

## RESEARCH ARTICLE

A 14<sup>th</sup> century CE *Brucella melitensis* genome and the recent expansion of the Western Mediterranean clade

George S. Long<sup>1,2</sup>\*, Jessica Hider<sup>2,3</sup>\*, Ana T. Duggan<sup>2,3</sup>, Jennifer Klunk<sup>1,2,4</sup>, Katherine Eaton<sup>2,3</sup>, Emil Karpinski<sup>1,2</sup>, Valentina Giuffra<sup>5</sup>, Luca Ventura<sup>6,7</sup>, Tracy L. Prowse<sup>3</sup>, Antonio Fornaciari<sup>5</sup>, Gino Fornaciari<sup>8</sup>, Edward C. Holmes<sup>9</sup>, G. Brian Golding<sup>1</sup>, Hendrik N. Poinar<sup>2,3,10,11,12</sup>\*

**1** Department of Biology, McMaster University, Hamilton, Canada, **2** McMaster Ancient DNA Centre, Departments of Anthropology and Biochemistry, McMaster University, Hamilton, Canada, **3** Department of Anthropology, McMaster University, Hamilton, Canada, **4** Daicel Arbor Biosciences, Ann Arbor, Michigan, United States of America, **5** Division of Paleopathology, Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy, **6** Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy, **7** Division of Pathology, San Salvatore Hospital, Coppito, Italy, **8** Maria Luisa di Borbone Academy, Viareggio, Italy, **9** Sydney Institute for Infectious Diseases, School of Medical Sciences, University of Sydney, Sydney, Australia, **10** Department of Biochemistry, McMaster University, Hamilton, Canada, **11** Michael G. DeGroote Institute for Infectious Disease Research, McMaster University, Hamilton, Canada, **12** CIFAR Humans and the Microbiome Program, Toronto, Canada

\* These authors contributed equally to this work.

\* [longg2@mcmaster.ca](mailto:longg2@mcmaster.ca) (GSL); [hiderjessica@gmail.com](mailto:hiderjessica@gmail.com) (JH); [poinarh@mcmaster.ca](mailto:poinarh@mcmaster.ca) (HNP)



## OPEN ACCESS

**Citation:** Long GS, Hider J, Duggan AT, Klunk J, Eaton K, Karpinski E, et al. (2023) A 14<sup>th</sup> century CE *Brucella melitensis* genome and the recent expansion of the Western Mediterranean clade. PLoS Pathog 19(7): e1011538. <https://doi.org/10.1371/journal.ppat.1011538>

**Editor:** D. Ashley Robinson, University of Mississippi Medical Center, UNITED STATES

**Received:** February 12, 2023

**Accepted:** July 5, 2023

**Published:** July 31, 2023

**Copyright:** © 2023 Long et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Code used to analyze the mapped data can be found at <https://github.com/longg2/AncientBrucella>. The scripts used for quality control, the metagenomic analyses, mapping, and the phylogenetics can be found at <https://github.com/longg2/LongBioinformatics>. The raw sequencing data can be found on NCBI under the ID PRJNA966239 while the assembled genome is GCF 030370715.

**Funding:** We thank the Social Sciences and Humanities Research Council of Canada

## Abstract

Brucellosis is a disease caused by the bacterium *Brucella* and typically transmitted through contact with infected ruminants. It is one of the most common chronic zoonotic diseases and of particular interest to public health agencies. Despite its well-known transmission history and characteristic symptoms, we lack a more complete understanding of the evolutionary history of its best-known species—*Brucella melitensis*. To address this knowledge gap we fortuitously found, sequenced and assembled a high-quality ancient *B. melitensis* draft genome from the kidney stone of a 14<sup>th</sup>-century Italian friar. The ancient strain contained fewer core genes than modern *B. melitensis* isolates, carried a complete complement of virulence genes, and did not contain any indication of significant antimicrobial resistances. The ancient *B. melitensis* genome fell as a basal sister lineage to a subgroup of *B. melitensis* strains within the Western Mediterranean phylogenetic group, with a short branch length indicative of its earlier sampling time, along with a similar gene content. By calibrating the molecular clock we suggest that the speciation event between *B. melitensis* and *B. abortus* is contemporaneous with the estimated time frame for the domestication of both sheep and goats. These results confirm the existence of the Western Mediterranean clade as a separate group in the 14<sup>th</sup> CE and suggest that its divergence was due to human and ruminant co-migration.

(20008499 to HNP), and the Canadian Institute for Advanced Research for the funding to HNP. The Natural Sciences and Engineering Research Council of Canada also provided funding for this work through a grant awarded to GBG (RGPIN-2020-05733). We are grateful for the support provided by the Joseph-Armand Bombardier Canada Graduate Scholarships Program Doctoral Scholarships to JH. EH is funded by an NHMRC Investigator Award (GNT2017197). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Author summary

We describe the reconstruction of a high-quality ancient *B. melitensis* genome isolated from the kidney stone of a 14<sup>th</sup> C. CE Italian friar. We confirm the presence of a *B. melitensis* genome from the Western Mediterranean clade in Italy during this time period and refine previous estimates for the diversification of this phylogenetic group. In addition, we propose that the speciation event separating *B. abortus* and *B. melitensis* occurred during the domestication of sheep and goats approximately 9000 BCE.

By comparing the gene content of the ancient *B. melitensis* genome with modern isolates, we provide evidence that neither the *B. melitensis* phylogenetic groups nor their respective sequence types necessarily correspond with the genetic components of their strains. The accessory genome of *B. melitensis* appears to be influenced by its geographic location, more so than its sequence type. We propose splitting the Eastern Mediterranean clade into three smaller groups to better reflect these genomic differences.

## Introduction

Brucellosis is a zoonotic disease caused by the bacterium *Brucella* that is most often transmitted to humans by domestic livestock through the consumption of animal products and the interactions with infected animals [1]. It is one of the most common zoonotic diseases with an estimated 500,000 cases each year, although its incidence is hypothesized to be far higher (~ 5 to 12.5 million) [2, 3]. Three main species of *Brucella* are pathogenic to humans. *B. melitensis* is carried by sheep and goats while *B. suis* and *B. abortus* are found in pigs and cows, respectively [4]. Brucellosis is a serious infection causing fever, chills, extreme muscle pain, and long-term osteoarticular changes that can be observed on bones [1]. As a chronic disease it is of particular importance to public health due to its indirect, yet significant economic impacts, by decreasing the workforce and reducing livestock reproduction and their associated products [5]. Brucellosis is common today in Asia, the Middle East, Africa, South America, and the Mediterranean [3]. Southern European countries make up the bulk of brucellosis infections in Europe, with Greece accounting for 20.3% of human cases while Italy has the highest prevalence of cattle herds, sheep and goat flocks testing positive for *Brucella* within the European Union as of 2020 [6].

Brucellosis is thought to have been a common scourge in the Mediterranean in the past, especially during the Roman and Medieval periods before the advent of pasteurization, and when pastoral practices such as transhumance were commonplace [7, 8]. However, little historical or archaeological evidence exists to support this claim [9]. Purported cases from the Roman period were identified in 16 skeletons from Herculaneum dating to 79 CE through the presence of non-specific vertebral lesions [8]. While skeletal lesions are a common manifestation of brucellosis, occurring in 10 to 85% of modern clinical cases, the lesions tend to occur in similar locations and with a similar appearance as those caused by tuberculosis [10], leading to potential confusion in differentiating between the two diseases in the past. If the pattern of lesion occurrence in brucellosis cases was the same as it is today, we would expect that more differential diagnoses of brucellosis would be reported from ancient remains [10]. However, there are several reasons why brucellosis may be under-reported in paleopathological analyses of archaeological samples; it is likely mistaken for other diseases that similarly impact bone like tuberculosis, the bacterium had not left skeletal changes by the time of death [7, 11, 12], or the non-specific nature of the lesions precluded a diagnosis [12].

Support for the presence of brucellosis during the Medieval period includes an ancient genome identified from a pelvic nodule of an individual in Sardinia dating to *c.* 1300 CE; two individuals from Albania (*c.* 900—*c.* 1200 CE) with PCR products arising from lesioned ribs and vertebrae; and two French individuals from the 14<sup>th</sup> and 18<sup>th</sup> C. CE where an infection was identified via paleopathology [13–15]. This limited DNA evidence compared to other infectious diseases could be due to improper sample selection [16], poor DNA preservation [12, 17, 18], or its mistaken diagnosis [18]. It is also possible that a focus on other diseases such as tuberculosis has meant that researchers have neglected possible evidence for brucellosis in the archaeological record.

The Mediterranean region is thought to have been a common geographic location for the origin and divergence of all *Brucella* species (especially *B. melitensis*) due to the diversity of strains in this area and the basal position of the Western Mediterranean clade in the global *Brucella* phylogeny [19, 20]. Despite this, the evolutionary history of the disease in the region is poorly understood. Due to the widespread movement of infected animals across regional, national, and international borders, the phylogeography of *Brucella* on both local and larger scales has been challenging to elucidate using modern isolates [19]. Sheep and goats are the preferred hosts for *B. melitensis*, and while spillover infections can occur, these alternative hosts tend to result in dead-end infections [4, 21]. This is especially the case in some animals that require constant exposure to *B. melitensis* to sustain an infection [22]. Sheep and goat husbandry extended across the Mediterranean littoral zone (Cyprus to Atlantic Portugal) by 5000 BCE [23–25]. Based on the dates of expansion of ruminants into the region during the Neolithic and dates of divergence for *B. melitensis*, it is hypothesized that the bacterium entered the region with the introduction of ruminants and subsequently diversified [4, 26, 27]. Paleopathological diagnoses of brucellosis in human skeletal remains from Jordan and Palestine (2100–1550 BCE), and Bahrain (3000–1200 BCE) support this hypothesis [7, 28, 29].

Archaeological and genetic data on the spread of sheep into the Mediterranean illustrate the complexity of repeated expansions and contractions over multiple periods of domestication. Subsequent selective breeding, trade, and geographic expansion have impacted the genetic signatures of sheep causing them to have little resemblance to their early progenitors [25, 30, 31]. Wild sheep and goats were domesticated approximately 10,000–8000 BCE in the Fertile Crescent [25, 32–34]. First domesticated for meat, sheep were later kept for their secondary products such as milk and wool [33]. The first expansion of domesticated sheep and goats from the Zagros mountains (northwestern Iran to southeastern Türkiye) spread to the borders of Europe and Africa by 8000 BCE [25, 33] following two routes of expansion; the Danubian route ran from Türkiye through the Balkans to the interior of Europe, while a southern route went through North Africa (North; Türkiye, Calabria, Corsica, and France; South—Iraq, Jordan, Egypt, Morocco) [25, 34, 35].

Subsequent expansions related to secondary features (*e.g.*, fat tailed sheep, wool sheep) occurred between 4000 to 3000 BCE in Europe [32, 33, 35]. Wool sheep likely originated in Southwest Asia, spread west and then across Europe from the Iberian Peninsula [30, 35]. Wool producing sheep partially replaced the initial domesticated sheep (hair sheep) and earlier breeds derived from previous introductions via different routes (Mediterranean versus Danubian). Evidence for this replacement is supported by a larger number of ancestral SNPs in modern domesticated sheep from southeastern to northwestern Europe (*i.e.*, the proportion of ancestral genetic components decrease when one is further away from the Fertile Crescent) [30, 32]. Italy and the Balkans were hubs for migrations of early and late domesticates such as wool sheep [32, 35].

An acute lack of genome-scale data concerning *B. melitensis* has impeded our understanding of its origin, evolution, and spread. Until recently, public databases contained few *B.*

*melitensis* genomes, limiting inferences on epidemiological history. This lack of data has since changed due to the recent focus on the Western Mediterranean clade of the bacterium in several studies [20, 27, 36]. Although it was hypothesized that this clade has a long-standing presence in the Mediterranean it was underrepresented in genomic databases [26]. A recent study increased the representation of the diversity within Western Mediterranean clade by adding 339 Italian strains from humans and ruminants to public databases [26]. A lack of public health support for genome surveillance may also explain in part why other regions, such as Africa, are similarly under-sampled [37, 38].

In contrast to the paucity of genome-scale data, the distribution of sequence types (ST)—the allelic profile of core housekeeping genes in a bacterial species—within *B. melitensis* is well defined [39, 40]. Sequence types provide a more nuanced view of the diversity within the *Brucella* genus. These STs can be used to quickly identify the phylogenetic group of an unknown *B. melitensis* strain as they are relatively well conserved. For example, there are five STs (ST 9, 11, 87, 88, and 89) present in the Western Mediterranean clade [40]. Biovars—groups of isolates divided by their biochemical properties—were previously used to categorize *Brucella* strains. The Western and Eastern Mediterranean phylogenetic groups, however, contain genomes arising from all three known *B. melitensis* biovars, indicating that STs are superior for discriminating between strains [40].

Only one partial ancient *B. melitensis* genome has been reconstructed to date, although its coverage is unfortunately too low to allow for genome scale analyses and comparison here (see S1 Appendix) [13]. The lack of ancient genomes impedes analyses attempting to determine the most recent common ancestor of extant *B. melitensis* as well as its overall phylogeographic history and gene content. Here, we describe the successful reconstruction of a high quality *B. melitensis* genome from the kidney stone of a beatified 14<sup>th</sup> CE friar, Sante Brancorsini (1343–1394 CE). The high resolution and precise date of this new genome along with its geographic location and associated time period make it particularly useful for addressing questions about the evolution of *B. melitensis* in the Mediterranean.

## Results

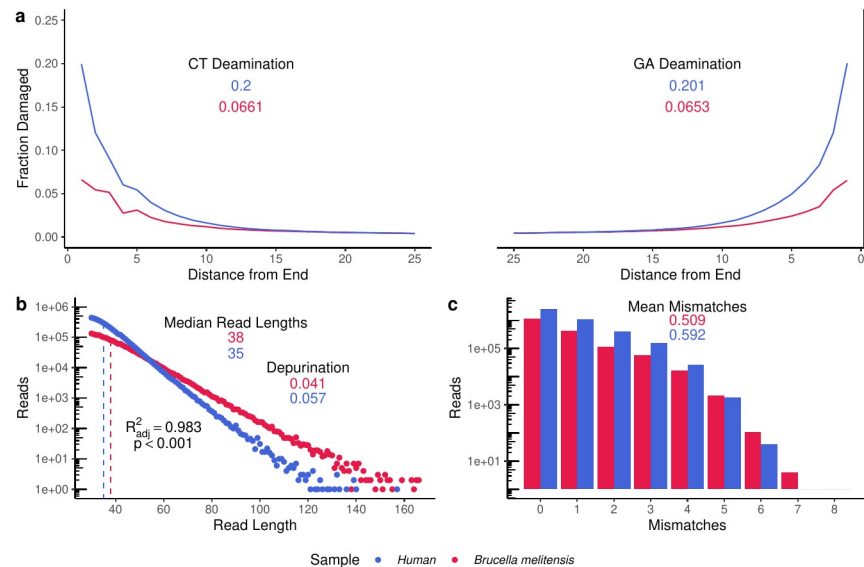
### Paleopathological overview

A paleopathological survey of the remains of Sante Brancorsini revealed several vertebrae with lesions characteristic of brucellosis [41]. The cavities caused by these lesions may have also caused spondylodiscitis—the inflammation of the nucleus pulposus—which while not typical of brucellosis, has also been reported [42]. Several calcified nodules were identified in the abdomen, which were later identified to be calcified kidney stones [43]. A more detailed look at the paleopathological examinations of Sante Brancorsini can be found in the original study [16].

### Validating the samples

A metagenomic analysis of DNA reads from the nodule (S1 Fig) libraries showed that a substantial fraction of the classified reads stemmed from the genus *Brucella* (2,777,352 reads or 19.501%) and *Homo sapiens* (8,576,720 or 67.740%). *Bradyrhizobium* was the dominant taxa identified in extraction blanks (1869 reads or ~ 25.363%), with *Brucella* being present in exceptionally low amounts (2 reads or ~ 0.03%) (S2 Fig).

The deamination pattern for reads mapping to *B. melitensis* appear to indicate authentic and relatively well preserved ancient DNA compared to the human DNA isolated from the same nodule (Fig 1a). This difference in preservation is supported by the fragment length distribution of the DNA reads, where *B. melitensis* has a longer median read length and a significantly (linear regression:  $p < 0.001$ ) shallower slope than its associated human DNA (Fig 1b).



**Fig 1. DNA authentication of the ancient *B. melitensis* genome.** a) Mapdamage plot showing the fraction of reads with deamination present on the 5' (left) and 3' (right) ends of the molecules. A hard masked hg38 human and the reference *B. melitensis* genomes were used to create this plot. b) FLDs of the reads mapped against their respective genomes. c) Edit distance plot for the mapped reads.

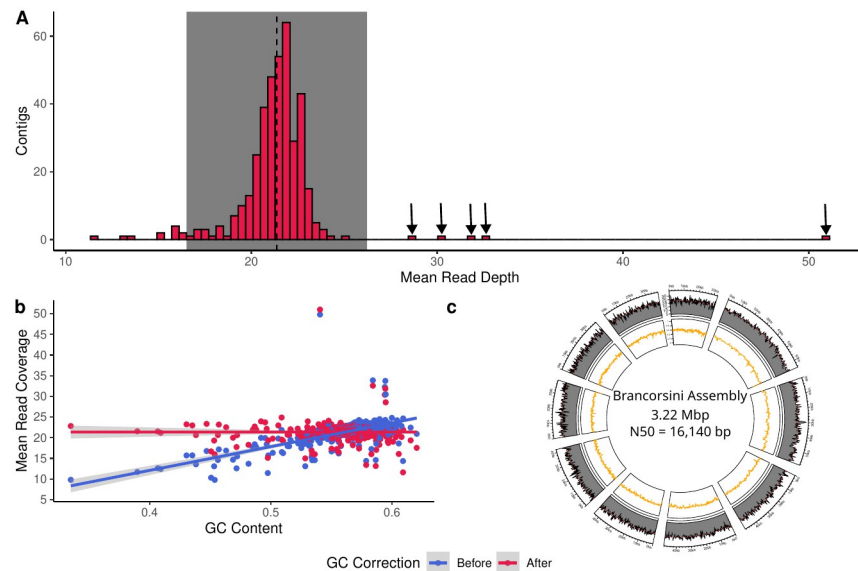
<https://doi.org/10.1371/journal.ppat.1011538.g001>

The edit distance distributions also support its authenticity, with the greater mean mismatch count for the human genome potentially caused by the higher reported deamination and other damage signals (Fig 1c). Discrepancies in preservation rates between the host and pathogen of interest are not a new phenomenon [44, 45] and it has been reported that GC content is a likely driver of depurination rates [46]. The difference in preservation rates were likely also magnified by the presence of a cell wall in *B. melitensis* [47].

### ***B. melitensis* genome reconstruction and phylogenetic analysis**

A total of 2996 genes were identified in the ancient *B. melitensis* assembly which had an overall mean coverage of 21.37 [21.13, 21.61]× (Fig 2a, Table 1). While GC correction did improve results (Fig 2b), five of the contigs still had significantly (two-sided t-test:  $p < 0.001$ ) deeper read depths than the rest of the assembly. A search for the genes associated with these higher copy number contigs identified the elongation factor *tuf* (*tuf\_1* in the pan-genome), an IS5 family transposase (*group\_1* in the pan-genome), and a porin related protein (*group\_1246* in the pan-genome) indicating that first two genes could potentially be transposons responsible for relocating the latter. The other two higher copy number contigs corresponded to fragments of an IS5 element and a tRNA gene per a *blastX* search of the RefSeq protein database (S1 File).

It is unlikely that these contigs originate from a phage or a plasmid as *Brucella* has been repeatedly shown to have a closed genome [48–50], furthermore, a search of the contigs with *MOB-suite* [51] did not identify any mobile elements. *Brucella* has not exhibit any signs of having recently undergone heterologous recombination [49, 50], thus the five higher copy number contigs must represent duplicated regions in the chromosome. This can be confirmed through a *blast* search of the nt database as there were multiple matches of these five contigs to the same accession (S1 File). Despite these outliers the overall coverage of the contigs is even with very few regions lacking reads (Fig 2c).



**Fig 2. Read coverage metrics of the *B. melitensis* pan-genome.** a) Distribution of mean contig coverages of the ancient *B. melitensis* assembly. The dashed line indicates the mean coverage while the shaded rectangle indicates  $\pm 2$  standard deviations from the mean. Arrows indicate contigs with a significantly greater read depth. b) GC bias correction results for the mapped contigs. c) Circos plot of the ten largest contigs in the ancient *B. melitensis* assembly. The outer ring indicates the mean read depth over a 0.5% sliding window whereas the inner ring is the GC content over the same period.

<https://doi.org/10.1371/journal.ppat.1011538.g002>

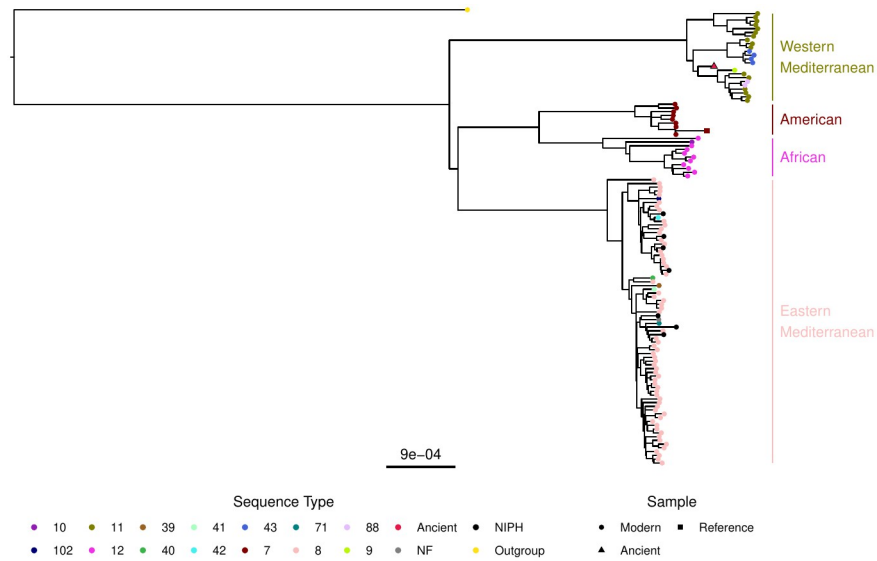
Maximum likelihood (ML) phylogenies were estimated for both the global and Western Mediterranean *B. melitensis* genome populations using 323 genomes obtained from NCBI on 2019–09–17. The global phylogeny—which used a trimmed dataset containing 17,510 core SNPs—firmly placed the ancient genome within the Western Mediterranean clade (Fig 3) recapitulating previously published trees [20, 26]. While the ancient strain is not ancestral to the entire phylogenetic group it is positioned basal to a specific sub-clade, indicating that these sub-groups were present in the 14<sup>th</sup> century. No temporal signal was present (linear regression:  $p = 0.740$ ,  $R^2_{adj} = -0.007$ ) in the global phylogeny when using TempEst [52].

We next performed a similar analysis of only the Western Mediterranean clade phylogeny using an alignment of 8262 core SNPs (Fig 4 and S5 Fig). The ancient genome fell in the same position as it did within the global phylogeny, and in this case there was strong evidence for a temporal (i.e. molecular clock) signal in both the ML (linear regression:

**Table 1. Summary Statistics of the ancient *B. melitensis* assembly.** An overview of the ancient *B. melitensis* assembly. The square brackets represent the 95% confidence interval.

Metric	<i>denovo</i> Assembly
Contigs	386
Assembly Length (bp)	3,220,175
N50 (bp)	16,140
Mean Contig Depth (x)	21.37 [21.13, 21.61]
GC Content (%)	56.1 [55.8, 56.5]
Genes Identified	2996
Core Genes Identified	2826
Proportion of Core (%)	97.5

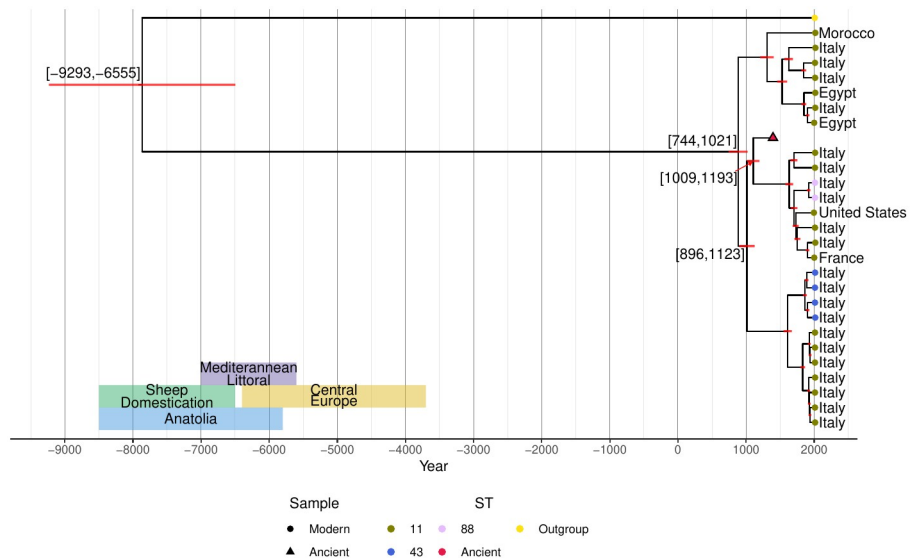
<https://doi.org/10.1371/journal.ppat.1011538.t001>



**Fig 3. Maximum Likelihood Phylogeny of *B. melitensis*.** Core SNP maximum likelihood phylogeny of the global distribution of *B. melitensis*. *B. abortus* 2308 was the outgroup used to root the tree while *B. melitensis* 16M is the reference genome.

<https://doi.org/10.1371/journal.ppat.1011538.g003>

$p < 0.001$ ,  $R^2_{adj} = 0.76$ ) and Bayesian phylogenies (log Bayes factor: 59.75). Under a strict molecular clock, the mean nucleotide substitution rate for the Western Mediterranean clade was  $9.322 \cdot 10^{-8}$  [ $8.062 \cdot 10^{-8}$ ,  $1.062 \cdot 10^{-7}$ ] subs/site/year, giving a mean divergence date of 1105 [1009, 1193] CE for the ancient strain. Under this model the diversification of the Western Mediterranean clade occurred about 884 [744, 1021] CE and that speciation from *B.*



**Fig 4. Phylogenetic analysis of the Western Mediterranean clade.** Core SNP Bayesian phylogeny of the Western Mediterranean phylogenetic group. The red bars at the nodes indicate the 95% highest posterior density of the estimated date for a node whereas the text indicates the boundaries of the interval. The text along with coloured shaded regions indicate the estimated dates for sheep domestication and migration to specific regions [105]. When available, the genomes are labelled by their country of isolation. *B. abortus* 2308 was the outgroup and reference genome.

<https://doi.org/10.1371/journal.ppat.1011538.g004>

*abortus* likely happened about 7865 [9293, 6555] BCE. We obtained very similar results using a relaxed molecular clock (i.e. a nucleotide substitution rate of  $9.463 \cdot 10^{-8}$  [ $7.625 \cdot 10^{-8}$ ,  $1.134 \cdot 10^{-7}$ ] subs/site/year) suggesting that these estimates are robust to rate variation.

## Functional analysis

The majority of the genes identified in the ancient strain (2826 of 2966) stemmed from the core genome of *B. melitensis*. Despite this, 72 core genes were absent from the ancient strain. A COG database search identified the functional categories for most of the missing core genes with a large number involved in metabolism or biogenesis (Table 2). These missing genes may have impacted the virulence of the ancient strain. Several of these genes were involved in the *Brucella* virulence related regulatory and sensory (BvrR/BvrS) mechanism responsible for directing the expression of several outer membrane proteins [53, 54]. The same results were also found with a STRING [55] database search (S1 File). A similar search for known *B. melitensis* virulence factors [56] was also performed, revealing that the ancient genome was likely pathogenic as genes from each virulence category were detected in the ancient strain (Table 3). AMR genes were also analyzed using the CARD database [57], although only four were identified in the ancient genome (Table 4) three of which were efflux pumps. Thus, the ancient genome did not contain any significant resistance mechanisms [58, 59].

The presence-absence variation (PAV) analysis of the accessory genome across modern and ancient genomes recapitulates the results of our phylogenetic analyses (Fig 5). Modern *B. melitensis* accessory genomes grouped into their respective geographic regions (Africa, the Americas, Eastern and Western Mediterranean). The ancient strain was similarly placed close to the Western Mediterranean phylogenetic group, again suggesting that it was an early member of this group. However, an exception to this discrete grouping was the intermixing of the American and African phylogenetic groups. This mixture of groups also supported by a SNP multiple correspondence analysis (MCA) suggests that these two phylogenetic groups are not representative of distinct functional differences (S7 Fig). The same reasoning applies to the

**Table 2. Identified COG categories of core genes absent from our ancient strain.** Genes were identified through a `blastx` search against the 2020 COG database. The gene counts are not exclusive as a gene may be identified in multiple categories.

Function	Gene Count
Cell wall/membrane/envelope biogenesis	8
Lipid transport and metabolism	5
Translation, ribosomal structure and biogenesis	4
Amino acid transport and metabolism	3
Replication, recombination and repair	3
Function unknown	3
Signal transduction mechanisms	3
Nucleotide transport and metabolism	2
Carbohydrate transport and metabolism	2
Coenzyme transport and metabolism	2
Transcription	2
General function prediction only	2
Posttranslational modification, protein turnover, chaperones	1
Secondary metabolites biosynthesis, transport and catabolism	1
Defense mechanisms	1
Mobilome: prophages, transposons	1

<https://doi.org/10.1371/journal.ppat.1011538.t002>



**Table 3. *Brucella* virulence factors present in our ancient strain.** Genes were found through a combination of gene name searches, blasting the COG database, and blasting against a specific gene.

Virulence Category	Genes
Urease	40
Cytochrome Oxidase	11
Lipopolysaccharide	11
Type IV Secretion System	10
BvrR/BvrS System	9
Superoxide Dismutase and Catalase	3
Alkyl Hydroperoxide Reductase	2
Cyclic $\beta$ -1–2-glucans	2
Nitric Oxide Reductase	2
Base Excision Repair	2
<i>Brucella</i> Virulence Factor A	1

<https://doi.org/10.1371/journal.ppat.1011538.t003>

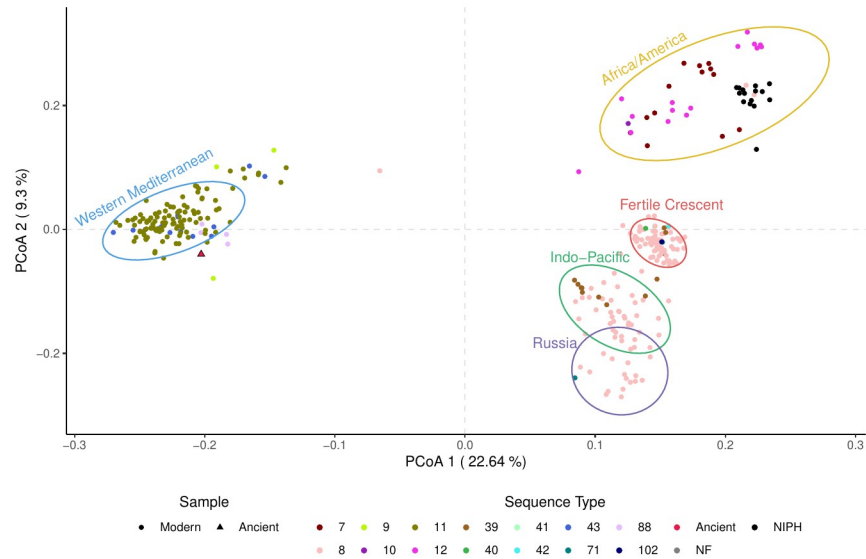
wayward sequence type 8 (ST) genomes clustering with the African and American strains. These strains were first identified as part of a Norwegian Institute of Public Health (NIPH) study analyzing the breadth of *B. melitensis* cases reported in the country [61]. They form a group of infections originating from Afghanistan, Georgia, Iraq, Norway, Syria, and Türkiye. While both their ST and core SNP phylogeny indicate that these strains are members of the Eastern Mediterranean clade, the PAV of the accessory genome places them within the African and American cluster. This further suggests that STs and phylogenetic clades do not convey sufficient information to convey functional differences between strains.

Naively clustering *B. melitensis* accessory genome strains (PAV) based on a principal coordinate analysis (PCoA) showed that the Eastern Mediterranean clade formed three distinct clusters (Fig 6). The first of these clusters contains the majority of genomes originating from the Fertile Crescent and is also primarily part of a sub-clade of the Eastern Mediterranean phylogenetic group in the core SNP phylogeny. There is substantial spillover of this cluster with the rest of the phylogeny, although they largely consist of genomes isolated outside of the Fertile Crescent (*i.e.* China and India). As it is hypothesized that *B. melitensis* originated within the Fertile Crescent [26], these results are of note as they suggest the presence of a distinct, stable population at its starting point. Also, the wayward ST 8 genomes in the African/American cluster are phylogenetically placed with the Fertile Crescent strains in the global core SNP phylogeny and form a sub-cluster within the African/American group. That most of these wayward strains were associated with infections in the Fertile Crescent suggests that loss of ten specific accessory genes is enough to be functionally similar to the African/American group (S1 File, S8 Fig). The other two clusters represent the remaining Eastern Mediterranean phylogenetic group, with one representing the Indo-Pacific region while the other consisting primarily of genomes isolated from Russia.

**Table 4. *Brucella* antibiotic resistance genes.** Genes were identified by searching the protein homolog models from the CARD database [57]. Gene name indicates the label assigned by Roary [60] whereas AMR Gene is the corresponding gene in the CARD database.

Gene Name	AMR Gene	ARO ID	Resistance Mechanism
group_2536	<i>mdtP</i>	3003550	Antibiotic efflux
group_88	<i>tetA</i>	3004639	Antibiotic efflux
group_970	<i>KpnH</i>	3004597	Antibiotic efflux
<i>mprF</i>	<i>mprF</i>	3003772	Antibiotic target alteration

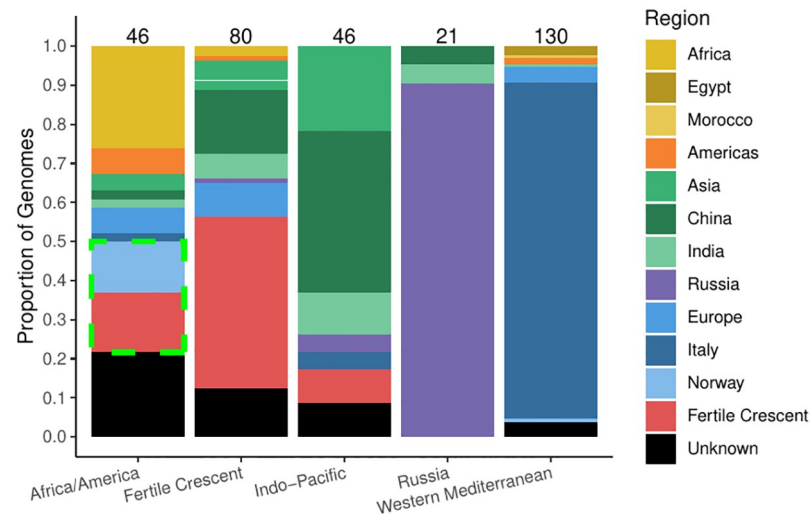
<https://doi.org/10.1371/journal.ppat.1011538.t004>



**Fig 5. Principal Coordinate Analyses (PCoA) of accessory gene PAV.** Sequence types are identified using SRST2, although the outlier ST 8 genomes (from NIPH) are in black. The genomes were also naively clustered using the clara algorithm with the number of clusters selected by searching for the local maximum in the silhouette plots.

<https://doi.org/10.1371/journal.ppat.1011538.g005>

The five genome clusters are the result of the PAV of a relatively small number of accessory genes (Table 5, S8 Fig). These genes are uniquely present or absent in their respective geographic clusters, and thus have an overly large impact on the positioning of the genomes in the PCoA. This is true regardless of whether or not the genes are filtered by their overall presence in the species. Furthermore, these genes are enough to impart additional gains or loss of function to the genomes (S1 File). The Russian and Western Mediterranean clusters are prime examples as they contain a large number of genes uniquely present in their genomes. The genes present in the Western Mediterranean impart a unique set of transferases and while the



**Fig 6. Isolation sources for the naive accessory gene PAV clusters.** Colours for the regions correspond with the PCoA clusters. The wayward ST 8 genomes are indicated by dashed green lines.

<https://doi.org/10.1371/journal.ppat.1011538.g006>

**Table 5. Number of accessory genes uniquely present or absent in the accessory genome clusters.** Results are shown for both the entire set of accessory genes as well as those present in at least five genomes.

Region	All Accessory Genes		Genes in $\geq 5$ Genomes	
	Genes Absent	Genes Present	Genes Absent	Genes Present
Africa/America	6	20	6	10
Fertile Crescent	5	4	5	1
Indo-Pacific	0	20	0	0
Russia	8	109	8	36
Western Mediterranean	10	44	10	22

<https://doi.org/10.1371/journal.ppat.1011538.t005>

Russian strains have a separate set of 30S ribosomal proteins and metal transporters (S1 File). The presence of these genes may be sufficient to determine the geographic region of a strain, especially given the clonal nature of *Brucella* [39].

## Discussion

### Authenticating the ancient *B. melitensis* genome

Metagenomic analyses indicated that the kidney stone from Sante Brancorsini contained *Brucella* sp. DNA (S2 Fig). The DNA was confirmed to arise from *B. melitensis* due to the high and relatively even coverage of reads across the assembly (21.37 [21.13, 21.61]x) (Fig 2a and 2c) and low number of mismatches (*i.e.*, edit distance) when mapping against a *B. melitensis* (GCF\_000007125) genome (Fig 1c and S9 Fig). Furthermore, approximately 92% of the recovered genes stem from the *Brucella* core genome.

In addition, the ancient *B. melitensis* and human DNA reads possess clear signals of deamination (Fig 1a) and depurination (Fig 1b) characteristic of ancient DNA damage. The fragment lengths in the mapped reads form an exponential distribution as expected (Fig 1b) [62].

*Brucella* is not a commensal genus of bacteria [63] and our identification of key virulence genes [50, 64, 65] confirms that the ancient genome was indeed pathogenic. The strain lacked resistance genes apart from the detection of efflux pumps (Table 4). Multi-drug efflux pumps, while also being involved in other mechanisms, do not provide strong antibiotic resistance on their own [50, 58, 59]. Modern *Brucella* genomes, while rare, can demonstrate signs of resistance to azithromycin [66], rifampicin [61], streptomycin, sulfamethoxazole, and ceftriaxone while susceptibility is decreasing for other antimicrobials [67].

### Time-scaled speciation of *Brucella*

The presence of several *B. melitensis* sequence types circulating in the modern Mediterranean littoral [19, 20, 26, 27] has led to the hypothesis that *B. melitensis* either originated in this region or was endemic, especially in Italy, for an extended period of time [19, 20]. The results of our Western Mediterranean *B. melitensis* phylogeny support the latter theory, which has also been supported by other recent studies [26, 27]. Our speciation date of *B. melitensis* is consistent with previously published dates of sheep and goat domestication in the Fertile Crescent [25]. Furthermore, with previous studies indicating that the Western Mediterranean clade diverged during the 4<sup>th</sup> or 5<sup>th</sup> millennium BCE [26, 33] it is likely that *B. melitensis* arose in the Fertile Crescent and was brought to the Western Mediterranean during the shift from hunter-gather practices to pastoralism [31, 33, 35, 68].

Although the process of moving from a hunter-gatherer society to an agriculturalist culture was complex, the spread was most often a shift to agro-pastoralism in the coastal regions of the Mediterranean with several colonies sprouting across the region. In contrast, the spread of the

Neolithic package [25, 69, 70] was much slower in inland locations. Generally, this variation in the speed of adoption meant that the coastal areas were more reliant on domestic livestock whereas the inland regions used a mix of domestication and hunter-gather practices during a similar time [25]. The process of adopting agricultural practices in Europe was completed by the 4<sup>th</sup> millennium BCE [71] with farming practices intensifying from nomadic pastoralism to transhumance. This transition from mixed hunter-gatherer and agricultural practices including transhumance pastoralism could have impacted the divergence of the Western Mediterranean *B. melitensis* clade during this time.

Specialization in wool sheep occurred between 5000 and 4000 BCE in the Fertile Crescent with the eventual replacement of the first domesticates (hair sheep) beginning approximately 4000 BCE. These sheep eventually spread to Europe sometime between 3000 and 2000 BCE [30, 32, 33, 35, 72]. Wool was a major innovation in Europe during the 2<sup>nd</sup> millennium BCE. Historical sources state that wool production thrived on the coasts of the northeastern Mediterranean [73]. The spread of wool sheep from Asia to Europe along with the level of production in the latter indicates there was a substantial amount of animal movement in these regions during this time [73]. The movement of wool sheep could have also contributed to the divergence of the Western Mediterranean clade as wool specialization also occurred during the same time as our estimated dates of divergence.

The placement of the *B. melitensis* genome in the Western Mediterranean clade supports the hypothesis of an early expansion starting in the 4<sup>th</sup> millennium BCE. Such results propose that the migration of sheep and goat populations are the driving force in the diversification of *B. melitensis*. This is especially true given the transportation of sheep and goats across Eurasia throughout their history as domesticates [30, 32, 73]. In combination with the clonal nature of *B. melitensis* [38], its highly conservative genome [48, 50], and the sequence typing results of the ancient genome (S1 Appendix) it is likely the phylogenetic structure is a direct result of these early migrations.

### Potential divisions in the Eastern Mediterranean phylogeny

The PAV of the accessory genes largely recapitulates the global *B. melitensis* phylogeny (Fig 4). The ancient genome shares the same general functions as strains from the Western Mediterranean clade. As *B. melitensis* rarely undergoes horizontal gene transfer [48], these results confirm that the Western Mediterranean clade has remained functionally stable for nearly 700 years (S8 and S10 Figs). The complete complement of virulence genes serves as an example of this long-term stability.

The information gained by analyzing the global phylogeny and the accessory PCoA together provide some insight into the evolution of this genus. A complete separation of the Western Mediterranean strains from all other strains in both analyses corroborates the hypothesis that these genomes were some of the first to diverge. Furthermore, the intermixing of the African and American strains suggests that they originated from a single clade that recently diverged. The Eastern Mediterranean phylogenetic group fell into three clusters that roughly correspond to different geographic regions within this area (Fig 6). The largest of the three clusters is a group of strains primarily from the Fertile Crescent (Iran, Iraq, Israel, Jordan, Kuwait, Syria, and Türkiye) and is of particular interest as it is thought that the practice of sheep husbandry originated in this region [25]. If this is the case, it is possible that this cluster represents the original population and accessory genome content of *B. melitensis*, as these strains have the fewest number of unique and overall accessory genes (S11 Fig). The next largest “Eastern Mediterranean” cluster comprises strains either originating from the Indo-Pacific region with strains primarily from China, India, and Southeast Asia while the final group are

genomes from Russia. While the silhouette algorithm fits five clusters, it is likely that the latter two groups (Indo-Pacific and Russia) are part of a larger cluster with global sampling bias artificially splitting them (Fig 6). If not replaced entirely, the “Eastern Mediterranean” group should at least be renamed as to better represent the full extent of its geographic spread.

Of particular interest are the regions of disconnection between the global phylogeny and the accessory genomes. While both data sets agree that the Western Mediterranean clade was the first to diverge from the rest of *B. melitensis* and that the African and American clades are closely related, the Eastern Mediterranean strains complicate this simple scenario. This issue is most obvious with the strains isolated and studied by the NIPH [61]. While these strains cluster with the American and African strains in the PCoA, in the phylogeny they are placed among those from the Fertile Crescent. This disconnect suggests that assigning function to either the ST or the phylogenetic group of strains can be inaccurate and potentially misleading.

The widespread movement of infected animals across regional, national, and international borders make the phylogenetics of *Brucella* challenging at both local and larger scales [19]. Our analyses, using the first high quality ancient *B. melitensis* genome, and applying core SNP and ST methods to compare with the now numerous contemporary genomes, provides new insights into the evolution of *B. melitensis* in the Mediterranean. We have confirmed that the Western Mediterranean lineage was present in the region since at least the 14<sup>th</sup> century CE, and we have narrowed the estimated time-scale for the expansion of the West Mediterranean clade thanks to strong date priors using the ancient genome [26, 27]. More genomes—both ancient and modern—from a wider geographic range are required to understand the nuances of how *B. melitensis* evolved and was disseminated across the Mediterranean and eventually the globe.

## Materials and methods

### Ethics statement

Ethics approval was obtained from the Hamilton Integrated Research Ethics Board associated with McMaster University (Project Number: 13841-T). Dr. Gino Fornaciari was given written legal and ethical permission to study Sante Brancorsini by the Congregation of Saints of the Catholic Church in Rome.

### Context

The sample used for this study is a kidney stone obtained from the mummified remains of Sante Brancorsini, an Italian noble born in 1343 CE [74]. Brancorsini became a Franciscan friar after reportedly killing someone in self-defence at the age of twenty [74, 75]. He died in 1394 and was initially buried in a common grave on church grounds. The remains of Brancorsini were then moved to the convent wall in 1395 after a lily grew on his grave. Brancorsini was beatified in 1769 and his remains were then transferred below an altar in a chapel named in his honour. Dr. Gino Fornaciari examined and sampled the partially skeletonized remains of Brancorsini in 1994 [75].

### Sample processing

A calcified nodule weighing 129.1 mg identified as a kidney stone was sent to the McMaster Ancient DNA Centre (McMaster University, Hamilton, Canada) for aDNA analysis. The entire stone was sampled in clean-room facilities and underwent six rounds of demineralization and digestion [76]. The supernatant from rounds 3 and 4 underwent purification,

extraction, and double-stranded library preparation following previously published aDNA procedures [77] with previously published modifications [78]. Two further modifications were made to the extraction protocols: a fifteen hour ligation period and MinElute purification instead of a bead-based method. The kidney stone was processed alongside extraction and library blanks. The libraries were then sequenced on an Illumina HiSeq 1500 platform with 90 bp paired-end reads at the Farncombe Family Digestive Health Research Institute. The library was further sequenced upon the discovery of *B. melitensis* during a preliminary metagenomic analysis. Additional information concerning the processing of the kidney stone can be found in the original study [16].

### Bioinformatics analysis

Adapter detection, read trimming, and merging were performed by `fastp` [79]. The data generated over multiple sequencing lanes and runs were pooled when applicable, and `Kraken 2` [80] was used to determine the overall metagenomic composition of the samples. A standard `Kraken 2` database with some slight modifications—a kmer size of 25bp and no minimizer—to account for the smaller read lengths of aDNA fragments was used [45]. The trimmed reads underwent string deduplication via `prinseq` [81] prior to the metagenomic analysis.

The reads were then mapped against the human genome—GRCh38 [82]—with `BWA aln` [83] using previously published parameters [84]. A minimum length of 30bp and mapping quality of 30 was required for a read to be successfully matched. Unsuccessful fragments were then separately mapped to the *B. melitensis* bv. 1 strain 16M (GCF\_000007125) genome to confirm its presence and determine the authenticity of the reads. `SPAdes 3.14.1` was used to *de novo* assemble the unsuccessful human fragments using kmer sizes of 15, 21, and 29 to account for the short read lengths [85]. Contigs smaller than 200bp were removed from the assembly with the rest being analyzed by `BlastN` [86] using the `nt` database to identify *B. melitensis* fragments. The `blast` output was analyzed using a previously published script and required that at least 50% of the identifications were of the same taxon [16]. The contigs were then analyzed with `MOB-suite` to identify any potential plasmids or mobile genetic elements [51]. A pan-genome was then created with 323 *B. melitensis* genomes obtained from NCBI on 2019-09-17 using `NCBImeta` [87] along with the ancient assembly. These sequences were annotated with `Prokka` [88] using proteins from GCF\_000007125 as the initial standard. The resulting genome annotations were then compiled by `Roary` [60] to create the *B. melitensis* pan-genome and calculate the PAV matrix. See the supplemental information for the pan-genome mapping results (S1 Appendix)

Successfully mapped human and *B. melitensis* reads were deduplicated based on their coordinates with `bam-rmdup` (<https://bitbucket.org/ustenzel/biohazard-tools/src/master/>). `mapDamage 2.0` [89] was used to estimate the level of deamination present across the mapped reads. Fragment length distributions (FLD) and mapping mismatches were also extracted from the mapped reads. The read depths were normalized based on the GC content of their respective genes to control for biases introduced during sequencing and DNA preservation [46, 90]. This was done by calculating the GC content over a sliding window and creating a linear regression in which the top 1% of residuals were filtered out. Once this was done, a correction factor was calculated by subtracting the mean of sequencing depths by the predicted value from the regression. This correction factor was then added to the original sequencing depth value [91].

### Phylogenetic analysis

A core SNP alignment of the 323 *B. melitensis* genomes from NCBI and the ancient *B. melitensis* assembly was created using `Snippy` (<https://github.com/tseemann/snippy>) with *B. abortus*

2308 (GCF\_000054005) as the outgroup and *B. melitensis* 16M (GCF\_000007125) as the reference strain. Recombinant regions were removed with Gubbins [93] and the resulting alignment used to create a maximum-likelihood (ML) phylogeny with IQ-TREE 2 [93]. The nucleotide substitution model (TVM+F+ASC+R5) was selected using ModelFinder [94] along with the added requirement that ascertainment bias correction was included. The initial phylogeny was then trimmed using TreeMmer [95] down to a root-to-tip length (RTL) of 95% to remove redundant genomes. An additional eight genomes were removed due to long branches and unclear isolation dates. The remaining 127 genomes then underwent the same phylogenetic pipeline, with Gubbins removing 2485 bp (12.42% from the alignment) prior to creating the phylogeny. Sequences were labelled based on their ST which was identified using SRST2 [96] and the *Brucella* MLST panel from PubMLST [97].

After identifying that the ancient genome was part of the Western Mediterranean clade, a second set of phylogenies only comprising the relevant strains from this location were estimated and rooted using *B. abortus* 2308. The Gubbins analysis removed 582 bp (6.58%) of the core SNP alignment. While the same settings were used for this phylogeny, a different nucleotide substitution model (K3Pu+F+ASC) was selected with ModelFinder [94] where K3Pu is the unequal bases version of K81 [98].

To determine if the sequence data contained any temporal signature, sampling dates (or sequencing dates when the former was unavailable) were assigned to the pruned phylogenies and a Root-to-tip (RTT) regression was performed for the global and Western Mediterranean clade phylogenies using TempEst [52]. A least squares dating analysis with LSD2 was also used to confirm the temporal signal within the phylogenies through previously published settings [45, 99]. This was done by running LSD2 on the individual trees from each bootstrap and summarizing the results with treeannotator [100]. The heights of each node set by the mean of the bootstrap distribution. The resulting phylogeny was then used to plot the dated trees.

BEAST 1.10.4 (Prerelease) [100] was also used to estimate Bayesian phylogenies of the trimmed global and Western Mediterranean phylogenies. A strict molecular clock model along with a coalescent exponential demographic model was used to create the phylogeny, in line with previous studies [26, 27]. Chain lengths varying between 50 million and 1 billion generations with a 20% burn-in was used to ensure that the estimated sample sizes were above 200. While invariant sites were excluded from the core SNP alignment, the number of invariant sites were manually inserted into the XML files as a constant pattern to ensure that ascertainment bias did not affect the clock model. The Bayesian phylogenies were tested for a temporal signature using the BETS method [101] by comparing the marginal likelihood estimates of a heterochronous and isochronous phylogeny (S1 Appendix). To ensure that the log marginal likelihoods for the isochronous phylogenies could be calculated, the clock rate priors were deleted from the XML files.

## Functional analysis

A gene PAV analysis was performed using the matrix calculated by Roary [60]. A 99% core genome was calculated and all the genes were functionally annotated using the 2020 COG database [102]. Core genes absent from the ancient genome were annotated using the STRING database [55]. *Brucella* virulence factors [56] were also identified to determine if the relevant genes were present in the ancient genomes. This was done by both searching the gene names in the pan-genome as well as by identifying the COG functions of the ambiguous genes. A search for AMR genes in the ancient *B. melitensis* assembly was also performed using the Resistance Gene Identifier [57].

We performed a PAV analysis on the accessory genes of the *B. melitensis* pan-genome. The PAV metrics of the strains were transformed into a binary distance matrix and clustered using a PCoA. To ensure that redundant genes were not included an exclusion criterion was created: genes not present in at least five genomes were removed from the analysis. Following this, naive clustering of the *B. melitensis* genomes using their PCoA coordinates was done with the `clara` algorithm [103]. A silhouette plot [104] was used to determine the optimal number of clusters. A multiple correspondence analysis (MCA) plot was also created using the SNPs of accessory genes from modern and ancient genomes relative to the reference *B. melitensis* strain (GCF\_000007125). Singleton SNPs were removed prior to creating the MCA as they would not provide any meaningful information.

## Supporting information

**S1 Fig. Kidney stone from Sante Brancorsini from which aDNA was obtained.** The diameter of the stone is 0.5mm.

(TIFF)

**S2 Fig. Metagenomic profiles of currently known ancient samples.** Metagenomic profiles were calculated at the genus taxonomic level. Numbers above each column indicate the total number of reads identified at the genus level or lower. The Geridu sample comes from a previously published ancient *B. melitensis* genome [13].

(TIF)

**S3 Fig. Global *B. melitensis* phylogeny.** Core SNP Bayesian phylogeny of the global *B. melitensis* genome diversity. The red bars at the nodes indicate the 95% highest posterior density of the estimated date for a node whereas the text indicates the boundaries of the interval. The text along with coloured shaded regions indicate the estimated dates for sheep domestication and migration to specific regions [105]. When available, the genomes are labelled by their country of isolation. *B. abortus* 2308 was the outgroup whereas *B. melitensis* 16M the reference genome.

(TIF)

**S4 Fig. Determining the temporal signal of the Global *B. melitensis* ML Phylogeny.** Root-to-tip distances were calculated using `TempEst` [52] by selecting the best root using the heuristic residual mean squared option.

(TIF)

**S5 Fig. Maximum Likelihood Phylogeny the Western Mediterranean clade.** Core SNP ML phylogeny of the Western Mediterranean phylogenetic group. The red bars at the nodes indicate the 95% highest posterior density of the estimated date for a node whereas the text indicates the boundaries of the interval. The text along with coloured shaded regions indicate the estimated dates for sheep domestication and migration to specific regions [105]. When available, the genomes are labelled by their country of isolation. *B. abortus* 2308 was the outgroup whereas *B. melitensis* 16M the reference genome.

(TIF)

**S6 Fig. Determining the temporal signal of the Western Mediterranean ML phylogeny.** Root-to-tip distances were calculated using `TempEst` [52] by selecting the best root using the heuristic residual mean squared option.

(TIF)

**S7 Fig. MCA of SNPs found in the accessory genome.** SNPs were created using the VCF files from the core SNP alignment for the global phylogeny. Singletons were removed from the



analysis prior to the MCA. **a)** contains the first two axes of the MCA whereas **b)** consists of the first and third.

(TIF)

**S8 Fig. Heatmap of PAV in the accessory genome of *B. melitensis*.** Red indicates that a gene is present whereas blue is absent. Rows and columns were clustered using the `Ward.D2` algorithm. The horizontal yellow lines represent the wayward ST 8 genomes whereas the vertical lines indicate genes that are uniquely present or absent in the Africa/America cluster.

(TIF)

**S9 Fig. Coverage plot of NC\_00317.1 and NC\_00318.1.** Coverage plots for the first (NC\_003317) and second (NC\_003318) chromosomes in *B. melitensis* (GCF\_000007125) for our **a)** ancient genome and **b)** a previously published [13] sample. A sliding window of 0.1% was used for illustrative purposes. The first track indicates the mean read depth with the red line representing the overall mean. The second track indicates the number of SNPs identified in the same window while the third is the GC content.

(TIF)

**S10 Fig. Heatmap of PAV in the Core Genome of *B. melitensis*.** Red indicates that a gene is present whereas blue is absent. Rows and columns were clustered using the `Ward.D2` algorithm.

(TIF)

**S11 Fig. Euler diagram of accessory genes in each PCoA cluster.** Genes were identified as part of a cluster if it was present in more than genome. The counts indicate how many genes are found by the intersecting groups.

(TIF)

**S12 Fig. Maximum Likelihood Phylogeny of the Western Mediterranean clade with all known ancient genomes.**

(TIF)

**S13 Fig. Pan-genome mapping thresholds. a)** Determining the number of additional genes identified at each CV threshold. The dashed red line indicates the chosen threshold ( $\leq 1.5$ ) for the analysis. **b)** Distribution of mean gene coverages for the mapping of the ancient *B. melitensis* strain to the pan-genome. The dashed black line indicates the mean coverage while the blue and orange lines represent  $-2\sigma$  and  $-3\sigma$  from the mean.

(TIF)

**S14 Fig. Principal Coordinate Analyses (PCoA) of accessory gene PAV from the Mapping Methodology.** Sequence types are identified using SRST2, although the outlier ST 8 genomes (from NIPH) are in black.

(TIF)

**S1 File. STRING results and missing core gene BLASTX results.** NIPH Compared to Africa and America: Cells in green represent genes present in Africa/America Cluster, red is otherwise. High Coverage Contigs CNV: Cells in green represent accessions with more than two hits; red indicates accessions with only one hit. High Coverage Contigs RefSeq Protein: Yellow cells indicate cases where multiple definitions were identified. Accessory Genes by Cluster: Green indicates that a gene is present whereas red is absent. Yellow cells indicate cases where multiple definitions were identified.

(XLSX)

**S1 Appendix. Supplemental information.**  
(PDF)

## Acknowledgments

We thank current and former members of the McMaster Ancient DNA Centre for their input into this project. Finally, we thank Father Giancarlo Mandolini O.F.M., the Rector for the convent of Santa Maria (Scotaneto at Mombaroccio) and the Congregation of Saints of the Catholic Church in Rome, who kindly gave Dr. Fornaciari legal and ethical permission to study the remains of Sante Brancorsini.

## Author Contributions

**Conceptualization:** Tracy L. Prowse, Gino Fornaciari, Hendrik N. Poinar.

**Data curation:** Katherine Eaton.

**Formal analysis:** George S. Long.

**Funding acquisition:** Hendrik N. Poinar.

**Investigation:** George S. Long, Jessica Hider, Ana T. Duggan, Jennifer Klunk, Katherine Eaton, Emil Karpinski, Valentina Giuffra, Luca Ventura, Edward C. Holmes.

**Methodology:** George S. Long, Jennifer Klunk.

**Resources:** Gino Fornaciari, G. Brian Golding.

**Software:** George S. Long.

**Supervision:** Tracy L. Prowse, G. Brian Golding, Hendrik N. Poinar.

**Validation:** Edward C. Holmes.

**Visualization:** George S. Long.

**Writing – original draft:** George S. Long, Jessica Hider.

**Writing – review & editing:** George S. Long, Jessica Hider, Ana T. Duggan, Jennifer Klunk, Emil Karpinski, Valentina Giuffra, Luca Ventura, Tracy L. Prowse, Antonio Fornaciari, Gino Fornaciari, Edward C. Holmes, G. Brian Golding, Hendrik N. Poinar.

## References

1. Pappas G, Akritidis N, Bosilkovski M, Tsianos E. Brucellosis. *N Engl J Med*. 2005; 352:2325–2336. <https://doi.org/10.1056/NEJMra050570> PMID: 15930423
2. Hull NC, Schumaker BA. Comparisons of brucellosis between human and veterinary medicine. *Infect Ecol Epidemiol*. 2018; 8:1500846. <https://doi.org/10.1080/20008686.2018.1500846> PMID: 30083304
3. Pappas G, Papadimitriou P, Akritidis N, Christou L, Tsianos EV. The new global map of human brucellosis. *Lancet Infect Dis*. 2006; 6:91–99. [https://doi.org/10.1016/S1473-3099\(06\)70382-6](https://doi.org/10.1016/S1473-3099(06)70382-6) PMID: 16439329
4. Moreno E. Retrospective and prospective perspectives on zoonotic brucellosis. *Front Microbiol*. 2014; 5:213. <https://doi.org/10.3389/fmicb.2014.00213> PMID: 24860561
5. Kam YW, Ahmed MY, Amrun SN, Lee B, Refaie T, Elgizouli K, et al. Systematic analysis of disease-specific immunological signatures in patients with febrile illness from Saudi Arabia. *Clin Transl Immunology*. 2020; 9:e1163. <https://doi.org/10.1002/cti2.1163> PMID: 32864128
6. European Food Safety Authority E. The European Union One Health 2020 Zoonoses Report. *EFSA J*. 2021; 19:e06971. <https://doi.org/10.2903/j.efsa.2021.6971>

7. D'Anastasio R, Staniscia T, Milia ML, Manzoli L, Capasso L. Origin, evolution and paleoepidemiology of brucellosis. *Epidemiol Infect.* 2011; 139:149–156. <https://doi.org/10.1017/S095026881000097X> PMID: 20447329
8. Capasso L. Brucellosis at Herculaneum (79 AD). *International Journal of Osteoarchaeology.* 1999; 9(5):277–288. [https://doi.org/10.1002/\(SICI\)1099-1212\(199909/10\)9:5%3C277::AID-OA489%3E3.0.CO;2-0](https://doi.org/10.1002/(SICI)1099-1212(199909/10)9:5%3C277::AID-OA489%3E3.0.CO;2-0)
9. Ortner DJ. Identification of pathological conditions in human skeletal remains. Academic press; 2003.
10. Turan H, Serephanoglu K, Karadeli E, Togan T, Arslan H. Osteoarticular involvement among 202 brucellosis cases identified in Central Anatolia region of Turkey. *Intern Med.* 2011; 50:421–428. <https://doi.org/10.2169/internalmedicine.50.4700> PMID: 21372451
11. Buzgan T, Karahocagil MK, Irmak H, Baran AI, Karsen H, Evirgen O, et al. Clinical manifestations and complications in 1028 cases of brucellosis: a retrospective evaluation and review of the literature. *Int J Infect Dis.* 2010; 14:e469–78. <https://doi.org/10.1016/j.ijid.2009.06.031> PMID: 19910232
12. Bendrey R, Cassidy JP, Fournié G, Merrett DC, Oakes RHA, Taylor GM. Approaching ancient disease from a One Health perspective: Interdisciplinary review for the investigation of zoonotic brucellosis. *International Journal of Osteoarchaeology.* 2020; 30(1):99–108. <https://doi.org/10.1002/oa.2837>
13. Kay GL, Sergeant MJ, Giuffra V, Bandiera P, Milanese M, Bramanti B, et al. Recovery of a medieval *Brucella melitensis* genome using shotgun metagenomics. *mBio.* 2014; 5:e01337–14. <https://doi.org/10.1128/mBio.01337-14> PMID: 25028426
14. Mutolo MJ, Jenny LL, Buszek AR, Fenton TW, Foran DR. Osteological and molecular identification of Brucellosis in ancient Butrint, Albania. *Am J Phys Anthropol.* 2012; 147:254–263. <https://doi.org/10.1002/ajpa.21643> PMID: 22212927
15. Meffray A, Biagini P, Rigeade C, Panuel M, Schneider L, Ardagna Y. À propos de deux cas de brucellose dans le sud de la France aux époques médiévale et moderne (Abbaye Saint-Sauveur, Aniane; La Closeraie, Aix-en-Provence). *Bulletins et mémoires de la Société d'Anthropologie de Paris BMSAP.* 2022; 34(34 (1)).
16. Hider J, Duggan AT, Klunk J, Eaton K, Long GS, Karpinski E, et al. Examining pathogen DNA recovery across the remains of a 14th century Italian friar (Blessed Sante) infected with *Brucella melitensis*. *International Journal of Paleopathology.* 2022; 39:20–34. <https://doi.org/10.1016/j.ijpp.2022.08.002> PMID: 36174312
17. Bouwman AS, Brown TA. The limits of biomolecular palaeopathology: ancient DNA cannot be used to study venereal syphilis. *Journal of Archaeological Science.* 2005; 32(5):703–713. <https://doi.org/10.1016/j.jas.2004.11.014>
18. von Hunnius TE, Yang D, Eng B, Waye JS, Saunders SR. Digging deeper into the limits of ancient DNA research on syphilis. *Journal of Archaeological Science.* 2007; 34(12):2091–2100. <https://doi.org/10.1016/j.jas.2007.02.007>
19. Garofolo G, DiGiannatale E, DeMassis F, Zilli K, Ancora M, Camma C, et al. Investigating genetic diversity of *Brucella abortus* and *Brucella melitensis* in Italy with MLVA-16. *Infect Genet Evol.* 2013; 19:59–70. <https://doi.org/10.1016/j.meegid.2013.06.021> PMID: 23831636
20. Tan KK, Tan YC, Chang LY, Lee KW, Nore SS, Yee WY, et al. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. *BMC Genomics.* 2015; 16:93. <https://doi.org/10.1186/s12864-015-1294-x> PMID: 25888205
21. Godfroid J, DeBolle X, Roop RM, O'Callaghan D, Tsois RM, Baldwin C, et al. The quest for a true One Health perspective of brucellosis. *Rev Sci Tech.* 2014; 33:521–538. <https://doi.org/10.20506/rst.33.2.2290> PMID: 25707181
22. Godfroid J. Brucellosis in livestock and wildlife: zoonotic diseases without pandemic potential in need of innovative one health approaches. *Arch Public Health.* 2017; 75:34. <https://doi.org/10.1186/s13690-017-0207-7> PMID: 28904791
23. MacKinnon M. Production and consumption of animals in Roman Italy: integrating the zooarchaeological and textual evidence. vol. 54. *Journal of Roman archaeology*; 2004.
24. De Rancourt M, Fois N, Lavín M, Tchakérián E, Vallerand F. Mediterranean sheep and goats production: An uncertain future. *Small Ruminant Research.* 2006; 62(3):167–179. <https://doi.org/10.1016/j.smallrumres.2005.08.012>
25. Zeder MA. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci U S A.* 2008; 105:11597–11604. <https://doi.org/10.1073/pnas.0801317105> PMID: 18697943
26. Janowicz A, DeMassis F, Zilli K, Ancora M, Tittarelli M, Sacchini F, et al. Evolutionary history and current distribution of the West Mediterranean lineage of *Brucella melitensis* in Italy. *Microb Genom.* 2020; 6. <https://doi.org/10.1099/mgen.0.000446> PMID: 33030422

27. Pisarenko SV, Kovalev DA, Volynkina AS, Ponomarenko DG, Rusanova DV, Zharinova NV, et al. Global evolution and phylogeography of *Brucella melitensis* strains. *BMC Genomics*. 2018; 19:353. <https://doi.org/10.1186/s12864-018-4762-2> PMID: 29747573
28. Brothwell D. The Paleopathology of the EB-MB and Middle Bronze Age Remains from Jericho (1957–58) Excavations. *Kenyon KM Excavations at Jericho II London*. 1965; p. 658–693.
29. Rashidi J, Ortner D, Frohlich B, Jonsdottir B. Brucellosis in Early Bronze Age Jordan and Bahrain: an analysis of possible cases of *Brucella* Spondylitis. In: *American Journal of Physical Anthropology*. WILEY-LISS DIV JOHN WILEY & SONS INC, 605 THIRD AVE, NEW YORK, NY 10158-0012 USA; 2001. p. 122–123.
30. Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, et al. Revealing the history of sheep domestication using retrovirus integrations. *Science*. 2009; 324:532–536. <https://doi.org/10.1126/science.1170587> PMID: 19390051
31. Larson G, Burger J. A population genetics view of animal domestication. *Trends Genet*. 2013; 29:197–205. <https://doi.org/10.1016/j.tig.2013.01.003> PMID: 23415592
32. Ciani E, Mastrangelo S, DaSilva A, Marroni F, Ferencakovic M, Ajmone-Marsan P, et al. On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genet Sel Evol*. 2020; 52:25. <https://doi.org/10.1186/s12711-020-00545-7> PMID: 32408891
33. Deng J, Xie XL, Wang DF, Zhao C, Lv FH, Li X, et al. Paternal Origins and Migratory Episodes of Domestic Sheep. *Curr Biol*. 2020; 30:4085–4095.e6. <https://doi.org/10.1016/j.cub.2020.07.077> PMID: 32822607
34. Alberto FJ, Boyer F, Orozco-terWengel P, Streeter I, Servin B, deVillemereuil P, et al. Convergent genomic signatures of domestication in sheep and goats. *Nat Commun*. 2018; 9:813. <https://doi.org/10.1038/s41467-018-03206-y> PMID: 29511174
35. Machova K, Malkova A, Vostry L. Sheep Post-Domestication Expansion in the Context of Mitochondrial and Y Chromosome Haplogroups and Haplotypes. *Genes (Basel)*. 2022; 13. <https://doi.org/10.3390/genes13040613> PMID: 35456419
36. Georgi E, Walter MC, Pfalzgraf MT, Northoff BH, Holdt LM, Scholz HC, et al. Whole genome sequencing of *Brucella melitensis* isolated from 57 patients in Germany reveals high diversity in strains from Middle East. *PLoS One*. 2017; 12:e0175425. <https://doi.org/10.1371/journal.pone.0175425> PMID: 28388689
37. Laine CG, Scott HM, Arenas-Gamboa AM. Human brucellosis: Widespread information deficiency hinders an understanding of global disease frequency. *PLoS Negl Trop Dis*. 2022; 16:e0010404. <https://doi.org/10.1371/journal.pntd.0010404> PMID: 35580076
38. Whatmore AM, Foster JT. Emerging diversity and ongoing expansion of the genus *Brucella*. *Infect Genet Evol*. 2021; 92:104865. <https://doi.org/10.1016/j.meegid.2021.104865> PMID: 33872784
39. Whatmore AM, Perrett LL, MacMillan AP. Characterisation of the genetic diversity of *Brucella* by multilocus sequencing. *BMC Microbiol*. 2007; 7:34. <https://doi.org/10.1186/1471-2180-7-34> PMID: 17448232
40. Whatmore AM, Koylass MS, Muchowski J, Edwards-Smallbone J, Gopaul KK, Perrett LL. Extended Multilocus Sequence Analysis to Describe the Global Population Structure of the Genus *Brucella*: Phylogeography and Relationship to Biovars. *Front Microbiol*. 2016; 7:2049. <https://doi.org/10.3389/fmicb.2016.02049> PMID: 28066370
41. Roberts Charlotte A and Buikstra Jane E. Chapter 11—Bacterial Infections. In: Buikstra Jane E, editor. *Ortner's Identification of Pathological Conditions in Human Skeletal Remains (Third Edition)*. Third Edition ed. San Diego: Academic Press; 2019. p. 321–439. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128097380000119>.
42. Esmailnejad-Ganji SM, Esmailnejad-Ganji SMR. Osteoarticular manifestations of human brucellosis: a review. *World journal of orthopedics*. 2019; 10(2):54. <https://doi.org/10.5312/wjo.v10.i2.54> PMID: 30788222
43. Ventura L, Arrizza L, Urbani V, Miranda G. Indagini macroscopiche, radiologiche e di laboratorio dei resti del Beato Sante Brancorsini (1343-1394); 2015.
44. Devault AM, Mortimer TD, Kitchen A, Kieseewetter H, Enk JM, Golding GB, et al. A molecular portrait of maternal sepsis from Byzantine Troy. *eLife*. 2017; 6:e20983. <https://doi.org/10.7554/eLife.20983> PMID: 28072390
45. Long GS, Klunk J, Duggan AT, Tapson M, Giuffra V, Gazze L, et al. A 16(th) century *Escherichia coli* draft genome associated with an opportunistic bile infection. *Commun Biol*. 2022; 5:599. <https://doi.org/10.1038/s42003-022-03527-1> PMID: 35710940

46. Mann AE, Sabin S, Ziesemer K, Vagene AJ, Schroeder H, Ozga AT, et al. Differential preservation of endogenous human and microbial DNA in dental calculus and dentin. *Sci Rep.* 2018; 8:9822. <https://doi.org/10.1038/s41598-018-28091-9> PMID: 29959351
47. Cardoso PG, Macedo GC, Azevedo V, Oliveira SC. *Brucella* spp noncanonical LPS: structure, biosynthesis, and interaction with host immune system. *Microb Cell Fact.* 2006; 5:13. <https://doi.org/10.1186/1475-2859-5-13> PMID: 16556309
48. Suarez-Esquivel M, Chaves-Olarte E, Moreno E, Guzman-Verri C. *Brucella* Genomics: Macro and Micro Evolution. *Int J Mol Sci.* 2020; 21. <https://doi.org/10.3390/ijms21207749> PMID: 33092044
49. Moreno E. Genome evolution within the alpha Proteobacteria: why do some bacteria not possess plasmids and others exhibit more than one different chromosome? *FEMS Microbiol Rev.* 1998; 22:255–275. <https://doi.org/10.1111/j.1574-6976.1998.tb00370.x> PMID: 9862123
50. Moreno E, Blasco JM, Letesson JJ, Gorvel JP, Moriyon I. Pathogenicity and Its Implications in Taxonomy: The *Brucella* and *Ochrobactrum* Case. *Pathogens.* 2022; 11. <https://doi.org/10.3390/pathogens11030377> PMID: 35335701
51. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom.* 2018; 4. <https://doi.org/10.1099/mgen.0.000206> PMID: 30052170
52. Rambaut A, Lam TT, MaxCarvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2016; 2:vew007. <https://doi.org/10.1093/ve/vew007> PMID: 27774300
53. Viadas C, Rodriguez MC, Sangari FJ, Gorvel JP, Garcia-Lobo JM, Lopez-Goni I. Transcriptome analysis of the *Brucella abortus* BvrR/BvrS two-component regulatory system. *PLoS One.* 2010; 5:e10216. <https://doi.org/10.1371/journal.pone.0010216> PMID: 20422049
54. Avila-Calderon ED, Medina-Chavez O, Flores-Romo L, Hernandez-Hernandez JM, Donis-Maturano L, Lopez-Merino A, et al. Outer Membrane Vesicles From *Brucella melitensis* Modulate Immune Response and Induce Cytoskeleton Rearrangement in Peripheral Blood Mononuclear Cells. *Front Microbiol.* 2020; 11:556795. <https://doi.org/10.3389/fmicb.2020.556795> PMID: 33193138
55. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/ measurement sets. *Nucleic Acids Res.* 2021; 49:D605–D612. <https://doi.org/10.1093/nar/gkaa1074> PMID: 33237311
56. Głowacka P, Zakowska D, Naylor K, Niemcewicz M, Bielawska-Drozd A. *Brucella*—Virulence Factors, Pathogenesis and Treatment. *Pol J Microbiol.* 2018; 67:151–161. <https://doi.org/10.21307/pjm-2018-029> PMID: 30015453
57. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research.* 2020; 48(D1):D517–D525. <https://doi.org/10.1093/nar/gkz935> PMID: 31665441
58. Benveniste R, Davies J. Mechanisms of antibiotic resistance in bacteria. *Annual Review of Biochemistry.* 1973; 42(1):471–506. <https://doi.org/10.1146/annurev.bi.42.070173.002351> PMID: 4581231
59. Piddock LJ. Multidrug-resistance efflux pumps? not just for resistance. *Nature Reviews Microbiology.* 2006; 4(8):629–636. <https://doi.org/10.1038/nrmicro1464> PMID: 16845433
60. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015; 31(22):3691–3693. <https://doi.org/10.1093/bioinformatics/btv421> PMID: 26198102
61. Johansen TB, Scheffer L, Jensen VK, Bohlin J, Feruglio SL. Whole-genome sequencing and antimicrobial resistance in *Brucella melitensis* from a Norwegian perspective. *Sci Rep.* 2018; 8:8538. <https://doi.org/10.1038/s41598-018-26906-3> PMID: 29867163
62. Kistler L, Ware R, Smith O, Collins M, Allaby RG. A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* 2017; 45:6310–6320. <https://doi.org/10.1093/nar/gkx361> PMID: 28486705
63. Lopez-Santiago R, Sanchez-Argaez AB, DeAlba-Nunez LG, Baltierra-Urbe SL, Moreno-Lafont MC. Immune Response to Mucosal *Brucella* Infection. *Front Immunol.* 2019; 10:1759. <https://doi.org/10.3389/fimmu.2019.01759> PMID: 31481953
64. Wattam AR, Foster JT, Mane SP, Beckstrom-Sternberg SM, Beckstrom-Sternberg JM, Dickerman AW, et al. Comparative phylogenomics and evolution of the Brucellae reveal a path to virulence. *J Bacteriol.* 2014; 196:920–930. <https://doi.org/10.1128/JB.01091-13> PMID: 24336939
65. Barquero-Calvo E, Chaves-Olarte E, Weiss DS, Guzman-Verri C, Chacon-Diaz C, Rucavado A, et al. *Brucella abortus* uses a stealthy strategy to avoid activation of the innate immune system during the

- onset of infection. PLoS One. 2007; 2:e631. <https://doi.org/10.1371/journal.pone.0000631> PMID: 17637846
66. Arapovic J, Kompes G, Dedic K, Teskeredzic S, Ostojic M, Travar M, et al. Antimicrobial resistance profiles of human *Brucella melitensis* isolates in three different microdilution broths: the first multicentre study in Bosnia and Herzegovina. J Glob Antimicrob Resist. 2022; 29:99–104. <https://doi.org/10.1016/j.jgar.2022.02.005> PMID: 35182775
  67. Wareth G, Dadar M, Ali H, Hamdy MER, Al-Talhy AM, Elkharsawi AR, et al. The perspective of antibiotic therapeutic challenges of brucellosis in the Middle East and North African countries: Current situation and therapeutic management. Transbound Emerg Dis. 2022; 69:e1253–e1268. <https://doi.org/10.1111/tbed.14502> PMID: 35244335
  68. Colli L, Milanese M, Talenti A, Bertolini F, Chen M, Crisa A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. Genet Sel Evol. 2018; 50:58. <https://doi.org/10.1186/s12711-018-0422-x> PMID: 30449284
  69. Biagi P, et al. A review of the Late Mesolithic in Italy and its implications for the Neolithic transition. The widening harvest. 2003; p. 133–156.
  70. Skeates R. Radiocarbon dating and interpretations of the Mesolithic-Neolithic transition in Italy. In: Ammerman AJ, Biagi P, editors. The widening harvest: the neolithic transition in Europe: looking back, looking forward. Boston: Archaeological Institute of America; 2003. p. 157–187. Available from: <http://dro.dur.ac.uk/3916/>.
  71. Barker G. 3. In: Agriculture, Pastoralism, and Mediterranean Landscapes in Prehistory. John Wiley & Sons, Ltd; 2005. p. 46–76. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470773536.ch3>.
  72. Rowley-Conwy P. Early domestic animals in Europe: imported or locally domesticated? In: Ammerman A, Biagi P, editors. The widening harvest: the neolithic transition in Europe: looking back, looking forward. vol. AIA Colloquia and Conference Papers; 6. Boston: Archaeological Institute of America; 2003. p. 99–117. Available from: <http://dro.dur.ac.uk/3840/>.
  73. Sabatini S, Bergerbrant S, Brandt LØ, Margaryan A, Allentoft ME. Approaching sheep herds origins and the emergence of the wool economy in continental Europe during the Bronze Age. Archaeological and Anthropological Sciences. 2019; 11(9):4909–4925. <https://doi.org/10.1007/s12520-019-00856-x>
  74. Mandolini Giancarlo. Il Beato Sante tra storia e scienza. Fonti storiche della vita e risultati scientifici della ricognizione del corpo del Beato Sante. Urbani: Arti Grafiche SITIBU, Italy; 2015.
  75. Bracci Silvano. Beato Sante Brancorsini e il Santuario di Mombaroccio. Torino: Elledici, Italy; 2015.
  76. Schwarz C, Debruyne R, Kuch M, McNally E, Schwarz H, Aubrey AD, et al. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res. 2009; 37:3215–3229. <https://doi.org/10.1093/nar/gkp159> PMID: 19321502
  77. Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences. 2013; p. 201314445. <https://doi.org/10.1073/pnas.1314445110> PMID: 24019490
  78. Glocke I, Meyer M. Extending the spectrum of DNA sequences retrieved from ancient bones and teeth. Genome Research. 2017. <https://doi.org/10.1101/gr.219675.116> PMID: 28408382
  79. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
  80. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019; 20:257. <https://doi.org/10.1186/s13059-019-1891-0> PMID: 31779668
  81. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. Bioinformatics. 2011; 27:863–864. <https://doi.org/10.1093/bioinformatics/btr026> PMID: 21278185
  82. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. 2017; 27:849–864. <https://doi.org/10.1101/gr.213611.116> PMID: 28396521
  83. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009; 25(14):1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
  84. Duggan AT, Perdomo MF, Piombino-Mascalci D, Marciniak S, Poinar D, Emery MV, et al. 17th century variola virus reveals the recent history of smallpox. Current Biology. 2016; 26(24):3407–3412. <https://doi.org/10.1016/j.cub.2016.10.061> PMID: 27939314
  85. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinformatics. 2020; 70:e102. <https://doi.org/10.1002/cpbi.102> PMID: 32559359

86. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10(1):421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
87. Eaton K. NCBImeta: efficient and comprehensive metadata retrieval from NCBI databases. *Journal of Open Source Software*. 2020; 5(46):1990. <https://doi.org/10.21105/joss.01990>
88. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–2069. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063
89. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 2013; 29(13):1682–1684. <https://doi.org/10.1093/bioinformatics/btt193> PMID: 23613487
90. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res*. 2012; 40:e72. <https://doi.org/10.1093/nar/gks001> PMID: 22323520
91. Brown CT, Olm MR, Thomas BC, Banfield JF. Measurement of bacterial replication rates in microbial communities. *Nat Biotechnol*. 2016; 34:1256–1263. <https://doi.org/10.1038/nbt.3704> PMID: 27819664
92. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research*. 2015; 43(3):e15–e15. <https://doi.org/10.1093/nar/gku1196> PMID: 25414349
93. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*. 2020; 37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015> PMID: 32011700
94. Kalyaanamoorthy S, Minh BQ, Wong TKF, vonHaeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017; 14:587–589. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
95. Menardo F, Loiseau C, Brites D, Coscolla M, Gygli SM, Rutaiwa LK, et al. Treemmer: a tool to reduce large phylogenetic datasets with minimal loss of diversity. *BMC Bioinformatics*. 2018; 19(1):1–8. <https://doi.org/10.1186/s12859-018-2164-8> PMID: 29716518
96. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med*. 2014; 6:90. <https://doi.org/10.1186/s13073-014-0090-6> PMID: 25422674
97. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res*. 2018; 3:124. <https://doi.org/10.12688/wellcomeopenres.14826.1> PMID: 30345391
98. Kimura M. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc Natl Acad Sci U S A*. 1981; 78:454–458. <https://doi.org/10.1073/pnas.78.1.454> PMID: 6165991
99. To TH, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. *Systematic Biology*. 2016; 65(1):82–97. <https://doi.org/10.1093/sysbio/syv068> PMID: 26424727
100. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus evolution*. 2018; 4(1):vey016. <https://doi.org/10.1093/ve/vey016> PMID: 29942656
101. Duchene S, Lemey P, Stadler T, Ho SYW, Duchene DA, Dhanasekaran V, et al. Bayesian Evaluation of Temporal Signal in Measurably Evolving Populations. *Mol Biol Evol*. 2020; 37:3363–3379. <https://doi.org/10.1093/molbev/msaa163> PMID: 32895707
102. Galperin MY, Wolf YI, Makarova KS, VeraAlvarez R, Landsman D, Koonin EV. COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*. 2021; 49:D274–D281. <https://doi.org/10.1093/nar/gkaa1018> PMID: 33167031
103. Kaufman L. Clustering large data sets. *Pattern Recognition in Practice*. 1986; p. 425–437. <https://doi.org/10.1016/B978-0-444-87877-9.50039-X>
104. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987; 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
105. Zeder MA. Out of the Fertile Crescent: The dispersal of domestic livestock through Europe and Africa. In: Boivin N, Crassard R, Petraglia M, editors. *Human Dispersal and Species Movement: From Prehistory to the Present*. Cambridge University Press; 2017. p. 261–303.