

UC Berkeley

UC Berkeley Previously Published Works

Title

Phylogeography of *Bacillus anthracis* in the Country of Georgia Shows Evidence of Population Structuring and Is Dissimilar to Other Regional Genotypes

Permalink

<https://escholarship.org/uc/item/6d27k31n>

Journal

PLOS ONE, 9(7)

ISSN

1932-6203

Authors

Khmaladze, Ekaterine
Birdsell, Dawn N
Naumann, Amber A
[et al.](#)

Publication Date

2014

DOI

10.1371/journal.pone.0102651

Peer reviewed



Phylogeography of *Bacillus anthracis* in the Country of Georgia Shows Evidence of Population Structuring and Is Dissimilar to Other Regional Genotypes

Ekaterine Khmaladze^{1,8*}, Dawn N. Birdsell^{2,9}, Amber A. Naumann², Christian B. Hochhalter², Meagan L. Seymour², Roxanne Nottingham², Stephen M. Beckstrom-Sternberg², James Beckstrom-Sternberg², Mikeljon P. Nikolich³, Gvantsa Chanturia¹, Ekaterine Zhgenti¹, Mariam Zakalashvili¹, Lile Malania¹, Giorgi Babuadze¹, Nikoloz Tsertsvadze¹, Natalia Abazashvili¹, Merab Kekelidze¹, Shota Tsanava^{1,8}, Paata Imnadze^{1,8}, Holly H. Ganz⁴, Wayne M. Getz⁴, Ofori Pearson⁵, Pawel Gajer⁶, Mark Eppinger^{6,7}, Jacques Ravel⁶, David M. Wagner², Richard T. Okinaka², James M. Schupp², Paul Keim², Talima Pearson^{2*}

1 National Center for Disease Control and Public Health, Tbilisi, Georgia, **2** Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona, United States of America, **3** Walter Reed Army Institute of Research, Silver Spring, Maryland, United States of America, **4** Department Environmental Science Policy and Management, University of California, Berkeley, California, United States of America, **5** US Geological Survey, Denver Federal Center, Denver, Colorado, United States of America, **6** Institute for Genome Sciences, Baltimore, Maryland, United States of America, **7** University of Texas at San Antonio, Texas, United States of America, **8** Ivane Javakishvili Tbilisi State University, Tbilisi, Georgia

Abstract

Sequence analyses and subtyping of *Bacillus anthracis* strains from Georgia reveal a single distinct lineage (Aust94) that is ecologically established. Phylogeographic analysis and comparisons to a global collection reveals a clade that is mostly restricted to Georgia. Within this clade, many groups are found around the country, however at least one subclade is only found in the eastern part. This pattern suggests that dispersal into and out of Georgia has been rare and despite historical dispersion within the country, for at least for one lineage, current spread is limited.

Citation: Khmaladze E, Birdsell DN, Naumann AA, Hochhalter CB, Seymour ML, et al. (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(7): e102651. doi:10.1371/journal.pone.0102651

Editor: Adam Driks, Loyola University Medical Center, United States of America

Received: December 30, 2013; **Accepted:** June 21, 2014; **Published:** July 21, 2014

Copyright: © 2014 Khmaladze et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the U.S. Department of Homeland Security S&T CB Division Bioforensics R&D Program (HSHQDC-10-C-00139) and by the Department of Defense's Defense Threat Reduction Agency (CBCALL12-DIAGB1-2-0194). Support for collecting the samples in Namibia was provided by National Institutes of Health GM083863 grant to Wayne M. Getz. The use of products/names does not constitute endorsement by the United States DHS. The findings and opinions expressed herein belong to the authors and do not necessarily reflect the official views of the Walter Reed Army Institute of Research, the U.S. Army or the Department of Defense. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* Email: Talima.Pearson@nau.edu

⁹ These authors contributed equally to this work.

Introduction

Bacillus anthracis, the causative agent of anthrax, continues to decimate livestock herds and cause concern over possible nefarious use. Source attribution of outbreaks is therefore essential for epidemiological and forensic investigations as well as control efforts. Accurate attribution requires determining evolutionary relationships among isolates and understanding how they fit into regional and global phylogeographic patterns. Genetic characterization of *B. anthracis* first gained considerable traction through multiple locus VNTR analysis (MLVA) [1] which identified major genetic groups. MLVA typing schemes have been expanded to include more loci; however, although these methods provide additional resolution, inferences on relationships among isolates are unreliable due to the prevalence of convergent alleles, negatively affecting the accuracy of downstream attribution conclusions. With the increased accessibility of whole genome sequencing, typing schemes based on single nucleotide polymor-

phisms (SNPs) provide both high resolution and highly accurate phylogenetic information [2–4] although phylogenetic discovery bias [2] must be taken into account.

We and others have previously used MLVA and SNP typing to characterize regional and global phylogeographic patterns of *B. anthracis*. Although ungulate grazers certainly play a role in the dissemination of this species, human-mediated dispersal has contributed to both ancient [5] and recent [6] spread. Contaminated animal products are frequently identified during outbreaks of disease in humans [6,7]; however, ecological establishment of exotic strains is rare [4,8]. None the less, distinguishing between endemic and non-indigenous strains is difficult and can obscure phylogeographic patterns and confuse attribution efforts. Intensive regional sampling studies are not only invaluable for identifying endemic strains and for defining regional patterns of dissemination and cycling, but also build the foundation for understanding global patterns of spread.

We report here the phylogeographic patterns of *B. anthracis* samples collected within the country of Georgia. Located in a geographic bottleneck between the Black and Caspian Seas, and between Europe and Asia, this region has been impacted by ancient and modern human movement and trade. Such anthropogenic influences may have shaped the distribution of *B. anthracis* in this region and may prove to be important in understanding global phylogeographic patterns.

Materials and Methods

Phylogenetic placement

To place the 272 Georgian isolates into the established global phylogeny [2,6–8], we screened all isolates with previously described canSNP assays [8]. All isolates were assigned to the Aust94 genetic group. To obtain more detailed resolution within this genetic group, we genotyped the Georgian isolates and 225 additional Aust94 group isolates from our global collection with previously published Aust94 assays [9](Table S1) as well as with novel assays (see CanSNP Selection and Analysis below). DNA templates were extracted using either chloroform [10], DNeasy blood and tissue kits (Qiagen, Valencia, CA) or heat soak.

Whole Genome Sequencing

To further resolve the genetic structure within Georgian Aust94 group, we sequenced the genomes of three Georgian isolates belonging to the Aust94 genetic group, using Illumina’s Genome Analyzer II (San Diego, CA). Library preparation for this isolate involved sonication of 5 µg genomic DNA, obtained through a standard chloroform extraction protocol [10] and shearing the DNA to an average fragment size of 350 bp. The library was quantified using SYBR-based qPCR and primers modified from the adaptor sequence. Paired-end read lengths were ~100 bp. The sequences of 52-G, 9080-G, and 8903-G were deposited into GenBank (PRJNA224563, PRJNA224558, and PRJNA224562, respectively) (Table S2).

SNP Discovery and Analysis

To identify putative SNPs, homologous genomic regions were identified using MUMmer [11] and aligned to search for SNPs using SolSNP (<http://sourceforge.net/projects/solsnp/>). We ensured site orthology by eliminating potential paralogs and requiring all genome alignments to include 100 bp flanking each side of the SNP. Furthermore, for analysis inclusion, all SNP loci were required to be present in all of the genomes analyzed. A maximum-parsimony tree was constructed by PAUP 4.0b10 software (Sinauer Associates, Inc., Sunderland, MA, USA) using all putative SNPs from this study and ten published genomes (Figure 1, panel A; Table S2).

CanSNP Selection and Analysis

WGS comparisons between Georgian (52-G) and Aust94 (Aust94, AAES00000000) strains revealed 50 putative SNPs specific to 52-G. Of these, twenty-six were incorporated into melt-MAMA genotyping assays, as previously described [9] and eight were selected as canonical SNP assays (Table 1). Allele-specific melt-MAMA primers were designed using Primer Express 3.0 software (Applied Biosystems, Foster City, CA). All other assay reagents and instrumentation were as previously described [9]. PCR reactions were first raised to 50°C for 2 min to activate the uracil glycolase, then raised to 95°C for 10 min to denature the DNA and then cycled at 95°C for 15 s and 55°C–60°C for 1 min for 33 cycles (Table 1). Immediately after the completion of the PCR cycle, amplicon melt dissociation was measured by ramping

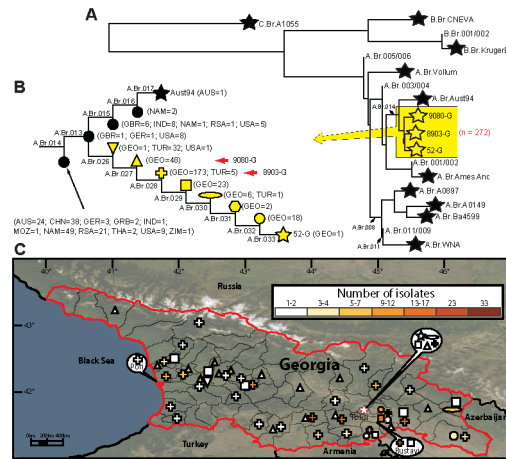


Figure 1. *Bacillus anthracis* phylogenetics in Georgia. A) Established phylogeny of *B. anthracis* [2,6–8]. Terminal subgroups representing sequenced strains are shown as stars, and intervening nodes representing collapsed branches appear as circles. The highlighted yellow box (part of the Aust94 lineage) indicate the phylogenetic location of Georgian strains. Stars within the highlighted yellow area represent the three Georgian strains sequenced for this study. The number of Georgian strains is indicated in red. B) Expansion of the Aust94 group and canSNP subgroups within the Georgian lineage. The number and origin of isolates are shown for each node and stars or red arrows indicate the locations of the sequenced strains. Shapes of nodes correspond to geographic location. C) Phylogeography of 272 *B. anthracis* isolates falling in the Aust94 group are mapped across the country of Georgia at a district level. The heat map legend indicates the number of isolates per subgroup found in a given district. doi:10.1371/journal.pone.0102651.g001

from 60°C to 95°C in 0.2°C/min increments and recording the fluorescent intensity.

Results and Discussion

Screening of 272 Georgian isolates with previously described canSNP assays [9] resulted in the assignment of all isolates to the group defined by the Australia94 (Aust94) genome [8] (Figure 1, panel A). The temporal, geographic and phylogenetic diversity of the Georgian Aust94 isolates, coupled with the detection of MLVA A3a [1] (roughly equivalent to the Aust94 lineage) isolates in the region by other researchers [12–15] is strong evidence for the ecological establishment of this clade (Table S3). Reference isolates from Turkey in the MLVA A3a group reported in Keim et al. [1] include six genotypes (genotypes 33, 36, 37, 41–43) and thus show regional diversity that is also reflected by the SNP genotyping here.

Members of the Aust94 group have been identified on five continents (Figure 1, panel B), suggesting extensive dispersal [8], however detailed phylogeography of any part of this lineage has not been previously described. The resulting whole genome SNP phylogenetic tree (Figure 1, panel A) (Table S2) drawn from 13 strains placed the three sequenced Georgian genomes within the Aust94 genetic group. Screening our global collection across previously published SNP assays (11) revealed five genetic groups along the Aust94 lineage (Figure 1, panel B) (Table S1 and S3) and eight novel groups along the lineage terminating in the 52-G genome.

The branches and topology leading to isolates that were genotyped, but not sequenced, remain unknown due to phylogenetic discovery bias and branch collapse, however clade membership is accurately estimated [2,4]. The node between branches A.Br.014 and A.Br.013 (named A.Br.014/013) forms the most

Table 1. Melt-MAMA primers targeting canonical SNPs for 8 new phylogenetic branches discovered in this study^a.

Branch	AmesAnc position ^b	Genome SNP state (D/A) ^c	Melt-MAMA primer sequence ^d	Conc, μM ^e	Annealing temp °C
A.Br.026	3,640,599	T/C	A:CTTCTTTTAATACATCTAAGTAAGTAAGCGTTgC D:cggggcggggcggggcggggcCTTCTTTTAATACA TCTAAGTAAGTAAGCGTTcTC:ATTGACCCAACAGCTACGAAATAC	0.45 0.15 0.15	60
A.Br.027	4,355,524	A/G	A:CCCATCCCAAGTGACACACTcG D:cggggcggggcggggcggggcCCCATTCCAAGTGACACACTgA C:AGCACTTGCTTATCTTGGAGCTT	0.60 0.15 0.15	60
A.Br.028	791,256	A/G	A:ACAGAGAAGGTTATAAGTCCAGAcGG D:cggggcggggcggggcggggcACAGAGAAGGTTATAAGTCCAGAAaGA C:CTCGCTTTTCCTGTTCTTTATTTCAC	0.15 0.15 0.15	60
A.Br.029	3,960,657	A/G	A:AGTATCCCAACCATTACTATAGTCACTaG D:ggggcggggcggggcggggcggggcAGTATTCCAAC CATTACTATAGTCACTcAC:GTACTATTGGTGTACTGCCAAATT	0.15 0.15 0.15	60
A.Br.030	3,528,668	A/G	A:CAATCCCTCGATTTACATATAAAATAAaG D:ggggcggggcggggcggggcggggcCAATCCCTCGATTTACA TATAAAATAAaAC:AGGTATGTATGAATTAGAAGGAAGAA	0.15 0.15 0.15	60
A.Br.031	3,018,054	C/T	A:ACTATCGCCAAAAGCAATTGaAT D:cggggcggggcggggcggggcACTATCGCCAAAAGCAATTGtAC C:TATTTTAGACAAGTACGAACTAGATAAATCAA	0.15 0.15 0.15	55
A.Br.032	3,520,170	G/A	A:CCACCAACAACGAATGGAAGtA D:cggggcggggcggggcggggcCCACCAACAACGAATGGAAGaG C:AGCATTAAATGAACGGCGTAAGTAATA	0.45 0.15 0.15	60
A.Br.033	3,610,151	C/T	A:TAAATAACCAAGGCGTCTTGCCAT D:ggggcggggcggggcggggcggggcCTAAATAACCAAGGCGTCTTGcTcAC C:TGTAGGACGTAGTATGGTGAAGTAGTAGAT	0.60 0.15 0.15	60

^aMelt-MAMA, melt-mismatch amplification mutation assay; SNP, single nucleotide polymorphism; con, concentration.

^bAmes Ancestor reference genome (NC_006570).

^cSNP states are presented according to the top strand in the Ames ancestor AE017334 D: Derived SNP state; A: Ancestral SNP state.

^dMelt-mismatch amplification mutation assay (MAMA), A: Ancestral; D: Derived; C: Common. Primer tails and antepenultimate or penultimate mismatch bases are in lower case.

^eFinal concentration of each primer in Melt-MAMA genotyping assays.

doi:10.1371/journal.pone.0102651.t001

basal subgroup and contains isolates collected from countries in five continents (Figure 1, panel B). The A.Br.013/015 subgroup contains isolates from Europe (n = 2) and USA (n = 8). It also gives rise to the lineages that contain all 272 Georgian isolates (Figure 1, panel B) as well as the lineage leading to the Aust94 genome. Despite the identification of isolates within this A.Br.013/015 node from Europe and the USA, this genetic group is not ecologically established in these regions and is thus not likely to be the source of the introduction into the Georgia/Turkey region. Rather, the presence of these isolates in Europe and the USA is likely to be due to the importation of contaminated animal products, possibly from the same geographic region responsible for the introduction into the Georgia/Turkey region. Additional phylogeography studies of these basal lineages are needed identify this source.

To further resolve the A.Br.013/015 group to understand the Georgian population structure, we screened all members of this group (including the 272 Georgian isolates) across the twenty-six SNP assays leading out to the 52-G strain, resulting in the identification of eight new groups within this lineage. All Georgian isolates fell within one of the eight new groups (Figure 1, panel B)(Table S3); however, some isolates from Turkey and one from the USA were also placed within these groups. The single USA isolate is probably from a contaminated animal product imported from the Turkey/Georgia region. Most Turkish isolates included in this study are assigned to the basal node along the 52-G lineage. However, as we have no knowledge of the phylogenetic topology within this basal node, it is impossible to determine if this group was first introduced into Turkey and subsequently dispersed into

Georgia or *vice versa*; both scenarios are equally parsimonious. The presence of Turkish isolates in two more recent nodes preceded by exclusively Georgian nodes suggests at least two dispersal events from Georgia to Turkey (Figure 1, panel B). Further sampling from neighboring countries will provide more details on the geographic limits of this 52-G clade and the impact of national boundaries on limiting the dispersal of *B. anthracis* in the area.

In any clade, members of the more ancient nodes have the greatest potential for widespread geographic dispersion. Indeed, along the lineage to the 52-G genome, members of the more basal nodes with multiple isolates from Georgia have been isolated across the country (Figure 1, panels B and C). However, without further sequencing of representative strains within each node, the phylogenetic topology cannot be known due to branch collapse [2]. Without phylogenetic knowledge, it is not possible to determine if phylogeographic clustering occurs at more recent evolutionary levels within these nodes. Conversely, isolates belonging to the clade after A.Br.030 which includes three nodes and the 52-G genome are only found in the southeast of Georgia (Figure 1, panel C), indicating more modern restrictions to dispersal. It is therefore likely that similar geographic structuring exists within the more basal nodes and would indicate that current dispersal of *B. anthracis* around Georgia is rare.

Conclusions

Our results are consistent with complex global dispersal patterns that have resulted in worldwide dispersal of the Aust94 group [6–

8]. This work now provides additional resolution and detail within the Aust94 group and shows it to be highly geographically structured with a group of closely related isolates being largely restricted to the region in and around Georgia and Turkey. Even in Georgia there is evidence of geographic structuring within the one lineage we characterized in detail, suggesting that that although anthrax has been dispersed throughout the country, current dispersal may be limited.

Supporting Information

Table S1 Melt-MAMAs published in Birdsell et al 2012 and used in this study.

(XLSX)

Table S2 Reference strains plus three new genomes used in this study and their NCBI accession numbers.

(XLSX)

References

- Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, et al. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 182: 2928–2936.
- Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, et al. (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.
- Pearson T, Hornstra HM, Sahl JW, Schaack S, Schupp JM, et al. (2013) When Outgroups Fail; Phylogenomics of Rooting the Emerging Pathogen, *Coxiella burnetii*. *Systematic Biology* 62: 752–762.
- Pearson T, Okinaka RT, Foster JT, Keim P (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. *Infection Genetics and Evolution* 9: 1010–1019.
- Kenefic LJ, Pearson T, Okinaka RT, Schupp JM, Wagner DM, et al. (2009) Pre-Columbian Origins for North American Anthrax. *Plos One* 4.
- Price EP, Seymour ML, Sarovich DS, Latham J, Wolken SR, et al. (2012) Molecular Epidemiologic Investigation of an Anthrax Outbreak among Heroin Users, Europe. *Emerging Infectious Diseases* 18: 1307–1313.
- Marston CK, Allen CA, Beaudry J, Price EP, Wolken SR, et al. (2011) Molecular Epidemiology of Anthrax Cases Associated with Recreational Use of Animal Hides and Yarn in the United States. *Plos One* 6.
- Van Ert MN, Easterday WR, Huynh LY, Okinaka RT, Hugh-Jones ME, et al. (2007) Global genetic population structure of *Bacillus anthracis*. *PLoS ONE* 2: e461.
- Birdsell DN, Pearson T, Price EP, Hornstra HM, Nera RD, et al. (2012) Melt Analysis of Mismatch Amplification Mutation Assays (Melt-MAMA): A Functional Study of a Cost-Effective SNP Genotyping Assay in Bacterial Models. *Plos One* 7.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*. 2nd ed: Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Kurtz S, Phillippy A, Delcher A, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biology* 5: R12.
- Durmaz R, Doganay M, Sahin M, Percin D, Karahocagil MK, et al. (2012) Molecular epidemiology of the *Bacillus anthracis* isolates collected throughout Turkey from 1983 to 2011. *Eur J Clin Microbiol Infect Dis* 31: 2783–2790.
- Eremenko EI, Ryazanova AG, Tsygankova OI, Tsygankova EA, Buravtseva NP, et al. (2012) Genotype diversity of *Bacillus anthracis* strains isolated from the Caucasus region. *Molecular Genetics Microbiology and Virology* 27: 74–78.
- Merabishvili M, Natidze M, Rigvava S, Brusetti L, Raddadi N, et al. (2006) Diversity of *Bacillus anthracis* strains in Georgia and of vaccine strains from the former Soviet Union. *Applied and Environmental Microbiology* 72: 5631–5636.
- Ortatatli M, Karagoz A, Percin D, Kenar L, Kilic S, et al. (2012) Antimicrobial susceptibility and molecular subtyping of 55 Turkish *Bacillus anthracis* strains using 25-loci multiple-locus VNTR analysis. *Comp Immunol Microbiol Infect Dis* 35: 355–361.

Table S3 Epidemiological and genotyping data for all *B. anthracis* strains included in this study.

(XLSX)

Acknowledgments

We thank Wendy C. Turner, Wolfgang Beyer, Bingxiang Wang and Alex Hoffmaster for sharing and shipment of strains included in this study. Collection of Namibia samples was authorized by the Namibian Ministry of Environment and Tourism under the auspices of Research/Collecting Permit 1448 issued to Holly H. Ganz.

Author Contributions

Conceived and designed the experiments: DNB MPN MK NT ST PI PK TP. Performed the experiments: EK AAN CBH MLS GC EZ MZ LM GB NA RN PG ME. Analyzed the data: EK DNB TP AAN CBH. Contributed reagents/materials/analysis tools: MPN MK NT ST PI HHG WMG OP PG ME JR DMW RTO JMS SMBS JBS PK TP. Wrote the paper: DNB TP.