Metabolically-versatile *Ca.* Thiodiazotropha symbionts of the deep-sea lucinid clam *Lucinoma kazani* have the genetic potential to fix nitrogen

Lina Ratinskaia[1,2], Stas Malavin[1,3], Tal Zvi-Kedem[1,2], Simina Vintila[4], Manuel Kleiner[4], Maxim Rubin-Blum[1,2]

[1]Biology Department, National Institute of Oceanography, Israel Oceanographic and Limnological Research (IOLR), Haifa, 3108000 Israel

[2]Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, 3498838 Israel

[3]Zuckerberg Institute for Water Research, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sde Boker Campus 8499000, Israel

[4]Department of Plant and Microbial Biology, North Carolina State University, Raleigh, NC, 27695, USA

Correspondence: Maxim Rubin-Blum – mrubin@ocean.org.il

Abstract

Lucinid clams are one of the most diverse and widespread symbiont-bearing animal groups in both shallow and deep-sea chemosynthetic habitats. Lucicnids harbor *Ca*. Thiodiazotropha symbionts that can oxidize inorganic and organic substrates such as hydrogen sulfide and formate to gain energy. The interplay between these key metabolic functions, nutrient uptake and biotic interactions in *Ca*. Thiodiazotropha is not fully understood. We collected *Lucinoma kazani* individuals from next to a deep-sea brine pool in the eastern Mediterranean Sea, at a depth of 1150 m and used Oxford Nanopore and Illumina sequencing to obtain high-quality genomes of their *Ca*. Thiodiazotropha gloverae symbiont. The genomes served as the basis for transcriptomic and proteomic analyses to characterize the *in situ* gene expression, metabolism and physiology of the symbionts. We found genes needed for N_2 fixation in the deep-sea symbiont's genome, which, to date, were only found in shallow-water *Ca*. Thiodiazotropha. However, we did not detect the expression of these genes and thus the potential role of nitrogen fixation in this symbiosis remains to be determined. We also found the high expression of carbon fixation and sulfur oxidation genes, which indicates chemolithoautotrophy as the key physiology of *Ca*. Thiodiazotropha. However, we also detected the expression of pathways for using methanol and formate as energy sources. Our findings highlight the key traits these microbes maintain to support the nutrition of their hosts and interact with them.

Introduction

Chemosynthesis, that is, the assimilation of inorganic carbon or methane into biomass using chemical energy rather than sunlight, is a key process driving productivity in some marine environments such as shallow water sediments and next to deep-sea vents and seeps [1]. Chemosynthetic bacteria and archaea established symbioses with an array of animal and protist hosts, allowing these organisms to thrive in marine habitats, such as hydrothermal vents, hydrocarbon seeps, deep organic falls and shallow-water seabed [2–4]. These associations differ in the taxonomic and functional diversity of the symbionts, the specificity of symbiont-host association, localization and transmission mechanisms [3, 5, 6].

Most symbionts are metabolic specialists, using a limited range of carbon and energy sources. Yet, some stand out exhibiting broader metabolic capacity. For example, the sulfur-oxidizing symbionts of bathymodioline mussels from hydrothermal vents can obtain energy from hydrogen [7] and can be mixotrophs, as they can use some organic carbon sources in addition to CO_2 [8]. The bona fide chemosynthetic symbionts often co-occur with other microbes that can use small organic compounds as energy and carbon sources. For example, *Cycloclasticus* use short-chain alkanes in *Bathymodiolus heckerae* and *Methylophaga* methylotrophs sustain growth using methanol, a byproduct of methane oxidation [9, 10].

One remarkable example of chemosynthetic symbioses are lucinid (Lucinidae) clams and their sulfur-oxidizing symbionts. Lucinidae is a species-rich clade, usually colonizing shallow-water sediments in the vicinity of seagrass meadows and coral reefs, but some species are found in deep-sea chemosynthetic habitats, including in the vicinity of vents and seeps [11]. Lucinids are characterized by their obligate association with intracellular chemosynthetic bacteria, most often with *Ca.* Thiodiazotropha (Chromatiales, Sedimaenticolaceae), and rarely with other gammaproteobacteria, such as Thiohalomonadales species [12, 13]. Similar to other Chromatiales symbionts, such as those of tubeworms [14–17], *Ca.* Thiodiazotropha fuels carbon fixation via the Calvin-Benson-Bassham (CBB) cycle primarily using the oxidation of sulfide, as well as other inorganic sulfur compounds. Yet, their energy sources expand beyond sulfur and may include the oxidation of hydrogen, methanol and formate using electron acceptors such as oxygen and nitrate [12, 13, 18, 19]. Several *Ca.* Thiodiazotropha genotypes are diazotrophs that can fix dinitriogen [20, 21], however, such diazotrophic symbionts were only described for shallow water and warm habitats [13].

While shallow-water lucinids are widely studied, the research of deep-sea lucinids is limited, given the difficulty of deep-sea sample collection. Here we focus on the symbionts of the deep-sea lucinid species *Lucinoma kazani*, endemic to the Mediterranean Sea [22, 23]. *L. kazani* is a close relative of Codakiinae lucinid *Lucinoma borealis* [11]. The latter was described in the North Atlantic at depths down to 165 m and consistently hosts *Ca*. T. gloverae symbionts [13, 24]. While shallow-water *Ca*. T. gloverae can fix dinitrogen, those found in *L. borealis* from the cold deep-sea waters lacked the genetic potential for diazotrophy. In contrast, *L. kazani* is found in the deep eastern Mediterranean Sea, where the water is generally warm (~14 °C) [25, 26]. Specifically, we found large populations of *L. kazani* near Palmahim Disturbance brine pools, which spew warm, ~ 22 °C brines [27]. Asking whether the unique conditions in the deep eastern Mediterranean Sea can lead to unique adaptations of *L. kazani* symbionts, we assembled high-quality metagenome-assembled genomes (MAGs) using both long (Oxford Nanopore) and short (Illumina) reads, and assessed their functionality using genome-centered transcriptomics and proteomics.

Materials and Methods

Sample collection: In April 2021, we collected four *L. kazani* specimens that thrived on the sediment surface in a brine pool area in Palmahim Disturbance offshore Israel at a water depth of approximately 1150 m (32° 13' 23.5" N 34° 10' 42.19" E), using a SAAB Seaeye Leopard remotely operated vehicle (ROV) 'Yona'. The specimens were dissected onboard upon retrieval. Approximately 3 hours passed between scoop collection and processing (the scoop with animals was placed in ROV's biobox), and gill tissue for consecutive DNA/RNA/protein extraction was preserved in RNAlater (4° for 24 hours, then RNAlater was decanted and specimens were kept at -80 °C). We recently showed that RNAlater is a good preservative for proteomic sample preservation in addition to the preservation of RNA/DNA [28].

Stable isotope analyses: Tissue samples from five frozen clam specimens were lyophilized for 24 hours, homogenized and weighed before stable isotope analysis at Cornell University Stable Isotope Laboratory. The isotopic composition of organic carbon (n=5), nitrogen (n=5) and sulfur (n=4) was determined using a Thermo Delta V isotope ratio mass spectrometer (IRMS) interfaced to a NC2500 elemental analyzer (Sisma-Ventura et al., 2022). The isotopic composition of each sample was expressed as the relative difference between isotopic ratios in the sample and that in conventional standards (Vienna Pee Dee Belemnite, atmospheric N_2 and Canyon Diablo Troilite

for carbon, nitrogen and sulfur, respectively). Measured isotope ratios are reported in the δ per mille (‰) notation representing the deviation from the standards:

$$\delta_{Sample} = (\ R_{Sample}\ /R_{standard} - 1\ ) \times 10^3$$

where, R = $^{15}N/^{14}N$, $^{13}C/^{12}C$ or $^{34}S/^{32}S$ ratios.

Preparation for omics: DNA, RNA and proteins were extracted from the four individuals using the AllPrep DNA/RNA/Protein Mini Kit (Cat. No. 80004, Qiagen). DNA libraries were constructed for these six individuals at HyLabs, Israel, and sequenced using 120 million 2x150 bp paired-end reads per sample with an Illumina NovaSeq at GENEWIZ, Germany. RNA libraries were constructed for these individuals at Novogene, Singapore, and sequenced using 100 million 2x150 bp paired-end reads per sample with an Illumina NovaSeq, following library construction with the NEBNext Ultra RNA Library Prep Kit for Illumina (Cat No. 7530) and ribosomal RNA depletion with the Ribo-Zero Plus rRNA Depletion Kit (Bacteria) (Cat No. 20037135) & Ribo-Zero Magnetic Kit (Plant Leaf). For long-read sequencing, DNA was extracted from an additional individual using DNeasy Blood & Tissue Kit (Qiagen). Long reads were produced by Minion (Oxford Nanopore), using R10.4.1 flow cells and SQK-NBD 114.24 Native Barcoding Kit 24 V14. Base calls from the generated raw Nanopore data were done using the super-accurate mode in the ONT Guppy base calling software version 6.4.2+97a7f06 with the dna_r10.4_e8.1_sup.cfg model.

Bioinformatics: Host mitochondrial genomes were assembled and annotated with MitoZ [29] and the 16S and 18S rRNA gene sequences were extracted using phyloFlash v3.4.2 [30]. We carried out adapter trimming and error correction with tadpole.sh, using the BBtools suite following read preparation with the BBtools suite (Bushnell, B, sourceforge.net/projects/bbmap/). Metagenomes were assembled from short reads using SPAdes V3.15 with –meta k=21,33,66,99,127 parameters [31]. We performed downstream mapping and binning of metagenome-assembled genomes (MAGs) using DAStool, Vamb, Maxbin 2.0 and Metabat2 [32–35] within the Atlas V2.9.1 framework [36]. MAG quality was assessed using Checkm2 [37] and QUAST [38]. This analysis yielded a single high-quality, dereplicated symbiont MAG (99.98% completeness and 0.47% contamination, N50=57,715, 110 contigs). Long Naonopre reads (N50=1,614, longest read of 86,787 bases) were aligned to this genome using Minimap2 [39], and a hybrid genome was constructed from these aligned long reads and short reads, using the 'isolate' module in Nanophase 0.2.2 [40]. For functional annotation, we used the Rapid Annotations using the Subsystems Technology (RAST) server [41], and verified key annotations by BLAST search against the NCBI

and UniProt databases. Prophage regions were identified with Phaster [42] and phage defense systems with DefenseFinder v1.2.0 [43].

RNA reads were quality-trimmed and mapped to the MAGs using BBmap with a 0.99 identity threshold (Bushnell, B, sourceforge.net/projects/bbmap/), and read counts were assigned to coding sequences using FeatureCounts [44]. The counts were normalized as transcripts per million (TPM) [45]. Phylogenomics was performed using the Codon Tree method in BV-BRC [46], which selects single-copy BV-BRC PGFams and analyzes single-copy genes using RAxML [47]. For selected features, genes/proteins were aligned with MAFFT [48], and treeing was performed with FastTree [49] and IQtree2 [50], using the best model.

Protein extraction and peptide preparation for metaproteomics: We resuspended the protein precipitates from the four individuals extracted with the AllPrep DNA/RNA/Protein Mini Kit kit in 60 µl of SDT lysis buffer [4% (w/v) SDS, 100 mM Tris-HCl pH 7.6, 0.1 M DTT] and heated to 95°C for 10 min. The SDT protein mixture was cleaned up, reduced, alkylated and digested using the filter-aided sample preparation (FASP) protocol as described previously [51]. We performed all centrifugation steps mentioned below at 14,000 x g. We combined lysates (60 µl) with 400 µl of urea solution (8 M urea in 0.1 M Tris-HCl pH 8.5) and loaded it onto 10 kDa MWCO 500 µl centrifugal filters (VWR International) followed by centrifugation for 20 min. We washed filters once by applying 200 µl of urea solution followed by 20 min of centrifugation to remove any remaining SDS. We performed protein alkylation by adding 100 µl IAA solution (0.05 M iodoacetamide in urea solution) to each filter and incubating for 20 min at room temperature followed by centrifugation for 20 min. The filters were washed three times with 100 µL of urea solution with 15 min centrifugations, followed by a buffer exchange to ABC (50 mM ammonium bicarbonate). Buffer exchange was accomplished by three cycles of adding 100 µl of ABC buffer and centrifuging for 15 min. For tryptic digestion, we added 1 µg of MS grade trypsin (ThermoFisher Scientific) in 40 µl of ABC buffer to each filter and incubated for 16 hours in a wet chamber at 37°C. We eluted tryptic peptides by adding 50 µl 0.5 M NaCl and centrifuging for 20 min. Peptide concentrations were determined with the Pierce Micro BCA assay (ThermoFisher Scientific) following the manufacturer's instructions.

LC-MS/MS: All proteomic samples were analyzed by 1D-LC-MS/MS as previously described [51] We loaded 1.2 µg of peptide from each sample onto a 5 mm, 300 µm ID C18 Acclaim®

PepMap100 pre-column (Thermo Fisher Scientific) with an UltiMateTM 3000 RSLCnano Liquid Chromatograph (Thermo Fisher Scientific) in loading solvent A (2% acetonitrile, 0.05% trifluoroacetic acid). Elution and separation of peptides on the analytical column (75 cm x 75 μm EASY-Spray column packed with PepMap RSLC C18, 2 μm material, Thermo Fisher Scientific; heated to 60°C) was achieved at a flow rate of 300 nl min^{-1} using a 140 min gradient going from 95% buffer A (0.1% formic acid) and 5% buffer B (0.1% formic acid, 80% acetonitrile) to 31% buffer B in 102 min, then to 50% B in 18 min, and finally to 99% B in 1 min and ending with 99% B. The analytical column was connected to a Q Exactive HF hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) via an Easy-Spray source. Eluting peptides were ionized via electrospray ionization (ESI). Carryover was reduced by a wash run (injection of 20 μl acetonitrile, 99% eluent B) between samples. MS1 spectra were acquired by performing a full MS scan at a resolution of 60,000 on a 380 to 1600 m/z window. MS2 spectra were acquired using data-dependent acquisition, selecting for fragmentation the 15 most abundant peptide ions (Top15) from the precursor MS1 spectra. A normalized collision energy of 25 was applied in the HCD cell to generate the peptide fragments for MS2 spectra. Other settings of the data-dependent acquisition included: a maximum injection time of 100 ms, a dynamic exclusion of 25 sec, and the exclusion of ions of +1 charge state from fragmentation. About 90,000 MS/MS spectra were acquired per sample.

Protein identification and quantification: We constructed a protein sequence database for protein identification using the protein sequences predicted from the metagenome-assembled genomes obtained in this study. To identify peptides from the host, we used the annotated protein sequences of *Loripes orbiculatus*[52]. We added sequences of common laboratory contaminants by appending the cRAP protein sequence database (http://www.thegpm.org/crap/). The final database contained 101,483 protein sequences and is included in the PRIDE submission (see data access statement) in fasta format. Searches of the MS/MS spectra against this database were performed with the Sequest HT node in Proteome Discoverer 2.3.0.523 as previously described [53]. The peptide false discovery rate (FDR) was calculated using the Percolator node in Proteome Discoverer and only peptides identified at a 5% FDR were retained for protein identification. Proteins were inferred from peptide identifications using the Protein-FDR Validator node in Proteome Discoverer with a target FDR of 5%. To estimate species abundances based on proteinaceous biomass using the metaproteomic data we followed the previously described

approach [54] with the added filter criterion of requiring two protein-unique peptides for a protein to be retained for biomass calculations.

Results and Discussion

Palmahim brine pool perimeter is inhabited by Lucinoma kazani

Clams formed a large, densely populated patch (several hundred square meters) near Palmahim brine pools (Figure 1). The clam shells were often covered with egg cases of *Galeus melastomus* sharks, which lay eggs in seeps likely due to the elevated temperatures that favor embryo development [55]. Thus, we assume that the temperature in shark/clam habitat is higher than the ambient 13.7 °C. In each scoop from this habitat, we found several live specimens among multiple empty shells.

Marker gene analysis indicated that the hosts were *Lucinoma kazani.* The mitochondrial *cytB* sequence was 99.15% similar to that of *L. kazani* (KF741674.1, the best hit in the NCBI). The 18S rRNA gene doesn't distinguish between *L. kazani* and *L. borealis*, as the full-length nuclear 18S rRNA gene sequences in the four collected individuals were identical to that of *L. borealis* (AM774501.1, no *L. kazani* 18S rRNA gene sequences are available in the NCBI database).

We assembled a single symbiont 16S rRNA gene sequence from each metagenome using reads that mapped to the Silva138.1 16S rRNA gene sequence database. These sequences were identical among the four collected individuals, and 99.5% similar to those of the best hit in NCBI, *L. kazani* symbiont sequence AM236336.1. Phylogenomics showed that the *Ca.* Thiodiazotropha gloverae symbiont of *L. borealis* (GenBank: JAHXIE000000000.1, from Devon, United Kingdom) is the closest relative of *L. kazani* symbionts with 96.55% average nucleotide identity (**Supplementary Figure 1**). The symbionts were abundant in the gills of all four individuals (73-83% of peptide spectral matches).

Stable isotope analyses confirmed the nutritional dependence of *L. kazani* on the chemosynthetic symbionts. Bulk $\delta^{13}C$=-30.8±0.8‰ and $\delta^{15}N$=0.9±1.9‰ (n=5, **Supplementary Table ST1**) values were similar to $\delta^{13}C$=−29.8‰±0.9‰ and $\delta^{15}N$=−2.5‰±0.6‰ (n=3) identified in the symbiont-bearing gills of *L. capensis* from the Namibian Shelf [56]. The negative $\delta^{34}S$ values between −18.9‰ and -1.6‰ indicated that chemosynthesis is driven by sulfur oxidation [57].

Hybrid assembly improved the contiguity of the symbiont genome and identified large adhesin-like proteins

The hybrid assembly resulted in a more contiguous ~4.9 Mb-long MAG consisting of 44 contigs with an N50=286,621 bp and the largest contig of 532,579 bp (**Table 1**). CheckM2 estimated completeness and contamination at 98.9% and 0.4%, respectively. Most importantly, the long reads were able to resolve long sequences (13,449 and 10,077 bp in length) encoding several FhaB (filamentous hemagglutinin, COG3210) domains of adhesins (**Supplementary Figure S2**). These proteins are usually secreted to outer bacterial membranes and serve as attachment factors to host cells[58], playing a role in pathogenicity [59] and symbioses [60]. *L. kazani* symbionts indeed may secrete these adhesin-like proteins, as we identified several genes encoding secretion-related features in the neighborhood of the two adhesin-like protein-encoding sequences (e.g, those encoding proteins with HylD/AcrA-like domains, two-component system response regulator and sensor histidine kinase, type I secretion system permease/ATPase, **Supplementary Figure S2**). Apart from FhaB, the putative adhesins comprised additional eukaryote-like domains, often resembling RTX/MARTX toxins [61, 62], which may be involved in interactions of symbionts and their hosts [63, 64], facilitating not only adhesion but also cell surface cleavage [65]. These domains comprised the toxin-related Ca_2^+-binding protein (COG2931) and choice-of-anchor K domains, cadherin-like and bacterial Ig domains, as well as M36-family metallopeptidases, among others (**Supplementary Figure S2**). BLASTing against the NCBI database identified proteins with highly similar domains among *Ca.* Thiodiazotropha symbionts of various lucinids. Examples include the 5907 aa protein in *Ca.* Thiodiazotropha taylori (MBT3016762.1) and a 5144 aa protein in *Ca*. Thiodiazotropha ex. *Troendleina suluensis* (MCU7840510.1). As full-length sequences are often missing in MAGs produced by the short-read assemblies, the overview of adhesin diversity in *Ca*. Thiodiazotropha is incomplete. Similar adhesin-like proteins were highly expressed by symbionts of *Phacoides pectinatus* lucinids [66]. We, however, observed a limited expression of these sequences at the RNA level (no proteins found), hinting that the respective genes may be downregulated in established symbioses, or expressed by a small fraction of the symbiont population. Given the key role of adhesins in interactions between metazoans and bacteria, as well as their reoccurrence and expression in *Ca*. Thiodiazotropha, we hypothesize that these large proteins play a role in host colonization.
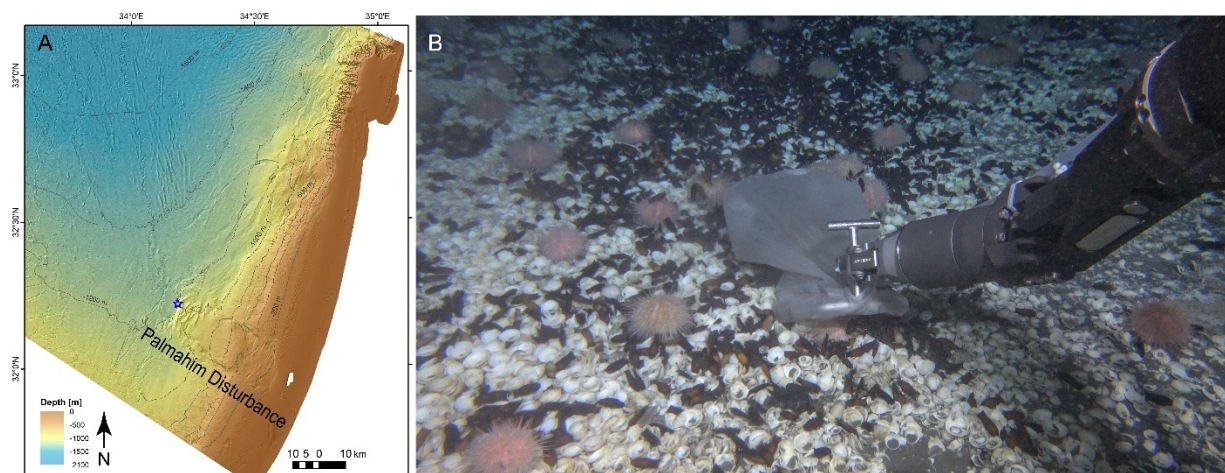
Figure 1: *Lucinoma kazani* inhabit the perimeter of deep-sea brine pools offshore Israel. A) The collection site location at the toe of Palmahim disturbance. B) Scoop collection of *L. kazani*.

Table 1: Assembly statistics for metagenome-assembled genomes of *Lucinoma kazani* symbionts.

	Short-read assembly	*Hybrid assembly*
Length (bp)	4,346,866	4,873,710
Contigs	110	44
N50 (bp)	57,715	286,621
Longest contig (bp)	202,093	532,579
Completeness (%)	99.98	98.81
Contamination (%)	0.47	0.44

Ca. T. gloverae ex L. kazani are metabolically versatile

The core, most well-expressed metabolic traits in *Ca.* T. gloveare ex. *L. kazani* included those involved in carbon fixation via the Calvin-Benson-Bassham (CBB) cycle and sulfur oxidation via the reverse dissimilatory sulfite reduction (DSR) pathway (**Figure 2, Supplementary Table ST2**). Similar to other Chromatiales symbionts [67, 68], *Ca.* T. gloverae ex. *L. kazani* uses the energy-efficient CBB pathway, driven by an H^+-PPase and a closely coupled PPi-dependent 6-phosphofructokinase (PPi-PFK). The *pfP* and *hppA* genes that encode these features are neighbors that appear to be co-expressed and have high expression values (**Figure 2, Supplementary Table ST2**). The symbiont encoded two forms of ribulose bisphosphate carboxylase/oxygenase (RubisCO), which differ in their affinity to oxygen and CO_2: form I has have high specificity factor, that is, functions better under low CO_2 and high O_2 conditions and form II enzymes, with low specificity factor [69–72]. This may reflect an adaptation to fluctuations in CO_2 and O_2

availability within the host's bacteriocyte. Form I RubisCO was highly expressed compared to form II (11-79 more RNA reads mapped, and 8-14 times more protein spectra), suggesting that either oxygen was not limiting, or CO_2 starvation. Other parts of the CBB pathway were also highly expressed. The CBB pathway is linked to glycogen storage, based on the substantial expression of the *glgABC* gene cluster. The symbiont is capable of gaining energy from the breakdown of these stored compounds, encoding and expressing a complete tricarboxylic acid cycle. We also found that several TRAP transporters were expressed, hinting at uptake of some organics, and thus mixotrophy.

The key enzymes involved in dissimilatory sulfide oxidation, including the adenylyl-sulfate (APS) reductase (AprAB), the sirohaem dissimilatory sulfite reductase (DSR) complex and sulfate adenylyltransferase (Sat), were detected as abundantly expressed at both the RNA and protein levels. Apart from Sat, the symbiont encoded the membrane-bound sulfite dehydrogenase (SoeABC). These two enzymes appear to synergistically catalyze the oxidation of sulfite in the purple sulfur bacterium *Allochromatium vinosum* [73]. Our expression analyses highlighted the important role of sulfur globule storage [74], as three types of sulfur globule proteins were highly expressed. In particular, SgpB was among the most well-expressed genes and most abundant proteins. The expression of the Sox system (SoxBXA and SoxYZ), needed to convert thiosulfate to sulfur globules, as well as the flavocytochrome c sulfide dehydrogenase FccAB, oxidizing sulfide to elemental sulfur [75], was substantial. We found three forms of sulfide: quinone reductases (Sqr1, 3 and 6), which differ in their affinity to sulfide [76], indicating the adaptability of the symbiont to unstable substrate concentrations. In summary, *Ca.* T. gloveare ex. *L. kazani* has a flexible rDSR pathway, typical of Chromatiales symbionts, such as those of siboglinid tubeworms [3].

Lucinoma symbionts use oxygen as a terminal electron acceptor but can reduce nitrate in the presence of oxygen [77]. In line with this observation, the expression of the CtaCDE terminal oxidase was high, yet the complete denitrification pathway was encoded and expressed. The symbionts may catalyze nitrate reduction using both low-affinity dissimilatory Nar and high-affinity periplasmic Nap nitrate reductases. Nap was expressed higher than Nar (also markedly higher at the protein level), potentially indicating nitrate limitation of denitrification [78, 79].

Unlike other *Ca.* T. gloverae [13], *L. kazani* symbiont encoded the canonical pathway of nitrate assimilation via Nas and NirBD. Yet, we found that the expression of assimilatory nitrate and

nitrite reduction, as well as that of ammonium uptake, was generally low. In turn, both glutamine synthetase (GlnA) and glutamate synthase (GltBD) were moderately expressed. The very low expression of P-II nitrogen regulatory protein (GlnK) may indicate that ammonium is not limiting *Ca.* T. gloverae ex. *L. kazani* growth [80], which is likely given the millimolar levels of ammonium in Palmahim brines [27]. Additional ammonium can be derived from cyanate and thiocyanate, given that both cyanate hydratase and thiocyanate hydrolase were encoded and moderately transcribed by the *L. kazani* symbiont (**Figure 2, Supplementary Table ST2**).
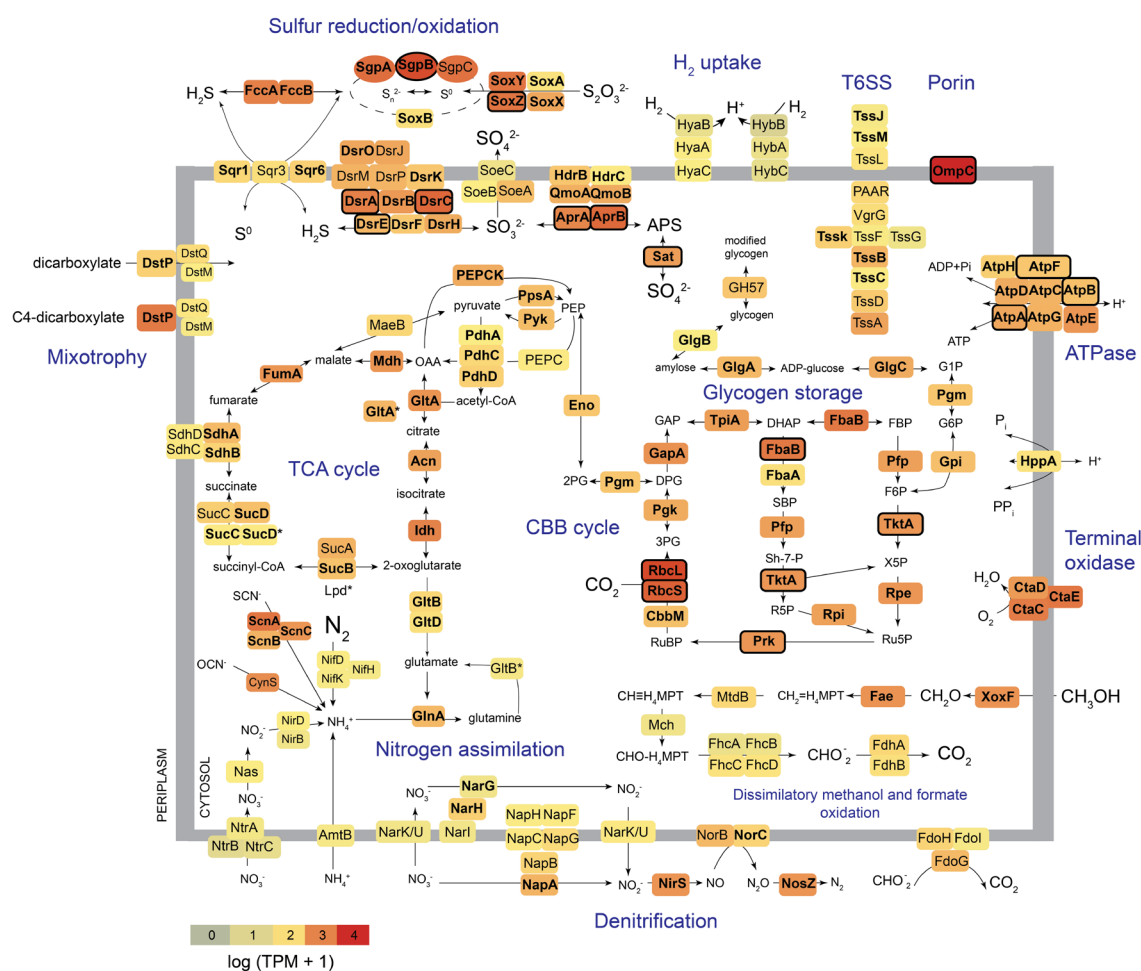

Figure 2: The key metabolic features in *Ca.* Thiodizaotropha gloverae ex. *Lucinoma kazani*. Color code is according to transcript abundance. Features detected at the protein level are marked in bold font. Those that constitute the top 50 most abundant proteins are framed. See Supplementary Table ST3 for protein annotations. TCA (tricarboxylic acid, Krebs) and CBB (Calvin-Benson-Bassham) cycles are abbreviated. * more than one gene copy present. ** The bona fide Lpd subunit of the 2-oxoglutarate dehydrogenase was not found, yet several variants of pyruvate dehydrogenase-associated lipoamide dehydrogenase were present, and may potentially complement the missing subunit.
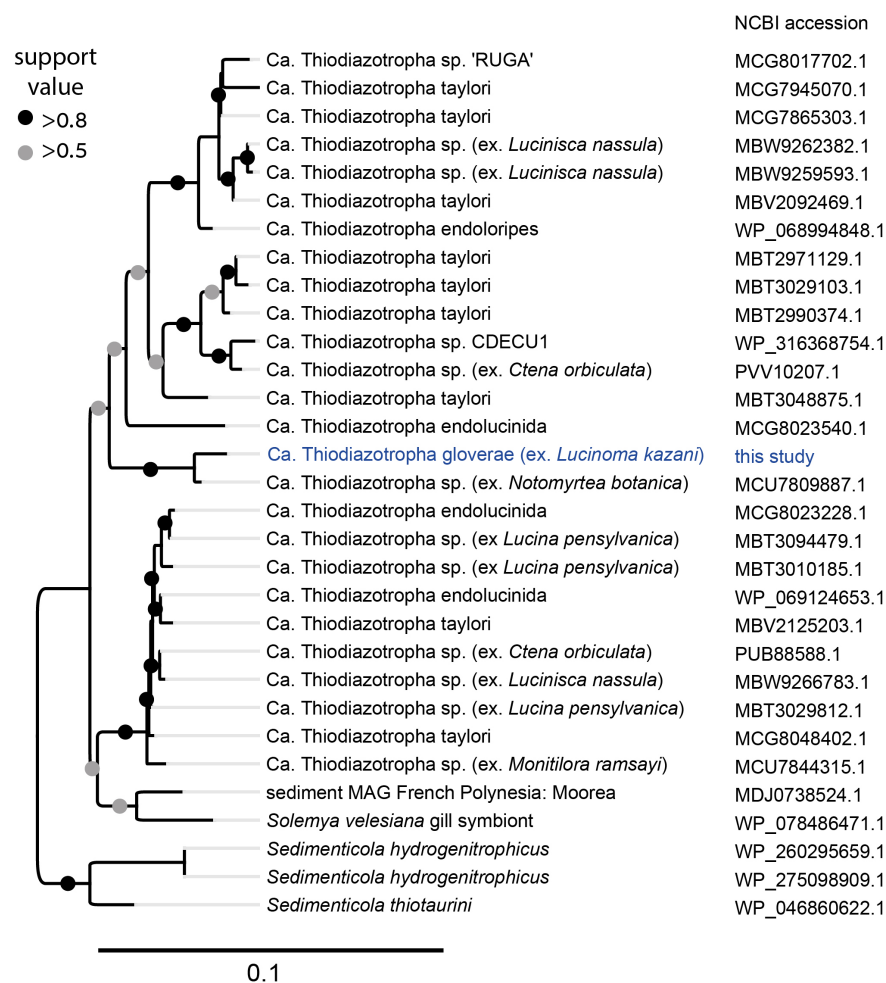
Figure 3: Phylogeny of 31 NifD amino acid sequences in the Thiodiazotropha clade. The maximum likelihood tree was constructed in FastTree, using the LG model. The tree was force-rooted at the *Sedimenticola* genus branch. The tree scale represents the number of substitutions per site. See Supplementary Figure S3 for an extended nucleotide tree.

While N_2 fixation was attributed to date only to the symbionts of lucinids from shallow habitats [13], we found a complete cluster of genes involved in nitrogen fixation, in particular, those that encode the core subunits of molybdenum-iron nitrogenase (NifDK) and nitrogenase reductase and maturation protein NifH. The phylogeny of these genes is not congruent with *Ca.* Thiodiazotropha phylogenomics (**Figure 3, Supplementary Figures 1 and 3**). Most strikingly, the best BLAST hits of the key *nifD* (98.4% identity) and *nifK* (96.4% identity) were most similar to the genes from the distant clade hosted by *Notomyrtea botanica* (Notomyrt1 species). Hits to *nifD* genes from other *Ca.* Thiodiazotropha were below 95 and 94%, respectively (**Figure 3**). This hints at an evolutionary history of the nitrogenase cluster in this clade, with widespread gene losses, and

potential secondary gains, which, however, are constrained within *Sedimenticolaceae*. Given that nitrogen fixation in lucinid symbionts is linked to shallow and warm habitats[13], temperature may play a role in the selection of this trait. Lucinid habitat in the deep easter Mediterranean Sea is characterized by aberrantly warm temperatures of ~14 °C of the deep water column, however, the actual temperature may be even higher, given the inflow of ~22 °C-warm brines. This may have played a role in retaining diazotrophy potential. Yet, we detected minimal expression values for this cluster at the RNA level, and no protein-level expression was found. Thus, nitrogen fixation may be limited in the host-associated *Thiodiazotropha*, for which other not as energetically-costly nitrogen sources such as ammonium and nitrate are available.

Following previous studies [12, 13, 18, 19], our data suggests that the use of methanol, formate and hydrogen may support the energy metabolism in *Ca.* T. gloverae ex. *L. kazani*. The symbionts can conserve energy by oxidizing methanol to formaldehyde, formate and CO_2, based on a gene cluster that encodes the lanthanide-dependent methanol dehydrogenase (XoxF), formaldehyde detoxification via the H_4MPT pathway and the tungsten-containing formate dehydrogenase. Methanol appears to be a source of energy, but not carbon, as the key genes in formaldehyde assimilation pathways were not found. These include genes encoding the key enzymes of ribulose monophosphate pathway 3-hexulose-6-phosphate synthase (*rmpA*) and 3-hexulose-6-phosphate isomerase (*rmpB*), as well as in serine pathway, such *sgaA*, encoding serine--glyoxylate aminotransferase. The symbiont also encoded and expressed the respiratory membrane-bound formate dehydrogenase FdoGHI, which may allow the conservation of energy under anaerobic conditions using nitrate as an electron acceptor [81].

We identified genomic clues for the use of dihydrogen by *L. kazani* symbionts. Similar to their relative *Sedimenticola hydrogenitrophicus* that can grow hydrogenotrophically [82], the symbiont genome contained a cluster of genes encoding the HyaABC subunits of the H_2-uptake [NiFe] hydrogenase, the accessory proteins, and hydrogen-sensing hydrogenase HoxCB. The *hyaA* and *hyaB* genes were present in two copies. The expression of these genes was low, indicating that similar to diazotrophy, host-associated symbionts were likely not using hydrogen under the *in situ* conditions at the time of sampling. Moreover, these hydrogenases can be directly linked to dinitrogen fixation, supplying electrons to sustain this process, as described in the purple bacterium *Rhodopseudomonas palustris* [83].

Potential features responsible for biotic interactions of Ca. T. gloverae ex. L. kazani

We observed traits that can mediate the interaction between *Ca.* T. gloverae ex. *L. kazani* with the host, or other bacteria/competing strains. Similar to the symbionts of tubeworms, the most well-expressed gene/protein was an OmpC superfamily porin [16, 84]. These prior studies suggested that such porins may be crucial for maintaining stable symbiotic associations. Together with adhesin-like proteins, these porins may play a role in maintaining the symbioses, yet their role has not been experimentally tested so far.

The symbionts encoded and expressed components of the type VI secretion system (T6SS), which facilitates virulence, host immunomodulation and bactericidal activity in many bacterial lineages [85–87], and were previously found to be expressed by lucinid symbionts [18, 19]. These contact-dependent interbacterial "weapons" [88] are most likely to be involved in interactions with competing symbiont genotypes or cheaters during colonization of the host [18, 84, 89]. For example, in bobtail squid symbionts *Vibrio fischeri*, T6SS was shown to facilitate strain incompatibility, eliminating closely related competitor strains [89, 90].

We identified only a few phage-defense systems in the genome, while no CRISPR-Cas systems were found. These defense systems included LmuB SMC Cap4 nuclease II and LmuA effector hydrolase from a Lamassu family [91], and very few restriction-modification and toxin-antitoxin systems (e.g., sanaTA, [92]), often integrated into poorly-expressed regions identified as partial inactive prophages (**Supplementary Table ST2**). This suggests a limited phage predation pressure on the *L. kazani* symbionts.

Conclusions

Lucinoma kazani, and their *Ca.* Thiodizotropha gloverae symbionts thrive in the warm deep-sea chemosynthetic habitat near the Mediterranean Sea brine pools. These conditions may determine their metabolic adaptations, for example, retaining the ability to fix dinitrogen. These deep-sea clam symbionts appear to maintain a versatile metabolism, modulating it via gene expression. For example, the preferential expression of cyanate/thiocyanate module over ammonium uptake, or changes in the expression of Sqr proteins with different affinities to sulfide. Further exploration of expression patterns under different environmental scenarios is needed to shed light on the interplay between different physiological strategies. Moreover, the heterogeneity of expression at a single symbiont level remains to be explored.

These genomes encode and express several features that can be involved in biotic interactions, including porins, secretions systems and adhesin-like proteins. We speculate that this versatility requires the maintenance of large genomes of ~5 Mb, which is larger than those previously suggested for lucinid symbionts from extreme environments, such as ~3 Mb genomes of *Ca*. T. "Aeq1" clade [13]. The discovery of key features such as toxin-like proteins in these large genomes was possible due to hybrid assemblies based on long-read sequencing.

Data availability

The raw metagenomic reads, as well as the final MAGs, were submitted to NCBI in PRJNA1078112. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [93] partner repository with the dataset identifier PXD051186.

Author contributions

MR-B, LR and MK conceived this study and acquired funding. LR, TZ-K and SM performed DNA/RNA/protein extraction and data analyses. SV and MK conducted proteomics. LR, MR-B, and MK wrote the paper with the contributions of all co-authors.

Acknowledgments

Funding

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Bibliography

1. Ricci F, Greening C. Chemosynthesis: a neglected foundation of marine ecology and biogeochemistry. *Trends Microbiol* 2023; 2235.

2. Dubilier N, Bergin C, Lott C. Symbiotic diversity in marine animals: the art of harnessing chemosynthesis. *Nat Rev Microbiol* 2008; **6**: 725–740.

3. Sogin EM, Kleiner M, Borowski C, Gruber-Vodicka HR, Dubilier N. Life in the dark: phylogenetic and physiological diversity of chemosynthetic symbioses. *Annu Rev Microbiol* 2021; **75**: 695–718.

4. Haber M, Burgsdorf I, Handley KM, Rubin-Blum M, Steindler L. Genomic insights into the lifestyles of Thaumarchaeota inside sponges. *Front Microbiol* 2021; 3441.

5. Russell SL, Corbett-Detig RB, Cavanaugh CM. Mixed transmission modes and dynamic genome evolution in an obligate animal–bacterial symbiosis. *ISME J* 2017; **11**: 1359–1371.

6. Sato Y, Wippler J, Wentrup C, Ansorge R, Sadowski M, Gruber-Vodicka H, et al. Fidelity varies in the symbiosis between a gutless marine worm and its microbial consortium. *Microbiome* 2022; **10**: 178.

7. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 2011; **476**: 176–80.

8. Ponnudurai R, Kleiner M, Sayavedra L, Petersen JM, Moche M, Otto A, et al. Metabolic and physiological interdependencies in the *Bathymodiolus azoricus* symbiosis. *ISME J* 2017; **11**: 463–477.

9. Rubin-Blum M, Antony CP, Borowski C, Sayavedra L, Pape T, Sahling H, et al. Short-chain alkanes fuel mussel and sponge Cycloclasticus symbionts from deep-sea gas and oil seeps. *Nat Microbiol* 2017; **2**: 17093.

10. Zvi-Kedem T, Vintila S, Kleiner M, Tchernov D, Rubin-Blum M. Metabolic handoffs between multiple symbionts may benefit the deep-sea bathymodioline mussels. *ISME Commun* 2023; **3**: 48.

11. Taylor JD, Glover EA, Smith L, Ikebe C, Williams ST. New molecular phylogeny of Lucinidae: Increased taxon base with focus on tropical Western Atlantic species (Mollusca: Bivalvia). *Zootaxa* 2016; **4196**: 381–398.

12. Osvatic JT, Wilkins LGE, Leibrecht L, Leray M, Zauner S, Polzin J, et al. Global

biogeography of chemosynthetic symbionts reveals both localized and globally distributed symbiont groups. *Proc Natl Acad Sci* 2021; **118**: e2104378118.

13. Osvatic JT, Yuen B, Kunert M, Wilkins L, Hausmann B, Girguis P, et al. Gene loss and symbiont switching during adaptation to the deep sea in a globally distributed symbiosis. *ISME J* 2023; **17**: 453–466.

14. Li Y, Liles MR, Halanych KM. Endosymbiont genomes yield clues of tubeworm success. *ISME J* 2018; **12**: 2785–2795.

15. Cavanaugh CM. Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. *Nature* 1983; **302**: 58–61.

16. Hinzke T, Kleiner M, Meister M, Schlüter R, Hentschker C, Pané-Farré J, et al. Bacterial symbiont subpopulations have different roles in a deep-sea symbiosis. *Elife* 2021; **10**: e58371.

17. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, et al. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science (80-)* 2007; **315**: 247–250.

18. Lim SJ, Alexander L, Engel AS, Paterson AT, Anderson LC, Campbell BJ. Extensive thioautotrophic gill endosymbiont diversity within a single *Ctena orbiculata* (bivalvia: lucinidae) population and implications for defining host-symbiont specificity and species recognition. *mSystems* 2019; **4**: 1–19.

19. Lim SJ, Davis B, Gill D, Swetenburg J, Anderson LC, Engel AS, et al. Gill microbiome structure and function in the chemosymbiotic coastal lucinid *Stewartia floridana*. *FEMS Microbiol Ecol* 2021; **97**: 1–15.

20. König S, Gros O, Heiden SE, Hinzke T, Thürmer A, Poehlein A, et al. Nitrogen fixation in a chemoautotrophic lucinid symbiosis. *Nat Microbiol* 2016; **2**: 1–10.

21. Petersen JM, Kemper A, Gruber-Vodicka H, Cardini U, Van Der Geest M, Kleiner M, et al. Chemosynthetic symbionts of marine invertebrate animals are capable of nitrogen fixation. *Nat Microbiol* 2016; **2**: 1–12.

22. Duperron S, Fiala-Médioni A, Caprais JC, Olu K, Sibuet M. Evidence for chemoautotrophic symbiosis in a Mediterranean cold seep clam (Bivalvia: Lucinidae): comparative sequence analysis of bacterial 16S rRNA, APS reductase and RubisCO genes. *FEMS Microbiol Ecol* 2007; **59**: 64–70.

23.   Salas C, Woodside J. Lucinoma kazani n. sp. (Mollusca: Bivalvia): Evidence of a living benthic community associated with a cold seep in the Eastern Mediterranean Sea. *Deep Res Part I Oceanogr Res Pap* 2002; **49**: 991–1005.

24.   Dando PR, Southward AJ, Southward EC. Chemoautotrophic symbionts in the gills of the bivalve mollusc Lucinoma borealis and the sediment chemistry of its habitat. *Proc - R Soc London, Ser B* 1986; **227**: 227–247.

25.   Basso D, Beccari V, Almogi-Labin A, Hyams-Kaphzan O, Weissman A, Makovsky Y, et al. Macro- and micro-fauna from cold seeps in the Palmahim Disturbance (Israeli offshore), with description of *Waisiuconcha corsellii* n.sp. (Bivalvia, Vesicomyidae). *Deep Res Part II Top Stud Oceanogr* 2020; **171**: 104723.

26.   Beccari V, Basso D, Spezzaferri S, Rüggeberg A, Neuman A, Makovsky Y. Preliminary video-spatial analysis of cold seep bivalve beds at the base of the continental slope of Israel (Palmahim Disturbance). *Deep Res Part II Top Stud Oceanogr* 2020; **171**: 104664.

27.   Herut B, Rubin-Blum M, Sisma-Ventura G, Jacobson Y, Bialik OM, Ozer T, et al. Discovery and chemical composition of the eastmost deep-sea anoxic brine pools in the Eastern Mediterranean Sea. *Front Mar Sci* 2022; **9**: 1040681.

28.   Jensen M, Wippler J, Kleiner M. Evaluation of RNA later as a field-compatible preservation method for metaproteomic analyses of bacterium-animal symbioses. *Microbiol Spectr* 2021; **9**: e01429-21.

29.   Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res* 2019; **47**: e63.

30.   Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: Rapid small-subunit rRNA profiling and targeted assembly from metagenomes. *mSystems* 2020; **5**: e00920-20.

31.   Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinforma* 2020; **70**: e102.

32.   Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2015; **32**: 605–607.

33.   Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 2018; **3**: 836–843.

34.   Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: An adaptive

binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019; **2019**: e7359.

35. Nissen JN, Johansen J, Allesøe RL, Sønderby CK, Armenteros JJA, Grønbech CH, et al. Improved metagenome binning and assembly using deep variational autoencoders. *Nat Biotechnol* 2021; **39**: 555–560.

36. Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics* 2020; **21**: 257.

37. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW, Chkolovski A, Parks DH, et al. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods* 2023; **20**: 1203–1212.

38. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013; **29**: 1072–5.

39. Li H. Minimap2: pairwise alignment for nucleotide sequences. 2017; 1–3.

40. Liu L, Yang Y, Deng Y, Zhang T. Nanopore long-read-only metagenomics enables complete and high-quality genome reconstruction from mock and complex metagenomes. *Microbiome* 2022; **10**: 209.

41. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 2008; **9**: 75.

42. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 2016; **44**: 1–6.

43. Tesson F, Georjon H, Vaysset H, Mordret E, Cury J, Hub B, et al. A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases. 2024.

44. Liao Y, Smyth GK, Shi W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2014; **30**: 923–930.

45. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci* 2012; **131**: 281–285.

46. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res* 2023; **51**: D678–D689.

47.   Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–1313.

48.   Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 2013; **30**: 772–780.

49.   Price MN, Dehal PS, Arkin AP. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009; **26**: 1641–1650.

50.   Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020; **37**: 1530–1534.

51.   Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods* 2009; **6**: 359–362.

52.   Yuen B, Polzin J, Petersen JM. Organ transcriptomes of the lucinid clam Loripes orbiculatus (Poli, 1791) provide insights into their specialised roles in the biology of a chemosymbiotic bivalve. *BMC Genomics* 2019; **20**: 820.

53.   Mordant A, Kleiner M. Evaluation of sample preservation and storage methods for metaproteomics analysis of intestinal microbiomes. *Microbiol Spectr* 2021; **9**.

54.   Kleiner M, Thorson E, Sharp CE, Dong X, Liu D, Li C, et al. Assessing species biomass contributions in microbial communities via metaproteomics. *Nat Commun* 2017; **8**: 1558.

55.   Treude T, Kiel S, Linke P, Peckmann J, Goedert JL. Elasmobranch egg capsules associated with modern and ancient cold seeps: A nursery for marine deep-water predators. *Mar Ecol Prog Ser* 2011; **437**: 175–181.

56.   Amorim K, Loick-Wilde N, Yuen B, Osvatic JT, Wäge-Recchioni J, Hausmann B, et al. Chemoautotrophy, symbiosis and sedimented diatoms support high biomass of benthic molluscs in the Namibian shelf. *Sci Rep* 2022; **12**: 9731.

57.   Carlier A, Ritt B, Rodrigues CF, Sarrazin J, Olu K, Grall J, et al. Heterogeneous energetic pathways and carbon sources on deep eastern Mediterranean cold seep communities. *Mar Biol* 2010; **157**: 2545–2565.

58.   Jacob-Dubuisson F, Locht C, Antoine R. Two-partner secretion in Gram-negative bacteria: A thrifty, specific pathway for large virulence proteins. *Mol Microbiol* 2001; **40**: 306–313.

59.   Holubova J, Stanek O, Juhasz A, Soumana IH, Makovicky P, Sebo P. The Fim and FhaB

adhesins play a crucial role in nasal cavity infection and *Bordetella pertussis* transmission in a novel mouse catarrhal infection model. *PLoS Pathogens* . 2022.

60. Cross KL, Leigh BA, Hatmaker EA, Mikaelyan A, Miller AK, Bordenstein SR. Genomes of gut bacteria from Nasonia wasps shed light on phylosymbiosis and microbe-assisted hybrid breakdown. *mSystems* 2021; **6**.

61. Satchell KJF. Structure and function of MARTX toxins and other large repetitive RTX proteins. *Annu Rev Microbiol* 2011; **65**: 71–90.

62. Linhartová I, Bumba L, Mašn J, Basler M, Osička R, Kamanová J, et al. RTX proteins: A highly diverse family secreted bya common mechanism. *FEMS Microbiol Rev* 2010; **34**: 1076–1112.

63. Sayavedra L, Ansorge R, Rubin-Blum M, Leisch N, Dubilier N, Petersen JM. Horizontal acquisition followed by expansion and diversification of toxin-related genes in deep-sea bivalve symbionts. *bioRxiv* 2019; 605386.

64. Sayavedra L, Kleiner M, Ponnudurai R, Wetzel S, Pelletier E, Barbe V, et al. Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *Elife* 2015; **4**: 1–39.

65. West-Roberts J, Valentin-Alvarado L, Mullen S, Sachdeva R, Smith J, Hug L, et al. Giant genes are rare but implicated in cell wall degradation by predatory bacteria. *BioRxiv* 2023; 568195.

66. Lim SJ, Davis BG, Gill DE, Walton J, Nachman E, Engel AS, et al. Taxonomic and functional heterogeneity of the gill microbiome in a symbiotic coastal mangrove lucinid species. *ISME J* 2019; **13**: 902–920.

67. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, et al. Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proceedings of the National Academy of Sciences* . 2012.

68. Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, et al. Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics* 2011; **11**: 3106–3117.

69. Robinson JJ, Scott KM, Swanson ST, O'Leary MH, Horken K, Tabita FR, et al. Kinetic isotope effect and characterization of form II RubisCO from the chemoautotrophic endosymbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Limnol Oceanogr*

2003; **48**: 48–54.

70.   Horken KM, Tabita FR. Closely related form I ribulose bisphosphate carboxylase/oxygenase molecules that possess different CO_2/O_2 substrate specificities. *Arch Biochem Biophys* 1999; **361**: 183–194.

71.   Berg IA. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. *Appl Environ Microbiol* 2011; **77**: 1925–1936.

72.   Kleiner M, Petersen JM, Dubilier N. Convergent and divergent evolution of metabolism in sulfur-oxidizing symbionts and the role of horizontal gene transfer. *Curr Opin Microbiol* 2012; **15**: 621–631.

73.   Dahl C, Franz B, Hensen D, Kesselheim A, Zigann R. Sulfite oxidation in the purple sulfur bacterium Allochromatium vinosum: Identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiol (United Kingdom)* 2013; **159**: 2626–2638.

74.   Brune DC. Isolation and characterization of sulfur globule proteins from *Chromatium vinosum* and *Thiocapsa roseopersicina*. *Arch Microbiol* 1995; **163**: 391–399.

75.   Grimm F, Franz B, Dahl C. Thiosulfate and Sulfur Oxidation in Purple Sulfur Bacteria. In: Dahl C, Friedrich CG (eds). *Microbial Sulfur Metabolism*. 2008. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 101–116.

76.   Duzs Á, Tóth A, Németh B, Balogh T, Kós PB, Rákhely G. A novel enzyme of type VI sulfide:quinone oxidoreductases in purple sulfur photosynthetic bacteria. *Appl Microbiol Biotechnol* 2018; **102**: 5133–5147.

77.   Hentschel U, Cary S, Felbeck H. Nitrate respiration in chemoautotrophic symbionts of the bivalve Lucinoma aequizonata. *Mar Ecol Prog Ser* 1993; **94**: 35–41.

78.   Moreno-Vivián C, Cabello P, Martínez-Luque M, Blasco R, Castillo F. Prokaryotic nitrate reduction: Molecular properties and functional distinction among bacterial nitrate reductases. *J Bacteriol* 1999; **181**: 6573–6584.

79.   Kraft B, Strous M, Tegetmeyer HE. Microbial nitrate respiration - Genes, enzymes and environmental distribution. *J Biotechnol* 2011; **155**: 104–117.

80.   Drepper T, Groß S, Yakunin AF, Hallenbeck PC, Masepohl B, Klipp W. Role of GlnB and GlnK in ammonium control of both nitrogenase systems in the phototrophic bacterium Rhodobacter capsulatus. *Microbiology* 2003; **149**: 2203–2212.
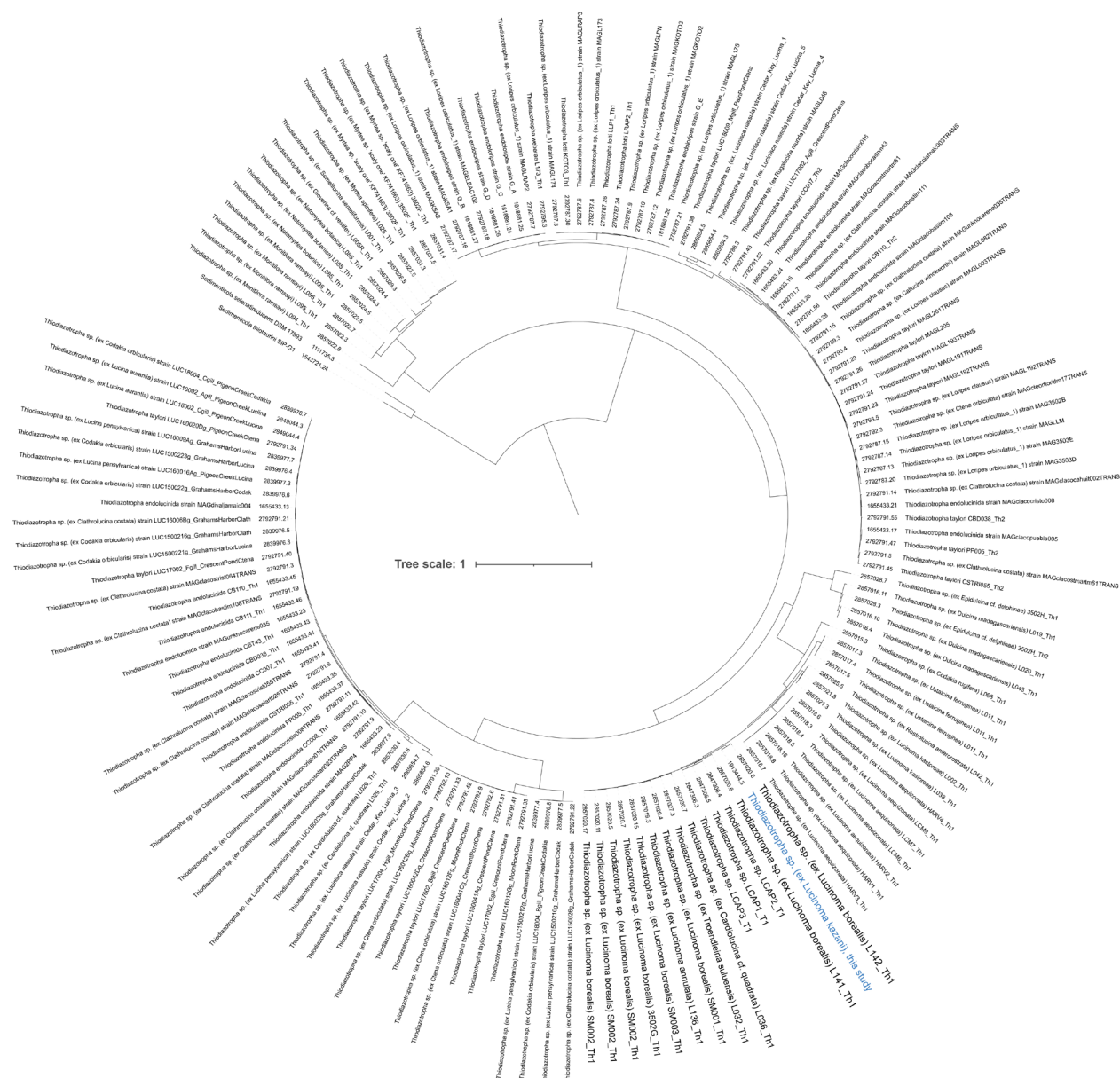
81.   Wang H, Gunsalus RP. Coordinate regulation of the Escherichia coli formate dehydrogenase *fdnGHI* and *fdhF* genes in response to nitrate, nitrite, and formate: Roles for NarL and NarP. *J Bacteriol* 2003; **185**: 5076–5085.

82.   Slobodkina G, Merkel A, Ratnikova N, Kuchierskaya A, Slobodkin A. *Sedimenticola hydrogenitrophicus* sp. nov. a chemolithoautotrophic bacterium isolated from a terrestrial mud volcano, and proposal of Sedimenticolaceae fam. nov. in the order Chromatiales. *Syst Appl Microbiol* 2023; **46**: 126451.

83.   Rey FE, Oda Y, Harwood CS. Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in *Rhodopseudomonas palustris*. *J Bacteriol* 2006; **188**: 6143–6152.

84.   Yang Y, Sun J, Sun Y, Kwan YH, Wong WC, Zhang Y, et al. Genomic, transcriptomic, and proteomic insights into the symbiosis of deep-sea tubeworm holobionts. *ISME J* 2020; **14**: 135–150.

85.   Perez M, Juniper SK. Insight into symbiont population structure among three vestimentifera tubeworm host species at eastern Pacific spreading centres. *Appl Environ Microbiol* 2016; **82**: AEM.00953-16.

86.   Ho BT, Dong TG, Mekalanos JJ. A view to a kill: The bacterial type VI secretion system. *Cell Host Microbe* 2014; **15**: 9–21.

87.   Jani AJ, Cotter PA. Type VI Secretion: Not just for pathogenesis anymore. *Cell Host Microbe* 2010; **8**: 2–6.

88.   Cianfanelli FR, Monlezun L, Coulthurst SJ. Aim, load, fire: The type VI secretion system, a bacterial nanoweapon. *Trends Microbiol* 2016; **24**: 51–62.

89.   Speare L, Cecere AG, Guckes KR, Smith S, Wollenberg MS, Mandel MJ, et al. Bacterial symbionts use a type VI secretion system to eliminate competitors in their natural host. *Proc Natl Acad Sci U S A* 2018; **115**: E8528–E8537.

90.   Guckes KR, Cecere AG, Wasilko NP, Williams AL, Bultman KM, Mandel MJ, et al. Incompatibility of Vibrio fischeri strains during symbiosis establishment depends on two functionally redundant hcp genes. *J Bacteriol* 2019; **201**: 1–14.

91.   Millman A, Melamed S, Leavitt A, Doron S, Bernheim A, Hör J, et al. An expanded arsenal of immune systems that protect bacteria from phages. *Cell Host Microbe* 2022; **30**: 1556-1569.e5.

92. Sberro H, Leavitt A, Kiro R, Koh E, Peleg Y, Qimron U, et al. Discovery of Functional toxin/antitoxin systems in bacteria by shotgun cloning. *Mol Cell* 2013; **50**: 136–148.

93. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, et al. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 2022; **50**: D543–D552.
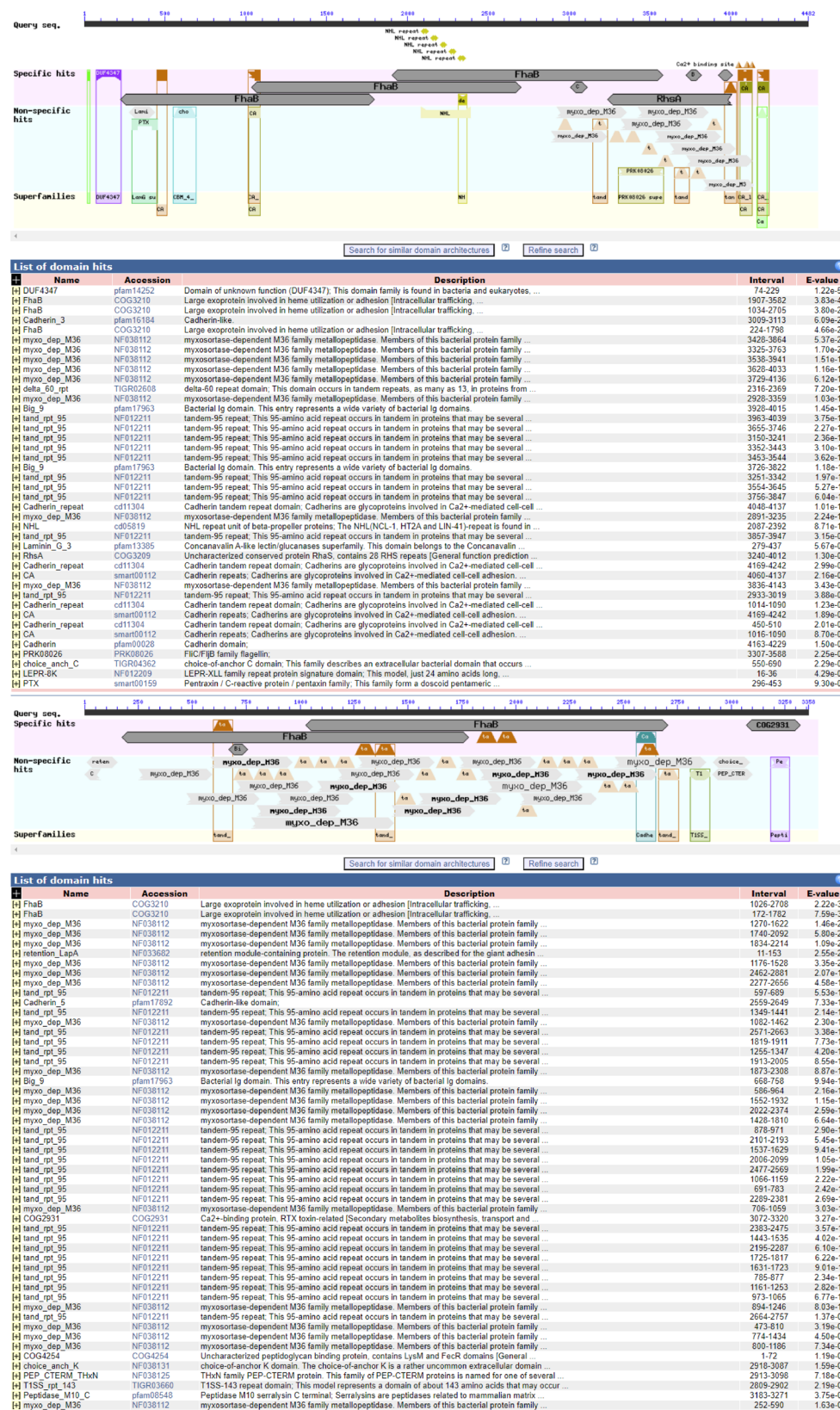
Supplementary Material

Metabolically-versatile *Ca.* Thiodiazotropha symbionts of the deep-sea lucinid clam *Lucinoma kazani* have the genetic potential to fix nitrogen
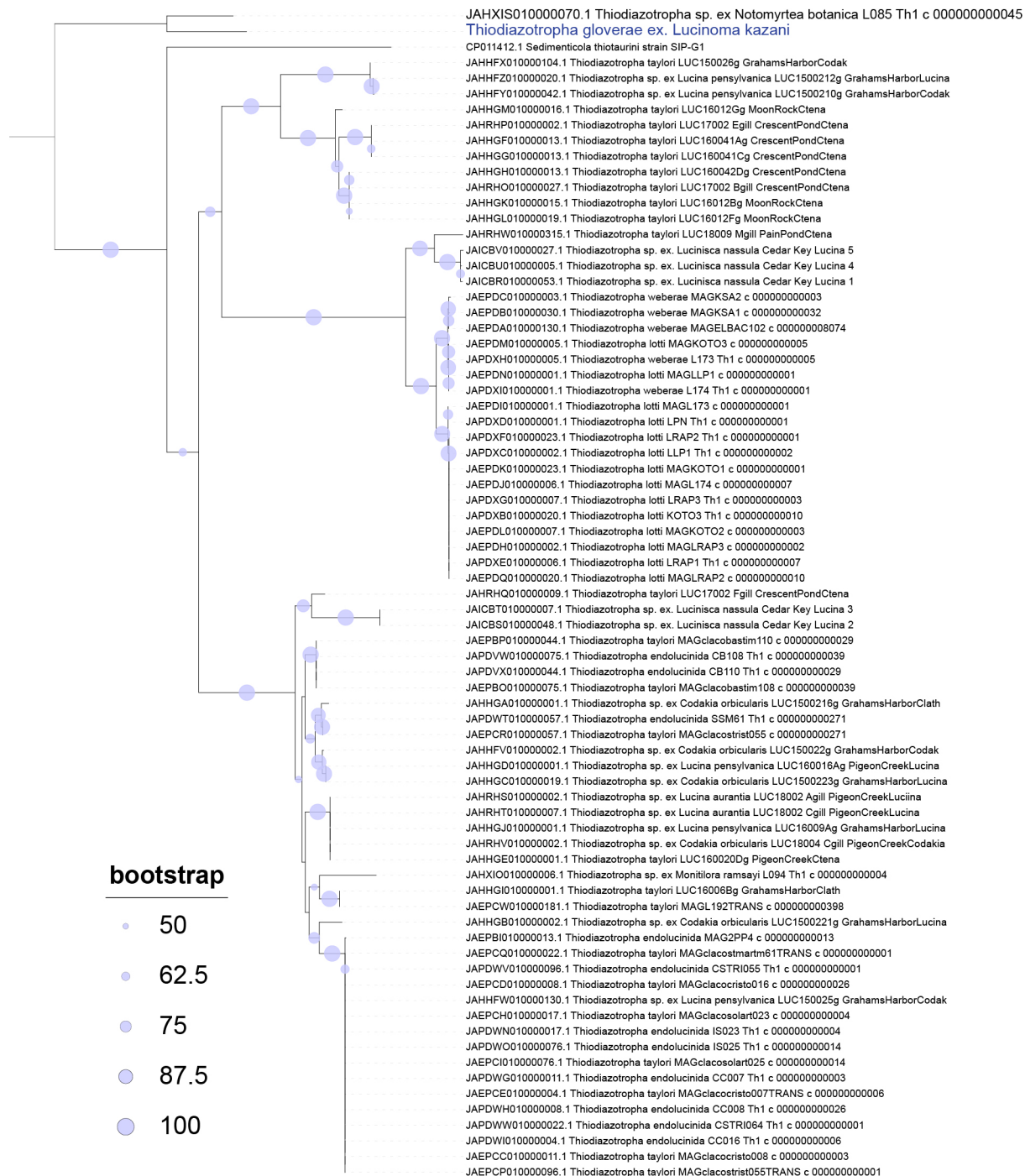
Lina Ratinskaia, Stanislav Malavin, Tal Zvi-Kedem, Simina Vintila, Manuel Kleiner, Maxim Rubin-Blum

Supplementary Figure S1: Phylohgenomic tree of 150 genomes in the Thiodiazotropha clade (2 *Sedimenticola* genomes were added for an outgroup). The tree is based on 411 single-copy genes and uses the JTT model in RAxML. The tree is drawn to scale and rooted at the midpoint. Numbers represent BV-BRC accessions.

Supplementary Figure S2: Domains in the trwo large toxin-like proteins in *Lucinoma kazani* genomes (CD-search, NCBI).

Supplementary Figure S3: Phylogeny of 73 nucleotide sequences encoding NifD in the Thiodiazotropha clade. The maximum likelihood tree is constructed in IQtree using the TN+F+I+G4 model. The tree is rooted at the midpoint. The tree scale represents the number of substitutions per site.

Supplementary Table Legends:

Supplementary Table ST1: Stable isotope analyses of carbon, nitrogen and sulfur in *Lucinoma kazani* tissues.

Supplementary Table ST2: Genomic features of *Ca.* Thiodoazotropha ex. *Lucinoma kazani*. Expression values as RNA read counts and protein abundance as % normalized spectral abundance factor (NSAF) in gills samples of four *L. kazani* specimens. Red fonts represent annotations based on homology with NCBI/UniProt databases (defined as hypotheticals by RAST-tk annotation). Green cells represent putative prophage regions.

Supplementary Table ST3: Protein annotations, Figure 2.