ECOLOGY

Nitrogen fixation in the widely distributed marine γ-proteobacterial diazotroph *Candidatus* Thalassolituus haligoni

Sonja A. Rose[1]*, Brent M. Robicheau[1]†, Jennifer Tolman[1], Debany Fonseca-Batista[1,2]‡, Elden Rowland[1], Dhwani Desai[1,3], Jenni-Marie Ratten[1], Ella Joy H. Kantor[1], André M. Comeau[3], Morgan G.I. Langille[3], Jon Jerlström-Hultqvist[4], Emmanuel Devred[5], Géraldine Sarthou[6], Erin M. Bertrand[1], Julie LaRoche[1]*

The high diversity and global distribution of heterotrophic bacterial diazotrophs (HBDs) in the ocean has recently become apparent. However, understanding the role these largely uncultured microorganisms play in marine N_2 fixation poses a challenge due to their undefined growth requirements and the complex regulation of the nitrogenase enzyme. We isolated and characterized *Candidatus* Thalassolituus haligoni, a member of a widely distributed clade of HBD belonging to the Oceanospirillales. Analysis of its *nifH* gene via amplicon sequencing revealed the extensive distribution of *Cand.* T. haligoni across the Pacific, Atlantic, and Arctic Oceans. Pangenome analysis indicates that the isolate shares >99% identity with an uncultured metagenome-assembled genome called Arc-Gamma-03, recently recovered from the Arctic Ocean. Through combined genomic, proteomic, and physiological approaches, we confirmed that the isolate fixes N_2 gas. However, the mechanisms governing nitrogenase regulation in *Cand.* T. haligoni remain unclear. We propose *Cand.* T. haligoni as a globally distributed, cultured HBD model species within this understudied clade of Oceanospirillales.

INTRODUCTION

Nitrogen (N) is an essential macronutrient limiting primary productivity in large areas of the surface ocean (*1*, *2*). Biological dinitrogen (N_2) fixation, a microbial process restricted to select groups of microorganisms (i.e., diazotrophs) that reduce inert N_2 gas to ammonia (NH_3), partially alleviates N limitation within specific oceanic regions (*1*, *3*–*7*). As a result, this globally important process is estimated to sustain up to 50% of ocean primary productivity (*1*–*6*). Historically, marine N_2 fixation studies predominantly targeted tropical, N-limited, and iron-replete oceanic regions due to nitrogenase energetics and enzymatic requirements, leading to a sampling bias toward large cyanobacterial diazotrophs such as *Trichodesmium* (*8*–*11*). The advent of DNA-based approaches to assess diazotroph diversity, first through the cloning and sequencing of the *nifH* marker gene (*12*) and later through high-throughput *nifH* amplicon sequencing [e.g., see (*13*)], led to the discovery of additional free-living and symbiotic unicellular cyanobacterial diazotrophs that are widely distributed in the surface ocean (*14*–*16*). *NifH* sequence data also revealed the ubiquitous presence of *nifH* genes belonging to noncyanobacterial diazotrophs that are phylogenetically affiliated with the phylum Proteobacteria and have been found to occupy both photic and aphotic oceanic regions (*13*, *17*–*20*). Subsequent global metagenomic surveys of marine microbes also confirmed the importance of noncyanobacterial diazotrophs throughout the world's oceans (*21*) and exposed a fundamental knowledge gap concerning their role within the marine N cycle (*20*, *22*).

Several globally important noncyanobacterial diazotrophs belong to the gammaproteobacteria [hereafter γ-proteobacteria; (*20*)]. This large noncyanobacterial diazotrophic group displays diverse lifestyles, ranging from free-living to particle-associated in both oxic and suboxic waters (*20*). Members of this group have also been observed in aphotic and photic zones of the tropical, subtropical, temperate, and polar regions in both open and coastal waters (*23*–*27*). To date, predominant noncyanobacterial diazotrophs have included the Gamma-A group, prevalent in surface waters across tropical/subtropical regions (*18*, *28*), and the Gamma-4 group (*29*), which has been recently recovered from both surface and deep waters of the temperate and tropical North Pacific (*26*). With most diazotrophic sampling being historically focused on gyre systems and warm water regions, there still remains a lack of comprehensive study of heterotrophic bacterial diazotroph (HBD) groups, including the dominant representatives (*20*, *26*). Furthermore, culturing efforts for globally important γ-proteobacterial diazotrophs have been hindered by knowledge gaps in their distribution, community dynamics, and growth requirements. As a result, our knowledge of the prevalent γ-proteobacterial diazotrophs is largely derived from the distribution of *nifH* sequence signatures and metagenome-assembled genomes (MAGs) in select environments, providing only a glimpse into their elusive lifestyle (*18*, *20*, *21*, *30*–*32*). With limited knowledge of cell-specific N_2 fixation rates (NFRs), growth condition preferences, and community interactions, further progress in understanding noncyanobacterial diazotroph ecophysiology and overall global significance in the N cycle critically depends on the availability of cultured representatives with cosmopolitan distribution (*20*).

[1]Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada. [2]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada. [3]Integrated Microbiome Resource (IMR) and Department of Pharmacology, Dalhousie University, Halifax, Nova Scotia, Canada. [4]Department of Cell and Molecular Biology, Uppsala Universitet, Uppsala, Sweden. [5]Fisheries and Oceans Canada, Bedford Institute of Oceanography, Halifax, Nova Scotia, Canada. [6]Univ Brest, CNRS, IRD, Ifremer, LEMAR, F-29280 Plouzane, France.
*Corresponding author. Email: s.rose@dal.ca (S.A.R.); julie.laroche@dal.ca (J.L.)
†Present address: Department of Biology, Clark University, Worcester, MA, USA.
‡Present address: DOTCAN Institute, Halifax, Nova Scotia, Canada.

Here, we describe *Candidatus* Thalassolituus haligoni, a previously uncultured HBD isolated from a coastal embayment in the Northwest Atlantic (NWA) Ocean. Phylogenomic analyses demonstrate that the bacterium belongs to one of the major clades of marine γ-proteobacterial diazotrophs (*20*, *21*, *32*) and, more specifically, is affiliated with a large clade of diazotrophs within Oceanospirillales. Through characterization of the genome, proteome, and physiology of *Cand*. T. haligoni, we describe this isolate's N_2 fixation capability and preferential use of C_4 carbon sources for heterotrophic growth. We further use several approaches [amplicon sequencing, quantitative polymerase chain reactions (PCRs), and comparative genomics] to demonstrate its cosmopolitan distribution, presence at high abundances in the tropical, temperate, and polar regions, and genomic similarities to a recently curated MAG from the Arctic Ocean. Given its wide distribution, versatile genome, cultivability, and N_2 fixation ability, *Cand*. T. haligoni is an important representative of a major clade of γ-proteobacterial diazotrophs within Oceanospirillales and will contribute to understanding the functional role of HBD's impact of N_2 fixation in a changing ocean.

RESULTS

Cand. T. haligoni—an HBD

The HBD, *Cand.* T. haligoni sp. nov. BB40, was initially isolated from seawater collected at 10 m depth in 2014 (from 44°41′30″N, 63°38′30″W Halifax, NS, Canada; fig. S1). The flagellated, rod-shaped bacterium grew as white colonies on amended f/2 medium [6:1, Nitrogen:Phosphorous (N:P)] with a cell diameter of ~0.5 μm (shortest axis) and ~2.5 μm (longest axis; Fig. 1A and fig. S2A). Globular formations consistently visible at polar ends of cells are likely storage compounds such as β-polyhydroxybutyrate (PHB). The pure cell culture, grown in artificial seawater (ASW)–amended f/2 medium (6:1, N:P), reached a maximum biomass of ~6×10^{10} cells liter^{-1} at growth rates of 0.7 to 0.9 day^{-1} before entering a prolonged stationary phase (>10 days) where cells formed small aggregates (figs. S1B and S2). The isolate grew best with C_4 dicarboxylic acid substrates as sole carbon sources, while equivalent concentrations of D-glucose did not support significant growth or biomass accumulation (fig. S3 and text S1). An initial alignment of the full-length *16S* rRNA (ribosomal RNA) gene in *Cand.* T. haligoni resulted in a pairwise identity of 96.54% to *Thalassolituus oleivorans* (a nondiazotroph), supporting the view that the isolate likely belongs to the same genus within the order Oceanospirillales (family: Oceanospirillaceae). We named the isolate *Cand.* T. haligoni. sp. nov. BB40 to note the geographic site of its initial isolation.

Cand. T. haligoni belongs to a widely distributed clade in the Oceanospirillales

The closed genome of *Cand.* T. haligoni is ~4.37 mega–base pairs (Mbp) with an average GC content of 53.4%. The genome encodes 4019 putative protein-coding sequences (CDS) corresponding to 463 SEED subsystems and 61 RNA genes (tRNAs and rRNAs; Fig. 1 and table S1). Phylogenomic analysis of closely related genomes and a *nifH* phylogenetic tree placed *Cand.* T. haligoni within a clade of γ-proteobacteria that are dominated by uncultured HBD MAGs recovered from large oceanographic metagenomic surveys [Fig. 2; (*21*, *32*)]. Analysis of average nucleotide identity (ANI) and amino acid identity (AAI) matrices (Fig. 3, A and B, and tables S2 and S3) also supported phylogenomic clustering within the genus *Thalassolituus* (Fig. 2A). Whole-genome comparisons between *Cand.* T. haligoni and closely related marine MAGs (HBD Gamma 04, 06, 08, 09, and 11; hereafter referred to collectively as γ-HBD) fell below the species threshold of 95% and above the genus threshold of 60% for ANI and AAI, respectively [tables S2 and S3; (*33*, *34*)], indicating that *Cand.* T. haligoni constitutes a new species within the *Thalassolituus* genus. In support of this, a phylogenomic tree places *Parathalassolituus penaei* as the closest neighbor to *Cand.* T. haligoni, within a monophyletic *Thalassolituus* clade (Fig. 2A). Given the ANI guideline whereby new genera have <60% AAI compared to other known genera, and since *P. penaei* had an AAI of 62% to *T. oleivorans* (table S3), it is more parsimonious to place *P. penaei* within the *Thalassolituus* genus (*33*, *34*). In addition, re-assembly of a recently curated MAG from the Arctic Ocean (*32*), Arc-Gamma-03 (98% complete; 87 contigs; 4.28 Mbp), shared 99% ANI identity with our isolate (table S2 and Fig. 3). Although the Arc-Gamma-03 MAG is not fully complete, it shares 100% pairwise identity with *Cand.* T. haligoni for both its *16S* rRNA and *nifH* genes (data S1A). We therefore propose that the Arctic-recovered MAG and *Cand.* T. haligoni belong to the same species.

The pangenome analysis between Arc-Gamma-03, *P. penaei*, HBD Gamma 04, 08, 11, and *Cand.* T. haligoni identified a total of 21,956 genes belonging to the diazotrophic *Thalassolituus* clade, resulting in 9047 total gene clusters, a core genome of 7309 shared gene clusters, and 4597 gene clusters shared between two or more, but not all lineages (Fig. 3C). The core genome included genes belonging to replication, lipid transport, transcription, energy production, posttranslational modification, and translation (Fig. 3D and data S1B). All *nif* genes from the six genomes analyzed were recovered within the core genome. Of the 1042 gene clusters shared only between Arc-Gamma-03 and *Cand*. T. haligoni, most belonged to replication, mobilomes, DNA repair, secondary metabolism, cell wall biogenesis, transposases, and DNA binding transcriptional regulators or were of unknown function (data S1C). Analysis of *Cand.* T. haligoni singleton genes identified 125 gene clusters, most of which were functionally annotated to genes involved in defense, transcription, mobilomes, energy conversion, general function, or unknowns (Fig. 3E and data S1D).

Genome-predicted metabolic pathways in *Cand*. T. haligoni

The complete N_2 fixation pathway, including 26 *nif* genes, is present in the genome of *Cand*. T. haligoni and is distributed across three gene clusters. The structural *nif* cluster includes the *nifHDK* genes (Fig. 1C). The regulatory *nif* cluster contains the transcriptional regulator *nifA* and repressor *nifL* and is located ~15,000 nucleotides (nt) upstream of the structural *nif* cluster. The stability cluster, which contains stability and accessory genes such as *nifENXUWSVMZ* and *sufA*, is located 200,000 nt downstream from the structural cluster. Genes embedded between the various *nif* clusters code for C_4 transporters (*dct*PQM), electron transport (*rnfABCDEG*), molybdenum transporters (*moaABCDE*), and O_2 defense [hemerythrin (*hr*)]. Overall, the three clusters constitute 20,000 nt of the genome, with the stability cluster being the largest (Fig. 1C).

Several carbon degradation pathways for compounds such as catechols, benzoate derivatives, lignin, PHBs, fermentation, and genes for the central carbohydrate metabolism are predicted from the genome annotation (Fig. 1D and table S1). Genes for the uptake of diverse carbon sources were also abundant. In particular, the genome contains multiple genes for C_4 tripartite ATP-independent

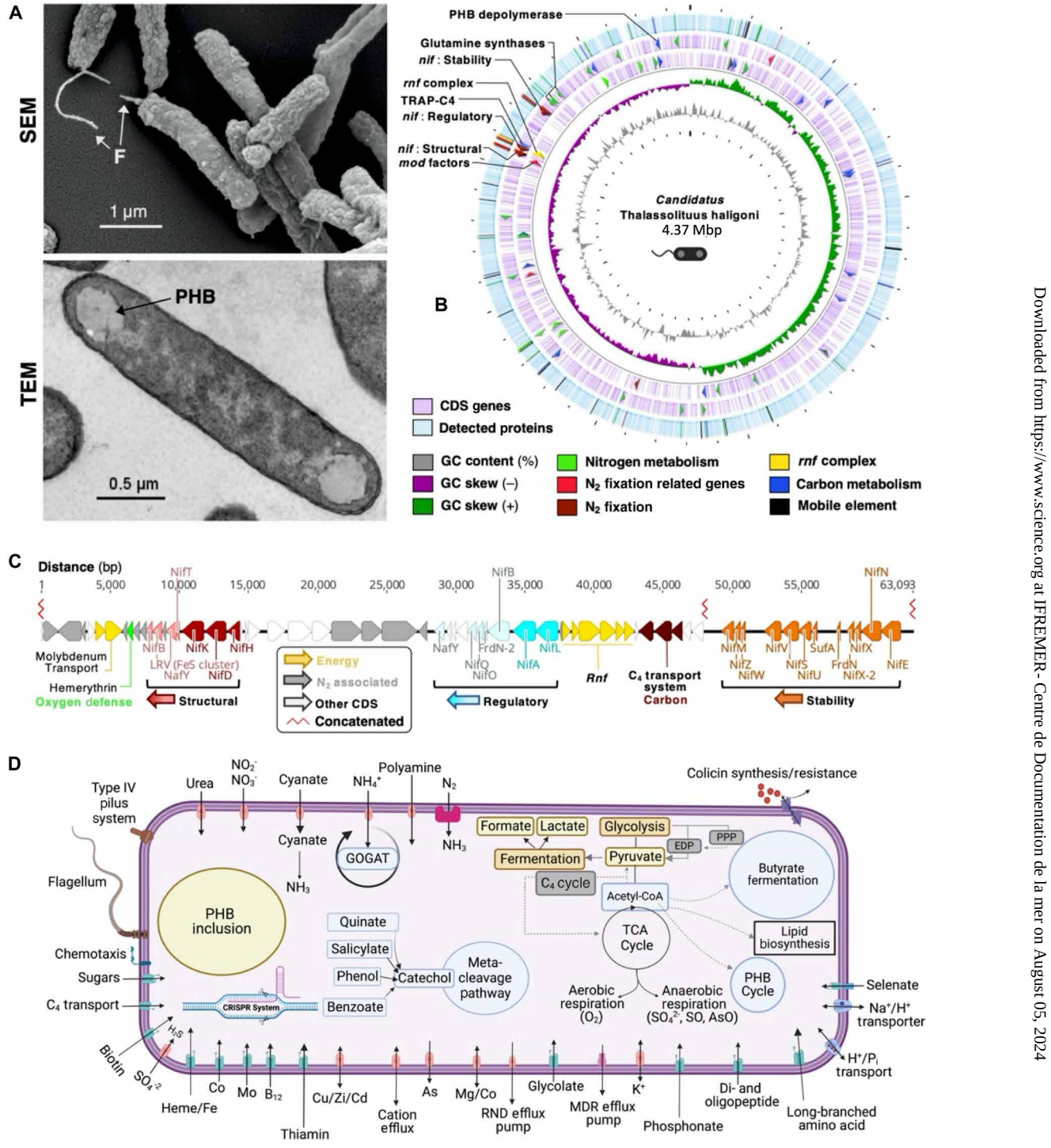

Fig. 1. Scanning electron microscopy and genomic, proteomic, and genome-assisted metabolic models of *Cand.* T. haligoni. (**A**) Cell morphology of the isolate under N_2-fixing conditions using scanning electron microscopy (SEM; top) and transmission electron microscopy (TEM; bottom). SEM magnification = 17.71×10^3. Arrows indicate flagella (F) and PHB granules (PHB). (**B**) Proteome-genome circular map under NO_3^- depleted conditions. Starting from the innermost ring: GC content (gray), GC skew for forward (green) and reverse strands (magenta), leading and lagging strand CDSs (light purple), and detected proteins (blue). Arrows in the genome indicate genes of interest [*nif* clusters, N_2 fixation–related genes are molybdate uptake and transformation genes (*mod* factors), TRAP-C_4 uptake, and glutamine synthesis] and have corresponding detected proteins highlighted in respective colors. White spaces in the proteome indicate regions not detected. (**C**) *Nif* clusters (red, blue, and orange) and associated N_2 fixation genes (gray). N_2 fixation–associated genes include glycogen phosphorylase and glycogen uptake genes. Numbers along the line indicate base-pair locations within the genome. A nucleotide break in the center of the diagram from 3,598,007 nt to 3,737,467 nt is represented by the zig-zag line in red (concatenated). Clusters are color-coded to the associated *nif* clusters, with accessory genes to that cluster indicated by respective color hues. (**D**) Simplified metabolic schematic of the isolate based on the genome annotation. The schematic was created using the RAST server (*143*) and BioRender.

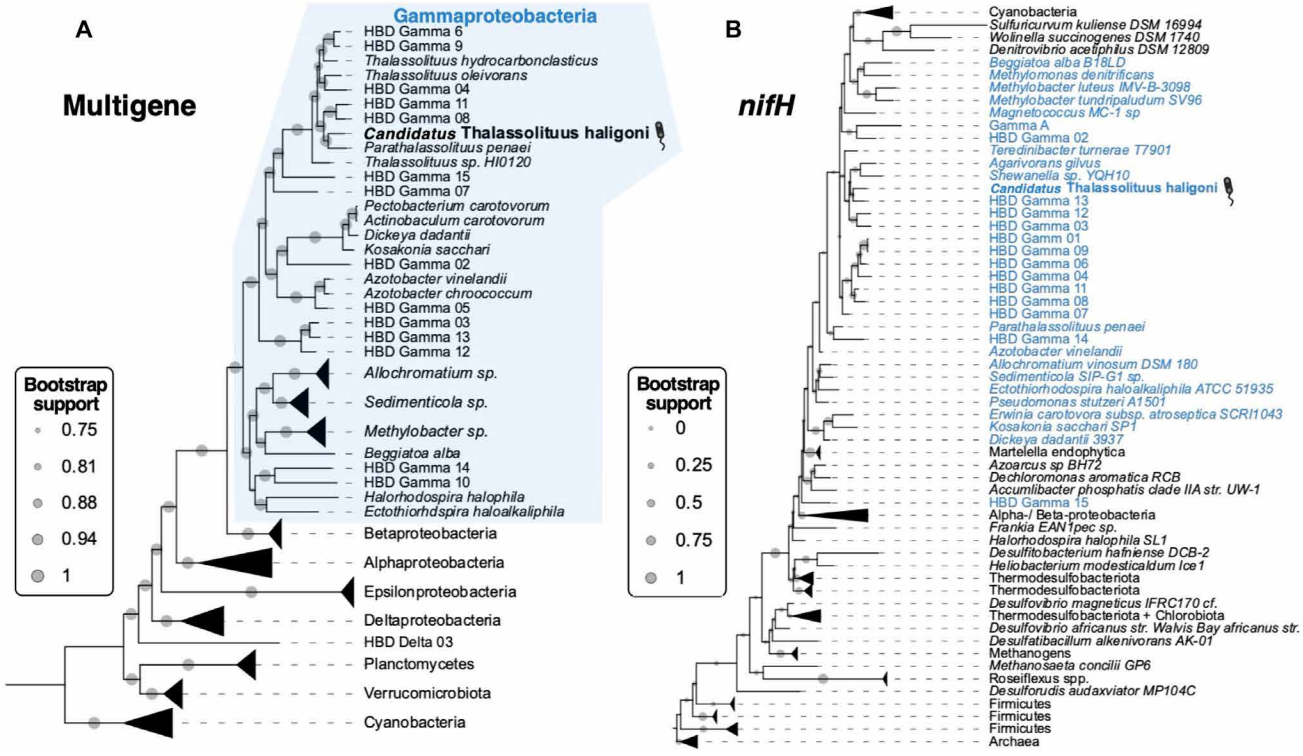

Fig. 2. Phylogenetic relationships of *Cand.* T. haligoni. (**A**) Phylogenomic analysis of 71 concatenated genes from 74 genomes. The maximum likelihood tree was constructed using Anvi'o and the Jones-Taylor-Thorton model. The concatenated protein sequences were subjected to 1000 bootstraps and visualized in FastTree. (**B**) *NifH* maximum likelihood phylogenetic tree of 166 diazotroph sequences using K2+G model with 1000 bootstraps. Collapsed clades are scaled to the number of sequences within the clade. γ-Proteobacteria in both trees are highlighted in blue.

periplasmic (TRAP) transporters (51 genes), glycolate and glyoxylate interconversions (7 genes), glycerol-3 phosphate uptake (7 genes), and mannose/sucrose uptake (1 gene; table S1). The isolate contains well-developed systems for regulating intracellular O_2 levels such as heme oxygenases, bacterial hemoglobin, hemerythrin, and several superoxide dismutases. The genome also contains genes for the CRISPR-Cas9 system, antibiotic resistance, a type IV pilus system, and chemotaxis receptors (Fig. 1D and table S1). Genes for the uptake and assimilation of fixed N (e.g., NH_4^+, urea, and NO_3^-) were also distributed throughout the genome (Fig. 1, B and D).

N_2 fixation and nitrogenase are detected over prolonged nitrate exposure in *Cand.* T. haligoni cultures

Shotgun proteomic analyses confirmed the expression of 3144 proteins in pure cultures of *Cand.* T. haligoni grown in a range of NO_3^- concentrations, accounting for 78.8% of the genome-predicted proteome (Fig. 1B and data S2A). Hypothetical proteins accounted for approximately 30% of detected proteins (930 proteins), while ribosomal and N_2 fixation proteins accounted for 3% (83 proteins) and 1.6% (52 proteins) of the total proteins, respectively (data S2B). Of the 21.2% of proteins (848 proteins) that were not detected, most were hypothetical proteins (464 proteins) and mobile elements (59 proteins; data S2B). Other undetected proteins included motility- and chemotaxis-related proteins and some DNA repair proteins. As cells were harvested in stationary phase, these proteins were expected to be in relatively low abundance (fig. S4 and data S2C).

Expression of NifHDK and 40 of the 55 other proteins involved in N_2 fixation regulation were detected across all NO_3^- conditions and replicates (Fig. 4A). Proteins that were detected in some but not all biological replicates included MoaD, NifNVWMB, C4_supp, NafY, hemerythrin (hr), RnfA_2, NtrC, and MoaB_2 (Fig. 4A). Duplicate copies of the electron shuttle complex, Rnf, encoded by two small operons of six genes in the genome, were recovered within the proteome, with 8 of the 12 proteins individually detected as a unique peptide group (Fig. 4A and data S2D). NFR measurements ranged from 0.004 to 0.014 fmol N cell^{-1} day^{-1}, with no statistically significant differences observed between treatments ($P = 0.16$, $\alpha = 0.05$). Overall, the detectable NFR across the four sampling points in our experiments indicated that a functional nitrogenase was present under all growth conditions (Fig. 4B). Pairwise comparisons of protein relative abundances across the four treatments also showed no statistically significant differences ($P > 0.05$), suggesting that in our experiments NO_3^- concentrations did not play a major role in the regulation of nitrogenase expression in *Cand.* T. haligoni (data S2E). Together, the expression of NifHDK proteins and the measured NFR confirm that the isolate was fixing N_2 when under prolonged NO_3^- exposure.

Cand. T. haligoni has a cosmopolitan distribution in the ocean

The *nifH* amplicon sequence variant (ASV) of *Cand.* T. haligoni was detected at 100% identity in field samples covering a broad geographic range (Fig. 5). *NifH* distribution within a compiled dataset

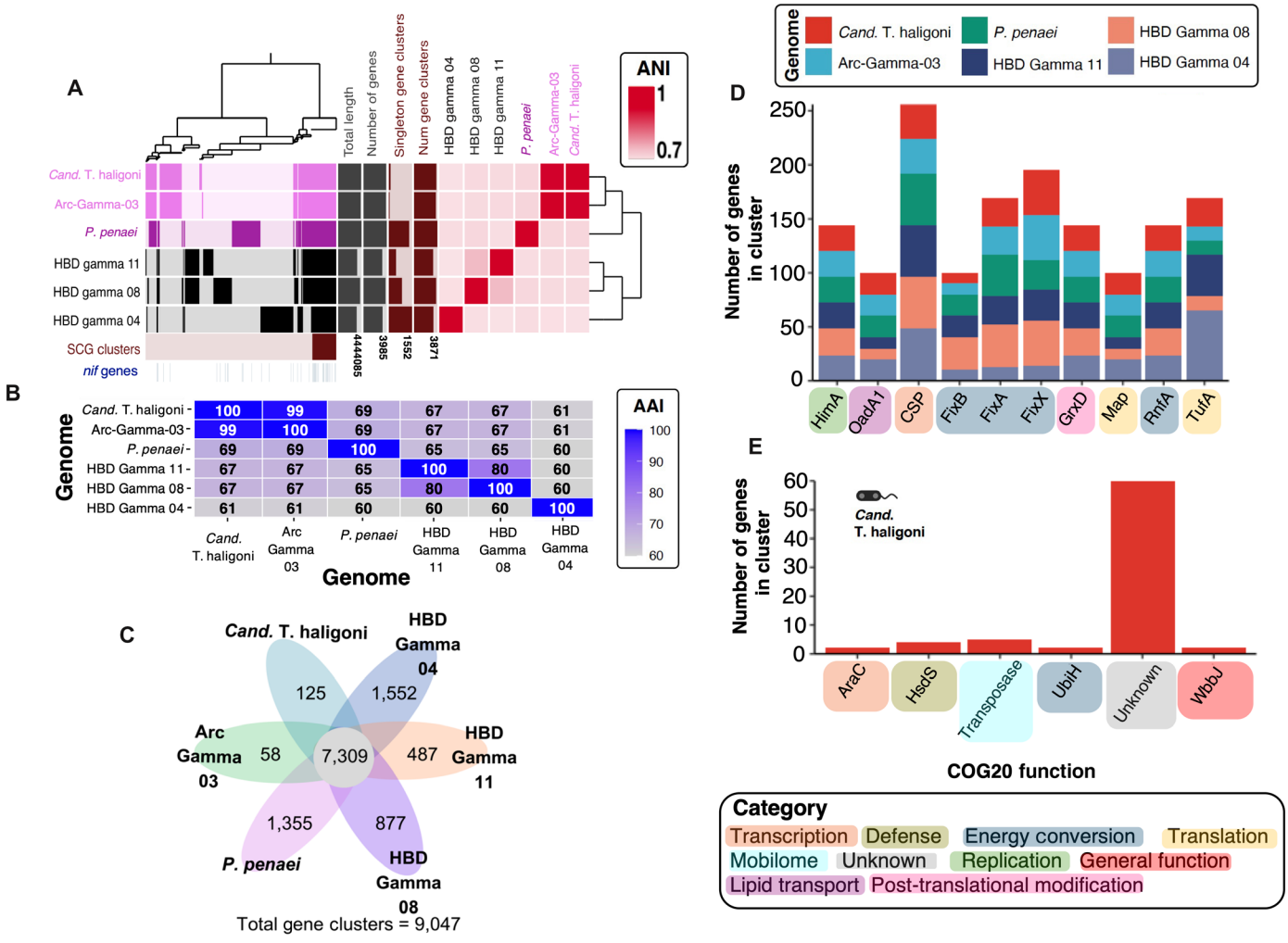

Fig. 3. Pangenome analysis of *Cand.* T. haligoni and close Oceanospirillales relatives. (**A**) Pangenome phylogram sorted based on the presence and absence of gene clusters across all six genomes. Genomes are colored based on close relationships determined by ANI comparisons (shown on a red scale). Phylogeny of ANI tree is based on the full identity between genomes. The % cutoff range for genus classification with ANI was set at 75% according to (*34*). Pangenome analysis was conducted according to the Anvi'o guidelines found in (*93*). (**B**) AAI of the six genomes. The accepted cutoff of 60% for classification of the same genus was according to (*34*). (**C**) Venn diagram of the number of shared gene clusters between all six genomes. Note that numbers only show genes that are unique to the genome. (**D**) Most abundant core gene clusters shared between the six genomes: Flavoproteins (*FixXAB*), NAD^+ oxidoreductase RNF (*RnfA*), glutaredoxin-related protein (*GrxD*), DNA binding protein IHF-α (*HimA*), cold shock protein (*Csp*), methionine aminopeptidase (*Map*), pyruvate/oxaloacetate carboxyltransferase (*OadA1*), and translation elongation factor (Ef-Tu) (*TufA*). (**E**) Gene clusters only found in *Cand.* T. haligoni. Clusters displayed only include gene clusters containing >1 gene for simplicity (see data S5 for full list). Gene cluster functions include transposases (transposase), 2-polyprenyl-6-methoxyphenol hydroxylase (*UbiH*), restriction endonuclease S subunit (*HsdS*), acetyltransferase/isoleucine patch superfamily (*WbbJ*), AraC-type DNA binding domain (*AraC*), and unassigned genes (unknown). Clusters in both (C) and (D) are organized based on their function and colored according to their category based on COG20 annotation.

of 865 eDNA samples established the presence of *Cand.* T. haligoni in the Arctic Ocean, in the North and South Atlantic and North Pacific Ocean, as well as along the shelf break off Japan and the Eastern China Sea (Fig. 5C). *Cand.* T. haligoni *nifH* sequence was also detected at depths of 1200 m (fig. S5A). Attempts to detect variants of the isolate by lowering the *nifH* gene ASV identity cutoff to 98% did not yield significant results, suggesting that *Cand.* T. haligoni is the dominant ecotype with respect to its *nifH* ASV signature (fig. S5). On the basis of its *nifH* gene sequence, the geographic distribution of *Cand.* T. haligoni is much broader than other MAGs of γ-proteobacterial diazotrophs recovered from large metagenomic surveys of marine microbes (e.g., HBD Gamma 04).

A redundancy analysis (RDA) of qPCR data confirmed that *Cand.* T. haligoni has a wide tolerance to various environmental conditions, as it showed no distinct preference for dissolved inorganic nitrogen (DIN) concentrations, chlorophyll a, or temperature (fig. S5B). For example, *nifH* gene copies reached values up to 1.25×10^5 copies $liter^{-1}$ in 0.03 μM NO_3^- at 16°C during the GEOVIDE cruise and up to 9×10^3 *nifH* gene copies $liter^{-1}$ in 16 μM NO_3^- at 1°C along the coast of Baffin Bay during the Canadian Arctic GEOTRACES cruise (fig. S5A and data S3A). Although sampling frequency was limited in deep water, *nifH* gene copies $liter^{-1}$ ranged from ~1×10^2 to 1×10^3 copies $liter^{-1}$ (fig. S5A; GEOTRACES). A coastal weekly time series from the NWA showed concentrations

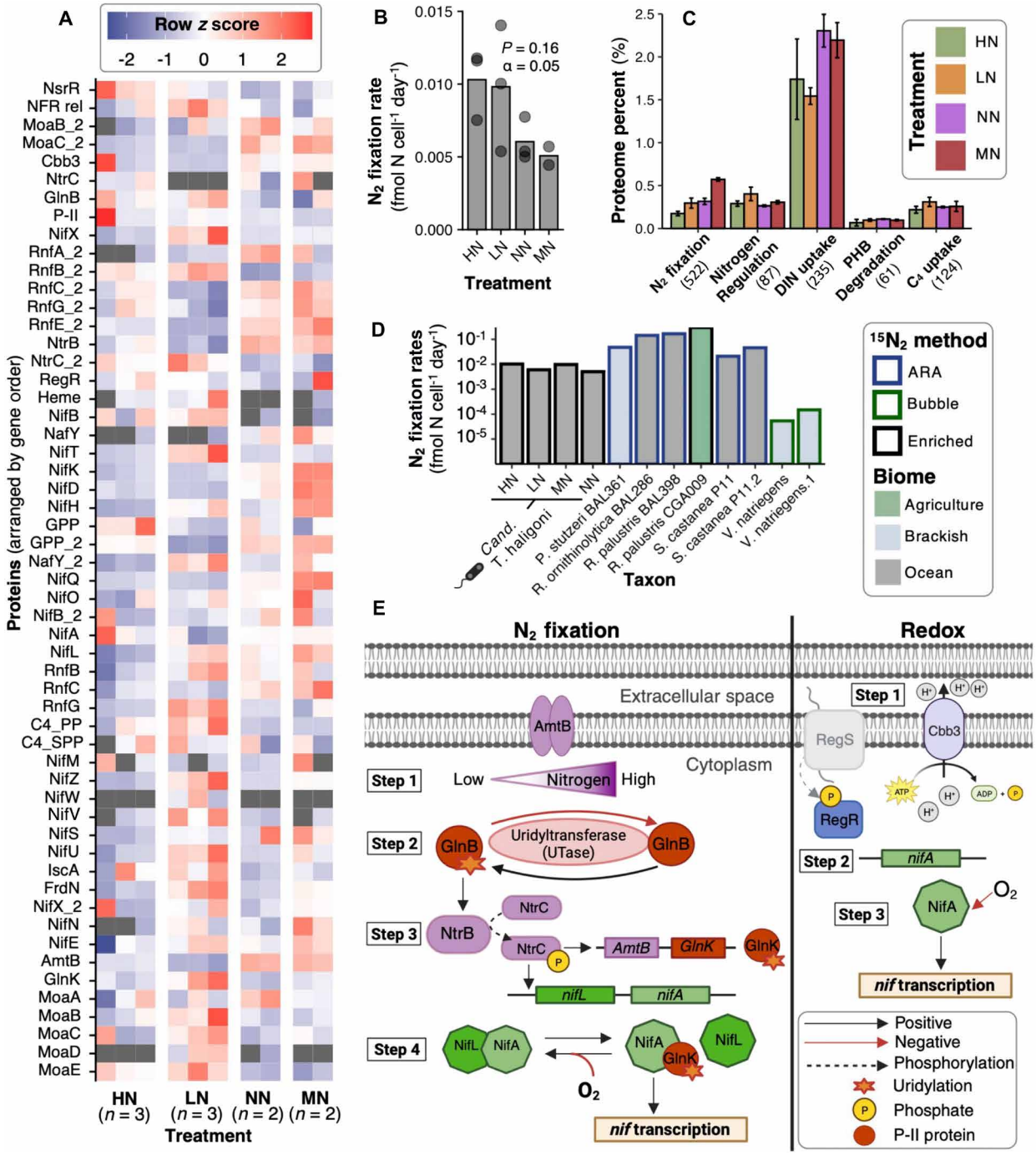

Fig. 4. N_2 fixation protein regulation in *Cand*. T. haligoni. (**A**) *Z* scores of N_2 fixation protein expression under prolonged NO_3^- exposure. Samples were taken in a time course (refer to methods for details). Gray squares represent proteins that were not detected. Individual *Z* score = (*X* − mean of *X* across treatments) ÷ SD of the mean for *X*, where *X* = raw protein relative abundance. (**B**) NFR measurements from samples in (A). Bar heights show average NFRs, with gray dots representing individual biological replicates. *P* value (not significant) was determined across all treatments using a Kruskal-Wallis test. (**C**) Percentage of *Cand*. T. haligoni's proteome attributed to key cellular functions under different NO_3^- conditions. The bars represent the average sum across biological replicates ±1 SE. Numbers in brackets represent the number of peptides attributed to the cellular function. (**D**) Comparison of NFRs for other HBDs from various biome locations with rates determined using acetylene reduction assay (ARA), the $^{15}N_2$ bubble method (bubble), or the $^{15}N_2$ dissolution method (enriched). Values are displayed on a logarithmic scale and are from (*23, 70, 79, 80, 144*) (*Sagittula castanea*, *Pseudomonas stutzeri* BAL361, *Rhodopseudomonas palustris* CGA009, *Rhosopseudomonas palustris* BAL398 & *Raoultella ornithinolytica* BAL286, *V. natriegens*). Note that for the species *S. castaena* PII and *V. natriegens*, replicate NFR measurements were measured and are reported here. (**E**) N_2 fixation regulation pathways detected in the proteome data of *Cand*. T. haligoni. Pathways include standard N_2 fixation and redox balancing. RegS, displayed in gray, was not detected. Proteins not detected in the pathway are pictured in gray. Proteins that are used in the N status of the cell are shown in purple, fixed N regulation proteins (GlnBK) are shown in red, and *nif* regulatory proteins (NifL and NifA) are shown in green.

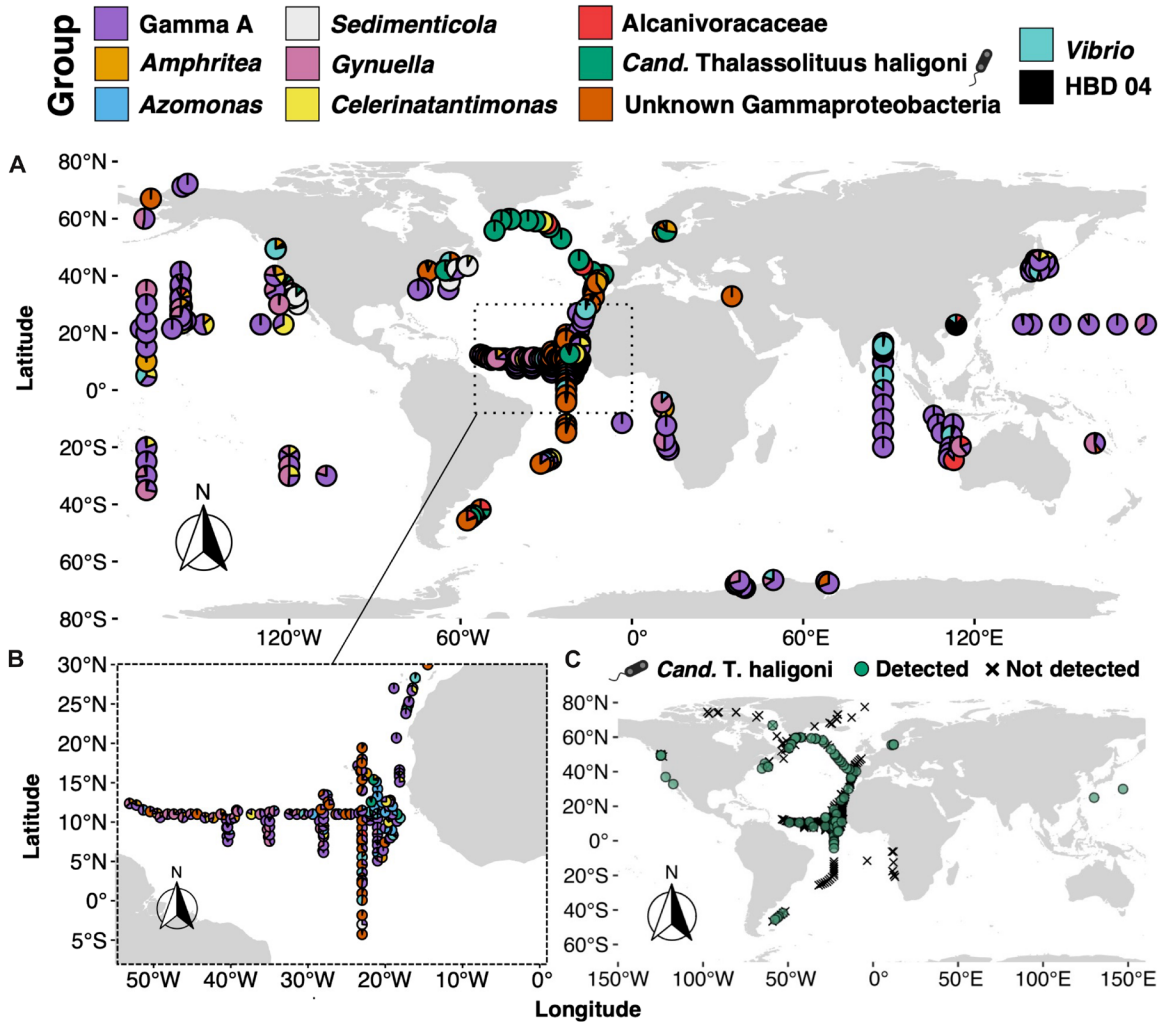

Fig. 5. Global surface distribution (0 to 30 m) of *Cand.* T. haligoni. (**A**) Global surface distribution of *Cand.* T. haligoni and other dominant γ-proteobacterial HBD *nifH* ASVs detected by Illumina amplicon sequencing. (**B**) Map inset of (A) showing dominant surface γ-proteobacterial ASVs in equatorial regions. Criteria for considering an ASV as dominant required its detection in at least 80 of the 865 surface water samples (0 to 30 m) and having at least 50 reads across all samples. Pie charts show the relative abundances as a percentage for dominant γ-proteobacterial HBD ASVs within each sample rather than the relative abundances of the entire *nifH* reads per sample. (**C**) *nifH* qPCR presence or absence for *Cand.* T. haligoni. Note that data points within the Canadian Arctic are not in (A) as the samples did not meet the cutoff criteria.

reaching up to 10^7 and 10^6 *nifH* gene copies liter^{-1} in euphotic and aphotic zones, respectively (fig. S7). In this coastal time series, the peak abundance of *nifH* gene copies per liter (>10^7) occurred in the fall of 2015 and 2017 in surface waters with approximately 2 μM DIN. During this same period, >10^6 *nifH* copies liter^{-1} were detected at 60 m depth, where DIN concentrations reached ~30 μM (fig. S7). Overall *nifH* qPCR and amplicon sequencing demonstrated that *Cand.* T. haligoni is broadly distributed geographically and is present in euphotic and aphotic regions of the water column (Fig. 5 and fig. S5A).

The *nifH* ASV signature of the isolate was recovered in both >3-μm and 0.2- to 3-μm size fractions. However, in both the Bedford Basin and Canadian Arctic samples, higher relative abundances for the isolate's *nifH* ASV were found in the 0.2- to 3-μm size fraction than in the >3-μm size fraction (figs. S8 and S9), suggesting that the isolate is free-living at least part of the time. The presence of the *nifH* ASV signature within the >3-μm size fraction suggests that the isolate may also be associated with larger particles through an epibiont lifestyle, self-aggregation, or colonization of organic matter; however, it is also possible that retention on >3-μm filters could simply be due to biomass accumulation during the filtration of larger volumes of seawater. Overall, *Cand.* T. haligoni appears to have a heterotrophic lifestyle, is widely distributed throughout different ocean basins, and is found throughout the entire water column, irrespective of the DIN concentration.

DISCUSSION

Cand. T. haligoni belongs to an important clade of marine diazotrophs within Oceanospirillales

Our study establishes *Cand.* T. haligoni as a widely distributed HBD within the Oceanospirillaceae family—a notable clade in the Oceanospirillales order [class γ-Proteobacteria] (Fig. 2). The closed genome of *Cand.* T. haligoni suggests that it has high metabolic

versatility for the degradation of simple and complex carbohydrates and oil byproducts, as is commonly found in Oceanospirillales (*35*). In addition to its N_2 fixation ability, the isolate has a full gene complement for the synthesis and degradation of PHB granules (Fig. 1). Oceanospirillaceae encompasses 26 diverse genera, representing largely marine species (*35*). Most of the Oceanospirillaceae genera, including *Cand.* T. haligoni, exhibit a wide temperature growth range (4° to 37°C; fig. S9) and have genome sizes ranging from 2.2 to 5.7 Mbp. At ~4.3 Mbp, the isolate's genome falls toward the upper limit of this size range (*35*–*37*).

Cand. T. haligoni is the only known cultured N_2-fixing representative of the Oceanospirillaceae family that is widely distributed throughout the open ocean (Fig. 5C and fig. S5). Our ability to culture the isolate has provided valuable insights into the lifestyle of other uncultured HBDs whose presence is only detected through metagenomic or *nifH* amplicon sequencing (*32*). The global importance of *Cand.* T. haligoni is further supported by the recent discovery of the Arctic-associated MAG known as Arc-Gamma-03 with 100% identity to the *nifH* and *16S* rRNA genes of our isolate (*32*). Whole-genome comparison (ANI/AAI = 99.37% identity) further confirmed that Arc-Gamma-03 and *Cand.* T. haligoni belong to the same species (table S3 and Fig. 3A). Hence, we propose that our isolate is a key model species to study the ecophysiology of the diazotrophic Oceanospirillales clade in polar and temperate conditions, as supported by the broad range of growth temperature tolerated by *Cand.* T. haligoni (fig. S10). Moreover, of the 3585 CDS in the Arc-Gamma-03 MAG, only 39 genes, mostly annotated as hypothetical, did not have a homolog in the genome of *Cand.* T. haligoni, highlighting the high similarity in metabolic potential between the two genomes. In contrast to Shiozaki *et al.* (*32*), *Cand.* T. haligoni was retained on a 0.2-μm filter in the NWA sector, Labrador Sea, and Baffin Bay regions. The discrepancies between our observations and those of Shiozaki *et al.* (*32*), who detected the ASV of the isolate in the ultrasmall (<0.2 μm) size fraction, are currently unresolved, but the size differences may be due to dormancy, spore formation, or nutrient limitation causing a reduction in cell size in a harsher environment (*38*, *39*).

Similar to other taxonomic groups that include diazotrophs (e.g., cyanobacteria), the ability to fix N_2 gas is not universal among Oceanospirillales, particularly in the genus *Thalassolituus*. Comparison between the phylogenomic and phylogenetic *nifH* trees shows that several *Thalassolituus* species (e.g., *T. oleivorans* and *T. hydrocarbonoclasticus*) do not contain the *nifH* gene, while other close relatives within this clade do (Fig. 2). Pangenome analysis indicated that *Cand.* T. haligoni, Arc-Gamma-03, *P. penaei*, and HBD Gamma 04, 08, and 11 share most of the *nif* genes (i.e., *nifHDK*) within the core genome (Fig. 3A). However, as the *nifH* gene of the six genomes was placed within different clades, it is possible that *nifH* acquisition occurred independently within each group from different ancestral species (Figs. 2 and 3A). Lateral gene transfer (LGT) has frequently been suggested as an explanation for the acquisition of diazotrophy (*40*, *41*), a pattern that may also apply to HBDs within the Oceanospirillales clade. Furthermore, the pangenome analysis revealed that the six genomes contain multiple copies of the homologous genes *nafY* and *nifX* (data S1B). Proteomic analysis confirmed the expression of these genes in *Cand.* T. haligoni, suggesting that their presence serves as a built-in redundancy mechanism to ensure proper nitrogenase synthesis, as genetic knockouts of either *nifX* or *nafY* in cyanobacterial diazotrophs significantly reduced NFR (*42*). As little is known about most HBDs, analysis of the core genome can provide better insight into the necessary genes required for clade classification of Oceanospirillales HBDs, niche occupancy, and nitrogenase expression.

Genome and proteome point to multiple oxygen protective strategies in *Cand.* T. haligoni

Although protective O_2 strategies and N_2 fixation regulation are relatively well understood in cyanobacterial (i.e., *Trichodesmium*, *Crocosphaera*, and *Anabaena*) and terrestrial diazotrophs (i.e., *Klebsiella*, *Bradyrhizobia*, and *Rhodobacter*), O_2 protection mechanisms in marine HBDs are not well characterized (*19*, *31*, *43*). Insights from our genomic and proteomic approaches suggest that alginate biosynthesis (AlgZ), hemerythrin (Hr), bacterial hemoglobin (HbO and Hmp), and cytochrome b may play a role in O_2 protective strategies used by *Cand.* T. haligoni during active N_2 fixation (fig. S11). Intracellular alginate serves dual roles in reducing O_2 by (i) promoting cell aggregation to create suboxic environments and (ii) impeding O_2 diffusion across the cell [fig. S2B; (*44*–*46*)]. We speculate that intracellular O_2 levels are likely then regulated by cytochrome b, a terminal oxidase in adenosine triphosphate (ATP) generation, and by the O_2-binding proteins, bacterial hemoglobin, and hemerythrin (*47*–*49*). The hypothesis for multiple O_2 protective strategies is further supported by a preliminary experiment where the level of these proteins in *Cand.* T. haligoni was up-regulated by one order of magnitude in a culture where the N_2 fixation machinery reached 10% of the protein mass fraction (data S4).

Nitrogenase activity may provide multiple metabolic benefits to *Cand.* T. haligoni

Current theories on the evolution of the nitrogenase enzyme suggest that abiotic sources of fixed N were likely abundant for early life in a primitive reducing atmosphere. As life diversified, the finite supply of fixed N could not support increasing demands as microorganisms increased in biomass (*50*), leading to the development of N_2 fixation as a means to meet the cellular N demand of an expanding biome in a dwindling fixed N supply (*51*, *52*). In the modern ocean where fixed N is a limiting nutrient for growth and cellular function of primary producers in large areas of the surface ocean (*1*), nitrogenase presently remains the only known enzyme complex that has evolved to alleviate N limitation in the presence of an oxygenic atmosphere (*52*, *53*). While the importance of N_2 fixation is recognized, the various metabolic strategies to balance homeostasis, nitrogenase expression, and O_2 protection remain enigmatic, especially in the widely distributed γ-proteobacterial diazotrophs.

The detection of several regulatory proteins in the proteome responsive to redox balance and N starvation suggests that nitrogenase expression may play a role in alleviating environmental stress, beyond N limitation within the isolate. This phenomenon has been observed in some terrestrial diazotrophs like *Azotobacter*, *Rhodopseudomonas*, and *Rhizobia* sp., to uphold cellular homeostasis and redox balance within the cell (*54*–*60*). As most research focuses on terrestrial HBDs, very few analyses have examined the function of nitrogenase in marine HBDs. The regulatory proteins recovered from the proteome of *Cand.* T. haligoni suggests two pathways for N_2 fixation regulation. The first pathway is common to most diazotrophs, concerning the regulation of N_2 fixation by the N-sensing NtrC/NtrB and the ammonium transporter AmtB signaling N starvation within the cell, and initiating a cascade of N_2 fixation

regulatory proteins starting with GlnB (Fig. 4E: N_2 fixation—step 2). The second predicted pathway governs cell redox balance similar to that of *Rhodobacter* and *Rhodopseudomonas* through the protein pair RegS and RegR (*43*, *61*). Upon redox potential imbalance, the enhancing binding protein, RegR, directly engages with RNA polymerase (σ^{54}) on *nifA*, initiating transcription of the *nif* regulatory protein [Fig. 4E: Redox—steps 2 and 3*;* (*61*)]. As our proteome analysis of nitrate-grown stationary cultures suggests that ~0.5% of the proteome (Fig. 4C) was allocated to nitrogenase and its associated proteins for diazotrophy, we hypothesize that these low cellular levels of nitrogenase may either have served to maintain redox balance or, alternatively, condition the cell for a rapid nitrogenase response to N limitation (Fig. 4E). The results from nitrate grown cultures (Fig. 4) are in contrast with a preliminary proteome obtained from N_2-grown cells (data S4), where the N_2 fixation machinery allocation was similar to that of *Pseudomonas stutzeri* and also within the range of the reported protein mass fraction for diazotrophic cyanobacteria actively fixing N_2 [(*62*–*64*)*;* text S2 and fig. S12]. While further investigation with the isolate is needed, additional proteins involved in potential N_2 fixation regulation (e.g., CbrA and CrcB) also increased in relative abundance under fixed N-free conditions [data S4; (*64*)]. This basal level of nitrogenase in HBDs may be more common than previously thought (*43*, *55*, *57*); however, the sparsity of proteomics data for HBDs is limiting the comparison with our isolate.

Cand. T. haligoni is widely distributed in global oceans

Within the few cultured marine γ-proteobacterial diazotrophs, *Cand.* T. haligoni is so far the only known member of this group with an observed global distribution (*17*, *20*, *23*). The isolate's *nifH* ASV signature has been recovered from polar, temperate, subtropical, and tropical regions. A strain-specific *nifH* qPCR assay also detected the isolate throughout the North Atlantic Ocean in coastal and open ocean habitats, in both surface and deep waters with varying concentrations of DIN and measurable NFR in overlapping regions (i.e., GEOVIDE; Geo-2, 13, 21) [figs. S5 and S7 and table S4; (*65*)]. NFR measurements and *nifH*-derived taxonomic information obtained from the GEOVIDE cruise demonstrated that γ-proteobacterial diazotrophs contributed 28% of the total *nifH* amplicon sequences (21,001 total reads), with a range in relative abundance between 4% and 98% across the GEOVIDE stations [see figures 5 and 6 from (*65*)]. Together, the above data thus highlight the importance of γ-proteobacterial diazotrophs' (including *Cand.* T. haligoni) contribution to the N cycle and the need to understand the environmental controls on their mode of diazotrophic growth.

Given that there are currently few direct measurements of cell densities for HBDs (*31*), TaqMan qPCR assays have been used to estimate the relative abundance of the dominant marine HBDs, despite the caveats associated with this approach (*11*, *20*). *NifH*-based qPCR abundance estimates of *Cand.* T. haligoni are higher than most other HBDs measured using similar qPCR-based approaches (*17*, *18*, *66*) and often comparable to those of the oligotrophic, surface-dominant Gamma-A (*18*), albeit with a different geographic distribution (Fig. 5A). In addition, *Cand.* T. haligoni *nifH* gene is even detected at levels comparable to those of the newly identified nitroplast of *Braarudosphaera bigelowii*, formally known as UCYN-A (approximately 10^6 *nifH* copies liter^{-1}), hence demonstrating its importance in the pelagic environment (*7*, *67*, *68*). Because of its versatile genome, ability to thrive on diverse carbon and N compounds, widespread distribution, and high *nifH* copy numbers, *Cand.* T. haligoni appears to be an important cultured representative for the exploration of the γ-HBD clade identified in our phylogenomic tree (Fig. 2).

Distribution patterns showing high coastal abundances in HBDs have been observed before and have been attributed to the presence of high concentrations of labile dissolved organic carbon (DOC) as fuel for heterotrophic bacteria (*30*). Marine planktonic aggregates are often suggested as potential habitats for HBDs, creating microenvironments favorable to N_2 fixation in waters, which would otherwise suppress nitrogenase activity [e.g., high dissolved O_2 and DIN concentrations; (*31*, *69*–*72*)]. During the GEOVIDE cruise, *Cand.* T. haligoni *nifH* qPCR copies liter^{-1} were highest during post-bloom periods in the North Atlantic Subtropical Gyre, and the Irminger and Labrador Sea—where DOC concentrations were also found to be highest [fig. S4A; (*73*)]. The *nifH* gene copies liter^{-1} from the coastal time series displayed similar patterns (fig. S6). The highest nifH copy detection (10^7 *nifH* copies liter^{-1}) occurred shortly after spring and fall phytoplankton blooms, with the isolate's presence detectable at all depths. The peak values seen in our time series data were not specifically associated with the large (>3 μm) or small (0.2 to 3 μm) size fractions, indicating that attachment to planktonic aggregates or self-aggregation (fig. S2) may be facultative, in contrast to Gamma-A (*74*). With the detection of *Cand.* T. haligoni throughout the water column and in regions where high remineralization is present, we suggest that the isolate acts as a key microbial player not only for DOC remineralization but also for supplementing fixed N to other phytoplankton and bacteria within these regions. The significance of *Cand.* T. haligoni appearing in higher abundances post-bloom is reminiscent of the colonization of young plants by the terrestrial HBD strain, *P. stutzeri* A1501, where abundances increase with older plants as DOC exudates become more complex and concentrated (*64*). By analogy, as the phytoplankton bloom progresses, more DOC exudates are released as N limitation sets in, making conditions more favorable for *Cand.* T. haligoni (*75*, *76*). Since various carbon compound mixtures and concentrations can elicit different N_2-fixing capability and growth behavior in HBDs, examining the relationships to phytoplankton DOC will be imperative for future culturing efforts and understanding the regulation of nitrogenase (*64*, *76*). Similarly, the observations of *Cand.* T. haligoni at depths greater than 1000 m suggest that this HBD may form aggregates and sink, contributing to the recently identified role of diazotrophs in carbon export to the deep sea (*77*, *78*).

Contribution of *Cand.* T. haligoni to the marine N cycle

In our experiments, NFR measurements obtained for *Cand.* T. haligoni are comparable to those of the marine α-proteobacterium, *Sagittula castaena P11* [~1×10^{-2} fmol N cell^{-1} day^{-1}] (*79*). This is in contrast to higher rates reported for *P. stutzeri* and *Rhodopseudomonas* sp. (*17*, *80*) and much lower rates (two orders of magnitude) for *Vibrio natriegens* [Fig. 4D; (*70*)]. However, reported NFR measurements for cyanobacterial diazotrophs are typically one to three orders of magnitude higher than any HBD rates published to date, in agreement with our results (*17*, *20*, *79*–*81*). Several hypotheses could explain these pronounced NFR differences, such as (i) cyanobacteria potentially outcompete HBDs due to their ability to use light energy, (ii) the enzymatic activity of nitrogenase varies between diazotrophic species occupying different temperature and nutrient regimes, and (iii) the nitrogenase enzyme might serve

alternative functions in HBDs, resulting in lower rate measurements when engaged in distinct metabolic roles (*61*, *66*, *82*–*85*). While NFR measurements from other HBD cultures were taken during the exponential growth phase, the results for *Cand.* T. haligoni presented here were obtained in stationary phase, where NFR may be lower than in exponential growth. As the relationship between NFR measurements and cell growth phase appears to be species specific among HBDs (*86*, *87*), further experimental work is needed to better assess NFRs that are more characteristic of the in situ lifestyle for this microorganism. In general, NFR measurements of *Cand.* T. haligoni fell within a mid-range compared to other cultured HBDs, and we predict that higher rates may be achieved in the absence of fixed N (data S4). While HBDs exhibit lower NFR per cell than cyanobacterial diazotrophs, their persistent presence in the water column and ability to fix N_2 irrespective of the surrounding fixed N concentrations underscore their significance in the N budget. Given the global distribution of *Cand.* T. haligoni, the low but sustained NFRs shown here thus warrant a re-evaluation of biological N_2 fixation input from HBDs into the N pool [(*6*); data S4]. Our study demonstrates that *Cand.* T. haligoni is present in the ocean at abundances similar to the uncultured and widely distributed Gamma-A and Gamma 4, and also to the haptophyte-associated complex UCYN-A. Given *Cand.* T. haligoni's abundance, cosmopolitan distribution, cultivability, genome availability, and representation in MAGs from the Arctic Ocean, we propose *Cand.* T. haligoni as a model marine diazotrophic representative for γ-proteobacterial HBDs, particularly from the Oceanospirillales clade. Leveraging the isolate's cultivation can allow for an in-depth exploration of the ecophysiology of Arc-Gamma-03 and identify strain-specific variations between the two genomes. Notably, the isolate defies current knowledge on the regulation of N_2 fixation, exhibiting constant N_2 fixation and Nif protein expression regardless of prolonged exposures to fixed N concentrations. Further work is needed to explore an alternative role for nitrogenase, such as a redox balancing hub, as previously proposed for terrestrial HBDs. As N_2 fixation in *Cand.* T. haligoni can take place in the presence and absence of nitrate, we propose that the Oceanospirillales clade of diazotrophs discussed here may be a major contributor to the global marine N_2 fixation budget.

MATERIALS AND METHODS

Isolation and genome characterization

Water samples were collected in January 2014 at 1, 5, 10, and 60 m depths from the Compass Buoy station in Bedford Basin (Halifax, Nova Scotia, Canada; 44°41′30″N, 63°38′30″W); the station, home to a long-term oceanographic time series (*88*), is located in the inner portion of the Halifax Harbour, surrounded by a large urban population center and connected to the main harbor through a shallow, narrow sill. The upper 30 m of the 71-m water column generally encompasses the photic zone (fig. S1). Natural seawater samples were aliquoted into 30-ml culture flasks and enriched with nutrients, except for fixed N according to table S5. To simulate a representative environment, samples from the photic zone (1, 5, and 10 m) were incubated with a 12-hour light/dark cycle (1 and 5 m at 12°C and 10 m at 5°C), while the 60-m sample was incubated at 4°C in the dark. Incubated samples were re-enriched monthly with the nutrient amendments as listed in table S5, and the presence of diazotrophs was monitored using nested *nifH* PCRs as described [see the "*NifH* community amplicon sequencing and qPCRs" section below; (*89*)]. A sample from 10 m enriched with phosphate and iron tested positive for the presence of *nifH* genes. Single bacterial cells from this sample were sorted in a grid on 1.2% agar plates prepared with N-free f/2 ASW (*90*) using a BD Influx cell sorter. Plates were incubated at 5°C (12-hour light/dark cycles) for 90 days. Single colonies that grew on the N-free seawater agar plates (see fig. S2A) were spread as mini-lawns on agar plates of the same composition and allowed to grow for several days before they were screened again for the presence of the *nifH* gene. Positive colonies were then transferred to diazotroph-selective YBC-II agar plates supplemented with 15 mM sodium acetate as a carbon source. The purity of the bacterial isolate was confirmed by sequencing of both the *nifH* gene and the 16S rRNA gene amplicons.

For short read sequencing, the bacterial DNA was extracted from 100 ml of prescreened *nifH*-positive culture using an AllPrep DNA/RNA Mini Kit (QIAGEN). Modifications included (i) resuspending the cell pellet in 50 µl of lysozyme (20 mg/ml) and incubating for 5 min at room temperature, (ii) adding 45 µl of proteinase K (20 mg/ml) and 600 µl of RLT buffer with 10 µl of β-mercaptoethanol and incubating for 15 min at 52°C, and (iii) eluting the DNA in 50 µl of AE buffer and RNA in 50 µl of ribonuclease (RNase)–free water. DNA for long-read sequencing was extracted using QIAGEN's Genomic-tips 20/G and genomic DNA buffer set according to the manufacturer's protocol using 10-ml culture aliquots that were subjected to three cycles of −20°C for 3 min and then 100°C for 5 min. DNA concentrations for short and long read sequencing were measured using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific). DNA size and integrity for long read sequencing were further verified by agarose gel electrophoresis, and a final concentration was determined using a Qubit 4 fluorometer (Invitrogen).

Genomic sequencing was conducted at the Integrated Microbiome Resource (IMR) at Dalhousie University (NS, Canada) using both an Illumina MiSeq and a Nanopore MinION device (*91*). Nanopore sequencing was conducted for both short and long reads to fully close the genome. The Illumina sequencing library was constructed using the Illumina Nextera XT kit with 5 to 10 ng of DNA extracted for short read sequencing, dual-indexed, and then run on a shared MiSeq run using 600-cycle v3 chemistry (300 + 300 bp paired-end). The first run on the MinION involved 1 µg of DNA extracted for short read sequencing that was sheared using a Covaris g-TUBE [5000*g* for 1-min + 1-min recovery spin; (*91*)] to obtain consistent DNA size. The short-read library was constructed using the Nanopore SQK-MAP006 kit and run on an Mk1 MinION device using an R7 flow cell for 48 hours. The second MinION sequencing run used the SQK-RAD003 protocol following the manufacturer's instructions with unsheared DNA extracted for long-read sequencing. The final hybrid genome assembly was constructed with combined Illumina and MinION reads using Nanopolish and the Canu Assembler [https://github.com/jts/nanopolish (*92*)]. The fully closed genome was deposited to the National Center for Biotechnology Institute (NCBI) and can be found under submission number PRJNA1046103.

Scanning and transmission electron microscopy

Cultures grown without fixed N in f/2 ASW were pelleted at 15,000*g*. The cell pellet was resuspended and fixed in 2.5% glutaraldehyde diluted with 0.1 M sodium cacodylate buffer overnight at 4°C. Samples for scanning electron microscopy (SEM) and transmission electron

microscopy (TEM) were processed at the Electron Microscope Facility and the Scientific Imaging Suite of the Biology Department, respectively, both at Dalhousie University (NS, Canada). For SEM, samples were pipetted onto poly-L-lysine–coated coverslips and allowed to adhere for 20 min. Samples were then rinsed with 0.1 M sodium cacodylate buffer (three 10-min washes) and secondarily fixed with 1% osmium tetroxide for 2 hours followed by a distilled water rinse. Dehydration occurred in a graded series of ethanol (1 × 10 min at 50%, 2 × 10 min at 70%, 2 × 10 min at 95%, and 2 × 10 min at 100%), left to dry for an additional 10 min, and then placed in a Leica CPD3000 Critical Point Dryer. Dried samples were attached to SEM stubs using a double-sided carbon tab and coated with 10-nm gold/palladium mixture (80/20) in a Leica EM Ace 600 sputter coater before analysis on a Zeiss Sigma 300 SEM (SE2 detector at 5 kV). For TEM, the samples were sequentially rinsed three times with 0.1 M sodium cacodylate buffer, fixed for 2 hours with 1% osmium tetroxide, rinsed with distilled water, and stained with 0.25% uranyl acetate at 4°C overnight. The samples were then dehydrated with a graduated series of acetone/H_2O mixtures: 50% acetone (10 min), 70% acetone (2 × 10 min), 95% acetone (2 × 10 min), 100% acetone (2 × 10 min), and anhydrous acetone (10 min). Then, the sample was gradually infiltrated in Epon-Araldite resin, sequentially as 3 parts anhydrous acetone to 1 part resin (3 hours), followed by 1 part anhydrous acetone to 3 parts resin (overnight), and finally with 100% Epon-Araldite resin (2 × 3 hours). After embedding in 100% Epon-Araldite resin, the sample was incubated at 60°C for 48 hours to cure. Ultrathin sections were cut using a Reichert-Jung Ultracut E Ultramicrotome with a diamond knife and placed on 300-mesh copper grids. Staining was performed as follows: 2% aqueous uranyl acetate (10 min), two rinses with distilled water, lead citrate (4 min), rinse with distilled water, and air-dried. Samples were viewed using a JEOL JEN 1230 transmission electron microscope at 80 kV, and images were captured using a Hamamatsu ORCA-HR digital camera.

Phylogenomic and phylogenetic tree construction

For the phylogenomic tree, genome sequences were retrieved from NCBI and MAGs from the *Tara* Oceans project (*21*). Table S6 provides a name comparison of assigned names from *Tara* Oceans to those used in this study to provide consistency across studies (*21*). A multigene alignment was prepared in Anvi'o v7.8 (*93*) using diazotroph genomes and close relatives to the isolate. Contigs for each genome were organized into an Anvi'o contigs database using anvi-gen-contigs-database. Genome contigs were then scanned using hidden Markov models (HMMs) for the Bacteria_71 collection of single-copy marker genes using the command anvi-run-hmms (*93*–*96*). The gene markers from each selected genome were concatenated and aligned using the command anvi-get-sequences-for-hmm-hits. FastTree was then used to generate a maximum likelihood tree from the aligned concatenated protein sequences using the Jones-Taylor Thorton model and was subjected to 1000 bootstraps (*97*). The *nifH* maximum likelihood phylogenetic tree was constructed in MEGA X (*98*) using a collection of 166 *nifH* sequences aligned via MUSCLE [codon; (*99*)], the Kimura two-parameter model (*100*), and 1000 bootstrap replicates. For this tree, evolutionary rate differences were also subjected to a discrete Gamma distribution. Both trees were visualized in iTOL (*101*). Phylogenomic and *nifH* phylogenetic trees were constructed with genomes and *nifH* sequences that shared >95% identity with *16S* rRNA and *nifH* genes of *Cand.* T. haligoni, respectively.

ARC-Gamma-03 MAG re-assembly and pangenome analysis

The Arctic metagenomes from the TARA project (*102*) were reprocessed, binned, and reanalyzed (*32*). One of the bins assembled in this process showed >99% identity with the *Cand.* T. haligoni genome. The genome of *Cand.* T. haligoni was therefore used as a reference to recruit metagenomic reads from the TARA arctic metagenomes. Briefly, 110 paired metagenome samples processed in (*32*) were downloaded from the NCBI SRA database and mapped to the genome of *Cand.* T. haligoni using bowtie2 (*103*). The mapped reads were extracted and assembled using megahit (*104*) with a minimum contig length of 1000 bases. The 223 contigs assembled by megahit were scaffolded using the program ragtag (*105*), resulting in 137 scaffolds (with a total length of 4,055,874 bases). The program prodigal (*106*) was then used to identify genes from the scaffolds. The genes were compared to the KEGG, COG, PFAM, and KOfam databases to annotate functions (*107*–*109*).

The pangenome analysis was conducted using Anvi'o v8 and workflows therein (*93, 110*). The contigs database for each genome was populated with functional annotations according to NCBI's COG database, the KOfam database of KEGG orthologs, and the Pfams protein family database (*107*–*109*). Taxonomy estimations were done using "anvi-run-scg-taxonomy" and "anvi-estimate-scg-taxonomy," which use the Genome Taxonomy Database (GTDB) and Diamond (*111, 112*). Next, a genome database was created, combining all the contigs databases for each genome, and the command anvi-pan-genome was used for pangenome computation with parameters "-minbit 0.5," "-mcl-inflation 10," "--use-ncbi-blast." ANI scores were calculated using anvi-compute-genome-similarity with PYANI (v0.2.7), and the pangenome was visualized using "anvi-display-pan" (*110, 113*).

Environmental DNA sampling and oceanographic data

Environmental DNA (eDNA) samples used for *nifH* amplicon sequencing and quantitative PCR (qPCR) assays were collected during field studies conducted on the Scotian Shelf during an Atlantic Zone Monitoring Program (AZMP) expedition [2014] (*114*), a GEOVIDE (GEOTRACES GA01) transect across the North Atlantic [2014], and a Canadian ArcticNet (GEOTRACES, GN03) transect through Baffin Bay and the Canadian Arctic Archipelago [2015]. Sample depths varied for each expedition but consisted of surface (<20 m) and deep water (>20 m) samples (see data S3A for depth details on each expedition). DNA samples were also collected from the Bedford Basin Monitoring Program (BBMP) time-series site weekly at 1, 5, 10, and 60 m throughout 2015–2019 [see (*115, 116*) for more details on this eDNA time series]. Briefly, 0.5 liter of seawater from each depth was filtered through 0.2-μm Isopore polycarbonate filters (Millipore; 2015–2017) or size-fractionated through 3- and 0.2-μm filters (2018–2019). Filtered seawater samples were maintained at −80°C until DNA extraction as described in (*115, 116*). Chlorophyll a, temperature, salinity, oxygen, and DIN for BBMP are available from the Bedford Institute of Oceanography (accessible at: https://www.bio.gc.ca/science/monitoring-monitorage/bbmp-pobb/bbmp-pobb-en.php). Oceanographic data from research expeditions (AZMP, GEOTRACES, and GEOVIDE) were obtained from Conductivity, Temperature, and Depth (CTD) sensors and are publicly available from datasets referred to in (*114*),

IDP2021 (GEOTRACES Intermediate Data Product Group, 2021), and https://doi.org/10.17600/14000200 (GEOVIDE). DIN concentrations were calculated as the total of $NO_3^- + NO_2^- + NH_3^-$.

NifH community amplicon sequencing and qPCRs

eDNA samples were subjected to *nifH* amplicon sequencing and qPCR analysis for in situ detection of *Cand.* T. haligoni. Nested *nifH* PCRs were used to amplify a portion of the *nifH* gene (327 nt) using primers *nifH3*/*nifH4* 5′-ATRTTRTTNGCNGCRTA-3′/5′-TTYTAYGGNAARGGNGG-3′ and *nifH2*/*nifH1* 5′-ADNGCCATCATYTCNCC-3′/5′-TGYGAYCCNAARGCNGA-3′ (*12*, *117*). Nested PCRs followed the same protocol as (*89*). Positive *nifH* samples were then subjected to next-generation sequencing on an Illumina MiSeq instrument following preparation methods listed in (*116*) at the IMR, Dalhousie University, Halifax, NS, Canada. *NifH* sequences were processed using QIIME 2 version 2019.7 (*118*) according to a modified protocol from (*119*). Primer sequences for *nifH1*/*nifH2* were removed using cutadapt (*118*) and denoised to ASVs using deblur (*120*) with a trim length of 325 nt. ASVs with an abundance of <0.1% were removed to account for MiSeq bleed-through. ASV taxonomies were assigned using a classifier trained from sequences collected in (*121*). *NifH* community sequences derived in-house are deposited under NCBI Bioprojects PRJNA930772 (BBMP), PRJNA325151 (AZMP), and PRJNA931255 (GN03). The sequence read archive (SRA) datasets used in this study can be found in data S5. These data were compiled through literature searches for *Cand.* T. haligoni, and nucleotide alignments of the isolate to the ASVs were generated from NCBI SRAs. Specifically, SRAs that contained *nifH* amplicon sequencing from the marine environment were downloaded and processed individually through QIIME 2 version 2019.7 following the same parameters as above. The *nifH* sequence of *Cand.* T. haligoni was then aligned against reference ASV *nifH* sequences generated from the SRAs to confirm publicly derived samples with the isolate (data S5). *NifH* ASVs matching ≥98.45% to *Cand.* T. haligoni were then traced back to the original SRA for the number of associated reads and metadata (*120*, *122*). ASVs identified as similar to *Cand.* T. haligoni and their accompanying read data were downloaded and used for later ecotype variant analysis (text S3).

TaqMan qPCRs were conducted on eDNA samples using primers and a probe specific to *Cand.* T. haligoni: forward 5′-AGCCCGGTGTTGGTTGTG-3′, reverse 5′-CGAGGAAGTTGATCGCTGTGA-3′, and MGB probe 6FAM-5′-TCGCGGTGTCATCACAGCGATCA-3′ [Applied Biosystems; (*91*)]. BLAST searches were conducted for primer and probe to ensure no other detection apart from the isolate (*91*). Tenfold dilution standard curves (1 to 10^7 *nifH* gene copies per reaction) were generated in triplicate using a double-stranded gBlock DNA fragment (250 nt) 5′-ATCATGGAGATGGCTGCTGAAGCGGGCACTGTGGAAGATCTGGAACTGGAAGACGTACTGAAAAGTGGTTACGGAGACATCCGTTGCGTTGAGTCCGGTGGTCCAGAGCCCGGTGTTGGTTGTGCCGGTCGCGGTGTCATCACAGCGATCAACTTCCTCGAAGAAGAAGGTGCTTACGAAGACGATCTGGACTTCGTGTTCTACGACGTACTGGGTGATGTTGTGTGTGGTGGTTTCGCCATGCCGATTC-3′ (IDT). gBlock concentration was measured on a Qubit 4 fluorometer before each use via a 1× HS dsDNA kit (Invitrogen). Each 16-µl qPCR contained 8 µl of 2 × TaqMan Universal PCR master mix (Applied Biosystems), 0.64 µl of 6FAM probe (0.1 µM final), 0.16 µl of forward primer (1 µM final) and 0.48 µl of reverse primer (3 µM final), 0.32 µl of bovine serum albumin (20,000 ng/µl; NEB), 1.4 µl of UltraPure water (Invitrogen), and 5 µl of template [either eDNA, gBlock DNA standard, or no-template control (water)]. Reactions were run on a ViiA7 real-time PCR using QuantStudio software (Applied Biosystems) using 45 cycles of 10 min (95°C), 15 s (95°C), and 1 min (60°C). Assay efficiencies were calculated using LinRegPCR V.11 (98% ± 8% SD), and the limit of quantification (LOQ) for each transect can be found in data S1 (B and C) (*123*, *124*).

Culture experiments with nitrate as a fixed N source

Culture experiments shown in fig. S1 were grown in 50-ml nonvented culture flasks with 220 µM NO_3^- at 15°C with 12-hour light/dark cycles. Cultures were sampled every 12 hours for cell density and nutrients (NO_3^-, NH_3^-, NO_2^-, and PO_4^{3-}). More detailed methods can be found in text S1. Growth rates and maximum biomass were compared across treatments to determine which carbon substrate and concentration provided the fastest growth and highest biomass. The favored substrate was then used for culture experiments below.

Samples used for shotgun proteomics and NFR in Fig. 4B were collected from sacrificial cultures that were in stationary phase under prolonged NO_3^- exposure. Nitrate concentrations in sampled culture bottles decreased over time across a 10-day sampling period. Sampling time points were conducted 3 days after inoculation [high nitrate "HN" (100 µM)] and 9 days [low nitrate "LN" (2 µM)]. New additions of NO_3^- were added (100 µM) after 9 days and cells were harvested 1 hour after NO_3^- addition (NN), and the medium nitrate "MN" (50 µM) treatment was sampled 12 hours later. The re-addition of NO_3^- was done to follow the effects of a fixed N source resupply on nitrogenase expression in the isolate. To ensure no interference from the heavy $^{15}N_2$ isotope in proteomic measurements, a total of nine biological replicates were used for each condition: three for proteomic work and six for NFR measurements. For all treatments, 285-ml cultures were grown in triplicate 300-ml polycarbonate bottles in f/2 medium prepared in ASW with a modified initial nitrate concentration of 220 µM NO_3^- (6:1 N:P) + 6.4 mM carbon (fumarate, 99% purity; Thermo Fisher Scientific) with 285 µl of resazurin dye (Alfa-Aesar; 1:1000 final dilution; data S6). Sacrificial culture bottles were weighed before and after adding inoculum and medium to identify any small variation in volume. The culture medium was filter-sterilized using a 0.2-µm Sterivex filtration unit (Millipore) and added to pre-sterilized incubation bottles. An inoculum of the pure culture was added to each flask (except blanks that contained no cells) to reach an initial cell density of 400 cells/µl within each bottle. Bottles were fitted with septum caps and stored at 15°C with 12-hour light/dark cycles. Cell density, biomass collection, and nutrient samples were taken every 12 hours.

Cell density was measured on an Accuri C6 flow cytometer (BD) using 1:1000 SYBR Green II dye (Invitrogen). Duplicate nutrient samples were filtered (0.2 µm) into 15-ml centrifuge tubes and stored at −20°C until later analysis on a Skalar San^{++} nutrient analyzer.

NFR measurements with cultures grown with nitrate as a fixed N source

NFR measurements were performed every 24 hours using the $^{15}N_2$ dissolution method on cultures as mentioned above (*125*, *126*). Enriched $^{15}N_2$ seawater and natural abundance (NA) medium were prepared from ASW identical in salt composition to the growth medium used for the cultures. Details concerning the preparation and

application of the $^{15}N_2$ method are similar to those of (*127*) with the exception that ASW was not degassed before the $^{15}N_2$ gas spike to limit changes to the water matrix. At the start of each incubation, 15 ml of enriched medium (5% final culture volume; 300 ml total volume) was added to each designated harvest bottle ensuring that no headspace remained. Cultures were sealed with septum caps and incubated at 15°C for 24 hours. Triplicate exetainers were sampled from each sacrificial culture of the NA condition to determine background N_2 pool within samples. After exetainer collection, samples were filtered onto 0.3-μm pre-combusted glass fiber filters (Advantec), dried for 12 hours at 50°C, and stored in the dark at room temperature until later analysis. Samples were analyzed with a Vario Micro Cube elemental analyzer (Elementar Americas) coupled to an Isoprime 100 Isotope Ratio Mass Spectrometer (Elementar Americas).

N content and isotope composition of each sample was calculated using standard curves prepared with in-house standard compounds calibrated against international reference materials USGS 40 and USGS 41 [United States Geological Survey (USGS)]. In-house standards for both NA and $^{15}N_2$-enriched particulate samples included L-glutamic acid (99% purity, Sigma-Aldrich), glycine (99% purity, VWR), and nicotinamide (99.5% purity, Sigma-Aldrich). The enriched certified reference material IAEA 305B (δ15N 375.3%) was used to calibrate the analysis of enriched samples only (*128*). Error rates, limits of detection (LODs), and LOQ were calculated according to (*129*) and data S7. LOD was determined using three times the SD of the isotopic composition of the corresponding NA sample for each experiment, while LOQ was determined using 10 times the same value. NFRs were calculated by measuring the isotopic tracer enrichment within the particulate material using the following equation

$$N_2\text{fixation rate(NFR)} = \frac{APN_{final} - APN_o}{AN_2 - APN_o} \times \frac{[PN]}{\Delta t} \quad (1)$$

where APN_{final} and APN_o represent the ^{15}N-Atom% in the particulate nitrogen (PN) incubated with and without the ^{15}N-N_2 tracer, respectively. AN_2 represents the ^{15}N-Atom% dissolved in seawater. Δt is the length of time for the incubation.

Protein extraction, digestion, and liquid chromatography–tandem mass spectrometry

Protein was extracted as described in (*130*) with slight modifications (text S4). Protein extracts were digested with trypsin using S-traps (Protifi) with a modified procedure (texts S4 and S5). Once desalted, samples were solubilized in 1% formic acid and 3% acetonitrile and analyzed by reverse-phase liquid chromatography–tandem mass spectrometry (LC-MS/MS). For each sample, 1 μg (3 μl) of total peptide was directly injected onto a 100 μm × 40 cm column containing 4-μm, 90-Å, Proteo C18 beads (Phenomenex, Torrance, CA), self-packed in a fused silica emitter tip (New Objective, Woburn, MA). Chromatography was performed using a Dionex Ultimate 3000 UHPLC (Thermo Fisher Scientific, San Jose, CA) at a flow rate of 0.25 μl/min. Peptides were separated using a gradient of 5 to 20% B over 57 min, then 20 to 50% B over 8 min, followed by 4 min at 95% B, where B solvent was 0.1% formic acid in acetonitrile and A solvent was 0.1% formic acid in high-performance liquid chromatography (HPLC)–grade water. Column outflow was interfaced to an Orbitrap LUMOS Tribrid with a Nanospray Flex ion source and FAIMS Pro ion mobility filter (Thermo Fisher Scientific, San Jose, CA). The ion spray voltage was 1800 V, and the ion transfer tube was 300°C. FAIMS was operated with standard resolution and a total carrier gas flow of 4.6 liters/min. Three FAIMS compensation voltages were applied over a fixed 3-s cycle time, 1.2 s at −50 V, 1.2 s at −65 V, and 0.6 s at −80 V. MS^1 scans were performed using the Orbitrap with a 400 to 1600 mass/charge ratio (*m*/*z*) mass range and 120,000 resolution (Profile mode, AGC target 100%—4 e5, Max IT 50 ms—Auto, RF lens 30%). Data-dependent MS/MS scans were acquired in parallel in the iontrap (Centroid mode, ACG target 100%—1 e4, Max IT 35 ms—Dynamic, Rapid scan rate) with a 1.6-Da quadrupole isolation window and high-energy collisional dissociation (HCD) fragmentation with 30% collision energy. The minimum intensity threshold for MS^2 selection was 5000, and ions were selected from most to least intense.

Protein identification and quantitative analysis

Raw spectral files from MS were processed using Proteome Discoverer (V.2.5) (Thermo Fisher Scientific, San Jose, CA). SEQUEST HT searches were performed against the genome of *Cand.* T. haligoni and combined with a list of common protein contaminants from UniProt (common Repository of Adventitious Protein). The precursor (MS^1) and fragment (MS^2) mass tolerance were 10 parts per million (ppm) and 0.6 Da, respectively. Decoy database searches were performed, and a strict false discovery rate of 1% and maximum ΔCn of 0.05 were applied and validated using Percolator (*131*). Precursor peak area quantification was performed using the Minora feature detector. Because of a low number of quantified peptides in two replicates (samples 29 and 32), these samples were removed from downstream analysis. All unique peptides matching identified proteins and razor peptides were used for quantification. Data are available at ProteomeXchange (accession number: PXD045743). Scripts to reproduce the dataset analyzed are available at: https://github.com/bertrand-lab/.

Protein normalization and functional group determination

Protein abundance was calculated by taking the sum of MS^1 peak areas for all unique peptides plus razor peptides matching to that protein. Individual protein abundances were normalized to total protein abundance (sum of all protein abundances) for each sample. Quantification for detected proteins can be found in data S2D. The row z score was calculated for each protein across the four treatments (Fig. 4A). The proteomic mass fraction displayed in Fig. 4C was calculated by taking the sum of all peptide MS^1 peak areas specific to the category of interest (i.e., N_2 fixation) and dividing by the sum of all peptide assigned MS^1 peak areas in the sample (*132*).

Statistical analysis and data visualization

All graphical and statistical analyses were conducted in R version 4.2.2 via R Studio 2022.7.2 (R Core Team; RStudio Team 2020) with the ggplot2 package therein to generate graphs (*133*). RDA was conducted on qPCR values of *Cand.* T. haligoni with corresponding oceanographic measurements (temperature, oxygen, chlorophyll, nitrate, and salinity) to assess for any environmental sensitivity (data S3A). R packages for RDA analysis included BiocManager, Tidyverse, and Vegan (*134*–*136*). Both qPCR and oceanographic data were first standardized using decostand in Vegan (method = standardize) (*136*). The envfit function in vegan was used to fit the environmental vectors onto the ordination plot with 999 permutations (data S3B). Pairwise comparisons were conducted between the

relative abundances of individual proteins for each treatment in Fig. 4A, and statistical values were adjusted using the Benjamini-Hochberg correction ($\alpha = 0.05$; data S2E). NFR measurements were subjected to a Kruskal-Wallis test ($\alpha = 0.05$) using the rstatix package (*137*).

Spatial distribution maps were made using Ocean Data View 5.6.3 using qPCR values and oceanographic data (Ocean Data View, https://odv.awi.de). To assess the dominance of the isolate compared to other γ-proteobacterial diazotrophs, *nifH* amplicon sequences were collected from the same database compiled for isolate detection (see the "*NifH* community amplicon sequencing and qPCRs" section). ASV identity cutoffs included an *E* value of −150 and a bit score of 600. ASVs considered related to the isolate included those with a percent identity >98.45%. These cutoff values were selected to account for any single-nucleotide polymorphisms and potential PCR amplification errors. Dominant γ-proteobacterial *nifH* ASVs were identified as those ASVs that (i) were not singletons, (ii) had a total of at least 50 reads across all 865 surface water samples (0 to 30 m; internal and SRA database), and (iii) were present in at least 80 stations. Relative abundances were calculated for the entire dataset, and all γ-proteobacterial *nifH* ASVs were subset from these data. To avoid overplotting of surface stations, only depths with the highest relative abundance of γ-proteobacteria ASVs were plotted. Maps of dominant ASV relative abundances and their geographic distributions were generated with ggplot2 using the geom_sf() command and the following packages: rnaturalearth, cowplot, naturalearthdata, Scatterpie, and ggspatial (*138*–*141*).The general genome schematic and nitrogenase regulation pathway of the isolate were created using BioRender (www.biorender.com). Proteome mapping of *Cand.* T. haligoni was annotated using CG viewer and the Proksee server (*142, 143*).

Note added in proof: We would like to acknowledge the recent findings concerning gamma-A published in B. Tschitschko *et al*., Rhizobia–diatom symbiosis fixes missing nitrogen in the ocean. *Nature* **630**, 899–904 (2024).

Supplementary Materials

This PDF file includes:

Supplementary Texts S1 to S5
Figs. S1 to S12
Tables S1 to S6
Legends for data S1 to S7
References

Other Supplementary Material for this manuscript includes the following:

Data S1 to S7

REFERENCES AND NOTES

1. C. M. Moore, M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, R. J. Geider, C. Guieu, S. L. Jaccard, T. D. Jickells, J. LaRoche, T. M. Lenton, N. M. Mahowald, E. Marañón, I. Marinov, J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, A. Tsuda, O. Ulloa, Processes and patterns of oceanic nutrient limitation. *Nat. Geosci.* **6**, 701–710 (2013).
2. X. Zhang, B. B. Ward, D. M. Sigman, Global nitrogen cycle: Critical enzymes, organisms, and processes for nitrogen budgets and dynamics. *Chem. Rev.* **120**, 5308–5351 (2020).
3. K. R. Arrigo, Marine microorganisms and global nutrient cycles. *Nature* **437**, 349–355 (2005).
4. D. Karl, A. Michaels, B. Bergman, D. Capone, E. Carpenter, R. Letelier, F. Lipschultz, H. Paerl, D. Sigman, L. Stal. "Dinitrogen fixation in the world's oceans" in *The Nitrogen Cycle at Regional to Global Scales* (Springer Netherlands, 2002), pp. 47–98.
5. M. Caffin, T. Moutin, R. A. Foster, P. Bouruet-Aubertot, A. M. Doglioli, H. Berthelot, C. Guieu, O. Grosso, S. Helias-Nunige, N. Leblond, A. Gimenez, A. A. Petrenko, A. de Verneil, S. Bonnet, N_2 Fixation as a dominant new N Source in the western tropical South Pacific Ocean (OUTPACE cruise). *Biogeosciences* **15**, 2565–2585 (2018).
6. T. D. Jickells, E. Buitenhuis, K. Altieri, A. R. Baker, D. Capone, R. A. Duce, F. Dentener, K. Fennel, M. Kanakidou, J. LaRoche, K. Lee, P. Liss, J. J. Middelburg, J. K. Moore, G. Okin, A. Oschlies, M. Sarin, S. Seitzinger, J. Sharples, A. Singh, P. Suntharalingam, M. Uematsu, L. M. Zamora, A reevaluation of the magnitude and impacts of anthropogenic atmospheric nitrogen inputs on the ocean. *Global Biogeochem. Cycles* **31**, 289–305 (2017).
7. T. H. Coale, V. Loconte, K. A. Turk-Kubo, B. Vanslembrouck, W. K. E. Mak, S. Cheung, A. Ekman, J.-H. Chen, K. Hagino, Y. Takano, T. Nishimura, M. Adachi, M. L. Gros, C. Larabell, J. P. Zehr, Nitrogen-fixing organelle in a marine alga. *Science* **384**, 217–222 (2024).
8. D. G. Capone, J. P. Zehr, H. W. Paerl, B. Bergman, E. J. Carpenter, *Trichodesmium*, a globally significant marine cyanobacterium. *Science* **276**, 1221–1229 (1997).
9. D. G. Capone, J. A. Burns, J. P. Montoya, A. Subramaniam, C. Mahaffey, T. Gunderson, A. F. Michaels, E. J. Carpenter, Nitrogen fixation by *Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochem. Cycles* **19**, (2005).
10. J. A. Sohm, E. A. Webb, D. G. Capone, Emerging patterns of marine nitrogen fixation. *Nat. Rev. Microbiol.* **9**, 499–508 (2011).
11. Y.-W. Luo, S. C. Doney, L. A. Anderson, M. Benavides, I. Berman-Frank, A. Bode, S. Bonnet, K. H. Boström, D. Böttjer, D. G. Capone, E. J. Carpenter, Y. L. Chen, M. J. Church, J. E. Dore, L. I. Falcón, A. Fernández, R. A. Foster, K. Furuya, F. Gómez, K. Gundersen, A. M. Hynes, D. M. Karl, S. Kitajima, R. J. Langlois, J. LaRoche, R. M. Letelier, E. Marañón, D. J. McGillicuddy, P. H. Moisander, C. M. Moore, B. Mouriño-Carballido, M. R. Mulholland, J. A. Needoba, K. M. Orcutt, A. J. Poulton, E. Rahav, P. Raimbault, A. P. Rees, L. Riemann, T. Shiozaki, A. Subramaniam, T. Tyrrell, K. A. Turk-Kubo, M. Varela, T. A. Villareal, E. A. Webb, A. E. White, J. Wu, J. P. Zehr, Database of diazotrophs in global ocean: Abundance, biomass and nitrogen fixation rates. *Earth Syst. Sci. Data* **4**, 47–73 (2012).
12. J. P. Zehr, L. A. McReynolds, Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* **55**, 2522–2526 (1989).
13. H. Farnelid, A. F. Andersson, S. Bertilsson, W. A. Al-Soud, L. H. Hansen, S. Sørensen, G. F. Steward, Å. Hagström, L. Riemann, Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLOS ONE* **6**, e19223 (2011).
14. J. P. Zehr, M. T. Mellon, S. Zani, New nitrogen-fixing microorganisms detected in oligotrophic oceans by amplification of nitrogenase (*nifH*) genes. *Appl. Environ. Microbiol.* **64**, 3444–3450 (1998).
15. C. Martínez-Pérez, W. Mohr, C. R. Löscher, J. Dekaezemacker, S. Littmann, P. Yilmaz, N. Lehnen, B. M. Fuchs, G. Lavik, R. A. Schmitz, J. LaRoche, M. M. M. Kuypers, The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nat. Microbiol.* **1**, 16163 (2016).
16. A. W. Thompson, R. A. Foster, A. Krupke, B. J. Carter, N. Musat, D. Vaulot, M. M. M. Kuypers, J. P. Zehr, Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**, 1546–1550 (2012).
17. H. Farnelid, J. Harder, M. Bentzon-Tilia, L. Riemann, Isolation of heterotrophic diazotrophic bacteria from estuarine surface waters. *Environ. Microbiol.* **16**, 3072–3082 (2014).
18. R. Langlois, T. Großkopf, M. Mills, S. Takeda, J. LaRoche, Widespread distribution and expression of gamma A (UMB), an uncultured, diazotrophic, γ-proteobacterial *nifH* phylotype. *PLOS ONE* **10**, e0128912 (2015).
19. J. P. Zehr, D. G. Capone, Changing perspectives in marine nitrogen fixation. *Science* **368**, eaay9514 (2020).
20. K. A. Turk-Kubo, M. R. Gradoville, S. Cheung, F. Cornejo-Castillo, K. J. Harding, M. Morando, M. Mills, J. P. Zehr, Non-cyanobacterial diazotrophs: Global diversity, distribution, ecophysiology, and activity in marine waters. *FEMS Microbiol. Rev.* **47**, fuac046 (2022).
21. T. O. Delmont, C. Quince, A. Shaiber, Ö. C. Esen, S. T. Lee, M. S. Rappé, S. L. McLellan, S. Lücker, A. M. Eren, Nitrogen-fixing populations of planctomycetes and proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804–813 (2018).
22. D. Bombar, R. W. Paerl, L. Riemann, Marine non-cyanobacterial diazotrophs: Moving beyond molecular detection. *Trends Microbiol.* **24**, 916–927 (2016).
23. M. Bentzon-Tilia, H. Farnelid, K. Jürgens, L. Riemann, Cultivation and isolation of N2-fixing bacteria from suboxic waters in the Baltic Sea. *FEMS Microbiol. Ecol.* **88**, 358–371 (2014).
24. D. Bombar, P. H. Moisander, J. W. Dippner, R. A. Foster, M. Voss, B. Karfeld, J. P. Zehr, Distribution of diazotrophic microorganisms and nifH gene expression in the Mekong River plume during intermonsoon. *Mar. Ecol. Prog. Ser.* **424**, 39–52 (2011).
25. W. Mohr, N. Lehnen, S. Ahmerkamp, H. K. Marchant, J. S. Graf, B. Tschitschko, P. Yilmaz, S. Littmann, H. Gruber-Vodicka, N. Leisch, M. Weber, C. Lott, C. J. Schubert, J. Milucka, M. M. M. Kuypers, Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine bacterium. *Nature* **600**, 105–109 (2021).
26. S. Cheung, J. Zehr, Gamma4: A genetically versatile Gammaproteobacterial *nifH* phylotype that is widely distributed in the North Pacific Ocean. *Environ. Microbiol.* **23**, 4246–4259 (2021).

27. C. Fernandez, L. Farías, O. Ulloa, Nitrogen fixation in denitrified marine waters. *PLOS ONE* **6**, e20539 (2011).
28. R. J. Langlois, D. Hümmer, J. LaRoche, Abundance of Gamma-A nifH genes in the Atlantic Ocean [dataset]. PANGAEA, doi:10.1594/PANGEAE.817863.
29. H. Halm, P. Lam, T. G. Ferdelman, G. Lavik, T. Dittmar, J. LaRoche, S. D'Hondt, M. M. Kuypers, Heterotrophic organisms dominate nitrogen fixation in the South Pacific Gyre. *ISME J.* **6**, 1238–1249 (2012).
30. C. R. Loescher, T. Großkopf, F. D. Desai, D. Gill, H. Schunck, P. L. Croot, C. Schlosser, S. C. Neulinger, N. Pinnow, G. Lavik, M. M. M. Kuypers, J. LaRoche, R. A. Schmitz, Facets of diazotrophy in the oxygen minimum zone waters off Peru. *ISME J.* **8**, 2180–2192 (2014).
31. L. Riemann, E. Rahav, U. Passow, H.-P. Grossart, D. de Beer, I. Klawonn, M. Eichner, M. Benavides, E. Bar-Zeev, Planktonic aggregates as hotspots for heterotrophic diazotrophy: The plot thickens. *Front. Microbiol.* **13**, 875050 (2022).
32. T. Shiozaki, Y. Nishimura, S. Yoshizawa, H. Takami, K. Hamasaki, A. Fujiwara, S. Nishino, N. Harada, Distribution and survival strategies of endemic and cosmopolitan diazotrophs in the Arctic Ocean, ISME J. **17**, 1340–1350 (2023). https://doi.org/10.1038/s41396-023-01424-x.
33. M. R. Olm, A. Crits-Christoph, S. Diamond, A. Lavy, P. B. Matheus Carnevali, J. F. Banfield, Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. *mSystems* **5**, e00731-19 (2020).
34. L. M. Rodriguez-R, K. T. Konstantinidis, Bypassing cultivation to identify bacterial species. *Microbe* **9**, 111–118 (2014).
35. M. Satomi, T. Fujii, "The family Oceanospirillaceae" in *The Prokaryotes: Gammaproteobacteria* (Springer, 2014), pp. 491–527.
36. M. M. Yakimov, L. Giuliano, R. Denaro, E. Crisafi, T. N. Chernikova, W.-R. Abraham, H. Luensdorf, K. N. Timmis, P. N. Golyshin, *Thalassolituus oleivorans* gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. *Int. J. Syst. Evol. Microbiol.* **54**, 141–148 (2004).
37. H. Chen, J. Dai, P. Yu, X. Wang, J. Wang, Y. Li, S. Wang, S. Li, D. Qiu, *Parathalassolituus penaei* gen. nov., sp. nov., a novel member of the family *Oceanospirillaceae* isolated from a coastal shrimp pond in Guangxi, PR, China. *Int. J. Syst. Evol. Microbiol.* **73**, (2023).
38. A. L. Müller, J. R. de Rezende, C. R. J. Hubert, K. U. Kjeldsen, I. Lagkouvardos, D. Berry, B. B. Jørgensen, A. Loy, Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents. *ISME J.* **8**, 1153–1165 (2004).
39. K. D. Young, The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).
40. K. J. Kechris, J. C. Lin, P. J. Bickel, A. N. Glazer, Quantitative exploration of the occurrence of lateral gene transfer by using nitrogen fixation genes as a case study. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9584–9589 (2006).
41. A. Koirala, V. S. Brözel, Phylogeny of nitrogenase structural and assembly components reveals new insights into the origin and distribution of nitrogen fixation across bacteria and archaea. *Microorganisms* **9**, 1662 (2021).
42. A. Nonaka, H. Yamamoto, N. Kamiya, H. Kotani, H. Yamakawa, R. Tsujimoto, Y. Fujita, Accessory proteins of the nitrogenase assembly, NifW, NifX/NafY, and NifZ, are essential for diazotrophic growth in the nonheterocystous cyanobacterium *Leptolyngbya boryana*. *Front. Microbiol.* **10**, 495 (2019).
43. R. Dixon, D. Kahn, Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* **2**, 621–631 (2004).
44. I. Klawonn, M. J. Eichner, S. T. Wilson, N. Moradi, B. Thamdrup, S. Kümmel, M. Gehre, A. Khalili, H.-P. Grossart, D. M. Karl, H. Ploug, Distinct nitrogen cycling and steep chemical gradients in *Trichodesmium* colonies. *ISME J.* **14**, 399–412 (2019).
45. J. C. Setubal, P. Dos Santos, B. S. Goldman, H. Ertesvåg, G. Espin, L. M. Rubio, S. Valla, N. F. Almeida, D. Balasubramanian, L. Cromes, L. Curatti, Z. Du, E. Godsy, B. Goodner, K. Hellner-Burris, J. A. Hernandez, K. Houmiel, J. Imperial, C. Kennedy, T. J. Larson, P. Latreille, L. S. Ligon, J. Lu, M. Mærk, N. M. Miller, S. Norton, I. P. O'Carroll, I. Paulsen, E. C. Raulfs, R. Roemer, J. Rosser, D. Segura, S. Slater, S. L. Stricklin, D. J. Studholme, J. Sun, C. J. Viana, E. Wallin, B. Wang, C. Wheeler, H. Zhu, D. R. Dean, R. Dixon, D. Wood, Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J. Bacteriol.* **191**, 4534–4545 (2009).
46. W. Sabra, A. P. Zeng, H. Lünsdorf, W. D. Deckwer, Effect of oxygen on formation and structure of *Azotobacter vinelandii Alginate* and its role in protecting nitrogenase. *Appl. Environ. Microbiol.* **66**, 4037–4044 (2000).
47. R. K. Poole, S. Hill, Respiratory protection of nitrogenase activity in *Azotobacter vinelandii*—Roles of the terminal oxidases. *Biosci. Rep.* **17**, 303–317 (1997).
48. N. Padmaja, H. Rajaram, S. K. Apte, A novel hemerythrin DNase from the nitrogen-fixing cyanobacterium *Anabaena sp.* strain PCC7120. *Arch. Biochem. Biophys.* **505**, 171–177 (2011).
49. C. Alvarez-Carreño, V. Alva, A. Becerra, A. Lazcano, Structure, function and evolution of the hemerythrin-like domain superfamily. *Protein Sci.* **27**, 848–860 (2018).
50. F. Mus, A. B. Alleman, N. Pence, L. C. Seefeldt, J. W. Peters, Exploring the alternatives of biological nitrogen fixation. *Metallomics* **10**, 523–538 (2018).
51. R. Navarro-González, C. P. McKay, D. N. Mvondo, A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature* **412**, 61–64 (2001).
52. E. S. Boyd, J. W. Peters, New insights into the evolutionary history of biological nitrogen fixation. *Front. Microbiol.* **4**, 201 (2013).
53. A. K. Garcia, D. F. Harris, A. J. Rivier, B. M. Carruthers, A. Pinochet-Barros, L. C. Seefeldt, B. Kaçar, Nitrogenase resurrection and the evolution of a singular enzymatic mechanism. *eLife* **12**, e85003 (2023).
54. A. Knapp, The sensitivity of marine N_2 fixation to dissolved inorganic nitrogen. *Front. Microbiol.* **3**, 374 (2012).
55. M. A. Trainer, T. C. Charles, The role of PHB metabolism in the symbiosis of rhizobia with legumes. *Appl. Microbiol. Biotechnol.* **71**, 377–386 (2006).
56. D. Tec-Campos, C. Zuñiga, A. Passi, J. Del Toro, J. D. Tibocha-Bonilla, A. Zepeda, M. J. Betenbaugh, K. Zengler, Modeling of nitrogen fixation and polymer production in the heterotrophic diazotroph *Azotobacter vinelandii* DJ. *Metab. Eng. Commun.* **11**, e00132 (2020).
57. K. Mandon, F. Nazaret, D. Farajzadeh, G. Alloing, P. Frendo, Redox regulation in diazotrophic bacteria in interaction with plants. *Antioxidants* **10**, 880 (2021).
58. E. K. Heiniger, Y. Oda, S. K. Samanta, C. S. Harwood, How posttranslational modification of nitrogenase is circumvented in *Rhodopseudomonas palustris* strains that produce hydrogen gas constitutively. *Appl. Environ. Microbiol.* **78**, 1023–1032 (2012).
59. A. B. Alleman, J. W. Peters, Mechanisms for generating low potential electrons across the metabolic diversity of nitrogen-fixing bacteria. *Appl. Environ. Microbiol.* **89**, e0037823 (2023).
60. K. Mandon, N. Michel-Reydellet, S. Encarnación, P. A. Kaminski, A. Leija, M. A. Cevallos, C. Elmerich, J. Mora, Poly-β-hydroxybutyrate turnover in *Azorhizobium caulinodans* is required for growth and affects *nifA* expression. *J. Bacteriol.* **180**, 5070–5076 (1998).
61. J. M. Dubbs, F. Robert Tabita, Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO_2 assimilation, nitrogen fixation, hydrogen metabolism and energy generation. *FEMS Microbiol. Rev.* **28**, 353–376 (2004).
62. C. Dingler, J. Kuhla, H. Wassink, J. Oelze, Levels and activities of nitrogenase proteins in *Azotobacter vinelandii* grown at different dissolved oxygen concentrations. *J. Bacteriol.* **170**, 2148–2152 (1988).
63. M. A. Saito, E. M. Bertrand, S. Dutkiewicz, V. V. Bulygin, D. M. Moran, F. M. Monteiro, M. J. Follows, F. W. Valois, J. B. Waterbury, Iron conservation by reduction of metalloenzyme inventories in the marine diazotroph *Crocosphaera Watsonii*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2184–2189 (2011).
64. F. Lv, Y. Zhan, W. Lu, X. Ke, Y. Shao, Y. Ma, J. Zheng, Z. Yang, S. Jiang, L. Shang, Y. Ma, L. Cheng, C. Elmerich, Y. Yan, M. Lin, Regulation of hierarchical carbon substrate utilization, nitrogen fixation, and root colonization by the Hfq/Crc/CrcZY genes in *Pseudomonas stutzeri*. *iScience* **25**, 105663 (2022).
65. D. Fonseca-Batista, X. Li, V. Riou, V. Michotey, F. Deman, F. Fripiat, S. Guasco, N. Brion, N. Lemaitre, M. Tonnard, M. Gallinari, H. Planquette, F. Planchon, G. Sarthou, M. Elskens, L. Chou, F. Dehairs, Evidence of high N_2 fixation rates in the temperate northeast Atlantic. *Biogeosciences* **16**, 999–1017 (2019).
66. Z. Shao, Y.-W. Luo, Controlling factors on the global distribution of a representative marine non-cyanobacterial diazotroph phylotype (Gamma A). *Biogeosciences* **19**, 2939–2952 (2022).
67. K. Turk, A. P. Rees, J. P. Zehr, N. Pereira, P. Swift, R. Shelley, M. Lohan, E. M. S. Woodward, J. Gilbert, Nitrogen fixation and nitrogenase (*nifH*) expression in tropical waters of the eastern North Atlantic. *ISME J.* **5**, 1201–1212 (2011).
68. P. H. Moisander, R. A. Beinart, I. Hewson, A. E. White, K. S. Johnson, C. A. Carlson, J. P. Montoya, J. P. Zehr, Unicellular cyanobacterial distributions broaden the oceanic N2 fixation domain. *Science* **327**, 1512–1514 (2010).
69. M. Simon, H.-P. Grossart, B. Schweitzer, H. Ploug, Microbial ecology of organic aggregates in aquatic ecosystems. *Aquat. Microb. Ecol.* **28**, 75–211 (2002).
70. E. Geisler, A. Bogler, E. Rahav, E. Bar-Zeev, Direct detection of heterotrophic diazotrophs associated with planktonic aggregates. *Sci. Rep.* **9**, 9288 (2019).
71. J. N. Pedersen, D. Bombar, R. W. Paerl, L. Riemann, Diazotrophs and N_2-fixation associated with particles in coastal estuarine waters. *Front. Microbiol.* **9**, 2759 (2018).
72. H. Farnelid, K. Turk-Kubo, H. Ploug, J. E. Ossolinski, J. R. Collins, B. A. S. Van Mooy, J. P. Zehr, Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *ISME J.* **13**, 170–182 (2019).
73. N. Lemaitre, H. Planquette, F. Planchon, G. Sarthou, S. Jacquet, M. I. García-Ibáñez, A. Gourain, M. Cheize, L. Monin, L. André, P. Laha, H. Terryn, F. Dehairs, Particulate barium tracing of significant mesopelagic carbon remineralisation in the North Atlantic. *Biogeosciences* **15**, 2289–2307 (2018).
74. F. M. Cornejo-Castillo, J. P. Zehr, Intriguing size distribution of the uncultured and globally widespread marine non-cyanobacterial diazotroph Gamma-A. *ISME J.* **15**, 124–128 (2021).
75. C. Lønborg, F. Baltar, C. Carreira, X. A. G. Morán, Dissolved organic carbon source influences tropical coastal heterotrophic bacterioplankton response to experimental warming. *Front. Microbiol.* **10**, 2807 (2019).
76. D. Li, J. Liu, R. Zhang, M. Chen, W. Yang, J. Li, Z. Fang, B. Wang, Y. Qiu, M. Zheng, N2 fixation impacted by carbon fixation via dissolved organic carbon in the changing Daya Bay, South China Sea. *Sci. Total Environ.* **674**, 592–602 (2019).

77. L. Bopp, O. Aumont, L. Kwiatkowski, C. Clerc, L. Dupont, C. Ethé, R. Séférian, A. Tagliabue, Diazotrophy as a key driver of the response of marine net primary productivity to climate change. *Biogeosciences* **19**, 4267–4285 (2022).
78. S. Bonnet, M. Benavides, F. A. C. Le Moigne, M. Camps, A. Torremocha, O. Grosso, C. Dimier, D. Spungin, I. Berman-Frank, L. Garczarek, F. M. Cornejo-Castillo, Diazotrophs are overlooked contributors to carbon and nitrogen export to the deep ocean. *ISME J.* **17**, 47–58 (2023).
79. C. Martínez-Pérez, W. Mohr, A. Schwedt, J. Dürschlag, C. M. Callbeck, H. Schunck, J. Dekaezemacker, C. R. T. Buckner, G. Lavik, B. M. Fuchs, M. M. M. Kuypers, Metabolic versatility of a novel N2-fixing alphaproteobacterium isolated from a marine oxygen minimum zone. *Environ. Microbiol.* **20**, 755–768 (2018).
80. M. Bentzon-Tilia, I. Severin, L. H. Hansen, L. Riemann, Genomics and ecophysiology of heterotrophic nitrogen-fixing bacteria isolated from estuarine surface water. *mBio* **6**, e00929 (2015).
81. K. A. Turk-Kubo, M. M. Mills, K. R. Arrigo, G. van Dijken, B. A. Henke, B. Stewart, S. T. Wilson, J. P. Zehr, UCYN-A/haptophyte symbioses dominate N2 fixation in the Southern California Current System. *ISME Commun.* **1**, 42 (2021).
82. D. Prévost, H. Antoun, L. M. Bordeleau, Effects of low temperatures on nitrogenase activity in sainfoin (*Onobrychis viciifolia*) nodulated by arctic rhizobia. *FEMS Microbiol. Ecol.* **3**, 205–210 (1987).
83. T. Shiozaki, A. Fujiwara, K. Inomura, Y. Hirose, F. Hashihama, N. Harada, Biological nitrogen fixation detected under Antarctic sea ice. *Nat. Geosci.* **13**, 729–732 (2020).
84. T. D. Niederberger, J. A. Sohm, J. Tirindelli, T. Gunderson, D. G. Capone, E. J. Carpenter, S. C. Cary, Diverse and highly active diazotrophic assemblages inhabit ephemerally wetted soils of the Antarctic Dry Valleys. *FEMS Microbiol. Ecol.* **82**, 376–390 (2012).
85. L. W. Von Friesen, L. Riemann, Nitrogen fixation in a changing arctic ocean: An overlooked source of nitrogen? *Front. Microbiol.* **11**, 596426 (2020).
86. J.-P. Bellenger, T. Wichard, Y. Xu, A. M. L. Kraepiel, Essential metals for nitrogen fixation in a free-living N2-fixing bacterium: Chelation, homeostasis and high use efficiency. *Environ. Microbiol.* **13**, 1395–1411 (2011).
87. R. Paerl, T. Hansen, N. Henriksen, A. Olesen, L. Riemann, N-fixation and related O2 constraints on model marine diazotroph Pseudomonas stutzeri BAL361. *Aquat. Microb. Ecol.* **81**, 125–136 (2018).
88. B. M. Robicheau, J. Tolman, E. M. Bertrand, J. LaRoche, Highly-resolved interannual phytoplankton community dynamics of the coastal Northwest Atlantic. *ISME Commun.* **2**, 38 (2022).
89. J. P. Zehr, J. B. Waterbury, P. J. Turner, J. P. Montoya, E. Omoregle, G. F. Steward, A. Hansen, D. M. Karl, Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* **412**, 635–638 (2001).
90. R. R. L. Guillard, J. H. Ryther, Studies of marine planktonic diatoms: I. Cyclotella nana hustedt, and detonula confervacea (cleve) gran. *Can. J. Microbiol.* **8**, 229–239 (1962).
91. J.-M. Ratten, "The diversity, distribution and potential metabolism of non-cyanobacterial diazotrophs in the North Atlantic Ocean," thesis, Dalhousie University, Halifax, Nova Scotia, Canada (2017).
92. S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).
93. A. M. Eren, E. Kiefl, A. Shaiber, I. Veseli, S. E. Miller, M. S. Schechter, I. Fink, J. N. Pan, M. Yousef, E. C. Fogarty, F. Trigodet, A. R. Watson, Ö. C. Esen, R. M. Moore, Q. Clayssen, M. D. Lee, V. Kivenson, E. D. Graham, B. D. Merrill, A. Karkman, D. Blankenberg, J. M. Eppley, A. Sjödin, J. J. Scott, X. Vázquez-Campos, L. J. McKay, E. A. McDaniel, S. L. R. Stevens, R. E. Anderson, J. Fuessel, A. Fernandez-Guerra, L. Maignien, T. O. Delmont, A. D. Willis, Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* **6**, 3–6 (2021).
94. C. Rinke, P. Schwientek, A. Sczyrba, N. N. Ivanova, I. J. Anderson, J.-F. Cheng, A. Darling, S. Malfatti, B. K. Swan, E. A. Gies, J. A. Dodsworth, B. P. Hedlund, G. Tsiamis, S. M. Sievert, W.-T. Liu, J. A. Eisen, S. J. Hallam, N. C. Kyrpides, R. Stepanauskas, E. M. Rubin, P. Hugenholtz, T. Woyke, Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
95. M. A. Campbell, W.-J. Chen, J. A. López, Molecular data do not provide unambiguous support for the monophyly of flatfishes (*Pleuronectiformes*): A reply to Betancur-R and Ortí. *Mol. Phylogenet. Evol.* **75**, 149–153 (2014).
96. M. D. Lee, GToTree: A user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164 (2019).
97. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLOS ONE* **5**, e9490 (2010).
98. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
99. R. C. Edgar, MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
100. M. Nei, S. Kumar, *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, 2000).
101. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
102. E. Karensti, S. G. Acinas, P. Bork, C. Bowler, C. De Varges, J. Raes, M. Sullivan, D. Arendt, F. Benzoni, J.-M. Claverie, M. Follows, G. Gorsky, P. Hingamp, D. Iudicone, O. Jaillon, S. Kandels- Lewis, U. Krzic, F. Not, H. Ogata, S. Peasant, E. G. Reynaud, C. Sardet, M. E. Sieracki, S. Speich, D. Velayoudon, J. Weissenbach, P. Wincker, A holistic approach to marine eco-systems biology. *PLOS Biol.* **9**, e1001177 (2011).
103. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie2. *Nat. Methods* **9**, 357–359 (2012).
104. D. Li, C.-M. Liu, R. Luo, K. Sadakane, T.-W. Lam, MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics* **31**, 1674–1676 (2015).
105. M. Alonge, L. Lebeigle, M. Kirsche, K. Jenike, S. Ou, S. Aganezov, X. Wang, Z. B. Lippman, M. C. Schatz, S. Soyk, Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* **23**, 258 (2022).
106. D. Hyatt, G.-L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
107. R. L. Tatusov, M. Y. Galperin, D. A. Natale, E. V. Koonin, The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
108. T. Aramaki, R. Blanc-Mathieu, H. Endo, K. Ohkubo, M. Kanehisa, S. Goto, H. Ogata, KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
109. J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G. A. Salazar, E. L. L. Sonnhammer, S. C. E. Tosatto, L. Paladin, S. Raj, L. J. Richardson, R. D. Finn, A. Bateman, Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
110. T. O. Delmont, A. M. Eren, Linking pangenomes and metagenomes: The *Prochlorococcus* metapangenome. *PeerJ* **6**, e4320 (2018).
111. D. H. Parks, M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil, P. Hugenholtz, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
112. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
113. L. Pritchard, H. R. Glover, S. Humphris, J. G. Elphinstone, I. K. Toth, Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* **8**, 12–24 (2016).
114. J. Zorz, C. Willis, A. M. Comeau, M. G. I. Langille, C. L. Johnson, W. K. W. Li, J. LaRoche, Drivers of regional bacterial community structure and diversity in the Northwest Atlantic Ocean. *Front. Microbiol.* **10**, 281 (2019).
115. S. Haas, B. M. Robicheau, S. Rakshit, J. Tolman, C. K. Algar, J. LaRoche, D. W. R. Wallace, Physical mixing in coastal waters controls and decouples nitrification via biomass dilution. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2004877118 (2021).
116. B. M. Robicheau, J. Tolman, S. Rose, D. Desai, J. LaRoche, Marine nitrogen-fixers in the Canadian Arctic Gateway are dominated by biogeographically distinct noncyanobacterial communities. *FEMS Microbiol. Ecol.* **99**, fiad122 (2023).
117. S. Zani, M. T. Mellon, J. L. Collier, J. P. Zehr, Expression of nifH genes in natural microbial assemblages in Lake George, New York, detected by reverse transcriptase PCR. *Appl. Environ. Microbiol.* **66**, 3119–3124 (2000).
118. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
119. A. M. Comeau, G. M. Douglas, M. G. I. Langille, Microbiome helper: A custom and streamlined workflow for microbiome research. *mSystems* **2**, e00127-16 (2017).
120. A. Amir, D. McDonald, J. A. Navas-Molina, E. Kopylova, J. T. Morton, Z. Z. Xu, E. P. Kightley, L. R. Thompson, E. R. Hyde, A. Gonzalez, R. Knight, Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* **2**, e00191-16 (2017).
121. J. C. Gaby, D. H. Buckley, A comprehensive evaluation of PCR primers to amplify the nifH gene of nitrogenase. *PLOS ONE* **7**, e42149 (2012).
122. E. Bolyen, J. R. Rideout, M. R. Dillon, N. A. Bokulich, C. C. Abnet, G. A. Al-Ghalith, H. Alexander, E. J. Alm, M. Arumugam, F. Asnicar, Y. Bai, J. E. Bisanz, K. Bittinger, A. Brejnrod, C. J. Brislawn, C. T. Brown, B. J. Callahan, A. M. Caraballo-Rodríguez, J. Chase, E. K. Cope, R. Da Silva, C. Diener, P. C. Dorrestein, D. M. Durall, C. Duvallet, C. F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J. M. Gauglitz, S. M. Gibbons, D. L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Huttenhower, G. A. Huttley, S. Janssen, A. K. Jarmusch, L. Jiang, B. D. Kaehler, K. B. Kang, C. R. Keefe, P. Keim, S. T. Kelley, D. Knights, I. Koester, T. Kosciolek, J. Kreps, M. G. I. Langille, J. Lee, R. Ley, Y.-X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B. D. Martin, D. McDonald, L. J. McIver, A. V. Melnik, J. L. Metcalf, S. C. Morgan, J. T. Morton, A. T. Naimey, J. A. Navas-Molina, L. F. Nothias, S. B. Orchanian, T. Pearson, S. L. Peoples, D. Petras, M. L. Preuss, E. Pruesse, L. B. Rasmussen, A. Rivers, M. S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S. J. Song, J. R. Spear, A. D. Swafford, L. R. Thompson, P. J. Torres, P. Trinh, A. Tripathi, P. J. Turnbaugh, S. Ul-Hasan, J. J. J. van der Hooft, F. Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K. C. Weber, C. H. D. Williamson, A. D. Willis, Z. Z. Xu, J. R. Zaneveld, Y. Zhang, Q. Zhu,

R. Knight, J. G. Caporaso, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
123. C. Ramakers, J. M. Ruijter, R. H. L. Deprez, A. F. M. Moorman, Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* **339**, 62–66 (2003).
124. J. M. Ruijter, C. Ramakers, W. M. H. Hoogaars, Y. Karlen, O. Bakker, M. J. B. van den Hoff, A. F. M. Moorman, Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* **37**, e45 (2009).
125. W. Mohr, T. Großkopf, D. W. R. Wallace, J. LaRoche, Methodological underestimation of oceanic nitrogen fixation rates. *PLOS ONE* **5**, e12583 (2010).
126. T. Großkopf, W. Mohr, T. Baustian, H. Schunk, D. Gill, M. M. M. Kuypers, G. Lavik, R. A. Schmitz, D. R. Wallance, J. LaRoche, Doubling of marine dinitrogen-fixation rates based on direct measurements. *Nature* **488**, 361–364 (2012).
127. D. Fonseca-Batista, F. Dehairs, V. Riou, F. Fripiat, M. Elskens, F. Deman, N. Brion, F. Quéroué, M. Bode, H., Nitrogen fixation in the eastern Atlantic reaches similar levels in the Southern and Northern Hemisphere. *J. Geophys. Res. Oceans* **122**, 587–601 (2017).
128. R. M. Parr, S. A. Clements, "Intercomparison of enriched stable isotope reference materials for medical and biological studies" (International Atomic Energy Agency, 1991).
129. A. E. White, J. Granger, C. Selden, M. R. Gradoville, L. Potts, A. Bourbonnais, R. W. Fulweiler, A. N. Knapp, W. Mohr, P. H. Moisander, C. R. Tobias, M. Caffin, S. T. Wilson, M. Benavides, S. Bonnet, M. R. Mulholland, B. X. Chang, A critical review of the $^{15}N_2$ tracer method to measure diazotrophic production in pelagic ecosystems. *Limnol. Oceanogr. Methods* **18**, 129–147 (2020).
130. M. Wu, J. S. P. McCain, E. Rowland, R. Middag, M. Sandgren, A. E. Allen, E. M. Bertrand, Manganese and iron deficiency in Southern Ocean *Phaeocystis antarctica* populations revealed through taxon-specific protein indicators. *Nat. Commun.* **10**, 3582 (2019).
131. M. The, M. J. MacCoss, W. S. Noble, L. Käll, Fast and accurate protein false discovery rates on large-scale proteomics data sets with Percolator 3.0. *J. Am. Soc. Mass Spectrom.* **27**, 1719–1727 (2016).
132. J. S. P. McCain, A. E. Allen, E. M. Bertrand, Proteomic traits vary across taxa in a coastal Antarctic phytoplankton bloom. *ISME J.* **16**, 569–579 (2022).
133. H. Wickham, Create elegant data visualisations using the grammar of graphics; https://ggplot2.tidyverse.org/.
134. M. Morgan, M. Ramos, B. P. Maintainer, BiocVersion: Set the appropriate version of Bioconductor packages. Bioconductor version: Release (3.18), 2023. doi:10.18129/B9.bioc.BiocVersion.
135. H. Wickham, Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
136. J. Oksanen, Vegan community ecology package version 2.6-2 (2022).
137. A. Kassambara, kassambara/rstatix: Pipe-friendly framework for basic statistical tests version 0.7.2.999; https://rdrr.io/github/kassambara/rstatix/.
138. P. Massicotte, World map data from natural Earth; https://docs.ropensci.org/rnaturalearth/.
139. C. Wilke, Introduction to cowplot; https://wilkelab.org/cowplot/articles/introduction.html.
140. G. Yu, S. Xu, Scatterpie: Scatter pie plot; https://cran.r-project.org/web/packages/scatterpie/index.html.
141. D. Dunnington, B. Thorne, D. Hernangómez, Ggspatial: Spatial data framework for ggplot2; https://cran.r-project.org/web/packages/ggspatial/index.html.
142. P. Stothard, D. S. Wishart, Circular genome visualization and exploration using CGView. *Bioinformatics* **21**, 537–539 (2005).
143. R. K. Aziz, D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
144. K. E. Luxem, A. M. L. Kraepiel, L. Zhang, J. R. Waldbauer, X. Zhang, Carbon substrate re-orders relative growth of a bacterium using Mo-, V-, or Fe-nitrogenase for nitrogen fixation. *Environ. Microbiol.* **22**, 1397–1408 (2020).
145. J. W. Leigh, D. Bryant, PopART: Full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).
146. B. M. Robicheau, J. Tolman, D. Desai, J. LaRoche, Microevolutionary patterns in ecotypes of the symbiotic cyanobacterium UCYN-A revealed from a Northwest Atlantic coastal time series. *Sci. Adv.* **9**, eadh9768 (2023).
147. H. Bandelt, P. Forster, A. Röhl, Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* **16**, 37–48 (1999).

Acknowledgments: We would like to thank W. Mohr for data collection on the Ant Polarstern ANT26. The GEOTRACES 2021 Intermediate Data Product (IDP2021) represents an international collaboration and is endorsed by the Scientific Committee on Oceanic Research (SCOR). The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to the IDP2021. We are grateful to the Atlantic Zone Monitoring Program—Maritimes and the Bedford Basin Monitoring Program, led by A. Cogswell and L. Beazley, at the Bedford Institute of Oceanography, along with the masters, officers, and crew of the CCGS *Hudson* and CCGS *Sigma T*. This work contains information licensed under the Open Government Licence—Canada. The *nifH* time series presented from the Bedford Basin involved the participation of many individuals who helped process the weekly samples (including I. Luddington, J.-M. Ratten, C. Willis, J. Zorz, C. Mackie, and M. Segura Guzman). We are greatly indebted to the master, G. Ferrand, the officers, and crew from the N/O *Pourquoi Pas?* for their logistic support during the GEOVIDE Voyage. The work on the GEOVIDE Voyage was supported in logistics by DT INSU and GENAVIR. Proteomic MS was conducted at Dalhousie's Biological Mass Spectrometry Facility, and we are grateful for technical assistance from A. Cohen and R. John. **Funding:** The funding of this work was supported through various individuals. J.L. was funded by Module C 37114 (OFI-CFREF) and NSERC Discovery Grant RGPIN-/04060-2021 for the funding of this project. E.M.B. acknowledges support from the OFI, NSERC Discovery Grant RGPIN-2015-05009, and Simons Foundation Grant 504183. B.M.R. was funded by an NSERC CGS-D Award, a Killam Predoctoral Scholarship, and an NS Graduate Scholarship. D.F.-B. was supported during the laboratory experimentation by funding from the Canada First Research Excellence Fund, through an International Postdoctoral Fellowship of the Ocean Frontier Institute (OFI) at Dalhousie University, and at the time of manuscript preparation by the Transatlantic Ocean System Science and Technology (TOSST) program. The French National Research Agency supported ANR GEOVIDE (ANR-13-BS06-0014). G.S. was supported by the French National Center for Scientific Research (CNRS-LEFE-CYBER), the LabexMER (ANR-10-LABX-19), and Ifremer (Institut français de recherche pour l'exploitation de la mer). **Author contributions:** Conceptualization: S.A.R., D.F.-B., E.M.B., and J.L. Methodology: S.A.R., J.T., D.F.-B., E.R., E.J.H.K., D.D., E.M.B., and J.L. Data curation: S.A.R., B.M.R., J.T., A.M.C., M.G.I.L., E.R., J.-M.R., J.J.-H., G.S., E.J.H.K., E.M.B., and J.L. Investigation: S.A.R., D.F.-B., E.R., E.D., and J.L. Visualization: S.A.R., B.M.R., E.M.B., and J.L. Supervision: E.M.B. and J.L. Writing—original draft: S.A.R., E.M.B., and J.L. Writing—review and editing: S.A.R., B.M.R., J.T., D.F.-B., E.R., D.D., J.-M.R., E.J.H.K., A.M.C., M.G.I.L., J.J.-H., G.S., E.D., E.M.B., and J.L. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. *NifH* community sequences used in this study can be found under bioprojects PRJNA930772 (BBMP), PRJNA325151 (AZMP), and PRJNA931255 (GN03) and a further extended list of SRAs in data S5. The full genome of *Cand.* T. haligoni can be found under project number PRJNA1046103. Proteomics data used for quantification in this study can be found in the ProteomeXchange database under the accession ID PXD045743. Scripts to reproduce the dataset analyzed can be found at: https://github.com/bertrand-lab/.

Submitted 28 November 2023
Accepted 25 June 2024
Published 31 July 2024
10.1126/sciadv.adn1476