Supplementary Table S1. Comparison of homology search results obtained by BLAST with default parameters and modifications of the search described in this work for 774 completely sequenced genome set.

Starting query sequence (toxin name and GI number)	Number of hits identified by PSI-BLAST with default parameters[a]	Exhaustive PSI-BLAST	Number of hits identified by TBLASTN with default parameters[b]	Exhaustive TBLASTN
HokC 49175991	94	117	n/a	n/a
TxpA 1730974	1	48[c]	n/a	n/a
LdrD 49176368	53	82[d]	n/a	n/a
Fst 256852532	0		n/a	n/a
IbsA 226524729	n/a	n/a	0 (identical self-hit has E-value=0.9)	149
TisB 162416093	n/a	n/a	31	35
ShoB 226524739	n/a	n/a	14	21

[a]PSI-BLAST default parameters: database = Refseq at NCBI; matrix = BLOSUM62; word size = 3; gap cost = existence 11, extension 1; PSI-BLAST inclusion threshold = 0.005; expect threshold = 100; no low complexity filtering; conditional compositional score matrix adjustment; no composition based statistics (1); iterations run until convergence (no more new hits detected above the inclusion threshold). Only hits with E-value less or equal 0.005 were considered as statistically significant.

[b]TBLAST with default parameters: Same as PSI-BLAST but no iterations run.

[c]This number only includes representatives found by exhaustive PSI-BLAST searches using the Refseq database and *Bacillus subtilis* TxpA as a starting query. 118 homologs, including the *Enterococcus faecalis* subfamily, are found when using the NR database.

[d]This number includes representatives of both Fst and Ldr families which we consider to be homologous. With the addition of homologs found by exhaustive TBLASTN the number of identified Ldr/Fst family members is twice as large (196 in total); see Supplementary Table 6.

Supplementary Table S2. Sequence and gene features of known type I toxins

Analysis of upstream and downstream distances

Family	Upstream distance	Downstream distance	Alignment core length
Ibs	1181	708	18
Hok	453	415	50
LdrD	650	702	33
TxpA	492	608	35
TisB	749	282	29
ShoB	877	534	26
Selected parameters	400	250	70

Analysis of 10 C-terminal amino acids for 175 proteins in selected genomes

Amino acid	Counts (absolute frequencies)
C	15
H	18
Q	24
D	35
G	36
T	41
P	52
W	68
F	71
E	79
Y	92
S	95
R	98
N	100
K	193

Supplementary Table S3. Strains and plasmids used in this study

Name	Relevant genotype or description	Source or reference
Strains		
Escherichia coli MG1655	*E. coli* F⁻ λ⁻ *ilvG⁻ rfb-50 rph-1*	lab stock
Eschericihia coli EDL933	Wild type EHEC, O157:H7	D. Friedman
Streptococcus pneumoniae R6	Avirulent laboratory strain	W. Haas
Bacillus subtilis PY79		E. Hobbs, (2)
	Wild type	
Bacillus subtilis sub. subtilis str. 168		K. Ramamurthi, (2)
	Wild type	
Enterococcus faecalis OG1RF	Laboratory strain; Rif^R, Fus^R	J. Lemos
Plasmids		
pAZ3	Cm^R	(3)
pDR111	Ap^R	(4)

Supplementary Table S4. Oligonucleotides used in this study

Name	Sequence (5'-3', restriction sites underlined)	Use
EF427	GCCCTTAACTATGGCACTAGTTAGGGGCTTTGG	Antisense to *fst-B S. pneumoniae* Northern analysis
EF507	CTTGCTGATTGTGCAACATTTCGTACC	sRNA-1 Northern analysis
EF514	CACCACAGCCTGTGCAACATCGTGTTC	sRNA-2 Northern analysis
EF512	TGTCCATCCGCTCTCCTTATTAAGGAGCGC	*z3289/z3290* Northern analysis
EF524	GTTGGTACGAAACGTTGCTCTCCG	*z3289* Primer extension
EF506	CCTGCTGGTGAGGCAGGTTCTTTTATTT	sRNA-1/sRNA-2 Primer extension
EF472	GATTATCATCGGCGCTTCTTGGCTGTAC	Antisense to *yhzE-2 B. subtilis* Northern analysis
EF470	CGGTCTACCCGAGAGCGTATGTCTAGTACGCTC	Antisense to *txpA E. faecalis* Northern analysis
EF486	TCCCCCTCCCCCTGGTGTTCTTAGTAAGCC	Antisense to *yonT B. subtilis* Northern analysis
EF398	TATACT<u>GAATTC</u>GTCATCGGTTACCGATTACGTTCC	pAZ3-*fst-B* PCR
EF399	GAACAC<u>AAGCTT</u>CCCTAACTAGTGCCATAGTTAAGGGC	pAZ3-*fst-B* PCR
EF390	CCGGAG<u>GAATTC</u>ATCGTGGTCCAACATACGTAAACC	pAZ3-*z3290* PCR
EF391	CGATGA <u>AAGCTT</u>AAATAAAAATGAGTTATATGATG	pAZ3-*z3290* PCR
EF441	CTGTAA<u>GCTAGC</u>TTTAAATAAAGGGCATTCGTCTAG	pDR111-*yhzE-2* PCR
EF494	GCATAG<u>GCATGC</u>ATTAAGCAGCTATGCTTTTTTATG	pDR111-*yhzE-2* PCR
EF437	GTATAATAG<u>GCTAGC</u>GAAGCTACCTTAGCTTATCCTC	pDR111-*txpA B. subtilis* PCR
EF438	CATTAC <u>GCATGC</u> ACCACTTGTTAATGTGTGTTAC	pDR111-*txpA B. subtilis* PCR
EF429	GTTATA<u>GAATTC</u>AAGTAGGAAGTTTATGACGGTGGCAATTTC	pAZ3-*txpA E.faecalis* PCR
EF430	GTATAA<u>AAGCTT</u>CGAAGAGAGCGTACTAGACATACGCTC	pAZ3-*txpA E.faecalis* PCR
EF473	CCATCC<u>GAATTC</u>ACACAAATAACTTACCACAAATAT	pAZ3-*yonT* PCR
EF474	CATTAA<u>AAGCTT</u>GCCAATGAATGTGATATGGGTAAACC	pAZ3-*yonT* PCR

Supplementary Table S5. Complete list of type I toxins in completely sequenced genomes identified by BLAST approach. (Separate .xls file)

Supplementary Table S6. Distribution of type I toxin families in 774 completely sequenced microbial genomes. (Separate .xls file)

Supplementary Table S7. Results of tandem ORF search in γ-Proteobacteria and Firmicutes. (Separate .xls file)

Supplementary Table S8. Results of search for characteristic sequence features of type I toxins in γ-Proteobacteria and Firmicutes. (Separate .xls file)

Supplementary Table S9. The data for predicted antitoxin RNA regions used for comparison with random sequences. (Separate .xls file)

REFERENCES

1. Schäffer, A.A., Aravind, L., Madden, T.L., Shavirin, S., Spouge, J.L., Wolf, Y.I., Koonin, E.V. and Altschul, S.F. (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.*, **29**, 2994-3005.
2. Zeigler, D.R., Prágai, Z., Rodriguez, S., Chevreux, B., Muffler, A., Albert, T., Bai, R., Wyss, M. and Perkins, J.B. (2008) The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J. Bacteriol.*, **190**, 6983-6995.
3. Kawano, M., Aravind, L. and Storz, G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Mol. Microbiol.*, **64**, 738-754.
4. van Ooi, j.C. and Losick, R. (2003) Subcellular localization of a small sporulation protein in *Bacillus subtilis. J. Bacteriol.*, **185**, 1391-1398.