# Chemosynthetic symbiont with a drastically reduced genome serves as primary energy storage in the marine flatworm *Paracatenula*

Oliver Jäckle[a], Brandon K. B. Seah[a], Målin Tietjen[a], Nikolaus Leisch[a], Manuel Liebeke[a], Manuel Kleiner[b,c], Jasmine S. Berg[a,d], and Harald R. Gruber-Vodicka[a,1]

[a]Max Planck Institute for Marine Microbiology, 28359 Bremen, Germany; [b]Department of Geoscience, University of Calgary, AB T2N 1N4, Canada; [c]Department of Plant & Microbial Biology, North Carolina State University, Raleigh, NC 27695; and [d]Institut de Minéralogie, Physique des Matériaux et Cosmochimie, Université Pierre et Marie Curie, 75252 Paris Cedex 05, France

Edited by Margaret J. McFall-Ngai, University of Hawaii at Manoa, Honolulu, HI, and approved March 1, 2019 (received for review November 7, 2018)

**Hosts of chemoautotrophic bacteria typically have much higher biomass than their symbionts and consume symbiont cells for nutrition. In contrast to this, chemoautotrophic *Candidatus* Riegeria symbionts in mouthless *Paracatenula* flatworms comprise up to half of the biomass of the consortium. Each species of *Paracatenula* harbors a specific *Ca*. Riegeria, and the endosymbionts have been vertically transmitted for at least 500 million years. Such prolonged strict vertical transmission leads to streamlining of symbiont genomes, and the retained physiological capacities reveal the functions the symbionts provide to their hosts. Here, we studied a species of *Paracatenula* from Sant'Andrea, Elba, Italy, using genomics, gene expression, imaging analyses, as well as targeted and untargeted MS. We show that its symbiont, *Ca*. R. santandreae has a drastically smaller genome (1.34 Mb) than the symbiont´s free-living relatives (4.29–4.97 Mb) but retains a versatile and energy-efficient metabolism. It encodes and expresses a complete intermediary carbon metabolism and enhanced carbon fixation through anaplerosis and accumulates massive intracellular inclusions such as sulfur, polyhydroxyalkanoates, and carbohydrates. Compared with symbiotic and free-living chemoautotrophs, *Ca*. R. santandreae's versatility in energy storage is unparalleled in chemoautotrophs with such compact genomes. Transmission EM as well as host and symbiont expression data suggest that *Ca*. R. santandreae largely provisions its host via outer-membrane vesicle secretion. With its high share of biomass in the symbiosis and large standing stocks of carbon and energy reserves, it has a unique role for bacterial symbionts—serving as the primary energy storage for its animal host.**

endosymbiont | chemoautotrophic | ecophysiology | energy storage | host nutrition

The discovery of dense animal communities at deep-sea hydrothermal vents in the late 1970s led to paradigm shifts in deep-sea ecology and animal physiology. Large macrofauna was shown to be independent of photoautotrophic production and instead lives off primary production performed by chemosynthetic bacterial symbionts (1–3). In the most extreme forms, animals such as the giant tubeworms have lost their mouth and digestive system and instead are nutritionally dependent on their symbionts (4).

Since then, chemosynthetic symbioses have been documented from diverse habitats and in various lineages of marine invertebrates and protists (5, 6). The symbionts are all *Proteobacteria*, typically *Gamma-* or *Epsilonproteobacteria*, that are either vertically transmitted or horizontally acquired from the environment (5–12). Genomes from chemosynthetic symbionts can range from 4.88 to 1.02 Mb, reflecting varying metabolic and ecological strategies (13–15). Surprisingly, some mouth- and gutless hosts take up their symbionts anew in every generation, although the lack of a digestive system suggests extreme dependence on their symbionts (12, 16). As these symbionts have to thrive in both free-living environmental and symbiotic states, it is difficult to attribute their genomic features to either functions they provide to their host, or traits that are necessary for environmental survival or to both.

The smallest genomes of chemoautotrophic symbionts have been observed for the gammaproteobacterial symbionts of vesicomyid clams that are directly transmitted between host generations (13, 14). Such strict vertical transmission leads to substantial and ongoing genome reduction. Reduced genomes of intracellular endosymbionts reflect the essential set of functions that the symbionts provide to their invertebrate hosts (17–23). Host provisioning by chemosynthetic bacteria is accomplished through a conveyor belt-like turnover of symbiont cells and compared with their hosts, they only have a small share of the total biomass in the consortium (24). The patterns of genome reduction in the symbionts of vesicomyid clams mirror this mode of nutrition since they

## Significance

Animals typically store their primary energy reserves in specialized cells. Here, we show that in the small marine flatworm *Paracatenula*, this function is performed by its bacterial chemosynthetic symbiont. The intracellular symbiont occupies half of the biomass in the symbiosis and has a highly reduced genome but efficiently stocks up and maintains carbon and energy, particularly sugars. The host rarely digests the symbiont cells to access these stocks. Instead, the symbionts appear to provide the bulk nutrition by secreting outer-membrane vesicles. This is in contrast to all other described chemosynthetic symbioses, where the hosts continuously digest full cells of a small and ideally growing symbiont population that cannot provide a long-term buffering capacity during nutrient limitation.

Author contributions: O.J., N.L., and H.R.G.-V. designed research; O.J., B.K.B.S., M.T., N.L., M.L., M.K., J.S.B., and H.R.G.-V. performed research; O.J., B.K.B.S., M.T., N.L., M.L., M.K., J.S.B., and H.R.G.-V. analyzed data; and O.J. and H.R.G.-V. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.



Data deposition: The assembled and annotated *Ca*. R. santandreae genome has been deposited in the European Nucleotide Archive (ENA) and is accessible under project no. PRJEB26644 and the assembly under accession no. LR026963. Raw total RNA library reads and the host assembly have been deposited in ENA under project no. PRJEB31702 via The German Federation for the Curation of Biological Data (GFBio). The MS proteomics data and the protein sequence databases have been deposited in the ProteomeXchange Consortium via the Proteomics Identifications (PRIDE) partner repository for the pure culture data with the dataset identifier PXD009856. The raw data for the EM based quantitative image analysis have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.7746806).

[1]To whom correspondence should be addressed. Email: hgruber@mpi-bremen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1818995116/-/DCSupplemental.

Published online April 8, 2019.

only retained limited metabolic and storage capacities and primarily serve as biosynthesis factories (13, 14, 18).

A second host group that lives in obligate symbiosis with vertically transmitted chemosynthetic endosymbionts is *Paracatenula*, a genus of marine flatworms that can be a dominant member of the meiofauna in shallow-water sediments (9). Unlike the vesicomyid clams that can still filter feed (3, 25), *Paracatenula* lack both a mouth and gut and are completely dependent on intracellular endosymbionts for nutrition (9, 26). All *Paracatenula* host species harbor a species-specific symbiont phylotype (9). These symbionts, called *Candidatus* Riegeria, comprise a third to a half of the total animal volume and are housed in vacuoles of specialized cells, the bacteriocytes, that make up the nutritive trophosome organ (9, 26, 27). *Ca.* Riegeria are the only known alphaproteobacterial chemosynthetic symbionts (9). They are the oldest extant clade of symbiotic chemoautotrophic bacteria, having been passed on in strict vertical transmission for more than 500 million years since the last common ancestor of all *Paracatenula* species (9).

In this study, we focus on a *Paracatenula* species that is highly abundant in the coarse shallow-water sediments of the bay Sant'Andrea on Elba, Italy (*SI Appendix*, *Supplementary Note 1* and Fig. S1). We propose the name *Ca.* Riegeria santandreae for its endosymbionts, in reference to the collection site. Based on the genome of *Ca.* R. santandreae that we generated using shotgun metagenomics, we analyzed its metabolism, ecophysiology, and evolution. Compared with free-living alphaproteobacterial relatives from the *Rhodospirillaceae*, the genome is highly reduced. We reconstructed the physiology of *Ca.* R. santandreae by integrating transcriptomics and correlative imaging approaches with highly sensitive MS-based methods for proteomics and metabolomics. Despite the reduced genome, the symbiont performs energy-efficient sulfide oxidation and couples it to versatile means of carbon fixation and carbon storage. The symbionts likely provision their host via outer-membrane vesicle (OMV) secretion. With its large share of the biomass in the consortium, the symbiont has a unique role in animal-bacteria associations—serving as the bulk storage of carbon and energy reserves for its animal host.

## Results and Discussion

**_Ca._ R. santandreae Belongs to _Rhodospirillaceae_ and Has a Reduced Genome.** The complete *Ca.* R. santandreae genome was assembled into one contig from the sequenced metagenome of a single *Paracatenula* sp. santandrea specimen. The genome had a size of 1.34 Mb, encoding 1,344 protein-coding and 56 RNA genes at a coding density of 83.9% and a genomic GC of 51.8% (*SI Appendix*, Fig. S2). The *Ca.* R. santandreae genome encoded pathways for sulfur oxidation based on reverse-acting dissimilatory sulfate reductase (rDSR) and for carbon fixation using the Calvin–Benson–Bassham (CBB) cycle, indicating that *Ca.* R. santandreae could acquire energy and carbon via thioautotrophy (*SI Appendix*, *Supplementary Note 2*).

*Ca.* R. santandreae was classified as a member of the family *Rhodospirillaceae* within the class *Alphaproteobacteria*, based on a phylogenetic analysis of 43 conserved single-copy marker genes (Fig. 1). It was phylogenetically nested among *Rhodospirillaceae* taxa that have genome sizes of 4.29–4.97 Mb, indicating that massive genome reduction has occurred in the lineage leading to *Ca.* R. santandreae (Fig. 1 and *SI Appendix*, Table S1). The *Rhodospirillaceae* are remarkably diverse and include photoheterotrophs and chemoheterotrophs but also chemoautotrophs, diazotrophs, and even magnetotactic genera (28–31). The taxa most closely related to the *Ca.* R. santandreae endosymbiont included *Rhodospirillum rubrum* [56.19% amino acid identity (AAI)], *Novispirillum itersonii* (56.22% AAI), *Caenispirillum salinarum* (56.99% AAI), and *Magnetospirillum magneticum* (57.01% AAI).

**_Ca._ R. santandreae Is a Chemoautotroph with a Versatile Carbon Metabolism Despite Its Reduced Genome.** To investigate the physiology of *Ca.* R. santandreae, we applied a combination of genomics,

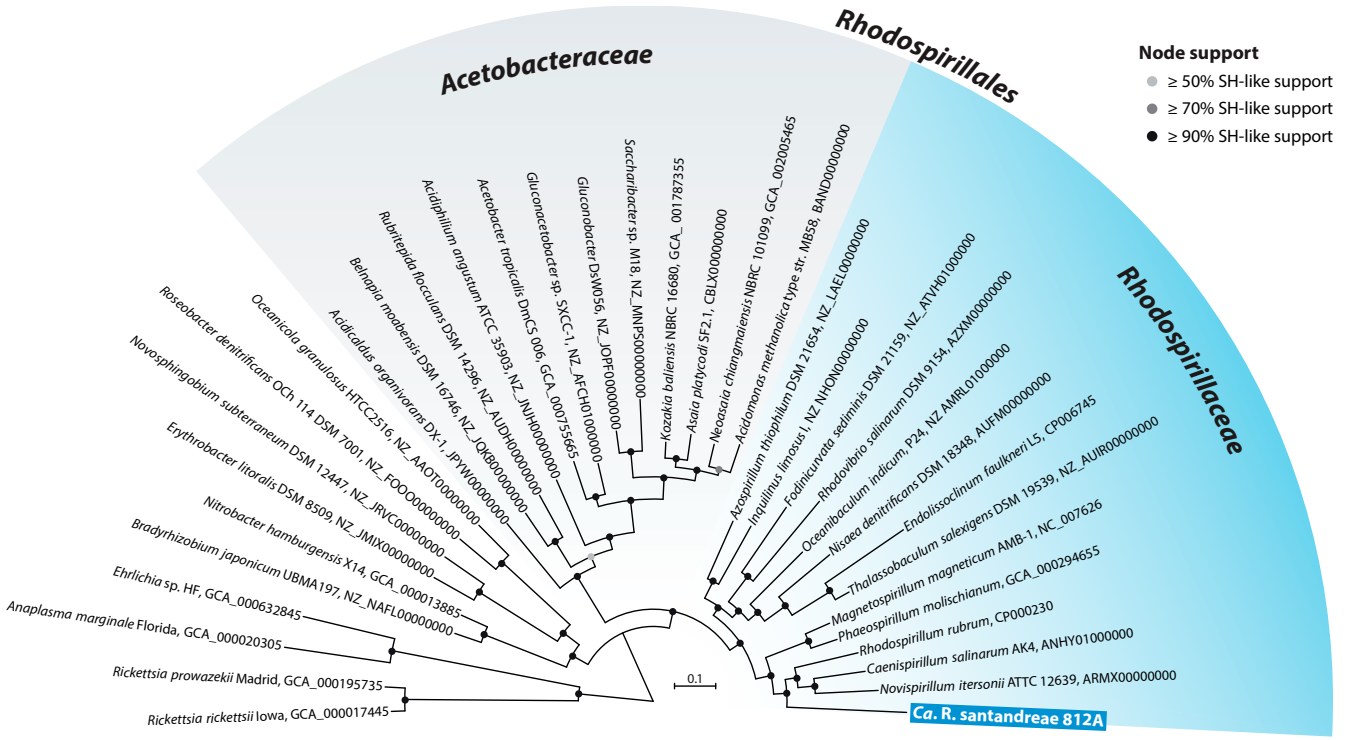


**Fig. 1.** The endosymbiont *Ca.* R. santandreae clusters deeply within *Rhodospirillaceae*. Phylogenetic reconstruction was based on a protein alignment of 43 conserved marker genes calculated using FastTree and rooted with *Rickettsiales* as the outgroup. Scale bar indicates substitutions per site.

transcriptomics, and imaging analyses with sensitive MS protocols that allow proteomic and metabolomic measurements from single microscopic animals. The expression data from specimens sampled directly from the environment covered up to 1,288 out of 1,400 predicted genes ($n = 3$ transcriptomes), of which 407 were also identified in the proteomes generated from single *Paracatenula* individuals as well as pools of up to three specimens ($n = 8$) (Fig. 2 and Dataset S1 A–*E*).

Genes of the rDSR sulfur oxidation pathway and CBB carbon fixation pathways were highly expressed and underline the thioautotrophic role of *Ca.* R. santandreae [sum of 8 CBB cycle proteins: 4.88 ± 1.17% (mean ± SD) normalized spectral abundance factor (NSAF); sum of 14 rDSR pathway proteins: 7.48 ± 0.95% NSAF; $n = 8$ proteomes] (Fig. 2 and *SI Appendix*, Figs. S3 and S4 and *Supplementary Note 2*). CBB cycle genes including the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) subunits *cbbL* and *cbbS* were among the top 10% of most highly represented genes in transcriptomes as well as proteomes (*SI Appendix*, Figs. S3 and S4).

Microautoradiography corroborated the gene expression data, as signals for carbon fixation were limited to the symbiont-bearing trophosome region. Without addition of an external electron donor supply, the *Ca.* R. santandreae symbionts fixed 9.48 ± 10.2 μmol C $g^{-1}$ $h^{-1}$ (maximum: 33 μmol C $g^{-1}$ $h^{-1}$) (*SI Appendix*, Figs. S5–S7 and Dataset S1F). These rates were in the range determined for other sulfur-oxidizing symbionts of, e.g., *Ifremeria nautilei* (0.7 μmol C $g^{-1}$ $h^{-1}$) and *Riftia pachyptila* (27 μmol C $g^{-1}$ $h^{-1}$) (32). The calculated fixation rates would be sufficient to sustain a doubling of the length of a *Paracatenula* sp. santandrea specimen in 10–34 d (Dataset S1F). Such growth rates, together with frequent asexual reproduction (33, 34), could explain their high abundances at the sampling site where they occurred in densities of 233–5,400 individuals per $m^3$ of sediment (1,809 ± 1,483; $n = 40$).

The CBB cycle encoded in *Ca.* R. santandreae was missing fructose-1,6-bisphosphatase and sedoheptulose-1,7-bisphosphatase, which could be replaced by a reversible and $PP_i$-dependent 6-phosphofructokinase (PFK) (35). *Ca.* R. santandreae encoded two PFKs, an ATP-dependent ATP-PFK and a $PP_i$-PFK (*SI Appendix*, Fig. S8). The $PP_i$-PFK had a 10-fold higher expression [transcripts per million (TPM): 5,073 ± 780; $n = 3$] than the ATP-PFK (TPM: 504 ± 128; $n = 3$) and also was among the top 10% of highest expressed genes (*SI Appendix*, Figs. S3 and S4). Together with the $PP_i$-energized proton pump ($PP_i$-$H^+$-pump) (*SI Appendix*, Fig. S9), which was among the 10% most expressed genes (TPM:



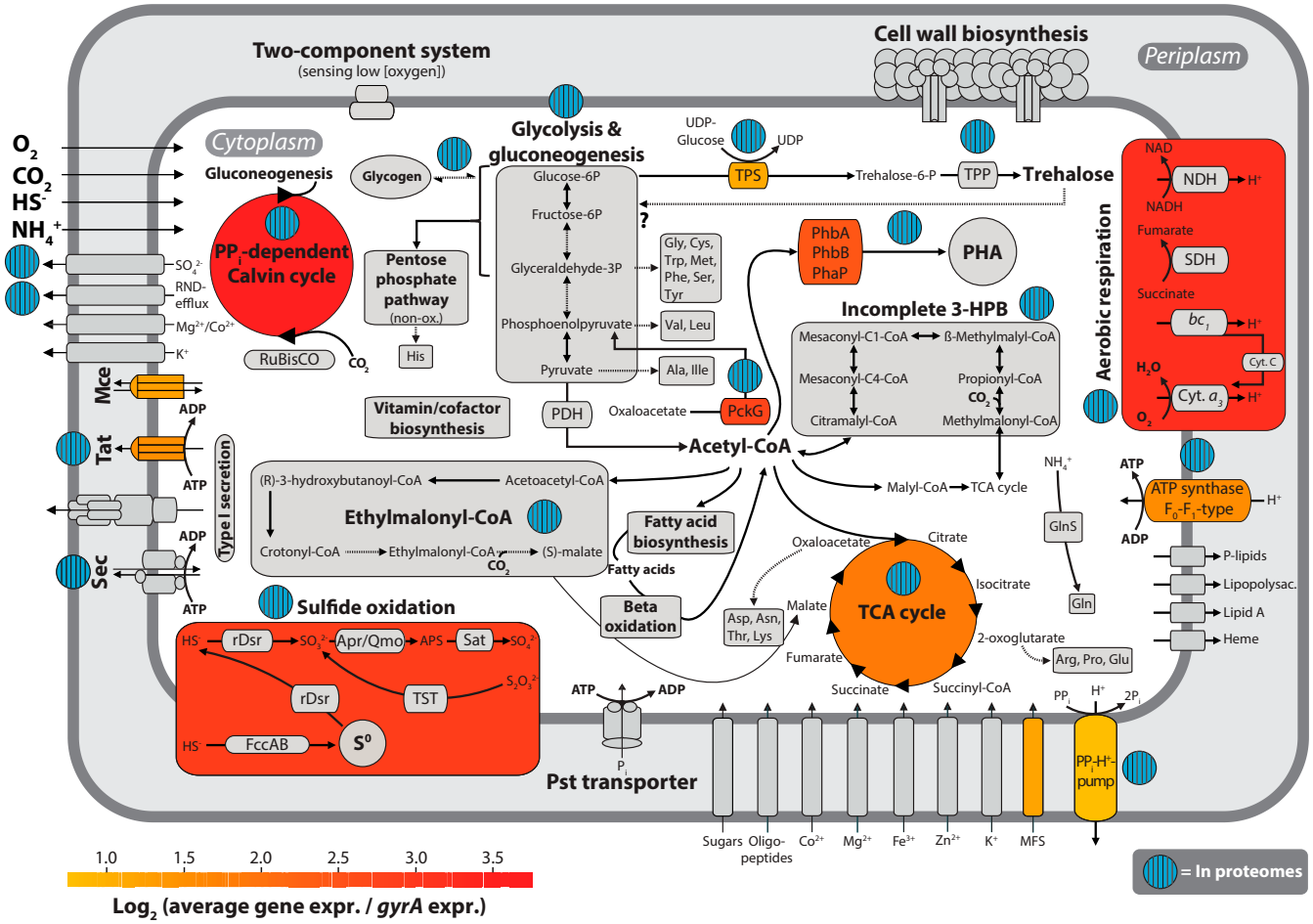


**Fig. 2.** The chemoautotroph *Ca.* R. santandreae expresses a versatile metabolism. Metabolic reconstruction of *Ca.* R. santandreae, pathways, or enzymes are colored by expression levels in the transcriptome (mean of three samples). Only the highest 10% of expressed genes were colored relative to DNA gyrase A (*gyr*A) mean expression. Expressed proteins detected in the proteome samples are indicated. Not all genes of certain pathways were transcribed, and not all transcribed genes were found in the proteomes (Dataset S1 B–E). Dotted arrows correspond to indirect synthesis of metabolites. Apr, sulfite reductase; $bc_1$, cytochrome C oxidase type $bc_1$; Cyt. $a_3$, cytochrome *c* oxidase type $a_3$; Cyt. C, cytochrome *c*; FccAB, flavocytochrome *c* sulfide dehydrogenase; Mce, mammalian cell entry proteins Mla; NDH, NADH-quinone oxidoreductase/dehydrogenase; PckG, phosphoenolpyruvate carboxykinase; PDH, pyruvate dehydrogenase complex; Pst, high-affinity phosphate transporter; Sat, ATP sulfurylase; SDH, succinate dehydrogenase; Sec, secretion pathway; Tat, twin-arginine translocation; TPP, trehalose-P phosphatase; TPS, trehalose-P-synthase; TST, thiosulfate sulfur transferase (rhodanese).

1,686 ± 163; $n = 3$), the $PP_i$-dependent variant of the CBB cycle can save up to 31.5% of the ATP invested in carbon fixation (35). We tested for a possible lateral gene transfer of the $PP_i$-PFK and $PP_i$-$H^+$-pump genes to *Ca*. R. santandreae as the $PP_i$-dependent CBB cycle has also been found in gammaproteobacterial chemosynthetic symbionts, including the symbionts of gutless oligochaetes, vestimentiferans tubeworms, and the bivalve *Solemya* (35–37). In contrast to the *Gammaproteobacteria*, the $PP_i$-PFK and $PP_i$-$H^+$-pump genes in *Ca*. R. santandreae did not form an operon (35). Both genes were of alphaproteobacterial origin and clustered with gene sequences of related *Rhodospirillaceae* (*SI Appendix*, *Supplementary Note 2* and Figs. S8 and S9). Other *Rhodospirillaceae* such as *Rhodospirillum rubrum* only lacked a sedoheptulose-1,7-bisphosphatase but still encoded a canonical fructose-1,6-bisphosphatase for the last step in the CBB cycle that has the highest energy-saving potential (35) (Dataset S1A). This suggests that the most energy-efficient variant of $PP_i$-dependent CBB cycle evolved independently in *Ca*. R. santandreae, possibly due to the same energy constraints that led to the selection of the most efficient sulfur oxidation pathway across the diversity of thiotrophic symbionts (38) (*SI Appendix*, *Supplementary Note 2*).

The *Ca.* R. santandreae symbiont encoded form IA RuBisCO for carbon fixation rather than the form II RuBisCO that was previously identified by PCR amplification in another *Ca.* Riegeria species, *Ca.* R. galateiae (9) (*SI Appendix*, Fig. S10). The presence of two different RuBisCO forms in the *Ca.* Riegeria clade that both clustered with sequences from *Alphaproteobacteria* suggests that the last common ancestor encoded both forms and that descendants differentially retained only one. Several members of the family *Rhodospirillaceae* are known to encode more than one RuBisCO form per genome, including forms IA and II (39). As form IA is adapted to higher oxygen and lower carbon dioxide levels compared with form II, the retention of form IA in *Ca*. R. santandreae could indicate a preference for habitats with higher oxygen concentrations for *Paracatenula* sp. santandrea compared with *Paracatenula galateia* (39). Niche differentiation via different forms of RuBisCO has been observed in free-living thiotrophic bacteria, e.g., in strains of *Ca.* Thiomargarita nelsonii (40), but not within a single clade of chemoautotrophic symbionts (15, 36, 41, 42).

Characteristic of obligate autotrophs such as the vesicomyid clam endosymbionts is an incomplete TCA cycle that lacks the alpha-ketoglutarate dehydrogenase complex (13, 14). In contrast, *Ca.* R. santandreae encoded a complete TCA cycle, which in combination with glycolysis, allows it to use sugars and other organic substrates such as fatty acids as carbon and energy sources (Fig. 2 and *SI Appendix*, Fig. S11). While these pathways might indicate a heterotrophic lifestyle, the results of our analysis of the transporters suggest that *Ca.* R. santandreae uses these pathways for internal cycling of carbon stocks (*SI Appendix*, *Supplementary Note 2*). We identified seven import transporters in the symbiont's genome with transmembrane domains and found only a very low number of importers of organic compounds (*SI Appendix*, Fig. S12 and Dataset S1 G and H). The *Ca.* R. santandreae symbiont can only take up selected peptides, but no sugars or fermentation end products that might be present in the sediments or originate from the host metabolism.

Still, TCA cycle genes were highly transcribed, and the corresponding proteins were abundant, constituting a total of 3 ± 0.49% NSAF (sum of seven proteins; $n = 8$ proteomes) of all measured proteins in the proteomes (Fig. 2 and *SI Appendix*, Figs. S3 and S4). We hypothesize that carbon fixation and the TCA cycle do not operate simultaneously in a single cell but are expressed in separate symbiont populations to avoid futile cycling. To replenish intermediates of the TCA cycle and to synthesize and convert acetyl-CoA into biomass, *Ca.* R. santandreae expresses the ethylmalonyl-CoA pathway and an incomplete 3-hydroxypropionate bicycle (3-HPB) pathway. The ethylmalonyl-CoA pathway (sum of five proteins, 2.25 ± 0.34% NSAF; $n = 8$ proteomes) is typical for *Alphaproteobacteria* (43–45) and has a similar function to the glyoxylate shunt that operates in many *Gammaproteobacteria* but additionally allows the coassimilation of carbon dioxide (43) (Fig. 2). This type of carbon fixation coupled to the replenishment of TCA cycle intermediates is called anaplerotic carbon fixation. For three molecules of acetyl-CoA that enter the ethylmalonyl-CoA pathway, two molecules of $CO_2$ can be fixed at the expense of only one ATP (46). A similarly highly expressed but incomplete 3-HPB pathway (sum of five proteins, 1.94 ± 0.42% NSAF; $n = 8$ proteomes) could allow the assimilation of organic compounds such as acetate, propionate, succinate, and malate, even though several key reactions that would permit it to function autotrophically were not predicted in *Ca.* R. santandreae (35, 47). The elements of the 3-HPB pathway that are encoded represent a second anaplerotic pathway, similar to the ethylmalonyl-CoA pathway, and consume one ATP for the fixation of one molecule of $CO_2$ (35, 47). Compared with the $4\ ^1/_9$ ATP that would be needed in the $PP_i$-dependent CBB cycle for the fixation of two molecules $CO_2$ (35), both the ethylmalonyl-CoA pathway and the 3-HPB pathway would allow for a cheaper maintenance and expansion of carbon stocks in *Ca.* R. santandreae. The total proteomic investment in both anaplerotic pathways (i.e., 3-HPB and ethylmalonyl-CoA) was similar to autotrophic carbon fixation via the $PP_i$-dependent CBB cycle (*SI Appendix*, Fig. S4). This suggests, in combination with the highly expressed TCA cycle, that the two pathways contribute substantially to carbon assimilation in *Ca*. R. santandreae. While this has been shown for other mixotrophs (46) or apparent heterotrophs like SAR11, one of the most successful clades of marine bacteria (48), it has not been observed in chemosynthetic symbioses. How much the possible anaplerotic fixation in *Ca*. R. santandreae contributes to the total carbon fixation in the symbiosis remains to be shown. Both pathways are not essential to replenish intermediates of the TCA cycle they synthetize. It is therefore tempting to speculate that without an essential benefit to the efficiency of the symbiont's carbon fixation or replenishment of TCA cycle intermediates, one or both pathways would have been lost in the process of reductive genome evolution.

The exact same gene set of the partial 3-HPB pathway described above has also been found in thiotrophic gammaproteobacterial symbionts from gutless oligochaetes and ciliates (35, 49). Our phylogenetic analyses showed that most of the *Ca*. R. santandreae homologs of genes characteristic for the 3-HPB pathway clustered with genes from these two gammaproteobacterial symbiont groups, except for one that clustered with sequences from other *Rhodospirillaceae* (*SI Appendix*, Fig. S13). For three out the five enzymes, the cluster of sequences from thiotrophic symbionts were the sister clade to homologs from *Rhodospirillaceae*, whereas two of them (Meh and Mct) had no homologs in *Rhodospirillales*. This extended phylogenetic distribution for several of the enzymes corroborates an origin of these genes for enzymes of the 3-HPB pathway outside of the *Chloroflexi*, as also suggested for the chemosynthetic symbionts of *Kentrophoros* (49), and points to the *Alphaproteobacteria* as one possible group of origin (50).

*Paracatenula* sp. santandrea shares the Sant'Andrea bay sediments with several groups of chemosynthetic meiofauna such as gutless oligochaetes, stilbonematinae nematodes, and the ciliate *Kentrophoros* (6, 51, 52). The most abundant taxon is the gutless oligochaete *Olavius algarvensis*, followed by *Paracatenula* sp. santandrea. They represent the most contrasting forms of these chemosynthetic symbioses. *O. algarvensis* hosts a mixotrophic consortium of five bacterial symbionts (53) with little signs of genome reduction in any of the symbiont groups (51), while *Paracatenula* relies on a single endosymbiont with a drastically reduced genome. These contrasting metabolic capabilities likely

reflect their different ecological niches. The *Olavius* symbiosis can use several energy sources in addition to sulfur oxidation, such as carbon monoxide, hydrogen, and small organic compounds (35, 54), whereas the *Paracatenula–Ca.* Riegeria symbiosis is an energy-efficient specialist with sulfur oxidization as its main energy source.

**Specialization and Convergence in Intracellular Thiotrophic Symbionts.** To test for deviating evolutionary patterns in the alphaproteobacterial *Ca.* R. santandreae symbiont, we compared its genome to those of 62 bacteria representing free-living relatives of *Ca.* Riegeria, parasites, nonthiotrophic mutualistic symbionts, as well as thiotrophic bacteria from other symbiotic and nonsymbiotic clades (for a list of genomes, see Dataset S1I). We analyzed the cellular processes encoded in each their genomes in a nonmetric multidimensional scaling (NMDS) ordination of the distribution of each their genes to clusters of orthologous genes (COGs) and to modules of the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Fig. 3 and *SI Appendix*, *Supplementary Note 3*).

Apart from *Ca.* R. santandreae, the thiotrophic symbiont genomes formed a spectrum that was dependent on their genome sizes, suggesting that COG- and KEGG-based profiles and genome contents in the other symbionts with small genomes were largely convergent. At the other end of this spectrum, the profiles of symbionts with nonreduced genomes overlapped with free-living thiotrophs such as *Allochromatium vinosum* and *Thiocapsa marina* (Dataset S1 J and K).

While *Ca.* R. santandreae had the most similar distribution of encoded functions to vesicomyid symbionts, another group of vertically transmitted chemosynthetic symbionts, its profile stood out in the observed NMDS patterns across the categories of COGs and KEGG modules (Fig. 3 *A* and *B*, *SI Appendix*, *Supplementary*

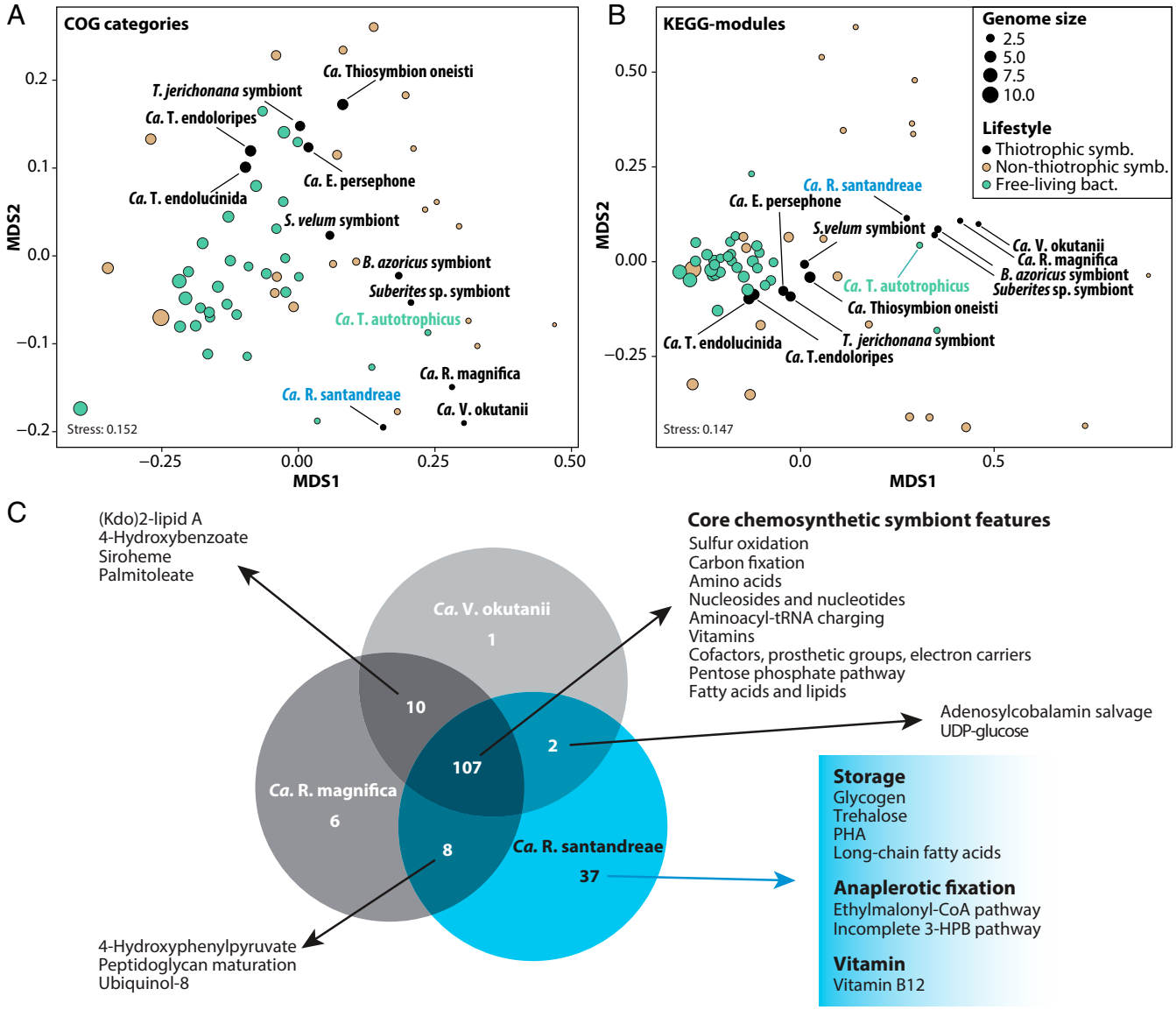


**Fig. 3.** The unique functions in *Ca.* R. santandreae allow efficient maintenance of large carbon and energy stocks. (*A* and *B*) Functional diversity of thiotrophic symbionts and selected reference bacteria. The diameter of each circle represents the genome sizes. Free-living bacteria include *Alpha-* and *Gammaproteobacteria* and one Cyanobacterium. A list of genomes used and the raw plots can be found in Dataset S1I ($n = 63$). (*A*) NMDS plot of COG category distributions with a 2D stress of 0.152. (*B*) NMDS plot of KEGG module distributions with a 2D stress of 0.147. (*C*) Unique pathways in vertically transmitted thiotrophic symbionts with reduced genomes. Metabolic pathways specific for *Ca.* R. santandreae or shared between two or more symbionts are highlighted. See details in Dataset S1L.

*Note 3* and Fig. S14, and Dataset S1 J and K; for a list of genomes, see Dataset S1I). *Ca*. R. santandreae deviated from the vesicomyid symbionts with an enrichment of genes coding for carbohydrate transport and metabolism (COG category G), lipid transport and metabolism (COG category I), and coenzyme transport and metabolism (COG category H) (Dataset S1J). We could corroborate this pattern in a comparison on the level of pathways as *Ca.* R. santandreae encoded 37 unique pathways that allow for flexible carbon fixation and the versatile storage of carbohydrates (Fig. 3*C* and *SI Appendix*, *Supplementary Note 3*). In particular, the capabilities to synthesize long-chain fatty acids and polyhydroxyalkanoates (PHAs) as well as trehalose and glycogen underline a possible dual role of the symbionts in the *Paracatenula* symbiosis, where the symbionts provide large storage capacities for the association (Fig. 3*C* and Dataset S1 L and M). This is in contrast to all other chemosynthetic symbioses and could be coupled to the high proportion of the total biomass that the *Ca.* Riegeria cells achieve in the *Paracatenula* consortium (9).

While *Ca*. R. santandreae has a substantial set of unique pathways, it also shared 107 pathways with the thiotrophic symbionts with the most reduced genomes, *Ca.* Ruthia magnifica, and *Ca.* Vesicomyosocius okutanii. Besides energy conservation through sulfur oxidation, these shared pathways were largely related to biosynthesis and include carbon fixation via the CBB cycle and pathways for amino acid, nucleoside, nucleotide, fatty acid, and lipid synthesis (*SI Appendix*, *Supplementary Note 3* and Dataset S1 L and M). These pathways represent the core functions of a nutritional chemosynthetic symbiont and reflect the strong selection based on the conserved metabolic needs of their invertebrate hosts. In such nutritional symbioses, essential biosynthetic pathways must be retained because their loss would lead to a decrease in fitness or even the extinction of the affected host (13, 14, 17). In contrast, a loss of catabolic pathways might not severely impact the symbionts but might, in return, even benefit the consortium as it leaves resources accessible to the host.

**_Ca_. R. santandreae Is a Major Carbon and Energy Buffer in the _Paracatenula_ Symbiosis.** Our comparative analyses showed that *Ca.* R. santandreae possessed and expressed an unexpectedly high number of pathways for a chemoautotrophic symbiont with a reduced genome, many connected to carbon metabolism and storage. The pathways included the synthesis of PHAs, trehalose, and glycogen (*SI Appendix*, Fig. S11 and Dataset S1L). We observed a large number of intracellular inclusions in the symbionts in both light microscopy and EM that caused the orange-white appearance of the trophosome region and further highlighted the symbionts storage capabilities (Figs. 4 and 5*E*). By combining Raman spectroscopic imaging with sensitive MS protocols, we were able to account for most of these storage inclusions that make up large proportion of the total *Ca.* R. santandreae volume (*SI Appendix*, Figs. S15 and S16 and *Supplementary Note 2*).

Two conspicuous inclusions were detectable in differential interference contrast microscopy. One of them, identified using Raman spectroscopy, was elemental sulfur that appears as light-refractile granules of 1–2 μm in diameter (Fig. 4*B* and *SI Appendix*, Fig. S15). The second type of inclusions was larger, with sizes of up to 3.5 μm in diameter, apparent in most cells and occupied 13.1 ± 9.4% of the cell volumes ($n = 40$ symbiont cells) (Fig. 4*C*). These were likely composed of PHA, as they were selectively stained by the lipophilic dye Nile Red, and both the polymer polyhydroxyvalerate and its monomer hydroxyvalerate could be identified by gas chromatography–MS (GC-MS) (*SI Appendix*, *Supplementary Note 2* and Fig. S16). Furthermore, genes for PHA synthesis and the phasin protein PhaP were highly transcribed, and PhaP was the most abundant protein in six out of eight proteomes (6.67 ± 1.83% NSAF; $n = 8$ proteomes) (*SI Appendix*, Figs. S3 and S4). Besides storing carbon, PHA might function as an electron sink under anoxic conditions (55–57). The host animal *Paracatenula* likely traverses the oxic and anoxic sediment layers on a regular basis to provide its symbionts with sulfide, a behavior shown for invertebrate meiofauna with thiotrophic symbionts (58, 59). The *Ca.* R. santandreae symbiont had the necessary genes for

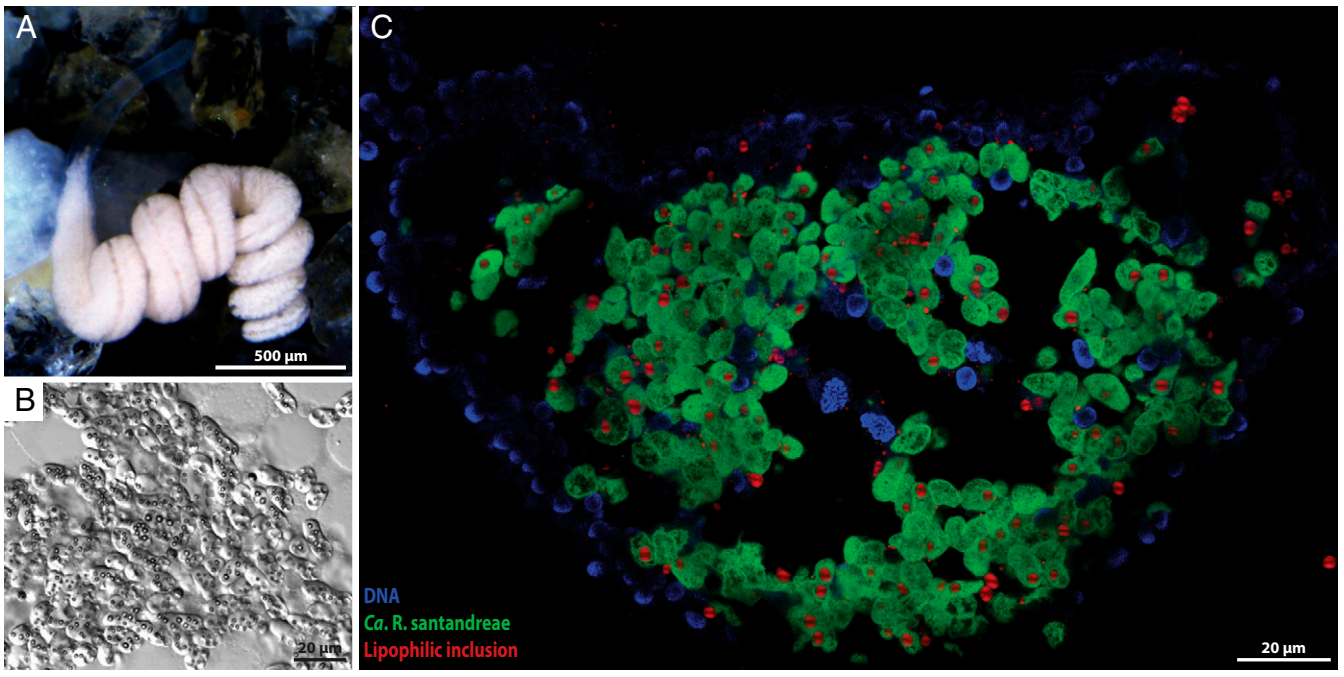


**Fig. 4.** *Ca.* R. santandreae are large endosymbiotic bacteria with massive storage inclusions. (*A*) Habitus of *Paracatenula* sp. santandrea. The white trophosome contains endosymbionts, while the anterior and transparent part of the worm (rostrum) is bacteria-free. (*B*) Differential interference contrast image of *Ca.* R. santandreae symbionts indicated multiple intracellular inclusions. (*C*) Confocal laser-scanning image of CARD-FISH combined with a lipophilic staining (Nile Red) on a transverse section of a *Paracatenula* specimen in the symbiont-bearing region. Overlay of DAPI signal (blue), Nile Red (red), and probe targeting the symbionts (green, EUB I–III) is shown.

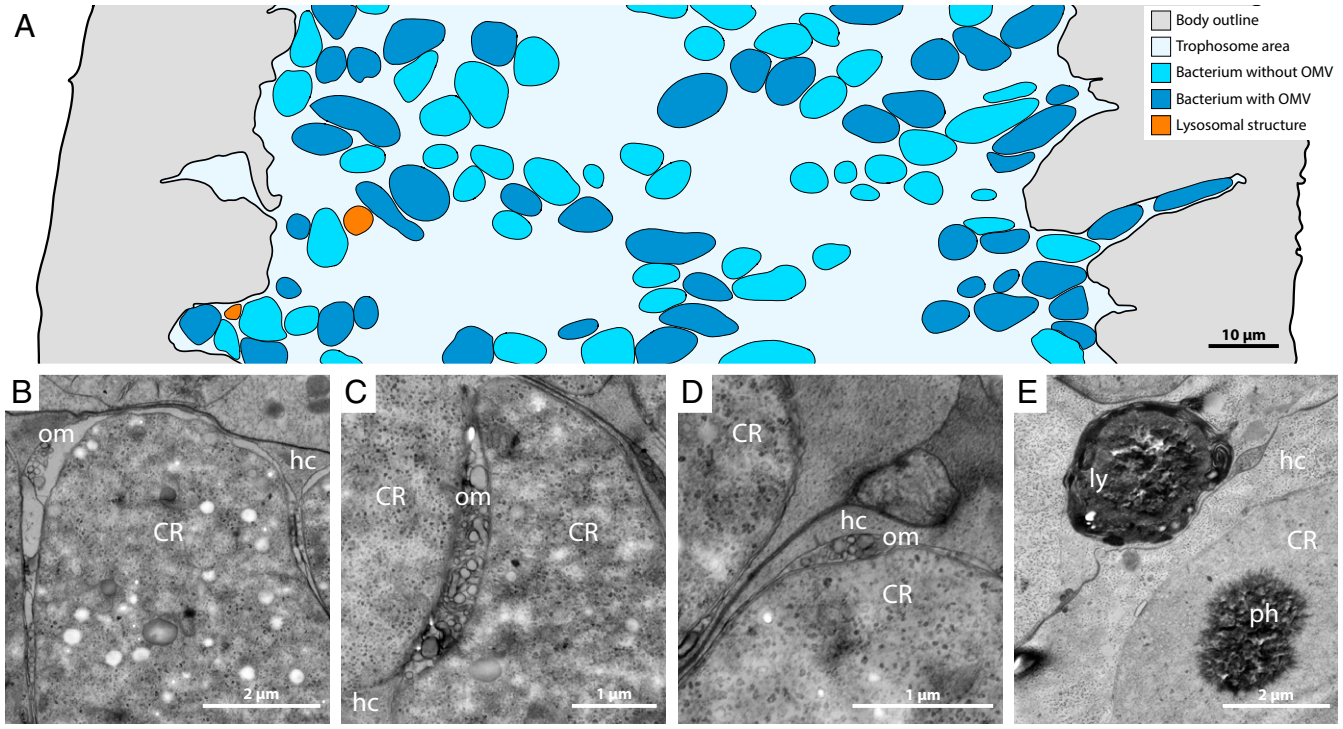


**Fig. 5.** *Ca.* R. santandreae is rarely digested and likely nourishes the host with OMVs. (*A–E*) TEM analysis of *Paracatenula* sp. santandrea. (*A*) Trace illustration of the trophosome region with the status of the intracellular bacterial symbiont *Ca.* R. santandreae indicated in dark blue (OMVs detected), light blue (no OMVs detected), and orange (digested—lysosomal bodies). See the raw overview image in *SI Appendix*, Fig. S22. (*B–D*) Representative image outcrops of the OMV-secreting symbiont population. CR, *Ca.* R. santandreae; hc, host cytosol; om, OMVs; ph, storage inclusion (PHA). (*E*) A symbiont cell that has undergone lysosomal digestion (ly). The enclosed structure in the lysosome resembled a PHA inclusion (ph) (Fig. 4*C* and *SI Appendix*, Fig. S21*A*).

aerobic respiration coupled with ATP synthesis but lacked terminal oxidases for utilization of nitrate or alternative inorganic electron acceptors (Fig. 2 and *SI Appendix*, *Supplementary Note 2*). A possible electron sink for the oxidation of sulfide to elemental sulfur when oxygen is absent could be the synthesis of PHA from acetyl-CoA, which would simultaneously function as both a store of energy and of carbon. The accumulation of PHA under anaerobic conditions has also been shown in other *Alphaproteobacteria* that live in changing redox conditions, such as the free-living *Ca.* Defluviicoccus tetraformis (60). A similar mechanism was shown for *Ca.* Accumulibacter phosphatis (*Betaproteobacteria*) and was also proposed for the unrelated gammaproteobacterial thiotrophic symbionts of *Olavius algarvensis* (55, 56, 61).

Electron-dense particles with the typical size and shape of glycogen inclusions were identified in both host and symbiont tissue (*SI Appendix*, *Supplementary Note 2* and Fig. S17). Glycogen is widely used to store energy and carbon in both bacteria and animals and is considered to be a long-term energy reserve under anaerobic conditions (35, 62–64). Proteins for glycogen metabolism were also detected in the symbiont proteomes, although their genes were not highly expressed (Fig. 2 and *SI Appendix*, Figs. S3 and S4). We detected proteins for both biosynthesis and degradation in the same samples, suggesting a dynamic switching between synthesis and utilization, possibly distributed in subpopulations of the symbionts (Dataset S1 D and E).

Unexpectedly, the disaccharide trehalose was by far the most abundant soluble metabolite measured in *Paracatenula* holobionts (*SI Appendix*, *Supplementary Note 2* and Figs. S16 and S18). We could link the trehalose pool to the symbionts as it was not detectable in host tissue without symbionts (anterior rostrum fragments; $n = 3$; *SI Appendix*, Fig. S18*A*). The *Ca.* R. santandreae symbiont uses the OtsAB pathway for trehalose synthesis from UDP-glucose and glucose-6-phosphate. The trehalose-P-synthase was among the 10% most highly expressed genes and was detectable in the proteome (*SI Appendix*, Figs. S3 and S4). Trehalose can be found in a wide spectrum of eukaryotes and bacteria and has multiple biological functions including carbon storage but also the protection from osmotic stress, e.g., due to varying salinities (65). *Paracatenula* sp. santandrea could experience lower salinities due to freshwater input into Sant'Andrea bay in periods of prolonged precipitation. Changes in salinity during salinity tolerance tests however had no effect on trehalose levels, indicating a storage rather than osmolyte function. Specimens survived 5-h treatments in seawater of a salinity ranging from 25 to 45‰, but no significant differences in trehalose concentrations could be measured (salinities from 20 to 50‰; $n = 5$; two-sided $t$ test; *SI Appendix*, Fig. S18*B*).

Based on expression data, chemical measurements, and visualizations, *Ca.* R. santandreae has a unique ability to accumulate massive amounts of multiple energy and carbon storage compounds. Our chemical characterization and quantification showed that all of these carbon and energy stocks are available with large standing stocks. The carbon stock is dominated by PHA and trehalose (each a few micrograms of C per worm), whereas glycogen is underrepresented (a few nanograms of C per worm) (*SI Appendix*, *Supplementary Note 2*). Elemental sulfur, PHA, and trehalose can be solely attributed to the symbionts. Taken together, our observations imply that the *Ca.* R. santandreae symbiont performs the primary energy storage for the whole animal–microbe symbiosis.

**OMVs Secreted by *Ca.* R. santandreae Likely Form the Basis of Host Nutrition.** Like many marine invertebrates associated with thiotrophic symbionts, *Paracatenula* lacks a mouth and a gut and depends on its endosymbionts for essential nutrients. This was reflected in the broad genomic repertoire and the expression of amino acid, vitamin, and cofactor biosynthesis pathways in the symbiont (*SI Appendix*, *Supplementary Note 2* and Figs. S19 and S20).

Active export across the symbiont membrane appears to be limited as we only identified 15 export related gene products in *Ca.* R. santandreae, 8 of which had at least 1 transmembrane (TM) domain based on the identification using the Transporter Classification Database (TCDB) (*SI Appendix*, *Supplementary Note 2* and Fig. S12 and Dataset S1 G–I). The main identified transporters were the Tat and Sec secretion systems that could translocate folded and unfolded proteins to the periplasm. A Type I secretion system could export unfolded proteins across the outer membrane and in the hosting vacuoles but no secretion system for the transfer of proteins into the host bacteriocyte could be detected. Although several transporters such as secretion systems have essential functions conserved across bacteria (66) and might be involved in the nutrition of the host, unexpectedly only the Tat secretion system for the translocation of folded proteins was highly expressed (Fig. 2 and Dataset S1B).

Given that we could not find expression based evidence of substantial export via transporters, we were surprised that we did not observe *Ca*. Riegeria cells undergoing lysosomal digestion in our transmission EM (TEM) analyses ($n = 5$ host specimens, 682 symbiont cells, 54,895 $\mu m^2$ total analyzed area) (Fig. 5 and *SI Appendix*, Fig. S21 and Table S2). We also only identified nine lamellar bodies containing inclusions of symbiont cells in the same dataset. The absence of active bacterial degradation and the presence of few lamellar bodies suggests a low level of symbiont digestion. In contrast, lysosomal digestion of symbiont cells is readily observed in bacteriocytes of other chemosynthetic hosts, e.g., members of the deep-sea mussel genus *Bathymodiolus* as well as in vesicomyid clams (67, 68). Host gene expression indicated lysosome formation and trafficking as well as proteolysis, all likely involved in the digestion of symbiont biomass (Dataset S1N). One of the most important enzymes in the lysosomal digestion of bacterial cells are lysozymes. They are glycoside hydrolases that cleave the 1,4-beta-linkages in peptidoglycan of bacterial cell walls and lead to cell lysis. Unexpectedly, we could not detect lysozyme transcripts in the host transcriptome [79.1% completeness identified by Benchmarking Universal Single-Copy Orthologs (69)], suggesting that lysozyme expression was below detection limit or absent. Several other types of digestive enzymes were expressed, such as the cysteine protease cathepsin B that has been shown to play a major role in the digestion of gut-bearing flatworms and is also expressed in the symbiotic gutless annelid *Olavius algarvensis* (70, 71) (Dataset S1N). Because the symbiont very likely has a peptidoglycan cell wall (Fig. 2 and *SI Appendix*, *Supplementary Note 4*), we conclude that complete symbiont cells are not the main target for the lysosomal digestion expressed in *Paracatenula*. This corroborates the ultrastructural observations and is in contrast to other animals with chemosynthetic bacterial symbionts. Hosts such as the *Bathymodiolus* deep-sea mussels have been shown to constitutively express lysozymes that play a key role in symbiont digestion and the control of symbiont populations (67).

Our data rather suggest that the main mode of symbiont-to-host transfer is based on the release of OMVs and their digestion by the host. OMVs naturally package lipids, proteins, and sugars and can also carry nucleic acids, but despite increased attention into their role in symbiotic interactions, they have not been shown to be broadly involved in host nutrition in animal-microbe interactions (72). We imaged random tissue sections from five individuals using TEM and quantified the number of bacteria, bacteria with OMVs, and lysosomal structures in the bacteriocyte region. On average, of the 136 ± 35 *Ca.* R. santandreae cells per section, 62.85 ± 10.94% showed OMV formation (total endosymbiont cells: 682; $n = 5$ specimens; Fig. 5*A* and *SI Appendix*, Table S2). We typically observed several clusters of small to large OMVs in close proximity to *Ca.* R. santandreae cells (Fig. 5 *A–D*). Correspondingly, we identified a highly expressed intermembrane lipid transport system (ILTS) [also known as mammalian cell entry proteins (*mlaED*)] of the endosymbiont that has been shown to play an important role in the formation of OMVs (73). The two most highly expressed genes of the ILTS (*mlaE* and *mlaD*; TPM: 2,536 ± 682 and 2,097 ± 531, respectively; $n = 3$) were among the top 10% of expressed genes and represent the subunits that form the transport pore in the inner membrane (74) (Fig. 2). In contrast, the outer-membrane protein of the ILTS, VacJ, had 60× lower expression (VacJ; TPM: 39 ± 13; $n = 3$). VacJ maintains outer-membrane asymmetry by removing phospholipids from the outer leaflet and supports their trafficking from the outer membrane (75). The observed low expression of VacJ has been linked to increased OMV formation in VacJ-deficient *Gammaproteobacteria* (73). Although we could not resolve the exact mode of OMV uptake by the host, their channeling into host lysosomal digestion is likely (76) and would also complement the observed expressions of digestive enzymes by the host.

Based on (*i*) the limited expression of transporter-based export by the endosymbionts, (*ii*) the observation of OMV formation in the majority of symbiont cells and corresponding symbiont expression profiles, (*iii*) the detection of host expression of lysosomal digestion but no expression of lysozymes, and (*iv*) the rare occurrence of digestion of full symbiont cells, we propose that digestion of symbiont material in the form of OMVs as the main mode of nutrition in the *Paracatenula* symbiosis. In this model, the *Ca.* R. santandreae symbiont builds up and efficiently maintains the primary energy storage for the whole animal–microbe consortium. The host then directly accesses all these stockpiles via the digestion of symbiont-derived OMVs and gains lipids, amino acids, and sugars, as well as many other possible nutritive constituents. Only a very small subpopulation of its symbionts is digested at any given time, in contrast to all other chemoautotrophic symbioses where symbiont cell digestion drives biomass transfer.

## Conclusion

After 500 million years of strict vertical transmission and reductive genome evolution, *Ca*. R. santandreae has a drastically reduced genome compared with its free-living relatives. Although the symbiont has been associated with its host much longer than, e.g., the oldest known insect symbionts (9, 77), it had to retain a much larger genome and metabolic toolkit, reflecting the multitude of biosynthetic functions that are essential for maintaining the symbiosis. It has maintained a versatile carbon metabolism and carbohydrate storage, in stark contrast to the similarly reduced vesicomyid clam symbionts that appear to be optimized only for autotrophic biomass production. The versatile function of *Ca.* R. santandreae might be connected to its ancestry in *Rhodospirillaceae*, with many of which they share a complete intermediary carbon metabolism as well as the ethylmalonyl-CoA pathway that is common in *Alphaproteobacteria* (43). In addition, *Ca*. R. santandreae has several unexpected features, even for a versatile alphaproteobacterial background that are shared with gammaproteobacterial symbionts—an incomplete 3-HPB pathway and a $PP_i$-dependent CBB cycle. In *Ca*. R. santandreae, the majority of the genes involved in these pathways are of alphaproteobacterial origin and not acquired via lateral gene transfer. Instead of showing an acquisition as proposed for the gammaproteobacterial symbionts, our analyses extended the phylogenetic distribution of these energy-efficient and versatile pathways and shed light on their origins and possible histories of transfer from the metabolically flexible *Alphaproteobacteria*.

The key elements in the physiology of *Ca*. R. santandreae such as enhanced modes of carbon fixation, intermediary carbon metabolism, and sulfur oxidation are optimized for high energy efficiency. The endosymbionts can create and maintain versatile carbon pools as they represent the bulk storage organ for the consortium and likely support its host via OMV secretion. This integration of storage and transfer into a single and streamlined chemoautotrophic organism with a reduced genome represents a unique type of bacteria–animal interaction. *Ca*. R. santandreae extends the spectrum of nutritional symbioses that range from

single vitamin or amino acid supplementation to full nutrition (17–23) and adds the function of primary energy and carbon storage that in all other animal hosts is performed by the animal itself.

## Materials and Methods

**Sample Collection and Light Microscopy.** *Paracatenula* sp. santandrea specimens were collected between 2013 and 2017 from the bay off Sant'Andrea, Elba, Italy (9). *Paracatenula* specimens were fixed in RNAlater (Ambion) and stored at 4 °C or fixed in methanol and stored at −20 °C. Specimens for TEM were kept for up to 7 months in 12 mL miniaquaria. Live specimens were imaged using a Nikon SMZ25 stereomicroscope (Nikon) and differential interference contrast on a Zeiss LSM 780 confocal laser microscope (Carl Zeiss AG).

**EM.** All specimens were fixed with PHEM (Pipes, Hepes, EGTA, $MgCl_2$)-buffered glutaraldehyde, processed with a high-pressure freezer, and freeze-substituted in acetone with 1% wt/vol osmium tetroxide. After low viscosity resin embedding and ultrathin sectioning (70 nm), mounted and contrasted sections were randomly chosen for quantitative image analysis of host and symbiont structures.

**CARD-FISH and Nile Red Staining.** CARD-FISH, DAPI, and Nile Red staining were performed on paraffin sections of *Paracatenula* sp. santandrea, and relative volumes of Nile Red-stained inclusions were estimated using quantitative image analysis.

**DNA Sequencing, Assembly, and Binning.** DNA was extracted using DNeasy Blood and Tissue Micro Kit (Qiagen). Illumina MiSeq reads were sequenced from paired-end libraries (Ovation Ultralow Library System V2; NuGEN), kmer-filtered for the high coverage symbionts, and assembled. The genome bin derived from the metagenomic assembly graph was improved using iterative read mapping and reassembly.

**Metabolic Reconstructions and Comparison.** The *Ca*. R. santandreae genome assembly was annotated, and the automated annotations were manually curated (transporters, PHA synthesis, and incomplete 3-HPB). Protein sequences were assigned to Clusters of Orthologous Genes (COGs) and KEGG Orthologs (KOs), and metabolic features were predicted from KOs and our curated annotations. The distribution of genes across (*i*) COG categories was quantified, and (*ii*) the completeness per KO module and genome was encoded into a matrix and analyzed using NMDS plots. Pathways were compared between *Ca*. R. santandreae, *Ca*. R. magnifica, and *Ca*. V. okutanii based on Pathway Tools and represented in a Venn diagram.

**RNA Extraction and Transcriptome Analyses.** Host and symbiont transcription was analyzed from total RNA extractions of single *Paracatenula* sp. santandrea specimens. Libraries (Ovation RNASeq System v2; NuGEN) were sequenced on the HiSeq 2500 platform, and transcript counts were generated using kallisto. The TPM were sorted based on the average for the three replicates.

For host transcriptomics, a de novo transcriptome assembly was generated from filtered mRNA reads using Trinity. After cleanup and deduplication, the final host transcriptome that consisted of 54,926 predicted genes was annotated with Trinotate.

**Protein Extraction and Proteomic Analyses.** For 1D liquid chromatography–tandem MS (LC-MS/MS) proteomics, tryptic digests from eight samples (four from one animal and four from three animals pooled) were separated using a 260-min gradient and roughly 110,000 MS/MS spectra were acquired per sample using the Orbitrap. The final protein identification database consisted of all protein sequences predicted from the *Ca*. R. santandreae genome and the database for common laboratory contaminants. A total of 407 symbiont proteins were identified using the Sequest HT node in Proteome Discoverer with an overall false discovery rate of <5%.

**Phylogenetic Analyses.** 16S rRNA and *dsrB* sequence identities between *Ca*. R. santandreae and *Ca*. R. galateiae were determined using single gene alignments. The average AAIs of *Ca*. R. santandreae and free-living relatives were calculated with the AAI calculator (enve-omics.ce.gatech.edu/aai).

The phylogenomic tree of *Ca*. R. santandreae and representatives from five alphaproteobacterial orders (*Rhodospirillales*, *Sphingomonadales*, *Rhodobacterales*, *Rhizobiales*, and *Rickettsiales*) was calculated based on 43 conserved marker and estimated from the amino acid alignment using FastTree.

Single-gene phylogenetic analyses based on amino acid alignments were conducted for RuBisCO, $PP_i$-PFK, $PP_i$-$H^+$-pump, and five genes of the incomplete 3-HPB pathway (Mch, Mcl, Smt, Mct, Meh).

**Physiological Experiments.** All physiological experiments were performed with biological replicates; the numbers are indicated in each experiment. Individual specimens served as technical replicates. Size normalization was based on images of each specimen.

Pulse and pulse/chase incubations of *Paracatenula* specimens with $^{14}C$-labeled bicarbonate were performed in either glass beads and artificial seawater or washed Sant'Andrea sediment and natural Mediterranean seawater. Samples were fixed in paraformaldehyde or glutaraldehyde, and the anterior half was used for microautoradiography and the posterior half for bulk measurements. Bulk $^{14}C$ incorporation was measured by liquid scintillation counting. Microautoradiography was performed on LR White (London Resin Company) sections.

For salinity tolerance experiments, specimens were incubated with filtered natural seawater of salinities ranging from 20 to 50‰ for 5 h and fixed in methanol. No specimens could be recovered from 20 and 50‰ setups.

**Metabolite Measurements.** Untargeted and PHA-targeted GC-MS–based metabolite measurements were performed on extractions from single *Paracatenula* specimens (trehalose quantification), five *Paracatenula* specimens (PHA quantification), and 25 *Paracatenula* specimens (bulk metabolome analysis). Additionally, three *Paracatenula* specimens were sectioned to separately analyze equal-sized fragments of rostrum (symbiont-free, anterior) and trophosome (symbiont-hosting, posterior).

A kit-based fluorometric glycogen assay (MAK016; Sigma-Aldrich) was used to quantify glycogen contents in tissue extracts from single worm specimens. Raman spectroscopy was performed on live *Paracatenula* sp. santandrea specimens to identify highly refractile inclusion in symbiont cells.

**Data Availability.** The assembled and annotated *Ca*. R. santandreae genome has been deposited in the European Nucleotide Archive (ENA) and is accessible under project no. PRJEB26644 and the assembly under accession no. LR026963. Raw total RNA library reads and the host assembly have been deposited in ENA under project no. PRJEB31702 via The German Federation for the Curation of Biological Data (GFBio) (78).

The MS proteomics data and the protein sequence databases have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository (79) for the pure culture data with the dataset identifier PXD009856.

The raw data for the EM-based quantitative image analysis have been deposited in Figshare (https://doi.org/10.6084/m9.figshare.7746806) (80).

**Code Availability.** The script that was used to classify the transporter families is available at https://github.com/kbseah/tcdbparse_sqlite.

**ACKNOWLEDGMENTS.** We acknowledge the Hydra Institute team on Elba, especially M. Weber, C. Lott, H. Kuhfuss, and M. Schneider for support during fieldwork. We thank N. Dubilier, K. Altendorf, and J. Ott for project discussions and editorial advice on an early version of this manuscript. We thank J. Wippler, A. Assié, C. P. Antony, and T. Wagner for project discussions. We acknowledge the Max Planck Genome Centre Cologne for library preparation and sequencing. We thank N. Tebeka and S. Dyksma for support with microautoradiography and M. Sadowski and R. Yemanaberhan for support during incubation experiments, M. Jung and G. Sondej for support during worm cultivations, J. Kesting and A. Chan for fieldwork help, A. Gruhl and B. Geier for support with microscopy, E. Puskás for support in GC-MS measurements and J. Zimmermann for technical suggestions. This study was funded by the Max Planck Society through N. Dubilier. M.K. was partially funded by the North Carolina State University Chancellor's Faculty Excellence Program Cluster on Microbiomes and Complex Microbial Communities. H.R.G.-V. was partially funded by Marie-Curie Intra-European Fellowship PIEF-GA-2011-301027 CARISYM.

1. Felbeck H (1981) Chemoautotrophic potential of the hydrothermal vent tube worm, *Riftia pachyptila* Jones (Vestimentifera). *Science* 213:336–338.
2. Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube wom *Riftia pachyptila* Jones: Possible chemoautotrophic symbionts. *Science* 213:340–342.
3. Cavanaugh CM (1983) Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. *Nature* 302:58–61.
4. Jones ML (1981) *Riftia pachyptila* Jones: Observations on the vestimentiferan worm from the Galapagos Rift. *Science* 213:333–336.
5. Dubilier N, Bergin C, Lott C (2008) Symbiotic diversity in marine animals: The art of harnessing chemosynthesis. *Nat Rev Microbiol* 6:725–740.
6. Seah BKB, et al. (2017) Specificity in diversity: Single origin of a widespread ciliate-bacteria symbiosis. *Proc Biol Sci* 284:20170764.
7. Dubilier N, et al. (2001) Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. *Nature* 411:298–302.
8. Haddad A, Camacho F, Durand P, Cary SC (1995) Phylogenetic characterization of the epibiotic bacteria associated with the hydrothermal vent polychaete *Alvinella pompejana*. *Appl Environ Microbiol* 61:1679–1687.

9. Gruber-Vodicka HR, et al. (2011) *Paracatenula*, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc Natl Acad Sci USA* 108:12078–12083.
10. Duperron S, et al. (2005) Dual symbiosis in a *Bathymodiolus* sp. mussel from a methane seep on the Gabon continental margin (Southeast Atlantic): 16S rRNA phylogeny and distribution of the symbionts in gills. *Nature* 71:1694–1700.
11. Assié A, et al. (2016) A specific and widespread association between deep-sea *Bathymodiolus* mussels and a novel family of *Epsilonproteobacteria*. *Environ Microbiol Rep* 8:805–813.
12. Russell SL, Corbett-Detig RB, Cavanaugh CM (2017) Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *ISME J* 11:1359–1371.
13. Newton ILG, et al. (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* 315:998–1000.
14. Kuwahara H, et al. (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* 17:881–886.
15. Petersen JM, et al. (2016) Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat Microbiol* 2:16195.
16. Nussbaumer AD, Fisher CR, Bright M (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 441:345–348.
17. McCutcheon JP, Moran NA (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26.
18. Kuwahara H, et al. (2008) Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena* clams. *Extremophiles* 12:365–374.
19. Bennett GM, Moran NA (2013) Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol* 5:1675–1688.
20. Moran NA, Bennett GM (2014) The tiniest tiny genomes. *Annu Rev Microbiol* 68: 195–215.
21. Anbutsu H, et al. (2017) Small genome symbiont underlies cuticle hardness in beetles. *Proc Natl Acad Sci USA* 114:E8382–E8391.
22. Hansen AK, Moran NA (2011) Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA* 108:2849–2854.
23. Moya A, Peretó J, Gil R, Latorre A (2008) Learning how to live together: Genomic insights into prokaryote–Animal symbioses. *Nat Rev Genet* 9:218–229.
24. Bright M, Sorgo A (2003) Ultrastructural reinvestigation of the trophosome in adults of *Riftia pachyptila* (Annelida, Siboglinidae). *Invertebr Biol* 122:347–368.
25. Le Pennec M, Beninger PG, Herry A (1995) Feeding and digestive adaptations of bivalve molluscs to sulphide-rich habitats. *Comp Biochem Physiol* 111A:183–189.
26. Ott J, Rieger G, Rieger R, Enderes F (1982) New mouthless interstitial worms from the sulfide system: Symbiosis with prokaryotes. *Mar Ecol (Berl)* 3:313–333.
27. Dirks U, et al. (2012) Bacterial symbiosis maintenance in the asexually reproducing and regenerating flatworm *Paracatenula galateia*. *PLoS One* 7:e34709.
28. Imhoff JF, Hiraishi A, Süling J (2005) Anoxygenic phototrophic purple bacteria. *Bergeys Manual of Systematic Bacteriology*, eds Brenner DJ, Krieg NR, Staley, JT, Garrity GM (Springer, Boston), 2nd Ed, pp 119–132.
29. Lindstrom ES, Tove SR, Wilson PW (1950) Nitrogen fixation by the green and purple sulfur bacteria. *Science* 112:197–198.
30. Blakemore R (1975) Magnetotactic bacteria. *Science* 190:377–379.
31. Baldani I, et al. (2014) The family *Rhodospirillaceae*. *The Prokaryotes*, eds Rosenberg E, DeLong EF, Stackebrandt E, Lory S, Thompson F (Springer, Berlin), pp 533–618.
32. Childress JJ, Girguis PR (2011) The metabolic demands of endosymbiotic chemoautotrophic metabolism on host physiological capacities. *J Exp Biol* 214:312–325.
33. Dirks U, Gruber-Vodicka HR, Leisch N, Sterrer W, Ott JA (2011) A new species of symbiotic flatworms, *Paracatenula galateia* sp. nov. (*Platyhelminthes*: *Catenulida*: *Retronectidae*) from Belize (Central America). *Mar Biol Res* 7:769–777.
34. Dirks U, Gruber-Vodicka HR, Egger B, Ott JA (2012) Proliferation pattern during rostrum regeneration of the symbiotic flatworm *Paracatenula galateia*: A pulse-chase-pulse analysis. *Cell Tissue Res* 349:517–525.
35. Kleiner M, et al. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA* 109:E1173–E1182.
36. Dmytrenko O, et al. (2014) The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: A blueprint for thriving in and out of symbiosis. *BMC Genomics* 15:924.
37. Markert S, et al. (2011) Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* 11:3106–3117.
38. Klatt JM, Polerecky L (2015) Assessment of the stoichiometry and efficiency of $CO_2$ fixation coupled to reduced sulfur oxidation. *Front Microbiol* 6:484.
39. Badger MR, Bek EJ (2008) Multiple Rubisco forms in proteobacteria: Their functional significance in relation to $CO_2$ acquisition by the CBB cycle. *J Exp Bot* 59:1525–1541.
40. Winkel M, et al. (2016) Single-cell sequencing of *Thiomargarita* reveals genomic flexibility for adaptation to dynamic redox conditions. *Front Microbiol* 7:964.
41. Kleiner M, Petersen JM, Dubilier N (2012) Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Curr Opin Microbiol* 15:621–631.
42. Nakagawa S, et al. (2014) Allying with armored snails: The complete genome of gammaproteobacterial endosymbiont. *ISME J* 8:40–51.
43. Alber BE (2011) Biotechnological potential of the ethylmalonyl-CoA pathway. *Appl Microbiol Biotechnol* 89:17–25.
44. Beck DAC, et al. (2015) Multiphyletic origins of methylotrophy in *Alphaproteobacteria*, exemplified by comparative genomics of Lake Washington isolates. *Environ Microbiol* 17:547–554.
45. Erb TJ, et al. (2007) Synthesis of $C_5$-dicarboxylic acids from $C_2$-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway. *Proc Natl Acad Sci USA* 104:10631–10636.
46. Bill N, et al. (2017) Fixation of $CO_2$ using the ethylmalonyl-CoA pathway in the photoheterotrophic marine bacterium *Dinoroseobacter shibae*. *Environ Microbiol* 19: 2645–2660.
47. Berg IA (2011) Ecological aspects of the distribution of different autotrophic $CO_2$ fixation pathways. *Appl Environ Microbiol* 77:1925–1936.
48. Moran MA, Miller WL (2007) Resourceful heterotrophs make the most of light in the coastal ocean. *Nat Rev Microbiol* 5:792–800.
49. Seah BKB, et al. (2019) Sulfur-oxidizing symbionts without canonical genes for autotrophic $CO_2$ fixation. bioRxiv:10.1101/540435.
50. Shih PM, Ward LM, Fischer WW (2017) Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. *Proc Natl Acad Sci USA* 114:10749–10754.
51. Woyke T, et al. (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955.
52. Zimmermann J, et al. (2016) Closely coupled evolutionary history of ecto- and endosymbionts from two distantly related animal phyla. *Mol Ecol* 25:3203–3223.
53. Kleiner M, et al. (2018) Metaproteomics method to determine carbon sources and assimilation pathways of species in microbial communities. *Proc Natl Acad Sci USA* 115:E5576–E5584.
54. Kleiner M, et al. (2015) Use of carbon monoxide and hydrogen by a bacteria-animal symbiosis from seagrass sediments. *Environ Microbiol* 17:5023–5035.
55. Kleiner M (2012) Metabolism and evolutionary ecology of chemosynthetic symbionts from marine invertebrates. PhD thesis (Universität Bremen, Bremen, Germany).
56. Burow LC, Mabbett AN, Blackall LL (2008) Anaerobic glyoxylate cycle activity during simultaneous utilization of glycogen and acetate in uncultured *Accumulibacter* enriched in enhanced biological phosphorus removal communities. *ISME J* 2:1040–1051.
57. Wilkinson JF (1959) The problem of energy-storage compounds in bacteria. *Exp Cell Res* 7:111–130.
58. Ott JA, et al. (1991) Tackling the sulfide gradient: A novel strategy involving marine nematodes and chemoautotrophic ectosymbionts. *Mar Ecol (Berl)* 12:261–279.
59. Giere O, Conway NM, Gastrock G, Schmidt C (1991) "Regulation" of gutless annelid ecology by endosymbiotic bacteria. *Mar Ecol Prog Ser* 68:287–299.
60. Nobu MK, Tamaki H, Kubota K, Liu W-T (2014) Metagenomic characterization of "*Candidatus* Defluviicoccus tetraformis strain TFO71", a tetrad-forming organism, predominant in an anaerobic-aerobic membrane bioreactor with deteriorated biological phosphorus removal. *Environ Microbiol* 16:2739–2751.
61. Martín HG, et al. (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat Biotechnol* 24:1263–1269.
62. Schöttler U, Bennet EM (1991) Annelids. *Metazoan Life Without Oxygen*, ed Bryant C (Chapman & Hall, London), 1st Ed, pp 165–185.
63. Sorgo A, Gaill F, Lechaire J-P, Arndt C, Bright M (2002) Glycogen storage in the *Riftia pachyptila* trophosome: Contribution of host and symbionts. *Mar Ecol Prog Ser* 231:115–120.
64. Ponnudurai R, et al. (2017) Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J* 11:463–477.
65. Zhang R, et al. (2011) Mechanistic analysis of trehalose synthase from *Mycobacterium smegmatis*. *J Biol Chem* 286:35601–35609.
66. Green ER, Mecsas J (2016) Bacterial secretion systems: An overview. *Microbiol Spectr* 4: VMBF-0012-2015.
67. Fiala-Médioni A, Michalski J-C, Jollès J, Alonso C, Montreuil J (1994) Lysosomic and lysozyme activities in the gill of bivalves from deem hydrothermal vents. *C R Acad Sci Paris, Sci la vie*. *Life Sci* 317:239–244.
68. Streams ME, Fisher CR, Fiala-Médioni A (1997) Methanotrophic symbiont location and fate of carbon incorporated from methane in a hydrocarbon seep mussel. *Mar Biol* 129:465–476.
69. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
70. Wippler J, et al. (2016) Transcriptomic and proteomic insights into innate immunity and adaptations to a symbiotic lifestyle in the gutless marine worm *Olavius algarvensis*. *BMC Genomics* 17:942.
71. Goupil LS, et al. (2016) Cysteine and aspartyl proteases contribute to protein digestion in the gut of freshwater planaria. *PLoS Negl Trop Dis* 10:e0004893.
72. Lynch JB, Alegado RA (2017) Spheres of hope, packets of doom: The good and bad of outer membrane vesicles in interspecies and ecological dynamics. *J Bacteriol* 199: e00012–e00017.
73. Roier S, et al. (2016) A novel mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. *Nat Commun* 7:10515.
74. Ekiert DC, et al. (2017) Architectures of lipid transport systems for the bacterial outer membrane. *Cell* 169:273–285.
75. Abellón-Ruiz J, et al. (2017) Structural basis for maintenance of bacterial outer membrane lipid asymmetry. *Nat Microbiol* 2:1616–1623.
76. O'Donoghue EJ, Krachler AM (2016) Mechanisms of outer membrane vesicle entry into host cells. *Cell Microbiol* 18:1508–1517.
77. Moran NA, Tran P, Gerardo NM (2005) Symbiosis and insect diversification: An ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. *Appl Environ Microbiol* 71:8802–8810.
78. Diepenbroek M (2014) Towards an Integrated Biodiversity and Ecological Research Data Management and Archiving Platform: The German Federation for the Curation of Biological Data (GFBio). *Informatik 2014 – Big Data Komplexität meistern. GI-Edition: Lecture Notes in Informatics (LNI) – Proceedings*, eds Plödereder E, Grunske L, Schneider E, Ull D (Köllen Verlag, Bonn), Vol 232, pp 1711–1724.
79. Vizcaíno JA, et al. (2016) 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* 44:D447–D456.
80. Leisch N (2019) Data from "Multi image tiling of the trophosome Area of *Paracatenula* sp. santandrea." Figshare. Available at https://figshare.com/s/df461220b5ca3282d817. Deposited February 21, 2019.