AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology

| Environmental Microbiology | Minireview

Widespread dissolved inorganic carbon-modifying toolkits in genomes of autotrophic *Bacteria* and *Archaea* and how they are likely to bridge supply from the environment to demand by autotrophic pathways

Kathleen M. Scott,[1] Ren R. Payne,[1] Arin Gahramanova[1]

AUTHOR AFFILIATION See affiliation list on p. 18.

ABSTRACT Using dissolved inorganic carbon (DIC) as a major carbon source, as autotrophs do, is complicated by the bedeviling nature of this substance. Autotrophs using the Calvin-Benson-Bassham cycle (CBB) are known to make use of a toolkit comprised of DIC transporters and carbonic anhydrase enzymes (CA) to facilitate DIC fixation. This minireview provides a brief overview of the current understanding of how toolkit function facilitates DIC fixation in *Cyanobacteria* and some *Proteobacteria* using the CBB and continues with a survey of the DIC toolkit gene presence in organisms using different versions of the CBB and other autotrophic pathways (reductive citric acid cycle, Wood-Ljungdahl pathway, hydroxypropionate bicycle, hydroxypropionate-hydroxybutyrate cycle, and dicarboxylate-hydroxybutyrate cycle). The potential function of toolkit gene products in these organisms is discussed in terms of CO_2 and HCO_3^- supply from the environment and demand by the autotrophic pathway. The presence of DIC toolkit genes in autotrophic organisms beyond those using the CBB suggests the relevance of DIC metabolism to these organisms and provides a basis for better engineering of these organisms for industrial and agricultural purposes.

KEYWORDS autotroph, carbon fixation, carbonic anhydrase, carbon dioxide concentrating mechanism

The first step of the biological carbon cycle is the fixation of dissolved inorganic carbon (DIC; CO_2 + HCO_3^- + $CO_3^=$) by organisms consuming it via autotrophic and anaplerotic pathways [reviewed in reference (1)]. The entry of DIC into the biological carbon cycle is complicated by aspects of DIC that make it a tricky growth substrate. The composition of DIC is sensitive to pH; CO_2 dominates at low pH, HCO_3^- at circumneutral pH, and $CO_3^=$ at alkaline pH. The different forms of DIC have profound differences in geometry and charge (linear neutral CO_2 vs. trigonal planar anions HCO_3^- and $CO_3^=$). Due to these differences in geometry and charge, enzymes are specific to different forms of DIC (Table 1). Many key autotrophic enzymes are specific to CO_2, which is problematic since HCO_3^- is the most abundant form at physiological, circumneutral pH (2). Using HCO_3^- has its own complications; CO_2 diffuses through cell membranes more rapidly than HCO_3^- (3), due to higher permeability in phospholipid bilayers (4) and aquaporins (5). These difficulties in using CO_2 or HCO_3^- are exacerbated by the slow rate of uncatalyzed interconversion between them, relative to metabolism (2). Nature has responded to these challenges with a toolkit consisting of several carbonic anhydrase enzymes [EC 4.2.1.1; (6)] and DIC transporters (7, 8).

The function of this DIC toolkit has been studied in greatest detail in autotrophs from phylum *Cyanobacteria* and to a lesser extent among a limited number of autotrophic

Editor Arpita Bose, Washington University in St. Louis, St. Louis, Missouri, USA

Address correspondence to Kathleen M. Scott, kmscott@usf.edu.

The authors declare no conflict of interest.

See the funding table on p. 18.

Published 1 February 2024

TABLE 1 Substrate specificities of DIC-metabolizing enzymes from autotrophic and anaplerotic pathways

Enzyme	EC	Substrate	References
Malic enzyme[a]	1.1.1.38 1.1.1.39 1.1.1.40	CO_2	(9, 10)
Isocitrate dehydrogenase	1.1.1.41	CO_2	(11)
Isocitrate dehydrogenase with carboxylating factor for IDH	6.4.1.7 and 1.1.1.41	HCO_3^-	(12)
Pyruvate synthase	1.2.7.1	CO_2	(13)
2-Oxoglutarate synthase	1.2.7.3	CO_2	(13)
Carbon monoxide dehydrogenase/ acetyl-CoA synthase	1.2.7.4/2.3.1.169	CO_2	(14)
Formylmethanofuran dehydrogenase	1.2.7.12	CO_2	(15)
Formate dehydrogenase	1.17.1.10	CO_2	(16)
Phosphoenolpyruvate carboxylase	4.1.1.31	HCO_3^-	(17–19)
Phosphoenolpyruvate carboxykinase[a]	4.1.1.32	CO_2	(20)
	4.1.1.38		
	4.1.1.49		
Ribulose 1,5-bisphosphate carboxylase/oxygenase	4.1.1.39	CO_2	(21)
Pyruvate carboxylase	6.4.1.1	HCO_3^-	(20)
Acetyl-CoA/propionyl-CoA carboxylase	6.4.1.2/6.4.1.3	HCO_3^-	(22)
Oxaloacetate decarboxylase (Na^+ extruding)[a]	7.2.4.2	HCO_3^-[b]	

[a]Though these enzymes generally operate under physiological conditions as decarboxylases (23), they have been shown to be capable of acting as carboxylases (24, 25).
[b]The DIC substrate for this enzyme have not been directly measured. However, since it is a biotin carboxylase (26), it is likely to use bicarbonate as a substrate (27, 28).

Proteobacteria (see below). This understanding of the DIC toolkit is likely to be quite narrow, given that it focuses on organisms from two phyla within domain *Bacteria* using a single pathway (the Calvin-Benson-Bassham cycle). Autotrophy is broadly distributed among multiple phyla of *Archaea* and *Bacteria*, with eight autotrophic DIC fixation pathways known and more likely to be discovered [reviewed in reference (29, 30)]. Besides the Calvin-Benson-Bassham cycle (CBB) (31), there are the reductive citric acid cycle (rTCA) (32), Wood-Ljungdahl pathway (WL) (33), dicarboxylate/4-hydroxybutyrate cycle (DCHB) (34), hydroxypropionate/4-hydroxbutyrate cycle (HPHB) (35), hydroxypropionate bicycle (HP) (36), reverse oxidative citric acid cycle (roTCA) (37, 38), and reductive glycine pathway (39). Our limited understanding of DIC toolkit function, given how critical it is to using DIC as a growth substrate, hinders our understanding of DIC fixation in the many habitats where non-CBB organisms from many phyla catalyze reactions of geochemical importance and contribute to primary productivity. These habitats include the open ocean, sediments and soils, sewage, digestive tracts (e.g., rumen and termite hindguts), terrestrial and marine hot springs, deep-sea hydrothermal vents, and the subsurface (Table S1) (40, 41). Some of these habitats have high CO_2 concentrations, which could make a DIC toolkit less necessary for autotrophic growth; however, CO_2 in these habitats can be erratic or low (40, 42–44), and some organisms isolated from them have elaborate DIC toolkits (45). Beyond hamstringing our understanding of primary productivity in a huge variety of habitats, this narrow understanding of DIC toolkit function likely compromises efforts to engineer DIC-fixing organisms and enzymes to enhance crop yields, synthesize compounds of industrial relevance, and incorporate them into carbon-capture technologies (46, 47).

To address this lacuna, this review begins with a description of DIC toolkit components and their function in systems from *Cyanobacteria* and *Proteobacteria* that have been characterized and continues with DIC toolkit presence and predicted function in other phyla based on finished genome sequences from autotrophs using multiple pathways from both *Archaea* and *Bacteria*. The roTCA and reductive glycine pathways are not included because of a lack of marker genes and uncertainties in their taxonomic distributions. *Cyanobacteria* are also excluded from the genome comparisons, as genome surveys of their DIC toolkits have been previously published (48, 49)

COMPONENTS OF THE DIC TOOLKIT

Carbonic anhydrase (CA) catalyzes the hydration of CO_2 (forming H_2CO_3) and dehydration of H_2CO_3 (forming CO_2). Since the protonation and deprotonation of H_2CO_3 are instantaneous, CA activity speeds the interconversion of CO_2 and HCO_3^-, bringing them to chemical equilibrium much more rapidly than in the enzyme's absence (50). For example, in the enzyme's absence, under conditions similar to surface seawater (2 mM HCO_3^-), the initial rate of CO_2 production from H_2CO_3 is 0.05 mol sec^{-1} L^{-1} (25°C, k_D = 26 sec^{-1}) (51). This rate would be doubled by adding just 0.6–3 µmol of CA per liter [17–334 mg, based on kinetic parameters from one of the fastest (52) and slowest (53) forms of CA]. Catalyzing this interconversion is beneficial for enzymes using either CO_2 or HCO_3^- and also facilitates DIC accumulation or dissipation in cells by minimizing diffusive limitation of CO_2 across membranes (54). As a result, CA is extremely useful to autotrophs and heterotrophs and is ubiquitous among organisms from all three domains of life (55). This ubiquity is accompanied by enzyme diversity. Currently, there are at least six known evolutionarily independent forms of CA: alpha (α) (56), beta (β) (57), gamma (γ) (58), delta (δ) (59), epsilon (ε, CsoSCA; deeply divergent βCA) (53, 60, 61), zeta (ζ, may be deeply divergent βCA) (62), eta (η, may be deeply divergent αCA) (63), theta (θ) (64), and iota (ι) (65). The taxonomic distribution, mechanism, and structure of these enzymes were recently reviewed (6).

DIC transporters are similarly diverse and, among autotrophic prokaryotes, have been described from *Cyanobacteria* and *Proteobacteria*. HCO_3^- transporters from *Cyanobacteria* include three evolutionarily independent forms: SbtA (66) and BicA (a member of the SulP transporter family) (67), which rely on membrane potential for transport, and an ABC transporter (CmpABCD) (68). SbtA-family and SulP-family transporters active on HCO_3^- have also been studied in autotrophic *Proteobacteria*, and a Chr-family transporter was also found to transport HCO_3^- (45). Two evolutionarily distinct types of multisubunit complexes have been described to be active on CO_2 and facilitate HCO_3^- accumulation in cells. *Cyanobacteria* have two homologous complexes that couple vectoral CA activity (CO_2 hydrating direction only) to membrane potential via NADH dehydrogenase complexes (69). The second complex, the DIC accumulating complex (DAC), was discovered in *Proteobacteria* and is widespread in many other phyla in *Archaea* and *Bacteria*. It uses extracellular CO_2 as a substrate to generate elevated intracellular DIC concentrations; the mechanism of this complex remains to be elucidated (70–73).

DIC TOOLKIT FUNCTION IN AUTOTROPHIC *PROTEOBACTERIA* AND *CYANOBACTERIA* AND PERHAPS ONE AUTOTROPHIC MEMBER OF *BACILLOTA*

The best-studied system with respect to DIC toolkit function is the CO_2-concentrating mechanism (CCM) present in *Cyanobacteria* and some autotrophic *Proteobacteria*. This system was first suggested in reference (74) and consists of transporters and CA acting in concert to facilitate the growth of cells under low CO_2 conditions [reviewed in references (75–78)]. Transporters (SbtA, BicA/SulP, and CmpABCD) and CO_2-active systems [NADH dehydrogenase-associated vectoral CA, or DAC] generate elevated intracellular HCO_3^- concentrations (70, 73, 79). HCO_3^- then enters carboxysomes, which are polyhedral microcompartments with protein shells permeable to HCO_3^- but not CO_2 (80). Carboxysomes contain ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) and CA (CsoSCA in *Proteobacteria* and some *Cyanobacteria*); carboxysomal CA converts some of the HCO_3^- to CO_2, which is then fixed by RubisCO [reviewed in reference (81)]. One important aspect of CCM function is the necessity of spatial segregation of HCO_3^- delivery to the cytoplasm from (non-vectoral) CA activity in the carboxysome. Heterologous expression of human CA in the cytoplasm of Cyanobacterium *Synechococcus elongatus* results in loss of the ability to grow under low CO_2 conditions and massive CO_2 leakage from cells (82), illustrating that intracellular DIC is not in chemical equilibrium; instead, it is dominated by HCO_3^-, which is the form delivered to the cytoplasm by HCO_3^- transporters and CO_2-active complexes. The presence of extracellular CA has been

documented in organisms with CCMs (83, 84), but its role in facilitating DIC uptake in these organisms is unclear.

In *Cyanobacteria*, CCMs are upregulated under low CO_2 conditions [reviewed in reference (77)]. This is also the case among the limited number of *Proteobacteria* for which CCMs have been studied (45, 85). Some *Proteobacteria* with CCMs also carry genes encoding noncarboxysomal RubisCO. In these organisms, genes encoding carboxysome components and DIC transporters are upregulated under low CO_2 conditions; under moderate or high CO_2 conditions, these CCM genes are downregulated, while genes encoding noncarboxysomal RubisCOs are upregulated (45, 86, 87). These noncarboxysomal RubisCOs are very diverse; some are form I enzymes, with large (CbbL) and small (CbbS) subunits (carboxysomal Rubisco is also form I), while others are form II, with a single type of subunit (CbbM), homologous to form I large subunits [reviewed in (88)].

A few studies explore DIC toolkit function beyond CCMs. CA plays a role in DIC supply for some *Proteobacteria* lacking carboxysomes. Facultative CBB autotrophs *Rhodopseudomonas palustris* and *Ralstonia eutropha* (89) both require CA activity to grow under low CO_2 conditions. For *R. palustris*, this CA activity is extracellular, and likely to facilitate CO_2 uptake by keeping the periplasmic DIC pool near equilibrium (90). For *R. eutropha*, CA activity is intracellular (89) and presumably functions to provide HCO_3^- for anaplerotic reactions. CA genes are present in many nonoxygenic photoautotrophs, and enzyme activity in some photosynthetic *Alphaproteobacteria* is higher when grown autotrophically (91).

The study of DIC toolkits has been sparse for organisms using pathways besides the CBB cycle. Perhaps, this is because a DIC toolkit seems particularly important to organisms relying on the CBB cycle because of RubisCO's lack of specificity as a catalyst. RubisCO can use both CO_2 and O_2 as substrates (92). When RubisCO acts as an oxygenase, this activity is not productive for cellular growth; cells must regenerate the ribulose 1,5-bisphosphate consumed by the oxygenase reaction using pathways that consume ATP (93). CCMs act to raise the ratio of CO_2:O_2 in the cellular microenvironment of RubisCO, favoring the carboxylase activity over oxygenase (92). However, if RubisCO oxygenase activity were the sole factor driving CCM evolution, one would not expect chemolithoautotrophic organisms living in low-O_2 habitats to have CCMs, but many do (8, 45). This suggests that DIC toolkits should be present beyond CBB autotrophs. The only study available of a possible DIC toolkit in a non-CBB autotroph is one noting the activity of cytoplasmic CA activity in *Acetobacterium woodii* when growing autotrophically, and the authors suggest CA could play a role in facilitating DIC fixation by the WL pathway in this organism (94). Given the widespread nature of CCMs in CBB-using autotrophs from a variety of habitats, some of which co-exist with autotrophs using other pathways [e.g., reference (95)], it seems likely that DIC toolkits are relevant beyond CBB organisms.

FREQUENCY OF DIC TOOLKIT GENES AMONG GENOMES FROM *BACTERIA* AND *ARCHAEA*

Genes likely to encode DIC transporters and CAs are widespread in finished genomes from *Bacteria* and *Archaea* (Fig. 1A). Some toolkit genes are less abundant in *Archaea*, e.g., those encoding some forms of CA (α, δ, ζ, θ, and CsoSCA). Given that new forms of CA continue to be uncovered, the possibility exists that there are novel types of this enzyme that remain to be found. If the genomes are limited to those organisms with a documented ability to grow as autotrophs (Table S1), the level of toolkit gene incidence is higher (Fig. 1B). This is particularly noteworthy, as this smaller sample specifically excludes *Cyanobacteria*, for which the DIC toolkit function has already been extremely well documented (see above). The only gene family that diminishes in abundance is Pfam10070, which includes the cytoplasmic subunits of DACs. This gene family is not present in the autotrophic members of *Archaea* represented in Fig. 1B. DACs are found in members of *Euryarchaeota*, class *Halobacteria* (70, 73); these members are heterotrophs and therefore are not included. Though widespread among both autotrophs

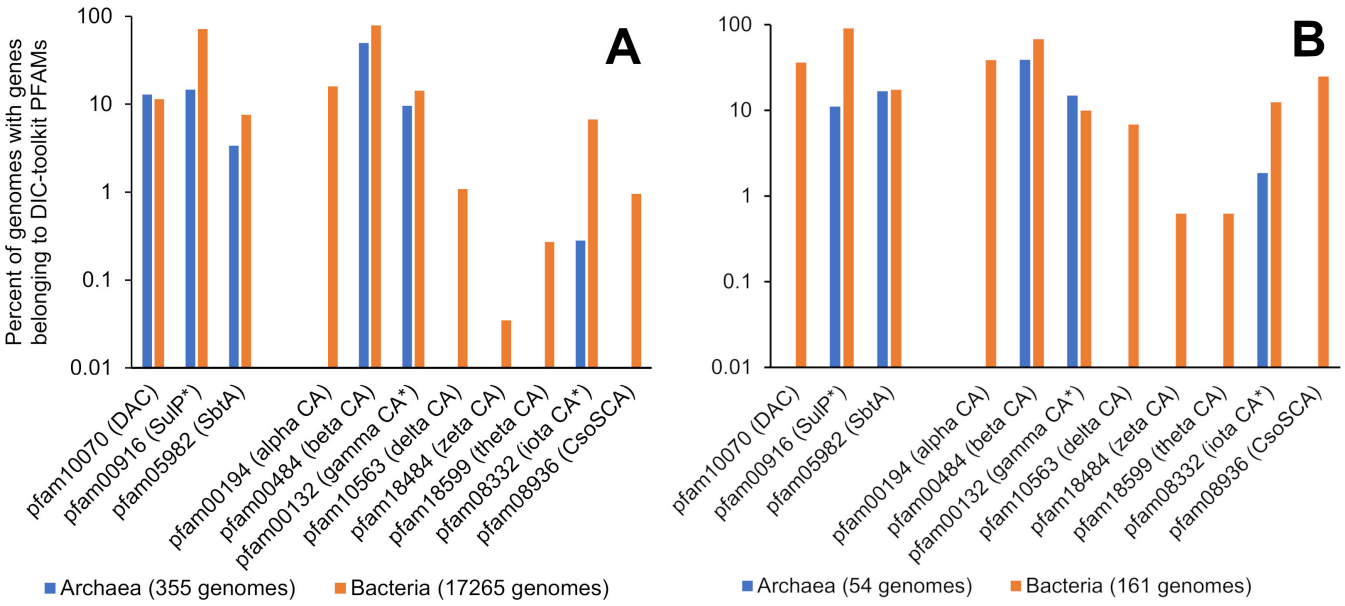

FIG 1 Prevalence of genes encoding DIC transporters (DAC, SulP, and SbtA) and carbonic anhydrase enzymes in finished genomes in the Integrated Microbial Genomes and Microbiomes database (https://img.jgi.doe.gov/) (96). (A) Percentage of all finished genomes in IMG with genes belonging to Pfams including DIC transporters and CA. Asterisks indicate Pfams that include members that do not metabolize DIC (Pfam00916 SulP includes sulfate transporters, Pfam00132 includes acyltransferases, and Pfam08332 includes protein kinases). (B) Percentage of all finished genomes in IMG from organisms capable of growing autotrophically, with genes belonging to Pfams including DIC transporters and CA. The genomes in B were the ones used for this study and represent organisms capable of fixing DIC via the Calvin-Benson-Bassham cycle, reductive citric acid cycle, Wool-Ljungdal pathway, hydroxypropionate bicycle, dicarboxylate-hydroxybutyrate cycle, or hydroxypropionate-hydroxybutyrate cycle. The procedure used for gathering these genomes is described in Supplemental Material.

and heterotrophs, the fact that toolkit gene abundance is particularly high among autotrophs strongly supports their relevance to autotrophic metabolism.

EVIDENCE THAT DIC TOOLKIT GENES ARE INVOLVED IN DIC FIXATION IN AUTOTROPHIC *BACTERIA* AND *ARCHAEA*

Prior study has provided many examples of the importance of DIC toolkit genes to autotrophic metabolism; genomic co-location of toolkit genes with those encoding steps of autotrophic DIC fixation pathways provides evidence for yet-to-be-studied connections between toolkit components and DIC fixation. The observation that genes encoding DIC toolkit components neighbor those encoding CBB pathway enzymes has precedence in the literature (8, 97), and only two of the many examples of this co-location are depicted here (Fig. 2). Carboxysome loci include *csoSCA* genes co-located with *cbbL* and *cbbS*, encoding the large and small subunits of carboxysomal form I RubisCO, and also commonly include DIC transporter genes (Fig. 2) (8, 45, 97). Noncarboxysomal RubisCO genes are also sometimes co-located with carbonic anhydrase genes (Fig. 2) (84, 86), which raises the possibility that CA facilitates carbon fixation by RubisCO.

There are some intriguing juxtapositions beyond those anticipated from prior study. Among organisms using the rTCA, DIC transporter or CA genes are co-located with genes encoding enzymes from the rTCA (Fig. 2), suggesting that there are yet-to-be-studied mechanisms for DIC toolkit interactions with this pathway. There is also a recurring juxtaposition in organisms using the WL pathway between genes encoding CA and formate hydrogenlyase or formate dehydrogenase (Fig. 2). It is hard to understand how CA in this context is used by these organisms to facilitate DIC fixation. Formate hydrogenlyase can oxidize formate to CO_2, while reducing protons to form hydrogen gas (98), though such a capability has yet to be measured in *Desulfocapsa sulfexigens*. Likewise, formate dehydrogenase in methanogenic *Archaea* functions in the formate oxidizing direction to reduce redox cofactor F_{420}, which is used primarily as a reductant

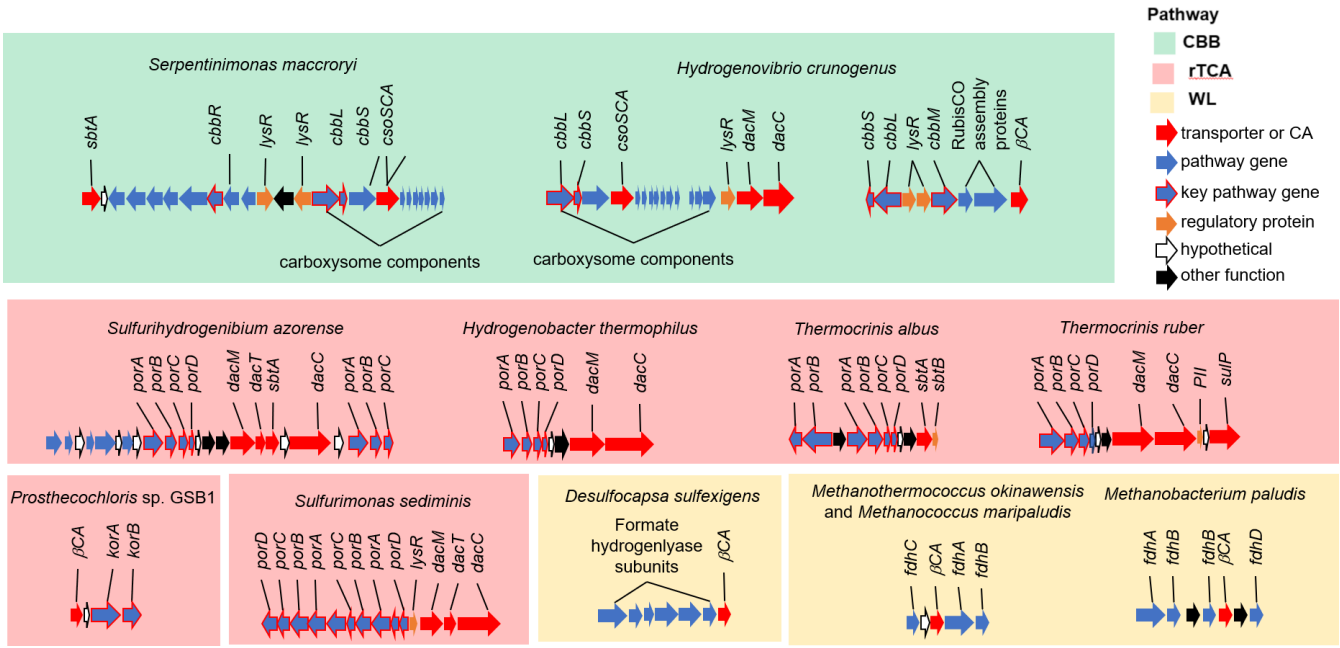

FIG 2 Colocation of genes encoding DIC transporters and carbonic anhydrase with genes from autotrophic DIC fixation pathways. "Pathway genes" encode enzymes catalyzing steps from autotrophic DIC fixation pathways. "Key pathway genes" encode enzymes catalyzing key steps from autotrophic DIC fixation pathways (e.g., CO_2- or HCO_3^--fixing enzymes: *cbbL*, *cbbS*: form I RubisCO; *cbbM*: form II RubisCO; *porABCD*, *korAB*: pyruvate or 2-oxoglutarate synthase; and *fdhABCD*: formate dehydrogenase).

for methanogenesis and to a minor degree by the WL pathway for cell biosynthesis (99). Instead of facilitating DIC fixation directly, perhaps these CAs facilitate the conversion of CO_2 produced from formate oxidation to HCO_3^-, which in turn could be used by a formate:bicarbonate antiporter to diminish the energetic expense of formate acquisition from the environment by making its acquisition electroneutral.

PHYLOGENETIC DISTRIBUTION OF DIC TOOLKIT GENES AMONG ORGANISMS CAPABLE OF AUTOTROPHIC GROWTH USING DIFFERENT PATHWAYS

The CBB, rTCA, WL, and HPHB are well represented among autotrophic organisms with finished genomes, while the HP and DCHB are much less so (Fig. 3 and 4). DIC toolkit genes are very broadly taxonomically distributed in autotrophic *Bacteria* (9 out of 10 phyla) and *Archaea* (all 3 phyla). Given that this sampling only includes finished genomes, which are a minority of sequenced genomes (~12% as of 4 August 2023; https://img.jgi.doe.gov/), it is likely that these genes are present in autotrophs from many other phyla. The toolkit is particularly well represented in organisms using the CBB in phyla *Proteobacteria*, *Bacillota*, and *Actinomycetota*, as well as organisms using the rTCA in phyla *Campylobacterota* and *Aquificota* and those using the HP in *Chloroflexota*.

Toolkit genes are less abundant among autotrophic *Archaea* (Fig. 4). βCA and γCA are represented among the phyla, while SbtA transporters are present in some members of *Nitrososphaerota*. Given the relative abundance of toolkit components in autotrophic *Bacteria* and the recent discoveries of novel forms of CA (65) and DIC accumulation (70–73), it seems likely that this paucity reflects the fact that domain *Archaea* is comparatively understudied.

Patterns of gene presence and absence sometimes follow organism taxonomy (Fig. 3 and 4). For example, all members of the *Ruthia/Vesicomyosocius/Thioglobus/Bathymodiolus* symbiont clade lack DIC transporter genes and carry βCA genes (Fig. 3). However, there are many departures from taxonomy. There is within-genus divergence. Both members of *Hydrogenovibrio* have genes encoding DAC, CsoSCA, and βCA, but genes encoding SulP, αCA, and ιCA are not present in both. Rather, extreme divergence

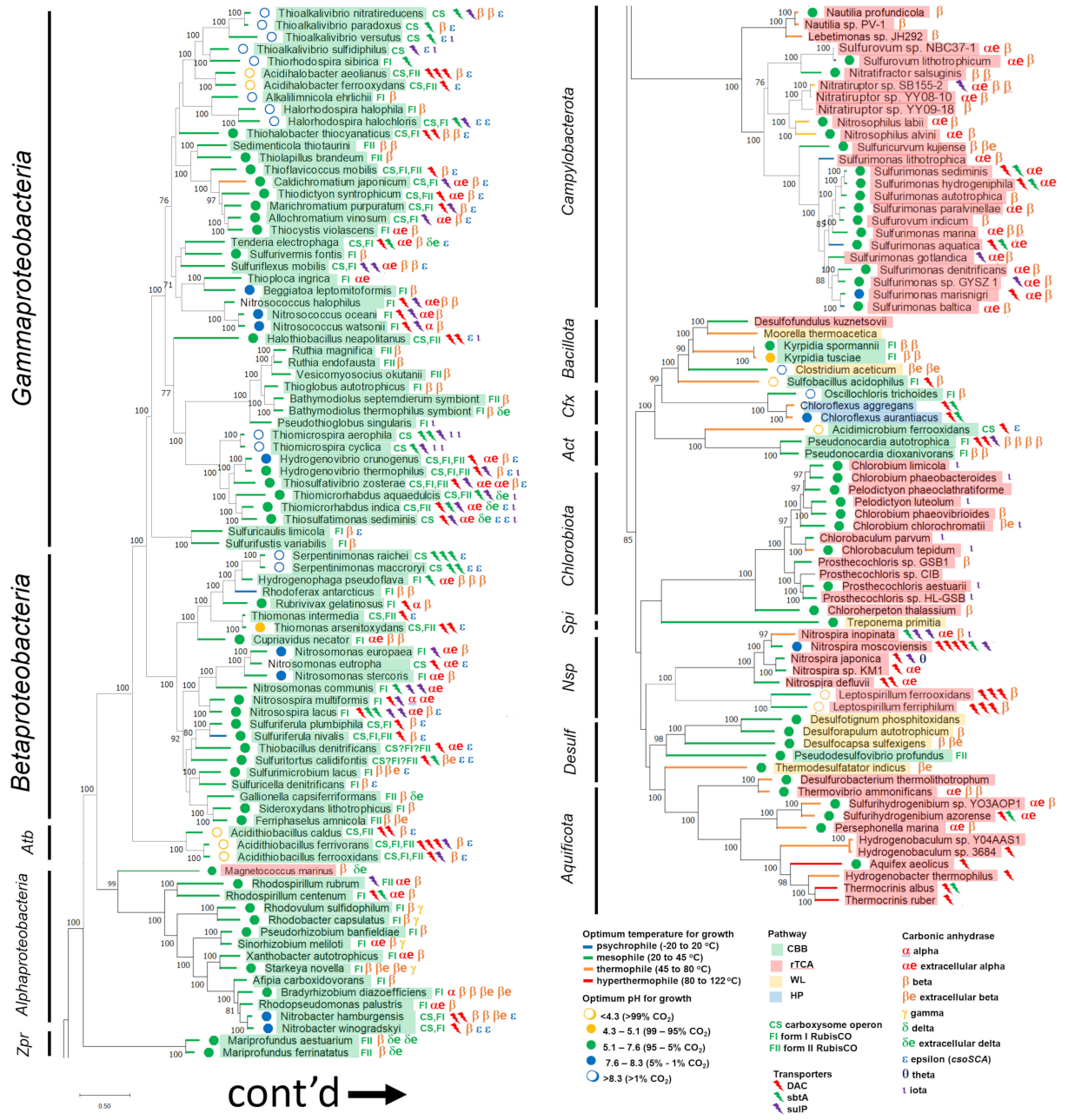

FIG 3 Taxonomic distribution of genes encoding DIC transporters (DAC, SulP, and SbtA) and carbonic anhydrase enzymes among members of *Bacteria*. Optimum growth conditions and autotrophic DIC fixation pathway are also provided. Maximum likelihood trees are based on concatenated alignments of amino acid sequences predicted from genes encoding ribosomal proteins. Genes were gathered, aligned, and concatenated from Ribosomal MLST [https://pubmlst.org/species-id (100)]). This alignment of 9,939 positions was used to generate a maximum likelihood tree in MEGA 11 (101) after finding the best model [Le-Gascuel (102), gamma distribution (five categories), and invariant sites]. Bootstrap values are based on 100 resamplings of the alignment. Phyla and classes were gathered from https://lpsn.dsmz.de/, the List of Prokaryotic names with Standing in Nomenclature, with the following exceptions: "*Desulfobacterota*" are based on reference (103), and *Candidatus* Zetaproteobacteria are based on (104). Autotrophic pathways were inferred from genome sequences and the literature, and optimum pH and temperatures for growth were gathered from the literature as well (Table S1). Predicted functions for gene products from genes encoding potential DIC transporters or carbonic anhydrase enzymes were verified using predictions of transmembrane helices (transporters) and conserved residues (carbonic anhydrase) as described in Table S2. Extracellular locations for carbonic anhydrase enzymes were predicted using SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) (105). The fraction of DIC present as CO_2 at optimal growth pH, when available, was calculated using $pK_1 = 6.35$ and $pK_2 = 10.33$ (2). *Act*, *Actinomycetota*; *Atb*, *Acidithiobacillia*; *Cfx*, *Chloroflexota*; *Desulf*, "*Desulfobacterota*"; *Nsp*, "*Nitrospirae*;" *Spi*, *Spirochaetota*; *Zpr*, *Candidatus* Zetaproteobacteria; CBB, Calvin-Benson-Bassham cycle; HP, hydroxypropionate bicycle; rTCA, reductive citric acid cycle; WL, Wood-Ljungdal pathway.

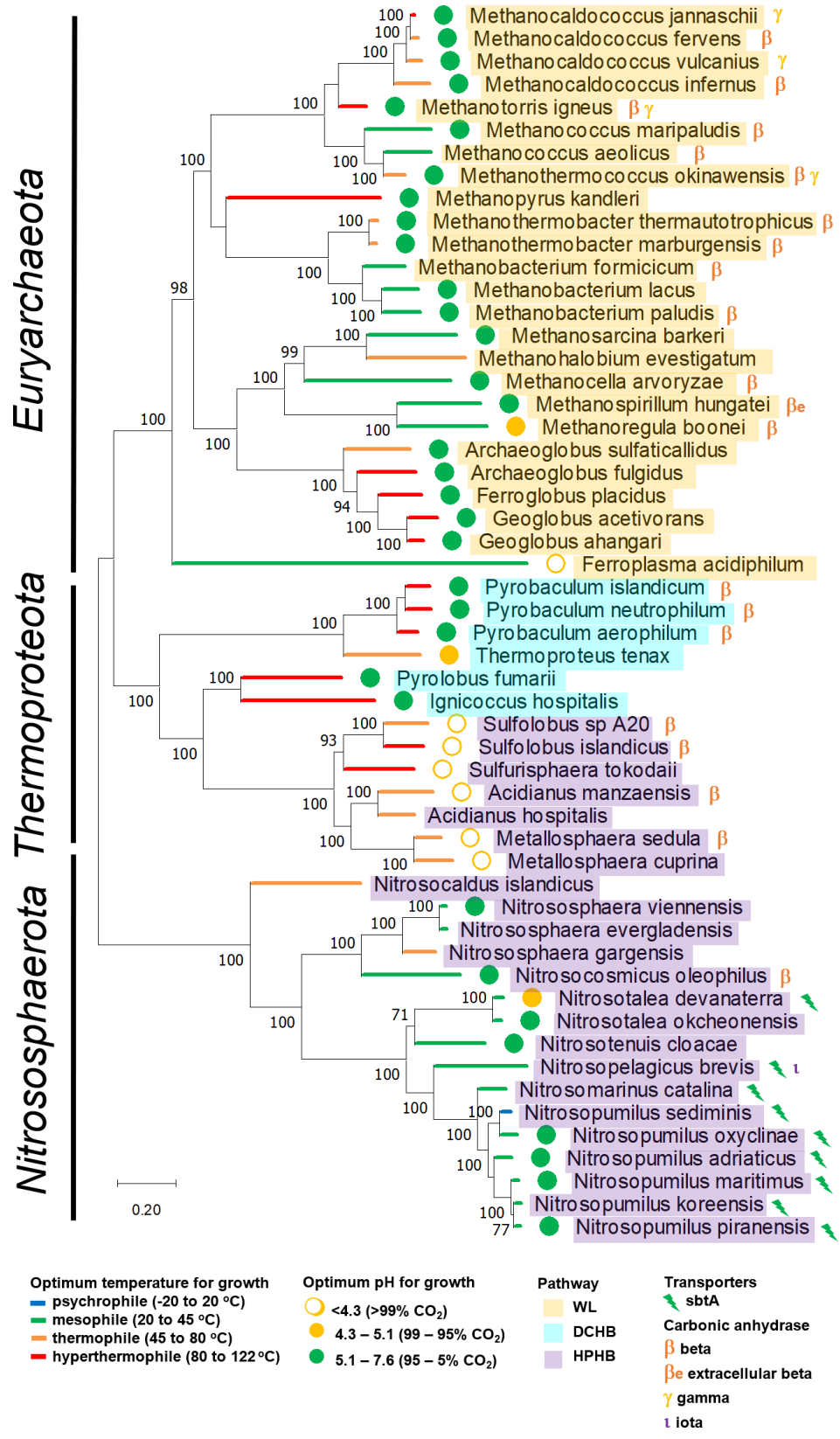

FIG 4 Taxonomic distribution of genes encoding DIC transporters (DAC, SulP, and SbtA) and carbonic anhydrase enzymes among members of *Archaea*. Optimum growth conditions and autotrophic DIC fixation pathway are also provided. Maximum likelihood trees are based on concatenated alignments of amino acid sequences predicted from genes encoding ribosomal

(Continued on next page)

FIG 4 (Continued)

proteins. Genes were gathered from genome sequences using COGs comprised of ribosomal large and small subunits. Amino acid sequences predicted from genes encoding each subunit were aligned via MUSCLE (MUltiple Sequence Comparison by Log-Expectation) (106) and concatenated using a script available from https://github.com/scooterboi85/Gene-concatenator, resulting in an alignment of 8,612 positions. Maximum likelihood analysis was implemented as described in Fig. 3. Phyla and classes were gathered from https://lpsn.dsmz.de/, the List of Prokaryotic names with Standing in Nomenclature. Autotrophic pathways were inferred from genome sequences and the literature, and optimum pH and temperatures for growth were gathered from the literature as well (Table S1). For members of genus *Pyrobaculum*, genome data suggest the DCHB pathway, but other evidence is less conclusive (107, 108). Predicted functions and cellular locations for gene products from genes encoding potential DIC transporters or carbonic anhydrase enzymes were verified as described in Fig. 3; Table S2. The fraction of DIC present as CO_2 at optimal growth pH, when available, was calculated using pK_1 = 6.35 and pK_2 = 10.33 (2). DCHB, dicarboxylate-hydroxybutyrate cycle; HPHB, hydroxypropionate-hydroxybutyrate cycle; WL, Wood-Ljungdal pathway.

is apparent within genus *Pseudonocardia*; both members of this genus carry genes encoding βCA, but one member (*autotrophica*) carries three DIC transporter genes, while the other (*dioxanivorans*) has none. The autotrophic DIC-fixing pathway and environment appear to play a role in DIC toolkit distribution. Organisms using the HPHB appear to have toolkit components that correlate with their optimal pH for growth, while organisms from multiple phyla using the CBB or rTCA are particularly "loaded," suggesting DIC toolkit distribution might be convergent with the autotrophic DIC fixation pathway (Fig. 3 and 4). Accordingly, the following sections explore the correlation between DIC toolkit components and the environment and autotrophic pathway.

DISTRIBUTION OF DIC TOOLKIT GENES RELATIVE TO ENVIRONMENTAL DIC SUPPLY

Autotrophs in this study have optimal pH values for growth ranging from 1.4 to 11 (Table S1) and therefore thrive in environments with dramatic differences in DIC composition. DIC composition is sensitive to pH, with CO_2 dominating below the pK_1 for carbonic acid (~pH 6.4), $CO_3^{=}$ dominating above the pK_2 (~10.3), and HCO_3^- dominating at circumneutral pH (2), where cytoplasmic pH is poised, even in acidophilic and alkaliphilic microorganisms (109, 110).

The ways in which CAs and DIC transporters could potentially facilitate growth in environments with differing DIC compositions is illustrated in Fig. 5A. DIC transporters using CO_2 or HCO_3^- could facilitate uptake at different environmental pH values, and extracellular CA could prevent the concentration of CO_2 or HCO_3^- from dropping below equilibrium values if consumed by the cell.

The presence of DIC toolkit components does correlate with pH (Fig. 5B). The distribution of different DIC transporters does seem to follow the environmental abundance of the form of DIC transported: DACs are absent in organisms with pH optima above 8.3 and SbtA transporters are absent in organisms with pH optima below 4.3, conditions where their substrates (CO_2 or HCO_3^-, respectively) are less than 1% of DIC. This trend mirrors what has been observed in metagenomes (73). Similar to SbtA, genes encoding SulP transporters likely to be active on HCO_3^- are more abundant in organisms growing at high pH (Fig. 5B). The ability to transport HCO_3^- by this type of transporter was predicted by phylogenetic analysis (Fig. S1); these predictions would be stronger if more SulP family transporters were biochemically characterized, since these transporters are active on a variety of compounds (111, 112).

Genes predicted to encode extracellular CA are absent from organisms growing below pH 5.1 or above 8.3 (Fig. 5B). One possibility is that this distribution indicates that these enzymes are pH labile. Another possibility is that these enzymes would not be particularly helpful at extremely acidic or alkaline pH; their ability to bring DIC to equilibrium would not facilitate CO_2 or HCO_3^- supply when taking place at pH values where either CO_2 or HCO_3^- are extremely scarce at equilibrium.

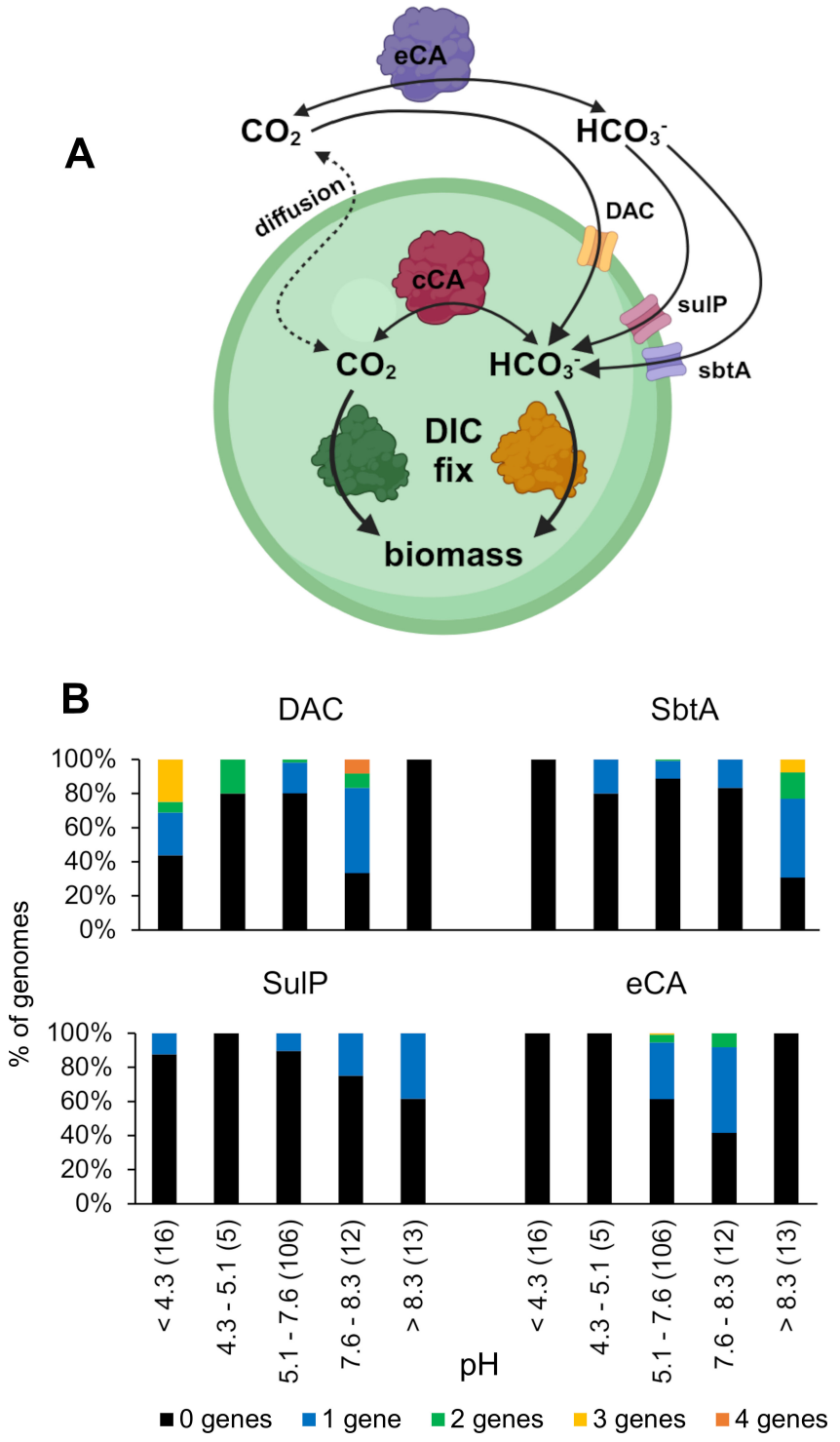

FIG 5 Potential functioning of DIC toolkit components within a cell and with the environment. (A) General model of an autotrophic cell, showing the location of DIC toolkit components. (B) Trends in DIC toolkit component presence and abundance with optimum growth pH for the host organism. pH ranges were chosen to reflect the following % of DIC that is in the form of CO_2: <4.3: at least 99% CO_2; 4.3–5.1: 99 – 95% CO_2; 5.1–7.6: 95 – 5% CO_2; 7.6–8.3: 5 – 1% CO_2; and >8.3: less than 1% CO_2. Numbers in parentheses are the numbers of genomes in each category. cCA, cytoplasmic carbonic anhydrase; DAC, DIC accumulating complex; DIC fix, DIC fixation; eCA, extracellular carbonic anhydrase; SbtA, SbtA family transporter; SulP, SulP family transporter.

CYTOPLASMIC CO_2 AND HCO_3^- DEMAND BY DIFFERENT AUTOTROPHIC PATHWAYS

Some of the carboxylases catalyzing autotrophic, anaplerotic, and biosynthetic DIC fixation use CO_2 as a substrate, while others use HCO_3^- (Table 1). As a result, organisms using different autotrophic DIC-fixing pathways have differing demands for cytoplasmic CO_2 and HCO_3^- for synthesizing the metabolic intermediates necessary for generating biomass (Fig. 6). Pathways which predominantly incorporate CO_2 into biomass include CBB, rTCA, and WL, though they also require HCO_3^- for oxaloacetate synthesis likely by phosphoenolpyruvate carboxylase or pyruvate carboxylase (Table 1; Fig. 6A). The contributions of CO_2 and HCO_3^- to the biomass of DCHB autotrophs are more evenly split, while HCO_3^- is the dominant form of DIC incorporated by organisms using the HP and HPHB pathways. HCO_3^- is also the dominant form of DIC incorporated by organisms with carboxysomes, even though they use the CBB cycle. In these organisms, RubisCO draws from the pool of CO_2 present in carboxysomes, which originated from cytoplasmic HCO_3^- that was dehydrated by carboxysomal CA after entering carboxysomes (74, 80). For organisms whose genomes encode both carboxysomal as well as noncarboxysomal RubisCO, the contributions of HCO_3^- and CO_2 to biomass will depend on whether the cells are growing under conditions when carboxysome synthesis is induced (e.g., low CO_2) or when noncarboxysomal RubisCO is predominant (e.g., high CO_2).

DISTRIBUTION OF DIC TOOLKIT GENES IN AUTOTROPHIC ORGANISMS RELYING PRIMARILY ON CO_2

Given the large differences in the demand for CO_2 and HCO_3^- predicted for organisms using different autotrophic DIC fixation pathways, it is not surprising that organisms using them have large differences in DIC toolkits (Fig. 7). For organisms relying primarily on CO_2 (CBB, rTCA, and WL), genes encoding DIC transporters are less abundant than for those organisms relying primarily on HCO_3^-.(Fig. 7A). Genes encoding CA are quite common and vary among the pathways (Fig. 7B through D). For cells without DIC transporters, provided that environmental pH is not alkaline enough to make extracellular CO_2 scarce, CO_2 can diffuse into cells through the membranes or aquaporins before fixation by CO_2-requiring carboxylases in the cytoplasm. However, these cells also require some HCO_3^- for oxaloacetate and pyrimidine synthesis, which could be provided by either cytoplasmic CA (cCA) from intracellular CO_2 or DIC transporters from extracellular DIC (Fig. 5A). Indeed, most organisms relying primarily on CO_2 have genes encoding either cCA or DIC transporters (Fig. 8). Some have genes encoding both, which could be a conundrum.

The simultaneous presence of both cCA and DIC transporters is problematic, as cCA would convert cytoplasmic HCO_3^- delivered by transporters into CO_2, which would diffuse out of the cell, dissipating the electrochemical gradients that DIC transporters couple to HCO_3^- acquisition (45, 66–68, 70–73). Indeed, when *Cyanobacteria* with DIC transporters are engineered to express cCA, massive amounts of CO_2 diffuse out of them (82). Perhaps organisms whose genomes encode both cCA and DIC transporters differentially express them, so that they are not present simultaneously. However, recent models indicate that low to moderate levels of co-expression of cCA and DIC transporters can facilitate the simultaneous supply of CO_2 and HCO_3^- for biosynthesis without CO_2 leakage (116).

Interestingly, 10 of the 14 CBB and rTCA organisms with genes encoding both DIC transporters and cCA are likely exposed to N_2O gas during growth. Six use ammonia as their electron donor (and produce N_2O as a by-product) (117), three use nitrite as an electron donor and likely are exposed to N_2O produced by the ammonia-oxidizing microorganisms with which they commonly co-occur (118), and one uses N_2O as an electron donor. Given that N_2O molecules are similar in size and shape to CO_2, perhaps, CO_2-dependent carboxylases and potentially also CA (but not αCA) (119) in these organisms are sensitive to this dissolved gas and the additional DIC toolkit compensates

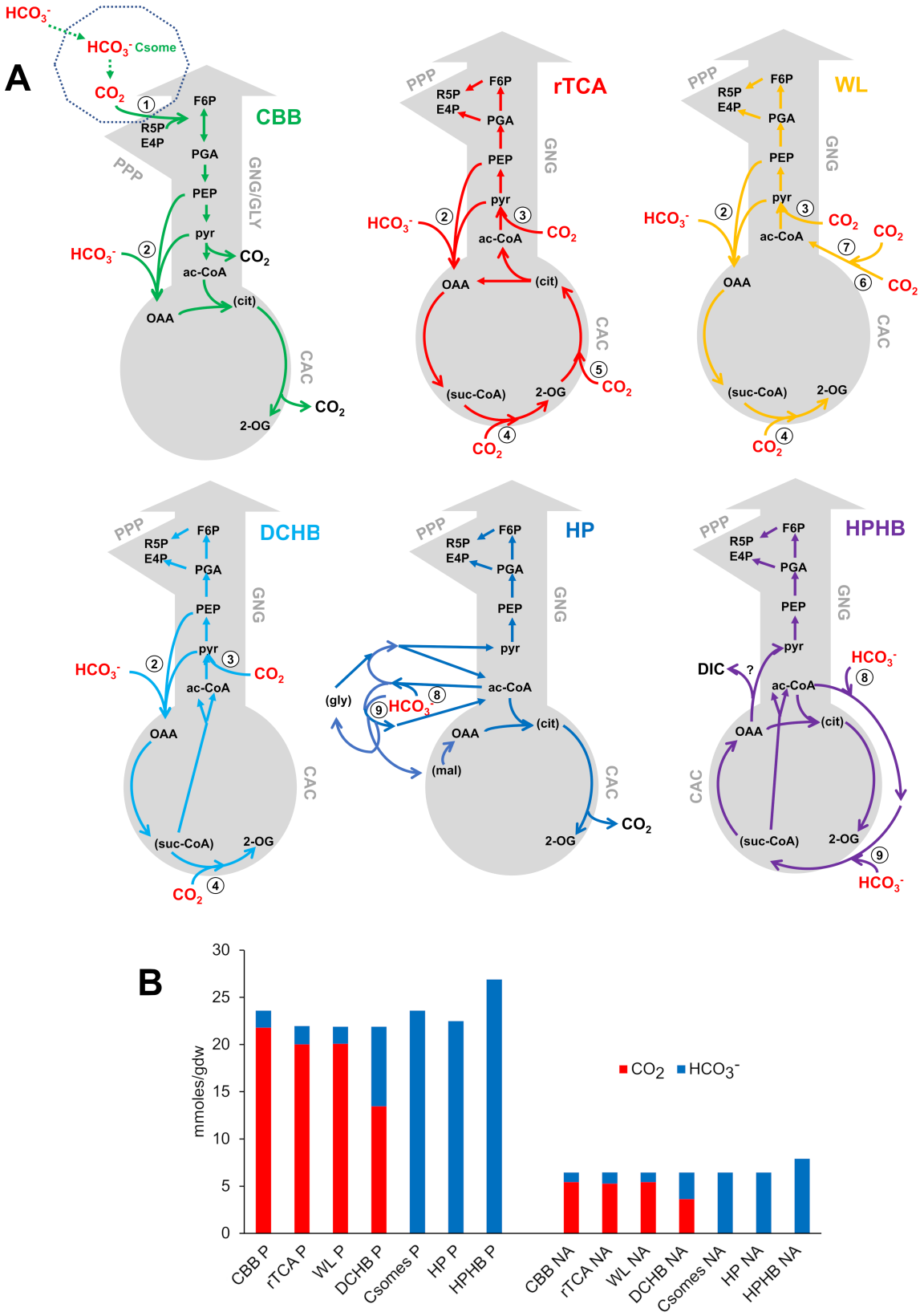

FIG 6 CO_2 and HCO_3^- consumption by organisms using different autotrophic DIC fixation pathways. (A) Overview of steps from autotrophic DIC fixation to the synthesis of metabolic intermediates necessary for protein and nucleotide biosyntheses. Some metabolic intermediates that are not themselves utilized for biosynthesis have been added for clarity and are enclosed in parentheses. Forms of DIC fixed by enzymes catalyzing autotrophic DIC fixation are from the

(Continued on next page)

FIG 6 (Continued)

references cited in Table 1. For CBB, rTCA, WL, and DCHB, arrows from both PEP and pyr reflect the variable distributions of phosphoenolpyruvate carboxylase and pyruvate carboxylase among *Bacteria* and *Archaea* (23, 113, 114). (B) Amounts of CO_2 and HCO_3^- necessary to synthesize protein (P) or nucleic acids (NA) for 1-gram dry weight of biomass of a generic cell using different DIC fixation pathways. An approach similar to (115) was used to calculate the contribution of CO_2 and HCO_3^- to the synthesis of macromolecules in autotrophic organisms using different autotrophic DIC fixation pathways (Supplemental Material). The mmoles of DIC consumed are greater for organisms using the CBB, Csomes, HP, and HPHB pathways due to losses during synthesis of metabolic intermediates (A) 1, RubisCO; 2, phosphoenolpyruvate carboxylase or pyruvate carboxylase; 3, pyruvate synthase; 4, 2-oxoglutarate synthase; 5, isocitrate dehydrogenase; 6, formate dehydrogenase (*Bacteria*) or formylmethanofuran dehydrogenase (*Archaea*); 7, carbon monoxide dehydrogenase/acetyl-CoA synthase; 8, 9, acetyl-CoA/propionyl-CoA carboxylase; ac-CoA, acetyl-coenzyme A; CAC, citric acid cycle; CBB, Calvin-Benson-Bassham cycle; cit, citrate; Csomes, carboxysome; DCHB, dicarboxylate-hydroxybutyrate cycle; DIC, dissolved inorganic carbon; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; GNG/GLY, gluconeogenesis/glycolysis; gly, glyoxylate; HP, hydroxypropionate bicycle; HPHB, hydroxypropionate-hydroxybutyrate cycle; mal, malate; OAA, oxaloacetate; 2-OG, 2-oxoglutarate; PEP, phosphoenolpyruvate; PGA, 3-phosphoglyceraldehyde; PPP, pentose phosphate pathway; pyr, pyruvate; R5P, ribose 5-phosphate; rTCA, reductive citric acid cycle; suc-CoA, succinyl-coenzyme A; WL, Wood-Ljungdahl pathway.

for this inhibition by increasing the concentration of cytoplasmic CO_2, which could mitigate competitive inhibition by N_2O.

Organisms using the CBB are generously endowed with genes encoding CA, and many of these are predicted to be extracellular, though almost all of these organisms are predicted to have cytoplasmic CA (Fig. 7B through D). Some of these organisms are intracellular chemolithoautotrophic symbionts of bivalves (*Ruthia magnifica* and *endofausta*, *Vesicomyosocius okutanii*, *Bathymodiolus septemdierum* and *thermophilus* symbionts) and lack DIC transporter genes. Their lack of DIC transporters and reliance on cCA for bicarbonate could be an adaptation to living in the high CO_2 habitat within actively metabolizing eukaryotic cells. For those organisms using the CBB that have DIC transporter genes, the majority also have eCA genes (Fig. 8). If coexpressed, the eCA could facilitate transporter activity as described above.

Organisms using the rTCA have DIC transporter and CA gene frequencies similar to those using CBB (Fig. 7 and 8); perhaps, this reflects their similarities in demand for CO_2 and HCO_3^- (Fig. 6). Interestingly, there does appear to be a bimodal distribution of DIC toolkit genes among organisms using the rTCA. Most of the organisms from phylum *Chlorobiota* encode a single carbonic anhydrase and no DIC transporters, while those from phyla *Campylobacterota*, "*Nitrospirae*," and *Aquificota* typically encode multiple CAs, at least one DIC transporter, or both (Fig. 3). This bimodal distribution of DIC toolkit genes may suggest that some organisms using the rTCA are adapted to lower CO_2 habitats (*Campylobacterota*, "*Nitrospirae*," and *Aquificota*) and others to higher (*Chlorobiota*), analogous to low CO_2-adapted (with carboxysomes) and higher CO_2-adapted (without carboxysomes) organisms using the CBB. Adaptation to low vs. high CO_2 habitats in organisms using the rTCA is also supported by their predicted mechanism for aminoimidazole ribonucleotide (AIR) carboxylation in purine biosynthesis. In most *Bacteria* and *Archaea*, two enzymes [5-(carboxyamino)imidazole ribonucleotide synthase, EC 6.3.4.18, encoded by *purK*, and N (5)-carboxyaminoimidazole ribonucleotide mutase, EC 5.4.99.18, encoded by *purE*] act together to carboxylate AIR (120)) PurK uses HCO_3^- as a substrate and passes it to PurE (121). When DIC concentrations are very high, PurE can carboxylate AIR in the absence of PurK, using CO_2 (122). Consistent with this observation, *Cyanobacteria* with mutations in *purK* require high CO_2 concentrations for growth (123). Many members of *Chlorobiota* only encode PurE; it is possible that AIR carboxylation is via CO_2 in these members of *Chlorobiota* (124), and as a result, their growth may require high CO_2 concentrations. Similar to PurK assisting PurE, a biotin carboxylase is present some members of *Aquificota* that assists isocitrate dehydrogenase by catalyzing the carboxylation of 2-oxoglutarate via HCO_3^-. In its absence, isocitrate dehydrogenase uses CO_2 (Table 1) (12). The biotin carboxylase could facilitate growth under low CO_2 conditions by diminishing the demand for intracellular CO_2.

Archaea and *Bacteria* using the WL pathway completely lack DIC transporters and have fewer genes encoding CA than CBB or rTCA organisms, though the cCA gene presence and number are similar to rTCA (Fig. 7). It is possible that organisms using

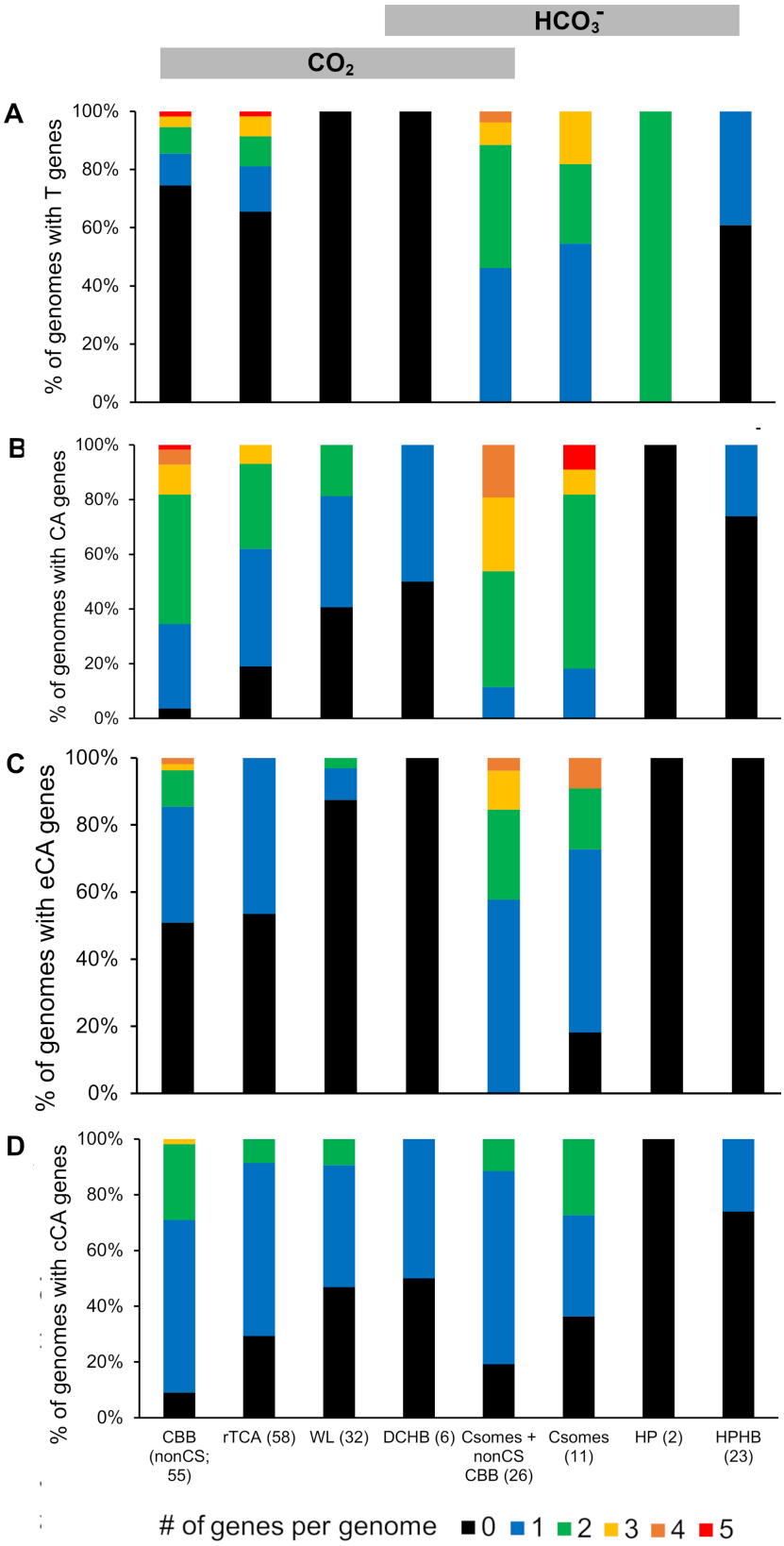

FIG 7 Number of genes encoding DIC transporters or carbonic anhydrase per genome in organisms using different autotrophic DIC fixation pathways. Pathways are positioned according to Fig. 6, with those relying predominantly on CO_2 toward the left, and those relying predominantly on HCO_3^-

(Continued on next page)

FIG 7 (Continued)

toward the right. Pathways are listed with the number (N) of genomes representing them. (A) Number of DIC transporter genes (T) per genome; (B) number of CA genes per genome (includes *csoSCA*); (C) number of extracellular CA genes (eCA) per genome; (D) number of cytoplasmic CA genes (cCA) per genome (excludes *csoSCA*). For both C and D, the CA location was predicted by SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) (105). CBB (nonCS), genomes encoding the CBB cycle, with only noncarboxysomal form I or form II RubisCO; rTCA, reductive citric acid cycle; WL, Wood-Ljungdahl pathway; DCHB, dicarboxylate-hydroxybutyrate cycle; Csomes + nonCS CBB, genomes encoding carboxysomes as well as noncarboxysomal form I and/or form II RubisCO; Csomes, genomes encoding carboxysomes, lacking noncarboxysomal RubisCO; HP, hydroxypropionate bicycle; HPHB, hydroxypropionate-hydroxybutyrate cycle.

the WL pathway are adapted to particularly high CO_2 habitats, which is also consistent with the majority of them having a purine biosynthetic pathway that requires high CO_2 (*purE*; see above). Indeed, the acetogens included in this group do require elevated CO_2 for growth (94). However, the absence of known DIC transporters does not rule out novel DIC transporters; it has been suggested that these organisms may use a yet-to-be characterized acetate-HCO_3^- antiporter (94).

DISTRIBUTION OF DIC TOOLKIT GENES IN AUTOTROPHIC ORGANISMS RELYING PRIMARILY ON HCO_3^-

In general, genes encoding DIC transporters are particularly abundant among autotrophs with DIC fixation pathways relying primarily on HCO_3^- (Fig. 7A), though HPHB organisms have fewer DIC transporter genes than the other HCO_3^--dependent autotrophs. However, among fellow members of *Archaea*, HPHB organisms have more DIC transporter genes than CO_2-dependent WL and DCHB organisms do (Fig. 7). The relative abundance of DIC transporters among autotrophs that predominantly fix HCO_3^- is particularly sensible, since the HCO_3^- that the transporters deposit in the cytoplasm could be used directly for HCO_3^- fixation. The abundance of cCA is broadly similar to CO_2-dependent autotrophs, though the HP and HPHB autotrophs have fewer than the others (Fig. 7D). The relative scarcity of cCA genes in HP and HPHB organisms would diminish loss of the cytoplasmic HCO_3^- pool that their transporters had delivered; the presence of cytoplasmic CA would convert a portion of this pool to CO_2, which could be lost by diffusion through the membrane (3).

Organisms with carboxysomal loci are very generously equipped with both DIC transporter genes and CA (Fig. 7). This observation is consistent with the model of CCM function constructed for *Cyanobacteria*, as described above. Many of these abundant CA genes are predicted to encode extracellular enzymes (Fig. 7C), which may function to supply HCO_3^- or CO_2 to DIC transporters. This would be particularly helpful if transporter activity is high enough to bring HCO_3^- or CO_2 concentrations below those present at equilibrium. The number of cCA genes (Fig. 7D) is similar to other organisms, which is a bit alarming, since carboxysomal carbonic anhydrase CsoSCA was not included in this tally, and cCA presents a risk to these cells by facilitating cytoplasmic HCO_3^- leakage by converting it to CO_2. Perhaps, some of these cCAs have been incorrectly assigned to the cytoplasm by SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/)(105).

The two HP organisms have genes encoding both DAC and SbtA DIC transporters and an absence of CA genes of any sort, which is quite interesting since the majority of the other organisms do have CA genes. An absence of CA, and presence of DIC transporter genes, is completely consistent with HCO_3^- use by the HP cycle (Fig. 6A). Given the small sample size (two finished genomes), it is not possible to know if this is typical for organisms using this pathway for autotrophic DIC fixation.

Genomes from organisms using the HPHB are the only members of *Archaea* in this study to have genes encoding DIC transporters (SbtA; Fig. 4 and 7). CA gene abundance is similar to other members of *Archaea* (Fig. 7B) and is predicted to be cytoplasmic (Fig. 7D). HPHB organisms with DIC transporters tend not to have cCA and vice-versa, though

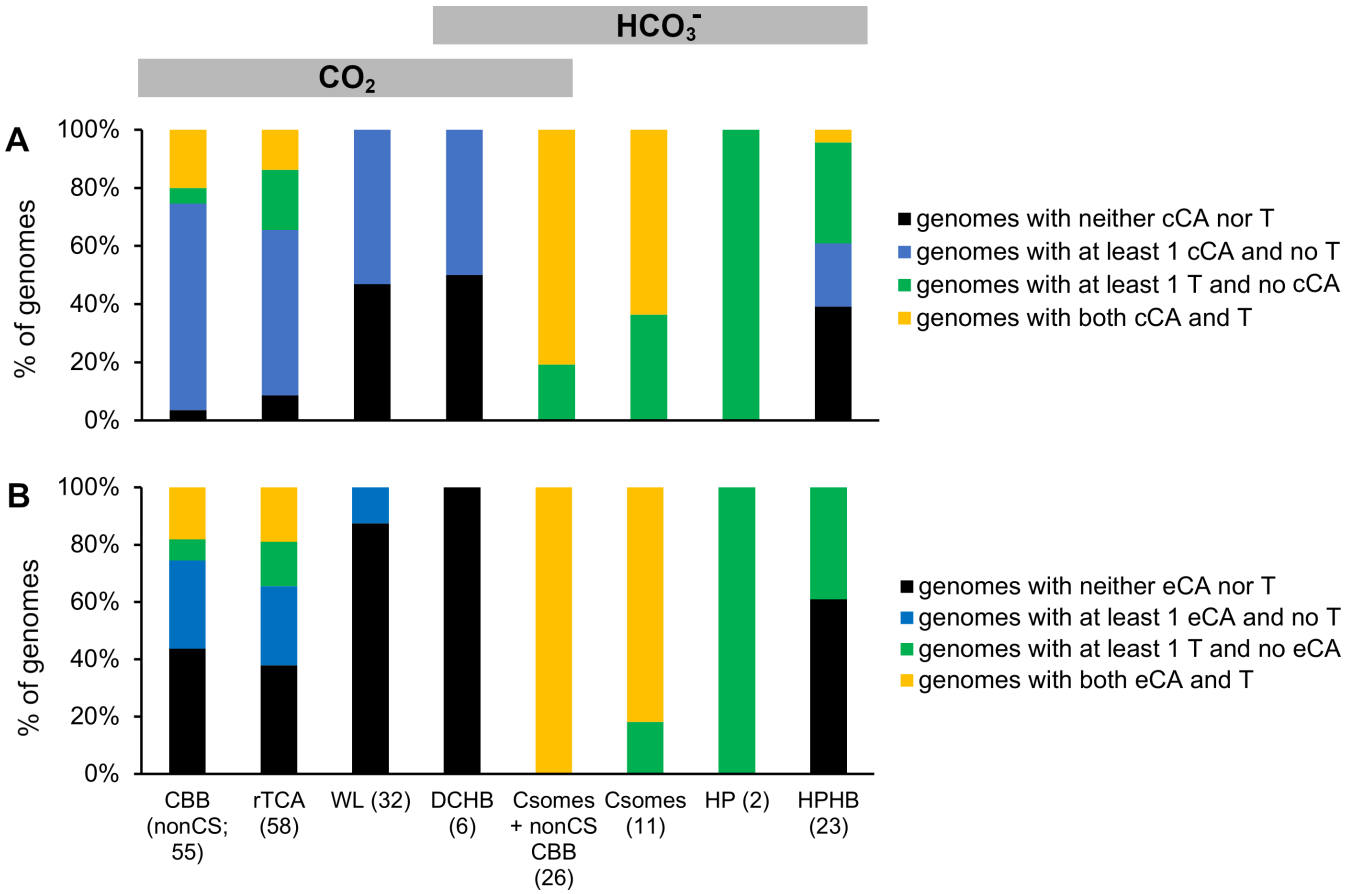

FIG 8 Coexistence of potential DIC transporter genes (T) and (A) cytoplasmic CA genes or (B) extracellular CA genes within genomes from organisms with different autotrophic DIC fixation pathways. Numbers in parentheses are the numbers of genomes in each category. The CA location was predicted by SignalP 6.0 (https://services.healthtech.dtu.dk/services/SignalP-6.0/) (105). eCA, extracellular carbonic anhydrase; cCA, cytoplasmic carbonic anhydrase; CBB (nonCS), genomes encoding the CBB cycle, with only noncarboxysomal form I or form II RubisCO; rTCA, reductive citric acid cycle; WL, Wood-Ljungdahl pathway; DCHB, dicarboxylate-hydroxybutyrate cycle; Csomes + nonCS CBB, genomes encoding carboxysomes as well as noncarboxysomal form I and/or form II RubisCO; Csomes, genomes encoding carboxysomes, lacking noncarboxysomal RubisCO; HP, hydroxypropionate bicycle; HPHB, hydroxypropionate-hydroxybutyrate cycle.

one organism does have both (Fig. 8A). This pattern of one-or-the-other (cCA vs. DIC transporter) for HCO_3^- supply is similar to what has been observed in *Firmicutes* (125), which minimizes leakage losses expected if both are highly expressed. Additionally, in this case, there appears to be an environmental component. Eight of nine of the HPHB organisms that have DIC transporter genes grow optimally at circumneutral pH, while four of five that have cCA are acidophilic (Fig. 4). SbtA transport requires HCO_3^-, which is not present at acidic pH, so *sbtA* gene absence from most of the acidophiles makes sense. Likewise, reliance on a cCA for cytoplasmic HCO_3^- in turn relies on diffusion of CO_2 from the environment, which is a better strategy in acidic environments than circumneutral ones, where the proportion of DIC as CO_2 is lower. The presence of both a DIC transporter and cCA gene in *Nitrosopelagicus brevis* is curious, as it is for the other organisms using other pathways.

DISTRIBUTION OF DIC TOOLKIT GENES IN AUTOTROPHIC ORGANISMS RELYING ON CO_2 AND HCO_3^-

Organisms using the DCHB, which requires nearly equal amounts of CO_2 and HCO_3^- simultaneously (Fig. 6), do not have genes encoding known DIC transporters, and half have genes encoding cCA (Fig. 7). Those with cCA belong to genus *Pyrobaculum*. Though genome data suggest these organisms use DCHB (107), biochemical data are

less conclusive, suggesting the rTCA could operate in these organisms (108). In this case, the presence of a different toolkit could reflect the use of a different pathway. The relative paucity of DIC toolkit genes may reflect the comparatively understudied nature of *Archaea*. If this paucity indeed reflects the actual abundance of DIC toolkit genes in these organisms, the ones with cCA are relying on diffusion of CO_2 from the environment for cytoplasmic CO_2 and (cCA-mediated) HCO_3^- supply. The organisms lacking both cCA and DIC transporters raise another possibility. Most of these organisms are hyperthermophiles (five of six); the remaining one is a thermophile (Fig. 4). All were isolated from hot springs (Table S1). Given that membrane permeability to CO_2 (126) and chemical (non-CA) DIC interconversion rates (2) both increase with temperature, perhaps, a DIC toolkit is less necessary for these organisms. However, it is important to note that thermophiles and hyperthermophiles using other pathways do have DIC toolkit components, including DIC transporters (e.g., members of phylum *Aquificota;* Fig. 3). The presence of DIC transporter genes in thermophilic and hyperthermophilic *Bacteria* suggests that transporters could be helpful for thermophilic and hyperthermophilic *Archaea*, especially since their cell membrane permeabilities have been found to be less sensitive to temperature than those present from hyperthermophilic *Bacteria* (127). Taken together, these observation suggest that *Archaea* using the DCHB pathway are likely to have novel DIC transporters.

Organisms whose genomes include both a carboxysome locus as well as noncarboxysomal RubisCO genes use both CO_2 and HCO_3^-, but unlike DCHB organisms, their use of these forms of DIC is not simultaneous but differentially regulated. Under low CO_2 conditions, they rely predominantly on HCO_3^- by upregulating carboxysome expression and repressing cytoplasmic RubisCO expression; under high CO_2 conditions, they rely predominantly on CO_2 by upregulating cytoplasmic RubisCO expression and repressing carboxysome expression (45, 86, 87). Since they alternate between carboxysomal and noncarboxysomal CBB use, their complement of DIC toolkit genes resembles a combination of both (Fig. 7), with high numbers of DIC transporter genes (similar to organisms that solely encode carboxysomal RubisCO), high numbers of CA genes (similar to both carboxysomal and noncarboxysomal CBB use), and an abundance of eCA genes.

FURTHER QUESTIONS

The analysis of DIC toolkit components encoded in the genomes of a variety of autotrophic organisms has raised some points of interest for autotrophs in general, as well as points specific to each pathway. One important unknown is the identity and prevalence of yet-to-be-described DIC transporters and CA. The latest additions to the lists of known DIC toolkit components and autotrophic pathways have been relatively recent [newest DIC transporter: 2017 (70); newest CA: 2019 (65); and newest autotrophic DIC fixation pathway: 2020 (39)], suggesting that there is much that remains to be uncovered. Undersampling issues are also apparent: comparatively few members of *Archaea* have been sequenced and studied, only two HP autotrophs have been completely sequenced, and organisms thriving at pH extremes and low temperatures are undersequenced (Fig. 3 to 5). Additionally, the interesting possibilities raised by genome data should be confirmed by measurements of gene expression and function under different growth conditions.

The presence of genes encoding both DIC transporters and cCA in organisms using carboxysomal and non-carboxysomal CBB, rTCA, and HPHB is also curious, given that their high-level coexpression in other organisms provides no growth advantage (125) or causes loss of growth under low CO_2 conditions (82). Differential expression and modulated expression (116) to minimize leakage are possible, as is a novel form of spatial segregation analogous to transporters and carboxysomes in organisms with CCMs.

The DIC toolkit is especially open for study among organisms using non-CBB pathways for DIC fixation, and the presence of toolkit genes in these organisms raises the possibility of studies of their function and expression. Are the rTCA organisms indeed taxonomically bimodal with respect to their adaptation to growth under low CO_2

conditions? Does the presence of DIC toolkit genes in rTCA organisms beyond phylum *Chlorobiota* enable them to grow better under low CO_2 conditions? Are there parallels in DIC transporter and CA expression with CBB organisms? Are WL organisms specifically adapted to high CO_2 conditions, or do some of them have yet-to-be-described transporter activities [e.g., acetate:HCO_3^- antiporters (94)] that could facilitate growth under low CO_2 conditions? Given that non-CBB autotrophs include many thermophiles and hyperthermophiles, they provide an opportunity to study the degree to which high temperatures influence the activity and necessity of DIC toolkit capabilities. The addition of more finished genomes from psychrophilic autotrophic organisms could extend these inferences as well (currently, only six are available).

The results of the *in silico* analyses presented here, as well as experimental studies of organisms with CCMs, strongly suggest that DIC toolkit genes could boost the performance of engineered autotrophic organisms in industrial contexts. If these engineered organisms are to be cultivated with air as the source of CO_2, DIC toolkit genes may be required for growth, as they are in *Cyanobacteria* and *Proteobacteria* with CCMs (70, 80, 128). The prevalence of DIC toolkit genes in autotrophic *Archaea* and *Bacteria* from habitats ranging from pH 1 to pH 11 (Fig. 3 to 5) using all six autotrophic pathways (Fig. 7 and 8) suggests that these genes provide a selective advantage to the organisms that carry them, which may translate into enhanced biomass in an industrial context. Given that organisms with disrupted DIC toolkit genes can be rescued when provided with extremely high CO_2 concentrations [1%–5% headspace CO_2, vol/vol (70, 80, 128)], it is possible that organisms in industrial environments with high CO_2 concentrations will not require DIC toolkit genes. However, many organisms that have been isolated from high CO_2 environments, as detailed above, have an elaborate collection of DIC toolkit genes, suggesting their utility even in these environments. The technologies for engineering microorganisms have only been available for 50 years (129). Given the fact that microorganisms have been evolving for 3.4–4.2 billion years (130–132), it seems that our attempts to engineer them are best informed by learning from existing organisms from multiple phyla and domains.

ACKNOWLEDGMENTS

We are grateful for the support from the National Science Foundation (NSF-MCB-1952676 to K.M.S.). We are also grateful to Gordon Cannon and Sabine Heinhorst for the discussion and comments on an earlier version of this manuscript and also to the anonymous reviewers for their suggestions, which significantly improved the manuscript.

AUTHOR AFFILIATION

[1]Integrative Biology Department, University of South Florida, Tampa, Florida, USA

AUTHOR ORCIDs

Kathleen M. Scott http://orcid.org/0000-0002-9407-518X

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	NSF-MCB-1952676	Kathleen M. Scott

AUTHOR CONTRIBUTIONS

Kathleen M. Scott, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review and editing | Ren R. Payne, Investigation, Writing

– review and editing | Arin Gahramanova, Investigation, software, Writing – review and editing

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental figures, tables, and details (AEM01557-23-s0001.docx). Supplemental material includes Tables S1-S3, Fig. S1, and details about gathering genomes, verifying predicted functions of genes, and calculating demand for CO2 and bicarbonate.

REFERENCES

1. Erb TJ. 2011. Carboxylases in natural and synthetic microbial pathways. Appl Environ Microbiol 77:8466–8477. https://doi.org/10.1128/AEM.05702-11
2. Zeebe RE, Wolf-Gladrow D. 2003. CO_2 in seawater: equilibrium, kinetics, isotopes. Elsevier, New York.
3. Raven JA, Beardall J. 2016. The ins and outs of CO_2. J Exp Bot 67:1–13. https://doi.org/10.1093/jxb/erv451
4. Gutknecht J, Bisson MA, Tosteson FC. 1977. Diffusion of carbon dioxide through lipid bilayer membranes: effects of carbonic anhydrase, bicarbonate, and unstirred layers. J Gen Physiol 69:779–794. https://doi.org/10.1085/jgp.69.6.779
5. Ding X, Matsumoto T, Gena P, Liu C, Pellegrini-Calace M, Zhong S, Sun X, Zhu Y, Katsuhara M, Iwasaki I, Kitagawa Y, Calamita G. 2013. Water and CO_2 permeability of SsAqpZ, the cyanobacterium *Synechococcus* sp. PCC7942 aquaporin. Biol Cell 105:118–128. https://doi.org/10.1111/boc.201200057
6. Matsuda Y, Nawaly H, Yoneda K. 2022. Carbonic anhydrase, p 167–195. In Blue planet, red and green photosynthesis
7. Price GD. 2011. Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. Photosynth Res 109:47–57. https://doi.org/10.1007/s11120-010-9608-y
8. Scott KM, Harmer TL, Gemmell BJ, Kramer AM, Sutter M, Kerfeld CA, Barber KS, Bari S, Boling JW, Campbell CP, Gallard-Gongora JF, Jackson JK, Lobos A, Mounger JM, Radulovic PW, Sanson JM, Schmid S, Takieddine C, Warlick KF, Whittaker R. 2020. Ubiquity and functional uniformity in CO_2 concentrating mechanisms in multiple phyla of bacteria is suggested by a diversity and prevalence of genes encoding candidate dissolved inorganic carbon transporters. FEMS Microbiol Lett 367:fnaa106. https://doi.org/10.1093/femsle/fnaa106
9. Häusler RE, Holtum JAM, Latzko E. 1987. CO2 is the inorganic carbon substrate of NADP malic enzymes from zea mays and from wheat germ. Eur J Biochem 163:619–626. https://doi.org/10.1111/j.1432-1033.1987.tb10911.x
10. Chang G-G, Tong L. 2003. Structure and function of malic enzymes, a new class of oxidative decarboxylases. Biochemistry 42:12721–12733. https://doi.org/10.1021/bi035251+
11. Dalziel K, Londesborough JC. 1968. The mechanisms of reductive carboxylation reactions. carbon dioxide or bicarbonate as substrate of nicotinamide-adenine dinucleotide phosphate-linked isocitrate dehydrogenase and malic enzyme. Biochem J 110:223–230. https://doi.org/10.1042/bj1100223
12. Aoshima M, Igarashi Y. 2006. A novel oxalosuccinate-forming enzyme involved in the reductive carboxylation of 2-oxoglutarate in *Hydrogenobacter thermophilus* TK-6. Mol Microbiol 62:748–759. https://doi.org/10.1111/j.1365-2958.2006.05399.x
13. Malandrinos G, Louloudi M, Hadjiliadis N. 2006. Thiamine models and perspectives on the mechanism of action of thiamine-dependent enzymes. Chem Soc Rev 35:684–692. https://doi.org/10.1039/b514511m
14. Ragsdale SW, Kumar M. 1996. Nickel-containing carbon monoxide dehydrogenase/acetyl-CoA synthase. Chem Rev 96:2515–2540. https://doi.org/10.1021/cr950058+
15. Vorholt JA, Thauer RK. 1997. The active species of 'CO2' utilized by formylmethanofuran dehydrogenase from methanogenic *Archaea*. Eur J Biochem 248:919–924. https://doi.org/10.1111/j.1432-1033.1997.00919.x
16. Thauer RK, Käufer B, Fuchs G. 1975. Active species of CO_2 utilized by reduced ferredoxin - CO_2 oxidoreductase from *Clostridium pasteurianum*. Eur J Biochem 55:111–117. https://doi.org/10.1111/j.1432-1033.1975.tb02143.x
17. Maruyama H, Easterday RL, Chang H-C, Lane MD. 1966. The enzymatic carboxylation of phosphoenolpyruvate: I. purification and properties of phosphoenolpyruvate carboxylase. J Biol Chem 241:2405–2412. https://doi.org/10.1016/S0021-9258(18)96634-2
18. Cooper TG, Wood HG. 1971. The carboxylation of phosphoenolpyruvate and pyruvate. II. The active species of "CO_2" utilized by phosphoenlpyruvate carboxylase and pyruvate carboxylase. J Biol Chem 246:5488–5490. https://doi.org/10.1016/S0021-9258(18)61931-3
19. O'Leary MH, Hermes JD. 1987. Determination of substrate-specificity of carboxylases by nuclear-magnetic-resonance. Anal Biochem 162:358–362. https://doi.org/10.1016/0003-2697(87)90404-0
20. Cooper TG, Tchen TT, Wood HG, Benedict CR. 1968. The carboxylation of phosphoenolpyruvate and pyruvate: I. The active species of "CO_2" utilized by phosphoenolpyruvate carboxykinase, carboxytransphosphorylase, and pyruvate carboxylase. J Biol Chem 243:3857–3863. https://doi.org/10.1016/S0021-9258(18)92022-3
21. Cooper TG, Filmer D, Wishnick M, Lane MD. 1969. The active species of "CO_2" utilized by ribulose diphosphate carboxylase. J Biol Chem 244:1081–1083. https://doi.org/10.1016/S0021-9258(18)91899-5
22. Kaziro Y, Hass LF, Boyer PD, Ochoa S. 1962. Mechanism of the propionyl carboxylase reaction:II. isotopic exchange and tracer experiments. J Biol Chem 237:1460–1468.
23. Sauer U, Eikmanns BJ. 2005. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. FEMS Microbiol Rev 29:765–794. https://doi.org/10.1016/j.femsre.2004.11.002
24. Dimroth P, Hilpert W. 1984. Carboxylation of pyruvate and acetyl coenzyme A by reversal of the sodium pumps oxaloacetate decarboxylase and methylmalonyl-CoA decarboxylase. Biochemistry 23:5360–5366. https://doi.org/10.1021/bi00317a039
25. Reiskind JB, Bowes G. 1991. The role of phosphoenolpyruvate carboxykinase in a marine macroalga with C4-like photosynthetic characteristics. Proc Natl Acad Sci U S A 88:2883–2887. https://doi.org/10.1073/pnas.88.7.2883
26. Dimroth P. 1982. The role of biotin and sodium in the decarboxylation of oxaloacetate by the membrane-bound oxaloacetate decarboxylase from *Klebsiella aerogenes*. Eur J Biochem 121:435–441. https://doi.org/10.1111/j.1432-1033.1982.tb05806.x
27. Chou C-Y, Yu LPC, Tong L. 2009. Crystal structure of biotin carboxylase in complex with substrates and implications for its catalytic mechanism. J Biol Chem 284:11690–11697. https://doi.org/10.1074/jbc.M805783200
28. Waldrop GL, Holden HM, St Maurice M. 2012. The enzymes of biotin dependent CO_2 metabolism: what structures reveal about their reaction mechanisms. Protein Sci 21:1597–1619. https://doi.org/10.1002/pro.2156

29. Berg IA. 2011. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. Appl Environ Microbiol 77:1925–1936. https://doi.org/10.1128/AEM.02473-10
30. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE, Fuchs G. 2010. Autotrophic carbon fixation in *Archaea*. Nat Rev Microbiol 8:447–460. https://doi.org/10.1038/nrmicro2365
31. Bassham JA, Benson AA, Kay LD, Harris AZ, Wilson AT, Calvin M. 1954. The path of carbon in photosynthesis.21. the cyclic regeneration of carbon dioxide acceptor. J Am Chem Soc 76:1760–1770. https://doi.org/10.1021/ja01636a012
32. Evans MC, Buchanan BB, Arnon DI. 1966. A new ferredoxin-dependent carbon reduction cycle in A photosynthetic bacterium. Proc Natl Acad Sci U S A 55:928–934. https://doi.org/10.1073/pnas.55.4.928
33. Ragsdale SW, Pierce E. 2008. Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. Biochim Biophys Acta 1784:1873–1898. https://doi.org/10.1016/j.bbapap.2008.08.012
34. Huber H, Gallenberger M, Jahn U, Eylert E, Berg IA, Kockelkorn D, Eisenreich W, Fuchs G. 2008. A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic archaeum *Ignicoccus hospitalis*. Proc Natl Acad Sci U S A 105:7851–7856. https://doi.org/10.1073/pnas.0801043105
35. Berg IA, Kockelkorn D, Buckel W, Fuchs G. 2007. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in *Archaea*. Science 318:1782–1786. https://doi.org/10.1126/science.1149976
36. Zarzycki J, Brecht V, Müller M, Fuchs G. 2009. Identifying the missing steps of the autotrophic 3-hydroxypropionate CO_2 fixation cycle in *Chloroflexus aurantiacus*. Proc Natl Acad Sci U S A 106:21317–21322. https://doi.org/10.1073/pnas.0908356106
37. Mall A, Sobotta J, Huber C, Tschirner C, Kowarschik S, Bačnik K, Mergelsberg M, Boll M, Hügler M, Eisenreich W, Berg IA. 2018. Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. Science 359:563–567. https://doi.org/10.1126/science.aao2410
38. Nunoura T, Chikaraishi Y, Izaki R, Suwa T, Sato T, Harada T, Mori K, Kato Y, Miyazaki M, Shimamura S, Yanagawa K, Shuto A, Ohkouchi N, Fujita N, Takaki Y, Atomi H, Takai K. 2018. A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. Science 359:559–563. https://doi.org/10.1126/science.aao3407
39. Sánchez-Andrea I, Guedes IA, Hornung B, Boeren S, Lawson CE, Sousa DZ, Bar-Even A, Claassens NJ, Stams AJM. 2020. The reductive glycine pathway allows autotrophic growth of *Desulfovibrio desulfuricans*. Nat Commun 11:5090. https://doi.org/10.1038/s41467-020-18906-7
40. Fullerton KM, Schrenk MO, Yücel M, Manini E, Basili M, Rogers TJ, Fattorini D, Di Carlo M, d'Errico G, Regoli F, Nakagawa M, Vetriani C, Smedile F, Ramírez C, Miller H, Morrison SM, Buongiorno J, Jessen GL, Steen AD, Martínez M, de Moor JM, Barry PH, Giovannelli D, Lloyd KG. 2021. Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. Nat Geosci 14:301–306. https://doi.org/10.1038/s41561-021-00725-0
41. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, Ramírez CJ, Barry PH, Yücel M, Selci M, Cordone A, Giovannelli D, Lloyd KG. 2023. Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME J 17:140–150. https://doi.org/10.1038/s41396-022-01331-7
42. Moser DP, Gihring TM, Brockman FJ, Fredrickson JK, Balkwill DL, Dollhopf ME, Lollar BS, Pratt LM, Boice E, Southam G, Wanger G, Baker BJ, Pfiffner SM, Lin LH, Onstott TC. 2005. *Desulfotomaculum* and *Methanobacterium* spp. dominate a 4- to 5-kilometer-deep fault. Appl Environ Microbiol 71:8773–8783. https://doi.org/10.1128/AEM.71.12.8773-8783.2005
43. Kelley DS, Karson JA, Früh-Green GL, Yoerger DR, Shank TM, Butterfield DA, Hayes JM, Schrenk MO, Olson EJ, Proskurowski G, et al. 2005. A serpentinite-hosted ecosystem: the lost city hydrothermal field. Science 307:1428–1434. https://doi.org/10.1126/science.1102556
44. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. 2006. Methane- and sulfur-metabolizing microbial communities dominate the lost city hydrothermal field ecosystem. Appl Environ Microbiol 72:6257–6270. https://doi.org/10.1128/AEM.00574-06
45. Scott KM, Leonard JM, Boden R, Chaput CD, Dennison C, Haller E, Harmer TL, Anderson A, Arnold T, Budenstein S, et al. 2019. Diversity in CO_2-concentrating mechanisms among chemolithoautotrophs from the genera *Hydrogenovibrio, Thiomicrorhabdus,* and *Thiomicrospira,* ubiquitous in sulfidic habitats worldwide. Appl Environ Microbiol 85. https://doi.org/10.1128/AEM.02096-18
46. Claassens NJ. 2017. A warm welcome for alternative CO_2 fixation pathways in microbial biotechnology. Microb Biotechnol 10:31–34. https://doi.org/10.1111/1751-7915.12456
47. Antranikian G, Streit WR. 2022. Microorganisms harbor keys to a circular bioeconomy making them useful tools in fighting plastic pollution and rising CO(2) levels. Extremophiles 26:10. https://doi.org/10.1007/s00792-022-01261-4
48. Badger MR, Price GD, Long BM, Woodger FJ. 2006. The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. J Exp Bot 57:249–265. https://doi.org/10.1093/jxb/eri286
49. Price GD, Badger MR, Woodger FJ, Long BM. 2008. Advances in understanding the cyanobacterial CO_2-concentrating-mechanism (CCM): functional components,Ci transporters, diversity, genetic regulation and prospects for engineering into plants. J Exp Bot 59:1441–1461. https://doi.org/10.1093/jxb/erm112
50. Smith KS, Ferry JG. 2000. Prokaryotic carbonic anhydrases. FEMS Microbiol Rev 24:335–366. https://doi.org/10.1111/j.1574-6976.2000.tb00546.x
51. Soli AL, Byrne RH. 2002. CO_2 system hydration and dehydration kinetics and the equilibrium CO_2/H_2CO_3 ratio in aqueous NaCl solution. Mar Chem 78:65–73. https://doi.org/10.1016/S0304-4203(02)00010-5
52. Sanyal G, Maren TH. 1981. Thermodynamics of carbonic anhydrase catalysis. a comparison between human isoenzymes B and C. J Biol Chem 256:608–612. https://doi.org/10.1016/S0021-9258(19)70016-7
53. Heinhorst S, Williams EB, Cai F, Murin CD, Shively JM, Cannon GC. 2006. Characterization of the carboxysomal carbonic anhydrase CsoCSA from *Halothiobacillus neapolitanus*. J Bacteriol 188:8087–8094. https://doi.org/10.1128/JB.00990-06
54. Badger MR, Price GD. 1994. The role of carbonic anhydrase in photosynthesis. Annu Rev Plant Physiol Plant Mol Biol 45:369–392. https://doi.org/10.1146/annurev.pp.45.060194.002101
55. Smith KS, Jakubzick C, Whittam TS, Ferry JG. 1999. Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. Proc Natl Acad Sci U S A 96:15184–15189. https://doi.org/10.1073/pnas.96.26.15184
56. Keilin D, Mann T. 1939. Carbonic anhydrase. Nature 144:442–443. https://doi.org/10.1038/144442b0
57. Guilloton MB, Korte JJ, Lamblin AF, Fuchs JA, Anderson PM. 1992. Carbonic-anhydrase in *Escherichia coli* - a product of the *cyn* operon. J Biol Chem 267:3731–3734. https://doi.org/10.1016/S0021--9258(19)50586-5
58. Alber BE, Ferry JG. 1994. A carbonic anhydrase from the archaeon *Methanosarcina thermophila*. Proc Natl Acad Sci U S A 91:6909–6913. https://doi.org/10.1073/pnas.91.15.6909
59. Roberts SB, Lane TW, Morel FMM. 1997. Carbonic anhydrase in the marine diatom *Thalassiosira weissflogii* (Bacillariophyceae). J Phycol 33:845–850. https://doi.org/10.1111/j.0022-3646.1997.00845.x
60. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, Yeates TO, Kerfeld CA. 2006. The structure of β-carbonic anhydrase from the carboxysomal shell reveals a distinct subclass with one active site for the price of two. J Biol Chem 281:7546–7555. https://doi.org/10.1074/jbc.M510464200
61. So AK-C, Espie GS, Williams EB, Shively JM, Heinhorst S, Cannon GC. 2004. A novel evolutionary lineage of carbonic anhydrase (ε class) is a component of the carboxysome shell. J Bacteriol 186:623–630. https://doi.org/10.1128/JB.186.3.623-630.2004
62. Xu Y, Feng L, Jeffrey PD, Shi Y, Morel FMM. 2008. Structure and metal exchange in the cadmium carbonic anhydrase of marine diatoms. Nature 452:56–61. https://doi.org/10.1038/nature06636
63. Del Prete S, Vullo D, Fisher GM, Andrews KT, Poulsen SA, Capasso C, Supuran CT. 2014. Discovery of a new family of carbonic anhydrases in the malaria pathogen *Plasmodium falciparum* - the η-carbonic anhydrases. Bioorg Med Chem Lett 24:4389–4396. https://doi.org/10.1016/j.bmcl.2014.08.015
64. Kikutani S, Nakajima K, Nagasato C, Tsuji Y, Miyatake A, Matsuda Y. 2016. Thylakoid luminal θ-carbonic anhydrase critical for growth and photosynthesis in the marine diatom *Phaeodactylum tricornutum*. Proc

Natl Acad Sci U S A 113:9828–9833. https://doi.org/10.1073/pnas.1603112113
65. Jensen EL, Clement R, Kosta A, Maberly SC, Gontero B. 2019. A new widespread subclass of carbonic anhydrase in marine phytoplankton. ISME J 13:2094–2106. https://doi.org/10.1038/s41396-019-0426-8
66. Shibata M, Katoh H, Sonoda M, Ohkawa H, Shimoyama M, Fukuzawa H, Kaplan A, Ogawa T. 2002. Genes essential to sodium-dependent bicarbonate transport in cyanobacteria. J Biol Chem 277:18658–18664. https://doi.org/10.1074/jbc.M112468200
67. Price GD, Woodger FJ, Badger MR, Howitt SM, Tucker L. 2004. Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. Proc Natl Acad Sci U S A 101:18228–18233. https://doi.org/10.1073/pnas.0405211101
68. Omata T, Price GD, Badger MR, Okamura M, Gohta S, Ogawa T. 1999. Identification of an ATP-binding cassette transporter involved in bicarbonate uptake in the cyanobacterium *Synechococcus* sp. strain PCC 7942. Proc Natl Acad Sci U S A 96:13571–13576. https://doi.org/10.1073/pnas.96.23.13571
69. Shibata M, Ohkawa H, Kaneko T, Fukuzawa H, Tabata S, Kaplan A, Ogawa T. 2001. Distinct constitutive and low-CO_2-induced CO_2 uptake systems in cyanobacteria: genes involved and their phylogenetic relationship with homologous genes in other organisms. Proc Natl Acad Sci U S A 98:11789–11794. https://doi.org/10.1073/pnas.191258298
70. Mangiapia M, Brown T-RW, Chaput D, Haller E, Harmer TL, Hashemy Z, Keeley R, Leonard J, Mancera P, Nicholson D, Stevens S, Wanjugi P, Zabinski T, Pan C, Scott KM, USF MCB4404L. 2017. Proteomic and mutant analysis of the CO_2 concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. J Bacteriol 199:e00871-16. https://doi.org/10.1128/JB.00871-16
71. Desmarais JJ, Flamholz AI, Blikstad C, Dugan EJ, Laughlin TG, Oltrogge LM, Chen AW, Wetmore K, Diamond S, Wang JY, Savage DF. 2019. DABs are inorganic carbon pumps found throughout prokaryotic phyla. Nat Microbiol 4:2204–2215. https://doi.org/10.1038/s41564-019-0520-8
72. Fan S-H, Ebner P, Reichert S, Hertlein T, Zabel S, Lankapalli AK, Nieselt K, Ohlsen K, Götz F. 2019. MpsAB is important for *Staphylococcus aureus* virulence and growth at atmospheric CO_2 levels. Nat Commun 10:3627. https://doi.org/10.1038/s41467-019-11547-5
73. Schmid S, Chaput D, Breitbart M, Hines R, Williams S, Gossett HK, Parsi SD, Peterson R, Whittaker RA, Tarver A, Scott KM. 2021. Dissolved inorganic carbon-accumulating complexes from autotrophic bacteria from extreme environments. J Bacteriol 203:e0037721. https://doi.org/10.1128/JB.00377-21
74. Reinhold L, Kosloff R, Kaplan A. 1991. A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. Can J Bot 69:984–988. https://doi.org/10.1139/b91-126
75. Badger MR, Hanson D, Price GD. 2002. Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. Funct Plant Biol 29:161–173. https://doi.org/10.1071/PP01213
76. Badger MR, Price GD. 2003. CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. J Exp Bot 54:609–622. https://doi.org/10.1093/jxb/erg076
77. Burnap RL, Hagemann M, Kaplan A. 2015. Regulation of CO_2 conentrating mechanism in cyanobacteria. Life (Basel) 5:348–371. https://doi.org/10.3390/life5010348
78. Long BM, Rae BD, Rolland V, Förster B, Price GD. 2016. Cyanobacterial CO_2-concentrating mechanism components: function and prospects for plant metabolic engineering. Curr Opin Plant Biol 31:1–8. https://doi.org/10.1016/j.pbi.2016.03.002
79. Scott KM, Leonard J, Boden R, Chaput D, Dennison C, Haller E, Harmer TL, Anderson A, Arnold T, Brand J, et al. 2019. Diversity in CO_2 concentrating mechanisms among chemolithoautotrophs from the genera *Hydrogenovibrio*, *Thiomicrorhabdus*, and *Thiomicrospira*, ubiquitous in sulfidic habitats worldwide. Appl Environ Microbiol 85:e02096-18. https://doi.org/10.1128/AEM.02096-18
80. Dou Z, Heinhorst S, Williams EB, Murin CD, Shively JM, Cannon GC. 2008. CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. J Biol Chem 283:10377–10384. https://doi.org/10.1074/jbc.M709285200
81. Kerfeld CA, Aussignargues C, Zarzycki J, Cai F, Sutter M. 2018. Bacterial microcompartments. Nat Rev Microbiol 16:277–290. https://doi.org/10.1038/nrmicro.2018.10
82. Price GD, Badger MR. 1989. Expression of human carbonic anhydrase in the cyanobacterium *Synechococcus* PCC7942 creates a high CO_2-requiring phenotype. Plant Physiol 91:505–513. https://doi.org/10.1104/pp.91.2.505
83. Soltes-Rak E, Mulligan ME, Coleman JR. 1997. Identification and characterization of a gene encoding a vertebrate-type carbonic anhdyrase in cyanobacteria. J Bacteriol 179:769–774. https://doi.org/10.1128/jb.179.3.769-774.1997
84. Dobrinski KP, Boller AJ, Scott KM. 2010. Expression and function of four carbonic anhydrase homologs in the deep-sea chemolithoautotroph *Thiomicrospira crunogena*. Appl Environ Microbiol 76:3561–3567. https://doi.org/10.1128/AEM.00064-10
85. Cai F, Heinhorst S, Shively JM, Cannon GC. 2008. Transcript analysis of the *Halothiobacillus neapolitanus cso* operon. Arch Microbiol 189:141–150. https://doi.org/10.1007/s00203-007-0305-y
86. Yoshizawa Y, Toyoda K, Arai H, Ishii M, Igarashi Y. 2004. CO_2-responsive expression and gene organization of three ribulose-1,5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in *Hydrogenovibrio marinus* strain MH-110. J Bacteriol 186:5685–5691. https://doi.org/10.1128/JB.186.17.5685-5691.2004
87. Dobrinski KP, Enkemann SA, Yoder SJ, Haller E, Scott KM. 2012. Transcription response of the sulfur chemolithoautotroph *Thiomicrospira crunogena* to dissolved inorganic carbon limitation. J Bacteriol 194:2074–2081. https://doi.org/10.1128/JB.06504-11
88. Tabita FR. 1999. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. Photosyn Res 60:1–28. https://doi.org/10.1023/A:1006211417981
89. Kusian B, Sültemeyer D, Bowien B. 2002. Carbonic anhydrase is essential for growth of *Ralstonia eutropha* at ambient CO_2 concentrations. J Bacteriol 184:5018–5026. https://doi.org/10.1128/JB.184.18.5018-5026.2002
90. Puskás LG, Inui M, Zahn K, Yukawa H. 2000. A periplasmic, α-type carbonic anhydrase from *Rhodopseudomonas palustris* is essential for bicarbonate uptake. Microbiology (Reading) 146 (Pt 11):2957–2966. https://doi.org/10.1099/00221287-146-11-2957
91. Ivanovsky RN, Keppen OI, Lebedeva NV, Gruzdev DS. 2020. Carbonic anhydrase in anoxygenic phototrophic bacteria. Microbiology 89:266–272. https://doi.org/10.1134/S0026261720020058
92. Hagemann M, Kern R, Maurino VG, Hanson DT, Weber APM, Sage RF, Bauwe H. 2016. Evolution of photorespiration from cyanobacteria to land plants, considering protein phylogenies and acquisition of carbon concentrating mechanisms. J Exp Bot 67:2963–2976. https://doi.org/10.1093/jxb/erw063
93. Montgomery BL, Lechno-Yossef S, Kerfeld CA. 2016. Interrelated modules in cyanobacterial photosynthesis: the carbon-concentrating mechanism, photorespiration, and light perception. J Exp Bot 67:2931–2940. https://doi.org/10.1093/jxb/erw162
94. Braus-Stromeyer SA, Schnappauf G, Braus GH, Gössner AS, Drake HL. 1997. Carbonic anhydrase in *Acetobacterium woodii* and other acetogenic bacteria. J Bacteriol 179:7197–7200. https://doi.org/10.1128/jb.179.22.7197-7200.1997
95. Hügler M, Sievert SM. 2011. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Ann Rev Mar Sci 3:261–289. https://doi.org/10.1146/annurev-marine-120709-142712
96. Chen I-M, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D, Varghese NJ, Reddy TBK, Mukherjee S, Ovchinnikova G, Nolan M, Seshadri R, Roux S, Visel A, Woyke T, Eloe-Fadrosh EA, Kyrpides NC, Ivanova NN. 2023. The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res 51:D723–D732. https://doi.org/10.1093/nar/gkac976
97. Axen SD, Erbilgin O, Kerfeld CA, Tanaka MM. 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. PLoS Comput Biol 10:e1003898. https://doi.org/10.1371/journal.pcbi.1003898
98. McDowall JS, Murphy BJ, Haumann M, Palmer T, Armstrong FA, Sargent F. 2014. Bacterial formate hydrogenlyase complex. Proc Natl Acad Sci U S A 111:E3948–E3956. https://doi.org/10.1073/pnas.1407927111

99. Wood GE, Haydock AK, Leigh JA. 2003. Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*. J Bacteriol 185:2548–2554. https://doi.org/10.1128/JB.185.8.2548-2554.2003
100. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, Wimalarathna H, Harrison OB, Sheppard SK, Cody AJ, Maiden MCJ. 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158:1005–1015. https://doi.org/10.1099/mic.0.055459-0
101. Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol Biol Evol 38:3022–3027. https://doi.org/10.1093/molbev/msab120
102. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307–1320. https://doi.org/10.1093/molbev/msn067
103. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, Loy A, Naganuma T, Nakai R, Whitman WB, Hahn MW, Kuever J, Hugenholtz P. 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. https://doi.org/10.1099/ijsem.0.004213
104. Emerson D, Rentz JA, Lilburn TG, Davis RE, Aldrich H, Chan C, Moyer CL, Reysenbach A-L. 2007. A novel lineage of proteobacteria involved in formation of marine Fe-oxidizing microbial mat communities. PLoS ONE 2:e667. https://doi.org/10.1371/journal.pone.0000667
105. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H. 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40:1023–1025. https://doi.org/10.1038/s41587-021-01156-3
106. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340
107. Ramos-Vera WH, Berg IA, Fuchs G. 2009. Autotrophic carbon dioxide assimilation in thermoproteales revisited. J Bacteriol 191:4286–4297. https://doi.org/10.1128/JB.00145-09
108. Hu YJ, Holden JF. 2006. Citric acid cycle in the hyperthermophilic archaeon *Pyrobaculum islandicum* grown autotrophically, heterotrophically, and mixotrophically with acetate. J Bacteriol 188:4350–4355. https://doi.org/10.1128/JB.00138-06
109. Matin A. 1990. Keeping a neutral cytoplasm - the bioenergetics of obligate acidophiles. FEMS Microbiol Lett 75:307–318. https://doi.org/10.1111/j.1574-6968.1990.tb04104.x
110. Koyama N, Nosoh Y. 1985. Effect of potassium and sodium ions on the cytoplasmic pH of an alkalophilic *Bacillus*. Biochimica et Biophysica Acta (BBA) - Biomem 812:206–212. https://doi.org/10.1016/0005-2736(85)90540-1
111. Alper SL, Sharma AK. 2013. The SLC26 gene family of anion transporters and channels. Mol Aspects Med 34:494–515. https://doi.org/10.1016/j.mam.2012.07.009
112. Karinou E, Compton ELR, Morel M, Javelle A. 2013. The *Escherichia coli* SLC26 homologue YchM (DauA) is a C4-dicarboxylic acid transporter. Mol Microbiol 87:623–640. https://doi.org/10.1111/mmi.12120
113. Mukhopadhyay B, Patel VJ, Wolfe RS. 2000. A stable archaeal pyruvate carboxylase from the hyperthermophile *Methanococcus jannaschii*. Arch Microbiol 174:406–414. https://doi.org/10.1007/s002030000225
114. Ettema TJG, Makarova KS, Jellema GL, Gierman HJ, Koonin EV, Huynen MA, de Vos WM, van der Oost J. 2004. Identification and functional verification of archaeal-type phosphoenolpyruvate carboxylase, a missing link in archaeal central carbohydrate metabolism. J Bacteriol 186:7754–7762. https://doi.org/10.1128/JB.186.22.7754-7762.2004
115. Mangiapia M, Scott K. 2016. From CO_2 to cell: energetic expense of creating biomass using the Calvin-Benson-Bassham and reductive citric acid cycles based on genome data. FEMS Microbiol Lett 363:fnw054. https://doi.org/10.1093/femsle/fnw054
116. Flamholz AI, Dugan E, Panich J, Desmarais JJ, Oltrogge LM, Fischer WW, Singer SW, Savage DF. 2022. Trajectories for the evolution of bacterial CO_2-concentrating mechanisms. Proc Natl Acad Sci U S A 119:e2210539119. https://doi.org/10.1073/pnas.2210539119
117. Arp DJ, Stein LY. 2003. Metabolism of inorganic N compounds by ammonia-oxidizing bacteria. Crit Rev Biochem Mol Biol 38:471–495. https://doi.org/10.1080/10409230390267446
118. Daims H, Lücker S, Wagner M. 2016. A new perspective on microbes formerly known as nitrite-oxidizing bacteria. Trends Microbiol. 24:699–712. https://doi.org/10.1016/j.tim.2016.05.004
119. Khalifah RG. 1971. The carbon dioxide hydration activity of carbonic anhydrase. I. Stop-flow kinetic studies on the native human isoenzymes B and C. J Biol Chem 246:2561–2573. https://doi.org/10.1016/S0021-9258(18)62326-9
120. Zhang Y, Morar M, Ealick SE. 2008. Structural biology of the purine biosynthetic pathway. Cell Mol Life Sci 65:3699–3724. https://doi.org/10.1007/s00018-008-8295-8
121. Firestine SM, Misialek S, Toffaletti DL, Klem TJ, Perfect JR, Davisson VJ. 1998. Biochemical role of the *Cryptococcus Neoformans* ADE2 protein in fungal *de novo* purine biosynthesis. Arch Biochem Biophys 351:123–134. https://doi.org/10.1006/abbi.1997.0512
122. Firestine SM, Poon S-W, Mueller EJ, Stubbe J, Davisson VJ. 1994. Reactions catalyzed by 5-aminoimidazole ribonucleotide carboxylases from *Escherichia coli* and *Gallus gallus*: a case for divergent catalytic mechanisms. Biochemistry 33:11927–11934. https://doi.org/10.1021/bi00205a031
123. Schwarz R, Lieman-Hurwitz J, Hassidim M, Kaplan A. 1992. Phenotypic complementation of high CO_2-requiring mutants of the cyanobacterium *Synechococcus* SP. strain PCC 7942 by inosine 5'-monophosphate. Plant Physiol 100:1987–1993. https://doi.org/10.1104/pp.100.4.1987
124. Brown AM, Hoopes SL, White RH, Sarisky CA. 2011. Purine biosynthesis in *Archaea*: variations on a theme. Biol Direct 6:63. https://doi.org/10.1186/1745-6150-6-63
125. Fan S-H, Matsuo M, Huang L, Tribelli PM, Götz F, Khursigara CM, Scott K. 2021. The MpsAB bicarbonate transporter is superior to carbonic anhydrase in biofilm-forming bacteria with limited CO_2 diffusion. Microbiol Spectr 9:e0030521. https://doi.org/10.1128/spectrum.00305-21
126. Blanco-Ameijeiras S, Stoll HM, Zhang HR, Hopkinson BM. 2020. Influence of temperature and CO_2 on plasma-membrane permeability to CO_2 and HCO_3^- in the marine haptophytes *Emiliania huxleyi* and *Calcidiscus leptoporus* (Prymnesiophyceae). J Phycol 56:1283–1294. https://doi.org/10.1111/jpy.13017
127. Konings WN, Albers S-V, Koning S, Driessen AJM. 2002. The cell membrane plays a crucial role in survival of bacteria and *Archaea* in extreme environments. Antonie Van Leeuwenhoek 81:61–72. https://doi.org/10.1023/a:1020573408652
128. Price GD, Badger MR. 1989. Isolation and characterization of high CO_2-requiring-mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO_2 within the carboxysome. Plant Physiol 91:514–525. https://doi.org/10.1104/pp.91.2.514
129. Cohen SN, Chang AC, Boyer HW, Helling RB. 1973. Construction of biologically functional bacterial plasmids *in vitro*. Proc Natl Acad Sci U S A 70:3240–3244. https://doi.org/10.1073/pnas.70.11.3240
130. Dodd MS, Papineau D, Grenne T, Slack JF, Rittner M, Pirajno F, O'Neil J, Little CTS. 2017. Evidence for early life in earth's oldest hydrothermal vent precipitates. Nature 543:60–64. https://doi.org/10.1038/nature21377
131. Schopf JW, Kitajima K, Spicuzza MJ, Kudryavtsev AB, Valley JW. 2018. SIMS analyses of the oldest known assemblage of microfossils document their taxon-correlated carbon isotope compositions. Proc Natl Acad Sci U S A 115:53–58. https://doi.org/10.1073/pnas.1718063115
132. Betts HC, Puttick MN, Clark JW, Williams TA, Donoghue PCJ, Pisani D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. Nat Ecol Evol 2:1556–1562. https://doi.org/10.1038/s41559-018-0644-x