Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*

Claire M. Fraser*, Sherwood Casjens†, Wai Mun Huang†, Granger G. Sutton*, Rebecca Clayton*, Raju Lathigra‡, Owen White*, Karen A. Ketchum*, Robert Dodson*, Erin K. Hickey*, Michelle Gwinn*, Brian Dougherty*, Jean-Francois Tomb*, Robert D. Fleischmann*, Delwood Richardson*, Jeremy Peterson*, Anthony R. Kerlavage*, John Quackenbush*, Steven Salzberg*, Mark Hanson‡, Rene van Vugt†, Nanette Palmer†, Mark D. Adams*, Jeannine Gocayne*, Janice Weidman*, Teresa Utterback*, Larry Watthey*, Lisa McDonald*, Patricia Artiach*, Cheryl Bowman*, Stacey Garland*, Claire Fujii*, Matthew D. Cotton*, Kurt Horst*, Kevin Roberts*, Bonnie Hatch*, Hamilton O. Smith* & J. Craig Venter*

** The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA*
† Division of Molecular Biology and Genetics, Department of Oncological Sciences, University of Utah, Salt Lake City, Utah 84132, USA
‡ MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, Maryland 20878, USA

The genome of the bacterium *Borrelia burgdorferi* B31, the aetiologic agent of Lyme disease, contains a linear chromosome of 910,725 base pairs and at least 17 linear and circular plasmids with a combined size of more than 533,000 base pairs. The chromosome contains 853 genes encoding a basic set of proteins for DNA replication, transcription, translation, solute transport and energy metabolism, but, like *Mycoplasma genitalium*, it contains no genes for cellular biosynthetic reactions. Because *B. burgdorferi* and *M. genitalium* are distantly related eubacteria, we suggest that their limited metabolic capacities reflect convergent evolution by gene loss from more metabolically competent progenitors. Of 430 genes on 11 plasmids, most have no known biological function; 39% of plasmid genes are paralogues that form 47 gene families. The biological significance of the multiple plasmid-encoded genes is not clear, although they may be involved in antigenic variation or immune evasion.

In the mid-1970s, a geographic clustering of an unusual rheumatoid arthritis-like condition was reported in Connecticut[1]. That cluster of cases focused attention on the syndrome that is now called Lyme disease. It was subsequently realized that a similar disorder had been known in Europe since the beginning of this century. Lyme disease is characterized by some or all of the following manifestations: an initial erythematous annular rash, 'flu-like symptoms, neurological complications, and arthritis in about 50% of untreated patients[2]. In the United States, the disease occurs primarily in northeastern and midwestern states, and in western parts of California and Oregon. These regions coincide with the ranges of various species of *Ixodes* ticks, the primary vector of Lyme disease. Lyme disease is now the most common tick-transmitted illness in the United States, and has been reported in many temperate parts of the Northern Hemisphere.

It was not until the early 1980s that a new spirochaete, *Borrelia burgdorferi*[3], was isolated and cultured from the midgut of *Ixodes* ticks, and subsequently from patients with Lyme disease[4,5]. Analysis of genetic diversity among individual *Borrelia* isolates has defined a closely related cluster containing at least 10 tick-borne species of Lyme disease agents, called '*B. burgdorferi* (*sensu lato*)'. *B. burgdorferi* resembles most other spirochaetes in that it is a highly specialized, motile, two-membrane, spiral-shaped bacterium that lives primarily as an extracellular pathogen. *Borrelia* is fastidious and difficult to culture *in vitro*, requiring a specially enriched media and low oxygen tension[6].

One of the most striking features of *B. burgdorferi* is its unusual genome, which includes a linear chromosome approximately one megabase in size[7–10] and numerous linear and circular plasmids[11–13], with some isolates containing up to 20 different plasmids. The plasmids have a copy number of approximately one per chromosome[10,14], and different plasmids often appear to share regions of homologous DNA[13,15,16]. Long-term culture of *B. burgdorferi* results in the loss of some plasmids, changes in protein expression profiles, and a loss in the ability of the organism to infect laboratory animals, suggesting that the plasmids encode important proteins involved in virulence[17–19].

Because of its importance as a pathogen of humans and animals, and the value of complete genome sequence information for understanding its life cycle and advancing drug and vaccine development, we sequenced the genome of *B. burgdorferi* type strain (B31), using the random sequencing method previously described[20–24]. Here we summarize the results from sequencing, assembly and analysis of the linear chromosome and 11 plasmids.

Chromosome analysis

The linear chromosome of *B. burgdorferi* has 910,725 base pairs (bp) and an average G+C content of 28.6%. Base pair one represents the first double-stranded base pair that we observed at the left telomere. Previous genome characterizations agree with the nucleotide sequence of the large chromosome[10,25–28]. The 853 predicted coding sequences (open reading frames; ORFs) have an average size of 992 bp, similar to that observed in other prokaryotic genomes, with 93% of the *B. burgdorferi* genome representing

Figure 1 Linear representations of the *B. burgdorferi* B31 chromosome and plasmids. The location of predicted coding regions colour-coded by biological role, RNA genes, and tRNAs is indicated. Arrows represent the direction of transcription for each predicted coding region. Numbers associated with tRNA symbols represent the number of tRNAs at a locus. Numbers associated with GES represent the number of membrane-spanning domains according to the Goldman, Engelman and Steitz scale as calculated by TopPred[49]. Only proteins with five or more GES are indicated. Members of paralogous gene families are identified by family number. Transporter abbreviations: mal, maltose; P, gly and bet, proline, glycine, betaine; glyc, glycerol; aa, amino acid; E, glutamate; fru, fructose; glu, glucose; s/p, spermidine/putrescine; pan, pantothenate; Pi, phosphate; lac, lactate; rib, ribose; ?, unknown. ►

Nature © Macmillan Publishers Ltd 1997

coding sequence. Biological roles were assigned to 59% of the 853 ORFs using the classification scheme adapted from Riley[29] (Fig. 1), 12% of ORFs matched hypothetical coding sequences of unknown function from other organisms, and 29% were new genes. The average relative molecular mass (M_r) of the chromosome-encoded proteins in *B. burgdorferi* is 37,529 ranging from 3,369 to 254,242, values similar to those observed in other bacteria including *Haemophilus influenzae*[20] *and Mycoplasma genitalium*[21]. The median isoelectric point (pI) for all predicted proteins is 9.7.

Analysis of codon usage in *B. burgdorferi* reveals that all 61 triplet codons are used. When both AU- and GC-containing codons specify a single amino acid, there is a marked bias (from 2-fold to more than 20-fold, depending on the amino acid) in the use of AU-rich codons. The most frequently used codons are AAA (Lys, 8.1%), AAU (Asn, 5.9%), AUU (Ile, 5.9%), UUU (Phe, 5.7%), GAA (Glu, 5.0%), GAU (Asp, 4.2%) and UUA (Leu, 4.2%). The most common amino acids are Ile (10.6%), Leu (10.3%), Lys (10.2%), Ser (7.8%) and Asn (7.2%). The high value for Lys is in agreement with the median calculated isoelectric point of 9.7.

Plasmid analysis

Analysis of the nucleotide sequence and Southern analyses on *B. burgdorferi* DNA indicate that, in addition to the large linear chromosome, isolate B31 contains linear plasmids of the following approximate sizes: 56 kilobase pairs (kbp) (lp56), 54 kbp (lp54), four plasmids of 28 kbp (lp28-1, lp28-2, lp28-3 and lp28-4), 38 kbp (lp38), 36 kbp (lp36), 25 kbp (lp25) and 17 kbp (lp17); and circular plasmids of the following sizes: 9 kbp (cp9), 26 kbp (cp26) and five or six homologous plasmids of 32 kbp (cp32). These include all of the plasmids previously identified in this strain, but comparisons with other B31 cultures suggest that this isolate may have lost one 21 kbp linear and one or two 32 kbp circular plasmids during growth in culture since its original isolation[11–14,19,30]. The sequences of all plasmids were assembled as part of this project. However, the assembled sequences of the cp32 and related lp56 plasmids could not be determined with a high degree of confidence because of DNA sequence similarity among them (≥99% in several regions of 3,000–5,000 bp per plasmid)[13,16] (Table 1). Improved assembly strategies are being tested to achieve closure on these plasmids (G. Sutton, unpublished). Plasmid lp17 is identical to that of lp16.9 from Barbour *et al.*[15].

The 11 plasmids we have described contain a total of 430 putative ORFs with an average size of 507 bp; plasmid G+C content ranges from 23.1% to 32.3%. Only 71% of plasmid DNA represents predicted coding sequences, a value significantly lower than that on the chromosome. This indicates that average intergenic distances are greater in the plasmids than in the chromosome, and that many potential ORFs contain authentic frameshifts or stops (see E29, for example), suggesting that they are decaying genes not encoding functional proteins. Of the 430 plasmid ORFs, only 70 (16%) could be identified and these include membrane proteins such as OspA-D, decorin-binding proteins, the VlsE lipoprotein recombination cassette, and the purine ribonucleotide biosynthetic enzymes GuaA and GuaB. We found that 100 ORFs (23%) match other hypothetical proteins from plasmids in this and related strains of *B. burgdorferi*[15,16,31]; 10 ORFs (2.3%) match hypothetical proteins from species other than *Borrelia*; and 250 ORFs (58%) have no database match.

We found that 47 paralogous gene families containing from 2 to 12 members account for 39% (169 ORFs) of the plasmid-encoded genes with no known biological role (Fig. 1). Paralogue families 32 and 50, typified by previously identified *B. burgdorferi* plasmid genes cp32 orfC and cp8.3 orf2, respectively, have some similarities to proteins involved in replication, segregation and control of copy number in other bacterial systems[16,31]. Previous studies have reported examples of plasmid gene duplication, but the extent of

Table 1 Genome features in *Borrelia burgdorferi*

Chromosome	910,725 bp (28.6% G+C)
Coding sequences (93%)	
RNAs (0.7%)	
Intergenic sequence (6.3%)	
853 coding sequences	
500 (59%) with identified database match	
104 (12%) match hypothetical proteins	
249 (29%) with no database match	
Plasmids	
cp9	9,386 bp (23.6% GC)
cp26	26,497 bp (26.3% GC)
lp17	16,828 bp (23.1% GC)
lp25	24,182 bp (23.3% GC)
lp28-1	26,926 bp (32.3% GC)
lp28-2	29,771 bp (31.5% GC)
lp28-3	28,605 bp (25.1% GC)
lp28-4	27,329 bp (24.4% GC)
lp36	36,834 bp (26.8% GC)
lp38	38,853 bp (26.1% GC)
lp54	53,590 bp (28.1% GC)
Coding sequences (71%)	
Intergenic sequence (29%)	
430 coding sequences	
70 (16%) with identified database match	
110 (26%) match hypothetical proteins	
250 (58%) with no database match	
Ribosomal RNA	Chromosome coordinates
16S	444581–446118
23S	438590–441508
5S	438446–438557
23S	435334–438267
5S	435201–435312
Stable RNA	
tmRNA	46973–47335
mpB	750816–751175
Transfer RNA	
34 species (8 clusters, 14 single genes)	

*The telomeric sequences of the nine linear plasmids assembled as part of this study were not determined; estimation of the number of missing terminal nucleotides by restriction analysis suggests that less than 1,200 bp is missing in all cases. Comparisons with previously determined sequences of lp 16.9 and one terminus of lp28-1 indicate that 25, 60 and 1,200 bp are missing, respectively.

◀ **Table 2** Gene identification numbers are listed with the prefix BB as in Fig. 2. Each gene identified is listed in its functional role category (adapted from Riley[29]). The percentage of similarity and a two-letter abbreviation for genus and species for the best match are also shown. An expanded version of this table with additional information is available on the World-Wide Web at http://www.tigr.org/tdb/mdb/bbdb/bbdb.htm. Abbreviations of gene names are: Ac, acetyl; BP, binding protein; biosyn, biosynthesis; cello, cellobiose; CPDase, carboxypeptidase; Dcase, decarboxylase; DHase, dehydrogenase; flgr, flagellar/flagellum; fru, fructose; GBP, glycine, betaine, L-proline; glu, glucose; Kase, kinase; mal, maltose; MC-methyl-accepting chemotaxis; MTase, methyltransferase; NAG, *N*-acetylglucosamine; OH, hydroxy; OP, oligopeptide; P, phosphate; PPTase, phosphotransferase; PPase, phosphatase; prt, protein; put, putative; RDase, reductase; RG, ribose/galactose; SAM, *S*-adenosylmethionine; Sase, synthetase/synthase; SP, spermidine/putrescine; ss, single-stranded; sub, subunit; Tase, transferase.
Abbrevation of genus and species are: Ah, *Aeromonas hydrophila*; Ar, *Agrobacterium radiobacter*; Al, *Alteromonas* sp.; Ab, *Anabaena* sp.; An, *Anacystis nidulans*; At, *arabidopsis thaliana*; Av, *Azotobacter vinelandii*; Bf, *Bacillus firmus*; Bl, *Cacillus licheniformis*; Bm, *Bacillus megaterium*; Bs, *Bacillus stearothermophilus*; Bs, *Bacillus subtilis*; Bb, *Borrelia burgdorferi*; Bc, *Borrelia coriaceae*; Bh, *Borrelia hermsii*; Ba, *Buchnera aphidicola*; Ca, *Clostridium acetobutylicum*; Cl, *Clostridium longisporum*; Cp, *Clostridium perfringens*; Cg, *Corynegacterium glutamicum*; Cb, *Coxiella burnetii*; Cp, *Cyanophora paradoxa*; Dd, *Dictyostelium discoideum*; Ec, *Escherichia coli*; Eh, *Entamoeba histolytica*; Ec, *Enterobacter cloacae*; El, *Enterococcus faecalis*; Eh, *Enterococcus hirae*; Ha, *Haemophilus aegyptius*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Hs, *Homo sapiens*; La, *Lactobacillus acidophilus*; Ll, *Lactococcus lactis*; Li, *Leptospira interrogans serovar lai*; Mj, *Methanococcus jannaschii*; Mb, *Methanosarcina barkeri*; Ml, *Mycobacterium leprae*; Mt, *Mycobacterium tuberculosis*; Mc, *Mycoplasma capricolum*; Mg, *Mycoplasma genitalium*; Mh, *Mycoplasma hominis*; Mh, *Mycoplasma hyorhinis*; Mm, *Mycoplasma mycoides*; Mp, *Mycoplasma pneumoniae*; Mx, *Myxococcus xanthus*; Ng, *Neisseria gonorrhoeae*; Nm, *Neisseria meningitidis*; Os, *Odontella sinensis*; Pt, *Paramecium tetraurelia*; Pa, *Pediococcus acidilactici*; Pf, *Plasmodium falciparum*; Pg, *Porphyromonas gingivalis*; Pv, *Proteus vulgaris*; Pa, *Pseudomonas aeruginosa*; Pm, *Pseudomonas mevalonii*; Pp, *Pseudomonas putida*; Rm, *Rhizobium meliloti*; Rc, *Rhodobacter capsulatus*; Rs, *Rhodobacter sphaeroides*; Rp, *Rickettsia prowazekii*; Sc, *Saccharomyces cerevisiae*; Sc, *Salmonella choleraesius*; St, *Salmonella typhimurium*; Sh, *Serpulina hyodysenteriae*; Sd, *Shigella dysenteriae*; So, *Spinacia oleracea*; Sc, *Staphylococcus camosus*; Se, *Staphylococcus epidermidis*; Sp, *Streptococcus pyogenes*; Sc, *Streptomyces coelicolor*; Ss, *Sulfolobus solfataricus*; Syn, *Synechococcus* sp.; Sp, *Synechocystis* PCC6803; Tt, *Thermoanaerobacterium thermosaccharolyticum*; Tb, Thermophilic bacterium RT8.B4.; Ttv, *Thermoproteus tenax* virus; Tm, *Thermotoga maritima*; Tat, *Thermus aquaticus thermophilus*; Ta, *Thermus aquaticus*; Td, *Treponema denticola*; Tp, *Treponema pallidum*; Ta, *Triticum aestivum*; Tb, *Trypanosoma brucei* mitochondrion; Vc, *Vibrio cholerae*; Vp, *Vibrio parahaemolyticus*; Zm, *Zymomonas mobilis*.

Nature © Macmillan Publishers Ltd 1997

Nature © Macmillan Publishers Ltd 1997

this redundancy has become even more apparent with the complete sequence of these 11 plasmids from isolate B31. Moreover, a preliminary search of 221 putative ORFs from the cp32s and lp56 indicates that at least 50% display ≥70% amino-acid similarity to ORFs from the other 11 plasmids presented here (data not shown). Although plasmid-encoded genes have been implicated in infectivity and virulence[17–19], the biological roles of most of these genes are not known. The significance of the large number of paralogous plasmid-encoded genes is not understood. These proteins may be expressed differentially in tick and mammalian hosts, or may undergo homologous recombination to generate antigenic variation in surface proteins. This hypothesis is supported by the identification of 63 plasmid-encoded putative membrane lipoproteins (Fig. 1).

Several copies of a putative recombinase/transposase similar to IS891-like transposases were identified in the *B. burgdorferi* plasmids. Linear plasmid 28-2 contains one full-length copy of this gene. Although no inverted repeats were found on either side of the transposase, there is a putative ribosome-binding site several nucleotides upstream of the apparent start codon, and a stem–loop structure ($-27\,\mathrm{kcal\,mol^{-1}}$) 195 bp downstream of the stop codon in an area with no ORFs. This transposase might represent a functional gene important for the frequent DNA rearrangements that presumably occur in *Borrelia* plasmids. There are other partial or nearly complete copies of the transposase gene that contain frame-destroying mutations elsewhere in the genome: two copies on lp17, one on lp36, one on lp38, one on lp28-3, two on lp28-1, and one near the right end of the large linear chromosome.

Origin of replication

The replication mechanism for the linear chromosome and plasmids in *B. burgdorferi* is not yet known. Replication possibly begins at the termini, as has been proposed for the poxvirus hairpin telomeres[32], or may begin from a single origin somewhere along the length of the linear replicon. Of the genes on the linear chromosome, 66% are transcribed away from the centre of the chromosome (Fig. 1), similar to the transcriptional bias observed for the genomes of *M. genitalium*[21] and *M. pneumoniae*[33]. It has been suggested that bacterial genes are optimally transcribed in the same direction as that in which replication forks pass over them, particularly for highly transcribed genes[34,35]

Given the transcriptional bias observed in *B. burgdorferi*, it seems likely that the origin of replication is near the centre of the chromosome. Because bacterial chromosomal replication origins are usually near *dnaA*[36], it is intriguing to note that this gene (BB437) lies almost exactly at the centre of the linear *B. burgdorferi* chromosome[10,27]. A centrally initiated, bi-directional replication fork would be equidistant from the two chromosome ends, and replication would traverse the rRNA genes in the same direction as transcription.

An analysis of GC skew, (G − C)/(G + C) calculated in 10-kilobase (kb) windows across the chromosome, shows a clear break at the putative origin of replication. The GC-skew values are uniformly negative from 0 to 450 kb (minus strand), and uniformly positive (plus strand) from 450 kb to the end of the chromosome (Fig. 2). Additional evidence for the location of the origin of replication comes from our discovery of an octamer, TTGTTTTT, whose skewed distribution in the plus versus the minus strand of the chromosome matches the GC skew (Fig. 2). The biological significance of this octamer has not yet been determined, although it may be analogous to the Chi site in *Escherichia coli* that is implicated in *recBCD* mediated recombination. No GC skew was observed in any of the plasmids, although the heptamer ATTTTTT displays a skewed distribution in the plus versus the minus strand of lp28-4 that changes at the approximate midpoint of the plasmid (not shown).

Transcription and translation

Genes encoding the three subunits (α, β, β') of the core RNA polymerase were identified in *B. burgdorferi* along with σ^{70} and two alternative σ factors, σ^{54} and *rpoS*. The role and specificity of each of these σ factors in transcription regulation in *B. burgdorferi* are not known. The *nusA*, *nusB* and *rho* genes, which are involved in transcription elongation and termination, were also identified.

A region of the genome with a significantly higher G + C content (43%), located between nucleotides 434,000 and 447,000, contains the rRNA operon. As previously reported, the rRNA operon in *B. burgdorferi* contains a 16S rRNA–Ala-tRNA–Ile-tRNA–23S rRNA–5S rRNA–23S rRNA–5S rRNA[37,38]. All of the genes are present in the same orientation, except for that encoding Ile tRNA. Four unrelated genes, encoding 3-methyladenine glycosylase, hydrolyase and two with no database match, are also present in the rRNA operon. Three of these genes are transcribed in the same direction as the rRNAs.

We identified in the chromosome 31 tRNAs with specificity for all 20 amino acids (Fig. 1). These are organized into 7 clusters plus 13 single genes. All tRNA synthetases are present except glutaminyl tRNA-synthetase. A single glutamyl tRNA synthetase probably aminoacylates both $tRNA^{Glu}$ and $tRNA^{Gln}$ with glutamate followed by transamidation by Glu-tRNA amidotransferase, a heterotrimeric enzyme present in *B. burgdorferi* and several Gram-positive bacteria and archaea[30]. The lysyl-tRNA synthetase (LysS) in *B. burgdorferi* is a class I type that has no resemblance to any known bacterial or eukaryotic LysS, but is most similar to LysS from the archaea[40].

Replication, repair and recombination

The complement of genes in *B. burgdorferi* involved in DNA replication is smaller than in *E. coli*, but similar to that in *M. genitalium*[21]. Three ORFs have been identified with high homology to four of the ten polypeptides in the *E. coli* DNA polymerase III: α, β and γ, and τ. In *E. coli*, the γ and τ proteins are produced by programmed ribosomal frameshifting. This observation suggests that DNA replication in *B. burgdorferi*, like that in *M. genitalium*, is accomplished with a restricted set of genes. *B. burgdorferi* has one

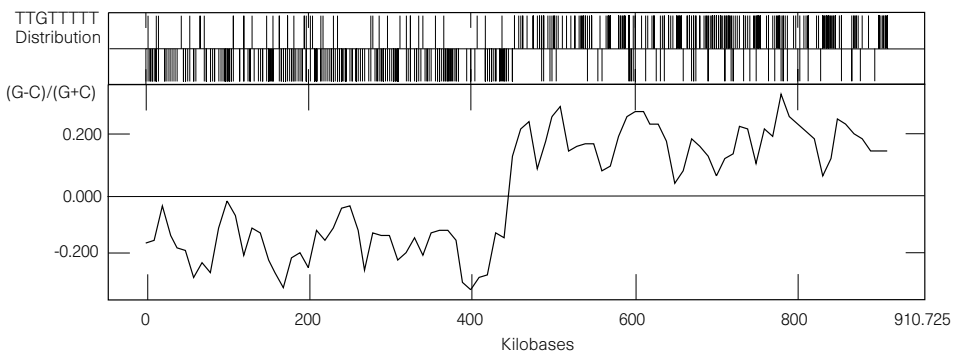


Figure 2 Distribution of TTGTTTTT and GC skew in the *B. burgdorferi* chromosome. Top, distribution of the octamer TTGTTTTT. The lines in the top panel represent the location of this octamer in the plus strand of the sequence, and those in the second panel represent the location of this oligomer in the minus strand of the sequence. Bottom, GC skew.

Nature © Macmillan Publishers Ltd 1997

type I topoisomerase (*topA*) and two type II topoisomerases (gyrase and topoisomerase IV) for DNA topology management and chromosome segregation, despite its linear chromosomal structure. This suggests that topoisomerase IV may be required for more than the separation of circular DNAs during segregation.

The DNA repair mechanisms in *B. burgdorferi* are similar to those in *M. genitalium*. DNA excision repair can presumably occur by a pathway involving endonuclease III, PolI and DNA ligase. The genes for two of three DNA mismatch repair enzyme (*mutS*, *mutL*) are present. The apparent absence of *mutH* is consistent with the lack of GATC (*dam*) methylation in strain B31 (S. Casjens, unpublished). Also present are genes for the repair of ultraviolet-induced DNA damage (*uvrA*, *uvrB*, *uvrC* and *uvrD*) (Table 2).

B. burgdorferi has a complete set of genes to perform homologous recombination, including *recA*, *recBCD*, *sbcC*, *sbcD*, *recG*, *ruvAB* and *recJ*. 3′-Exonuclease activity associated with *sbcB* in *E. coli* may be encoded by *exoA* (exodeoxynuclease III). Although *recA* is present, we found no evidence for *lexA*, which encodes the repressor that

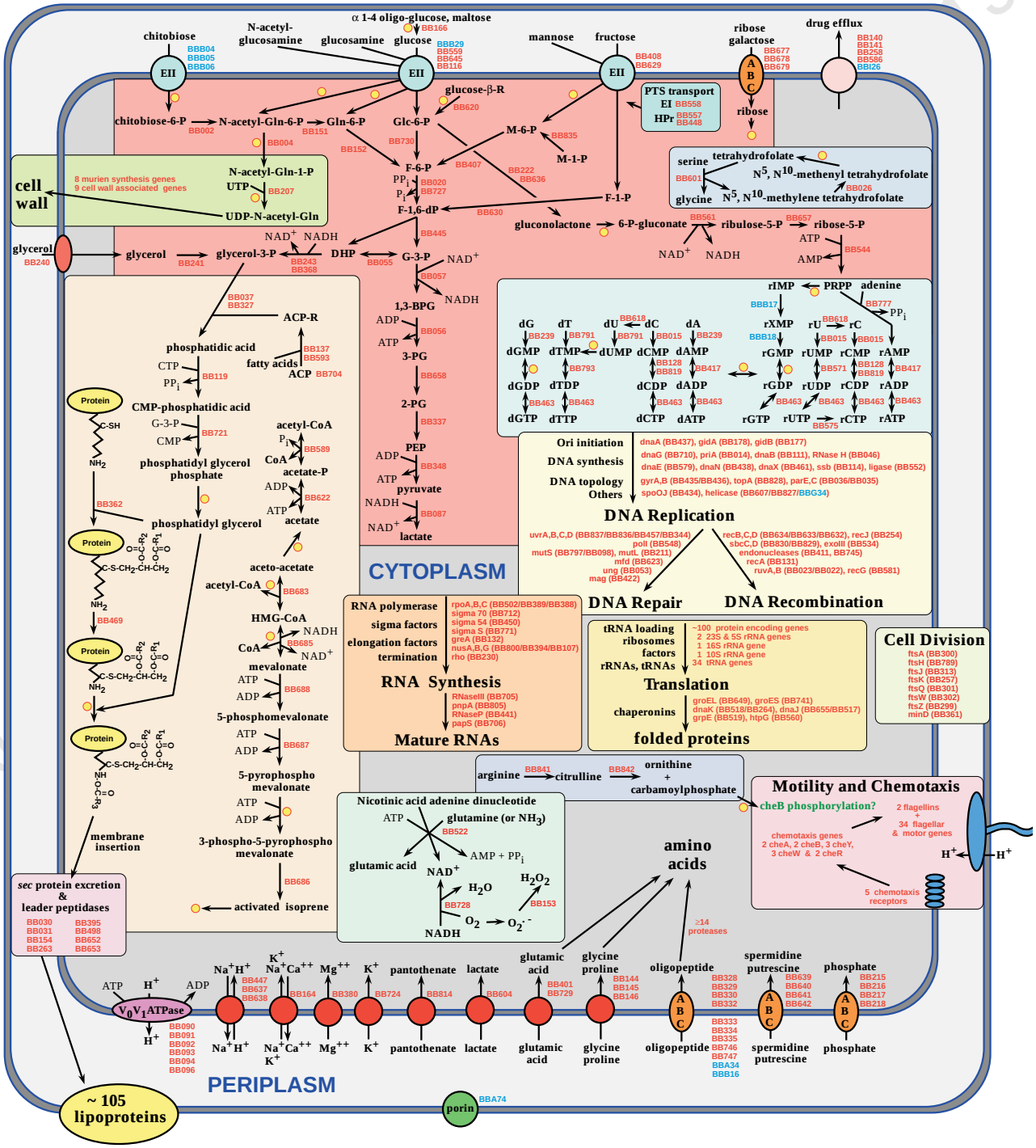


Figure 3 Solute transport and metabolic pathways in *B. burgdorferi*. A schematic diagram of a *B. burgdorferi* cell providing an integrated view of the transporters and the main components of the metabolism of this organism, as deduced from the genes identified in the genome. The ORF numbers correspond to those listed in Table 2 (red indicates chromosomal and blue indicates plasmid ORFs). Presumed transporter specificity is indicated. Yellow circles indicate: places where particular uncertainties exist as to the substrate specificity, subcellular location or direction of catalysis: or expected activities that were not found.

Nature © Macmillan Publishers Ltd 1997

regulates SOS genes in *E. coli*. No genes encoding DNA restriction or modification enzymes are present.

Biosynthetic pathways

The small genome size of *B. burgdorferi* is associated with an apparent absence of genes for the synthesis of amino acids, fatty acids, enzyme cofactors, and nucleotides, similar to that observed with *M. genitalium*[21] (Fig. 3, Table 2). The lack of biosynthetic pathways explains why growth of *B. burgdorferi in vitro* requires serum-supplemented mammalian tissue-culture medium. This is also consistent with previous biochemical data indicating that *Borrelia* lack the ability to elongate long-chain fatty acids, such that the fatty-acid composition of *Borrelia* cells reflects that present in the growth medium[6].

Transport

The linear chromosome of *B. burgdorferi* contains 46 ORFs and the plasmids contain 6 ORFs that encode transport and binding proteins (Fig. 3, Table 2). These gene products contribute to 16 distinct membrane transporters for amino acids, carbohydrates, anions and cations. The distribution of transporters between the four categories of functions in this section is similar to that observed in other heterotrophs (such as *Haemophilus influenzae*, *M. genitalium* and *H. pylori*), with most being dedicated to the import of organic compounds.

There are marked similarities between the transport capacity of *B. burgdorferi* and *M. genitalium*. Both genomes have a limited number of recognizable transporters, so it is not clear how they can sustain diverse physiological reactions. Several of the identified transporters in both genomes exhibit broad substrate specificity, exemplified by the oligopeptide ABC transporter (*opp* operon) or the glycine, betaine, L-proline transport system (*proVWX*). Therefore, these organisms probably compensate for their restricted coding potential by producing proteins that can import a wide variety of solutes. This is important because *B. burgdorferi* is unable to synthesize any amino acids *de novo*. We were unable to identify any transport systems for nucleosides, nucleotides, NAD/NADH or fatty acids, although they are likely to be present.

Glucose, fructose, maltose and disaccharides seem to be acquired by the phosphoenolpyruvate:phosphotransferase system (PTS). The two nonspecific components, enzyme 1 (*ptsI*) and Hpr (*ptsH*), are associated in one operon with an apparently glucose-specific, phosphohistidine-sugar phosphotransferase enzyme IIA (*crr*). Separate from this operon are four permeases (enzyme IIBC), *fruA* in two copies (fructose), *ptsG* (glucose) and *malX* (glucose/maltose) (Fig. 3, Table 2). The fructose-specific enzyme IIA is induced in the ORF with IIBC (*fruA*), as has been observed in *M. genitalium*[41]. Ribose may be imported by an ATP-binding cassette transporter (*rbsAC*). The *rbsAC* genes are transcribed in an operon with a methyl-accepting chemotaxis protein that may respond to β-galactosides, suggesting that movement of the organisms towards sugars may be coupled to the transport process.

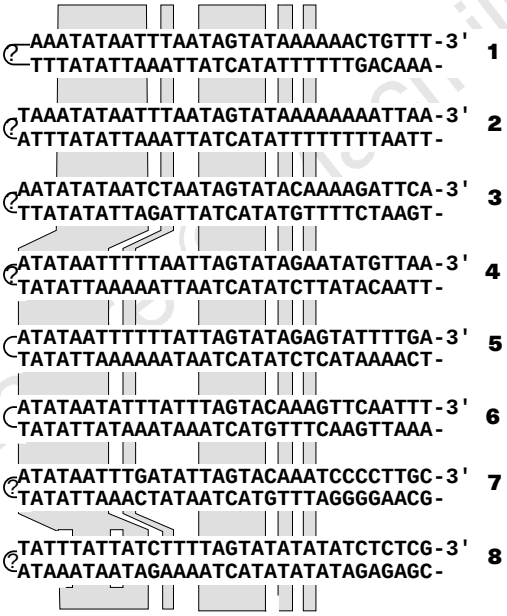


Figure 4 Telomere nucleotide sequences from *Borrelia* species. Nucleotide sequences are shown for known *Borrelia* telomeres as indicated: 1, *B. burgdorferi* Sh-2-82 chromosome left end; 2, *B. burgdorferi* B31 chromosome left end; 3, *B. afzelii* R-IP3 chromosome right end; 4, *B. burgdorferi* B31 chromosome right end; 5, *B. burgdorferi* B31 plasmid lp17 left end; 6, *B. burgdorferi* B31 plasmid lp17 right end; 7, *B. hermisii* plasmids bp7E and pb21E right ends; 8, *B. burgdorferi* B31 plasmid lp28-1 right end. In each case the telomere is at the left. Question marks (?) indicate locations where S1 nuclease was used to open terminal hairpins during the sequence determinations. Stippled areas highlight regions that appear to have been most highly conserved among these telomeres; no strong sequence conservation has been found near the right of the terminal 26 bp among the different sequences listed, except between the chromosomal left ends from strains B31 and Sh-2-82 (see text). The telomeric sequences of the strain B31 chromosome were determined in this report; the others are from references 14, 28, 30, 45, 46.

Energy metabolism

The limited metabolic capacity of *B. burgdorferi* is similar to that found in *M. genitalium* (Fig. 3, Table 2). Genes encoding all of the enzymes of the glycolytic pathway were identified. Analysis of the metabolic pathway suggests that *B. burgdorferi* uses glucose as a primary energy source, although other carbohydrates, including glycerol, glucosamine, fructose and maltose, may be used in glycolysis. Pyruvate produced by glycolysis is converted to lactate, consistent with the microaerophilic nature of *B. burgdorferi*. Generation of reducing power occurs through the oxidative branch of the pentose pathway. None of the genes encoding proteins of the tricarboxylic acid cycle or oxidative phosphorylation were identified. The similarity in metabolic strategies of two distantly related, obligate parasites, *M. genitalium* and *B. burgdorferi*, suggests convergent evolutionary gene loss from more metabolically competent, distant progenitors.

Addition of *N*-acetylglucosamine (NAG) to culture medium is required for growth of *B. burgdorferi*[6]. NAG is incorporated into the cell wall, and may also serve as an energy source. The cp26 plasmid encodes a PTS cellobiose transporter homologue that could have specificity for the structurally similar compound chitobiose (di-*N*-acetyl-D-glucosamine). A gene product on the chromosome with sequence similarity to chitobiase (BB2) may convert chitobiose to NAG. *B. burgdorferi* can metabolize NAG to fructose-6-phosphate, which then can enter the glycolytic cycle through the action of *N*-acetylglucosamine-6-phosphate deacetylase and glucosamine-6-phosphate isomerase. NAG is the primary constituent of chitin, which makes up the tick cuticle[6], and may be a source of carbohydrate for *B. burgdorferi* when it is associated with its tick host.

The parallels between *B. burgdorferi* and *M. genitalium* appear to extend to other aspects of their metabolism. Both organisms lack a respiratory electron transport chain, so ATP production must be accomplished by substrate-level phosphorylation. Consequently, membrane potential is established by the reverse reaction of the V_1V_0-type ATP synthase, here functioning as an ATPase to expel protons from the cytoplasm (Fig. 3, Table 2). The ATP synthase genes in *B. burgdorferi* appear to be transcribed as part of a seven-gene operon. They are not typical of those usually found in eubacteria, more closely resembling the eukaryotic vacuolar (V-type) and archaeal (A-type) H^+-translocating ATPases[42], both in size

Nature © Macmillan Publishers Ltd 1997

and sequence similarity, than the bacterial F_1F_0 ATPases. Genome analysis of *Treponema pallidum*, the pathogenic spirochaete that causes syphilis, has also revealed the presence of a V_1V_0-type ATP synthase (C. M. F. *et al.*, manuscript in preparation), suggesting that this may be a feature of spirochaetes.

Regulatory systems

Although the expression of *Borrelia* genes varies according to the current host species, temperature, host body location and other local factors, control of gene expression appears to differ from more well studied eubacteria. A typical set of homologues of heat-shock response genes is present (*groES*, *groEL*, *grpE*, *dnaJ*, *hslU*, *hslV*, *dnaK* and *htpG*), and *B. burgdorferi* is known to have such a response; however, it lacks the σ-32 that controls their transcription in *E. coli*. Only a few homologues to other eubacterial regulatory proteins are present, including only two response-regulator two-component systems.

Motility and chemotaxis

Like other spirochaetes, *B. burgdorferi* has periplasmic flagella that are inserted at each end of the cell and extend towards the middle of the cell body. The unique flagella allow the organism to move through viscous solutions, an ability that is presumed to be important in its migration to distant tissues following deposition in the skin layers[43]. Proteins involved in motility and chemotaxis are encoded by 54 genes, more than 6% of the *B. burgdorferi* chromosome, most of which are arranged in eight operons containing between 2 and 25 genes.

B. burgdorferi contains several copies of the chemotaxis genes (*cheR*, *cheW*, *cheA*, *cheY* and *cheB*) downstream of the methyl-accepting chemotaxis proteins. Other eubacteria also have duplications of some *che* genes, but those genes in *B. burgdorferi* are the most redundant set yet found. *B. burgdorferi* lacks recognizable virulence factors; thus, its ability to migrate to distant sites in the tick and mammalian host is probably dependent on a robust chemotaxis response. Multiple chemotaxis genes may provide redundancy in this system in order to meet such challenges or, alternatively, these genes may be differentially expressed under varied physiological conditions. Another speculative possibility is that the flagellar motors at the two ends of the *B. burgdorferi* cell are different and require different *che* systems. In support of this idea is the observation that one of the motor switch genes, *fliG*, is also present in two copies.

Membrane protein analysis

Much of the previous work on *B. burgdorferi* has focused on outer-surface membrane genes because of their potential importance in bacterial detection and vaccination. Nearly all *Borrelia* membrane proteins have been found to be typical bacterial lipoproteins. A search of *B. burgdorferi* ORFs for a consensus lipobox in the first 30 amino acids identified 105 putative lipoproteins, representing more than 8% of coding sequences. This contrasts with a total of only 20 putative lipoproteins in the 1.67-million base pair *H. pylori* genome (1.3% of coding sequences)[23]. The periplasmic binding proteins involved in transport of amino acids/peptides and phosphate in *B. burgdorferi* are candidate lipoproteins, suggesting that they may be anchored to the outer surface of the cytoplasmic membrane as in Gram-positive bacteria, rather than localized in the periplasmic space.

In better-characterized eubacteria, prolipoprotein diacylglycerol transferase (*lgt*), prolipoprotein signal peptidase (*lsp*), and apolipoprotein:phospholipid *N*-acyl transferase (*lnt*) are required for post-translational processing and addition of lipids to the amino-terminal cysteine. Genes for the first two of the enzymes (lgt and lsp) are present in the *B. burgdorferi* genome, but the gene for *lnt* was not identified, although biochemical evidence argues for all three activities in *B. burgdorferi*[44]. The sequence similarity of an *lnt* homologue in *B. burgdorferi* may be too low to be identified using our search methods, or its activity may be present in a new enzyme. In *E. coli* the Sec protein export system moves lipoproteins through the inner membrane, and *Borrelia* carries a complete set of these protein-secretion gene homologues (*secA/D/E/F/Y* and *tth*; only the non-essential *secB* is missing).

Analysis of telomeres

The two chromosomal telomeres of strain B31 have similar 26-bp inverted terminal sequences (Fig. 4). We found no other similarity between the two ends, and these 26-bp sequences are very similar to the previously characterized *Borrelia* telomeres. Terminal restriction fragments from both B31 chromosomal termini were shown to exhibit snapback kinetics (data not shown), strongly indicating that both terminate in covalently closed hairpins, like previously characterized *Borrelia* telomeres[28,45,46].

The left chromosomal telomere of strain B31 is identical to the previously characterized left telomere of strain Sh-2-82 (ref. 28), except for a 31 bp insertion in B31 26 bp from the end. The rightmost 7,454 bp contains surprisingly few ORFs, given the ORF density elsewhere on the chromosome. The function of this region is unknown, but it contains several unusual features. The right terminal 900 bp contains considerable homology to the left ends of lp17 and lp28-3. The region between 3,600 bp and 8,000 bp from the right end also contains several areas with similarity to plasmid sequences, including a portion of the transposase-like gene approximately 4,500 bp from the right end. The spacing between the two conserved motifs (ATATAAT and TAGTATA) in the right 26-bp terminal repeat is the same as most previously known plasmid telomeres but different from the previously known chromosomal telomeres. These findings support the idea that the right end of the *Borrelia* chromosome has historically exchanged telomeres with the linear plasmids[28].

Conclusions

The *B. burgdorferi* genome sequence will provide a new starting point for the study of the pathogenesis, prevention and treatment of Lyme disease. With the exception of a small number of putative virulence genes (haemolysins and drug-efflux proteins), this organism contains few, if any, recognizable genes involved in virulence or host–parasite interactions, suggesting that *B. burgdorferi* differs from better-studied eubacteria in this regard. It will be interesting to determine the role of the multi-copy plasmid-encoded genes, as previous work has implicated plasmid genes in infectivity and virulence. The completion of the genome sequence from a second spirochaete, *Treponema pallidum* (C.M.F. *et al.*, manuscript in preparation) will allow for the identification of genes specific to each species and to this bacterial phylum, and will provide further insight into prokaryotic diversity. □

Methods

Cell lines. A portion of a low-passage subculture of the original Lyme-disease spirochaete tick isolate[4] was obtained from A. Barbour. The type strain of *B. burgdorferi* (ATCC 35210)[3], B31, was derived from this isolate by limiting dilution cloning[5]. Cells were grown in Barbour–Stoenner–Kelly medium II (BSKII)[6], omitting the additions of antibiotics and gelatin, in tightly closed containers at 33–34 °C. Cells were subcultured three or fewer times *in vitro* between successive rounds of infection in C3H/HeJ mice to minimize loss of infectivity and plasmid content[17,18]. After four successive transfers of infection in mice, a primary culture of B31, established from infected ear tissue, was expanded to 2.5 l by four successive subcultures. All available evidence indicates that the B31 line used for preparation of genomic DNA was probably clonal, as genetic heterogeneity was undetectable by several criteria including macrorestriction analysis (S. Casjens, unpublished data) and plasmid analysis of clonal derivatives of the B31 line[13].

Sequencing. The *B. burgdorferi* genome was sequenced by a whole-genome random sequencing method previously applied to other microbial genomes[20–24].

Nature © Macmillan Publishers Ltd 1997

An approximately 7.5-fold genome coverage was achieved by generating 19,078 sequences from a small insert plasmid library with an average edited length of 505 bases. The ends of 69 large insert lambda clones were sequenced to obtain a genome scaffold; 50% of the genome was covered by at least one lambda clone. Sequences were assembled using TIGR Assembler as described[20–24], resulting in a total of 524 assemblies containing at least two sequences, which were clustered into 85 groups based on linking information from forward and reverse sequence reads. All *Borrelia* sequences that had been mapped were searched against the assemblies in an attempt to delineate which were derived from the various elements of the *B. burgdorferi* genome. Some contigs were also located on the existing physical map by Southern analysis. Sequence and physical gaps for the chromosome were closed as described[20–24]. At the completion of the project, less than 3% of the chromosome had single-fold coverage. The linear chromosome of *B. burgdorferi* has covalently closed hairpin structures at its termini that are similar to those reported for linear plasmids in this organism[11]. The telomeric sequences (106 and 72 bp, respectively, from the left and right ends) were obtained after nicking the terminal loop with S1 nuclease and amplifying terminal sequences by ligation-mediated polymerase chain reaction (PCR) as described[28]. The unknown terminal sequence was determined in both directions on four independent plasmid clones of the amplified DNA from each telomere. A minimum amount of S1 nuclease was used and, because of their sequence similarity to other *Borrelia* telomeres, it is likely that few, if any, nucleotides were lost from the B31 chromosomal telomeres in this process.

Identification of ORFs. Coding regions (ORFs) were identified using compositional analysis using an interpolated Markov model based on variable-length oligomers[47]. ORFs of >600 bp were used to train the Markov model, as well as *B. burgdorferi* ORFs from GenBank. Once trained, the model was applied to the complete *B. burgdorferi* genome sequence and identified 953 candidate ORfs. ORFs that overlapped were visually inspected, and in some cases removed. Non-overlapping ORFs that were found between predicted coding regions and >30 amino acids in length were retained and included in the final annotation. All putative ORFs were searched against a non-redundant amino-acid database as described[20–24]. ORFs were also analysed using 527 hidden Markov models constructed for several conserved protein families (PFAM v2.0) using HMMER[48]. Families of paralogous genes were constructed by pairwise searches of proteins using FASTA. Matches that spanned at least 60% of the smaller of the protein pair were retained and visually inspected. A total of 94 paralogous gene families containing 293 genes were identified (Fig. 1).

Identification of membrane-spanning domains (MSDs). TopPred[49] was used to identify potential MSDs in proteins. A total of 526 proteins containing at least one putative MSD were identified, of which 183 were predicted to have more than one MSD. The presence of signal peptides and the probable position of a cleavage site in secreted proteins were detected using Signal-P as described[23]; 189 proteins were predicted to have a signal peptide. Lipoproteins were identified by scanning for a lipobox in the first 30 amino acids of every protein. A consensus sequence relaxed from that used for *H. pylori*[23] was defined for the purpose of this search based on known or putative *B. burgdorferi* lipoprotein consensus sequences.

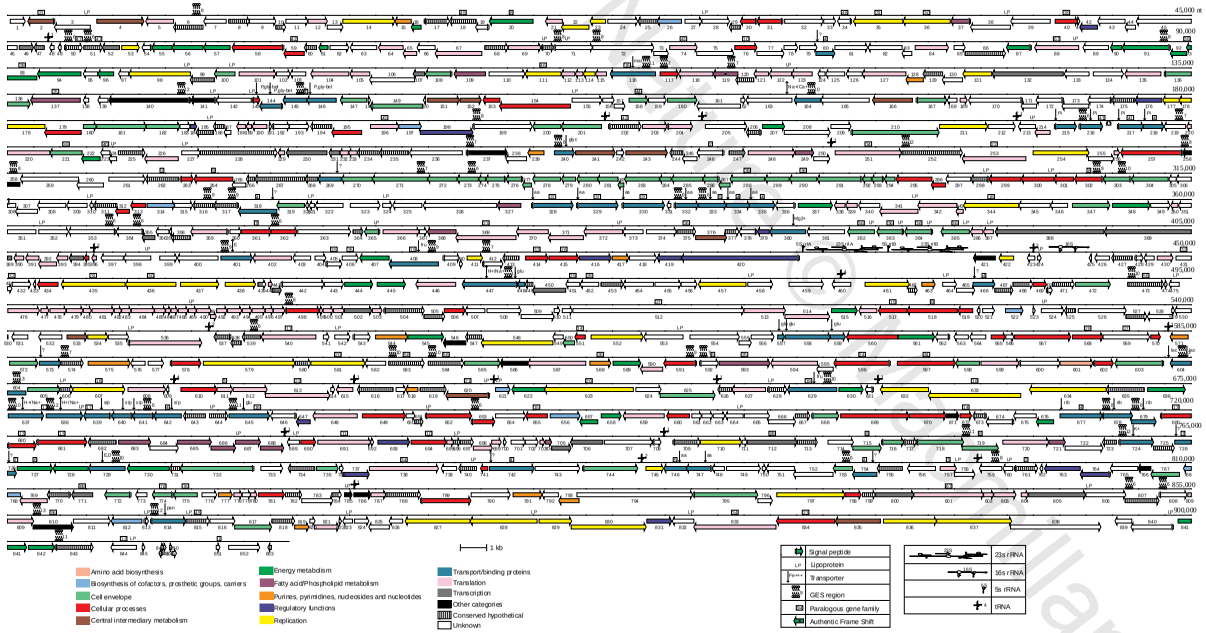
Received 3 November; accepted 18 November 1997.

1. Steere, A. C. *et al*. Lyme arthritis: an epidemic of oligoarticular arthritis in children and adults in three Connecticut communities. *Arthritis Rheum.* **20,** 7 (1977).
2. Steere, A. C. Lyme disease. *New Engl. J. Med.* **321,** 586–596 (1989).
3. Johnson, R. C., Schmidt, G. P., Hyde, F. W., Steigerwaalt, A. G. & Brenner, D. J. *Borrelia burgdorferi* sp. nov.: etiologic agent of Lyme disease. *Int. J. Syst. Bacteriol.* **34,** 496–497 (1984).
4. Burgdorfer, W. *et al*. Lyme disease—a tick-borne spirochetosis? *Science* **216,** 1317–1319 (1982).
5. Barbour, A. G., Burgdorfer, W., Hayes, S. F., Peter, O. & Aeschlimann, A. Isolation of a cultivable spirochete from *Ixodes ricinus* ticks of Switzerland. *Curr. Microbiol.* **8,** 123–126 (1983).
6. Barbour, A. & Hayes, S. F. Biology of *Borrelia* species. *Microbiol. Rev.* **50,** 381–400 (1986).
7. Baril, C., Richaud, C., Baranton, G. & Saint Girons, I. Linear chromosome of *Borrelia burgdorferi*. *Res. Microbiol.* **140,** 507–516 (1989).
8. Ferdows, M. S. & Barbour, A. G. Megabase-sized linear DNA in the bacterium *Borrelia burgdorferi*, the Lyme disease agent. *Proc. Natl Acac Sci. USA* **86,** 5969–5973 (1989).
9. Davidson, B., MacDougall, J. & Saint Girons, I. Physical map of the linear chromosome of the bacterium *Borrelia burgdorferi* 212, a causative agent of Lyme disease and localization of rRNA genes. *J. Bacteriol.* **174,** 3766–3774 (1992).
10. Casjens, S. & Huang, W. M. Linear chromosome physical and genetic map of *Borrelia burgdorferi*, the Lyme disease agent. *Mol. Microbiol.* **8,** 967–980 (1993).
11. Barbour, A. G. & Garon, C. F. Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science* **237,** 409–411 (1987).
12. Barbour, A. G. Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J. Clin. Microbiol.* **26,** 475–478 (1988).
13. Casjens, S., van Vugt, R., Stevenson, B., Tilly, K. & Rosa, P. Homology throughout the multiple 32-kilobase circular plasmids present in Lyme disease spirochetes. *J. Bacteriol.* **17,** 217–227 (1997).
14. Hinnebush, J. & Barbour, A. G. Linear- and circular-plasmid copy numbers in *Borrelia burgdorferi*. *J. Bacteriol.* **174,** 5251–5257 (1992).
15. Barbour, A. G., Carter, C. J., Bundoc, V. & Hinnebush, J. The nucleotide sequence of a linear plasmid of *Borrelia burgdorferi* reveals similarities to those of circular plasmids of other prokaryotes. *J. Bacteriol.* **178,** 6625–6639 (1996).
16. Zuckert, W. R. & Meyer, J. Circular and linear plasmids of Lyme disease spirochetes share extensive homology: characterization of a repeated DNA element. *J. Bacteriol.* **178,** 2287–2298 (1996).
17. Schwan, T. G., Burgdorfer, W. & Garon, C. Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of *in vitro* cultivation. *Infect. Immun.* **56,** 1831–1836 (1988).
18. Xu, Y., Kodner, C., Coleman, L. & Johnson, R. C. Correlation of plasmids with infectivity of *Borrelia burgdorferi sensu stricto* type strain B31. *Infect. Immun.* **64,** 3870–3876 (1996).
19. Norris, S. J., Howell, J. K., Garza, S. A., Ferdows, M. S. & Barbour, A. G. High- and low-infectivity phenotypes of clonal populations of *in vitro*-cultured *Borrelia burgdorferi*. *Infect. Immun.* **63,** 2206–2212 (1995).
20. Fleischmann, R. D. *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–512 (1995).
21. Fraser, C. M. *et al*. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
22. Bult, C. J. *et al*. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273,** 1058–1072 (1996).
23. Tomb, J.-F. *et al*. The complete genome sequence of the gastric pathogen *Heliocobacter pylori*. *Nature* **388,** 539–547 (1997).
24. Klenk, H.-P. *et al*. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390,** 364–370 (1997).
25. Casjens, S., Ley, H., DeLange, M., Rosa, P. & Huang, W. Linear chromosomes of Lyme disease agent spirochetes; genetic diversity and conservation of gene order. *J. Bacteriol.* **177,** 2769–2780 (1995).
26. Casjens, S. & Huang, W. in *Bacterial Genomes: Physical Structure and Analysis* (eds de Bruijn, F., Lupski, J. & Weinstock, G. J.) (Chapman and Hall, New York, in the press)
27. Old, I., MacDougall, J., Saint Girons, I. & Davidson, B. Mapping of genes on the linear chromosome of the bacterium *Borrelia burgdorferi*: possible locations for its origin of replication. *FEMS Microbiol. Lett.* **99,** 245–250 (1992).
28. Casjens, S. *et al*. Telomeres of the linear chromosomes of Lyme disease spirochetes: nucleotide sequence and possible exchange with linear plasmid telomers. *Mol. Microbiol.* **26,** 581–596 (1997).
29. Riley, M. Functions of gene products of *Escherichia coli*. *Microbiol. Rev.* **57,** 862–952 (1993).
30. Zhang, J., Hardham, J., Barbour, A. & Norris, S. Antigenic variation in Lyme disease *Borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89,** 275–285 (1997).
31. Dunn, J. J. *et al*. Complete nucleotide sequence of a circular plasmid from the Lyme disease spirochete, *Borrelia burgdorferi*. *J. Bacteriol.* **176,** 2706–2717 (1994).
32. Traktman, P. in *DNA Replication in Eukaryotic Cells* (ed. DePamphilis, M.) 775–798 (Cold Spring Harbor Laboratory Press, NY, 1996).
33. Himmelreich, R. *et al*. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24,** 4420–4449 (1996).
34. Brewer, B. When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* **53,** 679–686 (1988).
35. French, S. Consequences of replication fork movement through transcription units *in vivo*. *Science* **258,** 1362–1365 (1992).
36. Ogasawara, N. & Yoshikawa, H. Genes and their organization in the replication origin region of the bacterial chromosome. *Mol. Microbiol.* **6,** 629–634 (1992).
37. Gazumyan, A., Schwartz, J. J., Liveris, D. & Schwartz, I. Sequence analysis of the ribosomal RNA operon of the Lyme disease spirochete, *Borrelia burgdorferi*. *Gene* **146,** 57–65 (1994).
38. Ojaimi, C., Davidson, B., Saint Girons, I. & Old, I. Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of Lyme disease *spirochaetes Borrelia burgdorferi*, *B. garinii* and *B. afzelli*. *Microbiology* **140,** 2931–2940 (1994).
39. Cumow, A. W. *et al*. Glu-tRNA^{Gln} amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc. Natl Acad. Sci. USA* **94,** 11819–11826 (1997).
40. Ibba, M., Bobo, J. L., Rosa, P. A. & Soll, D. Archaeal-type lysyt-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl Acad. Sci. USA* (in the press).
41. Reizer, J., Paulsen, I. T., Reizer, A., Titgemeyer, F. & Saier, M. H. Jr Novel phosphotransferase system genes revealed by bacterial genome analysis: The complete complement of *pts* genes in *Mycoplasma genitalium*. *Microb. Comp. Genom.* **1,** 151–164 (1996).
42. Takase, K. *et al*. Sequencing and characterization of the *ntp* gene gluster for vacuolar-type Na+-translocating ATPase of *Enterococcurs hirae*. *J. Biol. Chem.* **269,** 11037–11044 (1994).
43. Sadziene, A., Rosa, P. A., Thompson, P. A., Hogan, D. M. & Barbour, A. G. Antibody-resistant mutants of *Borrelia burgdorferi*: *in vitro* selection and characterization. *J. Exp. Med.* **176,** 799–809 (1992).
44. Brandt, M. E., Riley, B. S., Radolf, J. D. & Norgard, M. V. Immunogenic integral membrane proteins of *Borrelia burgdorferi* are lipoproteins. *Infect. Immun.* **58,** 983–991 (1990).
45. Hinnebusch, J., Bergstrom, S. & Barbour, A. Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. *Mol. Microbiol.* **4,** 811–820 (1990).
46. Kitten, T. & Barbour, A. Juxtaposition of expressed variable antigen genes with a conserved telomere in the bacterium *Borrelia hermsii*. *Proc. Natl Acad. Sci. USA* **87,** 6077–6081 (1990).
47. Salzberg, S., Delcher, A., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* (in the press).
48. Sonnhammer, E. L. L., Eddy, S. R. & Durbin, R. Pfam: a comprehensive database of protein families based on seed alignments. *Proteins* **28,** 405–420 (1997).
49. Claros, M. G. & von Heijne, G. TopPred II: an improved software for membrane proteins structure predictions. *Comput. Appl. Biosci.* **10,** 685–686 (1994).

Acknowledgements. We thank A. G. Barbour for isolation of the *Borrelia burgdorferi* strain; A. Barbour, P. Rosa, K. Tilly, J. Riberio, B. Stevenson and D. Soll for discussions; N. K. Patel for technical assistance; M. Heaney, J. Scott and A. Saeed for software and database support; and V. Sapiro, B. Vincent and D. Maas for computer system support. This work was supported by a grant to J.C.V. and C.M.F. from the G. Harold and Leila Y. Mathers Charitable Foundation.

Correspondence and requests for materials should be sent to C.M.F. (e-mail: gbb@tigr.org). The annotated genome sequence and gene family alignments are available on the World-Wide Web at http://www.tigr.org/tdb/mdb/bbdb/bbdb.html. Sequences have been deposited with GenBank under the following accession numbers: AE00783 (chromosome); AE00784 (lp28-3); AE000785 (lp25); AE00786 (lp28-2); AE00787 (lp38); AE00788 (lp36); AE00789 (lp28-4); AE00790 (lp54); AE00791 (cp9); AE00792 (cp26); AE00793 (lp17); and AE00794 (lp28-1).

Nature © Macmillan Publishers Ltd 1997



Nature © Macmillan Publishers Ltd 1997

Nature © Macmillan Publishers Ltd 1997

Table 2. Identification of *B. burgdorferi* genes

BB #	Identification (Species)	%Sim.
Amino acid biosynthesis		
Serine family		
BB601	serine OHMTase (glyA) (Eo)	78
Biosynthesis of cofactors, prosthetic groups, and carriers		
Folic acid		
BB028	methylenetetrahydrofolate DHase (fdD) (Bs)	61
Heme and porphyrin		
BB197	protoporphyrinogen oxidase, prt (Bs)	58
BB658	oxygen-independent coproporphyrinogen III oxidase, prt (Bs)	51
Menaquinone and ubiquinone		
BB314	octaprenyl-dP Sase (sepB) (Eo)	57
Pantothenate		
BB812	pantothenate metabolism flavoprt (dfp) (Ec)	58
Pyridoxine		
BB768	pyridoxal kinase (pdxK) (Sc)	47
Thiamine		
BB821	4-methyl-5(b-OHethyl)-thiazole monoP kinase (thiK) (Ec)	58
Pyridine nucleotides		
BB522	NH(3)dep NAD+ Sase (Pa)	65
Cell envelope		
Membranes, lipoproteins, and porins		
BB332	basiomembrane prt B (bmpB) (Bb)	100
BB333	basiomembrane prt A (bmpA) (Bb)	100
BB334	basiomembrane prt C (bmpC) (Bb)	100
BB335	basiomembrane prt D (bmpD) (Bb)	100
BB108	basiomembrane prt (Tp)	50
BB319	exported prt (tpr38b) (Tp)	50
BB347	fibronectin/fibrinogen-BP, put (Sp)	53
BB442	inner membrane prt (Hi)	63
BB395	lipoprotein LA7 (Bb)	100
BB608	membrane-associated prt p66 (Bb)	100
BB753	membrane-spanning prt, put (Se)	49
BB795	outer membrane prt (Ng)	48
BB167	outer membrane prt (tpr50) (Tp)	48
BB795	rare lipoprt A (rlpA) (Hi)	58
BB110	surface-located membrane prt 1 (mp1) (Mh)	45
BB158	S2 prt (Bb)	60
lp54		
BBA24	decorin BP A (dbpA) (Bb)	94
BBA25	decorin BP B (dbpB) (Bb)	100
BBA38	lipoprotein (Bb)	54
BBA59	lipoprotein (Bb)	100
BBA62	lipoprotein (Bb)	100
BBA74	outer membrane porin (ome28) (Bb)	100
BBA15	outer surface prt A (ospA) (Bb)	99
BBA16	outer surface prt B (ospB) (Bb)	99
BBA03	outer membrane prt (Bb)	100
BBA52	outer membrane prt (Bb)	100
BBA05	S1 prt (Bb)	100
BBA04	S2 prt (Bb)	100
BBA60	surface lipoprt P27 (Bb)	81
lp56		
BBA09	outer surface prt D (ospD) (Bb)	100
BBA50	outer membrane prt, put (Bb)	61
BBA51	vlsE1 prt, put (Bb)	62
BBA52	vlsE1 prt, put (Bb)	55
lp28		
BBK53	outer membrane prt (Bb)	91
BBK52	prt p35 (Bb)	99
lp28-1		
BBF01	arpD prt, put (Bb)	53
BBF22	prt p35, put (Bb)	78
BBR32	vls recombination cassette Vls1-15 (Bb)	100
lp28-3		
BBH20	outer membrane porin (ome28), prt (Bb)	74
BBH21	outer membrane porin (ome28), prt (Bb)	74
lp36		
BBE09	prt p35 (Bb)	99
opp		
BBC09	exported prt A (oppA) (Bb)	100
Murein sacculus and peptidoglycan		
BB160	alanine racemase (alr) (Ec)	54
BB632	CPDase, put (Bs)	54
BB200	D-alanine:D-alanine ligase (ddlA) (Eo)	60
BB100	glutamate racemase (murI) (Ec)	58
BB625	N-Acmuramoyl-L-alanine amidase, prt (Ef)	53
BB732	penicillin-BP (pbp3) (Ng)	52
BB138	penicillin-BP (pbp-1) (Nm)	52
BB718	penicillin-BP (pbp2) (Hi)	52
BB303	phospho-N-Acmuramoyl-pentapeptideTase (mraY) (Bb)	100
BB715	rod shape-determining prt (mreB-1) (Ec)	78
BB716	rod shape-determining prt (mreC) (Bs)	51
BB719	rod shape-determining prt (mreB-2) (Hi)	61
BB805	serine-type D-Ala-D-Ala CPDase (dacA) (Hi)	55
BB472	UDP-N-Ac-G-1-carboxy-vinylTase (murA) (Ec)	59
BB598	UDP-N-Acmuramate DHase (murB) (Bs)	55
BB617	UDP-N-Acmuramate-alanine ligase (murC) (Hi)	54
BB595	UDP-N-Acmuramoylalanine-D-glutamate ligase (murD) (Bs)	52
BB201	UDP-N-Acmuramoylalanyl-D-glutamate-2,6-diaminopimelate ligase (murE) (Hi)	53
BB304	UDP-N-Acmuramoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (murF) (Bb)	100
BB767	UDP-N-Acglucosamine-N-Acmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol NAG TAse (murG) (Bs)	58
Surface polysaccharides, lipopolysaccharides and antigens		
BB744	antigen, p83/100 (Bb)	100
BB672	glycosyl TAse (lgtD) (Hi)	58
lp54		
BBA64	antigen, P35 (Bb)	100
BBA68	antigen, P35, put (Bb)	48
BBA73	antigen, P35, put (Bb)	52
lp28		
BBK41	antigen, P35, put (Bb)	47
lp28		
BBK15	antigen, P35, put (Bb)	50
BBK50	immunogenic prt P37 (Bb)	99
BBK32	immunogenic prt P35, put (Bb)	97
BBK37	immunogenic prt P37, put (Bb)	71
BBK45	immunogenic prt P37, put (Bb)	49
BBK46	immunogenic prt P37, put (Bb)	63
BBK48	immunogenic prt P37, put (Bb)	49
lp28-2		
BBH32	antigen, P35, put (Bb)	55
lp28-4		
BBK38	antigen, P35, put (Bb)	48
lp36		
BBE31	antigen, P35, put (Bb)	54
Surface structures		
BB239	flgr assembly prt (flbH) (Bb)	100
BB291	flgr basal-body rod prt (flgF) (Bb)	100
BB293	flgr basal-body rod prt (flgG) (Bb)	100
BB294	flgr basal-body rod prt (flgB) (Bb)	100
BB774	flgr basal-body rod prt (flgG) (So)	68
BB271	flgr biosyn prt (flhA) (Bb)	100
BB272	flgr biosyn prt (flhB) (Bb)	100
BB273	flgr biosyn prt (flhF) (Bb)	100
BB274	flgr biosyn prt (flhG) (Bb)	100
BB275	flgr biosyn prt (flhP) (Bb)	100
BB276	flgr biosyn prt (flhZ) (Bb)	100
BB147	flgr filament 41 kDa core prt (flaB) (Bb)	100
BB668	flgr filament outer layer prt (flaA) (Bb)	99
BB284	flgr hook assembly prt (flgD) (Bb)	100
BB283	flgr hook prt (flgE) (Bb)	100
BB181	flgr hook-associated prt (flgK) (Bb)	99
BB149	flgr hook-associated prt 2 (flID) (Bb)	99
BB182	flgr hook-associated prt 3 (flgL) (Bb)	99
BB775	flgr hook-basal body complex prt (flhO) (Bs)	58
BB292	flgr hook-basal body complex prt (flE) (Bb)	100
BB280	flgr motor rotation prt B (motB) (Bb)	100
BB281	flgr motor rotation prt A (motA) (Bb)	100
BB277	flgr motor-switch prt (fliN) (Bb)	100
BB278	flgr motor-switch prt (fliM) (Bb)	100
BB221	flgr motor-switch prt (fliG-1) (Td)	54
BB290	flgr motor-switch prt (fliG-2) (Bb)	100
BB772	flgr P-ring prt (flgI) (Ar)	51
BB279	flgr prt (flL) (Bb)	100
BB282	flgr prt (flbD) (Bb)	99
BB285	flgr prt (flbC) (Bb)	100
BB286	flgr prt (flbB) (Bb)	100
BB287	flgr prt (flbA) (Bb)	100
BB180	flgr prt, put (Bb)	100
BB550	flgr prt (flaJ) (Vp)	57
BB270	flgr-associated GTP-BP (flhF) (Bb)	100
BB288	flgr-specific ATP Sase (fliI) (Bb)	100
Cellular processes		
General		
BB567	chemotaxis histidine Kase (cheA-1) (Bb)	99
BB569	chemotaxis histidine Kase (cheA-2) (Bb)	100
BB040	chemotaxis prt MTase (cheR-1) (Re)	61
BB414	chemotaxis prt MTase (cheR-2) (Rm)	57
BB651	chemotaxis response regulator (cheY-1) (Tp)	74
BB570	chemotaxis response regulator (cheY-2) (Rm)	70
BB572	chemotaxis response regulator (cheY-3) (Bb)	98
BB660	GTP-BP (era) (Ec)	62
BB781	GTP-BP (obg) (Sym)	63
BB578	MC prt (mcp-1) (Tm)	52
BB598	MC prt (mcp-2) (Td)	57
BB597	MC prt (mcp-3) (Td)	52
BB830	MC prt (mcp-4) (Ec)	61
BB831	MC prt (mcp-5) (Ec)	52
BB495	prt-glutamate methylesterase (cheB-1) (Sc)	63
BB568	prt-glutamate methylesterase (cheB-2) (Sc)	59
BB312	purine-B chemotaxis prt (cheW-1) (Bb)	58
BB565	purine-B chemotaxis prt (cheW-2) (Ec)	68
BB670	purine-B-chemotaxis prt (cheW3) (Bb)	100
Cell division		
BB053	cell division contrl prt 27, put (Mj)	46
BB195	cell division contrl prt 27, put (Mj)	54
BB261	cell division inhibitor, put (Bs)	53
BB299	cell division prt (ftsZ) (Bs)	72
BB739	cell division prt (ftsH) (Bs)	72
BB078	cell division prt, put (Ec)	71
BB257	cell division prt, put (Ec)	71
BB300	cell division prt (ftsA) (Bb)	100
BB302	cell division prt (ftsW) (Bb)	100
BB301	cell division prt (divIB) (Bb)	100
BB313	cell division prt (ftsI) (Mj)	59
BB494	stage 0 sporulation prt J (spo0J) (Bs)	63

lp28-2
BBG08 stage 0 sporulation prt J (spoOJ) {Bb} 66

Cell killing
BB143 -hemolysin (hlyA) {Ah} 62
BB117 hemolysin III (yplQ) {Bs} 61
BB506 hemolysin (tlyA) {Sh} 59
BB059 hemolysin (tlyC) {Sh} 65
BB202 hemolysin, put {Syn} 54

Chaperones
BB741 chaperonin (groES) {Pg} 77
BB602 chaperonin, put {Cb} 72
BB519 grpE prt (grpE) {Bb} 100
BB295 heat shock prt (hslU) {Bb} 100
BB296 heat shock prt (hslV) {Bb} 100
BB649 heat shock prt (groEL) {Bb} 100
BB517 heat shock prt (dnaJ-1) {Bb} 100
BB655 heat shock prt (dnaJ-2) {Ca} 59
BB264 heat shock prt 70 (dnaK-1) {Bb} 61
BB518 heat shock prt 70 (dnaK-2) {Bb} 100
BB560 heat shock prt 90 (htpG) {Bb} 100

Detoxification
BB153 superoxide dismutase (sodA) {Hi} 68
BB690 neutrophil activating prt (napA) {Hi} 57
BB179 thiophene and furan oxidation prt (thdF) {Bb} 100

Protein and peptide secretion
BB154 preprt translocase sub (secA) {Bb} 100
BB395 preprt translocase sub (secE) {Bl} 62
BB498 preprt translocase sub (secY) {Sc} 64
BB362 prolipoprt diacylglyceryl Tase (lgt) {Ec} 56
BB652 prt-export membrane prt (secD) {Ec} 63
BB653 prt-export membrane prt (secF) {Hi} 63
BB030 signal peptidase I (lepB-1) {Bs} 51
BB031 signal peptidase I (lepB-2) {Syn} 57
BB263 signal peptidase I (lepB-3) {St} 57
BB469 signal peptidase II (lsp) {Sc} 60
BB694 signal recognition particle prt (ffh) {Bs} 70
BB610 trigger factor (tig) {Hi} 50

Transformation
BB591 competence locus E, put {Bs} 54
BB798 competence prt F, put {Hi} 52

Central intermediary metabolism

General
BB241 glycerol kinase (glpK) {Ec} 74
BB243 glycerol-3-P DHase, anaerobic (glpA) {Hi} 52
BB376 SAM Sase (metK) {Bs} 72

Amino sugars
BB152 glucosamine-6-P isomerase (nagB) {Hi} 79
BB151 N-Acglucosamine-6-P deAcase (nagA) {Hi} 54

Degradation of polysaccharides
BB620 -glucosidase, put {Syn} 58
BB002 -N-Achexosaminidase, put {As} 54

Phosphorus compounds
BB533 phnP prt (phnP) {Ec} 48

Polysaccharides - (cytoplasmic)
BB166 4- –glucanoTase (malQ) {Syn} 55
BB004 phosphoglucomutase (femD) {Mj} 52
BB835 phosphomannomutase (cpsG) {Hi} 57

Energy metabolism

Aerobic
BB728 NADH oxidase, water-forming (nox) {Sh} 59

Amino acids and amines
BB841 arginine deiminase (arcA) {Cp} 75
BB842 ornithine carbamoylTase (arcB) {Ng} 74

Anaerobic
BB016 glpE prt (glpE) {Hi} 53
BB087 L-lactate DHase (ldh) {Bs} 72

ATP-proton motive force interconversion
BB094 V-type ATPase, sub A (atpA) {Mb} 64
BB093 V-type ATPase, sub B (atpB) {Mb} 62
BB092 V-type ATPase, sub D (atpD) {Mj} 51
BB096 V-type ATPase, sub E (atpE) {Mj} 54
BB091 V-type ATPase, sub I (atpI) {Eh} 53
BB090 V-type ATPase, sub K (atpK) {Mj} 54

Electron transport
BB061 thioredoxin (trxA) {Ec} 59
BB515 thioredoxin RDase (trxB) {Bb} 99

Fermentation
BB622 acetate kinase (ackA) {Ec} 63
BB589 P AcTase (pta) {Tt} 65

Glycolysis
BB337 enolase (eno) {Bs} 79
BB445 fructose-bisP aldolase (fba) {Ec} 80
BB730 glucose-6-P isomerase (pgi) {Pf} 62
BB057 glyceraldehyde 3-P DHase (gap) {Bb} 99
BB630 1-phosphofructoKase (fruK) {Hi} 52
BB056 phosphoglycerate Kase (pgk) {Bb} 99
BB658 phosphoglycerate mutase (gpmA) {Ec} 79
BB348 pyruvate Kase (pyk) {Bs} 62
BB727 pyroP-fructose 6-P 1-PPTase (pfk) {Eh} 65
BB020 pyroP-fructose 6-P 1-PPTase, sub (pfpB) {Bb} 100
BB055 trioseP isomerase {Bb} 100

Pentose phosphate pathway
BB222 glucose-6-P 1-DHase, put {As} 48
BB636 glucose-6-P 1-DHase (zwf) {Hi} 64
BB561 phosphogluconate DHase (gnd) {Sd} 71
BB657 ribose 5-P isomerase (rpi) {Mj} 61

Sugars
BB407 mannose-6-P isomerase (manA) {Ec} 54
BB444 nucleotide sugar epimerase {Vc} 69
BB676 phosphoglycolate PPase (gph) {Hi} 50
BB207 UTP-glucose-1-P uridylylTase (gtaB) {Bs} 63
BB545 xylulokinase (xylB) {Bs} 43

Fatty acid and phospholipid metabolism

General
BB037 1-acyl-sn-glycerol-3-P AcTase (plsC) {Bb} 100
BB685 3-OH-3-methylglutaryl-CoA RDase (mvaA) {Pm} 52
BB683 3-OH-3-methylglutaryl-CoA Sase {At} 53
BB109 Ac-CoA C-AcTase (fadA) {Hi} 67
BB704 acyl carrier prt {Syn} 65
BB721 CDP-diacylglycerol-glycerol-3-P 3-phosphatidylTase {Bs} 55
BB327 glycerol-3-P O-acylTase, put {So} 50
BB368 glycerol-3-P DHase, NAD(P)+ (gpsA) {Bs} 54
BB137 long-chain-fatty-acid CoA ligase {Syn} 54
BB593 long-chain-fatty-acid CoA ligase {Syn} 56
BB688 melvalonate Kase {Mj} 51
BB686 mevalonate pyroP DCase {Sc} 52
BB119 phosphatidate cytidylylTase (cdsA), AFS{Ec} 61
BB249 phosphatidylTase {Hp} 52
BB687 phosphomevalonate Kase, put {Sc} 53

Purines, pyrimidines, nucleosides, nucleotides

Nucleotide and nucleoside interconversion
BB417 adenylate kinase (adk) {Bs} 64
BB128 cytidylate kinase (cmk-1) {Bs} 58
BB819 cytidylate kinase (cmk-2) {Mj} 57
BB463 nucleoside-diP kinase (ndk) {Bs} 70
BB793 thymidylate kinase (tmk) {Mj} 59
BB571 uridylate kinase (smbA) {Mj} 54

Purine ribonucleotide biosynthesis
BB544 phosphoribosyl pyroP Sase (prs) {Mp} 59

cp26
BBB18 GMP Sase (guaA) {Bb} 100
BBB17 IMP DHase (guaB) {Bb} 100

Pyrimidine ribonucleotide biosynthesis
BB575 CTP Sase (pyrG) {Mj} 71

Salvage of nucleosides and nucleotides
BB777 adenine phosphoribosylTase (apt) {Ta} 63
BB618 cytidine deaminase (cdd) {Mp} 61
BB239 deoxyguanosine/deoxyadenosine kinase(l) sub 2 (dck) {La} 59
BB375 pfs prt (pfs-1) {Ec} 64
BB588 pfs prt (pfs-2) {Hi} 59
BB791 thymidine kinase (tdk) {Bs} 47
BB015 uridine kinase (udk) {Bb} 100

lp36
BBK17 adenine deaminase (adeC) {Bs} 57

Regulatory functions

General
BB184 carbon storage regulator (csrA) {Hi} 63
BB647 ferric uptake regulation prt (fur) {Sp} 48
BB198 guanosine-3',5'-bis(diP) 3'-pyrophosphohydrolase (spoT) {Ec} 61
BB737 histidine phosphoKase/PPase, put {Ml} 49
BB176 methanol DHase regulator (moxR) {Bb} 99
BB416 pheromone shutdown prt (traB) {Ef} 61
BB042 P transport system regulatory prt (phoU) {Pa} 57
BB379 prt Kase C1 inhibitor (pkcI) {Bb} 100
BB419 response regulatory prt (rrp-1) {Syn} 57
BB763 response regulatory prt (rrp-2) {Ec} 67
BB764 sensory transduction histidine Kase, put {Bs} 60
BB420 sensory transduction histidine Kase, put {Syn} 61
BB693 xylose operon regulatory prt (xylR-1) {Th.} 48
BB831 xylose operon regulatory prt (xylR-2) {Syn} 51

lp54
BBA07 chpAI prt, put {Ec} 55

Replication

Degradation of DNA
BB411 endonuclease precursor (nucA) {As} 53

DNA replication, restriction, modification, recombination, and repair
BB422 3-methyladenine DNA glycosylase (mag) {At} 56
BB827 ATP-dep helicase (hrpA) {Ec} 61
BB437 chromosomal replication init prt (dnaA) {Bb} 100
BB435 DNA gyrase, sub A (gyrA) {Bs} 67
BB436 DNA gyrase, sub B (gyrB) {Bb} 99
BB344 DNA helicase (uvrD) {Ec} 55
BB552 DNA ligase (lig) {Ta} 56
BB211 DNA mismatch repair prt (mutL) {Hi} 55
BB797 DNA mismatch repair prt (mutS) {Hi} 57
BB098 DNA mismatch repair prt, put {Syn} 51
BB548 DNA polymerase I (polA) {Hi} 61
BB579 DNA polymerase III, sub (dnaE) {Ec} 62
BB438 DNA polymerase III, sub (dnaN) {Bb} 100
BB461 DNA polymerase III, sub / (dnaX) {Bs} 61
BB710 DNA primase (dnaG) {Bs} 56
BB581 DNA recombinase (recG) {Syn} 60
BB828 DNA topoisomerase I (topA) {Syn} 64
BB035 DNA topoisomerase IV (parC) {Bb} 58
BB036 DNA topoisomerase IV (parE) {Bb} 56
BB745 endonuclease III (nth) {Syn} 59
BB837 excinuclease ABC, sub A (uvrA) {Ec} 4
BB836 excinuclease ABC, sub B (uvrB) {Ec} 71
BB457 excinuclease ABC, sub C (uvrC) {Syn} 57
BB534 exodeoxyribonuclease III (exoA) {Bs} 67
BB632 exodeoxyribonuclease V, chain

Nature © Macmillan Publishers Ltd 1997

Nature © Macmillan Publishers Ltd 1997

(recD) {Ec} 54
BB633 exodeoxyribonuclease V, chain (recB) {Hi} 51
BB634 exodeoxyribonuclease V, chain (recC) {Hi} 51
BB829 exonuclease SbcD (sbcD) {Ec} 55
BB830 exonuclease SbcC (sbcC) {Ec} 52
BB177 glucose-inhibited div prt B (gidB) {Bb} 99
BB178 glucose-inhibited div prt A (gidA) {Bb} 100
BB022 Holliday junction DNA helicase (ruvB) {Bb} 100
BB023 Holliday junction DNA helicase (ruvA) {Bb} 100
BB014 primosomal prt N (priA) {Bb} 100
BB131 recA prt (recA) {Bb} 100
BB607 rep helicase, ss DNA-dep ATPase (rep) {Hi} 61
BB111 replicative DNA helicase (dnaB) {Ec} 58
BB114 ss DNA-BP (ssb) {Syn} 62
BB254 ss-DNA-specific exonuclease (recJ) {Hi} 52
BB623 transcription-repair coupling factor (mfd) {Hi} 60
BB053 uracil DNA glycosylase (ung) {Hi} 68

lp28-2
BBG32 replicative DNA helicase, put {Bs} 59

lp25
BBE29 adenine specific DNA MTase, put {Hp} 57

Transcription

General

BB052 spoU prt (spoU) {Ec} 54

Degradation of RNA

BB805 polyribonucleotide nucleotidylTase (pnpA) {Bs} 68
BB046 ribonuclease H (rnhB) {Hi} 66
BB705 ribonuclease III (rnc) {Bs} 62
BB441 ribonuclease P prt component (rnpA) {Bb} 100

DNA-dependent RNA polymerase

BB502 DNA-directed RNA polymerase (rpoA) {Bs} 64
BB389 DNA-directed RNA polymerase (rpoB) {Bb} 97
BB388 DNA-directed RNA polymerase (rpoC) {Ec} 71
BB771 RNA polymerase sigma factor (rpoS) {Pa} 61
BB712 RNA polymerase sigma-70 factor (rpoD) {Bb} 100
BB450 RNA polymerase sigma-54 factor (ntrA) {Av} 57

Transcription factors

BB107 N utilization substance prt B (nusB) {Ec} 62
BB800 N-utilization substance prt A (nusA) {Bs} 62
BB394 transcription antitermination factor (nusG) {Ec} 64
BB132 transcription elongation factor (greA) {Ec} 56
BB355 transcription factor, put {Mx} 47
BB230 transcription termination factor Rho (rho) {Bb} 100

RNA processing

BB706 polynucleotide adenylylTase (papS) {Bs} 57

Translation

General

BB590 dimethyladenosine Tase (ksgA) {Bs} 61
BB802 ribosome-B factor A (rbfA) {Bs} 62

Amino acyl tRNA synthetases

BB220 alanyl-tRNA Sase (alaS) {Ec} 62
BB594 arginyl-tRNA Sase (argS) {Mj} 55
BB101 asparaginyl-tRNA Sase (asnS) {Ec} 73
BB446 aspartyl-tRNA Sase (aspS) {Ec} 66
BB599 cysteinyl-tRNA Sase (cysS) {Hi} 58
BB372 glutamyl-tRNA Sase (gltX) {Rm} 63
BB371 glycyl-tRNA Sase (glyS) {Ta} 68
BB135 histidyl-tRNA Sase (hisS) {Mj} 59
BB833 isoleucyl-tRNA Sase (ileS) {Sc} 66
BB251 leucyl-tRNA Sase (leuS) {Bs} 70
BB659 lysyl-tRNA Sase {Mj} 54
BB587 methionyl-tRNA Sase (metG) {Sc} 67
BB514 phenylalanyl-tRNA Sase, sub (pheT) {Bb} 100
BB513 phenylalanyl-tRNA Sase, sub (pheS) {Bb} 100
BB402 prolyl-tRNA Sase (proS) {Sc} 65
BB226 seryl-tRNA Sase (serS) {Bs} 62
BB720 threonyl-tRNA Sase (thrZ) {Bs} 67
BB005 tryptophanyl-tRNA Sase (trsA) {Cl} 65
BB370 tyrosyl-tRNA Sase (tyrS) {Bs} 62
BB738 valyl-tRNA Sase (valS) {Bs} 67

Degradation of proteins, peptides, and glycopeptides

BB608 aminoacyl-histidine dipeptidase (pepD) {Hi} 55
BB366 aminopeptidase I (yscI) {Bb} 100
BB069 aminopeptidase II {Bs} 57
BB611 ATP-dep Clp protease proteolytic component (clpP-1) {Hi} 79
BB757 ATP-dep Clp protease proteolytic component (clpP-2) {Hi} 67
BB369 ATP-dep Clp protease, sub A (clpA) {Ec} 56
BB612 ATP-dep Clp protease, sub X (clpX) {Ec} 75
BB834 ATP-dep Clp protease, sub C (clpC) {Pp} 67
BB253 ATP-dep protease LA (lon-1) {Bb} 100
BB613 ATP-dep protease LA (lon-2) {Hi} 65
BB359 carboxyl-terminal protease (ctp) {Syn} 65
BB203 Lambda CII stability-governing prt (hflK) {Ec} 56
BB204 Lambda CII stability-governing prt (hflC) {Ec} 56
BB248 oligoendopeptidase F (pepF) {Ll} 58
BB067 peptidase, put {Sc} 56
BB104 periplasmic serine protease DO (htrA) {Hi} 60
BB430 proline dipeptidase (pepQ) {Hi} 49
BB769 sialoglycoprotease (gcp) {Hi} 60
BB627 vacuolar X-prolyl dipeptidyl aminopeptidase I (pepX) {Ml 55

BB118 zinc protease, put {Hi} 54
BB536 zinc protease, put {Hi} 52

Nucleoproteins

BB232 hbbU prt {Bb} 100

Protein modification

BB105 methionine aminopeptidase (map) {Bs} 68
BB065 polypeptide deformylase (def) {Syn} 67
BB648 serine/threonine kinase, put {Pf} 51

Ribosomal proteins: synthesis and modification

BB392 ribosomal prt L1 (rplA) {Bs} 71
BB481 ribosomal prt L2 (rplB) {Bb} 99
BB478 ribosomal prt L3 (rplC) {Bb} 99
BB479 ribosomal prt L4 (rplD) {Bb} 100
BB490 ribosomal prt L5 (rplE) {Hi} 80
BB493 ribosomal prt L6 (rplF) {Sc} 72
BB390 ribosomal prt L7/L12 (rplL) {Sc} 75
BB112 ribosomal prt L9 (rplI) {Ec} 57
BB391 ribosomal prt L10 (rplJ) {Bs} 61
BB393 ribosomal prt L11 (rplK) {Tm} 73
BB339 ribosomal prt L13 (rplM) {Hi} 72
BB488 ribosomal prt L14 (rplN) {Tm} 79
BB497 ribosomal prt L15 (rplO) {Bs} 68
BB485 ribosomal prt L16 (rplP) {Syn} 81
BB503 ribosomal prt L17 (rplQ) {Ec} 63
BB494 ribosomal prt L18 (rplR) {Bs} 69
BB699 ribosomal prt L19 (rplS) {Ec} 74
BB188 ribosomal prt L20 (rplT) {Ec} 70
BB778 ribosomal prt L21 (rplU) {Ec} 58
BB483 ribosomal prt L22 (rplV) {Bb} 100
BB480 ribosomal prt L23 (rplW) {Bb} 100
BB489 ribosomal prt L24 (rplX) {Ec} 64
BB780 ribosomal prt L27 (rpmA) {Hi} 82
BB350 ribosomal prt L28 (rpmB) {Ec} 62
BB486 ribosomal prt L29 (rpmC) {Bs} 65
BB496 ribosomal prt L30 (rpmD) {Bs} 60
BB229 ribosomal prt L31 (rpmE) {Bs} 69
BB703 ribosomal prt L32 (rpmF) {Bs} 62
BB396 ribosomal prt L33 (rpmG) {Bs} 76
BB440 ribosomal prt L34 (rpmH) {Bb} 100
BB189 ribosomal prt L35 (rpmI) {Ba} 74
BB499 ribosomal prt L36 (rpmJ) {Bs} 89
BB127 ribosomal prt S1 (rpsA) {Ec} 55
BB123 ribosomal prt S2 (rpsB) {Pa} 79
BB484 ribosomal prt S3 (rpsC) {Hi} 71
BB615 ribosomal prt S4 (rpsD) {Hi} 63
BB495 ribosomal prt S5 (rpsE) {Bs} 77
BB115 ribosomal prt S6 (rpsF) {Os} 50
BB386 ribosomal prt S7 (rpsG) {Sc} 75
BB492 ribosomal prt S8 (rpsH) {Syn} 66
BB338 ribosomal prt S9 (rpsI) {Hi} 71
BB477 ribosomal prt S10 (rpsJ) {Bb} 100
BB501 ribosomal prt S11 (rpsK) {Hi} 77
BB387 ribosomal prt S12 (rpsL) {An} 89
BB500 ribosomal prt S13 (rpsM) {Cp} 76
BB491 ribosomal prt S14 (rpsN) {Bs} 72
BB804 ribosomal prt S15 (rpsO) {Tt} 77
BB695 ribosomal prt S16 (rpsP) {Bs} 70
BB487 ribosomal prt S17 (rpsQ) {Mc} 76
BB113 ribosomal prt S18 (rpsR) {Bs} 78
BB482 ribosomal prt S19 (rpsS) {Bb} 99
BB233 ribosomal prt S20 (rpsT) {Bb} 100
BB256 ribosomal prt S21 (rpsU) {Mx} 68
BB516 rRNA methylase (yacO) {Mc} 66

tRNA modification

BB821 2-methylthio-N6-isopentyladenosine tRNA modification enzyme (miaA) {Ec} 53
BB084 AT (nifS) {Syn} 61
BB343 glu-tRNA amidoTase, sub C (gatC) {Bs} 56
BB341 glu-tRNA amidoTase, sub B (gatB) {Bs} 63
BB342 glu-tRNA amidoTase, sub A (gatA) {Bs} 61
BB064 methionyl-tRNA formylTase (fmt) {Ec} 56
BB787 peptidyl-tRNA hydrolase (pth) {Bb} 100
BB012 pseudouridylate Sase I (hisT) {Bb} 100
BB021 SAM: tRNA ribosylTase-isomerase {Bb} 96
BB809 tRNA-guanine transglycosylase (tgt) {Zm} 60
BB698 tRNA (guanine-N1)-MTase (trmD) {Mg} 68
BB803 tRNA pseudouridine 55 Sase (truB) {Ec} 57

Translation factors

BB088 GTP-B membrane prt (lepA) {Hi} 76
BB196 peptide chain release factor 1 (prfA) Hi} 73
BB074 peptide chain release factor 2 (prfB) {Sc} 70
BB121 ribosome releasing factor (frr) {Mt} 68
BB169 translation initiation factor 1 (infA) {Ec} 87
BB801 translation initiation factor 2 (infB) {Bs} 73
BB190 translation initiation factor 3 (infC) {Pv} 72
BB691 translation elongation factor G (fus-2) {Tm} 67
BB214 translation elongation factor P (efp) {Ec} 56
BB476 translation elongation factor TU (tuf) {Bb} 100
BB122 translation elongation factor TS (tsf) {Hi} 57
BB540 translation elongation factor G (fus-1) {Tm} 68

Transport and binding proteins

General

BB573 ABC transporter, ATP-BP {Bs} 53
BB742 ABC transporter, ATP-BP {Syn} 57
BB466 ABC transporter, ATP-BP {Hi} 74
BB754 ABC transporter, ATP-BP {Bl} 60
BB080 ABC transporter, ATP-BP {Mj} 63
BB269 ATP-BP (ylxH-1) {Bb} 100
BB726 ATP-BP (ylxH-2) {Bb} 54

lp38
BBJ26 ABC transporter, ATP-BP {Mj} 62

Amino acids, peptide, and amines

BB729 glutamate transporter (gltP) {Bs} 55
BB401 glutamate transporter, put {Bs} 53
BB146 GBP ABC transporter, ATP-BP

Nature © Macmillan Publishers Ltd 1997

(proV) {Sc} 71
BB145 GBP ABC transporter, permease prt (proW) {Ec} 66
BB144 GBP ABC transporter, BP (proX) {Ec} 43
BB334 OP ABC transporter, ATP-BP (oppD) {Bs} 75
BB335 OP ABC transporter, ATP-BP (oppF) {Bs} 80
BB332 OP ABC transporter, permease prt (oppB-1){Ec} 68
BB747 OP ABC transporter, permease prt (oppB-2){Bs} 54
BB333 OP ABC transporter, permease prt (oppC-1){Hi} 64
BB746 OP ABC transporter, permease prt (oppC-2){Bs} 52
BB328 OP ABC transporter, periplasmic BP (oppA-1) {Bb} 74
BB329 OP ABC transporter, periplasmic BP (oppA-2) {Bb} 94
BB330 OP ABC transporter, periplasmic BP (oppA-3) {Bb} 81
BB642 SP ABC transporter, ATP-BP (potA) {Ec} 69
BB641 SP ABC transporter, permease prt (potB) {Ec} 65
BB640 SP ABC transporter, permease prt (potC) {Ec} 63
BB639 SP ABC transporter, periplasmic BP (potD) {Ec} 53

lp54
BBA34 OP ABC transporter, periplasmic BP (oppA-4) {Bc} 66

cp26
BBB16 OP ABC transporter, periplasmic BP (oppA) {Bb} 78

Anions
BB218 P ABC transporter, ATP-BP (pstB) {Pa} 74
BB216 P ABC transporter, permease prt (pstC) {Ec} 58
BB217 P ABC transporter, permease prt (pstA) {Syn} 63
BB215 P ABC transporter, periplasmic P-BP (pstS) {Syn} 48

Carbohydrates, organic alcohols, and acids
BB240 glycerol uptake facilitator (glpF) {Bs} 57
BB604 L-lactate permease (lctP) {Ec} 57
BB318 methylgalactoside ABC transporter, ATP-BP (mglA) {Hi} 54
BB814 pantothenate permease (panF) {Ec} 63
BB448 phosphocarrier prt HPr (ptsH-1) {Mg} 56
BB557 phosphocarrier prt HPr (ptsH-2) {Hi} 69
BB558 phosphoenolpyruvate-prt PPase (ptsI) {Sc} 65
BB408 PTS system, fru-specific IIABC (fruA-1) {Ec} 65
BB629 PTS system, fru-specific IIABC (fruA-2) {Ec} 68
BB559 PTS system, glu-specific IIA (crr) {Bb} 100
BB645 PTS system, glu-specific IIBC (ptsG) {Sc} 67
BB116 PTS system, mal/glu-specific IIABC (malX) {Ec} 56
BB677 RG ABC transporter, ATP-BP (mglA) {Mg} 68
BB678 RG ABC transporter, permease prt (rbsC-1) {Mg} 51
BB679 RG ABC transporter, permease prt (rbsC-2) {Mp} 52

cp26
BBB04 PTS system, cello-specific IIC (celB) {Bs} 62
BBB05 PTS system, cello-specific IIA (celC) {Bs} 61
BBB06 PTS system, cello-specific IIB (celA) {Bs} 73
BBB29 PTS system, glu-specific IIBC, put {Ec} 70

Cations
BB724 K+ transport prt (ntpJ) {Eh} 60
BB380 Mg2+ transport prt (mgtE) {Bb} 100
BB164 Na+/Ca+ exchange prt, put {Mj} 59
BB447 Na+/H+ antiporter (napA) {Eh} 57
BB637 Na+/H+ antiporter (nhaC-1) {Bf} 48
BB638 Na+/H+ antiporter (nhaC-2) {Hi} 50

Other
BB451 chromate transport prt, put {Mj} 58

Other categories

Adaptations and atypical conditions
BB237 acid-inducible prt (act206) {Rm} 45
BB786 general stress prt (ctc) {Bs} 51
BB785 stage V sporulation prt G {Bm} 74
BB810 virulence factor mviN prt (mviN) {Hi} 51

Colicin-related functions
BB766 colicin V production prt, put {Hi} 52
BB546 outer membrane integrity prt (tolA) {Hi} 44

Drug and analog sensitivity
BB140 acriflavine resistance prt (acrB) {Hi} 53
BB258 bacitracin resistance prt (bacA) {Ec} 56
BB586 femA prt (femA) {Se} 47
BB141 membrane fusion prt (mtrC) {Hi} 47

lp28-4
BBI26 multidrug-efflux transporter {Hp} 55

lp25
BBE22 pyrazinamidase/nicotinamidase (pncA) {Mt} 56

Transposon-related functions

lp38
BBJ05 transposase-like prt, put {Bb} 89

lp36
BBK25 transposase-like prt, put {Bb} 80

lp28-1
BBF18 transposase-like prt, put {Bb} 96
BBF19 transposase-like prt, put {Bb} 96

lp28-2
BBG05 transposase-like prt {Bb} 99

lp28-3
BBH40 transposase-like prt, put {Bb} 57

lp17
BBD20 transposase-like prt, put {Bb} 99
BBD23 transposase-like prt, put {Bb} 88

Unknown
BB528 aldose RDase, put {Bs} 57
BB684 carotenoid biosyn prt, put {Ss} 58
BB671 chemotaxis operon prt (cheX) {Bb} 99
BB250 dedA prt (dedA) {Ec} 54
BB168 dnaK suppressor, put {Ec} 53
BB508 GTP-BP {Tp} 59
BB219 gufA prt {Mx} 54
BB421 hydrolase {Hi} 58
BB524 inositol monoPPase {Hs} 47
BB454 lipopolysaccharide biosyn-related prt {Mj} 49
BB702 lipopolysaccharide biosyn-related prt {Hi} 62
BB045 P115 prt {Mh} 53
BB336 P26 {Bb} 100
BB363 periplasmic prt {Bb} 100
BB033 small prt (smpB) {Rp} 70
BB297 smg prt {Bb} 100
BB443 spoIIIJ-associated prt (jag) {Bs} 56

lp54
BBA76 thy1 prt (thy1) {Dd} 68

lp28-4
BBI06 pfs prt (pfs) {Ec} 59

cp9
BBC09 rev prt (rev) {Bb} 62
BBC10 rev prt (rev) {Bb} 66

Nature © Macmillan Publishers Ltd 1997

Nature © Macmillan Publishers Ltd 1997