2009

Novel Features of the Polysaccharide-Digesting Gliding Bacterium *Flavobacterium johnsoniae* as Revealed by Genome Sequence Analysis

Mark J. McBride

Gary Xie

*See next page for additional authors*

Follow this and additional works at: [https://digitalcommons.uri.edu/cmb_facpubs](https://digitalcommons.uri.edu/cmb_facpubs)

Terms of Use
All rights reserved under copyright.

Citation/Publisher Attribution


Available at: [http://dx.doi.org/10.1128/AEM.01495-09](http://dx.doi.org/10.1128/AEM.01495-09)
Authors
**Novel Features of the Polysaccharide-Digesting Gliding Bacterium Flavobacterium johnsoniae as Revealed by Genome Sequence Analysis†**

Mark J. McBride,1* Gary Xie,2,3 Eric C. Martens,4,5 Alla Lapidus,3 Bernard Henrissat,6 Ryan G. Rhodes,1 Eugene Goltsman,3 Wei Wang,7 Jian Xu,3,7 David W. Hunnicutt,8 Andrew M. Staroscik,7 Timothy R. Hoover,10 Yi-Qiang Cheng,1 and Jennifer L. Stein10

University of Wisconsin—Milwaukee, Milwaukee, Wisconsin; Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico; Department of Energy Joint Genome Institute, Walnut Creek, California; Washington University School of Medicine, St. Louis, Missouri; University of Michigan Medical School, Ann Arbor, Michigan; Architecture et Fonction des Macromolécules Biologiques, UMR 6098, Centre National de la Recherche Scientifique and Universities of Aix-Marseille I & II, Marseille, France; BioEnergy Genome Center, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences (CAS-QIBET), Qingdao, China; St. Norbert College, DePere, Wisconsin; University of Rhode Island, Kingston, Rhode Island; and University of Georgia, Athens, Georgia

Received 25 June 2009/Accepted 22 August 2009

The 6.10-Mb genome sequence of the aerobic chitin-digesting gliding bacterium Flavobacterium johnsoniae (phylum Bacteroidetes) is presented. F. johnsoniae is a model organism for studies of bacteroidete gliding motility, gene regulation, and biochemistry. The mechanism of F. johnsoniae gliding is novel, and genome analysis confirms that it does not involve well-studied motility organelles, such as flagella or type IV pilus. The motility machinery is composed of Gld proteins in the cell envelope that are thought to comprise the “motor” and SprB, which is thought to function as a cell surface adhesin that is propelled by the motor. Analysis of the genome identified genes related to sprB that may encode alternative adhesins used for movement over different surfaces. Comparative genome analysis revealed that some of the gld and spr genes are found in nongliding bacteroidetes and may encode components of a novel protein secretion system. F. johnsoniae digests proteins, and 125 predicted peptidases were identified. F. johnsoniae also digests numerous polysaccharides, and 138 glycose hydrolyses, 9 polysaccharide lyases, and 17 carbohydrate esterases were predicted. The unexpected ability of F. johnsoniae to digest hemicelluloses, such as xylans, mannans, and xyloglucans, was predicted based on the genome analysis and confirmed experimentally. Numerous predicted cell surface proteins related to Bacteroides thetaiotaomicron SusC and SusD, which are likely involved in binding of oligosaccharides and transport across the outer membrane, were also identified. Genes required for synthesis of the novel outer membrane flexirubin pigments were identified by a combination of genome analysis and genetic experiments. Genes predicted to encode components of a multienzyme nonribosomal peptide synthetase were identified, as were novel aspects of gene regulation. The availability of techniques for genetic manipulation allows rapid exploration of the features identified for the polysaccharide-digesting gliding bacteroidete F. johnsoniae.

Flavobacterium johnsoniae (formerly Cytophaga johnsonae) is a member of the large and diverse phylum of gram-negative bacteria known as the Bacteroidetes. Members of this group of organisms have a number of unique characteristics that distinguish them from other bacteria. Some have novel cell surface machinery to utilize polysaccharides (85, 95, 96). Rapid gliding motility over surfaces is also common among these bacteria (59), as are unusual outer membrane sulfonolipids (29) and flexirubin pigments (78). Bacteroidete gene expression and regulation also have novel aspects (10, 11, 20, 39, 92). The many unusual features of these common but understudied bacteria provide numerous avenues for further exploration, which can be greatly aided by analysis of genome sequences.

F. johnsoniae digests many polysaccharides and proteins, but it is best known for its ability to rapidly digest insoluble chitin (87). Chitin is one of the most abundant biopolymers on earth (63). F. johnsoniae and other members of the Bacteroidetes phylum are thought to play important roles in the turnover of this compound in many environments (47). F. johnsoniae has become a model system for the study of bacteroidete gliding motility biochemistry and molecular biology (20, 27–29, 59, 72). This paper highlights novel features of the F. johnsoniae genome, with particular emphasis on genes and proteins likely to be involved in polysaccharide utilization, gliding motility, and the novel biochemistry of this organism.

**MATERIALS AND METHODS**

**Sequencing of the F. johnsoniae genome.** The random shotgun method was used to sequence the genome of F. johnsoniae UW101 (ATCC 17061). Large-insert (40-kb), medium-insert (8-kb), and small-insert (3-kb) random libraries were partially sequenced, and sequences were assembled with parallel phraps.

---

* Corresponding author. Mailing address: Department of Biological Sciences, University of Wisconsin—Milwaukee, 3209 N. Maryland Ave., 181 Lapham Hall, Milwaukee, WI 53211. Phone: (414) 229-5844. Fax: (414) 229-3926. E-mail: mcbride@uwm.edu.
† Supplemental material for this article may be found at http://aem.asm.org/.
‡ Published ahead of print on 28 August 2009.
Colonies were tested for the presence of flexirubin pigments by exposing them to 50 μl 10 N KOH, which resulted in a change from yellow to red if flexirubin pigments were present, followed by neutralization with 42 μl 12 N HCl, which resulted in a return to yellow pigmentation.

**Nucleotide sequence accession number.** The genome sequence of *F. johnsoniae* has been deposited in the GenBank database under accession number CP000685.

**RESULTS AND DISCUSSION**

**General genome features.** The *F. johnsoniae* genome consists of a single circular, 6,096,872 bp chromosome with a G+C content of 34.11% (Fig. 1). It is one of the largest sequenced bacteriophage genomes and is more than twice as large as the genome of the other sequenced member of the genus *Flavobacterium*, the fish pathogen *Flavobacterium psychrophilum* (24). Six rRNA operons were identified, and 5,056 encoding genes were predicted. The GC skew allowed prediction of the site of the origin of replication near nucleotide 32322321. Alignment of conserved genomic sequences of *F. johnsoniae* and *F. psychrophilum* with MAUVE revealed a 4.17-Mbp region between 5.43 Mbp and 3.5 Mbp that contained most of the conserved sequences (see Fig. S1 in the supplemental material). Not surprisingly, this region harbors most of the genes involved in central energy-generating metabolism (glycolysis, tricarboxylic acid cycle, oxidative phosphorylation) and the core information-processing genes involved in transcription and translation (Fig. 1). The remaining 1.93 Mbp of the genome harbor several transposon islands and a large number of genes encoding hypothetical proteins of unknown function. This region is also enriched for genes predicted to be involved in polysaccharide utilization, and 45% of
Predicted polysaccharide lyases.

of polysaccharides as nutrients. Analysis of the genome of bacteria (Fjoh_4175, Fjoh_4555, Fjoh_4757), animals (Fjoh_4175), and analysis of the predicted glycoside hydrolases and 67% of the polysaccharide lyases are encoded by genes in this region. In addition to genes predicted to be involved in chitin utilization, F. johnsoniae also has genes predicted to encode numerous glycohydrolases, polysaccharide lyases, and esterases that are likely involved in the digestion of other polysaccharides (see Tables S1, S2, and S3 in the supplemental material). It appears to have an arsenal of enzymes for digestion of plant cell wall polysaccharides, which may explain the prevalence of this organism in soil and rhizosphere habitats (71, 87). F. johnsoniae is known to digest pectin, and glycohydrolases, lyases, and esterases likely to be involved in this process were identified (see Tables S1, S2, and S3 and Fig. S2 in the supplemental material). Many genes predicted to encode glycohydrolases and esterases that could be involved in utilization of plant cell wall hemicelluloses, such as xylans (β-1,4-linked polymers of xylose often substituted with acetyl, arabinofuranoside, and glucuronosyl residues), mannan (heteropolysaccharides containing β-1,4-linked mannose residues), and xylolglycans (β-1,4-linked polymers of glucose substituted with xylose and other sugars), were also identified (see Tables S1 and S3). Fjoh_6074 was originally described as a chitinolytic bacterium (87), and analysis of the genome sequence identified genes encoding possible chitinolytic enzymes. These enzymes include five chitinases with known function resulted in “distantly related to” status. Because the threshold of similarity that correlates with a change of substrate specificity is variable from one to another, the criteria were tightened or loosened appropriately for several families. All analyses were conducted domain by domain with a known function resulted in “distantly related to” status. Because the threshold of similarity that correlates with a change of substrate specificity is variable from one to another, the criteria were tightened or loosened appropriately for several families. All analyses were conducted domain by domain to avoid problems arising from the modular structure of many of the proteins.

**Polysaccharide utilization.** F. johnsoniae utilizes a variety of polysaccharides as nutrients. Analysis of the genome identified 138 predicted glycoside hydrolases and nine predicted polysaccharide lyases. F. johnsoniae was originally described as a chitinolytic bacterium (87), and analysis of the genome sequence identified genes encoding possible chitinolytic enzymes. These enzymes include five chitinases that cut the long chitin polymers and five β-N-acetylglucosaminidases that release N-acetylglucosamine and/or chitobiose from the oligomers (Table 1). The predicted chitinases are diverse and include enzymes that are related to chitinases of bacteria (Fjoh_4175, Fjoh_4555, Fjoh_4757), animals (Fjoh_4560), and plants (Fjoh_2608). Other predicted enzymes that may be involved in chitin utilization include an N-acetylglucosamine kinase (Fjoh_4589) related to mouse NagK (34), an N-acetylglucosamine-6-phosphate deacetylase (Fjoh_3974) related to Vibrio cholerae NagA (98), and four glucosamine-6-phosphate isomerases/deaminases (Fjoh_2029, Fjoh_2381, Fjoh_4812, Fjoh_4557) related to E. coli NagB (79). NagK, NagA, and NagB are predicted to function in sequence to convert N-acetylglucosamine into fructose-6-phosphate for entry into the Embden-Meyerhof-Parnas pathway.

### Table 1. Predicted F. johnsoniae glycohydrolases involved in chitin digestion

<table>
<thead>
<tr>
<th>Gene</th>
<th>Predicted function</th>
<th>Homologs</th>
<th>Predicted localization</th>
<th>Molecular mass (kDa)</th>
<th>Enzymatic and other domains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fjoh_0674</td>
<td>Candidate β-N-acetylglucosaminidase</td>
<td>P. gingivalis NahA (38% identity over 774 amino acids [52])</td>
<td>Periplasmic</td>
<td>87.6 GH20</td>
<td></td>
</tr>
<tr>
<td>Fjoh_2039</td>
<td>Candidate β-N-acetylglucosaminidase</td>
<td>P. gingivalis NahA (40% identity over 610 amino acids [52])</td>
<td>Periplasmic</td>
<td>87.3 GH20</td>
<td></td>
</tr>
<tr>
<td>Fjoh_2118</td>
<td>Candidate β-glucosidase, related to N-acetylglucosaminidases</td>
<td>Pseudomonas piscida Chx60 (30% identity over 389 amino acids [91])</td>
<td>Cytoplasmic</td>
<td>59.9 GH3</td>
<td></td>
</tr>
<tr>
<td>Fjoh_2608</td>
<td>Distantly related to plant chitinases</td>
<td>Solanum tuberosum (potato) CHTB2 endochitinase 2 (24% identity over 231 amino acids [12])</td>
<td>Unknown</td>
<td>90.8 GH19</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4175</td>
<td>Candidate chitinase</td>
<td>Bacillus circulans ChiA (26% identity over 317 amino acids [94])</td>
<td>Unknown</td>
<td>57.8 GH18-CBM6-D5</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4555</td>
<td>Candidate chitinase</td>
<td>B. circulans ChiA (33% identity over 513 amino acids [94]), B. circulans ChiD (37% identity over 353 amino acids [93])</td>
<td>Outer membrane or extracellular</td>
<td>168.9 GH18-GH18</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4556</td>
<td>Candidate β-N-acetylglucosaminidase</td>
<td>P. gingivalis NahA (36% identity over 447 amino acids [52])</td>
<td>Periplasmic</td>
<td>77.3 GH20</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4560</td>
<td>Candidate chitinase</td>
<td>Bos taurus (bovine) ChiA (25% identity over 244 amino acids [89])</td>
<td>Lipoprotein</td>
<td>38.1 GH18</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4757</td>
<td>Candidate chitinase</td>
<td>B. circulans ChiA (37% identity over 298 amino acids [94])</td>
<td>Unknown, not cytoplasmic</td>
<td>41.2 GH18</td>
<td></td>
</tr>
<tr>
<td>Fjoh_4808</td>
<td>Candidate β-N-acetylglucosaminidase</td>
<td>P. gingivalis NahA (32% identity over 633 amino acids [52])</td>
<td>Periplasmic</td>
<td>94.8 GH20</td>
<td></td>
</tr>
</tbody>
</table>

*a Predicted functions were assigned by routines used for updating the Carbohydrate Active Enzymes database (http://www.cazy.org/) using the following criteria: typically, 70% or greater amino acid identity with a protein domain with a biochemically determined function at the time of analysis resulted in “candidate” status; 30% to 70% amino acid identity with a protein domain with a known function resulted in “related to” status; and less than 30% amino acid identity with a protein domain with a known function resulted in “distantly related to” status. Because the threshold of similarity that correlates with a change of substrate specificity is variable from one to another, the criteria were tightened or loosened appropriately for several families. All analyses were conducted domain by domain to avoid problems arising from the modular structure of many of the proteins.

*b Homologs were identified by a BlastP search with the Swiss-Prot database. Fjoh_4555 has two catalytic domains, so a homolog for each domain is listed. ChiA is similar to the N-terminal GH18 domain, and ChiD is similar to the C-terminal GH18 domain. The numbers in brackets are reference numbers.

*c Localization was predicted using the default settings of PSORTb (26). Predicted lipoproteins were identified using LipoP (42).

*d Predicted molecular mass of the primary product of translation, including any predicted signal peptide.

*e CBM6, family 6 CBM, as assigned by CAZY; D5, carboxy-terminal domain of R. marinus xylanases predicted to be involved in attachment to the cell surface (43); GH, glycoside hydrolase as assigned by CAZY (the numbers indicate families).
endoglucanase that could be involved in xylolgucan digestion were identified. Digestion of hemicelluloses has not been reported previously for *F. johnsoniae*. We tested growth of *F. johnsoniae* on a battery of carbohydrates and determined that it grows on many hemicellulosic substrates (Table 2). The highest growth rates were observed with three polysaccharides: glucomannan, polygalacturonate, and laminarin. *F. johnsoniae* also utilized the major monosaccharides in these glycans (mannose, galacturonic acid, and glucose, respectively), but the growth rates were lower. A similar phenomenon has been observed for *B. thetaiotaomicron*, which grows more rapidly on starch (21, 87) and is also known to utilize starch (21, 87) and xylanolytic enzymes (43). The *R. marinus* D5 domains have been postulated to be involved in attachment to the cell surface. One cluster of four genes (Fjoh_4174 to Fjoh_4177) has a high density of CBMs and D5 domains. The proteins encoded by these genes account for 6 of the 21 CBMs identified and 4 of the 11 D5 domains. The function of this cluster is not known, but it encodes a protein that lacks an obvious catalytic domain but contains two terminal domains that are similar to the “D5” domains of *Rhodothermus marinus* cell-associated xylanolytic enzymes (43). The *R. marinus* D5 domains have been postulated to be involved in attachment to the cell surface. One cluster of four genes (Fjoh_4174 to Fjoh_4177) has a high density of CBMs and D5 domains. The proteins encoded by these genes account for 6 of the 21 CBMs identified and 4 of the 11 D5 domains. The function of this cluster is not known, but it encodes a candidate chitinase, two possible β-1,3-glucanases, and a protein that lacks an obvious catalytic domain but contains two CBMs and a D5 domain.

Other predicted proteins that may be involved in polysaccharide utilization include proteins with similarity to the *B. thetaiotaomicron* outer membrane starch utilization proteins SusC and SusD. *B. thetaiotaomicron* is an anaerobic inhabitant of the human large intestine and is a distant relative of *F. johnsoniae*. The *B. thetaiotaomicron* polysaccharide-degrading enzymes are primarily cell associated (6, 81). The outer membrane lipoprotein SusD and the outer membrane protein SusC are involved in binding starch to the cell surface and in transport of oligomers across the outer membrane. Homologs of SusC and SusD are common in the phylum *Bacteroidetes* (31).

### Table 2. Growth of *F. johnsoniae* on various carbohydrates

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Substrate type</th>
<th>Relative growth rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>d-Glucose</td>
<td>Monosaccharide</td>
<td>1.0 (0.2)</td>
</tr>
<tr>
<td>d-Fructose</td>
<td>Monosaccharide</td>
<td>0.9 (0.1)</td>
</tr>
<tr>
<td>d-Galactose</td>
<td>Monosaccharide</td>
<td>0.7 (0.1)</td>
</tr>
<tr>
<td>d-Mannose</td>
<td>Monosaccharide</td>
<td>-</td>
</tr>
<tr>
<td>d-Xylose</td>
<td>Monosaccharide</td>
<td>1.2 (0.2)</td>
</tr>
<tr>
<td>d-Galacturonic acid</td>
<td>Monosaccharide</td>
<td>0.5 (0.1)</td>
</tr>
<tr>
<td>d-Glucosamine</td>
<td>Monosaccharide</td>
<td>0.6 (0.1)</td>
</tr>
<tr>
<td>N-Acetylglucosamine</td>
<td>Monosaccharide</td>
<td>±</td>
</tr>
<tr>
<td>Celllobiose</td>
<td>Disaccharide</td>
<td>1.0 (0.1)</td>
</tr>
<tr>
<td>Cellobiohexaose</td>
<td>Hexasaccharide</td>
<td>±</td>
</tr>
<tr>
<td>Pullulan</td>
<td>Linear water-soluble polymer of glucose with α-1,3 and α-1,6 glycosidic bonds</td>
<td>0.6 (0.1)</td>
</tr>
<tr>
<td>Laminarin (brown algae)</td>
<td>Linear polymer of glucose with β-1,3 and β-1,6 glycosidic bonds</td>
<td>1.3 (0.1)</td>
</tr>
<tr>
<td>Amylopectin (potato)</td>
<td>Insoluble component of plant starch; branched polymer of glucose with α-1,4 and α-1,6 glycosidic bonds</td>
<td>±</td>
</tr>
<tr>
<td>Dextran (from <em>L. mesenteroides</em>)</td>
<td>Branched polymer of glucose with primarily α-1,6 and α-1,4 glycosidic bonds</td>
<td>0.8 (0.1)</td>
</tr>
<tr>
<td>Polygalacturonate (citrus peel)</td>
<td>Pectic polysaccharide</td>
<td>1.5 (0.1)</td>
</tr>
<tr>
<td>Arabinoxylans (wheat)</td>
<td>Hemicellulosic polysaccharide</td>
<td>±</td>
</tr>
<tr>
<td>β-Glucan (barley)</td>
<td>Hemicellulosic polysaccharide</td>
<td>±</td>
</tr>
<tr>
<td>Methyl glucuronyl xylan</td>
<td>Hemicellulosic polysaccharide</td>
<td>±</td>
</tr>
<tr>
<td>Xylan (oat spelt, water-soluble fraction)</td>
<td>Hemicellulosic polysaccharide</td>
<td>0.7 (0.1)</td>
</tr>
<tr>
<td>Xyloglucan (tamarind)</td>
<td>Hemicellulosic polysaccharide</td>
<td>0.8 (0.1)</td>
</tr>
<tr>
<td>Galactomannan (carob)</td>
<td>Hemicellulosic polysaccharide</td>
<td>1.2 (0.1)</td>
</tr>
<tr>
<td>Glucomannan (konjac)</td>
<td>Hemicellulosic polysaccharide</td>
<td>1.9 (0.4)</td>
</tr>
</tbody>
</table>

* All carbohydrates were used at a concentration of 5 mg ml⁻¹ in SD minimal medium, except for rhamnogalacturonan I, which was used at a concentration of 0 mg ml⁻¹. The following carbohydrates did not support growth: d-arabinose, L-fucose, ribose, d-glucuronic acid, N-acetylgalactosamine, 1-rhamnose, sucrose, lactose, carboxymethyl cellulose, d-mannan (*Saccharomyces cerevisiae*), d-arabinan (sugar beet), arabinogalactan (larch), inulin (chicory), β-1,4-galactan (potato and lupin), α-carrageenan (scowed), levan (*Erwinia herbicola*), rhamnogalacturonan I (potato pectin), RNA, DNA (salmon sperm), and xanthan gum. The sources of some of the polysaccharides are indicated in parentheses.

* Relative growth rates were determined by dividing the doubling time (in min) in SD medium plus glucose by the doubling time in SD plus the test substrate. Thus, a relative growth rate greater than 1 for a substrate indicates that growth on the substrate was more rapid than growth on glucose. Each substrate was tested in quadruplicate, and standard deviations are indicated in parentheses. ±, cells grew too slowly to calculate the relative growth rate, but the optical density at 600 nm was at least 0.1 after 88 h of incubation; +, cells grew on the substrate, but the growth rate could not be determined accurately due to the high turbidity of the initial growth medium containing the insoluble polysaccharide.
Analysis of the *F. johnsoniae* genome revealed the presence of 44 *susC*-like genes and 42 *susD*-like genes (see Table S4 in the supplemental material). Each of the *susD*-like genes was located immediately downstream of a *susC*-like gene. In many cases the genes adjacent to the *susC*-like and *susD*-like genes are predicted to be involved in polysaccharide utilization (Fig. 2; see Fig. S2 in the supplemental material). In *Bacteroides* species such regions have been referred to as polysaccharide utilization loci (PULs) (13). The *SusC*-like and *SusD*-like proteins are likely involved in binding and uptake of the polysaccharides.
charide substrates that are attacked by the products of the neighboring genes. The SusC-like proteins are part of a larger family of TonB receptor-like proteins. In addition to the SusC-like proteins, 50 other TonB receptor-like proteins were identified using the genome information. Some of these proteins are predicted to function in Fe uptake, but the roles of the others are not known.

The SusC-like and SusD-like proteins that have been studied are involved in the utilization of soluble polysaccharides (56, 77). Analysis of the F. johnsoniae genome suggests that some SusC-like and SusD-like proteins may also function in utilization of insoluble polysaccharides, such as chitin and hemicelluloses. The region from Fjoh_4555 to Fjoh_4564 is likely to be involved in chitin utilization (Table 1 and Fig. 2). It contains genes encoding several candidate chitinases (Fjoh_4555, Fjoh_4560), a candidate 1,4-β-N-acetylglucosaminidase (Fjoh_4556), a glucosamine-6-phosphate isomerase/deaminase (Fjoh_4557), two SusC-like proteins (Fjoh_4559, Fjoh_4562), and two SusD-like proteins (Fjoh_4558, Fjoh_4561). The 1,4-β-N-acetylglucosaminidases thought to be involved in chitin digestion are all predicted to be periplasmic or cytoplasmic enzymes (Table 1), suggesting that chitin oligomers are transported across the outer membrane before digestion, as expected if SusC-like and SusD-like proteins are involved in chitin utilization. This suggests a model for chitin utilization in which cell surface proteins bind chitin and perform the initial digestion to form soluble oligomers that are transported into the periplasm for further digestion by 1,4-β-N-acetylglucosaminidases. This model may explain why intimate contact with insoluble chitin is needed for efficient utilization (61, 88). This strategy may also be used by F. johnsoniae for digestion of hemicelluloses, and it may be used by other bacteroidetes that digest insoluble substrates. For example, Cytophaga hutchinsonii has SusC-like and SusD-like proteins that may be involved in utilization of insoluble cellulose (95).

Proteases. Casein, gelatin, and other proteins are digested by F. johnsoniae and can serve as sole C, N, and energy sources (19, 21, 87). The MEROPS peptidase database recognized 124 predicted F. johnsoniae peptidases (see Table S5 in the supplemental material) and another 35 “nonpeptidase homolog” proteins that exhibit sequence similarity to peptidases but appear to lack residues critical for activity. The number of predicted peptidases is greater than that for any other member of the Bacteroidetes with a completed genome sequence. In addition to peptidases identified by MEROPS, F. johnsoniae has one additional predicted peptidase (Fjoh_0798), a homolog of the F. psychrophilum protease Fpp1 (83) that is not currently included in the MEROPS database. Ten of the predicted peptidase-encoding genes listed in Table S5 in the supplemental material are located near susC-like genes (see Fig. S2 in the supplemental material; data not shown). For eight of these genes, genes encoding obvious polysaccharide utilization proteins have not been found nearby, raising the possibility that these susC-like genes may function in protein utilization rather than in polysaccharide utilization. Proteases of the pathogenic bacteroidetes Porphyromonas gingivalis and F. psychrophilum have been suggested to play important roles in virulence (24, 69). F. johnsoniae has homologs of several of the extracellular proteases of these organisms, but it lacks homologs of the P. gingivalis gingipain proteases, which have been most clearly associated with virulence.

Protein export and secretion. As mentioned above, extracellular and cell surface proteins appear to be involved in polymer digestion by F. johnsoniae. Export of proteins across the cytoplasmic membrane is apparently mediated by SecA, SecE, SecY, SecDF, SecG, YidC, and YajC of the Sec system and TatA and TatC of the twin-arginine transport system. Several systems are available to mediate secretion of proteins across the outer membrane. Components of an apparent type II secretion system (GspD, GspE, GspF, GspG, and GspJ) are present. In addition, multiple copies of type IV secretion system genes encoding proteins related to VirB4 and VirD4 are also present and are associated with apparent conjugative transposons. These proteins may be involved in translocation of DNA and/or proteins. Possible ATP-binding cassette transporters that may transport specific proteins (type I transport) were also identified. Finally, F. johnsoniae has predicted components of a proposed bacteroidete-specific protein secretion system (K. Sato, M. Naito, H. Yukitake, H. Hirakawa, M. Shoji, M. J. McBride, R. G. Rhodes, and K. Nakayama, submitted for publication). This system, which is required for cell movement, is discussed below. The F. johnsoniae genome is predicted to encode 423 lipoproteins (8.37% of all proteins), as well as machinery for their processing (Lgt, LspA) and translocation (LolA).

Central metabolism and biosynthetic capabilities. F. johnsoniae carries out aerobic respiration of glucose, and genome analysis indicated the presence of a complete Embden-Meyerhof-Parnas pathway and a tricarboxylic acid cycle. Genes encoding each of the NADH dehydrogenase subunits, cytochrome c, cytochrome c oxidase, and components of ATP synthase were also present. F. johnsoniae uses menaquinones instead of ubiquinones as respiratory electron transport chain components, and genes encoding the enzymes for menaquinone synthesis were identified. F. johnsoniae does not have genes encoding the components of the bc1 complex (also known as complex III) that functions as a quinol:cytochrome c oxidoreductase in many aerobes. Instead, F. johnsoniae has genes encoding components of an alternative complex III that is thought to function as a menaquinol:cytochrome c oxidoreductase. The components of this complex were recently recognized in the distantly related bacteroidete R. marinus and in the photosynthetic bacterium Chloroflexus aurantiacus (70, 99). Involvement of menaquinones and replacement of the bc1 complex with the novel menaquinol oxidoreductase appear to be universal among aerobic bacteroidetes for which genome sequence data are available, including F. johnsoniae, F. psychrophilum, Gramella forsetii, C. hutchinsonii, and R. marinus. Some apparent metabolic abilities of F. johnsoniae were not anticipated. For example, unlike most other bacteroidetes, this species has a cluster of genes (Fjoh_3902 to Fjoh_3913) predicted to encode a nickel-dependent uptake hydrogenase complex and accessory proteins for maturation of the enzyme. The role of this hydrogenase complex in the energy metabolism of this aerobic respiratory heterotrophic bacterium is not yet clear. F. johnsoniae grows on minimal media with glucose as a sole carbon and energy source and thus has the ability to synthesize all of its organic components from this substrate. As expected, analysis of the genome sequence revealed genes en-
coding biosynthetic enzymes needed to synthesize amino acids, nucleotides, fatty acids, heme, and many vitamins and coenzymes. Novel outer membrane lipids and secondary metabolites. Many gliding bacteroidetes contain novel components in their outer membranes. These components include sulfonolipids, which may give the membrane increased fluidity and be important for gliding (1), and flexirubin pigments (3). Flexirubin pigments are found in many bacteroidetes, including *F. johnsoniae*, *F. psychrophilum*, and *C. hutchinsonii*, for which complete genome sequences are available. Flexirubin pigments typically consist of an α-phenyloctaenic acid chromophore esterified with resorcinol carrying two hydrocarbon chains (Fig. 3). Variations in the length of the polyenic acid, variations in the R group hydrocarbon chains, and substitutions on the phenyl rings result in production of a variety of different flexirubin pigments. A single strain of *F. johnsoniae* was reported to produce 25 different flexirubin pigments (2). Flexirubins are yellow at neutral pH but undergo a reversible switch to red under alkaline conditions. This allows simple identification of flexirubin-negative mutants.

A combination of genome analysis and genetic experiments resulted in identification of a cluster of genes involved in flexirubin synthesis. Comparative genome analysis identified two genes, Fjoh_1103 and Fjoh_1102, with likely roles in flexirubin synthesis. Homologs of these genes were identified in the genomes of the three flexirubin-producing bacteroidetes, *F. johnsoniae*, *F. psychrophilum*, and *C. hutchinsonii*, but not in the flexirubin nonproducers *Bacteroides fragilis*, *B. thetataomicron*, *Bacteroides vulgatus*, *G. forsetii*, *Parabacteroides distasonis*, and *P. gingivalis*. Fjoh_1103 and Fjoh_1102 are similar to *P. aurantiaca darA* and *darB*, respectively, which are involved in biosynthesis of the antifungal compound 2-hexyl-5-propyl-alkylresorcinol (68). Similar 2,5-dialkylresorcinol compounds are likely intermediates in *F. johnsoniae* flexirubin biosynthesis. An insertion mutation was constructed in *F. johnsoniae darB* (Fjoh_1102), and the resulting strain, CJ1702, lacked flexirubin pigments. Cells of the mutant were cream colored, whereas wild-type cells were yellow at neutral pH and red under alkaline conditions (Fig. 3B). The conversion of yellow wild-type cells to red cells by exposure to KOH was reversed by addition of HCl, as previously reported for flexirubin pigments (3). *F. johnsoniae darA* and *darB* are part of a large cluster of genes transcribed in the same direction and organized in what appear to be several operons (Fig. 3C). Genome comparisons revealed that in addition to *darA* and *darB*, homologs of seven other genes in this cluster (Fjoh_1080, Fjoh_1084, Fjoh_1095, Fjoh_1097, Fjoh_1098, Fjoh_1100, Fjoh_1108) are present in the flexirubin-positive bacteroidetes mentioned above but not in bacteroidetes that are flexirubin negative. Most of these genes encode hypothetical proteins of unknown function, but Fjoh_1095 encodes a predicted component of an ABC-2-type transporter that could be involved in localization of flexirubin pigments. Many other genes in this region are predicted to encode enzymes involved in lipid synthesis, such as a (3-oxoacyl)-(acyl carrier protein synthase (Fjoh_1087), beta-ketoacyl synthases (Fjoh_1088, Fjoh_1093, Fjoh_1106), and a beta-hydroxyacyl-(acyl carrier protein) dehydratase (Fjoh_1081), and some of these enzymes could have roles in flexirubin synthesis. *darA*, *darB*, and perhaps other genes in this region may also be involved in the synthesis of the mammalian cell growth-promoting dialkylresorcinol resorcinin, which is produced by strains of *F. johnsoniae* (38). Cosmid complementation of the spontaneous flexirubin-negative mutant *F. johnsoniae* UW102-154 (19, 37) confirmed that genes in this region other than *darA* and *darB* are involved in flexirubin synthesis, since the complementing plasmid, pMM340, carried a region spanning Fjoh_1078 to Fjoh_1089. Strains of *F. johnsoniae* are known to produce several secondary metabolites in addition to resorcinin, including monobactam and quinoline antibiotics (25, 44, 45), and numerous other secondary metabolites are produced by related bacteria (66).
Several clusters of genes likely to be involved in secondary metabolite production were identified in the *F. johnsoniae* genome. Most striking was a cluster of genes in the 70.7-kbp region spanning Fjoh_2083 to Fjoh_2104 that appear to encode components of a multienzyme nonribosomal peptide synthetase assembly line with a novel predicted product having an unknown function (see Fig. S3 in the supplemental material). *F. johnsoniae* also appears to have the capacity to produce an aerobactin-like siderophore. Genes located between Fjoh_3170 and Fjoh_3179 are predicted to be involved in the regulation, synthesis, and export of this compound.

**Signal transduction and regulation of gene expression.** *F. johnsoniae* genes encode a variety of proteins that are predicted to regulate gene expression in response to external or internal stimuli. The sigma factors include an RpoD (σ^70) homolog, an RpoN (σ^54) homolog, and 27 sigma factors belonging to the ECF subfamily. *F. johnsoniae* RpoD is similar in size (32.7 kDa) and sequence to other bacteriophage RpoD proteins and is much smaller than *E. coli* σ^70. Like the other bacteriophage RpoD proteins, *F. johnsoniae* RpoD lacks regions found in most nonbacteriophage RpoD proteins, such as N-terminal region 1.1 and the segment between regions 1.2 and 2.1 (92). The novel structure of bacteriophage RpoD sigma factors may account for some of the unusual features of bacteriophage housekeeping promoters (10, 20). For example, the −33/−7 consensus promoter sequence of *F. johnsoniae* and other members of the Bacteroidetes phylum that have been studied (TTG/TANNTTTG) differs from the −35/−10 consensus promoter sequences of other well-studied bacteria (20). A search of the *F. johnsoniae* genome with the promoter consensus sequence (with spacing between the two motifs set at the optimal value, 19 bp) revealed 109 exact matches in intergenic regions on the coding strand within 300 bp upstream of a start codon, whereas a similar search of the *E. coli* genome with the same consensus sequence revealed less than 10 sequences (20). Not surprisingly many of the putative *F. johnsoniae* promoters identified were associated with housekeeping genes that are likely to be highly expressed. Promoters with slight variations in the consensus sequence would have been missed by this analysis, so it is likely that many more RpoD-dependent promoters are present in the *F. johnsoniae* genome.

The functions of *F. johnsoniae* RpoN (σ^54) are not known, but σ^54-like proteins are required for transcription of genes involved in a wide variety of processes in other bacteria. In these other bacteria σ^54-RNA polymerase holoenzyme (σ^54-holoenzyme) binds to promoter sequences to form a closed complex, but it cannot proceed further in transcription initiation in the absence of an activator. Activators of σ^54-holoenzyme generally bind to sites located upstream of the promoter and contact σ^54-holoenzyme through DNA looping to stimulate transcription. Binding sites for σ^54-dependent activators are referred to as bacterial enhancers because some of them can function when they are moved several kilobases from their native locations.

Activators of σ^54-holoenzyme usually consist of an N-terminal regulatory domain, a central domain required for ATP hydrolysis and transcriptional activation, and a C-terminal DNA-binding domain. *F. johnsoniae* has six predicted σ^54-dependent activators (Fjoh_0470, Fjoh_0638, Fjoh_1332, Fjoh_1848, Fjoh_1977, and Fjoh_5047). Five of these activators have N-terminal response regulator receiver domains, whereas the remaining activator (Fjoh_1848) contains an N-terminal PAS domain. The stimuli controlling these activators are not known, but Fjoh_0470, Fjoh_0638, and Fjoh_1332 are adjacent to genes that likely encode the presumptive cognate sensor kinases.

Fjoh_5047 is unique among the *F. johnsoniae* σ^54-dependent activators in that it lacks a DNA-binding domain. When expressed in *E. coli*, Fjoh_5047 activated transcription from a σ^54-dependent lacZ reporter gene (data not shown), indicating that it is a functional activator. Examples of σ^54-dependent activators lacking DNA-binding domains are found in other members of the Bacteroidetes (11). A problem of specificity of gene activation arises for these unusual activators. In the few bacteria outside the Bacteroidetes phylum that are known to have such activators, specificity is conferred by one of two mechanisms. Some of these bacteria have only a single σ^54-dependent activator and thus avoid the specificity problem entirely, and specificity is conferred by the sigma factor (67). In other cases multiple σ^54-dependent activators are present, but these activators interact with different versions of σ^54 (74, 75). *F. johnsoniae* has multiple σ^54-dependent activators and a single σ^54, so it is unclear how specificity could be conferred without a DNA-binding domain. Fjoh_5047 may indiscriminately activate transcription of all of the genes in the RpoN regulon. Alternatively, Fjoh_5047 may gain specificity by interacting with an unidentified DNA-binding protein, or it may engage σ^54-holoenzyme in solution to alter the promoter specificity of the enzyme.

The genes regulated by RpoN and its activators are not known, but analysis of genes located near the genes encoding σ^54-dependent activators and identification of sequences that match the σ^54 consensus sequence suggest likely candidates (see Table S6 in the supplemental material). Two apparent operons (Fjoh_0471, Fjoh_0472, and Fjoh_0473; and Fjoh_1331, Fjoh_1330, and Fjoh_1329) are particularly good candidates for σ^54-dependent regulation because they are located adjacent to genes encoding σ^54-dependent activators and they have sequences that match the σ^54 promoter consensus sequence. In addition, multiple regions of dyad symmetry that may function as binding sites for the σ^54-dependent activators are located within 63 bp upstream of these putative promoters. Both of these operons are predicted to encode ABC-type antibacterial peptide transport systems, suggesting roles for σ^54 in resistance to such compounds.

The presence of large numbers of ECF sigma factors is a common property of members of the phylum Bacteroidetes (46). ECF sigma factors in other bacteria regulate expression of genes with extracellular functions. This regulation often involves transmembrane FecR-like anti-sigma factors that sense extracytoplasmic stimuli, and 12 of the 27 *F. johnsoniae* genes encoding ECF sigma factors are adjacent to genes predicted to encode such anti-sigma factors. Eight of the ECF sigma factors are encoded by genes that are adjacent to susC-like and susD-like genes (Fig. 2; see Fig. S2 in the supplemental material), and each of these genes is accompanied by a fecR-like anti-sigma factor gene. The eight SusC-like proteins encoded by these regions have an extra N-terminal signaling domain that is thought to interact with the relevant anti-sigma factor during signaling (Fig. 2) (48). Expression of nearby genes involved in
The genomes of other bacteroidetes known to exhibit gliding motility, such as *F. psychrophilum* and *C. hutchinsonii*, have homologs of each of the 14 *F. johnsoniae* motility genes described above (see Table S7 in the supplemental material). Not surprisingly, the amino acid sequences of the *F. johnsoniae* and *F. psychrophilum* motility proteins are similar, with identities ranging from 32% to 87%. *C. hutchinsonii* is not closely related to the two flavobacteria, and the levels of amino acid identity with the *F. johnsoniae* motility proteins are lower, ranging from 22% to 50%. In spite of the divergence in amino acid sequence, the organization of several gene clusters has been highly conserved. For example, *gldF* and *gldG*, which have been demonstrated to constitute an operon in *F. johnsoniae* (35), are also likely to be cotranscribed in *F. psychrophilum* and in *C. hutchinsonii* based on genome sequence analyses. GldF and GldG localize in the cytoplasmic membrane of *F. johnsoniae* and interact with GldA (35). Together, these proteins are thought to constitute an ATP-binding cassette transporter that is required for gliding. *C. hutchinsonii* *gldA* appears to be cotranscribed with *gldF* and *gldG*, which is consistent with the hypothesis that the products of these genes function together and form a complex. In *F. johnsoniae*, *gldK*, *gldL*, *gldM*, and *gldN* are clustered together on the genome and form two adjacent operons, and *gldK* is transcribed separately from the other three genes (14). A similar arrangement is found in *F. psychrophilum* and *C. hutchinsonii*. This may suggest that the protein products of these genes function together as part of a
complex and that coordinated expression may be important for assembly of the complex in the cell envelope.

Nonmotile bacteroidetes, such as P. gingivalis, B. fragilis, B. thetaiotaomicron, B. vulgatus, P. distasonis, and Salinibacter ruber, have homologs of some, but not all, of the F. johnsoniae motility genes (see Table S7 in the supplemental material). The SprA homolog in P. gingivalis, Sov, is required for secretion of gingipain proteases (80). Recent results suggest that homologs of GldK, GldL, GldM, and GldN are also required for P. gingivalis gingipain secretion and may be components of a novel protein secretion system (Sato et al., submitted). Conservation of many of the genes encoding these proteins in B. fragilis and P. distasonis suggests that similar protein secretion systems may function in many gliding and nongliding bacteroidetes. The F. johnsoniae secretion system may function in translocation of the motility protein SprB to the cell surface (Sato et al., submitted). The presence of a protein translocation machine at the heart of a motility apparatus is not unique, because the bacterial flagellum is built around a type III secretion system that is involved in secretion of flagellins and some other flagellar proteins (5). P. gingivalis lacks homologs of gldD, gldF, gldG, and gldJ. The proteins encoded by these genes may have functions that are essential for motility but not for protein secretion. The presence of homologs of each of the motility genes in the genome of the relatively unstudied marine bacteroidete G. forsetii (9) suggests that it may have the ability to glide, a property that has not been experimentally observed yet.

Most motile bacteria control their motility to move in a favorable direction. This typically involves a complex signal transduction system related to the E. coli chemotaxis system. F. johnsoniae has homologs of E. coli cheY (Fjoh_3353), cheB (Fjoh_3351), and cheR (Fjoh_3352). These genes are clustered together on the genome, supporting the idea that their products may function together. F. johnsoniae cheB is unusual in that it lacks the N-terminal response regulatory domain that is found in most CheB proteins. Genes encoding other expected components of a bacterial chemotaxis system, such as homologs of methyl-accepting chemotaxis proteins (MCPs), and of CheA, CheW, and CheZ, are apparently not present. The absence of obvious MCPs is surprising since CheB (methyl- terase) and CheR (methyltransferase) are expected to modify MCPs. The region surrounding F. johnsoniae cheB, cheR, and cheY contains several genes whose products are likely to be involved in signal transduction. These products include an apparent histidine kinase with a PAS domain (Fjoh_3360), a protein that contains a histidine kinase domain and a response regulatory domain (Fjoh_3355), and a large protein (Fjoh_3354) that contains a CHASE3 domain, a GAF domain, a histidine kinase domain, and a response regulatory domain. PAS, CHASE3, and GAF domains are commonly found in sensory or signal transduction proteins (7, 90, 100) and could have roles in F. johnsoniae tactic responses. However, the absence of central components of bacterial chemotaxis systems, such as CheA and MCPs, and the fact that the closely related gliding bacterium F. psychrophilum lacks cheB, cheR, and other obvious chemotaxis genes suggest that the genes and proteins involved in control of the bacteroidete motility machinery may be novel. Alternatively, cells of F. johnsoniae and F. psychrophilum may lack chemotaxis entirely, as has been suggested for several other motile bacteria that appear to lack critical chemotaxis genes (8, 22, 62). Negative chemotaxis of F. johnsoniae in response to H2O2 and several other chemicals has been reported based on observations of swimming patterns of groups of cells (51), but an analysis of the behavior of individual cells in response to potential chemoeffectors has not been conducted yet.

Numerous aspects of the physiology and molecular biology of F. johnsoniae were revealed in this initial analysis of its genome sequence. Many of these aspects, such as the unusual gliding motility machinery, protein secretion system, and polysaccharide utilization strategy, are shared by other members of the phylum Bacteroidetes. The genome sequence data and the availability of techniques to genetically manipulate F. johnsoniae should result in rapid progress toward a better understanding of the many novel features of these common but understudied bacteria.

ACKNOWLEDGMENTS

This work was performed under the auspices of the U.S. Department of Energy’s Office of Science, Biological and Environmental Research Program and at the University of California Lawrence Livermore National Laboratory under contract W-7405-Eng-48, at Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231, and at Los Alamos National Laboratory under contract DE-AC02-06NA25396. M.J.M. and R.G.R. were supported by National Science Foundation grant MCB-0641366. W.W. and I.X. were supported by a One Hundred Talents grant from the Chinese Academy of Sciences.

REFERENCES


