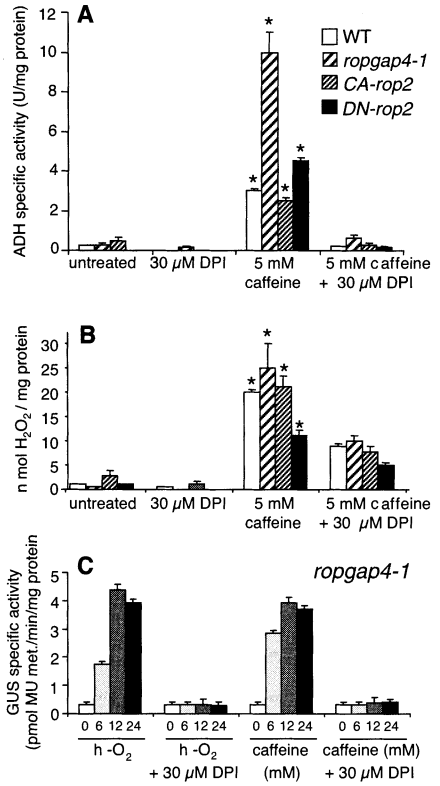

Fig. 3. ADH activity is stimulated by caffeine treatment via Rop-induced and DPI-sensitive H_2O_2 production. (**A**) ADH specific activity in seedlings treated with caffeine and/or DPI for 24 hours. (**B**) H_2O_2 levels in seedlings analyzed in (A). Values are mean ± SE of three independent experiments. Asterisk indicates significant difference from the maximal level detected after O_2 deprivation ($P < 0.01$; Student's t test). (**C**) GUS specific activity in *ropgap4-1* seedlings after O_2 deprivation, caffeine treatment, and DPI treatment. Values are mean ± SE of three independent experiments.

(Figs. 1A and 3C). (iii) Application of a H_2O_2 regenerating system elevated GUS activity in *ropgap4-1* seedlings (Fig. 4B). (iv) *RopGAP4* mRNA levels were constitutively elevated in *CA-rop2* seedlings (Fig. 1C).

Thus, a Rop rheostat regulates the production of H_2O_2 that is required to trigger the expression of beneficial genes (for example, *ADH*) and the avoidance of H_2O_2-induced cell death. Rop signaling is controlled by negative feedback regulation through the stimulation of *RopGAP4* transcription by H_2O_2. The termination of Rop signaling by RopGAP4 would alleviate oxidative stress and limit consumption of carbohydrate reserves via glycolysis and ethanolic fermentation. The reduced O_2 deprivation tolerance of the *DN-rop2*, *CA-rop2*, and *ropgap4-1* seedlings underscores the requirement for a fully functional Rop rheostat. We propose that a Rop rheostat is critical to developmental processes and environmental stress responses that use H_2O_2 as a second messenger or enhance H_2O_2 accumulation, including the response to abscisic acid, auxin, pathogen infection, and numerous abiotic stresses. Manipulation of the Rop signal transduction rheostat may enhance the productivity of crops that undergo transient submergence or soil waterlogging.

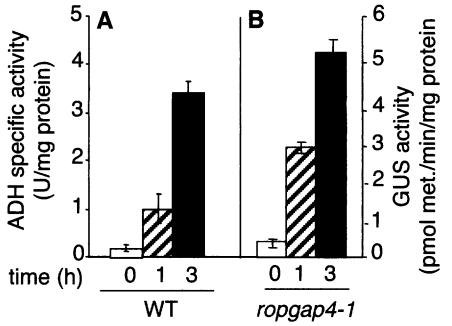

Fig. 4. ADH activity and *ROPGAP4* expression is induced by a H_2O_2-regenerating system. ADH specific activity in WT (**A**) and GUS specific activity in *ropgap4-1* seedlings (**B**) treated with glucose and glucose oxidase for up to 3 hours.

References and Notes

1. M. C. Drew, *Annu. Rev. Plant Physiol. Mol. Biol.* **48**, 223 (1997).
2. E. S. Dennis *et al.*, *J. Exp. Bot.* **51**, 89 (2000).
3. J. C. Sedbrook, P. J. Kronebusch, C. G. Borisy, A. J. Trewavas, P. H. Masson, *Plant Physiol.* **111**, 243 (1996).
4. C. C. Subbaiah, J. K. Zhang, M. M. Sachs, *Plant Physiol.* **105**, 369 (1994).
5. C. C. Subbaiah, D. S. Bush, M. M. Sachs, *Plant Cell* **6**, 1747 (1994).
6. H. J. Chung, R. J. Ferl, *Plant Physiol.* **121**, 429 (1999).
7. V. Sundaresan *et al.*, *Genes Dev.* **9**, 1797 (1995).
8. A. Baxter-Burrell, R. Chang, P. Springer, J. Bailey-Serres, *Ann. Bot.* **90**, in press.
9. G. Wu, H. Li, Z. B. Yang, *Plant Physiol.* **124**, 1625 (2000).
10. Y. K. Lin, Z. B. Yang, *Plant Cell* **9**, 1647 (1997).
11. H. Li, G. Wu, D. Ware, K. R. Davis, Z. B. Yang, *Plant Physiol.* **118**, 407 (1998).
12. H. Li, Y. K. Lin, R. M. Heath, M. X. Zhu, Z. B. Yang, *Plant Cell* **11**, 1731 (1999).
13. B. Kost, J. Mathur, N. H. Chua, *Curr. Opin. Plant Biol.* **2**, 462 (1999).
14. Z. L. Zheng, Z. B. Yang, *Plant Mol. Biol.* **44**, 1 (2000).
15. ———, *Trends Plant Sci.* **5**, 298 (2000).
16. Y. Fu, G. Wu, Z. Yang, *J. Cell Biol.* **152**, 1019 (2001).
17. A. J. Molendijk *et al.*, *EMBO J.* **20**, 2779 (2001).
18. J. Park, *et al.*, *Plant Physiol.* **124**, 725 (2000).
19. T. S. Potikha, C. C. Collins, D. I. Johnson, D. P. Delmer, A. Levine, *Plant Physiol.* **119**, 849 (1999).
20. T. Kawasaki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10922 (1999).
21. E. Lemichez *et al.*, *Genes Dev.* **15**, 1808 (2001).
22. H. Li, J. J. Shen, Z. L. Zheng, Y. K. Lin, Z. B. Yang, *Plant Physiol.* **126**, 670 (2001).
23. M. Jacobs, R. Dolferus, V. B. Van Den Bosshe, *Biochem. Genet.* **26**, 102 (1988).
24. G. Wu, Y. Gu, S. Li, Z. Yang, *Plant Cell* **13**, 2841 (2001).
25. P. G. Heyworth *et al.*, *Mol. Biol. Cell* **4**, 261 (1993).
26. M. Sagi, R. Fluhr, *Plant Physiol.* **126**, 1281 (2001).
27. I. M. Moller, *Annu. Rev. Plant Physiol.* **52**, 561 (2001).

28. Funded by grants from the U.S. Department of Agriculture (97-35100-419 to J.B.S.; 01-35304-09894 to Z.Y.), Southwest Consortium (99-N02 to P.S. and J.B.S.) and Department of Energy (DE-7G03-00ER15060 to Z.Y.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/vol/296/5575/2026/DC1
Materials and Methods
Fig. S1
References and Notes

4 March 2002; accepted 1 May 2002

Comparative Genome Sequencing for Discovery of Novel Polymorphisms in *Bacillus anthracis*

Timothy D. Read,[1] Steven L. Salzberg,[1] Mihai Pop,[1] Martin Shumway,[1] Lowell Umayam,[1] Lingxia Jiang,[1] Erik Holtzapple,[1] Joseph D. Busch,[2] Kimothy L. Smith,[2] James M. Schupp,[2] Daniel Solomon,[2] Paul Keim,[2] Claire M. Fraser[1]*

Comparison of the whole-genome sequence of *Bacillus anthracis* isolated from a victim of a recent bioterrorist anthrax attack with a reference reveals 60 new markers that include single nucleotide polymorphisms (SNPs), inserted or deleted sequences, and tandem repeats. Genome comparison detected four high-quality SNPs between the two sequenced *B. anthracis* chromosomes and seven differences among different preparations of the reference genome. These markers have been tested on a collection of anthrax isolates and were found to divide these samples into distinct families. These results demonstrate that genome-based analysis of microbial pathogens will provide a powerful new tool for investigation of infectious disease outbreaks.

On 4 October 2001, the Centers for Disease Control reported a highly unusual case of inhalational anthrax in a photo editor at a West Palm Beach, Florida, media organization (*1*). This turned out to be the first in a series of letter-based attacks over several weeks. The attacks resulted in five fatalities (including the first-diagnosed victim) and

several cases of severe inhalational anthrax. This wave of bioterrorism caused several billion dollars of loss to the U.S. economy and inflicted further anxiety on a population suffering after the World Trade Center tragedy.

Bacillus anthracis, the causative agent of anthrax, is a highly monomorphic species with genes from different isolates that typically have greater than 99% nucleotide sequence identity (*2*). The most accurate strain-typing tool available examines copy-number modulation of variable-number tandem repeat (VNTR) markers (*3*). VNTR analysis of the Florida, New York, and Washington, DC, isolates used in the letter attacks suggested that they were from the same source (*4*), namely, a strain originally isolated from a dead cow in Texas in 1981. This strain of *B. anthracis*, designated Ames, was subsequently sent to the U.S. Army Medical Research Institute (USAMRIID) in Fort Detrick, Maryland, where it was actively used in the U.S. defensive biological weapons program and also provided to other research laboratories in the United States and Europe (Fig. 1).

Though VNTR analysis examines genomic regions known to be highly variable, this method typically samples only a limited number of loci. The availability at TIGR (The Institute for Genomic Research) of a nearly complete genome sequence of a *B. anthracis* Ames isolate (the Porton isolate) that lacks the pXO1 and pXO2 virulence plasmids (*5*) provided an opportunity to use a comparative genomics approach to identify potential polymorphic sites in this genome (*6*).

The assembled sequences from the Florida isolate were aligned to the Porton chromosome and to previously sequenced *B. anthracis* plasmids (*6*) by MUMmer (*7*). The index pXO1 plasmid [181,654 base pairs (bp)] is from the Sterne strain, and the pXO2 plasmid (96,231 bp) is from the Pasteur strain (*8*, *9*). Alignments were processed further to identify three types of genetic differences: (i) SNPs, (ii) VNTRs, and (iii) inserted or deleted sequences (indels). These differences were compiled, and the underlying chromatogram files were examined. The average level of sequence coverage for the main Florida chromosome was 6-fold (i.e., each position in the genome was contained, on average, in six independent sequencing reads). The coverages for the pXO1 and pXO2 virulence plasmids were approximately 20-fold and 14-fold, respectively. This level of coverage suggests a molecular ratio of 3:2:1 for pXO1: pXO2:chromosome in the Florida isolate of *B. anthracis*.

The accuracy of any base pair in the assembled genome can be computed as a function of the accuracy of each of the underlying sequences. Because all sequence reads are independent samples of the genome, the probabilities derived from each read can be multiplied together to obtain a joint probability that the consensus sequence is correct. Base-calling software assigns an error probability to each base pair in a sequence read (*10*–*12*).

TIGR's error rate for finished genomes has been independently measured at less than 1 error in 88,000 nucleotides (*13*). At this rate, the 5.2-Mbp *B. anthracis* chromosome would be expected to have approximately 60 sequencing errors, and a comparison of the two isolates would be expected to yield 120 differences. The error rate in DNA sequencing can be dramatically reduced by eliminating regions of low sequence coverage, because error rate is not constant across the entire genome sequence or even across a single read (*14*). To limit the number of false positives due to sequencing errors, the comparative analysis in this study used only regions with at least 3-fold coverage in both the Florida and Porton sequences; within each

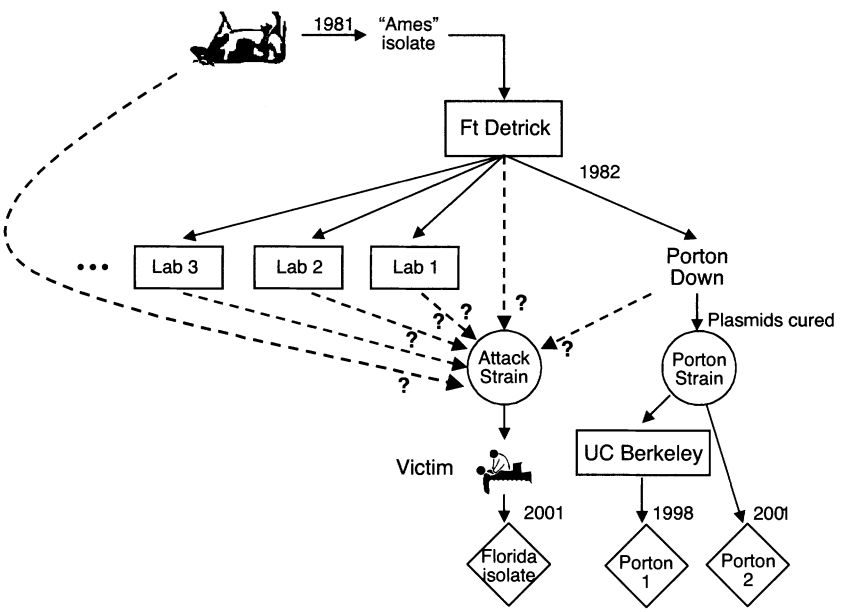

Fig. 1. Suggested relationship of Ames isolates. Known direct transfers of the isolates (and hence generations of growth) are shown as solid arrows. Diamond boxes are the sources of DNA for genomic sequencing. Hypothetical lines of descent are shown as dashed arrows. Porton1 and Porton2 are different cultures of the Porton Ames isolate (*5*).

Table 1. Polymorphisms between main chromosomes of Porton and Florida isolates of *B. anthracis*, Ames strain. Porton1 and Porton2 are derived from different cultures of the Porton isolate (*5*). CI-x indicates a chromosomal indel. X indicates depth of coverage. The probability that the polymorphism was correctly sequenced is computed as described and shown under *P*. NS indicates lack of evidence from particular isolate (i.e., all reads were from the other isolate). Not shown in those tables are more than 150 apparent SNPs in regions of low coverage; on the basis of statistical analysis, it is estimated that no more than three of these represent real differences.

Polymorphism ID	Porton1		Porton2		Florida isolate		*P*
	Genotype	X	Genotype	X	Genotype	X	
CS-1	A	8	A	8	G	3	$10^{-9.3}$
CS-2	ATATAT	5	NS		ATAT	6	10^{-13}
CS-3	NS		A	7	G	2	$10^{-7.2}$
CI-1	AAAA	14	AAAA	3	AAAAA	4	10^{-14}
	Intrareference strain polymorphisms						
CS-4	C	4	T	5	C	7	10^{-16}
CS-5	A	10	G	5	A	9	10^{-16}
CS-6	A	8	G	4	A	2	10^{-13}
CI-2	AAAAAAA	2	AAAAAA	5	AAAAAAA	5	10^{-16}
CS-7	T	14	G	5	G	7	10^{-25}
CS-8	T	5	G	7	G	3	10^{-15}
CI-3	AA	14	AAA	4	AAA	7	10^{-21}

[1]The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. [2]Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ 86011, USA.

*To whom correspondence should be addressed. E-mail: anthrax@tigr.org

isolate, only positions for which the sequence chromatograms agreed with each other were used. More precise estimates of accuracy were computed for each of the polymorphisms described (*12*).

Only four differences were discovered between the main chromosomes of the Florida and Porton isolates (Table 1). Two of these are SNPs and two are short indels. CS-2 (Table 1) is an AT indel that was seen only in particular clones from the Porton1 library. Seven additional polymorphisms were discovered within the Porton isolate chromosome. These differences can be explained by the fact that TIGR received DNA from two different cultures of the same isolate (*5*), separated chronologically by about 3 years.

Because the Porton isolate was cured of the pXO plasmids, we were only able to compare the Florida plasmids against the *B. anthracis* Sterne strain pXO1 (*8*) and Pasteur strain pXO2 (*9*). A total of 38 SNPs, 8 VNTRs, and 3 large indels (Table 2) were found. Fewer plasmid SNPs occur in pXO1 than in pXO2 (15 versus 23), though pXO1 is twice as large as pXO2. However, this data cannot be used to infer a higher mutation rate in pXO2, because the Ames strain is more closely related to Sterne than to the Pasteur strain (*3*).

Larger polymorphisms included both VNTRs (Table 3) and simple indels (Table 2). The 135-bp inserted sequence in pXO1, IX1-2, is homologous to a region from *Bacillus thuringiensis* (*15*), demonstrating that the Sterne lineage has lost this sequence. The 85-bp indel in pXO1, IX1-1, has 84% nucleotide identity (and identical flanking repeats) to a DNA sequence from *Bacillus cereus* [American Type Culture Collection (ATCC) 10987; T. Read, unpublished data] and is deleted in all but one of the Ames isolates tested (Table 4). This indicates that IX1-1 has been lost very recently within the Ames lineage. The conservation of as many as half of the pXO1 plasmid genes at greater than 40% nucleotide similarity in some *B. cereus* and *B. thuringiensis* strains has been reported in recent hybridization studies (*15*, *16*).

The largest difference is the 1.4-kbp IX2-7 indel, which may have been lost from the Ames pXO2 plasmid through recombination, as suggested by two flanking copies of a 143-bp direct repeat (*17*). Both IX1-1 and IX1-2 are also flanked by much shorter (9 to 10 bp) direct repeats. Among the VNTRs, VX2-1 has a more complex structure. This repeat has one extra copy in the Pasteur strain, but both strains contain another three tandem copies of the repeat approximately 100-bp downstream from the first set. The polymorphic variation occurs only in the first set of copies. VX2-1 and VX2-4 both contain an additional, inverted copy of the repeat adjacent to the tandem direct repeats. VX1-4 and VX2-5, the short TTA and AT repeat polymorphisms, were known previously (*3*); the remaining nine VNTRs and indels are novel.

SNP PS-33 from pXO2 occurs within a 42-bp repeat that occurs five times, in four variations. If we label these variants A to D, then the order of the repeats is ABCBD in the Florida isolate and ABCDD in the Sterne strain, with the SNP occurring in the fourth copy of the repeat.

In addition to the SNPs and VNTRs detected in this analysis, we identified two large inversions in pXO1 of the Florida isolate in relation to the previously sequenced Sterne strain. The largest (44.8 kbp) occurs between coordinates 117,178 and 162,008 (using the Sterne strain coordinates). It is flanked by inverted copies of an IS1627 sequence and is centered on the pXO1 "pathogenicity island" (*11*), which includes the genes for the tripartite lethal factor toxin (*18*). Inversion of the pathogenicity island has been described (*19*); our sequence coverage data and polymerase chain reaction (PCR) amplification across the junctions show that the Florida isolate contains a mixture of both orientations.

The second pXO1 inversion event occurs between coordinates 43,233 and 48,988 (using the Sterne strain coordinates). This 5755-bp region, flanked by two perfect 929-bp inverted repeats of degenerate transposase genes, contains genes pXO1-35 through pXO1-40 (*8*), all of which encode proteins of unknown function. PCR reactions across the junctions of this region indicate that the Florida isolate also contains plasmids with both orientations of this region. The inversions detected in this comparative analysis provide another marker for genotyping anthrax strains. In both cases, inversions appear to have been the result of recombination between nearly identical, more than 900-bp inverted segments of transposase genes.

Table 2. Insertions distinguishing the plasmids of the Florida isolate of *B. anthracis* from the plasmids in the Sterne and Pasteur strains. IX*y-z* indicates indel in plasmid P^{XO}*y*. X indicates depth of coverage. Repeat units are indicated in brackets. For IX1-1 and IX1-2, which are long insertions in pXO1, only the flanking 10- and 9-bp repeats are shown, with the length of the omitted sequence indicated by brackets (e g., 125 bp). Supporting online material includes a 1000-bp region centered on each of these polymorphisms.

ID	Strain	Repeat configuration	X
		pXO1	
IX1-1	Florida	[ATTCATCTC]	25
	Sterne	[ATTCATCTC] [76 bp] [ATTCATCTC]	
IX1-2	Florida	[ACATGTTTCT] [125 bp] [ACATGTTTCT]	27
	Sterne	[ACATGTTTCT]	
		pXO2	
IX2-7	Florida	[143 bp]	15
	Pasteur	[143 bp] [1.4 kbp] [143 bp]	

Table 3. VNTRs distinguishing the plasmids of the Florida isolate of *B. anthracis* from the plasmids in the Sterne and Pasteur strains. VX*y-z* indicates VNTR number *z* in plasmid pXO*y*. X indicates depth of coverage. Repeat units are indicated in brackets, and superscripts denote fractional copy number (e.g., VX1-4 consists of seven copies of the sequence TTA in one strain and nine copies in the other). For VX2-1, the 3.75 copy number refers to an additional copy containing only six of the eight bases. The probability that the polymorphism was correctly sequenced exceeds 10^{-97} for all entries. Supporting online material includes a 1000-bp region centered on each of these polymorphisms.

ID	Strain	Repeat configuration	X
		pXO1	
VX1-3	Florida	[AATAAATATATAATAGTTAA]^{1}	19
	Sterne	[AATAAATATATAATAGTTAA]^{2}	
VX1-4	Florida	[TTA]^{7}	9
	Sterne	[TTA]^{9}	
		pX02	
VX2-1	Florida	[ACATTTAC]$^{3.75}$	9
	Pasteur	[ACATTTAC]$^{4.75}$	
VX2-2	Florida	[ATTAA]^{4}	12
	Pasteur	[ATTAA]^{3}	
VX2-3	Florida	[A]^{35}	4
	Pasteur	[A]^{24}	
VX2-4	Florida	[TGTATCTT]^{3}	6
	Pasteur	{TGTATCTT]^{4}	
VS2-5	Florida	[AT]^{9}	19
	Pasteur	[AT]^{7}	
VX2-6	Florida	[ATTTTTAT]^{4}	4
	Pasteur	[ATTTTTAT]^{3}	

Table 4. Distribution of the polymorphisms presented in *B. anthracis* isolates. Primers were designed to sequences flanking SNPs and indels and were used to amplify from a range of isolates. VNTRs were analyzed by a direct sizing method (*3*). For reference, the sequence of the Florida isolate is shown under "Florida SEQ." The sequence of the locus for each isolate is shown. Entries not marked by an asterisk were not tested because the plasmid was missing from that strain. NS, no sequence of sufficient quality from the reaction available. Δ indicates that an indel was deleted from the sequence. The values in the VNTR columns are numbers of units. For VX2-1, the 3.75 copy number refers to an additional copy containing only six of the eight bases. Isolates A to D, 1925 Iowa cow, 2001 California cow, and 1997 Texas goat are described in the text.

Polymorphism ID	Florida SEQ	Other isolates						
		A	B	C	D	1997 TX goat	1925 IA cow	2001 CA cow
				Chromosome polymorphisms				
CS-1	G	G	G	G	G	G	G	G
CS-2	ATAT	NS	NS	NS	NS	NS	NS	NS
CS-3	G	NS	G	G	G	G	G	G
CI-1	AAAAA	NS	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA	AAAAA
				PXO1 VNTRs and insertions (versus Sterne)				
IX1-1	Δ		Δ	Δ	Δ	85-bp insert*		85-bp insert*
IX1-2	135-bp insert		135-bp insert	135-bp insert	135-bp insert	135-bp insert		135-bp insert
VX1-3	1		1	1	1	1		1
VX1-4	7		7	7	7	7		7
				pXO1 SNPs (versus Sterne)				
PS-1	C		C	C	C	C		T*
PS-2	C		C	C	C	C		C
PS-4	TATA		TATA	TATA	TATA	TATA		TATA
PS-5	C		C	C	C	NS		C
PS-21	AA		AA	AA	AA	AA		AA
PS-23	C		C	C	C	C		C
PS-24	G		G	G	G	G		G
PS-25	G		G	G	G	G		G
PS-26	G		G	NS	G	G		G
PS-27	A		NS	NS	NS	NS		NS
PS-28	T		T	T	T	T		T
PS-29	G		G	G	G	G		G
PS-30			A	A	A	A		A
PS-31	T		T	T	T	T		T
PS-32	G		G	G	G	G		A*
				pXO2 VNTRs and insertions (versus Pasteur)				
VX2-1	3.75	3.75	3.75	3.75	3.75	2.75*	4.75*	11.75*
VX2-2	4	4	4	4	4	4	4	4
VX2-3	35	37*	35	35	36*†	30*	14*	6*
VX2-4	3	3	3	3	3	3	4*	4*
VX2-5	9	9	9	9	9	9	7*	7*
VX2-6	4	4	4	4	4	4	3*	2*
IX2-7	NS	NS	NS	NS	NS	NS	NS	NS
				pXO2 SNPs (versus Pasteur)				
PI-I	T	T	T	T	T	T	T	T
PI-2	A	A	A	A	A	A	A	A
PS-33	C	C	C	C	C	C	T*	T*
PS-34	T	T	T	T	T	T	G*	G*
PS-35	T	T	T	T	T	T	T	T
PS-36	G	G	G	G	G	G	G	G
PS-37	A	A	A	A	A	NS	A	A
PS-38	G	G	G	G	G	G	G	G
PS-39	T	T	T	T	T	T	C*	C*
PS-40	C	C	C	C	C	C	T*	T*
PS-41	G	G	G	G	G	G	G	NS
PS-42	C	C	C	C	C	C	C	C
PS-43	A	A	A	A	A	A	A	A
PS-44	G	NS	NS	NS	NS	NS	NS	NS
PS-45	A	A	A	A	A	A	A	A
PS-47	C	C	C	C	C	C	C	C
PS-48	A	A	A	A	A	A	G*	G*
PS-49	T	T	T	T	T	T	T	T
PS-50	C	C	C	C	C	C	C	C
PS-51	C	C	C	C	C	C	A*	A*
PS-52	T	T	T	T	T	G*	G*	G*
PS-53	A	A	A	A	A	A	A	A
PS-54	G	G	G	G	G	G	G	G

*Sequence differs from Florida isolate. †Additionally validated by sequencing.

Although the differences between the Porton and Florida isolates are relatively few, many of the SNPs in Table 1 have a potential phenotypic effect. These changes are possibly a result of relaxed selection of *B. anthracis* in the laboratory environment. CI-1 and CS-2 result in frameshifts in ABC (ATP-binding cassette) transport genes in the Florida isolate. CS-3 results in nonsynonymous nucleotide change within a conserved domain of a histidine kinase sensor protein. SNP CS-1 causes a histidine (Porton) to tryptophan (Florida) residue change within a conserved "TGS" domain of the predicted guanine pyrophosphokinase (RelA) protein. This gene product plays a key role in response to environmental stress in bacteria (*20*). Another nonsynonymous mutation (CS-8) within the Porton strain is found in the cell differentiation regulator *spo0A*. Each of the eight plasmid VNTRs (Tables 2 and 3) occurs in an intragenic region. The VX1-4 mutation may be a major factor in pathogenicity because it occurs immediately upstream of the lethal toxin precursor gene.

To evaluate the utility of the novel polymorphisms as genetic markers for strain discrimination, we screened genomic DNA from a set of *B. anthracis* isolates (Table 4) (*21*). In addition to the Florida and Porton isolates used as controls, the set included five Ames isolates that were indistinguishable from each other on the basis of existing genotype information (*3*). Four are laboratory isolates obtained before the October 2001 attack that have been designated A to D. The fifth is a 1997 isolate from a goat in Texas that is the only other known natural occurrence of the Ames genotype. Two non-Ames *B. anthracis* isolates that caused North American bovine anthrax infections were included for comparison: one was isolated in Iowa in 1925, the other in California in 2001. The genotypes defined by the polymorphisms allowed us to divide these nine isolates into six distinct groups [Table 4 and Fig. 2; the Porton strain was cured of plasmids and is not represented in Fig. 2]. Data from typing of DNA from the Porton and Florida isolates validated the differences identified from genomic sequence comparisons (Table 4).

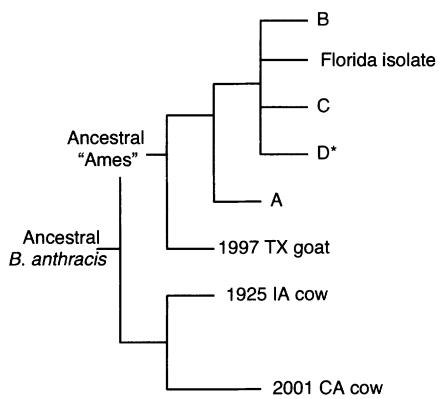

Fig. 2. Proposed tree of Ames isolates derived from strain typing on the basis of plasmid markers in Table 4. Isolates B and C are identical in all markers to the Florida isolate. Isolate D, which differs by only one additional adenine in the 35-bp VX2-3 polyadenylate tract, is shown in the same group as the Florida isolate, marked by an asterisk to indicate this possible difference. Isolate A has two additional bases in VX2-3 (versus the Florida group) and has also lost the pXO1 plasmid. On the basis of the number of differences in Table 4, the Texas goat, 1925 cow, and 2001 cow isolates are more diverged. Branch lengths do not take into account the differing mutation rates of SNPs and VNTRs. Also, the tree is based only on differences found between Ames and pXO plasmids from the Sterne and Pasteur strains.

Of the plasmid markers tested, nine SNPs and six VNTRs or indels varied between *B. anthracis* isolates (Table 4). All but three (PS-1, PS-32, and IX1-1) were found in the pXO2 plasmid, suggesting that this molecule is the source of much short-term variation in the *B. anthracis* genome. The chromosomal polymorphisms were identical in sequence to the Florida isolate in all Ames isolates tested. The fact that the chromosomal SNPs were unique to the Porton isolate is evidence that these mutations occurred after transfer of the strain to Porton Down. The *B. anthracis* strain isolated from a cow in 2001 has 15 differences from the Florida isolate (Table 4) with no recent genetic ties to the Florida or Porton isolates. The 1925 isolate (1925 cow, Table 4) has 12 differences from the Florida isolate, many of them matching the 2001 cow genotype. Within the Ames group, the Texas goat isolate varied at IX1-1 (retaining the 85-bp insert), VX2-1, VX2-3, and PS-52 by a transition mutation. Strain A formed a fifth distinct group, missing pXO1 and containing two extra adenines at VX2-3. Strain D differed only by single nucleotide at VX2-3 (36 adenines compared with 35 in Florida isolate) and cannot be considered a distinct group from B, C, and the controls. Figure 2 shows the relationships between five groups of isolates as a phylogenetic tree. The Porton strain was classified as a sixth group (Table 4) on the basis of its unique chromosomal SNPs. This comparison does not permit any conclusions about mutation rates. The calculation of mutation rates would require specific evidence about the number of generations separating the two samples, which is unknown, particularly with regard to the Florida isolate.

Comparative genomics approaches that are based on whole genome sequence data provide the most comprehensive picture of species diversity and lack the bias of methods such as multilocus sequence typing that sample only a small part of the genome (*22*). Genome-wide comparisons identify polymorphic loci that can be used for cost-efficient PCR-based sampling of variation within bacterial populations (*3*). Of the VNTRs identified in this study, only one with any variation among the strains tested, VX2-5, was known before comparative genomic sequencing. VX2-5 can distinguish the two most distantly related isolates, the 1925 Iowa cow and the 2001 California cow, from the Florida strain but not from each other (Table 4). In addition, comparative genomic sequencing can reveal novel, lineage-specific markers that enhance the resolution of genotyping schemes. In this study, we found 11 DNA sequence differences between the chromosomes of the Florida and Porton isolates of *B. anthracis* (Table 1). These isolates were indistinguishable by their VNTR profiles. Our results indicate that polymorphisms can appear after relatively few generations, because these strains were derived from a common ancestor in the mid-1980s and have been kept frozen.

Comparison of completed genome sequences also allows for elucidation of differences in gene content that may underlie observed phenotypic differences in infectivity and virulence between closely related strains or isolates. It may also facilitate the identification of genes that have been deliberately introduced into potential biowarfare agents. Intergenome comparisons are more efficient when at least one of the genomes is completely finished, rather than when the representations are the multiple, unordered assemblies typical of a "draft" project. When looking for differences between two very closely related strains, as we did here, both genomes must be sequenced to a high level of coverage in order to obtain accurate data. With the rapidly diminishing costs of DNA sequencing (currently at $0.03 per base pair in total costs to generate 8-fold coverage and $0.10 per base pair for a fully finished and annotated sequence), the increasing breadth of databases, and the likelihood of future advances in technology, this approach will become increasingly important.

To evaluate whether SNPs or other small genetic differences are genuine, the underlying quality values need to be available for every sequence that was used in an assembly. We are releasing on the TIGR Web site (www.tigr.org) four sets of *B. anthracis* Florida isolate genome data: (i) the set of all assembled contigs of all sizes, (ii) the set of individual sequences that were used to create those contigs, (iii) a file containing the quality values for each individual sequence, and (iv) a set of 1000-bp regions centered on each SNP that will allow investigators to identify and test these markers on any other anthrax strain. In addition, the genomic scaffold representing the main chromosome of the Florida isolate has been deposited with GenBank (accession number AAAC01000000), and the complete plasmid sequences are

available as accession numbers AE011190 and AE011191. Shotgun electropherogram traces from the chromosome and plasmids were deposited in the National Center for Biotechnology Information Trace Archive (trace IDs 107051265 to 107116771).

Genome-based analysis will provide a powerful new tool for investigation of unexpected outbreaks of infectious disease, whether these represent biological warfare attacks, emerging agents, or more familiar pathogens. Because even highly monomorphic species such as *B. anthracis* are known to exhibit phenotypic variation in infection potential, genomic comparisons can be used to investigate the genetic basis of pathogenesis. To lay the groundwork for future investigations, databases derived from genome-based surveys of natural variation in all major pathogens, not just potential biological warfare agents, should be constructed.

References and Notes

1. Centers for Disease Control, *Morb. Mortal. Wkly. Rep.* **50**, 877 (2001).
2. L. B. Price, M. Hugh-Jones, P. J. Jackson, P. Keim, *J. Bacteriol.* **181**, 2358 (1999).
3. P. Keim *et al.*, *J. Bacteriol.* **182**, 2928 (2000).
4. Centers for Disease Control, *Morb. Mortal. Wkly. Rep.* **50**, 909 (2001).
5. The index strain was sent from Fort Detrick to Porton Down, UK, in or after 1982 and was cured of the pXO1 and pXO2 virulence plasmids (Fig. 1). Two preparations of genomic DNA were sent to TIGR. The first was prepared in a laboratory at the University of California, Berkeley in 1998 ("Porton1"). The second was prepared at Porton Down in 2001 from a bacterium grown from the original frozen culture ("Porton2"). The reported history of storage of the bacteria used to create the cultures provides no obvious reasons for the seven intrastrain chromosomal SNPs observed (Table 1).
6. Materials and methods are available as supporting material on *Science* Online.
7. A. L. Delcher *et al.*, *Nucleic Acids Res.* **27**, 2369 (1999).
8. R. T. Okinaka *et al.*, *J. Bacteriol.* **181**, 6509 (1999).
9. R. T. Okinaka *et al.*, *J. Appl. Microbiol.* **87**, 261 (1999).
10. B. Ewing, L. Hillier, M. C. Wendl, P. Green, *Genome Res.* **8**, 175 (1998).
11. B. Ewing, P. Green, *Genome Res.* **8**, 186 (1998).
12. For a full discussion of quality value determinations, see supporting material on *Science* Online.
13. R. D. Fleischmann, *Emerg. Infect. Dis.* **7**, 487 (2001).
14. If a small portion of the genome has low coverage, this will have a disproportionate effect on the overall accuracy for the genome. For example, suppose that 10% of the genome has low coverage and an error rate of 1/10,000, whereas 90% of the genome has an error rate of 1/100,000,000. At a hypothetical average error rate per sequence of 1% (corresponding to a quality value of 20), 2-fold coverage will yield an error rate of 1/10,000, whereas 4-fold coverage will have an error rate of 1/100,000,000. This distribution would produce an overall sequencing accuracy of 1/100,000, despite the fact that the accuracy for 90% of the genome was essentially perfect.
15. J. Pannucci, R. T. Okinaka, R. Sabin, C. R. Kuske, *J. Bacteriol.* **184**, 134 (2002).
16. X. X. Peterson *et al.*, in preparation.
17. This deletion occurs in a complex repeat composed of four distinct units (labeled A to D). The order of these units in the Pasteur pXO2 plasmid is ACDCBACB, followed by another copy of D about 4.3-kbp downstream. The Florida isolate plasmid lacks the 1.4-kbp region between the two extreme copies of the C unit, leading to the ACB genotype.
18. D. L. Robertson, S. H. Leppla, *Gene* **44**, 71 (1986).
19. C. B. Thorne, in Bacillus subtilis *and Other Gram-Positive Bacteria* (American Society for Microbiology, Washington, DC, 1993), pp. 113–124.
20. A. B. Kolstø *et al.*, *J. Bacteriol.* **184**, 621 (2002).
21. Regions surrounding selected SNPs were amplified by PCR, and the products were sequenced. VNTRs were typed by sequencing and/or sizing on acrylamide gels (*3*). In cases where the SNPs or VNTRs differed from the Florida strain sequence (asterisk in Table 4) two additional sequence reactions were performed in order to obtain better overall quality values and at least 3-fold coverage for each SNP tested.
22. M. C. Enright, B. G. Spratt, *Trends Microbiol.* **12**, 482 (1999).

23. The sequencing of the Florida isolate was supported by National Science Foundation Small Grant for Exploratory Research (MCB-202304 to C.M.F. and T.D.R.). The Porton genome sequence was supported by a grant (N00149610604) from the Office of Naval Research, which was supplemented by the NIH, Department of Energy (grant DEFC0295ER6196), and funds from the United Kingdom Defence Science Technology Laboratory. S.L.S and M.P. were supported in part by NSF (grant KDI-9980088) and NIH (grant R01-LM06845).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1071837/DC1
Materials and Methods
Table S1

13 March 2002; accepted 25 April 2002
Published online 9 May 2002;
10.1126/science.1071837
Include this information when citing this paper.

AID Enzyme-Induced Hypermutation in an Actively Transcribed Gene in Fibroblasts

Kiyotsugu Yoshikawa,[1,2] Il-mi Okazaki,[1] Tomonori Eto,[1] Kazuo Kinoshita,[1] Masamichi Muramatsu,[1] Hitoshi Nagaoka,[1] Tasuku Honjo[1]*

Activation-induced cytidine deaminase (AID), a putative RNA-editing enzyme, is indispensable for somatic hypermutation (SHM), class switch recombination, and gene conversion of immunoglobulin genes, which indicates a common molecular mechanism for these phenomena. Here we show that ectopic expression of AID alone can induce hypermutation in an artificial *GFP* substrate in NIH 3T3 murine fibroblast cells. The frequency of mutations was closely correlated with the level of transcription of the target gene, and the distribution of mutations in NIH 3T3 cells was similar to those of SHM in B lymphocytes. These results indicate that AID is sufficient for the generation of SHM in an actively transcribed gene in fibroblasts, as well as B cells, and that any of the required cofactors must be present in these fibroblasts.

In order to protect against a huge number of pathogens, the vertebrate immune system increases the limited antigen receptor repertoire encoded in the genome by taking advantage of somatic DNA alterations. First, V(D)J recombination assembles two or three pieces of distant germ line segments to form a variable (V) exon of antigen receptor genes during the development of T and B lymphocytes. Subsequently, immunoglobulin genes of peripheral mature B lymphocytes are further modified by three types of genetic alterations: SHM (*1*) and gene conversion (GC) (*2*) in the V gene, and class switch recombination (CSR) in the heavy-chain constant region (C_H) gene (*3*). The immunoglobulins' specificity and affinity for antigen are augmented by either untemplated SHM or pseudogene-templated GC in the V region gene when coupled with selection by the antigen. Distinct from antigen specificity of the receptor, the Cμ gene is replaced with one of other CH genes by CSR, thereby changing immunoglobulin isotype and effector functions. Although molecular mechanisms for the three types of DNA alterations remain to be elucidated, recent isolation of a B cell–specific gene for AID and characterization of its function have shown that all three reactions are dependent on AID (*4–6*). We have shown that ectopic expression of AID induces CSR in an artificial construct introduced into NIH 3T3 murine fibroblasts (*7*); in this study, we ask whether AID expression induces SHM in nonlymphoid cells.

In order to examine hypermutation in nonlymphoid cells, we generated an NIH 3T3 cell line transfected with the tetracycline responsive (tet-off) transactivator and the pI plasmid (*8*), which carries a mutant *GFP* sequence driven by the inducible tetracycline (tet) promoter (Fig. 1A). The mutant *GFP* has a premature stop codon (TAG) in an RGYW [R, purine (A/G); Y, pyrimidine (C/T); W, A/T] SHM hotspot (*9*). Furthermore, the pI plasmid was shown to be useful for assaying SHM in a mouse pre–B cell line (*8*) and a

[1]Department of Medical Chemistry and Molecular Biology, Graduate School of Medicine; and [2]Radioisotope Research Center, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto, 606-8501, Japan.

*To whom correspondence should be addressed. E-mail: honjo@mfour.med.kyoto-u.ac.jp