A window into hydrothermal vent endosymbioses: the *Calyptogena magnifica* chemoautotrophic symbiont genome

Authors: I.L.G. Newton[1], T. Woyke[2], T.A. Auchtung[1], G.F. Dilly[1], R.J. Dutton[3], M.C. Fisher[1], K. M. Fontanez[1], E. Lau[1], F.J. Stewart[1], P.M. Richardson[2], K.W. Barry[2], J.C. Detter[2], D. Wu[4], J. A. Eisen[5], C.M. Cavanaugh[1*]

*corresponding author

The *Calyptogena magnifica* symbiont is the most metabolically capable intracellular endosymbiont, able to oxidize sulfur, fix carbon dioxide, assimilate nitrogen, and synthesize vitamins, cofactors, and 20 amino acids.

Keywords: chemosynthesis, maternal transmission, deep-sea, symbiosis

None of this material has been published or is under consideration elsewhere, including the Internet.

Mailing Addresses:

1 –Harvard University, 16 Divinity Ave, Biolabs 4080, Cambridge, MA 02138
Telephone number: 617-495-1138
Fax number: 617-496-6933

2 - DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
Telephone number: 530 752 3498
Fax number: 925-296-5850

3 – Harvard Medical School, Department of Microbiology and Molecular Genetics, 200 Longwood Ave., Boston, MA 02115
Telephone number: 617-432-1788
Fax number: 617-738-7664

4 – The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850
Telephone number: 301-795-7000
Fax number: 301-838-0208

5 – University of California, Davis Genome Center
Genome and Biomedical Sciences Facility Room 5311
451 East Health Sciences Drive
Davis, CA 95616-8816
Telephone number: 530 752 3498

Email addresses:

I.L.G. Newton – Garcia@fas.harvard.edu

T. Woyke – twoyke@lbl.gov

T.A. Auchtung – auchtung@fas.harvard.edu

G.F. Dilly – gdilly@fas.harvard.edu

R.J. Dutton – Rachel_dutton@hms.harvard.edu

M.C. Fisher – mfisher@oeb.harvard.edu

K. M. Fontanez – kfontanez@oeb.harvard.edu

E. Lau – elau@oeb.harvard.edu

F.J. Stewart - fstewart@fas.harvard.edu

P.M. Richardson - PMRichardson@lbl.gov

K.W. Barry – kwbarry@lbl.gov

J.C. Detter - cdetter@lanl.gov

D. Wu - dwu@tigr.org

J. A. Eisen – jaeisen@ucdavis.edu

C.M. Cavanaugh – cavanaug@fas.harvard.edu

Chemosynthetic endosymbionts are the metabolic cornerstone of hydrothermal vent communities, providing invertebrate hosts with nearly all of their nutrition. The *Calyptogena magnifica* (Bivalvia: Vesicomyidae) symbiont, *Candidatus Ruthia magnifica,* is the first intracellular chemosynthetic endosymbiont to have its genome sequenced, revealing an enormous suite of metabolic capabilities. The genome encodes the major chemosynthetic pathways as well as pathways for biosynthesis of vitamins, cofactors, and all 20 amino acids required by the host, indicating the host is entirely nutritionally dependent on *Ruthia*. This genome sequence will be invaluable in the study of these enigmatic associations and provides insights into the origin and evolution of autotrophic endosymbioses.

Miles below the surface of the ocean, where tectonic plates meet, the food-limited habitat of the deep-sea is punctuated by diverse communities of invertebrates and bacteria. Metazoans at these hydrothermal vents flourish thanks to the chemoautotrophy of symbiotic bacteria (*1*). Seawater here percolates into the crust, is heated as it reacts with oceanic basalt, and becomes enriched in the reduced sulfur and carbon dioxide that sulfur oxidizing chemoautotrophs require (*1*). The symbiotic bacteria use the energy gained in oxidation of these reduced sulfur compounds for carbon fixation. Analogous to photosynthetic chloroplasts, which are derived from cyanobacterial ancestors and use light energy to fix carbon for their plant and algal hosts, these chemosynthetic endosymbionts use chemical energy to provide their hosts with not only carbon but also a large array of additional necessary nutrients. The metazoan hosts, in turn, bridge the oxic-anoxic interface to provide their bacteria with the inorganic substrates necessary for chemosynthesis. Hosts often betray their nutritional dependence on these bacteria through their diminished or absent digestive systems. Although first discovered at hydrothermal vents, similar associations exist at mud flats, seagrass beds, and hydrocarbon seeps. In each case it is clear that these symbioses play major roles in community structuring and sulfur and carbon cycling. However, despite the widespread occurrence of these partnerships, little is known of the intricacies of host-symbiont interaction or symbiont metabolism due to their inaccessibility and our inability to culture either partner separately.

The giant clam, *Calyptogena magnifica* Boss and Turner (Bivalvia: Vesicomyidae) was one of the first organisms described following the discovery of hydrothermal vents (*2*). The vesicomyids are relatively old, with fossil records and phylogenies dating them at 50-100 Ma (*3*). *C. magnifica* grows to a large size (>26 cm in length), despite having a reduced gut and ciliary food groove (*2*), presenting a conundrum regarding how it acquires sufficient nutrients. The mystery of this clam's nutrition was solved when chemosynthetic, γ-proteobacterial symbionts, here named *Candidatus Ruthia magnifica* (in memory of Prof. Ruth Turner), were discovered within its gill bacteriocytes (*4, 5*) (Figure 1). The host depends largely on these symbionts for its carbon, as indicated by its anatomy and by stable carbon isotopic ratios (*6, 7*). However, how the host satisfies the rest of its nutritional needs remains unknown.

R. magnifica is the first intracellular chemosynthetic symbiont to have its genome sequenced. Here we describe analysis of this finished sequence. In particular we discuss how, despite a relatively small genome, the symbiont is predicted to convey a striking diversity of nutritional capabilities on the host. In addition, we consider how this symbiont's genome differs in fundamental ways from those of other nutritional endosymbionts.

Although, in some ways, the *R. magnifica* genome resembles that of other obligate mutualistic symbionts for which data are available, surprising differences were found. The genome has a low G+C content (34%) compared to free-living relatives (Table 1). In addition, the coding density (81.4%) and mean gene length (975 bp), though lower than commonly seen in free-living bacteria, are consistent with that in other endosymbiont genomes (*8*). These common features of endosymbionts are likely the result of genome reduction and degradation (rampant gene loss and mutation rate increases, respectively) that occur over evolutionary time across diverse symbiont species. This trend is evident in relatively recent symbioses such as the insect endosymbionts (30-250 Ma), as well as in chloroplasts (~1,800-2,100 Ma). Upon closer examination however, *R. magnifica* stands out in that its genome is large for a maternally transmitted endosymbiont (1.2 Mb). For example, the genomes of the γ-proteobacterial *Buchnera* species, which are endosymbionts of aphids, are some 80% smaller than closely related free-living species like *E. coli.* In contrast, *R. magnifica*'s genome is half the size of its relative's, *Thiomicrospira crunogena,* a free-living, γ-proteobacterial, sulfur-oxidizing chemoautotroph.

We propose that the limited genome reduction in *R. magnifica* is due to a fundamental difference in its biology compared to other nutritional endosymbionts characterized so far. Insect endosymbionts typically supplement the diet of their hosts, e.g., *Buchnera* provide essential amino acids that are missing in the phloem sap diet of aphids. Similarly, the γ-proteobacteria *Baumannia* and *Sulcia* together provide amino acids and vitamins for their sharpshooter hosts, but apparently not much more (*9*). These symbionts acquire much of what they need (e.g., sugars) from their host and thus can still survive with very small genomes (*10*). In contrast, and most strikingly, *R. magnifica* is predicted to encode all the metabolic pathways one would expect in free-living chemoautotrophs including carbon fixation, sulfur oxidation, nitrogen assimilation, and amino acid and cofactor/vitamin biosynthesis (Figure 2). Thus we conclude it provides the clam with the majority of its nutrition. In the following sections we discuss different aspects of the metabolic reconstruction of *R. magnifica* and what this might mean for the biology of its host. For simplicity, we refer to these reconstructions as though the pathways have been validated, although it should be emphasized that these are predictions.

R. magnifica's genome is largely dedicated to biosynthesis and energy metabolism, highlighting the importance of these pathways in the symbiosis (Figure 2). The *R. magnifica* genome also encodes enzymes for carbon fixation, sulfur oxidation, nitrogen assimilation and energy conservation. Genes encoding enzymes specific to the Calvin Cycle, a form II ribulose-1,5- bisphosphate carboxylase/oxygenase (RubisCO) and phosphoribulokinase (*11, 12*), were found in the *R. magnifica* genome (Figure 3). This pathway synthesizes phosphoglyceraldehyde from carbon dioxide and is the dominant form of carbon fixation in vent symbioses (*13*). However, the genome lacks homologs of sedoheptulose 1,7-bis-phosphatase (SBPase, EC 3.1.3.37) and fructose 1,6-bis-phosphatase (FBPase, EC 3.1.3.11), suggesting that the regeneration of ribulose 1,5-bisphosphate may not follow conventional routes. Instead, the *R. magnifica* genome contains a reversible pyrophosphate-dependent phosphofructokinase (EC 2.7.1.90) homolog that may use to generate fructose 6-phosphate (*14*).

Energy generation for carbon fixation in *R. magnifica* can result from sulfur oxidation via the *sox* (sulfur oxidation) and *dsr* (dissimilatory sulfite reductase) genes (Figure 3). The *R. magnifica sox* genes resemble those of the γ-proteobacteria *Thiobacillus denitrificans* and *Allochromatium vinosum,* and the green sulfur bacterium *Chlorobium tepidum* (*15-17*). Homologs of the *sox* genes are located in two positions in the *R. magnifica* genome with *soxXYZA* located in a single operon while *soxB* is elsewhere. The symbiont genome also contains homologs for many of the *dsr* genes which catalyze the oxidation of intracellularly stored sulfur in both *A. vinosum* and *Chlorobium limicola* (*16, 18*). Indeed, sulfur granules observed within *R. magnifica* cells may be a source of reduced sulfur when external sulfide is lacking (*19*). The symbiont's *dsr* genes were contained in a single cluster, *dsrABEFHCMKLOP*, missing *dsrJNRS*. As these latter proteins are not well characterized, it is not known how symbiont sulfur metabolism may be affected. Homologs encoding both a sulfide:quinone oxidoreductase and rhodanese are present, and along with the *dsr* and *sox* proteins, these enzymes can oxidize both thiosulfate ($S_2O_3^{2-}$) or sulfide (HS^-) to sulfite (SO_3^{2-}) (Figure 3). Sulfite can then be oxidized to sulfate (SO_4^{2-}) by the actions of APS reductase (AprAB) and ATP sulfurylase (Sat) before being exported from the cell via a sulfate transporter. This genomic evidence is supported by ATP sulfurylase activity detected in *C. magnifica* gill tissue (*7*), carbon dioxide uptake when sulfide or thiosulfate are provided to the clam (*20, 21*) and sulfide binding, zinc-containing lipoprotein in the host blood stream (*22*). Thus through the activities of the *sox* and *dsr* genes, the *R. magnifica* symbiont can generate energy from the oxidation of sulfide and thiosulfate.

Energy conservation, which involves creating a charged membrane, proceeds in *R. magnifica* through NADH dehydrogenase, a sulfide:quinone oxidoreductase, and an *rnf* complex, which in other bacteria has been shown to possess NADH and FMN:quinone oxidoreductase activity (*23*). The genome encodes a straightforward electron transport chain, thus the reduced quinone in the symbiont membrane could transfer electrons to cytochrome c via a bc_1 complex and a terminal cytochrome c oxidase could then transfer these electrons to oxygen.

Nitrogen assimilation is as important as carbon fixation in the context of this symbiosis as *Ruthia* appears to provide the majority if not all of the host's amino acids. In the predicted pathways, nitrate and ammonia enter the cell via a nitrate/nitrite (NarK) transporter and two ammonium permeases (AmtB1/2) and are then reduced via nitrate (NarB) and nitrite (NirA) reductase, and assimilated via glutamine synthetase (GlnA) and glutamate synthase (GltB/D), respectively (Figure 3). Although nitrate is the dominant form of nitrogen present at vents (*24*) and likely the source of nitrogen for the symbiosis, the symbiont may also assimilate ammonia via recycling of the host's amino acid waste products.

In keeping with the nutritional role of the symbionts, *R. magnifica*'s inferred intermediary metabolism can produce all necessary biosynthetic intermediates. The genome encodes a complete glycolytic pathway with a pyrophosphate-dependent phosphofructokinase homolog and the non-oxidative branch of the pentose phosphate pathway. The symbiont genome encodes a "horseshoe shaped" tricarboxylic acid (TCA) cycle, lacking alpha-ketoglutarate dehydrogenase. For other chemosynthetic bacteria, the lack of this enzyme has been suggested as an indicator of obligate autotrophy (*25*). Interestingly, the symbiont is also missing homologs of fumarate reductase, succinyl-coA synthase, and succinate dehydrogenase. However, the genome encodes isocitrate lyase, part of the glyoxylate shunt, and could produce succinate from isocitrate. Carbon fixed via the Calvin cycle can enter the TCA cycle through phosphoenolpyruvate and here could follow biosynthetic routes either to fumarate or alpha-ketoglutarate. All of the pathways for biosynthetic reagents required to support the metabolic capabilities of *R. magnifica* are thus encoded in the symbiont genome.

Unlike any other sequenced endosymbiont genome, *R. magnifica* encodes complete pathways for the biosynthesis of 20 amino acids. This full complement suggests that the symbiont can supply its host with the 9 essential amino acids or their precursors.   However, while *E. coli* has 16 essential amino acid biosynthesis regulatory genes (*26*), *metR* (involved in regulating methionine biosynthesis) is the only regulatory gene present in the *R. magnifica* genome. This lack of regulatory genes may be the result of the stability experienced by *R. magnifica* in its intracellular environment.

Animals are dependent on external sources for many of their vitamins and cofactors and bacterial symbionts often provide these nutrients (*10, 27*). The *R. magnifica* genome appears to have complete biosynthetic pathways for the majority of vitamins and cofactors *(39).* The only pathway conspicuously absent is that for cobalamin (B_{12}), a cofactor for methionine synthase (*27, 28*). Since *R. magnifica* encodes a cobalamin-independent methionine synthase, it is able to provide the host with methionine and the host is unlikely to require cobalamin.

As with other intracellular species, *R. magnifica* encodes a limited repertoire of transporters, however, those present reveal important details about the movement of metabolites between host and symbiont. Of the 58 proteins predicted to be involved in cell transport and binding in the *R. magnifica* genome, transporters involved in chemosynthesis (sulfate exporters), nitrogen assimilation (ammonium and nitrate importers), inorganic compounds (TrkAH, MgtE family, CaCA family and PiT family), and heavy metals (ZnuABC, RND superfamily, iron permeases) were identified. Surprisingly, few substrate-specific transporters and only two ABC transporter proteins of unknown substrate were found. As it is unlikely that these two ABC transporter proteins are translocating amino acids, vitamins, and cofactors to the host, perhaps the symbionts are "leaky" or the host is actively digesting symbiont cells. Indeed, the closest known relative to *Ruthia*, the bathymodiolid mussel symbionts, are digested intracellularly by their host (*29*). Although the vesicomyid clam and the bathymodiolid mussels are not closely related, electron micrographs suggest the presence of putative degradative stages of symbionts within *C. magnifica* bacteriocytes (Figure 1b).

Interestingly, the *R. magnifica* genome lacked the key cell division gene, *ftsZ*. FtsZ, a tubulin homolog, assembles as a ring within the bacterial cell, recruits the remaining cell division proteins and constricts to initiate cytokinesis (*30*). It is puzzling that *R. magnifica* lacked FtsZ given that it is almost universally conserved in bacteria, with the notable exception of the obligately intracellular pathogens in the Chlamydia division (*31*). In addition to the absence of *ftsZ*, *R. magnifica* and Chlamydia both lack the *murI* gene (*32*), required for the synthesis of D-glutamate, an essential component of the bacterial cell wall. The potential similarities in cell division and cell wall machinery between *R. magnifica* and Chlamydia may be responsible for the "elementary body" cell morphologies observed in both organisms inside the host cell (Figure 1b, *33*). In Chlamydia these bodies are the infectious, propagating form (*34*); their appearance in *R. magnifica* may reflect common mechanisms for adaptation to an obligately intracellular lifestyle.

Endosymbiont intracellular lifestyles have severe effects on genome evolution including genome reductions, skewed base compositions, and elevated rates of gene evolution (*8*). As noted above, *R. magnifica* does exhibit skewed composition and genome reduction, although these are minor shifts compared to those seen in insect endosymbionts. Previous studies have shown, however, that *R. magnifica* also exhibits faster nucleotide substitution rates than those of both free-living bacteria and environmentally transmitted chemosynthetic symbionts (*35*). The factors that contribute to these features of endosymbiont evolution are believed to be a combination of a relatively stable environment, population bottlenecks, and sequestration from free-living bacteria all of which likely occur in *R. magnifica*. In addition, as with some but not all other endosymbionts, *R. magnifica* has lost key genes in DNA repair processes that likely enhance the speed of genome degradation. For example, it is missing genes involved in induction of the SOS repair system and in recombinational repair, including the exonuclease complex genes *recB,C,D* and the highly conserved recombinase *recA* . Perhaps most importantly, it is also missing genes that could encode homologs of the MutSLH proteins, which, in other species greatly limit mutation rates by carrying out post-replication mismatch repair (*36*).

Given the apparent defects in DNA repair and the likely population forces pushing this organism's genome towards degradation it is particularly informative that it has retained genes that encode a full suite of chemosynthesis processes. For comparison, chloroplast genomes have lost over 90% of their content since their cyanobacterial ancestor entered endosymbiosis, with many of their genes having been transferred to the host nuclear genome (*37*). The more modern insect endosymbioses have lost between 70-80% of their genomes over a much shorter evolutionary time, and it is unknown if any of these pathways are encoded by the nucleus (*10, 38*). *R. magnifica*, in contrast, has the largest genome of any intracellular symbiont sequenced to date and may represent an early evolutionary intermediate towards a chemoautotrophic "plastid". The broad array of metabolic pathways encoded by *R. magnifica* expands prior knowledge of host nutritional dependency based on stable carbon isotopic ratios and host physiology and anatomy (*6, 7*). It is the extent of this dependency that may be preventing the loss of metabolic pathways in the *R. magnifica* genome. This selective pressure might be great enough to counter the forces of genome reduction and degradation seen in other endosymbionts and provides a novel framework for the study of endosymbiont evolution.

Methods:

Specimen collection and DNA extraction:

Calyptogena magnifica clams were collected using *DSV Alvin* at the East Pacific Rise, 9°N, during a December 2004 cruise on the *R/V Atlantis*. The symbiont-containing gills were dissected out of the clams, frozen in liquid nitrogen, and kept at -80°C until processed. They were then ground in liquid nitrogen, placed in lysis buffer (20 mM EDTA, 10 mM Tris-HCl, pH 7.9, 0.5 mg/ml lysozyme, 1% Triton X-100, 500 mM guanidine-HCl, 200 mM NaCl) and kept at 40ºC for 2 hr. After subsequent RNase (20 μg/ml, 37°C, 30 min) and proteinase K (20 μg/ml, 50ºC, 1.5 hr) treatments, the samples were centrifuged and the supernatant loaded onto a Qiagen genomic tip column and processed according to manufacturer's instructions.

Shotgun library construction

3 kb library. Briefly, 3 µg of DNA was randomly sheared to 2-4 kb fragments using a HydroShear® (GeneMachines) and end-repaired using T4 DNA polymerase and DNA Polymerase I, Large (Klenow) Fragment (New England Biolabs). The DNA was agarose gel separated and gel-purified using the QIAquick Gel Extraction Kit (Qiagen). Approximately 200 ng of sheared DNA was then ligated into 100 ng of linearized and dephosphorylated pUC18 vector

(Roche) at 24.5˚C for 90 min using the Fast-Link^{TM} DNA Ligation Kit (Epicentre). The ligation product was electroporated into ElectroMAX DH10B™ cells (Invitrogen) and plated on selective agar plates. Positive library clones were robotically picked using the Q-Bot multitasking robot (Genetix) and grown in selective media for sequencing.

8 kb library. Briefly, 10 µg of HMW DNA was randomly sheared to 6-8 kb fragments and end-repaired as described above. The DNA was agarose gel separated and filter tip gel-purified. Approximately 200 ng of DNA was blunt-end ligated into 100 ng of pMCL200 vector O/N at 16ºC using T4 DNA ligase (Roche Applied Science) and 10% (vol/vol) polyethylene glycol (Sigma). The ligation was phenol-chloroform extracted, ethanol precipitated, and resuspended in 20 µl TE. According to the manufacturers instructions, 1 µl of ligation product was electroporated into ElectroMAX DH10B™ Cells and processed as described above.

Fosmid library. The fosmid library was constructed using the CopyControl™ Fosmid Library Production Kit (Epicentre). DNA (~20 µg) was randomly sheared using a HydroShear, blunt-end repaired as described above and separated on an agarose pulse-field gel O/N at 4.5 V/cm. The 40 kb fragments were excised, gel-purified using AgarACE^{TM} (Promega) digestion followed by phenol-chloroform extraction and ethanol precipitation. DNA fragments were ligated into the pCC1Fos^{TM} Vector and the ligation packaged using MaxPlax^{TM} Lambda Packaging Extract and used to transfect TransforMax^{TM} EPI300 *E. coli*.

Transfected cells were plated on selective agar plates and fosmid clones picked using the Q-Bot multitasking robot and grown in selective media for sequencing.

End-sequencing

The pUC library was sequenced using using DyEnamic ET Terminators and resolved on MB4500 (MolecularDynamics/GeneralElectric). The pMCL and pCC1Fos libraries were sequenced with BigDye Terminators v3.1 and resolved with ABI PRISM 3730 (ABI) sequencers.

Processing and Assembly of Shotgun Data

A total of 22.15 Mb of phred Q20 sequence was generated from the three libraries; 9.43 Mb from 13755 reads from the small insert pUC library, 8.79 Mb from 13824 reads from the medium insert pMCL library, and 3.93 Mb from 9216 reads from the fosmid library. The DNA sequences derived from the *Ruthia magnifica* libraries were estimated to be 20% contaminated with the *Calyptogena magnifica* host genome. Although this level of contamination can confound finishing efforts, the bacterial genome was readily identifiable in our study. The 36,795 sequencing reads were blasted against a database containing all mollusk sequence available at NCBI and the 4X draft sequence available at the JGI for *Lottia gigantea*. A total of 498 reads were removed based on hits to this mollusk database. The remaining 24,595 reads were base called using phred version 0.990722.g, vector trimmed using crossmatch SPS-3.57, and assembled using parallel phrap compiled for SUNOS, version SPS - 4.18. One large, bacterial scaffold containing the *Ruthia magnifica* 16S rRNA gene resulted. The *Ruthia magnifica* scaffold consisted of only 2 contigs spanned by 33 fosmid clones, contained 17,307 reads, 1,156,121 consensus bp, was covered by an average read depth of 14X, and had a G+C content of 34%. The next largest scaffold was only 29 kb long, with an average read depth of ~7X and an average G+C content of 55%. BLASTn indicated that this scaffold encoded ribosomal genes closely related to those of *Caenorhabditis briggsae* and its binning (based on GC content and read depth) with a small scaffold containing the *Calyptogena magnifica* 18S rRNA gene confirmed its eukaryotic host origin.

Annotation and pathway reconstruction

Assembled sequence was first loaded into The Institute for Genomic Research's (TIGR) auto-annotation pipeline before being imported into MANATEE (http://manatee.sourceforge.net/), a web-based interface for manual annotation. Only after putative genes were computationally and manually validated were they assigned names and gene symbols. The TIGR guidelines for manual annotation based on annotator confidence in computational evidence were followed. The *Ruthia magnifica* genome was finished at the Joint Genome Institute and the assembly is currently being quality checked.

1. C. M. Cavanaugh, Z. P. McKiness, I. L. G. Newton, F. Stewart, in *The Prokaryotes*. (Springer-Verlag, Berlin, 2004).
2. K. J. Boss, R. D. Turner, *Malacologia* **20**, 161 (1980).
3. A. S. Peek, R. G. Gustafson, R. A. Lutz, R. C. Vrijenhoek, *Mar. Biol.* **130**, 151 (Dec, 1997).
4. H. Felbeck, G. Somero, *Trends Biochem. Sci.*, 201 (1982).
5. C. Cavanaugh, *Nature* **302**, 58 (1983).
6. J. J. Robinson, Ph.D., Harvard University (1997).
7. C. R. Fisher *et al.*, *Deep-Sea Res.* **35**, 1811 (1988).
8. J. J. Wernegreen, *Nature Rev. Genet.* **3**, 850 (Nov, 2002).
9. D. Wu *et al.*, *PLoS Biol.* **4**, e188 (Jun 6, 2006).
10. E. Zientz, T. Dandekar, R. Gross, *Microbiol. Mol. Biol. Rev.* **68**, 745 (Dec, 2004).
11. J. M. Shively, G. van Keulen, W. G. Meijer, *Annu. Rev. Microbiol.* **52**, 191 (1998).
12. J. J. Robinson, C.M. Cavanaugh, *Limnol. Oceanogr.* **40**, 1496 (1995).
13. C. M. Cavanaugh, J. J. Robinson, in *Microbial Growth on C^1 Compounds* M. E. Lidstrom, F. R. Tabita, Eds. (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996) pp. 285-292.
14. R. G. Kemp, R. L. Tripathi, *J. Bacteriol.* **175**, 5723 (Sep, 1993).
15. H. R. Beller *et al.*, *J. Bacteriol.* **188**, 1473 (Feb, 2006).
16. C. Dahl *et al.*, *J. Bacteriol.* **187**, 1392 (Feb, 2005).
17. J. A. Eisen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9509 (Jul 9, 2002).
18. F. Verte *et al.*, *Biochemistry* **41**, 2932 (Mar 5, 2002).
19. A. Fiala-medioni, C. Metivier, *Marine Biology* **90**, 215 (1986).
20. J. J. Childress, C. R. Fisher, J. A. Favuzzi, N. K. Sanders, *Physiol. Zool.* **64**, 1444 (Nov-Dec, 1991).
21. D. Nelson, C. Fisher, in *Microbiology of deep-sea hydrothermal vents* D. Karl, Ed. (CRC Press, Boca Raton, 1995).
22. F. Zal *et al.*, *Cah. Biol. Mar.* **41**, 413 (2000).
23. H. Kumagai, T. Fujiwara, H. Matsubara, K. Saeki, *Biochemistry* **36**, 5509 (May 6, 1997).
24. K. S. Johnson, J.J. Childress, R.R. Hessler, C.M. Sakamoto-Arnold, C.L. Beehler, *Deap-Sea Res.* **35**, 1723 (1988).
25. A. P. Wood, J. P. Aurikko, D. P. Kelly, *FEMS Microbiol. Rev.* **28**, 335 (Jun, 2004).
26. P. S. Fritsch, M. L. Urbanowski, G. V. Stauffer, *J. Bacteriol.* **182**, 5539 (Oct, 2000).
27. I. M. Keseler *et al.*, *Nucleic Acids Res.* **33**, D334 (Jan 1, 2005).
28. C. J. Krieger *et al.*, *Nucleic Acids Res.* **32**, D438 (Jan 1, 2004).
29. M. E. Streams, C. R. Fisher, A. Fiala-Medioni, *Mar. Biol.* **129**, 465 (Sep, 1997).

30. N. W. Goehring, J. Beckwith, *Curr. Biol.* **15**, R514 (Jul 12, 2005).

31. W. J. Brown, D. D. Rockey, *Infect. Immun.* **68**, 708 (Feb, 2000).

32. A. J. McCoy, A. T. Maurelli, *Trends Microbiol.* **14**, 70 (Feb, 2006).

33. unpublished observation, C.M. Cavanaugh.

34. M. R. Hammerschlag, *Semin. Pediatr. Infect. Dis.* **13**, 239 (Oct, 2002).

35. A. S. Peek, R. C. Vrijenhoek, B. S. Gaut, *Mol. Biol. Evol.* **15**, 1514 (Nov, 1998).

36. J. A. Eisen, P. C. Hanawalt, *Mutat. Res.* **435**, 171 (Dec 7, 1999).

37. W. Martin *et al.*, *Nature* **393**, 162 (May 14, 1998).

38. N. A. Moran, P. H. Degnan, *Mol. Ecol.* **15**, 1251 (Apr, 2006).

39. Table S2 available on Science Online

Acknowledgements

This research was funded by a US Department of Energy grant to Cavanaugh and Eisen and a Howard Hughes Medical Institute Predoctoral Fellowship to Newton. Sequencing was carried out at the Joint Genome Institute and we thank David Bruce and Eddy Rubin for project management. We thank Dan Fraenkel for his review of metabolic pathways in *Ruthia* and Peter Girguis and Kathleen Scott for their helpful comments on the manuscript.

Figure 1. Electron micrographs of *Ruthia magnifica* within host bacteriocytes. (A) Bacteriocyte containing many small (0.3 μm) coccoid-shaped symbionts. Scale bar = 5 μm (B) Higher magnification of *R. magnifica* showing the electron dense granules suggestive of *Chlamydia's* "elementary bodies." Scale bar = 2 μm *D.* symbiont in putative degradative state, *N*, bacteriocyte nucleus, *R, R. magnifica.*

Figure 2. The percentages of the genomes dedicated to different functional categories as predicted by annotation are shown for γ-proteobacterial symbionts (*Ruthia magnifica, Buchnera aphidicola*) and free-living relatives (*Thiomicrospira crunogena* and *Escherichia coli,* respectively).

Figure 3. Three major metabolic pathways are shown as inferred from the genomic content in *R. magnifica*. Enzymes or pathways present in the genome are colored while those not yet identified are either white or dashed. The Calvin cycle is used by the symbiont for carbon fixation and although missing fructose 1,6-bisphosphatase (FBPase) and sedoheptulose 1,7-bisphosphatase (SBPase), it could use a reversible phosphofructokinase to regenerate ribulose 5-phosphate. The sulfur oxidation pathway appeared similar to that of *Chlorobium tepidum*. The Sox proteins act in the periplasm to oxidize thiosulfate while sulfide may be oxidized intracellularly by the reversible dissimilatory sulfate reductase

(dsr) system. Nitrogen assimilation pathways via both ammonia and nitrate are present in the symbiont genome.

Figure 1.

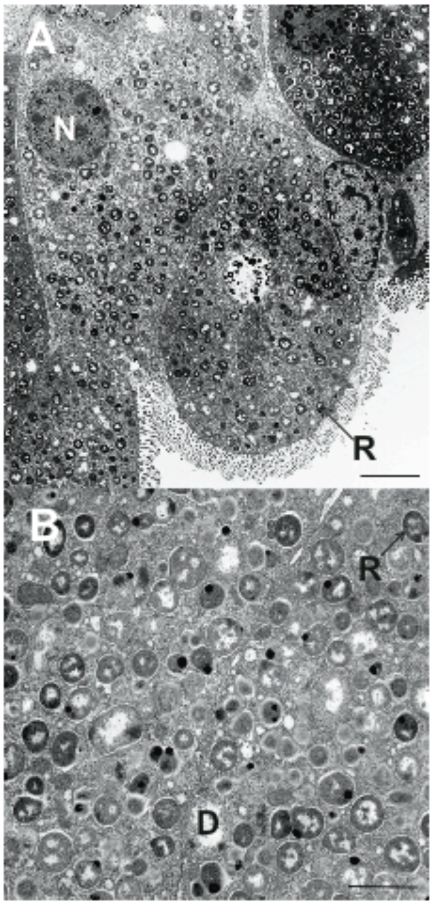
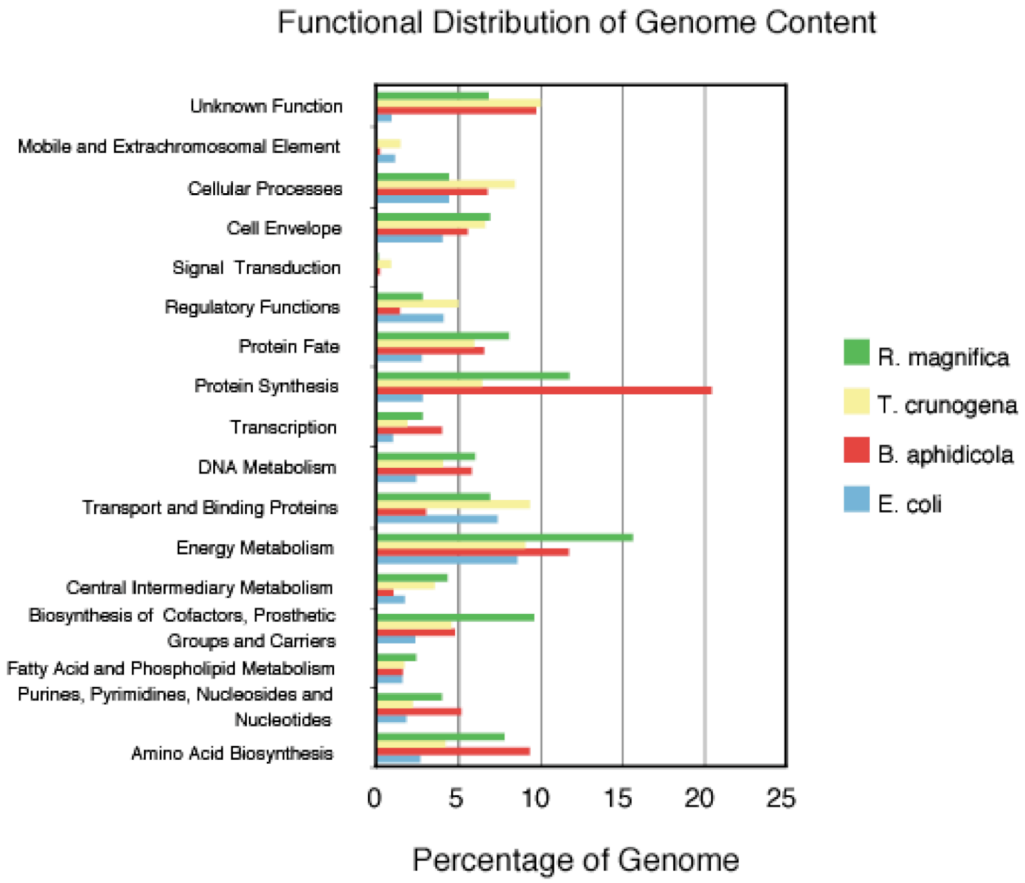


Figure 2


Functional Distribution of Genome Content
Unknown Function
Mobile and Extrachromosomal Element
Cellular Processes
Cell Envelope
Signal Transduction
Regulatory Functions
Protein Fate
Protein Synthesis
Transcription
DNA Metabolism
Transport and Binding Proteins
Energy Metabolism
Central Intermediary Metabolism
Biosynthesis of Cofactors, Prosthetic Groups and Carriers
Fatty Acid and Phospholipid Metabolism
Purines, Pyrimidines, Nucleosides and Nucleotides
Amino Acid Biosynthesis
0
5
10
15
20
25
Percentage of Genome
🟩 R. magnifica
🟨 T. crunogena
🟥 B. aphidicola
🟦 E. coli

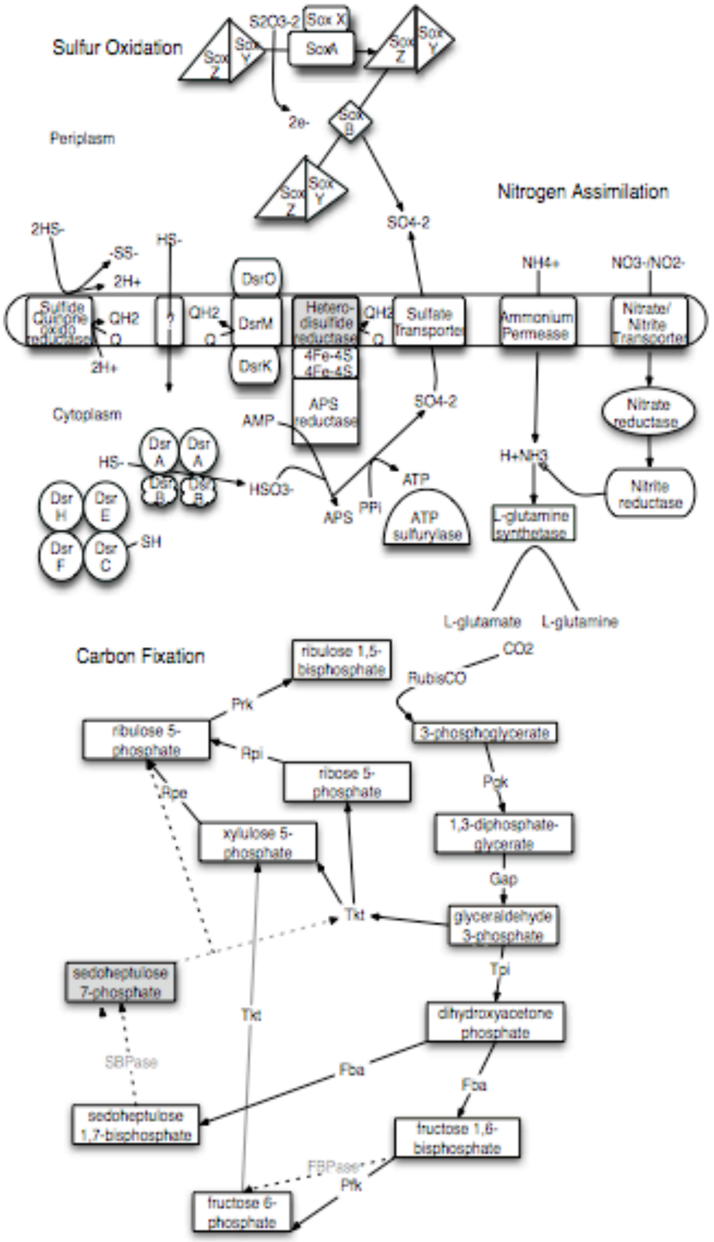

Sulfur Oxidation
S2O3-2
Sox X
SoxA
Sox Z
Sox Y
2e-
Sox B
Periplasm
Nitrogen Assimilation
SO4-2
2HS-
-SS-
2H+
HS-
DsrO
Sulfide Quinone oxido reductase
QH2
Q
?
DsrM
Hetero-disulfide reductase
Sulfate Transporter
Ammonium Permease
Nitrate/ Nitrite Transporter
NH4+
NO3-/NO2-
DsrK
4Fe-4S
4Fe-4S
APS reductase
Cytoplasm
AMP
Dsr A
Dsr B
HS-
HSO3-
APS
PPi
ATP
ATP sulfurylase
Dsr H
Dsr E
Dsr F
Dsr C
SH
Nitrate reductase
Nitrite reductase
H+NH3
L-glutamine synthetase
L-glutamate
L-glutamine
Carbon Fixation
ribulose 1,5-bisphosphate
CO2
RubisCO
Prk
ribulose 5-phosphate
3-phosphoglycerate
Rpi
ribose 5-phosphate
Pgk
Rpe
xylulose 5-phosphate
1,3-diphosphate-glycerate
Gap
Tkt
glyceraldehyde 3-phosphate
Tpi
sedoheptulose 7-phosphate
Tkt
dihydroxyacetone phosphate
SBPase
Fba
Fba
sedoheptulose 1,7-bisphosphate
fructose 1,6-bisphosphate
FBPase
Pfk
fructose 6-phosphate

Figure 3

Table 1. General genome features of the chemoautotrophic symbiont *Ruthia magnifica* compared with those of other γ-proteobacteria, including the free-living chemoautotroph, *Thiomicrospira crunogena,* an obligately intracellular aphid symbiont, *Buchnera aphidicola,* and a free-living relative of the aphid symbiont*,* *Escherichia coli*.

Features	*Ruthia magnifica*	*Thiomicrospira crunogena*	*Buchnera aphidicola*	*Escherichia coli*
Chromosome, Mb	1.2	2.4	0.6	4.6
Plasmids	0	0	1	0
G+C content, %	34	43	26	50
Total gene number	1248	2199	608	4289
rRNAs	3	9	3	22
tRNAs	36	43	32	88
Protein-coding, %	81.4	97.8	86.5	97.9
Mean gene length, bp	975	948	991	800

- *E. coli* is closely related to *B. aphidicola,* with 87.2% sequence identity in 16S rRNA; *T. crunogena* and *R. magnifica* share 83.3% 16S rRNA sequence identity.

Supporting online material

Supplementary Table 2. The *Ruthia magnifica* genome encodes pathways for many metabolic processes and biosynthesis of important amino acids, vitamins and cofactors. Complete pathways found in the genome are indicated by ‘+’ while absent pathways are indicated by ‘-‘.

Pathway	Prediction
Glycolysis	+
TCA cycle	+
Glyoxylate shunt	Partial
Respiration	+
Pentose phosphate pathway	Partial
Fatty acid biosynthesis	+
Cell wall biosynthesis	Partial
Biosynthesis of all 20 amino acids	+
Vitamin and Cofactor Biosynthesis	
Heme	+
Ubiquinone	+
Nicotinate and nicotinamide	+
Folate	+
Lipoate	+
Riboflavin	+
Pantothenate	+
Pyridoxine	+
Thiamine	+
Biotin	+
Cobalamin	-