AEM
Journals.ASM.org

Genetic Barcodes for Improved Environmental Tracking of an Anthrax Simulant

Patricia Buckley,[a] Bryan Rivers,[a,b] Sarah Katoski,[a,b] Michael H. Kim,[a] F. Joseph Kragl,[a] Stacey Broomall,[a] Michael Krepps,[a,c] Evan W. Skowronski,[a]* C. Nicole Rosenzweig,[a] Sari Paikoff,[d] Peter Emanuel,[a] and Henry S. Gibbons[a]

Biosciences Division, U.S. Army Edgewood Chemical Biological Center, Aberdeen Proving Ground, Maryland, USA[a]; Science Applications International Corporation, Aberdeen Proving Ground, Maryland, USA[b]; Excet, Inc., Aberdeen Proving Ground, Maryland, USA[c]; and Defense Threat Reduction Agency, Ft. Belvoir, Virginia, USA[d]

The development of realistic risk models that predict the dissemination, dispersion and persistence of potential biothreat agents have utilized nonpathogenic surrogate organisms such as *Bacillus atrophaeus* subsp. *globigii* or commercial products such as *Bacillus thuringiensis* subsp. *kurstaki*. Comparison of results from outdoor tests under different conditions requires the use of genetically identical strains; however, the requirement for isogenic strains limits the ability to compare other desirable properties, such as the behavior in the environment of the same strain prepared using different methods. Finally, current methods do not allow long-term studies of persistence or reaerosolization in test sites where simulants are heavily used or in areas where *B. thuringiensis* subsp. *kurstaki* is applied as a biopesticide. To create a set of genetically heterogeneous yet phenotypically indistinguishable strains so that variables intrinsic to simulations (e.g., sample preparation) can be varied and the strains can be tested under otherwise identical conditions, we have developed a strategy of introducing small genetic signatures ("barcodes") into neutral regions of the genome. The barcodes are stable over 300 generations and do not impact *in vitro* growth or sporulation. Each barcode contains common and specific tags that allow differentiation of marked strains from wild-type strains and from each other. Each tag is paired with specific real-time PCR assays that facilitate discrimination of barcoded strains from wild-type strains and from each other. These uniquely barcoded strains will be valuable tools for research into the environmental fate of released organisms by providing specific artificial detection signatures.

Spores of *Bacillus anthracis*, the causative agent of anthrax, have been successfully weaponized on large scales in at least two historical offensive biological weapons programs (1, 17, 40, 48). *B. anthracis* spores were disseminated through the mail in the well-documented 2001 anthrax attacks (5, 25–26, 38), and were alleged to have been used as a weapon in the former Rhodesia (29, 32). For this reason, *B. anthracis* remains classified as a category A biothreat agent. Their physical hardiness, their resistance to heat and environmental insults, and the relative ease with which spores can be refined, milled, and aerosolized without significant loss of viability make *B. anthracis* a significant concern as a potential weapon. Its historical use as a weapon or bioterrorism agent and the substantial potential economic consequences of anthrax releases have made understanding the behavior and dynamics of *Bacillus* spores a major focus of research. Knowledge of spore persistence, dissemination, and behavior in response to decontamination regimens is critical to developing accurate risk models and response regimens that are sufficiently robust while minimizing social and economic disruption.

Despite the clear need to acquire knowledge about *B. anthracis* itself, its virulent nature by multiple routes of infection makes the use of the actual agent (or even attenuated derivatives) in outdoor tests impossible. For this reason, initial efforts to develop nonpathogenic bacterial species as simulants focused on *Bacillus atrophaeus* subsp. *globigii*, a relative of *Bacillus subtilis* (12, 18). *B. atrophaeus* subsp. *globigii* has been used for many years as an outdoor simulant of *B. anthracis* (34). However, subsequent research has shown that, while *B. atrophaeus* subsp. *globigii* does mimic many of the properties of *B. anthracis*, it lacks an exosporium and has different thermal-kill properties (7, 11), which decreases its utility as a simulant for *B. anthracis*. The repertoire of *B. atrophaeus* subsp. *globigii* strains in use is quite small and is restricted to a single lineage with very few available polymorphisms that can discriminate between strains, many of which may affect strain and/or spore phenotypes (12).

The limitations of *B. atrophaeus* subsp. *globigii* as a surrogate for *B. anthracis* have prompted several groups to evaluate *Bacillus thuringiensis* subspecies as potential anthrax surrogates (11, 16). Like *B. atrophaeus* subsp. *globigii*, *B. thuringiensis* strains are not known to cause disease in humans, and many strains are available off-the-shelf as biological pesticides for widespread agricultural use in conventional and organic insect pest control (9). Following widespread outdoor applications in pest control scenarios, *Bacillus thuringiensis* subsp. *kurstaki* strains have been recovered from asymptomatic individuals following widespread aerial spray applications over populated areas (45, 46) without any concurrent epidemiological signs of associated disease (22). While *B. thuringiensis* and its pathogenic phylogenetic neighbors *Bacillus cereus* and *B. anthracis* share a highly conserved core genome, the accessory genome or pan-genome is quite variable (28, 36) and consists mainly of phages and plasmids, which encode most of the strain-specific functions that dictate host tropism (e.g., capsule and toxins). The crystalline toxins expressed by *B. thuringiensis* strains are

Received 25 June 2012 **Accepted** 12 September 2012
Published ahead of print 21 September 2012
Address correspondence to Henry S. Gibbons, henry.s.gibbons.civ@mail.mil.
* Present address: Evan W. Skowronski, TMG Biosciences, Incline Village, Nevada, USA.
Supplemental material for this article may be found at http://aem.asm.org/.

doi:10.1128/AEM.01827-12

TABLE 1 Strains and plasmids used in this work

Strain or plasmid	Description	Source or reference
Strains		
B. thuringiensis subsp. *kurstaki*		
ATCC 33679	HD-1 biopesticide strain	ATCC[a]
T1B1	ATCC 33679 ΔpHD1-XO1; barcoded at target 1 with common tag and specific tag 1	This work
T1B2	ATCC 33679 ΔpHD1-XO1; barcoded at target 1 with common tag and specific tag 2	This work
Foray	Commercial HD-1 biopesticide product dispersed in Fairfax County, VA	45
E. coli		
SM10	*E. coli* donor strain	24
SCS110	pSS4333 donor strain	24
Plasmids		
pRP1028	Allelic exchange vector, turbo-*rfp*, Spc^r	24
pSS4332	*I*-SceI expression vector, *gfp*, Kan^r	24
pT1B1	pRP1028 containing target 1 with common tag and specific tag 1	DNA2.0
pT1B2	pRP1028 containing target 1 with common tag and specific tag 2	DNA2.0

[a] ATCC, American Type Culture Collection.

specific to insects and are not known to affect mammalian hosts. Thus, *B. thuringiensis* spores share many of the important physical and biochemical characteristics of anthrax spores but do not pose a biological hazard to humans. While the use of *B. thuringiensis* as an anthrax simulant is not a novel idea (United Nations inspectors recovered a toxinless strain from a suspected bioweapons facility in Iraq in the late 1990s [8]), it has not yet been widely adopted. The widespread application of *B. thuringiensis* (particularly *B. thuringiensis* subsp. *kurstaki*) as a biopesticide has recently facilitated experimental studies of the persistence and transport of *B. thuringiensis* in the environment (46, 47). While those studies have provided extremely valuable information about the life cycle of deliberately released *B. thuringiensis* spores, the agricultural application of commercial *B. thuringiensis* preparations may not mimic the anticipated aerosol dissemination of an authentic biowarfare agent, confounding the ability to develop realistic models.

Gathering accurate information about organism behavior in the environment requires a combination of robust and reproducible sampling techniques, rigorous methods, and, optimally, a well-characterized input strain. Until now, only very limited numbers of suitable strains existed, limiting the number of possible studies in any given area or time until the recoverable signature returned to background levels. Particularly with persistent spores and in heavily used areas such as the U.S. Army's Dugway Proving Ground, low-level positive signals could be either authentic or spurious, potentially resulting from reaerosolization of spores left over from previous tests. In fact, the level of residual *B. atrophaeus* subsp. *globigii* spores in the soils at Dugway Proving Ground is as high as 10^5 spores/g soil (K. Omberg, personal communication). The lack of specific signatures for any given strain has made the differentiation of those events impossible.

As a potential solution to this problem, we describe here a new approach to simulant development whereby a stable genetic tag, or "barcode," is integrated directly into the chromosome of a *B. thuringiensis* subsp. *kurstaki* strain. Each barcode contains two tag modules, one common to all barcoded strains and one specific for each strain. To facilitate the detection and quantitation of each barcoded strain, tag-specific real-time PCR assays that can distinguish the strains from each other, from wild-type strains, and from a panel of near-neighbors and other potential interfering agents are described. We present data on the stability of the barcode during serial transfer and show that the insertion is neutral for *in vitro* growth kinetics. The development of new, specific strains will have a dramatic impact on the methodology of testing and analysis of environmental releases.

MATERIALS AND METHODS

Strains and plasmids. Strains and plasmids utilized in this study are shown in Table 1. Since it was expected that the tagged spore would be used in a broad range of indoor and outdoor test scenarios, strain ATCC 33679, an HD-1 strain (serotype 3a3b) that is registered with the United States Environmental Protection Agency as an approved biopesticide, was selected as the backbone for the barcoding efforts for its outstanding safety record in widespread gypsy moth control efforts, with annual outdoor applications of ~453 metric tons of *B. thuringiensis* subsp. *kurstaki* spores applied over >138,000 Ha in the United States alone with no significant medical issues recorded (44). The ATCC strain was confirmed to be an HD-1 strain of *B. thuringiensis* by comparison of plasmid profiles to previously published work (41) and by whole-genome sequence analysis with *in silico* multilocus sequence typing (MLST), amplified fragment length polymorphism (AFLP), and *cry* gene typing. Unless otherwise indicated, strains were grown on brain heart infusion agar (BHI) containing polymyxin B (50 U/ml) and either spectinomycin (250 μg/ml) or kanamycin (20 μg/ml). Unless otherwise noted, strains were incubated at 30°C.

Identification of a barcode insertion points. Potential insertion sites for the barcodes were identified based on a set of selection rules elaborated in Table 2. Insertion points were identified in the published genome sequence of *B. thuringiensis* subsp. *kurstaki* strains BMB171 (19) and T03a001 (RefSeq accession number BC_CM000751.1). Annotations of the BMB171 genome generated in RAST (4) and PATRIC (15) were compared. We also generated a draft genome sequence of ATCC 33679 (M. Krepps, S. Broomall, P. Roth, C. N. Rosenzweig, and H. S. Gibbons, unpublished data) and verified that the genome structure fulfilled the appropriate criteria. Of 294 intergenic regions >500 bp long (see Table S1 in the supplemental material), three potential target insertion points were identified (Table 3) that fulfilled all of the set criteria.

TABLE 2 Selection rules for barcode insertion points

Rule	Purpose
Target region must be located in the chromosome	Maximize stability by incorporation on major replicon
Insertion point must lie near the midpoint of an intergenic space larger than 500 bp	Minimize disruption of potential coding sequences or regulatory elements
No annotated genes or potential ORFs in the intergenic space	Minimize disruption of potential coding sequences or regulatory elements
Must lie between two convergently transcribed genes	Minimize disruption of potential coding sequences or regulatory elements
No repetitive structure in intergenic space	Facilitate synthesis and cloning of constructs and minimize potential issues with homologous recombination
No identical repetitive elements >200 bp in size within 10,000 bp	Minimize potential loss by deletion via homologous recombination between repeat elements (e.g., insertion sequences)
Target must be intact and consistently annotated in two or more available *B. thuringiensis* subsp. *kurstaki* sequences and in ATCC 33679 draft	Maximize likelihood of success in selected target strain
Target must be present in commercial *B. thuringiensis* subsp. *kurstaki* isolate	Maximize versatility and adaptability of barcode targeting vectors to different strains

Barcode module design. We appropriated a set of published 20-bp tags previously used in signature-tagged mutagenesis studies of pooled yeast strains (33). Tags were individually screened against the *B. thuringiensis* subsp. *kurstaki* genome sequences to eliminate sequences that had homology to any portion of the *B. thuringiensis* subsp. *kurstaki* chromosome. One tag was adopted as a common tag to be shared among multiple strains, while the others were used as strain-specific tags (S1, S2, etc.). The tags were flanked by an EcoRI restriction site to facilitate screening of recombinant strains. Figure 1 shows the general features of a barcode module and the design of associated real-time PCR assays.

Barcode insertion. Barcodes flanked by ~750 bp of chromosomal DNA sequence were generated synthetically (DNA2.0, Menlo Park, CA) and cloned into pRP1028, which was delivered by a protocol adapted from the work of Janes and Stibitz (24). The resulting plasmids were delivered by biparental mating into ATCC 33679. Replication of pRP1028 was suppressed by maintaining strains at 37°C. Strains which had integrated the plasmids by homologous recombination were selected on spectinomycin plates. Fluorescence of integrant colonies due to the turbo-*rfp* on pRP1028 was checked by transillumination. The *I*-SceI-expressing plasmid pSS4333 was delivered by triparental mating into the integrant strains. Green-fluorescing Spc^s colonies were screened for the presence of the barcode by PCR amplification and EcoRI digestion of the target locus. pSS4333 was cured by serial transfer on solid media in the absence of selection. The curing of the plasmids was verified by checking the strains for the absence of red or green fluorescence and by the lack of PCR amplification of plasmid-borne antibiotic resistance genes *spc* and *kan*.

Real-time PCR assays. Primers and concentrations used for the strain construction, verification, and detection of the barcodes are shown in Table S2 in the supplemental material. Barcodes were detected by real-time SYBR green PCR assays in 20-μl volumes in 384-well optical PCR plates. Amplification, data acquisition, and data analysis were carried out on an Applied Biosystems model 7900HT sequence detection system (Applied Biosystems, Foster City, CA). The barcode reactions were set up using SYBR green PCR master mix (catalog no. 4309155; Applied Biosystems, Foster City, CA), forward and reverse primers, nuclease-free sterile water, and 1 μl extracted DNA product. The thermocycler conditions for the common tag and barcode 2 were as follows: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min, followed by a dissociation stage of 95°C for 15 s, 60°C for 15 s, and 95°C for 15 s. The barcode 1 thermocycler conditions were set up similarly to the program above, with the following exception: annealing was at 55°C for 15 s (instead of 60°C for 15 s). The linear range for each reaction was determined by developing a standard curve for eight 10-fold serial dilutions of the corresponding genomic DNA. The efficiency of each reaction was calculated from the resulting graphs.

Barcode stability. A single starter culture of each strain was grown in 5 ml of medium. Three independent 50-ml cultures of each *B. thuringiensis* subsp. *kurstaki* strain (wild type and two strains containing different specific tags in target 1) were grown in BHI medium in shaking flasks at 30°C. Each day, at approximately the same time, cultures were diluted 1:1,000 in fresh medium. The process was repeated for 5 days, after which the cultures were allowed to incubate at 30°C for 3 days in order to induce sporulation. This cycle was repeated each week for 6 weeks, representing approximately 300 doublings. Where applicable, growth was monitored by optical density at 600 nm (OD_{600}).

Comparative growth of barcoded strains. Barcoded strains were grown either in a 20-liter fermentor or in parallel flask cultures. For parallel flask cultures, samples were withdrawn periodically for determination of the OD_{600}. For growth in the fermentors, wild-type and barcoded strains were grown in 20-liter volumes of NZ-Amine A medium in Micros 30 fermentors (New Brunswick Scientific, Enfield, CT). Starter cultures of 500 ml were grown in 2-liter flasks in a shaking incubator until the OD_{600} reached ~0.4. Seed cultures were aseptically transferred into the Micros 30 fermentor (with a 20-liter working volume) containing the NZ-Amine A medium. The operating conditions for the Micros 30-liter fermentor were controlled, with an agitation speed of 300 rpm and an airflow of one air volume per liquid volume per minute at 30°C and pH 7.0. Dissolved oxygen (percent saturation) and optical density (600 nm) were monitored using an in-line probe and by periodic sampling, respectively.

TABLE 3 Potential barcode insertion points

Target	Intergenic locus (BMB171 coordinates)	Flanking gene and product		Intergenic gap size (bp)
		5′	3′	
1	882263–882815	BMB171_C0768 acyl-coenzyme A synthetase	BMB171_C0769, hypothetical protein	552
2	1533619–1534178	BMB171_C1412, thiosulfate sulfurtransferase	VBIBacthu14800_1517,[a] hypothetical protein	559
3	2786602–2787119	BMB171_2615, hypothetical protein	BMB171_C2616, alkane-sulfonate monooxygenase (*ssuD*)	517

[a] PATRIC annotation; no NCBI locus tag called for this gene.

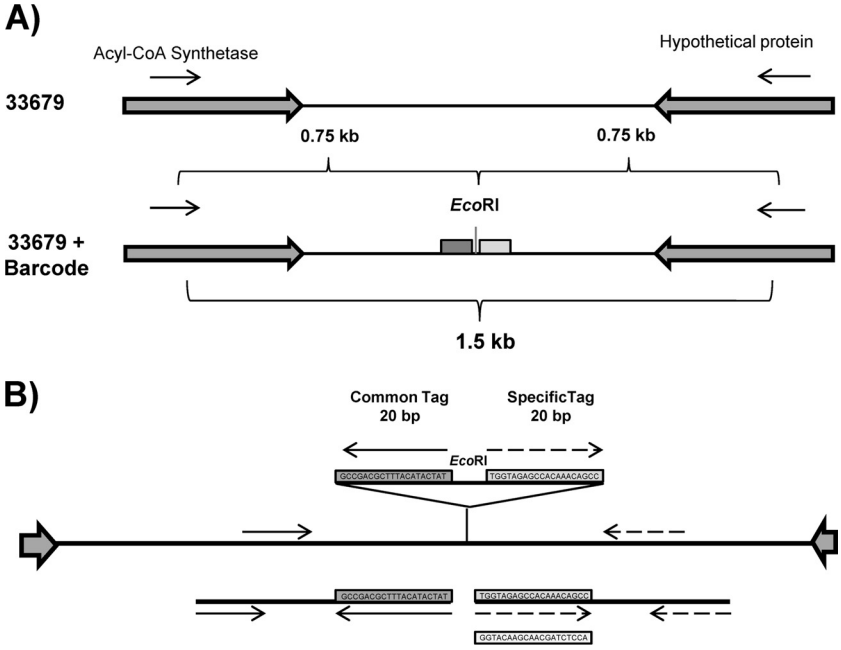

FIG 1 Barcode module design. (A) Barcode modules are integrated into the chromosome between convergently transcribed genes or operons (long filled arrows) in intergenic regions larger than 500 bp. Primers (arrows) are designed to amplify an ~1,500-bp flanking region surrounding the barcode to verify the insertion. (B) Barcode modules consist of two 20-nucleotide tags flanking an EcoRI restriction site. Real-time PCR assays are designed such that one of the tags serves as a primer binding site (arrows) to generate an amplicon using a second primer that anneals to a region of the chromosome flanking the barcode module. One of the tags (the common tag) is present in all barcoded strains, with a second tag (the specific tag) that serves as a unique identifier for each individual strain.

Genomic characterization of the barcoded strain. A 454 shotgun draft sequence of the barcoded strain was generated by standard methods using the 454 Titanium package (Roche/454, Branford, CT). Reads were mapped to the scaffolds from the parent strain that had been generated from melded 454 shotgun and paired-end libraries using Newbler v2.6 (Krepps et al., unpublished). Reads were mapped to the parent strain using the mapping algorithm in Genomics Workbench from CLC Bio (Aarhus, Denmark) using the default parameters. Regions of low or absent sequence coverage were identified, and deletion endpoints, where applicable, were identified by manual inspection of the mapped data.

RESULTS

Integration of barcodes. We selected *B. thuringiensis* subsp. *kurstaki* strain ATCC 33679, a prototypical HD-1 strain of *B. thuringiensis* subsp. *kurstaki*, for our barcoding efforts. Our selection was guided by the long history of the use of HD-1 strains as EPA-approved biopesticides, dating as far back as 1961 (23). *B. thuringiensis* subsp. *kurstaki* HD-1 is the active ingredient in Foray, a commercial *B. thuringiensis* subsp. *kurstaki* product. Target sequences were identified by PCR in both ATCC 33679 and a sample of Foray (Fig. 2A). Barcoded target constructs were synthesized and successfully integrated at two of the three identified loci (targets 1 and 2). PCR amplification of target sequences revealed the presence only of EcoRI-digestible product and not the parent product, indicating successful replacement of the parental allele (Fig. 2B). Similar results were obtained for target 2 (data not shown). Repeated attempts to integrate a barcode into target 3 were unsuccessful for reasons that are not clear at this time.

Real-time PCR detection assay. To allow easy detection of the barcoded strains, real-time SYBR green PCR assays specific to the common and specific tags were developed. The assay directed at the common tag recognized both barcoded strains, whereas the assays directed at the specific tags recognized only their cognate strains. Because one of the primers for each sequence is derived from endogenous genetic material, careful control over primer concentrations and PCR amplification conditions was found to be critical to avoid spurious false-positive signals due to asymmetric amplification. After careful optimization, none of the assays directed at the barcoded strains recognized the wild-type strain.

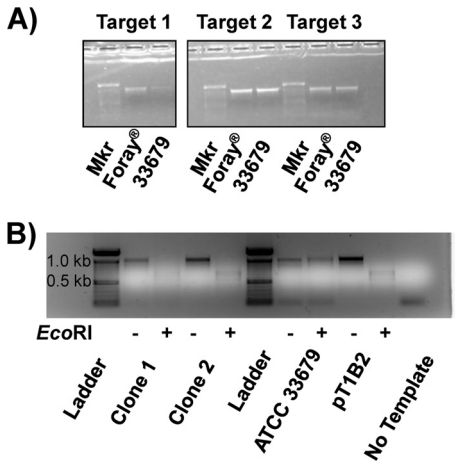

FIG 2 Integration of the barcodes. (A) Verification of the presence of target regions in two *B. thuringiensis* HD-1 variants. Target regions were amplified from genomic DNA from the wild-type strain used to make the barcoded constructs (ATCC 33679) and from Foray, a commercial *B. thuringiensis* subsp. *kurstaki* product. (B) PCR verification of barcode insertion. Target region amplicons from each strain and the suicide plasmid (pT1B1) were digested with EcoRI. The presence of the barcode renders the amplicon susceptible to digestion with EcoRI.

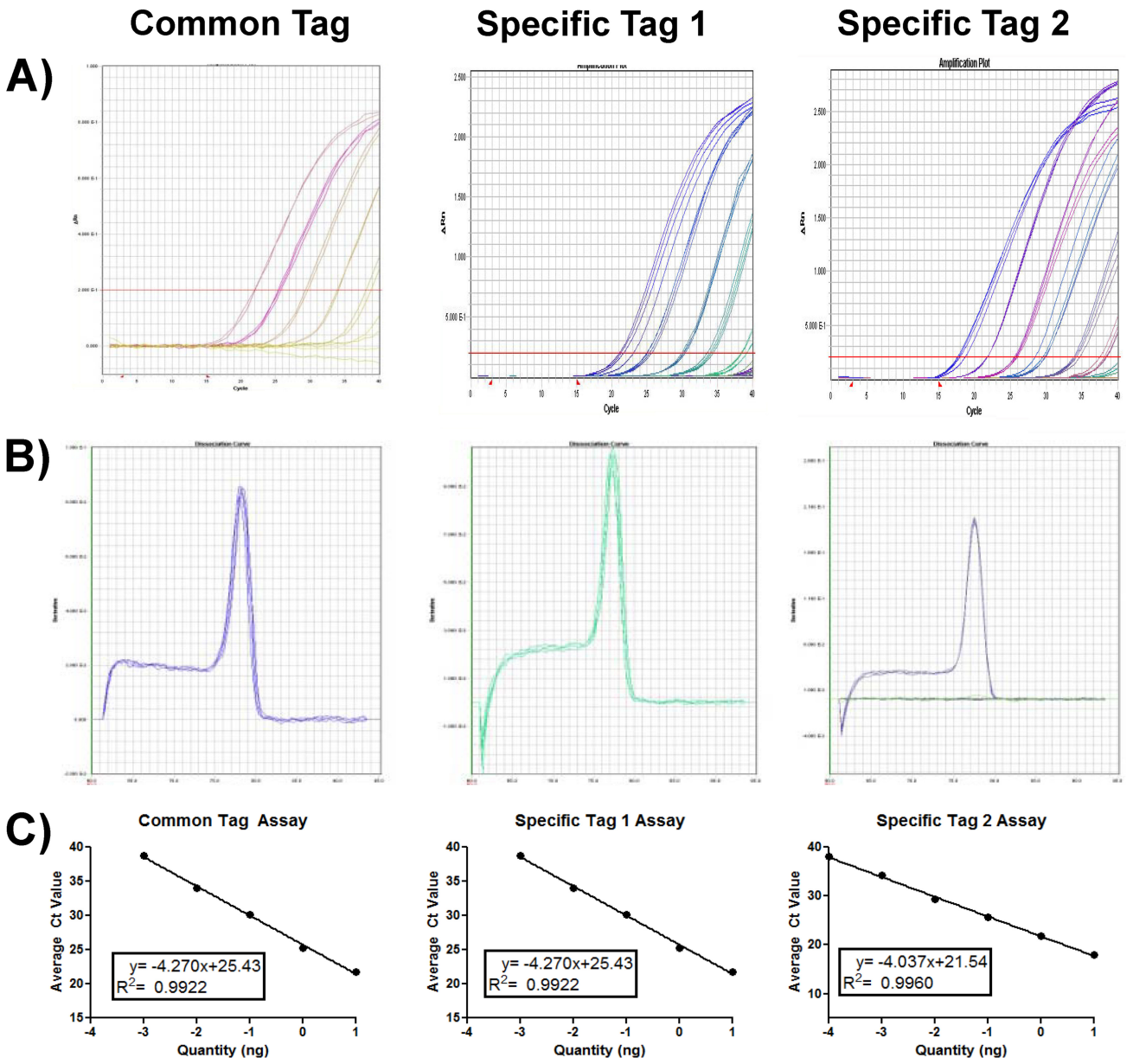

FIG 3 Real-time PCR assay for barcode detection. Genomic DNA from each cognate strain was used as the template for real-time PCR assays using barcode-specific primers. Cycle amplification plots (A), thermal dissociation curves (B), and standard curves (C) for each SYBR green real-time PCR assay are shown.

Figure 3A shows representative real-time PCR assay traces and standard curves (Fig. 3C) for each assay. Table 4 lists the linear range and limit of detection of each assay, along with its calculated efficiency. Based on the 11.2-Mbp estimated genome size obtained from the Newbler *de novo* assembly, which weights the sequence coverage of each element rather than the total size of the assembly, the detection limit for each assay is approximately 8 to 80 genome copies. The differences in the limit of detection between the assays are most likely attributable to the differences in GC content between the chromosomal primer binding sites.

TABLE 4 Sensitivity of PCR assays for *B. thuringiensis* subsp. *kurstaki* strains

Assay	Efficiency (%)	Linearity	Estimated LOD[a] (no. of genome copies)
Common tag	75	1 pg to 10 ng	83
Specific tag 1	72	1 pg to 10 ng	83
Specific tag 2	77	100 fg to 10 ng	8.3

[a] LOD, limit of detection.

Near-neighbor panel screens; inclusivity and exclusivity. Using the real-time SYBR green PCR assays discussed above, we tested the barcode PCR assays for specificity against the barcoded strains themselves, their wild-type parent strains, and a selection of related and unrelated bacterial strains (Table 5). The barcode assays were specific for their cognate targets and did not yield amplicons with the unmarked and near-neighbor strains. When nonspecific amplification was observed, these amplicons produced dramatically higher threshold cycle (C_T) values and no dis-

TABLE 5 Specificity of real-time PCR assays for *B. thuringiensis* subsp. *kurstaki* strains

Strain or material tested	Specificity[a]		
	Common tag	Specific tag 1	Specific tag 2
B. thuringiensis subsp. *kurstaki*	0/4	0/4	0/4
B. thuringiensis subsp. *kurstaki* T1B1	4/4 (22.0)	4/4 (25.2)	0/4
B. thuringiensis subsp. *kurstaki* T1B2	4/4 (27.2)	0/4	4/4 (22.0)
Bacillus anthracis VNR-Δ1	0/4	0/4	0/4
B. anthracis Ames	0/4	0/4	0/4
B. anthracis NNR-Δ1	0/4	0/4	0/4
B. anthracis ΔSterne	0/4	0/4	0/4
B. thuringiensis subsp. *israelensis* ATCC 35646	0/4	0/4	0/4
Bacillus cereus HER1414	0/4	0/4	0/4
B. subtilis ATCC 27370	0/4	0/4	0/4
B. atrophaeus subsp. *globigii*	0/4	0/4	0/4
Pseudomonas aeruginosa PAO-1	0/4	0/4	0/4
Streptococcus pyogenes ATCC 12384	0/4	0/4	0/4
Bordetella pertussis ATCC 9797	0/4	3/4 (39.0)[b]	0/4
Salmonella enterica serovar Typhimurium ATCC 14028	0/4	0/4	0/4
Escherichia coli ATCC 43985	0/4	0/4	0/4
Human placental DNA	0/4	1/4 (39.8)[b]	0/4
Escherichia coli O157:H7	0/4	0/4	0/4
Francisella tularensis SHU4	0/4	0/4	0/4
Yersinia pestis HARBIN35	0/4	0/4	0/4

[a] Number of samples crossing threshold per number of replicates tested. One nanogram of genomic DNA was tested per replicate. Values in parentheses are average C_T values.
[b] Negative dissociation curve data.

cernible thermal dissociation curves (see File S2 in the supplemental material).

Barcode stability and comparative growth experiments. To evaluate the stability of the integrated barcodes, tagged strains were grown in pure cultures for 6 weeks of daily passage in 50-ml shaking cultures during the week and were allowed to enter sporulation every 5 days. This serial transfer experiment was equivalent to approximately 300 bacterial doublings. Large population bottlenecks ($\sim 10^7$ cells) at each transfer were chosen to maximize the chance that, had the barcode inadvertently introduced an unfavorable phenotype, a strain containing a compensatory mutation (such as a deletion) would be randomly sampled during the passage experiment. For passaged and input strains, PCR assays directed at the barcodes yielded equivalent C_T values given identical quantities of input DNA (data not shown), indicating that the barcode had integrated stably into the chromosome and was not generating selective pressure against its retention.

We also compared the *in vitro* growth of barcoded and parental strains. Figure 4 shows the results of 20-liter fermentations of tagged and parent strains. Both optical density and oxygen consumption growth trajectories were very similar, and sporulation frequencies at the conclusion of both runs were identical and close to 100% (data not shown). Although we did not carry out replicate runs in the fermentor, we performed confirmatory growth curve experiments in the flask cultures during the early phases of the serial passage experiment described above. Like the growth in fermentation culture, the growth curves in flask culture were superimposable and were identical across three parallel cultures for each strain.

Genome resequencing. To identify any other potential genetic alterations to the barcoded strains and to verify unambiguously the location and uniqueness of the barcode insertion, we resequenced one of the barcoded strains and mapped the data onto a Newbler assembly of ATCC 33679. The barcode insertion points were evident at the specified locus (Fig. 5); our resequencing data indicated that approximately 350 kb of genetic material had been lost at some point during the strain construction process. Most of the deleted material corresponded to three of the scaffolded regions annotated as plasmids. Bioinformatic analysis of the annotated features generated in RAST and comparison with the deleted genes (see Table S3 in the supplemental material) revealed that most of the deleted material was likely one or more of the many plasmids present in *B. thuringiensis* subsp. *kurstaki* and *B. cereus* strains (3, 21, 35, 37). The genes lost included many homologues of genes on *B. anthracis* plasmid pXO1 (35). These plasmids, including pXO1 itself from *B. anthracis*, are readily cured during growth at higher temperatures (2, 39, 43), and given the requirement of prolonged 37°C incubation to suppress plasmid replication during the homologous recombination phase of strain construction, the loss of such material is not surprising.

DISCUSSION

We have successfully introduced small genetic barcodes—short, specific identifying signatures—into the genome of *B. thuringiensis* subsp. *kurstaki* and coupled the integration of those signatures to specific real-time PCR detection assays directed toward those barcodes. Our work differs from the widely used "signature tags" that uniquely identify transposon insertions and track abundance of individual mutant pools in large populations (20, 33), in that we aim to tag a single locus in multiple isolates with multiple stable chromosomal tags. Our work is similar in intent to the efforts in the synthetic biology community, which added specific "watermarks" to differentiate synthetic genomes from their natural counterparts (13, 14), in that it seeks to incorporate simple, neutral signatures into the chromosome as means of uniquely marking a strain. In fact, our efforts expand upon the idea of watermarking strains by developing specific detection assays based on real-time PCR. Indeed, as described in the accompanying article, these assays allow the detection and differentiation of barcoded *B. thuringiensis* subsp. *kurstaki* strains both in the laboratory and in the field (10).

The ability to assign a specific marker to a strain used in field

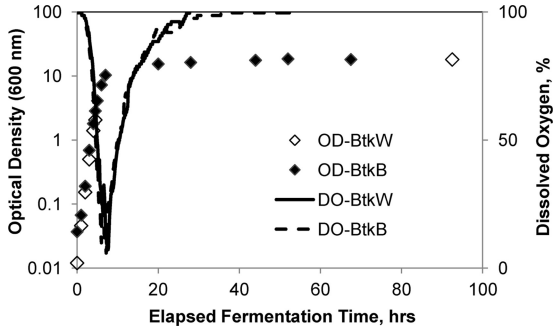

FIG 4 *In vitro* growth of wild-type and barcoded strains in 20-liter fermentors. Wild-type (BtkW) and barcoded (BtkB) strains were grown in 20-liter fermentors as described in Materials and Methods. Optical density (OD) and dissolved O_2 (DO) were monitored over the course of the fermentation.

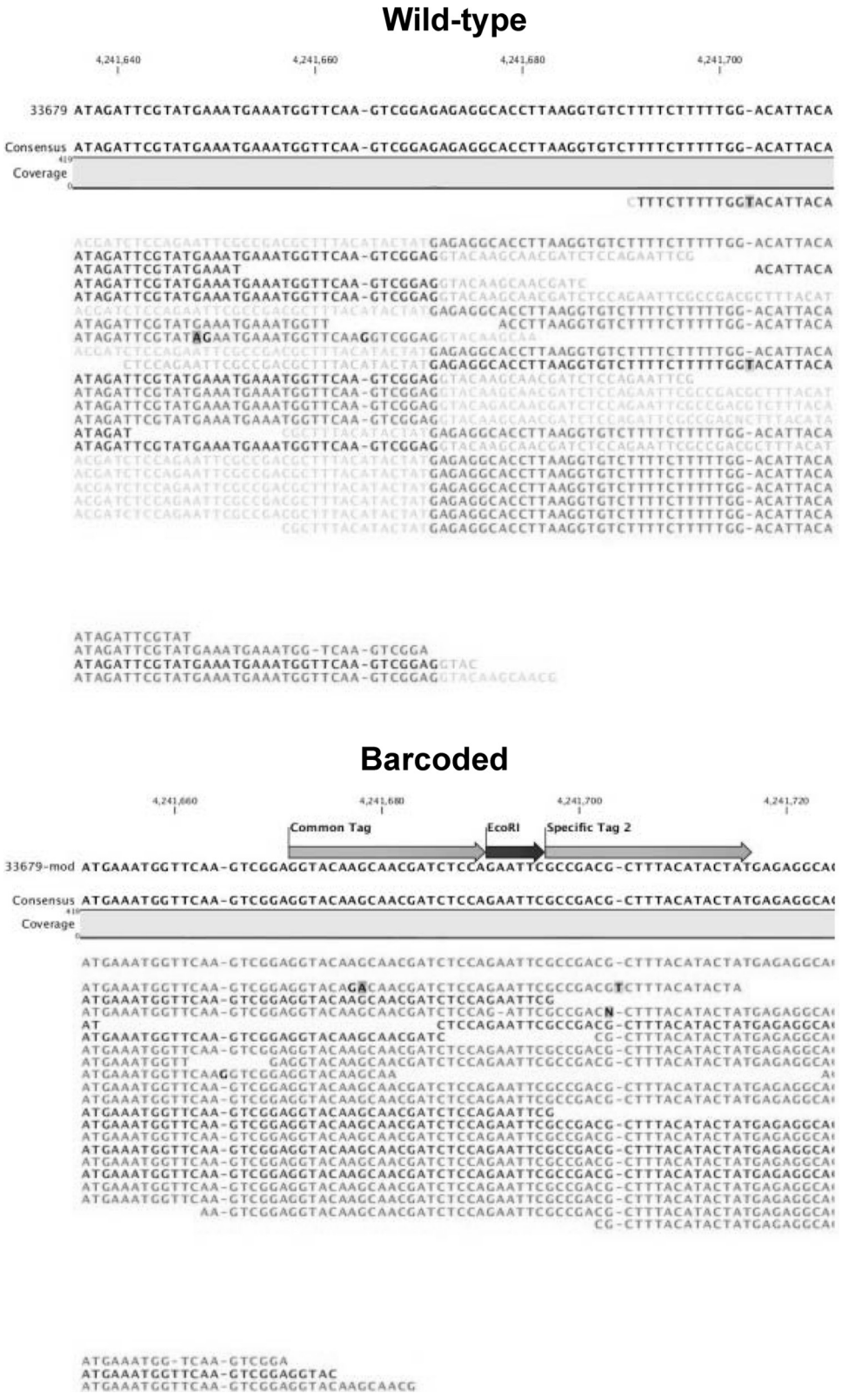

FIG 5 Whole-genome sequencing to verify barcode insertion. 454 shotgun sequencing reads were mapped onto the draft genome sequences of the wild-type (top) and *in silico*-modified barcoded (bottom) genomes. Perfect matches to reference sequence (top row in each panel) are indicated by dark letters; imperfect matches are indicated by lightly shaded letters.

release studies has additional potential advantages, particularly with viable bioinsecticides. Like simulant releases in heavily used proving ground areas, attempts to attribute infections with *Bacillus* spp. in areas of widespread *B. thuringiensis* subsp. *kurstaki* application to the *B. thuringiensis* subsp. *kurstaki* serotype actually applied to the treated area have been confounded by ubiquitous environmental *B. thuringiensis* isolates that cannot always be unambiguously distinguished from the biopesticide strain (reference 45 and references therein). The barcodes endow the new simulant strains with a unique signature that can allow us to definitively exclude simulant *B. thuringiensis* subsp. *kurstaki* strains as causative agents of suspected *B. thuringiensis* infections that might coincide with simulant releases, and if adopted by commercial pesticide manufacturers, these barcodes could serve as exclusionary markers for the biopesticide strains. Furthermore, insertion of the tag into the chromosome minimizes chances of transfer to other strains or species, as the rates of transfer of chromosomal loci between *B. thuringiensis* strains are quite low (2).

We acknowledge that the process used to insert the barcode may have caused the curing of one or more plasmids and/or the loss of chromosomal material. In previous years, the ability to perform *post hoc* genomic characterization of mutagenized strains was cost prohibitive. However, modern whole-genome sequencing allows the detection of such events and allows the precise inventory of genomic content of product strains. The loss of genetic material during *in vitro* culture of *Bacillus* strains is not surprising—plasmids are often unstable at high temperatures, and the strains carry a complement of genetic material that may be unfavorable during *in vitro* growth in rich medium. Recombination of T1B2 into the chromosome during the first step of homologous recombination occurred with much lower frequency than that of T1B1 (data not shown); in fact, the only successful integrant obtained was the deletion construct described here. In contrast, multiple successful integrants were obtained for T1B1. Together, these results suggest that one or more elements of T1B2 may be incompatible with a plasmid-encoded functionality, most likely putative restriction endonuclease encoded within the 400 kb of deleted material. While our barcode does not contain any obvious candidates for a restriction endonuclease recognition site, we cannot exclude the possibility that it may be sensitive to an endonuclease activity that is specific to a sequence motif present only in barcode 2; the most likely candidate at this time is the GATC consensus Dam methylation site present in barcode 2 (Fig. 1). While the effect of the loss of ~400 kb of genetic material (see Table S3 in the supplemental material) in our strains is not immediately evident, inferences might be gained from studies of plasmid loss in *B. anthracis*. In particular, loss of pXO1 from *B. anthracis* strains is associated with numerous phenotypic changes, including changes to sporulation kinetics, nutritional requirements, and phage sensitivity (42). Our strain also had lost a suite of genes involved in the biosynthesis of zwittermicin, a biologically active compound that, among other activities, potentiates the activity of crystal toxins in insect hosts (6, 27). The loss of this gene cluster containing large polyketide synthase modules is reminiscent of the early loss of surfactin biosynthesis during the domestication of *B. subtilis* and *B. atrophaeus* subsp. *globigii* strains (12, 30, 31). Based on our phenotypic analysis, the effects of the loss on *in vitro* growth, colony morphology, and sporulation of the barcoded strains appear to have been minimal. We are attempting to recreate the barcoded strains to retain and/or restore as much of the full complement of genes as possible.

The ability to differentiate two tagged strains based on the site-specific integration of specific genetic tags will allow controlled studies in situations where variables previously could not be controlled. For example, it is anticipated that two strains could be prepared or disseminated using different methods, released and collected simultaneously under identical environmental conditions, and then tracked independently in a single set of samples. Our data indicate that simultaneous detection and quantitation may be possible in mixtures containing wild-type and barcoded strains. Alternatively, a test area could be reused quickly without having to wait for the detection signals to return to background levels.

We believe that our barcoding strategy will be generally applicable to genetically tractable microorganisms, although the specific barcode sequences, the genetic tools required to deliver barcodes, and the actual insertion points for the barcodes themselves will differ from organism to organism. These variations will be based on the overall and local genetic structure of the target organism. We are currently automating the bioinformatic identification of barcode insertion points and the design of barcode modules to maximize specificity, sensitivity, and selectivity across a broad range of potential target organisms.

ACKNOWLEDGMENTS

This work was supported by the Defense Threat Reduction Agency (DTRA) project number CB3654 to H.S.G. Sequencing of the *B. thuringiensis* subsp. *kurstaki* isolate was supported by DTRA project number CB2847 to H.S.G., C.N.R, and E.W.S.

We thank Jason Edmonds and Vipin Rastogi for critical reading of the manuscript.

The opinions stated in this article are those of the authors and do not represent the official policy of the U.S. Army, Department of Defense, or the Government of the United States. Information in this report is cleared for public release.

REFERENCES

1. **Abramova FA, Grinberg LM, Yampolskaya OV, Walker DH.** 1993. Pathology of inhalational anthrax in 42 cases from the Sverdlovsk outbreak of 1979. Proc. Natl. Acad. Sci. U. S. A. **90**:2291–2294.
2. **Aronson AI, Beckman W.** 1987. Transfer of chromosomal genes and plasmids in *Bacillus thuringiensis*. Appl. Environ. Microbiol. **53**:1525–1530.
3. **Aronson AI, Beckman W, Dunn P.** 1986. *Bacillus thuringiensis* and related insect pathogens. Microbiol. Rev. **50**:1–24.
4. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75.
5. **Barakat LA, et al.** 2002. Fatal inhalational anthrax in a 94-year-old Connecticut woman. JAMA **287**:863–868.
6. **Broderick NA, Goodman RM, Raffa KF, Handelsman J.** 2000. Synergy between zwittermicin A and Bacillus thuringiensis subsp. kurstaki against gypsy moth (Lepidoptera: Lymantriidae). Environ. Entomol. **29**:101–107.
7. **Carrera M, Zandomeni RO, Fitzgibbon J, Sagripanti JL.** 2007. Difference between the spore sizes of Bacillus anthracis and other Bacillus species. J. Appl. Microbiol. **102**:303–312.
8. **Challacombe JF, et al.** 2007. The complete genome sequence of *Bacillus thuringiensis* Al Hakam. J. Bacteriol. **189**:3680–3681.
9. **Crickmore N.** 2006. Beyond the spore—past and future developments of Bacillus thuringiensis as a biopesticide. J. Appl. Microbiol. **101**:616–619.
10. **Emanuel PA, et al.** 2012. Detection and tracking of a novel genetically tagged biological simulant in the environment. Appl. Environ. Microbiol. **78**:8281–8288.
11. **Fricker M, Agren J, Segerman B, Knutsson R, Ehling-Schulz M.** 2011. Evaluation of Bacillus strains as model systems for the work on Bacillus anthracis spores. Int. J. Food Microbiol. **145**(Suppl 1):S129–S136.

12. **Gibbons HS, et al.** 2011. Genomic signatures of strain selection and enhancement in *Bacillus atrophaeus* var. *globigii*, a historical biowarfare simulant. PLoS One **6**:e17836. doi:10.1371/journal.pone.0017836.
13. **Gibson DG, et al.** 2008. Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science **319**:1215–1220.
14. **Gibson DG, et al.** 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. Science **329**:52–56.
15. **Gillespie JJ, et al.** 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. Infect. Immun. **79**:4286–4298.
16. **Greenberg DL, Busch JD, Keim P, Wagner DM.** 2010. Identifying experimental surrogates for Bacillus anthracis spores: a review. Invest. Genet. **1**:4.
17. **Haseltine WA.** 1999. Biohazard—the chilling true story of the largest covert biological weapons program in the world—told from the inside by the man who ran it. Science **285**:1019–1020.
18. **Hayward AE, Marchetta JA, Hutton RS.** 1946. Strain variation as a factor in the sporulating properties of the so-called *Bacillus globigii*. J. Bacteriol. **52**:51–54.
19. **He J, et al.** 2010. Complete genome sequence of *Bacillus thuringiensis* mutant strain BMB171. J. Bacteriol. **192**:4074–4075.
20. **Hensel M, et al.** 1995. Simultaneous identification of bacterial virulence genes by negative selection. Science **269**:400–403.
21. **Hu X, Van der Auwera G, Timmery S, Zhu L, Mahillon J.** 2009. Distribution, diversity, and potential mobility of extrachromosomal elements related to the *Bacillus anthracis* pXO1 and pXO2 virulence plasmids. Appl. Environ. Microbiol. **75**:3016–3028.
22. **Human Health Surveillance Scientific Committee.** 1999. Human health surveillance during the aerial spraying for control of North American gypsy moth on southern Vancouver Island, British Columbia, 1999. Capital Health Region, Canada.
23. **Ibrahim MA, Griko N, Junker M, Bulla LA.** 2010. Bacillus thuringiensis: a genomics and proteomics perspective. Bioeng. Bugs **1**:31–50.
24. **Janes BK, Stibitz S.** 2006. Routine markerless gene replacement in *Bacillus anthracis*. Infect. Immun. **74**:1949–1953.
25. **Jernigan DB, et al.** 2002. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. Emerg. Infect. Dis. **8**:1019–1028.
26. **Jernigan JA, et al.** 2001. Bioterrorism-related inhalational anthrax: the first 10 cases reported in the United States. Emerg. Infect. Dis. **7**:933–944.
27. **Kevany BM, Rasko DA, Thomas MG.** 2009. Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. Appl. Environ. Microbiol. **75**:1144–1155.
28. **Kolsto AB, Lereclus D, Mock M.** 2002. Genome structure and evolution of the Bacillus cereus group. Curr. Top. Microbiol. Immunol. **264**:95–108.
29. **Mangold T, Goldberg J.** 1999. Plague wars: the terrifying reality of biological warfare. St. Martin's Griffin, New York, NY.
30. **McLoon AL, Guttenplan SB, Kearns DB, Kolter R, Losick R.** 2011. Tracing the domestication of a biofilm-forming bacterium. J. Bacteriol. **193**:2027–2034.
31. **Nakano MM, Corbell N, Besson J, Zuber P.** 1992. Isolation and characterization of sfp: a gene that functions in the production of the lipopeptide biosurfactant, surfactin, in Bacillus subtilis. Mol. Gen. Genet. **232**:313–321.
32. **Nass M.** 1992. Anthrax epizootic in Zimbabwe, 1978–1980: due to deliberate spread? Phys. Soc. Responsib. Quart. **2**:198–209.
33. **Oh J, et al.** 2010. A universal TagModule collection for parallel genetic analysis of microorganisms. Nucleic Acids Res. **38**:e146.
34. **Page WF, Young HA, Crawford HM, Institute of Medicine (U.S.).** 2007. Long-term health effects of participation in Project SHAD (Shipboard Hazard and Defense). National Academies Press, Washington, DC.
35. **Pannucci J, Okinaka RT, Sabin R, Kuske CR.** 2002. *Bacillus anthracis* pXO1 plasmid sequence conservation among closely related bacterial species. J. Bacteriol. **184**:134–141.
36. **Papazisi L, et al.** 2011. Investigating the genome diversity of B. cereus and evolutionary aspects of B. anthracis emergence. Genomics **98**:26–39.
37. **Rasko DA, et al.** 2004. The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis pXO1. Nucleic Acids Res. **32**:977–988.
38. **Rasko DA, et al.** 2011. Bacillus anthracis comparative genome analysis in support of the Amerithrax investigation. Proc. Natl. Acad. Sci. U. S. A. **108**:5027–5032.
39. **Read TD, et al.** 2003. The genome sequence of Bacillus anthracis Ames and comparison to closely related bacteria. Nature **423**:81–86.
40. **Regis E.** 1999. The biology of doom: the history of America's secret germ warfare project. Henry Holt, New York, NY.
41. **Reyes-Ramírez A, Ibarra JE.** 2008. Plasmid patterns of *Bacillus thuringiensis* type strains. Appl. Environ. Microbiol. **74**:125–129.
42. **Thorne C.** 1993. *Bacillus anthracis*, p 113–124. *In* Sonenshein AL, Hoch JA, Losick R (ed), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, DC.
43. **Uchida I, Hashimoto K, Terakado N.** 1986. Virulence and immunogenicity in experimental animals of Bacillus anthracis strains harbouring or lacking 110 MDa and 60 MDa plasmids. J. Gen. Microbiol. **132**:557–559.
44. **U.S. Department of Agriculture.** 2008. Gypsy moth management in the United States: a cooperative approach. U.S. Department of Agriculture, Newtown Square, PA.
45. **Valadares de Amorim G, Whittome B, Shore B, Levin DB.** 2001. Identification of *Bacillus thuringiensis* subsp. *kurstaki* strain HD1-like bacteria from environmental and human samples after aerial spraying of Victoria, British Columbia, Canada, with Foray 48B. Appl. Environ. Microbiol. **67**:1035–1043.
46. **Van Cuyk S, et al.** 2011. Persistence of *Bacillus thuringiensis* subsp. *kurstaki* in urban environments following spraying. Appl. Environ. Microbiol. **77**:7954–7961.
47. **Van Cuyk S, Veal LA, Simpson B, Omberg KM.** 2011. Transport of Bacillus thuringiensis var. kurstaki via fomites. Biosecur. Bioterror. **9**:288–300.
48. **Walker DH, Yampolska O, Grinberg LM.** 1994. Death at Sverdlovsk: what have we learned? Am. J. Pathol. **144**:1135–1141.