

A *Bacillus anthracis* Genome Sequence from the Sverdlovsk 1979 Autopsy Specimens

Jason W. Sahl,^a Talima Pearson,^a Richard Okinaka,^a James M. Schupp,^b John D. Gillece,^b Hannah Heaton,^b Dawn Birdsell,^a Crystal Hepp,^{a,c} Viacheslav Fofanov,^{a,c} Ramón Nosedá,^d Antonio Fasanella,^e Alex Hoffmaster,^f David M. Wagner,^a  Paul Keim^{a,b}

Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona, USA^a; Division of Pathogen Genomics, The Translational Genomics Research Institute, Flagstaff, Arizona, USA^b; School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA^c; Laboratorio Azul, Azul, Buenos Aires, Argentina^d; Istituto Zooprofilattico Sperimentale di Puglia and Basilicata, Foggia, Italy^e; Centers for Disease Control and Prevention, Atlanta, Georgia, USA^f
 J.W.S. and T.P. contributed equally to this article.

ABSTRACT Anthrax is a zoonotic disease that occurs naturally in wild and domestic animals but has been used by both state-sponsored programs and terrorists as a biological weapon. A Soviet industrial production facility in Sverdlovsk, USSR, proved deficient in 1979 when a plume of spores was accidentally released and resulted in one of the largest known human anthrax outbreaks. In order to understand this outbreak and others, we generated a *Bacillus anthracis* population genetic database based upon whole-genome analysis to identify all single-nucleotide polymorphisms (SNPs) across a reference genome. Phylogenetic analysis has defined three major clades (A, B, and C), B and C being relatively rare compared to A. The A clade has numerous subclades, including a major polytomy named the trans-Eurasian (TEA) group. The TEA radiation is a dominant evolutionary feature of *B. anthracis*, with many contemporary populations having resulted from a large spatial dispersal of spores from a single source. Two autopsy specimens from the Sverdlovsk outbreak were deep sequenced to produce draft *B. anthracis* genomes. This allowed the phylogenetic placement of the Sverdlovsk strain into a clade with two Asian live vaccine strains, including the Russian Tsiankovskii strain. The genome was examined for evidence of drug resistance manipulation or other genetic engineering, but none was found. The Soviet Sverdlovsk strain genome is consistent with a wild-type strain from Russia that had no evidence of genetic manipulation during its industrial production. This work provides insights into the world's largest biological weapons program and provides an extensive *B. anthracis* phylogenetic reference.

IMPORTANCE The 1979 Russian anthrax outbreak resulted from an industrial accident at the Soviet anthrax spore production facility in the city of Sverdlovsk. Deep genomic sequencing of two autopsy specimens generated a draft genome and phylogenetic placement of the Soviet Sverdlovsk anthrax strain. While it is known that Soviet scientists had genetically manipulated *Bacillus anthracis* with the potential to evade vaccine prophylaxis and antibiotic therapeutics, there was no genomic evidence of this from the Sverdlovsk production strain genome. The whole-genome SNP genotype of the Sverdlovsk strain was used to precisely identify it and its close relatives in the context of an extensive global *B. anthracis* strain collection. This genomic identity can now be used for forensic tracking of this weapons material on a global scale and for future anthrax investigations.

Received 22 August 2016 Accepted 1 September 2016 Published 27 September 2016

Citation Sahl JW, Pearson T, Okinaka R, Schupp JM, Gillece JD, Heaton H, Birdsell D, Hepp C, Fofanov V, Nosedá R, Fasanella A, Hoffmaster A, Wagner DM, Keim P. 2016. A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens. *mBio* 7(5):e01501-16. doi:10.1128/mBio.01501-16.

Editor Julian Parkhill, The Sanger Institute

Copyright © 2016 Sahl 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 Paul Keim, PaulKeim@nau.edu.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. External solicited reviewers: Timothy Read, Emory University School of Medicine; Nicholas Loman, University of Birmingham; Matthew Meselson, Harvard University.

Anthrax is a zoonotic disease caused by *Bacillus anthracis* with a relatively small impact on global human health, but it has become notorious and widely feared due to its use and potential as a biological weapon. In its spore form, the bacterium represents a highly stable quiescent entity that is capable of surviving for decades, a critical part of its ecology, global distribution, evolution, and infectivity. The vegetative phase allows for cellular proliferation following spore germination in a host animal. The vegetative form expresses specific mechanisms for avoiding the innate host immunity, with some of these encoded on two large virulence plasmids—pXO1 and pXO2 (1). Adaptive immunity can be highly effective at preventing disease, and inter-

estingly, anthrax was the first bacterial disease mitigated with a vaccine (2). Vaccine development for this pathogen is an important veterinary and public health measure, but research with a potential weapon of mass destruction (WMD) unfortunately can also lead to highly similar research supporting pathogen weaponization. Therefore, the treaty created by the Biological Weapons Convention of 1975 with 175 State Parties prohibited all offensive efforts with any biological agent, including anthrax (3).

The *B. anthracis* spore's stability, potential for aerosolization, and its ability to cause acute pulmonary disease have historically led to multiple nations weaponizing this bacterium. It is well documented that large-scale production of spores was accomplished

by the United States, the United Kingdom, and the Soviet Union (4). Industrial spore production involves numerous quality control features to ensure spore stabilization, particle size, and the retention of virulence with extensive growth. These state-sponsored programs were to cease with the Biological Weapons Convention of 1975. However, there are at least two recent examples of anthrax spores being used in biological attacks: the Aum Shinrikyo cult's attempted liquid dispersal of *B. anthracis* in 1993 (5) and the 2001 United States anthrax letters that killed five and sickened an additional 17 (6).

The offensive anthrax weapons development programs were stopped in the United States and United Kingdom in the 1960s but continued covertly in the Soviet Union for at least another 20 years (4). Soviet, and later Russian, research on anthrax included projects to genetically modify *B. anthracis* strains. First, antibiotic resistance was genetically engineered into the vaccine strain STI-1 using recombinant DNA and a plasmid vector (7). This effort resulted in multidrug resistance to penicillin, rifampin, tetracycline, chloramphenicol, macrolides, and lincomycin with retention of normal colony morphology (7). The stated goal of this research was the development of novel vaccines that allowed the simultaneous use of a live vaccine strain and antibiotics in the case of human exposure. Without the drug-resistant live vaccine strain, long-term antibiotic therapy is required. Second, the program genetically engineered hemolytic properties from *Bacillus cereus* into *B. anthracis* by the transfer of cereolysin AB genes into the STI-1 strain, again via a recombinant plasmid (8). This genetic change resulted in a strain with unique pathogenic features that could overcome the standard STI-1 vaccine protection in animal studies. The generation of a hemolytic *B. anthracis* strain was ostensibly for research purposes to understand basic host immunomodulation during anthrax, yet yielded a strain and strategy that could defeat vaccine protection. Manipulation of the *B. anthracis* genome to change its phenotypic properties can and has been accomplished, raising concerns about dual use.

Evidence of the Soviet anthrax program's continuation and its scale were revealed by the 1979 industrial accident in Sverdlovsk, USSR (now known as Ekaterinburg), where at least 66 people died of inhalational anthrax (9). This event has been shrouded in mystery, with governmental denials and little public investigation, but it does represent one of the largest known human inhalational anthrax outbreaks in history (4). According to local sources (4, 10), in early April 1979 safety air filters were compromised during routine maintenance at the Ministry of Defense's (MOD) Scientific Research Institute of Microbiology (SRIM) spore production facility, known as Compound 19. This resulted in a plume of spores that spread downwind and caused human anthrax cases up to 4 km away and animal cases up to 50 km away (9). Russian pathologists investigated these deaths and generated formalin-fixed tissues from multiple victims for analysis. These specimens showed evidence of anthrax (11) and along with later PCR-based DNA analyses (12–14) that detected *B. anthracis* confirmed that this cluster of deaths was indeed due to anthrax.

Here we have continued the Sverdlovsk anthrax investigation through deep sequencing of the formalin-fixed tissues from two of the victims to generate a draft genomic sequence of the infecting *B. anthracis* strain. In this article, we also report the phylogenetic analysis of single-nucleotide polymorphisms (SNPs) discovered among 193 whole-genome sequences, which provided a phylogenetic context for analysis of the Sverdlovsk samples and can be

used for similar analysis of other samples of interest. This provides a high-resolution analysis with detailed clade and subclade structures defined by a curated SNP database. SNP genotyping accurately places the Sverdlovsk strain into a subclade defined by the Tsiankovskii vaccine strain. We also examine the genome sequences for evidence of genetic engineering and adaptation to large production biology. The results show the power of combining modern molecular biology methods with a high-resolution curated SNP database in order to analyze a *B. anthracis* strain involved in a historic anthrax incident.

RESULTS

A high-resolution reference phylogeny. We have constructed a high-resolution reference phylogeny from a large global *B. anthracis* strain collection. This is presented with collapsed clades (Fig. 1) to illustrate the overall phylogenetic structure but with complete branching details and annotated SNPs in the supplemental material (see Fig. S2 and S3). The global phylogeny is comprised of genomes from 193 strains (see Table S1 in the supplemental material) that represent the global diversity as defined by other subtyping methods such as multilocus variable-number tandem-repeat analysis (MLVA) (15) and canonical SNPs (16–19). Genomic sequence comparisons yielded 11,989 SNPs (5,663 parsimony informative) from orthologous genomic segments (see Table S2 in the supplemental material). This represents an average of only 1 SNP every ~500 bp across the entire genome and breadth of this species. A list of SNPs that define each branch and the homoplastic SNPs is provided in Table S3 in the supplemental material to facilitate efforts by other researchers to place their strains in these established clades.

The deeper phylogenetic relationships (Fig. 1A) are consistent with those reported previously with a more limited number of genomes (16–18, 20–24) as well as across different phylogenetic methods (maximum likelihood using the general time reversible [GTR] model of evolution and neighbor joining). There are three major clades, C being basal to the A and B clades (Fig. 1A). Members of the A clade are most frequently observed across the globe (~90%), with B (~10%) and C (<1%) members being much less frequent (16). The A clade can be divided into four major monophyletic subclades, with the “Ancient A” group being basal to the other subclades (Fig. 1A). Members of the “TransEurAsia” (TEA) subclade are most commonly observed as they have been highly successful across large and diverse geographic areas (16).

The unusually short lengths of the deepest branches of the TEA clade, coupled with the high frequency of isolates and geographic expansion, are indicative of a rapid and extensive evolutionary radiation (Fig. 1B). Many sublineages of this clade diverged before mutations occurred, leading to a lack of synapomorphic characters (shared alleles that could group some of these sublineages together) and the existence of a large polytomy (a node with 7 immediate descendant lineages: Tsiankovskii, STI, Pasteur, Heroin, TEA 011, and two lineages with 1 and 2 genomes each). The expansion of each of these lineages also leads to multiple distinct groups, also often with very little topological resolution in the deeper nodes. Given the number of isolates assigned to the TEA 011 group, the TEA clade can be divided into two main subgroups: paraphyletic TEA 008/011 (A.Br.008/011) and monophyletic TEA (A.Br.011).

Sverdlovsk specimen sequence analysis. By direct DNA sequencing, we generated metagenomic data from paraffin-

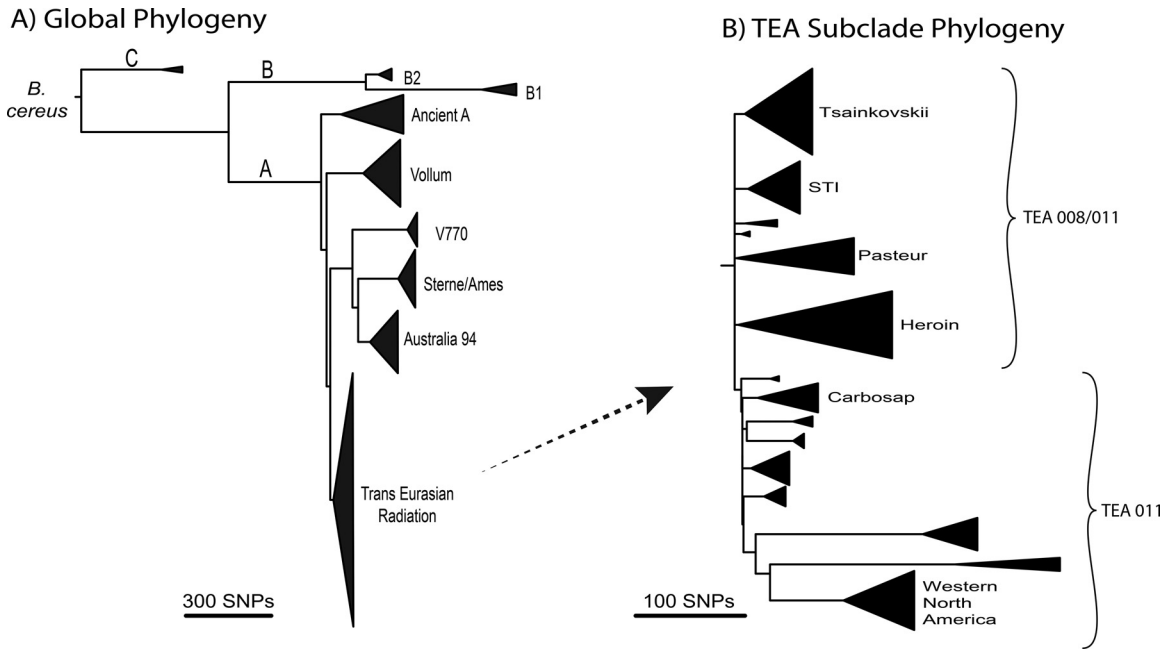


FIG 1 Phylogenetic structure of *B. anthracis*. Orthologous SNPs (11,989 total; 5,663 parsimony informative) from whole-genome sequences were analyzed by maximum parsimony to generate a phylogenetic tree. The major clades are collapsed in this figure, but the complete tree is available in the supplemental material (see Fig. S1). The overall consistency index is 0.98.

embedded formalin-fixed pathology specimens from two anthrax victims from the 1979 outbreak in Sverdlovsk, USSR. The presence of *B. anthracis* DNA in these specimens had been previously established (12), and targeted gene sequencing had also been successful (13, 14); however, until recent technological advances in DNA sequencing, this could only be accomplished by first PCR amplifying small portions of the genome. Sequencing across both the MiSeq and HiSeq Illumina platforms produced ~300 million reads and 20 Gb of nucleotide sequence data across both specimens. A direct mapping of reads against the finished genome of the Ames ancestor genome with BWA-MEM demonstrated that only 1.2% of the total sequence data mapped to the reference genome. This is expected as DNA is from human tissue. The *B. anthracis* coverage represented an average sequencing depth of 24× across the chromosome, with >100× coverage of pXO1 and pXO2 plasmids. These data covered 99% of the Ames ancestor genome, including both plasmids, with at least one read. Alignment stats are shown in Table 1.

From the reads, we assembled the Sverdlovsk genome into 128 contigs with an N_{50} size of 74 kb. A prediction of coding regions (CDSs) with Prodigal (25) on this assembly identified 5,579 CDSs; the same analysis on the Ames ancestor genome identified 5,756

CDSs. This demonstrates that while most of the genome was successfully assembled, parts of the genome may have been dropped from the assembly, most likely from insufficient coverage or collapsed repeats.

Data quality of the Sverdlovsk *B. anthracis* genome. Formalin fixation is known to damage nucleic acids, and this was demonstrated by the small size of the extracted DNA fragments (12), but its effect upon the validity of the Sverdlovsk genomic sequence was unknown. The intrinsic error rate in a sequencing project can be measured by mapping individual sequencing reads to a high-quality reference genome. This generates an estimate of the raw read error rate at each nucleotide and across the whole genome, representing a sequencing quality measurement particularly relevant to SNP identification. In a comparison of *B. anthracis* sequencing reads from Sverdlovsk pathology specimens to those from DNA isolated from culture, we observe a higher number of errors (see Fig. S1 in the supplemental material). The average rate per nucleotide was 0.2% for the culture-generated DNA versus 0.5% for the formalin-fixed tissue. In both cases, a true polymorphism would not be determined from a single read but rather from the consensus of multiple read coverage at any particular genomic position; however, see Sahl et al. (26) for a low-coverage

TABLE 1 Alignment status for Sverdlovsk *B. anthracis* genomes

Library	Type	Length (bp)	No. of reads (pairs)			Coverage (×)			Chromosome breadth at 1× (%)
			Total	Trimmed	Mapped	Chromosome	pXO1	pXO2	
Svd-1	HiSeq	93	6.4E07	4.5E07	5.4E05	8	32	55	96
Svd-1	MiSeq	300	1.0E07	3.4E06	1.5E05	1	3	27	30
Svd-2	HiSeq	93	7.5E07	6.6E07	1.0E06	15	85	104	99
Svd-2	MiSeq	300	1.2E07	3.6E06	1.7E05	1	6	28	36
Combined data			1.6E08	1.2E08	1.9E06	25	126	214	99

SNP-calling strategy. We further examined the consequences of this differential error rate by searching for the conservation of known SNPs along a particular phylogenetic path within these genomes. These were identified in the 193-genome phylogeny (Fig. 1), independent of the Sverdlovsk genome. There were 329 known SNP changes along the branches that connect the Ames ancestor reference to the composite Sverdlovsk genome (Fig. 1; see Table S3 and Fig. S2, S3D, and S3G in the supplemental material). All 329 SNP sites were present in the composite genome assembly. Excluding 29 SNP sites on the pXO1 and pXO2 plasmids because they have higher copy numbers, the coverage per SNP averaged $20\times$ at 273 of the remaining 300 genomic positions on the chromosome. Fourteen of the other chromosomal SNP sites contained less than 10 reads per site but still corresponded exactly to the expected base changes. Overall, we were able to discover and verify all of the known SNPs using the Sverdlovsk pathology specimen sequencing data. Based upon these two error estimations, we are confident that the sequenced genomes are of sufficient quality to justify our conclusions.

Phylogenetic position of the Sverdlovsk strain. Based upon shared SNPs, the Sverdlovsk genomes fall within the “Tsiankovskii” subclade of the TEA 008/011 group (Fig. 1B). Within this group, it is most closely related to two other Asian strains both of which are used as vaccines. There are only 13 SNPs on the branch to the Sverdlovsk genomes, 25 on the branch to Tsiankovskii, and 52 on the branch to Cvac02 (see Tables S2 to S4 in the supplemental material). These three genomes emerge from a polytomy, showing rapid divergence of these lineages before shared SNPs could arise. As this clade is comprised of laboratory strains, this divergence may be due to anthropogenic establishment of different lineages from a laboratory stock. Other clade members were isolated from anthrax-killed animals and are mostly Eastern European in origin, with the exception of one from China and one from Norway. Therefore, with the exception of the three “domesticated” strains, the clade members are naturally occurring wild-type strains.

Sverdlovsk *B. anthracis* genome-specific SNPs. The sequencing and analysis of Sverdlovsk genomes offer an opportunity to detect SNPs and to look for possible strain mixtures or contaminating DNA profiles from two of the tissue samples. To do this, nucleotides from individual reads are tabulated, and less than 100% agreement represents potential errors or mixtures at that genomic position. In particular, we are interested in the 13 SNPs that are unique to Sverdlovsk genomes as they allow a comparison to all other strains outside this group to identify mixtures. Table 2 shows the consensus read results from Sverdlovsk-specific SNPs, and overall there are only 7 variants, resulting in an error rate of 1.6%, which is only slightly higher than the overall error rate of 0.5%. In addition, we note that 6 of the 7 differences are located near the ends of reads, where the error rate is higher (data not presented). One SNP (NC_007530: 5138018) was detected between the two specimens, and this contrast appears to represent a real difference as it was supported by >18 reads. A small number of SNPs between these two specimens might be observed given the population size associated with large-scale production and subsequent amplification *in vivo*. Otherwise, we find no evidence in these two particular Sverdlovsk specimens for strain mixtures. It is important to recognize that these two specimens did not show mixed alleles at the *vrrA* locus analyzed by Jackson et al. (12).

TABLE 2 Read mixtures at Sverdlovsk genome-specific SNPs

SNP site	SNP	Read depth (\times)	% of consensus	Genomic element
28576	G to A	78	97.2	pXO2
59151	C to T	49	100	Chromosome
112138	A to G	36	100	Chromosome
602999	C to A	34	100	Chromosome
1359179	C to T	27	100	Chromosome
2138718	A to T	20	100	Chromosome
2979549	A to G	27	100	Chromosome
3593664	T to C	31	89.9	Chromosome
4034596	G to A	17	93.7	Chromosome
4236707	T to C	30	100	Chromosome
4504222	C to T	22	100	Chromosome
4896833	T to C	17	100	Chromosome
5186004	C to A	46	100	Chromosome

Genetic engineering evidence. Particular genes and SNP signatures in the Sverdlovsk genomes were examined for evidence of genetic manipulation of this strain. In the chromosome, fluoroquinolone resistance is known to be determined by amino acid changes in the *gyrA* and *parC* genes (27), rifampin resistance is associated with changes in the *rpoB* gene (28), and penicillin resistance is associated with changes in β -lactamase gene expression (29). With regard to amino acid changes in associated genes, the Sverdlovsk genomes contained wild-type drug-susceptible alleles. The cereolysin genes and plasmid sequences used by Russian scientists to alter *B. anthracis* phenotypes (7, 8) were not present. In addition, the read data were examined for other common genetic engineering vectors, which were not detected, from an alignment of raw reads against the NCBI UniVec database. The alignment of the 128 contigs to the Ames ancestor revealed no novel genes (Fig. 2), though this was not a closed genome. Hence, there is no evidence from this analysis of either molecular-based genetic engineering or classical bacteriological selection for altered drug resistance phenotypes.

DISCUSSION

The *B. anthracis* global phylogeny is one of the most robust evolutionary reconstructions available for any species. This is possible because orthologous SNPs represent highly stable evolutionary characters with very low homoplasy, and their rarity in this genome precludes any effects from mutational saturation. This species' evolutionary reconstruction is a function of its spore-vegetative cycle biology and, in particular, its ecological niche. The dormant spore stage is important for its dispersal and transmission, limiting evolutionary changes and restricting interactions with near-neighbor *Bacillus* species, making it resistant to horizontal gene transfer. Hence, the *B. anthracis* pan-genome is only slightly larger than the core genome, with variation primarily due to decay via gene deletion. Environmental growth outside the host is possible, but does not appear to represent a significant opportunity to shape this bacterium's genome and evolution. Long quiescent periods in the spore phase may create a “time capsule” where few or no mutations are generated, resulting in a highly homogeneous pathogen. In this sense, its niche differs from its close relative *B. cereus*, which is environmentally adapted with occasional pathogenic replication in a host (30). Fortunately, the genome variation that we can identify through whole-genome sequencing generates insights into anthrax history and allows predictions about its ecology.

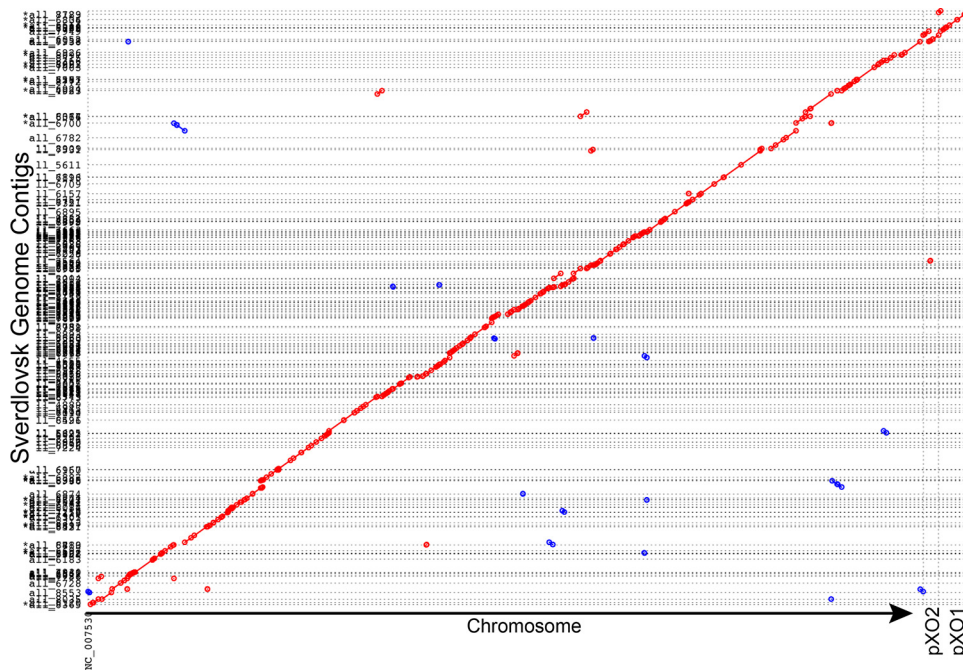


FIG 2 Sverdlovsk contigs aligned to the Ames ancestor genome. The reads from both autopsy specimens were combined for *de novo* assembly, which resulted in 128 contigs. These are aligned against the Ames ancestor chromosome and two plasmids, and the synteny was visualized with mummerplot. Greater than 99% of the Ames ancestor genome is represented by these 128 contigs.

The clade structure we observe with whole-genome sequencing is consistent with previous descriptions using lower-resolution methods or few genome sequences. What we add in this report is the precise definition of branching points, accurate branch length determinations, and the definition of canonical evolutionary characters for strain identification. Branch topology determination has been problematic with other molecular methods because of the abundance of short branches and polytomies at critical positions in the evolutionary structure. The A clade itself, but in particular its subclade TEA, is evidence for evolutionary radiations representing genetic bottlenecks, long-distance dispersal, and bursts in the fitness of these lineages. Even in a radiation, binary fission of replicating bacterial cells should result in phylogenetic structure that could be identified with sufficiently discriminatory methods. But in some cases, such as with the TEA clade, even whole-genome analysis does not yield topological phylogenetic structure, arguing for a very tight genetic expansion. This subclade contains a large portion of the world's anthrax burden (16), making this radiation event seminal. Molecular clock analyses for 106 subroot dated isolates (see Table S1 and Fig. S4 in the supplemental material) and the 48 dated TEA isolates (see Fig. S5 in the supplemental material) have revealed a complete lack of temporal signal among members of this relatively contemporary data set, leaving the exact timing of this radiation dependent upon phylogeographic hypotheses. These models are controversial and vary widely in their temporal predictions (22, 31). To ensure that the lack of molecular clock signal is not due to error arising from various sequencing methods, we pruned the phylogeny to clade A isolates with sister taxa that have dates of isolation within 5 years of each other. We then removed all non-parsimony informative sites, such that only shared SNPs were used to reconstruct the phylogeny as we assume that sequencing errors are unlikely to

occur on shared branches. As in the former root-to-tip analyses, a temporal signal was not evident (see Fig. S6 in the supplemental material). The lack of a consistent SNP substitution rate could be due to variation in the spore phase of the *B. anthracis* life cycle. However, other spore-forming bacteria have demonstrated temporal signal in their phylogenies (32), suggesting that specific ecological dynamics related to sporulation/infection rates must also be involved for anthrax. Ancient genomes from archeological sites would greatly assist in the temporal calibration of key branch points.

Detailed genome databases are a great resource for public health and forensic investigations of disease outbreaks (33). As disease events occur, they allow for the real-time matching of similar types and source identification. But pathogens are dynamic, and databases must be continually updated with isolates from contemporary outbreaks. For some pathogens, a few months can allow for genomic divergence that will make source tracking problematic (34, 35). The availability of high-quality reference databases sets the stage for further sampling (23). It is important to define the relevant subpopulation for additional investigative sampling (36), and this will not be possible prior to a disease outbreak.

Inspired by other preserved pathology tissue DNA analyses (37), two *B. anthracis* genome sequences from victims of the Soviet military accident in Sverdlovsk, USSR, were generated by deep sequencing of formalin-fixed autopsy specimens. Although only ~1.2% of the sequenced reads were associated with the pathogen, enough information was obtained for high-resolution phylogenetics and for draft genome assemblies. A higher than normal error rate was observed in the Sverdlovsk samples, likely due to the nature of the specimen preservation, but sufficient depth of coverage was still obtained to accurately genotype known SNP loci

and to identify strain-specific polymorphisms. Contigs assembled from the reads are syntenic with reference genomes and consistent with isolates from natural anthrax outbreaks, with no extraneous reads associated with cloning vectors or novel toxins. Additionally, there was no evidence of *B. anthracis* strain mixtures in these two particular specimens. Jackson et al. (12) reported mixed alleles at the *virrA* locus for some tissue samples, but not the two analyzed in this report. The *virrA* locus could not be assembled from these specimens due to its repeat structure, and the other victim specimens had very limited DNA that was prohibitive for metagenomic analysis. Hence, our analysis does not eliminate the possibility that mixed strains were involved in the Sverdlovsk anthrax outbreak.

The Soviet “battle strain” 836 was isolated from nature (10) and used for industrial spore production in the 1960s and 1970s, which was mostly prior to the advent of recombinant DNA methods. Traditional selection for mutants resistant to antibiotic resistance was certainly possible prior to 1979, but no such mutations are evident in the Sverdlovsk strain genomes. The great similarity of the genomes to other natural isolates argues for minimal laboratory manipulation. It is well established that *B. anthracis* attenuates with laboratory culturing, and selection for drug resistance frequently has secondary phenotypic consequences that would not be desirable for a weapons strain (27). All of this is highly suggestive of a weapons program that identified a suitable strain, maintained master cell stocks to avoid extensive passage, and performed minimal manipulations in order to maintain virulence. This strategy must have been used to produce large quantities of highly virulent material, as evidenced by the anthrax deaths in 1979.

MATERIALS AND METHODS

Sverdlovsk specimen DNA sequencing. DNA was extracted from paraffin-embedded formalin-fixed tissues from two victims as previously described (12). These extracts were characterized by quantitative PCR (qPCR) (13), and the two samples (Svd-1, 7.RA93.15.15, spleen; Svd-2, 21.RA93.38.4, lymph node) with the lowest threshold cycle (C_T) values were subjected to Illumina sequencing, first on a MiSeq and later on a HiSeq 2000. Sequencing libraries were constructed using the standard Kapa Biosystems Illumina NGS Library reagent kit (catalog no. KK8232; Kapa Biosystems, Boston, MA), using 12 cycles in the final amplification reaction. Due to the highly degraded nature of the input DNA, fragment size selection prior to library preparation targeted fragments that were <500 bp. Both samples yielded libraries with enough material for sequencing and were pooled and then sequenced using an entire MiSeq 600 cycle paired-end run with V3 chemistry. This same pool was subsequently sequenced on a HiSeq 2000, using two lanes.

Sequence analysis. Sequencing adapters were trimmed from reads with Trimmomatic (38). For SNP discovery, reads were aligned against the finished genome of the Ames ancestor (NC_007530, NC_007322, and NC_007323) with BWA-MEM (39), and SNPs were called with the UnifiedGenotyper method in GATK (40, 41). These methods were wrapped by the NASP pipeline (<http://tgennorth.github.io/NASP/>) (42). Functional information was applied to SNPs with SnpEff (43).

Error profile analysis. To understand the error profiles in the Sverdlovsk genomes, reads were aligned against the Ames ancestor with BWA-MEM, and for each position, the number of alleles that conflicted with the dominant allele was divided by the total number of bases at the position; this value was considered the per-base error rate. As a control, this procedure was also performed for another genome (A0362) in the same phylogenetic group. Error rates were binned into different categories and represented as a histogram (see Fig. S1 in the supplemental material).



FIG 3 The Tsiankovskii clade. A phylogenetic tree of the closest relatives to the Sverdlovsk genomes is shown. One SNP was discovered between Svd-1 and Svd-2. The Sverd_combined genotype is identical to Svd-2.

Genome assembly. To obtain a draft genome assembly, reads from both victims were combined and assembled with SPAdes v.3.6.0 (44). The first 200 bases of each contig were aligned against the GenBank (45) nucleotide database with BLASTN (46) to identify contigs not associated with *B. anthracis*; contigs that significantly aligned against human sequence were removed from the assembly. The contiguity of the assembly was then improved through a reference-guided approach with Align-Graph (47), using the Ames ancestor as the reference. The assembly was polished with Pilon v.1.3.0 (48), resulting in 128 contigs. A dot plot analysis using mummerplot (49) was used to examine the synteny against the Ames ancestor as the reference.

Phylogenetic reconstructions. We compared the genomes of 193 strains of *B. anthracis* (see Table S1 in the supplemental material) against the Ames ancestor to find SNPs (see Table S2 in the supplemental material) using the *in silico* Genotyper (50) and the Northern Arizona SNP Pipeline (42). All SNP loci—even those that are missing in some of the genomes—were retained for phylogenetic analyses. We used parsimony criteria and a heuristic search with default options using PAUP 4.0b10 (51) to infer phylogenetic trees. We report homoplasy using the consistency index as a measure of accuracy (52) as bootstrapping is a poor measurement of accuracy for trees with little homoplasy (53) in clonal organisms (20, 21). It should be noted, however, that the consistency index is influenced by the number of taxa, impacting direct comparisons across trees. The phylogeny for all *B. anthracis* genomes was rooted according to Pearson et al. (21). Trees of individual clades and subclades were rooted using a *B. anthracis* strain from another clade or the first strain to diverge from the rest of the group as determined by the overall phylogeny of *B. anthracis*. Phylogenetic branches were named according to precedent (16) and designated on trees (Fig. 3; see Fig. S2 and S3 in the supplemental material). In short, each branch contains a prefix “A.Br,” “B.Br,” “A/B.Br,” or “C.Br,” depending on the major clade designation, followed by an assigned number based upon the order of branch discovery within each of the major clades. This method maintains the branch name from previous publications and allows for the identification of novel branches. However, branch numbers of adjacent branch numbers will often not be contiguous. For each SNP, the branches on which character state changes occurred, as determined by PAUP (51) using the Describe-Trees command, is listed in the supplemental material (see Table S3 in the supplemental material).

For evolutionarily stable characters such as SNPs found in clonal organisms like *B. anthracis*, a single locus can define a branch and thus serve as a “canonical SNP” (16, 20, 21, 54). As such, the character states of only a small number of SNP loci need to be interrogated in order to place an unknown strain into the established phylogenetic order. The list of SNPs on each branch (see Table S3) thus serves as a resource of signatures that can be used to define a branch. However, new genome sequences will cause existing branches to be split, requiring additional branch names and updating the branch designation of these SNPs.

Accession number(s). All reads were submitted to the NCBI Sequence Read Archive for 21.RA93.38.4 (SRR2968141 and SRR2968216) and 7.RA93.15.15 (SRR2968143 and SRR2968198).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01501-16/-/DCSupplemental>.

Figure S1, PDF file, 0.3 MB.
 Figure S2, PDF file, 0.5 MB.
 Figure S3, PDF file, 0.4 MB.
 Figure S4, PDF file, 1 MB.
 Figure S5, PDF file, 0.1 MB.
 Figure S6, PDF file, 1.3 MB.
 Table S1, XLSX file, 0.03 MB.
 Table S2, XLSX file, 7 MB.
 Table S3, XLSX file, 0.3 MB.
 Table S4, XLSX file, 0.04 MB.

ACKNOWLEDGMENTS

We would like to thank the three reviewers who provided critical and constructive comments on the penultimate version of the manuscript.

This work was supported by a contract (HSHQDC-15-C-B0068) from the Department of Homeland Security Science and Technology Directorate.

FUNDING INFORMATION

This work, including the efforts of Paul Stephen Keim, was funded by U.S. Department of Homeland Security (DHS) (HSHQDC-15-C-B0068).

REFERENCES

- Mock M, Fouet A. 2001. Anthrax. *Annu Rev Microbiol* 55:647–671. <http://dx.doi.org/10.1146/annurev.micro.55.1.647>.
- Tigertt WD. 1980. Anthrax. William Smith Greenfield, M.D., F.R.C.P., Professor Superintendent, the Brown Animal Sanatory Institution (1878–81). Concerning the priority due to him for the production of the first vaccine against anthrax. *J Hyg* 85:415–420.
- United Nations Office for Disarmament Affairs. 2016. Biological weapons—the Biological Weapons Convention. United Nations Office for Disarmament Affairs, New York, NY. <https://www.un.org/disarmament/wmd/bio/>.
- Leitenberg M, Zilinskas RA, Kuhn JH. 2012. The Soviet biological weapons program: a history. Harvard University Press, Cambridge, MA.
- 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. <http://dx.doi.org/10.3201/eid1001.030238>.
- Jernigan DB, Raghunathan PL, Bell BP, Brechner R, Bresnitz EA, Butler JC, Cetron M, Cohen M, Doyle T, Fischer M, Greene C, Griffith KS, Guarner J, Hadler JL, Hayslett JA, Meyer R, Petersen LR, Phillips M, Pinner R, Popovic T. 2002. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings (bioterrorism-related anthrax). *Emerg Infect Dis* 8:1019–1029. <http://dx.doi.org/10.3201/eid0810.020353>.
- Stepanov AV, Marinin LI, Pomerantsev AP, Staritsin NA. 1996. Development of novel vaccines against anthrax in man. *J Biotechnol* 44:155–160. [http://dx.doi.org/10.1016/0168-1656\(95\)00092-5](http://dx.doi.org/10.1016/0168-1656(95)00092-5).
- Pomerantsev AP, Staritsin NA, Mockov YuV, Marinin LI. 1997. Expression of cereolysine AB genes in Bacillus anthracis vaccine strain ensures protection against experimental hemolytic anthrax infection. *Vaccine* 15:1846–1850. [http://dx.doi.org/10.1016/S0264-410X\(97\)00132-1](http://dx.doi.org/10.1016/S0264-410X(97)00132-1).
- Meselson M, Guillemin J, Hugh-Jones M, Langmuir A, Popova I, Shelokov A, Yampolskaya O. 1994. The Sverdlovsk anthrax outbreak of 1979. *Science* 266:1202–1208. <http://dx.doi.org/10.1126/science.7973702>.
- Alibek K, Handelman S. 1999. Biohazard: the chilling true story of the largest covert biological weapons in the world—told from the inside by the man who ran it. Dell Publishing, New York, NY.
- Abramova FA, Grinberg LM, Yampolskaya OV, Walker DH. 1993. Pathology of inhalational anthrax in 42 cases from the Sverdlovsk outbreak of 1979. *Proc Natl Acad Sci U S A* 90:2291–2294. <http://dx.doi.org/10.1073/pnas.90.6.2291>.
- Jackson PJ, Hugh-Jones ME, Adair DM, Green G, Hill KK, Kuske CR, Grinberg LM, Abramova FA, Keim P. 1998. PCR analysis of tissue samples from the 1979 Sverdlovsk anthrax victims: the presence of multiple Bacillus anthracis strains in different victims. *Proc Natl Acad Sci U S A* 95:1224–1229. <http://dx.doi.org/10.1073/pnas.95.3.1224>.
- Okinaka RT, Henrie M, Hill KK, Lowery KS, Van Ert M, Pearson T, Schupp J, Kenefic L, Beaudry J, Hofstadler SA, Jackson PJ, Keim P. 2008. Single nucleotide polymorphism typing of Bacillus anthracis from Sverdlovsk tissue. *Emerg Infect Dis* 14:653–656. <http://dx.doi.org/10.3201/eid1404.070984>.
- Price LB, Hugh-Jones M, Jackson PJ, Keim P. 1999. Genetic diversity in the protective antigen gene of Bacillus anthracis. *J Bacteriol* 181:2358–2362.
- Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within Bacillus anthracis. *J Bacteriol* 182:2928–2936. <http://dx.doi.org/10.1128/JB.182.10.2928-2936.2000>.
- 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. <http://dx.doi.org/10.1371/journal.pone.0000461>.
- Marston CK, Allen CA, Beaudry J, Price EP, Wolken SR, Pearson T, Keim P, Hoffmaster AR. 2011. Molecular epidemiology of anthrax cases associated with recreational use of animal hides and yarn in the United States. *PLoS One* 6:. <http://dx.doi.org/10.1371/journal.pone.0028274>.
- Khmaladze E, Birdsell DN, Naumann AA, Hochhalter CB, Seymour ML, Nottingham R, Beckstrom-Sternberg SM, Beckstrom-Sternberg J, Nikolich MP, Chanturia G, Zhgenti E, Zakalashvili M, Malania L, Babuadze G, Tsertsvadze N, Abazashvili N, Kekelidze M, Tsanova S, Imnadze P, Ganz HH, Getz WM, Pearson O, Gajer P, Eppinger M, Ravel J, Wagner DM, Okinaka RT, Schupp JM, Keim P, Pearson T. 2014. Phylogeography of Bacillus anthracis in the country of Georgia shows evidence of population structuring and is dissimilar to other regional genotypes. *PLoS One* 9:e102651. <http://dx.doi.org/10.1371/journal.pone.0102651>.
- Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, Mason J, Vincent G, Drees KP, Beckstrom-Sternberg SM, Phillippy AM, Koren S, Okinaka RT, Chung WK, Schupp JM, Wagner DM, Vipond R, Foster JT, Bergman NH, Burans J, Pearson T, Brooks T, Keim P. 2012. Molecular epidemiologic investigation of an anthrax outbreak among heroin users, Europe. *Emerg Infect Dis* 18:1307–1313. <http://dx.doi.org/10.3201/eid1808.111343>.
- Pearson T, Okinaka RT, Foster JT, Keim P. 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infect Genet Evol* 9:1010–1019. <http://dx.doi.org/10.1016/j.meegid.2009.05.014>.
- Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, Simonson TS, Kachur SM, Leadem RR, Cardon ML, Van Ert MN, Huynh LY, Fraser CM, Keim P. 2004. Phylogenetic discovery bias in Bacillus anthracis using single-nucleotide polymorphisms from whole-genome sequencing. *Proc Natl Acad Sci U S A* 101:13536–13541. <http://dx.doi.org/10.1073/pnas.0403844101>.
- Vergnaud G, Girault G, Thierry S, Pourcel C, Madani N, Blouin Y. 2016. Comparison of French and worldwide Bacillus anthracis strains favors a recent, post-Columbian origin of the predominant North-American clade. *PLoS One* 11:e0146216. <http://dx.doi.org/10.1371/journal.pone.0146216>.
- Keim P, Grunow R, Vipond R, Grass G, Hoffmaster A, Birdsell DN, Klee SR, Pullan S, Antwerpen M, Bayer BN, Latham J, Wiggins K, Hepp C, Pearson T, Brooks T, Sahl J, Wagner DM. 2015. Whole genome analysis of injection anthrax identifies two disease clusters spanning more than 13 years. *EBioMedicine* 2:1613–1618. <http://dx.doi.org/10.1016/j.ebiom.2015.10.004>.
- Pullan ST, Pearson TR, Latham J, Mason J, Atkinson B, Silman NJ, Marston CK, Sahl JW, Birdsell D, Hoffmaster AR, Keim P, Vipond R. 2015. Whole-genome sequencing investigation of animal-skin-drum-associated UK anthrax cases reveals evidence of mixed populations and relatedness to a US case. *Microb Genomics* <http://dx.doi.org/10.1099/mgen.0.000039>.

25. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
26. Sahl JW, Schupp JM, Rasko DA, Colman RE, Foster JT, Keim P. 2015. Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data. *Genome Med* 7:52. <http://dx.doi.org/10.1186/s13073-015-0176-9>.
27. Price LB, Vogler A, Pearson T, Busch JD, Schupp JM, Keim P. 2003. In vitro selection and characterization of *Bacillus anthracis* mutants with high-level resistance to ciprofloxacin. *Antimicrob Agents Chemother* 47:2362–2365. <http://dx.doi.org/10.1128/AAC.47.7.2362-2365.2003>.
28. Vogler AJ, Busch JD, Percy-Fine S, Tipton-Hunton C, Smith KL, Keim P. 2002. Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrob Agents Chemother* 46:511–513. <http://dx.doi.org/10.1128/AAC.46.2.511-513.2002>.
29. Ross CL, Thomason KS, Koehler TM. 2009. An extracytoplasmic function sigma factor controls beta-lactamase gene expression in *Bacillus anthracis* and other *Bacillus cereus* group species. *J Bacteriol* 191:6683–6693. <http://dx.doi.org/10.1128/JB.00691-09>.
30. Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK, Sozhamannan S, Mateczun AJ, Du L, Read TD. 2012. Genomic characterization of the *Bacillus cereus sensu lato* species: backdrop to the evolution of *Bacillus anthracis*. *Genome Res* 22:1512–1524. <http://dx.doi.org/10.1101/gr.134437.111>.
31. Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, Hoffmaster AR, Trim CB, Trim CP, Chung WK, Beaudry JA, Jiang L, Gajer P, Foster JT, Mead JI, Ravel J, Keim P. 2009. Pre-Columbian origins for North American anthrax. *PLoS One* 4:e4813. <http://dx.doi.org/10.1371/journal.pone.0004813>.
32. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CL, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TE, Walker AS. 2013. Diverse sources of *C. difficile* infection identified on whole-genome sequencing. *N Engl J Med* 369:1195–1205. <http://dx.doi.org/10.1056/NEJMoa1216064>.
33. Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J. 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. *Emerg Infect Dis* 18:e1. <http://dx.doi.org/10.3201/eid1811.120453>.
34. Eppinger M, Pearson T, Koenig SS, Pearson O, Hicks N, Agrawal S, Sanjar F, Galens K, Daugherty S, Crabtree J, Hendriksen RS, Price LB, Upadhyay BP, Shakya G, Fraser CM, Ravel J, Keim PS. 2014. Genomic epidemiology of the Haitian cholera outbreak: a single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 5:e01721-14. <http://dx.doi.org/10.1128/mBio.01721-14>.
35. Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Keim PS, Aarestrup FM. 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *mBio* 2:e00157-11. <http://dx.doi.org/10.1128/mBio.00157-11>.
36. Keim P, Pearson T, Budowle B, Wilson M, Wagner DM. 2011. Microbial forensic investigations in the context of bacterial population genetics, p 545–580. In Budowle B, Schutzer SE, Breeze RG, Keim PS, Morse SA (ed), *Microbial forensics*, 2nd ed. Elsevier, Philadelphia, PA.
37. Devault AM, Golding GB, Waglechner N, Enk JM, Kuch M, Tien JH, Shi M, Fisman DN, Dhody AN, Forrest S, Bos KI, Earn DJ, Holmes EC, Poinar HN. 2014. Second-pandemic strain of *Vibrio cholerae* from the Philadelphia cholera outbreak of 1849. *N Engl J Med* 370:334–340. <http://dx.doi.org/10.1056/NEJMoa1308663>.
38. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
39. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv arXiv:1303.3997v2 [q-bio.GN].
40. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491–498. <http://dx.doi.org/10.1038/ng.806>.
41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <http://dx.doi.org/10.1101/gr.107524.110>.
42. Sahl JW, Lemmer D, Travis J, Schupp J, Gillece J, Aziz M, Driebe E, Drees K, Hicks ND, Williamson C, Hepp C, Smith DE, Roe C, Engelthaler DM, Wagner DM, Keim P. 2016. The Northern Arizona SNP pipeline (NASP): accurate, flexible, and rapid identification of SNPs in WGS datasets. *bioRxiv* <http://dx.doi.org/10.1101/037267>.
43. Kent WJ. 2002. BLAT—the blast-like alignment tool. *Genome Res* 12:656–664. <http://dx.doi.org/10.1101/gr.229202>.
44. 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. <http://dx.doi.org/10.1089/cmb.2012.0021>.
45. Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res* 40:D48–D53. <http://dx.doi.org/10.1093/nar/gkr1202>.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2).
47. Bao E, Jiang T, Girke T. 2014. AlignGraph: algorithm for secondary de novo genome assembly guided by closely related references. *Bioinformatics* 30:i319–i328. <http://dx.doi.org/10.1093/bioinformatics/btu291>.
48. 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. <http://dx.doi.org/10.1371/journal.pone.0112963>.
49. Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. *Curr Protoc Bioinformatics* Chapter 10:Unit 10.3. <http://dx.doi.org/10.1002/0471250953.bi1003s00>.
50. Sahl JW, Beckstrom-Sternberg SM, Babic-Sternberg J, Gillece JD, Hepp CM, Auerbach RK, Tembe W, Wagner DM, Keim PS, Pearson T. 2015. The in silico Genotyper (ISG): an open-source pipeline to rapidly identify and annotate nucleotide variants for comparative genomics applications. *bioRxiv* <http://dx.doi.org/10.1101/015578>.
51. Wilgenbusch JC, Swofford D. 2003. Inferring evolutionary trees with PAUP*. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.4. <http://dx.doi.org/10.1002/0471250953.bi0604s00>.
52. Archie JW. 1996. Measures of homoplasy, p 153–188. In Sanderson MJ, Hufford L (ed), *Homoplasy: the recurrence of similarity in evolution*. Academic Press, San Diego, CA.
53. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791. <http://dx.doi.org/10.2307/2408678>.
54. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM. 2004. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect Genet Evol* 4:205–213. <http://dx.doi.org/10.1016/j.meegid.2004.02.005>.
55. Rambaut A, Lam TT, Carvalho LM, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using Tempest (formerly path-O-Gen). *Virus Evol* <http://dx.doi.org/10.1093/ve/vew007>.
56. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <http://dx.doi.org/10.1093/molbev/msw054>.