Microbiology (1999), **145**, 621–631 Printed in Great Britain

Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis*

Ole A. Økstad, Ida Hegna, Toril Lindbäck, Anne-Lise Rishovd and Anne-Brit Kolstø

Author for correspondence: Anne-Brit Kolstø. Tel: +47 22 95 84 60. Fax: +47 22 69 41 30. e-mail: annebko@biotek.uio.no

Biotechnology Centre of Oslo and School of Pharmacy, University of Oslo, Pb 1125 Blindern, 0316 Oslo, Norway

The opportunistic pathogen *Bacillus cereus* is the genetically stable member of a group of closely related bacteria including the insect pathogen *Bacillus thuringiensis* and the mammalian pathogen *Bacillus anthracis*. Physical maps of *B. cereus* and *B. thuringiensis* strains show considerable variations in discrete parts of the chromosome, suggesting that certain genome regions are more prone to rearrangements. *B. cereus* belongs to the same subgroup of *Bacillus* species as *Bacillus subtilis*, by both phenotypic and rRNA sequence classification. The analysis of 80 kb of genome sequence sampled from different regions of the *B. cereus* ATCC 10987 chromosome is reported. Analysis of the sequence and comparison of the localization of the putative genes with that of *B. subtilis* orthologues show the following: (1) gene organization is not conserved between *B. cereus* and *B. subtilis*; (2) several putative genes are more closely related to genes from other bacteria and archaea than to *B. subtilis*, or may be absent in *B. subtilis* 168; (3) *B. cereus* contains a 155 bp repetitive sequence that is not present in *B. subtilis*. By hybridization, this repeat is present in all *B. cereus* and *B. thuringiensis* strains so far investigated.

Keywords: comparative genome analysis, gene organization, repeated sequence, *Bacillus cereus*, *Bacillus subtilis*

INTRODUCTION

Several studies in comparative genomics using whole genomes have been performed. These have focused mainly on distantly related bacteria (Koonin *et al.*, 1996; Tatusov *et al.*, 1996; Kolstø, 1997; Smith *et al.*, 1997). In addition, the close relatives *Mycoplasma genitalium* and *Mycoplasma pneumoniae* have been analysed (Himmelreich *et al.*, 1997). By sequence comparison, the *Mycoplasma* genomes could be subdivided into six DNA segments which were ordered differently in the two species, suggested to have originated by translocation via homologous recombination. The order of the genes within each segment was conserved, and the increased size of the *M. pneumoniae* chromosome appeared to be due mainly to gene duplications (Himmelreich *et al.*, 1997).

Bacillus cereus is a Gram-positive, rod-shaped, endospore-forming bacterium commonly found in soil, air and water, and is a frequent cause of food-related diseases (Drobniewski, 1993). It is also an opportunistic pathogen involved in local and systemic infections. The pathogenicity of *B. cereus* is caused by its synthesis of several virulence factors, including a putative collagenase, phospholipases, haemolysins, diarrhoeagenic enterotoxins and emetic toxin (Drobniewski, 1993). *Bacillus thuringiensis* is distinguished from *B. cereus* by its production of insecticidal crystal toxins (reviewed by Aronson, 1993). Both species belong to the *B. cereus* group, which also includes the closely related *Bacillus anthracis* and *Bacillus mycoides*. Accumulated data have suggested that *B. cereus* should be considered the parent species of the group, and that the other species should be regarded as subspecies (Smith *et al.*, 1952; Gordon *et al.*, 1973). All members of the *B. cereus* group have been allocated to the *B. subtilis* group of bacilli, by both phenotypic and rRNA classification (Priest, 1993). More than 19 species were assigned to the group, for which *B. subtilis* was chosen as an appropriate representative.

Previously we have constructed physical maps of eight

The EMBL accession numbers for the sequences reported in this paper are specified in Table 1.

B. cereus and four *B. thuringiensis* strains using pulsed-field gel electrophoresis (PFGE) and primarily anonymous fragments of *Hin*dIII-digested chromosomal DNA as hybridization probes (Kolstø *et al.*, 1990; Carlson *et al.*, 1992, 1996a, b; Carlson & Kolstø, 1993, 1994; Lövgren *et al.*, 1998; K. Kvaløy & A.-B. Kolstø, unpublished results). The size of the chromosome was found to display a 2·6-fold size variation, and it was suggested that *B. cereus* and *B. thuringiensis* chromosomes contain one region in which the gene order is more conserved whereas other regions exhibit a more variable gene organization. Sequencing of the anonymous probes used for physical mapping has now enabled us to analyse and map 64 chromosomal genes from *B. cereus* ATCC 10987, and to compare the gene organization to its relative, *Bacillus subtilis* 168, for which the complete genome sequence is known (Kunst *et al.*, 1997). Contradictory to the two complete *Mycoplasma* genomes, the organization of orthologues in *B. cereus* and *B. subtilis* was not conserved, and no common arrangement pattern could be seen.

METHODS

Bacterial strain. *B. cereus* ATCC 10987 was obtained from the American Type Culture Collection, Manassas, VA, USA. The strain is a xylose-positive variant previously used in our laboratory due to its high level of phospholipase C production.

DNA sequencing and analysis. Plasmid DNA from previously prepared randomly selected clones of *Hin*dIII-digested genomic DNA from *B. cereus* ATCC 10987 (Kolstø *et al.*, 1990) was isolated using standard methods (Sambrook *et al.*, 1989) and sequenced by the dideoxynucleotide method (Sanger *et al.*, 1977) using the ThermoSequenase Kit (Amersham) and fluorescein end-labelled oligonucleotide primers on the ALF automated sequencer (Pharmacia). Oligonucleotide primers were prepared at the DNA Synthesis Laboratory, Biotechnology Centre of Oslo. DNA sequences were analysed using the GCG and EGCG sequence analysis programs (Devereux *et al.*, 1984). Sequences were edited using SEQED, ORFs and restriction sites were detected using MAP and translated using TRANSLATE. Translated ORFs were used to search the SWISS-PROT (release 34.0, 11/96), TREMBL (release 3.0, 05/97) and SubtiList (Data Release R14.2, 11/20/97) protein databases by FASTA (Pearson & Lipman, 1988), and by BLAST and GAPPED BLAST (Altschul *et al.*, 1990, 1997) to search the non-redundant database set of protein entries at NCBI, Bethesda, MD, USA. COMPOSITION was used to determine the G+C content of the DNA sequences, and multiple sequence alignments were constructed by PILEUP (Needleman & Wunsch, 1970; Feng & Doolittle, 1987; Higgins & Sharp, 1989).

PFGE, Southern blotting and DNA hybridization. Intact chromosomal DNA was prepared in agarose blocks and PFGE was run using the Bio-Rad CHEF Gene Mapper (Bio-Rad) or the Beckman GeneLine I apparatus (Beckman Instruments), as described by Carlson *et al.* (1992). Southern blotting and DNA hybridization using gene probes was performed as previously described (Carlson *et al.*, 1992). The oligonucleotide probe for *bcr1* (5′ CTC AAA ATT CAG CGG AAG CAA AGA ANT TAG GTG GGG G 3′, where N is A, G, C or T) was prepared at the DNA Synthesis Laboratory, Biotechnology Centre of Oslo. For hybridization, 100 ng of the probe (1 μl) was end-labelled using 2 μl [γ-^{32}P]dATP [5000 Ci ($18{\cdot}5 \times 10^{13}$ Bq) mmol^{-1}; Amersham] and 10 U T4 polynucleotide kinase (New England Biolabs) in 1 × T4 polynucleotide kinase reaction buffer and 10 μl reaction volume at 37 °C for 1 h. Membrane filters were pre-hybridized in 6 × SSC, 10 × Denhardt's solution, 0·2% SDS at 50 °C for 30 min, and hybridized for 1·5 h at room temperature following addition of the end-labelled probe. Filters were washed at room temperature for 30 min, followed by 15 min at 50 °C and 15 min at 62 °C, and exposed and scanned using a Storm 860 Phosphorimager (Molecular Dynamics).

RESULTS AND DISCUSSION

We have partially or completely sequenced 43 of the anonymous probes used for the construction of physical maps of *B. cereus* and *B. thuringiensis*. The probes covered a total sequence length of 80 kb, and 86 open reading frames (ORFs) were identified (Table 1). Seventy-eight ORFs showed similarity (>20% amino acid identity) to submitted entries in the protein sequence databases, and 54 were assigned a putative function by similarity to proteins of known or predicted function. Three ORFs were assigned functions in DNA repair, after complementing the DNA alkylation repair deficient *Escherichia coli* strain BK2118 (*alkC*, *alkD* and *alkE*; I. Alseth and others, unpublished results). Ten ORFs matched proteins of unknown function in the databases, eight of which were from *B. subtilis*. Fourteen ORFs displayed weak sequence similarity, whereas eight ORFs showed no similarity to any database entries. The G+C content was calculated to 35·4 mol%, in agreement with previous predictions from DNA re-association experiments (Priest, 1993).

Isolated genes novel to bacilli

Seven of the 54 putative genes of predicted function in *B. cereus* ATCC 10987 encoded proteins with no sequence homologues in other *Bacillus* spp. (Table 1). These included a type III restriction-modification system (*t3r*; Hegna *et al.*, 1992) resembling systems from the Gram-negative bacteria *Salmonella typhimurium* and *E. coli* phage P1, and one of the DNA alkylation repair enzymes (*alkD*; I. Alseth and others, unpublished results). Moreover, the protein encoded by *cysA* resembled sulfate transporters from cyanobacteria (*Synechococcus* PCC 7942) and eubacteria (*E. coli*) (Green *et al.*, 1989), but showed no similarity to *B. subtilis* proteins. No protein participating in sulfate transport has to our knowledge been identified in *B. subtilis*, although several sulfate metabolism enzymes are found (SubtiList, Data Release R14.2, 1998; www server v2.1.3: http://www.pasteur.fr/Bio/SubtiList.html). The *B. cereus glyS* gene encoded a homologue of class II aminoacyl-tRNA synthetases (Logan *et al.*, 1995). *B. subtilis* and *B. cereus* GlyS were of separate class II subtypes that are not sequence related (SubtiList, Data Release R14.2, 1998).

Recently, a number of bacteria have been identified that are capable of metabolizing atrazine, a triazine ring-containing herbicide that has been found to be less

Table 1. Analysis of 80 kb DNA sequence from *B. cereus* ATCC 10987

Accession no.	Fragment	Length (bp)	Gene homologues	Putative function*	Identity (%)†	Amino acid overlap	Organism/Accession no.
Y11217	agt85	763	*addA*	ATP-dependent nuclease subunit A	50	253/255	*B. subtilis* Sw: P23478
Y09433	bc522	1373	*rs1h*	30S ribosomal protein S1	63	381/381	*B. subtilis* Sw: P38494
Y09253	bc67	1099	*yfbM*	U (similarity to alkaline-phosphatase-like protein)	36	158/203	*B. subtilis* Gb: Z99105
Y09254	bc87	367	*bc87a*	U			
Y11139	bc204	2994	*bc204a*	U			
			bc204b	U			
			yor3	Putative *N*-ethylammeline chlorohydrolase	44	284/312	*Methanobacterium thermoautotrophicum* Gb: AE000910
Y08953	bc41	829	*cmk*	Cytidylate kinase	70	193/199	*B. subtilis* Sw: P38493
Y11141	bc532	431	*colA*	Putative collagenase precursor	36	78/143	*C. perfringens* Sw: P43153
Y09212	bc72	3478	*gltT*	Proton/sodium-glutamate symport protein	59	355/355	*B. subtilis* Sw: P24943
			ansB	Aspartate ammonia-lyase	79	467/478	*B. subtilis* Sw: P26899
			malS	Malate dehydrogenase	71	291/291	*B. subtilis* Em: Z99119
Y09252	bc92	1131	*purH*	Phosphoribosylamino-imidazolecarboxamide-formyltransferase	74	335/337	*B. subtilis* Sw: P12048
Y11138	bc205	2402	*bc205a*	U (similarity to capsule cluster gene *cap5J*)	19	358/405	*S. aureus* Gb: U81973
			bc205b	U			
			bc205c	U (similarity to galactosamine-containing minor teichoic acid biosynthesis protein GgaB)	64	14/21	*B. subtilis* Sw: P46918
Y10981	bc45	865	*proA*	γ-Glutamyl phosphate reductase	67	247/247	*B. subtilis* Sw: P39821
			yxjA	Putative pyrimidine nucleoside transport protein	86	22/22	*B. subtilis* Sw: P42312
Y11170	bc541	842	*ytdI*	U (similarity to *ytdI*)	58	177/184	*B. subtilis* Gb: AF008220
			ywjE	U (similarity to *ywjE*)	50	20/24	*B. subtilis* Sw: P45865
Y10909	bc73	1368	*sp5k_1*	Similarity to stage V sporulation protein K	40	275/455	*B. subtilis* Sw: P27643
Y11305	glpT	485	*glpT*	Glycerol-3-phosphate permease	71	153/161	*B. subtilis* Sw: P37948
Y09322	bc331	802	*nupC*	Nucleoside transport system protein	80	85/85	*E. coli* Sw: P33031

Table 1 *(cont.)*

Accession no.	Fragment	Length (bp)	Gene homologues	Putative function*	Identity (%)†	Amino acid overlap	Organism/Accession no.
			lytR	Putative negative transcriptional regulator for *lytABC* major autolysin operon	38	105/118	*B. subtilis* Sw: Q02115
Y11140	bc48	1497	*orB*	U (similarity to 2-oxoacid ferredoxin oxidoreductase)	29	119/120	*Sulfolobus acidocaldarius* PIR: JC4920
			ymcB	U (similarity to *ymcB*)	75	212/226	*B. subtilis* Em: Z99112
Y11171	bc542	2195	*yobN*	Putative L-amino acid oxidase	55	452/453	*B. subtilis* Gb: AF027868
			bc542b	U (Similarity to *ywcH* and *ytmO*)	62	21/21	*B. subtilis* Sw: P39606
					68	19/21	*B. subtilis* Gb: AF008220
Y10908	bc75	1911	*cysA*	Sulfate transport ATP-binding protein	46	233/233	*Synechococcus* PCC 7942 Sw: P14788
			wapA	Wall-associated protein	40	197/231	*B. stearothermophilus* Sw: P42018
Y09323	bc332	1606	*bc332a*	U (similarity to purine nucleoside phosphorylase homologue *pfs*)	26	204/216	*E. coli* Sw: P24247
			bc332b	U (similarity to *S. griseus phnO* regulatory protein for C-P lyase and to *B. subtilis ywnH*)	30	110/145	*Streptomyces griseus DDBJ*: D29661
					28	138/145	*B. subtilis* Em: Z99122
			IS231	Similarity (90% in 67/70 bp) to insertion sequence IS*231*G (embl: BACINSAA)			
			gerPF	Probable spore-germination protein F	89	39/39	*B. cereus* Gb: AF053927
Y11218	bc521	1796	*sp5k_2*	Similarity to stage V sporulation protein K	41	280/598	*B. subtilis* Sw: P27643
Y11172	bc543	90	*appC*	Oligopeptide ABC transport system permease protein	55	29/29	*B. subtilis* Sw: P42063
Y09255	bc86	245	*dnaI*	Primosomal protein	64	73/74	*B. subtilis* Sw: P06567
Y09719	bc69	2475	*mem1*	U (similar to transmembrane proteins)	33	193/196	*Saccharomyces cerevisiae* PIR: S69588
			bctL	Membrane transport protein, major facilitator superfamily (Økstad *et al.*, 1997)	29	401/441	*B. subtilis* Sw: P23054

Table 1 *(cont.)*

Accession no.	Fragment	Length (bp)	Gene homologues	Putative function*	Identity (%)†	Amino acid overlap	Organism/Accession no.
			glr	Glutamate racemase	59	267/269	*B. subtilis* Sw: P94556
X98455	bc203	5142	*ycbB*	Putative two-component system response regulator (Lindback *et al.*, 1997)	52	306/310	*B. subtilis* Sw: P40759
			snf2	Putative SNF2 helicase (Lindbäck *et al.*, 1997)	38	506/1066	*Synechocystis* sp. DDBJ: D90916
AJ010131	bc333	2303	*ykvW*	Putative heavy-metal-transporting ATPase	60	222/225	*B. subtilis* Em: Z99111
			yfkH	U (similarity to *yfkH*)	51	265/289	*B. subtilis* Em: Z99108
			bc333c	U			
AJ010111	bc9	6169	*pyc*	Pyruvate carboxylase	76	982/984	*B. stearothermophilus* DDBJ: D83706
			ctaA	Cytochrome aa_3 controlling protein	57	307/345	*B. stearothermophilus* DDBJ: D70843
			ctaB	Cytochrome *c* oxidase assembly factor	59	300/307	*B. subtilis* Sw: P24009
			ctaC	Cytochrome *c* oxidase subunit II	53	223/227	*B. stearothermophilus* DDBJ: D70843
AJ010134	bc8	408	*bc8a*	U (similarity to *nasR* nitrate regulatory protein)	31	63/130	*Klebsiella pneumoniae* PIR: A55859
AJ010135	bc121	380	*rpsB*	Ribosomal protein S2	98	48/51	*B. subtilis* Em: Z99112
			tsf	Elongation factor Ts	96	24/24	*B. subtilis* Em: Z99112
AJ010136	bc122	314	*ytiP*	U (similarity to *ytiP*)	66	102/104	*B. subtilis* Gb: AF00822
AJ010331	bc16	252	*bla1*	β-Lactamase I	88	60/68	*B. cereus* PIR: S03167
AJ010132	bc891	498	*bc89a*	U (similarity to inosine-uridine preferring nucleoside hydrolase/purine nucleosidase)	32	148/166	*Crithidia fasciculata* Sw: Q27546
AJ010133	bc892	452	*yvgN*	Putative dehydrogenase	78	127/130	*B. subtilis* Em: Z99121
AJ010112	bc85	125	*gerPF*	Probable spore germination protein F	100	33/33	*B. cereus* Gb: AF053927
AJ010129	bc421	615	*yxaA*	U (similarity to *yxaA*)	48	64/69	*B. subtilis* Sw: P42100
AJ010130	bc422	74	*hdh*	Putative homoserine dehydrogenase	60	23/24	*Synechocystis* PCC 6803 Sw: P52986
AJ010137	bc2	107	*ydhE*	Putative glycosyl transferase	51	29/35	*B. subtilis* Em: Z99107
AJ010138	bc201	3252	*spoIIQ*	Stage II sporulation; required for completion of engulfment	45	152/196	*B. subtilis* Em: Z99122

Table 1 *(cont.)*

Accession no.	Fragment	Length (bp)	Gene homologues	Putative function*	Identity (%)†	Amino acid overlap	Organism/Accession no.
			spoIIID	Stage III sporulation protein	98	90/90	*B. thuringiensis* DDBJ: D28169
			mbl	Cell shape-determination-like protein (*mreB*-like)	87	332/334	*B. subtilis* Sw: P39751
			bcr1	Bacillus cereus repeat 1; 155 nt sequence repeated in the *B. cereus* genome			
			fabZ	Putative hydroxymyristoyl-(acyl carrier protein) dehydratase	79	42/42	*B. subtilis* Em: Z83337
AJ010139	bc295	2866	*fumA*	Class I Fe–S fumarase	77	269/269	*B. stearothermophilus* Sw: Q04718
			yfjS	Polysaccharide deacetylase	57	255/260	*B. stearothermophilus* Sw: Q04729
			alkE	DNA-3-methyladenine glycosidase II	45	287/287	*B. subtilis* Em: Z99108
AJ010140	bc296	1356	*alkC*	DNA alkylation repair enzyme	40	243/256	*B. subtilis* Em: Y14080
AJ010128	bc297	2683	*bc297a*	U (similarity to *yvbW*)	23	196/203	*B. subtilis* SubtiList: *yvbW*
			alkD	DNA alkylation repair enzyme; shown by complementation, I. Alseth and others, unpublished results			
			glyS	Glycyl-tRNA synthetase	50	214/217	*M. tuberculosis* Em: Z95208
AJ000394	bc298	4392	*celR*	Putative transcriptional regulator of cellobiose phosphotransferase system operon	38	460/482	*B. stearothermophilus* PIR: A49898
			bc298b	U (similarity to zebrafish chordin)	30	126/137	*Danio rerio* Gb: AF034606
			chrA	Membrane transport protein. Similarity to chromate resistance proteins	25	379/393	*Synechocys.* PCC 6803 DDBJ: D90916
			celC	Cellobiose phosphotransferase	50	230/234	*B. stearothermophilus* PIR: D49898
			bcr1	*B. cereus* repeat 1; 155 nt sequence repeated in the *B. cereus* genome			
			celF	Putative phospho-β-glucosidase	78	9/9	*B. subtilis* SubtiList: *licH*

Table 1 (cont.)

Accession no.	Fragment	Length (bp)	Gene homologues	Putative function*	Identity (%)†	Amino acid overlap	Organism/Accession no.
					100	9/9	*B. cereus* Em: AJ007952
AJ007788	bc299	6006	*ilvD*	Isoleucine biosynthesis enzyme; dihydroxy-acid dehydratase	60	355/357	*B. subtilis* Sw: P05791
			ilvA	Isoleucine biosynthesis enzyme; threonine deaminase	64	417/420	*B. subtilis* Sw: P37946
			capA	U (similarity to *B. anthracis* CapA	29	307/368	*B. anthracis* Sw: P19579
				protein required for biosynthesis of type I capsular polysaccharide)	29	337/368	*B. subtilis* Em: Z92954
			bc299d	U			
			bc299e	U			
AJ007510	bc210	10915	*mcrB*	Putative DNA restriction enzyme	35	281/343	*M. thermoauto-trophicum* Gb: AE000833
			bc210b	U (similarity to SNF2 helicase)	20	200/257	*Plasmodium falciparum* Gb: AF003086
			t3mod	Type III restriction-modification system Mod protein	50	677/690	*S. typhimurium* Sw: P40814
			t3res	Type III restriction-modification system Res protein (Hegna *et al.*, 1992)	63	946/988	*S. typhimurium* Sw: P40815
			bc210e	U (similarity to *Borrelia burgdorferi* putative coding region BB0553)	23	222/431	*Bor. burgdorferi* Gb: AE001157
			bc210f	U (similarity to SNF2 helicases)	23	472/631	*B. cereus* Em: X98455

* U, unknown function (weak or no similarity to database entries, or similarity to proteins of unknown function).

† ORFs were used to search a non-redundant set of database entries from SWISS-PROT, GenPept (translated GenBank), PDB, SPUpdate and PIR using the BLAST search algorithm at the National Center for Biotechnology Information (NCBI), Bethesda, MD, USA (http://www.ncbi.nim.nih.gov).

biodegradable than other less substituted s-triazine ring compounds (University of Minnesota Biocatalysis/Biodegradation Database: http://www.labmed.umn.edu/umbbd/atr/atr_map.html). The presence in *B. cereus* of a *trzA* gene, encoding an *N*-ethylammeline chlorohydrolase homologue novel to bacilli, could allow *B. cereus* to degrade herbicides in its natural soil habitat. The role of the *orB* gene encoding a protein similar to the β-chain of 2-oxoacid:ferredoxin oxidoreductase, a key enzyme of the archaeal tricarboxylic acid cycle (Iwasaki *et al.*, 1995), is unclear. Putative *orB* homologues have been found in the archaea *Sulfolobus acidocaldarius* and *Methanobacterium thermoautotrophicum*, but have only been identified in one other eubacterium, the thermophile *Thermus thermophilus*.

B. subtilis gene organization is not conserved in *B. cereus*

Using previous allocations of the anonymous probes to the physical map (Carlson *et al.*, 1992; Carlson &

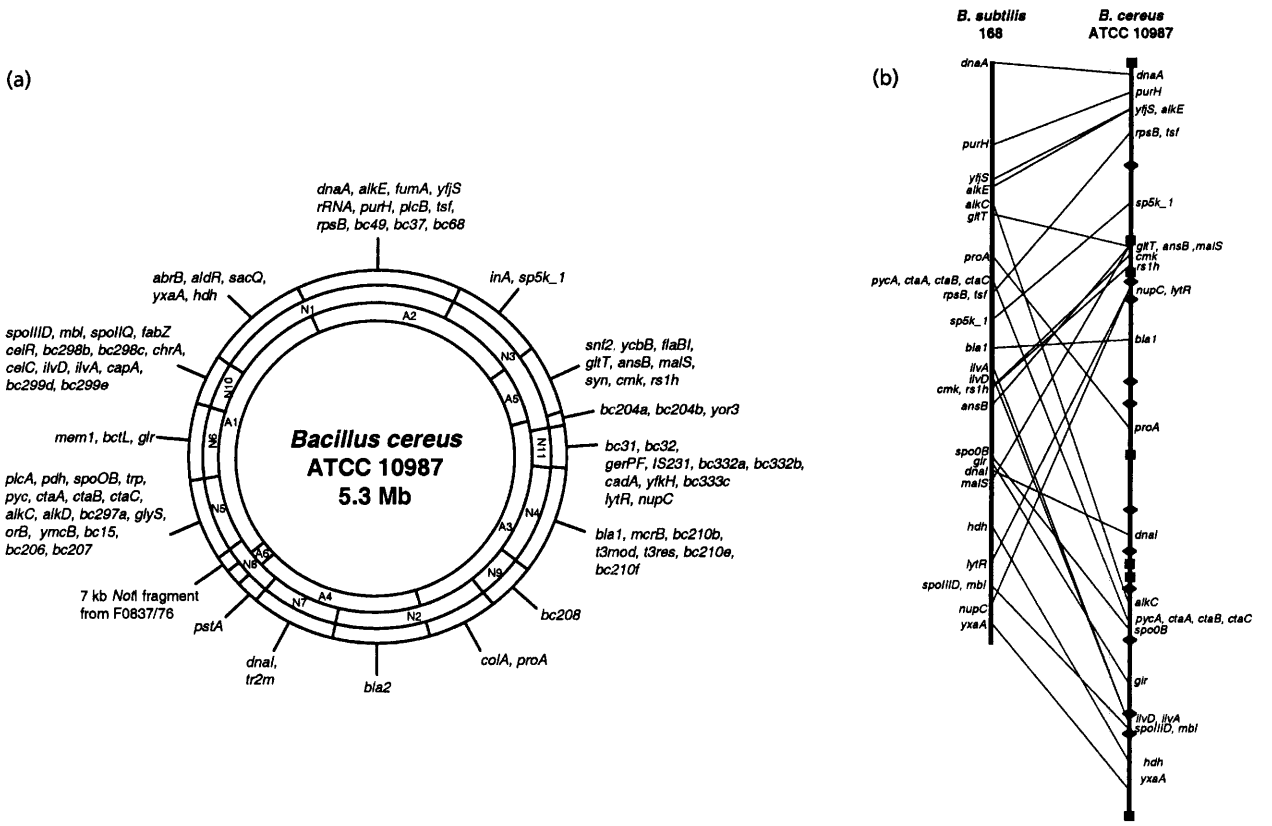

Fig. 1. (a) Physical map of the *B. cereus* ATCC 10987 chromosome. Inner circle, *Asc*I fragments; middle circle, *Not*I fragments. Restriction fragments are ordered according to size, with the lowest number corresponding to the largest fragment. (b) Comparison of gene localization on the *B. cereus* ATCC 10987 and *B. subtilis* 168 chromosomes. *Not*I sites and *Asc*I sites are indicated by diamonds and squares, respectively. Both chromosomes are shown linearized, *B. subtilis* with *dnaA* as the first gene on top, *B. cereus* with the region common to N1 and A1 restriction fragments on top. The comparison was drawn to obtain maximum similarity between the maps.

Kolstø, 1994), 64 of the identified genes/ORFs were assigned positions on the *B. cereus* ATCC 10987 chromosome (Fig. 1a). The updated map contains a total of 91 markers, of which 80 are genes/ORFs of known or unknown function, one is IS*231*, and 10 are anonymous. To examine whether the global gene order was conserved between *B. cereus* and *B. subtilis*, we compared the chromosomal location of 30 putative orthologues using the *B. cereus* ATCC 10987 physical map and the *B. subtilis* 168 genome sequence (Kunst *et al.*, 1997). The orthologues did not exhibit any apparent conservation of organization (Fig. 1b), but were, by contrast, arranged in an apparently random manner. This is in agreement with what has been observed when comparing the genomes of the distantly related bacteria *Haemophilus influenzae*, *Mycoplasma genitalium* and *Methanococcus jannaschii* (Kolstø, 1997), and may indicate that gene organization is generally not conserved even between members of the same genus, like *B. subtilis* and *B. cereus*. Similar results were found when comparing the position of nine equivalent genes in alkaliphilic *Bacillus firmus* OF4 with *B. subtilis* 168 (Grønstad *et al.*, 1998).

There were, however, examples where genes exhibited an identical clustering in *B. cereus* and *B. subtilis*: (1) *spoIIID* and *mbl*; (2) *pycA*, *ctaA*, *ctaB*, and *ctaC*; (3) *rpsB* and *tsf*. It is, however, interesting to note that although the local gene order was conserved, the chromosomal location of the three loci was very different in the two bacteria (Fig. 1b). Furthermore, the *rpsB–tsf* gene order encoding ribosomal protein S2 and elongation factor Ts is one of the very few examples of gene pairs for which orthologues are close to being universally clustered in eubacteria, being conserved in 10 out of 14 species where both genes have been sequenced. Additionally, putative homologues of two other *B. cereus* gene pairs were located in the vicinity of each other in *B. subtilis*, although lacking a tight physical coupling. (1) The branched chain amino acid synthesis genes *ilvD* and *ilvA* were in *B. subtilis* 168 separated by an 8·4 kb region harbouring nine genes. Similarly, (2) the putative polysaccharide deacetylase gene *yfjS* and the DNA-3-methyladenine glycosidase II gene *alkE* were 2·8 kb apart in *B. subtilis*, separated by two ORFs with similarity to 3-hydroxyisobutyrate dehydrogenase (*yfjR*) and a divalent cation transporter (*yfjQ*), respectively.

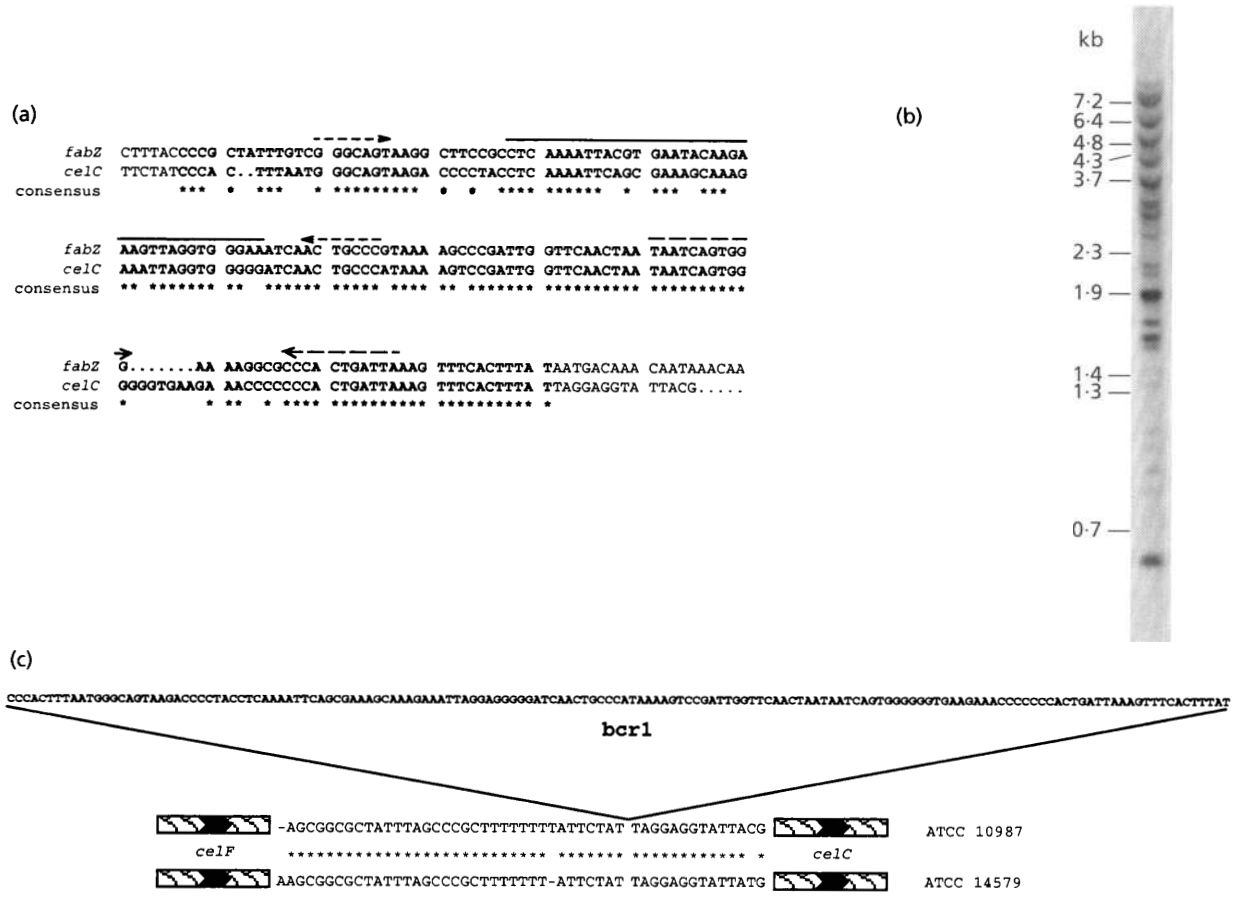

Fig. 2. (a) Two gene loci from *B. cereus* ATCC 10987 and ATCC 14579 containing the 154 bp *bcr1* repeat, with the defined *bcr1* sequence in bold. Inverted repeats within *bcr1* are marked by dashed arrows. The region covered by the oligonucleotide probe is indicated by the solid line above the sequences. (b) Analysis of *B. cereus* ATCC 10987 genomic DNA using 37 nt of the *bcr1* repeat as probe. Genomic DNA was digested with *Hin*cII, electrophoresed and blotted to a nylon membrane. Numbers indicate the positions of the molecular size marker (λ DNA digested with *Bst*EII, New England Biolabs). (c) Comparison of the *celF–celC* gene locus from *B. cereus* strains ATCC 10987 and ATCC 14579. Apart from the *bcr1* insertion, the intergenic regions were conserved in the two strains.

A novel repeat sequence in the *B. cereus* genome

Sequencing of the *celR-bc298a-chrA-celC-celF* gene locus showed that an intergenic 155 bp sequence in front of the *celC* gene was highly similar to a sequence upstream of *fabZ* in the *spoIIQ-spoIIID-mbl-fabZ* gene locus (Fig. 2a). We have named this novel repeated sequence *bcr1* (*Bacillus cereus* repeat 1). Repeats of *bcr1* were also found upstream of the *hbl* enterotoxin locus from strain ATCC 14579 (O. A. Økstad & A.-B. Kolstø, unpublished results), and upstream of *bctL* (Økstad *et al.*, 1997) and the *bct504a* gene from the same strain (data not shown). The *bcr1* sequences exhibited 78–93% pairwise identity, and a multiple alignment of the repeats showed that 63% of the nucleotide positions were conserved in all five loci. The sequences displayed an inverted repeat character, and were in all instances found in non-coding regions.

To investigate the degree of *bcr1* repetition, an oligonucleotide probe of 37 bp within *bcr1* was prepared, and blots of *Hin*cII-digested genomic DNA were hybridized with the probe. In strain ATCC 10987, 11 fragments hybridized strongly to the probe, while 11 fragments displayed weaker hybridization intensity (Fig. 2b). Similar hybridization results were obtained with 16 other *B. cereus* and *B. thuringiensis* strains tested, indicating that *bcr1* may be ubiquitous in *B. cereus* and *B. thuringiensis* strains. The locations of the repeat within the genome may vary, as shown by the presence of *bcr1* in the *celC–celF* locus in strain ATCC 10987, and its absence in the corresponding locus from strain ATCC 14579 (Fig. 2c). No hybridization by the *bcr1* probe to *B. subtilis* DNA was observed, and the sequence was not present when searching SubtiList or any other database. The *bcr1* repeat did not display similarity to any integron family sequences, including the repetitive sequence elements recently reported in *Vibrio cholerae* (Clark *et al.*, 1997).

Concluding remarks

E. coli and *S. typhimurium*, species thought to have differentiated 120–160 million years ago (Ochman & Wilson, 1987), have highly similar genetic maps. This led to the prediction that bacterial genomes would turn out to be generally stable, exhibiting a conserved chromosomal gene organization. Subsequent studies have indicated that this is generally not the case. Bacteria may be divided into two subgroups in terms of genome stability, where one group consists of stable species whereas members of the other subgroup display a higher level of heterogeneity within the species (Fonstein & Haselkorn, 1995). *B. cereus* clearly belongs to the latter group, with different strains showing a highly variable gene organization, particularly in certain regions of the genome (Carlson & Kolstø, 1994). *Salmonella typhi* (Liu & Sanderson, 1996), *Pseudomonas stutzeri* (Ginard *et al.*, 1997) and *Brucella suis* (Jumas-Bilak *et al.*, 1998) also belong to the high variability group, and homologous recombination between the *rrn* loci seems to be the major cause of their heterogeneity. It is interesting to note that the organization of the *rrn* gene operons within the clostridial chromosomes varies considerably (Cornillot *et al.*, 1997). In fact, the *rrn* operon organization of *Clostridium acetobutylicum* was more similar to that of *B. subtilis* than to other clostridia. By contrast, *B. subtilis* 168 has, by the examination of a number of derived strains, been found to have a generally stable chromosome structure in agreement with the high genetic stability of the species (Itaya, 1993).

We have previously suggested that *B. cereus* strains have been subject to a high frequency of genome rearrangements (Carlson *et al.*, 1992; Carlson & Kolstø, 1993, 1994). The present study shows that the chromosomal gene organization in *B. cereus* ATCC 10987 is widely different from *B. subtilis* 168. These results add to a growing number of studies indicating that bacteria exhibiting a conserved chromosomal gene organization may in fact be exceptions rather than the rule. We do not know if the *bcr1* sequence repeats may contribute to the chromosome variation and flexibility observed in strains of *B. cereus* and *B. thuringiensis*.

ACKNOWLEDGEMENTS

The authors are thankful to Ewa Jaroszewicz, Henning A. Johansen and Liv A. Bjørland for skilful technical assistance. The work was supported by grants from The Norwegian Research Council to A.B.K.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Aronson, A. I. (1993). The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival. *Mol Microbiol* **7**, 489–496.

Carlson, C. R. & Kolstø, A.-B. (1993). A complete physical map of a *Bacillus thuringiensis* chromosome. *J Bacteriol* **175**, 1053–1060.

Carlson, C. R. & Kolstø, A.-B. (1994). A small (2·5 Mb) *Bacillus cereus* chromosome corresponds to a conserved region of a larger (5·3 Mb) *Bacillus cereus* chromosome. *Mol Microbiol* **13**, 161–169.

Carlson, C. R., Grønstad, A. & Kolstø, A.-B. (1992). Physical maps of the genomes of three *Bacillus cereus* strains. *J Bacteriol* **174**, 3750–3756.

Carlson, C. R., Johansen, T. & Kolstø, A.-B. (1996a). The chromosome map of *Bacillus thuringiensis* subsp. *canadensis* HD224 is highly similar to that of the *Bacillus cereus* type strain ATCC 14579. *FEMS Microbiol Lett* **141**, 163–167.

Carlson, C. R., Johansen, T., Lecadet, M.-M. & Kolstø, A.-B. (1996b). Genomic organization of the entomopathogenic bacterium *Bacillus thuringiensis* subsp. *berliner* 1715. *Microbiology* **142**, 1625–1634.

Clark, C. A., Purins, L., Kaewrakon, P. & Manning, P. A. (1997). VCR repetitive sequence elements in the *Vibrio cholerae* chromosome constitute a mega-integron. *Mol Microbiol* **26**, 1137–1138.

Cornillot, E., Croux, C. & Soucaille, P. (1997). Physical and genetic map of the *Clostridium acetobutylicum* ATCC 824 chromosome. *J Bacteriol* **179**, 7426–7434.

Devereux, J., Haeberli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* **12**, 387–395.

Drobniewski, F. A. (1993). *Bacillus cereus* and related species. *Clin Microbiol Rev* **6**, 324–338.

Feng, D. F. & Doolittle, R. F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J Mol Evol* **25**, 351–360.

Fonstein, M. & Haselkorn, R. (1995). Physical mapping of bacterial genomes. *J Bacteriol* **177**, 3361–3369.

Ginard, M., Lalucat, J., Tummler, B. & Romling, U. (1997). Genome organisation of *Pseudomonas stutzeri* and resulting taxonomic and evolutionary considerations. *Int J Syst Bacteriol* **47**, 132–143.

Gordon, R. E., Haynes, W. C. & Pang C. H.-N. (1973). *The Genus Bacillus*. Agriculture Handbook no. 427. Washington, DC: United States Department of Agriculture.

Green, L. S., Laudenbach, D. E. & Grossman, A. R. (1989). A region of a cyanobacterial genome required for sulfate transport. *Proc Natl Acad Sci USA* **86**, 1949–1953.

Grønstad, A., Jaroszewicz, E., Ito, M., Sturr, M. G., Krulwich, T. A. & Kolstø, A.-B. (1998). Physical map of alkaliphilic *Bacillus firmus* OF4 and detection of a large endogenous plasmid. *Extremophiles* **2**, 447–453.

Hegna, I. K., Karlstrom, E. S., Lopez, R., Kristensen, T. & Kolsto, A.-B. (1992). A type-III DNA restriction and modification system in *Bacillus cereus*. *Gene* **114**, 149–150.

Higgins, D. G. & Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *Comput Appl Biosci* **5**, 151–153.

Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B. & Herrmann, R. (1997). Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res* **25**, 701–712.

Itaya, M. (1993). Stability and asymmetric replication of the *Bacillus subtilis* 168 chromosome structure. *J Bacteriol* **175**, 741–749.

Iwasaki, T., Wakagi, T. & Oshima, T. (1995). Ferredoxin-dependent redox system of a thermoacidophilic archaeon, *Sulfolobus* sp.

strain 7. Purification and characterization of a novel reduced ferredoxin-reoxidizing iron-sulfur flavoprotein. *J Biol Chem* **270**, 17878–17883.

Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., O'Callaghan, D. & Ramuz, M. (1998). Differences in chromosome number and genome rearrangements in the genus *Brucella*. *Mol Microbiol* **27**, 99–106.

Kolstø, A.-B. (1997). Dynamic bacterial genome organisation. *Mol Microbiol* **24**, 241–248.

Kolstø, A.-B., Grønstad, A. & Oppegaard, H. (1990). Physical map of the *Bacillus cereus* chromosome. *J Bacteriol* **172**, 3821–3825.

Koonin, E. V., Mushegian, A. R. & Rudd, K. E. (1996). Sequencing and analysis of bacterial genomes. *Curr Biol* **6**, 404–416.

Kunst, F., Ogasawara, N., Moszer, I. & 148 other authors (1997). The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256.

Lindbäck, T., Kolstø, A.-B. & Kolst, A. B. (1997). A *Bacillus cereus* member of the SNF2 family. *Microbiology* **143**, 171–174.

Liu, S. L. & Sanderson, K. E. (1996). Highly plastic chromosomal organization in *Salmonella typhi*. *Proc Natl Acad Sci USA* **93**, 10303–10308.

Logan, D. T., Mazauric, M. H., Kern, D. & Moras, D. (1995). Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J* **14**, 4156–4167.

Lövgren, A., Carlson, C. R., Eskils, K. & Kolstø, A.-B. (1998). Localization of putative virulence genes on a physical map of the *Bacillus thuringiensis* subsp. *gelechiae* chromosome. *Curr Microbiol* **37**, 245–250.

Needleman, S. B. & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**, 443–453.

Ochman, H. & Wilson, A. C. (1987). Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *Mol Evol* **26**, 74–86.

Økstad, O. A., Grønstad, A., Lindbäck, T. & Kolstø, A.-B. (1997). Insertional inactivation of a Tet(K)/Tet(L) like transporter does not eliminate tetracycline resistance in *Bacillus cereus*. *FEMS Microbiol Lett* **154**, 181–186.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**, 2444–2448.

Priest, F. G. (1993). Systematics and ecology of *Bacillus*. In *Bacillus subtilis and Other Gram-positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*, pp. 3–16. Edited by A. L. Sonenshein, J. A. Hoch & R. Losick. Washington, DC: American Society for Microbiology.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA-sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467.

Smith, D. R., Doucette-Stamm, L. A., Deloughery, C. & 34 other authors (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *J Bacteriol* **179**, 7135–7155.

Smith, N. R., Gordon, R. E. & Clark, F. E. (1952). In *Aerobic Sporeforming Bacteria*, pp. 56–67. Washington, DC: US Department of Agriculture.

Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, W. S., Borodovsky, M., Rudd, K. E. & Koonin, E. V. (1996). Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr Biol* **6**, 279–291.

Received 21 September 1998; revised 10 November 1998; accepted 18 November 1998.