Horizontal acquisition of a patchwork Calvin cycle by symbiotic and free-living Campylobacterota (formerly Epsilonproteobacteria)

Authors: Adrien Assié[1,§,$,*], Nikolaus Leisch[1*], Dimitri V. Meier[1#], Harald Gruber-Vodicka[1], Halina E. Tegetmeyer[1,9], Anke Meyerdirks[1], Manuel Kleiner[2,3], Tjorven Hinzke[2,4,5], Samantha Joye[6], Matthew Saxton[6], Nicole Dubilier[1,7,$] and Jillian M. Petersen[1,8,$]

1 Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany.
2 Department of Geoscience, University of Calgary, Calgary, 2500 University Drive Northwest, Alberta T2N 1N4, Canada
3 Department of Plant and Microbial Biology, North Carolina State University, Raleigh 27695, North Carolina, USA
4 Department of Pharmaceutical Biotechnology, University of Greifswald, Institute of Pharmacy, Greifswald D-17489, Germany
5 Institute of Marine Biotechnology, Greifswald D-17489, Germany
6 Department of Marine Sciences, The University of Georgia, Room 159, Marine Sciences Bldg. Athens, GA 30602-3636
7 MARUM - Zentrum für Marine Umweltwissenschaften, University of Bremen, Leobener Str. 2, 28359 Bremen, Germany
8 Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.
9 Center for Biotechnology, Bielefeld University, Universitaetsstrasse 27, 33615 Bielefeld, Germany
§ Current address: Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, United States.
# Current address: Division of Microbial Ecology, Department of Microbiology and Ecosystem Science, Research Network Chemistry Meets Microbiology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria.

* these authors contributed equally

$ Corresponding authors:

Adrien Assié	Adrien.Assie@bcm.edu	+1 713 1819156
Nicole Dubilier	ndubilie@mpi-bremen.de	+49 421 2028-932
Jillian M. Petersen	petersen@microbial-ecology.net	+43 1 4277 76606

Abstract

Although the majority of known autotrophs use the Calvin-Benson-Bassham (CBB) cycle for carbon fixation, all currently described autotrophs from the Campylobacterota (previously Epsilonproteobacteria) use the reductive tricarboxylic acid cycle (rTCA) instead. We discovered campylobacterotal epibionts ("*Candidatus* Thiobarba") of deep-sea mussels that have acquired a complete CBB cycle and lost key genes of the rTCA cycle. Intriguingly, the phylogenies of campylobacterotal CBB genes suggest they were acquired in multiple transfers from Gammaproteobacteria closely related to sulfur-oxidizing endosymbionts associated with the mussels, as well as from Betaproteobacteria. We hypothesize that "*Ca.* Thiobarba" switched from the rTCA to a fully functional CBB cycle during its evolution, by acquiring genes from multiple sources, including co-occurring symbionts. We also found key CBB cycle genes in free-living Campylobacterota, suggesting that the CBB cycle may be more widespread in this phylum than previously known. Metatranscriptomics and metaproteomics confirmed high expression of CBB cycle genes in mussel-associated "*Ca.* Thiobarba". Direct stable isotope fingerprinting showed that "*Ca.* Thiobarba" has typical CBB signatures, additional evidence that it uses this cycle for carbon fixation. Our discovery calls into question current assumptions about the distribution of carbon fixation pathways across the tree of life, and the interpretation of stable isotope measurements in the environment.

Introduction

All life on earth is based on carbon fixation, and its molecular machinery is increasingly becoming a focus of biotechnology and geo-engineering efforts due to its potential to improve crop yields and sequester carbon dioxide from the atmosphere[1]. Seven carbon fixation pathways have evolved in nature, and one purely synthetic pathway runs in vitro[2–4]. Of the six natural pathways, the Calvin-Benson-Bassham (CBB) cycle was the first discovered, and is believed to be the most widespread[5–7]. The CBB cycle is used by a diverse array of organisms throughout the tree of life, including plants and algae, cyanobacteria, and autotrophic members of the Alpha-, Beta- and Gammaproteobacteria. Its key enzyme, the ribulose bisphosphate carboxylase/oxygenase (RuBisCO) is thought to be the most abundant, as well as one of the most ancient enzymes on Earth[8,9].

The reductive tricarboxylic acid (rTCA) cycle was the second described carbon fixation pathway[10]. In short, it is a reversal of the energy-generating and oxidative TCA cycle. Instead of oxidizing acetyl-CoA and generating ATP and reducing equivalents, it reduces CO_2 at the expense of ATP and reducing equivalents[2,7,10]. Most of the enzymes are shared with the TCA cycle, except for those that catalyze irreversible reactions in the TCA, such as citrate synthase, which is catalyzed by ATP citrate lyase in the rTCA. However, given sufficiently high reactant to product ratios and enzyme concentrations, the citrate synthase reaction can be reversed to run the TCA cycle reductively, without any additional enzymes[11,12]. The rTCA pathway is widely distributed in nature, and has been described in diverse lineages of anaerobes and microaerobes such as the Chlorobi, Aquificae, Nitrospirae and is also commonly observed among the Proteobacteria, including the Deltaproteobacteria and the Campylobacterota, (formerly Epsilonproteobacteria)[13,14]. It is particularly prominent in the Campylobacterota, as all previously described autotrophic members of this class use the rTCA pathway for CO_2 fixation[2,13] .

Carbon fixation by chemoautotrophic microorganisms forms the basis of entire ecosystems at deep-sea hydrothermal vents and cold seeps[15,16]. Most of this carbon is fixed either via the CBB cycle, used by many gammaproteobacterial autotrophs, or the rTCA cycle, used by campylobacterotal autotrophs. This difference is reflected by the different niches colonized by these organisms at hydrothermal vents and seeps, with Gammaproteobacteria typically dominating habitats with higher oxygen and lower sulfide concentrations where the CBB cycle would be more efficient, and Campylobacterota typically thriving at lower oxygen and higher sulfide concentrations where the rTCA cycle could provide a selective advantage[17–23]. Experimental studies have linked substrate preferences in cultured Gammaproteobacteria and Campylobacterota to these ecological distributions[24–26]. Symbiotic invertebrates at hydrothermal vents and cold seeps associate with either gammaproteobacterial or campylobacterotal endosymbionts, which they rely on for most of their nutrition[27,28]. Some vent and seep invertebrates associate with both gammaproteobacterial and campylobacterotal symbionts simultaneously, which raises the question of how these co-occurring symbionts with differing habitat preferences can both be provided with suitable conditions[27,29,30].

Bathymodiolin mussels, a subfamily of mytilid bivalves, are found at most hydrothermal vents and cold seeps[31]. They have evolved mutualistic relationships with chemosynthetic bacteria, allowing them to colonize these extreme environments. Inside their gills, they host intracellular sulfide- or methane-oxidizing gammaproteobacterial endosymbionts. Many bathymodiolin species host both types in a ‘dual symbiosis’. Assié et al. recently discovered a novel family of Campylobacterota, which colonizes bathymodiolin mussels from around the world[32]. In contrast to the gammaproteobacterial endosymbionts of these mussels that are harbored inside gill cells called bacteriocytes, these Campylobacterota are filamentous epibionts that colonize the surface of the gill epithelia in dense patches.

In this study, we used a multi-omics approach to investigate the metabolism of these novel epibionts in two different species of bathymodiolin mussels. Surprisingly, the epibionts have, and express, all the genes required for the CBB cycle but are missing key genes of the rTCA cycle. These CBB cycle genes were most likely acquired by horizontal gene transfer from diverse sources. With a recently developed, highly sensitive, direct stable isotope fingerprinting technique[33], we show that the proteins of these epibionts have an isotopic signature typical of the CBB cycle, further demonstrating its importance for the metabolism of these epibionts. The discovery of Campylobacterota that employ the CBB cycle for CO_2 fixation has wide-reaching implications for understanding the evolution of carbon fixation pathways, and for interpreting stable isotope values in environmental samples.

Main text

Genome assemblies and annotations

We assembled Campylobacterota draft genomes from gill metagenomes of two mussel species: *"B." childressi* and *B. azoricus*. The draft genome from "*B." childressi* was 2.2 Mb, and estimated to be 95% complete. It had 30% GC, 2204 predicted protein-coding genes and 31 tRNA-encoding genes. The draft genome from *B. azoricus* was estimated to be 92% complete at 2.3 Mb. It had 30% GC, 2155 predicted protein-coding genes and 37 tRNAs (Table S1). The draft genomes had an average nucleotide sequence identity (ANI) of 83.1%, indicating that they represent different species belonging to the same genus[34–37].

Previous 16S ribosomal RNA gene (16S rRNA) sequence analysis identified this group of epibionts as a novel family-level deep-branching sister group of the *Sulfurovum* clade within the Campylobacterota[32]. The Campylobacterota draft genome from *"B." childressi* contained a partial 16S rRNA sequence (586 bp) that was 100% identical to the epibiont sequence previously published[32]. The 16S rRNA sequence identity and the ANI information allowed us to link the two draft genomes to the previously described epibiont[32].

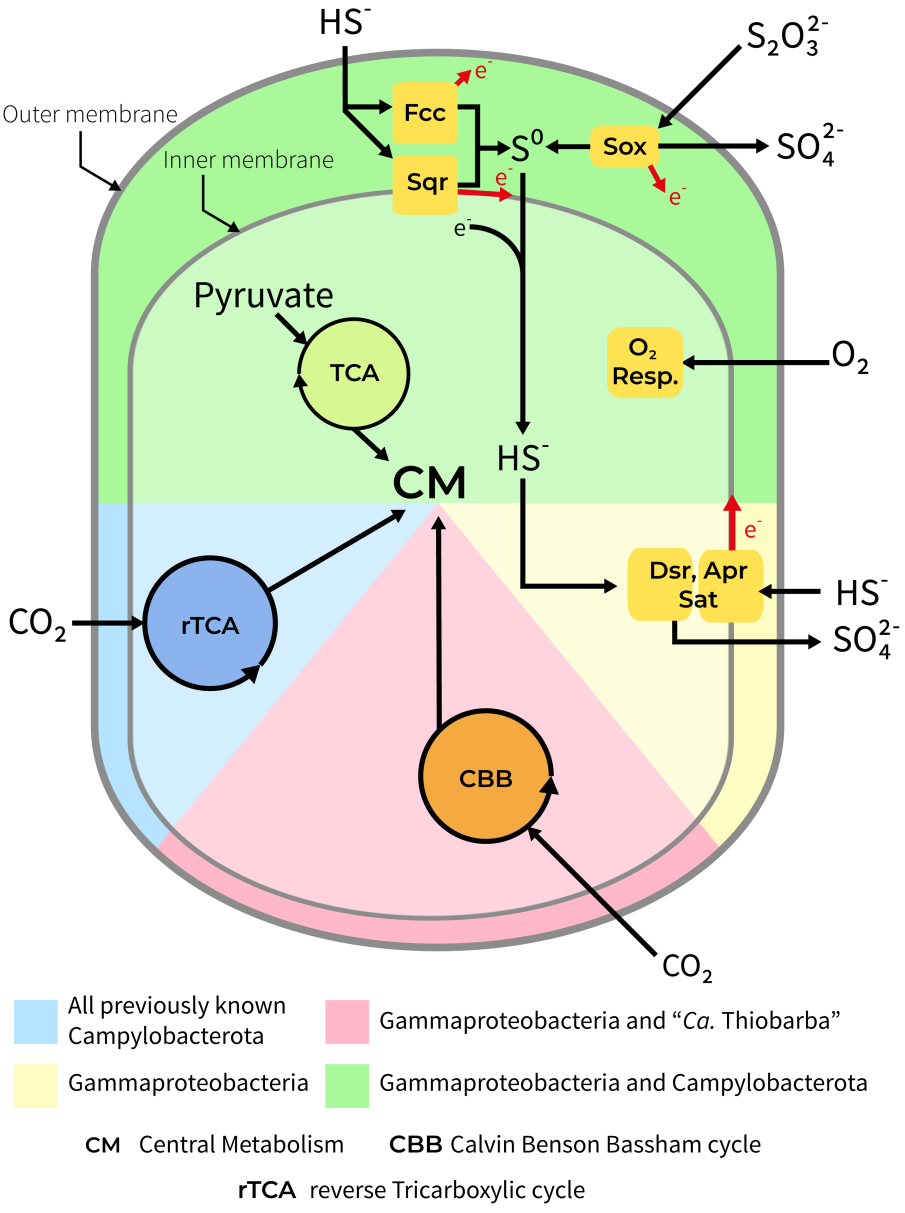
To better resolve the relationships of the mussel epibionts to other Campylobacterota we analyzed a set of 18 conserved marker genes from the two epibiont draft genomes and other publicly available Campylobacterota genomes (Figure S1). In contrast to the previous 16S rRNA based phylogeny[32], our analysis placed the mussel epibionts on a long branch, basal to the main Campylobacterota families. The long-branch formation for the genomes presented in this study is likely related to low amino acid sequence identity (AAI) values between these and the Campylobacterota representative genomes. AAI values were below 48% when comparing the Campylobacterota bins found in our bathymodiolin samples with their closest relative genomes, *Sulfurospirillum arcachonense* and *Arcobacter anaerophilus* (Supplementary Table 2). According to the guidelines of Rodriguez and Konstantinidis[36] , organisms with AAI values higher than 30% and lower than 55-60% are likely to belong to the same division, but not the same genus. This is corroborated on the 16S rRNA level[37], where our genomic data supports the previously published 16S rRNA based study[32] , indicating that the epibiont species belong to a novel family of Campylobacterota.

We therefore propose the new *Candidatus* family "Thiobarbaceae" (Campylobacterales, Campylobacterota), with the name composed of "*Thio-*" from the Greek word θεῖον, theîon for sulfur and "*barba*" from the Latin word for beard. The proposed family includes the novel *Candidatus* genus "Thiobarba" with two *Candidatus* species "*Ca.* T. azoricus" and "*Ca.* T. childressi", for the two epibiont species in reference to their respective hosts, *B. azoricus* and *B. childressi*. For more details on the aetiology see Supplementary note 1.

Unexpected carbon fixation pathways of "*Candidatus* Thiobarba spp."

Considering their phylogenetic relationship to free-living chemolithoautotrophic and mixotrophic Campylobacterota and their presence in sulfide-rich environments, we searched the epibiont draft genomes for metabolic pathways indicative of heterotrophy, autotrophy and sulfur oxidation. Both "*Ca.* Thiobarba" genomes encoded all the genes for the SOX multi-

157 enzyme pathway of sulfur oxidation, and are thus capable of lithotrophy using reduced sulfur
158 compounds as electron donors (Figure 1). Like other sulfur-oxidizing Campylobacterota, they
159 also appear capable of heterotrophic growth as their genomes contained a TCA cycle and a
160 partial glycogenesis/glycolysis pathway (Supplementary note 2 and Figure S2).



161

Figure 1. "*Ca*. Thiobarba spp." share metabolic features of Gammaproteobacteria and Campylobacterota. Figure shows overview of the main metabolic pathways for energy generation and carbon fixation in known chemosynthetic Gammaproteobacteria and Campylobacterota compared to the metabolism of "*Ca.* Thiobarba spp.".

All previously described sulfur-oxidizing Campylobacterota use the reverse TCA cycle for carbon fixation[2]. All of these bacteria have genes encoding the enzymes for this cycle including the pyruvate: ferredoxin oxidoreductase genes *porABCD,* the 2-oxoglutarate oxidoreductase genes *oorABDG,* and the ATP citrate lyase genes *aclAB*. Unexpectedly, we could not find most of these genes in the "*Ca.* Thiobarba" genomes. The "*Ca.* T. childressi" draft genome contained only the *porAB* genes, and the "*Ca* T. azoricus" draft genome contained *porABCD* and *aclA*, but not the *oorABDG* genes. To confirm that these genes were not missing because of errors in assembly, binning or annotation, we searched the genomes and unbinned metagenome assemblies with BLAST. No additional rTCA cycle genes could be found in the draft genomes or in the entire *"B." childressi* metagenome assembly (Supplementary note 3). The absence of rTCA cycle genes suggests that either a) the epibionts never had a complete rTCA cycle, or b) it has been lost over the course of evolution. The additional roles of the *por* genes in other metabolic pathways, such as pyruvate fermentation, could explain why these are present, at least in part, in both lineages[38].

Although the rTCA cycles were incomplete, both *"Ca.* Thiobarba" genomes contained all the genes required for carbon fixation via the CBB cycle (Figure 1). Most CBB cycle enzymes are used in other metabolic pathways, and are thus also found in heterotrophic bacteria, but two enzymes are unique to the cycle: Phosphoribulokinase (PRK) and ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO)[2]. In both "*Ca.* Thiobarba" species, 9 out of the 12 genes encoding PRK, RuBisCO and accessory proteins were grouped in two clusters, while three additional genes for the CBB cycle were scattered on separate contigs (Figure S3). The first cluster consisted of the RuBisCO Form I large and small subunits (*rbcL*

and *rbcS*), a conserved hypothetical protein, and the RuBisCO activation protein *cbbQ.* The order of these genes was conserved in both epibionts (Figure S4). "*Ca.* T. childressi" had an additional gene encoding the RuBisCO activation protein *cbbO* in this first cluster. In the "*Ca.* T. azoricus" genome, this gene was located on a separate contig. The second cluster included the genes coding for fructose-1,6-bisphosphatase, PRK, transketolase, phosphoglycolate phosphatase (not known to be involved in the CBB cycle), fructose-bisphosphate aldolase and ribulose-phosphate 3-epimerase (Figure S3). The order of the second gene cluster was consistent in both epibiont genomes, but the gene neighborhoods surrounding this cluster differed (Figure S4).

CBB cycle expression in "*Ca*. Thiobarba childressi"

To confirm expression of the CBB cycle by the epibionts, we analyzed the metatranscriptomes and -proteomes of *"B." childressi*, the mussel species with the highest abundance of these epibionts[32]. We found that all CBB cycle genes were expressed in the transcriptomes, including the *rbcL* and *rbcS*, which were among the most highly expressed genes of this epibiont (Table 1). Although "*Ca* T. childressi" was present in relatively low abundance in the metaproteome samples (~0.5% of the total sample proteinaceous biomass, calculated according to [39]), the RuBisCO small and large subunits were among the unique "*Ca.* T. childressi" proteins detected, further indicating high expression levels. The abundance of CBB cycle transcripts and proteins highlights their importance in the metabolism of "*Ca.* T. childressi" (for full transcription and expression information, see Tables S3 and S4).

211 **Table 1** Transcription and translation ranks for the detectable genes involved in the CBB

212 cycle of "*Ca.* T. childressi".

Gene_ID	Name	Transcription rank	Translation rank
BCM6EPS_1532	Ribulose bisphosphate carboxylase large chain EC 4.1.1.39	14	18
BCM6EPS_1531	Ribulose bisphosphate carboxylase small chain EC 4.1.1.39	17	19
BCM6EPS_1455	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase EC 1.2.1.12	75	Not detected
BCM6EPS_1028	Transketolase EC 2.2.1.1	80	Not detected
BCM6EPS_1030	Fructose-bisphosphate aldolase class II EC 4.1.2.13	90	Not detected
BCM6EPS_1031	Ribulose-phosphate 3-epimerase EC 5.1.3.1	150	Not detected
BCM6EPS_1027	Phosphoribulokinase EC 2.7.1.19	160	Not detected
BCM6EPS_1026	Fructose-1-6-bisphosphatase- type I EC 3.1.3.11	182	Not detected
BCM6EPS_1029	Phosphoglycolate phosphatase EC 3.1.3.18	236	Not detected
BCM6EPS_1456	Phosphoglycerate kinase EC 2.7.2.3	440	Not detected
BCM6EPS_514	Triosephosphate isomerase EC 5.3.1.1	461	Not detected

213

Direct stable isotope fingerprinting confirmed a CBB signature for "*Ca.* Thiobarba childressi"

The stable carbon isotope signatures of an environmental sample reflect the pathway that dominates inorganic carbon fixation in the chemoautotrophic members of the bacterial community[40]. Due to differences in kinetic isotope effects, the enzymes involved in the different carbon fixation pathways vary in the degree to which they discriminate against the heavier ^{13}C. This leads to a shift in the $^{12}C/^{13}C$ ratio between the inorganic carbon source and the generated biomass that is characteristic for the carbon fixation pathway. The CBB cycle generates a -13 to -26‰ shift of the $\delta^{13}C$ ratio, while the rTCA cycle leads to a much smaller -3 to -13‰ shift[40].

The average $\delta^{13}C$ value of bulk *"B." childressi* gill tissues was -47.1 ± 2.6‰, (Table S5). However, these values reflect the stable isotope composition of all members of the symbiotic community. As most of the biomass is from the host animal or the highly abundant methane-oxidizing gammaproteobacterial endosymbiont, the signal of the epibiont is greatly diluted[41]. To overcome this limitation and to distinguish between the stable carbon isotope values of the symbiotic partners, we employed the recently-developed direct Protein-SIF method (SIF = stable isotope fingerprinting) on our metaproteomic data set[33]. Direct Protein-SIF quantifies the stable isotopic composition of uncultivated members of a mixed community for which genomes or transcriptomes are available. Peptides from the methane-oxidizing symbionts had a $\delta^{13}C$ of -38.8 ± 0.7‰, and host peptides had -44.2 ± 0.6‰. These values are similar to those of the methane gas at this cold seep site. Thus, the methane-oxidizing symbionts likely obtain most of their carbon from methane[42,43]. The host values were similar to those of bulk measurements. However, they were unexpectedly light compared to the methane-oxidizing symbionts, considering that these mussels are thought to gain most of their nutrition from their methane-oxidizing symbionts, and would therefore be expected to have similar $\delta^{13}C$ values. As *B. childressi* is known to be capable of filter-feeding[44], these

values possibly reflect nutritional supplementation from filter-feeding on microorganisms with even lighter $\delta^{13}C$ values than the methane-oxidizing symbionts, that is from the seep environment (as phototrophic microorganisms from the surface would have heavier $\delta^{13}C$ values).

"Ca. T. childressi" had a much lower abundance in the metaproteomic dataset compared to the host and the methane-oxidizing symbionts. Nevertheless, we were able to detect 50 peptides that were unique to "*Ca.* T. childressi". This allowed us to estimate its natural $\delta^{13}C$ value, which was relatively light at -66.6 ± 12.5‰. There are two possible inorganic carbon sources for these epibionts: 1) ambient seawater inorganic carbon, which has a $\delta^{13}C$ value of +3‰[42], and 2) inorganic carbon produced as an end product of methane oxidation by methane-oxidizing bacteria or respiration by the host, which we expect to be around -39‰ for a gas hydrate site, similar to our collection site[41,42]. We calculated the expected values of biomass generated if either of these carbon sources were fixed through the rTCA cycle or the CBB cycle (Figure 2). Regardless of the inorganic carbon source, the $\delta^{13}C$ values of "*Ca.* T. childressi" peptides are far lighter than would be expected if they used the rTCA cycle. They are, however, consistent with the values for carbon fixation that could be expected when "*Ca.* T. childressi" used the CBB cycle, with inorganic carbon derived from symbiont methane oxidation or host respiration.

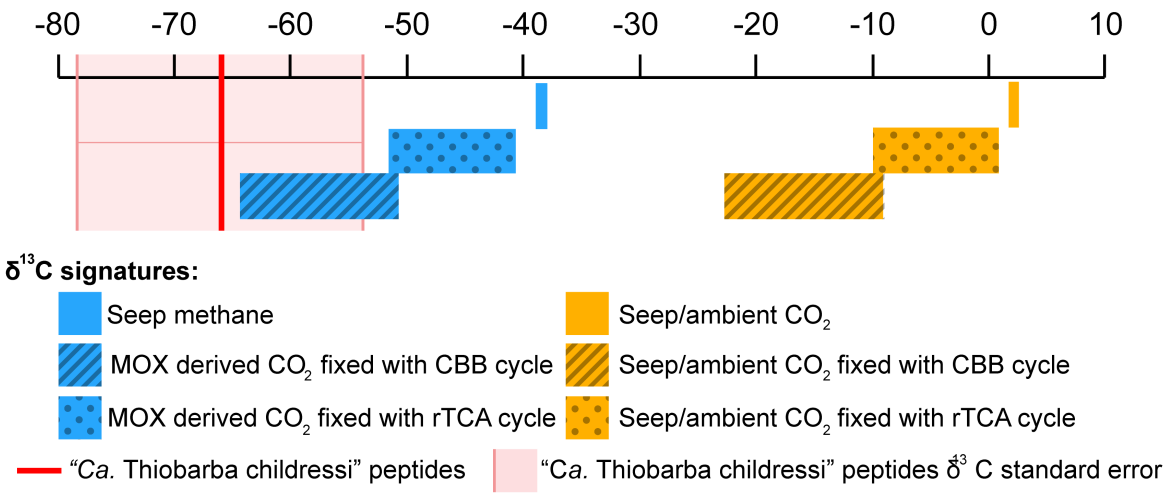


Figure 2. Stable carbon isotope values of "*Ca*. Thiobarba childressi" are consistent with carbon fixation via the CBB cycle. Model of $\delta^{13}C$ values of deep-sea carbon and the predicted influence of different inorganic fixation pathways on these values. The $\delta^{13}C$ values of CO_2 originating from ambient seawater are shown in yellow, and the expected $\delta^{13}C$ values of CO_2 originating from methane oxidation are shown in blue. The red line represents the average $\delta^{13}C$ value measured for "*Ca.* Thiobarba" peptides using direct Protein-SIF. Reference $\delta^{13}C$ values for "Seep methane" and "Seep/ambient CO_2" are based on Macavoy et al.[41] and Sassen et al.[42]. Transformations of $\delta^{13}C$ values for each metabolic pathway are estimated based on Pearson et al.[40].

Calvin cycle genes in free-living Campylobacterota

After discovering the CBB cycle in the mussel epibionts, we asked if other members of the Campylobacterota might also have acquired these genes. We discovered key CBB cycle genes in a Campylobacterota draft genome binned from a metagenomic library from diffuse hydrothermal fluids collected in the Manus Basin (Western Pacific)[23]. This draft genome was composed of 60 contigs with 29.1% GC content, and based on the CheckM single-copy genes set, was 92.2% complete[45]. Phylogenomic reconstruction placed this organism on a deep branch basal to the Arcobacteraceae family. AAI values showed between 55 and 58% similarity with the Arcobacteraceae, thus, this Campylobacterota bin might belong to a new genus within the Arcobacteraceae (Supplementary Table 2). Our phylogenetic analyses and calculated AAI values clearly show that this environmental bin belongs to a Campylobacteraota family distinct from "*Ca.* Thiobarba" (Figure S1). Although a

full rTCA cycle was present in the draft genome, we also found genes coding for a RuBisCO form I enzyme, a hypothetical gene and CbbQ in one cluster. This cluster shared the same gene order, as well as 84% nucleotide sequence identity, with the CBB cycle cluster we found in "*Ca.* Thiobarba" (Figure S4). The high sequence similarity between these clusters suggests a similar origin for both of them. If these genes and the enzymes they encode are active in the Manus Basin organism, then free-living Campylobacterota may also be able to use the CBB cycle to fix carbon. These bacteria could be using both cycles depending on the environmental setting, as suggested for the sulfide-oxidizing gammaproteobacterial symbionts of vestimentiferan tubeworms[46–48] found at hydrothermal vents, the large sulfur bacteria *Beggiatoa* and *Thiomargarita* spp.[49–51], and recently the cultivable sulfur oxidizer *Thioflavicoccus mobilis*[52]. The tubeworm symbiont *Ca.* E. persephone expresses both the CBB and the rTCA cycle in the same host individual, but it is still unclear how these two cycles are coordinated at the level of individual symbiont cells, or over time[47,48].

Possible evolutionary origins of Campylobacterota CBB genes

Considering the lack of CBB cycle genes in all Campylobacterota investigated prior to this study, it is most likely that this carbon fixation pathway was acquired by "*Ca.* Thiobarba" and free-living Campylobacterota through horizontal gene transfer, rather than being an ancestral pathway in this phylum. We investigated the evolutionary origins of the genes coding for CBB enzymes, including those with additional roles in other metabolic pathways, using BLAST analyses of nucleotide and protein sequences, and phylogenetic reconstruction of protein sequences. BLAST analyses revealed that only two of the "*Ca.* Thiobarba" CBB genes were affiliated with genes from other Campylobacterota. Of the other 10, five had best hits to Gammaproteobacteria, and five had best hits to Betaproteobacteria (Table S6).

Phylogenetic reconstruction further supported our hypothesis that the "*Ca.* Thiobarba" CBB cycle is a 'patchwork' of genes with evolutionary origins in the Betaproteobacteria, Gammaproteobacteria, and Campylobacterota (Figure 3). The RuBisCO large and small subunits *rbcL* and *rbcS*, their accessory proteins *cbbQ* and *cbbO*, as well as the glyceraldehyde-3-phosphate dehydrogenase proteins clustered with a clade of gammaproteobacterial sulfur-oxidizing chemolithoautotrophs. Many of the related sequences belonged to free-living sulfur oxidizers such as "*Ca.* Thioglobus autotrophicus" and the gammaproteobacterial sulfur-oxidizing endosymbionts of bathymodiolin mussels (Figure 4). Phylogenetic analysis of "*Ca.* Thiobarba" PRK proteins placed these on a long branch between gamma-, alpha- and betaproteobacterial clades, but this placement did not have high support (Figure 5). This could indicate that the "*Ca*. Thiobarba" PRK proteins truly belong to a Campylobacterota gene family, and because these are the first sequences available from this family, their phylogenetic placement is currently not well supported. Further sampling may help to clarify their evolutionary history. Four "*Ca.* Thiobarba" CBB cycle proteins consistently belonged to a sister branch to betaproteobacterial sequences (fructose 1,6-bisphophatase, 1,6-bisphophate aldolase, transketolase and ribulose phosphate 3-epimerase). Only two proteins were phylogenetically related to those from other Campylobacterota (phosphoglycerate kinase and triose phosphate isomerase) (Figures S5 to S24).

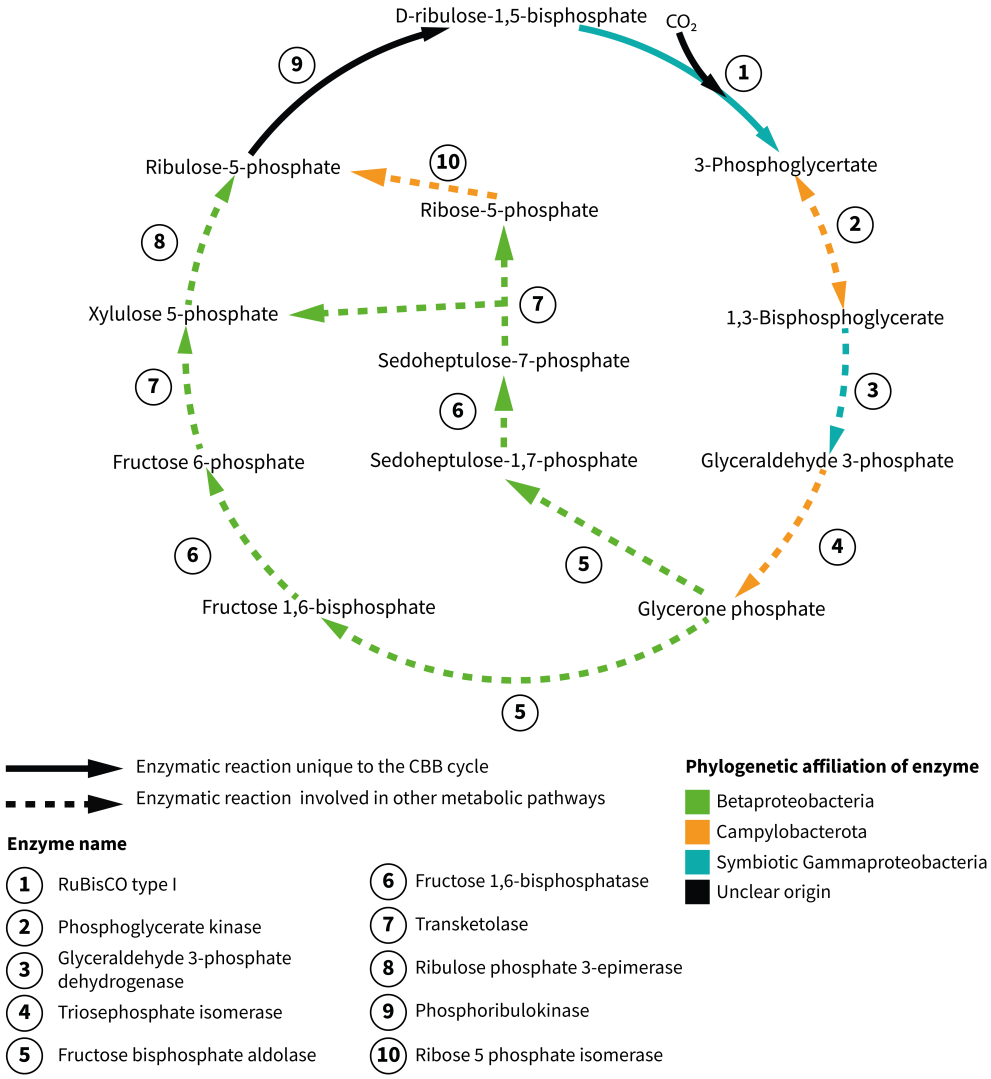
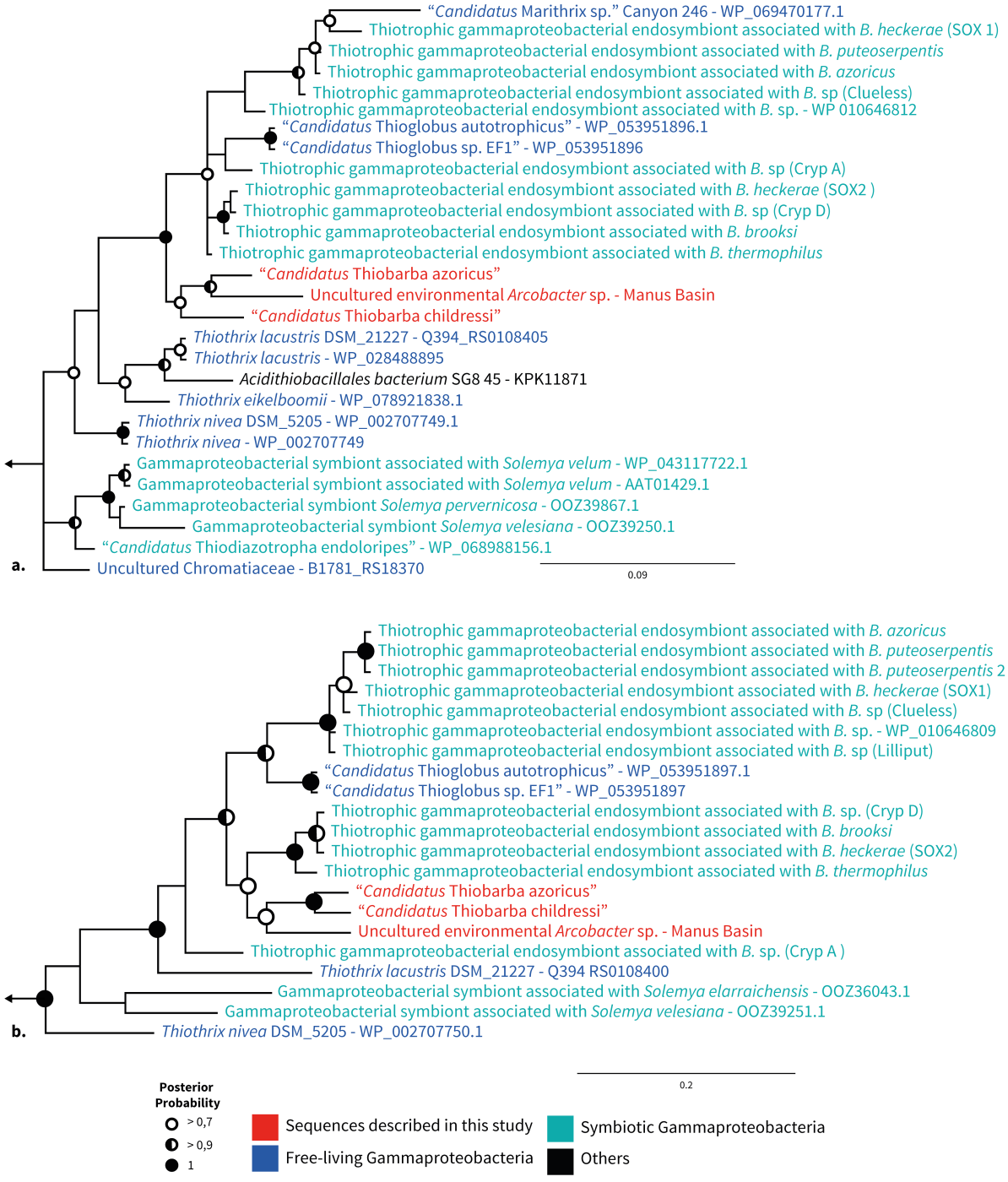


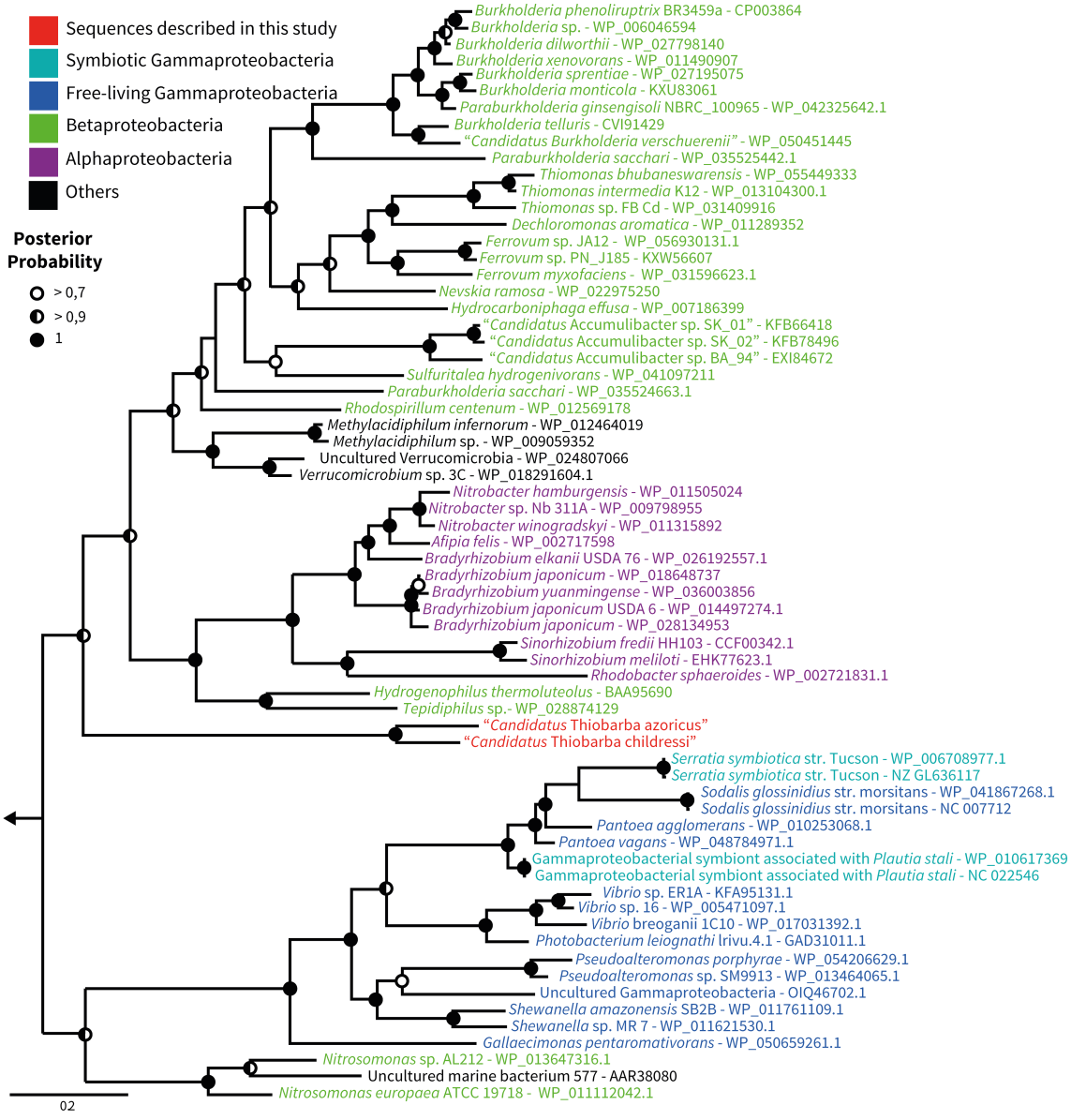
Figure 3. "*Ca*. Thiobarba" genomes encode a CBB cycle with genes affiliated to at least three phylogenetically distinct classes. The solid arrows indicate enzymatic reactions that are unique to the CBB cycle, while the dashed arrows indicate that the enzymes are also involved in other metabolic pathways. Enzyme names are shown in bold and the colors represent their phylogenetic affiliations.



334

335 **Figure 4**. "***Ca.* Thiobarba" RuBisCO proteins cluster with gammaproteobacterial**
336 **sequences.** Bayesian inference trees of RuBisCO large (a) and small (b) subunit amino acid
337 sequences under an LG model with Gamma-distributed rates of evolution. Analyses were
338 performed with 6 million generations using two parallel Monte Carlo Markov chains. Sample
339 trees were taken every 25000 generations. Left arrows indicate truncated tree, tree roots
340 were built from *Prochlorococcus* and *Synechococcus* sequences for (a) and *Planktothrix* and
341 *Synechococcus* sequences for (b). Full trees are displayed as Supplementary figures 14 and
342 15.
343

344



345

Figure 5. "*Ca*. Thiobarba" phosphoribulokinases are loosely affiliated with those from Betaproteobacteria, Alphaproteobacteria, and Verrucomicrobia. Bayesian inference tree of phosphoribulokinase amino acid sequences under an LG model with Gamma-distributed rates of evolution and a proportion of invariant sites. Analyses were performed with 6 million generations using two parallel Monte Carlo Markov chains. Sample trees were taken every 25000 generations. Left arrow indicates truncated root, the root is built from distant *Prochlorococcus* and *Synechococcus* sequences. Full tree is displayed as supplementary figure 23.

The CBB genes consistently fell into three phylogenetic groups: some CBB genes were most closely related to those of other Campylobacterota, some to those of Gammaproteobacteria, and some to those of Betaproteobacteria. The "*Ca*. Thiobarba" CBB genes that fell within the Gammaproteobacteria were similarly organized to genes with which they were most closely related, such as those from "*Ca.* Thioglobus autotrophicus" and the endosymbionts of bathymodiolin mussels (Figure S24 and Supplementary note 4). Similarly, the genes that were most closely related to Betaproteobacteria had a similar organization to genes found in free-living Betaproteobacteria such as *Paraburkholderia xenovorans* (NC_007651) and *Dechloromonas aromatica* (NC_007298) (Figure S24). This similarity further supports the hypothesis that these CBB cycle genes were acquired by "*Ca*. Thiobarba" at least twice, in independent horizontal gene transfer (HGT) events, with one possibly originating from Gammaproteobacteria, and another possibly from Betaproteobacteria. Alternatively, it is also possible that the Betaproteobacteria-like genes clustering on long branches, such as the PRK, are Campylobacterota genes that have not previously been sequenced. Regardless of the number of HGT events, the acquisition of these genes presumably happened in a common ancestor to the "*Ca.* Thiobarbaceae". Codon usage analysis supports our hypothesis of a relatively ancient acquisition, as the codon usage of the CBB genes was similar to that of the *"Ca.* Thiobarbaceae" core genome (Figure S25). After horizontal acquisition, the codon usage initially reflects the original donor's, but over time mutates to match its host's[53].

Gene order within each of the two CBB clusters was identical in "*Ca.* T. azoricus" and "*Ca.* T. childressi". This further supports our hypothesis of a single acquisition event for each cluster in a common ancient ancestor. This synteny also highlights the tendency of these clusters to resist genomic rearrangements. In contrast, the genomic neighborhoods of the CBB clusters differed between the two "*Ca.* Thiobarba", indicating that subsequent genome rearrangements occurred since the divergence of these two epibionts. Mobile element genes and transposases were the most highly expressed genes in "*Ca.* T. childressi" based on our transcriptomes, which, if active, could explain these rearrangements (Table S3)[54].

Evolutionary advantages of the CBB cycle

Members of the Campylobacterota occupy remarkably diverse habitats, and have a range of different lifestyles and metabolic capabilities, from chemolithoautotrophs that use a suite of electron donors and acceptors, to heterotrophic symbionts and pathogens of humans and other animals[13,55]. Evolutionary studies suggest that Campylobacterota emerged in deep-sea habitats, subsequently colonizing and diversifying across terrestrial and human-associated environments[13,56]. Considering the distribution of chemosynthetic potential within the Campylobacterota, it has been hypothesized that they evolved from an autotrophic common ancestor that first used the Wood–Ljungdahl pathway before switching to a more flexible rTCA cycle[13,17,57]. We hypothesize that in the symbiotic "*Ca*. Thiobarba" lineage, the rTCA cycle was replaced by yet another carbon fixation pathway, the CBB cycle.

Several environmental and genomic factors provide important clues as to why the CBB cycle was selected over the rTCA cycle in "*Ca*. Thiobarba". Both the CBB and rTCA cycles serve the same purpose, the fixation of inorganic carbon to provide building blocks for cell biomass. But a major difference between the two known carbon fixation pathways is their energy requirements. For example, if one molecule of pyruvate is synthesized from CO_2 via the CBB cycle seven molecules of ATP are used, while the rTCA cycle only requires two

ATP[2]. From an evolutionary point of view, exchanging a more energy-efficient carbon fixation pathway with a costlier one could best be explained if it comes with an additional advantage such as oxygen tolerance. The rTCA cycle relies on ferredoxin-based enzymes, which are quickly oxidized by oxygen, and as a result, most organisms with an rTCA cycle are anaerobes or microaerobes[58,59]. In contrast, CBB cycle enzymes are less affected by oxygen[60]. "*Ca.* Thiobarba" species colonize the gills of bathymodiolin mussels, a gas exchange organ that is exposed to oxygen and is typically dominated by gammaproteobacterial endosymbionts. The close phylogenetic relationship between some *"Ca.* Thiobarba" CBB genes with those from the sulfur-oxidizing gammaproteobacterial endosymbionts of bathymodiolin mussels suggests that either i) both symbionts acquired CBB genes from the same source or ii) *"Ca.* Thiobarba" acquired key genes from the gammaproteobacterial endosymbionts already adapted to the mussel gill niche.

Many deep-sea Campylobacterota grow attached to surfaces[29,61,62], thus, a "*Ca.* Thiobarba" ancestor might have colonized mussel gills prior to acquiring the CBB cycle. Living attached to the gills would bring these epibionts into close proximity to the gammaproteobacterial endosymbionts. Sharing a niche has been shown to be a stronger predictor of horizontal gene transfer than phylogenetic relatedness[63]. Moreover, Campylobacterota have remarkably flexible genomes, with rampant genomic rearrangement and DNA uptake[64–66]. This affinity for foreign DNA uptake, and the physical proximity of epibionts and endosymbionts support scenario ii) above. The acquisition of the CBB carbon fixation pathway may have enabled "*Ca.* Thiobarba" to thrive attached to an animal host, leading to the complete reliance on the CBB cycle for carbon fixation and the gradual loss of the rTCA cycle.

Another major difference between the CBB and rTCA cycles is the metabolic end product: Sugar precursors such as glyceraldehyde 3-phosphate for the CBB cycle, and acetyl-CoA for the rTCA cycle[2,60]. Intriguingly, the "*Ca.* Thiobarba" genomes encoded many

pathways requiring sugars, including N-linked glycosylation, capsular polysaccharides and lipooligosaccharide synthesis, some of which are not found in related chemolithoautotrophic Campylobacterota, but are present in heterotrophic host-associated Campylobacterota. “Ca. Thiobarba’s” ability to gain sugar precursors while fixing CO_2 using the CBB cycle might be advantageous, despite the higher energy requirements compared to the rTCA cycle. Many of the pathways requiring sugars are predicted to play a role in surface structures and extracellular polysaccharide capsule formation, which can be key mediators of host attachment, and thus may be essential for its epibiotic lifestyle[67,68].

Evolving a Calvin cycle in nature and the laboratory

The complex metabolic network that links carbon fixation and central carbon metabolism poses a massive challenge to switching carbon fixation pathways, either in nature or in the laboratory. These links are usually specific to each pathway and to each organism[60]. Efforts to introduce non-native carbon fixation pathways have mainly focused on the CBB cycle because theoretically, only two additional enzymes are needed to run this cycle, even in heterotrophs such as *E. coli*[69,70]. However, a number of challenges must be overcome to express ‘foreign’ carbon fixation pathways in new organisms. In addition to the challenges inherent in expressing horizontally acquired genes, such as non-native promoter and codon usage, and the need for chaperones and biosynthesis enzymes, gene expression must be tightly regulated to balance the production and consumption of intermediates and end products. Because of this, to run the CBB cycle in engineered *E. coli*, the CBB cycle had to be synthetically decoupled from gluconeogenesis by deleting the phosphoglycerate mutase gene[69]. Switching from one carbon fixation pathway to another may be simpler in chemolithoautotrophs than re-wiring a chemoorganoheterotroph such as *E. coli* to use the CBB cycle. In a chemolithoautotroph, production of energy and reducing equivalents are already decoupled from carbon fixation, as they are generated through oxidation of reduced

compounds such as sulfur. Nevertheless, switching from the rTCA to the CBB cycle is a major shift in cellular metabolism, requiring adaptation of diverse biosynthetic pathways linked to carbon fixation. As far as we are aware, this has not yet been observed in nature, but in the laboratory, *E. coli* required extensive fine-tuning of metabolic enzymes beyond the CBB cycle through experimental evolution to run a fully functional CBB cycle[3,69].

Conclusions

The environment is a potent driving force in structuring symbiotic and free-living microbial communities[27,30,71]. The distribution of gammaproteobacterial and campylobacterotal sulfur oxidizers is a typical example of adaptation to a geochemical niche; in a range of environments from hydrothermal vents[23] and cold seeps[21] to oxygen minimum zones[72] and coastal sediments[73], Gammaproteobacteria are usually associated with low-sulfide, high-oxygen environments, and campylobacterota with high-sulfide, low-oxygen environments. The horizontal acquisition of the CBB cycle genes may have allowed campylobacterotal "*Ca.* Thiobarba" to establish a symbiotic relationship in a niche that is usually dominated by Gammaproteobacteria.

The diverse origins of "*Ca*. Thiobarba's" CBB cycle genes showcases the modularity[74] of bacterial metabolism and demonstrates that in principle, fully functional metabolic cycles can be pieced together with enzymes from different organisms, both in the laboratory[3] and in nature. In addition to acquiring the two genes theoretically required by a heterotroph to encode a full CBB cycle, "*Ca*. Thiobarba" seems to have replaced an extensive set of additional CBB genes. This suggests that similar to laboratory models, this natural metabolic switch required 'tweaking' of further enzymes of this pathway, and possibly other pathways that siphon off intermediates. Metabolic modularity is considered one of the main factors organizing biological networks[74]. Understanding genome evolution in "*Ca.* Thiobarba" will shed light on the complex interplay between gene acquisition, expression and the selection

that caused the evolution of this major metabolic shift. Our findings highlight the central role that horizontal gene transfer plays in metabolic modularity and environmental adaptation.

Carbon isotope signatures are routinely used to assess the relative importance of the CBB and rTCA cycles in contemporary and past natural environments, and to infer the key organisms responsible for primary production[75–78]. Although stable isotope signatures may accurately reflect the relative importance of distinct carbon fixation pathways in environmental samples, our study shows that assigning these key ecological functions to particular microbial groups requires a deeper understanding of how the underlying metabolic pathways are distributed in nature.

Material & Methods

Sample collection

"B." childressi individuals were collected at cold seeps in the northern Gulf of Mexico at the GC246 and GC234 sites during the R/V Atlantis AT26-13 cruise in April 2014, Nautilus cruise NA044 in July 2014 and Nautilus NA058 cruise in May 2015. The *B. azoricus* individual was collected at the Lucky Strike hydrothermal vent field on the North Mid-Atlantic Ridge (NMAR) during the Biobaz cruise in 2013. A list of samples and fixation details are summarized in Table S7.

DNA and RNA extraction

DNA was extracted from mussel gill tissue according to Zhou *et al.* (1996)[79] with the following modifications: An initial overnight incubation step was performed at 37 °C in 360 µl of extraction buffer (100 mM Tris-HCl [pH 8.0], 100 mM sodium EDTA [pH 8.0], 100 mM sodium phosphate [pH 8.0], 1.5 M NaCl, 1% CTAB) and 40 µl of proteinase K (10 mg/ml). For transcriptome sequencing, RNA was extracted with an Allprep(R) DNA/RNA micro kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Concentrations of DNA and RNA were measured with a Qubit® 2.0 Fluorometer (Invitrogen, Eugen, USA).

Metagenome sequencing and assembly

DNA extracted from gill tissues of one *"B." childressi* individual was sequenced at the Center for Biotechnology at the University of Bielefeld (Bielefeld, Germany). A total of 471,459,598 paired-end reads (150 bp) and 7,739,150 paired-end reads (250 bp long) were generated on Illumina HiSeq 1500 and MiSeq machines, respectively. DNA extracted from gill tissues of one *"B." childressi* and one *B. azoricus* individual was sequenced by the Max Planck Genome Center (Cologne, Germany) and generated respectively 57,172,785 and 159,408,731 paired-end reads (150 bp long) on an Illumina HiSeq 2500.

We screened the metagenomic and metatranscriptomic libraries for the presence of campylobacterotal 16S rRNA sequences. The PhyloFlash 2.0 suite (https://github.com/HRGV/phyloFlash) was used to perform RNA small subunit (SSU) screening and reconstructions.

Metagenome assembly was performed as follows: First the raw reads were quality trimmed (Q=2) and Illumina adapters were removed using BBduk (BBmap suite v37.9 from Bushnell B. - sourceforge.net/projects/bbmap/). An initial assembly was performed with Megahit[80] using default settings. The resulting assembly file was then analyzed with metawatt V2.0 binning tools[81], and draft genome bins were generated by analyzing contig tetranucleotide frequency, differential coverage and GC content. Contigs belonging to bins with an Campylobacterota taxonomic signature were extracted. The quality-trimmed metagenomic reads were then mapped against the Campylobacterota contigs using Bbmap (BBmap suite v37.9), filtering reads with a minimum identity of 98%. The mapped reads were then used for a new assembly using SPAdes 3.4.2[82] with default settings. Additional details on the assembly process of "*Ca.* T. azoricus" are described in Supplementary note 3. The bin of the free-living Campylobacterota carrying CBB cycle genes was obtained from the Manus Basin metagenome "NSu-F5" as described in[23] with three rounds of read-mapping, re-assembly and binning for final bin completion of 92% and 11.7% contamination.

Bin quality was checked with CheckM[45] and a new iteration of taxonomic binning, mapping and assembly was performed until no contamination from other bacterial strains or host remained in the assembly. Contigs smaller than 900 bp were included in BLAST analysis but excluded from subsequent analyses because they were unlikely to have any relevant genetic information. Genomes were annotated with RAST and cross-checked with IMG ER web servers[83–85]. Genome average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated using the AAI and ANI calculator from the enveomics collection[86] with the default settings. The specific coverage for genomes and gene was calculated using BBmap.

Raw data was uploaded to the European Nucleotide Archive under the accession numbers: PRJEB19882, PRJEB23284, and PRJEB23286.

Transcriptome sequencing and processing

Transcriptomes of three *"B." childressi* individuals were sequenced at the Max Planck Genome Center (Cologne, Germany) details are in Table S7. Transcriptome reads were processed as in Rubin-Blum et al.[87]. Briefly, raw reads were mapped against the "*Ca.* T. childressi" draft genome with BBmap (BBmap suite v.37.09): reads were quality trimmed (Q=2), Illumina adapters removed and a minimum similarity of 98% used to map to the reference genome. The number of transcriptome reads mapping to each gene was estimated with featureCounts v1.5.2[88]. To compare the transcriptome libraries of each individual, a normalization factor was estimated with calcNormFactors based on the trimmed mean of M-values (TMM) implemented in the edgeR version 3.16.5[89]. The TMM normalized read counts were converted to reads per kilobases of exon per million reads mapped (RPKM) with edgeR (http://www.bioconductor.org).

rTCA cycle gene screening

To confirm presence or absence of the rTCA cycle in the metagenomic and transcriptomic libraries, we created a BLAST database containing published amino acid sequences of Campylobacterota rTCA key genes, citrate lyase, 2-oxoglutarate ferredoxin oxidoreductase and pyruvate ferredoxin kinase. The first metagenomic assembly iterations, as well as the final Campylobacterota bins, were screened using BLASTX against the respective database to detect the presence of potential rTCA cycle genes.

Phylogenomic reconstruction

Phylogenomic trees were calculated using Phylogenomics-tools (Brandon Seah, https://github.com/kbseah/phylogenomics-tools). The draft genomes "*Ca.* T. childressi" and "*Ca.* T. azoricus" and the free-living Campylobacterotum from Manus basin were compared

to the genomes of 41 Campylobacterota representatives. Five Deltaproteobacteria genomes were used as outgroup. Universal marker proteins conserved across all bacteria were screened using Amphora2[90]. Genes present in one copy in every draft genome were selected for the phylogenomic reconstruction (*rpsI , rplT , rpsB , rplM , rpsS , rplK , rplL , frr , rplP , rplA , rplB , pyrG ,rpsM , smpB*). Each gene set was aligned using MUSCLE[91]. The concat_align.pl script (phylogenomics-tools) was used for determining the best protein substitution model of each marker alignment (*rpsI ::*LG*, rplT ::*LG*, rpsB::*LG *, rplM::*LG *, rpsS::*LG *, rplK::*RTREV *, rplL::*LG *, frr ::*LG*, rplP ::*LG*, rplA::*LG *, rplB::*LG *, pyrG::*LG *,rpsM::*LG *, smpB::*LG). To calculate the multi-gene phylogeny, the marker genes from each genome were concatenated. The best tree with SH-like aLRT support value was calculated with RAxML[92] using the tree_calculations.pl script (phylogenomics-tools).

Phylogenetic analysis

The IMG ER pipeline detected genes with a gammaproteobacterial signature based on homologies to sequences in its database. We extracted and analyzed these sequences with the Geneious software version v 9.1.8[93] (http://www.geneious.com). Genes predicted by automated annotations were manually verified and curated using the public databases NCBI, Uniprot and Swissprot. Sequences of interest were compared to the NCBI nucleotide and amino acid databases using nucleotide- and amino acid-BLAST. We retrieved closely-related sequences from the BLASTX results on the NCBI non-redundant database. Additionally, other reference sequences were included in the analysis and all sequences were aligned using MUSCLE (v3.6.)[91]. To detect the best substitution model to use for phylogenetic reconstruction, we used the ProtTest3 package[94] (Model summarized in Supplementary Table 8). Phylogenetic analyses were then performed using Bayesian and Maximum likelihood analyses. Bayesian analysis was performed with MrBayes (v3.2)[95] under a General Time Reversible model with the best-fitted substitution model. Analyses were performed for two million generations using four parallel Monte Carlo Markov chains. Sample

590 trees were taken every 1000 generations. Maximum likelihood trees were calculated with
591 PHYML[96] using the best-fitted substitution model. We used 1000 bootstraps as support
592 values for nodes in the trees.

593 Codon usage analysis

594 The codon usage of "*Ca.* T. azoricus" and "*Ca.* T. childressi" genes was determined
595 with CodonW[53] using default parameters. The Principal Component Analysis was plotted
596 with R (version 3.4.0).

597 Bulk isotope analysis

598 Parts of *"B." childressi* gill tissues were used for bulk stable isotope analysis. Tissue pieces
599 were oven-dried overnight and ground to a fine powder. The dried tissue was weighed and
600 samples (0.3-0.7 mg dry weight) were packaged in tin capsules for mass spectrometry, and
601 analyzed using a Costech (Valencia, CA USA) elemental analyzer interfaced with a
602 continuous flow Micromass (Manchester, UK) Isoprime isotope ratio mass spectrometer (EA-
603 IRMS) for $^{15}N/^{14}N$ and $^{13}C/^{12}C$ ratios. Measurements are reported in δ notation [per mil (‰)
604 units] and ovalbumin was used as a routine standard. Precision for $\delta^{13}C$ and $\delta^{15}N$ was ±
605 0.2 ‰ and ± 0.4 ‰.

606 Protein extraction and peptide preparation

607 Parts of the gills (see Supplementary Table 7) of three "*B." childressi* specimen were used to
608 prepare tryptic digests following the filter-aided sample preparation (FASP) protocol of
609 Wisniewski et al.[97] with minor modifications[55]. Prior to FASP, cells were disrupted by beat-
610 beating samples in SDT lysis buffer (4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT)
611 using lysing matrix D tubes (MP Biomedicals) before heating to 95 °C for 10 minutes.

612 To allow binding of peptides to the SCX column for 2D-LC methods, peptides were desalted
613 using Sep-Pak C18 Plus Light Cartridges (Waters) according to the manufacturer's
614 instructions. A centrifugal vacuum concentrator was used to exchange acetonitrile after

615 peptide elution with 0.2% (v/v) formic acid. The Pierce Micro BCA assay (Thermo Scientific)
616 was used to determine peptide concentrations, following the manufacturer's instructions.

617 1D and 2D LC-MS/MS

618 All three samples were analyzed by 1D-LC-MS/MS and 2D-LC-MS/MS as described in
619 Kleiner et al.[39]. Briefly, sample analysis via 1D-LC-MS/MS was run twice. An UltiMate™ 3000
620 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) was used to load 1.5-3 µg
621 peptide with loading solvent A (2% acetonitrile, 0.05% trifluoroacetic acid) onto a 5 mm, 300
622 µm ID C18 Acclaim® PepMap100 pre-column (Thermo Fisher Scientific). Peptides were
623 eluted from the pre-column onto a 50 cm x 75 µm analytical EASY-Spray column packed
624 with PepMap RSLC C18, 2 µm material (Thermo Fisher Scientific) heated to 45 °C. An Easy-
625 Spray source connected the analytical column to a Q Exactive Plus hybrid quadrupole-
626 Orbitrap mass spectrometer (Thermo Fisher Scientific). Separation of peptides on the
627 analytical column was achieved at a flow rate of 225 nl min^{-1} using a 460 min gradient going
628 from 98% buffer A (0.1% formic acid) to 31% buffer B (0.1% formic acid, 80% acetonitrile) in
629 363 min, then to 50% B in 70 min, to 99% B in 1 min and ending with 99% B. Electrospray
630 ionization (ESI) was used to ionize eluting peptides. Carryover was reduced by two wash
631 runs (injection of 20 µl acetonitrile, 99% eluent B) and one blank run between samples. Data
632 acquisition with the Q Exactive Plus was done as in [98].

633 The 2D-LC-MS/MS experiments were performed as described by Kleiner et al.[39] with the
634 modification that pH plugs instead of NaCl salt plugs were used for peptide elution from the
635 SCX column. Briefly, 4.5 µg of peptide were loaded with loading solvent B (2% acetonitrile,
636 0.5% formic acid) onto a 10 cm, 300 µm ID Poros 10 S SCX column (Thermo Fisher
637 Scientific) at a flow rate of 5 µl min^{-1} using the same LC as for 1D-LC-MS/MS. Peptides that
638 did not bind to the SCX column were captured by the C18 pre-column (same as for 1D-LC),
639 which was in-line downstream of the SCX column. The C18 pre-column was then switched
640 in-line with the 50 cm x 75 µm analytical column (same as for 1D) and the breakthrough

separated using a gradient of eluent A and B (2% B to 31% B in 82 min, 50% B in 10 min, 99% B in 1 min, holding 99% B for 7 min, back to 2% B in 1 min, holding 2% B for 19 min). Peptides were eluted step-wise from the SCX to the C18 pre-column by injecting 20 µl of pH buffers with increasing pH (pH 2.5-pH 8, CTIBiphase buffers, Column Technology Inc.) from the autosampler. After each pH plug, the C18 pre-column was again switched in-line with the analytical column and peptides separated as above. Between samples, the SCX column was washed twice (injection of 20 µl 4 M NaCl in loading solvent B, 100% eluent B), the RP column once (injection of 20 µl acetonitrile, 99% eluent B) and a blank run was done to reduce carryover. Data was acquired with the Q Exactive Plus as in [98].

Protein identification and quantification

A database containing protein sequences predicted from the metatranscriptomic and -genomic data of the "*B." childressi* symbiosis generated in this study was used for protein identification as described in the 'Metagenome assembly' section above. The cRAP protein sequence database (http://www.thegpm.org/crap/), which contains sequences of common lab contaminants, was appended to the database. The final database contained 38,418 protein sequences. For protein identification, MS/MS spectra were searched against this database using the Sequest HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific) as in[33].

To quantify proteins, normalized spectral abundance factors (NSAFs)[99] were calculated per species and multiplied by 100, to give the relative protein abundance in %. For biomass calculations, the method described by Kleiner et al.[39] was used. Calculations of NSAFs and biomass for each sample were based on the combined data from both 1D-LC-MS/MS runs and the one 2D-LC-MS/MS run.

Direct Protein-SIF

Stable carbon isotope fingerprints (SIFs) for *"B." childressi* and its symbionts were determined as described by Kleiner et al.[33]. Human hair with a known $\delta^{13}C$ value was used as a reference to correct for instrument fractionation. A tryptic digest of the reference material was prepared as described above and with the same 1D-LC-MS/MS method as the samples. Due to the low abundance of the "*Ca.* Thiobarba" symbiont in terms of biomass the six 1D-LC-MS/MS datasets (technical replicate runs of three gill samples) were combined in one peptide identification search to obtain enough peptides for SIF estimation. For peptide identification, MS/MS spectra were searched against the database using the Sequest HT node in Proteome Discoverer version 2.0.0.802 (Thermo Fisher Scientific) and peptide spectral matches were filtered using the Percolator node as described by Petersen et al.[98]. The peptide-spectrum match (PSM) files generated by Proteome Discoverer were exported in tab-delimited text format. The 1D-LC-MS/MS raw files were converted to mzML format using the MSConvertGUI available in the ProteoWizard tool suite[100]. Only the MS^1 spectra were retained in the mzML files and the spectra were converted to centroided data by Vendor algorithm peak picking. The PSM and mzML files were used as input for the Calis-p software (https://sourceforge.net/projects/calis-p/) to extract peptide isotope distributions and to compute the direct Protein-SIF $\delta^{13}C$ value for each species[33]. The direct Protein-SIF $\delta^{13}C$ values were corrected for instrument fragmentation by applying the offset determined by comparing the direct Protein-SIF $\delta^{13}C$ value of the reference material with its known $\delta^{13}C$ value.

Data availability

The metagenomic and metatranscriptomic raw reads are available in the European Nucleotide Archive under Study Accession Number: ERZ772703, PRJEB23286, PRJEB23284 and PRJEB19882.

References

1. Rosgaard, L., de Porcellinis, A. J., Jacobsen, J. H., Frigaard, N.-U. & Sakuragi, Y. Bioengineering of carbon fixation, biofuels, and biochemicals in cyanobacteria and plants. *J. Biotechnol.* **162,** 134–147 (2012).

2. Hügler, M. & Sievert, S. M. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Ann. Rev. Mar. Sci.* **3,** 261–289 (2011).

3. Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S. & Erb, T. J. A synthetic pathway for the fixation of carbon dioxide in vitro. *Science* **354,** 900–904 (2016).

4. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO_2 fixation pathway. *Proc. Natl. Acad. Sci.* **115,** E92–E101 (2018).

5. Raven, J. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat. Microb. Ecol.* **56,** 177–192 (2009).

6. Falkowski, P. G., Fenchel, T. & Delong, E. F. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320,** 1034–1039 (2008).

7. Bassham, J. A. *et al.* The Path of Carbon in Photosynthesis. XXI. The Cyclic Regeneration of Carbon Dioxide Acceptor 1. *J. Am. Chem. Soc.* **76,** 1760–1770 (1954).

8. Raven, J. A. Rubisco: still the most abundant protein of Earth? *New Phytol.* **198,** 1–3 (2013).

9. Erb, T. J. & Zarzycki, J. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO_2 fixing enzyme. *Curr. Opin. Biotechnol.* **49,** 100–107 (2018).

10. Evans, M. C., Buchanan, B. B. & Arnon, D. I. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proc. Natl. Acad. Sci.* **55,** 928–934 (1966).

11. Mall, A. *et al.* Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* **359,** 563–567 (2018).

12. Nunoura, T. *et al.* A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* **359,** 559–563 (2018).

13. Waite, D. W. *et al.* Comparative Genomic Analysis of the Class Epsilonproteobacteria and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). *Front. Microbiol.* **8,** (2017).

14. Waite, D. W. *et al.* Addendum: Comparative Genomic Analysis of the Class Epsilonproteobacteria and Proposed Reclassification to Epsilonbacteraeota (phyl. nov.). *Front. Microbiol.* **9,** (2018).

15. Smith, C. Chemosynthesis in the deep-sea: life without the sun. *Biogeosciences Discuss.* **9,** 17037–17052 (2012).

16. Van Dover, C. L. *The Ecology of Deep-sea Hydrothermal Vents*. (Princeton University Press, 2000).

17. Campbell, B. J., Engel, A. S., Porter, M. L. & Takai, K. The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* **4,** 458–68 (2006).

18. Nakagawa, S. & Takai, K. Deep-sea vent chemoautotrophs: Diversity, biochemistry and ecological significance. *FEMS Microbiol. Ecol.* **65,** 1–14 (2008).

19. Sievert, S. & Vetriani, C. Chemoautotrophy at Deep-Sea Vents: Past, Present, and Future. *Oceanography* **25,** 218–233 (2012).

20. Reeves, E. P. *et al.* Microbial lipids reveal carbon assimilation patterns on hydrothermal sulfide chimneys. *Environ. Microbiol.* **16,** 3515–3532 (2014).

21. Pop Ristova, P., Wenzhöfer, F., Ramette, A., Felden, J. & Boetius, A. Spatial scales of bacterial community diversity at cold seeps (Eastern Mediterranean Sea). *ISME J.* **9,** 1–13 (2014).

22. Sievert, S. M., Hügler, M., Taylor, C. D. & Wirsen, C. O. Sulfur Oxidation at Deep-Sea Hydrothermal Vents in *Microbial Sulfur Metabolism* 238–258 (Springer Berlin Heidelberg, 2008).

23. Meier, D. V *et al.* Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. *ISME J.* **11,** 1545–1558 (2017).

24. Marshall, K. T. & Morris, R. M. Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J.* **7,** 452–455 (2013).

25. Inagaki, F., Takai, K., Nealson, K. H. & Horikoshi, K. *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the E-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. *Int. J. Syst. Evol. Microbiol.* **54,** 1477–1482 (2004).

26. Yamamoto, M. & Takai, K. Sulfur Metabolisms in Epsilon- and Gamma-Proteobacteria in Deep-Sea Hydrothermal Fields. *Front. Microbiol.* **2,** 192 (2011).

27. Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6,** 725–40 (2008).

28. Galkin, S. V. Structure of Hydrothermal Vent Communities in *Trace Metal Biogeochemistry and Ecology of Deep-Sea Hydrothermal Vent Systems* (eds. Galkin, S. V. & Demina, L. L.) 41–53 (Springer International Publishing Switzerland, 2016).

29. Petersen, J. M. *et al.* Dual symbiosis of the vent shrimp *Rimicaris exoculata* with filamentous gamma- and epsilonproteobacteria at four Mid-Atlantic Ridge hydrothermal vent fields. *Environ. Microbiol.* **12,** 2204–2218 (2010).

30. Beinart, R. A. *et al.* Evidence for the role of endosymbionts in regional-scale habitat partitioning by hydrothermal vent symbioses. *Proc. Natl. Acad. Sci.* **109,** E3241–E3250 (2012).

31. Duperron, S., Lorion, J., Samadi, S., Gros, O. & Gaill, F. Symbioses between deep-sea mussels (Mytilidae: Bathymodiolinae) and chemosynthetic bacteria: diversity, function and evolution. *C. R. Biol.* **332,** 298–310 (2009).

32. Assié, A. *et al.* A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of Epsilonproteobacteria. *Environ. Microbiol. Rep.* **8,** 805–813 (2016).

33. Kleiner, M. *et al.* Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc. Natl. Acad. Sci.* **115,** E5576–E5584 (2018).

34. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57,** 81–91 (2007).

35. Konstantinidis, K. T. & Tiedje, J. M. Towards a Genome-Based Taxonomy for Prokaryotes. *J. Bacteriol.* **187,** 6258–6264 (2005).

36. Yarza, P. *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat. Rev. Microbiol.* **12,** 635–645 (2014).

37. Rodriguez-R, L. M. & Konstantinidis, K. T. Bypassing cultivation to identify bacterial species. *Microbe* **9,** 111–118 (2014).

38. Rosenthal, B. *et al.* Evidence for the bacterial origin of genes encoding fermentation enzymes of the amitochondriate protozoan parasite *Entamoeba histolytica*. *J. Bacteriol.* **179,** 3736–3745 (1997).

39. Kleiner, M. *et al.* Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* **8,** 1558 (2017).

40. Pearson, A. Pathways of Carbon Assimilation and Their Impact on Organic Matter Values $\delta^{13}C$ in *Handbook of Hydrocarbon and Lipid Microbiology* (ed. Timmis, K. N.) 143–156 (Springer Berlin Heidelberg, 2010).

41. MacAvoy, S. E., Carney, R. S., Morgan, E. & Macko, S. a. Stable Isotope Variation Among the Mussel *Bathymodiolus childressi* and Associated Heterotrophic Fauna at Four Cold-Seep Communities in the Gulf of Mexico. *J. Shellfish Res.* **27,** 147–151 (2008).

42. Sassen, R. *et al.* Thermogenic gas hydrates and hydrocarbon gases in complex chemosynthetic communities, Gulf of Mexico continental slope. *Org. Geochem.* **30,** 485–497 (1999).

43. Petersen, J. M. & Dubilier, N. Methanotrophic symbioses in marine invertebrates. *Environ. Microbiol. Rep.* **1,** 319–335 (2009).

44. Riekenberg, P., Carney, R. & Fry, B. Trophic plasticity of the methanotrophic mussel *Bathymodiolus childressi* in the Gulf of Mexico. *Mar. Ecol. Prog. Ser.* **547,** 91–106 (2016).

45. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM : assessing the quality of microbial genomes recovered from isolates , single cells , and metagenomes. *Genome Res.* **25,** 1043–1055 (2015).

46. Li, Y., Liles, M. R. & Halanych, K. M. Endosymbiont genomes yield clues of tubeworm success. *ISME J.* (2018).

47. Robidart, J. C. *et al.* Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. *Environ. Microbiol.* **10,** 727–737 (2008).

48. Markert, S. *et al.* Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila*. *Science* **315,** 247–250 (2007).

49. Winkel, M. *et al.* Single-cell Sequencing of Thiomargarita Reveals Genomic Flexibility for Adaptation to Dynamic Redox Conditions. *Front. Microbiol.* **7,** (2016).

50. Flood, B. E. *et al.* Single-Cell (Meta-)Genomics of a Dimorphic Candidatus Thiomargarita nelsonii Reveals Genomic Plasticity. *Front. Microbiol.* **7,** (2016).

51. MacGregor, B. J., Biddle, J. F., Harbort, C., Matthysse, A. G. & Teske, A. Sulfide oxidation, nitrate respiration, carbon acquisition, and electron transport pathways suggested by the draft genome of a single orange Guaymas Basin Beggiatoa (Cand. Maribeggiatoa) sp. filament. *Mar. Genomics* **11,** 53–65 (2013).

52. Rubin Blum, M., Dubilier, N. & Kleiner, M. Genetic evidence for two carbon fixation pathways in symbiotic and free-living bacteria: The Calvin-Benson-Bassham cycle and the reverse tricarboxylic acid cycle. *bioRxiv* (2018).

53. Peden, J. F. Analysis of codon usage. *Biosystems.* **106,** 45–50 (2011).

54. Kleiner, M., Young, J. C., Shah, M., Verberkmoes, N. C. & Dubilier, N. Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *MBio* **4,** (2013).

55. Hamann, E. *et al.* Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* **534,** 254–258 (2016).

56. Zhang, Y. & Sievert, S. M. Pan-genome analyses identify lineage- and niche-specific markers of evolution and adaptation in Epsilonproteobacteria. *Front. Microbiol.* **5,** 110 (2014).

57. Hugler, M., Wirsen, C. O., Fuchs, G., Taylor, C. D. & Sievert, S. M. Evidence for Autotrophic CO_2 Fixation via the Reductive Tricarboxylic Acid Cycle by Members of the ε-Subdivision of Proteobacteria. *J. Bacteriol.* **187,** 3020–3027 (2005).

58. Ragsdale, S. W. Pyruvate ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* **103,** 2333–2346 (2003).

59. Imlay, J. A. Iron-sulphur clusters and the problem with oxygen. *Mol. Microbiol.* **59,** 1073–1082 (2006).

60. Berg, I. A. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl. Environ. Microbiol.* **77,** 1925–1936 (2011).

61. Pjevac, P. *et al.* Metaproteogenomic Profiling of Microbial Communities Colonizing Actively Venting Hydrothermal Chimneys. *Front. Microbiol.* **9,** (2018).

62. Kalenitchenko, D. *et al.* Ecological succession leads to chemosynthesis in mats colonizing wood in sea water. *ISME J.* **10,** 2246–2258 (2016).

63. Smillie, C. S. *et al.* Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* **480,** 241–244 (2011).

64. Gilbreath, J. J., Cody, W. L., Merrell, D. S. & Hendrixson, D. R. Change Is Good: Variations in Common Biological Mechanisms in the Epsilonproteobacterial Genera *Campylobacter* and *Helicobacter*. *Microbiol. Mol. Biol. Rev.* **75,** 84–132 (2011).

65. Porcelli, I., Reuter, M., Pearson, B. M., Wilhelm, T. & van Vliet, A. H. M. Parallel evolution of genome structure and transcriptional landscape in the Epsilonproteobacteria. *BMC Genomics* **14,** 616 (2013).

66. Anderson, R. E. *et al.* Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea hydrothermal vents. *Nat. Commun.* **8,** 1114 (2017).

67. Nakagawa, S. & Takaki, Y. Nonpathogenic Epsilonproteobacteria in *Encyclopedia of Life Sciences* 1–11 (John Wiley & Sons, Ltd, 2009).

68. Nakagawa, S. *et al.* Deep-sea vent -proteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl. Acad. Sci.* **104,** 12146–12150 (2007).

69. Antonovsky, N. *et al.* Sugar Synthesis from CO_2 in *Escherichia coli*. *Cell* **166,** 115–125 (2016).

70. Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci.* **107,** 8889–8894 (2010).

71. Zwirglmaier, K. *et al.* Linking regional variation of epibiotic bacterial diversity and trophic ecology in a new species of Kiwaidae (Decapoda, Anomura) from East Scotia Ridge (Antarctica) hydrothermal vents. *Microbiologyopen* **4,** 136–150 (2015).

72. Rogge, A. *et al.* Success of chemolithoautotrophic SUP05 and Sulfurimonas GD17 cells in pelagic Baltic Sea redox zones is facilitated by their lifestyles as K- and r -strategists. *Environ. Microbiol.* **19,** 2495–2506 (2017).

73. Pjevac, P., Kamyshny, A., Dyksma, S. & Mußmann, M. Microbial consumption of zero-valence sulfur in marine benthic habitats. *Environ. Microbiol.* **16,** 3416–3430 (2014).

74. Kreimer, A., Borenstein, E., Gophna, U. & Ruppin, E. The evolution of modularity in bacterial metabolic networks. *Proc. Natl. Acad. Sci.* **105,** 6976–6981 (2008).

75. Levin, L. A., Mendoza, G. F., Konotchick, T. & Lee, R. Macrobenthos community structure and trophic relationships within active and inactive Pacific hydrothermal sediments. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **56,** 1632–1648 (2009).

76. Zhang, C. L. *et al.* Lipid biomarkers and carbon-isotopes of modern travertine deposits (Yellowstone National Park, USA): Implications for biogeochemical dynamics in hot-spring systems. *Geochim. Cosmochim. Acta* **68,** 3157–3169 (2004).

77. Kelley, C. a, Coffin, R. B. & Cifuentes, L. a. Stable isotope evidence for alternative bacterial carbon sources in the Gulf of Mexico. *Limnol. Oceanogr.* **43,** 1962–1969 (1998).

78. Fry, B. & Sherr, E. B. $\delta^{13}C$ Measurements as Indicators of Carbon Flow in Marine and Freshwater Ecosystems in *Stable Isotopes in Ecological Research* (ed. Rundel P.W., Ehleringer J.R., N. K. A.) 196–229 (Springer, New York, NY, 1989).

79. Zhou, J., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62,** 316–322 (1996).

80. Li, D., Liu, C. M., Luo, R., Sadakane, K. & Lam, T. W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31,** 1674–1676 (2014).

81. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic

892 contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3,** 1–11 (2012).

893 82. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications
894 to Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

895 83. Aziz, R. K. *et al.* The RAST Server: Rapid Annotations using Subsystems Technology.
896 *BMC Genomics* **9,** 75 (2008).

897 84. Meyer, F. *et al.* The metagenomics RAST server – a public resource for the automatic
898 phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9,** 386
899 (2008).

900 85. Markowitz, V. M. *et al.* IMG 4 version of the integrated microbial genomes
901 comparative analysis system. *Nucleic Acids Res.* **42,** D560–D567 (2014).

902 86. Rodriguez-R, L. M. & Konstantinidis, K. T. The enveomics collection : a toolbox for
903 specialized analyses of microbial genomes and metagenomes. *Peer J Prepr.* **4,**
904 e1900v1 (2016).

905 87. Rubin-Blum, M. *et al.* Short-chain alkanes fuel mussel and sponge *Cycloclasticus*
906 symbionts from deep-sea gas and oil seeps. *Nat. Microbiol.* **2,** 17093 (2017).

907 88. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program
908 for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930
909 (2014).

910 89. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
911 differential expression analysis of digital gene expression data. *Bioinformatics* **26,**
912 139–140 (2010).

913 90. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with
914 AMPHORA2. *Bioinformatics* **28,** 1033–1034 (2012).

915 91. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high
916 throughput. *Nucleic Acid Res.* **32,** 1792–1797 (2004).

917 92. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of
918 large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

919 93. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software
920 platform for the organization and analysis of sequence data. *Bioinformatics* **28,** 1647–
921 9 (2012).

922 94. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-
923 fit models of protein evolution. *Bioinformatics* **27,** 1164–1165 (2011).

924 95. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model
925 Choice Across a Large Model Space. *Syst. Biol.* **61,** 539–542 (2012).

926 96. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood
927 Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59,** 307–321
928 (2010).

929 97. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation
930 method for proteome analysis. *Nat. Methods* **6,** 359–362 (2009).

931 98. Petersen, J. M. *et al.* Chemosynthetic symbionts of marine invertebrate animals are

capable of nitrogen fixation. *Nat. Microbiol.* **2,** 16195 (2016).

99. Zybailov, B. *et al.* Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5,** 2339–2347 (2006).

100. Chambers, M. C. *et al.* A cross-platform toolkit for mass spectrometry and proteomics. *Nat. Biotechnol.* **30,** 918–920 (2012).

101. Vizcaíno, J. A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44,** D447–D456 (2016).

Acknowledgements

We would like to thank the crew and captains of the scientific vessels of the following cruises: Atlantis cruise AT26-13, Nautilus cruises NA044 and NA058 and the “Pourquoi Pas?” Biobaz cruise. We also thank the Max Planck Genome Center in Cologne for the genome and transcriptome sequencing. We are grateful to Lizbeth Sayavedra for transcriptome assembly advice and support. We also thank Brandon Seah for his help with the naming of “*Ca*. Thiobarba”. We thank Marc Strous for access to proteomics equipment. The purchase of the proteomics equipment was supported by a grant of the Canadian Foundation for Innovation to Marc Strous.

This work was funded by the Max Planck Society, the DFG Cluster of Excellence ‘The Ocean in the Earth System’ at MARUM (University of Bremen), a European Research Council Advanced Grant (BathyBiome, Grant 340535) and a Gordon and Betty Moore Foundation Marine Microbiology Initiative Investigator Award through Grant GBMF3811 to ND, the European Union (EU) Marie Curie Actions Initial Training Network (ITN) SYMBIOMICS (contract number 264774) and a NSF award #0801741 to Dr. Samantha Joye to fund the AT26-13 expedition to the Gulf of Mexico. JMP was supported by the Vienna Science and Technology Fund (WWTF) through project VRG14-021. MK was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada through a Banting fellowship and the NC State Chancellor’s Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities. TH was supported through a fellowship of the German Academic Exchange Service DAAD.

Author contributions

A.A., N.L., J.P. and N.D. conceived and developed the study. A.A. and N.L analyzed data. A.A. and H.G.V. performed the metagenomic assemblies. A.A. performed transcriptome and genome analyses as well as phylogenetic reconstructions. M.K and T.H performed proteomic analyses. M.K. developed and performed isotopic fingerprinting method. N.L and A.A provided bulk Isotope analysis. N.L provided key support for isotope work and interpretations. A.M. and D.V.M. provided environmental genomic bin. H.E.T. performed genomic sequencing. S.J and M.S provided key biological samples. A.A. and N.L. wrote the manuscript with support from J.P. and N.D. All authors discussed the results and contributed to the final manuscript.