The average distance between the partition and the CRF centre was 5.6 cm and was always greater than 4 cm. Using results reported previously[6], we estimated the maximal CRF dimensions as follows. Average CRF centre and surround areas were summed, then doubled. The maximum distance from the centre of the receptive field to the CRF boundary was then estimated as the radius of a circle having this area (2.4 cm). This conservative estimate is lower than 4 cm. Along with the lack of pyramidal-cell responses to stimulation of the head chamber alone, this indicates that it is very unlikely that the CRF surround extends past the partition.

Received 20 January; accepted 17 March 2003; doi:10.1038/nature01590.

1. Rieke, F., Bodnar, D. A. & Bialek, W. Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc. R. Soc. Lond. B* **262,** 259–265 (1995).
2. Machens, C. K. *et al.* Representation of acoustic communication signals by insect auditory neurons. *J. Neurosci.* **21,** 3215–3227 (2001).
3. Nelson, M. E. & MacIver, M. A. Prey capture in the weakly electric fish *Apteronotus leptorhynchus*: sensory acquisition strategies and electrosensory consequences. *J. Exp. Biol.* **202,** 1195–1203 (1999).
4. Zupanc, G. K. H. & Maler, L. Evoked chirping in the weakly electric fish *Apteronotus leptorhynchus*: a quantitative biophysical analysis. *Can. J. Zool.* **71,** 2301–2310 (1993).
5. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160,** 106–154 (1962).
6. Bastian, J., Chacron, M. J. & Maler, L. Receptive field organization determines pyramidal cell stimulus-encoding capability and spatial stimulus selectivity. *J. Neurosci.* **22,** 4577–4590 (2002).
7. Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J. & Davis, J. Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* **378,** 492–496 (1995).
8. Vinje, W. & Gallant, J. L. Sparse Coding and decorrelation in primary visual cortex during natural vision. *Science* **287,** 1273–1276 (2000).
9. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24,** 1193–1216 (2001).
10. Voss, R. F. & Clarke, J. '1/*f* noise' in music: music from 1/*f* noise. *J. Acoust. Soc. Am.* **63,** 258–263 (1978).
11. Bastian, J. Electrolocation. I. How the electroreceptors of *Apteronotus albifrons* code for moving objects and other electrical stimuli. *J. Comp. Physiol. A* **144,** 465–479 (1981).
12. Gabbiani, F., Metzner, W., Wessel, R. & Koch, C. From stimulus encoding to feature extraction in weakly electric fish. *Nature* **384,** 564–567 (1996).
13. Maler, L., Sas, E. K. & Rogers, J. The cytology of the posterior lateral line lobe of high frequency weakly electric fish (*Gymnotoidei*): Dendritic differentiation and synaptic specificity in a simple cortex. *J. Comp. Neurol.* **195,** 87–139 (1981).
14. Rieke, F., Warland, D., de Ruyter van Steveninck, R. R. & Bialek, W. *Spikes: Exploring the Neural Code* (MIT, Cambridge, Massachusetts, 1996).
15. Borst, A. & Theunissen, F. Information theory and neural coding. *Nature Neurosci.* **2,** 947–957 (1999).
16. Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268,** 1503–1506 (1995).
17. de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R. & Bialek, W. Reproducibility and variability in neural spike trains. *Science* **275,** 1805–1808 (1997).
18. Maler, L. & Mugnaini, E. Correlating gamma-aminobutyric acidergic circuits and sensory function in the electrosensory lateral line lobe of a gymnotiform fish. *J. Comp. Neurol.* **345,** 224–252 (1994).
19. Berman, N. J. & Maler, L. Neural architecture of the electrosensory lateral line lobe: Adaptations for coincidence detection, a sensory searchlight and frequency-dependent adaptive filtering. *J. Exp. Biol.* **202,** 1243–1253 (1999).
20. Crampton, W. G. R. Electric signal design and habitat preferences in a species rich assembly of gymnotiform fishes from the upper Amazon basin. *Anais Acad. Bras. Cienc.* **70,** 805–847 (1998).
21. Zhang, H., Xu, J. & Feng, A. S. Effects of GABA-mediated inhibition on direction-dependent frequency tuning in the frog inferior colliculus. *J. Comp. Physiol.* **184,** 85–98 (1999).
22. Macleod, K. & Laurent, G. Distinct mechanisms for synchronization and temporal patterning of odor-encoding neural assembies. *Science* **274,** 976–979 (1996).
23. Doiron, B., Chacron, M. J., Maler, L., Longtin, A. & Bastian, J. Inhibitory feedback required for network burst responses to communication but not prey stimuli. *Nature* **421,** 539–543 (2003).
24. Carr, C. E., Maler, L. & Sas, E. Peripheral organization and central projections of the electrosensory organs in gymnotiform fish. *J. Comp. Neurol.* **211,** 139–153 (1982).
25. Heiligenberg, W. & Dye, J. Labelling of electrosensory afferents in a gymnotid fish by intracellular injection of HRP: The mystery of multiple maps. *J. Comp. Physiol. A* **148,** 287–296 (1982).
26. Metzner, W. & Heiligenberg, W. The coding of signals in the electric communication of the gymnotiform fish *Eigenmannia*: From electroreceptors to neurons in the torus semicircularis of the midbrain. *J. Comp. Physiol. A* **169,** 135–150 (1991).
27. Metzner, W. & Juranek, J. A sensory brain map for each behavior? *Proc. Natl Acad. Sci. USA* **26,** 14798–14803 (1997).
28. Heiligenberg, W. *Neural Nets in Electric Fish* (MIT, Cambridge, Massachusetts, 1991).
29. Bastchelet, E. *Circular Statistics in Biology* (Academic, New York, 1981).
30. Gabbiani, F. Coding of time varying signals in spike trains of linear and half-wave rectifying neurons. *Network Comput. Neural Sys.* **7,** 61–85 (1996).

Acknowledgements We thank A.-M. Oswald, J. Lewis and B. Lindner for their reading the manuscript. This research was supported by NSERC (M.J.C., B.D., A.L.), CIHR (L.M., A.L.) and NIH (J.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.C. (mchacron@physics.uottawa.ca).

The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria

Timothy D. Read*†, Scott N. Peterson*‡, Nicolas Tourasse§#, Les W. Baillie*†‖, Ian T. Paulsen*¶, Karen E. Nelson*, Hervé Tettelin*, Derrick E. Fouts*, Jonathan A. Eisen*¶, Steven R. Gill*, Erik K. Holtzapple*, Ole Andreas Økstad§#, Erlendur Helgason§#, Jennifer Rilstone*, Martin Wu*, James F. Kolonay*, Maureen J. Beanan*, Robert J. Dodson*, Lauren M. Brinkac*, Michelle Gwinn*, Robert T. DeBoy*, Ramana Madpu*, Sean C. Daugherty*, A. Scott Durkin*, Daniel H. Haft*, William C. Nelson*, Jeremy D. Peterson*, Mihai Pop*, Hoda M. Khouri*, Diana Radune*, Jonathan L. Benton*, Yasmin Mahamoud*, Lingxia Jiang*, Ioana R. Hance*, Janice F. Weidman*, Kristi J. Berry*, Roger D. Plaut*, Alex M. Wolf*, Kisha L. Watkins*, William C. Nierman*, Alyson Hazen*, Robin Cline*, Caroline Redmond†, Joanne E. Thwaite†, Owen White*, Steven L. Salzberg*¶, Brendan Thomason☆, Arthur M. Friedlander, Theresa M. Koehler††, Philip C. Hanna☆, Anne-Brit Kolstø§# & Claire M. Fraser*‡‡§§**

** The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA*
† Medical Biotechnology Center, University of Maryland Biotechnology Institute, Baltimore, Maryland 21201, USA
‡ Department of Biochemistry, ‡‡ Department of Microbiology and Tropical Medicine, §§ Department of Pharmacology, The George Washington University, Eye Street, Washington DC 20052, USA
§ School of Pharmacy, University of Oslo N-0316, Oslo, Norway
‖ Defence Science Technology Laboratory, Porton Down, Salisbury SP4 0JQ, UK
¶ Johns Hopkins University, Charles and 34th Streets, Baltimore, Maryland 21218, USA
 # The Biotechnology Center of Oslo, Oslo N-0317, Norway
☆ Department of Microbiology & Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA
*** US Army Medical Research Institute for Infectious Diseases, Frederick, Maryland 21702, USA*
†† Department of Microbiology and Molecular Genetics, University of Texas–Houston Health Science Center Medical School, University of Texas, Houston, Texas 77225, USA

Bacillus anthracis **is an endospore-forming bacterium that causes inhalational anthrax[1]. Key virulence genes are found on plasmids (extra-chromosomal, circular, double-stranded DNA molecules) pXO1 (ref. 2) and pXO2 (ref. 3). To identify additional genes that might contribute to virulence, we analysed the complete sequence of the chromosome of *B. anthracis* Ames (about 5.23 megabases). We found several chromosomally encoded proteins that may contribute to pathogenicity—including haemolysins, phospholipases and iron acquisition functions—and identified numerous surface proteins that might be important targets for vaccines and drugs. Almost all these putative chromosomal virulence and surface proteins have homologues in *Bacillus cereus*, highlighting the similarity of *B. anthracis* to near-neighbours that are not associated with anthrax[4]. By performing a comparative genome hybridization of 19 *B. cereus* and *Bacillus thuringiensis* strains against a *B. anthracis* DNA microarray, we confirmed the general similarity of chromosomal genes among this group of close relatives. However, we found that the gene sequences of pXO1 and pXO2 were more variable between strains, suggesting plasmid mobility in the group. The complete sequence of *B. anthracis* is a step towards a better understanding of anthrax pathogenesis.**

B. anthracis has become notorious as a bioweapon because of its tough, environmentally resistant endospore and its ability to cause lethal inhalational anthrax. During the course of the disease,

endospores are taken up by alveolar macrophages where they germinate in the phagolysosomal compartment[1]. Vegetative cells then escape from the macrophage, eventually infecting blood. Expression of the major plasmid-encoded virulence determinants, tripartite toxin and a poly-D-glutamic acid capsule, are essential for full pathogenicity[1]. Sequencing the chromosome of *B. anthracis* was undertaken to help identify additional genes that might contribute to virulence either by encoding functions necessary for the survival and escape from the mammalian macrophage or by enhancing evasion of the immune system and the extent of damage caused by the bacterium to its animal host.

The *B. anthracis* Ames chromosome sequenced in this work (5,227,293 base pairs, bp) derives from an isolate taken from a dead cow in Texas (Methods). This sequence differs in only 11 confirmed single nucleotide polymorphisms[5] (SNPs) from the 2001 Florida attack Ames isolate, verifying that the chromosome sequenced to completion is essentially identical to a virulent strain. The chromosome encodes 5,508 predicted protein-coding sequences (Table 1) with a pronounced bias for genes on the replication leading strand (Fig. 1), as has been seen in other low G + C Gram-positive replicons[6]. A feature shared with the chromosomes of other endospore-forming Gram-positive species of the genera *Bacillus* and *Clostridium*[7,8] is the concentration of the ribosomal RNA, transfer RNA and ribosomal protein genes around the replication origin. This arrangement may maximize protein synthesis during early rounds of DNA replication after germination from the dormant endospore phase. The chromosome also contains at least four prophages (Supplementary Information) as well as two type I introns, one of which disrupts the *recA* gene[9]. Housekeeping functions such as DNA replication and fatty-acid metabolism are overwhelmingly partitioned to the chromosome, whereas the pXO1 and pXO2 plasmids have a greater proportion of transposons, genes involved in toxicity and genes without function assigned (Table 1).

Most *B. anthracis* Ames chromosomal proteins have homologues to proteins encoded on the draft genome sequence of *B. cereus* ATCC 10987 (T.D.R., unpublished results), a closely related strain (Figs 1 and 2). There are only 141 proteins in *B. anthracis* for which a putative functional assignment could not be made that do not have a match in the protein set of *B. cereus* ATCC 10987 sequence (BLASTP[10] $E < 10^{-5}$). For the most part, these are encoded by genes of unknown function, are transposases or are present in phage regions. Almost all potential chromosomal virulence-enhancing genes have homologues in *B. cereus* ATCC 10987, suggesting that they are not specifically associated with the unique pathogenicity of *B. anthracis* but are part of the common arsenal of the *B. cereus* group of bacteria[11].

Table 1 **Features of the *B. anthracis* Ames genome**

Feature	Chromosome	pXO1*	pXO2*
Size (bp)	5,227,293	181,677	94,829
Number of genes	5,508	217	113
Replicon coding (%)	84.3	77.1	76.2
Average gene length (nt)	800	645	639
G+C content (%)	35.4	32.5	33.0
rRNA operons	11	0	0
tRNAs	95	0	0
sRNAs	3	2	0
Phage genes†	62	0	0
Transposon genes†	18	15	6
Disrupted reading frame‡	37	5	7
Genes with assigned function	2,762	65	38
Conserved hypothetical genes	1,212	22	19
Genes of unknown function	657	8	5
Hypothetical genes	877	122	51

*The complete, annotated pXO1 and pXO2 plasmids from an Ames strain isolated from the 2001 US bioterror attack were also resequenced recently at TIGR[5] and have been included in this analysis.
†According to TIGR role categories.
‡Genes with indels or point mutations resulting in early termination, confirmed by resequencing.

The chromosome of *B. anthracis* Ames contains several homologues of genes known to be involved in *B. cereus* and *B. thuringiensis* pathogenesis. These include two channel-forming type III haemolysins (BA5701, BA2241) and a complex of three non-haemolytic enterotoxins (BA1887–1889). Several *B. anthracis* Ames proteins have sequence homology to proteins that contribute to the virulence of the Gram-positive pathogen *Listeria monocytogenes*[12]. These include phosphatidyl-inositol-specific and phosphatidyl-choline-preferring phospholipase C (BA0677 and BA3891), internalin-like genes (BA1346 and BA1406), listeriolysin O (BA3355), sigma factor B (BA0992) and p60 extracellular protease (BA1952 and BA5474). The significance of these homologies may lie in the similarities in the pathways of intracellular survival and multiplication of *L. monocytogenes*, and the germination, survival and escape from macrophages by *B. anthracis*.

B. anthracis contains a gene encoding a homologue of the enhancin protein (BA3443), first described in baculoviruses that infect gypsy moths. Enhancin is a metalloprotease that boosts viral infectivity by degrading the mucin layer surrounding insect guts[13]. A homologue of *B. anthracis* enhancin is also found in the genome of *Yersinia pestis*, which survives in both mammals and insects[14]. *B. anthracis* also contains two homologues of *B. thuringiensis* immune inhibitor A metalloprotease (BA0672 and BA1295), which enhances virulence in insects through cleavage of bacteriocidal lectins[11]. The presence of these genes may be evidence of an insect-infecting lifestyle in a recent ancestor.

Germination of the anthrax endospore is a key initial event in the *B. anthracis* infectious cycle. *B. anthracis* has seven (six chromosomal and one plasmid-borne) paralogues of the *gerA* family of tricistronic operons utilized by endospores to recognize the presence of specific small molecules to initiate the germination process[15]. Protection of DNA during dormancy and efficient DNA repair during germination are also believed to be important factors in endospore viability. *B. anthracis* has several homologues of the *Bacillus subtilis* small acid soluble DNA protection proteins, and the full complement of DNA repair proteins found in *B. subtilis*. *B. anthracis* also appears to have additional DNA repair capabilities focused on UV-induced DNA damage, with a unique deoxyribodipyrimidine photolyase gene (BA3180) and two, rather than one, UV dimer endonucleases. The photolyase is more closely related to enzymes from proteobacteria than those from other Gram-positive bacteria. The *B. anthracis* genome encodes several proteins that mitigate damage by free-oxygen radicals, including five catalases and three Fe-Mn superoxide dismutases. Other detoxification functions for which no obvious homologues could be found in *B. subtilis* include bromoperoxidase, thiolperoxidase, multiple thioredoxin proteins and a cytoplasmic Cu-Zn superoxide dismutase (SodC; BA5139). SodC has been shown to have a key role in the virulence of certain other intracellular bacteria, counteracting nitric oxide-mediated killing in the macrophage[16].

The *B. anthracis* chromosome encodes a machinery for sporulation that is broadly similar to *B. subtilis*[7]. The proteins with the highest degree of sequence divergence between the species are endospore coat constituents and endospore polysaccharide biosynthesis components, suggesting altered composition of the outer surface. *B. subtilis* alternative sigma factors, which govern a cascade of events associated with cell development, are also generally conserved. One sigma factor missing in *B. anthracis* is *sigD*, which is essential for the expression of the flagellum operon[17]. However, *L. monocytogenes*, which is motile and carries a flagellum operon similar to *B. anthracis*, also lacks a *sigD* gene[18].

Despite having numerous predicted secreted proteins encoded in its genome (Supplementary Information), *B. anthracis* is notable for paucity of extracellular protease activity under standard laboratory conditions[19]. One reason for this lack of protein secretion may lie in a mutation that affects regulation of gene expression: a nonsense mutation in the *plcR* positive regulator gene[20]. In *B. thuringiensis*

and *B. cereus*, the *plcR* gene product is known to upregulate the production of numerous extracellular enzymes through binding at an upstream motif (TATGNAN$_4$TNCATA). Although the *B. anthracis plcR* homologue is truncated, there are 56 putative *plcR* binding motifs in the chromosome and 2 on pXO2. The extracellular protein genes downstream include phospholipases, enterotoxins and haemolysins (Table 2), and the *plcR* mutation has been shown to account for a dramatic reduction in lecithinase, protease and haemolysin production by *B. anthracis*[19]. However, it is possible that some PlcR-regulated gene products still contribute to virulence but are under alternative regulatory controls, as low-level expression of some of the genes in the PlcR regulon has been reported in *B. anthracis*[19]. There is another PlcR-family protein in the genome (BA0597) that might potentially function to complement expression under certain conditions.

The chromosome of *B. anthracis* contains three homologues of the sortase transpeptidase responsible for attachment of secreted proteins to peptidoglycans on the cell surface of Gram-positive bacteria[21], and also contains the *csaAB* genes for binding of proteins with S-layer homology (SLH) domains to polysaccharide. Using searches against models for the sortase attachment sites and SLH domains, 34 candidate surface proteins were identified (Supplementary Information). Two putative *B. anthracis* sortase-attached genes have internalin-like repeats[11]. The potential role of most proteins with SLH domains on the surface of *B. anthracis* is unknown at present. However, these surface proteins may mediate as-yet-unknown interactions between *B. anthracis* and its external environment, and could be targets for vaccine and drug design.

The broad similarity in metabolic and transport genes of *B. anthracis* and *B. subtilis* (the model aerobic Gram-positive organism)[7] suggests many common capabilities, yet there are a number of idiosyncrasies that may shed light on the ecology of *B. anthracis*. Compared to *B. subtilis*, *B. anthracis* appears to have an expanded capacity for amino-acid and peptide utilization. For instance, there are 17 ABC-type peptide binding proteins in

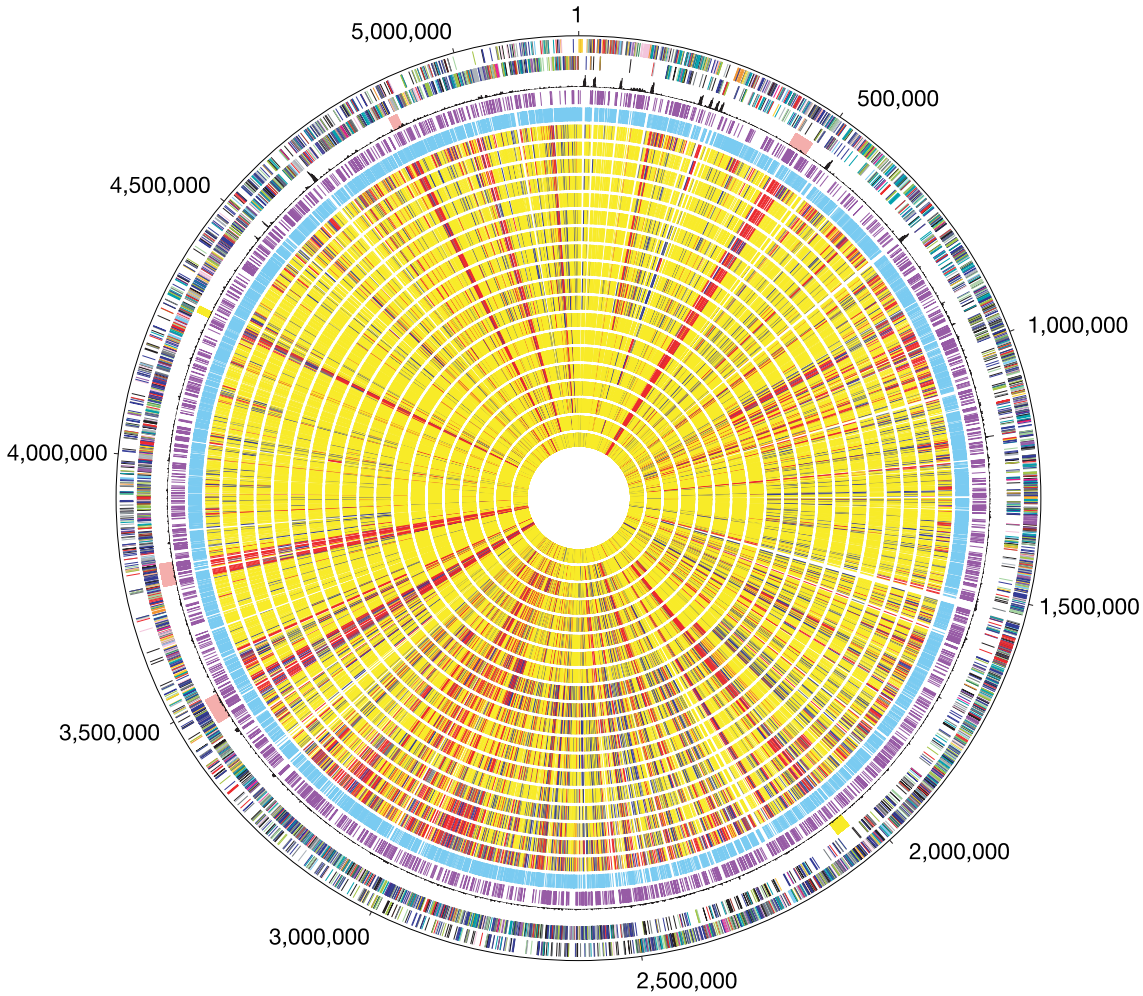

Figure 1 Circular representation of the *B. anthracis* chromosome and comparative genome hybridizations of *B. cereus* group strains. Outer circle, predicted coding regions on the plus strand colour-coded by role categories (see Supplementary Fig. 4). Circle 2, predicted coding regions on the minus strand colour-coded by role categories. Circle 3, atypical nucleotide composition curve. Salmon colour, phage regions; yellow, other unique regions located around positions 2.0 and 4.3 Mb (referred to as regions 5 and 6 in the text). Circle 4, genes not represented on the array. Circle 5, genes present on the array. Genes were classified into three groups: genes present in the query strain (shown yellow), genes absent in the query strain (red), and diverged genes (blue). Missing data are in grey. *B. cereus* group strains are displayed following the phylogeny of Fig. 2 (circle number, strain number): 6, *B.c.* 874; 7, *B.c.* 535; 8, *B.c.* 612; 9, *B.w.* 1143; 10, *B.t.* 248; 11, *B.t.* 442; 12, *B.c.* 14579; 13, *B.t.* 775; 14, *B.c.* 259; 15, *B.t.* 1031; 16, *B.t.* 251; 17, *B.c.* 607; 18, *B.c.* ATCC 10987; 19, *B.c.* 812; 20, *B.c.* 819; 21, *B.c.* 831; 22, *B.t.* 840; 23, *B.c.* 1123; 24, *B.c.* 816. Here we use *B.c*, *B.t.* and *B.w* to indicate *B. cereus*, *B. thuringiensis* and *Bacillus weihenstephanensis*, respectively.

B. anthracis compared with four in *B. subtilis*; and there are nine homologues of the BrnQ branched chain amino-acid transporter in *B. anthracis* and only two in *B. subtilis*. *B. anthracis* also has an expanded number of secreted proteases and peptidases relative to *B. subtilis* and a number of amino-acid utilization genes not found to date in other *Bacillus* genomes such as homogentisate dioxygenase (BA0242), involved in tyrosine degradation. Emphasizing the potential importance of peptides and amino acids for *B. anthracis* metabolism, there are six LysE/Rht amino-acid efflux systems compared with two in *B. subtilis*. These systems prevent accumulation of amino acids to bacteriostatic concentrations during growth on peptides[22]. *B. anthracis* may therefore be adapted for life in a protein-rich environment, such as decaying animal matter.

B. anthracis appears to have a reduced capacity for sugar utilization relative to *B. subtilis*. It lacks catabolic pathways for mannose, arabinose and rhamnose, and has reduced numbers of phosphotransferase systems and other types of sugar transporters. *B. anthracis* possesses genes for the cleavage of extracellular chitin and chitosan, and the utilization of N-acetylglucosamine constituents of these polymers. This may reflect some type of association with insects analogous to *B. thuringiensis*, or with polymers derived from plant or fungal material. *B. anthracis* contains a complete operon for polyester biosynthesis, which may function as an alternative energy storage compound for the organism. *B. anthracis* also has a multisubunit NADH hydrogenase not described before in Gram-positive bacteria.

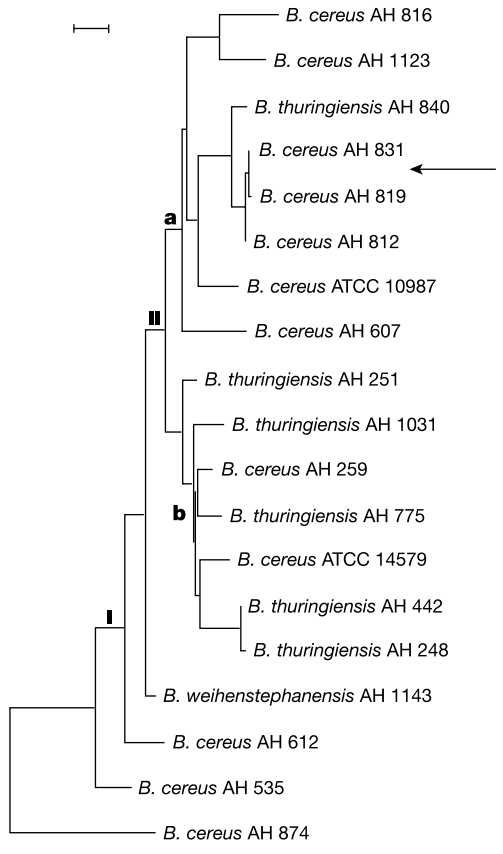

Figure 2 Phylogenetic relationships among 19 *B. cereus*/*B. thuringiensis* strains inferred from CGH results for 3,601 chromosomal *B. anthracis* genes. The tree was built by applying the neighbour-joining algorithm to a pairwise distance matrix of percentages of differences between the presence/absence patterns of all strains (diverged genes not taken into account). Similar trees were obtained using the maximum-parsimony method. The scale bar represents 2% divergence. The arrow indicates the position where *B. anthracis* would emerge by extrapolation from multilocus enzyme electrophoresis analysis[4].

B. anthracis possesses an expanded array of iron-acquisition genes compared to *B. subtilis* that may be important for iron scavenging in a mammalian host. These include 15 ABC uptake systems for iron siderophores or chelates, as well as two clusters of genes for the biosynthesis of siderophores. Two genes involved in synthesis of an aerobactin-like siderophore are not found in *B. subtilis* or the *B. cereus* ATCC 10987 sequence (BA1981, BA1982). Like *B. subtilis* and other soil bacteria, *B. anthracis* encodes a broad swathe of predicted drug efflux pumps, and a variety of other antibiotic-resistance genes are also present. However, it is unknown whether these contribute to resistance in a clinical setting[23].

We designed a *B. anthracis* DNA microarray on the basis of identifiable genes present at the conclusion of random phase sequencing. The microarray was used to compare *B. anthracis* to 19 members of the *B. cereus* group by comparative genome hybridization (CGH) (Fig. 1). Strains examined by CGH possessed 66–92% of their chromosomal genes in common with *B. anthracis*. Genes unique to *B. anthracis* in particular, and the *B. cereus* group in general, appear to be over-represented in the 2.0-Mb chromosomal region (coordinates 1,500,000 to 3,500,000 in Fig. 1) surrounding the presumed terminus of replication. Genome plasticity around the replication terminus has been seen in other comparisons of bacterial genomes[24]. Six smaller regions of the *B. anthracis* genome appeared to be absent from nearly all other *B. cereus* group strains tested (Fig. 1). Regions one to four correspond to the *B. anthracis* prophages (Supplementary Information), and region five centres on an IS110 family insertion element. Only region six does not bear obvious relationship to mobile elements. The magnitude of genomic variability based on CGH experiments comparing the *B. anthracis* Ames microarray and B. *cereus* group strains (Fig. 1) is 25–100 times greater than in similar experiments involving comparison of *B. anthracis* Ames to other *B. anthracis* strains (T. Blank and S.N.P., unpublished results). This reflects the very limited molecular diversity of the *B. anthracis* species[25].

Hybridization experiments indicate the presence of pXO1 homologues in half of the 19 strains examined (Supplementary Information), consistent with what has been shown in other studies[26]. Few genes from the pXO1 pathogenicity island, pXO1-96 to pXO1-127 (ref. 2), appeared to be present in the 19 *B. cereus* group strains. The toxin genes, central to anthrax aetiology, are found only in *B. anthracis* and not in any of the 19 *B. cereus* group strains sampled. In sharp contrast to pXO1, there were few pXO2 genes hybridizing with genomic DNA from the 19 *B. cereus* group bacteria.

Ratios obtained by CGH for chromosomal genes were used to infer the phylogenetic relationships among the *B. cereus* strains, producing three clusters (I, IIa and IIb; Fig. 2). The groupings were compatible with the phylogeny reconstructed using multilocus enzyme electrophoresis[4]. By extrapolation of the results from that study to the microarray-based phylogeny, *B. anthracis* would emerge within cluster IIa (arrow in Fig. 2). The presence of genes with pXO1 sequence identity in various branches covering all three clusters of the *B. cereus* group tree (Fig. 2), and the distribution of pXO1-like genes in the *B. cereus* group independent of the chromosomal relatedness among the strains (Supplementary Information), provides further evidence for mobility of pXO1 genes within the *B. cereus* group. Plasmid transfer within the *B. cereus* group is well established[27], and there are numerous mobility genes on pXO1[2]. Despite the evidence for genomic variability in the *B. cereus* group, the *B. anthracis* chromosome and virulence plasmids display little localized variation in G + C content and dinucleotide composition (generally associated with horizontally acquired genes from distantly related donors), suggesting that most genes are native to the *B. cereus* group.

The *B. anthracis* chromosome sequence portrays a soil-dwelling organism, possessing numerous potential virulence genes, which has possibly a preference for protein-rich environments. This is consistent with the evolution of *B. anthracis* from a *B. cereus* ancestor through acquisition of key plasmid-encoded toxin, capsule and regulatory loci. CGH data presented here demonstrate variability in plasmid gene content among the group as compared to chromosomal genes. Other major differences between *B. anthracis* and *B. cereus* may have been effected through altered gene expression rather than loss or gain of genes. Although both species contain genes associated with secreted proteases, haemolysins, extracellular chitinases[11], motility, tyrosine degradation and penicillin resistance[23], *B. anthracis* and *B. cereus* phenotypes differ with respect to the function of these genes. These changes in expression

Table 2 **Putative PlcR-regulated proteins**

Gene	Description	Motif-to-gene distance (nt)*	Signal peptide†
BA0401	Tellurium resistance protein	109	No
BA0400	Tellurium resistance protein		No
BA0399	Tellurium resistance protein, putative		No
BA0575	Methyl-accepting chemotaxis protein	101	Yes
BA0969	Hypothetical protein	55	No
BA0975	HD domain protein	143	No
BA0976	Hypothetical protein	219	No
BA0977	Hypothetical protein	157	Yes
BA1086	Sugar binding transcriptional regulator LacI family	73	No
BA1085	Acetyltransferase, GNAT family		No
BA1424	Histidyl-tRNA synthetase, putative	721	No
BA1470	Membrane protein, putative	747	No
BA1692	Conserved hypothetical protein	171	No
BA1888	Enterotoxin	520	Yes
BA1887	Enterotoxin		Yes
BA1889	Enterotoxin (point mutation)		Yes
BA2147	ScdA protein	16	No
BA2148	Hypothetical protein	268	No
BA2499	Hypothetical protein	123	Yes
BA2730	Neutral protease	114	Yes
BA3355	Thiol-activated cytolysin	248	Yes
BA3356	Membrane protein, putative	159	No
BA3370	Ribonuclease	269	Yes
BA3491	Conserved hypothetical protein	144	No
BA3635	Spore germination protein	36	Yes
BA3634	Spore germination protein		Yes
BA3633	Spore germination protein		No
BA3891	1-Phosphotydlinositol phosphodiesterase	106	Yes
BA3890	Hypothetical protein		No
BA3892	Serine protease, subtilase family	149	Yes
BA3893	Cell wall hydrolase, putative	146	Yes
BA4745	ABC transporter, ATP-binding protein	286	No
BA4744	Membrane protein, putative		Yes
BA4743	Rrf2 protein family protein, putative		No
BA4746	Acid phosphatase	58	Yes
BA4949	Metallo-β-lactamase family protein	84	No
BA5055	Conserved domain protein	194	No
BA5190	Acetyltransferase GNAT family	93	No
BA5191	NAD(P)H dehydrogenase, quinone family	518	No
BA5231	Hypothetical protein	14	No
BA5230	Hypothetical protein		No
BA5243	CAAX amino terminal protease family protein	139	Yes
BA5595	Transcriptional regulator plcR-related, putative, authentic point mutation	102	No
BA5594	plcR-associated protein		Yes
BA5596	Hypothetical protein	313	No
BA5605	Hypothetical protein	113	No
BA5606	Aminopeptidase, putative	113	Yes
BA5701	Channel protein, haemolysin III family	6	Yes
BA5700	UDP-galactose-4-epimerase		No
BA5702	Conserved hypothetical protein	179	No

*Distance from the 5′ end of the motif to the first nucleotide in the first gene of the operon. Genes in the same block are in the same predicted operon (that is, same transcription orientation and with no intervening rho-dependent terminator).
†See Supplementary Methods.

may reflect recent adaptations following acquisition of the pathogenicity island that contains the lethal toxin loci on pXO1. The *atxA* regulatory gene in this region controls toxin gene expression but is incompatible with the chromosomal regulator *plcR*, found in *B. cereus*[19]. The worldwide, near-clonal spread of the organism[25] suggests that expression of the toxin and capsule genes confers an advantage to *B. anthracis* that outweighs changes in the chromosomal gene expression. Findings from this genome sequence analysis raise further questions about the biology of *B. anthracis*; for instance, what are the roles of putative 'virulence' genes in close relatives of *B. anthracis* that do not cause anthrax, and do they actually contribute to virulence in *B. anthracis*? □

Methods

Genome sequencing and analysis of *B. anthracis* Ames (pXO1$^-$ pXO2$^-$)

B. anthracis Ames was cured of plasmid pXO1 by incubation at 43 °C and pXO2 subsequently cured by novobiocin treatment (Supplementary Methods). As previously described, the chromosome was sequenced using two DNA preparations. For the first (Porton1)[5], 2–3 kilobase (kb) and 4–7 kb random insert libraries in plasmid-derived vectors were constructed and end-sequenced following the standard strategy for TIGR microbial shotgun projects[6], achieving success rates of 74% and 64% and average high-quality read lengths of 559 nucleotides (nt) and 586 nt, respectively. For the Porton2 preparation[5], libraries of 2–3 kb and 6–8 kb were constructed with success rates of 89% and 85% and average high-quality read lengths of 609 nt and 645 nt. The completed chromosome sequence consisted of 73,806 and 6,052 reads from the Porton1 small and large insert libraries, and 3,532 and 32,430 from the Porton2 small and large insert libraries—achieving an average of 13-fold sequence coverage per base. After assembly, gaps between contigs were closed by editing, walking library clones, and linking assemblies by polymerase chain reaction (PCR). The Glimmer gene finder[28] was modified by enhancing its model of noncoding sequences. This improved its ability to exclude short open reading frames (ORFs), and substantially reduced the number of predicted small hypothetical proteins. Annotation was as described for a previous project[6]. BLASTP[10] was used for comparisons of the protein sets of *B. anthracis*, *B. cereus* ATCC 10987 (T.D.R., unpublished results; http://www.tigr.org/tdb/ufmg/) and other complete bacterial genomes (http://www.tigr.org/cmr2/ and http://genolist.pasteur.fr/Subtilist/). A predicted probability score of less than 10^{-5} was used as a standard cut-off to define a likely match.

DNA microarray preparation and analysis

Amplicons representing 79 of 217 and 41 of 122 genes from pXO1 and pXO2 respectively, and 3,601 of 5,753 chromosomal genes as predicted by Glimmer[28] (see Supplementary Methods) were arrayed onto glass microscope slides (Telechem Inc.). Redundant genes were generally represented once or a few times on the array. Genomic DNA was labelled with Cy3 and Cy5 according to J. DeRisi (http://www.microarrays.org/Pdfs/GenomicDNALabel_B.pdf), except that genomic DNA was not digested or sheared before labelling. Arrays were scanned with a GenePix 4000B scanner (Axon Inc.). Hybridization signals were quantified using TIGR SPOTFINDER (software available at http://www.tigr.org/softlab). Hybridization experiments were competitive using probes derived from *B. anthracis* Ames (reference) and a *B. cereus* group (query) strain. Normalized signal intensities were used to generate relative hybridization ratios (query/reference). Data representing weak signal were removed. The ratios from a maximum of six data points (duplicate spots, hybridizations performed in triplicate) were placed in three bins: <0.1, gene is absent in query strain; 0.1–0.3, present but diverged in query strain; and >0.3, gene is present in the query strain. A majority rule was applied to the data for binning such that more than 50% of ratios were in agreement as to assignment and that at least two data points were used (exceeded in 99% of the cases). In cases where less than two data points existed, the gene was treated as data missing.

The criteria for the numerical ranges of our bins were established in two ways. First, we determined the presence or absence of sequences homologous to 3,601 *B. anthracis* genes in the sequence of *B. cereus* ATCC 14579 (Integrated Genomics Inc.; http://www.integratedgenomics.com/) using BLASTN[10], and compared that to the assignments inferred from hybridization ratios. A threshold of 0.1 was found to be suitable for classifying a gene as absent (that is, agreement between sequence and CGH data in 99% of the cases), while a cut-off value of 0.3 was conservative for gene presence (agreement in 92% of the cases). Second, we used a set of 65 genes conserved in 26 bacterial genomes, NCBI COG database (http://www.ncbi.nlm.nih.gov/COG/). Genes judged as present in query strains using our selected cut-offs correctly binned data in 1,225 out of 1,235 total calls. There was a tendency for underprediction of plasmid homologues by CGH, when compared to results from the sequence analysis. Two possible explanations for this are variability in plasmid copy number in *B. cereus* strains relative to *B. anthracis1* and/or that the average divergence of plasmid genes is greater than chromosomal genes.

Other techniques and analysis

PCR amplification for microarray spotting, pulsed field gel electrophoresis and Southern blotting are described in Supplementary Information, as is the phylogenetic analysis of chromosomal data.

Received 14 August 2002; accepted 28 March 2003; doi:10.1038/nature01586.

1. Dixon, T. C., Meselson, M., Guillemin, J. & Hanna, P. C. Anthrax. *N. Engl. J. Med.* **341,** 815–826 (1999).
2. Okinaka, R. T. *et al.* Sequence and organization of pXO1, the large *Bacillus anthracis* plasmid harboring the anthrax toxin genes. *J. Bacteriol.* **181,** 6509–6515 (1999).
3. Okinaka, R. *et al.* Sequence, assembly and analysis of pXO1 and pXO2. *J. Appl. Microbiol.* **87,** 261–262 (1999).
4. Helgason, E. *et al. Bacillus anthracis, Bacillus cereus,* and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* **66,** 2627–2630 (2000).
5. Read, T. D. *et al.* Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296,** 2028–2033 (2002).
6. Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293,** 498–506 (2001).
7. Kunst, F. *et al.* The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–256 (1997).
8. Nolling, J. *et al.* Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183,** 4823–4838 (2001).
9. Ko, M., Choi, H. & Park, C. Group I self-splicing intron in the recA gene of *Bacillus anthracis*. *J. Bacteriol.* **184,** 3917–3922 (2002).
10. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
11. Guttmann, D. M. & Ellar, D. J. Phenotypic and genotypic comparisons of 23 strains from the *Bacillus cereus* complex for a selection of known and putative B. *thuringiensis* virulence factors. *FEMS Microbiol. Lett.* **188,** 7–13 (2000).
12. Cossart, P. Molecular and cellular basis of the infection by *Listeria monocytogenes*: an overview. *Int. J. Med. Microbiol.* **291,** 401–409 (2002).
13. Lepore, L. S., Roelvink, P. R. & Granados, R. R. Enhancin, the granulosis virus protein that facilitates nucleopolyhedrovirus (NPV) infections, is a metalloprotease. *J. Invertebr. Pathol.* **68,** 131–140 (1996).
14. Parkhill, J. *et al.* Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature* **413,** 523–527 (2001).
15. McCann, K. P., Robinson, C., Sammons, R. L., Smith, D. A. & Corfe, B. M. Alanine germination receptors of *Bacillus subtilis*. *Lett. Appl. Microbiol.* **23,** 290–294 (1996).
16. De Groote, M. A. *et al.* Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proc. Natl Acad. Sci. USA* **94,** 13997–14001 (1997).
17. West, J. T., Estacio, W. & Marquez-Magana, L. Relative roles of the *fla/che* P(A), P(D-3), and P(*sigD*) promoters in regulating motility and *sigD* expression in *Bacillus subtilis*. *J. Bacteriol.* **182,** 4841–4848 (2000).
18. Glaser, P. *et al.* Comparative genomics of *Listeria* species. *Science* **294,** 849–852 (2001).
19. Mignot, T. *et al.* The incompatibility between the PlcR- and AtxA-controlled regulons may have selected a nonsense mutation in *Bacillus anthracis*. *Mol. Microbiol.* **42,** 1189–1198 (2001).
20. Agaisse, H., Gominet, M., Okstad, O. A., Kolsto, A. B. & Lereclus, D. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis*. *Mol. Microbiol.* **32,** 1043–1053 (1999).
21. Pallen, M. J., Lam, A. C., Antonio, M. & Dunbar, K. An embarrassment of sortases—a richness of substrates? *Trends Microbiol.* **9,** 97–102 (2001).
22. Bellmann, A. *et al.* Expression control and specificity of the basic amino acid exporter LysE of *Corynebacterium glutamicum*. *Microbiology* **147,** 1765–1774 (2001).
23. Turnbull, P. C. Definitive identification of *Bacillus anthracis*—a review. *J. Appl. Microbiol.* **87,** 237–240 (1999).
24. Suyama, M. & Bork, P. Evolution of prokaryotic gene order: genome rearrangements in closely related species. *Trends Genet.* **17,** 10–13 (2001).
25. Keim, P. *et al.* Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* **182,** 2928–2936 (2000).
26. Pannucci, J., Okinaka, R. T., Sabin, R. & Kuske, C. R. Bacillus anthracis pXO1 plasmid sequence conservation among closely related bacterial species. *J. Bacteriol.* **184,** 134–141 (2002).
27. Thorne, C. B. Bacillus subtilis *and Other Gram-positive Bacteria* 113–124 (American Society for Microbiology, Washington DC, 1993).
28. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26,** 544–548 (1998).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We acknowledge the contributions of P. Turnbull, E. Saile, Y. Chen, J. Hunter-Cevera, N. McKinney, S. Cendrowski, M. Weiner, A. Fouet, A. Harrison, S. Leppla, M. Mock, C. Moran, G. Myers, G. Patra, J. Ravel, E. Reilly and T. Torok. The *B. anthracis* chromosome sequence was supported by funding from the Office of Naval Research (ONR), National Institutes of Allergy and Infectious Disease (NIAID), the Department of Energy and the UK Defence Science Technology Laboratory. Comparative genome hybridization experiments were supported by the ONR. A-B.K, N.T, O.A.O. and E.H were supported by the Norwegian Research Council. The sequencing of *B. cereus* ATCC 10987 was supported by the NIAID under the Pathogen Functional Genomics Resource Center contract, and the ONR.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.D.R. (anthrax@tigr.org). The *B. anthracis* genome sequence has been deposited at GenBank under accession number AE016879; the microarray data have been deposited at Gene Expression Omnibus (GEO) under accession number GSE341. The *B. cereus* 10987 unfinished genome sequence (GenBank accession number NC_003909) is available at http://www.tigr.org/tdb/ufmg).