Title

The exceptional form and function of the giant bacterium Ca. Epulopiscium viviparus revolves around its sodium motive force.

Permalink

Journal

Proceedings of the National Academy of Sciences, 120(52)

Authors

Sannino, David
Arroyo, Francine
Pepe-Ranney, Charles
et al.

Publication Date

2023-12-26

DOI

10.1073/pnas.2306160120

Peer reviewed

The exceptional form and function of the giant bacterium *Ca.* Epulopiscium viviparus revolves around its sodium motive force

David R. Sannino[a,1] [ID], Francine A. Arroyo[a,2], Charles Pepe-Ranney[b,3], Wenbo Chen[a,4], Jean-Marie Volland[c,d] [ID], Nathalie H. Elisabeth[e], and Esther R. Angert[a,5] [ID]

Epulopiscium spp. are the largest known heterotrophic bacteria; a large cigar-shaped individual is a million times the volume of *Escherichia coli*. To better understand the metabolic potential and relationship of *Epulopiscium* sp. type B with its host *Naso tonganus*, we generated a high-quality draft genome from a population of cells taken from a single fish. We propose the name *Candidatus* Epulopiscium viviparus to describe populations of this best-characterized *Epulopiscium* species. Metabolic reconstruction reveals more than 5% of the genome codes for carbohydrate active enzymes, which likely degrade recalcitrant host-diet algal polysaccharides into substrates that may be fermented to acetate, the most abundant short-chain fatty acid in the intestinal tract. Moreover, transcriptome analyses and the concentration of sodium ions in the host intestinal tract suggest that the use of a sodium motive force (SMF) to drive ATP synthesis and flagellar rotation is integral to symbiont metabolism and cellular biology. In natural populations, genes encoding both F-type and V-type ATPases and SMF generation via oxaloacetate decarboxylation are among the most highly expressed, suggesting that ATPases synthesize ATP and balance ion concentrations across the cell membrane. High expression of these and other integral membrane proteins may allow for the growth of its extensive intracellular membrane system. Further, complementary metabolism between microbe and host is implied with the potential provision of nitrogen and B vitamins to reinforce this nutritional symbiosis. The few features shared by all bacterial behemoths include extreme polyploidy, polyphosphate synthesis, and thus far, they have all resisted cultivation in the lab.

giant bacteria | polyploid | cellular bioenergetics | gut microbiota | ATPase

Bacteria are typically restricted to small sizes, as they rely on diffusion to encounter nutrients in the environment and to disperse intracellular biomolecules (1, 2). Further, it has been hypothesized that bacteria lack complex traits due to energetic constraints (3–5). Eukaryotes harbor scalable bioenergetic organelles (i.e., mitochondria and chloroplasts), which support increased genomic (nuclear) complexity. This in turn has allowed eukaryotes to expand diverse gene families, to explore genetic novelty, and hence to develop more intricate structures and traits. Despite having small genomes relative to eukaryotes, genomic innovation is central to bacterial diversity and evolution (6). Some bacteria have stretched their energetic and genomic resources to support large cell size and more complex lifestyles and morphologies. Mixotrophic or autotrophic giant bacteria, such as *Thiomargarita magnifica*, *T. namibiensis* and many cyanobacteria, break size constraints by using abundant and diffusible small molecules and/or light to meet their energy, carbon, and nitrogen needs (1, 7). Another important strategy of sulfur-oxidizing bacteria is discrete subcellular organization. Many giant sulfur bacteria, for instance, contain a large intracellular vacuole that takes up ~98% of the cell volume and press the active cytoplasm into a thin layer just beneath the cytoplasmic membrane (1). With this simple change, no part of the cytoplasm is more than a few microns away from environmental signals or nutrients. The membrane-bound vacuole also allows the cell to carry a ready supply of the electron acceptor nitrate to fuel its metabolism. Strategies for gigantism in heterotrophic bacteria are less known.

Epulopiscium spp. (a.k.a. "epulos") are intestinal symbionts of tropical marine surgeonfish (family Acanthuridae) (8–10) that are exceptional and unique in the bacterial world; they are the largest known heterotrophic bacteria. Individual cigar-shaped cells may reach up to 600 μm in length and 80 μm in width (11). The unicornfish *Naso tonganus* hosts *Epulopiscium* sp. morphotype B, the most extensively studied epulo (12). These cells are ~100 to 300 μm in length. They exhibit a number of complex traits including extreme polyploidy (12), with tens of thousands of copies of their chromosome located in a network

Significance

Giant bacteria are intriguing, underexplored biological enigmas. Many well-studied giants use abundant internal small-molecule stores and/or light to satisfy their oversized energy demands. However, the ways giant heterotrophs fulfill their expanded needs remain elusive. Using a comprehensive omics-guided approach, we reconstructed the metabolic potential of the giant heterotroph *Ca.* Epulopiscium viviparus, a marine surgeonfish intestinal symbiont, to determine how it fuels its robust metabolism. Through fermentation and coopting its sodium ion-rich environment, cells generate a sodium-motive force that drives ATP synthesis and flagellar motility. With high expression of ATPases and other membrane proteins, *Ca.* E. viviparus appears to have converged on a structure–function relationship previously described for the mitochondrion inner membrane, revealing another wrinkle in its unique biology.

[1]Present address: School of Molecular Biosciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom.

[2]Present address: Department of Biology, California State University–Fresno, Fresno, CA 93740.

[3]Present address: Gillings School of Global Public Health, University of North Carolina, Durham, NC 27599.

[4]Present address: Zhejiang Provincial Key Laboratory of Horticultural Plant Integrative Biology, Zhejiang University, Hangzhou, PR China 310058.

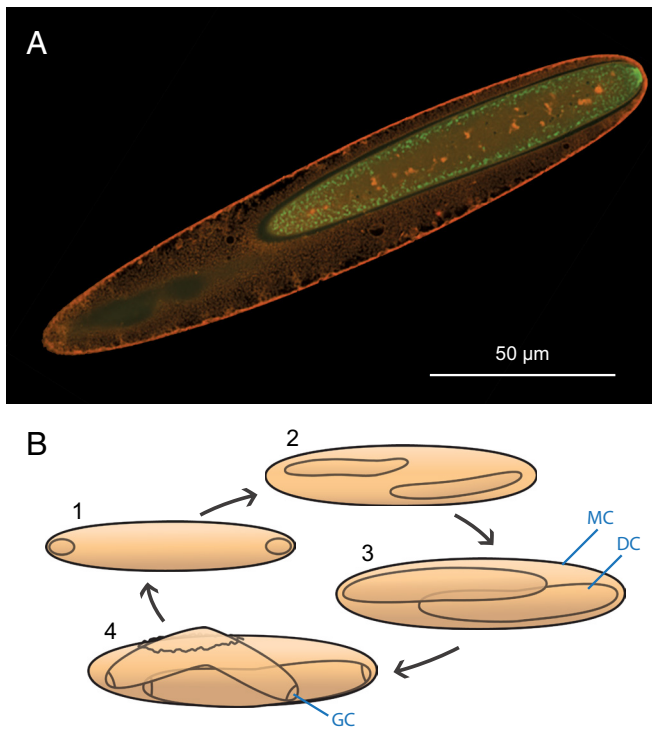[5]To whom correspondence may be addressed. Email: era23@cornell.edu.

Fig. 1. (*A*) Confocal micrograph of *Ca.* E. viviparus. This oblique optical section highlights the extensive intracellular membrane system (orange) and peripheral location of DNA (green) in offspring and mother cell. Note that only a portion of one developing offspring is visible in the plane of the section shown. Scale bar represents 50 μm. (*B*) The *Ca.* E. viviparus lifecycle. For simplicity, only cell outlines are illustrated. A cell begins its reproductive cycle with the extreme asymmetric division of a daughter cell (still contained in its mother cell, MC). The daughter cell (DC) membrane travels around the polar cell, fully engulfing this granddaughter cell (GC). Around this time, the daughter cell is released from the mother cell. The engulfed cell develops within a membrane-bound compartment, and the growth cycle continues. For RNAseq studies described here, cells at stage 1 predominate the early-morning (EM) samples, cells at stage 2 are present in the midday samples (MD) and cells at stage 3 are most abundant in the late afternoon (LA) samples. Evidence of binary fission has never been observed in *Ca.* E. viviparus.

of nucleoids at the periphery of the cytoplasm (13, 14). The nucleoids are adjacent to the heavily invaginated cell membrane (Fig. 1*A*) and chromosome number scales with cell volume (12, 14). Type B cells reproduce once every 24 h, and growth occurs primarily during daylight hours, when the host feeds actively. Adult *N. tonganus* often harbor large populations of type B epulos, which are more morphologically and genetically homogenous than other characterized epulo communities (15). Type B populations exhibit a synchronized, diurnal cell cycle (16) in which each mother cell produces 2 to 12 non-dormant, intracellular offspring in an extraordinary form of bacterial viviparity (Fig. 1*B*). The internal offspring replicate their genomes and grow inside the mother cell, until the offspring nearly fill the cytoplasm (16). Mother-cell lysis releases the offspring. This mode of propagation is derived from the highly conserved endosporulation program seen in diverse Bacillota/Firmicutes (15). Epulos are members of the Lachnospiraceae, which includes important gut symbionts of animals (17). Despite numerous studies on the cell biology of type B, a comprehensive understanding of the physiology of this uncultivated bacterial symbiont is still lacking. Microbial community metagenomic studies support the hypothesis that epulos help digest and ferment the carbohydrate-rich diet consumed by their host (18). However, these community-level metabolic predictions lack a clear picture of the physiological potential of individual populations and moreover have insufficient resolution to allow a refined analysis of the relationship of these symbionts with their host. Here, we provide insight into the metabolic and biosynthetic capabilities of *Epulopiscium* sp. type B, through the generation and analysis of a high-quality draft population assembled genome (PAG) from cells taken from a single host fish. Gene expression patterns in populations of type B cells, at three stages of their daily developmental cycle, offer remarkable structure–function insight into the unique biology of these bacteria. In our analysis of the RNAseq data, we focused on mechanisms of energy conservation, motility, and expression of potentially energy-taxing structures that might be produced by these giant heterotrophs.

Results and Discussion

Draft Genome Status, Nomenclature, and Analysis of Gene Transcription. We used genomics and transcriptomics to interrogate the metabolic and biosynthetic potential of *Epulopiscium* sp. type B. Through an iterative hybrid assembly approach and the use of a low-diversity population (19), we were able to generate a high-quality 3.28-Mb draft PAG, comprised of seven contigs, and conservatively estimated to be 92% complete (Table 1 and *SI Appendix*, Fig. S1). See *SI Appendix* for a description of how this low-diversity population was identified. The genome encodes at least one tRNA for each of the 20 amino acids (*SI Appendix*, Table S1). Consistent with previous studies (20), the type B genome appears to be a single, circular chromosome with no detectable plasmids. Further, to bolster our understanding of genome function, we conducted RNAseq analyses of type B populations from seven fish collected at three times of day (*SI Appendix*, Table S2). See *SI Appendix* for a description of the RNA extraction and transcript analysis. These collection time intervals correspond with different stages in the type B developmental cycle (Fig. 1*B*, stages 1 to 3) and correlate with changes in host feeding. *Naso tonganus* feeds during the day and does not eat at night. Fish begin to eat in the morning and continue to browse on algae throughout the daylight hours. Feeding slows in the late afternoon and fish cease feeding shortly before sunset.

Table 1. Statistics of the *Ca.* Epulopiscium viviparus genome

Genome feature	Value
Genome size (bp)	3,282,201
Contigs	7
Max contig length (bp)	905,387
N50 (bp)	602,975
G + C content	38.08%
Coding density	91.96%
ORFs	2,714
Protein coding genes	2,635
tRNAs	54
16S rRNA genes	6
23S rRNA genes	6
5S rRNA genes	6
Genome completeness based on CheckM (Bacillota)	97%
Genome completeness based on CheckM (Clostridia)	99%
Genome completeness based on conserved gene list	92%

Percent completeness based on the conserved gene list was determined using essential single-copy genes found in *Cellulosilyticum lentocellum* DSM 5427, the closest cultured, sequenced relative of *Ca.* Epulopiscium viviparus.

A recent study described a single-cell assembled genome (SCG-B10WGA) and metagenome bin (PG-AS2M_MBin01) as *Epulopiscium* sp. type B (18). These data were generated from *Acanthurus* species collected in the Red Sea. The single-cell genome yielded two bins, said to represent both a type B cell and a morphotype type A cell. However, one of the bins lacked 16S rRNA gene sequences and no morphological description of the cell used to generate the datasets was provided. Based on current criteria (7), these are medium-quality and low-quality drafts (*SI Appendix,* Fig. S2). The average percent nucleotide identity between these assemblies and our type B PAG draft is 77%, and likewise, BLAST analyses revealed low similarity between coding sequences. We therefore reject the nomenclature reported by Ngugi and colleagues (18) and propose the name "*Candidatus* Epulopiscium viviparus" (herein also referred to as type B) to represent *Epulopiscium* sp. type B symbionts of *N. tonganus*. The name reflects its unique intracellular reproductive strategy [(15) and described below], and ability to produce a greater number of non-dormant internal offspring than other known epulos (8, *SI Appendix,* Fig. S3).

Common Features of Bacterial Behemoths. Giant bacteria (*SI Appendix,* Fig. S4), including members of the candidate genera "Thiomargarita" and "Epulopiscium" share characteristics related to DNA management including extreme polyploidy and peripheral DNA localization patterns (21). With recently acquired genome data, we sought to uncover other overarching characteristics of large bacteria. Two draft genomes for *Ca.* Thiomargarita nelsonii strains (22, 23) and the draft genome of *Ca.* T. magnifica (24) served as the basis for comparison (Table 2).

Cell reproduction mechanisms for giant bacteria are diverse. Compared to other Pseudomonadota/Proteobacteria, *Ca.* T. magnifica encodes an unusual complement of cell division and peptidoglycan synthesis genes, which may contribute to cell elongation and terminal bud formation used for dispersal. Our analysis of the current draft genome of type B confirms and extends previous work that describes the conservation of sporulation genes in type B (*SI Appendix, Supplemental Text*), used to produce intracellular offspring (15). The apparent loss of genes for dipicolinic acid production and transport and spore photoproduct lyase suggests that type B is unable to produce endospores. Binary fission is rarely observed in *Thiomargarita* spp. and has never been observed in type B. This suggests that assembly of a functional divisome at the midcell may be as challenging in giant bacteria as it is for large *Escherichia coli* cells (25).

Likewise, mechanisms of energy conservation are distinct; there is not one metabolic strategy required to achieve these large cell

Table 2. Comparison of genomes of giant bacteria

Property	*Ca.* Epulopiscium viviparus	*Ca.* Thiomargarita nelsonii Thio36	*Ca.* Thiomargarita nelsonii Bud S10	*Ca. Thiomargarita magnifica* 75
Size	3.28 Mbp	5.3 Mbp	6.2 Mbp	12.2 Mbp
ORFs	2,714	7,596	7,525	11,788
Coding density (%)	92	72	82	77.6
tRNAs	54	23	46	72
GC%	38.08	42	41.3	42.21
Completeness (%) (CheckM)	97	70	89.8	92.86
Morphology: typical cell size	Cigar shape: 100 to 300 μm × 50 to 60 μm	Barrel shape, chains: 150 μm^2	Spherical: 150 μm in diameter	Sessile filaments up to 2 cm in length
Habitat	*Naso tonganus* intestinal symbiont	Sulfidic sediment surface water	Sulfidic sediment surface water	Sulfidic sediments of mangroves
Energy conservation	Fermentation	Lithotrophy	Lithotrophy	Lithotrophy
SMF generation	Oxaloacetate decarboxylase	Rnf complex	Rnf complex	Rnf complex
Nitrate reduction	Dissimilatory–non-respiratory	Dissimilatory (denitrification) and assimilatory	Dissimilatory (denitrification) and assimilatory	Incomplete pathways for dissimilatory and assimilatory
Carbon fixation	–	CBB cycle, reverse TCA	CBB cycle	CBB cycle
Degree of polyploidy	10,000s of chromosome copies	n.d.	n.d.	100,000s of chromosome copies
Polyphosphate accumulation	+	?	+	+
Reverse transcriptases	1	?	62	6
Presence of group I introns	–	+	+	+
Presence of group II introns	+ (2)	?	+ (57)	+ (3)
Presence of transposons/IS elements	+ (2)	?	+ (90)	+ (125)
Presence of MITEs	n.d.	?	+	+ (20)
Presence of inteins	n.d.	?	+ (4)	+ (4)
fdxN excision elements	n.d.	?	+ (4 *xisH*, 7 *xisI*, 8 serine type ssr)	+ (22 *xisH*, 32 *xisI*, 22 serine type ssr)

+ present; − absent; ? unknown; n.d. not determined; ssr is site-specific recombinase.

sizes. Both *Ca*. Thiomargarita nelsonii strains and *Ca*. T. magnifica are putative mixotrophs that can grow lithotrophically on reduced sulfur compounds, but have the genomic potential for H_2 oxidation (22, 23), with Bud S10 having the potential for arsenite oxidation (23). The *Ca*. Thiomargarita spp. respire and fix carbon (22, 23). In contrast, *Ca*. E. viviparus uses organic carbon for building biomass and lacks respiration. Our analysis of the genome of *Ca*. E. viviparus did not identify genes to produce cytochromes, or other components of an electron transport chain for aerobic or anaerobic respiration. The metabolic reconstruction based on the genome (described in more detail below) suggests that organic building blocks are imported directly or synthesized by transformation of imported precursors or products of central metabolism. No gene homologs coding for Rubisco or other proteins involved in carbon fixation were found.

A major difference between the genomes of these two groups of giant bacteria is the incidence of mobile genetic elements (Table 2). It has been hypothesized that both the *Ca*. Thiomargarita nelsonii Bud S10 and *Ca*. Thiomargarita magnifica genomes are malleable as they are littered with mobile genetic elements such as group I introns and inteins, some of which disrupt genes for polymerases, transcriptional regulators, and transporters (23, 24). Mobile elements are almost non-existent in the *Ca*. E. viviparus genome, which has a 92% coding density and only six pseudogenes (Table 2).

All described genomes of giant bacteria have repetitive elements. Like other Beggiatoaceae, the genomes of *Ca*. Thiomargarita spp. contain copies of the heptamer TAACTGA. Type B has large blocks of imperfect direct repeats (each unit typically 18 or 21 bp long, see for example *SI Appendix*, Fig. S5) scattered through the genome (*SI Appendix*, Fig. S1). In the type B genome, different blocks typically do not share the same sequence units. Repeats are found in intergenic regions, within genes coding for giant proteins, or the islands of repeats contain ORFs for small proteins of unknown function. Although the repeated elements are structurally distinct, they may be necessary for managing genetic resources in large polyploid bacteria. It is plausible that these repeats are recognized and bound by proteins that help maintain the DNA at the periphery of the cytoplasm. Repetitive sequences might facilitate recombination between chromosomes or during horizontal gene transfer. Using multilocus sequence analyses of type B cells, we found that recombination is important for maintaining diversity in type B populations (19).

Another shared characteristic is that all large bacteria have genes to accumulate polyphosphate. Although the exact function is unknown in these cells, polyphosphate can be used for energy and phosphate storage or a number of other metabolic and regulatory functions (26).

Together these observations suggest that polyploidy is vital for gigantism in bacteria and the ability to accumulate polyphosphate may be related to DNA metabolism (27) or energy storage. While energy conservation is important, no one means is essential to support a large bacterium. Cell morphology, modes of division, and other aspects of reproduction also vary.

Central Metabolism and Energy Conservation of *Ca*. Epulopiscium viviparus. To assess whether *Ca*. E. viviparus has unique energy conservation strategies to attain its giant size, we determined the key features of its central metabolism. Surprisingly, *Ca*. E. viviparus behaves like other Clostridia, relying on fermentation for energy conservation (28), as it lacks the genomic components for respiratory complexes. The central metabolism of *Ca*. E. viviparus, based on the type B genome, is summarized in Fig. 2. Pathways predicted to generate ATP via substrate-level phosphorylation include fermentation of pyruvate to acetate and fermentation of biosynthetic intermediates to propionate. In addition, ATP can be generated by the conversion of phosphoenol pyruvate (PEP) to oxaloacetate (OA) by PEP carboxykinase (29). Alternatively, pyruvate may be fermented to acetaldehyde and then ethanol, regenerating NAD^+. The genome encodes a complete set of genes for the non-oxidative branch of the pentose phosphate pathway (PPP). Like *Cellulosilyticum lentocellum* DSM 5427 and other Clostridia (30), type B codes for an incomplete TCA cycle, but has retained the ability to generate the key intermediates α-ketoglutarate (α-KG) and OA (Fig. 2). OA can enter the TCA cycle from Embden-Meyerhof-Parnas (EMP) to support biosynthesis, or it may be used to generate a sodium-motive-force (SMF) for energy conservation. A membrane-bound OA decarboxylase (OAD) complex couples decarboxylation with the translocation of Na^+ out of the cell (31). Further, the genome codes for the citrate lyase complex, which cleaves citrate to OA and acetate (32, 33). Citrate is taken up by the cell using CitM, a Mg^{2+}-citrate transporter, and likely by a putative tricarboxylate transporter, which is coded for in an operon with subunits for the citrate lyase complex. This combination of citrate transporters suggests extracellular citrate is used to generate OA which contributes to the SMF (34). Genes involved in citrate and OA metabolism are more highly expressed in populations collected in the early morning (EM) and midday (MD), during times of rapid offspring growth and development (Fig. 1*B* stages 1 and 2), than in late afternoon samples (Fig. 3).

The type B genome encodes both F- and V-type ATPases. For insight into their potential function, we looked at expression patterns (Fig. 3). In EM and MD samples, genes coding for the subunits of F-type ATPases are highly expressed, and the c subunit is among the 30 most abundant type B transcripts detected. EM and MD samples have at least 12- and 8-times higher ATPase transcript numbers, respectively, compared to those same genes from samples collected in the LA. As with citrate and OA metabolism, these results suggest that ATP synthesis is in demand during periods of offspring growth. When contrasted to other OA-fermenting Clostridia, type B cells collected in the EM have F-type ATPase subunit expression levels 12- to 92-times higher than *Clostridium beijerinckii* DSM 6423 (35) and Hungateiclostridium thermocellum ATCC 27405 (36), respectively (Fig. 3).

Based on the sequence of the c subunit in type B (*SI Appendix*, Fig. S6), its F_1F_0 ATP synthase is driven by the SMF (37, 38). Considering the metabolism of type B, specifically its production of acetate, and the concentration of sodium ions in the gut, we suspect that the V-type ATPase may serve as an ion pump that helps modulate intracellular ion concentrations. By using Na^+ ions to power its ATP synthases and flagella, type B may be taking advantage of the Na^+ levels within the *N. tonganus* intestine. Sodium ion concentrations in the region of the gut where *Ca*. E. viviparus resides are 0.18 to 0.25%. This is near the optimal concentration (0.25%) of extracellular sodium reported for motility of *Vibrio* spp. with Na^+-dependent flagella (39). We also uncovered an NADH-dependent nitrate reductase within an operon homologous to the *nar* operon of *Clostridium perfringens* (40). A potential ABC nitrate importer (41) was also identified. In *C. perfringens* and other Clostridia, NO_3^- serves as an electron sink (42), which drives forward substrate-level phosphorylation (43) and improves redox balancing (44). Due to the lack of a clear assimilatory nitrite reductase gene in the genome, we suspect the nitrate reductase serves energy conservation. Furthermore, we identified a putative *nirC*; in *E. coli*, NirC functions as a proton/NO_2^- antiporter for NO_2^- efflux (45). For *Ca*. E. viviparus, the antiporter may provide a means to remove this toxic byproduct. Based on the metabolic reconstruction, transcription data, and the high
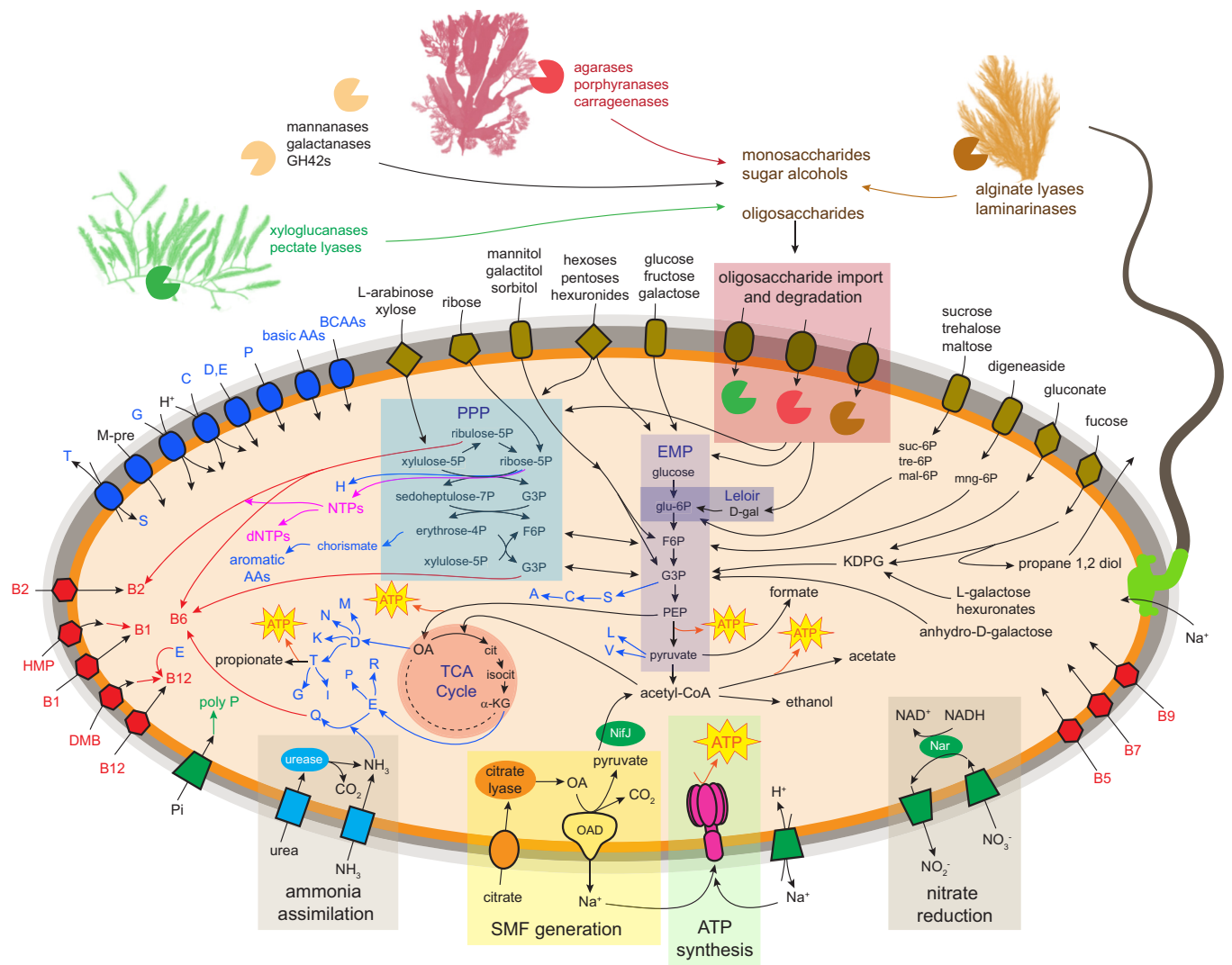
Fig. 2. The metabolic and biosynthetic landscape of *Ca*. E. viviparus. The gray and orange outlines represent the cell envelope. Specific metabolic pathways are highlighted in colored boxes with key intermediates shown. Catabolic steps or transport pathways are depicted by black lines with arrowheads, and dotted lines stand for missing steps of pathways. Carbohydrates grouped together produce the same intermediate, but not necessarily through the same pathways. The host algal diet consists primarily of red and green algae, although the fish may consume brown algae and some detritus. Exported or secreted polysaccharide degrading enzymes are depicted as Pacman symbols and classes of extracellular enzymes are listed. Color coding is used to indicate the major algal targets of these enzymes. The beige Pacman depicts the β-mannanases, β-galactanases, and GH42s active on both red and green algal polysaccharides. The oligosaccharides released are imported and degraded by intracellular enzymes. Carbohydrate transporters are colored brown. PTS are depicted as rectangles with rounded corners, diamonds depict multiple sugar transporters, and the hexagons and pentagons depict transporters for specific hexoses and pentoses, respectively. Blue ovals depict amino acid importers and exchangers, with amino acids coded by their single-letter abbreviations in blue text. Blue arrows depict amino acid biosynthesis pathways, and light blue rectangles represent transporters for urea and NH₃. Red hexagons are vitamin transporters, with the specific B vitamins labeled in red text, and red arrows depict their biosynthesis. Pink arrows illustrate biosynthesis of NTPs and dNTPs. Additional inorganic transport systems are in green. All carbohydrates are D form unless otherwise indicated. Abbreviations indicate PPP, non-oxidative pentose phosphate pathway; EMP, Embden–Meyerhof–Parnas pathway (glycolysis); TCA cycle, tricarboxylic acid cycle; OA, oxaloacetate; OAD oxaloacetate decarboxylate complex; α-KG, alpha-ketoglutarate; cit, citrate; isocit, isocitrate; BCAAs, branched-chain amino acids; M-pre, methionine precursors; poly P, polyphosphate; and KDPG, 2-keto-3-deoxy-6-phosphogluconate.

concentration of sodium ions present in the intestinal environment, we hypothesize that type B conserves energy through SMF generation which is dependent upon OA decarboxylation. It likely uses a distinct F_1F_0 ATPase to generate ATP and drives ATP generation via substrate-level phosphorylation forward via NO_3^- reduction.

Carbohydrate Metabolism in *Ca*. Epulopiscium viviparus. As fermentation is central to SMF generation and ATP production, we sought to uncover the carbon substrates that *Ca*. E. viviparus may use to satiate its large energy demands. A cluster of orthologous genes (COG) analysis was performed on the type B draft assembly (*SI Appendix,* Table S3) as well as on the genomes of other Lachnospiraceae, which included free-living, human-gut, and rumen-associated members of the Lachnospiraceae clusters XIVa and XIVb. *Ca*. E. viviparus devotes

15.43% of all proteins with COGs to carbohydrate transport and metabolism (Fig. 4 and *SI Appendix,* Fig. S7). This proportion was higher than any of the other genomes examined, even from other gut-associated species. It was on par with what was reported for community metagenomes from *Acanthurus* hosts (18). This suggests that enrichment for carbohydrate metabolism is not just a general product of being host associated, but rather, it has been selected for by the type B niche as a predominant gut symbiont of *N. tonganus*, with its diverse carbohydrate-rich diet.

To determine the variety of enzymes type B produces to breakdown and use the carbon sources available in the host diet, we conducted a Carbohydrate Active enZymes (CAZy) analysis. The CAZy analysis determined that the type B genome encodes 131 glycoside hydrolase (GH) CAZymes from 38 different families,
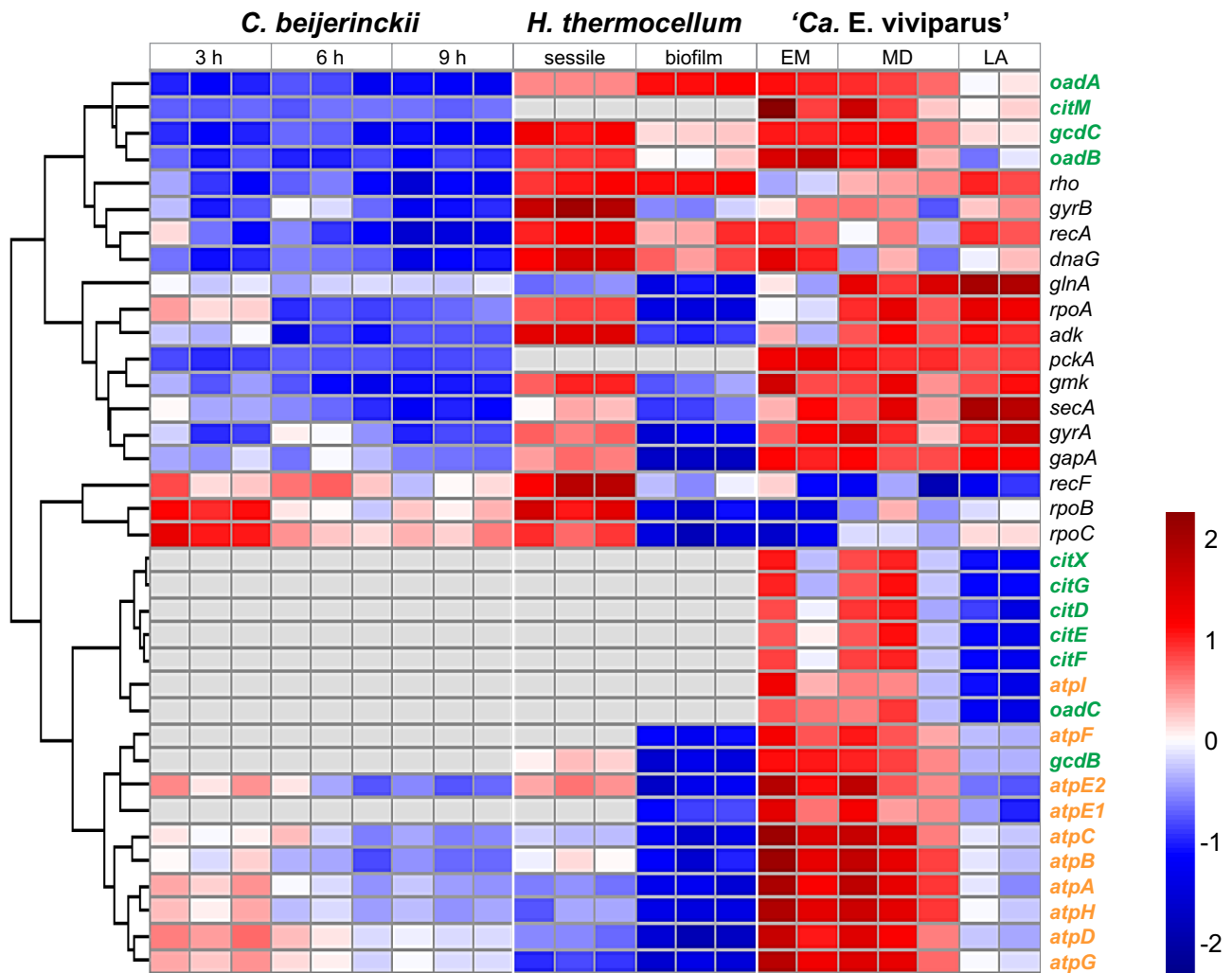
Fig. 3. Comparison of gene expression levels between *Clostridium beijerinckii* DSM 6423, *Hungateiclostridium thermocellum* ATCC 27405, and *Ca*. E. viviparus populations at different time points or growth conditions. Genes compared include subunits of the F_1F_0 ATP synthase (bold, orange text), citrate fermentation and oxaloacetate fermentation genes (bold, green text), and housekeeping genes (black text). Gene expression levels were taken at three different time points for *C. beijerinckii*, 3 h, 6 h, and 8 h, in triplicate (35). For *H. thermocellum*, expression was observed for both sessile and planktonic growth, in triplicate (36). For *Ca*. E. viviparus, different populations were assessed from the gut contents taken from different *Naso tonganus* specimens, caught either early morning (EM), midday (MD), or late afternoon (LA). In the wild, *N. tonganus* exhibits a diurnal feeding pattern. Samples here were selected to provide a view of *Ca*. E. viviparus activity at times when the fish host was feeding (details provided in *SI Appendix*, Table S2). RPKM values were log transformed and compared using pheatmap in R. The heatmap was scaled by rows (comparison between samples). Genes for which there were no expression data are in gray. The F_1F_0 ATPase subunits are more highly expressed in type B than the other species, but this only occurs at the EM and MD time points. Similar patterns emerge with the genes for OA and citrate fermentation in *Ca*. E. viviparus, though *H. thermocellum* planktonic cells have similar levels of expression as both EM and MD *Ca*. E. viviparus populations.

eight polysaccharide lyases (PL) from two families, and 26 carbohydrate esterases (CE) from nine different families (*SI Appendix,* Table S4). The degree and diversity of CAZymes present in the genome is slightly less than what is observed in the member of the Bacteroidota *Zobellia galactanivorans* Dsij, a heterotrophic marine bacterium, isolated from red algae, with the most CAZymes of any known bacterium (46). The abundance of COGs dedicated to carbohydrate metabolism, the diversity of CAZymes, and the comparable presence of these enzymes to free-living algal degrading bacteria, support the hypothesis that type B functions to degrade recalcitrant polysaccharides in algae consumed by its host.

The diet of *N. tonganus* is dominated by thallate and filamentous red algae (47, 48). The fish also feeds on green algae, such as *Caulerpa*. Brown algae are not a major component of the diet of *N. tonganus*; however, brown thallate and filamentous algae can be found in their gut. Consequently, the *N. tonganus* intestinal

microbiota have access to a diversity of algal structural polysaccharides (including agarose, porphyran, carrageenan, mannans, pectins, xyloglucans, arabinoglucans, and alginates), storage polysaccharides (laminarin, polyfructans), osmolytes, polyols, simple sugars, modified sugars, and sugar alcohols. The array of enzymes and transporters coded for in the type B genome can process and access this abundance of carbon and energy sources from all three algal types (Fig. 2). Imported oligosaccharides are degraded in the cytoplasm. Imported sugars and sugar alcohols feed the non-oxidative pentose phosphate pathway or are metabolized via the EMP or Leloir pathway. A more detailed analysis is presented in *SI Appendix, Supplemental Text* and summarized in *SI Appendix,* Table S5. An extensive study that characterized SCFA concentrations in the intestinal fluid of tropical marine herbivorous fishes reported mM levels of acetate (11.73 ± 0.94, n = 12) and propionate (0.63 ± 0.21) in regions inhabited by *Ca*. E. viviparus
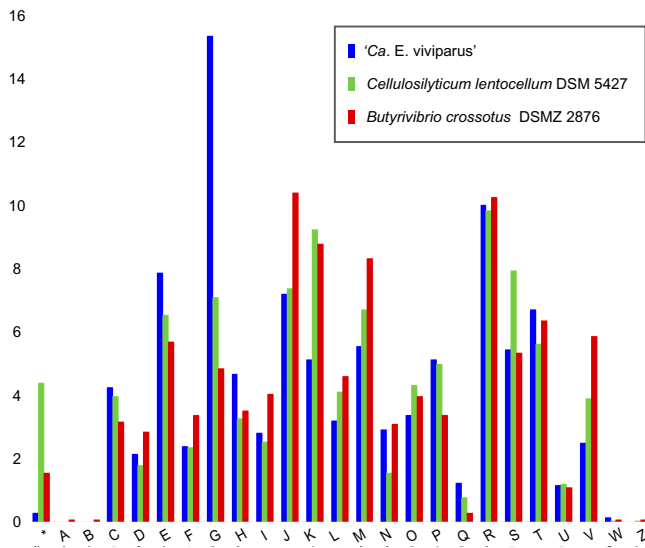
Fig. 4. COG comparison between *Ca*. E. viviparus and its relatives. *Butyrivibrio crossotus* DSM 2876 was used as a representative of the Lachnospiraceae XIVa and *C. lentocellum* DSM 5427 was used to represent the Lachnospiraceae XIVb. COG analysis reveals that *Ca*. E. viviparus is enriched for carbohydrate metabolism (COG G). Refer to *SI Appendix*, Table S3 for COG category codes and *SI Appendix*, Fig. S7 for the COG comparison against other Lachnospiraceae.

(47). Degradation and subsequent fermentation of a vast array of carbohydrates, in conjunction with citrate fermentation, likely supports the ability to attain such a large size and allows for the daily production of the greatest number of intracellular offspring of any known giant epulo.

Nitrogen Metabolism in *Ca*. Epulopiscium viviparus. *Ca*. E. viviparus has the potential to import inorganic nitrogen or take up amino acids (Fig. 2). The genome encodes the ability to produce all twenty proteinogenic amino acids, as nearly all the biosynthetic pathways are complete. It can synthesize L-lysine in a fashion typical of cyanobacteria (49), and L-methionine through a modified pathway (50). The pathway for L-serine lacks the haloacid dehalogenase (HAD) family hydrolase, phosphoserine phosphatase; however, three HAD hydrolase family proteins are present in the genome, so it is possible that one of these removes the phosphate instead. For further details, see *SI Appendix, Supplemental Text* (*SI Appendix*, Table S6).

To meet nitrogen requirements, the genome encodes an NH_3 transporter for uptake, as well as an NADPH-specific glutamate dehydrogenase to assimilate the NH_3 into L-glutamate. Furthermore, the genome contains a cluster of genes for urea uptake (*urtABCDE*) and catabolism (*ureABCEFGH*) via urease (51). This gives type B the ability to assimilate both forms of nitrogenous waste from either the host or other members of the microbiota. Teleost fish, like *N. tonganus*, typically produce NH_3, and this nitrogenous waste diffuses across the gills, while urea production is usually limited to the elasmobranchs, holocephalans, and coelacanths (52). However, there is precedent for urea cycling in teleost fish, as the microbiota of *Porichthys notatus* degrades urea in the intestine (53). The primary form of nitrogenous waste produced by *N. tonganus* is unknown, but the amount produced is probably lower than most carnivorous fishes. We hypothesize that the capacity for type B to metabolize urea and ammonia suggests a path for exchange of vital nutrients between fish and symbiont. For now, it is unclear whether the symbiont relies on acquisition of both inorganic and organic forms of nitrogen from the gut environment or whether amino acid uptake is primarily used by

intracellular offspring to support their rapid growth. Urease and urea transport gene expression is low in the early morning and higher in afternoon collections of *Ca*. E. viviparus. We also note that *glnA* is highly expressed in *Ca*. E. viviparus populations collected late in the growth cycle (Fig. 3), which suggests a shift toward NH_3 assimilation.

B Vitamin Metabolism of *Ca*. Epulopiscium viviparus. To meet its essential cofactor requirements, *Ca*. E. viviparus can synthesize multiple B vitamins (B_2, B_6, B_{12}) and has vitamin transporters (B_1, B_2, B_5, B_7, B_9) and salvage pathways (B_1, B_{12}). Details of vitamin acquisition systems are provided in *SI Appendix, Supplement Text* (*SI Appendix*, Table S7). The metabolism of vitamin B_3 (nicotinate) remains elusive. B_3 is used to make the essential cofactor NAD^+ (54). The lysis of mother cells may also contribute B-vitamins to the host, as the genomic capabilities to synthesize B_2, B_6, and B_{12} de novo are present. Microbiota provision of vitamins has been described in fish, such as in the freshwater Japanese eel, where B_{12} is provided by the gut microbes (55).

Phosphate Metabolism in *Ca*. Epulopiscium viviparus. *Ca*. E. viviparus has the genomic capacity to accumulate polyphosphate. The genome encodes a Na^+/P_i-dependent phosphate symporter, an exopolyphosphatase, and a polyphosphate kinase, with these genes being most highly expressed late in the day as offspring growth wanes. Like *Ca*. Thiomargarita spp., type B encodes multiple enzyme variants that use inorganic phosphate as a donor instead of ATP. This includes a copy of a polyphosphate-dependent NAD^+ kinase, as well as a pyrophosphate-dependent diphosphate-fructose-6-phosphate 1-phosphotransferase. Strikingly, nucleoside diphosphate kinase (*ndk*), used to generate the triphosphate forms of nucleotides and deoxynucleotides, is not present in the genome despite having the full complement of genes to make the diphosphorylated forms. Multiple Mollicutes lack *ndk* and instead use glycolytic kinases, and other nucleotide and deoxynucleotide diphosphates as phosphate acceptors (56). These kinases are present in the type B draft genome; therefore, it is plausible that *Ca*. E. viviparus generates these compounds in this fashion. In *Pseudomonas aeruginosa* and *E. coli* mutants lacking *ndk*, nucleoside triphosphates could still be generated from the polyphosphate-dependent polyphosphate kinase (Ppk) (27); type B might also employ this strategy.

The *Ca*. Epulopiscium viviparus Genome Encodes Giant Genes. It has been suggested that giant genes (>5 kb) are rare in bacteria and archaea due to the cost of producing large proteins (5), implying enhanced metabolism would be required to make them. Reva and Tümmler demonstrated that in fully sequenced bacterial and archaeal genomes, 0.2% of all the genes detected, were large (57). Strikingly, in *Ca*. E. viviparus 0.8% of the genes are >5 kb. Of these 21 genes, six are >10 kb and half of those are >15 kb. The largest gene encodes a 6,869-residue protein. The protein contains a secretion signal, transmembrane domain, as well as a fibronectin III domain (Fn3) (58), a vinculin conserved site, and a parallel β-helix repeat found in pectin and rhamnogalacturonan lyases (*SI Appendix*, Fig. S8). These features suggest that the protein is likely a surface-associated glycosyl hydrolase. The other large, secreted proteins (from genes over 9 kb in length) lack any discernable functional domains but appear to contain transmembrane helices. They are deplete in cysteine or arginine residues and are enriched for alanine, threonine, aspartate, and glutamate; these are signatures of giant surface proteins in other bacteria (57). Like other giant genes, some of the type B genes are highly repetitive (57). Proteins from four of the six genes that are over 10 kb in length lack an apparent secretion signal

or transmembrane domains, suggesting that these proteins are cytoplasmic. InterProScan identified immunoglobulin-like (Ig) fold domains in three of these four putative proteins. The second largest gene present encodes a non-secreted 6,833 amino acid protein with 31 Ig-fold domains (*SI Appendix,* Fig. S8). The third largest encodes a 5,531 amino acid protein with 21 Ig-fold domains. The repetition of Ig-like domains is reminiscent of the giant adhesin SiiE in *Salmonella enterica* (59). Expression levels of the giant genes are summarized in *SI Appendix,* Fig. S9.

Extracellular and Proposed Host-Association Structures and Functions of *Ca.* Epulopiscium viviparus. The *Ca.* E. viviparus genome codes for structures, such as flagella and secretion systems, that likely mediate interactions with its host. Unlike other giant bacteria, *Ca.* E. viviparus cells are rapid swimmers that proceed in either very long runs or in short runs where the leading pole reverses frequently (Movie S1). When observed on a microscope slide, a cell moves in a straight line while rotating on its long axis, and when the cell reverses direction, the cell rotates in the opposite direction (60). We have clocked swimming type B cells at speeds of 600 μm/s, rivaling speeds of the notoriously fast *Thiovulum majus* (61). Cells are covered with numerous peritrichous flagella and regularly spaced flagellar motors are observed in thin sections of the cell envelope (*SI Appendix,* Fig. S10). The genome encodes the full complement of genes to produce flagella (*SI Appendix,* Fig. S11) (62). Some *Bacillus* spp. have both H^+ translocating (MotAB) and Na^+ translocating (MotPS) stator proteins, and type B only has one set (63). Like the Na^+ powered flagella in *Bacillus* and *Vibrio* spp., *Ca.* E. viviparus MotB/PomB has a leucine rather than a valine at position 34, suggesting the type B flagella are fueled by SMF (64). We identified nine putative flagellin genes, with seven located adjacent to one another in an operon next to the *fla*/*che* operon. The other two flagellin genes are in separate operons, with one adjacent to a *fliS*. The presence of nine different flagellin proteins suggests a potential for them to play multiple roles for type B. Our RNAseq data show that not all flagellar filament genes (*fliC* or *hag*) are equally expressed in populations (*SI Appendix,* Fig. S12). It may be that some flagellins are better suited for different gut conditions, and the types of flagellin used are environmentally controlled (65). The variability in epitopes type B can create may also allow for it to adhere with greater avidity to a wider array of host-derived molecules or other gut contents (66).

The presence of nine different flagellin genes may be crucial for establishing host specificity and interactions with the fish immune system. Segmented filamentous bacteria (SFB), common members of vertebrate intestinal communities, have multiple flagellin genes in their genomes (67, 68). Flagella are present when newly emerged offspring establish their connection to the host epithelium. SFB flagellins interact with host immunoregulatory signaling pathways and mucosal proteins, and specific flagellin homologs have been demonstrated to confer host specificity of SFB strains (69). Likewise, the large variety of possible flagellin expression patterns in *Ca.* E. viviparus may modulate immune system signaling and allow cells to evade elimination by the host during colonization (65).

Type B appears to be chemotactic as it generally inhabits a small, defined region of the host intestinal tract (19). The genome contains all the genes necessary to employ the three chemotaxis adaptation systems exhibited in *B. subtilis* (70). Integral to chemotaxis are methyl-accepting chemotaxis proteins (MCPs), which are chemoreceptors that direct motility. Typically, MCPs have a ligand-binding domain (LBD), which can adopt a variety of folds, one or multiple transmembrane helices, a HAMP region, and a cytoplasmic signaling domain (71). We were able to identify three canonical MCPs. Two have calcium channels and chemotaxis receptors (Cache) domains in their LBD. The third has a taxis to aspartate and repellents (Tar) domain and a taxis to serine and repellents (Tsr) superfamily domain in its LBD. We identified seven other putative MCPs. The genomic content suggests that chemotaxis is important for type B to navigate the gut of *N. tonganus*.

Ca. E. viviparus likely produces complex cell wall–associated glycopolymers (*SI Appendix, Supplemental Text*) and encodes the ESX secretion system, which is only found in gram-positive bacteria, and functions in a species-specific manner. Wall associated polymers contribute to diverse functions such as virulence (72), spore formation (73), and bacterial competition (74, 75). In some species, like *B. subtilis,* the function of the ESX system is unknown. The *B. subtilis* system is characterized by having the FtsK/SpoIIIE family ATPase YukBA, which is encoded in the *yukEDCBA* operon containing structural genes and the secretion target, the small peptide YukE (76). The ESX-associated gene *yfiA* is also located in another location on the *B. subtilis* chromosome (77). We were able to uncover a modified form of this type VII secretion system operon in type B, which contains *yukBA, yukD,* and five *yukE* homologs. We were unable to find the *B. subtilis yueBC* homologs or *yukC* (76) but did locate a putative *yfiA* homolog. We were unable to detect any genes known to be involved in ESX-mediated competition. It is unclear why *Ca.* E. viviparus encodes this secretory system and why it has multiple copies of the secreted protein.

Gut microbiota can prevent the colonization of pathogens. The use of NO_3^- by *Ca.* E. viviparus may limit growth of some pathogens that use NO_3^- as a terminal electron acceptor. In mice, the microbiota regulates host NO_3^- production through the generation of butyrate, limiting colonization of pathogenic Enterobacteriaceae (78). The direct consumption of NO_3^- may serve an analogous function. The uncharacterized ESX system present in the genome may play some role in host–microbe interactions; as other bacteria use it in pathogenicity to blunt the host immune response (79), *Ca.* E. viviparus may use it in establishing itself as a predominant gut symbiont.

The Value of Efficient Energy Use. Giant bacteria have arisen in disparate phyla. The potential to become big does not rely on any singular environment, or metabolic strategy, nor is it supported by a specific genomic conformation or growth rate. Extreme polyploidy is one characteristic shared by giant bacteria. This certainly supports the transcriptional demands of these giant cells while working within the bounds of diffusion-limited transportation of biomolecules. Perhaps related to polyploidy, we found that *Ca.* E. viviparus and *Ca.* Thiomargarita spp. have the capacity to accumulate polyphosphate which may be important for DNA metabolism and energy conservation (26). A shift toward enzyme variants that use inorganic phosphate as a phosphate donor may be a way for bacterial giants to reserve ATP for a subset of cellular functions while supporting their replication and transcription needs. Further, the potential use of polyphosphate kinase to generate NTPs and dNTPs may have a twofold energy-saving effect. In addition to reducing the amount of ATP used to generate nucleotide triphosphates, it may make DNA synthesis more efficient. It is notable that Ppk typically localizes to the cell membrane (27), where in type B and *Thiomargarita* spp., the chromosomes are also located and replicated (20).

Additional energy-saving measures appear unique to *Ca.* E. viviparus. Some of these may have been selected for because *Ca.* E. viviparus metabolism relies on the activities of its fish host. If
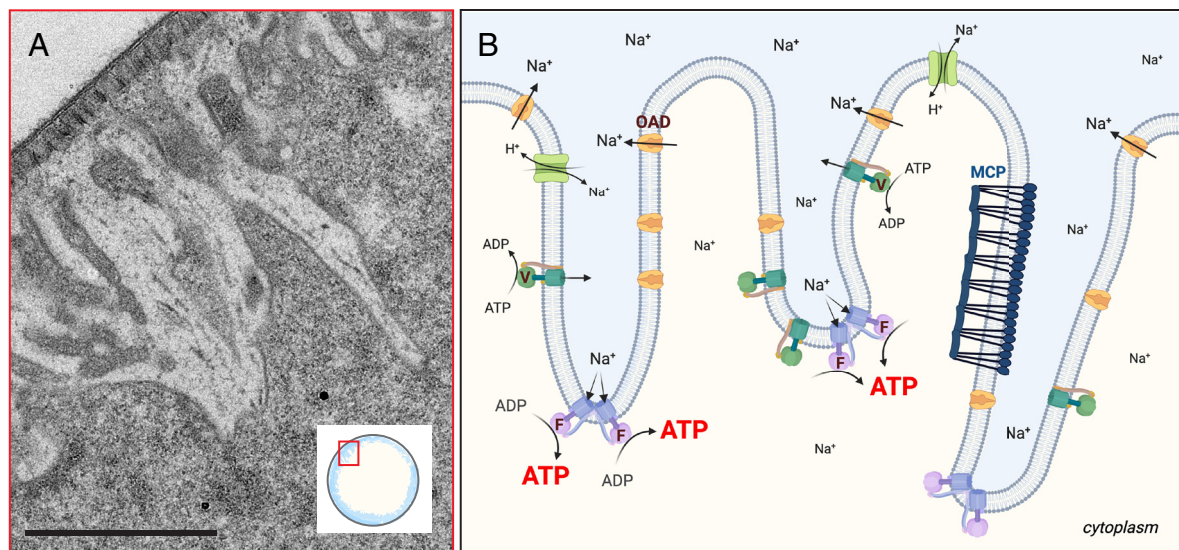
Fig. 5. *Ca.* E. viviparus membrane and key components of cell energetics. (*A*) transmission electron micrograph (TEM) of a portion of a cell. The insert illustrates the region of the cell cross-section (boxed) shown in the TEM. The TEM comprises part of the cell envelop, upper left, and invaginations of the cell membrane. (Scale bar, 1 μm.) (*B*) This drawing represents three cell membrane invaginations and integral membrane proteins that support the maintenance of the SMF. Features shown include the highly expressed OAD complex and a chemoreceptor array labeled "MCP." The two types of ATPases shown are also highly expressed; the Na$^+$-specific F$_1$F$_0$ ATP synthase (marked with "F") uses the SMF to generate ATP, while the V-type ATPase (designated with "V") likely serves as an ion pump. A proton/Na$^+$-antiporter encoded in the type B genome is also shown. Flagella are not shown but are powered by the SMF. We propose that the interactions between abundantly expressed membrane proteins, such as ATPases or MCPs, may drive membrane folding. Furthermore, we suggest that dimerization of the F-type ATPases may induce membrane invagination as reported in the mitochondrion (84). ATPase dimerization is shown here as overlapping F$_0$ regions of adjacent ATPases. Created with BioRender.

glycolytic kinases are playing the role of Ndk, cell growth would be directly coupled to central metabolism, as the cell would only be able to replicate DNA and transcribe genes during periods of fermentation. This may in part regulate the diurnal reproduction and growth cycle of *Ca.* E. viviparus populations, as DNA and RNA synthesis would be beholden on fish feeding activity to supply adequate fermentation substrates. It may explain why DNA deteriorates at the final stages of mother cell life (20), as they would no longer be able to make dNTPs for replication. It is also possible that during periods of high metabolic activity, glycolytic kinases, such as pyruvate, acetate, and PEP carboxykinase, which are all highly expressed when epulos are more metabolically active, can function as the Ndk, and during less active states, Ppk is used.

Multiple fused genes in biosynthetic pathways, improve pathway efficiency, reduce energy input (80), and reduce the number of proteins that have to co-localize at certain locations within the cell. Large bacterial cells would benefit from all of these. Comparing type B with its closest cultured relative, 5.82% of proteins encoded by the *Ca.* E. viviparus genome are fused according to IMG, which is in stark contrast to *C. lentocellum* DSM 5427, which has 0.97% of its proteins fused. The high recombination rates observed in type B populations (19) may have also selected for gene fusion events which would reduce the likelihood that metabolic pathways would be disrupted. Recombination may also account for why the type B mobilome is so reduced. Disrupted genes could be removed using intact copies located on nearby chromosomes or environmentally derived mother-cell DNA. This could be responsible for the differences in genome plasticity, coding density, and numbers of pseudogenes when compared to *Ca.* Thiomargarita spp.

The *Ca.* E. viviparus genome suggests that this microbial symbiont is tuned to its environment to maximize access to available organic carbon (e.g., algal carbohydrates) and nitrogen while providing nutritional benefits to the host. A versatile carbohydrate metabolism and the ability to synthesize essential resources (e.g., amino acids and vitamins) or salvage these resources when

available may allow these symbionts to maintain their standing with the host while not entirely relying on the host to provide more than just the raw materials (e.g., algal polysaccharides and inorganic nitrogenous waste). *Ca.* E. viviparus may also be contributing to host nutrition through the conversion of nitrogenous waste to protein. Microbial protein would supplement the unbalanced amino acid profile of marine algal proteins (81). Based on the daily lysis of mother cells, the release of protein from a resident *Ca.* E. viviparus population would happen when less food is entering the gut, therefore making that microbial protein more available to the host. The host may provide urea to its microbiota, which degrade it to NH$_3$ and incorporate it into biomass, which in turn gets recycled back to the host (82).

Cellular and Lifestyle Complexity of Giant Bacteria. Giant bacteria are exceptional in many ways. What is often underappreciated is the intracellular organization of *Epulopiscium* spp. and *Thiomargarita* spp. (4). Their complex cytoarchitectures are essential to support their metabolic vigor. Despite different morphologies, each of these bacteria maintains thousands of evenly spaced genome copies around the periphery of the cytoplasm. *Thiomargarita* spp. have cellular structures that allow them to cycle between sulfide-rich anoxic sediments and oxygen-rich sea water and survive these drastically fluctuating environmental conditions by using lithotrophy or mixotrophy. The proposed robust abilities for *Ca.* E. viviparus to breakdown and convert diverse carbohydrates fuels a remarkable fermentation scheme that supports rapid growth, reproduction, and motility. Based on metabolic reconstructions and RNAseq data, we suggest that OA decarboxylase and Na$^+$-dependent ATP synthase are key to the ability of type B to generate ATP using a SMF.

ATPase subunit genes and genes for other membrane proteins such as MCPs are among the most highly expressed genes of *Ca.* E. viviparus. We suggest that clustering of abundant membrane proteins is important for the generation of the internal membrane system. Membrane deformation in eukaryotes is driven by the insertion of BAR domain proteins (83), but we have been unable

to find any BAR domains in the predicted protein sequences coded for in the genome. The high expression levels of ATPases and MCPs, especially during offspring development, suggest an alternative means of developing the vast internal membrane system for large *Ca*. E. viviparus cells. As has been reported for the inner membrane of the mitochondrion, folds that make up cristae are formed by dimers of ATPases that bend the membrane and maintain the folds of the cristae (84). Overexpression of certain membrane proteins in *E. coli*, including ATPase subunits and the MCP Tsr, leads to a spontaneous increase in lipid production and membrane folds (85). We hypothesize that the abundant production of membrane proteins and possibly ATPase dimerization is occurring in the cell membrane of *Ca*. E. viviparus, thus providing a vast functional and yet cryptic surface area (Fig. 5). Thin sections of *Ca*. E. viviparus imaged using transmission electron microscopy (*SI Appendix*, Fig. S10) revealed bands of proteins that resemble chemoreceptor clusters (86) further supporting this hypothesis.

The membrane system provides additional surface area for the diverse transport proteins expressed by *Ca*. E. viviparus. The large volume of *Ca*. E. viviparus would accommodate the intracellular accumulation of oligo- and polysaccharides. Many of the GHs identified in the genome lack an obvious secretion signal; therefore, some polysaccharide degradation likely occurs intracellularly. Intracellular degradation would provide the added benefit of preventing other community members from accessing these compounds, giving a potential competitive advantage to type B, as has been described for other polysaccharide-degrading bacteria (87).

Conclusions

This genome-based study uncovered previously unseen advantages to being a giant bacterial symbiont. *Ca*. E. viviparus likely takes advantage of the Na$^+$-rich gut environment which is suitable to support a SMF for ATP synthesis using a fermentative metabolism and to drive flagellar rotation. Additionally, *Ca*. E. viviparus appears to have converged on a structure–function relationship previously described for the inner membrane of the mitochondrion. Furthermore, the identification of both exported and intracellular CAZymes suggests that the tremendous cell volume of *Ca*. E. viviparus may allow it to cache oligosaccharides to provide a substantial competitive advantage.

Advances in DNA sequencing technologies allow for the development of models to describe the metabolic potential of uncultivated bacteria, as we have done here for *Ca*. E. viviparus. Using transcriptomics, we were able to explore and confirm the expression of genes involved in many of the predicted key metabolic processes. However, additional studies are needed to further interrogate and assess these models and to extend our understanding of novel functions and their relationship to cell architecture. *Ca*. E. viviparus and other giant bacteria are exceptional candidates for microscopy-based analyses of metabolism. Tracking compounds, such as inorganic nitrogen species or polyphosphate, in situ with Raman microspectroscopy (88) could be used to confirm predicted metabolic pathways and the temporal importance of transformations. Further, stable-isotope probing (SIP) using labeled polysaccharides combined with metabolomics could potentially confirm how different algal and host-derived polysaccharides are catabolized and incorporated into the cell and may elucidate the crosstalk of metabolites between epulos and their fish host (89). *Ca*. E. viviparus genome codes for a remarkable array of CAZymes. How the cell employs these resources has yet to be fully appreciated. The biochemical characterization of these enzymes in vitro is needed to provide a more comprehensive view of potential polysaccharide use. Combined with in situ approaches, these data will inform a model of how *Ca*. E. viviparus prioritizes polysaccharide use which may impact its survival, its role in the larger microbial community, and its relationship with its host (90). Protein localization, as we have used previously for exploring the localization of cell division proteins (16), could allow for a more refined view of the role of integral membrane protein arrays in forming and maintaining membrane invaginations. These approaches may be used to observe a snapshot of cell metabolism, even in uncultivated bacteria, shortly after the subjects are taken from their natural environment.

We find it remarkable that the largest known bacteria have thus far eluded isolation. This suggests that bacterial behemoths are highly tuned to survive in the environments in which they evolved. The *Ca*. E. viviparus draft genome has now provided a foundation for understanding their growth requirements which may 1 d allow us to culture *Ca*. E. viviparus.

Methods Summary

Detailed methods are provided in *SI Appendix*.

Genomic DNA was extracted from ~35,000 handpicked and washed *Ca*. E. viviparus cells fixed in 80% ethanol, isolated from *N. tonganus* Nt_450 (19). Genomic DNA was prepared and sequenced using PacBio RSII with P6-C4 chemistry by the Yale Center for Genome Analysis. Illumina MiSeq sequencing (paired end, 2 × 250 bp) was performed at the Cornell University Biotechnology Resource Center. PacBio reads were either self-corrected using Canu 1.1 (91), or corrected with quality-controlled Illumina reads, using LoRDEC (92). Multiple assembly programs for genomes and metagenomes were tested and the LoRDEC corrected PacBio reads assembled with Canu 1.1 yielded the best assembly statistics and was subsequently polished. Select contigs generated from other assemblers were used to close gaps in the Canu contigs. The final assembly was assessed for completeness, heterogeneity, and contamination using CheckM (93) at both the phylum and class (Clostridia) levels.

The draft genome was annotated using IMG through the MGAP v4 pipeline (94) and through the Microscope platform of Genoscope (95). The genome was compared to other putative "type B" genomes from ref. 18 using BLAST Atlas in the GView server, and ANI was calculated using an ANI calculator (96). Metabolic reconstruction was based on KEGG and MicroCyc pathways. Missing genes in the pathways were uncovered by BLAST queries using homologs present in the *C. lentocellum* DSM 5427 and related genomes. To identify CAZymes, searches were performed using the CAZyme family-specific HMM database from dbCAN (97). Substrates for transporters were determined using the Transporter Classification Database (98). InterProScan was used to identify domains in the giant proteins (99). Tandem repeats were uncovered using Tandem Repeat Finder (100) as well as through manually scanning the genome. The Cluster of Orthologous Genes (COG) analysis was performed on the type B genome as well as on other draft or complete genomes of relatives.

RNA extraction and transcript analyses followed standard protocols which are described in *SI Appendix*.

Author affiliations: aDepartment of Microbiology, Cornell University, Ithaca, NY 14853; bSoil & Crop Sciences Section, School of Integrative Plant Sciences, Cornell University, Ithaca, NY 14853; cLaboratory for Research in Complex Systems, Menlo Park, CA 94025; dDepartment of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; and eDepartment of Energy Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

1. H. N. Schulz, B. B. Jørgensen, Big bacteria. *Annu. Rev. Microbiol.* **55**, 105–137 (2001).
2. K. D. Young, The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).
3. N. Lane, W. F. Martin, Mitochondria, complexity, and evolutionary deficit spending. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E666–E666 (2016).
4. N. Lane, How energy flow shapes cell evolution. *Curr. Biol.* **30**, R471–R476 (2020).
5. N. Lane, W. Martin, The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
6. P. C. Kirchberger, M. L. Schmidt, H. Ochman, The ingenuity of bacterial genomes. *Annu. Rev. Microbiol.* **74**, 815–834 (2020).
7. R. M. Bowers *et al.*, Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725 (2017).
8. K. Clements, D. Sutton, J. Choat, Occurrence and characteristics of unusual protistan symbionts from surgeonfishes (Acanthuridae) of the Great Barrier Reef, Australia. *Marine Biol.* **102**, 403–412 (1989).
9. S. Miyake, D. K. Ngugi, U. Stingl, Diet strongly influences the gut microbiota of surgeonfishes. *Mol. Ecol.* **24**, 656–672 (2015).
10. J. F. Flint, D. Drzymalski, W. L. Montgomery, G. Southam, E. R. Angert, Nocturnal production of endospores in natural populations of *Epulopiscium*-like surgeonfish symbionts. *J. Bacteriol.* **187**, 7460–7470 (2005).
11. E. R. Angert, K. D. Clements, N. R. Pace, The largest bacterium. *Nature* **362**, 239–241 (1993).
12. J. E. Mendell, K. D. Clements, J. H. Choat, E. R. Angert, Extreme polyploidy in a large bacterium. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6730–6734 (2008).
13. C. Robinow, E. R. Angert, Nucleoids and coated vesicles of "*Epulopiscium*" spp. *Arch. Microbiol.* **170**, 227–235 (1998).
14. E. Hutchison *et al.*, Developmental stage influences chromosome segregation patterns and arrangement in the extremely polyploid, giant bacterium *Epulopiscium* sp. type B. *Mol. Microbiol.* **107**, 68–80 (2018).
15. D. A. Miller, G. Suen, K. D. Clements, E. R. Angert, The genomic basis for the evolution of a novel form of cellular reproduction in the bacterium *Epulopiscium*. *BMC Genomics* **13**, 265 (2012).
16. E. R. Angert, K. D. Clements, Initiation of intracellular offspring in *Epulopiscium*. *Mol. Microbiol.* **51**, 827–835 (2004).
17. E. R. Angert, A. E. Brooks, N. R. Pace, Phylogenetic analysis of Metabacterium polyspora: Clues to the evolutionary origin of daughter cell production in *Epulopiscium* species, the largest bacteria. *J. Bacteriol.* **178**, 1451–1456 (1996).
18. D. K. Ngugi *et al.*, Genomic diversification of giant enteric symbionts reflects host dietary lifestyles. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7592–E7601 (2017).
19. F. A. Arroyo, T. E. Pawlowska, J. H. Choat, K. D. Clements, E. R. Angert, Recombination contributes to population diversification in the polyploid intestinal symbiont *Epulopiscium* sp. type B. *ISME J.* **13**, 1084–1097 (2019).
20. R. J. Ward, K. D. Clements, J. H. Choat, E. R. Angert, Cytology of terminally differentiated Epulopiscium mother cells. *DNA Cell Biol.* **28**, 57–64 (2009).
21. E. R. Angert, DNA replication and genomic architecture of very large bacteria. *Annu. Rev. Microbiol.* **66**, 197–212 (2012).
22. M. Winkel *et al.*, Single-cell sequencing of *Thiomargarita* reveals genomic flexibility for adaptation to dynamic redox conditions. *Front Microbiol.* **7**, 964 (2016).
23. B. E. Flood *et al.*, Single-cell (meta-) genomics of a dimorphic *Candidatus* Thiomargarita nelsonii reveals genomic plasticity. *Front Microbiol.* **7**, 603 (2016).
24. J.-M. Volland *et al.*, A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. *Science* **376**, 1453–1458 (2022).
25. M. Leaver, P. Dominguez-Cuevas, J. Coxhead, R. Daniel, J. Errington, Life without a wall or division machine in Bacillus subtilis. *Nature* **457**, 849–853 (2009).
26. E. Sanz-Luque, D. Bhaya, A. R. Grossman, Polyphosphate: A multifunctional metabolite in cyanobacteria and algae. *Front Plant Sci.* **11**, 938 (2020).
27. A. Kuroda, A. Kornberg, Polyphosphate kinase as a nucleoside diphosphate kinase in *Escherichia coli* and *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 439–442 (1997).
28. P. Dürre, Physiology and sporulation in *Clostridium*. *Microbiol. Spectrum* **2**, TBS-0010-2012 (2014).
29. J. Sridhar, M. A. Eiteman, J. W. Wiegel, Elucidation of enzymes in fermentation pathways used by *Clostridium thermosuccinogenes* growing on inulin. *Appl. Environ. Microbiol.* **66**, 246–251 (2000).
30. S. B. Crown *et al.*, Resolving the TCA cycle and pentose-phosphate pathway of *Clostridium acetobutylicum* ATCC 824: Isotopomer analysis, in vitro activities and expression analysis. *Biotechnol. J.* **6**, 300–305 (2011).
31. W. Buckel, Sodium ion-translocating decarboxylases. *Biochim. Biophys. Acta (BBA)-Bioenerg.* **1505**, 15–27 (2001).
32. Y.-T. Chen *et al.*, Genomic diversity of citrate fermentation in *Klebsiella pneumoniae*. *BMC Microbiol.* **9**, 168 (2009).
33. M. Bott, P. Dimroth, *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: Localization, sequencing, and expression. *Mol. Microbiol.* **14**, 347–356 (1994).
34. J. B. Warner, J. S. Lolkema, Growth of Bacillus subtilis on citrate and isocitrate is supported by the Mg2+-citrate transporter CitM. *Microbiology* **148**, 3405–3412 (2002).
35. H. M. de Gérando *et al.*, Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. *BMC Genomics* **19**, 242 (2018).
36. A. Dumitrache *et al.*, Specialized activities and expression differences for *Clostridium thermocellum* biofilm and planktonic cells. *Sci. Rep.* **7**, 43583 (2017).
37. P. Dimroth, B. Schink, Energy conservation in the decarboxylation of dicarboxylic acids by fermenting bacteria. *Arch. Microbiol.* **170**, 69–77 (1998).
38. W. Hilpert, B. Schink, P. Dimroth, Life by a new decarboxylation-dependent energy conservation mechanism with Na+ as coupling ion. *EMBO J.* **3**, 1665 (1984).
39. S. Kojima, K. Yamamoto, I. Kawagishi, M. Homma, The polar flagellar motor of Vibrio cholerae is driven by an Na+ motive force. *J. Bacteriol.* **181**, 1927–1930 (1999).
40. K. Fujinaga *et al.*, Analysis of genes involved in nitrate reduction in *Clostridium perfringens*. *Microbiology* **145**, 3377–3387 (1999).
41. C. Moreno-Vivián, P. Cabello, M. Martínez-Luque, R. Blasco, F. Castillo, Prokaryotic nitrate reduction: Molecular properties and functional distinction among bacterial nitrate reductases. *J. Bacteriol.* **181**, 6573–6584 (1999).
42. S. Seki, M. Satoh, Y. Seki, I. Yamamoto, H. Kondo, Nitrate reduction by pyruvate in *Clostridium perfringens*. *J. Gen. Appl. Microbiol.* **35**, 107–112 (1989).
43. S. Hasan, J. Hall, The physiological function of nitrate reduction in *Clostridium perfringens*. *Microbiology* **87**, 120–128 (1975).
44. M. Ishimoto, M. Umeyama, S. Chiba, Alteration of fermentation products from butyrate to acetate by nitrate reduction in *Clostridium perfringens*. *Zeitschrift für allgemeine Mikrobiologie* **14**, 115–121 (1974).
45. W. Jia, N. Tovell, S. Clegg, M. Trimmer, J. Cole, A single channel for nitrate uptake, nitrite export and nitrite uptake by *Escherichia coli* NarU and a role for NirC in nitrite export and uptake. *Biochem. J.* **417**, 297–307 (2009).
46. T. Barbeyron *et al.*, Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: Example of the model algae-associated bacterium *Zobellia galactanivorans* DsijT. *Environ. Microbiol.* **18**, 4610–4627 (2016).
47. K. Clements, J. Choat, Fermentation in tropical marine herbivorous fishes. *Physiol. Zool.* **68**, 355–378 (1995).
48. J. Choat, K. Clements, W. Robbins, The trophic status of herbivorous fishes on coral reefs. *Marine Biol.* **140**, 613–623 (2002).
49. A. J. Triassi *et al.*, L, L-diaminopimelate aminotransferase (DapL): A putative target for the development of narrow-spectrum antibacterial compounds. *Front Microbiol.* **5**, 509 (2014).
50. M. P. Ferla, W. M. Patrick, Bacterial methionine biosynthesis. *Microbiology* **160**, 1571–1584 (2014).
51. Y. H. Fong *et al.*, Structure of UreG/UreF/UreH complex reveals how urease accessory proteins facilitate maturation of *Helicobacter pylori* urease. *PLoS Biol.* **11**, e1001678 (2013).
52. A. Y. Ip, S. F. Chew, Ammonia production, excretion, toxicity, and defense in fish: A review. *Front Physiol.* **1**, 134 (2010).
53. C. Bucking, C. M. LeMoine, P. M. Craig, P. J. Walsh, Nitrogen metabolism of the intestine during digestion in a teleost fish, the plainfin midshipman (*Porichthys notatus*). *J. Exp. Biol.* **216**, 2821–2832 (2013).
54. O. Kurnasov *et al.*, NAD biosynthesis: Identification of the tryptophan to quinolinate pathway in bacteria. *Chem. Biol.* **10**, 1195–1204 (2003).
55. H. Sugita, C. Miyajima, Y. Deguchi, The vitamin B12-producing ability of the intestinal microflora of freshwater fish. *Aquaculture* **92**, 267–276 (1991).
56. J. D. Pollack, M. A. Myers, T. Dandekar, R. Herrmann, Suspected utility of enzymes with multiple activities in the small genome *Mycoplasma* species: The replacement of the missing "household" nucleoside diphosphate kinase gene and activity by glycolytic kinases. *Omics A J. Integrative Biol.* **6**, 247–258 (2002).
57. O. Reva, B. Tümmler, Think big-giant genes in bacteria. *Environ. Microbiol.* **10**, 768–777 (2008).
58. I. A. Kataeva *et al.*, The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. *Appl. Environ. Microbiol.* **68**, 4292–4300 (2002).
59. C. Wagner *et al.*, Functional dissection of SiiE, a giant non-fimbrial adhesin of *Salmonella enterica*. *Cell Microbiol.* **13**, 1286–1301 (2011).
60. E. R. Angert, "The enigmatic cytoarchitecture of *Epulopiscium spp*." in *Complex Intracellular Structures in Prokaryotes* (Springer, 2006), pp. 285–301.
61. F. Garcia-Pichel, Rapid bacterial swimming measured in swarming cells of *Thiovulum majus*. *J. Bacteriol.* **171**, 3560–3563 (1989).
62. S. Mukherjee, D. B. Kearns, The structure and regulation of flagella in *Bacillus subtilis*. *Annu. Rev. Genet.* **48**, 319–340 (2014).
63. M. Ito *et al.*, MotPS is the stator-force generator for motility of alkaliphilic *Bacillus*, and its homologue is a second functional Mot in *Bacillus subtilis*. *Mol. Microbiol.* **53**, 1035–1049 (2004).
64. N. Terahara, M. Sano, M. Ito, A *Bacillus* flagellar motor that can use both Na+ and K+ as a coupling ion is converted by a single mutation to use only Na+. *PLoS One* **7**, e46248 (2012).
65. J. R. McQuiston, P. I. Fields, R. V. Tauxe, J. M. Logsdon, Do *Salmonella* carry spare tyres? *Trends Microbiol.* **16**, 142–148 (2008).
66. Y. Rossez, E. B. Wolfson, A. Holmes, D. L. Gally, N. J. Holden, Bacterial flagella: Twist and stick, or dodge across the kingdoms. *PLoS Pathogens* **11**, e1004483 (2015).
67. I. Nkamba *et al.*, Intracellular offspring released from SFB filaments are flagellated. *Nat. Microbiol.* **5**, 34–39 (2020).

68. T. Kuwahara et al., The lifestyle of the segmented filamentous bacterium: A non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing. *DNA Res.* **18**, 291–303 (2011).

69. H. Chen, Y. Yin, Y. Wang, X. Wang, C. Xiang, Host specificity of flagellins from segmented filamentous bacteria affects their patterns of interaction with mouse ileal mucosal proteins. *Appl. Environ. Microbiol.* **83**, e01061-01017 (2017).

70. C. V. Rao, G. D. Glekas, G. W. Ordal, The three adaptation systems of *Bacillus subtilis* chemotaxis. *Trend Microbiol.* **16**, 480–487 (2008).

71. A. I. M. S. Ud-Din, A. Roujeinikova, Methyl-accepting chemotaxis proteins: A core sensing element in prokaryotes and archaea. *Cell Mol. Life Sci.* **74**, 3293–3303 (2017).

72. L. S. Ates, E. N. Houben, W. Bitter, Type VII secretion: A highly versatile secretion system. *Microbiol. Spectr.* **1**, (2016).

73. J. K. Fyans, D. Bignell, R. Loria, I. Toth, T. Palmer, The ESX/type VII secretion system modulates development, but not virulence, of the plant pathogen *Streptomyces scabies*. *Mol. Plant Pathol.* **14**, 119–130 (2013).

74. J. C. Whitney et al., A broadly distributed toxin family mediates contact-dependent antagonism between gram-positive bacteria. *eLife* **6**, e26938 (2017).

75. R. J. Ohr, M. Anderson, M. Shi, O. Schneewind, D. Missiakas, EssD, a nuclease effector of the Staphylococcus aureus ESS pathway. *J. Bacteriol.* **199**, e00528-00516 (2017).

76. L. A. Huppert et al., The ESX system in *Bacillus subtilis* mediates protein secretion. *PloS One* **9**, e96267 (2014).

77. D. Bottai, M. I. Gröschel, R. Brosch, Type VII secretion systems in gram-positive bacteria. *Curr. Topics Microbiol. Immunol.* **404**, 235–265 (2017).

78. M. X. Byndloss et al., Microbiota-activated PPAR-γ signaling inhibits dysbiotic Enterobacteriaceae expansion. *Science* **357**, 570–575 (2017).

79. M. I. Gröschel, F. Sayes, R. Simeone, L. Majlessi, R. Brosch, ESX secretion systems: Mycobacterial evolution to counter host immunity. *Nat. Rev. Microbiol.* **14**, 677 (2016).

80. C. S. Henry et al., Systematic identification and analysis of frequent gene fusion events in metabolic pathways. *BMC Genomics* **17**, 473 (2016).

81. M. Machado et al., Amino acid profile and protein quality assessment of macroalgae produced in an integrated multi-trophic aquaculture system. *Foods* **9**, 1382 (2020).

82. H. Mobley, R. Hausinger, Microbial ureases: Significance, regulation, and molecular characterization. *Microbiol. Rev.* **53**, 85–108 (1989).

83. M. Simunovic, E. Evergren, A. Callan-Jones, P. Bassereau, Curving cells inside and out: Roles of BAR domain proteins in membrane shaping and its cellular implications. *Annu. Rev. Cell and Dev. Biol.* **35**, 111–129 (2019).

84. T. B. Blum, A. Hahn, T. Meier, K. M. Davies, W. Kühlbrandt, Dimers of mitochondrial ATP synthase induce membrane curvature and self-assemble into rows. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 4250–4255 (2019).

85. J. Royes, V. Biou, N. Dautin, C. Tribet, B. Miroux, Inducible intracellular membranes: Molecular aspects and emerging applications. *Microbial Cell Factories* **19**, 1–26 (2020).

86. A. Briegel et al., Universal architecture of bacterial chemoreceptor arrays. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17181–17186 (2009).

87. G. Reintjes, C. Arnosti, B. M. Fuchs, R. Amann, An alternative polysaccharide uptake mechanism of marine bacteria. *ISME J.* **11**, 1640–1650 (2017).

88. R. Hatzenpichler, V. Krukenberg, R. L. Spietz, Z. J. Jay, Next-generation physiology approaches to study microbiome function at single cell level. *Nat. Rev. Microbiol.* **18**, 241–256 (2020).

89. D. Berry, A. Loy, Stable-isotope probing of human and animal microbiome function. *Trends Microbiol.* **26**, 999–1007 (2018).

90. C. Arnosti et al., The biogeochemistry of marine polysaccharides: Sources, inventories, and bacterial drivers of the carbohydrate cycle. *Annu. Rev. Marine Sci.* **13**, 81–108 (2021).

91. S. Koren et al., Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017).

92. L. Salmela, E. Rivals, LoRDEC: Accurate and efficient long read error correction. *Bioinformatics* **30**, 3506–3514 (2014).

93. D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).

94. V. M. Markowitz et al., IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2011).

95. D. Vallenet et al., MicroScope–an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Re.* **41**, D636–D647 (2012).

96. L. M. Rodriguez-R, K. T. Konstantinidis, The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ Preprints* **4**, e1900v1 (2016).

97. Y. Yin et al., dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **40**, W445–W451 (2012).

98. M. H. Saier et al., The transporter classification database (TCDB): Recent advances. *Nucleic Acids Res.* **44**, D372–D379 (2016).

99. P. Jones et al., InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).

100. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573 (1999).

101. D. R. Sannino, F. A. Arroyo, C. Pepe-Ranney, W. Chen, and E. R. Angert, *Epulopiscium* sp. 'N.t. morphotype B' strain M450 chromosome. GenBank. https://www.ncbi.nlm.nih.gov/nuccore/CP117982. Deposited 5 January 2023.

102. D. R. Sannino, F. A. Arroyo, W. Chen, and E. R. Angert, The transcriptional profile of 'Candidatus Epulopiscium viviparus' within *Naso tonganus* guts at different time points throughout its diurnal cycle Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE225926. Deposited 23 February 2023.

103. D. R. Sannino, Candidatus Epulopiscium viviparus assemblies. Ca.-Epulopiscium-viviparus. https://github.com/drs357/-Ca.-Epulopiscium-viviparus-/commit/b5c3b9546016eadc0fe532b7ed2094cac78dfbda. Deposited 28 November 2023.