The genome of the intracellular bacterium of the coastal bivalve, Solemya velum: A blueprint for thriving in and out of symbiosis

AUTHORS: Oleg Dmytrenko[1], Shelbi L. Russell[1], Wesley T. Loo[1], Kristina M. Fontanez[2], Li Liao[3], Guus Roeselers[4], Raghav Sharma[4], Frank J. Stewart[5], Irene L. G. Newton[6], Tanja Woyke[7], Dongying Wu[8], Jenna Morgan Lang[8], Jonathan A. Eisen[8]*, Colleen M. Cavanaugh[1]*

1. Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4081 Biological Laboratories, Cambridge, MA, 02138, USA

2. Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139, USA

3. SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, 200136, China

4. Microbiology & Systems Biology Group, TNO, Utrechtseweg 48, 3704HE, Zeist, The Netherlands

5. School of Biology, Georgia Institute of Technology,_Atlanta GA 30332-0230, USA

6. Indiana University, Department of Biology, 1001 East 3rd Street, Jordan Hall Bloomington, IN 47405, USA

7. DOE Joint Genome Institute, 2800 Mitchell Drive_Walnut Creek, CA 94598, USA

8. UC Davis Genome Center, 451 E. health Sciences Drive, David, CA 94616-8816

** To whom correspondence may be addressed. Colleen Cavanaugh, Harvard: cavanaug@fas.harvard.edu and/or Jonathan Eisen, University of California at Davis: jaeisen@ucdavis.edu*

September 25, 2014

ACKNOWLEDGMENTS:

This work was funded by grant 0412205 of the US National Science Foundation (NSF) and was made possible with the generous support of the U.S. Department of Energy Joint Genome Institute (JGI). Work by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We would like to express special thanks to Grace Pai for creating Sanger

The genome of the intracellular bacterium of the coastal bivalve, Solemya velum: A blueprint for thriving in and out of symbiosis

sequencing libraries and Shannon Smith and Terry Utterback for coordinating sequencing at TIGR

DISCLAIMER:

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: A blueprint for thriving in and out of symbiosis

Oleg Dmytrenko[1], Shelbi L. Russell[1], Wesley T. Loo[1], Kristina M. Fontanez[2], Li Liao[3], Guus Roeselers[4], Raghav Sharma[5], Frank J. Stewart[5], Irene L.G. Newton[6], Tanja Woyke[7], Dongying Wu[8], Jenna Morgan Lang[8], Jonathan A. Eisen[8]*, Colleen M. Cavanaugh[1]*

[1]Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4081 Biological Laboratories, Cambridge, MA, 02138, USA

[2] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139, USA

[3]SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai, 200136, China

[4]Microbiology & Systems Biology Group, TNO, Utrechtseweg 48, 3704HE, Zeist, The Netherlands

[5]School of Biology, Georgia Institute of Technology, Atlanta GA 30332-0230, USA

[6]Indiana University, Department of Biology, 1001 East 3rd Street, Jordan Hall Bloomington, IN 47405, USA

[7]DOE Joint Genome Institute, 2800 Mitchell Drive Walnut Creek, CA 94598, USA

[8]UC Davis Genome Center, 451 East Health Sciences Drive Davis, CA 95616-8816, USA

*Corresponding authors

Email addresses:

OD: odmytren@fas.harvard.edu

SLR: shelbirussell@fas.harvard.edu

WTL: wesleyloo@fas.harvard.edu

KMF: fontanez@mit.edu

LL: liaoli@pric.gov.cn

GR: guus.roeselers@tno.nl

FJS: frank.stewart@biology.gatech.edu

RS: rsharma61@gatech.edu

ILGN: irnewton@indiana.edu

TW: twoyke@lbl.gov

DW: dygwu@ucdavis.edu

JML: jnamorgan@ucd.edu

JAE: jaeisen@ucdavis.edu

CMC: cavanaug@fas.harvard.edu

Abstract

Background: Symbioses between chemoautotrophic bacteria and marine invertebrates are rare examples of living systems that are virtually independent of photosynthetic primary production. These associations have evolved multiple times in marine habitats, such as deep-sea hydrothermal vents and reducing sediments, characterized by steep gradients of oxygen and reduced chemicals. Due to difficulties associated with maintaining these symbioses in the laboratory and culturing the symbiotic bacteria, studies of chemosynthetic symbioses rely heavily on culture independent methods. The symbiosis between the coastal bivalve, *Solemya velum*, and its intracellular symbiont is a model for chemosynthetic symbioses given its accessibility in intertidal environments and the ability to maintain it under laboratory conditions. To better understand this symbiosis, the genome of the *S. velum* endosymbiont was sequenced.

Results: Relative to the genomes of obligate symbiotic bacteria, which commonly undergo erosion and reduction, the *S. velum* symbiont genome was large (2.86 Mb), GC-rich (50.4%), and contained a large number (78) of mobile genetic elements. Comparative genomics identified sets of genes specific to the chemosynthetic lifestyle and necessary to sustain the symbiosis. In addition, a number of inferred metabolic pathways and cellular processes, including heterotrophy, branched electron transport, and motility, suggested that besides the ability to function as an endosymbiont, the bacterium may have the capacity to live outside the host.

Conclusions: The physiological dexterity indicated by the genome substantially improves our understanding of the genetic and metabolic capabilities of the *S. velum* symbiont and the breadth of niches the partners may inhabit during their lifecycle.

Keywords: symbiosis, chemosynthesis, sulfur oxidation, respiratory flexibility, H^+/Na^+ - membrane cycles, Calvin cycle, pyrophosphate-dependent phosphofructokinase, TCA cycle, motility, mobile genetic elements.

Background

Symbiosis is one of the major driving forces of evolutionary adaptation. Chloroplasts and mitochondria are examples of ancient symbiotic partnerships which played key roles in the emergence and diversification of eukaryotic life on Earth [1]. Bacteria have been found in symbioses with organisms as diverse as plants, insects, marine invertebrates, and protists [2-5], expanding metabolic capabilities of the partners and allowing them to occupy otherwise unavailable ecological niches. Despite the ubiquity of such mutualistic associations and their importance to health and the environment, studies of many host-associated microorganisms have been complicated by difficulties in both the maintenance of symbiotic organisms in culture and the inability to genetically manipulate them. However, progress in culture-independent techniques has allowed for rapid advances in understanding symbiosis diversity, evolution, genetics, and physiology [6-8].

Symbioses between chemoautotrophic bacteria and invertebrates are ubiquitous in reducing marine habitats, such as deep-sea hydrothermal vents and coastal sediments. In these environments, the symbiotic bacteria derive energy by oxidizing reduced inorganic molecules (e.g., sulfide) and fix carbon dioxide for biomass production. Their hosts have evolved behavioral, physiological, and biochemical adaptations for capturing and delivering the required electron donors and acceptors to the symbionts. In return, these invertebrates obtain their nutrition from bacterial chemosynthesis [5, 9].

Solemya velum and its symbionts is one of the best-described chemoautotrophic symbioses. The host, a protobranch bivalve, lives in coastal nutrient-rich sediments

where it builds Y-shaped burrows that span the oxic-anoxic interface, allowing access to both reduced inorganic sulfur as an energy source and oxygen for use as a terminal oxidant [10]. The symbionts, which constitute a single 16S rRNA phylotype of γ-proteobacteria [11], are localized to specialized epithelial cells (bacteriocytes) in the gills, separated from the cytoplasm by a peribacterial membrane. Using energy from the oxidation of sulfide, the symbionts fix CO_2 via the Calvin-Benson-Bassham Cycle [12, 13]. Primary production in the symbionts sustains the host, which has only a rudimentary gut and cannot effectively filter-feed [14, 15]. Many key properties of this symbiosis still remain to be characterized, including the exchange of metabolites and signals between the symbiont and the host and the mechanism of symbiont acquisition at each new host generation (i.e., symbiont transmission mode).

The mode by which *S. velum* acquires its symbionts has important implications for understanding symbiont genome evolution. Symbiont-specific genes have been amplified from the host ovarian tissue of both *S. velum* and its congener, *S. reidi* [16, 17], raising the hypothesis that symbionts are transmitted maternally (vertically) between successive host generations via the egg. Vertical transmission has also been inferred in deep-sea clams of the Vesicomyidae [18, 19], in which symbionts have a reduced genome size (1.2 Mb) and appear to be obligately associated with their host [20-23]. In vesicomyid symbioses, host and symbiont phylogenies are largely congruent, a pattern consistent with vertical symbiont transmission [24]. Nonetheless, instances of lateral symbiont movement among some vesicomyids have been inferred based on decoupling of symbiont and host evolutionary trajectories [25], bringing diverse symbiont strains into contact and creating opportunities for symbiont genome evolution via recombination [26, 27]. In the Solemyidae, on the other hand, symbionts of different *Solemya* species are scattered across phylogenetic clades (i.e., polyphyly), indicating distinct evolutionary origins relative to the monophyly of the hosts [5, 28]. A preliminary analysis was unable

to definitively resolve the extent of genetic coupling between the *S. velum* host and its symbionts in populations along the southern New England coast [26]. These patterns may be the result of a physical decoupling of symbiont and host lineages, possibly due to lateral symbiont transmission between hosts.

It is therefore possible that transmission in solemyid symbioses, as in vesicomyids, involves a combination of both vertical passage through the maternal germ line and lateral acquisition of symbionts from the environment or other co-occurring host individuals. Such a mixed transmission mode could strongly impact symbiont genome evolution by creating opportunities for lateral gene transfer, relieving the constraints of genetic bottlenecks imposed by strict vertical transmission [29, 30], and imposing selective pressures for the maintenance of diverse functions in the symbiont genome that would mediate survival outside the host. The genome of the *S. velum* symbiont will provide insights into the transmission mode of this symbiont, define a framework for examining its physiological adaptations, and supply a reference sequence for future studies of the ecology and evolution of solemyid symbionts.

Here we present an analysis of the genome from the *S. velum* endosymbiont. First, genes that encode core metabolic functions are discussed. Emphasis is placed on bioenergetics, autotrophy, heterotrophy, and nitrogen metabolism, which indicate metabolic potential beyond strict chemolithoautotrophy. Genes encoding cellular functions that pertain to the symbiotic lifestyle are also analyzed. A special focus is on the processes, such as membrane transport, sensing, and motility that may be involved in interactions of the symbionts with the host and the environment. Wherever appropriate, the gene content is compared to that of free-living and host-associated bacteria, in particular the intracellular chemosynthetic symbionts of the vesicomyid clams, *Calyptogena magnifica* [22] and *Calyptogena okutanii* [20], the vestimentiferan tubeworms, *Riftia pachyptila* [31] and *Tevnia jerichona* [32], the scaly-foot snail,

Crysomollon squamiferum, [33] and the marine oligochaete worm, *Olavius algarvensis,* [34]. This comprehensive analysis defines the *S. velum* symbiont as a metabolically versatile bacterium adapted to living inside the host but also potentially capable of survival on the outside. It informs attempts to culture the symbionts and generates multiple intriguing hypotheses that now await experimental validation.

Results and discussion

General genome features

The genome of the *S. velum* symbiont consists of 10 non-overlapping scaffolds, totaling 2,702,453 bp, with an average G + C content of 51%. The three largest scaffolds (1.21 Mb, 0.89 Mb, 0.54 Mb) contain 97.8% of the total genomic sequence and 98.4% of the predicted genes (Table S1 in Additional file 1). Assembly of the scaffolds into a closed genome was prevented by stretches of single nucleotides or groups of a few nucleotides repeated up to 70 times that could not be spanned. However, the high depth of sequence coverage and the presence of all 31 core bacterial phylogenetic gene markers [35] suggest that most gene-coding regions were detected in the analysis. Nevertheless, as the genome is not closed, a definitive list of all symbiont genes could not be made.

An overview of the *S. velum* symbiont genome compared to selected symbiotic and free-living γ-proteobacteria, including other thiotrophs, is presented in Table 1. Briefly, 90.7% of the genome sequence encodes 2,757 genes, on average 885 bp long. 2,716 (98.5%) genes are protein-coding. Function was predicted for 1,988 (72.1%) of all the genes, while 769 (27.9%) were identified as encoding hypothetical proteins. 382 genes (13.8%) have one or more paralog in the genome, with the largest paralogous group encoding transposases associated with mobile elements. The genome contains a single ribosomal RNA (rRNA) operon and 38 transfer RNAs (tRNA) corresponding to the

20 standard proteinogenic amino acids. Due to the wobble base-pairing [36], tRNAs for each given amino acid can pair with any codon in the genome for that amino acid (Table S2).

A model of the symbiont cell based on functional predictions is presented in Figure 1. When grouped into COG categories [37], the largest number of genes within the genome of the *S. velum* symbiont was associated with metabolism of coenzymes, transcription, posttranscriptional modification of proteins, cell division, DNA replication, and energy metabolism (Figure 2). Based on a BLASTN [38] search against the NCBI-nr database analyzed by MEGAN [39], 1,735 of the genes in the genome were assigned to γ-proteobacteria, mainly other sulfur-oxidizing symbionts (197 genes) and bacteria from the order of Chromatiales (184 genes). Among the genes within γ-proteobacteria, 897 could not be assigned to a lower-level taxon in the NCBI taxonomy. 37 genes had the closest matches to eukaryotes and 6 to archaea. No taxa could be assigned to 29 genes, while 212 genes had no hits in the NCBI-nr database (Figure 3). The majority of the sequences designated as “eukaryotic” were hypothetical and produced low percent amino acid identity matches in the BLASTN search.

Metabolic Functions

1. Chemolithotrophy

The *S. velum* symbiont, and chemoautotrophic symbionts in general, are remarkable in their ability to support almost all the metabolic needs of their metazoan hosts with energy derived from thiotrophy. Present genome data illustrate the ability of the *S. velum* symbiont to oxidize both hydrogen sulfide and thiosulfate via diverse pathways, in agreement with previous measurements of symbiont gene expression [40] and *in vitro* experiments showing that both substrates can stimulate carbon fixation in the symbionts [10, 13]. The *S. velum* symbiont genes involved in the oxidation of

reduced sulfur species are most closely related to those of the purple sulfur γ-proteobacterium, *Allochromatium vinosum* (Figure 4), in which the genetic components and the biochemical mechanisms of sulfur metabolism have been well characterized [41].

Periplasmic sulfide and thiosulfate oxidation: In the periplasm of the *S. velum* symbiont, sulfide, thiosulfate, and, possibly, elemental sulfur, may be oxidized for energy by the Sox system, which is represented in the genome (Figure 4). The identified SoxYZAXB, flavocytochrome *c* dehydrogenase (FccAB), and type I and IV sulfide-quinone reductases (Sqr) potentially reduce cytochromes *c* and quinones, which along the course of the electron-transport chain translate into membrane-ion gradients, NADH, and ATP, ultimately fueling biosynthetic and other energy-requiring cellular processes, including autotrophy (Figure 1). In *A. vinosum* and the green non-sulfur bacterium, *Chlorobium tepidum*, SoxYZ, SoxAX, and SoxB proteins participate in the formation of transient sulfur deposits as intermediates during sulfur oxidation [42]. In fact, sulfur deposits are common to all known sulfur-oxidizing bacteria (SOBs) which, like the *S. velum* symbiont, lack SoxCD sulfur dehydrogenase (Figure 4) [43], including the symbionts of the hydrothermal vent tubeworm, *R. pachyptila* [31, 44], and the clam, *C. magnifica* [22, 45]. Microscopically-detectable intracellular or extracellular sulfur has not been observed either in the symbiont-containing gills of *S. velum* or directly within the symbionts (Cavanaugh, unpublished observation). Absence of sulfur deposits may be attributed to a very rapid consumption of any available reduced sulfur substrate. This agrees with the fact that the *S. velum* symbionts have the highest known carbon fixation rate, and, hence, demand for energy, of all the studied chemosynthetic symbionts, i.e., 65 μmol min^{-1} g of $protein^{-1}$ [13] compared to 0.45 μmol min^{-1} g of $protein^{-1}$ of the next highest rate measured in the symbionts of *R. pachyptila* [46]. Alternatively, in the *S. velum* symbionts intermediate sulfur may be stored in a chemical form that is not easily

observed microscopically.

Cytoplasmic sulfide oxidation: Energy generating oxidation of sulfide to sulfite may be catalyzed in the cytoplasm of the *S. velum* symbiont by the reverse-acting dissimilatory sulfite reductase (rDsr) pathway (Figure 1). All of the enzymes and accessory proteins required for this pathway are encoded in a *dsrABEFHCMKLJOPNRS* operon (Figure 4). While multiple homologues of *dsrC* were identified outside the *dsr* operon, these genes did not encode the two conserved C-terminal cysteines required for the protein to function [47, 48]. The DsrC enzyme likely mediates transfer of electrons from sulfide reductase, DsrAB, to a transmembrane electron transport complex DsrKMJOP, an entry point for electrons derived from cytoplasmic oxidation of sulfur into the electron transport chain [49]. rDsr may be the key energy-generating pathway in the symbiont, as sulfide has a six-fold higher effect on carbon fixation in *S. velum* [13] compared to thiosulfate oxidized by the Sox pathway.

Sulfite oxidation: Sulfite generated by rDsr may be further oxidized to sulfate in the cytoplasm by a sequential action of APS reductase (AprABM) and an ATP-generating ATP sulfurylase (Sat) (Figure 1 and Figure 4). Identification of these genes agrees with measured Apr and Sat activity in the symbiont-containing *S. velum* tissue [50]. Sulfate generated in this pathway may be exported from the cytoplasm via a sulfate-bicarbonate antiporter SulP (Figure 1). While electrons obtained from the oxidation of sulfide, thiosulfate, and, possibly, elemental sulfur by Sox and rDsr are shuttled into the electron transport chain, energy obtained from the oxidation of sulfite is immediately available in the form of ATP.

Bioenergetics

The *S. velum* symbiont is thought to harvest energy from reduced sulfur oxidation with oxygen. Interestingly, its genome also encodes other respiratory pathways suggestive of diverse metabolic strategies. Based on gene content, the symbiont may

utilize multiple electron donors such as hydrogen, pyruvate, malate, succinate, and formate, and use alternative electron acceptors such as nitrate and dimethyl sulfoxide (DMSO). Furthermore, unlike any chemosynthetic symbiont studied to date, the *S. velum* symbiont contains genes that may allow it to preferentially establish H^+ and Na^+ electrochemical membrane gradients during each step of respiration and to selectively utilize them for ATP synthesis, solute transport, and pH control. This high degree of respiratory flexibility encoded in the *S. velum* symbiont genome suggests that these bacteria are adapted to a highly variable environment.

Rnf complexes: The versatility of the respiratory electron transport chain in the *S. velum* symbiont is indicated by the presence of genes supporting electron donors like ferrodoxins, which have a redox potential as negative as -500mV [51], compared, for example, to -400 mV of $S_2O_3^{2-}$ and -270mV of H_2S. The reversible oxidation of ferrodoxins coupled to reduction of NAD^+ in the *S. velum* symbiont may be catalyzed by the H^+ or/and Na^+-motive Rnf complexes (Figure 1) encoded in the genome by two complete *rnfABCDGE* (*rnf1*) and *rnfBCDGEA* (*rnf2*) operons. The organization of genes in the operons is conserved with other bacteria, suggesting that these clusters did not arise from duplication. Previously, only *Axotobacter vinelandii* and *Desulfobacterium autotrophicum* were known to harbor two *rnf* operons [51]. Based on the presence of genes for pyruvate:ferredoxin oxidoreductase located between *rnfB2* and *rnfC2*, pyruvate may serve as an electron donor for at least one of the Rnf complexes. In general, *rnf* genes are distributed mainly among obligate and facultative anaerobes, including many pathogens that colonize oxygen-limited host tissues [51]. Together with the fact that ferrodoxins play a key role in anaerobic metabolism [52], this suggests that the *S. velum* symbiont, as well as other sequenced chemosynthetic symbionts, which all contain *rnf* genes, may be capable of facultative anaerobiosis.

***Hydrogenases*:** Hydrogen is another highly electron negative reductant (-420 mV)

that the *S. velum* symbiont may harness for the reduction of quinone and NAD^+ cellular pool (Figure 1). Hydrogen oxidation is suggested by the presence in the symbiont genome of *hup* and *hox2* operons encoding an uptake and a bidirectional hydrogenase, respectively. The two subunits of the symbiont [Ni-Fe]-uptake hydrogenase, HupSL, are most similar in amino acid sequence to HupS and HupL proteins from the symbionts of the tubeworms, *R. pachyptila* and the *T. jerichona*, (73% and 78% identity for the S and L subunits respectively), the sulfur bacterium, *Thiocapsa roseopersicina,* (68 and 74%), and the symbionts of the scaly-foot snail, *Crysomallon squamiferum*, (50 and 53%). In *T. roseopersicina,* HupSL has been experimentally demonstrated to reduce quinones of the respiratory chain with H_2 [53, 54]. Unlike all the other γ-proteobacteria containing HupSL, the *hup* operon in the *S. velum* symbiont does not encode the di-heme cytochrome *b*, which is necessary to link H_2 oxidation to quinone reduction in the cellular membrane [55]. However, a [Ni-Fe] hydrogenase cytochrome *b* homolog was found on a different genomic scaffold. Though this discordant gene organization is unlike that in other H_2 oxidizers, it is possible that the identified cytochrome *b* may act in tandem with HupSL to enable H_2 oxidation.

Apart from potentially reducing the respiratory quinone pool with H_2, the symbiont, by means of a bidirectional hydrogenase, may produce H_2 by oxidizing NAD^+. The *S. velum* symbiont Hox2FUYH enzyme complex is most similar in amino acid sequence (63-66%) to the recently-characterized NAD^+-reducing [Ni-Fe]-hydrogenase from *T. roseopersicina*, which can operate in reverse, generating H_2 when the high reduction state of the dinucleotide pool is growth-limiting [56]. As H_2 concentrations available to the *S.* velum symbiont have not been measured, it is unknown whether H_2 oxidation contributes to primary production to the degree that has been recently demonstrated in hydrothermal vent symbiosis [57].

Primary ion pumps: NADH (-320 mV), potentially derived from oxidation of H_2 or

heterotrophic metabolism (see TCA Cycle) in the *S. velum* symbiont, could be converted into an electrochemical gradient by two NADH:quinone oxidoreductases. The genome of the symbiont encodes the conventional H^+-translocating quinone-reducing NADH dehydrogenase (NdhABCDEFGHIJKLMN), a homolog of the mitochondrial Complex I, as well as an alternative Na^+-translocating NADH dehydrogenase (NqrABCDEF) (Figure 1). While Complex I is ubiquitous in bacteria, Nqr is found mainly in pathogenic and marine species [58]. Among symbiotic bacteria, *nqr* genes have so far been described only in *Buchnera* spp., an obligate endosymbiont of aphids [59]. The *S. velum* symbiont may be able to switch between Complex I and Nqr, preferentially generating either H^+ or Na^+ electrochemical gradients. Thus, depending on cellular requirements, the symbionts may synthesize ATP (see ATP Synthases) and regulate pH (see Ion Gradient Driven Transporters) independently from each other.

Quinone reductases: Apart from the electron donors such as sulfur and NADH, the *S. velum* symbiont may be able to directly reduce its quinone pool with a number of other substrates. This is suggested by the presence of genes encoding malate:quinone oxidoreductase (Mqo), succinate dehydrogenase (ShdCDAB), homologous to Complex II in mitochondria, and formate dehydrogenase-O (FdoGHI) (Figure 1). This is the first report of FdoGHI in a chemosynthetic symbiont genome. In *E. coli* this enzyme, which is common to facultative anaerobes [60], is used in formate-dependent oxygen respiration, allowing the bacteria to rapidly adapt to shifts from aerobiosis to anaerobiosis [61]. The presence of FdoGHI is additional evidence that the *S. velum* symbiont may be capable of facultative anaerobiosis (see Rnf Complexes).

The genome-encoded quinol-cytochrome-*c* oxidoreductase (bc_1, Complex III homologue) potentially links oxidation of quinols to the generation of a proton membrane gradient and the reduction of terminal electron acceptors (Figure 1), discussed next.

Terminal oxygen reductases: Similar to most aerobic and microaerophilic

bacteria, the genome of the *S. velum* symbiont encodes three types of H^+-motive terminal oxygen reductases (Figure 1), which suggest a capacity to respire O_2 over a wide range of concentrations. The genome contains a *ccoNOQP* operon encoding a cbb_3 cytochrome oxidase, which is known to function at nanomolar O_2 concentrations in the nitrogen-fixing plant symbionts, *Bradyrhizobium japonicum* [62], and in the microaerophilic human pathogens, *Campylobacter jejuni, Helicobacter pylori*, and *Neisseria meningitidis* [63]. The genome also encodes a aa_3 cytochrome oxidase (CoxAB), which is thought to function primarily under atmospheric oxygen concentrations [64] and is the only terminal oxidase in the symbionts of the bivalves *C. magnifica* [22] and *C. okutanii* [20]. The third terminal oxidase identified in the symbiont genome is a *cydAB*-encoded quinol oxidase, which is thought to oxidize quinols instead of cytochromes. CydAB may operate when an excess of reductants, potentially coming from the host, limits metabolic turnover and redox balance needs to be achieved [65]. The observed diversity of terminal oxygen reductases indicates that the supply of oxygen to the symbionts fluctuates over time or between free-living and symbiotic stages, necessitating adjustments in respiratory metabolism.

***Alternative terminal reductases*:** When oxygen is limited or unavailable, potentially either through competition for oxygen with the host or if the symbionts find themselves in the anoxic sediment that surrounds the burrow, the *S. velum* symbiont may be capable of using terminal electron acceptors other than oxygen. Although it is unknown whether the symbiont-containing gill bacteriocytes ever become anaerobic, the presence of genes for periplasmic NO_3^- reductase (*napFDAGHBC*) suggests that symbiont energy generation may involve electron transfer to nitrate, which is available in the porewater surrounding *S. velum* at concentrations of ~1-10 µM ([66], in preparation). The structure of the symbiont *napFDAGHBC* operon is consistent with that of enteric bacteria that are thought to use Nap for effectively scavenging nitrate during anaerobic

364 growth under nitrate-limited conditions (5 µM) [67]. The symbiont genome also encodes
365 a DMSO reductase (*dmsABC*), suggesting the ability to respire dimethyl sulfoxide
366 (DMSO), a breakdown product of dimethylsulfoniopropionate (DMSP) produced, for
367 example, by marine algae. DMSO is available at nanomolar concentrations in the
368 coastal eutrophic environments inhabited by *S. velum* [68], and Dms genes are common
369 to many marine sediment-dwelling bacteria, e.g., *Beggiatoa* and *Shewanella* [69, 70].
370 ***ATP synthases***: Based on the genome data, both H^+ and Na^+ membrane
371 gradients, established along the course of the electron transport chain during respiration,
372 may drive ATP synthesis in the *S. velum* symbiont via H^+- or Na^+-dependent ATP
373 synthases (Figure 1). The H^+-specificity of the F_0F_1-type ATP synthase is suggested by
374 the presence of two characteristic transmembrane helixes within the *c* subunit. In
375 contrast, an A_0A_1-type ATP synthase detected in the genome contains the characteristic
376 Na^+-binding PXXXQ motif I and ES motif II in the rotor subunit K. While proton-
377 translocating ATP synthases are predominant in bacteria, Na^+-coupled ATP synthesis
378 driven by respiration has recently been recognized in some marine and pathogenic
379 species [71, 72]. To our knowledge, this is the first instance of a Na^+-translocating ATP
380 synthase in a chemosynthetic symbiont.
381 ***Ion gradient driven transporters***: Cellular roles of the H^+ and Na^+ gradients in the
382 *S. velum* symbiont appear to extend beyond ATP synthesis. Besides ATP synthases, the
383 genome encodes diverse Na^+:substrate symporters and numerous Na^+:H^+ antiporters,
384 including the multi-subunit MrpEFGBBCDD complex (Figure 1). These transporters,
385 together with ATP synthases and respiratory ion pumps, may establish and consume
386 simultaneous transmembrane gradients of proton and sodium ions in the symbionts [71].
387 These parallel cycles of H^+ and Na^+ would allow the *S. velum* symbiont to synthesize
388 ATP and maintain pH homeostasis via two separate mechanisms.
389 **2. Carbon metabolism**

Autotrophic carbon fixation, fueled chiefly by sulfur oxidation, is the principal process in the *S. velum* symbiont, supplying both the symbionts and the host with organic carbon [14]. While previous studies focused primarily on RubisCO [10, 73], the key enzyme of the Calvin cycle for CO_2 fixation and the most highly expressed gene in the symbiont [40], our current analysis identified genes that encode catalytic components required for CO_2 fixation and storage, including the pyrophosphate-dependent phosphofructokinase, which has been hypothesized to command a more energy efficient variant of the cycle [22, 74-76]. Furthermore, the genome of the *S. velum* symbiont contains the gene for α-ketoglutarate dehydrogenase – the key enzyme of the tricarboxylic acid cycle (TCA), suggesting that the symbiont can respire organic carbon and may not be obligately autotrophic.

Autotrophy: The genome of the *S. velum* symbiont encodes a modified Calvin cycle that appears unique to chemosynthetic symbionts and some methanotrophic bacteria in that it lacks fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase. This alternative carbon fixation pathway may rely instead on a reversible pyrophosphate-dependent phosphofructokinase (PPi-PFK) (Figure 1). The ability of this enzyme to dephosphorylate fructose 1,6-phosphate and sedoheptulose 1,7-phosphate *in vitro* has been previously demonstrated for the PPi-PFK from *Methylococcus capsulatus* [74], which shares 73% sequence identity with the *S. velum* symbiont homologue. Compared to the classical Calvin cycle [77], the use of PPi-PFK can offer energy savings up to 9.25% [76]. While fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase enzymes release phosphate ions that cannot be used for energy gain, PPi-PFK produces energy-rich pyrophosphate. In *M. capsulatus* [74] and in the chemosynthetic symbionts of *R. pachyptila* [75] and the oligochete, *O. algarvensis* [76] it was proposed that the pyrophosphate could be converted into an electrochemical gradient by the membrane-bound proton-pumping pyrophosphatase

(H^+-PPase) co-encoded with PPi-PFK. Judging from the similar gene content, this mechanism may also be at work in the symbionts of the vent clams *C. magnifica* [22] and *C. okutanii* [20]. The *S. velum* symbiont genome also encodes a soluble pyrophosphatase (PPase) immediately upstream of the PPi-PFK gene. The PPase cannot conserve energy in a proton gradient, but may instead function to regulate the PPi-PFK, which may also participate in glycolysis as a kinase. This implies that in the *S. velum* symbiont the ability to control the directionality of carbon flux may be more critical than the energy saving aspect of this Calvin cycle variant.

Carbon Flux: Carbon fixed by the *S. velum* symbiont may be stored as polyglucose or fed into catabolic and anabolic reactions (Figure 1). The overall direction of the metabolic carbon flux in the symbionts can be controlled by at least two putative mechanisms. First, the reversible PPi-PFK, participating in the Calvin cycle as discussed above, may also phosphorylate fructose 6-bisphosphate during glycolysis. PPi-PFK appears to be the only enzyme encoded in the genome that could catalyze both the forward and the reverse reactions. The directionality of the catalysis may depend on the concentration of pyrophosphate in the cytoplasm [78], since this PPi-PFK is nonallosteric [74]. Second, the two glyceraldehyde 3-phosphate dehydrogenases, GapA and GapB, may be specific to glycolysis and the Calvin cycle/gluconeogenesis, respectively, by analogy to the homologous enzymes in *Staphylococcus aureus* [79]. In the symbiont genome, *gapB* is adjacent to the gene for transketolase, an enzyme in the Calvin cycle, further suggesting that these two Gap proteins may play a role in regulating the direction of the carbon flux either in the direction of glycolysis or Calvin cycle and gluconeogenesis. The symbionts of *C. magnifica, C. okutanii*, *R. pachyptila, T. jerichona*, and the scaly snail possess just a single *gap* gene, which has a much higher amino acid sequence identity to *gapB* than to *gapA* from the *S. velum* symbiont. The above evidence suggests that among the chemosynthetic symbionts, the symbiont of *S. velum*

is unique in placing an emphasis on more rigorously controlling the direction of the carbon flux.

Heterotrophy: The *S. velum* symbiont is the third chemosynthetic symbiont, along with the γ3-symbiont of *O. algarvensis* [34] and the intracellular γ-proteobacterial symbionts of the scaly-foot snail [33], known to encode all of the enzymes required for the complete TCA cycle, and, therefore, could oxidize organic carbon for energy (Figure 1). All of the other sequenced chemosynthetic symbionts lack genes for α-ketoglutarate dehydrogenase and citrate synthase, which suggests their obligate autotrophy [80].

Furthermore, genes for the glyoxylate bypass of the TCA cycle, encoding isocitrate lyase and malate synthase, were also found in the genome of the *S. velum* symbiont (Figure 1). These enzymes could allow the symbionts to grow on various carbon sources, including acetate and other two-carbon compounds, [81] or rapidly replenish intermediates of biosynthetic reactions. The presence of a glyoxylate bypass and the TCA cycle suggest that the symbiont may be a facultative mixo- or heterotroph. The adaptive role of having both heterotrophic pathways, however, is unclear, and may relate either to the intracellular conditions specific to this particular symbiosis or to the yet unconfirmed free-living existence of the symbionts in the environment.

3. Nitrogen metabolism

Ammonia, abundant in the sediment where *S. velum* burrows, is the main form of nitrogen assimilated by the symbiosis [82]. It has been suggested that the symbionts incorporate ammonia into biomass, which is then transferred to the host [66 in preparation], a process which has been described for the chemosynthetic symbionts of the hydrothermal vent tubeworm *Ridgeia piscesae* [83]. The presence of assimilatory nitrogen pathways in the *S. velum* symbiont genome corroborate this hypothesis.

Nitrogen assimilation**:** Extracellular ammonia may be imported by the symbionts via specific AmtB transporters and incorporated into glutamate and glutamine, which

serve as amino group donors for the other nitrogen-containing compounds in the cell (Figure 1). *S. velum* comes in contact with 20-100 µM concentration of ammonia in their coastal environment [66 in preparation]. Thus, it is not surprising that, unlike the chemosynthetic symbionts found at nitrate-rich (40 µM) hydrothermal vents [84, 85], the *S. velum* symbiont lacks *nar* genes for nitrate reductases capable of assimilatory nitrate reduction [32, 86-88]. Assimilation of ammonia has been previously demonstrated in the gills of *S. velum*, but was initially ascribed to the activity of the host glutamine synthetase (GS) [87]. The present analysis identified *glnA*, the gene that encodes GS, in the genome of the symbionts. A preliminary transcriptional study showed *glnA* to be one of the fifty most highly transcribed genes in the symbiont [40]. The biosynthetic pathways reconstructed on the basis of gene content suggest that the symbiont has the ability to make all of the 20 proteinogenic amino acids. The amino acid prototrophy of the symbionts is in keeping with their proposed role in providing most, if not all, of the host's nutrition [14, 15].

Urea metabolism: Host urea may serve as an additional source of assimilatory nitrogen for the *S. velum* symbionts. The identified *ureHABCEFG* operon encodes a cytoplasmic urease UreABC, which can hydrolyze urea, releasing ammonia that may be re-utilized by symbionts. Urea can enter the bacterial cell by passive diffusion [89], but under nitrogen starvation the symbionts may be able to take it up more rapidly via an ABC-transporter UrtABCDE, encoded directly upstream of the *ure* genes. Among chemosynthetic symbionts, urease genes have been previously described only in the γ-symbionts from the marine oligochaete worm *O. algarvensis* [34, 76], which, like *S. velum*, lives in coastal sediments. The sequenced chemosynthetic symbionts from hydrothermal vents lack urease genes, even though some of their host organisms, for instance *R. pachyptila* [90], are known to produce urea. This discrepancy may be accounted for by the fact that in coastal sediments urea is also present outside the host

494 in the pore water [66 in preparation].

495 ***Taurine synthesis***: The *S. velum* symbiont may also provide its host with
496 nitrogenous osmoregulants, such as the non-proteinogenic amino acid taurine [91]. In
497 the host tissues, taurine accounts for up to 70% of the total free amino acids and shows
498 an isotopic composition ($\delta^{13}C$, $\delta^{14}N$, $\delta^{34}S$) suggestive of symbiont origin [92]. Synthesis
499 of taurine may be accomplished by the two homologs of the reversible taurine
500 dioxygenase (TauD) encoded in the symbiont genome. Taurine could be actively
501 secreted to the host by the TauABCD ABC transporter, the genes for which were found
502 to contain a conserved binding domain for sulfonate, characteristic of the taurine
503 molecule. Since taurine synthesis requires sulfite [93], one of the final intermediates in
504 sulfur oxidation, this pathway could serve to dispose of SO_3^{2-}, and, thus, to drive forward
505 sulfur oxidation in the *S. velum* symbiont, benefiting both the host and the symbionts.
506

507 **Membrane-associated functions**

508 The diversity of membrane-associated functions encoded in the genome of the *S.*
509 *velum* symbiont suggests that the symbionts are fully autonomous of their host. Other
510 bacteria, which, like the symbionts, are thought to be obligately intracellular [17], have
511 lost genes required for the production of a cellular envelope, transport of solutes across
512 plasma membrane, sensing of the extracellular environment, as well as motility. These
513 bacteria instead rely on their hosts to perform these functions or no longer require them.

514 **1. Production of cellular envelope**

515 The *S. velum* symbiont appears capable of synthesizing and assembling a
516 cytoplasmic membrane, peptidoglycan layer (PG), and outer membrane populated by
517 lipopolysaccharides (LPS), which constitute a cellular envelope. While these abilities are
518 typical of the free-living γ-proteobacteria, two aspects in particular stand out in the
519 context of the symbiotic life-style. First, given the identified genes for the biosynthesis of

fatty-acids, the symbionts may build components of their plasma membrane mostly from *cis*-vaccenic acid (18 : lω7) (Figure 1). According to a previous analysis of lipid composition in *S. velum* [94], this unsaturated fatty acid and its derivatives are the main constituents of cellular membranes in the symbionts and the host alike. Furthermore, the isotopic signature of the host's lipids indicates that they are bacterial in origin [94]. Second, the identified genes for the synthesis of lipopolysaccharides (Figure 1) suggest that the symbiont may be able to assemble the LPS structures that are known to be sufficient for growth of *E. coli* [95]. Most intracellular symbionts that live within a host derived membrane, like the *S. velum* symbiont [10], lack LPS biosynthetic genes and are unable to replicate on their own [96]. However, the symbionts which have the genes to synthesize LPS tend to either live directly in the cytoplasm [96] and, have to make their own cellular envelope, or, like the symbionts of *R. pachyptila* [97], exist extracellularly for part of their life. Therefore, the symbiont of *S. velum* may not only be able to make a fully functional cellular envelope and supply some of its components to its host, but may also be capable of living outside the bacteriocytes.

2. Membrane transport

Transporters: The number of transporters encoded in the genome of the *S. velum* symbiont exceeds what has been found in other intracellular bacteria (Table 2). The diversity of genes for solute transport (Figure 1) suggests that the symbionts have an extensive chemical communication with their environment. The *S. velum* symbiont may use these transporters to import metabolic substrates and enzyme cofactors and export products of its biosynthesis to sustain the physiology of the host. It is known that fixed organic carbon is transferred from the symbiont to the host within minutes [98], which suggests a transport mechanism, since direct digestion of symbionts by host cells would likely take hours to days [99]. Such transport could be accomplished by exporters of amino acids (EamA), carboxylates (CitT), and fatty acids (FadLD), all of which are

encoded in the genome. Moreover, some of the importers found in the genome may also act as exporters, depending on the cellular environment [100]. Thus, the *S. velum* symbiont maintains a repertoire of transporters that, on the one hand, may negotiate diverse chemical exchanges with the environment and, on the other, allow it to provide nutrients to the host without being digested.

Multi-drug efflux pumps: The *S. velum* genome contains at least five sets of genes encoding multi-drug efflux pumps (AcrAB-TolC), suggesting the ability to expel host-derived antimicrobial agents. A comparable genetic capacity for the AcrAB-TolC efflux system has been found in bacteria, such as the plant symbiont *Rhizobium leguminosarum,* that have a free-living stage, but not in bacteria that are obligately intracellular (Table 2). The plant host of *R. leguminosarum* manipulates the cellular fate of its symbionts using antimicrobial-like peptide factors [101]. As a result, *R. leguminosarum* undergoes cell elongation and genome replication but looses its ability to divide. Only a small number of *R. leguminosarum* cells remain vegetative [102]. A very similar morphological differentiation of the symbionts has been observed in *S. velum* [103]. Assuming the bivalve host also uses peptide factors to control its symbionts, the *S. velum* symbiont may rely on the efflux pumps to maintain a small, undifferentiated population in the bacteriocytes for transmission to future host generations.

3. Sensory mechanisms and motility

The *S. velum* symbiont appears well equipped to sense extracellular chemical changes, consistent with its inferred ability to maintain a complex chemical exchange with the environment. Over forty transmembrane chemoreceptors are encoded in the genome of the symbionts. Almost half of them have one or more conserved PAS domains and therefore may play a role in sensing oxygen levels and redox potentials. To relay sensory information, the majority of the receptors contain either a diguanylate cyclase (GGDEF) or a histidine kinase (HisKA) signaling domain. Movement and surface

attachment using type IV pili, known as twitching motility, are the processes that may be regulated by chemosensory signal transduction in the *S. velum* symbiont (Figure 1). For example, in the genome of the symbiont, a chimeric gene containing PAS, GGDEF, and cyclic-diguanylate receptor (EAL) domains is co-located with *pilEY1XWVT* genes required to assemble a functional pilus. Furthermore, the symbiont genome contains *pilGIJ-cheAW* genes, which encode a transmembrane chemotaxis sensor protein, HisKA, and a DNA-binding response regulator, and are known to control twitching motility in other bacteria [104]. The symbiont may use the contractile pili to direct its movement in the environment with regard to chemicals gradients, and, potentially, also rely on the same mechanism to find and colonize new hosts.

Mobile genetic elements

The *S. velum* symbiont genome contains two major types of mobile elements, integrative and conjugative elements (ICEs) and insertion sequences (IS). The genome contains 25 insertions from 12 different ICE families (Table 3) as well as 53 copies of four different IS elements (Table 4). In total, these elements comprise 2.6% of the genome. No gene interruptions were associated with these elements. This large number and diversity of mobile elements suggest that these bacteria may come into contact with other bacterial lineages more often than expected for most vertically transmitted intracellular symbionts. Indeed, the abundance of mobile genetic elements in bacterial genomes has been shown to correlate with ecological niche. While there is considerable overlap between the amounts of mobile elements hosted by free-living and facultative intracellular bacteria, obligate intracellular bacteria that undergo faithful vertical transmission consistently have few or no mobile elements [105].

Two hypothesized life and evolutionary history scenarios may explain the observed mobile element content in the *S. velum* symbiont. One explanation is a relatively recent

shift in intracellularity, resulting in an expansion of mobile elements [106, 107]. Alternatively, the symbionts may undergo regular or occasional horizontal transmissions *between* hosts and at that time encounter opportunities for recombination between strains. For example, sporadic episodes of horizontal transmission in the primarily maternally transmitted insect symbiont *Wolbachia* have resulted in the acquisition and maintenance of novel mobile elements [108, 109]. In fact, horizontal transmission or host-switching has likely occurred in the history of symbionts of bivalves [110] including members of the genus *Solemya*, as 16S rRNA phylogenetic analyses show that these symbionts do not comprise a monophyletic clade [5, 11]. Additionally, many of the genes in the *S. velum* symbiont genome are most closely related to disparate bacterial taxa (Figure 3), suggesting that horizontal gene transfer may have occurred in the past. These preliminary lines of evidence support the hypothesis that horizontal symbiont transmission has occurred. However, more information is needed about the distribution and relationships of mobile elements among intra-host and inter-host *S. velum* symbiont populations before these hypotheses can be differentiated.

Conclusions

Many of the features commonly encoded in the genomes of chemosynthetic symbionts were observed in the genome of the *S. velum* symbiont alongside an array of genes unique to this bacterium. Potential adaptations to the symbiotic lifestyle, such as a more energy-efficient version of the Calvin cycle, were shared with the other sequenced chemosynthetic symbionts. The genes that set the *S. velum* symbiont apart from the others were those that encoded the TCA and the glyoxylate cycles, DMSO and urea reductases, as well as a highly branched electron transport chain. The presence of these functions may relate to the fact that *S. velum* lives in eutrophic sediment, unlike the oligotrophic environments inhabited by other chemosynthetic symbioses, e.g., *R.*

pachyptila, C. magnifica, and *O. algarvensis*.

The *S. velum* symbiont has long been considered to be vertically transmitted [17], but our genomic analyses of both its GC composition and gene content are inconsistent with predictions based on other vertically transmitted obligately-intracellular bacteria. The *Solemya velum* symbiont's genetic repertoire is replete with genes for chemosynthesis, heterotrophy, bioenergetics, nitrogen metabolism, cell maintenance, motility, communication, and exchange with the environment. With regard to genome size, GC content, and functional gene content, the genome is more similar to those of free-living sulfur-oxidizing bacteria (Table 1). Furthermore, the genome contains mobile elements that are comparable in numbers reported for horizontally-transmitted obligately-intracellular bacteria. These divergent lines of evidence suggest that the evolutionary life history of the *S. velum* symbiont may be more complicated than previously hypothesized. This could include, but may not be limited to, such scenarios as an opportunistic generalist lifestyle, a facultative symbiosis with a mixed transmission mode, or a very recent obligate association for this clade of bacteria with the host potentially on a path to a new type of a cellular organelle.

Materials and methods

Specimen collection and DNA preparation: *S. velum* individuals were collected by the staff of the Marine Resource Center of the Marine Biological Laboratory (MBL), Woods Hole, MA from reducing sediment of shallow eelgrass beds near Naushon Island, Woods Hole, MA in 2006, 2007, and 2012. The collection was performed in accordance with state collecting permit issued by the Division of Marine Fisheries and in compliance with all local, regional and federal regulations, including the Convention on Biological Diversity and the Convention on the Trade in Endangered Species of Wild Fauna and Flora. The excised gills were macerated in the laboratory using a dounce homogenizer

650 in 5 ml of 0.2 µm filtered seawater (FSW) per bivalve. Homogenates were passed
651 through 100 µm and 5 µm nylon filters (Small Parts Inc. #CMN-0105-A and CMN-0005-
652 A) and centrifuged at 5,000x g for 5 minutes at 4ºC. The pellet was resuspended in
653 FSW, pelleted, and resuspended in 1x TAE buffer. 50 g molten 2% agarose (SeaKem®
654 #50152) in 1x TAE was added to make plugs for genomic DNA extraction. The hardened
655 plugs were treated with DNAse I (0.25U/50 µl) at 37ºC for 10 minutes and then
656 equilibrated in TE buffer for 30 minutes at room temperature. Agarose plugs were further
657 processed using CHEF Mammalian Genomic DNA Plug Kit from Bio-Rad Laboratories
658 (#170-3591) according to the manufacturer's instructions. The protocol for pulse field gel
659 electrophoresis (PFGE) and isolation of the bacterial chromosomes from the agarose
660 plugs was adapted from Gil [111].
661 ***Genome sequencing and assembly***: Genomic bacterial DNA was sequenced at
662 the Institute for Genomic Research (TIGR), the Joint Genome Institute (JGI), and the
663 University of California, Davis, using a diversity of sequencing technologies. Two Sanger
664 libraries of 3-4 Kb and 10-12 Kb insert sizes were constructed as previously described
665 [112]. Sequencing of these Sanger libraries resulted in 110,187 reads with N50 of 969
666 bp and the average coverage depth of 8x. Subsequently, using Roche 454 technology,
667 387,143 sequencing reads with the N50 of 207 bp and the average coverage depth of
668 13x were obtained. Then, 25,635,107 Illumina sequencing reads were generated. The
669 Illumina sequences were 35 bp long and had the average coverage depth of 150x.
670 These Sanger, Roche 454, and Illumina sequences were assembled using the Paracel
671 Genome Assembler (Paracel Inc., Pasadena, CA) into 68 contigs. Next, symbiont DNA
672 was sequenced using Pacific Biosciences (PacBio) technology, resulting in 150,000
673 reads with N50 of 4,966 bp and 9x coverage depth. The insertion and deletion (indels)
674 errors, typical of the PacBio data [113], were reduced from 4% to 0.2% with Illumina
675 paired-end sequences (500x coverage) using PacBioToCA program [114] available as a

676 part of SMRT Analysis software package version 1.4 distributed by the Pacific
677 Biosciences [115]. The error correction step also removed any PacBio sequences of the
678 host origin, which, given the abundance of the symbionts in the gill tissue, had Illumina
679 coverage below 5x. The Illumina data used for the error correction were generated as
680 part of a different study and came from a specimen obtained at a different location (Point
681 Judith, RI) than the rest of the genomic data. Due to the extent of the intra-species
682 genomic sequence variation across geographic localities (Russell et al., in preparation),
683 these Illumina data could not be used to supplement the genome assembly but were
684 sufficient to correct the majority of sequencing indel errors in the PacBio reads. The
685 error-corrected 54,684 PacBio sequences with N50 of 1,409 bp were used to connect
686 the previous 68 genomic contigs into 30 larger scaffolds using the Automated Hybrid
687 Assembly (AHA) module of SMRT Analysis. The resulting 7 gaps within the scaffolds,
688 2,272 bp in total, were then filled in with the PacBio error-corrected sequences using the
689 PBJelly software tool [67], reducing the number of gaps to 4 and the total gap length to
690 100 bp. After discarding 20 of the smallest low coverage (2-9x) scaffolds that contained
691 mostly eukaryotic genes (>65%), identified as described below, only 10 scaffolds were
692 retained as a part of the symbiont genome.
693 ***Sequence analysis***: Open reading frames (ORFs) on *S. velum* symbiont scaffolds
694 were predicted using Glimmer [116], Prodigal [117], and GeneMarkS [118]. The software
695 parameters used to perform these analyses are listed in Table S4 in Additional file 4.
696 Once identified, the ORFs were translated into protein-coding sequences and queried
697 against the UniProt Reference Clusters (UniRef90) (20 November 2013) [119], National
698 Center for Biotechnology Information non-redundant (NCBI-nr) (4 November 2013)
699 [120], and M5 non-redundant (M5-nr) (27 January 2014) [121] databases for functional
700 annotation using BLASTP (e-value cutoff 0.001) [38]. UniRef90 gene entries sharing the
701 highest percent identity with the query and NCBI-nr and M5-nr entries with the highest

bit score match to the query were retained for annotation. Genes predicted by two or more methods (redundant) were considered the same and collapsed into a single entry if they shared the same start and stop position, orientation, and similar functional annotations. Non-redundant entries (i.e., gene predictions unique to a given software) were also retained. Finally, the above predictions and annotations were reconciled with the genes predicted and annotated through the Integrated Microbial Genomes Expert Review (IMG-ER) pipeline [122]. Selected origins of replication were verified by Ori-Finder [67]. The genes in the genome was assigned taxa in the NCBI taxonomy based on the BLASTN [38] searches (-best_hit_overhang 0.25, -best_hit_score_edge 0.05, -evalue 0.0001) against the NCBI-nr database (8 July 2014) computed with MEGAN 5.4.3 (maximum number of matches per read 100; LCA parameters: minimal support 5, minimal score 35, top percent 10) [39]. Selected promoters were identified with BPROM [123]. Signal peptides and transmembrane domains were predicted using SignalP 3.0 Server and TMHMM, respectively [124]. The Genomic Utility for Automated Comparison (GUAC) Python script (Additional file 5) was developed to inform comparative analyses of gene content across multiple genomes, in particular genes involved in sulfur-oxidation (Figure 4). The GUAC software first identified those target genes in the genomes of interest that were annotated with unambiguous gene symbols (e.g. *soxA*). Next, using amino acid sequences of these genes as queries, BLASTP searched for homologous sequences in the remaining target genomes (default cut-off values: bit score 50, identity 30%, alignment length over the source sequence 40%). These sequences were aligned using ClustalW [125]. The alignments were used to manually verify the results (e.g., based on known conserved domains, etc.). Mobile genetic elements were detected by type. Insertion sequences were found using OASIS [126]. Integrative conjugative elements and plasmid as well as phage sequences were identified by BLASTN [38] searches against the ICEberg [127] and ACLAME [128] databases, respectively (cut-off

values: 250 nucleotides alignment length and 90% identity). To determine whether mobile genetic elements interrupted open reading frames, the nucleotide regions before and after each element were concatenated and aligned to the NCBI-nr sequences using BLASTN.

Availability of supporting data

This genome project has been deposited at DDBJ/EMBL/GenBank under the accession JRAA00000000. The version described in this paper is version JRAA01000000.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

OD performed the DNA isolation and the final genome assembly, developed the Python GUAC script, carried out the sequence analysis and the manual annotation, and drafted the manuscript. SLR, WTL, KMF, LL, and GR participated in the sequence analysis, the manual annotation, and the drafting of the manuscript. FJS carried out the DNA isolation, coordinated and participated in the gene prediction and the automated annotation. RS performed the gene prediction and the automated annotation. ILGN carried out the DNA isolation and participated in the sequence analysis. TW and JAE coordinated and participated in the genome sequencing, the initial genome assembly, and the preliminary gene prediction and annotation. DW and JML performed the initial genome assembly, gene prediction, and annotation. CMC and JAE conceived of the

study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

Acknowledgments

This work was funded by grant 0412205 of the US National Science Foundation (NSF) and was made possible with the generous support of the U.S. Department of Energy Joint Genome Institute (JGI). The work conducted by JGI was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We would like to express special thanks to Grace Pai for creating Sanger sequencing libraries and Shannon Smith and Terry Utterback for coordinating sequencing at TIGR.

References

1. Sagan L: **On the origin of mitosing cells**. *J Theor Biol* 1967, **14**:225–274.

2. Gonzalez A, Clemente JC, Shade A, Metcalf JL, Song S, Prithiviraj B, Palmer BE, Knight R: **Our microbial selves: what ecology can teach us.** *EMBO Rep* 2011, **12**:775–784.

3. Dilworth MJ, James EK, Sprent JI: *Nitrogen-Fixing Leguminous Symbioses*. Kluwer Academic Pub; 2008.

4. Clark EL, Karley AJ, Hubbard SF: **Insect endosymbionts: manipulators of insect herbivore trophic interactions?** *Protoplasma* 2010, **244**:25–51.

5. Cavanaugh CM, McKiness Z, Newton I, Stewart FJ: **Marine Chemosynthetic Symbioses**. In *The Prokaryotes - Prokaryotic Biology and Symbiotic Associations*. 3rd edition. Edited by Rosenberg E; 2013:579–607.

6. Toft C, Andersson SGE: **Evolutionary microbial genomics: insights into bacterial host adaptation.** *Nat Rev Genet* 2010, **11**:465–475.

7. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng J-F: **One Bacterial Cell, One Complete Genome**. *PLoS ONE* 2010, **5**.

8. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U: **Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges.** *ISME J* 2013, **7**:2287–2300.

9. Dubilier N, Bergin C, Lott C: **Symbiotic diversity in marine animals: the art of harnessing chemosynthesis**. *Nat Rev Micro* 2008, **6**:725–740.

10. Cavanaugh CM: **Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats**. *Nature* 1983, **302**:58–61.

11. Eisen JA, Smith SW, Cavanaugh CM: **Phylogenetic relationships of chemoautotrophic bacterial symbionts of**

***Solemya velum* Say (Mollusca: Bivalvia) determined by 16S rRNA gene sequence analysis.** *J Bacteriol* 1992, **174**:3416–3421.

12. Cavanaugh CM, Abbott M, Veenhuis M: **Immunochemical localization of ribulose-1, 5-bisphosphate carboxylase in the symbiont-containing gills of *Solemya velum* (Bivalvia: Mollusca)**. *P Natl Acad Sci Usa* 1988, **85**:7786–7789.

13. Scott KM, Cavanaugh CM: **CO_2 uptake and fixation by endosymbiotic chemoautotrophs from the bivalve *Solemya velum***. *Appl Environ Microb* 2007, **73**:1174–1179.

14. Conway N, Capuzzo J, Fry B: **The role of endosymbiotic bacteria in the nutrition of *Solemya velum*: evidence from a stable isotope analysis of endosymbionts and host**. *Limnology and Oceanography* 1989, **34**:249–255.

15. Krueger DM, Gallager S, Cavanaugh CM: **Suspension feeding on phytoplankton by *Solemya velum*, a symbiont-containing clam**. *Mar Ecol-Prog Ser* 1992, **86**:145–151.

16. Cary SC: **Vertical transmission of a chemoautotrophic symbiont in the protobranch bivalve, *Solemya reidi*.** *Mol Marine Biol Biotechnol* 1994, **3**:121–130.

17. Krueger DM, Gustafson RG, Cavanaugh CM: **Vertical transmission of chemoautotrophic symbionts in the bivalve *Solemya velum* (Bivalvia: Protobranchia)**. *Biol Bull* 1996, **190**:195–202.

18. Peek A, Vrijenhoek R, Gaut B: **Accelerated evolutionary rate in sulfur-oxidizing endosymbiotic bacteria associated with the mode of symbiont transmission**. *Molecular Biology and Evolution* 1998, **15**:1514.

19. Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC: **Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calyptogena magnifica*.** *Appl Environ Microb* 2003, **69**:2058–2064.

20. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, Matsuyama K, Takishita K, Kawato M, Uematsu K: **Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii***. *Curr Biol* 2007, **17**:881–886.

21. Kuwahara H, Takaki Y, Yoshida T, Shimamura S, Takishita K, Reimer JD, Kato C, Maruyama T: **Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea *Calyptogena clams***. *Extremophiles* 2008, **12**:365–374.

22. Newton I, Woyke T, Auchtung T, Dilly G, Dutton R, Fisher M, Fontanez K, Lau E, Stewart FJ, Richardson P: **The *Calyptogena magnifica* chemoautotrophic symbiont genome**. *Science* 2007, **315**:998–1000.

23. Newton I, Girguis PR, Cavanaugh CM: **Comparative genomics of vesicomyid clam (Bivalvia: Mollusca) chemosynthetic symbionts.** *BMC Genomics* 2008, **9**:585.

24. Peek A, Feldman R, Lutz R, Vrijenhoek R: **Cospeciation of chemoautotrophic bacteria and deep sea clams**. *P Natl Acad Sci Usa* 1998, **95**:9962.

25. Stewart FJ, Young CR, Cavanaugh CM: **Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis**. *Molecular Biology and Evolution* 2008, **25**:673–687.

26. Stewart FJ, Young C, Cavanaugh CM: **Evidence for homologous recombination in intracellular chemosynthetic clam symbionts**. *Molecular Biology and Evolution* 2009, **26**:1391–1404.

27. Stewart FJ, Baik AHY, Cavanaugh CM: **Genetic subdivision of chemosynthetic endosymbionts of *Solemya velum* along the Southern New England coast.** *Appl Environ Microb* 2009, **75**:6005–6007.

28. Krueger DM, Cavanaugh CM: **Phylogenetic diversity of bacterial symbionts of *Solemya* hosts based on comparative sequence analysis of 16S rRNA genes**. *Appl Environ Microb* 1997, **63**:91.

29. Moran NA: **Accelerated evolution and Muller's rachet in endosymbiotic bacteria.** *P Natl Acad Sci Usa* 1996, **93**:2873–2878.

30. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA: **Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: a streamlined genome overrun by mobile genetic elements**. *PLoS Biol* 2004, **2**:E69.

830 31. Robidart J, Bench S, Feldman R, Novoradovsky A, Podell S, Gaasterland T, Allen E, Felbeck H: **Metabolic versatility**
831 **of the *Riftia pachyptila* endosymbiont revealed through metagenomics**. *Environ Microbiol* 2008, **10**:727–737.

832 32. Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thürmer A, Albrecht D, Wollherr A, Kabisch J, Le Bris N, Lehmann
833 R, Daniel R, Liesegang H, Hecker M, Schweder T: **Physiological homogeneity among the endosymbionts of *Riftia***
834 ***pachyptila* and *Tevnia jerichonana* revealed by proteogenomics.** *ISME J* 2012, **6**:766–776.

835 33. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S-I, Watanabe T, Fujiyoshi S, Mino S, Sawabe T, Maeda
836 T, Makita H, Nemoto S, Nishimura S-I, Watanabe H, Watsuji T-O, Takai K: **Allying with armored snails: the complete**
837 **genome of gammaproteobacterial endosymbiont**. *ISME J* 2014, **8**:40–51.

838 34. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro
839 HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N: **Symbiosis insights**
840 **through metagenomic analysis of a microbial consortium**. *Nature* 2006, **443**:950–955.

841 35. Wu M, Eisen JA: **A simple, fast, and accurate method of phylogenomic inference**. *Genome Biol* 2008, **9**.

842 36. Murphy FV, Ramakrishnan V: **Structure of a purine-purine wobble base pair in the decoding center of the**
843 **ribosome**. *Nat Struct Mol Biol* 2004, **11**:11251–1252.

844 37. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631–637.

845 38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *J Mol Biol* 1990, **215**:403–
846 410.

847 39. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC: **Integrative analysis of environmental sequences**
848 **using MEGAN4.** *Genome Res* 2011, **21**:1552–1560.

849 40. Stewart FJ, Dmytrenko O, DeLong E: **Metatranscriptomic analysis of sulfur oxidation genes in the**
850 **endosymbiont of *Solemya velum***. *Frontiers in Microbiology* 2011, **2**:1–10.

851 41. Frigaard N-U, Dahl C: **Sulfur metabolism in phototrophic sulfur bacteria**. *Adv Microb Physiol* 2009, **54**:103–200.

852 42. Dahl C, Prange A: **Bacterial sulfur globules: occurrence, structure and metabolism**. *Inclusions in Prokaryotes*
853 2006.

854 43. Friedrich C, Bardischewsky F, Rother D, Quentmeier A, Fischer J: **Prokaryotic sulfur oxidation**. *Curr Opin Microbiol*
855 2005, **8**:253–259.

856 44. Fisher C, Childress J, ARP A, BROOKS J, DISTEL D, Favuzzi J, Macko S, Newton A, Powell M, Somero G, SOTO T:
857 **Physiology, morphology, and biochemical composition of *Riftia pachyptila* at Rose Garden in 1985**. *Deep-Sea Res*
858 1988, **35**:1745–1758.

859 45. Vetter RD: **Elemental sulfur in the gills of three species of clams containing chemoautotrophic symbiotic**
860 **bacteria: a possible inorganic energy storage compound**. *Mar Biol* 1985, **88**:33–42.

861 46. Childress JJ, Girguis PR: **The metabolic demands of endosymbiotic chemoautotrophic metabolism on host**
862 **physiological capacities**. *J Exp Biol* 2011, **214**:312–325.

863 47. Cort JR, Selan U, Schulte A, Grimm F, Kennedy MA, Dahl C: ***Allochromatium vinosum* DsrC: Solution-state NMR**
864 **structure, redox properties, and interaction with DsrEFH, a protein essential for purple sulfur bacterial sulfur**
865 **oxidation**. *J Mol Biol* 2008, **382**:692–707.

866 48. Oliveira TF, Vonrhein C, Matias PM, Venceslau SS, Pereira IAC, Archer M: **Purification, crystallization and**
867 **preliminary crystallographic analysis of a dissimilatory DsrAB sulfite reductase in complex with DsrC**. *J Struct*
868 *Biol* 2008, **164**:236–239.

869 49. Ghosh W, Dam B: **Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and**
870 **ecologically diverse bacteria and archaea**. *Fems Microbiol Rev* 2009, **33**:999–1043.

871 50. Chen C, Rabourdin B, Hammen C: **The effect of hydrogen sulfide on the metabolism of *Solemya velum* and**
872 **enzymes of sulfide oxidation in gill tissue**. *Comparative Biochemistry and Physiology Part B: Biochemistry and*
873 *Molecular Biology* 1987, **88**:949–952.

874 51. Biegel E, Schmidt S, González JM, Müller V: **Biochemistry, evolution and physiological function of the Rnf**
875 **complex, a novel ion-motive electron transport complex in prokaryotes**. *Cellular and molecular life sciences* 2011,
876 **68**:613–634.

877 52. Bruschi M, Guerlesquin F: **Structure, function and evolution of bacterial ferredoxins.** *Fems Microbiol Rev* 1988,
878 **4**:155–175.

879 53. Kovács KL, Kovács AT, Maróti G, Mészáros LS, Balogh J, Latinovics D, Fülöp A, Dávid R, Dorogházi E, Rákhely G:
880 **The hydrogenases of *Thiocapsa roseopersicina*.** *Biochem Soc Trans* 2005, **33**:61–63.

881 54. Burgdorf T, Lenz O, Buhrke T, van der Linden E, Jones A, Albracht S, Friedrich B: **[NiFe]-hydrogenases of**
882 ***Ralstonia eutropha* H16: Modular enzymes for oxygen-tolerant biological hydrogen oxidation**. *J Mol Microb Biotech*
883 2005, **10**:181–196.

884 55. Vignais PM, Billoud B: **Occurrence, classification, and biological function of hydrogenases: an overview**. *Chem*
885 *Rev* 2007, **107**:4206–4272.

886 56. Maroti J, Farkas A, Nagy IK, Maroti G, Kondorosi E, Rakhely G, Kovacs KL: **A second soluble hox-type nife**
887 **enzyme completes the hydrogenase set in *Thiocapsa roseopersicina* BBS**. *Appl Environ Microb* 2010, **76**:5113–
888 5123.

889 57. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V,
890 Pelletier E, Fink D, Borowski C, Bach W, Dubilier N: **Hydrogen is an energy source for hydrothermal vent symbioses**.
891 *Nature* 2011, **476**:176–180.

892 58. Bogachev AV, Verkhovsky MI: **Na^+-translocating NADH: Quinone oxidoreductase: Progress achieved and**
893 **prospects of investigations**. *Biochemistry (Moscow)* 2005, **70**:143–149.

894 59. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial**
895 **symbiont of aphids *Buchnera* sp. APS**. *Nature* 2000, **407**:81–86.

896 60. Pickering BS, Oresnik IJ: **Formate-dependent autotrophic growth in *Sinorhizobium meliloti***. *J Bacteriol* 2008,
897 **190**:6409.

898 61. Benoit S, Abaibou H, Mandrand-Berthelot M-A: **Topological analysis of the aerobic membrane-bound formate**
899 **dehydrogenase of *Escherichia coli***. *J Bacteriol* 1998, **180**:6625.

900 62. Preisig O, Zufferey R, Thony-Meyer L, Appleby C, Hennecke H: **A high-affinity cbb_3-type cytochrome oxidase**
901 **terminates the symbiosis- specific respiratory chain of *Bradyrhizobium japonicum***. *J Bacteriol* 1996, **178**:1532.

902 63. Pitcher RS, Watmough NJ: **The bacterial cytochrome cbb_3 oxidases**. *Biochimica et Biophysica Acta (BBA) -*
903 *Bioenergetics* 2004, **1655**:388–399.

904 64. Nunoura T, Sako Y, Wakagi T, Uchida A: **Regulation of the aerobic respiratory chain in the facultatively aerobic**
905 **and hyperthermophilic archaeon *Pyrobaculum oguniense***. *Microbiology (Reading, Engl)* 2003, **149**:673–688.

906 65. Otten MF, Stork DM, Reijnders WN, Westerhoff HV, Van Spanning RJ: **Regulation of expression of terminal**
907 **oxidases in *Paracoccus denitrificans***. *European Journal of Biochemistry* 2001, **268**:2486–2497.

908 66. Krueger DM, Roeselers G, Sigman D, Cavanaugh CM: **Nitrogen Nutrition in the Symbiosis Solemya Velum**.

909 67. Potter LC, Millington P, Griffiths L, Thomas GH, Cole JA: **Competition between *Escherichia coli* strains**
910 **expressing either a periplasmic or a membrane-bound nitrate reductase: does Nap confer a selective advantage**
911 **during nitrate-limited growth?** *Biochem J* 1999, **344 Pt 1**:77–84.

912 68. Zemmelink H, Houghton L, Sievert S, Frew N, Dacey J: **Gradients in dimethylsuffide, dimethylsulfoniopropionate,**
913 **dimethylsulfoxide, and bacteria near the sea surface**. *Mar Ecol-Prog Ser* 2005, **295**:33–42.

914 69. Mussmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jorgensen BB, Huntemann M, Gloeckner FO, Amann R,
915 Koopman WJH, Lasken RS, Janto B, Hogg J, Stoodley P, Boissy R, Ehrlich GD: **Insights into the genome of large**
916 **sulfur bacteria revealed by analysis of single filaments**. *PLoS Biol* 2007, **5**:1923–1937.

917 70. McCrindle SL, Kappler U, McEwan AG: **Microbial dimethylsulfoxide and trimethylamine-N-oxide respiration.** *Adv*
918 *Microb Physiol* 2005, **50**:147–198.

71. Häse CC, Fedorova ND, Galperin MY, Dibrov PA: **Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons**. *Microbiology and Molecular Biology Reviews* 2001, **65**:353–70, table of contents.

72. Mulkidjanian AY, Dibrov P, Galperin MY: **The past and present of sodium energetics: may the sodium-motive force be with you.** *Biochim Biophys Acta* 2008, **1777**:985–992.

73. Robinson J, Cavanaugh CM: **Expression of form I and form II Rubisco in chemoautotrophic symbioses: implications for the interpretation of stable carbon isotope values**. *Limnology and Oceanography* 1995, **40**:1496–1502.

74. Reshetnikov AS, Rozova ON, Khmelenina VN, Mustakhimov II, Beschastny AP, Murrell JC, Trotsenko YA: **Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus* Bath**. *FEMS Microbiol Lett* 2008, **288**:202–210.

75. Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, Albrecht D, Thürmer A, Daniel R, Kleiner M, Hecker M, Schweder T: **Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont**. *Proteomics* 2011, **11**:3106–3117.

76. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang Y-J, Shah M, VerBerkmoes NC, Zarzycki J, Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier N: **Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use.** *Proceedings of the National Academy of Sciences* 2012, **109**:E1173–82.

77. Bassham J, Benson A, Calvin M: **The path of carbon in photosynthesis.** *J Biol Chem* 1950, **185**:781–787.

78. Fenton A, Paricharttanakul N, Reinhart G: **Identification of substrate contact residues important for the allosteric regulation of phosphofructokinase from *Eschericia coli***. *Biochemistry* 2003, **42**:6453–6459.

79. Purves J, Cockayne A, Moody PCE, Morrissey JA: **Comparison of the regulation, metabolic functions, and roles in virulence of the glyceraldehyde-3-phosphate dehydrogenase homologues gapA and gapB in *Staphylococcus aureus***. *Infection and Immunity* 2010, **78**:5223–5232.

80. Wood AP, Aurikko JP, Kelly DP: **A challenge for 21st century molecular biology and biochemistry: what are the causes of obligate autotrophy and methanotrophy?** *Fems Microbiol Rev* 2004, **28**:335–352.

81. Han SO, Inui M, Yukawa H: **Effect of carbon source availability and growth phase on expression of *Corynebacterium glutamicum* genes involved in the tricarboxylic acid cycle and glyoxylate bypass.** *Microbiology* 2008.

82. Lee R, Thuesen E, Childress J: **Ammonium and free amino acids as nitrogen sources for the chemoautotrophic symbiosis *Solemya reidi* Bernard (Bivalvia: Protobranchia)**. *J Exp Mar Biol Ecol* 1992, **158**:75–91.

83. Liao L, Wankel SD, Wu M, Cavanaugh CM, Girguis PR: **Characterizing the plasticity of nitrogen metabolism by the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*.** *Molecular Ecology* 2013.

84. Lee RW, Childress JJ: **Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and methanotrophic symbionts.** *Appl Environ Microb* 1994, **60**:1852–1858.

85. Bourbonnais A, Lehmann MF, Butterfield DA, Juniper SK: **Subseafloor nitrogen transformations in diffuse hydrothermal vent fluids of the Juan de Fuca Ridge evidenced by the isotopic composition of nitrate and ammonium**. *Geochemistry, Geophysics, Geosystems* 2012, **13**:n/a–n/a.

86. Hentschel U, Felbeck H: **Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila***. *Nature* 1993, **366**:338–340.

87. Lee R, Robinson J, Cavanaugh CM: **Pathways of inorganic nitrogen assimilation in chemoautotrophic bacteria-marine invertebrate symbioses: expression of host and symbiont glutamine synthetase**. *J Exp Biol* 1999, **202 (Pt 3)**:289–300.

88. Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospesel M, Felbeck H, Zal F: **Fate of nitrate acquired by the tubeworm *Riftia pachyptila*.** *Appl Environ Microb* 2000, **66**:2783–2790.

89. Beckers G, Bendt AK, Kramer R, Burkovski A: **Molecular identification of the urea uptake system and**

transcriptional analysis of urea transporter and urease-encoding genes in *Corynebacterium glutamicum*. *J Bacteriol* 2004, **186**:7645.

90. De Cian M, Regnault M, Lallier FH: **Nitrogen metabolites and related enzymatic activities in the body fluids and tissues of the hydrothermal vent tubeworm *Riftia pachyptila*.** *J Exp Biol* 2000, **203**:2907–2920.

91. Joyner JL, Peyer SM, Lee RW: **Possible roles of sulfur-containing amino acids in a chemoautotrophic bacterium-mollusc symbiosis**. *Biol Bull* 2003, **205**:331–338.

92. Conway N, Howes B, McDowell Capuzzo J, Turner R, Cavanaugh CM: **Characterization and site description of *Solemya borealis* (Bivalvia; Solemyidae), another bivalve-bacteria symbiosis**. *Mar Biol* 1992, **112**:601–613.

93. Eichhorn E, van der Ploeg JR, Kertesz MA, Leisinger T: **Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli***. *J Biol Chem* 1997, **272**:23031–23036.

94. Conway N, McDowell Capuzzo J: **Incorporation and utilization of bacterial lipids in the *Solemya velum* symbiosis**. *Mar Biol* 1991, **108**:277–291.

95. Karow M, Georgopoulos C: **Isolation and characterization of the *Escherichia coli msbB* gene, a multicopy suppressor of null mutations in the high-temperature requirement gene *htrB***. *J Bacteriol* 1992, **174**:702–710.

96. Moran N, McCutcheon J, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts**. *Annual Review of Genetics* 2008, **42**:165–190.

97. Nussbaumer AD, Fisher CR, Bright M: **Horizontal endosymbiont transmission in hydrothermal vent tubeworms**. *Nature* 2006, **441**:345–348.

98. Cavanaugh CM: **Symbiosis of chemoautotrophic bacteria and marine invertebrates**. *PhD Thesis*. Harvard University, Department of Organismic and Evolutionary Biology; 1985.

99. Fisher C, Childress J: **Organic carbon transfer from methanotrophic symbionts to the host hydrocarbon-seep mussel.** *Symbiosis* 1992, **12**:221–235.

100. Saurin W, Hofnung M, Dassa E: **Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters**. *J Mol Evol* 1999, **48**:22–41.

101. van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H: **Plant peptides govern terminal differentiation of bacteria in symbiosis**. *Science* 2010, **327**:1122–1125.

102. Paau AS, Bloch CB, Brill WJ: **Developmental fate of *Rhizobium meliloti* bacteroids in alfalfa nodules.** *J Bacteriol* 1980, **143**:1480–1490.

103. Stewart FJ, Cavanaugh CM: **Bacterial endosymbioses in *Solemya* (Mollusca: Bivalvia)—model systems for studies of symbiont–host adaptation**. *Antonie van Leeuwenhoek* 2006, **90**:343–360.

104. Whitchurch CB, Leech AJ, Young MD, Kennedy D, Sargent JL, Bertrand JJ, Semmler ABT, Mellick AS, Martin PR, Alm RA, Hobbs M, Beatson SA, Huang B, Nguyen L, Commolli JC, Engel JN, Darzins A, Mattick JS: **Characterization of a complex chemosensory signal transduction system which controls twitching motility in *Pseudomonas aeruginosa***. *Mol Microbiol* 2004, **52**:873–893.

105. Newton ILG, Bordenstein SR: **Correlations Between Bacterial Ecology and Mobile DNA**. *Current Microbiology* 2011, **62**:198–208.

106. Plague GR, Dunbar HE, Tran PL, Moran NA: **Extensive proliferation of transposable elements in heritable bacterial symbionts**. *J Bacteriol* 2008, **190**:777–779.

107. Gil R, Latorre A, Moya A: **Evolution of Prokaryote-Animal Symbiosis from a Genomics Perspective**. In *Microbiology Monographs*. *Volume 19*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010:207–233.

108. Cordaux R, Pichon S, Ling A, Pérez P, Delaunay C, Vavre F, Bouchon D, Grève P: **Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont**. *Molecular Biology and Evolution* 2008, **25**:1889–1896.

109. Chafee ME, Funk DJ, Harrison RG, Bordenstein SR: **Lateral phage transfer in obligate intracellular bacteria (wolbachia): verification from natural populations.** *Molecular Biology and Evolution* 2010, **27**:501–505.

1009 110. Roeselers G, Newton ILG: **On the evolutionary ecology of symbioses between chemosynthetic bacteria and**
1010 **bivalves.** *Applied microbiology and biotechnology* 2012, **94**:1–10.

1011 111. Gil R, Sabater-Muñoz B, Latorre A, Silva FJ, Moya A: **Extreme genome reduction in *Buchnera* spp.: toward the**
1012 **minimal genome needed for symbiotic life**. *P Natl Acad Sci Usa* 2002, **99**:4454–4458.

1013 112. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL,
1014 Moran NA, Eisen JA: **Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters**.
1015 *PLoS Biol* 2006, **4**:1079–1092.

1016 113. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K,
1017 Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner
1018 C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, et al.: **Real-time DNA sequencing from single**
1019 **polymerase molecules.** *Science* 2009, **323**:133–138.

1020 114. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis
1021 ED, Phillippy AM: **Hybrid error correction and de novo assembly of single-molecule sequencing reads.** *Nat*
1022 *Biotechnol* 2012.

1023 115. *Pacific Biosciences* [http://www.pacb.com]

1024 116. Delcher AL, Bratke KA, Powers EC, Salzberg SL: **Identifying bacterial genes and endosymbiont DNA with**
1025 **Glimmer**. *Bioinformatics* 2007, **23**:673–679.

1026 117. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and**
1027 **translation initiation site identification**. *BMC Bioinformatics* 2010, **11**.

1028 118. Besemer J, Lomsadze A, Borodovsky M: **GeneMarkS: a self-training method for prediction of gene starts in**
1029 **microbial genomes. Implications for finding sequence motifs in regulatory regions**. *Nucleic Acids Res* 2001,
1030 **29**:2607.

1031 119. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH: **UniRef: comprehensive and non-redundant UniProt**
1032 **reference clusters.** *Bioinformatics* 2007, **23**:1282–1288.

1033 120. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I: **RefSeq microbial genomes database: new representation**
1034 **and annotation strategy.** *Nucleic Acids Res* 2014, **42**:D553–9.

1035 121. Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, Mavrommatis K, Meyer F: **The M5nr: a novel non-**
1036 **redundant database containing protein sequences and annotations from multiple sources and associated tools**.
1037 *BMC Bioinformatics* 2012, **13**.

1038 122. **Standard operating procedure for the annotations of genomes and metagenomes submitted to the**
1039 **integrated microbial genomes expert review (IMG-ER) system** [http://img.jgi.doe.gov/w/doc/img_er_ann.pdf]

1040 123. *Bprom* [http://www.softberry.com]

1041 124. Emanuelsson O, Brunak S, Heijne von G, Nielsen H: **Locating proteins in the cell using TargetP, SignalP and**
1042 **related tools**. *Nature Protocols* 2007, **2**:953–971.

1043 125. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A,
1044 Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947–
1045 2948.

1046 126. Robinson DG, Lee M-C, Marx CJ: **OASIS: an automated program for global investigation of bacterial and**
1047 **archaeal insertion sequences**. *Nucleic Acids Res* 2012, **40**:–e174.

1048 127. Bi D, Xu Z, Harrison EM, Tai C, Wei Y, He X, Jia S, Deng Z, Rajakumar K, Ou H-Y: **ICEberg: a web-based**
1049 **resource for integrative and conjugative elements found in Bacteria.** *Nucleic Acids Res* 2012, **40**(Database
1050 issue):D621–6.

1051 128. Leplae R, Lima-Mendez G, Toussaint A: **ACLAME: A CLAssification of Mobile genetic Elements, update 2010**.
1052 *Nucleic Acids Res* 2010, **38**(Database issue):D57–D61.

1053 129. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, Tortell PD, Hallam SJ: **Metagenome of a**

versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 2009, **326**:578–582.

Figures

Figure 1. Predicted model of the *S. velum* symbiont cell.

The diagram, based on gene annotation of the symbiont genome, depicts key functional systems and metabolic pathways: sulfur oxidation, electron transport, ATP synthases, CO2-fixation via Calvin-Benson Cycle, gluconeogenesis, polyglucose synthesis, glycolysis, TCA and glyoxylate cycles, synthesis of amino acids, fatty acids, lipids, isoprenoids via non-mevalonate pathway, and the cell wall, solute transporters, protein secretion systems, and type IV pilus. Different protein categories are color-coded and the individual subunits indicated by shape symbols. The direction of substrate transport across the membrane is shown with arrows. Components of the electron transport chain are arranged from the lowest to the highest electronegativity of the electron donors (blue) and acceptors (red). The corresponding electronegativity values are listed next to the enzymes. Enzymes shared between the glycolysis, gluconeogenesis, and the Calvin cycles are designated in green. Enzymes unique to these pathways are designated in red. Enzymes shared between the Calvin cycle and pentose phosphate cycle are designated in blue. Amino acids which may be essential for the host are designated in red. Speculated pathways are designated with a question mark. The abbreviations used, the respective full gene names, and their genome reference numbers are listed in Table S3 in Additional file 3.

Figure 2. Comparison of the COG categories between the *S. velum* symbiont, selected symbiotic and free-living bacteria. The percentage of genes in each category is normalized to the percentage of those COG categories in the genome of *E. coli* K12 DH1, ATCC 33849. *NCBI accession PRJNA16744.

Figure 3. Taxa assigned to the genes in the *S. velum* symbiont genome. The insert chart shows breakdown of the genes in the taxa within the class of γ-proteobacteria (62.9%). The unassigned genes have not been assigned a lower taxon in this analysis. The unclassified genes have not been further classified in the NCBI taxonomy. All the taxa are mutually exclusive.

Figure 4. Comparison of the sulfur oxidation genes between the *S. velum* symbiont and other SOBs. (a) Presence of genes involved in chemotrophic sulfur oxidation in the symbionts of *S. velum*, other sulfur-oxidizing bacteria and archaea, and sulfate reduction in *D. autotrophicum*, which is included for comparison. Genes encoding pathways for reverse-acting dissimilatory sulfur-oxidation (rDsr) (Drs in *D. autotrophicum*) and periplasmic sulfur-oxidation (Sox), as well as auxiliary proteins, are listed. Numbers of gene homologs in each organism is designated with color. Presence of extra- or intracellular sulfur deposits, i.e., globules, in each organism, as obtained from literature, is indicated with hollow circles. The gene symbols used, the respective full gene names, and the reference numbers in the genome of the *S. velum* symbionts are listed in Table S3 in Additional file 3. (b) Presence of signal sequences and transmembrane domains in the sulfur-oxidations genes of the *S. velum* symbionts, followed by the list of organisms with the closest known homologs to those genes and their respective BLAST % identities (Avi - *Allochromatium vinosum*, Sup05 - uncultivated oxygen minimum zone microbe [129], Sli - *Sideroxydans lithotrophicus*, and Thia - *Thiocapsa marina*).

Tables

Table 1. **General genome features of the *S. velum* symbiont in comparison to other γ-proteobacteria.** The comparison includes genomes of the chemosynthetic symbionts of *R. pahyptila*, *C. magnifica*, and *C. okutanii*; a symbiont of psyllids (the smallest sequenced genome), *Carsonella ruddii*; an α-proteobacterial aphid symbiont, *B. aphidicola*; free-living sulfur-oxidizers, *T. crunogena* and *A. vinosum*, and enterobacterium *E. coli*. * NCBI Accession PRJNA16744 and PRJNA72967.

	***Solemya velum* endosymbiont**	*Riftia pachyptila* endosymbiont*	*Calyptogena magnifica* endosymbiont	*Calyptogena okutanii* endosymbiont	*Buchnera aphidicola* APS	*Ca. Carsonella ruddii* PV	*Thiomicrospira crunogena* XCL-2	*Allochromatium vinosum* DSM 180	*Escherichia coli* K12 DH1, ATCC 33849
Size, mb	**2.70**	3.20	1.20	1.02	0.65	0.16	2.40	3.60	4.63
G + C%	**51.0**	57.9	34.0	31.6	26.4	16.6	43.1	64.3	50.8
ORFs	**2757**	4182	1118	981	615	213	2263	3317	4273
Average ORF length, bp	**885**	354	874	897	935	737	974	1005	940
Percent coding	**90.7**	69,8	79,8	85,9	87,6	97,3	90,5	90,6	86,6
rRNA operons (16S-23S-5S)	**1**	1	1	1	1	1	3	3	7
tRNA genes	**38**	32	36	36	32	28	43	51	88
Proteins with predicted function	**1988**	2218	932	838	561	113	1785	2505	3506
Hypothetical and uncharacterized conserved proteins	**769**	3693	175	253	106	46	689	924	833
ORFs in paralogous families	**382**	292	27	19	7	0	159	413	794
Pseudogenes	**0**	0	100	2	1	0	8	81	178
Sigma factors	**9**	4	2	2	2	0	6	6	7
Mobile elements	**53**	10	0	0	0	0	10	19	39
	Symbionts						Free-living		

Table 2. **Comparison of extracellular transport genes in the *S. velum* symbiont, other symbiotic and free-living bacteria.**

Organism	Lifestyle	Transporter gene ratio to *S. velum* endosymbiont	Genome size (Mb)	Total number of genes involved in transport	Transporter genes per Mb of genome	ATP-dependent	Secondary	Phosphotransferase system	Ion channels	Unclassified	Protein secretion systems	Outer membrane transport
***Solemya velum* endosymbiont**	**Intracellular symbiont**	**1.00**	**2.7**	**224**	**75.2**	**100**	**70**	**1**	**5**	**5**	**17**	**26**
Calyptogena magnifica endosymbiont	Obligate intracellular symbiont	0.14	1.16	32	27.6	18	6	0	3	1	0	4
Calyptogena okutanii endosymbiont	Obligate intracellular symbiont	0.15	1.02	34	33.3	16	10	0	2	3	0	3
Buchnera aphidicola APS	Obligate intracellular symbiont	0.07	0.64	16	25.0	5	3	5	1	0	0	2
Sulcia muelleri GWSS	Obligate intracellular symbiont	0.03	0.25	7	28.0	4	2	0	0	0	0	1
Candidatus Blochmannia floridanus	Obligate intracellular symbiont	0.12	0.71	27	38.0	7	12	3	2	0	0	3
Wigglesworthia glossinidia	Obligate intracellular symbiont	0.11	0.70	25	35.7	9	14	0	2	0	0	0
Baumannia cicadellinicola	Obligate intracellular symbiont	0.13	0.69	28	40.6	11	10	3	1	0	0	3
Rhizobium leguminosarum bv. Viciae 3841	Free-living intracellular symbiont	2.47	7.75	553	71.4	281	203	7	18	2	13	29
Frankia alni ACN14a	Free-living intracellular symbiont	1.05	7.50	236	31,5	114	106	0	12	1	1	4
Vibrio fischeri MJ11	Extracellular symbiont	1.80	4.50	404	89.8	138	141	12	10	6	46	51
Wolbachia pipientis wSim	Obligate intracellular symbiont/parasite	0.21	1.06	48	45.3	19	28	0	0	1	8	0
Rickettsia prowazekii MadridE	Intracellular parasite	0.21	1.10	48	43.6	15	30	0	1	1	0	5
Escherichia coli K-12-MG1655	Commensal	1.58	4.64	354	76,3	74	235	29	13	2	3	35
Klebsiella pneumoniae kp342	Commensal	2.82	5.92	632	106.8	160	336	44	17	4	37	34
Thiomicrospira crunogena XCL-2	Free-living sulfide oxidizer	0.73	2.43	163	67.1	38	58	0	10	3	35	19
Allochromatium vinosum DSM 180	Free-living sulfide oxidizer	0.89	3.67	199	54.2	81	52	5	8	7	14	32

Sulfurimonas denitrificans DSM 1251	Free-living sulfide oxidizer	0.43	2.20	97	44.1	32	52	0	10	3	6	25
Methylococcus capsulatus Bath	Free-living methanotroph	0.76	3.30	171	51.8	56	60	0	6	2	16	31
Thermodesulfovibrio yellowstonii DSM 11347	Free-living sulfate reducer	0.39	2.00	88	44.0	31	34	0	3	2	3	15

Table 3. **ICE mobile genetic elements in the *S. velum* symbiont genome.**

ICE element	Copies	Length, bp
ICEVchLao1	1	834
ICEVchBan7	1	432
ICEVchBan9	2	429, 888
ICEVchInd5	1	282
ICEVchMex1	1	561
ICEVflind1	2	405, 729
ICEPalBan1	1	1389
ICEPdaSpa1	5	300, 387, 622, 939, 3568
ICESpuPO1	3	549, 627, 648
ICEPmiUSA1	1	1290

Table 4. **Insertion sequence mobile genetic elements in the *S. velum* symbiont genome.**

Family/Element	Copies	Length, bp	Terminal Inverted Repeats
IS30	30	1071	ATTCAA
IS3/IS407	18	1219	CCCCCA/CCCCCAA(C/T)AAGT
IS30	1	900	CAACCGTTTCAAT
IS5/IS5	1	1638	ACCCAAGGTA
IS481	1	1271	GAGACATCATTTACA
IS30	1	1137	TGATGTACGGGTCCGA
unknown	1	1848	CCCCTTCG

Additional files

Additional file 1. Table S1.

Length [bp], GC%, percentage of the total base pairs, and the number of genes in the scaffolds which constitute the genome of the *S. velum* symbiont.

Additional file 2. Table S2.

tRNA genes and the codon frequencies in the genome of the *S. velum* symbiont.

Additional file 3. Table S3.

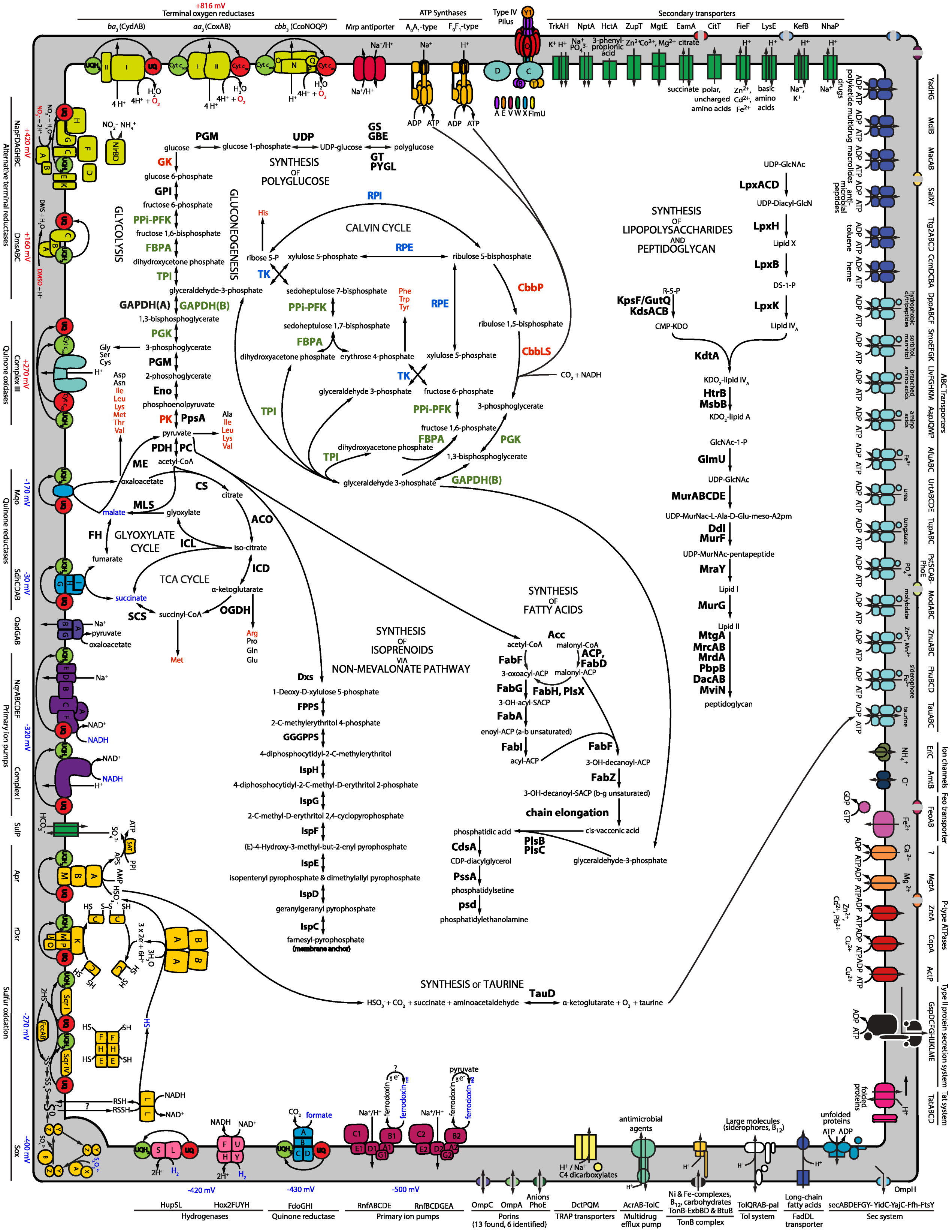
Gene product names used in Figure 1 and Figure 4, the corresponding NCBI protein ID reference numbers, and EC/TC numbers.

Additional file 4. Table S4.

Parameters of the gene prediction software.

Additional file 5. Genomic Utility for Automated Comparison (GUAC).

A Python script developed to inform comparative analyses of gene content across multiple genomes.



Terminal oxygen reductases
ATP Synthases
Type IV Pilus
Secondary transporters
Mrp antiporter
GLYCOLYSIS
GLUCONEOGENESIS
SYNTHESIS OF POLYGLUCOSE
CALVIN CYCLE
SYNTHESIS OF LIPOPOLYSACCHARIDES AND PEPTIDOGLYCAN
GLYOXYLATE CYCLE
TCA CYCLE
SYNTHESIS OF ISOPRENOIDS VIA NON-MEVALONATE PATHWAY
SYNTHESIS OF FATTY ACIDS
SYNTHESIS OF TAURINE
ABC Transporters
Ion channels
Feo transporter
P-type ATPases
Type II protein secretion system
Tat system
Hydrogenases
Quinone reductase
Primary ion pumps
Porins (13 found, 6 identified)
TRAP transporters
Multidrug efflux pump
TonB complex
Tol system
FadDL transporter
Sec system
Alternative terminal reductases
Quinone oxidases
Quinone reductases
Primary ion pumps
Sulfur oxidation

Figure 1

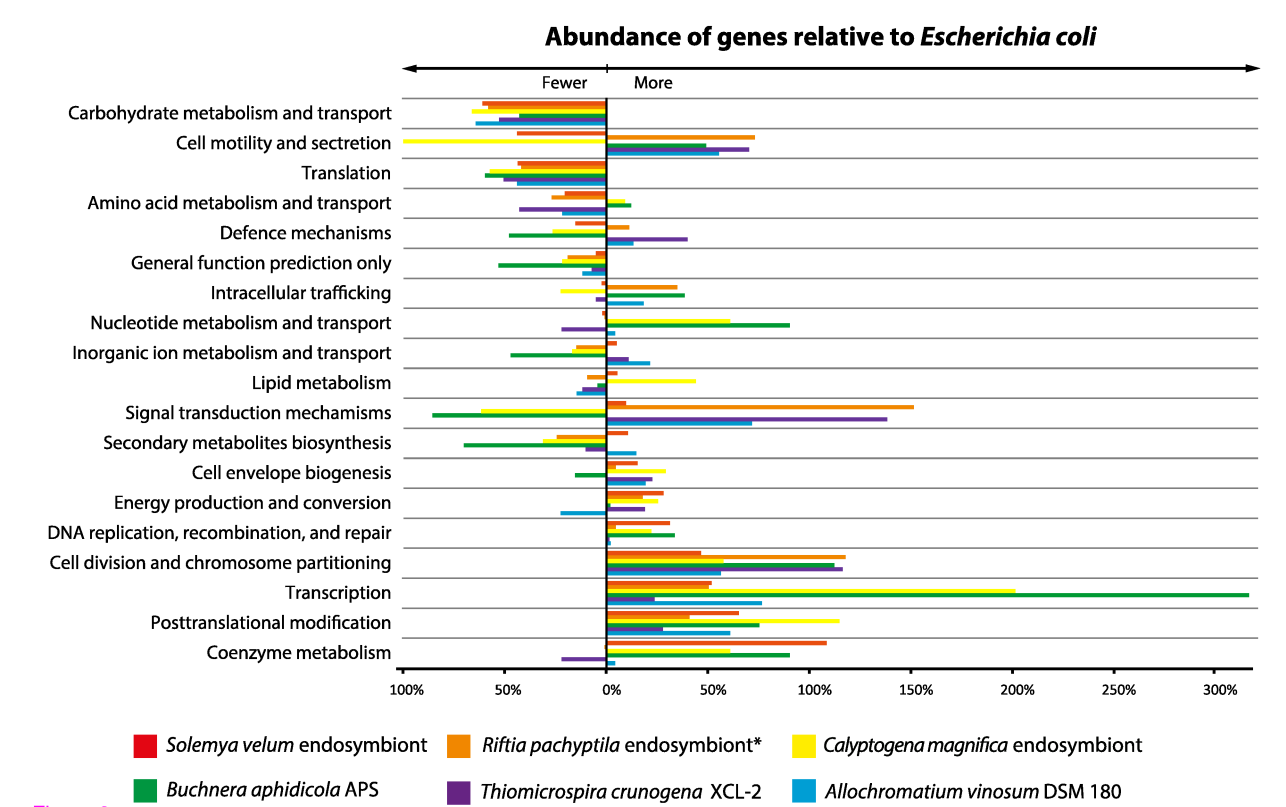

Abundance of genes relative to Escherichia coli
Fewer
More
Carbohydrate metabolism and transport
Cell motility and sectretion
Translation
Amino acid metabolism and transport
Defence mechanisms
General function prediction only
Intracellular trafficking
Nucleotide metabolism and transport
Inorganic ion metabolism and transport
Lipid metabolism
Signal transduction mechamisms
Secondary metabolites biosynthesis
Cell envelope biogenesis
Energy production and conversion
DNA replication, recombination, and repair
Cell division and chromosome partitioning
Transcription
Posttranslational modification
Coenzyme metabolism
100%
50%
0%
50%
100%
150%
200%
250%
300%
Solemya velum endosymbiont
Riftia pachyptila endosymbiont*
Calyptogena magnifica endosymbiont
Buchnera aphidicola APS
Thiomicrospira crunogena XCL-2
Allochromatium vinosum DSM 180

Figure 2

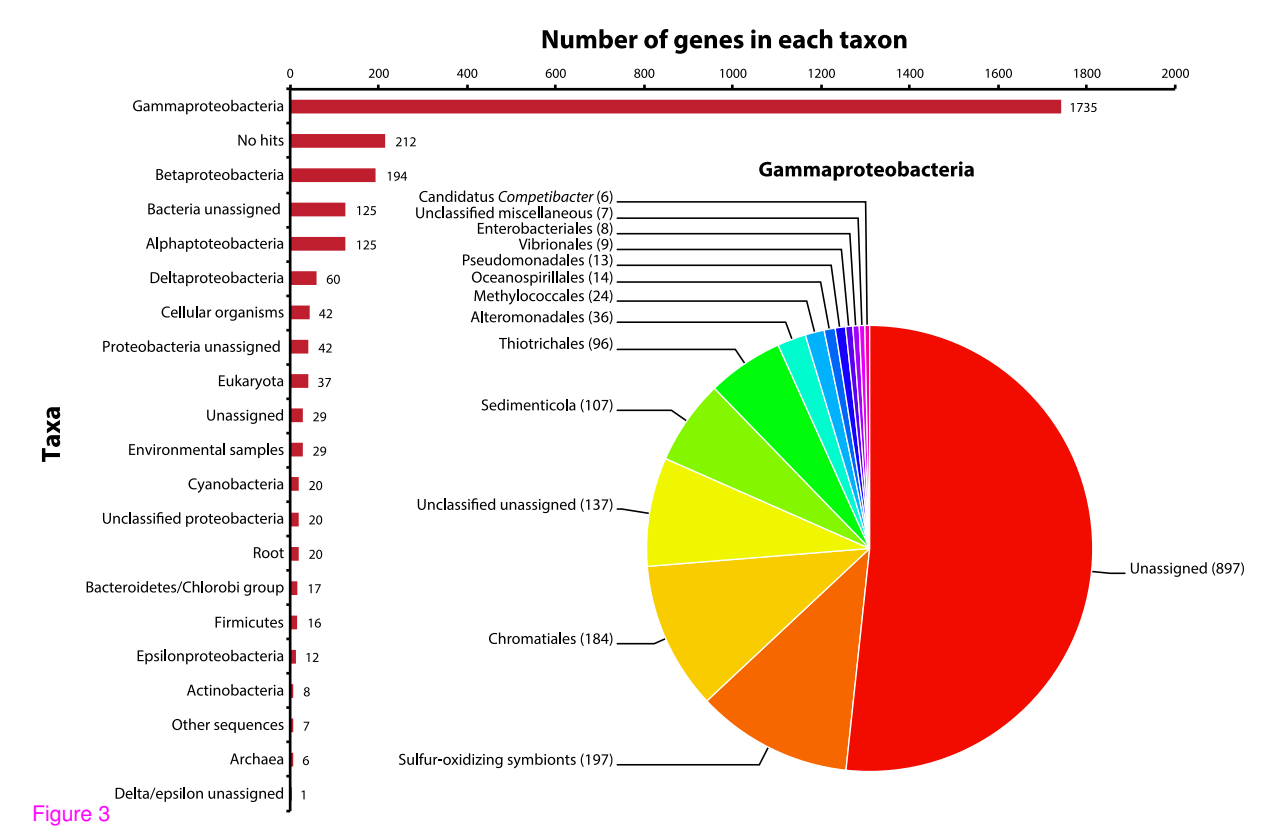

Number of genes in each taxon
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Taxa
Gammaproteobacteria 1735
No hits 212
Betaproteobacteria 194
Bacteria unassigned 125
Alphaptoteobacteria 125
Deltaproteobacteria 60
Cellular organisms 42
Proteobacteria unassigned 42
Eukaryota 37
Unassigned 29
Environmental samples 29
Cyanobacteria 20
Unclassified proteobacteria 20
Root 20
Bacteroidetes/Chlorobi group 17
Firmicutes 16
Epsilonproteobacteria 12
Actinobacteria 8
Other sequences 7
Archaea 6
Delta/epsilon unassigned 1
Gammaproteobacteria
Candidatus Competibacter (6)
Unclassified miscellaneous (7)
Enterobacteriales (8)
Vibrionales (9)
Pseudomonadales (13)
Oceanospirillales (14)
Methylococcales (24)
Alteromonadales (36)
Thiotrichales (96)
Sedimenticola (107)
Unclassified unassigned (137)
Chromatiales (184)
Sulfur-oxidizing symbionts (197)
Unassigned (897)

Figure 3

a
b
γ
β
α
ε
δ
Thermococcus gammatolerans
Sulfolobus solfataricus
Aquifex aeolicus
Thermus thermophilus
Sulfobacillus thermosulfidooxidans
Desulfobacterium autotrophicum
Sulfurimonas denitrificans
Chlorobium limicola
Chlorobium tepidum
Chlorobaculum parvum
Magnetospirillum magnetotacticum
Rhodopseudomonas palustris
Starkeya novella
Paracoccus denitrificans
Silicibacter pomeroyi
Roseobacter denitrificans
Rhodobacter capsulatus
Chromobacterium violaceum
Thiobacillus denitrificans
Beggiatoa sp
Halothiobacillus neapolitanus
Shewanella denitrificans
Calyptogena okutanii endosymbiont
Calyptogena magnifica endosymbiont
Thiothrix nivea
Halorhodospira halophila
Thioalkalivibrio sp
Allochromatium vinosum
Solemya velum endosymbiont
Riftia pachyptila endosymbiont
Thiomicrospira crunogena
Acidithiobacillus ferrooxidans
Solemya velum endosymbionts
% identity
Closest homolog
Signal sequence
Transmembrane domain
sulfur globules
dsrA
B
C
E
F
H
M
K
J
O
P
L
N
R
S
soxA*
A**
X
Y
Z
B
C
D
E
F
G
H
S
R
T
V
W
aprA
apr B
apr M
sat
sqr I
sqr III
sqr IV
sqr V
sqr VI
fccAB
sorA
sorB
qmoA
qmoB
qmoC
Avi 85
Avi 83
Avi 85
Avi 83
Sup05 71
Avi 52
Sup05 59
Avi 57
Avi 49
Avi 59
Avi 59
Avi 71
Sli 55
Avi 63
Avi 49
Avi 61
Sup05 56
Thia 56
Uncul 60
Avi 73
Tsul 91
Tsul 86
Tsul 66
Eper 81
Eper 75
Eper 70
Number of homologues
1 2 3 4 5 6 7 8 >=9
* - diheme ** - monoheme

Figure 4

Additional files provided with this submission:

Additional file 1: Additional file 1.xlsx, 51K
http://www.biomedcentral.com/imedia/1110977852141792/supp1.xlsx
Additional file 2: Additional file 2.xlsx, 43K
http://www.biomedcentral.com/imedia/1091587789141792/supp2.xlsx
Additional file 3: Additional file 3.xlsx, 75K
http://www.biomedcentral.com/imedia/1311208166141792/supp3.xlsx
Additional file 4: Additional file 4.xlsx, 39K
http://www.biomedcentral.com/imedia/2825041411417923/supp4.xlsx
Additional file 5: Additional file 5.cvs, 17K
http://www.biomedcentral.com/imedia/1556036798136870/supp5.cvs