Research
Article

Nature Biotechnology **22**, 55 - 61 (2004)
Published online: 14 December 2003 | doi:10.1038/nbt923

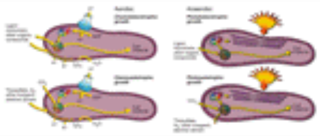
Complete genome sequence of the metabolically versatile photosynthetic bacterium *Rhodopseudomonas palustris*

Frank W Larimer[1,2], Patrick Chain[2,3], Loren Hauser[1,2], Jane Lamerdin[2,3,7], Stephanie Malfatti[2,3], Long Do[2,3,7], Miriam L Land[1,2], Dale A Pelletier[1,2], J Thomas Beatty[4], Andrew S Lang[4], F Robert Tabita[5], Janet L Gibson[5], Thomas E Hanson[5,7], Cedric Bobst[5], Janelle L Torres y Torres[6], Caroline Peres[6,7], Faith H Harrison[6], Jane Gibson[6] & Caroline S Harwood[6]

***Rhodopseudomonas palustris* is among the most metabolically versatile bacteria known. It uses light, inorganic compounds, or organic compounds, for energy. It acquires carbon from many types of green plant–derived compounds or by carbon dioxide fixation, and it fixes nitrogen. Here we describe the genome sequence of *R. palustris*, which consists of a 5,459,213-base-pair (bp) circular chromosome with 4,836 predicted genes and a plasmid of 8,427 bp. The sequence reveals genes that confer a remarkably large number of options within a given type of metabolism, including three nitrogenases, five benzene ring cleavage pathways and four light harvesting 2 systems. *R. palustris* encodes 63 signal transduction histidine kinases and 79 response regulator receiver domains. Almost 15% of the genome is devoted to transport. This genome sequence is a starting point to use *R. palustris* as a model to explore how organisms integrate metabolic modules in response to environmental perturbations.**

R. palustris is a purple photosynthetic bacterium that belongs to the alpha proteobacteria and is widely distributed in nature as indicated by its isolation from sources as diverse as swine waste lagoons, earthworm droppings, marine coastal sediments and pond water. It has extraordinary metabolic versatility and grows by any one of the four modes of metabolism that support life: photoautotrophic or photosynthetic (energy from light and carbon from carbon dioxide), photoheterotrophic (energy from light and carbon from organic compounds), chemoheterotrophic (carbon and energy from organic compounds) and chemoautotrophic (energy from inorganic compounds and carbon from carbon dioxide) (**Fig. 1**). *R. palustris* enjoys exceptional flexibility within each of these modes of metabolism. It grows with or without oxygen and uses many alternative forms of inorganic electron donors, carbon and nitrogen. It degrades plant biomass and chlorinated pollutants and it generates hydrogen as a product of nitrogen fixation[1, 2]. Thus *R. palustris* is a model organism to probe how the web of metabolic reactions that operates within the confines of a single cell adjusts and reweaves itself in response to changes in light, carbon, nitrogen and electron sources that are easily manipulated experimentally. As a critical step in the further development of this model we have sequenced and annotated the *R. palustris* genome. The genome comprises one circular chromosome that is 5.46 Mb in size. The sequenced strain also harbors a 8.4-kilobase (kb) circular plasmid.

Figure 1: Overview of the physiology of *R. palustris*.



Schematic representations of the four types of metabolism that support its growth are shown. The multicolored circle in each cell represents the enzymatic reactions of central metabolism.

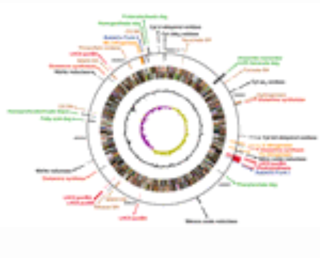
Full size image (75 KB)

Results

Major features of the genome

The *R. palustris* genome has very few repeat nucleotide sequences, insertion sequence elements or transposons. It has just 16 insertion sequence elements including representatives of the 'phage' integrase family, four ISR1-like elements and two *xerD* type elements. No horizontally transferred islands of DNA are apparent based on anomalous G + C content. *R. palustris* has 4,836 predicted protein-encoding genes (**Table 1** and **http://genome.ornl.gov/microbial/rpal/**). These include genes required for the biosynthesis of all its cellular components from carbon dioxide in keeping with its robust growth in media lacking organic carbon sources. *R. palustris* has many genes associated with energy metabolism, reflecting its metabolic versatility (**Fig. 2**). The chromosomal positions and numbered designations of these genes can be found in **Supplementary Table 1** online. There are genes allowing oxidation of hydrogen, thiosulfate and carbon monoxide as energy and reductant sources. Two homologous NADH dehydrogenase complexes that are encoded in the genome likely broker the catabolism of a wide variety of organic compounds, including fatty acids, dicarboxylic acids and lignin monomers. The conditions under which these two seemingly redundant enzyme systems are expressed have not been defined. Terminal oxidase genes should enable *R. palustris* to use nitrite, nitric oxide and nitrous oxide as electron acceptors during anaerobic respiration[3]. There are four sets of genes for terminal oxidases that can function with oxygen: a cytochrome aa_3 oxidase, a cytochrome cbb_3 oxidase, a cytochrome *d* quinol oxidase and a quinol *bd* oxidase. Photosynthesis genes enable the use of light as an energy source by cyclic photophosphorylation under anaerobic conditions.

Figure 2: The chromosome of *R. palustris* strain CGA009.



Major metabolic features and the locations of the genes that encode them are indicated on the outer circle. Progressing inward, the second circle depicts predicted coding regions on the plus strand colored by functional category: white, hypothetical; dark gray, unknown function; red, replication and repair; green, energy metabolism; blue, carbon and carbohydrate metabolism; cyan, lipid metabolism; magenta, transcription; yellow, translation; pale green, structural RNAs; sky blue, cellular processes; orange, amino acid metabolism; brown, general function prediction; pink, metabolism of cofactors and vitamins; light gray, conserved hypothetical; dark green, transport; lavender, signal transduction; light red, purine and pyrimidine metabolism. Third circle, predicted coding regions on minus strand (same color scheme as the second circle). Fourth circle, G + C content (deviation from average); fifth circle, G + C skew in purple and olive. Scale (in bp) is indicated along the outside of the circle.

Full size image (88 KB)

Table 1: General features of the *R. palustris* genome

Full table

Phototrophy

Genes *rpa1505–rpa1554* required for the generation of energy by photophosphorylation reside in a 55-kb region of the *R. palustris* chromosome. These include genes for bacteriochlorophyll and carotenoid biosynthesis as well as genes encoding the L, M and H polypeptides that form the membrane-bound reaction center complex, where light energy is absorbed to initiate electron transfer reactions. The reaction center genes *rpa1527*, *rpa1528* and *rpa1548* are the most highly conserved aspect of this region, sharing from 45 to 60% predicted amino acid identity with the corresponding genes from *Rhodobacter sphaeroides*, a model organism for the study of anoxygenic photosynthesis[4]. However the *R. palustris* reaction center proteins are most similar (on the order of 75% amino acid identity) to homologs in the unusual photosynthetic *Bradyrhizobium sp*. strain ORS278 (ref. 5). This strain forms nitrogen-fixing nodules on the stems of the plant *Aeschynomene sensitiva*, a tropical legume that grows in water logged soils[6]. In addition to a conserved arrangement of photosynthesis genes, the *A. sensitiva* symbiont and *R. palustris* each contain a bacteriophytochrome regulatory gene that is absent in other purple phototrophs. The symbiont's bacteriophytochrome absorbs far-red light and is required for expression of photosynthesis in response to illumination at 740 nm[7]. In our strain the homologous bacteriophytochrome gene *rpa1537* contains a frameshift mutation and is probably inactive. Analysis of rRNA sequences indicates that *R. palustris* is closely related to the *A. sensitiva* symbiont as well as to the soybean symbiont *B. japonicum*[8]. However, *R. palustris* has never been found in symbiotic association with plants, and its genome lacks nodulation genes.

R. palustris, like other purple phototrophic bacteria, responds to lowered light intensity by increasing the amount of light harvesting (LH) complexes. These consist of α and β polypeptides bound to bacteriochlorophyll and a carotenoid, to form a unit that oligomerizes to produce complexes that transfer light energy to the reaction center[9]. The pathway of light energy transfer is LH2 ⋄ LH1 ⋄ reaction center. *R. palustris* differs from other phototrophs in that it has multiple LH2 complexes that differ slightly in the wavelengths of light absorbed. It tunes its complement of LH2 complexes to harvest light of differing qualities and intensities[10]. The genome sequence reveals four complete sets of LH2 genes (*pucBA*) and one incomplete set (**Fig. 2** and **Supplementary Table 1** online). Two of the four complete sets of *pucBA* genes are located near bacteriophytochrome genes *rpa3015*, *rpa3016* and *rpa1490* that may function in the regulation of LH2 complex gene expression.

R. palustris has genes (*rpa0008* and *rpa0009*) that are similar to the circadian clock genes, *kaiB* and *kaiC* previously identified only in oxygenic photosynthetic bacteria[11]. *R. palustris* cells present in anoxic environments generate ample energy by photophosphorylation during daylight hours, but may be energy limited at night. Circadian regulation of energy consuming reactions such as nitrogen fixation would make sense, but has yet to be shown in *R. palustris*.

Carbon dioxide fixation

The *R. palustris* genome encodes two active forms of RubisCO, the key enzyme of the Calvin-Benson-Bassham (CBB) pathway of CO_2 fixation[12]. The form I (*cbbLS, rpa1559* and *rpa1560*) and form II (*cbbM, rpa4641*) RubisCO genes are located on almost opposite sides of the chromosome. The *cbbM* gene is linked to other CBB pathway genes in an arrangement that is similar, but not identical to form II *cbb* operons from other purple phototrophs. The *R. palustris* RubisCO form I gene cluster includes an expected divergently transcribed LysR type regulatory gene *cbbR*, but it differs from form I gene clusters in other species in that it includes three additional regulatory genes situated between *cbbR* and the *cbbLS* structural genes. These encode two predicted response regulators (Rpa1556 and Rpa1557) and a hybrid sensor kinase/response regulator (Rpa1558) that contains two PAS domains.

Inorganic compounds as a source of reducing power

R. palustris oxidizes inorganic compounds such as thiosulfate and hydrogen gas as energy sources for respiratory growth and as sources of reducing power for carbon dioxide and nitrogen fixation. *R. palustris* has a large cluster of genes (*rpa0959*–*rpa0979*) for the synthesis and assembly of a nickel-containing uptake hydrogenase. Its periplasmic thiosulfate:cytochrome *c* oxidoreductase complex is encoded by genes *rpa4459*–*rpa4467* that are very similar to *sox* genes that are found in many other sulfur oxidizing organisms[13]. Its use of reduced sulfur compounds as electron donors sets *R. palustris* apart from closely related phototrophic bacteria[14]. The genome also encodes carbon monoxide dehydrogenases and a formate dehydrogenase (**Fig. 2** and **Supplementary Table 1** online). These can potentially function to supply reductant and substrate for carbon dioxide fixation during anaerobic phototrophic growth or to supply reductant for both energy generation and carbon dioxide fixation under aerobic chemoautotrophic growth conditions.

RubisCO-like proteins

R. palustris is the only organism known to date that encodes two RubisCO-like proteins (RLPs)[12, 15]. RLPs contain varying numbers of substitutions in conserved active site residues. The single RLP from the green sulfur bacterium *Chlorobium tepidum* contains nine active site substitutions and cannot function as a RubisCO[15]. One of the *R. palustris* RLPs (RLP2, Rpa0262) is 66% identical to the *C. tepidum* RLP protein and contains the same pattern of active site substitutions. *R. palustris* RLP1 (Rpa2169) has seven active site substitutions dis-tinct from those in its RLP2. A *C. tepidum rlp* mutant is defective in its ability to oxidize reduced sulfur compounds and from this we infer that the *R. palustris* RLPs are probably involved in sulfur metabolism[16].

Biodegradation

Purple photosynthetic bacteria are a major component of microbial populations found in wastewater treatment facilities exposed to sunlight[17, 18]. *R. palustris* thrives in such environments because it metabolizes structurally diverse compounds found as components of degrading plant and animal wastes. These include lignin monomers, fatty acids and dicarboxylic acids of the types derived from green plants, animal fats and seed oils. *R. palustris* also degrades nitrogen-containing compounds including amino acids and heterocyclic aromatic compounds[2], and it dehalogenates and degrades chlorinated benzoates and chlorinated fatty acids[19, 20], compounds that are sometimes found in industrial wastes.

Although *R. palustris* has been studied for its biodegradation abilities and is a model for molecular studies of aromatic ring degradation in the absence of molecular oxygen[21], its genome has revealed a much larger inventory of degradation genes than expected. It encodes four distinct oxygenase-dependent ring cleavage pathways for the aerobic degradation of the aromatic compounds protocatechuate, homoprotocatechuate, homogentisate and phenylacetate (**Fig. 2** and **Supplementary Table 1** online). *R. palustris* has the potential to combine oxygen-sensitive and oxygen-requiring enzyme reaction sequences to accomplish complete degradation. An example is the anaerobic transformation of phenol to 4-hydroxyphenylacetate, which is then degraded aerobically via either the homogentisate or homoprotocatechuate pathways[22]. These types of transformations would be expected to occur in populations straddling oxic to anoxic transition zones. The genome contains 19 mono- or dioxygenase and four cytochrome P450 genes. Additional genes that may be useful in bioremediation or biocatalysis include nitrile hydratase (*rpa2805* and *rpa2806*) and amidase (*rpa2415*) genes, phosphonate utilization genes (*rpa0687–rpa0700*) and carboxylesterase genes (*rpa1568, rpa2627, rpa3893* and *rpa4646*). The *R. palustris* genome has 16 glutathione *S*–transferase genes, some of which may catalyze the cleavage of β-aryl ether bonds[23].

R. palustris encodes a complete tricarboxylic acid cycle, an Embden-Meyerhof pathway and a pentose phosphate pathway. A predicted glyoxylate shunt permits use of acetate as a sole carbon source, and the genome sequence indicates the synthesis of glycogen and poly β–hydroxyalkanoates as carbon storage polymers. Other genes encode enzymes to mobilize and degrade these polymers during times of carbon starvation. *R. palustris* has a limited ability to grow on sugars and this is reflected by the absence in its genome sequence of glucose or fructose transporters or a hexokinase gene. Genes of the Entner-Doudoroff pathway are absent.

Nitrogen fixation and nitrogen assimilation

We were surprised to find that *R. palustris* has structural genes for three different nitrogenases as well as the related cofactor and assembly genes for these nitrogenases (**Fig. 2** and **Supplementary Table 1** online). Previously, only *Azotobacter* sp., a heterotrophic obligate aerobe, had been found to encode three nitrogenases. *R. palustris* encodes a molybdenum-dependent nitrogenase, found in all nitrogen-fixing bacteria, and also a vanadium-dependent and an alternative iron nitrogenase. *R. palustris* encodes dinitrogenase reductase ADPribosyltransferase (DraT) (Rpa1431 and Rpa2405) and dinitrogenase reductase activating glycohydrolase (DraG) (Rpa2406) enzymes that likely modulate the activity of dinitrogenase reductase by reversible ADP ribosylation. Homologs of NifA (Rpa4632), VnfA (Rpa1374) and AnfA (Rpa1439) regulators are present to potentially activate their cognate clusters of nitrogenase genes in conjunction with the single RNA polymerase sigma factor, RpoN (Rpa0050).

Its genome sequence indicates that *R. palustris* incorporates ammonia exclusively through glutamine synthetase and glutamine:oxoglutarate aminotransferase reactions. It encodes four glutamine synthetases and genes for post-translational control of glutamine synthetase activity by reversible adenylylation are present. *R. palustris* has contiguous duplicated, although not identical, *amtB* genes *rpa0273* and *rpa0275* encoding ammonium transporters. Additional transport and metabolic capacity exists to use cyanate (*rpa2115*), urea (*rpa3658–rpa3664*) and ethanolamine (*rpa3747–rpa3749*) as potential nitrogen sources.

Regulation and signal transduction

Because it is a successful metabolic opportunist, *R. palustris* should be able to sense diverse environmental conditions to appropriately regulate gene expression for survival and growth. It also needs to integrate its metabolism and distribute limited pools of ATP and reductant to competing processes such as nitrogen fixation and carbon dioxide fixation. *R. palustris* has 451 potential regulatory and signaling genes, many of which encode multiple domain motifs (**Table 2**; see **Supplementary Table 2** online for a complete list)[24]. It devotes about the same proportion of its genes (9.3%) to regulation as do the soil bacteria *Pseudomonas putida*, *Streptomyces coelicolor* and *Streptomyces avermitilis* (**http://www.tigr.org/**). Regulatory genes comprise 5–6% of the genomes of most free-living bacteria. The great variety in the domain architecture of *R. palustris*' 63 signal transduction histidine kinases points to their involvement in regulating many different processes. Half of these genes encode from one to ten predicted transmembrane regions, 20 have PAS domains, 9 have GAF domains (which are characteristic of phytochromes) and 2 have very large, novel cytoplasmic domains. The genome has genes for 19 different RNA polymerase sigma factors, 16 of which are classified as extracytoplasmic function (ECF) sigma factors[25]. Two of the ECF sigma factor genes (*rpa0639* and *rpa1635*) are located near flagella biosynthesis genes and another (*rpa0550*) is translationally coupled to a gene resembling the cytochrome c_2 anti-sigma factor gene *chrR*[26], suggesting specific functions.

Table 2: *R. palustris* regulatory and signaling proteins

Full table

R. palustris has an acylhomoserine lactone (HSL) synthase gene (*rpa0320*) that is adjacent to the HSL-responsive regulator gene *rpa0321*. HSLs produced by gram-negative bacteria serve as intercellular signals that allow cells to monitor their population density. Generally, HSLs activate expression of genes that are advantageous to a species when cells of that species are at a population density perceived as a quorum. *R. palustris* genes that might be controlled by quorum sensing include genes *rpa1885–rpa1906* for a phage-like particle called a gene transfer agent[27], polyketide synthase gene *rpa3339*, and genes *rpa3342–rpa3357* for the production and export of exopolysaccharides[28, 29].

R. palustris has genes for three complete chemotaxis signal transduction complexes and it has 30 chemotaxis sensory transducer genes. All but five of the transducers are predicted to be membrane-bound proteins. Four of the transducer genes (*rpa4202, rpa4311, rpa4481* and *rpa4483*) are translationally coupled to or located just a few base pairs away from a sensor gene with a PAS domain. These gene pairs may have originally existed as single genes but have been translationally frameshifted. The existence of the same split genes in *Magnetospirillum magnetotacticum* and *Rhodospirillum rubrum* suggests that this arrangement may have been present in an ancestor common to these three organisms.

Transport

The genome of *R. palustris* encodes about 325 transport systems comprising at least 700 genes, adding up to almost 15% of the genome. Transport genes account for 5–6% of most bacterial genomes[30]. A complete listing, classified using the TC Number system[31] can be found as **Supplementary Table 3** online. There are 102 primary transport systems, defined as systems powered directly by ATP hydrolysis. These include 86 ATP-binding cassette (ABC) systems and 7 P-type ATPases and type II, III and IV secretion systems. The P-type ATPases likely confer resistance to heavy metals[32]. Separate *R. palustris* Type II secretion systems are likely used for the biogenesis of type IV pili and general protein secretion (the Sec system), with a type III secretion system for flagella biosynthesis. *R. palustris* has two sets of type IV secretion genes (*rpa2224–rpa2233* and *rpa4115–rpa4124*) similar to the *Trb* genes from *Agrobacterium tumefaciens* for conjugal transfer of DNA[33].

R. palustris encodes 137 secondary transport systems including 36 major facilitator superfamily (MFS) members, 22 resistance-nodulation-cell division (RND) pumps, 15 divalent metal transport (DMT) members and 8 tripartate ATP-independent periplasmic (TRAP) tranporters[34, 35]. All but two of the RND systems are classified as heavy metal and drug efflux pumps. This is the largest number of RND pumps observed in any bacterium to date and may explain the high intrinsic resistance of *R. palustris* to antibiotics. *R. palustris* has been isolated in high numbers from polluted environments[36]. Heavy metal efflux transporters should allow *R. palustris* to live in a variety of environments and still acquire the necessary nutrients while resisting heavy metal toxicity.

Of the 86 ABC systems, 20 are related to the branched chain amino acid uptake (*ilvFGHKL*) system of *E. coli*. Isoleucine, leucine and valine are hydrophobic amino acids and we speculate that other members of this amplified family are specific for other sorts of hydrophobic compounds such as lignin monomers, fatty acids and dicarboxylic acids derived from oils and fats. One system of this *ilv* ABC family (Rpa0665–Rpa0668) has tentatively been identified as a 4-hydroxybenzoate transport system[21]. Another (*rpa1789* and *rpa1791–1793*) lies adjacent to a feruloyl CoA ligase gene implying that it catalyzes the uptake of the lignin monomer ferulate. A third example is an *ilv* family ABC system (*rpa3719–3725*) that is next to genes for the degradation of the dicarboxylic acid pimelate. An analysis of 73 other microbial genomes shows that 34 of them have no *ilv*-like transport systems. Another 25 microbes have between one and five of these systems and 11 microbes have between six and ten *ilv* family ABC transporters. Only three other species, *Burkholderia fungorum* LB400 and *Ralstonia eutropha*, both β-proteobacteria, and *B. japonicum*, have 19 or more versions of the *ilv*-like ABC transport operon.

Iron acquisition appears to be particularly important for *R. palustris*. It encodes 24 outer membrane ferric iron siderophore receptors, and 7 TonB systems for powering these and other outer membrane receptors (**Supplementary Table 3** online). This implies that *R. palustris* uses a large number of different types of siderophores for iron acquisition. However, genes *rpa2388–rpa2390* to synthesize only one siderophore, rhizobactin[37], were detected suggesting that *R. palustris* may transport iron-loaded siderophores produced by other soil bacteria. As many as seven of the ECF sigma factors encoded by *R. palustris* are either translationally coupled to ferrisiderophore-like receptor genes or are located very close to genes involved in iron acquisition; in one case siderophore biosynthesis genes and in another, a predicted heme uptake system. This suggests a role for multiple alternative sigma factors in activating gene expression in response to iron starvation[38].

Discussion

R. palustris owes much of its metabolic versatility to known genes encoding metabolic modules of carbon dioxide fixation and photophosphorylation that act in concert with dehydrogenases, oxidoreductases and carbon degradation pathways to support its four modes of growth (**Fig. 1**). The number of options that *R. palustris* has within the major metabolic modes to take advantage of fluctuating supplies of carbon, nitrogen, light and oxygen is unusually large. The existence of genes for three nitrogenases, multiple aromatic degradation pathways and multiple oxidoreductases was not known before the genome sequence. Its large inventory of transport and chemotaxis genes implies that *R. palustris* is adept at sensing and acquiring diverse compounds from its environment. The groundwork has now been laid to explore regulatory strategies used by *R. palustris* to appropriately select and integrate its large number of metabolic choices.

R. palustris is ideally suited for use as a biocatalyst because it generates ample supplies of ATP from light thus catalyzing reactions that are thermodynamically unfavorable and beyond the potential of chemotrophic organisms. The metabolic group of purple phototrophic bacteria to which it belongs have been evaluated as sources of single cell protein, for the synthesis of polyhydroxyalkanoate 'bioplastics' and for the production of hydrogen, which they generate as a product of nitrogen fixation[39]. Its genome sequence reveals that *R. palustris* has additional capabilities, not shared by other purple bacteria, that enhance its potential for use in biotechnological applications. These include modulating photosynthesis according to light quality and degrading aromatic compounds that are typically found in agricultural and industrial wastes. That the genome encodes oxygen-requiring, as well as anaerobic reductive pathways, for the degrada-tion of aromatic rings, suggests the possibility of designing hybrid degradation pathways of broader substrate specificity than those that occur naturally. *R. palustris* has physical attributes that are well suited for process development. It undergoes asymmetric cell division and produces a cell surface adhesin at one end of the cell that causes cells to stick to solid substrates. *R. palustris* has especially good potential for use as a biocatalyst for hydrogen production. It is unique among purple phototrophic bacteria in encoding a vanadium-containing nitrogenase that catalyzes the production of approximately three times as much hydrogen as do molybdenum-containing nitrogenases[40]. *R. palustris* derives reductant for hydrogen generation from plant biomass, and energy captured from sunlight drives the process. Manipulating *R. palustris* to produce hydrogen efficiently will require a detailed knowledge of how each of its three nitrogenases is regulated. It will also be important to know in detail how the metabolic modules of photophosphorylation, biodegradation, carbon dioxide fixation and hydrogen uptake are regulated and how their activities are integrated.

Methods

Construction, isolation and sequencing of small-insert and large-insert libraries.

Genomic DNA, isolated from the *R. palustris* CGA009, was sequenced using a conventional whole genome shotgun strategy[41]. Briefly, random 2–3 kb-DNA fragments were isolated after mechanical shearing. These gel-extracted fragments were concentrated, end-repaired and cloned into pUC18. Double-ended plasmid sequencing reactions were carried out using PE BigDye Terminator chemistry (Perkin Elmer) and sequencing ladders were resolved on PE 3700 Automated DNA Sequencers. One round (117,510 reads) of small-insert library sequencing was done, generating roughly 9.6-fold redundancy.

A large insert (~30 kb) fosmid library was also constructed by *Sau*3AI partial digestion of genomic DNA and cloning into the pFos1 cloning vector[42]. End sequencing of ~300 fosmid clones (0.02-fold redundancy) generated roughly 2-fold genome scaffold coverage. The fosmids were fingerprinted with *Eco*RI to aid in assembly verification and determination of gap sizes and provided a minimal scaffold used for order and orientation across assembly gaps. The 8.4-kb plasmid was assembled from a total of 107 reads.

Sequence assembly and gap closure.

Sequence traces were processed with Phred[43; 44] for base calling and assessment of data quality before assembly with Phrap (P. Green, University of Washington, Seattle, Washington, USA) and visualization with Consed[45]. Gaps were closed by primer walking on gap-spanning library clones (identified using linking information from forward and reverse reads). Alternatively, some of the larger gaps, including the larger regions covered only by fosmid clones, were closed by primer walking on PCR products. Remaining physical (uncaptured) gaps were closed by combinatorial (multiplex) PCR. Sequence finishing and polishing added a total of 300 reads and assessment of final assembly quality was done as previously described[46].

Sequence analysis and annotation.

Gene modeling was done using the Critica[47], Glimmer[48] and Generation (**http://compbio.ornl.gov/generation/index.shtml**) modeling packages, the results were combined and a basic local alignment search tool (BLAST) for proteins (P) search of the translations versus GenBank's nonredundant database (NR) was conducted. The alignment of the N terminus of each gene model versus the best NR match was used to pick a preferred gene model. If no BLAST match was returned, the Critica model was retained. Gene models that overlapped by greater than 10% of their length were flagged, giving preference to genes with a BLAST match. The revised gene/protein set was searched against the KEGG GENES, InterPro (incorporating Pfam, TIGRFams, SmartHMM, PROSITE, PRINTS and ProDom) and Clusters of Orthologous Groups of proteins (COGs) databases, in addition to BLASTP versus NR. From these results, categorizations were developed using the KEGG and COGs hierarchies. Initial criteria for automated functional assignment required a minimum 50% residue identity over 80% of the length of the match for BLASTP alignments, plus concurring evidence from pattern or profile methods. Putative assignments were made for identities down to

30%, over 80% of the length. Automated assignments were reviewed and curated manually using a web-based editing environment.

Nucleotide sequence accession number.
The sequence of the complete genome of *R. palustris* CGA009 is available under GenBank/EMBL/DDBJ accession numbers **BX571963** (chromosome) and **BX571964** (plasmid).

Note: ***Supplementary information*** *is available on the Nature Biotechnology website.*

Acknowledgments

The Biological and Environmental Research program of the US Department of Energy's Office of Science funded this research. The Joint Genome Institute managed the overall sequencing effort. The University of California, Lawrence Livermore National Laboratory, carried out genome finishing under the auspices of the US Department of Energy (DOE). Computational annotation was carried out at the Oak Ridge National Laboratory, managed by UT-BATTELLE for the DOE. The DOE provided additional support to J.T.B., F.R.T. and C.S.H. The US Army Research Office provided support to C.S.H.

Competing interests statement

The authors declare no competing financial interests.

Received 24 September 2003; Accepted 3 November 2003; Published online 14 December 2003.

References

1. Barbosa, M.J., Rocha, J.M., Tramper, J. & Wijffels, R.H. Acetate as a carbon source for hydrogen production by photosynthetic bacteria. *J. Biotechnol.* **85**, 25–33 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |
2. Sasikala, C. & Ramana, C.V. Biodegradation and metabolism of unusual carbon compounds by anoxygenic phototrophic bacteria. *Adv. Microb. Physiol.* **39**, 339–377 (1998). | **PubMed** | **ISI** | **ChemPort** |
3. Philippot, L. Denitrifying genes in bacterial and archaeal genomes. *Biochim. Biophys. Acta.* **1577**, 355–376 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |
4. Hu, X., Ritz, T., Damjanovic, A., Autenrieth, F. & Schulten, K. Photosynthetic apparatus of purple bacteria. *Q. Rev. Biophys.* **35**, 1–62 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |
5. Giraud, E., Hannibal, L., Fardoux, J., Vermeglio, A. & Dreyfus, B. Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene sensitiva*. *Proc. Natl. Acad. Sci. USA* **97**, 14795–14800 (2000). | **Article** | **PubMed** | **ChemPort** |

6. Boivin, C. *et al*. Stem nodulation in legumes: diversity, mechanisms, and unusual characteristics. *Crit. Rev. Plant Sci.* **16**, 1–30 (1997). | **ISI** | **ChemPort** |
7. Giraud, E. *et al*. Bacteriophytochrome controls photosystem synthesis in anoxygenic bacteria. *Nature* **417**, 202–205 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |
8. van Berkum, P. *et al*. Discordant pylogenies within the *rrn* loci of Rhizobia. *J. Bacteriol.* **185**, 2988–2998 (2003). | **Article** | **PubMed** | **ISI** | **ChemPort** |
9. Cogdell, R.J. *et al*. How photosynthetic bacteria harvest solar energy. *J. Bacteriol.* **181**, 3869–3879 (1999). | **PubMed** | **ISI** | **ChemPort** |
10. Gall, A. & Robert, B. Characterization of the different peripheral light-harvesting complexes from high- and low-light grown cells from *Rhodopseudomonas palustris*. *Biochemistry* **38**, 5185–5190 (1999). | **Article** | **PubMed** | **ISI** | **ChemPort** |
11. Johnson, C.H. & Golden, S.S. Circadian programs in cyanobacteria: adaptiveness and mechanism. *Annu. Rev. Microbiol.* **53**, 389–409 (1999). | **Article** | **PubMed** | **ISI** | **ChemPort** |
12. Tabita, F.R. Microbial ribulose–1,5–bisphosphate carboxylase/oxygenase: a different perspective. *Photosynthesis Res.* **60**, 1–28 (1999). | **Article** | **ISI** | **ChemPort** |
13. Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A. & Fischer, J. Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Appl. Environ. Microbiol.* **67**, 2873–2882 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |
14. Rolls, J.P. & Lindstrom, E.S. Effect of thiosulfate on the photosynthetic growth of *Rhodopseudomonas palustris*. *J. Bacteriol.* **94**, 860–869 (1967). | **PubMed** | **ISI** | **ChemPort** |
15. Hanson, T.E. & Tabita, F.R. A ribulose–1,5–bisphosphate carboxylase/oxygenase (RubisCO)–like protein from *Chlorobium tepidum* that is involved with sulfur metabolism and the response to oxidative stress. *Proc. Natl. Acad. Sci. USA* **98**, 4397–4402 (2001). | **Article** | **PubMed** | **ChemPort** |
16. Hanson, T.E. & Tabita, F.R. Insights into the stress response and sulfur metabolism revealed by proteome analysis of a *Chlorobium tepidum* mutant lacking the RubisCO-like protein. *Photosynthesis Res.* **78**, 231–248 (2003). | **Article** | **ISI** | **ChemPort** |
17. Do, Y.S. *et al*. Role of *Rhodobacter* sp. strain PS9, a purple non-sulfur photosynthetic bacterium isolated from an anaerobic swine waste lagoon, in odor remediation. *Appl. Environ. Microbiol.* **69**, 1710–1720 (2003). | **Article** | **PubMed** | **ISI** | **ChemPort** |
18. Kobayashi, M. & Kobayashi, M. in *Anoxygenic Photosynthetic Bacteria* (eds. Blankenship, R.E., Madigan, M.T. & Bauer, C.E.) 1269–1282 (Kluwer Academic Publishers, Dordrecht, The Netherlands, 1995).
19. McGrath, J.E. & Harfoot, C.G. Reductive dehalogenation of halocarboxylic acids by the phototrophic genera *Rhodospirillum* and *Rhodopseudomonas*. *Appl. Environ. Microbiol.* **63**, 3333–3335 (1997). | **PubMed** | **ISI** | **ChemPort** |

20. Egland, P.G., Gibson, J. & Harwood, C.S. Reductive, coenzyme A–mediated pathway for 3–chlorobenzoate degradation in the phototrophic bacterium *Rhodopseudomonas palustris*. *Appl. Environ. Microbiol.* **67**, 1396–1399 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |

21. Egland, P.G., Pelletier, D.A., Dispensa, M., Gibson J. & Harwood, C.S. A cluster of bacterial genes for anaerobic benzene ring biodegradation. *Proc. Natl. Acad. Sci. USA* **94**, 6484–6489 (1997). | **Article** | **PubMed** | **ChemPort** |

22. Noh, U., Heck, S., Giffhorn, F. & Kohring, G.W. Phototrophic transformation of phenol to 4-hydroxyphenylacetate by *Rhodopseudomonas palustris*. *Appl. Microbiol. Biotechnol.* **58**, 830–835 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |

23. Masai, E. *et al.* Roles of the enantioselective glutathione S-transferases in cleavage of beta-aryl ether. *J. Bacteriol.* **185**, 1768–1775 (2003). | **Article** | **PubMed** | **ISI** | **ChemPort** |

24. Galperin, M.Y., Nikolskaya, A.N. & Koonin, E.V. Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiol. Lett.* **203**, 11–21 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |

25. Helmann, J.D. The extracytoplasmic function (ECF) sigma factors. *Adv. Microb. Physiol.* **46**, 47–110 (2002). | **PubMed** | **ISI** | **ChemPort** |

26. Newman, J.D., Anthony, J.R. & Donohue, T.J. The importance of zinc-binding to the function of *Rhodobacter sphaeroides* ChrR as an anti-sigma factor. *J. Mol. Biol.* **313**, 485–499 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |

27. Lang, A.S. & Beatty, J.T. The gene transfer agent of *Rhodobacter capsulatus* and "constitutive transduction" in prokaryotes. *Arch. Microbiol.* **175**, 241–249 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |

28. Marketon, M.M., Glenn, S.A., Eberhard, A. & Gonzalez, J.E. Quorum sensing controls exopolysaccharide production in *Sinorhizobium meliloti*. *J. Bacteriol.* **185**, 325–331 (2003). | **Article** | **PubMed** | **ISI** | **ChemPort** |

29. Schaefer, A.L., Taylor, T.A., Beatty, J.T. & Greenberg, E.P. Long-chain acyl-homoserine lactone quorum-sensing regulation of *Rhodobacter capsulatus* gene transfer agent production. *J. Bacteriol.* **184**, 6515–6521 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |

30. Paulson, I.T., Nguyen, L., Sliwinski, M.K., Rabus, R. & Saier, M.H. Jr. Microbial genome analysis: comparative transport capabilities in eighteen prokaryotes. *J. Mol. Biol.* **301**, 75–100 (2000). | **Article** | **PubMed** | **ISI** | **ChemPort** |

31. Saier, M.H. Jr. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* **64**, 354–411 (2000). | **Article** | **PubMed** | **ISI** | **ChemPort** |

32. Rosen, B.P. Transport and detoxification systems for transition metals, heavy metals and metalloids in eukaryotic and prokaryotic microbes. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **133**, 689–693 (2002). | **Article** | **PubMed** | **ISI** |

33. Cao, T.B. & Saier, M.H. Jr. Conjugal type IV macromolecular transfer systems of Gram–negative bacteria: organismal distribution, structural constraints and evolutionary conclusions. *Microbiology* **147**, 3201–3214 (2001). | **PubMed** | **ISI** | **ChemPort** |
34. Saier, M.H. Jr. & Paulsen, I.T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12**, 205–213 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |
35. Kelly, D.J. & Thomas, G.H. The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol. Rev.* **25**, 405–424 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |
36. Oda, Y. *et al.* Genotypic and phenotypic diversity within species of purple nonsulfur bacteria isolated from aquatic sediments. *Appl. Environ. Microbiol.* **68**, 3467–3477 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |
37. Lynch, D. *et al.* Genetic organization of the region encoding regulation, biosynthesis, and transport of rhizobactin 1021, a siderophore produced by *Sinorhizobium meliloti*. *J. Bacteriol.* **183**, 2576–2585 (2001). | **Article** | **PubMed** | **ISI** | **ChemPort** |
38. Visca, P., Leoni, L., Wilson, M.J. & Lamont, I.L. Iron transport and regulation, cell signalling and genomics: lessons from *Escherichia coli* and *Pseudomonas*. *Mol. Microbiol.* **45**, 1177–1190 (2002). | **Article** | **PubMed** | **ISI** | **ChemPort** |
39. Sasikala, C. & Ramana, C.V. Biotechnological potentials of anoxygenic phototrophic bacteria. II. Biopolyesters, biopesticide, biofuel, and biofertilizer. *Adv. Appl. Microbiol.* **41**, 227–278 (1995). | **PubMed** | **ISI** | **ChemPort** |
40. Eady, R.R. Structure–function relationships of alternative nitrogenases. *Chem. Rev.* **96**, 3013–3030 (1996). | **Article** | **PubMed** | **ISI** | **ChemPort** |
41. Fleischmann, R.D. *et al.* Whole genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496–512 (1995). | **Article** | **PubMed** | **ISI** | **ChemPort** |
42. Kim, U.J., Shizuya, H., deJong, P.J., Birren, B. & Simon, M.I. Stable propagation of cosmid sized human DNA inserts in an F factor based vector. *Nucleic Acids Res.* **20**, 1083–1085 (1992). | **Article** | **PubMed** | **ISI** | **ChemPort** |
43. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998). | **PubMed** | **ISI** | **ChemPort** |
44. Ewing, B., Hillier, L., Wendl, M.C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998). | **PubMed** | **ISI** | **ChemPort** |
45. Gordon, D., Abajian, C. & Green, P. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**, 195–202 (1998). | **PubMed** | **ISI** | **ChemPort** |
46. Chain, P. *et al.* Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J. Bacteriol.* **185**, 2759–2773 (2003). | **Article** | **PubMed** | **ISI** | **ChemPort** |

47. Badger, J.H. & Olsen, G.J. CRITICA: coding region identification tool invoking comparative analysis. *Mol. Biol. Evol.* **16**, 512–524 (1999). | PubMed | ISI | ChemPort |

48. Delcher, A.L., Harmon, D., Kasif, S., White, O. & Salzberg, S.L. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**, 4636–4641 (1999). | Article | PubMed | ISI | ChemPort |

1. Genome Analysis and Systems Modeling, Oak Ridge National Laboratory, One Bethel Valley Rd., Oak Ridge, Tennessee 37831, USA.
2. Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, California 94598, USA.
3. Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, California 94550, USA.
4. Department of Microbiology and Immunology, The University of British Columbia, 6174 University Blvd., Vancouver, British Columbia, Canada V6T 1Z3.
5. Department of Microbiology, The Ohio State University, 484 West 12th Ave., Columbus, Ohio 43210, USA.
6. Department of Microbiology, 3-432 Bowen Science Bldg., The University of Iowa, Iowa City, Iowa 52242, USA.
7. Present addresses: Odyssey Thera, 4550 Norris Canyon Rd., San Ramon, California 94583, USA (J.L.), Department of Biology, University of California San Diego, 9500 Gilman Dr., La Jolla, California 92093, USA (L.D.), Delaware Biotechnology Institute, The University of Delaware, 15 Innovation Way, Newark, Delaware 19711, USA (T.E.H.), Genencor International, 925 Page Mill Rd., Palo Alto, California 94304, USA (C.P.).

Correspondence to: Caroline S Harwood[6] e-mail: caroline-harwood@uiowa.edu

Nature Biotechnology ISSN 1087-0156 EISSN 1546-1696

Search: go