

RESEARCH ARTICLE

Open Access

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

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

## 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 (2.7 Mb), GC-rich (51%), 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<sup>+</sup>/Na<sup>+</sup> -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

\* Correspondence: jaeisen@ucdavis.edu; cavanaugh@fas.harvard.edu  
<sup>8</sup>UC Davis Genome Center, 451 East Health Sciences Drive, Davis, CA 95616-8816, USA

<sup>1</sup>Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4081 Biological Laboratories, Cambridge, MA 02138, USA  
Full list of author information is available at the end of the article

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 chemo-synthesis [5,9].

*Solemya velum* and its endosymbionts is one of the best-described chemoautotrophic symbioses. The host, a proto-branch 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  $\gamma$ -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<sub>2</sub> 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 Vesicomidae [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 vesicomid symbioses, host and symbiont phylogenies are largely congruent, a pattern consistent with vertical symbiont transmission [24]. Nonetheless, instances of lateral symbiont movement among some vesicomids 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 vesicomids, 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* symbiont. 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 symbiont 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 vesicomid clams, *Calyptogena magnifica* [22] and *Calyptogena okutanii* [20], the vestimentiferan tubeworms, *Riftia pachyptila* [31] and *Tevnia jerichona* [32], the scaly-foot snail, *Crysmollon 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 (Additional file 1: Table S1). 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  $\gamma$ -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 (Additional file 2: Table S2).

A model of the symbiont cell based on functional predictions is presented in Figure 1 (see Additional file 3: Table S3 for the list of the corresponding gene products). 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  $\gamma$ -proteobacteria, mainly other sulfur-oxidizing symbionts (197 genes) and bacteria from the order of Chromatiales (184 genes). Among the genes within  $\gamma$ -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

### 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 symbiont [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  $\gamma$ -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 encoded 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 [43]. In fact, sulfur deposits are common to all known sulfur-oxidizing bacteria (SOB) which, like the *S. velum* symbiont, lack SoxCD sulfur dehydrogenase (Figure 4) [44], including the symbionts of the hydrothermal vent tubeworm, *R. pachyptila* [31,45], and the clam, *C. magnifica* [22,46]. 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* symbiont have the highest known carbon fixation rate, and, hence, demand for energy, of all the studied chemosynthetic symbionts, i.e.,  $65 \mu\text{mol min}^{-1} \text{g of protein}^{-1}$  [13] compared to  $0.45 \mu\text{mol min}^{-1} \text{g of protein}^{-1}$  of the next highest rate measured in the symbionts of *R. pachyptila* [47]. Alternatively, in the *S. velum* symbiont 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*

**Table 1 General genome features of the *S. velum* symbiont in comparison to other  $\gamma$ -proteobacteria**

	<i>Solemya velum</i> endosymbiont	<i>Riftia pachyptila</i> endosymbiont*	<i>Calyptogenia</i> <i>magnifica</i> endosymbiont	<i>Calyptogenia</i> <i>okutanii</i> endosymbiont	<i>Buchnera</i> <i>aphidicola</i> APS	<i>Ca. Carsonella</i> <i>ruddii</i> PV	<i>Thiomicrospira</i> <i>crunogena</i> XCL-2	<i>Allochromatium</i> <i>vinosum</i> DSM 180	<i>Escherichia coli</i> 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
	Symbiont	Symbiont	Symbiont	Symbiont	Symbiont	Symbiont	Free-living	Free-living	Free-living

The comparison includes genomes of the chemosynthetic symbionts of *R. pachyptila*, *C. magnifica*, and *C. okutanii*; a symbiont of psyllids (the smallest sequenced genome), *Carsonella ruddii*; an  $\alpha$ -proteobacterial aphid symbiont, *B. aphidicola*; free-living sulfur-oxidizers, *T. crunogena* and *A. vinosum*, and enterobacterium *E. coli*. \*NCBI Accession PRJNA16744 and PRJNA72967.

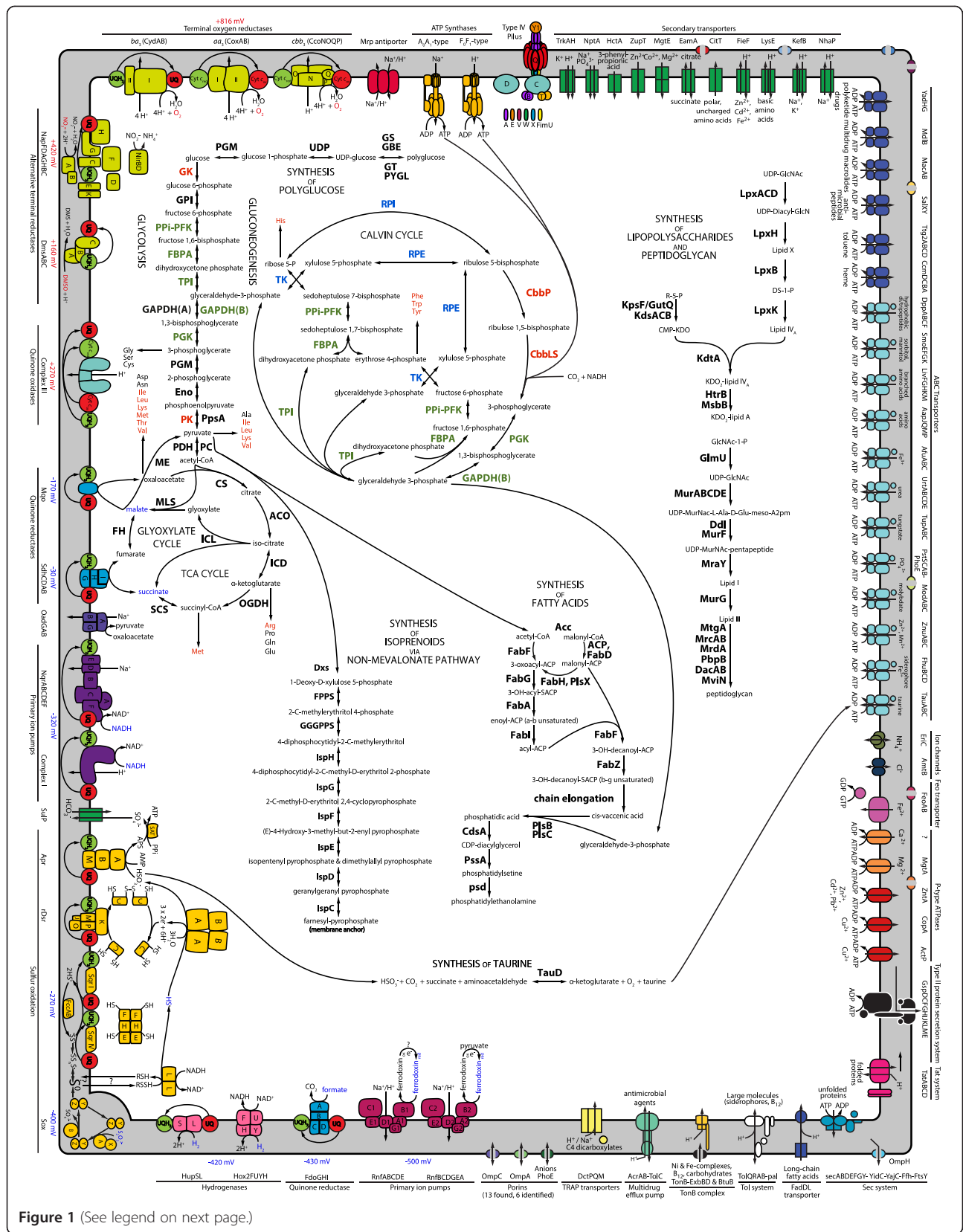


Figure 1 (See legend on next page.)

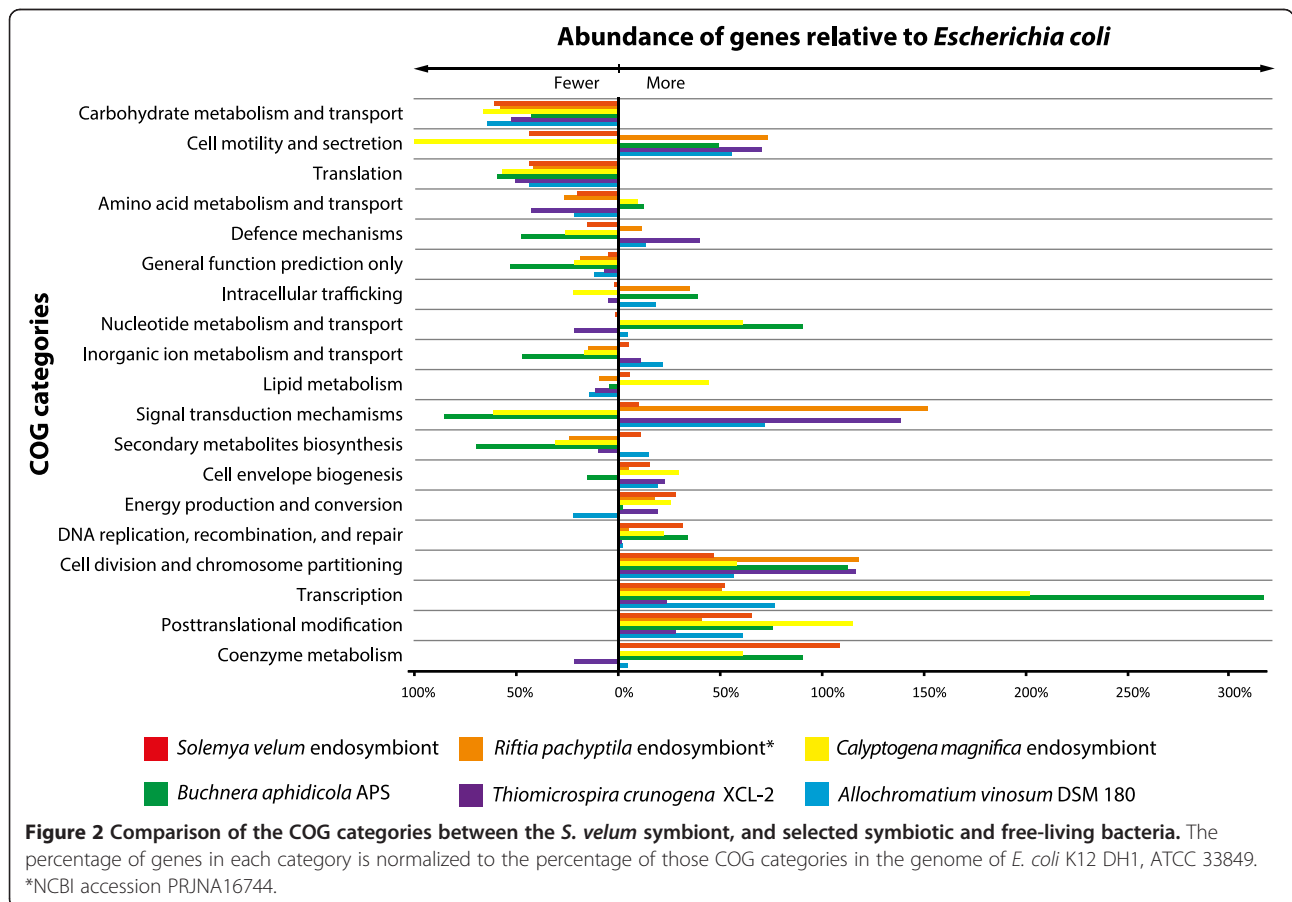


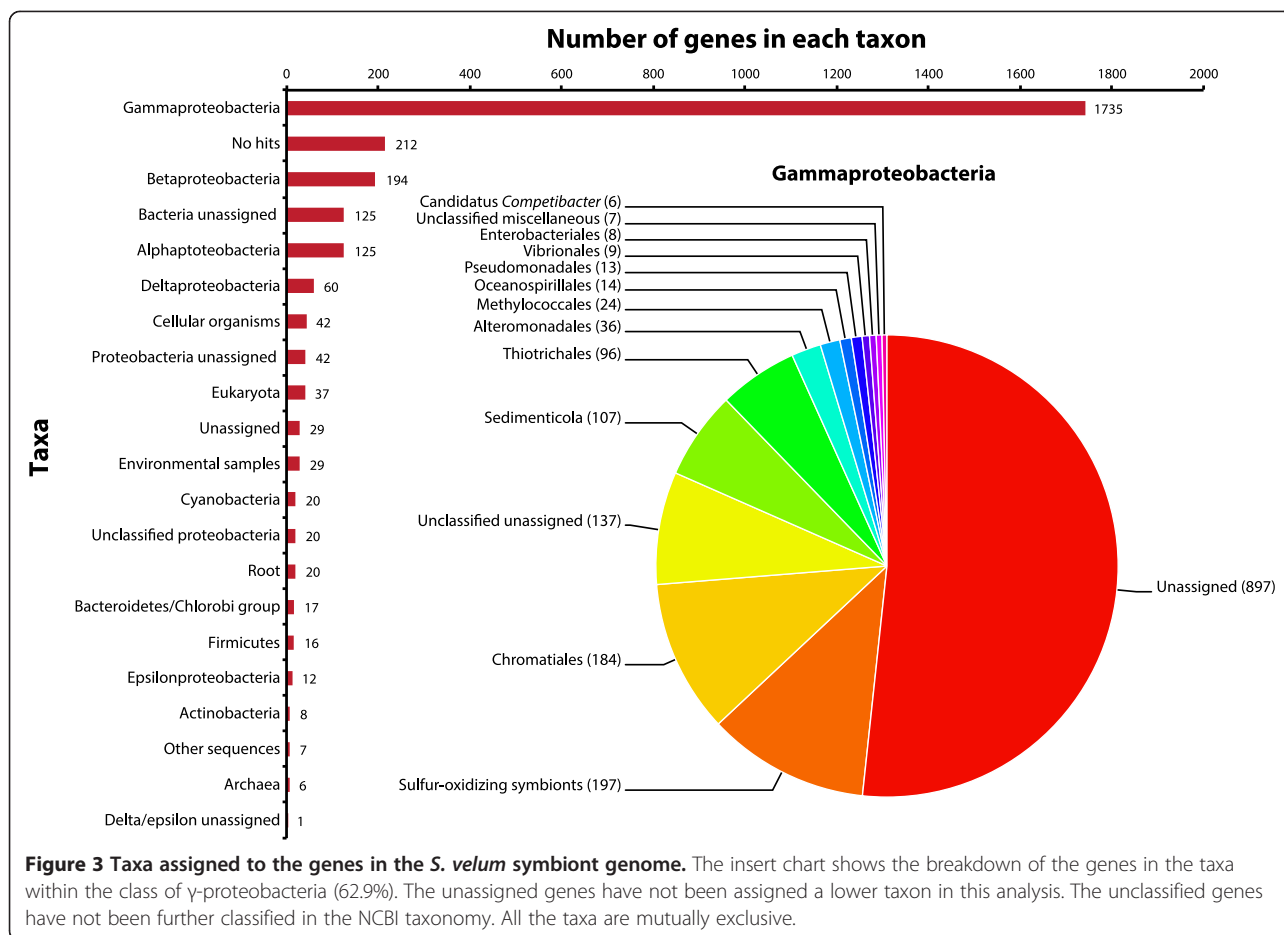
(See figure on previous page.)

**Figure 1 Predicted model of the *S. velum* symbiont cell.** The diagram, based on the gene annotation of the symbiont genome, depicts key functional systems and metabolic pathways: sulfur oxidation, electron transport, ATP synthases, CO<sub>2</sub>-fixation via the 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 the 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 respective enzymes. Enzymes shared between glycolysis, gluconeogenesis, and the Calvin cycle are designated in green. Enzymes unique to these pathways are designated in red. Enzymes shared between the Calvin cycle and the 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 product names, and the corresponding NCBI protein ID references are listed in Additional file 3: Table S3.

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 [48,49]. 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 [50]. rDsr may be the key energy-generating pathway in the symbiont, as sulfide has a six-fold higher effect on carbon fixation in the *S. velum* symbiosis [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) (Figures 1 and 4). Identification of the respective genes agrees with measured Apr and Sat activity in the symbiont-containing *S. velum* tissue [51]. 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 the 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 this bacterium is 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 ferredoxins, which have a redox potential as negative

as  $-500$  mV [52], compared, for example, to  $-400$  mV of  $S_2O_3^{2-}$  and  $-270$  mV of  $H_2S$ . The reversible oxidation of ferredoxins coupled to the 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 these 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 [52]. 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 [52]. Together with the fact that ferredoxins play a key role in anaerobic metabolism [53], 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.





not encode the di-heme cytochrome *b*, which is necessary to link H<sub>2</sub> oxidation to quinone reduction in the cellular membrane [56]. 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<sub>2</sub> oxidizers, it is possible that the identified cytochrome *b* may act in tandem with HupSL to enable H<sub>2</sub> oxidation.

Apart from potentially reducing the respiratory quinone pool with H<sub>2</sub>, the symbiont, by means of a bidirectional hydrogenase, may produce H<sub>2</sub> by oxidizing NAD<sup>+</sup>. The *S. velum* symbiont Hox2FUYH enzyme complex is most similar in amino acid sequence (63-66%) to the recently-characterized NAD<sup>+</sup>-reducing [Ni-Fe]-hydrogenase from *T. roseopersicina*, which can operate in reverse, generating H<sub>2</sub> when the high reduction state of the dinucleotide pool is growth-limiting [57]. As H<sub>2</sub> concentrations available to the *S. velum* symbiont have not been measured, it is unknown whether the H<sub>2</sub> oxidation contributes to primary production to the degree that has been recently demonstrated in a hydrothermal vent symbiosis [58].

**Primary ion pumps** NADH (-320 mV), potentially derived from oxidation of H<sub>2</sub> 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<sup>+</sup>-translocating quinone-reducing NADH dehydrogenase (NdhABCDEFGHIJKLMN), a homolog of the mitochondrial Complex I, as well as an alternative Na<sup>+</sup>-translocating NADH dehydrogenase (NqrABCDEF) (Figure 1). While Complex I is ubiquitous in bacteria, Nqr is found mainly in pathogenic and marine species [59]. Among symbiotic bacteria, *nqr* genes have so far been described only in *Buchnera* spp., an obligate endosymbiont of aphids [60]. The *S. velum* symbiont may be able to switch between Complex I and Nqr, preferentially generating either H<sup>+</sup> or Na<sup>+</sup> electrochemical gradients. Thus, depending on the cellular requirements, the symbiont 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 [61], is

used in formate-dependent oxygen respiration, allowing the bacteria to rapidly adapt to shifts from aerobiosis to anaerobiosis [62]. 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<sub>1</sub>*, 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<sup>+</sup>-motive terminal oxygen reductases (Figure 1), which suggest a capacity to respire O<sub>2</sub> over a wide range of concentrations. The genome contains a *ccoNOQP* operon encoding a *ccb<sub>3</sub>* cytochrome oxidase, which is known to function at nanomolar O<sub>2</sub> concentrations in the nitrogen-fixing plant symbionts, *Bradyrhizobium japonicum* [63], and in the microaerophilic human pathogens, *Campylobacter jejuni*, *Helicobacter pylori*, and *Neisseria meningitidis* [64]. The genome also encodes a *aa<sub>3</sub>* cytochrome oxidase (CoxAB), which is thought to function primarily under atmospheric oxygen concentrations [65] and is the only terminal oxidase in the symbionts of the bivalves *C. magnifica* [22] and *C. okutani* [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 [66]. 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<sub>3</sub><sup>-</sup> 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 ([67], 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 growth under nitrate-limited conditions (5 μM) [68]. The symbiont genome also encodes a DMSO

reductase (*dmsABC*), suggesting an ability to respire dimethyl sulfoxide (DMSO), a breakdown product of dimethylsulfoniopropionate (DMSP) produced, for example, by marine algae. DMSO is available at nanomolar concentrations in the coastal eutrophic environments inhabited by *S. velum* [69], and *Dms* genes are common to many marine sediment-dwelling bacteria, e.g., *Beggiatoa* and *Shewanella* [70,71].

**ATP synthases** Based on the genome data, both  $H^+$  and  $Na^+$  membrane gradients, established along the course of the electron transport chain during respiration, may drive ATP synthesis in the *S. velum* symbiont via either  $H^+$ - or  $Na^+$ -dependent ATP synthases (Figure 1). The  $H^+$ -specificity of the  $F_0F_1$ -type ATP synthase is suggested by the presence of two characteristic transmembrane helices within the *c* subunit. In contrast, an  $A_0A_1$ -type ATP synthase detected in the genome contains the characteristic  $Na^+$ -binding PXXXQ motif I and ES motif II in the rotor subunit K. While proton-translocating ATP synthases are predominant in bacteria,  $Na^+$ -coupled ATP synthesis driven by respiration has recently been recognized in some marine and pathogenic species [72,73]. To our knowledge, this is the first instance of a  $Na^+$ -translocating ATP synthase in a chemosynthetic symbiont.

**Ion gradient driven transporters** Cellular roles of the  $H^+$  and  $Na^+$  gradients in the *S. velum* symbiont appear to extend beyond ATP synthesis. Besides ATP synthases, the genome encodes diverse  $Na^+$ :substrate symporters and numerous  $Na^+$ : $H^+$  antiporters, including the multi-subunit MrpEFGBBCDD complex (Figure 1). These transporters, together with ATP synthases and respiratory ion pumps, may establish and consume simultaneous transmembrane gradients of proton and sodium ions in the symbiont [72]. These parallel cycles of  $H^+$  and  $Na^+$  would allow the *S. velum* symbiont to synthesize ATP and maintain pH homeostasis via two separate mechanisms.

#### Carbon metabolism

Autotrophic carbon fixation, fueled chiefly by sulfur oxidation, is the principal process in the *S. velum* symbiont, supplying both the symbiont and the host with organic carbon [14]. While previous studies focused primarily on RubisCO [10,74], 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 [22,75-77]. Furthermore, the genome of the *S. velum* symbiont contains the gene for  $\alpha$ -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 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-bisphosphate and sedoheptulose 1,7-bisphosphate *in vitro* has been previously demonstrated for the PPi-PFK from *Methylococcus capsulatus* [75], which shares 73% sequence identity with the *S. velum* symbiont homologue. Compared to the classical Calvin cycle [78], the use of a PPi-PFK can offer energy savings up to 9.25% [77]. While fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase enzymes release phosphate, PPi-PFK produces energy-rich pyrophosphate. In *M. capsulatus* [75] and in the chemosynthetic symbionts of *R. pachyptila* [76] and the oligochete, *O. algarvensis* [77] it was proposed that the pyrophosphate could be converted into an electrochemical gradient by a membrane-bound proton-pumping pyrophosphatase ( $H^+$ -PPase) co-encoded with the 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 convert the energy of a pyrophosphate into 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 symbiont 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-phosphate 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 [79], since this PPi-PFK is nonallosteric [75]. 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* [80]. 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 the 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 rigorously controlling the direction of the carbon flux.

**Heterotrophy** The *S. velum* symbiont is the third chemosynthetic symbiont, along with the  $\gamma$ -symbiont of *O. algarvensis* [34] and the intracellular  $\gamma$ -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  $\alpha$ -ketoglutarate dehydrogenase and citrate synthase, which suggests their obligate autotrophy [81].

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, [82] or rapidly replenish intermediates of biosynthetic reactions. The presence of the 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 host-free existence of the symbiont.

#### **Nitrogen metabolism**

Ammonia, abundant in the sediment where *S. velum* burrows, is the main form of nitrogen assimilated by the symbiosis [83]. It has been suggested that the symbionts incorporate ammonia into biomass, which is then transferred to the host ([67] in preparation), a process which has been described for the chemosynthetic symbionts of the hydrothermal vent tubeworm *Ridgeia piscesae* [84]. The presence of assimilatory nitrogen pathways in the *S. velum* symbiont genome corroborate this hypothesis.

**Nitrogen assimilation** Extracellular ammonia may be imported by the symbiont 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). The *S. velum* symbiosis comes in contact with 20–100  $\mu$ M concentration of ammonia in its coastal environment ([67] in preparation). Thus, it is not surprising that, unlike the chemosynthetic symbionts found at nitrate-rich (40  $\mu$ M) hydrothermal vents [85,86], the *S. velum* symbiont lacks *nar* genes for nitrate reductases capable of assimilatory nitrate reduction [32,87–89]. 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) [88]. The present analysis identified *glnA*, the gene that encodes GS, in the genome of the symbiont. 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 symbiont is in keeping with its 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 the symbiont. Urea can enter the bacterial cell by passive diffusion [90], but under nitrogen starvation the symbiont 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  $\gamma$ -symbionts from the marine oligochaete worm *O. algarvensis* [34,77], 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* [91], 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 in the pore water ([67] in preparation).

**Taurine synthesis** The *S. velum* symbiont may also provide its host with nitrogenous osmoregulators, such as the non-proteinogenic amino acid taurine [92]. In the host tissues, taurine accounts for up to 70% of the total free amino acids and shows an isotopic composition ( $\delta^{13}\text{C}$ ,  $\delta^{14}\text{N}$ ,  $\delta^{34}\text{S}$ ) suggestive of symbiont origin [93]. Synthesis of taurine may be accomplished by the two homologs of the reversible taurine dioxygenase (TauD) encoded in the symbiont genome. Taurine could be actively secreted to the host by the TauABCD ABC transporter, the genes for which were found to contain a conserved binding domain for sulfonate, characteristic of the taurine molecule. Since taurine synthesis requires

sulfite [94], one of the final intermediates in sulfur oxidation, this pathway could serve to dispose of  $\text{SO}_3^{2-}$ , and, thus, to drive forward sulfur oxidation in the *S. velum* symbiont, benefiting both the host and the symbiont.

#### Membrane-associated functions

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

#### Production of cellular envelope

The *S. velum* symbiont appears capable of synthesizing and assembling a cytoplasmic membrane, a peptidoglycan layer (PG), and an outer membrane populated by lipopolysaccharides (LPS), which constitute a cellular envelope. While these abilities are typical of the free-living  $\gamma$ -proteobacteria, two aspects in particular stand out in the context of the symbiotic life-style. First, given the identified genes for the biosynthesis of fatty-acids, the symbiont may build components of its plasma membrane mostly from *cis*-vaccenic acid (18:  $\omega$ 7) (Figure 1). According to a previous analysis of lipid composition in *S. velum* [95], this unsaturated fatty acid and its derivatives are the main constituents of cellular membranes in the symbiont and the host alike. Furthermore, the isotopic signature of the host's lipids indicates that they are bacterial in origin [95]. 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* [96]. 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 [97]. However, the symbionts which have the genes to synthesize LPS tend to either live directly in the cytoplasm [97] and, have to make their own cellular envelope, or, like the symbionts of *R. pachyptila* [98], 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.

#### 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 symbiont has 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 [99], which suggests a transport mechanism, since direct digestion of symbionts by host cells would likely take hours to days [100]. 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 [101]. 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* symbiont 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 [102]. As a result, *R. leguminosarum* undergoes cell elongation and genome replication but loses its ability to divide. Only a small number of *R. leguminosarum* cells remain vegetative [103]. A very similar morphological differentiation of the symbiont has been observed in *S. velum* [104]. Assuming the bivalve host also uses peptide factors to control its symbiont, 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.

#### 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 symbiont. 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.



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

Organism	Lifestyle	Transporter gene ratio to <i>S. velum</i> 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
<b><i>Solemya velum</i> endosymbiont</b>	<b>Intracellular symbiont</b>	<b>1.00</b>	<b>2.7</b>	<b>224</b>	<b>75.2</b>	<b>100</b>	<b>70</b>	<b>1</b>	<b>5</b>	<b>5</b>	<b>17</b>	<b>26</b>
<i>C. magnifica</i> endosymbiont	OIS*	0.14	1.16	32	27.6	18	6	0	3	1	0	4
<i>C. okutanii</i> endosymbiont	OIS	0.15	1.02	34	33.3	16	10	0	2	3	0	3
<i>Buchnera aphidicola</i> APS	OIS	0.07	0.64	16	25.0	5	3	5	1	0	0	2
<i>Sulcia muelleri</i> GWSS	OIS	0.03	0.25	7	28.0	4	2	0	0	0	0	1
<i>Ca. Blochmannia floridanus</i>	OIS	0.12	0.71	27	38.0	7	12	3	2	0	0	3
<i>Wigglesworthia glossinidia</i>	OIS	0.11	0.70	25	35.7	9	14	0	2	0	0	0
<i>Baumannia cicadellinicola</i>	OIS	0.13	0.69	28	40.6	11	10	3	1	0	0	3
<i>R. leguminosarum</i> bv. <i>Viciae</i> 3841	FIS*	2.47	7.75	553	71.4	281	203	7	18	2	13	29
<i>Frankia alni</i> ACN14a	FIS	1.05	7.50	236	31.5	114	106	0	12	1	1	4
<i>Vibrio fischeri</i> MJ11	Extracellular symbiont	1.80	4.50	404	89.8	138	141	12	10	6	46	51
<i>Wolbachia pipientis</i> wSim	Obligate intracellular symbiont/parasite	0.21	1.06	48	45.3	19	28	0	0	1	8	0
<i>Rickettsia prowazekii</i> MadridE	Intracellular parasite	0.21	1.10	48	43.6	15	30	0	1	1	0	5
<i>Escherichia coli</i> K-12-MG1655	Commensal	1.58	4.64	354	76.3	74	235	29	13	2	3	35
<i>Klebsiella pneumoniae</i> kp342	Commensal	2.82	5.92	632	106.8	160	336	44	17	4	37	34
<i>Thiomicrospira crunogena</i> XCL-2	Free-living sulfide oxidizer	0.73	2.43	163	67.1	38	58	0	10	3	35	19
<i>Allochromatium vinosum</i> DSM 180	Free-living sulfide oxidizer	0.89	3.67	199	54.2	81	52	5	8	7	14	32
<i>Sulfurimonas denitrificans</i> DSM 1251	Free-living sulfide oxidizer	0.43	2.20	97	44.1	32	52	0	10	3	6	25
<i>Methylococcus capsulatus</i> Bath	Free-living methanotroph	0.76	3.30	171	51.8	56	60	0	6	2	16	31
<i>Thermodesulfobivrio yellowstonii</i> DSM 11347	Free-living sulfate reducer	0.39	2.00	88	44.0	31	34	0	3	2	3	15

OIS - obligate intracellular symbiont; FIS - free-living intracellular symbiont.



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 *pilEYIXWVT* genes required to assemble a functional pilus. Furthermore, the symbiont genome contains *pilGII-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 [105]. 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 this bacterium 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 [106].

Two hypothesized life and evolutionary history scenarios may explain the observed mobile element content in the *S. velum* symbiont. One of them is a relatively recent shift to intracellularity, resulting in an expansion of mobile elements [107,108]. 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 [109,110]. In fact, horizontal transmission or host-switching has likely occurred in the history of symbionts of bivalves [111] 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*

**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

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 the 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 the highly branched electron transport chain. These functions may relate to the fact that the *S. velum* symbiosis lives in eutrophic sediment, unlike the oligotrophic environments inhabited by other chemosynthetic symbioses, e.g., those of *R. pachyptila*, *C. magnifica*, and *O. algarvensis*.

**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

The *S. velum* symbiont has long been considered to be vertically transmitted [17], but our genomic analyses are inconsistent with predictions based on other vertically transmitted obligately-intracellular bacteria. The *S. 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 the 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, 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.

## 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 in 5 ml of 0.2  $\mu$ m filtered seawater (FSW) per bivalve. Homogenates were passed through 100  $\mu$ m and 5  $\mu$ m nylon filters (Small Parts Inc. #CMN-0105-A and CMN-0005-A) and centrifuged at 5,000  $\times$  g for 5 minutes at 4°C. The pellet was resuspended in FSW, pelleted, and resuspended in 1x TAE buffer. 50 g molten 2% agarose (SeaKem® #50152) in 1x TAE was added to make plugs for genomic DNA extraction. The hardened plugs were treated with DNase I (0.25U/50  $\mu$ l) at 37°C for 10 minutes and then equilibrated in TE buffer for 30 minutes at room temperature. Agarose plugs were further processed using CHEF Mammalian Genomic DNA Plug Kit from Bio-Rad Laboratories (#170-3591) according to the manufacturer's instructions. The protocol for pulse field gel electrophoresis (PFGE) and isolation of the bacterial chromosomes from the agarose plugs was adapted from Gil [112].

### Genome sequencing and assembly

Genomic bacterial DNA was sequenced at the Institute for Genomic Research (TIGR), the Joint Genome Institute (JGI), and the University of California, Davis, using a diversity of sequencing technologies. Two Sanger libraries of 3–4 Kb and 10–12 Kb insert sizes were constructed as previously described [113]. Sequencing of these Sanger libraries resulted in 110,187 reads with N50 of 969 bp and the average coverage depth of 8x. Subsequently, using Roche 454 technology, 387,143 sequencing reads with the N50 of 207 bp and the average coverage depth of 13x were obtained. Then, 25,635,107 Illumina sequencing reads were generated. The Illumina sequences were 35 bp long and had the average coverage depth of 150x. These Sanger, Roche 454, and Illumina sequences were assembled using the Paracel Genome Assembler (Paracel Inc., Pasadena, CA) into 68 contigs. Next, symbiont DNA was sequenced using Pacific Biosciences (PacBio) technology, resulting in 150,000 reads with N50 of 4,966 bp and 9x coverage depth. The insertion and deletion (indels) errors, typical of the PacBio data [114], were reduced from 4% to 0.2% with Illumina paired-end sequences (500x coverage) using PacBioToCA program [115] available as a part of SMRT Analysis software package version 1.4 distributed by the Pacific Biosciences [116]. The error correction step also removed any PacBio sequences of the host origin, which, given the abundance of the symbionts in the gill tissue, had Illumina coverage below 5x. The Illumina data used for the error correction were generated as part of a different study and came from a specimen obtained at a different location (Point Judith, RI) than the rest of the genomic data. Due to the extent of the intra-species genomic sequence variation across geographic localities (Russell et al., in preparation), these Illumina data could not be used to supplement the genome assembly but were sufficient to correct the majority of sequencing indel errors in the PacBio reads. The error-corrected 54,684 PacBio sequences with N50 of 1,409 bp were used to connect the previous 68 genomic contigs into 30 larger scaffolds using the Automated Hybrid Assembly (AHA) module of SMRT Analysis. The resulting 7 gaps within the scaffolds, 2,272 bp in total, were then filled in with the PacBio error-corrected sequences using the PBJelly software tool [117], reducing the number of gaps to 4 and the total gap length to 100 bp. After discarding 20 of the smallest low coverage (2–9x) scaffolds that contained mostly eukaryotic genes (>65%), identified as described below, only 10 scaffolds were retained as a part of the symbiont genome.

### Sequence analysis

Open reading frames (ORFs) on *S. velum* symbiont scaffolds were predicted using Glimmer [118], Prodigal [119], and GeneMarkS [120]. The software parameters

used to perform these analyses are listed in Additional file 4: Table S4. Once identified, the ORFs were translated into protein-coding sequences and queried against the UniProt Reference Clusters (UniRef90) (20 November 2013) [121], National Center for Biotechnology Information non-redundant (NCBI-nr) (4 November 2013) [122], and M5 non-redundant (M5-nr) (27 January 2014) [123] databases for functional annotation using BLASTP (e-value cutoff 0.001) [38]. UniRef90 gene entries sharing the 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 [124]. Selected origins of replication were verified by Ori-Finder [125]. 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 [126]. Signal peptides and transmembrane domains were predicted using SignalP 3.0 Server and TMHMM, respectively [127]. 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 [128]. 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 [129]. Integrative conjugative elements and plasmid as well as phage sequences were identified by BLASTN [38] searches against the ICEberg [130] and ACLAME [131] 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.

#### 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 Figures 1 and 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.

#### 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 and coordinated 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.

#### Author details

<sup>1</sup>Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, 4081 Biological Laboratories, Cambridge, MA 02138, USA. <sup>2</sup>Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar Street, Cambridge, MA 02139, USA. <sup>3</sup>SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai 200136, China. <sup>4</sup>Microbiology & Systems Biology Group, TNO, Utrechtseweg 48, Zeist, Utrecht 3704HE, The Netherlands. <sup>5</sup>School of Biology, Georgia Institute of Technology, Atlanta, GA 30332-0230, USA. <sup>6</sup>Department of Biology, Indiana University, 1001 East 3rd Street, Jordan Hall, Bloomington, IN 47405, USA. <sup>7</sup>DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA. <sup>8</sup>UC Davis Genome Center, 451 East Health Sciences Drive, Davis, CA 95616-8816, USA.



Received: 3 April 2014 Accepted: 23 September 2014  
Published: 23 October 2014

## 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?** *Protospasma* 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**:1–8.
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<sub>2</sub> 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.** *Limnol Oceanogr* 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 Mar 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.** *Mol Biol Evol* 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 vesicomid 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.** *Proc Natl Acad Sci U S A* 1998, **95**:9962.
25. Stewart FJ, Young CR, Cavanaugh CM: **Lateral symbiont acquisition in a maternally transmitted chemosynthetic clam endosymbiosis.** *Mol Biol Evol* 2008, **25**:673–687.
26. Stewart FJ, Young C, Cavanaugh CM: **Evidence for homologous recombination in intracellular chemosynthetic clam symbionts.** *Mol Biol Evol* 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 ratchet in endosymbiotic bacteria.** *Proc Natl Acad Sci U S A* 1996, **93**:2873–2878.
30. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadijeh 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* wMel: a streamlined genome overrun by mobile genetic elements.** *PLoS Biol* 2004, **2**:E69.
31. Robidart J, Bench S, Feldman R, Novoradovsky A, Podell S, Gaasterland T, Allen E, Felbeck H: **Metabolic versatility of the *Riftia pachytila* endosymbiont revealed through metagenomics.** *Environ Microbiol* 2008, **10**:727–737.
32. Gardebrecht A, Markert S, Sievert SM, Felbeck H, Thürmer A, Albrecht D, Wollherr A, Kabisch J, Le Bris N, Lehmann R, Daniel R, Liesegang H, Hecker M, Schweder T: **Physiological homogeneity among the endosymbionts of *Riftia pachytila* and *Tevnia jerichonana* revealed by proteogenomics.** *ISME J* 2012, **6**:766–776.
33. Nakagawa S, Shimamura S, Takaki Y, Suzuki Y, Murakami S-I, Watanabe T, Fujiyoshi S, Mino S, Sawabe T, Maeda T, Makita H, Nemoto S, Nishimura S-I, Watanabe H, Watsujii T-O, Takai K: **Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont.** *ISME J* 2014, **8**:40–51.
34. Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehlend C, Rubin EM, Dubilier N: **Symbiosis insights through metagenomic analysis of a microbial consortium.** *Nature* 2006, **443**:950–955.
35. Wu M, Eisen JA: **A simple, fast, and accurate method of phylogenomic inference.** *Genome Biol* 2008, **9**:1–11.
36. Murphy FV, Ramakrishnan V: **Structure of a purine-purine wobble base pair in the decoding center of the ribosome.** *Nat Struct Mol Biol* 2004, **11**:11251–11252.
37. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631–637.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
39. Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC: **Integrative analysis of environmental sequences using MEGAN4.** *Genome Res* 2011, **21**:1552–1560.
40. Stewart FJ, Dmytrenko O, DeLong E: **Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of *Solemya velum*.** *Frontiers Microbiol* 2011, **2**:1–10.
41. Frigaard N-U, Dahl C: **Sulfur metabolism in phototrophic sulfur bacteria.** *Adv Microb Physiol* 2009, **54**:103–200.
42. 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.
43. Dahl C, Prange A: **Bacterial sulfur globules: occurrence, structure and metabolism.** In *Inclusions in Prokaryotes Microbiology Monographs*, Volume 1. 2006:21–51.
44. Friedrich C, Bardischewsky F, Rother D, Quentmeier A, Fischer J: **Prokaryotic sulfur oxidation.** *Curr Opin Microbiol* 2005, **8**:253–259.

45. Fisher C, Childress J, ARP A, BROOKS J, DISTEL D, Favuzzi J, Macko S, Newton A, Powell M, Somero G, SOTO T: **Physiology, morphology, and biochemical composition of *Riftia pachyptila* at Rose Garden in 1985.** *Deep-Sea Res* 1988, **35**:1745–1758.
46. Vetter RD: **Elemental sulfur in the gills of three species of clams containing chemoautotrophic symbiotic bacteria: a possible inorganic energy storage compound.** *Mar Biol* 1985, **88**:33–42.
47. Childress JJ, Girguis PR: **The metabolic demands of endosymbiotic chemoautotrophic metabolism on host physiological capacities.** *J Exp Biol* 2011, **214**:312–325.
48. Cort JR, Selan U, Schulte A, Grimm F, Kennedy MA, Dahl C: ***Allochrochromatium vinosum* DsrC: Solution-state NMR structure, redox properties, and interaction with DsrEFH, a protein essential for purple sulfur bacterial sulfur oxidation.** *J Mol Biol* 2008, **382**:692–707.
49. Oliveira TF, Vornrhein C, Matias PM, Venceslau SS, Pereira IAC, Archer M: **Purification, crystallization and preliminary crystallographic analysis of a dissimilatory DsrAB sulfite reductase in complex with DsrC.** *J Struct Biol* 2008, **164**:236–239.
50. Ghosh W, Dam B: **Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea.** *Fems Microbiol Rev* 2009, **33**:999–1043.
51. Chen C, Rabourdin B, Hammen C: **The effect of hydrogen sulfide on the metabolism of *Solemya velum* and enzymes of sulfide oxidation in gill tissue.** *Comp Biochem Physiol B Biochem Mol Biol* 1987, **88**:949–952.
52. Biegel E, Schmidt S, González JM, Müller V: **Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes.** *Cell Mol Life Sci* 2011, **68**:613–634.
53. Bruschi M, Guerlesquin F: **Structure, function and evolution of bacterial ferredoxins.** *Fems Microbiol Rev* 1988, **4**:155–175.
54. Kovács KL, Kovács AT, Maróti G, Mészáros LS, Balogh J, Latinovics D, Fülöp A, Dávid R, Dorogházi E, Rákhely G: **The hydrogenases of *Thiocapsa roseopersicina*.** *Biochem Soc Trans* 2005, **33**:61–63.
55. Burgdorf T, Lenz O, Buhrke T, van der Linden E, Jones A, Albracht S, Friedrich B: **[NiFe]-hydrogenases of *Ralstonia eutropha* H16: Modular enzymes for oxygen-tolerant biological hydrogen oxidation.** *J Mol Microb Biotech* 2005, **10**:181–196.
56. Vignais PM, Billoud B: **Occurrence, classification, and biological function of hydrogenases: an overview.** *Chem Rev* 2007, **107**:4206–4272.
57. Maroti J, Farkas A, Nagy IK, Maroti G, Kondorosi E, Rákhely G, Kovacs KL: **A second soluble hox-type nife enzyme completes the hydrogenase set in *Thiocapsa roseopersicina* BBS.** *Appl Environ Microbiol* 2010, **76**:5113–5123.
58. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, Hourdez S, Girguis PR, Wankel SD, Barbe V, Pelletier E, Fink D, Borowski C, Bach W, Dubilier N: **Hydrogen is an energy source for hydrothermal vent symbioses.** *Nature* 2011, **476**:176–180.
59. Bogachev AV, Verkhovskiy MI: **Na<sup>+</sup>-translocating NADH: quinone oxidoreductase: progress achieved and prospects of investigations.** *Biochem (Moscow)* 2005, **70**:143–149.
60. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS.** *Nature* 2000, **407**:81–86.
61. Pickering BS, Oresnik IJ: **Formate-dependent autotrophic growth in *Sinorhizobium meliloti*.** *J Bacteriol* 2008, **190**:6409.
62. Benoit S, Abaibou H, Mandrand-Berthelot M-A: **Topological analysis of the aerobic membrane-bound formate dehydrogenase of *Escherichia coli*.** *J Bacteriol* 1998, **180**:6625.
63. Preisig O, Zufferey R, Thony-Meyer L, Appleby C, Hennecke H: **A high-affinity *cbb*<sub>3</sub>-type cytochrome oxidase terminates the symbiosis-specific respiratory chain of *Bradyrhizobium japonicum*.** *J Bacteriol* 1996, **178**:1532.
64. Pitcher RS, Watmough NJ: **The bacterial cytochrome *cbb*<sub>3</sub> oxidases.** *Biochim Biophys Acta Bioenerg* 2004, **1655**:388–399.
65. Nunoura T, Sako Y, Wakagi T, Uchida A: **Regulation of the aerobic respiratory chain in the facultatively aerobic and hyperthermophilic archaeon *Pyrobaculum oguniense*.** *Microbiol (Reading, Engl)* 2003, **149**:673–688.
66. Otten MF, Stork DM, Reijnders WN, Westerhoff HV, Van Spanning RJ: **Regulation of expression of terminal oxidases in *Paracoccus denitrificans*.** *Eur J Biochem* 2001, **268**:2486–2497.
67. Krueger DM, Roeselers G, Sigman D, Cavanaugh CM: **Nitrogen nutrition in the symbiosis *Solemya velum* in preparation.**
68. Potter LC, Millington P, Griffiths L, Thomas GH, Cole JA: **Competition between *Escherichia coli* strains expressing either a periplasmic or a membrane-bound nitrate reductase: does Nap confer a selective advantage during nitrate-limited growth?** *Biochem J* 1999, **344**(Pt 1):77–84.
69. Zemmelink H, Houghton L, Sievert S, Frew N, Dacey J: **Gradients in dimethylsulfide, dimethylsulfoniopropionate, dimethylsulfoxide, and bacteria near the sea surface.** *Mar Ecol-Prog Ser* 2005, **295**:33–42.
70. Mussmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jorgensen BB, Hunte mann M, Gloeckner FO, Amann R, Koopman WJH, Lasken RS, Janto B, Hogg J, Stoodley P, Boissy R, Ehrlich GD: **Insights into the genome of large sulfur bacteria revealed by analysis of single filaments.** *PLoS Biol* 2007, **5**:1923–1937.
71. McCrindle SL, Kappler U, McEwan AG: **Microbial dimethylsulfoxide and trimethylamine-N-oxide respiration.** *Adv Microb Physiol* 2005, **50**:147–198.
72. Häse CC, Fedorova ND, Galperin MY, Dibrov PA: **Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons.** *Microbiol Mol Biol Rev* 2001, **65**:353–370. table of contents.
73. Mulikidjanian 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.
74. Robinson J, Cavanaugh CM: **Expression of form I and form II Rubisco in chemoautotrophic symbioses: implications for the interpretation of stable carbon isotope values.** *Limnol Oceanogr* 1995, **40**:1496–1502.
75. Reshetnikov AS, Rozova ON, Khmelelina 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.
76. 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.
77. 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.** *Proc Natl Acad Sci* 2012, **109**:E1173–E1182.
78. Bassham J, Benson A, Calvin M: **The path of carbon in photosynthesis.** *J Biol Chem* 1950, **185**:781–787.
79. Fenton A, Parichartanakul N, Reinhart G: **Identification of substrate contact residues important for the allosteric regulation of phosphofructokinase from *Escherichia coli*.** *Biochemistry* 2003, **42**:6453–6459.
80. 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*.** *Infect Immun* 2010, **78**:5223–5232.
81. 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.
82. 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, **154**:3073–3083.
83. 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.
84. 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 *Ridgella piscesae*.** *Mol Ecol* 2013.
85. Lee RW, Childress JJ: **Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and methanotrophic symbionts.** *Appl Environ Microb* 1994, **60**:1852–1858.
86. 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.** *Geochem Geophys Geosyst* 2012, **13**:1–23.
87. Hentschel U, Felbeck H: **Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*.** *Nature* 1993, **366**:338–340.
88. 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.
89. Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospeshil M, Felbeck H, Zal F: Fate of nitrate acquired by the tubeworm *Riftia pachytila*. *Appl Environ Microbiol* 2000, **66**:2783–2790.
90. 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.
91. 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 pachytila*. *J Exp Biol* 2000, **203**:2907–2920.
92. 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.
93. 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.
94. Eichhorn E, van der Ploeg JR, Kertesz MA, Leisinger T: Characterization of alpha-ketoglutarate-dependent taurine dioxigenase from *Escherichia coli*. *J Biol Chem* 1997, **272**:23031–23036.
95. Conway N, McDowell Capuzzo J: Incorporation and utilization of bacterial lipids in the *Solemya velum* symbiosis. *Mar Biol* 1991, **108**:277–291.
96. 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.
97. Moran N, McCutcheon J, Nakabachi A: Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 2008, **42**:165–190.
98. Nussbaumer AD, Fisher CR, Bright M: Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* 2006, **441**:345–348.
99. Cavanaugh CM: Symbiosis of chemoautotrophic bacteria and marine invertebrates. In *PhD Thesis*. Cambridge, MA, USA: Harvard University, Department of Organismic and Evolutionary Biology; 1985.
100. Fisher C, Childress J: Organic carbon transfer from methanotrophic symbionts to the host hydrocarbon-seep mussel. *Symbiosis* 1992, **12**:221–235.
101. 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.
102. 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.
103. Paa AS, Bloch CB, Brill WJ: Developmental fate of *Rhizobium meliloti* bacteroids in alfalfa nodules. *J Bacteriol* 1980, **143**:1480–1490.
104. 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.
105. 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.
106. Newton ILG, Bordenstein SR: Correlations between bacterial ecology and mobile DNA. *Curr Microbiol* 2011, **62**:198–208.
107. Plague GR, Dunbar HE, Tran PL, Moran NA: Extensive proliferation of transposable elements in heritable bacterial symbionts. *J Bacteriol* 2008, **190**:777–779.
108. 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.
109. 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. *Mol Biol Evol* 2008, **25**:1889–1896.
110. Chafee ME, Funk DJ, Harrison RG, Bordenstein SR: Lateral phage transfer in obligate intracellular bacteria (wolbachia): verification from natural populations. *Mol Biol Evol* 2010, **27**:501–505.
111. Roeselers G, Newton ILG: On the evolutionary ecology of symbioses between chemosynthetic bacteria and bivalves. *Appl Microbiol Biotechnol* 2012, **94**:1–10.
112. Gil R, Sabater-Muñoz B, Latorre A, Silva FJ, Moya A: Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc Natl Acad Sci U S A* 2002, **99**:4454–4458.
113. Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA: Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol* 2006, **4**:1079–1092.
114. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, et al: Real-time DNA sequencing from single polymerase molecules. *Science* 2009, **323**:133–138.
115. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM: Hybrid error correction and de novo assembly of single-molecule sequencing reads. *Nat Biotechnol* 2012.
116. *Pacific Biosciences*. [http://www.pacb.com]
117. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA: Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE* 2012, **7**:e47768.
118. Delcher AL, Bratke KA, Powers EC, Salzberg SL: Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 2007, **23**:673–679.
119. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ: Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010, **11**:1–11.
120. Besemer J, Lomsadze A, Borodovsky M: GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 2001, **29**:2607.
121. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH: UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 2007, **23**:1282–1288.
122. Tatusova T, Ciufu S, Fedorov B, O'Neill K, Tolstoy I: RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res* 2014, **42**:D553–D559.
123. Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, Mavrommatis K, Meyer F: The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* 2012, **13**:1–5.
124. Standard operating procedure for the annotations of genomes and metagenomes submitted to the integrated microbial genomes expert review (IMG-ER) system. [http://img.jgi.doe.gov/w/doc/img\_er\_ann.pdf]
125. Gao F, Zhang C-T: Ori-Finder: A web-based system for finding oriCs in unannotated bacterial genomes. *BMC Bioinformatics* 2009, **9**:1–6.
126. *Bprome*. [http://www.softberry.com]
127. Emanuelsson O, Brunak S, von Heijne G, Nielsen H: Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2007, **2**:953–971.
128. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**:2947–2948.
129. Robinson DG, Lee M-C, Marx CJ: OASIS: an automated program for global investigation of bacterial and archaeal insertion sequences. *Nucleic Acids Res* 2012, **40**:e174.
130. 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 resource for integrative and conjugative elements found in Bacteria. *Nucleic Acids Res* 2012, **40**(Database issue):D621–D626.
131. Leplae R, Lima-Mendez G, Toussaint A: ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* 2010, **38**(Database issue):D57–D61.

doi:10.1186/1471-2164-15-924

Cite this article as: Dmytrenko et al.: The genome of the intracellular bacterium of the coastal bivalve, *Solemya velum*: a blueprint for thriving in and out of symbiosis. *BMC Genomics* 2014 **15**:924.