The Genome of Deep-Sea Vent Chemolithoautotroph
*Thiomicrospira crunogena XCL-2*

<table>
<thead>
<tr>
<th>Item Type</th>
<th>Article</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DOI</strong></td>
<td>10.1371/journal.pbio.0040383</td>
</tr>
<tr>
<td><strong>Publisher</strong></td>
<td>Public Library of Science</td>
</tr>
<tr>
<td><strong>Journal</strong></td>
<td>PLoS Biology</td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>©2006 Scott et al. This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.</td>
</tr>
<tr>
<td><strong>Download date</strong></td>
<td>11/4/2019</td>
</tr>
<tr>
<td><strong>Link to Item</strong></td>
<td>URI: <a href="http://hdl.handle.net/20.500.11868/861">http://hdl.handle.net/20.500.11868/861</a></td>
</tr>
</tbody>
</table>
The Genome of Deep-Sea Vent Chemolithoautotroph *Thiomicrospira crunogena* XCL-2

Kathleen M. Scott\(^1\), Stefan M. Sievert\(^2\), Fereniki N. A brill\(^1\), Lois A. Ball\(^1\), Chantell J. Barrett\(^1\), Rodrigo A. Blake\(^1\), Amanda J. Boller\(^1\), Patrick S. G. Chain\(^3,4\), Justine A. Clark\(^1\), Carisa R. Davis\(^1\), Chris Detter\(^4\), Kimberly F. Do\(^1\), Kimberly P. Dobrinski\(^1\), Brandon I. Faza\(^1\), Kelly A. Fitzpatrick\(^1\), Sharyn K. Freyermuth\(^5\), Tara L. Harmer\(^6\), Loren J. Hauser\(^7\), Michael Hügler\(^2\), Cheryl A. Kerfeld\(^8\), Martin G. Klotz\(^9\), William W. Kong\(^1\), Miriam Land\(^7\), Alla Lapidus\(^4\), Frank W. Larimer\(^7\), Dana L. Longo\(^1\), Susan Lucas\(^4\), Stephanie A. Malfatti\(^3,4\), Steven E. Massey\(^1\), Darlene D. Martin\(^1\), Zoe McCuddin\(^10\), Folker Meyer\(^11\), Jessica L. Moore\(^1\), Luis H. Ocampo Jr.\(^1\), John H. Paul\(^12\), Ian T. Paulsen\(^13\), Douglas K. Reep\(^1\), Qinghu Ren\(^13\), Rachel L. Ross\(^1\), Priscila Y. Sato\(^1\), Phaedra Thomas\(^1\), Lance E. Tinkham\(^1\), Gary T. Zeruth\(^1\)

\(^1\) Biology Department, University of South Florida, Tampa, Florida, United States of America, \(^2\) Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, United States of America, \(^3\) Lawrence Livermore National Laboratory, Livermore, California, United States of America, \(^4\) Joint Genome Institute, Walnut Creek, California, United States of America, \(^5\) Department of Biochemistry, University of Missouri, Columbia, Missouri, United States of America, \(^6\) Division of Natural Sciences and Mathematics, The Richard Stockton College of New Jersey, Pomona, New Jersey, United States of America, \(^7\) Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America, \(^8\) Molecular Biology Institute, University of California Los Angeles, Los Angeles, California, United States of America, \(^9\) University of Louisville, Louisville, Kentucky, United States of America, \(^10\) The Monsanto Company, Ankeny, Iowa, United States of America, \(^11\) Center for Biotechnology, Bielefeld University, Bielefeld, Germany, \(^12\) College of Marine Science, University of South Florida, St. Petersburg, Florida, United States of America, \(^13\) The Institute for Genomic Research, Rockville, Maryland, United States of America

Presented here is the complete genome sequence of *Thiomicrospira crunogena* XCL-2, representative of ubiquitous chemolithoautotrophic sulfur-oxidizing bacteria isolated from deep-sea hydrothermal vents. This gammaproteobacterium has a single chromosome (2,427,734 base pairs), and its genome illustrates many of the adaptations that have enabled it to thrive at vents globally. It has 14 methyl-accepting chemotaxis protein genes, including four that may assist in positioning it in the redoxcline. A relative abundance of coding sequences (CDSs) encoding regulatory proteins likely control the expression of genes encoding carboxysomes, multiple dissolved inorganic nitrogen and phosphate transporters, as well as a phosphonate operon, which provide this species with a variety of options for acquiring these substrates from the environment. *Thiom. crunogena* XCL-2 is unusual among obligate sulfur-oxidizing bacteria in relying on the Sox system for the oxidation of reduced sulfur compounds. The genome has characteristics consistent with an obligately chemolithoautotrophic lifestyle, including few transporters predicted to have organic allocrits, and Calvin-Benson-Bassham cycle CDSs scattered throughout the genome.


**Introduction**

Deep-sea hydrothermal vent communities are sustained by prokaryotic chemolithoautotrophic primary producers that use the oxidation of electron donors available in hydrothermal fluid (H\(_2\), H\(_2\)S, and Fe\(^{2+}\)) to fuel carbon fixation [1–3]. The chemical and physical characteristics of their environment are dictated largely by the interaction of hydrothermal fluid and bottom water. When warm, reductant- and CO\(_2\)-rich hydrothermal fluid is emitted from fissures in the basalt crust, it creates eddies as it mixes with cold, oxic bottom water. As a consequence, at areas where dilute hydrothermal fluid and seawater mix, a microorganism’s habitat is erratic, oscillating from seconds to hours between dominance by hydrothermal fluid (warm; anoxic; abundant electron donors; 0.02 to >1 mM CO\(_2\)) and bottom water (2 °C; oxic; 0.02 mM CO\(_2\)) [4,5].

Common chemolithoautotrophic isolates from these “mixing zones” from hydrothermal vents include members of the genus *Thiomicrospira*, a group that originally included all marine, spiral-shaped sulfur-oxidizing bacteria. Subsequent analyses of 16S rDNA sequences have revealed the polyphyletic nature of this group; members of *Thiomicrospira* are distributed among the gamma and epsilon classes of the Proteobacteria. *Thiomicrospira crunogena*, a member of the
Thiom. crunogena Genome

To provide the energy necessary for carbon fixation and cell maintenance, Thiom. crunogena XCL-2 and its close relatives Thiomicrospira spp. L-12 and MA-3 are capable of using hydrogen sulfide, thiosulfate, elemental sulfur, and sulfide minerals (e.g., pyrite and chalcopyrite) as electron donors; the only electron acceptor they can use is oxygen [6,10–12].

Given its temporally variable habitat, Thiom. crunogena XCL-2 is likely adapted to cope with oscillations in the availability of the inorganic nutrients necessary for chemolithoautotrophic growth. One critical adaptation in this habitat is its carbon-concentrating mechanism [13,14]. This species is capable of rapid growth in the presence of low concentrations of dissolved inorganic carbon, due to an increase in cellular affinity for both HCO3− and CO2 under low-CO2 conditions [14]. The ability to grow under low-CO2 conditions is likely an advantage when the habitat is dominated by relatively low-CO2 seawater. Further adaptations in nutrient acquisition and microhabitat sensing are likely to be present in this organism.

Thiom. crunogena XCL-2 [15] is the first deep-sea autotrophic hydrothermal vent bacterium to have its genome completely sequenced and annotated. Many other autotrophic bacterial genomes have been examined previously, including several species of cyanobacteria (e.g., [16,17]), nitrifiers [18], purple nonsulfur [19], and green sulfur [20] photosynthetic bacteria, as well as an obligately chemolithoautotrophic sulfur-oxidizer [21] and a hydrogen-oxidizer [22]. These genomes have provided insight into the evolution of autotrophy among four of the seven phyla of Bacteria known to have autotrophic members.

The genome of Thiom. crunogena XCL-2 was sequenced to illuminate the evolution and physiology of bacterial primary producers from hydrothermal vents and other extreme environments. It was of interest to determine whether any specific adaptations to thrive in an environment with extreme temporal and spatial gradients in habitat geochemistry would be apparent from the genome. It was predicted that comparing its genome both to the other members of the Gammaproteobacteria, many of which are pathogenic heterotrophs, and also to autotrophs from the Proteobacteria and other phyla, would provide insights into the evolution and physiology of autotrophs within the Gammaproteobacteria. Further, this genome provides a reference point for uncultivated (to date) chemoautotrophic sulfur-oxidizing gammaproteobacterial symbionts of various invertebrates.

Results/Discussion

Genome Structure

Thiom. crunogena XCL-2 has a single chromosome consisting of 2.43 megabase pairs (Mb), with a GC content of 43.1% and a high coding density (90.6%; Figure 1). The GC skew shifts near the gene encoding the DnaA protein (located at “noon” on the circular map; Tcr0001), and thus the origin of replication is likely located nearby. One region with a deviation from the average %GC contains a phosphonate operon and has several other features consistent with its acquisition via horizontal gene transfer (see “Phosphorus Uptake” below). Many genes could be assigned a function with a high degree of confidence (Table 1), and a model for cell function based on these genes is presented (Figure 2).

Three rRNA operons are present, and two of them, including their intergenic regions, are 100% identical. In the third rRNA operon, the 16S and 5S genes are 100% identical to the other two, but the 23S gene has a single substitution. The intergenic regions of this third operon also have several substitutions compared to the other two, with three substitutions between the tRNA-Ile-GAT and tRNA-Ala-TGC genes, six substitutions between the tRNA-Ala-TGC and 23S genes, and one substitution between the 23S and 5S genes. Having three rRNA operons may provide additional
A putative prophage genome was noted in the *Thiom. crunogena* chromosome. The putative prophage is 38,090 base pairs (bp) and contains 54 coding sequences (CDSs), 21 of which (38.9%) had significant similarity to genes in GenBank. The prophage genome begins with a tyrosine integrase (*Tcr0656*) and contains a cl-like repressor gene (*Tcr0666*), features common to lambdoid prophages (Figure 3 [24]). These genes define a probable *lysis* module [25] and are in the opposite orientation from the rest of the phage genome (the replicative or *lytic* module).

The lytic half of the prophage genome encodes putative genes involved in DNA replication and phage assembly (Figure 3). Beginning with a putative DNA primase (*Tcr0668*) is a cluster of genes interpreted to represent an active or remnant DNA replication module (including an exonuclease of DNA polymerase, a hypothetical DNA binding protein, and a terminase large subunit: *Tcr0669*, *0670*, and *0672*). Terminases serve to cut the phage DNA in genome-sized fragments prior to packaging. Beyond this are eight CDSs of unknown function, and then two CDSs involved in capsid assembly, including the portal protein (*Tcr0679*) and a minor capsid protein (*Tcr0680*) similar to GPC of λ. Portal proteins are ring-like structures in phage capsids through which the DNA enters the capsid during packaging [26]. In λ, the GPC protein is a peptidase (S49 family) that cleaves the capsid protein from a scaffolding protein involved in the capsid assembly process [27]. Although no major capsid protein is identifiable from bioinformatics, capsid proteins are often difficult to identify from sequence information in marine phages [28]. A cluster of P2-like putative tail assembly and structural genes follows the capsid assembly genes. The general organization of these genes (tail fiber, tail shaft and sheath, and tape measure: *Tcr0691*; *Tcr0690*; *Tcr0695*; and *Tcr0698*) is also P2-like [24]. The complexity of these genes (ten putative CDSs involved in tail assembly) and the strong identity score for a contractile tail sheath protein strongly argues that this prophage was a member of the Myoviridae, i.e., phages possessing a contractile tail. The final gene in the prophage-like sequence was similar to a phage late control protein D, gpD (*Tcr0700*). In λ, gpD plays a role in the expansion of the capsid to accommodate the entire phage genome [29].

The high similarity of the CDSs to lambdoid (lysogeny and replication genes) and P2-like (tail module) temperate coliphages is surprising and unprecedented in marine prophage genomes [30]. A major frustration encountered in marine phage genomics is the low similarity of CDSs to anything in GenBank, making the interpretation of the biological function extremely difficult. The lambdoid siphophages are generally members of the Siphoviridae, whereas the P2-like phages are Myoviridae, which the *Thiom. crunogena* XCL-2 prophage is predicted to be. Such a mixed heritage is often the result of the modular evolution of phages. The general genomic organization of the *Thiom. crunogena* XCL-2 prophage-like element (integrase, repressor, DNA replicative genes, terminase, portal, capsid, tail genes) is common to several known prophages, including those of *Staphylococcus aureus* (i.e., *ϕMu50B*), *Streptococcus pyogenes* (prophages 370.3 and 370.2), and *Streptococcus thermophilus* (prophage O1205 [31]).

### Redox Substrate Metabolism and Electron Transport

Genes are present in this genome that encode all of the components essential to assemble a fully functional Sox system that performs sulfite-, thiosulfate-, sulfur-, and hydrogen sulfide–dependent cytochrome c reduction, namely, SoxXA (*Tcr0604* and *Tcr0601*), SoxYZ (*Tcr0603* and *Tcr0602*), SoxB (*Tcr1549*), and SoxCD (*Tcr0156* and *Tcr1575*) [32,33]. This well-characterized system for the oxidation of reduced sulfur compounds has been studied in facultatively chemolithoautotrophic, aerobic, thiosulfate-oxidizing alphaproteobacteria, including *Paracococcus versutus* GB17, *Thiobacillus versutus*, *Starkeya novella*, and *Pseudoaminobacter salicylatoxidans* ([32,34] and references therein). This model involves a periplasmic multienzyme complex that is capable of oxidizing various reduced sulfur compounds completely to sulfate. Genes encoding components of this complex have been identified, and it has further been shown that these so-called "sox" genes form extensive clusters in the genomes of the aforementioned bacteria. Essential components of the Sox system have also

### Table 1. *Thiom. crunogena* XCL-2 Genome Summary

<table>
<thead>
<tr>
<th>Item</th>
<th>Genes</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosomes</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Base pairs</td>
<td></td>
<td>2,427,734</td>
</tr>
<tr>
<td>GC content (%)</td>
<td></td>
<td>43.1</td>
</tr>
<tr>
<td>Percent coding</td>
<td></td>
<td>90.6</td>
</tr>
<tr>
<td>RNA-encoding genes</td>
<td>tRNAs</td>
<td>43</td>
</tr>
<tr>
<td>Genes in each COG category</td>
<td>DNA replication, recombination, and repair</td>
<td>113</td>
</tr>
<tr>
<td>DNA replication, recombination, and repair</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>Transcription</td>
<td></td>
<td>84</td>
</tr>
<tr>
<td>Translation, ribosomal structure and biogenesis</td>
<td>153</td>
<td></td>
</tr>
<tr>
<td>Posttranslational modification, protein turnover, chaperones</td>
<td>115</td>
<td></td>
</tr>
<tr>
<td>Energy production and conversion</td>
<td>117</td>
<td></td>
</tr>
<tr>
<td>Carbohydrate transport and metabolism</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Amino acid transport and metabolism</td>
<td>167</td>
<td></td>
</tr>
<tr>
<td>Nucleotide transport and metabolism</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Lipid transport and metabolism</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Coenzyme transport and metabolism</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>Secondary metabolite biosynthesis, transport, catabolism</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Cell wall/membrane/envelope biogenesis</td>
<td>142</td>
<td></td>
</tr>
<tr>
<td>Inorganic ion transport and metabolism</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>Cell motility</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>Signal transduction mechanisms</td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>Cell cycle control, cell division, chromosome partitioning</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Intracellular trafficking, secretion, and vesicular transport</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>General function</td>
<td></td>
<td>188</td>
</tr>
</tbody>
</table>

DOI: 10.1371/journal.pbio.0040383.t001
been identified in genomes of other bacteria known to be able to use reduced sulfur compounds as electron donors, resulting in the proposal that there might be a common mechanism for sulfur oxidation utilized by different bacteria [32,34]. Interestingly, *Thiom. crunogena* XCL-2 appears to be the first obligate chemolithoautotrophic sulfur-oxidizing bacterium to rely on the Sox system for oxidation of reduced sulfur compounds.

Genome analyses also reveal the presence of a putative sulfide:quinone reductase gene (*Tcr1170*; SQR). This enzyme is present in a number of phototrophic and chemotrophic bacteria and is best characterized from *Rhodobacter capsulatus* [35]. In this organism, it is located on the periplasmic surface of the cytoplasmic membrane, where it catalyzes the oxidation of sulfide to elemental sulfur, leading to the deposition of sulfur outside the cells. It seems reasonable to assume that SQR in *Thiom. crunogena* XCL-2 performs a similar function, explaining the deposition of sulfur outside the cell under certain conditions (e.g., low pH or oxygen [36]).

**Figure 2.** Cell Model for *Thiom. crunogena* XCL-2, with an Emphasis on Ultrastructure, Transport, Energy, Carbon Metabolism, and Chemotaxis

Genes encoding virtually all of the steps for the synthesis of nucleotides and amino acids by canonical pathways are present in the bacterium, but are omitted here for simplicity. Electron transport components are yellow, and abbreviations are as follows: bc, bc1 complex; cb3, cb3-type cytochrome C oxidase; cytC, cytochrome C; NDH, NADH dehydrogenase; Sox, Sox system; UQ, ubiquinone. MCPs are fuchsia, as are MCPs with PAS domains or PAS folds. Influx and efflux transporter families with representatives in this genome are indicated on the figure, with the number of each type of transporter in parentheses. ATP-dependent transporters are red, secondary transporters are sky blue, ion channels are light green, and unclassified transporters are purple. Abbreviations for transporter families are as follows: ABC, ATP-binding cassette superfamily; AGCS, alanine or glycine-cation symporter family; AMT, ammonium transporter family; APC, amino acid-polyamine-organocation family; ATP syn, ATP synthetase; BASS, bile acid:Na+ symporter family; BCCT, betaine/carnitine/choline transporter family; CaCA, Ca2+:cation antiporter family; CDF, cation diffusion facilitator family; CHR, chromosome transporter family; CPA, monovalent cation:proton antiporter-1, –2, and –3 families; DAAC, dicarboxylate/amino acid:cation symporter family; DASS, divalent anion:sodium symporter-2 family; FeoB, ferrous iron uptake family; IRT, iron/lead transporter superfamily; MATE, multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily, MATE family; McsS, small conductance mechanosensitive ion channel family; MFS, major facilitator superfamily; MgtE, Mg2+-transporter-E family; MIT, CorA metal ion transporter family; NCS2, nucleobase:cation antiporter-2 family; NRAMP, metal ion transporter family; NSS, neurotransmitter:sodium symporter family; P-ATP, P-type ATPase superfamily; Pit, inorganic phosphate transporter family; PnuC, nicotinate mononucleotide uptake permease family; RhtB, resistance to homoserine/threonine family; RND, resistance-nodulation-cell division superfamily; SSS, solute:sodium symporter family; SulP, sulfate permease family; TRAP, tripartite ATP-independent periplasmic transporter family; TRK, K+ transporter family; VIC, voltage-gated ion channel superfamily.

DOI: 10.1371/journal.pbio.0040383.g002
interesting question in this regard will be to determine how *Thiom. crunogena* XCL-2 remobilizes the sulfur globules. The dependence on the Sox system, and possibly SQR, for sulfur oxidation differs markedly from the obligately autotrophic sulfur-oxidizing betaproteobacterium *Thiobacillus denitrificans*, which has a multitude of pathways for sulfur oxidation, perhaps facilitating this organism’s ability to grow under aerobic and anaerobic conditions [21].

In contrast to the arrangement in facultatively autotrophic sulfur-oxidizers [34], the sox components in *Thiom. crunogena* XCL-2 are not organized in a single cluster, but in different parts of this genome: soxXYZA, soxB, and soxCD. In particular, the isolated location of soxB relative to other sox genes has not been observed in any other sulfur-oxidizing organisms. The components of the Sox system that form tight interactions in vivo are collocated in apparent operons (soxXYZA and soxCD [37]), which is consistent with the “molarity model” for operon function (reviewed in [38]), in which cotranslation from a single mRNA facilitates interactions between tightly interacting proteins, and perhaps correct folding. Perhaps for obligate chemolithotrophs like *Thiom. crunogena* XCL-2 that do not have multiple sulfur oxidation systems, in which sox gene expression is presumably constitutive and not subject to complex regulation [39], sox gene organization into a single operon may not be strongly evolutionarily selected. Alternatively, the *Thiom. crunogena* XCL-2 sox genes may not be constitutively expressed, and may instead function as a regulon.

The confirmation of the presence of a soxB gene in *Thiom. crunogena* XCL-2 is particularly interesting, as it is a departure from previous studies with close relatives. Attempts to PCR-amplify soxB from *Thiom. crunogena* ATCC 700270T and *Thiom. pelophila* DSM 1534T were unsuccessful [40]. In contrast, a newly isolated *Thiomicrospira* strain obtained from a hydrothermal vent in the North Fiji Basin, *Thiom. crunogena* HY-62, was positive, with phylogenetic analyses further revealing that its soxB was most closely related those from Alphaproteobacteria, such as *Silicibacter pomeroyi* [40]. The soxB gene from *Thiom. crunogena* XCL-2 falls into a cluster containing the green-sulfur bacterium *Chlorobium* and the purple sulfur gammaproteobacterium *Allochromatium vinosum*, and separate from the cluster containing soxB from *Si. pomeroyi* and *Thiom. crunogena* HY-62 (Figure 4). This either indicates that *Thiom. crunogena* XCL-2 has obtained its soxB gene through lateral gene transfer from different organisms, or that the originally described soxB gene in *Thiom. crunogena* HY-62 was derived from a contaminant. The fact that both soxA and soxX from *Thiom. crunogena* XCL-2 also group closely with their respective homologs from *Chlorobium* spp argues for the latter (unpublished data). Also, the negative result for the two other *Thiomicrospira* strains is difficult to explain in light of the observation that sulfur oxidation in *Thiom. crunogena* XCL-2 appears to be dependent on a functional Sox system. It is

---

Figure 3. Prophage Genome within the *Thiom. crunogena* XCL-2 Genome

Lysogenic and lytic genes are delineated, as are predicted gene functions.

DOI: 10.1371/journal.pbio.0040383.g003

Figure 4. Phylogenetic Relationship of *Thiom. crunogena* XCL-2 SoxB to Sequences of Selected Bacteria

Sequences were aligned using the program package MacVector. Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10 [113]. Bootstrap values (1,000 replicates) are given for the neighbor-joining (first value) and parsimony analyses (second value).

DOI: 10.1371/journal.pbio.0040383.g004
possible that *Thiom. crunogena* ATCC 700270<sup>T</sup> and *Thiom. pelophila* DSM 1543<sup>T</sup> also have *soxB* genes, but that the PCR primers did not target conserved regions of this gene.

Up to this point, obligate chemolithoautotrophic sulfur oxidizers were believed to use a pathway different from the Sox system, i.e., the SH pathway [41] or a pathway that represents basically a reversal of dissimilatory sulfate reduction, by utilizing the enzymes dissimilatory sulfite reductase, APS reductase, and ATP sulfurylase [42]. In this context, it is interesting to note that *Thiom. crunogena* also seems to lack enzymes for the assimilation of sulfate, i.e., ATP sulfurylase, APS kinase, PAPS reductase, and a sirohaem-containing sulfite reductase, indicating that it depends on reduced sulfur compounds for both dissimilation and assimilation. *Thiom. crunogena* XCL-2 apparently also lacks a sulfite:acceptor oxidoreductase (SorAB), an enzyme evolutionarily related to SoxCD that catalyzes the direct oxidation of sulfite to sulfate and that has a wide distribution among different sulfur-oxidizing bacteria (Figure S1). The presence of the Sox system and the dependence on it in an obligate chemolithoautotroph also raises the question of the origin of the Sox system. Possibly, this system first evolved in obligate autotrophs before it was transferred into facultative autotrophs. Alternatively, *Thiom. crunogena* XCL-2 might have secondarily lost its capability to grow heterotrophically.

Genes for NiFe hydrogenase large and small subunits are present (*Tcr2037* and *Tcr2038*), as well as all of the genes necessary for large subunit metal center assembly (*Tcr2035–6* and *Tcr2039–2043*) [43]. Their presence and organization into an apparent operon suggest that *Thiom. crunogena* XCL-2 could use H<sub>2</sub> as an electron donor for growth, as its close relative *Hydrogenovibrio* does [44,45]. However, attempts to cultivate *Thiom. crunogena* with H<sub>2</sub> as the sole electron donor have not been successful [46]. A requirement for reduced sulfur compounds, even when not used as the primary electron donor, is suggested by the absence of genes encoding the enzymes necessary for assimilatory sulfate reduction (APS reductase and ATP sulfurylase), which are necessary for cysteine synthesis in the absence of environmental sources of thiosulfate or sulfate. Alternatively, this hydrogenase could act as a reductant sink under periods of sulfur and oxygen scarcity, when starch degradation could be utilized to replenish ATP and other metabolite pools (see Central Carbon Metabolism, below).

The reduct partner for the *Thiom. crunogena* XCL-2 hydrogenase is suggested by the structure of the small subunit, which has two domains. One domain is similar to other hydrogenase small subunits, whereas the other is similar to pyridine nucleotide-disulphide oxidoreductases and has both an FAD and NADH binding site. The presence of an NADH binding site suggests that the small subunit itself transfers electrons between H<sub>2</sub> and NAD(H), unlike other soluble hydrogenases, in which this activity is mediated by separate “diaphorase” subunits [43], which *Thiom. crunogena* XCL-2 lacks. The small subunit does not have the twin arginine leader sequence that is found in periplasmic and membrane-associated hydrogenases [47], suggesting a cytoplasmic location for this enzyme.

All 14 genes for the subunits of an electrogenic NADH:ubiquinone oxidoreductase (NDH-1) are present (*Tcr0817–0830*) and are organized in an apparent operon, as in other proteobacteria [48,49]. A cluster of genes encoding an RNF-type NADH dehydrogenase, which is evolutionarily distinct from NDH-1 [50], is present in the *Thiom. crunogena* XCL-2 genome (*Tcr1031–1036*), and may shuttle NADH-derived electrons to specific cellular processes (as in [51]).

In this species, ubiquinone ferries electrons between NADH dehydrogenase and the *bc1* complex; all genes are present for its synthesis, but not for menaquinone. Unlike most bacteria, *Thiom. crunogena* XCL-2 does not synthesize the isopentenyl diphosphate units that make up the lipid portion of ubiquinone via the deoxyxylulose 5-phosphate pathway. Instead, most of the genes of the mevalonate pathway (HMG-CoA synthase, *Tcr1719*; HMG-CoA reductase, *Tcr1717*; mevalonate kinase/phosphomevalonate kinase, *Tcr1732, Tcr1733*; and diphosphomevalonate decarboxylase, *Tcr1734*) [52] are present. The single “missing” gene, for acetyl-CoA acetyltransferase, may not be necessary, because HMG-CoA reductase may also catalyze this reaction as it does in *Enterococcus faecalis* [53]. Interestingly, the mevalonate pathway is found in Archaea and eukaryotes, and is common among Gram-positive bacteria [52,54]. Thus far, the only other proteobacterium to have this pathway is from the alpha class, *Paracoccus sungvaninificans* [55]. Examination of unpublished genome data from the Integrated Microbial Genomes Web page (http://img.jgi.doe.gov), and queries of Genbank (http://www.ncbi.nlm.nih.gov/Genbank) did not uncover evidence for a complete set of genes for the mevalonate pathway in other proteobacteria.

The three components of the *bc1* complex are represented by three genes in an apparent operon, in the typical order (Rieske iron-sulfur subunit; cytochrome b subunit; cytochrome *c1* subunit; *Tcr0991–1001*) [49].

Consistent with its microaerophilic lifestyle and inability to use nitrate as an electron acceptor [6], the only terminal oxidase present in the *Thiom. crunogena* XCL-2 genome is a *cbb*<sub>3</sub>-type cytochrome *c* oxidase (*Tcr1963–5*). To date, *Helicobacter pylori* is the only other sequenced organism that has solely a *cbb*<sub>3</sub>-type oxidase, and this has been proposed to be an adaptation to growth under microaerophilic conditions [49], since *cbb*<sub>3</sub>-type oxidase has a higher affinity for oxygen than *aa*<sub>3</sub>-type oxidase does [56].

In searching for candidate cytochrome proteins that facilitate electron transfer between the Sox system and the *bc1* complex and *cbb*<sub>3</sub> cytochrome *c* oxidase, the genome was analyzed to identify genes that encode proteins with heme-coordinating motifs (CxxCH). This search yielded 28 putative heme-binding proteins (Table S1), compared to 54 identified in the genome of *Thiob. denitrificans* [21]. Thirteen of these genes encode proteins that were predicted to reside in the periplasm, two of which (*Tcr0628* and *Tcr0629*) were deemed particularly promising candidates as they met the following criteria: (1) they were not subunits of other cytochrome-containing systems, (2) they were small enough to serve as efficient electron shuttles, (3) they were characterized beyond the level of hypothetical or conserved hypothetical, and (4) they were present in *Thiob. denitrificans*, which also has both a Sox system as well as *cbb*<sub>3</sub> cytochrome *c* oxidase, and had not been implicated in other cellular functions in this organism. *Tcr0628* and *Tcr0629* both belong to the COG2863 family of cytochrome c553, which are involved in major catabolic pathways in numerous proteobacteria. Interestingly, genes *Tcr0628* and *Tcr0629*, which are separated by a 147-bp spacer that includes a Shine-Delgarno
sequence, are highly likely paralogs, and a nearly identical gene tandem was also identified in the genome of *Thiothrix denitrificans* (Tbd2026 and Tbd2027). A recent comprehensive phylogenetic analysis of the cytochrome c555 proteins, including the mono-heme cytochromes from *Thiom. crunogena* and *Thiothrix denitrificans*, revealed existence of a large protein superfamily that also includes proteins in the COG4654 cytochrome c551/c552 protein family (M. G. Klotz protein superfamily that also includes proteins in the COG family) are the key electron carriers that connect *Noc0751*, *NE0736*, and *Neut2204*, of the cytochrome c electrons will be used for reverse fixation processes (FixP, subunit III) in the *bc1* complex and cytochrome c oxidase.

Given the presence of these electron transport complexes and electron carriers, a model for electron transport chain function is presented here (Figure 2). When thiosulfate or sulfide are acting as the electron donor, the Sox system will introduce electrons into the electron transport chain at the level of cytochrome c [32]. Most will be oxidized by the *cbb* type cytochrome c oxidase to create a proton potential. Some of the cytochrome c electrons will be used for reverse electron transport to ubiquinone and NADH by the *bc1* complex and NADH:ubiquinone oxidoreductase. The NADH created by reverse electron transport must contribute to the cellular NADPH pool for use in biosynthetic pathways. No apparent ortholog of either a membrane-associated [58] or soluble [59] transhydrogenase is present. A gene encoding a NADH kinase is present (Tcr1633), and it is possible that it is also capable of phosphorylating NADH, as some other bacterial NADH kinases are [60].

**Carbon Dioxide Uptake and Fixation**

*Thiom. crunogena* XCL-2, like many species of cyanobacteria [63], has a carbon-concentrating mechanism, in which active dissolved inorganic carbon uptake generates intracellular concentrations that are as much as 100× higher than extracellular [14]. No apparent homologs of any of the cyanobacterial bicarbonate or carbon dioxide transport systems are present in this genome. *Thiom. crunogena* XCL-2 likely recruited bicarbonate and perhaps carbon dioxide transporters from transporter lineages evolutionarily distinct from those utilized by cyanobacteria. Three carbonic anhydrase genes are present (one α-class: *Tcr1545*; and two β-class: *Tcr0421* and *Tcr0841* [64–66]), one of which (α-class) is predicted to be periplasmic and membrane-associated, and may keep the periplasmic dissolved inorganic carbon pool at chemical equilibrium despite selective uptake of carbon dioxide or bicarbonate. One β-class enzyme gene is located near the gene for a form II Rubisco (see below) and may be co-expressed with it when the cells are grown under high-CO2 conditions. The other β-class (formerly ε-class; [66]) carbonic anhydrase is a member of a carboxysome operon and likely functions in this organism’s carbon-concentrating mechanism. Unlike many other bacteria [67], the gene encoding the sole SulP-type ion transporter (Tcr1533) does not have a carbonic anhydrase gene adjacent to it.

The genes encoding the enzymes of the Calvin-Benson-Bassham (CBB) cycle are all present. Three ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) enzymes are encoded in the genome: two form I (Fl) Rubisco (s) (*Tcr0427–8* and *Tcr0838–9*) and one form II (FlI) Rubisco (O) (*Tcr0424*). The two FlI Rubisco large subunit genes are quite similar to each other, with gene products that are 80% identical at the amino acid level. The FlII Rubisco shares only 30% identity in amino acid sequence with both FlI enzymes. The operon structure for each of these genes is similar to *Hydrogenovibrio marinus* [68]: one FlI operon includes Rubisco structural genes (*cbbL* and *cbbS*) followed by genes encoding proteins believed to be important in Rubisco assembly (*cbbQ* and *cbbQ*; *Tcr429–30*) [69,70]. The other FlI operon is part of an α-type carboxysomal operon (*Tcr0840–6*) [71] that includes carboxysome shell protein genes (*csoS1, csoS2*, and *csoS3* (encoding a β-type carbonic anhydrase [65,66]). In the FlII Rubisco operon, *cbbM* (encoding FlII Rubisco) is followed by *cbbO* and *cbbQ* genes, which in turn are followed by a gene encoding a β-class carboxysomal anhydrase (*Tcr0421–3*) [64]. Differing from *Hy. marinus*, the noncarboxysomal FlI and FlII Rubisco operons are juxtaposed and divergently transcribed, with two genes encoding LysR-type regulatory proteins between them (*Tcr0425–6*).

The genes encoding the other enzymes of the CBB cycle are scattered in the *Thiom. crunogena* XCL-2 genome, as in *Hy. marinus* [68]. This differs from facultative autotrophic proteobacteria, in which these genes are often clustered together and co-regulated [72–74]. Based on data from dedicated studies of CBB operons from a few model organisms, it has been suggested that obligate autotrophs like *Hy. marinus* do not have CBB cycle genes organized into an apparent operon, because these genes are presumably constitutively expressed and therefore do not need to be coordinately repressed [68].

Experimental evidence suggests that the CBB cycle is...
constitutively expressed in *Thiom. crunogena* XCL-2. This species cannot grow chemoorganoheterotrophically with acetate, glucose, or yeast extract as the carbon and energy source ([10]: Table S2). When grown in the presence of thiosulfate and dissolved inorganic carbon, RubisCO activities were high both in the presence and absence of these organic carbon sources in the growth medium (Table S3).

Many sequenced genomes from autotrophic bacteria have recently become available and provide a unique opportunity to determine whether CBB gene organization differs among autotrophs based on their lifestyle. Indeed, for all obligate autotrophs, RubisCO genes are not located near the genes encoding the other enzymes of the CBB cycle (Figure 6; Table S4). For example, the distance on the chromosome of these organisms between the genes encoding the only two enzymes unique to the CBB cycle, RubisCO (cbbLS and/or cbbM) and phosphoribulokinase (cbbP), ranges from 139–899 kilobase pairs (kbp) in Proteobacteria, and 151–3,206 kbp in the Cyanobacteria. In contrast, for most facultative autotrophs, cbbP and cbbLS and/or cbbM genes are near each other (Figure 6); in most cases, they appear to coexist in an operon. In the facultative autotroph *Rhodospirillum rubrum*, the cbbM and cbbP genes occupy adjacent, divergently transcribed operons (cbbRM and cbbEFPT). However, these genes are coordinately regulated, since binding sites for the regulatory protein cbbR are present between the operons [75]; perhaps they are coordinately repressed by a repressor protein that binds there as well. The lack of CBB enzyme operons in obligate autotrophs from the Alpha-, Beta-, and Gammaproteobacteria, as well as the Cyanobacteria, may reflect a lack of selective pressure for these genes to be juxtaposed in their chromosomes for ease of coordinate repression during heterotrophic growth.

### Central Carbon Metabolism

3-Phosphoglyceraldehyde generated by the Calvin-Benson-Bassham cycle enters the Embden-Meyerhof-Parnass pathway in the middle, and some carbon must be shunted in both directions to generate the carbon “backbones” for lipid, protein, nucleotide, and cell wall synthesis (Figure 7). All of the enzymes necessary to direct carbon from 3-phosphoglyceraldehyde to fructose-6-phosphate and glucose are en-
coded by this genome, as are all of the genes needed for starch synthesis. To convert fructose 1,6-bisphosphate to fructose 6-phosphate, either fructose bisphosphatase or phosphofructokinase could be used, as this genome encodes a reversible PP\textsubscript{i}-dependent phosphofructokinase (Tcr1583)\cite{76,77}. This store of carbon could be sent back through glycolysis to generate metabolic intermediates to replenish levels of cellular reductant (see below). Genes encoding all of the enzymes necessary to convert 3-phosphoglyceraldehyde to phosphoenolpyruvate and pyruvate are present, and the pyruvate could enter the citric acid cycle (CAC) via pyruvate dehydrogenase, as genes encoding all three subunits of this complex are represented (Tcr1001–3)\cite{151}. Activity could be measured with cell-free extracts of cultures grown in the presence and absence of glucose (M. Hügler and S. M. Sievert, unpublished data).

All of the genes necessary for an oxidative CAC are potentially present, as in some other obligate autotrophs and methanotrophs\cite{18,78}. However, some exceptions from the canonical CAC enzymes seem to be present. The \textit{T. crunogena} XCL-2 genome encodes neither a 2-oxoglutarate dehydrogenase nor a typical malate dehydrogenase, but it does have potential substitutions: a 2-oxoacid:acceptor oxidoreductase (\(\alpha\ and \(\beta\) subunit genes in an apparent operon, Tcr1709–10)\cite{499} and malate:quinone oxidoreductase (Tcr1873). 2-Oxoacid:acceptor oxidoreductase is reversible, unlike 2-oxoglutarate dehydrogenase, which is solely oxidative\cite{79,81}. An overall oxidative direction for the cycle is suggested by malate:quinone oxidoreductase. This membrane-associated enzyme donates the electrons from malate oxidation to the membrane quinone pool and is irreversible, unlike malate dehydrogenase, which donates electrons to NAD\textsuperscript{+}\cite{80}. The 2-oxoacid:acceptor oxidoreductase shows high similarity to the well-characterized 2-oxoglutarate:acceptor oxidoreductase of \textit{Thauera aromatica}\cite{82}, suggesting that it might catalyze the conversion 2-oxoglutarate rather than pyruvate as a substrate. However, cell-free extracts of cells grown autotrophically in the presence and absence of glucose were unable to grow well as organoheterotrophs were boxed. Abbreviations and accession numbers for the 16S sequences used to construct the cladogram are as follows: \textit{A. ehrlichei}, \\textit{Alkalilimnicola ehrlichei}, AF406554; \textit{Brady. sp.}, \textit{Bradyrhizobium sp.}, AF338169; \textit{B. japonicum}, \textit{Bradyrhizobium japonicum}, D13430; \textit{B. xenovorans}, \textit{Burkholderia xenovorans}, U86373; D. aromatica, \textit{Declaromonas aromatica}, AV032610; M. magnetum, \textit{Magnetospirillum magnetum}, D17514; M. capsulatus, \textit{Methyllococcus capsulatus} \textit{BATH}, AF331869; N. hamburgensis, \textit{Nitrooccus hamburgensis}, L11663; N. winogradskyi, \textit{Nitrobacter winogradskyi}, L11661; N. oceani, \textit{Nitosococcus oceani}, AF363287; N. europaee, \textit{Nitrosonomas europaee}, BX321856; N. multiformis, \textit{Nitrospira multiformis}, L35509; P. denitrificans, \textit{Paracoccus denitrificans}, X69159; R. sphaeroides, \textit{Rhodobacter sphaeroides}, CP000144; R. ferrireducens, \textit{Rhodotherax ferrireducens}, AF435948; R. palustris, \textit{Rhodopsudomonas palustris}, NC 00296; R. rubrum, \textit{Rhodospirillum rubrum}, D30778; R. gelatinus, \textit{Rubrivivax gelatinus}, M60862; S. meliloti, \textit{Sinosrhizobium meliloti}, D14509; T. denitrificans, \textit{Thiobacillus denitrificans}, AJ43144; T. crunogena, \textit{Thiomicrospira crunogena}, AF064545.

The cladogram was based on an alignment of 1,622 bp of the 16S rRNA genes, and is the most parsimonious tree (length 2,735) resulting from a heuristic search with 100 replicate random step-wise addition and TBR branch swapping (PAUP\textsuperscript{4.0b10} \cite{113}). Sequences were aligned using ClustalW \cite{114}, as implemented in BioEdit. Percent similarities and identities for \textit{cbbL}, \textit{cbbM}, and \textit{cbbP} gene products, as well as gene locus tags, are provided as supporting information (Table S4). DOI: 10.1371/journal.pbio.0040383.g006
glucose have neither 2-oxoglutarate- nor pyruvate:acceptor oxidoreductase activity (M. Hugler and S. M. Sievert, unpublished data); thus, the CAC does not appear to be complete under these conditions.

A wishbone-shaped reductive citric acid pathway is suggested by this apparent inability to catalyze the interconversion of succinyl-CoA and 2-oxoglutarate. However, even though genes are present encoding most of the enzymes of the reductive arm of the reductive citric acid pathway, from oxaloacetate to succinyl-CoA (phosphoenolpyruvate carboxylase, Tcr1521; fumarate hydratase, Tcr1384; succinate dehydrogenase/fumarate reductase, Tcr2029–31; succinyl-CoA synthetase; Tcr1373–4), the absence of malate dehydrogenase and malic enzyme genes, and the presence of a gene encoding malate:quinone-oxidoreductase (MQO) suggests a blockage of the reductive path as well.

A hypothesis for glycolysis/gluconeogenesis/CAC function is presented here to reconcile these observations (Figure 7). Under conditions in which reduced sulfur compounds and oxygen are sufficiently plentiful to provide cellular reductant and ATP for the Calvin cycle and other metabolic pathways, some carbon would be directed from glyceraldehyde 3-phosphate through gluconeogenesis to starch, whereas some would be directed to pyruvate and an incomplete CAC to meet the cell’s requirements for 2-oxoglutarate, oxaloacetate, and other carbon skeletons. Succinyl-CoA synthesis may not be required, because in most bacteria [83], this genome encodes the enzymes of an alternative pathway for porphyrin synthesis via 5-amino levulinic (glutamyl-tRNA synthetase, Ter1216; glutamyl tRNA reductase, Ter0390; glutamate 1-semialdehyde 2,1 aminomutase; Ter0888). Should environmental conditions shift to sulfide scarcity, cells could continue to generate ATP, carbon skeletons, and cellular reductant by hydrolyzing the starch and sending it through glycolysis and a full oxidative CAC. Should oxygen become scarce instead, cells could send carbon skeletons derived from starch through the incomplete CAC and oxidize excess NADH via the cytoplasmic NiFe hydrogenase, which would also maintain a membrane proton potential via intracellular proton consumption. Clearly, the exact regulation of the CAC under different growth conditions promises to be an interesting topic for future research.

Genes encoding isocitrate lyase and malate synthase are missing, indicating the absence of a glyoxylate cycle, and consistent with this organism’s inability to grow with acetate as the source of carbon (Table S2).

Nitrogen uptake and assimilation are described in Protocol S1 and Table S5.

Phosphorus Uptake

T. crunogena XCL-2 has all of the genes for the low affinity Pit system (Tcr0543–4) and an operon encoding the high affinity Pst system for phosphate uptake (Tcr0537–9) [84]. T. crunogena XCL-2 may also be able to use phosphonate as a phosphorus source, as it has an operon, phnFDCE-GHIJKLMNP (Tcr2078–90), encoding phosphonate transporters and the enzymes necessary to cleave phosphorus-carbon bonds (Figure 8). This phosphonate operon is flanked on either side by large (>6,500 bp) 100% identical direct repeat elements. These elements encode three predicted CDSs (Tcr2074–6 and Tcr2091–3): a small hypothetical, and two large (each >2,500 amino acids [aa] in length) CDSs with limited similarity to a phage-like integrase present in Desulfuromonas acetoxidans, including a domain involved in breaking and rejoining DNA (DBR-1 and DBR-2). It is interesting to note that two homologs found in the draft sequence of the high GC (65%) Gammaproteobacterium Azotobacter vinelandii AvOP have a similar gene organization to the large putative integrases DBR-1/DBR-2. Directly downstream of the first copy of this large repeat element (and upstream of the phosphonate operon) lies another repeat, one of the four IS911-related IS3-family insertion sequences [85] present in this genome (Figure 1). Along with the presence of the transposasintegrate genes and the flanking large repeat element (likely an IS element), the strikingly different G+C of this entire region (39.6%) and the direct repeats (35.9%)
phosphorous. Any excess phosphate accumulated by that may periodically experience a scarcity of inorganic XCL-2 could be stored as polyphosphate granules, crunogena XCL-2 a competitive advantage in an environment crunogena because polyphosphate kinase and exopolyphosphatase genes dissolved organic phosphorus [87], might provide phosphonates, which constitute a substantial fraction of closest relative (Figure S2). The potential capability to use denitrificans, another sulfur-oxidizing bacterium, being the gamma- and betaproteobacteria, with the sequence of Trichodesmium erythraeum has been horizontally acquired. Interestingly, it appears that the phosphonate operon from the marine cyanobacterium Tr. erythraeum, the exception of PhnJ protein of Tr. erythraeum, the tRNA—and its promoter region—and the 5’ ends of the two direct repeat elements, particularly the closest element (Figure 8). Taken together, it is proposed that this entire region has been horizontally acquired. Interestingly, it appears that the phosphonate operon from the marine cyanobacterium Trichodesmium erythraeum was also acquired by horizontal gene transfer [86]. Phylogenetic analyses reveal that the PhnJ protein of Thiom. crunogena XCL-2 falls into a cluster that, with the exception of Tr. erythraeum, contains sequences from gamma- and betaproteobacteria, with the sequence of Thiob. denitrificans, another sulfur-oxidizing bacterium, being the closest relative (Figure S2). The potential capability to use phosphonates, which constitute a substantial fraction of dissolved organic phosphorus [87], might provide Thiom. crunogena XCL-2 a competitive advantage in an environment that may periodically experience a scarcity of inorganic phosphorous. Any excess phosphate accumulated by Thiom. crunogena XCL-2 could be stored as polyphosphate granules, because polyphosphate kinase and exopolypophatase genes are present (Tcr1891–2).

Regulatory and Signaling Proteins

Despite its relative metabolic simplicity as an obligate autotroph, Thiom. crunogena XCL-2 allocates a substantial fraction of its protein-encoding genes (8.9%) to regulatory and signaling proteins (Table 2). In order to determine whether this was typical for a marine obligately chemolithoautotrophic gammaproteobacterium, the numbers of regulatory and signaling protein-encoding genes from this organism were compared to the only other such organism sequenced to date, Nitrosoccus oceani ATCC 19707 [88]. It was of interest to determine whether the differences in their habitats (Thiom. crunogena: attached, and inhabiting a stochastic hydrothermal vent environment, vs. N. oceani: planktonic, in a comparatively stable open-ocean habitat; [89]) would affect the sizes and compositions of their arsenals of regulatory and signaling proteins. Noteworthy differences between the two species include a high proportion of genes with EAL and GGDEF domains in Thiom. crunogena XCL-2 compared to N. oceani (Table 2). These proteins catalyze the hydrolysis and synthesis of cyclic diguanylate, suggesting the importance of this compound as an intracellular signaling molecule in Thiom. crunogena XCL-2 [90]. In some species the abundance of intracellular cyclic diguanylate dictates whether the cells will express genes that facilitate an attached vs. planktonic lifestyle [90]. Given that Thiom. crunogena was isolated by collecting scrapings from hydrothermal vent surfaces [6,15], perhaps cyclic diguanylate has a similar function in Thiom. crunogena as well.

Many of these EAL and GGDEF-domain proteins, and other predicted regulatory and signaling proteins, have PAS domains (Table 2; Table S6), which often function as redox and/or oxygen sensors by binding redox or oxygen-sensitive ligands (e.g., heme and FAD [91]). Twenty PAS-domain proteins predicted from Thiom. crunogena XCL-2’s genome sequence include four methyl-accepting chemotaxis proteins (MCPs) (see below), three signal transduction histidine kinases, six diguanylate cyclases, and seven diguanylate cyclase/phosphodiesterases. N. oceani has 14 predicted gene products with PAS/PAC domains; notable differences from Thiom. crunogena XCL-2 are an absence of PAS/PAC domain MCPs, and fewer PAS/PAC domain proteins involved in cyclic diguanylate metabolism (seven diguanylate cyclase/phosphodiesterases).

Despite its metabolic and morphological simplicity, Thiom. crunogena XCL-2 has almost as many genes encoding transcription factors (52) as the cyst and zoogloea-forming N. oceani does (76; Table 2 [89]). Indeed, most free-living bacteria compared to the genome average (43.1%) suggest that this region may have been acquired by horizontal gene transfer.

Figure 8. Thiom. crunogena XCL-2 Phosphonate Operon

The DBR-1 genes are identical to each other, as are the DBR-2 genes. Gene abbreviations are: chp, conserved hypothetical protein; DBR-1 and -2, DNA breaking-rejoining enzymes; hyp, hypothetical protein; phnFDCGHIJKLMNP, phosphonate operon. An asterisk (*) marks the location of a region (within and upstream of tRNA-phe) with a high level of similarity to the 5’ ends of the two direct repeat sequences noted in the figure. The transposase and integrase are actually a single CDS separated by a frameshift.

DOI: 10.1371/journal.pbio.0040383.g008
have a considerably lower frequency of genes encoding regulatory and signaling proteins (5.6% in *N. oceani* [88]; 5%–6% in other species [19]). Other organisms with frequencies similar to *Thiom. crunogena* XCL-2 (8.6%) include the metabolically versatile *Rhodopseudomonas palustris* (9.3% [19]). Although *Thiom. crunogena* XCL-2 is not metabolically versatile, it has several apparent operons that encode aspects of its structure and metabolism that are likely to enhance growth under certain environmental conditions (e.g., carboxysomes, phosphonate metabolism, assimilatory nitrate reductase, and hydrogenase). Perhaps the relative abundance of regulatory and signaling protein-encoding genes in *Thiom. crunogena* XCL-2 is a reflection of the remarkable temporal and spatial heterogeneity of its hydrothermal vent habitat.

**Chemotaxis**

Genes encoding the structural, regulatory, and assembly-related components of *Thiom. crunogena* XCL-2’s polar flagellae are organized into *flg* (Tcr1464–77) and *fla/flflf* clusters, similar to *Vibrio* spp. [92]. However, the *fla/flflf* cluster is split into two separate subclusters in *Thiom. crunogena* XCL-2 (Tcr0739–47 and Tcr1431–53).

Fourteen genes encoding MCPs are scattered throughout the genome, which is on the low end of the range of MCP gene numbers found in the genomes of gammaproteobacteria. The function of MCPs is to act as nutrient and toxin sensors that communicate with the flagellar motor via the CheA and CheY proteins [93]. As each MCP is specific to a particular nutrient or toxin, it is not surprising that *Thiom. crunogena* XCL-2 has relatively few MCPs, because its nutritional needs as an autotroph are rather simple. Interestingly, however, the number of MCP genes is high for obligately autotrophic proteobacteria (Table 2; Figure 9), particularly with respect to those containing a PAS domain or fold (Figure 9). The relative abundance of MCPs in *Thiom. crunogena* XCL-2 may be an adaptation to the sharp chemical and redox gradients and temporal instability of *Thiom. crunogena* XCL-2’s hydrothermal vent habitat [4].

**Adhesion**

A cluster of genes encoding pilin and the assembly and secretion machinery for type IV pili is present (*flp* *tadE* *epaBCFEF* *tadCBD; Tcr1722–30). In *Actinobacillus actinomycetemcomitans* and other organisms, these fimbriae mediate tight

### Table 2. *Thiom. crunogena* XCL-2 and *N. oceani* ATCC 19707 Regulatory and Signaling Proteins

<table>
<thead>
<tr>
<th>Number:</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thiom. crunogena</strong></td>
<td><strong>N. oceani</strong></td>
</tr>
<tr>
<td>72</td>
<td>104</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>52</td>
<td>76</td>
</tr>
<tr>
<td>128</td>
<td>75</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>34</td>
<td>29</td>
</tr>
<tr>
<td>200</td>
<td>179</td>
</tr>
</tbody>
</table>

*A list of locus tags for these genes is present in Table S6.
DOI: 10.1371/journal.pbio.0040383.t002*
adherence to a variety of substrates [94]. *Thiom. crunogena* was originally isolated from a biofilm [6]. Adhesion within biofilms may be mediated by these fimbrae.

### Heavy Metal Resistance

Despite being cultivated from a habitat that is prone to elevated concentrations of toxic heavy metals including nickel, copper, cadmium, lead, and zinc [95,96], *Thiom. crunogena* XCL-2’s arsenal of heavy metal efflux transporter genes does not distinguish it from *Escherichia coli* and other Gammaproteobacteria. It has 11 sets of resistance-nodulation-cell division superfamily (RND)-type transporters, five cation diffusion facilitator family (CDF) transporters, and six P-type ATPases, far fewer than the metal-resistant *Ralstonia metallidurans* (20 RND, three CDF, and 20 P-type [97]), and lacking the arsenate, cadmium, and mercury detoxification systems present in the genome of hydrothermal vent heterotroph *Idiomarina loihiensis* [98]. To verify this surprising result, *Thiom. crunogena* XCL-2 was cultivated in the presence of heavy metal salts to determine its sensitivities to these compounds (Table 3). Indeed, *Thiom. crunogena* XCL-2 is not particularly resistant to heavy metals; instead, it is more sensitive to them than *E. coli* [99]. Similar results were found for hydrothermal vent archaea [100]; for these organisms, the addition of sulfide to the growth medium was found to enhance their growth in the presence of heavy metal salts, and it was suggested that, in situ at the vents, sulfide might “protect” microorganisms from heavy metals by complexing with metals or forming precipitates with them [100]. Potentially, this strategy is utilized by *Thiom. crunogena* XCL-2. Alternatively, hydrothermal fluid at its mesophilic habitat may be so dilute that heavy metal concentrations do not get high enough to necessitate extensive adaptations to detoxify them.

### Conclusions

Many abilities are apparent from the genome of *Thiom. crunogena* XCL-2 that are likely to enable this organism to survive the spatially and temporally complex hydrothermal vent environment despite its simple, specialized metabolism.
The genomes become available. These mentioned here and others as more autotrophic opportunity use genomics techniques to address the hypothesis of the genome has provided an unprecedented view into the scarcity during its transit. Differential expression of portions of the CAC may guide it to microhabitats with characteristics favorable to its growth. For both species, the concentration at which growth ceased is listed.

Instead of having multiple metabolic pathways, *Thiom. crunogena* XCL-2 appears to have multiple adaptations to obtain autotrophic substrates. Fourteen MCPs presumably guide it to microhabitats with characteristics favorable to its growth, and type IV pili may enable it to live in an attached lifestyle once it finds these favorable conditions. A larger-than-expected arsenal of regulatory proteins may enable this organism to regulate multiple mechanisms for coping with variations in inorganic nutrient availability. Its three RubisCO genes, three carbonic anhydrase genes, and carbon-concentrating mechanism likely assist in coping with oscillations in environmental CO2 availability, while multiple ammonium transporters, nitrate reductase, low- and high-affinity phosphate uptake systems, and potential phosphate use, may enable it to cope with uncertain supplies of these macronutrients.

In contrast, systems for energy generation are more limited, with only one, i.e., Sox, or possibly two, i.e., Sox plus SQR, systems for sulfur oxidation and a single low oxygen–adapted terminal oxidase (ebb2-type). Instead of having a branched electron transport chain with multiple inputs and outputs, this organism may use the four PAS-domain or -fold MCPs to guide it to a portion of the chemiole at which its simple electron transport chain functions. It is worth noting, in this regard, that *Thiob. denitrificans*, which has several systems for sulfur oxidation, has fewer MCPs than *Thiom. crunogena* XCL-2 (Figure 9). Differential expression of portions of the CAC may enable it to survive periods of reduced sulfur or oxygen scarcity during its “transit” to more favorable microhabitats.

Up to this point, advances in our understanding of the biochemistry, genetics, and physiology of this bacterium have been hampered by a lack of a genetic system. The availability of the genome has provided an unprecedented view into the metabolic potential of this fascinating organism and an opportunity use genomics techniques to address the hypotheses mentioned here and others as more autotrophic genomes become available.

## Materials and Methods

**Library construction, sequencing, and sequence quality.** Three DNA libraries (with approximate insert sizes of 3, 7, and 35 kb) were sequenced using the whole-genome shotgun method as previously described [18]. Paired-end sequencing was performed at the Production Genomics Facility of the Joint Genome Institute (JGI), generating greater than 50,000 reads and resulting in approximately 13× depth of coverage. Approximately 400 additional finishing reads were sequenced to close gaps and address base quality issues. Assemblies were accomplished using the PHRED/PHRAP/CONSEED suite [101–103], and gap closure of this assembly, resolution of some sequence errors, and sequence polishing were performed as previously described [18].

**Gene identification and annotation.** Two independent annotations were undertaken: one by the Genome Analysis and System Modeling Group of the Life Sciences Division at Oak Ridge National Laboratory (ORNL), and the other by the University of Bielefeld Center for Biotechnology (CeBiTec). After completion, the two annotations were subjected to a side-by-side comparison, in which discrepancies were examined and manually edited.

Annotation by ORNL proceeded similarly to [18] and is briefly described here. Genes were predicted using GLIMMER [104] and CRITICA [105]. The lists of predicted genes were merged with the start site from CRITICA being used when stop sites were identical. The predicted CDSs were translated and submitted to a BLAST analysis against the KEGG database [106]. The BLAST analysis was used to evaluate overlaps and alternative start sites. Genes with large overlaps where both had good (1e–40) BLAST hits were left for manual resolution. Remaining overlaps were resolved manually and a QA process was used to identify frameshifted, missing, and pseudogenes. The resulting list of predicted genes was translated, and these amino acid sequences were used to query the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Pfam and TIGRFam were run with scores > trusted cutoff scores for the hidden Markov models (HMMs). Product assignments were made based on the hierarchy of TIGRFam, Pfam, Smart (part of InterPro), UniProt, KEGG, and COG databases.

Annotation by CeBiTec began by calling genes using the REGANOR strategy [107], which is based on training GLIMMER [104] with a positive training set created by CRITICA [105]. Predicted CDSs were translated, and the amino acid sequences were used to query the NCBI nonredundant database, SwissProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. Results were collated and presented via GenDB [http://www.cebitec.uni-bielefeld.de/groups/brl/software/gendb"] for manual verification. For each gene, the list of matches to databases was examined to deduce the gene product. Specific functional assignments suggested by matches with SwissProt and the NCBI nonredundant database were only accepted if they covered over 75% of the gene’s length, had an e-value < 0.001, and were supported by hits to curated databases (Pfam or TIGRFam, with scores greater than the trusted cutoff scores for the HMMs), or were consistent with gene context in the genome (e.g., membership in a potential operon with other genes with convincing matches to curated databases). When it was not possible to clarify the function of a gene based on matches in SwissProt and the nonredundant database, related evolutionary relationships were membership in a Pfam with a score greater than the trusted cutoff score for the family HMM), genes were annotated as members of gene families.

When it was not possible to infer function or family membership, genes were annotated as encoding hypothetical or conserved hypothetical proteins. If at least three matches from three other species that covered >75% of the gene’s length were retrieved from SwissProt and the nonredundant database, the genes were annotated as encoding conserved hypothetical proteins. Otherwise, the presence of a Shine-Dalgarno sequence upstream from the predicted start site was verified and the gene was annotated as encoding a hypothetical protein. For genes encoding either hypothetical or conserved hypothetical proteins, the cellular localization of their potential gene products was inferred based on TMHMM and SignalP [109,110]. When transmembrane alpha helices were predicted by TMHMM, the gene product was annotated as a predicted membrane protein. When SignalP Sigpep probability and max cleavage site probability were both >0.75, and no other predicted transmembrane regions were present, the gene was annotated as a predicted periplasmic or secreted protein.

**Comparative genomics.** All CDSs for this genome were used to query the TransportDB database [111]. Matches were assigned to transporter families to facilitate comparisons with other organisms within the TransportDB database (http://www.membranetransport.org). To compare operon structure for genes encoding the Calvin-Benson cycle, amino acid biosynthesis, phosphate metabolism, and to find all of the genes encoding MCPs, BLAST-queries of the microbial genomes included in the Integrated Microbial Genomes database were conducted [112]. Comparison of operon structure was

<table>
<thead>
<tr>
<th>Heavy Metal Ion</th>
<th><em>Thiom. crunogena</em></th>
<th>E. coli</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hg&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Cu&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>Ag&lt;sup&gt;1+&lt;/sup&gt;</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Cd&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>0.05</td>
<td>0.5</td>
</tr>
<tr>
<td>Co&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Ni&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>0.1</td>
<td>1</td>
</tr>
<tr>
<td>Zn&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cr&lt;sup&gt;3+&lt;/sup&gt;</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mn&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

*Thiom. crunogena* XCL-2 was cultivated on solid thiosulfate-supplemented artificial seawater media with metal salts added to the final concentration listed (0.01 to 20 mM).

DOI: 10.1371/journal.pbio.0040383.t003

XCL-2 was cultivated on solid thiosulfate-supplemented artificial seawater media with metal salts added to the final concentration listed (0.01 to 20 mM).
Supporting Information

Figure S1. Phylogenetic Relationships of SoxC Sequences of *Thiom. crunogena* XCL-2 with SoxC/SorA Sequences of Selected Bacteria

Sequences were aligned using the program package MacVector. Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10. At the base of the three main groups, bootstrap values (1,000 replicates) are given for the branch grouping *Thiob. denitrificans* and *Thiom. crunogena*. Bootstrap values are depicted only at the base of the three main groups.

Neighbor-joining and parsimony trees based on the predicted amino acid sequences were calculated using PAUP 4.0b10. Bootstrap values (1,000 replicates; neighbor-joining/parsimony analyses) are depicted only at the base of the three main groups and for the branch grouping *Thiob. denitrificans* and *Thiom. crunogena*.

Found at DOI: 10.1371/journal.pbio.0040383.sg001 (759 KB TIF).

Figure S2. Phylogenetic Relationships of PhnJ Sequences of *Thiom. crunogena* with PhnJ Sequences of Selected Bacteria

Table S1. Proteins with a Heme-Coordinating Motif (CxxCH)

Table S2. Growth of *Thiom. crunogena* XCL-2 on Solid Artificial Seawater Medium Supplemented with Carbon and Electron Sources

Table S3. *Thiom. crunogena* XCL-2 RubisCO Activity When Grown in the Presence and Absence of Organic Carbon

Table S4. Percent Similarities and Identities of Proteobacterial *cbbL*, *cbbM*, and *cbbP* Genes

References


