WILEY

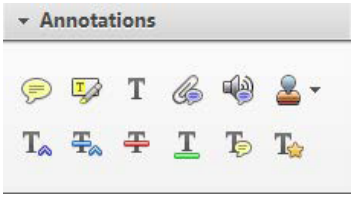
Online Proofing System Instructions

The Wiley Online Proofing System allows proof reviewers to review PDF proofs, mark corrections, respond to queries, upload replacement figures, and submit changes directly from the locally saved PDF proof.

1. For the best experience when reviewing your PDF proof ensure you are connected to the internet. This will allow the locally saved PDF proof to connect to the central Wiley Online Proofing System server. If you are connected to the Wiley Online Proofing System server you should see a green check mark icon above in the yellow banner.

Connected Disconnected

2. Please review the article proof on the following pages and mark any corrections, changes, and query responses using the Annotation Tools outlined on the next 2 pages.



3. Save your proof corrections by clicking the "Publish Comments" button in the yellow banner above. Corrections don't have to be marked in one sitting. You can publish comments and log back in at a later time to add and publish more comments before you click the "Complete Proof Review" button below.

4. If you need to supply additional or replacement files bigger than 5 Megabytes (MB) do not attach them directly to the PDF Proof, please click the "Upload Files" button to upload files:

Upload files

Click Here

5. When your proof review is complete and all corrections have been published to the server by clicking the "Publish Comments" button, please click the "Complete Proof Review" button below:

IMPORTANT: Did you reply to all author queries found on the first page of your proof?

IMPORTANT: Did you click the "Publish Comments" button to save all your corrections? Any unpublished comments will be lost.

IMPORTANT: Once you click "Complete Proof Review" you will not be able to add or publish additional corrections.

Complete Proof Review

Click Here

WILEY

Online Proofing System

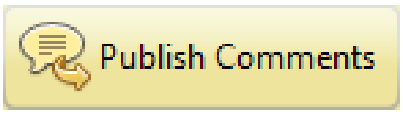
Enabling the Adobe PDF Viewer

In order to proof your article Adobe Reader or Adobe Acrobat needs to be your browser's default PDF viewer. See how to set this up for Internet Explorer, Firefox, and Safari at https://helpx.adobe.com/acrobat/using/display-pdf-in-browser.html

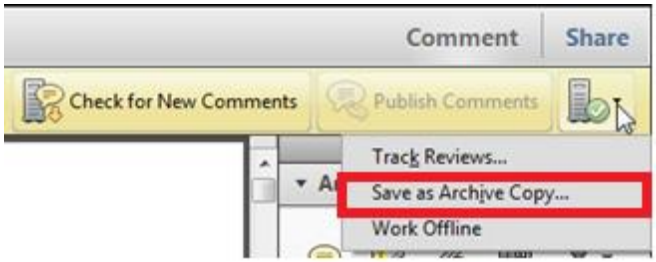
Google Chrome and Microsoft Edge do not support Adobe Reader or Adobe Acrobat as a PDF Viewer. We recommend using Internet Explorer, Firefox, or Safari.

1. Mark your corrections, changes, and query responses using the Annotation Tools outlined on the next 2 pages.

2. Save your proof corrections by clicking the "Publish Comments" button in the yellow banner above. Corrections don't have to be marked in one sitting. You can publish comments and log back in at a later time to add and publish more comments before you click the "Complete Proof Review" button.



3. When your proof review is complete we recommend you download a copy of your annotated proof for reference in any future correspondence concerning the article before publication. You can do this by clicking on the icon to the right of the 'Publish Comments' button and selecting 'Save as Archive Copy…'.



IMPORTANT: Did you reply to all queries listed on the Author Query Form appearing before your proof?

IMPORTANT: Did you click the "Publish Comments" button to save all your corrections? Any unpublished comments will be lost.

IMPORTANT: Once you click "Complete Proof Review" you will not be able to add or publish additional corrections.

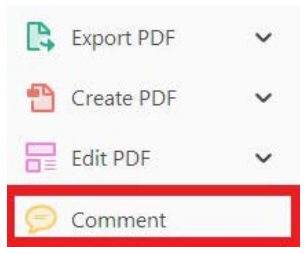
4. When your proof review is complete and all corrections have been published to the server by clicking the "Publish Comments" button, please click the "Complete Proof Review" button appearing above the proof in your web browser window.

WILEY

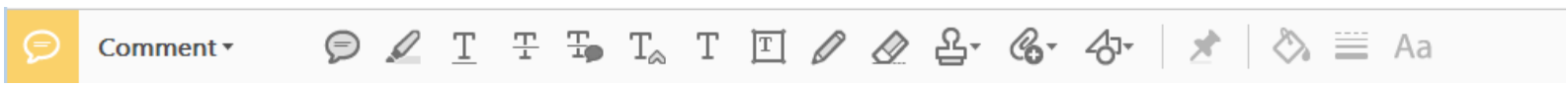
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-Annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 11 or above). (Note that this document uses screenshots from Adobe Reader DC.)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab (right-hand panel or under the Tools menu).



This will open up a ribbon panel at the top of the document. Using a tool will place a comment in the right-hand panel. The tools you will use for annotating your proof are shown below:

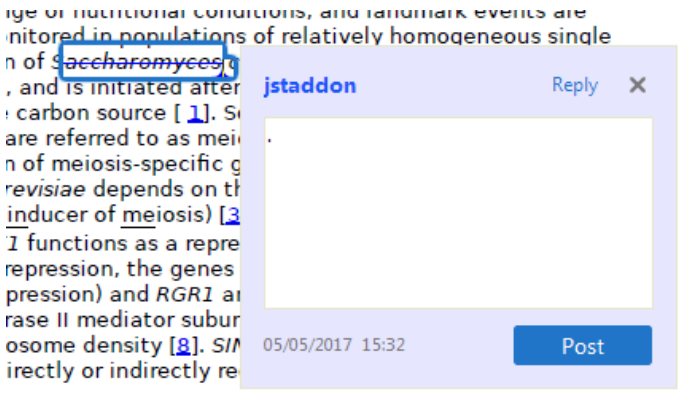


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it:

- Highlight a word or sentence.
- Click on .
- Type the replacement text into the blue box that appears.

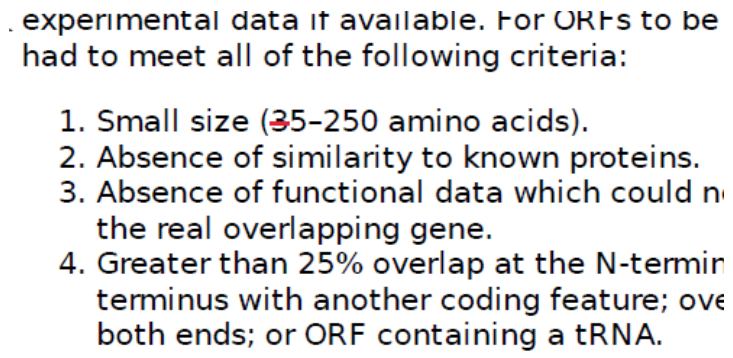


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it:

- Highlight a word or sentence.
- Click on .
- The text will be struck out in red.


experimental data if available. For ORFs to be had to meet all of the following criteria:

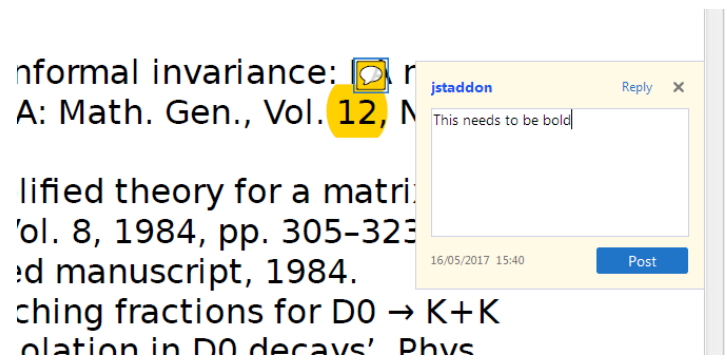
1. Small size (~~3~~5–250 amino acids).
2. Absence of similarity to known proteins.
3. Absence of functional data which could n the real overlapping gene.
4. Greater than 25% overlap at the N-termir terminus with another coding feature; ove both ends; or ORF containing a tRNA.

3. Commenting Tool – for highlighting a section to be changed to bold or italic or for general comments.

Use these 2 tools to highlight the text where a comment is then made.

How to use it:

- Click on .
- Click and drag over the text you need to highlight for the comment you will add.
- Click on .
- Click close to the text you just highlighted.
- Type any instructions regarding the text to be altered into the box that appears.

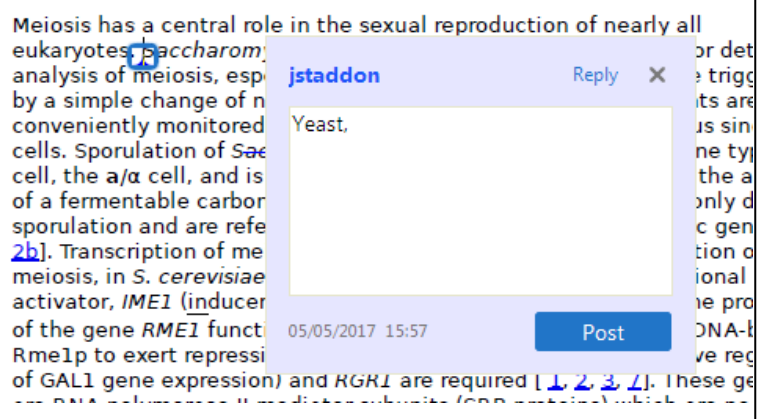


4. Insert Tool – for inserting missing text at specific points in the text.

Marks an insertion point in the text and opens up a text box where comments can be entered.

How to use it:

- Click on .
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the box that appears.



WILEY

USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

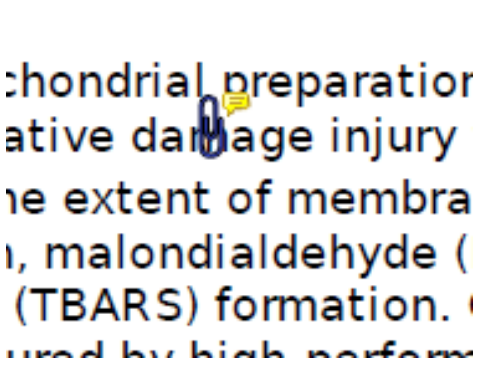
5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate place in the text.

How to use it:

- Click on  .
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.

  The attachment appears in the right-hand panel.

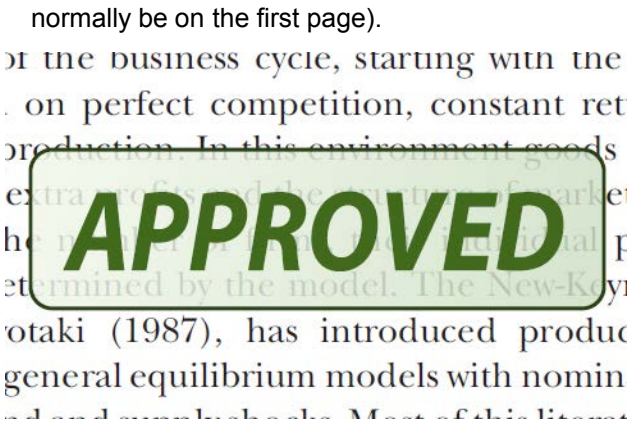


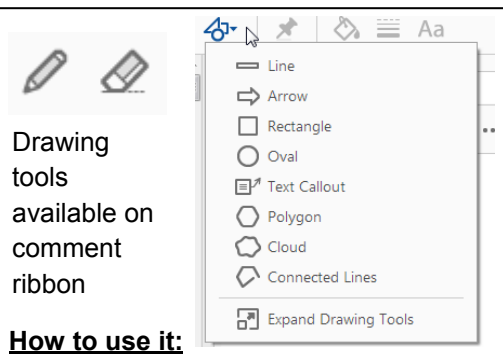
6. Add stamp Tool – for approving a proof if no corrections are required.

Inserts a selected stamp onto an appropriate place in the proof.

How to use it:

- Click on  .
- Select the stamp you want to use. (The Approved stamp is usually available directly in the menu that appears. Others are shown under *Dynamic*, *Sign Here*, *Standard Business*).
- Fill in any details and then click on the proof where you'd like the stamp to appear. (Where a proof is to be approved as it is, this would normally be on the first page).





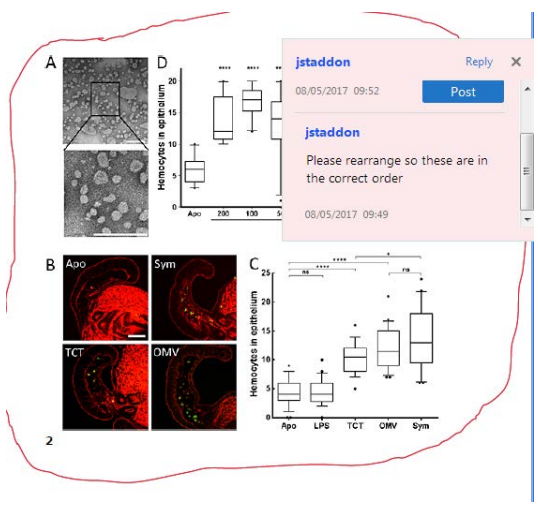
Drawing tools available on comment ribbon

7. Drawing Markups Tools – for drawing shapes, lines, and freeform annotations on proofs and commenting on these marks.

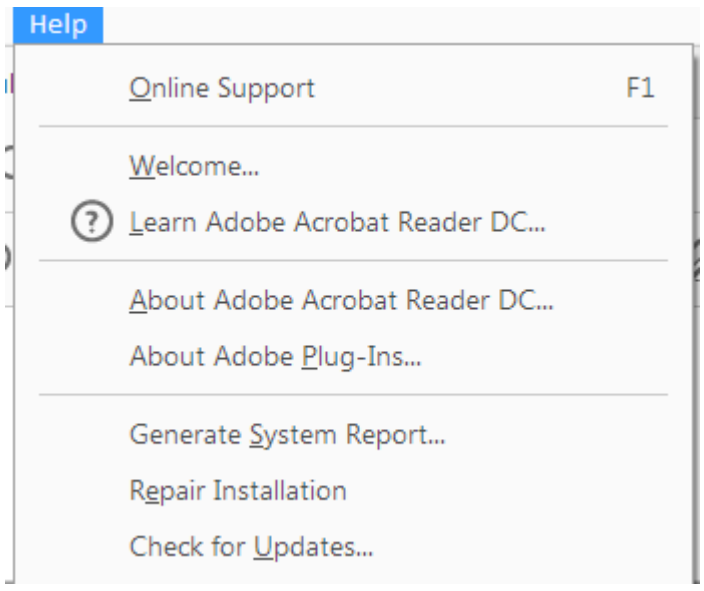
Allows shapes, lines, and freeform annotations to be drawn on proofs and for comments to be made on these marks.

How to use it:

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, right-click on shape and select *Open Pop-up Note.*
- Type any text in the red box that appears.



For further information on how to annotate proofs, click on the Help menu to reveal a list of further options:



WILEY

Author Query Form

Journal: EMI

Article: 14090

Dear Author,

During the copyediting of your manuscript the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond to each by marking the necessary comments using the PDF annotation tools.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

Query References	Query	Remarks?
AQ1	AUTHOR: Please confirm that given names (red) and surnames/family names (green) have been identified correctly.	Confirmed

Funding Info Query Form

Please confirm that the funding sponsor list below was correctly extracted from your article: that it includes all funders and that the text has been matched to the correct FundRef Registry organization names. If a name was not found in the FundRef registry, it may be not the canonical name form or it may be a program name rather than an organization name, or it may be an organization not yet included in FundRef Registry. If you know of another name form or a parent organization name for a "not found" item on this list below, please share that information.

FundRef name	FundRef Organization Name
USDA Higher Education Challenge Grant	U.S. Department of Agriculture
U.S. Department of Energy)?>U.S. Department of Energy Joint Genome Institute	U.S. Department of Energy Joint Genome Institute

This sponsor list is correct

environmental microbiology

Environmental Microbiology (2018) 00(00), 00–00 doi:10.1111/1462-2920.14090

Genomes of ubiquitous marine and hypersaline *Hydrogenovibrio*, *Thiomicrorhabdus* and *Thiomicrospira* spp. encode a diversity of mechanisms to sustain chemolithoautotrophy in heterogeneous environments

AQ1 Kathleen M. Scott,[1]* John Williams,[1]
Cody M. B. Porter,[1] Sydney Russel,[1]
Tara L. Harmer,[2] John H. Paul,[1] Kirsten M. Antonen,[1]
Megan K. Bridges,[1] Gary J. Camper,[1]
Christie K. Campla,[1] Leila G. Casella,[1] Eva Chase,[1]
James W. Conrad,[1] Mercedez C. Cruz,[1]
Darren S. Dunlap,[1] Laura Duran,[1]
Elizabeth M. Fahsbender,[1] Dawn B. Goldsmith,[1]
Ryan F. Keeley,[1] Matthew R. Kondoff,[1]
Breanna I. Kussy,[1] Marannda K. Lane,[1]
Stephanie Lawler,[1] Brittany A. Leigh,[1]
Courtney Lewis,[1] Lygia M. Lostal,[1] Devon Marking,[1]
Paola A. Mancera,[1] Evan C. McClenthan,[1]
Emily A. McIntyre,[1] Jessica A. Mine,[1] Swapnil Modi,[1]
Brittney D. Moore,[1] William A. Morgan,[1]
Kaleigh M. Nelson,[1] Kimmy N. Nguyen,[1]
Nicholas Ogburn,[1] David G. Parrino,[1]
Anangamanjari D. Pedapudi,[1] Rebecca P. Pelham,[1]
Amanda M. Preece,[1] Elizabeth A. Rampersad,[1]
Jason C. Richardson,[1] Christina M. Rodgers,[1]
Brent L. Schaffer,[1] Nancy E. Sheridan,[1]
Michael R. Solone,[1] Zachery R. Staley,[1]
Maki Tabuchi,[1] Ramond J. Waide,[1]
Pauline W. Wanjugi,[1] Suzanne Young,[1] Alicia Clum,[3]
Chris Daum,[3] Marcel Huntemann,[3] Natalia Ivanova,[3]
Nikos Kyrpides,[3] Natalia Mikhailova,[3]
Krishnaveni Palaniappan,[3] Manoj Pillay,[3]
T.B.K. Reddy,[3] Nicole Shapiro,[3] Dimitrios Stamatis,[3]
Neha Varghese,[3] Tanja Woyke,[3] Rich Boden,[4,5]
Sharyn K. Freyermuth[6] and Cheryl A. Kerfeld[7,8,9]

[1]*Department of Integrative Biology, University of South Florida, 4202 East Fowler Avenue, Tampa, FL, 33620, USA.*
[2]*Biology Program, Stockton University, Galloway, NJ, USA.*
[3]*Department of Energy Joint Genome Institute, Walnut Creek, CA, USA.*
[4]*School of Biological & Marine Sciences, University of Plymouth, Drake Circus, Plymouth, UK.*
[5]*Sustainable Earth Institute, University of Plymouth, Drake Circus, Plymouth, UK.*
[6]*Department of Biochemistry, University of Missouri, Columbia, MO, USA.*
[7]*MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI, USA.*
[8]*Department of Plant and Microbial Biology, University of California, Berkeley, CA, USA.*
[9]*MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.*

Received 13 September, 2017; revised 26 February, 2018; accepted 3 March, 2018. *For correspondence. E-mail kmscott@usf.edu; Tel. +813 974 5173; Fax +813 974 3263.

Summary

Chemolithoautotrophic bacteria from the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* are common, sometimes dominant, isolates from sulfidic habitats including hydrothermal vents, soda and salt lakes and marine sediments. Their genome sequences confirm their membership in a deeply branching clade of the *Gammaproteobacteria*. Several adaptations to heterogeneous habitats are apparent. Their genomes include large numbers of genes for sensing and responding to their environment (EAL- and GGDEF-domain proteins and methyl-accepting chemotaxis proteins) despite their small sizes (2.1–3.1 Mbp). An array of sulfur-oxidizing complexes are encoded, likely to facilitate these organisms' use of multiple forms of reduced sulfur as electron donors. Hydrogenase genes are present in some taxa, including group 1d and 2b hydrogenases in *Hydrogenovibrio marinus* and *H. thermophilus* MA2-6, acquired via horizontal gene transfer. In addition to high-affinity *cbb*$_3$ cytochrome *c* oxidase, some also encode cytochrome *bd*-type quinol oxidase or

ba_3-type cytochrome c oxidase, which could facilitate growth under different oxygen tensions, or maintain redox balance. Carboxysome operons are present in most, with genes downstream encoding transporters from four evolutionarily distinct families, which may act with the carboxysomes to form CO_2 concentrating mechanisms. These adaptations to habitat variability likely contribute to the cosmopolitan distribution of these organisms.

Introduction

Members of the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* are common in sulfidic habitats worldwide, present and often numerous at hydrothermal vents, terrestrial hot springs, soda and salt lakes and coastal sediments (Table 1; Kuenen and Veldkamp, 1972; Jannasch, 1985; Brinkhoff and Muyzer, 1997; Kato *et al*., 2009; Brazelton and Baross, 2010; Sorokin *et al*., 2011). These organisms, formerly classified in the genera *Hydrogenovibrio, Thiomicrospira* and *Thioalkalimicrobium*, have been taxonomically reassigned based on a polyphasic analysis, including the 16S rRNA gene and 53 ribosomal protein gene sequences (Table 1; Boden *et al*., 2017b). These obligate aerobes all grow chemolithoautotrophically on reduced sulfur compounds; some are also able to grow on molecular hydrogen (Table 1). All fix CO_2 using the transaldolase variant of the Calvin–Benson–Bassham cycle (Kuenen and Veldkamp, 1972; Nishihara *et al*., 1991; Muyzer *et al*., 1995; Ahmad *et al*., 1999; Brinkhoff *et al*., 1999b,c; Rainey *et al*., 2001; Sorokin *et al*., 2002; Knittel *et al*., 2005; Sorokin *et al*., 2006; 2007; Hansen and Perner, 2015; Zhang *et al*., 2016).

Hydrogenovibrio crunogenus XCL-2 was the first member of these genera to have its genome sequenced, and its genome has some puzzling features (Scott *et al*., 2006). Low numbers of membrane transporters reflect its obligately autotrophic lifestyle, although there are a surprisingly high number of genes encoding methyl-accepting chemotaxis proteins, and it was proposed that enhanced capability of sensing, and motility toward, microhabitats in which nutrients were available could compensate for this species' ability to use rather few nutrients for growth (Scott *et al*., 2006). The capability of using reduced sulfur compounds as electron donors is reflected by the presence of genes encoding the 'Sox' complex (*soxABCDXYZ*) associated with the Kelly–Friedrich pathway and sulfide:quinone oxidoreductase (EC 1.8.5.4) (Scott *et al*., 2006). Likewise, genes encoding a [NiFe] hydrogenase suggest the potential for growth on molecular hydrogen, although the presence of this gas does not facilitate growth by this strain under the provided conditions (Hansen and Perner, 2016).

The *H. crunogenus* XCL-2 genome suggests a departure from previous models of dissolved inorganic carbon (DIC) uptake in autotrophic microorganisms. *H. crunogenus* XCL-2 has a CO_2-concentrating mechanism (CCM) that facilitates growth under low DIC conditions (Dobrinski *et al*., 2005). Like the many species of the '*Cyanobacteria*' in which CCMs have been well characterized, *H. crunogenus* XCL-2 can generate an elevated concentration of intracellular DIC (Dobrinski *et al*., 2005; Price, 2011). This large pool of intracellular DIC drives carbon fixation by carboxysomes, intracellular microcompartments containing RubisCO (ribulose 1,5-bisphosphate carboxylase/oxygenase; EC 4.1.1.39) and carbonic anhydrase (EC 4.2.1.1; Cannon *et al*., 2010; Menning *et al*., 2016). The ability to generate intracellular DIC concentrations $>$ 100-fold higher than extracellular suggests active DIC uptake (Dobrinski *et al*., 2005), but genes orthologous to any of the DIC transporters characterized in the '*Cyanobacteria*' are absent in the *H. crunogenus* XCL-2 genome (Menning *et al*., 2016). In *H. crunogenus*, DIC uptake is facilitated by a novel two-component transporter. One subunit of this transporter is a member of a 'domain of unknown function' protein family (PFAM10070; Bateman *et al*., 2002), while the other subunit, which is predicted to be a transmembrane protein, belongs to a protein family that includes proton transporters (PFAM00361; Mangiapia *et al*., 2017).

Hydrogenovibrio crunogenus XCL-2 genome data also predict a peculiar citric acid cycle, in which 2-oxoglutarate dehydrogenase (EC 1.2.4.2/2.3.1.61/1.8.1.4) and malate dehydrogenase (EC 1.1.3.7) are absent (Scott *et al*., 2006). Malate:quinone oxidoreductase (EC 1.1.5.4) is encoded in the genome (Scott *et al*., 2006), is active in membrane fractions (Quasem *et al*., 2017), and could act as a replacement for malate dehydrogenase. The 2-oxoglutarate:ferredoxin oxidoreductase (EC 1.2.7.3) activity is absent, suggesting that this obligate autotroph may have a wishbone-shaped 'citric acid pathway', with an absence of interconversion of 2-oxoglutarate and succinyl-coA (Smith *et al*., 1967; Wood *et al*., 2004). However, genes encoding 2-oxoglutarate decarboxylase and succinic semialdehyde dehydrogenase are present, which could potentially close the citric acid cycle by converting 2-oxoglutarate to succinyl-CoA (Quasem *et al*., 2017).

At the time of its sequencing, placing these and other features of the *H. crunogenus* XCL-2 genome within an evolutionary context was complicated by the absence of genome sequences of close relatives among the basal *Gammaproteobacteria* (Williams *et al*., 2010). In order to determine whether the peculiar aspects of *H. crunogenus* XCL-2 described above are unique to the species, or are typical for the lineage, the genomes of eleven additional members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* were sequenced, representing the geographic, habitat and phylogenetic breadth of these genera, and

Table 1. Members of the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* with sequenced genomes.

Taxon	Basonym[a]	Habitat	Genome published	Electron donors[b,c]	Organic carbon oxidation or assimilation[c,d]	Opti-mum pH[c]	Opti-mum °C[c]	Opti-mum Na^+ (M)[c]
H. crunogenus XCL-2	*Thiomicrospira crunogena* XCL-2	East Pacific Rise deep-sea hydrothermal vent (Ahmad *et al.*, 1999)	(Scott *et al.*, 2006)	$S_2O_3^{-2}$, H_2S, S^0	–			
Hydrogenovibrio sp. XS5	*Thiomicrospira* sp. XS5	Red Sea deep brine-water interface (Zhang *et al.*, 2016)	(Zhang *et al.*, 2016)	$S_2O_3^{-2}$				
H. thermophilus JR2 DSM 25194	*Thiomicrospira* sp. JR2	Northeast Pacific deep-sea hydrothermal vent	This work	$S_2O_3^{-2}$				
H. thermophilus MA2–6	*Thiomicrospira* sp. MA2–6	Mid-Atlantic deep-sea hydrothermal vent (Muyzer *et al.*, 1995)	This work	$S_2O_3^{-2}$				
H. halophilus	*Thiomicrospira halophila*	Hypersaline lake (Russia) (Sorokin *et al.*, 2006)	This work	$S_2O_3^{-2}$, H_2S, S^0		7.5–7.8		1.5
Hydrogenovibrio sp. WB1	*Thiomicrospira* sp. WB1	Red Sea deep brine-water interface (Zhang *et al.*, 2016)	(Zhang *et al.*, 2016)	$S_2O_3^{-2}$				
H. kuenenii	*Thiomicrospira kuenenii*	Wadden Sea mud flat (Brinkhoff *et al.*, 1999a)	This work	$S_2O_3^{-2}$, H_2S, S^0, $S_4O_6^{-2}$	NG	6.0		0.42
H. marinus DSM 11271	N/A	Surface seawater near Japan (Nishihara *et al.*, 1991)	This work and (Jo *et al.*, 2014)	H_2, $S_2O_3^{-2}$, S^0, $S_4O_6^{-2}$	NG	6.5	37	0.5
Hydrogenovibrio sp. Milos-T1	*Thiomicrospira* sp. Milos-T1	Aegean shallow hydrothermal vent (Brinkhoff *et al.*, 1999b)	This work	$S_2O_3^{-2}$				
Tmr. frisia Kp2 DSM 25197	*Thiomicrospira* sp. Kp2	Northeast Pacific deep-sea hydrothermal vent	This work	$S_2O_3^{-2}$				
Thiomicrorhabdus sp. Milos-T2	*Thiomicrospira* sp. Milos-T2	Aegean shallow hydrothermal vent (Brinkhoff *et al.*, 1999b)	This work	$S_2O_3^{-2}$				
Tmr. arctica	*Thiomicrospira arctica*	Arctic Ocean coastal sediments near Svalbard (Knittel *et al.*, 2005)	This work	$S_2O_3^{-2}$, S^0, $S_4O_6^{-2}$	NG	7.3	11–13	0.25
Tmr. chilensis	*Thiomicrospira chilensis*	Chile continental shelf sediment (Brinkhoff *et al.*, 1999c)	This work	$S_2O_3^{-2}$, H_2S, S^0, $S_4O_6^{-2}$	NG	7.0	32–37	
Tms. aerophila	*Thioalkalimicrobium aerophilum* AL3	Hypersaline lake (Russia) (Rainey *et al.*, 2001)		$S_2O_3^{-2}$, H_2S, S^0, PS, $S_4O_6^{-2}$	A, NG	10		
Tms. cyclica	*Thioalkalimicrobium cyclicum*	Hypersaline lake (US) (Sorokin *et al.*, 2002)	(Kappler *et al.*, 2016)	$S_2O_3^{-2}$, H_2S, PS, $S_4O_6^{-2}$		9.5		0.3
Thiomicrospira sp. ALE5	*Thioalkalimicrobium* sp. ALE5	Hypersaline lake (Egypt) (Sorokin *et al.*, 2011)		$S_2O_3^{-2}$, H_2S, PS		9.5–10		
Tms. microaerophila				$S_2O_3^{-2}$, H_2S	A, NG	8–9	25–28	0.4–0.5

Table 1. *cont.*

Taxon	Basonym[a]	Habitat	Genome published	Electron donors[b,c]	Organic carbon oxidation or assimilation[c,d]	Opti-mum pH[c]	Opti-mum °C[c]	Opti-mum Na^+ (M)[c]
	Thioalkalimicrobium microaerophilum	Hypersaline lake (US) (Sorokin *et al.*, 2007)						
Tms. pelophila	N/A	Wadden Sea mud flat (Kuenen and Veldkamp, 1972)	This work	$S_2O_3^{-2}$, H_2S, S^0, $S_4O_6^{-2}$	A, NG	6.5–7.5	28–30	

a. Boden *et al.*, 2017b.
b. PS, polysulfide.
c. Data from (Kuenen and Veldkamp, 1972; Jannasch *et al.*, 1985; Nishihara *et al.*, 1991; Muyzer *et al.*, 1995; Brinkhoff *et al.*, 1999a,b,c; Rainey *et al.*, 2001; Sorokin *et al.*, 2002; Knittel *et al.*, 2005; Sorokin *et al.*, 2007; 2008; 2011; Zhang *et al.*, 2016; Ang *et al.*, 2017; Quasem *et al.*, 2017).
d.-, no assimilation of organic compounds, and no growth in the absence of reduced sulfur compounds; NG, no growth in the absence of inorganic electron donors; A, organic carbon assimilation.

genome comparisons among these organisms were expanded to include members of this lineage sequenced by other projects (Table 1).

Results and discussion

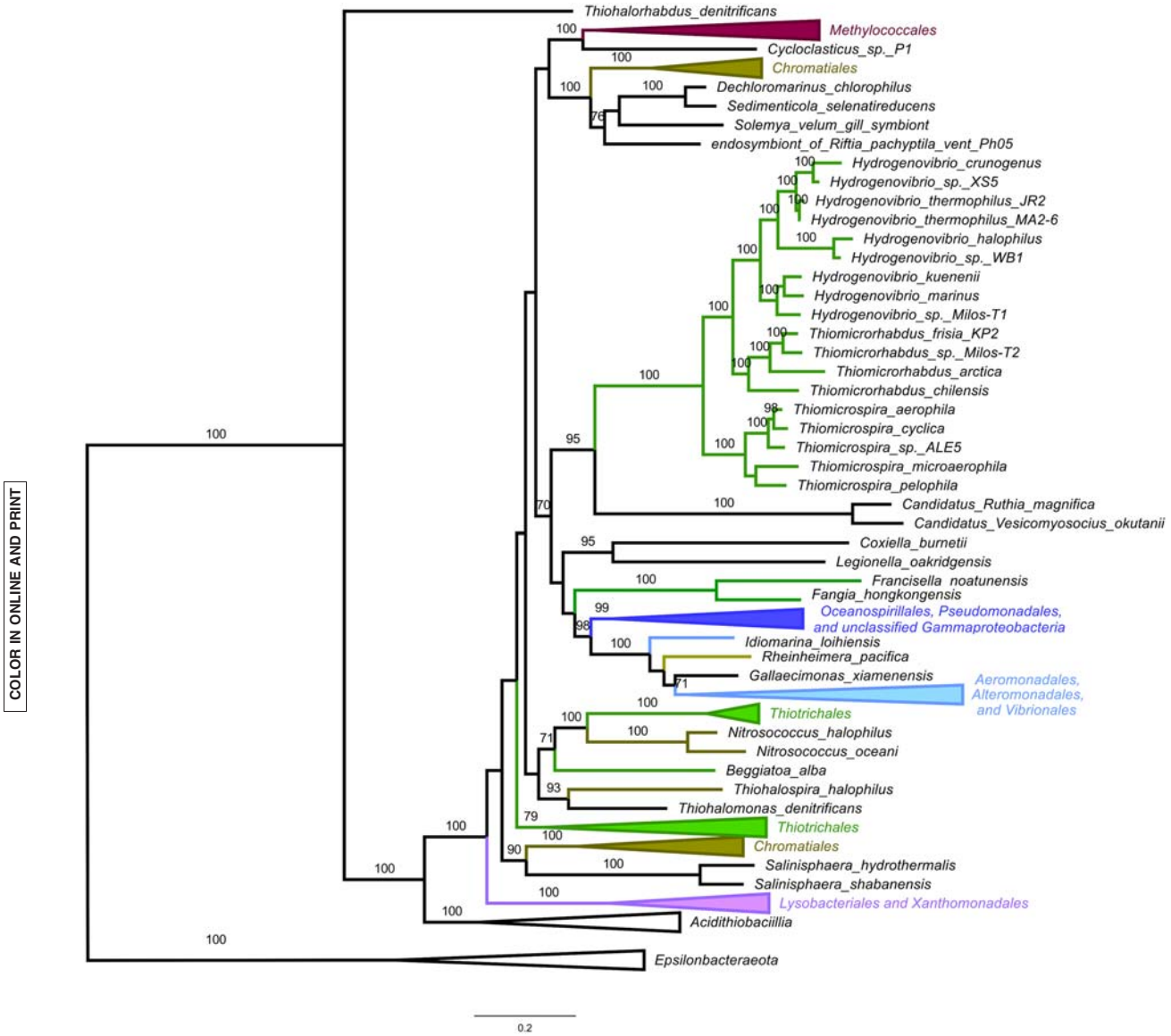
Ribosomal protein supertree

As previously described (Boden *et al.*, 2017b), the supertree constructed from amino acid sequences derived from ribosomal protein genes places the genera *Thiomicrospira, Thiomicrorhabdus* and *Hydrogenovibrio* together in a well-supported clade (Fig. 1). Since *Thiomicrospira* sp. Milos T1, XS-5 and WB-1 fall within genus *Hydrogenovibrio*, and *Thiomicrospira* sp. Milos-T2 falls within the genus *Thiomicrorhabdus*, they will be referred to here with those genus designations (Table 1). As also previously shown (Boden *et al.*, 2017b), on the basis of 16S rRNA gene sequence identities (100% vs. type strain I78) and genome-wide average nucleotide identities (96.5% between JR2 and MA2–6), *Thiomicrospira* sp. JR2 and MA2–6 are strains of *H. thermophilus* (values calculated relative to type strain I78), and will be referred to here as such (e.g., *H. thermophilus* JR2). *Thiomicrorhabdus* sp. KP2 is a strain of *Tmr. frisia* (99.5% 16S sequence identity with type strain JB-A2).

The order *Thiotrichales*, within which the genera *Thiomicrospira*, *Thiomicrorhabdus* and *Hydrogenovibrio* are circumscribed, is represented on the supertree by many well-supported clades, some of which are intermixed with clades of the *Chromatiales* (Fig. 1). This is also the case with many of the orders of the *Gammaproteobacteria* on this and other supertrees (Williams *et al.*, 2010; Ramulu *et al.*, 2014), suggesting that the taxonomy of the *Gammaproteobacteria* should be revised.

Genome structures and general features

Genome sizes among permanent draft and finished genome sequences of the genera *Thiomicrospira, Thiomicrorhabdus* and *Hydrogenovibrio* (2.1–3.1 Mb; Table 2) fall within the range of sizes of genomes of other free-living autotrophs from the phylum '*Proteobacteria*' (1.7–10.1 Mb; Supporting Information Table S1). Their genomes are larger than those of the vertically transmitted chemolithoautotrophic symbionts of vesicomyid clams (1.0–1.2 Mb), which reflects genome reduction in the symbionts (Kuwahara *et al.*, 2007; Newton *et al.*, 2007; 2008). Their genomes are smaller than those from organisms with additional capabilities such as photosynthesis, denitrification or heterotrophic growth (Supporting Information Table S1). The rRNA operon copy numbers (3–4) are at the high range of what is observed in other autotrophic members of the *Gammaproteobacteria* (Supporting Information Table S1), which may help them to respond more quickly to



COLOR IN ONLINE AND PRINT

Fig. 1. Maximum likelihood analysis of concatenated alignments of amino acid sequences derived from gene sequences encoding ribosomal proteins. Members of the order *Chromatiales* are indicated by olive-coloured clades and branches, while members of the *Thiotrichales* have green clades and branches. Bootstrap values > 70% from 200 resamplings of the alignment are shown, and the tree is unrooted. The scale bar represents the number of substitutions per site.

changes in their environment (Klappenbach *et al.*, 2000) and to have higher maximum specific growth rates (Roller *et al.*, 2016). Indeed, members of these genera are among the first to appear in media inoculated from environmental samples (K. Scott and R. Boden, unpubl. data).

Many members of *Hydrogenovibrio* and *Thiomicrorhabdus* have large tandem repeats preceding a gene encoding protein kinase (e.g., *Tcr_0106*; Fig. 2). *Hydrogenovibrio halophilus* lacks an ortholog to this protein kinase, as well as the repeated regions. For those organisms whose tandem repeat regions have been completely sequenced, repeat lengths range from 279 to 393 nucleotides in length, with 5 to > 27 copies in tandem (upper range is at the end of a contig; Fig. 2). Some members of these two genera whose draft genome sequences currently lack large tandem repeats are likely to also have them; in these organisms, orthologs to the protein kinases described above are present at the 5′ end of their contigs and the absence of tandem repeats may result from the difficulties in sequencing and assembly of these regions (Treangen and Salzberg, 2012). In other organisms, repeated regions such as these are hotspots for genome rearrangement that can serve to generate diversity of gene products (if falling within a gene) or diversity in gene

J_ID: **EMI** Customer A_ID: **EMI14090** Cadmus Art: **EMI14090** Ed. Ref. No.: **EMI-2018-0114.R1** Date: **9-March-18** Stage: Page: **6**

Table 2. Genome sequencing results.

Taxon	# Contigs	# Scaffolds	Average coverage (fold)	Estimated genome size (Mb)	Gene count	GC content (%)	# rRNA operons	tRNA count
H. thermophilus JR2	7	1	1031[a]	2.6	2444	50	3	44
H. thermophilus MA2–6	1	1	224.8	2.7	2546	50	3	45
H. halophilus	3	1	3302	2.4	2238	55	3	45
H. kuenenii	2	2	96.1	2.5	2289	42	3	43
H. marinus DSM 11271	3	3	284.5	2.6	2554	44	4	45
Hydrogenovibrio sp. Milos-T1	1	1	181.3	2.3	2253	44	3	44
Tmr. frisia Kp2	13	3	1455	2.7	2526	40	3	45
Thiomicrorhabdus sp. Milos-T2	8	2	1000.3	2.6	2382	38	3	46
Tmr. arctica	8	6	631.6	2.6	2337	42	3	45
Tmr. chilensis	5	2	679.5	2.4	2285	48	2	43
Tms. pelophila	1	1	178.4	2.1	2040	44	2	41

a. Illumina coverage; 454 coverage was 35.9-fold.

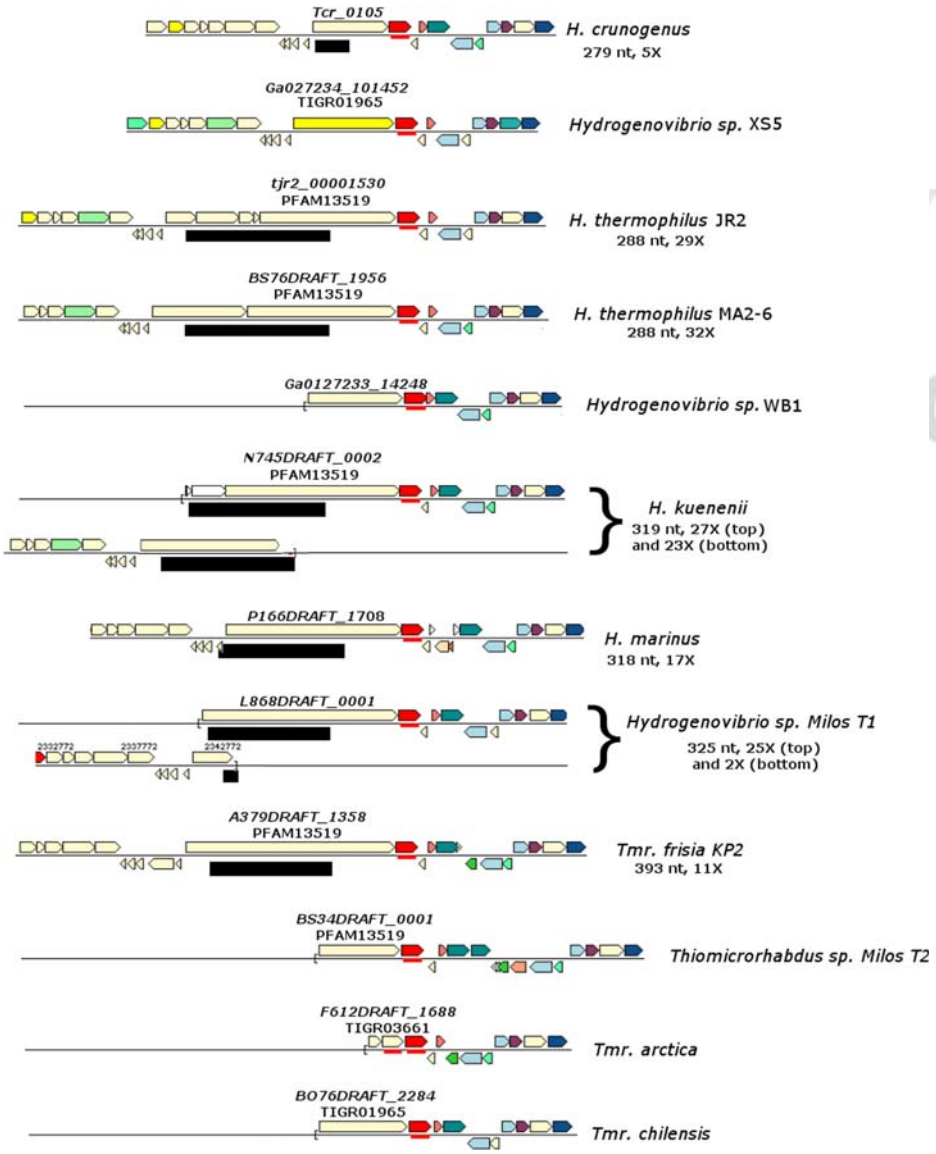


Fig. 2. Genome context for large tandem repeat regions. Genes are coloured by COG membership. Red genes encode protein kinase. Regions with large tandem repeats are indicated with black bars, and the size (in nucleotides) and number of tandem repeats is indicated next to the taxon name. Lines without genes indicate the ends of contigs. Locus tags and protein family membership for the large genes 5' to the protein kinase genes are indicated above the genes. PFAM13519— von Willebrand factor type A domain; TIGR01965—VCBS_repeat; TIGR03661—type I secretion C-terminal target domain.

COLOR IN ONLINE AND PRINT

ID: **nagarajulum** Time: 19:07 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

J_ID: **EMI** Customer A_ID: **EMI14090** Cadmus Art: **EMI14090** Ed. Ref. No.: **EMI-2018-0114.R1** Date: **9-March-18** Stage: Page: **7**

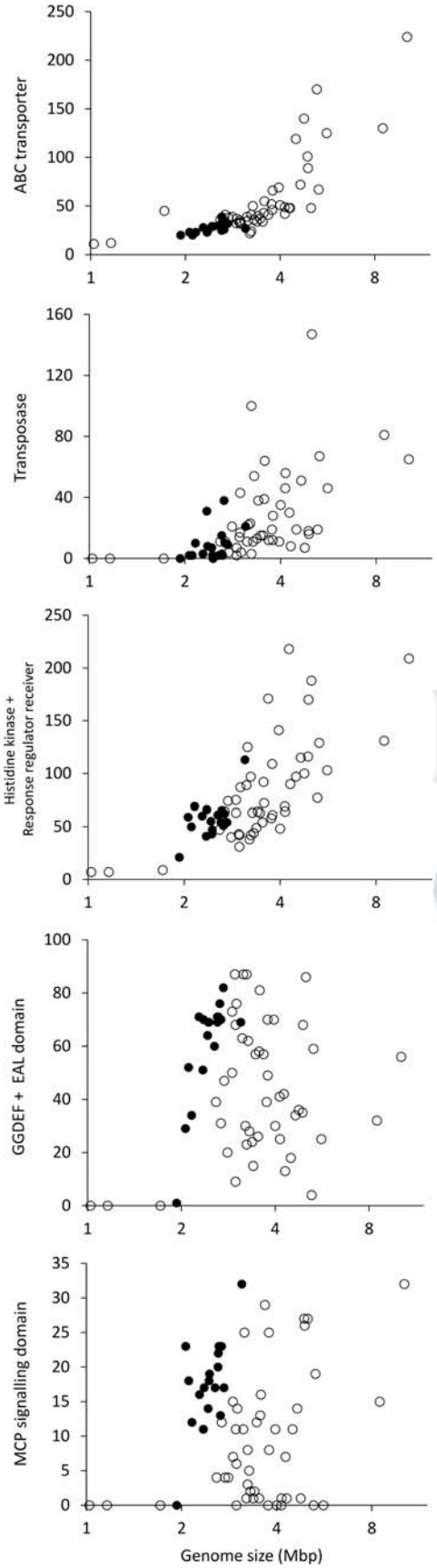


Fig. 3. Number of genes falling within selected Pfams in genomes of members of the genera *Hydrogenovibrio*, *Thiomicrorhabdus* and *Thiomicrospira* (solid black circles) and obligately or facultatively autotrophic members of the *Proteobacteria* (open circles). ABC transporter—members of Pfam00005; Transposase—members of Pfams 00872, 01527, 01609, 01797, 02371, 05598, 09299, 12762 and 12784; Histidine kinase + response regulator receiver—members of Pfams 02518 and 00072; GGDEF + EAL domain—members of Pfams 00990 and 00563; MCP signalling domain—members of Pfam00015.

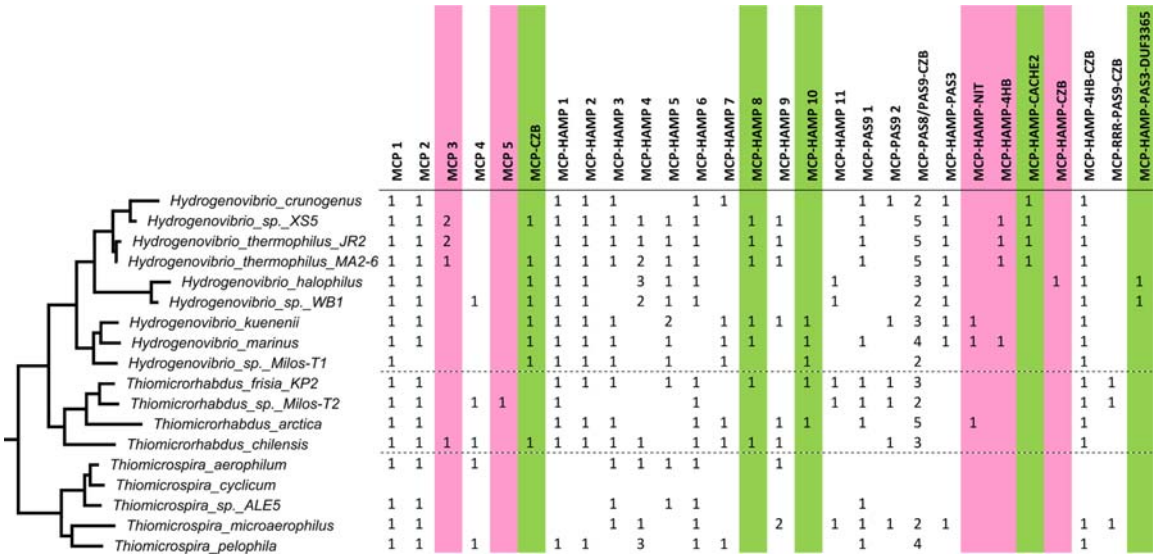
regulation (if falling upstream of a gene; Zhou *et al.*, 2014). Genes encoding large proteins (up to 4456 aa) have been annotated upstream from the protein kinase genes, and in many cases these large genes include the large tandem repeats (Fig. 2). These genes are variously annotated as Ig-like domain proteins or VCBS repeat containing proteins, and many fall within PFAM13519 (von Willebrand factor type A domain) or TIGR01965 (VCBS repeat). Both of these protein families include extracellular proteins involved in cell adhesion. When the amino acid sequences predicted from these genes are compared among the organisms studied here, they align poorly due to very low sequence similarities (~ 35%). Perhaps the tandem repeat regions serve to introduce variability in adhesion-related proteins, which could provide a mechanism for evading phage and other predators.

Putative prophages are present in the *Tmr. frisia* and *H. marinus* genomes (Supporting Information Fig. S1). These genome regions have G + C fractions that differ from their hosts (in mol%:mol% – *Tmr. frisia* KP2 prophage:host %GC = 44:40; *H. marinus* = 41:44), and they encode many phage-related proteins (Supporting Information Fig. S1). Perhaps the most tantalizing finding in the *H. marinus* prophage genome is alluded to by *P166DRAFT_0452-5*, a retroelement diversity generating mini-operon first discovered in *Bordatella* phage (Doulatov *et al.*, 2004). Diversity-generating retroelements (DGRs) are a family of genetic elements that function to diversify DNA sequences and the proteins they encode (Medhekar and Miller, 2007). Using an error-prone reverse transcriptase (*P166DRAFT_0452*) causes tropism switching that enables the phage to match the diversity generation of the host (Doulatov *et al.*, 2004). Thus, this phage has the putative capability to infect host cells of nearly limitless diversity. Usually the recipient of the diversity generation is genomically nearby, and in this case it may be the 'DUF3751: phage tail collar fibre protein-short tail fibre protein gp12' (*P166DRAFT_0457*).

Certain gene families are particularly well-represented in these genomes, when compared with other autotrophic members of the phylum '*Proteobacteria*' whose genomes have been completely sequenced (Fig. 3; Supporting Information Table S2). Based on membership in Pfams (Bateman *et al.*, 2002), genes encoding ABC transporters are abundant, though less so than in other autotrophic

ID: **nagarajulum** Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

J_ID: EMI Customer A_ID: EMI14090 Cadmus Art: EMI14090 Ed. Ref. No.: EMI-2018-0114.R1 Date: 9-March-18 Stage: Page: 8



COLOR IN ONLINE AND PRINT

Fig. 4. Clusters of genes encoding proteins with methyl-accepting chemotaxis protein signalling domains are sorted by taxa. Numbers of genes from each cluster present in each genome are listed for each taxon. Gene clusters that are absent from genus *Thiomicrospira*, but present in the genera *Thiomicrorhabdus* and *Hydrogenovibrio*, are shaded. Pink shading indicates clusters whose genes have no full length matches beyond these taxa, though some have many partial-length matches within them. Green shading indicates clusters with full-length matches in other members of *Gammaproteobacteria*. Gene clusters are named by Pfam domains present in their member's sequences: CACHE2—Pfam08269, Cache domain; CZB—Pfam13682, chemoreceptor zinc-binding domain; DUF3365—Pfam11845, Protein of unknown function; HAMP—Pfam00672, HAMP domain; MCP—Pfam00015, methyl-accepting chemotaxis protein signalling domain; NIT—Pfam08376, nitrate and nitrite sensing; PAS3—Pfam08447, PAS domain; PAS8—Pfam13188, PAS domain; PAS9—Pfam13426, PAS domain; RRR—Pfam00072, response regulator receiver domain; 4HB—Pfam12729, four helix bundle sensory module.

'Proteobacteria' with larger genomes (Fig. 3). Abundance of these transporters is a bit puzzling, given the requirements for chemolithoautotrophic growth consist of rather few nutrients, and also given this group's limited ability to assimilate organic carbon (Table 1). *Hydrogenovibrio* sp. Milos T1 and *Thiomicrorhabdus* sp. Milos T2 have the highest number of transposase genes of the genomes sequenced here (Fig. 3), which is interesting as these two organisms were isolated from the same habitat, but are not particularly closely related (Brinkhoff *et al.*, 1999b). In general, transposase gene frequencies are relatively low (Fig. 3), which may reflect the tendency for smaller genomes to contain fewer transposons (Touchon and Rocha, 2007).

Some gene families whose members facilitate sensing and responding to the environment are abundant in these genomes (Fig. 3). Genes encoding two-component regulatory systems (histidine kinase, response regulators) are slightly elevated in abundance based on genome sizes, but generally fall along an overall trend of increasing numbers with genome size (Galperin, 2005). Genes encoding GGDEF and EAL-domain proteins are particularly abundant in these taxa. In other organisms, these proteins synthesize and degrade the second-messenger cyclic di-GMP, which is involved in regulating motility, chemotaxis, biofilm formation and other interactions with organisms' biotic and abiotic environment (Römling *et al.*, 2013). The abundance of GGDEF and EAL-domain proteins, as well as methyl-accepting chemotaxis proteins, should make these organisms quite responsive to their spatially and temporally variable habitats.

Methyl-accepting chemotaxis proteins

Genes with MCP domains are very abundant in members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* (11–23 genes per genome; Fig. 3), particularly when compared with other members of family *Piscirickettsiaceae* (0–13 genes per genome, median = 4). The abundance of genes encoding MCP-domain proteins may result from inhabiting spatially and/or temporally variable sediment and hydrothermal vent communities; organisms inhabiting such environments tend to carry more MCP genes (Lacal *et al.*, 2010). The absence of such genes in *Tms. cyclica* is striking, and it also lacks genes encoding motility-related signalling and structures (*che* and *fli* operons), consistent with this species' nonmotile phenotype (Sorokin *et al.*, 2002). Perhaps nonmotility reflects the relative stability of the stratified hypersaline lake from which this species was isolated.

MCP genes among the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* fall into 28 clusters (Fig. 4). F4 Their predicted amino termini, which include their sensory domains, likely responsive to a variety of ligands, are less conserved than their carboxy termini, as has been noted

ID: nagarajulum Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

J_ID: EMI Customer A_ID: EMI14090 Cadmus Art: EMI14090 Ed. Ref. No.: EMI-2018-0114.R1 Date: 9-March-18 Stage: Page: 9

Table 3. Electron transport-related components encoded in genomes from the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*.

Taxon	SDH[a,b]	Sqr types	Sox	[NiFe]	[FeFe]	NDH	bc_1	cbb_3	*bd*	ba_3
H. crunogenus XCL-2		A, F	ABCDXYZ	1b		I	1	1		
Hydrogenovibrio sp. XS5	1	A, F	ABCDXYZ			I, II	1	1	1	
H. thermophilus JR2	1	A, F	ABCDXYZ			I, II	1	1	1	
H. thermophilus MA2–6	1	A, F	ABCDXYZ	1d, 2b		I, II	1	1	1	
H. halophilus	3	A, F	ABCDXYZ		1	I	1	1		
Hydrogenovibrio sp. WB1	3	A, F	ABCDXYZ		1	I	1	1		
H. kuenenii	1	A, F	ABCDXYZ		1	I, II	1	1		
H. marinus DSM 11271	1	A, F	ABCDXYZ	1d, 2b	1	I, II	1	1		
Hydrogenovibrio sp. Milos-T1	1	A, E, F	ABCDXYZ			I, II	1	1	1	
Tmr. frisia Kp2	4	A, F	ABCDXYZ		1	I	1	1		
Thiomicrorhabdus sp. Milos-T2	3	A, F	ABCDXYZ			I	1	1		1
Tmr. arctica		A, F	ABCDXYZ			I	1	1		
Tmr. chilensis	4	A, F	ABCDXYZ		1	I	1	1		
Tms. aerophila	3	B, F	ABCDXYZ			I	1	1		
Tms. cyclica	1	B, F	ABCDXYZ			I	1	1		
Thiomicrospira sp. ALE5	3	B, F	ABCDXYZ			I	1	1		
Tms. microaerophila	4	B, E, F	ABCDXYZ	1b		I	1	1		
Tms. pelophila	2	A, B, F	ABCDXYZ		1	I	1	1		

a. Abbreviations: SDH, flavocytochrome *c* sulfide dehydrogenase; Sqr types, sulfide:quinone oxidoreductases A, B, C, D, E or F; DSR, dissimilatory sulfite reductase system; Sox, Sox/thiosulfate-oxidizing multi-enzyme system, ABXYZ (incomplete) or ABCDXYZ (complete); Sgp, sulfur globule proteins A/B or C; Sor, sulfite dehydrogenase; Soe, sulfite oxidizing enzyme SoeABC; APS red, adenosine phosphosulfate reductase; [NiFe], group 1b, 1d or 2b [NiFe] hydrogenase; [FeFe], [FeFe] hydrogenase; NDH, type I and II NADH:quinone oxidoreductase; bc_1, cytochrome bc_1 complex; cbb_3, cbb_3-type cytochrome *c* oxidzse; *bd*, cytochrome *bd*-type quinol oxidase; ba_3, ba_3-type cytochrome *c* oxidase.
b. Numbers indicate the number of copies of a particular complex that are encoded by the genome.

for MCPs in general (Wuichet *et al*., 2007). Many of the MCP genes carried by these taxa encode amino-terminal PAS domains that might be involved in energy taxis, in which the redox state of components of the electron transport chain is sensed by redox-sensitive MCP proteins, which results in motility to microenvironments with an optimal concentration of reductant and oxidant. Organisms capable of energy taxis often have large numbers of MCP genes (Alexandre *et al*., 2004), and this is the case for the genera studied here. MCP genes with PAS domains comprise the largest of the 28 clusters (MCP-PAS8/PAS9-CZB; Fig. 4). This cluster includes multiple genes from many of these genomes. Phylogenetic analysis of these genes does not cluster them by species, suggesting duplication prior to divergence of these species (data not shown). Eighteen of the clusters, including this large PAS domain cluster, contain genes from genus *Thiomicrospira* and either *Hydrogenovibrio* or *Thiomicrorhabdus*, and likely reflect the presence of these genes in the shared ancestor of these three genera (Fig. 4). Ten clusters present in the genera *Hydrogenovibrio* and *Thiomicrorhabdus* do not have full-length matches within genus *Thiomicrospira*. Five of these clusters have full-length matches elsewhere in the *Gammaproteobacteria* and appear to have been acquired by the lineage leading to the genera *Hydrogenovibrio* and *Thiomicrorhabdus* subsequent to its divergence from genus *Thiomicrospira*. The other five clusters of genes absent from genus *Thiomicrospira* have many high-similarity partial-length hits that fell within the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*. Based on these observations, no genes appear to have been horizontally transferred from distantly related organisms. Perhaps this is because the MCP domain needs to be sufficiently conserved to communicate with the other components of the chemotaxis apparatus (e.g., CheW, CheA; Wadhams and Armitage, 2004).

Electron transport chains

Similar to other chemolithoautotrophs (e.g., Dmytrenko *et al*., 2014; Flood *et al*., 2016), genomes from members of the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* encode a variety of complexes to introduce electrons stripped from inorganic compounds into the electron transport chain, and to act as terminal oxidases. Given that several genomes from within this clade of microorganisms have been sequenced, they provide a unique opportunity for illuminating how these complexes were acquired. All 18 of the species examined carry the genetic potential to oxidize reduced sulfur compounds via sulfide:quinone oxidoreductases (EC 1.8.5.4; subdivided into clades sqrA–sqrF and sqrX; Gregersen *et al*., 2011). All carry *sqrF*, while *sqrA, sqrB* and *sqrE* are distributed less evenly among taxa (classified by comparison to sequences in Gregersen *et al*., 2011; Table 3). As a result, several taxa carry multiple sulfide:quinone oxidoreductase genes, and this is also the case for other members of *'Proteobacteria'* (Supporting Information Table S3). In

ID: nagarajulum Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

Chlorobaculum tepidum, genes encoding multiple forms of this enzyme are differentially induced by sulfide concentrations (Chan *et al.*, 2009); this may also be the case for these organisms. Most of the members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* also carry one or multiple copies of flavocytochrome *c* sulfide dehydrogenase (SDH; EC 1.8.1.19; Table 3). These genomes encode a particularly large number of copies of this enzyme (up to 4); other members of *'Proteobacteria'* typically carry fewer (1–2; Supporting Information Table S3). Expression of these paralogs may be differentially induced, as has been described above for *sqr* genes. Members of the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* carry genes encoding the complete thiosulfate-oxidizing multienzyme system ('Sox complex', encoded by *soxABCDXYZ*) associated with the Kelly–Friedrich pathway, which can oxidize thiosulfate completely to sulfate without the formation of any free intermediates in the cell or locale. These genes are not collocated in a single operon, similar to *H. crunogenus* (Scott *et al.*, 2006), suggesting that they may be differentially regulated. Most of the the *'Proteobacteria'* surveyed here carry *soxABXYZ* instead (Supporting Information Table S3); this 'incomplete' Sox complex may produce polysulfide or elemental sulfur (Frigaard and Dahl, 2009). Given the possibility for differential regulation of *sox* genes in *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*, their Sox complexes may not always include SoxCD proteins, or could use alternative cytochromes to fulfill their role. The presence of Sqr and SDH genes (and the potential to synthesize SoxABXYZ systems) is consistent with elemental sulfur production by these cells (Jannasch *et al.*, 1985; Brinkhoff *et al.*, 1999a,c; Takai *et al.*, 2004; Knittel *et al.*, 2005). Genes encoding sulfur globule proteins homologous to those present in *Allochromatium vinosum* are absent, which may indicate that other proteins encase sulfur globules, as has been observed in (Hanson *et al.*, 2016), or that the globules do not have protein coats.

Known pathways for sulfite metabolism are absent from the genomes of members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*, which distinguishes them from other *'Proteobacteria'* (Supporting Information Table S3) and is consistent with an inability to use sulfite as their sole electron donor (Brinkhoff *et al.*, 1999a,c; Rainey *et al.*, 2001; Sorokin *et al.*, 2002; Takai *et al.*, 2004; Knittel *et al.*, 2005). As has been noted for *H. crunogenus* (Scott *et al.*, 2006), members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* lack dissimilatory sulfite reductase (DSR); this enzyme is common in photosynthetic sulfur oxidizing bacteria, which use it to oxidize sulfide and elemental sulfur to sulfite (Frigaard and Dahl, 2009). This complex is also encoded in the genomes of many chemolithoautotrophic members of *'Proteobacteria'* (Supporting Information Table S3). Its absence in *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* is consistent with previous studies which noted that the DSR complex is absent from organisms carrying genes encoding a complete Sox complex that includes SoxCD (Frigaard and Dahl, 2009; Gregersen *et al.*, 2011). Genes encoding enzymes to oxidize sulfite directly (SorAB, SoeABC) or indirectly (dissimilatory APS reductase) are also absent from their genomes, though these genes are common among other *'Proteobacteria'* (Supporting Information Table S3).

The presence of genes encoding the Sox complex and sulfide:quinone oxidoreductase in all sequenced members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* suggests that their common ancestor was a sulfur-oxidizing chemolithotroph. All of them also carry genes encoding NDH-1-type NADH:quinone oxidoreductase (NADH dehydrogenase, EC 1.6.5.3), cytochrome bc_1 complex (EC 1.10.2.2) and a cbb_3-type cytochrome c oxidase (EC 1.9.3.1), which facilitates growth under low oxygen conditions (Pitcher and Watmough, 2004) and has been purified and characterized from *Tms. aerophila* (Sorokin *et al.*, 1999). The presence of a high-affinity cbb_3 cytochrome *c* oxidase in all species examined suggests that the common ancestor to these organisms was capable of growth under low oxygen conditions as well.

Other components of electron transport chains are not universally encoded by the genomes of these organisms. Genes encoding hydrogenases are unevenly distributed among *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*. Among these are genes encoding four types of hydrogenase (EC 1.12.1._; [FeFe], as well as [NiFe] groups 1b, 1d and 2b) (Hansen and Perner, 2016). [FeFe] hydrogenases are present in some of the taxa sequenced here (Table 3), though absent from other members of the *'Proteobacteria'* that were examined (Supporting Information Table S4). Genome contexts of these [FeFe]-hydrogenase genes do not suggest a function for this hydrogenase, or a direction to the catalytic activity (i.e., either producing or consuming H_2). Group 1b [NiFe]-hydrogenase genes are present in *H. crunogenus* (Scott *et al.*, 2006) and *Tms. microaerophila* ASL8-2. Matches to these genes largely fall within genomes of members of the phylum *Epsilonbacteraeota*, and group 1b hydrogenases are absent from the other *Gammaproteobacteria* queried here (Supporting Information Table S4). In *Tms. microaerophila* ASL8-2, as has been described in *H. crunogenus*, the small-subunit-encoding gene includes an NADH-binding domain; this, plus the absence of a *b*-type subunit-encoding gene, suggests that the redox partner for this enzyme could be cytoplasmic (Scott *et al.*, 2006). Hydrogenases using NADH as the redox partner are often reversible (Lubitz *et al.*, 2014), so the enzyme encoded by these genes may function either to use H_2 as an electron donor, or to produce this gas to maintain redox balance,

J_ID: EMI Customer A_ID: EMI14090 Cadmus Art: EMI14090 Ed. Ref. No.: EMI-2018-0114.R1 Date: 9-March-18 Stage: Page: 11

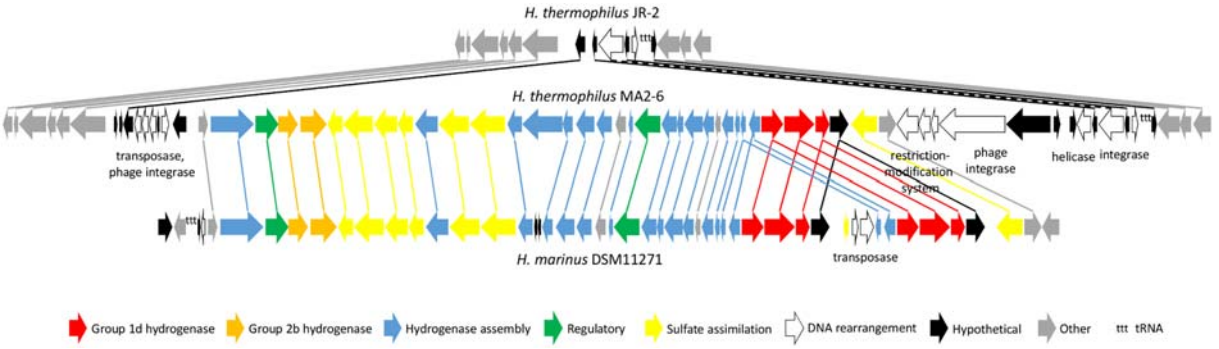


Fig. 5. Genome context and synteny of genes encoding and associated with hydrogenase enzymes in *H. thermophilus* MA2–6 and *H. marinus*. A region of the genome from *H. thermophilus* JR-2, which lacks hydrogenase, is provided for comparison, as it shows the likely site of insertion for this hydrogenase genomic island subsequent to the divergence of these two strains of *H. thermophilus*. Lines link orthologous genes.

which would be consistent with life in a highly variable environment. Group 1b [NiFe]-hydrogenase genes are also present in *H. crunogenus* TH55. Molecular hydrogen use by this organism is suggested by hydrogen consumption from culture headspace, and increased cell counts in the presence of this gas, demonstrating chemolithoautotrophic growth on molecular hydrogen (Hansen and Perner, 2015; 2016). Hydrogen use by *H. crunogenus* XCL-2 was not detected under these or a variety of other cultivation conditions, and there are no mutations in its hydrogenase or hydrogenase-associated genes that explain this lack of activity (Hansen and Perner, 2016). The function of the group 1b [NiFe]-hydrogenase remains mysterious in *H. crunogenus* XCL-2.

Genes encoding group 1d (oxygen-tolerant) and 2b (sensory) hydrogenases are present in both *H. marinus* and *H. thermophilus* MA2–6, and both are capable of growth with hydrogen gas as the electron donor (Hansen and Perner, 2016). Genes encoding group 1d hydrogenase are common among the members of *Gammaproteobacteria* (Supporting Information Table S4). For both *H. marinus* and *H. thermophilus* MA2–6, genes nearby encode enzymes for assimilatory sulfate reduction (subunits 1 and 2 of sulfate adenylyltransferase, thioredoxin-dependent adenylylsulfate APS reductase; Bick *et al*., 2000); perhaps this collocation assists with the synthesis of iron sulfur clusters for hydrogenase enzymes. Together, these genes liberate *H. marinus* and *H. thermophilus* MA2–6 from any requirement for reduced sulfur compounds either as electron donors or biosynthetic substrates; indeed, both strains grow on hydrogen gas in the absence of reduced sulfur sources (Nishihara *et al*., 1991, K. Scott, unpublished). This ability could be an advantage as some vent habitats are hydrogen-rich and sulfide poor (Kelley *et al*., 2005).

The chromosome regions encoding the 1d and 2b hydrogenases in *H. marinus* and *H. thermophilus* MA2–6 are strikingly similar; the order of the genes encoding these hydrogenases and the many proteins needed to assemble them are nearly identical (Fig. 5), despite a lack of conservation of gene order in other genomes encoding these hydrogenases. The predicted amino acid sequences of the large and small subunits of these hydrogenases share high sequence identities between these two species (e.g., for 1d, 92% and 96% identical respectively). Given the absence of group 2b and 1d hydrogenases in the other members of genus *Hydrogenovibrio*, including species closely related to *H. marinus* and *thermophilus* MA2–6, it seems likely that these two taxa acquired these genes independently via horizontal gene transfer. Given the similarity in gene order and sequence, they both likely acquired this hydrogenase gene cluster from similar organisms. Phylogenetic analysis of the group 2b and 1d hydrogenase gene sequences from genus *Hydrogenovibrio* places them together on a long branch (data not shown), precluding inference about the taxonomic affiliation of the donor organisms. For *H. thermophilus* MA2–6, there are many signs that acquisition of these genes was a relatively recent phage-mediated acquisition, as the hydrogenase gene cluster is surrounded by phage components (Fig. 5). *H. thermophilus* JR2, whose genome is otherwise syntenous with *H. thermophilus* MA2–6, lacks both these phage components and the gene cluster, suggesting hydrogenase acquisition by *H. thermophilus* MA2–6 subsequent to the *H. thermophilus* MA2–6/JR2 divergence. Hydrogenase acquisition by horizontal gene transfer may be widespread (Greening *et al*., 2016). *Tmr. hydrogeniphilia*, a recently isolated member of the *genus Thiomicrorhabdus*, grows chemolithoautotrophically on molecular hydrogen. *Tmr. hydrogeniphila* is very closely related to *Tmr. frisia*, and is the only member thus far of its genus to have this capability demonstrated (Watsuji *et al*., 2016; Boden *et al*., 2017a). The ability to grow with hydrogen as the electron donor is also observed in some species of

ID: nagarajulum Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

J_ID: EMI Customer A_ID: EMI14090 Cadmus Art: EMI14090 Ed. Ref. No.: EMI-2018-0114.R1 Date: 9-March-18 Stage: Page: 12

Acidithiobacillus; the distribution of this trait does not suggest vertical transmission from the ancestral *Acidithiobacillus* (Hedrich and Johnson, 2013).

The recent isolation of an organism falling within genus *Hydrogenovibrio* that can grow with ferrous iron as its electron donor (Barco *et al*., 2017) raises the possibility that the organisms sequenced here might also have this capability. However, their genomes do not include homologs of the outer membrane proteins canonically associated with iron oxidization or reduction [Cyc2 (and associated alternative complex III), MtoA, OmpB or OmcB; Liu *et al*., 2012; 2014; Barco *et al*., 2015; Kato *et al*., 2015]. Despite this absence of genetic evidence, it would be worthwhile to test these organisms for the ability to use ferrous iron as an electron donor, since mechanisms for iron oxidation have yet to be described for many organisms with this capability (He *et al*., 2017).

In addition to this diversity in mechanisms for input of reductant into the electron transport chains of these organisms, there is diversity in electron transport complexes as well (Table 3). All of these organisms carry genes encoding complex I (NADH:ubiquinone reductase, EC 1.6.5.3). Some members of *Hydrogenovibrio* carry genes encoding alternative (type II) NADH:ubiquinone reductase (EC 1.6.5.9), which is common among *'Proteobacteria'* (Marreiros *et al*., 2016), though not among the autotrophic organisms surveyed here (Supporting Information Table S4). Type II NADH:ubiquinone reductase oxidizes NADH without directly contributing to proton potential, which may function to maintain cellular NADH/NAD^+ ratios (Kerscher *et al*., 2008). In addition to the cbb_3 cytochrome c oxidase complex noted above, some members of *Hydrogenovibrio* also carry genes encoding cytochrome *bd*-type quinol oxidase (EC 1.10.3.10), while *Thiomicrorhabdus* sp. Milos T2 has ba_3-type cytochrome *c* oxidase (EC 1.10.3.11; Table 3). Biochemically characterized cytochrome *bd*-type quinol oxidases from other organisms, as well as cbb_3 cytochrome *c* oxidase, have high affinities for oxygen and facilitate growth under low-oxygen conditions (Pitcher and Watmough, 2004; Borisov *et al*., 2011), which would be helpful in many of the habitats where these genera are found. Indeed, both of these complexes are quite common among autotrophic *'Proteobacteria'*, which should facilitate growth in low-oxygen habitats where reduced inorganic electron donors are present (Supporting Information Table S4). Alternatively, the cytochrome *bd*-type oxidase and ba_3 complexes may provide mechanisms for redox balance. The ba_3 cytochrome *c* oxidase from '*Aquifex aeolicus*' uses both quinol and ferrocytochrome *c* as electron donors. Quinol oxidation by this complex may function to maintain redox balance when the quinone pool becomes excessively reduced (Gao *et al*., 2012). It has also been suggested that the *bd* complex may function to minimize oxidative damage when oxygen tensions rise (Poole and Hill, 1997; Ramel *et al*., 2015).

Much of the diversity in electron transport chain components is present in the genera *Hydrogenovibrio* and *Thiomicrorhabdus*, but not in genus *Thiomicrospira* (Table 3). The sequenced members of genus *Thiomicrospira* were all isolated from sediment habitats and hypersaline and soda lakes, while many of the sequenced members of the genera *Hydrogenovibrio* and *Thiomicrorhabdus* were isolated from hydrothermal vent environments (Table 1). The additional electron transport chain components found in these organisms could function to maintain electron transport activity when oxygen tensions or sulfide concentrations vary; perhaps they lend these organisms a degree of 'redox versatility' that provides an advantage in the extreme spatial and temporal heterogeneity of vent environments.

Carboxylases, carboxysomes and associated transporters

All of these organisms carry genes encoding the Calvin–Benson–Bassham (CBB) cycle; none carry genes encoding ATP-dependent citrate lyase (EC 2.3.3.8), or citryl-coA synthetase (EC 6.2.1.18)/citryl-coA ligase (EC 4.1.3.34), which are present in organisms using the reductive citric acid (Arnon-Buchanan) cycle for carbon fixation (Aoshima *et al*., 2004a,b; Hügler *et al*., 2007). Absence of a gene encoding sedoheptulose-bisphosphatase (EC 3.1.3.37) suggests that these organisms either have a bifunctional bisphosphatase, capable of acting both on fructose 1,6-bisphosphate as well as on sedoheptulose 1,7-bisphosphate (Yoo and Bowien, 1995), or that they use the transaldolase variant of the CBB cycle (Strøm *et al*., 1974; Boden *et al*., 2017b). Other organisms lacking these enzymes have been found to carry pyrophosphate-dependent phosphofructokinase (EC 2.7.1.90), which is reversible and can act on sedoheptulose 1,7-bisphosphate (Reshetnikov *et al*., 2008). This does not appear to be the case for the organisms studied here. Genes encoding pyrophosphate-dependent phosphofructokinase are often adjacent on the chromosome to genes encoding proton translocating pyrophosphate synthase (Kleiner *et al*., 2012); these genes are absent from the organisms studied here. It was originally suggested that enzyme from *H. crunogenus* might use pyrophosphate (Scott *et al*., 2006); however, amino acid sequences predicted from phosphofructokinase genes from this organism, as well as other members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* are consistent with ATP as a substrate; they have a glycine residue at position 104, and a lysine at 124 (positions relative to ATP-dependent enzyme from *Escherichia coli*), as do biochemically characterized ATP-dependent enzymes (Bapteste *et al*., 2003).

ID: nagarajulum Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

Table 4. Presence of genes encoding RubisCO, and enzymes for acetate assimilation and metabolism in the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*.

Taxon	Form IAc[a,b]	Form IAq	Form II	AK[c]	PTA	OFOR
H. crunogenus XCL-2	1	1	1			1
Hydrogenovibrio sp. XS-5	1	1	1	1	1	1
H. thermophilus JR2	1	1	1	1	1	1
H. thermophilus MA2–6	1	1	1	1	1	1
H. halophilus	1	1	1	1	1	1
Hydrogenovibrio sp. WB-1	1	1		1	1	1
H. kuenenii	1	1	1	1	1	
H. marinus	1	1	1	1	1	
Hydrogenovibrio sp. Milos-T1	1		1	1	1	
Tmr. frisia KP2	1	1	1			
Thiomicrorhabdus sp. Milos-T2		1	1			
Tmr. arctica		1	1			
Tmr. chilensis	1	1	1			
Tms. aerophila	1					
Tms. cyclica	1					
Thiomicrospira sp. ALE5	1					
Tms. microaerophila	1					
Tms. pelophila	1		1			

a. Form IAc = form IA carboxysomal RubisCO; Form IAq = form IA noncarboxysomal RubisCO; form II = form II RubisCO.
b. Numbers indicate the number of copies of a particular enzyme that are encoded by the genome.
c. AK, acetate kinase; PTA, phosphotransacetylase; OFOR, oxoacid:ferredoxin oxidoreductase.

Many of these organisms' genomes encode multiple RubisCO enzymes (form IAc, carboxysomal form IA RubisCO; form IAq, noncarboxysomal form IA RubisCO; form II; Table 4; Supporting Information Fig. S2; Boden *et al*., 2017b). Two, and often three, genes encoding RubisCO are present in genomes from members of *Hydrogenovibrio* and *Thiomicrorhabdus* (Table 4); it is more common among *'Proteobacteria'* for genomes to include one or two of these genes (Supporting Information Table S5). RubisCO enzymes have very large differences in K_{CO2} values and specificity for CO_2 versus O_2; noncarboxysomal form I enzymes typically have higher affinities for CO_2, as well as being more specific for CO_2 (Badger and Bek, 2008), while carboxysomal form I enzymes are packed into carboxysomes as part of CCMs (Menon *et al*., 2008). Encoding more versions of this enzyme may facilitate autotrophic growth in the presence of a broader range of CO_2 concentrations. For *H. marinus* and *H. crunogenus*, it has been demonstrated that these RubisCO genes are differentially transcribed in response to the concentration of CO_2 in the growth medium (Yoshizawa *et al*., 2004; Dobrinski *et al*., 2012), and having multiple enzymes is believed to facilitate stable rates of carbon fixation at a broad range of CO_2 and O_2 concentrations (Tabita, 1999; Badger and Bek, 2008).

Most genomes encode carboxysomal form IAc enzyme, consistent with the presence of carboxysomes and CO_2 concentrating mechanisms similar to what has been described in members of the '*Cyanobacteria'* as well as *Hydrogenovibrio crunogenus* (Dobrinski *et al*., 2005; Price *et al*., 2008; Boden *et al*., 2017b). Carboxysome presence has been previously verified via electron micrographs of *H. crunogenus* (Dobrinski *et al*., 2012; Menning *et al*., 2016), *H. marinus* (Yoshizawa *et al*., 2004), *Tms. aerophila* (Rainey *et al*., 2001) and *Tms. cyclica* (Sorokin *et al*., 2002). However, carboxysomes and CCMs are not inevitable among these genera; *Thiomicrorhabdus* sp. Milos-T2 and *Tmr. arctica* lack carboxysomal RubisCO genes, as well as genes encoding shell proteins and other components of carboxysomes. It is possible that these genes are encoded on portions of the genomes that remain to be sequenced. Alternatively, these taxa may have ~~alternative~~ mechanisms ~~beyond CCMs~~ for coping with periods of low CO_2 availability.

Genomes of most members of *Thiomicrospira* encode only carboxysomal RubisCO (form IAc) and not form II (Table 4), confirming results from PCR-based assays for these enzymes (Tourova *et al*., 2006). Most of the members of *Thiomicrospira* studied here were isolated from alkaline hypersaline and soda lakes (Sorokin *et al*., 2002; 2007; 2011). These habitats typically have high concentrations of bicarbonate and carbonate, but their high pH (9–10) likely renders CO_2 scarce (Sorokin *et al*., 2011). The hydrothermal vent or sediment habitats from which all of the other organisms in this study were isolated can have quite high CO_2 concentrations (e.g., Goffredi *et al*., 1997), which would make it an advantage for these organisms to be able to repress carboxysome synthesis. In hypersaline and soda lakes, noncarboxysomal form I RubisCO, or form II RubisCO, which typically has a low affinity for CO_2,

J_ID: EMI Customer A_ID: EMI14090 Cadmus Art: EMI14090 Ed. Ref. No.: EMI-2018-0114.R1 Date: 9-March-18 Stage: Page: 14

COLOR IN ONLINE AND PRINT

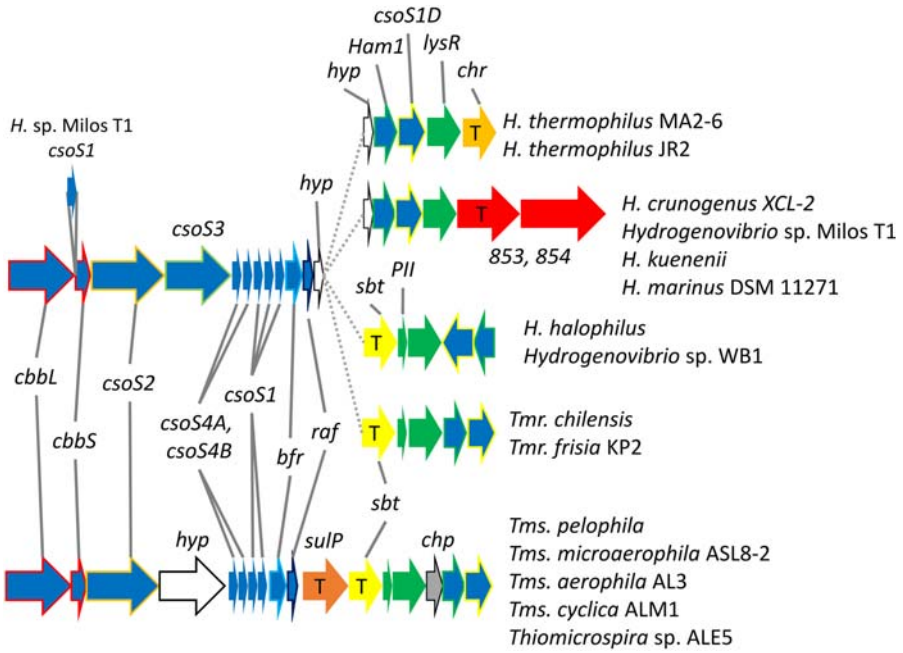


Fig. 6. Carboxysome operon synteny among members of the genera *Hydrogenovibrio, Thiomicrorhabdus*, and *Thiomicrospira*. 'T' marks genes that encode potential transporters. Gene abbreviations are: *bfr*—bacterioferritin; *cbbL*—RubisCO large subunit; *cbbS*—RibisCO small subunit; *chp*—conserved hypothetical protein; *chr*—chromate ion transporter family; *csoS1, csoS2, csos4A, csoS4b*—carboxysome shell proteins; *csoS3*—carboxysome carbonic anhydrase; *Ham1*—Ham1-domain protein; *hyp*—hypothetical protein; *lysR*—LysR-type transcriptional regulator; *PII*—PII signal transduction protein; *raf*—RubisCO assembly factor; *sbt*—sodium-dependent bicarbonate transporter family; *sulP*—sulfate transporter family; *853, 854*—homologs to *Tcr_0853, 0854*.

would function poorly and may explain the absence of these forms of the enzyme among the members of *Thiomicrospira* isolated from these habitats. Consistent with this, *Thiomicrospira pelophila*, the sole sequenced member of its genus to encode form II RubisCO, was isolated from coastal sediment, which is likely to have higher CO_2 concentrations, and grows optimally near neutral pH (Kuenen and Veldkamp, 1972).

Carboxysome operons among these organisms differ from those typically present in other chemolithoautotrophic bacteria (Axen *et al*., 2014) in that they lack *cbbO* and *cbbQ*, encoding RubisCO accessory proteins, as well as homologs to *parA,* which may play a role in carboxysome positioning in other organisms. Among the organisms studied here, carboxysome operon structure is generally well conserved (Fig. 6). An intriguing deviation is that *Tms. pelophila* and other members of *Thiomicrospira* lack *csoS3*, encoding a carboxysomal carbonic anhydrase (Cannon *et al*., 2010). In its place are genes that are unique to these taxa. *Tms. pelophila* gene *N746DRAFT_0321* and *Tms. microaerophila* ASL8-2 gene *NA59DRAFT_00206* have no homologs in IMG or NCBI non-redundant protein databases. *Thicy_1562, Thiae_1973* and *MZ34DRAFT_1177*, which are the corresponding genes in *Tms. cyclica* ALM1, *Tms. aerophila* AL3 and *Thiomicrospira* sp. ALE5, are homologous to each other. Given that carbonic anhydrase activity is key to the functioning of carboxysomes in other organisms (Cannon *et al*., 2010), its potential absence in these taxa is puzzling. One possibility is that the carboxysomal carbonic anhydrase for these organisms is encoded elsewhere on their chromosomes. Indeed, some of them do carry alpha carbonic anhydrase genes (*Thiae_1542, MZ34DRAFT_1567, N746DRAFT_1499, NA59DRAFT_02538*). Alternatively, the novel genes present in the carboxysome operons of these organisms may encode novel carbonic anhydrase enzymes. A third possibility is that these carboxysomes function in the absence of carbonic anhydrase activity, which would be particularly surprising, given the presence of this enzyme in all other carboxysomes characterized to date (Cannon *et al*., 2010).

Potential transporters are encoded 3′ to the carboxysome operons and may play a role in dissolved inorganic carbon uptake. These genes are quite diverse, encoding proteins that are members of four evolutionarily distinct transporter families (Saier *et al*., 2014): major facilitator superfamily SulP/BicA (MFS), HCO_3^-:Na^+ symporter (SbtA), chromate ion transporter (CHR). The fourth consists of a two-component DIC transporter in which one subunit is a homolog of the NADH dehydrogenase chain L (NDL). Members of three of these families have been implicated in DIC uptake: two in the '*Cyanobacteria*' [BicA (Price *et al*., 2004), SbtA (Shibata *et al*., 2002)], and one in *H. crunogenus* [encoded by *Tcr_0853* and *Tcr_0854;* Mangiapia *et al*. (2017)]. Some species carry representatives of several of these transporter families either downstream from the carboxysome operon or elsewhere on the chromosome (Fig. 6). Since the carboxysome operon falls at the end of a genome scaffold for *Hydrogenovibrio* sp. XS5, it was not possible to determine which, if any, potential transporter genes might be present 3′ to this locus.

If indeed genes from all four of these transporter families encode DIC transporters, it is interesting that such a

ID: nagarajulum Time: 19:08 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

diversity of transporters exists among these organisms. It can be anticipated that transporters from different families will differ from each other with respect to maximal rates and affinities, and in symported/antiported compounds and stoichiometries, providing advantages under certain growth conditions. They are particularly abundant and diverse among members of genus *Thiomicrospira*. Perhaps the added transporters provide a selective advantage in the extremely low CO_2 alkaline habitats from which most of the members of *Thiomicrospira* were isolated (Sorokin *et al*., 2011), or compensate for a lack of carboxysomal carbonic anhydrase in these species.

Central carbon metabolism

All of the target genomes carry genes encoding the same version of Embden–Meyerhof–Parnas glycolysis/gluconeogenesis and the citric acid cycle described for *H. crunogenus*, lacking all NADH-specific dehydrogenases (isocitrate EC 1.1.1.41, 2-oxoglutarate EC 1.2.4.2/2.3.1.61/1.8.1.4 and malate EC 1.1.3.7) except pyruvate dehydrogenase (EC 1.2.4.1/2.3.1.12/1.8.1.4; Quasem *et al*., 2017). Instead, they carry genes less common among members of the *'Proteobacteria'*:monomeric $NADP^+$ isocitrate dehydrogenase (EC 1.1.1.42; (Yasutake *et al*., 2002), and malate:quinone oxidoreductase (E.C. 1.1.5.4; Quasem *et al*., 2017). It is possible that they have incomplete citric acid cycles (Smith's horseshoe), as is often the case for obligate autotrophs (e.g., Boden *et al*., 2016; Hutt *et al*., 2017) that function primarily to provide biosynthetic intermediates (Smith *et al*., 1967; Wood *et al*., 2004). Alternatively, all of the genomes studied here encode homologs of the genes of the 2-oxoglutarate decarboxylase/succinic semialdehyde dehydrogenase bypass present in some members of the '*Cyanobacteria*' (Zhang and Bryant, 2011), so it is possible that they may be capable of catalysing a complete oxidative citric acid cycle.

Some of these organisms are capable of assimilating exogenous organic carbon (Table 1), and their genomes provide possible mechanisms for doing so. *H. thermophilus* I78 and *H. crunogenus* TH-55 and L-12 (Takai *et al*., 2004) can assimilate acetate and other organic compounds when inorganic electron donors are provided, as can *Tms. pelophila* and *Tms. aerophila* (Kuenen and Veldkamp, 1972; 1973; Rainey *et al*., 2001; Ang *et al*., 2017). All organisms sequenced here have genes encoding acetyl-CoA synthetase (EC 6.2.1.1), which could facilitate assimilation of acetate; some members of *Hydrogenovibrio* also have genes encoding acetate kinase (EC 2.7.2.1 or 2.7.2.12) and phosphotransacetylase (EC 2.3.1.8; Table 4). These organisms lack the genes necessary for the glyoxylate cycle, which, in the absence of a shunt between 2-oxoglutarate and succinyl-CoA would restrict the distribution of the acetate carbons to fatty acid biosynthesis and amino acids derived from 2-oxoglutarate (Wood *et al*., 2004). However, these organisms also all carry oxoacid:-ferredoxin oxidoreductase genes similar to those from *H. crunogenus*. *H. crunogenus* cell extracts have pyruvate:ferredoxin oxidoreductase activity (EC 1.2.7.10; Quasem *et al*., 2017), suggesting that its oxidoreductase genes encode pyruvate: ferredoxin oxidoreductase. Given the amino acid sequence similarities of orthologs from other members of *Hydrogenovibrio* (65%–77%, alpha subunits; 74%–78%, beta subunits), their oxidoreductases may also act on pyruvate. If the acetyl-CoA were converted to pyruvate by this enzyme, it would circumvent the need for a glyoxylate cycle.

Extremophile lifestyles

Two taxa sequenced here are extremophiles (Table 1), capable of growth at high salinity (*H. halophilus*; up to 3.5 M NaCl, optimum 1.5 M; Sorokin *et al*., 2006) or low temperature [*Tmr. arctica;* down to −2°C, optimum 11°C–13°C; (Knittel *et al*., 2005)]. Many halophiles and psychrophiles adapt to these conditions by accumulating compatible solutes such as ectoine, glycine betaine, proline or trehalose (Sleator and Hill, 2002; De Maayer *et al*., 2014). Genome evidence suggests that the ability to synthesize or transport of these molecules is a trait shared by all of the organisms studied here, and not specific to the two extremophiles. All sequenced members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* have genes encoding the enzymes necessary to synthesize ectoine (diaminobutyrate-2-oxoglutarate transaminase, EC2.6.1.76; diaminobutanoate acetyltransferase, EC 2.3.1.178 and ectoine synthase, 4.2.1.108). This observation is consistent with ectoine synthesis by *Thiomicrospira aerophila* when cultivated at elevated concentrations of Na^+ (Banciu *et al*., 2005). None carry genes encoding the enzymes necessary to synthesize glycine betaine or trehalose; instead, the genomes of *Tmr. arctica* and *Tmr.* Milos T2 carry genes encoding ABC transporters predicted to transport proline or glycine betaine (*F612DRAFT_1896–F612DRAFT_1898, BS34DRAFT_2245–BS34DRAFT_2247*), and all genomes studied here encode members of the BCCT betaine/carnitine/choline transporter family (Pfam02028); some members of this family transport glycine betaine (Ziegler *et al*., 2010). Neither *Tmr. arctica* nor *H. halophilus* appear to have acquired or amplified compatible solute-related genes to facilitate their psychrophilic or halophilic lifestyles.

The *Tmr. arctica* genome does have some aspects that may be adaptive to psychrophily. Growth at low temperature requires that proteins are especially flexible (Feller, 2013), making them particularly sensitive to denaturing. To facilitate correct folding of these proteins, particularly at warmer temperatures, some psychrophiles elevate expression of prolyl isomerase and molecular chaperones

(Williams *et al.*, 2011); some also carry added *dnaJ* genes (Riley *et al.*, 2008), presumably as added protection against denatured protein aggregation and facilitate refolding (Han and Christen, 2004). The *Tmr. arctica* genome carries slightly elevated numbers of both sorts of genes relative to other members of its genus. *Tmr. arctica* carries 14 genes belonging to prolyl isomerase PFAMs 00160, 00254, 00639, 13145 and 13616, compared with 12 in *Tmr. chilensis*, 13 in *Tmr. frisia* KP2, and 14 in *Tmr.* sp. Milos T2 and 5 genes encoding DnaJ proteins, compared with 2 in *Tmr.* sp. Milos T2, 4 in *Tmr. chilensis* and 5 in *Tmr. frisia* KP2.

The *Tmr. arctica* genome is unique among all of the organisms sequenced here in carrying a chromosome region encoding the enzymes necessary for extracellular polysaccharide (EPS) synthesis. Other psychrophiles produce EPS when cultivated at the low range of their growth temperatures; EPS also enhances their survival of freezing (Marx *et al.*, 2009). In *Tmr. arctica*, the chromosome region spanning from *F612DRAFT_0075* to *F612DRAFT_0094* includes genes homologous to those involved in EPS production in other organisms (*epsD–I, ExoZ)*, encoding glycosyltransferases, as well as genes encoding molecular machinery necessary to translocate EPS to the extracellular milleau (flippase, beta barrel porin). It is important to note, however, that the production of EPS-like substances has been reported in a nonpsychrophilic sulfur-oxidizing microorganism (Nunoura *et al.*, 2014); further study would be necessary to determine the role of these substances in *Tmr. arctica*.

In other respects, the *Tmr. arctica* genome does not depart from other members of *Thiomicrorhabdus*. When all genes from this organism are translated and amino acid frequencies are tallied, there are no differences in their frequencies, relative to other members of *Thiomicrorhabdus;* the elevated frequencies of lysine or diminished frequencies of proline, acidic or arginine observed in other psychrophiles (Ayala-del-Río *et al.*, 2010) are not apparent in this organism. Additionally, the fatty acid biosynthetic pathways are predicted to be identical in all of these organisms based on genome data, and all encode fatty acid desaturase; any adjustments to maintain membrane fluidity at colder temperatures likely occurs by differential expression of fatty acid biosynthetic genes shared by all members of *Thiomicrorhabdus*, as has been observed for *Shewanella* (Wang *et al.*, 2009).

Conclusions

Members of the genera *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira* all appear to have multiple adaptations to maintain aerobic chemolithoautotrophic metabolism in their heterogeneous environments. Rather than being able to express multiple physiologies (e.g., denitrification, heterotrophy via diverse carbon sources), they appear to have expanded their versatilities as chemolithoautotrophs by sensing their environment with a large arsenal of methyl-accepting chemotaxis proteins and responding to it via chemotaxis, as well as with an elaborate array of proteins that communicate with the second messenger cyclic di-GMP. They carry multiple complexes for introducing electrons removed from inorganic compounds into their electron transport chains, which eventually arrive at multiple terminal oxidases. Multiple RubisCO enzymes are present to facilitate CO_2 fixation at a variety of CO_2 and O_2 tensions, and multiple transporters are likely to facilitate CO_2 ~~or~~ bicarbonate uptake when these substrates are available at a range of concentrations.

Experimental procedures

Genome sequencing, assembly and annotation

Draft genome sequences were generated at the DOE Joint Genome Institute (JGI) and LANL Genome Science Group. The Pacific Biosciences (PacBio) technology (Eid *et al.*, 2009) was used for *H. thermophilus* MA2–6, *Hydrogenovibrio* sp. Milos-T1, *H. kuenenii, H. marinus* DSM 11271 and *Tms. pelophila*. A Pacbio SMRTbell™ library was constructed and sequenced on the PacBio RS platform. The raw reads were assembled using HGAP (version: 2.1.1; Chin *et al.*, 2013). Illumina technology was used to generate draft sequences for *H. halophilus, Tmr. chilensis, Tmr. arctica, Thiomicrorhabdus* sp. Milos-T2 and *Tmr. frisia* Kp2. An Illumina standard shotgun library and long insert mate pair library was constructed and sequenced using the Illumina HiSeq 2000 platform (Bennett, 2004). All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artefacts (L. Mingkun, A. Copeland, J. Han, unpubl.). Filtered Illumina reads were assembled using AllpathsLG (Gnerre *et al.*, 2011). The consensus was computationally shredded into 10 kbp overlapping fake reads (shreds). The Illumina draft data was also assembled with Velvet, version 1.1.05 (Zerbino and Birney, 2008), and the consensus sequences were computationally shredded into 1.5 kbp overlapping fake reads (shreds). The Illumina draft data was assembled again with Velvet using the shreds from the first Velvet assembly to guide the next assembly. The consensus from the second VELVET assembly was shredded into 1.5 kbp overlapping fake reads. The fake reads from the Allpaths assembly and both Velvet assemblies and a subset of the Illumina CLIP paired-end reads were assembled using parallel phrap, version 4.24 (High Performance Software, LLC). Possible mis-assemblies were corrected with manual editing in Consed (Ewing and Green, 1998; Ewing *et al.*, 1998; Gordon *et al.*, 1998). Gap closure was accomplished using repeat resolution software (Wei Gu, unpublished), and sequencing of bridging PCR fragments with Sanger and/or PacBio (Cliff Han, unpublished) technologies. The draft genome of *H. thermophilus* JR-2 was generated using a combination of Illumina (Bennett, 2004) and 454 technologies (Margulies *et al.*, 2005). Illumina libraries were constructed as above. A 454 Titanium standard

library and 2 paired end 454 libraries were also generated. The 454 Titanium standard data and the 454 paired end data were assembled together with Newbler, version 2.3-PreRelease-6/30/2009. The Newbler consensus sequences were computation ally shredded into 2 kb overlapping fake reads (shreds). The 454 Newbler consensus shreds, the Illumina VELVET consensus shreds, and the read pairs in the 454 paired end library were integrated using parallel phrap, version SOS-4.24 (High Performance Software, LLC). Illumina data were used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), or Dupfinisher (Han and Chain, 2006). All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov.

Genes were identified using Prodigal (Hyatt *et al*., 2010), followed by a round of manual curation using GenePRIMP (Pati *et al*., 2010) for finished genomes and draft genomes in fewer than 20 scaffolds. The predicted CDSs were translated and used to search the National Centre for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, KEGG, COG and InterPro databases. The tRNAScanSE tool (Lowe and Eddy, 1997) was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA (Pruesse *et al*., 2007). Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (Nawrocki *et al*., 2009). Additional gene prediction analysis and manual functional annotation was performed within the integrated microbial genomes (IMG) platform (Markowitz *et al*., 2014) developed by the Joint Genome Institute, Walnut Creek, CA (Markowitz *et al*., 2009). Tandem repeats were identified using Tandem repeat finder (Benson, 1999).

Ribosomal protein supertree

Representative taxa from the *Gammaproteobacteria*, the *'Epsilonbacteraeota'* (Waite *et al*., 2017), and the *Acidithiobacillia* were selected from the Integrated Microbial Genomes database (Markowitz *et al*., 2014). Genes encoding ribosomal proteins were gathered based on membership in Clusters of Orthologous Genes (COG) and Protein FAMilies (PFAM; (Tatusov *et al*., 2001; Bateman *et al*., 2002). Large ribosomal subunits 8, 13, 26 and 29 and small subunit 13 were omitted from this analysis as many strains were missing them. Genes encoding the subunits were aligned independently using the MUSCLE algorithm (Edgar, 2004) as implemented in MEGA 4.0 using default settings (Kumar *et al*., 2008), alignments were concatenated via FABOX (http://users-birc.au.dk/biopv/php/fabox/), and refined with GBLOCKS using stringent criteria (Talavera and Castresana, 2007). Phylogenetic trees were constructed in PhyML 3.0 (Guindon *et al*., 2010) using maximum likelihood (ML) analysis of amino acid sequences. The best fit model of evolution, estimated using smart model selection (SMS) in PhyML 3.0 (Guindon *et al*., 2010) was determined to be the amino acid replacement model of (Le and Gascuel, 2008) with four categories using a discrete Gamma distribution ($G = 0.884$) and a proportion of Invariant sites ($I = 0.143$; '$LG + G + I$'). The consensus tree was visualized using FigTree (Version 1.4.3; Rambaut, 2016).

Comparing genome traits of Hydrogenovibrio*,* Thiomicrorhabdus *and* Thiomicrospira *to other autotrophic members of* 'Proteobacteria'

To identify traits that distinguish members of *Hydrogenovibrio, Thiomicrorhabdus* and *Thiomicrospira*, these genomes were compared with those of other members of *'Proteobacteria'* that use the CBB cycle for carbon fixation. This comparison group was collected based on the presence of form I or form II RubisCO genes in their genomes, and demonstrated ability to grow with CO_2 as the major carbon source, as described in Quasem *et al*. (2017). To compare the distribution of electron transport chain components among these organisms, BLAST queries were undertaken, using genes encoding biochemically characterized enzymes (Brune, 1995; Gregersen *et al*., 2011; Dahl *et al*., 2013; Weissgerber *et al*., 2014). BLAST hits were evaluated based on sequence similarities, phylogenetic analysis, and chromosome collocation (Supporting Information Table S6). To compare the distribution and forms of genes encoding RubisCO, all genes belonging to Pfam00016 (Ribulose bisphosphate carboxylase large chain, catalytic domain) were collected from these genomes, and aligned as described above. A preliminary neighbour-joining tree was used to identify 'RubisCO-like proteins', which are not catalytically active as carboxylases (Tabita *et al*., 2008); these sequences were removed from the alignment. The remaining sequences were re-aligned via MUSCLE, the alignment was refined via GBLOCKS, and phylogenetic trees were constructed as described above ($G = 1.027$, $I = 0.07$; '$LG + G + I$').

Methyl-accepting chemotaxis proteins

All genes encoding methyl-accepting chemotaxis proteins were gathered from target genome sequences in IMG. These genes were identified based on the presence of a methyl-accepting chemotaxis protein signalling domain (PFAM 00015), since this portion of these proteins is conserved (Alexander and Zhulin, 2007). Amino acid sequences predicted from genes were clustered via CD-HIT (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi; Li *et al*., 2001). Clusters were generated with 30%–90% sequence identity, and a sequence identity of 40% was chosen for clustering, as this value was high enough to produce clusters with sequences that aligned well over their full length. Representative sequences from each cluster were used as query sequences for BLAST searches of IMG to find homologous genes from other organisms beyond the genera *Thiomicrospira, Hydrogenovibrio* and *Thiomicrorhabdus* whose sequences aligned well along their full length.

Acknowledgements

This research was supported by USDA Higher Education Challenge Grant, MOE-2008–02036 to S.K.F., C.A.K., and K.M.S., and NSF-IOS-1257532 to K.M.S. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office

of Science of the U.S. Department of Energy under Contract No. DE-AC02–05CH11231.

References

Ahmad, A., Barry, J.P., and Nelson, D.C. (1999) Phylogenetic affinity of a wide, vacuolate, nitrate-accumulating *Beggiatoa* sp. from Monterey Canyon, California, with *Thioploca* spp. *Appl Environ Microbiol* **65:** 270–277.

Alexander, R.P., and Zhulin, I.B. (2007) Evolutionary genomics reveals conserved structural determinants of signaling and adaptation in microbial chemoreceptors. *Proc Natl Acad Sci* **104:** 2885–2890.

Alexandre, G., Greer-Phillips, S., and Zhulin, I.B. (2004) Ecological role of energy taxis in microorganisms. *FEMS Microbiol Rev* **28:** 113–126.

Ang, W.K., Mahbob, M., Dhouib, R., and Kappler, U. (2017) Sulfur compound oxidation and carbon co-assimilation in the haloalkaliphilic sulfur oxidizers *Thioalkalivibrio versutus* and *Thioalkalimicrobium aerophilum*. *Res Microbiol* **168:** 255–265.

Aoshima, M., Ishii, M., and Igarashi, Y. (2004a) A novel enzyme, citryl-CoA lyase, catalyzing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Mol Microbiol* **52:** 763–770.

Aoshima, M., Ishii, M., and Igarashi, Y. (2004b) A novel enzyme, citryl-CoA synthetase, catalyzing the first step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6. *Mol Microbiol* **52:** 751–761.

Axen, S.D., Erbilgin, O., and Kerfeld, C.A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10:** e1003898.

Ayala-del-Río, H.L., Chain, P.S., Grzymski, J.J., Ponder, M.A., Ivanova, N., and Bergholz, P.W. (2010) The genome sequence of psychrobacter arcticus 273-4, a psychroactive siberian permafrost bacterium, reveals mechanisms for adaptation to low-temperature growth. *Appl Environ Microbiol* **76:** 2304–2312.

Badger, M.R., and Bek, E.J. (2008) Multiple RubisCO forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. *J Exp Bot* **59:** 1525–1541.

Banciu, H., Sorokin, D.Y., Rijpstra, W.I.C., Sinninghe DamstÃ©, J.S., Galinski, E.A., and Takaichi, S. (2005) Fatty acid, compatible solute and pigment composition of obligately chemolithoautotrophic alkaliphilic sulfur-oxidizing bacteria from soda lakes. *FEMS Microbiol Lett* **243:** 181–187.

Bapteste, E., Moreira, D., and Philippe, H. (2003) Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. *Gene* **318:** 185–191.

Barco, R.A., Emerson, D., Sylvan, J.B., Orcutt, B.N., Jacobson Meyers, M.E., Ramirez, G.A., *et al*. (2015) New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. *Appl Environ Microbiol* **81:** 5927–5937.

Barco, R.A., Hoffman, C.L., Ramirez, G.A., Toner, B.M., Edwards, K.J., and Sylvan, J.B. (2017) In-situ incubation of iron-sulfur mineral reveals a diverse chemolithoautotrophic community and a new biogeochemical role for Thiomicrospira. *Environ Microbiol* **19:** 1322–1337.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., *et al*. (2002) The Pfam protein families database. *Nucleic Acids Res* **30:** 276–280.

Bennett, S. (2004) Solexa Ltd. *Pharmacogenomics* **5:** 433–438.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27:** 573–580.

Bick, J.A., Dennis, J.J., Zylstra, G.J., Nowack, J., and Leustek, T. (2000) Identification of a new class of 5′-adenylylsulfate (aps) reductases from sulfate-assimilating bacteria. *J Bacteriol* **182:** 135–142.

Boden, R., Hutt, L.P., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., *et al*. (2016) Permanent draft genome of *Thermithiobaclillus tepidarius* DSM 3134(T), a moderately thermophilic, obligately chemolithoautotrophic member of the *Acidithiobacillia*. *Stand Genom Sci* **11:** 74.

Boden, R., Scott, K.M., Rae, A.W., Hutt, L.P., *et al*. (2017a) Reclassification of *Thiomicrospira hydrogeniphila* (Watsuji 2016) to *Thiomicrorhabdus hydrogenophila* comb. nov., with emended description of *Thiomicrorhabdus* (Boden). *Int J Syst Evol Microbiol* **67:** 4205–4209.

Boden, R., Scott, K.M., Williams, J., Russel, S., Antonen, K., Rae, A.W., and Hutt, L.P. (2017b) An evaluation of *Thiomicrospira, Hydrogenovibrio* and *Thioalkalimicrobium*: reclassification of four species of *Thiomicrospira* to each *Thiomicrorhabdus* gen. nov. and *Hydrogenovibrio*, and reclassification of all four species of *Thioalkalimicrobium* to *Thiomicrospira*. *Int J Syst Evol Microbiol* **67:** 1140–1151.

Borisov, V.B., Gennis, R.B., Hemp, J., and Verkhovsky, M.I. (2011) The cytochrome bd respiratory oxygen reductases. *Biochim Biophys Acta (BBA) - Bioenergetics* **1807:** 1398–1413.

Brazelton, W.J., and Baross, J.A. (2010) Metagenomic comparison of two *Thiomicrospira* lineages inhabiting contrasting deep-sea hydrothermal environments. *PLoS ONE* **5:** e13530.

Brinkhoff, T., and Muyzer, G. (1997) Increased species diversity and extended habitat range of sulfur-oxidizing *Thiomicrospira* spp. *Appl Environ Microbiol* **63:** 3789–3796.

Brinkhoff, T., Muyzer, G., Wirsesn, C.O., and Kuever, J. (1999a) *Thiomicrospira kuenenii* sp. nov. and *Thiomicrospira frisia* sp. nov., two mesophilic obligately chemolithoautotrophic sulfur-oxidizing bacteria isolated from an intertidal mud flat. *Int J Syst Evol Microbiol* **49:** 385–392.

Brinkhoff, T., Sievert, S.M., Kuever, J., and Muyzer, G. (1999b) Distribution and diversity of sulfur-oxidizing *Thiomicrospira* spp. at a shallow-water hydrothermal vent in the Aegean Sea (Milos, Greece). *Appl Environ Microbiol* **65:** 3843–3849.

Brinkhoff, T., Muyzer, G., Wirsen, C.O., and Kuever, J. (1999c) *Thiomicrospira chilensis* sp. nov., a mesophilic obligately chemolithoautotrophic sulfur-oxidizing bacterium isolated from a *Thioploca* mat. *Int J Syst Evol Microbiol* **49:** 875–879.

Brune, D.C. (1995) Isolation and characterization of sulfur globule proteins from *Chromatium vinosum* and *Thiocapsa roseopersicina*. *Arch Microbiol* **163:** 391–399.

Cannon, G.C., Heinhorst, S., and Kerfeld, C.A. (2010) Carboxysomal carbonic anhydrases: structure and role in microbial CO_2 fixation. *Biochim Biophys Acta-Prot Proteom* **1804:** 382–392.

Chan, L.-K., Morgan-Kiss, R.M., and Hanson, T.E. (2009) Functional analysis of three sulfide:quinone oxidoreductase homologs in *Chlorobaculum tepidum*. *J Bacteriol* **191:** 1026–1034.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., *et al*. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Meth* **10:** 563–569.

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

De Maayer, P., Anderson, D., Cary, C., and Cowan, D.A. (2014) Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* **15:** 508–517.

Dmytrenko, O., Russell, S.L., Loo, W.T., Fontanez, K.M., Liao, L., Roeselers, G., *et al*. (2014) The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genom* **15:** 924.

Dobrinski, K.P., Longo, D.L., and Scott, K.M. (2005) A hydrothermal vent chemolithoautotroph with a carbon concentrating mechanism. *J Bacteriol* **187:** 5761–5766.

Dobrinski, K.P., Enkemann, S.A., Yoder, S.J., Haller, E., and Scott, K.M. (2012) Transcription response of the sulfur chemolithoautotroph *Thiomicrospira crunogena* to dissolved inorganic carbon limitation. *J Bacteriol* **194:** 2074–2081.

Doulatov, S., Hodes, A., Dai, L., Mandhana, N., Liu, M., Deora, R., *et al*. (2004) Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. *Nature* **431:** 476–481.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res* **32:** 1792–1797.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., *et al*. (2009) Real-time DNA sequencing from single polymerase molecules. *Science* **323:** 133–138.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **8:** 175–185.

Feller, G. (2013) Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica* **2013:** 512840.

Flood, B.E., Fliss, P., Jones, D.S., Dick, G.J., Jain, S., Kaster, A.K., *et al*. (2016) Single-cell (meta-)genomics of a dimorphic candidatus *Thiomargarita nelsonii* reveals genomic plasticity. *Front Microbiol* **7:** 603.

Frigaard, N., and Dahl, C. (2009) Sulfur metabolism in phototrophic sulfur bacteria. *Adv Microb Physiol* **54:** 103–200.

Galperin, M. (2005) A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5:** 35–53.

Gao, Y., Meyer, B., Sokolova, L., Zwicker, K., Karas, M., Brutschy, B., *et al*. (2012) Heme-copper terminal oxidase using both cytochrome c and ubiquinol as electron donors. *PNAS* **109:** 3275–3280.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., *et al*. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A* **108:** 1513–1518.

Goffredi, S.K., Childress, J.J., Desaulniers, N.T., Lee, R.W., Lallier, F.H., and Hammond, D. (1997) Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external P-CO_2 and upon proton-equivalent ion transport by the worm. *J Exp Biol* **200:** 883–896.

Gordon, D., Abajian, C., and Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8:** 195–202.

Greening, C., Biswas, A., Carere, C.R., Jackson, C.J., Taylor, M.C., Stott, M.B., *et al*. (2016) Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. *Isme J* **10:** 761–777.

Gregersen, L.H., Bryant, D.A., and Frigaard, N.-U. (2011) Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. *Front Microbiol* **2:** 1–14.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Han, C., and Chain, P. (2006) Finishing repeat regions automatically with Dupfinisher. In *2006 International Conference on Bioinformatics & Computational Biology*. Arabnia, H.R. (ed.). Las Vegas: CSREA Press, pp. 141–146.

Han, W., and Christen, P. (2004) cis-Effect of DnaJ on DnaK in ternary complexes with chimeric DnaK/DnaJ-binding peptides. *FEBS Lett* **563:** 146–150.

Hansen, M., and Perner, M. (2015) Reasons for *Thiomicrospira crunogena's* recalcitrance towards previous attempts to detect its hydrogen consumption ability. *Environ Microbiol Reports* **8:** 53–57.

Hansen, M., and Perner, M. (2016) Hydrogenase gene distribution and H2 consumption ability within the *Thiomicrospira* lineage. *Front Microbiol* **7:** 1–13.

Hanson, T.E., Bonsu, E., Tuerk, A., Marnocha, C.L., Powell, D.H., and Chan, C.S. (2016) *Chlorobaculum tepidum* growth on biogenic S(0) as the sole photosynthetic electron donor. *Environ Microbiol* **18:** 2856–2867.

He, S., Barco, R.A., Emerson, D., and Roden, E.E. (2017) Comparative genomic analysis of neutrophilic iron(ii) oxidizer genomes for candidate genes in extracellular electron transfer. *Front Microbiol* **8:** 1584.

Hedrich, S., and Johnson, D.B. (2013) Aerobic and anaerobic oxidation of hydrogen by acidophilic bacteria. *FEMS Microbiol Lett* **349:** 40–45.

Hügler, M., Huber, H., Molyneaux, S.J., Vetriani, C., and Sievert, S.M. (2007) Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: evidence for two ways of citrate cleavage. *Environ Microbiol* **9:** 81–92.

Hutt, L.P., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., Varghese, N., *et al*. (2017) Permanent draft genome of *Thiobacillus thioparus* DSM 505T, an obligately chemolithoautotrophic member of the *Betaproteobacteria*. *Stand Genom Sci* **12:** 10.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* **11:** 119.

Jannasch, H., Wirsen, C., Nelson, D., and Robertson, L. (1985) *Thiomicrospira crunogena* sp. nov., a colorless, sulfur-oxidizing bacterium from a deep-sea hydrothermal vent. *Int J Syst Bacteriol* **35:** 422–424.

Jannasch, H.W. (1985) The chemosynthetic support of life and the microbial diversity at the deep-sea hydrothermal vents. *Proc R Soc Lond B* **225:** 277–297.

Jo, B.H., Hwang, B.H., and Cha, H.J. (2014) Draft genome sequence of *Hydrogenovibrio marinus* MH-110, a model organism for aerobic H2 metabolism. *J Biotechnol* **185:** 37–38.

Kappler, U., Davenport, K., Beatson, S., Lapidus, A., Pan, C., Han, C., *et al*. (2016) Complete genome sequence of the haloalkaliphilic, obligately chemolithoautotrophic thiosulfate and sulfide-oxidizing gamma-proteobacterium *Thioalkalimicrobium cyclicum* type strain ALM 1 (DSM 14477(T)). *Stand Genom Sci* **11:** 38.

Kato, S., Hara, K., Kasai, H., Teramura, T., Sunamura, M., Ishibashi, J-I., *et al*. (2009) Spatial distribution, diversity and composition of bacterial communities in sub-seafloor fluids at a deep-sea hydrothermal field of the Suiyo Seamount. *Deep Sea Res Part I: Oceanogr Res Papers* **56:** 1844–1855.

Kato, S., Ohkuma, M., Powell, D.H., Krepski, S.T., Oshima, K., Hattori, M., *et al*. (2015) Comparative genomic insights into ecophysiology of neutrophilic, microaerophilic iron oxidizing bacteria. *Front Microbiol* **6:** 1265.

Kelley, D.S., Karson, J.A., Fruh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A., *et al*. (2005) A serpentinite-hosted ecosystem: the lost city hydrothermal field. *Science* **307:** 1428–1434.

Kerscher, S., Dröse, S., Zickermann, V., and Brandt, U. (2008) The three families of respiratory NADH dehydrogenases. In *Bioenergetics: Energy Conservation and Conversion*. Schäfer, G., and Penefsky, H.S. (eds.). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 185–222.

Klappenbach, J.A., Dunbar, J.M., and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Appl Environ Microbiol* **66:** 1328–1333.

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., *et al*. (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci U S A* **109:** E1173–E1182.

Knittel, K., Kuever, J., Meyerdierks, A., Meinke, R., Amann, R., and Brinkhoff, T. (2005) *Thiomicrospira arctica* sp. nov. and *Thiomicrospira psychrophila* sp. nov., psychrophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacteria isolated from marine Arctic sediments. *Int J Syst Evol Microbiol* **55:** 781–786.

Kuenen, J.G., and Veldkamp, H. (1972) *Thiomicrospira pelophila, gen. n., sp. n.*, a new obligately chemolithoautotrohic colorless sulfur bacterium. *Antonie Van Leeuwenhoek* **38:** 241–256.

Kuenen, J.G., and Veldkamp, H. (1973) Effects of organic compounds on growth of chemostat cultures of *Thiomicrospira pelophila, Thiobacillus thioparus* and *Thiobacillus neapolitanus*. *Arch Mikrobiol* **94:** 173–190.

Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief Bioinform* **9:** 299–306.

Kuwahara, H., Yoshida, T., Takaki, Y., Shimamura, S., Nishi, S., Harada, M., *et al*. (2007) Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr Biol* **17:** 881–886.

Lacal, J., García-Fontana, C., Muñoz-Martínez, F., Ramos, J.-L., and Krell, T. (2010) Sensing of environmental signals: classification of chemoreceptors according to the size of their ligand binding regions. *Environ Microbiol* **12:** 2873–2884.

Le, S.Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25:** 1307–1320.

Li, W., Jaroszewski, L., and Godzik, A. (2001) Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **17:** 282–283.

Liu, J., Wang, Z., Belchik, S., Edwards, M., Liu, C., Kennedy, D., *et al*. (2012) Identification and characterization of mtoa: a decaheme c-type cytochrome of the neutrophilic fe(ii)-oxidizing bacterium *Sideroxydans lithotrophicus* ES-1. *Front Microbiol* **3:** 37.

Liu, Y., Wang, Z., Liu, J., Levar, C., Edwards, M.J., Babauta, J.T., *et al*. (2014) A trans-outer membrane porin-cytochrome protein complex for extracellular electron transfer by Geobacter sulfurreducens PCA. *Environ Microbiol Rep* **6:** 776–785.

Lowe, T.M., and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25:** 955–964.

Lubitz, W., Ogata, H., Rudiger, O., and Reijerse, E. (2014) Hydrogenases. *Chem Rev* **114:** 4081–4148.

Mangiapia, M., Brown, T.-R.W., Chaput, D., Haller, E., Harmer, T.L., Hashemy, Z., *et al*. (2017) Proteomic and mutant analysis of the CO_2 concentrating mechanism of hydrothermal vent chemolithoautotroph *Thiomicrospira crunogena*. *J Bacteriol* **199:** e00871.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Markowitz, V.M., Mavromatis, K., Ivanova, N.N., Chen, I.M., Chu, K., and Kyrpides, N.C. (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25:** 2271–2278.

Markowitz, V.M., Chen, I.A., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., *et al*. (2014) IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* **42:** D560–D567.

Marreiros, B.C., Sena, F.V., Sousa, F.M., Batista, A.P., and Pereira, M.M. (2016) Type II NADH:quinone oxidoreductase family: phylogenetic distribution, structural diversity and evolutionary divergences. *Environ Microbiol* **18:** 4697–4709.

Marx, J.G., Carpenter, S.D., and Deming, J.W. (2009) Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. *Can J Microbiol* **55:** 63–72.

Medhekar, B., and Miller, J.F. (2007) Diversity-generating retroelements. *Curr Opin Microbiol* **10:** 388–395.

Menning, K.J., Menon, B.B., Fox, G., and Scott, U.M.L.K.M. (2016) Dissolved inorganic carbon uptake in *Thiomicrospira crunogena* XCL-2 is Δp- and ATP-sensitive and enhances RubisCO-mediated carbon fixation. *Arch Microbiol* **198:** 149–159.

J_ID: **EMI** Customer A_ID: **EMI14090** Cadmus Art: **EMI14090** Ed. Ref. No.: **EMI-2018-0114.R1** Date: **9-March-18** Stage: Page: **21**

Menon, B.B., Dou, Z., Heinhorst, S., Shively, J.M., and Cannon, G.C. (2008) *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species. *PLoS One* **3:** e3570.

Muyzer, G., Teske, A., Wirsen, C.O., and Jannasch, H.W. (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deeop-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* **164:** 165–172.

Nawrocki, E.P., Kolbe, D.L., and Eddy, S.R. (2009) Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25:** 1335–1337.

Newton, I.L., Woyke, T., Auchtung, T.A., Dilly, G.F., Dutton, R.J., Fisher, M.C., *et al*. (2007) The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315:** 998–1000.

Newton, I.L., Girguis, P.R., and Cavanaugh, C.M. (2008) Comparative genomics of vesicomyid clam (Bivalvia: Mollusca) chemosynthetic symbionts. *BMC Genom* **9:** 585.

Nishihara, H., Igarashi, Y., and Kodama, T. (1991) *Hydrogenovibrio marinus* gen. nov., sp. nov., a marine obligately chemolithoautotrophic hydrogen-oxidizing bacterium. *Int J Syst Bacteriol* **41:** 130–133.

Nunoura, T., Takaki, Y., Kazama, H., Kakuta, J., Shimamura, S., Makita, H., *et al*. (2014) Physiological and genomic features of a novel sulfur-oxidizing gammaproteobacterium belonging to a previously uncultivated symbiotic lineage isolated from a hydrothermal vent. *Plos One* **9:** e104959.

Pati, A., Ivanova, N.N., Mikhailova, N., Ovchinnikova, G., Hooper, S.D., Lykidis, A., and Kyrpides, N.C. (2010) GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* **7:** 455–457.

Pitcher, R.S., and Watmough, N.J. (2004) The bacterial cytochrome cbb3 oxidases. *Biochim Biophys Acta* **1655:** 388–399.

Poole, R.K., and Hill, S. (1997) Respiratory protection of nitrogenase activity in *Azotobacter vinelandii*–roles of the terminal oxidases. *Biosci Rep* **17:** 303–317.

Price, G.D. (2011) Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. *Photosynth Res* **109:** 47–57.

Price, G., Woodger, F., Badger, M., Howitt, S., and Tucker, L. (2004) Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc Natl Acad Sci U S A* **101:** 18228–18233.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (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.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Quasem, I., Achille, A.N., Caddick, B.A., Carter, T.A., Daniels, C., Delaney, J.A., *et al*. (2017) Peculiar citric acid cycle of hydrothermal vent chemolithoautotroph *Hydrogenovibrio crunogenus*, and insights into carbon metabolism by obligate autotrophs. *FEMS Microbiol Let* **364:** fnx148.

Rambaut, A. (2016) *FigTree: Tree Figure Drawing Tool Version 1.4.3*. Edinburgh: CoInstitute of Evolutionary Biology, University of Edinburgh.

Ramel, F., Brasseur, G., Pieulle, L., Valette, O., Hirschler-Réa, A., Fardeau, M.L., and Dolla, A. (2015) Growth of the obligate anaerobe *Desulfovibrio vulgaris* Hildenborough under continuous low oxygen concentration sparging: impact of the membrane-bound oxygen reductases. *PLoS One* **10:** e0123455.

Ramulu, H.G., Groussin, M., Talla, E., Planel, R., Daubin, V., and Brochier-Armanet, C. (2014) Ribosomal proteins: toward a next generation standard for prokaryotic systematics?. *Mol Phylogenet Evol* **75:** 103–117.

Rainey, F.A., Jones, B.E., Lysenko, A.M., Robertson, L.A., Kuenen, G.J., Sorokin, D.Y., *et al*. (2001) *Thioalkalimicrobium aerophilum* gen. nov., sp. nov. and *Thioalkalimicrobium sibericum* sp. nov., and *Thioalkalivibrio versutus* gen. nov., sp. nov., *Thioalkalivibrio nitratis* sp.nov., novel and *Thioalkalivibrio denitrificans* sp. nov., novel obligately alkaliphilic and obligately chemolithoautotrophic sulfur-oxidizing bacteria from soda lakes. *Int J Syst Evol Microbiol* **51:** 565–580.

Reshetnikov, A.S., Rozova, O.N., Khmelenina, V.N., Mustakhimov, I.I., Beschastny, A.P., Murrell, J.C., and Trotsenko, Y.A. (2008) Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus* Bath. *FEMS Microbiol Lett* **288:** 202–210.

Riley, M., Staley, J.T., Danchin, A., Wang, T.Z., Brettin, T.S., Hauser, L.J., Land, M.L., and Thompson, L.S. (2008) Genomics of an extreme psychrophile, *Psychromonas ingrahamii*. *BMC Genom* **9:** 210.

Roller, B.R.K., Stoddard, S.F., and Schmidt, T.M. (2016) Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat Microbiol* **1:** 16160.

Römling, U., Galperin, M.Y., and Gomelsky, M. (2013) Cyclic di-GMP: the first 25 years of a universal bacterial second messenger. *Microbiol Mol Biol Rev* **77:** 1–52.

Saier, M.H., Jr., Reddy, V.S., Tamang, D.G., and Vastermark, A. (2014) The transporter classification database. *Nucleic Acids Res* **42:** D251–D258.

Scott, K.M., Sievert, S.M., Abril, F.N., Ball, L.A., Barrett, C.J., Blake, R.A., *et al*. (2006) The genome of deep-sea vent chemolithoautotroph *Thiomicrospira crunogena*. *PLoS Biol* **4:** e383.

Shibata, M., Katoh, H., Sonoda, M., Ohkawa, H., Shimoyama, M., Fukuzawa, H., *et al*. (2002) Genes essential to sodium-dependent bicarbonate transport in cyanobacteria. *J Biol Chem* **277:** 18658–18664.

Sleator, R.D., and Hill, C. (2002) Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol Rev* **26:** 49–71.

Smith, A.J., London, J., and Stanier, R.Y. (1967) Biochemical basis of obligate autotrophy in blue- green algae and thiobacilli. *J Bacteriol* **94:** 661–667.

Sorokin, D.Y., Cherepanov, A., Vries, S., and Kuenen, G.J. (1999) Identification of cytochrome c oxidase in the alkaliphilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium '*Thioalcalomicrobium aerophilum*' strain AL 3. *FEMS Microbiol Lett* **179:** 91–99.

Sorokin, D.Y., Gorlenko, V.M., Tourova, T.P., Tsapin, A.I., Nealson, K.H., and Kuenen, G.J. (2002) *Thioalkalimicrobium cyclicum* sp. nov. and *Thioalkalivibrio jannaschii* sp. nov., novel species of haloalkaliphilic, obligately chemolithoautotrophic sulfur-oxidizing bacteria from hypersaline

ID: **nagarajulum** Time: 19:09 I Path: //chenas03/Cenpro/ApplicationFiles/Journals/Wiley/EMI#/Vol00000/180055/Comp/APPFile/JW-EMI#180055

alkaline Mono Lake (California). *Int J Syst Evol Microbiol* **52:** 913–920.

Sorokin, D.Y., Tourova, T.P., Kolganova, T.V., Spiridonova, E.M., Berg, I.A., and Muyzer, G. (2006) *Thiomicrospira halphila* sp. nov., a moderately halophilic, obligately chemolithoautotrophic, sulfur-oxidizing bacterium from hypersaline lakes. *Int J Syst Evol Microbiol* **56:** 2375–2380.

Sorokin, D.Y., Foti, M., Pinkart, H.C., and Muyzer, G. (2007) Sulfur-oxidizing bacteria in Soap Lake (Washington State), a meromictic, haloalkaline lake with an unprecedented high sulfide content. *Appl Environ Microbiol* **73:** 451–455.

Sorokin, D.Y., Tourova, T.P., Muyzer, G., and Kuenen, G.J. (2008) *Thiohalospira halophila* gen. nov., sp. nov. and *Thiohalospira alkaliphila* sp. nov., novel obligately chemolithoautotrophic, halophilic, sulfur-oxidizing gammaproteobacteria from hypersaline habitats. *Int J Syst Evol Microbiol* **58:** 1685–1692.

Sorokin, D.Y., Kuenen, J.G., and Muyzer, G. (2011) The microbial sulfur cycle at extremely haloalkaline conditions of soda lakes. *Front Microbiol* **2:** 1–16.

Strøm, T., Ferenci, T., and Quayle, J.R. (1974) The carbon assimilation pathways of *Methylococcus capsulatus, Pseudomonas methanica* and *Methylosinus trichosporium* (OB3B) during growth on methane. *Biochem J* **144:** 465–476.

Tabita, F.R. (1999) Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res* **60:** 1–28.

Tabita, F.R., Hanson, T.E., Satagopan, S., Witte, B.H., and Kreel, N.E. (2008) Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Philos Trans R Soc B-Biol Sci* **363:** 2629–2640.

Takai, K., Hirayama, H., Nakagawa, T., Suzuki, Y., Nealson, K.H., and Horikoshi, K. (2004) *Thiomicrospira thermophila* sp nov., a novel microaerobic, thermotolerant, sulfur-oxidizing chemolithomixotroph isolated from a deep-sea hydrothermal fumarole in the TOTO caldera, Mariana Arc, Western Pacific. *Int J Syst Evol Microbiol* **54:** 2325–2333.

Talavera, G., and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56:** 564–577.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., and Rao, B.S. (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* **29:** 22–28.

Touchon, M., and Rocha, E.P.C. (2007) Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* **24:** 969–981.

Tourova, T.P., Spiridonova, E.M., Berg, I.A., Kuznetsov, B.B., and Sorokin, D.Y. (2006) Occurrence, phylogeny and evolution of ribulose-1,5-bisphosphate carboxylase/oxygenase genes in obligately chemolithoautotrophic sulfur-oxidizing bacteria of the genera *Thiomicrospira* and *Thioalkalimicrobium*. *Microbiology* **152:** 2159–2169.

Treangen, T.J., and Salzberg, S.L. (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13:** 36–46.

Wadhams, G.H., and Armitage, J.P. (2004) Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* **5:** 1024–1037.

Waite, D.W., Vanwonterghem, I., Rinke, C., Parks, D.H., Zhang, Y., Takai, K., *et al*. (2017) Comparative genomic analysis of the class *Epsilonproteobacteria* and proposed reclassification to *Epsilonbacteraeota* (phyl. nov.). *Front Microbiol* **8:** 682.

Wang, F., Xiao, X., Ou, H.Y., Gai, Y., and Wang, F. (2009) Role and regulation of fatty acid biosynthesis in the response of *Shewanella piezotolerans* WP3 to different temperatures and pressures. *J Bacteriol* **191:** 2574–2584.

Watsuji, T., Hada, E., Miyazaki, M., Ichimura, M., and Takai, K. (2016) *Thiomicrospira hydrogeniphila* sp. nov., an aerobic, hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a seawater tank containing a block of beef tallow. *Int J Syst Evol Microbiol* **66:** 3688–3693.

Weissgerber, T., Sylvester, M., Kröninger, L., and Dahl, C. (2014) A comparative quantitative proteomic study identifies new proteins relevant for sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. *Appl Environ Microbiol* **80:** 2279–2292.

Williams, K.P., Gillespie, J.J., Sobral, B.W.S., Nordberg, E.K., Snyder, E.E., Shallom, J.M., and Dickerman, A.W. (2010) Phylogeny of *gammaproteobacteria*. *J Bacteriol* **192:** 2305–2314.

Williams, T.J., Lauro, F.M., Ertan, H., Burg, D.W., Poljak, A., Raftery, M.J., and Cavicchioli, R. (2011) Defining the response of a microorganism to temperatures that span its complete growth temperature range (−2°C to 28°C) using multiplex quantitative proteomics. *Environ Microbiol* **13:** 2186–2203.

Wood, A.P., Aurikko, J.P., and Kelly, D.P. (2004) A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy?. *FEMS Microbiol Rev* **28:** 335.

Wuichet, K., Alexander, R.P., and Zhulin, I.B. (2007) Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. In *Methods in Enzymology.* Melvin I. Simon, B.R.C., and Alexandrine, C. (eds.). New York: Academic Press, pp. 3–31.

Yasutake, Y., Watanabe, S., Yao, M., Takada, Y., Fukunaga, N., and Tanaka, I. (2002) Structure of the monomeric isocitrate dehydrogenase: evidence of a protein monomerization by a domain duplication. *Structure* **10:** 1637–1648.

Yoo, J.-G., and Bowien, B. (1995) Analysis of the *cbbF* genes from *Alcaligenes eutrophus* that encode fructose-1,6-/sedoheptulose-1,7-bisphosphatase. *Curr Microbiol* **31:** 55–61.

Yoshizawa, Y., Toyoda, K., Arai, H., Ishii, M., and 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.

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821–829.

Zhang, S., and Bryant, D.A. (2011) The tricarboxylic acid cycle in Cyanobacteria. *Science* **334:** 1551–1553.

Zhang, G., Fauzi Haroon, M., Zhang, R., Hikmawan, T., and Stingl, U. (2016) Draft genome sequences of two *Thiomicrospira* strains isolated from the brine-seawater interface of kebrit deep in the red sea. *Genome Announc* **4:** e00109–e00116.

Zhou, K., Aertsen, A., and Michiels, C.W. (2014) The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiol Rev* **38:** 119–141.
Ziegler, C., Bremer, E., and Kramer, R. (2010) The BCCT family of carriers: from physiology to crystal structure. *Mol Microbiol* **78:** 13–34.

Supporting information

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

Table S1. Genome data of autotrophic microorganisms from the '*Proteobacteria*'.
Table S2. Numbers of genes belonging to selected Pfams.
Table S3. Inorganic sulfur-metabolizing systems encoded in genomes of autotrophic microorganisms from the '*Proteobacteria*'.
Table S4. Electron transport chain components encoded in genomes of autotrophic microorganisms from the '*Proteobacteria*'.
Table S5. Presence of genes encoding RubisCO in genomes of autotrophic microorganisms from the '*Proteobacteria*'.
Table S6. Queries and criteria used to find sulfur metabolizing complexes and electron transport chain components.
Fig. S1. Annotated ORFs of *H. marinus* and *Tmr. frisia* KP2 prophages.
Fig. S2. Maximum likelihood analysis of amino acid sequences predicted from genes encoding form II RubisCO (CbbM) and large subunits from form I (CbbL). The tree is rooted, and bootstrap values are from 1000 resamplings of the alignment. The scale bar represents the number of substitutions per site.