Essential *Bacillus subtilis* genes

K. Kobayashi[a], S. D. Ehrlich[b,c], A. Albertini[d], G. Amati[d], K. K. Andersen[e], M. Arnaud[f], K. Asai[g], S. Ashikaga[h], S. Aymerich[i], P. Bessieres[j], F. Boland[k], S. C. Brignell[l], S. Bron[m], K. Bunai[n], J. Chapuis[b], L. C. Christiansen[o], A. Danchin[p], M. Débarbouillé[f], E. Dervyn[b], E. Deuerling[q], K. Devine[e], S. K. Devine[e], O. Dreesen[p], J. Errington[r], S. Fillinger[i], S. J. Foster[k], Y. Fujita[s], A. Galizzi[d], R. Gardan[f], C. Eschevins[m], T. Fukushima[t], K. Haga[u], C. R. Harwood[l], M. Hecker[v], D. Hosoya[w], M. F. Hullo[p], H. Kakeshita[n], D. Karamata[x], Y. Kasahara[a], F. Kawamura[h], K. Koga[h], P. Koski[y], R. Kuwana[z], D. Imamura[w], M. Ishimaru[w], S. Ishikawa[t], I. Ishio[s], D. Le Coq[i], A. Masson[aa], C. Mauël[x], R. Meima[m], R. P. Mellado[bb], A. Moir[k], S. Moriya[a], E. Nagakawa[s], H. Nanamiya[h], S. Nakai[a], P. Nygaard[o], M. Ogura[cc], T. Ohanan[q], M. O'Reilly[e], M. O'Rourke[k], Z. Pragai[l], H. M. Pooley[x], G. Rapoport[f], J. P. Rawlins[r], L. A. Rivas[bb], C. Rivolta[x], A. Sadaie[u], Y. Sadaie[g], M. Sarvas[y], T. Sato[w], H. H. Saxild[o], E. Scanlan[e], W. Schumann[q], J. F. M. L. Seegers[aa], J. Sekiguchi[t], A. Sekowska[p], S. J. Séror[aa], M. Simon[dd], P. Stragier[dd], R. Studer[x], H. Takamatsu[z], T. Tanaka[cc], M. Takeuchi[w], H. B. Thomaides[r], V. Vagner[b], J. M. van Dijl[m], K. Watabe[z], A. Wipat[l], H. Yamamoto[t], M. Yamamoto[s], Y. Yamamoto[s], K. Yamane[n], K. Yata[ee], K. Yoshida[s], H. Yoshikawa[u], U. Zuber[v], and N. Ogasawara[a]

[a]Graduate School of Information Science, Nara Institute of Science and Technology, Nara 630-0101, Japan; [b]Génétique Microbienne, Institut National de la Recherche Agronomique, 78530 Jouy en Josas, France; [d]Genetica e Microbiologia, Università di Pavia, 1 via Ferrata, 27100 Pavia, Italy; [e]Genetics, Smurfit Institute, Trinity College, Dublin 2, Ireland; [f]Biochimie Microbienne, Institut Pasteur, 25 Rue du Dr. Roux, 75015 Paris, France; [g]Faculty of Science, Saitama University, Saitama 338-8570, Japan; [h]College of Science, Rikkyo (St. Paul's) University, Tokyo 171-8501, Japan; [i]Génétique Moléculaire et Cellulaire, Institut National de la Recherche Agronomique–Centre National de la Recherche Scientifique–Institut National Agronomique Paris-Grignon, 78850 Thiverval-Grignon, France; [j]Mathématiques Informatique Génomes, Institut National de la Recherche Agronomique, 78530 Jouy en Josas, France; [k]Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, United Kingdom; [l]Cell and Molecular Bioscience, Newcastle University Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, United Kingdom; [m]Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, 9750 AA, Haren, The Netherlands; [n]Institute of Biological Sciences, University of Tsukuba, Ibaraki 305-8572, Japan; [o]Biological Chemistry, Institute of Molecular Biology, Solvgade 83, 1307 K, Copenhagen, Denmark; [p]Genetique des Genomes Bacteriens, Institut Pasteur, Unité de Recherche Associée, Centre National de la Recherche Scientifique 2171, 75015 Paris, France; [q]Institute of Genetics, Bayreuth University, D-95440 Bayreuth, Germany; [r]Sir William Dunn School of Pathology, Oxford University, Oxford OX1 3RE, United Kingdom; [s]Faculty of Life Science and Biotechnology, Fukuyama University, Hiroshima 729-0292, Japan; [t]Faculty of Textile Science and Technology, Shinshu University, Nagano 386-8564, Japan; [u]Department of Bioscience, Tokyo University of Agriculture, Tokyo 156-8502, Japan; [v]Institute for Microbiology, Ernst-Moritz-Arndt-University, D-17487 Greifswald, Germany; [w]Department of International Environmental and Agricultural Science, Tokyo University of Agriculture and Technology, Tokyo 183-8509, Japan; [x]Institut de Genetique et de Biologie Microbiennes, CH-1005 Lausanne, Switzerland; [y]National Public Health Institute, 00300, Helsinki, Finland; [z]Faculty of Pharmaceutical Sciences, Setsunan University, Osaka 573-0101, Japan; [aa]Institut de Génétique et Microbiologie, Centre National de la Recherche Scientifique Unité Mixte de Recherche 8621, Université Paris-Sud, 91405 Orsay Cedex, France; [bb]Centro Nacional de Biotecnología, Campus de la Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain; [cc]School of Marine Science and Technology, University of Tokai, Shizuoka 424-8610, Japan; [dd]Institut de Biologie Physico-Chimique, 75005 Paris, France; and [ee]Radioisotope Center, National Institute of Genetics, Shizuoka 411-8540, Japan

Communicated by Richard M. Losick, Harvard University, Cambridge, MA, January 27, 2003 (received for review November 10, 2002)

To estimate the minimal gene set required to sustain bacterial life in nutritious conditions, we carried out a systematic inactivation of *Bacillus subtilis* genes. Among ≈4,100 genes of the organism, only 192 were shown to be indispensable by this or previous work. Another 79 genes were predicted to be essential. The vast majority of essential genes were categorized in relatively few domains of cell metabolism, with about half involved in information processing, one-fifth involved in the synthesis of cell envelope and the determination of cell shape and division, and one-tenth related to cell energetics. Only 4% of essential genes encode unknown functions. Most essential genes are present throughout a wide range of Bacteria, and almost 70% can also be found in Archaea and Eucarya. However, essential genes related to cell envelope, shape, division, and respiration tend to be lost from bacteria with small genomes. Unexpectedly, most genes involved in the Embden–Meyerhof–Parnas pathway are essential. Identification of unknown and unexpected essential genes opens research avenues to better understanding of processes that sustain bacterial life.

The definition of the minimal gene set required to sustain a living cell is of considerable interest. The functions specified by such a set are likely to provide a view of a ''minimal'' bacterial cell. Many functions should be essential in all cells and could be considered as a foundation of life itself. The determination of the range of essential functions in different cells should reveal possible solutions for sustaining life. Computational and experimental research has previously been carried out to define a minimal protein-encoding gene set. An upper-limit estimate of a minimal bacterial gene set was obtained from the sequence of the entire *Mycoplasma genitalium* genome, which contains only ≈480 genes (1). A computational approach, based on the assumption that essential genes are conserved in the genomes of *M. genitalium* and *Haemophilus influenzae*, led to a description of a smaller set of some 260 genes (2). More recently, an experimental approach involving high-density transposon mutagenesis of the *H. influenzae* genome led to a much higher estimate of ≈670 putative essential genes (3), whereas transposon mutagenesis of two mycoplasma species led to an estimate of 265–360 essential genes (4). Another experimental approach using antisense RNA to inhibit gene expression led to the identification of some 150 essential genes in *Staphylococcus aureus* (5). However, these approaches have limitations. Computation is likely to underestimate the minimal gene set because it takes into account only those genes that have remained similar enough during the course of evolution to be recognized as true orthologues. Transposon mutagenesis might overestimate the set by misclassification of nonessential genes that slow down the growth without arresting it but can also miss essential genes that tolerate transposon insertions (3, 6). Finally, the use of antisense RNA is limited to the genes for which an adequate expression of the inhibitory RNA can be obtained in the organism under study.

To obtain an independent and possibly more reliable estimate of a minimal protein-encoding gene set for bacteria, we systematically inactivated *Bacillus subtilis* genes. *B. subtilis* was chosen because it is one of the best studied bacteria (7) and is a model for low-G+C Gram-positive bacteria, which include both deadly pathogens, such as *Bacillus anthracis*, and bacteria widely used in food and industry, such as lactococci and bacilli. Because the essentiality of a gene depends on the conditions under which the organism is propagated, we used an environment likely to be optimal for *B. subtilis* and thus carried out inactivation on a

[c]To whom correspondence should be addressed. E-mail: ehrlich@jouy.inra.fr and ehrlich@is.aist-nara.ac.jp.

 www.pnas.org/cgi/doi/10.1073/pnas.0730515100

Table 1. Essential and nonessential *B. subtilis* genes

	Essential	Nonessential	Total
This study*	150	2,807	2,957
Previous studies†	42	614	656
Prediction‡	79	106	185
Phage genes	0	303	303
Total§	271 (6.6%)	3,830 (94.4%)	4,101

A list of the genes and their classifications can be accessed at http://bacillus.genome.ad.jp.

*We included 18 essential genes here that were inactivated in the course of this study and also studied previously.

†Carried out in *B. subtilis.*

‡Full list is presented as Table 3.

§Excluded are four genes that were not studied because of technical reasons (too short for insertional inactivation and too inconveniently placed for chloramphenicol replacement).

standard laboratory rich medium at 37°C. This choice also allowed for a comparison of the results obtained in many laboratories and many previous studies, nevertheless leaving open the possibility that a different gene set is essential under different growth conditions. Analysis of the mutants, in conjunction with the literature data, leads us to conclude that there are only 271 genes indispensable for growth in LB when inactivated singly. These fall into a relatively few large domains of cell physiology and are very broadly conserved in microorganisms.

Methods

The approach used for gene inactivation has been described (8). Briefly, it involved insertion of a nonreplicating plasmid into the target gene via a single crossover recombination. The expression of the downstream genes from the same operon was controlled by an isopropyl β-D-thiogalactoside (IPTG)-regulated promoter present on the inserted plasmid. A gene was deemed essential if it could not be inactivated by insertion (i.e., no transformants were obtained when competent recipient cells were mixed with the insertional plasmid) and if the strain became IPTG dependent when an intact copy of the gene was placed under control of the regulated promoter (8). IPTG-dependent strains could not be constructed for six essential genes, possibly because the regulated promoter was either not strong enough or not sufficiently tuned to provide appropriate gene expression levels. An alternative strategy was followed for ≈160 genes shorter than 300 bp, where insertional inactivation was limited by the insufficient gene length. These genes were replaced by a chloramphenicol resistance marker, and if replacement failed they were rendered IPTG-dependent. All mutations were made in the standard laboratory strain 168. Inactivation was not attempted for 656 genes studied previously in *B. subtilis*, and 185 genes having a high degree of similarity with genes well characterized in other bacteria or involved in well characterized processes, for which we could predict essentiality with confidence (Table 3, which is published as supporting information on the PNAS web site, www.pnas.org). Complete microbial genomes included in the Microbial Genome Database for Comparative Analysis (http://mbgd.genome.ad.jp/), comprising 54 bacteria, 16 archaea, and 2 yeasts, were analyzed for the presence of the *B. subtilis* essential gene homologs by using the default parameters, with 10^{-3} as a cut-off value.

Results

There are ≈4,100 annotated genes in the *B. subtilis* genome (9). Some 303 are encoded on prophages that can be eliminated from the genome and are not essential. Previous studies on 656 *B. subtilis* genes identified 42 that are essential (Table 1). Through predictions we propose that 79 other genes are essential, whereas 106 are not (Table 3). We inactivated all but 4 of the remaining genes and found that 150 are essential. This analysis leads us to conclude that there are 271 genes indispensable for growth when inactivated singly (Table 1). For ≈96% of these, we propose assignment to various domains of cell metabolism (Table 2; the complete list of genes is given in Table 4, which is published as supporting information on the PNAS web site).

Table 2. *B. subtilis* essential genes

DNA metabolism	27
Basic replication machinery	16
Packaging and segregation	9
Methylation	2
RNA metabolism	14
Basic transcription machinery	4
RNA modification	6
Regulation	4
Protein synthesis	95
Ribosomal proteins	52
Aminoacyl-tRNA synthetases	24
Translation factors	10
Protein folding and modification	3
Protein translocation	6
Cell envelope	44
Membrane lipids	16
Cell wall	28
Cell shape and division	10
Glycolysis	8
Respiratory pathways	22
Isoprenoids	8
Menaquinone	8
Cytochrome biogenesis	3
Thioredoxin	3
Nucleotides	10
Cofactors	15
CoA	1
Folate	3
NAD	4
S-Adenosylmethionine	1
Iron–sulfur cluster	6
Other	15
Unknown	11
Total	271

A complete list of genes and the evidence used to ascertain their essential nature are presented in Table 4.

Functional Assignment of Essential Genes. ***Information processing.*** About half of the essential genes are involved in DNA and RNA metabolism and protein synthesis. Sixteen genes encode the basic DNA replication machinery. They comprise five genes involved in the initiation of replication (*dnaA*, *B*, *D*, and *I*, and *priA*), eight genes encoding components of the replisome (*dnaC*, *E*, *G*, *N*, and *X*, *holA* and *B*, and *polC*), DNA ligase, and the Ssb protein. One gene, *pcrA*, has no clearly identified role, but could be involved in the progression of the replication fork (10). Among genes involved in DNA packaging and segregation, five encode topoisomerases (*topA*, *gyrA* and *B*, and *parD* and *E*), one encodes the general DNA-binding protein Hbsu, and three encode the proteins that act in the condensation of the nucleoid (*smc*, and *scpA* and *B*; ref. 11). The remaining two genes encode modification methylases, expected to be essential unless the cognate nucleases are inactivated.

Among 14 essential genes involved in RNA metabolism, four (*rpoA*, *B*, and *C*, and *sigA*) encode components of the basic transcription machinery, whereas six are involved in RNA modification. *rnc* and *rnpA* encode RNases, *cspR* and *trmD* and

GENETICS

U encode methylases, and *cca* encodes tRNA nucleotidyl transferase. Only four genes are involved in regulation of RNA synthesis: a two-component system *yycF* and *G* (12), a gene involved in the coupling between translation and termination of RNA synthesis, *nusA* (13), and an anti-sigma factor, YhdL (14).

The largest category, comprising 95 essential genes, is that involved in protein synthesis. Over half of the genes encode ribosomal proteins. Although there is no experimental evidence that they are essential in *B. subtilis*, we suggest that they belong to the essential set, because the ribosome itself is essential. This suggestion is supported by the observation that the inhibition of synthesis of 21 different ribosomal proteins is lethal in *S. aureus* (5). Among these are proteins such as L24, which was not absolutely essential in *E. coli*, but cells that lacked it grew very slowly and were thermosensitive (15). We suggest that there are 20 essential genes that encode aminoacyl-tRNA synthetases, corresponding to 18 amino acids. All but two are present in unique copies. We showed that one of the unique copy genes, *lysS*, is essential and assumed that others are too, without seeking further experimental evidence. There are two genes encoding tRNA-Tyr and tRNA-Thr synthetases. Only *tyrS* was essential when inactivated singly whereas either *thrS* or *thrZ* could assure the viability. We grouped with the synthetases three genes that are required for the conversion of the tRNA-Glu to tRNA-Gln (*gatC*, *B*, and *A*) and one gene that is required for the formylation of methionyl tRNA (*fmt*). Of the10 essential genes involved in mRNA translation, 3 are required for initiation (*infA*, *B*, and *C*), 3 are required for elongation (*tufA*, *tsf*, and *fusA*), and 4 are required for termination and ribosome recycling (*prfA* and *B*, *pth*, and *frr*). There is one essential gene involved in posttranslation modification, *map*, that encodes methionine aminopeptidase. Deformylation is also required, but can be carried out by products of two genes, *def* and *ykrB*, neither of which is essential when inactivated singly (16). Two essential genes, *groEL* and *ES*, are involved in protein folding. Finally, there are six essential genes that encode key components of the machinery for protein insertion into the membrane and secretion. These include the targeting factors Ffh and FtsY, the translocation motor SecA, two components of the translocation channel, SecY and E, and the folding catalyst PrsA. The essential DNA-binding protein Hbsu is also a part of the signal recognition particle (17).

Cell envelope, shape, and division. About one-fifth of the essential genes are required for these processes (Table 2). The synthesis of the cell envelope involves 44 essential genes, all required for membrane and cell wall formation. Membrane lipids, phospholipids, and glycolipids are synthesized from fatty acids. Fatty acid synthesis (Fig. 4, which is published as supporting information on the PNAS web site) is initiated by products of four genes, *accA*, *B*, *C*, and *D*, together with *acpA* and *fabD* gene products. *acpS* is required for the conversion of AcpA from the apo to the holo form, whereas *birA* is required for the addition of a biotinyl group to carboxylase. The fatty acid chains are elongated by the products of two essential genes, *fabFG*. The elongation cycle involves two additional steps that are catalyzed by pairs of genes with overlapping functions (*ycsD* and *ywpB*, and *fabI* and *L*), none of which is essential when inactivated singly (18). Two of the essential genes required for phospholipid synthesis (Fig. 5, which is published as supporting information on the PNAS web site), *gpsA* and *yhdO*, are involved in the conversion of dihydroxyacetone phosphate to phosphatidic acid, which is a precursor of complex lipids. Interestingly, *yerQ*, which encodes an enzyme with a diacylglycerol kinase catalytic domain found in eukaryotes and presumably catalyses synthesis of phosphatidic acid from another precursor (diacylglycerol), is also essential, whereas a homologue, *dgkA*, is not. Two essential genes, *cdsA* and *pgsA*, are required for synthesis of phosphatidylglycerol phosphate, which might be converted into phosphoglycerol by a nonspecific phosphatase. The remaining essential gene, *plsX*, appears to be required for both fatty acid and phospholipid biosynthesis in a way that is not well understood (19).

Synthesis of peptidoglycan, the main component of the cell wall, comprises two stages, the synthesis of the precursor molecules and the polymerization of peptidoglycan (20). All of the essential genes are involved in the first stage, which encompasses a variety of biosynthetic pathways: (*i*) Synthesis of aminosugars (Fig. 6, which is published as supporting information on the PNAS web site) by conversion of fructose-6-phosphate to UDP-*N*-acetyl-glucosamine and UDP-*N*-acetyl-manosamine. The first two steps, leading to glucosamine-1-phosphate, are catalyzed by the products of *glmS* and *ybbT* genes. The last two steps are carried out by the products of the *gcaD* and *yvyH*. More than one gene product seems to be able to acetylate glucosamine-1-phosphate, because there is no single essential gene for this step. (*ii*) Diaminopimelate (Fig. 7, which is published as supporting information on the PNAS web site) is synthesized from L-aspartate by eight successive reactions, six of which are carried out by products of essential genes *asd*, *dapA*, *B*, and *F*, and *ykuQ* and *R*. The first and the fifth step can be catalyzed by products of three (*dapG*, *lysC*, and *yclM*) and two genes (*mtnV* and *ywfG*), respectively; thus, none of the five is essential if inactivated singly. (*iii*) Two essential genes, *racE* and *alr*, encode racemases that convert L-glutamate and L-alanine into the corresponding D isomers. *racE* cannot be replaced by a homologue, *yrpC*. The essential *ddl* gene is required for synthesis of the dipeptide D-Ala-D-Ala. (*iv*) Eight essential genes, *murAA*, *murB*, *C*, *D*, *E*, *F*, and *G*, and *mraY*, are required for synthesis of the lipid-linked disaccharide-pentapeptide peptidoglycan precursor (Fig. 8, which is published as supporting information on the PNAS web site) from UDP-*N*-acetyl-glucosamine, phospho*enol*pyruvate, D-glutamine, diaminopimelate, D-ala dipeptide, and an isoprenylphosphate. Polymerization of peptidoglycan is carried out by the products of functionally redundant genes in *B. subtilis*. The cell wall of *B. subtilis* contains teichoic acid (21), and there are seven essential genes involved in its synthesis. Four, *tagA*, *B*, *D*, and *O*, are required for the synthesis of linkage units and three, *tagF*, *G*, and *H*, are required for chain polymerization, translocation, and linkage to peptidoglycan (Fig. 9, which is published as supporting information on the PNAS web site).

Ten essential genes are involved in cell shape and division. Septum formation requires seven (*ftsA*, *L*, *W*, and *Z*, *divIB* and *C*, and *pbpB*; ref. 21), whereas cell shape requires three (*rodA*, and *mreB* and *C*).

Embden–Meyerhof–Parnas (EMP) pathway and respiration. About 10% of essential genes, which have in common the provision of energy for the cell, are required for these processes. A majority of genes composing the ubiquitous EMP pathway are essential (Fig. 10, which is published as supporting information on the PNAS web site). The process can be viewed as consisting of two parts: the top, which converts hexose sugars to trioses, and the bottom, which converts these compounds to pyruvate, funneled into pyruvate dehydrogenase. The top part comprises four steps when glucose is the carbon source, the last two of which are catalyzed by products of essential genes *pfkA* and *fbaA*, whereas the bottom part comprises six steps, four of which are encoded by essential genes *tpiA*, *pgk*, *pgm*, and *eno*. The two remaining essential genes related to glycolysis are *tkt* and *prs*. The first encodes a transketolase, involved in the pentose pathway, whereas the second gene codes for a pyrophosphokinase that converts ribose-5-phosphate to 5-phospho-ribose-1-diphosphate, a common precursor of nucleotides and cofactors, such as NAD, which likely accounts for its essential role. Taken together, these results are rather unexpected. First, our experiments were carried out on a rich medium, which contains numerous compounds that could provide the energy and building blocks for cell life, the two known functions of the EMP pathway. Addition of glucose to LB did not restore growth of any of the nonviable EMP mutants. Second, in *B.*

subtilis a part of the EMP pathway can be bypassed via the pentose shunt, and it is surprising that both are simultaneously required for viability. Possibly, the enzymes revealed as essential have novel and unexpected functions in the cell. It should be noted that *pgm* and *eno* mutants have been isolated previously and had very slow growth (22), suggesting that the difference between lethal and almost-lethal mutation can be due to subtle differences in the experimental conditions and the strain background.

Respiration can provide energy for the cell, in the absence of glycolysis. We identified 22 essential genes involved in this process. Under the aerobic condition used in our experiments, respiration involves the transfer of electrons by various dehydrogenases to menaquinone and then to cytochromes (23). Menaquinone is synthesized from chorismate in seven steps, the last six of which are catalyzed by products of essential genes, *menA*, *B*, *C*, *D*, *E*, and *H* (Fig. 11, which is published as supporting information on the PNAS web site). Two genes, *menF* and *dhbC*, appear to be able to catalyze the first step, and neither is essential if inactivated singly. The penultimate step involves condensation of dihydroxynaphthoic acid with an isoprenoid biphosphate. Isoprenoids (Fig. 12, which is published as supporting information on the PNAS web site) are synthesized from pyruvate and glyceraldehyde-3-phosphate by a nonmevalonate pathway in *B. subtilis*. The first six steps, leading to isopentenyl diphosphate, involve seven essential genes, *dxs*, *dxr*, *ispE*, *yacM* and *N*, and *yqfP* and *Y*. Three other essential genes, *hepS* and *T* and *yqiD*, are required for the synthesis of farnesyl diphosphate and more complex compounds that are used for menaquinone synthesis. Altogether, of 22 essential genes involved in respiration, 16 are required for menaquinone synthesis. There are only three essential genes involved in cytochrome biogenesis, *resA*, *B*, and *C*. No cytochrome structural genes are essential, possibly reflecting overlapping functions of their products (24). We have included *trxA* and *B*, which encode thioredoxin and thioredoxin reductase with the respiration genes, because of the role of TrxA in electron transport, although this protein is involved in many other oxido-reduction reactions. We also included here a putative thioredoxin reductase gene, *yumC*.

Nucleotides and cofactors. Metabolism of these compounds requires ≈10% of the essential genes (Table 2). The metabolism of nucleotides is quite complex, comprising complementary *de novo* synthesis and salvage pathways (25). Nevertheless, we found 10 essential genes involved in this process. Among the four that participate in purine metabolism (Fig. 13, which is published as supporting information on the PNAS web site), two (*adk* and *gmk*) specify kinases, which phosphorylate AMP or GMP to the respective diphosphates. Absence of guanine from the medium accounts for the essential nature of *guaB*. Surprisingly, *hprT*, a gene from the purine salvage, is also essential, raising a possibility that its product has a second, unsuspected role in the cell. Two essential genes involved in pyrimidine metabolism (Fig. 14, which is published as supporting information on the PNAS web site), *cmk* and *tmk*, also encode kinases that phosphorylate CMP and TMP to corresponding diphosphates. The remaining essential gene, *pyrG*, encodes cytidylate synthetase, which converts UTP into CTP. This might reflect the paucity of cytidine in the rich medium. Interestingly, two *B. subtilis* essential genes encode enzymes present in the *E. coli* degradosome [*yjbN* (*ppnK*) and *eno*, a member of the EMP pathway], which provides CDP for DNA synthesis and further nucleotide metabolism, while controlling mRNA turnover (26). Finally, there are three essential genes involved simultaneously in purine and pyrimidine metabolism, *nrdE* and *F* and *ymaA*, that encode subunits of nucleoside-diphosphate reductase, which converts the ribose into deoxyribose derivatives.

Synthesis of only five cofactors, involving 16 genes, was required under our experimental conditions. NAD synthesis can take place *de novo* or by salvaging of precursors (Fig. 15, which is published as supporting information on the PNAS web site), and only the four genes involved in the salvage pathway (*yueK*, *yqeJ*, *nadE*, and *yjbN*) were essential. We speculate that the accumulation of nicotinate might repress *de novo* synthesis of nicotine mononucleotide in the absence of *yueK*, rendering this gene essential. There are three essential genes involved in folate metabolism (Fig. 16, which is published as supporting information on the PNAS web site). One, *dfrA*, codes for dihydrofolate reductase, which converts folate, presumably imported from the medium, to tetrahydrofolate. Two other genes, *glyA* and *folD*, are required for conversion of the latter compound to 10-formyl tetrahydrofolate, a one-carbon donor molecule for a number of reactions. *S*-adenosylmethionine (SAM) is another one-carbon donor, synthesized from ATP and methionine by SAM synthetase, encoded by the essential *metK* gene. There is only one essential gene involved in the biosynthesis of CoA, *ytaG*, that is required for the last step in the pathway (Fig. 17, which is published as supporting information on the PNAS web site), suggesting that the precursor, dephospho-CoA, is transported from the medium into the cell. The remaining cofactor is an iron–sulfur cluster, which forms part of proteins that participate in many aspects of the cell physiology, including redox and nonredox catalysis, as well as sensing for regulatory processes. There are five essential genes, *yurU*, *V*, *W*, *X*, and *Z*, involved in the synthesis of this cluster. We included here *yrvO*, a homologue of *yurV*.

Other processes. Only 15 essential genes that have a clear biochemical function were not associated with any of the large domains of cellular physiology discussed above. Among these are six GTP-binding proteins of the Era/Obg family. Only one, *obg*, has been studied previously in *B. subtilis* and been shown to affect the stress response mediated by σ^B. Five other genes, *mrpA*, *B*, *C*, *D* and *F*, encode a sodium–hydrogen antiporter, which is required to maintain pH homeostasis in the presence of sodium chloride concentrations similar to those found in LB (27). *ppaC* encodes the inorganic pyrophosphatase, which drives the anabolic fluxes by pyrophosphate hydrolysis in various biochemical reactions, whereas *gcp* encodes a sialopeptidase of

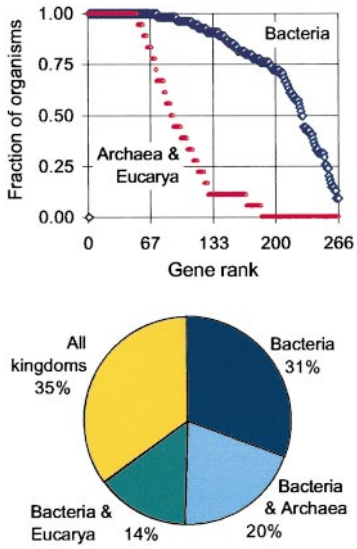

Fig. 1. *B. subtilis* essential gene homologues are widely conserved. (*Upper*) Genes are ordered by their relative abundance among 54 Bacteria (blue) and 18 Archaea and Eucarya (red). The position (rank) of a gene is shown on abscissa and the fraction of organisms in which a gene is present is shown on the ordinate. (*Lower*) Fraction of genes present in different kingdoms of life (a gene counted as "all kingdoms" is present in at least one archaeon and one eukaryote, in addition to bacteria, whereas a gene counted as "bacteria" is not present in any archae or eukaryote). The list of genes and organisms is presented in Table 4.

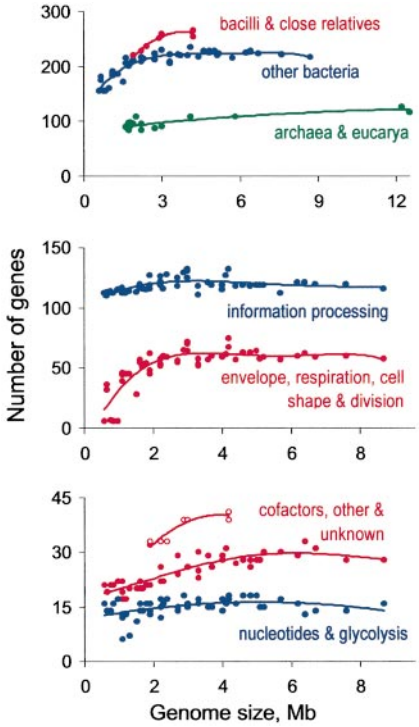

Fig. 2. The number of *B. subtilis* essential gene homologues depends on genome size. (*Top*) All genes. Bacilli and close relatives denote *Bacillus* species and other low-G+C Gram-positive bacteria, but not clostridia, mycoplasma, and ureaplasma. (*Middle* and *Bottom*) Different bacterial gene categories. Empty red circles in *Bottom* refer to Bacilli and close relatives, whereas filled red circles refer to other bacteria. Interpolated lines throughout the figure correspond to the best fitting polynomial of the second or the fourth order. The number of genes is: information processing, 136; envelope, respiration, cell shape, and division, 76; cofactors, other, and unknown, 41; and nucleotides and glycolysis, 18.

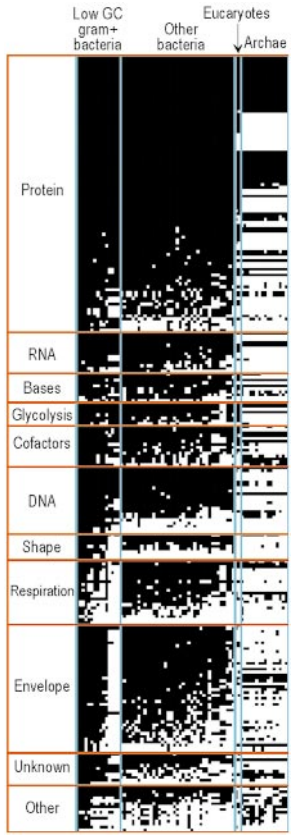

Fig. 3. Phylogenetic profiling of essential genes. The 271 *B. subtilis* genes were grouped in 266 clusters. Only one gene, *yhdL*, which encodes a possible anti-sigma protein, had no orthologues in the database and is not presented here. Each line and column corresponds to individual gene and organism, respectively. Presence and absence of a gene is indicated by a black and white square, respectively. The list of genes and organisms is given in Table 5, which is published as supporting information on the PNAS web site and the ordering is described in the text.

unknown role. The last two genes, *pdhA* and *odh*, encode subunits of pyruvate and 2-oxoglutarate dehydrogenase, respectively; growth of the mutants could be restored by addition to LB of the metabolites (acetate and succinate, respectively) related to the activity of the proteins they encode.

Unknown. The last category groups 11 essential genes for which we were unable to suggest a role in cell physiology. Biochemical functions, a protease and a hydrolase of the metallo-β-lactamase superfamily, can be suggested for products of two gene, *ydiC* and *ykqC*. One gene, *yneS*, encodes a putative membrane protein, and another, *ymdA*, encodes a protein with an HD domain of metal-dependent phosphohydrolases, whereas three, *yloQ*, *yqjK*, and *ywlC*, encode proteins with recognizable signatures, an ATP/GTP-binding site, a metallo-β-lactamase motif, and a putative RNA-binding motif, respectively. Four genes, *yacA*, *ydiB*, *ylaN*, and *yqeI*, have no easily recognizable features.

Conservation of Essential Genes. The average level at which homologues of essential *B. subtilis* genes are present in bacteria is rather high (approaching 80%), one-fourth being found in all bacteria and three-fourths in at least 75% (Fig. 1 *Upper*). The average is ≈36% in Eucarya and Archaea, but some 20% of the genes are nevertheless present in all 18 organisms we analyzed (Fig. 1 *Upper*). About one-third of the genes are found in all three kingdoms of life, and a further one-third are shared between Bacteria and either Archaea or Eucarya (Fig. 1 *Lower*).

The number of *B. subtilis* essential gene homologues present in an organism depends on at least two parameters: phylogenetic proximity to *B. subtilis* and genome size (Fig. 2 *Top*). The highest number is found in bacilli and close relatives, having genomes of >3 Mb (highlighted in red). Other bacteria with genomes of a similar size have, on average, slightly >80% of the *B. subtilis* essential gene homologues. This proportion drops to 57% with decreasing bacterial genome size, indicating progressive loss of essential genes. Archaea and Eucarya maintain, on average, 36% of the essential gene homologues, with the proportion varying between 33% and 44% almost linearly with genome size. In bacteria, gene loss occurs mainly in three categories (cell envelope, shape and division, and respiratory pathways) and to a lower extent in three other categories (cofactor synthesis, other processes, and unknown functions). In contrast, information processing, glycolysis, and nucleotide synthesis genes are largely retained (Fig. 2 *Middle* and *Bottom*).

Phylogenetic profiling of essential *B. subtilis* genes is summarized in Fig. 3. Organisms were grouped into four classes and ordered within each class on the basis of the number of essential gene homologues they share with *B. subtilis*, placing the organisms with fewest conserved genes at the right of each class. Genes were grouped in categories and ordered by abundance among all bacteria, which placed the less abundant genes at the bottom of each category. A number of general features are easily discernible from this analysis. (*i*) The five top categories are composed of genes present in >80% of Bacteria and at least 40% of Eucarya and Archaea, with the exception of RNA synthesis, which is less well conserved in the last two kingdoms. (*ii*) The next two categories, DNA metabolism and cell shape and division, contain genes

present in most bacteria and genes specific for Gram-positive organisms. This can most easily be seen from the appearance of the relatively broad horizontal white bars at the bottom of the two classes. (*iii*) The categories that contain genes missing from bacteria with small genomes are easily identified by the presence of the vertical white band at the right of the low-G+C Gram-positive bacteria class, corresponding to *Mycoplasma* and *Ureaplasma urealyticum*. In addition, there is an enlargement of the white zone at the right end of the "Other bacteria" class, noticeable for cell envelope, respiration, and unknown functions. (*iv*) Genes in the last two categories, unknown and other, although often found only in the closest relatives of *B. subtilis*, are nevertheless present in over a half of other bacteria.

Discussion

A Simple Bacterial Cell. Of some 4,100 genes of *B. subtilis*, only 271 are essential for growth under our experimental conditions when inactivated singly. About 80% of the functions they encode fall in a few large categories; namely, information processing, cell envelope, shape, division, and energetics. These observations lead to a view of a rather simple bacterial cell, consisting of a compartment, formed by a membrane and a wall, enclosing the elements necessary to synthesize proteins that carry out reactions required for (*i*) the duplication and inheritance of the genetic information; (*ii*) the division of the compartment; and (*iii*) the provision of energy. These processes do not appear to be coordinated by modulation of gene expression, because the expression regulators are by and large not essential. We suggest that the coordination might be carried out, at least in part, by the essential GTP-binding proteins, as appears to be the case in eukaryotes.

Broad Distribution of Essential Genes and Functions. Over 80% of essential *B. subtilis* gene homologues are present in all bacteria with genomes above ≈3 Mb, and 57% are found even in bacteria with the smallest genomes (mycoplasma). Almost 70% of genes are present in at least one kingdom other than Bacteria. Many organisms thus appear to rely on a similar set of essential functions, supporting the simple microbial cell view outlined above. The similarity might be even higher, because some of the genes might have diverged beyond recognition and some functions can be encoded by unrelated genes (28). However, genes involved in the synthesis of the cell envelope tend to be lost from bacteria with smaller genomes. Concomitantly, genes involved in the determination of cell shape, division, and respiration are also lost. This suggests that it may be possible to build, maintain, and reproduce the cell compartment in a simpler way than that used by bacteria with larger genomes, and that glycolysis can be sufficient to generate energy for the cell. A minimal essential gene set could thus be significantly smaller than the one present in bacteria with genomes larger than ≈3 Mb.

Unexpected Essential Genes. Notwithstanding the grouping of most essential functions in a few large categories, our study has revealed genes that were not expected to have an essential function under the experimental conditions used, such as eight EMP pathway genes and a gene involved in purine biosynthesis. These observations suggest previously unsuspected links between different domains of cell physiology.

Redundant Genes for Essential Functions. Our analysis does not detect essential functions encoded by redundant genes, because only a single gene was inactivated in each mutant strain. The list of the essential genes given here is thus likely to be underestimated, because synthetic lethal mutants are well known. A rigorous detection of the missing functions would require the systematic combination of all of the mutations in a single strain, which is beyond the present genetic technology. However, it is remarkable that single gene inactivation did reveal large categories of essential functions, suggesting that most of the vital cell processes are encoded by nonredundant genes. The presence of paralogues for ≈50% of *B. subtilis* genes (9) might thus allow the cell to respond to changing environmental conditions rather than provide back-up for vital processes.

Isogenic Mutant Collection. Finally, it should be noted that the isogenic set of ≈3,000 mutants that we have generated can be used to identify genes, and thus functions, that are essential under conditions different from those used here. Furthermore, the mutant set is a unique bacterial resource for studying various phenotypes and may thus lead to deeper insight into the metabolism of the bacterial cell.

This work was supported, in part, by European Union Grant BIO4-CT95-0278 and a Grant-in-Aid for Scientific Research on Priority Areas (C) "Genome Biology" from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

1. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., *et al.* (1995) *Science* **270,** 397–403.
2. Mushegian, A. R. & Koonin, E. V. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 10268–10273.
3. Akerley, B. J., Rubin, E. J., Novick, V. L., Amaya, K., Judson, N. & Mekalanos, J. J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 966–971.
4. Hutchison, C. A., Peterson, S. N., Gill, S. R., Cline, R. T., White, O., Fraser, C. M., Smith, H. O. & Venter, J. C. (1999) *Science* **286,** 2165–2169.
5. Ji, Y., Zhang, B., Van Horn, S. F., Warren, P., Woodnutt, G., Burnham, M. K. & Rosenberg, M. (2001) *Science* **293,** 2266–2269.
6. Gerdes, S. Y., Scholle, M. D., D'Souza, M., Bernal, A., Baev, M. V., Farrell, M., Kurnasov, O. V., Daugherty, M. D., Mseeh, F., Polanuyer, B. M., *et al.* (2002) *J. Bacteriol.* **184,** 4555–4572.
7. Sonenshein, A. L., Hoch, J. A. & Losick, R., eds. (2002) *Bacillus subtilis and Its Closest Relatives: From Genes to Cells* (Am. Soc. Microbiol., Washington, DC).
8. Vagner, V., Dervyn, E. & Ehrlich, S. D. (1998) *Microbiology* **144,** 3097–3104.
9. Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertero, M. G., Bessieres, P., Bolotin, A., Borchert, S., *et al.* (1997) *Nature* **390,** 249–256.
10. Petit, M. A. & Ehrlich, S. D. (2002) *EMBO J.* **21,** 3137–3147.
11. Soppa, J., Kobayashi, K., Noirot-Gros, M. F., Oesterhelt, D., Ehrlich, S. D., Dervyn, E., Ogasawara, N. & Moriya, S. (2002) *Mol. Microbiol.* **45,** 59–71.
12. Fabret, C. & Hoch, J. A. (1998) *J. Bacteriol.* **180,** 6375–6383.
13. Ingham, C. J., Dennis, J. & Furneaux, P. A. (1999) *Mol. Microbiol.* **31,** 651–663.
14. Horsburgh, M. J. & Moir, A. (1999) *Mol. Microbiol.* **32,** 41–50.
15. Nishi, K., Dabbs, E. R. & Schnier, J. (1985) *J. Bacteriol.* **163,** 890–894.
16. Haas, M., Beyer, D., Gahlmann, R. & Freiberg, C. (2001) *Microbiology* **147,** 1783–1791.
17. Nakamura, K., Yahagi, S., Yamazaki, T. & Yamane, K. (1999) *J. Biol. Chem.* **274,** 13569–13576.
18. Heath, R. J., Su, N., Murphy, C. K. & Rock, C. O. (2000) *J. Biol. Chem.* **275,** 40128–40133.
19. Morbidoni, H. R., de Mendoza, D. & Cronan, J. E., Jr. (1996) *J. Bacteriol.* **178,** 4794–4800.
20. Foster, S. J. & Popham, D. L. (2002) in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, eds. Sonenshein, A. L., Hoch, J. A. & Losick, R. (Am. Soc. Microbiol, Washington DC), pp. 21–41.
21. Errington, J. & Daniel, R. A. (2002) in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, eds. Sonenshein, A. L., Hoch, J. A. & Losick, R. (Am. Soc. Microbiol, Washington, DC), pp. 97–109.
22. Leyva-Vazquez, M. A. & Setlow, P. (1994) *J. Bacteriol.* **176,** 2788–2795.
23. von Wachenfeldt, C. & Hederstadt, L. (2002) in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, eds. Sonenshein, A. L., Hoch, J. A. & Losick, R. (Am. Soc. Microbiol, Washington, DC), pp. 163–179.
24. Winstedt, L. & von Wachenfeldt, C. (2000) *J. Bacteriol.* **182,** 6557–6564.
25. Switzer, R. L. (2002) in *Bacillus subtilis and Its Closest Relatives: From Genes to Cells*, eds. Sonenshein, A. L., Hoch, J. A. & Losick, R. (Am. Soc. Microbiol, Washington, DC), pp. 255–269.
26. Nitschké, P., Guerdoux-Jamet, P., Chiapello, H., Faroux, G., Henaut, C., Henaut, A. & Danchin, A. (1998) *FEMS Microbiol. Rev.* **22,** 207–227.
27. Ito, M., Guffanti, A. A., Oudega, B. & Krulwich, T. A. (1999) *J. Bacteriol.* **181,** 2394–2402.
28. Koonin, E. V. (2000) *Annu. Rev. Genomics Hum. Genet.* **1,** 99–116.