









Desulfovibrio diazotrophicus sp. nov., a sulfate-reducing bacterium from the human gut capable of nitrogen fixation

Lizbeth Sayavedra ^{1*}, Tianqi Li ^{1,2},
Marcelo Bueno Batista ³, Brandon K. B. Seah ⁴,
Catherine Booth ¹, Qixiao Zhai ², Wei Chen ²
and Arjan Narbad ¹

¹Gut Health and Microbes, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK.

²State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi, China.

³Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, UK.

⁴Max Planck Institute for Developmental Biology, Tübingen, Germany.

Summary

Sulfate-reducing bacteria (SRB) are widespread in human guts, yet their expansion has been linked to colonic diseases. We report the isolation, sequencing and physiological characterization of strain QI0027^T, a novel SRB species belonging to the class *Desulfovibrionia*. Metagenomic sequencing of stool samples from 45 Chinese individuals, and comparison with 1690 *Desulfovibrionaceae* metagenome-assembled genomes recovered from humans of diverse geographic locations, revealed the presence of QI0027^T in 22 further individuals. QI0027^T encoded nitrogen fixation genes and based on the acetylene reduction assay, actively fixed nitrogen. Transcriptomics revealed that QI0027^T overexpressed 42 genes in nitrogen-limiting conditions compared to cultures supplemented with ammonia, including genes encoding nitrogenases, a urea uptake system and the urease complex. Reanalyses of 835 public stool metatranscriptomes showed that nitrogenase genes from *Desulfovibrio* bacteria were expressed in six samples suggesting that nitrogen fixation might be active in the gut environment. Although frequently

thought of as a nutrient-rich environment, nitrogen fixation can occur in the human gut. Animals are often nitrogen limited and have evolved diverse strategies to capture biologically active nitrogen, ranging from amino acid transporters to stable associations with beneficial microbes that provide fixed nitrogen. QI0027^T is the first *Desulfovibrio* human isolate for which nitrogen fixation has been demonstrated, suggesting that some sulfate-reducing bacteria could also play a role in the availability of nitrogen in the gut.

Abbreviations

ANI	average nucleotide identity
<i>D.</i>	<i>Desulfovibrio</i>
FAME	fatty acid methyl ester
MAGs	metagenome-assembled genomes
MCP	methyl-accepting chemotaxis protein
SEM	scanning electron microscopy
SRB	sulfate-reducing bacteria
TEM	transmission electron microscopy
WGS	whole-genome sequencing

Introduction

Sulfate-reducing bacteria (SRB) are present in the mouth and the gut of ~50% of the human population (Nava *et al.*, 2012; Pasolli *et al.*, 2019). SRB thrive in the gut, releasing hydrogen sulphide (H₂S) as a product of sulfate reduction. H₂S is a potent genotoxin and has been linked to chronic colonic disorders and inflammation of the large intestine (Coutinho *et al.*, 2017). Likewise, the presence of some *Desulfovibrio* species has been implicated in chronic periodontitis, cell death and inflammatory bowel diseases such as ulcerative colitis and Crohn's disease (Loubinoux *et al.*, 2003). The detrimental role of SRB is, however, not firmly established. H₂S can also act as a signalling molecule or energy source for mitochondria (Gong *et al.*, 2011; Carbonero *et al.*, 2012). Moreover, by using hydrogen, SRB help in the efficient energy acquisition and complete oxidation of substrates produced by

Received 7 October, 2020; revised 24 March, 2021; accepted 16 April, 2021. *For correspondence. E-mail lizbeth.sayavedra@quadram.ac.uk; Tel: +44 (0)1603 25 1402.

fermentative bacteria (Fischbach and Sonnenburg, 2011). SRB could, therefore, have a dual role within the gut microbiome. Relatively few strains from healthy individuals have been characterized. Cultivated SRB isolated from humans include *Desulfovibrio piger*, *D. fairfieldensis*, *D. desulfuricans* and *D. legallii* (Loubinoux *et al.*, 2003; Vasoo *et al.*, 2014), with *D. piger* described as the most abundant in samples obtained from Western countries (Carbonero and Gaskins, 2015). The physiology and genomic potential of strains isolated from non-Western countries remain largely unexplored.

Biological nitrogen fixation is the process by which gaseous dinitrogen (N₂) is reduced to biologically available ammonia (NH₃) by diazotrophic microbes (Seefeldt *et al.*, 2009). Diazotrophy was reported in some *Desulfovibrio* species from free-living communities and the termite gut (Postgate and Kent, 1985; Kuhnigk *et al.*, 1996; Fröhlich *et al.*, 1999). All animals require access to a source of fixed nitrogen (Fuller and Reeds, 1998). Most of the readily available nitrogen in humans comes from food, but nitrogen fixation can occur in the microbiome of some human populations (Igai *et al.*, 2016). In the human microbiome, the potential to fix nitrogen has been so far linked to *Klebsiella* and *Clostridia* (Bergersen and Hipsley, 1970; Igai *et al.*, 2016). The human host can regulate the nitrogen available to the microbiome through the diet and intestinal secretions (Reese *et al.*, 2018), so bacteria that can fix nitrogen could potentially have a competitive advantage in the gut environment.

We report here the isolation and characterization of a novel species of the genus *Desulfovibrio*, strain QI0027^T, isolated from a stool sample provided by a Chinese male donor. Surprisingly this strain encoded the genes for

nitrogen fixation. The goals of this study were (i) to characterize strain QI0027^T, (ii) to identify if diazotrophy is commonly distributed among *Desulfovibrionaceae* associated with humans and (iii) to examine if the genes required for nitrogen fixation are expressed and functional. We used metagenomic sequencing of 45 human donors, as well as analysed comprehensive collections of genomes recovered from metagenomes to examine the distribution of QI0027^T. We used physiological tests to examine the nitrogenase activity. Finally, we used transcriptomics to determine the genes that are expressed by QI0027^T in response to the absence of a source of fixed nitrogen and reanalysed 835 stool metatranscriptomes to search for the expression of nitrogenases belonging to *Desulfovibrio*.

Results

Genome sequencing and phylogenomics

As part of a study to isolate and characterize SRB from the intestinal tract of healthy Chinese donors, we isolated strain QI0027^T. Using short-read and long-read sequencing, we sequenced and assembled a circular chromosome of 2.9 Mb genome with a 62.1% GC content. The closest relative of QI0027^T was *D. legallii* (Genome statistics of QI0027^T and *D. legallii* are shown in Table 1). These bacteria grouped within the *Desulfovibrio* genus based on a well-supported phylogenomic tree reconstructed using 30 marker genes (Shivani *et al.*, 2017) (Fig. 1). Even though the 16S rRNA gene of QI0027^T shared a 99% identity with the partial 16S rRNA sequence of *D. legallii*, the average nucleotide identity (ANI) of their whole genomes was 86.49%, the digital DNA–DNA hybridization (dDDH) was 31.4% (29%–33.9%

Table 1. Differentiation between *D. legallii* and QI0027^T.

		QI0027 ^T	<i>D. legallii</i> KHC7
Genome	GC (%) content	62.14	64.81
	Genome size (Mbp)	2.9	2.7
	N50 (bp)	2 906 404	124 328
	Number of scaffolds	1	36
	Genome completeness ^a (%)	100	100
	Genome contamination ^a (%)	0	0
	Strain heterogeneity (%)	0	0
Physiology	Antibiotic resistance		
	Ampicillin (25 µg)	S	S
	Chloramphenicol (10 µg)	S	S
	Nalidixic acid (30 µg)	S	S
	Erythromycin (30 µg)	S	S
	Tetracycline (30 µg)	S	S
	Streptomycin (25 µg)	R	R

For comparison, we used the *D. legallii* genome GCA_900102485. S = sensitive; R = Resistant; +/- = Light resistance.

^aGenome completeness and genome contamination were determined using Patric (Parrello *et al.*, 2019).

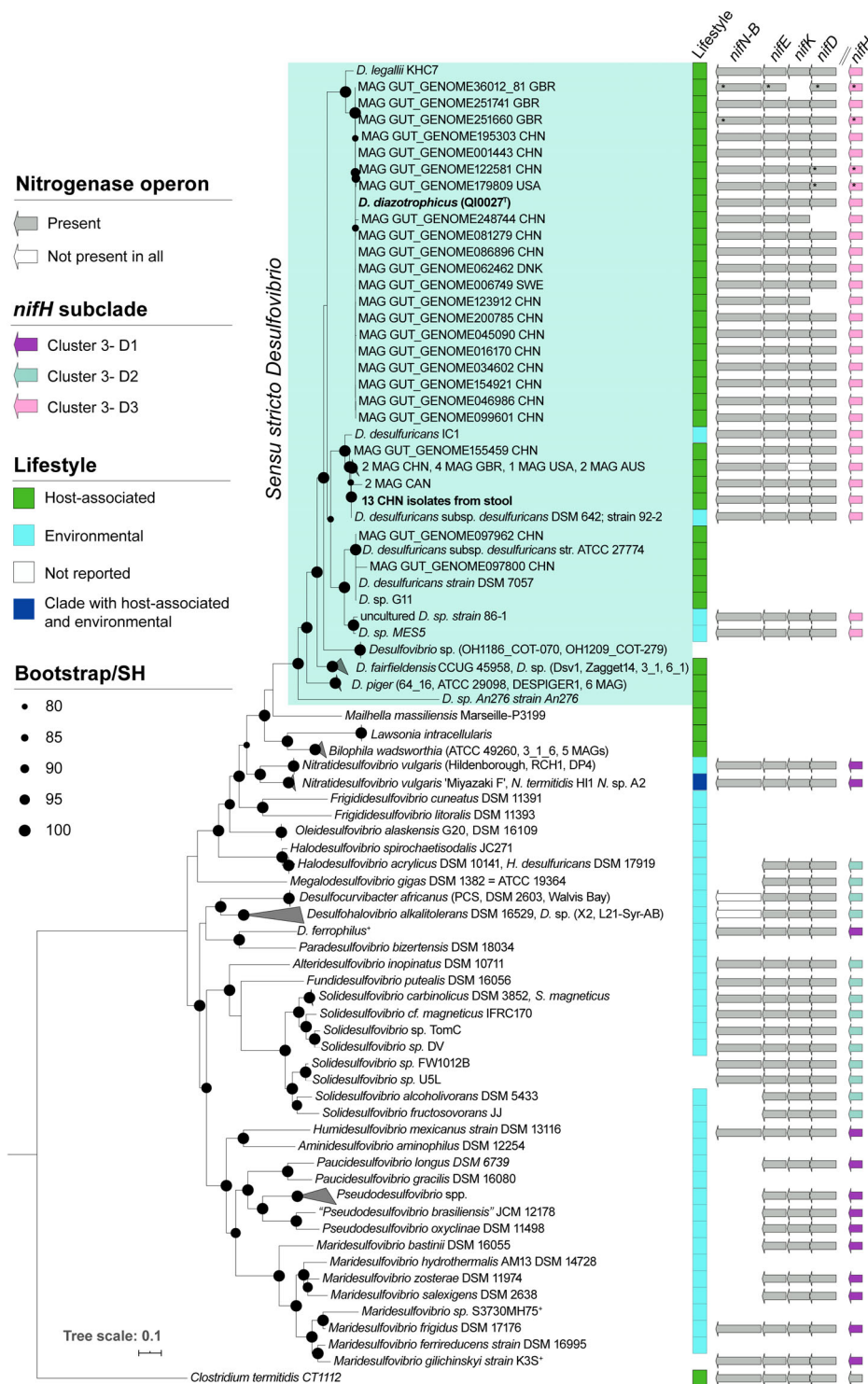


Fig. 1. Distribution of the nitrogenase gene cluster among *Desulfovibrionaceae* bacteria. A maximum-likelihood phylogenomic tree was reconstructed using the amino acid sequences encoded by 30 single-copy marker genes. The tree was rooted with *Clostridium termitidis* CT1112. Arrows represent the presence of genes encoding the dinitrogenase reductase (*nifH*), the molybdenum-iron (Mo-Fe) protein (dinitrogenase *nifD* and *nifK*) and the iron-molybdenum cofactor biosynthesis (*nifN-B* and *nifE*). The arrow representing *nifH* is colour coded depending on the *nifH* subclade where the gene clusters. The country of origin for each MAG is shown with the ISO3 code. *Genes recovered using targeted assembly for Q10027[†] MAGs. [†]Not reclassified by Waite *et al.* (2020).

confidence interval) and the difference in %G + C was 2.67. These values are well below the recommended threshold for species discrimination of <94%–96% ANI, <70% dDDH, as well as >1%G + C (Richter and Rossello-Mora, 2009; Meier-Kolthoff *et al.*, 2013; Meier-Kolthoff *et al.*, 2014). Thus, Q10027[†] and *D. legallii* are different

species. The family *Desulfovibrionaceae* has 19 described genera: *Desulfovibrio*, *Nitratidesulfovibrio*, *Paucidesulfovibrio*, *Oleidesulfovibrio*, *Pseudodesulfovibrio*, *Solidesulfovibrio*, *Maridesulfovibrio*, *Aminidesulfovibrio*, *Humidesulfovibrio*, *Frigididesulfovibrio*, *Megalodesulfovibrio*, *Alteridesulfovibrio*, *Fundidesulfovibrio*, *Paradesulfovibrio*, *Bilophila*, *Lawsonia*,

Halodesulfobivrio, *Desulfocurvus* and *Desulfobaculum*, although no representative genomes have been sequenced for the latter two genera (Fig. 1) (Waite *et al.*, 2020).

Physiology

Similar to other *Desulfovibrionaceae* bacteria, QI0027^T is a Gram-negative bacterium with a curved rod shape (Fig. 2). Cells were 1.4–5 µm long. QI0027^T grew at a temperature between 15 and 46°C, with an optimum growth temperature of 37°C. QI0027^T could grow with 0–0.51 M NaCl, with optimum growth at 0.068–0.171 M NaCl. This salinity concentration range is comparable with human physiological saline (0.9% wt./vol., 0.15 M) (Oxley *et al.*, 2010). Colonies on Postgate C plates were black due to the accumulation of iron sulphide precipitates (Supplementary Fig. 1a). Based on disc diffusion assays with antibiotics, QI0027^T was resistant to streptomycin (Table 1).

D. diazotrophicus from the gut can fix nitrogen 3167

Unlike the clearly motile *D. legallii*, QI0027^T was mostly non-motile but showed occasionally multiple lophotrichous flagella based on a spatial assay on soft agar and scanning and transmission electron microscopy (SEM and TEM) (see Fig. 2 and Supplementary Fig. 1b). SEM and negative staining combined with TEM showed that *D. legallii* showed a single polar flagellum with both EM techniques (Fig. 2). The flagella of QI0027^T were instead longer and thinner than those compared to *D. legallii*, which might explain why we could not observe them with SEM, as the sample processing is harsher on the cells compared to the negative staining (Fig. 2). QI0027^T encoded the genes for flagella synthesis, including the hook-basal body complex (*fliE*), the basal-body rod (*flgB*, *flgC*), the cytoplasmic components *fliI* and the membrane-bound components (*fliO*, *fliP*, *fliQ*, *flhA*, *flhB*, *fliR*) of the flagellar protein export apparatus (reviewed by Anderson *et al.*, 2010). QI0027^T also synthesized the negative

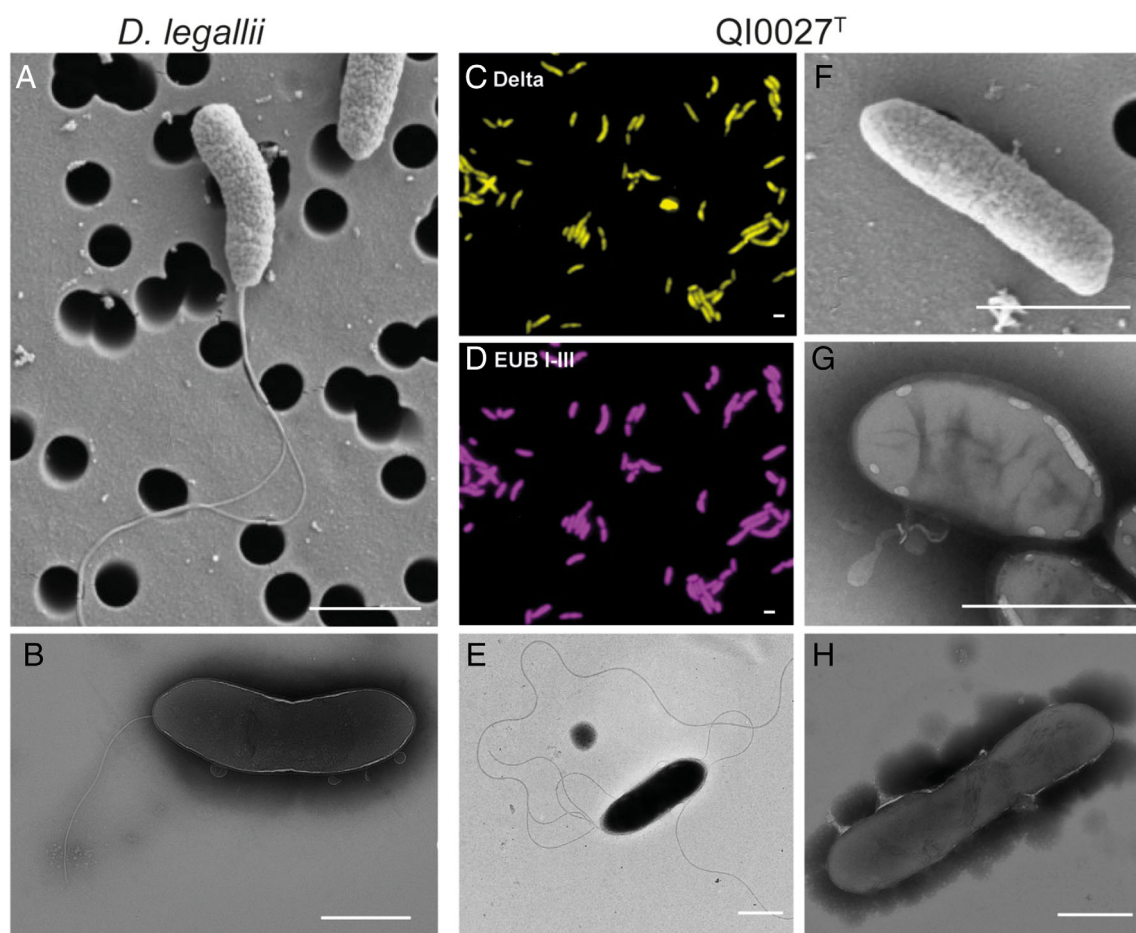


Fig. 2. Comparative microscopic characteristics of *D. legallii* and QI0027^T. Scanning electron microscopy (SEM) (A) and transmission electron microscopy (TEM) with negative staining (B) showed the presence of a single polar flagellum in *D. legallii*. Outer membrane vesicle-like structures could be observed in *D. legallii* (B). SEM (F) and TEM (G, H) show that most QI0027^T cells did not have flagella, with a few exceptions (E). Some cells also showed inclusion bodies, which might be polyhydroxyalkanoates (G) (Hai *et al.*, 2004). Scale bar = 1 µm.

regulator of flagellin synthesis *flgM*. We thus hypothesize that QI0027^T regulates the expression of flagella synthesis, which might be a key factor during the colonization and sensing of areas rich in nutrients in the gastrointestinal tract (Soutourina and Bertin, 2003).

Fatty acid analysis from QI0027^T and *D. legallii* showed that the main fatty acid was iso C_{17:1} (29% for QI0027^T and 24.2% for *D. legallii*), which is consistent with most *Desulfovibrio* species (Vainshtein et al., 1992). Other predominant fatty acids had also a similar composition and abundance for both strains: iso fatty acid methyl ester (FAME) C_{17:0} (24.9% for QI0027^T and 26.12% for *D. legallii*), FAME C_{15:0} (19.3% for QI0027^T and 20.6% for *D. legallii*) and iso C_{17:0} (17.2%–17.5% for QI0027^T and 19.9%–20.2% for *D. legallii*) (Supplementary Dataset 1). However, 10 fatty acids were detected at low abundance (<2%) only in QI0027^T, which likely reflects small differences in the metabolism of QI0027^T compared to *D. legallii* (Supplementary Dataset 1).

Distribution of QI0027-related species in human microbiome metagenomes

Metagenomic reads with >95% similarity to QI0027^T were detected in two further individuals based on sequencing of 45 stool metagenomes from Chinese donors using read mapping (see Supplementary Dataset 3 and Supplementary results and discussion – Identification of QI0027-related species in Chinese metagenomes). We successfully recovered a high-quality draft genome of QI0027^T species (ANI similarity of 99.49%) from metagenomic reads using a targeted assembly approach of *Desulfovibrionaceae* with the combined 45 stool metagenomic reads from Chinese donors. This targeted assembly approach was used to evaluate if the genome of this species could be recovered in high-complexity stool metagenomes. We were not able to recover any *D. desulfuricans* MAG, even though this was the species that could be isolated more often with our isolation efforts in Chinese stool samples (13 out of 14 *Desulfovibrionaceae* isolates).

We further identified 20 genomes from the same species as QI0027^T (ANI > 98%) using a large unified collection of metagenome-assembled genomes (MAGs) obtained from extensive diversity of lifestyles, ages and countries (Almeida et al., 2020). Samples included in the analyses were obtained from 170 samples from Africa, 3590 from Asia, 4586 from Europe, 3298 from North America, 262 from Oceania and 183 from South America (Almeida et al., 2020). The genome collection included 1690 *Desulfovibrionaceae* genomes classified as *D. piger*, *D. desulfuricans*, *D. fairfieldensis*, *Desulfovibrio* sp., *Bilophila* sp., *B. wadsworthia* and *Mailhella*

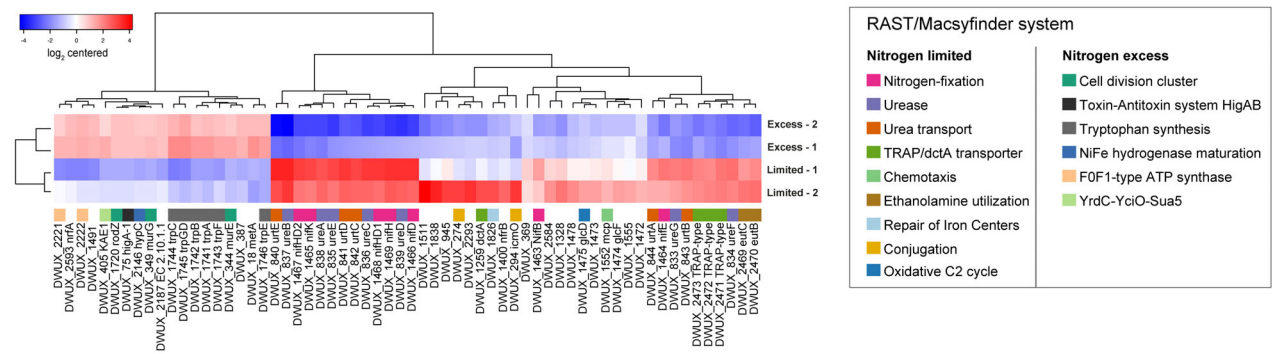
sp. Genomes from the same species as QI0027^T were assembled from stool metagenomes obtained from 14 donors residing in China, three in the UK, one in Sweden, one in Denmark and one in the USA (Metagenome studies from Qin et al., 2014; Bäckhed et al., 2015; Xie et al., 2016; He et al., 2017; Li et al., 2017; Wen et al., 2017). Since genomes of organisms that are more abundant are easier to recover from metagenomes (Albertsen et al., 2013), the distribution of QI0027^T among humans from diverse continents suggests that this species might be more prevalent in Asian individuals.

QI0027^T can fix nitrogen

Genome analysis of the strain QI0027^T revealed the presence of the minimal gene set for nitrogen fixation (Dos Santos et al., 2012). These genes included the key gene *nifH*, encoding the dinitrogenase reductase; *nifD* and *nifK*, encoding the MoFe dinitrogenase; as well as *nifE*, *nifN* and *nifB*, encoding the FeMo cofactor biosynthesis machinery. The *nifN* and *nifB* genes were fused into a single gene (*nifN-B*) and two genes encoding PII-like nitrogen regulatory proteins (*glnB1* and *glnB2*) were encoded between *nifH* and *nifD* as observed in *Clostridium* and other diazotrophic *Desulfovibrio* (Supplementary Fig. 2) (Chen et al., 2001; Seefeldt et al., 2009). To physiologically validate the nitrogen fixation ability of strain QI0027^T, we checked the expression pattern of the genes encoding the nitrogen fixation machinery under conditions of nitrogen limitation compared to nitrogen excess. As expected, we observed that all genes in the *nif* operon were significantly induced under nitrogen limiting conditions (Fig. 3A and B).

We demonstrated that strain QI0027^T was able to fix nitrogen based on an acetylene reduction assay performed in similar growth conditions used for the RNA-seq transcription analysis (Fig. 4). The specific nitrogenase activity rate for strain QI0027^T was estimated as 7.6 ± 5.2 nmol ethylene · mg protein⁻¹ · h⁻¹. As expected, no acetylene reduction was observed when the growth media was supplemented with yeast extract or with 18.54 mM NH₄Cl (referred to as nitrogen excess). To our knowledge, this is the first *Desulfovibrio* isolate from the human gut for which diazotrophy has been physiologically demonstrated.

Further evaluation of our whole transcriptome analysis revealed that the strain QI0027^T overexpressed 43 genes under nitrogen-limiting conditions, while an additional 19 genes were overexpressed under nitrogen excess conditions (Fig. 3A and Supplementary Dataset 2). Under nitrogen excess conditions, QI0027^T had a higher expression of genes related to cell cycle and cell division, RNA processing, glycolysis, respiration, as well as cobalamin, ATP, and tryptophan synthesis (Fig. 3A and B).

A Differentially expressed genes \log_2 , centered

B Differentially expressed genes

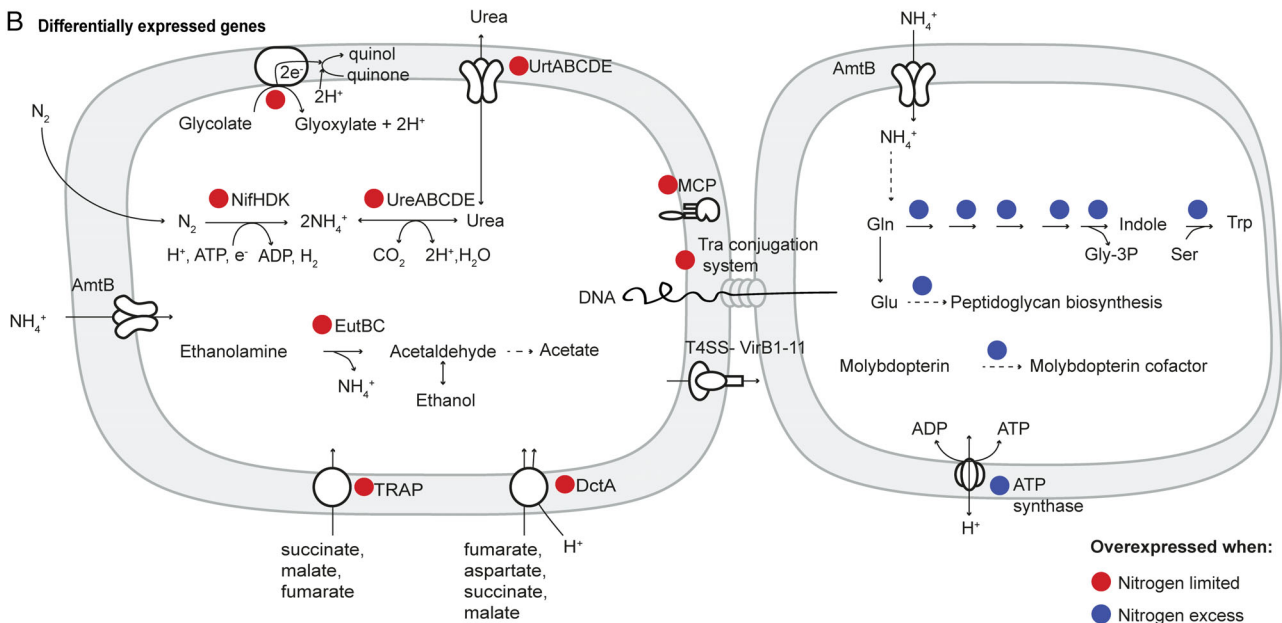


Fig. 3. Transcriptome expression under nitrogen limiting or excess conditions.

A. Differentially expressed genes by QI0027^T cultures with limiting or excess nitrogen. p -value cut-off for FDR = 0.05; minimum fold change = 3. B. Overview of differentially expressed pathways under nitrogen-limited or excess conditions by QI0027^T cultures. Under nitrogen limiting conditions, genes related to nitrogen fixation, urea production and urea transport are overexpressed. Under nitrogen excess conditions, tryptophan synthesis, peptidoglycan biosynthesis and molybdopterin cofactor are overexpressed. p -value cut-off for false discovery rate (FDR) = 0.05; minimum fold change = threefold. MCP = methyl-accepting chemotaxis protein; Gln = glutamine; Trp = Tryptophan; Gly-3P = Glyceraldehyde 3 phosphate. Reactions are not balanced.

Moreover, QI0027^T overexpressed the antitoxin HigA, a system that has been linked to survival response during environmental and chemical stresses, including amino acid starvation, growth and programmed cell death (Schureck *et al.*, 2014).

Besides the nitrogen fixation genes overexpressed under nitrogen limiting conditions discussed above, we observed overexpression of the urease enzyme complex *ureABCEFGD* and the reversible urea uptake system *urtABCDE*. *ureABCEFGD* is used to reversibly transform urea into ammonia and CO₂. Since QI0027^T cultures with limiting nitrogen did not have urea in the media, ammonia resulting from N₂ fixation was likely converted to urea and the excess excreted. Urea is less toxic than

ammonia, more soluble, and can be used as an intermediate molecule for nitrogen excretion (Levin *et al.*, 2009). Alternatively, some QI0027^T cells might have scavenged urea to use it as a nitrogen source. Two transporters, from the TRAP and DctA families, which can be used for the import of organic acids (Kelly and Thomas, 2001), were overexpressed under nitrogen-limiting conditions, suggesting that efficient transport of dicarboxylates is required to support or regulate nitrogen fixation (Dixon and Kahn, 2004; Huergo and Dixon, 2015; Bueno Batista and Dixon, 2019).

Interestingly, under nitrogen-limiting conditions, QI0027^T overexpressed one methyl-accepting chemotaxis protein (MCP) (p -value <0.05 and FDR <0.05) and

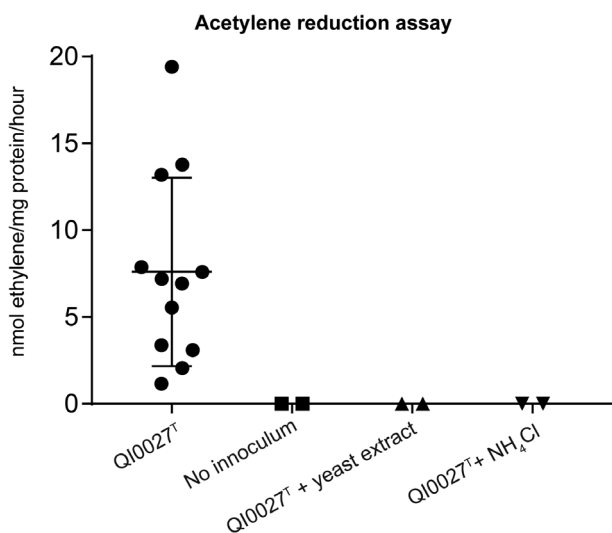


Fig. 4. Acetylene reduction assay under nitrogen limiting or excess conditions. Acetylene reduction assay was performed under anaerobic conditions in Hungate tubes.

three MCPs were overexpressed under nitrogen-excess conditions (p -value <0.05 but FDR >0.05) (Supplementary Dataset 2). MCP proteins can be used as chemical sensors (Ortega *et al.*, 2017). The MCP overexpressed under nitrogen-limiting conditions has a ligand-binding domain of Tar and Tsr chemoreceptors, which could potentially mediate responses towards optimal oxygen, redox and oxidizable substrates (Ortega *et al.*, 2017) that may support biological nitrogen fixation.

Prevalence of diazotrophy among *Desulfovibrionaceae*

Several members of the *Desulfovibrionaceae* family possessed the minimal *nif* gene set (*nifH*, *nifD*, *nifK*, *nifE* and *nifN-B*) for nitrogen fixation but in some cases, these genes may have been lost in several lineages (Fig. 1) (see Supplementary results and discussion - *nifH* was likely acquired through horizontal gene transfer). We searched for the genes related to nitrogen fixation among all available *Desulfovibrionaceae* genomes using tblastn and combined the gene presence/absence with our phylogenomic reconstruction of the *Desulfovibrionaceae* bacteria. Among the lineages that lack the nitrogen-fixing genes, we identified genera that have been linked to disease (e.g. *Bilophila* and *Lawsonia*). Within the *Desulfovibrionaceae*, neither the closed genome of *D. piger* nor other almost complete *D. piger* or *D. fairfieldensis* genomes encoded the key gene for nitrogen fixation *nifH*. These *Desulfovibrionaceae* species are the most abundant and widespread in Western individuals (Carbonero and Gaskins, 2015). *Desulfovibrionaceae* *desulfuricans* isolates obtained from free-living environments, as well as our

own 13 *D. desulfuricans* isolates obtained from eight Chinese donors, encoded the genes to fix nitrogen. All these *D. desulfuricans* strains grouped closer to *D. legallii* and QI0027^T.

From all *Desulfovibrionaceae* MAGs recovered from human metagenomes (Almeida *et al.*, 2020), only *D. desulfuricans* and *Desulfovibrio* MAGs that belong to the same species as QI0027^T based on an ANI $>98\%$ encoded the *nifHDKEN-B* genes (Fig. 1). Twelve *D. desulfuricans* MAGs, and 20 QI0027^T MAGs encoded *nif* genes. Four of these 20 QI0027^T MAGs originally assembled by Almeida *et al.* (2020) did not include the *nif* genes. However, using a targeted assembly approach with the metagenomic reads and the QI0027^T reference genome, we recovered the nitrogenase genes for the four QI0027^T MAGs. Based on phylogenetic analysis, the *nifH* gene grouped closer to *nifH* encoded by QI0027^T showing that using stringent mapping was enough to recover the *nifH* genes from QI0027^T. In agreement with our genome searches on the closed genome of *D. piger*, as well as on draft genomes of isolates from *D. fairfieldensis*, *Bilophila* and *Lawsonia*, none of the human-derived MAGs from these *Desulfovibrionaceae* species encoded the nitrogenase genes (Fig. 1). From the isolates and MAGs that have been sequenced thus far from *Desulfovibrionaceae* that encode the genomic repertoire for nitrogen fixation, only *D. legallii*, which was isolated from a shoulder joint infection, *D. desulfuricans* and QI0027^T species have been isolated from mammals (Fig. 1).

Expression of nitrogen fixation genes in human stool metatranscriptomes

To identify if *Desulfovibrio* bacteria can express the nitrogen fixation genes in the human gut environment, we searched for *nif* genes transcripts in 835 stool metatranscriptomes available through the Human Microbiome Project database (Consortium, 2019; Lloyd-Price *et al.*, 2019). Metatranscriptome reads more than 95% similar to *Desulfovibrio nif* genes were detected in six samples. From these six faecal samples, reads aligning to *nifH* from QI0027^T were present in two metatranscriptomes. In addition, reads aligning to *nifH* from *Desulfovibrio desulfuricans* subsp. *desulfuricans* DSM 642 were found in one sample (Supplementary Dataset 4). The low abundance of *Desulfovibrio nif* genes transcripts detected in these metatranscriptome datasets may reflect the abundance of diazotrophic *Desulfovibrio* in the gut microbiota. As the expression of nitrogen fixation genes is often controlled by oxygen, oxygen exposure during stool sample preparation may have also contributed to the low abundance of *nif* transcripts detected. The detection of

Desulfovibrio nif transcripts in the human gut metatranscriptome might indicate that nitrogen fixation can occur by *Desulfovibrio* in the gut.

Discussion

We have presented evidence that QI0027^T is a novel species with nitrogen-fixing ability. This novel species was misclassified by commonly used taxonomic classifiers that rely on clade-specific marker genes from metagenomes and could only be identified using whole-genome sequencing and assembly (of MAGs or pure cultures). Our study combines cultivation, genome sequencing, physiological characterization, differential expression analysis and metagenome sequencing to overcome the bias inherent to each technique.

Based on our acetylene reduction assay and differential expression analyses, strain QI0027^T fixed nitrogen only when we did not provide a source of fixed nitrogen in the media, such as yeast extract or NH₄Cl, which are known to inhibit the metabolically expensive process of nitrogen fixation (Hardy *et al.*, 1968). Our results show therefore that the genes for nitrogen fixation are functional in strain QI0027^T but that their activity can be inhibited depending on fixed nitrogen availability. The rate of acetylene reduction from QI0027^T was in the range of 1.16–13.18 nmol ethylene · mg protein⁻¹ · h⁻¹, which would be equivalent to ~0.001–0.261 nmol ethylene · g stool⁻¹ · h⁻¹ (for assumptions see Materials and methods). Igai *et al.* (2016) showed a rate of acetylene reduction in unfrozen stool samples of 0.055–1.608 nmol ethylene · g stool⁻¹ · h⁻¹, while Bergersen and Hipsley (1970) detected 0.75 nmol ethylene · g stool⁻¹ · h⁻¹ acetylene reducing activity in one European stool sample. Our extrapolation of the rate of acetylene reduction per gram of stool shows that the nitrogen fixation ability of QI0027^T could contribute at rates similar to those observed in human stool samples attributed previously to only other taxonomic groups (Igai *et al.*, 2016).

The rate of nitrogen fixation by QI0027^T was lower compared to the environmental isolates *D. desulfuricans*, *Nitratidesulfovibrio vulgaris*, *Megalodesulfovibrio gigas*, *Maridesulfovibrio salexigens*, *Desulfocurvibacter africanus* and *Dissulfuribacter thermophilus* (7.6 ± 5.2 vs. 42–918 nmol ethylene · mg protein⁻¹ · h⁻¹) (Postgate and Kent, 1985). Using this assay, three out of 15 environmental strains belonging to the *Desulfovibrionaceae* family were shown not to fix nitrogen. The authors hypothesized that the right conditions for nitrogen fixation might have yet to be found. However, this pattern can be explained now as a true absence of nitrogen fixation ability since the key gene for nitrogen fixation, *nifH*, has been gained and lost several times in the *Desulfovibrionaceae*

(Fig. 1, see Supplementary results and discussion – *nifH* was likely acquired through horizontal gene transfer).

Only a few members of the *Desulfovibrio* encoded *nifH*. The ability to fix nitrogen is not present in the *Desulfovibrionaceae* that are the most abundant in the gut but rather prevails in *Desulfovibrio* species that tend to be in lower abundance. *Desulfovibrio desulfuricans* is often recovered from human stool samples (Coutinho *et al.*, 2017). Some of these *D. desulfuricans* strains might also fix nitrogen, such as the case of our 13 *D. desulfuricans* Chinese gut isolates that encoded the nitrogenase operon. However, not all strains of *D. desulfuricans* encode the genes to fix nitrogen (Fig. 1). Indeed, *D. desulfuricans* formed two major clusters with an 80%–85% ANI between the two clusters, indicating that these species will need to be reclassified in future studies. The contribution to nitrogen fixation in the human gut by *Desulfovibrionaceae* is therefore likely not limited to QI0027^T.

Nitrogen fixation in a nutrient-rich environment

The human gut is a nutrient-rich environment in which most bioavailable nitrogen comes from amino acids present in food. Expression of nitrogen fixation-related genes affiliated to *Desulfovibrio* bacteria was found on six stool samples. However, read counts were low and it is difficult to relate gene expression in stool, which is often exposed to oxygen, to *in situ* activity within the gut for an oxygen-sensitive process like nitrogen fixation. Active nitrogen fixation in the human gut would suggest that this process is not only a potential survival strategy when outside a mammal host. If nitrogen fixation is a high-energy demanding process (Rascio and La Rocca, 2013), and the gut is nutrient-rich, which conditions could favour the persistence of *Desulfovibrio* species that can fix nitrogen?

It is commonly assumed that N₂ fixation can occur only when bioavailable nitrogen is limiting (<1 μM) (Knapp, 2012). However, nitrogen fixation can occur in environments with relatively high input of fixed nitrogen such as oceanic waters with high nitrate or ammonia concentrations (30 μM NO₃⁻; 200 μM NH₄⁺) as well as the human gut (12–30 mM ammonia in stool, which can be increased with higher protein consumption) (Knapp, 2012; Scott *et al.*, 2013; Igai *et al.*, 2016; Shiozaki *et al.*, 2018). Humans belonging to a population from Papua New Guinea that had low protein intake (e.g. fixed nitrogen), as well as Japanese individuals that had a higher protein intake, were shown to have a gut microbiome that was actively fixing nitrogen (Igai *et al.*, 2016). Nitrogen fixation was independent of the habitual nitrogen intake and was confirmed based on the acetylene reduction assay or ¹⁵N incorporation (Bergersen and

Hipsley, 1970; Igai *et al.*, 2016). In these human populations, nitrogen fixation by the microbiome corresponded to only 0.01% of the standard nitrogen requirement for humans. Nitrogen fixation was attributed to *Klebsiella* and *Clostridia* bacteria based on *nifH* sequences recovered from cloning and metagenomic analysis, including *Clostridia* encoded *nifH*-Cluster 3 sequences recovered from Chinese faeces (Igai *et al.*, 2016). However, the microbial contribution to the nitrogen intake from humans could be population specific depending on the dietary habits, especially when low protein diets are consumed.

Based on our acetylen reduction assay, 18.54 mM ammonium inhibited nitrogen fixation, which is within the range of ammonium detected in human stool (12–30 mM) (Scott *et al.*, 2013). Ammonium is, however, rapidly absorbed and metabolized to urea by the liver. Thus, although the gut may have on average an excess of nitrogen because of the host diet, there may be micro-niches where bioavailable nitrogen is locally limiting or a specific nitrogen-limited diet may result in sub-inhibitory levels of ammonia. More than 90% of nitrogen is absorbed in the small intestine and studies using germ-free animals have shown that most gut microbes increase the protein requirement of the host (Fuller and Reeds, 1998; Reese *et al.*, 2018). In the large intestine, the bacterial load increases which could create micro-niches where fixed-nitrogen becomes limiting (Dai *et al.*, 2011). The host might secrete nitrogen in the large intestine to control the microbial load and select for preferred microbial taxa (Reese *et al.*, 2018). Therefore, having nitrogen fixation ability would still be a selective advantage. SRB must compete for resources and space to thrive in the gut. Hydrogen, the most used energy source by SRB, is also used by acetogens and methanogens. Niche partitioning could be a mechanism to cope with competitors, as has been observed for even members of the same species (Sayavedra *et al.*, 2015; Ailloud *et al.*, 2019; Ansoorge *et al.*, 2019). Fixed nitrogen could then be used for the synthesis of amino acids, which in turn could be used as a common exchange currency with other members of the gut microbiome or the host (Lesser *et al.*, 2007; Mohamed *et al.*, 2008).

An alternative explanation for the presence of diazotrophs in environments with nitrogen excess is that nitrogen fixation is used to maintain an ideal intracellular redox state, analogous to bacteria which can grow heterotrophically but still retain energetically expensive CO₂ fixation pathways like the Calvin–Benson cycle (Tichi and Tabita, 2000; McKinlay and Harwood, 2010). This has been interpreted as a means to regenerate electron carriers like NAD⁺ under anaerobic conditions, where there can be an excess of reducing equivalents (Bombar *et al.*, 2016). Nitrogen fixation is also energetically

expensive and consumes reducing equivalents. So far, the use of N fixation as an electron sink has only been proposed for phototrophs, but perhaps nitrogen fixation could fulfil a similar function in anaerobic niches of the gut, with the additional benefit of producing bioavailable nitrogen. Further work will be required to disentangle which factors trigger nitrogen fixation by *Desulfovibrio* in the human gut and to understand the contribution of bioavailable nitrogen through nitrogen fixation, compared to other nitrogen assimilation pathways such as the import of free amino acids, ammonium, urea, nitrate and DNA.

Description of *Desulfovibrio diazotrophicus* sp. nov.

Based on the phylogenomic and physiological data presented in this study, we propose that strain QI0027^T is a novel species which we named *Desulfovibrio diazotrophicus* (di.a'zo.tro'phi.cus. Gr. prefix *dis* two, double; N.L. n. *azotum* nitrogen; Gr. adj. *trophikos* nourishment, relating to nutrition; N.L. masc. adj. *diazotrophicus* one that grows on dinitrogen). QI0027^T cells are non-spore-forming, Gram-negative rods to spirochete shaped bacteria from the *Desulfovibrionales* order, *Desulfovibrionia* class. Multiplies by binary fission. Cells are 1.4–5 µm long and show only sometimes lophotrichous flagella. After 2–7 days at 37°C on Postgate C media, colonies are 0.4–1.9 mm in diameter, pale brown with slightly grey hints to black in colour, round and with a raised elevation. QI0027^T grows at 15–46°C, (optimum at 37°C), at pH 4.5–6.5 (optimum pH 5.9), and with 0–0.51 M NaCl (optimum 0.068–0.171 M NaCl). An obligate anaerobe that respire sulfate using either hydrogen, carbon monoxide or alcohols as electron donors and that can fix nitrogen (Supplementary Fig. 3). Reaction to catalase was negative. QI0027^T can utilize L-rhamnose, D-fructose, L-fucose, D-galactose, D-galacturonic acid, palatinose, D,L-lactic acid, L-lactic acid, D-lactic acid methyl ester, L-malic acid, pyruvic acid, methyl pyruvate, L-glutamic acid, glucose and D-mannose (Supplementary results and discussion). The predominant cellular fatty acids are iso C_{17:1ω9c} (29%), iso fatty acid methyl ester (C_{17:0} (24.9%) and C_{15:0} (19.3%)), and iso C_{17:0} (17.2%–17.5%).

The type strain, QI0027^T (=DSMZ 109475^T = NCIMB 15228^T = NCTC 14354^T), was isolated from a faecal sample donated by a 25-year-old male healthy human in Wuxi, China. The G + C content of the type strain is 62.1%.

Materials and methods

Isolation

Strain QI0027^T was isolated from a human stool sample in Wuxi, China in June 2016. The 25-year-old male donor was healthy. The donor was provided with a stool

collection kit, which included an anaerobic bag, an anaerobic pack (Mitsubishi Gas Chemical, Japan) and ice. The sample was placed inside an anaerobic cabinet within 30 min of collection. A 1 g stool sample was homogenized and used for dilution (ranging from 1:10 to $1:1 \times 10^7$) using anaerobic $1 \times$ PBS. Dilutions were plated onto anaerobic Postgate C agar plates, which contained per litre of distilled water: 6 g sodium lactate; 4.5 g Na_2SO_4 , 1 g NH_4Cl , 1 g yeast extract, 0.5 g KH_2PO_4 , 0.3 g sodium citrate, 0.06 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.004 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 4 ml resazurin (0.02% wt./vol.), 0.04 g $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.5 g L-Cysteine hydrochloride and 15 g agar. pH was adjusted to 7.5 ± 0.1 using 5 M HCl and the media was sterilized by autoclaving. Plates were incubated at 37°C for 2 weeks. Single black colonies were selected and subcultured two more times until a pure culture of QI0027^T was obtained. The purity of the isolate was confirmed by colony morphology, light microscopy, and by checking that all cells in the culture hybridized to the deltaproteobacteria-specific fluorescence *in situ* hybridization probe Delta495a (Lucker *et al.*, 2007) as described by Amann and Fuchs (2008). Additionally, 16S rRNA sequencing from colonies that showed slightly different pigmentation or size was done to confirm the purity of QI0027^T. The 16S rRNA gene was amplified with the AMP_F (5'-GAG AGT TTG ATY CTG GCT CAG) and AMP_R (5'-AAG GAG GTG ATC CAR CCG CA) set of primers according to the method from Baker *et al.* (2003). Purified amplicons were sequenced using the Mix2Seq Kit Overnight service (Eurofins, Germany). To confirm the identity of QI0027^T, the 16S rRNA sequence was compared against the Silva SSU r138 database (Quast *et al.*, 2013). For long-term storage, strain QI0027^T was stored at -80°C in Protect Select tubes for anaerobes with cryobeads (TS/73-AN80, Technical Service Consultants, UK).

DNA extraction, whole-genome sequencing and assembly

DNA was extracted using the GenElute Bacterial Genomic DNA kit (NA2100, Sigma-Aldrich, UK) following the manufacturer's instructions. For whole-genome sequencing, DNA was sent to Earlham Institute, Norwich, UK, where a DNA library was constructed using a modified version of Illumina's Nextera protocol known as Low Input Transposase Enabled libraries (Beier and Gross, 2006). 947 409 paired-end 150 bp reads were generated using the Illumina HiSeq 4000 platform. Reads were quality trimmed with bbdduk (v.38.06) (sourceforge.net/projects/bbmap) using a minimum quality of 3. For long-read sequencing, a DNA library was prepared using the native barcoding genomic DNA expansion 1–12 kit (with EXP-NBD104, EXP-NBD114 and SQK-LSK109,

Nanopore, UK) according to the manufacturer's instructions and sequenced using a MiniION platform (Quadram Institute Bioscience, Norwich, UK). To obtain a high-quality closed genome, we used the hybrid assembler Unicycler (v.0.4.8) and polished the genome with Pilon (v.1.23 (Walker *et al.*, 2014; Wick *et al.*, 2017)). The hybrid assembly resulted in a single chromosome. We confirmed that the 5' and 3' extremes of the scaffold were connected, and hence the genome was circular, by visualization of the assembly graph with Bandage (Wick *et al.*, 2015). Genome statistics and quality checks for QI0027^T and *D. legallii* KHC7 were done with Patric (see Table 1) (Parrello *et al.*, 2019). Genome sequence-based species delimitation metrics were determined using fastANI (Jain *et al.*, 2018) and GGDC (v.2.1) (Meier-Kolthoff *et al.*, 2013; Meier-Kolthoff *et al.*, 2014). We report the dDDH value estimated with the recommended formula 'identities/HSP length'. The genome was annotated with PATRIC RASTtk (Brettin *et al.*, 2015).

Phylogenomics

Genomes from the *Desulfovibrionaceae* family were obtained from PATRIC (Wattam *et al.*, 2014) and NCBI. Phylogenomic analyses were conducted with the scripts available from phylogenomics-tools (github.com/kbseah/phylogenomics-tools, doi:10.5281/zenodo.46122). Conserved marker proteins that are mostly conserved across bacteria were extracted with the Amphora2 pipeline (Wu and Scott, 2012). The amino acid sequences encoded by the following 30 marker genes were aligned with Muscle (Edgar, 2004): *frr*, *infC*, *nusA*, *pgk*, *pyrG*, *rplA*, *rplB*, *rplC*, *rplD*, *rplE*, *rplF*, *rplK*, *rplL*, *rplM*, *rplN*, *rplP*, *rplS*, *rplT*, *rpmA*, *rpoB*, *rpsB*, *rpsC*, *rpsE*, *rpsI*, *rpsJ*, *rpsK*, *rpsM*, *rpsS*, *smpB* and *tsf*. Positions of the alignment that were not present in at least 75% of the sequences were removed with Geneious V9 (Kearse *et al.*, 2012). The best evolutionary amino acid substitution model for each marker gene was estimated with RaxML (Stamatakis, 2006). Single amino acid alignments were concatenated to reconstruct a phylogenomic tree with SH-like aLRT support values using RaxML and a partitioned alignment (Stamatakis, 2006). Pairwise ANI between the *Desulfovibrionaceae* genomes was estimated with fastANI (Jain *et al.*, 2018).

Metagenome sequencing, microbial profiling of stool samples and detecting the presence of QI0027^T in Chinese metagenomes

Forty five humans from China were provided with a stool collection kit as described above. Samples were stored at -80°C upon 48 h of collection. DNA was extracted from the stool samples using the FastDNA SPIN kit for

faeces (MP Biomedicals, Shanghai, China) according to the manufacturer's instructions using approximately 50 mg of stool material per individual. DNA libraries were prepared using the NEB Next Ultra DNA Library Prep kit for Illumina (New England BioLabs). Approximately 10 Gb were sequenced with a HiSeq X Ten instrument as paired-end 150 bp reads. Library preparation and sequencing was done by Novogene (China). Metagenomic reads were quality trimmed with a minimum quality of 2 and human host reads were removed using as reference the human genome GCA_000001405 with bbdduk. These clean reads were used to search for QI0027^T by mapping the reads against QI0027^T genome assembly with bbmap (v.38.43) using >95% mapping identity. We considered that QI0027^T species was present when >70% of the reference genome was covered by the reads.

Recovery of QI0027^T from Chinese metagenomes and MAGs

Clean reads of the 45 stool metagenomes were combined and mapped against all publicly available and our own *Desulfovibrionaceae* genomes. The resulting assembly from the combined reads was separated into genomes using an unsupervised binning method based on nucleotide composition, differential coverage and linkage data from paired-end reads (Alneberg *et al.*, 2014). The resulting reads were then assembled with metaSPAdes (v.3.13.0) (Nurk *et al.*, 2016) and binned with CONCOCT (v.1.1.0) (Alneberg *et al.*, 2014). Genome completeness was estimated using CheckM (v.1.0.13) with the marker lineage *Deltaproteobacteria* (UID3218) (Parks *et al.*, 2015) and identity was determined with GTDB-Tk (v.1.0.2) using the database release 89 (Parks *et al.*, 2019). Binning allowed us to recover a total of four *Desulfovibrionaceae* genomes with a completeness $\geq 97.34\%$ and a contamination $\leq 8.67\%$, including QI0027, *Bilophila wadsworthia*, *D. fairfieldensis* and *D. piger*. The relative abundance of reads mapping to the same species as QI0027^T was determined using bbmap (v.38.43) as the number of reads mapping to the genome/total number of reads.

The comprehensive compilation of MAGs from human microbiome sequencing efforts by Almeida *et al.* (2020) was also exploited to recover *Desulfovibrionaceae* bacteria genomes (opendata.lifebit.ai/table/?project = SGB and ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes, accessed on March 2020). These MAGs were obtained from the following countries: Australia, 24; Austria, 156; Bangladesh, 179; Canada, 300; China, 2133; Denmark, 1093; El Salvador, 87; Estonia, 219; Fiji, 238; Finland, 384; France, 157; Germany, 345; Hungary, 1; Iceland, 1; Ireland, 6; Israel, 954; Italy, 28; Japan, 31;

Kazakhstan, 172; Madagascar, 111; Mongolia, 109; Netherlands, 526; Norway, 1; Peru, 96; Russia, 285; Singapore, 12; Slovakia, 3; Spain, 395; Sweden, 640; United Kingdom, 345; United Republic of Tanzania, 59; and United States, 2997 (scripts to summarize the number of samples used by Almeida *et al.* (2020) are available under github.com/lSayaved/NitFix). Further metadata for these genomes was retrieved using the *curatedMetagenomicData* R/Bioconductor package (Pasolli *et al.*, 2017). To find genomes from the same species as QI0027^T, we used ANI similarity of all the *Desulfovibrionaceae* MAGs against the QI0027^T genome with fastANI (Jain *et al.*, 2018).

Detection of genes related to nitrogen fixation and phylogenetic analysis

We searched for the nitrogen fixation related genes among the *Desulfovibrionaceae* genomes included in our phylogenomic tree, as well as the *Desulfovibrionaceae* MAGs from the collection by Pasolli *et al.* (2019). We used as query the amino acid sequences encoded by *nifH*, *nifD*, *nifK*, *nifE* and *nifB* from QI0027^T against the isolate genomes and MAGs with tblastn (v.2.2.31+) to be able to detect partial genes at the end of scaffolds (>40% similarity, >40% query coverage, minimum alignment length >50). To improve the genome assembly of the four MAGs that did not encode the nitrogenase cluster, we used read mapping against the QI0027^T reference genome using bbmap (v.38.43) with a similarity $\geq 95\%$ against the reads from projects YSZC12003_3554, SRR3736997, H1M313811 and YSZC12003_36012 (Short Read Archive). Results were integrated with the phylogenomic tree with iTol (Letunic and Bork, 2016).

For phylogenetic analysis, we annotated all genomes with Prokka (v.1.14.0) and retrieved *nifH* amino acid sequences. Representative *nifH* amino acid sequences from Clusters 1–4 were retrieved from Gaby and Buckley (2014). Sequences were aligned using MAFFT v. 1.3.7 and realigned with ClustalW v2.1. The alignment was masked to remove alignment positions with >75% gaps. A maximum-likelihood tree reconstruction was obtained using FastTree (v. 2.1.5) with SH-like branching support values (Price *et al.*, 2009).

Physiology of QI0027^T

pH, salinity, temperature and motility. Physiological testing was done by adjusting Postgate C media without agarose in Hungate tubes (SciQuip, UK) and diluting cells 100-fold. To identify the pH range at which QI0027^T could grow, pH was adjusted by injecting deoxygenated 1.34 M HCl for acidic pHs, or 0.16 M NaOH with 10% NaHCO₃ (wt./vol.) for alkaline pHs before autoclaving. Final pH

was measured before inoculation (range from 3.98 to 8.83) and after the growth of the cells. QI0027^T grew in media with a pH range between 4.5 and 6.5, and the pH of the media became more basic after cell growth ranging from 6.8 to 7.2. To determine the tolerance of QI0027^T to different salinity concentrations, growth at 0–0.513 M NaCl was investigated in liquid Postgate C medium as described previously (Lopez-Cortes *et al.*, 2006). pH and salinity tolerance were investigated at 37°C for 5 days. To test for the effect of temperature on growth, duplicate cultures of strain QI0012 were incubated in 10 ml Hungate tubes at a temperature range of 10–59°C for up to 7 days. Growth was monitored with a turbidimeter CO 8000 (Biochrom, UK). Motility of QI0027^T was tested using semisolid Postgate C agar (5 g L⁻¹ of agar), using as reference for diffusion the motile *D. legallii*. The semi-solid media was stabbed with disposable inoculating loops and the diffusion was observed after 3–5 days.

Antibiotic resistance. Antibiotic resistance was examined by disc diffusion assays on Postgate C agar according to the standard procedure outlined in the National Committee for Clinical Laboratory Standards guidelines (Wayne, 1999). Inside an anaerobic cabinet, a sterile swab was dipped in a bacterial cell suspension grown overnight and spread on reduced Postgate C agar media. Antibiotic discs (Oxoid, Thermo Fisher Scientific, UK) were placed on the surface using sterile forceps. Plates were incubated for 48–72 h at 37°C. The following antibiotics were tested: ampicillin (25 µg), chloramphenicol (10 µg), nalidixic acid (30 µg), erythromycin (30 µg), tetracycline (30 µg) and streptomycin (25 µg).

Cell morphology, FISH and transmission electron microscopy. Morphological properties, such as cell shape, cell size and motility were observed by phase-contrast light microscopy (Olympus BX60 microscope). For fluorescence microscopy, a Zeiss Axio Imager M2 microscope was used with a 100× objective lens. Gram staining was done with the Gram staining kit (Sigma-Aldrich) according to the manufacturer's instructions. To investigate the presence of flagella, negative staining combined with transmission electron microscopy (TEM) was used. Briefly, a small drop of bacterial cell suspension was applied to a formvar/carbon-coated copper TEM grid (Agar Scientific, Stansted, UK) and left for 1 min. Excess liquid was removed with filter paper. A drop of 2% aqueous phosphotungstic acid was applied to the grid surface and left for 1 min. Excess stain was removed with filter paper and the grids were left to dry thoroughly before viewing in the transmission electron microscope. Grids were examined and imaged in an FEI Talos F200C TEM at 200 kV with a Gatan One-View digital camera. For scanning electron microscopy (SEM), cells were

prepared by critical-point drying (CPD) onto filter membranes (Kaláb *et al.*, 2008). Briefly, 100 µl of bacterial cell suspension was added dropwise onto an isopore membrane polycarbonate filter (HTTP01300, Millipore, UK). The filter was trimmed beforehand with a razor blade to identify the surface with inoculum. Cells were left to adhere to the surface for 5 min after which the filters were placed in 2.5% glutaraldehyde in 0.1 M PIPES buffer (pH 7.2) and fixed for 1 h. After washing with 0.1 M PIPES buffer, each sample was carefully inserted into a CPD capsule and dehydrated in a series of ethanol solutions (30, 50, 70, 80, 90, 3× 100% including a final dry ethanol change). Samples were dried in a Leica EM CPD300 Critical Point Dryer using liquid carbon dioxide as the transition fluid. The filters were carefully mounted onto SEM stubs using sticky tabs, ensuring that the inoculated surface was facing upwards. The samples were coated with platinum in an Agar high-resolution sputter-coater apparatus. SEM was carried out using a Zeiss Supra 55 VP FEG SEM, operating at 3 kV.

Fatty acid analyses. For the determination of cellular fatty acids from the strains *D. legallii* (DSM 19129) and QI0027^T, cells were grown in Postgate C media at 37°C for 18 h. Cells in the exponential phase were pelleted by centrifugation at 4000g for 10 min. Fatty acid identification was carried out by the Identification Service of the DSMZ (Braunschweig, Germany). Fatty acids were converted to FAMES by saponification, methylation and extraction according to Miller (1982) and Kuykendall *et al.* (1988). Samples were prepared according to the standard method given in MIDI Technical Note 101. Profiles of fatty acids were determined with an Agilent 6890N gas chromatograph and processed with the MIDI Inc Sherlock MIS software version 6.1.

Utilization of carbon substrates. To determine which carbon substrates could be utilized by QI0027^T and *D. legallii*, we used AN microplates from the OmniLog™ system (1007, Technopath, Ireland). We followed the manufacturer's instructions (AN Microplate™ procedure for testing carbon metabolism of anaerobic bacteria, draft of 3.11.2017 and (Takii *et al.*, 2008)). Briefly, cells of the two species were grown on Biolog Universal Anaerobe agar (Technopath) inside an anaerobic cabinet until visible growth could be observed (2–3 days). Cells were collected with cotton swabs and used for preparing a cell suspension in AN inoculating fluid (Technopath) with turbidity between 0.4 and 0.48. These suspensions were used for the inoculation of the AN microplate in triplicates. To maintain an anaerobic atmosphere during the experiment, AN microplates were placed inside plastic bags which had an atmosphere of 90% N₂, 5% CO₂ and 5% H₂. The plates were incubated at 37°C for 36 h and the

colour change caused by the response to the oxidation of the different substrates in the plate were measured with an Omnilog plate reader (Biolog). The resulting data were analysed with the *opm* package (v.1.3.63) (Vaas *et al.*, 2013) implemented in R. Reactions were defined as positive or negative using k-means partitioning as implemented in the function *do_disc*.

Determination of nitrogenase activity

To determine the nitrogenase activity of QI0027^T to fix nitrogen, we used the acetylene reduction assay (Dilworth, 1966; Schöllhorn and Burris, 1967). The cells were first grown in media free of added fixed nitrogen (ammonia, NH₃), modified from Postgate B media (Postgate, 1970), which contained per liter of distilled water: 3.5 g sodium lactate, 2 g MgSO₄·7H₂O, 0.25 g CaSO₄, 0.5 g K₂HPO₄, 0.004 g FeSO₄·7H₂O, 9 g NaCl, 4 ml resazurin (0.02% wt./vol.), 0.1 g ascorbic acid and 0.1 g thioglycolic acid. pH was adjusted to 7.5 ± 0.1 using 5 M HCl. The media was dispensed into Hungate tubes (6.2 ml gas phase) and sterilized by autoclaving. Media was supplemented with 1 ml L⁻¹ of trace element solution SL-10 (DSMZ). Growth could be observed after 5–7 days. These cells were used to inoculate 12 tubes with a 1:100 dilution in 10 ml of media. As nitrogenase activity is inhibited when fixed nitrogen is available, we included as negative control duplicate tubes supplemented with either 1 g L⁻¹ of yeast extract or 1 g L⁻¹ NH₄Cl (Hardy *et al.*, 1968; Postgate, 1970; Riederer-Henderson and Wilson, 1970). Duplicate blank tubes without inoculum were also included as negative controls. All tubes were injected 10% vol./vol. of the gas phase with acetylene and incubated at 37°C. After 7 days of incubation, ethylene formation was quantified by using a Perkin Elmer Clarus 480 gas chromatograph equipped with a HayeSep® N (80–100 MESH) column. The injector and oven temperatures were kept at 100°C, while the FID detector was set at 150°C. The carrier gas flow was set at 8–10 ml min⁻¹. The ethylene standard was prepared from ethephon decomposition based on Zhang and Wen (2010). Nitrogenase activity is reported as nmol of ethylene produced per mg protein per hour. To determine whole-cell protein concentration, 5 ml of the culture was centrifuged at 3200g for 10 min at 4°C. The pellet was resuspended in 500 µl NaOH 0.1 M and incubated for 3 h at 30°C. Cells were then sonicated using an MSE Soniprep 150 instrument for seven cycles of 15 s sonication followed by at least 30 s rest on ice. Protein content from the lysates was finally measured with the Bradford assay (BioRad). To compare our results with the ARAs previously described in human stool by Igai *et al.* (2016) and Bergersen and Hipsley (1970), we assumed the following: (i) the bacterial load is 10¹¹ cells per gram

(Sender *et al.*, 2016), (ii) bacterial cells have 60–330 fg of protein per cell (Zubkov *et al.*, 1999) and (iii) QI0027^T can be at an abundance from 0.02% to 0.06% from the total bacterial population (Supplementary Dataset 3).

Transcriptional response to the presence and absence of ammonia

To identify the genes that are expressed by QI0027^T when fixing nitrogen, we used whole transcriptome sequencing of cultures grown in modified Postgate B minimal media without any source of fixed nitrogen or with 1 g L⁻¹ NH₄CL. QI0027^T cells grown for 48 h in Postgate C media were centrifuged at 4000g for 3 min. Cells were resuspended in modified Postgate B media without NH₄CL. These cells were used to inoculate six Hungate tubes with 10 ml Postgate B media, six tubes supplemented with 1 g L⁻¹ NH₄Cl, as well as duplicate control tubes supplemented with 1 g L⁻¹ of yeast extract using a 1:100 dilution. Duplicate negative controls of Postgate B media supplemented with yeast extract without QI0027^T inoculum were also included. As expected, control tubes with yeast extract showed growth in 48 h. After 7 days, we divided the media from six tubes per condition into two to increase the input biomass for RNA extraction. Cells were centrifuged at 4000g for 10 min and immediately frozen in dry ice for 10 min. RNA was extracted using the RNeasy Plus Minikit (Qiagen, UK) according to the manufacturer's instructions with the following modifications. 350 µl RNeasy lysis buffer and 10 µl β-mercaptoethanol were added to the cell pellets. Cells were disrupted in a FastPrep (MP Bio, UK) with 3 cycles lasting 20 s at 4 m s⁻¹ in Lysing Matrix E tubes (MP Bio). Samples were centrifuged at 13 000g for 3 min, and the supernatant was used for further RNA extraction. The resulting RNA was used to synthesize cDNA using the SMARTer Universal Low Input RNA kit for sequencing (Takara, Korea), followed by library construction using the TruSeq Total RNA (NEB Microbe, New England BioLabs, USA) kit without rRNA depletion. 3 Gb per sample were generated as 100 bp paired-end reads using a NovaSeq instrument at Macrogen (South Korea).

For differential expression analyses, adapters and rRNA sequences were removed from transcriptome reads using *bbduk* (v.38.06). Clean reads were mapped to the QI0027^T reference genome with a minimum identity of 95% and all ambiguous mapping reported. The number of transcripts per gene was estimated with *featureCounts* (v.2.0) (Liao *et al.*, 2014). Differentially expressed genes were detected with *edgeR* with TMM normalization (Robinson *et al.*, 2010; Robinson and Oshlack, 2010). Pathway differentially expressed were examined with

Pathway Tools (v.23.5). Operons were predicted with Rockhopper. Secretion and conjugation systems were predicted with Maccsyfinder (v.1.0.5) using the CONJScan and TXSS models (Abby *et al.*, 2014; Abby *et al.*, 2016). Functional domains of MCP proteins were predicted with CD-Search tool (Marchler-Bauer *et al.*, 2014).

Expression of nitrogenase genes by Desulfovibrio in human stool metatranscriptomes

Metatranscriptome sequencing data were obtained from the Human Microbiome Project Data Portal (<https://portal.hmpdacc.org>, accessed in February 2021). 809 metatranscriptomes were sequenced with ~13 million paired-end reads each (Lloyd-Price *et al.*, 2019) and 26 had between ~1 and 45 million paired-end reads each (Consortium, 2019). Reads were mapped against QI0027^T, *Desulfovibrio legallii* strain KHC7 (FNBX01000016), *Desulfovibrio desulfuricans* subsp. *desulfuricans* DSM 642 (ATUZ01000015) and *Desulfovibrio desulfuricans* strain IC1 (CP036295) genomes using bmap (v.38.06) with a minimum similarity of 95%. Read counts per gene were estimated with featureCounts (v.2.0) (Liao *et al.*, 2014).

Acknowledgements

We thank Gemma Langridge, Claire Hill, Maria Diaz and Barry Bochner for technical support using the Omnilog machine. We thank the JIC Bioimaging facility for access to electron microscopes and David Baker for MinION sequencing. We thank the three anonymous reviewers whose suggestions helped to improve this manuscript.

Author Contributions

L.S. and A.N. designed the study. L.S. conducted experiments and analysed data. T.L. isolated SRB, collected stool samples and extracted nucleic acids for metagenomes. M.B.B. and L.S. did acetylene reduction assays. B.K.B.S. suggested the impact of nitrogen fixation as electron sink and critically commented on the manuscript. C.B. did electron microscopy. Q.Z. and W.C. coordinated sample collection in China. L.S. wrote the manuscript with contributions from all co-authors.

Data Availability Statement

All sequencing data were submitted to the European Nucleotide Archive (<http://www.ebi.ac.uk/ena/data/view/>). Sequencing reads and the genome assembly of QI0027^T are available with the accession number PRJEB34368. Metagenomic reads obtained from Chinese stool samples without human reads are available with the

D. diazotrophicus from the gut can fix nitrogen 3177

accession number PRJEB38832. Transcriptomic reads are available under the accession number PRJEB38833.

Ethics Statement

This study was approved by the Ethics Committee in Jiangnan University, China (SYXK 2012–0002). All the faecal samples from healthy persons were for public health purposes and these were the only human materials used in the present study. Written informed consent for the use of their faecal samples was obtained from the participants or their legal guardians. All of them conducted health questionnaires before sampling and no human experiments were involved. The collection of faecal samples had no risk of predictable harm or discomfort to the participants.

References

- Abby, S.S., Cury, J., Guglielmini, J., Néron, B., Touchon, M., and Rocha, E.P.C. (2016) Identification of protein secretion systems in bacterial genomes. *Sci Rep* **6**: 23080.
- Abby, S.S., Néron, B., Ménager, H., Touchon, M., and Rocha, E.P.C. (2014) MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS One* **9**: e110726.
- Ailloud, F., Didelot, X., Woltemate, S., Pfaffinger, G., Overmann, J., Bader, R.C., *et al.* (2019) Within-host evolution of *Helicobacter pylori* shaped by niche-specific adaptation, intragastric migrations and selective sweeps. *Nat Commun* **10**: 2273.
- Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G.W., and Nielsen, P.H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533–538.
- Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., *et al.* (2020) A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* **39**: 105.
- Alneberg, J., Bjarnason, B.S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., *et al.* (2014) Binning metagenomic contigs by coverage and composition. *Nat Methods* **11**: 1144–1146.
- Amann, R., and Fuchs, B.M. (2008) Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6**: 339–348.
- Anderson, J.K., Smith, T.G., and Hoover, T.R. (2010) Sense and sensibility: flagellum-mediated gene regulation. *Trends Microbiol* **18**: 30–37.
- Ansorge, R., Romano, S., Sayavedra, L., Porras, M.A.G., Kupczok, A., Tegetmeyer, H.E., *et al.* (2019) Functional diversity enables multiple symbiont strains to coexist in deep-sea mussels. *Nat Microbiol* **4**: 2487–2497.
- Bäckhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P., *et al.* (2015) Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**: 690–703.

- Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.
- Beier, D., and Gross, R. (2006) Regulation of bacterial virulence by two-component systems. *Curr Opin Microbiol* **9**: 143–152.
- Bergersen, F., and Hipsley, E. (1970) The presence of N₂-fixing bacteria in the intestines of man and animals. *Microbiology* **60**: 61–65.
- Bombar, D., Paerl, R.W., and Riemann, L. (2016) Marine non-cyanobacterial diazotrophs: moving beyond molecular detection. *Trends Microbiol* **24**: 916–927.
- Brettin, T., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Olsen, G.J., et al. (2015) RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* **5**: 8365.
- Bueno Batista, M., and Dixon, R. (2019) Manipulating nitrogen regulation in diazotrophic bacteria for agronomic benefit. *Biochem Soc Trans* **47**: 603.
- Carbonero, F., Benefiel, A.C., Alizadeh-Ghamsari, A.H., and Gaskins, H.R. (2012) Microbial pathways in colonic sulfur metabolism and links with health and disease. *Front Physiol* **3**: 448.
- Carbonero, F., and Gaskins, H.R. (2015) Sulfate-reducing bacteria in the human gut microbiome. In *Encyclopedia of Metagenomics: Environmental Metagenomics*, pp. 617–619. Boston, MA: Springer.
- Chen, J., Toth, J., and Kasap, M. (2001) Nitrogen-fixation genes and nitrogenase activity in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *J Ind Microbiol Biotechnol* **27**: 281–286.
- Coutinho, C., Coutinho-Silva, R., Zinkevich, V., Pearce, C. B., Ojcius, D.M., and Beech, I. (2017) Sulphate-reducing bacteria from ulcerative colitis patients induce apoptosis of gastrointestinal epithelial cells. *Microb Pathog* **112**: 126–134.
- Dai, Z.-L., Wu, G., and Zhu, W.-Y. (2011) Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. *Front Biosci* **16**: 1768–1786.
- Dilworth, M. (1966) Acetylene reduction by nitrogen-fixing preparations from *Clostridium pasteurianum*. *Biochim Biophys Acta (BBA)-Gen Subj* **127**: 285–294.
- Dixon, R., and Kahn, D. (2004) Genetic regulation of biological nitrogen fixation. *Nat Rev Microbiol* **2**: 621–631.
- Dos Santos, P.C., Fang, Z., Mason, S.W., Setubal, J.C., and Dixon, R. (2012) Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC Genomics* **13**: 162.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Fischbach, M.A., and Sonnenburg, J.L. (2011) Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **10**: 336–347.
- Fröhlich, J., Sass, H., Babenzien, H.-D., Kuhnigk, T., Varma, A., Saxena, S., et al. (1999) Isolation of *Desulfovibrio intestinalis* sp. nov. from the hindgut of the lower termite *Mastotermes darwiniensis*. *Can J Microbiol* **45**: 145–152.
- Fuller, M.F., and Reeds, P.J. (1998) Nitrogen cycling in the gut. *Annu Rev Nutr* **18**: 385–411.
- Gaby, J.C., and Buckley, D.H. (2014) A comprehensive aligned *nifH* gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database* **2014**: bau001.
- Gong, Q.H., Shi, X.R., Hong, Z.Y., Pan, L.L., Liu, X.H., and Zhu, Y.Z. (2011) A new hope for neurodegeneration: possible role of hydrogen sulfide. *J Alzheimers Dis* **24**: 173–182.
- Hardy, R.W., Holsten, R., Jackson, E., and Burns, R. (1968) The acetylene-ethylene assay for N₂ fixation: laboratory and field evaluation. *Plant Physiol* **43**: 1185–1207.
- He, Q., Gao, Y., Jie, Z., Yu, X., Laursen, J.M., Xiao, L., et al. (2017) Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* **6**: gix050.
- Huergo, L.F., and Dixon, R. (2015) The emergence of 2-oxoglutarate as a master regulator metabolite. *Microbiol Mol Biol Rev* **79**: 419–435.
- Igai, K., Itakura, M., Nishijima, S., Tsurumaru, H., Suda, W., Tsutaya, T., et al. (2016) Nitrogen fixation and *nifH* diversity in human gut microbiota. *Sci Rep* **6**: 31942.
- Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018) High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**: 5114.
- Kaláb, M., Yang, A.-F., and Chabot, D. (2008) Conventional scanning electron microscopy of bacteria. *Infocus Mag* **10**: 42–61.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.
- Kelly, D.J., and Thomas, G.H. (2001) The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol Rev* **25**: 405–424.
- Knapp, A.N. (2012) The sensitivity of marine N₂ fixation to dissolved inorganic nitrogen. *Front Microbiol* **3**: 374.
- Kuhnigk, T., Branke, J., Krekeler, D., Cypionka, H., and König, H. (1996) A feasible role of sulfate-reducing bacteria in the termite gut. *Syst Appl Microbiol* **19**: 139–149.
- Kuykendall, L., Roy, M., O'Neill, J., and Devine, T. (1988) Fatty acids, antibiotic resistance, and deoxyribonucleic acid homology groups of *Bradyrhizobium japonicum*. *Int J Syst Evol Microbiol* **38**: 358–361.
- Lesser, M.P., Falcón, L.I., Rodríguez-Román, A., Enríquez, S., Hoegh-Guldberg, O., and Iglesias-Prieto, R. (2007) Nitrogen fixation by symbiotic cyanobacteria provides a source of nitrogen for the scleractinian coral *Montastraea cavernosa*. *Mar Ecol Prog Ser* **346**: 143–152.
- Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.
- Levin, E.J., Quick, M., and Zhou, M. (2009) Crystal structure of a bacterial homologue of the kidney urea transporter. *Nature* **462**: 757–761.

- Li, J., Zhao, F., Wang, Y., Chen, J., Tao, J., Tian, G., *et al.* (2017) Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* **5**: 14.
- Liao, Y., Smyth, G.K., and Shi, W. (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.
- Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., *et al.* (2019) Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**: 655–662.
- Lopez-Cortes, A., Fardeau, M.L., Fauque, G., Joulian, C., and Ollivier, B. (2006) Reclassification of the sulfate- and nitrate-reducing bacterium *Desulfovibrio vulgaris* subsp. *oxamicus* as *Desulfovibrio oxamicus* sp. nov., comb. nov. *Int J Syst Evol Microbiol* **56**: 1495–1499.
- Loubinoux, J., Jaulhac, B., Piemont, Y., Monteil, H., and Le Faou, A.E. (2003) Isolation of sulfate-reducing bacteria from human thoracoabdominal pus. *J Clin Microbiol* **41**: 1304–1306.
- Lucker, S., Steger, D., Kjeldsen, K.U., MacGregor, B.J., Wagner, M., and Loy, A. (2007) Improved 16S rRNA-targeted probe set for analysis of sulfate-reducing bacteria by fluorescence *in situ* hybridization. *J Microbiol Methods* **69**: 523–528.
- Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., *et al.* (2014) CDD: NCBI's conserved domain database. *Nucleic Acids Res* **43**: D222–D226.
- McKinlay, J.B., and Harwood, C.S. (2010) Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proc Natl Acad Sci U S A* **107**: 11669–11675.
- Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., and Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**: 60.
- Meier-Kolthoff, J.P., Klenk, H.-P., and Göker, M. (2014) Taxonomic use of DNA G+C content and DNA–DNA hybridization in the genomic age. *Int J Syst Evol Microbiol* **64**: 352–356.
- Miller, L.T. (1982) Single derivatization method for routine analysis of bacterial whole-cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**: 584–586.
- Mohamed, N.M., Colman, A.S., Tal, Y., and Hill, R.T. (2008) Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environ Microbiol* **10**: 2910–2921.
- Nava, G.M., Carbonero, F., Croix, J.A., Greenberg, E., and Gaskins, H.R. (2012) Abundance and diversity of mucosa-associated hydrogenotrophic microbes in the healthy human colon. *ISME J* **6**: 57–70.
- Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. (2016) metaSPAdes: a new versatile de novo metagenomics assembler. *arXiv:160403071 [q-bio]*.
- Ortega, Á., Zhulin, I.B., and Krell, T. (2017) Sensory repertoire of bacterial chemoreceptors. *Microbiol Mol Biol Rev* **81**: e00033-17.
- Oxley, A.P., Lanfranchi, M.P., Würdemann, D., Ott, S., Schreiber, S., McGenity, T.J., *et al.* (2010) Halophilic archaea in the human intestinal mucosa. *Environ Microbiol* **12**: 2398–2410.
- Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2019) Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *BioRxiv*: 771964.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.
- Parrello, B., Butler, R., Chlenski, P., Olson, R., Overbeek, J., Pusch, G.D., *et al.* (2019) A machine learning-based service for estimating quality of genomes using PATRIC. *BMC Bioinformatics* **20**: 486.
- Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., *et al.* (2019) Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* **176**: 649–662.e20.
- Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D.T., *et al.* (2017) Accessible, curated metagenomic data through ExperimentHub. *Nat Methods* **14**: 1023–1024.
- Postgate, J. (1970) Nitrogen fixation by sporulating sulphate-reducing bacteria including rumen strains. *Microbiology* **63**: 137–139.
- Postgate, J.R., and Kent, H.M. (1985) Diazotrophy within *Desulfovibrio*. *Microbiology* **131**: 2119–2122.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**: 1641–1650.
- Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., *et al.* (2014) Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**: 59–64.
- Quast, C., Priesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al.* (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.
- Rascio, N., and La Rocca, N. (2013) Biological nitrogen fixation. *Nat Educ Knowl* **3**: 15.
- Reese, A.T., Pereira, F.C., Schintlmeister, A., Berry, D., Wagner, M., Hale, L.P., *et al.* (2018) Microbial nitrogen limitation in the mammalian large intestine. *Nat Microbiol* **3**: 1441–1450.
- Richter, M., and Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* **106**: 19126–19131.
- Riederer-Henderson, M.-A., and Wilson, P. (1970) Nitrogen fixation by sulphate-reducing bacteria. *Microbiology* **61**: 27–31.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Robinson, M.D., and Oshlack, A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**: R25.
- Sayavedra, L., Kleiner, M., Ponnudurai, R., Wetzels, S., Pelletier, E., Barbe, V., *et al.* (2015) Abundant toxin-related genes in the genomes of beneficial symbionts from deep-sea hydrothermal vent mussels. *eLife* **4**: e07966.

- Schöllhorn, R., and Burris, R. (1967) Acetylene as a competitive inhibitor of N₂ fixation. *Proc Natl Acad Sci U S A* **58**: 213–216.
- Schureck, M.A., Maehigashi, T., Miles, S.J., Marquez, J., Cho, S.E., Erdman, R., and Dunham, C.M. (2014) Structure of the *Proteus vulgaris* HigB-(HigA)2-HigB toxin-antitoxin complex. *J Biol Chem* **289**: 1060–1070.
- Scott, K.P., Gratz, S.W., Sheridan, P.O., Flint, H.J., and Duncan, S.H. (2013) The influence of diet on the gut microbiota. *Pharmacol Res* **69**: 52–60.
- Seefeldt, L.C., Hoffman, B.M., and Dean, D.R. (2009) Mechanism of Mo-dependent nitrogenase. *Annu Rev Biochem* **78**: 701–722.
- Sender, R., Fuchs, S., and Milo, R. (2016) Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol* **14**: e1002533.
- Shiozaki, T., Fujiwara, A., Ijichi, M., Harada, N., Nishino, S., Nishi, S., et al. (2018) Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnol Oceanogr* **63**: 2191–2205.
- Shivani, Y., Subhash, Y., Sasikala, C., and Ramana, C.V. (2017) *Halodesulfobacterium spirochaetisodalis* gen. nov. sp. nov. and reclassification of four *Desulfobacterium* spp. *Int J Syst Evol Microbiol* **67**: 87–93.
- Soutourina, O.A., and Bertin, P.N. (2003) Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiol Rev* **27**: 505–523.
- Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* **27**: 1009–1010.
- Takii, S., Hanada, S., Hase, Y., Tamaki, H., Uyeno, Y., Sekiguchi, Y., and Matsuura, K. (2008) *Desulfobacterium marinisediminis* sp. nov., a novel sulfate-reducing bacterium isolated from coastal marine sediment via enrichment with Casamino acids. *Int J Syst Evol Microbiol* **58**: 2433–2438.
- The Integrative HMP (iHMP) Research Network Consortium. (2019) The integrative human microbiome project. *Nature* **569**: 641–648.
- Tichi, M.A., and Tabita, F.R. (2000) Maintenance and control of redox poise in *Rhodobacter capsulatus* strains deficient in the Calvin-Benson-Bassham pathway. *Arch Microbiol* **174**: 322–333.
- Vaas, L.A.I., Sikorski, J., Hofner, B., Fiebig, A., Buddrhu, N., Klenk, H.P., and Goker, M. (2013) opm: an R package for analysing OmniLog(R) phenotype microarray data. *Bioinformatics* **29**: 1823–1824.
- Vainshtein, M., Hippe, H., and Kroppenstedt, R.M. (1992) Cellular fatty acid composition of *Desulfobacterium* species and its use in classification of sulfate-reducing bacteria. *Syst Appl Microbiol* **15**: 554–566.
- Vasoo, S., Mason, E.L., Gustafson, D.R., Cunningham, S.A., Cole, N.C., Vetter, E.A., et al. (2014) *Desulfobacterium legallii* prosthetic shoulder joint infection and review of antimicrobial susceptibility and clinical characteristics of *Desulfobacterium* infections. *J Clin Microbiol* **52**: 3105–3110.
- Waite, D.W., Chuvochina, M., Pelikan, C., Parks, D.H., Yilmaz, P., Wagner, M., et al. (2020) Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int J Syst Evol Microbiol* **70**: 5972–6016.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.
- Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., et al. (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* **42**: D581–D591.
- Wayne, P. (1999) NCCLS: National Committee for Clinical Laboratory Standards. *Performance Standards for Antimicrobial Susceptibility Testing*.
- Wen, C., Zheng, Z., Shao, T., Liu, L., Xie, Z., Le Chatelier, E., et al. (2017) Quantitative metagenomics reveals unique gut microbiome biomarkers in ankylosing spondylitis. *Genome Biol* **18**: 142.
- Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E. (2017) Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**: e1005595.
- Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015) Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* **31**: 3350–3352.
- Wu, M., and Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* **28**: 1033–1034.
- Xie, H., Guo, R., Zhong, H., Feng, Q., Lan, Z., Qin, B., et al. (2016) Shotgun metagenomics of 250 adult twins reveals genetic and environmental impacts on the gut microbiome. *Cell Syst* **3**: 572–584.e573.
- Zhang, W., and Wen, C.-K. (2010) Preparation of ethylene gas and comparison of ethylene responses induced by ethylene, ACC, and ethephon. *Plant Physiol Biochem* **48**: 45–53.
- Zubkov, M.V., Fuchs, B.M., Eilers, H., Burkill, P.H., and Amann, R. (1999) Determination of total protein content of bacterial cells by SYPRO staining and flow cytometry. *Appl Environ Microbiol* **65**: 3251–3257.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supporting Information.

Supplementary Dataset 1. Cellular fatty acid composition (%) of QI0027^T and its closest relative *D. legallii* (DSM 19129) grown in Postgate C media. FAME, fatty acid methyl ester; ND, not detected

Supplementary Dataset 2. Differential gene expression analysis of QI0027^T cultures with limiting or excess nitrogen. Differentially expressed (DE) genes were identified using edgeR. Fold change expression values were estimated using (log₂(expression with nitrogen excess)-log₂(expression with nitrogen limitation)). Thus, negative log₂ fold change values indicate genes overexpressed under nitrogen limiting conditions, while positive values indicate genes overexpressed under excess conditions. Read counts were normalized with trimmed mean of M values (TMM). For clustering analysis (Fig. 3B), log₂ centred values were

estimated from the TMM normalized values. Genes overexpressed with nitrogen deplete conditions are shown in blue. Genes overexpressed with nitrogen replete conditions are shown in orange. **p-value significant corrected using false discovery rate (FDR) correction. *p-value <0.05 and FDR <0.1. DE = Differentially expressed. DE = Differentially expressed.

Supplementary Dataset 3. Relative abundance of QI0027^T-related species in 45 Chinese stool metagenomes. The abundance was estimated based on read mapping against QI0027^T scaffolds ≥ 1000 bp. The relative abundance of the two *Desulfovibrionaceae* species predicted by MetaPhlan2 is also shown.

Supplementary Dataset 4. From 835 stool metatranscriptomes, six had transcript reads matching a gene from the nitrogenase gene cluster with a similarity >95%.

Supplementary Dataset 5. Carbon compounds used by *D. legallii* and QI0027^T as determined with the AN microplate system (Biolog). The use of each carbon substrate was assigned as being used or not based on the function do_disc from the opm package, using k-means, with no weak reactions for each of three replicates.

Supplementary Fig. 1. Colony morphology of QI0027^T (a) and motility assay (b). For the motility assay, soft agar was stabbed with an inoculum of QI0027 or *D. legallii* anaerobically. Cells were grown in Postgate C media with 5 g L⁻¹ agarose for 72 h. When bacteria are non-motile, they will show a sharp edge where the inoculation stab is made. In contrast, if they are motile, they will show a cloudy growth around the stabbing site. *D. legallii* showed a diffuse area of growth, in agreement with the previously reported presence of a flagellum (Vasoo *et al.*, 2014). In contrast, QI0027^T had a sharp edge area of growth. However, there were slight signs of iron sulphide outside the area of the stabbing site, which could be due to some cells being motile or the diffusion of the hydrogen sulphide gas produced by QI0027^T.

Supplementary Fig. 2. The *nif* gene cluster is highly conserved among QI0027^T, *D. legallii* and *D. desulfuricans*

subsp. *desulfuricans* DSM 642. Similarity between genomic regions was estimated with blastn incorporated into Easyfig (Sullivan *et al.*, 2011).

Supplementary Fig. 3. Metabolic reconstruction of QI0027^T based on genome information. Based on our genome reconstruction, QI0027^T can reduce sulfate to hydrogen sulphide, using either hydrogen, carbon monoxide or alcohols as electron donors. For the oxidation of alcohols, an alcohol dehydrogenase could be used for the conversion of ethanol to acetate, similar to *Megalodesulfovibrio gigas* (Kremer *et al.*, 1988). Lactate could be oxidized to pyruvate by a membrane-bound lactate dehydrogenase. Ammonia could be further used for the biosynthesis of glutamine, glutamate, asparagine, NAD and deazapurine. QI0027^T encoded the genes for the amino acid biosynthesis of methionine, asparagine, alanine, arginine, threonine, isoleucine, aspartate, serine, glycine, leucine, and valine. Additionally, QI0027^T could synthesize the following vitamins and cofactors: biotin (B7), folate (B9), phosphopantothenate (B5), thiamine (B1), pyridoxal (B6), adenosylcobalamin (B12) and flavin. H₂ase = hydrogenase; ADH = alcohol dehydrogenase; CODH = carbon monoxide dehydrogenase.

Supplementary Fig. 4. Maximum likelihood of *nifH* gene product tree reconstruction. 581 amino acid sequences spanning 300 amino acid positions were used to reconstruct the phylogeny. All non-collapsed *nifH* sequences belong to Cluster 3 *nifH*, including the *nifH* encoded by *Desulfovibrionaceae* (D1, D2, D3) and the *Clostridia*. Cluster 4 contains paralogous genes that are not involved in nitrogen fixation. Solid circles in phylogeny represent SH-like support between 0.8–1.0, with the size proportional to the support value. Bootstrap support (a) or SH support (b) between 80%–100% is represented as filled circles scaled to size. *nifH* phylogeny. Some *Desulfovibrionaceae* encoded a paralog nitrogenase *anfH* that grouped within cluster 2. Cluster 2 groups alternative nitrogenases that use an Fe-Fe cofactor instead of an Fe-Mo cofactor (Gaby and Buckley, 2014).