Supporting Information

González *et al.* 10.1073/pnas.0712027105

SI Methods

Transporters. The low number of transporters was apparent in most of the main transporters families. Thus, together with *Pelagibacter ubique*, MED152 has the lowest number of ATP-dependent transporters [a majority belonging to the ATP-binding Cassette (ABC) superfamily] of the marine bacteria sequenced so far (Table 1). The same holds true for the secondary transporters, which account for 50–70% of the transporters in the marine bacterial genomes sequenced until now. MED152 has only a total of 62 secondary transporters – similar to the number found in *Pelagibacter ubique*– while other bacteria may have from 100–200. In particular, the number of transporters in the drug/metabolite transporter (DMT) superfamily and in the major facilitator superfamily [MFS; used for transport of a diverse set of "small solutes in response to chemiosmotic ion gradients" (1)] was low, while the number of multidrug/oligosaccharidyl-lipid/polysaccharide (MOP) flippase superfamily transporters was similar to that in other marine bacteria (despite its smaller genome).

Like *Pelagibacter ubique*, MED152 (and other marine flavobacteria) lacks phosphotransferase system (PTS) genes for active transport of sugars. PTS genes are present in all major bacterial groups except cyanobacteria and epsilonproteobacteria (2). The lack thereof is consistent with the aerobic lifestyle of MED152 (PTS are particularly important for sugar uptake under anaerobic conditions).

Comparative Analysis of Transporters. Transporter families in MED152 were compared to the following marine genomes in the TransportDB database (details on transporter families and substrate specificity of particular transporters in specific bacteria are available at www.membranetransport.org): alphaproteobacteria *Bradyrhizobium japonicum* USDA 110, "*Candidatus* Pelagibacter ubique" HTCC1062, *Caulobacter crescentus* CB15, *Hyphomonas neptunium* ATCC 15444, *Jannaschia* sp. CCS1, *Loktanella vestfoldensis* SKA53, *Maricaulis maris* MCS10, *Roseobacter* sp. MED193, *Roseobacter* sp. TM1040, *Roseovarius nubinhibens* ISM, and *Silicibacter pomeroyi* DSS-3; gammaproteobacteria *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, *E. coli* K12-MG1655, *Marinomonas* sp. MED121, *Photobacterium profundum* SS9, *Pseudoalteromonas haloplanktis* TAC125, *Pseudomonas aeruginosa* PAO1, *Shewanella oneidensis* MR-1, and *Vibrio* sp. MED222; Bacteroidetes/Chlorobi phylum *Bacteroides fragilis* NCTC9343, *Bacteroides fragilis* YCH46, *Bacteroides thetaiotaomicron* VPI-5482, *Chlorobium tepidum* TLS, *Dokdonia* sp. MED134, *Leeuwenhoekiella blandensis* MED217, and *Polaribacter* sp. MED152; and Firmicutes *Bacillus subtilis* 168.

Bicarbonate Uptake. For the bicarbonate uptake experiment, MED152 was grown at 16°C in Marine Broth (Difco) diluted 1:8 with artificial seawater (35 practical salinity units, prepared from Sea Salts; Sigma) in light (180 μmol photons m^{-2} s^{-1}) or dark conditions (duplicate 200 ml cultures in each condition). Cultures were maintained without shaking to reduce aggregation and flock formation. After 50 h incubation (during exponential growth), four subsamples from one culture of each incubation condition (light and dark) were transferred to 25 ml glass bottles. From each original incubation condition, two subsamples (treated as duplicates) were placed in transparent bottles and two subsamples were placed in dark bottles (bottles covered with aluminum foil and black plastic). All bottles were then incubated under white light (180 μmol photons m^{-2} s^{-1}; no light entered the dark bottles) for 2 h with 20 μl of $H^{14}CO_3^-$ (3 μCi; DHI). Controls from each treatment were treated with 10% trichloroacetic acid (final concentration). After incubation, 3 ml aliquots from each replicate bottle were filtered through 0.2 μm pore size filters (25 mm diameter, Supor-200, Pall), and the filters were exposed to HCl 0.7 M fumes for 2 h. Finally, the filters were placed in vials with 10 ml scintillation mixture (Perkin–Elmer) and kept at least 24 h in the dark before counting. Bicarbonate uptake rates were calculated according to the standard radioactive carbon assimilation technique procedure (3). At the time of the experiment, bacterial abundance was $\approx 5.1 \times 10^8$ cells ml^{-1}, as determined by epifluorescence microscopy of SYBR Gold stained cells.

Pigment Analysis. MED152 was grown in Marine Broth 2216 (Difco) and filtered onto Whatman GF/F filters. Pigments were extracted by placing the filters in 3 ml of 90% acetone (with 0.01% of butylated hydroxytoluene to prevent chlorophyll allomerization) and vortexing them vigorously for 45 s. After 24 h at -20°C, samples were sonicated for 1 min and vortexed again for 45 s. The extracts were cleared by filtration through Poretics 0.8 μm polycarbonate filters. For pigment chromatography, 150 μl of a mixture of 0.5 ml extract plus 0.1 ml H_2O was injected into a Thermo HPLC system and run under the conditions described in ref. 4. Standards of myxoxanthophyll, zeaxanthin and β-carotene (DHI) were used for pigment identification and quantification.

1. Pao SS, Paulsen IT, Saier MH (1998) Major facilitator superfamily. *Microbiol Mol Biol Rev* 62:1–34.
2. Barabote RD, Saier MH (2005) Comparative genomic analyses of the bacterial phosphotransferase system. *Microbiol Mol Biol Rev* 69:608–634.
3. Parsons TR, Maita Y, Lalli CM (1984) *A Manual of Chemical and Biological Methods for Seawater Analysis* (Pergamon Press, Oxford, UK).
4. Latasa M, *et al.* (2001) Losses of chlorophylls and carotenoids in aqueous acetone and methanol extracts prepared for RPHPLC analysis of pigments. *Chromatographia* 53:385–391.
5. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
6. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
7. Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164–166.

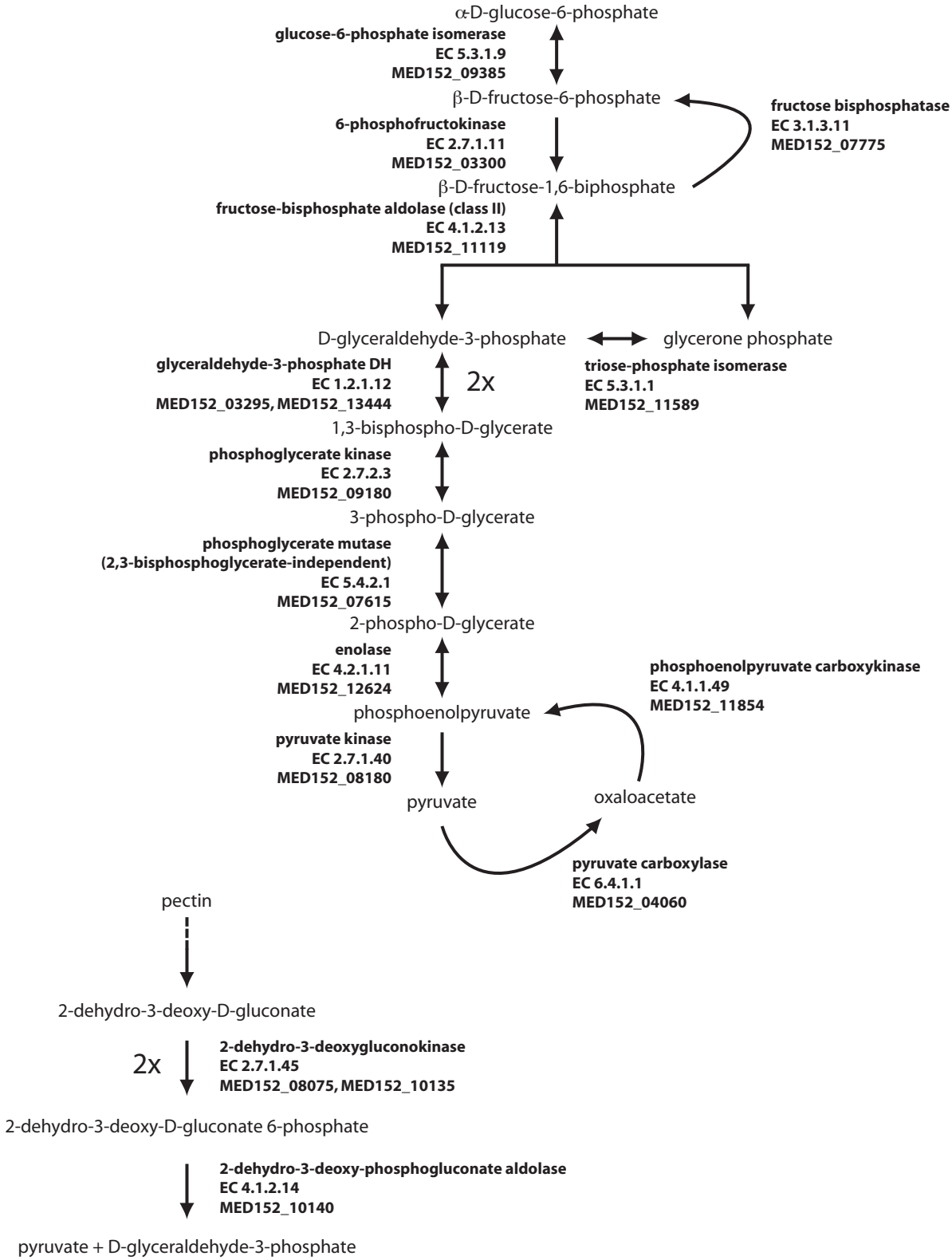

Fig. S1. Glycolysis/gluconeogenesis.

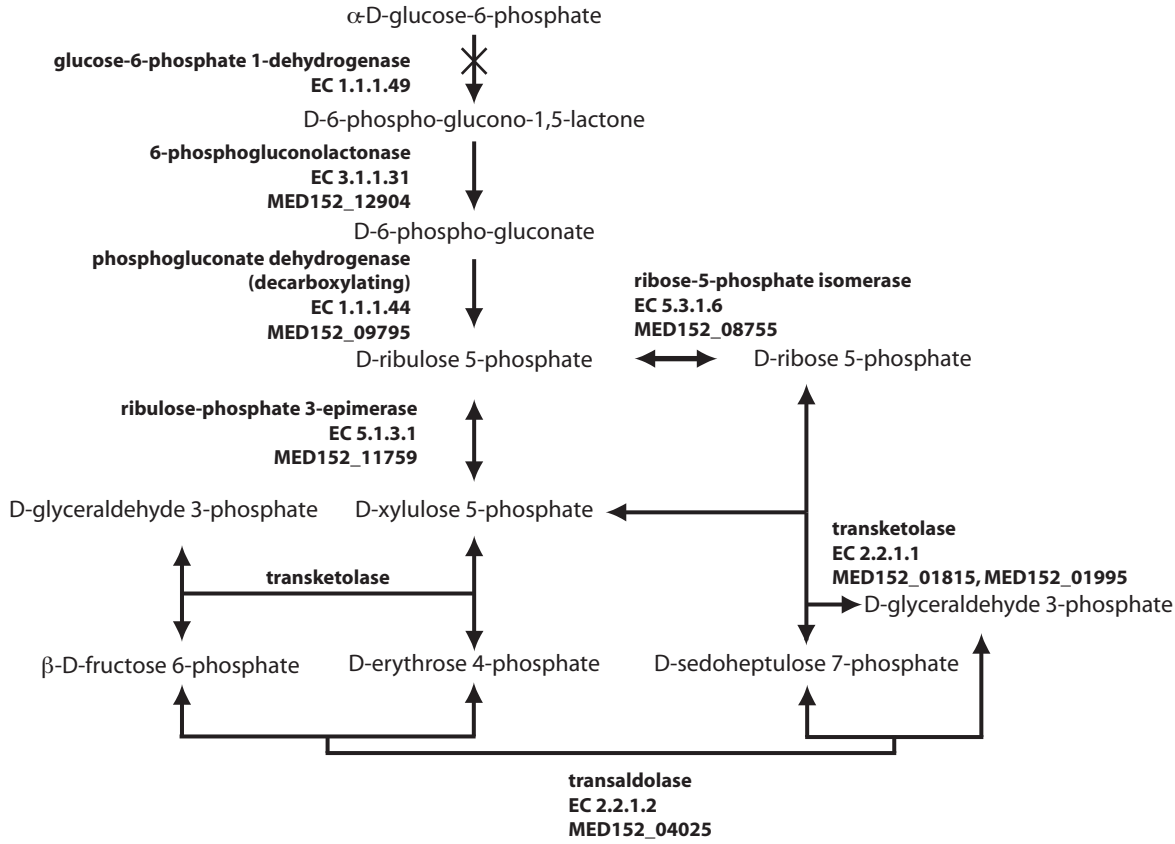

Fig. S2. Pentose phosphate pathway.

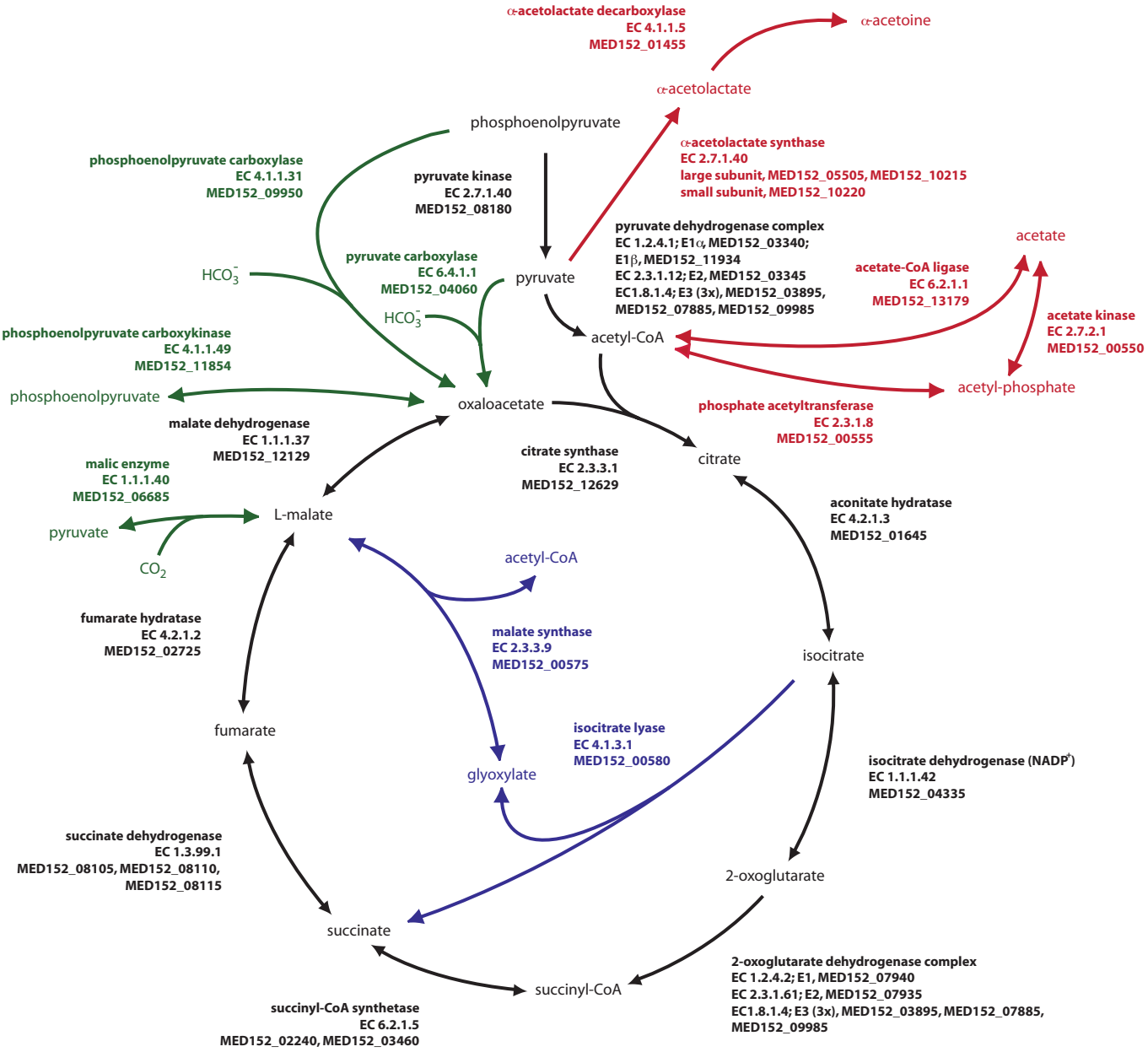

Fig. S3. TCA cycle and glyoxylate shunt.

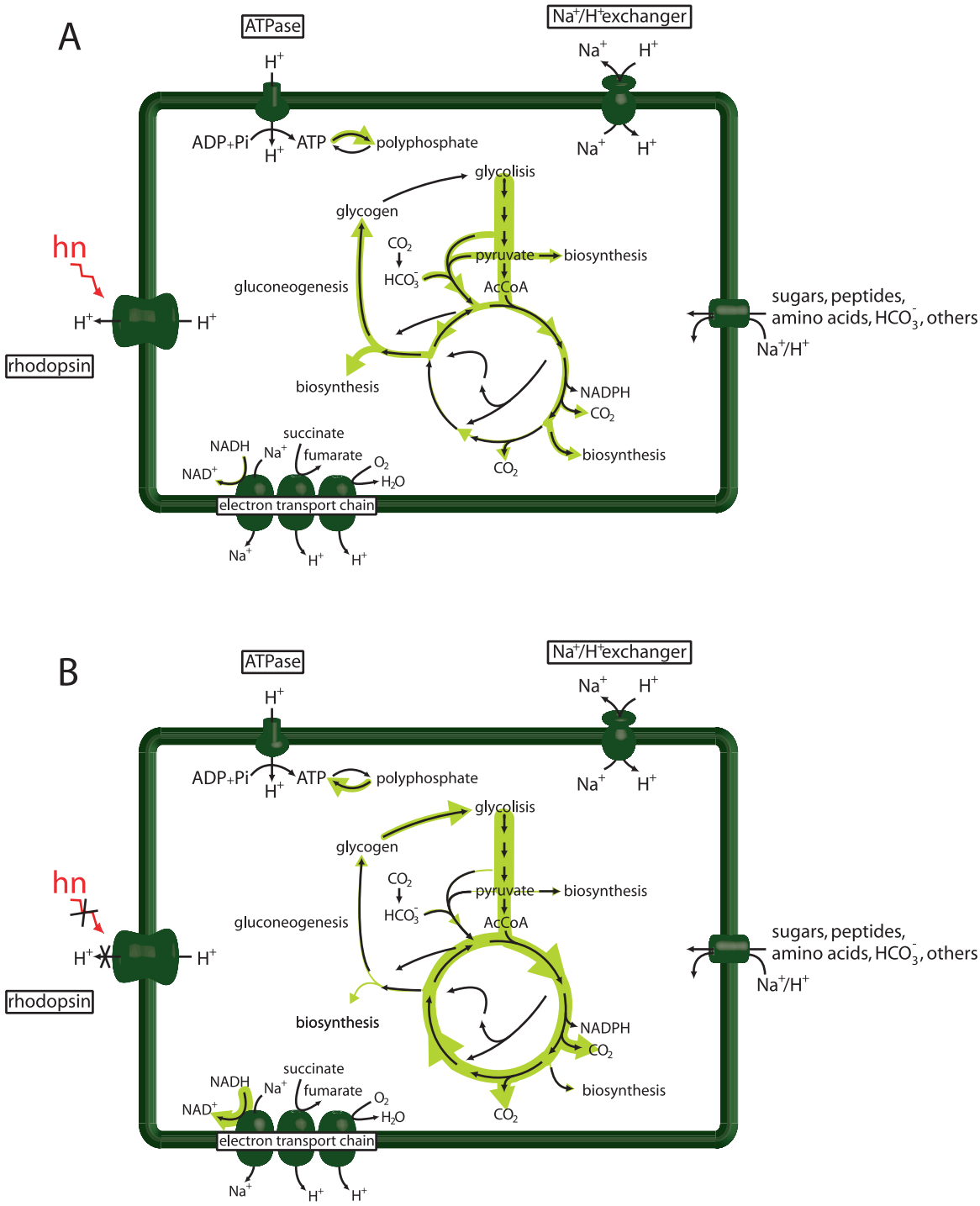

Fig. S4. Proposed carbon flow in MED152 in the presence and absence of light. When the cell is exposed to light (*A*), the prevailing role of the main metabolic pathways is biosynthetic. In the dark (*B*), the cell must respire more organic carbon than in the light to meet its energy requirements.

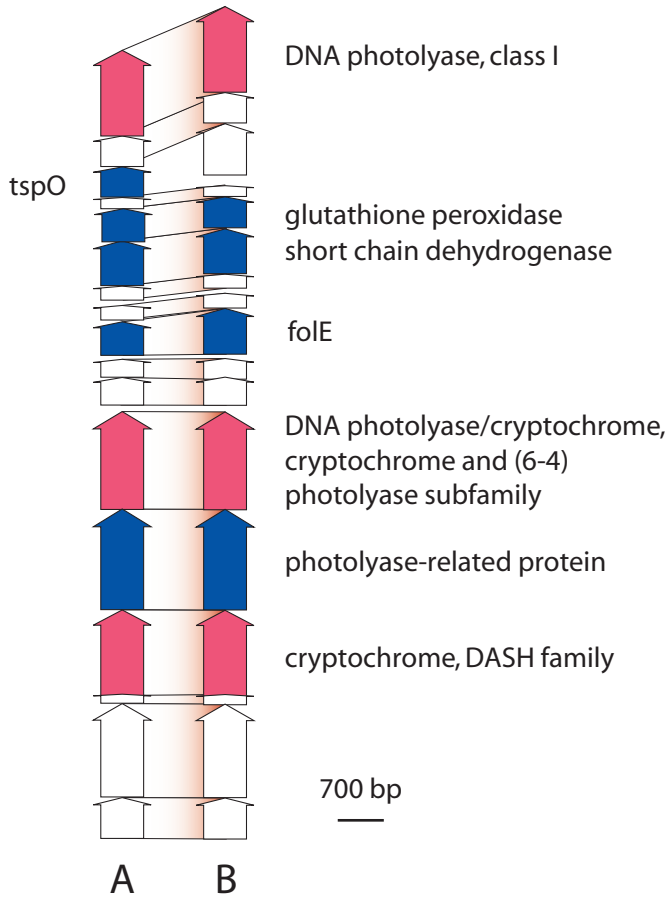

Fig. S5. *Polaribacter irgensii* 23-P (*A*) and *Polaribacter* sp. MED152 (*B*) gene clusters that contain the three copies of the DNA photolyase/cryptochrome genes (in red). Genes in white are conserved hypothetical. TspO is a Trp-rich sensory protein that acts as a transcriptional regulator and is only found in the genome of *Polaribacter irgensii* 23-P but not in MED152.

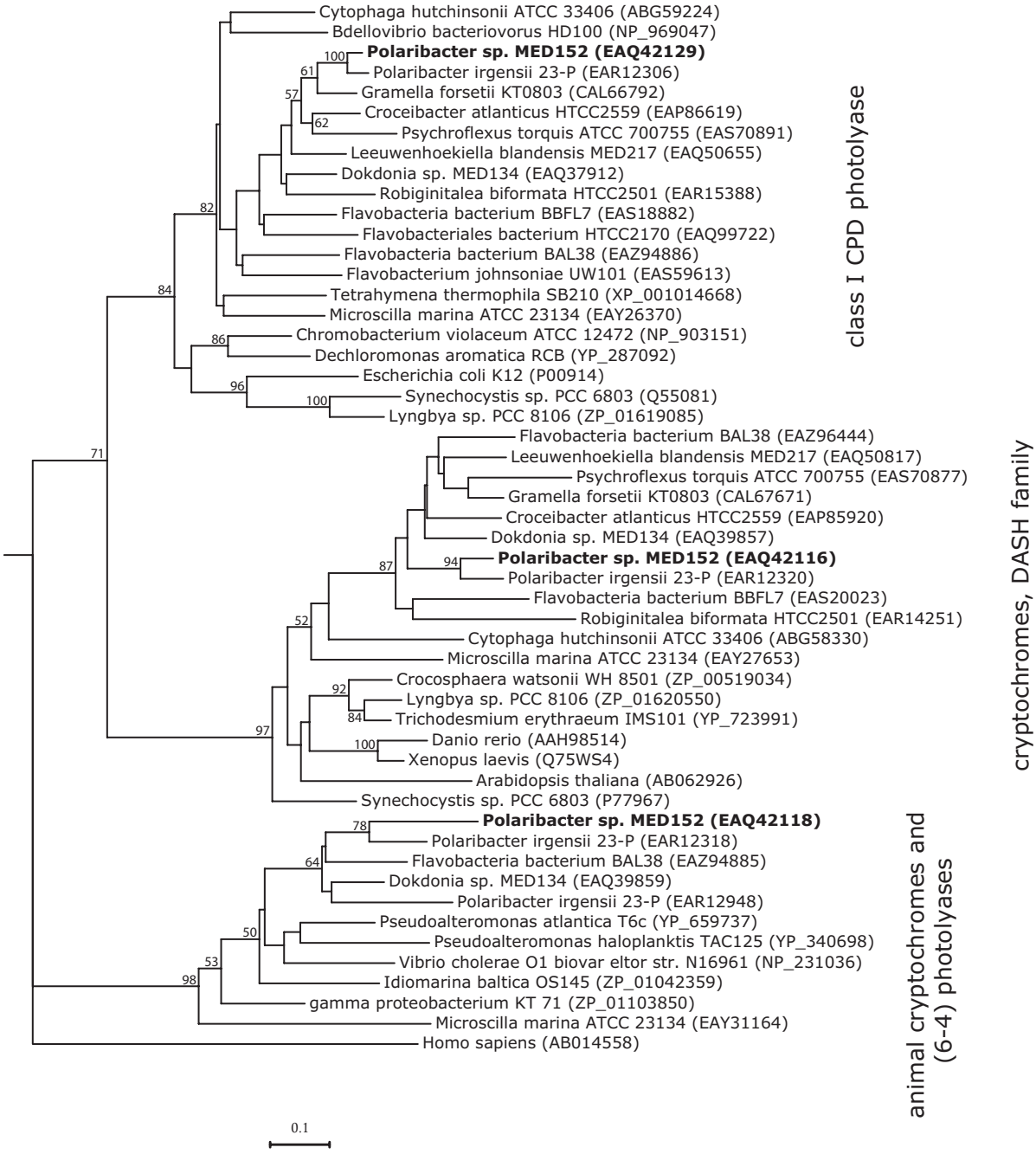

Fig. S6. Evolutionary relationships of cryptochrome/photolyase protein family from the marine Bacteroidetes and representatives from other organisms. Subfamilies are indicated on the right. A multiple alignment was constructed with the software package CLUSTAL W 1.74 (5). The alignment was edited with Gblocks (Version 0.91b) to identify conserved regions (6) with a minimum block of ten and without gaps. The tree was constructed based on a Kimura's distance matrix and the Neighbor-Joining method using the PHYLIP package (Version 3.2) (7). The statistical significance of the tree topology was evaluated by bootstrap analysis with 1000 iterative constructions of the neighbor-joining tree. The numbers at the nodes are bootstrap values higher than 50%. The scale bar represents the Kimura distance.

Fig. S7. Clusters of genes putatively involved in the attachment and degradation of polymeric compounds. Only those that contain the tandem TonB dependent/ligand-gated channel genes and *susD* homologs are represented.

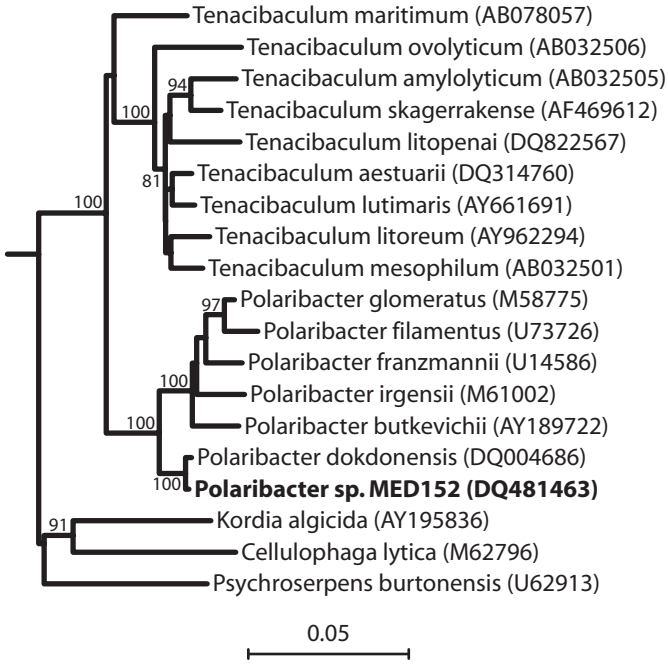

Fig. S8. Neighbor-joining phylogenetic tree based on 16S rRNA gene sequences of MED152 and closely related type strains for each species. A multiple alignment was constructed with the software package CLUSTAL W (Version 1.83; ref. 5). The alignment was edited with Gblocks (Version 0.91b) to identify conserved regions (6) with a minimum block of 10 and without gaps. The tree was constructed based on a Jukes-Cantor distance matrix and the Neighbor-Joining method using the PHYLIP package (Version 3.2) (7). The sequence of *Cytophaga hutchinsonii* (M58768) served as the outgroup. Bootstrap values greater than 70% confidence are shown at branching points (percentage of 1,000 resamplings). GenBank accession numbers are given in parentheses. The scale bar represents Jukes-Cantor distance.

Table S1. Summary of MED152 genome

Parameter	Value
No. of bp	2,967,150
GC content, %	30.61
Coding density, %	92.6
No. of predicted protein-coding genes	2,646
No. of predicted proteins with putative function	1,750
No. of conserved hypothetical proteins	593
No. of unknown proteins unique to MED152	303
No. of tRNAs	37
No. of *rrn*	3

Table S2. Key enzymes and metabolic pathways identified in MED152

Pathway/enzyme (gene copy when greater than one)	ORFs
Main pathways	
Pyruvate carboxylase. Anaplerotic pathway	MED152_04060
PEP carboxylase. Anaplerotic pathway	MED152_09950
Glyoxylate bypass. Anaplerotic pathway	MED152_00580 (*aceA*), MED152_00575 (*aceB*)
Malic enzyme. Anaplerotic pathway	MED152_06685
Phosphate acetyltransferase. Acetate metabolism	MED152_00555
Acetate kinase. Acetate metabolism	MED152_00550
Acetate CoA ligase. Acetate metabolism	MED152_13179
Acyl-CoA dehydrogenase. β-Oxidation (8)	MED152_01705, MED152_04870, MED152_05285, MED152_05295, MED152_05785, MED152_07525, MED152_07535, MED152_10560
Enoyl-CoA hydratase. β-Oxidation (4)	MED152_05285, MED152_06790, MED152_09440, MED152_11649
Electron transfer protein. β-Oxidation	MED152_11949 (*etfA*), MED152_11944 (*etfB*)
Glycerol kinase. Lipid degradation	MED152_08680
Glycerol-3-phosphate dehydrogenase ($NAD(P)^+$). Lipid degradation (2)	MED152_05430, MED152_08715
FAD-linked glycerol-3-phosphate dehydrogenase. Lipid degradation	MED152_08690
Glycogen biosynthesis	MED152_05855 (*glgA*), MED152_05860 (*glgC*), MED152_05865, MED152_05870, MED152_05875 (*glgB*), MED152_05880, MED152_05885
Electron transport chain	
Na^+-translocating NADH:quinone oxidoreductase	MED152_11704 (*nqrA*), MED152_11709 (*nqrB*), MED152_11714 (*nqrC*), MED152_11719 (*nqrD*), MED152_11724 (*nqrE*), MED152_11729 (*nqrF*)
Type II NADH dehydrogenase (FAD-dependent)	MED152_12814
Succinate dehydrogenase	MED152_08110 (*sdhA*), MED152_08115 (*sdhB*), MED152_08105 (*sdhC*)
Cytochrome c oxidase	MED152_10275 (*coxM*), MED152_10280 (*coxN*), MED152_06775 (*coxO*), MED152_06770 (*coxP*)
Cytochrome oxidase, cbb3 type	MED152_03190 (*ccoNO*), MED152_03195 (*ccoS*)
Cytochrome c class I	MED152_10265
Cytochrome c	MED152_10245
Cytochrome c assembly protein	MED152_05520
Cytochrome oxidase assembly protein	MED152_06660
Cytochrome oxidase assembly factor	MED152_06780
60 KD inner membrane protein OxaA homolog	MED152_13299
SenC (3)	MED152_03785, MED152_06550, MED152_06755
Nitrogen assimilation	
Glutamate synthase	MED152_05815 (*gltB*), MED152_05820 (*gltD*)
Ammonium channel (*amtB*) (2)	MED152_05800, MED152_05810
Nitrogen regulatory protein P-II	MED152_05805
Glutamine synthetase, type II	MED152_05925
Glutamine synthetase, type III	MED152_05920
Glutamate dehydrogenase	MED152_08660
Phosphate metabolism	
Phosphate permease	MED152_09510
Polyphosphate kinase (3)	MED152_00300, MED152_00305, MED152_05065
Exopolyphosphatase	MED152_04590
H^+-translocating pyrophosphatase	MED152_11924
Soluble pyrophosphatase	MED152_11929
Sulfur assimilation	
APS/PAPS pathway	MED152_06170 (*cysD*), MED152_06175 (*cysH*), MED152_06165 (*cysN*), MED152_06160 (*cysI*), MED152_09765 (*cysJ*), MED152_00895 (*cysK*), MED152_06130 (*cysM*), MED152_06135 (*cysE*)
Sulphate permease (3)	MED152_09030, MED152_10290, MED152_12914
Iron assimilation	
Fe^{3+}-binding periplasmic protein (2)	MED152_01595, MED152_11574
ABC-type Fe^{3+}-siderophore transporter (2)	MED152_00690, MED152_00695
Mn^{2+}/Fe^{2+} transporter, NRAMP family (2)	MED152_03840, MED152_11859
Ferritin-like protein (2)	MED152_01695, MED152_03075
ATP-dependent Fe^{2+} transport system FeoAB	MED152_06540, MED152_06545
Fur (2)	MED152_00170, MED152_01940
Repressor	MED152_10760
Bicarbonate uptake	
SulP type transporter (BicA)	MED152_09030
SbtA	MED152_03855
Carbonic anhydrase	MED152_09035

The following additional pathways were complete and are not shown on the table: glycolysis, gluconeogenesis, TCA cycle, pentose phosphate pathway, purine biosynthesis and salvage, pyrimidine biosynthesis and salvage, thimidylate biosynthesis, amino acid metabolism, fatty acid biosynthesis, NAD biosynthesis, riboflavin and FAD biosynthesis, siroheme biosynthesis, quinone biosynthesis, H^+-ATPase, pyridoxal phosphate biosynthesis, pantothenate and CoA biosynthesis. Complete cobalamine, biotin and thiamine biosynthetic pathways were not found. Entner–Doudoroff pathway is incomplete.

Table S3. Key stress response-related genes in MED152

Pathway/protein (copy number when greater than one)	Type of stress	ORFs
Glycine betaine biosynthesis	Osmotic	MED152_05190 (*betA*), MED152_04445 (*betB*)
OsmC (2)	Osmotic	MED152_04480, MED152_07905
Large Conductance Mechanosensitive Ion Channel (MscL) family (2)	Osmotic	MED152_07755*, MED152_08290*
Small Conductance Mechanosensitive Ion Channel (MscS) family (6)	Osmotic	MED152_03080*, MED152_05265*, MED152_06730*, MED152_10700*, MED152_11349*, MED152_12279*
Voltage-Gated K^+ Channel superfamily (2)	Osmotic	MED152_01430*, MED152_04890*
K^+ uptake protein TrkH	Osmotic	MED152_12369*
Monovalent Cation:Proton Antiporter-2 (CPA2) family K^+ transporter	Osmotic	MED152_02125*
Cl^- Channel (ClC) family	Osmotic	MED152_07855*
ATP-Binding Cassette (ABC) type Na^+ transporter (3)	Osmotic	MED152_05270*, MED152_05275*, MED152_13214*
Phage shock protein C (PspC)	Osmotic/cryoprotection	MED152_03405
Fatty acid desaturase (3)	Cryoprotection	MED152_06405, MED152_06505, MED152_09405
NhaC Na^+:H^+ Antiporter (NhaC) family (2)	Internal pH regulation/Na^+ extrusion	MED152_06020*, MED152_08675*
NhaD Na^+:H^+ Antiporter (NhaD) family	Internal pH regulation/Na^+ extrusion	MED152_05770*
Monovalent Na^+/H^+ Antiporter-1 (CPA1) family (2)	Internal pH regulation/Na^+ extrusion	MED152_02960*, MED152_03050*
Sterol desaturase (3)	Cryoprotection	MED152_02375, MED152_07040, MED152_10110
DEAD box RNA helicase (12)	Cryoprotection	MED152_00240, MED152_01675, MED152_01850, MED152_02405, MED152_02435, MED152_02810, MED152_03055, MED152_03355, MED152_07660, MED152_08085, MED152_10165, MED152_13039
Cold-shock DNA-binding protein (3)	Cryoprotection	MED152_01260, MED152_02750, MED152_06655
Stationary-phase survival acid phosphatase SurE	Starvation	MED152_06290
Guanosine polyphosphate (ppGpp) pyrophosphohydrolase/synthetase	Starvation	MED152_01945
Phosphate starvation-inducible protein, PhoH	Starvation	MED152_13139
Catalase	Oxidative	MED152_05245
Cu/Zn superoxide dismutase	Oxidative	MED152_04650
Fe/Mn superoxide dismutase	Oxidative	MED152_07780
Cytochrome-c peroxidase (2)	Oxidative	MED152_03805, MED152_08020
Protein-Met-S-oxide reductase, type A (3)	Oxidative	MED152_03040, MED152_07435, MED152_07450
Protein-Met-S-oxide reductase, type B (2)	Oxidative	MED152_01175, MED152_07430
Glutathione peroxidase	Oxidative	MED152_05385
AhpC/TSA family protein (3)	Oxidative	MED152_02100, MED152_05240, MED152_09055
Redoxin (16)	Oxidative	MED152_00520, MED152_00525, MED152_01395, MED152_01805, MED152_01910, MED152_03415, MED152_05235, MED152_06110, MED152_06230, MED152_06900, MED152_07620, MED152_08425, MED152_10120, MED152_10405, MED152_10550, MED152_11584
Arsenate oxidoreductase ArsC (3)	Metals	MED152_00320, MED152_02710, MED152_07715
Arsenite resistance protein ArsB	Metals	MED152_02140
Heavy metal transport/detoxification protein (2)	Metals	MED152_09525, MED152_11224
MerTP Mercuric Ion (Hg^{2+}) Permease (MerTP) family transporter	Metals	MED152_00110*
Arsenical Resistance-3 (ACR3) family transporter	Metals	MED152_02140*
Mn^{2+} Transporter (Nramp) family (2)	Metals	MED152_03840*, MED152_11859*
Integral membrane protein TerC family (2)	Metals	MED152_01215, MED152_13169
β-Lactamase (8)	Drugs	MED152_03350, MED152_03530, MED152_07260, MED152_07965, MED152_08005, MED152_09070, MED152_10295, MED152_10385
Chloramphenicol acetyltransferase	Drugs	MED152_03570
ATP-Binding Cassette (ABC) type transporter (4)	Drugs	MED152_00590*, MED152_00620*, MED152_11514*, MED152_13184*
Major Facilitator Superfamily (MFS) transporter (2)	Drugs	MED152_08705*, MED152_11639*
Multidrug/Oligosaccharidyl-lipid/Polysaccharide (MOP) Flippase superfamily transporter (3)	Drugs	MED152_01915*, MED152_02820*, MED152_09425*
Resistance-Nodulation-Cell Division (RND) superfamily transporter	Drugs	MED152_08525*

ORFs with an asterisk also appear in Table S8 showing transporters.

Table S4. Genes coding for replication, repair, and recombination mechanisms in MED152

Gene name	ORFs	Protein description; pathway
uvrA	MED152_01780, MED152_04900	ATPase, DNA binding; nucleotide excision repair
uvrB	MED152_03455	Helicase; nucleotide excision repair
uvrC	MED152_09500	Nuclease; nucleotide excision repair
uvrD	MED152_01615, MED152_10775	UvrD/REP helicase; nucleotide excision repair
mfd	MED152_10740	Transcription-repair coupling factor; nucleotide excision repair
ung	MED152_00800	Uracyl DNA glycosylase; base-excision repair
nth	MED152_01800, MED152_09640	HhH-GPD superfamily; base excision repair
tag	MED152_11989	DNA-3-methyladenine glycosidase I; base excision repair
xth	MED152_10450	Exodeoxyribonuclease III; base excision repair
	MED152_00700	Non-canonical purine NTP pyrophosphatase, rdgB/HAM1 family; DNA repair
radA	MED152_08245	DNA repair protein RadA/Sms
radC	MED152_12379	DNA repair protein RadC
mutL	MED152_01325	DNA mismatch repair protein
mutS	MED152_00795, MED152_08905, MED152_09620	DNA mismatch repair protein
recA	MED152_08735	Recombinase; DNA recombination
recD	MED152_00290	ATP-dependent exoDNAse (exonuclease V), alpha subunit; DNA recombination
recF	MED152_01345	DNA replication; DNA recombination
recJ	MED152_07910	Single-stranded-DNA-specific exonuclease; DNA degradation
recN	MED152_13259	DNA repair protein RecN
recO	MED152_10500	Recombination protein RecO
recR	MED152_04560	Recombination protein RecR
ruvA	MED152_06690	DNA helicase RuvA; DNA recombination
ruvB	MED152_10285, MED152_12424	DNA helicase RuvB; DNA recombination
ruvC	MED152_12024	Endodeoxyribonuclease RuvC; DNA recombination
	MED152_00210, MED152_01115, MED152_02515, MED152_04860, MED152_07400, MED152_09890	DNA recombination
cinA	MED152_03985	Competence/DNA damage-inducible protein CinA
	MED152_00230, MED152_02730, MED152_05735, MED152_07035, MED152_09645	DNA binding protein
polA	MED152_10105	DNA polymerase I
holA	MED152_00675	DNA polymerase III, delta subunit
dnaA	MED152_13404	Chromosomal replication initiator protein DnaA
dnaB	MED152_01985	DnaB replicative DNA helicase
dnaE	MED152_07245, MED152_11814	DNA polymerase III, alpha subunit
dnaG	MED152_11074	DNA primase
dnaN	MED152_04670	DNA polymerase III, beta subunit
dnaQ	MED152_01490, MED152_05060, MED152_11754	DNA polymerase III, epsilon subunit
dnaX	MED152_03045	DNA polymerase III, subunit gamma/tau
holB	MED152_09190	DNA polymerase III, delta' subunit
ligA	MED152_13454	DNA ligase, NAD-dependent
topA	MED152_11284	DNA topoisomerase I
exc	MED152_01555	DNA topoisomerase III
parC	MED152_01590	DNA topoisomerase IV, subunit A
parE	MED152_01585	DNA topoisomerase IV, subunit B
gyrA	MED152_04790	DNA gyrase, subunit A
gyrB	MED152_12139	DNA gyrase, subunit B
ssb	MED152_09160, MED152_09635	Single-stranded DNA-binding protein
rnhB	MED152_05995	Ribonuclease HII
	MED152_09655	Ribonuclease, Rne/Rng family
mfd	MED152_10740	Transcription-repair coupling factor
dcm	MED152_07385	DNA (cytosine-5-)-methyltransferase; restriction/modification
ogt	MED152_07670	Methylated DNA-protein Cys methyltransferase
	MED152_09340	6-O-methylguanine DNA methyltransferase
	MED152_07920	Helicase
cspD	MED152_01260, MED152_02750, MED152_06655	'Cold-shock' DNA-binding protein
	MED152_01310, MED152_04910, MED152_07835, MED152_09660, MED152_11354	Endonuclease/exonuclease/phosphatase family protein
	MED152_00240, MED152_01675, MED152_01850, MED152_02405, MED152_02435, MED152_02810, MED152_03055, MED152_03355, MED152_07660, MED152_08085, MED152_10165, MED152_13039	DEAD/DEAH box helicase
	MED152_05340, MED152_05350, MED152_05405	DNA photolyase/cryptochrome
	MED152_12919, MED152_13194	NUDIX hydrolase; base-excision repair

Table S5. Distribution of the anaplerotic enzymes pyruvate and PEP carboxylases, as well as carbonate anhydrase in cultured marine heterotrophic bacteria that contain the proteorhodopsin gene

Organism	Genbank accession no.	Genome size, bp	Taxonomy	Pyruvate carboxylase	PEP carboxylase	Carbonate anhydrase
Polaribacter sp. MED152	AANA00000000	2,967,150	Bacteroidetes	+	+	+
Polaribacter irgensii 23-P	AAOG00000000	2,745,458	Bacteroidetes	+	+	+
Dokdonia sp. MED134	AAMZ00000000	3,301,953	Bacteroidetes	+	+	+
Flavobacterium BAL38	AAXX00000000	2,806,989	Bacteroidetes	-	+	+
Pelagibacter ubique HTCC1002	AAPV00000000	1,328,618	Alphaprot.	-	-	-
Pelagibacter ubique HTCC1062	CP000084	1,308,759	Alphaprot.	-	+	-
Octadecabacter antarcticus 307	Moore F.	4,909,025	Alphaprot.	+	-	+
BAL199	ABHC00000000	6,162,950	Alphaprot.	-	+	++
Methylophilales HTCC2181	AAUX00000000	1,304,428	Betaprot.	-	-	+
Vibrio harveyi ATCC BAA-1116	CP000789	6,058,377	Gammaprot.	-	+	+
Vibrio sp. S14	AAOJ00000000	5,169,214	Gammaprot.	-	+	+
HTCC2143	AAVT00000000	3,925,629	Gammaprot.	-	+	++
HTCC2207	AAPI00000000	2,620,870	Gammaprot.	-	-	+
Photobacterium sp. SKA34	AAOU00000000	4,991,572	Gammaprot.	-	+	+
Marinobacter sp. ELB17	AAXY00000000	4,928,595	Gammaprot.	-	+	++

The presence of the proteorhodopsin gene is based on hits against PF01036. Further, only those peptides that contained an Asp85 and Asp96, or a different carboxylate residue, were considered (Asp85 and Asp96 functions as the H^+ acceptor and donor, respectively, during the rhodopsin photocycle in the H^+ pump type of rhodopsins). Number of pluses indicates the number of genes with the same putative function. Pyruvate carboxylase (EC 6.4.1.1) catalyzes the following reaction: pyruvate + ATP + HCO_3^- <=> ADP + phosphate + oxaloacetate. PEP carboxylase (EC 4.1.1.31) catalyzes the reaction: PEP + HCO_3^- + H_2O <=> phosphate + oxaloacetate. Carbonate anhydrase (EC 4.2.1.1) catalyzes the reaction: HCO_3^- <=> CO_2 + H_2O.

Table S6. Domains and peptides with a putative role in light absorption and response

Domain/peptide	ORFs	Function
Light sensing		
Phytochrome region	MED152_00100	Red and far-red light sensor; chromophore is a linear tetrapyrrole bound via a Cys residue
PAS and PAC domains	MED152_00100, MED152_03150, MED152_04250, MED152_06325	Flavin-binding; component of phytochromes; sensor of oxygen, redox and light
GAF domain	MED152_00100, MED152_02465, MED152_03150, MED152_05910, MED152_11204	Binds small-molecules (cAMP or cGMP) that act as second messengers; component of phytochromes
DNA photolyase/cryptochrome, animal cryptochrome and (6–4) photolyase family	MED152_05350	Flavin-binding; senses short wavelengths of visible light
Cryptochrome, DASH family	MED152_05340	Flavin-binding; senses short wavelengths of visible light
Deoxyribodipyrimidine photo-lyase, class I	MED152_05405	Flavin-binding; senses short wavelengths of visible light
BLUF domain	MED152_08960	Flavin-binding; blue light sensor
Synthesis of β-carotene		
Isopentenyl-diphosphate delta-isomerase (*idi*)	MED152_03605	Synthesis of β-carotene from isopentenyl diphosphate and dimethylallyl diphosphate
Farnesyl-diphosphate synthase (*ispA*)	MED152_07135 or MED152_08510	
Geranylgeranyl diphosphate synthase (*crtE*)	MED152_07135 or MED152_08510	
Phytoene synthase (*crtB*)	MED152_02670	
Phytoene desaturase (*crtI*)	MED152_02575 or MED152_02675	
Lycopene β-cyclase (*crtY*)	MED152_02660	
15,15′-β-Carotene dioxygenase (*blh*)	MED152_12849	Final step in the synthesis of retinal from β-carotene
Synthesis of proteorhodopsin apoprotein		
Opsin	MED152_12844	
Additional carotenoids		
Spheroidene monooxygenase (*crtA*)	MED152_02580	
Methoxyneurosporene dehydrogenase (*crtD*)	MED152_02575 or MED152_02675	
β-Carotene hydroxylase (*crtZ*)	MED152_02565 or MED152_02665	Synthesis of zeaxanthin from ß-carotene

Table S7. Distribution of putative light sensors in proteorhodopsin-containing marine bacteria (as defined in Table S5)

Organism	BLUF	Phytochrome	Animal cryptochrome and (6–4) photolyase family	Cryptochrome, DASH family	Photolyase, class I
Polaribacter sp. MED152	+	+	+	+	+
Polaribacter irgensii 23-P	-	-	++	+	+
Dokdonia sp. MED134	+	+	+	+	+
Flavobacterium BAL38	-	-	+	+	+
P. ubique HTCC1002	-	-	-	-	+
P. ubique HTCC1062	-	-	-	-	+
Octadecabacter antarcticus 307	-	-	+	-	+
Methylophilales HTCC2181	-	-	-	+	+
Vibrio harveyi ATCC BAA-1116	-	-	+	+	+
Vibrio sp. S14	-	-	+	+	+
HTCC2143	-	-	+	+	+
HTCC2207	-	-	-	+	+
Photobacterium sp. SKA34	-	-	+	+	+
Marinobacter sp. ELB17	-	-	++	+	+

Presence of light sensors are predicted by hits to specific PFAMs or TIGRFAMs. DNA photolyase/cryptochrome subfamilies are based on phylogenetic analysis as in Fig. S6, as well as on custom-built PFAMs. Of 80 Moore genomes, 30 contain at least one BLUF domain, 18 contain at least one phytochrome domain, and 12 contain three cryptochrome/photolyase peptides (most contain one or two and no other contain four cryptochrome/photolyase peptides).

Table S8. Transporters identified in the genome of MED152

Substrate	Type of transporter	ORFs
	Ion channels	
Cl^-	Chloride Channel (ClC) family	MED152_07855
Mg^{2+}/Co^{2+}	CorA Metal Ion Transporter (MIT) family	MED152_12054
	Large Conductance Mechanosensitive Ion Channel (MscL) family	MED152_07755, MED152_08290
	Small Conductance Mechanosensitive Ion Channel (MscS) family	MED152_03080, MED152_05265, MED152_06730, MED152_10700, MED152_11349, MED152_12279
K^+	Voltage-Gated Ion Channel (VIC) superfamily	MED152_01430, MED152_04890
	ATP-Binding Cassette (ABC) superfamily	
bacitracin		MED152_00590
multidrug		MED152_00620
Fe^{3+}		MED152_00695, MED152_01595, MED152_11574
unknown		MED152_01685
unknown		MED152_02455
unknown		MED152_02840, MED152_02845, MED152_02850
lipoprotein releasing		MED152_03470, MED152_03875
lipid A export		MED152_04055
lipoprotein releasing		MED152_04345
unknown		MED152_04415
Na^+ efflux		MED152_05270, MED152_05275
Fe-S assembly/SufBCD system		MED152_07995, MED152_08000, MED152_08010
cell division		MED152_12239
lipoprotein releasing		MED152_10825
multidrug		MED152_11514
toluene tolerance		MED152_03275, MED152_12664, MED152_12669
gliding motility		MED152_11809
gliding motility		MED152_13119
multidrug		MED152_13184
Na^+ efflux		MED152_13214
	Membrane pyrophosphatase	
H^+		MED152_11924
	P-type ATPase superfamily	
Cu^{2+}		MED152_00135
Zn^{2+}		MED152_00155
Cation		MED152_03205
	Secondary transporters	
$H_2AsO_3^-$	Arsenical Resistance-3 (ACR3) family	MED152_02140
Na^+/alanine	Ala or Gly:Cation Symporter (AGCS) family	MED152_04880
NH_4^+	Ammonium Transporter (Amt) family	MED152_05800, MED152_05810
amino acid	Amino Acid-Polyamine-Organocation (APC) family	MED152_09705
Na^+/unknown	Bile Acid:Na^+ Symporter (BASS) family	MED152_05040
Na^+/Ca^{2+}	Ca^{2+}:Cation Antiporter (CaCA) family	MED152_11634
nucleosides	Concentrative Nucleoside Transporter (CNT) family	MED152_11959
Na^+/H^+	Monovalent Cation:Proton Antiporter-1 (CPA1) family	MED152_02960, MED152_03050
K^+	Monovalent Cation:Proton Antiporter-2 (CPA2) family	MED152_02125
H^+/Na^+:glutamate	Dicarboxylate/Amino Acid:Cation (Na^+ or H^+) Symporter (DAACS) family	MED152_02010
Na^+/dicarboxylate	Divalent Anion:Na^+ Symporter (DASS) family	MED152_09770
drug/metabolite	Drug/Metabolite Transporter (DMT) Superfamily	MED152_02080, MED152_09260, MED152_09315, MED152_09530, MED152_10735
H^+/Na^+:sugar	Glycoside-Pentoside-Hexuronide (GPH):Cation Symporter family	MED152_00365, MED152_00460, MED152_05120
branched-chain amino acid	Branched Chain Amino Acid:Cation Symporter (LIVCS) family	MED152_08060
unknown	Major Facilitator Superfamily (MFS)	MED152_01190, MED152_03745, MED152_05095
nucleosides	Major Facilitator Superfamily (MFS)	MED152_05175
glucose/galactose	Major Facilitator Superfamily (MFS)	MED152_08460, MED152_13289
multidrug efflux	Major Facilitator Superfamily (MFS)	MED152_08705, MED152_11639
arabinose	Major Facilitator Superfamily (MFS)	MED152_09090

Substrate	Type of transporter	ORFs
hexuranate	Major Facilitator Superfamily (MFS)	MED152_09855
multidrug efflux	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_01915, MED152_02820, MED152_09425
oligosaccharide repeat unit	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_03130, MED152_10600
polysaccharide	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_04985
O-antigen	Multidrug/Oligosaccharidyl-lipid/ Polysaccharide (MOP) Flippase superfamily	MED152_12399
Na^+/H^+	NhaC Na^+:H^+ Antiporter (NhaC) family	MED152_06020, MED152_08675
Na^+/H^+	NhaD Na^+:H^+ Antiporter (NhaD) family	MED152_05770
Mn^{2+}	Metal Ion (Mn^{2+}-iron) Transporter (Nramp) family	MED152_03840, MED152_11859
60-kDa inner membrane protein OxaA homolog	Cytochrome Oxidase Biogenesis (Oxa1) family	MED152_13299
dipeptide/tripeptide: H^+	H^+-dependent Oligopeptide Transporter (POT) family	MED152_06255, MED152_06260
amino acid	Resistance to Homoserine/Thr (RhtB) family	MED152_01410
Cu^{2+}/Ag^+ efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_00145
multidrug efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_08525
cation/multidrug efflux	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_10530
unknown	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_11044
protein-export (SecDF)	Resistance-Nodulation-Cell Division (RND) superfamily	MED152_12119
Na^+/HCO_3^-	Na^+-dependent Bicarbonate Transporter (SBT) family	MED152_03855
Na^+/solute	Solute:Na^+ Symporter (SSS) family	MED152_04565, MED152_04660, MED152_05085
$SO_4^=$	$SO_4^=$ Permease (SulP) family	MED152_09030, MED152_10290, MED152_12914
protein export	Twin Arginine Targeting (Tat) family	MED152_02445, MED152_08785
K^+	K^+ Transporter (Trk) family	MED152_12369
	Unclassified transporters	
Ca^{2+}	Ca^{2+}-ATPase	MED152_12659
Fe^{2+}	Fe^{2+} Uptake (FeoB) family	MED152_06540
Hg^{2+}	MerTP Mercuric Ion (Hg^{2+}) Permease (MerTP) family	MED152_00110
Mg^{2+}	Mg^{2+} Transporter-E (MgtE) family	MED152_08915
nicotinamide mononucleotide	Nicotinamide Mononucleotide (NMN) Uptake Permease (PnuC) family	MED152_09745

Table S9. Genes and domains with a potential role in adhesion

Domain	ORFs	Protein family
Fasciclin	MED152_12639, MED152_13164	PF02469
Thrombospondin type 3 repeat	MED152_00310, MED152_03360, MED152_04530, MED152_10785, MED152_07970	PF02412
Fibronectin type III	MED152_07970	PF00041
PKD	MED152_00430, MED152_03360, MED152_04535, MED152_04570, MED152_04575, MED152_09845	PF00801
FG-GAP	MED152_05130, MED152_05135, MED152_05145, MED152_08145	PF01839
ASPIC and UnbV	MED152_05145, MED152_08145, MED152_13274	PF07593
Lectin C-type domain	MED152_07900	PF00059
Cadherin	MED152_00085	IPR002126
HYR	MED152_00085	PF02494
Leu rich repeat	MED152_10960, MED152_11029	PF00560
β-Propellor repeat	MED152_02285	IPR013519
Cellulose binding motif	MED152_00085	PF00553
DUF11	MED152_07970	PF01345
Von Willebrand factor type A	MED152_08995, MED152_09005, MED152_09010	PF00092
Cna protein B-type	MED152_10890	PF05738
Carbohydrate binding domain	MED152_00430, MED152_03360, MED152_04575, MED152_09845, MED152_11564	PF02018
Carbohydrate binding module (family 6)	MED152_00755	PF03422
Cellulose binding domain	MED152_00870	PF00942
DUF638	MED152_03735	PF04829
F5/8 type C domain	MED152_08630, MED152_13279	PF00754
Starch-binding domain	MED152_11969	PF00686