2010

Metabolic flexibility revealed in the genome of the cyst-forming α-1 proteobacterium Rhodospirillum centenum

Yih-Kuang Lu  
*Arizona State University at the Tempe Campus*

Jeremiah Marden  
*Indiana University - Bloomington*

Mira Han  
*Indiana University - Bloomington*

Wesley D. Swingley  
*University of California - Merced*

Stephen D. Mastrian  
*Translational Genomics Research Institute*

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.wustl.edu/open_access_pubs](https://digitalcommons.wustl.edu/open_access_pubs)

Part of the [Medicine and Health Sciences Commons](https://digitalcommons.wustl.edu/open_access_pubs)

**Recommended Citation**

Lu, Yih-Kuang; Marden, Jeremiah; Han, Mira; Swingley, Wesley D.; Mastrian, Stephen D.; Chowdhury, Sugata Roy; Hao, Jicheng; Helmy, Tamer; Kim, Sun; Kurdoglu, Ahmet A.; Matthies, Heather J.; Rollo, David; Stothard, Paul; Blankenship, Robert E.; Bauer, Carl E.; and Touchman, Jeffrey W., "Metabolic flexibility revealed in the genome of the cyst-forming α-1 proteobacterium Rhodospirillum centenum." *BMC Genomics*, 325. (2010).  
[https://digitalcommons.wustl.edu/open_access_pubs/50](https://digitalcommons.wustl.edu/open_access_pubs/50)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
Metabolic flexibility revealed in the genome of the cyst-forming α-1 proteobacterium *Rhodospirillum centenum*

Yih-Kuang Lu, Jeremiah Marden, Mira Han, Wesley D Swingley, Stephen D Mastrian, Sugata Roy Chowdhury, Jicheng Hao, Tamer Helmy, Sun Kim, Ahmet A Kurdoglu, Heather J Matthies, David Rollo, Paul Stothard, Robert E Blankenship, Carl E Bauer and Jeffrey W Touchman*

**Abstract**

**Background:** *Rhodospirillum centenum* is a photosynthetic non-sulfur purple bacterium that favors growth in an anoxygenic, photosynthetic N2-fixing environment. It is emerging as a genetically amenable model organism for molecular genetic analysis of cyst formation, photosynthesis, phototaxis, and cellular development. Here, we present an analysis of the genome of this bacterium.

**Results:** *R. centenum* contains a singular circular chromosome of 4,355,548 base pairs in size harboring 4,105 genes. It has an intact Calvin cycle with two forms of Rubisco, as well as a gene encoding phosphoenolpyruvate carboxylase (PEPC) for mixotrophic CO2 fixation. This dual carbon-fixation system may be required for regulating internal carbon flux to facilitate bacterial nitrogen assimilation. Enzymatic reactions associated with arsenate and mercuric detoxification are rare or unique compared to other purple bacteria. Among numerous newly identified signal transduction proteins, of particular interest is a putative bacteriophytochrome that is phylogenetically distinct from a previously characterized *R. centenum* phytochrome, Ppr. Genes encoding proteins involved in chemotaxis as well as a sophisticated dual flagellar system have also been mapped.

**Conclusions:** Remarkable metabolic versatility and a superior capability for photoautotrophic carbon assimilation is evident in *R. centenum*.

**Background**

*Rhodospirillum centenum* (also known as *Rhodocista centenaria*) is a thermotolerant α-1 proteobacterium that is closely related to species of the *Azospirillum* genus [1-4]. *R. centenum* is one of the few known thermotolerant purple bacteria species. It has an optimal growth temperature of 44°C, and is capable of differentiating into metabolically dormant cysts that can survive at temperatures as high as 65°C [1-3,5]. Consequently, *R. centenum* can often be cultivated from hot springs such as those found at Yellowstone National Park [3]. *R. centenum* metabolizes a unique set of carbon sources, but is unable to utilize malate or other C4 dicarboxylic acids [2]. Unlike *Rhodobacter capsulatus* and other known purple non-sulfur bacteria, *R. centenum* does not repress photosystem synthesis in the presence of molecular oxygen [6].

Three morphologically distinct cell types are observed during the *R. centenum* life cycle; swim cells, swarm cells and metabolically dormant cyst cells [1,7]. Proteobacterial encystment has been reported in diverse species, and is one of several prokaryotic resting cell strategies employed for surviving environmental stress. Physiological aspects of cyst cell development have been well studied in *R. centenum*, *Azotobacter vinelandii* and *Azospirillum brasilense*, several features of which are shared by these species [4,5,8-10]. Environmental stresses, including nutrient deprivation, trigger vegetative cells to undergo a multi-stage transition into rounded, immotile cells encapsulated by complex, protective outer coats (Figure 1B). These cysts are additionally typified by the presence of large intracellular granules of the induc-
trially significant polymer poly-hydroxybutyrate (PHB) [11]. The resulting cells have extreme desiccation resistance and also afford modest protection from stresses such as heat and UV light. While the morphological and resistive aspects of such cysts have been well studied, mechanisms that underlie the regulation of this process remain largely unknown.

Members of the cyst forming Azospirillum genera have significant agricultural importance. Specifically, the aerobic nitrogen fixing species A. brasilense and A. lipoferum (both closely related to R. centenum) are known to associate with, and stimulate the growth of, numerous grasses and cereals [12]. These bacteria may benefit such plants through their ability to aerobically fix nitrogen [12]. Interestingly, mutations that affect swim cell to swarm cell differentiation [13,14] and cyst-cell development [15] also affect plant root colonization. Thus, a better understanding of these cellular differentiation events is clearly warranted. A genome sequence advances the tools available for studies of these processes.

Results and Discussion

Genome Properties
R. centenum strain SW has a single 4,355,548-bp circular chromosome containing 4,105 genes including 4,002 open reading frames (ORFs), 12 rRNA genes, 52 tRNA genes, but no native plasmids (Table 1 and Figure 1A). The genomic G+C content of R. centenum (70.5%) is higher than the genome of the purple non-sulfur bacterium Rhodospirillum rubrum ATCC 11170 (65.4%; GenBank accession CP000230), its closest sequenced relative. Additionally, we found that 76% percent of R. rubrum proteins have a corresponding match in the R. centenum genome (e-value = 0.0001; Figure 1A). The total protein-coding content of the chromosome is 86% and the aver-

Table 1: General genome characteristics

| Characteristic                      | Value       |
|------------------------------------|-------------|
| Genome size                        | 4,355,548 bp|
| G+C content                        | 70.5%       |
| Genes                              | 4,105       |
| Proteins                           | 4,002       |
| Pseudogenes                        | 39          |
| Coding density                     | 86%         |
| Average gene length                | 945 nt      |
| Structural RNAs                    | 64          |
| Number of CDS with a predicted function | 2,633 |
| Number of CDS similar to proteins of unknown function | 1,135 |
| Number of CDS without similarities | 234         |
| Number of insertion sequences      | 153         |
age gene length is 945 nucleotides. There are 2,633 predicted proteins that have sequence similarity to other known proteins, while 1,135 (28%) are homologous to proteins of unknown function and a further 234 (6%) are unique to *R. centenum*. The entire genome sequence has been fully annotated and a summary of genes in major functional categories is shown in Table 2. GC skew analysis failed to present a strong inflection point indicative of an origin of replication (Figure 1A). However, the replication-related genes *dnaA* and *dnaN* begin at base pair positions 3,127,698 and 3,134721, respectively, which may suggest the general location of an origin.

Detailed metabolic pathway construction for *R. centenum* was generated from the annotated genome using the Pathway Tools suite of programs [16,17]. The predicted metabolic networks were further validated and adjusted with information from the literature that contained experimental data from taxonomically related bacteria. Based on the constructed metabolic scheme, 1,295 known enzymatic reactions involved in 263 pathways are predicted. A total of 985 putative metabolic compounds are predicted to exist in the metabolism of *R. centenum*, which appears to possess many metabolic pathways typical of purple photosynthetic bacteria in the α-proteobacteria class. Despite the existence of multiple nutrient metabolisms, no sulfide metabolism was identified, which verifies an earlier observation that *R. centenum* is incapable of reducing sulfide [18,19]. Nevertheless, genes required for anaerobic respiration involved in nitrate and fumarate reduction were identified.

### Table 2: Functional categories of *Rhodospirillum centenum* genes

| Gene functional category | No. of genes |
|--------------------------|--------------|
| Energy and central intermediary metabolisms | 559 |
| Fatty acid and phospholipid metabolism | 110 |
| Purines, pyrimidines, nucleosides, and nucleotides | 70 |
| Transport | 448 |
| Regulatory function and Signal transduction | 396 |
| Protein synthesis and fate | 340 |
| Cofactor biosynthesis | 166 |
| Amino acid synthesis | 125 |
| Cellular processes and Envelope proteins | 628 |
| DNA metabolism | 78 |
| Transcription | 71 |
| Unknown function and hypothetical proteins | 1142 |

*Some genes are included in more than one category.*

**Carbon fixation**

The genome sequence of *R. centenum* reveals a versatile capacity for carbon fixation. Two groups of Rubisco encoding genes (tentatively named *cbbL1S1* and *cbbL2S2*) are present within two putative operons located at distal positions of the chromosome. The first is linked with genes coding for proteins typically involved in the Calvin-Benson cycle, such as phosphoribulokinase (*prk*), whereas the second is associated with two genes (*cbbO* and *cbbQ*) that encode Rubisco activation proteins (Figure 2A). Both *cbb* operons are likely regulated by LysR-family transcription factors (*CbbR1* and *CbbR2*), whose corresponding genes are located immediately upstream of each respective operon. Though seemingly rare, some species of bacteria have multiple forms of Rubisco [reviewed in [20]]. A phylogenetic analysis of 18 phototrophic bacteria demonstrates that *R. centenum* possesses Rubisco subtypes IAq and IC, both of which are found predominantly in proteobacteria (Figure 2B). The IC form of Rubisco is found primarily among α/β-proteobacteria while form IAq is predominantly found in chemolithotrophic β/γ-proteobacteria, with the exception of *Rhodopseudomonas palustris* BisB5.

Different kinetic properties reported for these various Rubisco forms have led to speculation that each is adapted for use in specific environmental CO2 concentrations, and that possession of multiple Rubisco forms may be advantageous [20]. For instance, Form IC Rubisco has a slightly lower reaction rate (*k_{cat}~2-3.2 s^{-1}* ) than IAq (*k_{cat}~3.7 s^{-1}*) , suggesting that they are adapted to medium-to-high and medium-to-low [CO2] environments, respectively [21]. The need for such extreme metabolic flexibility is reflected by the wide range of environments that non-sulfur purple bacteria inhabit. The mechanism of switching between two Rubisco forms and the roles of the dedicated regulatory proteins CbbR1 and CbbR2 are unclear in these bacteria.

Phosphoenolpyruvate carboxylase (PEPC, *RC1_2446*) and pyruvate orthophosphate dikinase (*Pdk, RC1_1667*) are present in the *R. centenum* genome. These enzymes are widespread in plants and bacteria. PEPC and Pdk (with malic enzyme) are thought to be responsible for heterotrophic carbon dioxide assimilation in *Roseobacter denitrificans* since this bacteria does not contain Rubisco [22]. However, PEPC and Pdk are also found in species that primarily use Rubisco for carbon fixation, such as *R. palustris*. Thus, the function that PEPC and Pdk perform in *R. centenum*, as well as in related Rubisco-containing purple bacteria, is unclear. An *R. palustris* PEPC-deficient strain does exhibit a slower doubling time compared with the wild-type strain grown anaerobically in the light and aerobically in the dark when pyruvate is used as a carbon source [23]. Thus, autotrophic bacteria...
like *R. centenum* that cannot acquire C₄ dicarboxylic acids heterotrophically may have evolved an anaplerotic assimilation to ensure a continuous replenishment of C₄-dicarboxylic acids needed for amino acid biosynthesis.

An analysis of the α-proteobacteria class shows that only four anoxygenic photosynthetic species are known to possess Pk, Pdk, and PEPS (phosphoenolpyruvate synthase) together (Additional File 1, Figure S1). These include two members of the *Bradyrhizobium* family, *Hoe-
flea phototropica*, and *R. centenum*. The others contain only Pk and Pdk, or just Pdk (e.g., the mutualistic parasites of *Rickettsiales*). The four containing Pk, Pdk, and PEPS are reported to grow poorly with pyruvate, malate, and various other dicarboxylic acids, indicating a strong...
dependence on carbon source such as we observe with *R. centenum* [2,24,25].

Figure 3 presents a proposed scheme in which the pyruvate/PEP interconversion driven by pyruvate kinase (Pk), pyruvate orthophosphate dikinase (Pdk), and phosphoenolpyruvate synthase (PEPS) collaborate functionally to modulate carbon flux in *R. centenum*. It is based on previous experimental data illustrating the control of pyruvate/PEP interconversion under different trophic conditions in *R. denitrificans*, *R. rubrum*, and Archaea [22,26,27]. When aerobic respiration is suppressed due to the lack of oxygen, Pk is functionally replaced by Pdk that continuously supplies pyruvate [26,27]. We speculate that PEPS collaborates with both Pk and Pdk under nitrogen-fixing conditions, where it supplies a stable supplement of dicarboxylic acids through an internal pyruvate pool for amino acid biosynthesis. PEPS-driven gluconeogenesis in *R. centenum* may contribute to the balance of carbon flux for the non-oxidative pentose phosphate pathway when the rate of CO₂ fixation is limited.

**Metabolism**

*R. centenum* is an active nitrogen fixer. The *nif* genes of *R. centenum* are located in two distant regions: the first region consists of 22 genes that essentially include *nif*IX-ENKDHTZBA VW for nitrogenase biosynthesis, *modBC* for molybdenum transport, and *fixABCX* for electron transport. Phylogenetic analysis using a concatenated alignment of the nitrogenase structural genes *nifHDK* shows that *R. centenum* falls within a clade containing *R. sphaeroides*, *R. rubrum* and several Rhizobiales (not shown). This result suggests that this group of nitrogenases originated from a common ancestor.

Neither iron nor vanadium nitrogenases, coded by *anf* and *vnfVHDK*, respectively, were identified in the *R. centenum* genome. In *R. rubrum* the regulatory enzymes dinitrogenase reductase ADP-ribosyl transferase (DRAT) and dinitrogenase reductase-activating glycohydrolase (DRAG) regulate nitrogenase activity by reversible ADP-ribosylation of NifH [28,29]. An absence of these genes in *R. centenum* suggests that these organisms have different environmental requirements for nitrogen fixation. While *R. centenum* possesses a nitrogenase for nitrogen fixation, utilization of inorganic nitrogen compounds as an alternative nitrogen source is restricted. Unlike *R. denitrificans*, which has a large complement of genes for nitrogen metabolism (but not nitrogenase), the *R. centenum* genome does not contain genes which encode an assimilatory nitrite reductase (*nirSCFDGHJN*) for the denitrification pathway. Yet, the nitrate reduction by periplasmic nitrate reductase (*napABCDEF*) seems to be intact.

We identified the presence of two cytosolic detoxifying enzymes, arsenate and mercuric reductase. To our knowledge, the latter has never been reported in a purple bacterium. Two copies of arsenate reductase, encoded by *arsC1* and *arsC2*, are present (RC1_2995 and RC1_3700). One copy is associated with genes encoding an arsenate efflux pump (*arc3*) and an arsenic resistance repressor (*arsR*). In both ArsC proteins, the cysteine residues that presumably bind arsenate are conserved, but the protein sequences share less than 20% overall identity. The presence of *arc3* and *arsR* implies that arsenate reductase is an inducible enzyme when arsenate is present. There is one copy of mercuric reductase encoding *merA* (RC1_2279). Generally, these two detoxifying systems are energy dependent, with arsenate reductase using either...
thioredoxin or glutaredoxin while mercuric reductase uses NADPH [30].

Cyst formation
To date, the only concerted effort at finding regulators of encystment (Figure 1B) in any bacterial species has been undertaken in *R. centenum*, where a screen for Tn5 mutants displaying de-repressed encystment on nutrient-rich media uncovered several such components [31]. A subset of identified elements lay within an operon of chemotaxis-like genes (*che3*), individual in-frame deletions of which exhibit opposing premature ‘hyper-cyst’ or delayed ‘hypo-cyst’ phenotypes [32]. Predicted signaling components of note in this cluster include a small receiver domain protein (CheY3; RC1_2133), a hybrid sensor kinase-receiver CheA homolog (CheA3; RC1_2127) and a similarly hybridized kinase-receiver protein (CstS3; RC1_2124) that also contains a PAS sensory domain. As ultimate control over cyst cell development no doubt occurs at the transcriptional level, the phosphorylation of a classic response regulator, or does so by means of a more indirect mechanism remains to be elucidated. This screen also identified two genomically orphaned, cytoplasmic sensor kinases (CstS1, RC1_2847 and CstS2, RC1_2047), both predicted to contain PAS and PAC sensory motifs with CstS1 also containing a GAF and a receiver domain [31]. Deletion of *cstS1* and *cstS2* have contrasting respective hypo-cyst and hyper-cyst phenotypes, and epistatic analyses of these genes and *che3* components indicate a complex signaling hierarchy into which contributions are undoubtedly made by hitherto unknown regulatory elements (Berleman and Bauer, Unpublished Data). In fact further screens in our lab have uncovered several such regulators, including two sensor kinases, RC1_0896 and RC1_3465 (Marden and Bauer, Unpublished Data). These genes were independently identified by a similar screen in a separate lab where they are currently the subject of genetic characterization (Bird, Manuscript in preparation).

Photosynthetic and chlorophyll biosynthetic proteins
Despite striking dissimilarity in the genomic organization of photosynthesis genes in different photosynthetic species, most of the genes that carry out bacterial chlorophyll and carotenoid biosynthesis in *R. centenum* are found in a single photosynthetic gene cluster (PGC; Figure 4). The photosynthesis genes are organized into seven major operons. The gene cluster *hemA-puhH-acsF-puh-CBA-bch-lhaAb-chLMHBNF* is located immediately downstream of *pufMLAB-bchZYXC-crtFEDCBI* in *R. centenum*. This is in contrast to *Bradyrhizobium* sp. where it maps immediately upstream of *aerR-ppsR1-bchG2P*. The carotenoid biosynthesis gene *crtA* found in *Rhodobacter capsulatus* (among others) is not found in *R. centenum* (or *Bradyrhizobium* sp). The bch/heme biosynthesis genes *acsFl, puhE, hemA*, and the *cyc2* gene encoding cytochrome *c*, are present in the genome. Thus, the overall organization of the *R. centenum* PGC is similar to the PGC of *Bradyrhizobium* sp. but not closely related to that of *Rhodobacter* species. We also found that *R. centenum* and *R. rubrum* do not share contiguity of their PGCs, where the *R. rubrum* PGC is separated into two clusters in distant regions of the chromosome.

Photoreceptors
The *R. centenum* *ppr* gene represents the first bacteriophytochrome identified in purple bacteria [33]. Distinctive characteristics of Ppr, with respect to other bacteriophytochromes, have been discussed in detail elsewhere [reviewed in [34]]. In addition to *ppr*, a second gene (RC1_3803) is predicted to encode an additional bacteriophytochrome. RC1_3803 does not possess a photoactive yellow protein (YPY) domain, and has 45% and 55% similarity to both the photosensory core domain (PCD) and histidine kinase domain (HKD) of Ppr, respectively (not shown). A search of public protein databases identified a number of bacteriophytochromes that show homology to RC1_3803 (not shown). Based on characteristics of other bacteriophytochromes, we hypothesize that RC1_3803 may absorb near far-red light, as that wavelength of light is reported to promote negative phototaxis of *R. centenum* and is in the region of the spectrum where other bacteriophytochromes exhibit spectral absorbance [7], reviewed in [34].

Finally, there are two genes coding for flavin-binding photoreceptors. One gene (RC1_2193) putatively codes for a small blue light photoreceptor utilizing a flavin (BLUF) and a second protein (RC1_0351) putatively encodes a histidine kinase containing a light-oxygen-volt-age (LOV) domain. Both of these putative photoreceptors likely utilize FAD or FMN as a chromophore to absorb blue light to promote a conformational change to elicit an output response. Neither protein has been genetically disrupted, but they may play a role in controlling light regulated physiology or behavior in *R. centenum*.

Flagella
*R. centenum* synthesizes two flagella, a constitutive polar flagellum for swimming motility and inducible lateral flagella required for swarming motility on viscous or solid media [35]. We identified 72 flagella genes in the *R. centenum* genome distributed among five major flagellar gene clusters (FGCs) at various regions along the chromosome (Figure 5). Most structural genes are duplicated while
flgC, flgF, flhF, fliL, and fliN have either three or four copies each. Several genes involved in regulation, export or assembly (flhF, fliO, fliX, flaA, flaG, flbD, and fleN) are present as a single copy.

Lateral and polar flagellar systems have diverged twice in α-proteobacteria as well as in the common ancestor of β/γ-proteobacteria [36]. Many of the duplicated fla genes in R. centenum exhibit poor sequence similarity to their reciprocal pair indicating a high degree of diversity among the lateral and polar flagella genes (Additional File 1, Table S1). It also does not rule out that either the polar or lateral flagella genes may have been derived by lateral gene transfer. A phylogenetic tree using a concatenated alignment of eleven flg genes from R. centenum indicates that the lateral flagellar system of R. centenum indeed has a distinct origin from the polar flagellar system (Figure 6). The four small clusters (FGC2, FGC3, FGC4 and FGC5) that map to different positions on the chromosome have subunits predicted to constitute components of the polar flagellum. Structural components of the lateral flagella are predicted to reside among the large FGC1 cluster.

We have obtained several mutations in the transcription factor FlbD that disrupt synthesis of the polar, but not lateral flagella (D. Rollo and C. Bauer, unpublished results). These results indicate that expression of the polar and lateral flagella genes are distinct. Also, several insertion mutations that affect synthesis of both the lateral and polar flagellum map to components of the type III export system comprised of FliI, FlhA, FlhB and FliL, indicating that this export apparatus is used by both flagella systems.

There are 13 additional genes that include three functionally identified, four functionally predicted, and six of unknown function distributed among the fla clusters. These include genes encoding a CheY-like receiver protein (RC1_0209), lytic muramyl transglycoylase (RC1_0192), DNA polymerase III (RC1_0787), PPE-repeat proteins (RC1_0178), tetratricopeptide TPR_2 (RC1_0215), DNA binding protein (RC1_0222), and an ATPase involved in DNA repair (RC1_0187).

Chemotaxis and Signal Transduction

The three previously identified operons (Che1, Che2 and Che3) encoding chemotaxis-like proteins were confirmed to represent the entirety of such ‘Che-like’ clusters in R. centenum [32,37,38]. Each contains homologs of the E. coli chemotactic proteins CheA, CheY, CheW, CheB and CheR, with an additional CheW in the Che3 cluster (CheW3a and CheW3b) and a tripartate CheW in the Che2 cluster composed of three distinct CheW domains. Atypical CheA sensor kinases are also present in the Che1 and Che2 clusters, as each is hybridized to a receiver domain. Also of note are genes encoding non-canonical chemotactic components, two small proteins of unknown function in the Che2 cluster (RC1_0336 and...
RC1_0341) and an additional sensor kinase receiver domain hybrid (CstS3; RC1_2124) in the Che3 cluster. The functions of all three chemotaxis operons have been elucidated. Both chemotactic and phototactic behavior in *R. centenum* are under Che1 control, as strains with Che1 component disruptions have motility phenotypes similar to those of *E. coli* chemotactic mutants [39]. The Che2 cluster is involved in lateral flagella biosynthesis, as strains deleted of Che2 components are either hyper-flagellated or lack flagella completely, but remain chemotactic [32]. Lastly, the Che3 cluster directs timing of encystment, as deletions of Che3 components produce strains that are either early or delayed in cyst cell development [37]. The molecular mechanisms by which the latter two clusters achieve such altered functions are yet to be elucidated.

Whereas CheA, CheW, CheB and CheR homologs were only identified within these chemotaxis clusters (a CheB-CheR hybrid, RC1_3878, was discovered), 19 genes encode for small, stand-alone CheY-like receiver domain proteins. Besides the three Che cluster-associated CheY encoding homologs, two of these *cheY*-like genes are associated with neighboring genes encoding chemosensory proteins. The first (RC1_0955) is adjacent to one of two putative CheZ phosphatases (RC1_0954 and RC1_0882). The second (RC1_0353) is part of a potential operon with genes encoding a small hypothetical protein (RC1_0352), a methyl accepting chemotaxis domain protein (MCP; RC1_0354) and a homolog of the response regulator *fixJ* (RC1_0355). Whether these stand-alone receiver domains play a part in chemotactic signal transduction will require significant genetic and biochemical characterization.

A total of 33 genes encoding MCP domains were identified in our analysis, only three of which have been previously genetically characterized. The first identified, Ptr, was shown to be responsible for photosensory perception in *R. centenum* [40]. MCP2 and MCP3, which are respectively and functionally associated with the aforementioned Che2 and Che3 clusters, have functions independent of taxis, since gene deletion strains are chemotactic but instead are either hyper-flagellated (Δmcp2) or delayed in encystment (Δmcp3) [32,37]. As discussed above, a methyl-accepting sensory transducer is also tightly linked with a homolog of *fixJ*, which is known to be a transcription activator for nitrogen fixation in microaerobic conditions [41]. We speculate that this MCP may thus have a role in altering bacterial swimming to approach the microaerobic conditions that are optimal for nitrogen fixation. The relatively large number of new and uncharacterized MCP domain-containing proteins suggests a capacity to sense and respond to a
wide variety of extracellular signals, either with a tactic response, or with the alternate functions controlled by the Che2 and Che3 pathways.

In our analysis of two-component signal transduction, 55 identified genes are predicted to encode sensor kinases, 16 of which are hybridized to receiver domains. Excluding the latter group, 54 proteins are predicted to contain receiver domains. Of these 19 are stand-alone receiver domains with the remainder fused to assorted sensory and/or output domains. Lastly, 8 predicted proteins contain a histidine-containing phosphotransfer domain; three Class II histidine kinases (all R. centenum CheA proteins), two within hybrid Class I histidine kinases (RC1_0633 and RC1_2262), two stand-alone domains (RC1_2126 and RC1_3033) and one associated with a receiver domain (RC1_1779).

Conclusions
Analysis of the R. centenum genome demonstrates that both Rubisco- and PEPC- derived carbon assimilation can compensate for the inability to utilize malate or other C_4 dicarboxylic acids in the R. centenum environment. Many newly identified genes that are discussed in this report have advanced our knowledge of the structure and origin of the R. centenum PGC, the complex life cycle involving differentiation from swim to swarm cells, and the differentiation into heat and desiccation resistant resting cysts. R. centenum also contains many sensory

---

**Figure 6 Maximum likelihood tree based on conserved flagellar proteins depicting the position of R. centenum flagellar systems among other flagellated bacteria.** The tree is based on a concatenated alignment of 11 flagellar proteins (Flg) that are present in all primary and secondary flagellar systems [36]. A blue “1” or “2” after the species name indicates primary and secondary flagella systems, respectively. Alpha I (alphaproteobacteria group I) and Alpha II (alphaproteobacteria group II) are boxed. The primary (polar) and secondary (lateral) flagella of R. centenum are in red.
proteins such as bacteriophytochromes that control gene expression in response to complex environmental stimuli. The completion of the *R. centenum* genome impacts the study of cyst cell development in particular, already allowing the identification of an *A. brasilense* *flcA* homolog. FlcA is a transcriptional regulator of *A. brasilense* encystment, and appears to have a role in *R. centenum* encystment (Marden and Bauer, Unpublished Data). The sequenced genome has also allowed for the discovery of an additional sensor kinase involved in encystment, RC1_2747. This gene was originally disrupted and identified in a hyper-cyst screen, however the transposition occurred in a region with low sequence similarity and was placed in a class of unknown genes [5].

*R. centenum* is emerging as a model organism for molecular genetic analysis of cyst formation, photosynthesis, phototaxis, and cellular development. This species is genetically amenable, with a variety of genetic tools already developed to explore these processes. The generation of a complete and annotated genome sequence establishes the genetic infrastructure for such studies, provides a framework to organize all the genetic information about the organism, and catalyzes future 'omics' research.

**Methods**

*R. centenum* strain SW (ATCC 51521) originated from hot spring mud in Wyoming, United States. A single colony was grown anaerobically and total DNA was isolated using proteinase K treatment followed by phenol extraction. The DNA was fragmented by kinetic shearing, and three shotgun libraries were generated: small and medium insert libraries in the plasmid pOTW13 (using size fractions of 2-3 kb and 6-8 kb, respectively), and a large insert fosmid library in pEpFOS-5 (insert sizes ranging from 28-47 kb), which was used as a scaffold. The relative amount of sequence coverage obtained from the small, medium, and large insert libraries was approximately 8x, 1x, and 1x, respectively. The whole genome sequence was established from 55,014 end sequences (giving 9.7x coverage) derived from these libraries using dye terminator chemistry on ABI 3730xl automated sequencers. The sequence was assembled with the program Arachne [42] and finished as described previously [43]. The complete and annotated genome sequence of *R. centenum* has been deposited at DDBJ/EMBL/GenBank under the accession number CP000613.

Initial automated annotation of the genome was performed with the TIGR/JCVI Annotation Engine [44]. The Pathway-Tools software was employed for predicting and comparing metabolic pathways of *R. centenum* [16,17]. The initial process of metabolic construction for *R. centenum* was automatic and involved building each pathway based on genome annotation results and the presence of each pathway in the MetaCyc database [17]. A further step to validate the accuracy of the constructed metabolic network was carried out based on supporting information from the scientific literature.

**Electron microscopy**

*R. centenum* cultures were harvested and washed three times in phosphate buffer and then pipetted onto CENBA plates in 5 μl aliquots. After 1, 2 and 3 days incubation, the cell spots were harvested, fixed in 5% glutaraldehyde/100 mM HEPES/2 mM MgCl₂ and analyzed by transmission electron microscopy as described previously [18]. Mature *R. centenum* colonies were analyzed by scanning electron microscopy, performed as described previously [49].

**Additional material**

**Additional file 1** Neighbor-joining 16S rDNA phylogeny of the alphaproteobacteria class indicating the distribution of Pk, Pdk, and PEPS

A phylogenetic analysis of alpha-proteobacteria taxa that are annotated further to indicate phototrophism and the presence (or absence) of genes for Rubisco, Pk, Pdk, and PEPS. Characterization of *R. centenum* flagella genes. A table describing the gene name, copy number, similarity, and predicted function of all *R. centenum* flagella-associated genes.

**Authors’ contributions**

REB, CEB, and JWT designed research; MH, JM, SDM, PS and SRC conducted experiments and contributed analytic tools; Y-KL, MH, JM, WDS, JH, TH, SK, AAK, HJM, DR, REB, CEB and JWT analyzed data; and Y-KL, JM, CEB and JWT wrote the paper.

**Acknowledgements**

We would like to dedicate this study to the memory of Jeffrey Favinger who, along with Howard Gest, was the first to isolate *R. centenum*. We’d like to thank The Institute for Genomics Research and the J. C. Venter Institute for providing the Annotation Engine Service that provided us with first-pass automated annotation data and the manual annotation tool Manatee free of charge. Amber L. Conrad, Lisa C. Dejesa, and Heather L. Taylor provided excellent technical assistance with genome sequencing and finishing.
References

1. Favinger J, Stadwald R, Gest H. *Rhodospirillum centenum*, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. *Antonie van Leeuwenhoek* 1989, 55(3):291-296.

2. Stadwald-Demchick R, Turner FR, Gest H. Physiological properties of the thermotolerant photosynthetic bacterium, *Rhodospirillum centenum*. *FEMS Microbiol Lett* 1990, 67:139-144.

3. Nickens D, Fry CJ, Ragatz LR, Bauer CE, Gest H. Biotype of the nonsulfur purple photosynthetic bacterium, *Rhodospirillum centenum*. *Arch Microbiol* 1996, 165:91-96.

4. Stoffels M, Castellanos T, Hartmann A. Design and application of new 16S rRNA-targeted oligonucleotide probes for the *Azosporillum-Skernanera-Rhodosthodos-cluster*. *Syst Appl Microbiol* 2001, 24(1):83-97.

5. Berlemann JE, Bauer CE. Characterization of cyst cell formation in the purple photosynthetic bacterium *Rhodospirillum centenum*. *Trends Microbiol* 2004, 120(3):583-390.

6. Yildiz FH, Gest H, Bauer CE. Attenuated effect of oxygen on photographism in *Rhodospirillum centenum*. *J Bacteriol* 1991, 173(17):5502-5506.

7. Ragatz L, Jiang ZY, Bauer CE, Gest H. Macroscopic phototactic behavior of the purple photosynthetic bacterium *Rhodospirillum centenum*. *Arch Microbiol* 1995, 163(1):1-6.

8. Fani R, Bandi C, Bazicalufo M, Ceccherini MT, Fancelli S, Gallori E, Gerace L, Gripolla A, Micalia N, Damiani G. Phylogeny of the genus *Azospirillum* based on 16S rDNA sequence. *FEMS Microbiol Lett* 1995, 129(2-3):192-200.

9. Sadasivan L, Neys JA. Cyst production and brown pigment formation in aging cultures of *Azospirillum brasilense* ATCC 29145. *J Bacteriol* 1987, 169(4):1670-1677.

10. Stevensen LH, Socolosky MD. Cyst formation and poly-beta-hydroxybutyric acid accumulation in *Azotobacter*. *J Bacteriol* 1966, 91(1):304-310.

11. Anderson AJ, Dawes EA. Occurrence, metabolism, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiol Reviews* 1990, 54(4):450-472.

12. Steenhoudt D, Vanderleyden J. *Azospirillum*, a free-living nitrogen-fixing bacterium closely associated with grasses: genetic, biochemical and ecological aspects. *FEMS Microbiol Rev* 2002, 24(4):487-506.

13. Moens S, Schloter M, Vanderleyden J. Expression of the structural gene, flaF, encoding the flagellin of the lateral flagella in *Azospirillum brasilense* Sp2. *J Bacteriol* 1996, 178(16):5017-5019.

14. Moens S, Vanderleyden J. Functions of bacterial flagella. *Crit Rev Microbiol* 1996, 22(2):67-100.

15. Pereg Gek L, Gilchrist K, Kennedy IR. Mutants with enhanced nitrogenase activity in hydroponic *Azospirillum brasilense*-wheat associations. *Appl Environ Microbiol* 2000, 66(1):2175-2184.

16. Karp PD, Paley S, Romero P. *The Pathway Tools* software. *Bioinformatics* 2002, 18(1):225-232.

17. Caspi R, Foerst H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Sherar AG, Tissier C, Wall TC, Zhang P, Karp PD. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* 2008, D623-631.

18. Favinger J, Stadwald R, Gest H. *Rhodospirillum centenum*, sp. nov., a thermotolerant cyst-forming anoxygenic photosynthetic bacterium. *Antonie van Leeuwenhoek* 1989, 53(3):291-296.

19. Nickens D, Fry CJ, Ragatz LR, Bauer CE, Gest H. Biotype of the nonsulfur purple photosynthetic bacterium, *Rhodospirillum centenum*. *Arch Microbiol* 1996, 165:91-96.

20. Badger MR, Bek EJ. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *FEMS Microbiol Rev* 2008, 32(7):1257-1261.

21. Tabita FR. Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: a different perspective. *Photosynth Res* 1999, 60:21-38.

22. Tang KH, Feng X, Tang YJ. Blankenship RE. Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* OCH114. *PlcO One* 2009, 4(10):1-23.

23. Inoue M, Lumay V, Zahn K, Yamagata H, Yukawa H. Structural and functional analysis of the phoshophoenolpyruvate carboxylase gene from the purple nonsulfur bacterium *Rhodopseudomonas palustris* No. 7. *J Bacteriol* 1997, 179(13):4942-4945.

24. Biebl H, Tindall BJ, Pukall R, Lunsdorf H, Allgaier M, Wagner-Dobler I. *Hoeaella phototrophica* sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll a. *Int J Syst Evol Microbiol* 2006, 56(Pt 4):821-826.

25. Fleischman D, Kramer D. Photosynthetic rhizobia. *Biochim Biophys Acta* 1998, 1364(1):17-36.

26. Tjaden B, Plagens A, Door C, Siebers B, Hensel R. Phosphoenolpyruvate synthetase and pyruvate, phosphate dikinase of Thermoproteus tenax: key pieces in the puzzle of archaean carbohydrate metabolism. *Mol Microbiol* 2006, 60(2):287-298.

27. Grammel H, Gillies ED, Ghosh R. Microaerophilic cooperation of reductive and oxidative pathways allows maximal photosynthetic membrane biosynthesis in *Rhodospirillum rubrum*. *Appl Environ Microbiol* 2003, 69(11):6577-6586.

28. Schneider K, Goitan U, Sehemesier-Voigt S, Plass W, Muller A. Rapid purification of the protein components of a highly active ‘iron only’ nitrogenase. *Naturovissenschaften* 1994, 81(9):405-408.

29. Wang H, Noren A. Metabolic regulation of nitrogen fixation in *Rhodospirillum rubrum*. *Biochem Soc Trans* 2006, 34(1Pt 1):160-161.

30. Silver S, Phung Le T. A bacterial view of the periodic table: genes and proteins for toxic inorganic ions. *J Ind Microbiol Biotechnol* 2005, 32(11-12):587-605.

31. Berlemann JE, Hassellbring BM, Bauer CE. Hypercyct mutants in *Rhodospirillum centenum* identify regulatory loci involved in cyst cell differentiation. *J Bacteriol* 2004, 186(17):5834-5841.

32. Berlemann JE, Bauer CE. A che-like signal transduction cascade involved in controlling flagella biosynthesis in *Rhodospirillum centenum*. *Microb Biotechnol* 2005, 55(5):390-1402.

33. Jiang Z, Swem LR, Rushing BG, Devanathan S, Tollin G, Bauer CE. Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* 1999, 285(5426):405-409.

34. Graud E, Vermeglio A. Bacteriophytochromes in anoxygenic photosynthetic bacteria. *Photosynth Res* 2008, 97(2):141-153.

35. McClain J, Rollo DR, Rushing BG, Bauer CE. *Rhodospirillum centenum* utilizes separate motor and switch components to control lateral and polar flagellum rotation. *J Bacteriol* 2002, 184(9):2429-2438.

36. Liu R, Ochman H. Origins of flagellar gene operons and secondary flagellar systems. *J Bacteriol* 2007, 189(19):7098-7104.

37. Berlemann JE, Bauer CE. Involvement of a Che-like signal transduction cascade in regulating cyst cell development in *Rhodospirillum centenum*. *J Mol Microbiol Biotechnol* 2005, 6(6):1457-1466.

38. Jiang ZY, Bauer CE. Analysis of a chemotaxis operon from *Rhodospirillum centenum*. *J Bacteriol* 1997, 179(16):5712-5719.

39. Jiang ZY, Gest H, Bauer CE. Chemosensory and photosensory perception in purple photosynthetic bacteria utilize common signal transduction components. *J Bacteriol* 1997, 179(16):5720-5727.

40. Jiang ZY, Bauer CE. Component of the *Rhodospirillum centenum* photosensory apparatus with structural and functional similarity to methyl-accepting chemotaxis protein chemoreceptors. *J Bacteriol* 2001, 183(1):171-177.

41. Ton-Hoang B, Salhi M, Schumacher J, De Re S, Kahn D. Promoter-specific involvement of the Fdxr receiver domain in transcriptional activation. *J Mol Biol* 2001, 312(4):583-589.
42. Batzoglou S, Jaffe DB, Stanley K, Butler J, Niever S, Mauceli E, Berger B, Mesirov JP, Lander ES: ARACHNE: a whole-genome shotgun assembler, Genome Res. 2002, 12(1):177-189.

43. Swingley WD, Sadekar S, Mastrian SD, Matthies HJ, Hao J, Ramos H, Acharya CR, Conrad AL, Taylor HL, Dejeza LC, Shah MK, O’Huallachain ME, Lince MT, Blankenship RE, Beatty JT, Touchman JW: The complete genome sequence of Roseobacter denitrificans reveals a mixotrophic rather than photosynthetic metabolism, J Bacteriol 2007, 189(3):683-690.

44. Stothard P, Wishart DS: Circular genome visualization and exploration using CGView, Bioinformatics 2005, 21(4):537-539.

45. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res 2004, 32(5):1792-1797.

46. Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol Biol Evol 2000, 17(4):540-552.

47. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Systematic Biol 2003, 52(5):696-704.

48. Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0, Mol Biol Evol 2007, 24(8):1596-1599.

49. Nickens D, Fry C, Ragatz L, Bauer CE, Gest H: Biotype of the purple nonsulfur photosynthetic bacterium, Rhodospirillum centenum, Arch Microbiol 1996, 165:91-96.