The small genome of an abundant coastal ocean methylotroph

Stephen J. Giovannoni,1,† Darin H. Hayakawa,2† H. James Tripp,1† Ulrich Stingl,1† Scott A. Givan,3 Jang-Cheon Cho,4 Hyun-Myung Oh,1 Joshua B. Kitner,1 Kevin L. Vergin1 and Michael S. Rappé2
1Department of Microbiology and 3Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA.
2Hawaii Institute of Marine Biology, SOEST, University of Hawaii at Manoa, PO Box 1346, Kaneohe, HI 96744, USA.
4Division of Biology and Ocean Sciences, Inha University, Incheon, Korea.

Summary
OM43 is a clade of uncultured β-proteobacteria that is commonly found in environmental nucleic acid sequences from productive coastal ocean ecosystems, and some freshwater environments, but is rarely detected in ocean gyres. Ecological studies associate OM43 with phytoplankton blooms, and evolutionary relationships indicate that they might be methylotrophs. Here we report on the genome sequence and metabolic properties of the first axenic isolate of the OM43 clade, strain HTCC2181, which was obtained using new procedures for culturing cells in natural seawater. We found that this strain is an obligate methylotroph that cannot oxidize methane but can use the oxidized C1 compounds methanol and formaldehyde as sources of carbon and energy. Its complete genome is 1304 428 bp in length, the smallest yet reported for a free-living cell. The HTCC2181 genome includes genes for xanthorhodopsin and retinal biosynthesis, an auxiliary system for producing transmembrane electrochemical potentials from light. The discovery that HTCC2181 is an extremely simple specialist in C1 metabolism suggests an unanticipated, important role for oxidized C1 compounds as substrates for bacterioplankton productivity in coastal ecosystems.

Introduction
The OM43 clade was first discovered by sequencing ribosomal RNA genes from coastal waters of the Western Atlantic (Rappé et al., 1997), and subsequently shown to be common in coastal ecosystems (Rappé et al., 2000). An ecological study by Morris and colleagues (2006) indicated that the abundance of OM43 might be linked to phytoplankton populations and primary productivity, which are often much higher over the continental shelves. The first appearance of the OM43 clade in culture was described by Connon, who reported isolates from the Oregon coast obtained by dilution-to-extinction culturing in media made from natural seawater (Connon and Giovannoni, 2002). Follow-up studies resulted in the propagation of strain HTCC2181 from the Oregon coast, and a related strain, HIMB624, obtained from a much warmer coastal environment in Hawaii. In our laboratory, which routinely propagates many strains of oligotrophic bacterioplankton, OM43 strains are regarded as particularly sensitive to the variation in natural seawater media and are difficult to maintain. So far, they have not been propagated successfully on agar plates or artificial media, and can only be grown on natural seawater.

The OM43 clade is related to Type I methylotrophs of the family Methylophilaceae (Lidstrom, 2001). Methylophilaceae are aerobic, obligate methylotrophs that cannot oxidize methane, but can use methylated compounds, such as methanol, methylamine or formate as their sole source of carbon and energy (Anthony, 1982). The strain most closely related to the OM43 clade that has been described is Methylotenera mobilis (Kalyuzhnaya et al., 2006). Methylotenera mobilis is the only member of the Methylophilaceae described so far that cannot oxidize methanol. It is restricted to growth on a single C1 compound – methylamine.

A number of marine methylotrophs have been isolated using enrichment culturing methods. The Type I methanotroph Methylomonas pelagica was isolated from Sargasso Sea samples by enrichment with CH4 (Sieburth et al., 1987). Several studies reported the isolation of Type I methanotrophs from sewage outfalls (Lidstrom, 1988) or shoreline samples (Holmes et al., 1996), but the prevalence of these organisms in marine systems has not been measured, other than the assurance, which enrichments provide, that they are present. Environmental data
suggest that Methylophaga spp., which are unrelated to the Type I methanotrophs and cannot oxidize methane, are widespread in marine sediments, but numerically rare (Janvier et al., 2003). Stable isotope probing with $^{13}$C-labelled compounds implicated Methylophaga spp. and novel $\gamma$-proteobacteria in the metabolism of methanol and methylamine, but the long incubation times (4 days) left uncertainty about the prevalence of the strains in the native microbial community (Neufeld et al., 2007).

Methanol is a major component of oxygenated volatile organic chemicals (OVOC) in the oceans and atmosphere (Heikes et al., 2002; Singh et al., 2003). Air measurements over the Pacific indicate that ocean surface methanol concentrations are about 100 nM and that central ocean regions are net sinks for methanol deposited from the atmosphere (Heikes et al., 2002). The primary source of atmospheric methanol is terrestrial plants, but anthropogenic sources, methane oxidation, plant decay and fires are significant. However, sources and sinks of methanol are not well defined for ocean surface waters (Heikes et al., 2002). Phytolankton is postulated to be a source of methanol (Sinha et al., 2007), but measurements of productive phytoplankton communities in mesocosms suggest that even under these conditions seawater is a sink for atmospheric methanol (Sinha et al., 2007).

Here we give the first account of the propagation in culture and genome sequencing of a member of the OM43 clade. Physiological studies and genome annotation of strain HTCC2181 revealed it to be an obligate methylotroph with a narrow substrate range and a very small genome. Indeed, the genome is the smallest yet reported for a free-living cell, and only 44% of the size of the Pelagibacter genome, the next smallest genome reported for a free-living heterotrophic cell; that is, a cell that in nature is not found in direct association with other cells. The HTCC2181 genome now stands as the smallest genome reported for a free-living heterotrophic cell, and only 44% of the size of the $\beta$-Proteobacteria – Burkholderiales.

**Growth on C1 compounds**

Growth of HTCC2181 on C1 compounds was studied in seawater media amended with 0–30 $\mu$M and 0–3 mm methanol, formaldehyde, formate and methylamine (Fig. 2). The addition of methanol or formaldehyde to sterilized seawater media amended with 10 $\mu$M NH$_4$Cl and 1 $\mu$M K$_2$HPO$_4$ resulted in a stoichiometric increase in cell yield (Fig. 2), while the addition of methylamine, choline, dimethylsulfonyl propionate, dimethyl sulfide or glycine betaine did not enhance growth. Control experiments without added carbon and energy sources incubated in the dark or in the light did not show a difference in growth rates or yields. The doubling time of HTCC2181 in exponential growth phase was in the range of 14–18 h, independent of light or addition of substrates.

**Comparative genomics**

The HTCC2181 genome now stands as the smallest genome reported for a free-living heterotrophic cell; that is, a cell that in nature is not found in direct association with other cells. The genome size of HTCC2181 is 1304 428 bp, about 4 kb less than the Pelagibacter genome, the next smallest genome reported for a free-living cell, and only 44% of the size of the Methylobacillus flagellatus genome, the next largest genome from a methylotroph (Fig. 3D). Despite a very small genome size, HTCC2181 appears to have the pathways needed for a

### Table 1. 16S rRNA gene similarities between HTCC2181 and other methylotrophs.

<table>
<thead>
<tr>
<th>Strain</th>
<th>% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methylotenera mobilis (DQ287786)</td>
<td>94.4</td>
</tr>
<tr>
<td>Methylobacillus methylotrophus (AB193724)</td>
<td>94.3</td>
</tr>
<tr>
<td>Methylophilus leisingeri (AB193725)</td>
<td>94.1</td>
</tr>
<tr>
<td>Methylophilus quaylei (AY772089)</td>
<td>92.3*</td>
</tr>
<tr>
<td>Methylobacillus flagellatus (DQ287787)</td>
<td>92.9</td>
</tr>
<tr>
<td>Methylovorus mays (AY466132)</td>
<td>92.1*</td>
</tr>
<tr>
<td>Methylibium petroleiphilum (AF176594)</td>
<td>86.3</td>
</tr>
</tbody>
</table>

a. These sequences may contain errors.

**Methylophilaceae**, which are all obligate methylotrophs, has long prompted informal speculation that members of the OM43 clade might also be Type I methylotrophs. Table 1 shows 16S rRNA gene similarities between HTCC2181 and related methylotrophs. Similarities to *M. mobilis, Methylophilus methylotrophus* and *Methylophilus leisingeri* range from 94.1 to 94.4, suggesting that HTCC2181 represents a related genus of methylotrophs. *Methylibium petroleiphilum* (AF176594) (Kane et al., 2007), which has a much lower sequence similarity (86.3%), is in a different order of the $\beta$-Proteobacteria – Burkholderiales.

**Results and discussion**

**Phylogeny**

The phylogenetic relationships of OM43 clade strains HTCC2181 and HIMB624 to other members of the family *Methylophilaceae* are depicted in the 16S rRNA tree shown in Fig. 1. The OM43 clade is diverse, and includes both freshwater and marine representatives. The phylogenetic position of the OM43 clade among the family

© 2008 The Authors
Fig. 1. Phylogenetic relationships between strain HTCC2181 and representatives of the β-proteobacteria Type I methylotroph clade (Methylophilaceae) inferred from 16S rRNA gene sequence comparisons. All known members of this clade are obligate methylotrophs. Branch points supported by bootstrap values for all inference methods are indicated by solid circles (>90%) or open circles (>70%), while those supported by some inference methods are indicated by grey circles (>70%). The scale bar corresponds to 0.02 substitutions per nucleotide position. A variety of α-proteobacteria were used as outgroups. 16S rRNA gene sequence similarities ranged from 92.9% to 94.4% within the clade, but dropped to 86.3% for Methylibium petroleiphilum (AF176594), which is in the order Burkholderiales of the β-Proteobacteria.

Fig. 2. A. Stoichiometric increase in cell yield of strain HTCC2181 after addition of different concentrations of formaldehyde, methanol and formate compared with controls without amendments. Addition of methylamine did not result in an increase in cell yield compared with the controls. The increase in cell yield per μM of added methanol and formaldehyde was $1.24 \times 10^6$ and $1.29 \times 10^6$ cells per ml respectively. Increase in cell yield per μM formate was $2.70 \times 10^4$ cells per ml. Data points are mean values of duplicate flasks. Open circles, formaldehyde; closed circles, methanol; open squares, methylamine; closed squares, formate.

B. Growth curves for addition of different concentrations of methanol. Data points are mean values of duplicate flasks. Treatments and controls all received excess N, P, Fe and vitamins. Treatments varied in amount of added methanol only. Squares, control; circles, 100 μM; triangles, 300 μM; diamonds, 1 mM; stars, 3 mM.

© 2008 The Authors
relatively autonomous existence. Complete pathways were annotated for oxidative phosphorylation and biosynthesis of all 20 amino acids. The only vitamin biosynthetic pathway not encoded by the genome was that for B12. In contrast, Pelagibacter requires five vitamins (Giovannoni et al., 2005). No genes associated with motility were found. The G+C content of HTCC2181 is low (37.95%), though somewhat higher than that of Pelagibacter (29.7%) and a high-light adapted Prochlorococcus (30.8%). Published genomes from other methylotrophs have much higher G+C contents (M. flagellatus, 55.71%; Methylococcus capsulatus, 63.58%; M. petroleiphilum, 69.2%).

As illustrated by the MUMmer plots shown in Fig. 3A–C, the HTCC2181 genome is syntenic with the M. flagellatus genome over significant expanses of the genome, but shows little synteny with M. petroleiphilum (Kane et al., 2007), a distantly related methylotroph belonging to the order Burkholderiales of the β-proteobacteria (see legend to Fig. 1). The pattern shown in Fig. 3C suggests the genome of HTCC2181 was derived from genomes of methylotrophic relatives, like M. flagellatus, by genome reduction. The lost coding capacity is reflected in the reduced metabolic versatility of HTCC2181, which is evident in Table 2. For example, HTCC2181 lacks the mau gene cluster for methylamine utilization, genes for H4MPT biosynthesis and the H4MPT-linked formaldehyde oxidation pathway.

A plot of genome size versus the amount of non-coding DNA in all available sequenced microbial genomes reveals that the genome of HTCC2181 is very small and the proportion of its genome dedicated to non-coding DNA falls within the lowest 2.5% (Fig. S1) of all currently sequenced microbial genomes (Fig. 4). We interpret this as evidence of streamlining selection. The first reports of genome sequences from the cyanobacterium Prochlorococcus and the α-proteobacterium SAR11 (Pelagibacter) established that these very abundant marine bacterioplankton clades have unusually small genomes (Rocap et al., 2003; Giovannoni et al., 2005). The genome of Pelagibacter (1.31 Mbp) was the smallest reported genome for a free-living heterotrophic cell before this report (Giovannoni et al., 2005). Prochlorococcus genomes, which range in size from 1.66 to 2.41 Mbp, are the smallest cyanobacterial genomes reported (Rocap et al., 2003). A theory now referred to as genome

Fig. 3. Whole genome comparisons for three selected methylotrophs. A–C. MUMmer dot plots (Delcher et al., 1999b) showing gene order and content comparisons between HTCC2181, M. flagellatus (Chistoserdova et al., 2007) and M. petroleiphilum (Kane et al., 2007). Note that gene order is most conserved between HTCC2181 and M. flagellatus. The horizontal white space across the top of (B) reflects the presence of a plasmid for aromatic hydrocarbon utilization found only in M. petroleiphilum. D. Venn diagram showing the relative sizes of the three genomes being compared and the number of genes shared or unique.
streamlining has been invoked to explain the small genomes of some marine bacterioplankton (Rotthauwe et al., 1997; Dufresne et al., 2005; Giovannoni et al., 2005). The essence of this theory is that selection is most efficient in microbial populations with large effective population sizes, and therefore the elimination of unnecessary DNA from genomes will be most pronounced in organisms, like bacterioplankton, that meet this criterion. Moreover, there is the added factor that bacterioplankton live in a habitat that is frequently limited in the macronutrients N and P, which are stochiometrically high in nucleic acids. Thus, while it can be argued that all bacteria in some sense have ‘streamlined’ genomes (Lynch, 2006), basic theoretical considerations predict that marine bacterioplankton might exhibit streamlining to a greater extent than other bacterial populations. Alternatively, it could be argued that the marine habitat is simply more uniform than other habitats, requiring less versatility and allowing organisms to dispense with much of their genomes. This concept is belied by the observation that many marine bacterioplankton, albeit those with smaller population sizes, have relatively average genome sizes (Moran et al., 2004; Fuchs et al., 2007). Thus, ‘Candidatus Pelagibacter ubique’, and HTCC2181 are anomalous, and may represent an unusual evolutionary strategy of abandoning metabolic versatility in favour of specialization on carbon compounds that are present at low concentrations in the ambient nutrient field.

Table 2. Comparison of selected genome features of three methylotrophs.

<table>
<thead>
<tr>
<th>Gene cluster/function</th>
<th>HTCC2181</th>
<th>M. flagellatus</th>
<th>M. petroleiphilum</th>
</tr>
</thead>
<tbody>
<tr>
<td>mxl/mox methanol dehydrogenase cluster</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>xox cluster of methanol dehydrogenase homologues</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>PQQ synthesis cluster</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>RuMP cycle for carbon fixation</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>H,F biosynthesis and C1 transfer</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Multiple TonB-dependent receptors</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Methylamine utilization (mau cluster)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>H,MPT biosynthesis and C1 transfer</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Complete TCA cycle</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Gluconeogenesis</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Transposases</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Flagellum functions</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Serine cycle for carbon fixation</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Methyl tert-butyl ether (MTBE) utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Benzene utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Toluene utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Xylene utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Phenol utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Catechol utilization</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

Fig. 4. The total length of non-coding (spacer) DNA versus genome length for all published bacterial and archaeal genomes. Data points for methylotrophs are circled. The small genomes of Candidatus Pelagibacter ubique, and HTCC2181 are anomalous. Regression lines for obligate symbionts and other bacteria differ significantly, with symbionts tending to have small genomes but proportionately more non-coding DNA. Some marine bacteria, like Candidatus Pelagibacter ubique, and HTCC2181, have small genomes and fall below the regression lines, but other marine bacteria, including the abundant marine bacterium Prochlorococcus, fall close to the regression line. Mycoplasmas, which are heterotrophs with much smaller genomes than HTCC2181, sometimes can be grown axenically in media containing extracts of eukaryotic cells, but are always found in nature in association with multicellular eukaryotic organisms. Therefore, they are included among the obligate symbionts.
Metabolic reconstruction of pathways of C1 metabolism

HTCC2181 provides the second example of an organism with an ability to grow on methanol, but which lacks two genes, \textit{mxaF} and \textit{mA}, that code for large and small subunits of a confirmed methanol dehydrogenase (MDH). \textit{Methylibium petroleiphilum} PM1 also oxidizes methanol but lacks close homologues of these genes. While both HTCC2181 and \textit{M. petroleiphilum} are altogether missing the small subunit, which has no catalytic activity and whose function is unknown (Anthony and Williams, 2003), both have genes related to a paralogue of \textit{mxaF}, called \textit{xoxF}, that is found in \textit{Methylobacterium extorquens}, \textit{Paracoccus denitrificans} and \textit{M. flagellatus}. It has long been doubted that \textit{xoxF} genes have methanol dehydrogenase activity (Ras et al., 1991; Van Spanning et al., 1991; Harms et al., 1996; Chistoserdova and Lidstrom, 1997), primarily based on the fact that inactivation of the known methanol dehydrogenase, \textit{mxaF}, by gene disruption, renders \textit{P. denitrificans} unable to grow on methanol (Van Spanning et al., 1991). Presumably, if the \textit{xoxF} gene product in \textit{P. denitrificans} had methanol dehydrogenase activity, the ability to grow on methanol would be retained when the paralogous gene, \textit{mxaF}, was disrupted. However, very recently an \textit{xoxF} gene in \textit{Rhodobacter sphaeroides} with high similarity to the \textit{xoxF} in \textit{P. denitrificans} (80% amino acid identity) has been shown to be required for the use of methanol as a sole photosynthetic carbon source, thus implicating it in the oxidation of methanol to formaldehyde (Wilson et al., 2008).

![Phylogenetic tree of the large subunit of methanol dehydrogenase.](image)

**Fig. 5.** Phylogenetic tree of the large subunit of methanol dehydrogenase. The HTCC2181 gene is related to a cluster of \textit{M. flagellatus} paralogues that are thought to be unessential for methanol oxidation in \textit{M. flagellatus}. Direct experimental evidence has shown that the genes shown in green have methanol dehydrogenase activity and the genes shown in red do not. The tree was derived using the neighbour joining method. Bootstrap values were higher than 90% for both the parsimony and neighbour joining methods unless otherwise indicated. Open circles indicate bootstrap values > 70% for both inference methods, while those supported by one method (> 70%) are indicated by grey circles. Asterisks indicate genes from non-methylotrophs.
able to find a mxaED gene cluster in HTCC2181, which contains two genes also found in the long operon from *M. extorquens*. The simplest explanation for all of these facts is that the MDH and PQQ homologues in HTCC2181 enable this organism to oxidize methanol. This seems more likely than the alternative hypothesis that these genes encode proteins that oxidize an unidentified substrate and that HTCC2181 independently evolved a novel analogue of methanol dehydrogenase.

Genome annotation indicates that HTCC2181 is probably an obligate methylotroph. HTCC2181 is missing the E1 (*sucA*, EC 1.2.4.2) and E2 (*sucB*, EC 2.3.1.61) subunits of the α-ketoglutarate dehydrogenase complex. It has been shown that the absence of these enzymes results in an incomplete tricarboxylic acid cycle, which may be the biochemical basis for obligate methylotrophy, although genomic evidence confirming the results of the negative tests for enzyme activity remains incomplete (Taylor and Anthony, 1976; Anthony, 1982; Wood et al., 2004). Subsequently, it was observed that this enzyme complex is repressed in facultative methylotrophs during growth on C1 compounds, resulting in an incomplete TCA cycle that plays a strictly assimilatory role (Chistoserdova et al., 2003). Indeed, an incomplete TCA cycle lacking α-ketoglutarate dehydrogenase activity has been found in nearly all Type I methylotrophs (Hanson and Hanson, 1996), although homologues of α-ketoglutarate dehydrogenase were found in *M. capsulatus* (Ward et al., 2004).

Evidence that HTCC2181 is highly dependent on C1 compounds also came from an analysis of genes encoding transport functions, which revealed relatively few genes that could be convincingly assigned to protein families associated with organic molecule transport (Table S1). An interesting exception was an MFS family gene that apparently encodes a glycine betaine/proton symporter; however, in growth experiments glycine betaine failed to support growth of HTCC2181. Although it is likely that HTCC2181 can assimilate some organic molecules in addition to C1 compounds, the absence of evidence for many organic molecule transport functions and the absence of many enzymes that are associated with metabolic versatility in methylotrophs suggest that this organism is a C1 specialist.

Similar to other Type I methylotrophs, HTCC2181 uses the RuMP pathway for the assimilation of carbon from C1 compounds (Fig. 6). Based on the genome annotation, HTCC2181 employs the cleavage route from fructose-6-phosphate (F-6-P) to glyceraldehyde-3-phosphate and pyruvate, rather than the route that leads from F-6-P to dihydroxy acetone phosphate and glyceraldehyde-3-phosphate. In the predicted pathway, one molecule of F-6-P is cleaved for every three molecules of formaldehyde that are condensed, and one molecule of pyruvate is routed to biosynthesis. HTCC2181 appears to have the cyclic formaldehyde oxidation pathway and the linear tetrahydrofolate formaldehyde oxidation pathway, both of which can, in principle, produce energy from C1 compounds (Marison and Atwood, 1982; Chistoserdova and Lidstrom, 1994). However, it does not possess the complete suite of genes necessary for the alternate formaldehyde oxidation pathway: linear oxidation via tetrahydromethanopterin (H4MPT) (Chistoserdova et al., 1998) (Fig. 6). The cyclic formaldehyde oxidation pathway utilizes several enzymes from the RuMP pathway; however, formaldehyde is oxidized rather than assimilated by this pathway, to generate reducing power in the form of NADH and NADPH. In the tetrahydrofolate-linked formaldehyde oxidation pathway formaldehyde is oxidized to formate, accompanied by production of one ATP and reduction of an NADP⁺ to NADPH. Formate is then oxidized to CO₂, releasing electrons to the menaquinone pool. Formate can also be used for biosynthesis, as the tetrahydrofolate pathway is not completely reversible to formaldehyde and there is no serine cycle to incorporate C1 compounds from 5,10-methylene THF. In addition to energy generation, these pathways may also serve an important role by detoxifying formaldehyde.

Like the obligate methylotrophs *Bacillus* 4B6 and C2A1 (Colby and Zatman, 1975a), HTCC2181 uses the Entner–Doudoroff pathway instead of the Embden–Meyerhof–Parnas pathway for the sugar cleavage step during C1 compound assimilation. This means that it is missing two reversible Embden–Meyerhof–Parnas enzymes used for gluconeogenesis: fructose-bisphosphate aldolase (EC 4.1.2.13) and fructose-bisphosphatase (EC 3.1.3.11). In addition to TCA cycle lesions, the absence of fructose-bisphosphate aldolase and fructose-bisphosphatase was offered (Colby and Zatman, 1975b) as a possible explanation for the inability of the strains they studied to use alternate compounds as sole carbon sources.

**Xanthorhodopsin and associated genes**

The genome of HTCC2181 encodes an unusual rhodopsin. Rhodopsin proton pumps in bacteria (proteorhodopsins) were first found in a group of abundant marine γ-proteobacteria known as the SAR86 clade (Béjà et al., 2000). This exciting discovery suggested that marine bacteria in oligotrophic environments might have the potential for photoheterotrophy. Interestingly, in phylogenetic trees inferred from amino acid sequences, the HTCC2181 rhodopsin did not cluster with proteorhodopsins, nor with rhodopsins of halophilic archaea, but formed a clade with the xanthorhodopsin of *Salinibacter ruber* (*Bacteroidetes*) and rhodopsins from divergent organisms like *Gloeobacter violaceus* (Cyanobacteria), *Roseiflexus* sp. RS-1 (*Chloroflexi*), *Pyrocystis lunula* (*Dinoflagellate*) and *Fulvimarina*...
pelagi (<i>α</i>-proteobacteria) (data not shown). The <i>S. ruber</i> xanthorhodopsin was shown to bind carotenoid molecules that serve as a light harvesting antenna, thus extending the xanthorhodopsin action spectrum (Balashov et al., 2005). The functions of the other rhodopsins in this clade are unknown. The rhodopsin gene of HTCC2181 is followed downstream by <i>idsA</i> and <i>crtIBY</i>, which encode for important enzymes in the biosynthesis of retinal, the chromophore of rhodopsin holoenzymes. This arrangement is identical to other PR-containing strains (Sabehi et al., 2005), including the SAR92 clade (Stingl et al., 2007a).

The HTCC2181 rhodopsin gene contains the conserved amino acids Asp (102) and Glu (113), which are hypothesized to be the proton acceptor and donor of rhodopsin proton pumps. Position 105, which was shown to be important for wavelength tuning in proteorhodopsins (Man et al., 2003), is occupied by a leucine, indicating an absorbance maximum in the green. As phytoplankton preferentially absorb blue and red light, the prediction of a green absorption maximum for the HTCC2181 rhodopsin is consistent with the hypothesis that the OM43 clade is adapted to conditions found in highly productive regions of the oceans.

<ref>Fig. 6. A comparison of the central pathways of carbon metabolism in HTCC2181 and <i>M. flagellatus</i> following the metabolic scheme proposed by Chistoserdova and colleagues (2007). The HTCC2181 pathways were reconstructed from genome sequence. Coding sequence designations for <i>M. flagellatus</i> that are missing in HTCC2181 are indicated by italics. Six of the instances of gene reduction indicated by the italics are the result of the elimination of paralogues. Genes for methylamine oxidation are entirely absent (<i>mauFBEAGLMN</i>) from HTCC2181. C1 transfer and oxidation is mediated by tetrahydrofolate in HTCC2181 instead of tetrahydromethanopterin.</ref>
bacter in its annotated regulatory functions, having *rpoD*, σ32, the heat-shock transcription factor, and a third sigma factor homologous to *rpoE2*, a member of a family of sigma factors that are implicated in sensory responses (Table S2). Also present were *spol* genes, including the histidine kinases implicated in the regulation of sigma factors. Other two-component systems implicated in osmotic regulation and control of phosphate, iron and nitrogen uptake were also present. As with *Nitrosococcus oceani* (Klotz et al., 2006) and other Proteobacteria, only three of six units for the phosphotransferase (PTS) system were found, a configuration speculatively associated with regulation rather than transport (Boel et al., 2006).

Conclusions

Most attention to C1 metabolism in the oceans has focused on methane oxidation. The unanticipated discovery of a successful metabolic specialist that uses oxidized C1 compounds suggests that the biogeochemical cycling of these molecules in the oceans needs continuing scrutiny. It is interesting to note that nine 16S rRNA genes related to the OM43 clade (similarity > 0.97) show up in the Global Ocean Survey metagenomic data, mainly in samples from the north-eastern continental shelf of North America. This information is consistent with other data suggesting that members of this clade are common in continental shelf seas, but are not a dominant fraction of the bacterioplankton community. As OM43 abundance appears to be linked to phytoplankton populations, it is tempting to speculate that phytoplankton are a source of methanol, like land plants. Heikes observed methanol in the headspace of phytoplankton cultures, but information about the production of methanol and other C1 compounds by phytoplankton is scarce (Heikes et al., 2002). In future work it will be of interest to more accurately measure the sources and turnover rates of methanol and other C1 compounds in seawater, and to study the relationships between OM43 populations and the flux of these compounds. It will also be of considerable interest to understand the role of the HTCC2181 xanthorhodopsin in the metabolism of this organism.

Experimental procedures

Isolation and growth conditions

Strain HTCC2181 was isolated from surface waters 5 miles off the Newport, Oregon, hydroline (44°N; hydroline NH-5) using dilution-to-extinction culturing in low-nutrient medium as described previously (Connon and Giovannoni, 2002). HTCC2181 was grown in sterilized and amended seawater media LNHM (Connon and Giovannoni, 2002). The media contained 10 μM NH₄Cl, 1 μM K₂HPO₄ and a vitamin mixture (Davis and Guillard, 1958). Experiments were performed in duplicate 100 ml acid-washed polycarbonate flasks incubated at 16°C in the dark unless mentioned otherwise. The growth of cultures was measured daily by staining cells with the double-stranded DNA binding dye SYBR Green I and counting by flow cytometry with a Guava flow cytometer, as described previously (Stingl et al., 2007b).

Genome sequencing and annotation

The genome of strain HTCC2181 was sequenced by the J. Craig Venter Institute (http://www.jcvi.org) as a part of the Moore Foundation Microbial Genome Sequencing Project (http://www.moore.org/microgenomie). We used a custom annotation pipeline that relied on four phases: gene prediction, gene refinement, data collection and function prediction. Glimmer2 (Delcher et al., 1999a) was used to predict genes in the raw genome sequence. Gene refinement occurred twice; the first phase refined the gene start position predicted by Glimmer2 based on potential upstream ribosome binding site motifs using RBSfinder (TIGR). Data collection proceeded after the first round of gene refinement; every gene was subjected to BLAST searches of NCBI, SWISSPROT and KEGG as well as Hidden Markov Model (HMM) searches of Pfam. The second phase of gene refinement used information from the database searches to further refine gene start locations. Start positions were confirmed or updated based on statistical models of potential upstream or downstream start locations constructed from comparisons to similar database sequences. Data from similarity searches were also used to resolve significant overlaps between neighbouring coding regions. The final phase of the pipeline collected data from the database searches to infer a functional assignment for every gene. Function prediction relied on ranking database hits based on a scoring matrix generated dynamically for every gene. The scoring matrix was constructed by natural language processing of database hits to score words and phrases based on the qualities of the hits. The distribution of scores provided a statistical estimation of the quality of the prediction made by the pipeline. Automated function prediction was an unsupervised process. However, often the automated annotations were accurate enough to provide an initial estimation of potential physiological and metabolic pathways.

Phylogenetic analyses

We aligned 16S rDNA sequences against those in a database of over 20 000 16S rDNA sequences maintained with the ARB software package. We carried out phylogenetic analyses with the program PAUP* 4.0 beta 8, and included 1047 unambiguously aligned nucleotide positions (Swofford, 2002). The tree topology was inferred from evolutionary distances calculated with the Kimura two-parameter model for nucleotide change, a transition/transversion ratio that was estimated from the data, and neighbour joining. Bootstrap proportions from 1000 resamplings were calculated using parsimony and evolutionary distance methods. Parsimony analyses employed a heuristic search, tree bisection-reconnection, and a starting tree obtained by stepwise addition with random sequence addition. Similar approaches...
were used for the analyses of protein coding sequences (proteorhodopsin and methanol dehydrogenase), except that CLUSTAL W was used for the alignment and standard distances with among site variation were used to estimate evolutionary distances from the predicted amino acid sequences.

Nucleotide sequences accession number

The sequence reported in this study has been deposited in GenBank under the Accession No. AAUX00000000.

Acknowledgements

This research was supported by grants from the Gordon and Betty Moore Foundation Marine Microbiology Initiative, National Science Foundation Grant DEB-0207085, and NSF Science and Technology Center Award EF-0424599. We are grateful to the anonymous reviewers who commented extensively and expertly on our manuscript, leading to many improvements.

References


data between these lines are estimated to include 95% of all future values based on the current population of values. 

**Table S1.** Transporters. 
**Table S2.** Regulatory proteins.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the Correspondence for the article.