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

Simon Rasmussen[1*], Martin Sikora[2*], Peter De Barros Damgaard[2*], Rasmus Nielsen[2,3], Eske Willerslev[2,4,5,6]

[1]*Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Denmark.*
[2]*Lundbeck Foundation Center for GeoGenetics, University of Copenhagen, Denmark.*
[3]*Department of Integrative Biology, UC Berkeley, USA.*
[4]*Lundbeck Foundation Center for GeoGenetics, University of Cambridge, UK.*
[5]*Wellcome Trust Sanger Institute, UK.*
[6]*The Danish Institute for Advanced Study, University of Southern Denmark, Denmark.*
**Joint first authors*
For correspondence: ewillerslev@bio.ku.dk (EW); simon.rasmussen@cpr.ku.dk (SR)

Abstract

A manuscript recently posted by Keller et al. proposes that the finding of two individuals with evidence of plague infection from our study of "137 ancient human genomes from across the Eurasian steppes" does not inform about the Justinianic Plague. We are puzzled by their conclusion, as no where in our paper we claim that the *Y. pestis* strains we discovered were direct ancestors of the Justinianic Plague outbreak. Their criticism appears to be based on simple mis-reading, and do not contradict our findings. Rather, we emphasize that the results in another manuscript from the same group confirm and strengthen our finding of a most recent common ancestor between DA101 and the Justinianic plague strains.

General comments

In our study Damgaard et al. "137 ancient human genomes from across the Eurasian steppes" we used ancient genomics to investigate the population history of the Eurasian steppe over 4,000 years of prehistory. As part of this effort we identified two individuals with evidence for infection with ancient *Yersinia pestis*, one of which (DA101) yielded sufficient genomic coverage to situate it basal to an ancient strain recovered from a Justinianic plague victim from Aschheim (Germany)[1].

In the recently posted manuscript on bioRxiv "*Ancient Yersinia pestis genomes provide no evidence for the origins or spread of the Justinianic Plague*", Keller et al performed a re-analysis of the genomes reported in Damgaard et al. The main conclusion from their analyses is that neither of those genomes can "contribute to our understanding of the Justinianic Plague that began in 541 AD in the southeast Mediterranean". We are surprised by this conclusion, as we never made the claim in Damgaard et al., 2018 (ref. 2) that the DA101 *Y. pestis* strain was a direct ancestor of the Justinianic Plague outbreak. This is obvious as the age of the sample predates the Justinianic epidemic by three centuries, and therefore any such claim would be superfluous, as also noted by

Keller et al. Rather, our analyses identify DA101 as diverging basal to the Justinianic plague strain from Aschheim, and we report it as such in the paper:

"The genome of the Y. pestis strain DA101, which we name 0.ANT5, branches off from the main plague lineage just basal to the Justinian plague strain 0.ANT4, identified from an individual in Aschheim (Germany) and dated to about AD 530".

We further provide a simple speculation that has very little boldness to it, namely that the eventual ancestor of the Justinianic outbreak was introduced to Europe from Central Asia into Europe through human mobility, for which we document ample evidence through the analysis of the human genomes in our study. This is not a contradiction to the scenario of the Justinianic Plague epidemic itself being first introduced into Europe via the southeast Mediterranean basin. In particular, we report:

"Given that the most basal strains of present-day plague (0.PE7 clade) originate in Qinghai and the clade basal to the Justinian plague (0.ANT1) is from Xinjiang (China), two areas close to the Tian Shan mountains, 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."

Keller et al also provide a novel analysis of a second plague strain (DA147) we reported in our study. Due to the uncertainties associated with its low coverage (0.24X) we made no attempt to reconstruct its phylogenetic position, and therefore did not further discuss its association with the Justinianic Plague. The new results by Keller et al suggest a phylogenetic position close to the polytomy for branches 1-4. We welcome these new results and agree that an association with the Justinianic plague clade (including DA101) appears unlikely. Future studies on a higher coverage genome would be needed to further resolve the uncertainty in the inferred phylogenetic position.

The criticism in the commentary of Keller et al. therefore appears to be based on simple mis-interpretation, and do not contradict our findings. Rather, we emphasize that the results by Keller al. confirm and strengthen our finding of a most recent common ancestor between DA101 and the Justinianic plague strains. Finally, another manuscript by Keller et al. (ref. 3) reports additional novel genomes of the first plague pandemic, again confirming the basal position of DA101 and concordantly referring to our interpretation as an ancestral divergence event within Central Asia.

Specific point by point responses

Lines 82 - 89
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).

Reply:
We called the variation in the Aschheim genome using our own pipelines that take the false positive SNPs into account and used those for the analysis. However, we do observe an increased number of SNPs in Aschheim and acknowledge that a certain amount of the accelerated substitution rate might be caused by false SNPs. We also note that this has no effect on the overall conclusions of the manuscript other than the analysis of increased mutation rate and relation of that to hypotheses of Cui et al. on increased replication cycles associated with pandemics. The inclusion of the Alternerding strain does not change the position of the DA101 as basal to the Justinianic Plague strains.

Lines 98 – 103
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 153 AD) for their divergence time. 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.

Reply:
We appreciate the BEAST analysis; however, this is not different from our own results with an emergence of the shared lineage of 5 AD (95% HPD: 180 BC to 132 AD) and 49 AD (95% HPD: 127 BC to 161 AD) for the divergence times. We did not include this in the manuscript and do not see an important relevance of this for our interpretation as we do not suggest it to be part of the Justinianic Plague.

Lines 111 – 120
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).

Reply:
We did not claim introduction to Europe through Hunnic expansions, but rather state "*that the pandemic was brought to Europe towards the end of the Hunnic period through the Silk Road".* The end of the Hunnic period dates to the end of the 5th century AD, in close temporal proximity to the

Justinianic plague pandemic. Further, our paper itself disintegrated the notion of Huns as single ethnic group. Finally, in the manuscript we only outlined that we find a strain basal to the ones recovered from victims of the Justinian plague outbreak itself in the Tian Shan - a region that was connected to the Mediterranean through trade in the late Hunnic period. Thus, we do not see an introduction caused by human mobility as a very controversial claim.

Lines 121 – 126
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 accumulating little to no 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).

Reply:
As described above, we were not discussing the actual outbreak of the Justinianic plague (541 – 543 AD), but rather the ancestor of the strains. As the DA101 strain was found in Tian Shan and the closest modern lineages that are basal or derived (0.ANT1, 0.ANT2, 0.ANT3 and 0.ANT5 (sensu Eroshenko[4])) have a Central or East Asian origin (Chinese or Kyrgyzstan) we simply stated that it added to the evidence of a Central or East Asian origin of the strains eventually causing the Justinianic plague.

Lines 127 – 131
Since individual DA101 comes from a geographical location that today features a variety of plague foci with modern lineages 0.ANT1, 0.ANT2 and the newly described 0.ANT5 (sensu Eroshenko et al. 2017, Fig. 2), it may even be queried whether the sampled individual fell victim to an epidemic event or a to sporadic transmission.

Reply:
We did not propose that the Justinianic Plague was on-going already at the time when the DA101 individual perished and it is not impossible that the genome recovered from DA101 could be a sporadic zoonotic transmission. This does not in any way contradict that the origin of the strains which later cause the Justinianic Plague was of Central Asian origin. In fact it would support a zoonotic reservoir for the eventual ancestor of the Justinianic outbreak in that region.

Lines 132 – 164
The second *Y. pestis* genome from individual DA147 from North Ossetia, supposedly 6th to 9th century, could substantiate a spread of plague along the “southern fringes of the steppe”, although its phylogenetic position 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 shared 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.

Reply:
We only mentioned the DA147 strain twice in the main text where we state the facts of the sample and does not propose anything of the phylogenetic relationship or age of the sample other than the archaeological context: "*We find that two individuals, DA101 and DA147 (see Supplementary Information section 7), show detectable levels of Yersinia pestis DNA, compatible with the characterization of the full genome sequence at 8.7× and 0.24× coverage. The first individual (DA101) is a Hun from the Tian Shan mountains and dates to approximately AD 180, and the second individual (DA147) is from the Alan culture from North Ossetia and is estimated archaeologically to date to the sixth– ninth century AD.*"

Lines 200-205
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.

Reply:
Keller et al. employed a non-standard approach for genotyping the low coverage strain DA147, which involves multiplying the same read data five times. The UnifiedGenotyper does not require a minimum coverage for SNP calling, and artificially multiplying data violates its underlying assumptions that each read corresponds to an independent observation of the underlying sequence. As the probability of a specific base being covered more than once in a genome with an average coverage of 0.24X is very small, the observed read sequence is the only possible genotype at the vast majority of genomic positions.

Lines 165 – 169
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.

Reply:
We followed the naming scheme given by Achtman 2016[5] where the Aschheim strain was named 0.ANT4 and following this named DA101 as 0.ANT5. We apologize for naming it in conflict with the Eroshenko et al.[4] study. We acknowledge that the nomenclature is based on both phylogenetic and metabolic features that cannot be characterized in ancient strains. However, given the phylogenetic position between 0.ANT1 (basal), 0.ANT2 (derived) and 0.ANT3 (derived) it is very unlikely to be of another population.

Lines 169 – 172
Second, the "Justinianic Plague" is named after the Roman emperor Justinian I (c. 482-565 AD) who reigned during the onset of this pandemic. 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 Justinian dynasty.

Reply:
We are aware that the term "Justinianic Plague" is associated with the particular plague outbreak. We used it in the manuscript because the DA101 strain was found to be of the same lineage, but basal to the two strains isolated from remains of the Justinianic plague. We do not at any given time describe our strain as anything else than "basal to the Justinian" precisely for this reason.

Lines 173 – 182
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 phylogenetically, temporally and geographically distant positions. 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 that are 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.

Reply:
We want to emphasize that we did not associate the DA101 strain with the Justinianic plague (541-543 AD) itself. The strain is consequently named "basal to the Justinian" for this exact reason. Furthermore, as discussed above we do provide a speculative but plausible suggestion: "*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.*"
This would imply that the strain basal to the one that was later introduced to Europe through the Mediterranean originated in East Asia and was spread into Europe through increased trade.

We find this to be a very plausible suggestion that does not conflict with Keller et al.'s documented view of a connection to India. We therefore conclude that Keller et al.'s retortion of our findings is based on a misunderstanding – perhaps caused by the conciseness of this particular sentence, and our purposefully chosen limited space given to speculation.

Our interpretation is that the DA101 strain is of the same lineage and basal to the Justinianic plague strains: "*The genome of the Y. pestis strain DA101, which we name 0.ANT5, branches off from the main plague lineage just basal to the Justinian plague strain 0.ANT4, identified from an individual in Aschheim (Germany) and dated to about AD 530 (Extended Data Fig. 9).*" Second, the phylogenetic most similar clades (0.ANT1, 0.ANT2, 0.ANT3 and 0.ANT5 (sensu Eroshenko)) all have a Central or East Asian origin (Chinese or Kyrgyzstan). Therefore, a strain that predates and is basal to the Justinianic plague from the Tian Shan region is in our understanding compliant with an origin of the Justinianic plague from Central or East Asia. We write: "*Given that the most basal strains of present-day plague (0.PE7 clade) originate in Qinghai and the clade basal to the Justinian plague (0.ANT1) is from Xinjiang (China), two areas close to the Tian Shan mountains, 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.*"

Additionally, Eroshenko et al. identify an origin in the Tian Shan mountains as well for the 0.ANT branch of strains as "Diversity of *Y. pestis* strains and extensive distribution of the strains which belong to the 0.ANT branch testify to the antiquity of the mentioned above plague foci and suggest that strains of the 0.ANT branch, which serve as precursors for all highly virulent *Y. pestis* strains, had their origin in the Tien Shan mountains." (from Eroshenko et al., 2017 ref. 4).

With more space to elaborate on speculation we could have re-phrased the last part of the sentence as: ", we find provisional support for the hypothesis that the strain that eventually evolved into the strains causing the Justinianic Plague had a Central Asian origin and was found at the Silk Road on the southern fringes of the steppe."

Finally, the additional strains presented in Keller et al.,[3] further confirm our analyses. They confirm that the DA101 strain is basal compared to all other strains from the first pandemic (541 – 750)

and that all other outgroup sequences to that lineage are from China or Kyrgyzstan. As written above, this does in fact provide provisional support that the Justinian plague evolved from Chinese strains as hypothesized in Damgaard et al.

References

1. Wagner, D. M. *et al.* Yersinia pestis and the Plague of Justinian 541 – 543 AD : a genomic analysis. *Lancet Infect. Dis.* **14**, 319–326 (2014).
2. De Barros Damgaard, P. *et al.* 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, (2018).
3. Keller, M. *et al.* 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 (2019).
4. Eroshenko, G. A. *et al.* Yersinia pestis strains of ancient phylogenetic branch 0.ANT are widely spread in the high-mountain plague foci of Kyrgyzstan. *PLoS One* **12**, e0187230 (2017).
5. Achtman, M. How old are bacterial pathogens? *Proc. R. Soc. B Biol. Sci.* **283**, 20160990 (2016).