

The Complete Genome Sequence of *Bacillus licheniformis* DSM13, an Organism with Great Industrial Potential

Birgit Veith^a Christina Herzberg^a Silke Steckel^a Jörg Feesche^b
Karl Heinz Maurer^b Petra Ehrenreich^a Sebastian Bäumer^a Anke Henne^a
Heiko Liesegang^a Rainer Merkl^a Armin Ehrenreich^a Gerhard Gottschalk^a

^aGöttingen Genomics Laboratory and Competence Centre for Genome Research on Bacteria, Institute of Microbiology and Genetics, University of Göttingen, Göttingen, and ^bHenkel KGaA, VBT Enzymtechnologie, Düsseldorf, Germany

Key Words

Bacillus licheniformis DSM13 · Genome sequence ·
Bacillus subtilis · Industrial enzymes · Industrial
fermentation

Abstract

The genome of *Bacillus licheniformis* DSM13 consists of a single chromosome that has a size of 4,222,748 base pairs. The average G+C ratio is 46.2%. 4,286 open reading frames, 72 tRNA genes, 7 rRNA operons and 20 transposase genes were identified. The genome shows a marked co-linearity with *Bacillus subtilis* but contains defined inserted regions that can be identified at the sequence as well as at the functional level. *B. licheniformis* DSM13 has a well-conserved secretory system, no polyketide biosynthesis, but is able to form the lipopeptide lichenysin. From the further analysis of the genome sequence, we identified conserved regulatory DNA motifs, the occurrence of the glyoxylate bypass and the presence of an anaerobic ribonucleotide reductase explaining that *B. licheniformis* is able to grow on acetate and 2,3-butanediol as well as anaerobically on glucose. Many new genes of potential interest for biotechnological applications were found in *B. licheniformis*; candidates include proteases, pectate lyases, lipases and various polysaccharide degrading enzymes.

Copyright © 2004 S. Karger AG, Basel

Introduction

Bacillus licheniformis is a Gram-positive endospore-forming organism that can be isolated from soils and plant material all over the world [Sneath et al., 1986]. The organism was never reported to be pathogenic for either animals or plants and is used extensively for large-scale industrial production of exoenzymes as it can secrete large quantities of proteins of up to 20–25 g/l [Schallmeyer et al., 2004].

The alkaline serine proteases (subtilisins) that are manufactured with *B. licheniformis* and also with *Bacillus pumilus* and *Bacillus subtilis* have a primary application as additives to household detergents. Their annual output has been estimated to about 500 metric tonnes of pure enzyme protein [Schallmeyer et al., 2004]. Other products that can be produced by fermentation of *B. licheniformis* strains are amylases [Declerck et al., 2000; Yuuki et al., 1985] and the topical antibiotic bacitracin [Froyshov and Laland, 1974]. The hosts *B. licheniformis* and *Bacillus clausii* are also extremely important for commercial processes for heterologous exoenzymes as they frequently exhibit higher enzyme yields than *B. subtilis* [Schallmeyer et al., 2004].

B. licheniformis belongs to the *B. subtilis* group (group II) of the genus *Bacillus* together with other well-known species whose complete genome sequence has been deter-

Table 1. General genomic features of *B. licheniformis* as compared to *B. subtilis*, *B. halodurans*, *B. cereus* and *B. anthracis**

	<i>B. licheniformis</i>	<i>B. subtilis</i>	<i>B. halodurans</i>	<i>B. cereus</i>		<i>B. anthracis</i>		
	chromosome	chromosome	chromosome	chromosome	pBc10987**	chromosome	pXO1**	pXO2**
Size, bp	4,222,748	4,214,810	4,202,353	5,224,283	208,369	5,227,293	181,677	94,829
Number of genes	4,286	4,112	4,066	5,642	242	5,508	217	113
% coding	87.9	87.0	85.0	85.0	80.9	84.3	77.1	76.2
% G+C	46.2	43.5	43.7	35.6	33.5	35.4	32.5	33.0
rRNA operons	7	10	8	12	0	11	0	0
tRNA genes	72	86	78	98	0	95	0	0

* *B. subtilis* [Kunst et al., 1997]; *B. halodurans* [Takami et al., 2000]; *B. cereus* [Ivanova et al., 2003]; *B. anthracis* [Read et al., 2003].

** Plasmids of *B. cereus* and *B. anthracis*, respectively.

mined. These are *Bacillus anthracis* [Read et al., 2003], *Bacillus cereus* [Ivanova et al., 2003; Rasko et al., 2004], *Bacillus thuringiensis*, the alkaliphilic species *Bacillus halodurans* [Takami and Horikoshi, 2000; Takami et al., 2000] and *B. subtilis* [Kunst et al., 1997]. In the context of an extensive comparative genomics of this group of organisms and because of the biotechnological importance of the organism, we sequenced the genome of *B. licheniformis* and present a first analysis of data derived from the annotated sequence.

Results and Discussion

General Features of the *B. licheniformis* Genome

The genome of *B. licheniformis* is a single circular chromosome consisting of 4,222,748 base pairs (bp). Some of the features of this genome in comparison to genomes of other *Bacillus* species are summarized in table 1.

Clearly, the *B. licheniformis/subtilis/halodurans* group is quite distinct of the *B. cereus/anthracis* group [Rasko et al., 2004]. The genomes of the latter are larger by approximately 1 Mbp and plasmids are present. In addition, the G+C content is lower by approximately 10%. The *B. licheniformis* genome is lowest with respect to rRNA operons (7) and tRNA genes (72). The origin of replication was determined from the GC-skew [Lobry, 1996; Tillier and Collins, 2000] (supplementary fig. 1, available in online version only). A sharp point of inflection indicates the initiation site for chromosomal replication near the *dnaA* gene (BLi00001). In the surrounding region of the *dnaA* gene characteristic arrangements of DnaA sequence boxes are located (supplementary fig. 2, available in online version only). Organisms of the genera *Bacillus*

and *Clostridium* apparently perform preferably a co-directional replication and transcription. 74.3% of the ORFs are located on the leading strand. This compares well with *B. subtilis* but is significantly lower than the 82% in *Clostridium tetani* [Brüggemann et al., 2003] or *Clostridium perfringens* [Shimizu et al., 2002]. The pronounced bias of transcription is visible in figure 1 where the outermost orange circle marks genes transcribed clockwise while the blue circle visualizes genes directed counter clockwise.

Genome Comparison of Group II Bacilli

An automatic comparison of *B. licheniformis*, *B. subtilis* and *B. halodurans* was done on the ORF level and the results are summarized in figure 2. All three organisms do have a core genome of 2,323 orthologous proteins in common. There are 902 genes unique to *B. licheniformis*, 771 unique to *B. subtilis* and 1,408 unique to *B. halodurans*. *B. licheniformis* shares with *B. subtilis* 872 genes and only 189 with *B. halodurans*. For these comparisons a threshold e-value of 10^{-15} was used. This value was deduced from the manual inspection of BLAST alignments of *B. licheniformis* and *B. subtilis* ORFs and then applied to the automatic analysis of the *B. halodurans* data. This procedure could be used because the genomes of the three bacilli are of about the same size.

The ORFs *B. licheniformis* and *B. subtilis* have in common were characterized by BLAST analysis. Since the BLAST algorithm looks for local alignments, BLAST hits may result from conserved domains and necessarily not from the full length of a protein. Therefore, the BLAST alignments were inspected manually before an orthologous pair was defined. 74.6% of all *B. licheniformis* ORFs have more than 30% identical amino acids when compared with an orthologous ORF of *B. subtilis*. 43.9% have

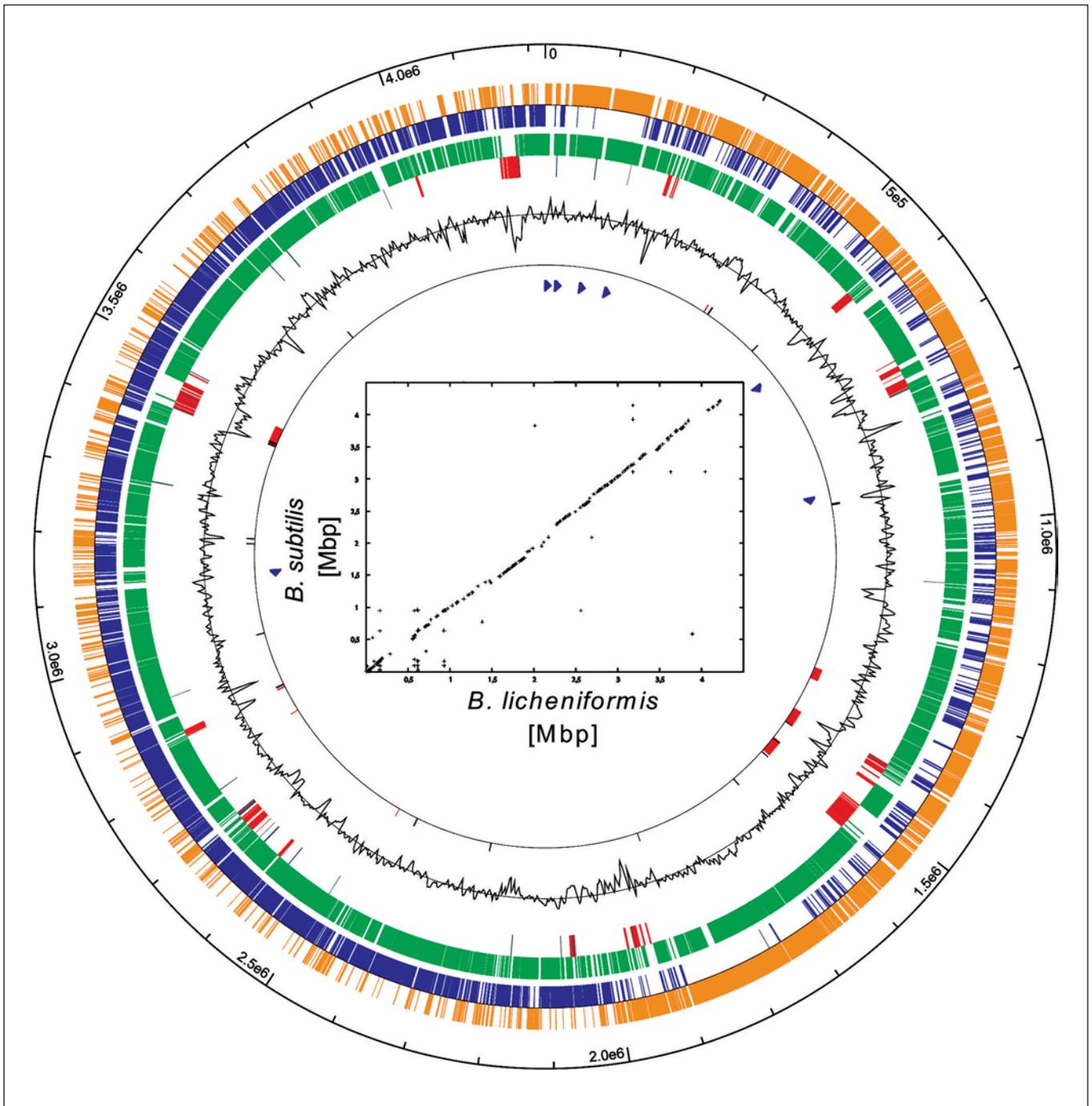


Fig. 1. Circular map of the chromosome of *B. licheniformis*. The coding sequence of the chromosome is shown in orange or blue, depending on strand orientation. ORFs of *B. licheniformis* that have homologous proteins in *B. subtilis* are depicted in green. Genes of *B. licheniformis* with a differing codon usage (alien genes) are shown in red. The variation of the G+C content is indicated as black graph (higher values outward). Prophages are highlighted in red boxes with IS elements depicted as short black lines. In the innermost ring, rRNA clusters are marked by blue arrows. The graph in the center of the circular map shows the MUMmer plot of *B. licheniformis* against *B. subtilis* based on nucleotide sequences.

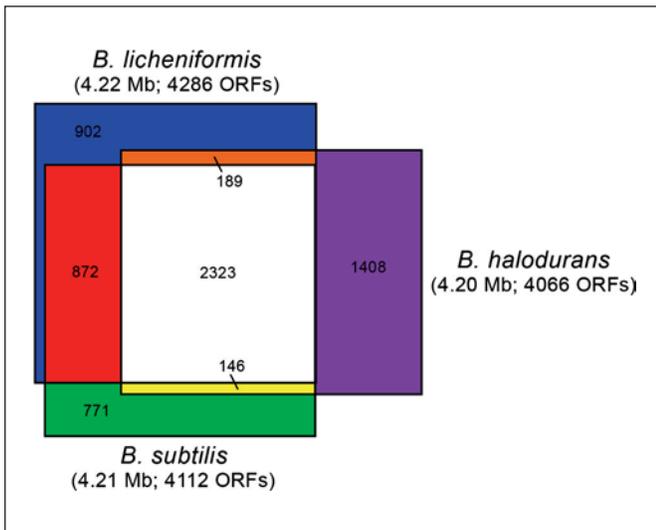


Fig. 2. Occurrence of homologous ORFs in *B. licheniformis*, *B. subtilis* and *B. halodurans*. Core ORFs, white; ORFs in *B. licheniformis* with homologues in *B. subtilis* but not in *B. halodurans*, red; *B. licheniformis* ORFs present in *B. halodurans* but absent in *B. subtilis*, orange; *B. subtilis* ORFs present in *B. halodurans* but absent in *B. licheniformis*, yellow. *B. licheniformis* ORFs not present in *B. subtilis* and *B. halodurans*, blue. *B. subtilis* ORFs not present in *B. licheniformis* and *B. halodurans*, green; *B. halodurans* ORFs not present in *B. subtilis* and *B. licheniformis*, purple. The size of the rectangles is proportional to the number of ORFs in each organism.

an ortholog with more than 70% identity and 14.8% of them even share more than 85% of amino acids with a counterpart in *B. subtilis*.

The third, green circle of figure 1 represents the location of ORFs that are orthologous to ORFs in *B. subtilis*. A mosaic pattern is apparent, stretches of *B. licheniformis* genes also present in *B. subtilis* are interrupted by stretches unique to *B. licheniformis*. Since the presence of genes tends to be more conserved than their position [Huynen and Bork, 1998; Lathe et al., 2000] it is not clear whether the segments of conserved functionality also represent co-linearity of the genomes. Therefore, we compared the *B. licheniformis* genome to the *B. subtilis* genome by MUMmer plot analysis [Delcher et al., 2002; Kurtz et al., 2004].

The plot (center of fig. 1) indicates segments of pronounced similarity of the genomes at the nucleotide level and suggests that patches of external DNA have been added to a conserved genomic core. Sequence conservation in regions of core genes even allows the identification of described recognition motives for sigma factors. We identified the described consensus motives that are recognized

by the SigD (BLi01868) sigma factor responsible for a regulon including late flagellar genes, several chemotaxis proteins and the major vegetative autolysins [Helmann and Moran, 2002] as shown in supplementary table 1, available in online version only. It was possible to identify conserved recognition sequences for this sigma factor upstream of the *hag* (BLi03780), *yvyC* (BLi03779), *yvyF* (BLi03788), *motA* (BLi01524), *cheV* (BLi01614), *mcpA* (BLi03295), *mcpC* (BLi01604) and *lytD* (BLi03821) genes. In another round of analysis, we searched for motives described to interact with SigL, a sigma factor that controls a regulon including genes for the catabolism of several amino acids including arginine, ornithine, leucine and valine and for the acetoin catabolism operon. We did identify the described consensus upstream of the *levD* (BLi02831), *rocD* (BLi00422), *ptb* (BLi02586), *yveP* (BLi03689) and *acoA* (BLi00849) genes (supplementary table 2, available in online version only). This is evidence for high functional conservation of the regulatory mechanisms of genes of the core genome.

To gain insight into the evolutionary mechanisms that shaped the genome of *B. licheniformis* we analyzed it with the program SIGI [Merkl, 2004]. The analysis calculates the codon frequency of each gene and compares it to the average codon usage model of the whole genome. This comparative codon usage evaluation identifies genes, which were possibly acquired by lateral gene transfer. The positions of genes predicted to have a significantly different codon usage are shown as red bars in the fourth circle of figure 1. It can be seen that they map well within areas that are unique to *B. licheniformis*, supporting the view that these DNA segments have been inserted into the core genome relatively recently as compared to the evolution of the core genome. This is also supported by the course of the G+C ratio. Three of the identified prophages fall in segments of potentially exogenous DNA. The fourth identified prophage is devoid of the statistical anomalies. This phage seems to be orthologous to a prophage also inserted into the *B. subtilis* genome at the same position. Therefore, this phage might reside in the genome for a long time, maybe it was already there prior to the evolutionary split of *B. licheniformis* and *B. subtilis*.

The genes that are common in *B. subtilis* and *B. licheniformis* include the central pathways of glycolysis, pentose phosphate cycle and the tricarboxylic acid cycle with the noted exception of the glyoxylate bypass that is missing from *B. subtilis*. Both organisms can synthesize all amino acids and vitamins, a fact supported by their ability to grow in mineral medium with a defined carbon and energy source. All the genes for protein secretion via

Table 2. Selected exoenzymes identified in the genome of *B. licheniformis* and corresponding orthologs in *B. subtilis*

Gene ID	Function	Gene designation in <i>B. subtilis</i>
BLi00656	α -Amylase precursor (EC 3.2.1.1)	
BLi03543	α -Glucosidase	
BLi02117	α -Glucosidase (EC 3.2.1.20)	
BLi03021	α -L-Arabinofuranosidase	<i>abfA</i>
BLi01295	Arabinan endo-1,5-L-arabinase	<i>abnA</i>
BLi04220	Arabinan endo-1,5- α -L-arabinosidase	<i>yxjA</i>
BLi04276	Arabinogalactane endo-1,4- α -galactosidase	<i>yvfO</i>
BLi00447	β -Galactosidase	<i>lacA</i>
BLi04214	β -Glucosidase	<i>bglH</i>
BLi01882	Cellulase (EC 3.2.1.4)	
BLi01881	Cellulose 1,4- β -cellobiosidase	
BLi00338	Chitinase (EC 3.2.1.14)	
BLi00339	Chitinase (EC 3.2.1.14)	
BLi02088	Endo-1,4- β -glucanase	<i>bglC</i>
BLi01883	Endo-1,4- β -mannosidase	
BLi00655	Endo-1,4- β -xylanase	<i>yjeA</i>
BLi01880	Endo-1,4-glucanase (EC 3.2.1.4)	
BLi00545	Esterase/lipase	
BLi00340	Glutamic acid-specific protease	<i>mpr</i>
BLi02827	Levanase	<i>sacC</i>
BLi03707	Levanase	<i>yveB</i>
BLi03706	Levansucrase	<i>sacB</i>
BLi03370	Lipase/esterase	
BLi02821	Lipase/esterase	
BLi00658	Maltogenic α -amylase (EC 3.2.1.1)	
BLi04019	Minor extracellular serine protease	<i>vpr</i>
BLi01123	Minor extracellular serine protease	<i>epr</i>
BLi01404	Pectate lyase	<i>pel</i>
BLi03053	Pectate lyase	<i>pelB</i>
BLi03741	Pectate lyase	<i>yvpA</i>
BLi04129	Pectate lyase	
BLi03498	Pectin methylesterase	
BLi04177	Peptidase T	<i>pepT</i>
BLi01399	Polysugar-degrading enzyme	<i>ykjC</i>
BLi02863	Protease	<i>yrrN</i>
BLi02862	Protease	<i>yrrO</i>
BLi01109	Subtilisin Carlsberg precursor (EC 3.4.21.62)	
BLi01909	Zinc protease (EC 3.4.99.-)	

the Sec system are present in both organisms: *secA* (BLi03773), *secDF* (BLi02891), *secE* (BLi00118), *secG* (BLi03643), *secY* (BLi00153), *ffh* (BLi01818) and *ftsY* (BLi01816). Four signal peptidases *sipS* (BLi00675), *sipT* (BLi01655), *sipV* (BLi01122) and *sipW* (BLi02638) could be identified [Tjalsma et al., 2000]. An ortholog to *sipU*, a fifth signal peptidase in *B. subtilis*, could not be detected. The genes for the sporulation system have mostly been identified in *B. licheniformis* with no notable difference. This is of no surprise as *B. licheniformis* sporulates readily at the end of its logarithmic growth.

Genes Present either in *B. subtilis* or in *B. licheniformis*

The most striking difference on the *B. subtilis* side is the presence of polyketide synthetases that comprise almost 4% of the genome. These synthetases are apparently not present in *B. licheniformis*. From the 1,091 genes unique to *B. licheniformis* in comparison with *B. subtilis*, 52% are hypothetical, 11% are involved in transport functions, 8% in regulation and 17% in various enzymatic reactions (supplementary table 3, available in online version only). Four genes could be identified in *B. licheniformis* (BLi00401, BLi00402, BLi00403,

BLi00404) coding for the non-ribosomal peptide synthetase complex for lichenysin biosynthesis [Konz et al., 1999]. Genes for bacitracin synthesis are absent in strain DSM13 but have been described for other *B. licheniformis* strains. *B. licheniformis* forms a reddish brown pigment on carbohydrate media containing sufficient iron [Sneath et al., 1986]. There are several candidate genes for the biosynthesis of this potential siderophore.

The ORFs BLi01186, BLi01185, BLi01187 and BLi01188 might be involved in its synthesis [Martinez et al., 1994; Moss et al., 1999]. Two operon structures with type I restriction systems were detected (BLi04315, BLi04316, BLi04318 and BLi00743, BLi00746, BLi00745, BLi00744), the presence of these systems might account for the reported difficulties in transforming this organism with heterologous DNA [O'Sullivan et al., 2000].

B. licheniformis is an important producer of exoenzymes. Prominent genes in this context have been identified, several of them are summarized in table 2. They include genes coding for the well-studied protease Subtilisin Carlsberg precursor (BLi01109) [Jacobs et al., 1985; Smith et al., 1968; Syed et al., 1993], the glutamic acid-specific protease (BLi00340) [Kakudo et al., 1992; Svendsen and Breddam, 1992], the maltogenic α -amylase (BLi00658) [Kim et al., 1992] and the heat- and pH-stable α -amylase (BLi00656) [Gray et al., 1986; Kandra et al., 2002; Stephens et al., 1984; Yuuki et al., 1985].

B. licheniformis contains genes with strong homology to isocitrate lyase (BLi04207) [Sandeman et al., 1991] and malate synthase (BLi04208) [Sharma et al., 2000] together forming the glyoxylate bypass enabling the organism to grow on C-2 substrates [Gottschalk, 1986]. Acetate and 2,3-butanediol which is assimilated via C-2 units [Fründ et al., 1989; Oppermann et al., 1989] are end products of incomplete oxidations or fermentations as carried out by bacilli. *B. subtilis* cannot grow on acetate or 2,3-butanediol but *B. licheniformis* does, which was confirmed [Veith and Ehrenreich, unpubl. data]. *B. licheniformis* might therefore be better adapted to gain additional energy from incompletely oxidized products accumulated during growth under insufficient oxygen supply. In this connection it is of interest that *B. licheniformis* is able to grow anaerobically on glucose [Veith and Ehrenreich, unpubl. data].

At least one system important for anaerobic growth was detected on the genome, genes for anaerobic ribonucleotide reductase and the accompanying activating enzyme (BLi03824, BLi03823, BLi04172) [Torrents et al., 2000]. These enzymes are also present in *B. cereus*,

which has been reported to grow anaerobically [Sneath et al., 1986], whereas they are lacking in *B. subtilis*.

It can be concluded that *B. licheniformis* is very effective in utilizing carbohydrates under conditions of varying oxygen tensions. This might be responsible for the high growth yields that have been described in the literature [Schallmeyer et al., 2004]. Genes for denitrification were not observed and experimentally denitrification could not be demonstrated with strain DSM13 and some others (DSM14, DSM1913, DSM1969, DSM12369, DSM12370). This is in contrast to textbooks in which the ability of denitrification is listed as a characteristic of *B. licheniformis* [Sneath et al., 1986].

Experimental Procedures

Genome Sequencing

Total genomic DNA of type strain DSM13 was prepared, sheared randomly and fractionated by 0.8% agarose gel electrophoresis. A small insert shotgun library was constructed by blunt cloning the dephosphorylated size fraction from 2 to 2.5 kb into the *Sma*I site of pTZ19R-Cm. A large insert library was constructed from *Sau*3AI partially digested genomic DNA in the SuperCos1 vector system. Insert ends of the recombinant plasmids were sequenced by using dye-terminator chemistry with Mega-BACE 1000/4000 and ABI Prism 377 DNA automated sequencers (Amersham Bioscience and Applied Biosystems). 46,100 sequence reads of an average length of 693 bp resulting in a mean coverage of 7.2-fold were produced. The sequence reads from the small insert library were processed with the PHRED software [Ewing et al., 1998], screened for vector sequence and poor quality regions by the PREGAP4 software [Staden et al., 2000]. The processed reads were assembled by GCPHRAP into a Gap4 database for further sequence editing. For closing remaining gaps the contigs were ordered by a chromosome comparison to *B. subtilis* using the ERGO software [Overbeek et al., 2003] to identify potential long PCR reactions. The gaps that could not be closed by this approach were closed by combinatorial multiplex PCR, vectorette technique and primer walking.

ORF Prediction, Annotation and Comparative Genomics

ORF prediction was done by using the YACOP software [Tech and Merkl, 2003]. Automatic annotation was done first by using the ERGO annotation tool and was verified and refined by manual annotation, considering FASTA [Pearson, 1994] and BLAST [Altschul et al., 1990] search against Swissprot, GenBank/European Molecular Biology Laboratory databases [Benson et al., 2004; Kulikova et al., 2004]. All similarity-based assignments were confirmed by checking the protein sequences with TMPred for transmembrane regions [Hofmann and Stoffel, 1993]. Comparisons to Pfam [Bateman et al., 2004], COGs and Prosite [Hulo et al., 2004] were done to confirm the occurrence of the functional motives that correspond to the assigned function and the presence of ribosome-binding sites.

Final annotation of ORFs that are orthologous in *B. licheniformis* and *B. subtilis* was done by protein/protein BLAST search of each ORF in *B. licheniformis* against a *B. subtilis* database and visually inspecting the resulting BLAST alignments that cover the whole length of every single ORF in question by the aid of internally developed software.

For comparative genomics of *B. licheniformis*, *B. subtilis* and *B. halodurans*, we did the following BLAST comparisons with each ORF of each of the first organisms against a database of the second one: *B. licheniformis*/*B. halodurans*, *B. licheniformis*/*B. subtilis*, *B. halodurans*/*B. subtilis*. The threshold e-value of 10^{-15} used in this analysis was derived from the number of ORFs found to be orthologous by the final annotation described above.

Growth Experiments

For determination of some growth capabilities of *B. licheniformis* we used a mineral medium consisting of 40 mM Na-K-phosphate buffer, pH 7.5, 10 mM NH₄Cl, 0.1 mM CaCl₂, 0.5 mM MgSO₄ and trace element solution SL9 [Ehrenreich and Widdel, 1994]. This medium was supplemented depending on the type of experiment with 30 mM glucose, 30 mM 2,3-butanediol or 30 mM

sodium acetate, the NH₄Cl was replaced by 10 mM NaNO₃ when testing possible nitrogen sources and by 50 mM NaNO₃ when testing denitrification. Additionally, denitrification was tested on NB-Medium (Merck, Darmstadt, Germany) supplemented with 50 mM NaNO₃. Anaerobic growth was assayed in completely filled screw-cap tubes. For the denitrification assay, inverted Durham tubes were included in the anaerobic screw-cap tubes. Cultures of *Pseudomonas putida* DSM 50906 were used as positive control.

Data deposition: The sequence reported in this paper has been deposited in the Genbank database (accession No. AE017333).

Acknowledgements

This work was supported by a grant of the Niedersächsisches Ministerium für Wissenschaft und Kultur to the Göttingen Genomics Laboratory and by funds of the Competence Network Göttingen 'Genome Research on Bacteria' financed by the German Federal Ministry of Education and Research (BMBF).

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. 2004. The Pfam protein families database. *Nucl Acids Res* 32 Database issue, D138–D141.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. 2004. GenBank: update. *Nucl Acids Res* 32 Database issue, D23–D26.
- Brüggemann, H., Bäumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A., Gottschalk, G. 2003. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc Natl Acad Sci USA* 100: 1316–1321.
- Declerck, N., Machius, M., Wiegand, G., Huber, R., Gaillardin, C. 2000. Probing structural determinants specifying high thermostability in *Bacillus licheniformis* α -amylase. *J Mol Biol* 301:1041–1057.
- Delcher, A.L., Phillippy, A., Carlton, J., Salzberg, S.L. 2002. Fast algorithms for large-scale genome alignment and comparison. *Nucl Acids Res* 30:2478–2483.
- Ehrenreich, A., Widdel, F. 1994. Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. *Appl Environ Microbiol* 60:4517–4526.
- Ewing, B., Hillier, L., Wendl, M.C., Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
- Froyshov, O., Laland, S.G. 1974. On the biosynthesis of bacitracin by a soluble enzyme complex from *Bacillus licheniformis*. *Eur J Biochem* 46:235–242.
- Fründ, C., Priefert, H., Steinbüchel, A., Schlegel, H.G. 1989. Biochemical and genetic analyses of acetoin catabolism in *Alcaligenes eutrophus*. *J Bacteriol* 171:6539–6548.
- Gottschalk, G. 1986. *Bacterial Metabolism*, ed 2. New York, Springer.
- Gray, G.L., Mainzer, S.E., Rey, M.W., Lamsa, M.H., Kindle, K.L., Carmona, C., Requadt, C. 1986. Structural genes encoding the thermophilic α -amylases of *Bacillus stearothermophilus* and *Bacillus licheniformis*. *J Bacteriol* 166: 635–643.
- Helmann, J.D., Moran, C.P., Jr. 2002. RNA polymerase and sigma factors; in Sonenhein A.L. (ed): *Bacillus subtilis and Its Closest Relatives*. Washington, ASM Press.
- Hofmann, K., Stoffel, W. 1993. TMbase – A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler* 374:166.
- Hulo, N., Sigrist, C.J., Le Saux, V., Langendijk-Genevaux, P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., Bairoch, A. 2004. Recent improvements to the PROSITE database. *Nucl Acids Res* 32 Database issue, D134–D137.
- Huynen, M.A., Bork, P. 1998. Measuring genome evolution. *Proc Natl Acad Sci USA* 95:5849–5856.
- Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., et al. 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* 423:87–91.
- Jacobs, M., Eliasson, M., Uhlen, M., Flock, J.I. 1985. Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. *Nucl Acids Res* 13:8913–8926.
- Kakudo, S., Kikuchi, N., Kitadokoro, K., Fujiwara, T., Nakamura, E., Okamoto, H., Shin, M., Tamaki, M., Teraoka, H., Tsuzuki, H., et al. 1992. Purification, characterization, cloning, and expression of a glutamic acid-specific protease from *Bacillus licheniformis* ATCC 14580. *J Biol Chem* 267:23782–23788.
- Kandra, L., Gyemant, G., Remenyik, J., Hovanszki, G., Liptak, A. 2002. Action pattern and subsite mapping of *Bacillus licheniformis* α -amylase (BLA) with modified maltooligosaccharide substrates. *FEBS Lett* 518:79–82.
- Kim, I.C., Cha, J.H., Kim, J.R., Jang, S.Y., Seo, B.C., Cheong, T.K., Lee, D.S., Choi, Y.D., Park, K.H. 1992. Catalytic properties of the cloned amylase from *Bacillus licheniformis*. *J Biol Chem* 267:22108–22114.
- Konz, D., Doekel, S., Marahiel, M.A. 1999. Molecular and biochemical characterization of the protein template controlling biosynthesis of the lipopeptide lichenysin. *J Bacteriol* 181: 133–140.
- Kulikova, T., Aldebert, P., Althorpe, N., Baker, W., Bates, K., Browne, P., van den Broek, A., Cochrane, G., Duggan, K., Eberhardt, R., et al. 2004. The EMBL Nucleotide Sequence Database. *Nucl Acids Res* 32 Database issue, D27–D30.
- 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. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256.

- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
- Lathe, W.C. 3rd, Snel, B., Bork, P. 2000. Gene context conservation of a higher order than operons. *Trends Biochem Sci* 25:474–479.
- Lobry, J.R. 1996. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol Biol Evol* 13:660–665.
- Martinez, J.L., Herrero, M., de Lorenzo, V. 1994. The organization of intercistronic regions of the aerobactin operon of pColV-K30 may account for the differential expression of the *iucABCD iutA* genes. *J Mol Biol* 238:288–293.
- Merkel, R. 2004. Score-based identification of genomic islands. *BMC Bioinformatics* 5:22.
- Moss, J.E., Cardozo, T.J., Zychlinsky, A., Groisman, E.A. 1999. The *selC*-associated SHI-2 pathogenicity island of *Shigella flexneri*. *Mol Microbiol* 33:74–83.
- Oppermann, F.B., Steinbüchel, A., Schlegel, H.G. 1989. Evidence for oxidative thiololytic cleavage of acetoin in *Pelobacter carbinolicus* analogous to aerobic oxidative decarboxylation of pyruvate. *FEMS Microbiol Lett* 51:113–118.
- O'Sullivan, D., Twomey, D.P., Coffey, A., Hill, C., Fitzgerald, G.F., Ross, R.P. 2000. Novel type I restriction specificities through domain shuffling of HsdS subunits in *Lactococcus lactis*. *Mol Microbiol* 36:866–875.
- Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov, E. Jr., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., et al. 2003. The ERGO genome analysis and discovery system. *Nucl Acids Res* 31:164–171.
- Pearson, W.R. 1994. Using the FASTA program to search protein and DNA sequence databases. *Methods Mol Biol* 25:365–389.
- Rasko, D.A., Ravel, J., Okstad, O.A., Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., et al. 2004. The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nucl Acids Res* 32:977–988.
- Read, T.D., Peterson, S.N., Tourasse, N., Baillie, L.W., Paulsen, I.T., Nelson, K.E., Tettelin, H., Fouts, D.E., Eisen, J.A., Gill, S.R., et al. 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 423:81–86.
- Sandeman, R.A., Hynes, M.J., Fincham, J.R., Connerton, I.F. 1991. Molecular organisation of the malate synthase genes of *Aspergillus nidulans* and *Neurospora crassa*. *Mol Gen Genet* 228:445–452.
- Schallmeyer, M., Singh, A., Ward, O.P. 2004. Developments in the use of *Bacillus* species for industrial production. *Can J Microbiol* 50:1–17.
- Sharma, V., Sharma, S., Hoener zu Bentrup, K., McKinney, J.D., Russell, D.G., Jacobs, W.R. Jr., Sacchettini, J.C. 2000. Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nat Struct Biol* 7:663–668.
- Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., Hayashi, H. 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci USA* 99:996–1001.
- Smith, E.L., DeLange, R.J., Evans, W.H., Landon, M., Markland, F.S. 1968. Subtilisin Carlsberg. V. The complete sequence: Comparison with subtilisin BPN^o; evolutionary relationships. *J Biol Chem* 243:2184–2191.
- Sneath, P.H.A., Mair, N.S., Sharpe, M.E., Holt, J.G. 1986. *Bergey's Manual of Systematic Bacteriology*, vol 2. Baltimore, Williams & Wilkins.
- Staden, R., Beal, K.F., Bonfield, J.K. 2000. The Staden package, 1998. *Methods Mol Biol* 132:115–130.
- Stephens, M.A., Ortlepp, S.A., Ollington, J.F., McConnell, D.J. 1984. Nucleotide sequence of the 5' region of the *Bacillus licheniformis* α -amylase gene: Comparison with the *B. amyloliquefaciens* gene. *J Bacteriol* 158:369–372.
- Svendsen, I., Breddam, K. 1992. Isolation and amino acid sequence of a glutamic acid specific endopeptidase from *Bacillus licheniformis*. *Eur J Biochem* 204:165–171.
- Syed, R., Wu, Z.P., Hogle, J.M., Hilvert, D. 1993. Crystal structure of selenosubtilisin at 2.0-Å resolution. *Biochemistry* 32:6157–6164.
- Takami, H., Horikoshi, K. 2000. Analysis of the genome of an alkaliphilic *Bacillus* strain from an industrial point of view. *Extremophiles* 4:99–108.
- Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., et al. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucl Acids Res* 28:4317–4331.
- Tech, M., Merkel, R. 2003. YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol* 3:441–451.
- Tillier, E.R., Collins, R.A. 2000. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *J Mol Evol* 50:249–257.
- Tjalsma, H., Bolhuis, A., Jongbloed, J.D., Bron, S., van Dijk, J.M. 2000. Signal peptide-dependent protein transport in *Bacillus subtilis*: A genome-based survey of the secretome. *Microbiol Mol Biol Rev* 64:515–547.
- Torrents, E., Buist, G., Liu, A., Eliasson, R., Kok, J., Gibert, I., Graslund, A., Reichard, P. 2000. The anaerobic (class III) ribonucleotide reductase from *Lactococcus lactis*. Catalytic properties and allosteric regulation of the pure enzyme system. *J Biol Chem* 275:2463–2471.
- Yuuki, T., Nomura, T., Tezuka, H., Tsuboi, A., Yamagata, H., Tsukagoshi, N., Udaka, S. 1985. Complete nucleotide sequence of a gene coding for heat- and pH-stable α -amylase of *Bacillus licheniformis*: Comparison of the amino acid sequences of three bacterial liquefying α -amylases deduced from the DNA sequences. *J Biochem (Tokyo)* 98:1147–1156.