The current state of *Subti*Wiki, the database for the model organism *Bacillus subtilis*

Tiago Pedreira, Christoph Elfmann and Jörg Stülke^{®*}

Department of General Microbiology, Institute of Microbiology and Genetics, Georg August University of Göttingen, Grisebachstr. 8, D-37077 Göttingen, Germany

Received September 09, 2021; Revised September 28, 2021; Editorial Decision September 29, 2021; Accepted October 01, 2021

ABSTRACT

Bacillus subtilis is a Gram-positive model bacterium with extensive documented annotation. However, with the rise of high-throughput techniques, the amount of complex data being generated every year has been increasing at a fast pace. Thus, having platforms ready to integrate and give a representation to these data becomes a priority. To address it, SubtiWiki (http://subtiwiki.uni-goettingen.de/) was created in 2008 and has been growing in data and viewership ever since. With millions of requests every year, it is the most visited *B. subtilis* database, providing scientists all over the world with curated information about its genes and proteins, as well as intricate protein-protein interactions, regulatory elements, expression data and metabolic pathways. However, there is still a large portion of annotation to be unveiled for some biological elements. Thus, to facilitate the development of new hypotheses for research, we have added a Homology section covering potential protein homologs in other organisms. Here, we present the recent developments of SubtiWiki and give a guided tour of our database and the current state of the data for this organism.

INTRODUCTION

With the rise of high-throughput techniques, the complexity and amount of information have been increasing significantly over the past years. Importantly, the new techniques not only allow the acquisition of one-dimensional datasets such as new sequences or protein catalogs, but can also generate bi- or even multidimensional datasets that include two or multiple partners such as in global protein–protein, protein–RNA or protein–metabolite interactomes (1–3). A possible approach to catalog and organize the data is by resorting to specialized databases. One type of these databases is dedicated to model organisms, which play a fundamental role in the advance of knowledge in science as they are widely used across the world (4,5). The databases

can widely differ in their intended audience and purpose, and accordingly, the presentation of the data varies substantially. Clearly, databases with a general audience in mind need a more intuitive structure and way of presentation as compared to databases that are designed for informaticians who are mainly interested in the technical possibilities to extract information.

The Gram-positive model bacterium *Bacillus subtilis* is one of the most intensively studied organisms. This bacterium has gained much interest due to the simple developmental program of sporulation, its genetic competence and the ease of genetic manipulation, because of the wide applications in biotechnology and agriculture, and since it is a close relative of many important pathogens such as Staphylococcus aureus, Clostridioides difficile or Listeria monocytogenes (6,7). We have previously developed SubtiWiki, a database focused on the functional annotation of B. sub*tilis* (8–10). In *Subti*Wiki, the user has access to individual gene and protein pages, complemented by SubtiApps, which give access to integrative metabolic pathways, regulatory networks and manually curated protein-protein interaction maps (10). As of today (September 2021), SubtiWiki contains data on 6121 protein- and RNA-coding genes, 2271 operons, 8355 literature references and >2660 documented protein interactions. Additionally, 49 metabolic pathways are displayed in an interactive map, a genomic interactive map is provided for the full extent of the B. subtilis genome and gene expression pattern profiles are presented for 121 experimental conditions, which can be on the transcript or protein levels.

Despite the current advances in the research on *B. subtilis*, there is still a large portion of genes to be further annotated. This is the case for $\sim 30\%$ of all *B. subtilis* genes/proteins. We try to address this issue by providing the user with a broad list of protein homologs in different organisms. To complement this, we also integrated the Cluster of Ortholog Genes (COG) database for every available protein (11). Thus, beyond providing the community with an up-to-date and free to use platform, we also contributed to the annotation of this model organism. Here, we report the current state of *Subti*Wiki and the integration of new data, and describe the most recent addition, the homology analysis.

*To whom correspondence should be addressed. Tel: +49 551 3933781; Fax: +49 551 3933808; Email: jstuelk@gwdg.de

© The Author(s) 2021. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

⁽http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

				Pathway	Interaction	Expression	Genome	Regulation	Regulon	History	Log in
{SubtiWiki}											
lJ											
				unt in Drom		ataa watii taha Qi					
The European Conference	conference w	ebsite!	:n - 27th Aug	ust in Pragu	le - please regi	ster until the 20	uth of July,		llights		
The 21st International Confer	ence on Bacil	li has been pos	tponed to 20)22 and will	take place in F	Prague.		• Pa	aper of the mo acillus labs	nth	
citB							16	• Al	l categories andom gene		
aconitase, trigger enzyme									ease cite us ^_ redits	^	
Locus	BSU_18000							Spec	cial pages		
Molecular weight	99.14 kDa								ene export wiz (ports	ard	
Isoelectric point	4.9								ser list istory		
Protein length	909 aa	Sequence	Blast					• St	atistics ownloads		
Gene length	2730 bp	Sequence	Blast						iniBacillus Cor	npendium	
Function	TCA cycle							Sear			_
Product	aconitase, 1	rigger enzyme						Gen	e / locus	tag	Go
Essential	no										
E.C.	4.2.1.3							Publ	Med		
Synonyms	citB							Tit	le, author	, id	Go
Outlinke	KEOO	BaubOue	Cubtil int	l la Dr		ession		Stru	cture		
Outlinks	KEGG	BsubCyc	SubtiList	t UniPr	ot Brow	ser			5		
Genomic Context										57 Jus	
									2	E B	
				□Hide	small RNAs	Zoom:	-				
rA CcdA	yneJ		citB			yneN				A STATE	
0 0 0 0	• •	000				00			La B		
1924000	•	1926000	0 00	19280	000	19300					
61								Expr	ression		
									otB [15	26680,1929409,+1) clust	ers A1042 size: 4 B210 size: 29 CM size: 19D
	Gene / ne	w RNA feature	Upshift	(transcript	5' end) 🕴 Do	ownshift (trans	cript 3' end)	0	My	WA Pr	M
								*- *-	1935 x x <s x="" }<br="">1935 x x x x x x }</s>	112 5 0 5 0 4 0 1 1 1 5 5 5 5 28 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	101 - 1011-10 10 - 1011-10
Categories containing th			TOA								
 Metabolism, Carbon metabolism, Amino acid/ ni 					ition of branch	ed-chain amin	o acids	Inter	actions		
	· · · · ·		- ·	- ·			r			CitG	

Figure 1. The *citB* gene page display. All gene pages share the same layout and follow the same structure, with changes that depend on the type of available data.

SubtiWIKI GENE PAGES

Similar to previous versions, version 4 of *Subti*Wiki keeps the general design and page layout. Upon searching for a gene or protein, the respective biological element page is presented in a simple and intuitive way (Figure 1). Specific and factual information is promptly presented in the form of a table, such as data on locus name, and in the case of protein-encoding loci, isoelectric point, molecular weight and function. Additionally, it is possible to directly access NCBI Blast tools with either DNA or protein sequences and the user gets access to other relevant database links that also have information on the queried protein, such as KEGG, UniProt or BsubCyc. Moreover, curated interactive information on specific data such as protein structure, protein–

Genome Browser

High und Wild Zorre Inco	yneB SirA CocdA yneJ CitB yneB II III IIII IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	1934000
High and MARK V Correction of the control o	ent sequence: cit8 wing + strand techter and transcort T	1934000
List	ent sequence: citB wing + strand citils arbstilfs 168 BST_18000 citB CARCG ACCMARAGE TOTOCOTATE TECCARGO TECTARGONAR ACCTTACCA COMARGE GAAATTEG CACTTACAA TACTACCTTACT TAMAAGCGT TAGAAATTEG CACTTACAA TACTACCTTACTACAA ACCATA TACCACCTOCAA OTECTTACTA AATCACATACA TACTACACTGC CATACCACATE CACAAATGGGG AACCTTACAA TACTACACCTTACAAATACGA TECTACACACT TECTAAAATCACACCTCCAA AACCTACCAAA GAAAATCACA CACAAACGGG TACCACACTGC CACACCTACAAC TACTACACAAAT CACAAATTATA TCAACCTGCAA AACCTACACAAA CACAATCAAATCA	1934000
List	B Ht BIA Dirac CCCIA Ynely CliB Ynely Up 100000	1934000
NUMBER	100000 1020000 1020000 1020000 <td>1934000</td>	1934000
Note: Note: <td< td=""><td>192000 1920000 1920000 1920000 1920000 1920000 1920000 1920000 19200000 19200000 19200000 19200000 19200000 192000000 1920000000 1920000000 1920000000 19200000000 1920000000000</td><td></td></td<>	192000 1920000 1920000 1920000 1920000 1920000 1920000 1920000 19200000 19200000 19200000 19200000 19200000 192000000 1920000000 1920000000 1920000000 19200000000 1920000000000	
SSG1 C C C C C C C C C C C C C C C C C C C	S661	
	Gene / new RNA feature Upshift (transcript 5' end) Downshift	transcript 3' en
	nt sequence: citB ing + strand illus subjells 168 BSU_18000 citB MARCA RACAMANAC TOCAGGAMA GACGTITITIC MACGATATA CAMATGGA AMACATATA TRACTACTOT TIAMAAGGT TAGAAGATTC AGGTATAGGA MACGA RACAMANAC TOCAGGAMA GACGTITITIC AMACGAGAAA AACGTITACT ACAMATGGA AMACATATA TIACTACTOT TIAMAAGGT TAGAAGATTC AGGTATAGGA MARCA RACAMANAC TOCAGGAMA GACGTITITIC AMACGAGAAA AACGTITACT ACAMATGGA AMACATATA TIACTACTOT TIAMAAGGT TAGAAGATTC GOGGAGATA TICGA MACTICCTTA TICCATCAAA GITCTTTTA AMACGAGTAT GACGACTITACT ACAMATGGGA AMACATATA TAGAAGATTC GOGGAGATA MARCA RACAMATTTT MACTOCGTA COTCOCGTA TICTACAGAG GATTAGAGATT GACGACTATC GACAAGATTA GACAGACTAC COMACAAAA TICATAGAAA Search in DNA sequence MARA TICAACCTGAGA MACGAATATA TICTAGGAGAA CACAGTAC CACACAGACT AGATTAGAG TACCACTAGAT TICTIGGACAA TICTITIGGCAAA GACACACTA Search in DNA sequence Search in DNA sequence MARA TICAACCTGAG AMACGAATTA TICTICGAACA CACACAACAA CAGACTACAC GAACAACAA GACACAGTA CACACAGAT TICTIGACAGAC COTTOTIACAAAA TICTICGGAAC COGGATAATTA TICTICGAACACGAA CAGACTACACA GATATATGG GAACAACACA GACTACGACA TICCGAACAACAA Sector AMAAAAGGGGT TICTICGAACAC COAGACAAACA CACACAACAA GATTAAGGG CACACAGACT CACGACAGACT CACACACAAA	transcript 3' en
	the sequence: citB ing + strand illus subcilis 168 BSU_18000 citB acacca AccAnamata Tractacacta Concentrate Concentrate Contractact Concentrate C	transcript 3' en
	the sequence: citB ing + strand illus subcilis 168 BSU_18000 citB acacca AccAnamata Tractacacta Concentrate Concentrate Contractact Concentrate C	transcript 3' er
ng + stand ilius subcliis 168 BU 18000 018 MAGE AGAAAAAA TUGAGGAMAA GAGGETTECE AAGGAGAMA AAGGETTACE ACAAAGGA MAACAATACA TACTACTCE TITANAAGGET AGAGETAGA GAGGATAGA GAGAGATACA AGAGETAGA GAGATAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	ing + strand illus subcliss 168 BSU_18000 citB MARCA RACAMAMAR TOCARGAMA GACGTITICO ANGORAMA AACGTITACT ACAANGGGA AAACATATCA TTACTACTCT TTAAAAGCGT TAGAAGATTC AGGTATAGGA RAGCITCOTTA TICCATCARA GITCITTITA AATCAGTATT GCGTCAGATT GACGGCTTCG THATCAAATA GAGAACAGCG GAAAATTIGG CAAANGGGG AAACCGCCGA RATA TCCAGCCGCG TCCGTGGTAT TITTACAAAT TITTACAAAA CTTCAGAGA GATAAGGGG GACAGAGA TAGCAGTCG GAAAATTIGG CAAANGGGG AACTGCCCGA RATA TCCAGCCGCG TCCGGGTAT TITTACAAATTITTACTAACAATTG GATAAAGGGG GACAGACA CTCCATRAGA CGCGGGGAGAT RAAA TCCAGCCGCG AAGAAAGGGCT TCCGGGTATA TCCAGACACCC ACACACAACA TGATAAGCGG TACGCAGGTC GAAAATTIG CGAACAACCA RAGA MAGGGGGA CUTTGTAGA ATCCTGGTA CGACAGACTA CACACACAA TGATTAGCG TATCGGGTATG TCCGGGGGAGAT GGGTA ACAAAGGGGGA CUTCTATACAATTA TCTGGGAACACCA CACACACAA TGATTAGCG TATCGGGTATT GCGCAACAACA TGCTTGGACA CCCTTCTTAC TTCCCCAGTC CACAGGTAAT CCCCCCGGAACAACA CTTCTGGCGCAAGGTT TGCGGGATGGG TATCGAGACACCAA GGGTA ACAACGCGGG TGTCGGTAAA TTCTTCGGAAC CGGAACAACA CACACCAAG GCTACGAACA GCTACGGACA TGCGGGAACACAACAC	
ng + stand Lius subilitis 168 BSU 1800 c18 AAGG ACCAMADAG TECGEGAMAA GACETTATE AAGGAGAMA AAGETTAGE AAAATAGAA TIAAACAACTE TIAAAAGGG TAGGAGAATAGAAGAA AAACAACTE TECAATAGAA TECGEGAMAA GACETTATE AAGAAGAGA MACETTAGE AAAATAGAA TIAAACAACTE TIAAAAGGG AAAGCAGGE AAGGATAGAA AAAATAGACTE TECAATAGAA TECCATTA AATAGAACATE CONACAAAAA AAAAACAACTE TIAAAAGGAG TAGGAGAAAAA AAAAAACAACTE CONACAAAAA AAAAAAATTETT AAACTGOGG AAAAAGGE TITAACAACTE CONACAAAAAA AAAAACAACTE GAAAATTEG GAAAAATTEG GAAAATTEG GAAAATTEG AATAACTAACTE TIAAAACTEG CONACATATEG GAAAATTEG AATAACTAACTE TIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	ng + strand illus subjijs 168 BSU_18000 citB MAGCA GACAMAANAG TGCARCAMAA GACGTITACT ACAMATGGGA AMACATATCA TTACTACTCT TTAMAGCGT TAGAAGATTC AGGTATAGGA AGCATCCCTTA TTCCARCAMA GACGTITTCC AAGCGAGAMA AACGTITACT ACAMATGGGA TAACACATCA GAAAATGGGA CAAAATGGGGA AACTGCCGAA GACTTCCTTA TTCCARCAMA GTCCTTTTAG AAACAGTATT GCGCACAAGT GACGGCTTCG TAAACAAGAGATC GGCTAATGCA GAAATATGG CAAAATGGGA AATGCCTGAA MAAA TCCAACCCTGA ANTICCTGGT AATCCAGTTA TCGAGCAGTA CGGCCTGGA AGGTATGGG GAACAGACG GTAAGGAGA TCGCGGAAA TGGA AGGCGGGA GCTTGTAAGT TATCCAGGGG GTAAAGCGG GTACAGAAAC GGTACTGGG TCTGGGCAAGT CGGAAAGAGT CGCAAGAAAT GGCT ACAAACTTTT AAGCTGGGCA AAGAAAGCGT TTACCAAGTA CGCTCTGCGAA AGGTATGTG GCACCAGGT CTGCGGTGA AATGAAACAA GGCT ACAAACTTCTT ATCCCAGTTC CAGGGAGAAA CTGTCGGAAA CTGTCCGAAC AGGTATGTG GCAACAAGA GTACTGGAAT GGCGA AAAAGGCGGA GCTTGTAAAG TATCCTGGGCA GGAAACGCAC GAAACTACG GTACTGGACAGA GGCGA AAAAGGCGGA GCTTCTTAAA TTGTCTGGACC GGGAAAGAAC CTACCGAAACAA GCTACTGGACT CTGGGGTAAAA GTACACAAA	
ng + stand Lius subilitis 168 BSU 1800 c18 AAGG ACCAMADAG TECGEGAMAA GACETTATE AAGGAGAMA AAGETTAGE AAAATAGAA TIAAACAACTE TIAAAAGGG TAGGAGAATAGAAGAA AAACAACTE TECAATAGAA TECGEGAMAA GACETTATE AAGAAGAGA MACETTAGE AAAATAGAA TIAAACAACTE TIAAAAGGG AAAGCAGGE AAGGATAGAA AAAATAGACTE TECAATAGAA TECCATTA AATAGAACATE CONACAAAAA AAAAACAACTE TIAAAAGGAG TAGGAGAAAAA AAAAAACAACTE CONACAAAAA AAAAAAATTETT AAACTGOGG AAAAAGGE TITAACAACTE CONACAAAAAA AAAAACAACTE GAAAATTEG GAAAAATTEG GAAAATTEG GAAAATTEG AATAACTAACTE TIAAAACTEG CONACATATEG GAAAATTEG AATAACTAACTE TIAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	ng + strand illus subjijs 168 BSU_18000 citB MAGCA GACAMAANAG TGCARCAMAA GACGTITACT ACAMATGGGA AMACATATCA TTACTACTCT TTAMAGCGT TAGAAGATTC AGGTATAGGA AGCATCCCTTA TTCCARCAMA GACGTITTCC AAGCGAGAMA AACGTITACT ACAMATGGGA TAACACATCA GAAAATGGGA CAAAATGGGGA AACTGCCGAA GACTTCCTTA TTCCARCAMA GTCCTTTTAG AAACAGTATT GCGCACAAGT GACGGCTTCG TAAACAAGAGATC GGCTAATGCA GAAATATGG CAAAATGGGA AATGCCTGAA MAAA TCCAACCCTGA ANTICCTGGT AATCCAGTTA TCGAGCAGTA CGGCCTGGA AGGTATGGG GAACAGACG GTAAGGAGA TCGCGGAAA TGGA AGGCGGGA GCTTGTAAGT TATCCAGGGG GTAAAGCGG GTACAGAAAC GGTACTGGG TCTGGGCAAGT CGGAAAGAGT CGCAAGAAAT GGCT ACAAACTTTT AAGCTGGGCA AAGAAAGCGT TTACCAAGTA CGCTCTGCGAA AGGTATGTG GCACCAGGT CTGCGGTGA AATGAAACAA GGCT ACAAACTTCTT ATCCCAGTTC CAGGGAGAAA CTGTCGGAAA CTGTCCGAAC AGGTATGTG GCAACAAGA GTACTGGAAT GGCGA AAAAGGCGGA GCTTGTAAAG TATCCTGGGCA GGAAACGCAC GAAACTACG GTACTGGACAGA GGCGA AAAAGGCGGA GCTTCTTAAA TTGTCTGGACC GGGAAAGAAC CTACCGAAACAA GCTACTGGACT CTGGGGTAAAA GTACACAAA	
Substiles 168 BSU 1800 cits MAGE MACANAARC TECACAAAA GACOTTECE AAGCAGAAA AACGATACT ACAAATGGGA AAACAATAC TAACAAGGT GAAAATGGGA ACAAATGGGA AACAATGGGA AACAAGGT GAAATGGGA AACAATGGGA AACAAGGT GAAATGGGA AACAAGGT GAAATGGGA AACAAGGT GAAATGGGA AACAAGGT GAAATGGGA AACAAGGT GAAAGAGGA GAAAGGAGA CTATGGAAGAAGA TECACAGGAAATGGA AACAAGGAT GAAAGAGGA CAATGGAAGAAGA TECACAGGAAATGGA AACAAGAGA TECACAGGAAATGGA AACAAGAAGAAGA AACAAGAAGA AACAAGAAGA AACAAGAAGA AACAAGAAGA AACAAGAAGA AACAAGAAGA AAGAAGAAGA AAGAAGAAGA AACAAGAAGA AAGAAGAAGAAAAAAAA	Search in DNA sequence Search in DNA sequence Search in DNA sequence Search in DNA sequence Address and a control of the search in DNA sequence and a second second second search a control of the search in DNA sequence and the second	
AAACCA AAAACC TOCAGGATA GACCTTTTCC AACCGATA AACCTTTCC AACCGATAGA AACCTTACT A TACAATGGA AACCTTACT TAAAACGAT TACAGT GACCATACA TACAATGGA AACCTTACAA GACCATACA TAGACATCA GACTAGATACA TAGACATCA GACTAGATACA AACCGATACA TAGACATCA GACTAGATACA AACCGATACA TAGACATCA GACTAGATACA AACCGATACA AAACCGATACA AAACCGATACA AAAACCGATACA AAAACCGATACA AAAACCGATACA AAAACCGATACA AAAAAAAAAA	AAACG AGCAMAAMAC TGCAGGAAA GACGTTTTCC AMGCGGAAA AACGTTTACT ACAAATGGGA AMACHATCA TTACTACTCT THAAAGCGT TAGAAGATTG GGAAATGGGA TTCGA AGCTTCCTTA TCCATCAAA GTTCTTTTG AMTCAGTATT GCCGTCAAGTT GACGGGCTTCG TAATAAAGCAGT GGAAAATTGG CAACTACGAGA GGATA TCCAGCGTCC GTTCGATCCACCC GTCCGTGTTA TTTTACAAGA CTTCACAGGG GTTCCGGCA TAGTAGAACGCGT GGAAAATTGG CAAATATGG AGACGCGT TAAAA TCAACCCTCA AMTTCCTGTT GTTCACACGG GTCCGGCAT TAGTAGACGCG GTACGGGGT CGCATAGAGA TGGCAGCTGC GGAAAATTGG CAACTAGAGA GGGAT ACCAACGGCG ACAAAAAGCG AAGAAACGCG TACCGGCGTG CAACTAGAGA TGCACTAGAGA TGCAGAAGAC GGGAT ATCCAGCGCG AAGAAAGCGCG AAGAACGCG AACGACGCA AGCTAGCGG TACTGGGGT CTCGGCGGGG GATTGGAGCGT TGAAG AMGACGGGGA GCTTGTAAAGT TACCGATTG CGGCGCGGAA CTTGTCGGCA AGCTAGGAG TACGGGGTT CTCGGCGGGG GATTGAAGCT GGGAT GCCTAGAAGCGCG GCTTGTAAACTATCG CAAGAGCTG CGGAACTGCC ACCTGGAGCA GCTCGGACGTC TCCGGCTGGG GATTGAAGCT GGGAT GCCTGGAAGCGCT TGCCGGTAGTA TCTCGGGACG CGGAACTGCC GGAACAACGA GCTACCGACT TGCGGCTGTA AGCTAGAAC GGGAT GCCTGGAAGGCT GTCGGTAAT TCTCGGACGC GGAATTGC GGGACTGCG GCGACAACG GCTCCGGATA GCTAGAACGC GGGAT GCCTGGAAA ATTGTCGGACG CGGGATTGCT GGGGACTGCG GGAACGCG GCCGGAAACG GCTCCGGATAGCG GGGAT GCCTGGAAGGCT GTCGGGATAT TCTCGGGACCGGGAA CTTCTCGGCGACAACG GCCACGAACAG GCTCCGGACAACG GGGAT GAAGGCGCG TCCGGAATAT TCTCTGGACC GGGAATTGC GGGACTGGC GCGACAACG CCTCGGATAT GCGAATTGCG GGGAT GAAAAGCGGCT TCCGGAATAT TCTCGGACGCG GGAATTGCG GGGACTAGG GCGACAACGC GCTCGGAATAGGA GGGAT GAAAGCGGCT TCCGGATAT TCTCGGACGCGGAACGCA GCTGCGGAAT GCGAATTGG GCAACAACAA	
DEDUX TOCACTICO CITCUANCC TOTIONTA TITUTCUANC CITCUCADG DIVOGOLA TARANTC GUITANACA TOCATAGO CITAAACAA TAAAACAA TO CONTACT CITAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAA	SGATA TCGACGTTCC GTTCAAACCG TCTCGTGTA TTTTACAAGA CTTCACAGGG GTTCCGGCA TAGTAGATCT GGCTTCACTG GGTTCACTG CGGAGAGAT TRAAA TCAACCCTCA AATTCCTGTT GATCCGTTA TCGATCACTC TGTACAGGTA GATAAGGGG GTACAGAGA TGCATTGCG GTTCACTGGAGT GGAGAAAT GGCT ACAAATTTTT AACTGGGGCA AAGAAAGCG TAACAGAGTA GATAAGGGG GTACAGAGA TGCATTGCG GTTCACGATG TCTTGCGAAT TGGAG GTCGAAACAGCGCA ACGTAGAGCG TAACCAGTAT TCGGCGTGTA GATGAGGATT CTGGGGAGG GTTCGGGCGGGAAACGA GGGA GGCTGGAAACAGCGCA GCCTTGTAACCTTTCTGGGAGA CTGTCGGGCG AACGAACAACA GGAACAACA GGCAACAACA GGCACGACA GGGA GGCTGGAAACAGCGCT GTCGGTAA TTGTTGGAAT TCTTGGGCAGT GAACGGCGCGGAAACGACA GGGA GGCTGGAAACGGCG GCCTGCTAACC GGAAGTAGCT GAACGCGCG TGTCGCGAACGACAACA GGCACAAAT GGCAACTAGG GTCTCGGCGTGA AGTAGACCAA GGGGA AGGCGGCG GCTTCTTACACTTCTCGGACC GGAAGTAGCT GAACGCGCG TGTCGCGAACGACACACA GGCACAACA GGCACAACA GGGAA GGCTGGAAACA TTGTGGAACT TCTCGGACC GGAAGTGCC GGACGACGACA GCTACGACAAT GGCACTATAG GCTCCGGATAA AGTAACCAA	
DANA TCACCCTGA ANTICCTGT GATCOGTAT CONTACTC TOTACHOCT OF CARAGED GATABAGGE GIACAMAGE COTAMATAGE ACTIGANT COLAMAGES GGCT MAANTITY ANGAGGEGA ANAMAGGET TARACAMATE CARGEGGAR CORCOGADA GATATTEG COCCAGGEGAR COTTOCTOGTA GGGA MAANTITY ANGATGEGA ANAMAGGET TARACAMATE CARGEGGAR COTACGAGGAR COTACGAGGEGA COTTOCAGA GGGA MAANTITY TARACAMATE COTTOTAGE ANAGAGATCA CARACANATE ACTIVICADA COGARATE COGAGGAR COTTOCAGA UNIT TOTCCARE TTOTAGAGE TARCCONT COGACUMA COGACUTA CARACANATE ACTIVICADA COGARATE COGACUMA COGACUTA COGACUTA GGGA MAAGGATTAT TATUTTYNT TOTTOGAR COGACUTAC TOGCCOTAL COGACUTAC COCACUMA COGACUTA COGACUTA COGACUTA COGACUA MARGE ANTAGTAGE COTTOCACUTA COCCCTCA COGACUMA TOGCCOGA COGACUTATE COCACUTA TAGACUCCA GGGA MAACCACUTAT ACTURATOR COCCOCCTCA ACCAGGATTAC COCACUTA TOGACUTA TAGACUTA TAGACUTA GGGA ACCCCTACACUTACUTACUTA COGACUTA ANAGE ANTAGTAGE COCTOCACUTA TACUTACUTA TOGCCOCAU TACACUTACUTA COGACUTA ANAGE ANTAGTAGE COCTOCACUTA TACUTACUTA TOGCCOCAU TACACUTACUTA TOGCCAU TACACUTACUTA COGACUTA TACAUACUTA TOTAGATA ACCOUNCE ANTAGUA ANACCUTAT TACUNATOR COCCOCAU TOGACUTAANATI ACTURATOR COCCOCAU TOGACUTAANATI ACTURATOR COCCOCAU ANAGE ANTAGTAGE COCTOCAU ANAGE ANTAGTAGE COCTOCAU ANAGE ANTAGTAGE COCTOCAU ANAGE ANTAGTAGE COCTOCAU ANAGE ANTAGTAGE TACAUCUTA TOGACUTAANATI ACTURATOR COCCOCAU ANAGE ANTAGTAGE ACCUTACUTA COGACUTA ANACCUTATE TACAUCUTA TOGACUTAANATI ACTURATOR COCCOCAU ANACCUTATE TACAUCUTA TOGACUTAANATI ACTURATOR COCCOCAU ANACCUTACUTA ACCUTACUTA TOGACUTAANATI ACCUTACUTA TOGACUTAANATI ACCUTACUTA TOGACUTACUTA AACCUTACU	IRANA TCACACCTCA ANTICTCTUT GATECTETA TCATEACACTE COTACAGETA GATEANAGEGA GTACAGAGA TOCACTIGACA GTANATATEGA ACTIGARATE CGAANAGAANE GEGET ACAANTITT AAGETGEGEA AAGAAAGEGET THACAATTA TCAGGEGAGTA CEGECTECAL AGGETATEGT GEACLAGETA CATEGOGAGA GETTEGTAGE AACCTAGAAC CACAGAACA AAGAAAACA AAGAACAACAA TGATTATACG TATCGEGGETTE CEGEGEGEG GATEGGEGA AAGTAACCAA GRAGA AAGAGGEGA GETTEGTAACE TATCCCAGATA CGETTEGTGG AACAGACTA CACACACAA TGATTACGG TATCGEGGETTE CEGEGGEGE GATEGGEGG AATGAAGET GEGAT AGAAAGGEGET TEGTEGGATCCCAGATT TETTEGGACC GGAACTACCA CEGAACAACAA TGATTATAG GGAACTAGE CICCGGAATA GATAACACAA GEGTA AAAAAGGEGET TETEGGATTAT TETTEGGACC GGAATTGGT GGAACTACGE CICCGAATTAT GGAATTATAG GGAACTACAA	ear
RAMA GAAGGGGA GCTTETTAAGT TACCTGATA GCGTTETTGG AACAACTA CACACACA TATTATAGG TATCGGTTET CTGGGTGGTTE GTGGGTTAA GTACAACTA GGGAT GGGTGGGGGGG GTTACAGGG GTGACAGGAC GGGGGGGAA GTATGGGG GAAGTGGGG GTGGGGGGAA GTGGGGGGTAA GGGATAGG TAGTAGGGGT GTGGGGTGAT TGGGGGGGG GTAGTGGG GGGGGGGG	reans anancegora ectretrance tarcegnata costificiende ancagarca cancenara trattances tarcegnate eregostes de antenarce Segan tectrograf ectretrar treccastre creargetaat egescegnaa ettetecena cegancanca egeancanca eccastregat teges Seges anangecet tetegetaat tittetegaat tettegege egescegnates etechentes cocancate gegantate costectaca	
Segge AAAAAGede Tercegraam tregterer conserve and the segment of the second terce and	GCGTG ARARAGEGET TETEGEGRARA TITETIGRAT TETEGERAEC GEGRATIGET GRACIECEC TIGERERICE GEGRACARTI GEGRATATEG ETECEGRARA EEGENERA	
NTET TECCAGTAGA TAGAGAGGG GTTAATGAC GEGGEGGEGE TEGGEGGEAT CETGAAGATA TEAGTETET TAGAGGATA TEGAGGATA TEGAGGATA GAGGAGAG AGGGAGAT GAGGAGAG GAGGAGAT GAGGAGAG GAGGAGAG GAGGAGAG GAGGAGAG GAGGAG		
AANCE ANTIGETEGE CECTEGEAGET AACAAGETE TEGETTAAA TEETENGGGA GAAAKINAG AATTAGET TAAACTECET AACGEGGAA GAAKEGAGT GECCA TIGETEGER TACAAGETE AACAAGTET AACAACTET AACAAGETTA GEGEGEGGAE TEGTAAGEGGE TEGTAAGEGGET GAATAGEGGAT GAACGEGAA GECCA TIGETEGER TACAAGETE TAACAAGETE TAAAGAGE TIGEGEGEGGAE CETECTAAGE TEGAAGAGEGET GEGEGEGAE CETETAAGEGET TAAAGTEGAT AACAAGEGAT TETAGEGGET TETAAAGTE GTAACAGEGAT ACTITETEGAA TEAGGECTE CETECATACA TGAAGAGEG TEGTAAGEGAT ACCONTAC TATAAGEGAT ACCONTAC AACAGETTAA TETATE TEGTECGEGATE GAAGAGEGE TIGEGAAAAA TAATTITETE ATTAAGEGATE CETECTAAGEAA TIGAAGEGAE TETAAAGEGATA ACCONTAC TATAAGEGAT ACCONTACT TAATTITAC CECCGETTE ACCOGAAAT GAAGEGATAA ATTAATTAA AAACGEATCA TATACEGATAC TITAAGEGAE GEAATTACCE AACAGTETAA TIGGE CETCATEGE GUAAATCAAT GEACTTOTTA ACCAAACTE TACCEGCAA AACAGTEGA AAACAGTEGA TAACCEGATCA TATACTTAA AAACAGTETAA TAAATGGAE ACGAATCAAT GEACGATTAC TITACGGAE GAAGTETGGA AAACATTAG AAAGTEGAA CATACTEGATAA GEACATTACT TACCEGGGA GEAATGGEGA AAACATTAGE AAACATTAGA TAAACTATAA TAAAGTEGAAAAGAGE TITACACAACA AGGEGGAATG AACAGTEGA TAAAT GAACTITATA TAAATGGAA GETAACTAGATCA GAAGCATCAG TATTITEGAA AAACATAGC TITACCEGA GAAGTEGGA AAAAAGAGE CATTAGAACAA GGGGCI CECCGCGAGAACAATGAA CATACATCGGGG GAAGTEGGGA AAACATAGC CECCGGAAA GAATTAGAA AAAAGATG TITACAGAAGA GAAGTEGGGAATG TAAAGTEGAAAATGAA GAACATCACTAGA GAAAAAGAATAA AAGAAAGATTA TACAGAGAGA CATACGGATAA CAACGATAGA GAAGGTGAAGA CECTET CGGGAAAGAA GACACTCACTGGA GAAAAAAAAAAAA		
SECE TECTECTECT ALAMETET ADAMETET CANACCENTA CENECOLATA CETETACIA TECTORECA ALAMEGET GACTAGES ATTACTES ALAMEGET ATATACTAS SECONDAL CANACCENTA CETERARGE TAMESES THACETES ALAMETES CANATASA CANATASA CANTETES ATACATAS ALAGESTA CETERARGES TECHNOLOGY ACCESSION CETERARGES ALAGESTA CANTACTASA CANTERA CANTESES CETERARGES ALAGESTA CANTENAN CETERARGES ALAGESTA CANTESA ALAGESTA CANTESA CANTESA ALAGESTA CANTESA CANTESA ALAGESTA CANTESA CANTESA CANTESA CANTESA ALAGESTA CANTESA ALAGESTA CANTESA ALAGESTA CANTESA ALAGESTA CANTESA CANTESA ALAGESTA ALAGESTA ALAGESTA ALAGESTA ALAGESTA ALAGESTA ALAGESTA CANTESA CANTESA ALAGESTA ALGESTA CANTESA CANTESA ALAGESTA CANTESA CANTESA CANTESA ALAGESTA ALGESTA ALAGESTA ALAGESTA ALGESTA ALAGESTA ALAGESTA ALAGESTA ALA		
TENG GECCETTE ALCGGGARAGE GARANATGE GAGAGAGE TEGGARANA TGATETEE ATTACGTETE TETTEEGE ALGGETAGE TITEGGGARAGE GEATETEGE GECANAGE GEGEGEGEGE GEGEGEGEGE ALGGEGARAGE GEGEGEGEGE GEGARAGEGE GEGARAGEGE GEGEGEGEGE TEGARAGEGE ALGGEGEGE TEGARAGEGEGE ALGGEGEGE TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE GEGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE GALGEGEGE TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE GALGEGEGE TEGARAGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGEGET TEGARAGEGE ALGGEGEGET TEGARAGEGEGET TEGARAGEGEGET TEGARAGEGEGEGET TEGARAGEGEGET TEGARAGEGEGET TEGARAGEGEGEGET TEGARAGEGEGET TEGARAGEGEGE		
THEOT THEOTENCE GEORETTETT GEORGATATE GEOTEGETEG AAGGETAALE ATTANCTAA AAACGATCE AATGETETE GEOLAAGATE GTCAAAACT ATACTTAAT TEGEE GEOLAAGGA GAAATCAA GACAATCATE GACCATAGT TACGCCAGA AAGAGETATCA AAAGGATCE GACAAGATE GACGAAGATE ATACTTAAT TEGEE GEOLAAGGA GAAATCAA GACAATCAA CTAACATCGA AAGACCAACCA TECTITGAA GAAGGATGA AAAGGATGE GACAAGACA AGCGCTGGA CAAATTGAA MARTO AAGCTTAAT TAAATGGAT AACGATCAA CTAACATCGA AAACCAACGA TECTITGAA GAAGATCGA CATCATTAGA GAAGATGGA CATAAAAGA ATGCGTATE TEGEE GEOLAAGGA GAAATCAA GACAATATE CECTEGEGG GGCATCGGA AAAGATGGE CECCCGGAA AFAGATTGA GAAGATGGA CATATAAGGA CATCATTAGA GGCGTTCA TOCACAACT GACCATATT CECTEGEGG GGCATCGGA AAAGATAGG CECCCGGAAA GATTATGCAA GAAGGAGGAA CATCATTAGA GGCGT CACCACTGA GACAATATE CECCGGAGA GAGAACATT TOCCAACAT GCCACATAGA CACGGAGATGA CACGGGAGAAC TACGGCGAGA GGCT CCGCCGCTGG AAACCATGGAT GAGAGACATT TOCCAACAT GCCACATAGA CACGGACAAAA CACGGAGAAAT GATCAATCATTA TGATGCATGC ATGAAATACA AAGAAGATA TOCGGCTTG TGCGATATA CAGGGAAAAACA GACGGATAAGA AGTCATCGAA GTCAATGTG ATGAAAGAACAGT TOCGCCTOT GACGTTAGTAA CATCATCATTA TGATGAAGA AGTCATCGAA GACAATTG CAAAGAAGAA CTITA GGGGATAAGA AGTCATCGAA GTAAATGATA TACAAGACG TOCGCCTOT GACTTAGAA CATCATCATTA TGAAAGAGT TGAAAACAGT TOCAACAGT TOCGCCTOT GACCTTGTGA CATGGAGAC AATCAATGGA GACGGCAATG TAACAACTTT TGAAGCAGTC CTITA GGCGATAGA AGTCATCGAA GTCAATGTA ACGAGAGAC TOCGCCCTOT GACCTTGTGA CATGATGAA GACGGCAATG TAACAACTTT TGAAGCAGTC CTITA GAGGATAAGA AGTCATCGAC ATGGCGGCAT CCTTCAAATG GGCTTCGTG AAAAATGGA CAGTCCTGA <i>Illus subtilis</i> 168 BU_1800 citb <i>KTAAK DUFQAKTTT</i> TGAKTHYYS LKALEDSGIG VSKLPYSIK VLLESVIRQV DGFVIKKEY ENLAKKCTAE LKDIDVPFKF SRVILQDFG VEAVVDLAEL RKAMANGGD PDKINPEIPY DLUIDHSVQV DAGCTEDALA VNHOLEFERN <i>PLSMA KKANNYQAY PPATGIVHQY NLEFLASVYH AIEEOGGLIY YPDTINCTO</i> HALANCTAE LKDIDVPFKF SRVILQDFG VEAVVDLAEL RKAMANGGD PDKINPEIPY DLUIDHSVQV DAGCTEDALA VNHOLEFERN <i>PLSMA KKANNYQAY PPATGIVHQY NLEFLASVYH NLEFLASVYH NLEGGGUYY VULDSGI KUGGYGIFA AGGAGGAGAG FYKNGUSAAG GGGTAGAE ENELKKIL KGETVKKG AAAAATGG TITSNPYVLI GAGUAKAAY ELGLAVENYH <i>PLSMA KKANNYQAY PPATGIVHQY NLEFLASVYH NLEF</i></i>		
NGARC MAGNETERAR TAAANGGGAT AACGATECAA CETACATACA AMACCCACCA TECTETGAA GAGATGEGE TGAGCCAGGE AACGTEGGE CATEMAAAGG ACTGCGETE TAAAT TCGGGGATE GATCACTAGE GACCATATE CECCTCGGG GGCATCGGA AMAGATAGE CECCGGGAT TATTICEAA GAAMAAGGT TETCACCTE TGACTETTAAC GGGCT CCGGCGTCG AACCATGAE GACGATATE TCCCTCGGG GGCATCGGA AMAGATAGE CECCGGGAT CATATTAGE GAAGAAGGA CEGGCCGAE AGTAA CATCATETA TGATGATGE ATGAAAAGAE TACCAGTET GCCGATETA CAGGAAAGA CATGEGGTATE CAGGAAAGAE TATGCGATAGE CACCGGATAG GCTC CCGGCTACAG GTAGATAGE CAAGAAGAET TCCAACAAGE AATCGGATAGE CACCGGATAG AACGAAATGE CACCGGATAG TACGAATGAE AGCGGAAAGA CETTA CGGGTAAAGA AGTCATCGAE GTAGATGETE ATGAAAGAE TCCAACGGG GAATGETTAGE CATGGGGTG CCACCGGATG TAACAACTET TGGAAGGAC CETTE CGGCGAACA AGTCATCGE GTAGATGETE ATGAAAGAE TCCATCCGE GACCTEGGA CTCTAGGAA GACGACAGAA GACGCGAAAG AGCGGAAGG TAACAACTET TGGAAGGAC CETTE GGGGAAGA AGTCATCGE GTAGATGETE ATGAAAGAE TCCATCCGE GACCTEGGA CTCTGGAA CCAATGAA GACGGCAATG TAACAACTET TGGAAGGC CETTE GGGCGAAGT TGAAAATGET TACTAACCGGC GGCAT CCTTCAAATG GGCGTCTGGA AAAAATGAA GCAGTCCTGA		
ENANT FUGGGGATTE GETACEARET GECCEREGGG GECARTCEGG ADAGETEGG CTGECGGAN ATTERTECHA GANANAGETE TTECHCTEG GEACTTEAC SEGET CEGEGGETE GANACEMENT GECARATE TECCARACTE TECCARACTE CEGARAGET ATTECCA CEGEGGATE TECHCEGG TETCHCETCE GEACTERAC SEGET CEGEGARET GANAGENERATE ANGAMARTA ANCOMATE CEGARAMA ACCMATECE ACCEGGARA TACCAGET CETTE TEGGEARATE GANAGETET GEAMAGET TEGANAGAT TECCARAMA ACCMATECE ACCEGGART TACCAGEGT TEACGACETA CETTE TEGGEARATE GANAGETATE GEAMAGEAT TECARAMAN ACCMATECE ACCEGGART TACCAGEGT TEACAACTE TECARACEA CETTE TEGGEARATE GANAGETTE GEAMAGEAT TECARAMAN ACCMATETTE TEATEGGET GEOTOCOCETE CASTITAAC ANGOTGANA TEGGARACA TITA CEGGEARAGE GEARATETET AGAAACTE TEGGARAGAT TEACAGAMA ACCATETTATA CANGOTAMA AGOTGANAA TEGGARACA TITA CEGGEARAGE GEARATETET AGAAACTE TEGTETTET GEOTOTTET GEATAGAGE ANTENTAGE GARCETTATE GAACACTE TAACAACTE TITA CEGGEARAGE TEGARATEGA TACTACEGCE ATEGECEGECA CETTERA TEATAGAACA ANTENAGA GACGATE TAACAACTET TEGAAGCAET TITA CEGGEARATE TEGAARTEGA TACTACEGCE ATEGECEGCA CETTERA TEATAGAACA ANTENAGA GACGATETA LILUS SUBFLIIS 168 BSU_18000 citB ETAM DYGARKETET TEGARTETATE TEATARTEGA TEATACEGCE ATEGECEGCE TEATAGETAGA CANTERAGE AKENGUGAE EKENGUGAE TEATAGETE TEATACEACE LILUS AKENTENGAN DEGEN TEATAGETATE TEATATEGAT TEATACEGCE ATEGECEGCET CETTERA TEATAGETA EKENGUGAES TEATEGAT TACAACTET TEAACTACE LIRUS AKATENYA PANTINGY NIELEDGEITY DETUNCTO HETTINICIGU LEGEVERICE ALMINGTAE LEXIDUPFKE SEVILOPPKE SEVILOPPKE SEVILOPPKE AVENALANVEGD PUNINEEDIALA VINDLEFEEN LIRUS AKATENYA PANTINGY NIELEDGEITY DETUNCTO HETTINICIGU LEGEVERICE TEAVANNOL LEGEVERED LIRUS AKENTE ANALAYOG DIALATISE TENENYULI GAGUAKAA ELGLAVPHY DEGEN VICULUNGEL LEPHAECHT DUCCTECI GISGELESEI ENANNE TEAVANNOL TEGRARDET ATUGALATUTE LERAMANAGE AUGUETAE ENANNEL LIRUS AKATENTER PERILEVER VICUSTO VICUSTORI TIGAACTET TEANACACET VICUNCAGA LEGILVARAY ELGLAVPHY GEEN VICULUNGEL LIPHAECHT DUCCTECI GISGELESEI ELAVANNOL ITSULGARET MILGOPARTE VICUNCAGATI ANALATICE TENENTUU LEGILVARATEL LIRUK BUNTUR DESTINGTA VICUSTORI VIGTEGOSTI DINFORMETE ANALANGET VICUNCAGESENE TEANANALI TISUSAGUR		
AGTAA CATCCATCTA TGATGGATGG ATGAAATACA AAGAAGATAA AACCGGTGTT GTGGTATTAG CAGGAAAAGA CTATGGTATG GGATCTGGG CTGGCACTGGGG TGGCAAAAGGA CCTTC TGGGATGGA ACCGCTCAT GGCAGGAGCT TGGAAGAAGA ATCTGGTAT TGATGGGATG GGTGCGGTT GAGTTGAAGAC AAGCAGTAG AAGGAGAATG TAGAAGATACA CTTTA GGGGATAGA AFCCATGGAG GTAGATGTG TGGCAGGAGC ATGCATGTAT TGATGAAGCA AATCATGGTA GGGGGGATG TAACAACTTT TGAAGACAGT CTTTG ATAGGGAAGT TGAAATGAT TACTACCGCC ATGGCGGGAT CCTTCAAATG GTGGTTGGTG AAAAAATGAA GCAGGCAATG TAACAACTTT TGAAGACAGT CTTTG ATAGGGAAGT TGAAATGAT TACTACCGCC ATGGCGGGAT CCTTCAAATG GTGGTTGGTG AAAAAATGAA GCAGTCCGA		
ICTIC TOGGGTCAG AACGGTCAT GCAGAGGGT TOGAAAGAAT TOACAGAAG AATOTTGTAT TOATGGGTT GOTGCCGCTT CASTTAAAC AAGGTGAAAA TOGGGATAGA		
CTTMA GEGGGTAMAGA AGTCATCGAG GTMGARGTTA ATGAMACAGT TOGTCCTOGT GACCTTGTGA CTGTAMAGAG AATGAATGAA GACGGCAATG TAACAACTTT TGAMGCAGTC CTTTG ATAGTGAMGT TGAMATTGAT TACTACCGCC ATGGCGGCAT CCTTCAMATG GTGCTTCGTG AMAMATGAA GCAGTCCTGA illus subtilis 168 BEU_18000 citB KTAMA DVGQARKTET TMGKTYHYYS LKALEDSGIG KVSKLPYSIK VLLESVLRQV DGFVIKKEHV ENLAKWCTAE LKDIDVPFKP SRVILQDFTG VPAVVDLASL RKAMANVGGD PDKINPEIPV DLVIDHSVQV DKAGTEDALA VNHDLEFERN PLSMA KKAFNNYQAV PPATGIVHQV NLEFLASVVH AIEEOGGLIV YPDTLVGTDS HTTMINGIGV LGWGVGGIEA EAGHLGQPSY FPVEVJGAK LVGKLPMGTT ATDLALKVTQ VLREKGVVGK FVEFFGFGIA ELPLADRATI ANNAPEYGAT DDEAL ANLLAIGND FEHIDVVEAN CENSIGLFTP DAEDPQFTDV VEILDSGIEA NLSGFRRPQD LIFLSAMQET FKKQLVSFAG NGGFGLMAE EMERIKFKLL NGEETVAKTG ALALAINISG TVTSNPYVLI GAGLVAKAV ELGLKVPNY DGEKN VICYLVNSGL LFYMKELGEN LVGYGCTTCI GNGGUSJEPE EEAVKKNDLI ITSVLGGMRN FEGHIEFUK GVALSFPEV VALLAGTVN INLKTDPIGV GKOGVNYPN DINFSNDEIM ALVGYDTVE LFKKYETVP DONKMANEIE DUKNN DMSTVINDP FFEENDEGK VERFLAKLV VGKFDDSYT DISFPAGKI G KUSPERDFN SYGSRKOME GVALSFPEV VALLAGTVN INLKTDPIGV GKOGVNYPN DINFSNDEIM ALVGYDTVE LFKKXENTOV DONKMANEIE		
illus subfilis 168 BSU_18000 citB KTAAK DVEQARKTFT TNKKYTHYS LKALEDSGIG KVSKLPYSIK VLLESVLRQV DGFVIKKEHV ENLAKWGTAE LKDIDVPFKP SRVILODFTG VPAVVDLASL RKAMAAVGGD PDKINPEIPV DLVIDHSVQV DKAGTEDALA VNNDLEFERN LIGAR KKAMPANYOV NEPKTGVMVO HLEFLASVH ALEEDGELVT YPDTLVGTDS HTTMINGIGV LGKGVGGTA EAGHLGOPSY FPVEVIGAK LVGKLPKGT ATDLALKVTQ VLAEKGVVGK FVEFEGGGTA ELFLADRATI ANNAPEYGAT JDEEL LNYIRLTGRD PEHIDVVEAY CRSNGLFYTP DAEDPQFTDV VEIDLSQIEA NLSGPKRPQD LIPLSAMQET FKKQLVSPAG NQGFGLNAEE ENKEKFKLL NGEETVHKTG AIAIAAITSC TNTSNPYVLI GAGLVAKKAV ELGLKVPNYV GGKV TVGVLNNSGL LPYMKELGFN LVGYGCTFCI GNSGPLEPET ERAVARNDLI ITSVLSGNRM FEGRIFELKU GNILASPFLV VAXLAGTVN INLKTOPIGV GKOGONYTPN DIWPSMDEIN ALVKGYTPE LERNETETVF DDNKRMNEIE LYNGN DDSTYLDPF FYEENSVEG KVETKLGLVU VGKFGSVT DHISPAGKIC AGVSGPADEN SYGSRROME WYMGGTPAN TIKKQILAGT EGFTYMFT GETTVADA WYKYENGTU JVGKGGSNTEVGA GSSRDWAAKG		
TAAK DYEARKTET TIKKTYETYE LKALEBGGIG KVEKLEYSIK VLLESVLAQU DGYLIKKENY EMLAKKGAE LKAUGUSTE SAVLADETG VAAVUDASL EKAMAAVGGD POKIPEIPY DUIDIBGVQU DKAGTEDALA VINDLEFERN IJSAK KKAFNNYQAY PANGIVHQY MLEFLASVYH AIEEDGELTY YPDITUGTUS ETTHINIGU LGKVGGIFA EAGHLGQFSY FYMEYIGKAL LVALHAGTY ATLALKYTQ ULEKGVUCK FVEFFGPGIA ELFLARATI ANMAPEKGAT DEEL LNILLIGGD PEHIDVVEAY CRSNGLFIFD DAEDGYTTYV VSIDLSQIEN HISGFKRPQD LIPLGANQET FKGUISPAG NQGFGLARE ENKEIFKLI NGETVNKTG ALALAITSC THISFYDIL GAGUXAKAN ELGLKYPHY GGEKV TVGLINNSGL LFYMMELGFN LVGVCTFGI GNSGPLEPET ERAVARNULL ITSULGGNRT FEGRIFHLIK GNILASPFLV VAYLAGTVN INLKTDFIGU GKGQNYTFN DIMPEMDEIN ALVGYTTPE LFMRETETYD DDNKRMMEIE JYNK NDSTITLANDE FYEENSUSEG KVEFLKLURV VGKFGGVTT DHISFAGAIG KONGSPANT RYSGRROMBE WANGTPANT INLKTDFIGU GKGQDYTFN DIMPEMDEIN ALVGYGTTEL GNGVGKGGGSKV GKGGGGAAGG MAGTANA DAEDGYGGGANT DHISFAGAIG KOTGANT DAGTSFRACH FYGGIFSANT SYGSRROMBE WANGTPANT INLKTDFIGU GKGGGNYTFN DIMPEMDEIN ALVGYGTTE LFMRETETYD DDNKRMMEIE	TTTE ATAGTGAAGT TGAAATTGAT TACTACCGCC ATGGCGGCAT CUTTCAAATG GTGCTTCGTG AAAAAATGAA GCAGTCCTGA	
KRAAK DVPQARKTFT TRKKTYHYYS LKALEDSGIG KVSKLEPYSIK VLESVLAGV DGFVIKKENV ENLANKORAE LKDIDVPFKF SAVILOPTG VPAVVDLASL KRAMANGGD POKINDELPV DLVDHSVOV DKACTEDALA VNHOLFERN FLSMA KKAPINYGAV PPARGIVNGV NLEFLASVVH AIEBOGLIVY VEIDLSGIEM HUGBYGGIEM ALGORGGIEM AEMGLOGESY FPVEPUIGAK LUGKLEPKGT ATDALALKYV ULEKKUVGK FVEFFGGIA ELPLADRAFI ANMAPEYGAT VDEEM LNILIGRAD PEHIDVEAY CESNGLETIP DAEDQOFTOV VEIDLSGIEM NLSGPKRPQD LIPLSANGET FKKQUSSFAG NGGFGLAMEE ENRIKFKLL NGEETVAKTG AIATAITSC THYSNPYVLI GGLVARKAV ELGLAVFNY DEGKV VYGYLVNSGL LEPYMKELGEN LUGYCGTTCI GNGGLSPEI EEAVAKNDLI ITSVLGGNRNF FEGHIFELVK GNLASPFDI VAXLAGTVM INLKTDPIGV GKOGNYVYN DLWPSMDEIM ALVKQTTVE LEFKRYETVF DDNRMMEHE LINKON DDSTVLNDF FFERMEDEG KVEPLKGLVY GVGFDGSYTT DDISPANGIG KUSSFARDIG SVGSFAGNE ENGLEFVENKE GVMGGTFANT, EIGYLTADTE GVENTSLICAK KUKKAVGUG GSSBADAKG		
FLSWA KKAFNNYQAV PPATGIVHQV NLEFLASVVH AIEEDGELVT YPDTLVGTDS HTTMINGIGV LGWGVGGIEA EAGHLGQPSY FPVPEVIGAK LVGKLPNGTT ATDLALKVTQ VLREKGVVGK FVEFFGPGIA ELPLADRATI ANNAPEYGAT VDEBA LNYLRLGGD PEHLDVVEAY CRSNGLFYTP DAEDGYTDV VEIDLSGIEA NLSGPRRYDD LIFLSANQET KRKQLVSPAG NQGGGLNAEE ENKELKKLL NGEETVMKTG AIAIAAITSC TNTSNPYULI GAGLVAKKAV ELGLKVTMYV PGSKV VTGYLVNSGL LPYMKELGFN LVGYGCTTCI GNSGPLSPEI EEAVAKNDLL ITSVLSGNRN FEGRIHPLVK GNYLASPPLV VAYALAGTVN INLKTDPIGV GKOGQNVYFN DIMPSMDEIN ALVKQTVTPE LFRKEYETVF DDNKRMNEE FLSMD NDSTYLONPF FFERNSVEFG KVEPLKGLRV VGKFGDSYTD BISPAGAIG KDTPAGKILG EKGVSFPDFN SYGSRRGNE VMRGFFANI RIKNQIAFGT EGGFTTWFF GEVTSIYDAG MKYKEDKTGL VVLAGKDYGM GSKDMAAKG		
VDERA LNYLELTGRD PERIDVVEAY CRSNLFYTP DAEDOPCTDV VEIDLSQIEA NLSGERKEDOL LIPLSANGET FEKGLYSENG NQGFGLANEE ENKEIKFIL KGEETVEKTG ALALTSC THYSNFYLI GAGLVAKKAV ELGLKVPHYV PGSKV VTGYLVNSGL LPYMKELGFN LVGYGCTTCI GNSGELSPEI EEAVANDLI ITSVLSGNRN FEGRIHPLVG GYLLASPPLV VAIALGTVN INLKTDPIGV GKDGGNVYFN DIWPSMDEIN ALVKQTVTPE LFRKEVETVP DDNKRMNEIE JUNN DNSTYLTQNPP FFERSVERGE KVEPLKGLAV VGKTGGSVT DHISPAGAIG KOTPACKL) EKKUSPADT SVGSRRGNME VMRGTPANI TIKKQILAGT GGGFTYMFT GGTFISDAC MKYKENGTA VGKTGGSVT DDNKAMAKG		
LYKWD NDSTYIQNPP FFEEMSVEPG KVEPLKGLRV VGKFGDSVTT DHISPAGAIG KDTPAGKYLQ EKGVSPRDFN SYGSRRGNHE VMMRGTFANI RIKNQIAPGT EGGFTTYWPT GEVTSIYDAC MKYKEDKTGL VVLAGKDYGM GSSRDWAAKG		
	LYMW DUSTYIQNP FFEENSVERG KVERLAGLAV VGKFOSSVTT DHISPAGAIG KDTPAGKILG EKGYSPENDT NSGSRRGMHE VMMRGTANI RIKNGIAPOT EGGFTYMPT GEVTSIYDAG MKYKEDKTGL VVLAGKDYGH GSSRDMAAKG Intyl assfrintes Nlympolicil gyrggenand Lgichckryle Udvervand Gyntyterav Vredsbeveld Yyrnggildg Vlaebangg	

Figure 2. Genome Browser focusing on the citB gene. Tabs at the top allow to search for specific genes, including flanking regions, or simply to query by region. The main frame includes an interactive scrollable genome, where all elements are clickable for a redirect to the specific gene page. At the bottom, both protein and DNA sequences are shown (on the left) with the possibility to show the reverse complement or to search for specific patterns within a sequence (on the right).

protein interactions, regulatory elements and functional annotation is also available. Each gene/protein is assigned to one or more functional categories organized in a tree-like structure and to regulons. The pages then provide detailed information on the gene, the protein and the regulation of its expression. To complement the data, biological materials available in the community, laboratories that study the gene or proteins, and publications are presented in a categorized way for easy reading and organization.

SubtiAPPS

For a better representation of information, different types of data are presented under specific browsers in *SubtiApps*. The Expression Browser fully explores expression patterns under 121 experimental conditions, while the Genomic Browser is an interactive map where one can scroll through the *B. subtilis* genome in an intuitive and easy manner. In addition to genes' coordinates, their respective DNA and protein sequences are loaded to provide a complete view over the genome (Figure 2).

One of the highlights of the previous version was the biological network visualizer for interactions and regulation (10). Currently, we have expanded the data to >5950 and >2660 experimentally validated regulatory and protein-protein interactions, respectively. Here, we offer a clear representation of the documented physical interactions with the possibility to expand the neighborhood (Figure 3A). Focusing on the regulatory data, regulatory networks can be interrogated for 2271 documented operons. There are three different types of annotation depending on the mechanism of regulation: green for positive regulation (or activa-

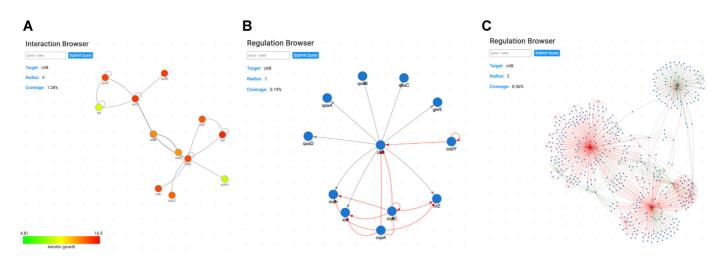


Figure 3. The Interaction and Regulation Browsers. (A) Interaction Browser for the *citB* gene. Nodes represent proteins and edges represent physical interactions, commonly referred to as protein–protein interactions. Using the settings, transcriptomic and proteomic data can be integrated using a color code (this feature is also available for the Regulation Browser). Upon selection of an experimental condition to be visualized, each node will be masked with a color depending on the expression value of the gene in the selected dataset. Green represents lower values of expression and red represents higher values of expression. (B) Regulation Browser for the *citB* gene. Nodes represent genes and edges represent regulatory interactions. The interaction type is represented with a color scheme: red for repression, green for activation and gray for sigma factor. (C) Regulation Browser for the *citB* gene with an increased radius of distance of 2. This setting allows to expand all regulations from the initial representation (B) with the distance of 2 elements [thus including the regulatory interactions of the primary partners that are shown in panel (B)], creating a more complex view over the regulatory network of *citB*.

tion), red for negative regulation (or inhibition) and gray for sigma factor interactions (Figure 3B). Similar to the Interaction Browser, the Regulation Browser also allows to expand the view to the neighborhood of the target gene, giving the user a view over the full known regulatory network of the gene of interest, but also of its neighbors, even allowing the visualization of regulatory subnetworks (Figure 3C). The gravity model (12) in which nodes (genes/proteins) are considered as mass points and edges (physical or regulatory interactions) as springs was preserved for both browsers and the user can find a clear and intuitive integration of the expression data within networks (Figure 3A). Notably, all references of such data can be found at the bottom of each respective gene page in the References section.

PROTEIN HOMOLOGY INTEGRATION

According to the SubtiWiki data collection, there are 1819 biological elements with unknown function or lacking annotation, which represents $\sim 30\%$ of all elements. To mitigate this lack of annotation, we decided to integrate information on potential protein homologies for all B. subtilis proteins. For this, we created a proteome library of 16 relevant bacteria, some close relatives of B. subtilis as well as important representatives of other bacterial groups such as the proteobacteria or cyanobacteria (13) (Figure 4A) by extracting the corresponding sequences from UniProt (14). To assess potential homologs, we developed a pipeline using the well-known sequence similarity search tool FASTA (15). The pipeline is composed of two rounds of alignments. The first one compares every protein of *B. subtilis* with the library, and the second one compares the best hits found in the previous step with the *B. subtilis* proteome. If the resulting protein from the second alignment is the same as the initial input of the first step, then the proteins are considered bidirectional best hits (Figure 4B). Although performing this second round of alignments gives more confidence in the prediction of homologs, it was still necessary to consider the difference of protein sizes when running alignments. Disregarding this step may cause false positives, as small proteins can have a high identity score for multiple regions of larger proteins. To avoid this, we paid special attention to the alignment of smaller proteins with larger ones, and manually excluded cases where strong inconsistencies were found. All resulting pairs showing an *E*-value ≤ 0.01 with a minimum identity of 40% were considered. However, lower values of identity were also considered as long as high values of similarity were retained.

Currently, *Subti*Wiki provides protein homology results for every protein, with the exception of 254 proteins that did not meet the identity/similarity and/or *E*-value cutoff criteria for any of the representative relatives. The results can be accessed on the gene page and are presented in an easy-to-read table (Figure 5). For each queried *B. subtilis* protein, there is a list of organisms and their respective best hit protein homologs. Each of these proteins is linked to its respective UniProt page and the metrics for identity, similarity and bidirectional best hit are shown. If no protein homolog for a certain organism was found, then this is stated (Figure 5).

Finally, to complement the new protein homology section, the COG database is now fully integrated in *Subti*Wiki. COG is a tool for comparative and functional genomics of prokaryotes and identifies orthologs in a representative set of different bacteria and archaea (11). By taking advantage of the optimized COG database and fully integrating their IDs, *Subti*Wiki now provides a strong complementary section of homologs and orthologs (Figure 6).

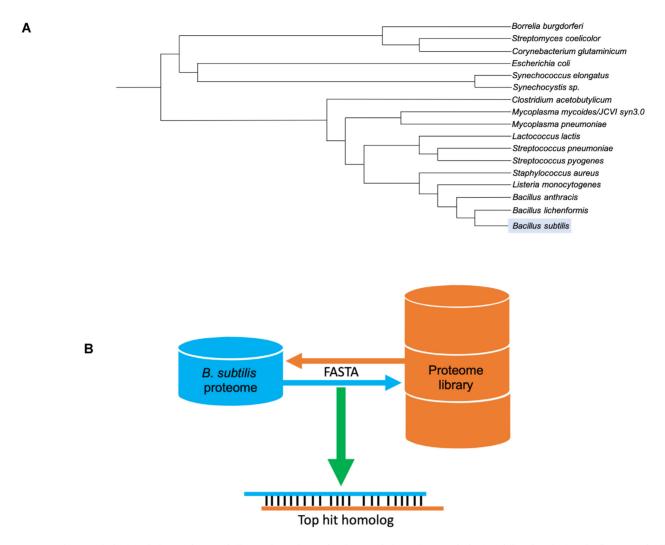


Figure 4. Homology analysis. (A) Phylogenetic tree of all organisms chosen for the protein homology analysis. (B) Bidirectional FASTA alignment pipeline representation.

NEW DATA INTEGRATION

A hallmark of this *Subti*Wiki update is the integration of new datasets. Recently, we have added a compendium of genome minimized strains derived from *B. subtilis* 168. Some of these strains have already been proven superior for biotechnological applications such as the production and secretion of difficult proteins and lantibiotics (16,17). The MiniBacillus Compendium section (http://www.subtiwiki. uni-goettingen.de/v4/minibacillus) (18) contains a table listing the name of each strain, the respective download link of Geneious and GenBank files, genomic details (genome size, deletion steps and percentage of genome reduction) and a list of publications associated with the respective strain.

Furthermore, we have added a large set of protein expression information. This information is based on a large-scale study that sought to compare proteomic responses to 91 different antibiotics and comparator compounds in an attempt to elucidate antibacterial modes of action (19). To supplement this representation, we also provide the possibility to download the entire dataset under the new Downloads section (http://www.subtiwiki.uni-goettingen.de/v4/ download). In addition to the previously implemented expression data, the new dataset can be accessed in the Expression Browser. Here, the user has full access to the available data for each biological element in a specific representation and it is possible to overlay the expression data for each protein in the Interaction Browser.

The latest updates reflect intensive research to identify novel transporters in *B. subtilis*. These include transporters for bicarbonate (NdhF-YbcC) (20,21) and several amino acids, including alanine (AlaP), glutamate and serine (both AimA), and the toxic analogon glyphosate (GltT) (22–25).

In addition, two new categories of genes/proteins have been included. These are the quasi-essential genes and genes that encode proteins that are required for the detoxification of toxic metabolites. Quasi-essential genes are those genes whose inactivation results in the immediate acquisition of suppressor mutations. Interestingly, essentiality of such quasi-essential genes has so far been controversial (26–28). The new concept of quasi-essentiality supports the idea that often there is no clear binary answer in this issue but rather a continuum from clearly dispensable genes, genes that prove

Potential protein homologs

Description

The following list presents the potential protein homologs for CitB using FASTA alignment tool. For all top hits, a bidirectional analysis was performed and those showing the same pair of proteins are considered to be potential orthologs. The column *Best hit bidirectional* informs if CitB was found to be best hit in the alignment with the protein of the respective organism.

CitB

Organism	Protein name	Identity	Similarity	Bidirectional best hit
Bacillus licheniformis	CitB	90.9%	97.7%	Yes
Bacillus anthracis	AcnA	78.7%	92.5%	Yes
Listeria monocytogenes	CitB	73.3%	91.5%	Yes
Staphylococcus aureus	SAOUHSC_01347	72.0%	90.4%	Yes
Lactococcus lactis	CitB	54.9%	80.1%	Yes
Streptococcus pneumoniae		No significant homolog	No significant homolog	
Streptococcus pyogenes		No significant homolog	No significant homolog	
Clostridium acetobutylicum	CitB	25.7%	48.7%	Yes
Mycoplasma pneumoniae		No significant homolog	No significant homolog	
JCVI-syn3A		No significant homolog	No significant homolog	
Corynebacterium glutamicum	Acn	50.5%	75.5%	Yes
Streptomyces coelicolor	CAC37548	55.1%	79.4%	Yes
Escherichia coli	AcnA	56.6%	80.8%	Yes
Synechocystis sp.	LeuC	27.5%	52.4%	No
Synechococcus elongatus	LeuC	27.1%	51.6%	No
Borrelia burgdorferi		No significant homolog	No significant homolog	

Figure 5. Protein homology page for the CitB protein. At the top is a brief description of the analysis, containing details on the FASTA alignment tool and on the table. The table is composed of five columns: Organism, lists all 16 organisms used; Protein name, best hit protein found in the respective organism with a hyperlink to UniProt; Identity and Similarity, scores from FASTA alignment tool; and Bidirectional best hit (yes/no), result of the bidirectional alignment.

essential only after long-term cultivation, genes that require genetic suppression upon deletion to those that cannot be deleted under any circumstances. As an example, the *topA* gene encoding DNA topoisomerase I is responsible for the relaxation of negatively supercoiled DNA behind the RNA polymerase. Upon deletion of *topA*, the bacteria immediately acquire mutations resulting in an increased expression of DNA topoisomerase IV that can then take over the role of TopA (29). Similarly, deletion of the *rny* gene encoding RNase Y results in the acquisition of suppressor mutations that result in a globally reduced level of transcription activity (30) and deletion of the unknown gene yqhY triggers mutations that often affect conserved residues of acetyl-CoA synthetase indicating that YqhY prevents a potentially toxic accumulation of malonyl-CoA or fatty acids (31). Proteins involved in detoxification reactions have recently been rec-

Categories containing this gene/protein

- Metabolism, Carbon metabolism, Carbon core metabolism, TCA cycle
- · Metabolism, Amino acid/ nitrogen metabolism, Utilization of amino acids, Utilization of branched-chain amino acids
- Information processing, Regulation of gene expression, Trigger enzyme, Trigger enzyme that acts by binding of a specific RNA element
- Information processing, Regulation of gene expression, RNA binding regulators

List of homologs in different organisms, belongs to COG1048

This gene is a member of the following regulons

• SigA regulon, CcpA regulon, CodY regulon, CcpC regulon, CitB regulon

Figure 6. Homology and COG database implementation in *Subti*Wiki on the *citB* gene page. The section highlighted in blue is shown when a corresponding COG entry is available for the gene and redirects to the respective COG database page.

ognized as very important for cellular metabolism as they remove harmful by-products or intermediates. Examples for such toxic by-products are 4-phosphoerythronate, a byproduct of erythrose-4-phosphate oxidation in the pentose phosphate pathway that inhibits the phosphogluconate dehydrogenase GndA, and 5-oxoproline, an unavoidable damage product formed spontaneously from glutamine. These toxic metabolites are disposed of by the CpgA GTPase that moonlights in dephosphorylation of 4-phosphoerythronate (32) and by the PxpABC 5-oxoprolinase (33), respectively. It has recently become obvious that these detoxification mechanisms are very important for the viability of any living cell. The limited knowledge on these mechanisms is an important bottleneck in all genome reduction projects (34,35). Thus, both these new categories are very important to get a more comprehensive knowledge about the minimal requirements to sustain bacterial life.

FUTURE PERSPECTIVES

With millions of accesses on a yearly basis, *Subti*Wiki has become one of the most popular database that provides upto-date and curated data for the model organism *B. subtilis*. Although this bacterium is one of the best-studied organisms, there is still a vast portion of data to be annotated and a lot to be discovered. By providing daily data and feature updates to the community, we expect *Subti*Wiki to continue to grow. We hope for *Subti*Wiki to expand in viewership and further establish itself as a prominent tool helping researchers develop new hypotheses. Future aims include the addition of new basis entities such as metabolites and corresponding new types of datasets such as protein–metabolite, protein–RNA and RNA–metabolite interactions that will consolidate the role of *Subti*Wiki as a trendsetter for model organism databases.

ACKNOWLEDGEMENTS

We are grateful to all lab members for testing the new features and for valuable feedback.

FUNDING

No funding agency. *Conflict of interest statement.* None declared.

REFERENCES

- O'Reilly,F.J., Xue,L., Graziadei,A., Sinn,L., Lenz,S., Tegunov,D., Blötz,C., Singh,N., Hagen,W.J.H., Cramer,P. *et al.* (2020) In-cell architecture of an actively transcribing–translating expressome. *Science*, **369**, 554–557.
- Link, H., Kochanowski, K. and Sauer, U. (2013) Systematic identification of allosteric protein–metabolite interactions that control enzyme activity *in vivo*. *Nat. Biotechnol.*, **31**, 357–361.
- 3. Gerovac, M., Vogel, J. and Smirnov, A. (2021) The world of stable ribonucleoprotein complexes and its mapping with Grad-seq and related approaches. *Front. Mol. Biosci.*, **8**, 661448.
- 4. Dietrich, M.R., Ankeny, R.A. and Chen, P.M. (2014) Publication trends in model organism research. *Genetics*, **198**, 787–794.
- 5. Oliver, S.G., Lock, A., Harris, M.A., Nurse, P. and Wood, V. (2016) Model organism databases: essential resources that need the support of both funders and users. *BMC Biol.*, **14**, 49.
- Errington, J. and van der Aart, L.T. (2020) Microbe profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse. *Microbiology*, **166**, 425–427.
- 7. Kovács, Á.T. (2019) Bacillus subtilis. Trends Microbiol., 27, 724–725.
- Michna, R., Commichau, F.M., Tödter, D., Zschiedrich, C.P. and Stülke, J. (2014) *SubtiWiki*—a database for the model organism *Bacillus subtilis* that links pathway, interaction and expression information. *Nucleic Acids Res.*, 42, D692–D698.
- 9. Michna, R., Zhu, B., Mäder, U. and Stülke, J. (2016) *Subti*Wiki 2.0—an integrated database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.*, 44, D654–D662.
- Zhu, B. and Stülke, J. (2018) SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. Nucleic Acids Res., 46, D743–D748.
- Galperin, M.Y., Wolf, Y.I., Makarova, K.S., Vera Alvarez, R., Landsman, D. and Koonin, E.V. (2021) COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.*, 49, D274–D281.
- Fruchterman, T.M.J. and Reingold, E.M. (1991) Graph drawing by force-directed placement. *Software Pract. Exper.*, 21, 1129–1164.
- Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.*, 49, W293–W296.
- UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res., 49, D480–D489.
- Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D. et al. (2019) The

EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.

- Aguilar Suárez, R., Stülke, J. and van Dijl, J.M. (2019) Less is more: towards a genome-reduced *Bacillus* cell factory for 'difficult proteins'. *ACS Synth. Biol.*, 8, 99–108.
- van Tilburg,A.Y., van Heel,A.J., Stülke,J., de Kok,N.A.W., Rueff,A.S. and Kuipers,O.P. (2020) MiniBacillus PG10 as a convenient and effective production host for lantibiotics. *ACS Synth. Biol.*, 9, 1833–1842.
- Michalik, S., Reder, A., Richts, B., Faßhauer, P., Mäder, U., Pedreira, T., Poehlein, A., van Heel, A., van Tilburg, A.Y., Altenbuchner, J. et al. (2021) The Bacillus subtilis minimal genome compendium. ACS Synth. Biol., https://doi.org/10.1021/acssynbio.1c00339.
- Senges, C.H.R., Stepanek, J.J., Wenzel, M., Raatschen, N., Ay, Ü., Märtens, Y., Prochnow, P., Vázquez-Hernández, M., Yayci, A., Schubert, B. *et al.* (2021) Comparison of proteomic responses as global approach to antibiotic mechanism of action elucidation. *Antimicrob. Agents Chemother.*, 64, e01373-20.
- 20. Fan,S.H., Ebner,P., Reichert,S., Hertlein,T., Zabel,S., Lankapalli,A.K., Nieselt,K., Ohlsen,K. and Götz,F. (2019) MpsAB is important for *Staphylococcus aureus* virulence and growth at atmospheric CO₂ levels. *Nat. Commun.*, **10**, 3627.
- Fan,S.H., Matsuo,M., Huang,L., Tribelli,P.M. and Götz,F. (2021) The MpsAB bicarbonate transporter is superior to carbonic anhydrase in biofilm-forming bacteria with limited CO₂ diffusion. *Microbiol. Spectr.*, 9, e00305-21.
- Sidiq,K.R., Chow,M.W., Zhao,Z. and Daniel,R.A. (2021) Alanine metabolism in *Bacillus subtilis*. Mol. Microbiol., 115, 739–757.
- Klewing, A., Koo, B.M., Krüger, L., Poehlein, A., Reuß, D., Daniel, R., Gross, C.A. and Stülke, J. (2020) Resistance to serine in *Bacillus subtilis*: identification of the serine transporter YbeC and of a metabolic network that links serine and threonine metabolism. *Environ. Microbiol.*, 22, 3937–3949.
- 24. Krüger, L., Herzberg, C., Rath, H., Pedreira, T., Ischebeck, T., Poehlein, A., Gundlach, J., Daniel, R., Völker, U., Mäder, U. et al. (2021) Essentiality of c-di-AMP in *Bacillus subtilis*: bypassing mutations converge in potassium and glutamate homeostasis. *PLoS Genet.*, **17**, e1009092.
- 25. Wicke, D., Schulz, L.M., Lentes, S., Scholz, P., Poehlein, A., Gibhardt, J., Daniel, R., Ischebeck, T. and Commichau, F.M. (2019)

Identification of the first glyphosate transporter by genomic adaptation. *Environ. Microbiol.*, **21**, 1287–1305.

- Commichau, F.M., Pietack, N. and Stülke, J. (2013) Essential genes in Bacillus subtilis: a re-evaluation after ten years. Mol. Biosyst., 9, 1068–1075.
- Figaro,S., Durand,S., Gilet,L., Cayet,N., Sachse,M. and Condon,C. (2013) *Bacillus subtilis* mutants with knockouts of the genes encoding ribonucleases RNase Y and RNase J1 are viable, with major defects in cell morphology, sporulation, and competence. *J. Bacteriol.*, **195**, 2340–2348.
- Koo,B.M., Kritikos,G., Farelli,J.D., Todor,H., Tong,K., Kimsey,H., Wapinski,I., Galardini,M., Cabal,A., Peters,J.M. *et al.* (2017) Construction and analysis of two genome-scale deletion libraries for *Bacillus subtilis. Cell Syst.*, 4, 291–305.
- Reuß,D.R., Faßhauer,P., Mroch,P.J., Ul-Haq,I., Koo,B.M., Pöhlein,A., Gross,C.A., Daniel,R., Brantl,S. and Stülke,J. (2019) Topoisomerase IV can functionally replace all type 1A topoisomerases in *Bacillus subtilis*. *Nucleic Acids Res.*, 47, 5231–5242.
- Benda,M., Woelfel,S., Faßhauer,P., Gunka,K., Klumpp,S., Poehlein,A., Kálalová,D., Šanderová,H., Daniel,R., Krásný,L. *et al.* (2021) Quasi-essentiality of RNase Y in *Bacillus subtilis* is caused by its critical role in the control of mRNA homeostasis. *Nucleic Acids Res.*, 49, 7088–7102.
- Tödter, D., Gunka, K. and Stülke, J. (2017) The highly conserved Asp23 family protein YqhY plays a role in lipid biosynthesis in *Bacillus subtilis. Front. Microbiol.*, 8, 883.
- Sachla,A. and Helmann,J.D. (2019) A bacterial checkpoint protein for ribosome assembly moonlights as an essential metabolite-proofreading enzyme. *Nat. Commun.*, 10, 1526.
- Niehaus, T.D., Elbawadi-sidhu, M., de Crécy-Lagard, V., Fiehn, O. and Hanson, A.D. (2017) Discovery of a widespread prokaryotic 5-oxoprolinase that was hiding in plain sight. J. Biol. Chem., 292, 16360–16367.
- Reuß,D.R., Commichau,F.M., Gundlach,J., Zhu,B. and Stülke,J. (2016) The blueprint of a minimal cell: *MiniBacillus. Microbiol. Mol. Biol. Rev.*, 80, 955–987.
- Reuß,D.R., Altenbuchner,J., Mäder,U., Rath,H., Ischebeck,T., Sappa,P.K., Thürmer,A., Guérin,C., Nicolas,P., Steil,L. *et al.* (2017) Large-scale reduction of the *Bacillus subtilis* genome: consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.*, 27, 289–299.