

# An 8000 years old genome reveals the Neolithic origin of the zoonosis *Brucella melitensis*

Received: 14 May 2024

Louis L'Hôte<sup>1</sup>, Ian Light<sup>1</sup>, Valeria Mattiangeli<sup>1</sup>, Matthew D. Teasdale<sup>1,3</sup>, Áine Halpin<sup>1</sup>, Lionel Gourichon<sup>1</sup>, Felix M. Key<sup>1</sup> & Kevin G. Daly<sup>1,5</sup> 

Accepted: 15 July 2024

Published online: 20 July 2024

 Check for updates

*Brucella melitensis* is a major livestock bacterial pathogen and zoonosis, causing disease and infection-related abortions in small ruminants and humans. A considerable burden to animal-based economies today, the presence of *Brucella* in Neolithic pastoral communities has been hypothesised but we lack direct genomic evidence thus far. We report a 3.45X *B. melitensis* genome preserved in an ~8000 year old sheep specimen from Menteşe Höyük, Northwest Türkiye, demonstrating that the pathogen had evolved and was circulating in Neolithic livestock. The genome is basal with respect to all known *B. melitensis* and allows the calibration of the *B. melitensis* speciation time from the primarily cattle-infecting *B. abortus* to approximately 9800 years Before Present (BP), coinciding with a period of consolidation and dispersal of livestock economies. We use the basal genome to timestamp evolutionary events in *B. melitensis*, including pseudogenization events linked to erythritol response, the supposed determinant of the pathogen's placental tropism in goats and sheep. Our data suggest that the development of herd management and multi-species livestock economies in the 11<sup>th</sup>–9<sup>th</sup> millennium BP drove speciation and host adaptation of this zoonotic pathogen.

Brucellosis is the most prevalent bacterial zoonosis globally<sup>1</sup> infecting 1.6–2.1 million people annually, and is caused by members of the non-motile *Brucella* genus. *Brucella* species are typically characterised by their mammalian host range and zoonotic potential, with four species considered to infect humans<sup>2</sup>. Of these, *Brucella melitensis* is the most frequent cause of brucellosis in humans and has a host reservoir of goats and sheep. The remaining three species linked to zoonotic infection are also domestic animal-associated: *B. abortus*, closely related to *B. melitensis* and primarily infecting cattle; *B. suis*, infecting pigs; and *B. canis*, dogs. These and other *Brucella* species comprise a “core” clade of high genomic similarity, also known as “classic” *Brucella* species, distinguished from the genetically distinct non-core or “atypical” *Brucella* which are rarely associated with zoonotic infection<sup>2</sup>. Brucellosis has higher incidence among cohorts frequently in contact

with livestock such as pastoralist communities<sup>3</sup>, and in regions most dependent on small ruminant husbandry<sup>1</sup>.

Considered intracellular pathogens, *Brucella* species evade and suppress the animal host immune system to establish replicative niches within cells such as macrophages<sup>4</sup>. Following systemic distribution, *Brucella* show tropism towards the mammary gland, reproductive organs and, in females, the placental trophoblasts. Placental invasion is followed by massive bacterial replication<sup>5</sup> and disruption of the placenta in pregnant animals, leading to spontaneous abortion. The placenta, foetal remains, and vaginal discharges are vectors of transmission, as is unpasteurised milk of infected animals<sup>6</sup>. Additional symptoms of *B. melitensis* or *B. abortus* livestock infection include weak offspring and low milk yield, contributing to a substantial economic and animal welfare burden. Brucellosis in humans is typically

<sup>1</sup>Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland. <sup>2</sup>Max Planck Institute for Infection Biology, Evolutionary Pathogenomics, 10117 Berlin, Germany. <sup>3</sup>Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. <sup>4</sup>Université Côte d'Azur, CNRS, CEPAM, Nice, France. <sup>5</sup>School of Agriculture and Food Science, University College Dublin, Dublin 4, Ireland.  e-mail: [kevin.daly@ucd.ie](mailto:kevin.daly@ucd.ie)

due to *B. melitensis* infection and has low mortality, causing a range of symptoms including undulating fever, general malaise, and osteoarthritis<sup>7</sup>. However, infection in pregnant women is associated with 18.6–73.3% risk of miscarriage and an increased risk of preterm birth<sup>8</sup>.

The intensification of sheep and goat management in 11th–10th millennium Before Present (BP) Southwest Asia, has been hypothesised as a fulcrum for *Brucella* zoonotic and epizootic transmission<sup>9</sup> (the latter the spread of disease from animal to animal). High incidence of caprine foetal and perinatal remains are found in Neolithic settlements across Southwest Asia including at Ganj Dareh in the Iranian Zagros<sup>10</sup>, where evidence for goat management in the 11th millennium BP is robust<sup>11</sup>. As infected females represent the main transmission source<sup>6</sup>, selective culling of young male goats practised in early herding communities could have created conditions permissive for sustained *Brucella* enzooticity<sup>9</sup>, potentiating human-spillover events. Skeletal remains of these early herders are suggestive of *Brucella* exposure: they feature a high incidence (>80%) of cranial porotic hyperostosis and a single case of vertebrate lytic lesions<sup>12</sup>, potential consequences of chronic brucellosis. Skeletal evidence of human and equine brucellosis (see examples<sup>13,14</sup>) is better attested in more recent contexts, with several cases of possible genetic confirmation (see refs. 15,16).

Genomic recovery of ancient *Brucella* species is inherently challenging, with just two genomes from the mediaeval period reported<sup>15,16</sup>. *Brucella* can persist in the environment for ~1.5 months<sup>17</sup> but lack features which facilitate macromolecular preservation such as capsules, thick cell walls or endospores. Unlike their soil and water-living *Ochrobactrum* relatives, *Brucella* appear to require an intracellular environment to replicate<sup>18</sup>. The localization of *Brucella* during infection is also not amenable to ancient DNA survival due to its greater localization in the reproductive organs, mammary glands, spleen, liver, and bone marrow<sup>19</sup>, tissues rarely preserved for palaeogenomic study. Finally the core *Brucella* species are highly similar at a genomic level with an average nucleotide identity (ANI) of >99.5%, in contrast to greater divergence from atypical non-core species (ANI with *B. melitensis*: ~97.7%) and the soil-ubiquitous sister genus *Ochrobactrum* (~81% ANI)<sup>20</sup>. Together, these complicate the retrieval and validation of ancient, intracellular *Brucella* genomes.

Despite these challenges, we report and validate a 3.45X *B. melitensis* genome recovered from a sheep specimen dating to ~8000 BP, from Menteşe Höyük in Northwest Türkiye, the oldest reported livestock pathogen and zoonotic genome to date.

## Results

### Detection and authentication of Neolithic *Brucella*

We screened for core *Brucella* species published ancient ruminant genomics datasets (representing 146 specimens, see Methods) and six newly-reported ovicaprine petrosal specimens (Supplementary Data 1) from Menteşe Höyük (40.278013°N, 29.523417°E; hereafter referred to as Menteşe) using a *k*-mer matching approach<sup>21</sup> correcting for read count and coverage<sup>22</sup>. A single female sheep specimen from Menteşe (Supplementary Figs. 1–3) produced a signal for *B. melitensis* (Fig. 1a), with 71 assigned reads representing a diversity of unique genome regions (*E* value = 0.00023; Supplementary Data 2). We generated two radiocarbon age estimates for the sample (Supplementary Data 3), and used Oxcal4.4<sup>23</sup> to produce a combined calibrated age of 8007–7863 cal BP (2 $\sigma$  range) (Supplementary Fig. 4), securely within the upper layer of Menteşe's Neolithic inhabitation.

We performed additional shotgun sequencing for Menteşe6 from libraries treated with Uracil-DNA Glycosylase (UDG) to remove post-mortem damage, and also libraries lacking UDG treatment, thus preserving native damage patterns (Supplementary Data 4). We then assessed the authenticity of the *Brucella* reads using HOPS<sup>24</sup>, an ancient DNA pipeline which utilises the alignment-based metagenomic

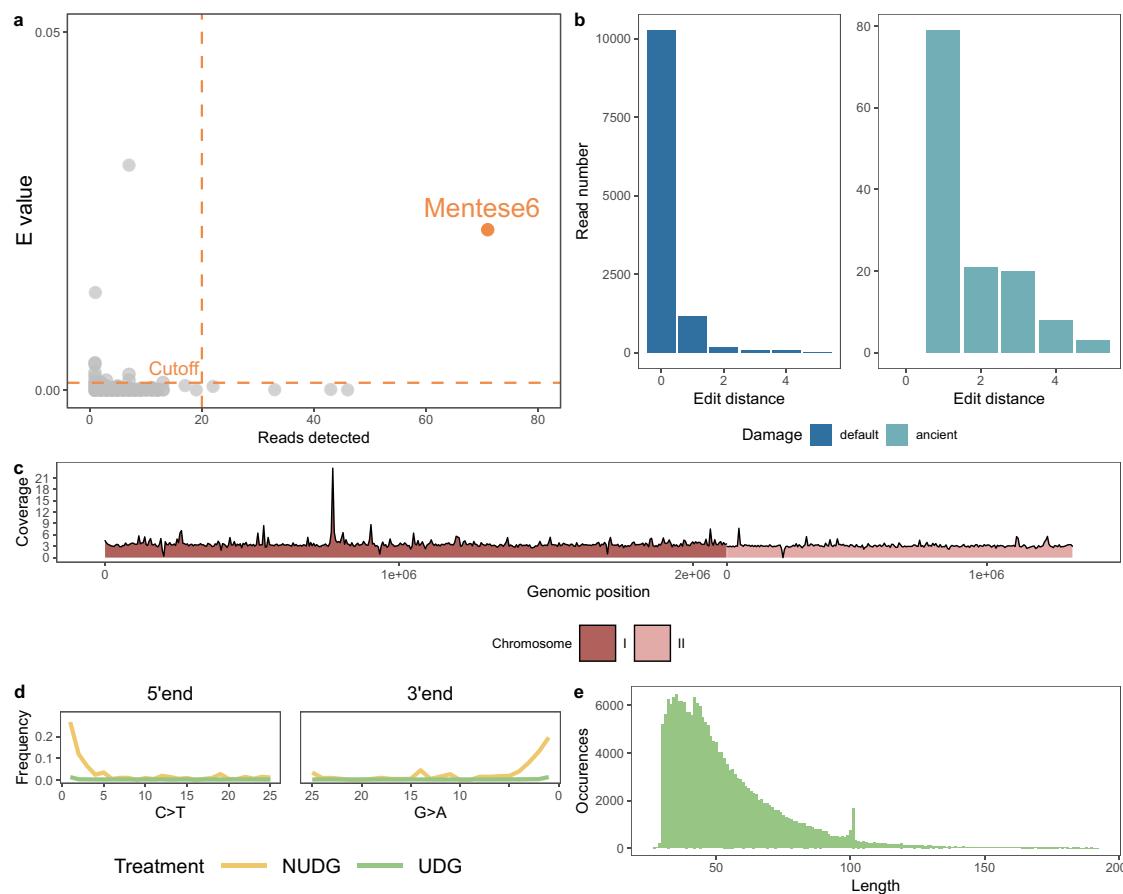
assigner MALT<sup>25</sup>. Using a subset of sequencing data with sheep-aligning reads removed (9,739,417 reads; see Methods), MALT assigned 20,040 UDG-treated unique reads to the *Brucella* node, with a *B. melitensis* reference showing the highest alignment rate (Supplementary Fig. 5). Over 86% of aligning reads showed zero mismatches (Fig. 1b), suggesting correct assignment to the *Brucella* node rather than spurious alignments from a related genus. A similar high alignment accuracy is evident when restricted to reads with persisting post-mortem molecular damage (Supplementary Fig. 5), supporting an authentic ancient origin of the *Brucella*-assigned reads. Together, these suggest the Menteşe6 genome represents a Neolithic *B. melitensis*.

Aligning UDG-repaired reads to the *B. melitensis* reference genome (GCF\_000007125.1) produced a 3.45X alignment (Table 1) after removal of sheep-aligned reads and other quality controls (see Methods). Approximately 94% of the *B. melitensis* genome was covered by at least one read (the 1X genome breadth; Fig. 1c), further supporting this as a genuine *B. melitensis* genome rather than a false positive signal due to alignment from related environmental species. Such spurious alignment can result in “read stacking” at homologous genome regions, leading to poor genome breadth relative to the average depth of coverage across the reference. Nevertheless, we observed some limited regions of stacking in the Menteşe6 alignment, which are filtered from downstream analyses through a per-site read maximum and sliding window coverage filter (see Methods). Lastly, aligning reads were short (Fig. 1e, mean 54 bp) and showed terminal transitions damage in both UDG repaired and non-repaired libraries (Fig. 1d), consistent with authentic ancient *Brucella* DNA recovery.

### Menteşe6 *Brucella* genome is basal to all known *Brucella melitensis*

We evaluated the phylogenetic position of the Menteşe6 *Brucella* genome using several complementary approaches. A maximum likelihood (ML) phylogeny of *B. melitensis*, representing the clades observed today<sup>26</sup> (Supplementary Data 5), and the cattle-associated sister species *B. abortus*, aligned to and rooted on the outgroup *B. suis* places the Menteşe6 *Brucella* as a basal member of the *B. melitensis* clade (Fig. 2a). The genome falls close to the ancestral node of *B. melitensis* and *B. abortus*, suggesting it existed relatively soon after separation of the two lineages and their speciation.

The genomes of *B. melitensis* and *B. abortus* are highly similar (pairwise >99% ANI<sup>20</sup>), and there remains a potential effect of soil-living *Ochrobactrum* reads in the Menteşe6 alignment. To assess this and confirm the *B. melitensis* species assignment we determined variant sites which discriminate specific lineages or branches within the *Brucella* genus (Supplementary Data 6), where all other genus members carry the alternative allele (Supplementary Fig. 6), using data aligned to *B. melitensis*. Across these clade-defining sites and applying strict filtering (see Methods), the Menteşe6 base matched 99.7% of variants fixed in and common to *B. melitensis* and *B. abortus*, and 28.1% variants defining the *B. melitensis* branch (Supplementary Fig. 7, Supplementary Data 7). In comparison, Menteşe6 matched 0% variants specific to *B. abortus*, 0% variants diagnostic of *B. canis*, and 0.001% of *B. inopinata* variants. We obtain similar results when aligning to a reference genome of a species (*B. suis*) equally related to *B. melitensis* and *B. abortus*, or using random read sampling rather than a consensus approach to assess possible heteroplasmy (Supplementary Data 7). Finally, ANI values<sup>27</sup> (Fig. 2b, Supplementary Data 8) indicate that the Menteşe6 genome has greater mean similarity with *B. melitensis* than *B. abortus* (mean 99.870 vs 99.856; Wilcox rank sum test *p*-value < 2.2 × 10<sup>-16</sup>). We therefore conclude that the Menteşe6 genome is a basal representative of *B. melitensis* existing soon after the species split from *B. abortus*, before it accrued many of the mutations fixed in *B. melitensis* today.



**Fig. 1 | Authentication of ancient *Brucella* DNA in Neolithic sheep specimen**

**Mentes6.** **a** kmer-based *E* value<sup>22</sup> relative to the number of reads detected for core *Brucella* species by KrakenUniq. Highlighted in orange is the screened library (from Mentes6) passing the threshold of > 20 reads and an *E* value > 0.001, indicated by the cutoff dashed lines. **b** HOPS edit distance for Mentes6 Uracil DNA-glycosylase (UDG) treated for reads assigned to the *Brucella* node, merged from seven single-

end libraries. Dark blue represents all reads (“default”) and light blue reads showing post-mortem damage (“ancient”). **c** Non-overlapping sliding windows (5 kb) of coverage of Mentes6 UDG libraries aligned on *B. melitensis*. **d** mapDamage<sup>23</sup> damage patterns for Mentes6 reads aligned to *B. melitensis*. Green represents UDG libraries and yellow represents non-UDG libraries (NUDG). **e** Read length distribution of the Mentes6 UDG libraries aligned on *B. melitensis*.

**Table 1 | Summary of Mentes6 alignment to *Brucella melitensis* genome from UDG-treated sequencing libraries**

Sample ID	Reads aligned	Alignment rate (%)	Coverage	1X Breadth	3X Breadth	Combined calibrated age (2σ)
Mentes6	212,160	0.009	3.453X	0.939	0.600	8007–7863 cal BP

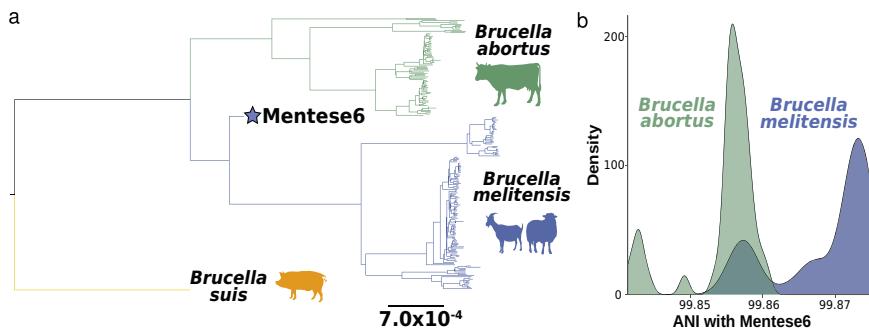
Values are after removal of duplicate reads and mapping quality 30 filter. 1X and 3X breadth are defined as the proportion of the genome covered by at least 1 or 3 reads respectively.

### Calibrating the *Brucella melitensis* speciation time

We leverage this ~8000 year old genome to refine the speciation time of *B. melitensis* and the primarily cattle-infecting *B. abortus*. Time-calibrated phylogenetic trees are premised on the consistent accrual of mutations through time, creating a relationship between genome sampling time and genetic change: this permits the estimation of timing of individual nodes within a phylogenetic tree. Similar to a previous study<sup>16</sup>, we find weak evidence for such a temporal signal<sup>28</sup> in both species (Supplementary Fig. 8). This discrepancy in sampling age of modern *Brucella* genomes and their measured genetic differentiation appears to be driven by clade-specific differences in mutation rate, indicated by clustering of lineages by their root-to-tip distance values. This is further illustrated by inflated ANI values of the *B. melitensis* Western Mediterranean lineage with the Mentes6 genome (Supplementary Fig. 8), leading to Mentes6 being relatively closer to some *B. abortus* sequences (Fig. 2b). Despite this apparent rate heterogeneity, marginal likelihoods supported a strict clock mutation model over a relaxed model which would allow rate to vary

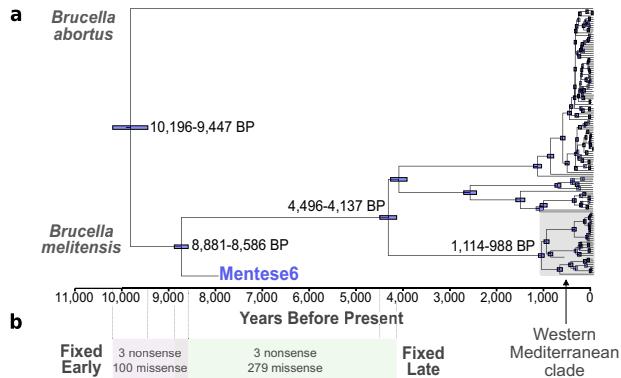
among branches (Supplementary Data 9). We proceeded with Bayesian time tree computation<sup>29</sup> using a data set of *B. melitensis* and a single *B. abortus* as an outgroup, and achieved good model stabilization (all parameter Effective Sample Sizes  $\geq 1260$ ; Supplementary Data 9). Our estimate of the *B. melitensis*-*B. abortus* speciation time is 9816 BP (95% Highest Posterior Density (HPD): 10,196–9447 BP; Fig. 3a), a ~750 year interval within the Neolithic era of Southwest Asia. This period overlaps with the emergence of herding<sup>30</sup>, and subsequent consolidation and dispersal of multi-species livestock economies in the region following more than a millennia of animal management experimentation<sup>31</sup>.

We infer a common ancestor time of *B. melitensis* today of 4317 BP (4496–4137 BP), within the Bronze Age of Western Eurasia, relatively recent compared to the beginnings of livestock domestication in the 11th millennium BP. A potential explanation is demographic processes in either of its main hosts, sheep and goats: *Brucella* is an intracellular, monoclonal genus, with a closed genome showing limited evidence of recombination<sup>32</sup> susceptible to host-mediated bottlenecks<sup>33</sup>.



**Fig. 2 | Genome based assignment of Mentese6 to *Brucella melitensis*.** **a** ML phylogeny (IQ-TREE2<sup>67</sup>) for *B. melitensis* and *B. abortus* aligned to outgroup *B. suis*. The blue star indicates Mentese6. The typical animal hosts of each lineage are shown in silhouette. **b** Distributions of Average Nucleotide Identity (ANI) between

Mentese6 and *B. melitensis* ( $n=144$ ) or *B. abortus* ( $n=83$ ). *B. melitensis* genomes with lower ANI with Mentese6 (<99.86%) derive from the Western Mediterranean clade, which shows elevated root-to-tip distances (Supplementary Fig. 8) indicative of an elevated mutation rate within that lineage.



**Fig. 3 | Time-calibrated phylogeny and gene evolution of *Brucella melitensis*.** **a** Beast<sup>29</sup> phylogeny using a strict molecular clock and Coalescent Bayesian Skyline tree prior. The 95% Highest Posterior Distribution (HPD) are indicated blue, and annotated if mentioned in the text. The *B. abortus* reference genome was included to estimate the *B. melitensis*-*B. abortus* speciation time. Samples were aligned to the *B. melitensis* reference genome. x-axis is Years Before Present (BP - relative to the year 1950). **b** Missense and nonsense mutations fixed in *B. melitensis* today are either early evolving or late evolving, occurring before or after the divergence of the Mentese6 Neolithic genome from the ancestor of modern *B. melitensis*. Dashed lines indicate the extent of the early or late evolving epochs, based on the 95% HPD of node ages from the Beast phylogeny.

This dating predates the occurrence of sheep wool use in Europe<sup>34</sup> and thus could be driven by demographic processes in the host ruminants, although other explanations are possible. Selection may have acted on beneficial mutations and lead to clonal expansion, a selective sweep<sup>35</sup>, explaining the relatively recent common ancestor of modern *B. melitensis*. We additionally infer a narrow diversification time (1114–988 BP; 836–962 Common Era) of the Western Mediterranean clade, basal to other modern *B. melitensis* lineages<sup>16</sup>. Our mutation rate estimate of  $8.418 \times 10^{-8}$  (95% HPD:  $8.770 \times 10^{-8}$ – $8.074 \times 10^{-8}$ ) is a more precise refinement of previous estimates for *B. melitensis*<sup>16</sup>.

#### Time-stamping functional evolution of *Brucella melitensis*

The ~8000 year old *B. melitensis* genome offers the opportunity to directly time-stamp key evolutionary steps of this zoonotic pathogen. We annotated single nucleotide variants fixed in modern *B. melitensis* relative to other *Brucella*, and identified 6 positions with nonsense mutations covered by at least one read in the Mentese6 Neolithic *Brucella* (Fig. 3b, Supplementary Data 10). At 3 of these 6 nonsense mutation sites, Mentese6 already carried the derived allele, while the remaining 3 lacked the stop codon-introducing allele. This establishes three pseudogenization events (evolutionary changes often associated

with host specificity<sup>32</sup>) before the divergence of Mentese6 from the ancestor of *B. melitensis* today ~8731 BP (95% HPD: 8881–8586 BP) but after the *B. melitensis*-*B. abortus* split (10,196–9447 BP).

The three pseudogenization events occurring prior to the split of Mentese6 from other *B. melitensis* have well defined molecular roles: *nosZ*, *iolG\_6* and *hipO*. Of these, *nosZ* is associated with the microbe's mode of living and pathogenesis. *nosZ* encodes nitrous-oxide reductase and is the final step in the denitrification cascade, a respiratory pathway employed under anaerobic conditions. *B. melitensis* is particularly limited in its capacity for anoxic growth compared to other *Brucella* species, a suspected consequence of the *nosZ* pseudogenization and resulting build-up of nitrous oxide and other toxic by-products of anoxic nitrogen reduction<sup>36</sup>. The intracellular environment and compartments have a lower oxygen concentration than the extracellular space<sup>37</sup>. Despite *B. melitensis*' reduced capacity for respiration within the cell, it, along with other core *Brucella* species, is traditionally considered to be a facultative intracellular parasite. *B. melitensis* preferentially replicates extracellularly in the presence of erythritol, a 4 C sugar to which the bacteria shows tropism<sup>38</sup>. Erythritol is produced by ruminant placental trophoblasts, and has been recently shown to be available in human trophoblast culture<sup>39</sup>, a potential mechanism of its placental colonisation in humans. The reduced ability of *B. melitensis* to grow under anaerobic conditions, fixed early in its evolution by the *nosZ* pseudogenization, may contribute to its pathogenesis by weighting replication towards the extracellular aerobic spaces when induced by placental erythritol.

For the remaining three pseudogenes (Supplementary Data 10) we observed the ancestral allele in the Mentese6 *Brucella* genome (i.e. the gene lacks the stop codon mutation), and thus postdate its divergence from the ancestor of *B. melitensis* today (95% HPD: 8881–8586 BP). Among these later-fixing nonsense mutations is one affecting BME\_RS15065 (previously BMEIII1002), truncating this hemerythrin domain-containing protein by 20 residues (137aa compared to 157aa in *B. abortus* 2308). The gene is predicted to have calcium/iron binding activity and is upregulated during growth in response to erythritol<sup>38</sup>; this pseudogenization may have become fixed in *B. melitensis* as a consequence of its erythritol tropism and induced extracellular replication. The remaining two nonsense mutations affect a gene encoding a hypothetical peptidoglycan-binding domain protein (BME\_RS05080) and a known glycosyltransferase pseudogene (BME\_RS17595). We additionally assess the Mentese6 genome's complement of infection-related genes based on gene-level coverage and breadth relative to the genome average (Supplementary Data 11). These suggest that the Neolithic *B. melitensis* had a sufficient repertoire of genes to establish chronic or congenital infection in its sheep host, enabling genome recovery from the Mentese6 temporal bone.

We similarly identified 379 fixed protein-altering missense mutations, 100 of which were present in the Neolithic Menteşe6 *Brucella* genome (Supplementary Data 10) and thus occurring early in *B. melitensis* evolution. Among these are three affecting infection-related genes *fliF*, *lptA*, and *cysB*. *FliF* is a flagellum-related pseudogene in *B. melitensis* today, but we lack reads overlapping the diagnostic non-sense mutation (II:164,232) to establish if this ring monomer protein gene was functional in Mentes6. *LptA* is involved in the modification of the “lipid A” cell envelope component, but this may not affect virulence<sup>40</sup>. Finally, *cysP* (BMEI0673) produces a sulfate/thiosulfate-binding protein and is regulated by the key infection-related operon *virB*<sup>41</sup>. We additionally find missense mutations in four erythritol (*ery*) operon genes, fixed in the Neolithic genome and *B. melitensis* today (Supplementary Data 10). Contrasting these are 279 missense mutations fixed later in *B. melitensis* evolution, affecting key genes involved in infection (*virB4*, *virB6*, *virB11*, *virB12*), intracellular defence evasion (*sodB*) and survival in the stomach’s low pH environment (*ureC*, *ureD\_2* and *ureF\_2*).

## Discussion

We report an approximately 8000 year old *B. melitensis* genome that confirms previous genomic<sup>16</sup>, palaeopathological<sup>12</sup>, and transmission model inferences<sup>9</sup> that livestock-infecting *Brucella* were present and sustained in Southwest Asian managed herds already during the Neolithic period. The Neolithic intensification of the human-animal relationship and shift towards pastoralism is considered the first epidemiological transition due to the increased opportunities for zoonotic transmission compared to hunter-gatherer societies<sup>42</sup>. This is supported by human-derived palaeomicrobial evidence<sup>43</sup>, although a broad ancient pathogen DNA screen suggests zoonosis frequency only accelerated during the last 7000 years<sup>44</sup>. Here we present direct evidence that livestock host-specific zoonotic pathogens evolved along with the development of livestock herding and multi-species animal-based economies 10,000-9000 years ago (Fig. 3a).

Menteşe is a Pottery Neolithic site located in the eastern Marmara region and was inhabited during the second half of the 9<sup>th</sup> millennium BP by farming communities primarily engaged in raising sheep and cattle, and to a lesser extent herding goats and pigs<sup>45</sup>. They further supplemented their diets through hunting wild mammals and birds, and gathering terrestrial and freshwater molluscs<sup>45</sup>. After an early focus on cattle, the proportion of sheep at Menteşe rose substantially, indicating small ruminants had a greater role in the food economy at the time when the Menteşe6 sheep lived. Mortality profiles, commonly used in zooarchaeology to infer livestock management strategies, indicate that sheep and cattle were exploited not only for meat but also for milk production. Dairy consumption at that time is also supported by lipid residue analysis of pottery sherds from neighbouring early Pottery Neolithic sites<sup>46</sup>, suggesting an emphasis across northwestern Anatolia on this practice. Sustained enzootic *Brucella* transmission is likely due to the social and economic interactions between western Anatolian communities<sup>47,48</sup>, which would have maintained reservoir populations<sup>9</sup>. These farming communities would have been at risk of brucellosis due to herd-circulating *B. melitensis*: human skeletons from Menteşe show a high frequency of osteoarthritis and anemia<sup>49</sup> indicative of infection, possibly brucellosis. While it is possible the *Brucella* genome recovered from this site was not infection-related, the genus is not considered to be free-living<sup>18</sup>. As the European Neolithic farming tradition stems from Anatolia<sup>30</sup>, *B. melitensis* likely also affected the first livestock of Europe.

We refine previous *B. melitensis* and *B. abortus* speciation time estimates<sup>16,50</sup> to a narrow ~750 year window (10,196–9447 BP) during the Neolithic era of Southwest Asia. Our inferred speciation time is consistent with recent estimates drawn from *B. melitensis*<sup>16</sup> and *B. abortus*<sup>50</sup>, falling in the early 10th millennium BP, a period documenting the consolidation of livestock-based economies in

Southwest Asia. In the Iranian Zagros Mountains, our estimate overlaps evidence of established, multi-generation communities of goat herders ~10,000 years BP, with the appearance of domestic-morphology sheep hundreds of years later<sup>11</sup>. In Central Anatolia, sheep and goat management was practised in the 11th millennium BP<sup>51</sup>; the management of cattle herds is documented by the mid-9th millennium BP, and likely earlier<sup>52</sup>. By the early 9th millennium BP, pastoralist communities managing cattle, sheep, goat, and pig herds spread rapidly along the southern and western coasts of Anatolia<sup>47</sup>, eventually into Europe.

Given our estimated *Brucella* diversification times, the intensification and spread of livestock husbandry may have also precipitated an epizootic transition: an increase in the potential for pathogen transmission within the herd and host switching between livestock species. Bacterial shedding due to brucellosis-induced abortion may have been particularly consequential in the context of greater host densities of Neolithic herds<sup>53</sup>. Within-settlement penning and confinement<sup>10</sup>, suboptimal hygiene practices in incipient herding communities, and the mixing of previously allopatric species may have promoted transmission and pathogen diversification. The common ancestor of *B. melitensis* and *B. abortus* evolved in this context, presumably diverging from a generalist infection capacity to more host-adapted lineages via circulation in mixed ruminant herds.

The specific genes with fixed nonsense in *B. melitensis* may reflect its pathogenesis within animals and humans. The early-evolving *nosZ* nonsense variant may be a consequence of the *B. melitensis*’s prolific extracellular replication induced by erythritol, a step in *Brucella*’s transmission via placental and foetal remains. As *B. abortus* and other genus members show erythritol sensitivity<sup>54</sup>, the *nosZ* pseudogenization likely occurred in the context of already-evolved erythritol tropism or response. Selection for respiratory activity under anaerobic conditions could therefore be relaxed in *B. melitensis*, given the ten-fold bias to extracellular replication in the presence of erythritol<sup>38</sup>. The late-occurring BME\_RS15065 pseudogenization may similarly have been fixed in this context. Additional palaeogenomes, of greater completeness and representing a range of temporal periods, would further refine the tempo of these evolutionary events underlying *B. melitensis*’s host adaptation and zoonotic capacity.

## Methods

### Sample provenance

Menteşe Höyük is found in the Yenişehir plain south of Lake Iznik (approximately 40.277 N, 29.524E), and is one of the earliest Neolithic sites in northwest Anatolia, occupied from the 9th millennium BP<sup>55</sup>. The Menteşe Höyük specimens analysed here derive from late-9th millennium BP settlement layers, from the 2000 test pits was exported in 2002 to Lionel Gourichon with the permission of the director of the Iznik Museum, Taylan Sevil, and the site’s excavation directors, Jacob Roodenberg and Songül Alpaslan-Roodenberg. Consent for the genetic analysis, as part of the ERC CODEX project, was given in August 2013 by Ömer Eren, then director of the Iznik Museum, with the approval of the excavation directors.

### Sampling and extraction

All laboratory work was performed in dedicated ancient DNA facilities in Trinity College Dublin. We sampled five ovicaprine petrosal specimens from Menteşe Höyük (Supplementary Data 1). Petrous bones were decontaminated by 30 minute exposure to UV light and removal of the surface dirt using a drill bit. A dremel saw was then used to subsample the petrous bones at the densest region. Bone subsamples were pulverised using a MixerMill at 30 f, until complete powderization.

100–150 mg of bone powder was used for DNA extraction. Initial screening double-stranded DNA (dsDNA) libraries were constructed from DNA extracted according to silica-based extraction protocols<sup>56</sup>. We performed an additional extraction round on bone pellets

remaining from the original extraction incubation. Bone pellets were subject to an EDTA-proteinase K (Fisher, 10628203; Sigma-Aldrich, 3115828001) digest at 37 °C for 24 hours using a modified pH PB buffer<sup>57</sup>.

### Library building and shotgun sequencing

All libraries were dsDNA libraries constructed following the Meyer and Kircher protocol<sup>58</sup>. One library using 5 µl starting DNA was prepared without uracil DNA glycosylase treatment, to capture native damage patterns. All other prepared libraries were pre-treated with 5 µl UDG glycosylase (New England Biolabs, M5505L) at 37 °C for 1 hour. Following dual indexed-PCR amplification for 12 cycles, libraries were shotgun sequenced on HiSeq 2000 (Macrogen, Seoul), MiSeq and NovaSeq 6000 platforms (TrinSeq, Dublin) (Supplementary Data 4).

### Radiocarbon dating

We generated two radiocarbon age estimates for Menteşe: UBA-47124: 7066 + / 42, and OxA-43559: 7117 + / 25 (Supplementary Data 3). We then used R\_Combine and the Oxcal4.4<sup>23</sup> to produce a combined calibrated age distribution. The resulting calibrated age 2σ range was 6057-5913 cal BCE (8007-7863 cal BP).

### Species and sex determination

Screening libraries from Menteşe were assigned species based on Fastq Screen<sup>59</sup>. Menteşe6 was assessed to be a sheep based on total assigned reads, and the “one read, one genome” metric (Supplementary Fig. 3). We assessed karyotypic sex based on the ratio of reads aligning to chromosomes and chromosome length (Supplementary Fig. 2).

### Screening using KrakenUniq and *E* value

After deduplication with PRINSEQ<sup>60</sup>, we screened one fastq file per sample from available ancient ruminant datasets (PRJEB26011, PRJEB31621, PRJEB40573, PRJEB43881) and from Menteşe specimens. Following adaptor removal with cutadapt<sup>61</sup> (single end reads) or AdapterRemoval v2<sup>62</sup> (paired end reads), fastq files were cleaned from host reads by aligning libraries with bowtie2<sup>63</sup> on a concatenated fasta mix of different animal and human reference genomes (GRCh38, Sscrofa11.1, ARS-UCD1.2, mCerEla1.1, ASM170441v1, EquCab3.0, ARS-UI\_Ramb\_v2.0) with no mapping quality filter applied. Then, the metagenomic profile of the non-aligned reads was computed with KrakenUniq<sup>21</sup> (v1.0.4) using a custom database containing a microbial version of NCBI nt combined with human and complete eukaryotic reference genomes<sup>64</sup>.

The KrakenUniq algorithm is effective at identifying low abundant taxa reads; notably, KrakenUniq reports unique *k*-mer assigned to reads and an estimate of the breadth of coverage. Guellil and colleagues<sup>22</sup> developed a score to discriminate true and false positives in taxonomic assignment using KrakenUniq based on an *E* value calculated as:  $E = (K/R)*C$ , where *K* is the unique *k*-mer count, *R* is the read count and *C* is the estimate of breadth of coverage. The *E* value of each KrakenUniq metagenomic assignment was calculated and the results were filtered at the species taxonomic level for a list of core *Brucella* (*B. abortus*, *B. canis*, *B. inopinata*, *B. melitensis*, *B. suis*, *Brucella microti*, *B. ovis*, *B. pinnipedalis*, *B. neotomae* and *B. ceti*). We further investigated Menteşe6, the only library passing our threshold of minimum 21 reads assigned at species level and *E* value > 0.001 (Supplementary Data 2).

### HOPS

As Menteşe6 passed the *E* value and minimum read number threshold, we assessed the authenticity of the *Brucella* signal with the HOPS pipeline<sup>24</sup>, using seven merged single-end fastq files due to memory limitations. We used as input the default pathogen list ([https://github.com/rhuebler/HOPS/blob/external/Resources/default\\_list.txt](https://github.com/rhuebler/HOPS/blob/external/Resources/default_list.txt)), in addition to all core *Brucella* species listed above and other *Brucella* genus

members. A reduced RefSeq database was constructed which minimises the inclusion of redundant sequences while maintaining the observed diversity within species. All available assemblies at the chromosome and whole genome level were considered for bacteria, archaea and viruses. For species which had more than one genome, all genomes were subjected to the MeshClust2 software, which utilises the mean-shift algorithm to select representative sequences that encompass distinct sections of the nucleotide diversity<sup>65</sup>. Selected genomes were subjected to a blast search on the first 10,000 bases to check for mislabelled genomes. Blast-matches to the sequence of interest were ignored and the genome was removed if any of the top 3 bitscore hits were not within the same species taxid. Several eukaryotic genomes were included as representation of species likely associated with domestication contexts. Conterminator was run to identify and remove eukaryotic contigs that were very short or significantly (>5%) contaminated across kingdoms<sup>66</sup>. The database was built using a step size of 4 to decrease the necessary computational resources.

### Lineage/diagnostic variant assessment

To assess the species-level identity of the Menteşe6 *Brucella*, we downloaded published genomes<sup>26</sup> and confirmed species-level assignment with a core genome alignment relative to *B. melitensis* (GCF\_000007125.1) using snippy (<https://github.com/tseemann/snippy>) and a IQTREE-2<sup>67</sup> ML phylogeny (parameters: -T 8 -m MF). We amended species assignment based on phylogeny inspection, and then reduced the dataset to the closely related “classic” or core *Brucella* species (*B. suis*, *B. canis*, *B. ovis*, *B. abortus*, *B. melitensis*, *B. ceti*, *B. pinnipedalis*, *B. microti*, *B. neotomae*) and a clade containing *B. inopinata*. For *B. suis*, *B. abortus*, and *B. melitensis*, we randomly subsampled genomes to 23 representatives, matching the number of *B. canis* genomes (Supplementary Data 6). The resulting 141 genomes were used to generate a core SNP alignment relative to either the *B. melitensis* or *B. suis* (GCF\_000007505.1) reference genomes, using the software snippy. For both reference-based alignments, a custom python script was used to identify variants which were fixed for a derived allele in a given lineage (e.g. *B. canis*; the *B. melitensis*-*B. abortus* clade), and the ancestral allele was observed in all other *Brucella* (extract\_lineage\_sites.py; all custom scripts available at [https://github.com/LouisLhote/Neolithic\\_Brucella\\_paper/](https://github.com/LouisLhote/Neolithic_Brucella_paper/)). For these sites, a pileup file was generated for the Menteşe6 (samtools mpileup -q 30 -Q 20 -B). A custom python script (pseudohaploidize\_pileup.py) was then used to randomly sample a base at each site. In parallel, we applied a more stringent call approach by removing sites with more than one allele observed (i.e. possible heteroplasmy) and with 10 or more covering reads, using a custom python script (pseudohaploidize\_pileup\_no-heteroplasmy-max10.py). Finally, the observed Menteşe6 alleles were compared with the derived and ancestral alleles to calculate a lineage match rate (Supplementary Data 7), using a custom R script (compare\_lineage\_sites\_pseudohap-pileup.r).

### Alignment of Menteşe6 data to host and *Brucella* genomes

Adaptor trimming, and quality filtering of demultiplexed reads was performed using AdapterRemoval v 2.3.2<sup>62</sup>, removing reads with <30 bp length and combining overlapping read pairs (-collapse -minadapteroverlap1 -adapter1 AGATCGGAAGAGCACACGTCTGAACCTCCAGTCAC -adapter2 AGATCGGAAGAGCGTCGTAGGGAAAGAGTGT -minlength 30 -trimns -trimqualities).

To remove host reads, the collapsed reads were aligned on the sheep reference genome oviAri3 (GCA\_000298735.1) using the bwa aln algorithm with relaxed parameters (-t1024 -n 0.01 -o 2). All non-aligned reads were then aligned on the *B. melitensis* reference genome ASM712v1 (GCF\_000007125.1) with the same parameters as above. Reads with mapping quality under 30 were filtered out with samtools v1.19.2, and duplicates were removed with picard MarkDuplicates v2.26.11. An additional alignment was performed of all non sheep-

aligned reads with the same parameters on the *B. suis* reference genome ASM750v1 (GCF\_000007505.1). We calculated post-mortem damage patterns on the *B. melitensis* alignment using mapDamage<sup>68</sup>. For downstream analysis only USER-treated data were used.

### Gene level coverage and breadth

Coverage and breadth (the proportion of bases covered by at least one read) was estimated for each gene in the *B. melitensis* bv. 1 str. 16 M strain reference genome (GCF\_000007125.1) using the RefSeq annotation and bedtools v2.31.1 coverage (Supplementary Data 11). A 5kbp sliding window of coverage was also computed using bedtools coverage (Fig. 1c). The coverage peak of ~21X on chromosome I overlaps with two RNA polymerase genes, removed in downstream analyses through a per-site maximum coverage and sliding window percentile coverage filter. Low coverage regions both overlap ribosomal RNA and tRNA genes.

### Phylogenetic analysis and reference datasets

We first assessed the phylogenetic position of Mentese6 using alignment to *B. suis*, an outgroup to *B. melitensis* and *B. abortus*. We downloaded a reference dataset representing global diversity of *B. melitensis* and *B. abortus* genomes<sup>26</sup> and then removed sequences which lacked a known sampling year and after thinning the dataset with treemmer (-RTL 0.95; available at <https://github.com/fmenardo/Treemmer>) following the approach of Long and colleagues<sup>15,16</sup>, leaving a final set of 227 genomes (Supplementary Data 5). The 227 genomes were converted into fastq reads using ART<sup>69</sup> with a read length of 150 bp and a read coverage of 20 fold. The same parameters for the Mentese6 read alignment step were used to align those reads on the *B. suis* reference genome ASM750v1 (GCF\_000007505.1). A vcf file was produced for every genome and for the Mentese6 bam file with GATK package (available at <https://gatk.broadinstitute.org/hc/en-us>), using the “EMIT\_ALL\_SITES” and “PLOIDY = 2” option that generated a call for all positions in the reference genome.

As intracellular *Brucella* have many free-living, closely-related relatives, we aimed to minimise false positive variant calling due to spurious alignment and retain only high confidence variants for Mentese6. Each vcf was separated by chromosome (I and II) using bcftools (<https://samtools.github.io/bcftools/bcftools.html>). We then removed sites covered by a maximum of twice the mean chromosomal coverage using vcftools (<https://vcftools.sourceforge.net/>). For the Mentese6 alignment to *B. suis*, this was 6 or 5 reads for chromosome I and II, due to their differing mean coverage (2.62X and 2.26X respectively). We then masked any sites called as heterozygous in Mentese6 using bcftools.

To detect recombination in the modern data, we used MultiVCFAnalyzer v0.85.2<sup>70</sup> to produce a full genome alignment (including non-variable sites) of the modern genomes analysed ( $n = 226$ ) and the *B. suis* reference for both chromosomes (Supplementary Data 5). The following parameters were used: a minimal genotyping quality of 30, a minimal coverage of 10 for base calls of homozygous alleles, and a minimal allele frequency of 80% for heterozygous calls. If any criteria was not met, an “N” was used in the respective genomic position. Recombination was detected in these modern data using Gubbins v3.1.0<sup>71</sup>.

We then determined variant positions in *B. suis*-aligned modern and ancient genomes (including Mentese6; see Supplementary Data 5) with MultiVCFAnalyzer v0.85.2 with the following parameters: a minimal genotyping quality of 30, a minimal coverage of 3 for base calls, and a minimal allele frequency of 90% for heterozygous call. If any of the criteria were not met, an “N” was used in that respective genomic position. Recombinant regions detected by Gubbins in the previous step were masked using MultiVCFAnalyzer. We additionally computed coverage of Mentese6 in non-overlapping 50 bp sliding windows using bedtools coverage and masked regions in the SNP call with greater

than the 99<sup>th</sup> percentile coverage (windows with mean coverage  $> 10X$ ). A total of 426 variant sites were masked by MultiVCFAnalyzer. Variant alignments produced by MultiVCFAnalyzer for chromosome I and II were concatenated using seqkit concat. We removed all positions in the resulting alignment where modern data contained an “N” (i.e. only sites where all modern samples had a base call were retained). The resulting alignment containing 30,364 varying sites and the reference genome was used to create a maximum-likelihood phylogeny with IQ-TREE2<sup>67</sup> using the nucleotide substitution model (TVM + F + ASC + R3) determined with modelfinder.

### Average nucleotide identity (ANI) using *B. suis* alignment

ANI calculation was performed using aniclustermapper<sup>27</sup> on the full genome alignment of the *B. suis*-aligned sequences of *B. melitensis* and *B. abortus* after removing all sites with missing data (“N”). ANI values are reported in Supplementary Data 8, and Fig. 2 for values concerning Mentese6.

### Temporal signal

We tested for a temporal signal in our dataset of *B. suis*-aligned *B. melitensis* and *B. abortus* genomes using TempEst<sup>28</sup>. For the age of samples, the sampling year for modern genomes was used; for the 14th century mediaeval *B. melitensis*<sup>16</sup>, the age of the death of the host was used; the Mentese6, the midpoint of its combined calibrated age was used. We found weak evidence for a temporal signal ( $R^2 = 0.3837$ , Supplementary Fig. 8). However, when sequences were assessed by their cladal identity within either species, there was a clear signal of local clock rates specific to subclades as expressed by clustering of clades by their root-to-tip distance. This rate variation appears to lead to a non-significant result in a dataset of limited pre-modern sequences (here, two). The Mentese6 sequence clearly shows a reduced root-to-tip distance (Fig. 2A and Supplementary Fig. 8), as expected by the radiocarbon-estimated age of the Mentese6 temporal bone.

### Beast time tree using *B. melitensis* alignment

To estimate the age of the most recent common ancestor of *B. melitensis* and *B. abortus* we produced a variant alignment of the Mentese6 sample, the *B. abortus* reference genome (GCF\_000054005.1), and the 141 *B. melitensis* used in the “Phylogenetic analysis and reference datasets” section (Supplementary Data 5). We used the same pipeline as the used in the ML phylogeny step but instead aligned on *B. melitensis* (GCF\_000007125.1), with the same filtering steps and alignment parameters as described for the *B. suis* alignment under “Phylogenetic analysis and reference datasets”.

For genotyping, as coverage was higher for the *B. melitensis* alignment (mean coverage 3.62X and 3.17X for chromosomes I and II), we retained sites where Mentese6 had at least 3 reads and a maximum of 7 or 6 reads for chromosomes I and II respectively. We masked recombinant regions detected in modern genomes by Gubbins, regions with  $\geq 99$ th percentile coverage as calculated (bedtools coverage) in non-overlapping 50 bp sliding windows (mean coverage  $> 11X$ ), and sites where any modern genomes lacked a base call (i.e. “N” was allowed only in Mentese6). A total of 389 variants were masked by MultiVCFAnalyzer, producing an alignment with 18,300 varying sites.

BEAST v2.6.7<sup>29</sup> was used for divergence dating of the *B. melitensis*-*B. abortus* phylogeny. For modern genomes, collection dates were used as tip dates with a uniform distribution of 0.1 year around the collection date. One *B. melitensis* genome in the dataset (GCA\_030370715.1) was obtained from a mediaeval Franciscan friar who died in 1394<sup>16</sup>. His death date was used as a tip date with a uniform distribution of 0.1 year around his death date. A lognormal distribution was used for the Mentese6 ancient sample tip date. We used the  $C_{14}$  calibration (6057–5913 cal BCE/8007–7863 cal BP) as the M parameter and we varied the S parameter so the 95%/2 $\sigma$  lognormal range matched

the range in real space. The number of invariant sites was directly supplied in the xml file.

Beast model test<sup>72</sup> was used to test two different clock models, a lognormal relaxed clock and a strict clock, and two demographics models, coalescent exponential population and coalescent Bayesian skyline. We estimated marginal likelihoods to determine the best clock and demographic model for our dataset using path sampling. We estimated that a strict clock and coalescent Bayesian skyline demographic model were the best fit for our datasets by comparing each clock model and demographic model against each other (Supplementary Data 9). We then ran 3 replicates of 250 millions of chains using the mutation rate from a recent study<sup>16</sup> as the rate prior and with 1% burn-in. The BEAST parameters log file were assessed in Tracer v1.7.2 using 20% burn-in. This showed excellent mixing (ESS >300) and model convergence between runs. The tree files produced by BEAST were combined with logcombiner using 20% burn-in. Then we used TreeAnnotator, to produce a maximum clade credibility tree with 0% burn-in (20% burn-in was already applied by logcombiner).

### Variant effect annotation

*B. melitensis*-defining variants covered in Mentes6 and identified above (“Lineage analysis”) were annotated using Ensembl Variant Effect Predictor (<https://www.ensembl.org/info/docs/tools/vep/index.html>) using the *B. melitensis* bv. 1 str. 16 M strain. Mentes6 genotype calls were first converted to an appropriate format and then uploaded to the Ensembl Variant Effect Predictor server (Supplementary Data 10). As variants were relative to this reference genome, we filtered for “missense\_variant” and “stop\_lost”, the latter identifying nonsense mutations which had occurred in the *B. melitensis* lineage relative to other *Brucella*.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Collapsed sequencing data with host reads removed and *B. melitensis*-aligned bam file for Mentes prior to data QC are available at ENA accession PRJEB75678. Reference sequences used in this study are previously published and available for *B. melitensis* [[https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_000007125.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000007125.1/)], *B. abortus* [[https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_000054005.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000054005.1/)], and *B. suis* [[https://www.ncbi.nlm.nih.gov/datasets/genome/GCF\\_000007505.1/](https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000007505.1/)]. We employed sequences previously analyzed by Abdel-Gilil and colleagues [<https://doi.org/10.1128/JCM.00311>]; ENA accessions for each are presented in Supplementary Data 6. Source data are provided with this paper. The screened Menteşe specimens are under the temporary stewardship of Kevin G. Daly (UCD, Dublin) and permanent stewardship of Lionel Gourichon, Université Côte d’Azur, Nice. Applications to access and re-examine the material should be made to Dr. Gourichon. Specimen identifiers are provided in Data S1.

### Code availability

Custom code and scripts used in this manuscript are available on github [[https://github.com/LouisLhote/Neolithic\\_Brucella\\_paper/](https://github.com/LouisLhote/Neolithic_Brucella_paper/)] and via zenodo<sup>73</sup>.

### References

1. Laine, C. G., Johnson, V. E., Scott, H. M. & Arenas-Gamboa, A. M. Global Estimate of Human Brucellosis Incidence. *Emerg. Infect. Dis.* **29**, 1789–1797 (2023).
2. Whatmore, A. M. & Foster, J. T. Emerging diversity and ongoing expansion of the genus *Brucella*. *Infect. Genet. Evol.* **92**, 104865 (2021).
3. Hassan, L. et al. Risk Factors for Acute Brucellosis in Patients on the Day of Admission at Selected Hospitals of Abbottabad, Pakistan. *Front Public Health* **9**, 669278 (2021).
4. von Bargen, K., Gorvel, J.-P. & Salcedo, S. P. Internal affairs: investigating the *Brucella* intracellular lifestyle. *FEMS Microbiol. Rev.* **36**, 533–562 (2012).
5. Alexander, B., Schnurrenberger, P. R. & Brown, R. R. Numbers of *Brucella* abortus in the placenta, umbilicus and fetal fluid of two naturally infected cows. *Vet. Rec.* **108**, 500 (1981).
6. Scientific Committee on Animal Health and Animal Welfare. *Brucellosis in Sheep and Goat (Brucella Melitensis)*. [https://food.ec.europa.eu/document/download/e377914a-810a-4afe-a84c-147f6f12a636\\_en?filename=sci-com\\_scah\\_out59\\_en.pdf](https://food.ec.europa.eu/document/download/e377914a-810a-4afe-a84c-147f6f12a636_en?filename=sci-com_scah_out59_en.pdf) (2001).
7. Pappas, G., Akritidis, N., Bouskouki, M. & Tsianos, E. Brucellosis. *N. Engl. J. Med.* **352**, 2325–2336 (2005).
8. Arenas-Gamboa, A. M. et al. Human Brucellosis and Adverse Pregnancy Outcomes. *Curr. Trop. Med. Rep.* **3**, 164–172 (2016).
9. Fournié, G., Pfeiffer, D. U. & Bendrey, R. Early animal farming and zoonotic disease dynamics: modelling brucellosis transmission in Neolithic goat populations. *R. Soc. Open Sci.* **4**, 160943 (2017).
10. Yeomans, L., Bangsgaard, P. & Ahadi, G. Perinatal Remains of Livestock: An Under-utilised Line of Evidence for Animal Penning in the Neolithic of Southwest Asia. *Environ. Archaeol.* **28**, 207–221 (2021).
11. Zeder, M. A. A view from the Zagros: new perspectives on livestock domestication in the Fertile Crescent. in *The First Steps of Animal Domestication: New Archaeozoological Approaches* (eds. Vigne, J.-D., Peters, J. & Helmer, D.) 125–146 (Oxbow Books, Oxford, 2005).
12. Merrett, D. C. Bioarchaeology in Early Neolithic Iran: Assessment of Health Status and Subsistence Strategy. (University of Manitoba, 2004).
13. Bendrey, R., Taylor, G. M., Bouwman, A. S. & Cassidy, J. P. Suspected bacterial disease in two archaeological horse skeletons from southern England: palaeopathological and biomolecular studies. *J. Archaeol. Sci.* **35**, 1581–1590 (2008).
14. Mutolo, M. J., Jenny, L. L., Buszek, A. R., Fenton, T. W. & Foran, D. R. Osteological and molecular identification of Brucellosis in ancient Butrint, Albania. *Am. J. Phys. Anthropol.* **147**, 254–263 (2012).
15. Kay, G. L. et al. Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. *MBio* **5**, e01337–14 (2014).
16. Long, G. S. et al. A 14th century CE *Brucella melitensis* genome and the recent expansion of the Western Mediterranean clade. *PLoS Pathog.* **19**, e1011538 (2023).
17. Aune, K., Rhyan, J. C., Russell, R., Roffe, T. J. & Corso, B. Environmental persistence of *Brucella abortus* in the Greater Yellowstone Area. *J. Wildl. Manag.* **76**, 253–261 (2012).
18. Moreno, E., Blasco, J. M., Letesson, J. J., Gorvel, J. P. & Moriyón, I. Pathogenicity and Its Implications in Taxonomy: The *Brucella* and *Ochrobactrum* Case. *Pathogens* **11**, 377 (2022).
19. González-Espinoza, G., Arce-Gorvel, V., Mémet, S. & Gorvel, J.-P. *Brucella*: Reservoirs and Niches in Animals and Humans. *Pathogens* **10**, 186 (2021).
20. Ashford, R. T., Muchowski, J., Koylass, M., Scholz, H. C. & Whatmore, A. M. Application of Whole Genome Sequencing and Pan-Family Multi-Locus Sequence Analysis to Characterize Relationships Within the Family Brucellaceae. *Front. Microbiol.* **11**, 1329 (2020).
21. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 198 (2018).
22. Guellil, M. et al. An invasive *Haemophilus influenzae* serotype b infection in an Anglo-Saxon plague victim. *Genome Biol.* **23**, 22 (2022).
23. Ramsey, C. B. Development of the Radiocarbon Calibration Program. *Radiocarbon* **43**, 355–363 (2001).

24. Hübler, R. et al. HOPS: automated detection and authentication of pathogen DNA in archaeological remains. *Genome Biol.* **20**, 280 (2019).

25. Herbig, A. et al. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. <https://doi.org/10.1101/050559> (2016).

26. Abdel-Gil, M. Y. et al. Core Genome Multilocus Sequence Typing Scheme for Improved Characterization and Epidemiological Surveillance of Pathogenic Brucella. *J. Clin. Microbiol.* **60**, e0031122 (2022).

27. Shimoyama, Y. A Tool for Drawing ANI Clustermap between All-vs-All Microbial Genomes. (Github, 2022).

28. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).

29. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

30. Zeder, M. A. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc. Natl Acad. Sci. USA.* **105**, 11597–11604 (2008).

31. Arbuckle, B. S. & Atici, L. Initial diversity in sheep and goat management in Neolithic south-western Asia. *Levantina* **45**, 219–235 (2013).

32. Suárez-Esquivel, M., Chaves-Olarte, E., Moreno, E. & Guzmán-Verri, C. Brucella Genomics: Macro and Micro Evolution. *Int. J. Mol. Sci.* **21**, 7749 (2020).

33. Toft, C. & Andersson, S. G. E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nat. Rev. Genet.* **11**, 465–475 (2010).

34. Sabatini, S., Bergerbrant, S., Brandt, L. Ø., Margaryan, A. & Allentoft, M. E. Approaching sheep herds origins and the emergence of the wool economy in continental Europe during the Bronze Age. *Archaeol. Anthropol. Sci.* **11**, 4909–4925 (2019).

35. Keim, P. S. & Wagner, D. M. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nat. Rev. Microbiol.* **7**, 813–821 (2009).

36. Freddi, L., de la Garza-García, J. A., Al Dahouk, S., Occhialini, A. & Köhler, S. Brucella spp. are facultative anaerobic bacteria under denitrifying conditions. *Microbiol Spectr.* **11**, e0276723 (2023).

37. Köhler, S. et al. The analysis of the intramacrophagic virulome of *Brucella suis* deciphers the environment encountered by the pathogen inside the macrophage host cell. *Proc. Natl Acad. Sci.* **99**, 15711–15716 (2002).

38. Petersen, E. et al. Erythritol triggers expression of virulence traits in *Brucella melitensis*. *Microbes Infect.* **15**, 440–449 (2013).

39. Barbier, T. et al. Erythritol Availability in Bovine, Murine and Human Models Highlights a Potential Role for the Host Aldose Reductase during Brucella Infection. *Front. Microbiol.* **8**, 1088 (2017).

40. Conde-Álvarez, R. et al. Identification of lptA, lpxE, and lpxO, Three Genes Involved in the Remodeling of Brucella Cell Envelope. *Front. Microbiol.* **8**, 2657 (2017).

41. Wang, Y. et al. Comparative proteomics analyses reveal the virB of *B. melitensis* affects expression of intracellular survival related proteins. *PLoS One* **4**, e5368 (2009).

42. Barrett, R., Kuzawa, C. W., McDade, T. & Armelagos, G. J. EMERGING AND RE-EMERGING INFECTIOUS DISEASES: The Third Epidemiologic Transition. *Annu. Rev. Anthropol.* **27**, 247–271 (1998).

43. Key, F. M. et al. Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat. Ecol. Evol.* **4**, 324–333 (2020).

44. Sikora, M. et al. The landscape of ancient human pathogens in Eurasia from the Stone Age to historical times. *bioRxiv* 2023.10.06.561165 (2023) <https://doi.org/10.1101/2023.10.06.561165>.

45. Gourichon, L. & Helmer, D. Etude de la faune néolithique de Menteşe (Turquie). in *Life and death in a prehistoric settlement in Northwest Anatolia. The İlipinar excavations, volume III: with contributions on Hacilar tepe and Menteşe* (eds. Roodenberg, J. & Alpaslan-Roodenberg, S) 435–448 (Nederlands Instituut voor het Nabije Oosten, Leiden, 2008).

46. Evershed, R. P. et al. Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* **455**, 528–531 (2008).

47. Arbuckle, B. et al. Data Sharing Reveals Complexity in the Westward Spread of Domestic Animals across Neolithic Turkey. *PLoS One* **9**, e99845 (2014).

48. Horejs, B. Long and short revolutions towards the Neolithic in western Anatolia and Aegean. *Doc. Praehist.* **46**, 68–83 (2019).

49. Roodenberg, J. and Alpaslan-Roodenberg, S. İlipinar and Menteşe - Early Farming Communities in the Eastern Marmara. *The Neolithic in Turkey. New Excavations & New Research. Volume 5: Northwestern Turkey and Istanbul*. Edited by M. Özdoğan, N. Başgelen and P. Kuniholm Archaeology and Art Publications, Istanbul, 2013 - p. 69–91 (2013).

50. Janke, N. R. et al. Global phylogenomic diversity of *Brucella abortus*: spread of a dominant lineage. *Front. Microbiol.* **14**, 1287046 (2023).

51. Stiner, M. C., Munro, N. D., Buitenhuis, H., Duru, G. & Özbaşaran, M. An endemic pathway to sheep and goat domestication at Aşıklı Höyük (Central Anatolia, Turkey). *Proc. Natl. Acad. Sci. USA.* **119**, e2110930119 (2022).

52. Arbuckle, B. S. & Kassebaum, T. M. Management and domestication of cattle (*Bos taurus*) in Neolithic Southwest Asia. *Anim. Fron* **11**, 10–19 (2021).

53. Foufopoulos, J., Wobeser, G. A. & McCallum, H. Disease Epidemiology in Natural Systems. in (Oxford University Press, 2022).

54. Smith, H. et al. Foetal Erythritol: A Cause of the Localization of *Brucella Abortus* in Bovine Contagious Abortion. *Nature* **193**, 47–49 (1962).

55. *Life and Death in a Prehistoric Settlement in Northwest Anatolia: The İlipinar Excavations, Volume III: With Contributions on Hacilar tepe and Menteşe*. 110 (Nederlands Instituut voor het Nabije Oosten, Leiden, 2008).

56. Mattiangeli, V., Cassidy, L. M., Daly, K. G., Mullin, V. E. & Verdugo, M. *Multi-Step Ancient DNA Extraction Protocol For Bone And Teeth. dx*. (2023) <https://doi.org/10.17504/protocols.io.6qpvr45b2gmk/v1>.

57. Mattiangeli, V., Cassidy, L. M., Daly, K. G. & Mullin, V. E. *Bleach Extraction Protocol: Damaged or Degraded DNA Recovery from Bone or Tooth Powder*. <https://doi.org/10.17504/protocols.io.8epv5j88nl1b/v1> (2023).

58. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* **2010**, db.prot5448 (2010).

59. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res.* **7**, 1338 (2018).

60. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

61. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).

62. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).

63. Poullet, M. & Orlando, L. Assessing DNA Sequence Alignment Methods for Characterizing Ancient Genomes and Methylomes. *Front. Ecol. Evol.* **8**, 105 (2020).

64. Pochon, Z. et al. aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *Genome Biol.* **24**, 1–30 (2023).

65. James, B. T. & Girgis, H. Z. MeShClust2: Application of alignment-free identity scores in clustering long DNA sequences. <https://doi.org/10.1101/451278> (2018).
66. Steinegger, M. & Salzberg, S. L. Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biol.* **21**, 115 (2020).
67. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
68. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
69. Huang, W., Li, L., Myers, J. R. & Marth, G. T. ART: a next-generation sequencing read simulator. *Bioinformatics* **28**, 593–594 (2012).
70. Bos, K. I. et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* **514**, 494–497 (2014).
71. Croucher, N. J. et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15 (2015).
72. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17**, 42 (2017).
73. L'Hôte, L. & Daly, K. *Neolithic\_Brucella\_paper-v1.0*. <https://doi.org/10.5072/zenodo.80001>.

## Acknowledgements

We thank Daniel Bradley, Marco Rosario Capodiferro, Lara Cassidy, Evangelos Dimopoulos, Louis Du Plessis, Iseult Jackson, and Anahit Hovhannisan for their advice on analysis and the manuscript. We thank Songül Alpaslan-Roodenberg, Daniel Helmer, and Jacob Roodenberg for their contributions as co-directors of excavations at Menteşe Höyük and for co-analysis of the faunal assemblage. We thank Andrew Hare for his contribution to laboratory work. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 21/PATH-S/9515(T) (to K.G.D.). M.D.T. was supported by European Research Council (ERC) Investigator Grant 295729-CodeX. Á.H. and V.M. are supported by the European Research Council under the European Union's Horizon 2020 research and innovation programme (885729-AncestralWeave). F.M.K. is supported by the Klaus Tschira Foundation (GSO/KT030) and together with I.L. supported by the Max Planck Society.

## Author contributions

K.G.D. conceived and supervised the study; L.L., V.M., M.D.T. and A.H. performed laboratory work; L.L. curated generated sequencing data and

reference datasets; I.L. and F.M.K. conceived and developed the MALT database; L.L. processed and analysed the data with contributions from K.G.D. and F.M.K.; L.L. and K.G.D. wrote custom scripts and visualised data; L.G. provided the specimens; K.G.D. and L.L. wrote the initial draft with review and editing contributions from F.M.K., I.L., L.G., and all other co-authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-50536-1>.

**Correspondence** and requests for materials should be addressed to Kevin G. Daly.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024