



In-Depth Analysis of *Bacillus anthracis* 16S rRNA Genes and Transcripts Reveals Intra- and Intergenomic Diversity and Facilitates Anthrax Detection

Peter Braun,^a Fee Zimmermann,^a  Mathias C. Walter,^a Sonja Mantel,^a Karin Aistleitner,[†] Inga Stürz,^a  Gregor Grass,^a  Kilian Stoecker^a

^aBundeswehr Institute of Microbiology, Munich, Germany

Gregor Grass and Kilian Stoecker contributed equally.

ABSTRACT Analysis of 16S rRNA (rRNA) genes provides a central means of taxonomic classification of bacterial species. Based on presumed sequence identity among species of the *Bacillus cereus sensu lato* group, the 16S rRNA genes of *B. anthracis* have been considered unsuitable for diagnosis of the anthrax pathogen. With the recent identification of a single nucleotide polymorphism in some 16S rRNA gene copies, specific identification of *B. anthracis* becomes feasible. Here, we designed and evaluated a set of *in situ*, *in vitro*, and *in silico* assays to assess the unknown 16S state of *B. anthracis* from different perspectives. Using a combination of digital PCR, fluorescence *in situ* hybridization, long-read genome sequencing, and bioinformatics, we were able to detect and quantify a unique 16S rRNA gene allele of *B. anthracis* (16S-BA-allele). This allele was found in all available *B. anthracis* genomes and may facilitate differentiation of the pathogen from any close relative. Bioinformatics analysis of 959 *B. anthracis* SRA data sets inferred that abundances and genomic arrangements of the 16S-BA-allele and the entire rRNA operon copy numbers differ considerably between strains. Expression ratios of 16S-BA-alleles were proportional to the respective genomic allele copy numbers. The findings and experimental tools presented here provide detailed insights into the intra- and intergenomic diversity of 16S rRNA genes and may pave the way for improved identification of *B. anthracis* and other pathogens with diverse rRNA operons.

IMPORTANCE For severe infectious diseases, precise pathogen detection is crucial for antibiotic therapy and patient survival. Identification of *Bacillus anthracis*, the causative agent of the zoonosis anthrax, can be challenging when querying specific nucleotide sequences such as in small subunit rRNA (16S rRNA) genes, which are commonly used for typing of bacteria. This study analyzed on a broad genomic scale a cryptic and hitherto underappreciated allelic variant of the bacterium's 16S rRNA genes and their transcripts using a set of *in situ*, *in vitro*, and *in silico* assays and found significant intra- and intergenomic heterogeneity in the distribution of the allele and overall rRNA operon copy numbers. This allelic variation was uniquely species specific, which enabled sensitive pathogen detection on both DNA and transcript levels. The methodology used here is likely also applicable to other pathogens that are otherwise difficult to discriminate from their less harmful relatives.

KEYWORDS 16S rRNA, *Bacillus anthracis*, fluorescence *in situ* hybridization, anthrax, digital PCR, genomics, pathogen detection

Anthrax, caused by the spore-forming bacterium *Bacillus anthracis*, is a disease of animals but can also affect humans either through contact with infected animals

Editor Pedro H. Oliveira, Génomique Métabolique, Genoscope, Institut Francois Jacob, CEA, CNRS, Université Paris-Saclay

Copyright © 2022 Braun et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Gregor Grass, gregor_grass@hotmail.com.

The authors declare no conflict of interest.

[†]Deceased 31 March 2019.

Received 17 November 2021

Accepted 20 December 2021

Published 25 January 2022

and their products or as a consequence of deliberate acts of bioterrorism (1, 2). Because of its high pathogenicity, rapid, sensitive, and unambiguous identification of the pathogen is vital. However, diagnostic differentiation of *B. anthracis* from its closest relatives of the *Bacillus cereus sensu lato* group is challenging. Phenotypic properties are not species specific, and nearly identical derivatives of the anthrax virulence plasmids can also be found in related bacilli (2).

In spite of earlier work (3), rRNA gene sequences have not been deemed discriminatory for unambiguous distinction of *B. anthracis* from its closest relatives due to the lack of specific sequence variations. Recent analysis of 16S rRNA gene alleles of *B. anthracis* and relatives, however, revealed an unexpected SNP (single-nucleotide polymorphism) at position 1110 (position 1139 in reference 4; 1110 according to the *B. anthracis* strain Ames ancestor, [NC_007530](#)) in some of the 16S rRNA gene copies (4). This SNP has previously been missed, most likely because it is present only in some of the eleven 16S rRNA gene copies (4). Despite the high abundance of more than 1,000 publicly available short-read genomic data sets and more than 260 genome assemblies, reliable information about sequence variations within *B. anthracis* rRNA operons is still scarce due to the limitations of short-read whole-genome sequencing (WGS) and subsequent reference mapping to detect sequence variations in paralogous, multicopy genes. Producing high-quality genomes, e.g., through hybrid assemblies of long- and short-read approaches, would help bridge this gap.

In this study, we validated a species-discriminatory SNP within the 16S rRNA genes of *B. anthracis* using a set of different *in situ*, *in vitro*, and *in silico* approaches on both genomic and transcript levels. Through this work, we established new diagnostic tools for *B. anthracis*, including a fluorescence *in situ* hybridization (FISH) assay and a digital PCR (dPCR) test for both genomic and transcript identification and quantification. While these new tools do not replace existing diagnostic approaches for identification of *B. anthracis*, they are a valuable addition to the toolbox for its detection and characterization. Finally, we expanded our analysis on all short-read *B. anthracis* data sets available in the NCBI Short Read Archive (SRA) and calculated the rRNA operon copy numbers and allele frequencies using a coverage ratio-based bioinformatics approach.

RESULTS

An SNP in transcripts of 16S rRNA genes enables specific microscopic detection of *B. anthracis* by FISH. Triggered by earlier data on a unique SNP position in some copies of the 16S rRNA gene of *B. anthracis* (guanine-to-adenine transition at position 1110) (4), we aimed at developing a new FISH assay for the identification of *B. anthracis*. Previous work has introduced a probe set for the FISH-based identification of *B. anthracis* (5). Evaluation of the probe sequences revealed, however, that they are unsuitable for unambiguous *B. anthracis* identification due to unspecific probe binding (6). Thus, we designed a FISH probe for discriminating *B. anthracis* from all of its close relatives targeting this specific SNP in 16S rRNA genes (probe BA_SNP_Cy3). Additionally, we developed probe BC_SNP_FAM, which binds to 16S rRNA sequence found in all *B. cereus sensu lato* strains, including *B. anthracis* (see Table S2 in the supplemental material). No other bacterial or archaeal 16S rRNA gene in the SILVA database had a full match for both of the newly designed probes (accessed 1 March 2021). To increase signal intensity and stringency (7), we incorporated two locked nucleic acids (LNA) in probe BA_SNP_Cy3 and one LNA in probe BC_SNP_FAM. Optimum formamide concentrations in the hybridization buffer of this FISH assay were titrated and finally set at 30% (vol/vol) formamide for species differentiation (Fig. S1).

For assay validation, the 16S rRNA probes were tested against a broad panel of *B. cereus sensu lato* strains, including *B. cereus* biovar *anthracis* (Table S1). The FISH assay allowed differentiation of *B. anthracis* from all other *B. cereus sensu lato* group members. *B. anthracis* cells displayed red fluorescence Cy3 signals after hybridization of the specific 16S rRNA variation at position 1110 and green fluorescence 6-carboxyfluorescein (FAM) signals resulting from hybridization to the divergent 16S rRNA featuring no

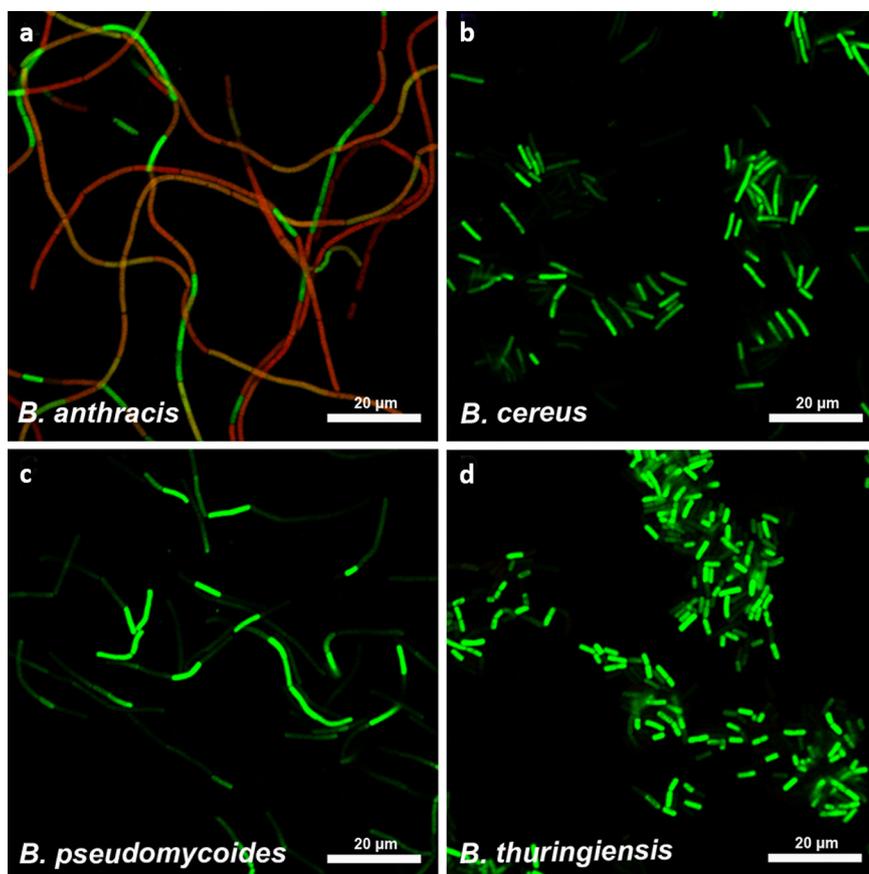


FIG 1 FISH-based microscopic differentiation of *B. anthracis* from other *B. cereus sensu lato* group species. Representative images for *B. anthracis* (a; strain Bangladesh 28/01) *B. cereus* (b; strain ATCC 6464), *B. pseudomycooides* (c; strain WS 3119), and *B. thuringiensis* (d; strain WS 2614) are shown as overlay images of red (probe BA_SNP_Cy3/568 nm) and green (fluorescent channels probe BC_SNP_FAM/520 nm).

B. anthracis-specific SNP (Fig. 1). No red Cy3-signals were detected in any of the non-*B. anthracis* *B. cereus sensu lato* group strains.

While we found Cy3 FISH signals for all *B. anthracis* strains, we discovered broad variations in Cy3 fluorescence signal intensities for different cells of the same and between different *B. anthracis* strains. Even for cells of the same chain, there were individual cells showing almost uniquely either the Cy3 or the FAM signal, resulting in a mosaic-like pattern (Fig. 1). Total fluorescence intensities varied between different *B. anthracis* strains from very strong Cy3 signals to the extreme cases of *B. anthracis* strains ATCC 4229 Pasteur, SA20, and A3783, for which Cy3 signals were very weak (for signal intensities see Table S1). These findings strongly indicate that the 16S rRNA of *B. anthracis* can be used for microscopy-based specific pathogen detection. Notably, variations in fluorescence intensities suggest differences in the rRNA expression level. As these differences might be caused by a gene dose effect, we decided to analyze the genomic distribution of the *B. anthracis*-specific SNP in 16S rRNA genes.

Genomic analysis of *B. anthracis* genomes reveals variations in 16S-BA-allele frequencies. We correlated FISH results with the abundance of 16S rRNA gene copies harboring the *B. anthracis*-specific SNP within different *B. anthracis* genomes. Despite the significant number of *B. anthracis* genomes published, the vast majority of sequences have been generated using short-read sequencing with subsequent mapping to the reference genome (Ames Ancestor; GenBank accession no. [NC_007530](https://www.ncbi.nlm.nih.gov/nuccore/NC_007530) [8]). Due to multiple copies of the rRNA operons, conventional short-read sequencing and mapping approaches do not allow for reliable detection of allele variations. During *de novo* assembly of short reads, nearly identical regions like rRNA operons are collapsed into

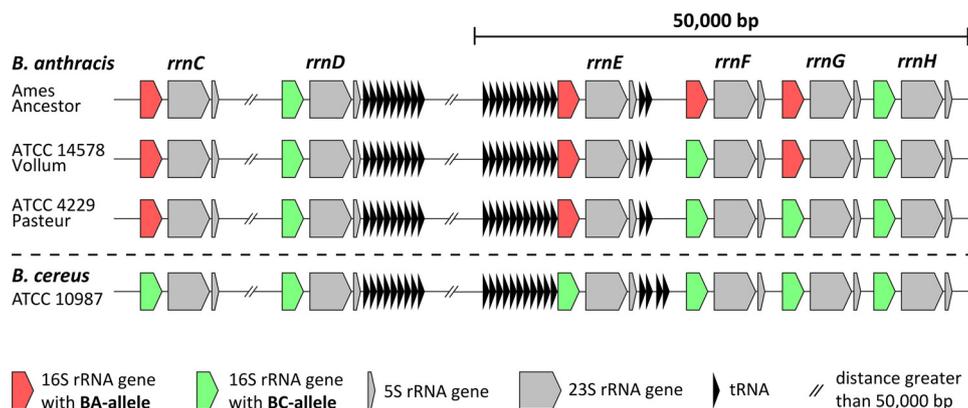


FIG 2 Schematic illustration of the genomic organization of rRNA operons and distribution of 16S alleles in *B. anthracis*. Depicted are the 16S, 23S, and 5S ribosomal subunits and tRNA genes from operons *rrnC* to *-H* in strains Ames Ancestor, ATCC 14578 Vollum, ATCC 4229 Pasteur, and *B. cereus* ATCC 10987. The 16S rRNA genes are either displayed in red for 16S-BA-alleles or in green for 16S-BC-alleles. Not shown are operons *rrnA*, *-B*, *-I*, *-J*, and *-K* exclusively carrying the 16S-BC allele in any strain. Distances are not to scale.

one contig representing only a consensus sequence missing any minor allele variations. Thus, potential differences in allele frequencies can easily be missed. Because of mapping to the reference genome, consensus sequences always feature 16S rRNA allele distribution identical to that of the reference. Hence, there is a need for high-quality genomes generated by hybrid assemblies using long- and short-read sequences for obtaining insights into the real distribution and diversity of 16S rRNA alleles in *B. anthracis* genomes.

To start meeting this need, we analyzed and compared the 16S gene sequences and locations in all available high-quality genomes of *B. anthracis* (accessed at the end of 2020) that are based on long-read sequencing and *de novo* assembly. Figure 2 shows a schematic illustration of the genomic organization of rRNA operons, including 16S, 23S, and 5S ribosomal subunits as well as tRNA genes from operons *rrnC* to *-H* (outlying operons A, B, I, J, and K are not shown) of representative strains for different 16S rRNA genotypes (Ames Ancestor, [NC_007530](#) [8]; ATCC 14578 Vollum [in-house sequenced; this work; Table S3]; ATCC 4229 Pasteur, [NZ_CP009476](#) [9]) and closely related *B. cereus* strain ATCC 10987, [NC_003909](#) (10).

We found that all 16S rRNA gene copies featuring the *B. anthracis*-specific SNP have 100% sequence identity, representing a distinct allele. For simplification, copies featuring this guanine-to-adenine transition at position 1110 were termed 16S-BA-(*B. anthracis*)-alleles, while all other variants lacking this transition were designated 16S-BC-(*B. cereus sensu lato*)-alleles.

The three *B. anthracis* strains, Ames Ancestor, ATCC 14578 Vollum, and ATCC 4229 Pasteur, harbored different 16S-BA/BC-allele frequencies, with 4/7, 3/8, and 2/9 copies, respectively (Fig. 2). No 16S-BA-alleles were found in *B. cereus* ATCC 10987 or any other non-*B. anthracis* strain. In all three *B. anthracis* strains, rRNA operons *rrnA*, *-B*, *-D*, *-H*, *-I*, *-J*, and *-K* carried 16S-BC-alleles, while for *rrnC* and *rrnE* exclusively the 16S-BA-allele was identified. Only two rRNA operons, *rrnF* and *rrnG*, were found to be variable, with strain Ames Ancestor harboring two 16S-BA-alleles and strain ATCC 4229 Pasteur only the BC-alleles for *rrnF* and *rrnG*. Strain ATCC 14578 Vollum exhibited an intermediate state with a 16S-BA-allele in *rrnG* and a BC-allele in *rrnF* (Fig. 2). Thus, it is possible that these differences in 16S rRNA allele distributions caused the observed variations in *B. anthracis*-specific FISH signals (Fig. 1) by gene dose-mediated differences in rRNA transcription levels.

A tetraplex dPCR assay enables the absolute quantification of species-specific 16S rRNA gene allele numbers in *B. anthracis*. To verify this finding and to quantify the ratios of each allele in a diverse panel of *B. anthracis* strains, we designed and tested a hydrolysis probe-based digital PCR (dPCR) assay (Fig. 3a). This assay utilized

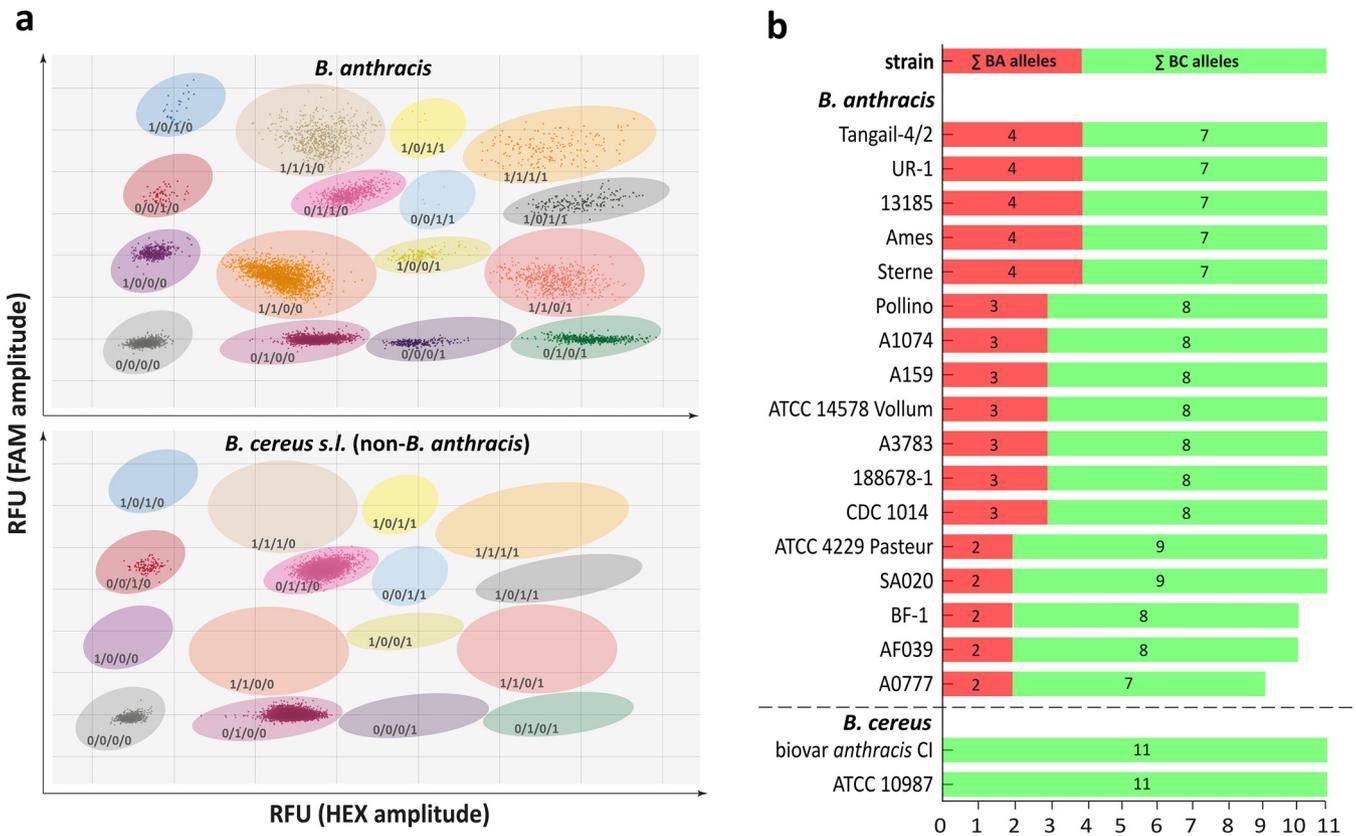


FIG 3 Detection and quantification of 16S rRNA gene alleles in *B. anthracis* and *B. cereus sensu lato* strains. (a) Typical results of a tetraplex dPCR assay using *B. anthracis* template DNA (upper) and DNA of a non-*B. anthracis* member of the *B. cereus sensu lato* group (lower). With each dot representing a droplet plotted according to its FAM signal amplitude (RFU, relative fluorescence units) on the y axis and HEX signal amplitude on the x axis, a total of 16 (for *B. anthracis*, upper) or 4 (non-*B. anthracis* members of the *B. cereus sensu lato* group, lower) clusters (defined by shaded areas) can be assigned to a certain dPCR marker combination of *gyrA* (FAM high signal), *PL3* (HEX high signal), 16S-BA-allele (FAM low signal), and 16S-BC-allele (HEX low signal). Each cluster is labeled with 1 (positive for respective marker) or 0 (negative for respective marker) according to its marker combination (*gyrA*/*PL3*/16S-BA-allele/16S-BC-allele). Since both the *PL3* gene and the 16S-BA-allele are exclusively found in *B. anthracis*, the 16S-BA- and 16S-BC-allele copy numbers can be calculated from the positive droplets of single-copy genes (*PL3* and *gyrA*) and multicopy 16S rRNA genes. All dPCR patterns lacking both the *PL3* gene and the 16S-BA-allele clusters represent DNA of a non-*B. anthracis* member of the *B. cereus sensu lato* group. (b) Copy numbers for 16S-BA- and 16S-BC-alleles for all *B. anthracis* strains and *B. cereus* biovar *anthracis* CI tested.

hexachlorofluorescein (HEX) (green) and FAM (blue) fluorescent dye-labeled allele-specific probes for the 16S-BC-allele and -BA-allele, respectively, with both probes targeting SNP 1110 of the 16S rRNA genes (Table S2). In parallel, a previously published second hydrolysis probe-based PCR assay using HEX dye was adopted for dPCR. This assay targets the *B. anthracis*-specific chromosomal *PL3* gene (11). Finally, a pan-*B. cereus sensu lato* hydrolysis probe-based PCR assay on the *gyrA* (*gyrase* gene) marker using FAM dye was designed, facilitating the detection and quantification of *B. cereus sensu lato* species (including *B. anthracis*) chromosomes. In these dPCR assays, the *PL3* and *gyrA* dPCR tests served as internal controls (for *B. anthracis* and *B. cereus sensu lato*, respectively), each positive for *B. anthracis* genomic DNA versus negative for *PL3* and positive for *gyrA* using genomic DNA of other members of the *B. cereus sensu lato* group.

These four assays were combined into a single tetraplex dPCR assay. To achieve the required signal separation of the four individual dPCR reactions (on our dPCR analysis instrument featuring only two channels, FAM and HEX), we deliberately altered the signal output levels by titrating concentrations of probes labeled with the same dye (Fig. 3a). Thus, the *PL3* marker assay was tuned to produce high HEX signals versus low HEX signals coming from 16S-BC-alleles. Likewise, the *gyrA* marker assay was set to produce high FAM signals versus low FAM signals originating from 16S-BA-alleles. Since both *PL3* and *gyrA* are single-copy genes located on the chromosome of *B. anthracis*,

these markers should result in very similar quantitative outputs when individual *B. anthracis* DNA samples are analyzed. Therefore, these markers served as internal quantification controls in this work.

A typical analysis output of this tetraplex dPCR assay is exemplified in Fig. 3a. In a two-dimensional plot (FAM signal amplitude on the y axis and HEX signal amplitude on the x axis) of such tetraplex dPCR data, one can discriminate a specific fluorescence pattern after dPCR, representing 16 clusters (when *B. anthracis* DNA was used as a template). Each of the droplets within a cluster contained a certain target combination of *gyrA*, *PL3*, 16S-BA-allele, and/or 16S-BC-allele (for example, *gyrA*⁺/*PL3*⁺/16S-BC-allele⁺/16S-BA-allele⁺ or *gyrA*⁻/*PL3*⁻/16S-BC-allele⁻/16S-BA-allele⁻). Using template DNA originating from a non-*B. anthracis* member of the *B. cereus sensu lato* group (i.e., not harboring any 16S-BA allele) resulted in the expected formation of only four droplet clusters, i.e., lacking all signals of *B. anthracis*-specific clusters containing combinations of the *PL3* marker or the 16S-BA-allele (Fig. 3a).

Testing the assay on the reference strains Ames, ATCC 14578 Vollum, and ATCC 4229 Pasteur, we found four, three, and two 16S-BA-alleles, respectively, and eleven 16S rRNA total copies per cell in all three strains. This agreed with the values determined by genomic analysis and, therefore, validated the dPCR assay being able to accurately quantify 16S rRNA alleles in *B. anthracis*.

Using the validated tetraplex dPCR assay, we analyzed the same strain panel as that tested by FISH (Table S1). Similar to FISH, there was no signal for 16S-BA-alleles in the 32 non-*B. anthracis* strains of the *B. cereus sensu lato* group. This panel included *B. cereus* biovar *anthracis*, which can cause anthrax-like disease due to the presence of both virulence plasmids. Its chromosomal background is closer to *B. cereus*, and, as expected, no 16S-BA-allele signal was detected. However, all of the 17 *B. anthracis* strains harbored at least two (up to four) copies of the 16S-BA-allele per cell (Fig. 3b). The majority of *B. anthracis* strains exhibited the genotype 4/7 or 3/8 (16S-BA/BC-alleles; six and seven strains, respectively). These predominant genotypes, together with genotype 2/9 (strain ATCC 4229 Pasteur and strain SA020), were all found to harbor 11 rRNA operons in total, which agrees with previously determined numbers of rRNA operons in these strains. Conversely, strains A182 and BF-1 harbored only ten 16S gene copies in total (genotype 2/8). Notably, strain A0777 exhibited just nine rRNA copies, two of which contained the *B. anthracis*-specific SNP (genotype 2/7).

16S-BA-allele frequencies and total rRNA operon copy numbers vary between different *B. anthracis* strains. To further confirm dPCR results and to exclude underestimation by dPCR as a possible cause of the unexpectedly low number of total rRNA operons in strains A182, BF-1, and A0777, we conducted a combination of long- and short-read sequencing on these and 32 additional *B. anthracis* strains (Table S1). A mean read length of about 15 kb generated by Nanopore sequencing combined with Illumina 2 × 300 bp paired-end sequencing allowed for the precise assembly of complete genomes, including correct positioning of rRNA operons on the chromosome. Coverage values of more than 200-fold enabled the accurate quantification of SNPs; therefore, genotypes based on 16S-BA/BC-allele distribution could be reliably determined. The results matched those obtained from dPCR, confirming the accuracy and reliability of the tetraplex assay. We found that strain A0777 lacked rRNA operons *rrnG* and *rrnH*. rRNA operon *rrnG* was not present in strains AF039, SA020, and BF-1. The genome regions downstream of the missing rRNA operons and upstream of the next rRNA operon were also absent.

To extend our analysis of 16S rRNA allelic states to more *B. anthracis* strains, we expanded our investigation on all publicly available short-read sequence data for *B. anthracis* generated using Illumina sequencing technology. Starting from our newly generated high-quality hybrid assemblies, we developed a *k*-mer- and coverage ratio-based tool to calculate the rRNA operon copy numbers and allele frequencies from all SRA data sets published until the end of 2020. These numbers of rRNA operons and 16S-BA-alleles (from short-read data sets) were identical to the long-read data of the same genomes (Data Set S1). After this method validation, we analyzed 986 SRA

TABLE 1 16S rRNA genotypes obtained from *k*-mer-based SRA analysis^a

16S-BA-allele	No. (%) of strains	No. of rRNA operon copies per genome	16S rRNA genotype (16S-BA-alleles/BC-alleles)	No. (%) of strains
1	3 (0.31)	9	1/8	1 (0.10)
		10	1/9	2 (0.21)
2	221 (23.0)	9	2/7	28 (2.92)
		10	2/8	116 (12.10)
		11	2/9	77 (8.03)
3	560 (58.39)	9	3/6	3 (0.31)
		10	3/7	39 (4.10)
		11	3/8	518 (54.01)
4	164 (17.10)	9	4/5	3 (0.31)
		10	4/6	32 (3.34)
		11	4/7	129 (13.45)
5	11 (1.15)	11	5/6	11 (1.15)

^aNumbers of 16S-BA-alleles, overall rRNA operon numbers, and 16S rRNA genotypes resulting from these values are listed with their respective frequencies.

Illumina sequenced data sets for 16S rRNA operon and BA/BC-allele distribution. After assembly and filtering, 959 genomes remained for a detailed comparison. The majority ($n = 735$, 76.64%) contained 11 rRNA operons, 189 genomes (19.71%) harbored 10 rRNA operons, and only 35 genomes (3.65%) contained 9 rRNA operons (Table 1). This ratio is comparable to that found in our initial strain set tested with FISH and dPCR (11 copies, 82.35%; 10 copies, 11.76%; 9 copies, 5.88%). Of these 959 genomes, the 16S-BA-allele distributions showed that 23.04% had 2, 58.39% had 3 and 17.10% had 4 copies (Table 1). As with the rRNA operon copy numbers, this distribution correlated with the 16S-BA-allele distribution in our strain set analyzed by dPCR and WGS (2, 29.41%; 3, 41.17%; 4, 29.41%). Notably, a few strains were calculated to possess 1 (0.31%) or 5 (1.15%) 16S-BA-alleles. The overall diversity of 16S rRNA genotypes (BA alleles/BC alleles) was higher than that in our initial strain set (genotypes 4/7, 3/8, and 2/9). Additional major genotypes (frequency, >5) obtained from SRA analysis comprised 16S-BA/BC-allele ratios of 2/8 and 2/7 and minor genotypes were 5/6, 4/6, 4/5, 3/7, 3/6, 1/9, and 1/8, each with frequencies of <5 .

Interestingly, 10 of the genomes that were calculated to possess five BA alleles are from the same originating lab and were sequenced with a 100 bp single-end technique only (Data Set S1). Thus, without genomic context it is hardly possible to validate the presence of a fifth 16S-BA-allele from single-end short reads. The same applies to the only other strain (BC038/2000031523) sequenced with 2×100 bp paired-end reads and a mean insert size of 520 bp. Along with three strains putatively containing a single 16S-BA-allele only, strains with five BA-alleles should be resequenced using long-read technology for validation.

Finally, we tested to which degree 16S rRNA genotypes fit the phylogenetic placement of strains. For this, we correlated established phylogeny of *B. anthracis* based on a number of canonical SNPs (12) with the distribution of 16S-BA-alleles within 10 major canonical SNP groups of the three branches, A, B, and C, of *B. anthracis*. Figure S2 shows that there is limited correlation. Notably, B-branch featured a small set of genotypes besides the major 2/8 type. The few C-branch strains all had the 3/7 genotype. A-branch (comprising the majority of isolates) was the most diverse, dominantly showing the 2/9 genotype (with the exception of canSNP group Ames, 4/7). Although the 16S rRNA genotypes did not follow the established phylogeny of *B. anthracis*, the newly developed tools (tetraplex dPCR and *k*-mer-based SRA analysis) might still be harnessed as an alternative typing system for *B. anthracis* strains.

Expression of 16S-BA-alleles is proportional to gene copy number. Various ratios of 16S-BA/BC-alleles constitute possible explanations for differences in FISH signals of cells of diverse *B. anthracis* strains (Fig. 1). Indeed, we found a significant correlation between 16S-BA/BC-allele ratios in sequenced genomes and mean intensities of the

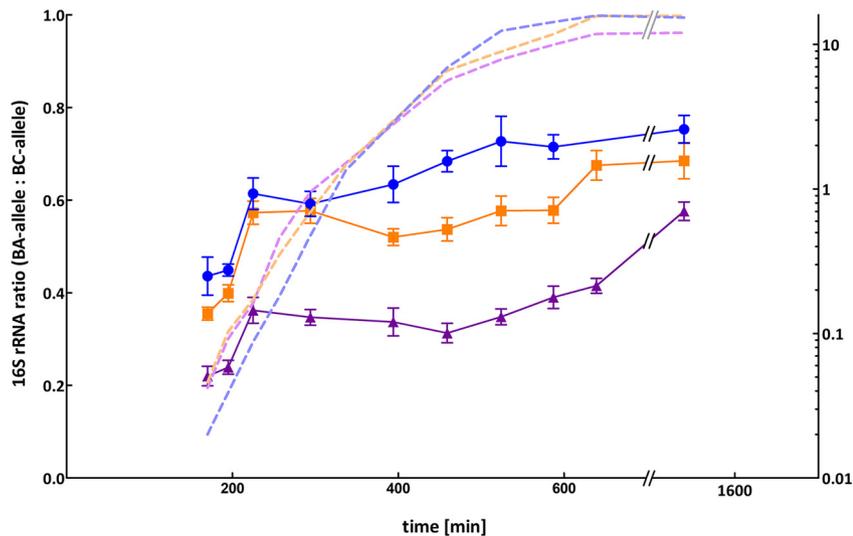


FIG 4 Expression ratios of 16S-BA- and -BC-alleles in three different *B. anthracis* strains at different growth phases. Expression level ratios of 16S-BA-alleles relative to 16S-BC-alleles were calculated from absolute target concentrations obtained by RT-dPCR. Values were plotted against time points of each sample taken during growth from early exponential to stationary phase for *B. anthracis* Sterne (blue), CDC1014 (orange), and Pasteur ATCC 4229 (purple) representing three major 16S rRNA genotypes (BA/BC), 4/7, 3/8, and 2/9, respectively. Error bars indicate the Poisson 95% confidence intervals for each copy number ratio. Dotted lines depict cell densities over time.

Cy3 FISH signals targeting the 16S-BA-allele (tested with the `cor.test` function in R, Pearson's $r = 0.61$, $P = 0.009$), confirming this assumption.

To investigate whether the 16S-BA-alleles are differentially expressed throughout different growth phases of *B. anthracis*, we quantified 16S rRNA from growth experiments (Fig. 4). For this, culture samples of *B. anthracis* strains Sterne, CDC1014, and Pasteur ATCC 4229, representing three major 16S-BA/BC-allele genotypes, 4/7, 3/8, and 2/9, respectively, were taken for total RNA extraction at several time points during lag, log, and stationary growth phase. To compare rRNA levels with FISH signals, we also took parallel samples from six of these time points for FISH analysis. By a one-step reverse transcription duplex dPCR, the two 16S allele targets were interrogated for the expression ratios of the 16S-BA- to 16S-BC-alleles. *B. anthracis* RNA yielded four clusters of drop-lets in two-dimensional analysis plots, namely, 16S-BC-allele⁻/16S-BA-allele⁻, 16S-BC-allele⁺/16S-BA-allele⁻, 16S-BC-allele⁻/16S-BA-allele⁺, and 16S-BC-allele⁺/16S-BA-allele⁺ (compare Fig. 3). RNA of other *B. cereus sensu lato* strains produced only two cluster types lacking 16S-BC-allele⁻/16S-BA-allele⁺ and 16S-BC-allele⁺/16S-BA-allele⁺. Absolute quantification of the two initial target concentrations of 16S-BA-alleles/BC-alleles in samples from growth cultures made it possible to determine their ratios representing the expression levels of the 16S-BA-alleles relative to those of 16S-BC-alleles (Fig. 4). Notably, 16S-BA/BC-allele rRNA ratios varied during growth and showed similar expression patterns in all three tested *B. anthracis* strains. Starting from a relatively low 16S-BA/BC-allele ratio in early log phase, the fraction of 16S-BA-allele expression increased in early log phase and decreased in mid-log phase with a final increase toward the stationary phase. While shifts in 16S-BA/BC-allele expression patterns in these strains were similar, differences were observed in numerical expression ratios. *B. anthracis* Sterne showed the highest 16S-BA/BC-allele expression ratio, ranging from 0.44 (early exponential phase) up to 0.75 (stationary phase), compared to CDC1014 with 0.36 to 0.69 and Pasteur ATCC 4229 with 0.22 to 0.58, which was found to have the lowest 16S-BA-allele expression in all growth phases. The largest differences in expression levels between all strains were observed in late log phase (Fig. 4). The observed diverging levels of 16S-BA-allele expression in the three tested strains can easily be explained by the different numbers of 16S-BA-allele copies per genome (2, 3, or 4). Nevertheless, the proportion of 16S-BA-allele rRNA in late

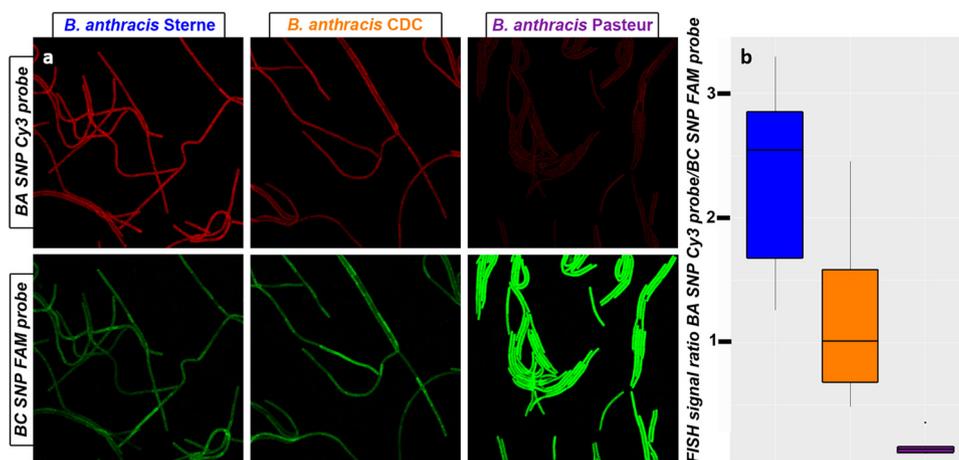


FIG 5 FISH of *B. anthracis* strains harboring different numbers of 16S-BA-alleles. (a) Representative FISH images showing signal intensities of *B. anthracis* strains with diverging genomic 16S (BA/BC) allele profiles Sterne (4/7), CDC1014 (3/8), and Pasteur ATCC 4229 (2/9). Samples were taken and processed after 460 min of continuous growth. (b) Boxplot of BA_SNP_Cy3 and BC_SNP_FAM FISH signal ratios across all sampled time points for *B. anthracis* Sterne (blue), CDC1014 (orange), and Pasteur ATCC 4229 (purple).

exponential *B. anthracis* cells is quite disproportionate. If all rRNA operons were transcribed at a constant and equal rate, one would expect a ratio of 0.22 (Pasteur 2/9), 0.38 (CDC/8), and 0.57 (Sterne 4/7). Instead, we measured ratios that correlate to a 1.57 (Pasteur)-, 1.46 (CDC)-, and 1.09 (Sterne)-fold 16S-BA-allele overrepresentation on average throughout all growth phases and up to 2.59 (Pasteur)-, 1.83 (CDC)-, and 1.32 (Sterne)-fold in stationary phase.

The shift toward elevated expression of the 16S-BA-allele genes over time was not significantly reflected in FISH signal intensities, possibly due to the general decrease of FISH signals over time. However, if cells were sampled and fixed at identical time points, 16S-BA/BC-allele ratios were always highest for *B. anthracis* Sterne and lowest for *B. anthracis* Pasteur, which reflects their 16S-BA/BC-allele ratios on the genomic and transcript levels (Fig. 5). Also, sampled across all time points, 16S-BA/BC-allele FISH signal ratios correlated well with allele distributions in the three different strains (analysis of variance in R, $P = 0.0002$) (Fig. 5).

DISCUSSION

Using a combination of newly developed *in situ*, *in vitro*, and *in silico* approaches, we unraveled the elusive heterogeneity of 16S rRNA genes in the biothreat agent *B. anthracis*. Results consistently delineate the organism's intragenomic diversity of 16S rRNA genes, their differential expression across growth phases, and their intergenomic heterogeneity in publicly available and newly sequenced genomes. Intragenomic microdiversity within 16S rRNA genes has long been known from other species (13, 14) and was found to increase with higher copy numbers of rRNA operons (15). Thus, the species-wide intra- and intergenomic microdiversity related to SNP 1110 in 9 to 11 copies of the 16S rRNA gene of *B. anthracis* is not totally unsurprising (3, 4). Whereas some such polymorphic sites are associated with a distinct phenotypic trait (e.g., stress resistance) (16, 17), the functional assignment for the majority of these sequence variations (including those in *B. anthracis* 16S rRNA genes) remains elusive.

Although discovered before using Sanger sequencing (4), the specific SNP in the 16S rRNA genes of *B. anthracis* was disregarded despite the availability of numerous published genomes. Generally, sequence variations in multicopy genes such as 16S rRNA genes can hardly be detected when relying on conventional short-read WGS and subsequent reference mapping (18), which was used to generate the majority of publicly available *B. anthracis* whole-genome sequences. SNP calling in different rRNA operons or other paralogous genes gives ambiguous results, since assemblers tend to

interpret low-frequency sequence variations as sequencing errors and correct them prior to assembly (19). Even if detected, distances of the SNP to unique flanking regions up- and downstream of the multicopy gene may be >1,000 bases and, thus, are larger than typical library fragment sizes of 500 to 800 bases. In such cases, chromosomal locations of SNPs cannot be reconstructed. Instead, all rRNA gene-related reads are assembled into one contig with diverse fringes (20). The average read length of Nanopore sequencing is typically larger than 5 kb and therefore can cover complete rRNA operons. Thus, any unique SNP occurring in a single or a few rRNA gene alleles can be precisely allocated to a specific chromosome position, especially when combined with short-read sequencing and hybrid assembly as used here. Therefore, the challenges described above will become rather minor for future genomic analysis of *B. anthracis*. Such work is facilitated by the additional 33 complete high-quality genomes we have contributed here. These genomes cover all three major phylogenetic lineages (canSNP groups), all bona fide 16S-BA-allele frequencies (2, 3, and 4), and all known rRNA operon copy numbers.

On the *B. anthracis* chromosome, the 16S rRNA operons *rrnE*, *-F*, *-G*, and *-H* are located in close proximity to each other, with only 15.8, 8.5, and 5.2 kb, respectively, between them (forming a genomic region with a high density of four 16S rRNA operons within less than 50 kb). Conversely, the other 16S rRNA genes are rather dispersed, with distances greater than 50 kb between them. The 16S-BA-allele is present in operons *rrnC* and *rrnE* in all strains analyzed with long-read WGS, while *rrnF* and *rrnG* seem to be variable. Since the four operons *rrnE*, *-F*, *-G*, and *-H* are relatively close to each other in the *B. anthracis* chromosome, homologous recombination and gene duplications might be the reason for this allelic variation. Compared to all other 16S rRNA alleles on the *B. anthracis* genome, the 16S rRNA copies in this region (*rrnE* to *rrnH*) seem to differ from each other only in SNP position 1110. This finding promotes the explanation that the 16S rRNA copies in this region of high rRNA operon density are subject to an increased recombination rate between alleles with and without *B. anthracis*-specific SNP 1110. This notion is also supported by the fact that only operons *rrnG* and *-H* seem to be affected by deletion events in all strains analyzed by long-read WGS. The alternative explanation, horizontal gene transfer of a divergent allele, seems unlikely. We were unable to identify any 16S rRNA gene in public databases matching the 16S-BA-allele outside *B. anthracis*.

Recombination and deletion events in 16S rRNA operons of *B. anthracis* do occur. These events were experimentally shown in a study on bacitracin resistance. Two deletion events, DelFG and DelGH, were described that caused elimination of gene clusters between rRNA operons *rrnF*, *-G*, and *-H* (18). These DelFG and DelGH events describe a possible origin of *B. anthracis* strains with 10 16S rRNA gene copies, i.e., 21% of all strains (Table 1). Random gene duplication and gene elimination by recombination also might explain another observation: the newly defined 16S rRNA genotypes did not convincingly reflect the established *B. anthracis* phylogeny (Fig. S2). Instead, some 16S rRNA genotypes seem to be dominant yet not exclusive in separate branches, e.g., 2/8 copies in B-branch or 3/8 in A-branch (Fig. S1).

The recognition of intra- and intergenomic 16S rRNA allele diversity in *B. anthracis* opens possibilities to harness unique SNPs in 16S rRNA gene alleles and their transcripts. This finding strongly highlights the great potential of such genomic variations for both identification of *B. anthracis* and for diagnostics of anthrax disease. This approach is probably also applicable to other pathogens that are otherwise difficult to discriminate from their less notorious relatives.

MATERIALS AND METHODS

Cultivation of bacteria. The cultivation of the virulent *B. anthracis* strains was performed in a biosafety level 3 laboratory (BSL3) (see Table S1 in the supplemental material). All *Bacillus* strains were cultivated overnight on Columbia blood agar plates (containing 5% sheep blood; Becton, Dickinson, Heidelberg, Germany) at 37°C.

For isolation of DNA, a 1- μ l loop of colonies was transferred to a 2-ml screwcap microcentrifuge tube, inactivated with 2% terralin PAA (Schülke & Mayr GmbH, Norderstedt, Germany) for 30 min, and washed three times with phosphate-buffered saline (PBS) as described previously (21).

For FISH, 50-ml centrifuge tubes containing 5 ml of tryptic soy broth (TSB; Merck KGaA, Darmstadt, Germany) were inoculated with one colony from an overnight culture (described above) and incubated at 37°C with shaking at 150 rpm. After 4 h of growth, bacteria were pelleted by centrifugation at 5,000 × *g* for 10 min, washed with PBS, and fixed with 3 ml 4% (vol/vol) formaldehyde for 1 h at ambient temperature. After fixation, cells were washed three times with PBS, resuspended in a 1:1 mixture of absolute ethanol and PBS, and stored at –20°C until further use. To ensure sterility, 1/10 of the inactivated material was incubated in thioglycolate medium (Merck KGaA, Darmstadt, Germany) for 7 days without growth before material was taken out of the BSL3 laboratory.

For growth-phase analysis, 1 ml of overnight cultures of attenuated *B. anthracis* (Sterne, CDC1014, and ATCC 4229 Pasteur) in TSB was used to inoculate 100 ml of fresh TSB in 1-liter baffled flasks and incubated at 37°C with shaking at 100 rpm. Every 30 min, turbidity was measured as the optical density at 600 nm (OD₆₀₀), and 1-ml samples were taken for FISH and RNA isolation, respectively. After pelleting by centrifugation, samples for RNA isolation were resuspended and inactivated using 2% terralin PAA for 30 min and washed three times with PBS. FISH samples were treated as described above.

Design of primers and probes. Primers and probes were designed using Geneious 10.1.3 (Biomatters, Auckland, New Zealand), and numerous probe variations were tested to identify the best combination and number of locked nucleic acids for differentiation of *B. anthracis* and the other *B. cereus sensu lato* group species based on the SNP (position 1110) detected previously (4). The final probes for FISH included 2 and 1 locked nucleic acid, while dPCR probes contained 5 and 6 for the *B. anthracis* (BA) and the *B. cereus sensu lato* (BC) probe, respectively (Table S2). For sequences of positive (EUB338 [22]) and negative (nonEUB [23]) control probes for FISH, see Table S2. Primers as well as probes labeled with 6-carboxyfluorescein (FAM), hexachlorofluorescein (HEX), indocarbocyanine (Cy3), and indodicarbocyanine (Cy5) were purchased commercially (TIB Molbiol, Berlin, Germany).

To determine the ideal formamide concentration for the FISH hybridization buffer, the fluorescence signals of probe BA_SNP_Cy3 and probe BC_SNP_FAM were assessed with *B. anthracis* Sterne and *B. cereus* ATCC 10987 at different formamide concentrations (0, 10, 20, 25, 30, 35, 40, 45, and 50% FA concentration in the hybridization buffer) as described elsewhere (24). Hybridization at 30% formamide was determined to be ideal for differentiation of *B. anthracis* and *B. cereus sensu lato* group (Fig. S1).

FISH and image processing. FISH was carried out as described elsewhere (24). A positive-control probe targeting eubacteria (EUB338 [22]) and a nonsense probe targeting no known bacterial species (nonEUB [23]) as a control for unspecific probe binding were included in each hybridization experiment. Briefly, 2 μl of fixed cells was spotted on Teflon-coated slides (Marienfeld, Lauda-Königshofen, Germany) and dried at 46°C. The cells then were permeabilized using 10 ml of 15 mg/ml lysozyme (no. 62970; Merck KGaA, Darmstadt, Germany) per well at 46°C for 12 min. After dehydration in an ascending ethanol series (50, 80, and 96% [vol/vol] ethanol), cells were covered with 10 μl hybridization buffer (0.9 M NaCl, 20 mM Tris-HCl [pH 8.0], 0.01% SDS, 30% formamide) with probes at a concentration of 10 μM and incubated in a humid chamber in the dark at 46°C for 1.5 h. Slides were washed in 50 ml prewarmed washing buffer (0.1 M NaCl, 20 mM Tris-HCl [pH 8.0], 5 mM EDTA [pH 8.0]) for 10 min at 48°C in a water bath. Finally, slides were dipped in ice-cold double-distilled water and carefully dried with compressed air. For each strain, FISH was performed in duplicate and two pictures were taken per well, so that the resulting fluorescence intensity was the mean of four images. To increase accuracy in the growth curve assay, five pictures were taken per well, so that the resulting fluorescence intensity was the mean of 10 images. All images were recorded with a confocal laser-scanning microscope (LSM 710; Zeiss, Jena, Germany). Excitations for FAM, Cy3, and Cy5 were at 490, 560, and 630 nm, respectively. Emission was measured within the following ranges: FAM, 493 to 552 nm; Cy3, 561 to 630 nm; and Cy5, 638 to 724 nm. Images were processed with Daime (25), using the area of the EUB signal as a mask to measure average fluorescence intensity for BA_SNP_Cy3 and BC_SNP_FAM. The EUB images were segmented and unspecific fluorescence excluded with default threshold settings, and this object layer was transferred to BA_SNP_Cy3 and BC_SNP_FAM images.

Isolation of nucleic acids. DNA isolation from inactivated cells was carried out using a MasterPure Gram-positive DNA purification kit (Lucigen, Middleton, WI, USA) according to the manufacturer's protocol. DNA samples were quantified using the Qubit dsDNA HS assay kit protocol (Thermo Scientific, Dreieich, Germany). For RNA isolation from inactivated cells, an RNeasy Protect bacterial minikit (Qiagen, Hilden, Germany) was used according to the supplier's protocol for enzymatic lysis and proteinase K digestion of bacteria. To eliminate residual DNA, RNA samples were purified twice using an RNeasy MinElute cleanup kit (Qiagen, Hilden, Germany) and quantified using the Qubit RNA HS assay kit protocol (Thermo Scientific, Dreieich, Germany). The absence of DNA in the final RNA preparation was verified by conducting PCR on marker *dhp61* (26) with negative results.

Tetraplex droplet dPCR assay for quantification of 16S rRNA gene alleles. Digital PCR (dPCR) allows for absolute quantification of DNA or RNA template concentrations (27). For 16S rRNA gene analysis, the 20-μl dPCR predroplet mix consisted of 10 μl dPCR supermix for probes (Bio-Rad Laboratories, Munich, Germany), 1 μl 20× 16S SNP primer mix (final concentrations, 900 nM), 0.6 μl of 20× mix of 16S SNP BC probe (final concentration, 150 nM), 0.6 μl of 20× mix of 16SSNP BA probe (final concentration, 150 nM), 0.9 μl of 20× PL3 primer-probe mix (final concentrations, probe, 225 nM; primers, 810 nM), 1.5 μl of GyrA 20× primer-probe mix (final concentrations, probe, 375 nM; primers, 1,350 nM), 4.4 μl of nuclease-free water (Qiagen, Hilden, Germany), and 1 μl of template DNA freshly diluted to a concentration of 0.05 ng/μl. To ensure independent segregation of the 16S rRNA gene copies from the bacterial chromosome and the reference genes into droplets, template DNA was digested (no cut sites within 16S rRNA genes) prior to dPCR by BsiWI-HF, BsrGI-HF, and HindIII-HF (New England Biolabs GmbH, Frankfurt am Main, Germany) in 1× Cutsmart buffer (New England Biolabs GmbH, Frankfurt am Main, Germany)

for 60 min, and then the enzymes were heat inactivated at 80°C for 20 min according to the manufacturer's protocol.

Partitioning of the reaction mixture into up to 20,000 individual droplets was achieved using a QX200 dPCR droplet generator (Bio-Rad Laboratories, Munich, Germany). A two-step PCR was performed on a Mastercycler Pro instrument (Eppendorf, Wesseling-Berzdorf, Germany) with the following settings: one DNA polymerase activation step at 95°C for 10 min was followed by 40 cycles of denaturation at 94°C for 30 s and annealing/extension at 58°C for 1 min. Final enzyme inactivation was performed at 98°C for 10 min before the samples were cooled down and held at 4°C. All steps were carried out with a temperature ramp rate of 2°C/s. After completion, droplets were analyzed using the QX100 droplet reader (Bio-Rad), and absolute concentrations for each target were quantified using Poisson statistics as implemented in the QuantaSoft Pro Software (Bio-Rad Laboratories, Munich, Germany).

The absolute concentrations of *PL3* and *gyrA* were compared. To ensure assay integrity, samples with a deviation range greater than 10% within the two markers were excluded and had to be repeated. If deviation was below 10%, both targets were set as a reference with a copy number of one. The software then automatically takes the mean concentration of both references to calculate the copy numbers of BC and BA alleles. According to the recommendations provided previously (28), all samples with copy numbers between 0.35 and 0.65 deviations from an integer number or with a confidence interval greater than 1 were excluded from analysis and were repeated. All valid runs were rounded to the next integer number.

Duplex one-step reverse transcription dPCR to compare expression levels of 16S BC- and 16S-BA-allele. The 20 μ l reverse transcription-dPCR reaction mixture consisted of 5 μ l one-step RT-dPCR advanced supermix for probes (Bio-Rad, Laboratories, Munich, Germany), 2 μ l of reverse transcriptase (final concentration, 20 U/ μ l; Bio-Rad), 0.6 μ l of dithiothreitol (DTT) (final concentration, 10 nM; Bio-Rad Laboratories, Munich, Germany), 1.5 μ l 20 \times 16S SNP primer mix (final concentration, 1,350 nM), 1.5 μ l of 20 \times mix of 16S SNP BC probe (final concentration, 375 nM), 1.5 μ l of 20 \times mix of 16S SNP BA probe (final concentration, 375 nM), 6.9 μ l of nuclease-free water (Qiagen, Hilden, Germany), and 1 μ l of template RNA. Reverse transcription was achieved within droplets prior to dPCR. Partitioning of the reaction mixture into up to 20,000 droplets was carried out using a QX200 dPCR droplet generator (Bio-Rad Laboratories), and PCR was performed on a Mastercycler Pro (Eppendorf, Wesseling-Berzdorf, Germany) with the following settings. The initial reverse transcription step was performed at 48°C for 60 min. An enzyme activation step at 95°C was carried out for 10 min, followed by 40 cycles of a two-step program of denaturation at 94°C for 30 s and annealing/extension at 58°C for 1 min. Final enzyme inactivation was performed at 98°C for 10 min before the samples were cooled down and held at 4°C. All steps were carried out with a temperature ramp rate of 2°C/s. After completion, droplets were analyzed using the QX100 droplet reader (Bio-Rad Laboratories, Munich, Germany), and results were quantified with the QuantaSoft Pro Software (Bio-Rad Laboratories).

Library preparation, sequencing, and assembly of genomes. The libraries for the Illumina sequencing were prepared using the NEBNext Ultra II FS DNA library prep kit for Illumina (New England Biolabs GmbH, Frankfurt am Main, Germany) according to the protocol for large fragment sizes of >550 bp but with a minimal fragmentation time of only 30 s. Afterwards, libraries were pooled equimolarly and sequenced on an Illumina MiSeq device (Illumina Inc., San Diego, CA) using the MiSeq reagent kit v3 (2 \times 300 bp).

The libraries for the nanopore sequencing were prepared using the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies, Oxford, UK) combined with the Native Barcoding Expansion EXP-NBD104 and sequenced as one pool on a MinION flowcell FLO-MIN106D (type R9.4.1; Oxford Nanopore Technologies, Oxford, UK) for 48 h. Basecalling and demultiplexing were done separately using Guppy v3.2.10 (Oxford Nanopore Technologies, Oxford, UK) with the high-accuracy basecalling model. Quality (≥ 10) and length ($\geq 1,000$ bp) filtering was done using FilTlong version 0.2.0 (<https://github.com/rwrick/FilTlong>).

Hybrid assemblies were constructed in two stages. First, nanopore reads were assembled using Flye version 2.7 (29) with default parameters and two iterations of polishing. Second, Illumina reads were assembled together with the nanopore raw reads and the nanopore assembly as trusted contigs using SPAdes version 3.14 (30) with parameters “-k 55,77,99,113,127 -careful.” Afterwards, the assembled contigs were reverse complemented, if necessary, and rotated to the same start sequence as strain Ames Ancestor. Finally, the contigs were polished once more using Pilon version 1.23 (31).

Bioinformatics analyses. For the long-read assemblies, ribosomal operons were annotated using barrnap version 0.9 (<https://github.com/tseemann/barrnap>). SNP alleles were searched using USEARCH version 11 (32) and the 16S SNP BA/BC probe sequences (Table S2) as an oligonucleotide sequence database. To investigate the frequency and distribution of the alleles of 16S rRNA genes in the *B. anthracis* species comprehensively, we downloaded all available short-read Illumina data sets (at the end of 2020) from the NCBI Sequence Read Archive (33). These data sets were then assembled using SPAdes v1.14 (30) with parameters “-k 55,77,99,113,127 -careful.” The contigs of the resulting assemblies were extended using tadpole from the BBTools package (34) and with parameters “el=1000 er=1000 mode=extend.” Afterwards, blastn (35) with parameters “-evaluate 1e-10 -word_size 9” was used to align the 23S rRNA sequence against each extended contig end. For each assembly, the number of contigs ending with a 23S rRNA fragment were counted, and CanSNPer (36) was used to determine the canonical SNPs and likely position in the CanSNP tree. In a next step, kmercountexact from the BBTools package was used with the parameters “fastadump=f mincount=2 k=16” to count all *k*-mers of size 16 from error-corrected reads. From these *k*-mers, the frequencies of the two allelic *k*-mers (sequences of 16 nucleotides used for the dPCR probes) were extracted. kmercountexact also reports a *k*-mer-based coverage

estimation of the sequenced reads, which is used to filter the assemblies by coverage (minimum of 20×), number of contigs (maximum of 200), number of potential rRNAs (>8), and success of CanSNPer prediction. For each remaining assembly, the number of rRNAs carrying the SNP of the 16S BA allele was estimated by determining the ratio of the allelic *k*-mers multiplied by the total number of rRNAs, rounded to a whole number. To validate this estimation, we applied the same algorithm to every assembly where both short and long reads and/or dPCR results were available and compared the estimated number of BA alleles to the counted number in the long-read assembly or to the measured number from the dPCR experiments. They were consistent across different sequencing coverages, total number of rRNA operons, and known BA allele frequencies.

Data availability. All genomic data generated or analyzed prior to or during this study can be accessed via the NCBI BioProject number [PRJNA695105](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA695105). Individual accession numbers are listed in Tables S3 and Data Set S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.2 MB.

FIG S1, TIF file, 0.2 MB.

FIG S2, TIF file, 0.1 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank Linda Dobrzykowski and Josua Zinner for technical assistance. For providing bacterial strains, we are grateful to Fabian Leendertz, Silke Klee, and Roland Grunow (RKI), Wolfgang Beyer (University of Hohenheim), and Paul Keim (Northern Arizona University). We also thank Olfert Landt and his team at TIB Molbiol (Berlin) for technical support in designing LNA-based probes.

This study was supported by funds from the German Federal Ministry of Defense (Sonderforschungsprojekt 36Z1-5-431618 and STAN 48-2009-23).

The research described here is part of the Medical Biological Defense Research Program of the Bundeswehr Joint Medical Service. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by any governmental agency, department, or other institutions.

P.B., F.Z., G.G., and K.S. designed the study and interpreted the results. M.W. contributed the bioinformatics analysis. I.S., F.Z., K.A., and S.M. performed FISH experiments. P.B. and I.S. performed digital PCR experiments. G.G., P.B., F.Z., and K.S. wrote the first draft manuscript, and all authors edited the manuscript. The authors dedicate this work to our dear late colleague Karin Aistleitner, who left us too soon and unexpectedly.

We declare no competing interests.

REFERENCES

1. Takahashi H, Keim P, Kaufmann AF, Keys C, Smith KL, Taniguchi K, Inouye S, Kurata T. 2004. *Bacillus anthracis* incident, Kameido, Tokyo, 1993. *Emerg Infect Dis* 10:117–120. <https://doi.org/10.3201/eid1001.030238>.
2. Turnbull PC (ed). 2008. Anthrax in humans and animals, 4th edition. WHO Press, Geneva, Switzerland.
3. Candelon B, Guilloux K, Ehrlich SD, Sorokin A. 2004. Two distinct types of rRNA operons in the *Bacillus cereus* group. *Microbiology (Reading)* 150: 601–611. <https://doi.org/10.1099/mic.0.26870-0>.
4. Hakovirta JR, Prezioso S, Hodge D, Pillai SP, Weigel LM. 2016. Identification and analysis of informative single nucleotide polymorphisms in 16S rRNA gene sequences of the *Bacillus cereus* group. *J Clin Microbiol* 54: 2749–2756. <https://doi.org/10.1128/JCM.01267-16>.
5. Weerasekara MLMAW, Ryuda N, Miyamoto H, Okumura T, Ueno D, Inoue K, Someya T. 2013. Double-color fluorescence in situ hybridization (FISH) for the detection of *Bacillus anthracis* spores in environmental samples with a novel permeabilization protocol. *J Microbiol Methods* 93:177–184. <https://doi.org/10.1016/j.mimet.2013.03.007>.
6. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
7. Koshkin AA, Singh SK, Nielsen P, Rajwanshi VK, Kumar R, Meldgaard M, Olsen CE, Wengel J. 1998. LNA (locked nucleic acids): synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition. *Tetrahedron* 54:3607–3630. [https://doi.org/10.1016/S0040-4020\(98\)00094-5](https://doi.org/10.1016/S0040-4020(98)00094-5).
8. Ravel J, Jiang L, Stanley ST, Wilson MR, Decker RS, Read TD, Worsham P, Keim PS, Salzberg SL, Fraser-Liggett CM, Rasko DA. 2009. The complete genome sequence of *Bacillus anthracis* Ames "Ancestor." *J Bacteriol* 191: 445–446. <https://doi.org/10.1128/JB.01347-08>.
9. Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT, Broomall SM, Bishop-Lilly KA, Bruce DC, Gibbons HS, Coyne SR, Lo C-C, Meincke L, Munk AC, Koroleva GI, Rosenzweig CN, Palacios GF, Redden CL, Minogue TD, Chain PS. 2015. Complete genome sequences for 35 biothreat assay-relevant bacillus species. *Genome Announc* 3:e01501-15. <https://doi.org/10.1128/genomeA.00151-15>.
10. Rasko DA, Altherr MR, Han CS, Ravel J. 2005. Genomics of the *Bacillus cereus* group of organisms. *FEMS Microbiol Rev* 29:303–329. <https://doi.org/10.1016/j.femsre.2004.12.005>.

11. Ellerbrok H, Nattermann H, Ozel M, Beutin L, Appel B, Pauli G. 2002. Rapid and sensitive identification of pathogenic and apathogenic *Bacillus anthracis* by real-time PCR. *FEMS Microbiol Lett* 214:51–59. <https://doi.org/10.1111/j.1574-6968.2002.tb11324.x>.
12. Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, Ravel J, Zanecki SR, Pearson T, Simonson TS, U'Ren JM, Kachur SM, Leadem-Dougherty RR, Rhoton SD, Zinser G, Farlow J, Coker PR, Smith KL, Wang B, Kenefic LJ, Fraser-Liggett CM, Wagner DM, Keim P. 2007. Global genetic population structure of *Bacillus anthracis*. *PLoS One* 2:e461. <https://doi.org/10.1371/journal.pone.0000461>.
13. Cilia V, Lafay B, Christen R. 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. *Mol Biol Evol* 13:451–461. <https://doi.org/10.1093/oxfordjournals.molbev.a025606>.
14. Acinas SG, Marcelino LA, Klepac-Ceraj V, Polz MF. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrm* operons. *J Bacteriol* 186:2629–2635. <https://doi.org/10.1128/JB.186.9.2629-2635.2004>.
15. Vetrovský T, Baldrian P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One* 8:e57923. <https://doi.org/10.1371/journal.pone.0057923>.
16. Werner G, Bartel M, Wellinghausen N, Essig A, Klare I, Witte W, Poppert S. 2007. Detection of mutations conferring resistance to linezolid in *Enterococcus* spp. by fluorescence in situ hybridization. *J Clin Microbiol* 45:3421–3423. <https://doi.org/10.1128/JCM.00179-07>.
17. Kurylo CM, Parks MM, Juetta MF, Zinshteyn B, Altman RB, Thibado JK, Vincent CT, Blanchard SC. 2018. Endogenous rRNA sequence variation can regulate stress response gene expression and phenotype. *Cell Rep* 25:236–248. <https://doi.org/10.1016/j.celrep.2018.08.093>.
18. Furuta Y, Harima H, Ito E, Maruyama F, Ohnishi N, Osaki K, Ogawa H, Squarre D, Hang'ombe BM, Higashi H. 2018. Loss of bacitracin resistance due to a large genomic deletion among *Bacillus anthracis* strains. *mSystems* 3:e00182-18. <https://doi.org/10.1128/mSystems.00182-18>.
19. Heydari M, Miclotte G, Demeester P, Van de Peer Y, Fostier J. 2017. Evaluation of the impact of Illumina error correction tools on de novo genome assembly. *BMC Bioinformatics* 18:374. <https://doi.org/10.1186/s12859-017-1784-8>.
20. Salzberg SL, Phillippy AM, Zimin A, Puiu D, Magoc T, Koren S, Treangen TJ, Schatz MC, Delcher AL, Roberts M, Marçais G, Pop M, Yorke JA. 2012. GAGE: a critical evaluation of genome assemblies and assembly algorithms. *Genome Res* 22:557–567. <https://doi.org/10.1101/gr.131383.111>.
21. Braun P, Grass G, Aceti A, Serrecchia L, Affuso A, Marino L, Grimaldi S, Pagano S, Hanczaruk M, Georgi E, Northoff B, Schöler A, Schlöter M, Antwerpen M, Fasanella A. 2015. Microevolution of anthrax from a young ancestor (M.A.Y.A.) suggests a soil-borne life cycle of *Bacillus anthracis*. *PLoS One* 10:e0135346. <https://doi.org/10.1371/journal.pone.0135346>.
22. Stahl DA, Flesher B, Mansfield HR, Montgomery L. 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. *Appl Environ Microbiol* 54:1079–1084. <https://doi.org/10.1128/aem.54.5.1079-1084.1988>.
23. Manz W, Amann R, Ludwig W, Wagner M, Schleifer K-H. 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of proteobacteria: problems and solutions. *Syst Appl Microbiol* 15:593–600. [https://doi.org/10.1016/S0723-2020\(11\)80121-9](https://doi.org/10.1016/S0723-2020(11)80121-9).
24. Daims H, Stoecker K, Wagner M. 2005. Fluorescence in situ hybridization for the detection of prokaryotes, p 213–239. *In* Osborn AM, Smith CJ (ed), *Advanced methods in molecular microbial ecology*. Bio-Garland, Abingdon, United Kingdom.
25. Daims H, Lückner S, Wagner M. 2006. Daime, a novel image analysis program for microbial ecology and biofilm research. *Environ Microbiol* 8:200–213. <https://doi.org/10.1111/j.1462-2920.2005.00880.x>.
26. Antwerpen MH, Zimmermann P, Bewley K, Frangoulidis D, Meyer H. 2008. Real-time PCR system targeting a chromosomal marker specific for *Bacillus anthracis*. *Mol Cell Probes* 22:313–315. <https://doi.org/10.1016/j.mcp.2008.06.001>.
27. Kuypers J, Jerome KR. 2017. Applications of digital PCR for clinical microbiology. *J Clin Microbiol* 55:1621–1628. <https://doi.org/10.1128/JCM.00211-17>.
28. Bell AD, Usher CL, McCarroll SA. 2018. Analyzing copy number variation with droplet digital PCR. *Methods Mol Biol* 1768:143–160. https://doi.org/10.1007/978-1-4939-7778-9_9.
29. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
30. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
31. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
32. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
33. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. *Nucleic Acids Res* 39:D19–D21. <https://doi.org/10.1093/nar/gkq1019>.
34. Bushnell B, Rood J, Singer E. 2017. BBMerge—accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056. <https://doi.org/10.1371/journal.pone.0185056>.
35. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
36. Lärkeryd A, Myrtenäs K, Karlsson E, Dwibedi CK, Forsman M, Larsson P, Johansson A, Sjödin A. 2014. CanSNPer: a hierarchical genotype classifier of clonal pathogens. *Bioinformatics* 30:1762–1764. <https://doi.org/10.1093/bioinformatics/btu113>.