

# Lawrence Berkeley National Laboratory

## Lawrence Berkeley National Laboratory

### Title

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

### Permalink

<https://escholarship.org/uc/item/8tk6v01c>

### Authors

Scott, Kathleen M.

Sievert, Stefan M.

Abril, Fereniki N.

et al.

### Publication Date

2006-08-23

Peer reviewed



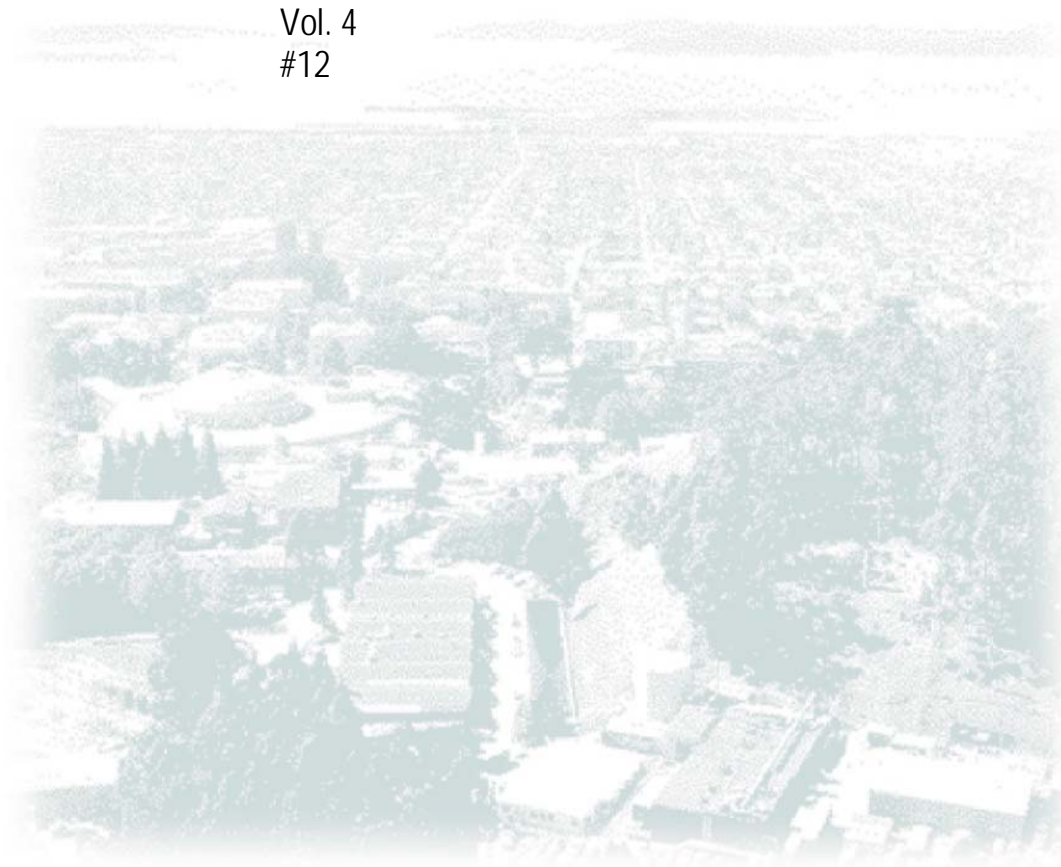
## ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

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

Author(s): Kathleen M. Scott, Stefan M. Sievert, et al

Division: Genomics

Journal Name: PLOS Biology  
November 14, 2006  
Vol. 4  
#12



1  
2           The Genome of Deep-Sea Vent Chemolithoautotroph  
3           *Thiomicrospira crunogena* XCL-2  
4  
5

6 Running head: *T. crunogena* genome  
7

8 **Kathleen M. Scott<sup>\*1</sup>, Stefan M. Sievert<sup>2</sup>, Fereniki N. Abril<sup>1</sup>, Lois A. Ball<sup>1</sup>,**  
9 **Chantell J. Barrett<sup>1</sup>, Rodrigo A. Blake<sup>1</sup>, Amanda J. Boller<sup>1</sup>, Patrick S. G.**  
10 **Chain<sup>3,4</sup>, Justine A. Clark<sup>1</sup>, Carisa R. Davis<sup>1</sup>, Chris Detter<sup>4</sup>, Kimberly F. Do<sup>1</sup>,**  
11 **Kimberly P. Dobrinski<sup>1</sup>, Brandon I. Faza<sup>1</sup>, Kelly A. Fitzpatrick<sup>1</sup>, Sharyn K.**  
12 **Freyermuth<sup>5</sup>, Tara L. Harmer<sup>6</sup>, Loren J. Hauser<sup>7</sup>, Michael Hügler<sup>2</sup>, Cheryl A.**  
13 **Kerfeld<sup>8</sup>, Martin G. Klotz<sup>9</sup>, William W. Kong<sup>1</sup>, Miriam Land<sup>7</sup>, Alla Lapidus<sup>4</sup>,**  
14 **Frank W. Larimer<sup>7</sup>, Dana L. Longo<sup>1</sup>, Susan Lucas<sup>4</sup>, Stephanie A. Malfatti<sup>3,4</sup>,**  
15 **Steven E. Massey<sup>1</sup>, Darlene D. Martin<sup>1</sup>, Zoe McCuddin<sup>10</sup>, Folker Meyer<sup>11</sup>,**  
16 **Jessica L. Moore<sup>1</sup>, Luis H. Ocampo Jr.<sup>1</sup>, John H. Paul<sup>12</sup>, Ian T. Paulsen<sup>13</sup>,**  
17 **Douglas K. Reep<sup>1</sup>, Qinghu Ren<sup>13</sup>, Rachel L. Ross<sup>1</sup>, Priscila Y. Sato<sup>1</sup>,**  
18 **Phaedra Thomas<sup>1</sup>, Lance E. Tinkham<sup>1</sup>, and Gary T. Zeruth<sup>1</sup>**  
19  
20

21 Biology Department, University of South Florida, Tampa, Florida USA<sup>1</sup>; Biology Department,  
22 Woods Hole Oceanographic Institution, Woods Hole, Massachusetts USA<sup>2</sup>; Lawrence Livermore  
23 National Laboratory, Livermore, California USA<sup>3</sup>; Joint Genome Institute, Walnut Creek, California  
24 USA<sup>4</sup>; Department of Biochemistry, University of Missouri, Columbia, Missouri USA<sup>5</sup>; Division of  
25 Natural Sciences and Mathematics, The Richard Stockton College of New Jersey, Pomona, New  
26 Jersey USA<sup>6</sup>; Oak Ridge National Laboratory, Oak Ridge, Tennessee USA<sup>7</sup>; Molecular Biology  
27 Institute, University of California, Los Angeles, California USA<sup>8</sup>; University of Louisville, Louisville  
28 USA<sup>9</sup>; The Monsanto Company, Ankeny, IA USA<sup>10</sup>; Center for Biotechnology, Bielefeld  
29 University, Germany<sup>11</sup>; College of Marine Science, University of South Florida, St. Petersburg,  
30 Florida USA<sup>12</sup>; The Institute for Genomic Research, Rockville, Maryland USA<sup>13</sup>  
31

32 \*Corresponding author. Mailing address: 4202 East Fowler Avenue; SCA 110; Tampa, FL  
33 33620. Phone: (813)974-5173. Fax: (813)974-3263. E-mail: [kscott@cas.usf.edu](mailto:kscott@cas.usf.edu).

34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58

**(Summary)**

**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 bp), 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 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. *T. crunogena* XCL-2 is unusual among obligate sulfur oxidizing bacteria in relying on the Sox system for the oxidation of reduced sulfur compounds. A 38 kb prophage is present, and a high level of prophage induction was observed, which may play a role in keeping competing populations of close relatives in check. 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.**

59

## 60 Introduction

61

62 Deep-sea hydrothermal vent communities are sustained by prokaryotic  
63 chemolithoautotrophic primary producers that use the oxidation of electron donors  
64 available in hydrothermal fluid ( $H_2$ ,  $H_2S$ ,  $Fe^{+2}$ ) to fuel carbon fixation [1,2,3]. The  
65 chemical and physical characteristics of their environment are dictated largely by the  
66 interaction of hydrothermal fluid and bottom water. When warm, reductant- and  $CO_2$ -  
67 rich hydrothermal fluid is emitted from fissures in the basalt crust, it creates eddies as it  
68 mixes with cold, oxic bottom water. As a consequence, at areas where dilute  
69 hydrothermal fluid and seawater mix, a microorganism's habitat is erratic, oscillating  
70 from seconds to hours between dominance by hydrothermal fluid (warm; anoxic;  
71 abundant electron donors; 0.02 to > 1mM  $CO_2$ ) and bottom water (2°C; oxic; 0.02 mM  
72  $CO_2$ ) [4,5].

73 Common chemolithoautotrophic isolates from these "mixing zones" from  
74 hydrothermal vents include members of the genus *Thiomicrospira*, a group which  
75 originally included all marine, spiral-shaped sulfur oxidizing bacteria. Subsequent  
76 analyses of 16S rDNA sequences have revealed the polyphyletic nature of this group;  
77 members of *Thiomicrospira* are distributed among the gamma and epsilon classes of the  
78 Proteobacteria. *T. crunogena*, a member of the cluster of *Thiomicrospiras* in the gamma  
79 class, was originally isolated from the East Pacific Rise [6]. Subsequently, *T. crunogena*  
80 strains were cultivated or detected with molecular methods from deep-sea vents in both  
81 the Pacific and Atlantic, indicating a global distribution for this phylotype [7]. Molecular  
82 methods in combination with cultivation further confirmed the ecological importance of  
83 *Tms. crunogena* and closely related species at deep-sea and shallow-water hydrothermal  
84 vents [8,9].

85 To provide the energy necessary for growth and cell maintenance, *T. crunogena*  
86 XCL-2 and its close relatives *Tms. spp.* L-12 and MA-3 are capable of using hydrogen  
87 sulfide, thiosulfate, elemental sulfur, and sulfide minerals (e.g., pyrite, chalcopyrite) as  
88 electron donors; the only electron acceptor they can use is oxygen [6,10,11,12]. A  
89 substantial portion of the proton motive force and ATP generated by sulfur oxidation is  
90 used by this autotrophic species for carbon fixation via the Calvin-Benson-Bassham cycle  
91 (K. Scott, unpubl. data). This genus was originally described as obligately autotrophic,  
92 based on the observations that: 1. growth was not observed when organic compounds  
93 were the sole source of carbon and energy, and 2. carbon fixation rates when grown in the  
94 presence of thiosulfate were not affected by the presence of organic compounds [10].  
95 *Thiomicrospira crunogena* XCL-2 shares these traits as well (K. Scott, unpubl. data).  
96 Interestingly, *T. crunogena* TH-55, which was isolated from the Western Pacific, might  
97 be capable of strict mixotrophic growth on thiosulfate-supplemented liquid media in the  
98 presence of yeast extract, glucose, and acetate, to which no dissolved inorganic carbon  
99 had been added [13]. Perhaps there are substantial differences in carbon metabolism  
100 within the *T. crunogena* phylotype.

101 As an obligate autotroph, *T. crunogena* XCL-2 is likely adapted to cope with  
102 oscillations in the availability of carbon dioxide, reduced sulfur compounds, oxygen,  
103 dissolved inorganic nitrogen and phosphorus. One critical adaptation in this habitat is its  
104 carbon concentrating mechanism [14,15]. This species is capable of rapid growth in the

105 presence of low concentrations of dissolved inorganic carbon, due to an increase in  
106 cellular affinity for both  $\text{HCO}_3^-$  and  $\text{CO}_2$  under low  $\text{CO}_2$  conditions [15]. The ability to  
107 grow under low  $\text{CO}_2$  conditions is likely an advantage when the habitat is dominated by  
108 relatively low  $\text{CO}_2$  seawater. Further adaptations in nutrient acquisition and microhabitat  
109 sensing are likely to be present in this organism.

110 *T. crunogena* XCL-2 [16] is the first deep-sea autotrophic hydrothermal vent  
111 bacterium to have its genome completely sequenced and annotated. Many other  
112 autotrophic bacterial genomes have been examined previously, including several species  
113 of cyanobacteria (e.g., [17,18], nitrifiers [19], purple nonsulfur [20] and green sulfur [21]  
114 photosynthetic bacteria, as well as an obligately chemolithoautotrophic sulfur-oxidizer  
115 [22] and a hydrogen-oxidizer [23]. These genomes have provided insight into the  
116 evolution of autotrophy among four of the seven phyla of Bacteria known to have  
117 autotrophic members.

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

129  
130

## 131 **Results/Discussion**

132  
133

### 133 **Genome structure**

134 *T. crunogena* XCL-2 has a single chromosome consisting of 2.43 Mbp, with a GC  
135 content of 43.1% and a high coding density (90.6 %; Figure 1). The GC skew shifts near  
136 the gene encoding the DnaA protein (located at ‘noon’ on the circular map; *Tcr0001*),  
137 and thus the origin of replication is likely located nearby. One region with a deviation  
138 from the average %GC contains a phosphonate operon and has several other features  
139 consistent with its acquisition via horizontal gene transfer (see ‘Phosphorus Uptake’  
140 below). Many genes could be assigned a function with a high degree of confidence  
141 (Table 1), and a model for cell function based on these genes is presented (Figure 2).

142 Three rRNA operons are present, and two of them, including their intergenic  
143 regions, are 100% identical. In the third rRNA operon, the 16S and 5S genes are 100%  
144 identical to the other two, but the 23S gene has a single substitution. The intergenic  
145 regions of this third operon also has several substitutions compared to the other two, with  
146 three substitutions between the tRNA-Ile-GAT and tRNA-Ala-TGC genes, six  
147 substitutions between the tRNA-Ala-TGC and 23S genes, and one substitution between  
148 the 23S and 5S genes. Having three rRNA operons may provide additional flexibility for  
149 rapid shifts in translation activity in response to a stochastic environment, and may  
150 contribute to this species’ rapid doubling times [6]. Forty-three tRNA genes were

151 identified by tRNA-scan SE [24] and Search For RNAs. An additional region of the  
152 chromosome was identified by Search For RNAs, the 3' end of which is 57% identical  
153 with the sequence of the tRNA-Asn-GTT gene, but has a 47 nucleotide extension of the  
154 5' end, and is a likely tRNA pseudogene.

155

## 156 Phylogeny

157 The majority of the predicted genes in the *T. crunogena* XCL-2 genome are most  
158 similar to genes from other members of the Proteobacteria (79%). As expected based on  
159 its membership in the gammaproteobacteria, the majority of its genes have highest  
160 identities with genes present in other members of this class (57%). Interestingly, a  
161 substantial number have closest matches in the Betaproteobacteria (13%), which reflects  
162 the basal position, within the Gammaproteobacteria, of the phylogenetic branch leading  
163 to *T. crunogena* XCL-2 and its relatives [8,16].

164

## 165 Prophage

166 A putative prophage genome was noted in the *T. crunogea* chromosome. This  
167 cluster of phage-like genes was flanked by pseudouridine synthase genes (*rluD*; *Tcr0655*  
168 and *Tcr0704*). The putative prophage is 38,090 bp and contains 54 CDSs, 21 of which  
169 (38.9%) had significant similarity to genes in GenBank. The prophage genome begins  
170 with a tyrosine integrase (*Tcr0656*), and contains a cI-like repressor gene (*Tcr0666*),  
171 features common to lambdoid prophages (Figure 3; [25]). These genes define a probable  
172 “lysogeny module” [26] and are in the opposite orientation from the rest of the phage  
173 genes (the replicative or “lytic module”). Also contained in the lysogeny module is a  
174 cytidine C5 DNA methylase (*Tcr0658*) and a NAD-dependent DNA ligase (*Tcr0663*).  
175 The former gene could be part of a component of a phage-encoded restriction  
176 modification system. Lytic phages often methylate their DNA to protect it from  
177 degradation by host restriction systems. Alternatively it may serve to methylate host  
178 DNA to protect it from degradation by alternate phage-encoded restriction systems.  
179 Phage-encoded ligases have been thought to be involved in non-homologous DNA end-  
180 joining events as part of illegitimate recombination mechanisms. Such mechanisms may  
181 contribute to the mosaic nature of phage genomes [27].

182 The lytic half of the prophage genome encodes putative genes involved in DNA  
183 replication and phage assembly (Figure 3). Beginning with a putative DNA primase  
184 (*Tcr0668*) is a cluster of genes interpreted to represent an active or remnant DNA  
185 replication module (including an exonuclease of DNA polymerase, a hypothetical DNA  
186 binding protein, and a terminase large subunit; *Tcr0669*, *0670*, *0672*). Terminases serve  
187 to cut the phage DNA in genome sized fragments prior to packaging. Beyond this are  
188 eight CDSs of unknown function, and then two CDSs involved in capsid assembly,  
189 including the portal protein (*Tcr0679*) and a minor capsid protein (*Tcr0680*) similar to  
190 GPC of  $\lambda$ . Portal proteins are ring-like structures in phage capsids through which the  
191 DNA enters the capsid during packaging [28]. In  $\lambda$ , the GPC protein is a peptidase (S49  
192 family) that cleaves the capsid protein from a scaffolding protein involved in the capsid  
193 assembly process [29]. Although no major capsid protein is identifiable from  
194 bioinformatics, capsid proteins are often difficult to identify from sequence information  
195 in marine phages [30]. A cluster of P2-like putative tail assembly and structural genes  
196 follows the capsid assembly genes. The general organization of these genes (tail fiber, tail

197 shaft and sheath, and tape measure; *Tcr0691*; *Tcr0690*; *Tcr0695*; *Tcr0698*) is also P2-like  
198 [25]. The complexity of these genes (10 putative CDSs involved in tail assembly) and the  
199 strong identity score for a contractile tail sheath protein strongly argues that this prophage  
200 was a member of the *Myoviridae*, ie. phages possessing a contractile tail. The final gene  
201 in the prophage-like sequence was similar to a phage late control protein D, gpD  
202 (*Tcr0700*). In  $\lambda$ , gpD plays a role in the expansion of the capsid to accommodate the  
203 entire phage genome [31].

204 Three additional CDSs are found between the putative gpD-encoding gene and  
205 the pseudouridine synthase gene, two of which (a protein tyrosine phosphatase, *Tcr0702*;  
206 ribonuclease E, *Tcr0703*) had strong similarities to bacterial sequences in GenBank. The  
207 ribonuclease E gene showed great sequence identity (>50%, e-value = 0.0) to a wide  
208 range of marine bacterial homologs, while the tyrosine phosphatase was similar to those  
209 from alkalophilic, thermophilic, or marine bacteria. It is hypothesized that these genes  
210 were transduced by the temperate phage through specialized transduction events.

211 The high similarity of the CDSs to lambdoid (lysogeny and replication genes) and  
212 P2-like (tail module) temperate coliphages is surprising and unprecedented in marine  
213 prophage genomes [32]. A major frustration encountered in marine phage genomics is the  
214 low similarity of CDSs to anything in GenBank, making the interpretation of the  
215 biological function extremely difficult. The lambdoid siphophages are generally members  
216 of the *Siphoviridae* whereas the P2-like phages are *Myoviridae*, which the *T. crunogena*  
217 XCL-2 prophage is predicted to be. Such a mixed heritage is often the result of the  
218 modular evolution of phages. The general genomic organization of the *T. crunogena*  
219 XCL-2 prophage-like element (integrase, repressor, DNA replicative genes, terminase,  
220 portal, capsid, tail genes) is common to several known prophages, including those of  
221 *Staphylococcus aureus* (ie.  $\phi$ Mu50B), *Streptococcus pyogenes* (prophages 370.3 and  
222 370.2), and *S. thermophilus* (prophage O1205; [33]).

223 Prophages are abundant in bacterial genomes (approx. 60% of the sequenced  
224 bacterial genomes contain prophages; Rob Edwards, personal communication). Often  
225 considered dangerous molecular time bombs that can kill the host upon induction, they  
226 also confer advantageous traits to the host through the process of conversion. Although  
227 the beneficial function of the *T. crunogena* XCL-2 prophage-like element is not known, it  
228 may result from the DNA methylating gene it contains. Alternatively, a relatively high  
229 level of spontaneous prophage induction was observed in this isolate ( $\sim 10^8 - 10^9$ /ml),  
230 though it is important to note that the majority of cells remain intact (Mobberly, Paul and  
231 Scott, unpubl.). It is possible that the prophage may serve to lyse closely related  
232 competitors sharing the same environment; *T. crunogena* XCL-2 that remain lysogenic  
233 will not be impacted by the released virus as the prophage confers immunity to  
234 superinfection.

235

## 236 Redox substrate metabolism and electron transport

237 Genes are present in this genome that encode all of the components essential to  
238 assemble a fully functional Sox-system that performs sulfite-, thiosulfate-, sulfur-, and  
239 hydrogen-sulfide dependent cytochrome c reduction, namely, SoxXA (*Tcr0604*,  
240 *Tcr0601*), SoxYZ (*Tcr0603*, *Tcr0602*), SoxB (*Tcr1549*), and SoxCD (*Tcr0156*, *Tcr0157*)  
241 [34,35]. This well-characterized system for the oxidation of reduced sulfur compounds  
242 has been studied in facultatively chemolithoautotrophic, aerobic, thiosulfate-oxidizing



243 alphaproteobacteria, including *Paracoccus versutus* GB17, *Thiobacillus versutus*,  
244 *Starkeya novella* and *Pseudoaminobacter salicylatoxidans* [34,36] and references  
245 therein). This model involves a periplasmic multienzyme complex that is capable of  
246 oxidizing various reduced sulfur compounds completely to sulfate. Genes encoding  
247 components of this complex have been identified, and it has further been shown that  
248 these so-called “sox” genes form extensive clusters in the genomes of the aforementioned  
249 bacteria. Essential components of the Sox-system have also been identified in genomes  
250 of other bacteria known to be able to use reduced sulfur compounds as electron donors,  
251 resulting in the proposal that there might be a common mechanism for sulfur oxidation  
252 utilized by different bacteria [34,36]. Interestingly, *T. crunogena* XCL-2 appears to be  
253 the first obligate chemolithoautotrophic sulfur-oxidizing bacterium to rely on the Sox  
254 system for oxidation of reduced sulfur compounds.

255 Genome analyses also reveal the presence of a putative sulfide:quinone reductase  
256 gene (*Tcr1170*; SQR). This enzyme is present in a number of phototrophic and  
257 chemotrophic bacteria and is best characterized from *Rhodobacter capsulatus* [37]. In  
258 this organism it is located on the periplasmic surface of the cytoplasmic membrane,  
259 where it catalyzes the oxidation of sulfide to elemental sulfur, leading to the deposition of  
260 sulfur outside the cells. It seems reasonable to assume that SQR in *T. crunogena* XCL-2  
261 performs a similar function, explaining the deposition of sulfur outside the cell under  
262 certain conditions (e.g., low pH or oxygen; [38]). The Sox system, on the other hand, is  
263 expected to result in the complete oxidation of sulfide to sulfate. Switching to the  
264 production of elemental sulfur rather than sulfate has the advantage that it prevents  
265 further acidification of the medium, which ultimately would result in cell lysis. An  
266 interesting question in this regard will be to determine how *T. crunogena* XCL-2  
267 remobilizes the sulfur globules. The dependence on the Sox system, and possibly SQR,  
268 for sulfur oxidation differs markedly from the obligately autotrophic sulfur-oxidizing  
269 betaproteobacterium *Thiobacillus denitrificans*, which has a multitude of pathways for  
270 sulfur oxidation, perhaps facilitating this organism’s ability to grow under aerobic and  
271 anaerobic conditions [22].

272 In contrast to the arrangement in facultatively autotrophic sulfur-oxidizers [36], the  
273 *sox* components in *T. crunogena* XCL-2 are not organized in a single cluster, but in  
274 different parts of this genome: *soxXYZA*, *soxB*, and *soxCD*. In particular, the isolated  
275 location of *soxB* relative to other *sox* genes has not been observed in any other sulfur-  
276 oxidizing organisms. The components of the Sox system that form tight interactions *in*  
277 *vivo* are collocated in apparent operons (*SoxXYZA*, *SoxCD*; [39]), which is consistent  
278 with the ‘molarity model’ for operon function (reviewed in [40]), in which cotranslation  
279 from a single mRNA facilitates interactions between tightly-interacting proteins, and  
280 perhaps correct folding. Perhaps for obligate chemolithotrophs like *T. crunogena* XCL-2  
281 that do not have multiple sulfur oxidation systems, in which *sox* gene expression is  
282 presumably constitutive and not subject to complex regulation [41], *sox* gene  
283 organization into a single operon may not be strongly evolutionarily selected.  
284 Alternatively, the *T. crunogena* XCL-2 *sox* genes may not be constitutively expressed,  
285 and may instead function as a regulon.

286 The confirmation of the presence of a *soxB* gene in *T. crunogena* XCL-2 is  
287 particularly interesting, as it is a departure from previous studies with close relatives.  
288 Attempts to PCR-amplify *soxB* from *T. crunogena* ATCC 700270<sup>T</sup> and *T. pelophila*

289 DSM 1534<sup>T</sup> were unsuccessful [42]. In contrast, a newly isolated *Thiomicrospira* strain  
290 obtained from a hydrothermal vent in the North Fiji Basin, *T. crunogena* HY-62, was  
291 positive, with phylogenetic analyses further revealing that its *soxB* was most closely  
292 related those from alphaproteobacteria, such as *Silicibacter pomeroyi* [42]. The *soxB*  
293 gene from *T. crunogena* XCL-2 falls into a cluster containing the green-sulfur bacterium  
294 *Chlorobium* and the purple-sulfur gammaproteobacterium *Allochromatium vinosum*, and  
295 separate from the cluster containing *soxB* from *S. pomeroyi* and *T. crunogena* HY-62  
296 (Figure 4). This either indicates that *T. crunogena* XCL-2 has obtained its *soxB* gene  
297 through lateral gene transfer from different organisms, or that the originally described  
298 *soxB* gene in *T. crunogena* HY-62 was derived from a contaminant. The fact that both  
299 *soxA* and *soxX* from *T. crunogena* XCL-2 also group closely with their respective  
300 homologs from *Chlorobium* spp argues for the latter (data not shown). Also, the negative  
301 result for the two other *Thiomicrospira* strains is difficult to explain in light of the  
302 observation that sulfur oxidation in *T. crunogena* XCL-2 appears to be dependent on a  
303 functional Sox system. It is possible that *T. crunogena* ATCC 700270<sup>T</sup> and *T. pelophila*  
304 DSM 1543<sup>T</sup> also have *soxB* genes, but that the PCR primers did not target conserved  
305 regions of this gene.

306 Up to this point, obligate chemolithoautotrophic sulfur oxidizers were believed to use  
307 a pathway different from the Sox system, i.e., the SI4 pathway [43] or a pathway that  
308 represents basically a reversal of dissimilatory sulfate reduction, by utilizing the enzymes  
309 dissimilatory sulfite reductase, APS reductase, and ATP sulfurylase [44]. In this context,  
310 it is interesting to note that *T. crunogena* also seems to lack enzymes for the assimilation  
311 of sulfate, i.e., ATP sulfurylase, APS kinase, PAPS reductase, and a sirohaem-containing  
312 sulfite reductase, indicating that it depends on reduced sulfur compounds for both  
313 dissimilation and assimilation. *T. crunogena* XCL-2 apparently also lacks a  
314 sulfite:acceptor oxidoreductase (SorAB), an enzyme evolutionarily related to SoxCD that  
315 catalyzes the direct oxidation of sulfite to sulfate and that has a wide distribution among  
316 different sulfur-oxidizing bacteria (see Supporting Information). The presence of the Sox  
317 system and the dependence on it in an obligate chemolithoautotroph also raises the  
318 question of the origin of the Sox system. Possibly, this system first evolved in obligate  
319 autotrophs before it was transferred into facultative autotrophs. Alternatively, *T.*  
320 *crunogena* XCL-2 might have secondarily lost its capability to grow heterotrophically.

321 Genes for Ni/Fe hydrogenase large and small subunits are present (*Tcr2037*;  
322 *Tcr2038*), as well as all of the genes necessary for large subunit metal center assembly  
323 (*Tcr2035 - 6*; *Tcr2039 - 2043*) [45]. Their presence and organization into an apparent  
324 operon suggest that *T. crunogena* XCL-2 could use H<sub>2</sub> as an electron donor for growth, as  
325 its close relative *Hydrogenovibrio* does [46,47]. However, attempts to cultivate *T.*  
326 *crunogena* XCL-2 with H<sub>2</sub> as the sole electron donor have not been successful ([48]; K.  
327 Scott, unpubl.data). A requirement for reduced sulfur compounds, even when not used as  
328 the primary electron donor, is suggested by the absence of genes encoding the enzymes  
329 necessary for assimilatory sulfate reduction (APS reductase; ATP sulfurylase), which are  
330 necessary for cysteine synthesis in the absence of environmental sources of thiosulfate or  
331 sulfide. Alternatively, this hydrogenase could act as a reductant sink under periods of  
332 sulfur and oxygen scarcity, when starch degradation could be utilized to replenish ATP  
333 and other metabolite pools (see “Central Carbon Metabolism”, below).

334 The redox partner for the *T. crunogena* XCL-2 hydrogenase is suggested by the  
335 structure of the small subunit, which has two domains. One domain is similar to other  
336 hydrogenase small subunits, while the other is similar to pyridine nucleotide-disulphide  
337 oxidoreductases and has both an FAD and NADH binding site. The presence of a NADH  
338 binding site suggests that the small subunit itself transfers electrons between H<sub>2</sub> and  
339 NAD(H), unlike other soluble hydrogenases, in which this activity is mediated by  
340 separate “diaphorase” subunits [45], which *T. crunogena* XCL-2 lacks. The small  
341 subunit does not have the twin arginine leader sequence that is found in periplasmic and  
342 membrane-associated hydrogenases [49], suggesting a cytoplasmic location for this  
343 enzyme.

344 All 14 genes for the subunits of an electrogenic NADH:ubiquinone  
345 oxidoreductase (NDH-1) are present (*Tcr0817 - 0830*) and are organized in an apparent  
346 operon, as in other proteobacteria [50,51]. A cluster of genes encoding an RNF-type  
347 NADH dehydrogenase, which is evolutionarily distinct from NDH-1 [52], is present in  
348 the *T. crunogena* XCL-2 genome (*Tcr1031 - 1036*), and may shuttle NADH-derived  
349 electrons to specific cellular processes (as in [53]).

350 In this species, ubiquinone ferries electrons between NADH dehydrogenase and  
351 the bc<sub>1</sub> complex; all genes are present for its synthesis, but not for menaquinone. Unlike  
352 most bacteria, *T. crunogena* XCL-2 does not synthesize the isopentenyl diphosphate units  
353 that make up the lipid portion of ubiquinone via the deoxyxylulose 5-phosphate pathway.  
354 Instead, most of the genes of the mevalonate pathway (HMG-CoA synthase, *Tcr1719*;  
355 HMG-CoA reductase, *Tcr1717*; mevalonate kinase/phosphomevalonate kinase, *Tcr1732*,  
356 *Tcr1733*; and diphosphomevalonate decarboxylase, *Tcr1734* [54]) are present. The  
357 single “missing” gene, for acetyl-CoA acetyltransferase, may not be necessary, as HMG-  
358 CoA reductase may also catalyze this reaction as it does in *Enterococcus faecalis* [55].  
359 Interestingly, the mevalonate pathway is found in Archaea, eukaryotes, and is common  
360 among gram positive bacteria [54,56]. Thus far, the only other proteobacterium to have  
361 this pathway is from the alpha class, *Paracoccus zeaxanthinifaciens* [57]. Examination  
362 of unpublished genome data from the Integrated Microbial Genomes webpage  
363 (<http://img.jgi.doe.gov/v1.1/main.cgi>), and queries of Genbank did not uncover evidence  
364 for a complete set of genes for the mevalonate pathway in other proteobacteria.

365 The three components of the bc<sub>1</sub> complex are represented by three genes in an  
366 apparent operon, in the typical order (Rieske iron-sulfur subunit; cytochrome *b* subunit;  
367 cytochrome *c1* subunit; *Tcr0991 - 3*; [51]).

368 Consistent with its microaerophilic lifestyle and inability to use nitrate as an  
369 electron acceptor [6], the only terminal oxidase present in the *T. crunogena* XCL-2  
370 genome is a *cbb*<sub>3</sub>-type cytochrome *c* oxidase (*Tcr1963 - 5*). Neither *bd*- nor *bo*<sub>3</sub>-type  
371 quinol oxidases are present, nor is an *aa*<sub>3</sub>-type cytochrome oxidase. To date,  
372 *Helicobacter pylori* is the only other sequenced organism that has solely a *cbb*<sub>3</sub>-type  
373 oxidase, and this has been proposed to be an adaptation to growth under microaerophilic  
374 conditions [51], since *cbb*<sub>3</sub>-type oxidase has a higher affinity for oxygen than *aa*<sub>3</sub>-type  
375 oxidase does [58].

376 In searching for candidate cytochrome proteins that facilitate electron transfer  
377 between the Sox system and the bc<sub>1</sub> complex and *cbb*<sub>3</sub> cytochrome *c* oxidase, the genome  
378 was analyzed to identify genes that encode proteins with heme-coordinating motifs  
379 (CxxCH). This search yielded 28 putative heme-binding proteins (Table S1), compared to

380 54 identified in the genome of *T. denitrificans* [22]. Thirteen of these genes encode  
381 proteins that were predicted to reside in the periplasm, five were predicted to integrate  
382 into the plasma membrane, and the remaining ten proteins were predicted to reside in the  
383 cytoplasm and are thus not considered as candidates for transferring electrons between  
384 the Sox system and the *bc<sub>1</sub>* complex and *cbb<sub>3</sub>* cytochrome c oxidase (Fig. 2). The five  
385 predicted membrane cytochrome proteins were discarded as well, as three of them  
386 contribute to assembly and function of the *bc<sub>1</sub>* complex and *cbb<sub>3</sub>* cytochrome c oxidase,  
387 one is predicted to be a diguanylate cyclase/phosphodiesterase, while the remaining one  
388 is a conserved hypothetical protein and likely has no role in catabolic electron transfer.

389 Of the thirteen predicted periplasmic proteins encoded by the remaining genes,  
390 two (*Tcr0628*; *Tcr0629*) were deemed particularly promising candidates as they met the  
391 following criteria: 1) they were not subunits of other cytochrome-containing systems, 2)  
392 they were small enough to serve as efficient electron shuttles, 3) they were characterized  
393 beyond the level of hypothetical or conserved hypothetical, and 4) they were present in  
394 *Thiobacillus denitrificans*, which also has both a Sox system as well as *cbb<sub>3</sub>* cytochrome  
395 c oxidase, and had not been implicated in other cellular functions in this organism.  
396 *Tcr0628* and *Tcr0629* both belong to the COG2863 family of cytochromes c553, which  
397 are involved in major catabolic pathways in numerous proteobacteria. Interestingly, genes  
398 *Tcr0628* and *Tcr0629*, which are separated by a 147-pb spacer that includes a Shine-  
399 Delgarno sequence, are highly likely paralogues and a nearly identical gene tandem was  
400 also identified in the genome of *T. denitrificans* (*Tbd2026*, *Tbd2027*). A recent  
401 comprehensive phylogenetic analysis of the cytochrome c553 proteins, including the  
402 mono-heme cytochromes from *T. crunogena* and *T. denitrificans*, revealed existence of a  
403 large protein superfamily that also includes proteins in the COG4654 cytochrome  
404 c551/c552 protein family (M.G. Klotz and A.B. Hooper, unpublished results). In  
405 ammonia-oxidizing bacteria, representatives of this protein superfamily (*NE0102*,  
406 *Neut2204*, *Nmula0344* in the COG4654 protein family; *Noc0751*, *NE0736*, *Neut1650* in  
407 the COG2863 protein family) are the key electron carriers that connect the *bc<sub>1</sub>* complex  
408 with complex IV as well as NO<sub>x</sub>-detoxifying reductases (i.e., NirK, NirS) and oxidases  
409 (i.e., cytochrome P460, cytochrome c peroxidase) involved in nitrifier denitrification  
410 ([59] and references therein). In Epsilonproteobacteria such as *Helicobacter pylori* and  
411 *hepaticus*, cytochromes in this family (*jhp1148*; *HH1517*) interact with the terminal  
412 cytochrome *cbb<sub>3</sub>* oxidase. Therefore, we propose that the expression products of genes  
413 *Tcr0628* and *Tcr0629* likely represent the electronic link between the Sox system and the  
414 *bc<sub>1</sub>* complex and *cbb<sub>3</sub>* cytochrome c oxidase in *T. crunogena*. It appears worthwhile to  
415 investigate experimentally whether the small difference in sequence between these two  
416 genes reflects an adaptation to binding to interaction partners with sites of different redox  
417 potential, namely cytochrome *c<sub>1</sub>* in the *bc<sub>1</sub>* complex and cytochrome FixP (subunit III) in  
418 *cbb<sub>3</sub>* cytochrome c oxidase.

419 Given the presence of these electron transport complexes and electron carriers, a  
420 model for electron transport chain function is presented here (Figure 2). When  
421 thiosulfate or sulfide are acting as the electron donor, the Sox system will introduce  
422 electrons into the electron transport chain at the level of cytochrome c [34]. Most will be  
423 oxidized by the *cbb<sub>3</sub>*-type cytochrome c oxidase to create a proton potential. Some of the  
424 cytochrome c electrons will be used for reverse electron transport to ubiquinone and  
425 NAD<sup>+</sup> by the *bc<sub>1</sub>* complex and NADH:ubiquinone oxidoreductase. The NADH created

426 by reverse electron transport must contribute to the cellular NADPH pool, for use in  
427 biosynthetic pathways. No apparent ortholog of either a membrane-associated [60] or  
428 soluble [61] transhydrogenase is present. A gene encoding a NAD<sup>+</sup> kinase is present  
429 (*Tcr1633*), and it is possible that it is also capable of phosphorylating NADH, as some  
430 other bacterial NAD<sup>+</sup> kinases are [62].

431

#### 432 Transporters and nutrient uptake

433 One hundred sixty nine transporter genes from 40 families are present in the *T.*  
434 *crunogena* XCL-2 genome (Figure 5), comprising 7.7% of the CDSs. This low  
435 frequency of transporter genes is similar to other obligately autotrophic proteobacteria  
436 and cyanobacteria as well as intracellular pathogenic bacteria such as *Xanthomonas*  
437 *axonopodis*, *Legionella pneumophila*, *Haemophilus influenzae*, and *Francisella*  
438 *tularensis* (Figure 5; [63,64]). Most heterotrophic gammaproteobacteria have higher  
439 transporter gene frequencies, up to 14.1% (Figure 5), which likely function to assist in the  
440 uptake of multiple organic carbon and energy sources, as suggested when transporters for  
441 sugars, amino acids and other organic acids, nucleotides and cofactors were tallied  
442 (Figure 5).

443

#### 444 Carbon dioxide uptake and fixation

445 *T. crunogena* XCL-2, like many species of cyanobacteria [65], has a carbon  
446 concentrating mechanism, in which active dissolved inorganic carbon uptake generates  
447 intracellular concentrations that are as much as 100X higher than extracellular [15]. No  
448 apparent homologs of any of the cyanobacterial bicarbonate or carbon dioxide uptake  
449 systems are present in this genome. *T. crunogena* XCL-2 likely recruited bicarbonate  
450 and perhaps carbon dioxide transporters from transporter lineages evolutionarily distinct  
451 from those utilized by cyanobacteria. Three carbonic anhydrase genes are present (one  
452  $\alpha$ -class, *Tcr1545*; two  $\beta$ -class, *Tcr0421*, *Tcr0841* [66,67,68], one of which ( $\alpha$ -class) is  
453 predicted to be periplasmic and membrane-associated, and may keep the periplasmic  
454 dissolved inorganic carbon pool at chemical equilibrium despite selective uptake of  
455 carbon dioxide or bicarbonate. One  $\beta$ -class enzyme gene is located near the gene for a  
456 form II RubisCO (see below) and may be coexpressed with it when the cells are grown  
457 under high-CO<sub>2</sub> conditions. The other  $\beta$ -class (formerly  $\epsilon$ -class; [68]) carbonic  
458 anhydrase is a member of a carboxysome operon and likely functions in this organism's  
459 carbon concentrating mechanism. Unlike many other bacteria [69], the gene encoding  
460 the sole SulP-type ion transporter (*Tcr1533*) does not have a carbonic anhydrase gene  
461 adjacent to it.

462 The genes encoding the enzymes of the Calvin-Benson-Bassham (CBB) cycle are  
463 all present. Three ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO) enzymes  
464 are encoded in the genome: two form I (FI) RubisCOs (*Tcr0427-8* and *Tcr0838-9*) and  
465 one form II (FII) RubisCO (*Tcr0424*). The two FI RubisCO large subunit genes are quite  
466 similar to each other, with gene products that are 80% identical at the amino acid level.  
467 The FII RubisCO shares only 30% identity in amino acid sequence with both FI enzymes.  
468 The operon structure for each of these genes is similar to *Hydrogenovibrio marinus* [70]:  
469 one FI operon includes RubisCO structural genes (*cbbL* and *cbbS*) followed by genes  
470 encoding proteins believed to be important in RubisCO assembly (*cbbO* and *cbbQ*;  
471 *Tcr429 - 30*) [71,72]. The other FI operon is part of an  $\alpha$ -type carboxysome operon

472 (*Tcr0840-6*) [73] that includes carboxysome shell protein genes *csoS1*, *csoS2*, and *csoS3*  
473 (encoding a  $\beta$ -class carbonic anhydrase; [67,68]. In the FII RubisCO operon, *cbbM*  
474 (encoding FII RubisCO) is followed by *cbbO* and *cbbQ* genes, which in turn are followed  
475 by a gene encoding a  $\beta$ -class carbonic anhydrase (*Tcr0421 – 3*) [66]. Differing from *H.*  
476 *marinus*, the noncarboxysomal FI and FII RubisCO operons are juxtaposed and  
477 divergently transcribed, with two genes encoding LysR-type regulatory proteins between  
478 them (*Tcr0425-6*).

479 The genes encoding the other enzymes of the CBB cycle are scattered in the *T.*  
480 *crunogena* XCL-2 genome, as in *H. marinus* [70]. This differs from facultative  
481 autotrophic proteobacteria, in which these genes are often clustered together and  
482 coregulated [74,75,76]. Based on data from dedicated studies of CBB operons from a  
483 few model organisms, it has been suggested that obligate autotrophs like *H. marinus* do  
484 not have CBB cycle genes organized into an apparent operon because these genes are  
485 presumably constitutively expressed, and therefore do not need to be coordinately  
486 repressed [70].

487 Experimental evidence suggests that the CBB cycle is constitutively expressed in  
488 *T. crunogena* XCL-2. This species cannot grow heterotrophically ([10]; K. Scott, unpubl.  
489 data). When both thiosulfate and dissolved inorganic carbon are provided, growth yields  
490 are enhanced by glucose or yeast extract (K. Scott, unpubl. data). However, even when  
491 these organic carbon sources are available, RubisCO activity is high (K. Scott, unpubl.  
492 data).

493 Many sequenced genomes from autotrophic bacteria have recently become  
494 available and provide a unique opportunity to determine whether CBB gene organization  
495 differs among autotrophs based on their lifestyle. Indeed, for all obligate autotrophs,  
496 RubisCO genes are not located near the genes encoding the other enzymes of the CBB  
497 cycle (Figure 6). For example, the distance on the chromosome of these organisms  
498 between the genes encoding the only two enzymes unique to the CBB cycle, RubisCO  
499 (*cbbLS* and/or *cbbM*) and phosphoribulokinase (*cbbP*), ranges from 139 – 899 kbp in  
500 Proteobacteria, and 151 – 3206 kbp in the Cyanobacteria. In contrast, for most  
501 facultative autotrophs, *cbbP* and *cbbLS* and/or *cbbM* genes are near each other (Figure 6);  
502 in most cases, they appear to coexist in an operon. In the facultative autotroph  
503 *Rhodospirillum rubrum*, the *cbbM* and *cbbP* genes occupy adjacent, divergently  
504 transcribed operons (*cbbRM* and *cbbEFPT*). However, these genes are coordinately  
505 regulated, since binding sites for the regulatory protein *cbbR* are present between the  
506 operons [77]; perhaps they are coordinately repressed by a repressor protein that binds  
507 there as well. The lack of CBB enzyme operons in obligate autotrophs from the Alpha-  
508 Beta-, and Gammaproteobacteria, as well as the cyanobacteria, may reflect a lack of  
509 selective pressure for these genes to be juxtaposed in their chromosomes for ease of  
510 coordinate repression during heterotrophic growth.

511

## 512 Central carbon metabolism

513 3-phosphoglyceraldehyde generated by the Calvin-Benson-Bassham cycle enters  
514 the Embden-Meyerhoff-Parnass pathway in the middle, and some carbon must be shunted  
515 in both directions to generate the carbon “backbones” for lipid, protein, nucleotide, and  
516 cell wall synthesis (Figure 7). All of the enzymes necessary to direct carbon from 3-  
517 phosphoglyceraldehyde to fructose-6-phosphate and glucose are encoded by this genome,

518 as are all of the genes needed for starch synthesis. To convert fructose 1,6-bisphosphate  
519 to fructose 6-phosphate, either fructose bisphosphatase or phosphofructokinase could be  
520 used, as this genome encodes a reversible PP<sub>i</sub>-dependent phosphofructokinase (*Tcr1583*)  
521 [78,79]. This store of carbon could be sent back through glycolysis to generate metabolic  
522 intermediates to replenish levels of cellular reductant (see below). Genes encoding all of  
523 the enzymes necessary to convert 3-phosphoglyceraldehyde to phosphoenolpyruvate and  
524 pyruvate are present, and the pyruvate could enter the citric acid cycle via pyruvate  
525 dehydrogenase, as genes encoding all three subunits of this complex are represented  
526 (*Tcr1001 - 3*) and activity could be measured with cell-free extracts of cultures grown in  
527 the presence and absence of glucose (Hügler and Sievert, unpublished data).

528 All of the genes necessary for an oxidative citric acid cycle (CAC) are potentially  
529 present, as in some other obligate autotrophs and methanotrophs [19,80]. However, some  
530 exceptions from the canonical CAC enzymes seem to be present. The *T. crunogena*  
531 XCL-2 genome encodes neither a 2-oxoglutarate dehydrogenase nor a typical malate  
532 dehydrogenase, but it does have potential substitutions: a 2-oxoacid:acceptor  
533 oxidoreductase ( $\alpha$  and  $\beta$  subunit genes in an apparent operon, *Tcr1709 - 10*), and malate:  
534 quinone-oxidoreductase (*Tcr1873*), as in *Helicobacter pylori* [81,82]. 2-oxoacid:acceptor  
535 oxidoreductase is reversible, unlike 2-oxoglutarate dehydrogenase, which is solely  
536 oxidative [81,83]. An overall oxidative direction for the cycle is suggested by malate:  
537 quinone oxidoreductase. This membrane-associated enzyme donates the electrons from  
538 malate oxidation to the membrane quinone pool and is irreversible, unlike malate  
539 dehydrogenase, which donates electrons to NAD<sup>+</sup> [82]. The 2-oxoacid:acceptor  
540 oxidoreductase shows high similarity to the well-characterized 2-oxoglutarate:acceptor  
541 oxidoreductase of *Thauera aromatica* [84], suggesting that it might catalyze the  
542 conversion 2-oxoglutarate rather than pyruvate as a substrate. However, cell-free extracts  
543 of cells grown autotrophically in the presence and absence of glucose have neither 2-  
544 oxoglutarate- nor pyruvate:acceptor oxidoreductase activity (Hügler and Sievert, unpubl.  
545 data); thus, the citric acid cycle does not appear to be complete under these conditions.

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

554 A hypothesis for glycolysis/gluconeogenesis/citric acid cycle function is  
555 presented here to reconcile these observations (Figure 7). Under conditions where  
556 reduced sulfur compounds and oxygen are sufficiently plentiful to provide cellular  
557 reductant and ATP for the Calvin cycle and other metabolic pathways, some carbon  
558 would be directed from glyceraldehyde 3-phosphate through gluconeogenesis to starch,  
559 while some would be directed to pyruvate and an incomplete citric acid cycle to meet the  
560 cell's requirements for 2-oxoglutarate, oxaloacetate, and other carbon skeletons.  
561 Succinyl-CoA synthesis may not be required, as in most bacteria [85], this genome  
562 encodes the enzymes of an alternative pathway for porphyrin synthesis via 5-amino  
563 levulinate (glutamyl-tRNA synthetase, *Tcr1216*; glutamyl tRNA reductase, *Tcr0390*;

564 glutamate 1-semialdehyde 2,1 aminomutase; *Tcr0888*). Should environmental  
565 conditions shift to sulfide scarcity, cells could continue to generate ATP, carbon  
566 skeletons, and cellular reductant by hydrolyzing the starch and sending it through  
567 glycolysis and a full oxidative citric acid cycle. Should oxygen become scarce instead,  
568 cells could send carbon skeletons derived from starch through the incomplete citric acid  
569 cycle and oxidize excess NADH via the cytoplasmic Ni/Fe hydrogenase, which would  
570 also maintain a membrane proton potential via intracellular proton consumption. Clearly,  
571 the exact regulation of the CAC under different growth conditions promises to be an  
572 interesting topic for future research.

573 Genes encoding isocitrate lyase and malate synthase are missing, indicating the  
574 absence of a glyoxylate cycle, and consistent with this organism's inability to grow with  
575 acetate as the source of carbon (K. Scott, unpubl. data).

576

### 577 Nitrogen and uptake and assimilation

578 *Thiomicrospira crunogena* XCL-2 is capable of growing with nitrate or ammonia  
579 as its nitrogen source ([6]; K. Scott, unpubl. data). Accordingly, it has an apparent  
580 operon encoding the components of a NasFED-type nitrate transporter (*Tcr1153 - 5*) [86],  
581 cytoplasmic assimilatory nitrate (*nasA*; *Tcr1159*) and nitrite reductase (*nirBD*; *Tcr1157-*  
582 *8*) genes, as well as four Amt-family ammonia transporters (*Tcr0954*; *Tcr1340*; *Tcr1500*;  
583 *Tcr2151*).

584 Ammonia originating from environmental sources or produced from nitrate  
585 reduction is incorporated into the *T. crunogena* XCL-2 organic nitrogen pool by  
586 glutamine synthetase and NADPH-dependent glutamate synthase. *T. crunogena* XCL-2  
587 has three different glutamine synthetase genes: one encodes a GlnA-type enzyme  
588 (*Tcr0536*) while the others are both GlnT-type (*Tcr1347*, *Tcr1798*) [87]. Perhaps these  
589 three glutamine synthetase genes are differentially expressed under different nitrogen  
590 conditions. Genes encoding the majority of the enzymes necessary to synthesize all 20  
591 L-amino acids and all five nucleobases were detected (see Supporting Information).

592

### 593 Phosphorus uptake

594 *T. crunogena* XCL-2 has all of the genes for the low affinity PiT system (*Tcr0543*  
595 - 4) and an operon encoding the high affinity Pst system for phosphate uptake (*Tcr0537 -*  
596 *9*) [88]. *T. crunogena* XCL-2 may also be able to use phosphonate as a phosphorus  
597 source, as it has an operon, *phnFDCEGHIJKLMNP* (*Tcr2078 - 90*), encoding  
598 phosphonate transporters and the enzymes necessary to cleave phosphorus-carbon bonds  
599 (Figure 8). This phosphonate operon is flanked on either side by large (>6500bp) 100%  
600 identical direct repeat elements. These elements encode three predicted coding sequences  
601 (*Tcr2074 - 6*; *Tcr2091 - 3*): a small hypothetical, and two large (>2500 aa in length)  
602 coding sequences with limited similarity to a phage-like integrase present in  
603 *Desulfuromonas acetoxidans*, including a domain involved in breaking and rejoining  
604 DNA (DBR-1, DBR-2). It is interesting to note that two homologs found in the draft  
605 sequence of the high GC (~65%) gammaproteobacterium *Azotobacter vinelandii* AvOP  
606 have a similar gene organization to the large putative integrases DBR-1/DBR-2. Directly  
607 downstream of the first copy of this large repeat element (and upstream of the  
608 phosphonate operon) lies another repeat, one of the four IS911-related IS3-family  
609 insertion sequences [89] present in this genome (Figure 1). Along with the presence of



610 the transposase/integrase genes and the flanking large repeat element (likely an IS  
611 element), the strikingly different G+C of this entire region (39.6%) and the direct repeats  
612 (35.9%) compared to the genome average (43.1%) suggest that this region may have been  
613 acquired by horizontal gene transfer.

614 Interestingly, immediately downstream of this island lies another region of  
615 comparatively low G+C (39.6%) that encodes a number of products involved in metal  
616 resistance (e.g., copper transporters and oxidases, heavy metal efflux system). Directly  
617 downstream of this second island lies a phage integrase (*Tcr2121*) adjacent to two  
618 tRNAs, which are known to be common phage insertion sites. Strikingly, there is a high  
619 level of similarity between the 5' region of the first tRNA – and its promoter region – and  
620 the 5' regions of the large repeat elements, particularly the closest element (Figure 8).  
621 Taken together, it is proposed that this entire region has been horizontally acquired.  
622 Interestingly, it appears that the phosphonate operon from the marine cyanobacterium  
623 *Trichodesmium erythraeum* was also acquired by horizontal gene transfer [90].  
624 Phylogenetic analyses reveal that the PhnJ protein of *T. crunogena* XCL-2 falls into a  
625 cluster that, with the exception of *Trichodesmium erythraeum*, contains sequences from  
626 gamma- and betaproteobacteria, with the sequence of *Thiobacillus denitrificans*, another  
627 sulfur-oxidizing bacterium, being the closest relative (see supporting information). The  
628 potential capability to use phosphonates, which constitute a substantial fraction of  
629 dissolved organic phosphorus[91], might provide *T. crunogena* XCL-2 a competitive  
630 advantage in an environment that may periodically experience a scarcity of inorganic  
631 phosphorous. Any excess phosphate accumulated by *T. crunogena* XCL-2 could be  
632 stored as polyphosphate granules, as polyphosphate kinase and exopolyphosphatase  
633 genes are present (*Tcr1891* - 2).

634

### 635 Regulatory and signaling proteins

636 Despite its relative metabolic simplicity as an obligate autotroph, *T. crunogena*  
637 XCL-2 allocates a substantial fraction of its protein-encoding genes (8.4%) to regulatory  
638 and signaling proteins (Table 2). In order to determine whether this was typical for a  
639 marine obligately chemolithoautotrophic gammaproteobacterium, the numbers of  
640 regulatory and signaling protein-encoding genes from this organism were compared to  
641 the only other such organism sequenced to date, *Nitrosococcus oceani* ATCC 19707 [92].  
642 It was of interest to determine whether the differences in their habitats (*T. crunogena*:  
643 attached, and inhabiting a stochastic hydrothermal vent environment, vs. *N. oceani*:  
644 planktonic, in a comparatively stable open ocean habitat; [93]) would affect the sizes and  
645 compositions of their arsenals of regulatory and signaling proteins. Noteworthy  
646 differences between the two species include a high proportion of genes with EAL and  
647 GGDEF domains in *T. crunogena* XCL-2 compared to *N. oceani* (Table 2). These  
648 proteins catalyze the hydrolysis and synthesis of cyclic diguanylate, suggesting the  
649 importance of this compound as an intracellular signaling molecule in *T. crunogena*  
650 XCL-2 [94]. In some species the abundance of intracellular cyclic diguanylate dictates  
651 whether the cells will express genes that facilitate an attached vs. planktonic lifestyle  
652 [94]. Given that *T. crunogena* was isolated by collecting scrapings from hydrothermal  
653 vent surfaces [6,16], perhaps cyclic diguanylate has a similar function in *T. crunogena* as  
654 well.

655 Many of these EAL and GGDEF-domain proteins, and other predicted regulatory  
656 and signaling proteins, have PAS domains (Table 2), which often function as redox  
657 and/or oxygen sensors by binding redox or oxygen-sensitive ligands (e.g., heme, FAD;  
658 [95]). Nineteen PAS-domain proteins predicted from *T. crunogena* XCL-2's genome  
659 sequence include 4 methyl-accepting chemotaxis proteins (see below), 3 signal  
660 transduction histidine kinases, 5 diguanylate cyclases, and 7 diguanylate  
661 cyclase/phosphodiesterases. *N. oceani* has 14 predicted gene products with PAS/PAC  
662 domains; notable differences from *T. crunogena* XCL-2 are an absence of PAS/PAC  
663 domain methyl-accepting chemotaxis proteins, and fewer PAS/PAC domain proteins  
664 involved in cyclic diguanylate metabolism (7 diguanylate cyclase/phosphodiesterases).

665 Despite its metabolic and morphological simplicity, *T. crunogena* XCL-2 has  
666 almost as many genes encoding transcription factors (52) as the cyst and zoogloea-  
667 forming *N. oceani* does (76; Table 2; [93]). Indeed, most free-living bacteria have a  
668 considerably lower frequency of genes encoding regulatory and signaling proteins (5.6%  
669 in *N. oceani* [92]; 5-6% in other species [20]). Other organisms with frequencies similar  
670 to *T. crunogena* XCL-2 (8.6%) include the metabolically versatile *Rhodospseudomonas*  
671 *palustris* (9.3%; [20]). Although *T. crunogena* XCL-2 is not metabolically versatile, it  
672 has several apparent operons that encode aspects of its structure and metabolism that are  
673 likely to enhance growth under certain environmental conditions (e.g., carboxysomes;  
674 phosphonate metabolism; assimilatory nitrate reductase; hydrogenase). Perhaps the  
675 relative abundance of regulatory and signaling protein-encoding genes in *T. crunogena*  
676 XCL-2 is a reflection of the remarkable temporal and spatial heterogeneity of its  
677 hydrothermal vent habitat.

678

## 679 Chemotaxis

680 Genes encoding the structural, regulatory, and assembly-related components of *T.*  
681 *crunogena* XCL-2's polar flagellae are organized into *flg* (*Tcr1464* - 77) and *fla/fli/flh*  
682 clusters, similar to *Vibrio* spp. [96]. However, the *fla/fli/flh* cluster is split into two  
683 separate subclusters in *T. crunogena* XCL-2 (*Tcr0739* - 47; *Tcr1431* - 53).

684 Fourteen genes encoding methyl-accepting chemotaxis proteins (MCPs) are  
685 scattered throughout the genome, which is on the low end of the range of MCP gene  
686 numbers found in the genomes of gammaproteobacteria. The function of MCPs is to act  
687 as nutrient and toxin-sensors that communicate with the flagellar motor via the CheA and  
688 CheY proteins [97]. As each MCP is specific to a particular nutrient or toxin, it is not  
689 surprising that *T. crunogena* XCL-2 has relatively few MCPs, as its nutritional needs as  
690 an autotroph are rather simple. Interestingly, however, the number of MCP genes is high  
691 for obligately autotrophic proteobacteria (Table 2; Figure 9), particularly with respect to  
692 those containing a PAS domain or fold (Figure 9). The relative abundance of MCPs in *T.*  
693 *crunogena* XCL-2 may be an adaptation to the sharp chemical and redox gradients and  
694 temporal instability of *T. crunogena* XCL-2's hydrothermal vent habitat [4].

695

## 696 Adhesion

697 A cluster of genes encoding pilin and the assembly and secretion machinery for  
698 type IV pili is present (*flp tadE cpaBCEF tadCBD*; *Tcr1722* - 30). In *Actinobacillus*  
699 *actinomycetemcomitans* and other organisms, these fimbriae mediate tight adherence to a  
700 variety of substrates [98]. When cultivated in the presence of low pH or oxygen

701 concentrations, *T. crunogena* XCL-2 forms clumps with the elemental sulfur globules  
702 that it excretes under these conditions ([38]; K. Scott, pers. obs.). Furthermore, *T.*  
703 *crunogena* was originally isolated from a biofilm [6]. Adhesion within biofilms may be  
704 mediated by these fimbriae.

705

## 706 Heavy metal resistance

707 Despite being cultivated from a habitat that is prone to elevated concentrations of  
708 toxic heavy metals including nickel, copper, cadmium, lead, and zinc [99,100], *T.*  
709 *crunogena* XCL-2's arsenal of heavy metal efflux transporter genes does not distinguish  
710 it from *E. coli* and other gammaproteobacteria. It has eleven sets of Resistance-  
711 Nodulation-Cell Division superfamily (RND)-type transporters, five Cation Diffusion  
712 Facilitator family (CDF) transporters, and six P-type ATPases, far fewer than the metal-  
713 resistant *Ralstonia metallidurans* (20 RND, 3 CDF, 20 P-type; [101]), and lacking the  
714 arsenate, cadmium, and mercury detoxification systems present in the genome of  
715 hydrothermal vent heterotroph *Idiomarina loihiensis* [102]. To verify this surprising  
716 result, *T. crunogena* XCL-2 was cultivated in the presence of heavy metal salts to  
717 determine its sensitivities to these compounds (Table 3). Indeed, *T. crunogena* XCL-2 is  
718 not particularly resistant to heavy metals; instead, it is more sensitive to them than *E. coli*  
719 [103]. Similar results were found for hydrothermal vent archaea [104]; for these  
720 organisms, the addition of sulfide to the growth medium was found to enhance their  
721 growth in the presence of heavy metal salts, and it was suggested that, *in situ* at the vents,  
722 sulfide might "protect" microorganisms from heavy metals by complexing with metals or  
723 forming precipitates with them [104]. Potentially, this strategy is utilized by *T. crunogena*  
724 XCL-2. Alternatively, hydrothermal fluid at its mesophilic habitat may be so dilute that  
725 heavy metal concentrations do not get high enough to necessitate extensive adaptations to  
726 detoxify them.

727 Indeed, some of these 'metal sequestering' proteins encoded in this genome may  
728 function instead in maintaining a stable supply of metals for enzyme active sites. This  
729 appears to be the case for the copper-binding CopA and B proteins, cytochrome c, and  
730 multicopper oxidase proteins encoded by *Tcr1573 – 6* and *2116 – 3*, which occupy two  
731 ~5900 bp regions that are ~92% identical to each other. The multicopper oxidase genes  
732 encoded by *Tcr1576* and *Tcr2113* may function in the reduction of reactive oxides such  
733 as NO<sub>x</sub> species [105,106,107,108]. Due to the juxtaposition of these four genes, we  
734 hypothesize that the CopA (*Tcr1575* and *Tcr2114*) and CopB (*Tcr1574* and *Tcr2115*)  
735 proteins may function to bind copper to ensure a steady supply of this metal cofactor for  
736 the multicopper oxidase/cytochrome c complex.

737

738

## 739 Conclusions

740 Many abilities are apparent from the genome of *T. crunogena* XCL-2 that are  
741 likely to enable this organism to survive the spatially and temporally complex  
742 hydrothermal vent environment despite its simple, specialized metabolism. Instead of  
743 having multiple metabolic pathways, *T. crunogena* XCL-2 appears to have multiple  
744 adaptations to obtain autotrophic substrates. Fourteen methyl-accepting chemotaxis  
745 proteins presumably guide it to microhabitats with characteristics favorable to its growth,  
746 and type IV pili may enable it to live an attached lifestyle once it finds these favorable

747 conditions. A larger-than-expected arsenal of regulatory proteins may enable this  
748 organism to regulate multiple mechanisms for coping with variations in inorganic  
749 nutrient availability. Its three RubisCO genes, three carbonic anhydrase genes, and  
750 carbon concentrating mechanism likely assist in coping with oscillations in  
751 environmental CO<sub>2</sub> availability, while multiple ammonium transporters, nitrate reductase,  
752 low- and high- affinity phosphate uptake systems, and potential phosphonate use, may  
753 enable it to cope with uncertain supplies of these macronutrients.

754 In contrast, systems for energy generation are more limited, with only one, i.e.,  
755 Sox, or possibly two, i.e., Sox plus SQR, systems for sulfur oxidation and a single low-  
756 oxygen adapted terminal oxidase (*cbb<sub>3</sub>*-type). Instead of having a branched electron  
757 transport chain with multiple inputs and outputs, this organism may use the four PAS-  
758 domain or -fold methyl-accepting chemotaxis proteins to guide it to a portion of the  
759 chemocline where its simple electron transport chain functions. It is worth noting, in this  
760 regard, that *Thiobacillus denitrificans*, which has several systems for sulfur oxidation, has  
761 fewer MCPs than *T. crunogena* XCL-2 (Figure 9). Differential expression of portions of  
762 the citric acid cycle may enable it to survive periods of reduced sulfur or oxygen scarcity  
763 during its ‘transit’ to more favorable microhabitats.

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

770  
771

## 772 **Materials and Methods**

773

774 **Library construction, sequencing, and sequence quality.** Three DNA libraries  
775 (with approximate insert sizes of 3, 7, and 35 kb) were sequenced using the whole-  
776 genome shotgun method as previously described [19]. Paired-end sequencing was  
777 performed at the Production Genomics Facility of the Joint Genome Institute (JGI),  
778 generating greater than 50,000 reads and resulting in approximately 13X depth of  
779 coverage. An additional ~400 finishing reads were sequenced to close gaps and address  
780 base quality issues. Assemblies were accomplished using the PHRED/PHRAP/CONSED  
781 suite [109,110,111], and gap closure, resolution of repetitive sequences and sequence  
782 polishing were performed as previously described [19].

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

789 Annotation by ORNL proceeded similarly to [19] and is briefly described here.  
790 Genes were predicted using GLIMMER [112] and CRITICA [113]. The lists of  
791 predicted genes were merged with the start site from CRITICA being used when stop  
792 sites were identical. The predicted coding sequences were translated and submitted to a

793 BLAST analysis against the KEGG database [114]. The BLAST analysis was used to  
794 evaluate overlaps and alternative start sites. Genes with large overlaps where both had  
795 good ( $1e-40$ ) BLAST hits were left for manual resolution. Remaining overlaps were  
796 resolved manually and a QA process was used to identify frameshifted, missing, and  
797 pseudogenes. The resulting list of predicted coding sequences were translated and these  
798 amino acid sequences were used to query the NCBI nonredundant database, UniProt,  
799 TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. PFam and TIGRFam  
800 were run with scores  $>$  trusted cutoff scores for the HMMs. Product assignments were  
801 made based on the hierarchy of TIGRFam, PRIAM, Pfam, Smart (part of InterPro),  
802 UniProt, KEGG, and COGs.

803 Annotation by CeBiTec began by calling genes using the REGANOR strategy  
804 [115], which is based on training GLIMMER [112] with a positive training set created by  
805 CRITICA [113]. Predicted coding sequences were translated and these amino acid  
806 sequences were used to query the NCBI nonredundant database, SwisProt, TIGRFam,  
807 Pfam, KEGG, COG, and InterPro databases. Results were collated and presented via  
808 GenDB [116] for manual verification. For each gene, the list of matches to databases  
809 was examined to deduce the gene product. Specific functional assignments suggested by  
810 matches with SwisProt and the NCBI nonredundant database were only accepted if they  
811 covered over 75% of the gene length, had an e-value  $<$  0.001, and were supported by hits  
812 to curated databases (Pfam or TIGRFam, with scores  $>$  trusted cutoff scores for the  
813 HMMs), or were consistent with gene context in the genome (e.g., membership in a  
814 potential operon with other genes with convincing matches to curated databases). When  
815 it was not possible to clarify the function of a gene based on matches in SwissProt and  
816 the nonredundant database, but evolutionary relatedness was apparent (e.g., membership  
817 in a Pfam with a score  $>$  trusted cutoff score for the family HMM), genes were annotated  
818 as members of gene families.

819 When it was not possible to infer function or family membership, genes were  
820 annotated as encoding hypothetical or conserved hypothetical proteins. If at least three  
821 matches from three other species that covered  $>$ 75% of the gene's length were retrieved  
822 from SwissProt and the nonredundant database, the genes were annotated as encoding  
823 conserved hypothetical proteins. Otherwise, the presence of a Shine-Dalgarno sequence  
824 upstream from the predicted start codon was verified and the gene was annotated as  
825 encoding a hypothetical protein. For genes encoding either hypothetical or conserved  
826 hypothetical proteins, the cellular location of their potential gene products was inferred  
827 based on TMHMM and SignalP [117,118]. When transmembrane alpha helices were  
828 predicted by TMHMM, the gene product was annotated as a predicted membrane protein.  
829 When SignalP Sigpep probability and max cleavage site probability were both  $>$ 0.75, and  
830 no other predicted transmembrane regions were present, the gene was annotated as a  
831 predicted periplasmic or secreted protein.

832 **Comparative genomics.** All CDSs for this genome were used to query the  
833 TransportDB database [119]. Matches were assigned to transporter families to facilitate  
834 comparisons with other organisms within the TransportDB database  
835 (<http://www.membranetransport.org/>). To compare operon structure for genes encoding  
836 the Calvin-Benson-Bassham cycle, amino acid biosynthesis, phosphonate metabolism,  
837 and to find all of the genes encoding methyl-accepting chemotaxis proteins, BLAST-  
838 queries of the microbial genomes included in the Integrated Microbial Genomes database

839 were conducted [120]. Comparison of operon structure was greatly facilitated by using  
840 the “Show Neighborhoods” function available on the IMG website  
841 (<http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>).

842 **Nucleotide sequence accession number.** The complete sequence of the *T.*  
843 *crunogena* XCL-2 genome is available from the nonredundant database (GenBank  
844 accession number CP000109).  
845

Acknowledgment:

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396

846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891

## References

1. Karl DM, Wirsen CO, Jannasch HW (1980) Deep-sea primary production at the Galápagos hydrothermal vents. *Science* 207: 1345-1346.
2. Edwards KJ, Rogers DR, Wirsen CO, McCollom TM (2003) Isolation and characterization of novel psychrophilic, neutrophilic, Fe-oxidizing, chemolithoautotrophic alpha- and gamma-Proteobacteria from the deep sea. *Applied and Environmental Microbiology* 69: 2906-2913.
3. Kelley DS, Karson JA, Fruh-Green GL, Yoerger DR, Shank TM, et al. (2005) A serpentinite-hosted ecosystem: The lost city hydrothermal field. *Science* 307: 1428-1434.
4. Johnson KS, Childress JJ, Beehler CL (1988) Short term temperature variability in the Rose Garden hydrothermal vent field. *Deep-Sea Res* 35: 1711-1722.
5. Goffredi SK, Childress JJ, Desaulniers NT, Lee RW, Lallier FH, et al. (1997) Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external P-CO<sub>2</sub> and upon proton-equivalent ion transport by the worm. *Journal of Experimental Biology* 200: 883-896.
6. Jannasch H, Wirsen C, Nelson D, 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.
7. Wirsen CO, Brinkhoff T, Kuever J, Muyzer G, Molyneaux S, et al. (1998) Comparison of a new *Thiomicrospira* strain from the Mid-Atlantic Ridge with known hydrothermal vent isolates. *Appl Environ Microbiol* 64: 4057-4059.
8. Muyzer G, A. Teske, C.O. Wirsen, H.W. Jannasch (1995) Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. *Arch Microbiol* 164: 165-172.
9. Brinkhoff T, Sievert SM, Kuever J, Muyzer G (1999) 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.
10. Ruby EG, Wirsen CO, Jannasch HW (1981) Chemolithotrophic Sulfur-Oxidizing Bacteria from the Galapagos Rift Hydrothermal Vents. *Appl Environ Microbiol* 42: 317-324.
11. Ruby EG, Jannasch HW (1982) Physiological characteristics of *Thiomicrospira* sp. strain L-12 isolated from deep-sea hydrothermal vents. *J Bacteriol* 149: 161-165.
12. Wirsen CO, Brinkhoff T, Kuever J, Muyzer G, Jannasch HW, et al. (1998) Comparison of a new *Thiomicrospira* strain from the mid-atlantic ridge with known hydrothermal vent isolates. *Applied and Environmental Microbiology* 64: 4057-4059.
13. Takai K, Hirayama H, Nakagawa T, Suzuki Y, Nealson KH, et al. (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. *International Journal of Systematic and Evolutionary Microbiology* 54: 2325-2333.
14. Scott KM, Bright M, Fisher CR (1998) The burden of independence: Inorganic carbon utilization strategies of the sulphur chemoautotrophic hydrothermal vent



- 892 isolate *Thiomicrospira crunogena* and the symbionts of hydrothermal vent and  
893 cold seep vestimentiferans. *Cah Biol Mar* 39: 379-381.
- 894 15. Dobrinski KP, Longo DL, Scott KM (2005) A hydrothermal vent  
895 chemolithoautotroph with a carbon concentrating mechanism. *J Bacteriol* 187:  
896 5761-5766.
- 897 16. Ahmad A, Barry JP, Nelson DC (1999) Phylogenetic affinity of a wide, vacuolate,  
898 nitrate-accumulating *Beggiatoa* sp. from Monterey Canyon, California, with  
899 *Thioploca* spp. *Appl Environ Microbiol* 65: 270-277.
- 900 17. Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann I, et al. (2003)  
901 Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a  
902 nearly minimal oxyphototrophic genome. *Proc Natl Acad Sci USA* 100: 10020-  
903 10025.
- 904 18. Palenik B, Brahmsha B, F L, Land M, Hauser L, et al. (2003) The genome of a  
905 motile marine *Synechococcus*. *Nature* 424: 1037-1042.
- 906 19. Chain P, Lamerdin J, Larimer F, Regala W, Lao V, et al. (2003) Complete genome  
907 sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph  
908 *Nitrosomonas europaea*. *J Bacteriol* 185: 2759-2773.
- 909 20. Larimer F, Chain P, Hauser L, Lamerdin J, Malfatti S, et al. (2004) Complete genome  
910 sequence of the metabolically versatile photosynthetic bacterium  
911 *Rhodospseudomonas palustris*. *Nature Biotechnology* 22: 55-61.
- 912 21. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, et al. (2002) The complete  
913 genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic,  
914 green-sulfur bacterium. *Proceedings of the National Academy of Sciences of the*  
915 *United States of America* 99: 9509-9514.
- 916 22. Beller HR, Chain PSG, Letain TE, Chakicherla A, Larimer FW, et al. (2006) The  
917 Genome Sequence of the Obligately Chemolithoautotrophic, Facultatively  
918 Anaerobic Bacterium *Thiobacillus denitrificans*. *J Bacteriol* 188: 1473-1488.
- 919 23. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, et al. (1998) The  
920 complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature*  
921 392: 353-358.
- 922 24. Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of  
923 transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955-964.
- 924 25. Casjens S (2003) Prophages and bacterial genomics: What have we learned so far?  
925 *Molec Microbiol* 49: 277-300.
- 926 26. Lucchini S, Desiere F, Brussow H (1999) Similarly organized lysogeny modules in  
927 temperate Siphoviridae from low GC content gram positive bacteria. *Virology*  
928 263: 427-435.
- 929 27. Brussow HC, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial  
930 pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol*  
931 *Molec Biol Rev* 68: 560-602.
- 932 28. Weigle PR, Sampson L, Winn-Stapley D, Casjens SR (2005) Molecular genetics of  
933 bacteriophage P22 scaffolding protein's functional domains. *J Molec Biol* 348:  
934 831-844.
- 935 29. Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB (1982) Nucleotide sequence  
936 of bacteriophage lambda DNA. *J Mol Biol* 162: 729-773.

- 937 30. Rohwer F, Segall A, Steward G, Seguritan V, Breitbart M, et al. (2000) The complete  
938 genomic sequence of the marine phage Roseophage SIO1 shares homology with  
939 non-marine phages. *Limnol Oceanogr* 42: 408-418.
- 940 31. Sternberg N, Weisberg R (1977) Packaging of coliphage lambda DNA. II. The role  
941 of gene D protein. *J Mol Biol* 117: 733-759.
- 942 32. Paul JH, Sullivan MB (2005) Marine phage genomics: What have we learned? *Curr*  
943 *Opin Biotech* 16: 299-307.
- 944 33. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H (2003) Prophage genomics.  
945 *Microbiol Molec Biol Rev* 67: 238-276.
- 946 34. Friedrich CG, Quentmeier A, Bardischewsky F, Rother D, Kraft R, et al. (2000)  
947 Novel genes coding for lithotrophic sulfur oxidation of *Paracoccus pantotrophus*  
948 GB17. *J Bacteriol* 182: 4677-4687.
- 949 35. Rother D, Henrich HJ, Quentmeier A, Bardischewsky F, Friedrich CG (2001) Novel  
950 genes of the sox gene cluster, mutagenesis of the flavoprotein SoxF, and evidence  
951 for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *Journal*  
952 *of Bacteriology* 183: 4499-4508.
- 953 36. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J (2005)  
954 Prokaryotic sulfur oxidation. *Current Opinion in Microbiology* 8: 253-259.
- 955 37. Schutz M, Maldener I, Griesbeck C, Hauska G (1999) Sulfide-quinone reductase  
956 from *Rhodobacter capsulatus*: Requirement for growth, periplasmic localization,  
957 and extension of gene sequence analysis. *Journal of Bacteriology* 181: 6516-6523.
- 958 38. Javor BJ, Wilmot DB, Vetter RD (1990) pH-Dependent metabolism of thiosulfate  
959 and sulfur globules in the chemolithotrophic marine bacterium *Thiomicrospira*  
960 *crunogena*. *Arch Microbiol* 154: 231-238.
- 961 39. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J (2001) Oxidation  
962 of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common  
963 Mechanism? *Appl Environ Microbiol* 67: 2873-2882.
- 964 40. Fani R, Brilli M, Lio P (2005) The origin and evolution of operons: The piecewise  
965 building of the proteobacterial histidine operon. *J Mol Evol* 60: 378-390.
- 966 41. Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-  
967 regulation and not by horizontal gene transfer. *Genome Res* 15: 809-819.
- 968 42. Petri R, Podgorsek L, Imhoff JF (2001) Phylogeny and distribution of the soxB gene  
969 among thiosulfate-oxidizing bacteria. *Fems Microbiology Letters* 197: 171-178.
- 970 43. Kelly DP, Shergill JK, Lu WP, Wood AP (1997) Oxidative metabolism of inorganic  
971 sulfur compounds by bacteria. *Antonie Van Leeuwenhoek International Journal of*  
972 *General and Molecular Microbiology* 71: 95-107.
- 973 44. Nelson DC, Hagen KD (1995) Physiology and Biochemistry of Symbiotic and Free-  
974 Living Chemoautotrophic Sulfur Bacteria. *American Zoologist* 35: 91-101.
- 975 45. Schwartz E, Friedrich B (2005) The H<sub>2</sub>-metabolizing prokaryotes. In: Dworkin M,  
976 editor. *The Prokaryotes: An evolving electronic resource for the microbiological*  
977 *community*, release 314, <http://linkspringer-nycom/link/service/books/10125/>.  
978 New York: Springer-Verlag.
- 979 46. Nishihara H, Yaguchi T, Chung SY, Suzuki K, Yanagi M, et al. (1998) Phylogenetic  
980 position of an obligately chemoautotrophic, marine hydrogen-oxidizing  
981 bacterium, *Hydrogenovibrio marinus*, on the basis of 16S rRNA gene sequences  
982 and two form I RubisCO gene sequences. *Arch Microbiol* 169: 364-368.

- 983 47. Nishihara H, Miyata Y, Miyashita Y, Bernhard M, Pohlmann A, et al. (2001)  
984 Analysis of the molecular species of hydrogenase in the cells of an obligately  
985 chemolithoautotrophic, marine hydrogen-oxidizing bacterium, *Hydrogenovibrio*  
986 *marinus*. Biosci Biotechnol Biochem 65: 2780-2784.
- 987 48. Nishihara H, Igarashi Y, Kodama T (1991) *Hydrogenovibrio marinus* gen. nov., sp.  
988 nov., a marine obligately chemolithoautotrophic hydrogen-oxidizing bacterium.  
989 Int J Syst Bacteriol 41: 130-133.
- 990 49. Dross F, Geisler V, Lenger R, Theis F, Krafft T, et al. (1992) The quinone-reactive  
991 Ni/Fe-hydrogenase of *Wollinella succinogenes*. Eur J Biochem 206: 93-102.
- 992 50. Friedrich T, Scheide D (2000) The respiratory complex I of bacteria, archaea and  
993 eukarya and its module common with membrane-bound multisubunit  
994 hydrogenases. FEBS Letters 479: 1-5.
- 995 51. Smith M, Finel M, Korolik V, Mendz G (2000) Characteristics of the aerobic  
996 respiratory chains of the microaerophiles *Campylobacter jejuni* and *Helicobacter*  
997 *pylori*. Arch Microbiol 174: 1-10.
- 998 52. Steuber J (2001) Na<sup>+</sup> translocation by bacterial NADH:quinone oxidoreductases: an  
999 extension to the complex-I family of primary redox pumps. Biochimica et  
1000 Biophysica Acta 1505: 45-56.
- 1001 53. Kumagai H, Fujiwara T, Matsubara H, Saeki K (1997) Membrane localization,  
1002 topology, and mutual stabilization of the rnfABC gene products in *Rhodobacter*  
1003 *capsulatus* and implications for a new family of energy-coupling NADH  
1004 oxidoreductases. Biochemistry 36: 5509-5521.
- 1005 54. Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: The  
1006 evolution of two ancient and distinct pathways across genomes. Proc Natl Acad  
1007 Sci USA 97: 13172-13177.
- 1008 55. Hedl M, Sutherlin A, Wilding E, Mazzulla M, McDevitt D, et al. (2002)  
1009 *Enterococcus faecalis* acetoacetyl-coenzyme A thiolase/3-hydroxy-3-  
1010 methylglutaryl-coenzyme A reductase, a dual-function protein of isopentenyl  
1011 diphosphate biosynthesis. J Bacteriol 184: 2116.
- 1012 56. Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, et al. (2000)  
1013 Identification, Evolution, and Essentiality of the Mevalonate Pathway for  
1014 Isopentenyl Diphosphate Biosynthesis in Gram-Positive Cocci. J Bacteriol 182:  
1015 4319-4327.
- 1016 57. Humbelin M, Thomas A, Lin J, Li J, Jore J, et al. (2002) Genetics of isoprenoid  
1017 biosynthesis in *Paracoccus zeaxanthinifaciens*. Gene 297: 129-139.
- 1018 58. Preisig O, Zufferey R, Thony-Meyer L, Appleby C, Hennecke H (1996) A high-  
1019 affinity *cbb3*-type cytochrome oxidase terminates the symbiosis-specific  
1020 respiratory chain of *Bradyrhizobium japonicum*. J Bacteriol 178: 1532-1538.
- 1021 59. Hooper AB, Arciero DM, Bergmann D, Hendrich MP (2005) The oxidation of  
1022 ammonia as an energy source in bacteria Dordrecht, the Netherlands: Springer.
- 1023 60. Jackson JB (2003) Proton translocation by transhydrogenase. FEBS Letters 545: 18-  
1024 24.
- 1025 61. Boonstra B, French CE, Wainwright I, Bruce NC (1999) The *udhA* Gene of  
1026 *Escherichia coli* Encodes a Soluble Pyridine Nucleotide Transhydrogenase. J  
1027 Bacteriol 181: 1030-1034.

- 1028 62. Mori S, Kawai S, Shi F, Mikami B, Murata K (2005) Molecular conversion of NAD  
1029 kinase to NADH kinase through single amino acid residue substitution. Journal of  
1030 Biological Chemistry 280: 24104-24112.
- 1031 63. Paulsen IT, Nguyen L, Sliwinski MK, Rabus R, Saier MH (2000) Microbial genome  
1032 analyses: comparative transport capabilities in eighteen prokaryotes. J Mol Biol  
1033 301: 75-100.
- 1034 64. Ren Q, Paulsen IT (2005) Comparative analysis of fundamental differences in  
1035 membrane transport capabilities in prokaryotes and eukaryotes. PLOS  
1036 computational biology 1: 190-201.
- 1037 65. Badger MR, Price GD, Long BM, Woodger FJ (2006) The environmental plasticity  
1038 and ecological genomics of the cyanobacterial CO<sub>2</sub> concentrating mechanism. J  
1039 Exp Bot 57: 249-265.
- 1040 66. Smith KS, Ferry JG (2000) Prokaryotic carbonic anhydrases. FEMS Microbiol Rev  
1041 24: 335-366.
- 1042 67. So AK, Espie GS, Williams EB, Shively JM, Heinhorst S, et al. (2004) A novel  
1043 evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the  
1044 carboxysome shell. J Bacteriol 186: 623-630.
- 1045 68. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, et al. (2006) The  
1046 Structure of beta-Carbonic Anhydrase from the Carboxysomal Shell Reveals a  
1047 Distinct Subclass with One Active Site for the Price of Two. J Biol Chem 281:  
1048 7546-7555.
- 1049 69. Felce J, Saier MH (2004) Carbonic anhydrase fused to anion transporters of the SulP  
1050 family: Evidence for a novel type of bicarbonate transporter. J Mol Microbiol  
1051 Biotechnol 8: 169-176.
- 1052 70. Yoshizawa Y, Toyoda K, Arai H, Ishii M, Igarashi Y (2004) CO<sub>2</sub>-responsive  
1053 expression and gene organization of three ribulose-1,5-bisphosphate  
1054 carboxylase/oxygenase enzymes and carboxysomes in *Hydrogenovibrio marinus*  
1055 strain MH-110. Journal of Bacteriology 186: 5685-5691.
- 1056 71. Hayashi NR, Arai H, Kodama T, Igarashi Y (1997) The novel genes, *cbbQ* and *cbbO*,  
1057 located downstream from the RubisCO genes of *Pseudomonas*  
1058 *hydrogenothermophila*, affect the conformational states and activity of RubisCO.  
1059 Biochem Biophys Res Commun 241: 565-569.
- 1060 72. Hayashi NR, Arai H, Kodama T, Igarashi Y (1999) The *cbbQ* genes located  
1061 downstream of the form I and form II RubisCO genes, affect the activity of both  
1062 RubisCOs. Biochem Biophys Res Commun 266: 177-183.
- 1063 73. Badger M, Hanson D, Price GD (2002) Evolution and diversity of CO<sub>2</sub> concentrating  
1064 mechanisms in cyanobacteria. Functional Plant Biology 29: 161-173.
- 1065 74. Gibson JL, Tabita FR (1996) The molecular regulation of the reductive pentose  
1066 phosphate pathway in Proteobacteria and Cyanobacteria. Arch Microbiol 166:  
1067 141-150.
- 1068 75. Kusian B, Bowien B (1997) Organization and regulation of *cbb* CO<sub>2</sub> assimilation  
1069 genes in autotrophic bacteria. FEMS Microbiol Rev 21: 135-155.
- 1070 76. Shively JM, Van Keulen G, Meijer WG (1998) Something form almost nothing:  
1071 carbon dioxide fixation in chemoautotrophs. Annual Reviews in Microbiology 52:  
1072 191-230.

- 1073 77. Falcone DL, Tabita FR (1993) Complementation analysis and regulation of CO<sub>2</sub>  
1074 fixation gene expression in a ribulose 1,5-bisphosphate carboxylase-oxygenase  
1075 deletion strain of *Rhodospirillum rubrum*. *J Bacteriol* 175: 5066-5077.
- 1076 78. Ding YR, Ronimus RS, Morgan HW (2000) Sequencing, Cloning, and High-Level  
1077 Expression of the pfp Gene, Encoding a P<sub>PPi</sub>-Dependent Phosphofructokinase  
1078 from the Extremely Thermophilic Eubacterium *Dictyoglomus thermophilum*. *J*  
1079 *Bacteriol* 182.
- 1080 79. Ronimus RS, Morgan HW (2001) The biochemical properties and phylogenies of  
1081 phosphofructokinases from extremophiles. *Extremophiles* 5: 357-373.
- 1082 80. Wood AP, Aurikko JP, Kelly DP (2004) A challenge for 21st century molecular  
1083 biology and biochemistry: what are the causes of obligate autotrophy and  
1084 methanotrophy? *FEMS Microbiol Rev* 28: 335-352.
- 1085 81. Hughes NJ, Clayton C, Chalk P, Kelly D (1998) *Helicobacter pylori* porCDAB and  
1086 oorDABC Genes Encode Distinct Pyruvate: Flavodoxin and 2-Oxoglutarate:  
1087 Acceptor Oxidoreductases Which Mediate Electron Transport to NADP. *J*  
1088 *Bacteriol* 180: 1119-1128.
- 1089 82. Kather B, K. Stingl, M. Van der Rest, K. Altendorf, and D. Molenaar (2000) Another  
1090 Unusual Type of Citric Acid Cycle Enzyme in *Helicobacter pylori*: the  
1091 Malate:Quinone Oxidoreductase. *J Bacteriol* 182: 3204-3209.
- 1092 83. Gehring U, Arnon DI (1972) Purification and Properties of alpha-Ketoglutarate  
1093 Synthase from a Photosynthetic Bacterium. *J Biol Chem* 247: 6963-6969.
- 1094 84. Breese K, Boll M, Alt-Morbe J, Schagger H, Fuchs G (1998) Genes coding for the  
1095 benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium  
1096 *Thauera aromatica*. *Eur J Biochem* 256: 148-154.
- 1097 85. Jahn D, Verkamp E, Soll D (1992) Glutamyl-transfer RNA: a precursor of heme and  
1098 chlorophyll biosynthesis. *Trends in Biochemical Sciences* 17: 215-218.
- 1099 86. Lin JT, Goldman BS, Stewart V (1994) The Nasfedcba Operon for Nitrate and Nitrite  
1100 Assimilation in *Klebsiella-Pneumoniae* M5al. *J Bacteriol* 176: 2551-2559.
- 1101 87. Merrick MJ, Edwards RA (1995) Nitrogen control in bacteria. *Microbiol Rev* 59:  
1102 604-622.
- 1103 88. van Veen HW (1997) Phosphate transport in prokaryotes: molecules, mediators and  
1104 mechanisms. *Antonie Van Leeuwenhoek International Journal of General and*  
1105 *Molecular Microbiology* 72: 299-315.
- 1106 89. Prere MF, Chandler M, Fayet O (1990) Transposition in *Shigella dysenteriae*:  
1107 isolation and analysis of IS911, a new member of the IS3 group of insertion  
1108 sequences. *J Bacteriol* 172: 4090-4099.
- 1109 90. Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, et al. (2006)  
1110 Phosphonate utilization by the globally important marine diazotroph  
1111 *Trichodesmium*. *Nature* 439: 68-71.
- 1112 91. Kolowitz LC, Ingall ED, Benner R (2001) Composition and cycling of marine  
1113 organic phosphorus. *Limnology and Oceanography* 46: 309-320.
- 1114 92. Klotz MG, Arp DJ, Chain PSG, El-Sheikh AF, Hauser LJ, et al. (2006) The complete  
1115 genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing  
1116 bacterium *Nitrosococcus oceani* ATCC19707. *Appl Envir Microbiol* In press.
- 1117 93. Watson SW (1965) Characteristics of a marine nitrifying bacterium, *Nitrosocystis*  
1118 *oceanus* Sp. N. *Limnol Oceanogr* 10: 274-289.

- 1119 94. Romling U, Gomelsky M, Galperin MY (2005) C-di-GMP: the dawning of a novel  
1120 bacterial signalling system. *Molec Microbiol* 57: 629-639.
- 1121 95. Zhulin I, Taylor B, Dixon R (1997) PAS domain S-boxes in Archaea, bacteria and  
1122 sensors for oxygen and redox. *Trends in Biochemical Sciences* 22: 331-333.
- 1123 96. McCarter LL (2001) Polar Flagellar Motility of the Vibrionaceae. *Microbiol Mol Biol*  
1124 *Rev* 65: 445-462.
- 1125 97. Wadhams GH, Armitage JP (2004) Making sense of it all : Bacterial chemotaxis.  
1126 *Nature Reviews Molecular Cell Biology* 5: 1024-1037.
- 1127 98. Kachlany SC, Planet PJ, DeSalle R, Fine DH, Figurski DH (2001) Genes for tight  
1128 adherence of *Actinobacillus actinomycescomitans*: from plaque to plaque to  
1129 pond scum. *Trends in Microbiology* 9: 429-437.
- 1130 99. Jannasch HW, Mottl MJ (1985) Geomicrobiology of deep-sea hydrothermal vents.  
1131 *Science* 229: 717-725.
- 1132 100. McCollom TM, Shock EL (1997) Geochemical constraints on  
1133 chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal  
1134 systems. *Geochimica et cosmochimica acta* 61: 4375-4391.
- 1135 101. Nies DH (2003) Efflux-mediated heavy metal resistance in prokaryotes. *FEMS*  
1136 *Microbiol Rev* 27: 313-339.
- 1137 102. Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, et al. (2004) Genome sequence of  
1138 the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid  
1139 fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* 101:  
1140 18036-18041.
- 1141 103. Nies DH (1999) Microbial heavy-metal resistance. *Appl Microbiol Biotechnol* 51:  
1142 730-750.
- 1143 104. Edgcomb VP, Molyneaux SJ, Saito MA, Lloyd K, Boer S, et al. (2004) Sulfide  
1144 ameliorates metal toxicity for deep-sea hydrothermal vent archaea. *Appl Environ*  
1145 *Microbiol* 70: 2551-2555.
- 1146 105. Beaumont HJE, Hommes NG, Sayavedra-Soto LA, Arp DJ, Arciero DM, et al.  
1147 (2002) Nitrite reductase of *Nitrosomonas europaea* is not essential for production  
1148 of gaseous nitrogen oxides and confers tolerance to nitrite. *J Bacteriol* 184: 2557-  
1149 2560.
- 1150 106. Beaumont HJE, Lens SI, Westerhoff HV, VanSpanning RJM (2005) Novel nirK  
1151 cluster genes in *Nitrosomonas europaea* are required for NirK-dependent  
1152 tolerance to nitrite. *J Bacteriol* 187: 6849-6851.
- 1153 107. Nakamura K, Go N (2005) Function and molecular evolution of multicopper blue  
1154 proteins. *Cell Molec Life Sci* 62: 2050 - 2066.
- 1155 108. Rock JD, Moir JW (2005) Microaerobic denitrification in *Neisseria meningitis*.  
1156 *Biochem Soc Trans* 33: 134-136.
- 1157 109. Ewing BL, Hillier M, Wendl P, Green P (1998) Basecalling of automated sequencer  
1158 traces using phred. I. Accuracy assessment. *Genome Res*  
1159 8: 175-185.
- 1160 110. Ewing B, Green P (1998) Basecalling of automated sequencer traces using phred. II.  
1161 Error probabilities. *Genome Res* 8: 186-194.
- 1162 111. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence  
1163 finishing. *Genome Res* 8: 195-202

1164 112. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial  
1165 gene identification with GLIMMER. *Nucleic Acids Res* 27: 4636-4641.

1166 113. Badger JH, Olsen GJ (1999) CRITICA: coding region identification tool invoking  
1167 comparative analysis. *Molecular Biology and Evolution* 16: 512-524.

1168 114. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes.  
1169 *Nucleic Acids Research* 28: 27-30.

1170 115. McHardy AC, Goesmann A, Puhler A, Meyer F (2004) Development of joint  
1171 application strategies for two microbial gene finders. *Bioinformatics* 20: 1622-  
1172 1631.

1173 116. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, et al. (2003) GenDB--An  
1174 open source genome annotation system for prokaryotic genomes. *Nucleic Acids*  
1175 *Res* 31: 2187-2195.

1176 117. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting  
1177 transmembrane protein topology with a hidden Markov model: Application to  
1178 complete genomes. *Journal of Molecular Biology* 305: 567-580.

1179 118. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of  
1180 signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340: 783-795.

1181 119. Ren Q, Kang KH, Paulsen IT (2004) TransportDB: a relational database of cellular  
1182 membrane transport systems. *Nucleic Acids Res* 32: D284-D288.

1183 120. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, et al. (2006)  
1184 The integrated microbial genomes (IMG) system. *Nucl Acids Res* %R  
1185 101093/nar/gkj024 34: D344-348.

1186 121. Swofford DL (2002) PAUP\*. *Phylogenetic Analysis Using Parsimony (\*and Other*  
1187 *Methods)*. . Version 4 ed. Sunderland, Massachusetts: Sinauer Associates.

1188 122. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The  
1189 CLUSTAL X windows interface: Flexible strategies for multiple sequence  
1190 alignment aided by quality analysis tools. *Nucl Acids Res* 25: 4876-4882.

1191

1192 **Acknowledgments**

1193

1194 This work was performed under the auspices of the United States Department of Energy  
1195 by Lawrence Livermore National Laboratory, University of California, under contract W-  
1196 7405-ENG-48. Genome closure was funded in part by a USF Innovative Teaching Grant  
1197 (KMS). SKF, CAK, and KMS gratefully acknowledge support from the United States  
1198 Department of Agriculture Higher Education Challenge Grants Program (Award #  
1199 20053841115876). SMS kindly acknowledges support through a fellowship received  
1200 from the Hanse Wissenschaftskolleg in Delmenhorst, Germany (<http://www.h-w-k.de>).  
1201 MH was supported by a WHOI postdoctoral scholarship. We would like to thank  
1202 Hannah Rutherford for her assistance in studies to ascertain the sensitivity of *T.*  
1203 *crunogena* to heavy metals, Marian Arada for her help in preparing genomic DNA,  
1204 Jennifer Mobberly for her assistance with inducing phage, and Shana K. Goffredi and  
1205 Shirley A. Kowalewski for their thoughtful suggestions on this manuscript. Doug Nelson  
1206 and three anonymous reviewers provided constructive comments that substantially  
1207 improved the manuscript.



1208

## 1209 **Figure Captions**

1210

1211 **Figure 1.** Circular map of the *Thiomicrospira crunogena* XCL-2 genome.

1212 The outer two rings are protein-encoding genes, which are color-coded according to COG  
1213 category. Rings 3 and 4 are tRNA and rRNA genes. Ring 5 indicates the location of a  
1214 prophage (magenta), phosphonate/heavy metal resistance island (cyan), and four insertion  
1215 sequences (red; two insertions at 2028543 and 2035034 are superimposed on this figure).  
1216 The black circle indicates the deviation from the average %GC, and the purple and green  
1217 circle is the GC skew ( $= [G-C]/[G+C]$ ). Both the %GC and GC skew were calculated  
1218 using a sliding window of 10,000 bp with a window step of 100.

1219

1220 **Figure 2.** Cell model for *Thiomicrospira crunogena* XCL-2, with an emphasis on  
1221 ultrastructure, transport, energy, carbon metabolism, and chemotaxis.

1222 Genes encoding virtually all of the steps for the synthesis of nucleotides and amino acids  
1223 by canonical pathways are present, and are omitted here for simplicity. Electron transport  
1224 components are yellow, and abbreviations are: NDH—NADH dehydrogenase; UQ—  
1225 ubiquinone; bc<sub>1</sub>—bc<sub>1</sub> complex; Sox—Sox system; cytC—cytochrome C; cbb<sub>3</sub>—cbb<sub>3</sub>-  
1226 type cytochrome C oxidase. Methyl-accepting chemotaxis proteins (MCP) are fuchsia, as  
1227 are MCP's with PAS domains or PAS folds. Influx and efflux transporter families with  
1228 representatives in this genome are indicated on the figure, with the number of each type  
1229 of transporter in parentheses. ATP-dependent transporters are red, secondary transporters  
1230 are sky blue, ion channels are green, and unclassified transporters are purple.

1231 Abbreviations for transporter families are as follows: ABC – ATP-binding cassette  
1232 superfamily; AGCS—Alanine or glycine:cation symporter family; AMT—Ammonium  
1233 transporter family; APC—amino acid-polyamine-organocation family; ATP syn—ATP  
1234 synthetase; BASS—Bile acid:Na<sup>+</sup> symporter family; BCCT—Betaine/carnitine/choline  
1235 transporter family; CaCA—Ca<sup>2+</sup>:cation antiporter family; CDF—cation diffusion  
1236 facilitator family; CHR—Chromate ion transporter family; CPA—Monovalent  
1237 cation:proton antiporter-1, -2, and -3 families; DAACS—Dicarboxylate/amino  
1238 acid:cation symporter family; DASS—Divalent anion:Na<sup>+</sup> symporter family; DMT—  
1239 Drug/metabolite transporter superfamily; FeoB—Ferrous iron uptake family; IRT—  
1240 Iron/lead transporter superfamily; MATE—multidrug/oligosaccharidyl-  
1241 lipid/polysaccharide (MOP) flippase superfamily, MATE family; McsS—Small  
1242 conductance mechanosensitive ion channel family; MFS—Major facilitator superfamily;  
1243 MgtE—Mg<sup>2+</sup> transporter-E family; MIT—CorA metal ion transporter family; NCS2—  
1244 Nucleobase:cation symporter-2 family; NRAMP—Metal ion transporter family; NSS—  
1245 Neurotransmitter:sodium symporter family; P-ATP—P-type ATPase superfamily; Pit—  
1246 Inorganic phosphate transporter family; PNaS—Phosphate:Na<sup>+</sup> symporter family;  
1247 PnuC—Nicotamide mononucleotide uptake permease family; RhtB—Resistance to  
1248 homoserine/threonine family; RND—Resistance-nodulation-cell division superfamily;  
1249 SSS—Solute:sodium symporter family; SulP—Sulfate permease family; TRAP—  
1250 Tripartite ATP-independent periplasmic transporter family; TRK—K<sup>+</sup> transporter family;  
1251 VIC—Voltage-gated ion channel superfamily.

1252

1253

1254 **Figure 3.** Prophage genome within the *Thiomicrospira crunogena* XCL-2  
1255 genome.  
1256 Lysogenic and lytic genes are delineated, as are predicted gene functions.  
1257

1258 **Figure 4.** Phylogenetic relationship of *Thiomicrospira crunogena* XCL-2 SoxB to  
1259 sequences of selected bacteria.  
1260 Sequences were aligned using the program package MacVector. Neighbor-joining and  
1261 parsimony trees based on the predicted amino acid sequences were calculated using  
1262 PAUP 4.0b10. Bootstrap values (1,000 replicates) are given for the neighbor-joining (first  
1263 value) and parsimony analyses (second value).  
1264

1265 **Figure 5.** Transporter gene frequencies within the genomes of *Thiomicrospira*  
1266 *crunogena* XCL-2 (marked with an arrow) and other proteobacteria.  
1267 *N. winogradskyi* is an alphaproteobacterium, *N. europaea* is a betaproteobacterium, and  
1268 *N. oceani* and *M. capsulatus* are gammaproteobacteria. Bars for intracellular pathogens  
1269 are lighter red than the other heterotrophic gammaproteobacteria.  
1270

1271 **Figure 6.** Calvin-Benson-Bassham cycle gene organization in Proteobacteria.  
1272 Rubisco genes (*cbbLS* and *cbbM*) are green, phosphoribulokinase genes (*cbbP*) are red,  
1273 other genes encoding Calvin-Benson-Bassham cycle enzymes are black, and  
1274 carboxysome structural genes are grey. For species where *cbbP* is not near *cbbLS* or  
1275 *cbbM*, the distance from the Rubisco gene to *cbbP* in kbp is indicated in parentheses.  
1276 *Thiobacillus denitrificans* has two *cbbP* genes, so two distances are indicated for this  
1277 species. Names of organisms that are unable to grow well as organoheterotrophs are  
1278 boxed. Abbreviations and accession numbers for the 16S sequences used to construct the  
1279 cladogram are as follows: *A. ehrlichei*--*Alkalilimnicola ehrlichei*, AF406554; *Brady.*  
1280 *sp.*--*Bradyrhizobium* sp., AF338169; *B. japonicum*--*Bradyrhizobium japonicum*, D13430;  
1281 *B. xenovorans*--*Burkholderia xenovorans*, U86373; *D. aromatica*--*Dechloromonas*  
1282 *aromatica*, AY032610; *M. magneticum*--*Magnetospirillum magneticum*, D17514; *M.*  
1283 *capsulatus*--*Methylococcus capsulatus* BATH, AF331869; *N. hamburgensis*--*Nitrobacter*  
1284 *hamburgensis*, L11663; *N. winogradskyi*--*Nitrobacter winogradskyi*, L11661; *N. oceani*--  
1285 *Nitrosococcus oceani*, AF363287; *N. europaea*--*Nitrosomonas europaea*, BX321856;  
1286 *N. multiformis*--*Nitrosospira multiformis*, L35509; *P. denitrificans*--*Paracoccus*  
1287 *denitrificans*, X69159; *R. sphaeroides*--*Rhodobacter sphaeroides*, CP000144; *R.*  
1288 *ferrireducens*--*Rhodoferax ferrireducens*, AF435948; *R. palustris*--*Rhodopseudomonas*  
1289 *palustris*, NC 005296; *R. rubrum*--*Rhodospirillum rubrum*, D30778; *R. gelatinosus*--  
1290 *Rubrivivax gelatinosus*, M60682; *S. meliloti*--*Sinorhizobium meliloti*, D14509; *T.*  
1291 *denitrificans*--*Thiobacillus denitrificans*, AJ43144; *T. crunogena*--*Thiomicrospira*  
1292 *crunogena*, AF064545. The cladogram was based on an alignment of 1622 bp of the 16S  
1293 rRNA genes, and is the most parsimonious tree (length 2735) resulting from a heuristic  
1294 search with 100 replicate random step-wise addition and TBR branch swapping  
1295 (PAUP\*4.0b10; [121]Swofford, 2003). Sequences were aligned using ClustalW [122], as  
1296 implemented in BioEdit. Percent similarities and identities for *cbbL*, *cbbM*, and *cbbP*  
1297 gene products, as well as gene locus tags, are provided as supporting information (Table  
1298 S2).  
1299

1300 **Figure 7.** Models for glycolysis, gluconeogenesis, and the citric acid cycle in  
1301 *Thiomicrospira crunogena* XCL-2.

1302 Models for central carbon metabolism for cells under environmental conditions with A.  
1303 sufficient reduced sulfur and oxygen; B. sulfide scarcity; C. oxygen scarcity; Green  
1304 arrows represent the two ‘non-canonical’ citric acid cycle enzymes, 2-oxoglutarate  
1305 oxidoreductase (2-OG OR) and malate: quinone oxidoreductase (MQO).

1306

1307 **Figure 8.** *Thiomicrospira crunogena* XCL-2 phosphonate operon.

1308 The *DBR-1* genes are identical to each other, as are the *DBR-2* genes. Gene abbreviations  
1309 are: *DBR-1* and *2*—DNA breaking-rejoining enzymes; *hyp*—hypothetical protein;

1310 *phnFDCEGHIJKLMNP*—phosphonate operon; *chp*—conserved hypothetical protein.

1311 An asterisk marks the location of a region (within and upstream of tRNA-phe) with a  
1312 high level of similarity to the 5’ ends of the two direct repeat sequences noted in the  
1313 figure. The transposase and integrase are actually a single CDS separated by a  
1314 frameshift.

1315

1316 **Figure 9.** Numbers of methyl-accepting chemotaxis protein genes in  
1317 *Thiomicrospira crunogena* XCL-2 and other proteobacteria.

1318 *T. crunogena* is marked with an arrow. (A) depicts the total number of CDS’s predicted  
1319 to encode methyl-accepting chemotaxis protein genes (MCPs), while those CDSs with a  
1320 PAS domain or fold are tallied in (B).

1321

1322

1323  
1324

**TABLE 1.** *Thiomicrospira crunogena* XCL-2 genome summary

<b>Item</b>	<b>Value</b>
Chromosomes	1
Basepairs	2,427,734
GC content (%)	43.1
% coding	90.6
RNA-encoding genes	
tRNAs	43
16S-Ile tRNA <sub>GAT</sub> -Ala tRNA <sub>TGC</sub> -23S-5S RNA operons	3
Genes in each COG category	
DNA replication, recombination, and repair	113
Transcription	84
Translation, ribosomal structure and biogenesis	153
Posttranslational modification, protein turnover, chaperones	115
Energy production and conversion	117
Carbohydrate transport and metabolism	79
Amino acid transport and metabolism	167
Nucleotide transport and metabolism	50
Lipid transport and metabolism	39
Coenzyme transport and metabolism	102
Secondary metabolite biosynthesis, transport, catabolism	37
Cell wall/membrane/envelope biogenesis	142
Inorganic ion transport and metabolism	120
Cell motility	79
Signal transduction mechanisms	147
Cell cycle control, cell division, chromosome partitioning	18
Intracellular trafficking, secretion, and vesicular transport	65
General function	188

1325

1326  
1327  
1328

**TABLE 2.** *Thiomicrospira crunogena* XCL-2\* and *Nitrosococcus oceani* ATCC 19707 regulatory and signaling proteins

Number:		
<i>T. crunogena</i>	<i>N. oceani</i>	
72	104	Transcription/Elongation/Termination Factors
6	9	Sigma Factors
6	11	Anti/Anti-Anti Sigma Factors
6	6	Termination/Antitermination Factors
2	2	Elongation Factors
52	76	Transcription factors
123	75	Signal Transduction proteins
		Chemotaxis Signal Transduction proteins (24 total, <i>T. crunogena</i> )
14	1	Methyl-accepting chemotaxis proteins
2	1	CheA signal transduction histidine kinase
3	2	CheW protein
2	0	Response regulator receiver modulated CheW protein
1	1	MCP methyltransferase, CheR-type
0	1	Response regulator receiver, CheY
1	1	Response regulator receiver modulated CheB methylesterase
1	0	CheD, stimulates methylation of MCP proteins
		Non-Chemotaxis Signal Transduction (99 total, <i>T. crunogena</i> )
17	18	Signal Transduction Histidine Kinase
5	0	Diguanylate phosphodiesterase
1	0	Response regulator receiver modulated diguanylate phosphodiesterase
16	2	Diguanylate cyclase
5	0	Diguanylate cyclase with PAS/PAC sensor
9	4	Diguanylate cyclase/phosphodiesterase
0	1	Periplasmic sensor hybrid histidine kinase and response regulator receiver modulated diguanylate cyclase/phosphodiesterase
7	5	Diguanylate cyclase/phosphodiesterase with PAS/PAC sensor
1	2	Response regulator receiver modulated diguanylate cyclase/phosphodiesterase
2	1	Cyclic nucleotide-binding protein
1	0	Cyclic-AMP phosphodiesterase
0	1	Adenylate/guanylate cyclase
1	4	PTS NTR Regulator proteins
34	29	Miscellaneous

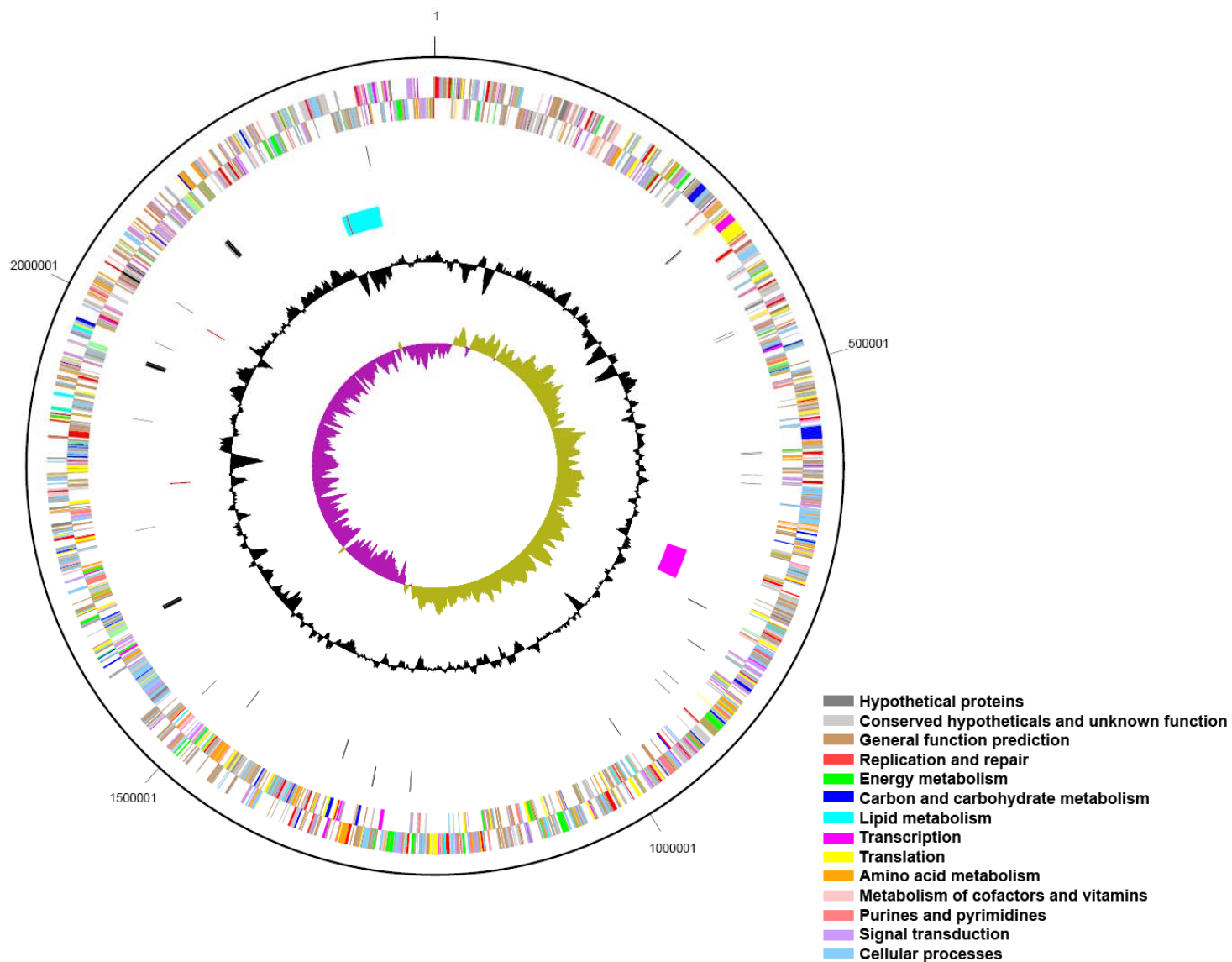
	195	179	Total
1329	*A list of locus tags for these genes is present in Table S4.		

**TABLE 3.** Growth-inhibiting concentrations (mM) of heavy metals for *Thiomicrospira crunogena* XCL-2 and *Escherichia coli*.

Heavy metal ion	<i>T. crunogena</i> <sup>a</sup>	<i>E. coli</i> <sup>b</sup>
Hg <sup>+2</sup>	0.01	0.01
Cu <sup>+2</sup>	0.02	1
Ag <sup>+1</sup>	0.02	0.02
Cd <sup>+2</sup>	0.05	0.5
Co <sup>+2</sup>	0.1	1
Ni <sup>+2</sup>	0.1	1
Zn <sup>+2</sup>	1	1
Cr <sup>+2</sup>	1	5
Mn <sup>+2</sup>	2	20

<sup>a</sup> *T. 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). For both species, the concentration at which growth ceased is listed.

<sup>b</sup>Data from [103].



**Figure 1**



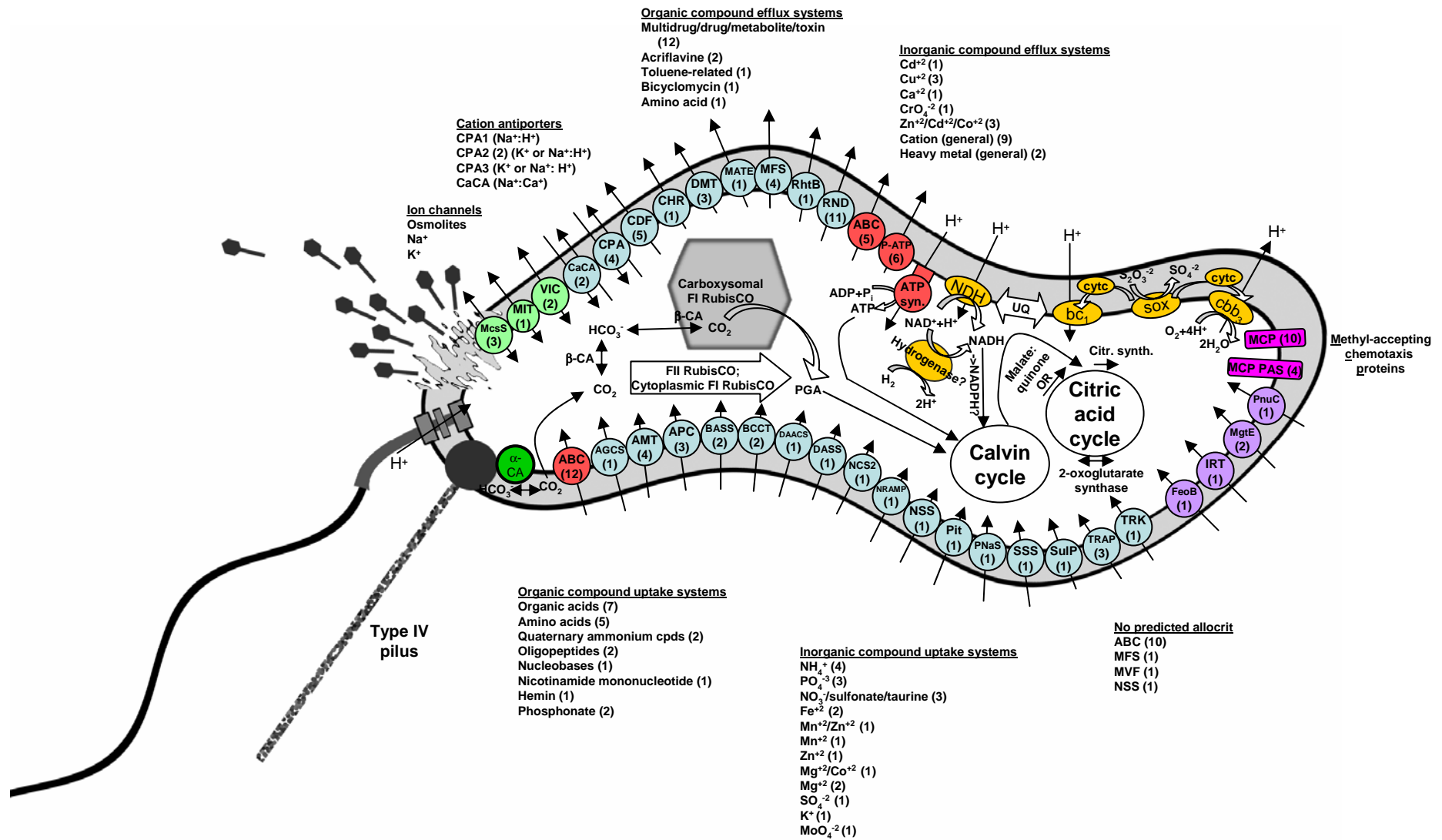
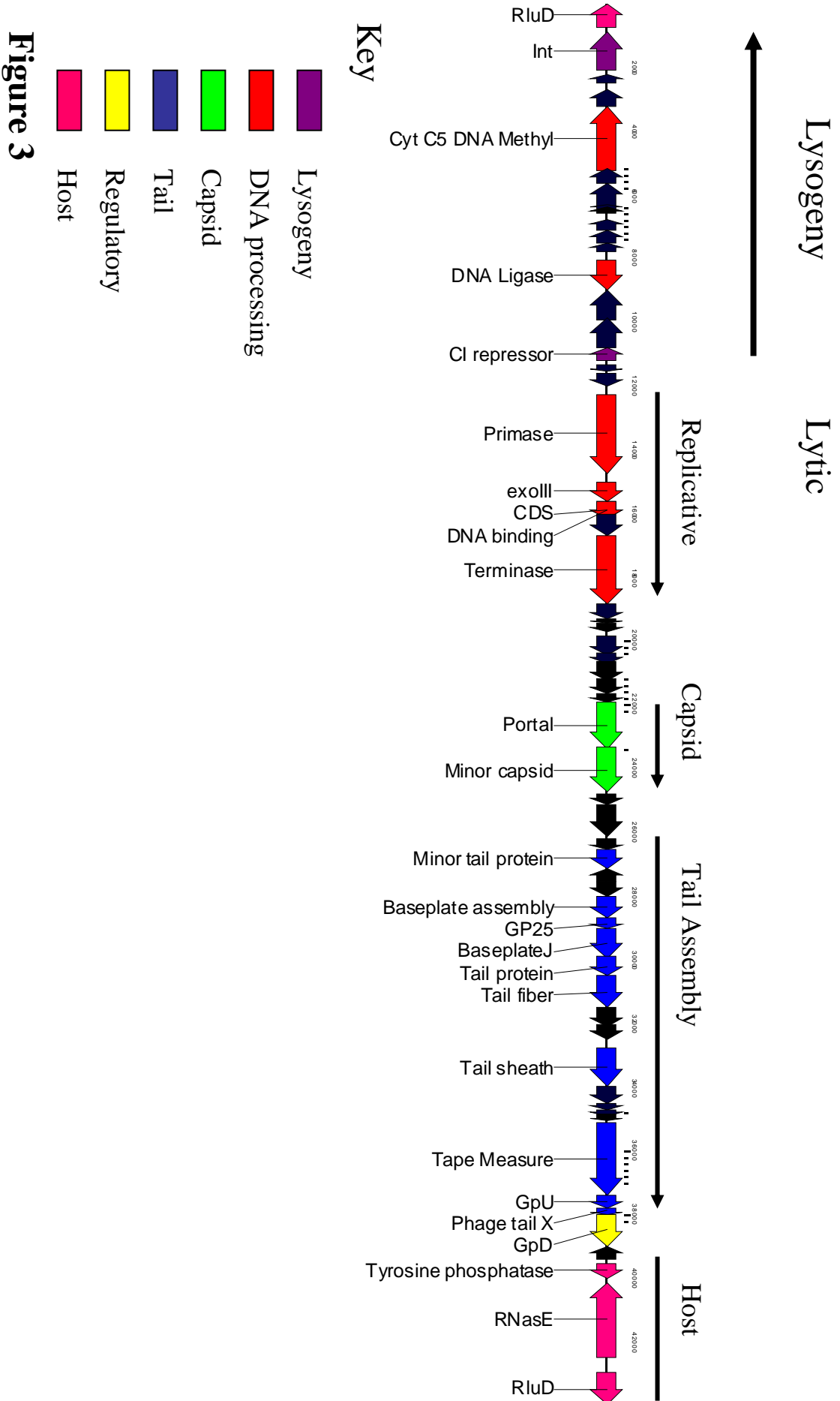
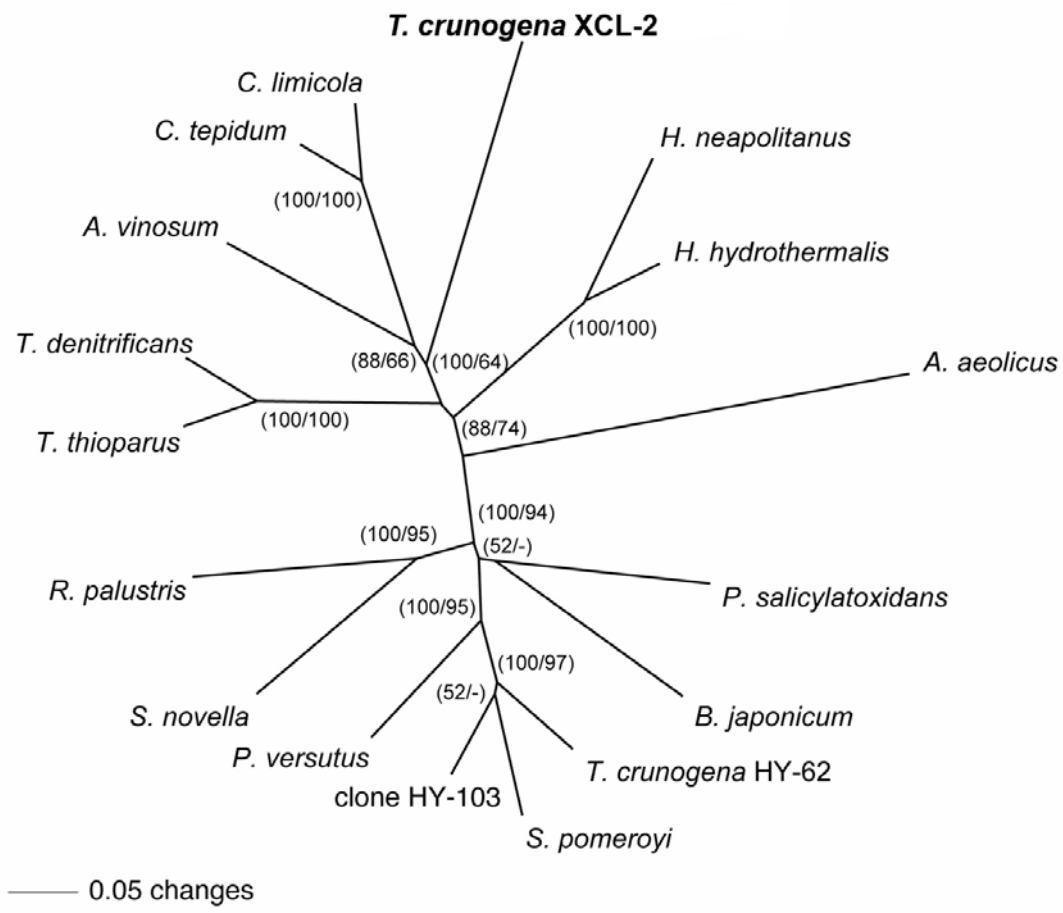


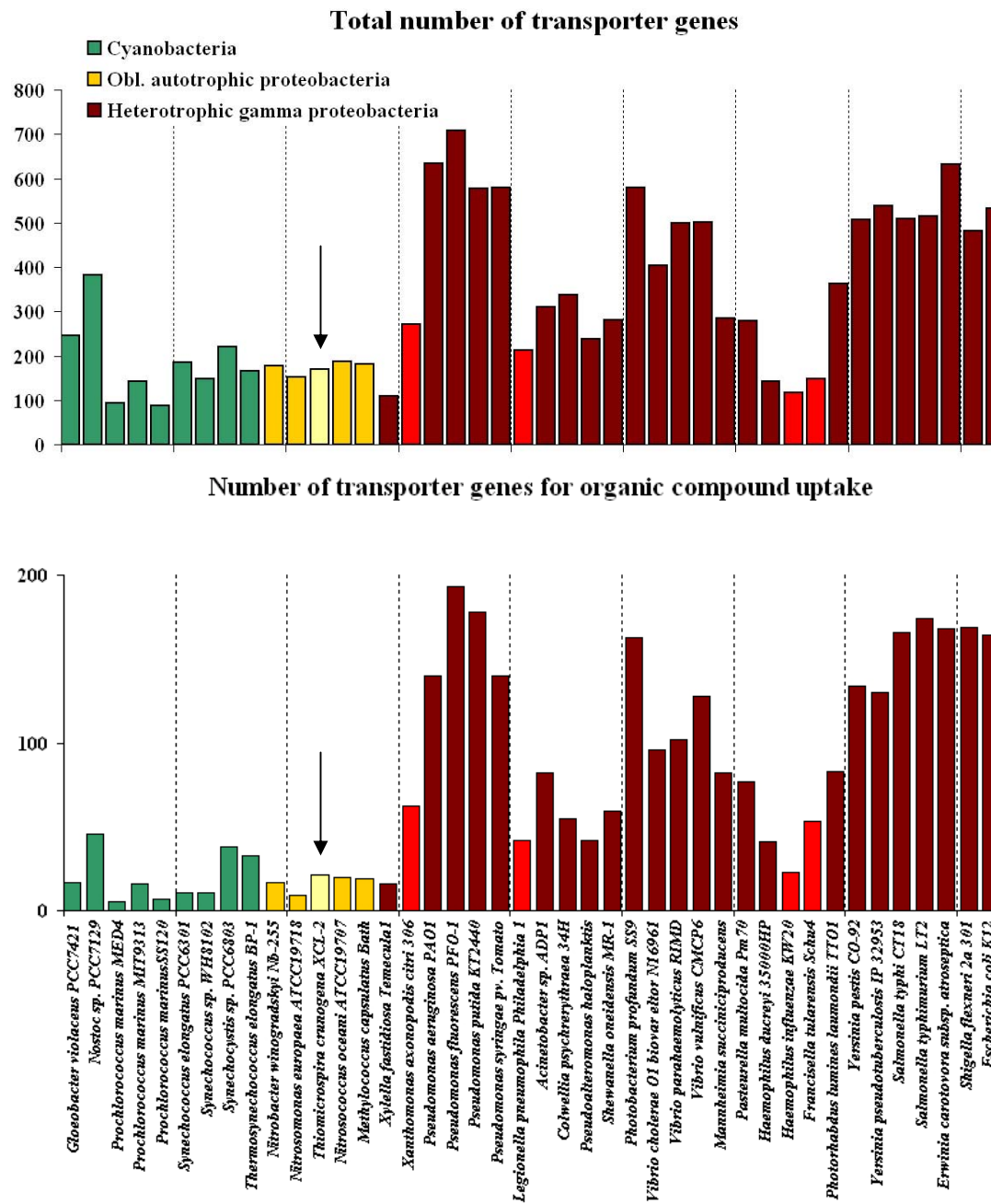
Figure 2



**Figure 3**



**Figure 4**



**Figure 5**

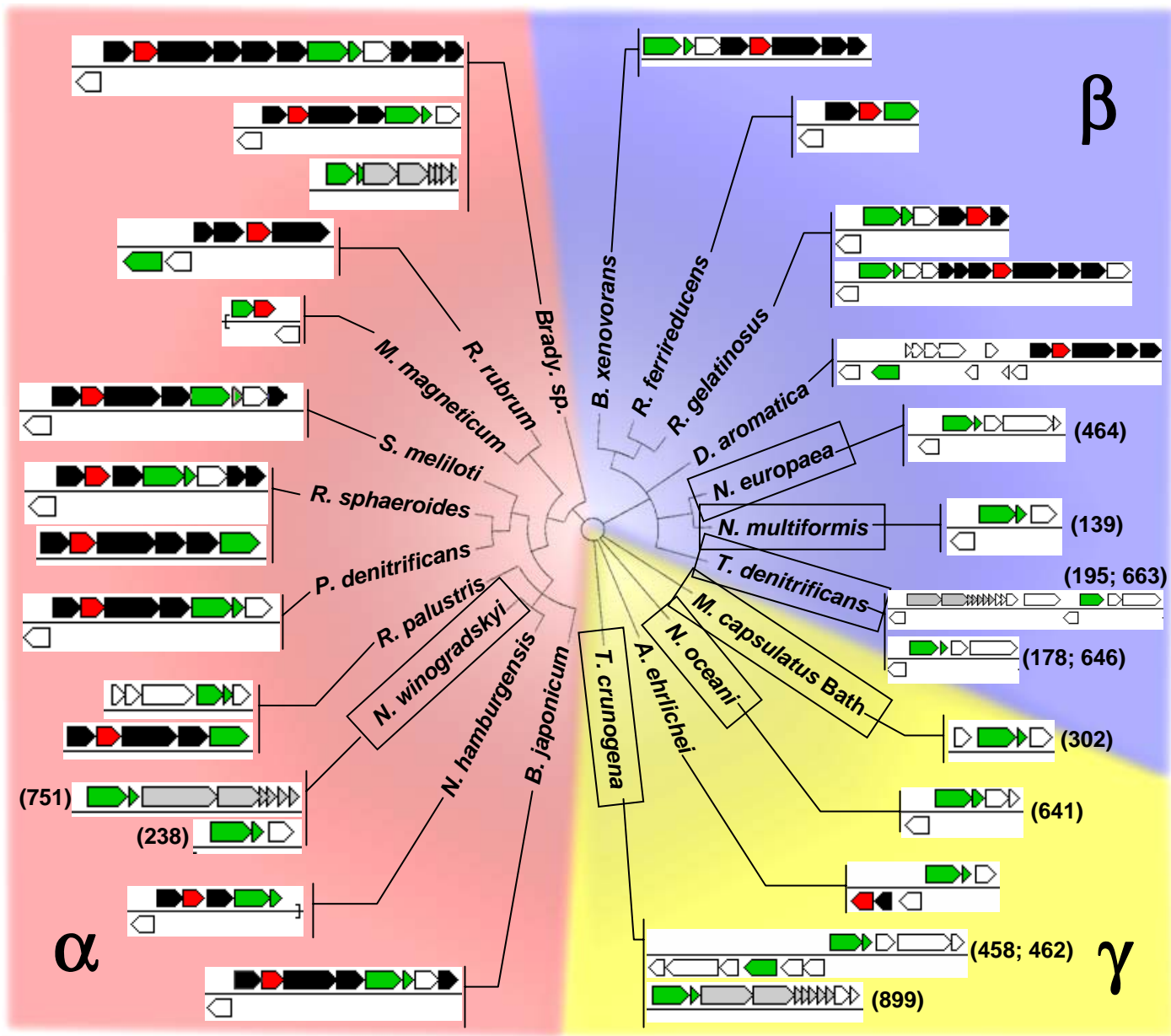
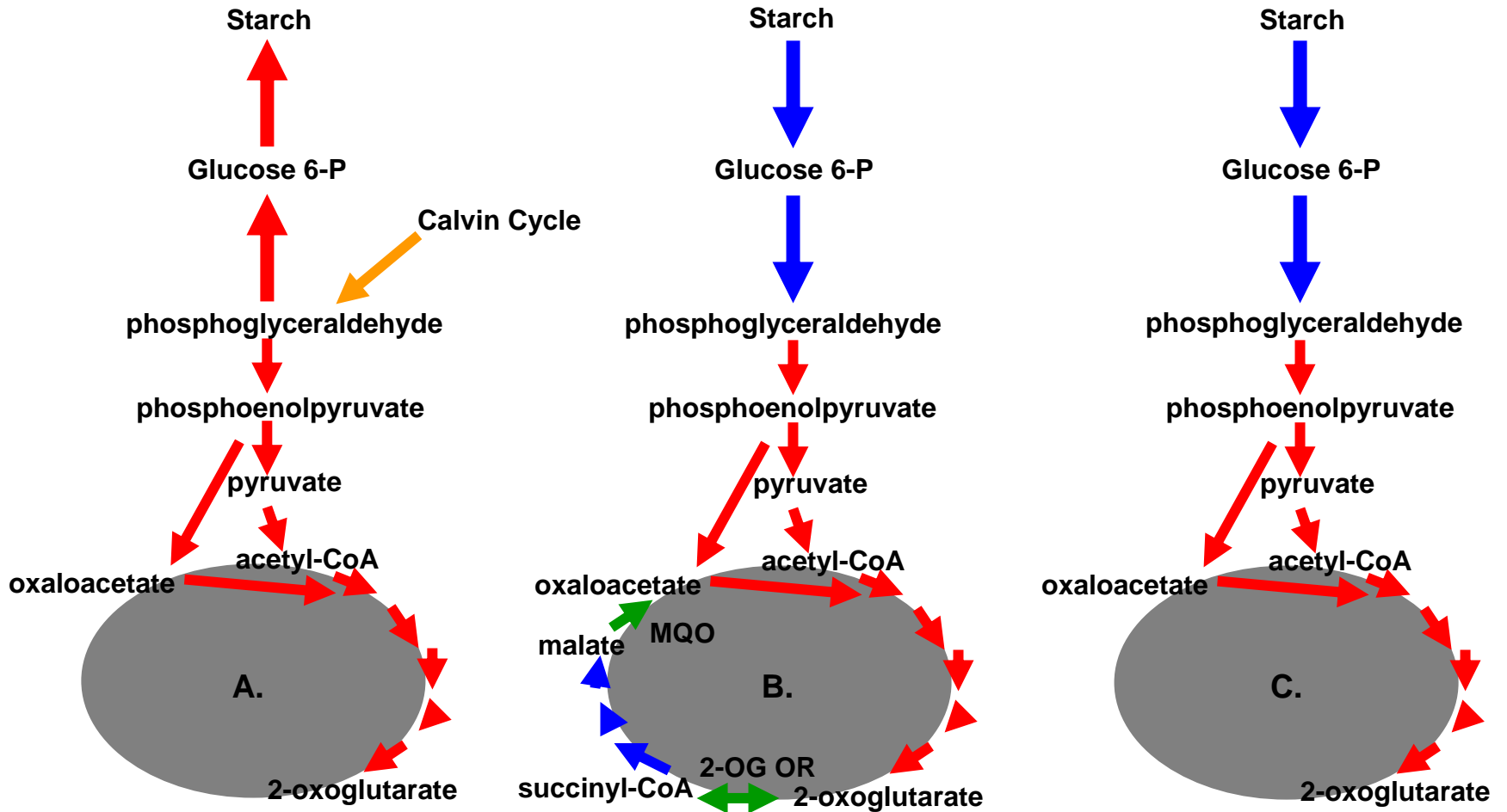
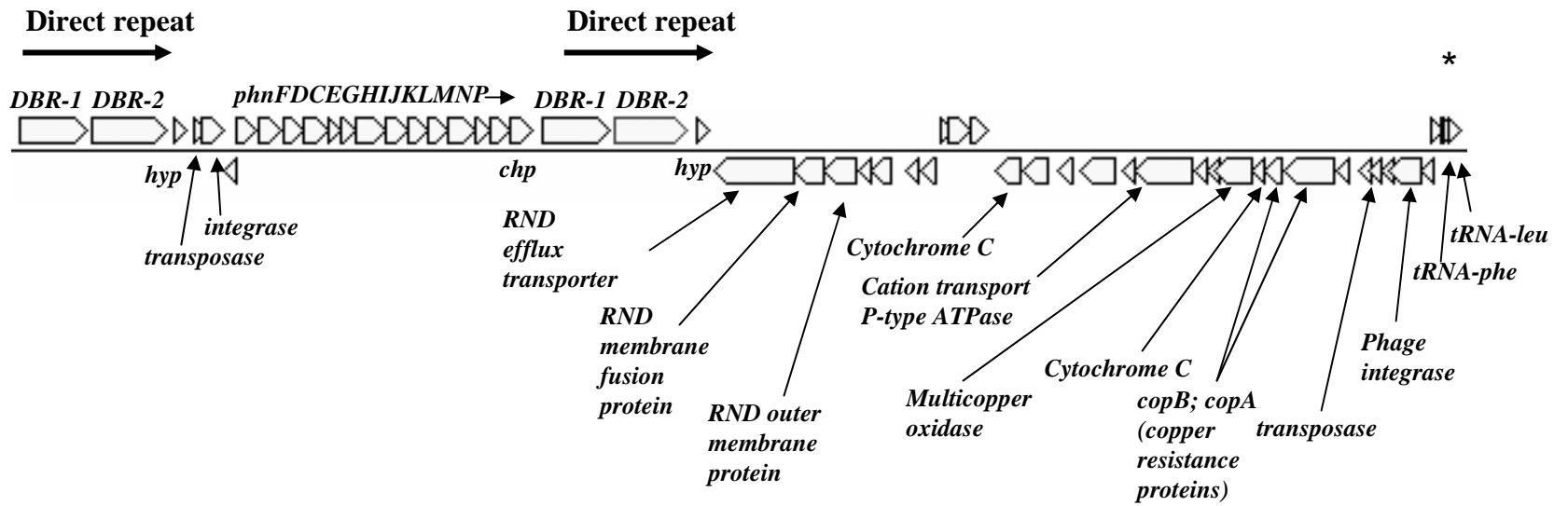


Figure 6



**Figure 7**



**Figure 8**

**Figure 9**

