

Supplementary material

Towards Exploring Toxin-Antitoxin Systems in *Geobacillus*: A Screen for Type II Toxin-Antitoxin System Families in A Thermophilic Genus

Rawana N. Alkhalili ^{1,*}, Joel Wallenius ² and Björn Canbäck ²

¹ Biotechnology, Department of Chemistry, Lund University, SE-221 00 Lund, Sweden;
rawana.alkhalili@biotek.lu.se

² Department of Biology, Lund University, SE-221 00 Lund, Sweden; J.W. joelwallenius@gmail.com; B.C.
Bjorn.Canback@biol.lu.se

* Correspondence: rawana.alkhalili@biotek.lu.se; Tel.: +46 46 222 94 19

13 **Table S1.** TAs identified in the current study. The used nucleotide and amino acid sequences belong to the
14 RefSeq genome record of the respective strain, except for GK1498 and GK1499 where the TA pair annotated
15 in the original genome record was used as it showed better alignment.

TA family	Toxin/Antitoxin (T/AT)	Strain	Genome accession #	Locus tag	Protein ID	Protein description
GacTA (GNAT-HTH)	wHTH_AT	<i>Gd</i> [†]	NC_009328	GTNG_1350	WP_008879365	IcI family transcriptional regulator
	GNAT_T	<i>Gd</i> [†]	NC_009328	GTNG_1349	WP_008879366	GNAT family N-acetyltransferase
	GNAT_T ¹	<i>Gd</i> [†]	NC_009328	GTNG_1577	WP_008879126	GNAT family N-acetyltransferase
	GNAT_T ¹	<i>Gd</i> [†]	NC_009328	GTNG_1578	WP_008879125	GNAT family N-acetyltransferase
	HTH_AT ¹	<i>Gd</i> [†]	NC_009328	GTNG_1575	WP_008879127	HTH-transcriptional regulator
	wHTH_AT	<i>Gk</i> [‡]	BA000043	GK1499	BAD75784	Transcriptional regulator (icI family)
	GNAT_T	<i>Gk</i> [‡]	BA000043	GK1498	WP_011230994 ²	GNAT family N-acetyltransferase
	wHTH_AT	<i>Gt</i> [§]	NC_016593	*_17290	WP_013145506	IcI family transcriptional regulator
	GNAT_T	<i>Gt</i> [§]	NC_016593	*_17280	WP_011230994	GNAT family N-acetyltransferase
	wHTH_AT	ZG [¶]	LDPD00000000	Contig_16_18	Table S2 ³	Table S2 ³
	GNAT_T	ZG [¶]	LDPD00000000	Contig_16_17	Table S2 ³	Table S2 ³
MazEF (I)	MazE (I)_AT	<i>Gk</i> [‡]	NC_006510	GK1647	WP_011231142	AbrB/MazE/ SpoVT family DNA-binding domain-containing protein
	MazF (I)_T	<i>Gk</i> [‡]	NC_006510	GK1648	WP_011231143	mRNA-degrading endonuclease
	MazE (I)_AT	<i>Gt</i> [§]	NC_016593	*_19080	WP_011231142	AbrB/MazE/ SpoVT family DNA-binding domain-containing protein
	MazF (I)_T	<i>Gt</i> [§]	NC_016593	*_19090	WP_011231143	mRNA -degrading endonuclease
	MazE (I)_AT	ZG [¶]	LDPD00000000	Contig_16_161	Table S2 ³	Table S2 ³
	MazF (I)_T	ZG [¶]	LDPD00000000	Contig_16_162	Table S2 ³	Table S2 ³
	MazE (I)_AT	ZG [¶]	LDPD00000000	Contig_16_162	Table S2 ³	Table S2 ³
MazEF (II)	RHH (II)_AT	<i>Gd</i> [†]	NC_009328	GTNG_0206	WP_008881474	Hypothetical protein
	MazF (II)_T	<i>Gd</i> [†]	NC_009328	GTNG_0207	WP_003253417	Type II toxin-antitoxin system endoribonuclease ndoa
	RHH (II)_AT	<i>Gk</i> [‡]	NC_006510	GK0232	WP_011229742	Hypothetical protein
	MazF (II)_T	<i>Gk</i> [‡]	NC_006510	GK0233	WP_003253417	Type II toxin-antitoxin system endoribonuclease ndoa
	RHH (II)_AT	<i>Gt</i> [§]	NC_016593	*_2490	WP_011229742	Hypothetical protein
	MazF (II)_T	<i>Gt</i> [§]	NC_016593	*_2500	WP_003253417	Type II toxin-antitoxin system endoribonuclease ndoa
	RHH (II)_AT	ZG [¶]	LDPD00000000	Contig_4_60	Table S2 ³	Table S2 ³
	RHH (II)_AT	ZG [¶]	LDPD00000000	Contig_4_60	Table S2 ³	Table S2 ³

	MazF (II)_T	ZG ¶	LDPD00000000	Contig 4_61	Table S2 ³	Table S2 ³
MNT-HEPN (I)	MNT (I) A_AT	Gt §	NC_016593	*_10710	WP_015374218	Putative DNA polymerase beta domain protein region
	HEPN (I)_T	Gt §	NC_016593	*_10720	WP_014195232	DUF86 domain-containing protein
	MNT (I) B_AT ¹	Gt §	NC_016593	Unannotated	Table S2 ³	Table S2 ³
	MNT (I) A_AT	ZG ¶	LDPD00000000	Contig 12_19	Table S2 ³	Table S2 ³
	MNT (I) B_AT ¹	ZG ¶	LDPD00000000	Contig 12_18	Table S2 ³	Table S2 ³
	HEPN (I)_T	ZG ¶	LDPD00000000	Contig 12_20	Table S2 ³	Table S2 ³
	KNTase_AT	Gt §	NC_016593	*_11510	WP_014195291	Nucleotidyltransferase domain-containing protein
MNT-HEPN (II)	HEPN (II)_T	Gt §	NC_016593	*_11500	WP_014195290	DUF86 domain-containing protein
	KNTase_AT	ZG ¶	LDPD00000000	Contig 12_84	Table S2 ³	Table S2 ³
	HEPN (II)_T	ZG ¶	LDPD00000000	Contig 12_83	Table S2 ³	Table S2 ³
ParDE	AbrB (I)_AT	Gk †	NC_006510	GK2355	WP_015375348	Hypothetical protein
	ParE_T	Gk †	NC_006510	GK2354 ⁴	WP_020278248	Hypothetical protein
	AbrB (I)_AT	Gt §	NC_016593	*_26570	WP_014196297	Hypothetical protein
	ParE_T	Gt §	NC_016593	*_26560	WP_014196296	Hypothetical protein
	AbrB (I)_AT	ZG ¶	LDPD00000000	Contig 23_243	Table S2 ³	Table S2 ³
	ParE	ZG ¶	LDPD00000000	Contig 23_242	Table S2 ³	Table S2 ³
PhD-Doc	AbrB (II)_AT	Gk †	NC_006510	GK1845 ⁴	WP_122381439	AbrB/MazE/SpoVT family DNA-binding domain-containing protein
	Doc_T	Gk †	NC_006510	GK1846	WP_011231336	Type II toxin-antitoxin system death-on-curing family toxin
	AbrB (II)_AT	Gt §	NC_016593	*_21520	WP_014195946	AbrB/MazE/ SpoVT family DNA-binding domain-containing protein
	Doc_T	Gt §	NC_016593	*_21530	WP_014195947	Type II toxin-antitoxin system death-on-curing family toxin
	AbrB (II)_AT	ZG ¶	LDPD00000000	Contig 18_126	Table S2 ³	Table S2 ³
	Doc_T	ZG ¶	LDPD00000000	Contig 18_127	Table S2 ³	Table S2 ³
RelBE	XRE	Gk †	NC_006510	GK3105	WP_011232575	Hypothetical protein
	RelE	Gk †	NC_006510	GK3104	WP_011232574	Type II toxin-antitoxin system RelE/ParE family toxin
	XRE	Gt §	NC_016593	*_34820	WP_014196753	Hypothetical protein
	RelE	Gt §	NC_016593	*_34810	WP_014196752	Type II toxin-antitoxin system RelE/ParE family toxin
	XRE	ZG ¶	LDPD00000000	Contig 25_196	Table S2 ³	Table S2 ³
	RelE	ZG ¶	LDPD00000000	Contig 25_195	Table S2 ³	Table S2 ³
VapBC	UPF0175	Gk †	NC_006510	GK1950	WP_013523968	UPF0175 family protein
	COG2405	Gk †	NC_006510	GK1949 ⁴	WP_021322439	DUF3368 domain-containing protein
	UPF0175	Gt §	NC_016593	*_22490	WP_013523968	UPF0175 family protein

	COG2405	<i>Gt</i> §	NC_016593	*_22480	WP_021322439	DUF3368 domain-containing protein
	HTH	<i>Gk</i> †	NC_006510	GK3185	WP_011232655	Hypothetical protein
	COG2856	<i>Gk</i> †	NC_006510	GK3184	WP_011232654	ImmA/IrrE family metallo-endopeptidase
	HTH	<i>Gt</i> §	NC_016593	*_35630	WP_014196828	Hypothetical protein
XRE_COG2856	COG2856	<i>Gt</i> §	NC_016593	*_35620	WP_014196827	ImmA/IrrE family metallo-endopeptidase

16 † *G. thermodenitrificans* NG80-2.

17 † *G. kaustophilus* HTA426.

18 § *G. thermoleovorans* CCB_US3_UF5.

19 ¶ *Geobacillus* sp. ZGt-1.

20 * stands for “GTCCBUS3UF5” which is part of the locus tags in *G. thermoleovorans* CCB_US3_UF5.

21 ¹ Solo toxin/antitoxin

22 ² The amino acid sequence of the protein product of GK1498 annotated in the original genome record of the strain is 100% identical to WP_011230994 over its entire length.

23 ³ The nucleotide and amino acid sequences are presented in Table S2.

24 ⁴ In the RefSeq genome record of *Gk* strain, only the RefSeq locus tags, GK_RS12120, GK_RS09630, and GK_RS10125 instead of GK2354, GK1845, and GK1949, respectively are shown.

27

28 **Table S2.** Nucleotide and amino acid sequences of the putative TAs of *Geobacillus* sp. ZGt-1 based on
29 the genome sequence annotation of the strain in [26], and of the unannotated putative MNT (II)
30 antitoxin of strain Gts. Genes encoded on the reverse DNA strand have their gene start position value
31 larger than that of the end position. The top hit of blastp results of the TA amino acid sequences is
32 also presented.

Locus tag	nt sequence	aa sequence	Start position	End position	Blastp ¹
Contig 16_18	ATGAATAAACGGTCTTAAAAACAAAAGAAC TGCTTGATTGTTTCTCGATTGCGAACGGTTG ACGCTGCCGGAATGGTCGAGCGGCTTCGGA TGCCCAAAACGTCGGTGTACCGGATGGCGCA GTCGCTTGTTGTCTTGGTTTTTGAAAAAC GAGGCGATTACTATGAGCTCGGTTTAGCCTTC TTGACGTTTGGCGCGCTCGTCGCCGAGCGGC TCGATATTCGCCGGGCGGCGCTGCCGGTGAT GAAGCGGCTGAAAGAGGAGACGAACGAAGC GGTGAATCTTGTCATTGCGACGGTGATGAG GCGCTGTATATTGAAAAAGTCGAGACGTCCG AGCCGGTGCGCGTCTATACGAAAGTCGGGCG GCGCGCCCCTCTGTATGCCGGGCGTGCCCT CGTGTCTTGCTCGCGTTTATGGACAAGCGCG ATCGGGAACGTTATTTAGAACAAAGTCGAGCT TGTCAAAATCGCCAACATACGGTGACCGAC AAAGAAGCGCTGCGCCGGTTGTTNGAAGAA GACCGGGACCGCGGCTATACGGTCAGCTATT CTGAGCTGGAAACTATTCGGCCGCCGTTGC GGTGCCGATTTTCAACCATGAAGGCGCGGCG GTTGCCGGGCTGAGCGTTGCCGGGCGCGAA CAGCGCTTTTCGCCAGATGATGTGCGCGCA TCGTTCCGCGATTGAAGCAGGCGGCGATGGA CATTTTCGCGCGAACTCGGCTTTCGGGGGAAG GGATGA	MNKTVLKTKELDLFLDCERLTL PEMVERLRMPKTSVYRMAQSL VVLGFLQKRGDYELGLAFLTFG ALVAERLDIRRAALPVMKRLKEE TNEAVNLVIRGDEALYIEKVET SEPVRVYTKVGRRAPLYAGACP RVLLAFMDKADRERYLEQVELV KIAKHTVTDKEALRRLXEEDRDR GYTVSYSELENYSAAVAVPIFNH EGAAGVAGLSVAGPEQRFSPDD VARIVPRLKQAAMDISRELGFR GKG	17392	18141	WP_047757836; IclR family transcriptional regulator [<i>Geobacillus</i> sp. ZGt-1]

Contig 16_17	ATGCCGCAATTTGTTGGCTCGAGACGGAGG AGGAAGTAAGAAGCGCTTTCCGGTCATGCG GGAGCTGCGCACTCATTGGATGAAGAAACG TATGTCGCGCTTGCGTGGAAGCGCAAGAAA AAGAAGGGTATAAGCTTGGCGCGTTATATGA TCAGGACAAAATGGTTGCCGTTGTCGGATT ATGCCGATGATCAGCTCTATAACGCGCGTT TATTTGGGTTTGGCATTGGTCACGACATCAG CCGAACGTTGAAAGGGTATGGAAGCGTT GCTGTCGCATGTGCACGAATGGGCGAAGGA GCAAGGCTACGGGATTGTCTCGCTGCATCC GGCCTGCAGCGGGTTGACGCCCATCGTTCT ATGAGGAAAAATGGAATATCAAAAGTGA GTTATGTGTTTTGAAACGCTTATCGATTGA	MPQFVWLETEEEVRSAPVMR ELRTHLDEETYVALVREAQEKEG YKLAALYDQDKMVAVVGFMP MITLYNGRFIWWCDLVTTAERS KGYGKALLSHVHEWAKEQGYGI VSLSSGLQRVDAHRFYEEKMEY QKVSFVFLKRLS	16826	17260	WP_021322570; GNAT family N- acetyltransferase [<i>Geobacillus</i>]
Contig 16_161	ATGGAATTAACAAACAGAAAGGAGTGTCAA TCATGACAATAACAGTTCAAAAATGGGGAAA CAGCCTTGGGTTGCGATCCCAAGCGTGATT GCTGAACGTTTAGCGCTTCATCAAGGATCAG AAGTGGAGATGATCGTTGAGAACCAAGCGAT CAAGTTGATCCGAAAAAGAAAAAGCCGACA TTGGAGGAACTTTGGCTAAAATCACGCCGG AAAATCGCCATGCTGAAATTGATTTTGAAC AGAAGGGAATGAATTGTTCTGA	MELNKQKGVSIMTITVQKWGN SLAVRIPSVIAERLALHQGSEVE MIVENQAIKLIPKKKPTLEELLA KITPENRHAIEDFGTEGNELF	156459	156728	WP_011231142; AbrB/MazE/SpoVT family DNA-binding domain- containing protein [Bacillaceae]
Contig 16_162	ATGCAAGCGCCGATCGTGGGGATCTTGTTT ACGTCAATTTCAATCCACAAGCAGGGCATGA GCAGGCAGGAAAAAGCCCGCATCGTTCTA TCACCAAGCGGTTCAATCAATTGACAGGGT TTGCGGTGCTTGGCCGATCACCCGGCAACA AAAAGGATATCCATTTGAAGTGGAATTGCCA TCAGGCTTGGCCGTTGAAGGCGTGATTTAA CCGATCAGGTCAAAAGTTTAGATTGGCGCGC CCGACAGCTTCAAATAGTAGGACGGGCGCCC AATGAAGTTGTTTCGGATTGCTTGATCTGAT TCATACTTTTCTTCGTGA	MQAPDRGLVYVFNFPQAGH EQAGKRPGIVLSPKRFNQLTGF AVLCPITRQQKGYPFVELPSGL AVEGVILTDQVKSLDWRARQL QIVGRAPNEVSDCLDLIHTFLS	156728	157057	WP_011231143; mRNA-degrading endonuclease [Bacillaceae]
Contig 4_60	GTGTCGGAATCTGGCGCAACAGCGAAATCG TCGTTCTGTTGCCGAGTCGCTGCTGACAGAA CTGGACGTGCTCGTAAAGCAGGAAAACGGC AACCGBAATGAACATTTATCAAGCGACGA AAATGTACATTGCGAGCGGAAGAAACGGCA AATTCGCGAGGCGATGAGACGAGGCTACAT GGAAATGGCGAAAATCAATTTATCTATCGCTT CTGAAGCGTTTCATGCTGAATACGAGGCCGA CCACACCGTTGAACGCTTAGTTAGCGGGGGG TAA	MSESGATAEIVRLPQSLTELD VLVKQENGNRNELIYQATKMYI RERKKRQIREAMRRGYMEMAK INLSIASEAFHAEYEADHTVERLV SGG	65656	65937	WP_011229742; hypothetical protein [Bacillaceae]
Contig 4_61	TTGATTGTCAAACGTGGCGACGTGATTTTGC GGACCTTTCCCGGTTGTTGGCTCGGAGCAG GGCGGCGTGCGCCCCGTGTTGGTGATCCAAA ACGATATCGGCAATCGTTTTAGCCCACGGT GATTGTCGCGCGATTACGGCGCAAATCCAA AAAGCGAAGCTGCCGACGCATGTCGAGATTG ACGCGAAACGCTACGGGTTTGAACGCGATT GGTCATTTTGCTTGAGCAAATTCGCACGATCG ACAAGCAACGGCTGACCGATAAAATCACTCA TTTGGACGATGAAATGATGGATAAAGTCGAT	MIVKRGDVYFADLSPVVGSEQG GVRPVLVIQNDIGNRFSPTVIVA AITAQIQKAKLPTHVEIDAKRYG FERDSVILLEQIRTIDKQLRDKIT HLDDMMDKVDEALQISLGLID F	65942	66292	WP_003253417; type II toxin- antitoxin system endoribonuclease NdoA [Bacillaceae]

	GAGGCGCTGCAAATTAGCTTAGGGCTGATCG ACTTTTGA				
Contig 12_19	ATGGATAATGCGGTGACGTTTGATCATTATAT GGAGTTGAAACTGAGTTTGGGAAGATCATT CAAAAACCAAGTGGATTTAGTGATTGTAGATG ATATCAAACCAGGATTAAGCCAACATTTTA AGGAGTGTTAAGTATGCAGAGGGATCCTAG	MDNAVTFDHYMELKLSLEDHF QKPVDLVVDDIKPGLKPTILRSV KYAEGS	17787	17942	WP_015374218; putative DNA polymerase beta domain protein region [<i>Geobacillus</i>]
Contig 12_18	ATGAAAATAACAGGCGCGGAAAAAGGGGGG GGTTATTTGCTATCACAAAGACATCCTTAA TGAAATATCGAAATGTTTAGATTGTGGAAG AAAAAGTATGGTGTGAAGCGAATCGGGTTAT TTGGCTCGTACAGTCGGGGGGGAACAAAGG GAATCAAGTGA	MKITGAEKGGGYLLSQDILNEI SKCLDLWKKYGVKRIGLFGSYS RGGTKGIK	17600	17764	WP_013146011; nucleotidyltransferase [<i>Geobacillus</i>] ²
Contig 12_20	ATGCAGAGGGATCCTAGCGTCTTTTGAAG ATATTTGGCTGCTGCAGAGAAATTGAGAA GTACACCCAAGGGCTTCTATGATGACTTTT TAGATAATGATTTAGTATCAGACGCAGTCATT AAAAACATATTAGTGATTGGAGAAGCAGCCA AAAACATTCCAGACGAAATCAGGCAAGCATC CCCATATATCGAATGGAGAAAGATGGCTGGC ATGAGGGATATGCTGATCCACAGTTATTCTC CATTAATTATCGAATTGTGTAG	MQRDPSVFLQDILAAAEKIEKYT QGLSYDDFLDNDLVSDAVIKNIL VIGEAANKNIPDEIRQASPIEWR KMAGMRDMLIHSYFSINYRIV	17926	18198	WP_014195232; DUF86 domain- containing protein [Bacillaceae]
Contig 12_84	TTGCCAAACGAAATGGAGACGATCATCATT AGACGCTCCGCCCGCTCTTACCCGTTTCGTC ATCTACCTGTTCCGTTACGCCCGCGTGGGAC GCTGCGCCAGACAGCGATGTCGACATCGCC TTTGTACAGCGACGGCGAACC GCATGATCCGT ATGAGCTGTTTCGCTCGCCGGGGAGTTGGC TGACAAGTTAGGGCGAGATGTCGATCTTGTC GATTTGCGCCAAGCCAGCACCGTGTTC AAG CGCAAGTCGTCTCGACGGGAAAAGCCATTGA TTGCCGCGACGAGCGGAAACGGGCTGAGTTC GAAATGAAAACATTGAAATGTATGTGAAAC TAAACGAAGAGAGAGACCCGGTGTTAAACA AATTACGGAAGCGGGAGCATATATGAAA GTGA	MPNEMETIIQTLR PALHPFVIYL FGSAARGTLRPDSDVDIAFVSD GEPHDPYELFRLAGELADKLGR DVDLVDLRQASTVFQAQVVST GKAIDCRDERKRAEFEMKTLKM YVKLNEERAPVLKQITESGSIYEK	84321	83914	WP_014195291; nucleotidyltransferase domain- containing protein [<i>Geobacillus</i>]
Contig 12_83	ATGAAAAGTGATGTCATTTAAACAAGATCA GCGTGATCGAACGCTGCCTGAAACGAATTCG TGAGAATATAACGGCGATCCAAAAATTTA CAAAATTACACAAAACAAGATTCGATCGTTCT GAATTTACAGCGGGCGTGCGAGGCATGCATC GATTTGGCCATGCATATTGTGGCCGAGCAGA AATTCGATTGCCGAGCATAGCCGCGATGC ATTCGCTCTCCTCGAAGAACATGGGGTCATCT CCCCTTCCATAAGCAAAAAGATGAAGGCGAT GGTCGGATTCCGCAACATCGCCGTTACGAC TATCAACAACCTGAACCTTGGCATCTTGCAAGC CATCGTCGAACACCATCTTGATGATTTCAAAC AATTTACGAAAGCCATCCTCGATTATGCTAAG AAAAACAGCTAG	MKSDVILNKISVIERCLKRIREEY NGDPKNLQNYTKQDSIVLNLQR ACEACIDLAMHIVAEQKFGLPQ HSRDAFALLEEHGVISPSISKKM KAMVGFRNIAVHDYQQLNLGIL QAIVEHHLDLFKQFTKAILDYAK KNS	83924	83505	WP_014195290; DUF86 domain- containing protein [Bacillaceae]
Contig 23_243	ATGGTACCATCAGATTCGATGACAAAGGAGA AGGTACCGATGGACAATATCCCTAACCGTAA GAAAAACCTTGTAACGAATTGCCGTATCG AGCAAAAAACAAATTACGATCCCAAAGGATT	MVPSDSMTKEKVPMDNIPNRK KNLVKRIAVSSKKQITIPKDFYEQ LGIGNEVLIELADNKLIHPIHED HFDFSDLILKDIEEGYTGELYK	238003	237596	WP_015375348; hypothetical protein [<i>Geobacillus</i>]

	TTTATGAGCAATTAGGAATTGGAAACGAGGT GTTAATTGAGTTAGCGGATAATAAGTTAATTA TTCATCCTATTCATGAGGATCATTTTGACTTT CCGATCTGATTTTAAAAGACTTGATCGAAGA AGGATATACAGGCGAAGAACTATACAAAGAA TTTGATACCGTAAATCGCAAATTGCCCTGC ATTCACGCGATGATTTCTGAGGAAAGACCG AAGGCAAAAACGTATACCGCGGATACGTTAG AGGAGTTATTCGCGAAGATGATGAACAATA A	EFVYRKSQIAPAFNAMISEERPK AKTYTADTLEELFGEDDEQ			
Contig 23_242	ATGATGAACAATAATCGGCTGCAATTGCTCCC AAAAGCAGAAAAAACCATCAAAAAGCTGACA AAGAAAGATCCGGTGCTGAAACAGCGTTTGA AAGAAGCCTTGCGGGAGATTCTTCCATCCA ACGGAAGCAGGAGAAGCGAAAACAGGGGAT TTGGCAGGGATTTACGGTTACGACATTTATCA TCAAGGAGTCAATTACGAAATCGCCTATTTTA TTGATCAGGATGAAAGCGGGAATATGGTGGT CGTTGTCTTGGCGGAACCCGCGAAAATTTT ATGATGAGCTGAAGCGTTATATGAAAGCGAA CAAAGCGAGACTTCTAAACCATAG	MMNNNRLQLLPAEKTIKKLTK KDPVLKQRLKEALREILSHPTA GEAKTGDLAGIYGDIYHQGVN YEIAYFIDQDESGNMVVVVLG TRENFYDELKRYMKANKARLPK P	237609	237271	WP_020278248; hypothetical protein [<i>Geobacillus</i>]
Contig 18_126	ATGGACGCGAAATGTTGGGAAGCAAGGAT GACAAAACGTATACACGTCGGATTAGCCAAG TCGGGAATAGTTTGCCGTTAGCATCCCGAA AGATTTAGCCACGATGCTAAACCTAAATAAA GGCGATGAAATCGAAATATATTACGACAAGG AACGAGGGGAAATCGTGATGAAACGCGCAA ACCGAATTCCAAAAGAAGTCCGTCCTGAAGT CGTGATGGCGATGAACCGCGCATCTCCAAA TATGACGAAGCGCTGCGCAACTTGAAATATA GATAA	MDAEMLGSKDDKTYRRISQV GNSLSVSIPKDLATMLNLNKGD EIEIYDKERGEIVMKRANRIPKE VRPEVVMAMNRAISKYDEALR NLKYR	124862	125143	WP_013523845; AbrB/MazE/SpoVT family DNA-binding domain- containing protein [<i>Geobacillus</i>]
Contig 18_127	ATGGTTTATTATTGACAGCGGAAGAAATCAT ATTTATCCATTACACGGTCATGGAAATGTACG ACGACGCAGAACAAAGCGGGATCCAATCCC CGATAAATTCGCGTGATGTTGGAAAGACCG AAAACGAAATTGTTGCGGGAGGAACAGTTCC CTTCGATTATTGAAAAGCCTTGCTGCTATTAC CATTCCATCGCGACAGGTCATATTTCCACAA CGGCAATAAGCGGACGGCTTTAACCGTATTC GTCACGTTCTTGATTTGAATGGATATGAGTT CACTATGACCAATAAAGAAGCGGAAGATTTT ACGGTGATCCTTGTCGAGGATGCCAAGTTCC GGGGAAATGATTGCATCCAGCATCTTGCCCA CGAATTAGAGAGTTACATCCGTCCCATCCAAA AACGAAAAGAATGA	MVYYLTAEIIFIHYTMEMYD DAEQAGIQFPDKFAWMLERP TKLFGEEQFPSIIEKACCYYHSIAT GHIFHNGNKRTALTVFVTFDL NGYEFTMTNKEAEDFTVYLVED AKFRGNDICIQLHAHELESYIRPI QKRKE	125164	125586	WP_014195947; type II toxin- antitoxin system death-on-curing family toxin [Bacillaceae]
Contig 25_196	cntgatattTTGATAGGAGGacgcntTTGGCTGT GCGCAAACAATTGCTTTACGAACTGATTGAA CGGTTGGATGAAACGGATCATCAACAGCGT ACGATTTTTTGATGTACTTGCTTGATCGGTCC AGAAAGGAACGAATGGTATGGGAGCGGATT GACGAAACAGATGAAGAAGAAGCGTTGACG GAAGAAGAGCGCCAGCAGCTGCAAAGCGAT GAAGGATATATCACCGGGGAGAAGCCAAG	MAVRKQLLYELIERLDETDHQT AYDFLMYLLDRSRKERMVWERI DETDEEEALTEERQQLQSDG YITGGEAKREFGLQVDLP	191064	190810	WP_082218538; hypothetical protein [<i>Geobacillus</i>] ³

	CGTGAATTCGGGCTACAAGTTGATTACCGTA				
	G				
Contig 25_195	GTGAATTCGGGCTACAAGTTGATTACCGTA GGGCCGCAGTCAAATTCATCGCTAGGCAAGA AAAAGAGGTTCAAGAACGTTGGCCTCTGGG TTGCAAGGCTGCTTGCGATCCACCGCAGG GGGATATAAAAAAGTTGAAGGGGCAGGATG GATTATATCGGCTCGGGTCGGAACATATCG TGTTTTGTTTCGCATCGATCATGATGAACGAA TCATCTATATTGAGGCGATCGGCAACCGCGG GGATGTGTATTGA	MNSGYKLIYRRAAVKFIARQEKE VQERLASGLQGLLAIPPGDIKK LKGQDGLYRLRVGTYRVLFRIDH DERIIYIEAIGNRGDVY	190841	190581	WP_047758268; type II toxin- antitoxin system RelE/ParE family toxin [<i>Geobacillus</i>]
Unannotated putative MNT (II); strain Gts ⁴	ATGAAAATAACAGGCGCGGAAAAAGGGGGG GGTTATTTGCTATCACAACAAGACATCCTTAA TGAAATATCGAAATGTTTAGATTTGTGGAAG AAAAAGTATGGTGTGAAGCGAATCGGGTTAT TTGGCTCGTACAGTCGGGGGGGAACAAAGG GAATCAAGTGA	MKITGAEKGGGYLLSQDILNEI SKCLDLWKKYGVKRIGLFGSYS RGGTKGIK	997713	997877	WP_013146011; nucleotidyltransferase [<i>Geobacillus</i>] ²

33

34

35

36

37

38

39

40

41

¹ Blastp results for all the sequences, except for contig 12_18, and the unannotated putative MNT (II), showed 100% matching identity to the presented protein ID over the entire length and with e-values very close to zero.
² The aa sequences of contig 12_18 and the unannotated putative MNT (II) are identical and showed 90% matching identity to WP_013146011, with 94% protein coverage. E-value was very close to zero.
³ The aa sequence of contig 25_196 showed 99% identity to WP_082218538 over the entire length of the protein. E-value was very close to zero.
⁴ The nucleotide and amino acid sequences were predicted by the Operon-Mapper tool. Gene position was retrieved by the tblastn tool.

42 **Table S3.** TA operon-sharing genes and their protein products as predicted in this study.

43

44 [†] *G. thermodenitrificans* NG80-2.

TA family	Strain	Toxin/Antitoxin (T/AT)	Locus tag	Total no. of genes/operon ¹	Locus tag of sharing genes	ID of sharing proteins ²	Identification of sharing proteins ²
GacTA	<i>Gd</i> [†]	wHTH_AT	GTNG_1350	5	GTNG_1350	WP_008879364	5-oxoprolinase subunit PxpB
					GTNG_1352	WP_035499192	Biotin-dependent carboxyltransferase
					GTNG_1353	WP_008879362	LamB/YcsF family protein
					GTNG_1354	WP_008879361	Divalent metal cation transporter
	<i>Gd</i> [†]	GNAT_T	GTNG_1349	1	- ³	- ³	- ³
	<i>Gd</i> [†]	HTH_AT_solo	GTNG_1575	2	GTNG_1574	WP_008879128	Hypothetical protein
	<i>Gk</i> [†]	wHTH_AT	GK1499	8	GK1500	WP_013145505	5-oxoprolinase subunit PxpB
					GK1501	WP_014195657	biotin-dependent carboxyltransferase
					GK1502	WP_011230998	LamB/YcsF family protein
					GK1503	WP_031206563 ⁴	DUF523 domain-containing protein
					GK1504	WP_011231000	Hypothetical protein
					GK1505	WP_011231001	AMP-binding protein
					GK1506	WP_011231002	Phosphotriesterase-related protein
	<i>Gk</i> [†]	GNAT_T	GK1498	1	- ³	- ³	- ³
	<i>Gt</i> [§]	wHTH_AT	*_17290	5	*_17300	WP_013145505	5-oxoprolinase subunit PxpB;
					*_17310	WP_014195657	Biotin-dependent carboxyltransferase
					*_17320	WP_011230998	LamB/YcsF family protein
					*_17330	WP_015374692 ⁵	DUF523 domain-containing protein
	<i>Gt</i> [§]	GNAT_T	*_17280	1	- ³	- ³	- ³
	ZG [¶]	wHTH_AT	Contig 16_18	8	contig16_19	WP_013145505 ⁵	5-oxoprolinase subunit PxpB
					contig16_20	WP_014195657 ⁴	Biotin-dependent carboxyltransferase
					contig16_21	WP_011230998 ⁴	LamB/YcsF family protein
					contig16_22	WP_031206563 ⁵	DUF523 domain-containing protei
					contig16_23	WP_011231000 ⁶	Hypothetical protein
					contig16_24	WP_011231001 ⁵	AMP-binding protein
					contig16_25	WP_021322572 ⁵	Phosphotriesterase-related protein
	ZG [¶]	GNAT_T	Contig 16_17	1	- ³	- ³	- ³
XRE- COG2856	<i>Gk</i> [†]	HTH- COG2856	GK3185 GK3184	3	GK3183	WP_011232653	Hypothetical protein
	<i>Gt</i> [§]	HTH- COG2856	*_35630 *_35620	3	*_35610	WP_044741960 ⁴	Hypothetical protein

45 [†] *G. kaustophilus* HTA426.

46 [§] *G. thermoleovorans* CCB_US3_UF5.

47 [¶] *Geobacillus* sp. ZGt-1.

48 * Stands for "GTCCBUS3UF5" which is part of the locus tags in *G. thermoleovorans* CCB_US3_UF5.

49 ¹ As predicted by the Operon Mapper tool, except for *_35630 and *_35620, whose putative operons were detected using the
50 ProOpDB.

51 ² Proteins other than TAs, whose genes are sharing operons with TA-coding genes. The sharing protein IDs and identifications
52 were derived mainly from the NCBI RefSeq genome files with some exceptions which are stated in the next footnotes.

53 ³ (-) indicates there are no genes, and thus no proteins, sharing the putative operon with the solo T/AT.

54 ⁴ Protein ID was derived using blastp; the matching identity was > 99.5% over the entire length, the e-value was equal (or very
55 close) to zero.

56 ⁵ Protein ID was derived using blastp; the matching identity was 100% over the entire length, the e-value
57 was equal (or very close) to zero.

58 ⁶ Protein ID was derived using blastp; the matching identity was 100% and the protein coverage was 89%, the e-value was very
59 close to zero.

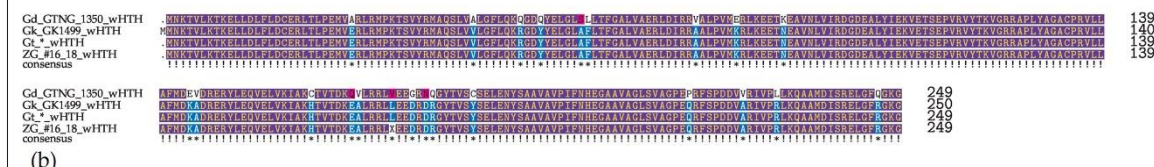
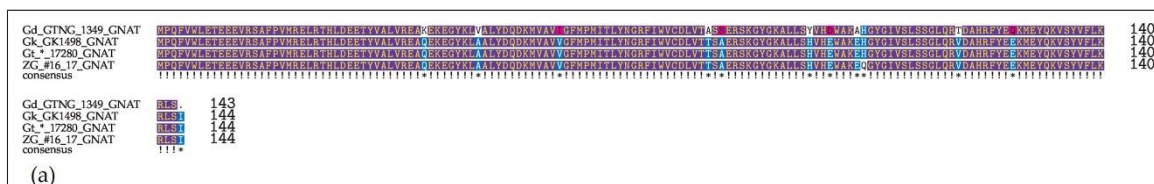


Figure S1. The GacTA TA family. (a) Global alignment of the GNAT toxins; (b) Global alignment of the wHTH domain-harboring antitoxins.

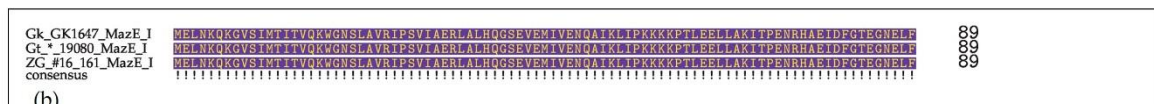


Figure S2. The MazEF (I) TA family. (a) Global alignment of the MazF (I) domain-harboring toxins; (b) Global alignment of the MazE (I) domain-harboring antitoxins.

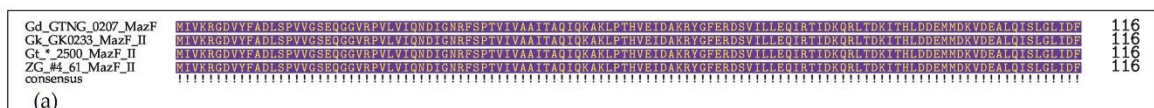


Figure S3. The MazEF (II) TA family. (a) Global alignment of the MazF (II) domain-harboring toxins; (b) Global alignment of the RHH domain-harboring antitoxins.



Figure S4. The MNT-HEPN (I) TA family. (a) Global alignment of the COG2361 domain-harboring toxins; (b) Global alignment of the COG1669 (A) domain-harboring antitoxins; (c) Global alignment of the COG1669 (B) domain-harboring antitoxins.

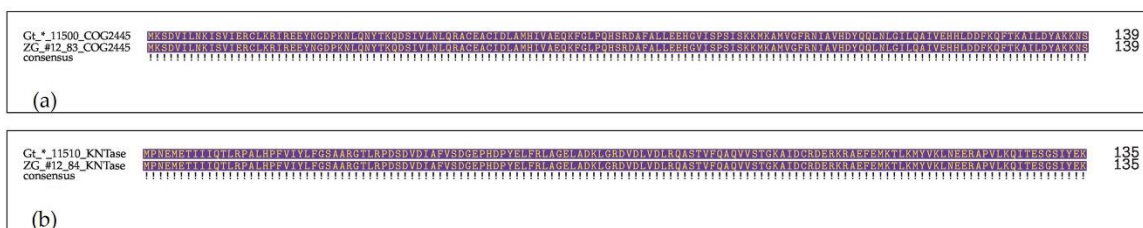


Figure S5. The MNT-HEPN (II) TA family. (a) Global alignment of the COG2445 domain-harboring toxins; (b) Global alignment of the KNTase domain-harboring antitoxins.



Figure S6. The ParDE TA family. (a) Global alignment of the ParE domain-harboring toxins; (b) Global alignment of the AbrB domain-harboring antitoxins.

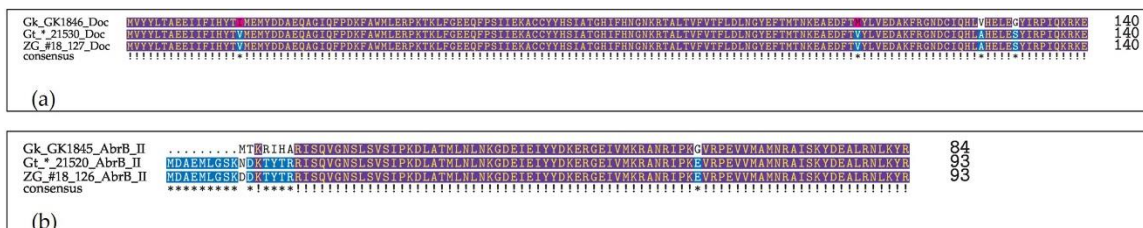


Figure S7. The Phd-Doc TA family. (a) Global alignment of the Doc domain-harboring toxins; (b) Global alignment of the AbrB domain-harboring antitoxins.



Figure S8. The RelBE TA family. (a) Global alignment of the RelE domain-harboring toxins; (b) Global alignment of the HTH domain-harboring antitoxins.

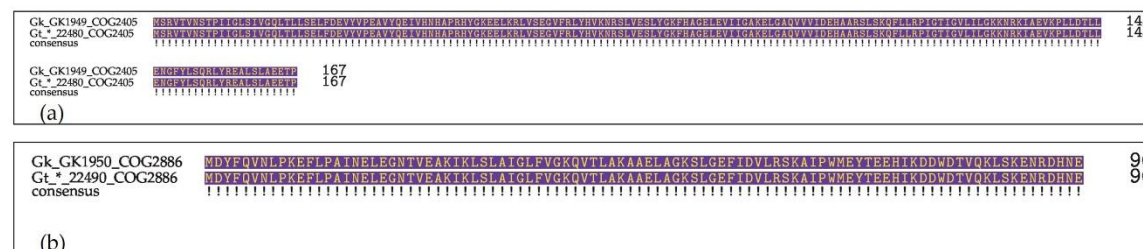


Figure S9. The VapBC TA family. (a) Global alignment of the COG2405 domain-harboring toxins; (b) Global alignment of the COG2886 domain-harboring antitoxins.

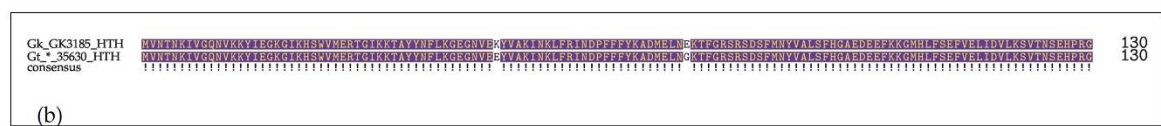
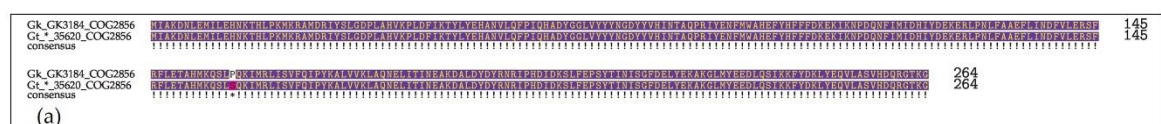
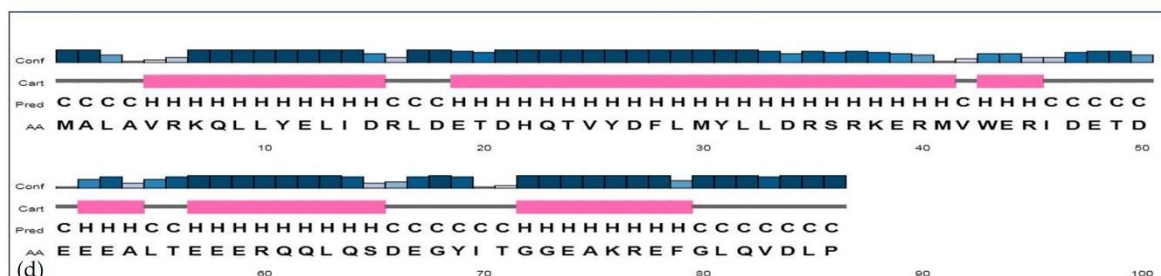
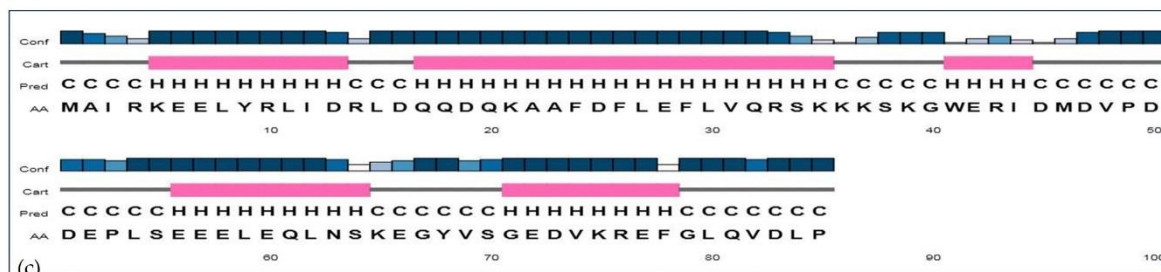
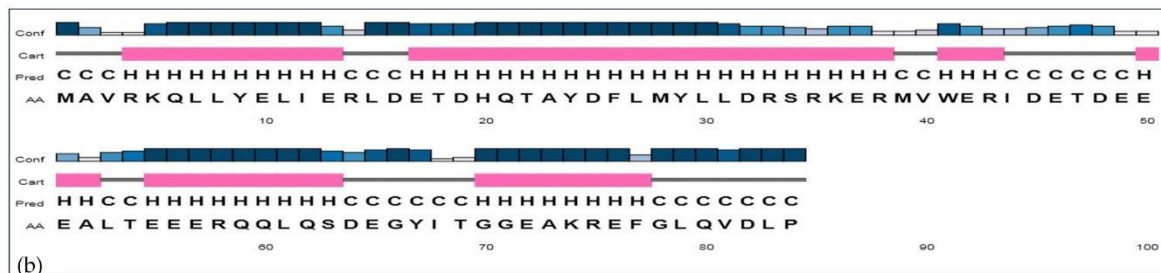
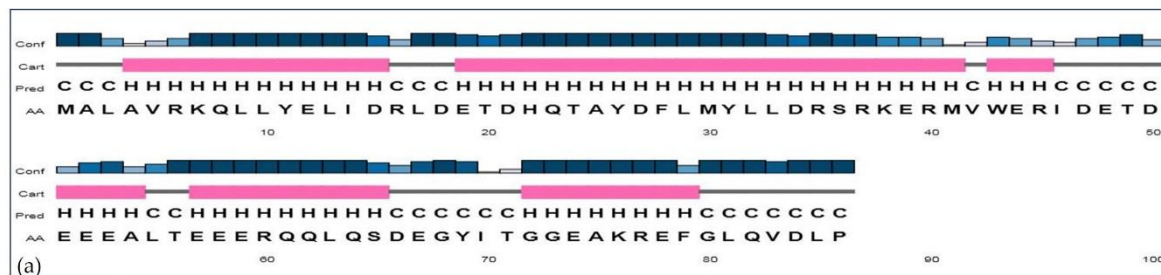


Figure S10. The XRE-COG2856 TA family. (a) Global alignment of the COG2856 domain-harboring toxins; (b) Global alignment of the HTH domain-harboring antitoxins.



Strand
Helix
Coil

Conf: - + Confidence of prediction
Cart: 3-state assignment cartoon
Pred: 3-state prediction
AA: Target Sequence

Figure S11. Secondary structure prediction of (a) GK3105; (b) contig 25_196; (c) WP_066367164; (d) *_34820 using the PSIPRED workbench tool [145].