1 The Stone Age Plague: 1000 years of

2 Persistence in Eurasia

3 Aida Andrades Valtueña[1], Alissa Mittnik[1,2], Felix M. Key[1], Wolfgang Haak[1,3], Raili Allmäe[4],

4 Andrej Belinskij[5], Mantas Daubaras[6], Michal Feldman[1,2] Rimantas Jankauskas[7], Ivor Janković[8,9],

5 Ken Massy[10,11], Mario Novak[8], Saskia Pfrengle[2], Sabine Reinhold[12], Mario Šlaus[13], Maria A.

6 Spyrou[1,2], Anna Szecsenyi-Nagy[14], Mari Tõrv[15], Svend Hansen[12], Kirsten I. Bos[1,2], Philipp W.

7 Stockhammer[1,10], Alexander Herbig[1,2]* and Johannes Krause[1,2]*

8

9 [1]Max Planck Institute for the Science of Human History, Jena, Germany

10 [2]Institute for Archaeological Sciences, Archaeo- and Palaeogenetics, University of Tübingen, Tübingen, Germany

11 [3]School of Biological Sciences, The University of Adelaide, Adelaide SA-5005, South Australia, Australia

12 [4]Archaeological Research Collection, Tallinn University, Tallinn, Estonia

13 [5] "Nasledie" Cultural Heritage Unit, Stavropol, Russia

14 [6]Department of Archaeology, Lithuanian Institute of History, Vilnius

15 [7]Department of Anatomy, Histology and Anthropology, Vilnius University, Vilnius, Lithuania

16 [8]Institute for anthropological research, Zagreb, Croatia

17 [9]Department of Anthropology, University of Wyoming, Laramie, USA

18 [10]Institute for Pre- and Protohistoric Archaeology and Archaeology of the Roman Provinces, Ludwig-Maximilians-University

19 Munich, Munich, Germany

20 [11]Heidelberg Academy of Sciences, Heidelberg, Germany

21 [12]Eurasia Department, German Archaeological Institute, Berlin, Germany

22 [13]Anthropological Center, Croatian Academy of Sciences and Arts, Zagreb, Croatia

23 [14]Institute of Archaeology, Research Centre for the Humanities, Hungarian Academy of Sciences, Budapest

24 H-1097, Hungary

25 [15]Independent researcher, Estonia

26

27 * To whom correspondence should be addressed

Abstract

28

29 Molecular signatures of *Yersinia pestis* were recently identified in prehistoric Eurasian

30 individuals, thus suggesting *Y. pestis* caused some form of disease in humans prior to the first

31 historically documented pandemic. Here, we present six new *Y. pestis* genomes spanning from

32 the European Late Neolithic to the Bronze Age (LNBA) dating from 4,800 to 3,700 BP. We show

33 that all currently investigated LNBA strains form a single genetic clade in the *Y. pestis*

34 phylogeny that appears to be extinct. Interpreting our data within the context of recent ancient

35 human genomic evidence, which suggests an increase in human mobility during the LNBA, we

36 propose a possible scenario for the spread of *Y. pestis* during the LNBA: *Y. pestis* may have

37 entered Europe from Central Eurasia during an expansion of steppe people, persisted within

38 Europe until the mid Bronze Age, and moved back towards Central Eurasia in parallel with

39 subsequent human population movements.

Introduction

40

41 Plague pandemics throughout human history caused unprecedented levels of mortality that

42 contributed to profound socioeconomic and political changes. Conventionally it is assumed that

43 plague affected human populations in three pandemic waves. The first, the Plague of Justinian,

44 starting in the 6th century AD, was followed by multiple epidemic outbreaks in Europe and the

45 Mediterranean basin, and has been associated with the weakening and decay of the Byzantine

46 empire (Russell, 1968). The second plague pandemic first struck in the 14th century with the

47 infamous 'Black Death' (1347-1352), which again spread from Asia to Europe seemingly along

48 both land and maritime routes (Zietz and Dunkelberg, 2004). It is estimated that this initial

49 onslaught killed 50% of the European population (Benedictow, 2004). It was followed by

50 outbreaks of varying intensity that lasted until the late eighteenth century (Cohn JR, 2008). The

51   most recent plague pandemic started in the 19th century and began in the Yunnan province of

52   China. It reached Hong Kong by 1894 and followed global trade routes to achieve a near

53   worldwide distribution (Stenseth et al., 2008). Since then plague has persisted in rodent

54   populations in many areas of the world and continues to cause both isolated human cases and

55   local epidemics (http://www.who.int/mediacentre/factsheets/fs267/en/).

56   Plague is caused by a systemic infection with the Gram-negative bacterium *Yersinia*

57   *pestis.* Advances in ancient DNA (aDNA) research have permitted the successful reconstruction

58   of a series of *Y. pestis* genomes from victims of both the first and second plague pandemic, thus

59   confirming a *Y. pestis* involvement and providing new perspectives on how this bacterium

60   historically spread through Europe (Bos et al., 2016, 2011; Feldman et al., 2016; Spyrou et al.,

61   2016; Wagner et al., 2014). Most recently, a study by Spyrou et al. (2016) suggested that during

62   the second pandemic, a European focus was established from where subsequent outbreaks,

63   such as the Ellwangen outbreak (16th century Germany) or the Great Plague of Marseille (1720-

64   1722 France), were derived (Spyrou et al., 2016). The authors also proposed that a descendant

65   of the Black Death strain travelled eastwards in the late 14th century, became established in

66   East Asia and subsequently gave rise to the most recent plague pandemic that spread the

67   pathogen around the globe.

68   Our perception of the evolutionary history of *Y. pestis* was changed substantially by a

69   recent report of two reconstructed genomes from Bronze Age individuals found in the Altai

70   region (Southern Siberia, dating to ~4,729 cal BP and ~3,635 cal BP, respectively) and

71   molecular *Y. pestis* signatures in an additional five individuals from Eurasia (~4,500 to 2,800

72   BP) suggesting the presence of plague in human populations over a diffuse geographic range

73   prior to the first historically recorded pandemics. Phylogenetic analysis of the two reconstructed

74   *Y. pestis* genomes from the Altai region shows that they occupy a phylogenetic position

75   ancestral to all extant *Y. pestis* strains, though this branch was not adequately resolved

76   (bootstrap lower than 95% (Rasmussen et al., 2015)). Further open questions remain regarding

77   *Y. pestis'* early association with humans. It is not currently known whether the *Y. pestis* lineages

78   circulating in Europe during the Late Neolithic and Bronze Age were all descended from the

79   ~5,000 BP Central Eurasian strain or whether there were multiple strains circulating in Europe

80   and Asia. Furthermore, how did plague spread over such a vast territory during the period

81   comprising the Late Neolithic and Bronze Age? Could these bacterial strains have been

82   associated with certain human groups and their respective subsistence strategies and cultures?

83   The Late Neolithic and Early Bronze Age in Western Eurasia (ca. 4,900-3,700/3,600 BP;

84   cf. Stockhammer et al., 2015) was a time of major transformative cultural and social changes

85   that led to cross-European networks of contact and exchange (Vandkilde, 2016). Intriguingly,

86   recent studies on ancient human genomes suggested a major expansion of people from the

87   Eurasian Steppe westwards into Central Europe as well as eastwards into Central Eurasia and

88   Southern Siberia starting around 4,800 BP (Allentoft et al., 2015; Haak et al., 2015). These

89   steppe people with a predominantly pastoral economy carried a genetic component that is

90   present in all Europeans today but was absent in early and middle Neolithic farmers in Europe

91   prior to their arrival. The highest amount of genetic 'steppe ancestry' in ancient Europeans was

92   found in individuals associated with the Late Neolithic Corded Ware Complex (Figure 1) around

93   4,500 BP (Haak et al., 2015), who show a genetic makeup close to the steppe people

94   associated with the 'Yamnaya' complex, suggesting a strong genetic link between those two

95   groups. Furthermore, it could be shown that the 'middle Neolithic farmer' genetic component

96   also appears in individuals associated with the Andronovo culture in the Altai region around

97   4,200 BP (Allentoft et al., 2015). These genetic links between humans ranging from Western

98   Eurasia to Southern Siberia highlight the dimensions of mobility and connectedness at the time

99   of the Bronze Age.

100   The reasons for the magnitude of the genetic turnover that occurred in Central Europe

101   around 4800 BP, where around 75% of the local middle Neolithic farmer genetics was replaced

102   (Haak et al., 2015), have yet to be explained. As in other episodes of human history, infectious

103   diseases may have played a significant role in triggering or catalyzing those major cultural shifts

104   and human migrations. Here we present six novel *Y. pestis* genomes from Central Europe and

105   the North Caucasus steppe spanning from the Late Neolithic to the Bronze Age (LNBA).

106   Through comparative analyses with other ancient and modern *Y. pestis* lineages (Bos et al.,

107   2016, 2011; Cui et al., 2013; Feldman et al., 2016; Kislichkina et al., 2015; Spyrou et al., 2016;

108   Zhgenti et al., 2015), we show that all LNBA strains form a single clade in the *Y. pestis*

109   phylogeny. This indicates a common origin of all currently identified *Y. pestis* strains circulating

110   in Eurasia during the Late Neolithic and Bronze Age, and reveals a distribution pattern that

111   parallels human movements in time and space.

Results

112

Screening

113

114   A total of 563 tooth and bone samples dating from the Late Neolithic to the Bronze Age from

115   Russia (122), Hungary and Croatia (139), Lithuania (27), Estonia (45), Latvia (10), and

116   Germany (Althausen 4, Augsburg 83, Mittelelbe-Saale 133) were screened for *Y. pestis* by

117   mapping DNA sequencing reads ranging in numbers from 700,000 to 21,000,000 against a

118   multi-fasta reference consisting of the genomes of 12 different *Yersinia* species (Table 1).

119   To assess if an individual was positive for *Y. pestis*, we calculated a score based on the

120   number of specific reads mapping to *Y. pestis* in comparison to the number of reads mapping to

121   other *Yersinia* species (See methods). Following this metric, all individuals with a positive score

122   were identified as possible candidates. Individuals that had a score higher than 0.005, and had

123   reads mapping to all the three plasmids present in *Y. pestis* were considered 'strong' positives.

124   In our dataset we identified five strong candidates, all of them tooth samples, from three

125   different locations spanning from the Late Neolithic to the Early Bronze Age: one individual from

126   the site Rasshevatskiy (RK1001) in the North Caucasus (Russia), one individual from the

127   Lithuanian site Gyvakarai (Gyvakarai1), one individual from the Estonian site Kunila (KunilaII)

128   and two individuals from Augsburg, Germany (Haunstetten, Unterer Talweg 85 Feature 1343

129   (1343UnTal85), and Haunstetten, Postillionstraße Feature 6 (6Post). Additionally, one individual

130   from the Croatian site Beli Manastir - Popova Zemlja (GEN72, also a tooth sample), which did

131   not pass the criteria for a strong candidate, was taken along as a potential candidate since it

132   had the highest number of reads mapping to the *Y. pestis* chromosome and all the plasmids

133   (chromosome=993, pCD1=243, pMT1=111, pPCP1=22). For a detailed description of all

134   samples, individuals and archaeological sites see Table 2 and SI.

135

Genome reconstruction

137   The five strong positive individuals identified during the screening step (Gyvakarai1, KunilaII,

138   1343Untal85, 6Post, RK1001) were shotgun sequenced to a depth of 379,155,741 to

139   1,529,935,532 reads. In addition to the shotgun sequencing, RK1001 and GEN72 were

140   enriched for *Y. pestis* DNA following an in-solution approach (See Methods). After mapping to

141   the reference genome (*Y. pestis* CO92, NC_003143.1), we reconstructed genomes from all six

142   potential candidates with a mean coverage from 3.7 to 12-fold with 86-94% of the reference

143   genome covered at least 1-fold (Table 2). The reads were independently mapped to the three

144   plasmids of *Y. pestis* CO92, and we reconstructed the three plasmids for our ancient samples

145   with mean coverage of: pCD1 7 to 24-fold, pMT1 3 to 14-fold and pPCP1 18 to 43-fold

146   (Supplementary Table 1).

147   In order to authenticate the ancient origin of the bacterial genomes, we evaluated the

148   damage patterns of terminal deamination common to ancient DNA (Briggs et al., 2007). All our

149   samples present typical damage profiles (Supplementary Figure 1). GEN72, RK1001, Post6 and

150   1343UnTal85 only retain damage in the last two bases as these libraries were prepared using a

151   'UDG-half' protocol (Rohland et al., 2015, See Methods).

152   The six reconstructed genomes and their plasmids were compared to the two Bronze

153   Age genomes reported previously (Rasmussen et al., 2015). After visual inspection of aligned

154   reads, our prehistoric genomes from Europe showed similar coverage of the reference genome

155   CO92, and all regions were also covered in the Bronze Age Altai *Y. pestis* genomes (Figure

156   2A). The six reconstructed genomes in this study lack the same region of the pMT1 plasmid,

157   which contains the *ymt* gene (Figure 2), as already identified in the Altai genomes (Rasmussen

158   et al., 2015). The *ymt* gene codes for the *Yersinia* murine toxin, which is an important virulence

159   factor in *Y. pestis* related to transmission via the flea vector (Hinnebusch et al., 2002, 2000).

160   The expression of *ymt* protects the bacteria from toxic blood digestion by-products in the flea's

161   gut and thus functions to aid in colonization of the flea midgut (Hinnebusch et al., 2002).

162

163   Phylogeny and Dating

164   To assess the phylogenetic positioning of the six European LNBA *Y. pestis* genomes with

165   respect to the modern and ancient *Y. pestis* genomes, Neighbour Joining (NJ, Supplementary

166   Figure 2A), Maximum Parsimony (MP, Supplementary Figure 2B) and Maximum Likelihood (ML,

167   Figure 3, Supplementary Figure 2C) trees were computed. Our samples form a distinct clade in

168   the *Y. pestis* phylogeny together with the previously reconstructed Southern Siberian Bronze

169   Age *Y. pestis* genomes (Rasmussen et al., 2015). This topology has a high bootstrap support of

170   >95% in all three methods. The branching point of the LNBA genomes with the main branch

171   leading towards the modern *Y. pestis* strains represents the most recent common ancestor

172   (MRCA) of all the extant and ancient *Y. pestis* genomes currently available.

7

173   To date the MRCA of *Y. pestis* we performed a 'tip dating' analysis using BEAST

174   (Drummond et al., 2012). The MRCA of all *Y. pestis* was dated to 6,078 years (95% HPD

175   interval: 5,036-7,494 years) suggesting a Holocene origin for plague, which is in agreement with

176   previous estimates (5,783 years, 95% HPD interval: 5,021–7,022 years, Rasmussen et al.,

177   2015). The time to the MRCA of *Y. pestis* and *Y. pseudotuberculosis* strain IP 32953 was

178   estimated to 28,258 years (95% HPD interval: 13,200-44,631 years). A maximum clade

179   credibility tree was computed (Supplementary Figure 3) supporting the same topology as the

180   NJ, MP and ML with high statistical support of the branching points of the LNBA plague clade.

181   # Genetic makeup

182   The effects of Single Nucleotide Polymorphisms (SNPs) detected in our dataset were

183   determined using the software *snpEff* (Cingolani et al., 2012) and an in-house program

184   (*MultiVCFAnalyzer*). A total of 423 SNPs were found in the LNBA branch including strain-

185   specific and shared SNPs. A total of 114 synonymous and 202 non-synonymous SNPs are

186   present in the LNBA branch. All the LNBA genomes share five SNPs: four non-synonymous and

187   one stop mutation (Supplementary Table 2).

188   Additionally, nine stop mutations were detected in the ancient branch, which were not

189   shared by all the LNBA genomes. Most of these mutations were found in the terminal part of the

190   LNBA branch with six being specific to the youngest Early Bronze Age *Y. pestis* strain

191   (RISE505), one being shared between RISE505 and Post6, one being GEN72-specific and one

192   being Gyvakarai1-specific (Supplementary Table 3). Additionally, RISE505 misses the start

193   codon of the YPO0956 gene, which is involved in iron transport, and one stop codon in

194   YPO2909, which is a pseudogene.

8

195   To identify potential homoplasies, a table of all variable SNPs was examined for any that

196 contradict the tree topology. Nine homoplasies were detected (Supplementary Table 4).

197 Furthermore, a tri-allelic site was detected at nucleotide position 4,104,762 (A,T,C).

198   The percentage of the gene covered in the LNBA plague genomes was calculated for a

199 set of genes that are related to virulence, flea transmission and colonization and dissemination

200 (Figure 2B). We observed the absence of YpfΦ (Derbise et al., 2007), a filamentous prophage,

201 in all LNBA plague genomes. While YpfΦ is found in some *Y. pestis* strains of branch 0, branch

202 1 and branch 2 as a free phage, it has only been fully integrated and stabilized into the

203 chromosome of the strains 1.ORI which are responsible for the third pandemic (Derbise and

204 Carniel, 2014). Additionally, the *yapC* gene was lost in the three younger LNBA strains

205 (1343UnTal85, Post6 and RISE505). YapC was initially thought to be involved in the adhesion

206 to mammalian cells, autoagglutination and biofilm formation when expressed in *E. coli* (Felek et

207 al., 2008). However, the *yapC* knockout in *Y. pestis* does not affect those functions. Felek and

208 colleagues have thus suggested that this is due to either low expression of *yapC in vitro* or by

209 compensation through other genes (Felek et al., 2008). The only virulence factor located in the

210 plasmids missing in all the LNBA *Y. pestis* strains is *ymt* (Figure 2). Other virulence factors,

211 such as *pla* and *caf1,* were already present in the LNBA*Y. pestis* genomes. The *pla* gene is

212 involved in the dissemination of the bacteria in the mammalian host by promoting the migration

213 of the bacteria to the lymphatic nodes (Lathem et al., 2007; Sebbane et al., 2006), while the

214 *caf1* gene encodes the F1 capsular antigen, which confers phagocytosis resistance to the

215 bacterium (Du et al., 2002). Both genes are absent in the closest relative *Y. pseudotuberculosis.*

216   Urease D (*ureD*) is an important gene that plays a role in flea transmission. When *ureD*

217 is expressed in the flea vector it causes a toxic oral reaction to the flea killing around 30-40%

218 (Chouikha and Hinnebusch, 2014). While *ureD* is functional in *Y. pseudotuberculosis*, it is a

219 pseudogene in *Y. pestis.* The pseudogenization of this gene is caused by a frameshift mutation

220 (insertion of a G in a six G-stretch) in *Y. pestis* (Sebbane et al., 2001). The LNBA *Y. pestis*

221   genomes were inspected in the search of this specific frameshift mutation. This insertion is not

222   present in those genomes indicating that this gene was still functional in *Y. pestis* at that time,

223   suggesting that it was as toxic to fleas as its ancestor *Y. pseudotuberculosis*.

224   Large-scale insertions and deletions (indels) were evaluated by comparison of mapped

225   data for the LNBA *Y. pestis* genomes, branch 0 strains (0.PE7, 0.PE2-F. 0.PE3, 0.PE4), KIM,

226   and CO92 using *Y. pseudotuberculosis* IP 32953 (NC_006155.1) as a reference. Regions larger

227   than 1 kb were explored as possible indels. We detected two regions present in the LNBA *Y.*

228   *pestis* genomes that are absent in all the other strains analyzed: a 1kb region (2,587,386-

229   2,588,553) that contains a single gene (YPTB0714) encoding an aldehyde dehydrogenase, part

230   of the R3 *Y. pseudotuberculosis*-specific region identified by Pouillot et al., 2008 and a second

231   region (1.5kb, 3,295,644-3,297,223) that contains a single gene (YPTB2793) encoding a

232   uracil/xanthine transporter being part of the region orf1 defined by Pouillot et al., 2008, which

233   was also characterized as *Y. pseudotuberculosis*-specific. Additionally, two missing regions

234   were detected: one region of 34kb is missing in the three younger genomes of the LNBA lineage

235   (Post6, 1343UnTal85 and RISE505) and another 36kb region, which contains flagella genes, is

236   missing in the youngest sample RISE505, as shown by Rasmussen et al., 2015, which contains

237   flagella genes. These two missing regions contain multiple membrane proteins, which could be

238   potential virulence factors or antigens recognized by the immune system of the host.

239   Discussion

240   The six prehistoric genomes presented here are the first complete *Y. pestis* genomes spanning

241   from the Late Neolithic to the Bronze Age in Europe. They form a distinct clade with the

242   previously reconstructed Southern Siberian Bronze Age *Y. pestis* genomes, confirming that all

243   LNBA genomes identified so far originate from a common ancestor. The previous reported

244   genome RISE509 (Rasmussen et al., 2015) together with the reconstructed RK1001 genome

245   reported here occupy the most basal position of all *Y. pestis* genomes sequenced to date. This

246   suggests that Central Eurasia rather than Eastern Asia should be considered as the region of

247   potential plague origin.

248   The temporal and spatial distribution of the Late Neolithic and Bronze Age *Y. pestis*

249   genomes allows us to evaluate the evolution and dissemination of plague in prehistory. We

250   propose two contrasting scenarios to explain the phylogenetic pattern observed in the LNBA *Y.*

251   *pestis* branch:

252   1. **Plague was introduced multiple times to Europe** from a common reservoir between

253   5,000 to 3,000 BP. Here, the bacterium would have been spread independently from a

254   source, most likely located in Central Eurasia, to Europe at least four times during a

255   period of over 1,000 years (Figure 1A), once to Lithuania and Croatia, once to Estonia,

256   and two times to Southern Germany. A similar "multiple wave" proposal has been made

257   for the second pandemic, where climatic fluctuation was considered as driving changes

258   in rodent populations (Schmid et al., 2015). We do not have such data for the time

259   periods in question here, and thus cannot speculate on the mechanism.

260   2. **Plague entered Europe once during the Neolithic**. From here it established a

261   reservoir within or close to Europe from which it circulated, and then moved back to

262   Central Eurasia and the Altai region/East Asia during the Bronze Age (Figure 1B). This

263   parallels the scenario of local persistence and eastward movement during the second

264   pandemic that is gaining support as more genetic data become available (Seifert et al.,

265   2016; Spyrou et al., 2016).

266   With just a few genomes available it is difficult to disentangle the two hypotheses; however,

267   interpreting our data in the context of what is known from human genetics and archaeological

268   data can offer some resolution. Ancient human genomic data point to a change in mobility and a

269   large scale expansion of people from the Caspian-Pontic Steppe related to individuals

270   associated with the 'Yamnaya' complex, both to the East and the West starting around 4,800

271   BP. These people carried a distinct genetic component that first appears in Central European

272   individuals from the Corded Ware Complex and then forms/becomes part of the genetic

273   composition of most subsequent and all modern day European populations (Allentoft et al.,

274   2015; Haak et al., 2015). It was furthermore shown that there is a close genetic link between the

275   highly mobile groups of people associated to the Southern Siberian 'Afanasievo Complex', the

276   'Yamnaya', and the Central and Eastern European Corded Ware Complex (Allentoft et al.,

277   2015).

278   Our earliest indication of plague in Europe is found in Croatia and the Baltic region and

279   coincides with the time of the arrival of the genetic steppe component (Allentoft et al., 2015).

280   The two Late Neolithic *Y. pestis* genomes from the Baltic in this study were reconstructed from

281   individuals associated with the Corded Ware Complex (Gyvakarai1 and KunilaII). The Baltic and

282   Croatian *Y. pestis* genomes are genetically derived from a common ancestor of the strain

283   RK1001, reconstructed from an individual associated to the 'Yamnaya' complex and RISE509

284   from the 'Afanasievo' complex from the Altai region, suggesting that the pathogen might have

285   spread with steppe people from Central Eurasia to Eastern and Central Europe during their

286   large scale expansion. Furthermore, human genomic analyses indicate that the individuals

287   RISE509, Gyvakarai1, KunilaII and GEN72 carry 'steppe ancestry' (Mathieson et al., 2017;

288   Mittnik et al., 2017) Evidence for these long distance contacts is also present in the

289   archaeological record. For example, the Gyvakarai1 burial is characterised by both a specific

290   set of grave inventory (hammer headed pin) and distinct skeletal morphology, which have no

291   analogues in earlier local populations (Tebelškis and Jankauskas, 2006).

292   The younger Late Neolithic *Y. pestis* genomes from Southern Germany are genetically

293   derived from the Baltic strains and are found in individuals associated with the Bell Beaker

294   Complex. Previous analyses have shown that Bell Beaker individuals from Germany also carry

295   'steppe ancestry' (Allentoft et al., 2015; Haak et al., 2015). This suggests that *Y. pestis* may

296   have been spread further southwestwards analogous to the human steppe component. The

297   youngest of the LNBA *Y. pestis* genomes (RISE505), found also in the Altai region, associated

298   with the Central Eurasian 'Andronovo' complex, descends from the Central European strains,

299   which suggests a spread back into Southern Siberia. Interestingly, genome-wide human data

300   shows that human individuals associated to the Sintashta, Srubnaya and Andronovo cultural

301   complexes in the Eurasian steppes (dating from around 3,700-3,500 BP) carried mixed ancestry

302   of middle Neolithic European farmers and Bronze Age steppe people, suggesting a backflow of

303   human genes from Europe to Central Eurasia (Allentoft et al., 2015). Form an archeological

304   perspective there is a close connection of the Abashevo cultural complex and Sintashta, that

305   might also have included population shifts West to East. In particular, the post-Sintashta

306   Andonovo cultural complex is an epoch of massive population shifts affecting all the area east of

307   the Urals to the Western borders of China including populations with European origin

308   (Koryakova and Epimakhov, 2007; Kuzmina, 2008). The steppe, as a natural corridor

309   connecting people and their livestock throughout Central and Western Eurasia, might have

310   facilitated the spread of strains closely related to the European Early Bronze Age *Y. pestis* back

311   to the Altai region, where RISE505 was found. The patterns in human genetic ancestry and

312   admixture, in combination with the temporal series within the LNBA *Y. pestis* branch, therefore

313   support scenario 2, suggesting that *Y. pestis* was introduced to Europe from the steppe around

314   4,800 BP. Thereafter, the pathogen became established in a local reservoir within or in close

315   proximity to Europe, from where the European *Y. pestis* strain was disseminated back to the

316   Altai region in a process connected to the backflow of human genetic ancestry from Western

317   Eurasia into Southern Siberia. The pathogen diversity, therefore, mirrors the archaeological

318   evidence, which indicates a strong intensification of Eurasian networks since the beginning of

319   the Bronze Age (Vandkilde, 2016).

320   Even though *Y. pestis* seems to have been spread following human movements, its

321   mode of transmission during this early phase of its evolution cannot be easily determined. Most

322   contemporary cases of *Y. pestis* infection occur via an arthropod vector and stem from a sylvatic

13

323   rodent population that has resistance to the bacterium. The flea transmission can be

324   accomplished by one of two mechanisms: the classical blockage-dependent flea transmission

325   (Hinnebusch et al., 1998) and the recently proposed early-phase transmission (EPT) (Eisen et

326   al., 2006). In the blockage-dependent model, *Y. pestis* causes an obstruction in the flea

327   digestive system by producing a biofilm that blocks the pre-gut of the flea within 1-2 weeks after

328   infection. This blockage prevents a blood meal from reaching the flea's gut, and regurgitation of

329   the blood by a hungry flea in repeated attempts to feed sheds several live bacteria into the

330   blood stream of the host (Chouikha and Hinnebusch, 2012; Hinnebusch et al., 1998). It has

331   been shown that the blockage-dependent transmission requires a functional *ymt* gene and *hms*

332   locus, and non-functional *rcsA, pde2* and *pde3* genes (Sun et al., 2014). *ymt* protects *Y. pestis*

333   from toxic by-products of blood digestion and allows the bacterium to colonise the mid-gut of the

334   flea. The *hms* locus is involved in biofilm formation and *rcsA*, *pde2* and *pde3* are the down-

335   regulators of biofilm formation. However, evidence is emerging that *Y. pestis* can be transmitted

336   efficiently within the first 1-4 days after entering the flea prior to biofilm formation (Eisen et al.,

337   2015, 2006), in a process known as the EPT model. Unfortunately this model is currently less

338   well understood molecularly and physiologically than blockage-dependent transmission, but has

339   been shown to be biofilm (Vetter et al., 2010) and *ymt* independent (Johnson et al., 2014).

340   Based on the genetic characteristics of the LNBA genomes (i.e. lack of *ymt*, still

341   functional *pde2* and *rcsA* as shown by previous work (Rasmussen et al., 2015), functional *ureD*

342   which will kill 30-40% of the flea vectors) it seems most parsimonious that *Y. pestis* was not able

343   to use a flea vector in a blockage-dependent model. However, since none of these genes seem

344   to be required for EPT, it remains possible that LNBA *Y. pestis* was transmitted by a flea vector

345   via this transmission mode. Under this assumption, the transmission would have been

346   presumably less efficient since a functional Urease D would have reduced the number of fleas

347   transmitting the bacteria.

348   The presence of genes involved in virulence in the mammalian host such as *pla* and

349   *caf1*, which are absent in *Y. pseudotuberculosis*, indicates that LNBA *Y. pestis* was already

350   adapted to mammalian hosts to some extent. *pla* aids in *Y. pestis* infiltration of the mammalian

351   host (Lathem et al., 2007; Sebbane et al., 2006). The *pla* gene present in the LNBA *Y. pestis*

352   strains has the ancestral I259 variant, which has been shown to be less efficient than the

353   derived T259 form (Haiko et al., 2009). *Y. pestis* with the ancestral variant is able to cause

354   pneumonic disease, however, it is less efficient in colonising other tissues (Zimbler et al., 2015).

355   This indicates that LNBA *Y. pestis* could potentially cause a pneumonic or a less virulent

356   bubonic form. In addition to the above noted changes, we detected two regions missing in the

357   LNBA genomes: a ~34kb region that contains genes encoding membrane proteins missing in

358   the three youngest *Y. pestis* strains (Post6, 1343UnTal85 and RISE505) and a ~36kb region

359   containing genes encoding proteins involved in flagellin production and iron transporters missing

360   in the youngest sample RISE505, as observed elsewhere (Rasmussen et al., 2015). This

361   genome decay affecting membrane and flagellar proteins potentially involved in interactions with

362   the host's immune system, can be an indication of adaptation to a new host pathogenic lifestyle

363   (Ochman and Moran, 2001).

364   Our common understanding is that plague is a disease adapted to rodents, where

365   commensal species such as *Rattus rattus* and their fleas play a central role as disease vectors

366   for humans (Perry and Fetherston, 1997). While a rodent-flea mediated transmission model is

367   compatible with the genomic makeup of the LNBA strains, disease dynamics may well have

368   differed in the past. The most parsimonious explanation would be that LNBA plague indeed

369   traveled with rodent species commensal to humans, in keeping with the orthodox model of

370   plague transmission. The Neolithic is conventionally considered to be a time period where new

371   diseases were introduced into human groups as they made the transition from a nomadic

372   lifestyle to one of sedentism, and where the adoption of agriculture and increased population

373   density acted synergistically to change the disease landscape (Ronald Barrett et al., 1998).

374   Whether commensal rodent populations were large enough to function as reservoir populations

375   for plague during human migrations at this time is unknown. In central Eurasian Bronze Age

376   cultures, agriculture, i.e. large scale food storage, is mostly absent (Ryabogina and Ivanov,

377   2011) However, contact between steppe inhabiting rodents, pastoralists and their herds might

378   have been frequent when moving within these environments. Alternative models of transmission

379   involving different host species, perhaps even humans or their domesticates, might carry some

380   traction, as the ancient disease may have behaved rather differently from the form we know

381   today.

382   Here, we present the first LNBA *Y. pestis* genomes from Europe. We show that all LNBA

383   genomes reconstructed so far form a distinct lineage that potentially entered Europe following

384   the migration of steppe people around 4,800 BP. We find striking parallels between the *Y. pestis*

385   dispersal pattern and human population movements during this time period. We propose two

386   scenarios for presence of the bacteria in Europe: a multiple introduction hypothesis from a

387   Central Eurasian source, or the establishment of a local *Y. pestis* focus within or close to

388   Europe from where a resident strain ultimately moved back towards Central Asia in the Bronze

389   Age. On account of the chronology and the tight synergy between the ancient *Y. pestis*

390   phylogeny and known patterns of human mobility, we find stronger support for the second

391   scenario.

392   The LNBA period was a time of increased mobility and cultural change. The presence of

393   *Y. pestis* may have been a promoting factor for the increase in mobility of human populations

394   (Rasmussen et al., 2015). The manifestation of the disease in Europe could have played a

395   major role in the processes that led to the genetic turnover observed in the European human

396   populations, who may have harbored different levels of immunity against this newly introduced

397   disease. Testing these hypotheses will require more extensive assessment of both human and

398   *Y. pestis* genomes from the presumed source population before and after migration from the

399   steppes, as well as in Europe during this period of genetic turnover.

Authors contribution

J.K., A.H. and A.A.V. conceived the study. K.M., R.A., M.D., R.J., M.T., P.W.S., A.B., I.J., M. N., S.R., M.S., A.S., S.H. provided the samples and performed archaeological assessment. A.M., S.P., M.F., A.A.V. performed laboratory work. A.A.V., A.H., M.A.S., F.M.K. and J.K. analysed the data. A.A.V, A.H., J.K., P.W.S., K.I.B., W.H. and A.M. wrote the manuscript with contributions from all co-authors. All authors read and approved the final manuscript.

Competing financial interests

The authors declare no competing financial interests.

Data availability

Raw sequencing data have been deposited at the European Nucleotide Archive under accession PRJEBXXXXX

Acknowledgements

17

429   References

430   Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup,

431   J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S.,

432   Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa,

433   P.D., Dąbrowski, P., Duffy, P.R., Ebel, A.V., Epimakhov, A., Frei, K., Furmanek,

434   M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R.,

435   Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi,

436   C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtchyan, R.,

437   Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag,

438   L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G.,

439   Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L.,

440   Sicheritz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015.

441   Population genomics of Bronze Age Eurasia. Nature 522, 167–172.

442   doi:10.1038/nature14507

443   Benedictow, O.J., 2004. The Black Death, 1346-1353: The Complete History. Boydell

444   Press.

445   Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J.,

446   Schuenemann, V.J., Poinar, D., Kuch, M., Golding, G.B., Dutour, O., Keim, P.,

447   Wagner, D.M., Holmes, E.C., Krause, J., Poinar, H.N., 2016. Eighteenth century

448   Yersinia pestis genomes reveal the long-term persistence of an historical plague

449   focus. eLife 5, e12994. doi:10.7554/eLife.12994

450   Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N.,

451   Coombes, B.K., McPhee, J.B., DeWitte, S.N., Meyer, M., Schmedes, S., Wood,

452   J., Earn, D.J.D., Herring, D.A., Bauer, P., Poinar, H.N., Krause, J., 2011. A draft

453   genome of Yersinia pestis from victims of the Black Death. Nature 478, 506–510.

454   doi:10.1038/nature10549

455   Briggs, A., Heyn, P., 2012. Preparation of Next-Generation Sequencing Libraries from

456   Damaged DNA, in: Shapiro, B., Hofreiter, M. (Eds.), Ancient DNA, Methods in

457   Molecular Biology. Humana Press, pp. 143–154. doi:10.1007/978-1-61779-516-

458   9_18

459   Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M.,

460   Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in

461   genomic DNA sequences from a Neandertal. PNAS 104, 14616–14621.

462   doi:10.1073/pnas.0704665104

463   Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K.,

464   Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics

465   10, 421. doi:10.1186/1471-2105-10-421

466   Chouikha, I., Hinnebusch, B.J., 2014. Silencing urease: A key evolutionary step that

467   facilitated the adaptation of Yersinia pestis to the flea-borne transmission route.

468   PNAS 111, 18709–18714. doi:10.1073/pnas.1413209111

469   Chouikha, I., Hinnebusch, B.J., 2012. Yersinia–flea interactions and the evolution of the

470   arthropod-borne transmission route of plague. Current Opinion in Microbiology,

471   Ecology and industrial microbiology/Special section: Microbial proteomics 15,

472   239–246. doi:10.1016/j.mib.2012.02.003

473   Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X.,

474   Ruden, D.M., 2012. A program for annotating and predicting the effects of single

475   nucleotide polymorphisms, SnpEff. Fly 6, 80–92. doi:10.4161/fly.19695

476   Cohn JR, S.K., 2008. 4 Epidemiology of the Black Death and Successive Waves of

477   Plague. Med Hist Suppl 74–100.

478   Cui, Y., Yu, C., Yan, Y., Li, D., Li, Y., Jombart, T., Weinert, L.A., Wang, Z., Guo, Z., Xu,

479   L., Zhang, Y., Zheng, H., Qin, N., Xiao, X., Wu, M., Wang, X., Zhou, D., Qi, Z.,

480   Du, Z., Wu, H., Yang, X., Cao, H., Wang, H., Wang, J., Yao, S., Rakin, A., Li, Y.,

481   Falush, D., Balloux, F., Achtman, M., Song, Y., Wang, J., Yang, R., 2013.

482   Historical variations in mutation rate in an epidemic pathogen, Yersinia pestis.

483   PNAS 110, 577–582. doi:10.1073/pnas.1205750110

484   Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B.,

485   Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete

486   mitochondrial genome sequence of a Middle Pleistocene cave bear

487   reconstructed from ultrashort DNA fragments. PNAS 110, 15758–15763.

488   doi:10.1073/pnas.1314445110

489   Derbise, A., Carniel, E., 2014. YpfΦ: a filamentous phage acquired by Yersinia pestis.

490   Front Microbiol 5. doi:10.3389/fmicb.2014.00701

491   Derbise, A., Chenal-Francisque, V., Pouillot, F., Fayolle, C., Prévost, M.-C., Médigue,

492   C., Hinnebusch, B.J., Carniel, E., 2007. A horizontally acquired filamentous

493   phage contributes to the pathogenicity of the plague bacillus. Molecular

494   Microbiology 63, 1145–1157. doi:10.1111/j.1365-2958.2006.05570.x

495   Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics

496   with BEAUti and the BEAST 1.7. Mol Biol Evol 29, 1969–1973.

497   doi:10.1093/molbev/mss075

498   Du, Y., Rosqvist, R., Forsberg, Å., 2002. Role of Fraction 1 Antigen of Yersinia pestis in

499   Inhibition of Phagocytosis. Infect. Immun. 70, 1453–1460.

500   doi:10.1128/IAI.70.3.1453-1460.2002

501   Eisen, R.J., Bearden, S.W., Wilder, A.P., Montenieri, J.A., Antolin, M.F., Gage, K.L.,

502   2006. Early-phase transmission of Yersinia pestis by unblocked fleas as a

503   mechanism explaining rapidly spreading plague epizootics. PNAS 103, 15380–

504   15385. doi:10.1073/pnas.0606831103

505   Eisen, R.J., Dennis, D.T., Gage, K.L., 2015. The Role of Early-Phase Transmission in

506   the Spread of Yersinia pestis. J Med Entomol 52, 1183–1192.

507   doi:10.1093/jme/tjv128

508   Feldman, M., Harbeck, M., Keller, M., Spyrou, M.A., Rott, A., Trautmann, B., Scholz,

509   H.C., Päffgen, B., Peters, J., McCormick, M., Bos, K., Herbig, A., Krause, J.,

510   2016. A high-coverage Yersinia pestis Genome from a 6th-century Justinianic

511   Plague Victim. Mol Biol Evol msw170. doi:10.1093/molbev/msw170

512   Felek, S., Lawrenz, M.B., Krukonis, E.S., 2008. The Yersinia pestis autotransporter

513   YapC mediates host cell binding, autoaggregation and biofilm formation.

514   Microbiology 154, 1802–1812. doi:10.1099/mic.0.2007/010918-0

515   Fu, Q., Meyer, M., Gao, X., Stenzel, U., Burbano, H.A., Kelso, J., Pääbo, S., 2013. DNA

516   analysis of an early modern human from Tianyuan Cave, China. PNAS 110,

517   2223–2227. doi:10.1073/pnas.1221359110

518   Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010.

519   New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies:

520   Assessing the Performance of PhyML 3.0. Syst Biol 59, 307–321.

521   doi:10.1093/sysbio/syq010

522   Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G.,

523   Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E.,

524   Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F.,

525   Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov,

526   O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M.A., Roth,

527   C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D.,

528   Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a

529   source for Indo-European languages in Europe. Nature 522, 207–211.

530   doi:10.1038/nature14317

531   Haiko, J., Kukkonen, M., Ravantti, J.J., Westerlund-Wikström, B., Korhonen, T.K., 2009.

532   The Single Substitution I259T, Conserved in the Plasminogen Activator Pla of

533   Pandemic Yersinia pestis Branches, Enhances Fibrinolytic Activity. J. Bacteriol.

534   191, 4758–4766. doi:10.1128/JB.00489-09

535   Hinnebusch, B.J., Fischer, E.R., Schwan, T.G., 1998. Evaluation of the Role of the

536   Yersinia pestis Plasminogen Activator and Other Plasmid-Encoded Factors in

537   Temperature-Dependent Blockage of the Flea. J Infect Dis. 178, 1406–1415.

538   doi:10.1086/314456

539   Hinnebusch, B.J., Rudolph, A.E., Cherepanov, P., Dixon, J.E., Schwan, T.G., Forsberg,

540   Å., 2002. Role of Yersinia Murine Toxin in Survival of Yersinia pestis in the

541   Midgut of the Flea Vector. Science 296, 733–735. doi:10.1126/science.1069972

542   Hinnebusch, J., Cherepanov, P., Du, Y., Rudolph, A., Dixon, J.D., Schwan, T.,

543   Forsberg, Å., 2000. Murine toxin of Yersinia pestis shows phospholipase D

544   activity but is not required for virulence in mice. International Journal of Medical

545   Microbiology 290, 483–487. doi:10.1016/S1438-4221(00)80070-3

546   Johnson, T.L., Hinnebusch, B.J., Boegler, K.A., Graham, C.B., MacMillan, K.,

547   Montenieri, J.A., Bearden, S.W., Gage, K.L., Eisen, R.J., 2014. Yersinia murine

548   toxin is not required for early-phase transmission of Yersinia pestis by Oropsylla

549   montana (Siphonaptera: Ceratophyllidae) or Xenopsylla cheopis (Siphonaptera:

550   Pulicidae). Microbiology 160, 2517–2525. doi:10.1099/mic.0.082123-0

551   Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013.

552   mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage

553   parameters. Bioinformatics 29, 1682–1684. doi:10.1093/bioinformatics/btt193

554   Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in

555   multiplex sequencing on the Illumina platform. Nucleic Acids Res 40, e3–e3.

556   doi:10.1093/nar/gkr771

557 Kislichkina, A.A., Bogun, A.G., Kadnikova, L.A., Maiskaya, N.V., Platonov, M.E.,

558   Anisimov, N.V., Galkina, E.V., Dentovskaya, S.V., Anisimov, A.P., 2015.

559   Nineteen Whole-Genome Assemblies of Yersinia pestis subsp. microtus,

560   Including Representatives of Biovars caucasica, talassica, hissarica, altaica,

561   xilingolensis, and ulegeica. Genome Announc 3. doi:10.1128/genomeA.01342-15

562 Koryakova, L., Epimakhov, A.V., 2007. The Urals and Western Siberia in the Bronze

563   and Iron Ages. Cambridge University Press.

564 Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones,

565   S.J., Marra, M.A., 2009. Circos: An information aesthetic for comparative

566   genomics. Genome Res. doi:10.1101/gr.092759.109

567 Kuzmina, E.E., 2008. The Prehistory of the Silk Road. University of Pennsylvania Press.

568 Lathem, W.W., Price, P.A., Miller, V.L., Goldman, W.E., 2007. A plasminogen-activating

569   protease specifically controls the development of primary pneumonic plague.

570   Science 315, 509–513. doi:10.1126/science.1137195

571 Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler

572   transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

573 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,

574   Abecasis, G., Durbin, R., Subgroup, 1000 Genome Project Data Processing,

575   2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25,

576   2078–2079. doi:10.1093/bioinformatics/btp352

577 Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A.,

578   Harney, E., Stewardson, K., Fernandes, D., Novak, M., Sirak, K., Gamba, C.,

579   Jones, E.R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J.L., de Castro,

24

580   J.M.B., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M.,

581   Meller, H., Mochalov, O., Moiseyev, V., Guerra, M.A.R., Roodenberg, J., Vergès,

582   J.M., Krause, J., Cooper, A., Alt, K.W., Brown, D., Anthony, D., Lalueza-Fox, C.,

583   Haak, W., Pinhasi, R., Reich, D., 2015. Genome-wide patterns of selection in 230

584   ancient Eurasians. Nature 528, 499–503. doi:10.1038/nature16152

585 Mathieson, I., Roodenberg, S.A., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick,

586   S., Olade, I., Broomandkhoshbacht, N., Cheronet, O., Fernandes, D., Ferry, M.,

587   Gamarra, B., Fortes, G.G., Haak, W., Harney, E., Krause-Kyora, B.,

588   Kucukkalipci, I., Michel, M., Mittnik, A., Nägele, K., Novak, M., Oppenheimer, J.,

589   Patterson, N., Pfrengle, S., Sirak, K., Stewardson, K., Vai, S., Alexandrov, S., Alt,

590   K.W., Andreescu, R., Antonović, D., Ash, A., Atanassova, N., Bacvarov, K.,

591   Gusztáv, M.B., Bocherens, H., Bolus, M., Boroneanţ, A., Boyadzhiev, Y., Budnik,

592   A., Burmaz, J., Chohadzhiev, S., Conard, N.J., Cottiaux, R., Čuka, M., Cupillard,

593   C., Drucker, D.G., Elenski, N., Francken, M., Galabova, B., Ganetovski, G., Gely,

594   B., Hajdu, T., Handzhyiska, V., Harvati, K., Higham, T., Iliev, S., Janković, I.,

595   Karavanić, I., Kennett, D.J., Komšo, D., Kozak, A., Labuda, D., Lari, M., Lazar,

596   C., Leppek, M., Leshtakov, K., Vetro, D.L., Los, D., Lozanov, I., Malina, M.,

597   Martini, F., McSweeney, K., Meller, H., Menđušić, M., Mirea, P., Moiseyev, V.,

598   Petrova, V., Price, T.D., Simalcsik, A., Sineo, L., Šlaus, M., Slavchev, V., Stanev,

599   P., Starović, A., Szeniczey, T., Talamo, S., Teschler-Nicola, M., Thevenet, C.,

600   Valchev, I., Valentin, F., Vasilyev, S., Veljanovska, F., Venelinova, S.,

601   Veselovskaya, E., Viola, B., Virag, C., Zaninović, J., Zäuner, S., Stockhammer,

602   P.W., Catalano, G., Krauß, R., Caramelli, D., Zariņa, G., Gaydarska, B., Lillie, M.,

603   Nikitin, A.G., Potekhina, I., Papathanasiou, A., Borić, D., Bonsall, C., Krause, J.,

604   Pinhasi, R., Reich, D., 2017. The Genomic History Of Southeastern Europe.

605   bioRxiv 135616. doi:10.1101/135616

606 Meyer, M., Kircher, M., 2010. Illumina Sequencing Library Preparation for Highly

607   Multiplexed Target Capture and Sequencing. Cold Spring Harb Protoc 2010,

608   pdb.prot5448. doi:10.1101/pdb.prot5448

609 Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R.,

610   Khartanovich, V., Moiseyev, V., Furtwängler, A., Valtueña, A.A., Feldman, M.,

611   Economou, C., Oinonen, M., Vasks, A., Tõrv, M., Balanovsky, O., Reich, D.,

612   Jankauskas, R., Haak, W., Schiffels, S., Krause, J., 2017. The Genetic History of

613   Northern Europe. bioRxiv 113241. doi:10.1101/113241

614 Ochman, H., Moran, N.A., 2001. Genes lost and genes found: evolution of bacterial

615   pathogenesis and symbiosis. Science 292, 1096–1099.

616 Okonechnikov, K., Conesa, A., García-Alcalde, F., 2016. Qualimap 2: advanced multi-

617   sample quality control for high-throughput sequencing data. Bioinformatics 32,

618   292–294. doi:10.1093/bioinformatics/btv566

619 Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., Nieselt, K., 2016.

620   EAGER: efficient ancient genome reconstruction. Genome Biology 17, 60.

621   doi:10.1186/s13059-016-0918-z

622 Perry, R.D., Fetherston, J.D., 1997. Yersinia pestis--etiologic agent of plague. Clin.

623   Microbiol. Rev. 10, 35–66.

624   Pouillot, F., Fayolle, C., Carniel, E., 2008. Characterization of Chromosomal Regions

625   Conserved in Yersinia pseudotuberculosis and Lost by Yersinia pestis. Infect.

626   Immun. 76, 4592–4599. doi:10.1128/IAI.00568-08

627   Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing

628   genomic features. Bioinformatics 26, 841–842. doi:10.1093/bioinformatics/btq033

629   R Development Core Team, 2008. R: A Language and Environment for Statistical

630   Computing. R Foundation for Statistical Computing, Vienna, Austria.

631   Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.-G.,

632   Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., Nielsen, H.B.,

633   Brunak, S., Avetisyan, P., Epimakhov, A., Khalyapin, M.V., Gnuni, A., Kriiska, A.,

634   Lasak, I., Metspalu, M., Moiseyev, V., Gromov, A., Pokutta, D., Saag, L., Varul,

635   L., Yepiskoposyan, L., Sicheritz-Pontén, T., Foley, R.A., Lahr, M.M., Nielsen, R.,

636   Kristiansen, K., Willerslev, E., 2015. Early Divergent Strains of Yersinia pestis in

637   Eurasia 5,000 Years Ago. Cell 163, 571–582. doi:10.1016/j.cell.2015.10.009

638   Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D., 2015. Partial uracil-DNA-

639   glycosylase treatment for screening of ancient DNA. Philos. Trans. R. Soc.

640   Lond., B, Biol. Sci. 370, 20130624. doi:10.1098/rstb.2013.0624

641   Ronald Barrett, Christopher W. Kuzawa, Thomas McDade, Armelagos, and G.J., 1998.

642   EMERGING AND RE-EMERGING INFECTIOUS DISEASES: The Third

643   Epidemiologic Transition. Annual Review of Anthropology 27, 247–271.

644   doi:10.1146/annurev.anthro.27.1.247

645   Russell, J.C., 1968. That earlier plague. Demography 5, 174–184.

646   doi:10.1007/BF03208570

647   Ryabogina, N.E., Ivanov, S.N., 2011. Ancient agriculture in Western Siberia: problems

648   of argumentation, paleoethnobotanic methods, and analysis of data.

649   Archaeology, Ethnology and Anthropology of Eurasia 39, 96–106.

650   doi:10.1016/j.aeae.2012.02.011

651   Schmid, B.V., Büntgen, U., Easterday, W.R., Ginzler, C., Walløe, L., Bramanti, B.,

652   Stenseth, N.C., 2015. Climate-driven introduction of the Black Death and

653   successive plague reintroductions into Europe. PNAS 112, 3020–3025.

654   doi:10.1073/pnas.1412887112

655   Schuenemann, V.J., Bos, K., DeWitte, S., Schmedes, S., Jamieson, J., Mittnik, A.,

656   Forrest, S., Coombes, B.K., Wood, J.W., Earn, D.J.D., White, W., Krause, J.,

657   Poinar, H.N., 2011. Targeted enrichment of ancient pathogens yielding the

658   pPCP1 plasmid of Yersinia pestis from victims of the Black Death. PNAS 108,

659   E746–E752. doi:10.1073/pnas.1105107108

660   Sebbane, F., Devalckenaere, A., Foulon, J., Carniel, E., Simonet, M., 2001. Silencing

661   and Reactivation of Urease inYersinia pestis Is Determined by One G Residue at

662   a Specific Position in the ureD Gene. Infect. Immun. 69, 170–176.

663   doi:10.1128/IAI.69.1.170-176.2001

664   Sebbane, F., Jarrett, C.O., Gardner, D., Long, D., Hinnebusch, B.J., 2006. Role of the

665   Yersinia pestis plasminogen activator in the incidence of distinct septicemic and

666   bubonic forms of flea-borne plague. Proc Natl Acad Sci U S A 103, 5526–5530.

667   doi:10.1073/pnas.0509544103

668   Seifert, L., Wiechmann, I., Harbeck, M., Thomas, A., Grupe, G., Projahn, M., Scholz,

669   H.C., Riehm, J.M., 2016. Genotyping Yersinia pestis in Historical Plague:

670   Evidence for Long-Term Persistence of Y. pestis in Europe from the 14th to the

671   17th Century. PLOS ONE 11, e0145194. doi:10.1371/journal.pone.0145194

672   Spyrou, M.A., Tukhbatova, R.I., Feldman, M., Drath, J., Kacki, S., Beltrán de Heredia,

673   J., Arnold, S., Sitdikov, A.G., Castex, D., Wahl, J., Gazimzyanov, I.R., Nurgaliev,

674   D.K., Herbig, A., Bos, K.I., Krause, J., 2016. Historical Y. pestis Genomes Reveal

675   the European Black Death as the Source of Ancient and Modern Plague

676   Pandemics. Cell Host & Microbe 19, 874–881. doi:10.1016/j.chom.2016.05.012

677   Stenseth, N.C., Atshabar, B.B., Begon, M., Belmain, S.R., Bertherat, E., Carniel, E.,

678   Gage, K.L., Leirs, H., Rahalison, L., 2008. Plague: Past, Present, and Future.

679   PLOS Medicine 5, e3. doi:10.1371/journal.pmed.0050003

680   Stockhammer, P.W., Massy, K., Knipper, C., Friedrich, R., Kromer, B., Lindauer, S.,

681   Radosavljević, J., Wittenborn, F., Krause, J., 2015. Rewriting the Central

682   European Early Bronze Age Chronology: Evidence from Large-Scale

683   Radiocarbon Dating. PLOS ONE 10, e0139705.

684   doi:10.1371/journal.pone.0139705

685   Sun, Y.-C., Jarrett, C.O., Bosio, C.F., Hinnebusch, B.J., 2014. Retracing the

686   Evolutionary Path that Led to Flea-borne Transmission of Yersinia pestis. Cell

687   Host Microbe 15, 578–586. doi:10.1016/j.chom.2014.04.003

688   Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular

689   Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30, 2725–2729.

690   doi:10.1093/molbev/mst197

691 Tebelškis, P., Jankauskas, R., 2006. The Late Neolithic grave at Gyvakarai in Lithuania

692   in the context of current archaeological and anthropological knowledge.

693   Archaeologia Baltica 8–20.

694 Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer

695   (IGV): high-performance genomics data visualization and exploration. Brief

696   Bioinform 14, 178–192. doi:10.1093/bib/bbs017

697 Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-

698   Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E.,

699   Garimella, K.V., Altshuler, D., Gabriel, S., DePristo, M.A., 2013. From FastQ

700   Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best

701   Practices Pipeline, in: Current Protocols in Bioinformatics. p. 43:11.10.1-

702   11.10.33.

703 Vandkilde, H., 2016. Bronzization: The Bronze Age as Pre-Modern Globalization.

704   Praehistorische Zeitschrift 91, 103–123. doi:10.1515/pz-2016-0005

705 Vetter, S.M., Eisen, R.J., Schotthoefer, A.M., Montenieri, J.A., Holmes, J.L., Bobrov,

706   A.G., Bearden, S.W., Perry, R.D., Gage, K.L., 2010. Biofilm formation is not

707   required for early-phase transmission of Yersinia pestis. Microbiology 156, 2216–

708   2225. doi:10.1099/mic.0.037952-0

709 Wagner, D.M., Klunk, J., Harbeck, M., Devault, A., Waglechner, N., Sahl, J.W., Enk, J.,

710   Birdsell, D.N., Kuch, M., Lumibao, C., Poinar, D., Pearson, T., Fourment, M.,

711   Golding, B., Riehm, J.M., Earn, D.J.D., DeWitte, S., Rouillard, J.-M., Grupe, G.,

712   Wiechmann, I., Bliska, J.B., Keim, P.S., Scholz, H.C., Holmes, E.C., Poinar, H.,

713   2014. Yersinia pestis and the Plague of Justinian 541–543 AD: a genomic

714   analysis. The Lancet Infectious Diseases 14, 319–326. doi:10.1016/S1473-

715   3099(13)70323-2

716   Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New

717   York.

718   Zhgenti, E., Johnson, S.L., Davenport, K.W., Chanturia, G., Daligault, H.E., Chain, P.S.,

719   Nikolich, M.P., 2015. Genome Assemblies for 11 Yersinia pestis Strains Isolated

720   in the Caucasus Region. Genome Announc. 3, e01030-15.

721   doi:10.1128/genomeA.01030-15

722   Zietz, B.P., Dunkelberg, H., 2004. The history of the plague and the research on the

723   causative agent Yersinia pestis. International Journal of Hygiene and

724   Environmental Health 207, 165–178. doi:10.1078/1438-4639-00259

725   Zimbler, D.L., Schroeder, J.A., Eddy, J.L., Lathem, W.W., 2015. Early emergence of

726   Yersinia pestis as a severe respiratory pathogen. Nat Commun 6.

727   doi:10.1038/ncomms8487

728

729 Figures and table legends

730 **Figure 1: Map of proposed *Yersinia pestis* circulation throughout Eurasia.** A) Entrance of

731 *Y. pestis* into Europe from Central Eurasia with the expansion of Yamnaya pastoralists around

732 4,800 years ago. B) Circulation of *Y.pestis* back into the Altai from Europe. Only complete

733 genomes are shown.

734 **Figure 2:** A) Average coverage plot for the chromosome and plasmids of *Yersinia pestis*, from

735 the outer ring to the inner ring: *Y. pestis* CO92 (NC_003143.1, reference), RISE509, RK1001,

736 GEN72, Gyvakarai1, KunilaII, 6Post, 1343UnTal85 and RISE505. Colours correspond to the

737 regions where the genomes were recovered from: Altai region (purple), Russia (red), Croatia

738 (dark yellow), Gyvakarai, Lithuania (blue), Kunila, Estonia (orange), Augsburg, Germany

739 (green). The average depth of coverage was calculated for 1kb regions for the chromosome and

740 100bp for the plasmids, each ring represents a maximum of 20X coverage. The figure was

741 generated with Circos (Krzywinski et al., 2009). B) Percentage covered of virulence factors

742 located on the *Yersinia pestis* chromosome and plasmids, plotted with in R using the ggplot2

743 package. [1] *ymt* gene, [2] *pla*, [3] deletion of flagelin genes, [4] filamentous prophage YpfΦ, [5]

744 *Y. pestis*-specific genes, [*] region mask in pPCP1 due to high similarity to expression vectors

745 during enzyme production (Schuenemann et al., 2011).

746 **Figure 3: Maximum Likelihood tree of all *Yersinia pestis* genomes including 1,867 SNPs**

747 **positions with complete deletion.** Nodes with support equal or higher than 95% are marked

748 with an asterisk. The colours represent different branches in the *Y. pestis* phylogeny: branch 0

749 (black), branch 1 (red), branch2 (green), branch 3 (blue), branch 4 (orange) and LNBA *Y. pestis*

750 branch (purple). *Y. pseudotuberculosis*-specific SNPs were excluded from the tree for

751 representative matters.

752 **Table 1: Genomes from the NCBI (RefSeq/Nucleotide) database, used in the multi-species**

753 **reference panel for screening for *Y. pestis* aDNA.**

754 **Table 2: Statistic of the *Y. pestis* genome reconstruction. BP = Before Present.**

STAR Methods

Sampling and extraction

Sampling of a total of 563 tooth and bone samples (Russia (122), Hungary and Croatia (139), Lithuania (27), Estonia (45), Latvia (10), and Germany (Althausen 4, Augsburg 83, Mittelelbe-Saale 133)) took place in the clean room facilities of the Institute for Archaeological Sciences at the University of Tu□bingen, the Institute of Archaeology RCH HAS in Budapest and of the MPI-SHH in Jena. After irradiation with UV light to remove surface contamination, teeth were sawed apart transversally at the border of crown and root, and dentine from inside the crown was sampled and powdered using a sterile dentistry drill. For the samples processed in Budapest, whole teeth were powdered. For bone samples, the surface layer from the sampling area was removed with a dentistry drill prior to obtaining bone powder from the inside of the bone by drilling. For each specimen we gathered between ~30 and 120 mg of powder to be used for DNA extraction.

Extraction was performed following a protocol optimized for the recovery of small ancient DNA molecules (Dabney et al., 2013), resulting in 100μl of DNA extract per sample. An aliquot of 20μl of extract was used to generate double-indexed libraries (Kircher et al., 2012; Meyer and Kircher, 2010). Negative controls were included in the extraction and library preparation and taken along for all further processing steps.

Shotgun screening

Libraries were PCR-amplified and quantified using an Agilent 2100 Bioanalyzer DNA 1000 chip

33

775   and pooled at equimolar concentrations prior to paired-end sequencing on a NextSeq500 with

776   2x101+8+8 and a HiSeq2500 with 2x101+8+8 cycles according to the manufacturer's

777   instructions to a depth of ~1.5 million reads per library.

In-silico screening

779   The sequencing data for the 170 samples was preprocessed with ClipAndMerge (Peltzer et al.,

780   2016) to remove adaptors, base quality-trim (20) and merging and filtering for only merged

781   reads. Reads were mapped using the BWA aln algorithm (Li and Durbin, 2009) to a multi-

782   species reference panel, containing various representatives of the genus *Yersinia* (Table 2) and

783   the plasmids of *Yersinia pestis*: pCD1, pMT1 and pPCP1 from *Y. pestis* CO92. The region

784   comprising 3000-4200bp of the *Y. pestis* specific plasmid pPCP1 was masked in the reference,

785   since it is highly similar to an expression vector used during the production of enzyme reagents

786   (Schuenemann et al., 2011).

787

788   Mapped files were then filtered for reads with a mapping quality higher than 20 with

789   Samtools (Li et al., 2009). PCR duplicates were removed using the MarkDuplicates tool in

790   Picard (1.140, http://broadinstitute.github.io/picard/). The number of reads mapping specifically

791   to each genome and to the plasmids were retrieved from the bam files using Samtools (Li et al.,

792   2009) idxstats. An endogenous based score was used to assess the potential of the sample

793   being 'positive' for *Y. pestis*. It was calculated as follows:

794

$$\frac{(YPS - \max(YS))}{M}$$

795 where YPS is the number of reads specifically mapping to *Y. pestis*; YS is the maximum number

796 of reads mapping specifically to a *Yersinia* species with the exception of *Y. pestis* and M is the

797 total number of merged reads in the sample. By using the maximum number of reads mapping

798 to another species of the genus *Yersinia*, the score takes in account different source of

799 contamination other than *Y. pseudotuberculosis.* Five samples (RK1001, Gyvakarai1, KunilaII,

800 6Post and 1343UnTal85) fulfilled the criteria for being considered strong candidates (score

801 higher than 0.005 and reads mapping to all plasmids). Another samples, GEN72, was also

802 included in further processing and analysis since it had higher numbers mapping the *Y. pestis*

803 chromosome and plasmids even though it did not full-fill the score requirements. For a detailed

804 description of the archaeological sites and individuals see the SI.

Deep shotgun sequencing

806 The five strong candidate samples detected in screening of the shotgun data were processed

807 for deep shotgun sequencing as following: For Gyvakarai1 the screening library described

808 above was pair-end sequenced on two lanes of a HiSeq4000 for 100 cycles, and on a full run of

809 a NextSeq500 for 75 cycles. The screening library for KunilaII was pair-end sequenced deeper

810 on 80% of one lane of a HiSeq4000 for 100 cycles. Additionally, 40 µl of DNA extract of KunilaII

811 was converted in to a library treated with UDG and endonuclease VIII to remove deaminated

812 bases (Briggs and Heyn, 2012), and pair-end sequenced on one lane of a HiSeq4000 for 75

813 cycles.

814 For RK1001, Post6 and 1343UnTal85, 60 µl of DNA extract each were converted into DNA

815 libraries using so-called UDG-half treatment, whereby deaminated bases are partially removed

816 and retained mostly at the ends of the molecule (Rohland et al., 2015). The library of RK1001

817 was deep shotgun pair-end sequenced in 8 lanes of a HiSeq4000 for 55 cycles. The libraries of

818 6Post and 1343UnTal85 were deep shotgun single-end sequenced on 2 and a half lanes of a

819 HiSeq4000 for 75 cycles. Post6 was additionally pair-end sequenced on a full run of a

820   NextSeq500 for 75 cycles.

Y.pestis in-solution capture

822   *Y. pestis* whole-genome DNA capture probes were designed using as template sequences the

823   *Y. pestis* CO92 chromosome (NC_003143.1), *Y. pestis* CO92 plasmid pMT1 (NC_003134.1), *Y.*

824   *pestis* CO92 plasmid pCD1 (NC_003131.1), *Y. pestis* KIM 10 chromosome (NC_004088.1), *Y.*

825   *pestis* Pestoides F chromosome (NC_009381.1) and *Y. pseudotuberculosis* IP 32953

826   chromosome (NC_006155.1). We used a 6 bp tiling with a probe length of 52 bp with an

827   additional 8 bp 3' linker sequence as described in (Fu et al., 2013). Low complexity regions

828   were masked using dustmasker (Camacho et al., 2009, version 2.2.32+). Redundant probes as

829   well as probes with more than 20% masked nucleotides were discarded. The procedure

830   resulted in 816,413 unique probe sequences. A second probe set was created with a coordinate

831   offset of 3 bp resulting in 827,438 unique probe sequences. In combination the two probe sets

832   represent an effective tiling density of 3 bp. The two probe sets were ordered on two 1 million

833   feature Agilent SureSelect DNA Capture Arrays. The full capacity of the arrays was filled up with

834   randomly selected probes. The two arrays were turned into in-solution DNA capture libraries as

835   described elsewhere (Fu et al., 2013).

836   For GEN72, 25 µl of DNA extract was converted into DNA libraries using so-called UDG-half

837   treatment as described above[10]. The UDG-half libraries of RK1001 and GEN72 were enriched

838   for *Y. pestis* DNA using in-solution DNA capture probes (see above) as described elsewhere

839   (Fu et al., 2013; Haak et al., 2015; Mathieson et al., 2015). The capture products of RK1001

840   and GEN72, were sequenced on 1 and 0.6 of the lane, respectevely, of the HiSeq4000 for 75

841   cycles.

Genome reconstruction and authentication

All samples were processed with the EAGER pipeline (Peltzer et al., 2016). Sequencing quality for each sample was evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and adaptors clipped using the ClipAndMerge module in EAGER. For paired-end data, the reads were also merged with ClipAndMerge and only the merged reads were kept for further analysis.

Due to variability in the laboratory preparation and sequencing strategies, the sequencing reads for each sample were treated as follows:

- Gyvakarai1: two HiSeq lanes and one Next-Seq run paired-end of the non-UDG treated library were combined and reads mapped to *Y. pestis* CO92 reference with BWA aln (-l 16, -n 0.01, hereby referred to as non-UDG parameters). Reads with mapping quality scores lower than 37 were filtered out. PCR duplicates were removed with MarkDuplicates. MapDamage (Jónsson et al., 2013, v2.0) was used to calculate damage plots. Coverage was calculated with Qualimap (Okonechnikov et al., 2016, v2.2).

- KunilaII: UDG and the non-UDG libraries were sequenced in 2 HiSeq pair-end lanes and processed separately until calculation of the coverage. The non-UDG treated libraries were mapped with non-UDG parameters while the UDG treated library reads were mapped with more stringent parameters (-l 32, -n 0.1, referred to as UDG parameters). Reads with mapping qualities less than 37 were filtered out and duplicates were removed with MarkDuplicates as before. The non-UDG bam file was used to calculate damage plots as indicated above. After duplicate removal, the UDG- and non-UDG treated BAM files were merged together and used to calculate the coverage as above.

- GEN72, Post6 and 1343UnTal85: the UDG-half treated libraries were sequenced in two HiSeq lanes for Post6 and 1343UnTal85 and 19,777,683 reads were generated in the HiSeq for GEN72, and two different runs were performed. For the first run, reads without

37

867   clipping were used to retain miscoding lesions indicative of aDNA. BWA aln was used for

868   mapping with non-UDG parameters (-l 16 and -n 0.01). Reads with mapping qualities

869   lower than 37 were filtered and PCR duplicates were removed with MarkDuplicates as

870   described above. Coverage and damage plots were calculated as above. After clipping

871   the last two bases with the module ClipAndMerge in eager, potentially affected by

872   damage, the samples were mapped with UDG parameters.

873   •   RK1001: UDG-half library was shotgun sequenced pair-end in 8 HiSeq lanes and in-

874   solution captured and sequenced single end to a depth of 303,148,884 reads sequenced

875   in the HiSeq. Shotgun and captured data were combined in a fastq file and processed as

876   described above for GEN72, Post6 and 1343UnTal85.

877   SNP calling & phylogenetic analysis

878 Prior to SNP calling in order to avoid false SNP calling due to aDNA damage, the quality scores

879 of damaged sites in the non-UDG treated samples were downscaled using MapDamage

880 (Jónsson et al., 2013, v2.0), as performed in previous analysis (Rasmussen et al., 2015). For

881 the UDG-half data, the files with the two last bases clipped, hence removing the damage signal,

882 and mapped with UDG parameters were used for SNP calling (see above).

883 SNP calling was performed with GATK UnifiedGenotyper (Van der Auwera et al., 2013) in

884 EAGER[43] with default parameters and the 'EMIT_ALL_SITES' output mode.

885 VCF files of the new ancient samples, along with the two complete genomes from Rasmussen

886 et al., 2015, the Black Death (Bos et al., 2011), Justinianic Plague (Feldman et al., 2016),

887 Bolgar (Spyrou et al., 2016) and Observance (Bos et al., 2016) genomes, were combined with a

888 curated dataset of 130 modern genomes (Cui et al., 2013) in addition to 11 samples from the

889 Former Soviet Union (Zhgenti et al., 2015) and 19 draft genomes of *Y. pestis* subsp. *microtus*

890 strains (Kislichkina et al., 2015).

891   The VCF files were processed with an in-house program (*MultiVCFAnalyser*) that

892   produced a SNP table and an alignment file containing all variable positions in the dataset, in

893   respect to the reference *Y. pestis* CO92. In order to call a SNP a minimum genotyping quality

894   (GATK) of 30 was required, with a minimum coverage of 3X, and with a minimal allele frequency

895   of 90% for a homozygous call. No heterozygous calls were included in the output files.

896   The SNP alignment was curated by removing all alignment columns with missing data

897   (complete deletion). The curated SNP alignment was then used to compute NJ and MP trees

898   with MEGA6 (Tamura et al., 2013) and a ML tree using PhyML 3.0 (Guindon et al., 2010) with

899   the GTR model used in previous *Y. pestis* work (Cui et al., 2013; Rasmussen et al., 2015), with

900   4 gamma categories and the best of NNI and SPR as tree branch optimization. The specific

901   variants of *Y. pseudotuberculosis* were removed from the analysis to improve the visual

902   resolution of the tree.

903   # Dating analysis

904   The SNP alignment after complete deletion was used for molecular dating using BEAST 1.8.2

905   (Drummond et al., 2012). The modern sample 0.PE3, also called Angola, was removed from the

906   dataset due to its long branch.

907   For tip dating, all modern genomes were set to an age of 0. The dates of the ancient

908   samples presented in this study plus the two complete genomes from Rasmussen et al., 2015

909   were recalibrated with Calib 7.1 (http://calib.qub.ac.uk/calib/) to the IntCal13 calibration curve.

910   The ancient samples were given the median calibrated probability as their age, and the 2 sigma

911   interval was used as the boundaries for a uniform prior sampling (Supplementary Table 5). The

912   dates published for previous historical genomes were transformed to cal BP assuming 1950 as

913   age 0 and given the mean as the age with the interval as the boundaries of a prior uniform

914   distribution: Black Death 603 (602-604, Bos et al., 2011); Observance 229 (228-230, Bos et al.,

915   2016), Bolgar 569 (550-588, Spyrou et al., 2016) and Justinian 1453 (1382-1524, Feldman et

916   al., 2016).

917   The molecular clock was tested and rejected using MEGA6. Therefore, we followed

918   previous work and used an uncorrelated relaxed clock with lognormal distribution (Cui et al.,

919   2013; Rasmussen et al., 2015) with the substitution model GTR+G4. Tree model was set up to

920   coalesce assuming a constant population size and a rooted ML tree was provided as a starting

921   tree. Two independent 1,000,000,000 MCMC chains were computed sampling every 5,000

922   steps. The two chains were then combined using LogCombiner from BEAST 1.8.2 (Drummond

923   et al., 2012) with a 10 percent burn-in (100,000,000 steps per chain). The ESS of the posterior,

924   prior, treeModel.rootHeight, tMRCA_allpestis are 4,589, 4,087, 1,054 and 7,571 respectively.

925   The trees files for the 2 chains were combined with LogCombiner with 100,000,000 of burning

926   and resampled every 20,000 steps giving a total number of 90,000 trees, that were used to

927   produce a Maximum Clade Credibility tree using TreeAnnotator from BEAST 1.8.2 (Drummond

928   et al., 2012).

929   # SNP effect analysis and virulence factors analysis

930   The SNP table from *MultiVCFAnalyzer* was provided to *SnpEff* (Cingolani et al., 2012) and the

931   effect of the SNPs within genes present in the dataset was evaluated. Additionally the SNP

932   table was manually assessed for possible homoplasies.

933   For the virulence factors, the samples were mapped as indicated above but without

934   applying quality filtering and the percentage of coverage was calculated for each region using

935   bedtools (Quinlan and Hall, 2010) and plotted using the package ggplot2 (Wickham, 2009) in R

936   (R Development Core Team, 2008). Additionally, *ureD* was manually explored for SNPs using

937   IGV (Thorvaldsdóttir et al., 2013).

938 ## Indel analysis

939 The samples including the two complete Bronze Age genomes (Rasmussen et al., 2015) were

940 mapped against *Y. pseudotuberculosis* IP 32953 with bwa with non-UDG parameters (-n 0.01, -l

941 16), except for RK1001, GEN72, 1343UnTal85 and 6Post that were mapped with bwa with UDG

942 parameters (-n 0.1, -l 32),. The modern genomes from branch 0 (0.PE7, 0.PE2, 0.PE3 and

943 0.PE4), *Y. pestis* CO92 and *Y. pestis* KIM10 were *in-silico* cut in 100 bp fragments with 1bp

944 tiling and mapped to *Y. pseudotuberculosis* reference using bwa with UDG parameters (-n 0.1, -

945 l 32). The non-covered regions were extracted using the bedtools genomecov function. Missing

946 regions larger than 1kb were comparatively explored in order to identify indels. Using the

947 bedtools intersect function, we extracted regions missing in the Neolithic genomes and present

948 in the modern ones and also the regions missing in the modern ones but still present in the

949 Neolithic genomes. The results were check by manual inspection in IGV (Thorvaldsdóttir et al.,
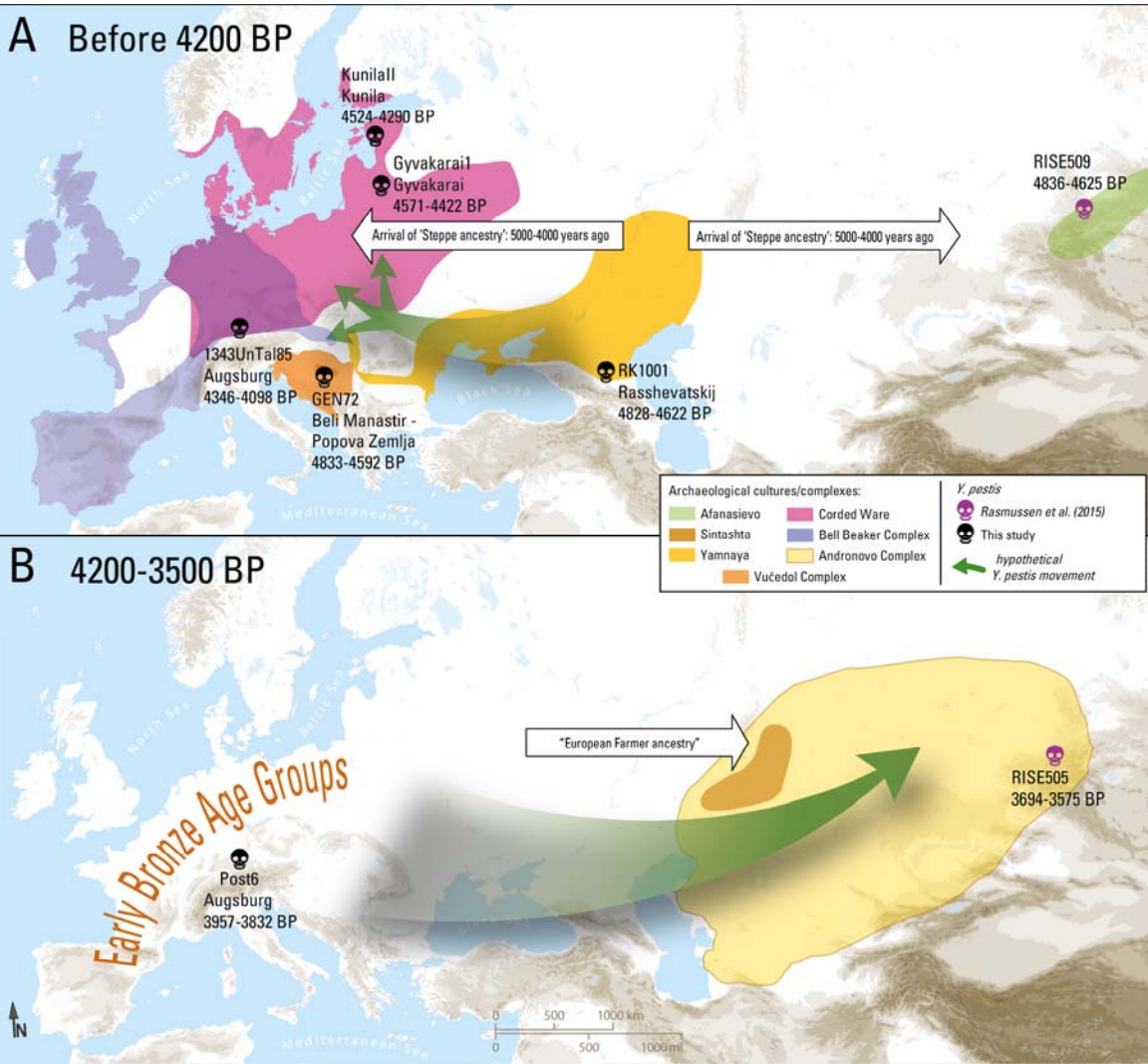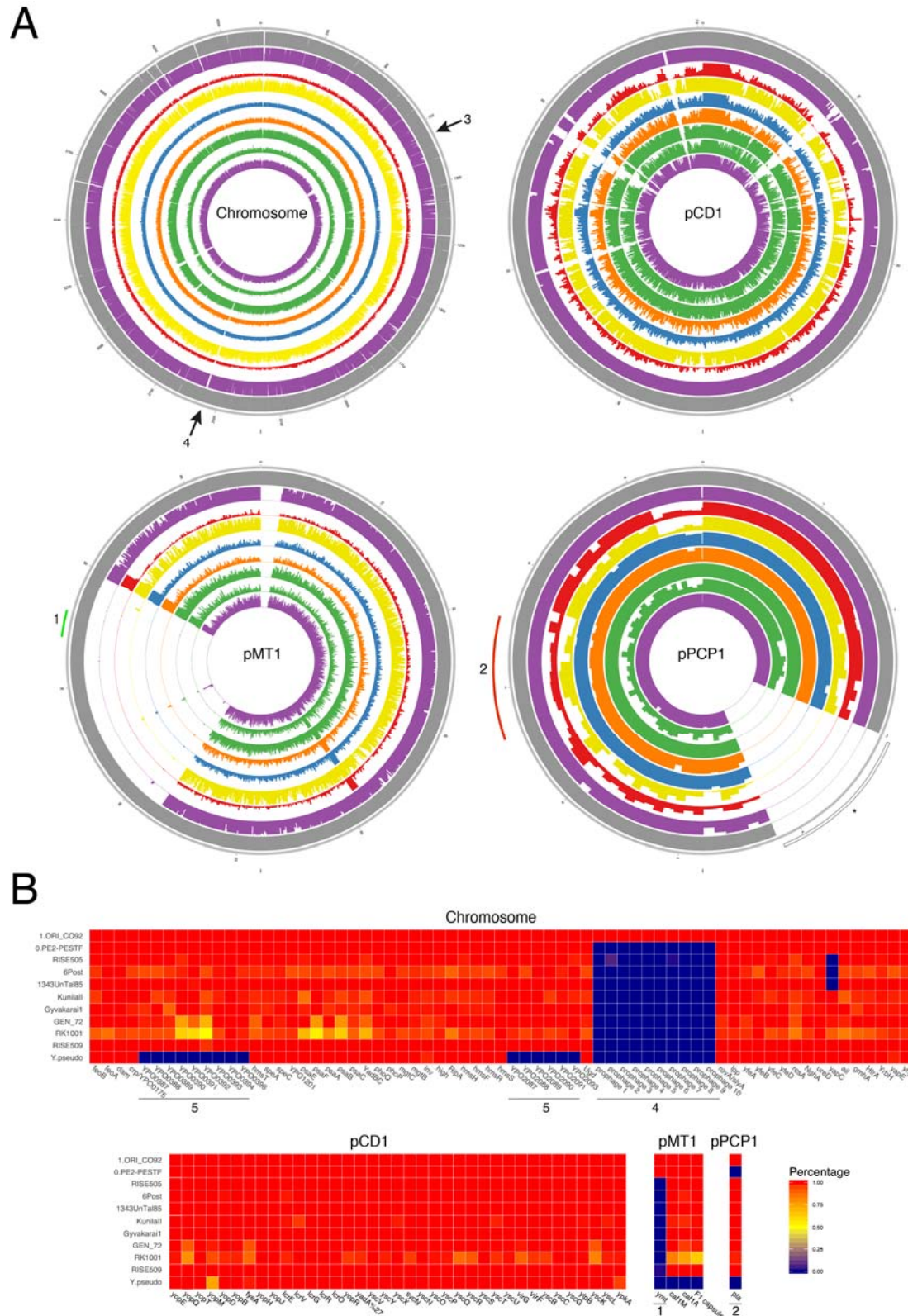
950 2013).

951 # Figures and Tables

952

953 Figure 1

954

955

956

957   Figure 2

958

959

960

961
962   Figure 3

963

964

965

966 **Table 1**

Species name	Strain	NCBI Accession number
Y. pestis	CO92	NC_003143.1
Y. pseudotuberculosis	IP 32953	NC_006155.1
Y. enterocolitca	subsp. enterocolitica 8081	NC_008800.1
Y. aldovae	ATCC 35236	NZ_ACCB01000210.1
Y. bercovieri	ATCC 43970	NZ_AALC02000229.1
Y. frederiksenii	ATCC 33641	NZ_AALE02000161.1
Y. intermedia	ATCC 29909	NZ_AALF02000123.1
Y. kristensenii	ATCC 33638	NZ_ACCA01000153.1
Y. mollaretii	ATCC 43969	NZ_AALD02000179.1
Y. rohdei	ATCC 43380	NZ_ACCD01000141.1
Y. ruckeri	ATCC 29473	NZ_ACCC01000174.1

967

968 **Table 2**

Individual	Tissue sampled	Site	Country	Dating (Median cal BP)	In-solution Capture	Clipped, merged and quality-filtered reads before mapping	Unique reads mapping to *Y. pestis* reference	Endogenous DNA (%)	Mean Coverage	Coverage (%)				
										>=1X	>=2X	>=3X	>=4X	>=5X
RK1001	Tooth	Rasshevatskiy	Russia	4720	no	1,529,935,532	119,540	0.01	1.0213	58.11	27.24	10.87	3.90	
					yes	303,148,884	383,900	0.85	3.3984	82.83	69.42	55.3	41.95	
					Combined shotgut/capture	1,833,084,416	418,581	0.17	3.6816	86.16	73.67	59.63	45.88	
GEN72	Tooth	Beli Manastir-Popova Zemlja	Croatia	4721	yes	19,777,683	1,321,320	24.36	12.6549	91.65	89.15	86.61	83.84	80.9
Gyvakarai1	Tooth	Gyvakarai	Lithuania	4427	no	1,021,452,137	473,207	0.05	5.2245	94.07	90.96	84.12	73.1	59
KunilaII	Tooth	Kunila	Estonia	4203	no	379,155,741	546,243	0.16	5.5418	92.48	86.65	77.58	66.49	54.7
1343UnTal85	Tooth	Augsburg	Germany	3873	no	1,174,989,269	1,165,435	0.14	10,5745	93.69	93.29	92.59	91.29	89.0
6Post	Tooth	Augsburg	Germany	3635	no	419,717,299	598,030	0.17	5.3062	89.71	81.66	71.14	59.82	48.8

969