

Population genomics of *Bacillus anthracis* from an anthrax hyperendemic area reveals transmission processes across spatial scales and unexpected within-host diversity

Taya L. Forde^{1*}, Tristan P. W. Dennis¹, O. Rhoda Aminu¹, William T. Harvey¹, Ayesha Hassim², Ireen Kiwelu³, Matej Medvecký¹, Deogratus Mshanga⁴, Henriette Van Heerden², Adeline Vogel¹, Ruth N. Zadoks^{1‡}, Blandina T. Mmbaga^{3,5}, Tiziana Lembo^{1†} and Roman Biek^{1†}

Abstract

Genomic sequencing has revolutionized our understanding of bacterial disease epidemiology, but remains underutilized for zoonotic pathogens in remote endemic settings. Anthrax, caused by the spore-forming bacterium *Bacillus anthracis*, remains a threat to human and animal health and rural livelihoods in low- and middle-income countries. While the global genomic diversity of *B. anthracis* has been well-characterized, there is limited information on how its populations are genetically structured at the scale at which transmission occurs, critical for understanding the pathogen's evolution and transmission dynamics. Using a uniquely rich dataset, we quantified genome-wide SNPs among 73 *B. anthracis* isolates derived from 33 livestock carcasses sampled over 1 year throughout the Ngorongoro Conservation Area, Tanzania, a region hyperendemic for anthrax. Genome-wide SNPs distinguished 22 unique *B. anthracis* genotypes (i.e. SNP profiles) within the study area. However, phylogeographical structure was lacking, as identical SNP profiles were found throughout the study area, likely the result of the long and variable periods of spore dormancy and long-distance livestock movements. Significantly, divergent genotypes were obtained from spatio-temporally linked cases and even individual carcasses. The high number of SNPs distinguishing isolates from the same host is unlikely to have arisen during infection, as supported by our simulation models. This points to an unexpectedly wide transmission bottleneck for *B. anthracis*, with an inoculum comprising multiple variants being the norm. Our work highlights that inferring transmission patterns of *B. anthracis* from genomic data will require analytical approaches that account for extended and variable environmental persistence, as well as co-infection.

DATA SUMMARY

Isolate Sequence Read Archive (SRA) accession numbers, along with metadata and sequence quality metrics, are available at the University of Glasgow's data repository, Enlighten: <http://dx.doi.org/10.5525/gla.researchdata.1217>. Scripts used in this study can be accessed on GitHub at <https://github.com/matejmedvecký/anthraxdiversityscripts> and https://github.com/tristanpwdennis/anthrax_diversity.

Received 21 September 2021; Accepted 10 December 2021; Published 21 February 2022

Author affiliations: ¹Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, UK; ²Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa; ³Kilimanjaro Clinical Research Institute, Kilimanjaro Christian Medical Centre, Moshi, Tanzania; ⁴Tanzania Veterinary Laboratory Agency, Northern Zone, Arusha, Tanzania; ⁵Kilimanjaro Christian Medical University College, Moshi, Tanzania.

***Correspondence:** Taya L. Forde, taya.forde@glasgow.ac.uk

Keywords: *Bacillus anthracis* endemic anthrax; phylogenomics; simulation modelling; within-host diversity.

Abbreviations: IQR, interquartile range; LMIC, low- and middle-income country; NCA, Ngorongoro Conservation Area; SRA, Sequence Read Archive; WGS, whole-genome sequencing.

‡Present address: Sydney School of Veterinary Science, University of Sydney, Sydney, Australia.

Sequence reads for all the isolates sequenced in this study are available from the European Nucleotide Archive SRA under BioProject accession number PRJEB45684.

†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary tables, nine supplementary figures and one supplementary file are available with the online version of this article.

000759 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Impact Statement

Pathogens transmitted between animals and people affect the health and livelihoods of farmers, particularly in developing countries dependent on livestock. Understanding over what distances these pathogens are transmitted and how they evolve is important to inform control strategies towards reducing disease impacts. Information on the circulation of *Bacillus anthracis*, which causes the often-lethal disease anthrax, is lacking for settings where the disease is commonplace. Consequently, we examined its genetic variability in an area in Tanzania where anthrax is widespread, including cases occurring closely both in space and time, and within individual livestock carcasses. We found no clear link between how closely cases were sampled and their genetic similarity. We suspect this lack of congruence is primarily driven by large-scale livestock movements, which control efforts should take into consideration, and long periods during which infectious bacteria lie dormant. Another significant finding was the co-occurrence of multiple *B. anthracis* types within individual hosts, suggesting animals are commonly infected with a mixture of variants. This needs to be accounted for when investigating possible connections between cases.

INTRODUCTION

Genomic data have the potential to transform our understanding of the evolution and epidemiology of pathogens of public-health importance [1]. However, this potential has yet to be fully harnessed for many zoonotic diseases that occur in hard-to-reach areas. Anthrax remains endemic in many low- and middle-income countries (LMICs) worldwide [2]. It is a disease characterized by sudden deaths in herbivorous livestock and wildlife, and can also cause serious, potentially fatal disease in people [3]. Anthrax is classified among the neglected zoonoses: a group of diseases shared by animals and people that, due to their occurrence in remote, disadvantaged communities, collectively receive less than 0.1% of international global health assistance [4]. As for many neglected zoonoses, there is little genomic data for *Bacillus anthracis*, the bacterium that causes anthrax, from endemic LMIC settings where surveillance tends to be limited. Such data could help to improve our understanding of transmission processes, such as how *B. anthracis* is spread within and between outbreaks, and ultimately contribute to more informed disease management.

The genomic diversity of *B. anthracis* has been well-described at a global scale. Isolates can be broadly divided into three major clades (A, B, C), of which the A clade is the most widespread and globally dominant [5–7]. Isolates from most *B. anthracis* lineages have been found across geographically widespread areas, often spanning multiple continents [7]. While particular variants often predominate regionally, high lineage diversity has also been reported, including co-circulation of strains from multiple lineages [8, 9]. How *B. anthracis* diversity is structured at smaller scales is less well defined. The pathogen has limited genomic diversity compared to other bacterial species (i.e. is genetically monomorphic), rendering standard genotyping methods such as multilocus sequence typing insufficiently discriminatory [10]. Therefore, a hierarchical genotyping scheme known as PHRANA was developed specifically for *B. anthracis*, based on quickly evolving repetitive regions nested within more phylogenetically stable markers – canSNPs – that distinguish among the major lineages [5, 11]. Variants of this scheme have been used to examine the diversity of *B. anthracis* in several endemic settings globally, including in a few African countries [8, 12–14]. Genome-wide SNP data would offer higher resolution for discriminating among closely related isolates. However, whole-genome sequencing (WGS) features in only a few studies of local *B. anthracis* diversity [15–18], and has rarely been conducted outside Europe [19], so the potential for phylogenomic data to be used to understand transmission patterns within hyperendemic areas has yet to be explored.

Transmission of *B. anthracis* occurs primarily through the environment. After causing the death of the animal host, vegetative bacteria are released into the environment via bodily fluids. Here, upon exposure to oxygen and cues related to a lack of nutrients, these bacteria sporulate and can persist in a dormant yet infectious state for several decades [3, 20]. While the viability of spores decreases over time, new *B. anthracis* infections could theoretically arise from recent cases or cases that occurred several years or even decades previously. How this environmental persistence shapes the spatio-temporal diversity of *B. anthracis* in endemic settings has never been investigated.

In molecular epidemiological studies of bacterial pathogens, a single isolate is typically sequenced from each individual case. However, this approach fails to recognize the bacterial population diversity that may exist within the host [21, 22]. Such diversity can either result from mutations that arise during infection, or from heterogeneity (multiple variants) in the inoculum, either through co-infection (exposure to multiple variants simultaneously) or superinfection (multiple exposures) [23, 24]. Under those scenarios, a single isolate is unlikely to represent the overall diversity of the pathogen within the host, and as a result this approach can lead to erroneous inferences about transmission pathways [25]. The importance of capturing within-host diversity is, therefore, increasingly recognized [26, 27]. In the case of anthrax, multiple genotypes of *B. anthracis* have been previously found within individual hosts [28–30], but it remains unclear whether this represents a more widespread phenomenon.

The objective of this study was to quantify the genomic diversity of *B. anthracis* at hierarchical spatial scales within the livestock population of a hyperendemic setting. This was accomplished using a unique dataset including: (i) isolates collected throughout a large (~8300 km²) area of northern Tanzania where anthrax is widespread; (ii) among spatio-temporally linked cases; and (iii)

within individual hosts, assessing multiple isolates from the same and different sample types associated with a case (e.g. tissue, blood, soil).

METHODS

Study area

This study was conducted in the Ngorongoro Conservation Area (NCA), part of the Serengeti ecosystem in northern Tanzania, which covers 8292 km². This multiple-land-use area is inhabited by roughly 87000 people [31] and one million livestock (sheep, goats and cattle) (C. C. Chubwa, Veterinary Officer for Ngorongoro District, personal communication). Northern Tanzania remains hyperendemic for anthrax [32], and prior to this study the NCA was recognized as a potential hotspot for this disease [33].

Sample collection

Samples were collected between May 2016 and April 2017 inclusive through active surveillance by a dedicated field team. Sudden deaths in animals reported by community members throughout the NCA were investigated and samples were collected when anthrax was suspected (File S1, available with the online version of this article). When available, the following samples were collected: a piece of skin tissue (tip of the ear if the carcass was still intact, or a piece of hide if the carcass had already been opened); whole blood; swab of blood or body fluid at natural orifices; blood or body-fluid soaked soil from below the carcass; and insects found on or around the carcass. Various metadata were recorded, including the species of animal affected and the location of sampling (File S1). All samples were stored at ambient temperature for up to 6 months at local veterinary facilities until transport to the Kilimanjaro Clinical Research Institute (KCRI) in Moshi, Tanzania, for molecular diagnostics, as previously described [34], with aliquots shipped to the University of Pretoria, South Africa, for selective culture and DNA extraction from *B. anthracis* isolates.

Selective culture, DNA extraction and sequencing

Sample pre-treatment (i.e. to inhibit competition from heat-sensitive bacteria) is described in File S1. Sample homogenates (100 µl) were plated onto both polymyxin-EDTA thallos acetate (PET) selective medium and 5% sheep blood agar (SBA). These were incubated at 37 °C overnight and the plates inspected for growth after 15–24 h incubation. The PET medium was then further incubated and inspected at 48 h. Suspect *B. anthracis* colonies based on typical morphological characteristics were sub-cultured onto SBA for purification and identification (File S1). In parallel, a single colony was streaked onto a new purity plate for nucleic acid extraction. In some instances, multiple isolates were selected from the same sample where the colonies demonstrated differences in morphology but were identified on the same plate and met the selection criteria (File S1).

DNA extracts from 75 isolates from 33 carcasses were submitted for library preparation and sequencing at MicrobesNG (Birmingham, UK). Libraries were prepared using the Nextera XT v2 kit (Illumina) and sequenced on the Illumina HiSeq platform, generating 250 bp paired-end reads.

Bioinformatics and genomic analyses

Reads were adapter trimmed by MicrobesNG using Trimmomatic v0.30 [35], and basic statistics determined using QUAST [36]. Bacterial species identification was confirmed using Kraken [37]. Based on these quality metrics, sequences from two isolates were excluded from further analyses: one due to a low number of reads (<40000), and another that was identified as *Bacillus cereus*. There were some further indications that not all cultures were pure *B. anthracis*, despite multiple rounds of sub-culture (File S1). A reference-based mapping approach and strict variant filtering criteria were implemented to minimize the issues associated with the sequence quality, while making use of as much of the data as possible.

Read mapping and variant calling were performed on the CLIMB computing platform for microbial genomics [38]. Trimmed reads were aligned to the chromosome of the Ames Ancestor reference genome (accession no. NC_007530) using BWA-MEM (version 0.7.17). Picard was used to mark and remove duplicate reads, add read group information, and index the BAM files [39]. Quality metrics for read mapping were obtained using Qualimap [40]. Mapped reads had a mean depth of coverage of 85× across isolates, ranging from 24× to 245× (median 72×). SNPs were detected in individual isolates by VarScan v2.4.4 [41] with parameters set as follows: minimum read depth of 4; minimum base quality of 20; variant allele frequency ≥0.95. Subsequent SNP curation steps are described in File S1. Custom Python scripts for the assessment of read mapping SNP metrics data, variant site filtering and generation of variant call and alignment files (source codes with description of their functionality and usage), along with the final variant call and multiple sequence alignment files, are available on GitHub (<https://github.com/matejmedvecky/anthraxdiversityscripts>).

The alignment of concatenated SNPs was analysed using ModelFinder to determine the most appropriate model of nucleotide substitution [42]. Subsequently, a maximum-likelihood phylogeny was estimated in IQ-TREE [43] under the Kimura-3-parameters (K3P) model, using 1000 ultrafast bootstrap replicates [44]. A distance matrix detailing SNP differences between isolates was

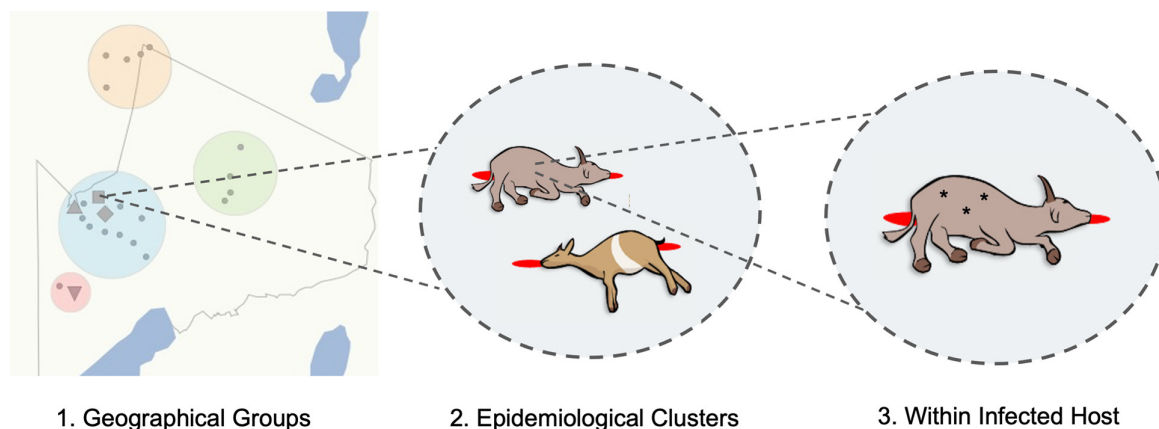


Fig. 1. Hierarchical levels at which the genomic diversity of *B. anthracis* was studied. (1) Sequenced isolates originating from livestock carcasses sampled throughout the anthrax hyperendemic NCA, northern Tanzania, (shown as grey dots in this panel) were categorized into four geographical groups (coloured circles); (2) a subset of these isolates was from spatio-temporally linked pairs of carcasses ($n=4$, represented by grey shapes in panel 1), referred to as epidemiological clusters; (3) multiple isolates ($n=2-4$; represented by asterisks) were sequenced from individual carcasses, either originating from multiple sample types (e.g. tissue and blood), and/or multiple isolates sequenced from a single sample. The shape file for the NCA was provided by Tanzania National Parks (TANAPA) [60].

constructed using `snp-dists` v0.6 (<https://github.com/tseemann/snp-dists>). The distance between GPS points was calculated using the `pointDistance` command in the R package *raster* [45]. Isolation by distance was tested using Mantel test to assess the correlation between SNP distance and Euclidean geographical distance within the R package *ade4* [46]. All program versions and commands used, the distance matrix, as well as small custom scripts are available on GitHub (https://github.com/tristanpwdennis/anthrax_diversity).

To place the newly sequenced isolates within the global phylogeny of *B. anthracis*, 80 WGS data sets from GenBank were accessed (Table S1), and a core-genome alignment generated using `Parsnp` v1.1.2 [47]. The resulting phylogeny indicated that NCA isolates belong to the ancient A lineage. To further resolve the diversity among NCA isolates compared with other publicly available isolates from the same lineage, reads from eight additional isolates available from the Sequence Read Archive (SRA) from a study by Bruce *et al.* [7] were accessed using `Fastq-dump` from the SRA-toolkit: all isolates ($n=4$) belonging to the 3.2 lineage, and two arbitrarily selected isolates from each of the 3.1 and 3.3 lineages (Table S2). These were run through our SNP-calling pipeline as described above, resulting in a sequence alignment file free of `-/N` characters, which was used to infer a phylogeny in `RAXML` v8.2.11 [48] using a GTR (general time reversible) model of nucleotide substitution, and using the Ames Ancestor reference genome as an outgroup.

Assessment of *B. anthracis* diversity at hierarchical spatial scales

The diversity of *B. anthracis* was assessed at multiple hierarchical spatial scales (Fig. 1). Carcasses for which detailed sampling location data were available ($n=32$) could be grouped into four distinct areas within the NCA (central, north, south, east), referred to herein as ‘geographical groups’. To assess the genomic relatedness among spatio-temporally linked cases, on four occasions, samples were collected from two carcasses either from the same or neighbouring households on the same or consecutive days, which we refer to as ‘epidemiological clusters’. Three of these clusters were in the central geographical group, while the fourth was in the southern group of carcasses sampled. Finally, multiple isolates ($n=2-4$) were sequenced from a single carcass for 21 carcasses. These were either (i) isolates from multiple sample types (i.e. tissue, blood, swabs, soil and/or insects) from a single carcass ($n=16$ carcasses) and/or (ii) multiple isolates from the same sample ($n=15$ carcasses).

Simulation modelling

To assess the likelihood of different levels of within-host diversity arising during the course of infection, we performed simulation modelling of *B. anthracis* infection with a homogenous inoculum of varying size (1, 2, 5, 10, 20, 50 and 100 bacterial genomes) to reflect uncertainty in anthrax infectious dose, and across a range of replication cycles (up to 25 generations). In each bacterial generation, genomes underwent a round of replication followed by cell division and accordingly population size doubled in each generation. With an infectious dose d , discrete generations and no die-off, the population size in generation i is therefore $d \times 2^{(i-1)}$. The number of mutations occurring in the replication of each genome was drawn from a Poisson distribution parameterized to reflect the estimated *B. anthracis* mutation rate (i.e. $\lambda=0.0043$). This genome-level mutation rate is based on the genome size of 5.2 million bp and a mutation rate of 8.3×10^{-10} mutations per site [29]; this represents an upper estimate of

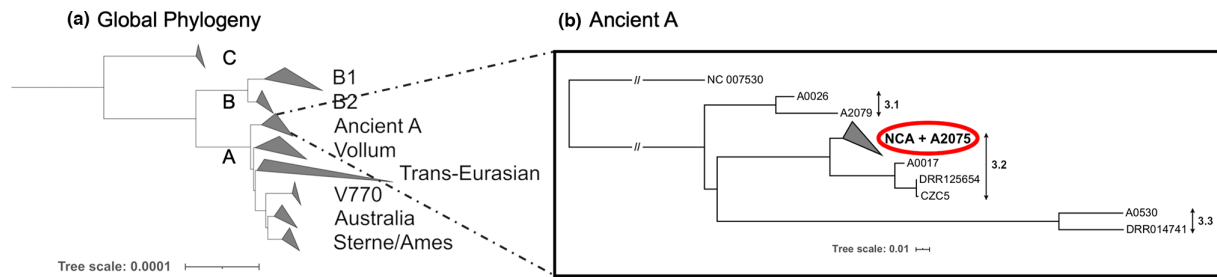


Fig. 2. Phylogenetic position of *B. anthracis* isolated from the NCA within the global population. (a) Global phylogeny of *B. anthracis*, showing the major clades (A, B and C) and sub-lineages. This tree was estimated based on a core SNP phylogeny of 80 publicly available genomes (Table S1). (b) Maximum-likelihood phylogenetic tree of the ancient A lineage (cluster 3 based on work from Bruce *et al.* [7]). All isolates from the NCA form a monophyletic lineage within cluster 3.2, along with the publicly available isolate A2075 (accession no. SRR2968187) isolated in 1999 from a baboon in central Tanzania. This tree was inferred using the general time reversible (GTR) model of nucleotide substitution, using the Ames Ancestor reference genome (accession no. NC_007530) as an outgroup. Tree scales reflect the number of substitutions per site.

the mutation rate and was chosen as we were interested in estimating the upper limits on reasonable expectation of diversity emerging during the course of an infection initiated by a homogenous dose. Results were extrapolated to 40 generations, proposed to be the upper limit on the number of replication cycles during an infection [11]. Details of simulations run are provided in File S1. Pairs of genomes were repeatedly sampled from these simulated populations and the count of mutations separating each pair was calculated. Sampling was performed 100 times per generation for each simulation. Simulations, sampling of simulated populations, and linear models summarizing trends in the outcome of these processes were performed in R [49].

RESULTS

B. anthracis within the NCA is limited to a single subgroup lacking a clear phylogeographical signal

WGS data were obtained from 73 *B. anthracis* isolates from a total of 33 carcass sites sampled throughout the NCA; all isolate metadata are available at <http://dx.doi.org/10.5525/gla.researchdata.1217>. Carcasses were of the following species: sheep ($n=18$), cattle ($n=7$), goats ($n=4$), donkey ($n=1$) or unknown host ($n=3$). All *B. anthracis* isolates from the NCA were found to belong to the ancient A subgroup of clade A (Fig. 2). Within this subgroup, the NCA isolates formed a monophyletic clade within cluster 3.2 as defined by Bruce *et al.* [7], which also contained the isolate A2075 (GenBank accession no. SRR2968187), isolated in 1999 from a baboon in Muhesi Game Reserve, central Tanzania.

A total of 125 SNPs that differed among the NCA isolates and A2075 were retained for analysis, of which 13 were unique to A2075. Twenty-two unique genotypes (i.e. SNP profiles) were found among the 73 NCA isolates (Fig. 3a–c). Based on a rarefaction analysis, the observed genotypic diversity was close to that present throughout the study area (i.e. further sampling would have been unlikely to reveal additional genotypes; Fig. S1). The maximum pairwise nucleotide difference between any two NCA-derived isolates was 49 SNPs [median=24, interquartile range (IQR)=10–35] (Fig. S2). There was no clear relationship between the pairwise nucleotide differences and the geographical distance between sampling locations (Fig. S3), as confirmed by a test for isolation by distance ($r=0.04$, $P=0.214$). Isolates from the central, eastern and southern geographical groups were observed throughout the phylogenetic tree, while all isolates from the northern sampling area were restricted to a single clade that contained the majority of NCA isolates (Fig. 3). In eight instances, identical *B. anthracis* genotypes were found in carcasses from different geographical groups, all of which involved isolates from the central geographical group and one of the other areas, with all areas implicated (Table S3). Identical or nearly identical SNP profiles (one SNP difference) were obtained from carcasses sampled 3–5 months apart on six occasions, and 10 months apart on one occasion (Table S4).

B. anthracis isolates from spatio-temporally linked cases rarely have identical SNP profiles

Between one and four isolates were sequenced from each carcass sampled as part of an epidemiological cluster (i.e. pair of spatio-temporally linked carcasses), resulting in 4–8 isolates per cluster. In most cases, isolates deriving from the same epidemiological cluster were phylogenetically unrelated. Only in one of the four clusters examined (C2) did both carcasses contain isolates with identical core SNP profiles (Fig. 3c). Overall, isolates from different carcasses within the same epidemiological cluster had similar numbers of SNP differences when compared with randomly selected carcasses (linked cases – median=21, IQR=0–33; unlinked cases – median=23, IQR=10–35) (Fig. 4).

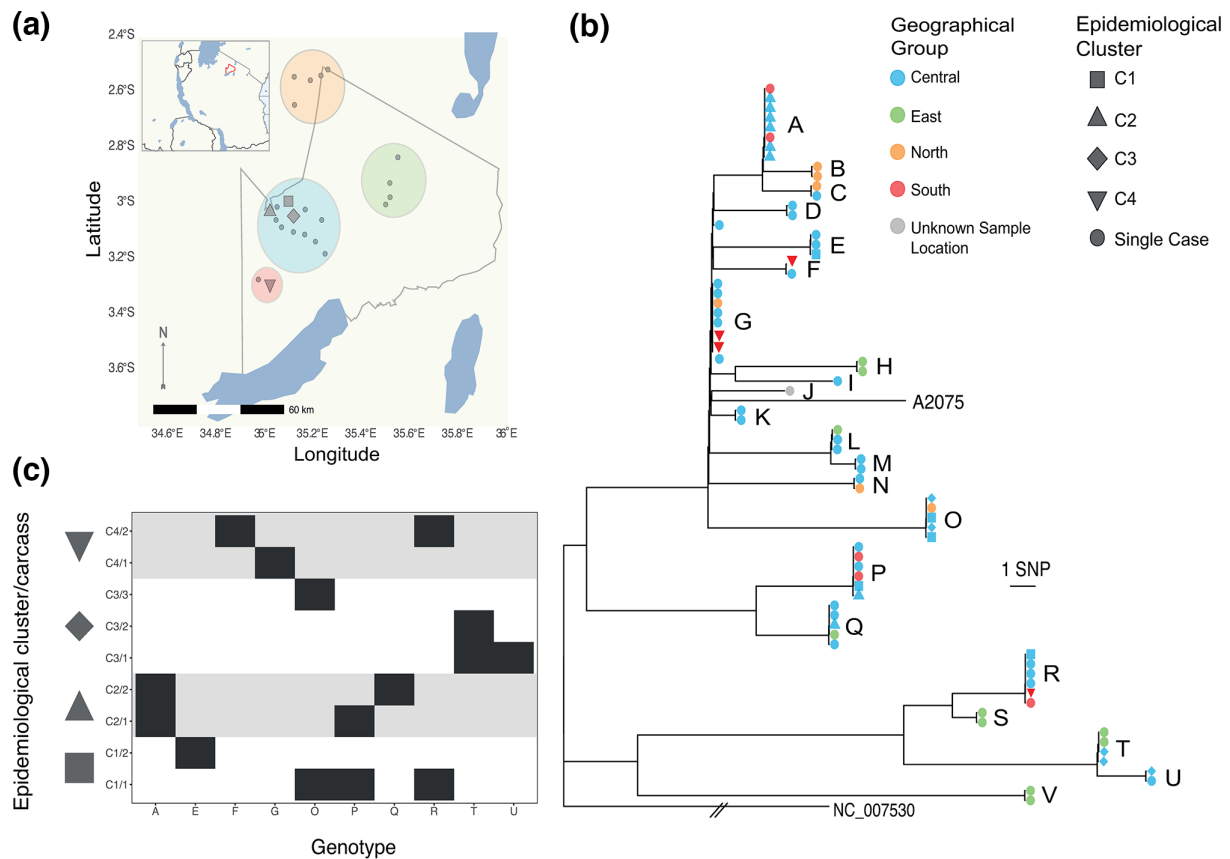


Fig. 3. Phylogeography of *B. anthracis* in the hyperendemic area of the NCA, Tanzania. (a) Spatial distribution of carcasses from which *B. anthracis* isolates were obtained. The map outlines the NCA and shows its location in northern Tanzania (outlined in red in inset). Carcasses were assigned to four geographical groups within the NCA based on spatial proximity, shown by coloured circles. (b) Maximum-likelihood tree estimating the phylogenetic relationship among *B. anthracis* isolates from the NCA. This tree is based on an un-gapped alignment of 125 high-quality core SNPs across the whole chromosome, rooted to the Ames Ancestor reference sequence (accession no. NC_005730) and including the publicly available isolate A2075 (accession no. SRR2968187) originating from central Tanzania. Using the more closely related isolates from cluster 3.2 as an outgroup produced the same root position. Isolates are coloured on the tree based on their collection site (geographical group) within the NCA. Epidemiological clusters of cases (pairs of carcasses sampled from the same or neighbouring households on the same or consecutive days) are distinguished by symbol shape. Letters distinguish the 22 unique genotypes (SNP profiles) detected. (c) Genotypes of *B. anthracis* isolated from within and between pairs of carcasses from the same epidemiological clusters (C1–C4). Individual carcasses are numbered /1 or /2. In cluster C3, two isolates were from a soil sample (C3/2 and C3/3); genotype T from this soil sample was shared with an isolate from C3/2. Otherwise, only in C2 was there evidence of a shared genotype between pairs of carcasses (genotype A). Thus, the level of sampling conducted here (1–4 isolates per carcass) did not produce evidence for the same combinations of genotypes being found among linked carcasses. Isolate-labelled versions of (a) and (b) can be found as Figs S7 and S8. The base earth, river and lake data for the map were downloaded from Natural Earth (<https://www.naturalearthdata.com/>). This figure was plotted in R v.3.6.1 with *ggplot2* [61], with the addition of the *sf* [62] and *ggtree* [63] packages.

Within-host diversity of *B. anthracis* is similar to between-host diversity in the NCA

Overall, the number of pairwise SNP differences between isolates from the same carcass was lower than that found between isolates from different carcasses (same – median=11, IQR=0–34; different – median=23, IQR=10–35) (Fig. 4). However, isolates with multiple distinct genotypes were obtained on 15 of 21 occasions wherein multiple isolates were sequenced from the same carcass, with as many as 43 SNP differences between isolates (Fig. 5). A high level of divergence was seen, regardless of whether isolates were from the same or different sample type (e.g. multiple isolates from a tissue sample versus isolates from tissue and soil samples; Fig. 5a). Divergent genotypes were observed in all carcasses from which three or more isolates were sequenced (12/12), and in a third (3/9) of carcasses for which two isolates were sequenced. A single SNP difference separated one isolate from two others within one carcass (AN16-83) and two SNP differences separated two isolates from a single soil sample (LNA). All other pairwise within-host SNP differences were eight or above. All of the within-host SNP profiles detected were shared with those from other carcasses.

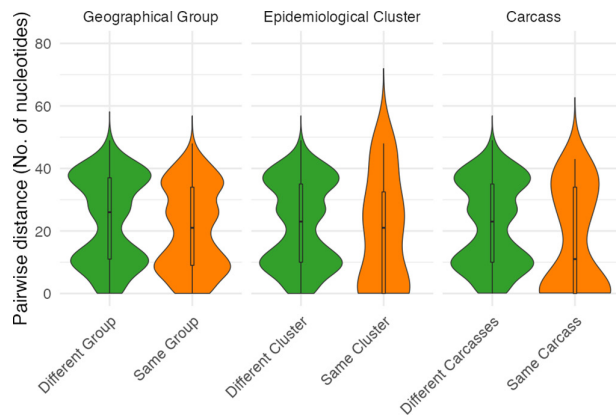


Fig. 4. Comparative numbers of single nucleotide differences among *B. anthracis* isolates from hierarchical spatial scales. Violin plots comparing pairwise nucleotide (SNP) differences between all *B. anthracis* isolates from the NCA versus SNP differences between isolates from: (i) within the same geographical group; (ii) within the same epidemiological cluster (but not from the same carcass); and (iii) within a single carcass. The plot width at each point along the y-axis reflects the number of observations of different pairwise distances. The central boxplot shows median and IQR, with whiskers showing minimum and maximum values up to 1.5 times the IQR.

Observed within-host diversity is unlikely to have arisen during the course of infection

Averaging across simulations with varying initial doses, in the 25th generation, 90.0% of the population was identical to the infecting dose, 9.53% differed by one SNP and 0.489% by two SNPs, with higher numbers of mutations very rare (<0.01%). While stochasticity was present in early generations and particularly at low inoculum doses (Figs S4–S6), the mean proportion of genomes with various numbers of SNPs changed predictably after the initial generations (Fig. 6a, b). Extrapolating to the 40th generation, then ~84% of the population would be expected to be identical to the infecting dose, ~15% differing by one SNP and ~0.9% by two SNPs.

To illustrate how this within-host generated diversity would be captured in our sampling, pairs of genomes were repeatedly sampled from simulated populations and pairwise differences calculated (Fig. 6c, d, File S1). Overall, pairwise differences greater

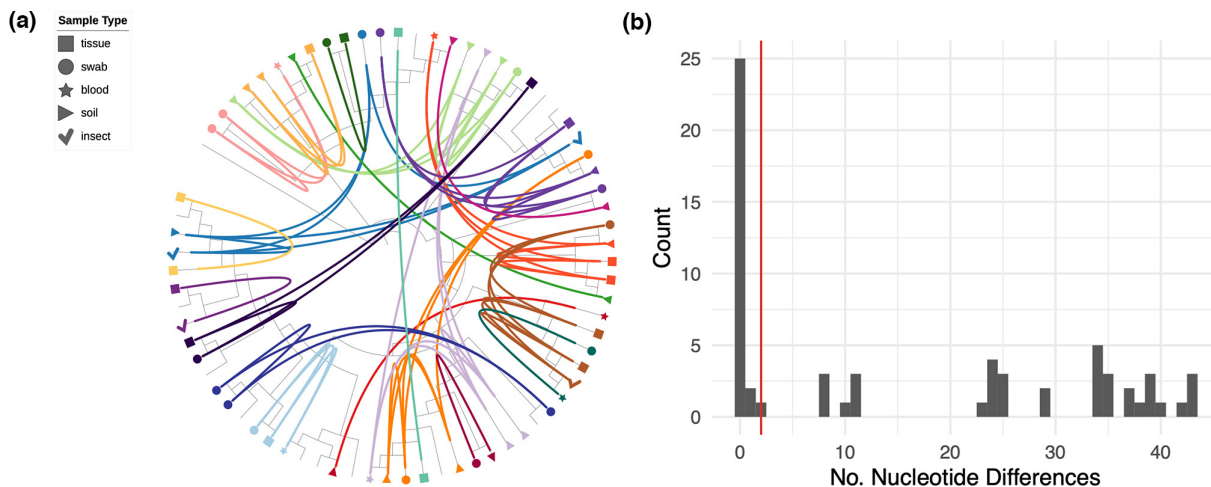


Fig. 5. Within-host diversity of *B. anthracis* among livestock in the NCA. (a) Circularized maximum-likelihood tree – based on high-quality core SNPs – displayed as a cladogram (branch-lengths ignored), rooted to the Ames Ancestor reference genome (accession no. NC_005730). Isolates from the same carcass are shown in the same colour and are linked by inner connecting lines. Isolates without labels are singletons (i.e. only one isolate sequenced per carcass site). Sample type is shown by the different symbol shapes indicated in the legend. The figure was prepared using iTOL [64]. For labelled taxa, see Fig. S9. (b) Histogram showing the relative frequency of pairwise SNP differences among *B. anthracis* isolates collected from the same carcass. The red line shows 99% upper limit of nucleotide differences observed among sampled pairs of genomes based on simulation of within-host evolution. Results suggest that almost all diversity observed within the same infected host is the result of a heterogenous inoculum.

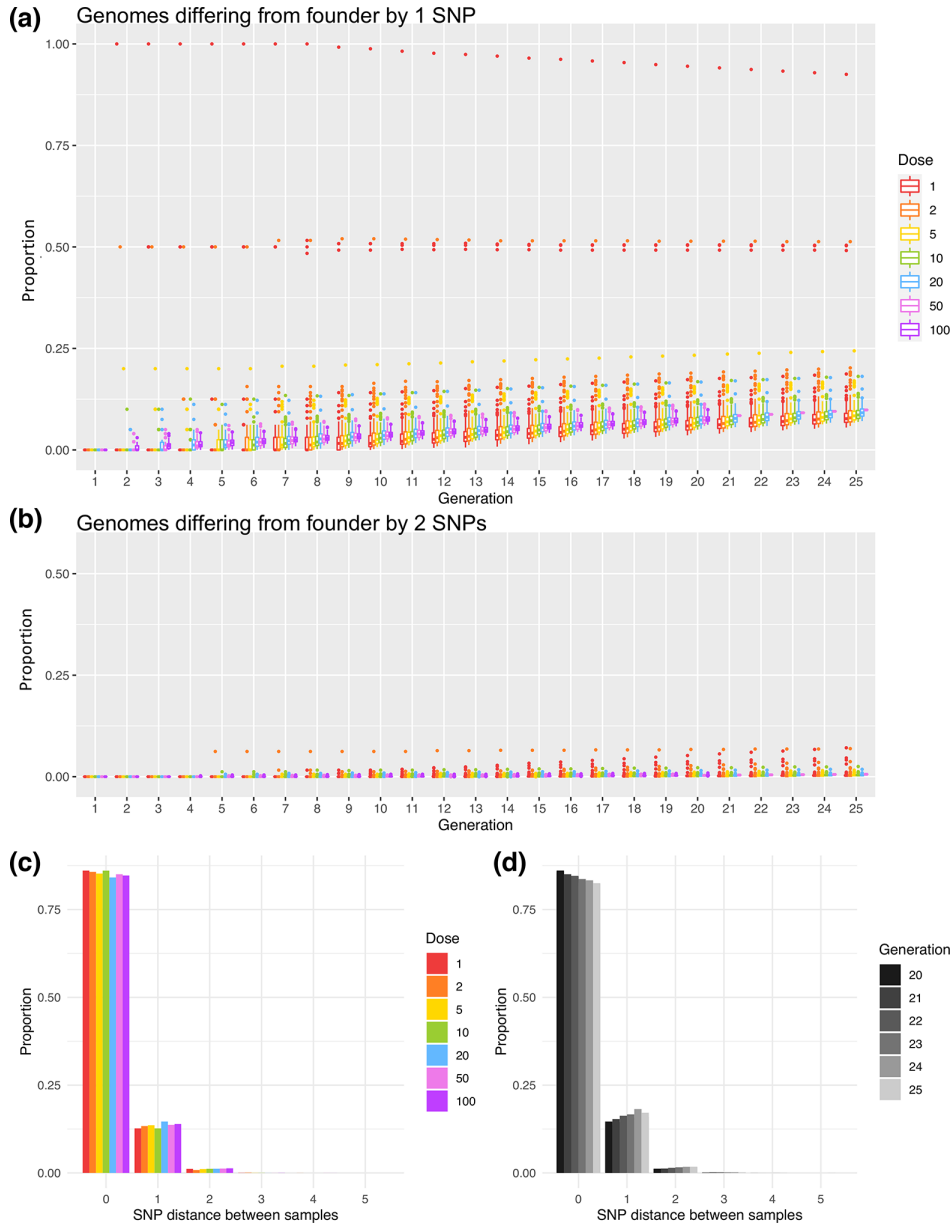


Fig. 6. Simulation of within-host populations and sampling of resulting genetic diversity. Box and whisker plots show the proportion of genomes across simulated populations that differ from the founding genome by either one nucleotide (a) or by two nucleotides (b). Boxes represent the IQR with a line showing the median and with outliers shown as points. Both boxes and outliers are coloured by inoculum size (dose) according to the legend. Simulations were run for 25 generations, or 20 generations for larger inoculum sizes. Greater stochastic heterogeneity was observed for lower starting doses. Note that in (b), only outliers are visible as the proportions observed were very low. (c) Bar plot showing the relationship between inoculum size (dose) and SNP distances between pairs of sampled genomes. Bars represent mean proportions of pairwise SNP differences between genomes sampled from simulated within-host populations. Populations were simulated from various inoculum sizes and were sampled in the 20th generation after 19 replication cycles. Fifty simulations were run from an initial dose of 20, and 100 simulations were run from each of the other inoculum sizes, with each simulated population sampled 100 times in each generation. (d) Bar plot showing the relationship between the number of replication cycles and SNP distances between pairs of sampled genomes. Populations were simulated for 25 generations and samples taken in generations 20–25 are represented in shades of grey according to the legend. Proportions are averaged across simulations initiated with inoculum sizes of 1, 2, 5, 10 and 20 and sampled repeatedly.

than two SNPs occurred in less than 0.2% of all simulated populations (Table S5). Higher SNP distances such as those observed (Fig. 5b) are, therefore, unlikely to arise within-host following infection with a homogenous inoculum.

DISCUSSION

Genomic data for understanding the population structure and transmission patterns of bacterial zoonoses has been limited for LMIC settings where these diseases tend to have the greatest impact. Despite *B. anthracis* having limited genomic diversity in comparison with other bacterial species, WGS provided sufficient resolution to discriminate among isolates collected from within a relatively small geographical area of a few thousand square kilometres. The way in which the genomic diversity was partitioned across hierarchical spatial scales within this area has a series of novel implications for our understanding of how the pathogen is transmitted and evolves during endemic circulation.

***B. anthracis* diversity within the hyperendemic NCA region is limited to a single clade A sub-group**

The NCA is a region where anthrax has likely been endemic for decades, if not centuries. Local community members claim that it has been an issue for their health and that of their livestock throughout living memory. Despite WGS providing sufficient discriminatory power to differentiate among individual *B. anthracis* isolates in this setting, fewer than 50 SNP differences were found among NCA isolates across the 5.2 Mb chromosomal genome, highlighting the degree to which this bacterial pathogen is monomorphic. All 73 sequenced isolates formed a monophyletic cluster within the ancient A lineage – also known as canSNP group A.Br.005/006 or A.Br.034 [6] – a subgroup of clade A comprised mostly of isolates from south-eastern Africa. This contrasts with some previous studies of anthrax diversity in chronically endemic areas reporting co-circulation of isolates from multiple lineages using lower-resolution markers [8, 9]. Our study, which only included isolates collected over 1 year, can be considered a snapshot of recent diversity only. That being said, Tanzanian isolates genotyped in a previous study ($n=17$) were also all found to belong to the ancient A clade [5], demonstrating that this particular lineage is dominant and well-established in this country. The most closely related publicly available isolate (A2075), sampled in central Tanzania approximately 300 km away from our study area, differed from some NCA isolates by only 13 SNPs, which illustrates that highly related *B. anthracis* isolates can be geographically widespread. Broader, longitudinal WGS studies of *B. anthracis* across different regions of Africa will be needed to assess the actual range of individual genotypes over space and time.

No phylogeographical signal observed despite considerable SNP diversity

Despite the considerable diversity observed (i.e. 22 unique genotypes within the sampled population), a phylogeographical signal was not detected at the scale of this study area. The finding of identical SNP profiles across distances of tens of kilometres likely reflects the ecology of anthrax in general and in our study system. First, there are few opportunities for genetic diversity to arise within the *B. anthracis* life cycle, which is characterized by long periods of environmental dormancy in spore form, punctuated by brief interludes of a few days where it develops into its vegetative state and replicates within an infected host. It is estimated that *B. anthracis* undergoes only 20–40 replications per infection [11]. Given its low mutation rate ($5.2\text{--}8.3 \times 10^{-10}$ mutations per site per generation) [5, 29, 50], novel mutations would likely arise in only a small proportion of *B. anthracis* infections, as also supported by our simulations. The short time of active replication within a host also means there is minimal opportunity for horizontal gene transfer with other bacteria [11], which further restricts the ability of *B. anthracis* to diversify. It is believed that *B. anthracis* rarely multiplies outside of a host, although there is some evidence for limited environmental replication by passing through the vegetative state [51, 52]. Viability of *B. anthracis* decays exponentially over time, although infectious spores remain detectable at carcass sites at least 4 years after the death of the animal [53], and under favourable conditions, spores can remain viable for up to several decades [20]. It is, therefore, reasonable to expect that spores from a single anthrax carcass with few or no novel mutations could be the source of subsequent infections over highly variable time periods, a phenomenon referred to by Sahl *et al.* [6] as a ‘time capsule’. This situation violates typical molecular clock assumptions [54], wherein SNPs would be expected to arise at a relatively steady rate within the pathogen population over time. The observation that contemporary isolates from the NCA are phylogenetically basal to isolate A2075, collected nearly two decades earlier, highlights this issue. In order for molecular clock models to be applied to environmentally persistent pathogens such as *B. anthracis* – for instance to estimate how long particular lineages have been in circulation – current analytical frameworks will need to be extended.

In addition to long and variable environmental persistence, animal movements and spatial admixture likely contribute to the lack of phylogeographical structure of *B. anthracis* within this study area. Potential sources of infection are quite spatially restricted, as the highest concentrations of viable spores are found within only a few metres of anthrax carcasses [55], in what have been termed ‘localized infectious zones’ [56]. However, given that the incubation period of anthrax in livestock typically ranges from 1 to 14 days and potentially longer [3], extensive movement can occur between the time of infection and the animal’s death, meaning the site of sampling likely does not reflect the site of infection. In parts of rural Africa where pastoralism is the main form of agriculture, livestock are moved for various reasons, including to access water, pastures and minerals. Such daily and/or seasonal livestock movements could contribute to the observation of identical SNP profiles over distances of tens of kilometres in our study area and the lack of relationship between genetic distance and geographical distance between sampling locations.

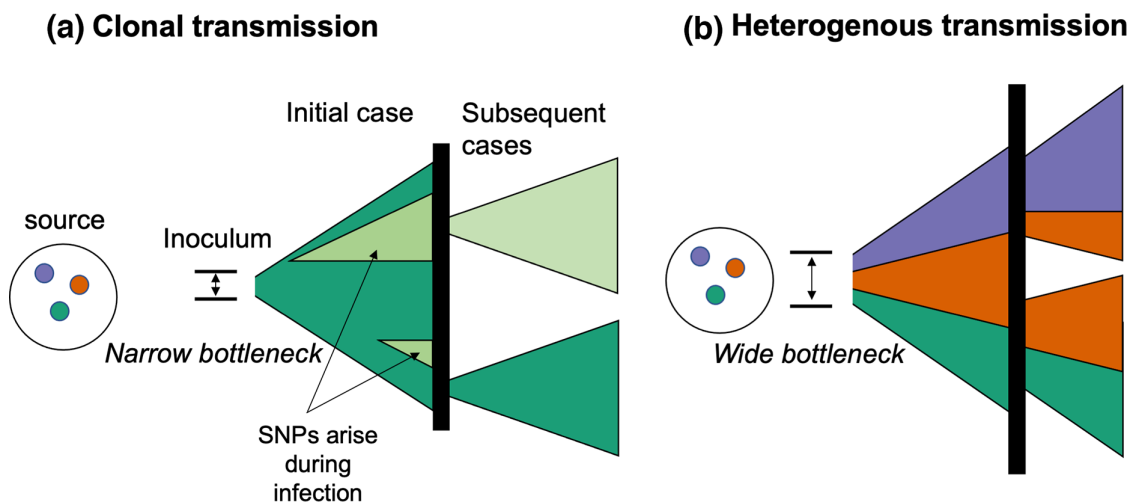


Fig. 7. Conceptual framework for the acquisition of within-host diversity of *B. anthracis*. Colours represent different genotypes of *B. anthracis* (i.e. with multiple nucleotide differences), while different shades (i.e. light green) represent single nucleotide variants arising during the course of infection. Vertical black lines represent the environment (e.g. soil, water) from which subsequent cases of anthrax arise through exposure to *B. anthracis* spores. (a) Example of infections stemming from small transmission bottleneck (inoculum of few spores), resulting in primarily clonal transmission. Where the bottleneck is narrow, a limited number of spores (and consequently genotypes) could comprise the inoculum, regardless of the number of variants present in the environment. A small number of SNPs may arise during the course of infection (see Figs 5b and 6), but are unlikely to be transmitted unless they arise early during the course of infection. These variants rarely differ from the founding genotype by >1–2 SNPs (Table S5). (b) Example of infections stemming from a wide bottleneck (infectious dose with multiple spores), wherein sufficient numbers of spores comprise the inoculum such that multiple genotypes present in the environmental source may seed initial and subsequent infections. In this example, the inoculum of the first case comprises three genotypes; this seeds two subsequent cases, with different combinations of these genotypes. Our results strongly support heterogenous transmission (b), either from single or multiple carcass sites, and a large transmission bottleneck. Figure adapted from Ågren *et al.* [29].

While livestock appear to be the primary drivers of *B. anthracis* transmission in the NCA, movement of infected wildlife and scavengers acting as carriers of *B. anthracis* spores could represent an additional mechanism contributing to our observations [57, 58]. Further comparative genomic studies across wider areas will be essential for elucidating the geographical scales at which transmission occurs. This would help to delineate areas across which coordinated livestock vaccination campaigns should occur to avoid regular re-incursions through animal movement and trade.

High within-host diversity is the result of simultaneous infection with multiple variants, not within-host evolution

We observed high *B. anthracis* diversity within individual hosts, essentially indistinguishable from levels of diversity found throughout the study area. Smaller numbers of SNPs (1–2) could potentially have arisen during culture; however, based on previous passaging experiments with *B. anthracis* [29], we do not expect this to have contributed in our case due to the limited number of passages performed. Alternatively, small numbers of SNPs could have arisen during the course of infection within the host. The simulations we performed suggest that isolates with >2 SNP differences between them are unlikely to be the product of within-host evolution. The regular occurrence of such differences or greater among isolates from the same carcass indicates that animals are commonly infected with a heterogenous infectious dose (i.e. ingestion of a mixture of *B. anthracis* genotypes from single or possibly multiple grazing or watering points; Fig. 7). Multiple SNP profiles were present among isolates from the same individual soil samples, supporting the occurrence of heterogeneity in a single infection source. In principle, post-mortem contamination from the environment prior to sampling could have contributed to the diversity of profiles detected. However, the probability of this occurring is extremely low, given that detectable levels of spores tend to be spatially restricted close to previous carcass site locations [55]. Therefore, we believe that our finding of high within-host diversity cannot be explained by environmental contamination. It is also noteworthy that all within-host SNP profiles observed in this study were shared with isolates from at least one other sampled carcass (i.e. none were unique); this strongly suggests that most of the observed within-host diversity is the result of multiple genotypes having been present in the inoculum rather than having been generated *de novo*. Taken together, this points to a wide transmission bottleneck, since a small inoculum comprising only a few spores would limit the possible diversity that could be transmitted.

Our findings make an important contribution to the ongoing debate about the size of the transmission bottleneck for naturally occurring anthrax in animals (i.e. the number of spores that give rise to a case). Recent work has proposed that founding populations may be as small as 1–3 individual spores [30]. However, our findings are clearly at odds with such narrow bottlenecks,

since the multiple genotypes we regularly observed within individual hosts could not have been transmitted within such a small inoculum. While our results provide no information about the exact size of the infectious dose, they align more closely with earlier suggestions of higher infectious doses, which are biologically plausible given that animals grazing at carcass sites might ingest hundreds of thousands of spores with each bite [53].

Spatio-temporally linked anthrax cases are rarely genetically linked

Limiting phylogenetic analyses to a single isolate per host often leads to incorrect inference of transmission events [22], including the potential to overlook important epidemiological connections [59]. This issue is exemplified in the current study by the spatio-temporally linked pair of cases in cluster 2: whereas several of the isolates from both carcasses ($n=3$ each) had identical SNP profiles, supporting a transmission link, both carcasses also harboured non-identical genotypes (Fig. 3). Under these circumstances, transmission links may be missed, even with complete sampling. As noted by Ågren *et al.* [29], in the case of multi-clonal infections, subsequent cases may stem from different genotypes from within the founding population, masking the fact that these cases stemmed from a common source, regardless of the number of isolates sequenced among the subsequent cases. This could be the case for the other three spatio-temporally linked pairs of anthrax carcasses investigated in this study (Fig. 3c), in which we only detected isolates with distinct SNP profiles. Alternatively, cases without a genetic link could be temporally linked for reasons other than exposure to a common source. For instance, animals might be more susceptible or at greater risk of exposure to infection at particular times of year, e.g. due to lower immune function related to their nutritional status and/or associated with weather extremes including prolonged rains or droughts [8]. More extensive sampling of within-carcass diversity would be necessary to investigate these hypotheses and to determine whether co-occurrence of genotypes could be used to track transmission patterns.

Conclusions

In this study, the genomic diversity of *B. anthracis* was quantified at various spatial scales within a hyperendemic setting. While WGS could discriminate among isolates within a relatively small geographical area, there was a lack of phylogeographical signal and limited genetic relatedness was observed among isolates from spatio-temporally linked cases. We hypothesize that this lack of spatial structure reflects the long-term persistence of *B. anthracis* spores in the environment, combined with extensive livestock movements related to local pastoralist practices. Based on simulations, the high within-host heterogeneity we observed points to an inoculum comprised of diverse genotypes, suggestive of a wide transmission bottleneck. Our work paves the way for studying *B. anthracis* genomic diversity and evolution within anthrax-endemic areas more broadly and to confirm the temporal and spatial scales over which genomic data are most informative for inferring transmission dynamics.

Funding information

T.L.F. was supported by a Marie Skłodowska-Curie Individual Fellowship (659223), a fellowship from the Natural Sciences and Engineering Research Council of Canada (PDF-471750-2015), the Wellcome Trust through a Springboard award to T.L. by the Academy of Medical Sciences (SBF002\1168), a Biotechnology and Biological Sciences Research Council (BBSRC) Discovery Fellowship (FORDE/BB/R012075/1), and a Lord Kelvin Adam Smith Leadership Fellow Fund from the University of Glasgow. O.R.A. was supported by a studentship from the Bill and Melinda Gates Foundation through the Program for Enhancing the Health and Productivity of Livestock (project reference ID 1083453). Establishment and maintenance of the Zoonoses Laboratory at the Kilimanjaro Clinical Research Institute (KCRI) (B.T.M.) was supported by the BBSRC (BB/J010367/1), BBSRC Zoonoses and Emerging Livestock Systems (BB/L017679, BB/L018926, BB/L018845), and the Wellcome Trust (096400/Z/11/Z). The funders had no role in study design, data collection and analysis, nor preparation of the manuscript.

Acknowledgements

We are grateful for all the support received for this research, particularly from the NCA community and authorities. We thank the Ngorongoro District Council, Ngorongoro Conservation Area Authority, District Veterinary Office, Tanzania Wildlife Research Institute and members of our field team – Sabore Ole Moko, Sironga Nanjicho, Kadogo Lerimba and Godwin Mshumba – for assistance with this study. We also thank the Directorate of Veterinary Services, Ministry of Agriculture, Livestock and Fisheries, and Ministry of Health, Community Development, Gender, Elderly and Children for their support. Finally, we thank Yi Xuan Chew and Nichith Kollanandi Ratheesh for initial assessments of the publicly available genomic data.

Author contributions

Conceptualization: T.L.F., T.L., R.B. Data curation: T.L.F., T.P.W.D. Formal analysis: T.L.F., T.P.W.D., W.T.H., M.M., A.V., R.B. Funding acquisition: T.L.F., B.T.M., T.L., R.B. Investigation: T.L.F., O.R.A., A.H., D.M., T.L. Methodology: T.L.F., T.P.W.D., W.T.H., M.M., R.B. Project administration: T.L.F., B.T.M., T.L. Resources: I.K., H.V.H. Software: T.P.W.D., W.T.H., M.M. Supervision: T.L., R.B. Visualization: T.L.F., T.P.W.D., W.T.H., M.M., A.V., R.B. Writing – original draft: T.L.F. Writing – review and editing: T.P.W.D., O.R.A., W.T.H., A.H., R.N.Z., T.L., R.B.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

This study received ethical approval from the Kilimanjaro Christian Medical University College ethics review committee (certificate no. 2050); the National Institute for Medical Research, Tanzania (NIMR/HQ/R.8a/vol. IX/2660); the Tanzanian Commission for Science and Technology (2016-95-NA-2016-45); and the College of Medical Veterinary and Life Sciences ethics committee at the University of Glasgow (200150152). It also received permission under section 20 of the Animal Diseases Act 35 (1984) at the University of Pretoria, South Africa (reference 12/11/1/1/6). The Bill and Melinda Gates Foundation and the National Institute for Medical Research, Tanzania, approved submission of this manuscript.

References

- Kao RR, Haydon DT, Lycett SJ, Murcia PR. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends Microbiol* 2014;22:282–291.
- Carlson CJ, Kracalik IT, Ross N, Alexander KA, Hugh-Jones ME, et al. The global distribution of *Bacillus anthracis* and associated anthrax risk to humans, livestock and wildlife. *Nat Microbiol* 2019;4:1337–1343.
- WHO. *Anthrax in Humans and Animals*. Geneva: World Health Organization; 2008.
- WHO. *The Control of Neglected Zoonotic Diseases. Community-Based Interventions for Prevention and Control. Report of the Third Conference Organized with ICONZ, DFID-RIU, Gates Foundation, SOS, EU, TDR and FAO with the Participation of ILRI and OIE*. Geneva: World Health Organization; 2011.
- Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, et al. Global genetic population structure of *Bacillus anthracis*. *PLoS One* 2007;2:e461.
- Sahl JW, Pearson T, Okinaka R, Schupp JM, Gillette JD, et al. A *Bacillus anthracis* genome sequence from the Sverdlovsk 1979 autopsy specimens. *mBio* 2016;7:e01501-16.
- Bruce SA, Schiraldi NJ, Kamath PL, Easterday WR, Turner WC. A classification framework for *Bacillus anthracis* defined by global genomic structure. *Evol Appl* 2020;13:935–944.
- Smith KL, DeVos V, Bryden H, Price LB, Hugh-Jones ME, et al. *Bacillus anthracis* diversity in Kruger National Park. *J Clin Microbiol* 2000;38:3780–3784.
- Simonson TS, Okinaka RT, Wang B, Easterday WR, Huynh L, et al. *Bacillus anthracis* in China and its relationship to worldwide lineages. *BMC Microbiol* 2009;9:71.
- Achtman M. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annu Rev Microbiol* 2008;62:53–70.
- Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, et al. Anthrax molecular epidemiology and forensics: using the appropriate marker for different evolutionary scales. *Infect Genet Evol* 2004;4:205–213.
- Fasanella A, Serrecchia L, Chiaverini A, Garofolo G, Muuka GM, et al. Use of Canonical Single Nucleotide Polymorphism (CanSNPs) to characterize *Bacillus anthracis* outbreak strains in Zambia between 1990 and 2014. *PeerJ* 2018;6:e5270.
- Beyer W, Bellan S, Eberle G, Ganz HH, Getz WM, et al. Distribution and molecular evolution of *Bacillus anthracis* genotypes in Namibia. *PLoS Negl Trop Dis* 2012;6:e1534.
- Maho A, Rossano A, Hächler H, Holzer A, Schelling E, et al. Antibiotic susceptibility and molecular diversity of *Bacillus anthracis* strains in Chad: detection of a new phylogenetic subgroup. *J Clin Microbiol* 2006;44:3422–3425.
- Girault G, Blouin Y, Vergnaud G, Derzelle S. High-throughput sequencing of *Bacillus anthracis* in France: investigating genome diversity and population structure using whole-genome SNP discovery. *BMC Genomics* 2014;15:288.
- Derzelle S, Girault G, Roest HIJ, Koene M. Molecular diversity of *Bacillus anthracis* in the Netherlands: investigating the relationship to the worldwide population using whole-genome SNP discovery. *Infect Genet Evol* 2015;32:370–376.
- Derzelle S, Girault G, Kokotovic B, Angen Ø. Whole genome-sequencing and phylogenetic analysis of a historical collection of *Bacillus anthracis* strains from Danish cattle. *PLoS One* 2015;10:e0134699.
- Lienemann T, Beyer W, Pelkola K, Rossow H, Rehn A, et al. Genotyping and phylogenetic placement of *Bacillus anthracis* isolates from Finland, a country with rare anthrax cases. *BMC Microbiol* 2018;18:102.
- Lekota KE, Hassim A, Madoroba E, Hefer CA, van Heerden H. Phylogenomic structure of *Bacillus anthracis* isolates in the Northern Cape Province, South Africa revealed novel single nucleotide polymorphisms. *Infect Genet Evol* 2020;80:104146.
- Carlson CJ, Getz WM, Kausrud KL, Cizauskas CA, Blackburn JK, et al. Spores and soil from six sides: interdisciplinarity and the environmental biology of anthrax (*Bacillus anthracis*). *Biol Rev Camb Philos Soc* 2018;93:1813–1831.
- Döpfer D, Buist W, Soyer Y, Munoz MA, Zadoks RN, et al. Assessing genetic heterogeneity within bacterial species isolated from gastrointestinal and environmental samples: how many isolates does it take? *Appl Environ Microbiol* 2008;74:3490–3496.
- Worby CJ, Lipsitch M, Hanage WP. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. *PLoS Comput Biol* 2014;10:e1003549.
- Futse JE, Brayton KA, Dark MJ, Knowles DP, Palmer GH. Superinfection as a driver of genomic diversification in antigenically variant pathogens. *Proc Natl Acad Sci USA* 2008;105:2123–2127.
- Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. *BMJ* 2015;350:h1314.
- Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. *Nat Rev Microbiol* 2016;14:150–162.
- Paterson GK, Harrison EM, Murray GGR, Welch JJ, Warland JH, et al. Capturing the cloud of diversity reveals complexity and heterogeneity of MRSA carriage, infection and transmission. *Nat Commun* 2015;6:6560.
- Alamil M, Hughes J, Berthier K, Desbiez C, Thébaud G, et al. Inferring epidemiological links from deep sequencing data: a statistical learning approach for human, animal and plant diseases. *Philos Trans R Soc Lond B Biol Sci* 2019;374:20180258.
- Beyer W, Turnbull PCB. Co-infection of an animal with more than one genotype can occur in anthrax. *Lett Appl Microbiol* 2013;57:380–384.
- Ågren J, Finn M, Bengtsson B, Segerman B. Microevolution during an anthrax outbreak leading to clonal heterogeneity and penicillin resistance. *PLoS One* 2014;9:e89112.
- Easterday WR, Ponciano JM, Gomez JP, Van Ert MN, Hadfield T, et al. Coalescence modeling of intrainfection *Bacillus anthracis* populations allows estimation of infection parameters in wild populations. *Proc Natl Acad Sci USA* 2020;117:4273–4280.
- Tanzania National Bureau of Statistics, Office of Chief Government Statistician. *Population and Housing Census*. Dodoma: Tanzania National Bureau of Statistics; 2012.
- Mwakapeje ER, Høgset S, Fyumagwa R, Nonga HE, Mdegela RH, et al. Anthrax outbreaks in the humans – livestock and wildlife interface areas of Northern Tanzania: a retrospective record review 2006–2016. *BMC Public Health* 2018;18:106.
- Lembo T, Hampson K, Auty H, Beesley CA, Bessell P, et al. Serologic surveillance of anthrax in the Serengeti ecosystem, Tanzania, 1996–2009. *Emerg Infect Dis* 2011;17:387–394.
- Aminu OR, Lembo T, Zadoks RN, Biek R, Lewis S, et al. Practical and effective diagnosis of animal anthrax in endemic low-resource settings. *PLoS Negl Trop Dis* 2020;14:e0008655.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
- Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 2014;15:R46.
- Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;2:e000086.
- Broad Institute. Picard toolkit; 2019. <http://broadinstitute.github.io/picard>
- García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, et al. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 2012;28:2678–2679.

41. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–576.
42. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
43. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
44. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013;30:1188–1195.
45. Hijmans RJ. Raster: Geographic Data Analysis and Modeling Version 3.0-12 from CRAN (<https://rdrr.io/cran/raster/>); 2020
46. Jombart T. *adegenet*: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 2008;24:1403–1405.
47. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15:524.
48. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
49. R Core Team. R: a Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2018. <https://www.R-project.org/>
50. Vogler AJ, Busch JD, Percy-Fine S, Tipton-Hunton C, Smith KL, et al. Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrob Agents Chemother* 2002;46:511–513.
51. Saile E, Koehler TM. *Bacillus anthracis* multiplication, persistence, and genetic exchange in the rhizosphere of grass plants. *Appl Environ Microbiol* 2006;72:3168–3174.
52. Braun P, Grass G, Aceti A, Serrecchia L, Affuso A, et al. Microevolution of Anthrax from a Young Ancestor (M.A.Y.A.) suggests a soil-borne life cycle of *Bacillus anthracis*. *PLoS One* 2015;10:e0135346.
53. Turner WC, Kausrud KL, Beyer W, Easterday WR, Barandongo ZR, et al. Lethal exposure: an integrated approach to pathogen transmission via environmental reservoirs. *Sci Rep* 2016;6:27311.
54. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. *Trends Ecol Evol* 2015;30:306–313.
55. Dragon DC, Bader DE, Mitchell J, Woollen N. Natural dissemination of *Bacillus anthracis* spores in northern Canada. *Appl Environ Microbiol* 2005;71:1610–1615.
56. Blackburn JK, Ganz HH, Ponciano JM, Turner WC, Ryan SJ, et al. Modeling R_0 for pathogens with environmental transmission: animal movements, pathogen populations, and local infectious zones. *Int J Environ Res Public Health* 2019;16:954.
57. De Vos V. Vultures as carriers of anthrax. *J S Afr Vet Assoc* 1979;50:35.
58. Stears K, Schmitt MH, Turner WC, McCauley DJ, Muse EA, et al. Hippopotamus movements structure the spatiotemporal dynamics of an active anthrax outbreak. *Ecosphere* 2021;12:e03540.
59. Forde TL, Orsel K, Zadoks RN, Biek R, Adams LG, et al. Bacterial genomics reveal the complex epidemiology of an emerging pathogen in arctic and boreal ungulates. *Front Microbiol* 2016;7:1759.
60. UNEP-WCMC. Protected Area Profile for Ngorongoro Conservation Area; 2020. www.protectedplanet.net
61. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2009.
62. Pebesma E. Simple features for R: standardized support for spatial vector data. *R Journal* 2018;10:439–446.
63. Yu G, Smith DK, Zhu H, Guan Y, Lam T-Y. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 2017;8:28–36.
64. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, et al. Toward automatic reconstruction of a highly resolved tree of life. *Science* 2006;311:1283–1287.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.