Ancient *Yersinia pestis* genomes provide no evidence for the origins or spread of the Justinianic Plague

Marcel Keller[1,2]*, Maria A. Spyrou[1], Michael McCormick[3,4], Kirsten I. Bos[1], Alexander Herbig[1]*, Johannes Krause[1,4]*

[1]Department of Archaeogenetics, Max Planck Institute for the Science of Human History, Jena, Germany

[2]Institute of Genomics, University of Tartu, Tartu, Estonia

[3]Initiative for the Science of the Human Past, Department of History, Harvard University, Cambridge, USA

[4]Max Planck-Harvard Research Center for the Archaeoscience of the Ancient Mediterranean

*For correspondence: marcel.keller@ut.ee (MK); herbig@shh.mpg.de (AH); krause@shh.mpg.de (JK)

Competing interests: The authors declare that no competing interests exist.

Abstract

Along with the publication of 137 ancient human genomes retrieved from archaeological remains of the Eurasian steppe, Damgaard et al., 2018 identified two individuals infected with *Yersinia pestis*, yielding one genome with 0.24x average coverage (DA147, 6th–9th c. AD) and another with 8.7x (DA101, 2nd–3rd c. AD). A phylogenetic analysis performed on the latter placed it in a position ancestral to a 6th-century Justinianic genome from Aschheim, Germany. These results are used to fuel an argument that the Justinianic Plague (541–544 AD) "was brought to Europe towards the end of the Hunnic period through the Silk Road along the southern fringes of the steppes" in contrast to the leading hypothesis of introduction via the Red Sea that is supported by historical accounts. In our reanalysis, we question the contested historical context of the presented genomes with the Justinianic Plague and show that the lower coverage genome might be rather related to the Black Death (1346–1353 AD).

Introduction

The recent sequencing of dozens of pathogen genomes reconstructed from ancient DNA enabled increased-resolution phylogeographic studies on the spread of infectious diseases in prehistoric and historic times, especially in the context of human migration, mobility and trade (Andrades Valtueña et al., 2017; Bos et al., 2016; Keller et al., 2019; Namouchi et al., 2018; Rascovan et al., 2019; Rasmussen et al., 2015; Spyrou et al., 2019, 2016; Vågene et al., 2018). This is especially true for plague with its long and richly documented history and the abundance of published ancient genomes of its causative agent, *Yersinia pestis*. Interpretation of phylogenetic data in the context of human history requires careful assessment of tree topologies, branch lengths and mutation rates as well as thoughtful consilient approaches in integrating historical and archaeological data to prevent overly simplistic, deterministic or even erroneous interpretations.

For the Second Pandemic, the geographic origin and the possible persistence within Europe after the Black Death (1346–1352 AD) are the subject of ongoing scientific and scholarly discussion (Bos et al., 2016; Namouchi et al., 2018; Schmid et al., 2015; Spyrou et al., 2016; Spyrou et al., 2019). Whereas the first comprehensive phylogenetic study on *Y. pestis* favoured an East Asian origin (Cui et al., 2013), other scenarios assume an origin in Central Asia or the Caucasus (Benedictow, 2004; Namouchi et al., 2018; Sussman, 2011). Similarly, the origin of the Justinianic Plague (541–544 AD) has long been hypothesized to have originated in Africa (Achtman et al., 1999; Cui et al., 2008; Sarris, 2002). More recent studies however agree that the strains causing the First Pandemic (541–750 AD) likely emerged in Central Asia (Eroshenko et al., 2017; Harper, 2017; Wagner et al., 2014). The fact that the first outbreak of the Justinianic Plague is reported for Pelusium, Egypt nevertheless raises questions about the history and itinerary of the causative *Y. pestis* strain prior to this outbreak. The currently favoured scenario is an introduction via the Red Sea from India (Harper, 2017; Tsiamis et al., 2009) since there are no historical sources supporting a land route via the Levant or the Arabian peninsula (Schamiloglu, 2016). As such, discrepancies that arise from different analytical approaches raise questions about the history of *Y. pestis* prior to the first documented Justinianic Plague outbreak.

In a recent publication, Damgaard et al., 2018 presented two ancient *Y. pestis* genomes: one from the Tian Shan region (DA101, 2nd–3rd c. AD, 8.7-fold average coverage), branching ancestral to the First Pandemic lineage; and one from the Caucasus (DA147, 6th–9th c. AD, 0.24-fold average coverage) which was not further investigated. Damgaard et al.'s phylogenetic analysis places DA101 ancestral to the published genome from Aschheim. This

positioning is supported by a single SNP shared between the two genomes. An additional five SNPs are unique to DA101 compared to 95 in Aschheim, which is provisionally consistent with Aschheim's younger age (reported in the SI and in Extended Data Fig. 9, though the latter does not present the tree at full resolution). The identified shared ancestry was interpreted as setting DA101 within the context of the "Justinian plague" (*sic*). The longer branch in the Aschheim genome is explained by its younger age and a seemingly accelerated substitution rate, which is supposedly indicative of an epidemic context.

Although the Justinianic Plague was previously thought to represent the first major onslaught of plague in humans (i.e., the First Pandemic), plentiful examples of human infections of *Y. pestis* are surfacing as far back as the Neolithic (Andrades Valtueña et al., 2017; Rascovan et al., 2019; Spyrou et al., 2018). Here, we present a reanalysis of both DA101 and DA147 genomes which does not seem to support the arguments made by Damgaard et al., 2018. Instead, the analysis of DA101 suggests it to be yet another example of a pre-Justinianic human infection. Furthermore, in contrast to its suggested archaeological dating, we show DA147 to occupy a phylogenetic position much closer to the Black Death (1346–1353 AD) than the Justinianic Plague (Fig. 1C). As such, neither genome can address the origin of the First Pandemic.

Results and Discussion

We reanalysed both presented genomes with a more extensive dataset of published modern and ancient *Y. pestis* genomes (Fig. 1A, Table S1). We opted to include the genome from Altenerding (Feldman et al., 2016) as a representative for the Justinianic Plague: though genetically identical to Aschheim (Wagner et al., 2014), its higher coverage makes it less prone to false positive SNPs that are common in metagenomic data with high environmental backgrounds. Of note, the Aschheim genome has been shown to carry a high number of false positive SNPs (Feldman et al., 2016), which might in part account for its longer branch and accelerated substitution rate observed by Damgaard et al. (see SI).

Analysis of the DA101 genome revealed a minimum of 3 SNPs shared with Altenerding and a minimum of 9 that are unique. By contrast, Altenerding has 51 unique SNPs (Table S2). This sets both nodes, i.e., that giving rise to the shared Justinianic/DA101 branch and the one separating them, deeper in time compared to what is presented in the original publication (Damgaard et al., 2018).

Further to this, we attempted a molecular dating analysis, though the age of individual DA101 proved difficult to determine given discrepancies in the text and SI, ranging from

"approximately 180 AD" (main text) to 214–261 calAD/1701 BP (Damgaard et al., 2018 Supplementary Table 2). Ultimately, we opted to use the calibrated radiocarbon interval, which yielded a mean age of 154 BC (95% HPD: 527 BC to 153 AD) for the emergence of the shared lineage and 9 BC (95% HPD: 318 BC to 221 AD) for their divergence time (Table S4). For comparison, dating results without the recently published RT5 genome (Spyrou et al., 2018) are shown in Table S4. This strongly supports a pre-Justinianic provenience for the DA101 genome. A number of shared or unique SNPs might be undetected for DA101 due to low coverage, hence the estimated divergence dates are conservative and might be even older.

Regarding the substitution rate, we do not observe a notable acceleration on the Altenerding branch (mean 2.67E-08) compared to the overall mean (1.48E-08) across the tested dataset (Fig. S2), particularly since both estimates show overlapping 95 % HPD intervals (Table S5). Nevertheless, we do observe an overdispersion of substitution rates across different *Y. pestis* lineages (described previously in Cui et al., 2013 and Spyrou et al., 2019) with the highest estimate here yielding an 17-fold deviation from the mean (2.46E-07).

Damgaard et al. do not discuss the fact that DA101 predates the onset of the Justinianic Plague by three centuries according to its radiocarbon date. This fact, however, is incompatible with their hypothesis of a 6th-century pandemic disease introduction to Europe through Hunnic expansion based on this genome alone, as argued suggestively multiple times in their work: in the abstract ("Scythians […] moved westward in about the second or third century BC, forming the Hun traditions in the fourth–fifth century AD, and carrying with them plague that was basal to the Justinian plague."), the subheader ("Origins and spread of the Justinian plague", p. 372) and the concluding sentence ("[…], we find provisional support for the hypothesis that the pandemic was brought to Europe towards the end of the Hunnic period through the Silk Road along the southern fringes of the steppes.", p. 373).

Previously published data demonstrating the absence of detectable genetic changes in *Y. pestis* and its extremely rapid movement during the Black Death in Europe (1347–1353 AD; (Namouchi et al., 2018; Spyrou et al., 2019, 2016) clearly indicate that this pathogen is able to travel vast geographic expanses quickly, without accumulating genetic diversity in the process. As such, the depth of the time interval for the coalescence of DA101 and the Justinianic genomes offers little to no evidence on the temporal or geographic origin of the Justinianic Plague (beginning in 541 AD, Fig. 1B). Since individual DA101 comes from a geographical location that today houses multiple plague foci including modern lineages 0.ANT1, 0.ANT2 and the newly described 0.ANT5 (*sensu* Eroshenko et al., 2017; Fig. 2), it may even be

surmised whether the sampled individual fell victim to an epidemic event or a to a sporadic individual infection.

The second *Y. pestis* genome from individual DA147 from North Ossetia, supposedly 6th–9th centuries, could substantiate a spread of plague along the "southern fringes of the steppe" (p. 373), although its phylogenetic placement was not investigated by Damgaard et al. Even though the coverage is low, our re-analysis of the raw sequence data from this individual and an assessment of phylogenetically informative positions reveals that it does not share any derived SNPs with Altenerding or DA101 (Table S2). None of the positions shared between Altenerding and DA101 are covered in DA147, but 2 out of the 9 unique SNPs of DA101 are covered and show the ancestral state. Of the unique Altenerding SNPs, 9 are covered in DA147 with 8 showing the ancestral state. The only SNP possibly shared with DA147 is a C>T change that is potentially caused by DNA damage, as it appears only in a single read.

Such initial results motivated a further exploration of DA147's possible phylogenetic position. For this, we used MultiVCFAnalyzer v0.85 for a comparative SNP analysis against our dataset of ancient and modern *Y. pestis* genomes (Table S1), while omitting all private calls in DA147 since their vast majority will represent DNA damage and sequencing errors due to the genome's low coverage. The remaining SNPs forming the branch of DA147 in Fig. S1 (red) are an artefact caused by homoplastic or triallelic sites. We computed a maximum likelihood phylogenetic tree that, unexpectedly, placed DA147 closest to the previously described polytomy of Branches 1–4 (Fig. S1). The genomes's placement was further investigated by visual inspection of all diagnostic SNPs separating Branches 1, 2, 3&4 and Branch 0 (see Table S3). Our analysis reveals several potential placements for DA147: (1) it is one SNP ancestral to the polytomy but derived with respect to the 0.ANT3 node, (2) it is directly on the polytomy, (3) it is one SNP ancestral to the Black Death strain (Bos et al., 2011) on Branch 1, or (4) it is one to 16 SNPs basal on Branch 2 (Fig. 1C; Table S3). The third scenario is of particular interest in the context of a recently discovered genome from Laishevo, Russia (Spyrou et al., 2019) which could be identical to DA147. Therefore, DA147 might instead offer currently unexplored insights into the origin of the Black Death.

Furthermore, this finding raises doubts about the precision in the archaeological dating of this specimen (6th–9th centuries; Damgaard et al., 2018). Unfortunately, the provenience of this genome cannot be further investigated since metadata from this individual are absent in Table S2 in Damgaard et al., 2018. Based on our molecular dating analysis, the node giving rise to 0.ANT3, which is basal to all possible placements of DA147, is dated to a mean age of 1030

AD (95% HPD: 732 AD to 1274 AD), thus placing this low coverage genome within the diversity that has accumulated within the last millennium.

Finally, we would like to correct two inaccuracies in nomenclature in the study: First, the label "0.ANT5" has already been given to a modern clade of *Y. pestis* strains reported by Eroshenko et al., 2017. In general, we recommend against applying nomenclature combining phylogenetic and metabolic features to ancient genomes (Achtman, 2016), since their metabolic profile has not yet been characterized. Second, the "Justinianic Plague" is named after the Roman emperor Justinian I (c. 482–565 AD) who reigned during the onset of this pandemic (Little et al., 2007). The term "Justinian Plague" as used by the authors is misleading, since it suggests a connection to either Justin I or Justin II of the Justinianic dynasty.

Overall, we argue that the two presented *Y. pestis* genomes cannot contribute to our understanding of the Justinianic Plague that began in 541 AD in the southeast Mediterranean basin due to their phylogenetic, temporal and geographical distance. Moreover, these genomes offer no support for a connection between the Justinianic Plague and the Hunnic expansion, or for a spread through the southern steppe, both of which are also in conflict with the leading, document-based hypothesis of a plague introduction via trade routes linking India to the Red Sea (Harper, 2017; Fig. 2). The low coverage genome might rather hold clues for the onset of the Black Death or on the origins of Branch 2. We suggest a redirected focus here, especially if higher coverage data from this or a similar archaeological sample becomes available in the future.

Materials and Methods

Sequencing data for the samples DA101 and DA147 were retrieved from ENA with the provided accession numbers (Damgaard et al., 2018) and processed with the EAGER pipeline (Peltzer et al., 2016), including Illumina adapter removal, sequencing quality filtering (minimum base quality of 20) and length filtering (minimum length of 30 bp).

For the DA101 sample with higher coverage, reads were clipped on both ends by 3 bases to remove the majority of damaged sites and subsequently filtered again for length using the same parameter. Mapping against the CO92 reference genome (chromosome NC_003143.1) was done with BWA (-l 32, -n 0.1, -q 37), reads with low mapping quality (-q 37) were removed with Samtools and duplicates were removed with MarkDuplicates.

SNP calling was performed with the UnifiedGenotyper within the Genome Analysis Toolkit (GATK) using the 'EMIT_ALL_SITES' option to generate a call for every position in the reference genome.

For the DA147 sample with low coverage, mapping was performed without prior damage clipping and with less stringent parameters in BWA (-l 16, -n 0.01, -q 37) to retrieve a maximum of coverage. Reads with low mapping quality were removed with Samtools (-q 37) and duplicates were removed with MarkDuplicates. For a phylogenetic analysis of the low coverage DA147 genome (0.24-fold), the bam-file was converted into a fastq-file using bedtools, multiplied by 5 and mapped again with identical parameters but without duplicate removal to reach the necessary coverage of positions for SNP calling. SNP calling was performed with the UnifiedGenotyper within the Genome Analysis Toolkit using 'EMIT_ALL_SITES' to generate calls for all positions in the reference genome.

For the phylogenetic analyses, we used 166 previously published modern *Y. pestis* genomes (Cui et al., 2013; Eroshenko et al., 2017; Kislichkina et al., 2015; Zhgenti et al., 2015), a *Y. pseudotuberculosis* reference genome (IP32953; Chain et al., 2004) as an outgroup and the following ancient genomes: nine genomes from Neolithic/Bronze Age contexts (Andrades Valtueña et al., 2017; Rasmussen et al., 2015; Spyrou et al., 2018), one genome of the Justinianic Plague (Altenerding; Feldman et al., 2016), one genome representing Black Death (8291-11972-8124; (Bos et al., 2011), and six genomes of the subsequent second plague pandemic (Observance OBS116, OBS137, OBS110, OBS107, OBS124; Bos et al., 2016); Bolgar, (Spyrou et al., 2016)). A complete list of all *Y. pestis* genomes used is given in Table S1. Previously identified problematic regions as well as regions annotated as repeat regions, rRNAs, tRNAs and tmRNAs were excluded for all subsequent analyses (Cui et al., 2013; Morelli et al., 2010). MultiVCFAnalyzer v0.85 (Bos et al., 2014) was used for generating a SNP table with the following settings: Minimal coverage for base call of 5 with a minimum genotyping quality of 30 for homozygous positions, minimum support of 90% for calling the dominant nucleotide in a 'heterozygous' position. The sample DA147 was processed in the outgroup mode in MultiVCFAnalyzer to remove all singletons, for the most part representing damaged sites called due to the prior multiplication of reads. The unfiltered SNP alignment produced by MultiVCFAnalyzer was used for all following analyses. Additionally, all phylogenetically informative positions were visually inspected in IGV v2.4 (Thorvaldsdóttir et al., 2013). Maximum likelihood trees were generated with RAxML v8 (Stamatakis, 2014) using the GTR substitution model based on a partial deletion (95 %) SNP alignment for DA101 (3673 SNPs) and a full SNP alignment for DA147 (3885 SNPs). Robustness of all trees was tested by the bootstrap methods using 1000 pseudo-replicates.

For the estimation of divergence times and substitution rates with BEAST 1.10 (Drummond and Rambaut, 2007), we used the coalescent Bayesian skyline model with a setup identical to that published in Spyrou et al., 2018 with the following modifications: Integration of the DA101 sample (Damgaard et al., 2018), 95% partial deletion SNP alignment and 800,000,000 states as chain length. A second run was performed without the recently published Bronze Age genome RT5 (Spyrou et al., 2018) to investigate whether the results are affected by previously unavailable data. The recently published genomes by Eroshenko et al., 2017 were not included in the dating analysis due to exceptionally long branches and unavailability of raw data to address potential mismapping. An MCC tree was produced using TreeAnnotator of BEAST v1.10, showing the relative mean substitution rates (Fig. S2). All trees were visualized in FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

For Fig. 2, we used the coordinates of DA146 and DA160, since they are identical and frame the sample DA147 which is not part of the Table S2 of Damgaard et al., 2018. The first outbreaks of the Justinianic Plague come from Harper, 2017 and (Stathakopoulos, 2004). For the trade routes, we used the "Indian and Persian trade routes with the West 50 BCE - 300 CE" and "Silk Road routes 1–1400 CE" from OWTRAD (http://www.ciolek.com/owtrad.html). Important trade centres are adopted from Harper, 2017.

Author Contributions

M.K., A.H., K.I.B and J.K. planned and designed the study. M.K. performed data processing and phylogenetic analyses; M.A.S. and M.K. performed dating analyses. M.M. provided and reviewed historical context information. M.K. wrote the manuscript with contributions from M.A.S., K.I.B. and A.H. and edits from all co-authors.

253 **References**

254 Achtman M. 2016. How old are bacterial pathogens? *Proc Biol Sci* **283**:e1003471–e1003471.
255 Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A, Carniel E. 1999. Yersinia pestis, the
256 cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. *Proc Natl*
257 *Acad Sci U S A* **96**:14043–14048.
258 Andrades Valtueña A, Mittnik A, Key FM, Haak W, Allmäe R, Belinskij A, Daubaras M,
259 Feldman M, Jankauskas R, Janković I, Massy K, Novak M, Pfrengle S, Reinhold S,
260 Šlaus M, Spyrou MA, Szécsényi-Nagy A, Tõrv M, Hansen S, Bos KI, Stockhammer
261 PW, Herbig A, Krause J. 2017. The Stone Age Plague and Its Persistence in Eurasia.
262 *Curr Biol* **27**:3683–3691.e8.
263 Benedictow OJ. 2004. The Black Death 1346–1353: The Complete History. Woodsbridge:
264 Boydell & Brewer.
265 Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM,
266 Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA,
267 Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson
268 D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J. 2014.
269 Pre-Columbian mycobacterial genomes reveal seals as a source of New World human
270 tuberculosis. *Nature* **514**:494–497.
271 Bos KI, Herbig A, Sahl J, Waglechner N, Fourment M, Forrest SA, Klunk J, Schuenemann
272 VJ, Poinar D, Kuch M, Golding GB, Dutour O, Keim P, Wagner DM, Holmes EC,
273 Krause J, Poinar HN. 2016. Eighteenth century Yersinia pestis genomes reveal the long-
274 term persistence of an historical plague focus. *Elife* **5**:e12994.
275 Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, McPhee
276 JB, DeWitte SN, Meyer M, Schmedes S, Wood J, Earn DJD, Herring DA, Bauer P,
277 Poinar HN, Krause J. 2011. A draft genome of Yersinia pestis from victims of the Black
278 Death. *Nature* **478**:506–510.
279 Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM,
280 Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M,
281 Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise
282 A, Hauser LJ, Garcia E. 2004. Insights into the evolution of Yersinia pestis through
283 whole-genome comparison with Yersinia pseudotuberculosis. *Proceedings of the*
284 *National Academy of Sciences* **101**:13826–13831.
285 Cui Y, Li Y, Gorgé O, Platonov ME, Yan Y, Guo Z, Pourcel C, Dentovskaya SV,

Balakhonov SV, Wang X, Song Y, Anisimov AP, Vergnaud G, Yang R. 2008. Insight into microevolution of Yersinia pestis by clustered regularly interspaced short palindromic repeats. *PLoS One* **3**:e2652.

Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, Weinert L a., Wang Z, Guo Z, Xu L, Zhang Y, Zheng H, Qin N, Xiao X, Wu M, Wang X, Zhou D, Qi Z, Du Z, Wu H, Yang X, Cao H, Wang H, Wang J, Yao S, Rakin A, Li Y, Falush D, Balloux F, Achtman M, Song Y, Wang J, Yang R. 2013. Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis. *Proc Natl Acad Sci U S A* **110**:577–582.

Damgaard P de B, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E, Baimukhanov N, Loman V, Hedeager L, Pedersen AG, Nielsen K, Afanasiev G, Akmatov K, Aldashev A, Alpaslan A, Baimbetov G, Bazaliiskii VI, Beisenov A, Boldbaatar B, Boldgiv B, Dorzhu C, Ellingvag S, Erdenebaatar D, Dajani R, Dmitriev E, Evdokimov V, Frei KM, Gromov A, Goryachev A, Hakonarson H, Hegay T, Khachatryan Z, Khaskhanov R, Kitov E, Kolbina A, Kubatbek T, Kukushkin A, Kukushkin I, Lau N, Margaryan A, Merkyte I, Mertz IV, Mertz VK, Mijiddorj E, Moiyesev V, Mukhtarova G, Nurmukhanbetov B, Orozbekova Z, Panyushkina I, Pieta K, Smrčka V, Shevnina I, Logvin A, Sjögren K-G, Štolcová T, Taravella AM, Tashbaeva K, Tkachev A, Tulegenov T, Voyakin D, Yepiskoposyan L, Undrakhbold S, Varfolomeev V, Weber A, Wilson Sayres MA, Kradin N, Allentoft ME, Orlando L, Nielsen R, Sikora M, Heyer E, Kristiansen K, Willerslev E. 2018. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**:369–374.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* **8**. doi:10.1186/1471-2148-7-214

Eroshenko GA, Nosov NY, Krasnov YM, Oglodin YG, Kukleva LM, Guseva NP, Kuznetsov AA, Abdikarimov T, Dzhaparova AK, Kutyrev VV. 2017. Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high- mountain plague foci of Kyrgyzstan. *PLoS One* **12**:e0187230–e0187230.

Feldman M, Harbeck M, Keller M, Spyrou MA, Rott A, Trautmann B, Scholz HC, Päffgen B, Peters J, McCormick M, Bos K, Herbig A, Krause J. 2016. A High-Coverage Yersinia pestis Genome from a Sixth-Century Justinianic Plague Victim. *Mol Biol Evol* **33**:2911–2923.

Harper K. 2017. The Fate of Rome: Climate, Disease, and the End of an Empire. Princeton: Princeton University Press.

Keller M, Spyrou MA, Scheib CL, Neumann GU, Kröpelin A, Haas-Gebhard B, Päffgen B, Haberstroh J, Ribera I Lacomba A, Raynaud C, Cessford C, Durand R, Stadler P, Nägele K, Bates JS, Trautmann B, Inskip SA, Peters J, Robb JE, Kivisild T, Castex D, McCormick M, Bos KI, Harbeck M, Herbig A, Krause J. 2019. Ancient Yersinia pestis genomes from across Western Europe reveal early diversification during the First Pandemic (541–750). *Proc Natl Acad Sci U S A* **116**:12363–12372.

Kislichkina AA, Bogun AG, Kadnikova LA, Maiskaya NV, Platonov ME, Anisimov NV, Galkina EV, Dentovskaya SV, Anisimov AP. 2015. Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus, Including Representatives of Biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica: TABLE 1. *Genome Announc* **3**:e01342–15.

Little LK, Hays JN, Morony MG, Kennedy HN, Stathakopoulos D, Sarris P, Stoclet AJ, Kulikowski M, Maddicott J, Dooley A, Sallares R, McCormick M. 2007. Plague and the End of Antiquity: The Pandemic of 541–750. Cambridge University Press.

Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, Cui Y, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JM, Balloux F, Keim P, Wirth T, Ravel J, Yang R, Carniel E, Achtman M. 2010. Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* **42**:1140–1143.

Namouchi A, Guellil M, Kersten O, Hänsch S, Ottoni C, Schmid BV, Pacciani E, Quaglia L, Vermunt M, Bauer EL, Derrick M, Jensen AØ, Kacki S, Cohn SK, Stenseth NC, Bramanti B. 2018. Integrative approach using *Yersinia pestis* genomes to revisit the historical landscape of plague during the Medieval Period. *Proceedings of the National Academy of Sciences* **115**:E11790-E11797.

Peltzer A, Jäger G, Herbig A, Seitz A, Kniep C, Krause J, Nieselt K. 2016. EAGER: efficient ancient genome reconstruction. *Genome Biol* **17**:60–60.

Rascovan N, Sjögren K-G, Kristiansen K, Nielsen R, Willerslev E, Desnues C, Rasmussen S. 2019. Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline Article Emergence and Spread of Basal Lineages of Yersinia pestis during the Neolithic Decline. *Cell* **176**:295–305.e10.

Rasmussen S, Allentoft ME, Nielsen K, Orlando L, Sikora M, Sjögren K-G, Pedersen AG, Schubert M, Van Dam A, Kapel CMO, Nielsen HB, Brunak S, Avetisyan P, Epimakhov A, Khalyapin MV, Gnuni A, Kriiska A, Lasak I, Metspalu M, Moiseyev V, Gromov A, Pokutta D, Saag L, Varul L, Yepiskoposyan L, Sicheritz-Pontén T, Foley RA, Lahr MM,

Nielsen R, Kristiansen K, Willerslev E. 2015. Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. *Cell* **163**:571–582.

Sarris P. 2002. The Justinianic plague: origins and effects. *Contin Chang* **17**:169–182.

Schamiloglu U. 2016. The Plague in the Time of Justinian and Central Eurasian History: An agenda for Research In: Zimonyi I, Karatay O, editors. Wiesbaden: Harrasowitz Verlag.

Schmid BV, Büntgen U, Easterday WR, Ginzler C, Walløe L, Bramanti B, Stenseth NC. 2015. Climate-driven introduction of the Black Death and successive plague reintroductions into Europe. *Proc Natl Acad Sci U S A* **112**:3020–3025.

Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, Neumann GU, Walker D, Alterauge A, Carty N, Cessford C, Fetz H, Gourvennec M, Hartle R, Henderson M, von Heyking K, Inskip SA, Kacki S, Key FM, Knox EL, Later C, Maheshwari-Aplin P, Peters J, Robb JE, Schreiber J, Kivisild T, Castex D, Lösch S, Harbeck M, Herbig A, Bos KI, Krause J. 2019. Phylogeography of the second plague pandemic revealed through analysis of historical Yersinia pestis genomes. *Nat Commun* **10**:4470.

Spyrou MA, Tukhbatova RI, Feldman M, Drath J, Kacki S, Beltrán de Heredia J, Arnold S, Sitdikov AG, Castex D, Wahl J, Gazimzyanov IR, Nurgaliev DK, Herbig A, Bos KI, Krause J. 2016. Historical Y. pestis Genomes Reveal the European Black Death as the Source of Ancient and Modern Plague Pandemics. *Cell Host Microbe* **19**:874–881.

Spyrou MA, Tukhbatova RI, Wang C-C, Andrades Valtueña A, Lankapalli AK, Kondrashin VV, Tsybin VA, Khokhlov A, Kühnert D, Herbig A, Bos KI, Krause J. 2018. Analysis of 3800-year-old Yersinia pestis genomes suggests Bronze Age origin for bubonic plague. *Nat Commun* **9**:2234.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.

Stathakopoulos D. 2004. Famine and Pestilence in the Late Roman and Early Byzantine Empire: A Systematic Survey of Subsistence Crises and Epidemics. Aldershot: Ashgate.

Sussman GD. 2011. Was the Black Death in India and China? *Bull Hist Med* **85**:319–355.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178–192.

Tsiamis C, Poulakou-Rebelakou E, Petridou E. 2009. The Red Sea and the Port of Clysma. A Possible Gate of Justinian's Plague. *Gesnerus* **66**:209–217.

Vågene ÅJ, Herbig A, Campana MG, Robles García NM, Warinner C, Sabin S, Spyrou MA,

388 Andrades Valtueña A, Huson D, Tuross N, Bos KI, Krause J. 2018. Salmonella enterica
389 genomes from victims of a major sixteenth-century epidemic in Mexico. *Nature Ecology*
390 *and Evolution* **2**:520–528.
391 Wagner DM, Klunk J, Harbeck M, Devault A, Waglechner N, Sahl JW, Enk J, Birdsell DN,
392 Kuch M, Lumibao C, Poinar D, Pearson T, Fourment M, Golding B, Riehm JM, Earn
393 DJD, Dewitte S, Rouillard J-M, Grupe G, Wiechmann I, Bliska JB, Keim PS, Scholz
394 HC, Holmes EC, Poinar H. 2014. Yersinia pestis and the plague of Justinian 541–543
395 AD: a genomic analysis. *Lancet Infect Dis* **14**:319–326.
396 Zhgenti E, Johnson SL, Davenport KW, Chanturia G, Daligault HE, Chain PS, Nikolich MP.
397 2015. Genome Assemblies for 11 Yersinia pestis Strains Isolated in the Caucasus
398 Region. *Genome Announc* **3**:1030–1015.

Fig. 1: A. Maximum likelihood tree based on 3673 SNPs of 167 modern and 18 ancient genomes. Main branches are collapsed for clarity. Numbers on nodes indicate bootstrap support. Highlighted are the Justinianic genome from Altenerding (green), the investigated Tian Shan genome DA101 (blue) and the recently characterized modern strains of clade 0.ANT5 (brown). Relevant parts of the tree are highlighted in pink (Fig. 1B) and light blue (Fig. 1C). **B.** Schematic tree of the common branch of Altenerding and DA101 showing the minimum number of SNPs (italics) and divergence dates (bold, mean and 95 % HPD). **C.** Schematic tree of the highlighted part of A (light blue) for the positioning of DA147 with numbers of SNPs (italics, # of diverged SNPs in DA147/# of covered SNPs in DA147/# of total SNPs on branch) and estimated divergence date of clade 0.ANT3 (bold, mean and 95 % HPD) based on the dating analysis with BEAST 1.10 (see Materials and Methods and SI). Possible placements of DA147 in the phylogeny are highlighted in red.

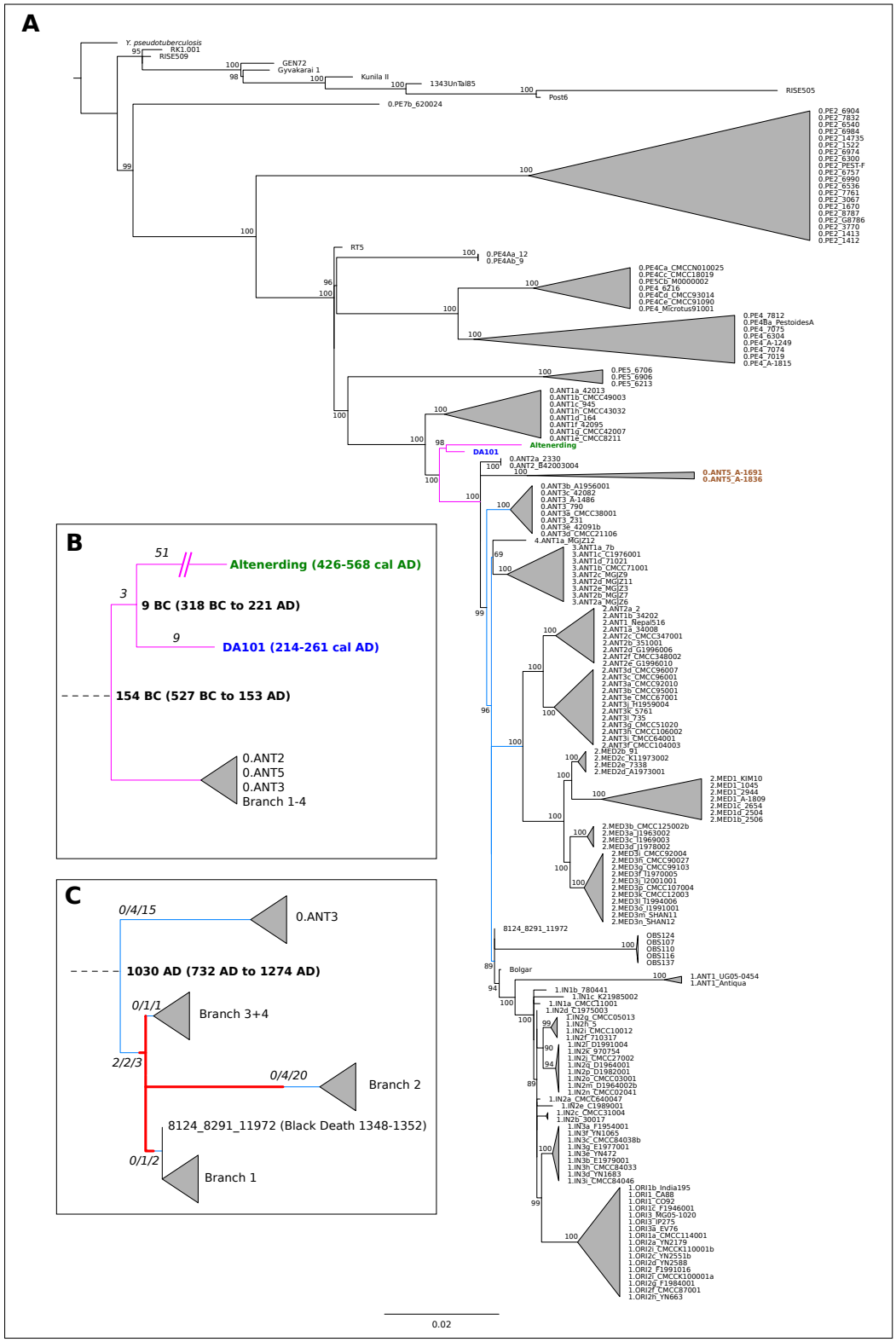

Fig. 2: Map of the origin of selected modern and ancient *Y. pestis* genomes in relation to the first outbreaks of the Justinianic Plague as well as important trade centres and routes. An introduction of plague via the Indian Ocean and the Red Sea is supported by well-established sea communications connecting Egyptian and Indian trade centres. The hypothesis of an overland transport through the steppe is in conflict with the directionality of the first outbreaks of the Justinianic Plague and lacks historical support. Although DA147 presumably dates later than 750 AD, it is shown for comparison.