, 1992), 340af (Gantner et al., 2011), 344af (Casamayor et al., 2002), A751f (Baker et al., 2003). Reverse: 909r (Brunk and Eis, 1998), 915r (Stahl and Amann, 1991), 1000R (Gantner et al., 2011), 1100R (Embley et al., 1992), 1119ar (Burggraf et al., 1997) , UA1406r (Baker et al., 2003), 1391r (Brunk and Eis, 1998), 1406ur (Lane, 1991), 1492ur (Lane, 1991).

Cloning of 16S rRNA gene PCR products, sequencing and analysis was performed as follows: PCR products were cloned in TOPO10 competent cells (Invitrogen, TOPO® TA Cloning® Kit, pCR®2.1 vector), inserts of positive clones were screened using two (*Hinf*I; *Bsu*RI) and four (*Alu*I; *Hha*I; *Hinf*I; *Rsa*) restriction enzymes after PCR amplification with the above mentioned primer pairs. Clones carrying unique inserts (52 unique patterns) were Sanger-sequenced (100 inserts sequenced), trimmed, quality checked (chimera slayer), aligned (Pruesse et al., 2012) and grouped at 1% difference level with the mothur software package (Schloss et al., 2009) and are referred to as operational taxonomic units (OTU) in the manuscript. Representative sequences from each OTU were then classified using the Bayesian classifier (Schloss et al., 2009) against a manually curated GreenGenes database that contains representatives of 98% identical clusters and an updated taxonomy (McDonald et al., 2012) (available at http://www.secondgenome.com/go/2011-greengenes-taxonomy/). The derived archaeal 16S rRNA gene sequences have been submitted to GenBank (Acc. No. JX865653-JX865767).

Phylogenetic tree of representative OTU sequences (SINA aligned; 28) was computed using ARB (Ludwig et al., 2004) and the SILVA database release SSU111 (Pruesse et al., 2007) by applying a maximum-likelihood algorithm. Sequences were trimmed to the same alignment length before tree calculation (*E. coli* position 346–943). The tree topology confirmed the classification gained from the Bayesian method (see above).

Detection of amoA genes was performed as described in Tourna et al. (Tourna et al., 2008) using the primers amoA104F-1d (GCA GGA GAC TAC ATM TTC TA) and amoA616R (GCC ATC CAT CTG TAT GTC CA). Amplicons were cloned and Sanger sequenced using M13 primers. A combined sample of all human subjects investigated in this study was used as a template for PCR (5 µl per 20 µl PCR reaction). Thirty-two sequences were obtained, 28 were of high quality (good chromatogram quality, >800 bps) and 21 of them were identified to be *amoA* genes via BLAST (Altschul et al., 1990) against the NCBI database (deposited under accession numbers KC582378-KC582398). Seven sequences (of approximately 510 bps) showed high similarity to sequences from *Staphylococcus epidermidis* (lipase precursor, *gehD*) and were not included in the analysis. *AmoA* genes were clustered at 97% similarity using mothur (Schloss et al., 2009) and a maximum likelihood tree was computed in ARB with the aid of the *amoA* database from Pester et al. (Ludwig et al., 2004; Pester et al., 2011).

Molecular analysis controls for each single step (sampling, DNA extraction, PCR setup) and at least one blank control was carried along, which underwent all detection procedures including regular archaeal PCR, nested-PCR, qPCR, and bacterial qPCR. All Archaea-directed controls were negative, no visible band occurred in any PCR amplification applied. BiSKit blanks exhibited a certain amount of detectable bacterial 16S rRNA genes (samples from intensive care unit Graz and clean room Friedrichshafen; archaeal control was negative). Therefore all human samples and those from the intensive care unit in

Regensburg were taken with DNA-free wipes (dry-heat treated for 24 hrs, 170°C). All wipe extractions blanks were negative in every archaeal and bacterial (q)PCR.

Fluorescence in situ hybridization (FISH) was performed on one human skin sample. The sample was fixed in paraformaldehyde (3% (w/v), final concentration). FISH was conducted as described earlier (Moissl et al., 2002), using probes ARC915 (directed against Archaea, rhodamine green labeled) and EUB338/I (directed against Bacteria, CY3 labeled), 20% (v/v) formamide and 0.01% (w/v) SDS. DAPI was used as counterstain. A NONEUB-Probe was used as a nonsense negative control. Microscopy was performed using an Olympus BX53 microscope (Olympus, Hamburg, Germany; camera: Olympus XM10, software: CelSens Standard 1.5). Fluorescence microscopy was performed using the following filters. U-FUN (excitation 360–370 nm, emission 420 nm IF) for DAPI, U-MINB3 (excitation 470–495 nm, 510 nm IF) for rhodamine green, and U-FRFP (excitation 535–555 nm HQ, emission 570–625 HQ) for CY3. FISH samples in general exhibited a high amount of particles (also from the wipe, see extraction procedure). These particles were either fluorescence active with CY3 filter or with all three filters, clearly distinguishable from microorganisms and not included in the analysis. These fibers made quantitative FISH not feasible.

Results and Discussion

In order to investigate the abundance and diversity of Archaea on human skin, wipe-samples from 13 human individuals (7 female and 6 male, age range 20–40, entire front torsos, Supplementary Table 1) were taken and analyzed with sensitive molecular techniques. All human subjects revealed the presence of archaeal 16S rRNA genes on their skin, accounting for up to 4.23% of the entire recovered prokaryotic microbiome (0.60% on average, *Figure III.2-1*). As Bacteria are likely to possess more ribosomal genes per genome (4.17 on average) compared to Archaea [1.69; (Klappenbach et al., 2001)], the average proportion of archaeal cells could be even greater (1.40% on average, max. 9.86%).

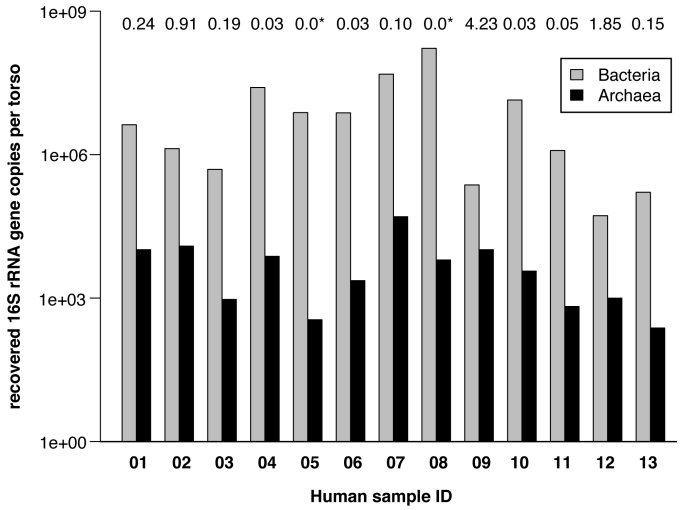

Figure III.2-1 *Abundance of bacterial and archaeal 16S rRNA gene copies retrieved from front torsos of 13 people. Values above bar graphs give percent of archaeal gene copies in the entire prokaryotic microbiome detected. Asterisks indicate an archaeal percentage lower than 0.01. X-axis gives human sample number (Table III.2-1), Y-axis shows log-transformed abundances of 16S rRNA genes.*

Five samples were selected for a deeper analysis of the archaeal diversity present on human skin. The archaeal community structure comprised OTUs of the phyla Thaumarchaeota (88% of all OTUs) and Euryarchaeota (12%) with 17 different taxonomic OTUs in total. All human subjects exhibited sequences of Thaumarchaeota. Phylogenetic analysis of these archaeal skin sequences placed them close to ammonia-oxidizing archaea from soil (Thaumarchaota group I.1b, *Figure III. 2-2*), but interestingly also close to sequences from built environments (clean rooms, intensive care units) found in this and earlier studies [(Moissl-Eichinger, 2011); *Figure III.2-2]*. Besides thaumarchaeal signatures, one human subject revealed two euryarchaeal sequences, which belonged to the taxon *Methanosarcina*, a putative methanogen reported previously for agricultural environments, but also the human intestine (Conway de Macario and Macario, 2009). Sequence classifications are summarized in *Table III.2-1*.

Table III.2-1 Analysis of clone libraries and classification of taxonomic OTUs.

Sample	Hum_01	Hum_04	Hum_07	Hum_08	Hum_10	Graz_12	Graz_13	Rgbg_1	Rgbg_3	Rgbg_F	Rgbg_F	Rgbg_3_ PMA	CR5	CR8	Classification	
Sample type	Human skin, wipe	Human skin, wipe	Human skin, wipe	Human skin, wipe	Human skin, wipe	Intensive care unit, BiSKit	Intensive care unit, BiSKit	Intensive care unit, wipe	Intensive care unit, wipe	Intensive care unit, wipe	Intensive care unit, wipe	Intensive care unit, wipe, PMA	Clean room ISO5, BiSKit	Clean room ISO8, BiSKit		
Detection method	Nested PCR	Nested PCR	Direct PCR	Nested PCR	Nested PCR	Nested PCR	Nested PCR	Direct PCR	Direct PCR	Nested PCR	Direct PCR	Direct PCR	Nested PCR	Nested PCR		
Coverage / # of clones screened	90.5% / 42	100% / 46	98.1% / 53	93.8% / 32	97.7% / 44	97.6% / 41	97.9% / 48	100% / 45	97.4% / 39	100% / 45	97.2% / 36	95.1% / 41	86.5% / 37	94.9% / 39	Phylum	Genus
CR5_31_CC, JX865688	-	-	-	-	-	-	97.92%	-	-	-	2.78%	-	-	-	Euryarchaeota	Halococcus
CR8_2_DB, JX865696	-	-	-	12.50%	-	-	-	-	-	-	-	-	-	-	Euryarchaeota	unclassified Halobacteriaceae
CR5_8_CD, JX865692	-	-	-	3.13%	-	-	-	-	-	-	-	-	-	-	Euryarchaeota	unclassified Halobacteriaceae
FlD_6_EA, JX865727	-	-	-	-	-	-	-	-	-	-	8.33%	-	-	-	Euryarchaeota	unclassified Halobacteriaceae
Zi1DN_20_FE, JX865743	-	-	11.32%	-	-	-	-	-	-	-	-	-	-	-	Euryarchaeota	Methanomethylovorans
Gr13_H02, JX865738	-	-	-	-	-	-	-	-	7.69%	-	-	-	-	-	Euryarchaeota	Methanosarcina
Gr13_H08, JX865739	-	-	3.77%	-	-	-	-	-	-	-	-	-	-	-	Euryarchaeota	Methanosarcina
MM_2_BA, JX865665	-	-	-	-	-	-	-	-	15.38%	-	-	-	-	-	Euryarchaeota	Methanosarcina
MM_6_BE, JX865673	-	-	-	-	-	-	-	8.89%	-	-	-	-	-	-	Euryarchaeota	Methanosarcina
Zi3D_28_BC, JX865756	-	-	-	-	-	-	-	-	-	-	-	-	-	51.28%	Euryarchaeota	unclassified Euryarchaeota
CMD_28_CE, JX865701	-	-	1.89%	-	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
FL_41_AC, JX865715	-	-	-	-	-	-	-	-	-	-	-	2.44%	-	-	Thaumarchaeota	Candidatus Nitrososphaera
Zi3D_26_BA, JX865755	-	-	-	-	-	-	-	-	-	-	-	-	-	30.77%	Thaumarchaeota	Candidatus Nitrososphaera
AA_15_AE, JX865655	2.38%	-	-	-	2.27%	-	-	42.22%	-	-	22.22%	51.22%	-	2.56%	Thaumarchaeota	Candidatus Nitrososphaera
AF_14_AB, JX865661	2.38%	-	9.43%	34.38%	-	-	-	44.44%	5.13%	-	-	-	2.70%	-	Thaumarchaeota	Candidatus Nitrososphaera
CMD_41_EA, JX865706	-	26.09%	28.30%	-	-	2.44%	-	-	64.10%	-	33.33%	19.51%	10.81%	2.56%	Thaumarchaeota	Candidatus Nitrososphaera
CR5_14_EA, JX865684	-	-	-	-	-	-	-	-	-	100.00%	-	19.51%	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CR5_4_BB, JX865690	-	-	-	-	-	21.95%	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CR8_1_AA, JX865693	-	-	-	3.13%	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CR8_10_CA, JX865694	-	-	-	12.50%	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
FLD_48_FD, JX865726	-	-	3.77%	-	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
Gr12_E3, JX865735	-	-	-	-	-	-	-	-	-	-	-	-	5.41%	-	Thaumarchaeota	Candidatus Nitrososphaera
Gr13_C7, JX865736	-	-	5.66%	-	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
Zi3P_15_CA, JX865760	-	-	-	-	-	68.29%	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CMD_3_CB, JX865702	-	-	22.64%	-	-	-	-	-	-	-	19.44%	4.88%	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CMD_40_CA, JX865705	-	-	-	-	-	-	-	-	-	-	-	-	2.70%	-	Thaumarchaeota	Candidatus Nitrososphaera
AA_1_AA, JX865653	2.38%	-	-	-	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
AA_10_CA, JX865654	2.38%	-	-	-	-	-	-	-	-	-	-	-	-	-	Thaumarchaeota	Candidatus Nitrososphaera
CMD2_17_AC, JX865709	-	-	-	-	-	-	-	-	-	-	-	-	8.11%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
CMD2_30_FC, JX865710	-	-	13.21%	28.13%	-	7.32%	-	4.44%	2.56%	-	-	2.44%	-	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
CMD2_35_CD, JX865711	-	-	-	-	-	-	-	-	-	-	-	-	2.70%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
CR5_12_AB, JX865683	-	-	-	-	-	-	-	-	-	-	-	-	2.70%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
CR5_47_DB, JX865691	-	-	-	-	-	-	2.08%	-	-	-	-	-	-	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
MM_13_DD, JX865663	90.48%	73.91%	-	6.25%	97.73%	-	-	-	-	-	-	-	37.84%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
MM_8_EC, JX865674	-	-	-	-	-	-	-	-	-	-	13.89%	-	8.11%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
CMD_38_DA, JX865704	-	-	-	-	-	-	-	-	-	-	-	-	2.70%	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
Gr12_D2, JX865733	-	-	-	-	-	-	-	-	5.13%	-	-	-	-	-	**Thaumarchaeota**	**Candidatus Nitrososphaera**
MM_32_DC, JX865670	-	-	-	-	-	-	-	-	-	-	-	-	-	12.82%	**Thaumarchaeota**	**Candidatus Nitrososphaera**
Zi3D_12_CB, JX865752	-	-	-	-	-	-	-	-	-	-	-	-	16.22%	-	unclassified Archaea	unclassified Archaea

Abbreviations: ICU (intensive care units); CR (clean room)

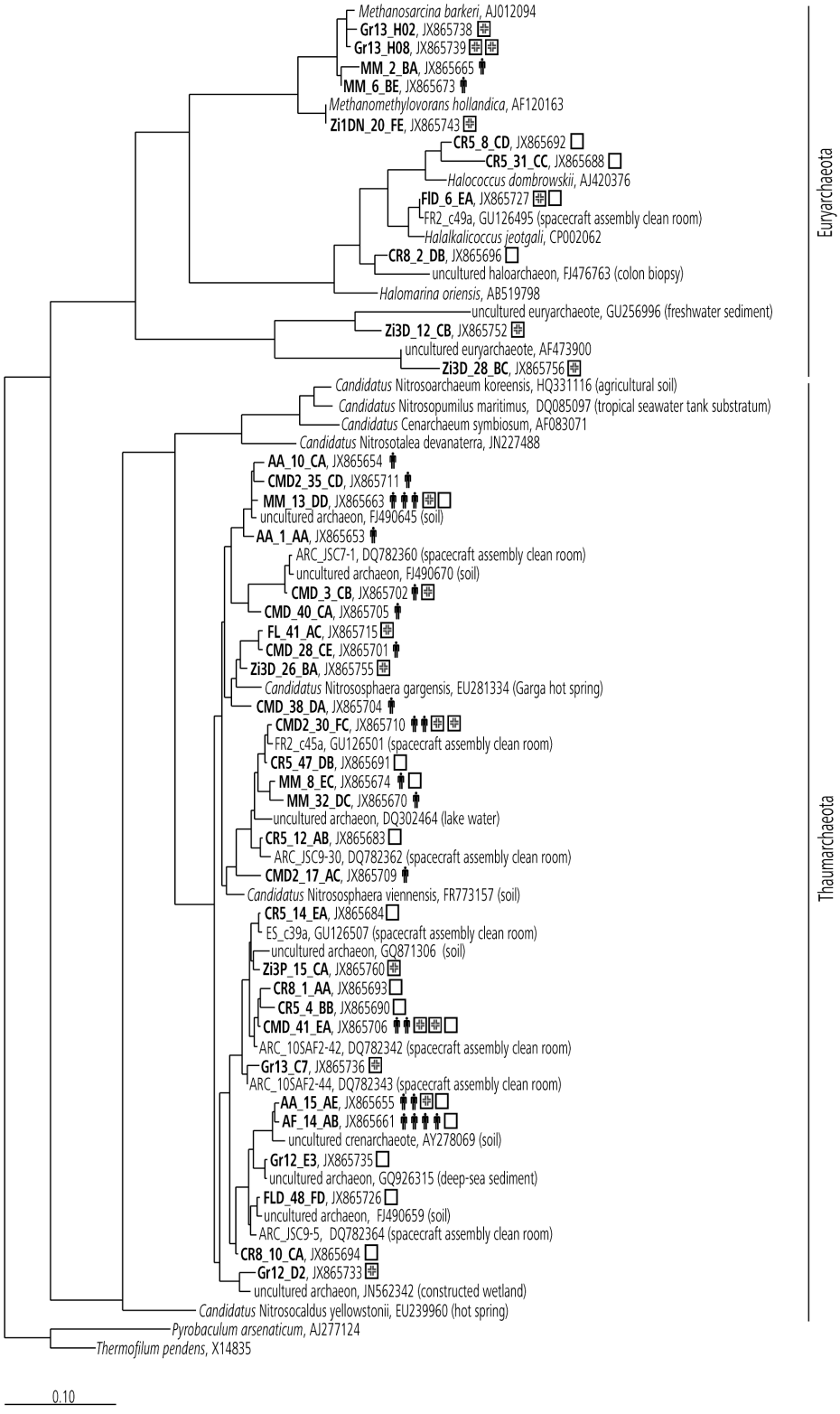

Figure III.2-2 *Maximum likelihood tree displaying all detected OTUs from human skin, intensive care unit, and clean room environments. Symbol "man": phylotype retrieved from human skin (the number of symbols gives the number of individuals carrying this phylotype; 5 subjects were screened with respect to the archaeal 16S rRNA gene pool). Symbol "hospital" (square with cross): phylotype detected in intensive care unit (two intensive care units were screened). Symbol "square": detected in a spacecraft assembly clean room (one facility was analyzed). Symbol "star" highlights phylotypes that were also found in the propidium monoazide (PMA)-treated sample, i.e. from cells with intact membranes. Scale bar refers to 10% nucleotide substitutions. Pyrobaculum arsenaticum and Thermofilum pendens (Crenarchaeota) were used as an outgroup.*

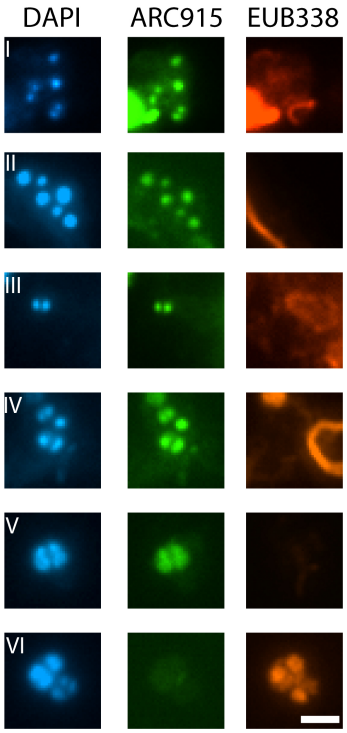

Figure III.2-3 *Fluorescence in situ hybridization, performed on a human skin wipe-sample for visualization of Archaea. DNA-containing cell (DAPI stain): blue, Archaea: green, Bacteria: red. I-V: Examples of positive archaeal signals (small cocci, probe ARC915 labeled with rhodamine green) are shown, which give a positive signal with DAPI and no signal with the Bacteria-directed probe (EUB338/I labeled with CY3). VI: Example of a positive bacterial signal. Bar: 2 µm.*

The presence of Archaea on human skin was further confirmed by fluorescence *in situ* hybridization experiments (*Figure III.2-3*). Unfortunately, the relative abundance of archaeal compared to bacterial cells could not be calculated due to the high amount of fluorescence-active particulates and fibers in the specimens (see Material and Methods for details), but archaeal signals were obvious and easily detectable. Archaeal cells were visualized as small cocci (approx. 0.5 µm in diameter) in the skin microbiome. Their shape and size was similar to thaumarchaeal cells previously detected in sludge samples (Mussmann et al., 2011).

Because human-dominated environments reflect the microbial diversity associated with the human skin and body (Flores et al., 2011), we can now propose a logical reason for the earlier discovery of archaeal signatures in controlled clean room facilities around the world (La Duc et al., 2012; Moissl et al., 2008; Moissl-Eichinger, 2011). The presence of (thaum-)archaeal signatures in clean rooms was confirmed in this study (EADS clean room facility) and further expanded to hospitals. Two intensive care units were sampled and the presence of Thaumarchaeota in these artificial environments was affirmed (*Figure III.2-2*). So far, all man-made environments studied by the authors have revealed the presence of archaeal 16S rRNA genes, belonging to two different phyla, Thaumarchaeota and Euryarchaeota, which can be attributed to the sensitive and appropriate assays employed. The integrity and therefore probable viability of archaeal cells in floor wipe-samples were also proven by FISH (Moissl-Eichinger, 2011) and with a molecular-based viability assay (Supplementary Table 1, *Figure III.2-2*). The clean room euryarchaeota included methanogens and different halophiles; both groups have been associated with human mouth and intestinal flora (Conway de Macario and Macario, 2009; Oxley et al., 2010), and were recently also found in navels (Hulcr et al., 2012). However, the aforementioned study on navels did not reveal any thaumarchaeal signatures and only three euryarchaeal phylotypes in a very small subset of the samples (Caporaso et al., 2011). Euryarchaeota (again methanogens and halophiles), Crenarchaeota and Thaumarchaeota were detected in a study by Caporaso et al., 2011, who analyzed the microbial community of human gut, tongue and palms of only two individuals but at 396 timepoints. In particular, the palm microbiome revealed the (fluctuating) presence of *Nitrososphaera* related sequences. Based on their statistical analysis, the authors considered those organisms insignificant and transient members of the human microbiome. Certainly, the palm skin represents one of the major contact surfaces of humans to their biotic and abiotic environment, and transient microorganisms can be found there more than elsewhere on human skin. Consequently, we excluded human palms and focused on human torso skin, which might harbor a more typical, less influenced microbial

diversity. We argue based on our results of 13 human skin samples, that Caporaso et al. may have underestimated the importance of Archaea and particularly Thaumarchaeota on human skin. To emphasize the finding of a general presence of Archaea on human skin (all samples revealed archaeal signatures), we were even able to visualize archaeal cells, indicating their active physiological status and their obvious presence in samples from human skin. We can conclude that previous studies have either (methodically) overlooked the archaeal diversity associated with human skin or underestimated their abundance.

Our study was the first to systematically show that Archaea, in particular Thaumarchaeota, are consistently present on human skin. This finding leads to a number of questions that cannot be answered at the current status of knowledge. For instance, the role, metabolism, infestation rate or also the origin of archaea associated with skin are unclear to date and need to be tackled in subsequent studies.

Interestingly, representatives of Thaumarchaeota cluster I.1b, were found in soil and aquatic environments, but also in wastewater treatment plants, as reported recently (Mussmann et al., 2011; Sauder et al., 2012). The detection of such thaumarchaeal sequences and cells on human skin and engineered environments could point to novel, currently unknown roles and metabolic capabilities besides chemolithoautotrophy (Mussmann et al., 2011). However, we note that all so far cultivated thaumarchaeal species are ammonia oxidizers (Konneke et al., 2005; Lehtovirta-Morley et al., 2011; Pester et al., 2011; Tourna et al., 2008) and the human skin is constantly emanating low amounts of ammonia (Nose et al., 2005). We were able to amplify and sequence *amoA* genes from a pooled sample of all 13 human subjects. Megablast analyses against the NCBI nucleotide collection showed clearly that these sequences belong to *amoA* genes (E-value = 0). A phylogenetic tree of the retrieved *amoA* sequences is depicted in *Figure III.2-4*. The sequences obtained were not closely related to *Nitrososphaera viennensis* or *N. gargensis amoA* genes, but were affiliated to two *Nitrososphaera* subcluster (4 and 6, *Figure III.2-4*), similar to the majority of the 16S rRNA genes (*Figure III.2-2*).

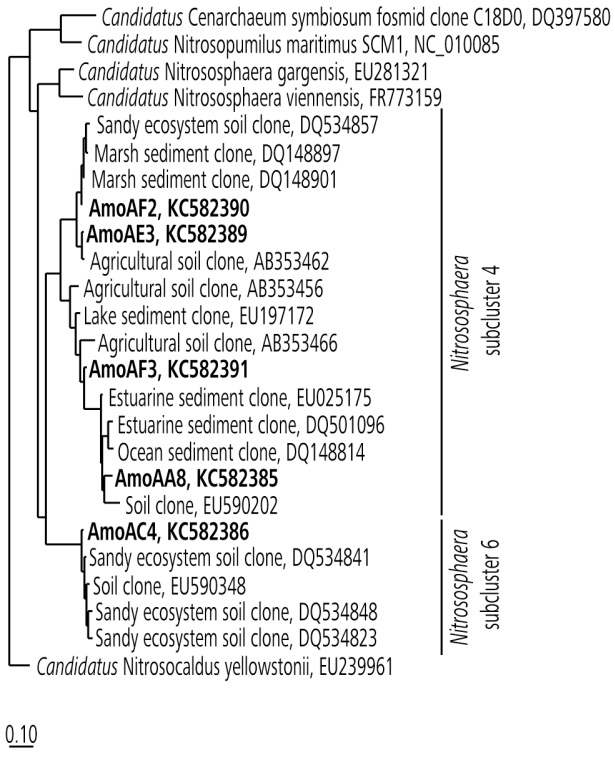

Figure III.2-4 *Maximum likelihood tree based on archaeal amoA gene sequences. Sequences recovered in this study are shown in bold. Information in parenthesis gives the number of retrieved sequences. Bar refers to 10% nucleotide substitutions per site.*

This finding suggests at least one possible explanation for the presence of these microorganisms. It can be hypothesized, that a chemolithotrophic ammonia turnover by Thaumarchaeota could influence the pH regulation of the human skin and therefore the natural protective layer, but this remains to be proven. Additionally, the interaction of humans with Archaea seems not to be restricted to a passive and/or indirect methanogenic activity in colon and mouth cavity, but could be an

active and direct relationship. Because protocols as used for the Human Microbiome Project apparently underestimated the presence of Archaea in general, screening methods in science and medicine should be better geared towards improved detection of Archaea, which will then lead to a comprehensive understanding of their beneficial or potentially pathogenic role in the human (skin) microbiome.

Acknowledgments

We thank Robert Huber, Reinhard Wirth, Emma Gagen, Harald Huber, Shariff Osman for valuable discussion and suggestions. We are grateful to the teams in Friedrichshafen (EADS Astrium), Graz (Gabriele Berg, Universitätsklinik Graz) and Regensburg (Caritas-Krankenhaus St. Josef) for the possibility to access and sample clean rooms and intensive care facilities. We are grateful to people who provided skin samples. Funding provided by the Institute for Microbiology and Archaea Center (Michael Thomm), the University of Regensburg and the European Space Agency [Contract No. 4000103794/11/NL/PA under subcontract with DLR (D/316/67130713)] is acknowledged. AJP was supported by the National German Academic Foundation (Studienstiftung des deutschen Volkes).

Supplementary Information

Supplementary information can be found online http://s3-eu-west-1.amazonaws.com/files.figshare.com/1083863/Table_S1.docx or on the supporting DVD.

3. Tackling the minority: sulfate-reducing bacteria in an archaea-dominated subsurface biofilm

Alexander J Probst[1,2], Hoi-Ying N Holman[2], Todd Z DeSantis[3], Gary L Andersen[2], Giovanni Birarda[2], Hans A Bechtel[4], Yvette M Piceno[2], Maria Sonnleitner[1], Kasthuri Venkateswaran[5] and Christine Moissl-Eichinger[1]

[1]Institute for Microbiology and Archaea Center, University of Regensburg, Regensburg, Germany; [2]Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [3]Department of Bioinformatics, Second Genome Inc., San Bruno, CA, USA; [4]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [5]Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Correspondence: C Moissl-Eichinger, University of Regensburg, Department for Microbiology and Archaea Center, Universitaetsstrasse 31, 93053 Regensburg, Germany. E-mail: christine.moissl-eichinger@ur.de

Publication information:
The ISME Journal (2013) 7, 635–651; doi:10.1038/ismej.2012.133; published online 22 November 2012
Received 23 October 2011; Revised 3 September 2012; Accepted 24 September 2012
Link: http://www.nature.com/ismej/journal/v7/n3/full/ismej2012133a.html

Abstract

Archaea are usually minor components of a microbial community and dominated by a large and diverse bacterial population. In contrast, the SM1 Euryarchaeon dominates a sulfidic aquifer by forming subsurface biofilms that contain a very minor bacterial fraction (5%). These unique biofilms are delivered in high biomass to the spring outflow that provides an outstanding window to the subsurface. Despite previous attempts to understand its natural role, the metabolic capacities of the SM1 Euryarchaeon remain mysterious to date. In this study, we focused on the minor bacterial fraction in order to obtain insights into the ecological function of the biofilm. We link phylogenetic diversity information with the spatial distribution of chemical and metabolic compounds by combining three different state-of-the-art methods: PhyloChip G3 DNA microarray technology, fluorescence in situ hybridization (FISH) and synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy. The results of PhyloChip and FISH technologies provide evidence for selective enrichment of sulfate-reducing bacteria, which was confirmed by the detection of bacterial dissimilatory sulfite reductase subunit B (dsrB) genes via quantitative PCR and sequence-based analyses. We further established a differentiation of archaeal and bacterial cells by SR-FTIR based on typical lipid and carbohydrate signatures, which demonstrated a co-localization of organic sulfate, carbonated mineral and bacterial signatures in the biofilm. All these results strongly indicate an involvement of the SM1 euryarchaeal biofilm in the global cycles of sulfur and carbon and support the hypothesis that sulfidic springs are important habitats for Earth's energy cycles. Moreover, these investigations of a bacterial minority in an Archaea-dominated environment are a remarkable example of the great power of combining highly sensitive microarrays with label-free infrared imaging.

Introduction

Although the Archaea-scientific community is evolving fast, the lack of knowledge with respect to mesophilic and cold-loving archaea is still enormous. The recent cultivation success of thaumarchaeal representatives is revealing novel and fascinating information, as are alternative procedures that allow *in situ* studies of archaea in their natural environment or in microcosm experiments (Dekas et al., 2009; Hatzenpichler et al., 2008; Tourna et al., 2011; Walker et al., 2010). One major challenge in understanding the ecological role of archaea is that they are underrepresented in most natural systems, typically accounting for much less than 50% of the microbial cells present. Although some reports have revealed a predominance of (cren-) archaeal cells in marine water columns, reaching numbers of up to 90% Archaea versus Bacteria (Karner et al., 2001), the archaeal part is composed of a broad diversity in these settings (DeLong, 1998). Natural environments that are predominated by one single species of Archaea are rare; the most famous are the anaerobic methane-oxidizing (AMO) consortium (Orphan et al., 2001) and the 'string-of-pearls community' (Rudolph et al., 2001). Both of these consortia seem to be based on syntrophy, in which both partners are mutually dependent on each other for nutrient exchange (Moissl-Eichinger and Huber, 2011). The AMO consortium has been the subject of numerous analyses and is currently fairly understood but the string-of-pearls community, and in particular the archaeal partner therein (SM1 Euryarchaeon), is still mysterious in many aspects.

The SM1 Euryarchaeon is found in sulfide-containing fresh and marine waters all over Europe (Rudolph et al., 2004), but only two sites (close to Regensburg, Bavaria, Germany) were studied extensively during the past 10 years: The Sippenauer Moor and the Muehlbacher Schwefelquelle ["Islinger Muehlbach"; (Henneberger et al., 2006)]. Both of these sites are characterized by a main, sulfidic spring, emanating into a streamlet where whitish mats of sulfide-oxidizing bacteria cover the submerged surfaces. These aquifers are very similar to sulfidic cave springs that are rich in sulfide, ammonia and sulfate (Engel et al., 2004) but poor in dissolved organic carbon, suggesting that the major microbial community of the biotopes are chemolithoautothrophs (Engel et al., 2003; Kodama and Watanabe, 2004). Although, sulfidic springs represent <10% of terrestrial fresh water springs (Palmer, 1991), they are believed to have an important role in global sulfur-cycling (Engel et al., 2003), as they can spawn huge amounts of microbial biomass mainly consisting of sulfur-oxidizing bacteria such as *Thiothrix*, *Beggiatoa* and *Sulfuricuvum*. These filamentous bacteria live as microbial mats or streamers in fluctuating gradients of sulfide and oxygen, and may also be responsible for the environmental success of the SM1 Euryarchaeon under oxygen-rich conditions (Moissl et al., 2002; Rudolph et al., 2004). Surrounding the archaeal colony, *Thiothrix* (Sippenauer Moor) and *Sulfuricurvum* (Muehlbacher Schwefelquelle) form the string-of-pearls community and possibly interact with the archaeon through an inter-species sulfur cycle (Moissl et al., 2002). In these communities, the bacterial partner and the SM1 Euryarchaeon are present in almost equal abundance, pointing at a 'real' partnership and possibly at a symbiotic/syntrophic relation. The proposed sulfur cycle suggests the SM1 Euryarchaeon being an anaerobic sulfate reducer surrounded by sulfur-oxidizing bacteria. The latter metabolize products from sulfate reduction (H_2S), and provide simultaneously the educts (sulfate) for the SM1 Euryarchaeon. In addition, the sulfur-oxidizing bacteria protect the SM1 Euryarchaeon from oxygen exposure by respiration (Moissl et al., 2002).

In contrast to other sulfidic springs that have been microbiologically studied, samples at the Muehlbacher Schwefelquelle can also be taken from ~1 m below the water table, where the upwelling water is not yet mixed with atmospheric oxygen. By placing an *in situ* trapping system in this subsurface setting, slime-like biofilm structures consisting almost exclusively of SM1 euryarchaeal cells can be caught from the water stream, in stark contrast to the abovementioned string-of-pearls community (Henneberger et al., 2006). This second life-style of the SM1 Euryarchaeon differs also from other described microbial systems, in which archaea are involved in biofilm formation (Frols et al., 2008; Lapaglia and Hartzell, 1997; Tyson et al., 2004): first, the SM1 euryarchaeal biofilm represents the only known naturally occurring Archaea-dominated biofilm, revealing a purity of up to 95%based on microscopic counts (Henneberger et al., 2006). Second, the small archaeal cocci form porous colonies with defined distances between the single cells mediated by their unique cell surface appendages (Henneberger et al., 2006; Moissl et al., 2005b). Third, bacteria in the biofilm are either randomly distributed or form dense microcolonies, and their varied morphological appearance hints at a broader genetic diversity. Lastly, no other archaea have been detected within the biofilm, using fluorescence *in situ* hybridization (FISH) or conventional cloning strategies, suggesting that the SM1 euryarchaeal biofilm is a natural 'archaeal monospecies biofilm' (Henneberger et al., 2006). The Muehlbacher

Schwefelquelle spring therefore represents an extraordinary window to an anoxic subsurface biotope of an unusual archaeon.

Using basic biochemical analyses, the water content of the SM1 Euryarchaeon biofilm has been determined to be extraordinary high (99.6%; Amann T et al., unpublished data) and the composition of the extracellular polymeric substance has been shown to have a high ratio of protein versus carbohydrates (1.5:1). No nucleic acids, however, were found in the matrix surrounding the SM1 Euryarchaeon cells (Henneberger et al., 2006). The protein content is mainly owing to its extraordinary cell surface structures, called *hami*, which are highly-complex, filamentous attachment tools with a nano-sized grappling hook at their end (Moissl et al., 2005b).

The biochemical analyses performed to date have been based on protocols that necessitate a complete extraction of chemical compounds from the biofilm and do not allow the assignment of organic and inorganic compounds to the different microbes in the biofilm (for example, to Bacteria or Archaea). Moreover, for the underrepresented bacteria in the biofilm neither their metabolic role nor their (possible metabolic) interaction with the archaea are defined or have been subject to deeper studies besides FISH (Henneberger et al., 2006). Hence, it is uncertain, whether bacterial key species coexist with the SM1 Euryarchaeon, or the detected bacterial diversity is randomly attached to the biofilm.

In order to understand the bacterial (and archaeal) diversity in the biofilm and a possible occurrence of certain key species therein, we have conducted highly sensitive PhyloChip analyses based on the 16S rRNA gene pool of the biofilm. In addition, we used synchrotron radiation-based Fourier transform infrared (SR-FTIR) spectromicroscopy to provide a nucleic-acid independent method to link the phylogenetic diversity information with the spatial distribution of the chemical composition and metabolic activity of the bacterial and archaeal cells within the biofilm. SR-FTIR is a non-invasive and label-free molecular imaging technique capable of micrometer spatial resolution (Holman et al., 2010). In this study, the capability of SR-FTIR to differentiate Bacteria from archaeal cells has been evaluated.

Material and Methods

Sampling site and physical characteristics. Biofilm samples were collected from the cold (~10.5 °C), sulfidic spring Muehlbacher Schwefelquelle. Its physical characteristics (pH and water composition) have already been described previously (Henneberger et al., 2006; Rudolph et al., 2004), and are found to be very constant over several years of measurement (including sulfate 16 mg l^{-1}, thiosulfate 14 mg l^{-1}, ammonia 0.33 mg l^{-1}). Oxygen concentrations at different locations of the spring and the stream were re-measured using a highly sensitive oxygen dipping probe (PSt6) coupled with temperature measurement (Fibox 3, LCD trace; PreSens, Regensburg, Germany).

Sample collection. An *in situ* biofilm trapping system was used to catch biofilm pieces washed up from the deeper subsurface. The nets were incubated for 3 days as deep as possible in the spring bore. Sampling was performed as described earlier (Henneberger et al., 2006). Samples for FISH analysis were incubated in phosphate-buffered saline-containing paraformaldehyde [final concentration 3% (wt/vol)] for 1 h at room temperature (22 °C±2 °C); samples for PhyloChip G3 assays were frozen at –20 °C and samples for SR-FTIR spectromicroscopy necessitated air-drying of the biofilm on gold screens (G225G1, Plano GmBH, Wetzlar, Germany). In addition, 25 ml of spring water were collected as a field control for PhyloChip experiments. Anaerobic sampling for incubation experiments was performed as follows: a double-opened Schott flask was placed on a funnel letting almost all water of the spring pass through. The flask had several layers of polyethylene nets to filter the spring water and catch biofilm fragments. After an incubation of 4 days the bottle was closed with rubber stoppers under water (oxygen-free conditions). All samples were kept on ice during the transport from the site to the laboratory.

Metagenomic DNA extraction. A measure of 250 µl of each biofilm sample were used for individual extraction procedures. Spring water was concentrated via a Millipore amicon 50 kDa cutoff centrifugal filter (Millipore, Billerica, MA, USA), according to manufacturer's specifications before undergoing DNA extraction as described previously (Moissl-Eichinger, 2011; Tillett and Neilan, 2000). Concentrations of double-stranded DNA in the samples were determined using Qubit Quantitation Platform (Invitrogen, Carlsbad, CA, USA).

Quantitative PCR and cloning of dsrB genes. Quantitative PCR (qPCR) was carried out in triplicates with 1 µl of metagenomic DNA as described previously (Moissl-Eichinger, 2011), and the following primer sets were used. Archaeal 16S rRNA genes: 345aF-517uR (Burggraf et al., 1992; Lane, 1991; Moissl-Eichinger, 2011); bacterial 16S rRNA genes: 338bF-517uR (Lane, 1991); dissimilatory sulfite reductase subunit B (*dsrB*) genes: DSRp2060F (Geets et al., 2006) and DSR4R (Wagner et al., 1998). 16S rRNA gene standards were developed from PCR products of *Methanococcus aeolicus* (DSM 17 508) and *Bacillus safensis* (DSM 19 292).
dsrB gene standard was generated from an environmental biofilm sample. After PCR-amplification of *dsrB* genes with the abovementioned primers the amplicons were cloned into pCR2.1-Topo vector. Fifty-two clones were randomly picked and inserts were sequenced using M13F and M13R primers. Forty-eight clones revealed high quality and were vector-trimmed, clustalW aligned and grouped into operational taxonomic units (OTU) at a 0.01 hard cutoff (Schloss et al., 2009). One representative sequence of each OTU was submitted to GenBank (Acc. no. JX515394–7); a representative clone of the dominant OTU (JX515394) was used for generating a qPCR standard (PCR amplicon generated with M13 primers). The coverage of the library was calculating according to Good (Good, 1953).

16S rRNA gene amplification. The template concentration for PCR was set to 3 ng for biofilm samples but DNA isolated from spring water revealed no measurable concentrations (<0.05 ng) due to low biomass. Consequently, 1 µl of template was used for single PCR; the same settings were also applied for the extraction blank (see below). Bacterial 16S rRNA genes were amplified in a gradient PCR using primers B27f and 1492r as described elsewhere (Hazen et al., 2010), and 30 cycles were run. For amplification of archaeal 16S rRNA genes the degenerated primer pair 345af [5′-CGGGGYGCASCAGGCGCGAA-3′ (Burggraf et al., 1992)] and 1406ur [5′-ACGGGCGGTGTGTRCAA-3′ (Lane, 1991)] with an annealing temperature of 60 °C were chosen. Running only 25 PCR cycles and an evaluation of the primers via RDP II (Cole et al., 2009) in comparison to previous Archaea-directed primers (Hazen et al., 2010) promised an increase of the detectable archaeal biodiversity (coverage of these and previous primers evaluated via RDP II, Supplementary Table S1). PCR products were gel-purified prior to cloning or microarray analysis (QIAquick Gel Extraction Kit, Qiagen, Germany).

Archaeal clone library. The PCR products of one biofilm sample were used to generate an archaeal 16S rRNA gene clone library by using the TOPO TA cloning kit with TOP 10′ cells (Invitrogen). Colonies were manually picked and inserts were amplified using the abovementioned archaeal primer pair. For screening, restriction-fragment-length polymorphisms were performed using two restriction enzymes [*Hae*III and *Hin*fI, Promega, Madison, WI, USA (Moissl-Eichinger, 2011; Vaneechoutte et al., 1992)]. Plasmids of clones with unique sequences were purified (Plasmid Mini DNA Purification Kit, Invitrogen) and bi-directionally sequenced using M13 primers (University of California, DNA Sequencing Facility, Berkeley, USA). After chimera check via Bellerophon [version 3 (http://greengenes.lbl.gov/)] and Pintail (Ashelford et al., 2005) sequences were compared with publicly available sequences and among each other using BLAST [http://blast.ncbi.nlm.nih.gov; (Altschul et al., 1990)].

16S rRNA gene microarray (PhyloChip G3) analysis. The PhyloChip G3 design, performance and analysis were already described (Hazen et al., 2010). Here, 500 ng of bacterial and 100 ng of archaeal 16S rRNA gene amplicons were used for PhyloChip analysis of biofilm samples. Hence, only 100 ng of bacterial PCR amplicons and 10 µl of archaeal amplicons (below detection limit) were hybridized on the chip for the background water sample. DNA extraction blanks yielded no quantifiable amounts and 14.5 µl of bacterial and 10.0 µl of archaeal post-PCR were used for PhyloChip assay. After combining amplicons, they were spiked with known amounts of non-16S rRNA genes (total 202 ng). Fluorescence intensities of these positive controls were used to normalize total array intensities among samples. Target fragmentation, biotin labeling, PhyloChip hybridization, scanning and staining, as well as background subtraction, noise calculation, detection and quantification criteria were performed as reported (Hazen et al., 2010).

PhyloChip data processing. Stage 1 and 2 analysis were performed as described elsewhere (Hazen et al., 2010) and thus, the threshold for identifying a bacterial OTU in a sample was set to a minimum of 18 perfect match probes. Quartiles of the ranked *r-scores* (response score to determine the potential of a probe pair responding to a target and not to the background) had to meet the following criteria: $rQ_1 \geq 0.70$, $rQ_2 \geq 0.95$, $rQ_3 \geq 0.98$. In addition, subfamilies that had an r_xQ_3

value (cross-hybridization adjusted response score) of ≥0.48, were considered as present but also requirement for the OTUs within this subfamily to be present.

For analysis of the archaeal OTUs the aforementioned parameters were adjusted to the smaller ~1000 bp 16S rRNA gene amplicons. As shown in Supplementary Figure S1A, the number of probe pairs that could possibly be scored with these amplicons, varied among the archaeal OTUs on the chip. Hence, the criterion to call an archaeal OTU present was adjusted to a probe pair score of 14. Consequently, 92/639 of the archaeal OTUs present on the PhyloChip could not be included in the analysis (Supplementary Figure S1A); however, these OTUs were not restricted to one specific phylum and spread within the domain of the Archaea.

Subfamily based analysis was done by picking one representative sequence within an OTU per subfamily that was detected at least in 2/3 of the biofilm samples or in the background water. These OTUs were classified to family level using the Greengenes (DeSantis et al., 2006) database in combination with SILVA (Pruesse et al., 2007) and RDP II (Cole et al., 2009). Trees based on multiple sequence alignments were generated by retrieving 70 000 character alignments from SILVA database and the neighbor joining method [MEGA 4, (Tamura et al., 2007)]. Trees with heatmaps were rendered in iTOL (Letunic and Bork, 2007).

Identification of significantly enriched OTUs in the biofilm. For identification of OTUs that were significantly enriched in the biofilm, three additional samples were included in the PhyloChip analysis that were taken from the Sippenauer Moor, where the SM1 Euryarchaeon can be cultivated *in situ* as a string-of-pearls community together with filamentous sulfur-oxidizing bacteria (Moissl et al., 2002; Rudolph et al., 2001). The samples from the Sippenauer Moor were collected as described previously (Moissl et al., 2003), and underwent the same molecular PhyloChip assay as described for the biofilm samples in this study. A two-tailed, homoscedastic Student's *t*-test was performed on abundance values of OTUs detected in at least one of the three biofilm or string-of-pearl-community samples. An adjusted *P*-value of 0.002 was chosen in order to avoid type one errors as 8114 different OTUs were included in the analysis. The resulting OTUs that met this requirement and had a higher average abundance in biofilm samples were then grouped in subfamilies by picking the OTU with the most drastic increase in abundance. Results of the string-of-pearls community analysis were used as a reference data set of an oxygen-exposed environment of the SM1 Euryarchaeon and were not included in this manuscript. Heatmaps of selected OTUs were generated in the R programming environment (http://www.r-project.org/).

Tracking the SM1 Euryarchaeon with PhyloChip technology. As the full 16S rRNA sequence of the SM1 Euryarchaeon is still not publicly available (Rudolph et al., 2001), this Euryarchaeon had not been included in the PhyloChip G3 design (Hazen et al., 2010). In order to track the abundance of the SM1 Euryarchaeon in samples analyzed with PhyloChip G3, the 1019 bp-long sequence of the dominant SM1 Euryarchaeon clone (IM-A1, JN861739) was bioinformatically broken up into 995 25-mers and compared with all probes present on the PhyloChip. Nine different probes were identified to perfectly match with the A1 clone sequence but only one of them revealed high specificity. By using RDP II probe match (Cole et al., 2009) the probe 5′-TGTGCAAGGAGCGGGGACATATTCA-3′ on the microarray (x=651, y=188; oligonucleotide sequences © 2011 Second Genome Inc.) was identified to match with only seven different archaeal sequences in the database, three of them belonging to SM1 euryarchaeal sequences. The relative hybridization intensities of this probe for biofilm, and water samples as well as positive and negative controls are given in Supplementary Figure S1B. As a matter of fact, the biofilm showed higher relative hybridization intensities (100 ng of archaeal PCR product) than the positive control (50 ng of purified PCR product of clone A1), reflecting the high abundance of the SM1 Euryarchaeon in the biofilm. The negative control and the extraction blank (see below) revealed very low intensities. Furthermore, the background water sample from the spring showed a very weak, relative hybridization intensity of 324 compared with an average value of 6913 retrieved from biofilm samples indicating a latent presence of the SM1 Euryarchaeon in the spring water (Supplementary Figure S1B).

Molecular analysis controls. Controls were included in each step mentioned above. For DNA extraction, a negative control performed with PCR grade water was used. The same control was then included in archaeal and bacterial 16S rRNA gene amplification and in PhyloChip analysis. Probes with positive response were masked in PhyloChip analysis of the actual samples in order to avoid false positives. In addition, a negative control of the PhyloChip analysis was performed that included nuclease- and nucleic-acid-free water as well as the spike genes only. However, no OTUs met the threshold

requirement neither in the extraction blank nor in the negative control. As a positive control, 50 ng of SM1-Euryarcheon clone IM-A1 (JN861739) was run in microarray analysis in order to see probe responses to the SM1-Euryarchaeon amplicon in comparison to environmental samples, which detected also no other OTUs.

FISH and fluorescence microscopy. Whole-cell hybridizations were performed as mentioned elsewhere (Rudolph et al., 2001), using domain- and species-directed probes [Bacteria: EUB338 (Amann et al., 1990b)], Archaea: ARCHmix (Henneberger et al., 2006; Moissl et al., 2002), SM1 Euryarchaeon: SMARCH714 (Moissl et al., 2003). For the detection of bacteria involved in sulfate reduction, the sulfate-reducing bacteria (SRB) 385 probe (Amann et al., 1990a) and the Delta495a/b/c probe mix was applied (Loy et al., 2002). Probes were labeled with Rhodamine Green (RG), Cy3 and Texas Red. Specimens were afterwards analyzed using either confocal laser scanning microscopy (CLSM, LSM 510 Meta, Zeiss, Oberkochen, Germany; exc. 488 and 543/ em. LP 505 and LP 585; multi-track for RG and Cy3) or epifluorescence microscopy (Olympus BX-60, Hamburg, Germany). For controls, a fluorescent dye-labeled nonsense probe (NONEUB338) were applied to the samples, and separate bacterial controls were also included (*Bacillus atrophaeus* DSM7264, *Escherichia coli* K12 DSM30083). Theoretical coverage of FISH probes and representative sequences of PhyloChip OTUs (*in silico* FISH) was assessed using the ARB software package (Ludwig et al., 2004).

CTC-FISH to measure activity of specific microorganisms. Biofilm samples were sampled anaerobically as described above and handled in an anaerobic glove box (Coy, Grass Lake, MI, USA). Biofilms were supplemented with 100 µl spring water and 10 µl 50 mM CTC (5-cyano-2,3-ditoryl tetrazolium chloride; (Stellmach, 1984; Stellmach and Severin, 1987; Yoshida and Hiraishi, 2004), which was prepared under anaerobic conditions (N_2 gas phase). After an anaerobic incubation of 2 h at 11 °C in a waterbath, the biofilms were removed and underwent fixation, FISH in suspension [Delta495a/b/c mix, RG, performed similar to Wallner et al. (Wallner et al., 1993)] and subsequent DAPI (4′,6-diamidino-2-phenylindole) staining.

SR-FTIR spectromicroscopy imaging and data analysis. SR-FTIR spectromicroscopy is a non-invasive and label-free chemical imaging technology that provides molecular information at micrometer spatial resolution (Carr et al., 1995; Dumas et al., 2009). SR-FTIR takes advantage of three technologies: (i) the well-known sensitivity of infrared spectroscopy to the bond vibration frequencies in a molecule for determining molecular functional groups, (ii) the convenience of a light microscope to locate areas for molecular and composition analysis, and (iii) the 100- to 1000-fold increase in signal-to-noise provided by a bright SR-based infrared light source. Using photons in the mid-infrared region (~2.5 to ~15.5 µm wavelength, or ~4000 to ~650 wavenumber in cm^{-1}), SR-FTIR spectromicroscopy has been successfully used to characterize microbial activities in geological materials and in both hydrated and dried biofilms (Hazen et al., 2010; Holman et al., 2010; Holman et al., 2009), in spite of the limitation that some signals may be ambiguous.

Freshly harvested samples (four replicates) were gently air dried onto gold-coated copper disks. Although drying affects the three-dimensional structure of the biofilms, prior microscopy experiments with other biofilms suggest that the two-dimensional structure is largely unaffected. Therefore, the measured spatial distribution of Bacteria, Archaea and the biogeochemical features could represent their native two-dimensional distribution within the biofilm. All SR-FTIR spectromicroscopy measurements were performed in the transflectance mode at the infrared beamline of the Advanced Light Source (http://infrared.als.lbl.gov/), where mid-infrared photons emitted from the synchrotron are focused with a 0.65 numerical aperture objective in a Nicolet Nic-Plan infrared microscope. In transflectance mode, the beam is transmitted through the sample, reflected off the gold-coated copper surface and then transmitted through the sample a second time before striking the mercury cadmium telluride detector. Each spectrum is an average of eight scans at a spectral resolution of 4 cm^{-1}. Background spectra were obtained on the cell-free area of the discs.

For each SR-FTIR imaging measurement, the 200 µm × 150 µm field-of-view for the biofilm was divided into equal-sized 2 µm × 2 µm pixels before raster scanning. The resulting data cube, which consists of position-associated FTIR spectra, was subjected to data preprocessing and processing calculations, including spectrum baseline removal, using both Thermo Scientific Omnic version 7.3 (Thermo Scientific, Madison, WI, USA) and Matlab (MathWorks, Nattick, MA, USA). The absorption spectra were then subjected to univariate and unsupervised multiple curve resolution (MCR) image analyses. The univariate approach, which integrates infrared absorbance of an individual peak of interest, relates the absorbance intensity to the relative concentration of a particular chemical component through the Beer–Lambert law. The unsupervised MCR

approach, on the other hand, is based on the principal component analysis (PCA) of the entire fingerprint region (1800–700 cm^{-1}) and of the C-H region (3100–2800 cm^{-1}) instead of individual peaks. MCR analysis of SR-FTIR spectra was applied to reveal the distributions of Archaea, Bacteria and chemical variations in the biofilms, which were hidden in the univariate approach. In this study, the unsupervised MCR analysis was performed with non-negative constraints on both concentration and spectral values (Budevska et al., 2003).

Validation of SR-FTIR for differentiating Archaea and Bacteria in biofilms. Our SR-FTIR approach assumed that Bacteria can be distinguished from Archaea by comparing spectral features of their lipids in the C-H region due to differences in cell envelope compositions. To confirm this, we performed validation experiments using the following four strains of archaea and bacteria: the archaeon *Sulfolobus solfataricus* DSMZ 1616^T (grown at 80 °C in 0.25 × SME medium) with glycosylated surface layer protein on its surface, the archaeon *Methanopyrus kandleri* DSMZ 6324^T (98 °C, in SME medium) with a pseudopeptidoglycan-containing cell wall covered by a proteinaceous layer, the Gram-negative bacterium *E. coli* K12 DSMZ $30\,083^T$ (37 °C, in LB medium) with a comparably less amount of peptidoglycan but large amount of lipopolysaccharides in its cell envelope, and the Gram-positive bacterium *B. atrophaeus* DSMZ 7264^T [32 °C, in TSB (tryptic soy broth) medium] with a high amount of peptidoglycan in its cell wall. We first made SR-FTIR measurements on the four archaea and bacteria strains, and results were compared and summarized in *Figure III.3-1*.

Baseline corrected and vector-normalized spectra in the C–H region between 3000 and 2800 cm^{-1} were then subjected to the multivariate PCA and then linear discriminant analysis (LDA) using MathLab (7.0). PCA and LDA were used to generate new variables (factors) that were linear combinations (that is, weighted sum) of the original variables (wavenumbers). PCA was first applied to the spectra to reduce the hundreds of absorbance intensities at different wavenumbers to just a few factors that could capture more than 95% of the variance. We typically selected seven components based on the 95% percentage of variance explained and on the spectral features of the loading plot. LDA was then applied to maximize the 'inter-class' variance over the 'intra-class' variance of the factors. We visualized the multivariate analysis results in the form of score plots (*Figure III.3-1B*, left panel) and cluster vector plots (*Figure III.3-1B*, right panels). In this study, score plots were three-dimensional plots where the first three PC-LDA components were the x-, y- and z-axes; the nearness between classes (clusters) indicates the similarity, whereas the distance between classes implies dissimilarity.

Detailed analyses (*Figure III.3-1B*, left and right panels) revealed that, in spite of the significant variations in the cell envelope (including cell wall) compositions, Bacteria can be distinguished from Archaea solely by comparing spectral features of their lipids in the C–H regions (3100–2800 cm^{-1}). As expected, bacterial membrane lipids consist of fatty acids with long alkylic ($–CH_2–$) chains which have only one to two terminal methyl ($CH_3–$) groups. In contrast, archaeal membrane lipids generally consist of branched and saturated hydrocarbon isoprene, and therefore relatively less $CH_2–$ and more $CH_3–$ groups (Mancuso et al., 1986). Our earlier study showed that the SM1 Euryarchaeon possesses a typical CH_3-rich lipid [archaeol, (Rudolph, 2003)]. In this context, the ratio of CH_2 to CH_3 could be used to detect Bacteria in an Archaea-dominated biofilm (Supplementary Figures S2A and S2B).

To confirm this observation further, we made measurements and performed spatial correlation analysis on FISH and MCR SR-FTIR images (Supplementary Figures S3A and S3B) using the image processing software ImageJ (http://www.macbiophotonics.ca/imagej/) and Manders approach of interpretation (Manders et al., 1993). Field-collected biofilm samples were labeled with Archaea-directed probes [ARCHmix; (Henneberger et al., 2006; Moissl et al., 2002)]. The samples' fluorescence and the corresponding MCR SR-FTIR images were acquired by means of a Nicolet Continuum XL infrared microscope equipped with a fluorescence attachment and a WG fluorescence cube. Although the MCR recovered image and the fluorescence image have different 'brightness', 'intensity' and effective spatial resolution (microns for infrared and hundreds of nanometers for fluorescence microscopy), Pearson's coefficient was 0.504, Manders' overlap coefficient 0.798, the co-localization coefficient M1 was 0.988 and M2 was 1.000 (Manders et al., 1993). This demonstrated that the two images were quite similar (Manders et al., 1993).

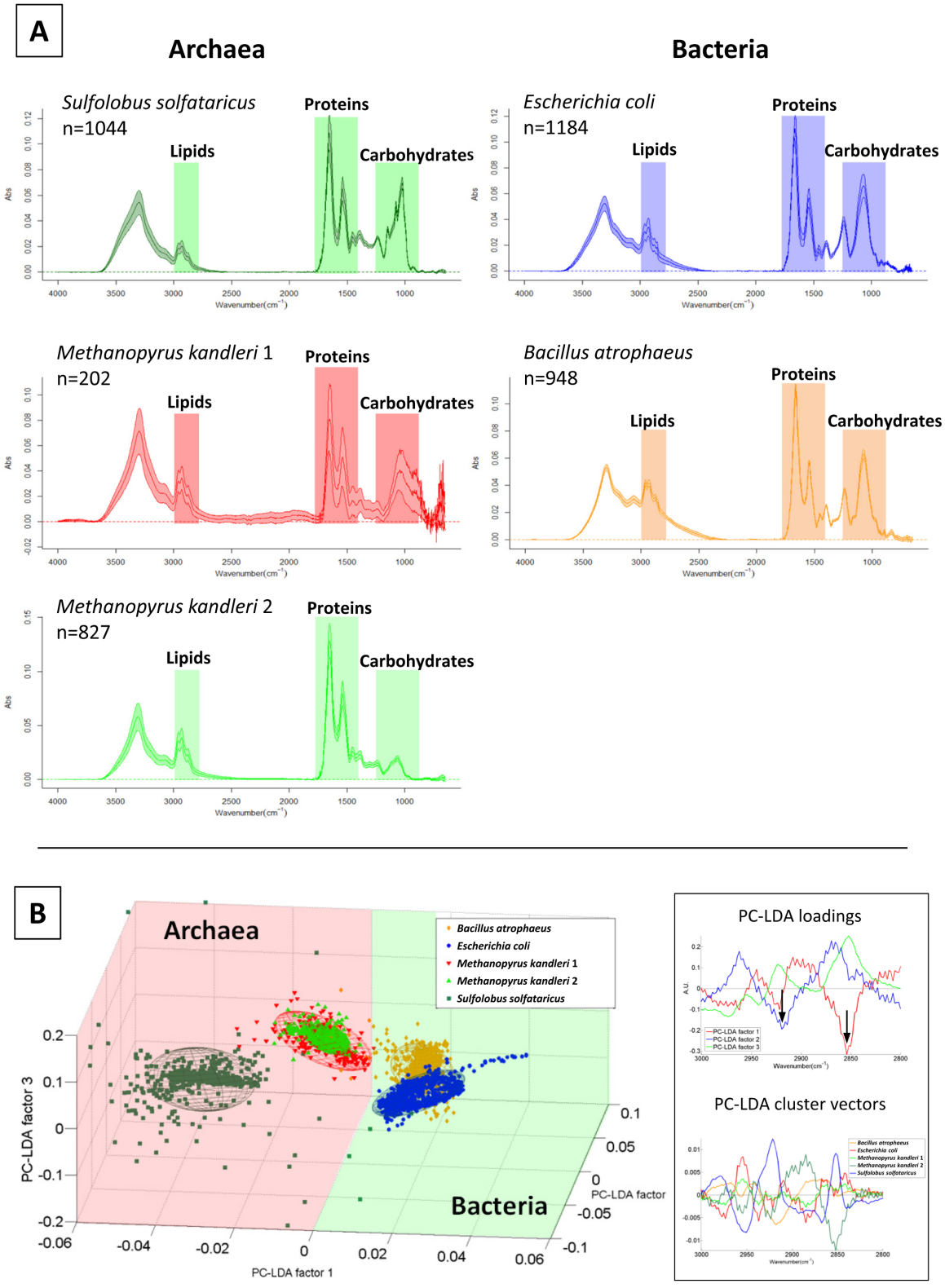

Figure III.3-1 *SR-FTIR validation experiments: comparison of reference archaea and bacteria.* ***A****) Comparison of SR-FTIR spectra of reference archaea and bacteria in the 4000–650 cm⁻¹ region reflecting individual membrane lipids and cell envelope compositional characteristics (see Materials and methods section). All spectra are mean ± standard deviation (colored area). Archaea: S. solfataricus (glycosilated surface layer); M. kandleri [pseudopeptidoglycan and proteinaceous sheath; please note, M. kandleri exhibits two types of spectra, depending on the observed accumulation of extracellular material (type 1: with extracellular material, type 2: without extracellular material)]. Bacteria: E. coli (Gram-negative cell wall), B. atrophaeus (Gram-positive cell wall). The numbers of reference spectra per species measured are given (n).* ***B****) PC-LDA of the same spectra in the CH vibration region (3000–2800 cm⁻¹). Left: three-dimensional PC-LDA score plots reveal an excellent separation of archaea and bacteria along the first PC-LDA factor; each ellipse covers an area of 95% confidence level. The three components explain 92.7% of the variance. Right: The first PC-LDA loading spectrum has two distinct peaks at 2920 cm⁻¹ and 2850 cm⁻¹ (see arrows), which are associated with CH_2 bond stretching. The corresponding cluster vector spectra reveal more specific membrane lipids composition and organization variations among the reference strains.*

Results

Using an *in situ* trapping system (Henneberger et al., 2006), fragments of the SM1-Euryarchaeon biofilm from the subsurface were collected for an in-depth characterization of the unique subsurface biotope that can be accessed through the Muehlbacher Schwefelquelle.

Oxygen concentration in the spring area revisited. Using a highly sensitive oxygen probe, a chemocline could be detected with decreasing oxygen concentrations towards the spring (Supplementary Figure S4). No oxygen could be measured in the subsurface water before mixing with the atmosphere in the streamlet, indicating a complete oxygen-free environment in the subsurface. These on-site measurements are in contrast to previous investigations that reported low amounts of oxygen in the spring water (Henneberger et al., 2006; Rudolph et al., 2004).

Dominance of the SM1 Euryarchaeon in the subsurface biofilm. FISH with Archaea-directed and SM1 Euryarchaeon-specific probes confirmed previous results showing the archaeal dominance within the biofilm. The predominance of the SM1 Euryarchaeon was additionally confirmed by using domain-specific qPCR. The ratio of archaeal and bacterial 16S rRNA gene copy numbers was 97:3 (*Table III.3-1*), which is similar to previously reported ratio of Archaea:Bacteria being 95:5 (Henneberger et al., 2006). A newly constructed clone library of archaeal 16S rRNA gene sequences generated from biofilm samples resulted in four different restriction-fragment-length polymorphism patterns after analyzing 48 clones. The dominant clone sequence (88% of all clones, IM-A1, JN861739) and two others (2%, IM-C8, JN861741; 2% IM-4–1, JN861742) showed high similarity to publicly available 16S rRNA gene sequences of the SM1 Euryarchaeon and among each other (>99%). One clone sequence (8%, IM-C4, JN861740) was closely related to the environmental clone sequence SMK5 (Rudolph et al., 2004), which was retrieved from Sippenauer Moor string-of-pearls community in 2005 (99% similarity) but shows a genetic distance of 20% to the SM1 Euryarchaeon sequence.

Table III.3-1 Quantification of archaeal/bacterial 16S rRNA and *dsrB* gene sequences in 1 ng metagenomic DNA from SM1 Euryarchaeon biofilm samples taken at two different sampling times

Biofilm replicate	Number of gene copies		
	Archaeal 16S rRNA	Bacterial 16S rRNA	*dsrB*
1	2.26E+06	6.88E+04	5.68E+03
2	3.33E+06	8.40E+04	5.46E+03
3	3.09E+06	7.49E+04	4.22E+03
Mean	2.89E+06	7.59E+04	5.12E+03

Abbreviation: dsrB, dissimilatory sulfate reductase subunit B.

The currently most sensitive method available [PhyloChip G3 16S rRNA gene microarray technology, detection limit 2 pM of 16S rRNA PCR product; (Hazen et al., 2010)] was used for characterizing the archaeal and bacterial composition in the biofilm and the spring water itself based on 16S rRNA gene analysis. Besides a comprehensive detection of Bacteria, the setup of the PhyloChip G3 technology was geared towards the identification of (also underrepresented) archaeal signatures. To avoid primer mismatches of typical, Archaea-directed primers binding to the front region of the 16S rRNA gene (Rudolph et al., 2001), a different primer set for amplification of (SM1 eury-)archaeal 16S rRNA genes was used and an adjusted bioinformatical approach for the shorter PCR amplicons was necessary. Although the SM1 Euryarchaeon was originally not included in PhyloChip G3 design (Hazen et al., 2010), we developed a method to track its abundance in our samples based on the hybridization intensity of a specific probe on the microarray. With the aid of PhyloChip technology, the SM1 Euryarchaeon was detected highly enriched in the biofilm samples (~2130% increase in abundance) compared with the spring water.

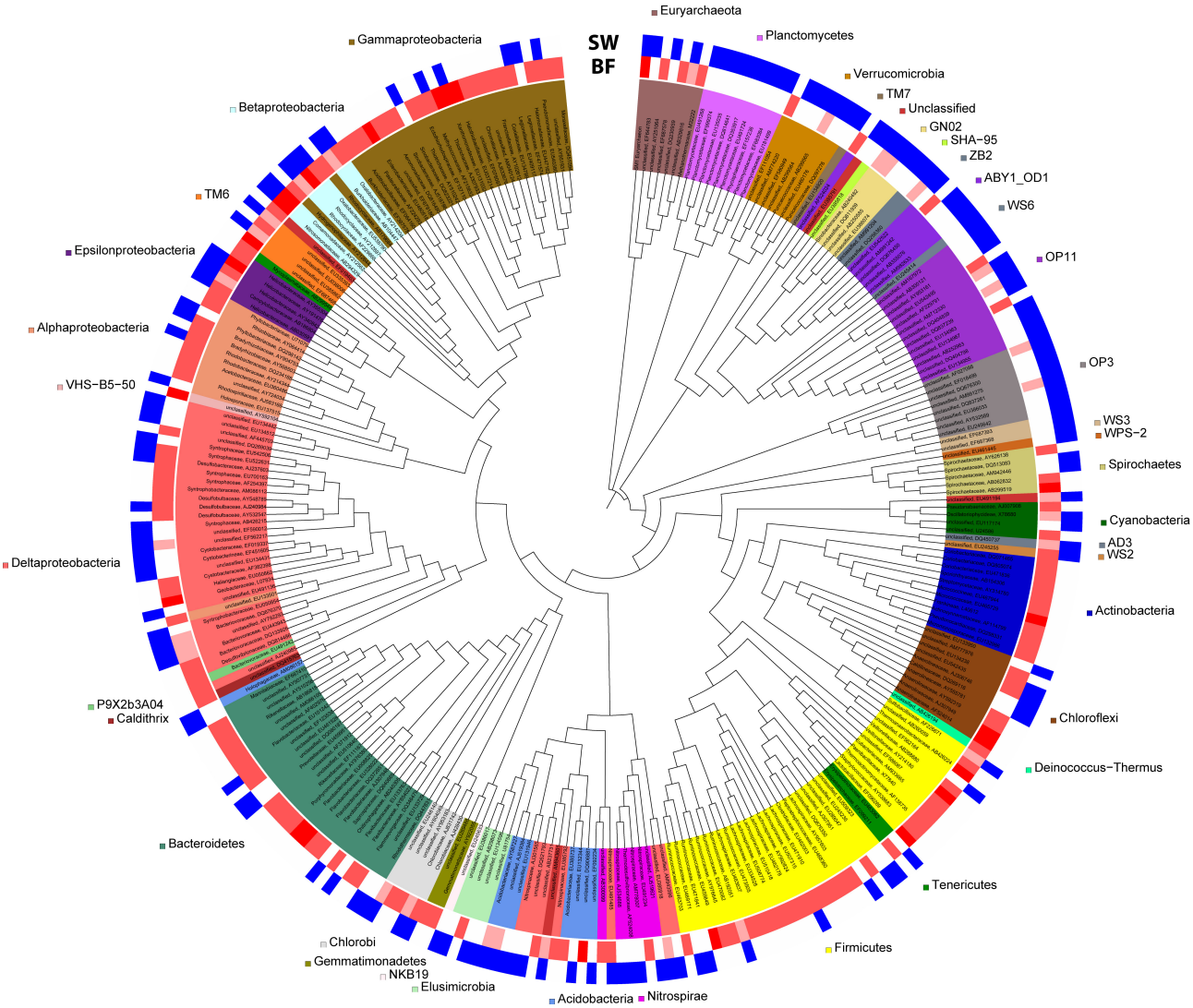

Figure III.3-2 *Difference in microbial richness between the spring water and the biofilm: presence and absence of subfamilies in spring water (SW) and biofilm samples (BF). Color intensities (red) of the biofilm samples reflect the number of times the subfamily was called present in one of the three replicates. As the SW sample was not replicated, heatmap reflects presence (blue label) or absence only. The Neighbor Joining tree was constructed with one representative OTU per subfamily (branch length is ignored). Leaf IDs give the classification on family level and the accession number of the representative OTU. Only those subfamilies that occurred in the water sample or in at least 2/3 of the biofilm replicates are shown.*

Microbial diversity in spring water and biofilm. In the spring water and biofilm microbiome, in total 4444 OTUs in 869 different subfamilies were detected by PhyloChip analyses with 10 OTUs (in 10 subfamilies) belonging to the archaeal domain in addition to the SM1 Euryarchaeon. The overall distribution of the microbial taxa on subfamily level ranged from 14% for Firmicutes to 0.1%, for example, for Aquificales. Only 36% of the subfamilies detected in the water were also present in at least one of the three biofilm replicates (*Figure III.3-2*).

The diversity of the spring water was dominated by Deltaproteobacteria (14%); however, signatures of Methanomicrobia and Thermosplasmata of the archaeal domain were also retrieved (distribution of spring water diversity in Supplementary Figure S5A). The diversity of Firmicutes, Gammaproteobacteria and Bacteroidetes increased in the biofilm, whereas members of the OP11-group and the Planctomycetes were less diverse than in the spring water. Considering the diversity of subfamilies that occurred in at least 2/3 of the biofilm samples, Firmicutes were again the most diverse taxon, followed by Gamma-proteobacteria and Bacteroidetes (Supplementary Figure S5B).

Core microbiome of the biofilm. Biofilm subfamilies detectable in 2/3 replicates were then analyzed in depth at the OTU level: If the coefficient of variation of the abundance values of a single OTU was <10% among biofilm replicates, the OTU was assumed to be non-fluctuating, and thus a constant member (potential key species) of the biofilm. Abundance values

and a detailed description of these constant OTUs (263, including the SM1 Euryarchaeon) are presented in Supplementary Figure S6. This community is considered to represent the core microbiome of the biofilm.
As OTUs such as *Thiothrix* clone sipK4 (AJ307941) and the *Sulfuricurvum* clone IMB1 (AJ307940) were found to fluctuate, they were not considered as representatives of the core microbiome. However, both of these OTUs have been identified as key species in the string-of-pearls communities at various sampling sites (Moissl et al., 2002; Rudolph et al., 2001). Eleven other string-of-pearls community related OTUs, which had been reported but not as key species, were also identified and found to be mostly fluctuating (Supplementary Table S2).

Significantly enriched OTUs in the biofilm compared with the string-of-pearls community. PhyloChip G3 analyses revealed 2139 OTUs that increased in abundance in the biofilm samples compared with the reference sample set (string-of-pearls community, data not shown). Eighty-three OTUs met the requirement of being highly significantly enriched (adjusted *P*-value 0.002), which were grouped into 44 subfamilies and are displayed in *Figure III.3-3*. The OTU with the greatest increase in abundance (accession number of representative sequence AJ831749; increase in abundance: 4559) was also the OTU with the most significant *P*-value of 4.60E-06. The representative sequence grouped this OTU in the Deltaproteobacteria, genus *Desulfobacula*. However, also OTUs of other phyla and genera were detected as significantly enriched, among those many Chloroflexi and Spirochetes.

Detection of SRB via FISH and correlation with PhyloChip data. The presence and amount of (potentially) SRB in the biofilm was further confirmed by FISH with two different (sets of) probes targeting bacterial sulfate reducers: bacterial sulfate-reducer probe SRB385 and Delta495 probe mix. Each approach was backed up by probes directed for Bacteria (Eub 338/I, Texas Red) or Archaea (ArchMix, RG), and DAPI staining or combinations thereof. The specificity of FISH experiments was confirmed by using appropriate controls and a nonsense probe (NONEUB338), which showed no signal when applied to biofilm samples.
The morphology of the bacteria in the biofilm was diverse, ranging from single cocci to aggregates, filaments, oval-, rod- and helix-shaped. The percentage of bacteria was estimated at 5%, confirming results from qPCR and previous studies (Henneberger et al., 2006). Interestingly, 85.4% (±4.7% s.d.; 15 biofilm samples analyzed) of cells stained with the bacterial probe also exhibited signals for the SRB385 probe (*Figure III.3-4a*). This percentage of SRB was confirmed by the usage of the Delta495 probe mix, which revealed an amount of 89.2% (±0.9% standard deviation; four biofilm samples analyzed) SRB (*Figure III.3-4b*).
In order to correlate FISH data with enriched OTUs detected by PhyloChip analysis, the theoretical coverage of the Bacteria- and sulfate-reducer directed FISH probes was analyzed *in silico* (*Figure III.3-3*). All probes (SRB385, Delta495 probe mix and EUB 338/I) showed theoretical coverage of the target group (SRB), and therefore confirmed our FISH results. Interestingly, 29.4% of all bacterial cells that did not stain with the Delta495 probe mix exhibited a typical Spirochaeta-like morphology. This genus was also found to be highly enriched in the biofilm, but whose 16S rRNA reveals >2 mismatches for the Delta495 probe mix (*Figure III.3-3*).
Based on the high percentage of the SRB385 probe and the Delta495 probe mix stained bacteria, the PhyloChip and *in silico* FISH analysis it can be concluded that the major part of bacteria in the biofilm can be affiliated to members of the Deltaproteobacteria, most likely to one specific, enriched OTU (genus *Desulfobacula* AJ831749). Cultivated members within the genus *Desulfobacula* were described as oval-shaped; bacteria with this morphology were positively stained with SRB and Delta495 probes and formed aggregates in the biofilm (*Figure III.3-4b*).

Detection of dsrB genes in biofilm samples. In order to further prove the presence of SRB and their metabolic capability, qPCR with *dsrB*-directed primers was performed. We were able to specifically detect the presence of genes encoding *dsrB* and to quantify their amount (*Table III.3-1*). The abundance of detectable *dsrB* genes in biofilm samples allowed the conclusion that these signatures were derived from bacteria and not from the dominant SM1 Euryarchaeon (three-log difference in archaeal 16S rRNA and *dsrB* gene abundance). Moreover, the one-log difference of bacterial 16S rRNA and *dsrB* genes can be attributed to the fact that ribosomal genes can have up to 15 copies per genome (Klappenbach et al., 2001; Lee et al., 2009), whereas *dsrB* genes generally appear once (Heidelberg et al., 2004).

A clone library generated from the *dsrB* amplicons showed four different OTUs of *dsrB* genes belonging to the Deltaproteobacteria cluster, whereas one OTU was dominant (Accession no. JX515394: 33 clones; Accession no. JX515395: 11 clones; Accession no. JX515396: 3 clones, Accession no. JX515397: 1 clone). The coverage of the library was determined as 98%.

Metabolic activity of SRB in the biofilm. A combination of CTC staining and FISH analysis showed an overlap of signals from CTC and SRB-directed FISH probes (Delta495 probe mix) in biofilm samples that were incubated in spring water anaerobically (Supplementary Figure S7). The formation of CTC-formazan precipitates can be attributed to biological redox reactions, for example, respiratory electron transport, and thus provide evidence for the metabolic activity (Stellmach, 1984; Stellmach and Severin, 1987; Yoshida and Hiraishi, 2004).

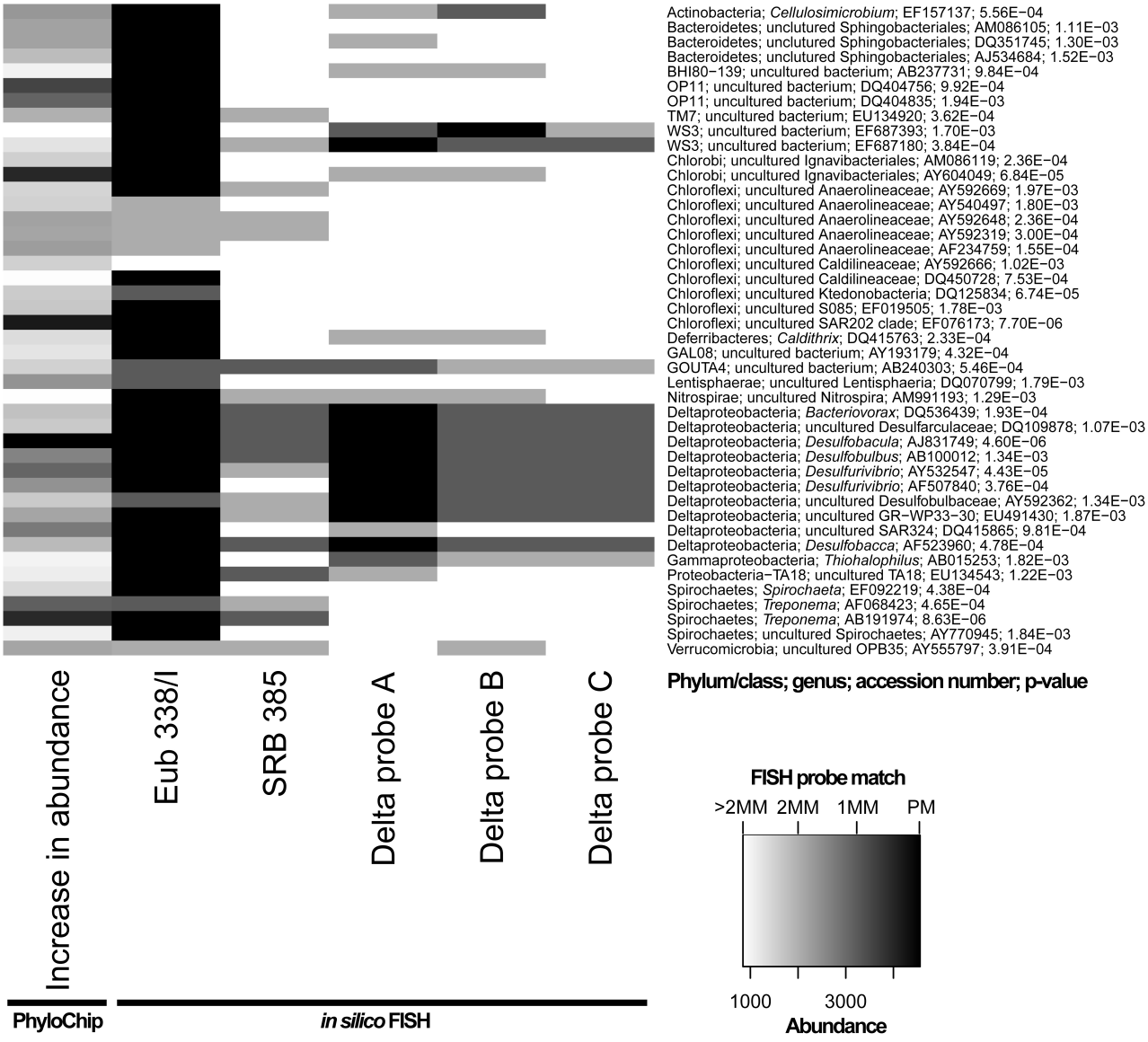

Figure III.3-3 *Significantly enriched OTUs (one representative of each subfamily) in the SM1 Euryarchaeon biofilm and in silico FISH-probe match. Heatmap of OTUs that increased highly significantly (P<0.002) in biofilm compared with string-of-pearls community samples and were called present in at least one of the samples (first column). Probes used for FISH experiments in this study were in silico matched to representative sequences of the enriched OTUs using the ARB software package. The theoretical coverage of the FISH probes is displayed in columns 2–5; the decreasing heatmap intensity reflects the number of mismatches of each probe per OTU (MM=mismatch, PM=perfect match).*

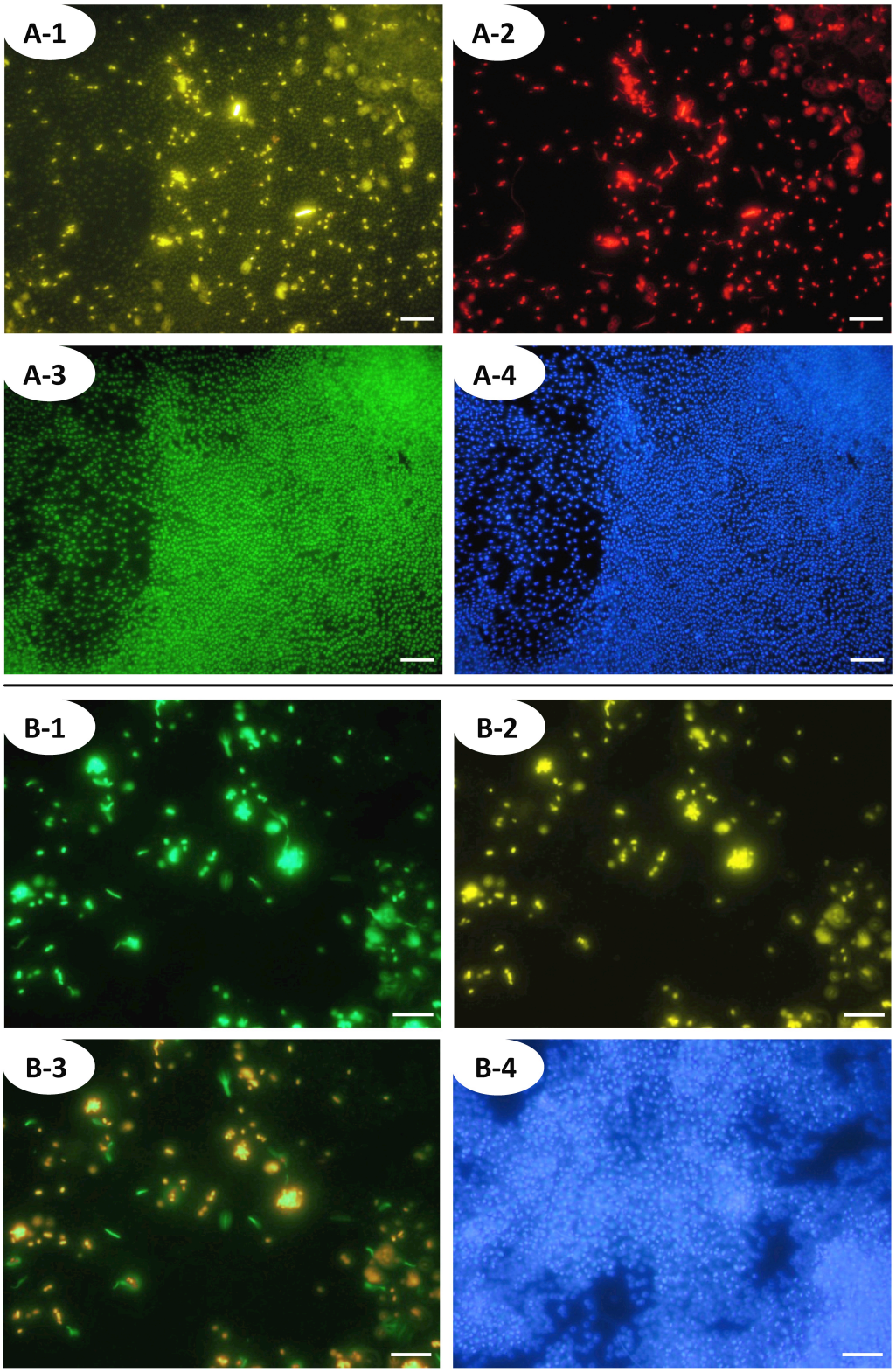

***Figure III.3-4 A**) Ternary FISH analysis of the subsurface biofilm with SRB-, Bacteria and Archaea-specific probes. Analysis reveals a dominance of archaeal cocci (SM1 Euryarchaeon) and of SRB-385-stained bacteria in the bacterial minority: ~85% of the detected Bacteria revealed a signal with the sulfate-reducer specific probe. Scale bars=10 µm. **A1**: biofilm, FISH-stained with probe SRB385 CY3 (targeting SRB, yellow). **A2**: same detail, stained with probe EUB 338/I Texas Red (targeting Bacteria, red). **A3**: same detail, stained with probe mixture ArchMix RG (targeting Archaea, green). **A4**: same detail, reference-stained with DAPI (blue). **B**) biofilm sample FISH-stained with SRB-directed Delta495 probe mix. The overwhelming majority of bacteria in the biofilm showed signals with the Delta495 probe mix (89.2%). Scale bars=10 µm. **B1**: probe 338/I RG (targeting Bacteria, green). **B2**: same detail, stained with probe mix Delta495 (targeting SRB, yellow). **B3**: overlay of details B1 and B2. **B4**: same detail, reference-stained with DAPI (blue).*

SR-FTIR measurements of Bacteria, Archaea, and metabolic intermediates distributions in biofilms. The high brightness of SR-FTIR spectromicroscopy enabled us to identify the presence of Bacteria, Archaea and a number of metabolic intermediates at a spatial resolution between 2 and 10 µm. Biofilm samples were first examined with crossed polarized microscopy and fluorescence microscopy because the biofilm fraction often exhibits visually interesting biogeochemical structures (Supplementary Figure S8). Then the spatial distribution of molecular composition and possible metabolites in the biofilm were analyzed by SR-FTIR spectromicroscopy. *Figure III.3-5A.I* shows examples of a range of typical SR-FTIR spectra collected on the biofilm samples (*Figure III.3-5A.II*, white circles in bright field). A striking feature was that these spectra, although obtained at locations merely several tens of micrometer apart (see circles in *Figure III.3-5A.II*), contained distinctly different signatures known to be associated with organic and inorganic markers typical of biogeochemical systems (*Table III.3-2*). Spatial distributions of the infrared absorption intensities (from univariate analysis) of these molecular markers are shown in *Figure III.3-5A.II*. Notice that the infrared absorption intensity ratio of CH_2 to CH_3 was ~30% higher in the biofilm regions occupied by large, filamentous-shaped, *Beggiatoa*-like bacteria, compared with the surrounding Archaea-dominated area.

MCR analysis confirmed that Bacteria can be distinguished from Archaea by their spectral features (*Figure III.3-5B.I* versus *Figure III.3-1A and B*; *Figure III.3-5B.II* versus Supplementary Figure S3, Supplementary Figure S2 versus *Figure III.3-1B.I*). Furthermore, specific metabolites as well as biogeochemical materials were found associated with these prokaryotic groups (*Figure III.3-5B)*. For example, in *Figure III.3-5A.II*, a combined univariate and MCR SR-FTIR analysis revealed strikingly overlapping infrared signals of organic sulfate products (R-S=O) and carbonate minerals with Bacteria-rich areas. Similar results were observed in other samples collected during this field experimental period (Supplementary Figures S9 and S10). This implies the presence of microscale mixtures of bacteria that are involved in subsurface sulfur and carbon turnover.

Table III.3-2 Tentative band assignments of the fundamental vibrational modes used in SR-FTIR spectromicroscopy

Molecules	Frequency (cm^{-1})	Assignment	References
Calcium carbonates	~1798	Multiphonon band	(Beniash et al., 1997; White, 1974)
	~874, ~845, ~710	Coupling between CO_3^{2-} groups in adjacent layers of calcite	(Beniash et al., 1997; Huang and Kerr, 1960; White, 1974)
Carbohydrates	1200 - 900	C–O–C, C–O ring vibrations of carbohydrates	(Naumann, 2000)
Clay	~3695, ~3620	OH groups in the inner-surface of kaolinite	(Ledoux and White, 1964)
Fatty acids / lipids	~2959, ~2872	C–H asym and sym stretching modes of $-CH_3$ in fatty acids	(Naumann, 2000; Sinclair et al., 1952; Whittaker et al., 2003)
	~2920, ~2852	C–H asymmetric and symmetric stretching modes of $>CH_2$ in fatty acids	(Naumann, 2000; Sinclair et al., 1952; Whittaker et al., 2003)
	~1468, ~1420	C–H deformation of $>CH_2$	(Naumann, 2000; Parker, 1983)
	~1378	C–H deformation of $-CH_3$	(Parker, 1983)
Proteins	1695 - 1600	Protein amide I	(Naumann, 2000)
	1580 - 1510	Protein amide II	(Naumann, 2000)
Sulfate: inorganic	~1150, ~1130, ~1105	S–O of the SO_4^{2-}	(Adler and Kerr, 1965; Peak et al., 1999; Smith, 1998)
	800 - 850	C-O-S vibration of sulfate in carbonates	(Takano, 1985)
Sulfate: organic	~ 1250, ~1240	S=O stretching vibration of sulfate ester	(Goren, 1970; Mayers et al., 1969; Mohsen et al., 2007; Percival and Wold, 1963)

Abbreviation: SR-FTIR, synchrotron radiation-based Fourier transform infrared. Assignments were used for identification of bio- and geochemical molecules (of dominance) measured in the investigated biofilms.

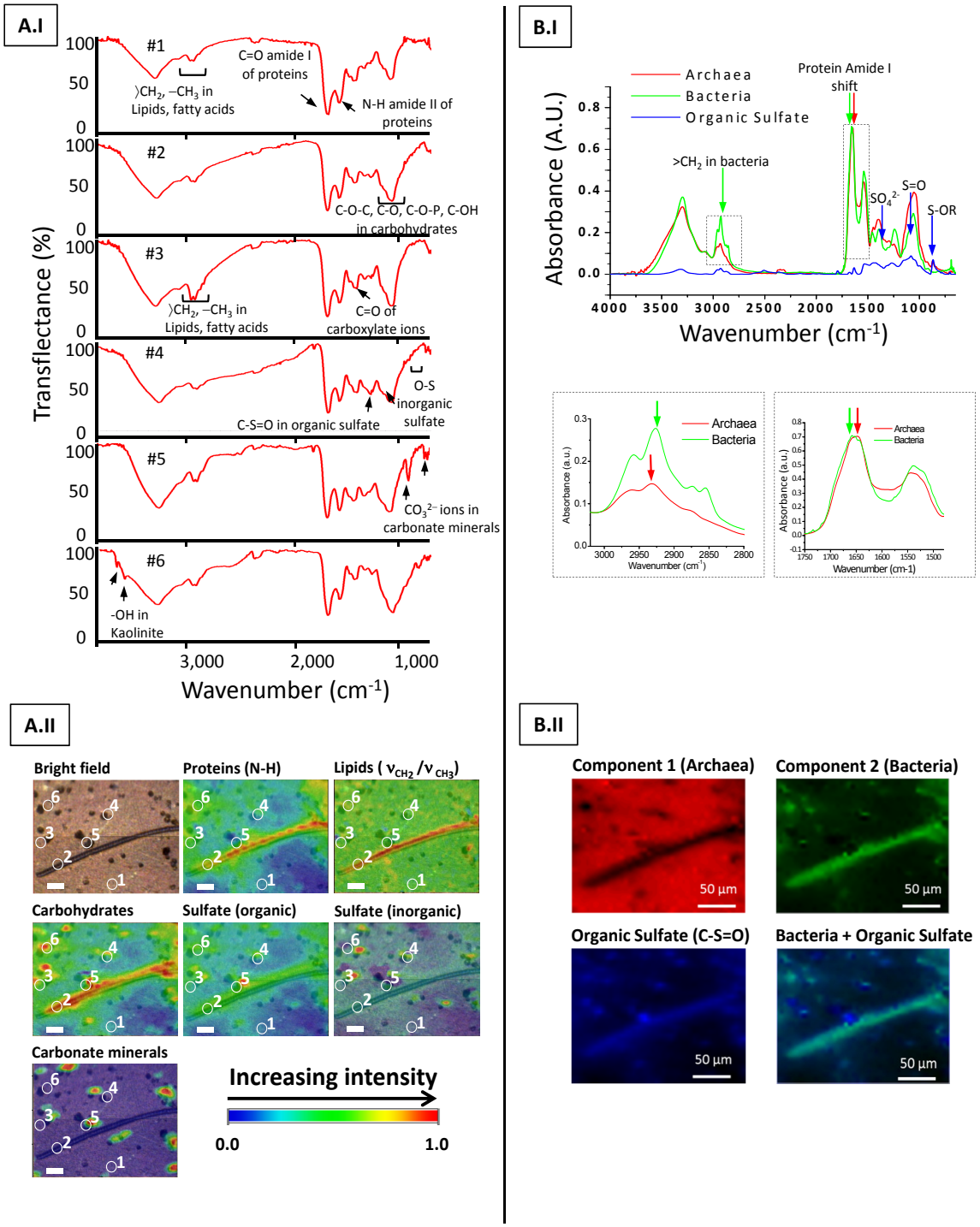

***Figure III.3-5 A**) SR-FTIR images of Bacteria in the Archaea-dominated biofilm. SR-FTIR images (220 µm by 180 µm) showing the distribution of microorganisms and biogeochemical products in an Archaea-dominated biofilm. (A.I and A.II) Distribution heatmap (from univariate analysis) of the relative abundance of total proteins (based on the peak area centered at ~1548 cm^{-1}), of bacterial lipids (the ratio of the peak area of CH_2 centered at ~2852 cm^{-1} to the peak area of CH_3 at ~2872 cm^{-1}), carbohydrates (the peak area centered at ~1089 cm^{-1}), sulfur/carbon biochemical cycling products (S=O from organic sulfate products, centered at ~1240 cm^{-1}, S=O of inorganic sulfate centered at ~1130 cm^{-1} and CO_3^{2-} groups of carbonate minerals in the 880–840 cm^{-1} region. The OH of clay centered at ~3695 cm^{-1} is not shown here). The white circles with numbers (1–6) in the bright field and in the SR-FTIR images (A.II) correspond to the transflectance spectra (A.I). The circles represent pixels where the spectra were recorded. Note: Filamentous bacterial structures in the biofilm were rarely observed but specifically presented here in order to illustrate the lipid signatures of Bacteria and Archaea (for more samples please see Supplementary Figure S9). Scale bars=25 µm. **B**) Multivariate curve resolution analysis to differentiate Archaea and Bacteria. B.I) Spectra of the three components extracted from the MCR, in red component 1 (Archaea), in green component 2 (Bacteria), in blue component presenting sulfate spectral features, with arrows pinpointing the spectral markers used in the analysis, in the panels below highlighted the spectral region of lipids, important region for the distinction of Bacteria and Archaea since their different membrane composition, and protein region where a shift is observable in Amide I band, index of a different protein content in Bacteria and Archaea. (B.II): Relative concentration images (220 µm by 180 µm) of Archaea (component 1) and Bacteria (component 2) recovered by the MCR analysis and the chemical distribution maps of organic sulfate (C-S=O) in blue. Merging the relative bacterial concentration image (in green color) with the organic sulfate distribution map (in blue) reveals the co-localization of bacteria and organic sulfate. Scale bars=50 µm.*

Discussion

Life in the subsurface is highly diverse and comprises an enormous fraction of Earth's biomass. However, the microbial community living in this extreme environment remains largely mysterious, as subsurface biotopes are hardly accessible, which makes it also difficult to understand ongoing geochemical processes in these environments (Onstott et al., 2010). The Muehlbacher Schwefelquelle, however, provides an extraordinary window to the subsurface and allowed the discovery of a highly unusual, Archaea-dominated microbial community (the SM1 euryarchaeal biofilm), which is continuously being washed up from the subsurface and can be harvested from the spring water. As the dominant SM1 Euryarchaeon still resists efforts to be cultivated and metabolically understood, this study focused on the bacterial minority thriving in this type of biofilms. Using a combined approach of molecular techniques and SR-FTIR spectromicroscopy, we demonstrated that the interplay between the underrepresented bacterial fraction and geologically important chemicals could be analyzed in order to obtain insights into a possible ecological role of this extraordinary microbial community.

The dominance of one specific archaeon, the SM1 Euryarchaeon, was revisited and confirmed in this study by qPCR techniques and FISH, proving the constancy of this subsurface system over several years (Henneberger et al., 2006). The sensitive PhyloChip technology also confirmed the abundance of the SM1 Euryarchaeon, but additionally identified the presence of other archaea disproving the initial statement of an 'archaeal mono-species biofilm' (Henneberger et al., 2006). Nevertheless, because none of these alternate archaea was visualized either in FISH or in the archaeal clone library, it can be assumed that they represent only a very minor fraction of the biofilm.

We demonstrated that the spectral features of membrane lipids can be used to distinguish Archaea from Bacteria even in complex samples without using either a MS- or a nucleic acid-based approach (Elvert et al., 2000; Sprott, 1992; Sturt et al., 2004). Furthermore, we also could use SR-FTIR to map the distribution of biogeochemical compounds, and to relate this molecular information to certain dominant ecological functions of even underrepresented microbial groups.

Raman microspectroscopy has often been used to characterize spatial distribution and molecular composition of biological samples (Beier et al., 2010; Hall et al., 2011; Li et al., 2012; Wagner, 2009). However, to date, Raman microscopy has often been used together with FISH (Raman-FISH) to differentiate microbial populations such as Bacteria and Archaea (Huang et al., 2007). On the contrary, SR-FTIR does not require cell labeling. SR-FTIR is also non-destructive, and therefore allows additional *in situ* studies of chemical composition changes in microbes on the same sample (Holman et al., 2010). As demonstrated in this study SR-FTIR spectromicroscopy imaging could associate the distribution of Archaea and Bacteria with biogeochemical compounds, giving us the opportunity to gain a more in-depth insight of the underpinning biogeochemical processes. In *Figure III.3-5*, for example, large, filamentous bacterial cells were observed along with increases in organic sulfate intensities, which suggests that these bacteria could belong to an sulfur-accumulating and -oxidizing bacterium such as *Beggiatoa* (Larkin and Strohl, 1983), a genus also identified in the biofilm core microbiome.

In other biofilm samples, such as those presented in Supplementary Figures S9 and S10, a majority of the areas that exhibited infrared spectral signatures of bacterial cells coincides with signals that are indicative of an accumulation of organic sulfate. The increasing sulfate signals could imply the presence of compounds such as adenosine-5′-phosphosulfate, 3′phosphoadenosine-5′phosphosulfate or sulfolipids (Goren, 1970). However, adenosine-5′-phosphosulfate is a typical intermediate of metabolically active either assimilatory or dissimilatory SRB and sulfur-oxidizing bacteria.

qPCR assays were able to detect a high amount of *dsrB* genes of Deltaproteobacteria, which were also identified by the PhyloChip G3 technology to be highly enriched in the biofilm. In FISH analyses of multiple biofilms a vast majority of the bacteria showed a positive signal after hybridization with two different (sets of) SRB-directed probes (the 385 probe and the Delta495a/b/c probe mix). This observation was supported by the CTC-FISH assay, which showed metabolic activity of bacteria stained with the Delta495 probe mix (see Supplementary Figure S7). These investigations confirmed that the bacterial microbiome of the SM1 Euryarchaeon biofilm is comprised mostly of Deltaproteobacteria, involved in sulfate reduction.

The fact that little sulfate signals were detected in the Archaea-rich regions implies two possible scenarios for the samples taken. In the first scenario, the supposed sulfate-reducing SM1 Euryarchaeon might be alive but metabolically inactive, having already reduced most of the sulfate compounds in its direct vicinity of the biofilm. In the second scenario, the SM1 Euryarchaeon might not be capable of sulfate reduction, a conclusion which is in stark contrast to the previous hypothesis (Moissl et al., 2002).

A number of metabolic pathways of SRB-associated archaea have already been reported in literature. For instance, in the AMO consortium SRB have a key role for archaeal, anaerobic methane oxidation (Orphan et al., 2001). However, it still remains unclear if the SM1 Euryarchaeon is capable of methane-oxidation or methanogenesis, or if it performs a completely different metabolism. Nevertheless it can be speculated, that a classical methanogen would quickly be outcompeted by SRB for hydrogen or organic substrates in sulfate-rich, anoxic environments such as the Muehlbacher Schwefelquelle (Lovley and Klug, 1983). Previous investigations have failed to detect F_{420}, a key co-enzyme for methanogenesis, showing no positive amplification of the according gene, nor a positive chemical detection based on chromatography (Moissl et al., 2003).

Possible metabolic functions of the SM1 Euryarchaeon remain speculative but may be responsible for the environmental success of this organism. As the SM1 Euryarchaeon is currently the only known archaeon to absolutely predominate one specific biotope, combined with its appearance in hot spots in Europe and maybe even beyond (Rudolph et al., 2004), a larger (ecological) role can be assumed, which is currently still mysterious. However, a metagenomic study of the biofilm is currently performed, for which the knowledge about the microbial diversity is an important and very helpful prerequisite. This approach may reveal the metabolic capabilities of the SM1 Euryarchaeon in the biofilm.

Although a broad diversity of microbes is detectable in the Muehlbacher Schwefelquelle biotope, the accumulation of SRB, which represent the overwhelming majority of the minor bacterial part, appears to not be an accident; rather it is clear that these bacteria provide a valuable function within the biofilm as their presence in the biofilm was monitored as its discovery more than 8 years ago. However, the open question is if and how the SM1 Euryarchaeon influences the (bacterial) diversity in the biofilm. Does it—as it seems to be obvious for living together with (selected) filamentous sulfide-oxidizers in surface waters—actively recruit SRB to the biofilm, or is this phenomenon a passive enrichment? How and why does the SM1 Euryarchaeon switch from biofilm to string-of-pearls community status and are transition states detectable? These and many more questions will have to be answered in future studies and promise astonishing insights into this fascinating natural archaeal system.

Acknowledgments

Technical assistance by Lauren Tom as well as review and discussion provided by Robert Huber and Reinhard Wirth are much appreciated. Work at University of Regensburg was performed under the DFG grant MO19773–1 given to Christine Moissl-Eichinger. Phylogenetic work at Lawrence Berkeley National Laboratory was performed under the auspices of the U.S. Department of Energy under contract no. DE-AC02–05CH11231. The SR-FTIR and associated imaging work were performed under the Berkeley Synchrotron Infrared Structural Biology (BSISB) Program and the Subsurface Science Scientific Focus Area funded by the U.S. Department of Energy, Office of Science and Office of Biological and Environmental Research through contracts DE-AC02–05CH11231. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract no. DE-AC02–05CH11231. The authors are grateful to PreSens (Germany, Regensburg) for providing the oxygen dipping probe PSt6 and the Fibox 3, LCD trace. Alexander J Probst was supported by the German National Academic Foundation (Studienstiftung des deutschen Volkes).

Supplementary information

Supplementary information can be found online http://www.nature.com/ismej/journal/v7/n3/extref/ismej2012133x1.pdf or on the supporting DVD.

4. Coupling genetic and chemical microbiome profiling reveals heterogeneity of archaeome and bacteriome in subsurface biofilms that are dominated by the same archaeal species

Alexander J Probst[1,2,3,7], Giovanni Birarda[2,7], Hoi-Ying N Holman[2], Todd Z DeSantis[3], Gerhard Wanner[4], Gary L Andersen[2], Alexandra K Perras[1], Sandra Meck[1], Jörg Völkel[5], Hans A. Bechtel[6], Reinhard Wirth[1], and Christine Moissl-Eichinger[1]

[1]Institute for Microbiology and Archaea Center, University of Regensburg, Regensburg, Germany; [2]Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, USA; [3]Department for Bioinformatics, Second Genome Inc., South San Francisco, USA; [4]Department of Biology I, Biozentrum, LMU Munich, Planegg-Martinsried, Germany; [5]Department of Geomorphology and Soil Science, Technische Universität München, Center of Life and Food Sciences Weihenstephan, Freising, Germany; [6]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[7]These authors contributed equally to this work.

Correspondence: C Moissl-Eichinger, University of Regensburg, Department for Microbiology and Archaea Center, Universitaetsstrasse 31, 93053 Regensburg, Germany. E-mail: christine.moissl-eichinger@ur.de

Publication information:
PLOS ONE, Received: 24 February 2014, in review with subject Category “Microbial ecology and functional diversity of natural habitats”

Abstract

Earth harbors an enormous portion of subsurface microbial life, which remains mainly unexplored due to the difficult access to samples. The unique hydrogeological conditions of two vicinal, sulfidic springs in southeast Germany provide accessible windows into the microbial and molecular diversity of subsurface biofilms dominated by the same uncultivated archaeal species called SM1 Euryarchaeon. Although both springs are fed by one deep groundwater current and have similar physical and chemical parameters, our multidisciplinary approach revealed that site-specific hydrogeological conditions altered the microbiome at various levels, from the community profile down to the strain level, and may even create different ecological niches for the biofilm-forming archaea. The analyses of infrared imaging spectra demonstrated great variations in archaeal membrane composition, suggesting different SM1 euryarchaeal strains at both aquifer outlets. This observation is supported by ultrastructural and metagenomic analyses of the archaeal biofilms. However, at 16S rRNA gene level, PhyloChip G3 DNA microarray detected similar biofilm communities for archaea, but not for bacteria. Although the biofilms showed an enrichment of different deltaproteobacteria, their function in sulfate-reduction appeared to be congruent. Consequently, the biofilms revealed striking differences due to hydrogeological variations despite their appearance at similar locations and dominance by the same archaeal species. These various facets of microbiome differences could only be obtained by the pioneering approach of coupling multivariate statistics of synchrotron-based infrared profiling to 16S rRNA gene amplicon analysis. The results of this communication provide deep insight into the dynamics of subsurface microbial life and warrant its future investigation with regard to metabolic and genomic analyses.

Introduction

The subsurface biosphere harbors an enormous portion of the Earth's microbiome. It is estimated, that up to 2.9 x 10^{29} and 2.2 x 10^{30} prokaryotic cells reside below the surface layer in marine habitats and terrestrial sediments, respectively (Whitman et al., 1998). Sampling, and thus exploration of the subsurface microbiomes by deep drilling is difficult, since each sample is subject to a possible contamination by surface microorganisms (Whitman et al., 1998). Currently, subsurface biotopes remain poorly understood as microbial and biogeochemical "dark matter" but have substantial contribution to carbon, nitrogen and sulfur cycling (Ulrich et al., 1998; Whitman et al., 1998; Wrighton et al., 2012). However, important windows to the subsurface are provided by aquifers and their natural and artificial springs (Castelle et al., 2013; Wrighton et al., 2012), delivering possibly 10^{3}-10^{6} prokaryotic cells/ml to the surface (Kristjansson and Stetter, 1992; Whitman et al., 1998). Although sulfidic springs are rather rare [10% of all terrestrial aquifers; (Palmer, 1991)], they contain excellent energy sources for subsurface and also surface life: Once mixed with oxygen as terminal electron acceptor nutrients from sulfidic subsurface aquifers can lead to high amounts of biomass in the outflow region (Engel et al., 2003; Engel et al., 2004; Koch et al., 2006; Moissl et al., 2002; Moissl et al., 2003; Rudolph et al., 2004; Rudolph et al., 2001). These biomasses, which are mostly complex microbial communities such as bacterial or archaeal/bacterial biofilms, have been the focus of many studies, yet the oxygen-free subsurface environment of sulfidic springs is lacking information concerning its biodiversity and variation over geographical location (Probst et al., 2013b).

In southeast Germany near Regensburg, sulfur springs rise out of the subsurface karst system in the Jurassic carbonate settings. Due to the hydrogeological conditions, atmospheric oxygen mixed with the cold (~10°C), anoxic sulfidic groundwater leads to a sudden increase in biomass and the appearance of the microbial "string-of-pearls community", observed in two distinct spring areas, the Sippenauer Moor and the Mühlbacher Schwefelquelle. In the "pearls", the uncultivated, phylogenetically deep-branching SM1 Euryarchaeon resides, surrounded by sulfur-oxidizing bacteria [*Thiothrix* sp., site Sippenauer Moor, (Moissl et al., 2002; Rudolph et al., 2001), for graphical illustration the reader is referred to *Figure III.4-1*]. This specific partnership of microbial members of two domains of life is highly specific and stable, and a syntrophic interaction is hypothesized (Morris et al., 2013). Although an inter-species sulfur-cycle between the archaea and the bacteria (sulfur-oxidizers) had been proposed (Moissl et al., 2002), no evidence for sulfate-reduction by the SM1 Euryarchaeon could be collected during a recent study (Probst et al., 2013b). A detailed characterization of the archaeal/bacterial community revealed a remarkable trait of the SM1 Euryarchaeon: the *hami* (singular *hamus*). These cell surface appendages are on average 2 µm long, occur as hundreds per cell and consist of a barbwire structure ("prickle region") with a nano-grappling hook at the distal end (Moissl et al., 2005b). The *hami* are a unique feature of the SM1 Euryarchaeon and considered a potential biomarker since they have never been observed for any other organism.

In the subsurface of the Mühlbacher Schwefelquelle, the SM1 Euryarchaeon was found to form an almost pure biofilm (Henneberger et al., 2006), consisting of a dense network of cells mediated by the *hami*. Such biofilm pieces are constantly washed up from deeper Earth layers and can be harvested at the spring outflow. In the biofilm, the archaea are associated with a minor bacteriome, dominated by sulfate-reducing bacteria (Probst et al., 2013b). The constant predominance of the SM1 Euryarchaeon (>95%) in the subsurface biofilms was demonstrated by different methods and in hundreds of samples taken between 2005 and 2013 from the Mühlbacher Schwefelquelle (Henneberger et al., 2006; Probst et al., 2013b). Minor investigations also included samples from the Sippenauer Moor, where the appearance of these subsurface biofilms was also observed but not further documented (Henneberger et al., 2006). Both biofilm systems provide a window to subsurface biotopes and are a model system for cold-loving archaea and biofilm-forming archaea as well as subsurface research in general (Probst et al., 2013b). What is missing in the current body of literature is a comparative study on the biofilms that can be harvested from the two different sampling sites in order to shed light onto the biodiversity variation of microbial subsurface life over geographical location.

In this study, we investigated the background of the hydrogeology of the two sulfidic springs that appeared to be supplied by the same deep waterflow. We consequently investigated microbial differences of subsurface biofilm samples from both sampling sites and specifically focused on the variation in the microbial composition, biochemical properties, the surface ultrastructure and fingerprinting of *hamus* gene occurrence (Moissl et al., 2005b). Empirically determined operational taxonomic units (eOTUs) derived from PhyloChip G3 data were used for microbial community profiling on 16S rRNA gene level. To add an extra dimension to the knowledge of the biochemistry of the SM1 Euryarchaeon biofilms, we applied

multivariate statistics to the chemical imaging data acquired by means of synchrotron radiation-based Fourier transform (SR-FTIR) spectromicroscopy.

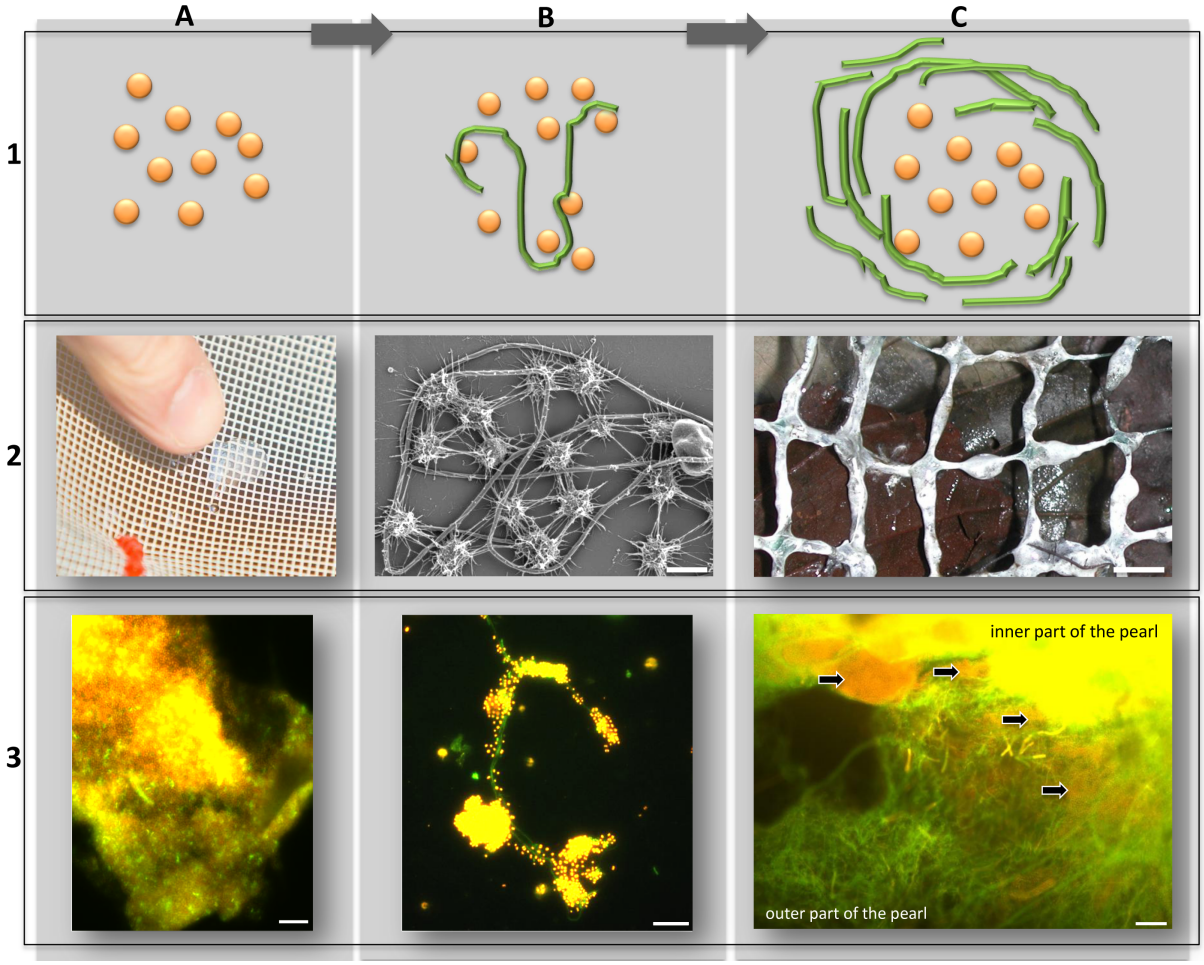

Figure III.4-1 *The conversion of biofilm to string-of-pearls community in the spring water originating from the subsurface.* ***A****) Biofilm;* ***B****) Intermediate transition state;* ***C****) String-of-pearls community;* ***Row 1****) Schematic drawings. Orange: SM1 euryarchaeal cocci, Green: Filamentous, sulfide-oxidizing bacteria.* ***Row 2****) Photographs and scanning electron micrograph (2B) of different stages.* ***Row 3****) FISH images of different stages [for MSI samples please see Probst et al., 2013b; Archaea orange (CY3), Bacteria green (RG)]. A: SM-BF, showing high dominance of Archaea. B: Attachment of archaea to filamentous bacteria. C: String-of-pearls communities with large archaeal colony and bacterial mantle. Arrows point to archaeal microcolonies, manteled by filamentous bacteria. It is proposed that attachment of SM1 Euryarchaeota to filamentous bacteria (B) mediates the transition from biofilm (A) to the string-of-pearls community (C). Scale bars: A3: 10 µm, B2: 1 µm B3: 10 mm, C3: 25 µm.*

Materials and Methods

The hydrogeology of Sippenauer Moor (SM) and Mühlbacher Schwefelquelle (MSI). The sulfur springs at the Sippenauer Moor (SM) near Kelheim (Lower Bavaria) rise out of the subsurface karst system developed in the Jurassic carbonate setting (Abele, 1950). To the contrary, the Mühlbacher Schwefelquelle at Isling (MSI) near Regensburg (Upper Palatonia) is not a natural spring (map in Fig. S1). It is a well drilled to a depth of 36.5 m in the year 1925, but has never been used because of a strong sulfur odor (information provided by REWAG Regensburger Energie- und Wasserversorgung AG & Co KG, the electricity and water supply institute at Regensburg, Germany). The MSI site is situated in the transition area of a river terrace of the Danube from pre-Eemian times (Riss glaciation), covered with Wuermian Loess and Loess Loam. The drill log from 1925 notes the 2.6 m thick Loess layer, the interlayering fluvial sediments and the Quaternary base, which overlies the top of the sedimentary Cretaceous bedrock (“Regensburger Grünsandstein”). At a depth of 23.45 m, the well reached an artesian groundwater table (aquifer) with strong discharge and sulfuric odor. It has to be assumed that the well reached the stratigraphic boundary between Cretaceous and Jurassic sediments (Malm), which are both described as (calciferous) sandstones.

The springs at both locations are connected to the deep water flow within the pre-alpine Tertiary Molasse basin (Lamcke, 1976). The deep aquifer is developed in the karst and fracture system of the underlying Jurassic sediments. Stable isotope geochemistry at comparable sites within the region (Lower Bavaria) point out the constant drainage of pore water from the hanging Molasse and Cretaceous layers into the underlying karstified Jurassic layers (Andres and Frisch, 1981). Isotope mixing ratios that have remained unchanged for decades and one radiocarbon age (^{14}C) of 30,000 years (uncalibrated) revealed the long term runoff of water within the deep Malm karst system within the Molasse basin towards the Danube valley downstream of Regensburg (Andres and Frisch, 1981).
Even though the sources for hydrogen sulfide of sulfur springs elsewhere in Lower and Upper Bavaria are bituminous Mesozoic sediments, pyrite rich Jurassic sediments (Lias) or Tertiary brown coal deposits (Baumann, 1981; Nielsen, 1981), the sulfide at SM and MSI comes from microbial sulfate reduction. Due to the cold temperature inorganic reduction processes can be excluded (Nielsen, 1981). The sulfates are set free out of salinar formations (Zechstein) located at the alpine rim of the Molasse basin. The sulfidic sulfur develops from the sulfate reduction of substances set free out of salinar formations (Zechstein) located at the alpine rim of the Molasse basin (Nielsen, 1981). Consequently, the reduction must be triggered by microorganisms. Anorganic reduction processes can be excluded because of the lack of higher temperatures (Nielsen, 1981). Even though other sources for hydrogen sulfide like bituminous Mesozoic sediments, pyrite rich Jurassic sediments (Lias) or Tertiary brown coal deposits exist (Baumann, 1981; Nielsen, 1981), being the reason for sulfur springs elsewhere in Lower and Upper Bavaria, the sulfate bound is the only explanation approach for the origin of the hydrogen sulfide in the deep underground waters at both sites, Sippenauer Moor (SM) and Isling (MSI).

Sample collection. The sampling permit for the Sippenauer Moor was issued by the Regensburgische Botanische Gesellschaft von 1790 e.V., Regensburg. The sampling permit for the Mühlbacher Schwefelquelle was obtained from Gartenamt Regensburg. The field studies did not involve endangered or protected species.
Samples from the Sippenauer Moor were collected by polyethylene nets. These nets were either used to filter water directly at a spring outlet in order to harvest milky, slimy biofilms washed up from deeper earth layers or to provide attachment and growth conditions for string-of-pearls community at 0.65 m to 0.80 m distance from the spring (under oxygen-enriched conditions). The collected samples included three Sippenauer Moor biofilm (SM-BF) samples taken directly from the spring outflow where the oxygen and H_2S concentrations were 0.02 mg/l and 0.85 mg/l, respectively, and six string-of-pearls community (SOPC) samples taken at a location where the outflows of three springs mix (oxygen concentrations were 0.89 and 1.10 mg/l respectively; sulfide concentration at mixing area: 0.5 mg/l).
The second sampling location was the Mühlbacher Schwefelquelle nearby Isling (MSI; linear distance 20 km to Sippenauer Moor; N48°59.140′ E12°07.631′ (Henneberger et al., 2006; Rudolph et al., 2004); Fig. S1). The sampling permit was issued by the Gartenbauamt Regensburg. Three MSI biofilm (MSI-BF) samples were harvested in a similar manner under oxygen-free conditions (samples from Probst et al. 2013b); here, the H_2S concentration was 0.85 mg/l and the oxygen concentration was below detection limit (details see: Probst et al. 2013b). The water composition of the two springs was documented earlier and showed high similarity (Rudolph et al., 2004).

Sample preparation. Biofilm samples were removed from the polyethylene nets with syringes/pipettes and transported to the laboratory on ice. Samples for DNA extraction were stored at -20°C, and samples for SR-FTIR analyses were air-dried on gold-coated grids after removing the liquid (Probst et al., 2013b). Samples for whole-cell fluorescence *in situ* hybridization (FISH) were prepared as described earlier (Rudolph et al., 2001).

DNA extraction and quantitative PCR (qPCR). DNA was extracted from samples using the XS-buffer protocol (Moissl-Eichinger, 2011). Bacterial and archaeal 16S rRNA genes as well as *dsrB* (dissimilatory sulfite-reductase subunit B) genes were quantified by qPCR as described previously (Moissl-Eichinger, 2011; Probst et al., 2013b). For each sample type (MSI biofilm, SM biofilm, SOPC) three independent biological replicates were individually measured three times.

Amplicon generation for microarray analysis (Probst et al., 2013b). Here, bacterial 16S rRNA genes were amplified from metagenomic DNA samples with primer pair 27F and 1492R (Hazen et al., 2010), whereas archaeal 16S rRNA gene with 345af and 1406ur (Burggraf et al., 1992; Lane, 1991). Amplicons were purified by agarose-gel electrophoresis as performed

earlier (Probst et al., 2013b). All 15 samples for microarray analyses were taken within one day, PCR amplified and moved forward for microarray hybridization: 3x MSI biofilm, 3x SM biofilm and 6x SOPC, 1x MSI water, 1x SM water, and 1x extraction blank. Water samples and extraction blank were used for control purposes.

PhyloChip G3 data acquisition. The PhyloChip G3™ Assay (Second Genome, South San Francisco, CA) and analysis were carried out as described (Hazen et al., 2010). Briefly, bacterial (500 ng) and archaeal (100 ng) 16S rRNA gene amplicons were combined, spiked with a known amount of non-16S rRNA genes for standardization, fragmented and biotin labeled. After hybridization on DNA microarrays, images were scanned, background and noise was determined, and fluorescence intensity was scaled to the spike-in internal controls (Hazen et al., 2010).

Empirical OTU (eOTU) discovery from PhyloChip data (Probst et al., 2014). The 25-mer 16S probes were compared to their mismatch controls and 24154 were found to be responsive (Hazen et al., 2010) in at least three microarray datasets (biological samples). Taxonomically related probes were clustered into probe-sets where pair-wise correlations ≥0.85 between $\log_2$ transformed fluorescence intensities (FI) were discovered as described by Probst et al 2014. A total of 1380 probe-sets were found and the empirical operational taxonomic units (eOTU) tracked by each probe-set were taxonomically annotated against the 2012 taxonomy using a Naïve Bayesian scoring and >80% bootstrapped confidence cutoff (DeSantis et al., 2006; McDonald et al., 2012). The mean $\log_2$ FI among the multiple probes for each eOTU was calculated for each sample. These values are referred to as the hybridization score (HybScore) used in PhyloChip abundance-based analysis. For details please see supplementary information or Probst et al. 2014. Thirty-two eOTUs detected in the DNA extraction blank were removed from further analyses as were eOTUs only present in spring water samples, resulting in 1337 eOTU considered for microbiome analyses.

Statistical analysis of microarray data. Second Genome's Microbial Profiling Analysis Pipeline (PhyCA-Stats™) was used for univariate and multivariate statistics of abundance scores (hybridization scores) of all eOTUs that were called present in at least one of the samples. The analyses included hierarchical clustering (average neighbor), NMDS (non-metric multidimensional scaling) and Adonis testing based on weighted UniFrac distance measure (Lozupone and Knight, 2005; Lozupone et al., 2011). We identified eOTUs that were significantly enriched in a sample category by applying a Welch-test on eOTU trajectories. The same test was applied for microbiome changes at family level, where abundances of each eOTU were summarized per family prior to significance testing.

Performance of microarray data for improved OTU calling. In this study, we used the well-established PhyloChip G3 DNA microarray for deciphering community relationships (Cooper et al., 2011; DeSantis et al., 2007; Hazen et al., 2010; Mendes et al., 2011; Vaishampayan et al., 2013). Originally, microarray technology designed on the basis of a reference dataset of 16S rRNA genes does not allow the detection of precluded / unknown 16S rRNA genes when using a reference database for OTU calling (Brodie et al., 2007; La Duc et al., 2009). However, the approach used herein differs by the mean of empirical OTU identification (Probst et al., 2014) and detected one eOTU affiliated to the SM1 Euryarchaeon (bootstrap 70%), although this archaeon has not been included in the original probe design of the array (method for classification of concatenated, interrupted probe sets identical to method described below for 16S rRNA gene classification / intergenic spacer region sequencing, see below). Consequently, this approach allowed inclusion of 16S rRNA genes not included in the chip design for microbial community relationship calculations. For details on empirical, non-supervised OTU calling the reader is referred to the Supplementary Information and Probst et al., 2014.

SR-FTIR spectromicroscopy and data analysis. SR-FTIR spectromicroscopy using photon energy in the mid-infrared region [4000 to 650 cm^{-1}; (Holman et al., 2010)] was used to obtain chemical information of the biofilm samples. Band assignment and spectra interpretation was done as described earlier (Mantsch and Chapman, 1996). More than 70,000 SR-FTIR spectra were collected for all biofilms at the infrared beamline (http://infrared.als.lbl.gov/) as described in Probst et al. 2013b. For each spectrum, the membrane methyl (-CH3) to the methylene (-CH2) absorbance ratio was computed, and a threshold value of 0.75 was used to designate the spectrum to be archaeal (≥ 0.75) or bacterial [<0.75; (Probst et al., 2013b)]. Univariate analysis was used to obtain semi-quantitative information on sample biogeochemical composition (see supplementary information). Principal Component-Linear Discriminant Analysis (PC-LDA) was performed in the Matlab (The

MathWorks, Inc., Massachusetts USA) environment using lipid spectral window (2800-3100 cm^{-1}) for archaeal communities, and lipids, plus carbohydrates (1280-900cm^{-1}) for bacterial communities. Both datasets were vector normalized by the Amide II (1550±10 cm^{-1}) absorption intensity.

Fluorescence in situ hybridization (FISH). Whole-cell hybridization was carried out as described in Rudolph et al. 2001 with following probes [Rhodamine Green (RG) or CY3 labeled]: EUB338/I (Amann et al., 1990b), ARCH344 (Moissl et al., 2003), SMARCH714 [SM1 Euryarchaeon; (Moissl et al., 2003)] SRB385 (Amann et al., 1990a) and Delta495a/b/c probe mix (Loy et al., 2002). Bacterial positive controls (strain *Escherichia coli* K12, DSM 30083) and negative controls (non-sense probe NONEUB338) were used to validate the experiments. Thereafter the samples were analyzed as described (Probst et al., 2013b).

Scanning electron microscopy (SEM) and transmission electron microscopy (TEM). For SEM, drops of fixed samples (0.1% glutardialdehyde; w/v) were placed onto glass slides, covered with a coverslip, and rapidly frozen with liquid nitrogen. The coverslip was removed with a razor blade and the glass slide was immediately fixed with 2.5% (w/v) glutardialdehyde in fixative buffer, washed, postfixed with 1.0% osmium tetroxide, washed with buffer, followed by deionized water, dehydrated in a graded series of acetone solutions, and critical-point dried after transfer to liquid CO_2. Specimens were mounted on stubs, coated with 3 nm of platinum using a magnetron sputter coater, and examined with a Zeiss Auriga scanning electron microscope operated at 1 kV. For TEM, fresh, unfixed biofilm pieces were deposited on a carbon-coated copper grid and negatively stained with 2% (w/v) uranyl acetate, pH 4.5 or 2.0% (w/v) phosphotungstic acid (PTA), pH 7.0. Samples were examined using a CM12 transmission electron microscope (Philips) operated at 120 keV.

Southern blotting of metagenomic DNA. Probes targeting the *hamus* gene were generated via amplification with the pili1f and pili2r primers (pili1f: 5'-CAGCATCAAAACAGGCGGGTGC-3', pili2r: 5'-GTTCCTCTGAATTTGTATACGG-3') and labeled with DIG High Prime as described in the manufacturer's instructions (Roche Diagnostics GmbH, Mannheim). 1 µg of metagenomic DNA of each biofilm type (Mühlbacher Schwefelquelle, Sippenauer Moor) was individually digested using the enzymes *Hinc*II and *Kpn*I, then electrophoresed and blotted on a nylon membrane. After hybridization with the *hamus*-specific probe, the membrane was blocked in TBST-B-buffer followed by an antibody reaction with Anti-Digoxygenin-AP conjugate (dilution up to 1:10000, Roche Diagnostics GmbH, Mannheim). The blot was washed thoroughly in 2xSSC (saline-sodium citrate buffer) / 0.1% SDS at RT followed by a second washing step in 0.5x SSC/ 0.1% SDS at 68°C. The detection was carried out through NBT/BCIP (nitro blue tetrazolium/ 5-bromo-4-chloro-3-indolyl-phosphate) reaction.

Intergenic spacer region sequencing of SM1 euryarchaeal strains. Parts of the archaeal rRNA operon were amplified from metagenomic biofilm DNA samples (MSI-BF and SM-BF) using 16S-345af [5'-CGGGGYGCASCAGGCGCGAA-3' (Burggraf et al., 1997)] and 23S-64R [5'-GCCNRGGCTTATCGCAGCTT-3' (Summit and Baross, 2001)]. Amplicons were cloned in *E. coli* (TOPO TA cloning kit, TOP 10' cells, Invitrogen) and 48 inserts per sample were bi-directionally sequenced (LGC Genomics, Berlin). Reverse sequences were trimmed to 16S rRNA genes and classified using the Naive Bayesian algorithm implemented in mother (Schloss et al., 2009; Wang et al., 2007) against an updated and 98%-clustered GreenGenes database [http://www.secondgenome.com/go/2011-greengenes-taxonomy/ (McDonald et al., 2012)] supplemented with known archaeal 16S rRNA gene sequences from sulfidic springs. Sequences classified as SM1 Euryarchaeota (bootstrap >90%) were then trimmed to the intergenic spacer region using the full reverse sequence. Multiple sequence alignments were generated using MUSCLE (Edgar, 2004).

Results

Dominance of Archaea in subsurface biofilms confirmed by molecular approaches. PhyloChip G3, qPCR and FISH revealed the predominance of Archaea in the subsurface biofilms samples from Mühlbacher Schwefelquelle (MSI-BF) and Sippenauer Moor (SM-BF) samples (*Table III.4-1*). QPCR showed that >97% of all 16S rRNA genes in MSI-BF and SM-BF samples were archaeal, but only 26% in the surface string-of-pearls community (SOPC). Cell counting after FISH staining showed 93% and 86% archaea in MSI-BF and SM-BF, respectively. These abundances were confirmed via SR-FTIR image analysis, which typically showed that archaea occupied >97% of the areas in MSI-BF and SM-BF, but only ~38% in SOPC (*Figure III.4-2*, left panels).

Table III.4-1 Quantification of archaeal and bacterial signatures via qPCR, FISH and SR-FTIR (values in brackes give standard deviation).

Method	Measurand	MSI-BF	SM-BF	SOPC
	Archaeal *16S rRNA* genes	2.89E+06 (±5.63E+05)*	2.09E+06 (±1.08E+06)	2.03E+05 (±5.48E+04)
QPCR (per ng DNA)	Bacterial *16S rRNA* genes	7.48E+04 (±7.65E+03)*	2.20E+04 (±2.68E+03)	5.75E+05 (±1.09E+05)
	DsrB genes	5.12E+03 (±7.87E+02)*	1.97E+03 (±9.82E+02)	3.46E+02 (±1.25E+02)
	Percent Archaea	97.44%*	98.96%	26.09%
	Percent archaeal cells	92.96 (±2.16)	86.41 (±7.02)	ND
FISH	Percent SRB385 stained Bacteria	85.4 (±4.7)*	39.32 (±11.81)	ND
	Percent DeltaMix stained Bacteria	89.2 (±0.9)*	63.87 (±14.70)	ND
SR-FTIR	Percent archaeal biomass	97.0 (±6.0)	97.1 (±4.4)	38.7 (±13.8)

*Abbreviation: ND, not determined; *data from Probst et al. 2013b*

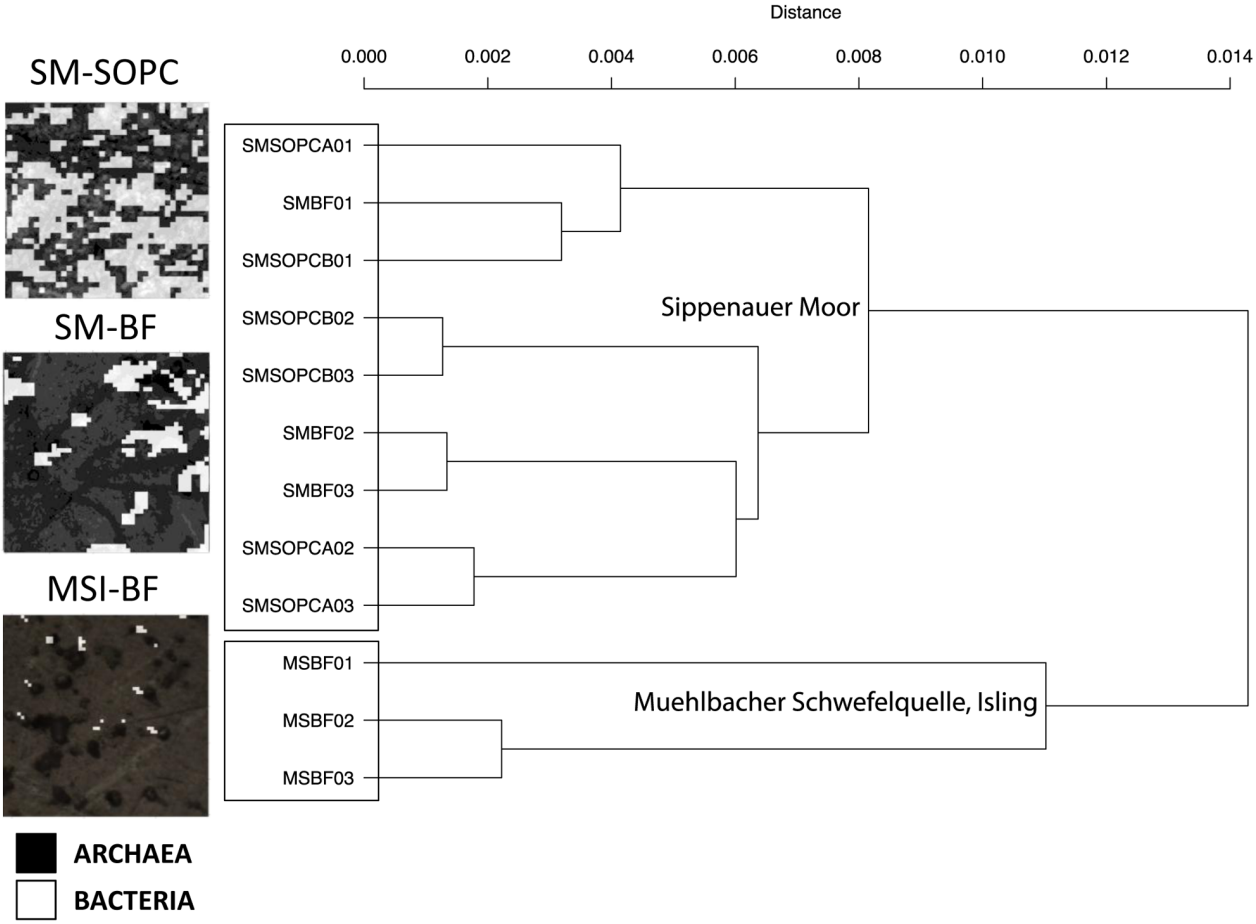

Figure III.4-2 *Abundance of Archaea and Bacteria in samples and the overall community relationship. Small panels present binary images of infrared data collected for three sample types, SM-BF (Sippenauer Moor, biofilm), SM-SOPC (Sippenauer Moor, string-of-pearls community) and MSI-BF (Mühlbacher Schwefelquelle, Isling, biofilm). Infrared maps show the distribution of Archaea and Bacteria in the samples. One pixel corresponds to 2 µm. Hierarchical clustering based weighted UniFrac of abundance values of eOTUs (Bacteria and Archaea). Two different clusters separating the samples based on hydrogeology were observed.*

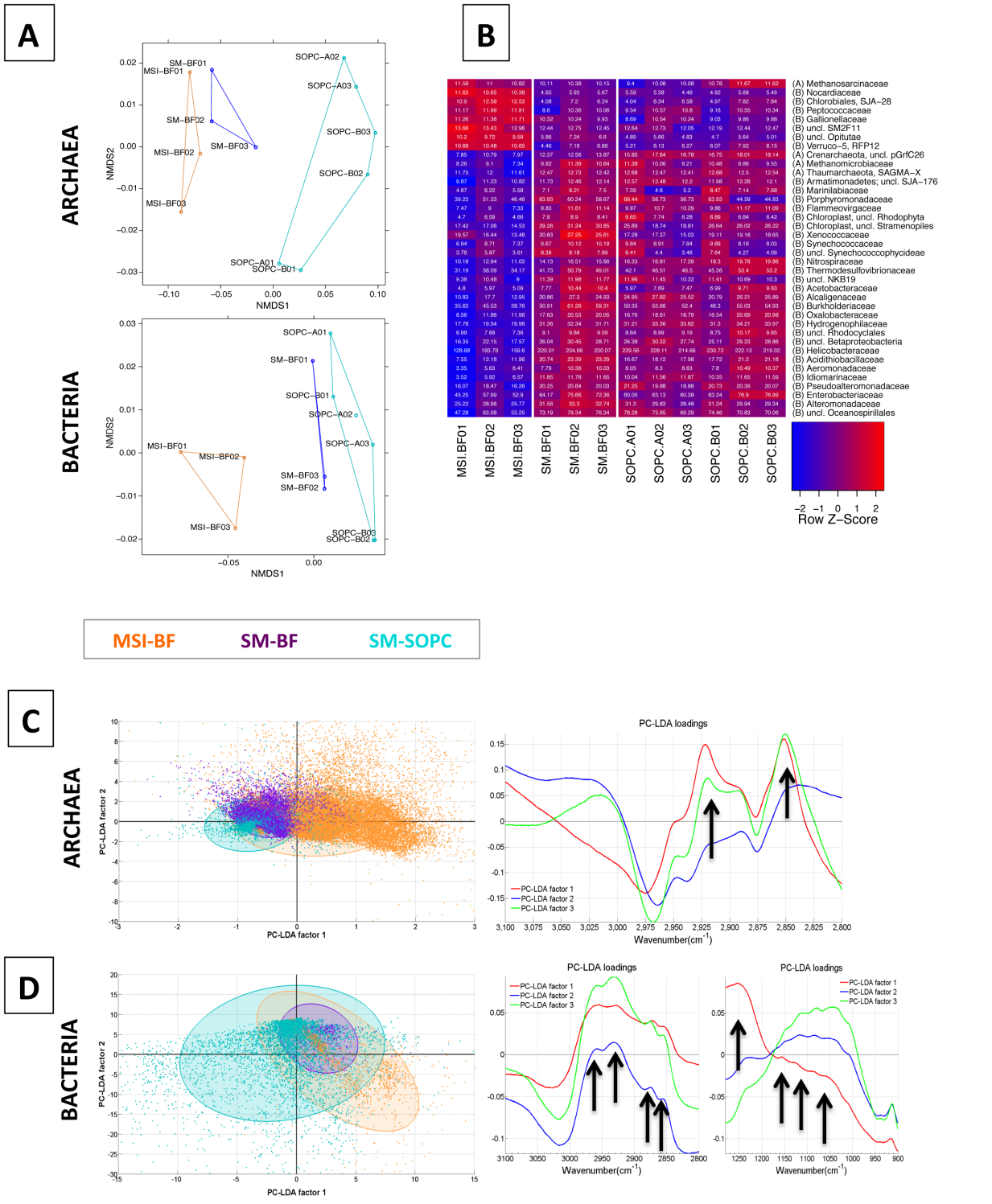

Figure III.4-3 *Detailed community profiling using PhyloChip G3™ and SR-FTIR.* ***A****) Ordination analysis of PhyloChip G3™ data based on weighted UniFrac measure of eOTU abundances followed by non-metric multidimensional scaling (NMDS). Stress for NMDS of archaeal eOTUs (#37): 0.0088. Stress for NMDS of bacterial eOTUs (#1300): 0.0223.* ***B****) Heatmap displaying significantly different families found between the two biofilm types, MSI-BF and SM-BF by PhyloChip G3™ assay. Significance is based on aggregated HybScores of eOTUs on family level followed by a Welch-test. For false discovery detection please see Fig. S6.* ***C****) Ordination analysis of SR-FTIR data based on a linear discriminant analysis and principal component analysis (PCA-LDA) in the spectral region of 2800-3100 cm^{-1} on the archaea spectra extracted from the maps from the three different locations. On the right there is the plot of PC-LDA loadings. PC-LDA1 explains for the 93.4% of the variance, PC-LDA2 for 5.3% and PC-LDA3 for 0.9%. Arrows point to the infrared signals used to explain the difference between the samples: 2975 cm^{-1}, 2965 cm^{-1}, 2924 cm^{-1} and 2850 cm^{-1}.* ***D****) PCA-LDA in the spectral regions of 900-1280 cm^{-1} and 2800-3100 cm^{-1} on SR-FTIR spectra of the bacteria "pixels" from the chemical maps of the samples at the three different locations. On the right there is plot of PC-LDA loadings in the two spectral region of interest. PC-LDA1 explains for the 54.5% of the variance, PC-LDA2 for 28.6% and PC-LDA3 for 7.3%. Arrows point to the main infrared signals used to explain the difference between the samples: 2958 cm^{-1}, 2925 cm^{-1}, 2870 cm^{-1} and 2850 cm^{-1}, in the second panel 1250 cm^{-1}, 1110 cm^{-1}, 1080 cm^{-1} and 1045 cm^{-1}.*

Site-specific microbiomes. PhyloChip G3 DNA microarray technology identified a total of 1300 bacterial and 37 archaeal eOTUs. Hierarchical analyses based on weighted UniFrac dissimilarities revealed clusters of samples based on their geographical regions (SM versus MSI; *Figure III.4-2*, right panel), which is supported by a highly significant Adonis p-value (0.008). The macroscopic appearance of samples (BF vs. SOPC) was also identified to have a significant influence on the observed microbiome relationships (p-value = 0.003) for biofilm samples from both locations. Considering samples from SM only, the microbiome differences between SM-BF and SOPC were insignificant (p-value = 0.058) indicating a site-specific microbiome.

Spatial dynamics of archaeome and bacteriome relationships. Aiming to analyze the microbial community relationships in detail, NMDS were performed on PhyloChip G3 derived 16S rRNA gene profiles of Bacteria and Archaea separately. For Archaea, the NMDS plot (*Figure III.4-3A*, upper panel) separated MSI-BF and SM-BF from SM-SOPC along the NMDS1 axis. This implies a greater similarity in the archaeal community relationship among samples from the anoxic subsurface (biofilms) than among samples from same hydrogeological regions but of different oxygen content (SM-BF versus SOPC). These differences went along with an increased richness of archaeal eOTUs in SOPC samples, which also included e.g. Thaumarchaeota representatives. For Bacteria, however, the NMDS plot (*Figure III.4-3A*, lower panel) separated the MSI-BF samples from the SM-BF and the SM-SOPC samples, suggesting that the bacterial community relationship was affected more strongly by the sampling location and its hydrogeology (additional information in Table S1).
Notably, the bacterial composition of the Sippenauer Moor biofilm (SM-BF) and string-of-pearls community (SM-SOPC) appeared very similar (*Figure III.4-3A*, lower panel), although there was a strong increase in oxygen content from the biofilm sampling area to the SOPC sampling area. It might be concluded, that the bacteriome of the SM-BF tended to be maintained in the short travel distance to the SOPC (*Figure III.4-1*) along with oxygen increase.

Both biofilms carried sulfate-reducing bacteria (SRB) with different taxonomic affiliation. 290 of 1337 eOTUs were significantly different in their relative abundance when comparing MSI-BF samples with SM-BF samples (Fig. S2A, S3; p-values <0.05) resulting in separated microbiomes (Fig. S2B). We observed that eOTUs of certain phyla like Verrucomicrobia or Spirochaeta and two eOTUs classified as *Desulfobacteraceae* belonging to the twelve most significant eOTUs (Fig. S4) were significantly enriched in MSI-BF vs. SM-BF samples. Other members of this family of SRB were also significantly enriched in the SM-BF samples but with higher p-values [0.003:0.050[.
Notably, SOPC samples clustered with SM-BF samples in hierarchal dendrograms (Fig. S2B) reflecting the similarity of these populations observed in other multivariate statistics mentioned earlier (*Figure III.4-2*, Table S1).
Considering summarized hybScores at family level (hybridization scores of eOTUs were aggregated across families based on taxonomic affiliation), 38 of 227 families had significant changes across aggregated trajectories between biofilm categories (*Figure III.4-3B*, Fig. S5). The signatures of the designated SRB families like *Desulfobacteraceae*, *Desulfobulbaceae*, *Desulfovibrionaceae* did not show a significant variation between the MSI-BF and SM-BF (p-values were 0.37, 0.30, and 0.36, respectively). However, the abundance of *Desulfobacteraceae* displayed a significant difference between the two biofilms, MSI-BF/SM-BF, and the SOPC samples (p-value = 0.01). The eOTU and family level analysis allowed the conclusion that SM1 Euryarchaeon biofilms support an enrichment of members of SRB, with differences at eOTU level but similarities on their family level. In other words, SRB – which dominated the bacteriome based on FISH analyses – were members of the same family, but of different species or strains.
These data are in accordance with microscopic FISH data, and quantitative PCR of *dsrB* genes, which showed an increase of one order of magnitude in copy numbers for biofilm samples compared to SOPC samples (*Table III.4-1*).

Similarity and variations of archaeal lipid signatures in biofilms. We applied PC-LDA analysis to the SR-FTIR spectra previously categorized as archaeal or bacterial (*Table III.4-1*) to gain insight into the biochemical differences in the composition at a functional group level of each microbiome. For the archaeal spectra, the two-dimensional PC-LDA score plot revealed that the first PC-LDA factor separated archaea in SM-BF and SM-SOPC samples from the majority (~70%) of the archaea in the MSI-BF samples (*Figure III.4-3C*, left panel). The first loading vector (the red trace in *Figure III.4-3C*, right panel) showed that positive features near 2924 cm^{-1} and 2850 cm^{-1} were responsible for this separation. These frequencies correspond to the infrared absorption signals of the asymmetric and symmetric vibrations of CH_2 in fatty acid chains of the

membrane amphiphiles. Additional peaks at 2975-2965 cm^{-1} are associated to the methoxy CH stretching of -OCH_3 and -OCH_2 ethers (Socrates, 2004). Therefore, the PC-LDA loadings plot suggested that the SM-BF and SM-SOPC archaea shared a similar membrane lipid composition, but differed from over 70% of the MSI-BF archaea. This could be explained by differences in the alkyl chain branching and in the polar heads (Ulrih et al., 2009). Meanwhile, the two-dimensional PC-LDA score plot of the bacterial spectra in both the lipid and the overall fingerprint region, were similar to microbiome relationships as revealed by PhyloChip analysis (see above).

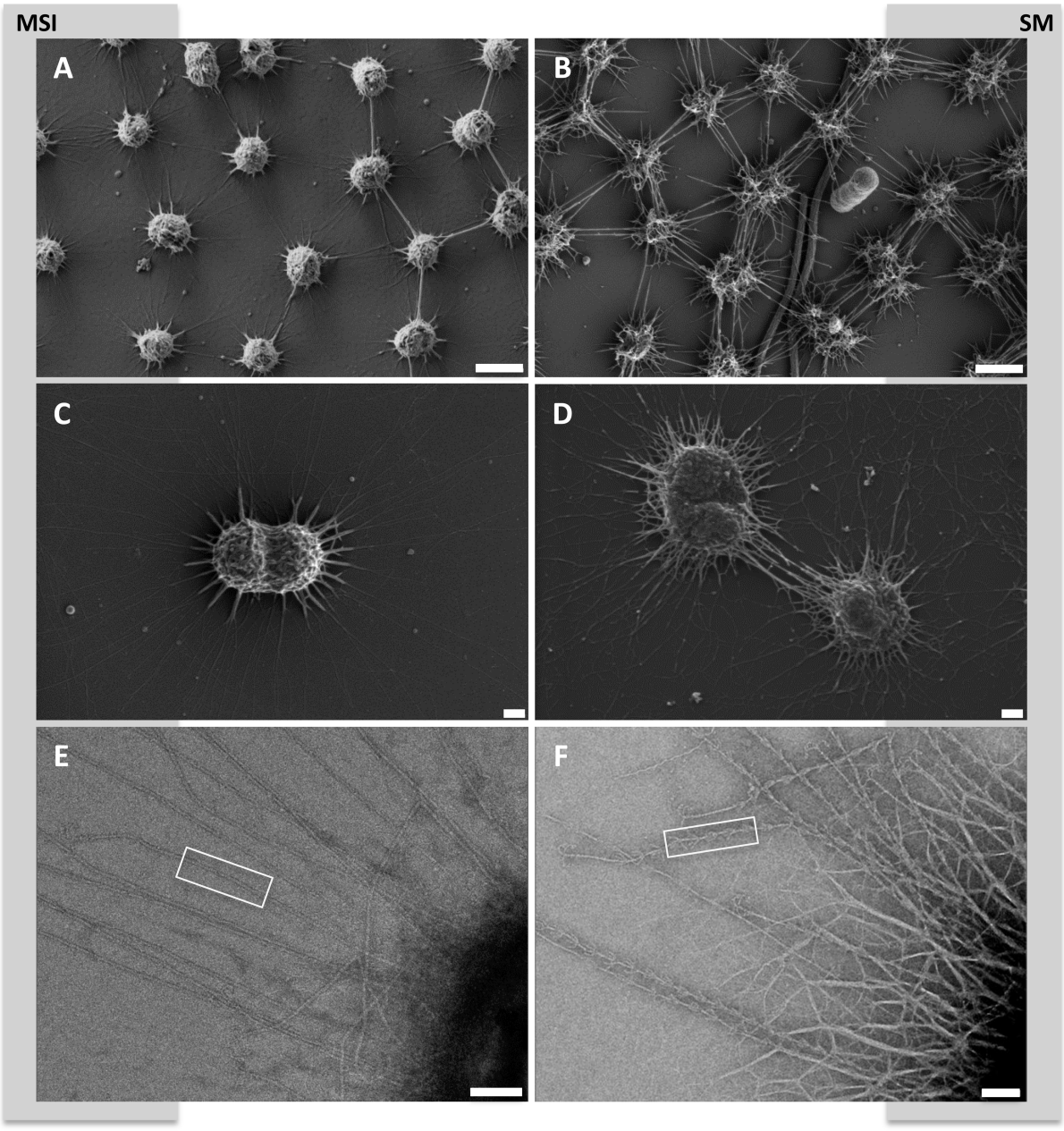

Figure III.4-4 *Scanning and transmission electron micrographs of biofilms, cells and hami. Left panels: MSI, right panel: SM.* ***A****) Scanning electron micrograph of MSI biofilm, showing SM1 euryarchaeal cells with defined distances and cell-cell connections. Bar: 1 µm.* ***B****) Scanning electron micrograph of SM biofilm, showing SM1 euryarchaeal cells with defined distances and fine-structured cell-cell connections. In-between: Bacterial filamentous and rod-shaped cells. Bar: 1 µm.* ***C****) Scanning electron micrograph of dividing SM1 euryarchaeal cell (MSI) with cell surface appendages. Bar: 200 nm.* ***D****) Scanning electron micrograph of dividing SM1 euryarchaeal cell (SM) with cell surface appendages. Bar: 200 nm.* ***E****) Transmission electron micrograph of cell surface appendages (hami) of SM1 euryarchaeal cells from the MSI biofilm. The hami carry the nano-grappling hooks, but besides that appear bare (square), without prickles (Moissl et al. 2005b). Bar: 100 nm.* ***F****) Transmission electron micrograph of cell surface appendages and matrix of SM1 euryarchaeal cells from the SM biofilm. The hami reveal the typical ultrastructure, with nano-grappling hooks and barbwire-like prickle region (square, Moissl et al 2005b). Bar: 100 nm.*

Ultrastructural differences exhibited in SM1-Euryarchaeon biofilms and hami appearance. A univariate analysis of the infrared absorption bands of the biomacromolecules (Fig. S6) confirmed that MSI-BF and SM-BF had the highest protein and lipid contents, whereas SM-SOPC the highest carbohydrate content. Consequently, samples from both biofilms were analyzed further using SEM and TEM to look into the ultrastructural differences.

SM1 archaeal cells appeared as single or dividing cocci (*Figure III.4-4*), connected via a network of cell appendages and extracellular matrices. Considering one layer of SM1 Euryarchaea (*Figure III.4-4A*, *4B*), most cells revealed regular distances to six neighbors in a hexagonal manner. Cells in the MSI biofilms were significantly larger than those in the SM biofilms (average diameter of 0.72 µm versus 0.60 µm, homoscedastic student's t-test of 40 cells each: p-value < 1.0E-07). Additionally, cell surfaces were napped and connections were smoother in MSI biofilms (*Figure III.4-4C*), whereas cell surfaces in the SM biofilms appeared fluffy with more connections between cells (*Figure III.4-4D*).

Cells in both biofilms carried *hami* with distal hooks that appeared correctly folded (*Figure III.4-4E-F*). Nevertheless, SM1 Euryarchaea in the MSI biofilm revealed only a low percentage of such correctly folded *hami* structures with respect to the 'prickle region' (Moissl et al., 2005b). Prickle regions seemed absent in most MSI *hami* (*Figure III.4-4E*), whereas such "bare" *hami* were sparsely observed for SM biofilm cells (*Figure III.4-4F*).

Two strains of the same archaeal species dominated the two biofilms. It was believed that the SM1 Euryarchaeota from SM and MSI were identical based on analysis at the 16S rRNA gene level (Henneberger et al., 2006). Yet both showed such strong variations in the membrane lipid composition (*Figure III.4-3C*) and ultrastructure (*Figure III.4-4*), implicating possible differences between the two archaeal populations at genomic level. Under this observation a comparative Southern blot analysis of biofilm DNA (MSI versus SM) with probes specifically designed to target the *hamus* gene, that encodes for the major protein of the unique cell surface appendages, was performed (Moissl et al., 2005b). Using different restriction enzymes (*Hinc*II and *Kpn*I) hybridization signals of several distinct bands were retrieved (Fig. S7). This result indicated at least the presence of more than one *hamus* gene in both samples. Moreover, there is a reliable difference in the restriction pattern between the metagenomic DNA from both biofilm types. Notably, SM1 cells purified from the SM-SOPC (Moissl et al., 2003) produced the same pattern as SM-BF. Additionally, sequencing of 96 clones of intergenic spacer regions (between 16S rRNA gene and 23S rRNA gene), without pre-selection of the clones via RFLP (Henneberger et al., 2006), showed that six single nucleotide polymorphisms existed between the two dominant sequences from the MSI-BF and SM-BF (Fig. S8), while the 16S rRNA gene sequences were identical providing evidence for different dominant strains of SM1 Euryarchaeota at the two sampling sites.

Discussion

Subsurface microbial life exists in an environment that is challenging in many ways: lack of sunlight, mostly cold temperatures, low nutrient levels and often anoxic conditions demand alternative ways of carbon assimilation and energy production. This includes the usage of other electron receptors than oxygen, resulting in anaerobic respiration or fermentation (Lovley and Chapelle, 1995; Lovley and Coates, 2000). To date, information on subsurface life is very limited. This is either due to the restricted accessibility of subsurface biotopes or due to the detection of many unexplored microbial taxa therein, which remain uncultivated and thus largely not understood (Castelle et al., 2013; Ortiz et al., 2013; Wrighton et al., 2012). In this regard, the two vicinal sulfidic springs studied here provide a stable and well accessible window to the subsurface and allow the exploration and comparison of archaeal biofilms delivered to the surface. While microbiome profiling on 16S rRNA gene sequences revealed similar archaeomes, the SR-FTIR approach uncovered striking differences in archaeal lipid signatures at a molecular level. These variations were either caused by the presence of different organisms (at strain level) or altered gene expression of the same organism, most likely reflecting adaptive responses to different environmental conditions. In contrast, PhyloChip data of the two biofilms revealed the enrichment of designated SRB of different taxonomic affiliation at OTU but not at family level, which, however, seemed to share widely diverse lipid composition as revealed by SR-FTIR ordination.

The constant co-appearance and bacterial predominance of such actively sulfate-reducing deltaproteobacteria within both SM1 biofilms suggests a possible syntrophic relationship (Morris et al., 2013; Probst et al., 2013b). Even more, the presence

of sulfate-reducers could also reflect environmental conditions prevailing in the biofilms' original biotopes and thus the growth conditions of the SM1 Euryarchaeon.

Generally, SRB's sulfate-reducing activity is linked to the oxidation of organic compounds or molecular hydrogen and to the formation of H_2S, an important biogenic compound found in considerable amount (0.85 mg/l) in both spring waters. As a requirement for the SRB catalyzed reactions, the biotope, or respective environment, needs to fulfill at least the following criteria: a) anoxic conditions, b) sulfate as an electron acceptor, and c) an electron donor, most likely either organic molecules or hydrogen. It could therefore be hypothesized, that the SM1 Euryarchaeon thrives under these conditions or even provides such an environment, creating a convenient biotope for SRB.

When biofilm pieces are washed up into oxygen-mixed areas of the surface spring water and attach to rigid material, the entire community is transformed into a string-of-pearls-like macroscopic appearance. The archaeal diversity increased, as shown for instance by the detection of Thaumarchaeota, and the bacteria originally being part of the biofilm, are absorbed into the string-of-pearls community. This process is completed by the most likely intentional settling of filamentous, sulfide-oxidzing bacteria (*Thiothrix*, *Sulfuricurvum*), which cover the archaeal microcolony and become an equal partner of the SM1 Euryarchaeon (Moissl et al., 2002; Rudolph et al., 2004). Supporting evidence for this hypothesis comes from the fact that filamentous bacteria were cocooned by cell surface appendages of SM1 Euryarchaeota in biofilm samples and similar bacteriomes were found for the biofilm and the SOPC at the Sippenauer Moor. The biofilms can therefore be considered precursors of the string-of-pearls community (*Figure III.4-1*).

Based on hydrogeology we can exclude a direct exchange of biomaterial between both aquifer outlets with respect to the subsurface water current (both aquifers are artesian). Additionally, even though both wells are fed by the same deep groundwater flow within the pre-alpine Tertiary Molasse basin, based on our studies, we can exclude a parallel transport of microbial communities from these regions to both springs, Sippenauer Moor and Mühlbacher Schwefelquelle: If delivered simultaneously to both biotopes from the same origin, one would expect similar patterns in (bio-) geochemical profiles and microbial diversity, since the biofilms analyzed were sampled in parallel (within one day). It appears that the local hydrogeology and geochemistry in the subsurface of the individual springs are responsible for creating different biotopes and causing differences in the observed microbiome structure.

Although a number of details with respect to archaeal 16S rRNA gene sequences, prevalence of SRB and general biofilm-structure are in agreement, the communities from both locations, and also the archaea themselves, reveal severe differences at various levels. For instance, SR-FTIR detected location dependent shifts in lipid profiles of biofilm associated archaea. In general, lipid variations can be growth phase dependent (Thirkell and Gray, 1974), point to a specific biotope-adaptation (De Rosa and Gambacorta, 1988; Sprott, 1992) and thus reflect influences from environmental parameters in both biotopes – or simply mirror strain-specific properties. The latter possibility is supported by detectable differences in fingerprint experiments with metagenomic DNA and within the archaeal SM and MSI 16S-23S rRNA gene intergenic spacer regions. Consequently, Southern-blotting and intergenic spacer analysis, together with the above-mentioned SR-FTIR analysis, and ultrastructural analyses suggested that two different SM1 euryarchaeal populations dominate the biofilms that can be found at the Mühlbacher Schwefelquelle and at the Sippenauer Moor. To our knowledge, this is the first report of a natural divergence of one archaeal species in nature.

Studying these archaeal communities, which still remain dark matter with regard to biochemical cycling, provided insight into the hydrogeological impact on microbiome variation and into potential microbial niche differentiation. Our multifarious results, based on the commingling of established and novel methods, have added another piece to the puzzle in order to understand the dynamics of subsurface microbial life in such a great, dark and little explored environment.

Acknowledgements

We thank Prof. Dr. Robert Huber for critical discussions. This research was funded by the DFG (grant MO 1977 3-1). AJP was supported by the German National Academic Foundation (Studienstiftung des Deutschen Volkes). The SR-FTIR spectromicroscopy work was conducted through the Berkeley Synchrotron Infrared Structural Biology (BSISB) Program at the Advanced Light Source, which is supported by the Director, Office of Biological and Environmental Research's Structural Biology Program, Office of Science of the U.S. Department of Energy through contract DE-AC02-05CH11231 with Lawrence Berkeley National Laboratory.

Conflict of interest

TZD and AJP are paid employees of Second Genome, Inc. (www.secondgenome.com). Their association with Second Genome does not alter our adherence to all journal policies. Second Genome is a gastrointestinal therapeutics company with a pipeline of intestinal microbiome modulators that impact metabolic diseases and IBD. The publication of archaeal microbiome dynamics in sulfidic groundwater does not influence the value of any of Second Genome's therapeutic assets.

Supplementary Information

Supplementary information, including PhyloChip G3 data, can be found on the supporting DVD.

5. Grappling with dark matter: Biology of an uncultivated subsurface archaeon

Alexander J. Probst[1], Thomas Weinmaier[2], Kasie Raymann[3], Alexandra Perras[1], Joanne B. Emerson[4], Thomas Rattei[2], Gerhard Wanner[5], Andreas Klingl[6], Ivan Berg[7], Bernhard Viehweger[8], Marcos Yoshinaga[8], Kai-Uwe Hinrichs[8], Simonetta Gribaldo[3], Brian C. Thomas[4], Sandra Meck[1], Anna Auerbach[1], Matthias Heise[9], Jillian F. Banfield[4] and Christine Moissl-Eichinger[1]

[1]University of Regensburg, Department for Microbiology and Archaea Center, Universitaetsstrasse 31, 93053 Regensburg, Germany; [2]University of Vienna, Division for Computational Systems Biology, Althanstraße 14, A-1090 Wien, Austria; [3]Institut Pasteur, Unité Biologie Moléculaire du Gene chez les Extrêmophiles, Département de Microbiologie, Paris, 75724 Cedex 15, France; [4]Department of Earth and Planetary Science, University of California, Berkeley 94720, USA; [5]University of Munich, Biozentrum, Department of Biology I, Großhadenerstrasse 4, 82152 Planegg-Martinsried, Germany; [6]Cell Biology and LOEWE Research Centre for Synthetic Microbiology (Synmikro), Karl-von-Frisch-Str. 8, 35043 Marburg, Germany; [7]Department of Microbiology, Faculty of Biology, University of Freiburg, Schänzlestr. 1, 79104 Freiburg, Germany; [8]Organic Geochemistry Group, MARUM-Center for Marine Environmental Sciences and Department of Geosciences, University of Bremen, D-28359 Bremen, Germany; [9]Institute of Experimental and Applied Physics, University of Regensburg, Universitaetsstrasse 31, 93053 Regensburg, Germany

Correspondence: C Moissl-Eichinger, University of Regensburg, Department for Microbiology and Archaea Center, Universitaetsstrasse 31, 93053 Regensburg, Germany. E-mail: christine.moissl-eichinger@ur.de; J F Banfield, University of California, Department of Earth and Planetary Sciences, 369 McCone Hall, Berkeley, CA 94720, US

Publication information:
In preparation.

Abstract

Subsurface microbial life appears to be vastly unexplored. However, recent advances in metagenomics and single cell sequencing have shed light onto the genetic specialties of its dark matter. Although this has expanded our understanding of Earth's subsurface microbiomes, most analyses are solely based on DNA sequence information. Here, we linked genome information to biology via transcriptomics, immunological and ultrastructural analyses for the uncultivated SM1 Euryarchaeon. Populations exhibited significant strain variation, and phylogenomic analysis of genomic data of this archaeon and its relatives suggested that they are representatives of a diverse and widespread novel euryarchaeal order (Candidatus "Altiarchaeales") that can dominate subsurface biotopes. Detailed analyses of the recovered genomes of "Candidatus Alticharchaeum hamiconnexum" suggested a strictly anaerobic, acetoclastic metabolism with a novel archaeal, Factor$_{420}$-free reductive acetyl-CoA pathway. Ultrastructural analyses of the archaeal cell walls revealed a double-membrane system composed of glycosidic diether lipids. Approximately 100 cell surface appendages resembling nano-grappling hooks are anchored in the membranes and protrude from each cell. A structural assembly model could be reconstructed for the grappling hooks from the genomic information. Biofilm formation mediated by these hooks and by various sugars may enable the archaea to filter for substrates and simultaneously bar other microorganisms from the precious nutrients like acetate that seep into the groundwater. This communication demonstrates how deep novel, cultivation-resistant microbial lineages can be examined in their environmental context.

Introduction

Archaea are important players in global biochemical cycles. They can utilize a variety organic and inorganic electron donors and acceptors and contribute to nitrogen cycling in soil and sea via ammonia oxidation (Konneke et al., 2005; Leininger et al., 2006; Offre et al., 2013). Unique mechanisms for carbon turnover by archaea can lead to production or consumption of important greenhouse gases, via e.g. the anaerobic oxidation of methane (Orphan et al., 2001). Archaea have also the ability to assimilate CO_2 by at least three different fixation pathways (Berg et al., 2010) and are apparently adapted to special microbial niches out-competing their bacterial counterpart. However, biotopes dominated by one single archaeal species are rare and have not been studied in the past concerning their metabolic capabilities and contribution to Earth's biochemical cycles (Chapelle et al., 2002; Henneberger et al., 2006; Probst et al., submitted; Probst et al., 2013b).

Although environmental genomics has led to the characterization of many novel archaeal lineages, like Korachaeota (Elkins et al., 2008), ARMAN (Baker et al., 2010; Baker et al., 2006) and other members of the DPANN superphylum (Rinke et al., 2013), there is still a great level of novelty in nature that remains unexplored (Baker and Dick, 2013). These linages often escape regular 16S rRNA screenings (Baker et al., 2006) and may harbor an unexpected metabolic diversity and unusual cell structure (Baker et al., 2010; Baker et al., 2006; Comolli et al., 2009). An example of a genetically unexplored archaeal lineage is represented by the SM1 Euryarchaeon. This archaeon was discovered to either live in close association with sulfur-oxidizing bacteria in sulfuric streamlets (Rudolph et al., 2004; Rudolph et al., 2001) or as an almost single species biofilm in oxygen-free subsurface aquifers (Henneberger et al., 2006; Probst et al., submitted; Probst et al., 2013b). Similar to ARMAN (Baker et al., 2006), the SM1 Euryarchaeon may have escaped regular 16S rRNA gene profiling but was found to be widespread in sulfidic springs in Europe when appropriate 16S rRNA gene primers were used (Rudolph et al., 2004).

Novel archaeal lineages were often shown to have an unusual biology as assessed by genomics, cryo-TEM and other technologies directly applied to environmental samples (Baker et al., 2010; Baker et al., 2006; Comolli et al., 2009; Moissl et al., 2005b). ARMAN archaea for instance, revealed to have a highly unexplored genetic repertoire (Baker et al., 2010) and were the one of the first archaea described to carry a double-membrane system (Comolli et al., 2009). Metagenomics can nowadays deliver near-complete to complete genomes of uncultivated organisms from many environmental samples (Albertsen et al., 2013; Baker et al., 2010; Castelle et al., 2013; Di Rienzi et al., 2013; Erkel et al., 2006; Kantor et al., 2013; Mondav et al., 2014; Nunoura et al., 2011; Tyson et al., 2004; Wrighton et al., 2012). This cultivation-independent technology appeared very useful for studying microbial life in nature particularly when coupled to other technologies for inferring structural and genetic information simultaneously (Baker et al., 2010; Baker et al., 2006).

Here, we integrated metagenomics, metatranscriptomics, focused-ion beam scanning electron microscopy, transmission electron microscopy and immunological analyses to investigate the structural and metabolic features of the SM1 Euryarchaeon. Our analysis targeted the Mühlbacher Schwefelquelle (Germany), where previous investigations revealed SM1 Euryarchaeon to be the dominant organism via biofilm formation [(Henneberger et al., 2006; Probst et al., 2013b); designated SM1-MSI]. Given the similarity and high abundance of a close relative to the SM1 Euryarchaeon from a second site (Crystal Geyer, USA), we additionally conducted a genomic analysis of these organisms (SM1-CG). Our results provide insight into the structural organization and genomic information of a widespread, novel archaeal lineage, which has the capability to outcompete other microorganisms in its subsurface biotope.

Results and Discussion

Different strains of the SM1 Euryarchaeon dominate a sulfidic aquifer in the subsurface. The uncultivated, so-called SM1 Euryarchaeon forms an almost single species biofilm in a cold, sulfidic spring (Mühlbacher Schwefelquelle, close to Regensburg, Germany). The biofilm is constantly washed up from the subsurface (Henneberger et al., 2006). Within the biofilms, the coccoid-shaped archaeal cells were spaced fairly regularly (*Figure III.5-1*), with each cell linked to 1-7 cells (mostly 6 cells) by a dense web of cell-cell contact threads. These connections occasionally appeared like tubes or bars, depending on the amount of surrounding extracellular polymeric substance (EPS) covering the thin surface structures (*Figure III.5-1*). The combination of cells and threads formed a spacious, penetrable, but strongly connected cell-to-cell network.

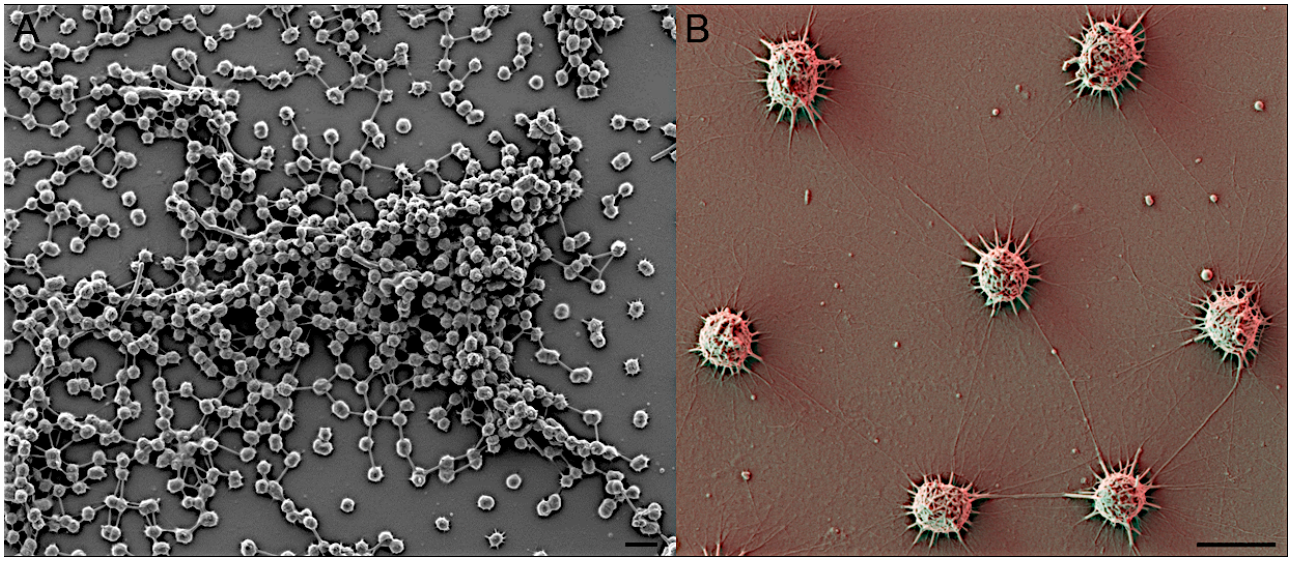

Figure III.5-1 *Scanning electron microscopy of SM1 euryarchaeal biofilm samples.* ***A****) Overview of the biofilm. Cells are connected by cell surface appendages, which appear like tubes due to the surrounding EPS. Bar: 2 µm.* ***B****) Detailed micrograph of SM1 Euryarchaeon cells showing a typical hexagonal pattern in the biofilm. The cell surface appendages are visible due to the low amount of EPS present. Bar: 1 µm.*

The biofilm contained no more than 5% bacteria, mainly sulfate-reducers (Probst et al., 2013b). Besides the SM1 Euryarchaeon, another archaeon [IM-C4 Euryarchaeon (Probst et al., 2013b)] was detected at a low relative abundance level. Cell numbers of IM-C4 were estimated by specific qPCR in nine independent samples to be 0.69±0.67% of all archaeal cells (assuming one 16S rRNA gene per cell). Fluorescence *in situ* hybridization of IM-C4 confirmed that abundances were so low as to preclude microscopic quantification (<1/100 cells) but revealed a coccoid morphology similar to SM1-MSI (Supplementary Figure 1).

As expected based on 16S rRNA gene profiling (Probst et al., 2013b), metagenomic shotgun sequencing revealed two archaeal bins, SM1 Euryarchaeon and IM-C4 Euryarchaeon. Relative abundances of these organisms, based on depth of sampling of the reconstructed 16S rRNA gene sequence, were estimated to be 87.7% for SM1-MSI and 1.3% for IM-C4. Bacterial signatures were mainly affiliated with the Deltaproteobacteria clade, with 3.7% relative abundance belonging to *Desulfocapsa* (phylogenetic tree with relative abundance estimates are depicted in Supplementary Figure 2), in agreement with previous studies (Probst et al., submitted; Probst et al., 2013b). Taxonomic binning of the whole assembly revealed a total of 3,333 kbs of genomic sequence for SM1-MSI (Supplementary Figure 3). The average GC-content was 32.09%, in accordance with experimental data (Henneberger et al., 2006). The SM1-MSI genome bin encoded for 3,294 proteins, a size that greatly exceeds the experimentally determined genome size of <1.9 Mbps (data not shown). However, most of the single-copy genes were present two or more times, indicating the presence of several strains. The analysis of K_aK_s-ratio (synonymous and non-synonymous mutations) in duplicated single-copy genes revealed a statistically significant difference to randomly mutated genes (Supplementary Figure 4, p-value = 0.0004) and provided evidence for the presence of at least two strains in the genomic bin. Consequently, the SM1-MSI bin represents a genomic bin of multiple SM1 euryarchaeal strains present in subsurface biofilms (Supplementary Figure 5). Almost the entire SM1-MSI genome was sampled at least once (completeness 98% based on marker genes; Supplementary Figure 6).

From the Mühlbacher Schwefelquelle site, we also recovered 1.374 Mb of genome sequence of the IM-C4 (GC-content 48.48%; Supplementary Figure 3), which showed a completeness of 90% (Supplementary Figure 6).

Genome reconstruction for a highly abundant SM1 Euryarchaeon from a high CO_2 aquifer. Given 16S rRNA gene sequence evidence that a SM1 Euryarchaeon was abundant in Crystal Geyser (Utah, USA; 17.5 °C), we also analyzed metagenomic datasets from this site. Shotgun paired short read sequences derived from samples collected onto 3 µm filters and post-3.0 µm onto 0.2 µm filters (Emerson et al., submitted). Interestingly, the SM1 Euryarchaeon was the most or second most abundant organism (depending on filter size fraction) in the planktonic microbial community. Despite multiple attempts using different data subsets and samples, all assemblies were quite fragmented; however, only one population was

sampled, based on the marker gene inventory. We attribute the fragmentation to strain sequence variation. Contigs were assigned to the SM1 Euryarchaeon based on the combination of high coverage and overall similarity between predicted protein sequences and previously reported archaeal sequences. The assembled fragments encoded ribosomal protein blocks, the 16S rRNA gene sequence, as well as a significant fraction of the genome. Based on the combination of fraction of recovered marker genes in the best assembly of SM1-CG and length of reconstructed sequence (1.5 Mb), we estimate the genome size to be 1.6 Mb with a completeness of 98%.

SM1 Euryarchaeon is the first characterized representative of a novel and diverse lineage of Euryarchaeota widespread in anaerobic, aquatic subsurface environments. The availability of genomic data for SM1 Eurarchaeon from two sites and one relative, IM-C4, enabled us to investigate their placement in the phylogeny of the archaea by using a phylogenomic approach. A Bayesian phylogeny issued of concatenation of 56 ribosomal proteins from 142 archaeal genomes indicated that SM1-MSI, SM1-CG and IM-C4 form a robust monophyletic cluster, sister to the Methanococcales (*Figure III.5-2*). However, their placement was unstable as frequently observed for very fast evolving lineages with few available close relatives such as the ARMANs or Nanoarchaeota (Brochier-Armanet et al., 2011), although the availability of the relatively distant IM-C4 likely helped avoiding a well-known tree reconstruction artifact by breaking the long-branch leading to SM1-MSI/CG (*Figure III.5-2*). We therefore applied a number of strategies known to alleviate potential tree reconstruction artifacts such as varying taxonomic sampling, recoding strategies, and use of alternative markers (Supplementary Figure 7 and 8). In all these analyses the SM1 lineage consistently branched within Euryarchaeota and displayed only a few alternative placements around early emerging orders (shown in *Figure III.5-2*). Such placement, and the lack of close relationship to a particular archaeal lineage are consistent with a specific profile of DNA replication components (Supplementary Figure 9), including the presence/absence pattern of a few markers that have been recently shown to provide phylogenetic information (Raymann et al., 2014). The distinctiveness of the SM1 lineage and its phylogenetic placement around a few deep euryarchaeal nodes was also confirmed by homology analysis of the whole proteome predicted for SM1-MSI (Supplementary Figure 10). In fact, taxonomic distribution of the closest hits showed a clear dominance of Euryarchaeota, and in particular methanogens.

Interestingly, we estimated that a large proportion of the SM1-MSI genes may have been acquired via horizontal gene transfer from bacteria (Supplementary Figure 11). For 2411 protein sequences maximum likelihood trees were calculated with potential homologues sequences detected in NCBI's non-redundant database (NR). 851 of these SM1-MSI protein sequences showed a monophyletic branch with bacterial sequences only, indicating that ~25.8% of genes were potentially acquired via horizontal gene transfer.

16S rRNA analysis indicated that SM1-MSI, SM1-CG and IM-C4 are representatives of a widespread and diverse archaeal group. The SM1 Euryarchaeon clusters phylogenetically with environmental 16S rRNA gene sequences from hot and cold springs, aquifers, ponds and deep-sea environments (*Figure III.5-3*). 16S rRNA gene sequence surveys targeting other springs close to the Mühlbacher Schwefelquelle detected other nearly single species biofilms dominated by SM1-MSI [with 100% 16S rRNA gene identity (Probst et al., submitted)]. Moreover, the SM1-MSI sequence was most closely related to another SM1 Euryarchaeon representative (~98% gene identity) derived from an archaeal population recovered from a CO_2 driven cold-water geyser termed SM1-CG (see above). More distant relatives of the SM1-MSI (13% difference in 16S rRNA gene sequence) were detected in hot springs (58.5°C) in the USA, described to be rich in hydrogen (Chapelle et al., 2002), and in Bulgarian hot springs [(Tomova et al., 2011) 79°C; 11% difference in 16S rRNA gene sequence]. The 16S rRNA gene sequence of the IM-C4 Euryarchaeon places this organism with sequences from archaea from deep-sea sediments, hot springs, aquifers, freshwater ponds and plant reservoirs, fairly close to SM1-MSI (75.9% 16S rRNA gene identity; *Figure III.5-3*).

Altogether, our phylogenomic analyses indicate that the SM1-MSI, SM1-CG, IM-C4 represent a new euryarchaeal order-level lineage of anaerobic, aquatic subsurface archaea, for which we propose the candidatus name "Altiarchaeales" (candidatus family "Altiarchaeaceae") indicating their origin in the subsurface. For the SM1 Euryarchaeon, the name "*Candidatus* Altiarchaeum hamiconnexum" is proposed.

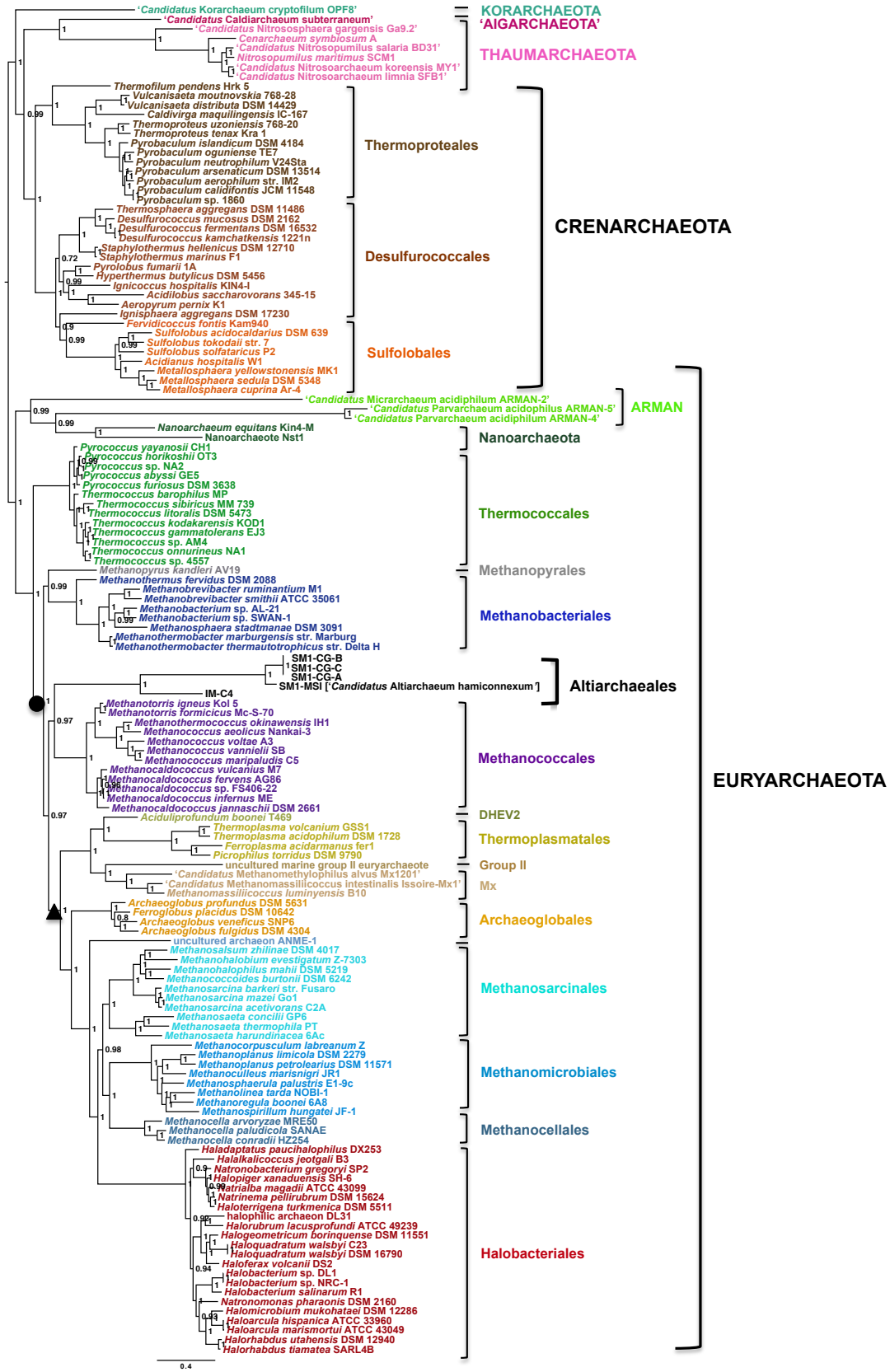

Figure III.5-2 *Bayesian phylogeny based on 56 concatenated ribosomal proteins (for details see Material and Methods and Results and Discussion). The SM1, IM-C4 and CGR-A/B/C represent a distinct euryarchaeal lineage related to Methanococcales. Numbers at nodes represent posterior probabilities and the scale bar represent average number of substitutions. Two alternative placements for the SM1 lineage are indicated: one obtained by Dayhoff6 recoding of the r-protein dataset (triangle), and one from concatenation of 10 universal markers (circle; please see Supplementary Figure 7 and 8).*

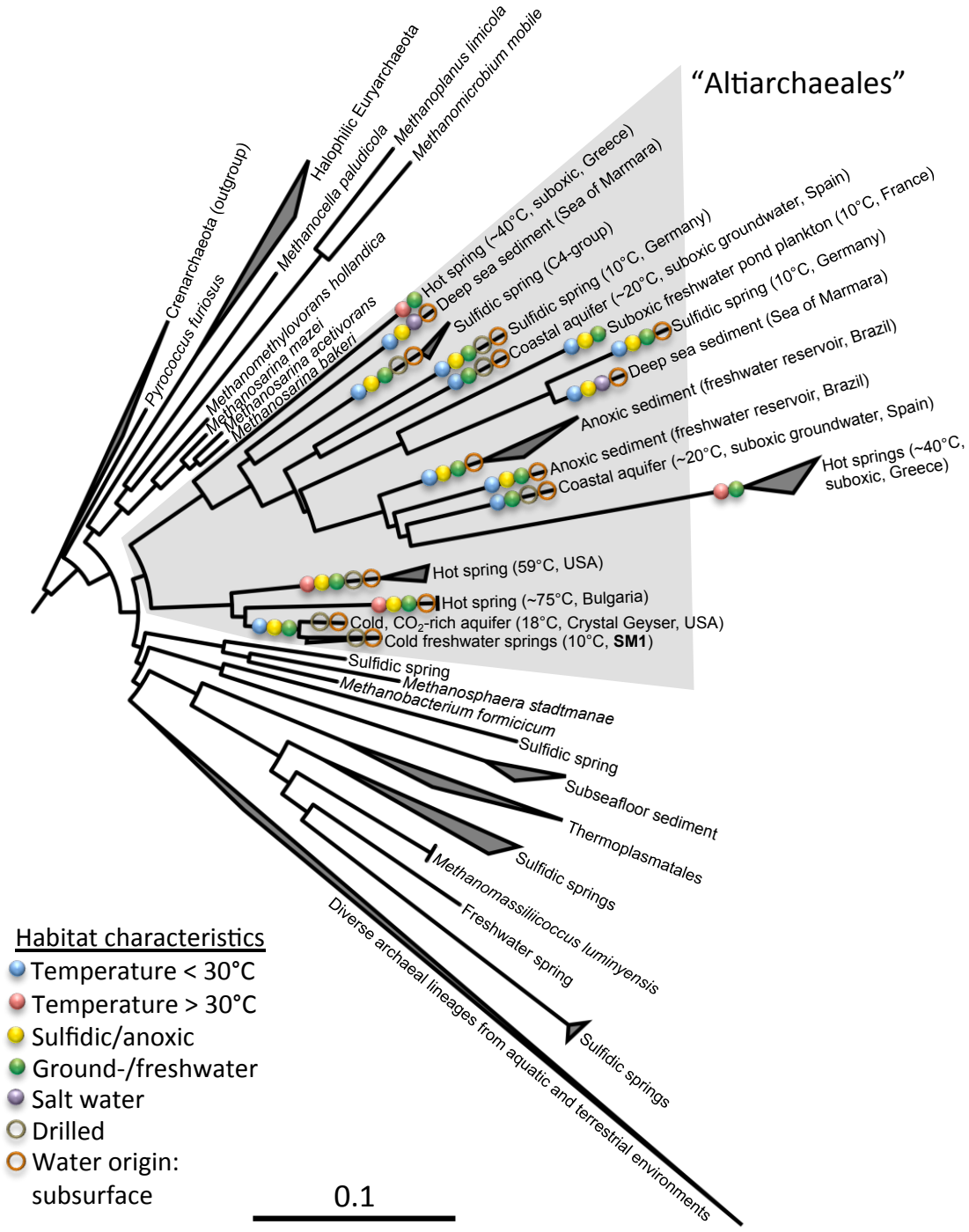

Figure III.5-3 *Distribution and diversity of SM1-related 16S rRNA sequences. The phylogenetic tree is based on 16S rRNA gene sequences of various studies (for accession numbers please see Supplementary Table 1). The SM1 Euryarchaeon is the representative of a novel archaeal candidate order termed "Altiarchaeales", shadowed in grey.*

"Candidatus Altiarchaeum hamiconnexum" features a distinct glycosidic diether lipid profile. In order to understand the biology of the candidatus order "*Altiarchaeales*" we linked genomic information of the SM1 Euryarchaeon to intact polar lipid, (ultra)structural, immunological and transcriptomic analyses of the SM1-MSI biofilm. Unless stated otherwise, the genomic information presented holds true for both genomes, SM1-MSI, from the Mühlbacher Schwefelquelle, and SM1-CG, from Crystal Geyser. The major pathways discussed in this manuscript are summarized in *Figure III.5-4* and corresponding annotations are provided in Supplementary Table 2.

The SM1 Euryarchaeon genome encoded for the synthesis of ether lipids via the mevalonate pathway (*Figure III.5-4* and Supplementary Table 2). Lipid analysis via HPLC-ESI-MS of intact polar lipids (IPLs) in biofilm sample taken from the Mühlbacher Schwefelquelle revealed a dominance of archaeal (94%) versus bacterial lipids (6%, see below). The archaeal IPLs consist of glycosidic diether lipids with (i) two phytanyl chains (C_{20}-C_{20} archaeol, Ar) and (ii) a combination of one phytanyl and one sesterterpanyl chain (C_{20}-C_{25} extended archaeol, Ext-Ar), which is in accordance with genomic information depicted in *Figure III.5-4*. Archaeal diethers were composed of monoglycosyl (1G) and diglycosyl (2G) headgroups, and a tentatively identified pentose-hexose headgroup. The major IPLs are 1G-AR and 2G-AR, representing 34% and 57% of total archaeal lipids (Supplementary Figure 12). The detected lipid profile is neither congruent with these from methanogens [(Koga et al., 1993); including ANME (Rossel et al., 2008)] nor with those from Thermoplasmatales (Huber and Stetter, 2006).

"*Candidatus Altiarchaeum hamiconnexum": A continuously dividing archaeon with double-membrane and nano-grappling hooks.* SM1 Euryarchaeon genomes did not encode for genes involved in motility, flagellation, S-layer or *pili* production. However, both genomes from the German and USA sites encoded genes for *hamus* subunits (plural *hami*). *Hami* are specialized cell surface appendages with barb-wire like filamentous structures and nano-grappling hooks at their distal end (diameter approx. 60 nm; *Figure III.5-5*). These appendages have only been described for SM1-MSI so far (Moissl et al., 2005b; Probst et al., submitted) and possess no homologs in public databases. Genes potentially encoding for the major *hamus* subunit were identified by searching predicted protein sequences with experimentally determined *hamus* protein sequences (Moissl et al., 2005b). The gene sequences of corresponding subunits revealed the presence of sec signal peptides and thus seem to be secreted via the Sec pathway into the periplasm (*Figure III.5-4*, Supplementary Table 2).
The anchoring of the *hami* in SM1-MSI archaea was assessed by different electron microscopy-based techniques. Thin sectioning and preliminary electron tomographic analysis revealed that they are most likely anchored within the cell wall, the cytoplasm end perhaps even in an additional basal structure within the cytoplasm (*Figure III.5-5*). As it could be shown, the cell wall consists of an outermost cellular membrane upon the cytoplasmic membrane, spanning a periplasmic space of about 30 nm (*Figure III.5-6*). No peptidoglycan- or pseudopeptidoglycan-layer was identified in ultrastructural analysis, nor any associated biosynthesis proteins were detected in the genomic dataset. The SM1 euryarchaeal cell wall is thus divided into inner membrane, periplasm and outermost membrane. Thin sections of the cells further revealed the presence of a thick extracellular polymeric substance (EPS-)layer and the *hami* forming a dense network around the cells. The *hami* were anchored in both membranes, with a protrusion into the cytoplasm ("thorn", *Figure III.5-5 D/E*). Based on electron microscopic analysis of membrane-released *hami*, we propose a model for the assembly of the *hamus* structure (*Figure III.5-5A*), which is highly similar to the type IV *pili* formation in bacteria, involving the outer membrane (Albers and Pohlschroder, 2009). Although it seems obvious that other proteins are involved in *hamus*-structure assembly (e.g. outer membrane channel, thorn), these proteins could not be identified via genomic analyses so far.
Several genes involved in cell replication were identified (*Figure III.5-4*), of which the *FtsZ* gene was of special interest as it is typical for Bacteria and Euryarchaeota (Makarova et al., 2010). FtsZ proteins are involved in cell division, constantly present but only aggregated when cells divide (Den Blaauwen et al., 1999). These proteins may give rise to electron-dense rings localized in the cytoplasm underneath the cell membrane (*Figure III.5-7A*) in focused-ion-beam scanning electron microscopy tomography (FIB-SEM) images of the SM1-MSI biofilm cells (*Figure III.5-7B*). Nearly all cells reconstructed using FIB-SEM were observed to possess these rings. In case of dividing cells, these rings were associated with the septum of the SM1 euryarchaeal cells indicating a relation to cell division or even being FtsZ. In subsequent immuno-staining experiments with antibiodies against FtsZ homologues, the SM1 Euryarchaeon cells were shown to accumulate FtsZ between separating cells (*Figure III.5-7C/D/E*). Consequently, SM1 euryarchaeal cells in the biofilm samples were under active division when harvested. Indeed, 73.2±9.6% of the cells in the biofilm showed the state of division (data not shown). Due to the presence of FtsZ rings and the fact that 95% of the Archaea are alive at the time of sampling (Henneberger et al., 2006), one can assume that the entire biofilm archaeome is highly active in metabolism and proliferation. The continuous cell division could be the reason for the aggregation of mainly silent DNA replication errors in coding and non-coding sequences. This may account for the genetic versatility of the SM1-MSI cells and result in the strain variation that complicated genome recovery (see above).

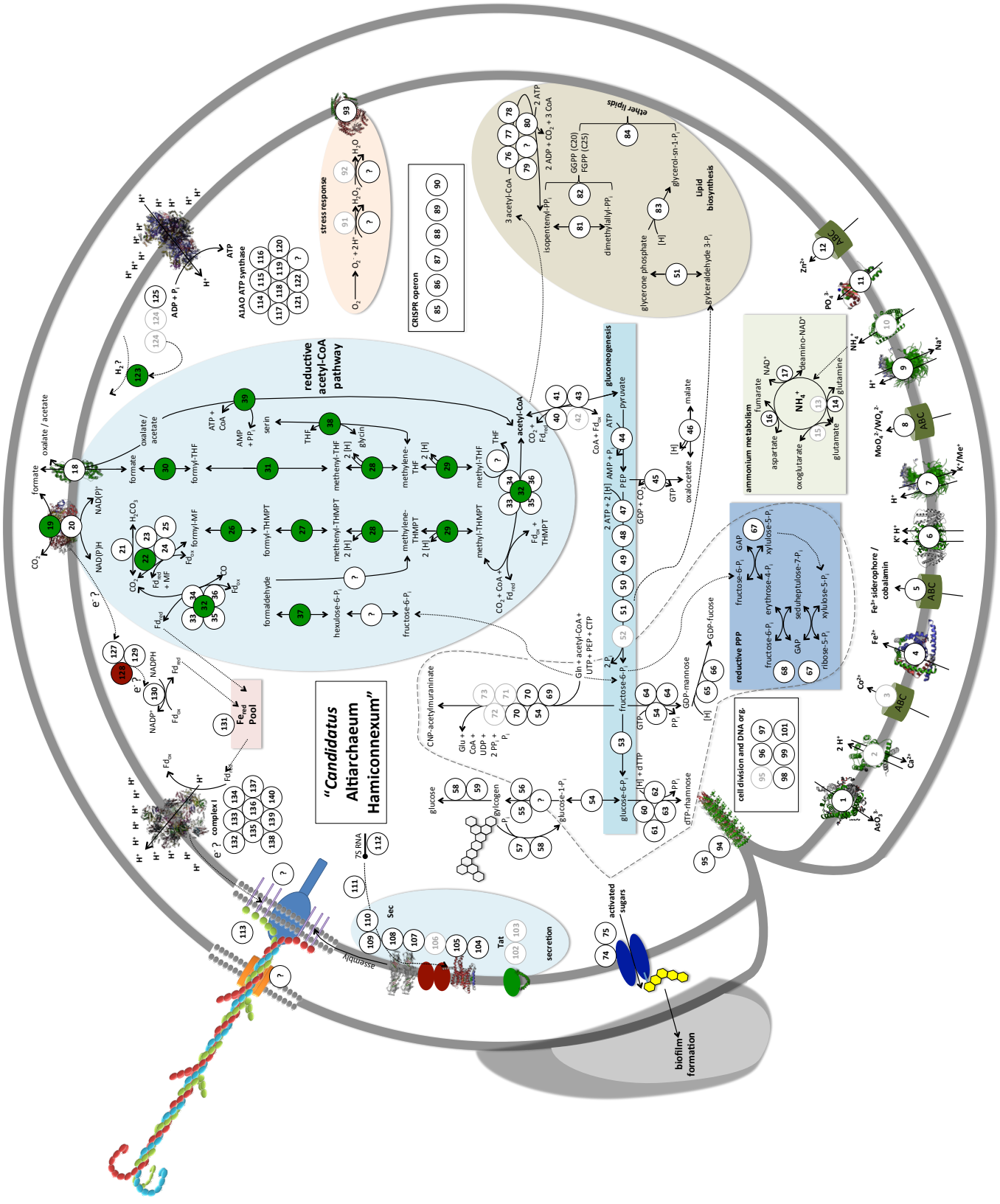

Figure III.5-4 *Metabolic pathway map of genes encoded in the SM1 euryarchaeal pangenome. Membrane proteins are indicated by symbols or protein structures of subunits (taken from Pfam database). Predicted proteins are indicated by numbers, which can be found in Supplementary Table 2. Specific enzymes were tested for presence of corresponding mRNA: Enzymes labeled in green were also detected in the mRNA pool of the biofilm (SM1-MSI), while enzymes with red color were absent in mRNA pool. Interrogation marks indicate potential metabolic pathway reactions that are likely present but are lacking evidence in the fragmented genomic bins. Black numbers indicated the presence of predicted enzymes in both SM1-MSI and SM1-CG. Grey numbers indicate presence in SM1-MSI only. A version with high resolution is available in the supplementary PDF version on the supporting DVD.*

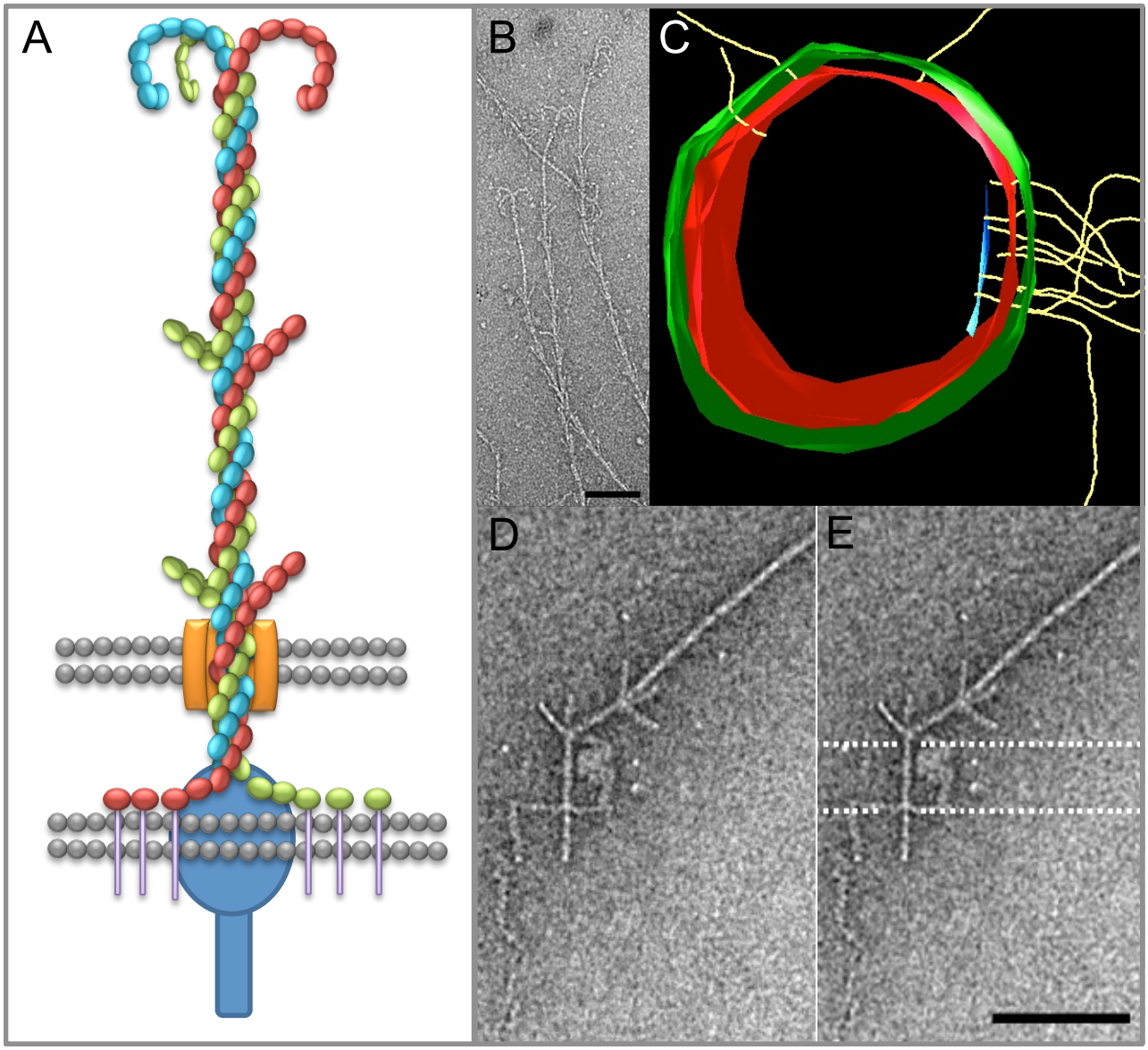

Figure III.5-5 *Hamus formation, structure and anchorage in SM1 euryarchaeal double membrane.* ***A****) Assembly and formation model, based on microscopical and genetic data. The individual threads are formed at the inner membrane, and the hamus structure is built in the periplasm, before being pushed through the outer membrane. The first prickle appears right outside of the cell.* ***B****) Electron micrograph of hamus filaments (negatively contrasted) as found in the biofilm. Bar: 100 nm.* ***C****) 3D model resulting from an electron tomogram of a SM1 euryarchaeal cell, showing the arrangement of hamus fibres and their protrusion ("thorn") into the cytoplasm. Blue, anchoring structure; green, outer cellular membrane; red, cytoplasmic membrane; yellow, hami fibres.* ***D, E****) Hamus anchoring structure as visualized after release of the hami by dissolving the membrane. In E, the positions of the two membranes are indicated (white line). Bar: 100 nm.*

"*Candidatus Altiarchaeum hamiconnexum" performing an acetoclastic metabolism?* Since the Mühlbacher sulfide spring, the natural environment of SM1 Euryarchaeon, is poor in organic substances, it is not surprising that genes for the transport and assimilation of complex organic compounds, amino acids and sugars are missing in the genomes (*Figure III.5-4*, for details on transporter and ammonia assimilation please see supplementary information). Instead, SM1 metabolism pivot on activated acetic acid, acetyl-CoA, which appears to be a starting point of the biosynthesis of cellular building blocks (*Figure III.5-4*), similarly to many other anaerobes (Fuchs, 2011).

SM1 genomes possessed most of the genes for the reductive acetyl-CoA (Wood-Ljungdahl) pathway, and these genes appear to be transcribed *in vivo* as depicted in *Figure III.5-4* (green: mRNA detected, red: no mRNA detected). In contrast, genes for the key enzymes of other autotrophic CO_2 fixation pathways are missing. As the reductive acetyl-CoA pathway in methanogenic archaea, the SM1 pathway makes use of methanofuran and tetrahydromethanopterin (THMPT) as the C_1-carrieres. However, the SM1-MSI and SM-CG contain neither genes involved in the methane production from methyl-THMPT nor genes for the biosynthesis of the key cofactors involved in this process (coenzyme M and coenzyme B).

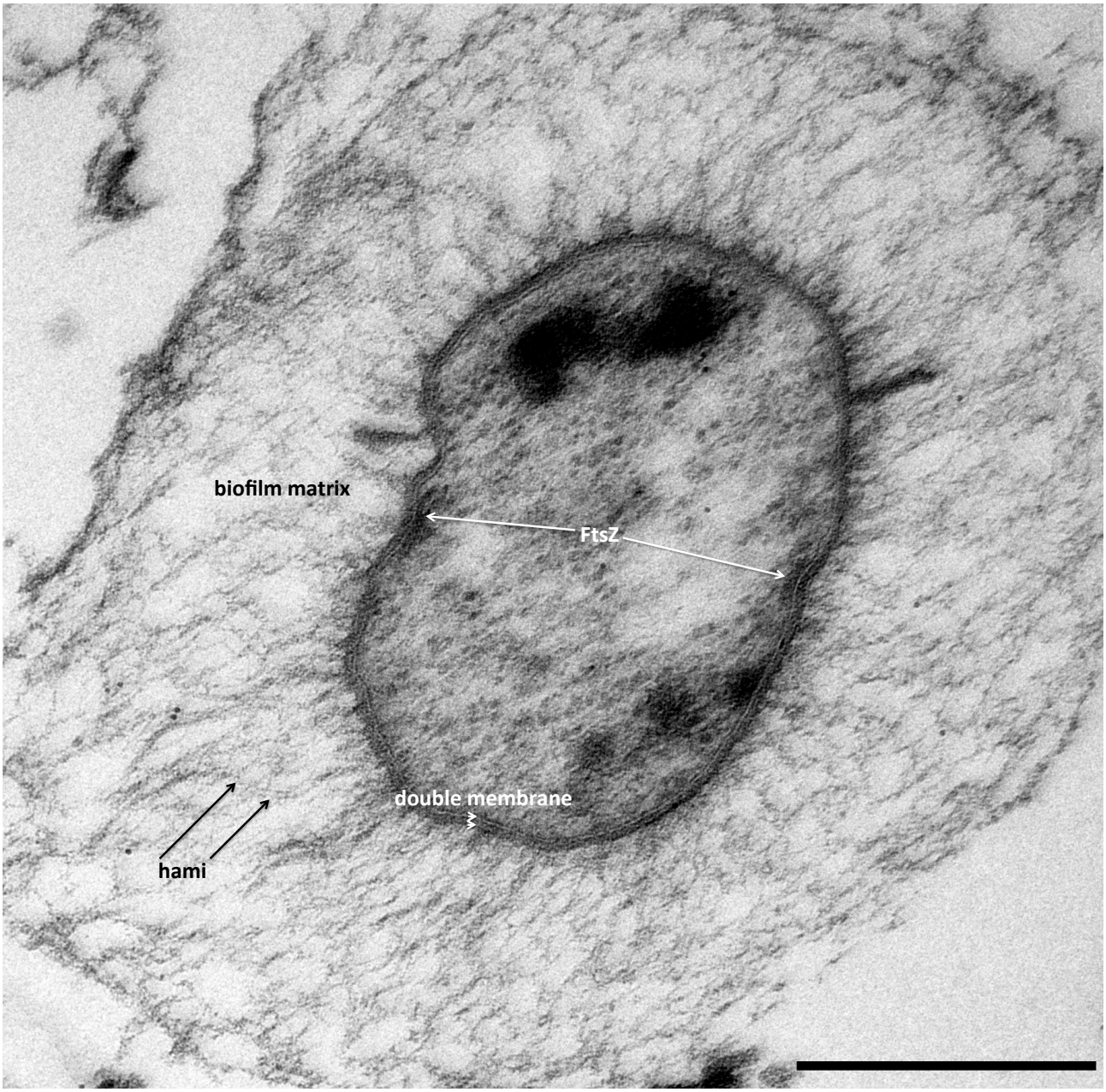

Figure III.5-6 *Electron micrograph (ultrathin section) of a SM1 euryarchaeal cell in the biofilm after high-pressure freezing (Wohlwend HPF Compact 02). The cell is surrounded by an EPS matrix and cell surface appendages (hami), which go beyond the matrix. The cells show two membranes with a faint periplasm. FtsZ rings (Figure III.5-7) are located at the inner membrane. Bar: 500 nm.*

Furthermore, the genes for the biosynthesis of factor 420 (F_{420}) are also missing in the pangenome, and the presence of F_{420} could not be detected in the SM1 cells (Moissl et al., 2003). In methanogenic archaea, $F_{420}H_2$ is used in methylene-THMPT-dehydrogenase and methylene-THMPT-reductase reactions. In SM1, the archaeal F_{420}-dependent THMPT dehydrogenase is replaced by a NAD(P)-dependent enzyme of potential bacterial origin, known mainly from methylotrophic bacteria [(Vorholt et al., 1998); the other identified genes, including the key enzyme CO dehydrogenase/acetyl-CoA synthase, are of archaeal origin, please see Supplementary Table 3]. For the reduction of methylene-THMPT, NAD(P)-dependent enzymes have not been shown yet but indications were found that this reaction is compensated by a methylene-THF reductase, also present in SM1 Euryarchaeon's genome (Stokke et al., 2012). The "classical version" of the reductive acetyl-CoA pathway is fully reversible and can function for both acetyl-CoA synthesis and oxidation (Schauder et al., 1988).

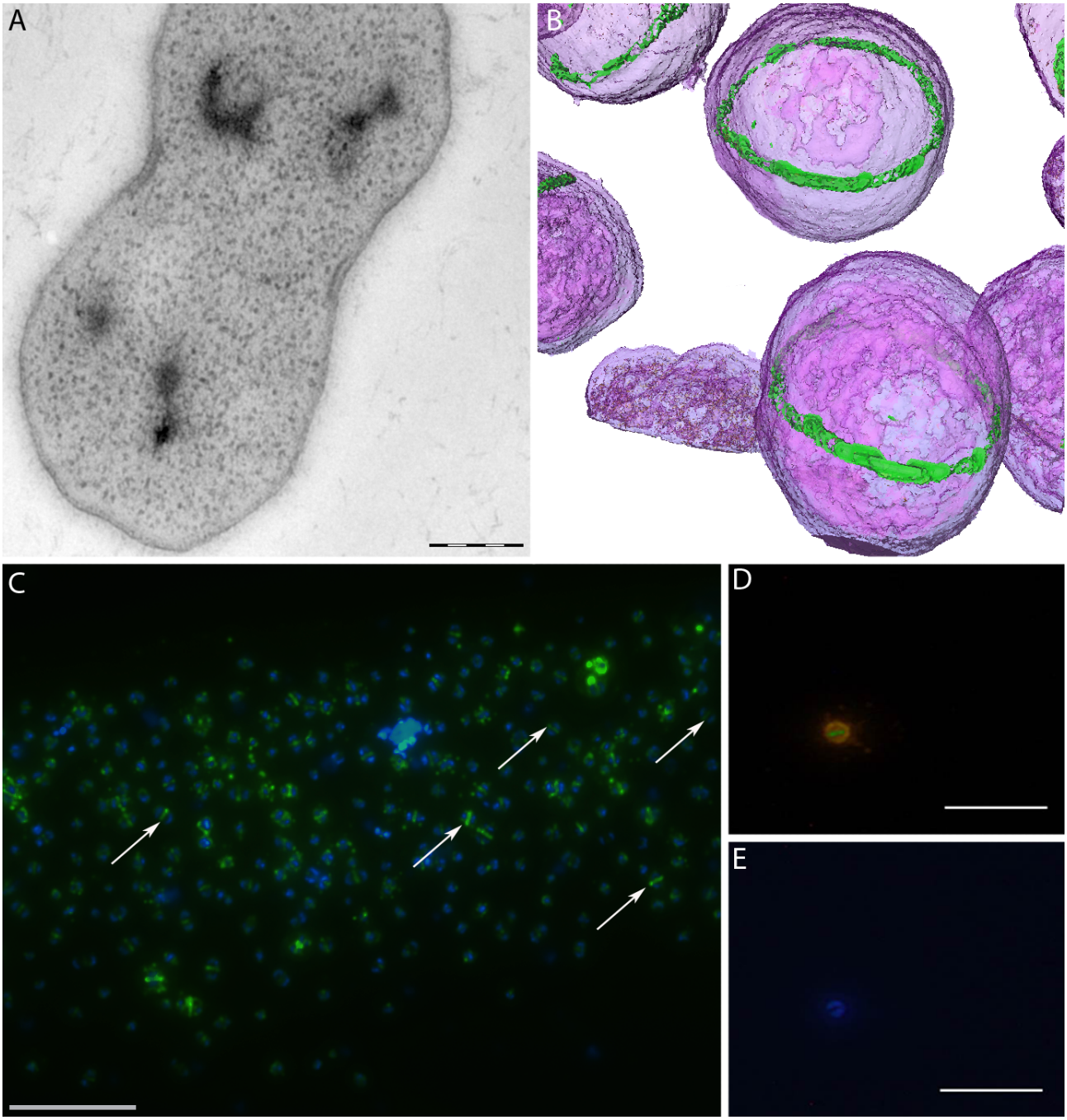

Figure III.5-7 *FtsZ and its localization in SM1 euryarchaeal cells indicating a continuous division of cells in the subsurface biotope.* ***A****) Thin section of a dividing SM1 cell. The high-pressure freezing was performed with the Leica HPM 100. An electron-dense zone is observed beneath the cytoplasmic membrane at the septum. Bar: 200 nm.* ***B****) Three-dimensional reconstruction (FIB-SEM) of SM1 cells in the biofilm. Electron-dense parts are labeled in green, cells walls are violet.* ***C****) Fluorescence micrograph of a biofilm sample showing DNA in blue (DAPI) and FtsZ (labeled with FtsZ antibody) in green. FtsZ is located in the septum area of dividing cells. For an image with higher contrast please see PDF version in supplementary. Bar: 10 µm.* ***D****) Single SM1 Euryarchaeon cell, labeled with anti-FtsZ (green) and anti-Hami (yellow) antibodies showing that FtsZ is expressed in SM1 Euryarchaeon cells. Bar: 5 µm.* ***E****) Same micrograph as in D but DAPI staining is indicated in blue. Bar: 5 µm.*

The redox potential of $F_{420}/F_{420}H_2$ (–360 mV) is 40 mV more negative than that of $NAD(P)^+/NAD(P)H$ [–320 mV; (Walsh, 1986)]. Therefore, the substitution of F_{420}-dependent reactions with the NAD(P)-dependent ones suggests that the reductive acetyl-CoA pathway in SM1 Euryarchaeon functions rather for acetyl-CoA catabolism than for autotrophic CO_2 fixation. This hypothesis is supported by the fact that a homolog of the methylene-THMPT dehydrogenase was shown to catalyze the oxidation of the methylene group in *Methylobacterium extorquens* AM1 (Vorholt et al., 1998). Alternative pathways of acetyl-CoA oxidation, the tricarboxylic acid (TCA) cycle and the glyoxylate cycle, are incomplete in SM1 bins.

The functioning of the reductive acetyl-CoA pathway requires strictly anaerobic conditions. Groundwater often has a high fluctuation of oxygen concentrations, and rubrerythrin and associated oxygen stress response enzymes are encoded in SM1 genome, although no genes for catalase or superoxide dismutase were detected (*Figure III.5-4*). The SM1 biofilm occurs in an oxygen-depleted groundwater flow, which is protected from oxygen input by the high sulfidic content enabling the Archaeon to perform a strictly anaerobic metabolism (Probst et al., 2013b; Rudolph et al., 2004).

Where could acetyl-CoA come from? The conversion of fatty acids, carbohydrates or amino acids into acetyl-CoA could be ruled out: The SM1 pangenome did not code for β-oxidation enzymes, because the genes for electron transferring flavoprotein, acyl-CoA dehydrogenases and crotonases are missing. No indications for a fermentation-based metabolism or for amino acids utilization were found. Furthermore, the genome lacks enzymes of carbohydrate utilization. Instead, the presence of a unidirectional gluconeogenetic fructose-1,6-bisphosphate aldolase/phosphatase (Say and Fuchs, 2010) suggests that the metabolic fluxes are oriented from acetyl-CoA to phosphoenolpyruvate (via ferredoxin-dependent pyruvate synthase and phosphoenolpyruvate synthase) and further to fructose-6-phosphate, and not the other way around (*Figure III.5-4*).

Yet, SM1 genomes contain a homologue of AMP- and pyrophosphate-forming acetyl-CoA synthetase, which was transcribed in biofilm samples taken from Mühlbacher Schwefelquelle. This is the only enzyme responsible for the activation of a carboxylic acid encoded in the SM1 genome. Neither genes for ADP-forming synthetases (including the TCA cycle enzyme succinyl-CoA synthetase) nor CoA-transferase and phosphotransacetylase/acetate kinase homologues were identified. Pyrophosphate formed in the acetyl-CoA synthetase reaction could only be hydrolyzed by a soluble pyrophosphatase, whereas genes for a membrane-bound pyrophosphatases are missing. Therefore, an activation of acetate is accompanied by a conversion of ATP into AMP and two phosphates, i.e. by a hydrolysis of two high-energy bonds of ATP. This makes this reaction irreversible. Interestingly, acetoclastic methanogens utilizing AMP-forming acetyl-CoA synthetases (*Methanosaeta* sp.) have very high affinity to acetate (5-20 μM) in comparison to those utilizing acetate kinase/phosphotransacetylase like *Methanosarcina* [1 high energy bond of ATP pro 1 acetyl-CoA formed, minimal threshold concentration of acetate ~1 mM (Smith and Ingram-Smith, 2007)]. The concentration of dissolved organic carbon (including acetate) in sulfidic spring waters, where the SM1 Euryarchaeon flourishes, was in general very low (1.3 mg/l on average), thus probably stipulating the necessity to invest more energy in acetate activation.

The gene for acetyl-CoA synthetase is located next to a gene encoding a transporter of a major facilitator superfamily with the similarity to a formate/oxalate anitporter. It tends to speculate that this protein might be involved in acetate transport, possibly associated to formate antiport. Formate production from CO_2 could be coupled, e.g., with CO dehydrogenase reaction [(Schuchmann and Muller, 2013); note, that CO is known to be globally present in groundwater ecosystems (Chapelle and Bradley, 2007)]. Formate transported in the periplasmatic space might be oxidized by a membrane-bound formate dehydrogenase present in the SM1 pangenome but its catalytic site appears to face the periplasm. This would lead to the liberation of two protons and formation of a proton gradient.

Energy conservation in "Candidatus Altiarchaeum hamiconnexum" and possible electron acceptor. The oxidation of reduced ferredoxin formed during acetyl-CoA (or CO) oxidation might also lead to the generation of the proton gradient mediated by a complex-I-like oxidoreductase complex similar to the Rnf complex [(Hess et al., 2013) (*Figure III.5-4*)]. The genes of this complex (missing the NADH accepting module) had been discovered in groundwater sediment microbial genomes (Castelle et al., 2013) and are hypothesized to function with ferredoxin (Battchikova et al., 2011; Pereira et al., 2011). Another pathway of energy conservation via reduced ferredoxin is electron bifurcation (Buckel and Thauer, 2013; Pereira et al., 2011). The presence in SM1 genome of homologues of heterodisulfide reductase genes (of a bacterial origin) encoded next to a gene for a putative ferredoxin-NADPH-reductase suggests a possibility of electron bifurcation. However, these genes were neither present in MSI-CG nor expressed in the biofilm samples (SM1-MSI), as shown for the subunit B of heterodisulfide reductase gene (*Figure III.5-4*).

Besides acetate, hydrogen could be a potential electron donor for SM1 growth. However, the SM1 genome did not possess genes encoding hydrogenases [just like an ANME-1 genome (Meyerdierks et al., 2010)]. Although homologues of the electron shuttling [NiFe]-hydrogenase subunits were encoded (and transcribed) in SM1-MSI (*Figure III.5-4*), neither catalytic subunits of [NiFe]-hydrogenase nor any [FeFe]-hydrogenases were identified. Nevertheless, genes for hydrogenase

maturation enzymes like HypA, which help to incorporate nickel (Hube et al., 2002; Olson et al., 2001), were present in the SM1-MSI pangenome. They might be involved in the maturation of CO-dehydrogenase/acetyl-CoA synthase.

What is a possible electron acceptor for the suggested acetate oxidation in the SM1 Euryarchaeon? It is compelling to believe that the cell surface appendages (*hami*) of SM1, whose genes are encoded next to metabolic genes (CO dehydrogenase/acetyl-CoA synthase), may act as a potential electron transporter to the outside of the cell. There, potential electron acceptors could be ferrous iron [as it was shown by *Geobacter sulfurreducens* (Reguera et al., 2005)] or other microorganisms like sulfate-reducing bacteria closely associated with SM1-MSI (Probst et al., 2013b). However, experiments with SM1 biofilms did not show an increased conductivity compared to the spring water (Supplementary Figure 13) enabling to conclude that *hami* may not mediate electron transfer to an external acceptor.

Mühlbacher spring water has high concentrations of sulfate and thiosulfate, although the concentrations of nitrate and nitrite were below detection limit (Rudolph et al., 2004). In the SM1 Euryarchaeon genome, we identified neither membrane-bound enzymes transferring electrons to a potential final electron acceptor (like nitrate or nitrate) nor homologs of cytochromes. With regard to dissimilatory or assimilatory sulfate reduction, only a homologue of the adenylyl sulfate kinase was found. In this respect, SM1 Euryarchaea are similar to anaerobic methanotrophic archaea of the ANME-1 group. Moreover, bacterial lipids in the biofilm were diacyl- and acyl-ether-glycerol lipids, which have been reported to constitute major core lipids of cultured sulfate-reducing bacteria (Sturt et al., 2004), and of those inhabiting methanotrophic sediments (Rossel et al., 2008), where electron transfer between ANME and sulfate-reducing bacteria is still enigmatic (Meyerdierks et al., 2010; Milucka et al., 2012; Stokke et al., 2012). Note that ANME-1 is the lineage with the greatest number of similar protein sequences compared to SM1-MSI (13% of all proteins, Supplementary Figure 10). Interestingly, anaerobic methanotrophs of ANME-2 group are capable of sulfate reduction to zero-valent sulphur compounds via an unknown mechanism (Milucka et al., 2012). Apparently, at least one of the great majority of proteins that remain uncharacterized in SM1, may encode for an enzyme transferring electrons to a final acceptor.

Biosynthetic reactions in "Candidatus Altiarchaeum hamiconnexum". Gluconeogenesis products glucose-6-phosphate and fructose-6-phosphate are used for the synthesis of a variety of activated forms of sugars, like rhamnose, mannose, fucose and acetylmuraninate. These sugars could either serve the many glycosyltransferases (44 in total) found in the genome or be precursors for the synthesis of the surrounding biofilm matrix (*Figure III.5-6*), which was shown to be comprised of carbohydrates (Henneberger et al., 2006). Noteworthy, glucose-6-phosphate is a precursor for a storage product glycogen, which is a great resource for activated sugars for biofilm synthesis. Glycogen can afterwards be either degraded by an amylase or a phosphorylase homologue. Since the SM1 bins do not code for a potential hexokinase, the degradation of glycogen via amylase appears to be a metabolic impasse. Phosphorolytic cleavage of glycogen is more reasonable than hydrolytic cleavage, even though phosphorylases encoded in the SM1 genome have only low similarity with the characterized ones (*Figure III.5-4*). Pentose phosphates are synthesized in SM1 through either a reversed ribulose monophosphate pathway (Kato et al., 2006; Sakai et al., 1999; Yanase et al., 1996; Yurimoto et al., 2005) or via transketolase and transaldolase reactions (*Figure III.5-4*). Formaldehyde synthesized in the hexulose-6-phosphate synthase reaction of the ribulose monophosphate pathway can be scavenged in spontaneous reaction with THMPT or tetrahydrofolate (THF) or through the oxidation with one of the aldehyde dehydrogenases present in SM1 genomes. Apart from the genes for THMPT-dependent enzymes, SM1 genome encodes several THF-dependent enzymes, which are transcribed in the biofilm samples. The corresponding enzymes probably provide C_1-units for purine biosynthesis as in case of *Methanosarcina barkeri* (Buchenau and Thauer, 2004). The synthesis of glutamate precursor 2-oxoglutarate proceeds probably through *re*-citrate synthase, aconitase and isocitrate dehydrogenase.

Conclusion

Subsurface microbial life harbors a great proportion of unexplored lineages. Our study expanded this knowledge by the means of a novel archaeal lineage that was genomically and ecophysiologically characterized. "*Candidatus* Altiarchaeum hamiconnexum", is a representative of an archaeal clade comprised solely by environmental 16S rRNA gene sequences. These sequences were discussed to belong to methanogenic archaea that dominate a subsurface environments (Chapelle and Bradley, 2007). However, the metagenomic and metatranscriptomic analysis presented herein enabled the exclusion of

methanogenesis for this lineage of archaea and are indicative of an acetoclastic metabolism adapted to low substrate concentrations by using a novel archaeal acetyl-CoA pathway. Their adaptation to strictly anaerobic environments empowers them to dominate subsurface microbial life, which has so far not been observed for any other archaeal lineage. Particularly biofilm formation as observed for the Mühlbacher Schwefelquelle archaea, appears to bring major advantages. The biofilm could act as a filter for substrates seeping into the subsurface water current from higher sediment layers. The regular distance of "*Candidatus* Altiarchaeum hamiconnexum" cells in the biofilm mediate by biofilm matrix and *hami* may simultaneously act as an improved way for product removal by the underwater current. Accessibility of the biofilm by other microorganisms appears to be prevented retaining them from "*Candidatus* Altiarchaeum hamiconnexum's" precious nutrients.

Material and Methods

16S rRNA gene based analysis of diversity and distribution of the SM1 group. During a survey of 10 sulfidic and non-sulfidic springs in Bavaria, several archaeal 16S rRNA gene sequences were obtained (details not shown) and all of them were submitted to GenBank (KJ566428 - KJ566522). These sequences, as well as previously obtained archaeal sequences from biofilms and microbial communities from Sippenauer Moor and Mühlbacher Schwefelquelle and relative sequences available in public databases [SILVA (Pruesse et al., 2007) and NCBI's NR] were used to assess the phylogenetic position and ecological distribution of SM1 Euryarchaeon relatives. All sequences used (Accession numbers are provided in Supplementary Table 1) were aligned to a 50,000 character alignment using SINA (Pruesse et al., 2012). Tree reconstruction limited to the alignment positions 8,194 to 28,438 and only sequences spanning this region were included. Calculation was performed using the Maximum Likelihood algorithm (Stamatakis et al., 2005) as implemented in the ARB software (Ludwig et al., 2004). The resulting tree was rendered in iTOL (Letunic and Bork, 2007).

Sampling, DNA extraction, quality control for shotgun sequencing. Samples were taken from the sulfidic spring called Mühlbacher Schwefelquelle, Isling, (MSI) near Regensburg (coordinates: 48.98 N, 12.13 E) at a depth of 1 m through a time period of four months (Probst et al., 2013b). A two-side opened Schott flask allowing a continuous flow through of spring water (approximately 5000 liters per hour) was used to filter biofilm droplets with a polyethylene net (Probst et al., 2013b). Samples were transported to the laboratory on ice, concentrated via centrifugation (10 min, 13.8k x g) and stored at -80°C. Metagenomic DNA was extracted using MoBio Power Biofilm DNA kit (MoBio, Carlsbad, USA). Quality with regard to length of metagenomic DNA was verified using agarose gel electrophoresis and the quantity of archaea and bacteria in the samples was determined via qPCR as described earlier (Probst et al., 2013b). Metagenomic DNA was subjected to paired-end sequencing at LGC Genomics, Berlin (Germany). Using two independent sequencing platforms, Illumina HiSeq 2000 and Roche 454 FLX Titanium, 426,869,764 (average length: 100 bases) and 246,234 reads (average length 342 bases) were generated, respectively (LGC Genomics, Berlin, Germany). For the Illumina reads LGC Genomics carried out clipping of sequencing adapters, quality filtering and filtering for paired reads.

Reconstruction of 16S rRNA genes and abundance estimation. 16S rRNA genes were reconstructed from Illumina sequences using Emirge (Miller et al., 2011). The reference database of Emirge [SILVA (Pruesse et al., 2007) database SSURef_108_NR, sequences clustered by uclust (Edgar, 2010) to 97% similarity] was supplemented with archaeal 16S rRNA genes of SM1-MSI and IM-C4 that could be successfully recovered during metagenome assembly (see below). 16S rRNA genes successfully reconstructed by Emirge were first classified using a Bayesian classifier [(Schloss et al., 2009; Wang et al., 2007) against a manually curated 16S rRNA gene database (DeSantis et al., 2006; McDonald et al., 2012) available at http://www.secondgenome.com/go/2011-greengenes-taxonomy supplemented with various sequences from sulfidic freshwater spring] and taxonomic affiliations were compared to those of Emirge's sister 16S rRNA gene sequences in order to ensure the correct classifications. Reconstructed 16S rRNA genes were aligned using SINA (Pruesse et al., 2012), which removed low quality sequences, and checked for chimeric content (Schloss et al., 2009). Chimeric sequences and suspicious sequences with lower 16S rRNA gene query coverage than 80% [blastn against NCBI's NR (Altschul et al., 1990)] were removed from the data set. A tree of recovered and sister sequences was computed using FastTree v2.0 (Price et al., 2010) and rendered in iTOL (Letunic and Bork, 2007) to display relative abundances calculated by Emirge.

Absolute abundance of IM-C4 Euryarchaeon in biofilm samples was measured by quantitative PCR [344af 5'-ACGGGGYGCAGCAGGCGCGA-3' (Casamayor et al., 2002); IM-C4_r 3'-CCTGCAAGCCCTACCGTT-5'] as described elsewhere (Probst et al., 2013b). In addition, fluorescence in situ hybdridization with the above mentioned specific oligo-nulceotide was performed as described (Rudolph et al., 2001).

Assembly, binning and completeness estimation of genomes. The filtered illumina paired-end reads were digitally normalized to a maximal k-mer coverage of 130 and a minimal k-mer coverage of 3, resulting in a total of 54,853,038 paired reads (Brown et al., 2012). A 10% subset of the normalized reads together with the entire set of 454 paired-end reads was used for a hybrid assembly with MIRA (Chevreux, 2005). Resulting contigs were filtered by a minimal average coverage cutoff of 10 and a minimal length cutoff of 3 kpbs. These quality cutoffs were shown to be very conservative in order to avoid chimeric sequences in datasets (Mende et al., 2012). Filtered contigs were searched against a custom set of marker genes (Supplementary Figure 5) from the eggNOG database (Powell et al., 2012) and blast results were phylogenetically analyzed using MEGAN4 [(Altschul et al., 1990; Huson et al., 2011); best hit, bit score cutoff 150]. Contigs classified as archaeal were extracted and grouped into separate archaeal. Bins were further constrained based on coverage/GC plots (R-Development-Core-Team, 2005) and phylogenetic placement of the marker genes, i.e. 16S rRNA and *hamus* genes, which were identified using blastn (Altschul et al., 1990). These bins were used as custom training sets for Phymm and the PhymmBL algorithm was used for taxonomic classification of the complete assembly (Brady and Salzberg, 2009). This re-classification of the metagenome expanded the bins by recruiting further contigs based on nucleotide composition. Coding sequences were predicted based on a house-internal workflow that integrates *ab initio* predictions from Glimmer (Delcher et al., 1999), GeneMark (Borodovsky and Mclninch, 1993), Prodigal (Hyatt et al., 2010), and Critica (Badger and Olsen, 1999) with homology information derived from a BLASTp (Altschul et al., 1990) search against the NCBI non-redundant database. Noncoding RNAs were identified using tRNAscanSE (Lowe and Eddy, 1997) and RNAmmer (Lagesen et al., 2007) and by searching against the Rfam database (Griffiths-Jones et al., 2005). Functional annotation of the CDSs was based on InterProScan (Mulder and Apweiler, 2007; Quevillon et al., 2005) and the Swiss-Prot and trEMBL databases (Boeckmann et al., 2003). To estimate completeness of genomic bins predicted protein sequences were searched against a custom database of COGs that occur in at least 99% of the bacteria and archaea in the eggNOG database [(Powell et al., 2012); Supplementary Figure 5] with an e-value cutoff of E-10.
For a number of known single-copy genes multiple copies were found in the SM1-MSI bin. To rule out sequencing errors as the reason for these multiple copies, the K_a/K_s ratio of the number of non-synonymous substitutions per non-synonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s) was calculated (Wang et al., 2010) for these genes and compared to a null model. Therefore, nucleotide sequences of single-copy genes that were present twice were extracted and aligned (Zhang et al., 2012). For each pair of genes with at least five variable positions the K_a/K_s ratio was calculated (Wang et al., 2010). To generate a null-model, a similar amount of mutations as observed in the corresponding pair was introduced randomly into one of the two genes using an in-house script. Then the randomly non-mutated gene and the mutated version were aligned (Zhang et al., 2012) and the K_a/K_s ratio was computed (Wang et al., 2010). A paired t-test of log-transformed K_a/K_s ratios paired by housekeeping gene was used to test for significant differences between observed variations of single-copy genes and randomly introduced mutations.

Annotation of archaeal genomes for metabolic pathway prediction. Bins of archaeal contigs were re-annotated using the synteny-supported annotation platform MaGe (Vallenet et al., 2013; Vallenet et al., 2006). Specific tools that supported the annotation and manual curation process were blastn, blastp (Altschul et al., 1990), Microcyc (Vallenet et al., 2013), KEGG (Kanehisa and Goto, 2000), SWISS-PROT (Boeckmann et al., 2003), TrEMBL (Boeckmann et al., 2003), TCDB (Saier et al., 2006), COGnitor (Tatusov et al., 2000), FigFam (Meyer et al., 2009), InterProScan (Mulder and Apweiler, 2007; Quevillon et al., 2005), Pfam (Punta et al., 2012), PsortB (Nancy et al., 2010), HMMsearch and HMMblast (Eddy, 2009).

Phylogenomic analysis. Previous datasets of archaeal ribosomal proteins were used to create HMM profiles (Finn et al., 2011). HMM searches (Finn et al., 2011) were then performed on a local database of 137 complete archaeal genomes and SM1, IM-C4, CGR-A CGR-B and CGR-C genomes. Each protein dataset was aligned using Muscle (Edgar, 2004) with default parameters, and unambiguously aligned positions were automatically selected by using the BMGE software for

multiple-alignment trimming (Criscuolo and Gribaldo, 2010) with a BLOSUM70 substitution matrix. Trimmed alignments were then concatenated by allowing a maximum of 15 missing proteins per data set resulting in a final dataset of 56 ribosomal proteins with 6546 amino acid positions for phylogenetic analysis. PhyloBayes 3.3b (Lartillot et al., 2009) was used to perform Bayesian analysis using the CAT model and a gamma distribution with four categories of evolutionary rates. The concatenated datasets were also recoded using Dayhoff6 and Dayhoff4 recoding schemes as implemented in PhyloBayes 3.3b (Lartillot et al., 2009) and analyzed with the same model parameters. For each dataset, two independent chains were run in parallel until convergence. The first 25% of trees were discarded as 'burn in' and the posterior consensus was computed by selecting one tree out of every two to compute the 50% majority consensus tree.
Considering universally conserved proteins (10 different: Tef1p, Rpa190p, Rpa135p, Sec61p, Eft1p, Kae1p, Fun12p, Srp54p, Rli1p, Vma2p), analyses were identical to those for ribosomal protein analyses as mentioned above but using a BLOSUM62 substitution matrix resulting in 4537 amino acid positions.
Previous datasets of archaeal DNA replication proteins (Raymann et al., 2014) were used to search for homologs in the SM1 and IM-C4 genomes using HMM profiles and searches (Finn et al., 2011). Phylogenetic analyses were performed for each protein to assess orthology and define the presence/absence pattern of DNA replication markers.
Further evidence for the phylogenetic placement of the SM1 cluster was gathered by identifying best-hit homologs in the 137 archaeal genomes [see above; blastp (Altschul et al., 1990), e-value cutoff E-4].

Estimation of horizontal gene transfer. Protein sequences predicted for the SM1-MSI bin were searched against the NCBI's NR database (Altschul et al., 1990) using Phylogenie (Frickey and Lupas, 2004) in order to identify potential homologs. Maximum likelihood trees of each potential homologue set of protein sequences were calculated (Stamatakis et al., 2005) and analyzed using Phylogenie [phat; (Frickey and Lupas, 2004)]. In addition, the taxonomic distribution of hits in NR was analyzed using MEGAN4 (Huson et al., 2011) with a bitscore cutoff of 50.
More in depth analysis was carried out with proteins involved in C_1 cycling of the SM1 genome. These were used as seeds to perform exhaustive homology searches using blastp (Altschul et al., 1990) and HMMER (Finn et al., 2011) on a local database of 211 genomes representative of the main bacterial phyla and the 137 currently available complete archaeal genomes (one per species). For each protein, the top ten hits — retrieved from the non-redundant sequence database on NCBI using the blastp (Altschul et al., 1990) — were added. Final single protein datasets were trimmed using the software BMGE (Criscuolo and Gribaldo, 2010) with default parameters and subjected to phylogenetic analyses by Maximum Likelihood and Bayesian methods. Single-matrix substitution models were chosen using the ProteinModelSelection script available from the RAxML website (http://www.exelixis-lab.org/). Maximum likelihood analyses were performed with RAxML (Stamatakis et al., 2005). Bayesian analyses were run with MrBayes 3.2 (Ronquist et al., 2012), using the mixed amino acid substitution model and four categories of evolutionary rates. Two independent runs were performed for each data set, and runs were stopped when they reached a standard deviation of split frequency below 0.01 or the log likelihood values reached stationary. The majority-rule consensus trees were obtained after discarding first 25% samples as 'burn-in'. Each protein was assigned as potentially arising from horizontal gene transfer from a bacterial source if it robustly clustered within a bacterial clade, as opposed to those that clustered within archaea, and those whose origin remained uncertain.

Transcriptomics. For genes for certain key enzymes or for single subunits thereof, transcription was tested via specific mRNA detection in biofilm samples (Supplementary Table 4 containing list of genes and primers). Total RNA was isolated using the PowerBiofilm™ RNA Isolation Kit (Mobio Laboratories Inc., Carlsbad, USA) according to manufacturers' instructions (DNA digestion was performed for 30 min). After precipitation of nucleic acids (Moissl-Eichinger, 2011) the DNAse treatment was repeated, followed by subsequent reverse-transcription to cDNA (QuantiTect Rev. Transcription Kit, Qiagen, Hilden, Germany). Specific primers were designed using the web tool Primer3v.0.4.0 software (http://biotools.umassmed.edu/bioapps/primer3_www.cgi; parameters: product size: optimum 400bp, GC% 40-60%, annealing temperature: 60°C optimum). Specificity of primers was tested using blast (Altschul et al., 1990) against NCBI NR and the metagenome. CDNA was used for amplification with designed primer pairs individually (denaturation time: 5 min 95°C; 30 cycles: 45s 94°C, 45s 60°C, 90s 72°C; final elongation: 10 min 72°C). Positive PCR products were purified (HiYield® Gel/

PCR DNA Fragments Extraction Kit; Sued-Laborbedarf GmbH, Gauting, Germany) and Sanger sequenced (LGC Genomics GmbH, Berlin, Germany). Experiments were carried out in duplicates.

Fluorescence immuno-labeling. For the production of *hami*-specific antibodies, *hami* filaments were released from the cells as follows: Biofilm samples were incubated in KPH buffer [NaCl 0.7 mM, $MgCl_2$ 0.1 mM, $CaSO_4$ 1.6 mM, HEPES 1.0 mM, supplemented with 0.1% SDS (v/v)] for 25 minutes and periodically vortexed. At that time, SM1 euryarchaeal cells were completely dissolved. The resulting suspension was centrifuged (30 minutes, 5.500 g, 20°C) to remove larger precipitates and the supernatant containing the *hami* was ultracentrifuged (1 h, 92387.1 g, Beckman OPTIMA LE 80 K, 70.1 Ti-Rotor, 4°C). The pellet was re-suspended in 2 ml of KPH buffer and applied on a sucrose-gradient [10-70% sucrose (w/v) in sterile KPH] and centrifuged for 17 hours (309k x g, Beckman OPTIMA LE 80 K, SW 60 Rotor, 4°C). The band appearing in the lower third of the tube was removed using a sterile syringe. After confirming the presence of *hami* via transmission electron microscopy (see below) the sample was sent to Davids biotechnology (Regensburg, Germany) for antibody production. A chicken was pre-immunized with the *hami*-fraction (0.22 mg/ml) three times over the course of 21 days. 28 days after the first immunization, eggs were collected and the IgG fraction ("anti-*hami*"; 15.1 mg/ml in 0.02% sodium-azid) was harvested.
For immuno-staining, collected biofilms were fixed with paraformaldehyd [5% (v/v)] at room temperature (1 h) and washed three times with 1x PBS (phosphate buffered saline). Afterwards, fixed cells were incubated for 15 minutes in PBST [PBS including Tween20 0.05 (v/v) and 0.1% SDS (v/v)] at 30°C, followed by a centrifugation step (15 min, 14.500 g, 20°C). The first primary antibody, anti-Anabaena FtsZ [(Kuhn et al., 2000); AS07217, Agrisera; dilution 1:200] was added and incubated for 2 hours at 30°C. Cells were centrifuged again, followed by incubation in PBST [+0.1% SDS (v/v)] for 15 minutes and another centrifugation step. After incubation for one hour with the conjugated goat anti-rabbit IgG (dilution 1:200), the cells were washed twice with PBST [+0.1% SDS (v/v)], spread within a well of a gelatine-coated slide (P. Marienfeld KG, Lauda-Koenigshofen, Germany) and fixed by air-drying. After incubation with 16 µl of PBST [+0.1% SDS (v/v)] at 37°C, the PBST buffer was replaced with 16 µl PBST buffer containing the anti-*hami* IgG (dilution 1:2000) and the cells were labeled for 1 hour at 37°C. Subsequently, the slide was washed 15 min in PBST [+0.1% SDS (v/v)], rinsed with H_2O and air dried. The second antibody (goat anti-chicken, Cy3-labeled; 0.64 mg/ml, dilution 1:500) was added and incubated for 1 hour at 37°C. After washing two times with PBST [+0.1% SDS (v/v)], the slide was rinsed with H_2O, DAPI stained and analyzed by fluorescence microscopy [Olympus (BX53F, Hamburg, Germany) with epifluorescence equipment and imaging software cellSens].

Transmission and scanning electron microscopy. For TEM, fresh, unfixed biofilm pieces were deposited on a carbon-coated copper grid and negatively stained with 2% (w/v) uranyl acetate, pH 4.5 or 2.0% (w/v) phosphotungstic acid (PTA), pH 7.0. In a second approach, biofilms were treated with 1% SDS (w/v) for 30 min, causing destruction of the cell wall and a potential release of cell appendages. These samples were examined using a CM12 transmission electron microscope (FEI, Eindhoven, The Netherlands) operated at 120 keV. All images were digitally recorded using a slow-scan charge-coupled device camera that was connected to a computer with TVIPS software (TVIPS GmbH, Gauting, Germany).
For conventional fixation, freshly taken biofilms were fixed in original spring water including 0.1% glutardialdehyde (w/v). Samples were rinsed several times in fixative buffer and postfixed at room temperature for 1 h with 1% (w/v) osmium tetroxide. After two washing steps in water, the cells were stained for 30 min with 1% (w/v) uranyl acetate in 20% (v/v) acetone. Dehydration was performed by a graded acetone series. Samples were then infiltrated and embedded in Spurr's low-viscosity resin. For high-pressure freezing experiment samples were frozen either with a Leica HPM100 (Leica Microsystems GmbH, Wetzlar, Germany) or a Wohlwend HPF Compact 02 (Engineering Office M-Wohlwend GmbH, Sennwald, Switzerland). In the first case aluminum platelets were used which were filled with one piece of biofilm. Freeze substitution was performed in acetone with 2% (w/v) osmium tetroxide and 0.2% uranyl acetate (w/v), including 5% (v/v) water. After embedding the samples in Spurr's low-viscosity resin, ultrathin sections were cut with a diamond knife and mounted onto uncoated copper grids. The sections were poststained with aqueous lead citrate (100 mM, pH 13.0). Transmission electron micrographs of samples prepared this way were taken with an EM 912 electron microscope (Zeiss) equipped with an integrated OMEGA energy filter operated at 80 kV in the zero loss mode.

For SEM, drops of the sample were placed onto a glass slide, covered with a coverslip, and rapidly frozen with liquid nitrogen. The coverslip was removed with a razor blade and the glass slide was immediately fixed with 2.5% (w/v) glutaraldehyde in 10 mM cacodylate buffer (pH 7.0), postfixed with 1% osmium tetroxide in fixative buffer, dehydrated in a graded series of acetone solutions, and critical-point dried after transfer to liquid CO_2. Specimens were mounted on stubs, coated with 3 nm platinum using a magnetron sputter coater, and examined with a Zeiss Auriga scanning electron microscope operated at 1-2 kV. When using the Wohlwend HPF Compact 02, biofilm aggregates were placed in the centre of 3 mm gold carriers and high pressure frozen with a and subsequently freeze substituted in an automatic EM AFS 2 unit (Leica Microsystems GmbH, Wetzlar, Germany). The substitution medium consisted of acetone in combination with 0.2% osmium tetroxide, 0.25% uranyl acetate and 5% water. The substitution program including washing steps and the following epon embedding, thin sectioning and post staining was carried out as described previously (Gonzalez et al., 2011; Klingl, 2011; Peschke et al., 2013). For these samples, microscopy was performed on a JEOL JEM-2100 (JEOL, Tokyo, Japan) also operated at 120 kV and equipped with a 2k x 2k fast scan CCD camera F214 combined with EM Menu4 (TVIPS GmbH, Gauting, Gemany).

Focused-ion-beam scanning electron microscopy tomography (FIB-SEM tomography). The focused ion beam FIB block face serial sectioning was performed by a Zeiss-Auriga workstation. The focused ion beam consisted of Ga^+ ions accelerated by a voltage of 30 kV. In the cut-and-view mode, sections ranging in thickness between 5 nm and 10 nm (dependent on the magnification) were produced with the FIB and field emission scanning electron microscopy (FESEM) images, which were recorded at 1.5 kV using the in-lens energy selective backscattered (EsB) detector set to -1200V. Specimens were tilted to an angle of 54°; images were tilt corrected for undistorted surface view.

Lipid Extraction and HPLC-MS analysis. Total lipid extracts (TLEs) were obtained from samples of the biofilm using a modified Bligh and Dyer protocol (Sturt et al., 2004), after adding an internal standard (phosphatidylcholine C21:0/21:0) and 3 g of combusted sea sand. Approximately 10^8 to 10^9 cells were subjected to extraction. The samples were treated with dichloromethane/methanol/buffer (DCM/MeOH/buffer; 1:2:0.8; v/v/v) and ultrasonicated. This procedure was conducted in four steps, using two times phosphate buffer (pH 7.4), followed by two times trichloroacetic acid buffer (pH 2). In each step after sonication, the mixture was centrifuged and the supernatant collected in a separatory funnel. DCM and deionized Milli-Q water (1:1; v/v) were added to the combined supernatants, to a ratio of DCM/MeOH/buffer (2:1:0.8; v/v/v). Thereafter, the organic phase was separated and the remaining aqueous phase washed three times with DCM. The collected organic phase (TLE) was subsequently washed three times with Milli-Q water and finally reduced to dryness under a stream of nitrogen at a constant temperature of 37°C. The obtained TLEs were stored at -20°C.
Separation of IPLs was achieved on a Dionex Ultimate 3000 UHPLC equipped with a Waters Acquity UPLC Amide column (150 x 2.1 mm, 1.8 µm particle size). Chromatographic conditions, according to a previously published method (Wörmer et al., 2013), were as follows: constant flow rate of 0.4 ml/min with eluent A [75% acetonitrile; 25% DCM; 0.01% formic acid; 0.01% ammonium hydroxide solution (NH_3)] and eluent B (50% MeOH; 50% Milli-Q water; 0.4% formic acid; 0.4% NH_3). Under a constant flow, the HPLC routine applied: 99% A and 1% B for 2.5 min, increasing to 5% B at 4 min, followed by a linear gradient to 25% B at 22.5 min and then to 40% B at 26.5 min. Thereafter a 1 min washing step with 40% B followed and afterwards reset to the initial conditions for 8 min to achieve column re-equilibration. Compound detection was conducted on a Bruker maXis Ultra-High Resolution qToF-MS, equipped with an electrospray interface (ESI). IPLs were measured in positive ionization mode, while scanning a mass-to-charge (m/z) range of 150-2,000, with automated data-dependent MS/MS fragmentation of base peak ions. Compound identification was achieved by monitoring exact masses of possible parent ions (present mainly as H^+ and NH_4^+ adducts) in combination with characteristic fragmentation patterns (Sturt et al., 2004; Yoshinaga et al., 2011). The reported relative distribution of microbial lipids is based on the peak areas of the respective molecular ions without differentiating for potential differences in response factors; the data should therefore be viewed as semi-quantitative.

Conductivity analysis of biofilms. Conductivity was measured at a probe station (Cascade Microtech PM8, Beaverton, USA). A 100-µl drop of springwater including a biofilm flock was put on a microscope slide. Two probes with a distance of l = 1 mm were dipped into the drop, so they touched the biofilm. A Yokogawa 7651 voltage source was used to apply a voltage

to the probes. The voltage was sweeped between -0.3 V and +0.3 V. The resulting current was amplified with an Ithaco 1211 current amplifier, its output was measured with a Agilent 34401A voltagemeter. The conductivity of the solution was calculated by

$$X = (I_+ - I_-)/(U_+ - U_-) * l/A$$

using the current values I_+ and I_- for maximum and minimum voltage U_+ and U_-. A is the surface area of the probes immerged in the solution and is estimated to be 1.5 mm².

Description of "Candidatus Altiarchaeum hamiconnexum". We here propose the name "*Candidatus* Altiarchaeum hamiconnexum sp. nov." for the SM1 Euryarchaeon found in the Mühlbacher Schwefelquelle, Sippenauer Moor and Crystal Geyser.

[Altiarchaeum gen. nov. (Al.ti.archae'um. L. adj. altus, high, deep; N.L. neutr. n. archaeum, a taxonomic unit; N.L. neutr. n. Altiarchaeum, an Archaeum from the deep.]

[A. hamiconexum spec. nov. (ha.mi.con.ex'um. L. n. hamum, a hook; L. part. adj. conexus, connected; N.L. neutr. adj. hamiconexum connected by a hook).]

Biofilm-forming small coccoid cells (0.4-0.7 µm in diameter) with double membrane. Non-motile. Formation of highly structured cell surface appendages (*hami*) and extracellular polymeric substances (EPS). Lipids are mainly mono- or diglycosyl diehter with two phytanyl chains (C_{20}-C_{20}) or a combination of one phytanyl chain and one sesterterpanyl chain (C_{20}-C_{25}). The biofilms originate in the deep subsurface of cold, anoxic freshwater springs. In the surface water of sulfidic springs, "*Candidatus* Altiarchaeum hamiconnexum" forms specific microbial communities together with sulfide-oxidizing bacteria [string-of-pearls community; (Henneberger et al., 2006; Moissl et al., 2002; Rudolph et al., 2004; Rudolph et al., 2001)]. Cell division is based on FtsZ. Non-methanogenic, may grow on acetate, formate or CO.

We also propose the candidatus order "*Altiarchaeales*" (order nov.), and the candidatus family "*Altiarchaeaceae*" (fam. nov.).

Acknowledgements

Research on SM1 from MSI and IM-C4 was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft), grant no. MO 1977 3-1 given to CM-E. AJP was supported by the German National Academic Foundation (Studienstiftung des deutschen Volkes). Research on SM1 CG was supported by US DOE EFRC grant given to Jillian Banfield. Research of IAB was supported by the German Research Foundation (grant no. BE 4822/3-1 and Heisenberg fellowship). Kasie Raymann is a scholar from the Pasteur–Paris University (PPU) International PhD program and received a stipend from the Paul W. Zuccaire Foundation. We thank Charlotte Völkel for experimental assistance and Prof. Dr. Reinhard Wirth, Prof. Dr. Robert Huber and Dr. Tamas Torok for critical discussion. Pierre Offre and Dmitrij Turaev are acknowledged for support in annotation software handling.

Supplementary Information

Supplementary information can be found on the supporting DVD.

File	Content
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_genome.fasta	Genomic bin of SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_CDS.fasta	CDS predicted for SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_protein.fasta	Protein sequences for SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_ncRNA.fasta	Non-coding RNAs predicted for SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_annotation_table.txt	Annotations of predicted proteins for SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-MSI_genbank_genome.txt	Genbank file for SM1-MSI
./Publications/05_Probst_in-prep/Additional_data/SM1-CG_protein.fasta	Protein sequences for SM1-CG
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_genome.fasta	Genomic bin of IM-C4
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_CDS.fasta	CDS predicted for IM-C4
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_protein.fasta	Protein sequences for IM-C4
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_ncRNA.fasta	Non-coding RNAs predicted for IM-C4
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_annotation_table.txt	Annotations of predicted proteins for IM-C4
./Publications/05_Probst_in-prep/Additional_data/IMC4-MSI_genbank_genome.txt	Genbank file for IM-C4

IV. General discussion

In this thesis the archaeome and bacteriome of two independent biotopes were investigated using next-generation-based, cultivation-independent technologies. The first biotope comprised spacecraft assembly cleanrooms that were shown to harbor a very small fraction of viable bacteriome (Vaishampayan et al., 2013). A study on the archaeome in cleanrooms empowered us to conclude on the origin of archaea in these facilities: The human skin was identified to carry an intrinsic proportion of archaea that are potentially transported into cleanrooms via human activity (Probst et al., 2013a).

The second biotope comprised sulfidic subsurface aquifers near Regensburg, where almost mono-species archaeal biofilms occur [SM1 Euryarchaeon; (Henneberger et al., 2006)]. These biofilms were shown to carry mainly sulfate-reducing bacteria in their minor bacterial fraction (Probst et al., submitted; Probst et al., 2013b) and strain-specific variations in their archaeome across geographic location (Probst et al., submitted). A metagenomics survey enabled inferring the phylogenomic position of the dominant archaea in the biofilm and predicting a metabolic model for the SM1 Euryarchaeon, which potentially grows on acetate and/or oxidizable C_1 compounds in the subsurface biotope (Probst et al., in preparation).

1. Planetary protection research shaping the human skin archaeome

1.1. The cleanroom microbiome research: How to respect the living?

In 2001, Venkateswaran published the first 16S rRNA gene amplicon-based analysis of a cleanroom microbiome, which opened a new era of planetary protection research (Venkateswaran et al., 2001). Many follow-up studies elucidated the microbiome of cleanroom and spacecraft hardware by applying molecular approaches (Cooper et al., 2011; La Duc et al., 2004; La Duc et al., 2003; La Duc et al., 2009; La Duc et al., 2012; Moissl et al., 2008; Moissl et al., 2007; Moissl-Eichinger, 2011; Schwendner et al., 2013; Stieglmeier et al., 2012; Vaishampayan et al., 2010). However, these studies never accounted for the viable proportion in their data, although strict and rigorous cleaning efforts, which may kill a great proportion of microorganisms, are applied in these nutrient-depleted biotopes. Nevertheless, by using standard and alternative cultivation approaches, few studies have shed light on the proportion of microorganisms that can survive harsh cleanroom conditions (Probst et al., 2010b; Puleo et al., 1977; Schwendner et al., 2013; Stieglmeier et al., 2009). The isolates retrieved in the aforementioned studies are a valuable resource for resistance testing in order to estimate the potential planetary protection risk of cleanroom microbiomes (Moissl-Eichinger et al., 2012b). Recently, we gauged the percentage of culturable genera in cleanroom microbiomes to be ca. 35%

(Moissl-Eichinger et al., 2013), which is extraordinary high compared to the regular 1% estimate for the entire biosphere (Amann et al., 1995; Colwell, 1997). However, the calculation was performed at genus level, and the uncultivated diversity was estimated solely based on 16S rRNA gene cloning efforts. Moreover, potential recovery efficiency errors were not taken into account either (Probst et al., 2010a; Probst et al., 2011). Considering the rare microbiome that can be generated with more sensitive microbiome profiling methods may decrease the cultivated percentage drastically (DeSantis et al., 2007; Hazen et al., 2010; Sogin et al., 2006). As a result, it remained inevitable to explore the remaining >65% of not-yet cultured microorganisms with regard to A) viability, B) phylogeny, and C) metabolic capabilities for estimating planetary protection risk.

In the course of a recent study, we coupled a molecular viability assay to next generation technology platforms to perform in depth 16S rRNA gene analyses via 454 pyrotagsequencing and PhyloChip G3 (Vaishampayan et al., 2013). This study was the first ever elucidating the viable molecular diversity of cleanrooms and the first ever to link viability testing to PhyloChip technology. The viability assay was based on propidium monoazide (PMA), a chemical that is added prior to DNA extraction, passes compromised cell walls, and binds covalently to DNA after photoactivation. The PMA-DNA complex is then inaccessible for a DNA-polymerase for downstream polymerase chain reaction (PCR). It should be noted, that PMA does not enter microbes with intact cell membranes, which are consequently considered viable (Nocker et al., 2007). This PMA procedure has led to diverse publications but has also been subject to limitations concerning cells with thick, murein-rich cell walls like Actinobacteria or endospores, which may be dead but escape PMA labeling (Andorra et al., 2010; Bae and Wuertz, 2009; Cawthorn and Witthuhn, 2008; Fittipaldi et al., 2011; Lin et al., 2011; Nocker et al., 2009; Nocker et al., 2007; Parshionikar et al., 2010; Probst et al., 2013a; Vaishampayan et al., 2013; van Frankenhuyzen et al., 2011; Vesper et al., 2008; Wagner et al., 2008; Yanez et al., 2011). Actinobacteria were identified to be non-viable in cleanroom samples indicating that the PMA procedure applied herein was not biased toward murein-rich, vegetative cells, whose disrupted cell walls apparently allowed PMA to pass (Vaishampayan et al., 2013). However, signatures of some bacilli were retrieved in the viable fraction of the microbiome, indicative of their viability or of their dormant status as endospores. Although proper de-coating procedures can be used to even discriminate between viable and non-viable endospores in quantitative PCR and microscopy (Probst et al., 2012; Rawsthorne et al., 2009), the procedures also disrupt viable vegetative cells. Ergo, a detection of viable endospores and vegetative cells in the same sample preparation via PMA-treatment is currently unfeasible, and potential endospores interfere with the observed viable microbiome. Nevertheless, the study presented herein is encouraging to such a degree as spacecraft cleaning procedures appeared to work well and to decrease the microbiome load 10-fold, when comparing the proportion of viable and total microbiome (Vaishampayan et al., 2013). At the same time, the study enabled researches to thoroughly assess the viable microbial cleanroom community more than cultivation assays had done before.

As indicated earlier, scientist should not only study the phylogeny of the viable cleanroom microbiome but also its potential metabolic capabilities. Diverse technologies of the ‘omics’ era are available for studying microbiomes. Unfortunately, not every technology can be applied to cleanroom

samples. The biomass that can be harvested from cleanrooms is very low (Moissl-Eichinger et al., 2013; Vaishampayan et al., 2013; Venkateswaran et al., 2001) making metatranscriptomics and metaproteomics inapplicable, although these technologies would give direct evidence for viability of the corresponding microorganism. Hence, metagenomics is the only existing alternative to study microbiome function in this biotope. But how should future studies be designed to elucidate the metagenome of viable microbes in cleanrooms? In general, metagenomics is the application of sequencing technology on extracted DNA or DNA ligated into vectors (Tyson et al., 2004; Venter et al., 2004). Consequently, the application of PMA to samples prior to DNA extraction would not be possible unless coupled to multiple displacement amplification (MDA; whole genome amplification). MDA commonly used for low biomass samples is, however, introducing great bias to artificially designed but also natural microbial communities (Abulencia et al., 2006; Direito et al., 2014; Yilmaz et al., 2010). In particular, dominant organisms appeared to be altered drastically in their relative abundances, as revealed by pyrotagsequencing analysis of 16S rRNA genes (Yilmaz et al., 2010). In conclusion, metagenomics coupled to PMA/MDA treatment may be biased with regard to the reconstructed microbial community composition but could in turn still reveal unseen metabolic capabilities of not-yet cultivated species. Recently, a project set out with the aim of unraveling the metagenome of a cleanroom has been launched by a group around Dr. Vaishampayan. The resulting information on the cleanroom microbiome function is eagerly desired by the planetary protection community (pers. com. Parag Vaishampayan, Jet Propulsion Laboratory, CA, US, funded by NASA-ROSES 2011). The data could have profound implications on planetary protection considerations of future life detection missions.

1.2. Discovery and definition of the human skin archaeome

The bacteriome in cleanrooms resembles the human microbiome as most of the signatures and isolates were closely related to human-associated microorganisms (Moissl et al., 2007; Stieglmeier et al., 2009). Putative influence by the geographic location of cleanrooms may just arise from different scientist working in these facilities as the human microbiome itself is versatile (Caporaso et al., 2011; La Duc et al., 2007; Moissl-Eichinger et al., 2013; Yatsunenko et al., 2012). Consequently, humans are considered the major source of microbial contaminants in cleanroom environments due to the loss of skin particles that bear skin microbiota (Moissl-Eichinger et al., 2013). Signatures of e.g. *Staphylococcus* — a common cleanroom contaminant — are very likely to originate from skin rather than from the surrounding environment (Stieglmeier et al., 2009), yet a direct comparison of the cleanroom microbiome and human skin microbiome is missing in the scientific literature. There are indeed extraordinary microorganisms in cleanrooms, whose origin remains speculative as they had previously not been reported for human skin. For instance, the occurrence of Thaumarchaeota was argued to be due to human activity (Moissl-Eichinger, 2011), although the human skin was believed to be free of Archaea, as stated by Grice and Segre in April 2011 (Grice and Segre, 2011). Two months after the review by Grice and Segre appeared in the literature, Thaumarchaeota were found on the palms of two human subjects but were categorized as transient members of the skin microbiome (Caporaso et al., 2011). One year later, Hulcr et al. found signatures of archaea in human navels (Hulcr et al., 2012). However, these archaea were not classified Thaumarchaeota but methanogens and

haloarchaea and were most likely oral or fecal contaminants (Probst et al., 2013a). Published in 2013, we determined that 16S rRNA gene signatures of cleanroom archaea originate from human skin and that thaumarchaeal operational taxonomic units (OTU; grouped at 1% dissimilarity) are shared between skin microbiome, hospital microbiome and cleanroom microbiome. Ergo, the origin of archaea in cleanrooms was consequently attributed to human activity and to skin particles that are shed in such facilities (Probst et al., 2013a). However, the natural biotope, where the archaea flourish, may not be the cleanroom environment as PMA-treated samples have so far not yielded any archaeal signatures, neither from NASA nor from ESA facilities (data not shown; pers. com. Parag, Vaishampayan, Jet Propulsion Laboratory, CA, US). Also the human skin as a natural biotope for Thaumarchaeota may be questioned. A greater survey with more than 13 human subjects would be necessary in order to solidify the statement that archaea are a continuous part of the human skin microbiota. In addition, metadata should be collected for each human subject in order to link the abundance (or presence/absence) of archaeal signatures with age, gender, disease status or other factors. Although thaumarchaeal signatures/cells possibly originate from the surrounding environment and stick to human skin, their signatures on 13 human subjects confirms their contribution to the human microbiome.

With the detection of Thaumarchaeota on human skin and with the rejection of the hypothesis that archaea are only transient members of the skin microbiome, we have identified another phylum of microorganisms that are a part of the human microbiome. As a result, the human archaeome consists now of three different phyla, the Euryarchaeota, the Crenarchaeota, and the Thaumarchaeota. Euryarchaeota, mainly methanogens, have been detected in the human gut and oral cavity (Brusa et al., 1993; Lepp et al., 2004), as have Crenarchaeaota in human gut (Rieu-Lesme et al., 2005). Thaumarchaeota are a part of the human skin microbiome (Probst et al., 2013a) and were recently also found in human stool samples (Hoffmann et al., 2013).

The origin and classification of archaea in cleanroom environments was determined but their function remains enigmatic. So far, all cultivated representatives of the thaumarchaeal phylum are ammonia oxidizers (Hatzenpichler et al., 2008; Konneke et al., 2005; Tourna et al., 2011) but they also appear to be mixotrophic as their growth can be supported by pyruvate addition (Tourna et al., 2011). Although the human skin is constantly emanating ammonia (Nose et al., 2005), the quantity of archaeal *amoA* genes compared to 16S rRNA genes was low on human skin (Probst et al., 2013a). This is in congruence with the fact that the capacity of ammonia-oxidation may not hold true for all members of this phylum. Isolation and characterization of more representatives of the phylum Thaumarchaeota — and in particular of those from human skin — are the only chance to test for the interaction and relationship of Thaumarchaeota with humans.

Although the function of the archaea in cleanroom biotopes and on human skin could not be deciphered entirely, the results of this thesis represent a milestone in planetary protection research. The viable bacteriome in cleanrooms was shown to be of minimal diversity and abundance (Vaishampayan et al., 2013). This is highly important for present and future life detection missions as the current cleaning protocols appear to function well. At the same time, the origin of the archaeome in cleanrooms was determined to be the human (skin) microbiome, whose composition was in turn

extended by another phylum, the Thaumarchaeota. These findings may have great impact on the design of medical skin (microbiome) surveys, which should from now on contain the detection of archaea in regular diagnostics (Probst et al., 2013a).

2. Microbiome profiling beyond nucleic acids?

2.1. 16S rRNA gene analysis: The tip of the iceberg or 'a proxy of the proxy'

With its 9 (hyper-)variable regions, the universal 16S rRNA gene has emerged as golden standard for inferring microbiome profiles. The analysis of 16S rRNA gene pools from the environment has proven an appositely tool to unveil entire microbial communities and their changes over time and space. Historically, the approximate length of 1.5 kb was found very suitable for Sanger sequencing analysis, only necessitating two sequencing runs per cloned 16S rRNA gene. With the rise of next generation sequencing platforms, more 16S rRNA gene molecules could be assayed allowing the identification of the rare biosphere in environmental samples (Sogin et al., 2006; Zhou et al., 2011). Based on a recent study on screening the Western English Channel (Caporaso et al., 2012), the term 'Everything is everywhere, but the environment selects' (Baas-Becking, 1934) needed to be changed to 'Everything is everywhere, but the environment selects its abundance' (Probst et al., 2014), which was without a doubt a milestone in (16S-rRNA based) microbial ecology research.

However, 16S rRNA gene profiling is suffering from a variety of methodological bottlenecks that can lead to illegitimate conclusions concerning A) phylogenetic relationships, B) single taxa abundance, and C) entire community structure. Prior to sequencing 16S rRNA genes, enough starting material, i.e. 16S rRNA gene copies, is required. Usually, this is circumvented by performing a PCR with so-called 'universal' primers. First of all, there are no 'universal' primers that can cover all Bacteria and/or Archaea. Ergo, there is always an uncovered portion of microbial members that are not considered in any downstream analysis [the reader is referred to two recent studies on primer specificity (Gantner et al., 2011; Klindworth et al., 2013)]. For instance, the SM1 Euryarchaeon studied in this thesis would have never been identified to play a role in the biotope it is dominating if someone had used standard 16S rRNA gene primers that cover the near-complete 16S rRNA gene (Rudolph et al., 2001). Second, the PCR amplification step introduces a bias towards selecting dominant genes and those that appear to have better scores with the applied primers. The latter bias has also been observed in the current thesis when analyzing the intergenic spacer region (between 16S rRNA gene and 23S rRNA gene) of archaea in the SM1-Euryarchaeon dominated biofilm. More than 50% of the sequences in the clone library were identified as the IM-C4 Euryarchaeon, although its abundance was estimated to be <1% using quantitative PCR with specific primers (Probst et al., in preparation). These intra-sample comparisons of reads are error-prone, and taxa abundances are semi-quantitative (a 'proxy of the proxy') when derived from amplicon sequencing analysis. Other technologies like lipid extraction-based profiling can fairly complement microbiome research and give more confidence in observed community structures as for instance performed in Hazen et al. (Hazen et al., 2010). Third, the short reads generated in the above-mentioned next-generation sequencing of 16S rRNA genes are limiting the resolution of the technique. While the majority of sequences do get classified at family

level, and at least 50% at genus level, the species level can hardly be achieved irrespective of the (hyper-)variable region used (Wang et al., 2007). Last, the phylogenetic relationships calculated from 16S rRNA genes can be error-prone as Thaumarchaeota, for instance, were for a long time believed to be Crenarchaeota due to the low capacity of single-gene derived evolutionary relationships (Brochier-Armanet et al., 2008; Pester et al., 2011; Spang et al., 2010).

2.2. A new technique for microbiome profiling using Synchrotron radiation-based Fourier transform infrared spectromicroscopy

In the course of this thesis, SR-FTIR spectromicroscopy was used for the first time to decipher microbiomes and their function. We first applied this technology to one sample type and provided evidence that Archaea and Bacteria can be differentiated in their IR-spectra due to carbohydrate content and the different ratios of $-CH_2-$ and $-CH_3$ in their lipids (Probst et al., 2013b). Apart from this taxonomic classification of infrared signals, the technology enabled demonstrating an increased content of organic sulfate in bacteria-rich regions of the biofilm. Hence, we coupled metabolomic and taxonomic information of a sample by using one single technology. In a second publication, biofilm samples from different sampling sites (Mühlbacher Schwefelquelle and Sippenauer Moor) were compared with regard to their infrared signals across multiple samples (Probst et al., submitted) and moved consequently from alpha- to beta-diversity analysis. The approach described in the above-mentioned publication is, however, based on treating the spectrum of each single pixel separately for microbial community analysis. Overlapping confidence intervals in ordination analysis are then interpreted as microbiome similarity (Probst et al., submitted). However, this may be only applicable for simple microbial communities as studied herein, and more robust measures of microbiome relationships need to be established, when complex samples with very dissimilar lipid profiles are analyzed.

A comparison of the spectra before plotting them in an ordination analysis is a potential solution for the analysis of complex microbial infrared data. The spectra can be grouped into so-called operational spectral units (OSU, in analogy to OTU, operational taxonomic unit) used for downstream multivariate statistics. Using a novel approach, we performed such an analysis on the data presented in Probst et al. (Probst et al., submitted). We first calculated a hierarchical clustering based on an Euclidean-distance measure and Ward's method by taking into account the lipid and carbohydrate spectra of each pixel (R-Development-Core-Team, 2005; Ward Jr, 1963). The nodes of the resulting dendrogram were then grouped at a certain height (dissimilarity) and each group was defined as one individual OSU with a certain abundance (number of pixels per node). Such a dendrogram is depicted in *Figure IV.2-1A*.

The resulting OSUs were exported and their abundances were used for regular multivariate statistics. *Figure IV.2-1B* shows an ordination plot, in which the two major sample types could be differentiated validating the observation made by PhyloChip G3 analysis (Probst et al., submitted). In future, this novel analysis technique could be improved by for instance applying an Unifrac algorithm using the Ward-linkage dendrogram of *Figure IV.2-1* as a tree (Lozupone and Knight, 2005). This would no longer necessitate the grouping of spectra into OSUs, and multivariate statistics could be directly applied to datasets.

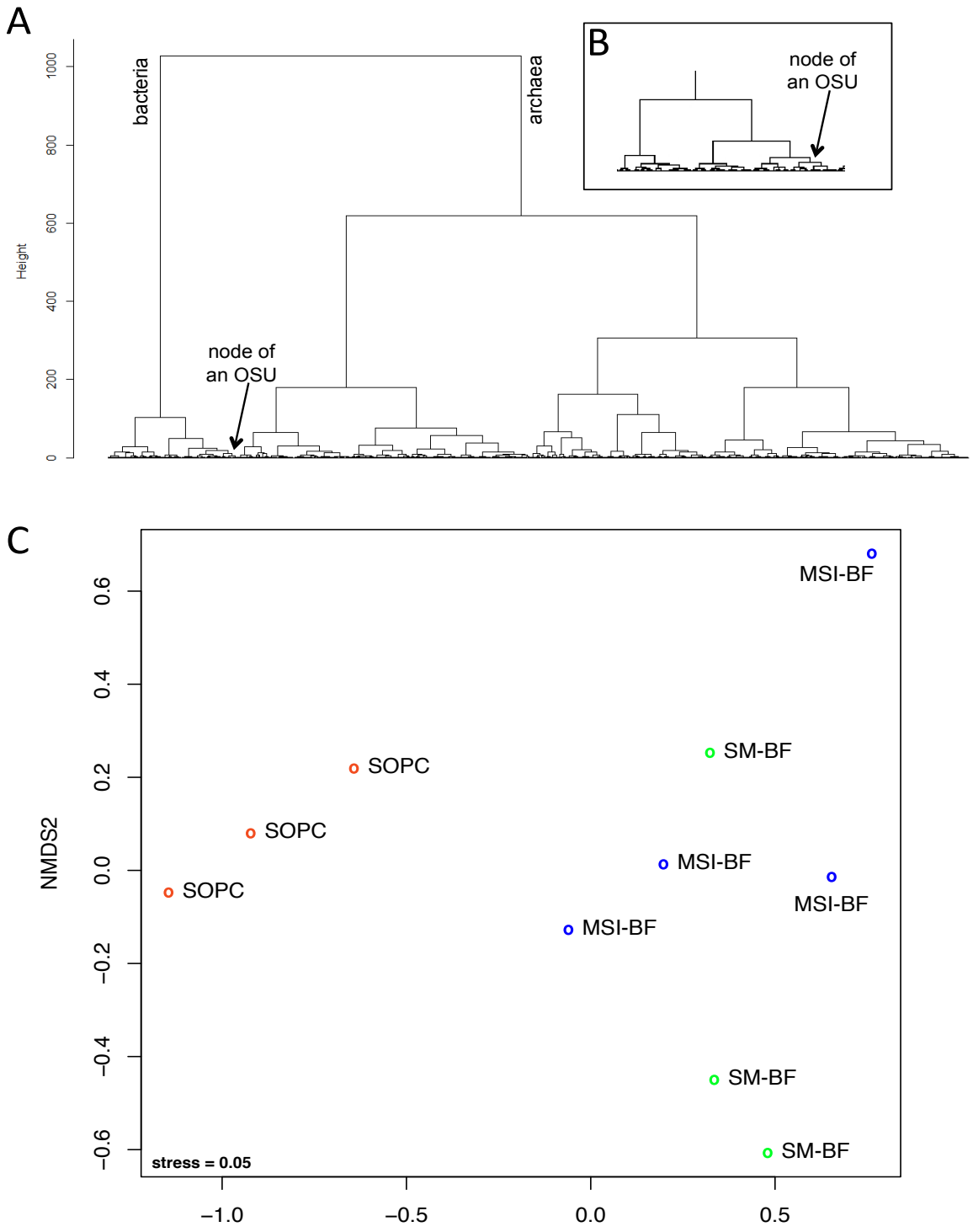

Figure IV.2-1 *OSU-based microbiome analysis of infrared-imaging data.* ***A****) Hierarchical clustering of spectra (lipid and carbohydrate regions) of biofilm and string-of-pearls community samples. Two major clusters are observed, one comprised of bacteria, the other one comprised of archaea. The dendrogram was cut at a height of 10 (dissimilarity value) for grouping of spectra into OSUs.* ***B****) Enlargement of a part of A. Arrow points to the same node as presented in A.* ***B****) Ordination analysis (NMDS, non-metric multidimensional scaling) of OSU data reveals a separation of string-of-pearls-community samples from the Sippenauer Moor (SOPC) from biofilm samples (MSI-BF from Mühlbacher Schwefelquelle and SM-BF from Sippenauer Moor). The separation is supported by a highly significant Adonis p-value (0.001) and a highly significant delta (0.001) in multi-response permutation procedure, which supported the distant grouping with a chance-corrected within group agreement of A=0.20.*

The downside of infrared-based microbiome profiling is the lack of databases with reference spectra. In contrast, 16S rRNA genes in public databases are increasing tremendously in size (McDonald et al., 2012). However, for every other microbiome profiling technique currently available (including comparative metagenomics and lipidomics), the extraction of chemical compounds like nucleic acids, proteins or lipids from the microbial sample is mandatory. This is not only destructive to the sample, it also creates biases as some cells may get lysed more easily than others increasing their signals in the profile. In contrast, infrared-based analysis appeared to be highly congruent with quantitative PCR resulting in absolute abundances of biomass (Probst et al., submitted), and potentially displaying an improved microbiome relationship that is sought after in 16S rRNA gene based analysis. An ambiguous aspect of infrared-imaging is capturing the change of lipid composition of the same microorganism as a response to environmental conditions (Meador et al., 2014). Ergo, infrared microbiome profiling may be more sensitive in detecting minor environmental changes than nucleic acid-based technologies. These changes do, however, not always result from differences in the microbial community but from differential gene expression.

In summary, the approach of using infrared spectra for microbiome profiling has advantages and disadvantages but represents an attractive alternative to destructive microbiome profiling methods (an according patent has been submitted[1]). Future studies may combine regular, 16S rRNA gene-based, microbiome profiling with infrared analysis to more comprehensively elucidate microbiome changes in environmental samples.

3. "*Candidatus* Altiarchaeum hamiconnexum" a.k.a. SM1 Euryarchaeon

3.1. The time after the SM1 Euryarchaeon 'pangenome'

The metagenomic survey on biofilm samples taken from the Mühlbacher Schwefelquelle enabled the reconstruction of a genomic bin for the SM1 Euryarchaeon. Analysis of single-copy genes revealed, however, that the bin consisted of several genomes of different strains of SM1 Euryarchaeon (Probst et al., in preparation). In order to harvest sufficient DNA for metagenomics, multiple samples from different sampling events were pooled. In case different strains occur at different sampling times, the observed strain variation could result in a heterogenous genome bin. However, this hypothesis can be rejected as analyses of the intergenic spacer region (between 16S and 23S rRNA gene) have shown that different strains of the SM1 Euryarchaeon are present in one sample, i.e. spring water filtered for biofilm droplets over a period of three days (Probst et al., submitted). Consequently, the biofilm washed up from deeper Earth layers may have a heterogenous composition of multiple SM1 Euryarchaeon strains. Nevertheless, further analyses with single biofilm droplets from one sampling event are necessary. These analyses should be designed to either verify or falsify the contribution of two or more SM1 Euryarchaeon strains to the same subsurface biofilm. Methodological approaches to determine the strain variation of SM1 Euryarchaea in biofilm samples are manifold, but a combination of them may be necessary to exclude biases arising from one single technique. First,

[1] 'Rapid and label-free infrared imaging for microbial community screening and profiling' Giovanni Birarda, Alexander J Probst, Hoi-Ying N Holman. U.S. patent application serial number 61/908,014 filed on 22 November 2013

existing SR-FTIR data should be revisited, and the different lipid spectra observed for the strains at the two sampling sites should be explored. A rough estimate of the overlapping spectra from another sampling site (Sippenauer Moor) suggests a strain differentiation with a proportion of 1:2 for biofilm samples from the Mühlbacher Schwefelquelle indicative of one dominant SM1 Euryarchaeon strain. Second, single biofilm droplets should be individually used for DNA extraction and intergenic spacer region sequencing. The diversity of intergenic spacer regions of one biofilm droplet may provide conclusions on the heterogeneity. Using those recovered sequences of one single biofilm droplet, FISH-probes that target these different intergenic spacer regions can be designed and applied. Quantitative FISH of individual biofilm pieces could potentially reveal the presence and abundance of one or two individual SM1 Euryarchaeon strains (Schmid et al., 2001).

Having a genomic bin of several strains at hand enabled us to predict a metabolism for the SM1 Euryarchaeon (Probst et al., in preparation). However, it remains speculative if every gene of this metabolic model is present in every strain of the SM1 Euryarchaeon or if a niche differentiation occurred. Although the main metabolic pathways were confirmed with a second genome of one population from the Crystal Geyser aquifer in the USA (Probst et al., in preparation), it is desirable to reconstruct a complete genome of a single SM1 Euryarchaeon strain, preferably of the dominant one. Since the reconstructed SM1 Euryarchaeon genome from the Crystal Geyser dataset appeared to be fragmented as well, other technologies than illumina and 454 pyrosequencing are necessary in order to account for strain sequence variations. Attractive alternatives resulting in longer reads than the aforementioned sequencing technologies are PacBio sequencing or the cloning and sequencing of larger fragments as performed by Banfield and co-workers in 2004 (McCarthy, 2010; Tyson et al., 2004). Another approach may be single cell sorting, whole genome amplification, and sequencing, which would, however, result in an incomplete genome (Lasken, 2007; Zhang et al., 1992). Performing single cell genomics multiple times and combining the sequencing data of identical strains of the same population could in turn yield a complete genome from one strain. Furthermore, the combination of different DNA extraction methods and separate sequencing as presented by Albertsen and co-workers may be the method of choice for any further SM1 Euryarchaeon-related metagenomics projects (Albertsen et al., 2013). Nevertheless, gaps in the genome data may have to be closed via elaborative methods in any of the aforementioned cases (Beigel et al., 2001; Guo and Xiong, 2006; Reddy et al., 2008).

The SM1 Euryarchaeon biofilm was also discovered at a second sampling site near Regensburg, the Sippenauer Moor (Henneberger et al., 2006), where potentially less strain variation existed but sampling under anaerobic conditions is impossible (Probst et al., submitted). Sequencing efforts of those biofilms could, however, also lead to a better genome reconstruction. Moreover, the strains found at the Sippenauer Moor and the Mühlbacher Schwefelquelle are different (Probst et al., submitted) enabling the discovery of the core gene set of the SM1 Euryarchaeon via comparative genomics.

Taken together, it is desirable to improve the existing quality of the SM1 Euryarchaeon genome in order to create larger contigs or even one single one. As a result, a complete or near-complete genome would empower researchers to look into the genome organization of the SM1 Euryarchaeon

as well as potentially improve the existing function prediction of proteins encoded in the genome using synteny-based approaches available in e.g. MaGe (Vallenet et al., 2013; Vallenet et al., 2006).

Regardless of the completeness of the genome, a metagenome-wide transcriptomic approach should be applied. The metatranscriptomic data could lead to a better understanding of the essential genes of the SM1 Euryarchaeon contributing to its environmental success. With regard to a limited set of twelve metabolic genes, we conducted such a survey (Probst et al., in preparation), which helped us to affirm central metabolic predictions. A large metatranscriptomic analysis of biofilm samples would, however, drastically restrict the amount of genes that are important for pathway predictions. The relative abundance of gene expression profiles should be further investigated using quantitative PCR (of cDNA) as some genes may be of low expression and their contribution to the metabolic profile may be accordingly low. This approach could further limit the amount of potential metabolic pathways performed by the SM1 Euryarchaeon in the subsurface and strengthen the understanding of its ecology.

3.2. From metabolic predictions to cultivation of the SM1 Euryarchaeon

The major goal of transcriptomic and genomic investigations of the SM1 Euryarchaeon biofilms is to better understand the metabolism of this archaeon. This could not only enable its cultivation under laboratory conditions but help to understand its contribution to global geochemical cycles. From the genomic data we concluded that the SM1 Euryarchaeon has two independent C_1-carrier pathways, a.k.a. acetyl-CoA (CoA: Coezyme A) pathways, that use either tetrahydromethanopterin or tetrahydrofolate as cofactors (Berg et al., 2010; Maden, 2000; Probst et al., in preparation). Of all available archaeal genomes, there are only few that encode for both, the tetrahydromethanopterin and the tetrahydrofolate pathway. Those archaea with both pathways use one for bioenergetics to create sufficient redox potential for proton gradient generation. In similar manner, methylotrophic bacteria utilize both C_1-carrier systems, whereas a C_1-groups bound to tetrahydromethanopterin exhibit a greater potential for oxidation (Chistoserdova et al., 1998; Maden, 2000). Therefore, it is likely for the SM1 Euryarchaeon that the pathway with tetrahydromethanopterin is used for bioenergetics, while the one with tetrahydrofolate is involved in anabolism (Probst et al., in preparation).

The SM1 Euryarchaeon has a novel, archaeal acetyl-CoA pathway with tetrahydro-methanopterin as C_1-carrier. The novelty of this pathway lies in its independence of Factor$_{420}$ (F_{420}): Enzymes that necessitate the reduced F_{420} in the acetyl-CoA pathway are replaced with potential xenologs that use NAD(P)H (NADPH: reduced nicotinamide adenine dinucleotide phosphate; NADH: reduced nicotinamide adenine dinucleotide) as cofactors. The different redox potentials of F_{420} and NAD(P)H favor the oxidation of C_1-compounds in the pathway similar to the metabolism of methylotrophs (Chistoserdova et al., 1998), although the (unknown) intracellular concentration of the reduced and oxidized cofactors must be considered. Ergo, this is further evidence that the tetrahydromethanopterin pathway is used for bioenergetics in SM1 Euryarchaeon but oxidizable C_1-compounds must be provided to the system. The presence and transcription of an acetyl-CoA synthetase in the SM1 Euryarchaeon genome is indicative of an acetate-based metabolism, where acetate is activated to acetyl-CoA. The carbon monoxide dehydrogenase/acetyl-CoA synthase cleaves the acetyl-CoA to a methyl group and carbon monoxide. The methyl group is likely to be

transferred to tetrahydromethanopterin, and the two products, methyl-tetrahydromethanopterin and carbon monoxide, may then be both oxidized to carbon dioxide resulting in two NAD(P)H and two reduced ferredoxins (Probst et al., in preparation). The latter are used to generate a proton gradient via a corresponding oxidoreductase, but the terminal electron acceptors are unknown as no potential genes mediating electron transfer in the membrane were identified (Probst et al., in preparation). Although no catalytic subunit could be identified, electron shuttling subunits for hydrogenases were identified in the genome. The electron shuttling subunit was shown to be transcribed, and according maturation enzymes of hydrogenases were encoded in the SM1 Euryarchaeon genome (Probst et al., in preparation). However, if acetate is oxidized to carbon dioxide, protons cannot be the final electron acceptors when considering bioenergetics. Hence, hydrogen may be a side-product or a side-educt of SM1 Euryarchaeon, and other electron acceptors must be considered (Probst et al., in preparation).

Under anaerobic conditions, inorganic, oxidized compounds like nitrate, nitrite, sulfate or sulfite may serve as electron acceptors, but genes encoding for according enzymes were not detected in the genome. In general, sulfate reduction may be possible without activation to adenylylsulfate if a sufficient amount of reduced ferredoxin is present in cells (Thauer, 2011). Moreover, the non-thermophilic sulfate-reduction to elemental sulfur performed by archaea was shown to take place in ANME (anaerobic methanotrophic archaea), but a lack of knowledge on any enzyme involved must be stated (Milucka et al., 2012). Due to the absence of nitrate and nitrite in the spring water, the high concentration of sulfate is indicative of being the terminal electron acceptor for the SM1 Euryarchaeon. Nevertheless, genes encoding for enzymes pinpointing to potential electron acceptors may be either missing due to the fragmentation of the genome or have no available homologs in public databases and simply escape identification (similar to the enzymes for the above-mentioned non-thermophilic sulfate reduction performed by ANME). Consequently, two potential electron donors for bioenergetics, namely carbon monoxide and acetate, of the SM1 Euryarchaeon are known but the terminal electron acceptor remains enigmatic.

Apart from bioenergetics, acetate may serve in the anabolism of the SM1 Euryarchaeon after activation to acetyl-CoA. Beyond acetate, three potential carbon sources could be envisaged from the genomic data. Formate, carbon dioxide and carbon monoxide could be used for generation of acetyl-CoA, which would, however, necessitate the reductive direction of the acetyl-CoA pathway. Since an oxidation of C_1-compounds is more likely (see above), carbon dioxide and carbon monoxide fixation are rather unlikely. Instead, carbon monoxide and formate could be utilized to produce redox equivalents like reduced ferredoxin and NAD(P)H, respectively. Carbon monoxide is therefore a potential educt for bioenergetics (Probst et al., in preparation).

In the future, further evidence for the carbon source of the SM1 Euryarchaeon could be gathered either via stable-isotope probing or via nano-SIMS (secondary ion mass spectrometry) with ^{13}C-labeled substrates (Neufeld et al., 2007; Radajewski et al., 2000), whereas acetate is the most attractive compound in the entire scenario. Although all cultivation attempts have failed hitherto, the identification of potential carbon and nitrogen sources for SM1 Euryarchaeon is an important prerequisite, which scientists have now access to. Hence, future cultivation or enrichment experiments should be systematically designed based upon the available genome data in order to

evaluate potential electron acceptors. These cultivation/enrichment experiments should meet the following criteria:

i) The spring water of one sulfidic spring, where the SM1 Euryarchaeon is found, should serve as the basis. **ii**) A suitable carbon source, either acetate, formate or combinations thereof, and sufficient ammonia as nitrogen source must be provided. **iii**) Nitrate, nitrite, sulfate, sulfite or thiosulfate should be added as potential electron acceptors, and their concentration along with according reduction products in the medium should be monitored. **iv**) Gas phases must be adjusted accordingly concerning H_2, CO_2, and CO. The gas phase should contain either H_2, assuming hydrogen consumption, or no H_2 at all assuming hydrogen production. The same applies for carbon dioxide, assuming either carbon fixation or acetate/C_1-oxidation. CO could be a potential electron donor. The gas composition must be determined regularly throughout the cultivation experiments. **v**) Trace elements should be added, since the formly methanofuran dehydrogenase was shown to be tungsten/molybdate dependent (Schmitz et al., 1992a; Schmitz et al., 1992b; Schmitz et al., 1994; Schmitz et al., 1992c; Vorholt et al., 1996) and according transporters are encoded in the SM1 Euryarchaeon genome (Probst et al., in preparation). **vi**) Assuming that the isolation of the SM1 Euryarchaeon is not possible throughout the experiments, the enrichment cultures should be carried out at least five times each in order to conclude on consumption or production of the aforementioned substances. **vii**) The growth and contribution of other microorganisms to potential energy cycling must be controlled rigorously to suppress e.g. sulfate-reducing bacteria as described previously (Saleh et al., 1964), unless a syntrophic relationship is considered. **viii**) The growth of SM1 Euryarchaeon should be monitored regularly with either rRNA (cDNA) or PMA-treated DNA as template for quantitative PCR using highly specific primers solely targeting the SM1 Euryarchaeon. Two other archaea (IM-C4 and one other genomic bin) were detected in metagenomics and PhyloChip G3 revealed signatures of methanogens (Probst et al., 2013b; Probst et al., in preparation). These other archaea must be excluded from any assay targeting SM1 Euryarchaeal cell number enumeration. This small amount of possible cultivation attempts adds up to 90 different combinations. The complex setting makes potential cultivation attempts very time consuming. However, based on the predicted metabolism in Probst et al., there is one combination of substances that is most attractive (Probst et al., in preparation): Nitrogen gas phase, acetate, carbon monoxide, formate and (thio)sulfate.

In case future cultivation and isolation attempts continue to fail, other technologies must be applied to further understand the metabolism of the SM1 Euryarchaeon. For the current metabolic model we used metagenomics to predict genes and their according function, but this technique gives no information regarding gene transcription. Metatranscriptomics in turn provides information regarding the mRNA pool (Poretsky et al., 2005), but the translation and full expression of genes can only be assessed by metaproteomics (Wilmes and Bond, 2004). Although the application of metaproteomics was for instance successful in the case of ANME-1, it necessitates a complete or near-complete genome of the organism of interest (Stokke et al., 2012). Consequently, the genome of the SM1 Euryarchaeon needs to be improved before metaproteomics can be performed. However, there are other techniques that scientists could utilize to generate a more complete picture of the SM1

Euryarchaeon metabolism. For instance, recombinant expression and subsequent enzymatic analysis or crystal structure determination may enable improved function prediction for genes that are transcribed in the SM1 Euryarchaeon. Such a gene of interest may be the putative methyl-tetrahydrofolate reductase. Homologs of this enzyme act in pterin pathways of bacteria and archaea and are in general essential for purin and amino acid anabolism (Maden, 2000). Many archaea possess complementary pterin pathway, where tetrahydromethanopterin is the carrier for C_1-groups instead of tetrahydrofolate (Maden, 2000). From the genomic data available in public databases it can be concluded that only few organisms utilize both pathways, like *Methanosarcina barkeri* (Buchenau and Thauer, 2004). In contrast, analysis of the genome of "*Candidatus* Methanomassiliicoccus intestinalis" (Borrel et al., 2013) indicates that this organism has the entire tetrahydrofolate pathway but is missing the tetrahydromethanopterin pathway. This is, however, rather the exception than the norm for archaea. On the other hand, methylotrophic bacteria possess both C_1-carrier systems, and consequently, neither one of the two pathways appears to be restricted to one of the two domains of Prokaryotes (Chistoserdova et al., 1998). Indeed, some enzymes were even shown to be catalytically active with both coenzymes (Vorholt et al., 1998). Hence, the missing methyl-tetrahydromethanopterin reductase in SM1 Euryarchaeon is potentially replaced by a putative methyl-tetrahydrofolate reductase (Probst et al., in preparation). This compensation would be similar to the one proposed for ANME-1, where the same enzymatic step is missing and only one putative enzyme in the tetrahydrofolate pathway is present, namely the methyl-tetrahydrofolate reductase (Hallam et al., 2004; Meyerdierks et al., 2010). This hypothesis is supported by the fact that this enzyme shows a similar relative quantity in protein mass to enzymes involved in reverse methanogenesis (Stokke et al., 2012). In future experiments, the corresponding gene of interest in the SM1 Euryarchaeon genome should be artificially designed to meet the codon-usage requirements of *E. coli*. After its recombinant expression, enzymatic assays should be designed to determine if the methyl tetrahydrofolate reductase can really compensate for the predicted metabolic reaction. Such an experiment would also have direct implications on anaerobic methane oxidation research, where the complete reverse methanogenesis via the tetrahydromethanopterin pathway has only been shown for ANME-2 but not for ANME-1 (Haroon et al., 2013).

Another interesting candidate to elucidate the metabolic predictions of SM1 Euryarchaeon may be a potential hydrogenase. The according gene is located next to a putative coenzyme factor 420 hydrogenase/dehydrogenase subunit B, which may act as electron shuttle for a hydrogenase. The protein of unknown function is predicted to be located in the cytoplasmic membrane and shows some similarity (~25%) to NADH dehydrogenases. Ergo, it may be an interesting candidate for the recombinant expression and determination of its 3D structure, which may help to predict a function for the protein.

It is hypothesized that the SM1 Euryarchaeon is a potential sulfate-reducer (Moissl et al., 2002), and current investigations support this hypothesis (Probst et al., in preparation). The archaeon showed a great proportion of its proteins to be homologues to those of ANME-1 and may perform a similar sulfate reduction as determined for a recently described ANME strain (Milucka et al., 2012). However, no genomic information of this ANME strain studied by Milucka et al. is currently publicly

available. Once accessible by the public, the genomic information will enable comparative genomics with the SM1 Euryarchaeon to identify potential genes of interest that could be studied *in vitro* after recombinant expression. In sum, analyzing recombinantly expressed genes of the SM1 Euryarchaeon could help to understand its metabolism and should be pursued in case cultivation attempts continue to fail.

3.3. The interspecies relationship of the string-of-pearls community revisited

The string-of-pearls community was defined to be a close association of two microorganisms, the SM1 Euryarchaeon and sulfide-oxidizing bacteria like *Thiothrix* sp. or *Sulfuricurvum* sp. (Moissl et al., 2002; Moissl et al., 2003; Rudolph et al., 2004; Rudolph et al., 2001). The consortium is the result of SM1 Euryarchaeon biofilm samples washed up from deeper Earth layers and their exposure to oxygen (Probst et al., submitted). It appears to be obvious that the SM1 Euryarchaeon is seeking the partnership with a sulfur-oxidizing bacterium (Moissl et al., 2002; Rudolph et al., 2004) most likely mediated by the *hami* (Probst et al., submitted).

Moissl et al., 2002 pointed out that the SM1 Euryarchaeon is a potential sulfate-reducer, which can currently not be denied. This assumption is based on the fact that members of the genus *Thiothrix* and *Sulfiricurvum* are able to oxidize sulfur compounds. However, speculations deriving from 16S rRNA gene phylogeny can also lead to different conclusions. For instance, members of the genus *Thiothrix* were also reported to be mixotrophic (Larkin, 1980) and could feed on dead SM1 Euryarchaea cells. In another scenario, the SM1 Euryarchaeon could be a nitrite oxidizer, since members of both genera, *Thiothrix* and *Sulfuricurvum*, have been reported to be able to reduce nitrate to nitrite (Kodama and Watanabe, 2004; Larkin and Shinabarger, 1983). The latter hypothesis would include an interspecies nitrogen cycle. Nevertheless, nitrate and nitrite were below detection limit in the spring water of the Mühlbacher Schwefelquelle and the Sippenauer Moor and can thus most likely be excluded as terminal electron acceptors for the SM1 Euryarchaeon.

Currently, it is unknown if the SM1 Euryarchaeon and the sulfur-oxidizing bacteria are the only microbes involved in this interaction. Indeed, some results pinpoint to another type of microorganisms, namely sulfate-reducing bacteria. In this paragraph, current knowledge on the metabolism of SM1 Euryarchaeon and of the sulfate-reducing bacteria in biofilm samples is adduced to provide an new perspective on the string-of-pearls community.

In the course of this study, results evidenced the different bacteriomes of the biofilm samples and the spring water at the Mühlbacher Schwefelquelle. While the spring water has many different bacteria and none of them appeared to be dominant, the biofilms carried — apart from SM1 Euryarchaeon — mainly sulfate-reducing bacteria (Probst et al., 2013b). Similar results were retrieved for the second biofilm found at the Sippenauer Moor, which represents an independent system with a different SM1 Euryarchaeon strain compared to the Mühlbacher Schwefelquelle (Probst et al., submitted). However, a random attachment of bacteria from the spring water to the biofilm would lead to similar bacteriome compositions of the water and biofilm. Ergo, the SM1 Euryarchaeon appears to provide a habitat suitable for sulfate-reducers. A competition of the latter with the former may also be excluded due to their appearance in relatively high numbers (5%) in the community; in comparison, the IM-C4 Euryarchaeon was revealed to have a similar metabolism as the SM1 Euryarchaeon and

showed an abundance of only <0.1% pointing to a real competition in the biotope (Probst et al., in preparation). As a result, the SM1 Euryarchaeon biofilm seems to enrich sulfate-reducing bacteria, which were shown to be active in sulfate-reduction for respiration (Probst et al., 2013b). However, sulfate-reducing bacteria can use a variety of (organic and inorganic) electron donors including molecular hydrogen or acetate (Liamleam and Annachhatre, 2007). Acetate may play an important role for SM1 Euryarchaeon presumably performing an acetoclastic metabolism (Probst et al., in preparation). The terminal electron acceptor is currently unknown but indications for genes encoding hydrogenases were found in the SM1 Euryarchaeon genome (Probst et al., in preparation). These hydrogenases could either mediate hydrogen consumption or transfer electrons to protons resulting in hydrogen production. Considering the balance of oxidizing one molecule of acetate via the acetyl-CoA pathway present in SM1 Euryarchaeon, two reduced ferredoxins and two reduced NAD(P)H are gained. While ferredoxin can be used to build a proton gradient via an oxidoreductase encoded in the genome, NAD(P)H may only be used for anabolism by SM1 Euryarchaeon. However, exceeding amounts NAD(P)H must be re-oxidized to regenerate $NAD(P)^+$ to keep reactions energetically favorable, which could be performed via an aforementioned hydrogenase. The result would be low amounts of hydrogen emitted by the SM1 Euryarchaeon, which could be consumed by sulfate-reducing bacteria and enhance their enrichment in e.g. biofilm samples compared to the surrounding spring water. In this scenario, hydrogen would be a side-product of the SM1 Euryarchaeon, sufficiently enough to provide growth for a minor bacteriome under oxygen-free conditions. Considering the ratio of archaeal 16S rRNA genes and *dsrB* genes of sulfate-reducing bacteria in biofilm samples and string-of-pearls community samples, these genes appeared in similar proportions in both sample types [biofilm at Mühlbacher Schwefelquelle: 0.18, biofilm at Sippenauer Moor: 0.09, string-of-pearls community at Sippenauer Moor: 0.17 per µg DNA (Probst et al., submitted; Probst et al., 2013b)]. Furthermore, the presence of sulfate-reducing bacteria in the string-of-pearls community is supported by the detection of their 16S rRNA gene signatures (Probst et al., submitted; Probst et al., 2013b). Although their relative abundance in 16S rRNA gene sequences decreases in the string-of-pearls community [due to the overwhelming abundance of *Thiothrix* sp. (Moissl et al., 2002)], their absolute abundance with regard to archaea persists. Consequently, sulfate-reducing bacteria appear to be a persistent member of the potentially syntrophic relationship between SM1 Euryarchaeon and sulfur-oxidizing bacteria in the string-of-pearls community.

A new model based on the interaction of SM1 Euryarchaeon with sulfate-reducing and sulfur-oxidizing bacteria is predicted and depicted in *Figure IV.3-1*. Here, sulfur-oxidizing bacteria (*Thiothrix* sp. or *Sulfuricurvum* sp.) provide sulfate and an oxygen free environment to the interior of the pearl, which was indeed shown to be oxygen-free (Rudolph, 2003). Both the oxygen-depleted environment and the sulfate are of considerable importance for the consortium of SM1 Euryarchaeon and sulfate-reducing bacteria: The SM1 Euryarchaeon was shown to be a strictly anaerobic microorganism (Probst et al., in preparation) as are sulfate-reducing bacteria. The latter potentially use the sulfate produced by sulfur-oxidizing bacteria for anaerobic respiration. The electron donor, most likely molecular hydrogen, is provided by the SM1 Euryarchaeon, which has the only benefit of an anaerobic environment in this model. However, other benefits like acetate from the surrounding

microorganisms cannot be excluded at this point. Based on this new model, the interaction of microorganisms in the string-of-pearls community predicted by Moissl et al. 2002 needs to be extended by one more group of microorganisms, which are sulfate-reducing bacteria.

The model depicted in *Figure IV.3-1* does not include the possibility of the SM1 Euryarchaeon being a sulfate-reducer similar to ANME-2 [see above (Milucka et al., 2012)]. However, assuming the SM1 Euryarchaeon performs sulfate-reduction resulting in intracellular molecular sulfur, polysulfide would be the intermediate between SM1 Euryarchaeon and sulfate-reducing bacteria. The latter would perform a disproportionation with sulfide and sulfate as terminal products (Milucka et al., 2012), which are then metabolized by SM1 Euryarchaeon and sulfur-oxidizing bacteria, respectively. Similar to the hydrogen-based model depicted in *Figure IV.3-1* this model involves the necessity of sulfate-reducing bacteria but has different/additional intermediates.

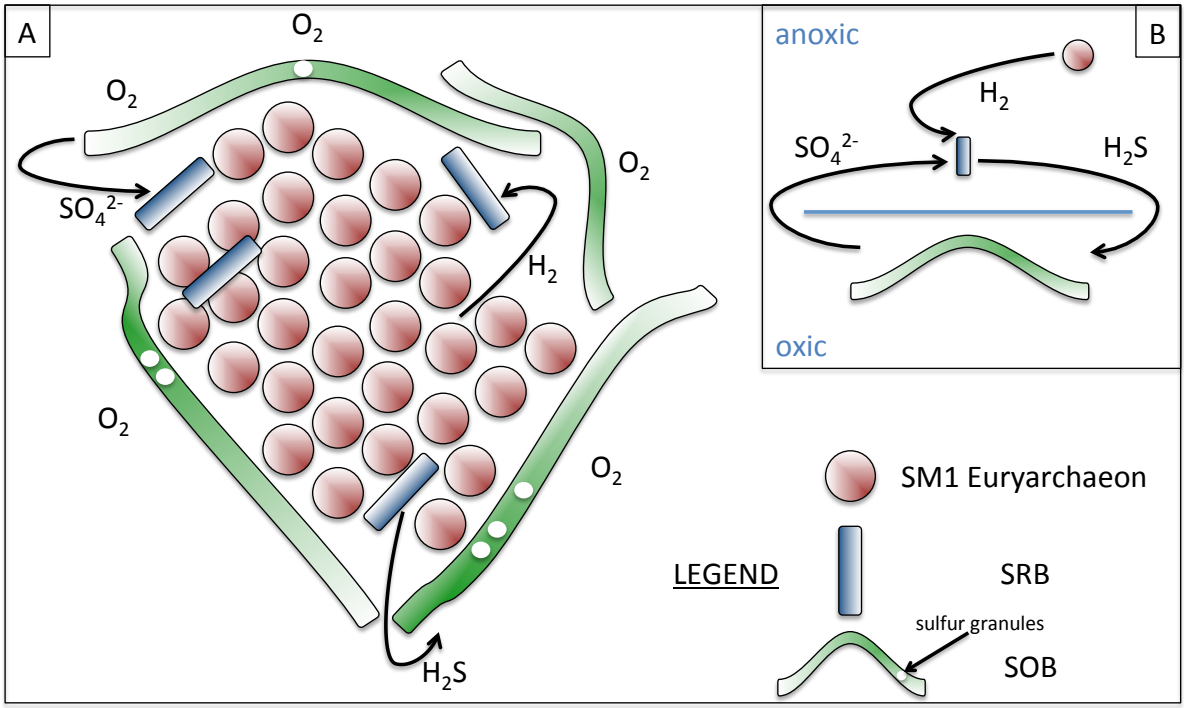

Figure IV.3-1 *Model of the string-of-pearls community based on the syntrophic interaction of three different types of microorganisms.* ***A****) Model of a pearl. The sulfate-reducing bacteria (SOB) surround the archaea (SM1 Euryarchaeon) and the sulfate-reducing bacteria (SRB) protecting them from oxygen via respiration and sulfide oxidation. SOB provide sulfate to the SRB. SM1 Euryarchaeon provides hydrogen (or potentially acetate, see main text for discussion) for the SRB, which in turn provide sulfide for the SOB. Bright circles in SOB symbolize elemental sulfur, S_8.* ***B****) simplified model of A) dividing the process into oxic and anoxic metabolic processes.*
Please note, that this is a simplified model based on the current knowledge on the metabolism of SRB (sulfate-reduction) and the coding potential for a hydrogenase of the SM1 Euryarchaeon. Further involvements of other intermediates (e.g. polysulfide, please see main text) can currently not be excluded.

In order to understand the interaction between sulfate-reducing bacteria, SM1 Euryarchaeon and sulfide-oxidizing bacteria, future investigations should include the recovery of genomic and transcriptomic information of all major bacteria in the systems, i.e. sulfate-reducing bacteria and sulfur-oxidizing bacteria. In the course of this study, we confirmed that biofilms samples are precursors of the string-of-pearls community (Probst et al., submitted) and the same SM1 Euryarchaeon strain persists in both life styles. This situation enables researchers to explore the

biology of the same organism in two different biotopes: the subsurface and the surface. Different conditions like oxygen-increase and the interaction with different microorganisms may alter the gene expression of the SM1 Euryarchaeon. Future investigations should aim for understanding the differences between biofilm and string-of-pearls community on a transcriptomic level. These studies would pioneers in describing the differential gene expression of the very same archaeon in two different natural biotopes.

3.4. Dominance in the dark: conclusions from SM1 Euryarchaeon research

The metagenomic research performed herein, enabled the reconstruction of nine genomic bins, six different bacteria and three different archaea (Probst et al., in preparation). As 16S rRNA genes are generally barely recovered from metagenomic assemblies, only three out of nine genomes exhibited 16S rRNA genes: SM1-MSI (SM1 Euryarchaeon), IM-C4 (IM-C4 Euryarchaeon) and Bac08-MSI. The latter genome was identified to belong to a microbe of the OD1 candidate phylum, whose members are designated sulfate-reducers (Wrighton et al., 2012) pinpointing the importance of sulfate-reducers in the SM1 Euryarchaeon's biofilm. Besides the SM1-MSI and the IM-C4 bin, the genome of one more archaeon with at G/C content of approximately ~38% was successfully recovered from the sequencing data set. Although the genome has less than 1 Mbps in size, the completeness was estimated to be 96%. Since no 16S rRNA gene could be reconstructed for this genome, its phylogenetic position has not yet been determined, although a first glimpse at its genome revealed some interesting findings. Phylogenetically, this archaeon may belong to the Euryarchaeota without a cultivated representative close by. No genes for any sugar metabolism including gluconeogenesis were identified. The presence of encoded amino-acid transporters is suggestive of a heterotrophic metabolism and someone may speculate that this genome belongs to a parasitic archaeon. In similar manner, the genome of the IM-C4 Euryarchaeon has not yet been analyzed thoroughly, although methanogenesis can here be ruled out due to missing key enzymes in the genome. Ergo, apart from the SM1 Euryarchaeon bin, no other of the nine reconstructed genomes has been analyzed in depth, and future investigations tailored to determine the phylogeny and metabolic capability of these microorganisms are warranted. In addition, a functional and taxonomic classification of unassembled reads should be performed in case no genomes of the potentially syntrophic sulfate-reducing bacteria could be reconstructed.

The SM1-MSI Euryarchaeon, the IM-C4 Euryarchaeon (Probst et al., 2013b), and the SM1-CG Euryarchaeon (SM1 Euryarchaeon genome recovered from the Crystal Geyser, Utah, US) were phylogenetically characterized using concatenated ribosomal protein sequences (Probst et al., in preparation). The archaea are considered representatives of a novel, non-methanogenic euryarchaeal lineage, sibling to the Methanococcales. A paper by Lovley's group published in *Nature* described archaea of the SM1 Euryarchaeon clade to dominate a subsurface water current [Lidy Hot Springs, US (Chapelle et al., 2002)]. The subsurface biotope was discussed to be analogous to extraterrestrial subsurface settings on Mars and Europa. Due to the low amounts of complex carbon in the biotope, the archaea were designated methanogens without any further biological evidence than *ca*. 500-bps short 16S rRNA gene sequences. However, there is no evidence that members of the SM1 clade are methanogenic and, based on our results, the entire study by Lovely and co-workers must be

questioned. Taking together the dominance of the SM1 Euryarchaea clade at several different biotopes (Mühlbacher Schwefelquelle, Sippenauer Moor, Crystal Geyser, Lidy Hot Springs), the SM1 clade appears to represent the only archaeal lineage reported in the literature that is capable of dominating one biotope and of out-competing other microorganisms in a so far unexplored manner.

One of our articles about the Mühlbacher Schwefelquelle was featured by *Astrobiology Magazine*[1] as sulfidic springs may resemble extraterrestrial biotopes due to the oxygen-depleted environment (Probst et al., 2013b). However, the metabolism of the SM1 Euryarchaeon may not fulfill all the requirements allowing a chemolithoauotrophic growth in extraterrestrial settings. As outlined above, the SM1 Euryarchaeon may have an acetoclastic metabolism and thus grow on organic substances that are usually not available without primary producers in a biotope. Acetate in the Mühlbacher Schwefelquelle may derive from acetogenic microorganisms in anoxic sediment layers above the water current. *Hami*-mediated biofilm formation by the SM1 Euryarchaeon may enable the archaea to filter acetate and oxidizable C_1-compounds that seep through the sediment layers into the groundwater. The necessity to filter for substrates would explain the regular distance of cells in the biofilm: The constant space between the cells may be a tradeoff between stability mediated by the *hami* and maximum distance in order to optimize the filtering mechanism and nutrient uptake. At the same time, the maximized distance between cells in the biofilm may enable SM1 Euryarchaea to efficiently dispose metabolic products like hydrogen or polysulfide. The high pressure and constant water-flow in the subsurface may support the withdrawal of these products. Other than the *hami*, the biofilm matrix may enable the SM1 Euryarchaeon biofilms to become barely accessible for other microorganisms. Indeed, it may bar others from the precious nutrients seeping into the biotope and exported proteases encoded in the genome may additionally help in suppressing the surrounding bacteria (Probst et al., in preparation). Moreover, biofilm formation may empower the SM1 Euryarchaeon to cope with stress conditions. Consequently, the biological costs of biofilm formation via sugars and *hami* may be worth their price and lead to the dominance of the SM1 Euryarchaeon in its biotope.

Although the metabolism of the SM1 Euryarchaeon could not be entirely decoded in the course of this thesis, the results appear encouraging so far. The metagenomic data enabled us to identify potential carbon and nitrogen sources as well as electron donors for future cultivation attempts of the SM1 Euryarchaeon in the laboratory. Meanwhile, analyses regarding the diversity and the function of the associated bacteriome in the SM1 Euryarchaeon biofilm were successfully deciphered showing a dominance of sulfate-reducing bacteria in the bacterial minority (Probst et al., submitted; Probst et al., 2013b). Including a second sulfidic spring, different SM1 Euryarchaeon biofilms from different biotopes appeared to have differences in their microbiome at various levels, from phylum down to strain level. These results highlight the complexity and versatility of subsurface microbial life making metagenomics challenging (Probst et al., submitted): The metagenomic survey of biofilms from one biotope resulted in a ‘pangenome’ of the SM1 Euryarchaeon, but also yielded in many other genomic bins. Three different archaea and several different bacteria of diverse phyla were successfully

[1] http://www.astrobio.net/pressrelease/5280/synchrotron-infrared-unveils-a-mysterious-microbial-community
last access: 17 March 2014

reconstructed from the environment and are still waiting to be explored. The dominant organism, the SM1 Euryarchaeon was phylogenomically placed into the tree of life as a separate lineage in the Euryarchaeota with several genomic specialties. The list of the extraordinary characteristics of the SM1 Euryarchaeon (its natural biofilm formation, the domination of a biotope, and its cell surface appendages) must be extended by a novel factor$_{420}$-free, archaeal acetyl-CoA pathway and a rare archaeal double-membrane system (Probst et al., in preparation).

The results of this thesis have contributed to the understanding of the complex and vastly unexplored subsurface microbial life and shed light on the so-far enigmatic metabolic capabilities of the SM1 Euryarchaeon, "*Candidatus* Altiarchaeum hamiconnexum".

V. Zusammenfassung

Archaeen, die dritte Domäne des Lebens, sind ubiquitär in Ökosystemen und leisten einen großen Beitrag zum Energiezyklus der Erde. Nichtsdestotrotz ist ihre Verbreitung und ihr metabolisches Potential nur gering erforscht, da nur ein kleiner Anteil dieser Organismen im Labor kultiviert und studiert werden kann. Ziel der vorliegenden Arbeit war es, das Vorkommen und die Funktion von Archaeen und des begleitenden bakteriellen Mikrobioms in zwei noch relativ unerforschten Biotopen zu entschlüsseln. Da es bis heute nicht möglich war Archaeen aus diesen Biotopen zu kultivieren, wurden für diesen Ansatz Metagenomstudien und andere molekulare Methoden der mikrobiellen Ökologie herangezogen.

Das erste untersuchte Biotop waren sogenannte *Spacecraft Assembly Facilties*. Diese Reinräume, in welchen Raumfahrzeuge gebaut werden, sind durch kontinuierliche Reinigung, regulierte Luftfeuchtigkeit, und wenig Nährstoffe gekennzeichnet. Obwohl das Auftreten von Thaumarchaeota in diese Reinräumen bereits dokumentiert worden war, blieb deren Herkunft und Eintritt in den Reinraum gänzlich unverstanden und unerforscht. Im Verlauf dieser Dissertation konnte gezeigt werden, dass Thaumarchaeota auf menschlicher Haut vorkommen und menschliche Aktivität somit für das Auftreten dieser Mikroorganismen in Reinräumen verantwortlich ist. Parallel dazu wurden molekulare Methoden der Mikrobiom-Analyse mit einer Lebend-Markierung verknüpft um Einblick in das potentiell lebende Bakteriom eines Reinraums zu gewinnen. Diese Analysen sind ein Meilenstein in der mikrobiellen Reinraum-Forschung, da sie zeigen, dass weniger als ein Zehntel aller Bakterien in Reinräumen potentiell am Leben und ihre Diversität entsprechend niedrig ist.

Das zweite Ökosystem, das im Verlauf dieser Arbeit untersucht wurde, waren unterirdische Biotope, die durch sulfidische, anaerobe Schwefelquellen zugänglich waren. Vorangegangene Studien hatten gezeigt, dass die Biodiversität und das metabolische Potential von Mikroorganismen, deren Ökosysteme sich im Untergrund befinden, nur geringfügig verstanden sind. In der Tat stellen diese Biotope ein großes Reservoir an unbekannten Mikroorganismen dar, die sukkzessive erforscht werden. In dieser Arbeit wurden vor allem Schwefelquellen um Regensburg untersucht, in welchen das SM1 Euryarchaeon endeckt worden war. Dieses — in Hinblick auf Metabolismus und Genomik — noch wenig erforschte Archaeon formt einen beinhahe Monospezies-Biofilm (95% Reinheit) im Untergrund, dessen Bruchstücke kontinuierlich durch sulfidische Quellen zur Oberfläche befördert werden. Die Analysen der Biofilme in der vorliegenden Arbeit zeigten, dass das Archaeon überweigend mit aktiven Sulfat-reduzierenden Bakterien im Biofilm vergesellschaftet war. Vergleichende Analysen von zwei Biofilmen aus unterschiedlichen Quellen erbrachten den Nachweis, dass sich deren Mikrobiom zwar auf vielerlei taxonomischen Ebenen unterschied, jedoch die Funktion der Bakterien in der Gemeinschaft identisch war. Um die metabolischen Eigenschaften und die phylogenomische Position des SM1 Euryarchaeons zu entschlüsseln wurden Metagenomstudien durchgeführt. Deren Ergebnisse zeigten eindeutig eine distinkte phylogenomische Position des SM1 Euryarchaeons, für welches somit der Name "*Candidatus* Altiarchaeum hamiconnexum" vorgeschlagen wurde (Candidatus Familie "*Altiarchaeaceae*", Candidatus Ordnung "*Altiarchaeales*"). Das Genom dieses neuen Taxons kodierte für Enzyme eines neuartigen, archaeellen Acetyl-CoA-Weg ohne $Faktor_{420}$. Dieser Weg ermöglicht den Archaeen mit Doppelmembran ein anaerobes Wachstum, vermutlich durch Acetat-Oxidation. Zudem wurden Gene für die Biofilmbildung identifiziert, welche für die Synthese von Zellanhängen (kleine Enterhaken) und von Zuckern (Biofilmmatrix) kodieren. Es wird vermutet, dass der so gebildete Biofilm einen kompetitiven Vorteil gegenüber anderen Organismen darstellt, mit dessen Hilfe die Archaeen kostbare Nährstoffe (z.B. Acetat) filtern, die in das Habitat des Archaeons einsickern.

Die Ergebnisse dieser Arbeit haben zum besseren Verständnis von Archaeen und ihren koexistierenden Bakterien in zwei bis dato wenig erforschten Biotopen beigetragen. Gleichzeitig ist diese Studie ein herausragendes Beispiel dafür, wie gut die Ökophysiologie eines neuen Archaeons mit aktuellen Forschungsmethoden in der Natur studiert werden kann.

VI. Bibliography

A

Abele, G. (1950). Die Heil- und Mineralquellen Südbayerns. Suppenau bei Saal an der Donau. Geologica Bavarica *2*, 94-95.

Abulencia, C.B., Wyborski, D.L., Garcia, J.A., Podar, M., Chen, W., Chang, S.H., Chang, H.W., Watson, D., Brodie, E.L., Hazen, T.C*., et al.* (2006). Environmental whole-genome amplification to access microbial populations in contaminated sediments. Appl Environ Microbiol *72*, 3291-3301.

Adler, H.H., and Kerr, P.F. (1965). Variations in infrared spectra molecular symmetry and site symmetry of sulfate minerals. American Mineralogist *50*, 132.

Albers, S.V., and Pohlschroder, M. (2009). Diversity of archaeal type IV pilin-like structures. Extremophiles *13*, 403-410.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol *31*, 533-538.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J Mol Biol *215*, 403-410.

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., and Stahl, D.A. (1990a). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Appl Environ Microbiol *56*, 1919-1925.

Amann, R.I., Krumholz, L., and Stahl, D.A. (1990b). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J Bacteriol *172*, 762-770.

Amann, R.I., Ludwig, W., and Schleifer, K.H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev *59*, 143-169.

Andorra, I., Esteve-Zarzoso, B., Guillamon, J.M., and Mas, A. (2010). Determination of viable wine yeast using DNA binding dyes and quantitative PCR. Int J Food Microbiol *144*, 257-262.

Andres, G., and Frisch, H. (1981). Hydrogeologie und Hydraulik im Malmkarst des Molassebeckens und der angrenzenden Fränkischen-Schwäbischen Alb. Schriftenreihe Bayer Landesamt f Wasserwirtschaft *15*, 108-117.

Arrhenius, S. (1908). Worlds in the making: the evolution of the universe (Harper & brothers).

Ash, C., Farrow, J.A., Dorsch, M., Stackebrandt, E., and Collins, M.D. (1991). Comparative analysis of Bacillus anthracis, Bacillus cereus, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int J Syst Bacteriol *41*, 343-346.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol *71*, 7724-7736.

B

Baas-Becking, L.G.M. (1934). Geobiologie of inleiding tot de milieukunde (Den Haag, the Netherlands).

Badger, J.H., and Olsen, G.J. (1999). CRITICA: coding region identification tool invoking comparative analysis. Molecular Biology and Evolution *16*, 512-524.

Bae, S., and Wuertz, S. (2009). Discrimination of viable and dead fecal Bacteroidales bacteria by quantitative PCR with propidium monoazide. Appl Environ Microbiol *75*, 2940-2944.

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., and Banfield, J.F. (2010). Enigmatic, ultrasmall, uncultivated Archaea. Proc Natl Acad Sci U S A *107*, 8806-8811.

Baker, B.J., and Dick, G.J. (2013). Omic Approaches in Microbial Ecology: Charting the Unknown. ASM Microbe magazine *8*, 353-359.

Baker, B.J., Tyson, G.W., Webb, R.I., Flanagan, J., Hugenholtz, P., Allen, E.E., and Banfield, J.F. (2006). Lineages of acidophilic archaea revealed by community genomic analysis. Science *314*, 1933-1935.

Baker, G.C., Smith, J.J., and Cowan, D.A. (2003). Review and re-analysis of domain-specific 16S primers. J Microbiol Methods *55*, 541-555.

Barns, S.M., Delwiche, C.F., Palmer, J.D., and Pace, N.R. (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A *93*, 9188-9193.

Battchikova, N., Eisenhut, M., and Aro, E.-M. (2011). Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. Biochimica et Biophysica Acta (BBA)-Bioenergetics *1807*, 935-944.

Baumann, M. (1981). Hydrogeologische, hydrochemische und erschließungstechnische Verhältnisse der Schwefelquellen. . Schriftenreihe Bayer Landesat f Wasserwirtschaft *15*, 4-13.

Beier, B.D., Quivey, R.G., and Berger, A.J. (2010). Identification of different bacterial species in biofilms using confocal Raman microscopy. Journal of biomedical optics *15*, 066001-066001-5.

Beigel, R., Alon, N., Kasif, S., Apaydin, M.S., and Fortnow, L. (2001). An optimal procedure for gap closing in whole genome shotgun sequencing. Paper presented at: Proceedings of the fifth annual international conference on Computational biology (ACM).

Ben-Amor, K., Heilig, H., Smidt, H., Vaughan, E.E., Abee, T., and de Vos, W.M. (2005). Genetic diversity of viable, injured, and dead fecal bacteria assessed by fluorescence-activated cell sorting and 16S rRNA gene analysis. Appl Environ Microbiol *71*, 4679-4689.

Beniash, E., Aizenberg, J., Addadi, L., and Weiner, S. (1997). Amorphous calcium carbonate transforms into calcite during sea urchin larval spicule growth. Proceedings of the Royal Society of London Series B: Biological Sciences *264*, 461-465.

Bennett, S. (2004). Solexa ltd. Pharmacogenomics *5*, 433-438.

Bennett, S., Barnes, C., Cox, A., Davies, L., and Brown, C. (2005). Toward the 1,000 dollar human genome. Pharmacogenomics *6*, 373-382.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hugler, M., Alber, B.E., and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. Nat Rev Microbiol *8*, 447-460.

Blaxter, M. (2003). Molecular systematics: Counting angels with DNA. Nature *421*, 122-124.

Blaxter, M., and Floyd, R. (2003). Molecular taxonomics for biodiversity surveys: already a reality. Trends in Ecology & Evolution *18*, 268-269.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., and Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society B: Biological Sciences *360*, 1935-1943.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I.*, et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res *31*, 365-370.

Borodovsky, M., and Mclninch, J. (1993). GENMARK: parallel gene recognition for both DNA strands. Computers & chemistry *17*, 123-133.

Borrel, G., Harris, H.M., Parisot, N., Gaci, N., Tottey, W., Mihajlovski, A., Deane, J., Gribaldo, S., Bardot, O., Peyretaillade, E.*, et al.* (2013). Genome Sequence of "*Candidatus* Methanomassiliicoccus intestinalis" Issoire-Mx1, a Third Thermoplasmatales-Related Methanogenic Archaeon from Human Feces. Genome Announc *1*, e00453-13.

Boulos, L., Prevost, M., Barbeau, B., Coallier, J., and Desjardins, R. (1999). LIVE/DEAD BacLight : application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. J Microbiol Methods *37*, 77-86.

Brady, A., and Salzberg, S.L. (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated

Markov models. Nat Methods *6*, 673-676.

Brescia, C.C., Griffin, S.M., Ware, M.W., Varughese, E.A., Egorov, A.I., and Villegas, E.N. (2009). Cryptosporidium propidium monoazide-PCR, a molecular biology-based technique for genotyping of viable Cryptosporidium oocysts. Appl Environ Microbiol *75*, 6856-6863.

Brochier, C., Gribaldo, S., Zivanovic, Y., Confalonieri, F., and Forterre, P. (2005). Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? Genome Biol *6*, R42.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008). Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. Nat Rev Microbiol *6*, 245-252.

Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. Curr Opin Microbiol *14*, 274-281.

Brodie, E.L., Desantis, T.Z., Joyner, D.C., Baek, S.M., Larsen, J.T., Andersen, G.L., Hazen, T.C., Richardson, P.M., Herman, D.J., Tokunaga, T.K.*, et al.* (2006). Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl Environ Microbiol *72*, 6288-6298.

Brodie, E.L., DeSantis, T.Z., Parker, J.P., Zubietta, I.X., Piceno, Y.M., and Andersen, G.L. (2007). Urban aerosols harbor diverse and dynamic bacterial populations. Proc Natl Acad Sci U S A *104*, 299-304.

Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., and Brom, T.H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. arXiv preprint arXiv:12034802.

Brunk, C.F., and Eis, N. (1998). Quantitative measure of small-subunit rRNA gene sequences of the kingdom korarchaeota. Appl Environ Microbiol *64*, 5064-5066.

Brusa, T., Canzi, E., Allievi, L., Del Puppo, E., and Ferrari, A. (1993). Methanogens in the human intestinal tract and oral cavity. Current Microbiology *27*, 261-265.

Buchenau, B.r., and Thauer, R.K. (2004). Tetrahydrofolate-specific enzymes in Methanosarcina barkeri and growth dependence of this methanogenic archaeon on folic acid or p-aminobenzoic acid. Arch Microbiol *182*, 313-325.

Buckel, W., and Thauer, R.K. (2013). Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. Biochim Biophys Acta *1827*, 94-113.

Budevska, B.O., Sum, S.T., and Jones, T.J. (2003). Application of multivariate curve resolution for analysis of FT-IR microspectroscopic images of in situ plant tissue. Appl Spectrosc *57*, 124-131.

Burggraf, S., Huber, H., and Stetter, K. (1997). Reclassification of the crenarchaeal orders and families in accordance with 16S rRNA sequence data. Int J Syst Bacteriol *47*, 657-660.

Burggraf, S., Olsen, G.J., Stetter, K.O., and Woese, C.R. (1992). A phylogenetic analysis of *Aquifex pyrophilus*. Syst Appl Microbiol *15*, 352-356.

Burggraf, S., Stetter, K., Rouviere, P., and Woese, C. (1991). *Methanopyrus kandleri*: An archaeal methanogen unrelated to all other known methanogens. Syst Appl Microbiol *14*, 346-351.

C

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I.*, et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods *7*, 335-336.

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N.*, et al.* (2011). Moving pictures of the human microbiome. Genome Biol *12*, R50.

Caporaso, J.G., Paszkiewicz, K., Field, D., Knight, R., and Gilbert, J.A. (2012). The Western English Channel contains a persistent microbial seed bank. ISME J *6*, 1089-1093.

Carr, G., Reffner, J., and Williams, G. (1995). Performance of an infrared microspectrometer at the NSLS. Review of Scientific Instruments *66*, 1490-1492.

Casamayor, E.O., Massana, R., Benlloch, S., Ovreas, L., Diez, B., Goddard, V.J., Gasol, J.M., Joint, I., Rodriguez-Valera, F., and Pedros-Alio, C. (2002). Changes in archaeal, bacterial and eukaryal assemblages along a salinity gradient by comparison of genetic fingerprinting methods in a multipond solar saltern. Environ Microbiol *4*, 338-348.

Castelle, C.J., Hug, L.A., Wrighton, K.C., Thomas, B.C., Williams, K.H., Wu, D., Tringe, S.G., Singer, S.W., Eisen, J.A., and

Banfield, J.F. (2013). Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. Nat Commun *4*, 2120.

Cawthorn, D.M., and Witthuhn, R.C. (2008). Selective PCR detection of viable *Enterobacter sakazakii* cells utilizing propidium monoazide or ethidium bromide monoazide. J Appl Microbiol *105*, 1178-1185.

Chapelle, F.H., and Bradley, P.M. (2007). Hydrologic significance of carbon monoxide concentrations in ground water. Ground Water *45*, 272-280.

Chapelle, F.H., O'Neill, K., Bradley, P.M., Methe, B.A., Ciufo, S.A., Knobel, L.L., and Lovley, D.R. (2002). A hydrogen-based subsurface microbial community dominated by methanogens. Nature *415*, 312-315.

Chevreux, B. (2005). MIRA: an automated genome and EST assembler. Ruprecht-Karls University, Heidelberg, Germany.

Chistoserdova, L., Vorholt, J.A., Thauer, R.K., and Lidstrom, M.E. (1998). C1 transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic Archaea. Science *281*, 99-102.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M.*, et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res *37*, D141-145.

Colson, P., de Lamballerie, X., Fournous, G., and Raoult, D. (2012). Reclassification of giant viruses composing a fourth domain of life in the new order Megavirales. Intervirology *55*, 321-332.

Colson, P., Gimenez, G., Boyer, M., Fournous, G., and Raoult, D. (2011). The giant *Cafeteria roenbergensis* virus that infects a widespread marine phagocytic protist is a new member of the fourth domain of Life. PLoS One *6*, e18935.

Colwell, R.R. (1997). Microbial diversity: the importance of exploration and conservation. J Ind Microbiol Biotechnol *18*, 302-307.

Comolli, L.R., Baker, B.J., Downing, K.H., Siegerist, C.E., and Banfield, J.F. (2009). Three-dimensional analysis of the structure and ecology of a novel, ultra-small archaeon. ISME J *3*, 159-167.

Conley, C.A., and Rummel, J.D. (2013). Appropriate protection of Mars. Nature Geoscience *6*, 587-588.

Contursi, P., Cannio, R., Prato, S., Fiorentino, G., Rossi, M., and Bartolucci, S. (2003). Development of a genetic system for hyperthermophilic Archaea: expression of a moderate thermophilic bacterial alcohol dehydrogenase gene in *Sulfolobus solfataricus*. FEMS microbiology letters *218*, 115-120.

Conway de Macario, E., and Macario, A.J. (2009). Methanogenic archaea in health and disease: a novel paradigm of microbial pathogenesis. Int J Med Microbiol *299*, 99-108.

Cooper, M., La Duc, M.T., Probst, A., Vaishampayan, P., Stam, C., Benardini, J.N., Piceno, Y.M., Andersen, G.L., and Venkateswaran, K. (2011). Comparison of innovative molecular approaches and standard spore assays for assessment of surface cleanliness. Appl Environ Microbiol *77*, 5438-5444.

COSPAR (2002). COSPAR planetary protection policy. Commitee on Space Research, Huston, TX, USA.

Crick, F.H., and Orgel, L.E. (1973). Directed panspermia. Icarus *19*, 341-346.

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*, 210.

D

DasSarma, P., and DasSarma, S. (2008). On the origin of prokaryotic" species": the taxonomy of halophilic Archaea. Saline Systems *4*.

De Rosa, M., and Gambacorta, A. (1988). The lipids of archaebacteria. Prog Lipid Res *27*, 153-175.

DeAngelis, K.M., Wu, C.H., Beller, H.R., Brodie, E.L., Chakraborty, R., DeSantis, T.Z., Fortney, J.L., Hazen, T.C., Osman, S.R., Singer, M.E.*, et al.* (2011). PCR amplification-independent methods for detection of microbial communities by the high-density microarray PhyloChip. Appl Environ Microbiol *77*, 6313-6322.

Dekas, A.E., Poretsky, R.S., and Orphan, V.J. (2009). Deep-sea archaea fix and share nitrogen in methane-consuming microbial consortia. Science *326*, 422-426.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. (1999). Improved microbial gene identification with GLIMMER. Nucleic Acids Res *27*, 4636-4641.

DeLong, E.F. (1992). Archaea in coastal marine environments. Proc Natl Acad Sci U S A *89*, 5685-5689.

DeLong, E.F. (1998). Everything in moderation: archaea as 'non-extremophiles'. Curr Opin Genet Dev *8*, 649-654.

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nat Rev Genet *6*, 361-375.

Den Blaauwen, T., Buddelmeijer, N., Aarsman, M.E., Hameete, C.M., and Nanninga, N. (1999). Timing of FtsZ assembly in *Escherichia coli*. J Bacteriol *181*, 5167-5175.

DeSantis, T.Z., Brodie, E.L., Moberg, J.P., Zubieta, I.X., Piceno, Y.M., and Andersen, G.L. (2007). High-density universal 16S rRNA microarray analysis reveals broader diversity than typical clone library when sampling the environment. Microb Ecol *53*, 371-383.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol *72*, 5069-5072.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F*., et al.* (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. Elife *2*, e01102.

Direito, S.O., Zaura, E., Little, M., Ehrenfreund, P., and Roling, W.F. (2014). Systematic evaluation of bias in microbial community profiles induced by whole genome amplification. Environ Microbiol *16*, 643-657.

Dowd, S.E., Sun, Y., Secor, P.R., Rhoads, D.D., Wolcott, B.M., James, G.A., and Wolcott, R.D. (2008). Survey of bacterial diversity in chronic wounds using pyrosequencing, DGGE, and full ribosome shotgun sequencing. BMC Microbiol *8*, 43.

Dumas, P., Miller, L., and Tobin, M. (2009). Challenges in biology and medicine with synchrotron infrared light. Acta Physica Polonica-Series A General Physics *115*, 446.

E

Eddy, S.R. (2009). A new generation of homology search tools based on probabilistic inference. Genome Inform *23*, 205-211.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research *32*, 1792-1797.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461.

Edgar, R.C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Meth *10*, 996-998.

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*., et al.* (2008). A korarchaeal genome reveals insights into the evolution of the Archaea. Proc Natl Acad Sci U S A *105*, 8102-8107.

Elvert, M., Suess, E., Greinert, J., and Whiticar, M.J. (2000). Archaea mediating anaerobic methane oxidation in deep-sea sediments at cold seeps of the eastern Aleutian subduction zone. Organic Geochemistry *31*, 1175-1187.

Embley, T.M., Finlay, B.J., Thomas, R.H., and Dyal, P.L. (1992). The use of rRNA sequences and fluorescent probes to investigate the phylogenetic positions of the anaerobic ciliate *Metopus palaeformis* and its archaeobacterial endosymbiont. Journal of General Microbiology *138*, 1479-1487.

Emerson, J., Thomas, B., Alvarez, W., and Banfield, J. (submitted). Metagenomic reconstruction of a CO_2-driven geyser microbial community dominated by iron- and sulfur-oxidizing bacteria and novel microorganisms. Environ Microbiol.

Engel, A.S., Lee, N., Porter, M.L., Stern, L.A., Bennett, P.C., and Wagner, M. (2003). Filamentous "Epsilonproteobacteria" dominate microbial mats from sulfidic cave springs. Appl Environ Microbiol *69*, 5503-5511.

Engel, A.S., Porter, M.L., Stern, L.A., Quinlan, S., and Bennett, P.C. (2004). Bacterial diversity and ecosystem function of filamentous microbial mats from aphotic (cave) sulfidic springs dominated by chemolithoautotrophic "Epsilonproteobacteria". FEMS Microbiol Ecol *51*, 31-53.

Erkel, C., Kube, M., Reinhardt, R., and Liesack, W. (2006). Genome of Rice Cluster I archaea - the key methane producers in the rice rhizosphere. Science *313*, 370-372.

Euzeby, J.P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. Int J Syst Bacteriol *47*, 590-592.

F

Fairén, A.G., and Schulze-Makuch, D. (2013). The overprotection of Mars. Nature Geosci *6*, 510-511.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. Systematic Biology *27*, 401-410.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. Nucleic Acids Res *39*, W29-37.

Fittipaldi, M., Codony, F., Adrados, B., Camper, A.K., and Morato, J. (2011). Viable real-time PCR in environmental samples: can all data be interpreted directly? Microb Ecol *61*, 7-12.

Flores, G.E., Bates, S.T., Knights, D., Lauber, C.L., Stombaugh, J., Knight, R., and Fierer, N. (2011). Microbial biogeography of public restroom surfaces. PLoS One *6*, e28132.

Fox, G., Stackebrandt, E., Hespell, R., Gibson, J., Maniloff, J., Dyer, T., Wolfe, R., Balch, W., Tanner, R., and Magrum, L. (1980). The phylogeny of prokaryotes. Science *209*, 457.

Fox, G.E. (2013). Carl R. Woese, 1928-2012. Astrobiology *13*, 1201-1202.

Fox, G.E., Pechman, K.R., and Woese, C.R. (1977). Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. Int J Syst Bacteriol *27*, 44-57.

Frickey, T., and Lupas, A.N. (2004). PhyloGenie: automated phylome generation and analysis. Nucleic Acids Res *32*, 5231-5238.

Fröls, S. (2013). Archaeal biofilms: widespread and complex. Biochem Soc Trans *41*, 393-398.

Frols, S., Ajon, M., Wagner, M., Teichmann, D., Zolghadr, B., Folea, M., Boekema, E.J., Driessen, A.J., Schleper, C., and Albers, S.V. (2008). UV-inducible cellular aggregation of the hyperthermophilic archaeon *Sulfolobus solfataricus* is mediated by pili formation. Mol Microbiol *70*, 938-952.

Fuchs, G. (2011). Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? Annu Rev Microbiol *65*, 631-658.

Fuhrman, J.A., McCallum, K., and Davis, A.A. (1992). Novel major archaebacterial group from marine plankton. Nature *356*, 148-149.

G

Gantner, S., Andersson, A.F., Alonso-Saez, L., and Bertilsson, S. (2011). Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. J Microbiol Methods *84*, 12-18.

Geets, J., Borremans, B., Diels, L., Springael, D., Vangronsveld, J., van der Lelie, D., and Vanbroekhoven, K. (2006). DsrB gene-based DGGE for community and diversity surveys of sulfate-reducing bacteria. J Microbiol Methods *66*, 194-205.

Ghosh, S., Osman, S., Vaishampayan, P., and Venkateswaran, K. (2010). Recurrent isolation of extremotolerant bacteria from the clean room where Phoenix spacecraft components were assembled. Astrobiology *10*, 325-335.

Gilbert, J.A., Meyer, F., Antonopoulos, D., Balaji, P., Brown, C.T., Brown, C.T., Desai, N., Eisen, J.A., Evers, D., Field, D.*, et al.* (2010a). Meeting report: the terabase metagenomics workshop and the vision of an Earth microbiome project. Stand Genomic Sci *3*, 243-248.

Gilbert, J.A., Meyer, F., Jansson, J., Gordon, J., Pace, N., Tiedje, J., Ley, R., Fierer, N., Field, D., Kyrpides, N.*, et al.* (2010b). The Earth Microbiome Project: Meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6 2010. Stand Genomic Sci *3*, 249-253.

Gilles, A., Meglécz, E., Pech, N., Ferreira, S., Malausa, T., and Martin, J.-F. (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics *12*, 245.

Gonzalez, N.H., Felsner, G., Schramm, F.D., Klingl, A., Maier, U.G., and Bolte, K. (2011). A single peroxisomal targeting signal mediates matrix protein import in diatoms. PLoS One *6*, e25316.

Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika *40*, 237-264.

Goren, M.B. (1970). Sulfolipid I of *Mycobacterium tuberculosis*, strain H37Rv. II. Structural studies. Biochim Biophys Acta *210*, 127-138.

Gribaldo, S., and Philippe, H. (2002). Ancient phylogenetic relationships. Theor Popul Biol *61*, 391-408.

Gribaldo, S., Poole, A.M., Daubin, V., Forterre, P., and Brochier-Armanet, C. (2010). The origin of eukaryotes and their relationship with the Archaea: are we at a phylogenomic impasse? Nat Rev Microbiol *8*, 743-752.

Grice, E.A., and Segre, J.A. (2011). The skin microbiome. Nat Rev Microbiol *9*, 244-253.

Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res *33*, D121-124.

Guo, H., and Xiong, J. (2006). A specific and versatile genome walking technique. Gene *381*, 18-23.

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

Guy, L., Spang, A., Saw, J.H., and Ettema, T.J.G. (2014). 'Geoarchaeote NAG1' is a deeply rooting lineage of the archaeal order Thermoproteales rather than a new phylum. ISME J.

H

Hall, E.K., Singer, G.A., Polzl, M., Hammerle, I., Schwarz, C., Daims, H., Maixner, F., and Battin, T.J. (2011). Looking inside the box: using Raman microspectroscopy to deconstruct microbial biomass stoichiometry one cell at a time. ISME J *5*, 196-208.

Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. Science *305*, 1457-1462.

Hamady, M., Walker, J.J., Harris, J.K., Gold, N.J., and Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nature Meth *5*, 235-237.

Harold, R., and Stanier, R.Y. (1955). The genera Leucothrix and Thiothrix. Bacteriol Rev *19*, 49-64.

Haroon, M.F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., Yuan, Z., and Tyson, G.W. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. Nature *500*, 567-570.

Harris, J.K., Kelley, S.T., Spiegelman, G.B., and Pace, N.R. (2003). The genetic core of the universal ancestor. Genome Res *13*, 407-412.

Harrison, F., and Kennedy, M. (1922). The red discoloration of cured codfish. Trans R SOC Can Sect *III*, 101-152.

Hatzenpichler, R., Lebedeva, E.V., Spieck, E., Stoecker, K., Richter, A., Daims, H., and Wagner, M. (2008). A moderately thermophilic ammonia-oxidizing crenarchaeote from a hot spring. Proc Natl Acad Sci U S A *105*, 2134-2139.

Hazen, T.C., Dubinsky, E.A., DeSantis, T.Z., Andersen, G.L., Piceno, Y.M., Singh, N., Jansson, J.K., Probst, A., Borglin, S.E., Fortney, J.L.*, et al.* (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science *330*, 204-208.

Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M.*, et al.* (2004). The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol *22*, 554-559.

Hein, I., Flekna, G., Wagner, M., Nocker, A., and Camper, A.K. (2006). Possible errors in the interpretation of ethidium bromide and PicoGreen DNA staining results from ethidium monoazide-treated DNA. Appl Environ Microbiol *72*, 6860-6861; author reply 6861-6862.

Henneberger, R., Moissl, C., Amann, T., Rudolph, C., and Huber, R. (2006). New insights into the lifestyle of the cold-loving SM1 euryarchaeon: natural growth as a monospecies biofilm in the subsurface. Appl Environ Microbiol *72*, 192-199.

Hess, V., Schuchmann, K., and Müller, V. (2013). The ferredoxin: NAD+ oxidoreductase (Rnf) from the acetogen Acetobacterium woodii requires Na^+ and is reversibly coupled to the membrane potential. Journal of Biological Chemistry *288*, 31496-31502.

Hierro, N., Esteve-Zarzoso, B., Gonzalez, A., Mas, A., and Guillamon, J.M. (2006). Real-time quantitative PCR (QPCR) and reverse transcription-QPCR for detection and enumeration of total yeasts in wine. Appl Environ Microbiol *72*, 7148-7155.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G.D., Lewis, J.D., and Bushman, F.D. (2013). Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. PLoS One *8*, e66019.

Holman, H.Y., Bechtel, H.A., Hao, Z., and Martin, M.C. (2010). Synchrotron IR Spectromicroscopy: Chemistry of Living Cells.

Anal Chem *82*, 8757-8765.

Holman, H.Y., Wozei, E., Lin, Z., Comolli, L.R., Ball, D.A., Borglin, S., Fields, M.W., Hazen, T.C., and Downing, K.H. (2009). Real-time molecular monitoring of chemical environment in obligate anaerobes during oxygen adaptive response. Proc Natl Acad Sci U S A *106*, 12599-12604.

Huang, C., and Kerr, P. (1960). Infrared study of the carbonated minerals. The American Mineralogist *45*, 311-324.

Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., and Wagner, M. (2007). Raman-FISH: combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ Microbiol *9*, 1878-1889.

Hube, M., Blokesch, M., and Bock, A. (2002). Network of hydrogenase maturation in *Escherichia coli*: role of accessory proteins HypA and HybF. J Bacteriol *184*, 3879-3885.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C., and Stetter, K.O. (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature *417*, 63-67.

Huber, H., and Stetter, K.O. (2006). Thermoplasmatales. The Prokaryotes: Volume 3: Archaea Bacteria: Firmicutes, Actinomycetes, 101-112.

Hulcr, J., Latimer, A.M., Henley, J.B., Rountree, N.R., Fierer, N., Lucky, A., Lowman, M.D., and Dunn, R.R. (2012). A jungle in there: bacteria in belly buttons are highly diverse, but predictable. PLoS One *7*, e47712.

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207-214.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol *8*, R143.

Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res *21*, 1552-1560.

Hutchison, C.A., 3rd (2007). DNA sequencing: bench to bedside and beyond. Nucleic Acids Res *35*, 6227-6237.

Hutchison, C.A., and Venter, J.C. (2006). Single-cell genomics. Nature Biotech *24*, 657-658.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinf *11*, 119.

J

Jackson, S.A., Flemer, B., McCann, A., Kennedy, J., Morrissey, J.P., O'Gara, F., and Dobson, A.D. (2013). Archaea appear to dominate the microbiome of *Inflatella pellicula* deep sea sponges. PLoS One *8*, e84438.

Jensen, M.A., Webster, J.A., and Straus, N. (1993). Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms. Appl Environ Microbiol *59*, 945-952.

K

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

Kantor, R.S., Wrighton, K.C., Handley, K.M., Sharon, I., Hug, L.A., Castelle, C.J., Thomas, B.C., and Banfield, J.F. (2013). Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio *4*, e00708-00713.

Kaprelyants, A.S., Gottschal, J.C., and Kell, D.B. (1993). Dormancy in non-sporulating bacteria. FEMS Microbiol Rev *10*, 271-285.

Karner, M.B., DeLong, E.F., and Karl, D.M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature *409*, 507-510.

Kato, N., Yurimoto, H., and Thauer, R.K. (2006). The physiological role of the ribulose monophosphate pathway in bacteria and archaea. Biosci Biotechnol Biochem *70*, 10-21.

Keer, J.T., and Birch, L. (2003). Molecular methods for the assessment of bacterial viability. J Microbiol Methods *53*, 175-183.

Kircher, M., Stenzel, U., and Kelso, J. (2009). Improved base calling for the Illumina Genome Analyzer using machine learning strategies. Genome Biol *10*, R83.

Klappenbach, J.A., Saxman, P.R., Cole, J.R., and Schmidt, T.M. (2001). rrndb: the ribosomal RNA operon copy number

database. Nucleic Acids Res *29*, 181-184.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glockner, F.O. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res *41*, e1.

Klingl, A. (2011). Struktur und Funktion von S-layern acidophiler Bakterien und Archaeen, ihre Rolle bei der Pyrit-Oxidation sowie die Adhäsion an Oberflächen. In Department for Microbiology and Archaea Center (Regensburg: University of Regensburg).

Kluyver, A., and Van Niel, C. (1936). Prospects for a natural system of classification of bacteria. Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene Abteilung II *34*, 369-403.

Koch, M., Rudolph, C., Moissl, C., and Huber, R. (2006). A cold-loving crenarchaeon is a substantial part of a novel microbial community in cold sulphidic marsh water. FEMS Microbiol Ecol *57*, 55-66.

Kodama, Y., and Watanabe, K. (2004). *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. Int J Syst Evol Microbiol *54*, 2297-2300.

Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993). Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. Microbiol Rev *57*, 164-182.

Konneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature *437*, 543-546.

Kort, R., Nocker, A., de Kat Angelino-Bart, A., van Veen, S., Verheij, H., Schuren, F., and Montijn, R. (2010). Real-time detection of viable microorganisms by intracellular phototautomerism. BMC Biotechnol *10*, 45.

Kozubal, M.A., Romine, M., Jennings, R., Jay, Z.J., Tringe, S.G., Rusch, D.B., Beam, J.P., McCue, L.A., and Inskeep, W.P. (2013). Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. ISME J *7*, 622-634.

Kristjansson, J.K., and Stetter, K.O. (1992). Thermophilic bacteria (London: CRC Press, Inc.).

Kuczynski, J., Liu, Z., Lozupone, C., McDonald, D., Fierer, N., and Knight, R. (2010). Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. Nat Methods *7*, 813-819.

Kuhn, I., Peng, L., Bedu, S., and Zhang, C.C. (2000). Developmental regulation of the cell division protein FtsZ in *Anabaena* sp. strain PCC 7120, a cyanobacterium capable of terminal differentiation. J Bacteriol *182*, 4640-4643.

Kwan, K., Cooper, M., La Duc, M.T., Vaishampayan, P., Stam, C., Benardini, J.N., Scalzi, G., Moissl-Eichinger, C., and Venkateswaran, K. (2011). Evaluation of procedures for the collection, processing, and analysis of biomolecules from low-biomass surfaces. Appl Environ Microbiol *77*, 2943-2953.

L

La Duc, M.T., Dekas, A., Osman, S., Moissl, C., Newcombe, D., and Venkateswaran, K. (2007). Isolation and characterization of bacteria capable of tolerating the extreme conditions of clean room environments. Appl Environ Microbiol *73*, 2600-2611.

La Duc, M.T., Kern, R., and Venkateswaran, K. (2004). Microbial monitoring of spacecraft and associated environments. Microb Ecol *47*, 150-158.

La Duc, M.T., Nicholson, W., Kern, R., and Venkateswaran, K. (2003). Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility. Environ Microbiol *5*, 977-985.

La Duc, M.T., Osman, S., Vaishampayan, P., Piceno, Y., Andersen, G., Spry, J.A., and Venkateswaran, K. (2009). Comprehensive census of bacteria in clean rooms by using DNA microarray and cloning methods. Appl Environ Microbiol *75*, 6559-6567.

La Duc, M.T., Vaishampayan, P., Nilsson, H.R., Torok, T., and Venkateswaran, K. (2012). Pyrosequencing-derived bacterial, archaeal, and fungal diversity of spacecraft hardware destined for Mars. Appl Environ Microbiol *78*, 5912-5922.

Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res *35*, 3100-3108.

Lake, J.A., Henderson, E., Oakes, M., and Clark, M.W. (1984). Eocytes: a new ribosome structure indicates a kingdom with

a close relationship to eukaryotes. Proc Natl Acad Sci U S A *81*, 3786-3790.

Lake, J.A., and Rivera, M.C. (2004). Deriving the genomic tree of life in the presence of horizontal gene transfer: conditioned reconstruction. Mol Biol Evol *21*, 681-690.

Lamcke, K. (1976). Übertiefe Grundgewässer im Süddeutschen Alpenvorland. BuU Ver Schweiz Petroleum-Geol u -Ing *42*, 9-18.

Lane, D.J. (1991). 16S/23S rRNA sequencing. Nucleic acid techniques in bacterial systematics (Chichester: John Wiley & Sons).

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc Natl Acad Sci U S A *82*, 6955-6959.

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R.*, et al.* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol *31*, 814-821.

Lapaglia, C., and Hartzell, P.L. (1997). Stress-Induced Production of Biofilm in the Hyperthermophile *Archaeoglobus fulgidus*. Appl Environ Microbiol *63*, 3158-3163.

Larkin, J.M. (1980). Isolation of *Thiothrix* in pure culture and observation of a filamentous epiphyte on *Thiothrix*. Curr Microbiol *4*, 155-158.

Larkin, J.M., and Shinabarger, D.L. (1983). Characterization of *Thiothrix nivea*. Int J Syst Bacteriol *33*, 841-846.

Larkin, J.M., and Strohl, W.R. (1983). *Beggiatoa*, *Thiothrix*, and *Thioploca*. Annu Rev Microbiol *37*, 341-367.

Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol *7 Suppl 1*, S4.

Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286-2288.

Lasken, R.S. (2007). Single-cell genomic sequencing using multiple displacement amplification. Current opinion in microbiology *10*, 510-516.

Ledoux, R.L., and White, J.L. (1964). Infrared study of the OH groups in expanded kaolinite. Science *143*, 244-245.

Lee, Z.M., Bussema, C., 3rd, and Schmidt, T.M. (2009). rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. Nucleic Acids Res *37*, D489-493.

Legendre, M., Arslan, D., Abergel, C., and Claverie, J.M. (2012). Genomics of Megavirus and the elusive fourth domain of Life. Commun Integr Biol *5*, 102-106.

Lehtovirta-Morley, L.E., Stoecker, K., Vilcinskas, A., Prosser, J.I., and Nicol, G.W. (2011). Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. Proc Natl Acad Sci U S A *108*, 15892-15897.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G., Prosser, J., Schuster, S., and Schleper, C. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature *442*, 806-809.

Lepp, P.W., Brinig, M.M., Ouverney, C.C., Palm, K., Armitage, G.C., and Relman, D.A. (2004). Methanogenic Archaea and human periodontal disease. Proc Natl Acad Sci U S A *101*, 6176-6181.

Letunic, I., and Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinf *23*, 127-128.

Li, M., Canniffe, D.P., Jackson, P.J., Davison, P.A., FitzGerald, S., Dickman, M.J., Burgess, J.G., Hunter, C.N., and Huang, W.E. (2012). Rapid resonance Raman microspectroscopy to probe carbon dioxide fixation by single cells in microbial communities. ISME J *6*, 875-885.

Liamleam, W., and Annachhatre, A.P. (2007). Electron donors for biological sulfate reduction. Biotechnol Adv *25*, 452-463.

Lin, W.T., Luo, J.F., and Guo, Y. (2011). Comparison and characterization of microbial communities in sulfide-rich wastewater with and without propidium monoazide treatment. Curr Microbiol *62*, 374-381.

Lovley, D.R., and Chapelle, F.H. (1995). Deep subsurface microbial processes. Rev Geophys *33*, 365-381.

Lovley, D.R., and Coates, J.D. (2000). Novel forms of anaerobic respiration of environmental relevance. Curr Opin Microbiol *3*, 252-256.

Lovley, D.R., and Klug, M.J. (1983). Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. Appl Environ Microbiol *45*, 187-192.

Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res *25*, 955-964.

Loy, A., Lehner, A., Lee, N., Adamczyk, J., Meier, H., Ernst, J., Schleifer, K.H., and Wagner, M. (2002). Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. Appl Environ Microbiol *68*, 5064-5081.

Loy, A., Schulz, C., Lucker, S., Schopfer-Wendels, A., Stoecker, K., Baranyi, C., Lehner, A., and Wagner, M. (2005). 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "Rhodocyclales". Appl Environ Microbiol *71*, 1373-1386.

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol *71*, 8228-8235.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: an effective distance metric for microbial community comparison. ISME J *5*, 169-172.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G.*, et al.* (2004). ARB: a software environment for sequence data. Nucleic Acids Res *32*, 1363-1371.

M

Maden, B.E. (2000). Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C1 metabolism. Biochem J *350 Pt 3*, 609-629.

Makarova, K.S., Yutin, N., Bell, S.D., and Koonin, E.V. (2010). Evolution of diverse cell division and vesicle formation systems in Archaea. Nat Rev Microbiol *8*, 731-741.

Mancuso, C.A., Nichols, P.D., and White, D.C. (1986). A method for the separation and characterization of archaebacterial signature ether lipids. J Lipid Res *27*, 49-56.

Manders, E., Verbeek, F., and Aten, J. (1993). Measurement of co-localization of objects in dual-colour confocal images. J Microscopy *169*, 375-382.

Mantsch, H.H., and Chapman, D. (1996). Infrared spectroscopy of biomolecules (New York: Wiley-Liss, Inc).

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z.*, et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.

Mayers, G.L., Pousada, M., and Haines, T.H. (1969). Microbial sulfolipids. 3. The disulfate of (+)-1,14-docosanediol in *Ochromonas danica*. Biochem *8*, 2981-2986.

McCarthy, A. (2010). Third generation DNA sequencing: pacific biosciences' single molecule real time technology. Chemistry & biology *17*, 675-676.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J *6*, 610-618.

Meador, T.B., Gagen, E.J., Loscar, M.E., Goldhammer, T., Yoshinaga, M.Y., Wendt, J., Thomm, M., and 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.

Melosh, H. (1988). The rocky road to panspermia. Nature *332*, 687.

Mende, D.R., Waller, A.S., Sunagawa, S., Jarvelin, A.I., Chan, M.M., Arumugam, M., Raes, J., and Bork, P. (2012). Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One *7*, e31386.

Mendes, R., Kruijt, M., de Bruijn, I., Dekkers, E., van der Voort, M., Schneider, J.H., Piceno, Y.M., DeSantis, T.Z., Andersen, G.L., Bakker, P.A.*, et al.* (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science *332*, 1097-1100.

Metcalf, W.W., Zhang, J.K., Apolinario, E., Sowers, K.R., and Wolfe, R.S. (1997). A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. Proc Natl Acad Sci U S A *94*, 2626-2631.

Meyer, F., Overbeek, R., and Rodriguez, A. (2009). FIGfams: yet another set of protein families. Nucleic Acids Res *37*, 6643-

6654.

Meyerdierks, A., Kube, M., Kostadinov, I., Teeling, H., Glockner, F.O., Reinhardt, R., and Amann, R. (2010). Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. Environ Microbiol *12*, 422-439.

Miller, C.S., Baker, B.J., Thomas, B.C., Singer, S.W., and Banfield, J.F. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome Biol *12*, R44.

Milucka, J., Ferdelman, T.G., Polerecky, L., Franzke, D., Wegener, G., Schmid, M., Lieberwirth, I., Wagner, M., Widdel, F., and Kuypers, M.M. (2012). Zero-valent sulphur is a key intermediate in marine methane oxidation. Nature *491*, 541-546.

Mohsen, M., Mohamed, S.F., Ali, F., and El-Sayed, O.H. (2007). Chemical structure and antiviral activity of water-soluble sulfated polysaccharides from *Sargassum latifolium*. J Appl Sci Res *3*, 1178-1185.

Moissl, C. (2004). Molekularbiologische und strukturelle Untersuchungen zur Biologie des neuartigen, kälteliebenden SM1 Euryarchaeons und seiner verschiedenen Lebensgemeinschaften. In Department for Microbiology and Archaea Center (Regensburg: University of Regensburg).

Moissl, C., Briegel, A., Engelhardt, H., and Huber, R. (2005a). Enterhaken und Stacheldraht: Verbluffende Strukturen aus der archaeellen Nano-Welt. Biospektrum, Heidelberg *11*, 732.

Moissl, C., Bruckner, J.C., and Venkateswaran, K. (2008). Archaeal diversity analysis of spacecraft assembly clean rooms. ISME J *2*, 115-119.

Moissl, C., Osman, S., La Duc, M.T., Dekas, A., Brodie, E., DeSantis, T., and Venkateswaran, K. (2007). Molecular bacterial community analysis of clean rooms where spacecraft are assembled. FEMS Microbiol Ecol *61*, 509-521.

Moissl, C., Rachel, R., Briegel, A., Engelhardt, H., and Huber, R. (2005b). The unique structure of archaeal '*hami*', highly complex cell appendages with nano-grappling hooks. Mol Microbiol *56*, 361-370.

Moissl, C., Rudolph, C., and Huber, R. (2002). Natural communities of novel archaea and bacteria with a string-of-pearls-like morphology: molecular analysis of the bacterial partners. Appl Environ Microbiol *68*, 933-937.

Moissl, C., Rudolph, C., Rachel, R., Koch, M., and Huber, R. (2003). In situ growth of the novel SM1 euryarchaeon from a string-of-pearls-like microbial community in its cold biotope, its physical separation and insights into its structure and physiology. Arch Microbiol *180*, 211-217.

Moissl-Eichinger, C. (2011). Archaea in artificial environments: their presence in global spacecraft clean rooms and impact on planetary protection. ISME J *5*, 209-219.

Moissl-Eichinger, C., Henneberger, R., and Huber, R. (2012a). SM1: A cold-loving archaeon with powerful nano-grappling hooks. In Extremophiles: Microbiology and Biotechnology, 77 (Norfolk: Caister Academic Press).

Moissl-Eichinger, C., and Huber, H. (2011). Archaeal symbionts and parasites. Curr Opin Microbiol *14*, 364-370.

Moissl-Eichinger, C., Pukall, R., Probst, A.J., Stieglmeier, M., Schwendner, P., Mora, M., Barczyk, S., Bohmeier, M., and Rettberg, P. (2013). Lessons learned from the microbial analysis of the Herschel spacecraft during assembly, integration, and test operations. Astrobiol *13*, 1125-1139.

Moissl-Eichinger, C., Rettberg, P., and Pukall, R. (2012b). The first collection of spacecraft-associated microorganisms: a public source for extremotolerant microorganisms from spacecraft assembly clean rooms. Astrobiology *12*, 1024-1034.

Mondav, R., Woodcroft, B.J., Kim, E.H., McCalley, C.K., Hodgkins, S.B., Crill, P.M., Chanton, J., Hurst, G.B., Verberkmoes, N.C., Saleska, S.R.*, et al.* (2014). Discovery of a novel methanogen prevalent in thawing permafrost. Nat Commun *5*, 3212.

Morris, B.E., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. FEMS Microbiol Rev *37*, 384-406.

Morrison, D. (2001). The NASA astrobiology program. Astrobiology *1*, 3-13.

Mühl, H., Kochem, A.J., Disque, C., and Sakka, S.G. (2010). Activity and DNA contamination of commercial polymerase chain reaction reagents for the universal 16S rDNA real-time polymerase chain reaction detection of bacterial pathogens in blood. Diagn Microbiol Infect Dis *66*, 41-49.

Mulder, N., and Apweiler, R. (2007). InterPro and InterProScan. In Comparative genomics (Springer), pp. 59-70.

Mussmann, M., Brito, I., Pitcher, A., Sinninghe Damste, J.S., Hatzenpichler, R., Richter, A., Nielsen, J.L., Nielsen, P.H., Muller, A., Daims, H.*, et al.* (2011). Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. Proc Natl Acad Sci U S A *108*, 16771-16776.

N

Nagpal, M.L., Fox, K.F., and Fox, A. (1998). Utility of 16S-23S rRNA spacer region methodology: how similar are interspace regions within a genome and between strains for closely related organisms? J Microbiol Meth *33*, 211-219.

Nancy, Y.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S.b., Lo, R., Dao, P., Sahinalp, S.C., Ester, M., and Foster, L.J. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics *26*, 1608-1615.

Naumann, D. (2000). Infrared spectroscopy in microbiology (Chichester: John Wiley & Sons, Ltd.).

Neufeld, J.D., Wagner, M., and Murrell, J.C. (2007). Who eats what, where and when? Isotope-labelling experiments are coming of age. ISME J *1*, 103-110.

Nielsen, H. (1981). Schwefelisotope und ihre Aussage zur Entstehung der Schwefelquellen. Schriftenreihe Bayer Landesamt f Wasserwirtschaft *15*, 99-107.

Nocker, A., Cheung, C.Y., and Camper, A.K. (2006). Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. J Microbiol Methods *67*, 310-320.

Nocker, A., Mazza, A., Masson, L., Camper, A.K., and Brousseau, R. (2009). Selective detection of live bacteria combining propidium monoazide sample treatment with microarray technology. J Microbiol Meth *76*, 253-261.

Nocker, A., Richter-Heitmann, T., Montijn, R., Schuren, F., and Kort, R. (2010). Discrimination between live and dead cells in bacterial communities from environmental water samples analyzed by 454 pyrosequencing. Int Microbiol *13*, 59-65.

Nocker, A., Sossa-Fernandez, P., Burr, M.D., and Camper, A.K. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol *73*, 5111-5117.

Nose, K., Mizuno, T., Yamane, N., Kondo, T., Ohtani, H., Araki, S., and Tsuda, T. (2005). Identification of ammonia in gas emanated from human skin and its correlation with that in blood. Anal Sci *21*, 1471-1474.

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., Chee, G.-J., Hattori, M., Kanai, A., and Atomi, H. (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. Nucleic Acids Res *39*, 3204-3223.

Nyrén, P., Pettersson, B., and Uhlen, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. Anal Biochem *208*, 171-175.

O

Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. Annu Rev Microbiol *67*, 437-457.

Olson, J.W., Mehta, N.S., and Maier, R.J. (2001). Requirement of nickel metabolism proteins HypA and HypB for full activity of both hydrogenase and urease in *Helicobacter pylori*. Mol Microbiol *39*, 176-182.

Onstott, T., Colwell, F., Kieft, T., Murdoch, L., and Phelps, T. (2010). New horizons for deep subsurface microbiology. Issues.

Orell, A., Fröls, S., and Albers, S.V. (2013). Archaeal biofilms: the great unexplored. Annu Rev Microbiol *67*, 337-354.

Oren, A. (2008). Nomenclature and taxonomy of halophilic archaea — comments on the proposal by DasSarma and DasSarma for nomenclatural changes within the order Halobacteriales. Int J Syst Evol Microbiol *58*, 2245-2246.

Orphan, V.J., House, C.H., Hinrichs, K.U., McKeegan, K.D., and DeLong, E.F. (2001). Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. Science *293*, 484-487.

Ortiz, M., Legatzki, A., Neilson, J.W., Fryslie, B., Nelson, W.M., Wing, R.A., Soderlund, C.A., Pryor, B.M., and Maier, R.M. (2013). Making a living while starving in the dark: metagenomic insights into the energy dynamics of a carbonate cave. ISME J.

Osman, S., Peeters, Z., La Duc, M.T., Mancinelli, R., Ehrenfreund, P., and Venkateswaran, K. (2008). Effect of shadowing on survival of bacteria under conditions simulating the Martian atmosphere and UV radiation. Appl Environ Microbiol *74*, 959-970.

Oxley, A.P., Lanfranconi, M.P., Wurdemann, D., Ott, S., Schreiber, S., McGenity, T.J., Timmis, K.N., and Nogales, B. (2010).

Halophilic archaea in the human intestinal mucosa. Environ Microbiol *12*, 2398-2410.

P

Pace, N., Stahl, D., Lane, D., and Olsen, G. (1986). The analysis of natural microbial populations by ribosomal RNA sequences. Adv Microb Ecol *9*, 1-55.

Pace, N.R. (1997). A molecular view of microbial diversity and the biosphere. Science *276*, 734-740.

Palmer, A.N. (1991). Origin and morphology of limestone caves. Geol Soc Am Bull *103*, 1-21.

Parker, F. (1983). Applications of infrared, raman and resonance raman spectroscopy in biochemistry (New York: Plenum Press).

Parshionikar, S., Laseke, I., and Fout, G.S. (2010). Use of propidium monoazide in reverse transcriptase PCR to distinguish between infectious and noninfectious enteric viruses in water samples. Appl Environ Microbiol *76*, 4318-4326.

Parsons, P. (1996). Exobiology: dusting off panspermia. Nature *383*, 221-222.

Parte, A.C. (2014). LPSN — list of prokaryotic names with standing in nomenclature. Nucleic Acids Res *42*, D613-616.

Peak, D., Ford, R.G., and Sparks, D.L. (1999). An in Situ ATR-FTIR Investigation of sulfate bonding mechanisms on goethite. J Colloid Interface Sci *218*, 289-299.

Percival, E., and Wold, J. (1963). 1040. The acid polysaccharide from the green seaweed Ulva lactuca. Part II. The site of the ester sulphate. J Chem Soc (Resumed), 5459-5468.

Pereira, I.A., Ramos, A.R., Grein, F., Marques, M.C., da Silva, S.M., and Venceslau, S.S. (2011). A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. Front Microbiol *2*, 69.

Pereto, J., Lopez-Garcia, P., and Moreira, D. (2004). Ancestral lipid biosynthesis and early membrane evolution. Trends Biochem Sci *29*, 469-477.

Peschke, M., Moog, D., Klingl, A., Maier, U.G., and Hempel, F. (2013). Evidence for glycoprotein transport into complex plastids. Proc Natl Acad Sci U S A *110*, 10860-10865.

Pester, M., Schleper, C., and Wagner, M. (2011). The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. Curr Opin Microbiol *14*, 300-306.

Philippe, N., Legendre, M., Doutre, G., Coute, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C.*, et al.* (2013). Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science *341*, 281-286.

Podar, M., Makarova, K.S., Graham, D.E., Wolf, Y.I., Koonin, E.V., and 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.

Pointing, S.B., Chan, Y., Lacap, D.C., Lau, M.C., Jurgens, J.A., and Farrell, R.L. (2009). Highly specialized microbial diversity in hyper-arid polar desert. Proc Natl Acad Sci U S A *106*, 19964-19969.

Poretsky, R.S., Bano, N., Buchan, A., LeCleir, G., Kleikemper, J., Pickering, M., Pate, W.M., Moran, M.A., and Hollibaugh, J.T. (2005). Analysis of microbial gene transcripts in environmental samples. Appl Environ Microbiol *71*, 4121-4126.

Powell, S., Szklarczyk, D., Trachana, K., Roth, A., Kuhn, M., Muller, J., Arnold, R., Rattei, T., Letunic, I., Doerks, T.*, et al.* (2012). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucleic Acids Res *40*, D284-289.

Preheim, S.P., Perrotta, A.R., Martin-Platero, A.M., Gupta, A., and Alm, E.J. (2013). Distribution-based clustering: using ecology to refine the operational taxonomic unit. Appl Environ Microbiol *79*, 6593-6603.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 — approximately maximum-likelihood trees for large alignments. PLoS One *5*, e9490.

Probst, A., Facius, R., Wirth, R., and Moissl-Eichinger, C. (2010a). Validation of a nylon-flocked-swab protocol for efficient recovery of bacterial spores from smooth and rough surfaces. Appl Environ Microbiol *76*, 5148-5158.

Probst, A., Facius, R., Wirth, R., Wolf, M., and Moissl-Eichinger, C. (2011). Recovery of bacillus spore contaminants from rough surfaces: a challenge to space mission cleanliness control. Appl Environ Microbiol *77*, 1628-1637.

Probst, A., Lum, P., Bettina, J., Dubinsky, E., Piceno, Y., Tom, L., Andersen, G., He, Z., and DeSantis, T. (2014). Microarray of 16S rRNA gene probes for quantifying population differences across microbiome samples. In Microarrays: Cur-

rent Technology, Innovations and Applications, Z. He, ed. (Norfolk, UK: Horizon Scientific Press and Caister Academic Press).

Probst, A., Mahnert, A., Weber, C., Haberer, K., and Moissl-Eichinger, C. (2012). Detecting inactivated endospores in fluorescence microscopy using propidium monoazide. Intl J Astrobiol *11*, 117-123.

Probst, A., Vaishampayan, P., Osman, S., Moissl-Eichinger, C., Andersen, G.L., and Venkateswaran, K. (2010b). Diversity of anaerobic microbes in spacecraft assembly clean rooms. Appl Environ Microbiol *76*, 2837-2845.

Probst, A.J., Auerbach, A.K., and Moissl-Eichinger, C. (2013a). Archaea on human skin. PLoS One *8*, e65388.

Probst, A.J., Birarda, G., Holman, H.-Y.N., DeSantis, T.Z., Wanner, G., Andersen, G.L., Perras, A.K., Meck, S., Völkel, J., Bechtel, H.A.*, et al.* (submitted). Coupling genetic and chemical microbiome profiling reveals heterogeneity of archaeome and bacteriome in subsurface biofilms that are dominated by the same archaeal species. PLoS One.

Probst, A.J., Holman, H.Y., DeSantis, T.Z., Andersen, G.L., Birarda, G., Bechtel, H.A., Piceno, Y.M., Sonnleitner, M., Venkateswaran, K., and Moissl-Eichinger, C. (2013b). Tackling the minority: sulfate-reducing bacteria in an archaea-dominated subsurface biofilm. ISME J *7*, 635-651.

Probst, A.J., Weinmair, T., Raymann, K., Perras, A., Emerson, J.B., Rattei, T., Wanner, G., Klingl, A., Berg, I., Viehweger, B.*, et al.* (in preparation). Grappling with dark matter: Biology of an uncultivated subsurface archaeon. Nat Com.

Pruesse, E., Peplies, J., and Glockner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinf *28*, 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*, 7188-7196.

Puleo, J., Fields, N., Bergstrom, S., Oxborrow, G., Stabekis, P., and Koukol, R. (1977). Microbiological profiles of the Viking spacecraft. Appl Environ Microbiol *33*, 379-384.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J.*, et al.* (2012). The Pfam protein families database. Nucleic Acids Res *40*, D290-301.

Q

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., and Lopez, R. (2005). InterProScan: protein domains identifier. Nucleic Acids Res *33*, W116-W120.

Quince, C., Lanzen, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods *6*, 639-641.

R

R-Development-Core-Team (2005). R: A language and environment for statistical computing (ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria, 2013. url: http://R-project.org).

Radajewski, S., Ineson, P., Parekh, N.R., and Murrell, J.C. (2000). Stable-isotope probing as a tool in microbial ecology. Nature *403*, 646-649.

Raskin, L., Stromley, J.M., Rittmann, B.E., and Stahl, D.A. (1994). Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. App Environ Microbiol *60*, 1232-1240.

Rawsthorne, H., Dock, C.N., and Jaykus, L.A. (2009). PCR-based method using propidium monoazide to distinguish viable from nonviable *Bacillus subtilis* spores. Appl Environ Microbiol *75*, 2936-2939.

Raymann, K., Forterre, P., Brochier-Armanet, C., and Gribaldo, S. (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. Genome Biol Evol *6*, 192-212.

Reddy, P.S., Mahanty, S., Kaul, T., Nair, S., Sopory, S.K., and Reddy, M.K. (2008). A high-throughput genome-walking method and its use for cloning unknown flanking sequences. Anal Biochem *381*, 248-253.

Reguera, G., McCarthy, K.D., Mehta, T., Nicoll, J.S., Tuominen, M.T., and Lovley, D.R. (2005). Extracellular electron transfer via microbial nanowires. Nature *435*, 1098-1101.

Rieu-Lesme, F., Delbes, C., and Sollelis, L. (2005). Recovery of partial 16S rDNA sequences suggests the presence of Crenarchaeota in the human digestive ecosystem. Curr Microbiol *51*, 317-321.

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.*, et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431-437.

Rogers, G.B., Stressmann, F.A., Koller, G., Daniels, T., Carroll, M.P., and Bruce, K.D. (2008). Assessing the diagnostic importance of nonviable bacterial cells in respiratory infections. Diagnostic microbiology and infectious disease *62*, 133-141.

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., and Nyrén, P. I. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal Biochem *242*, 84-89.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol *61*, 539-542.

Rossel, P.E., Lipp, J.S., Fredricks, H.F., Arnds, J., Boetius, A., Elvert, M., and Hinrichs, K.-U. (2008). Intact polar lipids of anaerobic methanotrophic archaea and associated bacteria. Org Geochem *39*, 992-999.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., and Edwards, M. (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature *475*, 348-352.

Rudolph, C. (2003). Molekularbiologische Untersuchungen zur Verbreitung und Physiologie neuartiger, unkultivierter Archaeen in kalten Schwefelquellen. In Department for Microbiology and Archaea Center (Regensburg: University of Regensburg).

Rudolph, C., Moissl, C., Henneberger, R., and Huber, R. (2004). Ecology and microbial structures of archaeal/bacterial strings-of-pearls communities and archaeal relatives thriving in cold sulfidic springs. FEMS Microbiol Ecol *50*, 1-11.

Rudolph, C., Wanner, G., and Huber, R. (2001). Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology. Appl Environ Microbiol *67*, 2336-2344.

Rummel, J.D. (1989). Planetary protection policy overview and application to future missions. Adv Space Res *9*, 181-184.

S

Saier, M.H., Jr., Tran, C.V., and Barabote, R.D. (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. Nucleic Acids Res *34*, D181-186.

Sakai, Y., Mitsui, R., Katayama, Y., Yanase, H., and Kato, N. (1999). Organization of the genes involved in the ribulose monophosphate pathway in an obligate methylotrophic bacterium, *Methylomonas aminofaciens* 77a. FEMS Microbiol Lett *176*, 125-130.

Saleh, A., Macpherson, R., and Miller, J. (1964). The effect of inhibitors on sulphate reducing bacteria: a compilation. J Appl Microbiol *27*, 281-293.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., and Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. Nature *265*, 687-695.

Sauder, L.A., Peterse, F., Schouten, S., and Neufeld, J.D. (2012). Low-ammonia niche of ammonia-oxidizing archaea in rotating biological contactors of a municipal wastewater treatment plant. Environ Microbiol *14*, 2589-2600.

Say, R.F., and Fuchs, G. (2010). Fructose 1, 6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. Nature *464*, 1077-1081.

Schauder, R., Preuß, A., Jetten, M., and Fuchs, G. (1988). Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum*. Arch Microbiol *151*, 84-89.

Schiex, T., Gouzy, J.r.m., Moisan, A., and de Oliveira, Y. (2003). FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res *31*, 3738-3741.

Schloss, P.D., Gevers, D., and Westcott, S.L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS ONE *6*, e27310.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., and Robinson, C.J. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol *75*, 7537-7541.

Schmid, M., Schmitz-Esser, S., Jetten, M., and Wagner, M. (2001). 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and in situ detection. Environ Microbiol *3*, 450-

459.

Schmitz, R., Albracht, S., and Thauer, R. (1992a). Properties of the tungsten-substituted molybdenum formylmethanofuran dehydrogenase from *Methanobacterium wolfei*. FEBS letters *309*, 78-81.

Schmitz, R.A., Albracht, S.P., and Thauer, R.K. (1992b). A molybdenum and a tungsten isoenzyme of formylmethanofuran dehydrogenase in the thermophilic archaeon *Methanobacterium wolfei*. Europ J Biochem *209*, 1013-1018.

Schmitz, R.A., Bertram, P.A., and Thauer, R.K. (1994). Tungstate does not support synthesis of active formylmethanofuran dehydrogenase in *Methanosarcina barkeri*. Arch Microbiol *161*, 528-530.

Schmitz, R.A., Richter, M., Linder, D., and Thauer, R.K. (1992c). A tungsten containing active formylmethanofuran dehydrogenase in the thermophilic archaeon *Methanobacterium wolfei*. Europ J Biochem *207*, 559-565.

Schuchmann, K., and Muller, V. (2013). Direct and reversible hydrogenation of CO_2 to formate by a bacterial carbon dioxide reductase. Science *342*, 1382-1385.

Schwendner, P., Moissl-Eichinger, C., Barczyk, S., Bohmeier, M., Pukall, R., and Rettberg, P. (2013). Insights into the microbial diversity and bioburden in a South American spacecraft assembly clean room. Astrobiol *13*, 1140-1154.

Secker, J., Lepock, J., and Wesson, P. (1994). Damage due to ultraviolet and ionizing radiation during the ejection of shielded micro-organisms from the vicinity of 1M⊙ main sequence and red giant stars. Astrophys Space Sci *219*, 1-28.

Sharon, I., and Banfield, J.F. (2013). Genomes from metagenomics. Science *342*, 1057-1058.

Sinclair, R., McKay, A., Myers, G., and Jones, R.N. (1952). The Infrared Absorption Spectra of Unsaturated Fatty Acids and Esters1. Journal of the American Chemical Society *74*, 2578-2585.

Smith, B.C. (1998). Infrared spectral interpretation: a systematic approach (CRC press).

Smith, K.S., and Ingram-Smith, C. (2007). *Methanosaeta*, the forgotten methanogen? Trends Microbiol *15*, 150-155.

Snel, B., Huynen, M.A., and Dutilh, B.E. (2005). Genome trees and the nature of genome evolution. Annu Rev Microbiol *59*, 191-209.

Socrates, G. (2004). Infrared and raman characteristic group frequencies: Tables and charts (New York City: John Wiley & Sons).

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere”. Proc Natl Acad Sci U S A *103*, 12115-12120.

Spang, A., Hatzenpichler, R., Brochier-Armanet, C., Rattei, T., Tischler, P., Spieck, E., Streit, W., Stahl, D.A., Wagner, M., and Schleper, C. (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. Trends Microbiol *18*, 331-340.

Sprott, G.D. (1992). Structures of archaebacterial membrane lipids. J Bioenerg Biomembr *24*, 555-566.

Stackebrandt, E., and Goebel, B. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol *44*, 846-849.

Stahl, D., and Amann, R. (1991). Development and application of nucleic acid probes. In Nucleic acid techniques in bacterial systemaitcs, E. Stackebrandt, and M. Goodfellow, eds. (New Yord: John Wiley and Sons), pp. 205-248.

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinf *21*, 456-463.

Stellmach, J. (1984). Fluorescent redox dyes. 1. Production of fluorescent formazan by unstimulated and phorbol ester- or digitonin-stimulated Ehrlich ascites tumor cells. Histochem *80*, 137-143.

Stellmach, J., and Severin, E. (1987). A fluorescent redox dye. Influence of several substrates and electron carriers on the tetrazolium saltâ”formazan reaction of Ehrlich ascites tumour cells. Histochem J *19*, 21-26.

Stieglmeier, M., Rettberg, P., Barczyk, S., Bohmeier, M., Pukall, R., Wirth, R., and Moissl-Eichinger, C. (2012). Abundance and diversity of microbial inhabitants in European spacecraft-associated clean rooms. Astrobiology *12*, 572-585.

Stieglmeier, M., Wirth, R., Kminek, G., and Moissl-Eichinger, C. (2009). Cultivation of anaerobic and facultatively anaerobic bacteria from spacecraft-associated clean rooms. Appl Environ Microbiol *75*, 3484-3491.

Stokke, R., Roalkvam, I., Lanzen, A., Haflidason, H., and Steen, I.H. (2012). Integrated metagenomic and metaproteomic analyses of an ANME-1-dominated community in marine cold seep sediments. Environ Microbiol *14*, 1333-1346.

Sturt, H.F., Summons, R.E., Smith, K., Elvert, M., and Hinrichs, K.U. (2004). Intact polar membrane lipids in prokaryotes and

sediments deciphered by high-performance liquid chromatography/electrospray ionization multistage mass spectrometry--new biomarkers for biogeochemistry and microbial ecology. Rapid Commun Mass Spectrom *18*, 617-628.

Summit, M., and Baross, J.A. (2001). A novel microbial habitat in the mid-ocean ridge subseafloor. Proc Natl Acad Sci U S A *98*, 2158-2163.

Suzuki, M.T., Taylor, L.T., and DeLong, E.F. (2000). Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl Environ Microbiol *66*, 4605-4614.

T

Takano, B. (1985). Geochemical implications of sulfate in sedimentary carbonates. Chemical Geology *49*, 393-403.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol *24*, 1596-1599.

Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res *28*, 33-36.

Thauer, R.K. (2011). Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO2. Curr Opin Microbiol *14*, 292-299.

Thirkell, D., and Gray, E.M. (1974). Variation in the lipid and fatty acid composition in purified membrane fractions from Sarcina aurantiaca in relation to growth phase. Antonie van Leeuwenhoek *40*, 71-78.

Tillett, D., and Neilan, B.A. (2000). Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. J Phycol *36*, 251-258.

Tobias, C., and Todd, P.E. (1974). Space Radiation Biology and Related Topics (New York: Academic Press).

Tomova, I., Dimitrova, D., Stoilova-Disheva, M., Lyutskanova, D., and Kambourova, M. (2011). Archaeal diversity at two hot springs, Rupi Basin, Bulgaria. Biotechnology and Biotechnological Equipment *25*, 105-113.

Tourna, M., Freitag, T.E., Nicol, G.W., and Prosser, J.I. (2008). Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. Environ Microbiol *10*, 1357-1364.

Tourna, M., Stieglmeier, M., Spang, A., Konneke, M., Schintlmeister, A., Urich, T., Engel, M., Schloter, M., Wagner, M., Richter, A.*, et al.* (2011). *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. Proc Natl Acad Sci U S A *108*, 8420-8425.

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S., and Banfield, J.F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature *428*, 37-43.

U

Ulrich, G.A., Martino, D., Burger, K., Routh, J., Grossman, E.L., Ammerman, J.W., and Suflita, J.M. (1998). Sulfur cycling in the terrestrial subsurface: commensal interactions, spatial scales, and microbial heterogeneity. Microb Ecol *36*, 141-151.

Ulrih, N.P., Gmajner, D., and Raspor, P. (2009). Structural and physicochemical properties of polar lipids from thermophilic archaea. Appl Microbiol Biotechnol *84*, 249-260.

V

Vaishampayan, P., Osman, S., Andersen, G., and Venkateswaran, K. (2010). High-Density 16S Microarray and Clone Library-Based Microbial Community Composition of the Phoenix Spacecraft Assembly Clean Room. Astrobiology *10*, 499-508.

Vaishampayan, P., Probst, A.J., La Duc, M.T., Bargoma, E., Benardini, J.N., Andersen, G.L., and Venkateswaran, K. (2013). New perspectives on viable microbial communities in low-biomass cleanroom environments. ISME J *7*, 312-324.

Vaishampayan, P., Roberts, A., Augustus, A., Schwendner, P., Mayilraj, S., and Salmassi, T. (submitted). *Deinococcus phoenicis* sp. nov., an extreme ionizing radiation resistant bacterium isolated from the Phoenix Lander assembly facility. Int J Syst Evol Microbiol.

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fevre, F., Longin, C., Mornico, D., Roche, D.*, et al.* (2013). MicroScope — an integrated microbial resource for the curation and comparative analysis of genomic

and metabolic data. Nucleic Acids Res *41*, D636-647.

Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., and Medigue, C. (2006). MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res *34*, 53-65.

van Frankenhuyzen, J.K., Trevors, J.T., Lee, H., Flemming, C.A., and Habash, M.B. (2011). Molecular pathogen detection in biosolids with a focus on quantitative PCR using propidium monoazide for viable cell enumeration. J Microbiol Meth *87*, 263-272.

Vaneechoutte, M., Rossau, R., De Vos, P., Gillis, M., Janssens, D., Paepe, N., De Rouck, A., Fiers, T., Claeys, G., and Kersters, K. (1992). Rapid identification of bacteria of the *Comamonadaceae* with amplified ribosomal DNA-restriction analysis (ARDRA). FEMS Microbiol Lett *72*, 227-233.

Venkateswaran, K., Satomi, M., Chung, S., Kern, R., Koukol, R., Basic, C., and White, D. (2001). Molecular microbial diversity of a spacecraft assembly facility. Syst Appl Microbiol *24*, 311-320.

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D., Paulsen, I., Nelson, K.E., Nelson, W.*, et al.* (2004). Environmental genome shotgun sequencing of the sargasso sea. Science *304*, 66-74.

Ventosa, A., and Oren, A. (1996). Halobacterium salinarum nom. corrig., a name to replace *Halobacterium salinarium* (Elazari-Volcani) and to include *Halobacterium halobium* and *Halobacterium cutirubrum*. Int J Syst Bacteriol *46*, 347-347.

Vesper, S., McKinstry, C., Hartmann, C., Neace, M., Yoder, S., and Vesper, A. (2008). Quantifying fungal viability in air and water samples using quantitative PCR after treatment with propidium monoazide (PMA). J Microbiol Meth *72*, 180-184.

Vorholt, J.A., Chistoserdova, L., Lidstrom, M.E., and Thauer, R.K. (1998). The NADP-dependent methylene tetrahydromethanopterin dehydrogenase in *Methylobacterium extorquens* AM1. J Bacteriol *180*, 5351-5356.

Vorholt, J.A., Vaupel, M., and Thauer, R.K. (1996). A polyferredoxin with eight [4Fe4S] clusters as a subunit of molybdenum formylmethanofuran dehydrogenase from *Methanosarcina barkeri*. Europ J Biochem *236*, 309-317.

W

Wagner, A.O., Malin, C., Knapp, B.A., and Illmer, P. (2008). Removal of free extracellular DNA from environmental samples by ethidium monoazide and propidium monoazide. Appl Environ Microbiol *74*, 2537-2539.

Wagner, M. (2009). Single-cell ecophysiology of microbes as revealed by Raman microspectroscopy or secondary ion mass spectrometry imaging. Annu Rev Microbiol *63*, 411-429.

Wagner, M., Roger, A.J., Flax, J.L., Brusseau, G.A., and Stahl, D.A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. J Bacteriol *180*, 2975-2982.

Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., Brochier-Armanet, C., Chain, P.S., Chan, P.P., Gollabgir, A.*, et al.* (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. Proc Natl Acad Sci U S A *107*, 8818-8823.

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*, 136-143.

Walsh, C. (1986). Naturally occurring 5-deazaflavin coenzymes: biological redox roles. Acc Chem Res *19*, 216-221.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genom Proteom Bioinf *8*, 77-80.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol *73*, 5261-5267.

Ward, D.M., Weller, R., and Bateson, M.M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature *345*, 63-65.

Ward Jr, J.H. (1963). Hierarchical grouping to optimize an objective function. J Am Stat Assoc *58*, 236-244.

Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M.*, et al.* (2003). The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci U S A *100*, 12984-12988.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., and Ley,

R.E. (2012). Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. ISME J *6*, 94-103.

White, W. (1974). The carbonated minerals (London: Mineralogical Society).

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A *95*, 6578-6583.

Whittaker, P., Mossoba, M.M., Al-Khaldi, S., Fry, F.S., Dunkel, V.C., Tall, B.D., and Yurawecz, M.P. (2003). Identification of foodborne bacteria by infrared spectroscopy using cellular fatty acid methyl esters. J Microbiol Methods *55*, 709-716.

Williams, T.A., Embley, T.M., and Heinz, E. (2011). Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. PLoS One *6*, e21080.

Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. Nature *504*, 231-236.

Wilmes, P., and Bond, P.L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. Environ Microbiol *6*, 911-920.

Woese, C.R., and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A *74*, 5088-5090.

Woese, C.R., Kandler, O., and 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*, 4576-4579.

Wörmer, L., Lipp, J.S., Schröder, J.M., and Hinrichs, K.-U. (2013). Application of two new LC-ESI-"MS methods for improved detection of intact polar lipids (IPLs) in environmental samples. Organic Geochemistry *59*, 10-21.

Wortman, J., Giglio, M., Creasy, H., Chen, A., Liolios, K., Chu, K., Davidovics, N., Mazaitis, M., DeSantis, T., Singh, N.*, et al.* (2010). A data analysis and coordination center for the human microbiome project. Genome Biol *11*, 1-1.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H.*, et al.* (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science *337*, 1661-1665.

Wu, D., Wu, M., Halpern, A., Rusch, D.B., Yooseph, S., Frazier, M., Venter, J.C., and Eisen, J.A. (2011). Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. PLoS One *6*, e18011.

Y

Yanase, H., Ikeyama, K., Mitsui, R., Ra, S., Kita, K., Sakai, Y., and Kato, N. (1996). Cloning and sequence analysis of the gene encoding 3-hexulose-6-phosphate synthase from the methylotrophic bacterium, *Methylomonas aminofaciens* 77a, and its expression in *Escherichia coli*. FEMS Microbiol Lett *135*, 201-205.

Yanez, M.A., Nocker, A., Soria-Soria, E., Murtula, R., Martinez, L., and Catalan, V. (2011). Quantification of viable *Legionella pneumophila* cells using propidium monoazide combined with quantitative PCR. J Microbiol Meth *85*, 124-130.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P.*, et al.* (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222-227.

Yilmaz, S., Allgaier, M., and Hugenholtz, P. (2010). Multiple displacement amplification compromises quantitative analysis of metagenomes. Nat Methods *7*, 943-944.

Yoshida, N., and Hiraishi, A. (2004). An improved redox dye-staining method using 5-cyano-2, 3-ditoryl tetrazolium chloride for detection of metabolically active bacteria in activated sludge. Microb Environ *19*, 61-70.

Yoshinaga, M.Y., Kellermann, M.Y., Rossel, P.E., Schubotz, F., Lipp, J.S., and 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*, 3563-3574.

Yurimoto, H., Kato, N., and Sakai, Y. (2005). Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. Chem Rec *5*, 367-375.

Z

Zhang, L., Cui, X., Schmitt, K., Hubert, R., Navidi, W., and Arnheim, N. (1992). Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci U S A *89*, 5847-5851.

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiol *156*, 287-301.

Zhang, Z., Xiao, J., Wu, J., Zhang, H., Liu, G., Wang, X., and Dai, L. (2012). ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. Biochem Biophys Res Commun *419*, 779-781.

Zhou, H.W., Li, D.F., Tam, N.F., Jiang, X.T., Zhang, H., Sheng, H.F., Qin, J., Liu, X., and Zou, F. (2011). BIPES, a cost-effective high-throughput method for assessing microbial diversity. ISME J *5*, 741-749.

VII. Content of supporting DVD

Table VII-1 Data structure on the supporting DVD

Folder	Description
./List_of_Publications	Folder contains pdf files of all publications listed on pp. 6-8
./Publications/01_Vaishampayan-Probst-LaDuc_ISME_2013/	Contains data for the publication of Chapter III.1
./Publications/02_Probst_PLOS-ONE_2013/	Contains data for the publication of Chapter III.2
./Publications/03_Probst_ISMEJ_2013/	Contains data for the publication of Chapter III.3
./Publications/04_Probst-Birarda_PLOS-ONE_2014/	Contains data for the publication of Chapter III.4
./Publications/05_Probst_in-prep/	Contains data for the publication of Chapter III.5

Table VII-2 Data structure of subfolders in ./Publications/0*/

Folder	Description
~/Manuscript/	Contains a copy of the manuscript as pdf
~/Supporting_information/	Contains supporting information along with the manuscript
~/Additional_data/	Contains additional files like genome sequences or high resolution images that are or were not intended to be published

VIII. Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe des Literaturzitats gekennzeichnet.

Regensburg, 31.03.2014 _____

Alexander J. Probst