

# Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation

Jillian M. Petersen<sup>1,2\*</sup>, Anna Kemper<sup>2</sup>, Harald Gruber-Vodicka<sup>2</sup>, Ulisse Cardini<sup>1</sup>, Matthijs van der Geest<sup>3,4</sup>, Manuel Kleiner<sup>5</sup>, Silvia Bulgheresi<sup>6</sup>, Marc Mußmann<sup>1</sup>, Craig Herbold<sup>1</sup>, Brandon K.B. Seah<sup>2</sup>, Chakkiath Paul Antony<sup>2</sup>, Dan Liu<sup>5</sup>, Alexandra Belitz<sup>1</sup> and Miriam Weber<sup>7</sup>

**Chemosynthetic symbioses are partnerships between invertebrate animals and chemosynthetic bacteria. The latter are the primary producers, providing most of the organic carbon needed for the animal host's nutrition. We sequenced genomes of the chemosynthetic symbionts from the lucinid bivalve *Loripes lucinalis* and the stilbonematid nematode *Laxus oneistus*. The symbionts of both host species encoded nitrogen fixation genes. This is remarkable as no marine chemosynthetic symbiont was previously known to be capable of nitrogen fixation. We detected nitrogenase expression by the symbionts of lucinid clams at the transcriptomic and proteomic level. Mean stable nitrogen isotope values of *Loripes lucinalis* were within the range expected for fixed atmospheric nitrogen, further suggesting active nitrogen fixation by the symbionts. The ability to fix nitrogen may be widespread among chemosynthetic symbioses in oligotrophic habitats, where nitrogen availability often limits primary productivity.**

Symbioses between animals and chemosynthetic bacteria are widespread in Earth's oceans<sup>1</sup>. Animals from at least seven phyla have formed such symbioses, and even more chemosynthetic bacterial lineages have evolved symbioses with animal hosts<sup>1</sup>. Chemosynthetic symbionts can use a range of chemicals, such as sulfide, methane, hydrogen and carbon monoxide, to power their metabolism<sup>2–4</sup>. The hosts of chemosynthetic symbionts dominate some animal communities. For example, shallow-water lucinid bivalves, which host sulfur-oxidizing symbionts, often dominate the macrobenthic infaunal community in seagrass meadows, where they can reach densities greater than 3,500 individuals per square metre<sup>5,6</sup>. Their diversity in nature, their persistence over evolutionary timescales and their dominance in many habitats attest to the success of these symbiotic partnerships<sup>1</sup>.

Chemosynthetic symbionts are primarily considered 'nutritional symbionts', meaning their primary role is to provide nutrition for their hosts<sup>1,7</sup>. So far, most studies have focused on inorganic carbon fixation by the symbionts and the transfer of fixed organic carbon compounds to the hosts. In addition to organic carbon, all animals require a source of fixed nitrogen. However, nitrogen metabolism in chemosynthetic symbioses has received far less attention. Chemosynthetic symbionts have been shown to gain their nitrogen from ammonium or nitrate in their environment<sup>8–10</sup> and co-occurring nitrogen-fixing and chemosynthetic symbionts have been found in cold-water corals<sup>11</sup>. Nitrogen fixation by chemosynthetic symbionts has long been hypothesized, but so far not yet shown<sup>12–14</sup>.

Our study focused mainly on the endosymbiosis between bivalves of the family Lucinidae and sulfur-oxidizing bacteria. Lucinids are by far the most diverse and widespread group of bivalves that host chemosynthetic symbionts<sup>15</sup>. There are at least

400 living species, occupying a range of habitats including mangrove sediments, seagrass beds, coral reef sediments and coastal mud and sand<sup>16</sup>. In seagrass habitats, lucinid bivalves and their sulfur-oxidizing symbionts are part of a nested symbiosis with seagrasses, which may be essential to the health and ecological success of seagrasses<sup>6</sup>. We focused on the symbiosis between *Loripes lucinalis* (Lamarck, 1818) and its endosymbionts. We also investigated a second symbiosis, that between stilbonematid nematodes and their sulfur-oxidizing ectosymbionts, because these symbionts are associated with the family Chromatiaceae, which contains a number of diazotrophic sulfur oxidizers<sup>17,18</sup>. Nematodes of the subfamily Stilbonematinae (family Desmodoridae) can be found worldwide in marine sulfidic habitats<sup>19</sup>. All known species have a dense coating of ectosymbionts on their cuticle, which are hypothesized to contribute to their host's nutrition<sup>19</sup>. The name *Candidatus* Thiosymbion oneisti will be proposed elsewhere for the nematode symbionts (Gruber-Vodicka *et al.*, in preparation). We propose the name *Candidatus* Thiodiazotropha endoloripes for the symbiont of *Loripes lucinalis*, where 'Thiodiazotropha' refers to the sulfur-oxidizing ('thio') and nitrogen-fixing ('diazotroph') metabolism of the symbiont and 'endoloripes' ('Endo-', Greek from *ἔνδον* meaning 'within', 'loripes') refers to the endosymbiotic association with *Loripes lucinalis*, its bivalve host.

## Results and discussion

**Phylogenomics, and carbon and energy metabolism of *L. lucinalis* and *Laxus oneistus* symbionts.** The symbiont draft genomes from five clam individuals (*Ca.* Thiodiazotropha endoloripes A–E) were 100% complete for a set of 281 marker genes conserved across all Gammaproteobacteria and ranged in size from 4.46 to 4.88 megabase pairs (Mb) on 12–48 contigs

<sup>1</sup>Department of Microbiology and Ecosystem Science, Division of Microbial Ecology, Research Network Chemistry meets Microbiology, University of Vienna, Althanstrasse 14, Vienna 1090, Austria. <sup>2</sup>Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, Bremen 28359, Germany. <sup>3</sup>Centre for Marine Biodiversity, Exploitation and Conservation (MARBEC), UMR 9190, IRD-IFREMER-CNRS-UM, Université de Montpellier, Montpellier Cedex 5 34095, France. <sup>4</sup>Department of Coastal Systems and Utrecht University, NIOZ Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, Texel, The Netherlands. <sup>5</sup>Department of Geoscience, University of Calgary, 2500 University Drive Northwest, Alberta T2N 1N4, Canada. <sup>6</sup>Archaea Biology and Ecogenomics Division, Department of Ecogenomics and Systems Biology, University of Vienna, Althanstrasse 14, Vienna 1090, Austria. <sup>7</sup>HYDRA Institute for Marine Sciences, Elba Field Station, Campo nell'Elba, Livorno 54037, Italy. \*e-mail: [petersen@microbial-ecology.net](mailto:petersen@microbial-ecology.net)

**Table 1 | Features of bivalve and nematode symbiont genomes.**

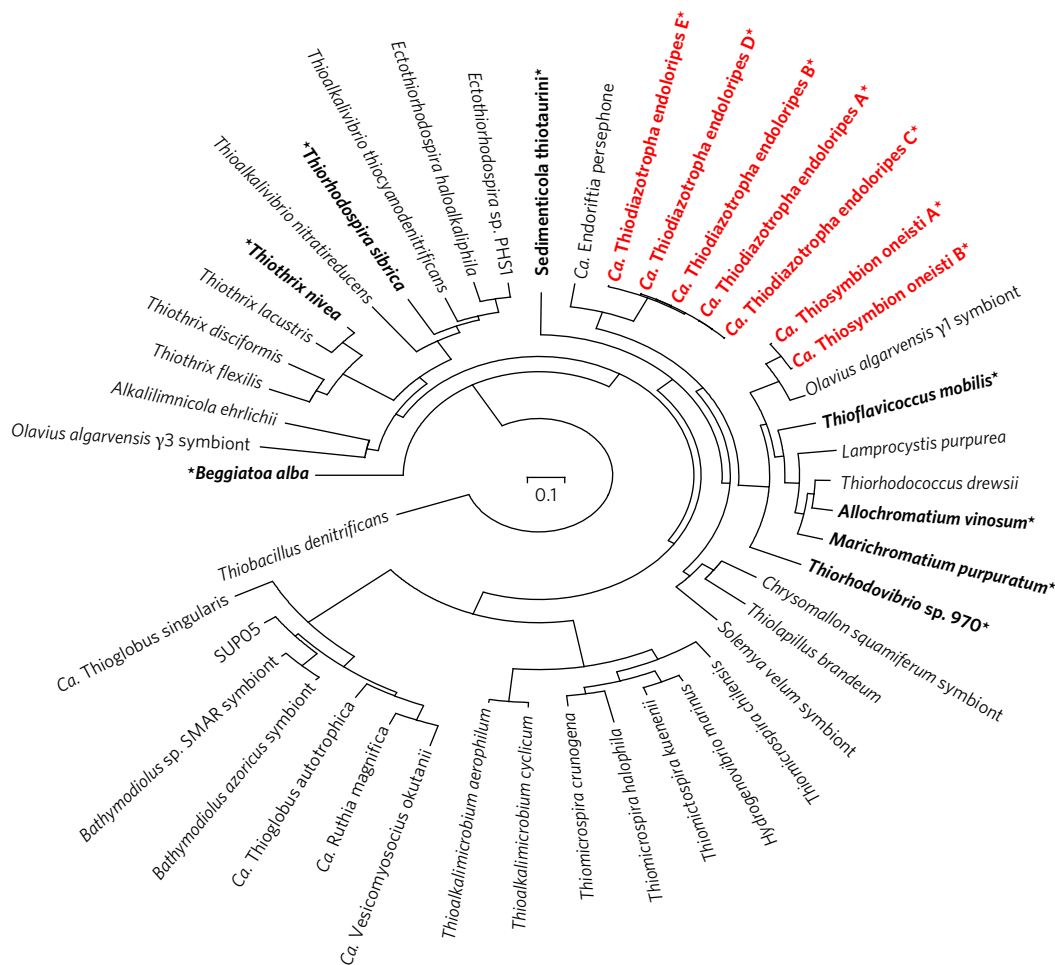
Genome	Size (Mb)	No. of contigs	No. of genes predicted	GC content (%)	Completeness estimate* (%)
<i>Ca. Thiodiazotropha endoloripes</i> A	4.46	15	4,193	52.1	100
<i>Ca. Thiodiazotropha endoloripes</i> B	4.61	18	4,381	51.9	100
<i>Ca. Thiodiazotropha endoloripes</i> C	4.46	18	4,226	52.1	100
<i>Ca. Thiodiazotropha endoloripes</i> D	4.5	12	4,301	52.0	100
<i>Ca. Thiodiazotropha endoloripes</i> E	4.88	48	4,685	51.7	100
<i>Ca. Thiosymbion oneisti</i> A	4.44	2,026	4,149	58.71	96.63
<i>Ca. Thiosymbion oneisti</i> B	4.33	1,891	4,050	58.84	96.07

\*Completeness estimates were calculated based on how many of 281 conserved gammaproteobacterial marker genes were present in each draft genome. See Methods for details.

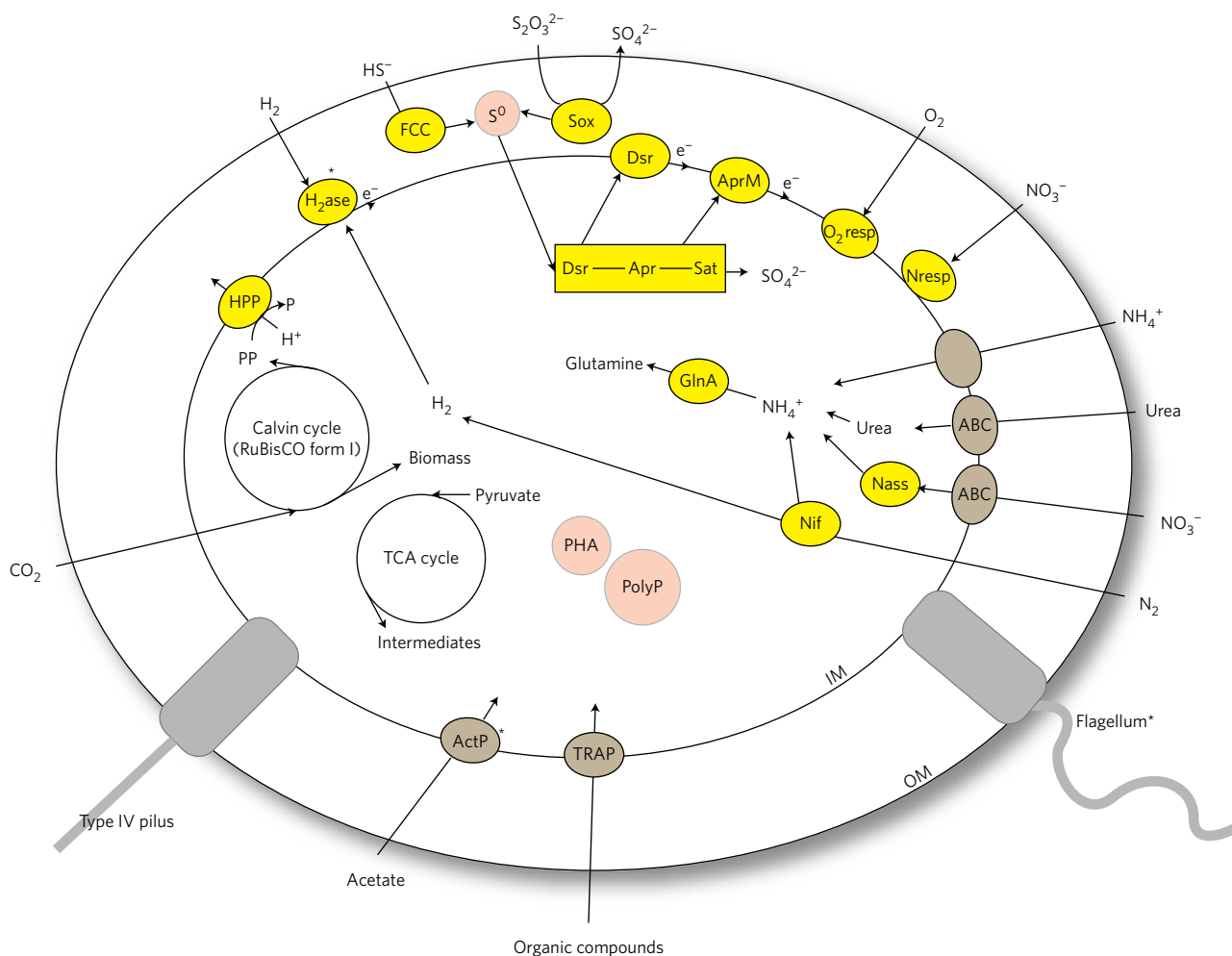
(Table 1). Those from two individuals of *L. oneistus* (*Ca. Thiosymbion oneisti* A–B) were 86.75 and 89.11% complete at sizes of 3.66 and 3.51 Mb on 183 and 193 contigs (Table 1). We conducted phylogenomic analyses to better understand the relationships between the bivalve and nematode symbionts and other symbiotic and free-living gammaproteobacterial sulfur oxidizers. Consistent with previous analyses based on 16S rRNA genes<sup>1</sup>, our phylogenomic analysis placed the lucinid symbionts *Ca. Thiodiazotropha endoloripes* together in a cluster with *Candidatus* Enderiffia persephone, the sulfur-oxidizing symbiont of the hydrothermal vent tubeworm *Riftia pachyptila* (Fig. 1). Also consistent with previous analyses<sup>20</sup>, the nematode symbionts

*Ca. Thiosymbion oneisti* clustered together with the sulfur-oxidizing endosymbionts of the gutless oligochaete worm *Olavius algarvensis* and were affiliated with free-living sulfur oxidizers from the family Chromatiaceae (Fig. 1). The *L. lucinalis* and *R. pachyptila* symbiont cluster was not clearly associated with any known free-living sulfur oxidizers, but formed a sister group to the clade containing the stilbonematid and oligochaete symbionts and members of the Chromatiaceae (Fig. 1).

As expected based on previous studies of symbiont metabolism, genes and pathways for sulfur oxidation and carbon fixation were found in all five *Ca. Thiodiazotropha endoloripes* and in both *Ca. Thiosymbion oneisti* draft symbiont genomes (Fig. 2). Nematode



**Figure 1 | Phylogenomic tree of sulfur-oxidizing Gammaproteobacteria.** Phylogenomic tree of free-living and symbiotic sulfur oxidizers from the Gammaproteobacteria. The 25 single-copy genes used in the analysis were defined based on the AMPHORA2 core bacterial phylogenetic marker database<sup>66</sup>. SUP05 bin refers to the genomic assembly from the metagenome study by Walsh and co-authors<sup>81</sup>. The betaproteobacterial sulfur oxidizer *Thiobacillus denitrificans* was used as the outgroup. SH-like support values were above 90% for all nodes of the tree. Genomes encoding nitrogenase genes are indicated with an asterisk and bold text. Sequences from this study are shown in red.



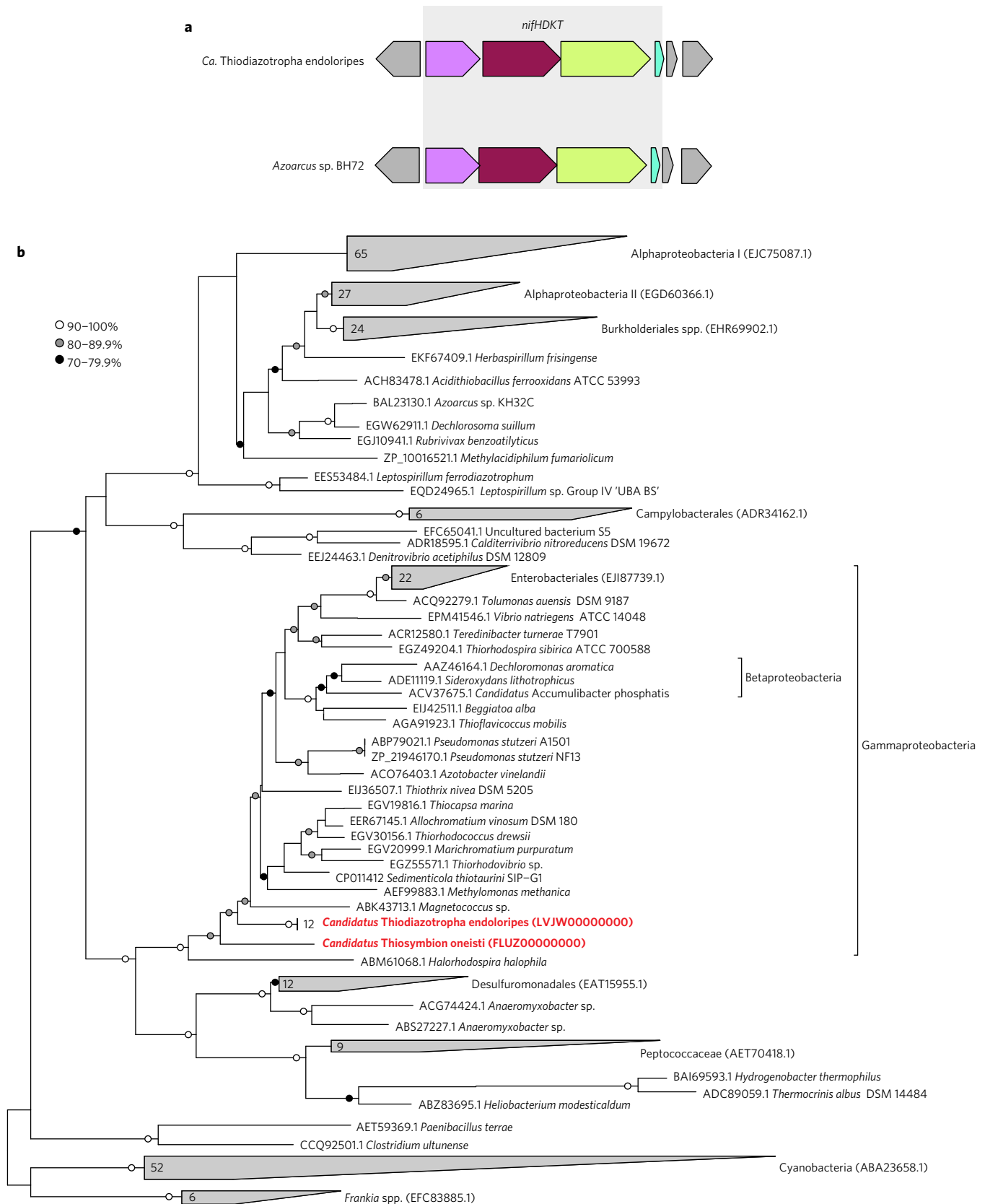
**Figure 2 | Overview of the major cellular features and metabolic pathways encoded in bivalve and nematode symbiont genomes.** Metabolic enzymes and enzyme systems are shown in yellow, transporters in brown, storage granules in pink and structural features in grey. Features only encoded in the draft genomes of *Ca. Thiodiazotropha endoloripes* and not yet found in the draft genomes of *Ca. Thiosymbion oneisti* are indicated by an asterisk. IM, inner membrane; OM, outer membrane; HPP, proton-translocating pyrophosphatase;  $H_2ase$ , uptake hydrogenase; FCC, flavocytochrome *c*;  $S^0$ , elemental sulfur granule; Sox, sox enzyme system for sulfur oxidation; Dsr, reverse dissimilatory sulfite reductase; Apr, adenosine phosphosulfate reductase; AprM, adenosine 5'-phosphosulfate membrane anchor; Sat, sulfate adenylyltransferase;  $O_2 resp$ , genes for oxygen respiration (cytochrome *c* oxidases); Nresp, nitrate respiration (denitrification, pathway complete to  $N_2$ ); Nass, assimilatory nitrate reduction; ActP, acetate transporter; ABC, ABC transporter; TRAP, TRAP transporter; PolyP, polyphosphate granule; PHA, polyhydroxyalkanoate granule; GlnA, glutamine synthetase; Nif, nitrogenase.

and bivalve symbionts encoded a complete tricarboxylic acid (TCA) cycle and transporters for uptake of organic compounds and thus have the potential for heterotrophic growth. Both are capable of using oxygen and oxidized nitrogenous compounds such as nitrate as terminal electron acceptors, but only the bivalve symbiont draft genomes encoded genes for uptake hydrogenases, which would allow them to use hydrogen as an energy source. Bivalve and nematode symbionts both produce intracellular elemental sulfur granules<sup>21,22</sup>. The genomes also revealed the potential to store organic carbon in the form of polyhydroxyalkanoates (PHAs) and phosphorous in the form of polyphosphate granules. A detailed comparative genomics study of the lucinid and stilbonematine symbionts is beyond the scope of this study and will be published elsewhere.

#### Symbionts of *L. lucinalis* and *L. oneistus* encode nitrogenases.

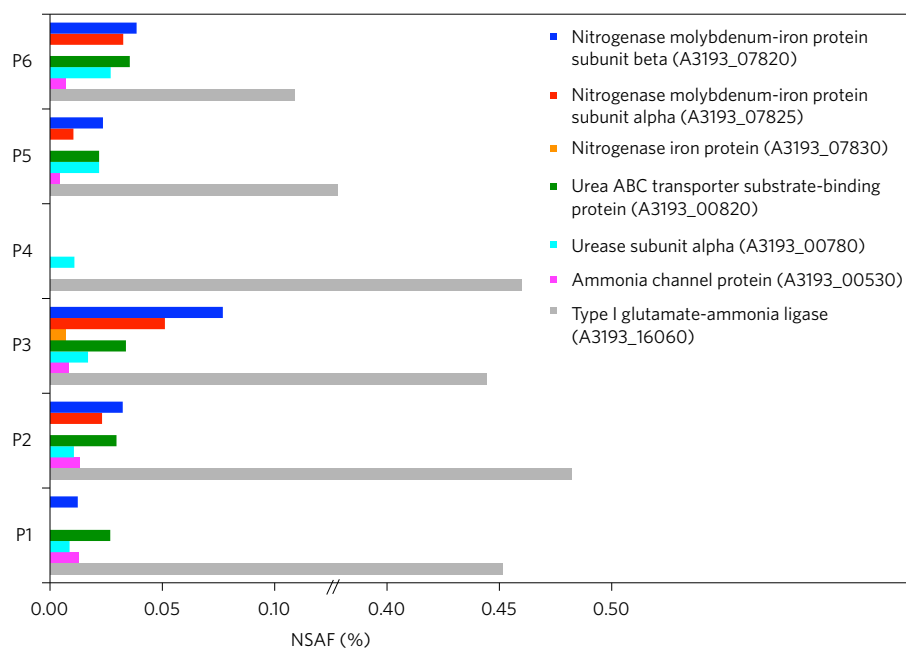
Surprisingly, the five draft genomes of *Ca. Thiodiazotropha endoloripes* and both draft genomes of *Ca. Thiosymbion oneisti* contained large clusters of genes involved in nitrogen fixation, including the structural genes for the iron-molybdenum dinitrogenase (*nifD* and *nifK*), the dinitrogenase reductase subunit (*nifH*), ferredoxins, and maturation and regulatory factors (Fig. 3,

Supplementary Discussion and Supplementary Figs 1 and 2). The *nifH* gene is commonly used as a functional marker for nitrogen fixation, so many *nifH* genes are available in public databases. Our phylogenetic analyses showed that the sulfur-oxidizing symbiont sequences clustered together with Group 1 molybdenum-dependent NifH sequences as defined by Raymond<sup>23</sup> (Fig. 3). The symbiont sequences fell into a clade containing mainly Gammaproteobacteria and some Betaproteobacteria. The NifH sequences from both *Ca. Thiodiazotropha endoloripes* and *Ca. Thiosymbion oneisti* grouped separately to those from other members of the Chromatiaceae, which could indicate that this gene was acquired by horizontal gene transfer in these two symbionts. To gain more insight into the evolutionary history of nitrogen fixation in *Ca. Thiodiazotropha endoloripes* and *Ca. Thiosymbion oneisti*, we analysed the phylogeny of the NifD proteins, which make up one of the two structural subunits of the nitrogenase enzyme. The placement of NifD proteins from the chemosynthetic symbionts was different to the placement of their NifH proteins (Supplementary Fig. 3). The NifD from *Ca. Thiodiazotropha endoloripes* grouped together with the NifD from *Sedimenticola thiotaurini* (Supplementary Fig. 3). This was similar to our phylogenomic



0.10

**Figure 3 | Clusters of *nif* genes and NifH phylogeny. a**, Schematic representation of the *nifHDKT* gene cluster in the *Ca. Thiodiazotropha endoloripes* draft genomes and in the model nitrogen-fixing bacterium *Azoarcus* sp. BH72. See Supplementary Fig. 2 for an overview of the complete *nif* cluster. **b**, Maximum-likelihood phylogeny of full-length Group 1 NifH proteins. Percentages refer to SH-like support values from aLRT. Sequences from this study are shown in bold red text. Numbers in wedges indicate how many sequences are contained in that collapsed clade. Brackets contain examples of specific protein sequences.



**Figure 4 | Expression of proteins for nitrogen assimilation in *Loripes lucinalis*.** Bar chart showing the abundance of seven symbiont proteins involved in assimilation of nitrogen from urea, ammonia and dinitrogen gas that were identified by proteomics in the gills of six *L. lucinalis* individuals (P1–P6). See Supplementary Data set 2 for the complete data set. NSAF, normalized spectral abundance factor.

results, which showed that *S. thiotaurini* grouped together with the Chromatiaceae (Fig. 1). The NifD of *Ca. Thiosymbion oneisti* grouped together with the NifD from *Methylomonas methanica* and not with those from other members of the Chromatiaceae (Supplementary Fig. 3). In summary, there appears to be a history of horizontal transfer of the genes encoding both NifH and NifD, so these proteins could have been horizontally acquired by the chemosynthetic symbionts of lucinid clams and stilbonematid nematodes.

Using PCR primers specific for *nifH*, we screened DNA extracted from six *L. lucinalis* individuals, three from France and three from Mauritania, and DNA from five additional lucinid species from sampling sites around the world (Supplementary Tables 1 and 2). We obtained bands of the correct size from all individuals of *L. lucinalis* and from three other lucinid species (*Euanodontia ovum*, *Codakia orbicularis* and *Clathrolucina costata*) (Supplementary Table 2). We could not obtain a PCR product from one individual each of *Lucinoma borealis* and *Epidulcina cf delphinae*, although the symbiont 16S rRNA gene could be amplified from these samples (Supplementary Table 2). The PCR products were directly sequenced and, although they were too short to determine their phylogenetic position reliably, they were highly similar to the sequences we obtained from the symbiont genomes. All *nifH* sequences from *L. lucinalis* symbionts were between 92 and 100% identical at the nucleotide level (97–100% identity at the amino acid level). Among symbionts of different lucinid species, the *nifH* sequence identity ranged from 83 to 88% (91–98% at the amino acid level). Genome sequencing of these symbiont species would confirm whether they also encode all genes necessary for nitrogen fixation. However, the presence of the *nifH* gene raises the possibility that the symbionts of many lucinid species might be capable of nitrogen fixation.

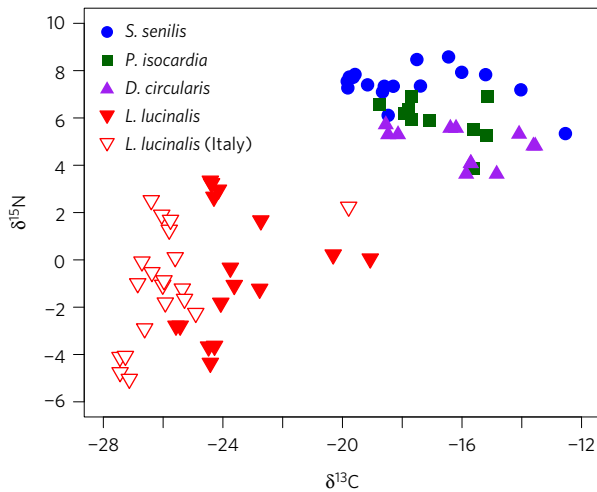
Nitrogen-fixing symbioses are common in marine ecosystems, particularly in habitats where nitrogen availability limits primary production, such as oligotrophic coral reefs and open ocean water<sup>24,25</sup>. Chemosynthetic symbionts are well known for their contribution to host nutrition through carbon fixation, but so far, no marine chemosynthetic symbiont was known to be capable of

nitrogen fixation. Nitrogen fixation has been hypothesized in the only known chemosynthetic symbiosis in the terrestrial environment between ectosymbiotic *Thiothrix*-related bacteria and *Niphargus* amphipods, because *nifH* transcripts could be PCR-amplified from the *Niphargus* ectosymbiotic community<sup>26</sup>. In contrast to previous PCR-based studies, we could unambiguously associate the *nifH* sequence with the rest of the symbiont genome. The presence of these genes in the symbionts of nematodes and bivalves, two unrelated hosts from different animal phyla, shows that the ability to fix nitrogen is not restricted to one phylogenetic group of symbionts, or to the symbionts of only one animal group, but may be widespread in chemosynthetic symbioses.

#### Nitrogen fixation genes are expressed by *L. lucinalis* symbionts.

To test whether nitrogen fixation genes are actively expressed by the lucinid symbionts when living in their hosts, we sequenced the gill metatranscriptomes of five individuals and analysed gill metaproteomes of another six individuals of *L. lucinalis* from Elba (Italy). Transcripts from genes involved in nitrogen fixation were among the 30 most abundantly expressed genes in two of these five individuals (Supplementary Fig. 4 and Supplementary Data set 1). In gill metaproteomes from six *L. lucinalis* individuals, between 892 and 1,377 symbiont proteins could be detected (Supplementary Data set 2). Nitrogenase proteins were detected in five of these six individuals (Fig. 4 and Supplementary Data set 2). Nitrogen fixation is therefore one of the metabolic pathways actively expressed by the symbionts in some *L. lucinalis* individuals (see Supplementary Discussion for further details).

It is remarkable that the symbionts in the animal tissue actively express nitrogenases and seem to be nitrogen-limited (Supplementary Discussion). Because nitrogen fixation is metabolically costly, it is often downregulated when other nitrogen sources are available in the environment<sup>27,28</sup>. As far as we are aware, the concentrations of nitrate, ammonium and urea have not yet been measured in lucinid tissues. Animals are known to produce ammonia and urea as nitrogenous waste products and both of these could be used as nitrogen sources by the lucinid and stilbonematid symbionts (Fig. 2). It is possible that symbiont nitrogen fixation is regulated



**Figure 5 | Carbon and nitrogen stable isotope ratios (‰) of symbiont-free and symbiont-hosting bivalves.** Both  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  signatures (‰) of *L. lucinalis* ( $N = 13$ ) are significantly more depleted compared to three symbiont-free bivalve species *S. senilis* ( $N = 17$ ), *P. isocardia* ( $N = 10$ ) and *D. circularis* ( $N = 5$ ) found on the intertidal flats of Banc d'Arguin, Mauritania. See Supplementary Table 4 for parameter estimates of the models, with carbon and nitrogen isotope signature (‰) modelled as a function of species. All samples were collected at Banc d'Arguin, Mauritania, unless otherwise specified.

by the availability of these nitrogen sources in the gill tissues, or in the immediate environment of the bivalves, similar to the scenario in tropical corals, where primary productivity by zooxanthellae benefits from fixed nitrogen when other nutrient sources are lacking<sup>29</sup>. It is interesting to note the parallel between carbon and nitrogen fixation in lucinid symbiosis. The symbionts clearly also fix carbon inside the animal tissue, although the environment experienced by the symbionts inside the host is presumably replete with organic compounds and the symbionts have the capability to grow heterotrophically. In the future, it would be intriguing to investigate how the presence of alternative carbon and nitrogen sources influences the regulation of nitrogen and carbon fixation by the symbionts.

#### Signature of nitrogen fixation in *L. lucinalis* stable isotope ratios.

Stable isotopes provide a valuable record of the nutrition sources used by organisms in their natural environment. The stable nitrogen and carbon isotopic composition of the biomass of primary producers is usually lighter than their inorganic carbon and nitrogen source due to the slight preference enzymes such as the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and nitrogenase have for the lighter isotope<sup>30</sup>. This shift is called fractionation. In consumers, fractionation leads to enrichment of  $^{13}\text{C}$  and  $^{15}\text{N}$  because of the preferential release of the lighter isotopes and induces a shift in the remaining biomass to heavier values<sup>31</sup>. Newly fixed nitrogen typically has a  $\delta^{15}\text{N}$  between  $-2$  and  $0$ ‰ (ref. 30). This is very similar to the composition of atmospheric nitrogen gas (by definition  $\delta^{15}\text{N} = 0$ ‰), because fractionation during microbial nitrogen fixation is minimal<sup>30</sup>. In contrast, values for heterotrophic marine filter-feeding bivalves are typically larger than  $6$ ‰ because they feed on material that is enriched in  $^{15}\text{N}$  (ref. 32). We analysed the stable carbon and nitrogen isotopic composition of *L. lucinalis* and three symbiont-free bivalve species, *Senilia senilis*, *Pelecypora isocardia* and *Diplodonta circularis* (see Supplementary Table 3 for an overview of stable isotope data). These species co-occur with *L. lucinalis* in intertidal flats of the Banc d'Arguin, Mauritania, and thus presumably have access to the same nutrient sources in their

environment<sup>33</sup>. For comparison, we also analysed the stable carbon and nitrogen isotopic composition of 20 *L. lucinalis* individuals from Elba (Supplementary Table 3).

The model including an effect of species was significantly better than the intercept model for both carbon (log ratio statistic = 72.83, degrees of freedom = 4,  $P < 0.0001$ ) and nitrogen stable isotope ratios (log ratio statistic = 98.89, degrees of freedom = 4,  $P < 0.0001$ ). As expected, the  $\delta^{13}\text{C}$  values of *L. lucinalis* ( $\delta^{13}\text{C}$  (mean  $\pm$  s.d.) =  $-23.4 \pm 1.8$ ‰) were significantly more depleted than those of the three symbiont-free bivalve species *S. senilis* ( $\delta^{13}\text{C} = -17.7 \pm 2.2$ ‰), *P. isocardia* ( $\delta^{13}\text{C} = -16.9 \pm 1.3$ ‰) and *D. circularis* ( $\delta^{13}\text{C} = -16.4 \pm 2.0$ ‰), consistent with a chemosynthetic source of carbon for *L. lucinalis* (Fig. 5 and Supplementary Table 4). The  $\delta^{13}\text{C}$  values of *L. lucinalis* from Elba ( $\delta^{13}\text{C} = -25.9 \pm 1.6$ ‰) were significantly more depleted than those collected at Banc d'Arguin ( $-2.6$ , standard error (SE) 1.0,  $P = 0.01$ ; Fig. 5 and Supplementary Table 4). *L. lucinalis* can also filter-feed and can vary the relative proportion of nutrition it gains from heterotrophic feeding depending on the environmental conditions<sup>34</sup>. The difference in  $\delta^{13}\text{C}$  signatures between Banc d'Arguin and Elba suggests that *L. lucinalis* from the subtidal Elba habitat rely on their endosymbionts for a larger proportion of their carbon nutrition than those from the intertidal Banc d'Arguin, although this difference could also be due to the different handling of samples from Italy and Mauritania, as those from Italy were frozen immediately after sampling and those from Mauritania were first kept in aquaria for 24 h.

The  $\delta^{15}\text{N}$  values of *L. lucinalis* ( $\delta^{15}\text{N}$  (mean  $\pm$  s.d.) =  $-0.5 \pm 2.8$ ‰) were also significantly more depleted than those of the three symbiont-free species *S. senilis* ( $\delta^{15}\text{N} = 7.4 \pm 0.8$ ‰), *P. isocardia* ( $\delta^{15}\text{N} = 5.9 \pm 0.9$ ‰) and *D. circularis* ( $\delta^{15}\text{N} = 5.1 \pm 0.7$ ‰) (Fig. 5 and Supplementary Table 4). The  $\delta^{15}\text{N}$  values of *L. lucinalis* from Banc d'Arguin and those from Elba ( $\delta^{15}\text{N} = -1.1 \pm 2.3$ ‰) were not significantly different ( $-0.61$ , SE = 1.03,  $P = 0.56$ , Supplementary Table 4). The mean  $\delta^{15}\text{N}$  value of *L. lucinalis* (Banc d'Arguin,  $-0.5$ ‰; Elba,  $-1.1$ ‰) fell exactly in the range expected for newly fixed nitrogen. This is consistent with active nitrogen fixation by the symbionts, a result that should be corroborated in future studies, for example by isotope tracer experiments.

The  $\delta^{15}\text{N}$  values for *L. lucinalis* from both Mauritania and Italy varied remarkably between individuals, and ranged from  $-4.4$ ‰ to  $+3.3$ ‰ (Mauritania) and from  $-5.0$ ‰ to  $+2.5$ ‰ (Elba) (Fig. 5). This is intriguing considering our transcriptomics and proteomics results, which showed that nitrogenases were not detectable in all *L. lucinalis* individuals analysed. The variability in  $\delta^{15}\text{N}$  signatures between individuals might reflect differences in the relative contribution of nitrogen fixation, the uptake of alternative nitrogen sources such as ammonium, urea or nitrate by the lucinid symbionts, and filter feeding by the host.

#### Concurrent aerobic and anaerobic metabolism in *L. lucinalis* symbiont populations.

All known nitrogenase enzymes are extremely sensitive to oxygen<sup>35</sup>. The bivalve gill, which is the site for aerobic gas exchange in the animal, therefore seems to be an unusual location for nitrogen-fixing bacteria, but this is not unprecedented—the nitrogen-fixing symbionts of wood-boring bivalves are housed in the host's gill<sup>36</sup>. The symbionts of wood-boring bivalves have been clearly shown to fix nitrogen when inside the gill<sup>37</sup>. Oxygen respiration by bacteria requires terminal cytochrome oxidases, of which a number of different types are known. For example, aa3 type terminal oxidases function best under atmospheric oxygen concentrations, while cbb3 type terminal oxidases function best under reduced oxygen concentrations<sup>38</sup>. *Bradyrhizobium japonicum*, the nitrogen-fixing symbiont of leguminous plants, switches to expressing a cbb3 type terminal oxidase to adapt to the low oxygen concentrations found in the root nodules of plants<sup>39</sup>. The plant host expresses

leghaemoglobins that bind oxygen, keeping concentrations low in root nodules and preventing damage to the nitrogenases expressed by their nitrogen-fixing symbionts<sup>40</sup>. *L. lucinalis* is not known to produce haemoglobins, but they have been reported in other lucinid species<sup>41</sup>. *Ca. Thiodiazotropha endoloripes* encodes both the low-affinity aa3 and the high-affinity cbb3 type oxidases. Transcripts for the aa3 and the high-affinity cbb3 type oxidases were among the 30 most abundant in all transcriptomes and were also detected in all six proteomes, suggesting that at least some symbionts experience relatively high oxygen concentrations (Supplementary Fig. 4 and Supplementary Data sets 1 and 2). The cbb3 type was also expressed by the symbionts in all five bivalves analysed for transcriptomics, and all six proteomes (Supplementary Data sets 1 and 2). It is also intriguing that a gene similar to anaerobic dimethyl sulfoxide reductases was one of the most highly expressed genes in all five transcriptomes and all six proteomes. Terminal molybdopterin oxidoreductases such as dimethyl sulfoxide reductases allow anaerobic respiration using terminal electron acceptors other than oxygen, in this case, possibly dimethyl sulfoxide (DMSO)<sup>42</sup>. Aerobic and anaerobic metabolism might therefore coexist in a single host individual, possibly by different symbiont subpopulations (Supplementary Discussion). Sub-oxic or anoxic niches in the gill tissue would provide ideal conditions for nitrogen-fixing symbionts. Alternatively, the bivalve host may control the oxygen concentrations experienced in the gill, temporally, by controlling the rate at which it pumps oxygenated sea water through its burrow. Finally, in the stilbonematid nematodes, the host's migratory behaviour would allow its symbionts to experience both oxic conditions in shallow sediment layers and anoxic conditions in the deeper sediment<sup>43</sup>. It is therefore possible that the nematode ectosymbionts fix nitrogen when the nematodes inhabit deep anoxic sediment layers.

**Potential roles of nitrogen fixation for the symbiosis and the ecosystem.** The discovery of nitrogen-fixing chemosynthetic symbionts was at first glance surprising, considering that marine animals produce ammonium and urea as nitrogenous waste products<sup>44</sup> and many chemosynthetic symbionts including those from *Ca. Thiodiazotropha endoloripes* and *Ca. Thiosymbion oneisti* have the genetic potential to use ammonium or urea as nitrogen sources (Fig. 2). Indeed, in the *Olavius algarvensis* (oligochaete annelid) chemosynthetic symbiosis, recycling of host nitrogenous waste products by the symbiotic bacteria may have resulted in the reduction of the host nephridia, the organs responsible for processing nitrogenous waste<sup>3</sup>. Although the recycling of host nitrogenous waste would undoubtedly result in an overall more efficient nitrogen use by the symbiosis, an external source of nitrogen would still be needed for net growth<sup>45</sup>. Just as carbon fixation by chemosynthetic symbionts provides a source of carbon for host nutrition, nitrogen fixation by chemosynthetic symbionts may also provide a source of nitrogen for their host.

Chemosynthetic symbioses are often found in oligotrophic habitats such as coral reef sands or sediments associated with seagrass meadows. Presumably, animal hosts living in such oligotrophic environments would benefit most from hosting nitrogen-fixing chemosynthetic symbionts. Although we do not currently know the concentration of nutrients at our sampling sites, inorganic nitrogen concentrations in seagrass ecosystems are typically one to two orders of magnitude higher in pore waters (~1–180  $\mu\text{M}$ ) than in the overlying water column<sup>46</sup>. However, in a chemosynthetic symbiosis, where carbon is rapidly fixed, organic carbon can be in excess in relation to the demand of the autotrophic microbial community<sup>47</sup>. Consequently, even if other sources of nitrogen are available in their environment, the symbiont population may be nitrogen-limited and thus benefit from nitrogen fixation.

Following on from this, nitrogen fixation by chemosynthetic symbionts could contribute a source of new nitrogen to the ecosystems they inhabit. Intriguingly, one study of seagrass sediment biogeochemistry showed that porewater ammonium concentrations were higher when lucinid bivalves were present<sup>48</sup>. Seagrasses have been shown to grow better in the presence of lucinids, which was hypothesized to be due to the removal of toxic sulfide by their symbionts<sup>6</sup>. Our results point to another possible mechanism for the promotion of seagrass growth: the provision of fixed nitrogen to the porewater of the sediments, where the seagrass roots could take it up. Seagrasses clearly benefit from active plant-associated and free-living diazotrophic microbial communities that provide them with a source of fixed nitrogen<sup>24,49</sup>. If lucinid bivalves and their symbionts are providing a net source of nitrogen to their environments, then they might also contribute to the ecosystem nitrogen budget in habitats where they are abundant.

## Methods

**Sample collection.** Lucinid bivalves of the species *L. lucinalis* were collected in the Bay of Fetovaia, Elba, Italy, for genome and transcriptome sequencing, proteomics and stable isotope analysis (see Supplementary Table 1 for more information on sampling). The bivalves were found 10–50 cm deep in sediment 4–10 m away from a seagrass meadow of *Posidonia oceanica*. Samples for DNA sequencing were preserved in RNAlater (Thermo Fisher Scientific). Samples for transcriptome sequencing were removed from the shells and placed in RNAlater immediately after sampling and transport to the field station, which took no longer than 45 min. Proteomics samples were transported to a boat on the surface in a plastic container immediately after being removed from the sediment. Once on the boat, the clams were frozen immediately on dry ice and stored on dry ice or at  $-80\text{ }^{\circ}\text{C}$  or colder until protein extraction. The time from collection from the sediment to freezing was no more than 10 min. The shell length (mm) of bivalves for stable isotope analysis was measured using calipers, then the flesh was removed from the shell and frozen at  $-20\text{ }^{\circ}\text{C}$ .

*Loripes lucinalis* and three co-occurring non-symbiotic (that is, heterotrophic) bivalve species (*Senilia senilis* (Linnaeus, 1758), *Pelecypora isocardia* (Dunker, 1845) and *Diplodonta circularis* (Dunker, 1846)) were collected from the intertidal flats surrounding Iwik Peninsula, Banc d'Arguin, Mauritania, for stable isotope analyses (Supplementary Table 1). A sediment core (15 cm diameter, 20 cm depth) was taken at 40 different sampling sites, some bare, some covered with *Zostera noltii* seagrasses, and sieved over a 1 mm mesh. The 40 sites were spread over an area of ~36 km<sup>2</sup>. From each benthic sample, bivalves were sorted within 12 h at the field station in Iwik. The shell length (mm) of each bivalve was measured with calipers and one individual per species was placed in an aquarium with 0.7- $\mu\text{m}$ -filtered, oxygenated sea water to clear their gut contents. Animals at least 4 mm in length were chosen to provide enough biomass for carbon and nitrogen isotope measurements. After 24 h, bivalves were taken out of the aquarium, then the flesh was removed from the shell and stored frozen at  $-20\text{ }^{\circ}\text{C}$ .

*L. lucinalis* for PCR amplification of *nifH* and 16S rRNA genes were collected from Banc d'Arguin, Mauritania, and Thau Lagoon, France (Supplementary Table 1). In Mauritania, bivalves were sieved from seagrass sediments and roots, and the shells were immediately opened before the bivalves were stored in RNAlater at  $-20\text{ }^{\circ}\text{C}$ . In France, *L. lucinalis* were sorted from sediment within 10 min after collection and immediately fixed in 96% ethanol. Bivalves were stored at room temperature until DNA extraction.

*Codakia orbicularis* (Linnaeus, 1758) was collected from St George's Cay, Bahamas, under permit from the Bahamas Government (N. Higgs, MAMR&LG/FIS/17). Gill tissues were dissected and fixed in RNAlater shortly after collection, then stored at  $-20\text{ }^{\circ}\text{C}$ .

*Euanodontia ovum* (Reeve, 1850) was collected from Mauritius, and *Clathrolucina costata* (d'Orbigny, 1842) was collected from Curaçao (Supplementary Table 1). Both species were sampled by digging in seagrass beds with a shovel and trowels. Seagrass rhizomes and root masses were sieved through 2 mm mesh sieves. After collection, the bivalves were placed in cool bags then later slightly opened in the laboratory with a scalpel blade and preserved in 100% ethanol. The ethanol was changed after one day. Samples were stored at room temperature until processing.

*Laxus oneistus* was collected from Carrie Bow Cay, Belize. Nematodes were extracted from the sand by shaking in sea water and pouring the supernatant through a mesh screen with a pore size of 63  $\mu\text{m}$ . Single individuals were then picked by hand under a dissecting microscope then stored in RNAlater.

**DNA extraction, genome sequencing and analysis.** DNA for metagenome sequencing was extracted from homogenates from a single whole gill from each of five *L. lucinalis* from Elba (*L. lucinalis* A–E) according to the method of Zhou and colleagues<sup>50</sup>. DNA was extracted from two whole *L. oneistus* individuals (*L. oneistus* A and B) with intact symbiotic coat using the Blood and Tissue Micro Kit (Qiagen) according to the manufacturer's instructions. After extraction, DNA concentration was measured with the fluorometric quantitation tool Qubit (Life Technologies).

From each lucinid individual, 1 µg DNA was used for PE-library preparations using the NEBNext kit (New England Biolabs). For each stilbonematid individual, 5 ng DNA was used with an Ovation Ultralow Library Systems kit (NuGEN). All libraries were size-selected on an agarose gel and sequenced at the Max Planck Genome Centre in Cologne, Germany. Illumina HiSeq 2 × 100 bp paired-end reads were sequenced for each lucinid individual (30 million reads each) and each stilbonematid (18 million reads each). The assembly for all samples was done with SPAdes 3.1 (ref. 51) after removing adapters and low-quality reads with bbduk (<https://sourceforge.net/projects/bbmap/>) by setting a minimum quality value of two and a minimum length of 36. Single reads were excluded from the analysis. To largely remove the host genomic reads for the lucinid samples, the reads were split based on a kmer frequency analysis performed with bbnorm. Only reads with average kmer frequencies of 30 for *L. lucinalis* A, B and D, 25 for *L. lucinalis* C, 50 for *L. lucinalis* E and 25 for *L. oneistus* A and B were kept using bbnorm (<https://sourceforge.net/projects/bbmap/>). The assembly was carried out using kmers 21, 33, 55 and 77. The initial binning was done using Metawatt 2.1 (ref. 52). Final binning was done interactively by collecting all contigs linked to this initial high confidence bin using the FASTG linkage information provided from the SPAdes assembly program with gbtools version 2.4.4 (ref. 53). We binned a draft symbiont genome from each of five individuals of *L. lucinalis*. No other bacterial genomes were identified in our five metagenomic libraries. The genome completeness for all samples was calculated using checkM version 1.05 (ref. 54) and the gammaproteobacterial marker gene set using the taxonomy workflow. The annotation was performed using RAST (ref. 55).

For PCR amplification of 16S rRNA and *nifH* genes from diverse lucinid species, DNA was extracted with a DNeasy Blood and Tissue kit from Qiagen according to the manufacturer's instructions.

**RNA extraction, transcriptome sequencing and analysis.** DNA and RNA were co-extracted from one whole gill of five *L. lucinalis* individuals from the Bay of Fetoavia, Elba (different from those used for genome sequencing), using an AllPrep DNA/RNA mini kit according to the manufacturer's instructions (Qiagen). From each sample, at least 10 ng RNA were used for paired-end library preparations. Total RNA was converted to double-stranded cDNA and amplified linearly with the Ovation RNA-Seq System V2 kit (NuGEN). Then, 300 ng double-stranded cDNA was fragmented to an average size of 400 bp (Covaris). Illumina-compatible libraries were generated with the NEBNext Ultra DNA Library Prep Kit for Illumina kit (NEB). Finally, fragments were enriched by a PCR step for six cycles. Quality assessment was carried out at various steps at the RNA or DNA level with an Agilent Bioanalyser. DNA was quantified by fluorometry (Qubit, Thermo Fisher Scientific). Libraries were quantified by fluorometry, immobilized and processed onto a flow cell with a cBot (Illumina), followed by sequencing with TruSeq v3 chemistry on a HiSeq2500 at the Max Planck Genome Centre in Cologne, Germany. All libraries were sequenced multiplexed on single lanes of two consecutive runs that yielded a total of 26 million Illumina HiSeq 2 × 100 bp paired-end reads per individual. Paired reads generated from mRNA transcripts were indexed and aligned to genes predicted from annotation of *Ca. Thiodiazotropha endoloripes* A using bwa version 0.7.12-r1039 (index, aln and sampe) with default parameters<sup>56</sup>. We mapped the transcriptome reads from each individual sequenced to all five symbiont draft genomes (generated from another five individual clams). Approximately 0.5% of the transcriptome reads from each individual mapped to symbiont protein-coding genes (Supplementary Fig. 5). We plotted the identity (%) of each transcriptome read mapped to each genome. This analysis showed similar patterns for *Ca. Thiodiazotropha endoloripes* A, B, C and D, with the majority of mapped reads being 100% identical to the genome to which they were mapped, but the identities to E were generally lower (Supplementary Fig. 5). These results may indicate that *Ca. Thiodiazotropha endoloripes* A, B, C and D and all five symbiont populations used for transcriptome sequencing were highly similar and *Ca. Thiodiazotropha endoloripes* E was by chance a slightly different strain. We therefore continued the transcriptome analysis with the results of mapping to one representative symbiont genome, *Ca. Thiodiazotropha endoloripes* A. Mapping files were sorted and the number of properly paired mapped reads specific to each gene was extracted with Samtools view (-f 3)<sup>57</sup>. To rank the genes detected in the transcriptome of each individual according to the abundance of their transcripts, we calculated the number of transcripts that mapped to each gene as a percentage of the total number of transcripts that mapped to all genes (% total counts). We then adjusted this percentage according to gene length and ranked the genes according to the percentage of all transcripts that mapped per kb per gene (% of total counts adjusted to gene length).

**Protein extraction and analysis.** We prepared tryptic digests from six biological replicates (one whole gill from each of six individuals frozen immediately after sampling, see section 'Sample collection') following the filter-aided sample preparation (FASP) protocol described by Wiśniewski *et al.*<sup>58</sup> with some small modifications as described by Hamann and co-authors<sup>59</sup>. Peptides were desalted using Sep-Pak C18 Plus Light Cartridges (Waters) according to the manufacturer's instructions. Approximate peptide concentrations were determined using the Pierce Micro BCA assay (Thermo Scientific Pierce) following the manufacturer's instructions.

Samples were analysed by one-dimensional liquid chromatography tandem mass spectrometry. For each sample a technical replicate was run. Two blank runs were done between samples to reduce carry over. For each run, 2,000 ng of peptide were loaded onto a 2 cm, 75 µm ID C18 Acclaim PepMap 100 pre-column (Thermo Fisher Scientific) using an EASY-nLC 1000 liquid chromatograph (Thermo Fisher Scientific) set up in two-column mode. The pre-column was connected to a 50 cm × 75 µm analytical EASY-Spray column packed with PepMap RSLC C18, 2 µm material (Thermo Fisher Scientific), which was heated to 35 °C using the integrated heating module. The analytical column was connected via an Easy-Spray source to a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were separated on the analytical column at a flow rate of 225 nl min<sup>-1</sup> using a 460 min gradient going from buffer A (0.2% formic acid, 5% acetonitrile) to 20% buffer B (0.2% formic acid in acetonitrile) in 354 min, then from 20 to 35% B in 71 min and ending with 35 min at 100% B. Eluting peptides were ionized with electrospray ionization (ESI) and analysed in the Q Exactive Plus. Full scans were acquired in the Orbitrap at 70,000 resolution. MS/MS scans of the 15 most abundant precursor ions were acquired in the Orbitrap at 17,500 resolution. The mass (*m/z*) 445.12003 was used as lock mass as described by Olsen *et al.*<sup>60</sup> with the modification that lock mass was detected in the full scan rather than by separate SIM scan injection. Lock mass use was set to 'best'. Ions with charge state +1 were excluded from MS/MS analysis. Dynamic exclusion was set to 30 s. Roughly 500,000 MS/MS spectra were acquired per sample (two technical replicates combined).

For protein identification, a database was created using all protein sequences predicted from the *Ca. Thiodiazotropha endoloripes* genomes published in this study. CD-HIT was used to remove redundant sequences from the database<sup>61</sup>. The cRAP protein sequence database (<http://www.thegpm.org/crap/>) containing protein sequences of common laboratory contaminants was appended to the database. The final database contained 6,816 protein sequences. The database was submitted to the PRIDE repository (see section 'Data availability'). For protein 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) with the following parameters: Trypsin (Full), max. 2 missed cleavages, 10 ppm precursor mass tolerance, 0.1 Da fragment mass tolerance and max. 3 equal dynamic modifications per peptide. The following three dynamic modifications were considered: oxidation on M (+15.995 Da), carbamidomethyl on C (+57.021 Da) and acetyl on the protein N terminus (+42.011 Da). False discovery rates (FDRs) for peptide spectral matches (PSMs) were calculated and filtered using the Percolator Node in Proteome Discoverer. The Percolator algorithm 'uses semi-supervised learning and a decoy database search strategy to learn to distinguish between correct and incorrect PSMs'<sup>62</sup>. Percolator was run with the following settings: maximum Delta Cn 0.05, a strict target FDR of 0.01, a relaxed target FDR of 0.05 and validation based on *q*-value. The Protein FDR Validator Node in Proteome Discoverer was used to classify protein identifications based on *q*-value. Proteins with a *q*-value of <0.01 were classified as high-confidence identifications and proteins with a *q*-value of 0.01–0.05 were classified as medium-confidence identifications. Only proteins identified with medium or high confidence were retained, resulting in an overall FDR of 5%. Based on these filtering criteria, between 892 and 1,377 proteins were identified per sample. The sample reports were then exported as a tab-delimited file for further processing.

For protein quantification, normalized spectral abundance factors (NSAFs) were calculated based on the number of PSMs per protein using the method described by Florens *et al.*<sup>63</sup> and multiplied by 100. The NSAFx100 gives the relative abundance of a protein in a sample in %.

**PCR amplification and sequencing of *nifH* and 16S rRNA genes.** DNA extracted from diverse lucinid species was used as a template for amplification using 16S rRNA- and *nifH*-specific primers and the DreamTaq DNA Polymerase (Thermo Fisher Scientific). Primers 27F and 1492R were used to amplify the 16S rRNA gene<sup>64</sup>. Degenerate primers IGK3/DVV were used to amplify the *nifH* gene<sup>65</sup>. The reaction conditions were as follows: one cycle at 94 °C (4 min); 32 cycles and 34 cycles (for the 16S rRNA and *nifH* genes, respectively) at 94 °C (30 s), 52 °C (45 s) and 72 °C (45 s); plus one final cycle at 72 °C (10 min). The resulting PCR products were purified with a QIAquick PCR Purification Kit from Qiagen and Sanger sequenced by Microsynth AG.

**Phylogenomic analysis.** Available genomes of gammaproteobacterial sulfur oxidizers were retrieved from GenBank and the Joint Genome Institute's Integrated Microbial Genomes database (JGI-IMG). Phylogenomic treeing of these genomes along with the lucinid and stilbonematid symbiont genomes was performed using scripts that were available at phylogenomics-tools (doi:10.5281/zenodo.46122). Marker proteins that are universally conserved across the bacterial domain were extracted from the sulfur-oxidizer genomes using the Amphora2 pipeline<sup>66</sup>. Twenty-five markers (*frr*, *infC*, *nusA*, *pgk*, *pyrG*, *rplB*, *rplC*, *rplD*, *rplE*, *rplF*, *rplM*, *rplN*, *rplP*, *rplS*, *rplT*, *rplA*, *rpsB*, *rpsC*, *rpsE*, *rpsI*, *rpsJ*, *rpsK*, *rpsM*, *rpsS* and *tsf*) that were identified to be occurring in single copy in all of the genomes were used for alignment on Muscle<sup>67</sup>. An alignment mask was generated using Zorro<sup>68</sup>. Poorly aligned regions or misaligned regions were visually identified and removed from the alignments. The marker alignments were further concatenated into a single



partitioned alignment and the best protein substitution model for each of the markers (*frt*, *nusA*, *pyrG*, *rplB*, *rplC*, *rplD*, *rplE*, *rplF*, *rplM*, *rplP*, *rplS*, *rplT*, *rpsB*, *rpsC*, *rpsM*, *tsf*, *lG*; *infC*, *rpsE*, *rps*, *rpsS*: JTT; *pgk*, *rpmA*: WAG; *rplN*, *rpsK*: RTREV; *rpsL*: JTTDCMUT) was predicted using the *concat\_align.pl* script (phylogenomics-tools). The best tree with SH-like aLRT support values<sup>69</sup> was finally defined on RAxML (ref. 70) using the *tree\_calculations.pl* script on phylogenomics-tools.

**Phylogenetic analysis of *NifH* and *NifD*.** Symbiont *nifH* and *nifD* gene sequences were obtained from the RAST annotations. Gene sequences from *S. thiotaurini* SIP-G1 (ref. 71) were downloaded from GenBank (CP011412). Sequences were imported using ARB<sup>72</sup> into curated databases for *nifH* ([http://www.zehr.pmc.ucsc.edu/nifH\\_Database\\_Public/](http://www.zehr.pmc.ucsc.edu/nifH_Database_Public/))<sup>73</sup> and *nifD* (<http://www.css.cornell.edu/faculty/buckley/nifH.htm>)<sup>74</sup> respectively and approximate placements were found with the parsimony quick-add feature in ARB. Group I *nifH* sequences were exported and dereplicated at 95% identity with Usearch 8.1 cluster-fast algorithm (length-sorted)<sup>75</sup>. *NifD* sequences were exported and those without nitrogenase alpha chain domains (TIGR01284, TIGR01862) were excluded as likely false positives. For both genes, the amino acid translation was aligned with Muscle 3.7 (ref. 67) and the amino acid alignment was used to guide the nucleotide alignment with TranslatorX (ref. 76). Maximum likelihood trees of amino acid and nucleotide (first and second codon positions only) alignments were calculated with PhyML 3.1 (ref. 77) using the WAG + Gamma and GTR + Gamma models, respectively, both with four Gamma rate categories and SH-like aLRT support values<sup>69</sup>.

**Nitrogen and carbon stable isotope analysis.** Whole individuals were used for stable isotope analysis, so different individuals were used for molecular and stable isotope analysis. Tissue samples of *L. lucinalis* ( $N = 20$ ; shell length (mean  $\pm$  s.d.) =  $9.0 \pm 3.3$  mm) from Elba (Italy) were freeze-dried for 72 h, ground to a fine powder and acidified in an HCl atmosphere to remove traces of carbonate. Dried tissues were weighed before stable isotope analysis on a continuous-flow elemental analyser-isotope ratio mass spectrometer (EA-IRMS) consisting of an elemental analyser (EA 1110, CE Instruments) coupled via a ConFlo III interface (Finnigan MAT, Thermo Fisher) to the IRMS (DeltaPLUS, Finnigan MAT, Thermo Fisher).

Tissue samples of *L. lucinalis* ( $N = 12$ ; shell length (mean  $\pm$  s.d.) =  $8.2 \pm 1.7$  mm), *S. senilis* ( $N = 17$ ; shell length =  $40.3 \pm 16.9$  mm), *P. isocardia* ( $N = 10$ ; shell length =  $11.0 \pm 4.8$  mm) and *D. circularis* ( $N = 5$ ; shell length =  $12.0 \pm 5.4$  mm) collected at Banc d'Arguin, Mauritania, were freeze-dried for 72 h, homogenized by mortar and pestle, weighed into tin cups. Their carbon and nitrogen stable isotope ratios were measured with a Thermo Scientific (Flash 2000) elemental analyser coupled to a Delta V isotope mass spectrometer as described above. Tissue samples from Banc d'Arguin were not acidified before stable isotope analysis, as previous studies showed no effect of tissue treatment with HCl on  $\delta^{13}\text{C}$  values of *L. lucinalis*<sup>34</sup>.

Carbon and nitrogen stable isotope ratios (‰) are expressed in delta notation as  $\delta^{13}\text{C}$  or  $\delta^{15}\text{N}$  ( $= (R_{\text{sample}}/R_{\text{ref}} - 1) \times 1,000$ ).  $R_{\text{sample}}$  is the ratio of the heavy to light isotope ( $^{13}\text{C}/^{12}\text{C}$  or  $^{15}\text{N}/^{14}\text{N}$ ) in the sample and  $R_{\text{ref}}$  is the same ratio for the reference material, the Vienna Pee Dee Belemnite standard for C ( $R_{\text{ref}} = 0.01118$ ) and atmospheric nitrogen for N ( $R_{\text{ref}} = 0.00368$ ).

**Statistical analysis of stable isotope ratios.** To analyse the effect of bivalve species on carbon and nitrogen stable isotope ratios, we used linear mixed-effects models, with random intercepts and with sampling site as a random effect and species as explanatory variable. On six occasions, an individual from two different focal species was collected from the same sampling site at Banc d'Arguin (Mauritania), while multiple individuals of *L. lucinalis* were collected from the same sampling site at Elba (Italy). Assumptions of normality and homogeneity of residuals were visually inspected by plotting quantile–quantile plots and by plotting the standardized residuals versus fitted values and the standardized residuals per species. This revealed species-dependent spread in the error variances in the model with nitrogen stable isotope ratios (‰) as dependent variable, for which we accounted by adding a 'varIdent' variance structure to the model<sup>78</sup>. We tested for the significance of the species effect on stable carbon and nitrogen isotope ratios using likelihood ratio tests. Reported  $P$  values are for a two-tailed test. All analyses were performed in program R (R Development Core Team 2015, version 3.2.2; <http://www.R-project.org>). For linear mixed-effects models, the R-package nlme was used<sup>79</sup>.

**Data availability.** Genome and transcriptome data from *L. lucinalis* have been submitted to NCBI in BioProject PRJNA314435. The draft genome sequences from lucinid symbionts are available under accession nos. LVJW000000000 (*Ca. Thiodiazotropha endoloripes A*), LVJX000000000 (*Ca. Thiodiazotropha endoloripes B*), LVJY000000000 (*Ca. Thiodiazotropha endoloripes C*), LVJZ000000000 (*Ca. Thiodiazotropha endoloripes D*) and LVKA000000000 (*Ca. Thiodiazotropha endoloripes E*). Transcriptome sequences can be found at the NCBI Short Read Archive under accession no. SRP073135. Draft genome sequences from stilbonematid nematodes have been submitted to ENA under project no. PRJEB14785 and can be found under accession nos. FLUZ000000000 (*Ca. Thiosymbiont oneisti A*) and FLUY000000000 (*Ca. Thiosymbiont oneisti B*). PCR-amplified *nifH* sequences can be found under accession nos. LT548937–LT548954 and partial 16S rRNA genes under LT548918–LT548936. The mass

spectrometry proteomics data and the protein sequence database have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD004536<sup>80</sup>.

Received 14 March 2016; accepted 7 September 2016;  
published 24 October 2016

## References

- Dubilier, N., Bergin, C. & Lott, C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat. Rev. Microbiol.* **6**, 725–739 (2008).
- Petersen, J. M. *et al.* Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**, 176–180 (2011).
- Kleiner, M. *et al.* 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 (2012).
- Kleiner, M. *et al.* Use of carbon monoxide and hydrogen by a bacteria–animal symbiosis from seagrass sediments. *Environ. Microbiol.* **17**, 5023–5035 (2015).
- van der Geest, M., van Gils, J. A., van der Meer, J., Olff, H. & Piersma, T. Suitability of calcein as an *in situ* growth marker in burrowing bivalves. *J. Exp. Mar. Biol. Ecol.* **399**, 1–7 (2011).
- van der Heide, T. *et al.* A three-stage symbiosis forms the foundation of seagrass ecosystems. *Science* **336**, 1432–1434 (2012).
- Sayavedra, L. *et al.* Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *eLife* **15**, e07966 (2015).
- Lee, R. W. & Childress, J. J. Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and methanotrophic symbionts. *Appl. Environ. Microbiol.* **60**, 1852–1858 (1994).
- Lee, R. W., Thuesen, E. V. & Childress, J. J. Ammonium and free amino acids as nitrogen sources for the chemoautotrophic symbiosis *Solemya reidi* Bernard (Bivalvia: Protobranchia). *J. Exp. Mar. Biol. Ecol.* **158**, 75–91 (1992).
- Liao, L., Wankel, S. D., Wu, M., Cavanaugh, C. M. & Girguis, P. R. Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*. *Mol. Ecol.* **23**, 1544–1557 (2014).
- Middelburg, J. J. *et al.* Discovery of symbiotic nitrogen fixation and chemoautotrophy in cold-water corals. *Sci. Rep.* **5**, 17962 (2015).
- Brooks, J. M. *et al.* Deep sea hydrocarbon seep communities—evidence for energy and nutritional carbon sources. *Science* **238**, 1138–1142 (1987).
- Fisher, C. R. Chemoautotrophic and methanotrophic symbioses in marine invertebrates. *Rev. Aquat. Sci.* **2**, 399–436 (1990).
- Kennicutt, M. C. *et al.* Stable isotope partitioning in seep and vent organisms—chemical and ecological significance. *Chem. Geol.* **101**, 293–310 (1992).
- Taylor, J. D. & Glover, E. A. Lucinidae (Bivalvia)—the most diverse group of chemosymbiotic molluscs. *Zool. J. Linn. Soc.* **148**, 421–438 (2006).
- Stanley, S. M. Evolutionary radiation of shallow-water Lucinidae (Bivalvia with endosymbionts) as a result of the rise of seagrasses and mangroves. *Geology* **42**, 803–806 (2014).
- Imhoff, J. F. in *The Prokaryotes* (eds Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 151–178 (Springer, 2014).
- Pende, N. *et al.* Size-independent symmetric division in extraordinarily long cells. *Nat. Commun.* **5**, 4803 (2014).
- Ott, J., Bright, M. & Bulgheresi, S. Symbioses between marine nematodes and sulfur-oxidizing chemoautotrophic bacteria. *Symbiosis* **36**, 103–126 (2004).
- Zimmermann, J. *et al.* Closely coupled evolutionary history of ecto- and endosymbionts from two distantly-related animal phyla. *Mol. Ecol.* **25**, 3203–3223 (2016).
- Himmel, D., Maurin, L. C., Gros, O. & Mansot, J.-L. Raman microspectrometry sulfur detection and characterization in the marine ectosymbiotic nematode *Eubostrius dianae* (Desmodoridae, Stilbonematidae). *Biol. Cell* **101**, 43–54 (2009).
- Johnson, M. A. & Fernandez, C. Bacterial symbiosis in *Loripes lucinalis* (Mollusca: Bivalvia) with comments on reproductive strategy. *J. Mar. Biol. Assoc. UK* **81**, 251–257 (2001).
- Raymond, J. The natural history of nitrogen fixation. *Mol. Biol. Evol.* **21**, 541–554 (2003).
- Cardini, U., Bednarz, V. N., Foster, R. A. & Wild, C. Benthic  $\text{N}_2$  fixation in coral reefs and the potential effects of human-induced environmental change. *Ecol. Evol.* **4**, 1706–1727 (2014).
- Fiore, C. L., Jarett, J. K., Olson, N. D. & Lesser, M. P. Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol.* **18**, 455–463 (2010).
- Desai, M. S., Assig, K. & Dattagupta, S. Nitrogen fixation in distinct microbial niches within a chemoautotrophy-driven cave ecosystem. *ISME J.* **7**, 2411–2423 (2013).
- Dixon, R. & Kahn, D. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**, 621–631 (2004).
- Karl, D. *et al.* *Dinitrogen Fixation in the World's Oceans* (Springer Business and Science Media, 2002).

29. Cardini, U. *et al.* Functional significance of dinitrogen fixation in sustaining coral productivity under oligotrophic conditions. *Proc. R. Soc. B* **282**, 20152257 (2015).
30. Fry, B. *Stable Isotope Ecology* (Springer Science & Business Media, 2007).
31. Adams, T. S. & Sterner, R. W. The effect of dietary nitrogen content on trophic level <sup>15</sup>N enrichment. *Limnol. Oceanogr.* **45**, 601–607 (2000).
32. Doi, H. *et al.* Spatial shifts in food sources for macrozoobenthos in an estuarine ecosystem: carbon and nitrogen stable isotope analyses. *Estuar. Coast. Shelf Sci.* **64**, 316–322 (2005).
33. Salem, M. V. A., van der Geest, M., Piersma, T., Saoud, Y. & van Gils, J. A. Seasonal changes in mollusc abundance in a tropical intertidal ecosystem, Banc d'Arguin (Mauritania): testing the 'depletion by shorebirds' hypothesis. *Estuar. Coast. Shelf Sci.* **136**, 26–34 (2014).
34. van der Geest, M. *et al.* Nutritional and reproductive strategies in a chemosymbiotic bivalve living in a tropical intertidal seagrass bed. *Mar. Ecol. Prog. Ser.* **501**, 113–126 (2014).
35. Gallon, J. R. Reconciling the incompatible: N<sub>2</sub> fixation and O<sub>2</sub>. *New Phytol.* **122**, 571–609 (1992).
36. Sipe, A. R., Wilbur, A. E. & Cary, S. C. Bacterial symbiont transmission in the wood-boring shipworm *Bankia setacea* (Bivalvia: Teredinidae). *Appl. Environ. Microbiol.* **66**, 1685–1691 (2000).
37. Lechene, C. P., Luyten, Y., McMahon, G. & Distel, D. L. Quantitative imaging of nitrogen fixation by individual bacteria within animal cells. *Science* **317**, 1563–1566 (2007).
38. García-Horsman, J. A., Barquera, B., Rumbley, J., Ma, J. & Gennis, R. B. The superfamily of heme-copper respiratory oxidases. *J. Bacteriol.* **176**, 5587–5600 (1994).
39. Preisig, O., Zufferey, R., Thöny-Meyer, L., Appleby, C. A. & Hennecke, H. A high-affinity cbb3-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*. *J. Bacteriol.* **178**, 1532–1538 (1996).
40. Appleby, C. A. Leghemoglobin and *Rhizobium* respiration. *Annu. Rev. Plant Physiol.* **35**, 443–478 (1984).
41. Frenkiel, L., Gros, O. & Mouëza, M. Gill structure in *Lucina pectinata* (Bivalvia: Lucinidae) with reference to hemoglobin in bivalves with symbiotic sulphur-oxidizing bacteria. *Mar. Biol.* **125**, 511–524 (1996).
42. Kappler, U. & Schäfer, H. in *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment* Vol. 14 (eds Kroneck, P. M. H. & Torres, M. E. S.) 279–313 (Springer, 2014).
43. Ott, J. A. *et al.* Tackling the sulfide gradient—a novel strategy involving marine nematodes and chemoautotrophic ectosymbionts. *Mar. Ecol.* **12**, 261–279 (1991).
44. Wright, P. A. Nitrogen excretion: three end products, many physiological roles. *J. Exp. Biol.* **198**, 273–281 (1995).
45. Nelson, D. C. & Fisher, C. R. in *The Microbiology of Deep-Sea Hydrothermal Vents* (ed. Karl, D. M.) 125–167 (CRC, 1995).
46. Touchette, B. W. & Burkholder, J. M. Review of nitrogen and phosphorus metabolism in seagrasses. *J. Exp. Mar. Biol. Ecol.* **250**, 133–167 (2000).
47. Sorgo, A., Gaill, F., Lechaire, J.-P., Arndt, C. & Bright, M. Glycogen storage in the *Riftia pachyptila* trophosome: contribution of host and symbionts. *Mar. Ecol. Prog. Ser.* **231**, 115–120 (2002).
48. Reynolds, L. K., Berg, P. & Ziemann, J. C. Lucinid clam influence on the biogeochemistry of the seagrass *Thalassia testudinum* sediments. *Estuar. Coasts* **30**, 482–490 (2007).
49. Welsh, D. T. Nitrogen fixation in seagrass meadows: regulation, plant–bacteria interactions and significance to primary productivity. *Ecol. Lett.* **3**, 58–71 (2000).
50. Zhou, J. Z., Bruns, M. A. & Tiedje, J. M. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**, 316–322 (1996).
51. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
52. Strous, M., Kraft, B., Bisdorf, R. & Tegetmeyer, H. E. The binning of metagenomic contigs for microbial physiology of mixed cultures. *Front. Microbiol.* **3**, 410 (2012).
53. Seah, B. K. B. & Gruber-Vodicka, H. R. gbtools: interactive visualization of metagenome bins in R. *Front. Microbiol.* **6**, 1451 (2015).
54. 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).
55. Aziz, R. K. *et al.* The RAST server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
56. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
57. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
59. Hamann, E. *et al.* Environmental Breviatea harbour mutualistic *Arcobacter* epibionts. *Nature* **534**, 254–258 (2016).
60. Olsen, J. V. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol. Cell. Proteomics* **4**, 2010–2021 (2005).
61. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
62. Spivak, M., Weston, J., Bottou, L., Käll, L. & Noble, W. S. Improvements to the Percolator algorithm for peptide identification from shotgun proteomics data sets. *J. Proteome Res.* **8**, 3737–3745 (2009).
63. Florens, L. *et al.* Analyzing chromatin remodeling complexes using shotgun proteomics and normalized spectral abundance factors. *Methods* **40**, 303–311 (2006).
64. Lane, D. J. in *Nucleic Acid Techniques in Bacterial Systematics* (eds Stackebrandt, E. & Goodfellow, M.) 125–175 (Wiley, 1991).
65. Ando, S. *et al.* Detection of *nifH* sequences in sugarcane (*Saccharum officinarum* L.) and pineapple (*Ananas comosus* [L.] Merr.). *Soil Sci. Plant Nutr.* **51**, 303–308 (2005).
66. Wu, M. & Scott, A. J. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**, 1033–1034 (2012).
67. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
68. Wu, M., Chatterji, S. & Eisen, J. A. Accounting for alignment uncertainty in phylogenomics. *PLoS ONE* **7**, e30288 (2012).
69. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
70. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313. (2014).
71. Flood, B. E., Jones, D. S. & Bailey, J. V. Complete genome sequence of *Sedimenticola thiotaurini* strain SIP-G1, a polyphosphate- and polyhydroxyalkanoate-accumulating sulfur-oxidizing gammaproteobacterium isolated from salt marsh sediments. *Genome Announc.* **3**, e00671 (2015).
72. Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**, 1363–1371 (2004).
73. Heller, P., Tripp, H. J., Turk-Kubo, K. & Zehr, J. P. ARBitrator: a software pipeline for on-demand retrieval of auto-curated *nifH* sequences from GenBank. *Bioinformatics* **30**, 2883–2890 (2014).
74. Gaby, J. C. & Buckley, D. H. A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database (Oxford)* **2014**, bau001 (2014).
75. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
76. Abascal, F., Zardoya, R. & Telford, M. J. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res.* **38**, W7–W13 (2010).
77. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
78. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A. & Smith, G. M. *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).
79. nlme: Linear and Nonlinear Mixed Effects Models v. 3.1–120 (Pinheiro *et al.*, 2015); <https://cran.r-project.org/web/packages/nlme/index.html>
80. Vizcaino, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **32**, 223–226 (2014).
81. Walsh, D. A. *et al.* Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326**, 578–582 (2009).

## Acknowledgements

The authors thank J. Taylor from the Natural History Museum in London and N. Higgs from the University of Plymouth for providing samples. Thanks go to the Parc National du Banc d'Arguin (PNBA) for permission to work in the park and for use of the facilities at the Iwik field station. The authors thank A. Richter, W. Wanek and M. Watzka from the Division of Terrestrial Ecosystem Research at the University of Vienna and M. Smit and J. Ossebaar from NIOZ, for stable isotope analyses. The authors also thank S. Wetzel and M. Sadowski for technical assistance, M. Smit, A. Cado van der Lely and the team at the HYDRA Institute for support during fieldwork, N. Dubilier for supporting the project and M. 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. M.v.d.G. was supported by a post-doctoral fellowship of the Mediterranean Centre for Environment and Biodiversity (LabEx-CeMEB) and an NWO-WOTRO Integrated Programme grant (W.01.65.221.00) to Theunis Piersma. M.K. was supported by a NSERC Banting Postdoctoral Fellowship. This work was funded by the Max Planck Society, the Vienna Science and Technology Fund (WWTF) through project VRG14-021 to J.M.P. and a Fast Track grant from the Robert Bosch Foundation to J.M.P. Finally, the authors thank three anonymous reviewers for their input.

## Author contributions

J.M.P., A.K., H.G.-V., U.C., M.v.d.G., M.K., M.M., S.B., D.L., A.B. and M.W. designed and carried out experiments. J.M.P., A.K., H.G.-V., U.C., M.v.d.G., M.K., C.P.A., C.H. and B.K. B.S. analysed data. J.M.P. wrote the paper with contributions from all authors.

## Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to J.M.P.

**How to cite this article:** Petersen, J. M. *et al.* Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat. Microbiol.* **2**, 16195 (2016).

### Competing interests

The authors declare no competing financial interests.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

## Author Correction: Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation

Jillian M. Petersen, Anna Kemper, Harald Gruber-Vodicka, Ulisse Cardini, Matthijs van der Geest, Manuel Kleiner, Silvia Bulgheresi, Marc Mußmann, Craig Herbold, Brandon K. B. Seah, Chakkiath Paul Antony, Dan Liu, Alexandra Belitz and Miriam Weber

Correction to: *Nature Microbiology* <https://doi.org/10.1038/nmicrobiol.2016.195>, published online 24 October 2016.

In this Article, the completeness and number of contigs for draft genomes from two individuals of *Laxus oneistus* are incorrect in the main text, although the correct information is included in Table 1. The original and corrected versions of the relevant sentence are shown in the correction notice.

Original sentence:

Those from two individuals of *L. oneistus* (*Ca. Thiosymbion oneisti* A–B) were 86.75 and 89.11% complete at sizes of 3.66 and 3.51 Mb on 183 and 193 contigs (Table 1).

Corrected sentence:

Those from two individuals of *L. oneistus* (*Ca. Thiosymbion oneisti* A–B) were 96.63 and 96.07% complete at sizes of 4.44 and 4.33 Mb on 2,026 and 1,891 contigs (Table 1).

Published online: 27 June 2018

<https://doi.org/10.1038/s41564-018-0196-5>