AEM Accepts, published online ahead of print on 25 May 2007
Appl. Environ. Microbiol. doi:10.1128/AEM.02580-06

Ecological Genomics of Marine Roseobacters

M. A. Moran[1*], R. Belas[2], M. A. Schell[3], J. M. González[4], F. Sun[1], S. Sun[1], B. J. Binder[1], J. Edmonds[1], W. Ye[1], B. Orcutt[1], E. C. Howard[3], C. Meile[1], W. Palefsky[5], A. Goesmann[6], Q. Ren[7], I. Paulsen[7], L. E. Ulrich[8], L. S. Thompson[10], E. Saunders[10], and A. Buchan[11]

[1]Department of Marine Sciences, University of Georgia, Athens, GA 30602, USA
[2]Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt St., Baltimore, MD 21202, USA
[3]Department of Microbiology, University of Georgia, Athens, GA 30602, USA
[4]Department of Microbiology, University of La Laguna, 38071 La Laguna, Tenerife, Spain
[5]Institute of Ecology, University of Georgia, Athens, GA 30602, USA
[6]Center for Biotechnology, Bielefeld University, D-33594 Bielefeld, Germany
[7]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA
[8]Joint Institute for Computational Sciences, The University of Tennessee - Oak Ridge National Laboratory, Oak Ridge, TN 37831-6173
[9]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA
[10]Bioscience Division and DOE Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, NM 87545
[11]Department of Microbiology, University of Tennessee, Knoxville, TN 37996-0845

*Correspondence: Mary Ann Moran, Department of Marine Science, University of Georgia, Athens, GA 30602; voice: 706-542-6481; fax: 706-542-5888; e-mail: mmoran@uga.edu.

Abstract

Bacterioplankton of the marine Roseobacter clade have genomes that reflect a dynamic environment and diverse interactions with marine plankton. Comparative genome sequence analysis of three cultured representatives suggests that cellular requirements for nitrogen are largely provided by regenerated ammonium and organic compounds (polyamines, allophanate, urea), while typical sources of carbon include amino acids, glyoxylate, and aromatic metabolites. An unexpectedly large number of genes are predicted to encode the production, degradation, and efflux of toxins and metabolites. A mechanism likely involved in cell-to-cell DNA or protein transfer was also discovered: *vir*-related genes encoding a Type IV secretion system typical of bacterial pathogens. These suggest a potential for interacting with neighboring cells and impacting the routing of organic matter into the microbial loop. Genes shared among the three roseobacters and also common in nine draft Roseobacter genomes include those for carbon monoxide oxidation, DMSP demethylation, and aromatic compound degradation. Genes shared with other cultured marine bacteria include those for utilizing sodium gradients, transport and metabolism of sulfate, and osmoregulation.

Introduction

In surface waters of the open ocean, one in ten bacterial cells is a member of the Roseobacter group (17). In coastal waters, the number of Roseobacter cells increases to one in five (11, 19). Despite their obvious ecological success, however, roseobacters do not fit the stereotype of a small, metabolically conservative, 'oligotrophic' bacterium (8, 18). Instead, they are large (0.08 μm^3) (38), easily cultured (19), and respond readily to increased substrate availability (7). Analysis of the first Roseobacter genome sequence,

24 that of *Silicibacter pomeroyi*, revealed a fairly large genome (4.5 Mb) housing abundant
25 and diverse transporters, complex regulatory systems, and multiple pathways for
26 acquiring carbon and energy in seawater. Roseobacters thus appear to be quite versatile
27 from metabolic and ecological standpoints (43), with an assortment of strategies for
28 obtaining carbon and nutrients and, directly or indirectly, affecting the biogeochemical
29 status of seawater.

30 The availability of two additional closed genome sequences of cultured
31 roseobacters provides the opportunity for an ecologically-based analysis of the genetic
32 capabilities of this bacterial taxon. The three organisms are assumed to have different
33 niches in the surface ocean based on the conditions of their isolation: *S. pomeroyi* is a
34 free-living heterotrophic bacterioplankter obtained from coastal seawater (43); congener
35 *Silicibacter* sp. TM1040 (96% 16S rRNA sequence identity to *S. pomeroyi*; Fig. 1) is an
36 associate of the marine dinoflagellate *Pfiesteria piscicida* (1, 40); and *Jannaschia* sp.
37 CCS1 (with 94% 16S rRNA sequence identity to the two *Silicibacter* species) represents
38 a recently discovered class of marine aerobic bacteriochlorophyll *a*-based phototrophs
39 (4). Our comparative analysis of the three Roseobacter genomes centered on three
40 questions: 1) What physiological and ecological traits of roseobacters can be inferred
41 from the genome sequences (focusing on genes that define ecological strategies and
42 biogeochemical roles)?; 2) What makes a Roseobacter a Roseobacter (focusing on genes
43 shared among the species)?; and 3) What makes a Roseobacter marine (focusing on genes
44 shared with other cultured marine bacteria from phylogenetically distant taxa)?

Materials and Methods

Organisms. *Silicibacter pomeroyi* strain DSS-3 was isolated from Atlantic coastal seawater (Georgia, USA; salinity = 31) in 1999 on a minimal seawater medium enriched with 10 μM dimethylsulfoniopropionate (DMSP). *Silicibacter* sp. TM1040 was isolated from a marine dinoflagellate culture (*Pfiesteria piscicida)* in 2000 on 0.5x Marine Agar 2216. *Jannaschia* sp. CCS1 was isolated from Pacific coastal seawater (Bodega Head, California, USA; salinity = 33) in 2003 on 1/10 strength YTSS agar.

Genomic sequencing and annotation. *Silicibacter* sp. TM1040 and *Jannaschia* sp. CCS1 were sequenced by the random shotgun method (see www.jgi.doe.gov for details of library construction and sequencing). Large, medium, and small insert random libraries (38 kb, 9.5 kb, and 3 kb for *Silicibacter* sp. TM1040; 38 kb, 13 kb, and 5.6 kb for *Jannaschia* sp. CCS1) were sequenced with an average success rate of 96% and an average high-quality read length of 685 nucleotides. Reads were assembled with parallel Phrap (High Performance Software, LLC), and possible misassemblies were corrected with Dupfinisher (23) or transposon bomb of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing, custom primer walking, or PCR amplification. The completed genome sequence of *Silicibacter* sp. TM1040 contains 59,153 reads, achieving an average of 13-fold sequence coverage per base. The completed genome sequence of *Jannaschia* sp. CCS1 contains 67,683 reads, achieving an average of 14-fold sequence coverage per base. The error rate was < 1 in 100,000. Sequencing and annotation of *S. pomeroyi* DSS-3 is described in Moran et al. (43).

Sequence analysis and annotation. The *Silicibacter* sp. TM1040 and *Jannaschia* sp. CCS1 genomes were annotated with the GenDB system (39).

Synteny analysis. Comparative genetic arrangements were analyzed by the Artemis Comparison Tool (50). Files were prepared using separate one-way tBLASTx analysis of chromosomes, megaplasmids, and plasmids with cut-offs of >45% identity and scores >35.

Signal transduction analysis. Signal transduction proteins were identified by searching against the Pfam (3) and SMART (35) domain databases and filtering for proteins that contained domains implicated in signal transduction (55). Transmembrane regions were predicted using Phobius (27). Cultured marine bacteria used in comparisons were *Cellulophaga* sp. MED134, *Croceibacter atlanticus* HTCC2559, *Flavobacterium* sp. MED217, *Polaribacter irgensii* 23-P, *Robiginitalea biformata* HTCC2501, *Tenacibaculum* sp. MED152, *Cytophaga hutchinsonii, Crocosphaera watsonii* WH8501, *Synechococcus* sp. WH8102, *Prochlorococcus marinus* MIT9313, *Bacillus* sp. NRRL B-14911, *Oceanobacillus iheyensis* HTE831, *Rhodopirellula* sp. 1, *Pelagibacter ubique* HTCC1062, *Erythrobacter litoralis* HTCC2594, *Erythrobacter* sp. NAP1, *Nitrobacter* sp. Nb-311A, *Oceanicaulis alexandrii* HTCC2633, *Parvularcula bermudensis* HTCC2503, *Sphingopyxis alaskensis* RB2256, *Alteromonas macleodii, Colwellia psychrerythraea* 34H, *Idiomarina baltica* OS145, *Idiomarina loihiensis* L2TR, *Marinobacter aquaeolei* VT8, *Marinomonas* sp. MED121, *Nitrosococcus oceani* ATCC 19707, *Oceanospirillum* sp. MED92, *Photobacterium profundum* SS9, *Photobacterium* sp. SKA34, *Pseudoalteromonas atlantica* T6c, *Pseudoalteromonas haloplanktis* TAC125, *Shewanella baltica* OS155, *Shewanella frigidimarina* NCIMB 400, *Vibrio fischeri* ES114, *Vibrio splendidus* 12B01, and *Magnetococcus* sp. MC-1. α-Proteobacteria used in comparisons were *Agrobacterium tumefaciens* C58 UWash,

Caulobacter crescentus CB15*, Mesorhizobium loti* MAFF303099*, Mesorhizobium* sp. BNC1, *Paracoccus denitrificans* PD1222, *Rhodobacter sphaeroides* 2.4.1, *Rhodopseudomonas palustris* BisB18, and *Sinorhizobium meliloti* 1021.

Transporter profiles and ANOSIMS analysis. Complete predicted protein sequences were searched against a curated set of proteins with family assignment for similarity to known or putative transporter proteins and against a non-redundant general protein database using a semi-automated pipeline (48). Manual annotation for final assignments was based on the number of hits to the transporter database, maximum, minimum, and average BLAST E-values, the description of top hits to the general protein database, and COG assignments. The roseobacters were compared to a selected group of bacteria for which similar transporter analysis was available, consisting of six closely-related non-marine α-Proteobacteria (*A. tumefaciens* C58 Uwash, *C. crescentus* CB15, *R. palustris* CGA009, *M. loti* MAFF303099, *R. sphaeroides* 2.4.1, *S. meliloti* 1021) and seven marine heterotrophic bacteria (*Vibrio vulnificus* CMCP6, *Desulfotalea psychrophila* LSv54, *Oceanobacillus iheyensis* HTE831, *Rhodopirellula* sp. 1, *Colwellia psychrerythraea* 34H, *Photobacterium profundum* SS9, and *Idiomarina loihiensis* L2TR). Using the 70 transporter families that had $\geq$2 occurrences in at least one genome, a Bray-Curtis similarity matrix was created. A one-way analysis of similarity (ANOSIM; ref 15) was used to test for significant differences in transporter composition among the three groups defined *a priori*. Differences among groups were visualized using non-metric multi-dimensional scaling of the similarity matrix (PRIMER 5 software; ref 15).

Identification of shared and unique genes. Genes shared among the three Roseobacter strains were identified by three-way Reciprocal Best Hit (RBH) analysis

using BLAST thresholds of E< 10^{-5} and amino acid identity of >30%. Core genes in the 12 total Roseobacter genome sequences were identified by sequential 2-way RBH analysis, beginning with the *S. pomeroyi* genome and adding the other 11 genomes one at a time. Core genes were defined as those found in all 12 genomes based on RBH to *S. pomeroyi* genes. As each genome was sequentially added to the analysis, the total gene count was incremented if a gene had no RBH in any previous genome. Both the core gene count and total gene count are likely to be underestimated with this approach because nine of the genomes are draft. A second approach, the "in silico genomic subtraction" (ISGS) method (16), was also used because the RBH method misses orthologs if ORFs are overlooked or miscalled. This method is based on sequential subtractions of ORFs based on one-way BLASTp and tBLASTn analysis. *S. pomeroyi* genes were blasted sequentially against the genomes of non-marine α-Proteobacteria relatives *R. sphaeroides* 2.4.2, *P. denitrificans* PD1222, *S. meliloti* 1021, and *M. loti* MAFF303099 and orthologous genes were removed from the dataset. Remaining ORFs were compared by tBLASTn to the genome sequences of *Silicibacter* sp. TM1040 and *Jannaschia* sp. CCS1 to obtain a final ISGS unique gene set consisting of 121 ORFs shared by all three Roseobacters but missing from the four non-marine α-Proteobacteria genomes. The final composition of this gene set is affected by the bioinformatic decision of which α-Proteobacteria relatives to use in the subtraction, but consists largely of genes more typical of roseobacters than their α-Proteobacteria relatives.

Identification of genes with a marine signal. We determined whether orthologs of Roseobacter genes were found more commonly in marine than in non-marine genomes. Twenty pairs of non-Roseobacter genomes were assembled, each consisting of

137 the genome of a marine isolate (most from surface seawater) paired with its closest non-
138 marine relative at the time of the analysis: *Synechococcus* sp. WH8102 and
139 *Synechococcus elongatus* PCC 7942; *Rhodopirellula baltica* SH 1 and *Moorella*
140 *thermoacetica* ATCC 39073; *P. ubique* HTCC1062 and *Wolbachia* sp. TRS; *P. marinus*
141 MIT9313 and *Gloeobacter violaceus* PCC 7421; *Pseudoalteromonas haloplanktis*
142 TAC125 and *Legionella pneumophila*; *Shewanella frigidimarina* NCIMB 400 and
143 *Shewanella putrefaciens* CN-32; *Magnetococcus* sp. MC-1 and *Rhodospirillum rubrum*;
144 *Sphingomonas alaskensis* and *Zymomonas mobilis* ZM4; *Erythrobacter litoralis*
145 HTCC2594 and *Novosphingobium aromaticivorans* F199; *Flavobacterium* MED217 and
146 *Bacteroides fragilis* ATCC25285; *C. atlanticus* HTCC2559 and *Porphyromonas*
147 *gingivalis* W83; *V. splendidus* 12B01 and *Escherichia coli* K12; *I. baltica* OS145 and
148 *Pseudomonas putida* KT2440; *Janibacter* sp. HTCC2649 and *Arthrobacter* sp. FB24;
149 *Parvularcula bermudensis* HTCC2503 and *Brucella abortus* 9-941; *Bacillus* sp. NRRL
150 B-14911 and *Bacillus halodurans* C-125; *Marinomonas* sp. MED121 and *Pseudomonas*
151 *syringae* B728a; *Nitrosococcus oceani* ATCC 19707 and *Pseudomonas fluorescens* Pf0-
152 1; *Oceanicaulis alexandrii* HTCC2633 and *R. sphaeroides* 2.4.2; and *Nitrobacter* sp.
153 Nb311A and *R. palustris* BisB18. Orthologs to Roseobacter genes were identified in
154 these genomes with BLASTp analysis and using the criteria of reciprocal best hit, an E
155 value of $\leq 10^{-5}$, and a percent identity of $\geq 30\%$. Whether the observed gene distribution
156 was different than expected (the null hypothesis being that the gene occurred in an equal
157 number of marine and non-marine genomes) was estimated by computing a Chi-square
158 statistic and associated p value.

159 **Accession Numbers**

The complete *Silicibacter* sp. TM1040 and *Jannaschia* sp. CCS1 genome sequences are available under GenBank accession numbers NC_008043 and NC_007802. The complete *Silicibacter pomeroyi* DSS-3 sequence was previously deposited under GenBank accession numbers NC_003911. Nine draft Roseobacter genomes available for comparison at the time of analysis were *Loktanella vestfoldensis* SKA53 (AAMS00000000), *Roseovarius* sp. 217 (AAMV00000000), *Roseobacter* sp. MED193 (AANB00000000), *Sulfitobacter* sp. EE-36 (AALV00000000), *Roseovarius nubinhibens* ISM (AALY00000000), *Sulfitobacter* sp. NAS-14.1 (AALZ00000000), *Oceanicola batsensis* HTCC2597 (AAMO00000000), *Oceanicola granulosus* HTCC2516 (AAOT00000000), and *Rhodobacterales bacterium* HTCC2654 (AAMT00000000).

Results and Discussion

General genome organization and content. All three Roseobacter genomes contain ~4,000 genes (Table 1), but there is considerable variability in genome organization and content (Fig. S1). *S. pomeroyi* DSS-3 has a 491 Kb megaplasmid (pSPD1), *Jannaschia* sp. CCS1 has an 80 Kb plasmid (pJCS1), and *Silicibacter* sp. TM1040 has both an 823 Kb megaplasmid (pSTM1) and a 131 Kb plasmid (pSTM2). Neither the *S. pomeroyi* nor *Jannaschia* sp. CCS1 plasmids have rRNA operons, but the two plasmids of *Silicibacter* sp. TM1040 together harbor four of its five rRNA operons (Table 1). Designation of pSTM1 and pSTM2 as plasmids rather than chromosomes is based on the finding that none of 11 single-copy genes essential for bacterial growth (51) are found on either and that each contains a plasmid-type replication origin (*rep, parAB*) (12). The relatively small plasmids of *Silicibacter* sp. TM1040 and *Jannaschia* sp. CCS1

(pSTM2 and pJCS1) do not appear closely related and only nine genes are similar (including *rep*, *parAB*, several genes involved in sugar or polysaccharide metabolism, and a single ABC transporter-like protein). A plasmid from an uncharacterized Roseobacter isolate (149 kb pSD25) (57) likewise has only 15 genes related to those of pSTM2 and three related to those of pJCS1, suggesting the presence of a diverse Roseobacter gene pool harbored on small (~100-150 Kb) plasmids (Fig. S1).

Each genome contains a number of genes that have no orthologs in the other two; these account for 31% of the *S. pomeroyi* DSS-3 genome, 27% of the *Silicibacter* sp. TM1040 genome, and 39% of the *Jannaschia* sp. CCS1 genome (Table 2). No prophages are found in the *S. pomeroyi* or *Jannaschia* sp. CCS1 genomes, but four are present in the *Silicibacter* sp. TM1040 genome (Table S1). Three of these can be experimentally induced by mitomycin C treatment (9) and the fourth lacks required structural proteins and may be defective. Variable and dynamic genome organization seems to be the rule for the Roseobacter lineage (46).

Inferred ecology and physiology. Because of the numerical and ecological importance of marine roseobacters in surface ocean bacterioplankton communities, our annotation efforts paid greatest attention to genes of biogeochemical interest.

i. Interfacing with the environment – Genes encoding signal transduction proteins comprise approximately 7% of the Roseobacter genomes, similar to the average for a marine comparison group (6.3 $\pm$ 2.7%) and a non-marine α-Proteobacteria comparison group (8.1 $\pm$ 0.9%; Fig. 2). Two-component systems constitute a low-percentage (< 20%) of the overall signal transduction machinery in these Roseobacter species. Consequently, the cells respond to environmental stimuli predominantly via one-component systems that

206 account for 86% (302 out of 353) of the signaling proteins in the *S. pomeroyi* genome,
207 and slightly lower percentages in the *Silicibacter* sp. TM1040 (70%) and *Jannaschia* sp.
208 CCS1 (80%) genomes. The Roseobacter genomes contain a particularly high number of
209 AsnC/Lrp regulatory proteins, at >3 times the average for the marine comparison group
210 (for which genome sizes are comparable; Fig. 2). This suggests that amino acid
211 metabolism is particularly important and closely regulated in roseobacters; indeed several
212 of the AsnC/Lrp regulators are located adjacent to amino acid transporters or amino acid
213 metabolism-related genes. The roseobacters also have a larger number of other one-
214 component systems relative to other cultured marine bacteria, including GntR, IclR,
215 MarR, and TetR-like repressors. Generally, the Roseobacter signal transduction profiles
216 are more similar to those of their closest (and non-marine) α-Proteobacteria relatives than
217 to marine bacteria (Fig. 2).
218 One-component signal transduction systems with BLUF domains sense blue light
219 based on a conformational change following photon absorption (26). Homologs are found
220 in the genomes of *Silicibacter* sp. TM1040 (TM1040_2025 and 2027) and *Jannaschia* sp.
221 CCS1 (Jann_2030, Jann_2780, Jann_3321), but not *S. pomeroyi*. A two-component
222 bacterial phytochrome system that may be involved in detection of red and far-red light is
223 also found in *Silicibacter* sp. TM1040 (TM1040_2684). The presence of systems for
224 sensing ambient light is noteworthy in a taxon largely restricted to surface waters, and
225 may allow light sensing for phototrophy (in *Jannaschia* sp. CCS1), protective pigment
226 formation (47), or synchronization of activity with that of photosynthetic plankton.
227 Genomes of many other marine heterotrophs representing diverse taxa (but few non-
228 marine heterotrophs) likewise were found to contain proteins with BLUF domains,

229 including other Roseobacters (*Sulfitobacter* NAS-14.1, *Sulfitobacter* sp. EE-36,
230 *Roseobacter* sp. MED193, *Roseovarius* sp. 217), other α-Proteobacteria (*Sphingopyxis*
231 *alaskensis* SKA58), γ-Proteobacteria (*Alteromonas macleodii* "deep ecotype",
232 *Pseudoalteromonas tunicata*, *Marinomonas* sp. MED121, *Oceanospirillum* sp. MED92,
233 *Reinekea* sp. MED297, *Vibrio* sp. MED222, and *V. splendidus* 12B01), and Bacteroidetes
234 (*Psychroflexus torquis* and *Tenacibaculum* sp. MED152).
235 Swimming motility and chemotaxis behavior determine the ability of roseobacter
236 cells to position themselves in a heterogeneous environment. Motility has been
237 demonstrated experimentally in the three strains, and flagellar structural genes were
238 indeed found for all (Table S2). *Silicibacter* sp. TM1040 has six copies of the flagellar
239 filament gene *fliC* compared to just one for *S. pomeroyi* and *Jannaschia* sp. CCS1.
240 Chemotaxis receptors (methyl-accepting chemotaxis proteins; MCPs) and signal
241 transducers (*che* gene orthologs) are abundant in *Silicibacter* sp. TM1040 and *Jannaschia*
242 sp. CCS1, but surprisingly are completely lacking in *S. pomeroyi* (43) (Table S3). Twenty
243 MCPs are available to bind chemical attractants or repellants in the membrane or
244 cytoplasm of *Silicibacter* sp. TM1040, and 14 cytoplasmic chemotaxis signal transducers
245 receive signals from these MCPs to control changes in swimming direction. *Jannaschia*
246 sp. CCS1 has four MCPs and eight signal transducers. Initial clues about which
247 compounds stimulate chemotaxis in roseobacters have emerged from experimental
248 studies of *Silicibacter* sp. TM1040: positive chemotactic responses are induced by amino
249 acids, dinoflagellate cell homogenates, the algal osmolyte dimethylsulfoniopropionate
250 (DMSP), and DMSP breakdown products (41).

ABC-type transporters average 95 per genome for the three roseobacters compared to 150 for an α-Proteobacteria comparison group and 66 for a marine heterotroph group; Major Facilitator Superfamily (MFS) transporters average 22 per genome for roseobacters compared to 39 for α-Proteobacteria and 19 for marine heterotrophs. The set of Roseobacter transporter genes is not statistically different from either comparison group, although α-Proteobacteria and marine heterotrophs differ from each other (ANOSIM; $p = 0.008$; Fig. S2). Two particular transporter families are more abundant in roseobacters than in either comparison group, however: tripartite ATP-independent periplasmic (TRAP) transporters often associated with uptake of dicarboxylic acids and sugars (averaging 22 genes per genome compared to only seven for α-Proteobacteria and six for marine heterotrophs), and the drug/metabolite (DMT) superfamily involved in cellular export of toxins and metabolites (averaging 30 genes per genome compared to 16 for α-Proteobacteria and 14 for marine heterotrophs). Both of these families use ion gradients instead of ATP to drive substrate uptake, which may be beneficial given the external salt concentrations roseobacters experience. If so, not all marine heterotrophs utilize this transport strategy.

ii. Acquiring carbon and energy – The Roseobacter genomes reveal a surprising number of pathways for catabolism of structurally diverse aromatic substrates. A pathway for the degradation of protocatechuate was identified previously in members of the Roseobacter clade (6). Now, five other distinct pathways representing both oxygenase-dependent and -independent ring cleaving mechanisms have been found. These encode the aerobic degradation of the aromatic intermediates gentisate, homoprotocatechuate, homogentisate, benzoate, and phenylacetate. *S. pomeroyi* DSS-3 and *Jannaschia* sp.

CCS1 possess genes for all six pathways; *Silicibacter* sp. TM1040 possesses genes for four (Fig. 3). Possible substrates for this wealth of catabolic pathways more typical of soil bacteria include phenolic metabolites (e.g. antioxidants, toxins, sunscreens, and predator deterrents) from marine plankton (13, 14), as well as lignin derivatives from coastal marshes (42).

All three Roseobacter genomes contain a gene for a demethylase that mediates the first step in DMSP degradation (*dmdA*) (24). Half of the cultured roseobacters for which a genome sequence is available have this gene, and the DMSP demethylating phenotype has been shown experimentally for *S. pomeroyi* and *Silicibacter* sp. TM1040 (24, 40) (Table 3). DMSP can be a major source of both carbon and sulfur to marine bacterioplankton in ocean surface waters (30).

S. pomeroyi harbors two *cox* operons that have been shown experimentally to mediate carbon monoxide oxidation at the low concentrations typical of ocean surface waters ($\leq$ 5 nM) (43). The other two Roseobacter genomes likewise contain *cox* clusters: two in *Jannaschia* sp. CCS1 (Jann_1763, 1765, 1766 and Jann_2095, 2096, 2097) and one in *Silicibacter* sp. TM1040 (TM1040_1764, 1765, 1766). Without any apparent mechanism for CO_2 fixation, this pathway may be used by the roseobacters as an energy supplement, potentially resulting in higher heterotrophic growth yields in ocean surface waters where organic matter is limiting but CO is ubiquitous (albeit at low concentrations). Including the nine available draft sequences, all but one of twelve Roseobacter genomes contain at least one *cox* operon (Table 3).

Members of the Roseobacter clade have been found to contribute to aerobic anoxygenic phototrophy (AAnP) in ocean surface waters (4, 46). The genome sequence

297 of *Jannaschia* sp. CCS1, along with the recently published genome of *Roseobacter*
298 *denitrificans* (52), provides insight into the genes of a marine AAnP. With no
299 recognizable carbon fixation pathways, phototrophy may be used by *Jannaschia* sp.
300 CCS1 to obtain energy and thereby increase efficiency of heterotrophic growth, as
301 hypothesized for carbon monoxide oxidation. Forty-three genes form a 45.8 Kb
302 photosynthetic cluster (Jann_0412-0184) that includes genes for carotenoid biosynthesis
303 (*crt*), bacteriochlorophyll synthesis (*bch*), reaction center and light-harvesting complex I
304 (*puf*), and regulation (*ppaA* and *ppsR*). A smaller group of genes located 146 Kb
305 downstream of the main cluster contains genes for the light-harvesting complex II
306 (*pucBACD*; Jann_1581-1584). Overall, there is considerable congruence in the
307 photosynthetic gene complements of *Jannaschia* sp. CCS1, the marine aerobic
308 anoxygenic phototroph *Erythrobacter* sp. NAP1 (which is not a Roseobacter), and
309 several anaerobic anoxygenic phototrophs (Table S4 and Fig. S3). Neither the *S.*
310 *pomeroyi* nor *Silicibacter* sp. TM1040 genomes contain phototrophy genes. However,
311 searches of nine additional Roseobacter genomes available in draft form (Fig. 1) showed
312 that 25% of these cultured roseobacters are predicted to be phototrophs (Table 3). None
313 of the Roseobacter genomes contain evidence of proteorhodopsin-based phtototrophy.
314 All three Roseobacter genomes encode a complete tricarboxylic acid (TCA) cycle
315 as well as peripheral pathways such as phosphoenolpyruvate (PEP) carboxylase and
316 (except for *Silicibacter* sp. TM1040) the glyoxylate shunt, permitting growth on a variety
317 of organic acids (e.g. acetate, pyruvate). All three strains also harbor genes for the
318 Embden-Meyerhof, Entner-Doudoroff, and pentose phosphate pathways.

iii. Acquiring nitrogen and phosphorus – Organic nitrogen and ammonium appear to serve as the primary sources of cellular nitrogen for marine roseobacters. All three genomes contain genes for the assimilation of urea, amino acids, other nitrogen-rich organic compounds often associated with eukaryotic plankton (polyamines, and the osmolytes taurine and glycine betaine), and ammonium (Fig. 4). This reliance on organic nitrogen and ammonium is in keeping with both the observed vertical distribution of roseobacters (highest abundance in surface waters where regenerated rather than "new" nitrogen dominates) and frequent physical associations with organic N-rich phytoplankton cells (25). Neither nitrate nor nitrite transporters are found in the two *Silicibacter* genomes, and experimental evidence confirms that *S. pomeroyi* cannot assimilate nitrate (Figure S4). However, *Jannaschia* sp. CCS1 has genes for uptake and assimilatory reduction of nitrate (*nrtA,B,D, nirB, nirA, nasA*) (Fig. 4).

Inorganic P appears to be acquired by all three species through the high-affinity phosphate transport system (*pstSCAB*). Both *Silicibacter* species also have genes for transport (*phnCDE*) and cleavage (*phnGHIJKLN*) of organic P in the form of phosphonates (C-P bonded phosphorus), one of two major P organic reservoirs in seawater (Figure S5). Only the *Jannaschia* sp. CCS1 genome and one other draft Roseobacter genome are lacking phosphonate transport genes (Table 3). *S. pomeroyi* but not *Silicibacter* sp. TM1040 can also obtain organic phosphorus from the other major organic P reservoir, phosphoesters (C-O-P bonded phosphorus; mediated by *phoA*) (Fig. 4). While *Jannaschia* sp. CCS1 has neither of the organic phosphorus uptake systems, the presence of genes for inorganic P storage as polyphosphate (*ppk*) (which is missing in the

341 other two genomes) may provide an alternative strategy for maintaining a reliable supply
342 of phosphorus.

343 *iv. Interfacing with other organisms* - Roseobacters harbor a number of genes that
344 may be involved in information, metabolite, and DNA exchange with Roseobacter
345 relatives, other bacteria, and marine eukaryotes.

346 Nonribosomal peptide synthetases (NRPSs) catalyze the synthesis of a wide array
347 of bacterial antibiotics, toxins, siderophores, and other bioactive compounds from
348 common or modified amino acids. *S. pomeroyi* and *Silicibacter* sp. TM1040, but not
349 *Jannaschia* sp. CSS1, harbor a very similar 10-gene cluster containing two adjacent
350 NRPSs, an NRPS loading enzyme (phosphopantetheinyl transferase-family), and four
351 glycosyl transferases. Close homologs of these genes are not found outside the
352 Roseobacter group, suggesting they may encode a novel glycosylated peptide that plays a
353 role in defense, signaling, or host-microbe interactions with marine plankton (7).

354 Roseobacter genes may encode an alternative synthesis pathway for the plant
355 auxin indoleacetic acid (IAA) via indole-3-acetonitrile, including indoleacetamide
356 hydrolase and nitrile hydratase (32) in *S. pomeroyi* and *Silicibacter* sp. TM1040
357 (SPO2938, 1314,1315; TM1040_1578, 1971, 1972), and arylacetone-specific
358 nitrilase (33) in *Jannaschia* sp. CCS1 (Jann_3735). If this is the case, the potential
359 benefits of harboring IAA-related genes would include initiating cell wall loosening and
360 sugar release from phytoplankton cells, as has been found for vascular plants (36).

361 Twelve *vir*-related ORFs on the pSTM2 plasmid of *Silicibacter* sp. TM1040
362 indicate the presence of a Type IV secretion system for translocating DNA or proteins
363 into other cells. These genes have homology to *virD2* and *virD4*, which encode for the

relaxase and coupling proteins (providing the energetics for export of DNA), and *virB1-11* (excluding *virB7*), which encode the inner membrane channel and pilus structure of the transfer machinery (10) (Table S5). In *Agrobacterium tumefaciens*, homologous *vir* genes transfer the T-DNA into plant cells to initiate gall formation. The association between *Silicibacter* sp. TM1040 and dinoflagellate *P. piscicida* (1, 40, 41) tempts speculation that Vir proteins enable roseobacters to transfer DNA or protein directly to phytoplankton cells. The *S. pomeroyi* and *Jannaschia* sp. CCS1 genomes do not have *vir* orthologs (Table S5), although half of the twelve complete and draft Roseobacter genome sequences do (Table 3).

Overall, the Roseobacter gene collection suggests members of this clade have multiple mechanisms for sensing and reacting to their environment while acquiring diverse substrates and nutrients for growth. This is a distinctly different ecological tactic than the "genome streamlining" strategy of other successful marine heterotrophs (18).

Which genes make a Roseobacter a Roseobacter? Whether or not microbial biogeochemical functions can be predicted from the taxonomic composition of natural communities has been debated among microbial ecologists for more than a decade. While the answer may turn out to be different for each taxon, it hinges on whether functions encoded by a set of unique and predictable genes can be ascribed to most members of a bacterial clade. The Roseobacter lineage is a coherent group based on 16S rRNA phylogeny (7), yet the diversity in physiology and habitat within the taxon is considerable (56). Thus we asked whether shared genes can be identified that form the basis of a predictable ecological role for marine roseobacters.

386 The three roseobacters share 1,939 genes, accounting for ~50% of each genome
387 (Table 2). To focus specifically on unique genes, we first narrowed down the three-way
388 ortholog gene set by excluding genes that were also found in non-marine α-
389 Proteobacteria relatives. We used both protein (BLASTp) and nucleic acid (tBLASTn)
390 homology searches to condense the 1,939 three-way shared genes to 121 unique
391 Roseobacter genes based on their presence in all three Roseobacter genomes but their
392 absence in the *Rhodobacter sphaeroides* 2.4.2, *Paracoccus denitrificans*, *Sinorhizobium*
393 *meliloti* 1021, and *Mesorhizobium loti* MAFF303099 genomes (Table S6).
394 The 121 Roseobacter unique genes are clearly dominated by transport proteins,
395 which account for 37 (31%) of the gene set. Since transporters average 9% of ORFs in
396 Bacteria (48) and 6% in the roseobacters, this disproportionately high percentage argues
397 for a central role for transporters in defining the taxon. The compounds predicted to move
398 across Roseobacter membranes by the unique transporters include glyoxylate (a product
399 of algal photorespiration and organic matter photoxidation), acetate, allophanate (a
400 breakdown product of urea in chlorophytes and bacteria) (28), arginine, branched-chain
401 amino acids, ammonia, and secondary metabolites. The presence of three sodium-based
402 transporters (one symporter, one antiporter, and one voltage-gated ion channel) is
403 consistent with adaptations to a high-salt environment, as is the abundance of TRAP
404 transporters that likely depend on sodium ions for activity (29). Other gene categories
405 well represented in the unique genes are involved in aromatic compound degradation
406 (four genes), sulfur transport and metabolism (four genes), and methyl group transfer
407 (two genes). There were also genes for extracellular peptide-mediated signaling
408 (rhomboid protein homolog), extracellular degradation of the storage compound

409 polyhydroxybutyrate (poly-3-hydroxybutyrate depolymerase), and resistance to peroxides
410 (hydroperoxide resistance protein) (Table S6). Fifty-one (42%) of the 121 unique genes
411 are in at least 10 of the 12 Roseobacter genomes (Table S6). Megaplasmids pSPD1 and
412 pSTM1 of *S. pomeroyi* and *Silicibacter* sp. TM1040, respectively, contain unique genes
413 in an unbiased proportion relative to the total gene distribution (e.g., pSPD1 contains
414 10.4% of total genes and 12.4% of unique genes). No unique genes were found on the
415 small *Silicibacter* sp. TM1040 or *Jannaschia* sp. CCS1 plasmids (pJCS1 and pSTM2),
416 consistent with the idea that the smaller Roseobacter plasmids harbor a diverse and non-
417 essential gene pool.
418 We also analyzed Roseobacter genes using a more standard concept of "core"
419 genes, defined here as those present in the three complete plus all nine draft Roseobacter
420 genomes using a reciprocal best hits criterion. Of the 14,287 different genes harbored
421 within the 12 Roseobacter genomes, 1,310 are core (Fig. 5). The core gene set includes
422 those for cellular metabolism, energy generation, and ribosomal protein synthesis, among
423 other critical processes (Table S7). The gene set does not include many of the
424 biogeochemically relevant pathways ascribed to roseobacters, such as carbon monoxide
425 oxidation (*cox* genes), sulfur oxidation (*sox* genes), DMSP demethylation (*dmdA*),
426 denitrification (*nirK, nirS, nor* genes*, nos* genes), and phosphonate utilzation (*phn*
427 genes). Thus while the total Roseobacter gene repertoire consists of a number of
428 ecologically relevant metabolic pathways, only a subset of these are present in any single
429 genome (Table 3). This "mix and match" genome arrangement may allow for adaptation
430 to a wide diversity of ecological niches, as has been observed for this group.

Which genes make a Roseobacter marine? Classic studies on the existence of distinctly marine bacteria in the 1960s concluded that metabolic pathways of cultured marine bacteria are quite similar to those of non-marine bacteria, and that transport systems based on sodium ions is the main difference between the two groups (37). We took a genomic approach to this question of "marine-biased" functions by identifying Roseobacter genes more likely to be found in genomes of other cultured marine taxa than in non-marine taxa. Using BLASTp, *S. pomeroyi* ORFs were compared against a set of 20 genome pairs each consisting of a cultured marine bacterioplankter outside the Roseobacter clade and its closest non-marine relative. Forty *S. pomeroyi* genes were identified as having a distribution statistically biased toward the marine bacteria (Table S8; Chi square $p \leq 0.05$). Three functional gene categories dominated this set: genes involved in sodium transport or utilizing sodium gradients, genes involved in sulfate transport and metabolism, and genes involved in osmoregulation. Sodium-related genes include two sodium symporters that co-transport sodium with other compounds (likely amino acids; SPO1810 and SPO2370), and a respiratory sodium pump (SPOA0030; a component of a multi-gene NADH:ubiquinone oxidoreductase) that establishes a sodium gradient for ATP synthesis. The sulfur-related genes include two sulfate permeases (SPO1956 and SPO3058) that transport sulfate across the cytoplasmic membrane, a predicted glyoxalase with carbon-sulfur lyase activity (SPO0721) and a methionine-γ-lyase (SPOA0318) potentially involved in metabolism of organic sulfur compounds, and an antioxidant that may be specific to sulfur-containing radicals (AhpC/Tsa family protein; SPO2103). Osmoregulation-related genes include osmotically inducible protein *osmC* (SPO2301) and a homolog of the glycine betaine synthesis gene *betA* (SPO0190)

(Table S8). The 44 genes in the *Silicibacter* sp. TM1040 genome and 36 in the *Jannaschia* sp. CCS1 genome also found significantly more often in other cultured marine than in non-marine genomes similarly have sodium-, sulfate-, and osmoregulation-related functions well represented (Table S8). In general, most of the Roseobacter genes on the marine-biased list are indeed those that permit the organisms to benefit from or manage the plentiful supply of ions in seawater, and few appear to mediate metabolic processes that are unique to or characteristic of marine bacteria. A recent culture-independent study of marine operons also found that genes involved in osmolyte transport systems and sodium-based respiratory pumps were significantly enriched in marine bacterial communities (54).

Marine bacteria in the SAR11 group are very abundant in ocean surface waters and commonly co-occur with Roseobacter cells (17). A comparison of the Roseobacter genomes to that of SAR11 member *P. ubique* HTCC1062 showed that the *P. ubique* genome contains 547 ORFs (out of 1,354 total) that are not shared with the Roseobacter genomes; over a quarter of these are for hypothetical proteins or proteins of unknown function. We looked in the *P. ubique* genome for selected biogeochemically relevant genes identified in Roseobacter genomes. Except for the presence of genes for DMSP catabolism (*dmdA*), there was no overlap, suggesting quite distinct ecological roles for these two abundant taxa (Table 3).

Hypothesis generation. Genome sequences of cultured marine bacterioplankton can serve to generate hypotheses about ecological activities and biogeochemical roles of their uncultured relatives. Aspects of the Roseobacter genomes suggest that they live in nutrient-replete plankton aggregates within the bulk oligotrophic ocean. Typical

components of eukaryotic cytosols can be transported by the roseobacters, such as polyamines, taurine, phosphoesters, phosphonates, aromatic metabolites, glyoxylate, allophanate, acetate, glycine betaine, DMSP, branched chain amino acids, and organic acids. Many of these compounds are produced by phytoplankton or zooplankton (23, 30, 34, 44), and may become available to roseobacters through exudation, diffusion-driven loss, or trophodynamic interactions including grazing and viral lysis. Regular observations of associations between roseobacters and marine phytoplankton in algal blooms (45, 49) and phycospheres of cultured dinoflagellates and diatoms (1, 25) are consistent with this idea. Members of the Roseobacter clade appear to exhibit a gradient in their reliance on "hot spots" in ocean waters, however. For example, *Jannaschia* sp. CCS1 differs from the other two Roseobacters in that its genome indicates it is phototrophic, transports and reduces nitrate, only assimilates inorganic forms of P, and has only one rRNA operon with which to gear up metabolic activity in response to transient nutrient availability (compared to three in *S. pomeroyi* and five in *Silicibacter* sp. TM1040) (31).

The repertoire of Roseobacter genes also generates the hypothesis that they frequently interact with neighboring cells, possibly to increase their own access to resources. Abundant toxin and metabolite (DMT) transporters, NRPS and possibly IAA synthesis pathways, extracellular polyhydroxybutyrate degradation, beta-lactamase related proteins, extracellular peptide signaling genes, and RND-MFP family proteins support this idea. Although speculative, it is possible that roseobacters directly capture organic matter from eukaryotic cells through Vir proteins, IAA production, or hemolysin-

499 type proteins. Such activity would influence the amount of carbon and nutrients entering
500 the microbial food web (2) and affect their availability to metazoans.

501 At least two successful ecological strategies for marine bacterioplankton thus
502 appear to operate in ocean surface waters. One tactic, typified by the roseobacters, is
503 based on metabolically versatile cells that can compete well with other organisms for
504 labile substrates within plankton-dense microzones. Another tactic, typified by SAR11
505 clade members, is based on metabolically conservative, free-living cells that scavenge
506 dilute organic matter dissolved in seawater (18). Roseobacter and SAR11 clade members
507 represent a significant fraction of marine bacterioplankton communities, together
508 accounting for ≥30% of the bacterioplankton in most ocean surface waters (17, 19, 20).
509 As more genomic information emerges for these and other key bacterioplankton clades,
510 including those yet to be cultured, insights into the biochemical capabilities of individual
511 cells will enrich understanding of bacterial ecology and ocean-scale biogeochemical
512 processes.
513

Acknowledgements

We thank M. Joye for unpublished data. This work was supported by grants from the National Science Foundation (MCB-0531476 to M.A.M., MCB-0446001 to R.B., and MCB-0534203 to A.B.) and the Gordon and Betty Moore Foundation. The Department of Energy provided genome sequences of *Silicibacter* sp. TM 1040 and *Jannaschia* sp. CCS1.

References

1. **Alavi, M., T. Miller, K. Erlandson, R. Schneide, and R. Belas.** 2001. Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. Environ. Microbiol. **3**:380-396.

2. **Azam, F.** 1998. Microbial control of oceanic carbon flux: The plot thickens. Science. **280**:694-696.

3. **Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, et al.** 2004. The Pfam protein families database. Nucleic Acids Res. **32**:D138-D141.

4. **Béjà, O., M. T. Suzuki, J. F. Heidelberg, W. C. Nelson, C. M. Preston, et al.** 2002. Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature. **415**:630-633.

5. **Buchan, A., L. S. Collier, E. L. Neidle, and M. A. Moran.** 2000. Key aromatic-ring-cleaving enzyme, protocatechuate 3,4-dioxygenase, in the ecologically important marine *Roseobacter* lineage. Appl. Environ. Microbiol. **66**:4662-4672.

6. **Buchan, A., E .L. Neidle, and M. A. Moran.** 2004. Diverse organization of genes of the ß-ketoadipate pathway in members of the marine *Roseobacter* lineage. Appl. Environ. Microbiol. **70**:1658-1668.

7. **Buchan, A., J. M. González, and M. A. Moran.** 2005. Overview of the marine Roseobacter lineage. Appl. Environ. Microbiol. **71**:5665-5677.

8. **Button, D. K.** 1991. Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the Michaelis Constant. Appl. Environ. Microbiol. **57**:2033-2038.

9. **Chen, F., K. Wang, J. Stewart, and R. Belas.** 2006. Induction of multiple prophages from a marine bacterium: a genomic approach. Appl. Environ. Microbiol. **72**:4995-5001.

10. **Christie, P. J., K. Atmakuri, V. Krishnamoorthy, S. Jakubowski, and E. Cascales.** 2005. Biogenesis, architecture, and function of bacterial Type IV secretion systems. Annu. Rev. Microbiol. **59**:451-485.

11. **DeLong, E. F.** 2005. Microbial community genomics in the ocean. Nat. Rev. Microbiol. **3**:459-469.

12. **Dolowy, P., J. Mondzelewski, R. Zawadzka, J. Baj, and D. Bartosik.** 2005. Cloning and characterization of a region responsible for the maintenance of megaplasmid pTAV3 of *Paracoccus versutus* UW1. Plasmid. **53**:239-250.

13. **Duval, B., K. Shetty, and W. H. J. Thomas.** 1999. Phenolic compounds and antioxidant properties in the snow alga *Chlamydomonas nivalis* after exposure to UV light. Appl. Phycol. **11**:559-566.

14. **Faulkner, D. J.** 2000. Marine natural products. Natural Product Rep. **17**:7-55.

15. **Field, J. G., K. R. Clarke, and R. M. Warwick.** 1982. A practical strategy for analysing multispecies distribution patterns. Mar. Ecol. Prog. Ser. 8:37-52.

16. **Gabriel, D. W., C. Allen, M. Schell, T. P. Denny, J. T. Greenberg, et al.** 2006. Identification of open reading frames unique to a select agent: *Ralstonia solanacearum* Race 3 Biovar 2. Mol. Plant-Microbe Inter. **19**:69-79.

17. **Giovannoni, S. J., and M .S. Rappé.** 2000. The uncultured microbial majority. pp. 47-84. In: D. L. Kirchman (ed), Microbial Ecology of the Oceans. John Wiley & Sons, New York, N.Y.

18. **Giovannoni, S. J., H. J. Tripp, S. Givan, M. Podar, K. L. Vergin, et al.** 2005. Genome streamlining in a cosmopolitan oceanic bacterium. Science. **309**:1242-1245.

19. **González, J. M., and M. A. Moran.** 1997. Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater. Appl. Environ. Microbiol. **63**:4237-4242.

20. **González, J. M., R. Simó, R. Massana, J. S. Covert, E. O. Casamayor, et al.** 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl. Environ. Microbiol. **66**:4237-4246.

21. **González, J. M., J. S. Covert, W. B. Whitman, J. Henricksen, F. Mayer, et al.** 2003. *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., DMSP demethylating bacteria from marine environments. Int. J. Syst. Evol. Microbiol. **53**:1261-1269.

22. **Han, C. S. and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher. pp. 141-146. In: H. R. Arabnia and H. Valafar (eds), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press.

23. **Hellebust, J. A.** 1965. Excretion of some organic compounds by marine phytoplankton. Limnol. Oceanogr. **10**:192-206.

24. **Howard, E. C., J. R. Henriksen, A. Buchan, C. R. Reisch, H. Bürgmann, et al.** 2006. Bacterial taxa limiting sulfur flux from the ocean. Science **314**:649-652.

25. **Jasti, S., M. E. Sieracki, N. J. Poulton, M. W. Giewat, and J. N. Rooney-Varga.** 2005. Phylogenetic diversity and specificity of bacteria closely associated with *Alexandrium* spp. and other phytoplankton. Appl. Environ. Microbiol. **71**:3483-3494.

26. **Jung, A., T. Domratcheva, M. Tarutina, Q. Wu, W. H. Ko, et al.** 2005. Structure of a bacterial BLUF photoreceptor: Insights into blue light-mediated signal transduction. Proc. Nat. Acad. Sci. **102**:12350-12355.

27. **Käll, L., A. Krogh, and E. L. L. Sonnhammer.** 2004. A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. **338**:1027-1036.

28. **Kanamori, T., N. Kanou, H. Atomi, and T. Imanaka.** 2004. Enzymatic characterization of a prokaryotic urea carboxylase. J. Bacteriol. **186**:2532-2539.

29. **Kelly, D. J., and G. H. Thomas.** 2001. The tripartite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. FEMS Microbiol. Rev. **25**:405-24.

30. **Kiene, R. P., L. J. Linn, and J. A. Bruton.** 2000. New and important roles for DMSP in marine microbial communities. J. Sea Res. **43**:209-224.

31. **Klappenbach, J. A., J. Dunbar, and T. M. Schmidt.** 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66:1328-1333.

32. **Kobayashi, M., H. Izui, T. Nagasawa, and H. Yamada.** 1993. Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole-3-acetonitrile: cloning of the *Alcaligenes* gene and site-directed mutagenesis of cysteine residues Proc. Natl. Acad. Sci. **90**:247-251.

33. **Kobayashi, M., T. Suzuki, T. Fujita, M. Masuda, and S. Shimizu.** 1995. Occurrence of enzymes involved in biosynthesis of indole-3-acetic acid from indole-3-acetonitrile in plant-associated bacteria, *Agrobacterium* and *Rhizobium*. Proc. Natl. Acad. Sci. USA. **92**:714-718.

34. **Lee, C., and N. O. G. Jorgensen.** 1995. Seasonal cycling of putrescine and amino-acids in relation to biological production in a stratified coastal salt pond. Biogeochem. **29**:131-157.

35. **Letunic, I., R. R. Copley, S. Schmidt, F. D. Ciccarelli, T. Doerks, et al.** 2004. SMART 4.0: towards genomic data integration. Nucleic Acids Res. **32**:D142-D144.

36. **Lindow, S. E., and M. T. Brandl.** 2003. Microbiology of the phyllosphere. Appl. Environ. Microbiol. **69**:1875-1883.

37. **MacLeod, R. A.** 1965. The question of the existence of specific marine bacteria. Bacteriological Rev. **29**:9-23.

38. **Malmstrom, R. R., R. P. Kiene, and D. L. Kirchman.** 2004. Identification and enumeration of bacteria assimilating dimethylsulfoniopropionate (DMSP) in the North Atlantic and Gulf of Mexico. Limnol. Oceanogr. **49**:597-606.

39. **Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, et al.** 2003. GenDB-an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. **31**:2187-2195.

40. **Miller, T. R., and R Belas.** 2004. Dimethylsulfoniopropionate Metabolism by *Pfiesteria*-Associated *Roseobacter* spp. Appl. Environ. Microbiol. **70**:3383-3391.

41. **Miller, T. R., and R. Belas.** 2006. Motility is involved in *Silicibacter* sp. TM1040 interaction with dinoflagellates. Appl. Environ. Microbiol. **8**:1648-1659.

42. **Moran, M. A., R. E. Hodson.** 1994. Dissolved humic substances of vascular plant origin in a coastal marine environment. Limnol. Oceanogr. **39**:762-771.

43. **Moran, M. A., A. Buchan, J. M. González, J. F. Heidelberg, W. B. Whitman, et al.** 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature. **432**:910-913.

44. **Nishibori, N., A. Yuasa, M. Sakai, S. Fujihara, and S. Nishio.** 2001. Free polyamine concentrations in coastal seawater during phytoplankton bloom. Fish. Sci. **67**:79-83.

45. **Pinhassi J., R. Simó, J. M. González, M. Vila, L. Alonso-Sáez, et al.** 2005. Dimethylsulfoniopropionate turnover is linked to the composition and dynamics of the bacterioplankton assemblage during a microcosm phytoplankton bloom. Appl. Environ. Microbiol. **71**:7650-7660.

46. **Pradella, S., M. Allgaier, C. Hoch, O. Päuker, E. Stackebrandt, et al.** 2004. Genome organization and localization of the *pufLM* genes of the photosynthesis reaction center in phylogenetically diverse marine *Alphaproteobacteria.* Appl. Environ. Microbiol. **70**:3360-3369.

47. **Quest, B., and W Gärtner.** 2004. Chromophore selectivity in bacterial phytochromes. Eur. J. Biochem. **271**:1117-1126.

48. **Ren, Q., and I. T. Paulsen.** 2005. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. PLoS Comput. Biol. **1(3)**:e27.

49. **Riemann, L., G. F. Steward, and F. Azam.** 2000. Dynamics of bacterial community composition and activity during a mesocosm diatom bloom. Appl. Environ. Microbiol. **66**:578-587.

50. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, et al.** 2000. Artemis: sequence visualization and annotation. Bioinformatics. **16**:944-945.

51. **Santos, S. R., and H. Ochman.** 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. Environ. Microbiol. **6**:754-759.

52. **Swingley, W. D., S. Sadekar, S. D. Mastrian, H. J. Matthies, J. Hao, et al.** 2007. The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. J. Bacteriol. 189:683-90.

53. **Swofford, D. L.** 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods) 4.0.

54. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, et al.** 2005. Comparative metagenomics of microbial communities. Science **308**:554-557.

55. **Ulrich, L. E., E. V. Koonin, and I. B. Zhulin.** 2005. One-component systems dominate signal transduction in prokaryotes. Trends Microbiol. **13**:52-56.

56. **Wagner-Döbler, I., and H. Biebl.** 2006. Environmental biology of the marine *Roseobacter* lineage. Ann. Rev. Microbiol. 60: 255-280.

57. **Zhong, Z., R. Caspi, D. Helinski, V. Knauf, S. Sykes, et al.** 2003. Nucleotide sequence based characterizations of two cryptic plasmids from the marine bacterium *Ruegeria* isolate PR1b. Plasmid. **49**:233-252.

Supporting Information Files

Tables S1-S8.

Figure S1-S6.

Table 1. Features of three Roseobacter genomes. n/a = not applicable.

Feature	*S. pomeroyi* DSS-3	*Silicibacter* sp. TM1040	*Jannaschia* sp. CCS1
Total number of coding sequences	4283	3863	4283
G+C Content (%)	64.0	60.0	62.2
Number of rRNA operons	3	5	1
Number of replicons	2	3	2
Molecule length by replicon (bp)			
Main chromosome	4,109,442	3,201,640	4,317,977
Megaplasmid	491,611	823,032	n/a
Plasmid	n/a	131,885	86,072
Coding sequences by replicon			
Main chromosome	3,838	3013	4212
Megaplasmid	445	747	n/a
Plasmid	n/a	103	71
G+C Content by replicon			
Main chromosome	64.2	60.4	62.3
Megaplasmid	62.8	59.3	n/a
Plasmid	n/a	55.3	57.8
rRNA operons by replicon			
Main chromosome	3	1	1
Megaplasmid	0	3	n/a
Plasmid	n/a	1	0

Table 2. Numbers (and percent) shared and unique genes in three Roseobacter genomes.

3-way shared genes are those with orthologs in all three genomes.

Method	3-way Shared	Unique to *S. pomeroyi* DSS-3	Unique to *Silicibacter* sp. TM1040	Unique to *Jannaschia* sp. CCS1
RBH Analysis	1939*	1336 (31%)	1027 (27%)	1685 (39%)
ISGS Analysis	121†	1260 (29%)	983 (25%)	1600 (37%)

*RBH analysis used the criterion of reciprocal best hits between pairs of genomes with a corresponding BLAST Expect value of $<10^{-5}$ and amino acid identity of >30%.

†ISGS (In Silico Genomic Subtraction) analysis used one-way BLASTp and tBLASTn analysis to identify genes shared by the three Roseobacters but missing from four non-marine α-Proteobacteria relatives (*R. sphaeroides*, *P. denitrificans*, *S. meliloti*, and *M. loti*).

Table 3. Survey of biogeochemically relevant genes in three complete (*S. pomeroyi* DSS-3, *Silicibacter* sp. TM1040, *Jannaschia* sp. CCS1) and nine draft Roseobacter genomes, including those for aerobic anoxygenic phototrophy (AAnP genes), carbon monoxide oxidation (*cox* genes), aromatic compound degradation (*pcaG/H* and *boxA*), sulfur oxidation (*soxB*), denitrification (*nirS*/*K*), nitrate assimilation (*nasA*), phosphonate use (*phn* genes), Type IV secretion (*vir* genes), and DMSP demethylase (*dmdA)*. A '+' indicates the presence of a homolog with an E value $\leq 10^{-40}$ and amino acid percent sequence identity ≥40%. Phenotypes that have been demonstrated experimentally for a Roseobacter strain are indicated by a 'yes'. n.d=not determined. The presence of an ortholog in the genome of SAR11 member *P. ubique* is indicated.

Genome	AAnP genes	*cox* genes	*pcaG H*	*boxC*	*soxB*	*nirS/K*	*nasA*	*phn* genes	*vir* genes	*dmdA*
Silicibacter pomeroyi DSS-3		+	+	+	+	+		+		+
Silicibacter sp. TM1040		+	+					+	+	+
Jannaschia sp. CCS1	+	+	+	+			+			+
Sulfitobacter sp. EE-36		+	+		+		+	+		
Sulfitobacter sp. NAS-14.1		+	+		+		+	+	+	
Roseovarius nubinhibens ISM			+		+			+	+	+
Roseovarius sp. 217	+	+	+		+	+	+	+	+	+
Oceanicola batsensis HTCC2597		+	+					+	+	
Oceanicola granulosus HTCC2516		+								
Rhodobacterales bacterium HTCC2654		+		+	+	+	+	+	+	
Loktanella vestfoldensis SKA53	+	+	+					+		
Roseobacter sp. MED193		+	+		+		+	+	+	+
Percent of Roseobacter genomes	25%	92%	83%	25%	58%	25%	50%	92%	50%	50%
Experimental evidence	n.d.	yes†	yes§	n.d.	yes∫	yes*	yes¥	yes¶	n.d.	yes£
Ortholog in *P. ubique* HTCC1062	no	no	no	no	no	no	no	no	no	yes

†=ref. 43; §=ref. 5; ∫=ref. 21; *=Figure S7; ¥=Figure S4; ¶=Figure S5; £=ref. 24.

Figure Captions

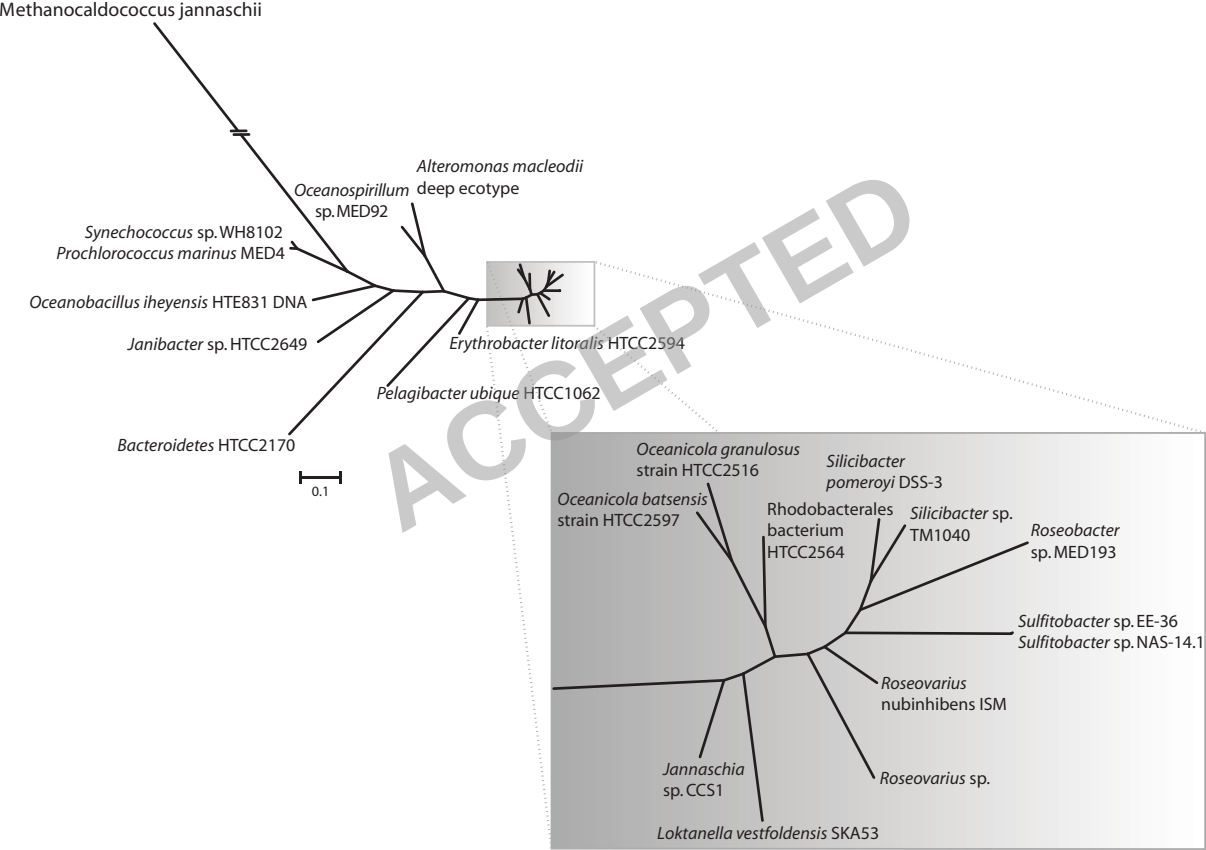
Fig. 1. Phylogenetic tree of 16S rRNA gene sequences from 12 roseobacters and other selected marine bacterioplankton for which a genome sequence is available. The tree is based on positions 21 to 1490 of the 16S rDNA (*E. coli* numbering system). The tree was constructed with the PAUP* package (53), version 4.10b, using the maximum likelihood method. The bar corresponds to the number of changes per nucleotide for the main tree.

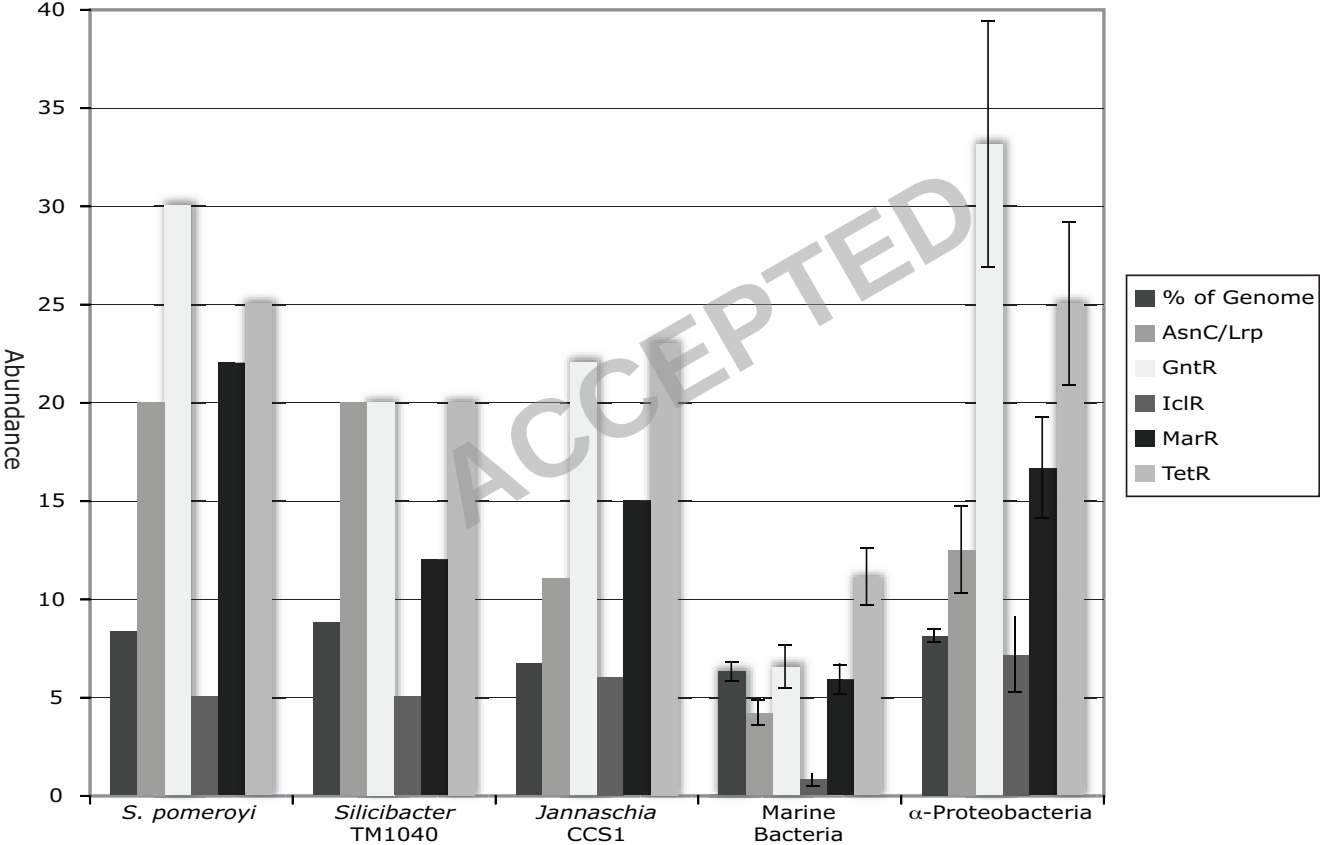
Fig. 2. Signal transduction proteins in three Roseobacter genomes and two comparison groups consisting of marine bacterioplankton or closely-related non-marine α-Proteobacteria. Abundance of genes encoding signal transduction proteins are shown as percent of total coding sequences (% of genome) or as numbers of the one-component systems AsnC/Lrp (likely involved in amino acid metabolism), GntR (repression of gluconate utilization), IclR (repression of the acetate operon), MarR (repression of antibiotic resistance or stress response), and TetR (repression of tetracycline resistance). The marine comparison group consists of 37 genomes with an average size of 4.0 Mb, while the non-marine α-Proteobacteria comparison group consists of 8 genomes with an average size of 5.5 Mb).

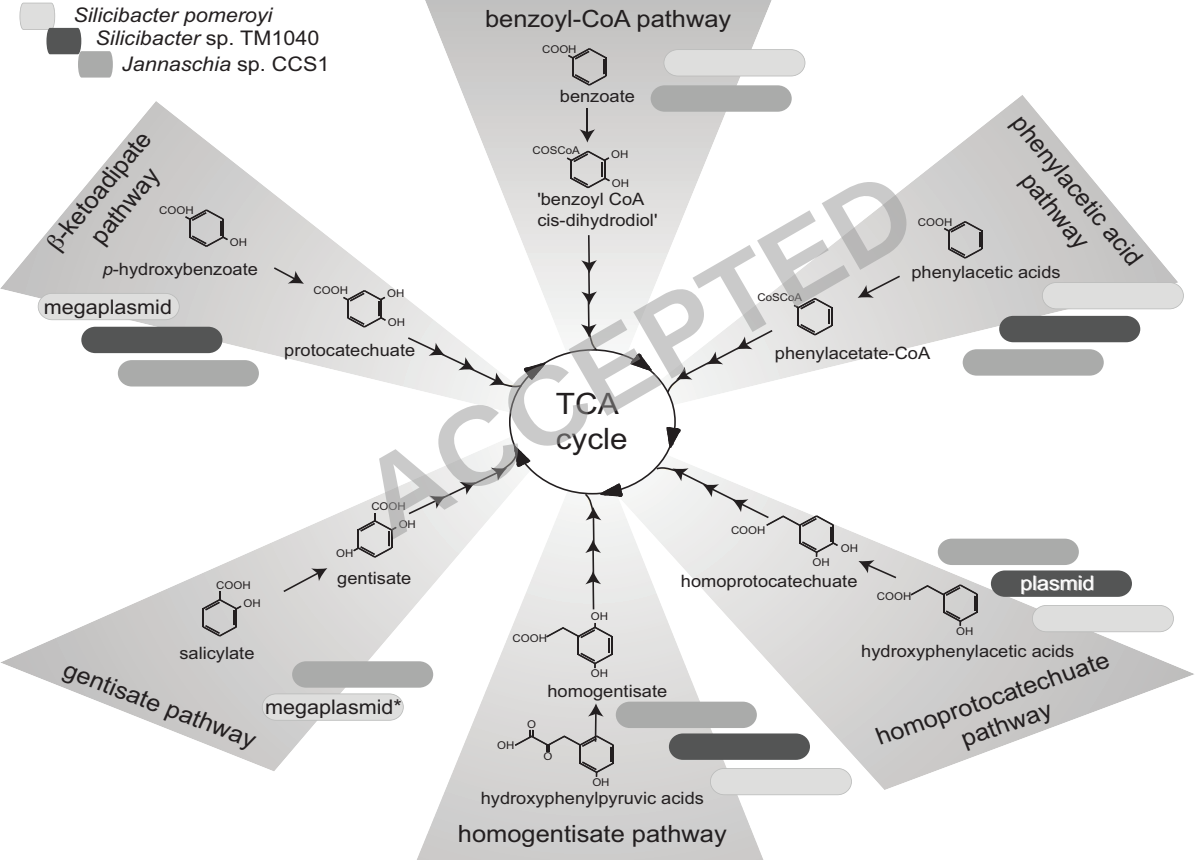
Fig. 3. Six distinct ring-cleaving pathways present in three Roseobacter genomes.
Shaded tabs indicate the presence of a pathway (light gray = *S. pomeroyi*, dark gray = *Silicibacter* sp. TM1040, medium gray = *Jannaschia* sp. CCS1). The gene(s) encoding the ring cleaving enzymes are chromosomally located unless otherwise indicated within the tab. **S. pomeroyi* contains two copies of the gentisate pathway, one on the chromosome and one on megaplasmid pSPD. The arrowheads leading into the TCA cycle indicate multiple steps.

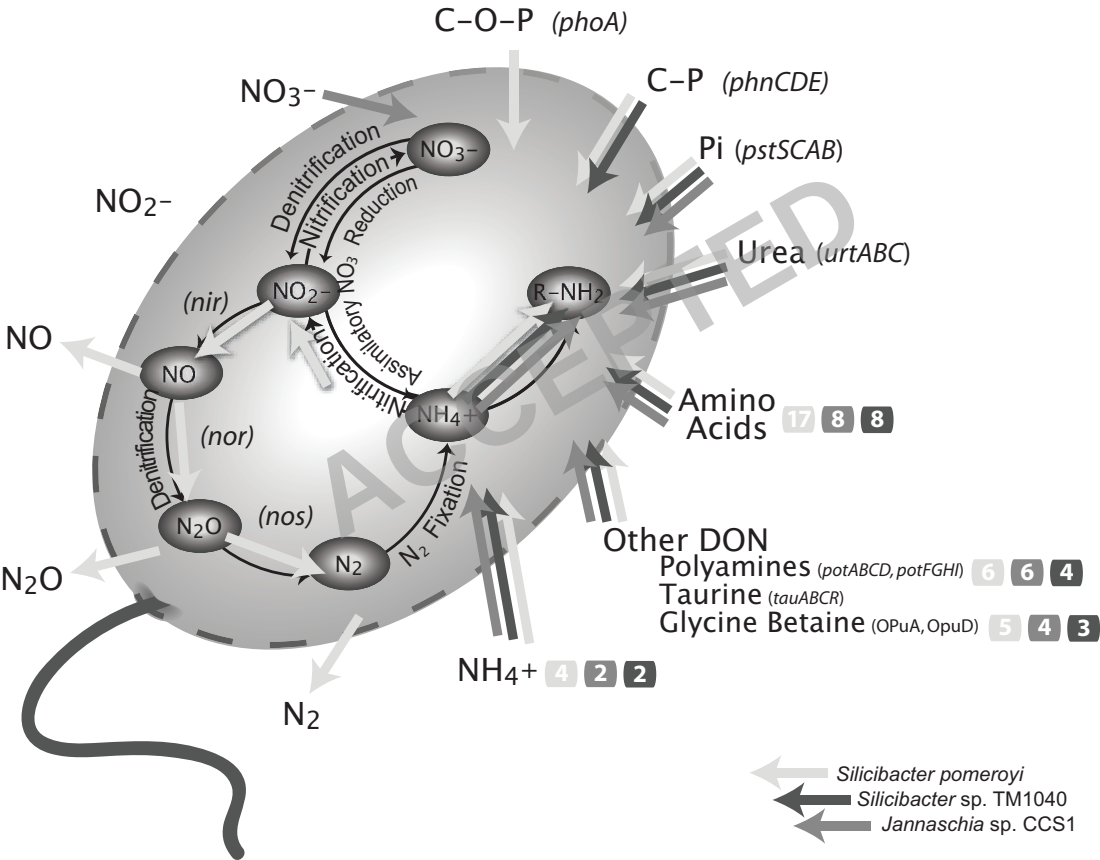
Fig. 4. Genes for acquisition of nitrogen and phosphorus in three Roseobacter genomes. Shaded arrows indicate the presence of a pathway (light gray = *S. pomeroyi*, dark gray = *Silicibacter* sp. TM1040, medium gray = *Jannaschia* sp. CCS1). Numbers in colored circles indicate the number of ORFs if $\geq$2 copies. Amino acid transporters and branched-chain amino acid transporters are summed. C-O-P = phosphoesters; C-P = phosphonates; Pi = phosphate.

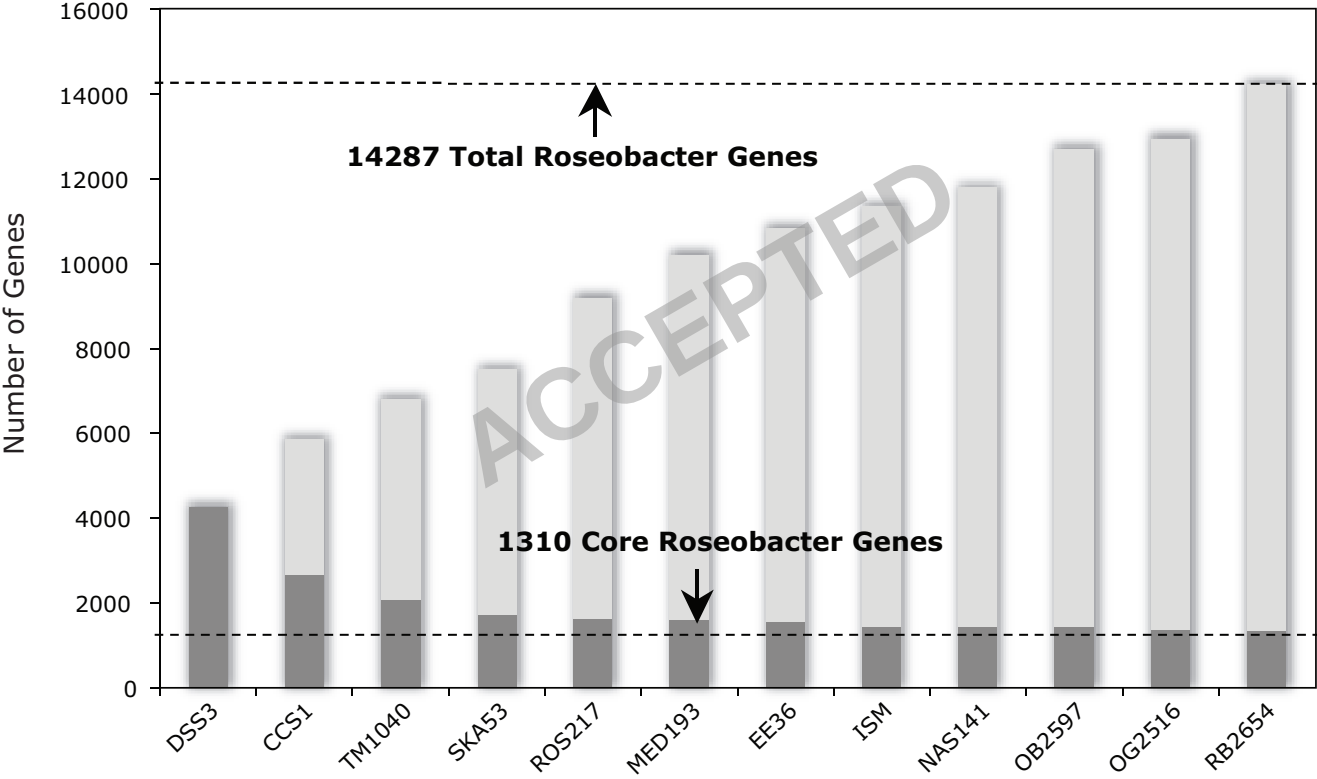
Fig. 5. Roseobacter core genes (dark) and total genes (dark plus light) as a function of increasing genome number. Core genes are those with a reciprocal best hit to the *S. pomeroyi* DSS-3 genome for the indicated genome plus all preceding genomes. Total genes are distinct genes in the Roseobacter group, incremented for each genome by the number of genes without a reciprocal best hit in any of the preceding genomes.


Methanocaldococcus jannaschii
Alteromonas macleodii deep ecotype
Oceanospirillum sp. MED92
Synechococcus sp. WH8102
Prochlorococcus marinus MED4
Oceanobacillus iheyensis HTE831 DNA
Janibacter sp. HTCC2649
Erythrobacter litoralis HTCC2594
Pelagibacter ubique HTCC1062
Bacteroidetes HTCC2170
0.1
Oceanicola granulosus strain HTCC2516
Silicibacter pomeroyi DSS-3
Oceanicola batsensis strain HTCC2597
Rhodobacterales bacterium HTCC2564
Silicibacter sp. TM1040
Roseobacter sp. MED193
Sulfitobacter sp. EE-36
Sulfitobacter sp. NAS-14.1
Roseovarius nubinhibens ISM
Jannaschia sp. CCS1
Roseovarius sp.
Loktanella vestfoldensis SKA53


40
35
30
25
20
15
10
5
0
Abundance
% of Genome
AsnC/Lrp
GntR
IclR
MarR
TetR
S. pomeroyi
Silicibacter TM1040
Jannaschia CCS1
Marine Bacteria
α-Proteobacteria
ACCEPTED


Silicibacter pomeroyi
Silicibacter sp. TM1040
Jannaschia sp. CCS1
benzoyl-CoA pathway
benzoate
'benzoyl CoA cis-dihydrodiol'
β-ketoadipate pathway
p-hydroxybenzoate
megaplasmid
protocatechuate
phenylacetic acid pathway
phenylacetic acids
phenylacetate-CoA
TCA cycle
gentisate
salicylate
gentisate pathway
megaplasmid*
homogentisate
hydroxyphenylpyruvic acids
homogentisate pathway
homoprotocatechuate
plasmid
hydroxyphenylacetic acids
homoprotocatechuate pathway
ACCEPTED


C–O–P (phoA)
C–P (phnCDE)
Pi (pstSCAB)
Urea (urtABC)
Amino Acids 17 8 8
Other DON
Polyamines (potABCD, potFGHI) 6 6 4
Taurine (tauABCR)
Glycine Betaine (OPuA, OpuD) 5 4 3
NH4+ 4 2 2
N2
N2O
NO
NO2–
NO3–
Denitrification
Nitrification
Reduction
(nir)
(nor)
(nos)
Assimilatory
Nitrification
N2 Fixation
R–NH2
NH4+
ACCEPTED
Silicibacter pomeroyi
Silicibacter sp. TM1040
Jannaschia sp. CCS1


16000
14000
12000
10000
8000
6000
4000
2000
0
Number of Genes
14287 Total Roseobacter Genes
1310 Core Roseobacter Genes
DSS3
CCS1
TM1040
SKA53
ROS217
MED193
EE36
ISM
NAS141
OB2597
OG2516
RB2654