Genomic analyses of ancient *Mycobacterium tuberculosis* complex strains from the Americas

Tanvi P. Honap[1], Åshild J. Vågene[2], Alexander Herbig[2], Jane E. Buikstra[3], Kirsten Bos[2], Johannes Krause[2], and Anne C. Stone[3]

[1] School of Life Sciences, Arizona State University, Tempe, USA, [2] Max Planck Institute for the Science of Human History, Jena, Germany, [3] School of Human Evolution and Social Change, Arizona State University, Tempe, USA

BACKGROUND

- Tuberculosis (TB) has affected humans for several millennia and in 2015, was the leading cause of death due to a single infectious agent[1]. It is caused by bacteria belonging to the *Mycobacterium tuberculosis* complex (MTBC).
- The MTBC evolved in Africa and spread to other parts of the world along with human migrations[2]. Animal-adapted MTBC species evolved from a human MTBC lineage, establishing that humans transmitted MTBC strains to other animals[2,3]. However, the source from which humans originally acquired the MTBC remains unknown.
- Earlier, Europeans were thought to have brought TB to the Americas during the Age of Exploration; all TB strains found in the Americas today are of European origin[4,5].
- Skeletal evidence for TB in the Americas dates back to ~160 CE South America[6,7]. Thus, it was hypothesized that TB was brought to the Americas along with human migrations out-of-Africa during the Pleistocene[8].
- Pre-Columbian TB strains were replaced by Lineage 4 *M. tuberculosis* strains after the arrival of Europeans[3]; however, it has not been determined how rapidly this replacement occurred.
- In 2014, our group published the genomes of human **MTBC strains recovered from three ~1000-year old individuals from coastal Peru which were most closely related to *M. pinnipedii*** strains which infect seals and sea mammals (Fig. 1) **[9].**
- Seals in the southern hemisphere are infected with *M. pinnipedii*[11]. Seals carried MTBC strains to the coast of South America, where the exploitation of seals for their meat and fur by coastal populations in Peru facilitated the transfer of these strains to humans.

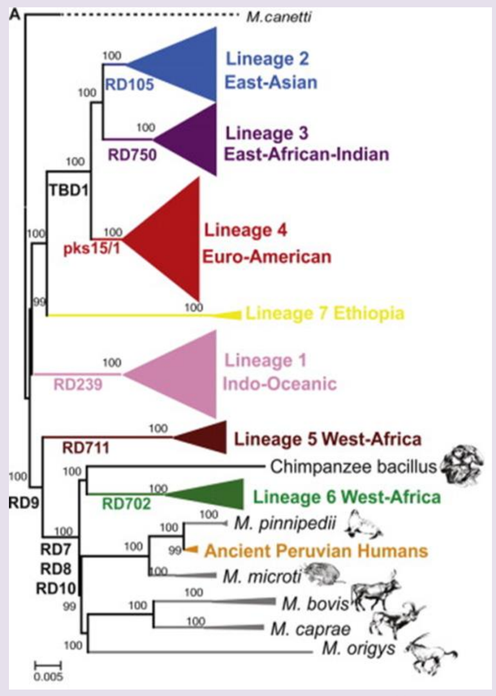

Fig. 1: Phylogeny of the MTBC[10]

OBJECTIVES

To recover genomes of ancient MTBC strains from non-coastal populations from the Americas so as to determine whether the pinniped-derived Peruvian MTBC strains adapted to humans and spread to other parts of the Americas via trade routes.

METHODS AND RESULTS

I. SAMPLING AND DNA EXTRACTIONS

- Samples (ribs, vertebrae, or teeth) collected from 68 individuals showing characteristic signs of skeletal TB (Fig. 2)
- DNA extracted using a silica column-based purification method[12]

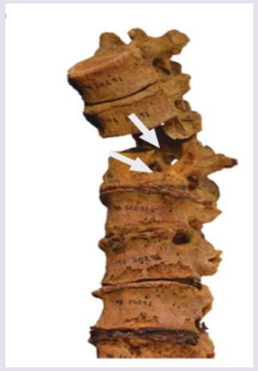

Fig. 2: Sample with skeletal TB[9]

II. DETECTING MTBC DNA USING QUANTITATIVE PCR ASSAYS

- Three quantitative PCR assays used to detect MTBC DNA[13]
- **17 samples tested positive for *rpoB* gene and IS6110 element, and 14 samples tested positive for IS1081 element**

III. LIBRARY PREPARATION AND ENRICHMENT CAPTURE

- Positive DNA extracts converted into DNA libraries
- Libraries target enriched the MTBC genes – *rpoB*, *gyrA*, *gyrB*, *katG*, and *mtp40*, using an in-solution hybridization capture protocol[9] and sequenced on Illumina MiSeq
- Reads processed using SeqPrep and mapped to *M. tuberculosis* H37Rv genome using bwa
- MapDamage[14] used to check for the presence of characteristic ancient DNA damage patterns (Fig. 3)

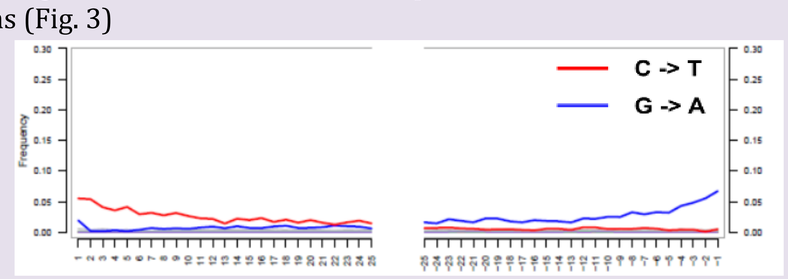

Fig. 3: DNA damage pattern for sample AD344

- **11 samples showed more than 50% coverage of all five targeted genes and were selected for whole-genome enrichment and sequencing of MTBC genomes**

IV. WHOLE-GENOME ENRICHMENT AND SEQUENCING

- 11 samples enriched for entire MTBC genome using synthetic baits and in-solution hybridization capture, followed by deep sequencing over two Illumina sequencing runs
- Reads trimmed and merged using SeqPrep and mapped to the MTBC ancestral reference genome using bwa
- SNPs called using GATK and the VCF2Genome program[14] used to construct a genome with the following parameters – at least Q30, minimum coverage of three reads, and minimum SNP allele frequency of 0.9
- Eight samples showed high amount of environmental mycobacterial DNA with mean coverage ranging from 2.4 – 9.5 X. Five of these samples are currently being resequenced to increase the mean coverage to 15X
- **Three samples from three different archaeological sites in Alaska (likely belonging to the post-contact era) showed high amounts of endogenous MTBC DNA (Table 1)**

V. DETERMINING LINEAGE SPECIFIC SNPS

- AD340 and AD351 strains contain SNP G2181026C specific to MTBC Lineage 4.2.1
- AD344 strain contains SNP A2789341C specific to MTBC Lineage 4.5

Sample	Archaeological site	Percentage of genome recovered	Mean coverage	MTBC Lineage
AD340	St. Michael, Alaska	96.2	19.2	4.2.1
AD344	Old Hamilton, Alaska	97.3	26.0	4.5
AD351	Ekwok, Alaska	97.6	26.6	4.2.1

Table 1: Sample information for the three samples from which ~20X coverage MTBC genomes were recovered

VI. PHYLOGENETIC ANALYSES

- Publically available whole-genome data for 64 MTBC strains downloaded from SRA and processed following pipeline given earlier
- A FASTA file containing multiple-genome alignment for all 67 strains and including *M. canettii* as an outgroup, used as input for MEGA6 (Fig. 4)

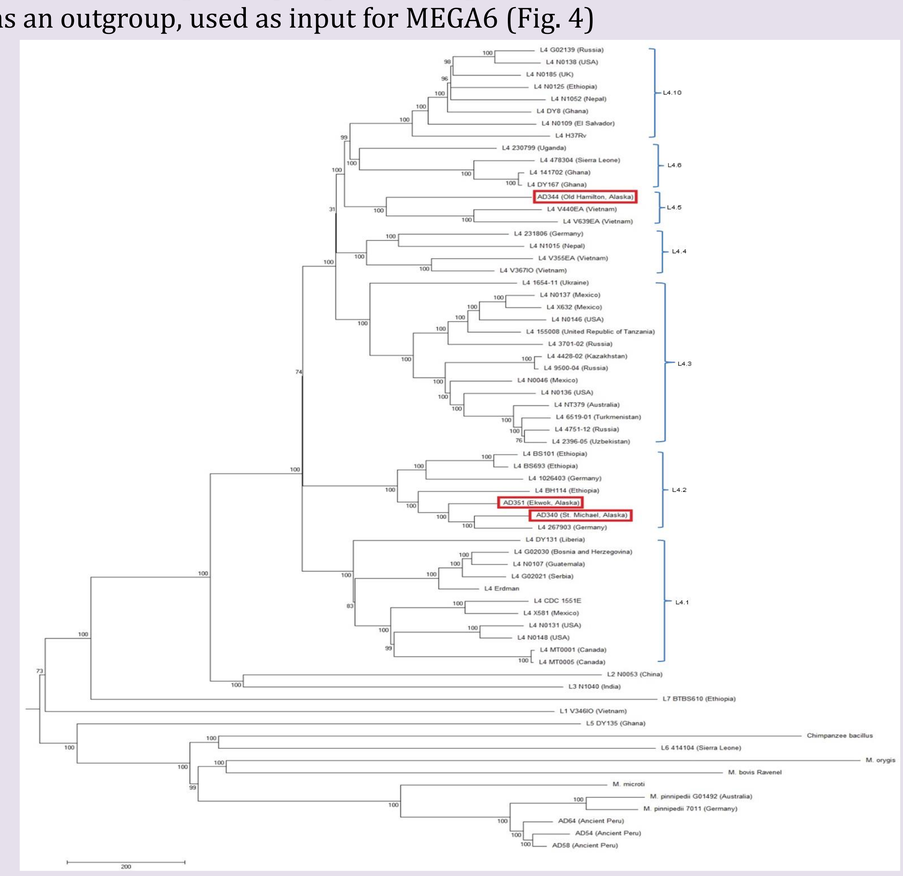

Fig. 4: A Maximum Parsimony tree generated using the Subtree-Pruning-Regrafting (SPR) method and 500 bootstrap replicates. All positions with less than 90% site coverage were eliminated. The final alignment length was 4,024,559 positions and comprised 32,408 variable nucleotide positions.

DISCUSSION

- Based on the archaeological context, the three Alaskan samples likely belong to the post-contact era. The consumption of a mostly marine diet, as is the case with these populations, leads to the Old Carbon effect which can affect the radiocarbon dates estimated from skeletal samples. This can result in a date which is centuries younger than the actual date[15,16]. Currently, we are in the process of recalibrating our radiocarbon dating data to account for the marine reservoir effect.
- European contact with Native Alaskans first began in 1741 with the Russians and was sustained thereafter[17]. TB deaths among Native Alaskans are documented as early as 1770 and by end of the 19th century, TB was a major health concern among these populations[18]. TB strains circulating in the Americas today belong mostly to *M. tuberculosis* Lineage 4.
- **AD344 (Old Hamilton) strain belongs to sub-lineage L4.5.** Strains from this sub-lineage are found in certain countries in Africa and Asia, but not commonly in the Americas[19].
- **AD340 (St. Michael) and AD351 (Ekwok) strains belong to sub-lineage L4.2.1 (Ural lineage).** Strains belonging to this sub-lineage have been found in many European countries, Russia, China, as well as in some African countries[19]. Our strains appear to be most closely related to *M. tuberculosis* strain 267903 from Germany, but further analyses incorporating whole-genome data from Russian-origin L4.2.1 strains might reveal interesting results.

Fig. 5: Locations of the three Alaskan sites

ONGOING WORK

- Sequencing of five whole MTBC genomes, which include three pre-contact era samples from non-coastal sites in the Americas

ACKNOWLEDGEMENTS

This work was supported by funding from the National Science Foundation, Wenner Gren Foundation, European Research Council, and Max Planck Society. T. Honap was funded by a travel grant from the Graduate and Professional Students Association, Arizona State University. References for this poster are available upon request.