The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

F. Kunst[1], N. Ogasawara[2], I. Moszer[3], A. M. Albertini[4], G. Alloni[4], V. Azevedo[5], M. G. Bertero[3,4], P. Bessières[5], A. Bolotin[5], S. Borchert[6], R. Borriss[7], L. Boursier[3], A. Brans[8], M. Braun[9], S. C. Brignell[10], S. Bron[11], S. Brouillet[3,12], C. V. Bruschi[13], B. Caldwell[14], V. Capuano[5], N. M. Carter[10], S.-K. Choi[15], J.-J. Codani[16], I. F. Connerton[17], N. J. Cummings[17], R. A. Daniel[18], F. Denizot[19], K. M. Devine[20], A. Düsterhöft[9], S. D. Ehrlich[5], P. T. Emmerson[21], K. D. Entian[6], J. Errington[18], C. Fabret[19], E. Ferrari[14], D. Foulger[18], C. Fritz[9], M. Fujita[22], Y. Fujita[23], S. Fuma[24], A. Galizzi[4], N. Galleron[5], S.-Y. Ghim[15], P. Glaser[3], A. Goffeau[25], E. J. Golightly[26], G. Grandi[27], G. Guiseppi[19], B. J. Guy[10], K. Haga[28], J. Haiech[19], C. R. Harwood[10], A. Hénaut[29], H. Hilbert[9], S. Holsappel[11], S. Hosono[30], M.-F. Hullo[3], M. Itaya[31], L. Jones[32], B. Joris[8], D. Karamata[33], Y. Kasahara[2], M. Klaerr-Blanchard[3], C. Klein[6], Y. Kobayashi[30], P. Koetter[6], G. Koningstein[34], S. Krogh[20], M. Kumano[24], K. Kurita[24], A. Lapidus[5], S. Lardinois[8], J. Lauber[9], V. Lazarevic[33], S.-M. Lee[35], A. Levine[36], H. Liu[28], S. Masuda[30], C. Mauël[33], C. Médigue[3,12], N. Medina[36], R. P. Mellado[37], M. Mizuno[30], D. Moestl[9], S. Nakai[2], M. Noback[11], D. Noone[20], M. O'Reilly[20], K. Ogawa[24], A. Ogiwara[38], B. Oudega[34], S.-H. Park[15], V. Parro[37], T. M. Pohl[39], D. Portetelle[40], S. Porwollik[7], A. M. Prescott[18], E. Presecan[3], P. Pujic[5], B. Purnelle[25], G. Rapoport[1], M. Rey[26], S. Reynolds[33], M. Rieger[41], C. Rivolta[33], E. Rocha[3,12], B. Roche[36], M. Rose[6], Y. Sadaie[22], T. Sato[30], E. Scanlan[20], S. Schleich[3], R. Schroeter[7], F. Scoffone[4], J. Sekiguchi[42], A. Sekowska[3], S. J. Seror[36], P. Serror[5], B.-S. Shin[15], B. Soldo[33], A. Sorokin[5], E. Tacconi[4], T. Takagi[43], H. Takahashi[28], K. Takemaru[30], M. Takeuchi[30], A. Tamakoshi[24], T. Tanaka[44], P. Terpstra[11], A. Tognoni[27], V. Tosato[13], S. Uchiyama[42], M. Vandenbol[40], F. Vannier[36], A. Vassarotti[45], A. Viari[12], R. Wambutt[46], E. Wedler[46], H. Wedler[46], T. Weitzenegger[39], P. Winters[14], A. Wipat[10], H. Yamamoto[42], K. Yamane[24], K. Yasumoto[28], K. Yata[22], K. Yoshida[23], H.-F. Yoshikawa[28], E. Zumstein[5], H. Yoshikawa[2] & A. Danchin[3]

[1] *Institut Pasteur, Unité de Biochimie Microbienne, 25 rue du Docteur Roux, 75724 Paris Cedex 15, France*
[2] *Nara Institute of Science and Technology, Graduate School of Biological Sciences, Ikoma, Nara 630-01, Japan*
[3] *Institut Pasteur, Unité de Régulation de l'Expression Génétique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France*
[4] *Dipartimento di Genetica e Microbiologia, Universita di Pavia, Via Abbiategrasso 207, 27100 Pavia, Italy*
[5] *INRA, Génétique Microbienne, Domaine de Vilvert, 78352 Jouy-en-Josas Cedex, France*
[6] *Institut für Mikrobiologie, J. W. Goethe-Universität, Marie Curie Strasse 9, 60439 Frankfurt/Maine, Germany*
[7] *Institut für Genetik und Mikrobiologie, Humboldt Universität, Chausseestrasse 17, D-10115 Berlin, Germany*
[8] *Centre d'Ingénierie des Protéines, Université de Liège, Institut de Chimie B6, Sart Tilman, B-4000 Liège, Belgium*
[9] *QIAGEN GmbH, Max-Volmer-Strasse 4, D-40724 Hilden, Germany*
[10] *Department of Microbiological, Immunological and Virological Sciences, The Medical School, University of Newcastle, Framlington Place, Newcastle upon Tyne NE2 4HH, UK*
[11] *Department of Genetics, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands*
[12] *Atelier de BioInformatique, Université Paris VI, 12 rue Cuvier, 75005 Paris, France*
[13] *ICGEB, AREA Science Park, Padriciano 99, I-34012 Trieste, Italy*
[14] *Genencor International, 925 Page Mill Road, Palo Alto, California 94304-1013, USA*
[15] *Bacterial Molecular Genetics Research Unit, Applied Microbiology Research Division, KRIBB, PO Box 115, Yusong, Taejon 305-600, Korea*
[16] *INRIA, Domaine de Voluceau, PB 105, 78153 Le Chesnay Cedex, France*
[17] *Institute of Food Research, Department of Food Macromolecular Science, Reading Laboratory, Earley Gate, Whiteknights Road, Reading RG6 6BZ, UK*
[18] *Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, OX1 3RE, UK*
[19] *Laboratoire de Chimie Bactérienne, CNRS BP 71, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 09, France*
[20] *Department of Genetics, Trinity College, Lincoln Place Gate, Dublin 2, Republic of Ireland*
[21] *Department of Biochemistry and Genetics, The Medical School, University of Newcastle, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK*
[22] *Radioisotope Center, National Insitute of Genetics, Mishima, Shizuoka-ken 411, Japan*
[23] *Department of Biotechnology, Faculty of Engineering, Fukuyama University, Higashimura-cho, Fukuyama-shi, Hiroshima 729-02, Japan*
[24] *Institute of Biological Sciences, Tsukuba University, Tsuiuba-shi, Ibaraki 305, Japan*
[25] *Faculté des Sciences Agronomiques, Unité de Biochimie Physiologique, Université Catholique de Louvain, Place Croix du Sud, 2-20 B-1348 Louvain-la-Neuve, Belgium*
[26] *Novo Nordisk Biotech, 1445 Drew Avenue, Davis, California 95616-4880, USA*
[27] *Eniricerche, Via Maritano 26, San Donato Milanese, 20097 Milan, Italy*
[28] *Institute of Molecular and Cellular Biology, The University of Tokyo, Bunkyo-ku, Tokyo 113, Japan*
[29] *Laboratoire Génome et Informatique, Université de Versailles, Bâtiment Buffon, 45 Avenue des États-Unis, 78035 Versailles Cedex, France*
[30] *Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu, Tokyo 183, Japan*
[31] *Mitsubishi Kasei Institute of Life Sciences, 11 Minamyiooa, Machida-shi, Tokyo 194, Japan*
[32] *Institut Pasteur, Service d'Informatique Scientifique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France*
[33] *Institut de Génétique et Biologie Microbiennes, Université de Lausanne, 19 rue César Roux, 1005 Lausanne, Switzerland*
[34] *Department of Molecular Microbiology, MBW/BCA, Faculty of Biology, Vrije Universiteit Amsterdam, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands*
[35] *Chongju University College of Science and Engineering, Chongju City, Korea*
[36] *Institut de Génétique et Microbiologie, Université Paris Sud, URA CNRS 2225, Université Paris XI–Bâtiment 409, 91405 Orsay Cedex, France*
[37] *Centro Nacional de Biotecnologia (CSIC), Campus Universidad Autonoma, Cantoblanco, 28049 Madrid, Spain*
[38] *National Institute of Basic Biology, 38 Nishigounaka, Myoudaiji-chou, Okazaki 444, Japan*
[39] *Gesellschaft für Analyse-Technik und Consulting mbH, Fritz-Arnold Straße 23, D-78467 Konstanz, Germany*
[40] *Department of Microbiology, Faculty of Agronomy, 6 Avenue du Maréchal Juin, B-5030 Gembloux, Belgium*
[41] *Biotech Research, BMF, Wilhelmsfeld, Klingelstrasse 35, D-69434 Hirschhorn, Germany*
[42] *Department of Applied Biology, Faculty of Textile Science and Technology, Shinshu University 3-15-1, Tokida, Ueda-shi, Nagano 386, Japan*
[43] *Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan*
[44] *Department of Marine Science, School of Marine Science and Technology, Tokai University, 3-20-1 Orido Shimizu, Shizuoka 424, Japan*
[45] *European Commission, DG XII-E-1, SDME 8/78, Rue de la Loi 200, B-1049 Brussels, Belgium*
[46] *AGOWA GmbH, Glienicker Weg 185, 12489 Berlin, Germany*

Bacillus subtilis **is the best-characterized member of the Gram-positive bacteria. Its genome of 4,214,810 base pairs comprises 4,100 protein-coding genes. Of these protein-coding genes, 53% are represented once, while a quarter of the genome corresponds to several gene families that have been greatly expanded by gene duplication, the largest family containing 77 putative ATP-binding transport proteins. In addition, a large proportion of the genetic capacity is devoted to the utilization of a variety of carbon sources, including many plant-derived molecules. The identification of five signal peptidase genes, as well as several genes for components of the secretion apparatus, is important given the capacity of** ***Bacillus*** **strains to secrete large amounts of industrially important enzymes. Many of the genes are involved in the synthesis of secondary metabolites, including antibiotics, that are more typically associated with** ***Streptomyces*** **species. The genome contains at least ten prophages or remnants of prophages, indicating that bacteriophage infection has played an important evolutionary role in horizontal gene transfer, in particular in the propagation of bacterial pathogenesis.**

Techniques for large-scale DNA sequencing have brought about a revolution in our perception of genomes. Together with our understanding of intermediary metabolism, it is now realistic to envisage a time when it should be possible to provide an extensive chemical definition of many living organisms. During the past couple of years, the genome sequences of *Haemophilus influenzae*, *Mycoplasma genitalium*, *Synechocystis* PCC6803, *Methanococcus jannaschii*, *M. pneumoniae*, *Escherichia coli*, *Helicobacter pylori*, *Archaeoglobus fulgidus* and the yeast *Saccharomyces cerevisiae* have been published in their entirety[1–8], and at least 40 prokaryotic genomes are currently being sequenced. Regularly updated lists of genome sequencing projects are available at http://www.mcs.anl.gov/home/gaasterl/genomes.html (Argonne National Laboratory, Illinois, USA) and http://www.tigr.org (TIGR, Rockville, Maryland, USA).

The list of sequenced microorganisms does not currently include a paradigm for Gram-positive bacteria, which are known to be important for the environment, medicine and industry. *Bacillus subtilis* has been chosen to fill this gap[9,10] as its biochemistry, physiology and genetics have been studied intensely for more than 40 years. *B. subtilis* is an aerobic, endospore-forming, rod-shaped bacterium commonly found in soil, water sources and in association with plants. *B. subtilis* and its close relatives are an important source of industrial enzymes (such as amylases and proteases), and much of the commercial interest in these bacteria arises from their capacity to secrete these enzymes at gram per litre concentrations. It has therefore been used for the study of protein secretion and for development as a host for the production of heterologous proteins[11]. *B. subtilis* (*natto*) is also used in the production of Natto, a traditional Japanese dish of fermented soya beans.

Under conditions of nutritional starvation, *B. subtilis* stops growing and initiates responses to restore growth by increasing metabolic diversity. These responses include the induction of motility and chemotaxis, and the production of macromolecular hydrolases (proteases and carbohydrases) and antibiotics. If these responses fail to re-establish growth, the cells are induced to form chemically, irradiation- and desiccation-resistant endospores. Sporulation involves a perturbation of the normal cell cycle and the differentiation of a binucleate cell into two cell types. The division of the cell into a smaller forespore and a larger mother cell, each with an entire copy of the chromosome, is the first morphological indication of sporulation. The former is engulfed by the latter and differential expression of their respective genomes, coupled to a complex network of interconnected regulatory pathways and developmental checkpoints, culminates in the programmed death and lysis of the mother cell and release of the mature spore[12]. In an alternative developmental process, *B. subtilis* is also able to differentiate into a physiological state, the competent state, that allows it to undergo genetic transformation[13].

General features of the DNA sequence

Analysis at the replicon level. The *B. subtilis* chromosome has 4,214,810 base pairs (bp), with the origin of replication coinciding with the base numbering start point[14], and the terminus at about 2,017 kilobases (kb)[15]. The average G + C ratio is 43.5%, but it varies considerably throughout the chromosome. This average is also different if one considers the nucleotide content of coding sequences, for which G and A (24% and 30%) are relatively more abundant than their counterparts C and T (20% and 26%). A significant inversion of the relative G − C/G + C ratio is visible at the origin of replication, indicating asymmetry of the nucleotide composition between the replication leading strand and the lagging strand[16]. Several A + T-rich islands are likely to reveal the signature of bacteriophage lysogens or other inserted elements (Fig. 1, see below).

We have analysed the abundance of oligonucleotides ('words') in the genome in various ways: absolute number of words in the genomic text, or comparison with the expected count derived from several models of the chromosome (for example, Markov models, or simulated sequences in which previously known features of the genome were conserved[17]). Comparing the experimental data with various models allowed us to define under- and overrepresentation of words in the experimental data set by reference to the model chosen. In general, the dinucleotide bias follows closely what has been described for other prokaryotes[18,19], in that the dinucleotides most overrepresented are AA, TT and GC, whereas those less represented are TA, AC and GT. Plots of the frequencies of AG, GA, CT and TC in sliding windows along the chromosome show dramatic decreases or increases around the origin and terminus of replication (data not shown). Trinucleotide frequency, directly related to the coding frame, will be discussed below. The distribution of words of four, five and six nucleotides shows significant correlations between the usage of some words and replication (several such oligonucleotides are very significantly overrepresented in one of the strands and underrepresented in the other one).

Setting a statistical cut-off for the significance of duplications at 10^{-3}, we expected duplication by chance of words longer than 24 nucleotides to be rare[20]. In fact, the genome of *B. subtilis* contains a plethora of such duplications, some of them appearing more than

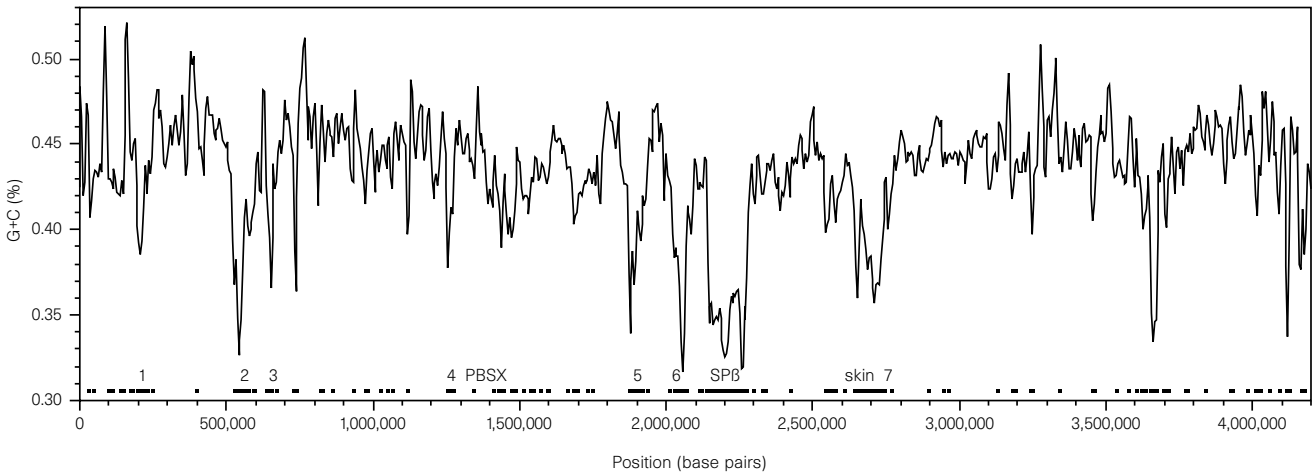


Figure 1 Distribution of A + T-rich islands along the chromosome of *B. subtilis*, in sliding windows of 10,000 nucleotides, with a step of 5,000 nucleotides. Location of genes from class 3 according to codon usage analysis (see Fig. 4) is indicated by dots at the bottom of the graph. Known prophages (PBSX, SP*β* and *skin*) are indicated by their names, and prophage-like elements are numbered from 1 to 7.

twice. Among the duplications, we identified, as expected, the ribosomal RNA genes and their flanking regions, but also regions known to correspond to genes comprising long sequence repeats (such as *pks* and *srf*). We also found several regions that were not expected: a 182-bp repetition within the *yyaL* and *yyaO* genes; a 410-bp repetition between the *yxaK* and *yxaL* genes; an internal duplication of 174 bp inside *ydcI*; and significant duplications in the regions involved in the transcriptional control of several genes (such as 118 bp repeated three times between *yxbB* and *yxbC*). Finally, we found several repetitions at the borders of regions that might be involved in bacteriophage integration.

The most prominent duplication was a 190-bp element that was repeated 10 times in the chromosome. Multiple alignment of the ten repeats showed that they could be classified into two subfamilies with six and three copies each, plus a copy of what appears to be a chimaera. Similar sequences have also been described in the closely related species *Bacillus licheniformis*[21,22]. A striking feature of these repeats is that they are only found in half of the chromosome, at either side of the origin of replication, with five repeats on each side. Furthermore, with the exception of the most distal repeat at position 737,062, they lie in the same orientation with respect to the movement of the replication fork (Figs 2 and 3). Putative secondary structures conserved by compensatory mutations, as well as an insert in three of the copies, suggest that this element could indicate a structural RNA molecule.

Analysis at the transcription and translation level. Over 4,000 putative protein coding sequences (CDSs) have been identified, with an average size of 890 bp, covering 87% of the genome sequence (Fig. 2). We found that 78% of the genes started with ATG, 13% with TTG and 9% with GTG, which compares with 85%, 3% and 14%, respectively, in *E. coli*[8]. Fifteen genes (eight in the predicted CDSs in bacteriophage SPβ) exhibiting unusual start codons (namely ATT and CTG) were also identified through their similarities to known genes in other organisms or because they had a good GeneMark prediction (see Methods). This has not yet been substantiated experimentally. However, in the case of the gene coding for translation initiation factor 3, the similarity with its *E. coli* counterpart strongly suggests that the initiation codon is ATT, as is the case in *E. coli*.

We have not annotated CDSs that largely or entirely overlap existing genes, although such genes (for example, *comS* inside *srfAA*) certainly exist. It is also likely that some of the short CDSs present in the *B. subtilis* genome have been overlooked. For these reasons and possible sequencing errors, the estimated number of *B. subtilis* CDSs will fluctuate around the present figure of 4,100.

In several cases, in-frame termination codons or frameshifts were confirmed to be present on the chromosome (for example, an internal termination codon in *ywtF*, or the known programmed translational frameshift in *prfB*), indicating that the genes are either non-functional (pseudogenes) or subject to regulatory processes. It will therefore be of interest to determine whether these gene features are conserved in related *Bacillus* species, especially as strain 168 is derived from the Marburg strain that was subjected to X-ray irradiation[23].

A few regions do not have any identifiable feature indicating that they are transcribed: they could be 'grey holes' of the type described in *E. coli*[24]. Preliminary studies involving all regions of more than 400 bp without annotated CDSs indicated that, of ~300 such regions, only 15% were likely to be really devoid of protein-coding sequences. One of the longest such regions, located between *yfjO* and *yfjN*, is 1,628 bp long. Grey holes seem generally to be clustered near the terminus of replication. However, a grey-hole cluster located at ~600 kb might be related to the temporary chromosome partition observed during the first stages of sporulation, when a segment of about one-third of the chromosome enters the prespore, and remains the sole part of the chromosome in the prespore for a significant transition period[25].

The codon usage of *B. subtilis* CDSs was analysed using factorial correspondence analysis[17]. We found that the CDSs of *B. subtilis* could be separated into three well-defined classes (Fig. 4). Class 1 comprises the majority of the *B. subtilis* genes (3,375 CDSs), including most of the genes involved in sporulation. Class 2 (188 CDSs) includes genes that are highly expressed under exponential growth conditions, such as genes encoding the transcription and translation machineries, core intermediary metabolism, stress proteins, and one-third of genes of unknown function. Class 3 (537 CDSs) contains a very high proportion of genes of unidentified function (84%), and the members of this class have codons enriched in A + T residues. These genes are usually clustered into groups between 15 and 160 genes (for example, bacteriophage SPβ) and correspond to the A + T-rich islands described above (Fig. 1). When they are of known function, or when their products display similarity to proteins of known function, they usually correspond to functions found in, or associated with, bacteriophages or transposons, as well as functions related to the cell envelope. This includes the region *ydc/ydd/yde* (40 genes that are missing in some *B. subtilis* strains[26]), where gene products showing similarities to bacteriophage and transposon proteins are intertwined. Many of these genes are associated with virulence genes identified in pathogenic Gram-positive bacteria, suggesting that such virulence factors are transmitted horizontally among bacteria at a much higher frequency than previously thought. If we include these A + T-rich regions as possible cryptic phages, together with known bacteriophages or bacteriophage-like elements (SPβ, PBSX and the *skin* element), we find that the genome of *B. subtilis* 168 contains at least 10 such elements (Figs 2 and 3). Annotation of the corresponding regions often reveals the presence of genes that are similar to bacteriophage lytic enzymes, perhaps accounting for the observation that *B. subtilis* cultures are extremely prone to lysis.

The ribosomal RNA genes have been previously identified and

Table 1 Functional classification of the *Bacillus subtilis* protein-coding genes

The genes of known function or encoding products similar to known proteins in *B. subtilis* or in other organisms have been classified into functional categories (2,379 genes). The total number of genes in each category is indicated after the category title. Genes are listed in alphabetical order within each category, and their positions (in kilobases) on the *B. subtilis* chromosome are indicated after the gene names. A brief description is given for each gene. In some cases, interacting proteins have been indicated between brackets (for example, histidine kinases and response regulator, phosphatases and their substrates). More detailed and constantly updated information is available in the SubtiList database (see Methods). A preliminary assessment of the significance of sequence similarities was obtained through an automated procedure involving a combination between the BLAST2P probability and the percentage of amino-acid identity. Matches considered significant were re-examined manually. It should be emphasized that functions assigned to '*y*' genes are based only on sequence similarity information with the best counterparts in protein databanks. Genes whose products are only similar to other unknown proteins, or not significantly similar to any other proteins in databanks (categories V and VI), were omitted.

Figure 2 General view of the *B. subtilis* chromosome. Arrows indicate the orientation of transcription. Genes are coloured according to their classification into six broad functional categories (blue, category I; green, category II; red, category III; orange, category IV; purple, category V; pink, category VI; see Table 1). Class 2 CDSs according to codon usage analysis are indicated by oblique hatches, and class 3 CDSs are indicated by vertical hatches. Ribosomal RNA genes are coloured in yellow. Transfer RNA genes are marked by triangles. Other RNA genes are represented as white arrows. Known genes (non-'*y*' genes) are printed in bold type. Putative transcription termination sites are represented as loops. Known prophages and prophage-like elements are indicated by brown hatches on the chromosome line. The 190-bp element repeated ten times is represented by hatched boxes.

shown to be organized into ten rRNA operons, mainly clustered around the origin of replication of the chromosome (Figs 2 and 3). In addition to the 84 previously identified tRNA genes, by using the Palingol[27] and tRNAscan[28] programs, we propose four putative new tRNA loci (at 1,262 kb, 1,945 kb, 2,003 kb and 2,899 kb), specific for lysine, proline and arginine (UUU, GGG, CCU and UCU anticodons, respectively). The 10S RNA involved in degradation of proteins made from truncated mRNA has been identified (*ssrA*), as well as the RNA component of RNase P (*rnpB*) and the 4.5S RNA involved in the secretion apparatus (*scr*).

There is a strong transcription orientation bias with respect to the movement of the replication fork: 75% of the predicted genes are transcribed in the direction of replication. Plotting the density of coding nucleotides in each strand along the chromosome readily identifies the replication origin and terminus (Fig. 3). To identify putative operons, we followed ref. 29 for describing Rho-independent transcription termination sites. This yielded ~1,630 putative terminators (340 of which were bidirectional). We retained only those that were located less than 100 bp downstream of a gene, or that were considered by the program to be 'very strong' (in order to account for possible erroneous CDSs). This yielded a total of ~1,250 terminators, with a mean operon size of three genes. A similar approach to the identification of promoters is problematical, especially because at least 14 sigma factors, recognizing different promoter sequences, have been identified in *B. subtilis*. Nevertheless, the consensus of the main vegetative sigma factor (σ^A) appears to be identical to its counterpart in *E. coli* (σ^{70}): 5′-TTGACA-n_{17}-TATAAT-3′. Relaxing the constraints of the similarity to sigma-specific consensus sequences led to an extremely high number of false-positive results, suggesting that the consensus-oriented approach to the identification of promoters should be replaced by another approach[17].

Classification of gene products

Genes were classified according to ref. 14, based on the representation of cells as Turing machines in which one distinguishes between the machine and the program (Table 1). Using the BLAST2P software running against a composite protein databank compound of SWISS-PROT (release 34), TREMBL (release 3, update 1) and *B. subtilis* proteins, we assigned at least one significant counterpart with a known function to 58% of the *B. subtilis* proteins. Thus for up to 42% of the gene products, the function cannot be predicted by similarity to proteins of known function: 4% of the proteins are similar only to other unknown proteins of *B. subtilis*; 12% are similar to unknown proteins from some other organism; and 26% of the proteins are not significantly similar to any other proteins in databanks. This preliminary analysis should be interpreted with caution, because only ~1,200 gene functions (30%) have been experimentally identified in *B. subtilis*. We used the '*y*' prefix in gene names to emphasize that the function has not been ascertained (2,853 '*y*' genes, representing 70%).

Regulatory systems. Transcription regulatory proteins. Helix–turn–helix proteins form a large family of regulatory proteins found in both prokaryotes and eukaryotes. There are several classes, including repressors, activators and sigma factors. Using BLAST searches, we constructed consensus matrices for helix–turn–helix proteins to analyse the *B. subtilis* protein library. We identified 18 sigma or sigma-like factors, of which nine (including a new one) are of the SigA type. We also putatively identified 20 regulators (among which 18 were products of '*y*' genes) of the GntR family, 19 regulators (15 '*y*' genes) of the LysR family, and 12 regulators (5 '*y*' genes) of the LacI family. Other transcription regulatory proteins were of the AraC family (11 members, 10 '*y*'), the Lrp family (7 members, 3 '*y*'), the DeoR family (6 members, 3 '*y*'), or additional families (such as the MarR, ArsR or TetR families). A puzzling observation is that several regulatory proteins display significant similarity to aminotransferases (seven such enzymes have been identified as showing similarity to repressors).

Two-component signal-transduction pathways. Two-component regulatory systems, consisting of a sensor protein kinase and a response regulator, are widespread among prokaryotes. We have identified 34 genes encoding response regulators in *B. subtilis*, most of which have adjacent genes encoding histidine kinases. Response regulators possess a well-conserved N-terminal phospho-acceptor domain[30], whereas their C-terminal DNA-binding domains share similarities with previously identified response regulators in *E. coli*, *Rhizobium meliloti*, *Klebsiella pneumoniae* or *Staphylococcus aureus*. Representatives of the four subfamilies recently identified in *E. coli*[31]

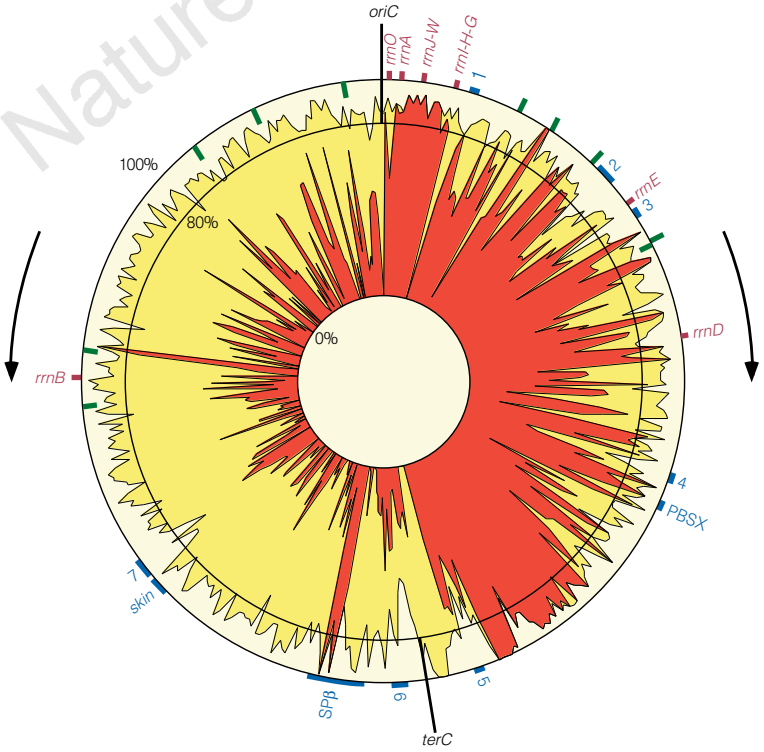


Figure 3 Density of coding nucleotides along the *B. subtilis* chromosome. Yellow stands for the density of coding nucleotides in both strands of the sequence; red indicates the density of coding nucleotides in the clockwise strand (nucleotides involved in genes transcribed in the clockwise orientation). The movement of the replication forks is represented by arrows. Ribosomal RNA operons are indicated by brown boxes. Known prophages and prophage-like elements are represented as blue lines. The 190-bp element repeated ten times is represented by green lines.

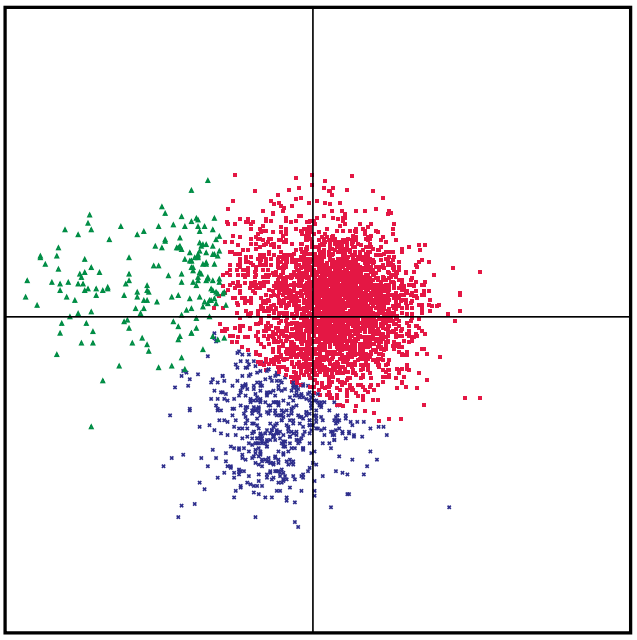


Figure 4 Factorial correspondence analysis of codon usage in the *B. subtilis* CDSs. Red dots, genes from class 1; green triangles, genes from class 2; blue crosses, genes from class 3. Class 2 contains genes coding for the translation and transcription machineries, and genes of the core intermediary metabolism. Class 3 genes correspond to codons strongly enriched in A or T in the wobble position; they generally belong to prophage-like inserts in the genome.

(OmpR, FixJ, CitB and LytR) have been identified in *B. subtilis*. In a fifth subfamily, CheY, the DNA-binding domain is absent. The DNA-binding domain of a single *B. subtilis* response regulator, YesN, shares similarity with regulatory proteins of the AraC family.

Quorum sensing. The *B. subtilis* genome contains 11 aspartate phosphatase genes, whose products are involved in dephosphorylation of response regulators, that do not seem to have counterparts in Gram-negative bacteria such as *E. coli*. Downstream from the corresponding genes are some small genes, called *phr*, encoding regulatory peptides that may serve as quorum sensors[32]. Seven *phr* genes have been identified so far, including three new genes (*phrG*, *phrI* and *phrK*).

Protein secretion. It is known that *B. subtilis* and related *Bacillus* species, in particular *B. licheniformis* and *B. amyloliquefaciens*, have a high capacity to secrete proteins into the culture medium. Several genes encoding proteins of the major secretion pathway have been identified: *secA*, *secD*, *secE*, *secF*, *secY*, *ffh* and *ftsY*. Surprisingly, there is no gene for the SecB chaperone. It is thought that other chaperone(s) and targeting factor(s), such as Ffh and FtsY, may take over the SecB function. Further, although there is only one such gene in *E. coli*, five type I signal peptidase genes (*sipS*, *sipT*, *sipU*, *sipV* and *sipW*) have been found[33]. The *lsp* gene, encoding a type II signal peptidase required for processing of lipo-modified precursors, was also identified. PrsA, located at the outer side of the membrane, is important for the refolding of several mature proteins after their translocation through the membrane.

Other families of proteins. ABC transporters were the most frequent class of proteins found in *B. subtilis*. They must be extremely important in Gram-positive bacteria, because they have an envelope comprising a single membrane. ABC transporters will therefore allow such bacteria to escape the toxic action of many compounds. We propose that 77 such transporters are encoded in the genome. In general they involve the interaction of at least three gene products, specified by genes organized into an operon. Other families comprised 47 transport proteins similar to facilitators (and perhaps sometimes part of the ABC transport systems), 18 amino-acid permeases (probably antiporters), and at least 16 sugar transporters belonging to the PEP-dependent phosphotransferase system.

General stress proteins are important for the survival of bacteria under a variety of environmental conditions. We identified 43 temperature-shock and general stress proteins displaying strong similarity to *E. coli* counterparts.

Missing genes. Histone-like proteins such as HU and H-NS have been identified in *E. coli*. We found that *B. subtilis* encodes two putative histone-like proteins that show similarity to *E. coli* HU, namely HBsu and YonN, but found no homologue to H-NS. It is known that the *hbs* gene encoding HBsu is essential, but we do not expect the *yonN* gene to be essential because it is present in the SPβ prophage. IHF is similar to HU, and it is not known whether HBsu plays a similar role to that of IHF in *E. coli*. Similarly, no protein similar to FIS could be found.

Genes encoding products that interact with methylated DNA, such as *seqA* in *E. coli*, involved in the regulation of replication initiation timing, or *mutH*, the endonuclease recognizing the newly synthesized strand during mismatch repair at hemi-methylated

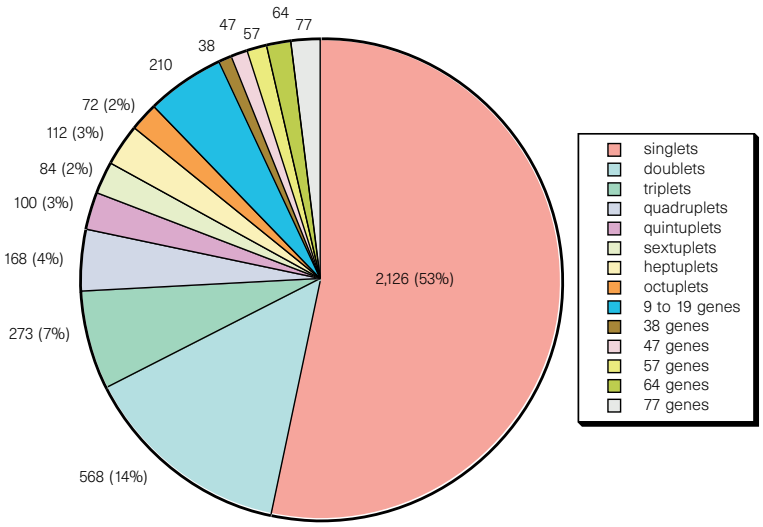


Figure 5 Gene paralogue distribution in the genome of *B. subtilis*. Each *B. subtilis* protein has been compared with all other proteins in the genome, using a Smith and Waterman algorithm. The baseline is established by making a similar comparison using 100 independent random shuffles of the protein sequence (Z-score > 13).

GATC sites, are also missing. This is in line with the absence of known methylation in *B. subtilis*, equivalent to Dam methylation in *E. coli*. Similarly, *E. coli sfiA*, encoding an inhibitor of FtsZ action in the SOS response, has no counterpart in *B. subtilis*. In contrast, *B. subtilis* replication initiation-specific genes, such as *dnaB* and *dnaD*, are missing in *E. coli*. The exact counterpart of the *E. coli mukB* gene, involved in chromosome partitioning, does not exist in *B. subtilis*, but genes *spo0J* and *smc* (Smc is weakly similar to MukB), which are suggested to be involved in partitioning of the *B. subtilis* chromosome, are missing in *E. coli*.

Turnover of mRNA is controlled in *E. coli* by a 'degradosome' comprising RNase E. It has a counterpart in *B. subtilis*, but we failed to find a clear homologue of RNase E in this organism. Whether this is related to the role of ribosomal protein S1 as an RNA helicase involved in mRNA turnover in *E. coli* requires further investigation. In particular, a homologue of *rpsA* (S1 structural gene), *ypfD*, might be involved in a structure homologous to the degradosome[34].

Structurally unrelated genes of similar function. Several genes encode products that have similar functions in *E. coli* and *B. subtilis*, but have no evident common structure. This is the case for the helicase loader genes, *E. coli dnaC* and *B. subtilis dnaI*; the genes coding for the replication termination protein, *E. coli tus* and *B. subtilis rtp*; and the division topology specifier genes, *E. coli minE* and *B. subtilis divIVA*. The situation may even be more complex in multisubunit enzymes: *B. subtilis* synthesizes two DNA polymerase III α chains, one having $3'-5'$ proofreading exonuclease activity (PolC) and the other without the exonuclease activity (DnaE); in *E. coli*, only the latter exists. *E. coli* DNA polymerase II is structurally related to DNA polymerase α of eukaryotes, whereas *B. subtilis* YshC is related to DNA polymerase β.

Metabolism of small molecules

The type and range of metabolism used for the interconversion of low-molecular-weight compounds provide important clues to an organism's natural environment(s) and its biological activity. Here we briefly outline the main metabolic pathways of *B. subtilis* before the reconstruction of these pathways *in silico*, the correlation of genes with specific steps in the pathway, and ultimately the prediction of patterns of gene expression.

Intermediary metabolism. It has long been known that *B. subtilis* can use a variety of carbohydrates. As expected, it encodes an Embden–Meyerhof–Parnas glycolytic pathway, coupled to a functional tricarboxylic acid cycle. Further, *B. subtilis* is also able to grow anaerobically in the presence of nitrate as an electron acceptor. This metabolism is, at least in part, regulated by the FNR protein, binding to sites upstream of at least eight genes (four sites experimentally confirmed and four putative sites). A noteworthy feature of *B. subtilis* metabolism is an apparent requirement of branched short-chain carboxylic acids for lipid biosynthesis[35]. Branched-chain 2-keto acid decarboxylase activity exists and may be linked to a variety of genes, suggesting that *B. subtilis* can synthesize and utilize linear branched short-chain carboxylic acids and alcohols.

Amino-acid and nucleotide metabolism. Pyrimidine metabolism of *B. subtilis* seems to be regulated in a way fundamentally different from that of *E. coli*, as it has two carbamylphosphate synthetases (one specific for arginine synthesis, the other for pyrimidine). Additionally, the aspartate transcarbamylase of *B. subtilis* does not act as an allosteric regulator as it does in *E. coli*. As in other microorganisms, pyrimidine deoxyribonucleotides are synthesized from ribonucleoside diphosphates, not triphosphates. The cytidine diphosphate required for DNA synthesis is derived from either the salvage pathway of mRNA turnover or from the synthesis of phospholipids and components of the cell wall. This means that polynucleotide phosphorylase is of fundamental importance in nucleic acid metabolism, and may account for its important role in competence[36]. Two ribonucleoside reductases, both of class I, NrdEF type, are encoded by the *B. subtilis* chromosome, in one case from within the SPβ genome. In this latter case, the gene corresponding to the large subunit both contains an intron and codes for an intein (V.L., unpublished data). The gene of the small subunit of this enzyme also contains an intron, encoding an endonuclease, as was found for the homologue in bacteriophage T4.

By similarity with genes from other organisms, there appears to be, in addition to genes involved in amino-acid degradation (such as the *roc* operon, which degrades arginine and related amino acids), a large number of genes involved in the degradation of molecules such as opines and related molecules, derived from plants. This is also in line with the fact that *B. subtilis* degrades polygalacturonate, and suggests that, in its biotope, it forms specific relations with plants.

Secondary metabolism. In addition to many genes coding for degradative enzymes, almost 4% of the *B. subtilis* genome codes for large multifunctional enzymes (for example, the *srf*, *pps* and *pks* loci), similar to those involved in the synthesis of antibiotics in other genera of Gram-positive bacteria such as *Streptomyces*. Natural isolates of *B. subtilis* produce compounds with antibiotic activity, such as surfactin, fengycin and difficidin, that can be related to the above-mentioned loci. This bacterium therefore provides a simple and genetically amenable model in which to study the synthesis of antibiotics and its regulation. These pathways are often organized in very long operons (for example, the *pks* region spans 78.5 kb, about 2% of the genome). The corresponding sequences are mostly located near the terminus of replication, together with prophages and prophage-like sequences.

Paralogues and orthologues

It is important to relate intermediary metabolism to genome structure, function and evolution. We therefore compared the *B. subtilis* proteins with themselves, as well as with proteins from known complete genomes, using a consistent statistical method that allows the evaluation of unbiased probabilities of similarities between proteins[37,38]. For *Z*-scores higher than 13, the number of proteins similar to each given protein does not vary, indicating that this cut-off value identifies sets of proteins that are significantly similar.

Families of paralogues. Many of the paralogues constitute large families of functionally related proteins, involved in the transport of compounds into and out of the cell, or involved in transcription regulation. Another part of the genome consists of gene doublets (568 genes), triplets (273 genes), quadruplets (168 genes) and quintuplets (100 genes). Finally, about half of the genome is made of genes coding for proteins with no apparent paralogues (Fig. 5). No large family comprises only proteins without any similarity to proteins of known function.

The process by which paralogues are generated is not well understood, but we might find clues by studying some of the duplications in the genome. Several approximate DNA repetitions, associated with very high levels of protein identity, were found, mainly within regions putatively or previously identified as prophages. This is in line with previous observations about PBSX and the *skin* element[39,40], and suggests that these prophage-like elements share a common ancestor and have diverged relatively recently. In addition, several protein duplications are in genes that are located very close to each other, such as *yukL* and *dhbF* (the corresponding proteins are 65% identical in an overlap of 580 amino acids), *yugJ* and *yugK* (proteins 73% identical), *yxjG* and *yxjH* (proteins 70% identical), and the entire *opuB* operon, which is duplicated 3 kb away (*opuC* operon, yielding ~80% of amino-acid identity in the corresponding proteins).

The study of paralogues showed that, as in other genomes, a few classes of genes have been highly expanded. This argues against the idea of the genome evolving through a series of duplications of ancestral genomes, but rather for the idea of genes as living organisms, subject to evolutionary constraints, some being sub-

mitted to expansion and natural selection, and others to local duplications of DNA regions.

Among paralogue doublets, some were unexpected, such as the three aminoacyl tRNA synthetases doublets (*hisS* (2,817 kb) and *hisZ* (3,588 kb); *thrS* (2,960 kb) and *thrZ* (3,855 kb); *tyrS* (3,036 kb) and *tyrZ* (3,945 kb)) or the two *mutS* paralogues (*mutS* and *yshD*). This latter situation is similar to that found in *Synechocystis*. In the case of *B. subtilis*, the presence of two MutS proteins could indicate that there are two different pathways for long-patch mismatch repair, possibly a consequence of the active genetic transformation mechanism of *B. subtilis*.

Families of orthologues. Because *Mycoplasma* spp. are thought to be derived from Gram-positive bacteria similar to *B. subtilis*, we compared the *B. subtilis* genome with that of *M. genitalium*. Among the 450 genes encoded by *M. genitalium*, the products of 300 are similar to proteins of *B. subtilis*. Among the 146 remaining gene products, a further 3 are similar to proteins of other *Bacillus* species, and 9 to proteins of other Gram-positive bacteria; 25 are similar to proteins of Gram-negative bacteria; and 19 are similar to proteins of other *Mycoplasma* spp. This leaves only 90 genes that would be specific to *M. genitalium* and might be involved in the interaction of this organism with its host.

The *B. subtilis* genome is similar in size to that of *E. coli*. Because these bacteria probably diverged more than one billion years ago, it is of evolutionary value to investigate their relative similarity. About 1,000 *B. subtilis* genes have clear orthologous counterparts in *E. coli* (one-quarter of the genome). These genes did not belong either to the prophage-like regions or to regions coding for secondary metabolism (~15% of the *B. subtilis* genome). This indicates that a large fraction of these genomes shared similar functions. At first sight, however, it seems that little of the operon structure has been conserved. We nevertheless found that ~100 putative operons or parts of operons were conserved between *E. coli* and *B. subtilis*. Among these, ~12 exhibited a reshuffled gene order (typically, the arabinose operon is *araABD* in *B. subtilis* and *araBAD* in *E. coli*). In addition to the core of the translation and transcription machinery, we identified other classes of operons that were well conserved between the two organisms, including major integrated functions such as ATP synthesis (*atp* operon) and electron transfer (*cta* and *qox* operons). As well as being well preserved, the murein biosynthetic region was partly duplicated, allowing creation of part of the genes required for the sporulation division machinery[41]. The amino-acid biosynthesis genes differ more in their organization: the *E. coli* genes for arginine biosynthesis are spread throughout the chromosome, whereas the arginine biosynthesis genes of *B. subtilis* form an operon. The same is true for purine biosynthetic genes. Genes responsible for the biosynthesis of coenzymes and prosthetic groups in *B. subtilis* are often clustered in operons that differ from those found in *E. coli*. Finally, several operons conserved in *E. coli* and *B. subtilis* correspond to unknown functions, and should therefore be priority targets for the functional analysis of these model genomes.

Comparison with *Synechocystis* PCC6803 revealed about 800 orthologues. However, in this case the putative operon structure is extremely poorly conserved, apart from four of the ribosomal protein operons, the *groES–groEL* operon, *yfnHG* (respectively in *Synechocystis rfbFG*), *rpsB-tsf*, *ylxS-nusA-infB*, *asd-dapGA-ymfA*, *spmAB*, *efp-accB*, *grpE-dnaK*, *yurXW*. The nine-gene *atp* operon of *B. subtilis* is split into two parts in *Synechocystis*: *atpBE* and *atpIHGFDAC*.

Conclusion

The biochemistry, physiology and molecular biology of *B. subtilis* have been extensively studied over the past 40 years. In particular, *B. subtilis* has been used to study postexponential phase phenomena such as sporulation and competence for DNA uptake. The genome sequences of *E. coli* and *B. subtilis* provide a means of studying the evolutionary divergence, one billion years ago, of eubacteria into the Gram-positive and Gram-negative groups. The availability of powerful genetic tools will allow the *B. subtilis* genome sequence data to be exploited fully within the framework of a systematic functional analysis program, undertaken by a consortium of 19 European and 7 Japanese laboratories coordinated by S. D. Ehrlich (INRA, Jouy-en-Josas, France) and by N. Ogasawara and H. Yoshikawa (Nara Institute of Science and Technology, Nara, Japan). □

Methods

Genome cloning and sequencing. An international consortium was established to sequence the genome of *B. subtilis* strain 168 (refs 9, 10, 42). At its peak, 25 European, seven Japanese and one Korean laboratory participated in the program, together with two biotechnology companies. Five contiguous DNA regions totalling 0.94 Mb, and two additional regions of 0.28 and 0.14 Mb, were sequenced by the Japanese partners, while the European partners sequenced a total of 2.68 Mb. A few sequences from strain 168 published previously were not resequenced when long overlaps did not indicate differences.

A major technical difficulty was the inability to construct in *E. coli* gene banks representative of the entire *B. subtilis* chromosome using vectors that have proved efficient for other sources of bacterial DNA (such as bacteriophage or cosmid vectors). This was due to the generally very high level of expression of *B. subtilis* genes in *E. coli*, leading to toxic effects. This limitation was overcome by: cloning into a variety of vectors[9,43,44]; using an *E. coli* strain maintaining low-copy number plasmids[44]; using an integrative plasmid/marker rescue genome-walking strategy[44]; and *in vitro* amplification using polymerase chain reaction (PCR) techniques[45,46].

Although cloning vectors were used in the early stages as templates for sequencing reactions, they were largely superseded in the later stages by long-range and inverse PCR techniques. To reduce sequencing errors resulting from PCR amplification artefacts, at least eight amplification reactions were performed independently and subsequently pooled. The various sequencing groups were free to choose their own strategy, except that all DNA sequences had to be determined entirely on both strands.

Sequence annotation and verification. The sequences were annotated by the groups, and sent to a central depository at the Institut Pasteur[14]. The Japanese sequences were also sent there through the Japanese depository at the Nara Institute of Science and Technology. The same procedures were used to identify CDSs and to detect frameshifts. They were embedded within a cooperative computer environment dedicated to automatic sequence annotation and analysis[39]. In a first step, we identified in all six possible frames the open reading frames (ORFs) that were at least 100 codons in length. In a second step, three independent methods were used: the first method used the GeneMark coding-sequence prediction method[47] together with the search for CDSs preceded by typical translation initiation signals (5′-AAGGAGGTG-3′), located 4–13 bases upstream of the putative start codons (ATG, TTG or GTG); the second method used the results of a BLAST2X analysis performed on the entire *B. subtilis* genome against the non-redundant protein databank at the NCBI; and the third method was based on the distribution of non-overlapping trinucleotides or hexanucleotides in the three frames of an ORF[48].

In general, frameshifts and missense mutations generating termination codons or eliminating start codons are relatively easy to detect. We shall devise a procedure for detecting another type of error, GC instead of CG or vice versa, which are much more difficult to identify. It should be noted that putative frameshift errors should not be corrected automatically. The sequences of the flanking regions of a 500-bp fragment centred around a putative error were sent to an independent verification group, which performed PCR amplifications using chromosomal DNA as template, and sequenced the corresponding DNA products.

Organization and accessibility of data. The *B. subtilis* sequence data have been combined with data from other sources (biochemical, physiological and genetic) in a specialized database, SubtiList[49], available as a Macintosh or Windows stand-alone application (4th Dimension runtime) by anonymous FTP at ftp://ftp.pasteur.fr/pub/GenomeDB/SubtiList. SubtiList is also accessible through a World-Wide Web server at http://www.pasteur.fr/Bio/SubtiList.html,

where it has been implemented on a UNIX system using the Sybase relational database management system. A completely rewritten version of SubtiList is in preparation to facilitate browsing of the information of the whole chromosome. Flat files of the whole DNA and protein sequences in EMBL and FASTA format will be made available at the above ftp address. Another *B. subtilis* genome database is also under development at the Human Genome Center of Tokyo University (http://www.genome.ad.jp), and SubtiList will also be available there.

Received 16 July; 29 September 1997.

1. Fleischmann, R. D. *et al*. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269,** 496–512 (1995).
2. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270,** 397–403 (1995).
3. Kaneko, T. *et al*. Sequence analysis of the genome of the unicellular Cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3,** 109–136 (1996).
4. Bult, C. J. *et al*. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273,** 1058–1073 (1996).
5. Himmelreich, R. *et al*. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24,** 4420–4449 (1996).
6. Goffeau, A. *et al*. The yeast genome directory. *Nature* **387,** 5–105 (1997).
7. Tomb, J.-F. *et al*. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
8. Blattner, F. R. *et al*. The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1462 (1997).
9. Kunst, F., Vassarotti, A. & Danchin, A. Organization of the European *Bacillus subtilis* genome sequencing project. *Microbiology* **389,** 84–87 (1995).
10. Ogasawara, N. & Yoshikawa, H. The systematic sequencing of the *Bacillus subtilis* genome in Japan. *Microbiology* **142,** 2993–2994 (1996).
11. Harwood, C. R. *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol.* **10,** 247–256 (1992).
12. Stragier, P. & Losick, R. Molecular genetics of sporulation in *Bacillus subtilis*. *Annu. Rev. Genet.* **30,** 297–341 (1996).
13. Solomon, J. M. & Grossman, A. D. Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* **12,** 150–155 (1996).
14. Moszer, I., Kunst, F. & Danchin, A. The European *Bacillus subtilis* genome sequencing project: current status and accessibility of the data from a new World Wide Web site. *Microbiology* **142,** 2987–2991 (1996).
15. Franks, A. H., Griffiths, A. A. & Wake, R. G. Identification and characterization of new DNA replication terminators in *Bacillus subtilis*. *Mol. Microbiol.* **17,** 13–23 (1995).
16. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13,** 660–665 (1996).
17. Hénaut, A. & Danchin, A. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds Neidhardt, F. *et al.*) 2047–2066 (ASM, Washington DC, 1996).
18. Nussinov, R. The universal dinucleotide asymmetry rules in DNA and amino acid codon choice. *Nucleic Acids Res.* **17,** 237–244 (1981).
19. Karlin, S., Burge, C. & Campbell, A. M. Statistical analyses of counts and distributions of restriction sites in DNA sequences. *Nucleic Acids Res.* **20,** 1363–1370 (1992).
20. Burge, C., Campbell, A. M. & Karlin, S. Over- and under-representation of short oligonucleotides in DNA sequences. *Proc. Natl Acad. Sci. USA* **89,** 1358–1362 (1992).
21. Kasahara, Y., Nakai, S. & Ogasawara, H. Sequence analysis of the 36-kb region between *gntZ* and *trnY* genes of *Bacillus subtilis* genome. *DNA Res.* **4,** 155–159 (1997).
22. Presecan, E. *et al*. The *Bacillus subtilis* genome from *gerBC* (311°) to *licR* (334°). *Microbiology* **143,** 3313–3328 (1997).
23. Burkholder, P. R. & Giles, N. H. Induced biochemical mutations in *Bacillus subtilis*. *Am. J. Bot.* **33,** 345–348 (1947).
24. Daniels, D. L., Plunkett, G. III, Burland, V. & Blattner, F. R. Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* **257,** 771–778 (1992).
25. Wu, L. J. & Errington, J. *Bacillus subtilis* SpoIIIE protein required for DNA segregation during asymmetric cell division. *Science* **264,** 572–575 (1994).
26. Itaya, M. Stability and asymmetric replication of the *Bacillus subtilis* 168 chromosome structure. *J. Bacteriol.* **175,** 741–749 (1993).
27. Billoud, B., Kontic, M. & Viari, A. Palingol: a declarative programming language to describe nucleic acids' secondary structures and to scan sequence database. *Nucleic Acids Res.* **24,** 1395–1403 (1996).
28. Fichant, G. A. & Burks, C. Identifying potential tRNA genes in genomic DNA sequences. *J. Mol. Biol.* **220,** 659–671 (1991).
29. d'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent *Escherichia coli* transcription terminators. A statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.* **216,** 835–858 (1990).
30. Stock, J. B., Surette, M. G., Levitt, M. & Park, P. in *Two-Component Signal Transduction* (eds Hoch, J. A. & Silhavy, T. J.) 25–51 (ASM, Washington DC, 1995).
31. Mizuno, T. Compilation of all genes encoding two-component phosphotransfer signal transducers in the genome of *Escherichia coli*. *DNA Res.* **4,** 161–168 (1997).
32. Perego, M., Glaser, P. & Hoch, J. A. Aspartyl-phosphate phosphatases deactivate the response regulator components of the sporulation signal transduction system in *Bacillus subtilis*. *Mol. Microbiol.* **19,** 1151–1157 (1996).
33. Tjalsma, H. *et al*. *Bacillus subtilis* contains four closely related type I signal peptidases with overlapping substrate specificities: constitutive and temporally controlled expression of different *sip* genes. *J. Biol. Chem.* **272,** 25983–25992 (1997).
34. Danchin, A. Comparison between the *Escherichia coli* and *Bacillus subtilis* genomes suggests that a major function of polynucleotide phosphorylase is to synthesize CDP. *DNA Res.* **4,** 9–18 (1997).
35. Suutari, M. & Laakso, S. Unsaturated and branched chain-fatty acids in temperature adaptation of *Bacillus subtilis* and *Bacillus megaterium*. *Biochim. Biophys. Acta* **1126,** 119–124 (1992).
36. Luttinger, A., Hahn, J. & Dubnau, D. Polynucleotide phosphorylase is necessary for competence development in *Bacillus subtilis*. *Mol. Microbiol.* **19,** 343–356 (1996).
37. Landès, C., Hénaut, A. & Risler, J.-L. A comparison of several similarity indices used in the classification of protein sequences: a multivariate analysis. *Nucleic Acids Res.* **20,** 3631–3637 (1992).
38. Glémet, E. & Codani, J.-J. LASSAP, a LArge Scale Sequence compArison Package. *Comput. Appl. Biosci.* **13,** 137–143 (1997).
39. Médigue, C., Moszer, I., Viari, A. & Danchin, A. Analysis of a *Bacillus subtilis* genome fragment using a co-operative computer system prototype. *Gene* **165,** GC37–GC51 (1995).
40. Krogh, S., O'Reilly, M., Nolan, N. & Devine, K. M. The phage-like element PBSX and part of the *skin* element, which are resident at different locations on the *Bacillus subtilis* chromosome, are highly homologous. *Microbiology* **142,** 2031–2040 (1996).
41. Daniel, R. A., Drake, S., Buchanan, C. E., Scholle, R. & Errington, J. The *Bacillus subtilis spoVD* gene encodes a mother-cell-specific penicillin-binding protein required for spore morphogenesis. *J. Mol. Biol.* **235,** 209–220 (1994).
42. Anagnostopoulos, C. & Spizizen, J. Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.* **81,** 741–746 (1961).
43. Azevedo, V. *et al*. An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. *Proc. Natl Acad. Sci. USA* **90,** 6047–6051 (1993).
44. Glaser, P. *et al*. *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb region from 325° to 333°. *Mol. Microbiol.* **10,** 371–384 (1993).
45. Ogasawara, N., Nakai, S. & Yoshikawa, H. Systematic sequencing of the 180 kilobase region of the *Bacillus subtilis* chromosome containing the replication origin. *DNA Res.* **1,** 1–14 (1994).
46. Sorokin, A. *et al*. A new approach using multiplex long accurate PCR and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome Res.* **6,** 448–453 (1996).
47. Borodovsky, M. & McIninch, J. GENMARK: parallel gene recognition for both DNA strands. *Comput. Chem.* **17,** 123–133 (1993).
48. Fichant, G. A. & Quentin, Y. A frameshift error detection algorithm for DNA sequencing projects. *Nucleic Acids Res.* **23,** 2900–2908 (1995).
49. Moszer, I., Glaser, P. & Danchin, A. SubtiList: a relational database for the *Bacillus subtilis* genome. *Microbiology* **141,** 261–268 (1995).

Acknowledgements. We thank C. Anagnostopoulos, R. Dedonder and J. Hoch for their pioneering efforts, and A. Bairoch for advice in annotating *B. subtilis* protein data. The main funding of the European network was provided by the European Commission under the Biotechnology program. The Japanese project was included in the Human Genome Program, and supported by a research grant from the Ministry of Education, Science and Culture, and the Proposal-Based Advanced Industrial Technology R&D Program from New Energy and Industrial Technology Development Organization. The Swiss and Korean projects were funded by the Swiss National Fund and the Korean government, respectively. An industrial platform was set up to facilitate contacts between participants of the European consortium and some European biotechnology companies: DuPont de Nemours (France, USA), Frimond (Belgium), Genencor (Finland, USA), Gist Brocades (The Netherlands), Glaxo-Wellcome (UK, Italy), Hoechst Marion Roussel (France, Germany), F. Hoffmann-La Roche AG (Switzerland), Novo Nordisk (Denmark), SmithKline Beecham (UK).

Correspondence and requests for materials should be addressed to F.K. (e-mail: fkunst@pasteur.fr), N.O. (nogasawa@bs.aist-nara.ac.jp), H.Y. (hyoshika:bs.aist-nara.ac.jp) or A.D. (adanchin@pasteur.fr). The sequence has been deposited in EMBL/GenBank/DDBJ with accession numbers from Z99104 to Z99124.

Table 1. Functional classification of the *Bacillus subtilis* protein-coding genes.

I		CELL ENVELOPE AND CELLULAR PROCESSES 866
I.1		CELL WALL 93
cwlA	2665	*N*-acetylmuramoyl-L-alanine amidase (minor autolysin)
cwlC	1873	*N*-acetylmuramoyl-L-alanine amidase (sporulation mother cell wall)
cwlD	157	*N*-acetylmuramoyl-L-alanine amidase (germination)
cwlJ	282	cell wall hydrolase (sporulation)
dacA	18	penicillin-binding protein 5 (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis)
dacB	2424	penicillin-binding protein 5* (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis) (spore cortex)
dacF	2445	penicillin-binding protein (D-alanyl-D-alanine carboxypeptidase) (peptidoglycan biosynthesis)
ddlA	508	D-alanyl-D-alanine ligase A (peptidoglycan biosynthesis)
dltA	3951	D-alanyl-D-alanine carrier protein ligase (lipoteichoic acid biosynthesis)
dltB	3953	D-alanine transfer from Dcp to undecaprenol-phosphate (lipoteichoic acid biosynthesis)
dltC	3954	D-alanine carrier protein (lipoteichoic acid biosynthesis)
dltD	3954	D-alanine transfer from undecaprenol-phosphate to the poly(glycerophosphate) chain (lipoteichoic acid biosynthesis)
dltE	3955	involved in lipoteichoic acid biosynthesis
gcaD	56	UDP-*N*-acetylglucosamine pyrophosphorylase (peptidoglycan and lipopolysaccharide biosynthesis)
ggaA	3670	galactosamine-containing minor teichoic acid biosynthesis
ggaB	3669	galactosamine-containing minor teichoic acid biosynthesis
gtaB	3665	UTP-glucose-1-phosphate uridylyltransferase
lytB	3662	modifier protein of major autolysin LytC (CWBP76)
lytC	3660	*N*-acetylmuramoyl-L-alanine amidase (major autolysin) (CWBP49)
lytD	3687	*N*-acetylglucosaminidase (major autolysin) (CWBP90)
lytE	1018	cell wall lytic activity (CWBP33)
mbl	3747	MreB-like protein
mraY	1587	phospho-*N*-acetylmuramoyl-pentapeptide transferase (peptidoglycan biosynthesis)
mreB	2861	cell-shape determining protein
mreBH	1517	cell-shape determining protein
mreC	2860	cell-shape determining protein
mreD	2859	cell-shape determining protein
murA	3778	UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (peptidoglycan biosynthesis)
murB	1592	UDP-*N*-acetylenolpyruvoylglucosamine reductase (peptidoglycan biosynthesis)
murC	3049	UDP-*N*-acetylmuramate-alanine ligase (peptidoglycan biosynthesis)
murD	1588	UDP-*N*-acetylmuramoylalanine-D-glutamate ligase (peptidoglycan biosynthesis)
murE	1586	UDP-*N*-acetylmuramoylananine-D-glutamate-2,6-diaminopimelate ligase (peptidoglycan biosynthesis)
murF	509	UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl-D-alanyl ligase (peptidoglycan biosynthesis)
murG	1591	UDP-*N*-acetylglucosamine-*N*-acetylmuramyl-(pentapeptide)pyrophosphoryl-undecaprenol *N*-acetylglucosamine transferase (peptidoglycan biosynthesis)
murZ	3806	UDP-*N*-acetylglucosamine 1-carboxyvinyltransferase (peptidoglycan biosynthesis)
pbp	1999	penicillin-binding protein (peptidoglycan biosynthesis)
pbpA	2583	penicillin-binding protein 2A (peptidoglycan biosynthesis) (spore outgrowth)
pbpB	1581	penicillin-binding protein 2B (peptidoglycan biosynthesis) (cell-division septum)
pbpC	463	penicillin-binding protein 3 (peptidoglycan biosynthesis)
pbpD	3233	penicillin-binding protein 4 (peptidoglycan biosynthesis)
pbpE	3535	penicillin-binding protein 4* (peptidoglycan biosynthesis) (spore cortex)
pbpF	1083	penicillin-binding protein 1A (peptidoglycan biosynthesis) (germination)
pbpX	1765	penicillin-binding protein (peptidoglycan biosynthesis)
ponA	2341	penicillin-binding proteins 1A/1B (peptidoglycan biosynthesis)
racE	2903	glutamate racemase (peptidoglycan biosynthesis)
spoVD	1584	penicillin-binding protein (peptidoglycan biosynthesis) (spore cortex)
tagA	3680	involved in polyglycerol phosphate teichoic acid biosynthesis
tagB	3681	involved in polyglycerol phosphate teichoic acid biosynthesis
tagC	3682	involved in polyglycerol phosphate teichoic acid biosynthesis
tagD	3680	glycerol-3-phosphate cytidylyltransferase (teichoic acid biosynthesis)
tagE	3679	UDP-glucose:polyglycerol phosphate glucosyltransferase (teichoic acid biosynthesis)
tagF	3677	CDP-glycerol:polyglycerol phosphate glycerophosphotransferase (teichoic acid biosynthesis)
tagG	3675	teichoic acid translocation (permease)
tagH	3674	teichoic acid translocation (ATP-binding protein)
tagO	3649	teichoic acid linkage unit synthesis
tuaA	3658	biosynthesis of teichuronic acid
tuaB	3657	biosynthesis of teichuronic acid
tuaC	3656	biosynthesis of teichuronic acid
tuaD	3655	biosynthesis of teichuronic acid (UDP-glucose 6-dehydrogenase)
tuaE	3653	biosynthesis of teichuronic acid
tuaF	3652	biosynthesis of teichuronic acid
tuaG	3651	biosynthesis of teichuronic acid
tuaH	3650	biosynthesis of teichuronic acid
wapA	4029	cell wall-associated protein precursor (CWBP200, 105, 62)
wprA	1153	cell wall-associated protein precursor (CWBP23 and serine protease CWBP52)
xlyA	1347	*N*-acetylmuramoyl-L-alanine amidase (PBSX prophage-mediated lysis)
xlyB	1317	*N*-acetylmuramoyl-L-alanine amidase (PBSX prophage-mediated lysis)
yfnG	799	CDP-glucose 4,6-dehydratase
yhdD	1013	cell wall-binding protein
ykuA	1467	penicillin-binding protein
ylbI	1569	lipopolysaccharide core biosynthesis
ymaG	1865	cell wall protein
yngB	1946	UTP-glucose-1-phosphate uridylyltransferase
yocH	2093	cell wall-binding protein
yodJ	2135	D-alanyl-D-alanine carboxypeptidase
yojL	2116	cell wall-binding protein
yomC	2263	*N*-acetylmuramoyl-L-alanine amidase
ypdQ	2310	cell wall enzyme
ypfP	2306	cell wall synthesis
ypjH	2357	lipopolysaccharide biosynthesis-related protein
yqeE	2649	*N*-acetylmuramoyl-L-alanine amidase
yqfY	2588	peptidoglycan acetylation
yqiI	2515	*N*-acetylmuramoyl-L-alanine amidase
yrhL	2771	acyltransferase
yrrR	2791	penicillin-binding protein
yrvJ	2818	*N*-acetylmuramoyl-L-alanine amidase
ytcC	3157	lipopolysaccharide *N*-acetylglucosaminyltransferase
ytkC	3135	autolytic amidase
ytxN	3161	lipopolysaccharide *N*-acetylglucosaminyltransferase
yubE	3191	*N*-acetylmuramoyl-L-alanine amidase
yvcE	3575	cell wall-binding protein
ywhE	3849	penicillin-binding protein
ywtD	3697	murein hydrolase
I.2		TRANSPORT/BINDING PROTEINS AND LIPOPROTEINS 381
aapA	2766	amino acid permease
alsT	1938	amino acid carrier protein
amyC	3099	maltose transport protein
amyD	3098	sugar transport
appA	1213	oligopeptide ABC transporter (oligopeptide-binding protein)
appB	1215	oligopeptide ABC transporter (permease)
appC	1216	oligopeptide ABC transporter (permease)
appD	1211	oligopeptide ABC transporter (ATP-binding protein)
appF	1212	oligopeptide ABC transporter (ATP-binding protein)
araE	3485	L-arabinose transport (permease)
araN	2942	L-arabinose transport (sugar-binding protein)
araP	2941	L-arabinose transport (integral membrane protein)
araQ	2940	L-arabinose transport (integral membrane protein)
azlC	2729	branched-chain amino acid transport
azlD	2728	branched-chain amino acid transport
bglP	4034	phosphotransferase system (PTS) β-glucoside-specific enzyme IIABC component
blt	2716	multidrug-efflux transporter
bmr	2494	multidrug-efflux transporter
braB	3027	branched-chain amino acid transporter
brnQ	2728	branched-chain amino acid transporter
citM	834	secondary transporter of the Mg^{2+}/citrate complex
csbX	2838	α-ketoglutarate permease
cydC	3976	ABC transporter required for expression of cytochrome *bd* (ATP-binding protein)
cydD	3974	ABC transporter required for expression of cytochrome *bd* (ATP-binding protein)
czcD	2724	cation-efflux system membrane protein
dppA	1360	dipeptide ABC transporter (sporulation)
dppB	1361	dipeptide ABC transporter (permease) (sporulation)
dppC	1362	dipeptide ABC transporter (permease) (sporulation)
dppD	1363	dipeptide ABC transporter (ATP-binding protein) (sporulation)
dppE	1364	dipeptide ABC transporter (dipeptide-binding protein) (sporulation)
ebrA	1865	multidrug resistance protein
ebrB	1864	multidrug resistance protein
ecsA	1077	ABC transporter (ATP-binding protein)
ecsB	1078	ABC transporter (membrane protein)
expZ	606	ATP-binding transport protein
feuA	183	iron-uptake system (binding protein)
feuB	182	iron-uptake system (integral membrane protein)
feuC	181	iron-uptake system (integral membrane protein)
fhuB	3417	ferrichrome ABC transporter (permease)
fhuC	3415	ferrichrome ABC transporter (ATP-binding protein)
fhuD	3418	ferrichrome ABC transporter (ferrichrome-binding protein)
fhuG	3416	ferrichrome ABC transporter (permease)
fruA	1509	phosphotransferase system (PTS) fructose-specific enzyme IIBC component
gabP	686	γ-aminobutyrate permease
glnH	2802	glutamine ABC transporter (glutamine-binding)
glnM	2803	glutamine ABC transporter (membrane protein)
glnP	2804	glutamine ABC transporter (membrane protein)
glnQ	2802	glutamine ABC transporter (ATP-binding protein)
glpF	1002	glycerol uptake facilitator
glpT	235	glycerol-3-phosphate permease
gltP	255	H^+/glutamate symport protein
gltT	1097	H^+/Na^+-glutamate symport protein
glvC	892	phosphotransferase system (PTS) arbutin-like enzyme IIBC component
gntP	4115	gluconate permease (gluconate utilization)
hisP	3004	histidine transport protein (ATP-binding protein)
hutM	4046	histidine permease
iolF	4077	inositol transport protein
kdgT	2322	2-keto-3-deoxygluconate permease (pectin utilization)
lctP	330	L-lactate permease
levD	2762	phosphotransferase system (PTS) fructose-specific enzyme IIA component
levE	2762	phosphotransferase system (PTS) fructose-specific enzyme IIB component
levF	2761	phosphotransferase system (PTS) fructose-specific enzyme IIC component
levG	2760	phosphotransferase system (PTS) fructose-specific enzyme IID component
licA	3959	phosphotransferase system (PTS) lichenan-specific enzyme IIA component
licB	3961	phosphotransferase system (PTS) lichenan-specific enzyme IIB component
licC	3960	phosphotransferase system (PTS) lichenan-specific enzyme IIC component
lmrB	290	lincomycin-resistance protein
lplA	779	lipoprotein
lplB	781	transmembrane lipoprotein
lplC	782	transmembrane lipoprotein
mdr	334	multidrug-efflux transporter (puromycin, nerfloxacin, tosufloxacin)
msmE	3097	multiple sugar-binding protein
msmX	3984	multiple sugar-binding transport ATP-binding protein
mtlA	449	phosphotransferase system (PTS) mannitol-specific enzyme IIABC component
narK	3833	nitrite extrusion protein
nasA	363	nitrate transporter
natA	296	Na^+ ABC transporter (extrusion) (ATP-binding protein)
natB	297	Na^+ ABC transporter (extrusion) (membrane protein)
nrgA	3756	ammonium transporter
nupC	4050	pyrimidine-nucleoside transport protein
oppA	1219	oligopeptide ABC transporter (binding protein) (initiation of sporulation, competence development)
oppB	1221	oligopeptide ABC transporter (permease) (initiation of sporulation, competence development)
oppC	1222	oligopeptide ABC transporter (permease) (initiation of sporulation, competence development)
oppD	1223	oligopeptide ABC transporter (ATP-binding protein) (initiation of sporulation, competence development)
oppF	1224	oligopeptide ABC transporter (ATP-binding protein) (initiation of sporulation, competence development)
opuAA	321	glycine betaine ABC transporter (ATP-binding protein) (osmoprotection)
opuAB	322	glycine betaine ABC transporter (permease) (osmoprotection)
opuAC	323	glycine betaine ABC transporter (glycine betaine-binding protein) (osmoprotection)
opuBA	3462	choline ABC transporter (ATP-binding protein) (osmoprotection)
opuBB	3461	choline ABC transporter (membrane protein) (osmoprotection)
opuBC	3460	choline ABC transporter (choline-binding protein) (osmoprotection)
opuBD	3460	choline ABC transporter (membrane protein) (osmoprotection)
opuCA	3470	glycine betaine/carnitine/choline ABC transporter (ATP-binding protein) (osmoprotection)
opuCB	3469	glycine betaine/carnitine/choline ABC transporter (membrane protein) (osmoprotection)
opuCC	3468	glycine betaine/carnitine/choline ABC transporter (osmoprotectant-binding protein) (osmoprotection)
opuCD	3467	glycine betaine/carnitine/choline ABC transporter (membrane protein) (osmoprotection)
opuD	3076	glycine betaine transporter (osmoprotection)
opuE	728	proline transporter (osmoprotection)
pbuX	2319	xanthine permease
ptsG	1457	phosphotransferase system (PTS) glucose-specific enzyme IIABC component
ptsI	1459	phosphotransferase system (PTS) enzyme I (general energy coupling protein of the PTS)
pyrP	1618	uracil permease (pyrimidine biosynthesis)
rbsA	3703	ribose ABC transporter (ATP-binding protein)
rbsB	3705	ribose ABC transporter (ribose-binding protein)
rbsC	3704	ribose ABC transporter (permease)
rbsD	3702	ribose ABC transporter (membrane protein)
rocC	3876	amino acid permease (arginine and ornithine utilization)
rocE	4143	amino acid permease (arginine and ornithine utilization)
sacP	3904	phosphotransferase system (PTS) sucrose-specific enzyme IIBC component
slp	1533	small peptidoglycan-associated lipoprotein
sunT	2269	sublancin 168 lantibiotic transporter
tetB	4188	tetracycline resistance protein
treP	850	phosphotransferase system (PTS) trehalose-specific enzyme IIBC component
trkA	2723	potassium uptake
yabM	65	amino acid transporter
ybaE	151	ABC transporter (ATP-binding protein)
ybbF	191	sucrose phosphotransferase enzyme II
ybcL	212	chloramphenicol resistance protein
ybdA	217	ABC transporter (binding protein)
ybdB	218	ABC transporter (permease)
ybeC	231	amino acid transporter
ybfS	257	phosphotransferase system enzyme II
ybgF	262	histidine permease
ybgH	264	sodium/proton-dependent alanine transporter
ybxA	150	ABC transporter (ATP-binding protein)
ybxG	227	amino acid permease
ycbE	270	glucarate transporter
ycbK	277	efflux system
ycbN	280	ABC transporter (ATP-binding protein)
yccK	298	ion channel
ycdI	309	ABC transporter (ATP-binding protein)
yceI	317	transporter
yceJ	320	multidrug-efflux transporter
ycgH	337	amino acid transporter
ycgO	347	proline permease
yckA	368	amino acid ABC transporter (permease)
yckB	368	amino acid ABC transporter (binding protein)
yckI	410	glutamine ABC transporter (ATP-binding protein)
yckJ	410	glutamine ABC transporter (permease)
yckK	411	glutamine ABC transporter (glutamine-binding protein)
yclF	417	di-tripeptide ABC transporter (membrane protein)
yclH	424	ABC transporter (permease)
yclI	426	transporter
yclN	432	ferrichrome ABC transporter (permease)
yclO	433	ferrichrome ABC transporter (permease)
yclP	434	ferrichrome ABC transporter (ATP-binding protein)
yclQ	435	ferrichrome ABC transporter (binding protein)
ycnB	437	multidrug resistance protein
ycnJ	448	copper export protein
ycsG	457	branched chain amino acids transporter
ydbA	493	ABC transporter (binding protein)
ydbE	497	C4-dicarboxylate binding protein
ydbH	500	C4-dicarboxylate transport protein
ydbJ	502	ABC transporter (ATP-binding protein)
ydeG	566	metabolite transport protein

ydeR	578	antibiotic resistance protein
ydfA	580	arsenical pump membrane protein
ydfI	589	antibiotic transport-associated protein
ydfL	595	multidrug-efflux transporter regulator
ydfM	596	cation efflux system
ydfO	597	ABC transporter (binding protein)
ydgF	608	amino acid ABC transporter (permease)
ydgH	609	transporter
ydgK	613	bicyclomycin resistance protein
ydhL	626	chloramphenicol resistance protein
ydhM	626	cellobiose phosphotransferase system enzyme II
ydhN	627	cellobiose phosphotransferase system enzyme II
ydhO	627	cellobiose phosphotransferase system enzyme II
ydiF	646	ABC transporter (ATP-binding protein)
ydjD	668	H^+-symporter
ydjK	676	sugar transporter
yeaB	687	cation efflux system membrane protein
yecA	712	amino acid permease
yesO	761	sugar-binding protein
yesP	762	lactose permease
yesQ	763	lactose permease
yfhA	921	iron(III) dicitrate transport permease
yfhI	926	antibiotic resistance protein
yfiB	893	ABC transporter (ATP-binding protein)
yfiC	895	ABC transporter (ATP-binding protein)
yfiG	900	metabolite transport protein
yfiL	905	ABC transporter (ATP-binding protein)
yfiM	906	ABC transporter (ATP-binding protein)
yfiN	907	ABC transporter (ATP-binding protein)
yfiS	913	multidrug resistance protein
yfiU	916	multidrug-efflux transporter
yfiY	920	iron(III) dicitrate transport permease
yfiZ	920	iron(III) dicitrate transport permease
yfjQ	872	divalent cation transport protein
yfkE	865	H^+/Ca^{2+} exchanger
yfkF	865	multidrug-efflux transporter
yfkH	862	transporter
yfkL	861	multidrug resistance protein
yflA	844	aminoacid carrier protein
yflE	844	anion-binding protein
yflF	840	phosphotransferase system enzyme II
yflS	829	2-oxoglutarate/malate translocator
yfmC	826	ferrichrome ABC transporter (binding protein)
yfmD	825	ferrichrome ABC transporter (permease)
yfmE	824	ferrichrome ABC transporter (permease)
yfmF	823	ferrichrome ABC transporter (ATP-binding protein)
yfmM	815	ABC transporter (ATP-binding protein)
yfmO	812	multidrug-efflux transporter
yfmR	809	ABC transporter (ATP-binding protein)
yfnA	806	metabolite transporter
ygaD	939	ABC transporter (ATP-binding protein)
ygaL	961	nitrate ABC transporter (binding protein)
ygaM	963	ABC transporter (permease)
ygbA	962	ABC transporter (binding lipoprotein)
yhaQ	1062	ABC transporter (ATP-binding protein)
yhaU	1060	Na^+/H^+ antiporter
yhcA	977	multidrug resistance protein
yhcG	981	glycine betaine/L-proline transport
yhcH	982	ABC transporter (ATP-binding protein)
yhcJ	984	ABC transporter (binding lipoprotein)
yhcL	986	sodium-glutamate symporter
yhdG	1023	amino acid transporter
yhdH	1024	sodium-dependent transporter
yheH	1047	ABC transporter (ATP-binding protein)
yheI	1045	ABC transporter (ATP-binding protein)
yheL	1044	Na^+/H^+ antiporter
yhfQ	1107	iron(III) dicitrate-binding protein
yhjB	1120	metabolite permease
yhjO	1133	multidrug-efflux transporter
yhjP	1133	transporter binding protein
yitG	1177	multidrug resistance protein
yitZ	1194	multidrug resistance protein
yjbQ	1240	Na^+/H^+ antiporter
yjdD	1272	fructose phosphotransferase system enzyme II
yjkB	1296	amino acid ABC transporter (ATP-binding protein)
yjmB	1301	Na^+:galactoside symporter
yjmG	1307	hexuronate transporter
ykaB	1350	low-affinity inorganic phosphate transporter
ykbA	1352	amino acid permease
ykcA	1353	ABC transporter (binding protein)
yktD	1368	oligopeptide ABC transporter (permease)
yknU	1499	ABC transporter (ATP-binding protein)
yknV	1501	ABC transporter (ATP-binding protein)
yknY	1505	ABC transporter (ATP-binding protein)
ykoD	1390	cation ABC transporter (ATP-binding protein)
ykoK	1395	Mg^{2+} transporter
ykpA	1512	ABC transporter (ATP-binding protein)
ykrM	1416	Na^+-transporting ATP synthase
ykuC	1476	macrolide-efflux protein
ykvW	1451	heavy metal-transporting ATPase
ylmA	1606	ABC transporter (ATP-binding protein)
ylnA	1630	anion permease
yloB	1637	calcium-transporting ATPase
ynaJ	1887	H^+-symporter
yncC	1896	metabolite transport protein
yocN	2098	permease
yocR	2106	sodium-dependent transporter
yocS	2106	sodium-dependent transporter
yodE	2129	aromatic metabolite transporter
yodF	2130	proline permease
yojA	2125	gluconate permease
ypqE	2337	phosphotransferase system enzyme II
yqeW	2620	Na^+/P_i cotransporter
yqgG	2581	phosphate ABC transporter (binding protein)
yqgH	2580	phosphate ABC transporter (permease)
yqgI	2579	phosphate ABC transporter (permease)
yqgJ	2578	phosphate ABC transporter (ATP-binding protein)
yqgK	2577	phosphate ABC transporter (ATP-binding protein)
yqiH	2515	lipoprotein
yqiX	2492	amino acid ABC transporter (binding protein)
yqiY	2491	amino acid ABC transporter (permease)
yqiZ	2491	amino acid ABC transporter (ATP-binding protein)
yqjV	2466	multidrug resistance protein
yqkI	2453	Na^+/H^+ antiporter
yraO	2745	citrate transporter
yrbD	2841	sodium/proton-dependent alanine carrier protein
ytbD	2968	antibiotic resistance protein
ytcP	3087	ABC transporter (permease)
ytcQ	3086	lipoprotein
yteQ	3082	sugar transport protein
ytgA	3145	ABC transporter (membrane protein)
ytgB	3144	ABC transporter (ATP-binding protein)
ytgC	3143	ABC transporter (membrane protein)
ythP	3071	ABC transporter (ATP-binding protein)
ytlC	3132	anion transport ABC transporter (ATP-binding protein)
ytlD	3133	ABC transporter (permease)
ytlP	3065	ABC transporter (permease)
ytmJ	3007	amino acid ABC transporter (binding protein)
ytmK	3006	amino acid ABC transporter (binding protein)
ytmL	3006	amino acid ABC transporter (permease)
ytmM	3005	amino acid ABC transporter (permease)
ytnA	3125	proline permease
ytrB	3118	ABC transporter (ATP-binding protein)
ytrE	3115	ABC transporter (ATP-binding protein)
ytsC	3111	ABC transporter (ATP-binding protein)
ytsD	3110	ABC transporter (permease)
yttB	3108	multidrug resistance protein
yubD	3192	multidrug resistance protein
yubG	3188	Na^+-transporting ATP synthase
yufN	3239	ABC transporter (lipoprotein)
yufO	3240	ABC transporter (ATP-binding protein)
yufR	3244	organic acid transport protein
yufU	3248	Na^+/H^+ antiporter
yufV	3249	Na^+/H^+ antiporter
yugO	3218	potassium channel protein
yunJ	3330	purine permease
yunK	3331	purine permease
yurJ	3345	multiple sugar ABC transporter (ATP-binding protein)
yurM	3348	sugar permease
yurN	3349	sugar permease
yurO	3350	multiple sugar-binding protein
yurY	3360	ABC transporter (ATP-binding protein)
yusC	3363	ABC transporter (ATP-binding protein)
yusP	3374	multidrug-efflux transporter
yusV	3379	iron(III) dicitrate transport permease
yutK	3307	Na^+/nucleoside cotransporter
yuxJ	3232	multidrug-efflux transporter
yvaE	3448	multidrug-efflux transporter
yvbW	3490	amino acid permease
yvcC	3579	ABC transporter (ATP-binding protein)
yvcR	3565	ABC transporter (ATP-binding protein)
yvcS	3565	ABC transporter (permease)
yvdB	3561	transporter
yvdG	3555	maltose/maltodextrin-binding protein
yvdH	3554	maltodextrin transport system permease
yvdI	3552	maltodextrin transport system permease
yveA	3538	permease
yvfH	3510	L-lactate permease
yvfK	3508	maltose/maltodextrin-binding protein
yvfL	3506	maltodextrin transport system permease
yvfM	3505	maltodextrin transport system permease
yvfR	3498	ABC transporter (ATP-binding protein)
yvgK	3424	molybdenum-binding protein
yvgL	3424	molybdate-binding protein
yvgM	3425	molybdenum transport permease
yvgW	3440	heavy metal-transporting ATPase
yvgX	3443	heavy metal-transporting ATPase
yvgY	3443	mercuric transport protein
yvkA	3618	multidrug-efflux transporter
yvmA	3605	transporter
yvqJ	3399	macrolide-efflux protein
yvrA	3402	iron transport system
yvrB	3403	iron permease
yvrC	3403	iron-binding protein
yvrO	3413	amino acid ABC transporter (ATP-binding protein)
yvsH	3420	ABC transporter (amino acid permease)
ywbA	3938	phosphotransferase system enzyme II
ywbF	3933	sugar permease
ywcA	3923	Na^+-dependent symport
ywcJ	3904	nitrite transporter
ywfA	3874	chloramphenicol resistance
ywfF	3869	efflux protein
ywhQ	3837	ABC transporter (ATP-binding protein)
ywjA	3821	ABC transporter (ATP-binding protein)
ywoA	3758	bacteriocin transport permease
ywoD	3754	transporter
ywoE	3753	permease
ywoG	3749	antibiotic resistance protein
ywpC	3743	large conductance mechanosensitive channel protein
ywrA	3721	chromate transport protein
ywrB	3720	chromate transport protein
ywrK	3712	arsenical pump membrane protein
ywtG	3693	metabolite transport protein
yxaM	4100	antibiotic resistance protein
yxcC	4087	metabolite transport protein
yxdL	4070	ABC transporter (ATP-binding protein)
yxdM	4069	ABC transporter (permease)
yxeB	4066	ABC transporter (binding protein)
yxeM	4059	amino acid ABC transporter (binding protein)
yxeN	4058	amino acid ABC transporter (permease)
yxeO	4058	amino acid ABC transporter (ATP-binding protein)
yxeR	4054	ethanolamine transporter
yxiQ	4009	Mg^{2+}/citrate complex transporter
yxjA	4005	pyrimidine nucleoside transport
yxkJ	3979	metabolite-sodium symport
yxlA	3970	purine-cytosine permease
yxlF	3968	ABC transporter (ATP-binding protein)
yxlH	3966	multidrug-efflux transporter
yyaJ	4194	transporter
yybF	4180	antibiotic resistance protein
yybJ	4175	ABC transporter (ATP-binding protein)
yybL	4174	ABC transporter (permease)
yybO	4169	ABC transporter (permease)
yycB	4159	ABC transporter (permease)
yydI	4125	ABC transporter (ATP-binding protein)
yyzE	4122	phosphotransferase systeme enzyme II

I.3 SENSORS (SIGNAL TRANSDUCTION) 38

cheA	1712	two-component sensor histidine kinase [CheB/CheY] chemotactic signal modulator
citS	830	two-component sensor histidine kinase [CitT]
comP	3255	two-component sensor histidine kinase [ComA] involved in early competence
degS	3646	two-component sensor histidine kinase [DegU] involved in degradative enzyme and competence regulation
kinA	1469	two-component sensor histidine kinase [Spo0F] involved in the initiation of sporulation
kinB	3229	two-component sensor histidine kinase [Spo0F] involved in the initiation of sporulation
kinC	1518	two-component sensor histidine kinase [Spo0A] involved in the initiation of sporulation (phospho-relay-independent)
lytS	2957	two-component sensor histidine kinase [LytT] involved in the rate of autolysis
phoR	2977	two-component sensor histidine kinase [PhoP] involved in phosphate regulation
resE	2416	two-component sensor histidine kinase [ResD] involved in aerobic and anaerobic respiration
ybdK	222	two-component sensor histidine kinase [YbdJ]
ycbA	266	two-component sensor histidine kinase [YcbB]
ycbM	279	two-component sensor histidine kinase [YcbL]
yccG	295	two-component sensor histidine kinase [YccH]
yclK	427	two-component sensor histidine kinase [YclJ]
ydbF	497	two-component sensor histidine kinase [YdbG]
ydfH	587	two-component sensor histidine kinase [YdfI]
yesM	758	two-component sensor histidine kinase [YesN]
yfiJ	903	two-component sensor histidine kinase [YfiK]
yhcY	1008	two-component sensor histidine kinase [YhcZ]
yhjL	1129	sensory transduction pleiotropic regulatory protein
ykoH	1392	two-component sensor histidine kinase [YkoG]
ykrQ	1419	two-component sensor histidine kinase
ykvD	1432	two-component sensor histidine kinase
yocF	2090	two-component sensor histidine kinase [YocG]
yrkQ	2704	two-component sensor histidine kinase [YrkP]
ytrP	3035	two-component sensor histidine kinase
ytsB	3112	two-component sensor histidine kinase [YtsA]
yufL	3236	two-component sensor histidine kinase [YufM]
yvcQ	3566	two-component sensor histidine kinase [YvcP]
yvfT	3497	two-component sensor histidine kinase [YvfU]
yvqB	3385	two-component sensor histidine kinase [YvqA]
yvqE	3395	two-component sensor histidine kinase [YvqC]
yvrG	3407	two-component sensor histidine kinase [YvrH]
ywpD	3741	two-component sensor histidine kinase
yxdK	4071	two-component sensor histidine kinase [YxdJ]
yxjM	3992	two-component sensor histidine kinase [YxjL]
yycG	4153	two-component sensor histidine kinase [YycF]

I.4 MEMBRANE BIOENERGETICS (ELECTRON TRANSPORT CHAIN AND ATP SYNTHASE) 78

atpA	3784	ATP synthase (subunit α)
atpB	3787	ATP synthase (subunit a)
atpC	3781	ATP synthase (subunit ε)
atpD	3782	ATP synthase (subunit β)
atpE	3786	ATP synthase (subunit c)
atpF	3786	ATP synthase (subunit b)
atpG	3783	ATP synthase (subunit γ)
atpH	3785	ATP synthase (subunit δ)
atpI	3787	ATP synthase (subunit i)
cccA	2599	cytochrome c_{550}
cccB	3625	cytochrome c_{551}
ccdA	1922	required for a late step of cytochrome *c* synthesis
ctaA	1558	cytochrome caa_3 oxidase (required for biosynthesis)
ctaB	1559	cytochrome caa_3 oxidase (assembly factor)
ctaC	1560	cytochrome caa_3 oxidase (subunit II)
ctaD	1561	cytochrome caa_3 oxidase (subunit I)
ctaE	1563	cytochrome caa_3 oxidase (subunit III)
ctaF	1563	cytochrome caa_3 oxidase (subunit IV)
cydA	3978	cytochrome *bd* ubiquinol oxidase (subunit I)
cydB	3977	cytochrome *bd* ubiquinol oxidase (subunit II)
etfA	2915	electron transfer flavoprotein (α subunit)
etfB	2916	electron transfer flavoprotein (β subunit)
fer	2409	ferredoxin
hmp	1372	flavohemoglobin
narG	3829	nitrate reductase (α subunit)
narH	3825	nitrate reductase (β subunit)
narI	3823	nitrate reductase (γ subunit)
narJ	3824	nitrate reductase (protein J)
ndhF	205	NADH dehydrogenase (subunit 5)
qcrA	2364	menaquinol:cytochrome *c* oxidoreductase (iron-sulphur subunit)
qcrB	2364	menaquinol:cytochrome *c* oxidoreductase (cytochrome *b* subunit)
qcrC	2363	menaquinol:cytochrome *c* oxidoreductase (cytochrome *b*/*c* subunit)
qoxA	3917	cytochrome aa_3 quinol oxidase (subunit II)
qoxB	3916	cytochrome aa_3 quinol oxidase (subunit I)
qoxC	3914	cytochrome aa_3 quinol oxidase (subunit III)
qoxD	3913	cytochrome aa_3 quinol oxidase (subunit IV)
resA	2421	essential protein similar to cytochrome *c* biogenesis protein
resB	2420	essential protein similar to cytochrome *c* biogenesis protein
resC	2418	essential protein similar to cytochrome *c* biogenesis protein
tlp	1930	thioredoxin-like protein
trxA	2912	thioredoxin
trxB	3573	thioredoxin reductase
ycgT	352	thioredoxin reductase
ycnD	439	NADPH-flavin oxidoreductase
ydbP	508	thioredoxin
ydeQ	576	NAD(P)H oxidoreductase
ydfQ	598	thioredoxin
ydgI	613	NADH dehydrogenase
yfkO	854	NAD(P)H-flavin oxidoreductase
yfmJ	818	quinone oxidoreductase
yjdK	1280	cytochrome *c* oxidase assembly factor
yjlD	1299	NADH dehydrogenase
ykuN	1486	flavodoxin
ykuP	1488	sulfite reductase
ykuU	1492	2-cys peroxiredoxin
ykvV	1450	thioredoxin
yneN	1929	thiol:disulfide interchange protein
yojN	2114	nitric-oxide reductase
yolI	2267	thioredoxin
yosR	2159	thioredoxin
ypdA	2401	thioredoxin reductase
yqiG	2516	NADH-dependent flavin oxidoreductase
yqjM	2475	NADH-dependent flavin oxidoreductase
yrkL	2708	NAD(P)H oxidoreductase
ythA	3139	cytochrome *d* oxidase subunit
ytpP	3054	thioredoxin H1
ytrC	3117	cytochrome *c* oxidase subunit
ytrD	3116	cytochrome *c* oxidase subunit
yufD	3249	NADH dehydrogenase (ubiquinone)
yufT	3246	NADH dehydrogenase
yumB	3300	NADH dehydrogenase
yumC	3301	thioredoxin reductase
yusE	3364	thioredoxin
yutJ	3308	NADH dehydrogenase
yvaB	3445	NAD(P)H dehydrogenase (quinone)
ywcG	3911	NADPH-flavin oxidoreductase
ywhN	3840	ubiquinol-cytochrome *c* reductase
ywrO	3708	NAD(P)H oxidoreductase

I.5 MOBILITY AND CHEMOTAXIS 55

cheC	1715	inhibition of CheR-mediated methylation of methyl-accepting chemotaxis proteins
cheD	1715	required for methylation of methyl-accepting chemotaxis proteins by CheR
cheR	2380	methyl-accepting chemotaxis proteins methyltransferase
cheV	1473	modulation of CheA activity in response to attractants (CheW and CheY similar domains)
cheW	1714	modulation of CheA activity in response to attractants
flgB	1691	flagellar basal-body rod protein
flgC	1691	flagellar basal-body rod protein

flgE	1700	flagellar hook protein
flgK	3639	flagellar hook-associated protein 1 (HAP1)
flgL	3637	flagellar hook-associated protein 3 (HAP3)
flgM	3640	flagellin synthesis regulatory protein (anti-sigma factor [σ^{D}])
flhA	1707	flagella-associated protein
flhB	1706	flagella-associated protein
flhF	1709	flagella-associated protein
flhO	3746	flagellar basal-body rod protein
flhP	3745	flagellar hook-basal body protein
fliD	3633	flagellar hook-associated protein 2 (HAP2)
fliE	1692	flagellar hook-basal body protein
fliF	1692	flagellar basal-body M-ring protein
fliG	1694	flagellar motor switch protein
fliH	1695	flagellar assembly protein
fliI	1695	flagellar-specific ATP synthase
fliJ	1697	flagellar protein required for formation of basal body
fliK	1698	flagellar hook-length control
fliL	1701	flagellar protein required for flagellar formation
fliM	1701	flagellar motor switch protein
fliP	1704	flagellar protein required for flagellar formation
fliQ	1705	flagellar protein required for flagellar formation
fliR	1705	flagellar protein required for flagellar formation
fliS	3632	flagellar protein
fliT	3632	flagellar protein
fliY	1702	flagellar motor switch protein
fliZ	1704	flagellar protein required for flagellar formation
hag	3635	flagellin protein
mcpA	3207	methyl-accepting chemotaxis protein (glucose and α-methyl-glucoside)
mcpB	3212	methyl-accepting chemotaxis protein (asparagine, glutamine and histidine)
mcpC	1463	methyl-accepting chemotaxis protein (cysteine, proline, threonine, glycine, serine, lysine, valine and arginine)
motA	1435	motility protein (flagellar motor rotation)
motB	1434	motility protein (flagellar motor rotation)
tlpA	3209	methyl-accepting chemotaxis protein
tlpB	3205	methyl-accepting chemotaxis protein
tlpC	374	methyl-accepting chemotaxis protein
yfmS	808	methyl-accepting chemotaxis protein
yhfV	1113	methyl-accepting chemotaxis protein
ylqH	1679	flagellar biosynthetic protein
ylxG	1699	flagellar hook assembly protein
ylxH	1710	flagellar biosynthesis switch protein
yoaH	2030	methyl-accepting chemotaxis protein
ytxD	3043	flagellar motor apparatus
ytxE	3042	motility protein
yvaQ	3457	transmembrane receptor taxis protein
yvyC	3634	flagellar protein
yvyF	3640	flagellar protein
yvyG	3639	flagellar protein
yvzB	3609	flagellin

I.6 PROTEIN SECRETION 18

csaA	2079	chaperonin involved in protein secretion
ffh	1672	signal recognition particle
ftsY	1670	signal recognition particle
lsp	1616	signal peptidase II
lytA	3662	secretion of major autolysin LytC
prsA	1071	protein secretion (post-translocation chaperonin)
secA	3630	preprotein translocase subunit
secE	118	preprotein translocase subunit
secF	2828	protein-export membrane protein (product also similar to SecD of *E. coli*)
secY	145	preprotein translocase subunit
sipS	2432	signal peptidase I
sipT	1511	signal peptidase I
sipU	454	signal peptidase I
sipV	1122	signal peptidase I
sipW	2554	signal peptidase I
yaaT	42	signal peptidase II
yacD	81	protein secretion PrsA homologue
yobE	2057	general secretion pathway protein

I.7 CELL DIVISION 21

divIB	1593	cell-division initiation protein (septum formation)
divIC	69	cell-division initiation protein (septum formation)
divIVA	1612	cell-division initiation protein (septum placement)
ftsA	1596	cell-division protein (septum formation)
ftsE	3625	cell-division ATP-binding protein
ftsH	77	cell-division protein / general stress protein (class III heat-shock)
ftsL	1581	cell-division protein (septum formation)
ftsX	3624	cell-division protein
ftsZ	1597	cell-division initiation protein (septum formation)
gid	1685	glucose-inhibited division protein
gidA	4211	glucose-inhibited division protein
gidB	4209	glucose-inhibited division protein
maf	2862	septum formation
minC	2859	cell-division inhibitor (septum placement)
minD	2858	cell-division inhibitor (septum placement) (ATPase activator of MinC)
yacA	75	cell-cycle protein
yfhF	925	cell-division inhibitor
yjoB	1314	cell-division protein FtsH homologue
ylaO	1552	cell-division protein
ylmH	1611	cell-division protein
ywcF	3912	cell-division protein

I.8 SPORULATION 139

bofA	30	inhibitor of the pro-σ^{K} processing machinery
bofC	2837	forespore regulator of the σ^{K} checkpoint
cgeA	2148	maturation of the outermost layer of the spore
cgeB	2148	maturation of the outermost layer of the spore
cgeC	2148	maturation of the outermost layer of the spore
cgeD	2147	maturation of the outermost layer of the spore
cgeE	2146	maturation of the outermost layer of the spore
cotA	685	spore coat protein (outer)
cotB	3715	spore coat protein (outer)
cotC	1905	spore coat protein (outer)
cotD	2332	spore coat protein (inner)
cotE	1774	spore coat protein (outer)
cotF	4166	spore coat protein
cotG	3716	spore coat protein
cotH	3716	spore coat protein (inner)
cotJA	755	polypeptide composition of the spore coat
cotJB	756	polypeptide composition of the spore coat
cotJC	756	polypeptide composition of the spore coat
cotK	1926	spore coat protein
cotL	1926	spore coat protein
cotM	1925	spore coat protein (outer)
cotN	2553	spore coat-associated protein
cotS	3160	spore coat protein
cotT	1280	spore coat protein (inner)
cotV	1251	spore coat protein (insoluble fraction)
cotW	1251	spore coat protein (insoluble fraction)
cotX	1251	spore coat protein (insoluble fraction)
cotY	1250	spore coat protein (insoluble fraction)
cotZ	1249	spore coat protein (insoluble fraction)
csgA	228	sporulation-specific SASP protein
jag	4213	SpoIIIJ-associated protein
kapB	3230	activator of KinB in the initiation of sporulation
kapD	3232	inhibitor of the KinA pathway to sporulation
kbaA	159	activation of the KinB signaling pathway to sporulation
obg	2853	GTP-binding protein involved in initiation of sporulation (Spo0A activation)
phrA	1316	phosphatase (RapA) inhibitor (imported by Opp)
phrC	430	phosphatase (RapC) regulator / competence and sporulation stimulating factor (CSF)
phrE	2660	phosphatase (RapE) regulator
phrF	3846	phosphatase (RapF) regulator
phrG	4141	phosphatase (RapG) regulator
phrI	548	phosphatase (RapI) regulator
phrK	2063	phosphatase (RapK) regulator
rapA	1315	response regulator aspartate phosphatase [Spo0F~P]
rapB	3771	response regulator aspartate phosphatase [Spo0F~P]
rapC	428	response regulator aspartate phosphatase
rapD	3743	response regulator aspartate phosphatase
rapE	2658	response regulator aspartate phosphatase
rapF	3845	response regulator aspartate phosphatase
rapG	4139	response regulator aspartate phosphatase
rapH	750	response regulator aspartate phosphatase
rapI	547	response regulator aspartate phosphatase
rapJ	304	response regulator aspartate phosphatase
rapK	2061	response regulator aspartate phosphatase
sinI	2552	antagonist of SinR
soj	4206	centromere-like function involved in forespore chromosome partitioning / inhibition of Spo0A activation
splB	1461	spore photoproduct lyase
spmA	2423	spore maturation protein (spore core dehydratation)
spmB	2422	spore maturation protein (spore core dehydratation)
spo0B	2854	sporulation initiation phosphoprotein (part of phosphorelay: Spo0F~P->Spo0B~P->Spo0A~P)
spo0E	1430	negative sporulation regulatory phosphatase [Spo0A~P]
spo0J	4206	chromosome positioning near the pole and transport through the polar septum / antagonist of Soj
spoIIAA	2444	anti-anti-sigma factor [SpoIIAB]
spoIIAB	2444	anti-sigma factor [σ^{F}(SpoIIAC)] and serine kinase [SpoIIAA]
spoIIB	2864	endospore development (oligosporogenous mutation)
spoIID	3777	required for complete dissolution of the asymmetric septum
spoIIE	71	serine phosphatase [SpoIIAA~P] (σ^{F} activation) / asymmetric septum formation
spoIIGA	1603	protease (processing of pro-σ^{E} to active σ^{E})
spoIIIAA	2537	mutants block sporulation after engulfment
spoIIIAB	2536	mutants block sporulation after engulfment
spoIIIAC	2535	mutants block sporulation after engulfment
spoIIIAD	2535	mutants block sporulation after engulfment
spoIIIAE	2535	mutants block sporulation after engulfment
spoIIIAF	2534	mutants block sporulation after engulfment
spoIIIAG	2533	mutants block sporulation after engulfment
spoIIIAH	2532	mutants block sporulation after engulfment
spoIIIE	1752	DNA translocase required for chromosome partitioning through the septum into the forespore
spoIIIJ	4214	essential for σ^{G} activity at stage III
spoIIM	2450	required for dissolution of the septal cell wall
spoIIP	2634	required for dissolution of the septal cell wall
spoIIQ	3760	required for completion of engulfment
spoIIR	3794	required for processing of pro-σ^{E}
spoIISA	1349	lethal when synthesized during vegetative growth in the absence of SpoIISB
spoIISB	1348	disruption blocks sporulation after septum formation
spoIVA	2387	required for proper spore cortex formation and coat assembly
spoIVB	2520	intercompartmental signalling of pro-σ^{K} processing/activation in the mother-cell
spoIVCA	2654	site-specific DNA recombinase required for creating the *sigK* gene (excision of the *skin* element)
spoIVFA	2857	inhibitor of SpoIVFB
spoIVFB	2856	protease (processing of pro-σ^{K} to active σ^{K})
spoVAA	2443	mutants lead to the production of immature spores
spoVAB	2442	mutants lead to the production of immature spores
spoVAC	2441	mutants lead to the production of immature spores
spoVAD	2441	mutants lead to the production of immature spores
spoVAE	2440	mutants lead to the production of immature spores
spoVAF	2439	mutants lead to the production of immature spores
spoVB	2829	involved in spore cortex synthesis
spoVC	60	thermosensitive mutant blocks spore coat formation
spoVE	1590	required for spore cortex synthesis
spoVFA	1744	dipicolinate synthase subunit A
spoVFB	1745	dipicolinate synthase subunit B
spoVG	56	required for spore cortex synthesis
spoVID	2872	required for assembly of the spore coat
spoVK	1873	disruption leads to the production of immature spores
spoVM	1655	required for normal spore cortex and coat synthesis
spoVR	1015	involved in spore cortex synthesis
spoVS	1769	required for dehydratation of the spore core and assembly of the coat
spsA	3892	spore coat polysaccharide synthesis
spsB	3891	spore coat polysaccharide synthesis
spsC	3890	spore coat polysaccharide synthesis
spsD	3889	spore coat polysaccharide synthesis
spsE	3888	spore coat polysaccharide synthesis
spsF	3887	spore coat polysaccharide synthesis
spsG	3886	spore coat polysaccharide synthesis
spsI	3885	spore coat polysaccharide synthesis
spsJ	3884	spore coat polysaccharide synthesis
spsK	3883	spore coat polysaccharide synthesis
sspA	3025	small acid-soluble spore protein (major α-type SASP)
sspB	1050	small acid-soluble spore protein (major β-type SASP)
sspC	2155	small acid-soluble spore protein (minor α/β-type SASP)
sspD	1413	small acid-soluble spore protein (minor α/β-type SASP)
sspE	937	small acid-soluble spore protein (major γ-type SASP)
sspF	53	small acid-soluble spore protein (minor α/β-type SASP)
usd	3748	required for translation of *spoIIID*
yknT	1495	sporulation protein σ^{E}-controlled
ykvU	1449	spore cortex membrane protein
ynzH	1901	spore coat protein
yobW	2083	membrane protein σ^{K}-controlled
yqgT	2568	γ-D-glutamyl-L-diamino acid endopeptidase I
yqjG	2483	lipoprotein SpoIIIJ-like
yraD	2754	spore coat protein
yraE	2754	spore coat protein
yraF	2752	spore coat protein
yraG	2752	spore coat protein
yrbA	2845	spore coat protein
yrbB	2844	spore coat protein
yrbC	2843	spore coat protein
ytaA	3161	spore coat protein
ytgP	3074	spore cortex protein
ytpT	3051	DNA translocase stage III sporulation protein
yyaA	4208	DNA-binding protein Spo0J-like

I.9 GERMINATION 23

gerAA	3390	germination response to L-alanine
gerAB	3391	germination response to L-alanine
gerAC	3392	germination response to L-alanine
gerBA	3688	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerBB	3689	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerBC	3690	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerCA	2384	heptaprenyl diphosphate synthase component I (menaquinone biosynthesis)
gerCB	2383	menaquinone biosynthesis methyltransferase (menaquinone biosynthesis)
gerCC	2382	heptaprenyl diphosphate synthase component II (menaquinone biosynthesis)
gerD	159	germination response to L-alanine and to the combination of glucose, fructose, L-asparagine, and KCl
gerKA	420	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerKB	423	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerKC	421	germination response to the combination of glucose, fructose, L-asparagine, and KCl
gerM	2902	germination (cortex hydrolysis) and sporulation (stage II, multiple polar septa)
gpr	2635	spore protease (degradation of SASPs)
sleB	2399	spore cortex-lytic enzyme
yfkQ	850	spore germination response
yfkR	848	spore germination protein
yfkT	847	spore germination protein
ykvT	1448	spore cortex-lytic enzyme
yndD	1907	spore germination protein
yndE	1908	spore germination protein
yndF	1909	spore germination protein

I.10 TRANSFORMATION/COMPETENCE 20

cinA	1763	competence-damage inducible protein
comC	2864	late competence protein required for processing and translocation of ComGC
comEA	2640	late competence operon required for DNA binding and uptake
comEB	2640	late competence operon required for DNA binding and uptake
comEC	2639	late competence operon required for DNA binding and uptake
comER	2640	non-essential gene for competence
comFA	3643	late competence protein required for DNA uptake
comFB	3641	late competence gene
comFC	3641	late competence gene
comGA	2559	late competence gene
comGB	2558	DNA transport machinery
comGC	2557	exogenous DNA-binding
comGD	2557	DNA transport machinery
comGE	2557	DNA transport machinery
comGF	2556	DNA transport machinery
comGG	2556	DNA transport machinery
comS	390	assembly link between regulatory components of the competence signal transduction pathway
comX	3255	competence pheromone precursor (activation of ComA)
mecA	1229	negative regulator of competence
ypbH	2403	negative regulation of competence MecA homologue

II INTERMEDIARY METABOLISM 742

II.1 METABOLISM OF CARBOHYDRATES AND RELATED MOLECULES 261

II.1.1 SPECIFIC PATHWAYS 214

abfA	2939	α-L-arabinofuranosidase
abnA	2949	arabinan-endo 1,5—L-arabinase (degradation of plant cell wall polysaccharide)
ackA	3015	acetate kinase
acoA	879	acetoin dehydrogenase E1 component (TPP-dependent α subunit)
acoB	880	acetoin dehydrogenase E1 component (TPP-dependent β subunit)
acoC	881	acetoin dehydrogenase E2 component (dihydrolipoamide acetyltransferase)
acoL	882	acetoin dehydrogenase E3 component (dihydrolipoamide dehydrogenase)
acsA	3039	acetyl-CoA synthetase
acuA	3039	acetoin utilization
acuB	3040	acetoin utilization
acuC	3040	acetoin utilization
adhA	2756	NADP-dependent alcohol dehydrogenase
adhB	2753	alcohol dehydrogenase
aldX	4093	aldehyde dehydrogenase
aldY	3985	aldehyde dehydrogenase
alsD	3709	α-acetolactate decarboxylase (acetoin biosynthesis)
alsS	3710	α-acetolactate synthase (acetoin biosynthesis)
amyE	327	α-amylase
amyX	3063	pullulanase
araA	2948	L-arabinose isomerase (L-arabinose utilization)
araB	2946	L-ribulokinase (L-arabinose utilization)
araD	2945	L-ribulose-5-phosphate 4-epimerase (L-arabinose utilization)
araL	2944	L-arabinose operon
araM	2943	L-arabinose operon
bglA	4122	6-phospho-—glucosidase
bglC	1940	endo-1,4—glucanase (cellulose degradation)

bglH	4033	β-glucosidase (cellulose degradation)
bglS	4011	endo-–1,3-1,4 glucanase (lichenan degradation)
crh	3569	catabolite repression HPr-like protein
csn	2748	chitosanase
csrA	3635	carbon storage regulator
fruB	1508	fructose 1-phosphate kinase
galE	3990	UDP-glucose 4-epimerase (galactose metabolism)
galK	3921	galactokinase (galactose metabolism)
galT	3919	galactose-1-phosphate uridyltransferase (galactose metabolism)
gdh	445	glucose 1-dehydrogenase
glcK	2571	glucose kinase
glgA	3167	starch (bacterial glycogen) synthase (glycogen biosynthesis)
glgB	3171	1,4–glucan branching enzyme (glycogen biosynthesis)
glgC	3169	glucose-1-phosphate adenylyltransferase (glycogen biosynthesis)
glgD	3168	required for glycogen biosynthesis
glgP	3165	glycogen phosphorylase (glycogen metabolism)
glpD	1004	glycerol-3-phosphate dehydrogenase (glycerol utilization)
glpK	1003	glycerol kinase (glycerol utilization)
glvA	890	6-phospho–glucosidase (arbutin fermentation)
gntK	4113	gluconate kinase (gluconate utilization)
gntZ	4116	6-phosphogluconate dehydrogenase (gluconate utilization)
gpsA	2389	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
gutB	667	sorbitol dehydrogenase
iolB	4082	myo-inositol catabolism
iolC	4081	myo-inositol catabolism
iolD	4080	myo-inositol catabolism
iolE	4078	myo-inositol catabolism
iolG	4076	myo-inositol 2-dehydrogenase (inositol catabolism)
iolH	4075	myo-inositol catabolism
iolI	4074	myo-inositol catabolism
iolS	4084	myo-inositol catabolism
kdgA	2323	deoxyphosphogluconate aldolase (pectin utilization)
kdgK	2324	2-keto-3-deoxygluconate kinase (pectin utilization)
kduD	2326	2-keto-3-deoxygluconate oxidoreductase (pectin utilization)
kduI	2325	5-keto-4-deoxyuronate isomerase (pectin utilization)
lacA	3504	β-galactosidase
lctE	329	L-lactate dehydrogenase
licH	3959	6-phospho–glucosidase
lplD	782	hydrolytic enzyme
melA	3100	α-D-galactoside galactohydrolase
mtlD	451	mannitol-1-phosphate dehydrogenase
nagA	3594	*N*-acetylglucosamine-6-phosphate deacetylase (*N*-acetyl glucosamine utilization)
nagB	3596	*N*-acetylglucosamine-6-phosphate isomerase (*N*-acetyl glucosamine utilization)
narQ	3773	required for formate dehydrogenase activity
pel	828	pectate lyase
pelB	2034	pectate lyase
pmi	3688	mannose-6-phosphate isomerase
pps	2053	phosphoenolpyruvate synthase
pta	3865	phosphotransacetylase
ptsH	1459	histidine-containing phosphocarrier protein of the phosphotransferase system (PTS) (HPr protein)
rbsK	3701	ribokinase (ribose metabolism)
sacA	3902	sucrase-6-phosphate hydrolase
sacB	3535	levansucrase
sacC	2759	levanase
sacX	3941	negative regulatory protein of SacY
treA	851	trehalose-6-phosphate hydrolase
xsa	2914	β-xylosidase / α-L-arabinosidase (xylan degradation)
xylA	1891	xylose isomerase (xylose metabolism)
xylB	1893	xylulose kinase (xylose metabolism)
xynA	2054	endo-1,4–xylanase (xylan degradation)
xynB	1888	xylan β-1,4-xylosidase (xylan degradation)
xynD	1945	endo-1,4–xylanase (xylan degradation)
ybaN	161	polysaccharide deacetylase
ybbD	188	β-hexosaminidase
ybcM	213	glucosamine-fructose-6-phosphate aminotransferase
ybfT	258	glucosamine-6-phosphate isomerase
ycbC	268	5-dehydro-4-deoxyglucarate dehydratase
ycbD	269	aldehyde dehydrogenase
ycbF	272	glucarate dehydratase
ycdF	305	glucose 1-dehydrogenase
ycdG	306	oligo-1,6-glucosidase
ycgS	352	aromatic hydrocarbon catabolism
yckE	370	β-glucosidase
yckG	375	D-arabino 3-hexulose 6-phosphate formaldehyde lyase
ycsN	466	aryl-alcohol dehydrogenase
ydaD	471	alcohol dehydrogenase
ydaF	473	acetyltransferase
ydaM	482	cellulose synthase
ydaP	488	pyruvate oxidase
ydhP	628	β-glucosidase
ydhR	631	fructokinase
ydhS	632	mannose-6-phosphate isomerase
ydhT	632	mannan endo-1,4–mannosidase
ydjE	670	fructokinase
ydjL	679	L-iditol 2-dehydrogenase
ydjP	682	arylesterase
yeaC	688	methanol dehydrogenase regulation
yesY	774	rhamnogalacturonan acetylesterase
yesZ	774	β-galactosidase
yfhM	929	epoxide hydrolase
yfhR	937	glucose 1-dehydrogenase
yfjS	869	polysaccharide deacetylase
yfmT	807	benzaldehyde dehydrogenase
yfnH	798	glucose-1-phosphate cytidylyltransferase
ygaK	958	reticuline oxidase
yhcW	997	phosphoglycolate phosphatase
yhdF	1022	glucose 1-dehydrogenase
yhdN	1030	aldo/keto reductase
yheN	1041	endo-1,4–xylanase
yhfE	1095	glucanase
yhxB	1006	phosphomannomutase
yhxC	1115	alcohol dehydrogenase
yhxD	1118	ribitol dehydrogenase
yisS	1164	myo-inositol 2-dehydrogenase
yitF	1175	mandelate racemase
yitY	1192	L-gulonolactone oxydase
yjdE	1274	mannose-6-phosphate isomerase
yjeA	1281	endo-1,4–xylanase
yjgC	1285	formate dehydrogenase
yjmA	1300	glucuronate isomerase
yjmD	1304	sorbitol dehydrogenase
yjmE	1305	D-mannonate hydrolase
yjmF	1306	2-deoxy-D-gluconate 3-dehydrogenase
yjmI	1309	tagaturonate reductase
yjmJ	1311	altronate hydrolase
ykcC	1356	dolichol phosphate mannose synthase
ykfB	1366	chloromuconate cycloisomerase
ykfC	1367	polysugar degrading enzyme
ykoT	1403	dolichol phosphate mannose synthase
ykrW	1427	ribulose-bisphosphate carboxylase
yktC	1537	myo-inositol-1(or 4)-monophosphatase
ykuF	1477	glucose 1-dehydrogenase
ykvO	1442	glucose 1-dehydrogenase
ykvQ	1445	chitinase
yloR	1653	ribulose-5-phosphate 3-epimerase
ylxY	1741	deacetylase
ynfF	1943	endo-xylanase
yngE	1951	propionyl-CoA carboxylase
yoaC	2023	xylulokinase
yoaD	2024	phosphoglycerate dehydrogenase
yoaE	2025	formate dehydrogenase
yoaI	2031	4-hydroxyphenylacetate-3-hydroxylase
yogA	2007	alcohol dehydrogenase
yqiQ	2507	phosphoenolpyruvate mutase
yqjD	2488	propionyl-CoA carboxylase
yrhE	2780	formate dehydrogenase
yrhG	2780	formate dehydrogenase
yrhH	2778	methyltransferase
yrhO	2768	cyclodextrin metabolism
yrpG	2742	sugar-phosphate dehydrogenase
ysdC	2950	endo-1,4–glucanase
ysfC	2932	glycolate oxidase subunit
ysfD	2934	glycolate oxidase subunit
ytbE	2969	plant metabolite dehydrogenase
ytcA	3155	NDP-sugar dehydrogenase
ytcB	3156	NDP-sugar epimerase
ytcI	3024	acetate-CoA ligase
ytdA	3155	UTP-glucose-1-phosphate uridylyltransferase
ytiB	3138	carbonic anhydrase
ytoP	3055	endo-1,4–glucanase
yttI	2989	acetyl-CoA carboxylase
yugF	3227	dihydrolipoamide *S*-acetyltransferase
yugJ	3224	NADH-dependent butanol dehydrogenase
yugK	3222	NADH-dependent butanol dehydrogenase
yugT	3215	exo–1,4-glucosidase
yulC	3200	rhamnulokinase
yulE	3198	L-rhamnose isomerase
yusZ	3382	retinol dehydrogenase
yutF	3318	*N*-acetyl-glucosamine catabolism
yuxG	3203	sorbitol-6-phosphate 2-dehydrogenase
yvaM	3455	hydrolase
yvcN	3568	*N*-hydroxyarylamine *O*-acetyltransferase
yvcT	3562	glycerate dehydrogenase
yvdA	3561	carbonic anhydrase
yvdF	3557	glucan 1,4–maltohydrolase
yvdL	3548	oligo-1,6-glucosidase
yvdM	3547	β-phosphoglucomutase
yveB	3537	levanase
yvfO	3502	arabinogalactan endo-1,4–galactosidase
yvfQ	3499	hydrolase
yvfV	3495	glycolate oxidase
yvgN	3427	plant-metabolite dehydrogenase
yvkC	3615	pyruvate,water dikinase
yvoE	3592	phosphoglycolate phosphatase
yvoF	3591	*O*-acetyltransferase
yvpA	3590	pectate lyase
yvyH	3664	UDP-*N*-acetylglucosamine 2-epimerase
ywdH	3895	aldehyde dehydrogenase
ywfD	3872	glucose 1-dehydrogenase
ywjI	3805	glycerol-inducible protein
ywqF	3730	NDP-sugar dehydrogenase
yxbG	4091	glucose 1-dehydrogenase
yxiA	4040	arabinan endo-1,5–L-arabinosidase
yxjF	4000	gluconate 5-dehydrogenase
yxnA	4107	glucose 1-dehydrogenase
yyaE	4202	formate dehydrogenase
yyaI	4196	galactoside acetyltransferase
yycR	4136	formaldehyde dehydrogenase

II.1.2 MAIN GLYCOLYTIC PATHWAYS 28

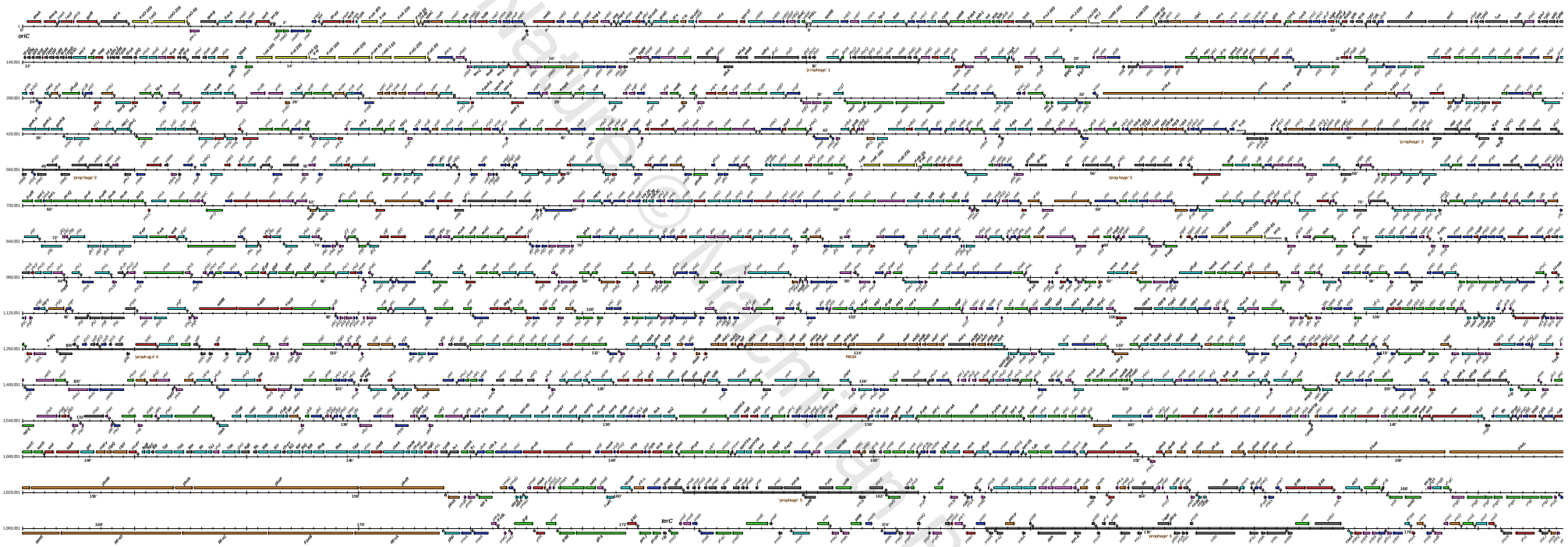
eno	3477	enolase (glycolysis)
fbaA	3808	fructose-1,6-bisphosphate aldolase (glycolysis)
fbp	4127	fructose-1,6-bisphosphatase (gluconeogenesis)
gap	3482	glyceraldehyde 3-phosphate dehydrogenase (glycolysis)
gapB	2967	glyceraldehyde 3-phosphate dehydrogenase (glycolysis)
iolJ	4073	fructose-1,6-bisphosphate aldolase (glycolysis)
pckA	3129	phosphoenolpyruvate carboxykinase
pdhA	1528	pyruvate dehydrogenase (E1 α subunit)
pdhB	1529	pyruvate dehydrogenase (E1 β subunit)
pdhC	1530	pyruvate dehydrogenase (dihydrolipoamide acetyltransferase E2 subunit)
pdhD	1531	pyruvate dehydrogenase / 2-oxoglutarate dehydrogenase (dihydrolipoamide dehydrogenase E3 subunit)
pfk	2987	6-phosphofructokinase (glycolysis)
pgi	3221	glucose-6-phosphate isomerase (glycolysis)
pgk	3480	phosphoglycerate kinase (glycolysis)
pgm	3478	phosphoglycerate mutase (glycolysis)
pycA	1554	pyruvate carboxylase
pykA	2986	pyruvate kinase (glycolysis)
tkt	1919	transketolase (pentose phosphate)
tpi	3479	triose phosphate isomerase (glycolysis)
ybbT	198	phosphoglucomutase (glycolysis)
ydeA	558	glyceraldehyde 3-phosphate dehydrogenase (glycolysis)
yhfR	1109	phosphoglycerate mutase (glycolysis)
yqeC	2651	6-phosphogluconate dehydrogenase (pentose phosphate)
yqiV	2501	dihydrolipoamide dehydrogenase
yqjI	2481	6-phosphogluconate dehydrogenase (pentose phosphate)
yqjJ	2478	glucose-6-phosphate 1-dehydrogenase (pentose phosphate)
ywjH	3807	transaldolase (pentose phosphate)
ywlF	3791	ribose 5-phosphate epimerase (pentose phosphate)

II.1.3 TCA CYCLE 19

citA	1021	citrate synthase I
citB	1926	aconitate hydratase
citC	2980	isocitrate dehydrogenase
citG	3389	fumarate hydratase
citH	2979	malate dehydrogenase
citZ	2981	citrate synthase II
malS	3058	malate dehydrogenase
mmgD	2510	citrate synthase III
odhA	2111	2-oxoglutarate dehydrogenase (E1 subunit)
odhB	2108	2-oxoglutarate dehydrogenase (dihydrolipoamide transsuccinylase, E2 subunit)
sdhA	2907	succinate dehydrogenase (flavoprotein subunit)
sdhB	2905	succinate dehydrogenase (iron-sulphur protein)
sdhC	2908	succinate dehydrogenase (cytochrome b_{558} subunit)
sucC	1680	succinyl-CoA synthetase (β subunit)
sucD	1681	succinyl-CoA synthetase (α subunit)
yjmC	1303	malate dehydrogenase
yqkJ	2452	malate dehydrogenase
ytsJ	2990	malate dehydrogenase
ywkA	3801	malate dehydrogenase

II.2 METABOLISM OF AMINO ACIDS AND RELATED MOLECULES 205

ald	3277	L-alanine dehydrogenase
ampS	1516	aminopeptidase
ansA	2456	L-asparaginase
ansB	2455	L-aspartase
aprE	1105	extracellular alkaline serine protease (subtilisin E)
aprX	1862	intracellular alkaline serine protease
argB	1197	*N*-acetylglutamate 5-phosphotransferase (arginine biosynthesis)
argC	1195	*N*-acetylglutamate γ-semialdehyde dehydrogenase (arginine biosynthesis)
argD	1198	*N*-acetylornithine aminotransferase (arginine biosynthesis)
argE	2142	acetylornithine deacetylase (arginine biosynthesis)
argF	1203	ornithine carbamoyltransferase (arginine biosynthesis)
argG	3013	argininosuccinate synthase (arginine biosynthesis)
argH	3012	argininosuccinate lyase (arginine biosynthesis)
argJ	1196	ornithine acetyltransferase / amino-acid acetyltransferase (arginine biosynthesis)
aroA	3046	3-deoxy-D-arabino-heptulosonate 7-phosphate synthase / chorismate mutase-isozyme 3 (shikimate pathway)
aroB	2378	3-dehydroquinate synthase (shikimate pathway)
aroC	2413	3-dehydroquinate dehydratase (shikimate pathway)
aroD	2645	shikimate 5-dehydrogenase (shikimate pathway)
aroE	2368	5-enolpyruvoylshikimate-3-phosphate synthase (shikimate pathway)
aroF	2380	chorismate synthase (shikimate pathway)
aroH	2377	chorismate mutase (isozymes 1 and 2) (aromatic amino acids biosynthesis)
aroI	340	shikimate kinase (shikimate pathway)
asd	1745	aspartate-semialdehyde dehydrogenase
ask	2910	aspartokinase II attenuator
asnB	3127	asparagine synthetase
asnH	4098	asparagine synthetase
aspB	2348	aspartate aminotransferase
bcsA	2317	naringenin-chalcone synthase (phenylalanine metabolism)
bfmBAA	2499	branched-chain α-keto acid dehydrogenase E1 (2-oxoisovalerate dehydrogenase α subunit)
bfmBAB	2498	branched-chain α-keto acid dehydrogenase E1 (2-oxoisovalerate dehydrogenase β subunit)
bfmBB	2497	branched-chain α-keto acid dehydrogenase E2 subunit (lipoamide acyltransferase)
bltD	2718	spermine/spermidine acetyltransferase
bpr	1599	bacillopeptidase F
cad	1535	lysine decarboxylase
carA	1199	carbamoyl-phosphate transferase-arginine (subunit A) (arginine biosynthesis)
carB	1200	carbamoyl-phosphate transferase-arginine (subunit B) (arginine biosynthesis)
ctpA	2133	carboxy-terminal processing protease
cysE	113	serine acetyltransferase (cysteine biosynthesis)
cysH	1630	phosphoadenosine phosphosulfate reductase (cysteine biosynthesis)
cysK	82	cysteine synthetase A (cysteine biosynthesis)
dal	517	D-alanine racemase
dapA	1748	dihydrodipicolinate synthase (diaminopimelate/lysine biosynthesis)
dapB	2359	dihydrodipicolinate reductase (diaminopimelate/lysine biosynthesis)
dapG	1747	aspartokinase I (α and β subunits)
def	1646	polypeptide deformylase
epr	3939	minor extracellular serine protease
glmS	200	L-glutamine-D-fructose-6-phosphate amidotransferase
glnA	1878	glutamine synthetase
gltA	2014	glutamate synthase (large subunit) (glutamate biosynthesis)
gltB	2009	glutamate synthase (small subunit) (glutamate biosynthesis)
glyA	3789	serine hydroxymethyltransferase (glycine/serine/threonine metabolism)
hisA	3584	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (histidine biosynthesis)
hisB	3585	imidazoleglycerol-phosphate dehydratase (histidine biosynthesis)
hisC	2371	histidinol-phosphate aminotransferase (histidine biosynthesis) / tyrosine and phenylalanine aminotransferase
hisD	3587	histidinol dehydrogenase (histidine biosynthesis)
hisF	3583	HisF cyclase-like protein (synthesis of D-erythro-imidazole glycerol phosphate)
hisG	3587	ATP phosphoribosyltransferase (histidine biosynthesis)
hisH	3585	amidotransferase (histidine biosynthesis)
hisI	3583	phosphoribosyl-AMP cyclohydrolase / phosphoribosyl-ATP pyrophosphohydrolase (histidine biosynthesis)
hom	3315	homoserine dehydrogenase (threonine/methionine biosynthesis)
hutG	4045	formiminoglutamate hydrolase (histidine utilization)
hutH	4041	histidase (histidine utilization)
hutI	4044	imidazolone-5-propionate hydrolase (histidine utilization)
hutU	4042	urocanase (histidine utilization)
ilvA	2293	threonine dehydratase (isoleucine biosynthesis)
ilvB	2896	acetolactate synthase (large subunit) (valine/isoleucine biosynthesis)
ilvC	2894	ketol-acid reductoisomerase (valine/isoleucine biosynthesis)
ilvD	2302	dihydroxy-acid dehydratase (valine/isoleucine biosynthesis)
ilvN	2894	acetolactate synthase (small subunit)



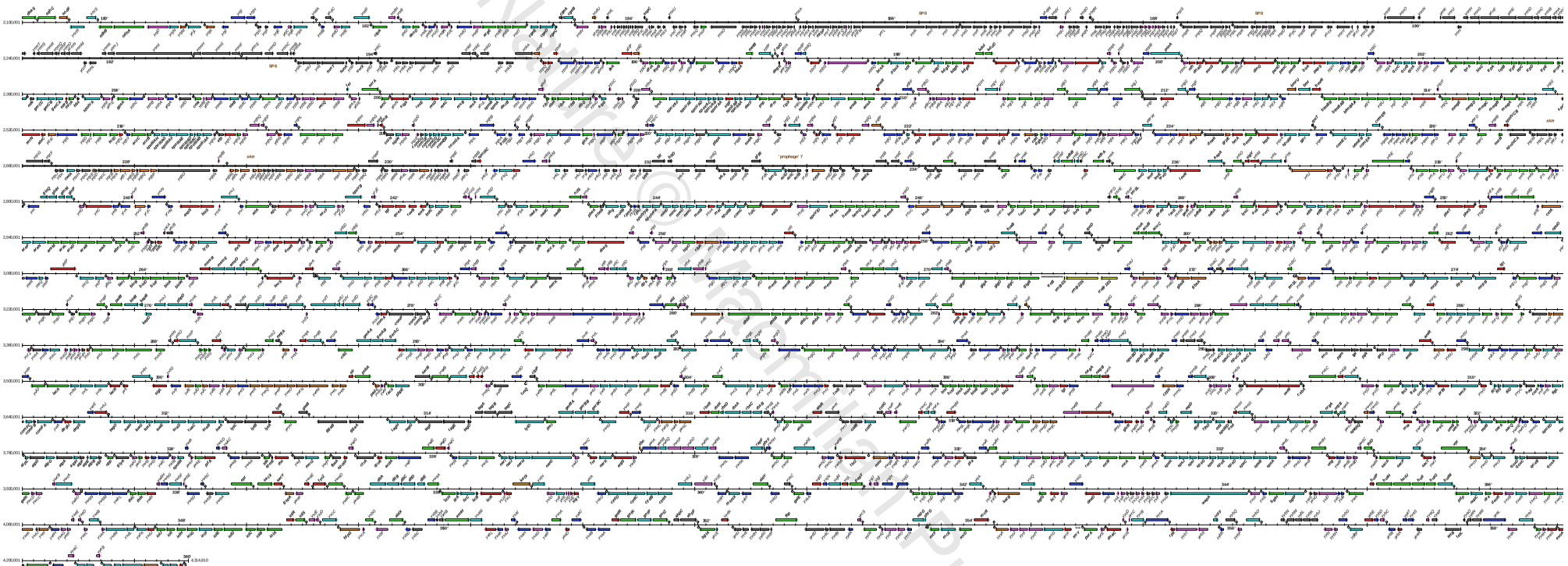


Table 1. (continuation) Functional classification of the *Bacillus subtilis* protein-coding genes.

		(valine/isoleucine biosynthesis)
iolA	4083	methylmalonate-semialdehyde dehydrogenase (valine metabolism)
ipi	1189	intracellular proteinase inhibitor
ispA	1386	major intracellular serine protease
kbl	1771	2-amino-3-ketobutyrate CoA ligase
leuA	2893	2-isopropylmalate synthase (leucine biosynthesis)
leuB	2891	3-isopropylmalate dehydrogenase (leucine biosynthesis)
leuC	2890	3-isopropylmalate dehydratase (large subunit) (leucine biosynthesis)
leuD	2889	3-isopropylmalate dehydratase (small subunit) (leucine biosynthesis)
lysA	2437	diaminopimelate decarboxylase (lysine biosynthesis)
lysC	2910	aspartokinase II (α and β subunits) (diaminopimelate/lysine biosynthesis)
metB	2305	homoserine *O*-succinyltransferase (methionine biosynthesis)
metC	1385	cobalamin-independent methionine synthase (methionine biosynthesis)
metK	3128	*S*-adenosylmethionine synthetase
mpr	245	extracellular metalloprotease
nasB	362	assimilatory nitrate reductase (electron transfer subunit)
nasC	360	assimilatory nitrate reductase (catalytic subunit)
nasD	358	assimilatory nitrite reductase (subunit)
nasE	355	assimilatory nitrite reductase (subunit)
nprB	1186	extracellular neutral protease B
nprE	1541	extracellular neutral metalloprotease
nrgB	3757	nitrogen-regulated PII-like protein
patA	1472	aminotransferase
patB	3228	aminotransferase
pepT	3994	peptidase T
pheA	2851	prephenate dehydratase (phenylalanine biosynthesis)
pheB	2852	chorismate mutase (phenylalanine biosynthesis)
proA	1379	γ-glutamyl phosphate reductase (proline biosynthesis)
proB	1378	γ-glutamyl kinase (proline biosynthesis)
proH	2017	involved in proline biosynthesis (salt-inducible)
proJ	2016	glutamate 5-kinase (proline biosynthesis)
racX	3533	amino acid racemase
rocA	3879	pyrroline-5 carboxylate dehydrogenase (arginine and ornithine utilization)
rocB	3878	involved in arginine and ornithine utilization
rocD	4145	ornithine aminotransferase (arginine and ornithine utilization)
rocF	4142	arginase (arginine and ornithine utilization)
serA	2410	phosphoglycerate dehydrogenase (serine biosynthesis)
serC	1076	phosphoserine aminotransferase (serine biosynthesis)
tdh	1770	threonine 3-dehydrogenase (threonine catabolism)
thrB	3313	homoserine kinase (threonine biosynthesis)
thrC	3314	threonine synthase (threonine biosynthesis)
trpA	2372	tryptophan synthase (α subunit) (tryptophan biosynthesis)
trpB	2373	tryptophan synthase (β subunit) (tryptophan biosynthesis)
trpC	2374	indol-3-glycerol phosphate synthase (tryptophan biosynthesis)
trpD	2375	anthranilate phosphoribosyltransferase (tryptophan biosynthesis)
trpE	2377	anthranilate synthase (tryptophan biosynthesis)
trpF	2373	phosphoribosyl anthranilate isomerase (tryptophan biosynthesis)
tyrA	2370	prephenate dehydrogenase (tyrosine biosynthesis)
ureA	3768	urease (γ subunit)
ureB	3768	urease (β subunit)
ureC	3767	urease (α subunit)
vpr	3907	minor extracellular serine protease
yaaO	38	lysine decarboxylase
ybgE	259	branched-chain amino acid aminotransferase
ybgJ	265	glutaminase
yccC	290	asparaginase
ycgM	344	proline oxidase
ycgN	345	1-pyrroline-5-carboxylate dehydrogenase
yclE	415	prolyl aminopeptidase
yclM	432	homoserine dehydrogenase
ycnG	441	4-aminobutyrate aminotransferase
ycnH	443	succinate-semialdehyde dehydrogenase
ycsA	452	3-isopropylmalate dehydrogenase
ycsJ	459	allophanate hydrolase
yerD	718	glutamate synthase (ferredoxin)
yerM	729	amidase
yhaA	1081	aminoacylase
yhdR	1034	aspartate aminotransferase
yheM	1041	D-alanine aminotransferase
yisK	1152	5-oxo-1,2,5-tricarboxilic-3-penten acid decarboxylase
yisO	1157	asparagine synthase
yisW	1167	opine aminotransferase
yjbG	1231	oligoendopeptidase
yjbR	1243	sarcosine oxidase
yjcI	1258	cystathionine γ-synthase
yjcJ	1259	cystathionine β-lyase
ykeA	1359	pyrroline-5-carboxylate reductase
ykrV	1425	aspartate aminotransferase
ykuQ	1488	tetrahydrodipicolinate succinylase
ykuR	1489	hippurate hydrolase
ylaM	1551	glutaminase
ylmB	1607	acetylornithine deacetylase
yloW	1658	phosphoglycerate dehydrogenase
ylpA	1658	L-serine dehydratase
ymfG	1757	processing protease
ymfH	1758	processing protease
ymxG	1742	processing protease
ynaI	1885	phosphoribosylanthranilate isomerase
yncD	1898	alanine racemase
yobN	2074	L-amino acid oxidase
yodT	2145	adenosylmethionine-8-amino-7-oxononanoate aminotransferase
ypcA	2403	glutamate dehydrogenase
ypwA	2321	carboxypeptidase
yqeI	2644	dihydrodipicolinate reductase
yqhI	2549	aminomethyltransferase
yqhJ	2547	glycine dehydrogenase
yqhK	2546	glycine dehydrogenase
yqhS	2539	3-dehydroquinate dehydratase
yqiT	2503	leucine dehydrogenase
yqjE	2486	tripeptidase
yqjN	2475	amino acid degradation
yqjO	2472	pyrroline-5-carboxylate reductase
yqjR	2470	D-serine dehydratase
yrbE	2839	opine catabolism
yrhA	2786	cysteine synthase
yrhB	2785	cystathionine γ-synthase
yrhP	2768	dihydrodipicolinate reductase
yrpC	2738	glutamate racemase
yrrN	2794	protease
yrrO	2793	protease
ysnE	2897	acetyltransferase
ytfD	3146	*N*-acylamino acid racemase
ytkP	3066	cysteine synthase
yubC	3193	cysteine dioxygenase
yugH	3226	aspartate aminotransferase
yurG	3341	aspartate aminotransferase
yurH	3343	*N*-carbamyl-L-amino acid amidohydrolase
yurL	3347	opine catabolism
yurP	3351	opine catabolism
yurR	3353	opine catabolism
yurT	3354	methylglyoxalase
yusH	3366	glycine cleavage system protein H
yusM	3373	proline dehydrogenase
yusX	3381	oligoendopeptidase
yutL	3306	diaminopimelate epimerase
yuxL	3312	acylaminoacyl-peptidase
yvaK	3454	carboxylesterase
yvfD	3516	serine *O*-acetyltransferase
yvjB	3623	carboxy-terminal processing protease
ywaA	3956	branched-chain amino acid aminotransferase
ywaD	3947	aminopeptidase
yweB	3881	glutamate dehydrogenase
ywfG	3868	aspartate aminotransferase
ywhF	3849	spermidine synthase
ywhG	3848	agmatinase
ywrD	3720	γ-glutamyltransferase
yxeP	4057	aminoacylase
II.3		METABOLISM OF NUCLEOTIDES AND NUCLEIC ACIDS 83
adeC	1521	adenine deaminase
adk	146	adenylate kinase
apt	2823	adenine phosphoribosyltransferase
cdd	2611	cytidine/deoxycytidine deaminase
cmk	2396	cytidylate kinase
ctrA	3811	CTP synthetase (pyrimidine biosynthesis)
deoD	2135	purine nucleoside phosphorylase (purine nucleoside salvage)
dra	4051	deoxyribose-phosphate aldolase (nucleotide/deoxyribonucleotide catabolism)
drm	2448	phosphodeoxyribomutase (purine nucleoside salvage)
guaA	692	GMP synthetase (GMP biosynthesis)
guaB	16	inositol-monophosphate dehydrogenase (GMP biosynthesis)
hipO	3000	hippurate hydrolase
hprT	76	hypoxanthine-guanine phosphoribosyltransferase (purine salvage)
ndk	2381	nucleoside diphosphate kinase
nin	372	inhibitor of the DNA degrading activity of NucA
nrdE	1868	ribonucleoside-diphosphate reductase (major subunit)
nrdF	1870	ribonucleoside-diphosphate reductase (minor subunit)
nucA	372	membrane-associated nuclease
nucB	2652	sporulation-specific extracellular nuclease
pdp	4049	pyrimidine-nucleoside phosphorylase
pnp	2446	purine nucleoside phosphorylase (purine nucleoside salvage)
pnpA	1739	polynucleotide phosphorylase
prs	58	phosphoribosyl pyrophosphate synthetase (nucleotide biosynthesis)
purA	4156	adenylosuccinate synthetase (AMP biosynthesis)
purB	700	adenylosuccinate lyase (purine biosynthesis)
purC	701	phosphoribosylaminoimidazole succinocarboxamide synthetase (purine biosynthesis)
purD	710	phosphoribosylglycinamide synthetase (purine biosynthesis)
purE	698	phosphoribosylaminoimidazole carboxylase I (purine biosynthesis)
purF	705	phosphoribosylpyrophosphate amidotransferase (purine biosynthesis)
purH	708	phosphoribosylaminoimidazole carboxy formyl formyltransferase / inosine-monophosphate cyclohydrolase (purine biosynthesis)
purK	699	phosphoribosylaminoimidazole carboxylase II (purine biosynthesis)
purL	702	phosphoribosylformylglycinamidine synthetase II (purine biosynthesis)
purM	706	phosphoribosylaminoimidazole synthetase (purine biosynthesis)
purN	708	phosphoribosylglycinamide formyltransferase (purine biosynthesis)
purQ	703	phosphoribosylformylglycinamidine synthetase I (purine biosynthesis)
purT	244	phosphoribosylglycinamide formyltransferase 2 (purine biosynthesis)
pyrAA	1622	carbamoyl-phosphate synthetase (glutaminase subunit) (pyrimidine biosynthesis)
pyrAB	1623	carbamoyl-phosphate synthetase (catalytic subunit) (pyrimidine biosynthesis)
pyrB	1620	aspartate carbamoyltransferase (pyrimidine biosynthesis)
pyrC	1621	dihydroorotase (pyrimidine biosynthesis)
pyrD	1627	dihydroorotate dehydrogenase (pyrimidine biosynthesis)
pyrDII	1626	dihydroorotate dehydrogenase (electron transfer subunit) (pyrimidine biosynthesis)
pyrE	1629	orotate phosphoribosyltransferase (pyrimidine biosynthesis)
pyrF	1628	orotidine 5′-phosphate decarboxylase (pyrimidine biosynthesis)
relA	2822	GTP pyrophosphokinase (stringent response)
smbA	1719	uridylate kinase (pyrimidine biosynthesis)
tdk	3802	thymidine kinase
thyA	1901	thymidylate synthase A (deoxyribonucleotide biosynthesis)
thyB	2297	thymidylate synthase B (deoxyribonucleotide biosynthesis)
tmk	39	thymidylate kinase
udk	2792	uridine kinase (pyrimidine salvage)
upp	3788	uracil phosphoribosyltransferase (pyrimidine salvage)
xpt	2319	xanthine phosphoribosyltransferase (purine biosynthesis)
yaaF	23	deoxypurine kinase subunit
yaaG	24	deoxypurine kinase subunit
yabR	70	polyribonucleotide nucleotidyltransferase
yerA	713	adenine deaminase
yfkN	859	2′,3′-cyclic-nucleotide 2′-phosphodiesterase
yhaM	1069	CMP-binding factor
yhcR	991	5′-nucleotidase
yirY	1144	DNA exonuclease
yjbM	1236	GTP pyrophosphokinase
yjbP	1240	diadenosine tetraphosphatase
ykkE	1377	formyltetrahydrofolate deformylase
ylbB	1565	IMP dehydrogenase
yloD	1641	guanylate kinase
ymaA	1868	ribonucleoprotein
yncB	1895	micrococcal nuclease
yncF	1899	deoxyuridine 5′-triphosphate pyrophosphatase
yosN	2165	ribonucleoside-diphosphate reductase (α subunit)
yosO	2164	ribonucleoside-diphosphate reductase (α subunit)
yosP	2161	ribonucleoside-diphosphate reductase (β subunit)
yosS	2159	deoxyuridine 5′-triphosphate nucleotidohydrolase
ypfD	2395	ribosomal protein S1 homologue
yqiB	2528	exodeoxyribonuclease VII (large subunit)
yqiC	2526	exodeoxyribonuclease VII (small subunit)
yrdF	2730	ribonuclease inhibitor
yrrU	2787	purine nucleoside phosphorylase
yumD	3302	GMP reductase
yunH	3328	allantoinase
yunL	3332	uricase
yurI	3343	ribonuclease
ywaC	3949	GTP-pyrophosphokinase
II.4		METABOLISM OF LIPIDS 77
accA	2988	acetyl-CoA carboxylase (α subunit) (long-chain fatty acid biosynthesis)
accB	2531	acetyl-CoA carboxylase (biotin carboxyl carrier subunit) (long-chain fatty acid biosynthesis)
accC	2531	acetyl-CoA carboxylase (biotin carboxylase subunit) (long-chain fatty acid biosynthesis)
acdA	3813	acyl-CoA dehydrogenase
acpA	1665	acyl carrier protein (fatty acid biosynthesis)
cdsA	1721	phosphatidate cytidylyltransferase (phospholipid biosynthesis)
dgkA	2611	diacylglycerol kinase (phospholipid biosynthesis)
fabD	1663	malonyl CoA-acyl carrier protein transacylase (fatty acid biosynthesis)
fabG	1664	3-ketoacyl-acyl carrier protein reductase (fatty acid biosynthesis)
glpQ	234	glycerophosphoryl diester phosphodiesterase (glycerol metabolism)
lcfA	2919	long chain acyl-CoA synthetase (fatty acid metabolism)
lipA	292	lipase
lipB	910	lipase
mmgA	2513	acetyl-CoA acetyltransferase
mmgB	2512	3-hydroxybutyryl-CoA dehydrogenase
mmgC	2511	acyl-CoA dehydrogenase
nap	593	carboxylesterase NA
pgsA	1762	phosphatidylglycerophosphate synthase (acidic phospholipid biosynthesis)
plsX	1662	involved in fatty acid/phospholipid synthesis
pnbA	3530	*p*-nitrobenzyl esterase
psd	249	phosphatidylserine decarboxylase (phospholipid biosynthesis)
pssA	248	phosphatidylserine synthase (phospholipid biosynthesis)
sqhC	2101	squalene-hopene cyclase (hopanoid metabolism)
ybfK	247	carboxylesterase
yclB	412	phenylacrylic acid decarboxylase
ydbM	505	butyryl-CoA dehydrogenase
ydcB	515	holo- acyl-carrier protein synthase
yfjR	871	3-hydroxyisobutyrate dehydrogenase
yhaR	1061	3-hydroxbutyryl-CoA dehydratase
yhdO	1031	1-acylglycerol-3-phosphate *O*-acyltransferase
yhdW	1038	glycerophosphodiester phosphodiesterase
yhfB	1093	3-oxoacyl- acyl-carrier protein synthase
yhfJ	1099	lipoate-protein ligase
yhfL	1100	long-chain fatty-acid-CoA ligase
yhfS	1110	acetyl-CoA *C*-acetyltransferase
yhfT	1111	long-chain fatty-acid-CoA ligase
yisP	1159	phytoene synthase
yjaX	1208	3-oxoacyl- acyl-carrier protein synthase
yjaY	1209	3-oxoacyl- acyl-carrier protein synthase
yjbW	1247	enoyl- acyl-carrier protein reductase
yjdA	1268	3-oxoacyl- acyl-carrier protein reductase
ykhA	1372	acyl-CoA hydrolase
ykwC	1465	3-hydroxyisobutyrate dehydrogenase
ymfI	1759	3-oxoacyl- acyl-carrier protein reductase
yngF	1951	3-hydroxbutyryl-CoA dehydratase
yngG	1952	hydroxymethylglutaryl-CoA lyase
yngI	1955	long-chain acyl-CoA synthetase
yngJ	1957	butyryl-CoA dehydrogenase
yocE	2089	fatty-acid desaturase
yocJ	2096	ACP phosphodiesterase
yodR	2143	butyrate-acetoacetate CoA-transferase
yodS	2144	3-oxoadipate CoA-transferase
yoxD	2019	3-oxoacyl- acyl-carrier protein reductase
yqiD	2526	geranyltranstransferase
yqiK	2514	glycerophosphodiester phosphodiesterase
yqiS	2504	phosphate butyryltransferase
yqiU	2502	branched-chain fatty-acid kinase
yqjQ	2471	ketoacyl reductase
ysiB	2917	3-hydroxbutyryl-CoA dehydratase
ytkK	3011	3-oxoacyl- acyl-carrier protein reductase
ytpA	3123	lysophospholipase
yusJ	3368	butyryl-CoA dehydrogenase
yusK	3369	acetyl-CoA *C*-acyltransferase
yusL	3372	3-hydroxyacyl-CoA dehydrogenase
yusQ	3376	acyloate catabolism
yusR	3376	3-oxoacyl- acyl-carrier protein reductase
yusS	3377	3-oxoacyl- acyl-carrier protein reductase
yvaG	3450	3-oxoacyl- acyl-carrier protein reductase
yvrD	3404	ketoacyl-carrier protein reductase
ywfH	3866	3-oxoacyl- acyl-carrier protein reductase
ywhB	3853	4-oxalocrotonate tautomerase
ywiE	3822	cardiolipin synthetase
ywjE	3816	cardiolipin synthetase
ywnE	3762	cardiolipin synthase

ywpB	3743	hydroxymyristoyl-(acyl carrier protein) dehydratase
yxjD	4001	3-oxoadipate CoA-transferase
yxjE	4001	3-oxoadipate CoA-transferase
II.5		METABOLISM OF COENZYMES AND PROSTHETIC GROUPS 99
bioA	3094	adenosylmethionine-8-amino-7-oxononanoate aminotransferase (biotin biosynthesis)
bioB	3091	biotin synthetase (biotin biosynthesis)
bioD	3091	dethiobiotin synthetase (biotin biosynthesis)
bioF	3092	8-amino-7-oxononanoate synthase (biotin biosynthesis)
bioI	3089	cytochrome P450-like enzyme (biotin biosynthesis)
bioW	3094	6-carboxyhexanoate-CoA ligase (biotin biosynthesis)
dfrA	2296	dihydrofolate reductase (glycine/purine/DNA precursor synthesis, conversion of dUMP to dTMP)
dhaS	2100	aldehyde dehydrogenase
dhbA	3291	2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase (2,3-dihydroxybenzoate biosynthesis)
dhbB	3288	isochorismatase (2,3-dihydroxybenzoate biosynthesis)
dhbC	3291	isochorismate synthase (2,3-dihydroxybenzoate biosynthesis)
dhbE	3289	2,3-dihydroxybenzoate-AMP ligase (enterobactin synthetase component E) (2,3-dihydroxybenzoate biosynthesis)
dhbF	3287	involved in 2,3-dihydroxybenzoate biosynthesis
folA	87	dihydroneopterin aldolase (folate biosynthesis)
folC	2866	folyl-polyglutamate synthetase (folate biosynthesis)
folD	2529	methylenetetrahydrofolate dehydrogenase / methenyltetrahydrofolate cyclohydrolase (purines and amino acids biosynthesis)
folK	87	7,8-dihydro-6-hydroxymethylpterin pyrophosphokinase (dihydrofolate biosynthesis)
ggt	2004	γ-glutamyltranspeptidase (glutathione metabolism)
gsaB	943	glutamate-1-semialdehyde aminotransferase
hemA	2878	glutamyl-tRNA reductase (porphyrin biosynthesis)
hemB	2874	δ-aminolevulinic acid dehydratase (porphyrin biosynthesis)
hemC	2876	porphobilinogen deaminase (porphyrin biosynthesis)
hemD	2875	uroporphyrinogen III cosynthase (porphyrin biosynthesis)
hemE	1086	uroporphyrinogen III decarboxylase (porphyrin biosynthesis)
hemH	1087	ferrochelatase (porphyrin biosynthesis)
hemL	2873	glutamate-1-semialdehyde 2,1-aminotransferase (porphyrin biosynthesis)
hemN	2630	coproporphyrinogen III oxidase (porphyrin biosynthesis)
hemX	2877	negative effector of the concentration of HemA
hemY	1088	protoporphyrinogen IX oxidase (porphyrin biosynthesis)
menB	3149	dihydroxynapthoic acid synthetase (menaquinone biosynthesis)
menD	3151	2-succinyl-6-hydroxy-2,4-cyclohexadiene-1-carboxylate synthase / 2-oxoglutarate decarboxylase (menaquinone biosynthesis)
menE	3148	*O*-succinylbenzoic acid-CoA ligase (menaquinone biosynthesis)
menF	3153	menaquinone-specific isochorismate synthase (menaquinone biosynthesis)
moaB	3014	molybdopterin precursor biosynthesis
moaD	1499	molybdopterin converting factor (subunit 1)
moaE	1498	molybdopterin converting factor (subunit 2)
mobA	1495	molybdopterin-guanine dinucleotide biosynthesis
mobB	1498	molybdopterin-guanine dinucleotide biosynthesis
moeA	1497	molybdopterin biosynthesis protein
moeB	1496	molybdopterin biosynthesis protein
mtrA	2385	GTP cyclohydrolase I (tetrahydrofolate biosynthesis)
nadA	2846	quinolinate synthetase (quinolinate biosynthesis)
nadB	2849	L-aspartate oxidase (quinolinate biosynthesis)
nadC	2847	nicotinate-nucleotide pyrophosphorylase (NAD/NADP biosynthesis)
nadE	338	NH_3-dependent NAD^+ synthetase (NAD biosynthesis)
narA	3772	molybdopterin precursor biosynthesis
nasF	355	uroporphyrin-III *C*-methyltransferase (porphyrin biosynthesis)
nifS	2849	required for NAD biosynthesis
pabA	84	*p*-aminobenzoate synthase glutamine amidotransferase (subunit B) / anthranilate synthase (subunit II) (folate and tryptophan biosynthesis)
pabB	83	*p*-aminobenzoate synthase (subunit A) (folate biosynthesis)
pabC	85	aminodeoxychorismate lyase (folate biosynthesis)
panB	2354	ketopantoate hydroxymethyltransferase (pantothenate biosynthesis)
panC	2353	pantothenate synthetase (pantothenate biosynthesis)
panD	2352	aspartate 1-decarboxylase (pantothenate biosynthesis)
ribA	2429	GTP cyclohydrolase II / 3,4-dihydroxy-2-butanone 4-phosphate synthase (riboflavin biosynthesis)
ribB	2429	riboflavin synthase (α subunit) (riboflavin biosynthesis)
ribC	1737	riboflavin kinase / FAD synthase (riboflavin biosynthesis)
ribG	2431	riboflavin-specific deaminase (riboflavin biosynthesis)
ribH	2428	riboflavin synthase (β subunit) (riboflavin biosynthesis)
ribT	2427	reductase (riboflavin biosynthesis)
sul	86	dihydropteroate synthase (dihydrofolate biosynthesis)
thiA	955	synthesis of the pyrimidine moiety of thiamin (thiamin biosynthesis)
thiC	3930	thiamine-phosphate pyrophosphorylase (thiamin biosynthesis)
thiD	3900	phosphomethylpyrimidine kinase (thiamin biosynthesis)
thiK	3931	hydroxyethylthiazole kinase (thiamin biosynthesis)
yaaI	26	isochorismatase
ydiA	640	thiamin-monophosphate kinase
ydiG	646	molybdopterin precursor biosynthesis
yhaV	1058	coproporphyrinogen III oxidase
yhcB	979	flavodoxin
yhfU	1112	biotin biosynthesis
yhxA	1000	adenosylmethionine-8-amino-7-oxononanoate aminotransferase
yjbT	1245	thiamin biosynthesis
yjbU	1245	thiamin biosynthesis
yjbV	1246	phosphomethylpyrimidine kinase
ykpB	1513	thiamin biosynthesis
ykvK	1440	6-pyruvoyl tetrahydrobiopterin synthase
ykvL	1440	coenzyme PQQ synthesis
ylbQ	1577	pyrimidine-thiamine biosynthesis
ylnD	1633	uroporphyrin-III *C*-methyltransferase
ylnF	1635	uroporphyrin-III *C*-methyltransferase
ylol	1642	pantothenate metabolism flavoprotein
yngH	1954	biotin carboxylase
yodC	2127	nitroreductase
yqgN	2574	5-formyltetrahydrofolate cyclo-ligase
yqjS	2469	pantothenate kinase
yrrL	2796	folate metabolism
yrrM	2795	caffeoyl-CoA *O*-methyltransferase
yueD	3265	sepiapterin reductase
yueJ	3261	pyrazinamidase/nicotinamidase
yueK	3260	nicotinate phosphoribosyltransferase
yuiG	3293	biotin metabolism
yurB	3335	4-hydroxybenzoyl-CoA reductase
yurC	3338	4-hydroxybenzoyl-CoA reductase
yurD	3338	4-hydroxybenzoyl-CoA reductase
yutB	3320	lipoic acid synthetase
ywaB	3950	quinone biosynthesis
ywkE	3796	protoporphyrinogen oxidase
ywoC	3755	isochorismatase
II.6		METABOLISM OF PHOSPHATE 9
phoA	1018	alkaline phosphatase A
phoB	621	alkaline phosphatase III
phoD	284	phosphodiesterase/alkaline phosphatase
phoH	2615	phosphate starvation-induced protein
xpaC	36	hydrolysis of 5-bromo 4-chloroindolyl phosphate
ybfM	248	alkaline phosphatase
ykoX	1409	alkaline phosphatase
ylaK	1549	phosphate starvation inducible protein
yngC	1947	alkaline phosphatase
II.7		METABOLISM OF SULPHUR 8
yisZ	1170	adenylylsulfate kinase
yitA	1171	sulfate adenylyltransferase
yitB	1172	phospho-adenylylsulfate sulfotransferase
ylnB	1632	sulfate adenylyltransferase
ylnC	1633	adenylylsulfate kinase
yuiH	3293	sulfite oxidase
yvgQ	3431	sulfite reductase
yvgR	3433	sulfite reductase
III		INFORMATION PATHWAYS 482
III.1		DNA REPLICATION 22
dnaA	0	initiation of chromosome replication
dnaB	2965	initiation of chromosome replication / membrane attachment protein
dnaC	4158	replicative DNA helicase
dnaD	2345	initiation of chromosome replication
dnaE	2994	DNA polymerase III (α subunit)
dnaG	2603	DNA primase
dnaI	2963	primosome component (helicase loader)
dnaN	2	DNA polymerase III (β subunit)
dnaX	27	DNA polymerase III (γ and τ subunits)
holB	41	DNA polymerase III (δ' subunit)
polA	2975	DNA polymerase I
polC	1727	DNA polymerase III (α subunit)
priA	1643	primosomal replication factor Y
rnh	1677	ribonuclease H
rtp	2018	replication terminator protein
ssb	4199	single-strand DNA-binding protein
yerF	719	ATP-dependent DNA helicase
yerG	721	DNA ligase
yoqV	2192	DNA ligase
yorL	2179	DNA polymerase III (α subunit)
ypcP	2311	5'-3' exonuclease
ywpH	3740	single-strand DNA-binding protein
III.2		DNA RESTRICTION/MODIFICATION AND REPAIR 39
adaA	204	methylphosphotriester-DNA alkyltransferase / transcriptional activator of the *adaAB* operon
adaB	204	O^6-methylguanine-DNA methyltransferase
alkA	203	DNA-3-methyladenine glycosylase
dat	1421	O^6-methylguanine DNA alkyltransferase
dinB	608	nuclease inhibitor
dinG	2352	ATP-dependent DNA helicase
exoA	4198	3'-exo-deoxyribonuclease
mtbP	2172	modification methylase Bsu
mutL	1778	DNA mismatch repair
mutM	2972	formamidopyrimidine-DNA glycosidase
mutS	1775	DNA mismatch repair (recognition)
mutT	488	mutator protein
nth	2345	endonuclease III (DNA repair)
sms	106	DNA repair protein homologue
ung	3897	uracil-DNA glycosylase
uvrA	3612	excinuclease ABC (subunit A)
uvrB	3614	excinuclease ABC (subunit B)
uvrC	2912	excinuclease ABC (subunit C)
uvrX	2271	UV-damage repair protein
ydiO	655	DNA-methyltransferase (cytosine-specific)
ydiP	656	DNA-methyltransferase (cytosine-specific)
ydiS	660	DNA restriction
yfhQ	935	A/G-specific adenine glycosylase
yfjP	872	DNA-3-methyladenine glycosidase II
yisT	1165	nuclease inhibitor
yjcD	1255	ATP-dependent DNA helicase
yjhB	1290	mutator MutT protein
yozK	2064	DNA repair protein
yprA	2336	ATP-dependent helicase
ypvA	2329	ATP-dependent helicase
yqfS	2593	endonuclease IV
yqjH	2483	DNA-damage repair protein
yqjW	2465	ATP/GTP-binding protein
yshC	2924	DNA polymerase β
yshD	2922	DNA mismatch repair protein
ysxA	2862	DNA repair protein
yvcI	3572	mutator MutT protein
ywjD	3817	UV-endonuclease
yxlJ	3964	DNA-3-methyladenine glycosidase
III.3		DNA RECOMBINATION 17
addA	1139	ATP-dependent deoxyribonuclease (subunit A)
addB	1136	ATP-dependent deoxyribonuclease (subunit B)
recA	1764	multifunctional protein involved in homologous recombination and DNA repair (LexA-autocleavage)
recF	3	DNA repair and genetic recombination
recN	2522	DNA repair and genetic recombination
recQ	2408	ATP-dependent DNA helicase
recR	29	DNA repair and genetic recombination
ruvA	2836	Holliday junction DNA helicase
ruvB	2835	Holliday junction DNA helicase
sbcD	1143	exonuclease SbcD homologue
ylpB	1659	ATP-dependent DNA helicase
yocI	2095	ATP-dependent DNA helicase
yorK	2180	single-strand DNA-specific exonuclease
yqhH	2549	SNF2 helicase
yrrC	2808	conjugation transfer protein
yrvE	2825	single-strand DNA-specific exonuclease
ywqA	3735	SNF2 helicase
III.4		DNA PACKAGING AND SEGREGATION. 10
grlA	1935	DNA gyrase-like protein (subunit A)
grlB	1933	DNA gyrase-like protein (subunit B)
gyrA	7	DNA gyrase (subunit A)
gyrB	5	DNA gyrase (subunit B)
hbs	2385	non-specific DNA-binding protein HBsu
smc	1666	chromosome segregation SMC protein homologue
smf	1682	DNA processing Smf protein homologue
topA	1683	DNA topoisomerase I
topB	476	DNA topoisomerase III
yonN	2225	HU-related DNA-binding protein
III.5		RNA SYNTHESIS 244
III.5.1		INITIATION 19
sigA	2601	RNA polymerase major sigma factor (σ^A)
sigB	522	RNA polymerase general stress sigma factor (σ^B)
sigD	1716	RNA polymerase flagella, motility, chemotaxis and autolysis sigma factor (σ^D)
sigE	1604	RNA polymerase sporulation mother cell-specific (early) sigma factor (σ^E) (SpoIIGB)
sigF	2443	RNA polymerase sporulation forespore-specific (early) sigma factor (σ^F) (SpoIIAC)
sigG	1605	RNA polymerase sporulation forespore-specific (late) sigma factor (σ^G) (SpoIIIG)
sigH	117	RNA polymerase vegetative and early stationary-phase sigma factor (σ^H) (Spo0H)
sigL	3513	RNA polymerase sigma factor (σ^L)
sigV	2769	RNA polymerase ECF-type sigma factor (σ^V)
sigW	195	RNA polymerase ECF-type sigma factor (σ^W)
sigX	2414	RNA polymerase ECF-type sigma factor (σ^X)
sigY	3970	RNA polymerase ECF-type sigma factor (σ^Y)
sigZ	2742	RNA polymerase ECF-type sigma factor (σ^Z)
spoIIIC	2701	RNA polymerase sporulation mother cell-specific (late) sigma factor (σ^K) (C-terminal half)
spoIVCB	2652	RNA polymerase sporulation mother-cell-specific (late) sigma factor (σ^K) (N-terminal half)
xpf	1324	RNA polymerase PBSX sigma factor-like
yhdM	1030	RNA polymerase ECF-type sigma factor
ykoZ	1411	RNA polymerase sigma factor
ylaC	1543	RNA polymerase ECF-type sigma factor
III.5.2		REGULATION 213
abh	1517	transcriptional regulator of transition state genes (AbrB-like)
abrB	45	transcriptional pleiotropic regulator of transition state genes (*aprE*, *comK*, *ftsAZ*, *hpr*, *motAB*, *nprE*, *pbpE*, *rbs*, *spo0H*, *spoVG*, *spo0E*, *tycA*)
acoR	883	transcriptional activator of the acetoin dehydrogenase operon (*acoABCL*)
ahrC	2522	transcriptional regulator of arginine metabolism expression (*roc* operons)
alsR	3711	transcriptional regulator of the α-acetolactate operon (*alsSD*)
ansR	2456	transcriptional repressor of the *ansAB* operon (Xre family)
araR	3485	transcriptional repressor of the arabinose operon (*araABDLMNPQ*)
azlB	2729	transcriptional repressor of the *azlBCD* operon
birA	2355	transcriptional repressor of the biotin operon (*bioWAFDBI*) / biotin acetyl-CoA-carboxylase synthetase
bltR	2716	transcriptional regulator of the *bltD* operon
bmrR	2495	transcriptional activator of the *bmrUR* operon
ccpA	3044	transcriptional regulator involved in carbon catabolite control
cheB	1711	two-component response regulator-like [CheA] / methyl-accepting chemotaxis proteins-glutamate methylesterase
cheY	1703	two-component response regulator [CheA] involved in modulation of flagellar switch bias (chemotaxis)
citR	1020	transcriptional repressor of the citrate synthase I gene (*citA*)
citT	832	two-component response regulator [CitS]
codY	1690	transcriptional pleiotropic repressor (expression of *srfA*, *comK*, *dpp*, *gabP*, *hut*, *ureABC*)
comA	3253	two-component response regulator [ComP] of late competence genes / surfactin production
comK	1117	competence transcription factor (CTF), final autoregulatory control switch prior to competence development
comQ	3256	transcriptional regulator of late competence operon (*comG*) and surfactin expression (*srfA*)
ctsR	101	transcriptional repressor of class III stress genes (*clpC*, *clpP*)
degA	1163	transcriptional activator involved in the degradation of glutamine phosphoribosylpyrophosphate amidotransferase
degU	3644	two-component response regulator [DegS] involved in degradative enzyme and competence regulation (*sacB*, *degQ*, *comK*)
deoR	4052	transcriptional repressor of the *dra/nupC/pdp* operon (deoxyribonucleoside)
fnr	3831	transcriptional regulator of anaerobic genes (*narK*, *narGHJI*)
fruR	1507	transcriptional repressor of the fructose operon (*fruRBA*)
gerE	2904	transcriptional regulator required for expression of late spore coat genes
glcR	3739	transcriptional repressor involved in the expression of the phosphotransferase system
glcT	1456	transcriptional antiterminator essential for the expression of the *ptsGHI* operon
glnR	1877	transcriptional repressor of the glutamine synthetase gene (*glnA*)
glpP	1001	transcriptional antiterminator and control of mRNA stability of *glpD*
gltC	2014	transcriptional activator of the glutamate synthase operon (*gltAB*)
gltR	2725	transcriptional repressor of the glutamate synthase operon (*gltAB*)
gntR	4113	transcriptional repressor of the gluconate operon (*gntRKPZ*)

gutR	667	transcriptional activator of the sorbitol dehydrogenase gene (*gutA*)
hpr	1073	transcriptional repressor of sporulation and extracellular proteases genes (*aprE*, *nprE*, *sin*)
hrcA	2629	transcriptional repressor of class I heat-shock genes (*dnaK*, *groESL*)
hutP	4040	transcriptional activator of the histidine utilization operon (*hutPHUIGM*)
iolR	4084	transcriptional repressor of the myo-inositol catabolism operon (*iolABCDEFGHIJ*/*iolRS*)
kdgR	2325	transcriptional repressor of the pectin utilization operon (*kdgRKAT*)
lacR	3509	transcriptional repressor of the β-galactosidase gene (*lacA*)
levR	2765	transcriptional activator of the levanase operon (*levDEFG*/*sacC*)
lexA	1918	transcriptional repressor of the SOS regulon
licR	3963	transcriptional regulator (antiterminator) of the lichenan operon (*licBCAH*)
licT	4012	transcriptional antiterminator required for substrate-dependent induction and catabolite repression of *bglPH*
lmrA	290	transcriptional repressor of the lincomycin operon (*lmrBA*)
lrpA	551	transcriptional Lrp-like regulator (repression of *glyA* transcription and KinB-dependent sporulation)
lrpB	552	transcriptional Lrp-like regulator (repression of *glyA* transcription and KinB-dependent sporulation)
lrpC	476	transcriptional regulator (Lrp/AsnC family)
lytR	3662	attenuator role for *lytABC* and *lytR* expression
lytT	2956	two-component response regulator [LytS] involved in the rate of autolysis
msmR	3096	transcriptional regulator (LacI family)
mta	3764	transcriptional activator of multidrug-efflux transporter genes (*bmr* and *blt*)
mtrB	2384	tryptophan operon RNA-binding attenuation protein (TRAP)
paiA	3304	transcriptional repressor of sporulation, septation and degradative enzyme genes (*aprE*, *nprE*, *phoA*, *sacB*)
paiB	3304	transcriptional repressor of sporulation and degradative enzyme genes
phoP	2978	two-component response regulator [PhoR] involved in phosphate regulation (*phoA*, *phoB*, *phoD*, *resABCDE*)
pksA	1781	transcriptional regulator of the polyketide synthase operon (*pks*)
purR	54	transcriptional repressor of the purine operon (*purEKBCLQFMNHD*)
pyrR	1618	transcriptional attenuation of the pyrimidine operon (*pyrPBCADFE*) / uracil phosphoribosyltransferase activity (minor) (pyrimidine biosynthesis)
rbsR	3700	transcriptional repressor of the ribose operon (*rbsRKDACB*)
resD	2417	two-component response regulator [ResE] involved in aerobic and anaerobic respiration (*resA*, *ctaA*, *qcrABC*, *fnr*)
ribR	3001	transcriptional regulator of riboflavin biosynthesis genes
rocR	4145	transcriptional activator of arginine utilization operons (*rocABC*, *rocDEF*)
sacT	3906	transcriptional antiterminator involved in positive regulation of *sacA* and *sacP*
sacV	532	transcriptional regulator of the levansucrase gene (*sacB*)
sacY	3942	transcriptional antiterminator involved in positive regulation of levansucrase and sucrase synthesis
senS	959	transcriptional regulator of extracellular enzyme genes (*amyE*, *aprE*, *nprE*)
sinR	2552	transcriptional regulator of post-exponential-phase responses genes (*aprE*, *comK*, *kinB*, *sigD*, *spo0A*, *spoIIA*, *spoIIE*, *spoIIG*)
slr	3529	transcriptional activator of competence development and sporulation genes
splA	1461	transcriptional regulator of the spore photoproduct lyase operon (*splAB*)
spo0A	2518	two-component response regulator [KinC] central for the initiation of sporulation (*spo0A*, *abrB*, *kinA*, *kinC*, *spoIIA*, *spoIIE*, *spoIIG*) (part of phosphorelay: Spo0B~P->Spo0A~P)
spo0F	3809	two-component response regulator [KinA, KinB] involved in the initiation of sporulation (part of phosphorelay: Spo0F~P->Spo0B~P)
spoIIID	3748	transcriptional regulator of σ^E- and σ^K-dependent genes
spoVT	64	transcriptional positive and negative regulator of σ^G-dependent genes
tenA	1242	transcriptional regulator of extracellular enzyme genes (*aprE*, *nprE*, *phoA*, *sacB*)
tenI	1243	transcriptional activator of extracellular enzyme genes
tnrA	1397	transcriptional pleiotropic regulator invoved in global nitrogen regulation (expression of *nrgAB*, *nasB*, *gabP*, *ureABC*, *glnRA*)
treR	853	transcriptional repressor of the trehalose operon (*trePAR*)
xre	1321	transcriptional repressor of PBSX genes
xylR	1891	transcriptional repressor of the xylose operon (*xylAB*)
yacF	88	transcriptional regulator (nitrogen regulation protein)
ybbB	185	transcriptional regulator (AraC/XylS family)
ybdJ	221	two-component response regulator [YbdK]
ybfI	244	transcriptional regulator (AraC/XylS family)
ybfP	251	transcriptional regulator (AraC/XylS family)
ybgA	258	transcriptional regulator (GntR family)
ycbB	267	two-component response regulator [YcbA]
ycbG	273	transcriptional regulator (GntR family)
ycbL	278	two-component response regulator [YcbM]
yccH	296	two-component response regulator [YccG]
yceK	320	transcriptional regulator (ArsR family)
ycgK	341	transcriptional regulator (LysR family)
yclA	412	transcriptional regulator (LysR family)
yclJ	426	two-component response regulator [YclK]
ycnC	438	transcriptional regulator (TetR/AcrR family)
ycnF	441	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
ycnK	449	transcriptional regulator (DeoR family)
ycsO	461	transcriptional regulator (IclR family)
ycxD	406	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
yczG	439	transcriptional regulator (ArsR family)
ydaA	467	transcriptional antiterminator (BglG family)
ydbG	499	two-component response regulator [YdbF]
ydcN	531	transcriptional regulator (phage-related) (Xre family)
ydeC	562	transcriptional regulator (AraC/XylS family)
ydeE	564	transcriptional regulator (AraC/XylS family)
ydeF	564	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
ydeL	571	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
ydeS	578	transcriptional regulator (TetR/AcrR family)
ydeT	579	transcriptional regulator (ArsR family)
ydfD	583	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
ydfI	589	two-component response regulator [YdfH]
ydgG	609	transcriptional regulator (MarR family)
ydgJ	613	transcriptional regulator (MarR family)
ydhC	616	transcriptional regulator (GntR family)
ydhQ	630	transcriptional regulator (GntR family)
yerO	732	transcriptional regulator (TetR/AcrR family)
yesN	760	two-component response regulator [YesM]
yesS	765	transcriptional regulator (AraC/XylS family)
yetL	790	transcriptional regulator (MarR family)
yezC	711	transcriptional regulator (Lrp/AsnC family)
yfiF	898	transcriptional regulator (AraC/XylS family)
yfiK	905	two-component response regulator [YfiJ]
yfiV	916	transcriptional regulator (MarR family)
yfmP	812	transcriptional regulator (MerR family)
ygaG	944	transcriptional regulator (Fur family)
yhbI	976	transcriptional regulator (MarR family)
yhcF	981	transcriptional regulator (GntR family)
yhcZ	1009	two-component response regulator [YhcY]
yhdI	1027	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
yhdQ	1033	transcriptional regulator (MerR family)
yhgD	1089	transcriptional regulator (TetR/AcrR family)
yhjM	1129	transcriptional regulator (LacI family)
yisR	1162	transcriptional regulator (AraC/XylS family)
yisV	1166	transcriptional regulator (GntR family) / aminotransferase (MocR-like)
yjdC	1270	transcriptional antiterminator (BglG family)
yjdI	1277	transcription regulation
yjmH	1308	transcriptional regulator (LacI family)
ykoG	1391	two-component response regulator [YkoH]
ykoM	1398	transcriptional regulator (MarR family)
ykuM	1485	transcriptional regulator (LysR family)
ykvE	1433	transcriptional regulator (MarR family)
ykvZ	1455	transcriptional regulator (LacI family)
ymfC	1754	transcriptional regulator (GntR family)
yneI	1923	two-component response regulator (CheY homologue)
yoaU	2045	transcriptional regulator (LysR family)
yobD	2056	transcriptional regulator (phage-related) (Xre family)
yobQ	2080	transcriptional regulator (AraC/XylS family)
yocG	2091	two-component response regulator [YocF]
yofA	2007	transcriptional regulator (LysR family)
yonR	2221	transcriptional regulator (phage-related) (Xre family)
yozA	2084	transcriptional regulator (ArsR family)
yozG	2043	transcriptional regulator
yplP	2294	transcriptional regulator (σ^L-dependent)
ypoP	2287	transcriptional regulator (MarR family)
yppQ	2287	transcriptional regulator (PilB family)
ypuN	2414	negative regulator of σ^X activity
yqaE	2698	transcriptional regulator (phage-related) (Xre family)
yqcJ	2657	transcriptional regulator (ArsR family)
yqfV	2591	transcriptional regulator (Fur family)
yqhN	2543	transcriptional regulator
yqiR	2506	transcriptional regulator (σ^L-dependent)
yqkL	2450	transcriptional regulator (Fur family)
yraB	2755	transcriptional regulator (MerR family)
yraN	2746	transcriptional regulator (LysR family)
yrdQ	2721	transcriptional regulator (LysR family)
yrhI	2777	transcriptional regulator (TetR/AcrR family)
yrhM	2770	anti-sigma factor [σ^V]
yrkP	2704	two-component response regulator [YrkQ]
ysiA	2918	transcriptional regulator (TetR/AcrR family)
ysmB	2904	transcriptional regulator (MarR family)
ytdP	3083	transcriptional regulator (AraC/XylS family)
ytlI	3008	transcriptional regulator (LysR family)
ytrA	3118	transcriptional regulator (GntR family)
ytsA	3113	two-component response regulator [YtsB]
ytzE	3071	transcriptional regulator (DeoR family)
yufM	3238	two-component response regulator [YufL]
yugG	3227	transcriptional regulator (Lrp/AsnC family)
yulB	3201	transcriptional regulator (DeoR family)
yurK	3345	transcriptional regulator (GntR family)
yusO	3374	transcriptional regulator (MarR family)
yusT	3377	transcriptional regulator (LysR family)
yvbA	3466	transcriptional regulator (ArsR family)
yvbU	3488	transcriptional regulator (LysR family)
yvcP	3567	two-component response regulator [YvcQ]
yvdE	3558	transcriptional regulator (LacI family)
yvdT	3540	transcriptional regulator (TetR/AcrR family)
yvfI	3509	transcriptional regulator (GntR family)
yvfU	3496	two-component response regulator [YvfT]
yvhI	3646	transcriptional regulator
yvkB	3617	transcriptional regulator (TetR/AcrR family)
yvoA	3596	transcriptional regulator (GntR family)
yvqA	3385	two-component response regulator [YvqB]
yvqC	3394	two-component response regulator [YvqE]
yvrH	3409	two-component response regulator [YvrG]
ywaE	3945	transcriptional regulator (MarR family)
ywbI	3932	transcriptional regulator (LysR family)
ywfK	3864	transcriptional regulator (LysR family)
ywhA	3853	transcriptional regulator (MarR family)
ywoH	3748	transcriptional regulator (MarR family)
ywqM	3723	transcriptional regulator (LysR family)
ywrC	3720	transcriptional regulator (Lrp/AsnC family)
ywtF	3693	transcriptional regulator
yxaD	4109	transcriptional regulator (MarR family)
yxdJ	4072	two-component response regulator [YxdK]
yxjL	3993	two-component response regulator [YxjM]
yxjO	3991	transcriptional regulator (LysR family)
yyaG	4197	transcriptional regulator (LacI family)
yyaN	4189	transcriptional regulator (MerR family)
yybA	4183	transcriptional regulator (MarR family)
yybE	4180	transcriptional regulator (LysR family)
yycF	4154	two-component response regulator [YycG]
yydK	4122	transcriptional regulator (GntR family)

III.5.3 ELONGATION....8

greA	2791	transcription elongation factor
mfd	60	transcription-repair coupling factor
papS	2356	poly(A) polymerase
rpoA	149	RNA polymerase (α subunit)
rpoB	122	RNA polymerase (β subunit)
rpoC	126	RNA polymerase (β' subunit)
rpoE	3812	RNA polymerase (δ subunit)
yvrE	3406	RNA polymerase

III.5.4 TERMINATION....4

nusA	1732	transcription termination
nusG	118	transcription antitermination factor
rho	3804	transcriptional terminator Rho
yqhZ	2529	transcription termination

III.6 RNA MODIFICATION....19

cspR	970	rRNA methylase homolog
deaD	4016	ATP-dependent RNA helicase
miaA	1866	tRNA isopentenylpyrophosphate transferase
queA	2834	*S*-adenosylmethionine tRNA ribosyltransferase (queuosine biosynthesis)
rncS	1665	ribonuclease III
rnpA	4214	ribonuclease P (protein component)
rph	2901	ribonuclease PH
tgt	2833	tRNA-guanine transglycosylase (queuosine biosynthesis)
trmD	1675	tRNA methyltransferase
truA	153	pseudouridylate synthase I
truB	1736	tRNA pseudouridine 5S synthase
ydbR	511	ATP-dependent RNA helicase
yefA	737	RNA methyltransferase
yfjO	873	RNA methyltransferase
yfmL	816	RNA helicase
yloM	1647	RNA-binding Sun protein
yqfR	2595	ATP-dependent RNA helicase
ysgA	2931	rRNA methylase
yugI	3225	polyribonucleotide nucleotidyltransferase

III.7 PROTEIN SYNTHESIS....96

III.7.1 RIBOSOMAL PROTEINS....56

rplA	119	ribosomal protein L1 (BL1)
rplB	137	ribosomal protein L2 (BL2)
rplC	136	ribosomal protein L3 (BL3)
rplD	136	ribosomal protein L4
rplE	141	ribosomal protein L5 (BL6)
rplF	142	ribosomal protein L6 (BL8)
rplI	4163	ribosomal protein L9
rplJ	120	ribosomal protein L10 (BL5)
rplK	119	ribosomal protein L11 (BL11)
rplL	121	ribosomal protein L12 (BL9)
rplM	154	ribosomal protein L13
rplN	140	ribosomal protein L14
rplO	144	ribosomal protein L15
rplP	139	ribosomal protein L16
rplQ	150	ribosomal protein L17 (BL15)
rplR	143	ribosomal protein L18
rplS	1675	ribosomal protein L19
rplT	2952	ribosomal protein L20
rplU	2855	ribosomal protein L21 (BL20)
rplV	138	ribosomal protein L22 (BL17)
rplW	137	ribosomal protein L23
rplX	141	ribosomal protein L24 (BL23) (histone-like protein HPB12)
rpmA	2854	ribosomal protein L27 (BL24)
rpmB	1655	ribosomal protein L28
rpmC	140	ribosomal protein L29
rpmD	144	ribosomal protein L30 (BL27)
rpmE	3802	ribosomal protein L31
rpmF	1575	ribosomal protein L32
rpmG	117	ribosomal protein L33
rpmH	4215	ribosomal protein L34
rpmI	2952	ribosomal protein L35
rpmJ	148	ribosomal protein L36 (ribosomal protein B)
rpsB	1717	ribosomal protein S2
rpsC	139	ribosomal protein S3 (BS3)
rpsD	3035	ribosomal protein S4 (BS4)
rpsE	143	ribosomal protein S5
rpsF	4199	ribosomal protein S6 (BS9)
rpsG	130	ribosomal protein S7 (BS7)
rpsH	142	ribosomal protein S8 (BS8)
rpsI	154	ribosomal protein S9
rpsJ	135	ribosomal protein S10 (BS13)
rpsK	148	ribosomal protein S11 (BS11)
rpsL	130	ribosomal protein S12 (BS12)
rpsM	148	ribosomal protein S13
rpsN	142	ribosomal protein S14
rpsO	1738	ribosomal protein S15 (BS18)
rpsP	1673	ribosomal protein S16 (BS17)
rpsQ	140	ribosomal protein S17 (BS16)
rpsR	4198	ribosomal protein S18
rpsS	138	ribosomal protein S19 (BS19)
rpsT	2635	ribosomal protein S20 (BS20)
rpsU	2620	ribosomal protein S21
ybxF	129	ribosomal protein L7AE family
yhzA	965	ribosomal protein S14
ylxQ	1733	ribosomal protein L7AE family
yvyD	3631	ribosomal protein S30AE family

III.7.2 AMINOACYL-TRNA SYNTHETASES....25

alaS	2800	alanyl-tRNA synthetase
argS	3834	arginyl-tRNA synthetase
asnS	2347	asparaginyl-tRNA synthetase
aspS	2816	aspartyl-tRNA synthetase
cysS	113	cysteinyl-tRNA synthetase
gltX	111	glutamyl-tRNA synthetase
glyQ	2608	glycyl-tRNA synthetase (α subunit)
glyS	2607	glycyl-tRNA synthetase (β subunit)
hisS	2817	histidyl-tRNA synthetase
hisZ	3588	histidyl-tRNA synthetase
ileS	1613	isoleucyl-tRNA synthetase
leuS	3104	leucyl-tRNA synthetase
lysS	89	lysyl-tRNA synthetase
metS	46	methionyl-tRNA synthetase
pheS	2930	phenylalanyl-tRNA synthetase (α subunit)
pheT	2929	phenylalanyl-tRNA synthetase (β subunit)
proS	1725	prolyl-tRNA synthetase
serS	21	seryl-tRNA synthetase
thrS	2960	threonyl-tRNA synthetase (major)
thrZ	3855	threonyl-tRNA synthetase (minor)
trpS	1219	tryptophanyl-tRNA synthetase
tyrS	3037	tyrosyl-tRNA synthetase (major)
tyrZ	3946	tyrosyl-tRNA synthetase (minor)
valS	2869	valyl-tRNA synthetase
ytpR	3052	phenylalanyl-tRNA synthetase (β subunit)

III.7.3 INITIATION....6

fmt	1646	methionyl-tRNA formyltransferase
infA	148	initiation factor IF-1
infB	1733	initiation factor IF-2
infC	2952	initiation factor IF-3
rbfA	1736	ribosome-binding factor A
ykrS	1423	initiation factor eIF-2B (α subunit)

III.7.4 ELONGATION....6

efp	2538	elongation factor P
fus	131	elongation factor G

lepA 2632 GTP-binding protein
tsf 1718 elongation factor Ts
tufA 133 elongation factor Tu
ylaG 1546 GTP-binding elongation factor

III.7.5 TERMINATION 3
frr 1720 ribosome recycling factor
prfA 3797 peptide chain release factor 1
prfB 3627 peptide chain release factor 2

III.8 PROTEIN MODIFICATION 27
amhX 325 amidohydrolase
lgt 3593 prolipoprotein diacylglyceryl transferase (lipoprotein biosynthesis)
map 147 methionine aminopeptidase
pcp 287 pyrrolidone-carboxylate peptidase
ppiB 2435 peptidyl-prolyl isomerase
prkA 973 serine protein kinase
tgl 3212 transglutaminase
ybdM 224 protein kinase
ydiC 642 glycoprotein endopeptidase
ydiD 643 ribosomal-protein-alanine *N*-acetyltransferase
ydiE 643 glycoprotein endopeptidase
yfkJ 862 protein-tyrosine phosphatase
yflG 840 methionine aminopeptidase
yjcK 1261 ribosomal-protein-alanine *N*-acetyltransferase
ykrB 1526 formylmethionine deformylase
ykvY 1453 Xaa-Pro dipeptidase
yloP 1651 protein kinase
yppP 2287 peptide methionine sulfoxide reductase
yqeT 2624 ribosomal protein L11 methyltransferase
yqhT 2539 Xaa-Pro dipeptidase
ytel 3020 protease IV
ytjP 3068 Xaa-His dipeptidase
ytvA 3105 protein kinase
ytxM 3150 prolyl aminopeptidase
yuiE 3297 leucyl aminopeptidase
ywlE 3791 protein-tyrosine-phosphatase
yxaL 4102 serine/threonine protein kinase

III.9 PROTEIN FOLDING 8
dnaK 2627 class I heat-shock protein (chaperonin)
groEL 650 class I heat-shock protein (chaperonin)
groES 650 class I heat-shock protein (chaperonin)
tig 2887 trigger factor (prolyl isomerase)
ykkC 1376 chaperonin
ykkD 1376 chaperonin
yvdR 3541 chaperonin
yvdS 3541 chaperonin

IV OTHER FUNCTIONS 289

IV.1 ADAPTATION TO ATYPICAL CONDITIONS 72
bsaA 2304 glutathione peroxidase
clpC 104 class III stress response-related ATPase (repressor of competence)
clpE 1437 ATP-dependent Clp protease-like
clpP 3545 ATP-dependent Clp protease proteolytic subunit (class III heat-shock protein)
clpQ 1688 β-type subunit of the 20S proteasome
clpX 2885 ATP-dependent Clp protease ATP-binding subunit (class III heat-shock protein)
clpY 1688 ATP-dependent Clp protease-like
csbB 930 stress response protein
cspB 984 major cold-shock protein
cspC 559 cold-shock protein
cspD 2307 cold-shock protein
cstA 2937 carbon starvation-induced protein
ctc 59 general stress protein
degQ 3256 degradative enzyme production
degR 2308 degradative enzyme production
dnaJ 2625 heat-shock protein (activation of DnaK)
dps 3136 stress- and starvation-induced gene controlled by σ^B
gbsA 3186 glycine betaine aldehyde dehydrogenase (osmoprotection)
gbsB 3184 alcohol dehydrogenase (osmoprotection)
grpE 2628 heat-shock protein (activation of DnaK)
gsiB 494 general stress protein
gspA 3944 general stress protein
hit 1076 Hit-like protein involved in cell-cycle regulation
htpG 4090 class III heat-shock protein (chaperonin)
htrA 1359 serine protease Do (heat-shock protein)
ispU 1387 activation of σ^H
lonA 2882 class III heat-shock ATP-dependent protease
lonB 2884 Lon-like ATP-dependent protease
mrgA 3383 metalloregulation DNA-binding stress protein
rsbR 519 positive regulator of σ^B activity (interaction with RsbS)
rsbS 520 negative regulator of σ^B activity (antagonist of RsbT)
rsbT 520 positive regulator of σ^B activity (switch protein/serine kinase [RsbS])
rsbU 521 indirect positive regulator of σ^B activity (serine phosphatase [RsbV~P])
rsbV 522 positive regulator of σ^B activity (anti-anti-sigma factor [RsbW])
rsbW 522 negative regulator of σ^B activity (switch protein/serine kinase [RsbV], anti-sigma factor [σ^B])
rsbX 523 indirect negative regulator of σ^B activity (serine phosphatase [RsbS~P])
ycdH 308 adhesion protein
ydaG 473 general stress protein
yfiQ 910 surface adhesion
ykrL 1414 heat-shock protein
ykzA 1381 general stress protein
yloA 1637 fibronectin-binding protein
yloU 1655 alkaline-shock protein
ynbA 1875 GTP-binding protein protease modulator
ynzF 1880 δ-endotoxin
yocK 2097 general stress protein
yocM 2098 small heat-shock protein
yodU 2151 capsular polysaccharide biosynthesis
yokG 2279 δ-endotoxin
ypqP 2286 capsular polysaccharide biosynthesis
ytxG 3047 general stress protein
ytxH 3047 general stress protein
ytxJ 3046 general stress protein
yveK 3529 capsular polysaccharide biosynthesis
yveL 3528 capsular polysaccharide biosynthesis
yveM 3527 capsular polysaccharide biosynthesis
yveN 3525 capsular polysaccharide biosynthesis
yveO 3524 exopolysaccharide biosynthesis
yveP 3523 capsular polysaccharide biosynthesis
yveQ 3522 capsular polysaccharide biosynthesis
yveR 3521 spore coat polysaccharide biosynthesis
yveT 3519 capsular polysaccharide biosynthesis
yvfC 3517 capsular polysaccharide biosynthesis
yvfE 3515 spore coat polysaccharide biosynthesis
yvtB 3384 serine protease Do
ywqC 3732 capsular polysaccharide biosynthesis
ywqD 3732 capsular polysaccharide biosynthesis
ywqE 3731 capsular polysaccharide biosynthesis
ywsC 3700 capsular polyglutamate biosynthesis
ywtA 3698 capsular polyglutamate biosynthesis
ywtB 3698 capsular polyglutamate biosynthesis
yyxA 4148 serine protease Do

IV.2 DETOXIFICATION 68
aadK 2736 aminoglycoside 6-adenylyltransferase
ahpC 4118 alkyl hydroperoxide reductase (small subunit)
ahpF 4119 alkyl hydroperoxide reductase (large subunit) / NADH dehydrogenase
bmrU 2493 multidrug resistance protein cotranscribed with *bmr*
cah 342 cephalosporin C deacetylase
cypA 2732 cytochrome P450-like enzyme
cypX 3603 cytochrome P450-like enzyme
katA 960 vegetative catalase 1
katB 4009 catalase 2
katX 3964 catalase
ksgA 51 dimethyladenosine transferase (kasugamycin resistance)
mmr 3857 methylenomycin A resistance protein
padC 3532 ferulate decarboxylase
penP 2048 β-lactamase
pksS 1859 hydroxylase of the polyketide produced by the *pks* cluster
sodA 2585 superoxide dismutase
sodF 2103 superoxide dismutase
tetL 4188 tetracycline resistance leader peptide
thdF 4212 thiophen and furan oxidation
tmrB 339 tunicamycin resistance
yaaN 36 toxic cation resistance
ybbE 190 β-lactamase
ybfO 250 erythromycin esterase
ybxI 229 β-lactamase
ycbJ 276 viomycin phosphotransferase
ycbR 283 toxic cation resistance protein
yceC 312 tellurium resistance protein
yceD 312 tellurium resistance protein
yceE 313 tellurium resistance protein
yceF 314 tellurium resistance protein
yceH 316 toxic anion resistance protein
ycsF 457 lactam utilization protein
ydbD 496 manganese-containing catalase
ydfB 581 antibiotic resistance protein
ydhE 618 macrolide glycosyltransferase
yerP 732 acriflavin resistance protein
yetM 790 salicylate 1-monooxygenase
yetO 792 cytochrome P450 / NADPH-cytochrome P450 reductase
yflM 836 nitric-oxide synthase
yfnC 804 fosmidmycin resistance protein
ygaF 943 thiol-specific antioxidant protein
yhjG 1122 monooxygenase
yisY 1169 chloride peroxidase
yjiB 1291 monooxygenase
yjiC 1292 macrolide glycosyltransferase
ykfA 1366 immunity to bacteriotoxins
ykkB 1375 *N*-acetyltransferase
ykoY 1410 toxic anion resistance protein
yndN 1916 fosfomycin resistance protein
yocD 2088 immunity to bacteriotoxins
yojK 2117 macrolide glycosyltransferase
yojM 2115 superoxide dismutase
yokD 2281 aminoglycoside *N3'*-acetyltransferase
yqcM 2655 arsenate reductase
yqfP 2596 penicillin tolerance
yrhJ 2776 cytochrome P450 / NADPH-cytochrome P450 reductase
yrpB 2736 2-nitropropane dioxygenase
ytgI 3017 thiol peroxidase
ytnJ 3002 nitrilotriacetate monooxygenase
yubB 3195 bacitracin resistance protein (undecaprenol kinase)
yusI 3366 arsenate reductase
yvbT 3487 alkanal monooxygenase
yvdP 3543 reticuline oxidase
ywcH 3910 monooxygenase
ywnH 3760 phosphinothricin acetyltransferase
yxeI 4062 penicillin amidase
yxeK 4061 monooxygenase
yyaR 4185 streptothricine acetyl-transferase

IV.3 ANTIBIOTIC PRODUCTION 30
pksB 1782 involved in polyketide synthesis
pksC 1783 involved in polyketide synthesis
pksD 1785 involved in polyketide synthesis
pksE 1785 involved in polyketide synthesis
pksF 1788 involved in polyketide synthesis
pksG 1789 involved in polyketide synthesis
pksH 1790 involved in polyketide synthesis
pksI 1791 involved in polyketide synthesis
pksJ 1792 involved in polyketide synthesis
pksK 1794 polyketide synthase
pksL 1808 polyketide synthase
pksM 1821 polyketide synthase
pksN 1834 polyketide synthase
pksP 1835 polyketide synthase
pksR 1850 polyketide synthase
ppsA 1997 peptide synthetase
ppsB 1990 peptide synthetase
ppsC 1982 peptide synthetase
ppsD 1974 peptide synthetase
ppsE 1963 peptide synthetase
sbo 3835 subtilosin A
sfp 408 surfactin production
srfAA 377 surfactin synthetase / competence
srfAB 387 surfactin synthetase / competence
srfAC 398 surfactin synthetase / competence
srfAD 402 surfactin synthetase / competence
sunA 2269 sublancin 168 lantibiotic antimicrobial precursor peptide
yomB 2264 bacteriocin
yukL 3282 antibiotic synthetase
yukM 3283 antibiotic synthetase

IV.4 PHAGE-RELATED FUNCTIONS 83
codV 1687 integrase/recombinase
ripX 2449 integrase/recombinase
xhlA 1346 involved in cell lysis upon induction of PBSX
xhlB 1346 hydrolysis of 5-bromo 4-chloroindolyl phosphate upon induction of PBSX (holin)
xkdA 1320 PBSX prophage
xkdB 1321 PBSX prophage
xkdC 1322 PBSX prophage
xkdD 1323 PBSX prophage
xkdE 1327 PBSX prophage
xkdF 1328 PBSX prophage
xkdG 1329 PBSX prophage
xkdH 1330 PBSX prophage
xkdI 1331 PBSX prophage
xkdJ 1331 PBSX prophage
xkdK 1332 PBSX prophage
xkdM 1333 PBSX prophage
xkdN 1334 PBSX prophage
xkdO 1334 PBSX prophage
xkdP 1338 PBSX prophage
xkdQ 1339 PBSX prophage
xkdR 1340 PBSX prophage
xkdS 1340 PBSX prophage
xkdT 1341 PBSX prophage
xkdU 1342 PBSX prophage
xkdV 1343 PBSX prophage
xkdW 1345 PBSX prophage
xkdX 1345 PBSX prophage
xkdY 1345 PBSX prophage lytic exoenzyme
xtmA 1325 PBSX terminase (small subunit)
xtmB 1325 PBSX terminase (large subunit)
xtrA 1324 PBSX prophage
ycdD 304 L-alanoyl-D-glutamate peptidase
ydcL 530 integrase
ydcM 531 immunity region protein in prophage
yhgE 1090 phage infection protein
yjbJ 1235 lytic transglycosylase
yjqB 1318 phage-related replication protein
ymaC 1863 phage-related protein
ymaH 1867 host factor-1 protein
ymfD 1755 phage-related protein
ymfE 1756 phage-related protein
yndL 1914 phage-related replication protein
yobO 2075 phage-related pre-neck appendage protein
yokA 2284 DNA recombinase
yokL 2274 phage-related protein
yolB 2272 phage-related protein
yomA 2264 holin
yomJ 2248 phage-related immunity protein
yomP 2243 phage-related protein
yomR 2242 phage-related protein
yomS 2241 phage-related lytic exoenzyme
yoqD 2200 phage-related DNA-binding protein anti-repressor
yoqZ 2190 phage-related protein
yosQ 2160 phage-related endodeoxyribonuclease
yqaB 2700 phage-related protein
yqaJ 2696 phage-related protein
yqaK 2695 phage-related protein
yqaM 2694 phage-related protein
yqaO 2692 phage-related protein
yqaS 2690 phage-related terminase small subunit
yqaT 2689 phage-related terminase large subunit
yqbA 2688 phage-related protein
yqbD 2684 phage-related protein
yqbE 2683 phage-related protein
yqbH 2682 phage-related protein
yqbI 2681 phage-related protein
yqbJ 2681 phage-related protein
yqbK 2680 phage-related protein
yqbL 2679 phage-related protein
yqbM 2679 phage-related protein
yqbN 2677 phage-related protein
yqbO 2677 phage-related protein
yqbP 2672 phage-related protein
yqbQ 2671 phage-related protein
yqbR 2670 phage-related protein
yqbS 2670 phage-related protein
yqbT 2670 phage-related protein
yqcA 2669 phage-related protein
yqcC 2668 phage-related protein
yqcD 2667 phage-related protein
yqcE 2666 phage-related protein
yqxG 2666 phage-related lytic exoenzyme
yqxH 2665 holin

IV.5 TRANSPOSON AND IS 10
ydcP 533 transposon protein
ydcQ 533 transposon protein
ydcR 535 transposon protein
yddB 537 transposon protein
yddE 538 transposon protein
yddH 544 transposon protein
yefB 739 site-specific recombinase
yefC 739 resolvase
yneB 1918 resolvase
yocA 2085 transposon-related protein

IV.6 MISCELLANEOUS 26
bex 2610 GTP-binding protein
csbA 3614 putative membrane protein
csfB 36 σ^E-transcribed gene
ctaG 1564 function unknown
eag 1430 small membrane protein
ecsC 1079 function unknown
mmgE 2509 function unknown
nifZ 3027 NifS protein homologue
sapB 726 mutant activates alkaline phosphatase during sporulation independently of σ^F and σ^E
sbp 1595 small basic protein
veg 53 function unknown
yacI 102 creatine kinase
ybaL 157 ATP-binding Mrp-like protein
ycbU 287 NifS protein homologue
yerN 730 pet112-like protein
yhdP 1033 hemolysin
yhdT 1035 hemolysin
yheG 1049 calcium-binding protein
yplQ 2295 hemolysin III homologue
yqxC 2523 hemolysin-like
yrkA 2720 hemolysin-like
yrvO 2811 NifS protein homologue
yuaG 3181 epidermal surface antigen
yurV 3357 NifU protein homologue
yurW 3358 NifS protein homologue
yutI 3309 NifU protein homologue

V SIMILAR TO UNKNOWN PROTEINS 668

V.1 FROM *B. SUBTILIS* 177

V.2 FROM OTHER ORGANISMS 491

VI NO SIMILARITY 1,053