Reports

Genome Streamlining in a Cosmopolitan Oceanic Bacterium

Stephen J. Giovannoni,1* H. James Tripp,1 Scott Givan,2 Mircea Podar,3 Kevin L. Vergin,1 Damon Baptista,3 Lisa Bibbs,3 Jonathan Eads,3 Toby H. Richardson,3 Michiel Noordewier,3 Michael S. Rappé,4 Jay M. Short,3 James C. Carrington,2 Eric J. Mathur3

The SAR11 clade consists of very small, heterotrophic marine α-proteobacteria that are found throughout the oceans, where they account for about 25% of all microbial cells. Pelagibacter ubique, the first cultured member of this clade, has the smallest genome and encodes the smallest number of predicted open reading frames known for a free-living microorganism. In contrast to parasitic bacteria and archaea with small genomes, P. ubique has complete biosynthetic pathways for all 20 amino acids and all but a few cofactors. P. ubique has no pseudogenes, introns, transposons, extrachromosomal elements, or inteins; few paralogs; and the shortest intergenic spacers yet observed for any cell.

Pelagibacter ubique, strain HTCC1062, belongs to one of the most successful clades of organisms on the planet (1), but it has the smallest genome (1,308,759 base pairs) of any cell known to replicate independently in nature (Fig. 1). In situ hybridization studies show that these organisms occur as unattached cells suspended in the water column (1). They grow by assimilating organic compounds from the ocean’s dissolved organic carbon (DOC) reservoir, and can generate metabolic energy either by a light-driven proteorhodopsin proton pump (2) or by respiration (3). The marine planktonic environment is poor in nutrients, and the availability of N, P, and organic carbon typically limits the productivity of microbial communities. P. ubique is arguably the smallest free-living cell that has been studied in a laboratory, and even its small genome occupies a substantial fraction (~30%) of the cell volume. The small size of the SAR11 clade cells fits a model proposed by Button (4) for natural selection acting to optimize surface-to-volume ratios in oligotrophic cells, such that the capacity of the cytoplasm to process substrates will be matched to steady-state membrane transport rates.

Surprisingly, this genome appears to encode nearly all of the basic functions of α-proteobacterial cells (Table 1). The small genome size is attributable to the nearly complete absence of nonfunctional or redundant DNA and the paring down of all but the most fundamental metabolic and regulatory functions. For example, P. ubique falls at the extreme end of the range for intergenic DNA regions, with a median spacer size of only three bases (Fig. 2). Intergenic DNA regions vary considerably among bacteria and archaea, even including parasites that have small genomes (5). No pseudogenes, phage genes, or recent gene duplications were found in P. ubique.

To further explore this trend, we investigated paralogous gene families by means of BLAST clustering with variable threshold limits. The genome had the smallest number of paralogous genes observed in any free-living cell (Fig. 1) (fig. S1). A steep slope in

---

*To whom correspondence should be addressed.
E-mail: steve.giovannoni@oregonstate.edu

Fig. 1. Number of predicted protein-encoding genes versus genome size for 244 complete published genomes from bacteria and archaea. P. ubique has the smallest number of genes (1354 open reading frames) for any free-living organism.

---

1Department of Microbiology, 2Center for Gene Research and Biotechnology, Oregon State University, Corvallis, OR 97331, USA. 3Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA. 4Hawaii Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawaii, Post Office Box 1346, Kaneohe, HI 96744, USA.

www.sciencemag.org
The streamlining hypothesis has been used to explain genome reduction in *Prochlorococcus*, a photoautotroph that reaches population sizes in the oceans that are similar to those of *Pelagibacter* (7–9). *Prochlorococcus* genomes range from 1.66 to 2.41 million base pairs (Mbp). Many organisms with reduced genomes, including some pathogens, also have very low G:C to A:T ratios (10) (fig. S3), which can be attributed to biases in mutational frequencies, but alternatively might convey a selective advantage by lowering the nitrogen requirement for DNA synthesis, thereby reducing the cellular requirement for fixed forms of nitrogen (7). N and P are both proportionately important constituents of DNA that are frequently limiting in seawater. The *P. ubique* genome is 29.7% G+C. Of four complete *Prochlorococcus* genome sequences, the two that lack the DNA repair enzyme 6-0-methylguanine-DNA methyltransferase also have very low G:C to A:T ratios. In the absence of this enzyme, the extent of accepted G:C to A:T mutations increases; however, the *P. ubique* genome encodes this enzyme, which suggests that other factors are the cause of its low G:C to A:T ratio.

Annotation revealed a spare metabolic network encoding a variant of the Entner-Duodoroff pathway, a tricarboxylic acid (TCA) cycle, a glyoxylate bypass, and a typical electron transport chain (Table 1). Anaplerotic pathways for cellular constituents, other than five vitamins, appeared to be complete, but genes that would confer alternate metabolic lifestyles, motility, or other complexities of structure and function were nearly absent. Conspicuous exceptions were genes for carotenoid synthesis, retinal synthesis, and proteorhodopsin. *P. ubique* constitutively expresses a light-dependent retinylidene proton pump and is the first cultured bacterium to exhibit the gene that encodes it (2). The genome also contained genes for type II secretion (including adhesion) and type IV pilin biogenesis. Examination of gene distributions among metabolic categories (fig. S4) supported the conclusion that genome reduction in *P. ubique* has spared genes for core proteobacterial functions while reducing the proportion of the genome devoted to noncoding DNA. Relative to other *α*-proteobacterial genomes, the proportions of *P. ubique* genes encoding transport functions, biosynthesis of amino acids, and energy metabolism were high (table S3).

The sheer size of *Pelagibacter* populations indicates that they consume a large proportion of the labile DOC in the oceans. The global DOC pool is estimated to be 6.85 × 10^{17} g C (11), roughly equaling the mass of inorganic C in the atmosphere (12). Examination of the *P. ubique* genome revealed that about half of all transporters, and nearly all nutrient-uptake transporters, are members of the ATP-binding cassette (ABC) family (table S1). ABC transporters typically have high substrate affinities and therefore provide an advantage at the cost of ATP hydrolysis. Inferred transport functions included the uptake of a variety of nitrogenous compounds: ammonia, urea, basic amino acids, spermidine, and putrescine. Broad-specificity transporters for sugars, branched amino acids, dicarboxylic and tricarboxylic acids, and a number of specialized osmolytes (including glycine betaine, proline, mannitol, and 3-dimethylsulfoniopropionate) were found in the genome. Autoradiography with native populations of SAR11 has demonstrated high uptake activity for amino acids and 3-dimethylsulfoniopropionate (13). Hence, efficiency is achieved in a low-nutrient system by reliance on transporters with broad substrate ranges (14) and a number of specialized substrate targets, in particular, nitrogenous compounds and osmolytes.

---

**Table 1. Metabolic pathways in Pelagibacter.**

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycolysis</td>
<td>Uncertain</td>
</tr>
<tr>
<td>TCA cycle</td>
<td>Present</td>
</tr>
<tr>
<td>Glyoxylate shunt</td>
<td>Present</td>
</tr>
<tr>
<td>Respiration</td>
<td>Present</td>
</tr>
<tr>
<td>Pentose phosphate cycle</td>
<td>Present</td>
</tr>
<tr>
<td>Fatty acid biosynthesis</td>
<td>Present</td>
</tr>
<tr>
<td>Cell wall biosynthesis</td>
<td>Present</td>
</tr>
<tr>
<td>Biosynthesis of all 20 amino acids</td>
<td>Present</td>
</tr>
<tr>
<td>Heme biosynthesis</td>
<td>Present</td>
</tr>
<tr>
<td>Ubiquinone</td>
<td>Present</td>
</tr>
<tr>
<td>Nicotinate and nicotinamide</td>
<td>Present</td>
</tr>
<tr>
<td>Folate</td>
<td>Present</td>
</tr>
<tr>
<td>Riboflavin</td>
<td>Present</td>
</tr>
<tr>
<td>Pantothenate</td>
<td>Absent</td>
</tr>
<tr>
<td>B₆</td>
<td>Absent</td>
</tr>
<tr>
<td>Thiamine</td>
<td>Absent</td>
</tr>
<tr>
<td>Biotin</td>
<td>Absent</td>
</tr>
<tr>
<td>B₁₂</td>
<td>Absent</td>
</tr>
<tr>
<td>Retinal</td>
<td>Present</td>
</tr>
</tbody>
</table>
The genome encoded two sigma factors, the heat shock factor σ^{32} and a σ^{70} (rpoD), but no homolog of rpoN, the gene for the nitrogen starvation factor σ^{54} (table S2). Only four two-component regulatory systems were identified, three of which match the only two-component regulatory systems in Rickettsia (15). The presence of homologs to PhoR/PhoB/PhoC, NtrY/NtrX, and envZ/OmpR suggested regulated responses to phosphate limitation, N limitation, and osmotic stress. The only additional two-component system, RegB/RegA, has been implicated in the regulation of cellular oxidation/reduction processes in phototrophic α-proteobacteria (16). A gene encoding a ferric iron uptake regulator was also present.

In its simplicity the P. ubique genome is unique among other heterotrophic marine bacteria, such as Vibrio sp. (17), Pseudalteromonas (18), Shewanella (19), and Silicibacter (20), which have considerably larger genomes (4.0 to 5.3 Mbp) and global regulatory systems that enable them to implement a variety of metabolic strategies in response to environmental variation. We hypothesize that P. ubique makes use of the ambient DOC field (21), whereas heterotrophic bacterioplankton with larger genomes are poised to rapidly exploit pulses of nutrients (22) at the expense of replication efficiency during the intervening periods (23). This hypothesis is consistent with the observation that P. ubique has a single ribosomal RNA (rRNA) operon and a low growth rate (0.40 to 0.58 cell divisions per day) that does not vary in response to nutrient addition. In contrast, heterotrophic marine bacteria with large genomes have some of the highest recorded growth rates and are very responsive to nutrient concentration.

Like some other α-proteobacteria and especially archaea, HTCC1062 has an alternate thymidylate synthase for thymine synthesis,

![Fig. 3. Maximum likelihood phylogenetic tree for the gene encoding RNA polymerase subunit B. Sequences represented by accession numbers are environmental sequences from the Sargasso Sea (19). The sequence indicated by a star is part of the 5.7-kb contig IBEA_CTG_2159647 that is part of a conserved gene cluster also present in Pelagibacter ubique. Numbers indicated by solid arrowheads represent amino acid percentage identity to the Pelagibacter gene. For comparison, the identity between two species of Mesorhizobium is also indicated (open arrowhead). Bootstrap support (100 maximum-likelihood replicates) is indicated for the major clades (* if less than 50).](image-url)
thyX (24). As in other strains that lack the most common thymidylate synthase (thyA) but have thyX, HTCC1062 also lacks the dihydrofolate reductase folA (25). Evidence suggests that the gene encoding thyX can substitute for folA (24). A full glycolytic pathway was not reconstructed because of the confounding diversity of glycolytic pathways (26). Five enzymes in the canonical glycolytic pathway were not seen, including two key enzymes involved in allosteric control: phosphofructokinase and pyruvate kinase. An enzyme thought to substitute for pyruvate kinase (27), known as PDPK (pyruvate-phosphate dikinase), was found. Some but not all of the enzymes for the nonphosphorylated Entner-Doudoroff pathway, considered more ancient than canonical glycolysis (26, 28), were detected, as well as a complete pathway for gluconeogenesis, also considered more ancient than canonical glycolysis (29). Sugar transporters with best BLAST hits to maltose/trehalose transport were found, so presumably a complete glycolytic pathway does function in this cell.

Whole-genome shotgun (WGS) sequence data from the Sargasso Sea segregated at high similarity values, relative to other α-proteobacteria and proteobacteria, in a BLASTN analysis of the P. ubique genome (fig. S4). Sequence diversity prevented Venter et al. (19) from reconstructing SAR11 genomes from the Sargasso Sea WGS data set, although SAR11 rRNA genes accounted for 380 of 1412 16S rRNA genes and gene fragments they recovered (26.9%), and the library was estimated to encode the equivalent of about 775 SAR11 genomes. Three Sargasso Sea contiguous sequences (contigs) that were long (5.6 to 22.5 kb) and highly similar to the P. ubique genome were analyzed in detail. Genes on these contigs were synteny with genes from the P. ubique genome, with amino acid sequence identities ranging from 68 to 96% (fig. S5). Phylogenetic analysis of four conserved genes from these contigs (those encoding RNA polymerase subunit B, Fig. 3; elongation factor G, fig. S6; DNA gyrase subunit B, fig. S7; and ribosomal protein S12, fig. S8) showed them to be associated with large, diverse environmental clades that branched within the α-proteobacteria. We hypothesize that evolutionary divergence within the SAR11 clade and the accumulation of neutral variation are the most likely explanations for the natural heterogeneity in SAR11 genome sequences.

Metabolic reconstruction failed to resolve why P. ubique will not grow on artificial media. When cultured in seawater, it attains cell densities similar to populations in nature, typically $10^5$ to $10^6$ ml$^{-1}$ depending on the water sample (3). No evidence of quorum-sensing systems was found in the genome, and experimental additions of nutrients supported the results from metabolic reconstruction, which suggests that an unusual growth factor may play a role in the ecology of this organism. P. ubique has taken a tack in evolution that is distinctly different from that of all other heterotrophic marine bacteria for which genome sequences are available. Evolution has divested it of all but the most fundamental cellular systems such that it replicates under limiting nutrient resources as efficiently as possible, with the outcome that it has become the dominant clade in the ocean.

References and Notes
19. J. C. Venter et al., *Science* 304, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
30. Supported by NSF grant EF0307223, Diversa Corporation, the Gordon and Betty Moore Foundation, and the Oregon State University Center for Gene Research and Biotechnology. We thank S. Wells, M. Hudson, D. Barofsky, M. Staples, J. Garcia, B. Buchner, P. Sammon, K. Li, and J. Ritter for technical assistance and J. Heidelberg for advice about genome assembly. We also acknowledge the crew of the R/V Elakha for assistance with sample and seawater collections, the staff of the Central Services Laboratory at Oregon State University for supplementary sequence analyses, and the staff of the Mass Spectrometry Laboratory at Oregon State University for proteomic analyses. The sequence reported in this study has been deposited in GenBank under accession number CP000084.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5738/1242/DC1
Materials and Methods
Tables S1 to S3
Figures S1 to S9
References
26 April 2005; accepted 11 July 2005
10.1126/science.1114057

Contact-Dependent Inhibition of Growth in *Escherichia coli*
Stephanie K. Aoki, Rupinderjit Pamma, Aaron D. Hernday, Jessica E. Bickham, Bruce A. Braaten, David A. Low*

Bacteria have developed mechanisms to communicate and compete with each other for limited environmental resources. We found that certain *Escherichia coli*, including uropathogenic strains, contained a bacterial growth-inhibition system that uses direct cell-to-cell contact. Inhibition was conditional, dependent upon the growth state of the inhibitory cell and the pili expression state of the target cell. Both a large cell-surface protein designated Contact-dependent inhibitor A (CdiA) and two-partner secretion family member CdiB were required for growth inhibition. The CdiAB system may function to regulate the growth of specific cells within a differentiated bacterial population.

Bacteria communicate with each other in multiple ways, including the secretion of signaling molecules that enable a cell population to determine when it has reached a certain density or that a potential partner is present for conjugation (1, 2). Cellular communication can also occur through contact between cells, as has been shown for *Myxococcus xanthus*, which undergoes a complex developmental pathway (3, 4). Here we describe a different type of intracellular interaction in which bacterial growth is regulated by direct cell-to-cell contact.

Wild-type *Escherichia coli* isolate EC93 inhibited the growth of laboratory *E. coli* K-12