Unraveling the critical growth factors for stable cultivation of (nano-sized) Micrarchaeota

Susanne Krause [1]*, Sabrina Gfrerer [2]*, Carsten Reuse [7,8], Nina Dombrowski [3], Laura Villanueva [3,4], Boyke Bunk [5], Cathrin Spröer [5], Thomas R. Neu [6], Ute Kuhlicke [6], Kerstin Schmidt-Hohagen [7,8], Karsten Hiller [7,8], Reinhard Rachel [9] Anja Spang [3,10] and Johannes Gescher [1,2]

[1] Department of Applied Biology, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[2] Institute for Biological Interfaces, Karlsruhe Institute of Technology (KIT), Eggenstein-Leopoldshafen, Germany
[3] Department of Marine Microbiology and Biogeochemistry, NIOZ, Royal Netherlands Institute for Sea Research, Den Burg, The Netherlands
[4] Department of Earth Sciences, Faculty of Geosciences, Utrecht University, Utrecht, The Netherlands
[5] Leibniz Institute DSMZ, Braunschweig, Germany
[6] Helmholtz-Centre for Environmental Research UFZ, Magdeburg, Germany
[7] Bioinformatics & Biochemistry, Technische Universität Braunschweig, Braunschweig, Germany
[8] Braunschweig Integrated Centre for Systems Biology (BRICS), Technische Universität Braunschweig, Braunschweig, Germany
[9] Center for Electron Microscopy, University of Regensburg, Regensburg, Germany
[10] Department of Cell- and Molecular Biology, Science for Life Laboratory, Uppsala University, SE-75123, Uppsala, Sweden

Corresponding author:
Prof. Dr. Johannes Gescher
Karlsruhe Institute of Technology
Institute of Applied Biosciences
Department of applied biology
Phone: ++49721-608-41940
FAX: ++49721-608-41941
E-mail: johannes.gescher@kit.edu

* These authors contributed equally to this work.

Abstract

Micrarchaeota are members of the archaeal DPANN superphylum. These so far poorly characterized archaea have been found to have reduced genomes and likely depend on interactions with host organisms for growth and survival. Here we report on the enrichment of the first stable co-culture of a member of the Micrarchaeota together with its host, as well as the isolation of the latter. Electron microscopic analysis suggest that growth is dependent on the physical interaction of the two organisms within a biofilm. The interaction seems to be ensured by the necessity to grow in form of a biofilm. Furthermore, transcriptomic analyses indicate a shift towards biofilm formation of the host as a result of co-cultivation. Finally, genomic, metabolomic, extracellular polymeric substance (EPSs) and lipid content analyses reveal that the Micrarchaeon symbiont relies on the acquisition of metabolites from its host and thereby provide first insights into the basis of symbiont-host interactions.

Introduction

In 2002, Huber and colleagues described a novel nano-sized archaeon, *Nanoarchaeum equitans* [1]. Later, metagenomic data of environmental samples revealed that the Nanoarchaeota are part of a tentative superphylum of nano-sized archaea now referred to as DPANN – an acronym on its first members lineages, the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, and Nanohaloarchaeota [2]. Most DPANN representatives have reduced genomes and are thought to comprise a diversity of potential archaeal symbionts. Besides the name-giving phyla, the DPANN also include the Woese- and Pacearchaeota [3], Huberarchaeota [4], Micrarchaeota [5], Altiarchaeota [6], Undinarchaeota [7] and Mamarchaeota [8,9] as well as several so far undefined phyla [8,10]. Nano-sized archaea are globally distributed and can comprise non-negligible proportions of microbial communities [11]. Yet only a few representatives have been enriched under laboratory conditions [1,12–17]. Current genomic data suggest that most DPANN archaea have reduced genomes, limited metabolic capabilities and various auxotrophies and might depend on interactions with other community members. The extent of genome reduction varies within the DPANN members. For instance, marine Nanoarchaeota are characterized by highly reduced genomes of about 0.5 Mbp and seem to represent ectoparasites that are strongly dependent on their host [1]. On the other hand, the first members of the Nanohaloarchaeota and Micrarchaeota [5] have larger genome sizes and seem metabolically more flexible [14–16,18]; yet cultivated Nanohaloarchaeota representatives are nevertheless host-dependent [16,17]. While recent work has provided more insights into symbiotic interactions characterizing certain representatives of the DPANN [16,19], additional model systems remain to be established.

We have recently succeeded in enriching a member of the Micrarchaeota in a community of four different microorganisms [15]. Here, we report the isolation of the first stable coculture of this Micrarchaeon together with its host, a previously unknown member of the Thermoplasmatales, as well as the isolation of the latter. This allowed us to conduct experiments

86 aiming to understand the interaction of the two organisms and the response of the
87 Thermoplasmatales member to growth in coculture with the Micrarchaeon.

Results and discussion

Isolation of the A_DKE/B_DKE co-culture and pure culture of B_DKE

Originally, we referred to the putative symbionts belonging to the Micrarchaeota and Thermoplasmatales as A_DKE and B_DKE, respectively [15]. Based on the herein purified co-culture, the isolation of the Thermoplasmatales host the reconstruction of genome sequences of both organisms, we propose the names '*Candidatus* Micrarchaeum harzensis' *sp. nov.* (N.L. masc./fem. adj. *harzensis*, pertaining to the German region of the Harz Mountains, where the organism was isolated) and '*Candidatus* Scheffleriplasma hospitalis' *gen. nov. sp. nov.* (Schef'fler.i.plas´ma. N.L. gen. masc. n. *scheffleri* of Scheffler, named in honor of the geologist Dr. Horst Scheffler and in recognition of his work on mine geology and commitment to our work; Gr. neut. n. *plasma* something shaped or molded; hos.pi.ta′lis. L. masc. adj. *hospitalis* relating to a guest, hospitable, referring to its ability to serve as a host for *Candidatus* 'Micrarchaeum harzensis'). For brevity we will stick to the abbreviation A_DKE and B_DKE throughout this manuscript.

In order to obtain a better understanding of the interactions between A_DKE and B_DKE we first aimed to obtain a co-culture of the putative symbiont and host organisms, i.e. purify the original enrichment from a *Cuniculiplasma*-related archaeon referred to as C_DKE and a fungus (with *Acidothrix acidophila* as its closest related organism) [15] C_DKE represented the minority of the archaea in the enrichment culture and is closely related to an organism reported to have a low pH optimum of 1.0–1.2 [20]. Hence, it was possible to eliminate C_DKE by transferring the culture in media with a pH of 2.5, which exceeds its optimal pH range, while still supporting the growth of A_DKE and B_DKE. Secondly, the fungus, which was isolated and inferred to be psychrophilic/psychrotolerant (data not shown), was successfully eliminated by incubating the enrichment cultures at 37 °C over three consecutive culture transfers. From the previous microscopy analysis, it was known that A_DKE thrives together with B_DKE in biofilm-like structures [15]. We discovered that it was possible to enhance the biofilm formation

of the host- or helper-organism B_DKE by lowering the pH-value of the medium from 2.5 to 2.0, which led to robust biofilm formation of the co-culture (Supplementary Figure S1). At the same time, we were able to isolate B_DKE by enriching at pH 2.5 for planktonic organisms.

Genomic potential of A_DKE and B_DKE

DNA of the co-culture containing A_DKE and B_DKE was sequenced using a combination of PacBio and Illumina sequencing. For comparison, Illumina sequencing was also performed on the B_DKE pure culture. The two organisms have circular chromosomes of 1 959 588 base pairs (bp) (B_DKE) and 989 838 bp (A_DKE), and GC contents of 44.4 % and 45.8 %, respectively. The genomes of the pure B_DKE isolate and the strain within the co-culture were 100 % identical, which was important for the later comparative transcriptomic analysis. An analysis of clusters of orthologous groups (COGs) revealed that A_DKE contains more proteins with unknown function (29 %; 300 putative proteins without an arCOG-assignment) relative to the overall number of genes compared to B_DKE (20 %; 419 putative proteins without an arCOG-assignment). The complete genome of A_DKE confirmed earlier findings [15], such as an extremely limited set of genes coding for proteins involved in central carbon metabolism. We could only detect one gene encoding a putative enzyme of the pentose phosphate pathway and two genes for putative enzymes of a glycolysis or gluconeogenesis pathways (Supplementary Table S1). However, we did identify a putative set of genes coding for enzymes for the conversion of glucose to glycerate, which together comprise four of the seven reactions of the non-phosphorylative Entner-Doudoroff pathway (Supplementary Table S1). The A_DKE genome also contains genes encoding enzymes for the conversion of pyruvate to acetyl-CoA, though we could not identify candidate proteins for reactions leading from glycerate to pyruvate. In agreement with our previous study, proteins for almost all steps of the tricarboxylic acid cycle (TCA) were detected in A_DKE, and may, through the production of NADH, fuel an electron transport chain and the generation of a proton gradient for ATP synthase-based ATP

production. Gene clusters encoding a full NADH dehydrogenase and an ATP synthase complex were discovered in the A_DKE genome. Moreover, we identified genes encoding one subunit of the cytochrome bc1 complex and two subunits of the cytochrome c oxidase (Supplementary Table S1). Although the organism might have the ability to generate energy and produce reducing equivalents, it will be highly dependent on building blocks acquired either from the environment (or culture medium) or from the partner organism B_DKE. For instance, A_DKE has major gaps in various biosynthesis pathways including for amino acids; the few steps encoded by A_DKE comprise the production of aspartate from oxaloacetate, glutamate from α-ketoglutarate and phenylalanine from phenylpyruvate. Phenylpyruvate could be produced from tyrosine, which was taken up from the medium by the co-culture (see below). Other amino acid biosynthesis pathways could not be detected. Furthermore, genes encoding known amino acid transporters seem to be absent [8,10] and DNA, RNA, and lipid biosynthesis pathways are incomplete (see below). Consequently, A_DKE may acquire certain metabolites or building blocks directly from its partner B_DKE through cell-cell interactions as seen in the *Ignococcus hospitalis/Nanoarchaeum equitans* system [21,22]. In turn, the dependency of A_DKE on growth in a biofilm (see above) may be due to the need to establish cellular contact with its host.

Impact of growth in co-culture on the B_DKE transcriptome

To get further insights into the effect of the symbiont on its host, we compared gene expression levels of B_DKE with and without co-cultivation with A_DKE under otherwise identical growth conditions. In particular, we compared three pure cultures with four co-cultures and filtered for differentially expressed genes with p-values lower than 0.05 and log2fold changes higher or lower than 2 or -2. This analysis revealed 15 genes that were differentially expressed based on these criteria (Table 1).

Table 1: Differentially expressed genes of B_DKE in pure culture in comparison to B_DKE in co-culture with A_DKE. The table indicates the gene ID, the description of the expressed protein, the log2-fold change, and the p-value.

gene ID	description	log2 fold change	p-value
Thermo_01860	hypothetical protein	-5,42	7,32E-201
Thermo_01798	hypothetical protein	-5,21	6,1E-97
Thermo_00025	hypothetical protein	-4,87	0
Thermo_00444	41 kDa flagellin	-4,4	0
Thermo_01859	41 kDa flagellin	-4,33	3,06E-87
Thermo_01877	Pilin/Flagellin family protein	-4,06	5,16E-93
Thermo_00680	putative aminopeptidase 1	-4,05	1,29E-292
Thermo_00445	Putative archaeal flagellar protein C	-4	0
Thermo_00502	Hexuronic acid methyltransferase AglP	-3,83	1,04E-232
Thermo_00446	Putative archaeal flagellar protein D/E	-3,59	0
Thermo_01799	hypothetical protein	-3,23	9,13E-80
Thermo_01144	Pink FeS protein	-2,88	2,75E-27
Thermo_01181	Trehalose/maltose import ATP-binding protein MalK	-2,26	1,73E-60
Thermo_00581	Trehalose/maltose import ATP-binding protein MalK	-2,22	2,51E-77
Thermo_00582	ABC-type transport system permease component	-2,13	9,45E-70

All of those 15 genes were downregulated in the host B_DKE in the co-culture compared to the pure culture of B_DKE and comprise five genes of or associated with the B_DKE archaellum complex[23]. This suggests a decreased cellular motility of the host in the co-culture, which is in line with the observed tendency of B_DKE to form a biofilm in the presence of A_DKE. The closest relative of B_DKE, *Cuniculiplasma divulgatum*, does not contain any flagella-related genes[24], whereas the related 'G-Plasmas' are described to contain the full *fla*-operon (*flaBCDEFGHIJ*)[25]. Furthermore, the gene for the hexuronic acid methyltransferase AglP, a component of the protein glycosylation was downregulated and may indicate an alteration of the glycosylation pattern of B_DKE, as well as influence cell-cell-interaction in the presence of A_DKE. It may also change the exopolysaccharide (EPS) matrix of B_DKE, which would explain the observed binding differences of some lectins (see below). Three other downregulated genes code for transport proteins that might be involved in the uptake of carbohydrate molecules. While the effect of the decreased expression level of these transporters is unclear, it may be speculated that it could lead to higher availability of certain metabolites in the medium and support growth of A_DKE. Other downregulated genes encode hypothetical

proteins, a putative aminopeptidase, and an iron-sulfur-protein. Thus far, their potential impact on the interaction of B_DKE with A_DKE remains unclear.

Metabolomic analysis in the presence and absence of A_DKE

Next, we performed a metabolomic analysis to be able to compare metabolites in co-culture to the pure culture and determine whether the presence of A_DKE changes the pattern of depleted and produced organic carbon compounds. The growth of B_DKE in pure culture, estimated on the change in ferrous iron concentration over time, was faster than in the co-culture (Supplementary Figure S2). Therefore, we compared samples with equal ferrous iron concentration (3, 4, and 5 weeks of growth for the pure culture; 5, 6, and 7 weeks of growth for the co-culture) as they indicate similar growth phases. Figure 1 shows a heatmap representing all significantly different metabolites ($p < 0.01$) present in the samples at the various time points. While the majority of the detected compounds changed simultaneously in the pure- and the co-culture throughout the growth phases, some metabolites showed different patterns.

Figure 1: Heatmap showing significantly different metabolite levels ($p < 0.01$) between corresponding growth phases of pure culture (B_DKE) and co-culture (A_DKE/B_DKE). Normalization was done using a z-score, and significance was calculated by a two-tailed t-test.

Tyrosine levels were found to decrease faster in the co-culture, whereas 4-hydroxyphenyl acetic acid, an intermediate of the tyrosine degradation pathway, increased. Hence, tyrosine degradation seems to be accelerated in the co-culture. Both A_DKE and B_DKE encode enzymes catalyzing the conversion of tyrosine to 4-hydroxyphenylpyuvate (Supplementary table S1) and the presence of the Micrarchaeon might be the reason for faster tyrosine depletion. Note, the specific genes for the carboxylase and dehydrogenase reaction from 4-hydroxyphenylpyruvate to 4-hydroxyphenyl acetic acid are currently unknown. While

2-phenylglycine, a degradation product of phenylalanine that enters the ketoadipate pathway, decreased, muconic acid, an intermediate of that pathway, increased in the co-culture. Still, we could neither identify genes involved in the degradation of 2-phenylglycine nor for a complete ß-ketoadipate pathway in A_DKE or B_DKE (Supplementary Table S1). Moreover, the ß-ketoadipate pathway operates under oxic conditions and the organisms were cultivated in the absence of oxygen. Hence, so far, we cannot explain the consumption and production of 2-phenylglycine and muconic acid, respectively. The analyses also revealed increased levels of gluconic acid in the co-culture, which may be a product of sugar-degradation, for instance from the biofilm EPS matrix [26]. Genomic information indicates that B_DKE is able to degrade glucose into glucono-1,5-lactone (KO: K18124), which can spontaneously be converted into gluconic acid (Supplementary Table S1). Both organisms possess the enzymes to convert gluconic acid to glycerate, as already discussed above.

Overall, the analysis reveals that the pattern of metabolites does not seem to deviate between isolated B_DKE and co-culture and that the kinetics of consumption show minor differences towards faster consumption of some compounds in the co-culture. Hence, either both organisms employ similar pathways and compounds or, perhaps more likely, A_DKE predominantly uses metabolites provided by B_DKE. The latter assumption is in agreement with the sparsity of transporters encoded in the A_DKE genome and indications for cell-cell interactions among the two organisms.

Membrane lipids of A_DKE and B_DKE and lipid biosynthetic pathways

An analysis of the intact polar lipids (IPLs) of co-cultures of A_DKE and B_DKE revealed the membrane archaeal isoprenoidal glycerol dibiphytanyl glycerol (GDGT) with zero cyclopentane rings (i.e. GDGT-0, also known as caldarchaeol) as the main lipid, making up to 97 % of the total intact polar lipids, together with a minor amount of archaeol (2,3-di-O-phytanyl glycerol diether) (Supplementary Table S2). Comparison of the results to the pure

238 B_DKE-culture revealed no differences in the relative abundance of the archaeal IPLs,
239 suggesting that the Micrarchaeon A_DKE has an identical membrane lipid composition as its
240 host B_DKE.

241 Archaeal membrane lipids are formed by isoprenoid side chains linked through ether
242 bonds to glycerol-1-phosphate (G1P, synthesized by the G1P-dehydrogenase[27]) either as a
243 bilayer of diethers (archaeols) or a monolayer of tetraethers (i.e. GDGTs). The isoprenoid
244 building blocks are synthesized by one of the four variants of the archaeal mevalonate (MVA)
245 pathway, which differ with regard to the enzymes mediating the last three enzymatic steps (see
246 [28] for a review). Figure 2 shows an overview of the different MVA pathways known. The
247 isoprenoid C20 units are linked to the G1P backbone through ether bonds by the
248 geranylgeranylglyceryl diphosphate (GGGP) synthase and (S)-2,3-di-O-
249 geranylgeranylglyceryl diphosphate (DGGGP) synthase.

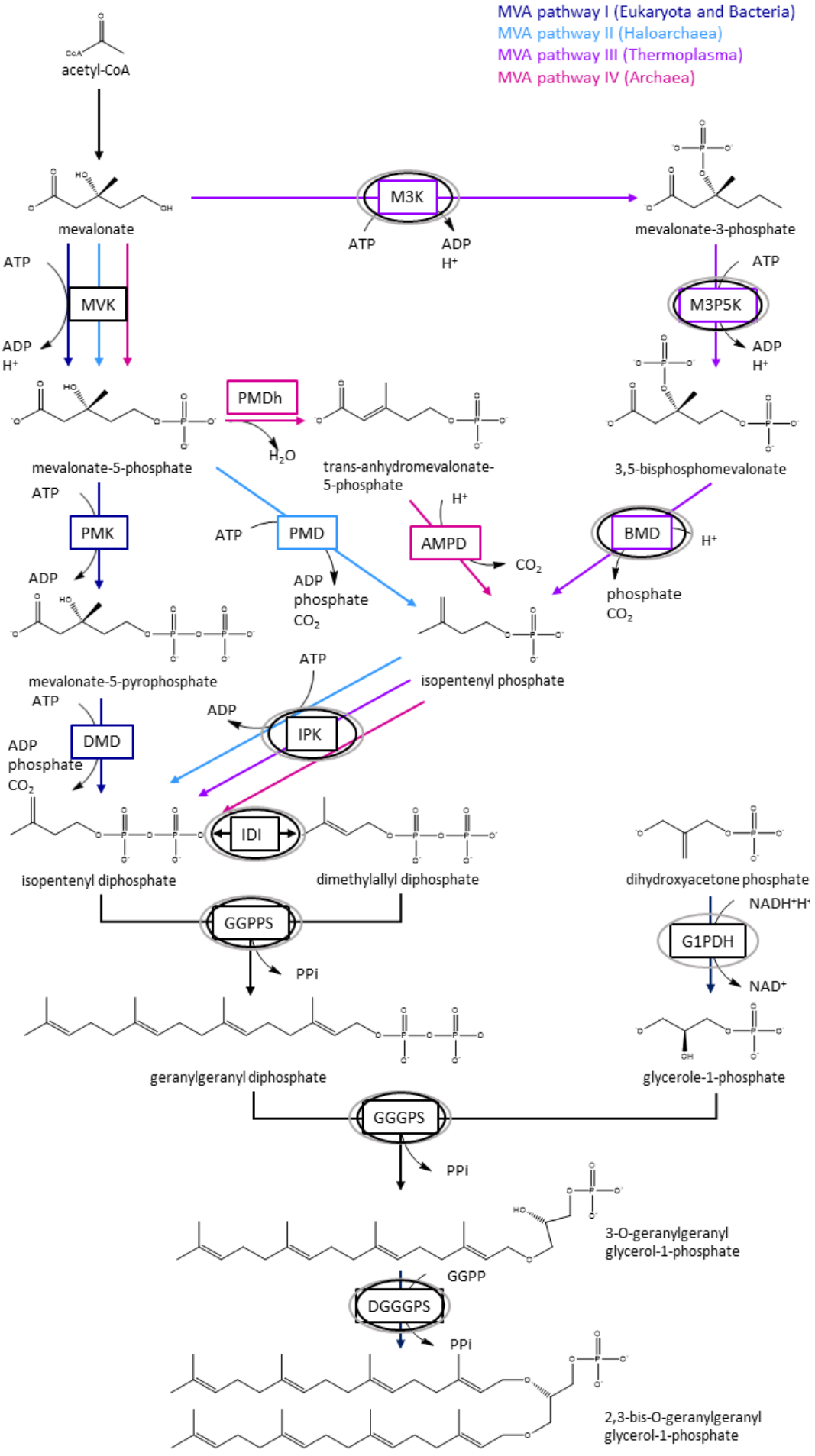


Figure 2: Schematic overview of mevalonate pathways and lipid metabolism. The different pathways are indicated with dark blue (MVA pathway I), light blue (MVA pathway II), violet (MVA pathway III) and magenta arrows (MVA pathway IV), respectively. Names of enzymes are boxed. Enzymes expressed in A_DKE and B_DKE are indicated with black and grey circles. Abbreviations are AMPD: anhydromevalonate phosphate decarboxylase; BMD: bisphosphomevalonate decarboxylase; DGGGPS: 2,3-bis-O-geranylgeranyl glycerol-1-phosphate synthase; DMD: diphosphomevalonate decarboxylase; GGGPS: 3-O-geranylgeranyl glycerol-1-phosphate synthase; GGPP: geranylgeranyl diphosphate; GGPPS: geranylgeranyl diphosphate synthase; G1PDH: glycerol-1-phosphate dehydrogenase; IDI: isopentenyl-diphosphate-delta-isomerase; IPK: isopentenyl phosphate kinase; MVK: mevalonate kinase; M3K: mevalonate 3-kinase; M3P5K: mevalonate 3-phosphate 5-kinase; PMD: phosphomevalonate decarboxylase; PMK: phosphomevalonate kinase.

The genome data (Supplementary Tables S3-S5) revealed that B_DKE uses variant-III of the MVA pathways first described in *Thermoplasma acidophilum* [29]. In particular, B_DKE encodes the three key enzymes mevalonate-3-kinase (arCOG02937), mevalonate-3-phosphate-5-kinase (COG02074), and mevalonate-3,5-bisphosphate-decarboxylase (arCOG02937), characterizing this pathway, while it lacks genes for a canonical mevalonate kinase (Supplementary Tables S3-S4) similar to other members of the acidophilic Thermoplasmatales (Vinokur *et al.*, 2016). Prenyltransferases found in the genome are farnesyl diphosphate synthase and geranyl-geranyl diphosphate synthase. Also, a G1P-DH (i.e. glycerol-1-phosphate dehydrogenase) could be identified. Genes encoding enzymes for the ether bond formation (i.e. GGGP and DGGGP synthase) and saturation of isoprenoids (i.e. geranylgeranyl reductases) are also encoded in the B_DKE genome (see Supplementary Table S5). In contrast, A_DKE has an incomplete mevalonate and archaeal lipid pathway (Supplementary Tables S3-S4). In particular, while an ancestor of A_DKE and some other Micrarchaeota are likely to have acquired three key genes of the variant-III mevalonate pathway from Thermoplasmatales archaea (i.e. mevalonate-3-kinase (arCOG02937), mevalonate-3-phosphate-5-kinase (COG02074), and mevalonate-3,5-bisphosphate-decarboxylase (arCOG02937)) (Figure 3), A_DKE lacks a homolog of the hydroxymethylglutaryl-CoA reductase (Supplementary Table

S3). Note, its genome does not provide any evidence for the presence of another variant of the mevalonate pathway (Supplementary Table S3). Furthermore, A_DKE does not appear to encode a G1PDH. One of the encoded geranylgeranyl reductase homologs of A_DKE, likely involved in lipid biosynthesis, also seems to be acquired by horizontal gene transfer (HGT) from Thermoplasmatales (arCOG00570) (Figure 3).

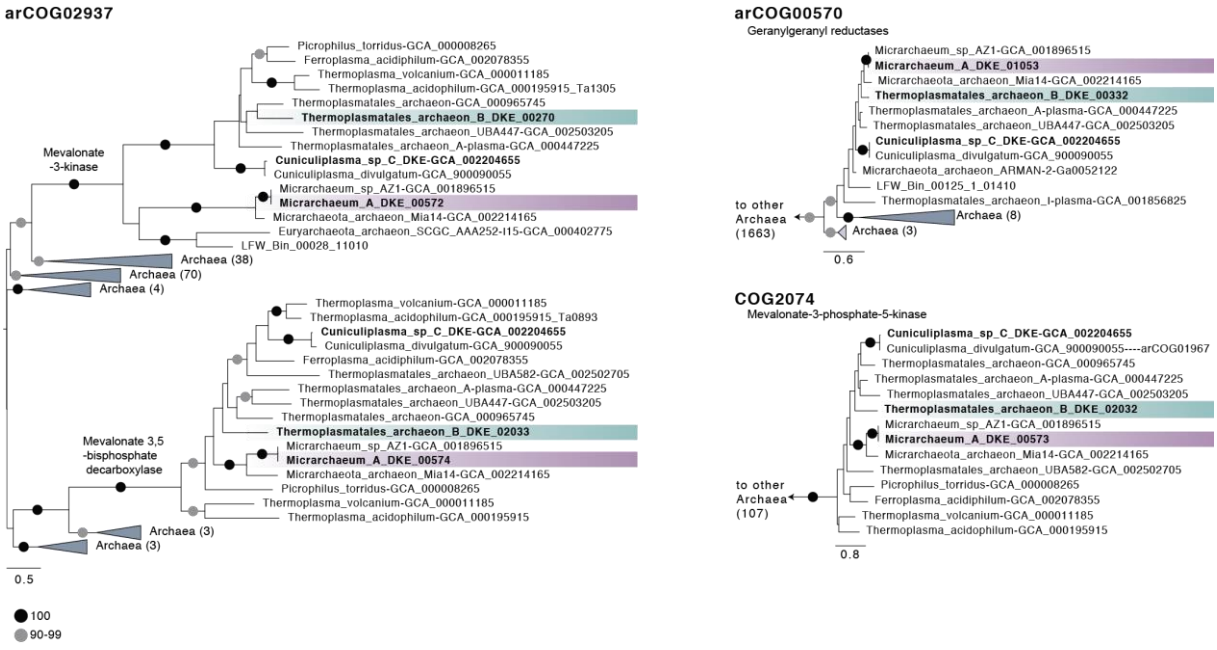
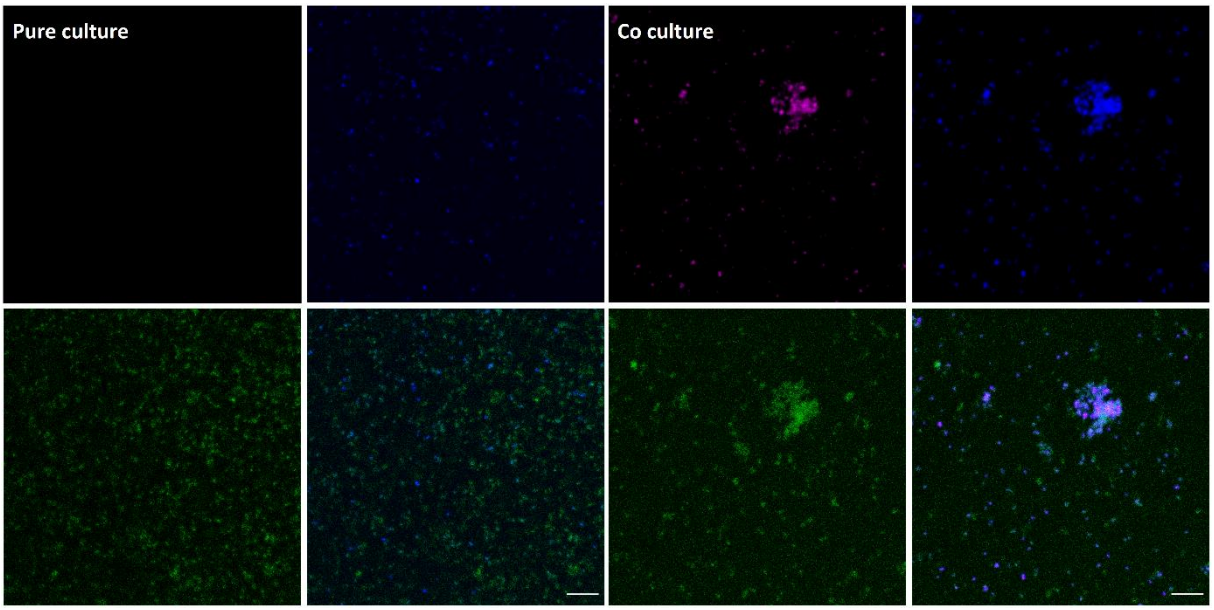


Figure 3. Schematic trees of three protein families with indications for gene transfers between acidophilic Micrarchaeota and Thermoplasmatales. The Maximum likelihood phylogenetic trees shown here included our archaeal backbone dataset (Supplementary Table S6) and were inferred for arCOG02937 (149 sequences with 304 amino acids), arCOG00570 (1877 sequences with 132 amino acids) and COG2074 (arCOG01968 and arCOG01967, 121 sequences with 171 amino acids) with the LG+C10+F+R model with an ultrafast bootstrap approximation run with 1000 replicates. Please note, that the inclusions of bacterial and eukaryotic homologs did not change the interpretation of our findings. Note that arCOG02937 and arCOG02074 represent the key enzymes of the Type-III mevalonate pathway characteristic of Thermoplasmatales. Only bootstrap support values above 90 are shown as indicated in the panel.

Together with our experimental data, the presence of an incomplete variant-III mevalonate and lipid biosynthesis pathways in A_DKE, indicates that this organism depends

297 on lipids or precursors thereof from its host, similar to what has been previously described in
298 the DPANN archaeon *N. equitans* [31] and likely other DPANN members such as
299 *Nanohaloarchaeum antarcticus* [16].
300
301 **Biofilm composition of pure and co-cultures**
302 As the isolation experiments and our previous results point towards the importance of
303 extracellular polymeric substances (EPS) for successful cultivation of A_DKE, we next
304 investigated the composition of the EPS matrix in the co-culture as compared to the pure culture
305 of B_DKE. To this end the glycoconjugates were analyzed with fluorescently labeled lectins,
306 and the signals were correlated to the individual cell type by CARD-FISH analysis (Figure 4).
307 Lectins are complex proteins, which bind specifically to carbohydrate structures [32]. In this
308 study, 70 different fluorescently labeled lectins, which represent all commercially available
309 lectins available so far (Supplementary Table S7), were used to analyze the EPS in pure and
310 co-cultures.
311



Lectin	Single-sugar binding specifity	Pure culture	Co-culture	
			B_DKE	A_DKE
AAL	α-Fucose	+	+	-
CA	Lactose>N-Acetyl-Galactosamine >Galactose and related sugars	++	++	-
GNA	Mannose	+	++	+
GS-I	Galactose, N-Acetyl-Galactosamine	+	+	-
HHA	Mannose	++	-	+
HPA	N-Acetyl-Galactosamine	++	+	+
IAA	not determined	++	++	-
PTA	Galactose, N-Acetyl-Galactosamine	+	-	-
RCA	ß-Galactose, Lactose	+	+	+
RPA	N-Acetyl-D-Galactosamine	++	++	-
SSA	α-N-Acetyl-Galactosamine	+	++	+
TKA	Galactose	-	++	-

Figure 4: Results of lectin staining of co-culture of A_DKE and B_DKE and B_DKE pure culture. Upper part shows an example of the microscopic analysis of the pure culture (left) and co-culture (right). The images show the results for staining with lectin CA (co-culture) and HHA (pure culture). Color allocation: B_DKE was stained with the general Archaea probe Arch915 (blue) which does not stain A_DKE. A_DKE was stained using the Micrarchaeota-specific ARMAN980 probe (red). Lectin staining is shown in green. Scale bars represent 10 µm. The table shows binding lectins, their carbohydrate-binding specificity, and the strength of the signal in pure and co-culture (with + as weak binding, ++ as binding, and - as no binding).

Among the tested lectins, only those specific to galactose- and mannose-related conjugates bound the extracellular matrix of the co-culture and of B_DKE cells in pure culture. Notably, lectins, IAA, HHA, and PTA seemed to discriminate between the extracellular matrix

of B_DKE in the presence or absence of A_DKE, suggesting a potential influence of A_DKE on the matrix chemistry or composition. Lectin staining of A_DKE cells was weak and only possible with some of the lectins that bound to B_DKE. This may reflect the inability of A_DKE to build carbohydrate polymers or the production of a sort of exotic polymer for which we did not have a lectin. It suggests that the detected signals on A_DKE are likely due to growth within the biofilm matrix of B_DKE.

Overall, the results indicate that B_DKE displays galactose and mannose on its cell surface and that these carbohydrates are also components of the co-culture EPS matrix. This is corroborated by the presence of transcriptionally expressed genes for metabolic pathways leading to UDP-glucose, UDP-galactose, GDP-mannose, UDP-N-acetylgalactosamine, and N-acetylglucosamine in the genome of B_DKE (see Supplementary Table S8 and Supplementary Figure S3).

Evidence for direct cell-cell interactions between A_DKE and B_DKE

Due to the pleomorphic morphology and great variability in cell size of members of the Thermoplasmatales including B_DKE, it was previously challenging to clearly distinguish symbiont and host cells on electron micrographs. Recently, [33] has shown that A_DKE cells are characterized by the presence of an S-layer that can be observed on electron micrographs of Platinum-Carbon shadowed samples. Hence, we conducted an electron microscopic study to test whether the two organisms indeed physically interact as indicated by our various analyses (see above). Electron micrographs revealed the attachment of several A_DKE to B_DKE cells, suggesting direct cell-cell interactions between these organisms, as was previously shown for *N. equitans* and *I. hospitalis* [21,22] (Figure 5). However, we also observed a large number of A_DKE cells that were not in contact with their potential host organism, which is in agreement with observations from microscopic images of CARD-FISH stained cultures. While we cannot exclude that this is (to some degree) a result of sample preparation, it is possible that growth in

the biofilm enables a more dynamic interaction between A_DKE and B_DKE than observed for *N. equitans* and *I. hospitalis* [21,22], as the risk of detaching from the host is mitigated by growth within the biofilm matrix. Moreover, A_DKE has a larger genome and in turn greater metabolic flexibility than *N. equitans* [34] and may in turn be less dependent on permanent attachment to host cells. Finally, we detected several unattached A_DKE cells in the process of cell division suggesting that A_DKE can store a sufficient amount of building blocks to divide without being in direct cell-cell contact with B_DKE.

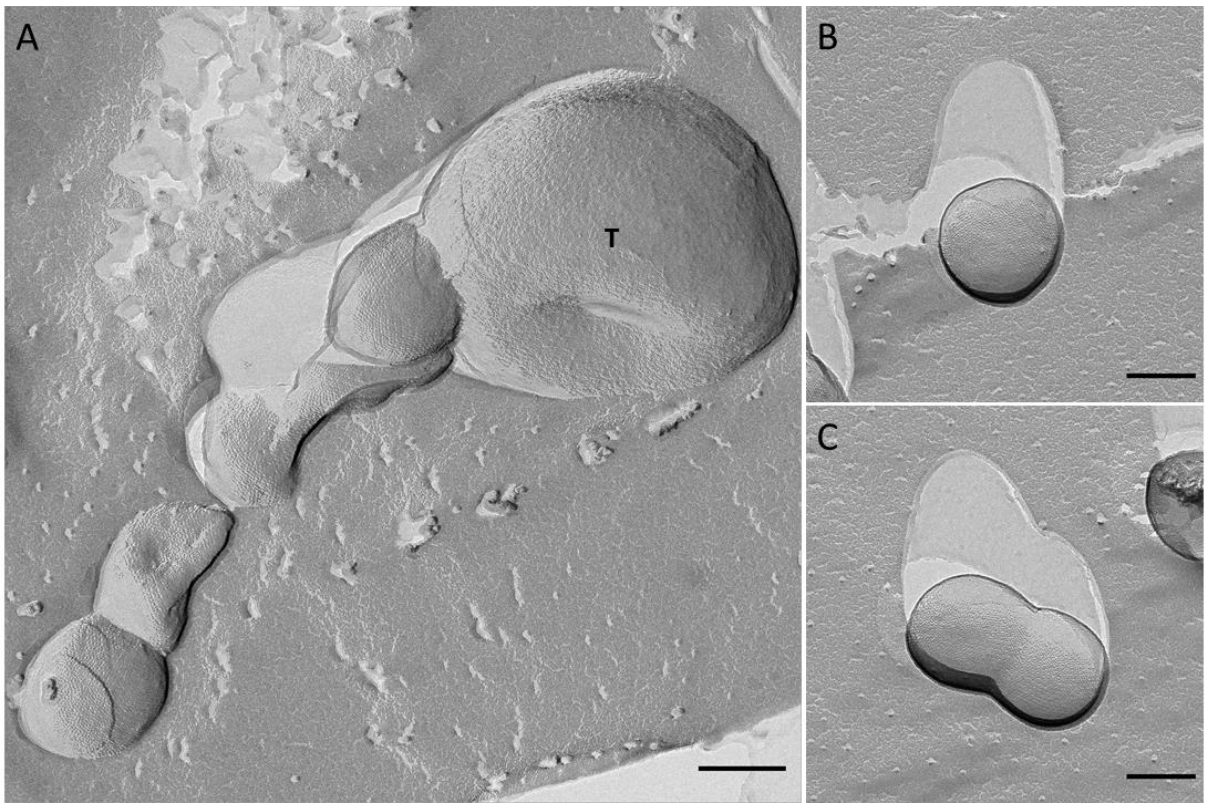


Figure 5: Electron micrographs of freeze-etched, Platinum-Carbon shadowed co-culture cells, containing '*Candidatus* Scheffleriplasma hospitalis' B_DKE and '*Candidatus* Micrarchaeum harzensis' A_DKE. A_DKE cells, displaying an S-Layer on their surface, were observed in physical contact with B_DKE cells (tagged with T) (A), as well as free living (B) and undergoing cell division (C). Scale bars equal 200 nm.

Conclusion

Previous attempts to purify members of the Micrarchaeota with their respective hosts yielded in relatively diverse enrichments or led to the disappearance of the symbiont after some time of incubation [14,18,19]. Here we used physiology informed strategies for deselecting against additional community members and show that selection for biofilm formation of the B_DKE host population was crucial for obtaining a stable co-culture. The detailed characterization of our co-culture in comparison with pure host cultures using for instance transcriptomic, metabolomic and lipid analyses indicate a specific regulatory response of B_DKE as a consequence of growth with A_DKE. These characterizations also suggest a dynamic physical interaction of the symbiont and host organisms that is likely crucial for the uptake of various metabolites and building blocks for among others membrane formation. The close cell-cell interactions between acidophilic Micrarchaeota and Thermoplasmatales may also provide a route for horizontal gene transfer among these DPANN symbionts and their hosts (Figure 4). The robust growth of our culture and the possibility to upscale the growth experiments to a fermenter scale provides the basis for prospective work aiming to unravel further insights into the details of the symbiosis between the acidophilic Micrarchaeota A_DKE and its Thermoplasmatales host. On a broader scale, it will be an important step to compare cell-cell interactions underlying this system to the various addition symbioses involving cultivated DPANN archaea and their hosts and establish unique and common characteristics [12–14,16,17].

Materials and Methods

Culturing conditions

All cultures were cultivated under anoxic conditions in a modified Picrophilus medium at a pH of 2.0 or 2.5 and at 22 °C as previously described [15]. Transfers of the cultures were conducted with 20 % of the pre-culture. The growth phase was assessed by following the reduction of ferric iron via the ferrozine assay [35]. Furthermore, the cultures were monitored regarding their activity via CARD-FISH (see below). The cultures typically reached the exponential growth phase after incubation for 2-8 weeks.

CARD-FISH and Lectin staining

Samples were fixed for 1 h in 4 % formaldehyde, washed twice in phosphate-buffered saline (PBS), and stored at -20 °C in 50:50 PBS/ethanol for CARD-FISH analysis or at 4 °C in 100 % PBS for lectin staining. The fixed cells were hybridized and further processed as described previously [15,36]. Labeling was conducted using HRP-labelled 16S rDNA probes TH1187 (Thermoplasmatales, GTA CTG ACC TGC CGT CGA C, 20 % formamide; [37]) and ARM980 (ARMAN, GCC GTC GCT TCT GGT AAT, 30 % formamide; [38]). For standard CARD-FISH staining, Alexa546 and Alexa488 fluorophores were used, and counterstaining was conducted with DAPI. CARD-FISH staining for lectin analysis was conducted using Alexa546 and Alexa647 fluorophores; see below for more details. Slides were visualized on a Leica DM 5500B microscope (objective lens 100×: HCX PL FLUOTAR, 1.4, oil immersion and objective lens 64x: HCX PL APO; eyepiece 10x: HC PLAN s (25) M), and images were taken with a Leica DFC 360 FX CCD camera and the corresponding Leica LAS AF software.

Lectin staining was conducted according to a protocol of Neu and Kuhlicke (2017). The positive lectin results were compared with CARD-FISH staining of the same slides, using a protocol of Bennke *et al.* (2013). CARD-FISH staining was performed as described above with the following modifications: 1) the low melting agarose and all ethanol washing steps were

omitted, as these would negatively affect the lectin staining [40]; 2) cells were not counterstained with DAPI. After fixation and staining via CARD-FISH, the samples were dried at 37 °C, and 100 µl lectin solution (0.1 mg/mL) was added and incubated for 30 min at room temperature in the dark. After washing with PBS solution, the slides were dried at 37 °C and mounted in the same embedding buffer as used for CARD-FISH. Please refer to Supplementary Table S7 for an overview of lectins used. Imaging was conducted as described above.

DNA/RNA isolation and quantitative PCR analysis

Isolation of DNA for quantitative (qPCR) analysis was conducted with the Invisorb Spin Forensic Kit following the manufacturer's instructions (Invitek, Berlin). DNA for Illumina and PacBio sequencing was isolated as described by Lo *et al.* (2017). RNA isolation and library preparation was conducted by IMGM laboratories GmbH using the RNeasy Micro Kit (Qiagen, Hilden) and TruSeq® Stranded total RNA LT kit according to the manufacturer`s instructions (illumina, Berlin). All sequencing analyses were conducted with samples of exponentially growing pure host cultures as well as symbiont-host co-cultures. The ratio of the Micrarchaeon A_DKE to the Thermoplasmatales archaeon B_DKE was calculated via qPCR using standard curves, as described in Krause *et al.* (2017). The primer sequences are listed in Supplementary Table S9.

Metagenome analysis

A SMRTbell™ template library was prepared according to the instructions from PacificBiosciences, Menlo Park, CA, USA, following the Procedure & Checklist – Greater Than 10 kb Template Preparation. Briefly, for preparation of 15 kB libraries, DNA was end-repaired and ligated overnight to hairpin adapters applying components from the DNA/Polymerase Binding Kit P6 from Pacific Biosciences, Menlo Park, CA, USA. Reactions were carried out according to the manufacturer's instructions. BluePippin™ Size-Selection to

436 greater than 4 kb was performed according to the manufacturer's instructions (Sage Science,
437 Beverly, MA, USA). Conditions for annealing of sequencing primers and binding of
438 polymerase to purified SMRTbell™ template were assessed with the Calculator in RS Remote,
439 PacificBiosciences, Menlo Park, CA, USA. One SMRT cell was sequenced on the PacBio RSII
440 (PacificBiosciences, Menlo Park, CA, USA), taking one 240-min movie. Libraries for
441 sequencing on the Illumina platform were prepared to apply the Nextera XT DNA Library
442 Preparation Kit (Illumina, San Diego, USA) with modifications according to Kishony *et al.* [42]
443 and sequenced on Illumina NextSeq™ 500 (Illumina, San Diego, USA).

444 Genome assembly was performed by applying the RS_HGAP_Assembly.3 protocol
445 included in the SMRT Portal version 2.3.0. The assembly revealed two major contigs.
446 Potentially misassembled artificial contigs with low coverage and included in other replicons
447 were removed from the assembly. Redundancies at the ends of the two major contigs allowed
448 them to be circularized. Replicons were adjusted to *smc* (chromosome partition protein Smc)
449 as the first gene. Error-correction was performed by mapping of the Illumina short reads onto
450 finished genomes using the Burrows-Wheeler Aligner bwa 0.6.2 in paired-end mode using
451 default settings [43] with subsequent variant and consensus calling using VarScan 2.3.6
452 (Parameters: mpileup2cns --min-coverage 10 --min-reads2 6 --min-avg-qual 20 --min-var-freq
453 0.8 --min-freq-for-hom 0.75 --p-value 0.01 --strand-filter 1 --variants 1 --output-vcf 1) [44].
454 Automated genome annotation was carried out using Prokka 1.8 [45].
455
456 **Genome annotations**
457 For further functional annotation, the protein files from the two complete genomes as well as
458 22 Micrarchaeota and 11 Thermoplasmatales reference genomes (Supplementary Table S6)
459 were compared against several databases, including the Archaeal Clusters of Orthologous
460 Genes (arCOGs [46]; version from 2018), the KO profiles from the KEGG Automatic Annotation
461 Server (KAAS [47]; downloaded April 2019), the Pfam database ([48] Release 31.0), the TIGRFAM

462 database ([49] Release 15.0), the Carbohydrate-Active enZymes (CAZy) database ([50] downloaded
463 from dbCAN2 in September 2019), the Transporter Classification Database (TCDB [51];
464 downloaded in November 2018), the hydrogenase database (HydDB [52]; downloaded in
465 November 2018) and NCBI_nr (downloaded in November 2018). Hmmsearch v3.1b298 was
466 used to read HMM profiles of the ArCOG, PFAM, TIGRFAM and CAZyme databases and
467 search against a protein database (settings: hmmsearch <hmmfile> <seqdb> -E 1e-4 [53]). The
468 best hit for each protein was selected based on the highest e-value and bitscore by using a
469 custom script (hmmsearchTable, available at https://zenodo.org/record/3839790 [7]). BlastP was
470 used with TCBD, HydDB and NCBI_nr as input databases and the protein sequences as query
471 (settings: -evalue 1e-20 -outfmt 6). Additionally, all proteins were scanned for protein domains
472 using InterProScan (v5.29-68.0; settings: --iprlookup –goterms [54]). For InterProScan we report
473 multiple hits corresponding to the individual domains of a protein using a custom script
474 (parse_IPRdomains_vs2_GO_2.py) (Supplementary Tables S3 and S4). All custom scripts are
475 available at https://zenodo.org/record/3839790 [7].
476
477 **Phylogenetic analyses**

478 We performed phylogenetic analyses of membrane lipid biosynthetic proteins, whenever
479 homologs of relevant arCOGs were present in both A_DKE and B_DKE to assess the extent of
480 horizontal gene transfer (HGT) affecting these proteins (Supplementary Table S5). In
481 particular, we extracted homologs of corresponding arCOGs for all of these proteins from
482 A_DKE, B_DKE, a reference set of 566 archaeal genomes (archaea-only analysis) as well as
483 from an additional of 3020 bacterial and 100 eukaryotic genomes (universal analysis)
484 (Supplementary Table S6). The reference genomes were annotated as described above. For the
485 archaea-only analysis, the individual homologs for each protein family were aligned using
486 MAFFT L-INS-i v7.407 (settings: --reorder [55]), trimmed with BMGE v1.12 (settings: -t AA -
487 m BLOSUM30 -h 0.55 [56]). Subsequently, phylogenetic trees were inferred using IQ-TREE

(v1.6.10, settings: -m LG+C10+F+R -wbtl -bb 1000 -bnni [57]). For the universal analysis, MAFFT L-INS-i v7.407 and MAFFT v7.407 were used to align protein families with Less/equal (≤) or more (>) than 1000 homologs, respectively. BMGE v1.12 was used for trimming all alignments (settings: -t AA -m BLOSUM30 -h 0.55 [56]) and phylogenetic trees were inferred using IQ-TREE (v1.6.10, settings: -m LG+C10+F+R -wbtl -bb 1000 -bnni [57]). Due to the large number of sequences affiliating with arCOG00570 when including bacterial and eukaryotic homologs (i.e. 13381), sequences for this protein family were aligned using MAFFT v7.407, trimmed with TrimAL (v1.2rev59, settings: -gappyout [58]). Sequences with => 90 % gaps were removed using a custom script (faa_drop.py) and used for a phylogenetic analysis with FastTree (v2.1.10, settings: -lg -gamma).

Transcriptomic analysis

Sequencing of RNA samples of three replicates of the pure culture and four replicates of the co-culture was performed on an Illumina NextSeq® 500 NGS system using 2 x 75 bp paired-end read chemistry. All samples were taken from cultures in exponential growth phase. Transcriptomic analyses were performed using Kallist v0.45.0 [59] and compared to the reference completed genomes (see above). Differential expression was assessed by using the R package DESeq2 (1.24.0; Love *et al.*, 2014). Log2fold change shrinkage for normalization was calculated by the ashr program [61].

Lipid analyses

For the analysis of membrane lipids, cells were derived from 20 mL pure and co-culture in the exponential growth phase. Samples were taken for monitoring of cells via CARD-FISH prior to filtering (Supplementary Figure S4). Cells were filtered onto 0.3 µm pore size 47 mm diameter glass fiber filters (GF75, Advantec MFS, Inc, CA, USA). Total lipids were extracted from the freeze-dried glass-fiber filters using a modified Bligh and Dyer method [62], as described

514 earlier [63]. The extracts were dried under nitrogen and split into two aliquots, one left untreated
515 and another hydrolyzed with 1.5 M HCl in methanol by reflux at 130 °C for 2 h to remove the
516 headgroups from the archaeal intact polar lipids (IPL) and release the core lipids (CLs). The pH
517 was adjusted to 7 by adding 2 M KOH/MeOH (1:1 v/v) and, after the addition of water to a
518 final 1:1 (v/v) ratio of H_2O-MeOH, extracted three times with dichloromethane (DCM). The
519 DCM fractions were collected and dried over sodium sulfate. The dried samples were dissolved
520 in hexane–2-propanol (99:1, vol/vol) and filtered over a 0.45-µm polytetrafluoroethylene filter.
521 The extracts after acid-hydrolysis contained the IPL-derived CLs plus CLs, while the non-
522 hydrolyzed extracts consisted only of the CLs. Extracts were analyzed by UHPLC–atmospheric
523 pressure chemical ionization (APCI) MS for archaeal CLs, including archaeol (diether, C_{20}
524 isoprenoid chains) and glycerol dialkyl glycerol tetraether (GDGTs, tetraether, C_{40} side chain),
525 according to Hopmans *et al*. (2016), with some modifications. Briefly, the analysis was
526 performed on an Agilent 1260 UHPLC coupled to a 6130 quadrupole MSD in selected ion
527 monitoring (SIM) mode. This allowed the detection of GDGTs with 0 to 4 cyclopentane
528 moieties, crenarchaeol as well as archaeol. The separation was achieved on two UHPLC silica
529 columns (BEH HILIC columns, 2.1 × 150 mm, 1.7 µm; Waters) in series, fitted with a 2.1 × 5-
530 mm pre-column of the same material (Waters) and maintained at 30 ˚C. Archaeal CLs were
531 eluted isocratically for 10 min with 10 % B, followed by a linear gradient to 18 % B in 20 min,
532 then a linear gradient to 100 % B in 20 min, where A is hexane and B is hexane:isopropanol
533 (9:1). The flow rate was 0.2 mL/min. Total run time was 61 min with a 20 min re-equilibration.
534 Source settings were identical to Schouten *et al*. (2007). The typical injection volume was 10
535 µl of a 1 mg/mL solution (weighted dried Bligh and Dyer extract dissolved in
536 hexane:isopropanol (99:1, v/v ratio). The m/z values of the protonated molecules of archaeol
537 and GDGTs were monitored. GDGTs were quantified by adding a C_{46} GTGT internal standard
538 [66]. A response factor derived from an archaeol:GDGT-0 standard (1:1) was used to correct for
539 the difference in ionization between archaeol and GDGTs.

The Bligh and Dyer extract (non-hydrolyzed) and the acid-hydrolyzed Bligh and Dyer extract were also analyzed using ultra-high-performance liquid chromatography coupled to positive ion atmospheric pressure chemical ionization/Time-of-Flight mass spectrometry (UHPLC-APCI/ToFMS) on an Agilent 1290 Infinity II UHPLC, equipped with an automatic injector, coupled to a 6230 Agilent TOF MS and Mass Hunter software. This additional analysis was performed to detect other archaeal lipids that were not included in the SIM method on the 6130 quadrupole MSD mentioned above. Separation of the archaeal lipids was achieved according to Hopmans *et al*. (2016) with some modifications using two silica BEH HILIC columns in series (2.1 × 150 mm, 1.7 µm; Waters) at a temperature of 25 °C. The injection volume was 10 µL. Compounds were isocratically eluted with 90 % A and 10 % B for the first 10 min, followed by a gradient to 18 % B in 15 min, a gradient to 30 % B in 25 min, and a linear gradient to 100 % B in 30 min. A = hexane and B = hexane/isopropanol (9:1, v/v) and the flow rate was 0.2 mL/min. The conditions for the APCI source were identical to Schouten *et al*. (2007) and Hopmans *et al*. (2016). Also, the fragmentor was set at 300 V. The ToFMS was operated in extended dynamic range mode (2 GHz) with a scan rate of 2 Hz. We assessed archaeal lipid distributions by monitoring m/z 600 to 1400. Archaeal lipids were identified by searching within 10 ppm mass accuracy for relevant [M+H]+ signals.

Metabolome analysis

A pure culture of B_DKE and a co-culture of A_DKE and B_DKE were inoculated as described above for 42 (pure culture) or 49 days (co-culture) until all available ferric iron was reduced and stationary phase was reached. Experiments were performed in triplicates. For metabolomic examination, 1 mL culture was sampled and stored at -80 °C until further analyses. Samples were taken every 7 days, along with samples for ferrous iron quantification to estimate growth, and samples for DNA extraction and CARD-FISH for further detection of A_DKE and B_DKE. Due to the low pH of the culture medium, 500 µl samples were amended by inducing sulfur

566 precipitation through the addition of a spatula tip of $CaCO_3$ to each sample, mixing for 5 min
567 at 2000 rpm at room temperature. This treatment also led to cell lysis so that the analysis
568 included also intracellular metabolites. After centrifugation for 5 min at 14000 rpm at room
569 temperature, 50 µl of the supernatant was transferred to glass vials and dried under vacuum at
570 4 °C. Dried samples were stored at -80 °C until further analysis.

571 Online metabolite derivatization was performed using a Gerstel MPS2 autosampler
572 (Muehlheim, Germany). Dried metabolites were dissolved in 15 µl of 2 % methoxyamine
573 hydrochloride in pyridine at 40 °C under shaking. After 90 min, an equal volume of N-Methyl-
574 N-(trimethylsilyl)trifluoroacetamide (MSTFA) was added and held for 30 min at 40 °C. One µl
575 of the sample was injected into an SSL injector at 270 °C in splitless mode. Gas
576 chromatography/mass spectrometry (GC/MS) analysis was performed using an Agilent 7890A
577 GC equipped with a 30-m DB-35MS # 5-m Duraguard capillary column. Helium was used as
578 carrier gas at a flow rate of 1.0 mL/min. The GC oven temperature was held at 100 °C for 2
579 min and increased to 300 °C at 10 K/min. After 3 min, the temperature was increased to 325
580 °C. The GC was connected to an Agilent 5975C inert XL MSD, operating under electron
581 ionization at 70 eV. The MS source was held at 230 °C and the quadrupole at 150 °C. The total
582 run time of one sample was 60 min. All GC/MS chromatograms were processed by using the
583 Metabolite Detector software [67].
584

585 **Structural analysis by electron microscopy**

586 For freeze-etching, both a pure culture of B_DKE and a co-culture of A_DKE and B_DKE
587 were concentrated by centrifugation (3,000 x g). The concentrated cell pellet (1.5 µL) was
588 applied onto a gold carrier, frozen in liquid nitrogen, and transferred into a freeze-etching device
589 (CFE-50, Cressington, Watford, UK; $p < 10^{-5}$ mbar). At T=176 K, samples were fractured using
590 a cold knife (T=90 K); after sublimation of about 400 nm of surface water, the samples were
591 shadowed with Platinum-Carbon at an angle of 45 degrees (1.5 nm), and an additional layer of
592 pure Carbon (about 15 nm; both by electron-beam evaporation). Replicas were cleaned for 15
593 hours on 70 % H2SO4, washed three times on bidistilled water, taken up on 700 mesh (hex)
594 grids and air-dried. For electron microscopy analysis at 200 kV, a transmission electron
595 microscope JEM-2100F (JEOL GmbH, Freising, Germany), equipped with a F416 CMOS
596 camera (TVIPS, Gauting, Germany) under control of SerialEM v. 3.8 [68] was used.

597

598 **Accession of data**

599 The genome sequences including annotations have been deposited at NCBI Genbank under
600 Accession Numbers CP060530 and CP060531. Raw reads of transcriptomic data are available
601 under SRA files SRX8933312-SRX8933318. All raw files phylogenetic trees can be found in
602 a repository (doi: 10.5281/zenodo.4725436).

Acknowledgements

The authors thank Carola Berg for excellent technical assistance.

References

1. Huber, H. *et al.* A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**, 63–67 (2002).

2. Rinke, C. *et al.* Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).

3. Castelle, C. J. *et al.* Genomic expansion of domain Archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).

4. Probst, A. J. *et al.* Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. *Nat. Microbiol.* **3**, 328–336 (2018).

5. Baker, B. J. *et al.* Enigmatic, ultrasmall, uncultivated Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 8806–8811 (2010).

6. Probst, A. J. *et al.* Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. *Nat. Commun.* **5**, 5497 (2014).

7. Dombrowski, N. *et al.* Undinarchaeota illuminate DPANN phylogeny and the impact of gene transfer on archaeal evolution. *Nat. Commun.* **11**, 3939 (2020).

8. Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 (2018).

9. Castelle, C. J. *et al.* Protein family content uncovers lineage relationships and bacterial pathway maintenance mechanisms in DPANN archaea. *bioRxiv* 2021.01.12.426361 (2021) doi:10.1101/2021.01.12.426361.

10. Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P. & Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* **366**, (2019).

11. Ghuneim, L.-A. J., Jones, D. L., Golyshin, P. N. & Golyshina, O. V. Nano-sized and filterable Bacteria and Archaea: Biodiversity and function. *Front. Microbiol.* **9**, 1971 (2018).

632 12. Wurch, L. *et al.* Genomics-informed isolation and characterization of a symbiotic
633 Nanoarchaeota system from a terrestrial geothermal environment. *Nat. Commun.* **7**, 1–
634 10 (2016).

635 13. St. John, E. *et al.* A new symbiotic nanoarchaeote (*Candidatus* Nanoclepta minutus) and
636 its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring.
637 *Syst. Appl. Microbiol.* **42**, 94–106 (2018).

638 14. Golyshina, O. V *et al.* Mysterious 'ARMAN' archaea depend on association with
639 euryarchaeal host in culture *in situ*. *Nat. Commun.* 1–11 (2017) doi:10.1038/s41467-017-
640 00104-7.

641 15. Krause, S., Bremges, A., Münch, P. C., McHardy, A. C. & Gescher, J. Characterisation
642 of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* **7**, 1–13 (2017).

643 16. Hamm, J. N. *et al.* Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc.*
644 *Natl. Acad. Sci.* **116**, 14661 LP – 14670 (2019).

645 17. La Cono, V. *et al.* Symbiosis between nanohaloarchaeon and haloarchaeon is based on
646 utilization of different polysaccharides. *Proc. Natl. Acad. Sci.* **117**, 20223 LP – 20234
647 (2020).

648 18. Chen, L.-X. *et al.* Metabolic versatility of small Archaea Micrarchaeota and
649 Parvarchaeota. *ISME J.* 1 (2017) doi:10.1038/s41396-017-0002-z.

650 19. Golyshina, O. V *et al.* Diversity of 'Ca. Micrarchaeota' in Two Distinct Types of Acidic
651 Environments and Their Associations with Thermoplasmatales. *Genes (Basel).* **10**, 1–11
652 (2019).

653 20. Golyshina, O. V *et al.* The novel extremely acidophilic, cell-wall-deficient archaeon
654 *Cuniculiplasma divulgatum* gen. nov., sp. nov. represents a new family,
655 Cuniculiplasmataceae fam. nov., of the order Thermoplasmatales. *Int. J. Syst. Evol.*
656 *Microbiol.* **66**, 332–340 (2016).

657 21. Jahn, U. *et al. Nanoarchaeum equitans* and *Ignicoccus hospitalis*: New insights into a
658 unique, intimate association of two Archaea. *J. Bacteriol.* **190**, 1743 LP – 1750 (2008).

659 22. Heimerl, T. *et al.* A complex endomembrane system in the Archaeon *Ignicoccus*
660 *hospitalis* tapped by *Nanoarchaeum equitans*. *Front. Microbiol.* **8**, 1072 (2017).

661 23. Jarrell, K. F. & Albers, S.-V. The archaellum: an old motility structure with a new name.
662 *Trends Microbiol.* **20**, 307–312 (2012).

663 24. Golyshina, O. V. *et al.* Biology of Archaea from a novel family Cuniculiplasmataceae
664 (Thermoplasmata) ubiquitous in hyperacidic environments. *Sci. Rep.* **6**, 39034 (2016).

665 25. Yelton, A. P. *et al.* Comparative genomics in acid mine drainage biofilm communities
666 reveals metabolic and structural differentiation of co-occurring Archaea. *BMC Genomics*
667 **14**, 485 (2013).

668 26. Kopp, D., Willows, R. & Sunnar, A. Characterisation of the first archaeal mannonate
669 dehydratase from *Thermoplasma acidophilum* and its potential role in the catabolism of
670 D-mannose. *Catalysts* 1–14 (2019).

671 27. Koga, Y. & Morii, H. Biosynthesis of ether-type polar lipids in Archaea and evolutionary
672 considerations. *Microbiol. Mol. Biol. Rev.* **71**, 97–120 (2007).

673 28. Jain, S., Caforio, A. & Driessen, A. Biosynthesis of archaeal membrane ether lipids.
674 *Front. Microbiol.* **5**, 641 (2014).

675 29. Vinokur, J. M., Korman, T. P., Cao, Z. & Bowie, J. U. Evidence of a novel mevalonate
676 pathway in archaea. *Biochemistry* **53**, 4161–4168 (2014).

677 30. Vinokur, J. M., Cummins, M. C., Korman, T. P. & Bowie, J. U. An adaptation to life in
678 acid through a novel mevalonate pathway. *Sci. Rep.* **6**, 1–11 (2016).

679 31. Jahn, U., Summons, R., Sturt, H., Grosjean, E. & Huber, H. Composition of the lipids of
680 *Nanoarchaeum equitans* and their origin from its host *Ignicoccus sp.* strain KIN4/I. *Arch.*
681 *Microbiol.* **182**, 404–413 (2004).

682 32. Neu, T. R., Swerhone, G. D. W. & Lawrence, J. R. Assessment of lectin-binding analysis
683 for in situ detection of glycoconjugates in biofilm systems. *Microbiology* **147**, 299–313
684 (2001).

685 33. Gfrerer, S. *et al.* Micrarchaeota are covered by a proteinaceous S-Layer. Unpublished

686 34. Waters, E. *et al.* The genome of *Nanoarchaeum equitans*: Insights into early archaeal
687 evolution and derived parasitism. *Proc. Natl. Acad. Sci.* **100**, 12984 LP – 12988 (2003).

688 35. Stookey, L. L. Ferrozine - a new spectrophotometric reagent for iron. *Anal. Chem.* **42**,
689 779–781 (1970).

690 36. Pernthaler, A., Pernthaler, J. & Amann, R. Sensitive multi-color fluorescence *in situ*
691 hybridization for the identification of environmental microorganisms. *Mol. Microb.*
692 *Ecol. Man.* **3**, 711–726 (2004).

693 37. Bond, P. L. & Banfield, J. F. Design and performance of rRNA targeted oligonucleotide
694 probes for *in situ* detection and phylogenetic identification of microorganisms inhabiting
695 acid mine drainage environments. *Microb. Ecol.* **41**, 149–161 (2001).

696 38. Baker, B. J. *et al.* Lineages of acidophilic Archaea revealed by community genomic
697 analysis. *Science* **314**, 1933–1935 (2006).

698 39. Neu, T. & Kuhlicke, U. Fluorescence lectin bar-coding of glycoconjugates in the
699 extracellular matrix of biofilm and bioaggregate forming microorganisms.
700 *Microorganisms* **5**, 5 (2017).

701 40. Bennke, C. M., Neu, T. R., Fuchs, B. M. & Amann, R. Mapping glycoconjugate-
702 mediated interactions of marine Bacteroidetes with diatoms. *Syst. Appl. Microbiol.* **36**,
703 417–425 (2013).

704 41. Lo, I. *et al.* Strain-resolved community proteomics reveals recombining genomes of
705 acidophilic Bacteria. *Nature* **446**, 537—541 (2007).

706 42. Baym, M. *et al.* Inexpensive multiplexed library preparation for megabase-sized
707 genomes. *PLoS One* **10**, 1–15 (2015).

708 43. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler
709 transform. *Bioinformatics* **26**, 589–595 (2010).

710 44. Wilson, R. K. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery
711 in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).

712 45. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
713 2069 (2014).

714 46. Makarova, K., Wolf, Y. & Koonin, E. Archaeal Clusters of Orthologous Genes
715 (arCOGs): An update and application for analysis of shared features between

716 Thermococcales, Methanococcales, and Methanobacteriales. *Life* **5**, 818–840 (2015).

717 47. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM
718 and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).

719 48. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–
720 D141 (2004).

721 49. Haft, D. H., Selengut, J. D. & White, O. The TIGRFAMs database of protein families.
722 *Nucleic Acids Res.* **31**, 371–373 (2003).

723 50. Yin, Y. *et al.* dbCAN: a web resource for automated carbohydrate-active enzyme
724 annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).

725 51. Saier, M. H., Tran, C. V & Barabote, R. D. TCDB: the Transporter Classification
726 Database for membrane transport protein analyses and information. *Nucleic Acids Res.*
727 **34**, D181–D186 (2006).

728 52. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: A web tool for hydrogenase
729 classification and analysis. *Sci. Rep.* **6**, 34212 (2016).

730 53. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence
731 similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).

732 54. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
733 *Bioinformatics* **30**, 1236–1240 (2014).

734 55. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid
735 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**,
736 3059–3066 (2002).

737 56. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a
738 new software for selection of phylogenetic informative regions from multiple sequence
739 alignments. *BMC Evol. Biol.* **10**, 210 (2010).

740 57. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and
741 effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol.*
742 *Biol. Evol.* **32**, 268–274 (2015).

743 58. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
744 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
745 (2009).

746 59. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
747 quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

748 60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
749 for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

750 61. Stephens, M. False discovery rates: A new deal. *Biostatistics* **18**, 275–294 (2017).

751 62. Bligh, E. G., Dyer, W. J. A rapid method of total lipid extraction and purification. *Can.*
752 *J. Biochem. Physiol.* **37**, 911–917 (1959).

753 63. Lengger, S. K. *et al.* Intact polar and core glycerol dibiphytanyl glycerol tetraether lipids
754 in the Arabian Sea oxygen minimum zone. Part II: Selective preservation and
755 degradation in sediments and consequences for the TEX86. *Geochim. Cosmochim. Acta*
756 **98**, 244–258 (2012).

757 64. Hopmans, E. C., Schouten, S. & Sinninghe Damsté, J. S. The effect of improved
758 chromatography on GDGT-based palaeoproxies. *Org. Geochem.* **93**, 1–6 (2016).

759 65. Schouten, S., Huguet, C., Hopmans, E. C., Kienhuis, M. V. M. & Sinninghe Damsté, J.
760 S. Analytical methodology for TEX86 paleothermometry by high-performance liquid
761 chromatography/atmospheric pressure chemical ionization-mass spectrometry. *Anal.*
762 *Chem.* **79**, 2940–2944 (2007).

763 66. Huguet, C. *et al.* An improved method to determine the absolute abundance of glycerol
764 dibiphytanyl glycerol tetraether lipids. *Org. Geochem.* **37**, 1036–1041 (2006).

765 67. Hiller, K. *et al.* Metabolite detector: Comprehensive analysis tool for targeted and
766 nontargeted GC/MS based metabolome analysis. *Anal. Chem.* **81**, 3429–3439 (2009).

767 68. Mastronarde, D. N. Automated electron microscope tomography using robust prediction
768 of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

769