Supporting Information

Alcaraz *et al.* 10.1073/pnas.0800981105

Supporting Text

Strains and Media. *Bacillus coahuilensis* was isolated by the group of V.S. from water column in the desiccation pond named Laguna Grande of the Churince system. NRRLB14911 was isolated from water columns in the Gulf of Mexico by J.S. Marine medium (US Biological) was used for general growth purposes and a modified marine medium was used to determine amino acid requirements and growth under different phosphate concentrations. This modified marine medium contained 0.5% casein hydrolysate (Difco), 5.48 mM sorbitol, 0.41 mM Fe (III) citrate, 40 mM $MgSO_4{\cdot}7H_2O$, 0.33 M NaCl, 9 mM $CaCl_2$, 0.67 mM KCl, 0.67 mM KBr, 0.47 mM NaF, 0.35 mM H_3BO_3, 0.19 mM $(NH_4)NO_3$, 4.7 μg/ml vitamin B complex. The concentration of Na_2HPO_4 was varied as indicated. pH was adjusted to 7 with 10 M NaOH.

Genome Sequencing. The genomic DNA of M44 was isolated, by using a standard technique, from pure cells cultured in marine medium (US Biological). Shotgun libraries were prepared using *Escherichia coli* DH12S, a host suited for the stable cloning of methylated DNA (Invitrogen). The entire genome sequence was obtained from a combination of 16,698 end sequences (providing 6-fold coverage) from a pUC18 genomic shotgun library (2–5 kb) using dye terminator chemistry on automated DNA sequencers (ABI3700, Applied Biosystems) and 454 technology with seven runs and a 29-fold coverage. Synteny-guided gap closure for some contigs was performed by PCR direct sequencing using primers designed to anneal to each end of the neighboring contigs. Tail PCR was carried out from ends of contigs for which we had no synteny information. Fifteen scaffolds were assembled. The ends of 43 contigs consisted of highly similar repeat sequences and we did not attempt to determine their exact order because these sequences did not provide new gene information. A pseudogenome was assembled for the annotation process.

Genome Analysis. Two independent assemblies were performed, one assembly of the 454 by using its Newbler assembler (http://www.454.com/enabling-technology/the-software.asp) and one assembly of the Sanger reads by using Phred-Phrap (1) with default parameters in both cases. With the two sets of assemblies, all Vs. all alignments were performed with the MUMmer package (2) by using regions with a minimum overlap of 100 bp. A new consensus assembly of the hybrid sequences was conducted by using the CAP3 assembler (3) and manually curated. To assess the orientation and synteny conservation of the newly assembled fragments, we made Promer alignments (2) and plots versus *B. cereus* E33L, *B. cereus* ATTC 10987, *B. cereus* ATTC 14579, *B. anthracis* str. 'Ames,' *B. anthracis* str. 'Ames ancestor,' *B. anthracis* str. Sterne, *B. thuringiensis* serovar konkukian, *B. licheniformis* ATCC 14580, *B. halodurans* C-125, *Geobacillus kaustophilus* HTA426, *B. subtilis* str. 168, *Oceanobacillus iheyensis* HTE831, *B. clausii* KSM-K16, and *Bacillus* sp. NRRLB-14911.

Gene prediction was done by using Glimmer v3.02 (4) and GeneMark.hmm (5) for prokaryotes, followed by the automatic annotation pipeline using BASys system (6), which retrieves sequence information from SWISS PROT, InterPro, Pfam, PROSITE, COG, GO, KEGG, and NCBI NR databases, among other databases. Transfer RNAs were detected by using tRNAscan-SE (7). Annotations were checked manually and frameshift corrections were performed by using FrameD (8) along with the original reads. Extra annotation of small (<100 bp) ORFs of known function was done manually. Sequence manipulation, parsing, and statistical features like genome size, GC%, codon usage, and CAI were obtained through the EMBOSS 3.0 package (9) and Perl scripts.

Phylogenomics analyses were done according to Ciccarelli *et al.* (10) after first searching all of the universally conserved COGs across all of the sequenced *Bacillus* spp. resulting in a list of 20 genes (see supporting information (SI) Table S3). Each translated sequence was aligned individually first with MUSCLE (11) and then concatenated. Phylogeny was reconstructed using tree-puzzle (12) with the following criteria: quartet puzzling, approximate quartet likelihood, 1,000 puzzling steps, exact parameter estimates, WAG (13) model of substitution, and gamma-distributed rates estimated from the dataset.

Orthologues between the *Bacillus* spp. were obtained through best bidirectional blastp (14) with a cutoff value of $1e^{-10}$ of at least 70% the length of the product. Paralogous analysis was done as described by Pushker *et al.* (15). Metabolic pathway reconstruction was done by using KEGG Automated Annotation Server (http://www.genome.ad.jp/kegg/kaas/help.html).

HMM profiles (16) were built and calibrated for 51 translated genes related to the nitrogen cycle comprising all of the series of assimilatory and dissimilatory reactions involving both inorganic and organic forms of nitrogen. The protein sequences used for the HMM search included: AllC, Alr, ArcA, ArcC, ArgD, ArgF, ArgG, ArgH, AspA, CarB, Chi, ChiA, ChiD, ChiHal, Chitin, CreA, CycA, DagA, GdhA.arch, GdhA.bacsu, GdhA.ecol, GlnA, GlnB, GltB, Hcp, Hmp, HutH, IlvE, LdcA, LytC, NagZ, NapA, NarG, NarH, NarI, NarJ, NasAt, NirA, NorB, NorZ, NtcA, NtrC, PbpC, RgI, RocF, TrpC, TrpF, TyrB, UreA, UreB, and UreC.

ABC importers' gene families were searched for by means of building and calibrating HMM profiles for each of the import families deposited in the ABCISSE (17) database (http://www.pasteur.fr/recherche/unites/pmtg/abc/database.iphtml). We also performed Wilcoxon's Signed-Ranks Matched Pairs Test between the ABC importers within the genomes of *B. coahuilensis, NRRL, B. subtilis,* and *O. iheyensis* and found differences between all of them ($P < 0.05$).

To identify windows that contained regions with unusual composition properties (RUCPs) within *B. coahuilensis* genome the Similarity Plot (S-plot) (18) application was used. This method was recently presented as an alternative means for identifying potential horizontally transferred elements. To assess the degree and pattern of similarity (or dissimilarity) between two genomic sequences of size M_1 and M_2, the genomes are divided into windows of length w slide along each genome with steps (the distance between the start of two neighboring windows) of size s. Similarity is quantified by using the Pearson correlation coefficient between the frequencies of n-mers (short subsequences of length n). By first comparing the genome of interest against itself, the degree of homogeneity of the genome can be determined as the average correlation value across all windows. Next, the degree of similarity of each window with respect to its own genome is calculated as the average of the correlation coefficients for each window against all other windows in the genome for which it is located. Because it is our intent to identify foreign DNA, windows that are unusually dissimilar to the rest of their genome are of particular interest.

Each of the 115 RUCPs of *B. coahuilensis* is compared with 430 complete as well as partial sequenced microbial genomes available from the NCBI database by using S-plot. Included in these genomes are 10 *B. anthracis* (3 complete, 7 whole genome shotguns), 3 *B. cereus*, 1 *B. clausii*, 1 *B. halodurans*, 2 *B. licheniformis*, 1 *B. subtilis*, 1 *B. thuringiensis*, 1 *Geobacillus kaustophilus*, and 1 *Oceanobacillus iheyensis* genomes. For each RUCP the most similar window in all of these genomes was identified. Seven RUCPs were identified as having a matching window, >0.7 in one of these other microbial genomes indicating a highly correlated/similar sequence. These windows are likely to share the same functionality with the identified windows in the other microbial genomes. Therefore, these windows are not likely to have been introduced by HGT. The remaining 108 windows may have been introduced into the *B. coahuilensis* genome as a result of HGT.

Additionally, a second search round for HGT elements was performed by means of best bidirectional blastp (14) hit with an evalue of $1e^{-10}$ by using as query all of the genes of *B. coahuilensis* against a cyanobacterium and Archaea database retrieved from KEGG database including the following species (KEGG organism): syn, syw, syc, syf, syd, sye, syg, cya, cyb, tel, gvi, ana, ava, pma, pmm, pmt, pmn, pmi, pmb, pmc, pmf, pmg, pme, ter, mja, mmp, mmq, mac, mba, mma, mbu, mtp, mhu, mla, mem, mth, mst, mka, afu, hal, hma, hwa, nph, tac, tvo, pto, pho, pab, pfu, tko, ape, smr, hbu, sso, sto, sai, pai, pis, pcl, tpe, and neq. Results from the bidirectional blastp were filtered to exclude all genes shared between the *Bacillus* spp. to discriminate potential housekeeping genes and parse a list of unique genes shared between *B. coahuilensis,* cyanobacteria, and Archaea.

Retrieval of *B. coahuilensis* rhodopsin, SQD and SQDX orthologous sequences was done through a blastp (e value $1e^{-10}$, length 70%, ID $>30\%$) versus NR and environmental databases of NCBI. Multiple alignments were made by using ClustalW (19) with the BLOSUM62 matrix. Phylogenetic analyses were done by using MEGA 3.1 (20) and a neighbor-joining reconstruction with the following parameters: 1,000 replicates bootstrap (seed = 24,054), by using Poisson correction substitution model for amino acids, gaps using complete deletion, assuming independent evolution for each amino acid sequence, and pairwise deletion was used when comparing translated sequences from different organisms.

Homology modeling of *B. coahuilensis* SQD and rhodopsin with crystal structures from *Arabidopsis* and *Anabaena* sp. PCC 7120, respectively, was done by using the SWISS-MODEL web server (http://swissmodel.expasy.org) (21) by Deep View (Swiss-Pdb Viewer) (22). PDB templates used were 1I24 for SQD and 1XIO for rhodopsin. WhatCheck summary of SWISS-MODEL reports the following values for SQD and rhodopsins, respectively, were as follows: structure *Z*-scores: first generation packing quality: -0.178, -0.242; second generation packing quality -1.094, -0.127; Ramachandran plot appearance: -0.075, 0.255; chi-1/chi-2 rotamer normality: 1.221, 0.209; backbone conformation: -0.464, -0.198; and root mean square (rms) *Z*-scores: bond lengths: 0.690, 0.810; bond angles: 0.950, 1.214; omega angle restrains: 0.874, 0.686; side chain planarity: 1.302, 1.318; improper dihedral distribution: 1.541, 1.302; inside/outside distribution: 1.147, 1.302. Complete WhatCheck and coordinates reports are available on request. Diagram images were produced in Pymol Version 0.99rc6 (http://www.pymol.org)**.**

RT-PCR. Semiquantitative RT-PCRs were carried out by using SuperScript One Step RT-PCR with Platinum Taq (Invitrogen Life Technologies) at 20, 25 ,and 30 cycles. Oligos for bacteriorhodopsin *bsr*: 5′ TCGCTATGGTCATCCCGTTGTGG (forward); 5′ AGAGGGACCTAATAGCCATGCAG (reverse). Oligos for *sqd1:* 5′ TGCGCCTTACAGTATGATTGACC (forward) and 5′ AAGCCCTTGTTTGTTCTCCTGAT (reverse). RNA was obtained by using TRIzol (Invitrogen Life Technologies) from strains grown in modified marine medium supplemented with phosphate at 0.001, 0.005, 0.05. 0.5, and 5 mM. For light/dark experiments, strain was grown on Petri dishes with marine medium grown at 37°C either under white or blue light or in the dark.

Lipid Extraction and Analysis. Lipids from *Arabidopsis*, *Cyanobacteria* spp., and *B. coahuilensis* were isolated (details are available on request). Lipid extracts were observed and isolated by using the TLC technique as described (23). For lipid footprint analysis, individual lipids were isolated from TLC plates, duplicates of each lipid spot were analyzed by MALDI-TOF MS technology. Spots corresponding to SQDG were isolated and eluted with 4 volumes of CH_3Cl: CH_3OH (2:1 V/V) and 1 volume of 0.9% NaCl. The chloroformic fraction was extracted and evaporated under a constant N_2 stream and resuspended in 100 μl of $CH_3Cl/CH_3OH/CH_3Cl/CH_3OONa$ (300:665:35 V/V) and analyzed by electro spray ionization MS-MS.

1. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities*. Genome Res* 8:186–194.
2. Kurtz S, *et al.* (2004) Versatile and open software for comparing large genomes. *Genome Biol*. 5:R12.
3. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868–877.
4. Delcher AL, Bratke KA, Powers EC, Salzberg SL (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer *Bioinformatics* btm009.
5. Lukashin AV, Borodovsky M (1998) GeneMark.hmm: New solutions for gene finding. *Nucleic Acids Res* 26:1107–1115.
6. Van Domselaar GH, *et al.* (2005) BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res* 33:W455–W459.
7. Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
8. Schiex T, Gouzy J, Moisan A, de Oliveira Y FrameD (2003) A flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* 31:3738–3741.
9. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16:276–277.
10. Ciccarelli FD, *et al.* (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311:1283–1287.
11. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
12. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
13. Goldman N, Whelan S (2000) Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Mol Biol Evol* 17:975–978.
14. Altschul SF, *et al.* (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
15. Pushker R, Mira A, Rodriguez-Valera F (2004) Comparative genomics of gene-family size in closely related bacteria. *Genome Biol* 5:R27.
16. Bateman A, Haft DH (2002) HMM-based databases in InterPro Brief. *Bioinformatics* 3:236–245.
17. Bouige P, Laurent D, Piloyan L, Dassa E (2002) Phylogenetic and functional classification of ATP-binding cassette (ABC) systems. *Curr Protein Pept Sci* 3:541–559.
18. Putonti C, *et al.* (2006) A computational tool for the genomic identification of regions of unusual compositional properties and its utilization in the detection of horizontally transferred sequences. *Mol Biol Evol* 23:1863–1868.
19. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
20. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5:150–163.
21. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385.
22. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714–2723.
23. Hartel H, Dormann P, Benning C (2000) .DGD1-independent biosynthesis of extraplastidic galactolipids after phosphate deprivation in Arabidopsis *Proc Natl Acad Sci USA* 97:10649–10654.
24. Cerritos R, *et al.* (2008) Bacillus coahuilensis sp. nov. A new moderately halophilic species from different pozas in the Cuatro Ciénegas Valley in Coahuila, México. *Int J Syst Evol Microbiol* 58:919–923.

Fig. S1. Phylogenetic analysis of the 16S rRNA from *B. coahuilensis* and related species. Neighbor-oining tree of 16S rRNA using Kimura's two-parameter substitution model and 1,000 bootstrap replicates. Shown are the sequenced *B. coahuilensis* (m4–4) as well as other *B. coahuilensis* isolates. Most of the strains shown in the phylogeny were isolated from marine environments, marine sediments, and estuaries. The exceptions are *Bacillus* sp. 19500 and *Bacillus* sp. 8-gw2-7 isolated from a mural paintings tomb in Seville, Spain, and from freshwater in Michigan, respectively. The accession numbers of the strains used in this analysis are: 54303774, 50235228, 118577870, 5524657, 92091064, 15420442, 116266516, 75991537, 16973341, otherwise sequences were retrieved from the whole genome sequence.

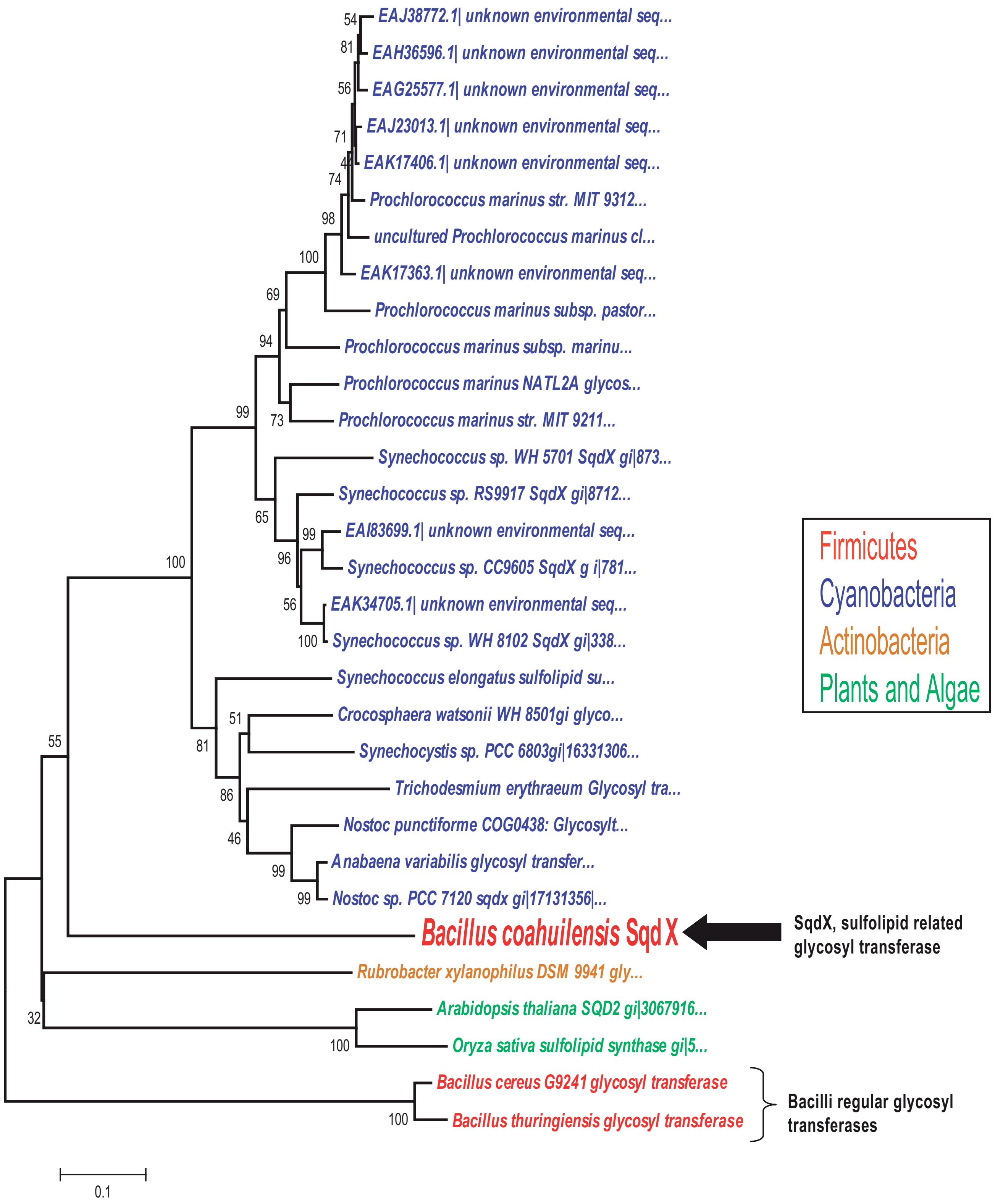


Fig. S2. Phylogenetic analysis of the glycosyltransferase coded in the *B. coahuilensis sqd* operon. Despite the presence in bacteria of different glycosyltransferases, we show that the glycosyltransferase SqdX in the *B. coahuilensis* SQD1 operon is phylogenetically closer to the plant and cyanobacterial proteins than it is to the bacterial glycosyltransferases, giving support to the horizontal transfer of the SQD1 operon.

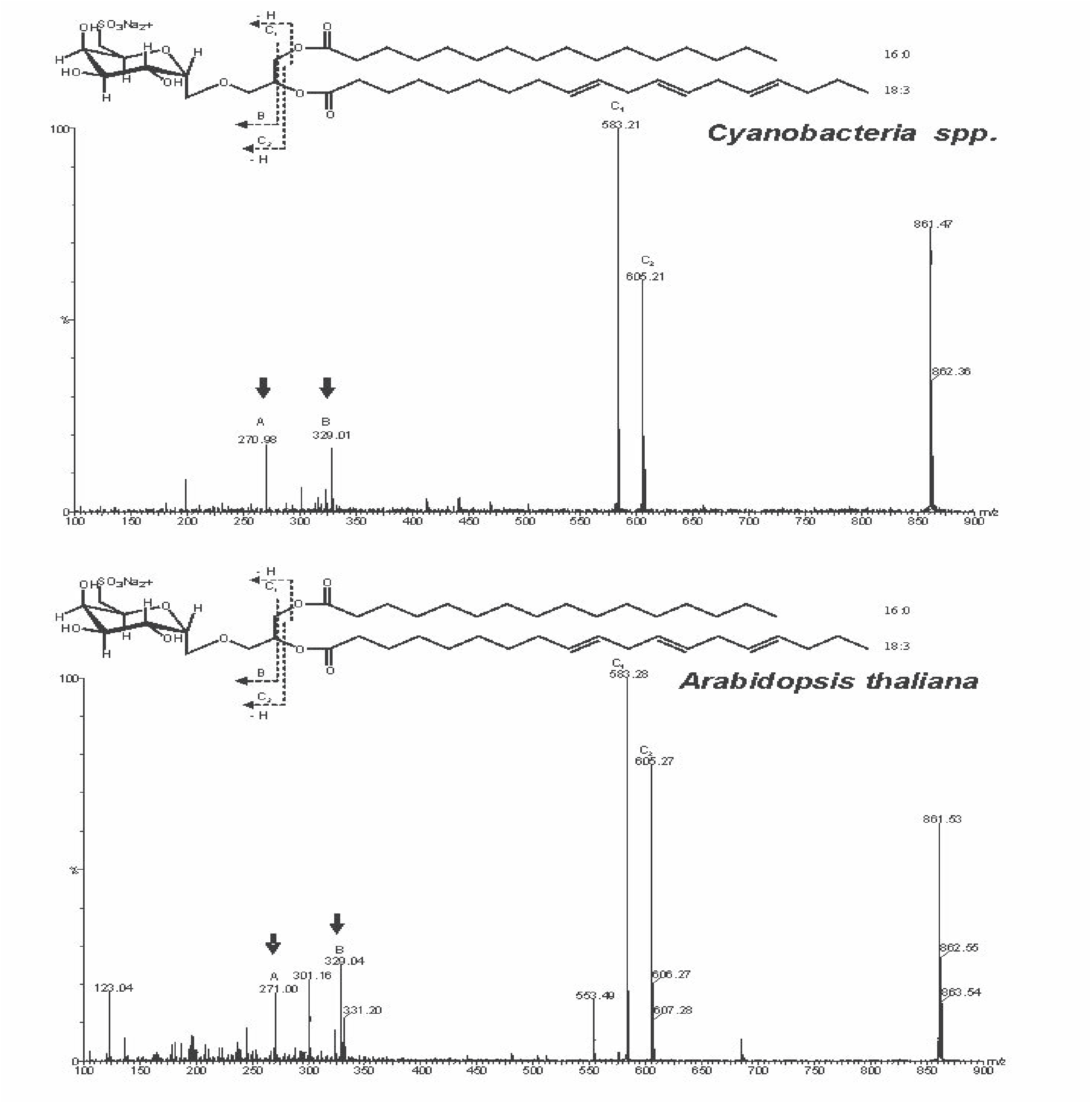
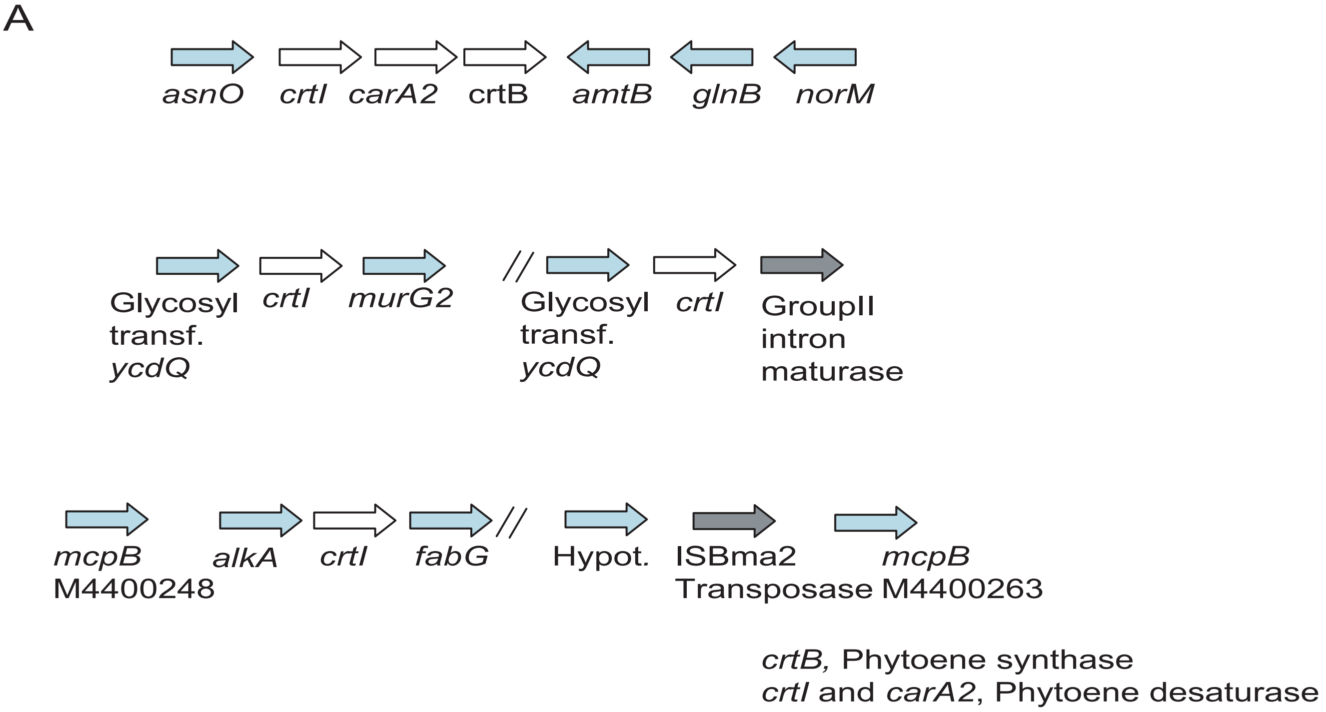


Fig. S3. ESI MS-MS mass spectrum of TLC-isolated SQDG spots from *Cyanobacteria* spp. and *A. thaliana* lipid extracts. The proposed cleavage pathways are shown above each spectrum. All fragments contain sodium. For the cleavages of the fatty acyl chains, the subscript number in the symbol represents the relative position (C1 and C2) in the fatty acyl group. Arrows signaling A and B indicate specific spectrums for sulfoquinovose moieties.



B

Methyl-accepting chemotaxis proteins in *B. coahuilensis*

M4400025	*mcpC*	M4402078	*tlpB*
M4400248	*mcpB*	M4402148	*mcpB*
M4400263	*mcpB*	M4402589	*mcpB*
M4400630	*tlpA*	M4402800	*tlpA*
M4400687	*mcpB*	M4402859	*tlpB*
M4400745	*tlpB*	M4403081	*mcpB*
M4401183	*mcpB*	M4403083	*mcpA*
M4401184	*mcpC*	M4403306	*mcpC*
M4401682	*tlpB*		

Chemotaxis response proteins

M4401449	cheB	Response regulator protein-glutamate methylesterase
M4401450	cheA	Two-component sensor histidine kinase
M4401452	cheC	Chemotaxis protein cheC
M4401861	cheR	Chemotaxis protein methyltransferase

Fig. S4. Carotenoid synthesis and methyl-accepting chemotaxis protein (MCP) coding genes in *B. coahuilensis.* (*A*) Carotenoid synthesis genes are distributed in the *B. coahuilensis* genome. One operon contains both a synthase (*carB*) and desaturase (*crtI*) genes. Other *crtI* genes are found at different locations. Two of them are close to genes encoding transposases, suggesting that these were acquired through HGT. (*B*) Methyl-accepting proteins in the genome of *B. coahuilensis*. Some sensory rhodopsins are known to transduce their signal through MCPs. *B. coahuilensis* has 17 MCP coding genes. We do not know, however, whether these are involved in the phototransduction signaling.

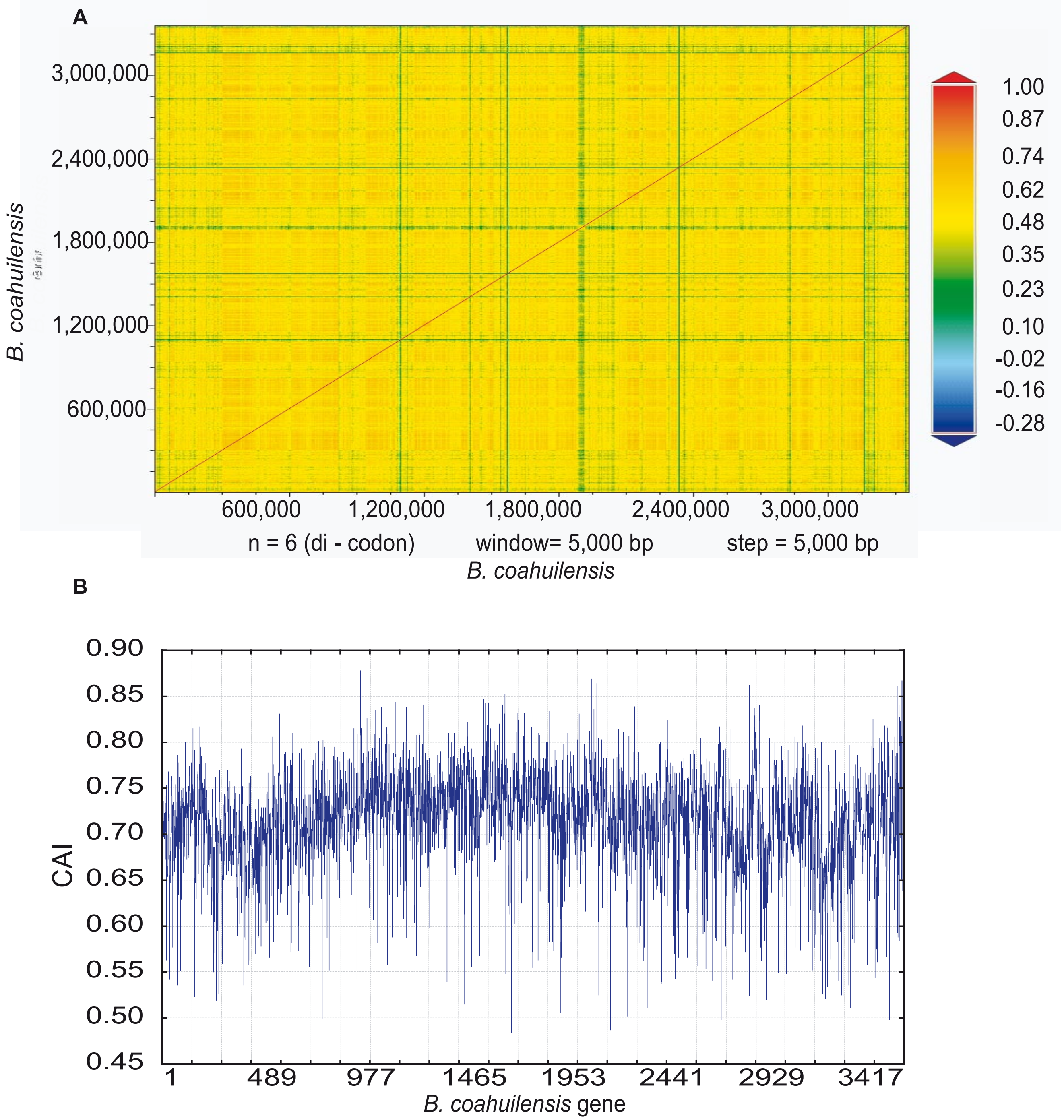
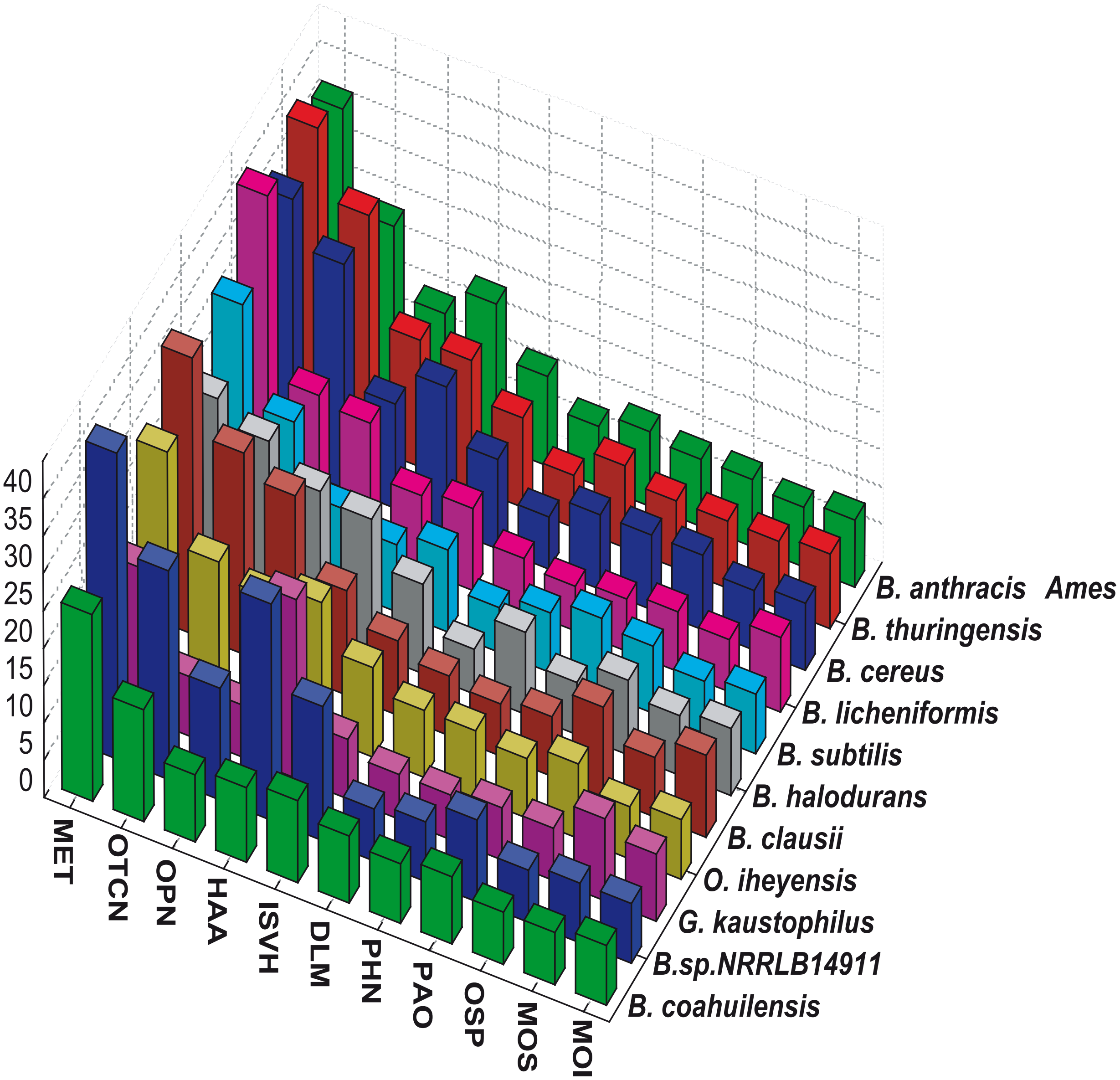


Fig. S5. Nucleotide composition analysis and Codon Adaptation Index analysis to detect Horizontal Gene Transfer events. (*A*) S-plot for *B. coahuilensis* versus itself. Different Pearson correlation coefficients are represented on the plot by different colors. The plot leads to the identification of Regions of Unusual Composition (RUCPs) (see *SI Text*.) within the genome of *B. coahuilensis*. (*B*) Codon Adaptation Index (CAI) of each predicted ORF of the genome of *B.coahuilensis*. The average CAI is 0.7147 ± 0.0537. An underaverage CAI could reflect recent insertion into the genome or function restrictions. Over average CAI probably means an adaptation to an effective transcription/translation rate.



MET	Metals
OTCN	Osmoprotectans Taurine Cyanate and Nitrate
OPN	Oligopeptides and Nickel
HAA	Hydrophobic amino acids and amides
ISVH	Iron-Siderophores Vitamin B-12 and Hemin
DLM	D- L-Methionine and derivates
PHN	Phosphonates and phosphites
PAO	Polar amino acid and opines
OSP	Oligosaccharides and polyols
MOS	Monosaccharides
MOI	Mineral and Organic

Fig. S6. Distribution of ABC importer families in the *Bacillus* spp. and closely related species. ABC importer gene families were searched by means of building and calibrating HMM profiles for each of the import families deposited in the ABCISSE database (http://www.pasteur.fr/recherche/unites/pmtg/abc/database.iphtml) to detect these genes in all of the sequenced *Bacillus* spp. Bars in different colors denote the presence of a gene predicted to code for a given importer with the height representing the number of genes present for any given category (see *SI Text*).

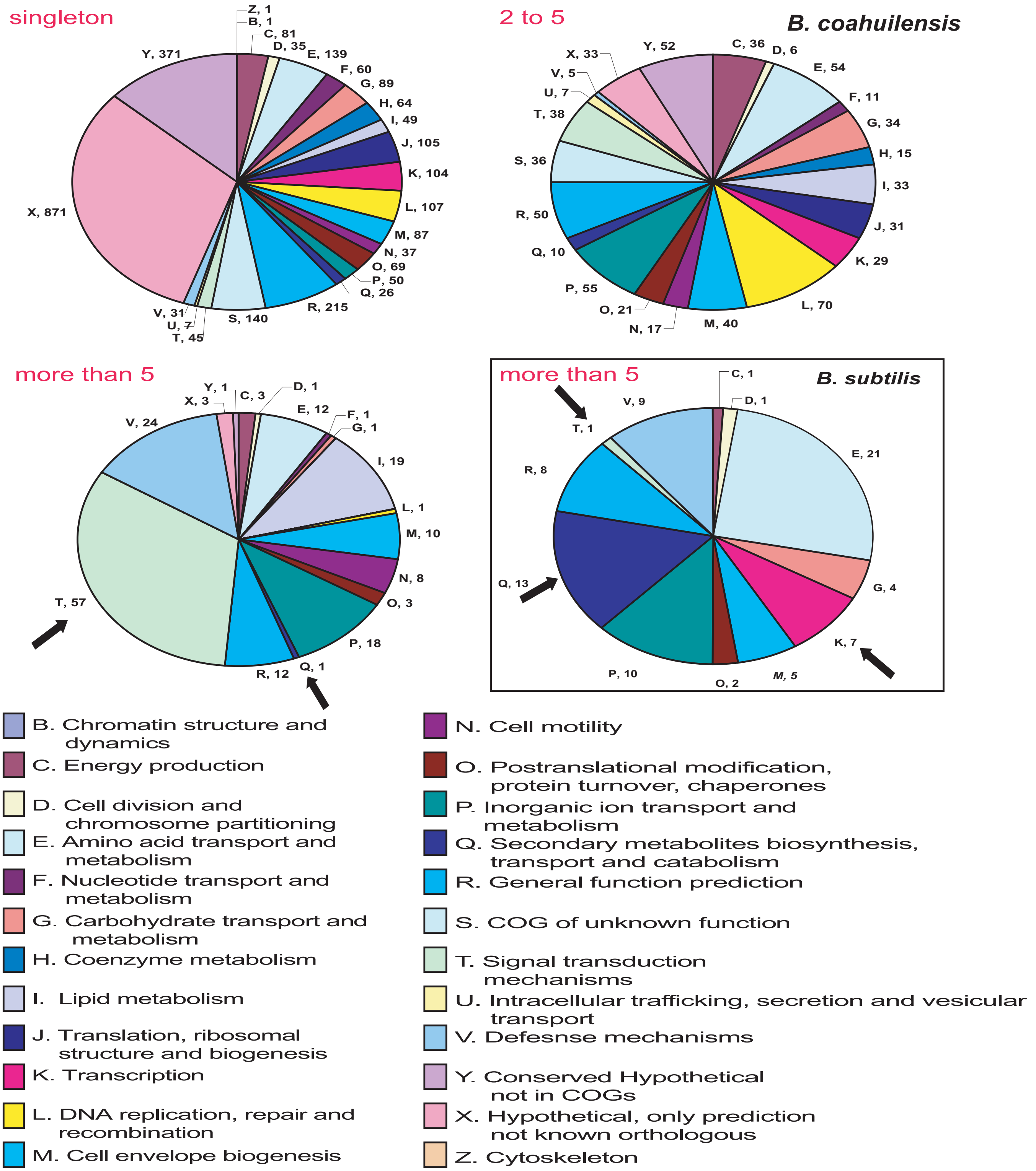


Fig. S7. Comparison of the distribution of paralogous genes in *B. subtilis* and *B. coahuilensis.* The analysis of paralogous genes in *B. coahuilensis* and *B. subtilis* was done as in Pushker *et al.* (15)

Table S1. Genomic sequencing results

Runs	Reads	Bases	Contigs	Total length, Mb	Average length, kb	N50, kb
1	158,013	16,348,073	1390	1.361	1	1
2	316,756	32,921,912	853	3.15	3.7	6.8
3	471,089	48,929,738	242	3.277	13.5	29.5
4	630,975	65,562,837	133	3.287	24.7	47.2
5	792,282	82,130,316	112	3.29	29.4	53.5
6	965,789	99,627,638	107	3.289	30.7	56.3
7	1,294,112	136,203,848	107	3.321	31.54	62.98

Assembly statistics	Sanger	454	454 + Sanger
Reads	16,698	1,294,112	-
Sequenced bases	20,709,240	136,203,848	-
Total Mb (assembly)	2.492	3.321	3.351
Number of contigs	876	107	73
Shortest contig, bp	46	498	1,379
Longest contig, bp	37,164	136,892	256,258
Non-ATCG bases (assembly)	1,803	0	284

We used a hybrid 454/Sanger sequencing strategy and in this table we provide a summary of the assembly data.

Table S2. Comparison of genomic features among sequenced *Bacillus* spp.

Strain	Ref. sequence	GenBank	Length, Mbp	GC content, %	Proteins	RNAs
Bacills coahuilensis str. M44	ABFU00000000	ABFU00000000	3.358	37.5	3,640	87
Bacillus sp. NRRL B-14911	NZ_AAOX00000000	AAOX00000000	5.086	45	5,691	106
Oceanobacillus iheyensis HTE831	NC_004193	BA000028	3.63	35	3,500	92
Geobacillus kaustophilus HTA426	NC_006510	BA000043	3.545	52	3,498	114
Bacillus subtilis subsp. *subtilis* str. 168	NC_000964	AL009126	4.215	43	4,105	119
Bacillus clausii KSM-K16	NC_006582	AP006627	4.304	44	4,096	96
Bacillus licheniformis ATCC 14580*	NC_006270	CP000002	4.222	46	4,152	93
Bacillus licheniformis ATCC 14580†	NC_006322	AE017333	4.223	46	4,196	93
Bacillus halodurans C-125	NC_002570	BA000004	4.202	43	4,066	105
Bacillus anthracis str. Ames	NC_003997	AE016879	5.227	35	5,311	128
Bacillus anthracis str. Sterne	NC_005945	AE017225	5.229	35	5,287	128
Bacillus cereus ATCC 10987	NC_003909	AE017194	5.224	35	5,603	133
Bacillus cereus ATCC 14579 project at INRA	NC_004722	AE016877	5.412	35	5,234	142
Bacillus cereus E33L	NC_006274	CP000001	5.301	35	5,134	135
Bacillus thuringiensis serovar *konkukian* str. 97–27	NC_005957	AE017355	5.238	35	5,117	144

*Novozymes Biotech.

†Project at Gottingen Genom. Lab.

Table S3. Twenty universally distributed Clusters of Orthologous Groups (COGs) used in the phylogenomic analysis

ID	Av. length	Annotation
COG0018	548	Arginyl-tRNA synthetase
COG0049	182	Ribosomal protein S7
COG0052	240	Ribosomal protein S2
COG0080	154	Ribosomal protein L11
COG0081	230	Ribosomal protein L1
COG0087	288	Ribosomal protein L3
COG0092	240	Ribosomal protein S3
COG0094	182	Ribosomal protein L5
COG0096	131	Ribosomal protein S8
COG0097	177	Ribosomal protein L6P/L9E
COG0098	220	Ribosomal protein S5
COG0100	145	Ribosomal protein S11
COG0172	442	Seryl-tRNA synthetase
COG0200	166	Ribosomal protein L15
COG0201	445	Preprotein translocase subunit SecY
COG0202	323	DNA-directed RNA polymerase, alpha subunit
COG0256	178	Ribosomal protein L18
COG0495	854	Leucyl-tRNA synthetase
COG0522	199	Ribosomal protein S4 and related proteins
COG0533	375	Metal-dependent proteases with chaperone activity

Table S4. ABC Importers proportion in a given category per genome and normalization relative to genome size

	Bsu	*Bha*	*Ban*	Oih	*Bth*	*Bli*	*Bce*	*Gka*	*Bcl*	*Bco*	*B14911*	Total
MET	28	21	32	25	35	37	31	15	32	20	36	312
OTCN	15	18	19	13	26	13	25	3	22	10	23	187
OPN	6	14	10	11	12	12	9	2	19	4	10	109
HAA	4	13	14	13	12	5	14	19	9	5	24	132
ISVH	6	7	7	7	7	6	7	3	5	6	13	74
DLM	1	1	3	4	2	2	2	1	3	4	2	25
PHN	3	6	5	4	6	1	5	1	2	3	3	39
PAO	5	2	4	3	4	2	5	2	3	4	6	40
OSP	4	5	4	5	4	3	5	2	7	2	2	43
MOS	2	3	3	2	4	2	3	6	3	2	3	33
MOI	3	4	4	3	5	5	4	4	6	3	3	44
TOTAL	77	94	105	90	117	88	110	58	111	63	125	1038
Genome size	4.22	4.2	5.23	3.63	5.24	4.22	5.22	3.54	4.3	3.36	5.09	
MET	6.63	5.00	6.12	6.89	6.68	8.76	5.93	4.23	7.43	5.95	7.08	
%	36.36	22.34	30.48	27.8	29.91	42.04	28.20	25.86	28.83	31.74	28.80	
OTCN	3.55	4.28	3.63	3.58	4.96	3.07	4.78	0.84	5.11	2.98	4.52	
%	19.48	19.15	18.09	14.4	22.22	14.77	22.73	5.17	19.82	15.87	18.40	
OPN	1.42	3.33	1.91	3.03	2.29	2.84	1.72	0.56	4.41	1.19	1.96	
%	7.79	14.89	9.52	12.2	10.25	13.63	8.18	3.45	17.11	6.35	8.00	
HAA	0.95	3.09	2.68	3.59	2.29	1.18	2.68	5.36	2.09	1.49	4.72	
%	5.19	13.83	13.33	14.4	10.25	5.68	12.73	32.76	8.11	8.11	19.20	
ISVH	1.42	1.66	1.34	1.93	1.33	1.42	1.34	0.84	1.16	1.79	2.55	
%	7.79	7.44	6.66	7.77	5.98	6.82	6.36	5.17	4.50	9.52	10.40	
DLM	0.24	0.24	0.57	1.10	0.38	0.47	0.38	0.28	0.70	1.19	0.39	
%	1.30	1.06	2.86	4.44	1.71	2.27	1.82	1.72	2.70	6.35	1.60	
PHN	0.71	1.43	0.96	1.10	1.14	0.24	0.96	0.28	0.46	0.89	0.59	
%	3.89	6.39	4.76	4.44	5.12	1.13	4.54	1.72	1.80	4.76	2.40	
PAO	1.18	0.47	0.76	0.82	0.76	0.47	0.96	0.56	0.70	1.19	1.18	
%	6.49	2.13	3.81	3.33	3.42	2.27	4.54	3.45	2.70	6.35	4.80	
OSP	0.95	1.19	0.76	1.38	0.76	0.71	0.96	0.56	1.62	0.59	0.39	
%	5.19	5.32	3.81	5.55	3.42	3.41	4.54	3.45	6.30	3.17	1.60	
MOS	0.47	0.71	0.57	0.55	0.76	0.47	0.57	1.69	0.70	0.59	0.59	
%	2.59	3.19	2.86	2.22	3.42	2.27	2.73	10.34	2.70	3.17	2.40	
MOI	0.71	0.95	0.76	0.82	0.95	1.18	0.76	1.13	1.39	0.89	0.59	
%	3.89	4.25	3.81	3.33	4.27	5.68	3.63	6.90	5.40	4.76	2.40	

Normalized data: No. of transporters of a given class per genome size. %, transporters of a given class as percent of total number of transporters in that species. *Bsu*, *B. subtilis*; *Bha*, *B. halodurans*; *Ba*, *B. anthracis* AMES; Oih, *O. iheyensis*; *Bth*, *B. thuringensis*; *Bli*, *B. licheniformis*; *Bce*, *B. cereus*; *Gka*, *G . kaustophilus*; *Bcl*, *B. clusii*; *Bco*, *B. coahuilensis*; *B14911*, *B*. sp. NRRLB 14911.

Table S5. Growth requirements of *B. coahuilensis* and *Bacillus* sp. NRRLB14911

Amino acid lacking	*B. coahuilensis*	NRRLB14911
L-Alanine	+	+
L-Aspartic	+	+
L-Glutamic	−/+	+
L-Asparagine	+	+
L-Glutamine	+	+
L-Arginine	−/+	+
L-Proline	−	−/+
L-Cysteine	+	+
L-Glycine	+	+
L-Serine	−	+
L-Lysine	+	+
L-Methionine	−/+	−/+
L-Threonine	−	+
L-Isoleucine	−/+	+
L-Leucine	−/+	+
L-Valine	−	+
L-Phenylalanine	−	+
L-Tryptophane	−	+
L-Tyrosine	−	+
L-Histidine	−	+
All amino acids added	+	+
No amino acids added	−	−/+

Modified marine medium containing the stated amino acids was inoculated with *B. coahuilensis* or NRRLB14911 and cultured at 37°C with agitation in nephelometric flasks. Absorbance was measured with a Klett-Summerson colorimeter at 24 and 48 h.

Table S6. Comparative analysis of the presence in *B. coahuilensis* orthologs for selected sporulation, germination, and competence genes

B. subtilis	Gene name	Function	*B. coahuilensis* ortholog	Orthologs in other *Bacillus* (when absent in *B. coahuilensis*)
Sporulation initiation				
BSU13990 BSU31450 BSU14490 BSU13660 BSU13530	*kinA* *kinB* *kinC* *kinD* *kinE*	two-component sensor histidine kinases	M4401017 (*kinA*) M4401019 (*kinC*) M4401702 (*kinA*) M4402663 (*kinA*) M4402848 (*kinA*) M4400900 (*kinB*) M4400627 (*kinC*)	
BSU16170	*codY*	transcriptional repressor CodY	M4401419	
BSU00370	*abrB*	transcriptional regulator	M4403643	
BSU12430 BSU36690 BSU03770 BSU36380 BSU25830 BSU37460 BSU40300 BSU06830 BSU05010 BSU02820 BSU18910	*rapABC DEFGHI*	response regulator aspartate phosphatase	Absent	Bs (11) Bl (24) Bcg ( ) Bh (5) Bcl (9)
Sporulation sigma factors				
BSU23450	*sigF*	sporulation sigma factor SigG	M4401792	
BSU15330	*sigG*	sporulation sigma factor SigE	M4401332	
BSU15320	*sigE (spoIIGB)*	RNA polymerase sporulation specific sigma factor (sigma-K) (C-terminal half)	M4401330	
BSU26390	*sigK (spoIIIC)*	sporulation sigma factor SigK	M4402263	
BSU00980	*sigH*	Transition state sigma factor SigH	M4400132	
spo0				
BSU24220	*spo0A*	two-component response regulator	M4402096	
BSU27930	*spo0B*	sporulation initiation phosphotransferase	M4402284	
BSU13640	*spo0E*	negative sporulation regulatory phosphatase	Absent	Bs Bl NR
BSU37130	*spo0F*	two-component response regulator	M4403280	
BSU40960	*spo0J*	site-specific DNA-binding protein	M4400017	
BSU11430	*spo0K (oppA)*	oligopeptide ABC transporter (binding protein)	Similar to BM4400987 (*dppE*)	
BSU08760	*spo0M*	sporulation-control gene	M4401200	
spoII				
BSU23470	*spoIIAA*	anti-anti-sigma factor	M4401790	
BSU23460	*spoIIAB*	anti-sigma F factor	M4403657	
BSU28060	*spoIIB*	required for complete dissolution of the asymmetric septum (stage II sporulation)	Absent	Bs Bl Bcg Bh NR
BSU36750	*spoIID*	serine phosphatase	M4403330	
BSU00640	*spoIIE*	protease	M4400073	
BSU15310	*spoIIGA*	required for dissolution of the septal cell wall (stage II sporulation)	M4401329	
BSU23530	*spoIIM*	required for dissolution of the septal cell wall (stage II sporulation)	M4401782	
BSU25530	*spoIIP*	required for completion of engulfment	M4402279	
BSU36550	*spoIIQ*	required for processing of pro-sigma-E (extracellular signal interacting with SpoIIGA?) (stage II sporulation)	BM4403333 (*yebA*)	
BSU36970	*spoIIR*	required for processing of pro-sigma-E	M4403297	
BSU12830	*spoIISA*	sporulation protein IISA	Absent	Bs Bl Bcl
BSU12820	*spoIISB*	sporulation protein IISB	Absent	Bs Bl
spoIII				

B. subtilis	Gene name	Function	*B. coahuilensis* ortholog	Orthologs in other *Bacillus* (when absent in *B. coahuilensis*)
BSU24430	*spoIIIAA*	mutants block sporulation after engulfment	M4402070	
BSU24420	*spoIIIAB*	stage III sporulation protein SpoAB	M4402071	
BSU24400	*spoIIIAD*	mutants block sporulation after engulfment (stage III sporulation)	M4402072	
BSU24390	*spoIIIAE*	mutants block sporulation after engulfment (stage III sporulation)	M4402073	
BSU24380	*spoIIIAF*	mutants block sporulation after engulfment (stage III sporulation)	M4402074	
BSU24370	*spoIIIAG*	mutants block sporulation after engulfment (stage III sporulation)	M4402076	
BSU24360	*spoIIIAH*	mutants block sporulation after engulfment (stage III sporulation)	M4402077	
BSU41040	*spoIIIJ*	OxaA-like protein precursor	M4400009 (*oxaI*)	
BSU36420	*spoIIID*	transcriptional regulator	M4403334	
BSU16800	*spoIIIE*	DNA translocase	M4401633 (*ftsK*)	
BSU41030	*jag*	SpoIIIJ-associated protein	M4400010	
spoIV				
BSU22800	*spoIVA*	required for proper spore cortex formation and coat assembly (stage IV sporulation)	M4401851	
BSU24230	*spoIVB*	serine peptidase of the SA class	M4402095	
BSU25770	*spoIVCA*	site-specific DNA recombinase		
	spoIVFA	stage IV sporulation protein FA	M4402368	
BSU27970	*spoIVFB*	membrane metalloprotease	M4402369	
BSU00230	*bofA*	inhibition of the pro-sigma-K processing machinery	Similar to BM4400117 and BM4400031	Bs Bl Bcg Bcl Bh Gk Oi NR
spoV				
BSU23440	*spoVAA*	sporulation protein VAA	M4401793	
BSU23430	*spoVAB*	sporulation protein VAB	Absent	Bs Bl Bcg Oi Gk NR
BSU23420	*spoVAC*	sporulation protein VAC	M4401923	M4403660
BSU23410	*spoVAD*	stage V sporulation protein SpoVAD	M4401924	M4403661
BSU23400	*spoVAE*	sporulation protein VAE	M4401795	
BSU23390	*spoVAF*	sporulation protein VAF	M4401796 and M4402028	
BSU27670	*spoVB*	involved in spore cortex synthesis (stage V sporulation)	M4402312	
BSU15170	*spoVD*	penicillin-binding protein	M4401140	
BSU15210	*spoVE*	required for spore cortex peptidoglycan synthesis (stage V sporulation)	M4401145	
BSU00490	*spoVG*	required for spore cortex synthesis	M4400089	
BSU17420	*spoVK*	sporulation protein VK	M4403653	
BSU15810	*spoVM*	required for normal spore cortex and coat synthesis (stage V sporulation)	M4403652	
BSU09400	*spoVR*	involved in spore cortex synthesis (stage V sporulation)	M4400764	
BSU16980	*spoVS*	required for dehydratation of the spore core and assembly of the coat (stage V sporulation)	M4401524	
BSU00560	*spoVT*	transcriptional regulator	M4400082	
BSU28110	*spoVID*	required for assembly of the spore coat (stage VI sporulation)	M4402335	
BSU00430	*yabG*	hypothetical protein	M4400097	
BSU04110	*ycsK*	hypothetical protein	M4400277	
BSU31470	*kapD*	sporulation inhibitor KapD	M4402705	
BSU23190	*dacB*	D-alanyl-D-alanine carboxypeptidase	M4401803	
Spore coat				
BSU06300	*cotA*	spore coat protein (outer)	Absent	Bs Bl Oi Bcl (Similar multicopper oxidases in Bh Gk Ba)
BSU36050	*cotB*	spore coat protein (outer)	Absent	Bs Gk Bcg
BSU17700	*cotC*	spore coat protein (outer)	Absent	Bs
BSU22200	*cotD*	spore coat protein (inner)	M4403659	

B. subtilis	Gene name	Function	*B. coahuilensis* ortholog	Orthologs in other *Bacillus* (when absent in *B. coahuilensis*)
BSU17030	*cotE*	morphogenic protein	M4401542 (*cotE*)	Bs Bl Bcg Bh Bcl Oi NR
BSU40530	*cotF*	spore coat protein	Similar to BM4403540 (*yhcQ*)	Bs Bl Bce Bt NR
BSU36070	*cotG*	morphogenetic protein	Absent	Bs
BSU36060	*cotH*	spore coat protein (inner)	Absent	Bs Bl Bcg
BSU06890	*cotJA*	polypeptide composition of the spore coat; required for the assembly of CotJC	M4401829 (*cotJA*)	Bs Bl Bcg NR
BSU06900	*cotJB*	polypeptide composition of the spore coat	M4403658 (*cotJB*)	Bs Bl Bcg Gk NR
BSU06910	*cotJC*	polypeptide composition of the spore coat	M4401828 (*cotJC*)	Bs Bl Bcg NR
BSU17970	*cotM*	spore coat protein (outer)	Similar to BM4401693	Bs Bl NR
BSU24620	*cotN (tasA)*	spore coat protein	BM4402955 (*cotN*)	Bs NR
BSU05550	*cotP*	spore coat protein	Absent	Bs NR
BSU30900	*cotS*	spore coat protein	Similar to M4403199	Bs Bl Bcg Bh NR
BSU30910	*cotSA*	spore coat protein	Absent	Similar glycosyltransferases in Bs Bl Bcg Gk Bh NR
BSU12090	*cotT*	spore coat protein (inner)	Absent	Bs
BSU11780	*cotV*	spore coat protein (insoluble fraction)	Absent	Bs Bl
BSU11770	*cotW*	spore coat protein (insoluble fraction)	Absent	Bs Bl
BSU11760	*cotX*	spore coat protein (insoluble fraction)	M4403656	Bs Bl Bce Bh Bcl Oi NR
BSU11750 BSU11740	*cotY cotZ*	spore coat protein (insoluble fraction)	Absent	Bs Bl Bcg Oi
BSU27830	*coxA*	spore cortex protein	Absent	Bs Bl
BSU37910	*spsA*	spore coat polysaccharide synthesis	Absent	Bs Bh
BSU37900	*spsB*	spore coat polysaccharide synthesis	Absent	Bs
BSU37890	*spsC*	spore coat polysaccharide synthesis	Absent	Bs Bl Bcg Gk NR
BSU37880	*spsD*	spore coat polysaccharide synthesis	Absent	Bs
BSU37870	*spsE*	spore coat polysaccharide synthesis	Absent	Bs Gk
BSU37860	*spsF*	spore coat polysaccharide synthesis	Absent	Bs
BSU37850	*spsG*	spore coat polysaccharide synthesis	Absent	Bs
BSU37840	*spsI*	spore coat polysaccharide synthesis (glucose-1 phosphate thymidylyltransferase)	Absent	Bs Bcg Bcl Bh Gk NR
BSU37830	*spsJ*	spore coat polysaccharide synthesis (dTDP glucose 4,6-dehydratase/epimerase)	Similar to BM4403005	Bs Bcg Bcl Gk NR
BSU37820	*spsK*	spore coat polysaccharide synthesis (dTDP-4-dehydrorhamnose reductase	Similar to BM4400707 and BM4403219 *strL*	Bs Bcg Bcl Bh Oi NR
BSU37810	*spsL*	spore coat polysaccharide synthesis (dTDP-4-dehydrorhamnose epimerase)	Absent	Bs Bcg Bh Bcl Oi NR
BSU02070	*csgA*	sporulation-specific SASP protein	Absent	Bs Bl NR
BSU22850	*seaA*	involved in spore envelope assembly	M4401847 (*yphB*)	Bs Bl Oi NR
BSU29570 BSU09750 BSU19950 BSU13470	*sspA sspB* *sspC sspD*	small acid-soluble spore protein (alpha/beta-type SASP)	M4403662 (*sspA*)	Bs (4) Bl (6) Bc (6) Bt (7) Ban (6) Bcl(3) Bh (3) Gk (1) Oi (1) NR (5)
BSU08660	*sspE*	small acid-soluble spore protein (gamma-type SASP)	BM4400675 (hypothetical)	Bs Bl Bcg Bcl Bh Oi NR
BSU32640	*sspG*	small acid-soluble spore protein	Absent	Bs
BSU00450	*sspF*	small acid-soluble spore protein (alpha/beta-type SASP)	M4400095 (*sspF*)	
BSU28660	*sspI*	small acid-soluble spore protein SspI	M4402408 (*sspI*)	
BSU33340	*sspJ*	small acid-soluble spore protein	Absent	Bs Bl
BSU22000	*sspL*	small acid-soluble spore protein	Absent	Bs Bl
BSU22290	*sspM*	small acid-soluble spore protein	Absent	Bs Bl
BSU18020	*sspN*	small acid-soluble spore protein	M4403655	
BSU17990	*sspO*	acid-soluble spore protein O	M4401690	
BSU18030	*tlp*	Tlp spore cortex-lytic enzyme	M4403654 (*tlp*)	
BSU22930	*sleB*	spore cortex-lytic enzyme	M4401837 (*sleB*)	
BSU13820	*ykvT*	hyp hypothetical proteinothetical protein	M4402316	
BSU23170	*spmB*	spore maturation protein	M4401806 (*spmB*)	
BSU23180	*spmA*	spore maturation protein	M4401805 (*spmA*)	
BSU13930	*splB*	spore photoproduct (thymine dimer) lyase	M4400797 (splB)	

B. subtilis	Gene name	Function	*B. coahuilensis* ortholog	Orthologs in other *Bacillus* (when absent in *B. coahuilensis*)
Germination				
		(earliest stage)	M4400597 (*gerAB*) M4401291(*gerAB*), M4401292, M4401293 (*gerIA*)	
BSU33070	*gerAC*	germination response to L-alanine and related amino acids (earliest stage)	BM4403467 (MLA incompleto)	
BSU35800	*gerBA*	probable component of a germinant receptor	Absent	Bs
BSU35810	*gerBB*	probable component of a germinant receptor	Absent	Bs
BSU35820	*gerBC*	germination response to L-alanine and to the combination of glucose, fructose, L-asparagine, and KCl (early stage)	Absent	Bs
BSU01550	*gerD*	germination response to L-alanine and to the combination of glucose, fructose, L-asparagine, and KCl (early stage)	M4400170	
BSU28410	*gerE*	transcriptional regulator	M4402392	
	gerIA	germination response to the combination of glucose, fructose, L-asparagine, and KCl	M4400595, M4401293	
BSU03700	*gerKA*	germination response to the combination of glucose, fructose, L-asparagine, and KCl	M4403644	
BSU03720	*gerKB*	germination response to the combination of glucose, fructose, L-asparagine, and KCl	Absent (in its place there is a distant *ger* gene, M4403466)	
BSU03710	*gerKC*	germination (cortex hydrolysis) and sporulation (stage II, multiple polar septa)	M4403467	
BSU28380	*gerM*	spore germination protein	M4402396	
BSU10720	*gerPA*	spore germination protein	M4403651	
BSU10710	*gerPB*	spore germination protein	M4403650	
BSU10700	*gerPC*	spore germination protein	M4403649	
BSU10690	*gerPD*	spore germination protein	M4403648	
BSU10680	*gerPE*	spore germination protein	M4403647	
BSU10670	*gerPF*	spore germination protein (*ywdL*)	M4403646	
BSU15090	*gerQ*	spore germination protein; Prespore Specific Transcriptional Activator (*ylbO*)	BM4401120	
BSU37620	*gerR*	transcriptional regulatory protein	BM4403368 (*rsfA*)	
BSU25540	*gpr*	Germination protease precursor	M4402278	
Competence				
		protein A		
BSU31690	*comP*	two-component sensor histidine kinase	Absent	Bs Bl Bc Bt
BSU16930	*cinA*	competence damage-inducible protein A	M4401517	
BSU03430	*nucA*	nuclease	Absent	Bs Bl Bcg
BSU03420	*nin*	inhibition of the DNA degrading activity of NucA	Absent	Bs Bl Bcg
	comC	DNA-binding protein	M4402351	
BSU03500	*comS*	regulation of genetic competence	Absent	Bs
BSU25590	*comEA*	unspecific high-affinity DNA-binding protein	M4402274	
BSU25580	*comEB*	required for DNA binding and uptake	M4402275	
BSU25570	*comEC*	putative integral membrane protein	M4402276	
BSU25600	*comER*	late competence protein	M4402272	
BSU35470	*comFA*	late competence protein	M4403036	
BSU35450	*comFC*	competence protein	M4403037	
BSU24730	*comGA*	probably part of the DNA transport machinery	M4402041	
BSU24720	*comGB*	probably part of the DNA transport machinery	M4402042	
BSU24710	*comGC*	probably part of the DNA transport machinery	M4402043	

B. subtilis	Gene name	Function	*B. coahuilensis* ortholog	Orthologs in other *Bacillus* (when absent in *B. coahuilensis*)
BSU24700	*comGD*	probably part of the DNA transport machinery	Absent	Bs Bl Bcg
BSU24690	*comGE*	probably part of the DNA transport machinery	Absent	Bs Bl
BSU24680	*comGF*	probably part of the DNA transport machinery	Absent	Bs Bl
BSU24670	*comGG*	competence transcription factor	Absent	Bs Bl
BSU10420	*comK*	competence transcription factor (CTF)	M4403645	
BSU11300	*med*	late competence gene	M4400979	
BSU31710	*comQ*	transcriptional regulator	Absent	Bs Bl
BSU31700	*comX*	pherormone	Absent	Bs
BSU11310	*comZ*		M4400980	
BSU00860	*clpC*	class III stress response-related ATPase	M4400120	
BSU34540	*clpP*	ATP-dependent Clp protease proteolytic subunit	M4403116	
BSU28220	*clpX*	ATP-dependent protease ATP-binding subunit	M4402322	
BSU11520	*mecA*	adaptor protein	M4400995	
	spxA	regulatory protein	M4400994	
BSU14990	*ylbF*	regulatory protein	M4403560	

[a] When no ortholog was found in *B. coahuilensis* we searched in the genomes of other *Bacillus* spp. and closely related species to determine how common this absence was. For genes coding for proteins of less that 100 residues, we looked for sinteny and searched at the corresponding location. Bs, *B. subtilis; Bl, B. licheniformis; Bce, B. cereus; Ba, B. anthracis; Bt, B. thuringiensis; Bcg, Bacillus cereus group (encloses Bce, Ba, and Bt); Bcl, B. clausii; Oi, O. iheyensis; Bk, B. kaustopilus; NR, B.* sp. *NRRL11194*.

Table S7. Cell envelope, biogenesis and cell division proteins in *B. coahuilensis* in comparison to *B. subtilis* and other *Bacillus* spp.

	*B. subtilis**	Function	Possible orthologs in *B. coahuilensis*	Orthologs in other *Bacillus*† (when absent in *B. coahuilensis*)
Fatty acid synthesis				
Initiation of fatty acid synthesis	*accA*‡, *B*‡, *C*‡, *D*‡, *acpA*‡, *fabD*‡, *birA*‡		M4402554 (*accA*), M4402080 (*accB*), M4402081 and M4401121 (*accC*), M4402553 (*accD*), M4401388 (*acpA*), M4401386 (*fabD*), M4401885 (*birA*)	
	fabHA, fabHB	Overlapping function	M4400982 (*fabHA*)	
Fatty acid chain elongation	*fabF*‡,*G*‡		M4400983 (*fabF*) M4401387(*fabG*) (seven more 3-oxoacyl-[acyl carrier-protein] reductase)	
	fabI fabL	Overlapping function	M4401919 (*fabI*)	
	ywpB (=*fabZ*), *ycsD*	Overlapping function	M4403340 (*fabZ*)	
Phospholipid synthesis	*gpsA*		M4401849 (*gpsA*)	
	yhdO‡	Similar to 1-acylglycerol-3-phosphate *O*-acyltransferase	Similar to BM4401843 LPAT1 and BM4401172 LPAT1 (chloroplast)	
	cdsA‡	Phosphatidate cytidylyltransferase	M4401481 (*cdsA*)	
	pgsA‡	CDP-diacylglycerol-glycerol-3-phosphate 3-phosphatidyltransferase	M4401516 (*pgsA*)	
	ywjE, ywnE, ywiE	Cardiolipin synthetase (overlapping function)	M4403386 (*ywjE*) M4400996 (*cls*), M4401628 (*cls*)	
	yerQ‡, *dgkA*	Diacylglycerol kinase, overlapping function	Similar to M4400357 (*bmrU*), M4400531 (*bmrU*) M4402174 (*dgkA*)	
	ugtP	Glycosyltransferase	Putative glycosyl transferase BM4401162 (*ypfP*)	
	pssA	Phosphatidylserine synthase	**Absent**	*Bs*, *Bl*, *Bcg*, *Bh*, Gk, NR
	psd	Phosphatidylserine decarboxylase	**Absent**	*Bs*, *Bl*, *Bcg*, *Bh*, Gk, NR
	plsX‡	Fatty acid/phospholipid synthesis protein	M4401385 (*plsX*)	
Peptidoglycan synthesis				
Synthesis of amino sugars	*glmS*‡	Glucosamine-fructose-6-phosphate aminotransferase [isomerizing]	M4400183 (*glmS*)	
	ybbT (glmM)‡	Phosphoglucosamine mutase	M4400182 (*glmM*)	
	gcaD‡	Bifunctional protein	M4400088 (*gcaD*)	
	yvyH (mnaA)‡	UDP-*N*-acetylglucosamine-2-epimerase	M4402980 (*mnnA*)	
	nagA	*N*-Acetylglucosamine-6-phosphate deacetylase	M4402950 and M4403240 (*nagA*)	
	nagB, gamA (ybfT)	Glucosamine-6-phosphate deaminase (overlapping function)	M4403241 (*nagB*)	
	pgi‡	Glucose-6-phosphate isomerase A	M4403532 (*pgiA*)	
	gtaB	UTP-glucose-1-phosphate uridylyltransferase	M4403211/M4403022 (*gtaB*)	
	tagE	UDP-glucose:polyglycerol phosphate glucosyltransferase	**Absent**	Bs, Bcl
	nagP	Ohosphotransferase system (PTS) *N*-acetylglucosamine-specific enzyme IICB component	Similar to M4403536 (*nagE*)	
	gamP	Probable PTS glucosamine-specific enzyme IICBA component	Similar to M4400583 (*ptsG*)	
Diaminopimelate	*dapG, lysC, yclM*	Overlapping function	M4401459 (*dapG*), M4402386 (*lysC*)	
	asd		M4401460 (*asd*)	

	B. subtilis[*]	Function	Possible orthologs in *B. coahuilensis*	Orthologs in other *Bacillus*[†] (when absent in *B. coahuilensis*)
	dapA, B		M4401457 (*dapA*), M4401880 (*dapB*)	
	ykuQ (dapD)[‡]		M4401047 (*dapD*)	
	ykrV, ywfG	Overlapping function	M4400772 Transaminase (*mtnE*), M4402696 Aspartate aminotransferase (*aspC*)	
	ykuR[‡]	Peptidase	Similar to M4401048 (*ytnL*) and M4400898 (*amaA*)	
	dapF[‡]		M4401227 (*dapF*)	
	lysA		M4400477 (*lysA*)	
	hom		M4401975 (*hom*)	
	spoVFA, B		M4401462, M4401461	
Racemases	*racE*[‡], *yrpC*	Overlapping function	M4402395 (*murI*)	
	Alr[‡]	Alanine racemase	M4400388 (*alr*); two more hypothetical: M4400440 and M4401269	
Synthesis of lipid-linked disaccharide pentapeptide	*Ddl*[‡]	D-alanine–D-alanine ligase	M4400378 (*ddlB*)	
	murAA[‡], *AB*	Overlapping function	M4403328 (*murAA*), M4403284 (*murAB*)	
	murB[‡]		M4401321 *murB*, M4402943 *murB2*	
	murC[‡],*D*[‡],*E*[‡]		M4402485 (*murC*), M4401144 (*murD*), M4400855 (*murE*)	
	murF[‡], *mraY*[‡], *murG*[‡]		M4400379 (*murF*), M4401143 (*mraY*), M4401217 (*murG2*)	
Teichoic acid biosynthesis	*tagO*[§]	Teichoic acid linkage unit synthesis	M4403025 (*tagO*)	
B. subtilis 168 poly (glycerol phosphate)	*tagA*	Polyglycerol phosphate assembly and export	**Absent**	Bs, Bl, Bce, Bt, Ban, Gk, NR
	tagB	Polyglycerol phosphate assembly and export	**Absent**	Bs, Bli, Bcl
	tagD	Glycerol-3-phosphate cytidylyltransferase	**Absent**	Bs, Bli, Bcl
	tagF	CDP-glycerol:polyglycerol phosphate glycerophosphotransferase	**Absent**	Bs, Bli, Bcl
	tagG	Teichoic acid translocation (permease)	**Absent**	Bs, Bl, Bcl, Bh, Ban
	tagH	Teichoic acid translocation (ATP-binding protein)	Hit to other transporters	
B. subtilis 23 poly (ribitol phosphate)	*tarA, D*	Teichoic acid linkage unit synthesis: *N*-acetylmannosamine transferase and -glycerol-3-P-cytidyltransferase		
	tarB, F	Glycerolphosphotransferases		
	tarK, L	Ribitoltransferases		
	tarI, J	5-P-cytidyltransferase 5-P-dehydrogenase		
Teichuronic acid biosynthesis	*tuaA*	(Lipid-carrier) sugar transferase	**Absent**	*Bs, Bl, Bt, Bcl, Bh, Gk*
	tuaB	Polymer export	M4403024 (*tuaB*)	
	tuaC	Sugar transferase	**Absent**	Bs, Bl
	tuaG	Sugar transferase	**Absent**	Bs, Bl, Bh, Bcl, Bcg
	tuaH	Sugar transferase	**Absent**	Bs, Bl
	tuaD	UDP-G-dehydrogenase	Similar to M4402984 (*tuaD*)	*Bs, Bl, Bt, Bcl, Bh, Gk, NR*
	tuaE	Repeating unit	**Absent**	*Bs, Bl*

	B. subtilis[*]	Function	Possible orthologs in *B. coahuilensis*	Orthologs in other *Bacillus*[†] (when absent in *B. coahuilensis*)
	tuaF	Membrane bound, unknown function	**Absent**	*Bs, Bl*
Cell shape and division				
Septum formation	*ftsA*[‡]		M4401327 (*ftsA*), M4402492 (*ftsA*)	
	ftsW[‡]		M4400657 (*ftsW*), M4401090 (*ylaO*)	
	ftsZ[‡]		M4401328 (*ftsZ*)	
	ftsL[‡]		Similar to M4402000 (*yqgB)*	
	divIB[‡], *C*[‡], *pbpB*[‡]	Penicillin-binding protein 2B	M4401322 *divIB*, (*divIC*, similar to M4400077), M4401139 (*pbpB*)	
Cell shape	*rodA*[‡], *mreB*[‡], *C*[‡]		M4400360 (*ywcF*), M4402361 (*mreB*), M4402362 (*mreC*)	
Capsule[¶]	*capA*	Poly-gamma-glutamic synthesis	M441683 (*capA*) M441686 (*capA*) (MLA roto? le faltan unos 100 aas se juntan los dos genes)	Bs (*pga*), Bcg, Oi
	capD	Capsular polysaccharide biosynthesis protein *capD*	M4403209 (*capD*)	Bs (*yveM*), Bl, Bce, Bh, Oi, NR
	capI	NAD-dependent epimerase/dehydratase	M44000268 (*capI*)	Bt, Bh, Oi, NR
	icaA	Biofilm PIA synthesis *N*-acetylglucosaminyltransferase *icaA*	M4402203 (*icaA*) 2977 M4403023	Bh, Gk, NR
	swrC		M4401621 M4401624	

---

[*] In bold, genes for which no ortholog is found in *B. coahuilensis.* In parentheses, alternative name given to a gene.

[†] Whenever no ortholog was found in *B. coahuilensis* we searched in the genomes of other *Bacillus* to determine how common this absence was in this genre. Bs*, B. subtilis; Bl, B. licheniformis; Bce, B. cereus; Ba, B. anthracis; Bt, B. thuringiensis; Bcg, Bacillus cereu*s group (encloses *Bce, Ba,* and *Bt*); *Bcl, B. clausii; Oi, O. iheyensis; Gk, Geobacillus kaustophilus* HTA426; *NR, B*. sp. NRRL11194).

[‡] Essential in *B. subtilis* [Kobayashi K, *et al.* (2003) Essential Bacillus subtilis genes. *Proc Natl Acad Sci USA* 100:4678–4683].

[§] Suggested to be involved in the synthesis of other cell wall polyanionic acids [Soldo B, Lazarevick V, Karamata D (2002) *tagO* is involved in the synthesis of all anionic cell-wall polymers in *Bacillus subtilis*. *Microbiology* 20, Vol. 148:2079–2087].

[¶] Genes that may be involved in capsule formation are not all present in *B. subtilis*. There are 137 genes annotated in this functional category of which we only show a selected set.

Table 8. Analysis of the presence in *B. coahuilensis* of the *B. subtilis* genes constituting the phosphate regulon

Induced in *Bs* by limiting phosphate	Gene/operon	Function	*B. coahuilensis* ortholog*	Orthologs in other *Bacillus*† (when absent in *B. coahuilensis*)
	phoP	Alkaline phosphatase synthesis transcriptional regulatory protein	M4402201 (*phoP*) M4402563 (*phoP*)	
	phoR	Alkaline phosphatase synthesis sensor protein	M4402564 (*phoR*)	
	phoB	Alkaline phosphatase III (secreted)	M4403423 (*phoB*)	
	ydhF	Unknown	**Absent**	*Bs*
	pstS	High-affinity phosphate ABC transporter	M4402010 (*sphX*)	
	pstC	Phosphate transport system permease	BM4402011 (*yqgH*)	
	pstA	Phosphate transport system permease	M4402012 (*pstA-1*)	
	pstBA	Phosphate ABC transporter (ATP-binding protein)	M4402013 (*pstB*)	
	pstBB	Phosphate ABC transporter (ATP-binding protein)	**Absent**	*Bs Bl*
	phoD	Phosphodiesterase possibly for teichoic acid turnover	**Absent**	*Bs Bl*
	tatAD	Similar to hypothetical proteins	**Absent**	*Bs Bl Ba Bcl Gk Oi NR*
	resA	Thiol-disulfide oxidoreductase	M4401809 (*resA*)	
	resB	Required for cytochrome *c* synthesis	M4401810 (*resB*)	
	resC	Required for cytochrome *c* synthesis	M4401811 (*resC*)	
	resD	Transcriptional regulatory protein	M4401812 (*resD*)	
	resE tuaABCDEFGH glpQ phoA	Sensor protein teichuronic acid biosynthesis glycerol phosphoryl diester phosphodiesterase (hydrolysis of deacetylated phospholipids; (secreted) Alkaline phosphatase	M4401813 and M4400914 (*resE*) **Absent** Similar to BM4402098 (*yqiK*) **Absent**	See Table S6 A single AP also in *Bh Bcl Gk Oi NR*
	tatCD ykoL	Twin arginine transporter, unknown function	**Absent** (has *tatCy*) **Absent**	*Bs Bl Bcg Bh Bcl Gk Oi NR Bs Bl*
	yttP	Probable HTH-type transcriptional regulator, unknown function	M4402509 (*yttP*)	
	ydbD yurI	Similar to manganese containing catalase Extracellular RNase	M4401506 (*ydbD*) M4400550 (*bsn*)	
	yjdB vpr	Unknown extracellular serine protease	**Absent** M4403315 (*vpr*)	*Bs*
	lytB rapA glcU cotP yfkN	Modifier protein of LytC response regulator Aspartate phosphatase glucose uptake Spore coat protein similar to 2′,3′ cyclic nucleotide 2′ phosphodiesterase	**Absent** **Absent** **Absent** **Absent** **Absent**	See Table S5 *Bs Bl Bcg NR*
Repressed	*tagAB tagDEF*	Polyglycerol teichoic acid	**Absent** **Absent**	See Table S6
Related genes present in *B. coahuilensis*	*resD phoR cpdB phoU*	Two-component response regulator Two-component sensor histidine kinase 2′,3′-cyclicnucleotide 2′-phosphodiesterase precursor negative regulator of the Pi regulon	*M4400235 (resD) M4400236 (phoR) M4400483 (cpdB) M4402014 (phoU)*	*See Table S6 Bcg Bh Bcl NR*

*In bold, genes for which no ortholog is found in *B. coahuilensis.* In parentheses, name given to the *B. coahuilensis* gene.

†When no ortholog was found in *B. coahuilensis* we searched in the genomes of other *Bacillus* spp. and closely related species to determine how common this absence was. For genes coding for proteins of $<$100 residues, we looked for sinteny and searched at the corresponding location. Bs, *B. subtilis; Bl, B. licheniformis; Bce, B. cereus; Ba, B. anthracis; Bt, B. thuringiensis; Bcg, Bacillus cereus* group (encloses *Bce*, *Ba*, and *Bt*); *Bcl, B. clausii; Oi, O. iheyensis; Bk, B. kaustopilus; NR, Bacillus* sp. NRRL11194.