LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-JRNL-227471

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

K. M. Scott, S. M. Sievert, et al

PLOS Biology, v 4, no. 12,

November 14, 2006

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

Running head: *T. crunogena* genome

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

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

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

(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.

Introduction

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

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

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

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

presence of low concentrations of dissolved inorganic carbon, due to an increase in cellular affinity for both HCO_3^- and CO_2 under low CO_2 conditions [15]. The ability to grow under low CO_2 conditions is likely an advantage when the habitat is dominated by relatively low CO_2 seawater. Further adaptations in nutrient acquisition and microhabitat sensing are likely to be present in this organism.

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

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

Results/Discussion

Genome structure

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

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

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

Phylogeny

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

Prophage

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

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

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

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

The high similarity of the CDSs to lambdoid (lysogeny and replication genes) and P2-like (tail module) temperate coliphages is surprising and unprecedented in marine prophage genomes [32]. A major frustration encountered in marine phage genomics is the low similarity of CDSs to anything in GenBank, making the interpretation of the biological function extremely difficult. The lambdoid siphophages are generally members of the *Siphoviridae* whereas the P2-like phages are *Myoviridae*, which the *T. crunogena* XCL-2 prophage is predicted to be. Such a mixed heritage is often the result of the modular evolution of phages. The general genomic organization of the *T. crunogena* XCL-2 prophage-like element (integrase, repressor, DNA replicative genes, terminase, portal, capsid, tail genes) is common to several known prophages, including those of *Staphylococcus aureus* (ie. ϕMu50B), *Streptococcus pyogenes* (prophages 370.3 and 370.2), and *S. thermophilus* (prophage O1205; [33]).

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

Redox substrate metabolism and electron transport

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

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

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

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

The confirmation of the presence of a *soxB* gene in *T. crunogena* XCL-2 is particularly interesting, as it is a departure from previous studies with close relatives. Attempts to PCR-amplify *soxB* from *T. crunogena* ATCC 700270^{T} and *T. pelophila*

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

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

Genes for Ni/Fe hydrogenase large and small subunits are present (*Tcr2037; Tcr2038*), as well as all of the genes necessary for large subunit metal center assembly (*Tcr2035 - 6; Tcr2039 - 2043*) [45]. Their presence and organization into an apparent operon suggest that *T. crunogena* XCL-2 could use H_2 as an electron donor for growth, as its close relative *Hydrogenovibrio* does [46,47]. However, attempts to cultivate *T. crunogena* XCL-2 with H_2 as the sole electron donor have not been successful ([48]; K. Scott, unpubl.data). A requirement for reduced sulfur compounds, even when not used as the primary electron donor, is suggested by the absence of genes encoding the enzymes necessary for assimilatory sulfate reduction (APS reductase; ATP sulfurylase), which are necessary for cysteine synthesis in the absence of environmental sources of thiosulfate or sulfide. Alternatively, this hydrogenase could act as a reductant sink under periods of sulfur and oxygen scarcity, when starch degradation could be utilized to replenish ATP and other metabolite pools (see “Central Carbon Metabolism”, below).

The redox partner for the *T. crunogena* XCL-2 hydrogenase is suggested by the structure of the small subunit, which has two domains. One domain is similar to other hydrogenase small subunits, while the other is similar to pyridine nucleotide-disulphide oxidoreductases and has both an FAD and NADH binding site. The presence of a NADH binding site suggests that the small subunit itself transfers electrons between H_2 and NAD(H), unlike other soluble hydrogenases, in which this activity is mediated by separate "diaphorase" subunits [45], which *T. crunogena* XCL-2 lacks. The small subunit does not have the twin arginine leader sequence that is found in periplasmic and membrane-associated hydrogenases [49], suggesting a cytoplasmic location for this enzyme.

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

In this species, ubiquinone ferries electrons between NADH dehydrogenase and the bc1 complex; all genes are present for its synthesis, but not for menaquinone. Unlike most bacteria, *T. crunogena* XCL-2 does not synthesize the isopentenyl diphosphate units that make up the lipid portion of ubiquinone via the deoxyxylulose 5-phosphate pathway. Instead, most of the genes of the mevalonate pathway (HMG-CoA synthase, *Tcr1719*; HMG-CoA reductase, *Tcr1717;* mevalonate kinase/phosphomevalonate kinase, *Tcr1732, Tcr1733;* and diphosphomevalonate decarboxylase, *Tcr1734* [54]) are present. The single "missing" gene, for acetyl-CoA acetyltransferase, may not be necessary, as HMG-CoA reductase may also catalyze this reaction as it does in *Enterococcus faecalis* [55]. Interestingly, the mevalonate pathway is found in Archaea, eukaryotes, and is common among gram positive bacteria [54,56]. Thus far, the only other proteobacterium to have this pathway is from the alpha class, *Paracoccus zeaxanthinifaciens* [57]. Examination of unpublished genome data from the Integrated Microbial Genomes webpage (http://img.jgi.doe.gov/v1.1/main.cgi), and queries of Genbank did not uncover evidence for a complete set of genes for the mevalonate pathway in other proteobacteria.

The three components of the bc1 complex are represented by three genes in an apparent operon, in the typical order (Rieske iron-sulfur subunit; cytochrome *b* subunit; cytochrome *c1* subunit; *Tcr0991 – 3;* [51]).

Consistent with its microaerophilic lifestyle and inability to use nitrate as an electron acceptor [6], the only terminal oxidase present in the *T. crunogena* XCL-2 genome is a cbb_3-type cytochrome c oxidase (*Tcr1963 - 5*). Neither *bd*- nor bo_3-type quinol oxidases are present, nor is an aa_3-type cytochrome oxidase. To date, *Helicobacter pylori* is the only other sequenced organism that has solely a cbb_3-type oxidase, and this has been proposed to be an adaptation to growth under microaerophilic conditions [51], since cbb_3-type oxidase has a higher affinity for oxygen than aa_3-type oxidase does [58].

In searching for candidate cytochrome proteins that facilitate electron transfer between the Sox system and the bc_1 complex and cbb_3 cytochrome c oxidase, the genome was analyzed to identify genes that encode proteins with heme-coordinating motifs (CxxCH). This search yielded 28 putative heme-binding proteins (Table S1), compared to

54 identified in the genome of *T. denitrificans* [22]. Thirteen of these genes encode proteins that were predicted to reside in the periplasm, five were predicted to integrate into the plasma membrane, and the remaining ten proteins were predicted to reside in the cytoplasm and are thus not considered as candidates for transferring electrons between the Sox system and the bc_1 complex and cbb_3 cytochrome c oxidase (Fig. 2). The five predicted membrane cytochrome proteins were discarded as well, as three of them contribute to assembly and function of the *bc1* complex and cbb_3 cytochrome c oxidase, one is predicted to be a diguanylate cyclase/phosphodiesterase, while the remaining one is a conserved hypothetical protein and likely has no role in catabolic electron transfer.

Of the thirteen predicted periplasmic proteins encoded by the remaining genes, two (*Tcr0628; Tcr0628*) were deemed particularly promising candidates as they met the following criteria: 1) they were not subunits of other cytochrome-containing systems, 2) they were small enough to serve as efficient electron shuttles, 3) they were characterized beyond the level of hypothetical or conserved hypothetical, and 4) they were present in *Thiobacillus denitrificans*, which also has both a Sox system as well as cbb_3 cytochrome c oxidase, and had not been implicated in other cellular functions in this organism. *Tcr0628* and *Tcr0629* both belong to the COG2863 family of cytochromes c553, which are involved in major catabolic pathways in numerous proteobacteria. Interestingly, genes *Tcr0628* and *Tcr0629*, which are separated by a 147-pb spacer that includes a Shine-Delgarno sequence, are highly likely paralogues and a nearly identical gene tandem was also identified in the genome of *T. denitrificans* (*Tbd2026, Tbd2027*). A recent comprehensive phylogenetic analysis of the cytochrome c553 proteins, including the mono-heme cytochromes from *T. crunogena* and *T. denitrificans*, revealed existence of a large protein superfamily that also includes proteins in the COG4654 cytochrome c551/c552 protein family (M.G. Klotz and A.B. Hooper, unpublished results). In ammonia-oxidizing bacteria, representatives of this protein superfamily (*NE0102*, *Neut2204*, *NmulA0344* in the COG4654 protein family; *Noc0751*, *NE0736*, *Neut1650* in the COG2863 protein family) are the key electron carriers that connect the *bc1* complex with complex IV as well as NO_x-detoxofying reductases (i.e., NirK, NirS) and oxidases (i.e., cytochrome P460, cytochrome c peroxidase) involved in nitrifier denitrification ([59] and references therein). In Epsilonproteobacteria such as *Helicobacter pylori* and *hepaticus*, cytochromes in this family (*jhp1148; HH1517*) interact with the terminal cytochrome cbb_3 oxidase. Therefore, we propose that the expression products of genes *Tcr0628* and *Tcr0629* likely represent the electronic link between the Sox system and the *bc1* complex and cbb_3 cytochrome c oxidase in *T. crunogena*. It appears worthwhile to investigate experimentally whether the small difference in sequence between these two genes reflects an adaptation to binding to interaction partners with sites of different redox potential, namely cytochrome c_1 in the *bc1* complex and cytochrome FixP (subunit III) in cbb_3 cytochrome c oxidase.

Given the presence of these electron transport complexes and electron carriers, a model for electron transport chain function is presented here (Figure 2). When thiosulfate or sulfide are acting as the electron donor, the Sox system will introduce electrons into the electron transport chain at the level of cytochrome c [34]. Most will be oxidized by the cbb_3-type cytochrome c oxidase to create a proton potential. Some of the cytochrome c electrons will be used for reverse electron transport to ubiquinone and NAD^+ by the bc1 complex and NADH:ubiquinone oxidoreductase. The NADH created

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

Transporters and nutrient uptake

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

Carbon dioxide uptake and fixation

T. crunogena XCL-2, like many species of cyanobacteria [65], has a carbon concentrating mechanism, in which active dissolved inorganic carbon uptake generates intracellular concentrations that are as much as 100X higher than extracellular [15]. No apparent homologs of any of the cyanobacterial bicarbonate or carbon dioxide uptake systems are present in this genome. *T. crunogena* XCL-2 likely recruited bicarbonate and perhaps carbon dioxide transporters from transporter lineages evolutionarily distinct from those utilized by cyanobacteria. Three carbonic anhydrase genes are present (one α-class, *Tcr1545*; two β-class, *Tcr0421, Tcr0841* [66,67,68], one of which (α-class) is predicted to be periplasmic and membrane-associated, and may keep the periplasmic dissolved inorganic carbon pool at chemical equilibrium despite selective uptake of carbon dioxide or bicarbonate. One β-class enzyme gene is located near the gene for a form II RubisCO (see below) and may be coexpressed with it when the cells are grown under high-CO_2 conditions. The other β-class (formerly ε-class; [68]) carbonic anhydrase is a member of a carboxysome operon and likely functions in this organism's carbon concentrating mechanism. Unlike many other bacteria [69], the gene encoding the sole SulP-type ion transporter (*Tcr1533*) does not have a carbonic anhydrase gene adjacent to it.

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

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

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

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

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

Central carbon metabolism

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

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

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

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

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

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

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

Nitrogen and uptake and assimilation

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

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

Phosphorus uptake

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

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

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

Chemotaxis

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

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

Adhesion

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

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

Heavy metal resistance

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

Indeed, some of these 'metal sequestering' proteins encoded in this genome may function instead in maintaining a stable supply of metals for enzyme active sites. This appears to be the case for the copper-binding CopA and B proteins, cytochrome c, and multicopper oxidase proteins encoded by *Tcr1573 – 6* and *2116 – 3*, which occupy two ~5900 bp regions that are ~92% identical to eachother. The multicopper oxidase genes encoded by *Tcr1576* and *Tcr2113* may function in the reduction of reactive oxides such as NO_x species [105,106,107,108]. Due to the juxtaposition of these four genes, we hypothesize that the CopA (*Tcr1575* and *Tcr2114*) and CopB (*Tcr1574* and *Tcr2115*) proteins may function to bind copper to ensure a steady supply of this metal cofactor for the multicopper oxidase/cytochrome c complex.

Conclusions

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

conditions. A larger-than-expected arsenal of regulatory proteins may enable this organism to regulate multiple mechanisms for coping with variations in inorganic nutrient availability. Its three RubisCO genes, three carbonic anhydrase genes, and carbon concentrating mechanism likely assist in coping with oscillations in environmental CO_2 availability, while multiple ammonium transporters, nitrate reductase, low- and high- affinity phosphate uptake systems, and potential phosphonate use, may enable it to cope with uncertain supplies of these macronutrients.

In contrast, systems for energy generation are more limited, with only one, i.e., Sox, or possibly two, i.e., Sox plus SQR, systems for sulfur oxidation and a single low-oxygen adapted terminal oxidase (*cbb*$_3$-type). Instead of having a branched electron transport chain with multiple inputs and outputs, this organism may use the four PAS-domain or –fold methyl-accepting chemotaxis proteins to guide it to a portion of the chemocline where its simple electron transport chain functions. It is worth noting, in this regard, that *Thiobacillus denitrificans,* which has several systems for sulfur oxidation, has fewer MCPs than *T. crunogena* XCL-2 (Figure 9). Differential expression of portions of the citric acid cycle may enable it to survive periods of reduced sulfur or oxygen scarcity during its 'transit' to more favorable microhabitats.

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

Materials and Methods

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

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

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

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

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

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

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

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

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

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

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-CO2 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 deeop-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

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

30. Rohwer F, Segall A, Steward G, Seguritan V, Breitbart M, et al. (2000) The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with non-marine phages. Limnol Oceanogr 42: 408-418.
31. Sternberg N, Weisberg R (1977) Packaging of coliphage lambda DNA. II. The role of gene D protein. J Mol Biol 117: 733-759.
32. Paul JH, Sullivan MB (2005) Marine phage genomics: What have we learned? Curr Opin Biotech 16: 299-307.
33. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H (2003) Prophage genomics. Microbiol Molec Biol Rev 67: 238-276.
34. Friedrich CG, Quentmeier A, Bardischewsky F, Rother D, Kraft R, et al. (2000) Novel genes coding for lithotrophic sulfur oxidation of *Paracoccus pantotrophus* GB17. J Bacteriol 182: 4677-4687.
35. Rother D, Henrich HJ, Quentmeier A, Bardischewsky F, Friedrich CG (2001) Novel genes of the sox gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in Paracoccus pantotrophus GB17. Journal of Bacteriology 183: 4499-4508.
36. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J (2005) Prokaryotic sulfur oxidation. Current Opinion in Microbiology 8: 253-259.
37. Schutz M, Maldener I, Griesbeck C, Hauska G (1999) Sulfide-quinone reductase from Rhodobacter capsulatus: Requirement for growth, periplasmic localization, and extension of gene sequence analysis. Journal of Bacteriology 181: 6516-6523.
38. Javor BJ, Wilmot DB, Vetter RD (1990) pH–Dependent metabolism of thiosulfate and sulfur globules in the chemolithotrophic marine bacterium *Thiomicrospira crunogena*. Arch Microbiol 154: 231–238.
39. Friedrich CG, Rother D, Bardischewsky F, Quentmeier A, Fischer J (2001) Oxidation of Reduced Inorganic Sulfur Compounds by Bacteria: Emergence of a Common Mechanism? Appl Environ Microbiol 67: 2873-2882.
40. Fani R, Brilli M, Lio P (2005) The origin and evolution of operons: The piecewise building of the proteobacterial histidine operon. J Mol Evol 60: 378-390.
41. Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. Genome Res 15: 809-819.
42. Petri R, Podgorsek L, Imhoff JF (2001) Phylogeny and distribution of the soxB gene among thiosulfate-oxidizing bacteria. Fems Microbiology Letters 197: 171-178.
43. Kelly DP, Shergill JK, Lu WP, Wood AP (1997) Oxidative metabolism of inorganic sulfur compounds by bacteria. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 71: 95-107.
44. Nelson DC, Hagen KD (1995) Physiology and Biochemistry of Symbiotic and Free-Living Chemoautotrophic Sulfur Bacteria. American Zoologist 35: 91-101.
45. Schwartz E, Friedrich B (2005) The H2-metabolizing prokaryotes. In: Dworkin M, editor. The Prokaryotes: An evolving electronic resource for the microbiological community, release 314, http://linkspringer-nycom/link/service/books/10125/. New York: Springer-Verlag.
46. Nishihara H, Yaguchi T, Chung SY, Suzuki K, Yanagi M, et al. (1998) Phylogenetic position of an obligately chemoautotrophic, marine hydrogen-oxidizing bacterium, *Hydrogenovibrio marinus*, on the basis of 16S rRNA gene sequences and two form I RubisCO gene sequences. Arch Microbiol 169: 364-368.

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

62. Mori S, Kawai S, Shi F, Mikami B, Murata K (2005) Molecular conversion of NAD kinase to NADH kinase through single amino acid residue substitution. Journal of Biological Chemistry 280: 24104-24112.
63. Paulsen IT, Nguyen L, Sliwinski MK, Rabus R, Saier MH (2000) Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. J Mol Biol 301: 75-100.
64. Ren Q, Paulsen IT (2005) Comparative analysis of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLOS computational biology 1: 190-201.
65. Badger MR, Price GD, Long BM, Woodger FJ (2006) The environmental plasticity and ecological genomics of the cyanobacterial CO2 concentrating mechanism. J Exp Bot 57: 249-265.
66. Smith KS, Ferry JG (2000) Prokaryotic carbonic anhydrases. FEMS Microbiol Rev 24: 335-366.
67. So AK, Espie GS, Williams EB, Shively JM, Heinhorst S, et al. (2004) A novel evolutionary lineage of carbonic anhydrase (epsilon class) is a component of the carboxysome shell. J Bacteriol 186: 623-630.
68. Sawaya MR, Cannon GC, Heinhorst S, Tanaka S, Williams EB, et al. (2006) The Structure of beta-Carbonic Anhydrase from the Carboxysomal Shell Reveals a Distinct Subclass with One Active Site for the Price of Two. J Biol Chem 281: 7546-7555.
69. Felce J, Saier MH (2004) Carbonic anhydrase fused to anion transporters of the SulP family: Evidence for a novel type of bicarbonate transporter. J Mol Microbiol Biotechnol 8: 169-176.
70. Yoshizawa Y, Toyoda K, Arai H, Ishii M, Igarashi Y (2004) CO2-responsive expression and gene organization of three ribulose-1,5-bisphosphate carboxylase/oxygenase enzymes and carboxysomes in Hydrogenovibrio marinus strain MH-110. Journal of Bacteriology 186: 5685-5691.
71. Hayashi NR, Arai H, Kodama T, Igarashi Y (1997) The novel genes, *cbbQ* and *cbbO*, located downstream from the RubisCO genes of *Pseudomonas hydrogenothermophila*, affect the conformational states and activity of RubisCO. Biochem Biophys Res Commun 241: 565-569.
72. Hayashi NR, Arai H, Kodama T, Igarashi Y (1999) The cbbQ genes located downstream of the form I and form II RubisCO genes, affect the activity of both RubisCOs. Biochem Biophys Res Commun 266: 177-183.
73. Badger M, Hanson D, Price GD (2002) Evolution and diversity of CO2 concentrating mechanisms in cyanobacteria. Functional Plant Biology 29: 161-173.
74. Gibson JL, Tabita FR (1996) The molecular regulation of the reductive pentose phosphate pathway in Proteobacteria and Cyanobacteria. Arch Microbiol 166: 141-150.
75. Kusian B, Bowien B (1997) Organization and regulation of *cbb* CO2 assimilation genes in autotrophic bacteria. FEMS Microbiol Rev 21: 135-155.
76. Shively JM, Van Keulen G, Meijer WG (1998) Something form almost nothing: carbon dioxide fixation in chemoautotrophs. Annual Reviews in Microbiology 52: 191-230.

77. Falcone DL, Tabita FR (1993) Complementation analysis and regulation of CO2 fixation gene expression in a ribulose 1,5-bisphosphate carboxylase-oxygenase deletion strain of *Rhodospirillum rubrum*. J Bacteriol 175: 5066-5077.
78. Ding YR, Ronimus RS, Morgan HW (2000) Sequencing, Cloning, and High-Level Expression of the pfp Gene, Encoding a PPi-Dependent Phosphofructokinase from the Extremely Thermophilic Eubacterium *Dictyoglomus thermophilum*. J Bacteriol 182.
79. Ronimus RS, Morgan HW (2001) The biochemical properties and phylogenies of phosphofructokinases from extremophiles. Extremophiles 5: 357-373.
80. Wood AP, Aurikkoa JP, Kelly DP (2004) A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy? FEMS Microbiol Rev 28: 335-352.
81. Hughes NJ, Clayton C, Chalk P, Kelly D (1998) *Helicobacter pylori* porCDAB and oorDABC Genes Encode Distinct Pyruvate: Flavodoxin and 2-Oxoglutarate: Acceptor Oxidoreductases Which Mediate Electron Transport to NADP. J Bacteriol 180: 1119-1128.
82. Kather B, K. Stingl, M. Van der Rest, K. Altendorf, and D. Molenaar (2000) Another Unusual Type of Citric Acid Cycle Enzyme in Helicobacter pylori: the Malate:Quinone Oxidoreductase. J Bacteriol 182: 3204-3209.
83. Gehring U, Arnon DI (1972) Purification and Properties of alpha-Ketoglutarate Synthase from a Photosynthetic Bacterium. J Biol Chem 247: 6963-6969.
84. Breese K, Boll M, Alt-Morbe J, Schagger H, Fuchs G (1998) Genes coding for the benzoyl-CoA pathway of anaerobic aromatic metabolism in the bacterium *Thauera aromatica*. Eur J Biochem 256: 148-154.
85. Jahn D, Verkamp E, Soll D (1992) Glutamyl-transfer RNA: a precursor of heme and chlorophyll biosynthesis. Trends in Biochemical Sciences 17: 215-218.
86. Lin JT, Goldman BS, Stewart V (1994) The Nasfedcba Operon for Nitrate and Nitrite Assimilation in Klebsiella-Pneumoniae M5al. J Bacteriol 176: 2551-2559.
87. Merrick MJ, Edwards RA (1995) Nitrogen control in bacteria. Microbiol Rev 59: 604-622.
88. van Veen HW (1997) Phosphate transport in prokaryotes: molecules, mediators and mechanisms. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 72: 299-315.
89. Prere MF, Chandler M, Fayet O (1990) Transposition in Shigella dysenteriae: isolation and analysis of IS911, a new member of the IS3 group of insertion sequences. J Bacteriol 172: 4090-4099.
90. Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, et al. (2006) Phosphonate utilization by the globally important marine diazotroph Trichodesmium. Nature 439: 68-71.
91. Kolowith LC, Ingall ED, Benner R (2001) Composition and cycling of marine organic phosphorus. Limnology and Oceanography 46: 309-320.
92. Klotz MG, Arp DJ, Chain PSG, El-Sheikh AF, Hauser LJ, et al. (2006) The complete genome sequence of the marine, chemolithoautotrophic, ammonia-oxidizing bacterium *Nitrosococcus oceani* ATCC19707. Appl Envir Microbiol In press.
93. Watson SW (1965) Characteristics of a marine nitrifying bacterium, *Nitrosocystis oceanus* Sp. N. Limnol Oceanogr 10: 274-289.

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

112. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) Improved microbial gene identification with GLIMMER. Nucleic Acids Res 27: 4636-4641.
113. Badger JH, Olsen GJ (1999) CRITICA: coding region identification tool invoking comparative analysis. Molecular Biology and Evolution 16: 512-524.
114. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research 28: 27-30.
115. McHardy AC, Goesmann A, Puhler A, Meyer F (2004) Development of joint application strategies for two microbial gene finders. Bioinformatics 20: 1622-1631.
116. Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, et al. (2003) GenDB--An open source genome annotation system for prokaryotic genomes. Nucleic Acids Res 31: 2187-2195.
117. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology 305: 567-580.
118. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology 340: 783-795.
119. Ren Q, Kang KH, Paulsen IT (2004) TransportDB: a relational database of cellular membrane transport systems. Nucleic Acids Res 32: D284-D288.
120. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, et al. (2006) The integrated microbial genomes (IMG) system. Nucl Acids Res %R 101093/nar/gkj024 34: D344-348.
121. Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). . Version 4 ed. Sunderland, Massachusetts: Sinauer Associates.
122. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res 25: 4876-4882.

Acknowledgments

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

Figure Captions

Figure 1. Circular map of the *Thiomicrospira crunogena* XCL-2 genome.
The outer two rings are protein-encoding genes, which are color-coded according to COG category. Rings 3 and 4 are tRNA and rRNA genes. Ring 5 indicates the location of a prophage (magenta), phosphonate/heavy metal resistance island (cyan), and four insertion sequences (red; two insertions at 2028543 and 2035034 are superimposed on this figure). The black circle indicates the deviation from the average %GC, and the purple and green circle is the GC skew (= [G-C]/[G+C]). Both the %GC and GC skew were calculated using a sliding window of 10,000 bp with a window step of 100.

Figure 2. Cell model for *Thiomicrospira crunogena* XCL-2, with an emphasis on ultrastructure, transport, energy, carbon metabolism, and chemotaxis.
Genes encoding virtually all of the steps for the synthesis of nucleotides and amino acids by canonical pathways are present, and are omitted here for simplicity. Electron transport components are yellow, and abbreviations are: NDH—NADH dehydrogenase; UQ—ubiquinone; bc_1—bc_1 complex; Sox—Sox system; cytC—cytochrome C; cbb_3—cbb_3-type cytochrome C oxidase. Methyl-accepting chemotaxis proteins (MCP) are fuchsia, as are MCP's with PAS domains or PAS folds. Influx and efflux transporter families with representatives in this genome are indicated on the figure, with the number of each type of transporter in parentheses. ATP-dependent transporters are red, secondary transporters are sky blue, ion channels are green, and unclassified transporters are purple. Abbreviations for transporter families are as follows: ABC – ATP-binding cassette superfamily; AGCS—Alanine or glycine:cation symporter family; AMT—Ammonium transporter family; APC—amino acid-polyamine-organocation family; ATP syn—ATP synthetase; BASS—Bile acid:Na^+ symporter family; BCCT—Betaine/carnitine/choline transporter family; CaCA—Ca^{2+}:cation antiporter family; CDF—cation diffusion facilitator family; CHR—Chromate ion transporter family; CPA—Monovalent cation:proton antiporter-1, -2, and -3 families; DAACS—Dicarboxylate/amino acid:cation symporter family; DASS—Divalent anion:Na^+ symporter family; DMT—Drug/metabolite transporter superfamily; FeoB—Ferrous iron uptake family; IRT—Iron/lead transporter superfamily; MATE—multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily, MATE family; McsS—Small conductance mechanosensitive ion channel family; MFS—Major facilitator superfamily; MgtE—Mg^{2+} transporter-E family; MIT—CorA metal ion transporter family; NCS2—Nucleobase:cation symporter-2 family; NRAMP—Metal ion transporter family; NSS—Neurotransmitter:sodium symporter family; P-ATP—P-type ATPase superfamily; Pit—Inorganic phosphate transporter family; PNaS—Phosphate:Na^+ symporter family; PnuC—Nicotamide mononucleotide uptake permease family; RhtB—Resistance to homoserine/threonine family; RND—Resistance-nodulation-cell division superfamily; SSS—Solute:sodium symporter family; SulP—Sulfate permease family; TRAP—Tripartite ATP-independent periplasmic transporter family; TRK—K^+ transporter family; VIC—Voltage-gated ion channel superfamily.

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

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

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

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

Figure 7. Models for glycolysis, gluconeogenesis, and the citric acid cycle in *Thiomicrospira crunogena* XCL-2.
Models for central carbon metabolism for cells under environmental conditions with A. sufficient reduced sulfur and oxygen; B. sulfide scarcity; C. oxygen scarcity; Green arrows represent the two 'non-canonical' citric acid cycle enzymes, 2-oxoglutarate oxidoreductase (2-OG OR) and malate: quinone oxidoreductase (MQO).

Figure 8. *Thiomicrospira crunogena* XCL-2 phosphonate operon.
The *DBR-1* genes are identical to eachother, as are the *DBR-2* genes. Gene abbreviations are: *DBR-1* and *2*—DNA breaking-rejoining enzymes; *hyp*—hypothetical protein; *phnFDCEGHIJKLMNP*—phosphonate operon; *chp*—conserved hypothetical protein. An asterisk marks the location of a region (within and upstream of tRNA-phe) with a high level of similarity to the 5' ends of the two direct repeat sequences noted in the figure. The transposase and integrase are actually a single CDS separated by a frameshift.

Figure 9. Numbers of methyl-accepting chemotaxis protein genes in *Thiomicrospira crunogena* XCL-2 and other proteobacteria.
T. crunogena is marked with an arrow. (A) depicts the total number of CDS's predicted to encode methyl-accepting chemotaxis protein genes (MCPs), while those CDSs with a PAS domain or fold are tallied in (B).

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

Item	Value
Chromosomes	1
Basepairs	2,427,734
GC content (%)	43.1
% coding	90.6
RNA-encoding genes	
tRNAs	43
16S-Ile $tRNA_{GAT}$-Ala $tRNA_{TGC}$-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

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

Number: *T. crunogena*	*N. oceani*	
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, *T. crunogena*)
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, *T. crunogena*)
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

*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	*T. crunogena*[a]	*E. coli*[b]
Hg^{+2}	0.01	0.01
Cu^{+2}	0.02	1
Ag^{+1}	0.02	0.02
Cd^{+2}	0.05	0.5
Co^{+2}	0.1	1
Ni^{+2}	0.1	1
Zn^{+2}	1	1
Cr^{+2}	1	5
Mn^{+2}	2	20

[a] *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.

[b]Data from [103].

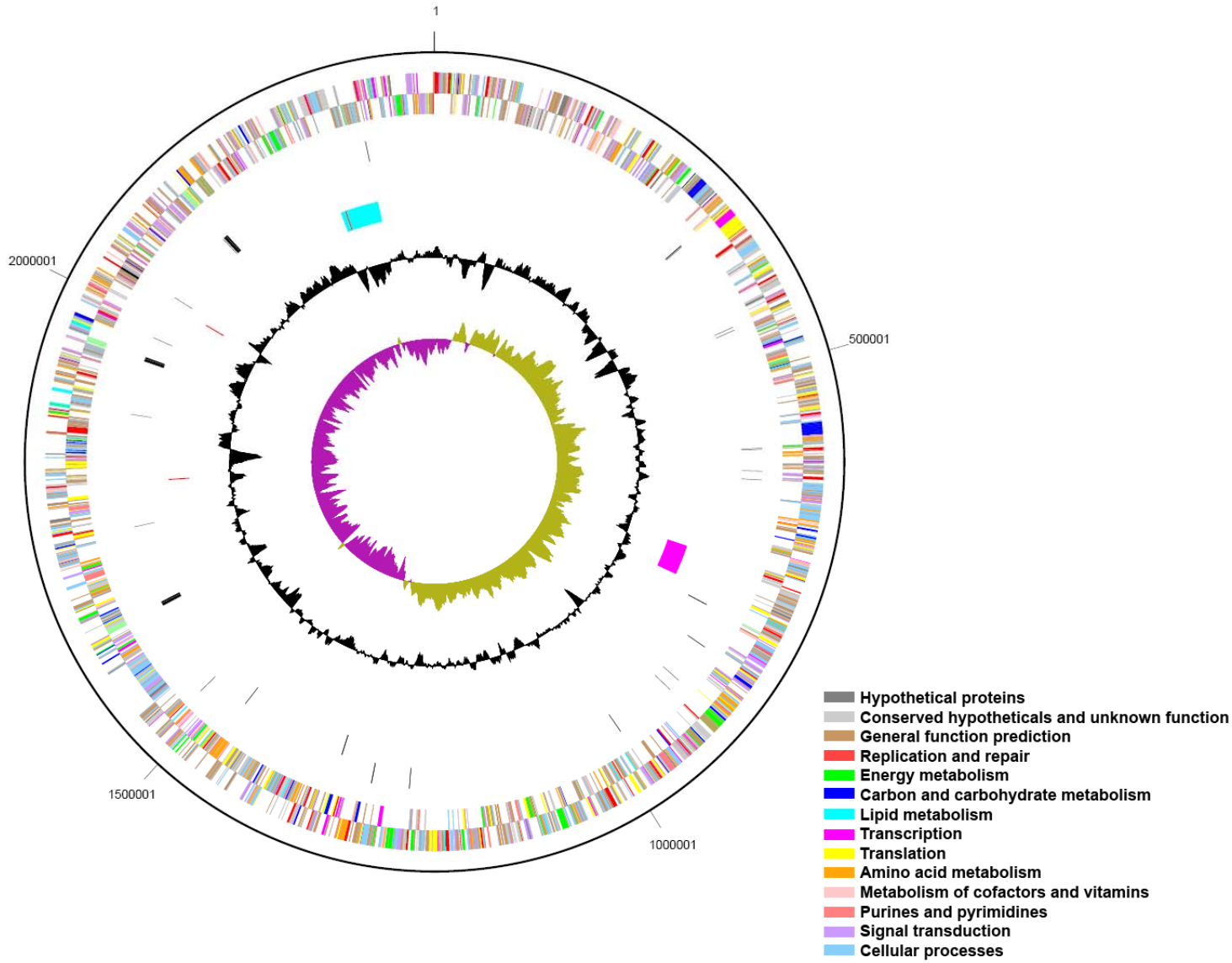

1
500001
1000001
1500001
2000001
Hypothetical proteins
Conserved hypotheticals and unknown function
General function prediction
Replication and repair
Energy metabolism
Carbon and carbohydrate metabolism
Lipid metabolism
Transcription
Translation
Amino acid metabolism
Metabolism of cofactors and vitamins
Purines and pyrimidines
Signal transduction
Cellular processes

Figure 1

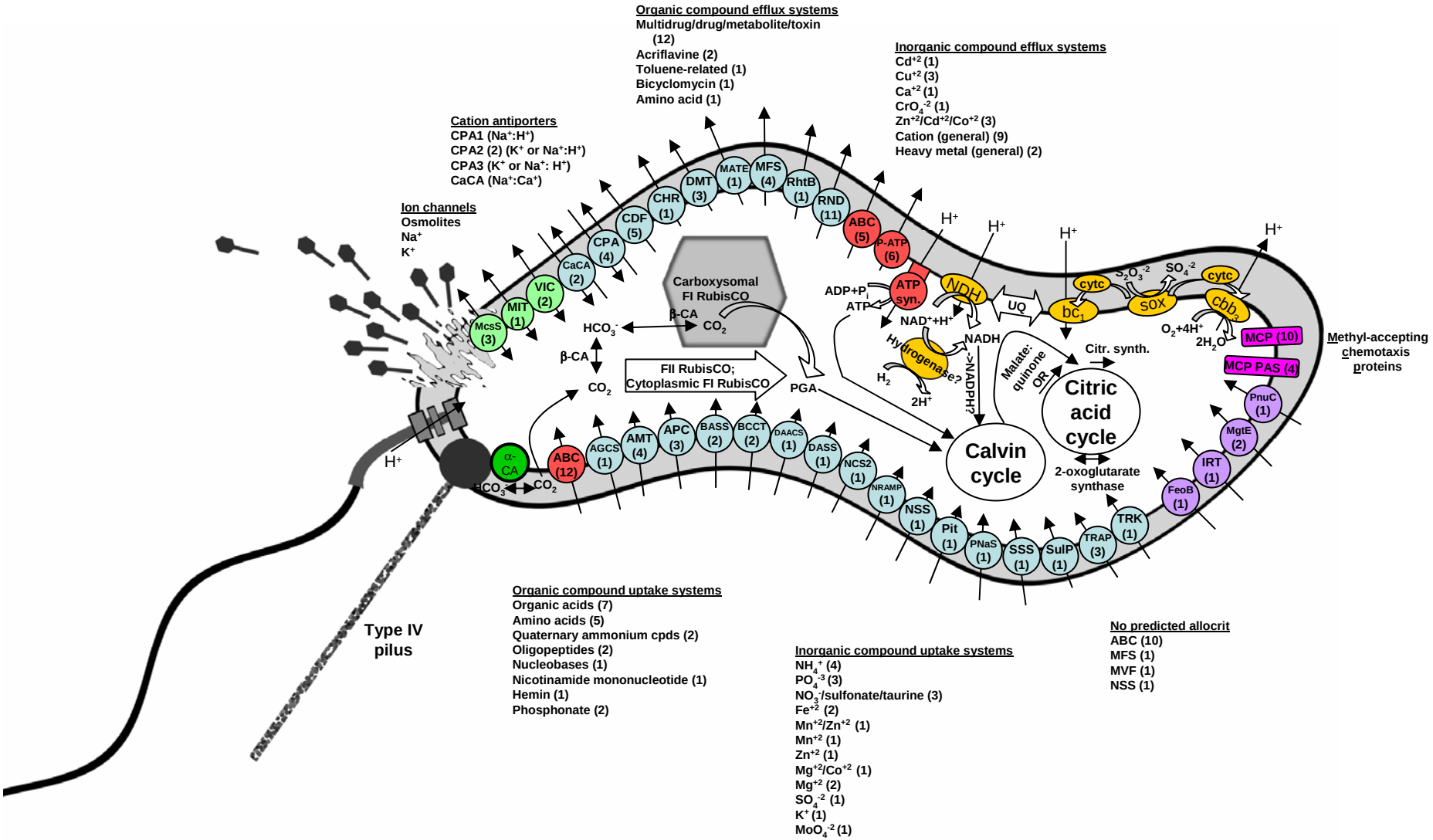
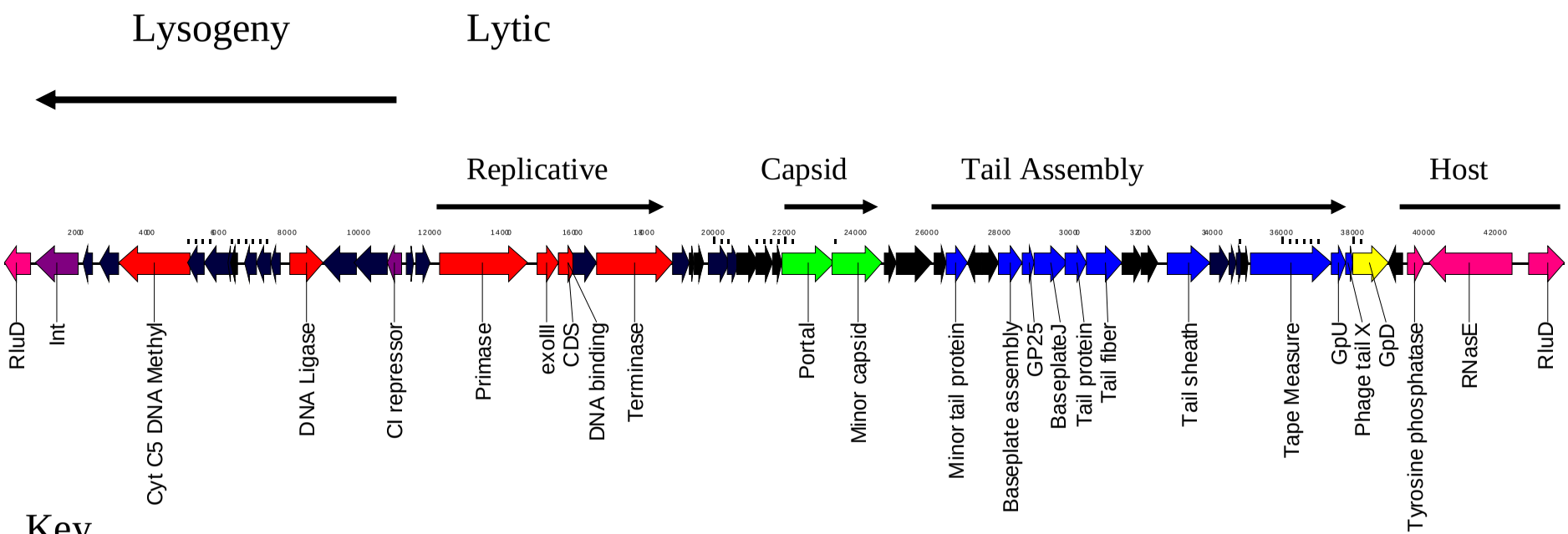

Organic compound efflux systems
Multidrug/drug/metabolite/toxin (12)
Acriflavine (2)
Toluene-related (1)
Bicyclomycin (1)
Amino acid (1)
Inorganic compound efflux systems
Cd+2 (1)
Cu+2 (3)
Ca+2 (1)
CrO4-2 (1)
Zn+2/Cd+2/Co+2 (3)
Cation (general) (9)
Heavy metal (general) (2)
Cation antiporters
CPA1 (Na+:H+)
CPA2 (2) (K+ or Na+:H+)
CPA3 (K+ or Na+: H+)
CaCA (Na+:Ca+)
Ion channels
Osmolites
Na+
K+
MATE (1)
MFS (4)
DMT (3)
RhtB (1)
CHR (1)
RND (11)
CDF (5)
ABC (5)
CPA (4)
P-ATP (6)
CaCA (2)
VIC (2)
MIT (1)
McsS (3)
Carboxysomal FI RubisCO
β-CA
HCO3-
CO2
ADP+Pi
ATP
ATP syn.
NDH
UQ
bc1
cytc
S2O3-2
SO4-2
SOX
cytc
cbb3
O2+4H+
2H2O
MCP (10)
MCP PAS (4)
Methyl-accepting chemotaxis proteins
NAD++H+
NADH
Hydrogenase?
H2
2H+
>NADPH?
Malate: quinone OR
Citr. synth.
Citric acid cycle
2-oxoglutarate synthase
FII RubisCO; Cytoplasmic FI RubisCO
PGA
Calvin cycle
α-CA
ABC (12)
AGCS (1)
AMT (4)
APC (3)
BASS (2)
BCCT (2)
DAACS (1)
DASS (1)
NCS2 (1)
NRAMP (1)
NSS (1)
Pit (1)
PNaS (1)
SSS (1)
SulP (1)
TRAP (3)
TRK (1)
FeoB (1)
IRT (1)
MgtE (2)
PnuC (1)
Type IV pilus
Organic compound uptake systems
Organic acids (7)
Amino acids (5)
Quaternary ammonium cpds (2)
Oligopeptides (2)
Nucleobases (1)
Nicotinamide mononucleotide (1)
Hemin (1)
Phosphonate (2)
Inorganic compound uptake systems
NH4+ (4)
PO4-3 (3)
NO3-/sulfonate/taurine (3)
Fe+2 (2)
Mn+2/Zn+2 (1)
Mn+2 (1)
Zn+2 (1)
Mg+2/Co+2 (1)
Mg+2 (2)
SO4-2 (1)
K+ (1)
MoO4-2 (1)
No predicted allocrit
ABC (10)
MFS (1)
MVF (1)
NSS (1)

Figure 2



Key

- Lysogeny
- DNA processing
- Capsid
- Tail
- Regulatory
- Host

Figure 3

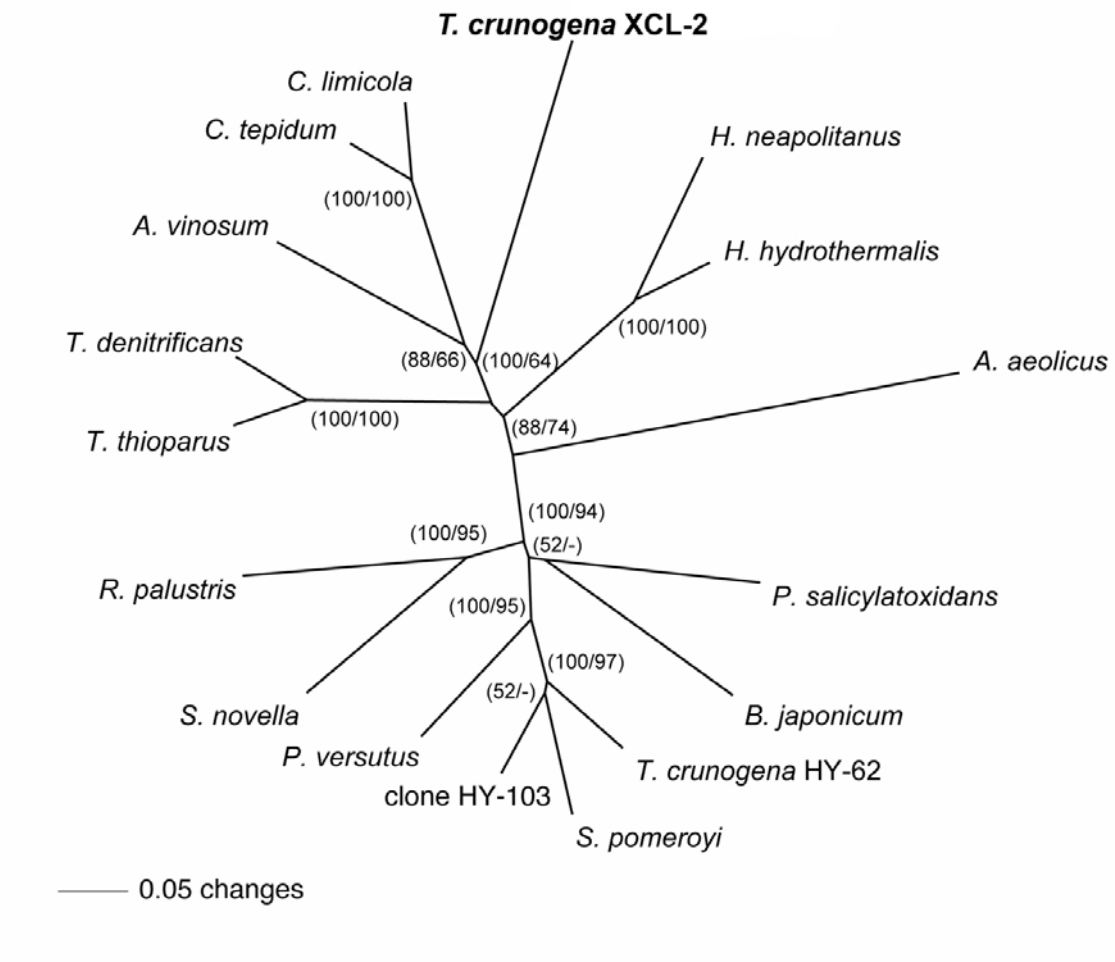

T. crunogena XCL-2
C. limicola
C. tepidum
(100/100)
A. vinosum
H. neapolitanus
H. hydrothermalis
(100/100)
T. denitrificans
(88/66)
(100/64)
A. aeolicus
(100/100)
T. thioparus
(88/74)
(100/94)
(100/95)
(52/-)
R. palustris
P. salicylatoxidans
(100/95)
(100/97)
(52/-)
S. novella
B. japonicum
P. versutus
T. crunogena HY-62
clone HY-103
S. pomeroyi
0.05 changes

Figure 4

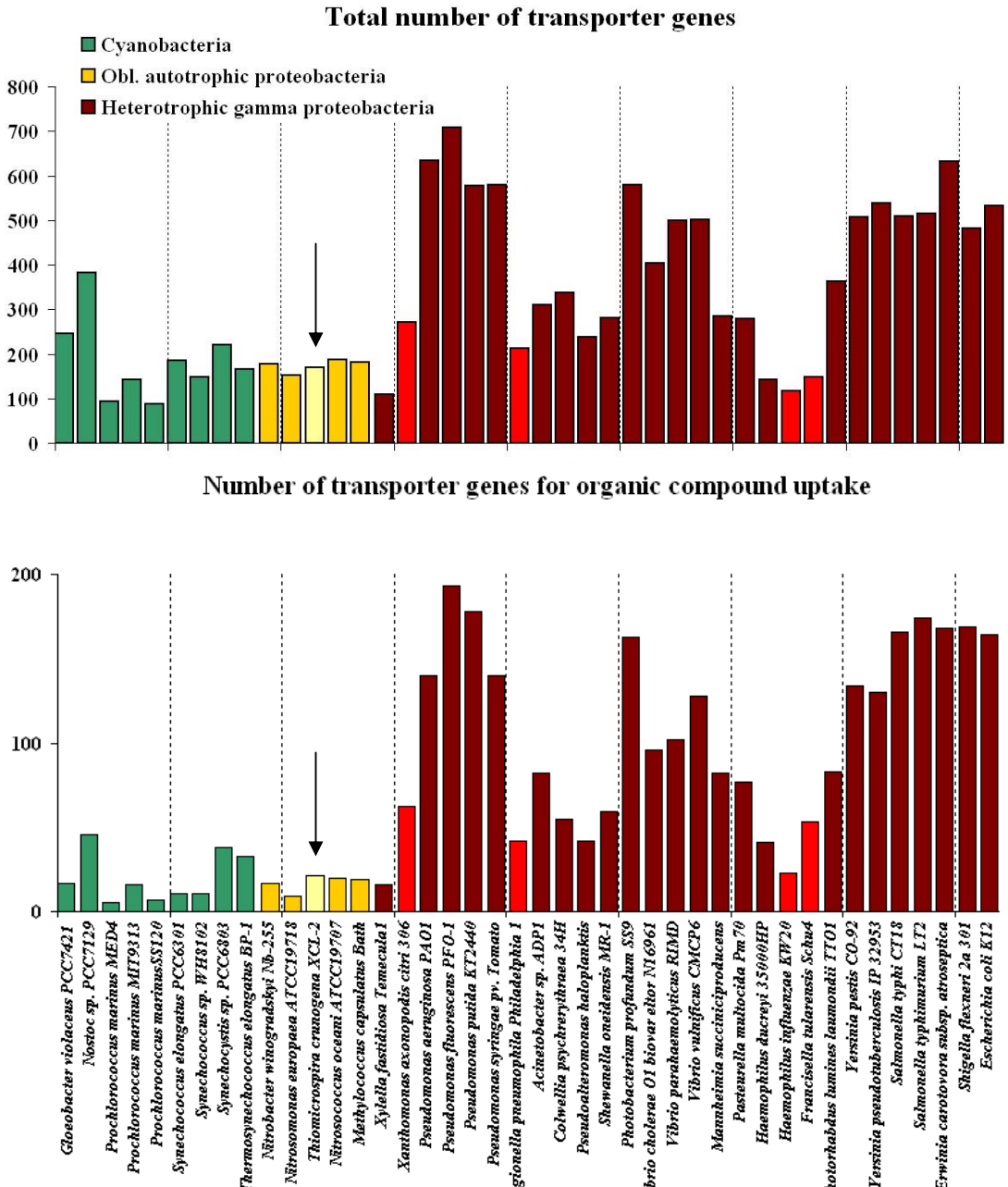

Total number of transporter genes
Cyanobacteria
Obl. autotrophic proteobacteria
Heterotrophic gamma proteobacteria
800
700
600
500
400
300
200
100
0
Number of transporter genes for organic compound uptake
200
100
0
Gloeobacter violaceus PCC7421
Nostoc sp. PCC7120
Prochlorococcus marinus MED4
Prochlorococcus marinus MIT9313
Prochlorococcus marinusSS120
Synechococcus elongatus PCC6301
Synechococcus sp. WH8102
Synechocystis sp. PCC6803
Thermosynechococcus elongatus BP-1
Nitrobacter winogradskyi Nb-255
Nitrosomonas europaea ATCC19718
Thiomicrospira crunogena XCL-2
Nitrosococcus oceani ATCC19707
Methylococcus capsulatus Bath
Xylella fastidiosa Temecula1
Xanthomonas axonopodis citri 306
Pseudomonas aeruginosa PAO1
Pseudomonas fluorescens PFO-1
Pseudomonas putida KT2440
Pseudomonas syringae pv. Tomato
Legionella pneumophila Philadelphia 1
Acinetobacter sp. ADP1
Colwellia psychrerythraea 34H
Pseudoalteromonas haloplanktis
Shewanella oneidensis MR-1
Photobacterium profundum SS9
Vibrio cholerae O1 biovar eltor N16961
Vibrio parahaemolyticus RIMD
Vibrio vulnificus CMCP6
Mannheimia succiniciproducens
Pasteurella multocida Pm70
Haemophilus ducreyi 35000HP
Haemophilus influenzae KW20
Francisella tularensis Schu4
Photorhabdus lumines laumondii TTO1
Yersinia pestis CO-92
Yersinia pseudotuberculosis IP 32953
Salmonella typhi CT18
Salmonella typhimurium LT2
Erwinia carotovora subsp. atrosepticа
Shigella flexneri 2a 301
Escherichia coli K12

Figure 5

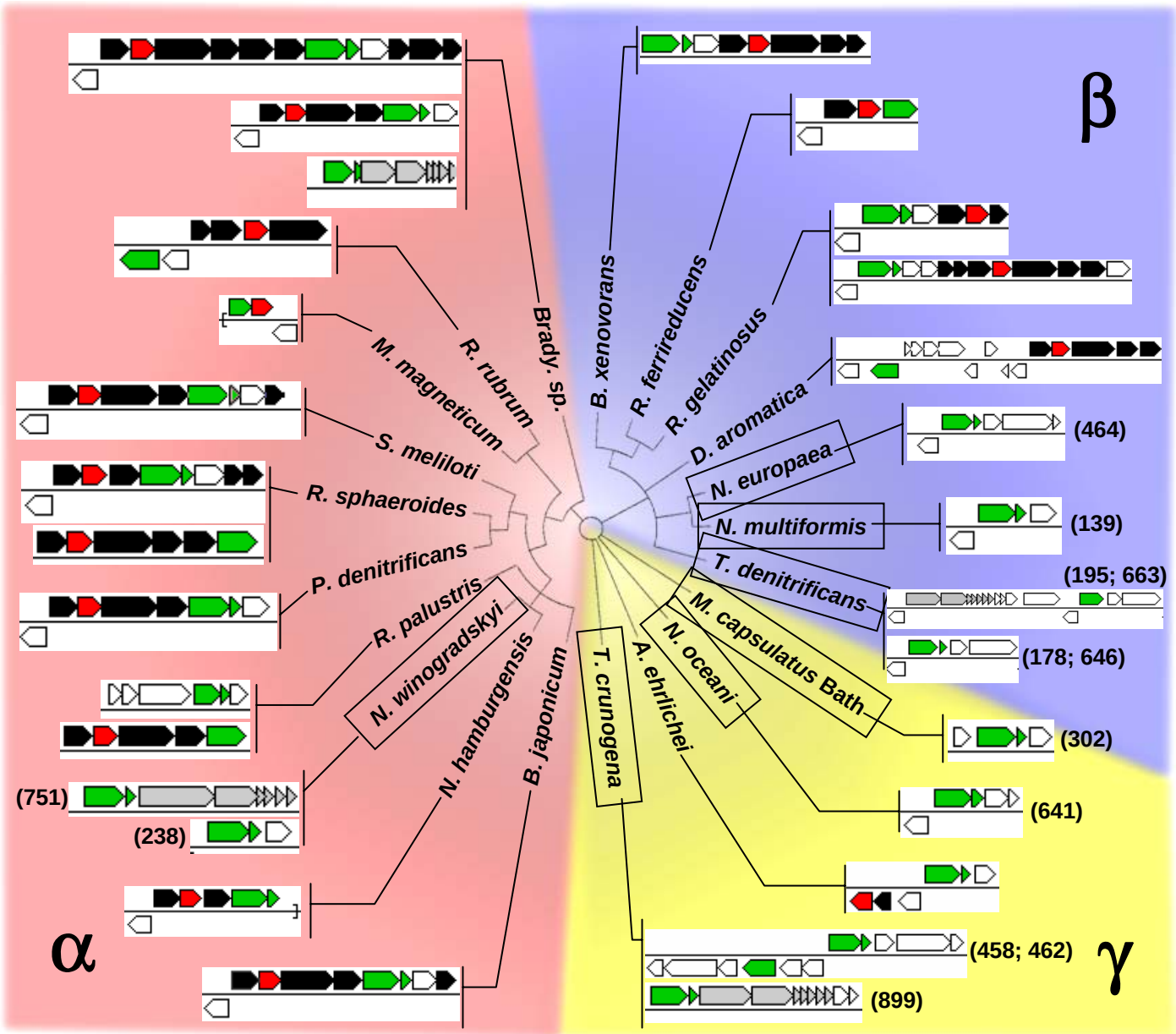

α
β
γ
Brady. sp.
R. rubrum
M. magneticum
S. meliloti
R. sphaeroides
P. denitrificans
R. palustris
N. winogradskyi
N. hamburgensis
B. japonicum
B. xenovorans
R. ferrireducens
R. gelatinosus
D. aromatica
N. europaea
N. multiformis
T. denitrificans
M. capsulatus Bath
N. oceani
A. ehrlichei
T. crunogena
(751)
(238)
(464)
(139)
(195; 663)
(178; 646)
(302)
(641)
(458; 462)
(899)

Figure 6

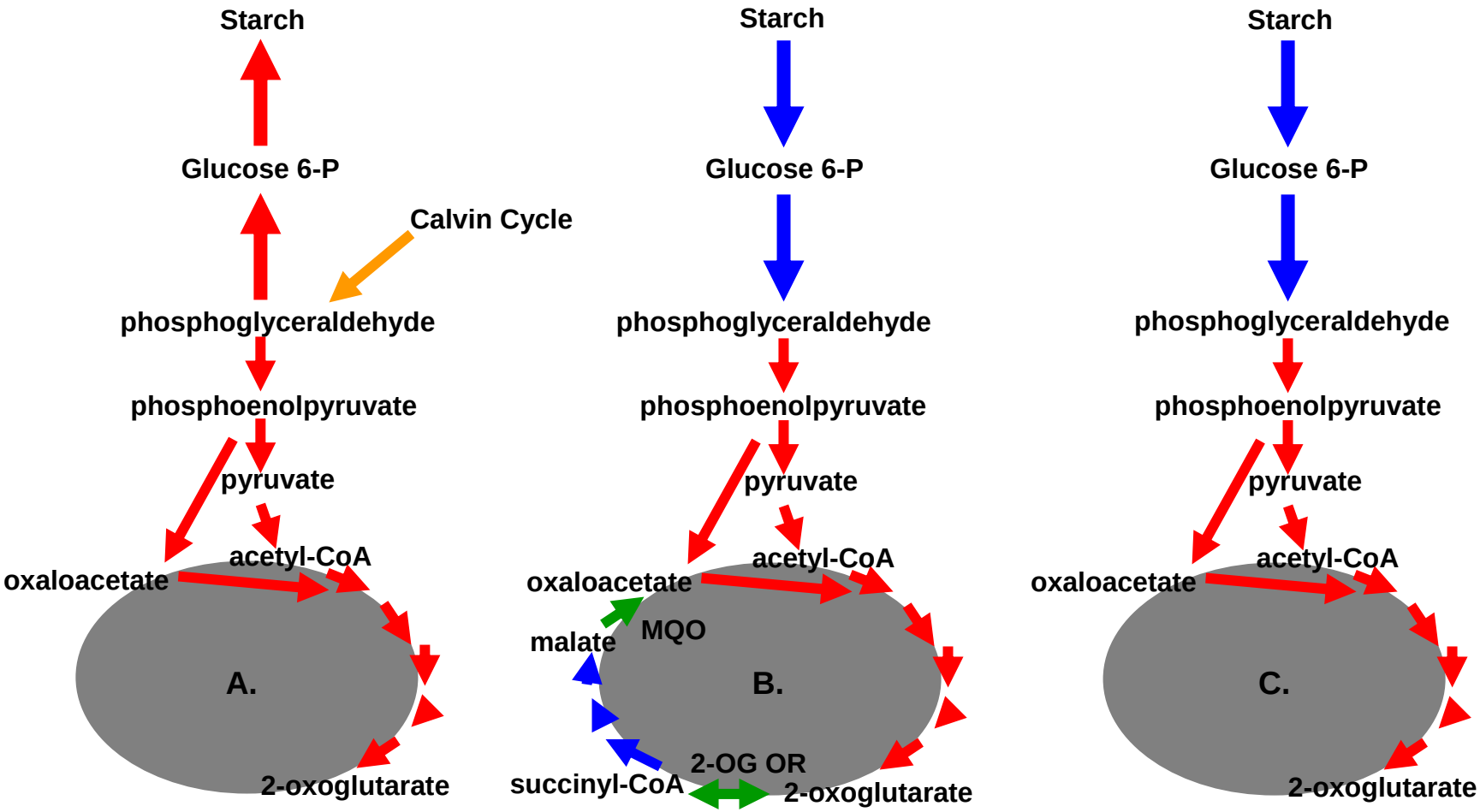

Starch
Glucose 6-P
Calvin Cycle
phosphoglyceraldehyde
phosphoenolpyruvate
pyruvate
acetyl-CoA
oxaloacetate
A.
2-oxoglutarate
Starch
Glucose 6-P
phosphoglyceraldehyde
phosphoenolpyruvate
pyruvate
acetyl-CoA
oxaloacetate
malate
MQO
B.
2-OG OR
succinyl-CoA
2-oxoglutarate
Starch
Glucose 6-P
phosphoglyceraldehyde
phosphoenolpyruvate
pyruvate
acetyl-CoA
oxaloacetate
C.
2-oxoglutarate

Figure 7

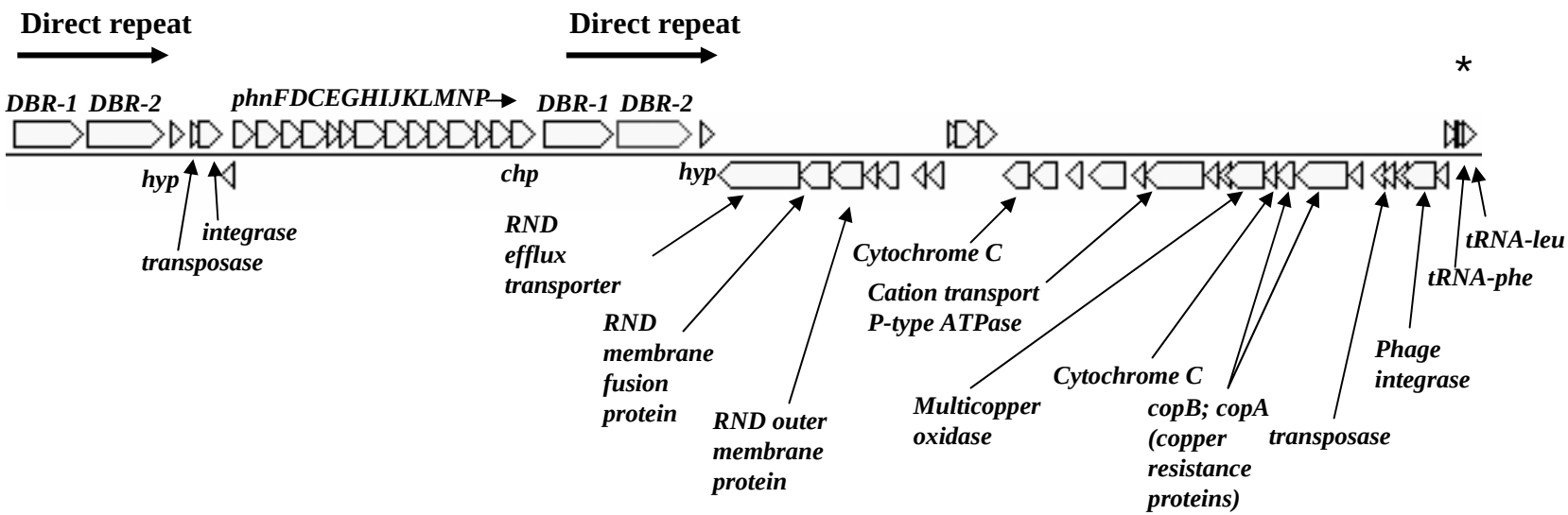

Direct repeat
Direct repeat
*
DBR-1 DBR-2
phnFDCEGHIJKLMNP
DBR-1 DBR-2
hyp
integrase
transposase
chp
hyp
RND efflux transporter
RND membrane fusion protein
RND outer membrane protein
Cytochrome C
Cation transport P-type ATPase
Multicopper oxidase
Cytochrome C
copB; copA (copper resistance proteins)
transposase
Phage integrase
tRNA-phe
tRNA-leu

Figure 8

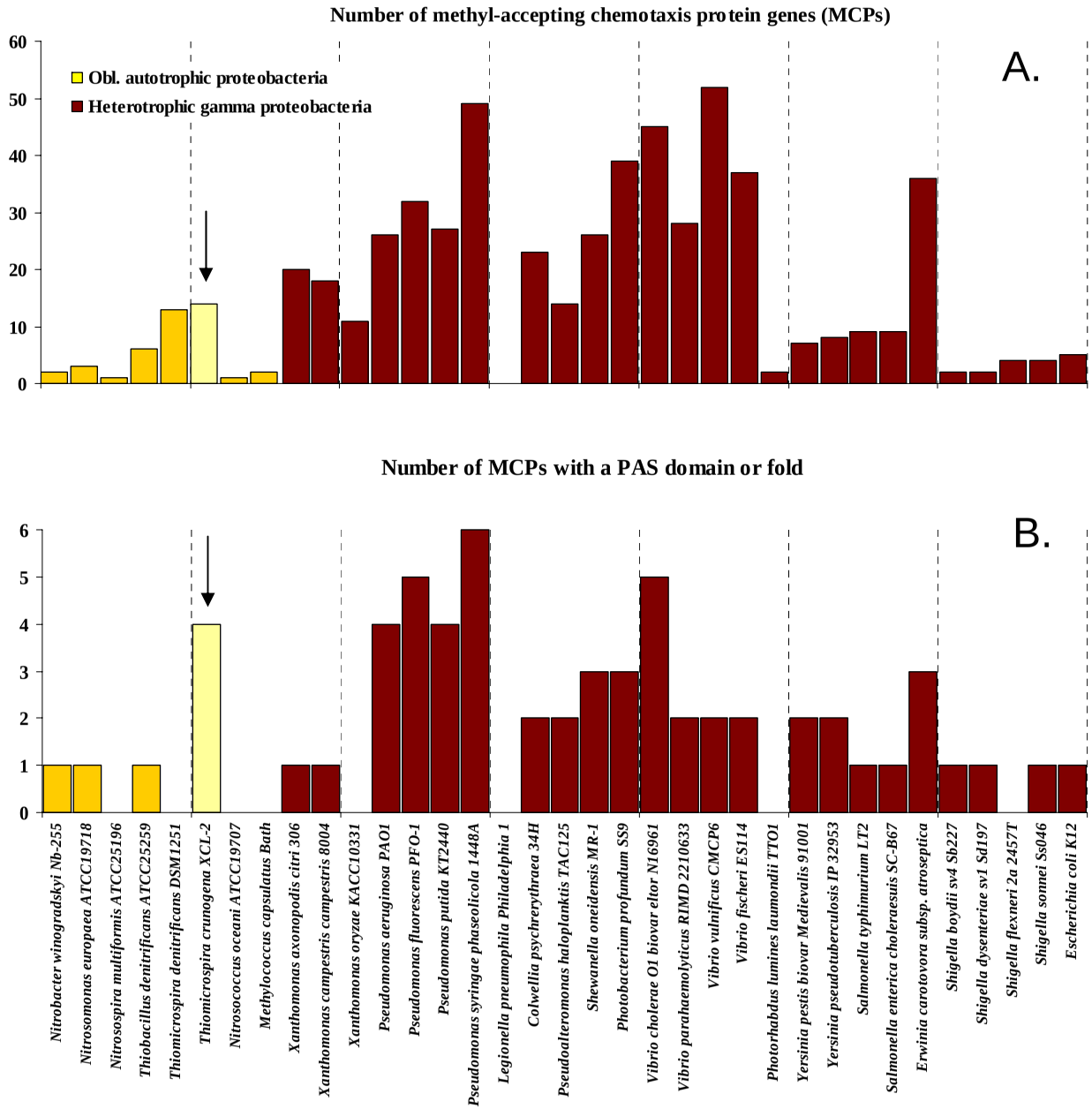

Number of methyl-accepting chemotaxis protein genes (MCPs)
A.
Obl. autotrophic proteobacteria
Heterotrophic gamma proteobacteria
60
50
40
30
20
10
0
Number of MCPs with a PAS domain or fold
B.
6
5
4
3
2
1
0
Nitrobacter winogradskyi Nb-255
Nitrosomonas europaea ATCC19718
Nitrosospira multiformis ATCC25196
Thiobacillus denitrificans ATCC25259
Thiomicrospira denitrificans DSM1251
Thiomicrospira crunogena XCL-2
Nitrosococcus oceani ATCC19707
Methylococcus capsulatus Bath
Xanthomonas axonopodis citri 306
Xanthomonas campestris campestris 8004
Xanthomonas oryzae KACC10331
Pseudomonas aeruginosa PAO1
Pseudomonas fluorescens PFO-1
Pseudomonas putida KT2440
Pseudomonas syringae phaseolicola 1448A
Legionella pneumophila Philadelphia 1
Colwellia psychrerythraea 34H
Pseudoalteromonas haloplanktis TAC125
Shewanella oneidensis MR-1
Photobacterium profundum SS9
Vibrio cholerae O1 biovar eltor N16961
Vibrio parahaemolyticus RIMD 2210633
Vibrio vulnificus CMCP6
Vibrio fischeri ES114
Photorhabdus lumines laumondii TTO1
Yersinia pestis biovar Medievalis 91001
Yersinia pseudotuberculosis IP 32953
Salmonella typhimurium LT2
Salmonella enterica choleraesuis SC-B67
Erwinia carotovora subsp. atroseptica
Shigella boydii sv4 Sb227
Shigella dysenteriae sv1 Sd197
Shigella flexneri 2a 2457T
Shigella sonnei Ss046
Escherichia coli K12

Figure 9