

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

AUTHORS: Oleg Dmytrenko¹, Shelbi L. Russell¹, Wesley T. Loo¹, Kristina M. Fontanez², Li Liao³, Guus Roeselers⁴, Raghav Sharma⁴, Frank J. Stewart⁵, Irene L. G. Newton⁶, Tanja Woyke⁷, Dongying Wu⁸, Jenna Morgan Lang⁸, 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, Davis, CA 94616-8816

** To whom correspondence may be addressed. Colleen Cavanaugh, Harvard: cavanaugh@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.

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

3

4 Oleg Dmytrenko¹, Shelbi L. Russell¹, Wesley T. Loo¹, Kristina M. Fontanez², Li Liao³,
5 Guus Roeselers⁴, Raghav Sharma⁵, Frank J. Stewart⁵, Irene L.G. Newton⁶, Tanja
6 Woyke⁷, Dongying Wu⁸, Jenna Morgan Lang⁸, Jonathan A. Eisen^{8*}, Colleen M.
7 Cavanaugh^{1*}

8

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

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

13 ³SOA Key Laboratory for Polar Science, Polar Research Institute of China, Shanghai,
14 200136, China

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

17 ⁵School of Biology, Georgia Institute of Technology, □Atlanta GA 30332-0230, USA

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

20 ⁷DOE Joint Genome Institute, 2800 Mitchell Drive □Walnut Creek, CA 94598, USA

21 ⁸UC Davis Genome Center, 451 East Health Sciences Drive □Davis, CA 95616-8816,
22 USA

23

24 *Corresponding authors

25

26 Email addresses:

27 OD: odmytren@fas.harvard.edu

28 SLR: shelbirussell@fas.harvard.edu

29 WTL: wesleyloo@fas.harvard.edu

30 KMF: fontanez@mit.edu

31 LL: liaoli@pric.gov.cn

32 GR: guus.roesellers@tno.nl

33 FJS: frank.stewart@biology.gatech.edu

34 RS: rsharma61@gatech.edu

35 ILGN: irnewton@indiana.edu

36 TW: twoyke@lbl.gov

37 DW: dygwu@ucdavis.edu

38 JML: jnamorgan@ucd.edu

39 JAE: jaeisen@ucdavis.edu

40 CMC: cavanaugh@fas.harvard.edu

41

42

43

44

45

46

47

48

49

50

51

52 **Abstract**

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

64

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

73

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

77

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

81

82 **Background**

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

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

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

104 where it builds Y-shaped burrows that span the oxic-anoxic interface, allowing access to
105 both reduced inorganic sulfur as an energy source and oxygen for use as a terminal
106 oxidant [10]. The symbionts, which constitute a single 16S rRNA phylotype of γ -
107 proteobacteria [11], are localized to specialized epithelial cells (bacteriocytes) in the gills,
108 separated from the cytoplasm by a peribacterial membrane. Using energy from the
109 oxidation of sulfide, the symbionts fix CO₂ via the Calvin-Benson-Bassham Cycle [12,
110 13]. Primary production in the symbionts sustains the host, which has only a rudimentary
111 gut and cannot effectively filter-feed [14, 15]. Many key properties of this symbiosis still
112 remain to be characterized, including the exchange of metabolites and signals between
113 the symbiont and the host and the mechanism of symbiont acquisition at each new host
114 generation (i.e., symbiont transmission mode).

115 The mode by which *S. velum* acquires its symbionts has important implications for
116 understanding symbiont genome evolution. Symbiont-specific genes have been
117 amplified from the host ovarian tissue of both *S. velum* and its congener, *S. reidi* [16,
118 17], raising the hypothesis that symbionts are transmitted maternally (vertically) between
119 successive host generations via the egg. Vertical transmission has also been inferred in
120 deep-sea clams of the Vesicomidae [18, 19], in which symbionts have a reduced
121 genome size (1.2 Mb) and appear to be obligately associated with their host [20-23]. In
122 vesicomid symbioses, host and symbiont phylogenies are largely congruent, a pattern
123 consistent with vertical symbiont transmission [24]. Nonetheless, instances of lateral
124 symbiont movement among some vesicomids have been inferred based on decoupling
125 of symbiont and host evolutionary trajectories [25], bringing diverse symbiont strains into
126 contact and creating opportunities for symbiont genome evolution via recombination [26,
127 27]. In the Solemyidae, on the other hand, symbionts of different *Solemya* species are
128 scattered across phylogenetic clades (i.e., polyphyly), indicating distinct evolutionary
129 origins relative to the monophyly of the hosts [5, 28]. A preliminary analysis was unable

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

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

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

156 *Cryosomollon squamiferum*, [33] and the marine oligochaete worm, *Olavius algarvensis*,
157 [34]. This comprehensive analysis defines the *S. velum* symbiont as a metabolically
158 versatile bacterium adapted to living inside the host but also potentially capable of
159 survival on the outside. It informs attempts to culture the symbionts and generates
160 multiple intriguing hypotheses that now await experimental validation.

161

162 **Results and discussion**

163 **General genome features**

164 The genome of the *S. velum* symbiont consists of 10 non-overlapping scaffolds,
165 totaling 2,702,453 bp, with an average G + C content of 51%. The three largest scaffolds
166 (1.21 Mb, 0.89 Mb, 0.54 Mb) contain 97.8% of the total genomic sequence and 98.4% of
167 the predicted genes (Table S1 in Additional file 1). Assembly of the scaffolds into a
168 closed genome was prevented by stretches of single nucleotides or groups of a few
169 nucleotides repeated up to 70 times that could not be spanned. However, the high depth
170 of sequence coverage and the presence of all 31 core bacterial phylogenetic gene
171 markers [35] suggest that most gene-coding regions were detected in the analysis.
172 Nevertheless, as the genome is not closed, a definitive list of all symbiont genes could
173 not be made.

174 An overview of the *S. velum* symbiont genome compared to selected symbiotic
175 and free-living γ -proteobacteria, including other thiotrophs, is presented in Table 1.
176 Briefly, 90.7% of the genome sequence encodes 2,757 genes, on average 885 bp long.
177 2,716 (98.5%) genes are protein-coding. Function was predicted for 1,988 (72.1%) of all
178 the genes, while 769 (27.9%) were identified as encoding hypothetical proteins. 382
179 genes (13.8%) have one or more paralog in the genome, with the largest paralogous
180 group encoding transposases associated with mobile elements. The genome contains a
181 single ribosomal RNA (rRNA) operon and 38 transfer RNAs (tRNA) corresponding to the

182 20 standard proteinogenic amino acids. Due to the wobble base-pairing [36], tRNAs for
183 each given amino acid can pair with any codon in the genome for that amino acid (Table
184 S2).

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

198

199 **Metabolic Functions**

200 **1. Chemolithotrophy**

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

208 reduced sulfur species are most closely related to those of the purple sulfur γ -
209 proteobacterium, *Allochromatium vinosum* (Figure 4), in which the genetic components
210 and the biochemical mechanisms of sulfur metabolism have been well characterized
211 [41].

212 **Periplasmic sulfide and thiosulfate oxidation:** In the periplasm of the *S. velum*
213 symbiont, sulfide, thiosulfate, and, possibly, elemental sulfur, may be oxidized for energy
214 by the Sox system, which is represented in the genome (Figure 4). The identified
215 SoxYZAXB, flavocytochrome *c* dehydrogenase (FccAB), and type I and IV sulfide-
216 quinone reductases (Sqr) potentially reduce cytochromes *c* and quinones, which along
217 the course of the electron-transport chain translate into membrane-ion gradients, NADH,
218 and ATP, ultimately fueling biosynthetic and other energy-requiring cellular processes,
219 including autotrophy (Figure 1). In *A. vinosum* and the green non-sulfur bacterium,
220 *Chlorobium tepidum*, SoxYZ, SoxAX, and SoxB proteins participate in the formation of
221 transient sulfur deposits as intermediates during sulfur oxidation [42]. In fact, sulfur
222 deposits are common to all known sulfur-oxidizing bacteria (SOBs) which, like the *S.*
223 *velum* symbiont, lack SoxCD sulfur dehydrogenase (Figure 4) [43], including the
224 symbionts of the hydrothermal vent tubeworm, *R. pachyptila* [31, 44], and the clam, *C.*
225 *magnifica* [22, 45]. Microscopically-detectable intracellular or extracellular sulfur has not
226 been observed either in the symbiont-containing gills of *S. velum* or directly within the
227 symbionts (Cavanaugh, unpublished observation). Absence of sulfur deposits may be
228 attributed to a very rapid consumption of any available reduced sulfur substrate. This
229 agrees with the fact that the *S. velum* symbionts have the highest known carbon fixation
230 rate, and, hence, demand for energy, of all the studied chemosynthetic symbionts, i.e.,
231 $65 \mu\text{mol min}^{-1} \text{g of protein}^{-1}$ [13] compared to $0.45 \mu\text{mol min}^{-1} \text{g of protein}^{-1}$ of the next
232 highest rate measured in the symbionts of *R. pachyptila* [46]. Alternatively, in the *S.*
233 *velum* symbionts intermediate sulfur may be stored in a chemical form that is not easily

234 observed microscopically.

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

247 **Sulfite oxidation:** Sulfite generated by rDsr may be further oxidized to sulfate in
248 the cytoplasm by a sequential action of APS reductase (AprABM) and an ATP-
249 generating ATP sulfurylase (Sat) (Figure 1 and Figure 4). Identification of these genes
250 agrees with measured Apr and Sat activity in the symbiont-containing *S. velum* tissue
251 [50]. Sulfate generated in this pathway may be exported from the cytoplasm via a
252 sulfate-bicarbonate antiporter SulP (Figure 1). While electrons obtained from the
253 oxidation of sulfide, thiosulfate, and, possibly, elemental sulfur by Sox and rDsr are
254 shuttled into the electron transport chain, energy obtained from the oxidation of sulfite is
255 immediately available in the form of ATP.

256 **Bioenergetics**

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

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

268 **Rnf complexes:** The versatility of the respiratory electron transport chain in the *S.*
269 *velum* symbiont is indicated by the presence of genes supporting electron donors like
270 ferredoxins, which have a redox potential as negative as -500mV [51], compared, for
271 example, to -400 mV of S₂O₃²⁻ and -270mV of H₂S. The reversible oxidation of
272 ferredoxins coupled to reduction of NAD⁺ in the *S. velum* symbiont may be catalyzed by
273 the H⁺ or/and Na⁺-motive Rnf complexes (Figure 1) encoded in the genome by two
274 complete *rnfABCDGE* (*rnf1*) and *rnfBCDGEA* (*rnf2*) operons. The organization of genes
275 in the operons is conserved with other bacteria, suggesting that these clusters did not
276 arise from duplication. Previously, only *Axotobacter vinelandii* and *Desulfobacterium*
277 *autotrophicum* were known to harbor two *rnf* operons [51]. Based on the presence of
278 genes for pyruvate:ferredoxin oxidoreductase located between *rnfB2* and *rnfC2*,
279 pyruvate may serve as an electron donor for at least one of the Rnf complexes. In
280 general, *rnf* genes are distributed mainly among obligate and facultative anaerobes,
281 including many pathogens that colonize oxygen-limited host tissues [51]. Together with
282 the fact that ferredoxins play a key role in anaerobic metabolism [52], this suggests that
283 the *S. velum* symbiont, as well as other sequenced chemosynthetic symbionts, which all
284 contain *rnf* genes, may be capable of facultative anaerobiosis.

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

286 that the *S. velum* symbiont may harness for the reduction of quinone and NAD⁺ cellular
287 pool (Figure 1). Hydrogen oxidation is suggested by the presence in the symbiont
288 genome of *hup* and *hox2* operons encoding an uptake and a bidirectional hydrogenase,
289 respectively. The two subunits of the symbiont [Ni-Fe]-uptake hydrogenase, HupSL, are
290 most similar in amino acid sequence to HupS and HupL proteins from the symbionts of
291 the tubeworms, *R. pachyptila* and the *T. jerichona*, (73% and 78% identity for the S and
292 L subunits respectively), the sulfur bacterium, *Thiocapsa roseopersicina*, (68 and 74%),
293 and the symbionts of the scaly-foot snail, *Crysmallon squamiferum*, (50 and 53%). In *T.*
294 *roseopersicina*, HupSL has been experimentally demonstrated to reduce quinones of the
295 respiratory chain with H₂ [53, 54]. Unlike all the other γ -proteobacteria containing HupSL,
296 the *hup* operon in the *S. velum* symbiont does not encode the di-heme cytochrome *b*,
297 which is necessary to link H₂ oxidation to quinone reduction in the cellular membrane
298 [55]. However, a [Ni-Fe] hydrogenase cytochrome *b* homolog was found on a different
299 genomic scaffold. Though this discordant gene organization is unlike that in other H₂
300 oxidizers, it is possible that the identified cytochrome *b* may act in tandem with HupSL to
301 enable H₂ oxidation.

302 Apart from potentially reducing the respiratory quinone pool with H₂, the symbiont,
303 by means of a bidirectional hydrogenase, may produce H₂ by oxidizing NAD⁺. The *S.*
304 *velum* symbiont Hox2FUYH enzyme complex is most similar in amino acid sequence
305 (63-66%) to the recently-characterized NAD⁺-reducing [Ni-Fe]-hydrogenase from *T.*
306 *roseopersicina*, which can operate in reverse, generating H₂ when the high reduction
307 state of the dinucleotide pool is growth-limiting [56]. As H₂ concentrations available to the
308 *S. velum* symbiont have not been measured, it is unknown whether H₂ oxidation
309 contributes to primary production to the degree that has been recently demonstrated in
310 hydrothermal vent symbiosis [57].

311 **Primary ion pumps:** NADH (-320 mV), potentially derived from oxidation of H₂ or

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

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

334 The genome-encoded quinol-cytochrome-c oxidoreductase (*bc₁*, Complex III
335 homologue) potentially links oxidation of quinols to the generation of a proton membrane
336 gradient and the reduction of terminal electron acceptors (Figure 1), discussed next.

337 **Terminal oxygen reductases:** Similar to most aerobic and microaerophilic

338 bacteria, the genome of the *S. velum* symbiont encodes three types of H⁺-motive
339 terminal oxygen reductases (Figure 1), which suggest a capacity to respire O₂ over a
340 wide range of concentrations. The genome contains a *ccoNOQP* operon encoding a
341 *cbb₃* cytochrome oxidase, which is known to function at nanomolar O₂ concentrations in
342 the nitrogen-fixing plant symbionts, *Bradyrhizobium japonicum* [62], and in the
343 microaerophilic human pathogens, *Campylobacter jejuni*, *Helicobacter pylori*, and
344 *Neisseria meningitidis* [63]. The genome also encodes a *aa₃* cytochrome oxidase
345 (CoxAB), which is thought to function primarily under atmospheric oxygen
346 concentrations [64] and is the only terminal oxidase in the symbionts of the bivalves *C.*
347 *magnifica* [22] and *C. okutani* [20]. The third terminal oxidase identified in the symbiont
348 genome is a *cydAB*-encoded quinol oxidase, which is thought to oxidize quinols instead
349 of cytochromes. CydAB may operate when an excess of reductants, potentially coming
350 from the host, limits metabolic turnover and redox balance needs to be achieved [65].
351 The observed diversity of terminal oxygen reductases indicates that the supply of
352 oxygen to the symbionts fluctuates over time or between free-living and symbiotic
353 stages, necessitating adjustments in respiratory metabolism.

354 **Alternative terminal reductases:** When oxygen is limited or unavailable,
355 potentially either through competition for oxygen with the host or if the symbionts find
356 themselves in the anoxic sediment that surrounds the burrow, the *S. velum* symbiont
357 may be capable of using terminal electron acceptors other than oxygen. Although it is
358 unknown whether the symbiont-containing gill bacteriocytes ever become anaerobic, the
359 presence of genes for periplasmic NO₃⁻ reductase (*napFDAGHBC*) suggests that
360 symbiont energy generation may involve electron transfer to nitrate, which is available in
361 the porewater surrounding *S. velum* at concentrations of ~1-10 μM ([66], in preparation).
362 The structure of the symbiont *napFDAGHBC* operon is consistent with that of enteric
363 bacteria that are thought to use Nap for effectively scavenging nitrate during anaerobic

364 growth under nitrate-limited conditions (5 μ M) [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 helices 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**

390 Autotrophic carbon fixation, fueled chiefly by sulfur oxidation, is the principal
391 process in the *S. velum* symbiont, supplying both the symbionts and the host with
392 organic carbon [14]. While previous studies focused primarily on RubisCO [10, 73], the
393 key enzyme of the Calvin cycle for CO₂ fixation and the most highly expressed gene in
394 the symbiont [40], our current analysis identified genes that encode catalytic
395 components required for CO₂ fixation and storage, including the pyrophosphate-
396 dependent phosphofructokinase, which has been hypothesized to command a more
397 energy efficient variant of the cycle [22, 74-76]. Furthermore, the genome of the *S.*
398 *velum* symbiont contains the gene for α-ketoglutarate dehydrogenase – the key enzyme
399 of the tricarboxylic acid cycle (TCA), suggesting that the symbiont can respire organic
400 carbon and may not be obligately autotrophic.

401 **Autotrophy:** The genome of the *S. velum* symbiont encodes a modified Calvin
402 cycle that appears unique to chemosynthetic symbionts and some methanotrophic
403 bacteria in that it lacks fructose 1,6-bisphosphatase and sedoheptulose 1,7-
404 bisphosphatase. This alternative carbon fixation pathway may rely instead on a
405 reversible pyrophosphate-dependent phosphofructokinase (PPi-PFK) (Figure 1). The
406 ability of this enzyme to dephosphorylate fructose 1,6-phosphate and sedoheptulose
407 1,7-phosphate *in vitro* has been previously demonstrated for the PPi-PFK from
408 *Methylococcus capsulatus* [74], which shares 73% sequence identity with the *S. velum*
409 symbiont homologue. Compared to the classical Calvin cycle [77], the use of PPi-PFK
410 can offer energy savings up to 9.25% [76]. While fructose 1,6-bisphosphatase and
411 sedoheptulose 1,7-bisphosphatase enzymes release phosphate ions that cannot be
412 used for energy gain, PPi-PFK produces energy-rich pyrophosphate. In *M. capsulatus*
413 [74] and in the chemosynthetic symbionts of *R. pachyptila* [75] and the oligochete, *O.*
414 *algarvensis* [76] it was proposed that the pyrophosphate could be converted into an
415 electrochemical gradient by the membrane-bound proton-pumping pyrophosphatase

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

424 **Carbon Flux:** Carbon fixed by the *S. velum* symbiont may be stored as
425 polyglucose or fed into catabolic and anabolic reactions (Figure 1). The overall direction
426 of the metabolic carbon flux in the symbionts can be controlled by at least two putative
427 mechanisms. First, the reversible PPI-PFK, participating in the Calvin cycle as discussed
428 above, may also phosphorylate fructose 6-bisphosphate during glycolysis. PPI-PFK
429 appears to be the only enzyme encoded in the genome that could catalyze both the
430 forward and the reverse reactions. The directionality of the catalysis may depend on the
431 concentration of pyrophosphate in the cytoplasm [78], since this PPI-PFK is nonallosteric
432 [74]. Second, the two glyceraldehyde 3-phosphate dehydrogenases, GapA and GapB,
433 may be specific to glycolysis and the Calvin cycle/gluconeogenesis, respectively, by
434 analogy to the homologous enzymes in *Staphylococcus aureus* [79]. In the symbiont
435 genome, *gapB* is adjacent to the gene for transketolase, an enzyme in the Calvin cycle,
436 further suggesting that these two Gap proteins may play a role in regulating the direction
437 of the carbon flux either in the direction of glycolysis or Calvin cycle and
438 gluconeogenesis. The symbionts of *C. magnifica*, *C. okutanii*, *R. pachyptila*, *T. jerichona*,
439 and the scaly snail possess just a single *gap* gene, which has a much higher amino acid
440 sequence identity to *gapB* than to *gapA* from the *S. velum* symbiont. The above
441 evidence suggests that among the chemosynthetic symbionts, the symbiont of *S. velum*

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

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

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

459 **3. Nitrogen metabolism**

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

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

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

482 **Urea metabolism:** Host urea may serve as an additional source of assimilatory
483 nitrogen for the *S. velum* symbionts. The identified *ureHABCEFG* operon encodes a
484 cytoplasmic urease UreABC, which can hydrolyze urea, releasing ammonia that may be
485 re-utilized by symbionts. Urea can enter the bacterial cell by passive diffusion [89], but
486 under nitrogen starvation the symbionts may be able to take it up more rapidly via an
487 ABC-transporter UrtABCDE, encoded directly upstream of the *ure* genes. Among
488 chemosynthetic symbionts, urease genes have been previously described only in the γ -
489 symbionts from the marine oligochaete worm *O. algarvensis* [34, 76], which, like *S.*
490 *velum*, lives in coastal sediments. The sequenced chemosynthetic symbionts from
491 hydrothermal vents lack urease genes, even though some of their host organisms, for
492 instance *R. pachyptila* [90], are known to produce urea. This discrepancy may be
493 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 ($\delta^{13}\text{C}$, $\delta^{14}\text{N}$, $\delta^{34}\text{S}$) 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

520 fatty-acids, the symbionts may build components of their plasma membrane mostly from
521 *cis*-vaccenic acid (18 : ω 7) (Figure 1). According to a previous analysis of lipid
522 composition in *S. velum* [94], this unsaturated fatty acid and its derivatives are the main
523 constituents of cellular membranes in the symbionts and the host alike. Furthermore, the
524 isotopic signature of the host's lipids indicates that they are bacterial in origin [94].
525 Second, the identified genes for the synthesis of lipopolysaccharides (Figure 1) suggest
526 that the symbiont may be able to assemble the LPS structures that are known to be
527 sufficient for growth of *E. coli* [95]. Most intracellular symbionts that live within a host
528 derived membrane, like the *S. velum* symbiont [10], lack LPS biosynthetic genes and are
529 unable to replicate on their own [96]. However, the symbionts which have the genes to
530 synthesize LPS tend to either live directly in the cytoplasm [96] and, have to make their
531 own cellular envelope, or, like the symbionts of *R. pachyptila* [97], exist extracellularly for
532 part of their life. Therefore, the symbiont of *S. velum* may not only be able to make a fully
533 functional cellular envelope and supply some of its components to its host, but may also
534 be capable of living outside the bacteriocytes.

535 **2. Membrane transport**

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

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

551 **Multi-drug efflux pumps:** The *S. velum* genome contains at least five sets of
552 genes encoding multi-drug efflux pumps (AcrAB-TolC), suggesting the ability to expel
553 host-derived antimicrobial agents. A comparable genetic capacity for the AcrAB-TolC
554 efflux system has been found in bacteria, such as the plant symbiont *Rhizobium*
555 *leguminosarum*, that have a free-living stage, but not in bacteria that are obligately
556 intracellular (Table 2). The plant host of *R. leguminosarum* manipulates the cellular fate
557 of its symbionts using antimicrobial-like peptide factors [101]. As a result, *R.*
558 *leguminosarum* undergoes cell elongation and genome replication but loses its ability to
559 divide. Only a small number of *R. leguminosarum* cells remain vegetative [102]. A very
560 similar morphological differentiation of the symbionts has been observed in *S. velum*
561 [103]. Assuming the bivalve host also uses peptide factors to control its symbionts, the
562 *S. velum* symbiont may rely on the efflux pumps to maintain a small, undifferentiated
563 population in the bacteriocytes for transmission to future host generations.

564 **3. Sensory mechanisms and motility**

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

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

582

583 **Mobile genetic elements**

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

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

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

613

614 **Conclusions**

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

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

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

640

641 **Materials and methods**

642 ***Specimen collection and DNA preparation:*** *S. velum* individuals were collected
643 by the staff of the Marine Resource Center of the Marine Biological Laboratory (MBL),
644 Woods Hole, MA from reducing sediment of shallow eelgrass beds near Naushon Island,
645 Woods Hole, MA in 2006, 2007, and 2012. The collection was performed in accordance
646 with state collecting permit issued by the Division of Marine Fisheries and in compliance
647 with all local, regional and federal regulations, including the Convention on Biological
648 Diversity and the Convention on the Trade in Endangered Species of Wild Fauna and
649 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 4°C. 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 37°C 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

702 bit score match to the query were retained for annotation. Genes predicted by two or
703 more methods (redundant) were considered the same and collapsed into a single entry if
704 they shared the same start and stop position, orientation, and similar functional
705 annotations. Non-redundant entries (i.e., gene predictions unique to a given software)
706 were also retained. Finally, the above predictions and annotations were reconciled with
707 the genes predicted and annotated through the Integrated Microbial Genomes Expert
708 Review (IMG-ER) pipeline [122]. Selected origins of replication were verified by Ori-
709 Finder [67]. The genes in the genome was assigned taxa in the NCBI taxonomy based
710 on the BLASTN [38] searches (-best_hit_overhang 0.25, -best_hit_score_edge 0.05, -
711 evalue 0.0001) against the NCBI-nr database (8 July 2014) computed with MEGAN
712 5.4.3 (maximum number of matches per read 100; LCA parameters: minimal support 5,
713 minimal score 35, top percent 10) [39]. Selected promoters were identified with BPROM
714 [123]. Signal peptides and transmembrane domains were predicted using SignalP 3.0
715 Server and TMHMM, respectively [124]. The Genomic Utility for Automated Comparison
716 (GUAC) Python script (Additional file 5) was developed to inform comparative analyses
717 of gene content across multiple genomes, in particular genes involved in sulfur-oxidation
718 (Figure 4). The GUAC software first identified those target genes in the genomes of
719 interest that were annotated with unambiguous gene symbols (e.g. *soxA*). Next, using
720 amino acid sequences of these genes as queries, BLASTP searched for homologous
721 sequences in the remaining target genomes (default cut-off values: bit score 50, identity
722 30%, alignment length over the source sequence 40%). These sequences were aligned
723 using ClustalW [125]. The alignments were used to manually verify the results (e.g.,
724 based on known conserved domains, etc.). Mobile genetic elements were detected by
725 type. Insertion sequences were found using OASIS [126]. Integrative conjugative
726 elements and plasmid as well as phage sequences were identified by BLASTN [38]
727 searches against the ICEberg [127] and ACLAME [128] databases, respectively (cut-off

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

732

733 **Availability of supporting data**

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

736

737 **Competing interests**

738 The authors declare that they have no competing interests.

739

740 **Authors' contributions**

741

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

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

754

755 **Acknowledgments**

756 This work was funded by grant 0412205 of the US National Science Foundation
757 (NSF) and was made possible with the generous support of the U.S. Department of
758 Energy Joint Genome Institute (JGI). The work conducted by JGI was supported by the
759 Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-
760 05CH11231. We would like to express special thanks to Grace Pai for creating Sanger
761 sequencing libraries and Shannon Smith and Terry Utterback for coordinating
762 sequencing at TIGR.

763

764 **References**

- 765 1. Sagan L: **On the origin of mitosing cells.** *J Theor Biol* 1967, **14**:225–274.
- 766 2. Gonzalez A, Clemente JC, Shade A, Metcalf JL, Song S, Prithiviraj B, Palmer BE, Knight R: **Our microbial selves:
767 what ecology can teach us.** *EMBO Rep* 2011, **12**:775–784.
- 768 3. Dilworth MJ, James EK, Sprent JI: *Nitrogen-Fixing Leguminous Symbioses.* Kluwer Academic Pub; 2008.
- 769 4. Clark EL, Karley AJ, Hubbard SF: **Insect endosymbionts: manipulators of insect herbivore trophic interactions?**
770 *Protospasma* 2010, **244**:25–51.
- 771 5. Cavanaugh CM, McKiness Z, Newton I, Stewart FJ: **Marine Chemosynthetic Symbioses.** In *The Prokaryotes -
772 Prokaryotic Biology and Symbiotic Associations.* 3rd edition. Edited by Rosenberg E; 2013:579–607.
- 773 6. Toft C, Andersson SGE: **Evolutionary microbial genomics: insights into bacterial host adaptation.** *Nat Rev Genet*
774 2010, **11**:465–475.
- 775 7. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald
776 BR, Moran NA, Bristow J, Cheng J-F: **One Bacterial Cell, One Complete Genome.** *PLoS ONE* 2010, **5**.
- 777 8. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U: **Single-cell genomics
778 reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges.** *ISME J* 2013,
779 **7**:2287–2300.
- 780 9. Dubilier N, Bergin C, Lott C: **Symbiotic diversity in marine animals: the art of harnessing chemosynthesis.** *Nat
781 Rev Micro* 2008, **6**:725–740.
- 782 10. Cavanaugh CM: **Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats.**
783 *Nature* 1983, **302**:58–61.
- 784 11. Eisen JA, Smith SW, Cavanaugh CM: **Phylogenetic relationships of chemoautotrophic bacterial symbionts of**

- 785 ***Solemya velum* Say (Mollusca: Bivalvia) determined by 16S rRNA gene sequence analysis.** *J Bacteriol* 1992,
786 174:3416–3421.
- 787 12. Cavanaugh CM, Abbott M, Veenhuis M: **Immunochemical localization of ribulose-1, 5-bisphosphate carboxylase**
788 **in the symbiont-containing gills of *Solemya velum* (Bivalvia: Mollusca).** *P Natl Acad Sci Usa* 1988, 85:7786–7789.
- 789 13. Scott KM, Cavanaugh CM: **CO₂ uptake and fixation by endosymbiotic chemoautotrophs from the bivalve**
790 ***Solemya velum*.** *Appl Environ Microb* 2007, 73:1174–1179.
- 791 14. Conway N, Capuzzo J, Fry B: **The role of endosymbiotic bacteria in the nutrition of *Solemya velum*: evidence**
792 **from a stable isotope analysis of endosymbionts and host.** *Limnology and Oceanography* 1989, 34:249–255.
- 793 15. Krueger DM, Gallager S, Cavanaugh CM: **Suspension feeding on phytoplankton by *Solemya velum*, a symbiont-**
794 **containing clam.** *Mar Ecol-Prog Ser* 1992, 86:145–151.
- 795 16. Cary SC: **Vertical transmission of a chemoautotrophic symbiont in the protobranch bivalve, *Solemya reidi*.** *Mol*
796 *Marine Biol Biotechnol* 1994, 3:121–130.
- 797 17. Krueger DM, Gustafson RG, Cavanaugh CM: **Vertical transmission of chemoautotrophic symbionts in the**
798 **bivalve *Solemya velum* (Bivalvia: Protobranchia).** *Biol Bull* 1996, 190:195–202.
- 799 18. Peek A, Vrijenhoek R, Gaut B: **Accelerated evolutionary rate in sulfur-oxidizing endosymbiotic bacteria**
800 **associated with the mode of symbiont transmission.** *Molecular Biology and Evolution* 1998, 15:1514.
- 801 19. Hurtado LA, Mateos M, Lutz RA, Vrijenhoek RC: **Coupling of bacterial endosymbiont and host mitochondrial**
802 **genomes in the hydrothermal vent clam *Calymptogena magnifica*.** *Appl Environ Microb* 2003, 69:2058–2064.
- 803 20. Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M, Matsuyama K, Takishita K, Kawato M, Uematsu
804 K: **Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calymptogena okutanii*.** *Curr*
805 *Biol* 2007, 17:881–886.
- 806 21. Kuwahara H, Takaki Y, Yoshida T, Shimamura S, Takishita K, Reimer JD, Kato C, Maruyama T: **Reductive genome**
807 **evolution in chemoautotrophic intracellular symbionts of deep-sea *Calymptogena* clams.** *Extremophiles* 2008,
808 12:365–374.
- 809 22. Newton I, Woyke T, Auchtung T, Dilly G, Dutton R, Fisher M, Fontanez K, Lau E, Stewart FJ, Richardson P: **The**
810 ***Calymptogena magnifica* chemoautotrophic symbiont genome.** *Science* 2007, 315:998–1000.
- 811 23. Newton I, Girguis PR, Cavanaugh CM: **Comparative genomics of vesicomyid clam (Bivalvia: Mollusca)**
812 **chemosynthetic symbionts.** *BMC Genomics* 2008, 9:585.
- 813 24. Peek A, Feldman R, Lutz R, Vrijenhoek R: **Cospeciation of chemoautotrophic bacteria and deep sea clams.** *P*
814 *Natl Acad Sci Usa* 1998, 95:9962.
- 815 25. Stewart FJ, Young CR, Cavanaugh CM: **Lateral symbiont acquisition in a maternally transmitted**
816 **chemosynthetic clam endosymbiosis.** *Molecular Biology and Evolution* 2008, 25:673–687.
- 817 26. Stewart FJ, Young C, Cavanaugh CM: **Evidence for homologous recombination in intracellular chemosynthetic**
818 **clam symbionts.** *Molecular Biology and Evolution* 2009, 26:1391–1404.
- 819 27. Stewart FJ, Baik AHY, Cavanaugh CM: **Genetic subdivision of chemosynthetic endosymbionts of *Solemya***
820 ***velum* along the Southern New England coast.** *Appl Environ Microb* 2009, 75:6005–6007.
- 821 28. Krueger DM, Cavanaugh CM: **Phylogenetic diversity of bacterial symbionts of *Solemya* hosts based on**
822 **comparative sequence analysis of 16S rRNA genes.** *Appl Environ Microb* 1997, 63:91.
- 823 29. Moran NA: **Accelerated evolution and Muller's ratchet in endosymbiotic bacteria.** *P Natl Acad Sci Usa* 1996,
824 93:2873–2878.
- 825 30. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N,
826 Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee
827 P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA:
828 **Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile**
829 **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: ***Allochrochromatium 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, Roubourdin 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 *cbb₃*-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 *cbb₃* 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. Zemelink H, Houghton L, Sievert S, Frew N, Dacey J: **Gradients in dimethylsulfide, dimethylsulfoniopropionate,**
913 **dimethylsulfoxide, and bacteria near the sea surface.** *Mar Ecol-Prog Ser* 2005, **295:33–42.**
- 914 69. Musmann 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.**

- 919 71. Häse CC, Fedorova ND, Galperin MY, Dibrov PA: **Sodium ion cycle in bacterial pathogens: evidence from cross-**
920 **genome comparisons.** *Microbiology and Molecular Biology Reviews* 2001, **65**:353–70, table of contents.
- 921 72. Mulikidjanian AY, Dibrov P, Galperin MY: **The past and present of sodium energetics: may the sodium-motive**
922 **force be with you.** *Biochim Biophys Acta* 2008, **1777**:985–992.
- 923 73. Robinson J, Cavanaugh CM: **Expression of form I and form II Rubisco in chemoautotrophic symbioses:**
924 **implications for the interpretation of stable carbon isotope values.** *Limnology and Oceanography* 1995, **40**:1496–
925 1502.
- 926 74. Reshetnikov AS, Rozova ON, Khmelenina VN, Mustakhimov II, Beschastny AP, Murrell JC, Trotsenko YA:
927 **Characterization of the pyrophosphate-dependent 6-phosphofructokinase from *Methylococcus capsulatus* Bath.**
928 *FEMS Microbiol Lett* 2008, **288**:202–210.
- 929 75. Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D, Albrecht D, Thürmer A, Daniel R, Kleiner M,
930 Hecker M, Schweder T: **Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont.**
931 *Proteomics* 2011, **11**:3106–3117.
- 932 76. Kleiner M, Wentrup C, Lott C, Teeling H, Wetzel S, Young J, Chang Y-J, Shah M, VerBerkmoes NC, Zarzycki J,
933 Fuchs G, Markert S, Hempel K, Voigt B, Becher D, Liebeke M, Lalk M, Albrecht D, Hecker M, Schweder T, Dubilier N:
934 **Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for**
935 **carbon and energy use.** *Proceedings of the National Academy of Sciences* 2012, **109**:E1173–82.
- 936 77. Bassham J, Benson A, Calvin M: **The path of carbon in photosynthesis.** *J Biol Chem* 1950, **185**:781–787.
- 937 78. Fenton A, Paricharttanakul N, Reinhart G: **Identification of substrate contact residues important for the allosteric**
938 **regulation of phosphofructokinase from *Eschericia coli*.** *Biochemistry* 2003, **42**:6453–6459.
- 939 79. Purves J, Cockayne A, Moody PCE, Morrissey JA: **Comparison of the regulation, metabolic functions, and roles**
940 **in virulence of the glyceraldehyde-3-phosphate dehydrogenase homologues gapA and gapB in *Staphylococcus***
941 ***aureus*.** *Infection and Immunity* 2010, **78**:5223–5232.
- 942 80. Wood AP, Aurikko JP, Kelly DP: **A challenge for 21st century molecular biology and biochemistry: what are the**
943 **causes of obligate autotrophy and methanotrophy?** *Fems Microbiol Rev* 2004, **28**:335–352.
- 944 81. Han SO, Inui M, Yukawa H: **Effect of carbon source availability and growth phase on expression of**
945 ***Corynebacterium glutamicum* genes involved in the tricarboxylic acid cycle and glyoxylate bypass.** *Microbiology*
946 2008.
- 947 82. Lee R, Thuesen E, Childress J: **Ammonium and free amino acids as nitrogen sources for the chemoautotrophic**
948 **symbiosis *Solemya reidi* Bernard (Bivalvia: Protobranchia).** *J Exp Mar Biol Ecol* 1992, **158**:75–91.
- 949 83. Liao L, Wankel SD, Wu M, Cavanaugh CM, Girguis PR: **Characterizing the plasticity of nitrogen metabolism by**
950 **the host and symbionts of the hydrothermal vent chemoautotrophic symbioses *Ridgeia piscesae*.** *Molecular*
951 *Ecology* 2013.
- 952 84. Lee RW, Childress JJ: **Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic**
953 **and methanotrophic symbionts.** *Appl Environ Microb* 1994, **60**:1852–1858.
- 954 85. Bourbonnais A, Lehmann MF, Butterfield DA, Juniper SK: **Subseafloor nitrogen transformations in diffuse**
955 **hydrothermal vent fluids of the Juan de Fuca Ridge evidenced by the isotopic composition of nitrate and**
956 **ammonium.** *Geochemistry, Geophysics, Geosystems* 2012, **13**:n/a–n/a.
- 957 86. Hentschel U, Felbeck H: **Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*.** *Nature* 1993,
958 **366**:338–340.
- 959 87. Lee R, Robinson J, Cavanaugh CM: **Pathways of inorganic nitrogen assimilation in chemoautotrophic bacteria-**
960 **marine invertebrate symbioses: expression of host and symbiont glutamine synthetase.** *J Exp Biol* 1999, **202** (Pt
961 **3**):289–300.
- 962 88. Girguis PR, Lee RW, Desaulniers N, Childress JJ, Pospesel M, Felbeck H, Zal F: **Fate of nitrate acquired by the**
963 **tubeworm *Riftia pachyptila*.** *Appl Environ Microb* 2000, **66**:2783–2790.
- 964 89. Beckers G, Bendt AK, Kramer R, Burkovski A: **Molecular identification of the urea uptake system and**

- 965 transcriptional analysis of urea transporter and urease-encoding genes in *Corynebacterium glutamicum*. *J*
966 *Bacteriol* 2004, **186**:7645.
- 967 90. De Cian M, Regnault M, Lallier FH: **Nitrogen metabolites and related enzymatic activities in the body fluids and**
968 **tissues of the hydrothermal vent tubeworm *Riftia pachyptila***. *J Exp Biol* 2000, **203**:2907–2920.
- 969 91. Joyner JL, Peyer SM, Lee RW: **Possible roles of sulfur-containing amino acids in a chemoautotrophic**
970 **bacterium-mollusc symbiosis**. *Biol Bull* 2003, **205**:331–338.
- 971 92. Conway N, Howes B, McDowell Capuzzo J, Turner R, Cavanaugh CM: **Characterization and site description of**
972 ***Solemya borealis* (Bivalvia; Solemyidae), another bivalve-bacteria symbiosis**. *Mar Biol* 1992, **112**:601–613.
- 973 93. Eichhorn E, van der Ploeg JR, Kertesz MA, Leisinger T: **Characterization of alpha-ketoglutarate-dependent**
974 **taurine dioxygenase from *Escherichia coli***. *J Biol Chem* 1997, **272**:23031–23036.
- 975 94. Conway N, McDowell Capuzzo J: **Incorporation and utilization of bacterial lipids in the *Solemya velum***
976 **symbiosis**. *Mar Biol* 1991, **108**:277–291.
- 977 95. Karow M, Georgopoulos C: **Isolation and characterization of the *Escherichia coli* *msbB* gene, a multicopy**
978 **suppressor of null mutations in the high-temperature requirement gene *htrB***. *J Bacteriol* 1992, **174**:702–710.
- 979 96. Moran N, McCutcheon J, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts**. *Annual Review*
980 *of Genetics* 2008, **42**:165–190.
- 981 97. Nussbaumer AD, Fisher CR, Bright M: **Horizontal endosymbiont transmission in hydrothermal vent tubeworms**.
982 *Nature* 2006, **441**:345–348.
- 983 98. Cavanaugh CM: **Symbiosis of chemoautotrophic bacteria and marine invertebrates**. *PhD Thesis*. Harvard
984 University, Department of Organismic and Evolutionary Biology; 1985.
- 985 99. Fisher C, Childress J: **Organic carbon transfer from methanotrophic symbionts to the host hydrocarbon-seep**
986 **mussel**. *Symbiosis* 1992, **12**:221–235.
- 987 100. Saurin W, Hofnung M, Dassa E: **Getting in or out: early segregation between importers and exporters in the**
988 **evolution of ATP-binding cassette (ABC) transporters**. *J Mol Evol* 1999, **48**:22–41.
- 989 101. van de Velde W, Zehirov G, Szatmari A, Debreczeny M, Ishihara H, Kevei Z, Farkas A, Mikulass K, Nagy A, Tiricz H:
990 **Plant peptides govern terminal differentiation of bacteria in symbiosis**. *Science* 2010, **327**:1122–1125.
- 991 102. Paau AS, Bloch CB, Brill WJ: **Developmental fate of *Rhizobium meliloti* bacteroids in alfalfa nodules**. *J Bacteriol*
992 1980, **143**:1480–1490.
- 993 103. Stewart FJ, Cavanaugh CM: **Bacterial endosymbioses in *Solemya* (Mollusca: Bivalvia)—model systems for**
994 **studies of symbiont–host adaptation**. *Antonie van Leeuwenhoek* 2006, **90**:343–360.
- 995 104. Whitchurch CB, Leech AJ, Young MD, Kennedy D, Sargent JL, Bertrand JJ, Semmler ABT, Mellick AS, Martin PR,
996 Alm RA, Hobbs M, Beatson SA, Huang B, Nguyen L, Commolli JC, Engel JN, Darzins A, Mattick JS: **Characterization of**
997 **a complex chemosensory signal transduction system which controls twitching motility in *Pseudomonas***
998 ***aeruginosa***. *Mol Microbiol* 2004, **52**:873–893.
- 999 105. Newton ILG, Bordenstein SR: **Correlations Between Bacterial Ecology and Mobile DNA**. *Current Microbiology*
1000 2011, **62**:198–208.
- 1001 106. Plague GR, Dunbar HE, Tran PL, Moran NA: **Extensive proliferation of transposable elements in heritable**
1002 **bacterial symbionts**. *J Bacteriol* 2008, **190**:777–779.
- 1003 107. Gil R, Latorre A, Moya A: **Evolution of Prokaryote-Animal Symbiosis from a Genomics Perspective**. In
1004 *Microbiology Monographs. Volume 19*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010:207–233.
- 1005 108. Cordaux R, Pichon S, Ling A, Pérez P, Delaunay C, Vavre F, Bouchon D, Grève P: **Intense transpositional activity**
1006 **of insertion sequences in an ancient obligate endosymbiont**. *Molecular Biology and Evolution* 2008, **25**:1889–1896.
- 1007 109. Chafee ME, Funk DJ, Harrison RG, Bordenstein SR: **Lateral phage transfer in obligate intracellular bacteria**
1008 **(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, Ciufu 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**

1054 versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 2009, **326**:578–582.

1055

1056 **Figures**

1057 **Figure 1. Predicted model of the *S. velum* symbiont cell.**

1058 The diagram, based on gene annotation of the symbiont genome, depicts key functional
1059 systems and metabolic pathways: sulfur oxidation, electron transport, ATP synthases,
1060 CO₂-fixation via Calvin-Benson Cycle, gluconeogenesis, polyglucose synthesis,
1061 glycolysis, TCA and glyoxylate cycles, synthesis of amino acids, fatty acids, lipids,
1062 isoprenoids via non-mevalonate pathway, and the cell wall, solute transporters, protein
1063 secretion systems, and type IV pilus. Different protein categories are color-coded and
1064 the individual subunits indicated by shape symbols. The direction of substrate transport
1065 across the membrane is shown with arrows. Components of the electron transport chain
1066 are arranged from the lowest to the highest electronegativity of the electron donors
1067 (blue) and acceptors (red). The corresponding electronegativity values are listed next to
1068 the enzymes. Enzymes shared between the glycolysis, gluconeogenesis, and the Calvin
1069 cycles are designated in green. Enzymes unique to these pathways are designated in
1070 red. Enzymes shared between the Calvin cycle and pentose phosphate cycle are
1071 designated in blue. Amino acids which may be essential for the host are designated in
1072 red. Speculated pathways are designated with a question mark. The abbreviations used,
1073 the respective full gene names, and their genome reference numbers are listed in Table
1074 S3 in Additional file 3.

1075

1076 **Figure 2. Comparison of the COG categories between the *S. velum* symbiont,**

1077 **selected symbiotic and free-living bacteria.** The percentage of genes in each
1078 category is normalized to the percentage of those COG categories in the genome of *E.*
1079 *coli* K12 DH1, ATCC 33849. *NCBI accession PRJNA16744.

1080

1081 **Figure 3. Taxa assigned to the genes in the *S. velum* symbiont genome.** The insert
1082 chart shows breakdown of the genes in the taxa within the class of γ -proteobacteria
1083 (62.9%). The unassigned genes have not been assigned a lower taxon in this analysis.
1084 The unclassified genes have not been further classified in the NCBI taxonomy. All the
1085 taxa are mutually exclusive.

1086

1087 **Figure 4. Comparison of the sulfur oxidation genes between the *S. velum***
1088 **symbiont and other SOB.** (a) Presence of genes involved in chemotrophic sulfur
1089 oxidation in the symbionts of *S. velum*, other sulfur-oxidizing bacteria and archaea, and
1090 sulfate reduction in *D. autotrophicum*, which is included for comparison. Genes encoding
1091 pathways for reverse-acting dissimilatory sulfur-oxidation (rDsr) (Drs in *D.*
1092 *autotrophicum*) and periplasmic sulfur-oxidation (Sox), as well as auxiliary proteins, are
1093 listed. Numbers of gene homologs in each organism is designated with color. Presence
1094 of extra- or intracellular sulfur deposits, i.e., globules, in each organism, as obtained
1095 from literature, is indicated with hollow circles. The gene symbols used, the respective
1096 full gene names, and the reference numbers in the genome of the *S. velum* symbionts
1097 are listed in Table S3 in Additional file 3. (b) Presence of signal sequences and
1098 transmembrane domains in the sulfur-oxidations genes of the *S. velum* symbionts,
1099 followed by the list of organisms with the closest known homologs to those genes and
1100 their respective BLAST % identities (Avi - *Allochromatium vinosum*, Sup05 - uncultivated
1101 oxygen minimum zone microbe [129], Sli - *Sideroxydans lithotrophicus*, and Thia -
1102 *Thiocapsa marina*).

1103

1104

1105 **Tables**

1106 **Table 1. General genome features of the *S. velum* symbiont in comparison to other**
 1107 **γ -proteobacteria.** The comparison includes genomes of the chemosynthetic symbionts
 1108 of *R. pachyptila*, *C. magnifica*, and *C. okutanii*; a symbiont of psyllids (the smallest
 1109 sequenced genome), *Carsonella ruddii*; an α -proteobacterial aphid symbiont, *B.*
 1110 *aphidicola*; free-living sulfur-oxidizers, *T. crunogena* and *A. vinosum*, and
 1111 enterobacterium *E. coli*. * NCBI Accession PRJNA16744 and PRJNA72967.
 1112

	<i>Solemya velum</i> endosymbiont	<i>Riftia pachyptila</i> endosymbiont*	<i>Calyptogena magnifica</i> endosymbiont	<i>Calyptogena okutanii</i> endosymbiont	<i>Buchnera aphidicola</i> APS	<i>Ca. Carsonella ruddii</i> PV	<i>Thiomicrospira crunogena</i> XCL-2	<i>Allochromatium 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
	Symbionts						Free-living		

1113
 1114 **Table 2. Comparison of extracellular transport genes in the *S. velum* symbiont,**
 1115 **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
<i>Solemya velum</i> endosymbiont	Intracellular symbiont	1.00	2.7	224	75.2	100	70	1	5	5	17	26
<i>Calyptogena magnifica</i> endosymbiont	Obligate intracellular symbiont	0.14	1.16	32	27.6	18	6	0	3	1	0	4
<i>Calyptogena okutanii</i> endosymbiont	Obligate intracellular symbiont	0.15	1.02	34	33.3	16	10	0	2	3	0	3
<i>Buchnera aphidicola</i> APS	Obligate intracellular symbiont	0.07	0.64	16	25.0	5	3	5	1	0	0	2
<i>Sulcia muelleri</i> GWSS	Obligate intracellular symbiont	0.03	0.25	7	28.0	4	2	0	0	0	0	1
<i>Candidatus Blochmannia floridanus</i>	Obligate intracellular symbiont	0.12	0.71	27	38.0	7	12	3	2	0	0	3
<i>Wigglesworthia glossinidia</i>	Obligate intracellular symbiont	0.11	0.70	25	35.7	9	14	0	2	0	0	0
<i>Baumannia cicadellincola</i>	Obligate intracellular symbiont	0.13	0.69	28	40.6	11	10	3	1	0	0	3
<i>Rhizobium leguminosarum</i> bv. Viciae 3841	Free-living intracellular symbiont	2.47	7.75	553	71.4	281	203	7	18	2	13	29
<i>Frankia alni</i> ACN14a	Free-living intracellular symbiont	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>Thermodesulfovibrio yellowstonii</i> DSM 11347	Free-living sulfate reducer	0.39	2.00	88	44.0	31	34	0	3	2	3	15

1116

1117 **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

1118

1119 **Table 4. Insertion sequence mobile genetic elements in the *S. velum* symbiont**
 1120 **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

1121 **Additional files**

1122 **Additional file 1. Table S1.**

1123 Length [bp], GC%, percentage of the total base pairs, and the number of genes in the
 1124 scaffolds which constitute the genome of the *S. velum* symbiont.

1125 **Additional file 2. Table S2.**

1126 tRNA genes and the codon frequencies in the genome of the *S. velum* symbiont.

1127 **Additional file 3. Table S3.**

1128 Gene product names used in Figure 1 and Figure 4, the corresponding NCBI protein ID

1129 reference numbers, and EC/TC numbers.

1130 **Additional file 4. Table S4.**

1131 Parameters of the gene prediction software.

1132 **Additional file 5. Genomic Utility for Automated Comparison (GUAC).**

1133 A Python script developed to inform comparative analyses of gene content across

1134 multiple genomes.

Abundance of genes relative to *Escherichia coli*

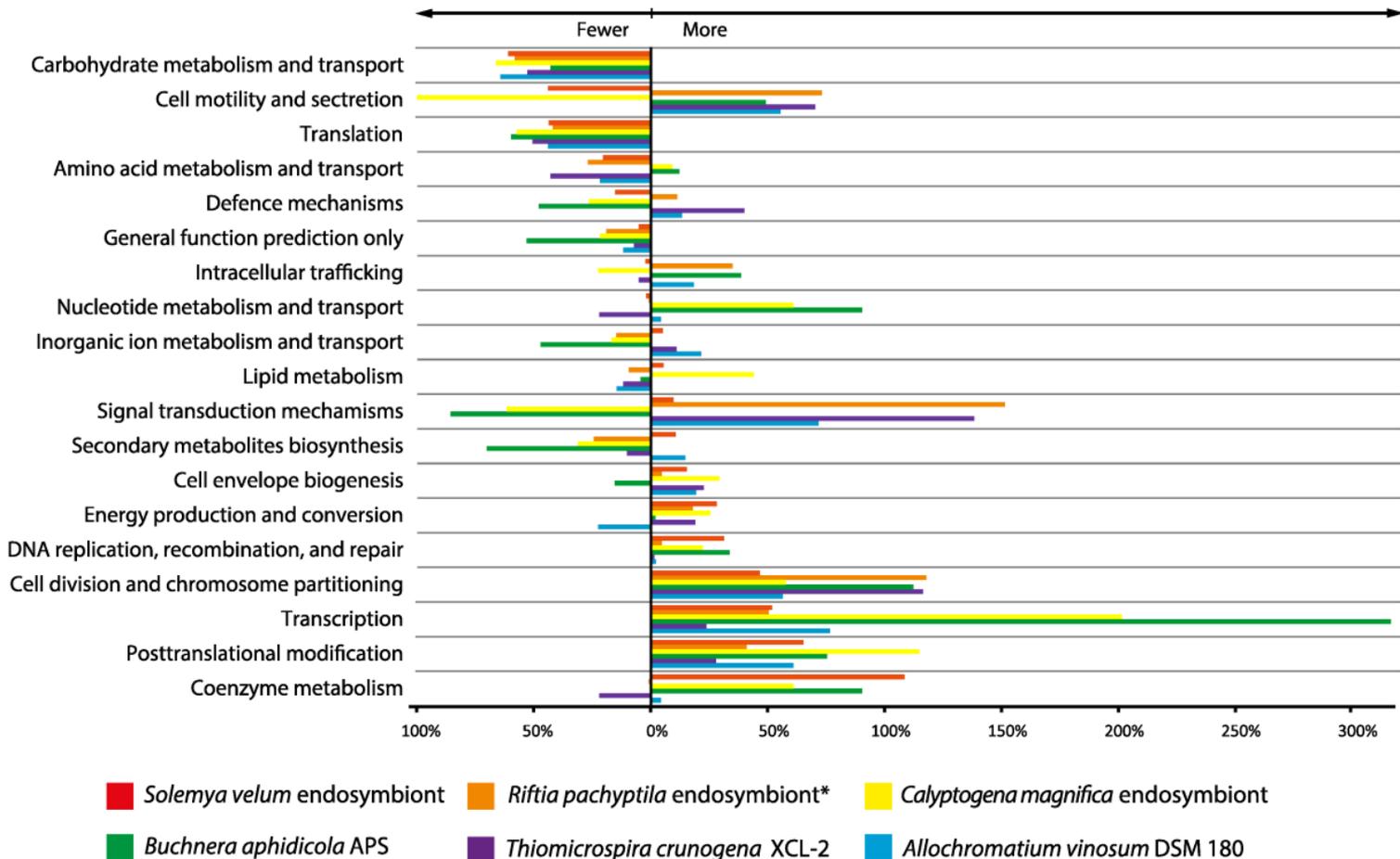


Figure 2

Number of genes in each taxon

0 200 400 600 800 1000 1200 1400 1600 1800 2000

Taxa

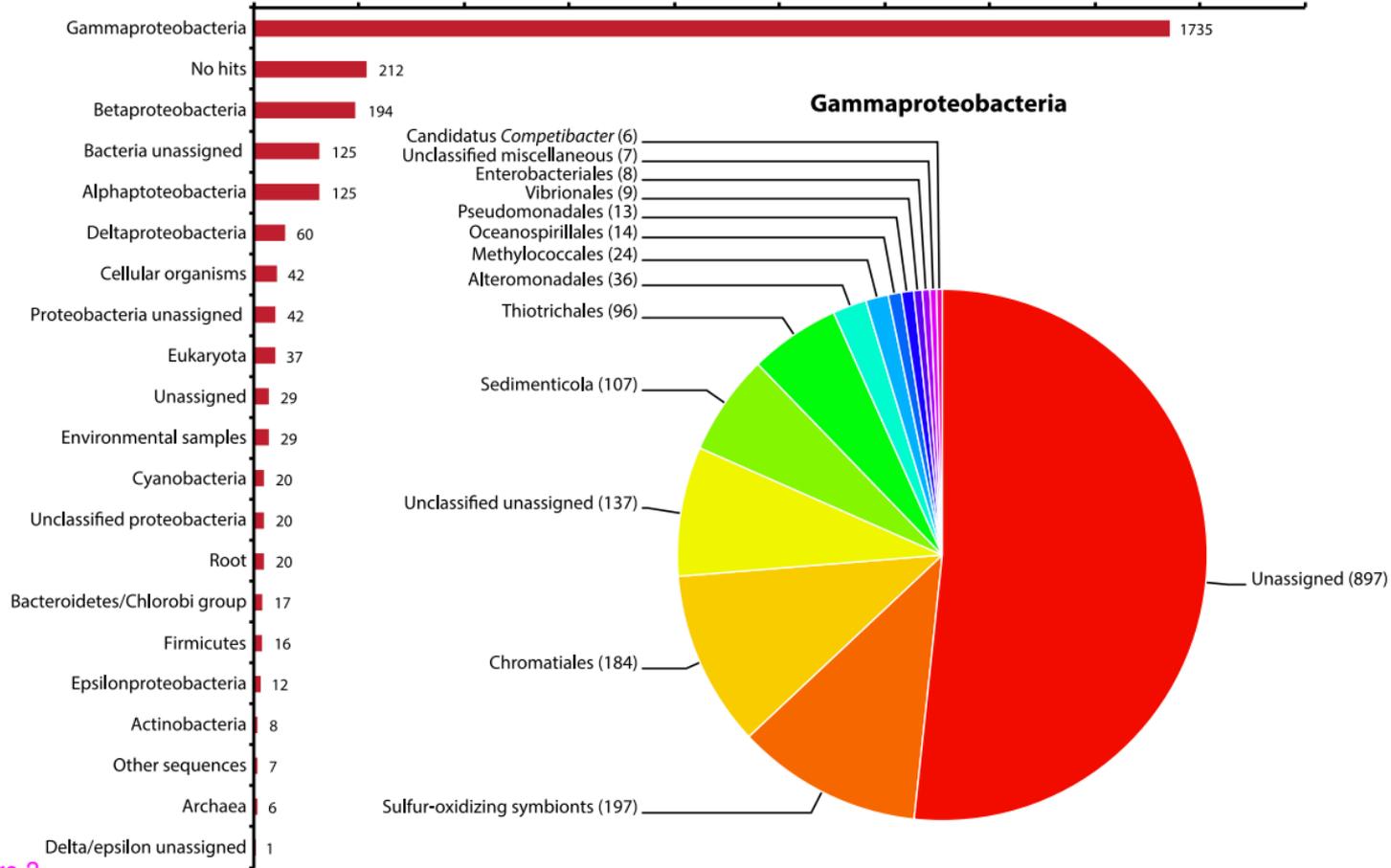


Figure 3

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>