MINI-REVIEW

V. Tosato · C. V. Bruschi

Knowledge of the *Bacillus subtilis* genome: impacts on fundamental science and biotechnology

Abstract The advent of genomics has greatly influenced fundamental and applied microbiology. This has become paradigmatic in the case of *Bacillus subtilis*, a primary model bacterium for research and biotechnology. Indeed, mining its genome has provided more fruitful information than classical approaches would have yielded in a longer period of time. Through advanced analysis of its genome and transcriptome, fundamental discoveries dealing with the informational architecture of the *B. subtilis* chromosome, as well as with the elucidation of its pathway-level regulation of gene expression, have been achieved. The possibility of performing a complete metabolic manipulation of the secretory pathway of *Bacillus* is promising important biotechnological fallouts. Similar emphasis exists for the possibility of controlling the cell in the formation of biofilms with specific physical and chemical characteristics. At the theoretical level, the new concept of genetic superinformation has been formulated and its analytical approach implemented, while the understanding of the minimal genetic requirements for the existence of a reproducing bacterial cell is being tackled. In summary, the impact of the *B. subtilis* genome has philosophically revolutionised the way that basic knowledge is translated into applied microbiology and biotechnology, making this bacterium the workhorse of post-genomic microbiology.

Introduction

Bacillus spp., generally aerobic gram-positive, endospore-forming, rod-shaped bacteria, have been considered for many years as suitable experimental organisms for molecular biology. Today, there are at least 65 validly

V. Tosato (✉) · C. V. Bruschi
Microbiology Group, International Centre for Genetic
Engineering and Biotechnology,
AREA Science Park, Padriciano 99,
34012 Trieste, Italy
e-mail: tosato@icgeb.org
Tel.: +39-040-3757318
Fax: +39-040-3757343

described species in the genus *Bacillus* mainly clustered in at least five different groups based on 16S rRNA sequence analysis (Priest 1993). *Bacillus subtilis*, in particular, has been chosen as the appropriate representative of Group II and has been studied intensely for more than 40 years as a model organism for genetics and, later, as a GRAS host for the production of heterologous proteins (Harwood 1992).

The release of the complete sequence of the 4,214,810 base pairs of the *B. subtilis* genome (Kunst et al. 1997) opened the era of the functional analysis of the gram-positive bacteria and, as presented on the the cover of the 20 November 1997 issue of *Nature*, "filled a gap in genomics". A multitude of studies was carried out on the *B. subtilis* genome, from the definition of the expression profiles using microarrays to the analysis of its DNA topology and secondary structures. Extensive analyses have also been performed on the chromosome of the pathogenic species of the genus *Bacillus*. The genome sequences of *Bacillus anthracis* Ames (Read et al. 2003) and *Bacillus cereus* (Ivanova et al. 2003) have recently been released and compared with that of *Bacillus thuringiensis* (Radnedge et al. 2003) showing a very close evolutionary relationship among them, which prevents an easy classification and detection of specific regions suitable as targets for vaccines against anthrax. The analysis of the common genes coding for pathogenicity factors in these three species leads us to believe that two different ancestors existed, one for the benign soil bacilli such as *B. subtilis*, and another for the so-called *cereus* group, which was an opportunistic insect pathogen. Surely, in the coming months these new data on the pathogenic *Bacillus*-spp. will drastically change the current beliefs regarding the potential of gram-positive organisms, leading to debates, technological developments and, hopefully, improvements in human health. The twilight of the sequencing era has coincided with the dawn of the microarrays era. Today, the benefits of the microarray-based transcription profiles of *B. subtilis* are beginning to be perceived in biotechnology.

The purpose of this review is to focus on the most recent progress related to the functional analysis of the

Bacillus spp. genomes, with a particular attention to *B. subtilis*, pointing out their revolutionary impact on scientific know-how and industrial applications. Furthermore, some new conclusions on the biological meaning of DNA secondary structures are provided within the framework of this novel, in-depth type of genome analysis.

The genomic features: how to investigate the meaning of the primary structure?

Our information about the *B. subtilis* genome is mainly related to strain 168trpC2, derived from Marburg strain after X-ray irradiation (Burkholder and Giles 1947); this strain was chosen for the complete genome sequencing project. The average G+C ratio of the *B. subtilis* chromosome is 43.5%, a value that is very high when compared to the G+C content of pathogenic *Bacillus* spp. such as *B. anthracis* A2012 (35.1%) (Read et al. 2003) and *B. cereus* ATCC 14579 (35.3%) (Ivanova et al. 2003), but similar to the alkaliphilic *Bacillus halodurans* (43.7%) (Takami et al. 2000). A common feature among these species is that the G+C content varies considerably throughout the chromosome. Moreover, as another common characteristic, A+T islands reveal the signature of bacteriophage lysogens or other inserted elements leading to the hypothesis that either the chromosome is a result of a phage-mediated recombination between the genomes of closely related bacteria (Ivanova et al. 2003) or that it results from DNA fragments horizontally transferred from source organisms that are richer in A+T than the host (Nicolas et al. 2002). On a smaller scale, horizontal gene transfer could also have been accomplished by transposases. Transposons and transposon-related proteins had an evolutionary role accounting for internal rearrangements of the genome. Ten are encoded in the *B. subtilis* genome while up to 112 are encoded in *B. halodurans* C-125 chromosome (Takami et al. 2000). Gene transposition and lateral transfer analysis has allowed the development of methods for reconstructing phylogenetic trees and branching orders of gram-positive bacteria (Kunisawa 2003). In B. *subtilis* the GC skew $(nG-nC)/(nG+nC)$ and the AT skew are positive on the replication leading strand (Lobry 1996). These skews are linked to the A+G enrichment of the coding strand, due to a high preference for encoding proteins on the leading strand in *B. subtilis*.

A large number of repeats has been found in those regions that flank loci putatively involved in bacteriophage integration. A 190-bp element is repeated ten times in the chromosome, five repeats on each side of the origin of replication (Kunst et al. 1997). Similar repeated sequences are also present in various strains of *Bacillus licheniformis* in which entire important genes, nearly identical to those of *B. subtilis*, have been discovered (Tye et al. 2002). Among duplications, the *B. subtilis* genome is also characterised by genes comprising long sequence repeats. These genes encode peptide synthetases involved in the multi-enzymatic mechanism of non-ribosomal synthesis of proteins such as surfactin (Cosmina et al. 1993) and

fengycin (Tosato et al. 1997). Clusters of dimers have been analysed leading to the development of algorithms able to identify regulatory motifs in the genome of *B. subtilis* (Mwangi and Saggia 2003). The over-represented dimers have been grouped to represent classes of motifs recognised by transcription factors. A drawback of this methodology, also recognised by the authors, is that a transcription factor can bind specific dimeric sequences not over-represented in the genome and therefore undetected by the algorithm. Among particular DNA motifs, palindromic sequences seem less frequent in the bacterial genome than expected on the basis of their G+C content. In particular, there is a consistent decrease in the number of palindromes within the Shine-Dalgarno regions of *B. subtilis* (Fuglsang 2003). We believe that this is consistent with their possible impairment of DNA transcription.

DNA secondary structures, genomic organisation and topology may be exploited as superinformation by the cell

The analysis of the DNA secondary structure of the whole *B. subtilis* genome is a first step in the direction of understanding whether a correlation exists between phenomena (i.e. cellular shape, existence of preferential sites of integration in transposon mutagenesis, non-coding DNA) that cannot be explained by the primary structure of nucleic acids and the presence of hairpins and bending along the chromosome (Tosato et al. 2003). A decreasing gradient of large DNA hairpins from the origin towards the *terC* end of chromosomal replication characterises the genome of *B. subtilis*, as well as a concentration of the most curved DNA in the intergenic regions rather than within the open reading frames (ORFs). The lack of hairpins around the *terC* locus (with the exception of hairpins A and B of the replication terminus) correlates with the low level of homologous recombination reported in this 2,110- to 2,960-kb region (Chedin et al 1998; ElKaroui et al. 1999). This observation is in agreement with the knowledge that the presence of secondary structures as hairpins may favour the initiation of recombination (Lobachev et al. 1998).

On the involvement of recombination in genome plasticity, *B. subtilis* has been found to have a generally stable chromosomal structure compared with other *Bacillus* spp. *B. cereus* and *B. thuringiensis* strains have been subjected to a high frequency of genome rearrangements (Carlson et al. 1992) causing an extensive genetic diversity among different environmental isolates (Carlson et al. 1994). In contrast to the genomic diversity within *B. cereus* and *B. thuringiensis*, *B. anthracis* appears to be genetically clonal (Keim et al. 1997). This pathogenic species has the same apparatus of DNA repair proteins found in *B. subtilis*, but it appears to have additional DNA repair capabilities acting mostly on UV-induced DNA damage with two, rather than one, UV dimer endonucleases (Read et al. 2003). Moreover, many genes coding

for detoxification proteins (catalases, bromoperoxydases, superoxide dismutases) involved in preventing oxidative DNA damage by free-oxygen radicals, have no homologues in *B. subtilis*. Perhaps this capability to repair its genome with higher efficiency can be correlated with the lower genetic diversity of *B. anthracis* among different isolates.

It is known that the genomic organisation may affect gene expression, as it is paradigmatically represented by the distinct strand-specific clusters of genes coding in opposite directions starting from a single short intergenic region in some chromosomes of *Leishmania major* and other kinetoplastids (Tosato et al. 2001). *B. subtilis* chromosomal gene organisation is very different from that of other *Bacillus* spp. of the same subgroup as *B. cereus* (Okstad et al. 1999) and *Bacillus firmus* (Gronstad et al. 1998). This fact indicates that, at least in *Bacillus* spp., gene organisation is generally not conserved even between members of the same group. In *B. subtilis*, for instance, the chromosomal position of the *spoIIR* gene regulates sigmaE activation allowing the mother cell-specific gene expression that is necessary for spore formation (Zupancic et al. 2001).

DNA supercoiling also plays important roles in gene expression regulation, even if the molecular mechanism is still unclear (Chen and Wu 2003). The topology of the genomic DNA in bacteria is controlled by DNA gyrase, the prototypical type II topoisomerase, and by topoisomerase IV, which is more efficient than gyrase in unknotting and decatenating DNA. Both these enzymes have been recently characterised in *B. subtilis* (Barnes et al. 2003). The DNA supercoiling in *B. subtilis* changes with variations in environmental factors such as osmolarity, temperature, oxygen and carbon sources supply, enabling the transcription activation or repression of some specific genes. For instance a highly negative supercoiled DNA structure involving the gyrase activity is also required to induce the osmotic response (Alice and Sanchez-Rivas 1997). Negative supercoiling also regulates homologous recombination in eukaryotes (Trigueros and Rocha 2002) as well as illegitimate recombination in bacteria (Shanado et al. 1998). These and many others findings indicate that DNA supercoiling is involved not only in the regulation of gene expression, but also in the major pathways of recombination, therefore representing, together with chromosomal gene organisation and DNA secondary structures, an instrument of the cell for the management and storage of genetic superinformation.

The essentiality of few genes and of many genetic interactions are requirements for life

After the completion of the genome sequencing, another ambitious project was carried out performing a systematic inactivation of the 4,100 genes of *B. subtilis* (Kobayashi et al. 2003). The final goal was to estimate the minimal gene set required to sustain bacterial life in nutritious conditions. In other words, this extensive knockout project was developed to define the minimal bacterial cell. Analysis *in silico* had already identified 260 genes by comparison of the small genomes of *Mycoplasma genitalium* and *Haemophilus influenzae*, which can be considered essential for bacterial life (Mushegian and Koonin 1996). Only 271 genes out of 4,100 in *B. subtilis* were found to be essential for growth when inactivated singly. Of course, this kind of analysis does not detect essential functions encoded by paralogs. On the other hand, it is known that, in *Bacillus* spp. as in many other organisms, it is not so rare to find genes that, if knocked out together, cause the death of the cell, but that are not essential for life if knocked out singly (synthetic lethal mutants). To define the complete panorama of interactions among genes and their proteins resulting in survival, a mutant with the systematic and progressive deletion of all its genes should be produced. The goal of this project is not such a utopia.

On this subject, an effort to optimise the *Bacillus* spp. cell factory has been the systematic removal of a few dispensable genomic regions such as prophages and AT-rich islands, which has lead to the construction of a new microorganism lacking 332 genes (Westers et al. 2003). It is foreseeable that, following this trend of whole-genome manipulation, a complete "minimal *Bacillus* spp." will be achieved in the near future, perhaps employing an approach similar to the one being implemented by the group of Craig Venter at the Institute for Biological Energy Alternatives, involving global transposon-based genome mutagenesis (Hutchison et al. 1999). Selection will be left to the achievement of complex, biotechnologically relevant processes, such as the utilisation of polluting carbon sources as substrates for growth.

The extensive and systematic gene knockout work carried out so far has led to the assignment of many genes relative to their expression level and essentiality. In the past, it was proposed that highly expressed genes be preferentially positioned in the leading strand to allow faster DNA replication and lower transcript losses (Brewer 1988). However, it has been demonstrated that essentiality and not the efficiency of expression drives gene-strand bias in *B. subtilis* (Rocha and Danchin 2003). Today, the simultaneous expression of genes on the same strand is considered as part of a chromosomal hyperstructure, leading to a new strand-specific model for genomic segregation in bacteria (Rocha et al. 2003). Furthermore, it has been found that the presence of gene clustering in operons does not affect this model, suggesting that an operon needs to contain only one essential gene to be preferentially positioned on the leading DNA strand. These results gives credit to the role of essentiality in the organisation of the bacterial chromosome confirming, moreover, the idea that DNA secondary structures, topology, or the genomic positions of essential genes are informative for the cells, in terms of survival, as much as the primary DNA sequence. As a consequence of this theory, lethality can be provoked not only by deletion, but also by strand-switching or re-positioning of an essential gene. This is a further indication that complex interactions exist among different regions of the chromosome and, in

particular, of coding regions, supporting the conviction that an extensive, systematic, gene-by-gene deletion should give an exhaustive answer to many questions arisen after the completion of the genomic sequencing.

Transcriptome comparison provides global information about regulons, bridging the gap between genomics and its industrial applications

From the genome project, over 4,000 putative protein-coding sequences have been identified by Gene Mark Prediction (Kunst et al. 1997). Their average size is 890 bp leading to a coding content of 87%. The operons are 1898 (Rocha and Danchin 2003) and they can be predicted also efficiently by a non-homology method (Moreno-Hagelsieb and Collado-Vides 2002). A data base of promoters and transcription factors has been constructed, revealing that the number of promoters with repeated homologous binding sites is significantly lower in *B. subtilis* that in *Escherichia coli*. In addition *B. subtilis* promoters can have binding regions of activators that are downstream of transcriptional initiation sites (Ishii et al. 2001).

Comparing the pattern of transcripts from wild-type cells with the pattern of mutants for a particular function it is possible to recognise and identify genes involved in the pathway responsible for that function. In transcriptome-related studies, the pathway is usually called regulon and the genes that are induced or repressed in the mutant belong to it. For instance, by an extensive transcriptional profiling, it was found that the expression of at least 586 genes, representing more than 10% of the ORFs in the *B. subtilis* genome, is related to the *SpoA* gene, which codes for a regulatory protein necessary to start sporulation (Fawcett et al. 2000). The response of the *B. subtilis* transcriptome to the presence of glucose has also been investigated as well as the role of the pleiotropic transcriptional regulator CcpA in this response (Blencke et al. 2003). Transcriptomic approaches allowed identification of new members of the Deg regulon of *B. subtilis*, which controls the transition from the exponential to the stationary growth phase (Mader et al. 2002) and of the RelA regulon which is involved in amino acid, glucose and oxygen starvation (Eymann et al. 2002). Moreover, a transcriptional profiling approach was used to investigate the adaptation of *B. subtilis* to extreme conditions such as high salinity (Steil et al. 2003).

Studies that are transcriptome-related allow obtaining useful information for industrial applications. One of these applications is related to a particular feature of *B. subtilis*, which exists in its ability to form biofilms. To gain insight into the pathway of the biofilm formation, microarrays were used to find changes in transcriptome when bacterial cells were transitioning from a planktonic to a biofilm state (Stanley et al. 2003). By using this technology, 519 genes have been recognised to affect biofilm formation, which is heavily used in biomedical and industrial applications. Therefore, transcription profiling, especially if coupled with proteome analysis, allows outlining the draft of the major pathways, offering an impressive amount of imputs and insights about functionality of regulons. With respect to DNA array techniques, proteomic studies provide more information about signal transduction, post-translational modifications and protein secretion. A proteomic view of the cell physiology of *B. subtilis* has been established using highly sensitive two-dimensional protein gel electrophoresis and microsequencing (Hecker 2003). Since *Bacillus* spp. are frequently used as hosts for the production of biomedical recombinant proteins such as pro-insulin (Olmos-Soto and Contreras-Flores 2003) or staphylokinase (Ye et al. 1999), DNA microarray techniques and two-dimensional polyacrylamide gel electrophoresis have been used to identify the genes involved in overexpression, to improve recombinant protein production. The knowledge of the secretory pathway is also very important for the production of peptides, in a market of proteins secreted from *Bacillus* spp. that is over U.S. $1 billion per year. For this reason genomics and proteomic approaches have been used to explore the Tat pathway of *B. subtilis* to increase the secretion of proteins from *Bacillus* spp., bypassing the general secretory pathway (Sec) (van Dijl et al. 2002). Recently, the physical map of *B. subtilis* (natto), which is used for the production of Natto, a traditional Japanese dish of fermented soy beans, has been constructed and compared to the closely related *B. subtilis* Marburg 168trpC2 strain to demonstrate its difference and to implement its productivity (Qiu et al. 2003).

These are some of a multitude of examples, in which the complete transcriptional profile, built from the genomic sequencing data, provides essential information relevant not only to scientific knowledge, but also to its biotechnological applications, with a consequent improvement of the industrial production of recombinant proteins from *Bacillus* spp.

Conclusions and future perspectives

The magnitude of data resulting from the extensive analysis of the *B. subtilis* genome and its comparison with the chromosomes of other *Bacillus* spp. is evidence of the vitality and the continuous *in fieri* status of functional genomics. The acquired knowledge and the implementation of the technical progress achieved in the last 6 years, following the release of the *B. subtilis* genome sequence, has allowed substantial development in industrial processes. Proteomics on one side and bioinformatics on the other, integrated with knew knowledge on the genomic information stored within secondary DNA stuctures, complete the panorama that the functional and comparative genomics have outlined. More genome-wide, technology-oriented approaches to the solution of complex biotechnologically related problems will be implemented in the future. This will help us to read more efficiently an additional new page from the ancient book of life everyday, as well as to continue to profit from it. However, the prospect deriving from these new ap-

proaches does raise some concern about the necessary control over a possible exaggerated use or misuse of this technology.

References

Alice AF, Sanchez-Rivas C (1997) DNA supercoiling and osmoresistance in *B. subtilis* 168. Curr Microbiol 35:309–315

Barnes MH, LaMarr WA, Foster KA (2003) DNA gyrase and DNA topoisomerase of *Bacillus subtilis*: expression and characterization of recombinant enzymes encoded by the gyrA, gyrB and parC, parE genes. Protein Expr Purif 29:259–264

Blenke HM, Homuth G, Ludwig H, Mader U, Hecker M, Stulke J (2003) Transcriptional profiling of gene expression in response to glucose in *Bacillus subtilis*: regulation of the central metabolic pathways. Metab Eng 5:133–149

Brewer BJ (1988) When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. Cell 53:679–686

Burkholder PR, Giles NH (1947) Induced biochemical mutations in *Bacillus subtilis.* Am J Bot 33:345–348

Carlson CR, Gronstad A, Kolsto AB (1992) Physical map of the genomes of three *Bacillus cereus* strains. J Bacteriol 174:3750–3756

Carlson CR, Caugant DA, Kolsto AB (1994) Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. Appl Environ Microbiol 60:1719–1725

Chedin F, Noirot P, Biaudet V, Ehrlich SD (1998) A five nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *B. subtilis.* Mol Microbiol 29:1369–1377

Chen CC, Wu HY (2003) Transcription-driven DNA supercoiling and gene expression control. Front Biosci 8:430–439

Cosmina P, Rodriguez F, de Ferra F, Grandi G, Perego M, Venema G, van Sinderen D (1993) Sequence and analysis of the genetic locus responsible for surfactin synthesis in *Bacillus subtilis.* Mol Microbiol 8:821–831

ElKaroui M, Biaudet V, Schbath S, Gruss A (1999) Characteristics of Chi distribution on different bacterial genomes. Res Microbiol 150:579–587

Eymann C, Homuth G, Scharf C, Hecker M (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J Bacteriol 184:2500–2520

Fawcett P, Eichenberger P, Losick R, Youngman P (2000) The transcriptional profile of early to middle sporulation in *Bacillus subtilis.* Proc Natl Acad Sci U S A 97:8063–8068

Fuglsang A (2003) Distribution of potential type II restriction sites (palindromes) in prokaryotes. Biochem Biophys Res Commun 310:280–285

Gronstad A, Jaroszewicz E, Ito M, Sturr MG, Krulwich TA, Kolsto AB (1998) Physical map of alkaliphylic *Bacillus firmus* OF4 and detection of a large endogenous plasmid. Extremophiles 2:447–453

Harwood CR (1992) *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. Trends Biotechnol 10:247–256

Hecker M (2003) A proteomic view of cell physiology of *Bacillus subtilis* bringing the genome sequence to life. Adv Biochem Eng Biotechnol 83:57–92

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999). Global transposon mutagenesis and a minimal *Mycoplasma* genome. Science 286:2165–2169

Ishii T, Yoshida K, Terai G, Fujita Y, Nakai K (2001) DBTBS: a database of *Bacillus subtilis* promoters and transcription factors. Nucleic Acids Res 29:278–280

Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, Bhattacharyya A, Reznik G, Mikhailova N, Lapidus A, Chu L, Mazur M, Goltsman E, Larsen N, D'Souza M, Walunas T, Grechkin Y, Pusch G, Haselkorn R, Fonstein M, Ehrlich SD, Overbeek R, Kyrpides N (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423:87–91

Keim P, Kalif A, Schupp J, Hill K, Travis SE, Richmond K, Adair DM, Hugh-Jones M, Kuske CR, Jackson P (1997) Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J Bacteriol 179:818–824

Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K, Chapuis J, Christiansen LC, Danchin A, Debarbouille M, Dervyn E, Deuerling E, Devine K, Devine SK, Dreesen O, Errington J, Fillinger S, Foster SJ, Fujita Y, Galizzi A, Gardan R, Eschevins C, Fukushima T, Haga K, Harwood CR, Hecker M, Hosoya D, Hullo MF, Kakeshita H, Karamata D, Kasahara Y, Kawamura F, Koga K, Koski P, Kuwana R, Imamura D, Ishimaru M, Ishikawa S, Ishio I, Le Coq D, Masson A, Mauel C, Meima R, Mellado RP, Moir A, Moriya S, Nagakawa E, Nanamiya H, Nakai S, Nygaard P, Ogura M, Ohanan T, O'Reilly M, O'Rourke M, Pragai Z, Pooley HM, Rapoport G, Rawlins JP, Rivas LA, Rivolta C, Sadaie A, Sadaie Y, Sarvas M, Sato T, Saxild HH, Scanlan E, W, Seegers JF, Sekiguchi J, Sekowska A, Seror SJ, Simon M, Stragier P, Studer R, Takamatsu H, Tanaka T, Takeuchi M, Thomaides HB, Vagner V, van Dijl JM, Watabe K, Wipat A, Yamamoto H, Yamamoto M, Yamamoto Y, Yamane K, Yata K, Yoshida K, Yoshikawa H, Zuber U, Ogasawara N (2003) Essential *Bacillus subtilis* genes. Proc Natl Acad Sci U S A 100:4678–4683

Kunisawa T (2003) Gene arrangements and branching orders of gram-positive bacteria. J Theor Biol 222:495–503

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A, et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature 390:249–256

Lobachev KS, Shor BM, Tran HT, Taylor W, Keen JD, Resnick MA, Gordenin DA (1998) Factors affecting inverted repeat stimulation of recombination and deletion in *Saccharomyces cerevisiae*. Genetics 148:1507–1524

Lobry JR (1996) Asymmetric substitution patterns in the two DNA strands of bacteria. Mol Biol Evol 13:660–665

Mader U, Antelmann H, Buder T, Dahl MK, Hecker M, Homuth G (2002) *Bacillus subtilis* functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. Mol Genet Genomics 268:455–467

Moreno-Hagelsieb G, Collado-Vides JA (2002) A powerful non-homology method for the prediction of operons in prokaryotes. Bioinformatics 18:329–336

Mushegian AR, Koonin EV (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A 93:10268–10273

Mwangi MM, Saggia ED (2003) Genome wide identification of regulatory motifs in *Bacillus subtilis.* BMC Bioinformatics 4:18–38

Nicolas P, Bize L, Muri F, Hoebeke M, Rodolphe F, Ehrlich SD, Prum B, Bessieres P (2002) Mining *Bacillus subtilis* chromosome heterogeneities using hidden Markov models. Nucleic Acids Res 30:1418–1426

Okstad OA, Hegna I, Lindback T, Rishovd A, Kolsto AB (1999) Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis.* Microbiology 145:621–631

Olmos-Soto J, Contreras-Flores R (2003) Genetic system constructed to overproduce and secrete proinsulin in *Bacillus subtilis*. Appl Microbiol Biotechnol 62:369–373

Priest FG (1993) Systematics and Ecology of *Bacillus*. In: Sonenshein AL, Hoch JA, Losick R (eds) *Bacillus subtilis* and other gram positive bacteria. American Society for Microbiology, Washington D.C., pp 3–16

Qiu D, Oshima H, Ohashi Y, Itaya M (2003) Construction of physical maps of *Bacillus subtilis* (natto) strains. Nucleic Acids Res [Suppl] 3:207–208

Radnedge L, Agron PG, Hill KK, Jackson PJ, Ticknor LO, Keim P, Andersen GL (2003) Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. Appl Environ Microbiol 69:2755–2764

Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, Tettelin H, Fouts DE, Eisen JA, Gill SR, Holtzapple EK, Okstad OA, Helgason E, Rilstone J, Wu M, Kolonay JF, Beanan MJ, Dodson RJ, Brinkac LM, Gwinn M, DeBoy RT, Madpu R, Daugherty SC, Durkin AS, Haft DH, Nelson WC, Peterson JD, Pop M, Khouri HM, Radune D, Benton JL, Mahamoud Y, Jiang L, Hance IR, Weidman JF, Berry KJ, Plaut RD, Wolf AM, Watkins KL, Nierman WC, Hazen A, Cline R, Redmond C, Thwaite JE, White O, Salzberg SL, Thomason B, Friedlander AM, Koehler TM, Hanna PC, Kolsto AB, Fraser CM (2003) The genome sequence of *B. anthracis* Ames and comparison to closely related bacteria. Nature 423:81–86

Rocha EP, Danchin A (2003) Essentiality, not expressiveness, drives gene strand bias in bacteria. Nature Genet 34:377–378

Rocha EP, Fralick J, Vediyappan G, Danchin A, Norris V (2003) A strand-specific model for chromosome segregation in bacteria. Mol Microbiol 49:895–903

Shanado Y, Kato J, Ikeda K (1998) *Escherichia coli* HU protein suppresses DNA-gyrase mediated illegitimate recombination and SOS induction. Genes Cells 3:511–520

Stanley NR, Britton RA, Grossman AD, Lazazzera BA (2003) Identification of catabolite repression as a physiological regulator of biofilm formation by *Bacillus subtilis* by use of DNA microarrays. J Bacteriol 185:1951–1957

Steil L, Hoffmann T, Budde I, Volker U, Bremer E (2003) Genome-wide transcriptional profiling analysis of adaptation of *Bacillus subtilis* to high salinity. J Bacteriol 185:6358–6370

Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K (2000) Complete genome sequence of the alkaliphylic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res 28:4317–4331

Tosato V, Albertini AM, Zotti M, Sonda S, Bruschi CV (1997) Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. Microbiology 143:3443–3450

Tosato V, Ciarloni L, Ivens AC, Rajandream MA, Barrell BG,-*Leishmania major* Friedlin chromosomes. Curr Genet 40:186–194

Tosato V, Gjuracic K, Vlahovicek K, Pongor S, Danchin A, Bruschi CV (2003) The DNA secondary structure of the *Bacillus subtilis* genome. FEMS Microbiol Lett 218:23–30

Trigueros S, Rocha J (2002) Failure to relax negative supercoiling of DNA is a primary cause of mitotic hyper-recombination in topoisomerase-deficient yeast cells. J Biol Chem 277:37207–37211

Tye AJ, Siu FK, Leung TY, Lim BL (2002) Molecular cloning and the biochemical characterization of two novel phytases from *Bacillus subtilis* 168 and *Bacillus licheniformis*. Appl Microbiol Biotechnol 59:190–197

Van Dijl JM, Braun PG, Robinson C, Quax WJ, Antelmann H, Hecker M, Muller J, Tjalsma H, Bron S, Jongbloed JD (2002) Functional genomic analysis of the *Bacillus subtilis* Tat pathway for protein secretion. J Biotechnol 98:243–54

Westers H, Dorenbos R, Van Dijl JM, Kabel J, Flanagan T, Devine KM, Jude F, Seror SJ, Beekman AC, Darmon E, Eschevins C, De Jong A, Bron S, Kuipers OP, Albertini AM, Antelmann H, Hecker M, Zamboni N, Sauer U, Bruand C, Ehrlich DS, Alonso JC, Salas M, Quax WJ (2003) Genome engineering reveals large dispensable regions in *Bacillus subtilis*. Mol Biol Evol (in press)

Ye R, Kim JH, Kim BG, Szarka S, Sihota E, Wong SL (1999) High level secretory production of intact, biologically active staphylokinase from *Bacillus subtilis*. Biotechnol Bioeng 62:87–96

Zupancic ML, Tran H, Hofmeister AE (2001) Chromosomal organization governs the timing of cell type-specific gene expression required for spore formation in *Bacillus subtilis*. Mol Microbiol 39:1471–1481