Applied and Environmental Microbiology, Dec. 2011, p. 8676–8686 Vol. 77, No. 24
0099-2240/11/$12.00 doi:10.1128/AEM.06152-11

Genomics of the Proteorhodopsin-Containing Marine Flavobacterium *Dokdonia* sp. Strain MED134▽†

José M. González,[1]* Jarone Pinhassi,[2] Beatriz Fernández-Gómez,[3] Montserrat Coll-Lladó,[3] Mónica González-Velázquez,[1] Pere Puigbò,[4] Sebastian Jaenicke,[5] Laura Gómez-Consarnau,[6] Antoni Fernàndez-Guerra,[7] Alexander Goesmann,[5] and Carlos Pedrós-Alió[3]

Department of Microbiology, University of La Laguna, ES-38206 La Laguna, Tenerife, Spain[1]; Marine Microbiology, School of Natural Sciences, Linnaeus University, SE-39182 Kalmar, Sweden[2]; Departament de Biologia Marina i Oceanografia, Institut de Ciències del Mar, CSIC, ES-08003 Barcelona, Spain[3]; National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894[4]; Center for Biotechnology, Bielefeld University, D-33594 Bielefeld, Germany[5]; Department of Biological Sciences, University of Southern California, Los Angeles, California 90089[6]; and Department of Continental Ecology-Biogeodynamics and Biodiversity Interactions, Centre d'Estudis Avançats de Blanes, CSIC, ES-17300 Blanes, Spain[7]

Received 11 July 2011/Accepted 4 October 2011

Proteorhodopsin phototrophy is expected to have considerable impact on the ecology and biogeochemical roles of marine bacteria. However, the genetic features contributing to the success of proteorhodopsin-containing bacteria remain largely unknown. We investigated the genome of *Dokdonia* sp. strain MED134 (*Bacteroidetes*) for features potentially explaining its ability to grow better in light than darkness. MED134 has a relatively high number of peptidases, suggesting that amino acids are the main carbon and nitrogen sources. In addition, MED134 shares with other environmental genomes a reduction in gene copies at the expense of important ones, like membrane transporters, which might be compensated by the presence of the proteorhodopsin gene. The genome analyses suggest *Dokdonia* sp. MED134 is able to respond to light at least partly due to the presence of a strong flavobacterial consensus promoter sequence for the proteorhodopsin gene. Moreover, *Dokdonia* sp. MED134 has a complete set of anaplerotic enzymes likely to play a role in the adaptation of the carbon anabolism to the different sources of energy it can use, including light or various organic matter compounds. In addition to promoting growth, proteorhodopsin phototrophy could provide energy for the degradation of complex or recalcitrant organic matter, survival during periods of low nutrients, or uptake of amino acids and peptides at low concentrations. Our analysis suggests that the ability to harness light potentially makes MED134 less dependent on the amount and quality of organic matter or other nutrients. The genomic features reported here may well be among the keys to a successful photoheterotrophic lifestyle.

Sunlight is the principal source of energy in the biosphere and organisms have evolved means to harvest it. In the photosynthetic complexes of plants and bacteria, chlorophyll and carotenoid molecules absorb light and convey the energy to reaction centers for energy transformation. Compared to these complexes, proteorhodopsins (PRs) are relatively simple membrane proteins, composed of a single protein with retinal as the light absorbing molecule, which function as H^+ pumps when exposed to light. The H^+ gradient can then be dissipated to generate ATP, propel flagella, or transport substrates inside the cell. Thus, an array of possible functions has been proposed for PR (18, 43, 82).

The gene for PR was first found in DNA fragments directly obtained from seawater and later found to be abundant and diverse (6, 74). Single-cell genomes have also established the presence of the PR gene in environmental marine bacteria (70). Its function as an H^+ pump for energy conservation was demonstrated in *Escherichia coli* cells transformed with the PR gene (43) and in *Shewanella* (29). However, few experimental studies have examined the physiological and ecological role of PR in bacteria. Among these, light did not make any difference in the growth of PR-containing alpha- and gammaproteobacteria (21, 71). Therefore, the activity of this membrane protein in marine prokaryotes may be other than growth enhancing, such as promoting survival during times of scarce nutrients (24, 69). Nevertheless, light-dark cycles have been shown to result in upregulation of PR gene expression and growth of natural populations of flavobacteria from coastal samples (38). Working on the PR-containing model organism *Dokdonia* sp. strain MED134, Gómez-Consarnau et al. (23) were the first to directly show that marine bacteria can use PR phototrophy to grow better in light than in the dark, although only under nutrient-deprived conditions. As expected, the genes that encode the apoprotein and enzymes needed to synthesize retinal were found. Key amino acid positions are conserved in MED134 PR to function as an H^+ pump that absorbs green light, as in environmental *Bacteroidetes* sequences. Laser flash photolysis and absorption spectra confirmed these predictions (23). Subsequent studies of *Dokdonia* sp. MED134 also support these findings and further demonstrate that the transcription of the opsin gene (*prd*) and retinal synthesis genes is also enhanced in response to light (34). *Dokdonia* sp. MED134 remains the only PR-containing bacterium thus far shown to

---

* Corresponding author. Mailing address: Department of Microbiology, University of La Laguna, ES-38206 La Laguna, Tenerife, Spain. Phone: 34 (922) 318515. Fax: 34 (922) 318666. E-mail: jgonzalez@ull.es.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 14 October 2011.

use light to increase its growth. Therefore, using *Dokdonia* as a model organism has the potential to shed light on the biological function of this highly abundant membrane protein in aquatic environments.

Dokdonia sp. MED134 is a representative of the phylum *Bacteroidetes*, an abundant, widespread, and diverse group of bacteria in the ocean. They seem especially abundant on suspended particles compared to free-living prokaryotic communities (14) and respond faster than other bacteria to sporadic nutrient increase, such as during algal bloom decays (52, 61). Members of the phylum have some distinctive features, such as gliding motility over surfaces, a characteristic that has made them model organisms for biochemistry and molecular biology studies (44–45). *Bacteroides* have also been model organisms for research into their unique mechanism of attachment and degradation of polymeric substances (2–3). This mechanism for the efficient degradation of particulate organic matter is common to marine *Bacteroidetes* based on whole-genome analysis of isolates (4, 25). Genome analysis has therefore brought forward hypotheses on the role of *Bacteroidetes* in marine biogeochemical cycling. However, these hypotheses have only been tested in a few cases.

The genome sequence of MED134 opens up the possibility of studying the effect of light on the growth and genetic repertoire of an organism with a PR-based metabolism. Genome analysis could provide information about the genes potentially involved in the strong light response of MED134, since it is crucial to identify genomic features characteristic of these phototrophs. It will also aid in understanding the light response, or lack of it, among other bacteria containing PR genes. Here we analyze the genome of MED134 to better understand the potential benefits of PR phototrophy in the ecology and biogeochemical role of marine bacteria.

MATERIALS AND METHODS

Isolation of flavobacteria. Bacteria were isolated from northwestern Mediterranean Sea surface water (0.5-m depth), collected 1 km off the coast of Catalonia at the Blanes Bay Microbial Observatory (41°40′N, 2°48′E; Spain). Strain MED134 was isolated on Zobell agar plates.

Genomic sequencing and annotation. Whole-genome sequencing was done by the J. Craig Venter Institute through the Gordon and Betty Moore Foundation initiative in Marine Microbiology (https://research.venterinstitute.org/moore/). Large (40 kb) and small (4 kb) insert random libraries were sequenced by the Sanger method as described by Goldberg et al. (22) with an average success rate of 92% and an average high-quality read length of 845 (large insert) and 874 (small insert) nucleotides. The completed genome sequence of MED134 contains 33,792 reads, achieving an average of 9-fold sequence coverage per base. The genome was finally assembled on one scaffold after alignment with the genome of *Krokinobacter* sp. strain 4H-3-7-5 (35) using the MUMmer software (37) before gap closure (Lifesequencing, Valencia, Spain).

Open reading frames (ORFs) were predicted and auto-annotated using GenDB (48) and manually curated. Hidden Markov Models (HMMs) were used to determine ORF membership in families and superfamilies using Pfam v25.0, run with HMMER3 (16). A hit was considered valid if its score was equal or bigger than the "gathering score" for the model.

Genes encoding candidate glycoside hydrolases, polysaccharide lyases, carbohydrate esterases, and carbohydrate binding modules were detected by searching for the PFAMs in the Carbohydrate Active Enzymes (CAZy; www.cazy.org) database (10). Peptidases were detected by running all PFAMs against the peptidase sequences in the MEROPS database (http://merops.sanger.ac.uk [59]) and manually extracting peptidase PFAMs. The resulting PFAMs were used to detect the peptidases in the genomes. *Bacteroidetes* for sequence comparisons are listed in Table S1 in the supplemental material. GC-Profile to detect variations in GC content (20, 81), SIGI-HMM based on hidden Markov model algorithm (75), and SeqWord Genome Browser to find oligonucleotide usage pattern anomalies (19) were used to identify potential foreign DNA. PSORTb version 3.0.0 (80) was used to predict the localization of each predicted protein. Lipoproteins were identified using LipoP (30). When necessary, analysis was carried out by customized Python scripts.

TABLE 1. General features of the *Dokdonia* MED134 genome

Parameter	Value
No. of bp	3,301,953
G+C content (%)	38
Coding density (%)	91
No. of:	
Predicted protein coding genes	3,008
Predicted proteins with putative function	1,907
Conserved hypothetical proteins	921
Unknown proteins unique to MED134	180
tRNAs	43
Structural RNAs	6
rrn operons	3
Genes in paralogous clusters	417
Lipopeptides	252
Proteins with signal peptide	451
Transporter-related proteins	109
Peptidases	106
Glycosyl hydrolases	18

Predicted highly expressed genes. The method to predict a group of highly expressed (PHX) genes in MED134 is based on the codon bias due to translational selection in these genes, as described in Puigbò et al. (56, 57). First, MED134 genes were evaluated for translational selection. Correspondence analysis of the Relative Synonymous Codon Usage (RSCU), traditionally used to determine translational selection (51), was used to test whether there is codon usage bias in the ribosomal protein genes compared to the rest of the genes in the genome. In MED134, the group of ribosomal protein genes forms an independent cluster in the first axis of the correspondence analysis (see Fig. S1 to S3 in the supplemental material). Thus, the genome of MED134 is under strong translational selection and a group of PHX genes were predicted. The group of ribosomal protein genes was used as a seed, together with an iterative algorithm, to define the PHX genes. The CAIrp (Codon Adaptation Index, using the mean codon usage of ribosomal protein genes as a reference [64]) of each gene was calculated (57). Genes with the highest CAIrp values were used as a reference set to recalculate the CAI (Codon Adaptation Index using the mean codon usage of PHX genes as a reference) values of all genes. This process was iteratively repeated until a homogeneous group of PHX genes was reached.

Nucleotide sequence accession number. The sequences of *Dokdonia* sp. MED134 can be accessed using GenBank accession number AAMZ00000000.

RESULTS AND DISCUSSION

Genome properties. Table 1 summarizes the properties of the MED134 genome. The genome size (3,301,953 bp) is in the lower average range for nonsymbiotic *Bacteroidetes*, which varies from 9.8 Mb (*Microscilla marina*) to ~3 Mb in the case of *Polaribacter* spp. Notably, the majority of free-living heterotrophic marine bacteria with smaller genomes are members of the SAR11 group (~1.3 Mbp), but also the environmental flavobacteria genomes MS024-2A and MS024-3C are estimated to have smaller genomes (1.9 and 1.5 Mbp, respectively; see Table S1 in the supplemental material) (77). The number of conserved genes in MED134, however, was similar to that of *Bacteroidetes* with larger genomes and consequently, COG analysis showed that MED134 contained 170 of the 213 core genes described for bacteria in Bratlie et al. (8), which is within the range of most *Bacteroidetes* (166 to 176). Our analyses indicate that MED134 and other marine *Bacteroidetes* with PR genes have a remarkably low fraction of paralogous genes

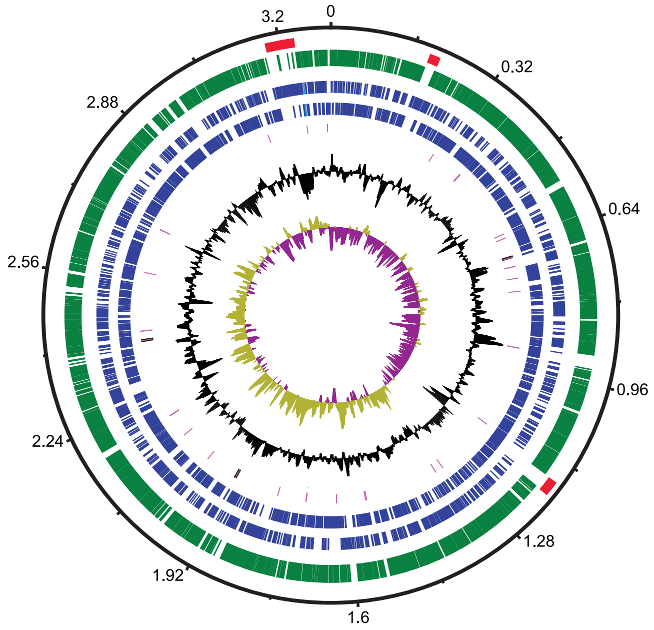

FIG. 1. Schematic circular representation of the *Dokdonia* sp. MED134 genome sequence. Outer ring, sequence address in nucleotides (Mbp). The green ring shows matches to *Krokinobacter* sp. 4H-3-7-5 based on BLASTP reciprocal best hits; red shows the location of the predicted genome islands and cyan the position of the *prd* and *blh*. The blue rings show the location of the identified genes in the positive and negative strands. The next ring shows tRNA in pink and rRNA in black. Black ring, deviation from the average %G+C. Inner ring, GC skew (G−C)/(G+C).

(average 13.7%) compared to other marine *Bacteroidetes* (see Table S1 in the supplemental material) (77). Thus, compared to nonphototrophic counterparts, the phototrophic *Bacteroidetes* have reduced genomes at the cost of a reduced number of genes within each paralogous gene family but retain a similar number of gene families. The MED134 genome contains three rRNA operons (Fig. 1), each with a copy of the 5S, 16S, and 23S rRNA genes, as well as 43 tRNA genes and six structural RNA genes. The latter include, as predicted by Rfam (http://rfam.sanger.ac.uk), one RNase P class A, one transfer-mRNA (tmRNA), one signal recognition particle RNA (SRP RNA), and three riboswitches, cobalamin, *S*-adenosylmethionine (SAM), and thiaminepyrophosphate (TPP). Both cobalamin and TPP riboswitches target TonB-dependent transporters (MED134_04109 and MED134_10161, respectively), and the SAM riboswitch targets *metK* (methionine adenosyltransferase; MED134_00225). The tmRNA in particular is expressed at high copy in *Dokdonia* sp. MED134 in response to low nutrients (34) and is likely to have a role in the recycling of proteins under stress conditions when the genes are not properly translated (15).

Physiology. Analysis of the MED134 genome confirmed a strictly aerobic type of heterotrophic metabolism. The Embden-Meyerhof-Parnas pathway, gluconeogenesis, and the tricarboxylic acid (TCA) cycle were complete (see Table S2 in the supplemental material). No mechanism for autotrophic growth or alternative sources of reducing power other than organics were identified. As expected for marine bacteria (36), MED134 has an Na^+-translocating NADH:quinone oxidoreductase as the main positive ion membrane pump. Thus, marine bacteria such as MED134 typically use an Na^+ gradient for energy conservation (i.e., in the first step of the respiratory chain) rather than using an H^+ gradient like in most nonmarine bacteria (26). This Na^+ gradient can also be used to transport substrates inside the cell or to exchange for H^+ through the Na^+/H^+ antiporters (two such antiporters are present in MED134). Na^+/H^+ exchangeability through antiporters is a common feature in bacteria able to pump both ions for energy conservation (26). Considering that the respiratory electron chain pumps mostly Na^+, while the PR pumps exclusively H^+ instead, the bioenergetics of the cell should be different when the bacterium grows in light versus darkness. For example, the ATPase activity relies on the Na^+/H^+ antiporter activity when the bacterium grows in darkness, since the respiratory chain is pumping both ions. Likewise, some substrates enter the cell through an Na^+- or H^+-gradient based transporter. Kimura et al. (34) recently provided evidence that Na^+-transport related genes, as well as *prd* and genes involved in retinal biosynthesis, are upregulated in the presence of light. Switching between light and darkness should change the balance of the Na^+/H^+ gradients for energy conservation, and the cell environment might thus have a regulatory role over genes not obviously otherwise connected to the light response.

In addition, MED134 does not encode a bc_1 complex (complex III) in the respiratory chain. Like other aerobic *Bacteroidetes*, instead it has the four genes that encode a menaquinol:cytochrome *c* oxidoreductase (alternative complex III; MED134_11366-MED134_11381) (46). The genes for the synthesis of menaquinones are also present, as are the genes encoding cytochrome *c*, cytochrome *c* oxidase, and ATP synthase. *ccoNOQP* encode a cbb_3-type cytochrome oxidase, which has a very high affinity for O_2 and allows respiration under low-O_2 levels (55). This would allow MED134 to manage transiently low-O_2 tension microniches, which have been suggested to form in particles due to active metabolism and low O_2 diffusion rates (53).

MED134 contains an unexpected number of enzymes involved in anaplerotic metabolism (Fig. 2). Anaplerotic enzymes replenish the TCA cycle by the carboxylation of three-carbon compounds (pyruvate or PEP) to four-carbon TCA cycle intermediates (malate or oxaloacetate). Restoration of TCA cycle intermediates is necessary when precursors are taken up for biosynthesis. This is the case when the bacterium has a surplus of energy in the form of an H^+ gradient. The PR activity generates this gradient, and the carbon flow should go toward biosynthesis and less to oxidation (with CO_2 as the final product), making the metabolism more efficient in the presence of light. Genes for four different anaplerotic enzymes are present in MED134 (i.e., PEP carboxykinase, pyruvate carboxylase, PEP carboxylase, and malic enzyme). A SulP-type Na^+-dependent bicarbonate transporter is also present, and a carbonic anhydrase interconverts CO_2 and HCO_3^- to ensure that they are not limiting substrates for the anaplerotic enzymes.

A mixotrophic CO_2 assimilation pathway was proposed in the bacteriochlorophyll *a* containing alphaproteobacterium *Roseobacter denitrificans* in which malic enzyme takes most of the carbon flow (72). The main anaplerotic enzyme in MED134 is not yet known, although it could vary depending on the conditions. Malic enzyme relies on $NADP^+/NADPH$ to

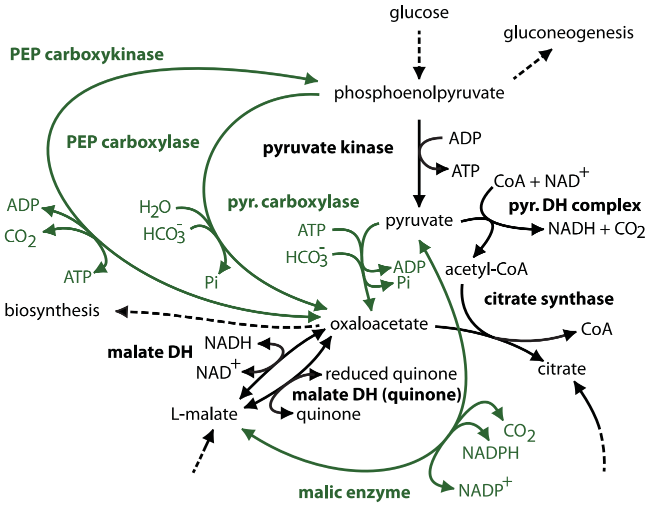

FIG. 2. Enzymes involved in MED134 anaplerotic metabolism. Enzymes that connect glycolysis/gluconeogenesis with the TCA cycle direct the carbon flux through the main metabolic pathways. These reactions are involved in the regeneration of precursors in the TCA cycle.

function while pyruvate carboxylase depends on ATP; therefore, the carbon flow might take one route or another according to the balance of energy and reducing power in the cell. In any case, anaplerotic reactions in MED134 should make the bacterium more versatile to adapt its carbon flux to different substrates and growth conditions and thus maximize resource utilization. Such conditions include growth in light versus darkness or on energy-rich substrates versus largely oxidized organic material. This would be potentially of major importance for efficient use of resources in a carbon-limited marine environment, since it would allow bacteria to use assimilated organic carbon economically, primarily for biosynthesis.

Utilization of high-molecular-weight compounds. Particulate material is an important component in the turnover and sinking flux of carbon in the ocean (65). Particles harbor a dense prokaryotic population in which *Bacteroidetes* are key players (14). *Bacteroidetes* in general, and flavobacteria in particular, are important in the degradation of organic matter during phytoplankton bloom conditions (52, 61). Accordingly, genome analysis implies that particle attachment and growth on high-molecular-weight organic matter is characteristic of marine *Bacteroidetes* isolates (4, 25, 58). *Bacteroidetes* genomes included in the present study stand out over those of other phyla for the large number of proteins predicted to be secreted and those that contain a lipoprotein signal. Attachment of marine *Bacteroidetes* may be through lipoproteins functioning as adhesins, and both lipoproteins and secreted proteins may participate in the degradation of polysaccharides and peptides for obtaining both carbon and energy. MED134, in particular, contains 451 proteins predicted to be secreted (15%) and 252 lipoproteins (8.4%). Among the secreted proteins, 39 are predicted to be peptidases (8.6%) and 12 are predicted to be glycosyl hydrolases (2.7%). The export of proteins across the cytoplasmic membrane is mediated by the Sec-dependent secretion pathway and twin-arginine transport system (complete set of genes present). Secretion of proteins across the outer membrane might also be mediated by the type II secretion system, since several components are found, i.e., *gspDEFG* and possibly *gspJ*, but not *gspCHKLMO*.

Proteins with cell surface adhesion domains could be involved in polymer breakdown by *Bacteroidetes*. The MED134 genome contains 41 proteins with at least one cell-cell or cell-surface interaction domain. Some of these adhesion domains are found in the predicted peptidases and glycosyl hydrolases, suggesting that attachment to the polymeric substrate and degradation are tightly linked. Proteins larger than 2,000 amino acids are likely to participate in attachment or gliding motility in *Bacteroidetes* (50). Notably, large proteins (13 of those predicted) covered 3.2% of the MED134 genome. These contain numerous internal repeats, adhesion domains or Ca^{2+}-binding motifs that are known to participate in attachment to surfaces. They are predicted to be extracellular, and their isoelectric point is acidic (2.2 to 4.2), which is consistent with protein stabilization and adaptation to salinity (average isoelectric point for the MED134 proteome was 6.8) (39). Two of them contain peptidase domains and therefore are likely to be involved in the degradation of extracellular peptides for carbon and energy. In comparison, the environmental genome MS024-2A contains six of these large proteins, whose genes cover 2.9% of its genome. This is an unusually high proportion for bacteria in general, but our analysis revealed that it is characteristic of other *Bacteroidetes*, including the marine species.

Exopolysaccharides produced by MED134 can also be expected to be involved in nonspecific attachment to particles; the genome encodes 44 predicted glycosyl transferases, of which 15 belong to family 1, 27 belong to family 2, and 1 belongs to the exostosin family. In addition, 12 proteins are predicted to encode polysaccharide biosynthesis and export proteins. These genes were arranged in clusters along with enzymes for the Wzy-dependent pathway for polysaccharide biosynthesis. Three such clusters span MED134_05924 to MED134_06039, MED134_09366 to MED134_09426, and MED134_13546 to MED134_13686. Genome analysis suggests that both receptor recognition and biofilm formation are involved in binding to surfaces.

The MED134 genome contains 106 predicted peptidases (3.5% of the proteins), although some of these are likely to be involved in processes such as internal protein turnover or peptidoglycan biosynthesis. A search for peptidases in 1,380 prokaryotic genomes in GenBank revealed that MED134 ranked at position 19 among these genomes, with the highest fraction of peptidases per total number of proteins (the average fraction of peptidases in all GenBank genomes was 2.3%). A total of 33% of the peptidases are predicted to be cytoplasmic, while the remainder contain a signal peptide, transmembrane domains, or a lipoprotein signal. Another bacterium with a proportion of peptidases as high as MED134 is *Idiomarina loihiensis*, which is known for its amino-acid-based metabolism (28). As expected, MED134 readily hydrolyzed gelatin as assayed as described by Baumann and Baumann (5). The environmental flavobacteria genomes MS024-2A and MS024-3C also contain a similar fraction of proteins predicted to be peptidases (3.6% in each case).

Examination of the MED134 genome identified 13 glycosyl hydrolase genes of diverse families. In addition to glycosyl hydrolases, MED134 contains 12 carbohydrate esterases responsible for hydrolysis of carbohydrate esters but lacks polysaccharide lyases. One protein with a sulfatase domain and a lipoprotein signal was also found in MED134, which may well be involved in the degradation of complex polysaccharides. Although not completely understood, the ability to degrade polysaccharides in the model *Bacteroidetes* species *Flavobacterium johnsoniae* and *Cytophaga hutchinsonii* is dependent on the ability to glide over surfaces (9, 12, 78). MED134 moves by gliding motility, although slowly on Zobell agar plates, and the gliding motility genes *glgABCDGHIJKLMO* were indeed identified.

A distinctive characteristic of well-studied *Bacteroidetes* is a system to degrade complex particulate organic matter in a manner where adhesion and degradation are tightly coupled for efficient utilization (2, 3). A potentially important set of proteins for degrading polysaccharides is the *sus* system, where the *susC* and *susD* elements are key components. SusC is a type of TonB-dependent receptor, which forms a channel through the outer membrane and SusD interacts with SusC and the polysaccharide. MED134 contains 31 *susC* homologs, with four of them located in tandem with *susD* homologs. It has been suggested that *susC* that are not in tandem with *susD* also have a function in sugar scavenging (7). In addition, TonB-dependent transporters were recently found to be abundant in oceanic bacterial communities, and it was put forward that energy for nutrient uptake through such transporters might be provided by the H$^+$-pumping activity of PR exposed to light (49). The strategy of *Dokdonia* sp. MED134 and other *Bacteroidetes* to efficiently utilize the carbon sources and harness light energy to take up scarce nutrients could thus be a widespread mechanism for nutrient acquisition in the ocean.

Self-transmissible elements. Horizontal gene transfer contributes to the adaptation and survival of prokaryotes in the natural environment. Transferred genes are frequently aggregated into genomic islands, some of which might be essential for the host. Three possible islands were found in the genome of MED134 (Fig. 1), one of which spans MED134_03214 to MED134_03314 and contains characteristics of a selfish element that avoids its loss through a toxin-antitoxin system. The second island (MED134_08411 to MED134_08546) includes the only transposase gene in the genome. It is of the IS*1* type and is next to type II secretion system genes that are likely to have been acquired by lateral gene transfer, since the closest BLAST hits are to the fish pathogen *Flavobacterium psychrophilum* JIP02/86, the alga killer *Kordia algicida* OT-1, and the flavobacterium ALC-1. The third and largest island spans 55 kbp in MED134 (MED134_06859 to MED134_07079) and contains genes present in conjugative transposons responsible for antibiotic resistance in pathogenic *Bacteroides* (see Table S3 in the supplemental material) (67). Conjugative transposons mediate their own transfer but also mobilize other genetic elements (63). This island contains genes involved in transposon maintenance and mobilization but appears to have lost most of the genes in the *tra* operon necessary for conjugative transfer. It also contains eight restriction/modification genes, which should prevent losing this region (loss would result in cell death). Our finding of the putative conjugative transposon shows that this mechanism for exchanging genetic information is present not only in *Bacteroides* spp. but also might be involved in tailoring the genomes of marine flavobacteria.

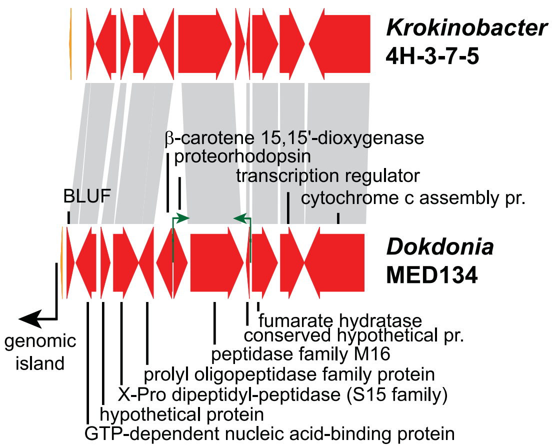

FIG. 3. PR gene neighborhood in the genome of *Dokdonia* sp. MED134. Gene labels show gene products. tRNAs are commonly involved in the integration of genomic islands (yellow). The BLUF domain detects blue light and is right next to the beginning of the genomic island that spans MED134_06859 to MED134_07079 and contains conjugative transposon genes and restriction and modification genes. Although the genomic island lacks the genes present in active conjugative transposons in *Bacteroides* spp., restriction and modification genes prevent the loss of this region. As in all other known *Bacteroidetes*, *prd* (PR) and *blh* (β-carotene 15,15′-dioxygenase) are in opposite directions. At the 3′ end of *prd*, a rho-independent transcriptional terminator was found ($\Delta G = -20.5$ kcal/mol), making it a monocistronic operon regulated at the transcriptional level. RpoD promoter consensus sequences are indicated by green arrows. The BLUF domain protein and GTP-dependent nucleic acid-binding protein genes are syntenic in bacteria. X-Pro dipeptidyl-peptidase (S15 family) is predicted to be noncytoplasmatic, since it contains a signal peptide. The prolyl oligopeptidase protein family contains one transmembrane domain, and the M16 family peptidase is predicted to be a lipoprotein. This suggests that these three types of peptidases are involved in the degradation of peptides for growth. The transcription regulator of the LuxR/FixJ family contains seven transmembrane domains and a 7TMR-DISM domain, which is found in cell surface receptors (1) and is therefore likely to respond to extracellular stimuli. Gene arrangement in *Krokinobacter* sp. 4H-3-7-5 shows the position of insertion of *blh* and *prd*.

MED134 *prd*, *blh*, and a blue light sensor gene are right next to the conjugative transposon remnant (Fig. 3). The PR genes in MED134, *Krokinobacter* 4H-3-7-5, and *Dokdonia* sp. PRO95 (60) have a different origin since the percent identity between MED134 and 4H-3-7-5 is 73% and that between 4H-3-7-5 and PRO95 is 97%. However, the 16S rRNA gene sequence identity among the three is 97.8% or higher (Fig. 4). This suggests lateral gene transfer among flavobacteria. In addition, the PR genes in MED134 and 4H-3-7-5 are located in different loci (Fig. 5). Any flavobacterial genome that harbors genes for the synthesis of carotenoids might become a phototroph by one single event involving the syntenic genes *blh* and *prd*.

Sigma factors. The primary sigma factor in bacteria (RpoD or σ^{70}), which binds to conserved sequences of strong promoters, plays a critical role in regulating the initiation of transcription and thereby also the expression of housekeeping genes such as those involved in transcription and translation during active growth. Somewhat unusually, MED134 encodes not one, but two RpoD homologs. The first homolog, MED134_12871,

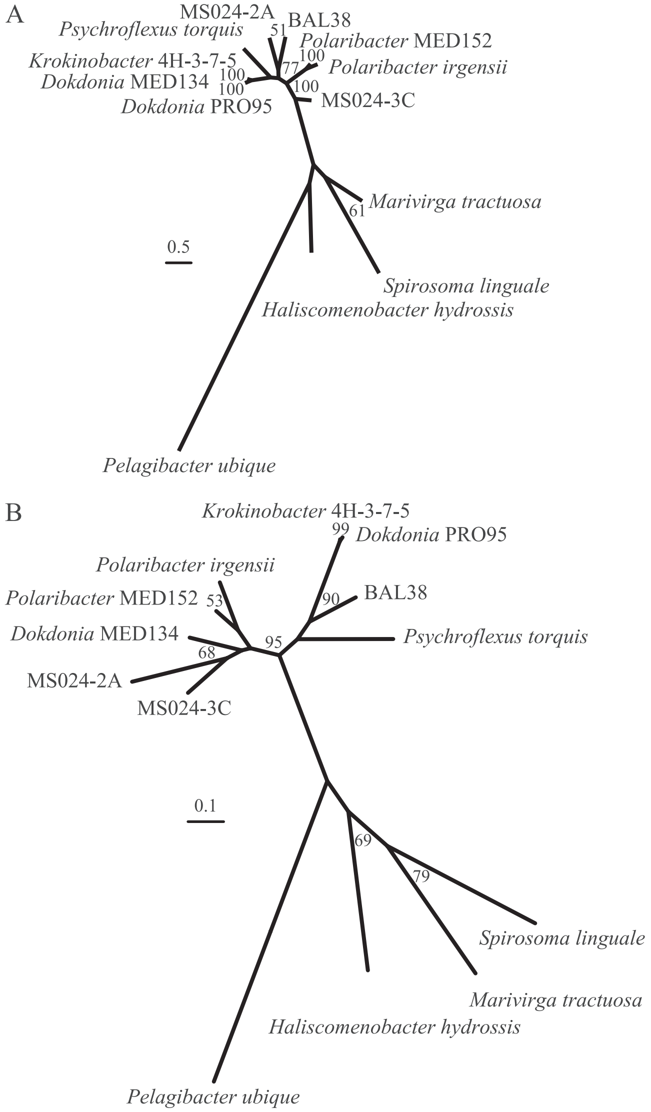

FIG. 4. Phylogenetic analyses of 16S rRNA gene and PR amino acid sequences. Maximum-likelihood trees of the 16S rRNA gene (A) and PR amino acid sequences (B) are shown. The trees show representatives of the phylum *Bacteroidetes* with *P. ubique* HTCC1062 as an outgroup. The divergent phylogeny of the 16S rRNA gene and PR protein sequences in *Dokdonia*/*Krokinobacter* suggests that the PR gene was acquired via lateral gene transfer. The alignment was generated using CLUSTAL W2 (40) and then further edited with Gblocks (11) to eliminate highly diverged regions. The maximum-likelihood tree was inferred with RAxML (68) using the WAG substitution matrix (implemented as "PROTCATWAGF") in the case of protein sequences and GTR substitution matrix (implemented as "GTRGAMMA") in 16S rRNA sequences. Numbers at nodes are bootstrap values greater than 50 (100 replicates). The scale bar indicates substitutions per site.

shares an amino acid sequence similarity as high as 93% with the RpoD in the genome of *F. johnsoniae* UW101 and, since key amino acids responsible for the recognition of the promoter sequence are also conserved, it is most likely to recognize the same −33/−7 element consensus promoter sequence as in *F. johnsoniae* UW101 (46). In *F. johnsoniae*, the conserved promoter sequence is found 109 times, in most cases upstream of housekeeping genes, but also upstream of genes encoding TonB receptors, ABC-type transporters, and genes related to polysaccharide degradation. Correspondingly, in MED134, the same promoter consensus sequence was found upstream of 75 genes, some of which were the expected housekeeping genes (see Table S4 in the supplemental material). Notably, however, the consensus promoter sequence was also found upstream of the PR gene (containing also the expected AT-rich regions within and upstream of the −33/−7 element consensus sequence in strong flavobacterial promoters [13, 17]). Further analysis showed that other PR-containing flavobacteria with the *F. johnsoniae* RpoD homolog (Fig. 6) had the promoter sequence upstream of carotenoid genes, which are necessary precursors for synthesis of the PR cofactor, retinal. The presence of the primary sigma factor for regulating the expression of the PR gene strongly suggests that obtaining energy through PR phototrophy is a central component in the physiology/metabolism of strain MED134 and could potentially also contribute to the observed strong ability of MED134 to utilize light for improving growth (23).

The second RpoD copy, MED134_05474, has a 69% amino acid sequence similarity to the *F. johnsoniae* RpoD (Fig. 6) and has some differences in key amino acids that recognize the promoter sequence, which precludes identifying what genes this RpoD copy regulates. Still, the mere fact that MED134 has two copies of the primary sigma factor is remarkable. The presence of several copies of RpoD is a characteristic of marine *Bacteroidetes* with large genomes, e.g., *Zunongwangia profunda* (two copies; 5.1 Mbp) and BAL39 (three copies; 5.8 Mbp), but not all. For example, *M. marina* only contains one copy in 9.8 Mbp. Indeed, multiples copies of RpoD, usually a single-copy gene, is typical of marine bacteria with complex developmental phases, such as cyanobacteria, myxococcales, planctomycetes, or actinobacteria. Moreover, MED134 contains the alternative σ factors RpoN (σ^{54} [47]) and 15 extracytoplasmic function family σ factors. Alternative σ factors recognize different sets of promoter sequences compared to primary σ factors and are thus important for determining which particular set of genes is expressed under particular environmental conditions. A search of the MED134 genome for the RpoN promoter consensus sequence revealed 31 matches (see Table S5 in the supplemental material). The list of genes potentially regulated by RpoN includes those encoding RecA, GrpE, GroES, GTP pyrophosphokinase, heat shock protein HtpG, transporters, and a tetrahydrofolate biosynthesis enzyme, as well as RpoN itself. This suggests that it is involved in survival under stress conditions. In summary, the presence of two different RpoD copies in MED134, which has a relatively small genome and few paralogs, together with the alternative sigma factors, could be a way to obtain metabolic flexibility and successfully manage changes in growth conditions, such as transient supplies of nutrients (e.g., following algal blooms) in otherwise challenging oligotrophic/low-nutrient marine environments.

Signal transduction and regulation of gene expression. The *Dokdonia* genome tentatively encodes 67 signal transduction proteins (2.23% of its proteins), similarly to other *Bacteroidetes*. Signal transduction proteins in MED134 include 57 two-component systems, eight HD-type phosphohydrolases and two guanylate/adenylate cyclases. The transcriptional reg-

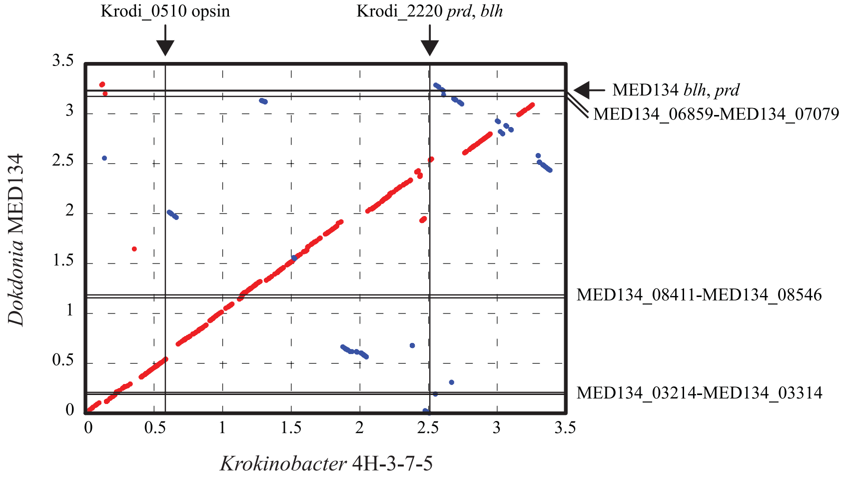

FIG. 5. Global alignment of *Dokdonia* sp. MED134 and *Krokinobacter* sp. 4H-3-7-5 as visualized by Mummerplot. Blue indicates a reverse complement match. Positions of the genomic islands detected in MED134 are indicated on the right. 4H-3-7-5 (Krodi_2220) and MED134 PR genes are not located at the same positions in their genomes. MED134 and 4H-3-7-5 share 2452 proteins based on BLASTP reciprocal best hits (82% of proteins in MED134; minimum percent identity set at 30, minimum coverage at 70% and maximum E-value at 1).

ulators in the genome made up 3.6% of its proteins, a proportion close to the average for marine *Bacteroidetes*.

MED134 would be expected to encode proteins to regulate gene expression in response to light. The role of phytochromes is well known in photosynthetic organisms. They respond to the red or far-red region of the visible light spectrum at the beginning of a regulatory cascade. MED134 has one phytochrome gene next to a two-component signal transduction system with PAS domains, which might also respond to light, and its cognate response regulator. These light regulators are right next to the RNA polymerase genes *rpoBC* and the ribosomal protein genes *rplKAJL* (Fig. 7), which form a conserved superoperon in bacteria (54). A search of GenBank genomes showed that regulatory genes in its proximity are common in prokaryotes. Adjacent genes in genomes, whose expressions are not obviously connected or are not part of an operon, can still be functionally linked (41–42). Therefore, finding a regulatory gene that responds to light in the phototrophic MED134 indicates that light could affect the expression of genes in the neighborhood. *prd* and *blh* in both environmental genomes MS024-2A and MS024-3C are also next to similar housekeeping genes, in this case within a superoperon that contains the RNA polymerase subunit gene *rpoA*, ribosomal proteins and additional genes known to be the largest superoperon in bacteria (62). Since the phytochrome genes in MED134 and the PR genes in the environmental genomes are next to highly expressed genes, this proximity strongly implies that light is an important factor in the lifestyle of these organisms.

Besides phytochromes, another candidate for gene regulation in response to light is TspO. This regulatory protein is involved in the expression of photosynthetic genes in organisms such as *Rhodobacter* or *Synechococcus*. It is an integral membrane protein found in all domains of life, including humans, although its function is enigmatic. It has five predicted transmembrane segments and regulates photosynthesis gene expression, although is also involved in nutrient stress in *Sinorhizobium meliloti*. It binds benzodiazepin, tetrapyrrol, or steroid-type molecules as part of its regulatory role but the regulation cascade is not known. The role of this domain in nonphotosynthetic organisms is not known. In MED134 *tspO* is next to three genes in the mevalonate pathway for isopentenyl diphosphate biosynthesis, a precursor of carotenoids, including retinal. There is synteny for the same pathway in the two environmental genomes MS024-2A and MS024-3C. In *Polaribacter irgensii* 23-P, *tspO* is not next to carotenoid genes but within a cryptochrome/photolyase gene cluster. TspO is thus likely to have a regulatory role that involves light response by different types of genes in marine *Bacteroidetes*.

MED134 also contains three types of cryptochromes/photolyases: the animal cryptochrome and (6-4) photolyase family, the DASH family, and the deoxyribodipyrimidine photo-lyase class I. However, ORFs annotated as cryptochromes in bacterial genomes are not likely to be the beginning of a regulatory mechanism in response to light, since this function has not been proven (73). Instead, these proteins most probably function as photolyases for DNA repair after UV damage, by using a photon of blue light to catalyze the reaction. This type of repair mechanism is frequently found in genomes of surface seawater prokaryotes (66). Finally, a BLUF domain that detects blue light is situated in a gene that is five genes away from *prd* and *blh* (Fig. 3). The proximity of *prd* and the BLUF domain gene might also indicate similar functional and regulatory contexts. Overall, the genome analysis indicates a number of mechanisms for responding to extracellular stimuli, light being a special case. Although the physiological responses that light induces are not yet known, as expected, light sensing and phototrophy seem to go hand in hand as we elucidate the lifestyle of MED134.

Predicted highly expressed genes. The prediction of such genes is an approach to investigate gene expression and regulation in specialized physiological groups of microorganisms. Of the ORFs expected to encode proteins larger than 100 amino acids, a total of 118 genes (4.4% of ORFs) in MED134 were PHX (see Table S6 in the supplemental material). This proportion is low compared to most previously analyzed genomes. Such a low proportion is only

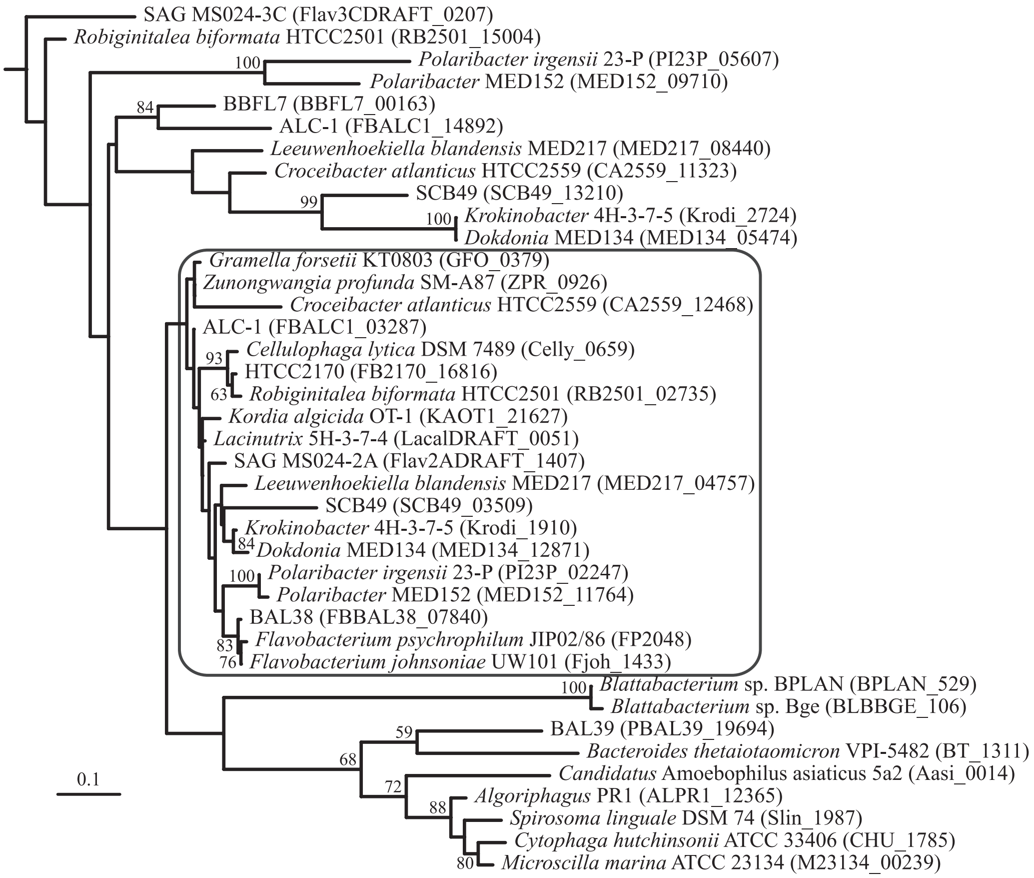

FIG. 6. Maximum likelihood tree with the RpoD homologs from representative *Bacteroidetes* sequences. Box indicates *F. johnsoniae* RpoD homologs that are likely to recognize the same −7/−33 element consensus sequences (TANNTTTG/TTTG) separated by a spacer of variable length, generally 19 to 21 nucleotides. Key amino acids that interact with the promoter sequence are conserved in the *F. johnsoniae* RpoD homologs. A greater variability is found in the RpoD sequences outside of the boxed area. Locus tags are shown between parentheses. The alignment was generated using CLUSTAL W2 (40) and edited with Gblocks (11) to eliminate highly diverged regions. The maximum-likelihood tree was inferred with RAxML (68) using the WAG substitution matrix. The sequence of *Cyanobium* sp. strain PCC 7001 served as outgroup. The numbers at the nodes are bootstrap values greater than 50 for 100 replicates. The scale bar indicates substitutions per site.

found in large genomes (31). For fast-growing bacteria, such as *E. coli*, a larger proportion of genes are PHX, including genes involved in glycolysis and TCA cycle genes (31–32). MED134 generation time is ~5 h in artificial seawater amended with yeast extract and peptone to 0.34 mM final C concentration at 21°C as in Gómez-Consarnau et al. (23). This is 8-fold faster than that reported for *Pelagibacter ubique* but slower than fast-growing species commonly used in the laboratory in rich media. The presence of six PHX genes in the TCA cycle of MED134, a slow-growing bacterium, may be because PR-containing bacteria depend on the TCA cycle not only for ATP production but also as a major source of precursors for carbon anabolism. The overall PHX pattern shows features that are found in both fast- and slow-growing bacteria. For example, only 27 ribosomal proteins were PHX, which is typical of slow-growing organisms, considerably lower than the average of 38 for bacterial genomes (33). However, genes that encode the RNA polymerase subunits RpoB and RpoC are PHX in MED134, which is a pattern typical of fast-growing organisms.

The MED134 genome contains three PHX genes for the uptake of ammonia: Glu dehydrogenase (MED134_09951) and Glu-ammonia ligase types II (MED134_03484) and III (MED134_03489). As many as 15 transport genes are PHX. Three *susC* and two *susD* genes are PHX. MED134_05219 (*susC*) and MED134_05214 (*susD*) are next to each other and are surrounded by a number of peptidase but not glycosyl hydrolase genes, suggesting that this *sus* system is involved in the efficient degradation of peptides instead of polysaccharides. The preprotein translocase, YajC subunit, and the protein-export membrane protein SecD genes were PHX, indicating that secretion plays a major role in the MED134 lifestyle.

Transporters. As was previously described for another PR-containing flavobacterium (25), a relatively low number of proteins seem to be involved in transport activity. A total of 109 ORFs were annotated as transporter related genes (3.6%; see Table S7 in the supplemental material), whereas the average percentage for the bacteria in the TransportDB database is 5.9%. Bacteria in the TransportDB database with ~100 transporter genes or fewer were obligate symbionts, pathogens, or autotrophs with reduced genome sizes. A large proportion of the transporters are secondary transporters (63 proteins) and ATP-binding cassette type (47 proteins). Transporters for which the substrate is predictable include those for sugars,

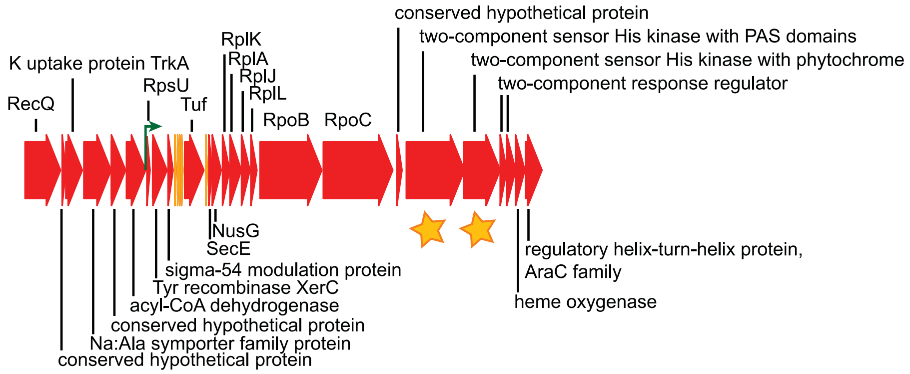

FIG. 7. Phytochrome gene neighborhood in the genome of MED134. Gene labels show gene products. *rps* and *rpl* encode ribosomal proteins, and *rpoBC* encode the RNA polymerase subunits β and β′, respectively. Genes such as *nusG*, *tuf*, and *secE* and tRNAs genes are frequently found to be associated with RNA polymerase and ribosomal protein operons in bacteria. The first two-component sensor contains the following domains: CheB methylesterase domain; the CheR methyltransferase, all-alpha domain; CheR methyltransferase, SAM binding domain; two PAS domains; His kinase A (phosphoacceptor) domain; and His kinase-like ATPase. The second two-component sensor contains the domains: PAS_2 domain, GAF domain, phytochrome, His kinase A (phosphoacceptor) domain, and His kinase-like ATPase. tRNAs are indicated in yellow. The RpoD promoter consensus sequence is indicated by a green arrow. Stars indicate light sensors. The heme oxygenase and phytochrome genes are syntenic in *Bacteria*. Heme oxygenase releases Fe^{2+} and biliverdin from heme and biliverdin becomes the prosthetic group in the phytochrome.

amino acids, peptides, nucleosides, nicotinamide mononucleotide, metals, and other ions, as well as drug efflux proteins.

TonB-dependent transporters are involved in absorption and concentration in the periplasm of substrates that are either poorly permeable through the porin channels or encountered at very low concentrations. Two of the TonB-dependent transporters seem to be involved in the uptake of vitamins. Cobalamin and TPP riboswitches target MED134_04109 and MED134_10161, respectively, both TonB-dependent transporter genes, which indicates that their gene products are regulated at the translational level by the amount of the vitamins cobalamin and thiamine, for which the biosynthetic pathways could not be reconstructed. MED134_10161 contains both the consensus promoter sequences for RpoD and RpoN, which indicates that its expression is important under different growth conditions. The other components that interact with TonB-dependent receptors and transmit energy from the H^+ motive force, TonB and ExbBD, are also present in the genome.

Iron is an essential nutrient and its acquisition is difficult due to low solubility of Fe^{3+} in aerobic environments. Consequently, it is scarce in the marine environment or of limited bioavailability. MED134 contains a number of mechanisms to make use of different forms of this element. Siderophore synthetic genes were not found but iron can enter the cell through an ATP-binding cassette-type transport system or NRAMP family type secondary transporter. In addition, two ORFs encode ferritin-like proteins for storage of iron and avoiding oxidative stress. Iron metabolism is regulated by three Fur proteins, one repressor, and FecR.

Heme is abundant in phytoplankton biomass, and bacteria living in association with algal cells or their detritus have acquired mechanisms to recycle iron in the form of heme (76). A TonB-dependent outer membrane receptor has a high percent identity with the heme uptake system of *M. marina*, an organism known to take up heme (27). Once it is concentrated in the periplasmic space, an ATP-binding cassette-type transport system specific for heme carries it to the cytoplasm. Heme can then be incorporated as it is or a heme oxygenase (MED134_02455) removes iron for other cell needs. *Dokdonia* sp. MED134 also contains a membrane attack complex considered to be a virulence factor in *Bacteroides* (79), which might participate in the breakdown of algal cells for the release of nutrients (Fig. 8). Mechanisms for the breakdown of eukaryotic cells and heme utilization in MED134 add further evidence of a close interaction between this flavobacterium and algal cells such as those in algal blooms, which may thus represent sporadic but at times abundant sources of not just biopolymers for growth but also iron.

Conclusions. Analysis of the MED134 genome revealed a wide spectrum of genomic adaptations and their putative ecological functions for marine *Bacteroidetes*. In addition to the genes directly involved in PR phototrophy (e.g., *prd* and *blh*), the analysis suggested that PR gene promoter strength, main metabolic pathways, and signal transduction mechanisms involving light, might all contribute to a positive response of MED134 when exposed to light. The relatively high number of peptidases suggests a preference for peptides over polysaccha-

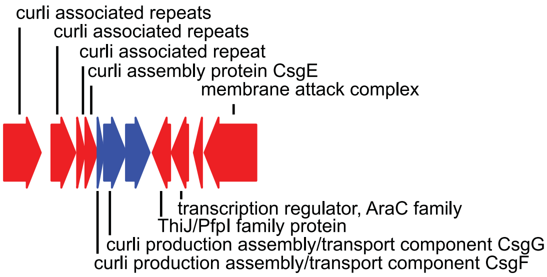

FIG. 8. Putative gene cluster involved in the attachment and breakdown of eukaryotic cells. Curli-associated repeat proteins and CsgEFG are involved in the assemblage of fibers for adhesion to surfaces and cells and biofilm formation. Membrane-attack complex/perforin protein is a transmembrane protein that forms a pore in the eukaryotic cell membranes described in pathogenic *Bacteroides*. The product of the cluster could be involved in the attachment and lysis of algal cells. Blue indicates predicted lipoproteins.

rides for growth. The main metabolic pathways are adapted to changes in the carbon flux when amino acids are converted in TCA intermediates or light is used as an alternative energy source. Interestingly, conjugative transposon genes were detected, indicating that these genetic elements shape the genomes of marine *Bacteroidetes*. Indeed, phylogenetic analysis of the PR gene suggested transfer events among flavobacteria. Apparently, their genomes are prone to take up the PR gene to improve their fitness. However, genome analysis also suggested that the genetic repertoire of the recipient bacterium is likely to determine the success of the new gene. The genomic analysis revealed unexpected and novel genetic features of PR-containing marine *Bacteroidetes* and should facilitate further research on light utilization mechanisms.

ACKNOWLEDGMENTS

J.M.G. and C.P.-A. were supported by grant CTM2010-11060-E from the Spanish Ministry of Science and Innovation, J.P. was supported by the Swedish Research Council and FORMAS, and P.P. was supported by the intramural funds of the U.S. Department of Health and Human Services (National Library of Medicine, National Institutes of Health).

REFERENCES

1. **Anantharaman, V., and L. Aravind.** 2003. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. BMC Genomics **4:**34.
2. **Anderson, K. L., and A. A. Salyers.** 1989. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch binding sites and periplasmic starch-degrading enzymes. J. Bacteriol. **171:** 3192–3198.
3. **Anderson, K. L., and A. A. Salyers.** 1989. Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. J. Bacteriol. **171:**3199–3204.
4. **Bauer, M. M., et al.** 2006. Whole genome analysis of the marine Bacteroidetes "*Gramella forsetii*" reveals adaptations to degradation of polymeric organic matter. Environ. Microbiol. **8:**2201–2213.
5. **Baumann, P., and L. Baumann.** 1981. The marine Gram-negative eubacteria: genera *Photobacterium*, *Beneckea*, *Alteromonas*, *Pseudomonas*, and *Alcaligenes*, p. 1302–1331. *In* M. P. Starr, H. Stolp, H. G. Trüper, A. Balows, and H. G. Schlegel (ed.), The prokaryotes. Springer-Verlag, Berlin, Germany.
6. **Béjà, O., et al.** 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289:**1902–1906.
7. **Blanvillain, S., et al.** 2007. Plant carbohydrate scavenging through TonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS One **2:**e224.
8. **Bratlie, M. S., J. Johansen, and F. Drabløs.** 2010. Relationship between operon preference and functional properties of persistent genes in bacterial genomes. BMC Genomics **11:**71.
9. **Braun, T. F., M. K. Khubbar, D. A. Saffarini, and M. J. McBride.** 2005. *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. J. Bacteriol. **187:**6943–6952.
10. **Cantarel, B. L., et al.** 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res. **37:**D233–238.
11. **Castresana, J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. **17:**540–552.
12. **Chang, L. E., J. L. Pate, and R. J. Betzig.** 1984. Isolation and characterization of nonspreading mutants of the gliding bacterium *Cytophaga johnsoniae*. J. Bacteriol. **159:**26–35.
13. **Chen, S., M. Bagdasarian, M. G. Kaufman, A. K. Bates, and E. D. Walker.** 2007. Mutational analysis of the *ompA* promoter from *Flavobacterium johnsoniae*. J. Bacteriol. **189:**5108–5118.
14. **DeLong, E., D. G. Franks, and A. L. Alldredge.** 1993. Phylogenetic diversity of aggregate-attached versus free-living marine bacterial assemblages. Limnol. Oceanogr. **38:**924–934.
15. **Dulebohn, D., J. Choy, T. Sundermeier, N. Okan, and A. W. Karzai.** 2007. Trans-translation: the tmRNA-mediated surveillance mechanism for ribosome rescue, directed protein degradation, and nonstop mRNA decay. Biochemistry **46:**4681–4693.
16. **Eddy, S. R.** 2008. A probabilistic model of local sequence alignment that simplifies statistical significance estimation. PLoS Comput. Biol. **4:**e1000069.
17. **Estrem, S. T., T. Gaal, W. Ross, and R. L. Gourse.** 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. U. S. A. **95:**9761–9766.
18. **Fuhrman, J. A., M. S. Schwalbach, and U. Stingl.** 2008. Proteorhodopsins: an array of physiological roles? Nat. Rev. Microbiol. **6:**488–494.
19. **Ganesan, H., A. S. Rakitianskaia, C. F. Davenport, B. Tümmler, and O. N. Reva.** 2008. The SeqWord Genome Browser: an online tool for the identification and visualization of atypical regions of bacterial genomes through oligonucleotide usage. BMC Bioinformatics **9:**333.
20. **Gao, F., and C. T. Zhang.** 2006. GC-Profile: a web-based tool for visualizing and analyzing the variation of GC content in genomic sequences, Nucleic Acids Res. **34:**W686–W691.
21. **Giovannoni, S. J., et al.** 2005. Proteorhodopsin in the ubiquitous marine bacterium SAR11. Nature **438:**82–85.
22. **Goldberg, S. M. D., et al.** 2006. A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc. Natl. Acad. Sci. U. S. A. **103:**11240–11245.
23. **Gómez-Consarnau, L., et al.** 2007. Light stimulates growth of proteorhodopsin-containing marine *Flavobacteria*. Nature **445:**210–213.
24. **Gómez-Consarnau, L., et al.** 2010. Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. PLoS Biol. **8:**e1000358.
25. **González, J. M., et al.** 2008. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (*Flavobacteria*). Proc. Natl. Acad. Sci. U. S. A. **105:**8724–8729.
26. **Häse, C. C., N. D. Fedorova, M. Y. Galperin, and P. A. Dibrov.** 2001. Sodium ion cycle in bacterial pathogens: evidence from cross-genome comparisons. Microbiol. Mol. Biol. Rev. **65:**353–370.
27. **Hopkinson, B. M., K. L. Roe, and K. A. Barbeau.** 2008. Heme uptake by *Microscilla marina* and evidence for heme uptake systems in the genomes of diverse marine bacteria. Appl. Environ. Microbiol. **74:**6263–6270.
28. **Hou, S., et al.** 2004. Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. Proc. Natl. Acad. Sci. U. S. A. **101:**18036–18041.
29. **Johnson, E. T., et al.** 2010. Enhancement of survival and electricity production in an engineered bacterium by light-driven proton pumping. Appl. Environ. Microbiol. **76:**4123–4129.
30. **Juncker, A. S., et al.** 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. **12:**1652–1662.
31. **Karlin, S., and J. Mrázek.** 2000. Predicted highly expressed genes of diverse prokaryotic genomes. J. Bacteriol. **182:**5238–5250.
32. **Karlin, S., J. Mrázek, A. Campbell, and D. Kaiser.** 2001. Characterizations of highly expressed genes of four fast-growing bacteria. J. Bacteriol. **183:** 5025–5040.
33. **Karlin, S., L. Brocchieri, A. Campbell, M. Cyert, and J. Mrázek.** 2005. Genomic and proteomic comparisons between bacterial and archaeal genomes and related comparisons with the yeast and fly genomes. Proc. Natl. Acad. Sci. U. S. A. **102:**7309–7314.
34. **Kimura, H., C. R. Young, A. Martinez, and E. F. DeLong.** 2011. Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. ISME J., in press.
35. **Klippel, B., et al.** 2011. Complete genome sequences of *Krokinobacter* sp. strain 4H-3-7-5 and *Lacinutrix* sp. strain 5H-3-7-4, polysaccharide-degrading members of the family *Flavobacteriaceae*. J. Bacteriol. **193:**4545–4546.
36. **Kogure, K.** 1998. Bioenergetics of marine bacteria. Curr. Opin. Biotechnol. **9:**278–282.
37. **Kurtz, S., et al.** 2004. Versatile and open software for comparing large genomes. Genome Biol. **5:**R12.
38. **Lami, R., M. T. Cottrell, B. J. Campbell, and D. L. Kirchman.** 2009. Light-dependent growth and proteorhodopsin expression by *Flavobacteria* and SAR11 in experiments with Delaware coastal water. Environ. Microbiol. **11:**3201–3209.
39. **Lanyi, J. K.** 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. Bacteriol. Rev. **38:**272–290.
40. **Larkin, M. A., et al.** 2007. CLUSTAL W and CLUSTAL X version 2.0. Bioinformatics **23:**2947–2948.
41. **Lathe, W. C., B. Snel, and P. Bork.** 2000. Gene context conservation of a higher order than operons. Trends Biochem. Sci. **25:**474–479.
42. **Lawrence, J. G.** 2003. Gene organization: selection, selfishness, and serendipity. Annu. Rev. Microbiol. **57:**419–440.
43. **Martinez, A., A. S. Bradley, J. R. Waldbauer, R. E. Summons, and E. F. DeLong.** 2007. Proteorhodopsin photosystem gene expression enables photophosphorylation in a heterologous host. Proc. Natl. Acad. Sci. U. S. A. **104:**5590–5595.
44. **McBride, M. J.** 2001. Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. Annu. Rev. Microbiol. **55:**49–75.
45. **McBride, M. J.** 2004. *Cytophaga-Flavobacterium* gliding motility. J. Mol. Microbiol. Biotechnol. **7:**63–71.
46. **McBride, M. J., et al.** 2009. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. Appl. Environ. Microbiol. **75:**6864–6875.
47. **Merrick, M. J.** 1993. In a class of its own–the RNA polymerase sigma factor sigma 54 (sigma N). Mol. Microbiol. **10:**903–909.

48. **Meyer, F., et al.** 2003. GenDB: an open source genome annotation system for prokaryote genomes. Nucleic Acids Res. **31:**2187–2195.
49. **Morris, R. M., et al.** 2010. Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. ISME J. **4:**673–685.
50. **Nelson, S. S., S. Bollampalli, and M. J. McBride.** 2008. SprB is a cell surface component of the *Flavobacterium johnsoniae* gliding motility machinery. J. Bacteriol. **190:**2851–2857.
51. **Perriere, G., and J. Thioulouse.** 2002. Use and misuse of correspondence analysis in codon usage studies. Nucleic Acids Res. **30:**4548–4555.
52. **Pinhassi, J., et al.** 2004. Changes in bacterioplankton composition under different phytoplankton regimens. Appl. Environ. Microbiol. **70:**6753–6766.
53. **Ploug, H., M. Kühl, B. Buchholz-Cleven, and B. B. Jørgensen.** 1997. Anoxic aggregates-an ephemeral phenomenon in the pelagic environment? Aquat. Microb. Ecol. **13:**285–294.
54. **Post, L. E., G. D. Strycharz, M. Nomura, H. Lewis, and P. P. Dennis.** 1979. Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit P in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. **76:**1697–1701.
55. **Preisig, O., D. Anthamatten, and H. Hennecke.** 1993. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. Proc. Natl. Acad. Sci. U. S. A. **90:**3309–3313.
56. **Puigbò, P., A. Romeu, and S. Garcia-Vallvé.** 2008. HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. Nucleic Acids Res. **36:**D524–D527.
57. **Puigbò, P., E. Guzmán, A. Romeu, and S. Garcia-Vallvé.** 2007. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. **35:**W126–W131.
58. **Qin, Q.-L., et al.** 2010. The complete genome of *Zunongwangia profunda* SM-A87 reveals its adaptation to the deep-sea environment and ecological role in sedimentary organic nitrogen degradation. BMC Genomics **11:**247.
59. **Rawlings, N. D., A. J. Barrett, and A. Bateman.** 2010. MEROPS: the peptidase database. Nucleic Acids Res. **38:**D227–D233.
60. **Riedel, T., et al.** 2010. Constitutive expression of the proteorhodopsin gene by a flavobacterium strain representative of the proteorhodopsin-producing microbial community in the North Sea. Appl. Environ. Microbiol. **76:**3187–3197.
61. **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.
62. **Rogozin, I. B., et al.** 2002. Connected gene neighborhoods in prokaryotic genomes. Nucleic Acids Res. **30:**2212–2223.
63. **Salyers, A. A., N. B. Shoemaker, A. M. Stevens, and L. Y. Li.** 1995. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiol. Rev. **59:**579–590.
64. **Sharp, P. M., and W. H. Li.** 1987. The codon adaptation index: a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
65. **Simon, M., H. P. Grossart, B. Schweitzer, and H. Ploug.** 2002. Microbial ecology of organic aggregates in aquatic ecosystems. Aquat. Microb. Ecol. **28:**175–211.
66. **Singh, A. H., T. Doerks, I. Letunic, J. Raes, and P. Bork.** 2009. Discovering functional novelty in metagenomes: examples from light-mediated processes. J. Bacteriol. **191:**32–41.
67. **Smith, C. J., G. D. Tribble, and D. P. Bayley.** 1998. Genetic elements of *Bacteroides* species: a moving story. Plasmid **40:**12–29.
68. **Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688–2690.
69. **Steindler, L., M. S. Schwalbach, D. P. Smith, F. Chan, and S. J. Giovannoni.** 2011. Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS One **6:**e19725.
70. **Stepanauskas, R., and M. E. Sieracki.** 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc. Natl. Acad. Sci. U. S. A. **104:**9052–9057.
71. **Stingl, U., R. A. Desiderio, J. C. Cho, K. L. Vergin, and S. J. Giovannoni.** 2007. The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. Appl. Environ. Microbiol. **73:**2290–2296.
72. **Tang, K.-H., X. Feng, Y. J. Tang, and R. E. Blankenship.** 2009. Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* OCh114. PLoS One **4:**e7233.
73. **van der Horst, M. A., J. Key, and K. J. Hellingwerf.** 2007. Photosensing in chemotrophic, non-phototrophic bacteria: let there be light sensing too. Trends Microbiol. **15:**554–562.
74. **Venter, J. C., et al.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66–74.
75. **Waack, S., et al.** 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7:**142.
76. **Weaver, R. S., D. L. Kirchman, and D. A. Hutchins.** 2003. Utilization of iron/organic ligand complexes by marine bacterioplankton. Aquat. Microb. Ecol. **31:**227–239.
77. **Woyke, T., et al.** 2009. Assembling the marine metagenome, one cell at a time. PLoS One **4:**e5299.
78. **Xie, G., et al.** 2007. Genome sequence of the cellulolytic gliding bacterium *Cytophaga hutchinsonii*. Appl. Environ. Microbiol. **73:**3536–3546.
79. **Xu, Q., et al.** 2010. Structure of a membrane-attack complex/perforin (MACPF) family protein from the human gut symbiont *Bacteroides thetaiotaomicron*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **66:**1297–1305.
80. **Yu, N. Y., et al.** 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes, Bioinformatics **26:**1608–1615.
81. **Zhang, C.-T., F. Gao, and R. Zhang.** 2005. Segmentation algorithm for DNA sequences. Phys. Rev. E **72:**041917(1–6).
82. **Zubkov, M. V.** 2009. Photoheterotrophy in marine prokaryotes. J. Plankton Res. **31:**933–938.