Open access • Journal Article • DOI:10.1126/SCIENCE.1107008

Extensive DNA inversions in the B. fragilis genome control variable gene expression — Source link

Ana Cerdeño-Tárraga, Sheila Patrick, Lisa Crossman, Garry W. Blakely ...+21 more authors

Published on: 04 Mar 2005 - Science (American Association for the Advancement of Science)

Topics: Gene, Bacteroides fragilis, Antigenic variation, Genome and Intergenic region

Extensive DNA inversions in the B. fragilis genome control variable gene expression

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M. A., Barron, A., Clark, L., Corton, C., Doggett, J., Holden, M. T., Larke, N., Line, A., Lord, A., Norbertczak, H., ... Parkhill, J. (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, *307(5714)*(5714), 1463-1465. https://doi.org/10.1126/science.1107008

Published in:
Science

Supporting Online Material

Materials and Methods

This sequence strain NCTC 9343 was originally isolated from an abdominal infection at St Bartholomew's Hospital, London in 1955. The culture from which DNA was produced for the genomic sequencing was enriched for the EDL phase, with minimum subculture from an early freeze dried stock culture, but remained antigenically mixed. Bacteria were grown in defined medium broth (*1*) in an anaerobic cabinet (MACS Anaerobic Workstation Don Whitley Scientific, Shipley, UK; 80% N2, 10% CO2 and 10% H2). Percoll density gradient enrichment was used to obtain populations that were non-capsulate by light microscopy as described previously (*2*).

DNA was isolated using a modification of the basic protocol for preparation of genomic DNA from bacteria detailed in Ausubel *et al.* (*3*). In brief, bacterial cells were lysed in 10 mM TrisHCl 1mM EDTA buffer (pH 8.0) containing SDS (0.5%), lysozyme (4mg/ml) and proteinase K (0.1mg/ml). Polysaccharide was precipitated using cetyltrimethylammonium bromide (CTAB) and DNA extracted with chloroform: isoamyl alcohol and phenol:choloroform:isoamyl alcohol mixtures. DNA was precipitated using isopropanol, spooled out with a glass rod and washed in ethanol. The initial genome assembly was obtained from 94,563 paired end sequences (giving 10-fold coverage) derived from four pUC18 genomic shotgun libraries (with insert sizes ranging from 2.0 to 4.0 kb) using big-dye terminator chemistry on ABI3700 automated sequencers. This was supplemented with 4,991 end-reads from an m13mp18 library with an insert size of 2.0-4.0 kb; 3,388 paired-end sequences

from a pBACe3.6 library with insert sizes of 10-18 kb (a clone coverage of 4.56-fold) were used as a scaffold. All identified repeats were bridged by read-pairs or end-sequenced PCR products. A further 4,594 sequencing reads were generated during finishing. The sequences were assembled, finished and annotated as described previously (*4*), using Artemis (*5*) to collate data and facilitate annotation. The DNA and encoded protein sequences of related species were compared using the Artemis Comparison Tool (ACT) (K. Rutherford, unpublished; http://www.sanger.ac.uk/Software/ACT/). Orthologous gene sets were calculated by reciprocal best match FASTA comparisons, with subsequent manual curation. Pseudogenes had one or more mutations that would ablate expression; each of the inactivating mutations was subsequently checked against the original sequencing data.

Genomic comparison of *Bacteroides fragilis* with *Bacteroides thetaiotaomicron*

The recently sequenced *B. thetaiotaomicron* strain VPI-5482 (ATCC 29148)(*6*), has a larger genome of 6.26 Mb containing 4779 predicted genes and an unusually low ratio of gene number to genome size with 0.763 genes/kb (*6*). The proportion of the genome devoted to protein coding (89.3%) is, however, not unusual. This is explained by a high average gene length (1173 bp), and abundant large proteins of over 600 aa. At 1087bp the average gene length in *B. fragilis* is slightly shorter, but it is still amongst the highest so far reported in bacteria.

Despite their con-generic taxonomic status, *B. fragilis* and *B. thetaiotaomicron* share remarkably few orthologous genes (as measured by reciprocal-best-match FASTA analysis). Only 2,384 predicted genes are shared, with *B. fragilis* containing

1890 unique genes (44.3 % of its coding capacity), and *B. thetaiotaomicron* containing 2337 (48.9 %). This is considerably greater than the number of genes unique to *Salmonella* vs. *Escherichia* (*7*), or to *Escherichia* vs. *Yersinia* (*8*) indicating the breadth of diversity amongst *Bacteroides*. This is underlined by the average amino-acid identity between orthologous gene pairs; just 76.6%.

A comparison of *B. thetaiotaomicron* with *B. fragilis* was carried out to determine the nature of the non-orthologous genes. A particularly interesting feature is the extent of diversity of cell envelope, transmembrane, polysaccharide and outer membrane proteins compared to *B. thetaiotaomicron* (Fig. S2); *B. fragilis* unique surface proteins include 128 inner membrane proteins, 2 peptidoglycan associated proteins and 57 predicted outer membrane proteins. In addition, there are 102 unique surface polysaccharide biosynthesis genes, arranged in 10 gene clusters (seven of which are controlled by invertible promoters). This variation in surface polysaccharides and other antigens may be significant in relation to the difference in the pathogenic potential between *B. fragilis* and *B. thetaiotaomicron*. In terms of transcriptional regulators, the most abundant class is the AraC family of transcriptional regulators, with 20 being unique compared to *B. thetaiotaomicron*. An additional point of interest in the *B. thetaiotaomicron* genome is the significantly increased number of sigma factors and anti-sigma factors relative to genome size; of a total of 45 sigma factors in *B. fragilis* there are 16 that are not conserved in *B. thetaiotaomicron*.

Additional DNA inversions

In addition to the restriction/modification intergenic shufflon, a further three different systems were identified. The simplest involves BF0335 and BF0336 (IR-P), which encode two parts of a sensor-regulator system. BF0336 encodes the putative sensor domain and transmembrane domain, and BF0335 encodes the histidine kinase, response receiver and DNA binding domains. The first CDS is flanked by inverted repeats su such that recombination will fuse the two into a single coding sequence. Other fused sensor-regulator genes, lacking invertible regions are present both in *B. fragilis* and *B. thetaiotaomicron* (*6*).

A further two more complex systems involve gene pairs similar to *ragA* (*susC* -like) and *ragB* of *P. gingivalis* (Fig. 1). In *P. gingivalis* RagB is a major immunodominant surface antigen and *ragA/B* positive strains are associated with sites of periodontal destruction (*9*) SusC is an outer membrane protein, which in association with SusD, binds starch at the bacterial surface in *B. thetaiotaomicron* (*10*). In one system, five *ragA/B* like gene pairs (BF1716/8, BF1719/20, BF1722/21, BF1798/7 and BF1803/2; IR-EE), at either end of a ~90 kb region of the chromosome, have an extensive repeat sequence (over 200bp) that overlaps the start codon of the *ragA* homologs (Fig. 1C). One of these (BF1803) is downstream of an invertible promoter (Table S2A, Group 2; IR-T), and recombination between these larger repeats could bring any one of the alternative gene pairs downstream of this promoter by inversion of the intervening sequence. Evidence that this was indeed occurring was found in the shotgun data. A potential recombinase (BF1795) is located close to BF1797/9 and is associated with the putative conjugative transposon encoded within the 90kb central region.

A second independent system involves a further group of four co-located *ragA/B* like gene pairs (BF0590, BF0592, BF0594 and BF0596; Table 1b, IR-CC), where three of the *susC* (*ragA*-like) homologues lack an appropriate start codon. Within the 5' end of the coding region of these genes is a 60bp inverted repeat, identical to that found 68 amino-acids downstream of the start codon of the fourth, BF0594 (Fig. 1C). Recombination between these inverted repeats would fuse the coding sequences of any of these genes to the promoter and translation initiation signals of BF0594. Again, evidence was seen that this was occurring in the shotgun. Another potential recombinase (BF0593) is encoded adjacent to these genes. Examples of similar multiple intergenic inversions include *omp1* of *Dichelobacter* (formerly *Bacteroides*) *nodosus* and the variable surface antigens of *Mycoplasma pulmonis* (*11*). There are also invertible promoter regions upstream of nine other *susC* homologues (Table S2A) and several other predicted outer membrane proteins belonging to a related family (Table S4). This indicates that there is a strong selective pressure in favour of the variation of these molecules, which suggests that they play an important role in the survival of *B. fragilis*.

Several invertible promoters are between divergent coding regions, thus potentially driving variable transcription of other genes. These include a sialoconjugate degradation operon (IR-T; BF1804-BF1817); a sigma factor (IR-W; BF2944), indicating the possibility of cascade regulation; and the *cpn10-cpn60* operon encoding the major chaperones GroES and GroEL (IR-M). These genes are known to be essential, and have not been seen to be phase-variable, in other organisms. Analysis of the promoter region for *cpn10* in *B. thetaiotaomicron* shows that it is

flanked by 29bp inverted repeats whose sequence is similar to that of *B. fragilis*, although it has not been identified as being invertible.

An 11kb region of the plasmid (pBF9343) containing partition, replication and mobilisation functions, and the invertase *finB*, undergoes inversion (Fig. S1B). There appears to be no clear functional consequence of this, although it may affect the transcription of *finB*, as one of the ends of the inversion is ~100 bp upstream of its start site. As both FinB and the chromosomally-encoded FinA (MpiA) may be involved in inversion of *hin*-like invertible promoters, this raises the possibility of a random variation in the rate of variation at those promoters in plasmid-containing strains.

Virulence associated genes

Iron is essential for the growth of most bacteria. Within the human host iron is sequestered such that the free iron concentration is estimated to be 10^{-18} M, well below the concentration necessary to support bacterial growth. To overcome this iron famine, pathogenic bacteria have evolved specific adaptive mechanisms for obtaining iron. One is the production of secreted iron chelating compounds termed siderophores that sequester iron and are subsequently re-imported through membrane receptors (*12*). Internalisation is generally performed by TonB-dependent outer-membrane receptors, of which *B. fragilis* contains at least 57, and is energised by the TonB/ExbBD complex (BF3737, BF3738 and BF3739). An FhuA (*Vibrio cholerae* ferrichrome receptor) homologue, together with a two-component sensor-regulator pair and a complete sigma, anti-sigma and anti-anti-sigma factor system (BF2844, BF2842 and BF2843) is evident. Whether *B. fragilis* synthesises siderophores or

utilizes those produced by other bacteria remains to be determined, although a non-ribosomal peptide synthase (BF2837), an enzyme family known to be involved in the biosynthesis pathway for other siderophores, is located nearby and may be part of a siderophore biosynthesis system. In addition, there are two periplasmic binding-protein dependent iron uptake systems, of the ferric citrate FecCD family (BF1185 and BF2247), each with an associated periplasmic iron-binding protein. The putative siderophore uptake systems appear to be independent of the already identified iron-repressible haem uptake protein HupA that forms part of a haem binding outer membrane protein complex (*13*). Although growth is severely limited in the absence of haem (*14*), *B. fragilis* does not produce zones of haemolysis on blood agar. There are, however, 13 CDSs, fewer than in the *B. thetaiotaomicron* genome, which may encode a haemolytic function. One of these (BF0270) is similar to the haemolysin A of *Prevotella melaninogenica (*formerly *Bacteroides melaninogenicus)*. Interestingly, within the associated CDSs (BF0266-0269) there is a putative haem receptor, suggesting that this might be a haem acquisition operon.

Extracellular enzymes, potentially capable of degrading components of the host's extracellular matrix, host cells and tissue, and therefore potentially involved in *B. fragilis* virulence, include hyaluronidase, chondroitin sulphatase, fibrinolysin, DNAase, lipases, proteases and neuraminidases (*15*). Bacterial sialidases or neuraminidases that remove the sialic acid residues from host oligosaccharides are implicated in bacterial virulence as they potentially interfere with normal host cell function. The neuraminidase NanH (BF1806) has previously been sequenced from *B. fragilis* and a second, highly similar, gene is present (BF4051), which could be equivalent to the second previously reported in *B. fragilis* strain SBT3182 by Tanaka,

et al. (*16*) The observed degradation of hyaluronic acid, a component of the host's extracellular matrix, may be due to BF3796, which is similar to the hyaluronidase encoded by a *Streptococcus pyogenes* bacteriophage (*17*). Three putative tricorn-like proteases (BF0080, 2517 and 3752) with similarity in the beta-propeller, PDZ and catalytic domains are evident, but as in the MdsD protein of *Prevotella* sp, they also contain a putative signal peptide (*18*). Which of the potentially secreted putative 23 peptidases and 2 lipases are involved in virulence rather than nutrition remains to be determined.

Attachment to host tissues is key to the virulence of many pathogenic bacteria, and can be promoted by a variety of adhesins. The non-capsulate/EDL population of *B. fragilis* haemagglutinates erythrocytes (*19*). One haemagglutinin (BF1428) that has a homologue in *B. thetaiotaomicron* is evident; however, no fimbrial genes have been identified, despite reports of their observation in other strains (*20*). There is no evidence for type III, IV, autotransporter or two-partner secretion systems nor flagellar biosynthesis systems in the *B. fragilis* genome, hence secreted virulence determinants are likely to be exported via Hly-type I secretion systems such as BF0010 and BF0011, BF0608 and BF0610 or BF3811 and BF3812, which are similar to the haemolysin type I secretion system HlyDB from *E. coli* (*21*), or via the type II general secretion pathway. *B. fragilis* produces copious quantities of enzyme-containing outer membrane vesicles (*22*). This may therefore be an important export mechanism.

DNA recombination and repair of metronidazole induced DNA damage

The current drug of choice in treating *B. fragilis* infections, metronidazole, is activated intracellularly via anaerobic reduction of the nitro group and interacts with DNA causing strand breaks. RecA-mediated strand exchange is required for repair of DNA breaks, the importance of which is shown by the sensitivity of a *B. thetaiotaomicron recA* mutant to metronidazole and other DNA damaging agents (*23*). RecA (BF1180) is 60% identical to *E. coli* RecA and can complement an *E. coli recA* mutant, demonstrating that DNA repair pathways are conserved in *B. fragilis*(*24*).

Two major pathways of recombination have been extensively studied in the model organisms *E. coli* and *Bacillus subtilis*; single-strand gap repair mediated by the RecFOR proteins and double-strand break repair (reviewed in ref. *25*) Specific DExx motif helicases, belonging to Superfamily I, are essential in cells expressing RecF; these are Rep and UvrD in Gram negative bacteria and PcrA in Gram positive bacteria (*26*). *B. fragilis* contains a total of 24 putative DNA helicases, compared to 11 in *E. coli*, the functional significance of which is unknown. RecFOR homologues are present in *B. fragilis* together with two homologues of the PcrA helicase usually found in Gram positive bacteria. The presence of single PcrA homologues has been suggested to be a hallmark of Gram positive bacteria (*26*). It remains to be determined whether the two homologues in *B. fragilis* have separate roles or overlapping functions analogous to the activities of Rep and UvrD in *E. coli*.

Double-strand break repair is mediated by RecBCD in *E. coli* and AddAB in *B. subtilis* (*25*). *B. fragilis* contains a CDS (BF0679) encoding a homologue of RecD but does not contain identifiable homologues of RecBC or AddAB. The RexAB ATP-dependent exonuclease/helicase has been implicated in repair of double-strand breaks

in the Gram positive bacterium *Lactobacillus lactis* (*27*). There is a single RexA (BF2192) homologue in *B. fragilis* but no identifiable RexB. A potential hypothesis is that double-strand break repair in *B. fragilis* is mediated by a novel mechanism where the joint action of the RexA and RecD helicases unwind the DNA duplex with concurrent degradation by the exonuclease function of RexA until the equivalent of a chi site is encountered. This combination of helicases of opposite polarities (RecD has 5'-3' while RexA has 3'-5' activity) would be analogous to the helicase functions present in the RecBCD complex (*28*).

The sequence and annotation of the genome are available with further details from http://www.sanger.ac.uk/Projects/B_fragilis/.

Supporting Figures:

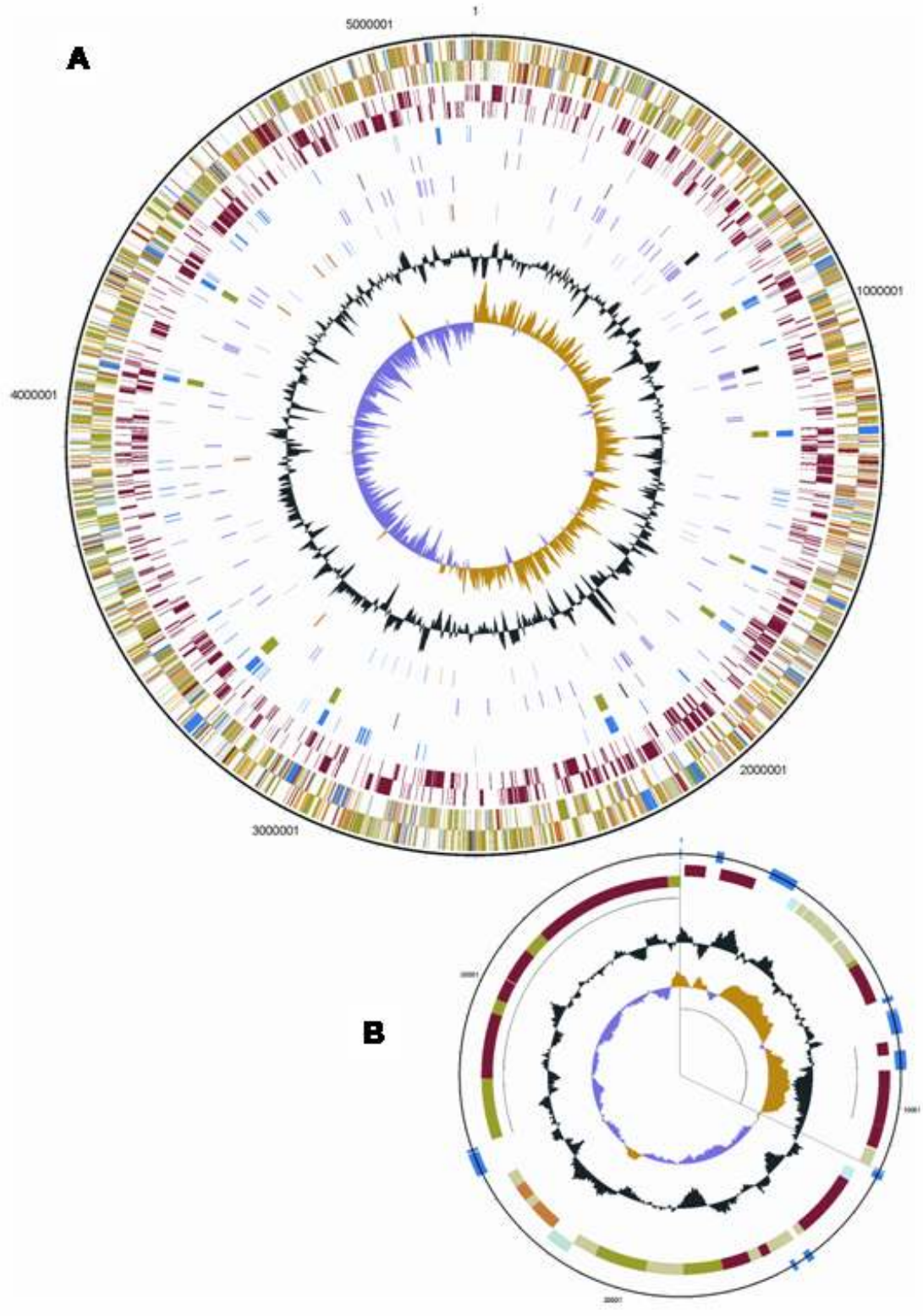

Figure S1: A) Circular representation of the *B. fragilis* NCTC9343 chromosome: From the outer to the inner circle: Circle 1: DNA coordinates (origin in base 1); Circles 2+3: all CDSs

(forward and reverse strands); Circles 4+5: Unique CDSs in *B. fragilis* as compared with *B. thetaiotaomicron* CDSs (forward and reverse strands); Circle 6: Pathogenicity related CDSs (blue); Circle 7: polysaccharide biosynthesis clusters (green) and invertible regions (black); Circle 8: SusC homologues (pink); Circle 9: rRNAs (orange) and tRNAs (blue); Circle 10: G+C content (plotted using a 10Kb window); Circle 11: GC skew ((G-C)/(G+C)) (plotted using a 10Kb window). Colour coding for circles 2 and 3: dark blue; pathogenicity/adaptation, black; energy metabolism, red; information transfer, dark green; surface associated, cyan; degradation of large molecules, magenta; degradation of small molecules, yellow; central/intermediary metabolism, pale green; unknown, pale blue; regulators, orange; conserved hypothetical, brown; pseudogenes, pink; phage+IS elements, grey; miscellaneous. **B)** Circular representation of the pBF9343 plasmid: From the outer to the inner circle: Circle 1: DNA coordinates with repeats marked as blue boxes; Circles 2+3: CDSs (forward and reverse strands); Circle 4: Transfer region marked with black line; Circle 5: Mobilisation region marked with black line; Circle 6: G+C content (plotted using a 1Kb window); Circle 7: GC skew ((G-C)/(G+C)) (plotted using a 1Kb window); Circle 8: Inverted region marked with black line.

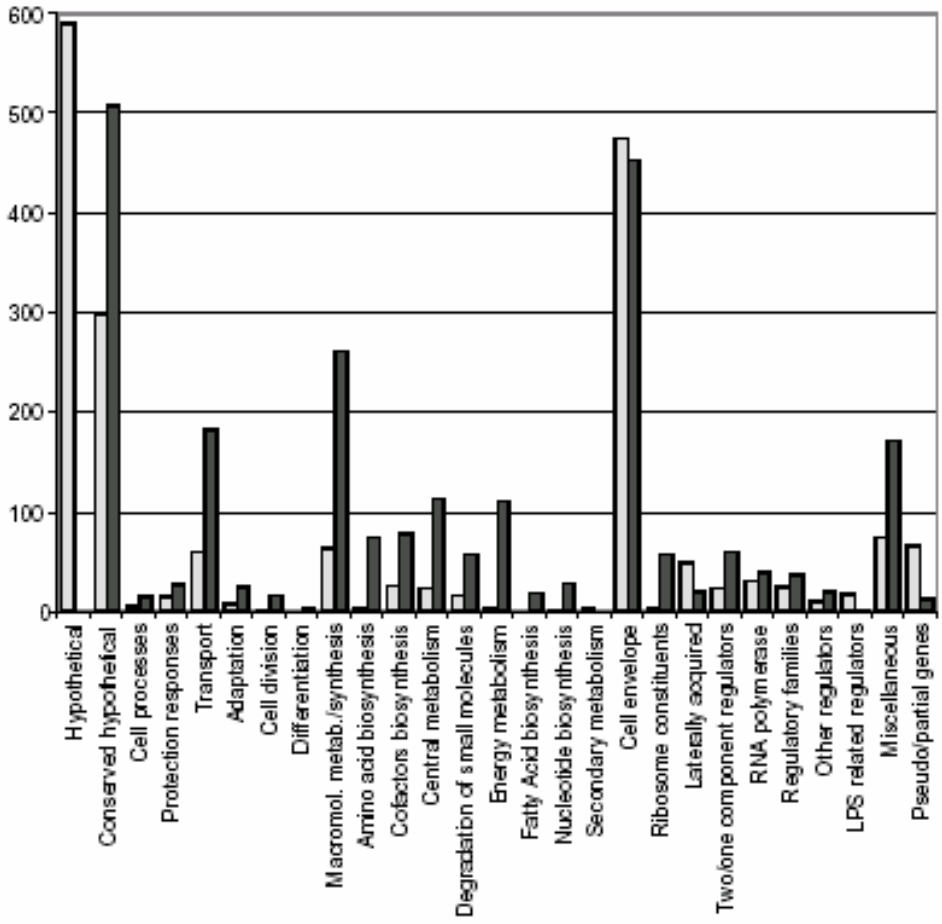

Figure S2: Numbers of orthologous genes between *B. fragilis* and *B. thetaiotaomicron* (dark grey), and genes unique to *B. fragilis* (light grey) within specific functional categories.

Supporting Tables:

Table S1 - General features of the *B. fragilis* genome.

Chromosome	Size (bp)	5,205,140
	G+C content (%)	43.19
	CDSs	4,274
	of which pseudogenes	70
	Coding density (%)	88.1
	Average gene length (bp)	1,091
	Ribosomal RNAs	19
	Transfer RNAs	73
	IS/transposon elements	24
Plasmid pBF9343	Size (bp)	36,560
	G+C content (%)	32.24
	CDSs	48
	Coding density (%)	85.0
	Average gene length (bp)	652

Table S2A - Invertible promoters.

Group	Average length of *fin* regions (bp)	Invertible region (IR)	Coordinates	Active in shotgun	Regulated CDSs
1	226	D	3032390.. 3032595	Yes	PS E polysaccharide biosynthesis region
		DD	894511.. 894721	No	PS G polysaccharide biosynthesis region.
		E	1634575.. 1634805	Yes	PS A polysaccharide biosynthesis region.
		F	4091660.. 4091889	Yes	PS H polysaccharide biosynthesis region.
		G	2211236.. 2211454	Yes	PS B polysaccharide biosynthesis region.
		H	4361354.. 4361586	Yes	PS D polysaccharide biosynthesis region.
		I	1806791.. 1807023	No	PS F polysaccharide biosynthesis region.
	161	A	89789.. 89949	No	Putative membrane protein.
		B	91906.. 92066	Yes	Hypothetical protein. Putative type I restriction-modification enzyme.

		C	129406.. 129566	Yes	Hypothetical protein. Membrane protein. Putative SusC homologue surface membrane protein.
		J	4866094.. 4866254	Yes	Conserved hypothetical protein. Hypothetical protein.
		K	4868436.. 4868596	No	Hypothetical protein. Putative membrane protein.
2	370	AA	603197.. 603530	Yes	Putative exported protein.
		L	5046156.. 5046613	Yes	Putative exported protein. Putative phosphoenolpyruvate carboxykinase.
		M	3790067.. 3790365	Yes	10kDa chaperonin GroES. Putative exported protein.
		N	5023725.. 5024147	Yes	Putative outer membrane protein. Putative pyruvate carboxylase biotin-containing subunit.
		O	2282840.. 2283132	Yes	Putative outer membrane protein.
		T	2093020.. 2094551	Yes	Putative gene cluster for degradation of sialoconjugates.
		V	4218300.. 4218701	No	Putative exported protein. Conserved hypothetical protein. Putative anti-sigma factor.
		W	3420744.. 3421217	No	Putative anti-sigma factor. Putative ECF-sigma factor, RpoE-like.
		X	3754426.. 3754869	No	Putative exported protein. Putative outer membrane protein.
		Y	3839227.. 3839616	Yes	Putative outer membrane protein. Putative outer membrane receptor protein.

		Z	1121569.. 1121954	No	Putative enoyl ACP-reductase.

Table S2B - Other invertible regions.

Invertible Region (IR)	Coordinates	Active in shotgun	CDSs involved	effect
BB (multiple inversions)	2147598.. 2151089	Yes	Putative type I restriction-modification endonuclease specificity subunit. BF1839, BF1840, BF1841, BF1842	Exchange of DNA-binding modules in specificity subunit
CC (multiple inversions)	708692.. 720999	Yes	Putative outer membrane proteins. BF0590, BF0592, BF0594, BF0596	Switching of alternative outer membrane proteins onto translational start signals (fixed promoter)
EE (multiple inversions)	1999949.. 2093942	Partially	Putative outer membrane proteins. BF1716, BF1719, BF1722, BF1798, BF1803	Switching of alternative outer membrane proteins to control by invertible promoter in IR-T
P	401492.. 403339	Yes	Putative two-component sensor histidine kinase/response regulator fusion.	Fusion and separation of sensor and phospho-relay components
Q	2831962.. 2833445	Yes	Hypothetical proteins. BF2439A, BF2439B	Alternative orientation of two hypothetical proteins (with or against direction of transcription of surrounding genes)
R	1073991.. 1079515	Yes	Putative outer membrane proteins. BF0865, BF0866	Alternative orientation of genes (with or against direction of transcription of surrounding genes
S	1079501.. 1087414	Yes	Putative outer membrane proteins. BF0867, BF0868 Putative RNA polymerase ECF-sigma factor and putative anti-sigma factor. BF0869, BF0870	Alternative orientation of genes

Table S3 – Polysaccharide biosynthesis operons.

Region	Coordinates	Variable promoter (IR)	Active in shotgun	Regulators	CDSs
PS A	1635470..1635940	E	Yes	UpaY / UpaZ	BF1367 – BF1377
PS B	2211582..2234865	G	Yes	UpbY / UpbZ	BF1893 – BF1914
PS C	1260915..1277396	Non-variable	-	UpcY / UpcZ	BF1009 – BF1026
PS D	4346276..4361140	H	Yes	UpdY / UpdZ	BF3683 – BF3699
PS E	3016482..3032253	D	Yes	UpeY / UpeZ	BF2591 – BF2606
PS F	1807211..1822424	I	No	UpfY / UpfZ	BF1549 – BF1565
PS G	894922..917651	DD	No	UpgY / UpgZ	BF0731 – BF0752
PS H	4076795..4091435	F	Yes	UphY / UphZ	BF3451 – BF3466
PS I	3249979..3278123	Non-variable	-	UpiY	BF2790 – BF2817
PS J	1987522..2004910	Non-variable	-	-	BF1706 - BF1718

Table S4 – Putative SusC homologues.

CDS	Coordinates	Product	Invertible region
BF0229	250479..253874	Putative SusC homologue outer membrane protein	No
BF0288	334274..337474	Putative SusC homologue TonB-dependent outer membrane protein	No
BF0334	397798..401148	Putative SusC homologue outer membrane protein	No
BF0341	412037..415084	Putative TonB-dependent outer membrane exported protein	No
BF0349	420462..423518	Putative TonB dependent outer membrane exported protein	No
BF0381	465995..469327	Putative exported protein	No
BF0501	581461..584628	Putative outer membrane protein	No
BF0518	603565..606753	Putative outer membrane protein	No
frrG (BF0536)	629760..633170	Putative outer membrane protein	No
BF0571	680148..683426	Putative TonB-linked outer membrane receptor protein	No
BF0578	694072..696408	Putative TonB-dependent outer membrane receptor protein	No

BF0590	705773..708751	Putative SusC homologue surface membrane protein	Yes
BF0592	710759..713614	Putative SusC homologue surface membrane protein	Yes
BF0594	715846..718878	Putative SusC homologue surface membrane protein	Yes
BF0596	720940..723876	Putative SusC homologue surface membrane protein	Yes
BF0661	797339..800476	Putative outer membrane protein	No
BF0681	828289..831702	Putative outer membrane protein	No
BF0711	871405..874152	Putative TonB-dependent outer membrane receptor protein	No
BF0759	924125..927259	Putative outer membrane protein	No
BF0807	989559..992741	Putative outer membrane protein	No
BF0864	1070528..1073572	conserved hypothetical protein	No
BF0866	1075945..1079352	Putative outer membrane protein	No
BF0868	1081426..1084734	Putative TonB-linked outer membrane protein	No
BF0871	1087602..1090943	Putative exported protein	No
BF0890	1111869..1115135	Putative outer membrane receptor protein	No
BF0893	1118124..1121387	Putative outer membrane receptor protein	No
BF0971	1213148..1216468	Putative outer membrane protein	No
BF0977	1222361..1225738	Putative TonB-dependent outer membrane receptor protein	No
BF1204	1463991..1467359	Putative outer membrane protein	No
BF1310	1570137..1573439	Putative outer membrane protein	No
BF1415	1679751..1683143	Putative outer membrane protein	No
BF1512	1771475..1774855	Putative outer membrane protein	No
BF1618	1886174..1889299	Putative outer membrane receptor protein	No
BF1716	1999994..2002084	Putative SusC homologue outer membrane protein	Yes
BF1719	2005661..2008927	Putative SusC homologue outer membrane protein	Yes
BF1722	2012226..2015546	Putative SusC homologue outer membrane protein	Yes
BF1798	2081856..2085086	Putative SusC homologue outer membrane protein	Yes
BF1803	2090601..2093867	Putative SusC homologue outer membrane protein	Yes
BF1804	2094629..2097796	Putative TonB-dependent outer membrane receptor protein	No
BF1816	2119766..2122507	Putative outer membrane protein	No

BF1956	2283190..2286564	Putative outer membrane protein	No
BF1992	2330851..2334225	Putative outer membrane protein	No
BF2044	2387138..2389924	Putative TonB-dependent outer membrane receptor protein	No
BF2084	2437288..2439492	Putative TonB-dependent outer membrane receptor protein	No
BF2195	2557432..2560734	Putative exported protein	No
BF2270	2654934..2658200	Putative exported protein	No
BF2697	3136811..3139111	Putative exported TonB-dependent receptor protein	No
BF2708	3149210..3151351	Putative exported TonB-dependent receptor protein	No
BF2907	3376170..3379502	Putative exported protein	No
BF2942	3416623..3419997	Putative exported protein	No
BF3024	3515789..3518941	Putative exported protein	No
BF3097	3612821..3616255	Putative exported protein	No
BF3146	3670784..3673789	Putative exported protein	No
BF3199	3754969..3758121	Putative exported protein	No
BF3258	3835857..3839033	Putative exported protein	No
BF3307	3899923..3903075	Putative exported protein	No
omp117 (BF3412)	4024110..4027376	Putative outer membrane protein	No
BF3444	4066003..4069050	Putative membrane protein	No
BF3572	4206914..4210357	Putative membrane protein	No
BF3576	4214892..4218104	Putative exported protein	No
BF3581	4222187..4225660	Putative membrane protein	No
BF3642	4297965..4300640	Putative exported protein	No
BF3712	4374951..4378244	Putative exported protein	No
BF3724	4392109..4395414	Putative exported protein	No
BF3746	4419207..4422644	Putative membrane protein	No
BF4056	4769105..4772347	Putative outer membrane protein	No
BF4062	4781654..4784647	Putative TonB-linked outer membrane protein	No
BF4132	4860945..4864271	Putative outer membrane protein	No
BF4169	4911410..4914403	Putative TonB-linked outer membrane protein	No
BF4178	4932416..4935583	Putative outer membrane protein	No
BF4246	5030783..5033878	Putative outer membrane protein	No

BF4248	5030783..5033878	Putative outer membrane protein	No
BF4256	5046824..5050081	Putative outer membrane protein	No
BF4268	5067453..5070962	Putative outer membrane protein (pseudogene)	No
BF4323	5135539..5138736	Putative TonB-dependent outer membrane receptor protein	No

Supporting References

1. R. L. Van Tassell, T. D. Wilkins, *Can J Microbiol* **24**, 1619-21 (1978).
2. S. Patrick, M. J. Larkin, in *Microbial biofilms: Formation and Control* S. P. Denyer, Ed. (Blackwell Scientific Publications, Oxford, 1993) pp. 109-131.
3. F. M. Ausubel *et al.*, *Short Protocols in Molecular Biology 2nd Edition*. J. Wiley, Ed. (Chichester, 1992).
4. J. Parkhill *et al.*, *Nature* **404**, 502-6 (2000).
5. K. Rutherford *et al.*, *Bioinformatics* **16**, 944-5 (2000).
6. J. Xu *et al.*, *Science* **299**, 2074-6 (2003).
7. J. Parkhill *et al.*, *Nature* **413**, 848-52 (2001).
8. J. Parkhill *et al.*, *Nature* **413**, 523-7 (2001).
9. S. A. Hanley, J. Aduse-Opoku, M. A. Curtis, *Infect Immun* **67**, 1157-71 (1999).
10. K. H. Cho, A. A. Salyers, *J Bacteriol* **183**, 7224-30 (2001).
11. T. Komano, *Annual Review of Genetics* **33**, 171-191 (1999).
12. G. S. Moeck, J. W. Coulton, *Mol Microbiol* **28**, 675-81 (1998).
13. B. R. Otto, J. G. Kusters, J. Luirink, F. K. de Graaf, B. Oudega, *Infect Immun* **64**, 4345-50 (1996).
14. V. H. Varel, M. P. Bryant, *Appl Microbiol* **28**, 251-7 (1974).
15. S. Patrick, in *Molecular Medical Microbiology* M. Sussman, Ed. (Academic Press, London, 2002) pp. 1921-1948.
16. H. Tanaka, F. Ito, T. Iwasaki, *J Biochem (Tokyo)* **115**, 318-21 (1994).
17. J. R. Baker, S. Dong, D. G. Pritchard, *Biochem J* **365**, 317-22 (2002).
18. M. J. Pallen, A. C. Lam, N. Loman, *Trends Microbiol* **9**, 518-21 (2001).
19. S. Patrick, Coffey, A., Emmerson, A.M., and Larkin, M.J., *FEMS Microbiology Letters* **50**, 67-71 (1988).
20. S. Patrick, B. I. Duerden, in *Principles and Practice of Clinical Bacteriology* S. H. G. Hawkey, Ed. (J. Wiley, London, 2004).
21. R. C. Wang *et al.*, *J Mol Biol* **217**, 441-54 (1991).
22. S. Patrick, J. P. McKenna, S. O'Hagan, E. Dermott, *Microb Pathog* **20**, 191-202 (1996).
23. A. J. Cooper, A. P. Kalinowski, N. B. Shoemaker, A. A. Salyers, *J Bacteriol* **179**, 6221-7 (1997).
24. H. J. Goodman, D. R. Woods, *Gene* **94**, 77-82 (1990).
25. G. A. Cromie, J. C. Connelly, D. R. Leach, *Mol Cell* **8**, 1163-74 (2001).
26. M. A. Petit, and Erlich, D., *EMBO Jounal* **21**, 3137-3147 (2002).
27. A. Quiberoni *et al.*, *Res Microbiol* **152**, 131-9 (2001).
28. M. S. Dillingham, M. Spies, S. C. Kowalczykowski, *Nature* **423**, 893-7 (2003).