

# Genomic Signatures of Strain Selection and Enhancement in *Bacillus atrophaeus* var. *globigii*, a Historical Biowarfare Simulant

Henry S. Gibbons<sup>1\*</sup>, Stacey M. Broomall<sup>1</sup>, Lauren A. McNew<sup>1,2</sup>, Hajnalka Daligault<sup>10</sup>, Carol Chapman<sup>7</sup>, David Bruce<sup>10</sup>, Mark Karavis<sup>1</sup>, Michael Krepps<sup>1,11</sup>, Paul A. McGregor<sup>1,3</sup>, Charles Hong<sup>1,8</sup>, Kyong H. Park<sup>1</sup>, Arya Akmal<sup>7</sup>, Andrew Feldman<sup>4</sup>, Jeffrey S. Lin<sup>4</sup>, Wenling E. Chang<sup>9</sup>, Brandon W. Higgs<sup>9</sup>, Plamen Demirev<sup>4</sup>, John Lindquist<sup>5</sup>, Alvin Liem<sup>1,6</sup>, Ed Fochler<sup>1,6</sup>, Timothy D. Read<sup>7<sup>aa</sup></sup>, Roxanne Tapia<sup>10</sup>, Shannon Johnson<sup>10</sup>, Kimberly A. Bishop-Lilly<sup>7</sup>, Chris Detter<sup>10</sup>, Cliff Han<sup>10</sup>, Shanmuga Sozhamannan<sup>7</sup>, C. Nicole Rosenzweig<sup>1</sup>, Evan W. Skowronski<sup>1<sup>ab</sup></sup>

**1** BioSciences Division, Edgewood Chemical Biological Center, Aberdeen Proving Ground, Maryland, United States of America, **2** Battelle Memorial Institute, Aberdeen Proving Ground, Maryland, United States of America, **3** Science Applications International Corporation, Aberdeen Proving Ground, Maryland, United States of America, **4** Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland, United States of America, **5** Department of Bacteriology, University of Wisconsin, Madison, Wisconsin, United States of America, **6** OptiMetrics Inc, Abingdon, Maryland, United States of America, **7** Naval Medical Research Center, Biological Defense Research Directorate, Silver Spring, Maryland, United States of America, **8** Defense Threat Reduction Agency, Fort Belvoir, Virginia, United States of America, **9** The MITRE Corporation, McLean, Virginia, United States of America, **10** Department of Energy Joint Genome Institute, Los Alamos National Laboratories, Los Alamos, New Mexico, United States of America, **11** Excet Inc., Aberdeen Proving Ground, Maryland, United States of America

## Abstract

**Background:** Despite the decades-long use of *Bacillus atrophaeus* var. *globigii* (BG) as a simulant for biological warfare (BW) agents, knowledge of its genome composition is limited. Furthermore, the ability to differentiate signatures of deliberate adaptation and selection from natural variation is lacking for most bacterial agents. We characterized a lineage of BG with a long history of use as a simulant for BW operations, focusing on classical bacteriological markers, metabolic profiling and whole-genome shotgun sequencing (WGS).

**Results:** Archival strains and two “present day” type strains were compared to simulant strains on different laboratory media. Several of the samples produced multiple colony morphotypes that differed from that of an archival isolate. To trace the microevolutionary history of these isolates, we obtained WGS data for several archival and present-day strains and morphotypes. *Bacillus*-wide phylogenetic analysis identified *B. subtilis* as the nearest neighbor to *B. atrophaeus*. The genome of *B. atrophaeus* is, on average, 86% identical to *B. subtilis* at the nucleotide level. WGS of variants revealed that several strains were mixed but highly related populations and uncovered a progressive accumulation of mutations among the “military” isolates. Metabolic profiling and microscopic examination of bacterial cultures revealed enhanced growth of “military” isolates on lactate-containing media, and showed that the “military” strains exhibited a hypersporulating phenotype.

**Conclusions:** Our analysis revealed the genomic and phenotypic signatures of strain adaptation and deliberate selection for traits that were desirable in a simulant organism. Together, these results demonstrate the power of whole-genome and modern systems-level approaches to characterize microbial lineages to develop and validate forensic markers for strain discrimination and reveal signatures of deliberate adaptation.

**Citation:** Gibbons HS, Broomall SM, McNew LA, Daligault H, Chapman C, et al. (2011) Genomic Signatures of Strain Selection and Enhancement in *Bacillus atrophaeus* var. *globigii*, a Historical Biowarfare Simulant. PLoS ONE 6(3): e17836. doi:10.1371/journal.pone.0017836

**Editor:** Jonathan Badger, J. Craig Venter Institute, United States of America

**Received:** October 20, 2010; **Accepted:** February 15, 2011; **Published:** March 25, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was made possible by the Defense Threat Reduction Agency Transformational Medical Technologies Initiative CB2847 to HSG and ES; by TMTI0068\_07\_NM\_T (to KBL and SS), and by Department of Homeland Security award number NBCHC070040 to AF, PD, and JL. The funders had no role in study design, data collection and analysis, or preparation of the manuscript. The opinions therein are those of the authors and are not necessarily those of the U.S. Government or any of its agencies. Information in this report is unclassified and cleared for public release.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: henry.gibbons@us.army.mil

These authors contributed equally to this work.

<sup>aa</sup> Current address: Emory University School of Medicine, Emory University, Atlanta, Georgia, United States of America

<sup>ab</sup> Current address: Operational Surveyors, Incline Village, Nevada, United States of America

## Introduction

*Bacillus atrophaeus* is a soil-dwelling, non-pathogenic, aerobic spore-forming bacillus related to *B. subtilis*. For more than six decades, this organism has played an integral role in the biodefense community as a simulant for biological warfare and bioterrorism events (BW) and is commonly referred to by its military two-letter designation “BG” [1,2]. *B. atrophaeus* has served in studies of agent dispersal [3], decontamination simulations [4,5] and large-scale process development [6]. In addition to its historical use as a BW simulant, it is currently in widespread commercial use as a surrogate for spore-forming bacteria [5,7] and is the basis of numerous assays for spore inactivation [8,9]. In addition to its role as a simulant, the organism plays an important role in the biotechnology industry as a source of restriction endonucleases and of the glycosylation inhibitor nojirimycin [10].

The taxonomic placement of *B. atrophaeus* has changed dramatically over the years. Originally isolated as *B. globigii* in 1900 (Migula) as a variant of *B. subtilis*, it was originally distinguished from *B. subtilis* by the formation of a black-tinted pigment on nutrient agar and by low rates of heterologous gene transfer from *B. subtilis* [11]. It has been alternately known as *B. subtilis* var. *niger*, *B. niger*, and has been confused with *B. licheniformis* [12]. Other than the formation of the dark pigment, it is virtually indistinguishable from *B. subtilis* by conventional phenotypic analysis [13], and the lack of distinguishing metabolic or phenotypic features has contributed to the confusion in the taxonomic placement of this organism. Low interspecies DNA transfer frequencies suggested substantial divergence [11]. Based on analysis of comparative DNA hybridization, phenotypic and biochemical tests, Nakamura advocated that pigment-producing *B. subtilis*-like isolates should be classified as a distinct species termed *B. atrophaeus* [13]. Recently, more sensitive typing methods such as amplified fragment length polymorphism analysis showed that *B. atrophaeus* strains could be classified into two major biovars: var. *globigii* encompassing the classical, commonly used BG isolates, and var. *atrophaeus* encompassing other closely related yet genetically distinct strains [14].

Here we report the definitive molecular typing of several BG strains using whole-genome sequences, and develop a plausible microevolutionary history of a commonly used lineage based on the accumulation of mutations over time and during transfer between laboratories. The selected strains span more than six decades of development, use, and transfer of BG between various institutions and laboratories and offer an unparalleled opportunity to investigate mutation under selection and drift over time. Phenotypic analysis revealed substantial heterogeneity both between and within strains, even in type strains, while high-throughput metabolic profiling revealed metabolic “enhancements” to a population that had returned to the University of Wisconsin (UW) from Camp Detrick in 1952. Whole-genome comparisons of single-nucleotide polymorphisms (SNPs), small insertion/deletion motifs (indels), and large-scale genomic architecture analysis by optical maps are combined to generate a plausible history of acquisition and use of operationally relevant strains by the American Type Culture Collection (ATCC) and by several laboratories within the biodefense community.

Finally, our analysis of mutation profiles revealed potential signatures of the deliberate selection of strains with properties of enhanced growth and spore yields, properties that were deemed desirable in a simulant [6]. We also report genetic differences between strains in use in the biodefense community and the commercial sector that argue for adoption of a more uniform standard for *B. atrophaeus* as a simulant.

## Materials and Methods

### Strains and growth conditions

*B. atrophaeus* strains and their sources are indicated in Table 1. Archival strains were maintained as spores in sterile soil at the University of Wisconsin (Figure 1). The 1013 lineage, originally founded from the 1942 strain, was extensively passaged by serial transfer every 12–18 months on agar slants for 30 years. Unless otherwise indicated, strains were grown using LB agar plates, LB agar broth or Tryptic Soy agar containing 5% sheep’s blood (SBA, HealthLink) at 37°C.

### Analysis of colony morphology variation

Spores were germinated by plating on LB media at 37°C. Plates were examined by stereomicroscopy using indirect lighting and imaged using a Nikon SMZ1500 with a total magnification of 16×. Colonies exhibiting distinct morphologies were repeatedly streaked to confirm stability of the phenotype.

### Whole genome sequencing

Genomic DNA was prepared from all isolates using the Blood and Cell Culture DNA Midi Kit for Bacteria (QIAGEN) from 10 ml overnight cultures in LB. BACI051-N was sequenced at the Naval Medical Research Center, while all other isolates were sequenced to >25-fold coverage at the US Army Edgewood Chemical Biological Center by massively parallel pyrosequencing on the Roche/454 GS-FLX using the Titanium reagent package. Draft genome sequences of all isolates were assembled *de novo* using Newbler [15] (Roche) and analyzed using both Newbler and Lasergene (DNASTar, Madison, WI). The 1942 Vogel isolate was designated as the reference strain and was brought to completion using standard finishing techniques.

The draft genome of *Bacillus atrophaeus* var. *globigii* was finished at the Department of Energy Joint Genome Institute (JGI) using a combination of Illumina [16] and 454 datasets [15]. For this genome, we constructed and sequenced an Illumina GAii shotgun library which generated 151,202,17 reads totaling 544 Mb, which was combined with 454 Titanium standard library which generated 387,327 reads totaling 137 Mb of 454 data. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. The initial draft assembly contained 25 contigs in 25 scaffolds. The 454 Titanium standard data were assembled with Newbler, version 2.3. The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data were assembled with VELVET, version 0.7.63 [17], and the consensus sequences were computationally shredded into 1.5 kb overlapping fake reads (shreds). We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). The software Consed [18,19,20] was used in the following finishing process. Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher [21], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 79 additional reactions and 10 shatter libraries were necessary to close gaps and to raise the quality of the finished sequence. The total size of the genome is 4 168 266 bp and the final assembly is based on 137 Mb of 454 draft data which provides an average 33.4× coverage of the genome and 544 Mb of Illumina draft data which provides an average 133× coverage of the genome. The complete sequence and WGS were deposited at DDBJ/EMBL/GenBank under accession numbers listed in Table 2. The

**Table 1.** Strains utilized in this study.

| Strain       | Description   | Source                 |
|--------------|---|------------------------|
| 1942 Vogel   | Archival isolate  | U. Wisconsin           |
| Detrick-1    | "Camp Detrick", morphology variant  | U. U. Wisconsin        |
| Detrick-2    | "Camp Detrick", morphology variant  | U. Wisconsin           |
| Detrick-3    | "Camp Detrick", morphology variant  | U. Wisconsin           |
| Dugway       | Bioferm Lot 10–88, large-scale fermentor strain   | Dugway Proving Grounds |
| BACI051-N    | US Department of Defense reference strain, Dominant morphotype  | USAMRIID               |
| BACI051-E    | BACI051, Minor morphotype   | USAMRIID               |
| ATCC 9372-1  | <i>B. atrophaeus</i> var. <i>globigii</i> strain [14], derived from "Camp Detrick Red Strain", Minor morphotype | ATCC                   |
| ATCC 9372-2  | 9372, Dominant morphotype   | ATCC                   |
| 1013-1       | 1942 Vogel <i>in vitro</i> passaged variant   | U. Wisconsin           |
| 1013-2       | 1942 Vogel <i>in vitro</i> passaged variant   | U. Wisconsin           |
| ATCC 49822-1 | <i>B. atrophaeus</i> var. <i>globigii</i> strain, [14], Dominant morphotype                                     | ATCC                   |
| ATCC 49822-2 | 49822, Minor morphotype   | ATCC                   |
| NRS-356      | Archival isolate  | U. Wisconsin           |

doi:10.1371/journal.pone.0017836.t001

WGS versions described in this paper are the first versions, e.g. AEFM01000000.

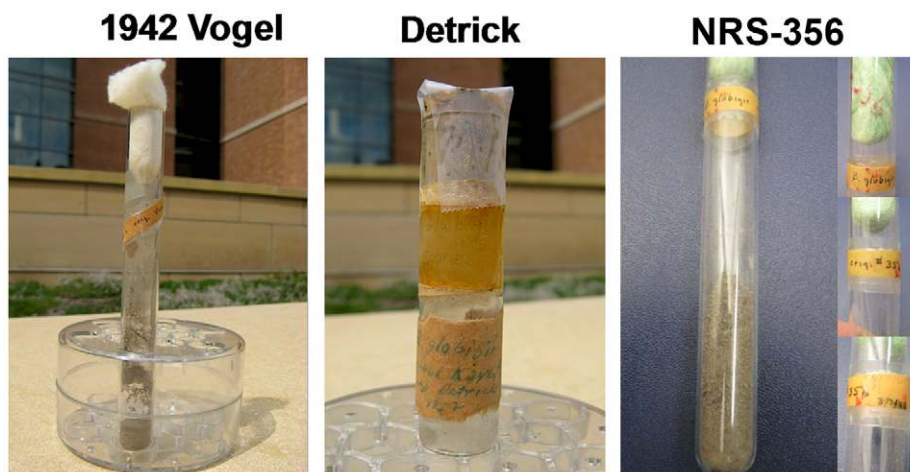
#### Identification of high-confidence mutations

Templated assembly of the remaining strains were mapped to the 1942 finished sequence using the GSMapper tool in Newbler (Roche). High-confidence mutations were selected from Newbler "HCDiffs" calls (Table S1) by applying additional selection criteria that mandated high quality scores in both reference and templated assemblies with >80% of the sequencing reads differing from the reference, elimination of mutation calls associated with homopolymer tracts (with the exception of tracts that were formed by a deletion – see below), and a minimum coverage depth of 5× with bidirectional sequence reads. Finally, the raw 454 reads from the 1942 isolate were mapped to the finished sequence to assess error

bias in the 454 process and to correct for residual sequencing errors in the finished sequence. Accession numbers of the relevant whole-genome shotgun sequences are found in Table 3. Phylogeny was calculated using PAUP 4.0b10. Fifty-eight nucleotide positions were used with gaps being treated as a "5th base" and all characters assuming equal weight. One thousand bootstrap replicates were computed using a heuristic search with the optimal criterion set to "parsimony". The tree was created using stepwise addition.

#### Confirmatory sequencing of SNP/Indels

Nineteen loci in which putative mutations were identified from the 454 dataset were re-sequenced from PCR products by standard Sanger dye-terminator methods. No false-negatives or false-positives were identified among the re-sequenced loci; however resequencing of the apparent mutation at position



**Figure 1. Archival samples of *B. atrophaeus* var. *globigii* ("*B. globigii*") from the University of Wisconsin Department of Bacteriology.** Samples had been maintained as suspensions of viable spores in sterile soil for approximately 60 years. The 1942 (left) and NRS-356 sample dated from 1944 (right) were found in the University of Wisconsin Department of Bacteriology strain collections. The 1952 sample (center) was returned to the Univ. of Wisconsin from Camp Detrick in 1952. doi:10.1371/journal.pone.0017836.g001

**Table 2.** Colony Morphology and Biochemical Data.

| Strain     | Morphology (LB) |           |              |       |        |          | Blood agar |            | Catalase |
|------------|-----------------|-----------|--------------|-------|--------|----------|------------|------------|----------|
|            | Form            | Elevation | Color        | Sheen | Size   | Margin   | Hemolysis  | Color      |          |
| 1942 Vogel | Circular        | Flat      | Orange       | Shiny | Normal | Undulate | ++         | Dark Brown | Y        |
| Detrick-1  | Circular        | Umbonate  | Orange       | Shiny | Normal | Undulate | ++         | Dark Brown | N        |
| Detrick-2  | Irregular       | Umbonate  | Orange       | Matte | Normal | Undulate | –          | Beige      | N        |
| Detrick-3  | Circular        | Convex    | Dark Orange  | Shiny | Small  | Undulate | +          | Brown      | N        |
| Dugway     | Circular        | Umbonate  | Orange       | Shiny | Normal | Undulate | –          | Beige      | N        |
| BAC1051-N  | Circular        | Raised    | Orange       | Matte | Normal | Undulate | –          | Beige      | N        |
| BAC1051-E  | Circular        | Raised    | Orange       | Matte | Normal | Undulate | +          | Beige      | N        |
| 9372-1     | Circular        | Raised    | Off-white    | Shiny | Normal | Undulate | –          | Grey       | N        |
| 9372-2     | Circular        | Raised    | Light orange | Shiny | Normal | Undulate | –          | Grey       | N        |
| 1013-1     | Irregular       | Raised    | Orange       | Matte | Normal | Undulate | –          | Tan        | Y        |
| 1013-2     | Circular        | Umbonate  | Off-white    | Matte | Normal | Undulate | +          | White      | Y        |
| 49822-1    | Circular        | Umbonate  | Orange       | Shiny | Normal | Undulate | +++        | ND         | Y        |
| 49822-2    | Circular        | Raised    | Orange       | Shiny | Normal | Undulate | +          | ND         | Y        |
| NRS-356    | Circular        | Flat      | Orange       | Shiny | Normal | Undulate | ++         | Dark Brown | Y        |

doi:10.1371/journal.pone.0017836.t002

1486408 revealed mixed genotypes in several isolates that are artifacts of a large duplication in the 1942 chromosome. Therefore, this signal cannot be considered a true SNP.

### Annotation, comparative genomic analysis, and multiple alignments

Preliminary annotations were generated using a combination of the RAST [22] algorithm (rast.nmpdr.org). Loci containing mutations were used to query the non-redundant (nr) databases and Refseq

protein databases at NCBI using directed BLASTx and BLASTp. The comparative BLAST tool from RAST was utilized for genome-wide protein sequence comparisons to *B. subtilis*. Results were filtered for bi-directional hits. Multiple alignments were generated by MegAlign from the LaserGene software package using the CUSTALW algorithm.

### Optical mapping

Genomic DNA was prepared from live bacteria on agar slants to maximize the yield of extremely high-molecular weight DNA.

**Table 3.** Genome Sequencing and *de novo* Assembly Statistics.

| Strain Name             | Median Depth Of Coverage | Number Of Large Contigs | N50 Contig Size | Largest Contig Size | Percent Q40 Plus Bases | Number of Contigs | Number Of Bases      | NCBI Accession number <sup>f</sup> | NCBI project ID |
|-------------------------|--------------------------|-------------------------|-----------------|---------------------|------------------------|-------------------|----------------------|------------------------------------|-----------------|
| <b>1942<sup>a</sup></b> | 32 <sup>a</sup>          | n/a                     | n/a             | n/a                 | 99.94 <sup>a</sup>     | n/a               | 4168266 <sup>b</sup> | CP002207                           | 46075           |
| <b>Detrick-1</b>        | 30                       | 33                      | 193257          | 809459              | 99.93                  | 38                | 4130179              | AEFP00000000                       | 46077           |
| <b>Detrick-2</b>        | 29                       | 50 <sup>c</sup>         | 152641          | 273988 <sup>c</sup> | 99.57                  | 67 <sup>c</sup>   | 4127403              | AEFQ00000000                       | 46079           |
| <b>Detrick-3</b>        | 42                       | 32                      | 193246          | 809488              | 99.96                  | 37                | 4131174              | AEFR00000000                       | 46081           |
| <b>Dugway</b>           | 49                       | 38                      | 205337          | 419767              | 99.85                  | 53                | 4130651              | AEFO00000000                       | 34819           |
| <b>ATCC 9372-1</b>      | 44                       | 29                      | 255796          | 943057              | 99.91                  | 36                | 4108235 <sup>d</sup> | AEFM00000000                       | 37683           |
| <b>ATCC 9372-2</b>      | 42                       | 34                      | 250623          | 546945              | 99.88                  | 40                | 4130094              | AEFU00000000                       | 46211           |
| <b>BAC1051-N</b>        | 47                       | 78                      | 86661           | 212948              | 99.95                  | 81                | 4130502              | AEFY00000000                       | 51595           |
| <b>BAC1051-E</b>        | 36                       | 35                      | 192166          | 630535              | 99.97                  | 42                | 4130076              | AEFX00000000                       | 48615           |
| <b>1013-1</b>           | 31                       | 30                      | 282215          | 914719              | 99.95                  | 39                | 4130375              | AEFS00000000                       | 46207           |
| <b>1013-2</b>           | 30                       | 33                      | 225488          | 512878              | 99.93                  | 40                | 4057611 <sup>e</sup> | AEFT00000000                       | 46209           |
| <b>ATCC 49822-1</b>     | 37                       | 35                      | 192016          | 638113              | 99.92                  | 50                | 4132165              | AEFV00000000                       | 46283           |
| <b>ATCC 49822-2</b>     | 34                       | 34                      | 213362          | 638120              | 99.89                  | 51                | 4135194              | AEFW00000000                       | 46285           |
| <b>NRS-356</b>          | 44                       | 42                      | 178925          | 527969              | 99.97                  | 44                | 4128298              | PENDING                            | PENDING         |

<sup>a</sup>Genome was finished and closed to a single contig. Depth of coverage and Q40+ metrics are for *de novo* assembly of 454 data only.

<sup>b</sup>454Draft sequences appear smaller than 1942 due to collapse of repeat regions.

<sup>c</sup>The average read length of this sample was considerably lower (239 versus >300) for other genomes in this dataset, resulting in larger number of contigs and decreased assembly quality.

<sup>d</sup>Deletion of 23 kb with clear join points (positions 4,022,138–4,045,817) verified by optical mapping (Figure 4B).

<sup>e</sup>Deletion of 73 kb with clear join points (positions 3,992,613 to 4,065,341) verified by optical mapping (Figure 4B).

doi:10.1371/journal.pone.0017836.t003

Optical maps were generated by digestion with *Nco*I of DNA arrayed linearly on glass slides and the resulting maps were aligned and compared with the MapSolver software package (OpGen, Inc., Gaithersburg MD).

### Information-based Genomic Distance (IBGD) analysis

Using an information-based method for genomic classification [23], the sequence contigs from BG isolates 1942, 1013-2 and 49822 were analyzed in order to map the phylogenetic relationships of these isolates to other *Bacillus* species. In this method, genomic content is characterized by the frequencies of occurrence of short n-mers contained within each sequence (n typically from 3 to 16). These n-mers are then rank ordered by genome. The pair-wise comparison of the rank of n-mers within two different genomes is then used to compute an information-based genetic distance (IBGD), where the sum of the differences in rank for all possible n-mers is weighted by an entropy factor that depends on the frequencies of occurrence of the respective n-mers in the two genomes. The pair-wise IBGD values are then used to construct a phylogenetic network [24]. *Bacilli* genomes were obtained from Genbank. This method for phylogenetic characterization enables computation even with the unassembled reads, and it can be applied to draft or partial genome sequence data, which was the case for the three *B. atrophaeus* genomes studied here.

### Phenotype microarray (PM) analysis

The first seven BG strains listed in Table 1 were streaked for single colonies on BHI plates and incubated at 33°C overnight, followed by subculturing a second time under the same conditions. Subsequently, cell suspensions were prepared according to Biolog specifications, with OD readings ranging between 0.35–0.45 at 600 nm. Biolog phenotypic microarray plates PM1 through PM20, were inoculated according to the manufacturer's specifications, and incubated at 37°C for 72 hours. Readings were taken every 15 minutes, and data processed by OmniLog Phenotype Microarray File Management/Kinetic Plot and Parametric modules. Two biological replicates of the experiment were conducted for each strain. PM1-10 contain single wells for each growth condition whereas PM11-20 contain quadruplicate wells for each condition.

### Statistical analyses and heatmap of phenotype microarray data

The area under the curve (AUC) values were computed by adding all OmniLog values at all time points for each of the 1200 distinct phenotypes produced from the OmniLog software. The AUC values from the two different biological replicates for each unique phenotype were averaged. The ratio for each AUC was calculated between the 6 query strains (Detrick-1, Detrick-2, Detrick-3, 1013-1, 1013-2, and Dugway) and reference parent strain (1942). For the purpose of visualization, 1920 phenotypes were included in the heatmap (i.e. this better represents the locations of the phenotypes which correspond to different modes of action categories). The same ratios were used for the phenotypes that have replicates. The ratio values were formatted as PM1 to PM20 for each strain across the columns and wells  $A_i$  to  $H_i$ , where  $i = 1$  to 12 for the rows. The results were plotted in a heatmap using R [25]. Positive growth wells are represented by green blocks while negative growth wells are represented by red blocks.

### Catalase assay

Catalase activity was assayed by spotting drops of hydrogen peroxide (3%) onto isolated colonies on LB agar plates. Colonies

were monitored for bubble formation, signifying the release of water and oxygen. A colony was considered to be catalase positive by observation of bubbles.

### Sporulation efficiency assays

Streaks of Detrick 1, Detrick 2, and 1013 strains were grown for two days on TSA plates containing SBA. Bacterial cell mass was scraped using an inoculating loop (1  $\mu$ l) from the streak and resuspended in PBS. Sporulation was evaluated by bright field phase-contrast microscopy. Phase-bright free spores and phase-dark vegetative cells were counted. Five representative viewing fields were counted from each strain for each experiment. This experiment was completed in triplicate by repeating once per day over the course of three consecutive days.

In order to compare the percent sporulation between Detrick 1 and Detrick 2, and Detrick 1 and 1013, a mixed analysis of variance (ANOVA) was used to complete the analysis. Strain and viewing field were evaluated as fixed factors, and replicate was included as a random factor. The natural log of the percent sporulation was taken to obtain a normal distribution of the residual error. Tukey's method was applied to compare the difference between the mean log percent sporulation.

## Results

### Historical investigations of BG provenance

We traced a potential provenance of the commonly used BG strains through an exhaustive search of the open literature and the archives of the University of Wisconsin, which suggested a possible lineage from which the "military" BG strains were derived. The original source of the strains were the collections at the University of Wisconsin during the 1930s and 1940s, from which the strains were transferred to Camp Detrick at the initiation of the US Army's BW program at the beginning of the Second World War [6,26]. At Camp Detrick, BG was used as a non-pathogenic surrogate in process development for spore-forming bacteria. It is tempting to speculate that the University of Wisconsin supplied BG to Porton Down: A note found in the archive of Dr. Baldwin's papers, dated February 19, 1943, contained an order from Dr. Fildes (presumably Sir Paul Fildes, a noted bacteriologist active in the British BW program at the time), for a batch of *B. subtilis* spores. It is not clear whether BG or *B. subtilis* subsp. *subtilis* was supplied, or whether this material was actually delivered. Unfortunately, original records describing in detail the maintenance of the strains during the period 1942–1955 were destroyed as per US Army policy at the time (Dr. Mark Wolcott, USAMRIID; personal communication), and the personnel who had first-hand knowledge of the strain passage histories and methods are deceased. Therefore, the actual source of the Camp Detrick isolates must be inferred from published work [6], limited available documentation (e.g. ATCC 9372) and the genome sequences presented here. From Camp Detrick the isolates were eventually transferred to ATCC as *B. subtilis* var. *niger* "red strain." The desire to maintain a phenotypically and genotypically uniform simulant throughout the biodefense community prompted us to elucidate whether significant phenotypic and/or genomic differences had accumulated in any of the commonly used isolates during the growth and transfer of strains to different institutions and to compare the isolates in broad use today to the so-called "Mil-Spec" strain (ATCC 9372). In contrast, the origin of ATCC 49822 prior to acquisition F. Young's laboratory (the depositor) is unclear.

We obtained isolates from archival spore suspensions in sterile soil from the University of Wisconsin with legible labels dating back

as far as 1942 (Figure 1; Table 1). These isolates included an archival stock dated 1942 that likely predated the transfer to Camp Detrick, as well as material that had been returned to the University of Wisconsin from Camp Detrick in 1952. A derivative of the 1942 strain that had been repeatedly passaged *in vitro* on agar slants over a period of approximately 30 years allowed us to compare the genomic signatures of deliberate selection with the effects of long-term *in vitro* passage. In addition, a sample of strain NRS-356 [13], which is mentioned as a possible parent strain in correspondence between various academic laboratories and Camp Detrick, was also obtained from the same source as the 1942 “Vogel” strain. These isolates were subsampled, germinated on LB plates, screened for colony morphology variation (see below). Genomic DNA was prepared from these isolates for sequencing.

### BG strains exhibit distinct colony morphologies

Upon initial plating of the archival and modern-day BG stocks, we noted distinct colony morphotypes for many of the strains, with some strains containing multiple variants (Figure 2, Table 2). Some of these morphotypes were consistent with those observed by Hayward *et al.* [6] who originally described the emergence of colony variants in “*B. globigii*.” As in the earlier report, individual morphotypes were stable and did not interconvert with high frequency (data not shown), suggesting that these morphotypes were the result of relatively rare chromosomal mutations, although 1013-1 occasionally threw off papillae in heavier streaks (not shown). Multiple morphotypes were noted for ATCC9372, ATCC 49822, Detrick, and 1013, while the archival 1942 isolate, the isolate obtained from Dugway Proving Ground (Dugway) and BACI051 appeared to be pure populations on LB. All strains tested positive for BG using Real Time-PCR primers specific to the *recF* gene (Methods S1) [27]. The appearance of multiple colony morphotypes even within single “strains” strongly suggested an as yet undescribed level of genetic diversity within these samples that likely affected the expression of cell-surface components and/or sporulation. The intra-strain colony morphology variation was particularly dramatic in the *in vitro* passaged 1013 and ATCC9372 isolates, in which one variant of each lineage had lost the production of color on LB or SB A plates (Figure 2), suggesting more dramatic alterations to the genome.

### Whole genome sequencing of BG isolates

Draft genome sequences were generated from several BG strains in our collection. A summary of the results from the sequenced isolates is indicated in Table 3. All of the “military” isolates (Detrick clones 1 through 3, BACI051, Dugway) were extremely closely related to each other and to both ATCC9372 variants. The ATCC isolates possessed additional mutations that were absent in the “military” isolates. The size of the finished and closed genome of *B. atrophaeus* var. *globigii* 1942 was 4,168,266 bp, and annotation using RAST [22] revealed 4433 features, including 4343 protein-coding sequences and 90 RNA molecules [28]. The preliminary annotations derived from RAST are available as Genbank .gbk files in the supplementary material.

### Bioinformatic analysis of sequence data

On average, the genome of *B. atrophaeus* is approximately 86% identical to *B. subtilis* on the nucleotide level, supporting its delineation as a distinct species and agreeing well with previous estimates [29]. Analysis of the IBGD using whole-genome sequences (N-mer length >4) supported the identification of *B. subtilis* 168 as the closest relative among sequenced bacterial genomes (Figure 3). For this particular case,  $n = 5$  (i.e., there were  $4^5 = 1024$  total 5-mers used to compute the IBGD). The IBGD

values were relatively insensitive to the choice of  $n$  over the range of 4–8. The three BG genomes analyzed grouped closely together, and our analysis of the *Bacillus*-wide phylogeny using IBGD revealed the phylogenetic distance of that *B. subtilis*/*B. atrophaeus* species from *B. anthracis*, supporting the inferences published elsewhere from rRNA sequence analysis (Figure 3) [30]. Primary amino acid sequences of RAST-annotated proteins are on average 72% (median 83%) identical between *B. atrophaeus* and *B. subtilis*. When only the proteins that yielded bidirectional BLAST hits in RAST are examined, the predicted proteome of *B. atrophaeus* is, on average, 83% identical (86% median) to *B. subtilis*.

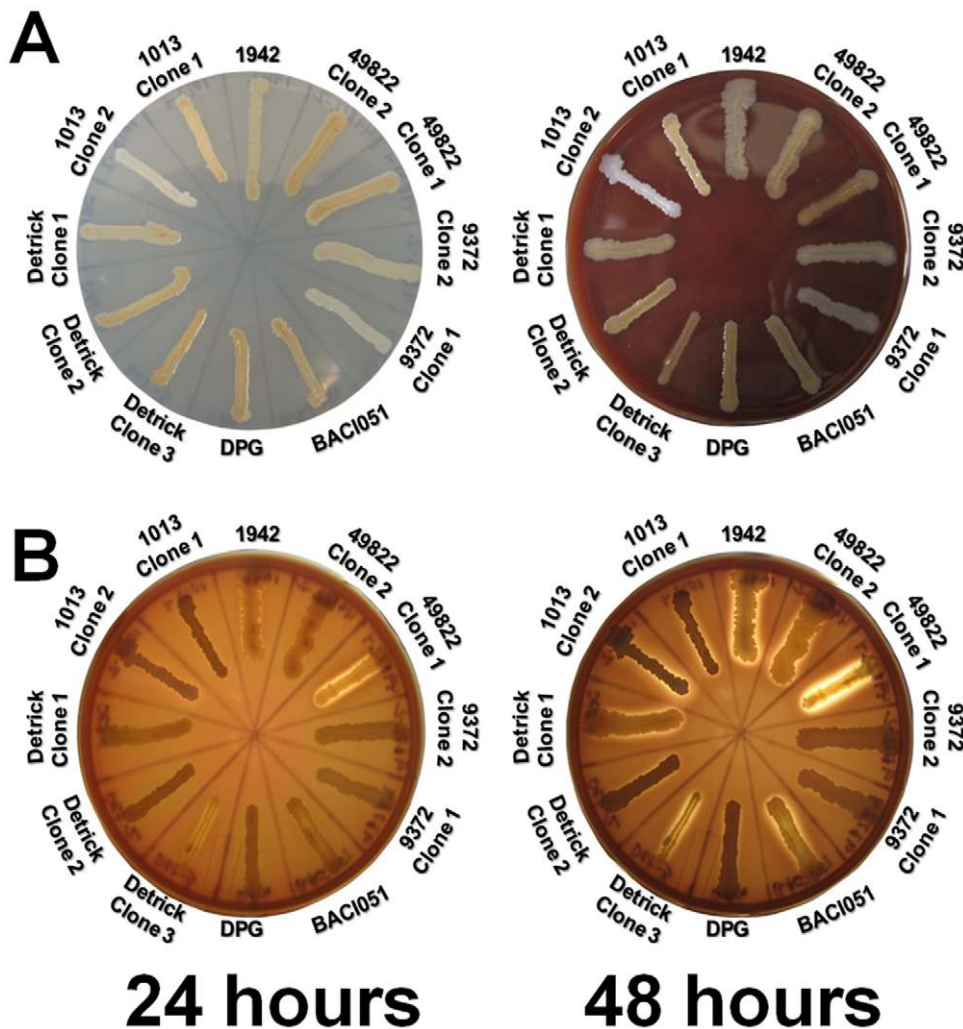
We utilized the finished sequence of the 1942 isolate as a reference strain for templated assembly of the remaining BG draft sequences. Two additional ATCC isolates of *B. atrophaeus* (49337 and 6537) were distinguishable from var. *globigii* on the basis of very high SNP/indel counts, lower coverage of and percentage of reads mapping to the 1942 reference, and unique genomic features which supported their proposed classification as var. *atrophaeus* [14]. The distinguishing genomic features of var. *atrophaeus* strains and the delineation of the *B. atrophaeus* clade from *B. subtilis* will be published elsewhere.

### Scaffolding of “military” BG genomes using optical maps

Optical restriction mapping [31,32,33] was used to compare the overall genomic structure of selected isolates. No differences in overall genome architecture between the “military” BG isolates, the archival 1942 isolate, or 1013-1 were observed (Figure 4, data not shown), suggesting that the global architecture of these isolates is relatively stable, even over 30 years of serial *in vitro* passage. However, the optical maps and sequence coverage analysis of 1013-2 and 9372-1 revealed substantial deletions of approximately 72,727 and 23,678 bases, respectively, of genomic material spanning from positions 3,992,613 to 4,065,341 (1013-2) or 4,022,138–4,045,817 (ATCC 9372-1) (Figure 4; Table S2). The genes within this deleted region are listed in Table S3 but notably contain genes encoding for nitrite reduction, germination (*gerKABC*), and biosynthesis of the lipopeptide surfactin (*sfvCAB*) [34,35]. A defect in surfactin production is a particularly intriguing candidate for the morphology and pigmentation variations in 1013-2 and ATCC 9372-1, since disruption of *sfvA* has been shown to have dramatic effects on spreading motility on semisolid media, on biofilm formation [34,36], and low-grade hemolytic activity.

### Mutation analysis of BG isolates

Using the *de novo* assembled draft sequence from the 1942 isolate as a template for subsequent analysis of SNPs and small indels in the other “military” isolates, we generated a list of high-confidence, discriminatory mutations that differentiate the strains (Figure 5A). The nature and annotation of the mutations are found in Table 4 and can be assigned an approximate temporal order in which they occurred (Figure 5B). Based on this analysis, 1942 is the most likely parental strain for all of the isolates in this study, with the 1013 lineage diverging earliest, followed by 49822, then the “military” lineage prior to the transfer to Camp Detrick. This conclusion is based on the observation that 49822 shares three SNPs with Detrick-1. The latter is the most likely progenitor of the other “military” isolates, since it has the fewest mutations relative to strain 1942. Detrick-1 can be differentiated from other “military” isolates by possessing the parental allele of *spo0F* rather than the H101R allele (position 3231470) that is characteristic of all of the other “military” BG isolates and the ATCC9372 strains. The two colony morphology variants of ATCC9372 each exhibited distinct mutation profiles indicating that the reference



**Figure 2. Appearance of *B. atrophaeus* strains on solid media.** A) Appearance of *B. atrophaeus* strains on LB or blood agar plates after 24 hours at 37°C. Plates were illuminated directly. B) β-Hemolysis of some *B. atrophaeus* strains. Transilluminated plates after 24 or 48 hours of growth on blood agar at 37°C.

doi:10.1371/journal.pone.0017836.g002

strain is in fact a mixed population of at least two genetically distinct substrains.

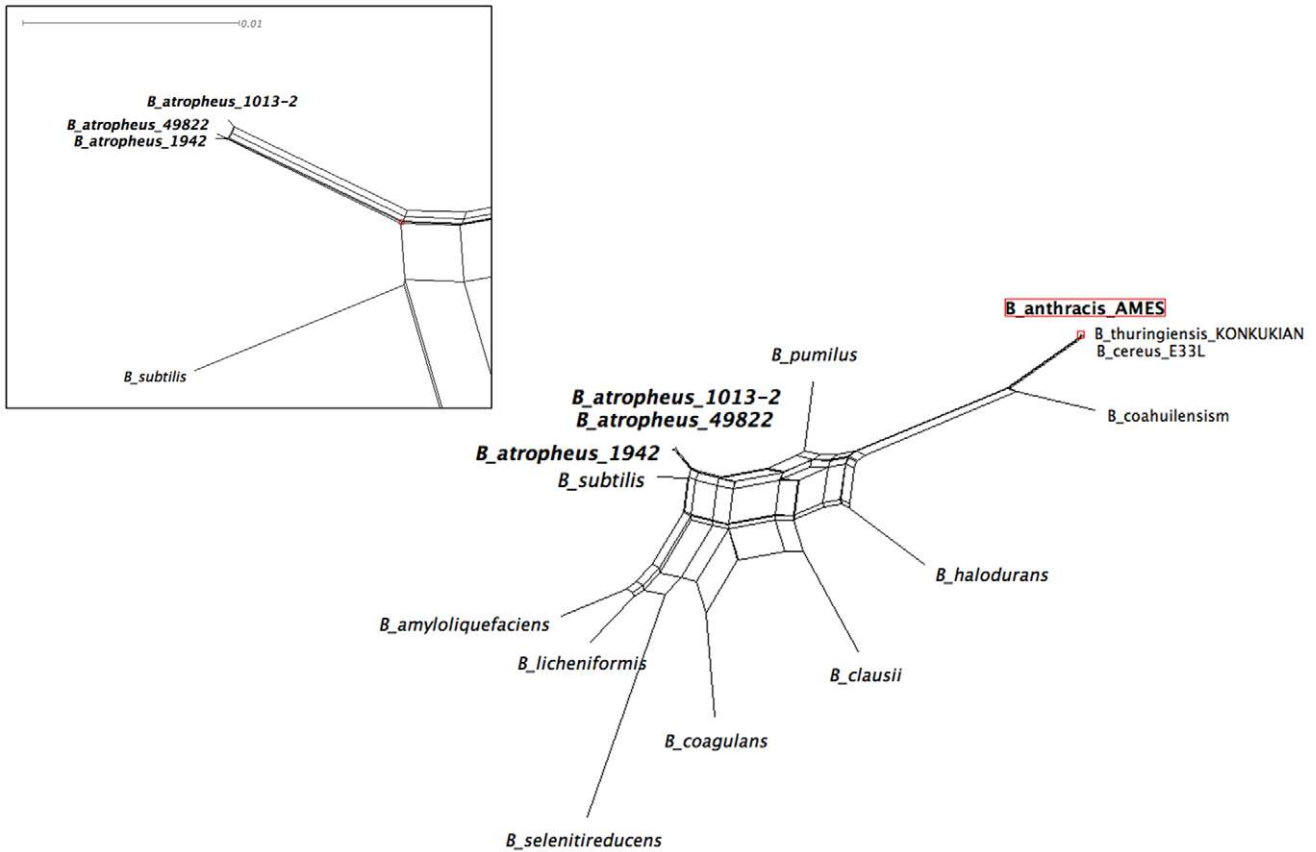
### Effects of genotype on strain phenotypes

The 72 kb deletion in 1013-2 included the structural genes for biosynthesis of surfactin, a cyclic lipopeptide with a mild hemolytic activity [34]. To test whether the “military” and *in vitro* passaged strains possessed low-grade hemolytic activity, we streaked these variants on rich agar media containing 5% sheep’s blood and looked for hemolysis. To our surprise, all strains exhibited striking variation in their coloration (Figure 2A), with the 1942, 9372-1 and Detrick-1 isolates considerably darker on blood agar than the other “military” and *in vitro* passaged isolates. In addition, on LB the 1013-2 and 9372-1 isolates appeared white and off-white, respectively. Pigmentation of *B. subtilis* colonies is associated with production of a melanin-like pigment by the CotA protein, a major component of the spore coat [37]. In addition to the variations in pigmentation, streaks of the 1942 and Detrick-1 isolates were consistently translucent under transillumination (Figure 2B). These zones of translucency are suggestive of weak β-hemolysis, which has previously been observed in *B. subtilis*

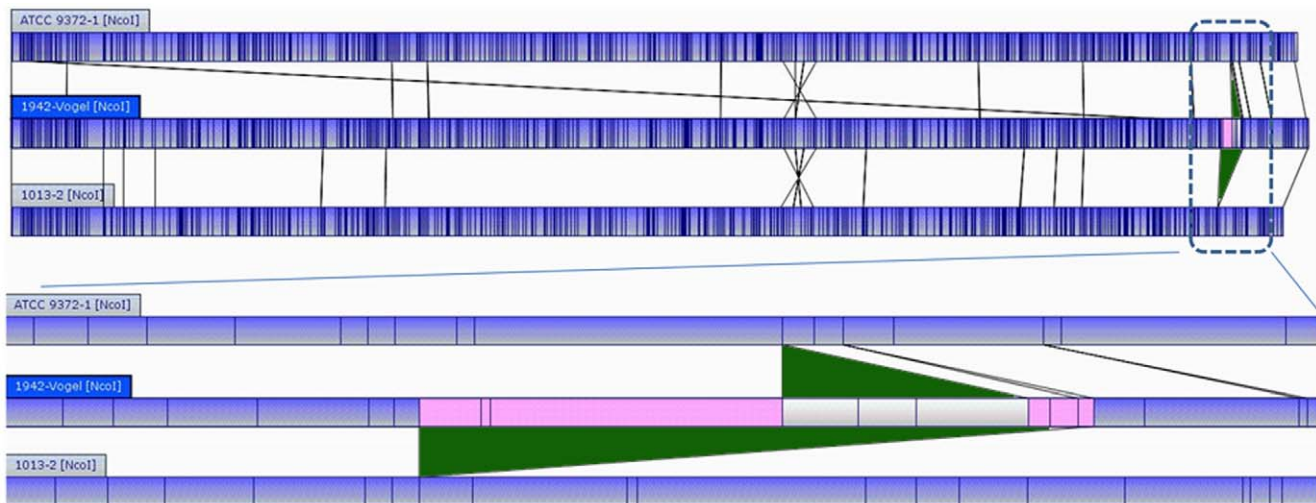
strains that produce high levels of surfactin [34,35,38]. The other strains exhibited either weak α-hemolysis or none at all, with the exception of the strongly hemolytic 49822-1 variant. At least in the “military” lineage, the quasi-hemolytic phenotype and dark-brown colony pigmentation correlated with the presence of a wild-type *spo0F* allele, suggesting that the ability of *B. atrophaeus* to lyse red blood cells may be regulated in part by *spo0F*. However this was not universally the case; the BACI051 strain had two discernible variants on SBA (not shown), one of which appeared to have recovered partial hemolytic activity (Figure 2B).

### BG strains have distinct metabolic profiles

To gain insight into the effects of genetic divergence of the adapted isolates on their metabolic capacity, the Detrick isolates and the separate 1013 isolates were compared by multiphenotype analysis using the Omnilog system, which allows the high-throughput comparison of 96×20 discrete growth conditions, including carbon, nitrogen, phosphorus, sulfate, nutrient supplements, pH, osmolytes as well as a broad class of growth inhibitors. The growth of the 1942 strain was used as a reference for determining relative growth rates of the other strains. The results of

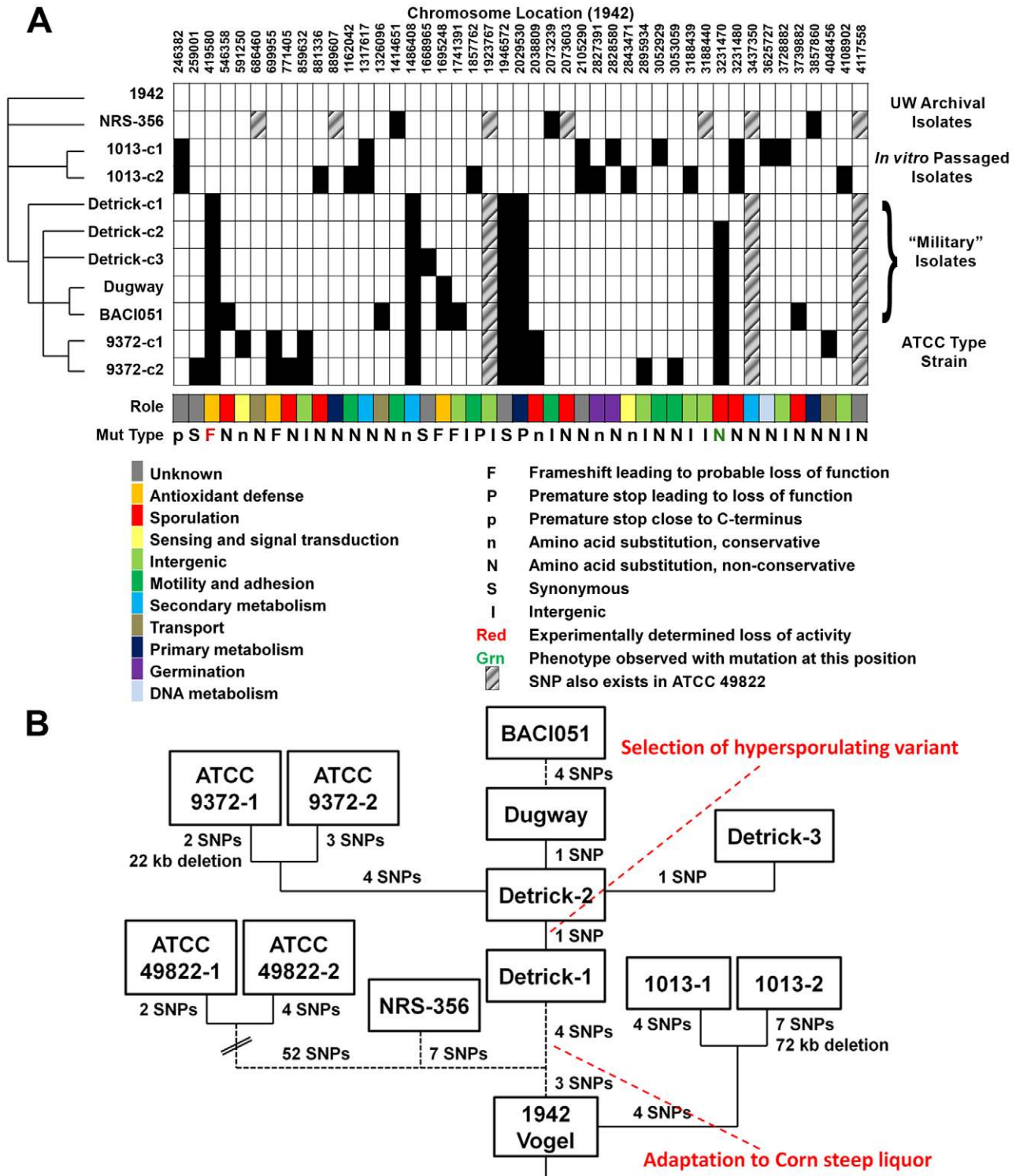


**Figure 3. Identification of *B. subtilis* as nearest-neighbor to *B. atropheus* var. *globigii* by whole-genome phylogenetic analysis of *Bacillus* genomes.** Information-based genomic distance (IBGD) was determined by comparing the relative distributions of n-mers within each genome to generate a pair-wise matrix of relative n-mer frequencies (see Materials and Methods). Variation of the n-mer length between 4 and 8 did not substantially affect the derived phylogeny. In this case n-mer length of 5 was utilized. For clarity, only three select species of the *B. cereus* group (of more than 30 that all cluster together) are labeled on the figure. The apparent divergence of isolate 1013-2 is due to alteration of the n-mer frequencies as a result of the deletion of 72 kb of genomic material.  
doi:10.1371/journal.pone.0017836.g003



**Figure 4. Optical mapping of *B. atropheus* var. *globigii* and detection of a 72 kb deletion.** **A)** Whole-genome consensus optical restriction maps (*NcoI*) of *B. atropheus* ATCC 9372-1 (Top), 1942 (middle) and 1013-2 (bottom) isolates. **B)** expanded view showing detail of the deleted regions in ATCC 9372-1 and 1013-2.  
doi:10.1371/journal.pone.0017836.g004





**Figure 5. Whole-genome mutation analysis and evolutionary history of the “military” lineage of *B. atrophaeus* var. *globigii*.** **A)** Whole-genome shotgun sequences of the other strains were mapped to the *de novo* assembled contigs of the 1942 strain using Newbler. Mutations exhibiting high quality scores in both reference and query sequences and with differences from the template exhibited in >85% of the individual sequencing reads are indicated as a blackened box. In one case (position 259001 in ATCC 9372-1) an initial false-negative due to the formation of a homopolymeric tract was found by direct inspection of the assemblies. The genes whose functions are altered by the given mutation are indicated in Table 4. **B)** Microevolutionary history of *B. atrophaeus* var. *globigii* strains. “Enhancement” events are indicated in red.  
doi:10.1371/journal.pone.0017836.g005

**Table 4.** SNP/Indel locations in *B. atrophaeus* genomes<sup>1</sup>.

| Genome location      | RAST Locus (NCBI Locus Tag #) | Types <sup>2</sup> | Gene affected <sup>3</sup>  | <i>B. subtilis</i> Homolog (% Identity) | Mutation       | AA change                      | Position in protein/total length |
|----------------------|-------------------------------|--------------------|---|---|----------------|--------------------------------|----------------------------------|
| 246382               | BG236 (01550)                 | PS                 | <i>yetF</i> – hypothetical protein  | BSU07140 (73.9)                         | C:T            | *                              | 217/231                          |
| 259001               | BG250 (12250)                 | SYN                | <i>yfnL</i> – hypothetical protein  | BSU07230 (91.5)                         | C:T            |                                |                                  |
| 419580               | BG405 (02000)                 | 1 bp INS           | <i>katA</i> – vegetative catalase   | BSU08820 (94.4)                         | +T             |                                | 197/483                          |
| 546358               | BG536 (02620)                 | NS                 | <i>hpr/scoC</i> – MarR family regulator (extracellular protease expression and sporulation) | BSU1000 (85.7)                          | C:G            | A:P                            | 13/200                           |
| 591250               | BG584 (02840)                 | NS                 | <i>hemAT</i> – heme-based aerotactic transducer   | BSU10380 (87.7)                         | C:T            | R:K                            | 215/432                          |
| 686460               | BG685 (03350)                 | NS                 | Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein OppA                 | BSU11430                                | T:G            | W:G                            | 520/545                          |
| 699955               | BG706 (03425)                 | PS                 | <i>yjbl</i> – truncated hemoglobin  | BSU11590 (96.1)                         | +C             | *                              | 89/132                           |
| 771405               | BG795 (03855)                 | NS                 | <i>rapA</i> – response regulator (Dephosphorylates Spo0F)                                   | BSU12430 (91.2)                         | T:G            | S:A                            | 261/378                          |
| 859632               | BG904/905 (04400/04405)       | IG                 |   |   | T:C            |                                |                                  |
| 881336               | BG929 (04515)                 | NS                 | <i>kinE</i> – 2-component histidine kinase; regulates Spo0F activity                        | BSU13530 (84.6)                         | C:T            | P:L                            | 577/739                          |
| 889607               | BG939 (04560)                 | NS                 | Methylthioribulose-1-phosphate dehydratase (EC 4.2.1.109), <i>mtnB</i>                      | BSU13610 (87.6)                         | A:G            | Y:C                            | 86/209                           |
| 1162042              | BG1225 (05960)                | INS                | <i>fljY</i> – flagellar switch motor protein  | BSU16320 (87.2)                         | +GTGGTC        | V <sub>3</sub> :V <sub>5</sub> | 355/379                          |
| 1317617              | BG1320 (06415)                | NS                 | <i>pksR</i> – Polyketide synthase   | BSU17720 (71.0)                         | G:C            | C:S                            | 1698/2574                        |
| 1326096              | BG1327 (06450)                | NS                 | <i>ebrB</i> – multidrug resistance family protein   | BSU1729 (77.8)                          | A:G            | S:P                            | 24/115                           |
| 1414651              | BG1226 (07000)                | NS                 | Chemotaxis regulator - transmits chemoreceptor signals to flagellar motor components CheY   | BSU17940 (86.0)                         | C:T            | A:V                            | 97/120                           |
| 1486408 <sup>4</sup> | BG1487 (07220)                | SYN                | <i>ppsA</i> – Plipeptin synthase/non-ribosomal peptide synthetase.                          | BSU18340                                | G:A            |                                | 1872/2560                        |
| 1668965              | BG1681 (08240)                | 1 bp DEL           | <i>yozB</i> – putative inner membrane protein   | BSU19140 (93.8)                         | –T             |                                | 55/178                           |
| 1695248              | BG1706 (08360)                | 5 bp INS           | <i>yojO</i> – hypothetical protein, potential nitricoxide synthase activation protein       | BSU19380 (87.5)                         | +GCTCT         |                                | 140/638                          |
| 1741391              | BG1764/1765 (08650/08655)     | IG                 |   |   | A:G            |                                |                                  |
| 1857762              | BG1903 (09335)                | PS                 | <i>ypfA</i> – possible pilZ homolog   | BSU22910 (60.1)                         | G:A            | *                              | 197/218                          |
| 1923767 <sup>5</sup> | BG1978/1979 (09750/09755)     | IG                 |   |   | C:T            |                                |                                  |
| 1946572              | BG1997 (09870)                | SYN                | Putative phage protein  | NA                                      | T:C            |                                | 283/584                          |
| 2029530              | BG2096 (10375)                | PS                 | 2-methylcitrate synthase  | BSU24140 (78.3)                         | +GA            | *                              | 197/218                          |
| 2038809              | BG2105 (10420)                | NS                 | <i>spo0A</i> – 2-component response regulator, controls initiation of sporulation           | BSU24220 (96.6)                         | G:A            | A:V                            | 225/265                          |
| 2073239              | BG2144/2145 (10615/10610)     | IG                 | Intergenic - possible promoter region of <i>sinI</i> gene (repressor of SinR)               |   | T:C            |                                |                                  |
| 2073603              | BG2146 (10620)                | NS                 | SinR, regulator of post-exponential-phase responses genes (competence and sporulation)      | BSU26100 (98.0%)                        | A:G            | K:E                            |                                  |
| 2105290              | BG2186 (10820)                | NS                 | <i>yqgE</i> – putative efflux transporter   | BSU25010 (92.5)                         | G:C            | H:D                            | 390/430                          |
| 2827391              | BG2944 (14450)                | NS                 | <i>gerAB</i> – spore germination protein  | BSU33060 (67.7)                         | G:A            | R:K                            | 35/365                           |
| 2828580              | BG2945 (14455)                | NS                 | <i>gerAC</i> – spore germination protein, governs germination in response to alanine        | BSU33070 (61.8)                         | G:T            | D:Y                            | 64/372                           |
| 2843471              | BG2960 (14530)                | NS                 | <i>yvrG</i> – sensor histidine kinase involved in cell wall processes                       | BSU33210 (75.4)                         | T:G            | I:L                            | 18/580                           |
| 2895934              | BG3014/3015 (14785/14790)     | IG                 |   |   | A:G            |                                |                                  |
| 3052929              | BG3176 (15575)                | NS                 | Flagellin   | NA                                      | A:G            | I:T                            | 279/320                          |
| 3053059              | BG3176 (15575)                | NS                 | Flagellin   | NA                                      | C:T            | A:T                            | 236/320                          |
| 3188439              | BG3318/3319 (16270/16275)     | IG                 | Intergenic  |   | 10 bp deletion |                                |                                  |

Table 4. Cont.

| Genome location      | RAST Locus (NCBI Locus Tag #)          | Types <sup>2</sup> | Gene affected <sup>3</sup>  | <i>B. subtilis</i> Homolog (% Identity) | Mutation | AA change | Position in protein/total length |
|----------------------|--|--------------------|---|---|----------|-----------|----------------------------------|
| 3188440              | BG3318/3319 (16270/16275)              | IG                 | Intergenic  | NA                                      | T:G      |           |                                  |
| 3231470              | BG3372 (16525)                         | NS                 | <i>Spo0F</i> – 2-component response regulator. Integrates signals from sensor kinases to trigger entry into sporulation | BSU37130 (98.4)                         | T:C      | H:R       | 101/124                          |
| 3231480              |  |                    |   |   | C:G      | A:P       | 98/124                           |
| 3437350 <sup>5</sup> | BG3610 (17625)                         | NS                 | Short chain dehydrogenase/reductase   | NA                                      | T:C      | T:A       | 149/247                          |
| 3625727              | BG3807 (18590)                         | NS                 | <i>trmE</i> – GTPase and tRNA-U34 5-formylation enzyme TrmE   | BSU41020 (93.3)                         | C:A      | D:Y       | 2/459                            |
| 3728882              | BG3897/3899 <sup>6</sup> (19040/19045) | IG                 |   |   | AA:TT    |           |                                  |
| 3739882              | BG3911 (19100)                         | NS                 | <i>sigH</i> ( <i>spo0H</i> ) RNA polymerase sigma factor  | BSU00980 (98.6)                         | C:T      | P:L       | 197/218                          |
| 3857860              | BG4037 (19695)                         | NS                 | Glycosyltransferase   | NA                                      | T:C      | G:R       | 374/402                          |
| 4048456              | BG4209 (20515)                         | NS                 | <i>yclF/dtpT</i> – putative di-tripeptide-proton ABC transporter  | BSU3670 (87.0)                          | A:G      | W:R       | 397/492 <sup>7</sup>             |
| 4108902              | BG4273/4274 (20825/20830)              | IG                 |   |   | C:A      |           |                                  |
| 4117558 <sup>5</sup> | BG4285 (20875)                         | NS                 | <i>ydaL</i> – uncharacterized membrane protein  | BSU04290 (76.5)                         | C:T      | P:L       | 9/574                            |

<sup>1</sup>In addition to the above mutations, putative SNPs at positions 39912, 45501, 61413, 611822, 1170114, 1251549, 2174033, 2771980, 3128882, 3269962, 3354410, 3528224, and 3880949 were common to all reference mapping experiments against the finished genome using 454 data, including the 1942 dataset. The commonality to all datasets suggest that these were errors in the finished sequence that are being verified.

<sup>2</sup>PS = Premature stop; SYN = Synonymous; INS = Insertion; NS = Non-synonymous; IG = Intergenic; DEL = Deletion.

<sup>3</sup>Annotations are a combination of RAST and directed tBLASTn and BLASTp searches vs *Bacillus* databases.

<sup>4</sup>Forms part of a large polypeptide synthase containing highly homologous regions.

<sup>5</sup>Also shared with strain ATCC 49822.

<sup>6</sup>The conserved start codon of the *radA* gene (BG3899) of *B. atrophaeus* and *B. subtilis* falls within the BG3898 ORF. Therefore BG3898 as called by RAST is not deemed likely to be a protein-coding gene.

<sup>7</sup>In putative transmembrane region.

doi:10.1371/journal.pone.0017836.t004

these experiments are summarized in Figure 6 and Table S4. In general, growth of the 1013 isolates was significantly diminished relative to the 1942 in many different growth conditions, most notably in the ability to use amino acids and peptides as carbon and nitrogen sources, to withstand osmotic stress, and to grow under reduced pH. In addition, the strains had developed sensitivity to beta-lactams, quinolones, and membrane-disrupting activities. These results suggested broad combined effects of several mutations on the phenotype of the strains. In addition to the *spo0F*(A98P) allele, which is a likely candidate for highly pleiotropic effects on the decision to sporulate under many different conditions, both strains contain substitutions in the *yetF* and *yqgE* genes that may be contributing to the phenotypes observed. The more pronounced defect in 1013-2 may be attributable to defects in the *gerAB* and *gerAC* genes and the large 72 kb deletion which contains several genes involved in germination.

By contrast, the Detrick isolates in general grew more robustly than the 1942 strain under multiple growth conditions. Increased relative growth rates were particularly pronounced for acidic conditions and media containing osmolytes, but particularly for wells containing sodium lactate [6].

Another isolate in the “military” lineage, Dugway, is clearly derived from the Detrick lineage by SNP/indel profiling yet has a metabolic profile that is much closer to the parental strain. Like the Detrick isolates, the Dugway strain grows better at low pH, but many of the other conditions do not promote elevated growth relative to 1942. Only one mutation differentiates that isolate from the Detrick-2 isolate – a 2-bp insertion in the *yajO* gene encoding a

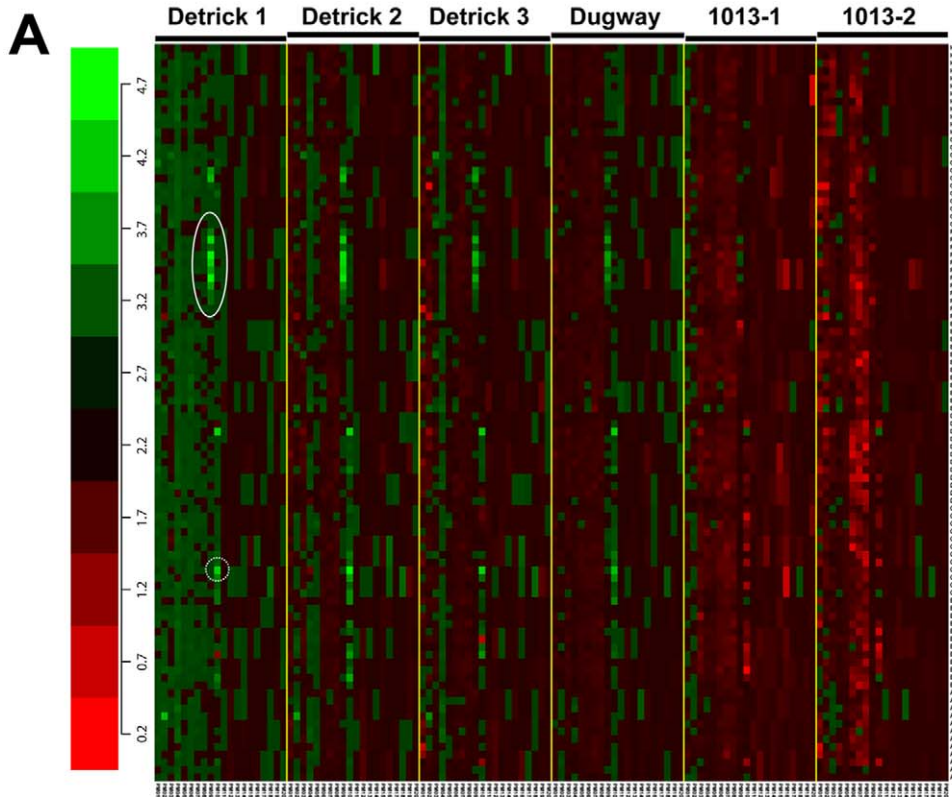
putative activator of nitric oxide (NO) synthesis. Again, the physiological role of this mutation is unclear, although nitric oxide synthesis plays a critical role in modulating antibiotic resistance in *Bacillus* spp. [39]. In addition to its role in promoting resistance to antibacterial drugs, NO is known to modulate *B. subtilis* genes involved in nitrate respiration when oxygen is limited [40]; thus the lowered growth in this strain may reflect the inability to grow to higher densities and overcome the resulting lower oxygen tension. An additional isolate, BACI051 is clearly derived from Dugway, yet two variants have accumulated additional mutations in *sigH* (*spo0H*), *hpr/scoC*, and *ebrB*. Notably, the phenotype of BACI051-E on plates more closely resembles the 1942 strain (Figure 2B).

### Catalase activity of BG

sequencing of the “military” isolates revealed a frameshift mutation in the *katA* gene encoding the major vegetative catalase [41]. The absence of catalase activity in “military” isolates was confirmed by adding a solution of 3% H<sub>2</sub>O<sub>2</sub> to smears of various strains. In contrast to the 1942 strain, which exhibited immediate and robust catalase activity, the strains containing the frameshift lacked this activity. A small amount of bubbling could be seen, probably due to the presence of a second catalase normally packaged in spores [42].

### Sporulation efficiency

To test whether the phenotype observed on blood agar was associated with differences in sporulation, selected strains were grown for two days as patches on blood agar, resuspended in PBS



**B**

| Strain    |           |           |        |        |        | Phenotype                         |
|-----------|-----------|-----------|--------|--------|--------|-----------------------------------|
| Detrick-1 | Detrick-2 | Detrick-3 | Dugway | 1013-1 | 1013-2 |                                   |
| 1.087     | 0.707     | 0.256     | 0.880  | 0.750  | 0.235  | Succinic Acid                     |
| 1.272     | 0.636     | 0.403     | 0.703  | 0.920  | 0.422  | α-Ketoglutaric Acid               |
| 1.185     | 0.612     | 0.246     | 0.747  | 0.663  | 0.657  | Citric Acid                       |
| 1.331     | 0.735     | 0.809     | 0.797  | 0.742  | 0.226  | Ala-Gly                           |
| 0.426     | 0.395     | 0.298     | 0.881  | 0.831  | 0.233  | D-Lactic Acid Methyl Ester        |
| 1.031     | 0.723     | 0.216     | 0.647  | 0.652  | 0.228  | L-Histidine                       |
| 0.829     | 0.473     | 0.638     | 0.601  | 0.384  | 0.452  | Ethylenediamine                   |
| 1.169     | 0.837     | 0.829     | 0.797  | 0.656  | 0.224  | Asn-Glu                           |
| 1.011     | 0.774     | 0.802     | 0.681  | 0.582  | 0.218  | Ser-Pro                           |
| 1.053     | 0.778     | 0.857     | 0.736  | 0.532  | 0.212  | Thr-Pro                           |
| 0.993     | 0.881     | 0.951     | 0.805  | 0.681  | 0.247  | Val-Glu                           |
| 0.964     | 0.916     | 0.895     | 0.781  | 0.269  | 0.634  | 7% Urea                           |
| 3.329     | 2.808     | 2.408     | 1.089  | 1.159  | 1.127  | Inulin                            |
| 4.715     | 4.272     | 3.126     | 2.328  | 1.295  | 0.791  | 6% Sodium Lactate                 |
| 2.873     | 3.612     | 1.582     | 1.689  | 0.871  | 0.810  | 7% Sodium Lactate                 |
| 3.673     | 3.924     | 2.963     | 2.708  | 1.122  | 0.848  | 8% Sodium Lactate                 |
| 4.497     | 4.119     | 3.182     | 2.813  | 1.297  | 0.848  | 9% Sodium Lactate                 |
| 3.987     | 3.613     | 2.896     | 2.701  | 1.394  | 0.890  | 11% Sodium Lactate                |
| 3.251     | 3.374     | 2.792     | 1.783  | 1.120  | 1.366  | 100mM Sodium Benzoate pH 5.2      |
| 2.693     | 2.716     | 2.396     | 1.028  | 0.917  | 1.150  | 200mM Sodium Benzoate pH 5.2      |
| 3.860     | 3.853     | 3.684     | 3.709  | 1.485  | 1.664  | pH 4.5 + D,L Diamino-Pimelic acid |
| 3.756     | 4.466     | 3.776     | 3.332  | 0.819  | 0.780  | pH 4.5 + L-Alanine                |
| 1.721     | 1.864     | 2.274     | 0.998  | 0.710  | 0.899  | D,L-α-Amino-Caprylic Acid         |
| 2.099     | 2.962     | 1.184     | 1.164  | 0.462  | 1.357  | pH 4.5 + L-Serine                 |
| 1.227     | 1.644     | 1.104     | 1.175  | 0.195  | 0.400  | pH 4.5 + L-Arginine               |
| 1.200     | 1.883     | 1.623     | 1.248  | 0.291  | 0.237  | pH 4.5 + L-Glutamic Acid          |
| 0.996     | 0.962     | 0.263     | 1.054  | 0.550  | 0.579  | pH 4.5 + L-Glutamine              |
| 1.894     | 2.371     | 1.016     | 2.208  | 1.009  | 0.923  | pH 4.5 + L-Lysine                 |
| 1.063     | 1.006     | 0.503     | 1.067  | 0.244  | 0.412  | pH 4.5 + L-Homoarginine           |
| 1.073     | 2.129     | 1.085     | 1.784  | 0.944  | 0.899  | pH 4.5 + α-Amino Malonate         |
| 1.597     | 2.067     | 1.915     | 0.987  | 0.562  | 0.901  | Sodium orthovanadate              |
| 0.871     | 0.750     | 0.830     | 0.974  | 0.309  | 0.563  | Sodium selenite                   |
| 1.366     | 1.367     | 0.916     | 2.016  | 1.169  | 1.055  | Hexachlorophene                   |

0.000 0.250 0.500 1.000 2.500 5.000  
Ratio AUC (Query strain/1942)

**Figure 6. Omnilog phenotypic arrays of *B. atrophaeus* subsp. *globigii* strains.** Six strains were each inoculated into twenty 96-well Omnilog plates and grown at 37°C. Reduction of tetrazolium dye by respiring cells was measured every 15 minutes by optical density. Dye reduction relative to the 1942 strain is shown; the red ratio values indicate less respiration while the green ratio values indicate more respiration as compared to the 1942 strain. Individual arrays or strains are displayed in each of the six major columns labeled Detrick 1, Detrick 2, Detrick 3, 1013-1, 1013-2, and Dugway. **A)** Heat map of all conditions for each strain. Each of the twenty plates for each strain is represented by the notation PM01-PM20 (left-to-right for each strain) along the x-axis. The rows represent the well position, and are denoted as A<sub>i</sub> to H<sub>i</sub> (i = 1 to 12) from the bottom to the top of the plot in each array along the y-axis. Each cell ratio value represents the average of two biological replicates for each strain. Plates PM01-PM10 contains single wells for each growth condition, while plates PM11-PM20 contain quadruplicate wells for each growth condition. Solid circle indicates wells containing sodium lactate; dotted circle indicates well containing L-serine at pH 4.5. The details of the 1920 growth conditions can be found in the first worksheet labeled “All strain AUC data” in Table S4. **B)** Most significant phenotypes for each of the six test strains as compared to the 1942 strain. The phenotypes with statistically significant increases and/or the decreases in ratio values for each of the six strains are presented. For the 1013 isolates only the conditions giving the five largest changes are presented. The number in each color block indicates the ratio for the test strain relative to the parent strain for the phenotype specified. The details of all significant phenotypes for each test strain can be obtained in Table S4. *Bold italic* font indicates p<0.05.  
doi:10.1371/journal.pone.0017836.g006

and counted directly. Strain Detrick-2 exhibited significantly higher percentages of phase-bright spores than the Detrick-1 strain (Figure 7, Mean +/- standard error of the mean). Similar results were observed for the 1942 and Dugway strains (not shown). The 1013-1 strain exhibited an even higher degree of sporulation than the Detrick-1 strain under identical conditions (Figure 7).

**Discussion**

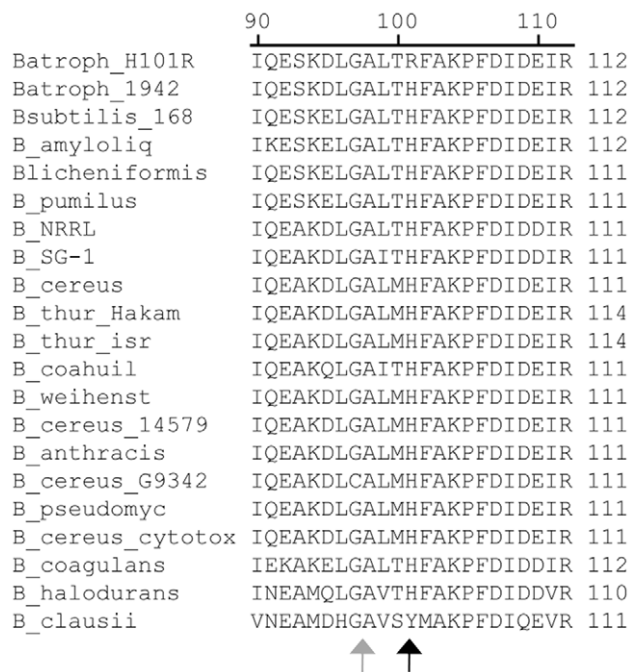
*Bacillus atrophaeus* has historically been grouped with *B. subtilis*, and is usually described as a black-pigmented variant (var. *niger*) because of its many phenotypic similarities to the better-characterized *B. subtilis*. Both organisms are soil-dwelling, non-pathogenic saprophytes, but have been differentiated by the ability

to produce pigment on nutrient media containing an organic nitrogen source [13]. The orange pigmentation of *B. atrophaeus* var. *globigii* spores made it an attractive simulant for *B. anthracis*, facilitating the detection of dispersed spores in complex environmental samples. Recently, more sensitive phylogenetic approaches using AFLP have delineated *B. atrophaeus* as a separate species [13,14]. The taxonomic confusion has arisen due to inadequately sensitive typing methods, and has led to misattribution of pathogenic qualities associated with some *B. licheniformis* strains to the *B. atrophaeus* strains currently in use as simulants [12], for which no direct evidence of pathogenicity exists. This report defines the genomic composition of *B. atrophaeus* var. *globigii* and clearly separates the species by whole-genome phylogenetic analysis.

In this study, we generated a high-quality, closed reference genome for the 1942 isolate using a combination of 454, Illumina, and directed Sanger sequencing. We expect the final genome to have an error rate of ~1 in 50,000 basepairs. When we mapped the 454 datasets for all of the isolates back to the finished sequence that was generated using the same DNA, we noted several putative SNPs that were common to all datasets (Table 4). We believe these represent errors introduced during generation of the final consensus sequence, as they did not appear when the isolates were mapped against draft sequence generated exclusively using the 454 platform; these are currently being verified and the final sequence will be updated.

Our sequences of multiple, closely related strains of this organism allow us to trace the derivation of the “military” BG isolates currently in use to a culture present at Camp Detrick during the 1940s and 1950s. The origin of ATCC 49822 is not as clear, but a publication from that era suggests a possible common origin at the University of Wisconsin [43]. While that strain is unlikely to be NRS-356 itself, given the presence of several strain-specific SNPs in our sequence, the SNPs common to both 49822 and the “military” lineage suggest a common ancestor that is not represented among the strains sequenced for this study. Given the lack of original records, it is unclear whether the NRS-356 variant in this study might have passed through Camp Detrick and been returned to the University of Wisconsin. However, given the date on the label and the general secrecy of operations at Camp Detrick during the Second World War [26] we consider this possibility unlikely.

During development of BG as a simulant for *B. anthracis*, strains were selected that exhibited the most desirable characteristics, those being rapid growth, high spore yield, and experimental reproducibility. Without being aware of the nature of the genetic alterations in their “optimized” strains, BW workers at Camp Detrick selected a mutant that provided dramatically higher total and relative spore yields, and generated consistent experimental results [6]. These strains were adopted into the inventories of numerous biodefense laboratories and have been used for many



**Figure 7. The *spo0F*(H101R) and *spo0F*(A98P) alleles are associated with hypersporulation.** Phase-contrast microscopy of BG strains after two days of growth on SBA. Vegetative cells appear as phase-dark rods, while spores appear as round, phase-bright globules. The mean percentage sporulation of each strain in a representative experiment is given ±SEM. The experiment was repeated on three consecutive days; representative results of a single experiment are shown. Statistical significance was determined by mixed ANOVA (Tukey’s method, p<0.05).  
doi:10.1371/journal.pone.0017836.g007

decades in simulations of decontamination and dispersal [12]. By applying a combination of genomic and biochemical profiling techniques, our data demonstrate that the BG isolates were “enhanced” by researchers at Camp Detrick during the development of the organism as a simulant.

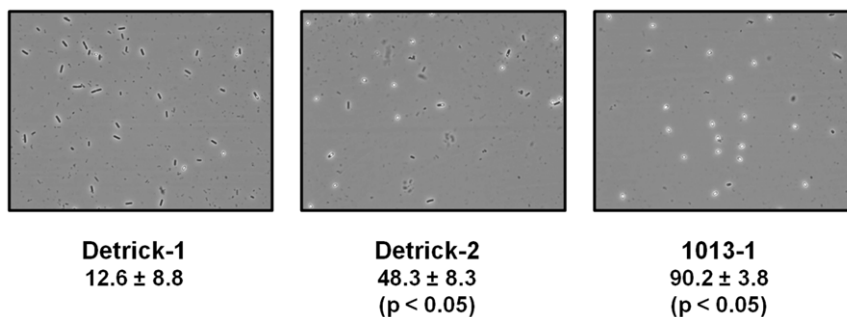
The selection of a strain with the desired properties appears to have occurred in at least two discrete steps, as shown by the genome sequences and metabolic profiles. The initial step appears to have been the adaptation of a strain to growth in corn steep liquor, an acidic medium rich in protein and lactate [44]. The robust growth of the Detrick strains relative to 1942 in low-pH medium containing high lactate levels is likely due to mutations in *mngD* (2-methylcitrate synthase, position 2029530), or a short-chain 3-oxoacyl-[acyl-carrier-protein] reductase (position 3437350), or both. The most likely candidate for a mutation in the Detrick isolates that increases growth is the frameshift in *mngD* that occurred following the divergence from the 49822 lineage and results in an altered C-terminus (Figure S1). The *mngD* gene encodes a 2-methylcitrate synthase that is expressed in the mother cell at the intermediate stages of sporulation [45]. A null mutation in *mngD* had no perceptible effect on sporulation, although other TCA-cycle enzymes when mutated led to a loss of sporulation [45]. The effects of the frameshift mutation on sporulation and cellular physiology on the function of the enzyme are not clear at this time. We speculate that the frameshift mutation alters the substrate specificity of MngD in favor of citrate, thus increasing the flux of lactate-derived intermediates through the tricarboxylic acid cycle. Evidence for this possibility includes the observations that 2-methylcitrate synthases can have partial citrate synthase activity [45] and that the *B. subtilis* *mngD* gene can complement a *glcA* (citrate synthase) mutant of *E. coli* [46]. Alternatively, alteration of function of *mngD* may have predisposed the lactate-adapted strain to acquisition of a hypersporulating phenotype, which is not readily isolated or stable in *B. subtilis* (see below); however the presence of a hypersporulating phenotype in an independently evolved lineage (1013) of BG indicates that the species may have an intrinsic predisposition to evolving such a phenotype *in vitro*.

The “military” strains also grow more readily on media containing *D,L*-diaminopimelic acid (*meso*-DAP), a major component of bacterial peptidoglycan. Corn steep liquor is derived from the incubation of corn in water at 42–55°C, during which a lactic fermentation by a community of wild organisms including numerous uncharacterized *Bacillus* spp. occurs. Total bacterial counts at the conclusion of CSL production can be quite high [44], thus the availability of such compounds for growth is not surprising. Another potential source of *meso*-DAP could be

bacterial autolysis during sporulation. The relative roles of each of the alleles in growth on lactate and/or *meso*-DAP is the subject of current investigation in our laboratory.

The second step in the development of BG as a simulant appears to have been the deliberate selection of a hypersporulating variant [6,47]. Importantly, the selection of a strain optimized for spore yield resulted in the fixation of a new *spo0F* allele that has no counterpart among the available *spo0F* sequences (Figure 8). The sole Spo0F sequence that differs at position 101 is that of *B. clausii*, in which tyrosine replaces histidine. Notably, the *spo0F(H101R)* mutation is distinct from a separate *spo0F(A98P)* mutation present in the *in vitro* passaged 1013 isolates. Given that the amino acid sequence of *B. atrophaeus* Spo0F is identical to that of *B. subtilis* but for two conservative substitutions, it is likely to have very similar if not identical biochemical properties. Detrick-1 and 1942 likely represent one of the two R colony morphotypes described by Hayward *et al.* [6], whereas the hypersporulating F morphotypes likely arose due to the emergence of the *spo0F(H101R)* mutation. However, the possibility that Detrick-1 represents a reversion mutant at this locus from Detrick-2 cannot formally be excluded, but since it represented the dominant morphotype in the 1952 Detrick vial we believe this is unlikely. The presence of the *spo0F(H101R)* allele in the ATCC 9372 strains suggests that these strains were acquired by ATCC after this mutation appeared within the Detrick lineage. Experiments to verify the roles of each allele in modulating sporulation are currently in progress. Preliminary results indicate that transformation of *B. subtilis*  $\Delta$ *spo0F* with *B. atrophaeus* DNA and selection of *spo+* cells dramatically alters colony morphology independently of the *spo0F* allele introduced; additional studies to verify the effects of each allele are currently in progress (James Hoch, personal communication).

The H101R and A98P alleles are likely to alter the response to signals promoting sporulation. A *spo0F(H101A)* allele results in a sporulation-proficient strain that throws off sporulation-deficient papillae [48], and the same mutation has been shown to suppress the *spo*<sup>-</sup> phenotype of a strain containing a defective *kinA* allele. H101 has been proposed as a potential metal-binding site with particular affinity for Cu<sup>2+</sup> [49]. Binding of Cu<sup>2+</sup> (or another divalent metal) at this site may modulate interaction with one or more sensor kinases that promote sporulation. Substitution of positively charged arginine at this position could potentially mimic the binding of a metal cation in the loop containing H101, resulting in altered sporulation of the strains due to a change in the interaction with the kinases governing sporulation. It is unclear why, given the proposed role of divalent Cu<sup>2+</sup> in suppressing



**Figure 8. Multiple alignment of Spo0F protein sequences.** The predicted protein sequences of Spo0F from multiple *Bacillus* species were aligned using ClustalW. Residues mutated in hypersporulating variants are indicated with grey (A98P) and black (H101R) arrows. Key: Batroph – *Bacillus atrophaeus*; Bsubtilis – *B. subtilis* 168; B\_amyloliq – *B. amyloliquefaciens*; B\_NRRL – *Bacillus* spp. NRRL; B\_SG-1 – *Bacillus* spp. SG-1; B\_thur – *B. thuringiensis* strains AI Hakam and var. Israelensis (Isr); B\_coahuil – *B. coahuilensis*; B\_weihenst – *B. weihenstephanensis*; B\_pseudomyc – *B. pseudomycoloides*.

doi:10.1371/journal.pone.0017836.g008

sporulation, H101R would result in a hypersporulation phenotype. The mechanistic relationship between *spo0F*(H101R) and the hypersporulation phenotype will be tested in future experiments.

Both variants in the 1013 lineage possess an A98P allele in *spo0F*. Although the presence of several other mutations within this lineage confounds the attribution of the hypersporulating phenotype to this allele at this time, the presence of a mutation in the same gene as another hypersporulating mutant is highly suggestive. The effect of proline substitution at position 98 on Spo0F function is not immediately obvious, but the relatively inflexible proline residue can disrupt alpha-helices in protein structures. The 1013-1 lineage exhibits a hypersporulating phenotype even more pronounced than *spo0F*(H101R) strains in the “military” lineage. The observation that hypersporulating phenotypes have emerged during cultivation of two independent *B. atrophaeus* lineages point to the possibility that certain *in vitro* selection pressures may actually favor hypersporulating variants.

The selection pressures acting on the sporulation pathway are highlighted by the sheer number of mutations discovered within the entire data set that occur in proteins known to play roles in sporulation. Nine of the 38 mutations (23%) found in all lineages were in genes that directly or indirectly regulate either entry into stationary phase or sporulation; this number exceeds the number that would be expected if mutations were to occur by chance, since less than 5% of *B. subtilis* genes are dedicated to regulatory processes of any kind [50,51]. In addition to the mutations found within the “military” lineage, the two variants of ATCC 49822 shown in Figure 2 differ by mutations in *rpoB* (Table S5) which also plays a role in entry into sporulation [52]. Null mutations in *spo0F* resulting in asporogenous phenotypes contribute to colony morphology variation in *B. anthracis*, *B. thuringiensis* and *B. subtilis* [53,54,55]. Enhanced *in vitro* “fitness” is also a likely driver behind the recovery of asporogenous *B. anthracis* mutants that were discovered during the investigation into the *B. anthracis* attacks of 2001 [56]. Because the process of sporulation is highly energy-intensive and irreversible once commenced, mutants that delay sporulation (or fail to sporulate altogether) to take advantage of remaining nutrients would out-compete wild-type cells during repeated passage *in vitro* in the absence of other selection pressures, as has been demonstrated in extended *in vitro* evolution studies with *B. subtilis* under relaxed sporulation conditions [57]. This may not be universally the case, since gain-of-function mutations in sporulation such as those observed in this study may compete favorably with wild-type cells if cannibalism of vegetative cells by sporulating bacteria is the dominant selective pressure [58]. Finally, horizontally transferred genetic elements can have dramatic effects on sporulation: for example, recent studies of phage lysogeny in *B. anthracis* have revealed the ability of several integrated phages to positively affect the kinetics of sporulation upon lysogeny of commonly used *B. anthracis* strains [59].

This study identifies the *spo0F*(H101R) allele as the signature of a deliberate selection during the development of *B. atrophaeus* as a simulant. However, without the knowledge of the history and the analysis of the phenotypes of the strains originating from “Camp Detrick” as published in the open literature, attribution of this genotype to a deliberate selection event would not have been definitive, since a similar phenotype is observed in the 1013 lineage which to our knowledge was not deliberately selected for any specific trait. Any study designed to determine genomic “signatures” of deliberate enhancement or selection is likely to require an analysis of the baseline likelihood that mutations conferring a similar phenotype would emerge and become fixed by natural processes within an evolutionary timeframe consistent with a known time interval or number of passages.

Available evidence suggests that hypersporulation is not easily evolved *in vitro*. Maughan and coworkers attempted to evolve populations of a laboratory strain of *B. subtilis* with a hypersporulating phenotype by repeatedly heat-shocking cultures. While their efforts to enrich for hypersporulators failed, other studies revealed that asporogenous mutants evolved readily [60,61], confirming many early studies ([62] and references therein). With the exception of the studies by Maughan *et al.*, most of these investigators applied selections intended to inhibit sporulation rather than to enrich for strains with elevated sporulation rates. The 1013 lineage was never heat-shocked during its many transfers; thus the adaptations seen in this work are the result of balancing sporulation versus vegetative growth for prolonged periods on agar slants. However, because undomesticated isolates were observed to sporulate to 98–100% [60], we cannot formally exclude the possibility that *in vitro* culture of the 1942 strain following its isolation for an unknown period by the University of Wisconsin might have selected for a hyposporulating variant. In this scenario, the H101R and A98P mutations would represent suppressor mutations. We consider this possibility unlikely, given the phenotypic similarity of two environmental isolates in the UW collection (1942 and NRS-356). Furthermore, a progression toward darker pigmentation and greater hemolysis is evident in the “military” lineage (Figure 2B). These phenotypic changes are associated with the accumulation of additional mutations including a P145L substitution mutation in *sigH*, a positive regulator of sporulation [63,64] and an A13P mutation in *scoC*, a negative regulator of sporulation [65]. Together, the strains analyzed in this study suggest strong selective pressures on the genes in the sporulation pathway, and more carefully controlled studies should be carried out to determine the dynamics of *in vitro* evolution and adaptation of spore-forming organisms, as has been done extensively in *E. coli* [66,67,68,69,70].

Unexpectedly, the “military” lineages were also marked by the loss of catalase activity, whose presence is an identifying feature of both *B. subtilis* and *B. atrophaeus* [13]. This activity was present in a separate lineage of *in vitro* passaged organisms, so it is not immediately clear why “military” isolates, i.e. those subjected to selection within the early days of the development of BG as a simulant organism, would have lost the catalase activity characteristic of the parental isolate. Because the KatA gene product is not found in spores [41,71], we consider it unlikely that the absence of this activity would impact the resistance of spores to decontamination reagents, and thus any antioxidant resistance phenotype exhibited by spores of “military” isolates would likely have gone unnoticed. However, direct comparisons of the “military” *B. atrophaeus* lineages to the progenitor strains have not been done, and pleiotropic effects of a *spo0F* mutation on spore physiology cannot currently be excluded.

Whole-genome approaches are becoming critical components of microbial forensics. The SNPs and indels identified in the analysis of evidentiary materials currently become the basis for higher-throughput assays to screen large numbers of samples [56,72]. Decreasing costs of whole-genome sequencing, and the comprehensive nature of the analysis, may make this the preferred method of forensic analysis of microbial samples in the future. With recently developed techniques of allele quantitation within populations by mass spectrometry [73], real-time PCR [74], and census-by-sequencing [68,75], it may be possible to quantitate accurately rare alleles within any given microbial population. We are particularly intrigued by the possibility that, given a mixture of different variants and sufficient sequencing power, ultra-high coverage sequencing may prove to be a more quantitative means of enumerating the relative populations in a sample even before the presence of variants has been established. The results from sequencing two strains of BACI051 in this study provide evidence of such hidden diversity.

The genomic basis of interlaboratory strain variation is only beginning to become evident, with recent studies tracing the histories of commonly used lab strains of *B. subtilis* 168, *E. coli*, *Salmonella enterica* serovar Typhimurium 14028s, *Pseudomonas aeruginosa* PA01 and *Mycobacterium tuberculosis* H37Rv [76,77,78,79,80,81]. These have revealed significant divergence of putatively identical strains from one laboratory to another, largely arising from mutations that accumulate during serial passage. Like the earlier work, our study highlights the utility of approaches based on whole-genome sequencing for the discrimination of closely related strains, especially when investigating the provenance for a given isolate. Tragically, at least 13 institutions are known to have destroyed archival collections of Select Agents [82] following the implementation of mandatory monitoring and reporting requirements, representing an incalculable loss of phenotypic and genomic diversity. This report underscores the importance of maintaining the genetic heritage preserved in the culture collections of individual investigators and institutions.

## Supporting Information

**Figure S1** Effect of frameshift mutation in the *mngD* gene on the C-terminus of the 2-methylcitrate synthase homolog of *B. atrophaeus* strain Detrick-1. Arrow indicates the location of the GA dinucleotide insertion. Multiple alignment was performed using the ClustalW algorithm in the MEGAlign module of LaserGene. (TIF)

**Table S1** Table S1 is a consolidated spreadsheet containing Newbler HCDiffs calls for each templated assembly to the finished sequence of the 1942 isolate. (XLSX)

**Table S2** Table S2 details the scaffolding of large contigs of the *de novo* assembly of 454 data for the 1942 strain based on optical maps. (XLSX)

## References

- Turnbough CL, Jr. (2003) Discovery of phage display peptide ligands for species-specific detection of *Bacillus* spores. *J Microbiol Methods* 53: 263–271.
- Sratis-Cullum DN, Griffin GD, Mobley J, Vass AA, Vo-Dinh T (2003) A miniature biochip system for detection of aerosolized *Bacillus globigii* spores. *Anal Chem* 75: 275–280.
- Kournikakis B, Ho J, Duncan S (2010) Anthrax letters: personal exposure, building contamination, and effectiveness of immediate mitigation measures. *J Occup Environ Hyg* 7: 71–79.
- Phillips CR (1949) The sterilizing action of gaseous ethylene oxide; sterilization of contaminated objects with ethylene oxide and related compounds; time, concentration and temperature relationships. *Am J Hyg* 50: 280–288.
- Sagripanti JL, Carrera M, Insalaco J, Ziemski M, Rogers J, et al. (2007) Virulent spores of *Bacillus anthracis* and other *Bacillus* species deposited on solid surfaces have similar sensitivity to chemical decontaminants. *J Appl Microbiol* 102: 11–21.
- 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.
- 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.
- Silva JM, Moreira AJ, Oliveira DC, Bonato CB, Mansano RD, et al. (2007) Comparative sterilization effectiveness of plasma in O<sub>2</sub>-H<sub>2</sub>O<sub>2</sub> mixtures and ethylene oxide treatment. *PDA J Pharm Sci Technol* 61: 204–210.
- United States Pharmacopeia (2000) Biological indicator for ethylene oxide sterilization, paper strip. The United States Pharmacopeia/The National Formulary. RockvilleMD: US Pharmacopeia. pp 231–232.
- Stein DC, Kopec LK, Yasbin RE, Young FE (1984) Characterization of *Bacillus subtilis* DSM704 and its production of 1-deoxynojirimycin. *Appl Environ Microbiol* 48: 280–284.
- Harris-Warrick RM, Lederberg J (1978) Interspecies transformation in *Bacillus*: sequence heterology as the major barrier. *J Bacteriol* 133: 1237–1245.
- Page WF, Young HA, Crawford HM, Institute of Medicine (U.S.). Advisory Panel for the Study of Long-Term Health Effects of Participation in Project SHAD (2007) Long-term health effects of participation in Project SHAD (Shipboard Hazard and Defense). Washington, D.C.: National Academies Press xviii, 123 p.
- Nakamura LK (1989) Taxonomic Relationship of Black-Pigmented *Bacillus subtilis* Strains and a Proposal for *Bacillus atrophaeus* sp. nov. *Int J Syst Bacteriol* 39: 295–300.
- Burke SA, Wright JD, Robinson MK, Bronk BV, Warren RL (2004) Detection of molecular diversity in *Bacillus atrophaeus* by amplified fragment length polymorphism analysis. *Appl Environ Microbiol* 70: 2786–2790.
- Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376–380.
- Bennett S (2004) Solexa Ltd. *Pharmacogenomics* 5: 433–438.
- Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18: 821–829.
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8: 186–194.
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175–185.
- Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* 8: 195–202.
- Han CS, Chain P. Finishing repeat regions automatically with DupFinisher. In: Arabnia HR, Valafar H, eds. 2006. CSREA Press. pp 141–146.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9: 75.
- Yang AC, Goldberger AL, Peng CK (2005) Genomic classification using an information-based similarity index: application to the SARS coronavirus. *J Comput Biol* 12: 1103–1116.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23: 254–267.

**Table S3** Table S3 contains RAST annotation files (.gtf format) of the 1942 strain and indicates the location of the large deletions in ATCC 9372-1 and 1013-2. (XLSX)

**Table S4** Table S4 contains Omnilog phenotypic array data normalized to 1942 strain. (XLSX)

**Table S5** Table S5 contains HCDiffs calls from templated assembly of ATCC 49822 variants using the finished sequence of the 1942 isolate. (XLSX)

**Methods S1** Confirmation of BG Identity by RT-PCR. (DOCX)

## Acknowledgements

We thank Dr. Kevin P. O’Connell for helpful discussions and insights into the manuscript and for facilitating the collaboration with the University of Wisconsin. We also thank Kristin Willner and Amy Butanifor assistance with sequencing. We thank Gary Ouellette for help with phylogenetic analysis and Drs. Mark Wolcott (USAMRIID) and James Hoch (Scripps) for helpful discussions. The opinions presented here are those of the authors and are not necessarily those of the U.S. Government or any of its agencies. Information in this report is unclassified and cleared for public release.

## Author Contributions

Conceived and designed the experiments: HSG TDR EWS SS CD. Performed the experiments: SMB C. Hong CC AA HD DB RT SJ C. Han PAM LAM M. Karavis KAB-L SS. Analyzed the data: HSG LAM HD CC M. Karavis KHP AF LAM HD CC M. Krepps KHP AF JSL WEC BWH PD AL EF RT SJ KAB-L CNR. Contributed reagents/materials/analysis tools: JL. Wrote the manuscript: HSG.



25. Team RDC (2009) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
26. Regis E (1999) The Biology of Doom: the History of America's Secret Germ Warfare Project. New York: Henry Holt. 259 p.
27. Buttner MP, Cruz P, Stetzenbach LD, Klima-Comba AK, Stevens VL, et al. (2004) Evaluation of the Biological Sampling Kit (BiSKit) for large-area surface sampling. *Appl Environ Microbiol* 70: 7040–7045.
28. McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, et al. (2007) The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation. *Nucleic Acids Res* 35: D347–353.
29. Fritze D, Pukall R (2001) Reclassification of bioindicator strains *Bacillus subtilis* DSM 675 and *Bacillus subtilis* DSM 2277 as *Bacillus atrophaeus*. *Int J Syst Evol Microbiol* 51: 35–37.
30. Greenberg DL, Busch JD, Keim P, Wagner DM (2010) Identifying experimental surrogates for *Bacillus anthracis* spores: a review. *Investig Genet* 1: 4.
31. Latreille P, Norton S, Goldman BS, Henkhaus J, Miller N, et al. (2007) Optical mapping as a routine tool for bacterial genome sequence finishing. *BMC Genomics* 8: 321.
32. Kotewicz ML, Mammel MK, LeClerc JE, Cebula TA (2008) Optical mapping and 454 sequencing of *Escherichia coli* O157 : H7 isolates linked to the US 2006 spinach-associated outbreak. *Microbiology* 154: 3518–3528.
33. Kotewicz ML, Jackson SA, LeClerc JE, Cebula TA (2007) Optical maps distinguish individual strains of *Escherichia coli* O157 : H7. *Microbiology* 153: 1720–1733.
34. Nakano MM, Marahiel MA, Zuber P (1988) Identification of a genetic locus required for biosynthesis of the lipopeptide antibiotic surfactin in *Bacillus subtilis*. *J Bacteriol* 170: 5662–5668.
35. Nakano MM, Magnuson R, Myers A, Curry J, Grossman AD, et al. (1991) *sfyA* is an operon required for surfactin production, competence development, and efficient sporulation in *Bacillus subtilis*. *J Bacteriol* 173: 1770–1778.
36. Kinsinger RF, Shirk MC, Fall R (2003) Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. *J Bacteriol* 185: 5627–5631.
37. Hullo MF, Moszer I, Danchin A, Martin-Verstraete I (2001) CotA of *Bacillus subtilis* is a copper-dependent laccase. *J Bacteriol* 183: 5426–5430.
38. Nakano MM, Zuber P (1989) Cloning and characterization of *sfyB*, a regulatory gene involved in surfactin production and competence in *Bacillus subtilis*. *J Bacteriol* 171: 5347–5353.
39. Gusarov I, Shatalin K, Starodubtseva M, Nudler E (2009) Endogenous nitric oxide protects bacteria against a wide spectrum of antibiotics. *Science* 325: 1380–1384.
40. Nakano MM (2002) Induction of ResDE-dependent gene expression in *Bacillus subtilis* in response to nitric oxide and nitrosative stress. *J Bacteriol* 184: 1783–1787.
41. Casillas-Martinez L, Setlow P (1997) Alkyl hydroperoxide reductase, catalase, MrgA, and superoxide dismutase are not involved in resistance of *Bacillus subtilis* spores to heat or oxidizing agents. *J Bacteriol* 179: 7420–7425.
42. Bagyan I, Casillas-Martinez L, Setlow P (1998) The *katX* gene, which codes for the catalase in spores of *Bacillus subtilis*, is a forespore-specific gene controlled by sigmaF, and KatX is essential for hydrogen peroxide resistance of the germinating spore. *J Bacteriol* 180: 2057–2062.
43. Young FE, Tipper DJ, Strominger JL (1964) Autolysis of Cell Walls of *Bacillus subtilis*. *Journal of Biological Chemistry* 239: PC3600–PC3602.
44. Liggett RW, Koffler H (1948) Corn steep liquor in microbiology. *Bacteriol Rev* 12: 297–311.
45. Bryan E, Beall B, Moran C, Jr. (1996) A sigma E dependent operon subject to catabolite repression during sporulation in *Bacillus subtilis*. *J Bacteriol* 178: 4778–4786.
46. Gerike U, Hough DW, Russell NJ, Dyal-Smith ML, Danson MJ (1998) Citrate synthase and 2-methylcitrate synthase: structural, functional and evolutionary relationships. *Microbiology* 144 (Pt 4): 929–935.
47. Roth NG (1955) Development of Processes for the Production and Use of *Bacillus globigii* Spores as a Simulant. *Camp Detrick*.
48. Tzeng YL, Hoch JA (1997) Molecular recognition in signal transduction: the interaction surfaces of the Spo0F response regulator with its cognate phosphorelay proteins revealed by alanine scanning mutagenesis. *J Mol Biol* 272: 200–212.
49. Kojetin DJ, Thompson RJ, Benson LM, Naylor S, Waterman J, et al. (2005) Structural analysis of divalent metals binding to the *Bacillus subtilis* response regulator Spo0F: the possibility for in vitro metalloregulation in the initiation of sporulation. *Biometals* 18: 449–466.
50. Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249–256.
51. Piggot PJ, Hilbert DW (2004) Sporulation of *Bacillus subtilis*. *Curr Opin Microbiol* 7: 579–586.
52. Asai K, Inaoka T, Nanamiya H, Sadaie Y, Ochi K, et al. (2007) Isolation and characterization of sporulation-initiation mutation in the *Bacillus subtilis* *prfB* gene. *Biosci Biotechnol Biochem* 71: 397–406.
53. Takahashi I (1965) Transduction of Sporogenesis in *Bacillus subtilis*. *J Bacteriol* 89: 294–298.
54. Chemerilova VI, Sekerina OA, Talalaeva GB (2007) [Analysis of the morphological variants arising during S <-> R dissociation in *Bacillus thuringiensis*]. *Mikrobiologiya* 76: 507–514.
55. Worsham PL, Sowers MR (1999) Isolation of an asporogenic (*spo0A*) protective antigen-producing strain of *Bacillus anthracis*. *Can J Microbiol* 45: 1–8.
56. Cummings CABC, C.A., Fang R, Barker M, Brzoska PM, Williamson P, Beudry JA, Matthews M, Schupp JM, Wagner DM, Furtado MR, Keim P, Budowle B (2009) Whole-genome typing of *Bacillus anthracis* isolates by next-generation sequencing accurately and rapidly identifies strain-specific diagnostic polymorphisms. *Forensic Sci Int* 2: 300–301.
57. Maughan H, Masel J, Birky CW, Jr., Nicholson WL (2007) The roles of mutation accumulation and selection in loss of sporulation in experimental populations of *Bacillus subtilis*. *Genetics* 177: 937–948.
58. Gonzalez-Pastor JE, Hobbs EC, Losick R (2003) Cannibalism by sporulating bacteria. *Science* 301: 510–513.
59. Schuch R, Fischetti VA (2009) The secret life of the anthrax agent *Bacillus anthracis*: bacteriophage-mediated ecological adaptations. *PLoS One* 4: e6532.
60. Maughan H, Nicholson WL (2004) Stochastic processes influence stationary-phase decisions in *Bacillus subtilis*. *J Bacteriol* 186: 2212–2214.
61. Maughan H, Callicotte V, Hancock A, Birky CW, Jr., Nicholson WL, et al. (2006) The population genetics of phenotypic deterioration in experimental populations of *Bacillus subtilis*. *Evolution* 60: 686–695.
62. Brunstetter BC, Magoon CA (1932) Studies on Bacterial Spores: III. A Contribution to the Physiology of Spore Production in *Bacillus mycoides*. *J Bacteriol* 24: 85–122.
63. Bai U, Lewandoski M, Dubnau E, Smith I (1990) Temporal regulation of the *Bacillus subtilis* early sporulation gene *spo0F*. *J Bacteriol* 172: 5432–5439.
64. Predich M, Nair G, Smith I (1992) *Bacillus subtilis* early sporulation genes *kinA*, *spo0F*, and *spo0A* are transcribed by the RNA polymerase containing sigma H. *J Bacteriol* 174: 2771–2778.
65. Perego M, Hoch JA (1988) Sequence analysis and regulation of the *hpr* locus, a regulatory gene for protease production and sporulation in *Bacillus subtilis*. *J Bacteriol* 170: 2560–2567.
66. Cooper TF, Lenski RE (2010) Experimental evolution with *E. coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol Biol* 10: 11.
67. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, et al. (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1247.
68. Barrick JE, Lenski RE (2009) Genome-wide Mutational Diversity in an Evolving Population of *Escherichia coli*. *Cold Spring Harb Symp Quant Biol*.
69. Ferenci T (2008) The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of *rpoS* mutations. *Heredity* 100: 446–452.
70. Maharjan R, Setto S, Notley-McRobb L, Ferenci T (2006) Clonal adaptive radiation in a constant environment. *Science* 313: 514–517.
71. Liu H, Bergman NH, Thomason B, Shallom S, Hazen A, et al. (2004) Formation and composition of the *Bacillus anthracis* endospore. *J Bacteriol* 186: 164–178.
72. Van Ert MN, Easterday WR, Simonson TS, U'Ren JM, Pearson T, et al. (2007) Strain-specific single-nucleotide polymorphism assays for the *Bacillus anthracis* Ames strain. *J Clin Microbiol* 45: 47–53.
73. Thomas RK, Baker AC, Debiase RM, Winckler W, Laframboise T, et al. (2007) High-throughput oncogene mutation profiling in human cancer. *Nat Genet* 39: 347–351.
74. Liu CM, Driebe EM, Schupp J, Kelley E, Nguyen JT, et al. (2010) Rapid quantification of single-nucleotide mutations in mixed influenza A viral populations using allele-specific mixture analysis. *J Virol Methods* 163: 109–115.
75. Holt KE, Teo YY, Li H, Nair S, Dougan G, et al. (2009) Detecting SNPs and estimating allele frequencies in clonal bacterial populations by sequencing pooled DNA. *Bioinformatics* 25: 2074–2075.
76. Srivatsan A, Han Y, Peng J, Tehranchi AK, Gibbs R, et al. (2008) High-Precision, Whole-Genome Sequencing of Laboratory Strains Facilitates Genetic Studies. *PLoS Genet* 4: e1000139.
77. Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF (2009) Tracing ancestors and relatives of *Escherichia coli* B, and the derivation of B strains REL606 and BL21(DE3). *J Mol Biol* 394: 634–643.
78. Klockgether J, Munder A, Neugebauer J, Davenport CF, Stanke F, et al. (2009) Genome Diversity of *Pseudomonas aeruginosa* PAO1 laboratory strains. *J Bacteriol*. pp JB.01515–01509.
79. Ferenci T, Zhou Z, Betteridge T, Ren Y, Liu Y, et al. (2009) Genomic sequencing reveals regulatory mutations and recombinational events in the widely used MC4100 lineage of *Escherichia coli* K-12. *J Bacteriol* 191: 4025–4029.
80. Jarvik T, Smillie C, Groisman EA, Ochman H (2010) Short-Term Signatures of Evolutionary Change in the *Salmonella enterica* Serovar *Typhimurium* 14028 Genome. *J Bacteriol* 192: 560–567.
81. Ioerger TR, Feng Y, Ganesula K, Chen X, Dobos KM, et al. (2010) Variation Among Genome Sequences of H37Rv Strains of *M. tuberculosis* from Multiple Laboratories. *J Bacteriol*. pp JB.00166–00110.
82. Casadevall A, Imperiale MJ (2010) Destruction of Microbial Collections in Response to Select Agent and Toxin List Regulations. *Biosecurity and Biodefense: Biodefense Strategy, Practice, and Science* 8: 151–154.