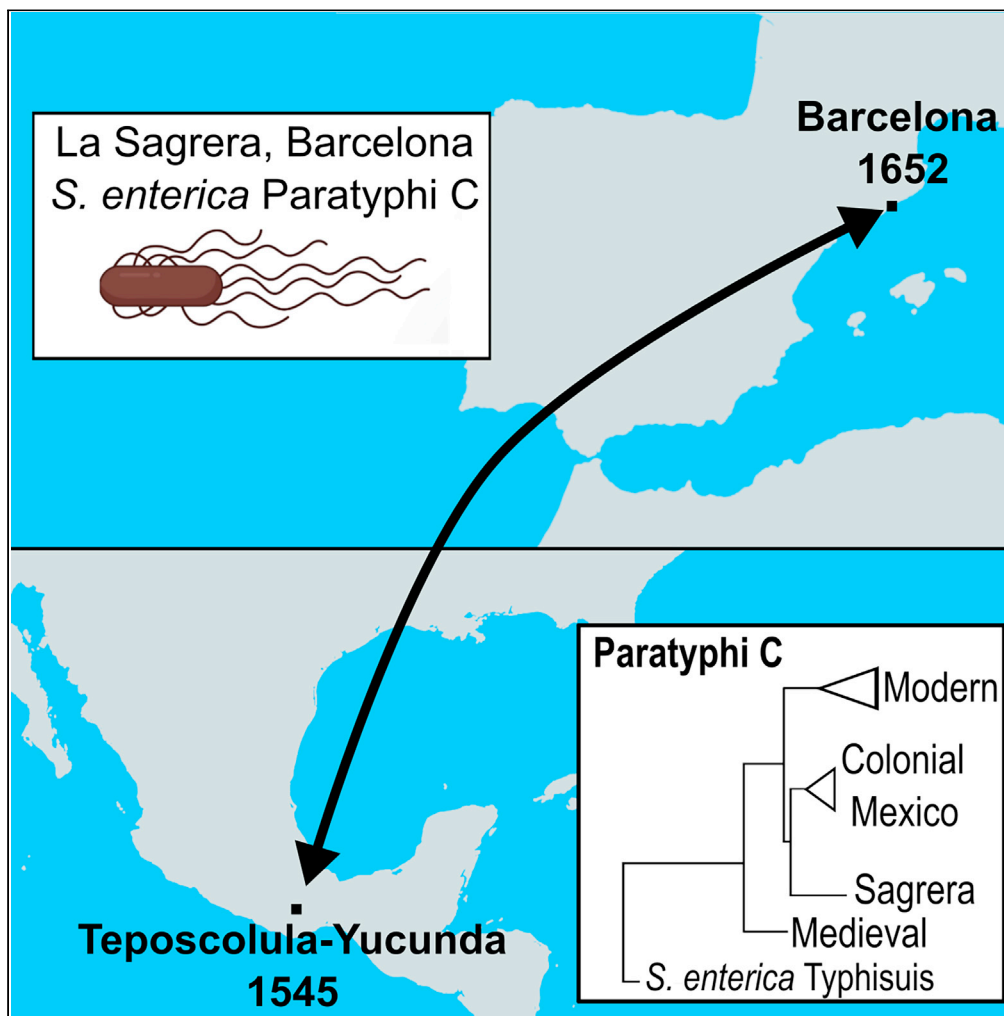


## Article

*Salmonella enterica* from a soldier from the 1652 siege of Barcelona (Spain) supports historical transatlantic epidemic contacts

Toni de-Dios,  
Pablo Carrión,  
Iñigo Olalde, ...,  
François Balloux,  
Lucy van Dorp,  
Carles Lalueza-Fox

lucy.dorp.12@ucl.ac.uk (L.v.D.)  
carles.lalueza.fox@gmail.com  
(C.L.-F.)

**Highlights**

Genome-wide data from  
two soldiers of the 1652  
siege of Barcelona (Spain)

Retrieval of an ancient  
*Salmonella enterica*  
serovar Paratyphi C strain

Seventeenth century  
European *S. enterica*  
Paratyphi C clusters with  
colonial Mexican strains

Further support for  
transatlantic dispersals of  
paratyphoid fever

de-Dios et al., iScience 24,  
103021  
September 24, 2021 © 2021  
The Author(s).  
[https://doi.org/10.1016/  
j.isci.2021.103021](https://doi.org/10.1016/j.isci.2021.103021)

## Article

# *Salmonella enterica* from a soldier from the 1652 siege of Barcelona (Spain) supports historical transatlantic epidemic contacts

Toni de-Dios,<sup>1</sup> Pablo Carrión,<sup>1</sup> Iñigo Olalde,<sup>1</sup> Laia Llovera Nadal,<sup>1</sup> Esther Lizano,<sup>1</sup> Dídac Pàmies,<sup>2</sup> Tomas Marques-Bonet,<sup>1,3,4,5</sup> François Balloux,<sup>6</sup> Lucy van Dorp,<sup>6,\*</sup> and Carles Lalueza-Fox<sup>1,7,\*</sup>

**SUMMARY**

**Ancient pathogen genomics is an emerging field allowing reconstruction of past epidemics. The demise of post-contact American populations may, at least in part, have been caused by paratyphoid fever brought by Europeans. We retrieved genome-wide data from two Spanish soldiers who were besieging the city of Barcelona in 1652, during the Reapers' War. Their ancestry derived from the Basque region and Sardinia, respectively, (at that time, this island belonged to the Spanish kingdom). Despite the proposed plague epidemic, we could not find solid evidence for the presence of the causative plague agent in these individuals. However, we retrieved from one individual a substantial fraction of the *Salmonella enterica* serovar Paratyphi C lineage linked to paratyphoid fever in colonial period Mexico. Our results support a growing body of evidence that Paratyphi C enteric fever was more prevalent in Europe and the Americas in the past than it is today.**

**INTRODUCTION**

Next generation sequencing techniques allow for the recovery of historic pathogens from a variety of sources including medical collections (De-Dios, 2019; Van Dorp et al., 2020), ancient parchments (de-Dios et al., 2020; Piñar et al., 2020), ancient "chewing gum" (Jensen et al., 2019), and human archaeological remains (Maixner et al., 2016; Mühlemann et al., 2018; Vågene et al., 2018; Zhou et al., 2018; Key et al., 2020). One of the most widely studied historic pathogen is *Yersinia pestis*, the bacterial agent of plague, for which numerous ancient genomes—almost exclusively sequenced from teeth—have been analyzed spanning different time periods and locations (Bos et al., 2011, 2016; Rasmussen et al., 2015; Andrades Valtueña et al., 2017; Rascovan et al., 2019).

More recently, ancient genomics has recovered DNA attributable to the pathogen *Salmonella enterica*, a species comprising over 2500 serovars. These include *Salmonella enterica* serotypes Typhi (or typhoid fever, which accounts for 75% of cases) and Paratyphi A, B, and C (or paratyphoid fevers) (Cash-Goldwasser and Barry, 2018). The former serovar is globally prevalent among paratyphoid fevers (Ochiai et al., 2005; Stanaway et al., 2019), while Paratyphi B and C are relatively scarce (Achtman et al., 2012; Cash-Goldwasser and Barry, 2018). Typhoid and paratyphoid fevers affect up to 14 million people and cause the death of 135,900 people annually (Stanaway et al., 2019). Although a non-life-threatening disease with appropriate antibiotic treatment, without it, mortality rates can reach 10–20% (Crump et al., 2008; World Health Organisation, 2018). Typhoid and paratyphoid fevers are particularly prevalent in developing countries of sub-Saharan Africa, South East Asia, and South Asia, where they represent one of the leading causes of death and disability (Naghavi et al., 2017; Vos et al., 2017).

Genomic analyses of ancient strains of *S. enterica* support its prevalence as a pathogen of humans in the past, with *S. enterica* directly observed in association with humans across Eurasia over at least the last 6500 years (Key et al., 2020). Together with other *Salmonella enterica* serotypes, Paratyphi C belongs to a subgroup of *S. enterica* serovars defined as the Ancient Eurasian Super Branch (AESB), a cluster that comprises phylogenetically close serotypes that infect different wild animals and livestock, as well as strictly human-adapted lineages. The process of host adaptation of the AESB likely accompanied the Neolithic transition, when changes in lifestyle and closer interactions with domesticated animals may have led to recurrent

<sup>1</sup>Institute of Evolutionary Biology (CSIC-UPF), 08003 Barcelona, Spain

<sup>2</sup>Antequem. Arqueologia- Patrimoni Cultural, 08301 Mataró, Spain

<sup>3</sup>Catalan Institution of Research and Advanced Studies (ICREA), 08010 Barcelona, Spain

<sup>4</sup>CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain

<sup>5</sup>Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193 Cerdanyola del Vallès, Spain

<sup>6</sup>UCL Genetics Institute, University College London, London WC1E 6BT, UK

<sup>7</sup>Lead contact

\*Correspondence: lucy.dorp.12@ucl.ac.uk (L.v.D.), carles.lalueza.fox@gmail.com (C.L.-F.)

<https://doi.org/10.1016/j.isci.2021.103021>



exposure and infections (Key et al., 2020). The characterization of the Paratyphi C genome “Ragna” identified from the sequenced remains of both the bones and the teeth of an 800-year-old skeleton from Trondheim, Norway, demonstrates the presence of Paratyphi C and invasive salmonellosis in past Medieval Europe (Zhou et al., 2018).

More recent evidence for the past epidemic potential of this bacterium is the discovery of Paratyphi C in human burials from colonial Mexico associated with the *Cocoliztli* epidemics (Vågene et al., 2018). The *Cocoliztli* (disease or plague in Nahuatl) was a series of epidemics that devastated the native populations of New Spain after the arrival of Spanish colonizers during the early 16<sup>th</sup> century (Malvido and Viesca, 1985). These epidemics were believed to have been caused by the introduction of infectious diseases by Europeans with suggested agents including measles, smallpox, malaria, and unknown haemorrhagic fever (Soliminos d’Ardois, 1970; Malvido and Viesca, 1985; Marr and Kiracofe, 2000). Detection of Paratyphi C in association with victims of the *Cocoliztli* lends support to the hypothesis that *S. enterica* was a contributor to epidemics during recent historical times. *S. enterica* has also been suggested as the causative agent of the famous Plague of Athens (430–426 BCE) following the amplification of two DNA fragments from individuals from the ancient *Kerameikos* mass grave, dating to that period (Papagrigorakis et al., 2006). However, subsequent phylogenetic assessment of these sequences could not authenticate them as *S. enterica* (Shapiro et al., 2006). This discrepancy can be partly explained by the difficulties of retrieving significant portions of ancient microbial genomes before the advent of second-generation sequencing technologies. Therefore, although it has been suggested that Paratyphi C was a “globally” distributed pathogen of the past, more evidence based on genome-wide data from different locations and time periods is needed to assess its putative role as an epidemic pathogen.

In this study, we generate and analyze DNA from the teeth of two individuals following archaeological excavations at La Sagrera (Sant Martí de Provençals), the Spanish army encampment that besieged Barcelona (August 1651 to October 1652) during the Reapers’ War (Monguiló et al., 2012). As well as characterizing the ancestry of these two soldiers, we retrieve from one individual a large number of sequence reads mapping to *Salmonella enterica* Paratyphi C. We place the resulting Paratyphi C genome in the context of previously published modern and ancient strains including from medieval Norway and colonial Mexico. The accurate dating of this wartime site, coupled with the observations provided by this genome, extends our understanding of the spatiotemporal breadth of Paratyphi C across Europe and the Americas.

## RESULTS

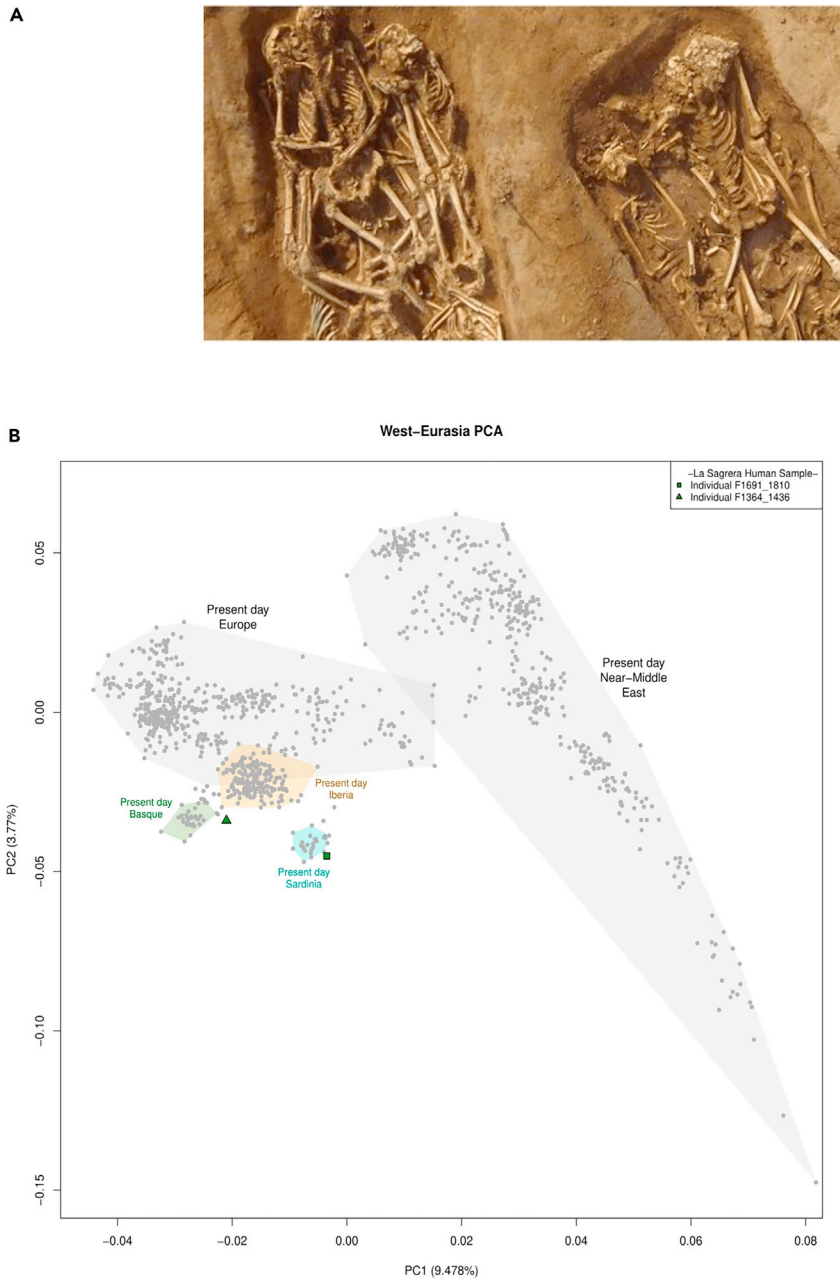
### Ancient DNA sequencing and characterization of La Sagrera soldiers

Archaeological excavations at La Sagrera uncovered 576 skeletons of soldiers grouped in shallow graves. Most of these contained between one and ten individuals, but they also included two large pits with 69 and 79 individuals, respectively. Based on the archaeological context of the burials, the soldiers were hastily buried—sometimes still dressed and with boots on—without signs of war injuries and within a short period of time. Archaeologists tend to associate such burial procedures with disease outbreaks (Signoli et al., 2004; Raoult et al., 2006). In contemporaneous chronicles, it is mentioned that both the defenders—within the city—and the besiegers suffered a bout of “pestilence” that has been previously considered to be a plague outbreak (Parets, 1660; Luis and Moya, 1990).

We sampled the teeth of two soldiers of the Spanish army, labeled F1691-1810 and F1364-1436, for ancient DNA sequencing (Figure 1A). From these samples, we obtained 246,473,297 and 99,018,404 DNA reads, respectively. The ends of the DNA reads obtained displayed typical ancient DNA (aDNA) damage patterns in a ratio consistent with the estimated age of the site (Figure S1). In both cases, the molecular sex was assigned to male (Table S1). Based on the presence of different mitochondrial DNA (mtDNA) haplogroups and on heterozygosity levels of the X chromosome, we found that one of the libraries was partially contaminated (around 9% of human reads). Therefore, we only considered the non-contaminated libraries for the subsequent human population genetics analysis. The mtDNA haplogroups were assigned as U5b1f1a and H2a5a for F1691-1810 and F1364-1436, respectively (Table S2). Y chromosome haplogroups were assigned with differing degrees of resolution to R1 and R1b1a1b1 for F1691-1810 and F1364-1436, respectively.

### Population genetic affinities of the analyzed individuals

Little is known about the geographic origins of soldiers enlisted in the Spanish army at the time, and uncertainties exist over whether these predominately comprised locally or foreign raised troops. Projection



**Figure 1. Population genetic affinities of the soldiers analyzed at the archaeological site at La Sagrera**

(A) Image of some of the soldiers' remains found at the site.

(B) Genetic data for both individuals (see legend at top right) are projected onto the two main principal components (PC) defined by 1431 present-day West-Eurasian individuals genotyped by human origins array (gray-points). Modern populations more closely related to the Siege of Barcelona soldiers are highlighted (present-day Basques (green); Spanish (yellow); Sardinians (cyan)).

of the diversity detected both individuals into a principal component analysis (PCA) conducted on modern West Eurasians placed F1364-1436 neighboring the diversity observed in present day Basques and F1691-1810 falling close to the genetic diversity observed in present day Sardinians (Figure 1B) (Álvarez-Iglesias et al., 2009; Cardoso et al., 2013). What is now the Basque Country and the island of Sardinia were under dominion of the Spanish crown during the 17<sup>th</sup> century, implying that all the regions of the Spanish realm contributed with manpower to the war effort. In fact, it is recorded that the fall of Barcelona in 1652 was

celebrated across the Spanish possessions, including Cagliari in Sardinia (Caredda, 2015). The analysis of further individuals could provide additional information on the composition of professional Spanish armies during the 17<sup>th</sup> century; for example, contemporaneous chronicles explain there was a contingent of Irish mercenaries (Morales Ó, 2004).

### Ancient pathogens screening of La Sagrera soldiers

Although contemporaneous assessments suggested that soldiers died of plague during the siege (Luis and Moya, 1990), metagenomic screening of the non-human DNA content from both individuals was unable to identify the presence of *Yersinia pestis* DNA reads; only 0.00003%–0.000003% of all DNA reads were assigned to pathogenic *Yersinia* species (Table S3). The absence of *Y. pestis* in these individuals does not exclude outbreaks of plague during the siege, noting that identification of *Y. pestis* tends to require large-scale screening of individuals. Interestingly, 0.01% of sequencing reads were assigned to *Salmonella* in one of the individuals F1691-1810, with only 0.00001% in the other. After mapping all reads in F1691-1810 against a comprehensive set of *Salmonella enterica* reference genomes, Paratyphi C was determined as the best representative based on the mean coverage, the percentage of the genome covered by at least a single read, and the mean read edit distance (Table S4) (Liu et al., 2009). Mapping solely against the Paratyphi C reference chromosome (see STAR Methods), we obtained 24,225 uniquely mapped DNA reads, accounting for 30% of the reference genome at an average depth of coverage of 0.38X (Table S5). In addition, we obtained 59.02% and an average depth of coverage of 0.91X over the ~5 kb Paratyphi C virulence plasmid (pSPCV), supporting its vertical inheritance and co-evolution with the Paratyphi C lineage (Zhou et al., 2018; Key et al., 2020). The expected coverage for both the chromosome and the plasmid matched the expected theoretical value (Figures 2A, 2B and 2D) (see STAR Methods). Postmortem damage at the end of the *Salmonella* DNA reads (up to 15%) and read length distribution, as well as their absence in two extraction blanks (Table S5), supported these sequences as authentically old rather than modern contaminants (Figures 2C, 2E, and 2F).

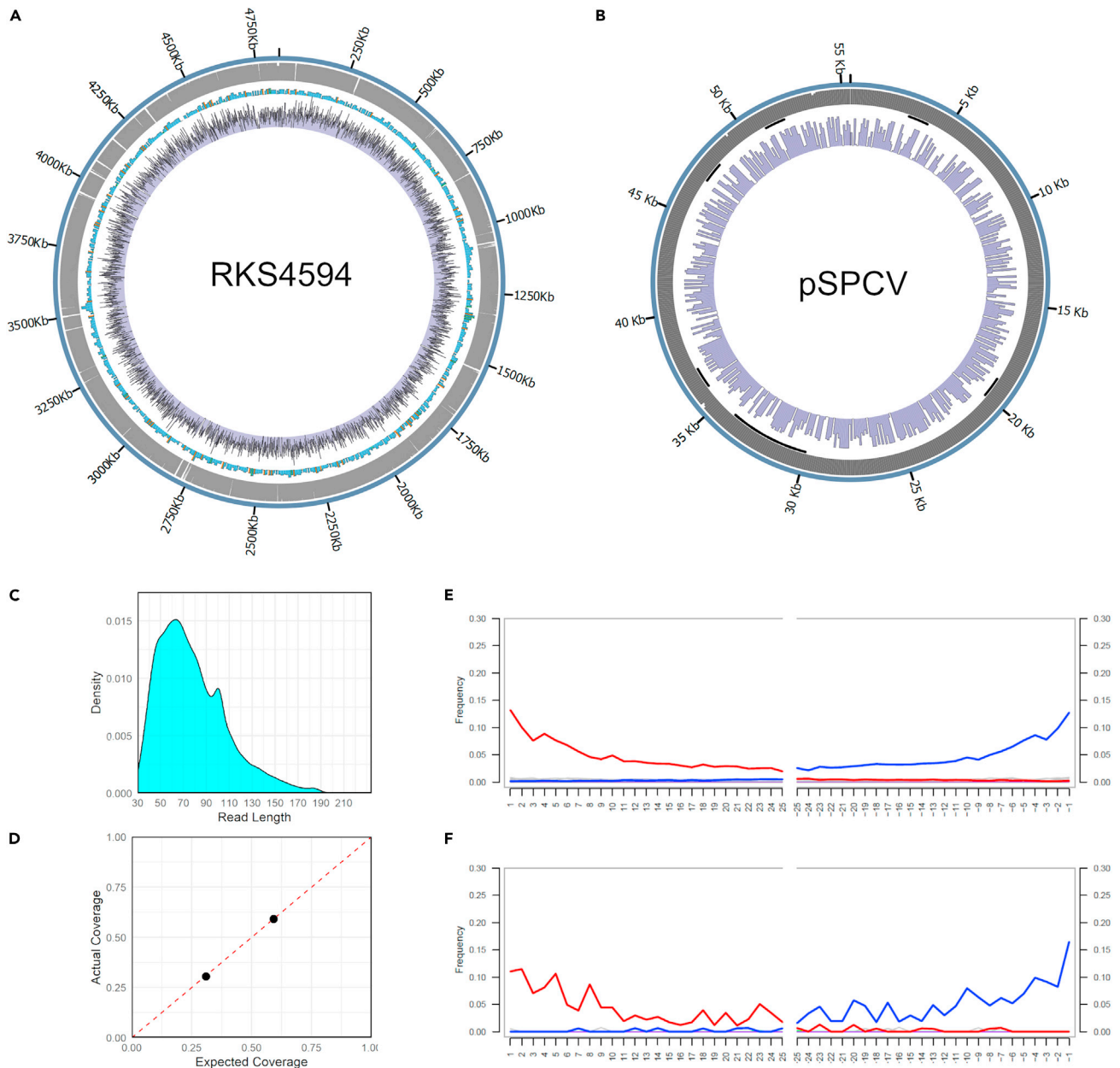
### Phylogenetic analysis of the *Salmonella enterica* Paratyphi C-retrieved sequences

After filtering out sequences by depth, heterozygosity, and the presence of low coverage transitions (see STAR Methods), we generated 1,146,808 confidently covered positions, of which 10,250 are single-nucleotide polymorphisms (SNPs). Finally, we explored the depth of coverage in virulence-related regions associated with the pSPCV (NC\_012124.1) plasmid (Liu et al., 2009) and other virulence-associated genes. No apparent variation in coverage over these genes was detected, suggesting that virulence gene composition of pSPCV has been largely maintained from the time of our sample to the present day (Figure S2; Table S6).

To place our strain into the wider context of *Salmonella enterica*, we built a maximum likelihood (ML) tree over the chromosomal alignment that included 415 isolates (both ancient and modern), encompassing the AESB clade and including *S. enterica* serovar Enteritidis as an outgroup. This placed our 17th-century Barcelona strain within the diversity of sampled Paratyphi C. Repeating the ML tree using only Paratyphi C strains and Typhisuis as an outgroup, F1691-1810 was positioned in a clade (with 100% bootstrap support) falling basal to all modern Paratyphi C diversity and including the colonial Mexican Paratyphi C strains associated with the Cocoliztli epidemics (Figures S3A and S4) (Vågene et al., 2018). The only strain basal to this clade is the medieval sample from Trondheim (Norway) dated to 1,200 CE (Zhou et al., 2018). Under the expectation of constant accumulation of mutations with time, the terminal branch lengths of the ancient samples are longer than seen in modern strains, suggesting the possibility of recombination events or downstream incorporation of SNPs which may be erroneously called due to low coverage or postmortem damage (Figure S3B).

Due to the high genetic affinity and close temporal range of the Paratyphi C genome obtained from the F1691-1810 La Sagrera individual and the ancient lineage associated with the Mexican Cocoliztli, we explored if the split time of these clades was more consistent with an introduction of *S. enterica* Paratyphi C into the Americas by Spanish colonizers or if the strain from Barcelona was imported into Spain from the American continent. To create a robust alignment suitable for the application of phylogenetic tip dating, we first set out to prune from the alignment all sites in conflict with the expectation under clonal evolution. In particular, we applied *ClonalFrameML* (Didelot et al., 2015) to identify all putative homologous recombination events between sets of donors and recipients in the Paratyphi C phylogeny. We identified putative recombinant tracts comprising a total of 104 different events totaling 30 kb in length (Figures S5 and S6; Table S7).





**Figure 2. Characteristics of the *S. enterica* Paratyphi C recovered sequences**

(A and B) Coverage plot of the Paratyphi C chromosome (A) and plasmid (B); the outer blue circle provides the genomic position, the second outermost circle in gray provides the reference mappability, the third most outermost ring provides the presence of genes (blue), pseudogenes (orange), and RNAs (green) along the reference, and the innermost purple circle provides the mean depth (binned) of coverage.

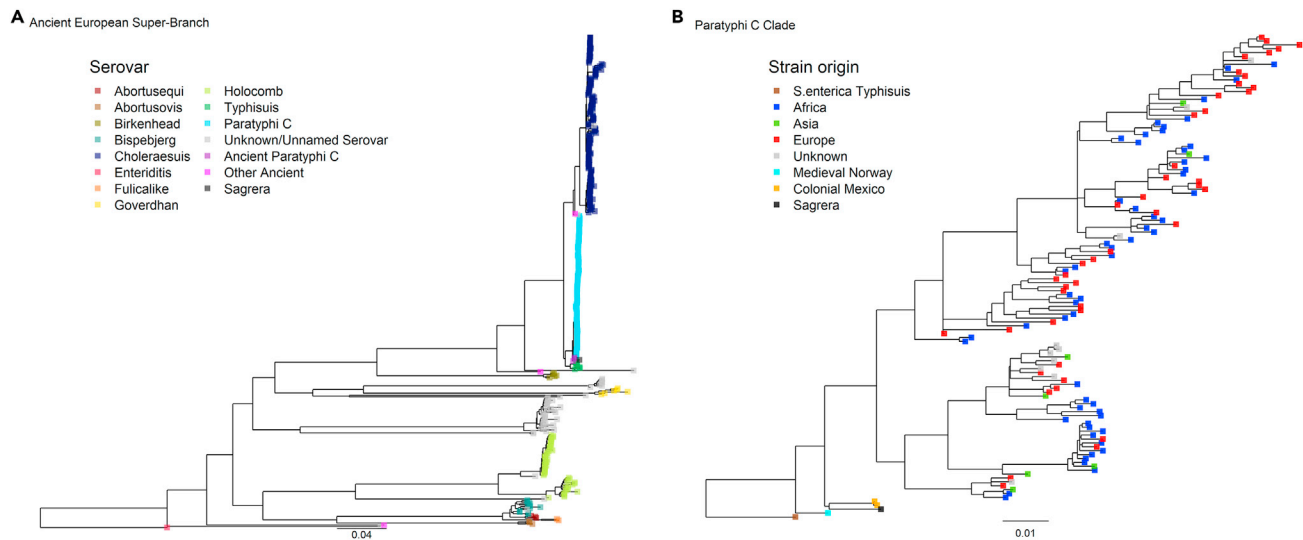
(C) Read length distribution of sequences mapped against Paratyphi C reference genome.

(D) Comparison between the calculated and expected coverage and the actual observed coverage for both the pSPCV plasmid and chromosome.

(E) aDNA damage patterns (site specific substitutions) at both ends of the DNA reads mapped against the Paratyphi C reference genome. (F) damage patterns at both ends of the DNA reads mapped to the pSPCV plasmid. For both (E) and (F), red provides the C to T substitution frequency; blue provides the G to A substitution frequency; and gray provides the frequency relative to all other substitutions.

### Temporal relationship of La Sagrera with other taxa in the Paratyphi C clade

In addition, we applied a further strict filtering technique, identifying all homoplastic positions in the alignment given the consensus relationships supported by the ML phylogeny. Such sites may derive from recombination events otherwise undetected by *ClonalFrameML* or correspond to spurious variant calls



**Figure 3. Maximum likelihood phylogeny of *S. enterica* falling within the Ancient Eurasian Super Branch (AESB).**

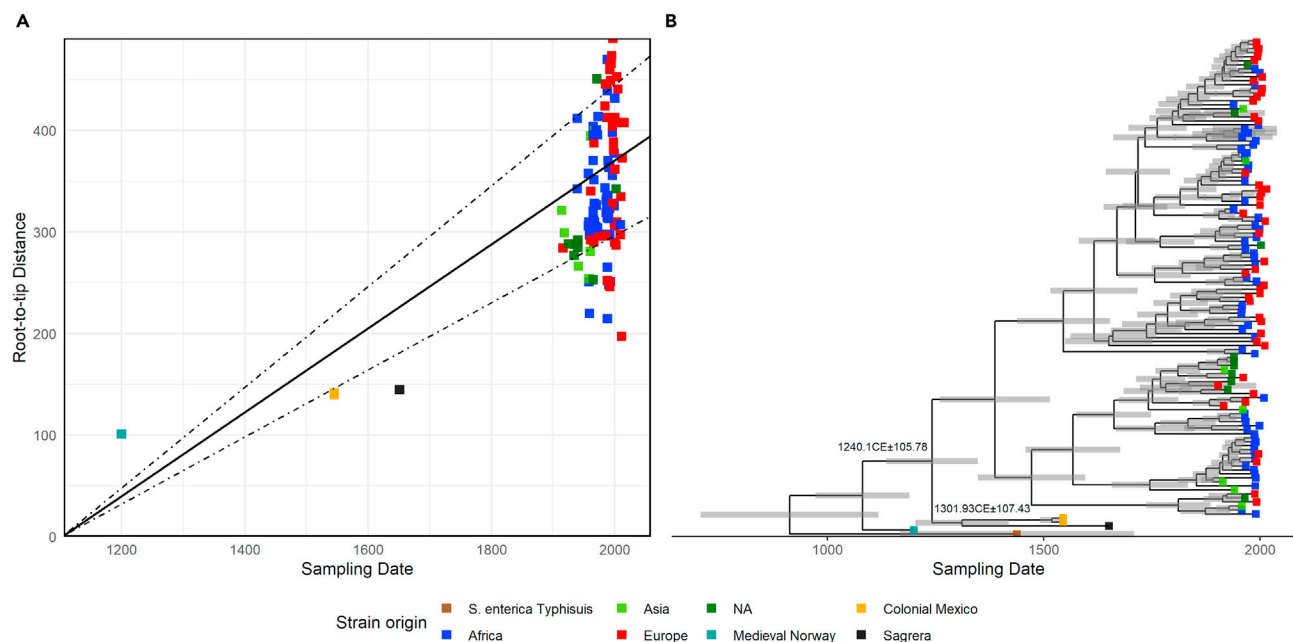
(A) ML tree of La Sagrera and 413 *S. enterica* strains representing the ancient and modern diversity of the AESB from previous studies (Vågene et al., 2018; Zhou et al., 2018; Key et al., 2020). The tree is rooted using *S. enterica* ser. Enteritidis from a previous study (Alikhan et al., 2018). Serovars are colored at the branch tip.

(B) Maximum likelihood phylogeny of the Paratyphi C clade including 124 Paratyphi C modern and ancient strains. The tree includes La Sagrera strain, 119 Modern Paratyphi C strains (Key et al., 2020), 2 strains from Colonial Mexico (Vågene et al., 2018), a strain from Medieval Norway (Zhou et al., 2018), and an *S. enterica* ser. Typhisuis as an outgroup (Key et al., 2020). Tip points are colored according to their origin as given by the legends. In both cases, phylogenies were constructed from clonal alignments with recombination events and homoplasies removed.

included as a result of low coverage or DNA damage. Applying *HomoplasyFinder* (Crispell et al., 2019), we identified homoplasies by calculating the consistency index, which, for each site, provides a measure of the observed number of changes divided by the minimum number of changes needed to achieve a certain state at tip of a tree. We detected 532 positions in the alignment with a consistency index  $<0.5$  and subsequently pruned these from the recombination-stripped Paratyphi C alignment (Crispell et al., 2019). The resulting filtered alignment comprised 3680 variant sites fully covered across all of the 124 Paratyphi C strains. We generated a further ML phylogeny using only those positions that passed the filtering criteria. The new tree was topologically congruent; however, the terminal branch length of the ancient tips in the tree was reduced, suggesting that the alignment filtering procedure removed recombinant and/or spurious sites (Figure 3).

Previous studies have utilized ancient Paratyphi C genomes to calibrate the evolutionary rate for members of the AESB (Zhou et al., 2018; Key et al., 2020). Similarly, we estimate the age of the most recent common ancestor (MRCA) (the split times) of our sample at relevant nodes by calibrating the recombination-filtered phylogenetic tree by time. To do so, we first confirmed the existence of a significant temporal signal over the alignment by computing the correlation between the root-to-tip genetic distance and the estimated time of sample collection. For our ancient Spanish sample, we set a prior on the tip date to 1652 to coincide with the timing of the siege of Barcelona and the archaeological dating of the site. We recovered a positive correlation ( $R^2 = 0.3$ ) between the root-to-tip distance and sampling time, which was significant following  $1 \times 10^4$  randomizations of the tip sampling date ( $p < 1.00 \times 10^{-4}$ ) (Figure 4A). We then applied *BactDating* to formally estimate the mutation rate over the alignment running until full convergence of the MCMC algorithm (Didelot et al., 2018) (Figure S7). We estimated a mutation rate for this clade over the clonal frame to  $1.02 \times 10^{-4}$  mutations per site per year. This mutation rate corresponds well with the rates reported for other *S. enterica* serovars (Okoro et al., 2012; Hawkey et al., 2013; Octavia et al., 2015) and is similar to recent estimates following equivalent approximations (Didelot et al., 2021).

Based on the application of an additive clock model, the estimated rate of evolution leads to an inferred date of the split between the genome identified in medieval Trondheim (Ragna) and the rest of the diversity of Paratyphi C to 1058.43 CE (CI: 949.85 CE–1167.0 CE), a timing consistent with an equivalent analysis by



**Figure 4. Temporal signal in the Paratyphi C clade**

(A) Regression of the root-to-tip distance and collection date of the samples present in the tree, with  $1 \times 10^7$  permutations.  $R^2 = 0.44$ ,  $p < 1.00 \times 10^{-4}$ . Dotted lines provide the 95% CI.

(B) Time-calibrated phylogenetic tree of the curated Paratyphi C dataset. X axis provides the time of collection date. Gray bars at each node denote the 95% CI of the estimated divergence times. Key divergence estimates are highlighted and discussed in the text.

Zhou et al. (2018) (1162–1526 ybp, CI: 833–1975 ybp), thereby supporting the robustness of temporal estimates when including our low coverage sample. We estimate the divergence time between the 1652 La Sagrera Spanish genome and samples from colonial Mexico from sampled modern diversity to 1240.11 CE (CI: 1134.32 CE–1345.89 CE). Finally, we estimated the split between our 1652 sample and those linked solely to New World epidemics in Mexico to 1301.93 CE (CI: 1194.5 CE–1409.35 CE) (Figure 4B), suggesting a common ancestor of Mexican and the Spanish Paratyphi C around the turn of the 14<sup>th</sup> century.

## DISCUSSION

Despite an expectation that the two soldiers whose remains we analyzed had died of the plague (Parets, 1660; Luis and Moya, 1990), no reliable traces of *Yersinia pestis* could be found. This does not preclude plague outbreaks during the Siege of Barcelona; a far larger number of skeletons will have to be screened before *Y. pestis* can be formally ruled out as the agent of the disease that affected Barcelona's besiegers. However, we did recover the partial genome of the pathogen *Salmonella enterica* Paratyphi C in one soldier, which may have contributed to his demise, possibly as part of a wider epidemic affecting the Spanish army. Indeed, outbreaks of enteric fever are plausible considering the sanitary conditions of a cramped military camp during a siege, as well as the surrounding marshes (that were drained in modern times (Martin-Vide, 2015)). Both of these factors and the presence of contaminated water may favor the rapid transmission of disease (Gunn et al., 2014). Nonetheless, it is not possible to rule out the co-existence of *Y. pestis* given that so far we have only analyzed two individuals and the fact that Plague was a well-known disease at the time with clear symptoms (Parets, 1660; Luis and Moya, 1990).

Recent studies have demonstrated the possible epidemic character of *S. enterica* Paratyphi C in historical times. For example, *Salmonella* Paratyphi C DNA was found in ancient remains from the New World (Vågene et al., 2018) associated with mass burials in Mexico dating back to the mid-16<sup>th</sup> century, plausibly attributed to the Cocoliztli epidemics. The discovery and characterization of the pathogen in a 12<sup>th</sup> century woman from Trondheim, Norway, further supports the notion that *S. Paratyphi C* may have noticeably contributed to past pandemics of enteric fever in recent human history (Zhou et al., 2018), despite being rare in Europe and the Americas today.



Our work, recovering a ~1652 *S. enterica* Paratyphi C genome from a Spanish soldier, extends the spatiotemporal distribution of *S. Paratyphi C* to Southern Europe to as recently as the mid-17<sup>th</sup> century and provides a further observation of the presence of this pathogen in epidemic burial context. The phylogenetic placement of our Paratyphi C genome from 1652 Barcelona basal to the diversity observed in post-contact Mexican strains supports one of two scenarios: (i) an introduction of *S. Paratyphi C* strains from Europe or (ii) an extensive geographic range of *S. Paratyphi C* prior to the colonial period. Given that other cases of introduction of pathogens in the Americas during the colonial period have been recently proposed, aided by the analysis of ancient microbial genomics, including parvovirus and hepatitis B virus, leprosy, syphilis, and malaria (Yalcindag et al., 2012; Schuenemann et al., 2013; Guzmán-Solis et al., 2020; Majander et al., 2020; Van Dorp et al., 2020), it is plausible that colonial migrations contributed to the distribution of past Paratyphi C observed. The continental origin of some infectious diseases putatively associated with colonial settlement, such as syphilis, has been debated for decades with little progress from the analysis of modern strains alone. Syphilis was once thought to have been introduced in Europe from the Americas; however, strains likely predating Columbian times have been recently discovered in Northern European skeletal remains (Schuenemann et al., 2018; Majander et al., 2020).

The presence of *S. Paratyphi C* in mid-17<sup>th</sup> century Spain is particularly interesting in the context of these historical epidemics due to the fact that these strains are no longer present in the current world diversity of *S. Paratyphi C*. Our estimated split between the Spanish and Mexican strains to the early 14<sup>th</sup> century predates Columbian times. Such an observation may be consistent with a possible European reservoir in Spain of the *Salmonella* lineages which are postulated to have caused widespread epidemics in the native populations of the Americas during the discovery epoch. Adding support to our estimates, the divergence dates we recover over the wider phylogeny, for instance, the split time of *S. enterica* Paratyphi C from *S. enterica* Typhisuis, are highly consistent with independent estimates obtained in studies providing the first characterizations of the timing of evolutionary events in this serovar using ancient DNA (Zhou et al., 2018; Key et al., 2020). This suggests the inclusion of even partial ancient genomes, as we obtained here, can be useful in temporal phylogenomic analyses. Despite uncertainties associated to phylogenetic tip-dating methods and to the low coverage data from ancient skeletal remains, our results reinforce the presence of *S. enterica* Paratyphi C in Europe for centuries, opening the possibility that this pathogen contributed to other historical pandemics of debated cause—including the one suffered by Spanish besiegers of Barcelona in 1652.

Nevertheless, the likely complex pattern of trans-Atlantic connectivity detected here may only be fully clarified with the study of additional ancient samples, both from the Americas, Africa, and Eurasia. One challenge caused by the lack of osteological signals associated with typhoid fevers means that such a survey will need to be conducted blindly, analyzing, like we did in this study, historical mass graves potentially attributed to other common past pandemics such as the plague.

### Limitations of the study

It has not been possible to retrieve the full *S. enterica* Paratyphi C genome, and the low coverage may have resulted in the inclusion of erroneous sites despite our stringent filters (Figure S8). It is possible this could subtly impact some of our temporal estimates. In the future, we would like to generate DNA from remains of additional soldiers from this site in the hope to identify further, higher coverage, representatives of the clade.

### STAR★ METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECTS DETAILS
  - The site
- METHOD DETAILS

- DNA extraction and library preparation
- Human DNA mapping
- Sex determination and uniparental markers analysis
- Contamination estimates
- Human population genetics analysis
- Pathogens' screening and Salmonella sequences mapping
- Coverage and mappability of Salmonella enterica Paratyphi C
- Virulence gene presence
- Variant calling and phylogenetics dataset creation
- Maximum likelihood tree
- Temporal signal exploration
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.103021>.

## ACKNOWLEDGMENTS

We are grateful to the Museum of the History of Barcelona (MUHBA) for allowing access to these samples. C.L.-F. is supported by a PGC2018-0955931-B-100 grant (MCIU/AEI/FEDER, UE) of Spain; T.M.-B. is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 864203), BFU2017-86471-P (MINECO/FEDER, UE), "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M), Howard Hughes International Early Career, Secretaria d'Universitats i Recerca, and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880). L.v.D. is supported by a UCL Excellence Fellowship.

## AUTHOR CONTRIBUTIONS

Conceptualization, C.L.-F.; investigation, D.P., L.L.N., E.L., and T.d.-D.; resources, T.M.-B. and C.L.-F.; supervision, F.B., L.v.D., T.M.-B., and C.L.-F.; writing-original draft, T.d.-D., L.v.D., and C.L.-F.; writing-review & editing, all authors.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 26, 2021

Revised: May 14, 2021

Accepted: August 19, 2021

Published: September 24, 2021

## REFERENCES

- Achtman, M., Wain, J., Weill, F.-X., Nair, S., Zhou, Z., Sangal, V., Krauland, M.G., Hale, J.L., Harbottle, H., Uesbeck, A., et al. (2012). Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* *8*, e1002776. <https://doi.org/10.1371/journal.ppat.1002776>.
- Acman, M., van Dorp, L., Santini, J.M., and Balloux, F. (2020). Large-scale network analysis captures biological features of bacterial plasmids. *Nat. Commun.* *11*, 2452. <https://doi.org/10.1038/s41467-020-16282-w>.
- Alikhan, N.F., Zhou, Z., Sergeant, M.J., and Achtman, M. (2018). A genomic overview of the population structure of *Salmonella*. *PLoS Genet.* *14*, e1007261.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410. <http://www.sciencedirect.com/science/article/pii/S0022283605803602>.
- Álvarez-Iglesias, V., Mosquera-Miguel, A., Cerezo, M., Quintáns, B., Zarrabeitia, M.T., Cuscó, I., Lareu, M.V., García, Ó., Pérez-Jurado, L., A, Carracedo, et al. (2009). New population and phylogenetic features of the internal variation within mitochondrial DNA macro-haplogroup R0. *PLoS One* *4*, e5112. <https://pubmed.ncbi.nlm.nih.gov/19340307>.
- Andrades Valtueña, A., Mittnik, A., Key, F.M., Haak, W., Allmãe, R., Belinskij, A., Daubaras, M., Feldman, M., Jankauskas, R., Janković, I., et al. (2017). The stone age plague and its persistence in Eurasia. *Curr. Biol.* *27*, 3683–3691.e8.
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA [5]. *Nat. Genet.* *23*, 147.
- Batzilla, J., Höper, D., Antonenka, U., Heesemann, J., and Rakin, A. (2011). Complete genome sequence of *Yersinia enterocolitica* subsp. palearctica serogroup O:3. *J. Bacteriol.* *193*, 2065–2067.
- Biagini, S.A., Solé-Morata, N., Matisoo-Smith, E., Zalloua, P., Comas, D., and Calafell, F. (2019). People from Ibiza: an unexpected isolate in the Western Mediterranean. *Eur. J. Hum. Genet.* *27*, 941–951. <https://doi.org/10.1038/s41431-019-0361-1>.

- Bos, K.I., Schuenemann, V.J., Golding, G.B., Burbano, H.A., Waglechner, N., Coombes, B.K., McPhee, J.B., Dewitte, S.N., Meyer, M., Schmedes, S., et al. (2011). A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* 478, 506–510. <https://doi.org/10.1038/nature10549>.
- Bos, K.I., Herbig, A., Sahl, J., Waglechner, N., Fourment, M., Forrest, S.A., Klunk, J., Schuenemann, V.J., Poinar, D., Kuch, M., et al. (2016). Eighteenth century *Yersinia pestis* genomes reveal the long-term persistence of an historical plague focus. *eLife* 5, e12994. Edited by R.A. Neher. <https://doi.org/10.7554/eLife.12994>.
- Broad Institute (2018). Picard, Available at: <http://broadinstitute.github.io/picard/>.
- Cardoso, S., Valverde, L., Alfonso-Sánchez, M.A., Palencia-Madrid, L., Elcoroaristizabal, X., Algorta, J., Catarino, S., Arteta, D., Herrera, R.J., Zarrabeitia, M.T., et al. (2013). The expanded mtDNA phylogeny of the Franco-Cantabrian region upholds the pre-neolithic genetic substrate of Basques. *PLoS ONE* 8, e67835. <https://doi.org/10.1371/journal.pone.0067835>.
- Caredda, S. (2015). Un agente de la Corona hispánica en Cerdeña: Pedro Martínez Rubio (1614-1667) y la relación de las fiestas calaritanas por la rendición de Barcelona (1652). In *Las relaciones de sucesos en los cambios políticos y sociales de la Europa Moderna*. Universitat Autònoma de Barcelona, Studia Aurea Monográfica, 6, J.G. López and S. Boadas, eds., pp. 259–269.
- Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S.S.T., Sinding, M.H.S., Samaniego, J.A., Wales, N., Sicheritz-Pontén, T., and Gilbert, M.T.P. (2018). Single-tube library preparation for degraded DNA. *Methods Ecol. Evol.* 9, 410–419.
- Cash-Goldwasser, S., and Barry, M. (2018). CDC Yellow Book 2018: Health Information for International Travel, Clinical Infectious Diseases, Edited by G.W. Brunette and J.B. Nemhauser. <https://wwwnc.cdc.gov/travel/yellowbook/2020>.
- Church, D.M., Schneider, V.A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.C., Agarwala, R., McLaren, W.M., Ritchie, G.R.S., et al. (2011). Modernizing reference genome assemblies. *PLoS Biol.* 9, e1001091.
- Connor, T.R., Owen, S.V., Langridge, G., Connell, S., Nair, S., Reuter, S., Dallman, T.J., Corander, J., Tabing, K.C., Le Hello, S., et al. (2016). 'What's in a name? Species-wide whole-genome sequencing resolves invasive and noninvasive lineages of *Salmonella enterica* serotype Paratyphi B'. *mBio* 7, 1–9.
- Crispell, J., Balaz, D., and Gordon, S.V. (2019). Homoplasifyfinder: a simple tool to identify homoplasies on a phylogeny. *Microb. Genomics* 5, e000245. <https://pubmed.ncbi.nlm.nih.gov/30663960>.
- Crump, J.A., Ram, P.K., Gupta, S.K., Miller, M.A., and Mintz, E.D. (2008). Part I. Analysis of data gaps pertaining to *Salmonella enterica* serotype Typhi infections in low and medium human development index countries, 1984–2005. *Epidemiol. Infect.* 136, 436–448. <https://pubmed.ncbi.nlm.nih.gov/17686194>.
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Pääbo, S., Arsuaga, J.L., et al. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U S A* 110, 15758–15763. <https://www.pnas.org/content/110/39/15758>.
- de-Dios, T., van Dorp, L., Charlier, P., Morfopoulou, S., Lizano, E., Bon, C., Le Bitouzé, C., Alvarez-Estape, M., Marqués-Bonet, T., Balloux, F., et al. (2020). Metagenomic analysis of a blood stain from the French revolutionary Jean-Paul Marat (1743–1793). *Infect. Genet. Evol.* 80, 104209.
- De-Dios, T., van Dorp, L., Gelabert, P., Carøe, C., Sandoval-Velasco, M., Fregel, R., Escosa, R., Aranda, C., Huijben, S., Balloux, F., et al. (2019). Genetic affinities of an eradicated European *Plasmodium falciparum* strain. *Microb. Genom.* 5, e000289. <https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000289>.
- Depristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–501.
- Didelot, X., Wilson, D.J., Bryant, D., Quail, M., and Cockfield, J. (2015). ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput. Biol.* 11, e1004041. Edited by A. Prlic. <https://doi.org/10.1371/journal.pcbi.1004041>.
- Didelot, X., Croucher, N.J., Bentley, S.D., Harris, S.R., and Wilson, D.J. (2018). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Res.* 46, e134. <https://doi.org/10.1093/nar/gky783>.
- Didelot, X., Siveroni, I., and Volz, E.M. (2021). Additive uncorrelated relaxed clock models for the dating of genomic epidemiology phylogenies. *Mol. Biol. Evol.* 38, 307–317.
- Fitch, W.M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* 20, 406. <http://www.jstor.org/stable/2412116>.
- Fortea, J., de la Rasilla, M., García-Tabernero, A., Gigli, E., Rosas, A., and Lalueza-Fox, C. (2008). Excavation protocol of bone remains for Neandertal DNA analysis in El Sidrón cave (Asturias, Spain). *J. Hum. Evol.* 55, 353–357.
- Gunn, J.S., Marshall, J.M., Baker, S., Dongol, S., Charles, R.C., and Ryan, E.T. (2014). *Salmonella* chronic carriage: epidemiology, diagnosis, and gallbladder persistence. *Trends Microbiol.* 22, 648–655. <https://pubmed.ncbi.nlm.nih.gov/25065707>.
- Guzmán-Solis, A., Blanco-Melo, D., Villa-Islas, V., Bravo-López, M., Sandoval-Velasco, M., Wesp, J., Gómez-Valdés, J., de la Luz Moreno-Cabrera, M., Meraz-Moreno, A., Solís-Pichardo, G., et al. (2020). Ancient viral genomes reveal introduction of HBV and B19V to Mexico during the transatlantic slave trade. *bioRxiv*. <http://biorxiv.org/content/early/2020/06/06/2020.06.05.137083.abstract>.
- Hawkey, J., Edwards, D.J., Dimovski, K., Hiley, L., Billman-Jacobe, H., Hogg, G., and Holt, K.E. (2013). Evidence of microevolution of *Salmonella* Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. *BMC Genom.* 14, 800. <https://doi.org/10.1186/1471-2164-14-800>.
- Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Gr. Stat.* 5, 299–314. <http://www.tandfonline.com/doi/abs/10.1080/10618600.1996.10474713>.
- Jensen, T.Z.T., Niemann, J., Iversen, K.H., Fotakis, A.K., Gopalakrishnan, S., Vågøne, Å.J., Pedersen, M.W., Sinding, M.H.S., Ellegaard, M.R., Allentoft, M.E., et al. (2019). A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nat. Commun.* 10, 5520. <https://doi.org/10.1038/s41467-019-13549-9>.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684.
- Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res.* 25, 918–925.
- Key, F.M., Posth, C., Esquivel-Gomez, L.R., Hübner, R., Spyrou, M.A., Neumann, G.U., Furtwängler, A., Sabin, S., Burri, M., Wissgott, A., et al. (2020). Emergence of human-adapted *Salmonella enterica* is linked to the Neolithization process. *Nat. Ecol. Evol.* 4, 324–333.
- Korneliusson, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinform.* 15, 356.
- Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M., et al. (2014). Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513, 409–413.
- Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* 536, 419–424. <https://doi.org/10.1038/nature19310>.
- Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, 256–259. <https://doi.org/10.1093/nar/gkz239>.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Liu, W.Q., Feng, Y., Wang, Y., Zou, Q.H., Chen, F., Guo, J.T., Peng, Y.H., Jin, Y., Li, Y.G., Hu, S.N., et al. (2009). *Salmonella* paratyphi C: genetic divergence from *Salmonella choleraesuis* and pathogenic convergence with *salmonella typhi*.

PLoS One 4, e4510. <https://doi.org/10.1371/journal.pone.0004510>.

Luis, J., and Moya, B. (1990). Sociedad y peste en la Barcelona de 1651. *Manuscripts Revista Història Modern.* 8, 255–282.

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M.R., Hallows, J.L., Kusebauch, U., Vigl, E.E., Malfertheiner, P., Megraud, F., et al. (2016). The 5300-year-old *Helicobacter pylori* genome of the Iceman. *Science* 351, 162–165. <https://science.sciencemag.org/content/351/6269/162>.

Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., du Plessis, L., Pla-Díaz, M., Arora, N., Akgül, G., Salo, K., Schats, R., et al. (2020). Ancient bacterial genomes reveal a high diversity of *Treponema pallidum* strains in early modern Europe. *Curr. Biol.* 30, 3788–3803. [e10](https://doi.org/10.1016/j.cub.2020.07.011).

Malvido, E., and Viesca, C. (1985). Epidemia de cocoliztli de 1576. In *Historias*, pp. 26–33. [https://www.estudioshistoricos.inah.gob.mx/revistaHistorias/wp-content/uploads/historias\\_11\\_24-33.pdf](https://www.estudioshistoricos.inah.gob.mx/revistaHistorias/wp-content/uploads/historias_11_24-33.pdf).

Marr, J.S., and Kiracofe, J.B. (2000). Was the Huey cocoliztli a haemorrhagic fever? *Med. Hist.* 44, 363–388. <https://pubmed.ncbi.nlm.nih.gov/10954969>.

Martín-Vide, J.P. (2015). Restauración del río Besòs en Barcelona. *Historia y lecciones*. *Ribagada* 2, 51–60. <https://doi.org/10.1016/j.riba.2015.07.001>.

Monguiló, E., Hernandez, J., and Molinas, R. (2012). Intervenció arqueològica a l'espai delimitat pels carrers d'Espronceda, ronda de Sant Martí, carrer de Josep Soldevila i passeig de la Verneda (Ave Sagrera), Mataró.

Morales Ó, R. (2004). “Una nación inclinada al ruido de las armas” La presencia irlandesa en los ejércitos españoles, 1580-1818: ¿La historia de un éxito? *Tiempos Modern.* 10, 1–15.

Mühlemann, B., Jones, T.C., De Barros Damgaard, P., Allentoft, M.E., Shevina, I., Logvin, A., Usmanova, E., Panyushkina, I.P., Boldgiv, B., Bazartseren, T., et al. (2018). Ancient hepatitis B viruses from the bronze age to the medieval period. *Nature* 557, 418–423. <https://doi.org/10.1038/s41586-018-0097-z>.

Naghavi, M., Abajobir, A.A., Abbafati, C., Abbas, K.M., Abd-Allah, F., Abera, S.F., Aboyans, V., Adetokunboh, O., Ärnlöv, J., Afshin, A., et al. (2017). Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1151–1210.

Ochiai, R.L., Wang, X., Von Seidlein, L., Yang, J., Bhutta, Z.A., Bhattacharya, S.K., Agtini, M., Deen, J.L., Wain, J., Kim, D.R., et al. (2005). Salmonella paratyphi A rates, Asia. *Emerg. Infect. Dis.* 11, 1764–1766. <https://pubmed.ncbi.nlm.nih.gov/16318734>.

Octavia, S., Wang, Q., Tanaka, M.M., Sintchenko, V., and Lan, R. (2015). Genomic variability of serial human isolates of *Salmonella enterica* serovar Typhimurium associated with prolonged carriage. *J. Clin. Microbiol.* 53, 3507–3514. Edited by N.A. Ledebauer. <http://jcm.asm.org/content/53/11/3507.abstract>.

Okonechnikov, K., Conesa, A., and García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32, 292–294. <https://pubmed.ncbi.nlm.nih.gov/26428292>.

Okoro, C.K., Kingsley, R.A., Connor, T.R., Harris, S.R., Parry, C.M., Al-Mashhadani, M.N., Kariuki, S., Msefula, C.L., Gordon, M.A., de Pinna, E., et al. (2012). Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat. Genet.* 44, 1215–1221. <https://doi.org/10.1038/ng.2423>.

Papagrigrakis, M.J., Yapijakis, C., Synodinos, P.N., and Baziotopoulou-Valavani, E. (2006). DNA examination of ancient dental pulp incriminates typhoid fever as a probable cause of the Plague of Athens. *Int. J. Infect. Dis.* 10, 206–214.

Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>.

Parets, M. (1660). De molts successos que han succeït dins Barcelona i molts altres llocs de Catalunya dignes de memòria (1626–1660), Barcelona.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genet.* 2, 2074–2093.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093. <http://www.genetics.org/content/192/3/1065>.

Piñar, G., Tafer, H., Schreiner, M., Miklas, H., and Sterflinger, K. (2020). Decoding the biological information contained in two ancient Slavonic parchment codices: an added historical value. *Environ. Microbiol.* 22, 3218–3233. <https://pubmed.ncbi.nlm.nih.gov/32400083>.

Plummer, M., et al. (2006). CODA: convergence diagnosis and output analysis for MCMC. *R. News* 6, 7–11. <http://cran.r-project.org/doc/Rnews/>.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.

Raoult, D., Doutour, O., Houhamdi, L., Jankauskas, R., Fournier, P.E., Ardagna, Y., Drancourt, M., Signoli, M., La, V.D., Macia, Y., et al. (2006). Evidence for louse-transmitted diseases in soldiers of Napoleon's Grand army in Vilnius'. *J. Infect. Dis.* 193, 112–120. <https://doi.org/10.1086/498534>.

Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., and Rasmussen, S. (2019). Emergence and spread of basal lineages of *Yersinia pestis* during the Neolithic decline. *Cell* 176, 295–305. [http://www.sciencedirect.com/science/article/pii/S0092867418314648](https://www.sciencedirect.com/science/article/pii/S0092867418314648).

Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M.O., et al. (2015). Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. *Cell* 163, 571–582.

Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol.* 16, 1–18.

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

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J.F., Al-Rasheid, K.A.S., Willerslev, E., Krogh, A., and Orlando, L. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genom.* 13, 178. <http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-13-178>.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* 9, 88. <https://www.ncbi.nlm.nih.gov/pubmed/26868221>.

Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., et al. (2013). Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341, 179–183.

Schuenemann, V.J., Kumar Lankapalli, A., Barquera, R., Nelson, E.A., Iraiz Hernández, D., Acuña Alonzo, V., Bos, K.I., Márquez Morfin, L., Herbig, A., and Krause, J. (2018). Historic *Treponema pallidum* genomes from Colonial Mexico retrieved from archaeological remains. *PLoS Negl. Trop. Dis.* 12, e0006447. <https://pubmed.ncbi.nlm.nih.gov/29927932>.

Shapiro, B., Rambaut, A., and Gilbert, M.T.P. (2006). No proof that typhoid caused the Plague of Athens (a reply to Papagrigrakis et al.). *Int. J. Infect. Dis.* 10, 334–335.

Signoli, M., Ardagna, Y., Adalian, P., Devriendt, W., Laly, L., Rigeade, C., Vette, T., Kuncevicus, A., Poskiene, J., Barkus, A., et al. (2004). ‘Découverte d’un charnier de la Grande Armée en Lituanie (Vilnius, 1812)’. *C. R. Palevol* 3, 219–227. <http://www.sciencedirect.com/science/article/pii/S163106830400065X>.

Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *J. Archaeol. Sci.* 40, 4477–4482.

Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proc. Natl. Acad. Sci. U S A* 111, 2229–2234. <http://www.pnas.org/content/111/6/2229.abstract>.

Somolinos de Ardois, G. (1970). La epidemia de Cocoliztli de 1545 señalada en un codice. *Tribuna Med.* 15, 85.

Sprague, L.D., and Neubauer, H. (2014). Genome sequence of *Yersinia similis* Y228T, a member of

the *Yersinia pseudotuberculosis* complex. *Genome Announc.* 2, e00216-14.

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

Stanaway, J.D., Reiner, R.C., Blacker, B.F., Goldberg, E.M., Khalil, I.A., Troeger, C.E., Andrews, J.R., Bhutta, Z.A., Crump, J.A., Im, J., et al. (2019). The global burden of typhoid and paratyphoid fevers: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect. Dis.* 19, 369–381.

Vågene, Å.J., Herbig, A., Campana, M.G., Robles García, N.M., Warinner, C., Sabin, S., Spyrou, M.A., Andrades Valtueña, A., Huson, D., Tuross, N., et al. (2018). *Salmonella enterica* genomes from victims of a major sixteenth-century epidemic in Mexico. *Nat. Ecol. Evol.* 2, 520–528.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* 43, 11.10.1–33.

Van Dorp, L., Gelabert, P., Rieux, A., De Manuel, M., De-Dios, T., Gopalakrishnan, S., Carøe, C., Sandoval-Velasco, M., Fregel, R., Olalde, I., et al.

(2020). *Plasmodium vivax* malaria viewed through the lens of an eradicated European strain. *Mol. Biol. Evol.* 37, 773–785. <https://academic.oup.com/mbe/article/37/3/773/5614438>.

Vos, T., Abajobir, A.A., Abbafati, C., Abbas, K.M., Abate, K.H., Abd-Allah, F., Abdulle, A.M., Abebo, T.A., Abera, S.F., Aboyans, V., et al. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1211–1259.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44, W58–W63.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag). <https://ggplot2.tidyverse.org>.

Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20, 257. <https://doi.org/10.1186/s13059-019-1891-0>.

World Health Organisation (2018). Typhoid vaccine: WHO position paper - March 2018. *Wkly.*

*Epidemiol. Rec.* 13, 153–172. [www.who.int/immunization/position\\_papers/](http://www.who.int/immunization/position_papers/).

Yalcindag, E., Elguero, E., Arnathau, C., Durand, P., Akiana, J., Anderson, T.J., Aubouy, A., Balloux, F., Besnard, P., Bogreau, H., et al. (2012). Multiple independent introductions of *Plasmodium falciparum* in South America. *Proc. Natl. Acad. Sci. U S A* 109, 511–516. <https://pubmed.ncbi.nlm.nih.gov/22203975>.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8, 28–36.

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.F., Sergeant, M.J., Langridge, G., Fotakis, A.K., Nair, S., Stenøien, H.K., et al. (2018). Pan-genome analysis of ancient and modern *Salmonella enterica* demonstrates genomic stability of the invasive Para C lineage for Millennia. *Curr. Biol.* 28, 2420–2428.e10.

Zhou, Z., Alikhan, N.F., Mohamed, K., Fan, Y., and Achtman, M. (2020). 'The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* genomic diversity'. *Genome Res.* 30, 138–152.

STAR★ METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological Samples</b>		
Tooth from a 17 <sup>th</sup> century soldier	La Sagrera site (Barcelona, Spain)	F1364-1436
Tooth from a 17 <sup>th</sup> century soldier	La Sagrera site (Barcelona, Spain)	F1691-1810
<b>Deposited Data</b>		
Sequencing data	This paper	ENA Project: <a href="https://ena.ebi.ac.uk/ena/record/PRJEB43276">PRJEB43276</a>
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	<a href="http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/">http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/</a>
Human Origins Dataset	(Lazaridis et al., 2016)	<a href="https://reich.hms.harvard.edu/datasets">https://reich.hms.harvard.edu/datasets</a>
Y chromosome haplogroup database version 15.73	International Society of Genetic Genealogy	<a href="http://www.isogg.org">http://www.isogg.org</a>
<i>Yersinia enterocolitica</i> assembly ASM25317v1	(Batzilla et al., 2011)	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCF_000253175.1/">https://www.ncbi.nlm.nih.gov/assembly/GCF_000253175.1/</a>
<i>Yersinia pestis</i> assembly ASM379822v1	Food and Drug Administration	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCF_003798225.1/">https://www.ncbi.nlm.nih.gov/assembly/GCF_003798225.1/</a>
<i>Yersinia pseudotuberculosis</i> assembly 51108_B01	Wellcome Sanger Institute	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCF_900637475.1/">https://www.ncbi.nlm.nih.gov/assembly/GCF_900637475.1/</a>
<i>Yersinia similis</i> assembly ASM58251v1	(Sprague and Neubauer, 2014)	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCF_000582515.1/">https://www.ncbi.nlm.nih.gov/assembly/GCF_000582515.1/</a>
<i>Salmonella enterica</i> serovar Paratyphi C assembly RKS4594	(Liu et al., 2009)	<a href="https://www.ncbi.nlm.nih.gov/assembly/GCF_000018385.1/">https://www.ncbi.nlm.nih.gov/assembly/GCF_000018385.1/</a>
Plasmid Database	(Acman et al., 2020)	<a href="https://github.com/macman123/plasmid_network_analysis">https://github.com/macman123/plasmid_network_analysis</a>
<i>Salmonella enterica</i> AESB strains whole genome data	(Key et al., 2020)	<a href="http://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:12971">http://enterobase.warwick.ac.uk/species/senterica/search_strains?query=workspace:12971</a>
<i>Salmonella enterica</i> Enteriditis whole genome data	(Alikhan et al., 2018)	NCBI Sample: ERS217420
Medieval Norway <i>Salmonella enterica</i> Paratyphi C whole genome data	(Zhou et al., 2018)	ENA Project: PRJEB19916
Colonial Mexico <i>Salmonella enterica</i> Paratyphi C whole genome data	(Vågene et al., 2018)	ENA Project: PRJEB23438
Phylogenetic tree of “la Sagrera” <i>Salmonella</i> sequences within the Paratyphi C diversity	This Paper	<a href="https://microreact.org/project/G7XdKYogeqhqkWBQLu34t">https://microreact.org/project/G7XdKYogeqhqkWBQLu34t</a>
Recombination tracks of “la Sagrera” – Paratyphi C phylogeny	This Paper	N/A
Dation of “la Sagrera” – Paratyphi C phylogeny	This Paper	N/A
<b>Software and algorithms</b>		
AdapterRemoval v.2.2.2	(Schubert et al., 2016)	<a href="https://github.com/MikkelSchubert/adaptremoval">https://github.com/MikkelSchubert/adaptremoval</a>
BWA v.0.1.17	(Li and Durbin, 2009)	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
SAMtools v.1.12	(Li et al., 2009)	<a href="http://www.htslib.org/">http://www.htslib.org/</a>
picard v.2.25.5	(Broad Institute, 2018)	<a href="https://broadinstitute.github.io/picard/">https://broadinstitute.github.io/picard/</a>
Qualimap v.2.2	(Okonechnikov et al., 2016)	<a href="http://qualimap.conesalab.org/">http://qualimap.conesalab.org/</a>
pmdtools v.0.60	(Skoglund et al., 2014)	<a href="https://github.com/pontussk/PMDtools">https://github.com/pontussk/PMDtools</a>
mapDamage2 v.2.0.8	(Jónsson et al., 2013)	<a href="https://ginolhac.github.io/mapDamage/">https://ginolhac.github.io/mapDamage/</a>
Ry_compute v.0.4	(Skoglund et al., 2013)	<a href="https://github.com/pontussk/ry_compute/blob/master/ry_compute.py">https://github.com/pontussk/ry_compute/blob/master/ry_compute.py</a>

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Haplogrep2 v.2.1.21	(Weissensteiner et al., 2016)	<a href="https://haplogrep.i-med.ac.at/app/index.html">https://haplogrep.i-med.ac.at/app/index.html</a>
schmutzi v.1.5.4	(Renaud et al., 2015)	<a href="https://github.com/grenaud/schmutzi/tree/master">https://github.com/grenaud/schmutzi/tree/master</a>
angsd v.0.925	(Korneliussen et al., 2014)	<a href="http://www.popgen.dk/angsd">http://www.popgen.dk/angsd</a>
EIGENSOFT v.7.2.1	(Patterson et al., 2006; Price et al., 2006)	<a href="https://www.hsph.harvard.edu/alkes-price/software/">https://www.hsph.harvard.edu/alkes-price/software/</a>
prinseq v.0.20.4	(Schmieder and Edwards, 2011)	<a href="https://sourceforge.net/projects/prinseq/files/">https://sourceforge.net/projects/prinseq/files/</a>
Kraken2 v.2.1.1	(Wood et al., 2019)	<a href="http://www.ccb.jhu.edu/software/kraken2/">http://www.ccb.jhu.edu/software/kraken2/</a>
bedtools v.2.24.0	(Quinlan and Hall, 2010)	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
GATK v.3.7	(Van der Auwera et al., 2013)	<a href="https://gatk.broadinstitute.org/hc/en-us">https://gatk.broadinstitute.org/hc/en-us</a>
BamUtil v.1.0.13	(Jun et al., 2015)	<a href="https://github.com/statgen/bamUtil">https://github.com/statgen/bamUtil</a>
RAxML v.8.2.4	(Stamatakis, 2014)	<a href="https://cme.h-its.org/exelixis/web/software/raxml/">https://cme.h-its.org/exelixis/web/software/raxml/</a>
R v.4.1.0	(Ihaka and Gentleman, 1996)	<a href="https://cran.r-project.org/">https://cran.r-project.org/</a>
Ggtree v.1.14.6	(Yu et al., 2017)	<a href="https://guangchuangyu.github.io/software/ggtree/">https://guangchuangyu.github.io/software/ggtree/</a>
iTOL v.6	(Letunic and Bork, 2019)	<a href="https://itol.embl.de/">https://itol.embl.de/</a>
ClonalFrameML v.1.2	(Didelot et al., 2015)	<a href="https://github.com/xavierdidelot/ClonalFrameML">https://github.com/xavierdidelot/ClonalFrameML</a>
HomoplasyFinder v. 0.0.0.9	(Crispell et al., 2019)	<a href="https://github.com/JosephCrispell/homoplasyFinder">https://github.com/JosephCrispell/homoplasyFinder</a>
BactDating v.1.0.11	(Didelot et al., 2018)	<a href="https://github.com/xavierdidelot/BactDating">https://github.com/xavierdidelot/BactDating</a>
ggplot2 v.3.0.0	(Wickham, 2016)	<a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a>
Ape v.5.4	(Paradis and Schliep, 2019)	<a href="https://cran.r-project.org/web/packages/ape/index.html">https://cran.r-project.org/web/packages/ape/index.html</a>
CODA v.0.19-4	(Plummer et al., 2006)	<a href="https://cran.r-project.org/web/packages/coda/index.html">https://cran.r-project.org/web/packages/coda/index.html</a>

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Carles Lalueza-Fox [carles.lalueza.fox@gmail.com](mailto:carles.lalueza.fox@gmail.com).

**Materials availability**

This study did not generate new unique reagents.

**Data and code availability**

- All bacterial and human shotgun sequencing data have been deposited at ENA (project PRJEB43276) and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#).
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**EXPERIMENTAL MODEL AND SUBJECTS DETAILS**

**The site**

A recently excavated archaeological site located in La Sagrera -currently a north-eastern quarter of the city of Barcelona- dates from 1651-1652 CE, during the conflict known as the Reapers' War (*Guerra dels Segadors*, 1640-1659), the last stage of the Thirty Years' War (Monguiló et al., 2012). During 1651, the city was under siege by Spanish forces commanded by Juan José de Austria. Despite the efforts of the local garrison and the arrival of French reinforcements, the city capitulated in the spring of 1652. The site consists of multiple mass graves containing mostly male individuals (118 out of 140 studied by physical

anthropologists so far), with ages ranging from 16 to 40 years and skeletal hallmarks of having undertaken intense physical activities. The individuals analyzed in this study come from the burial F1691 (individual 1810) and F1364 (individual 1436). An upper canine was extracted from each of the skulls and used for genetic analyses. The chronology was confirmed by a direct radiocarbon dating of one individual (F1334) that yielded a date of 360+/-30 years BP (Beta-573178) and by the analysis of 152 coins found across the site. Of these, 104 coins correspond to coinage minted in the first half of the 17th century from Philip IV of Spain and also Louis XIII and XIV of France. The remaining 48 coins are a diverse assembly of coins minted during the second half of the 16th century and the first half of the 17th century.

## METHOD DETAILS

### DNA extraction and library preparation

All DNA extraction and initial library preparation steps (prior to amplification) were performed in a dedicated ancient DNA laboratory, physically isolated from the laboratory used for post-PCR analyses. Strict protocols were followed to minimize the amount of human DNA in the ancient DNA laboratory, including the existence of positive air pressure in the clean rooms, the wearing of a full body suit, sleeves, shoe covers, clean shoes, facemask, hair net and double gloving. All lab surfaces, consumables, disposables, tools and instruments were wiped with bleach and ethanol, and UV irradiated before and after use. First, the teeth samples were UV irradiated (245 nm) for 10 minutes and the outermost surface of the teeth was scrapped off with a drill engraving cutter, followed by another UV irradiation for 10 more minutes, in order to exclude surface DNA contamination. Second, approximately 30 mg of tooth cementum were drilled into a fine powder by a Dremel drilling machine at low speed (5000 rpm). The fact that despite these precautions one initial library was contaminated suggests this took place at the excavation itself, probably during handling. Such an observation would suggest it is advisable to excavate, even recent sites such as this one, with anti-contamination protocols such as those implemented in far more ancient samples (Fortea et al., 2008).

DNA extraction from tooth powder was performed following the method proposed by Dabney et al. (2013). The teeth-powder samples, including an extraction blank, were added to 1 ml of extraction buffer (final concentrations: 0.45 M EDTA, 0.25 mg/mL Proteinase K, pH 8.0), resuspended by vortexing and incubated at 37°C overnight (24h) on rotation (750-900 rpm). Remaining tooth powder was then pelleted by centrifugation in a bench-top centrifuge for 2 min at maximum speed (16,100 × g). The supernatant was added to 10mL of binding buffer (final concentrations: 5 M guanidine hydrochloride, 40% (vol/vol) isopropanol, 0.05% Tween-20, and 90mM sodium acetate (pH 5.2)) and purified on a High Pure Extender column (Roche). DNA samples were eluted with 45µl of low EDTA TE buffer (pH 8.0).

A total of 35µl of each DNA extract were converted into Illumina sequencing libraries following the BEST protocol (Carøe et al., 2018). Each library was amplified by PCR using two uniquely barcoded primers. After index PCR, libraries were purified with a 1.5x AMPure clean (Beckman Coulter) and eluted in 25µl of low EDTA TE buffer (pH 8.0). Libraries were quantified using BioAnalyzer and sequenced by HiSeq 4000 (Illumina).

### Human DNA mapping

Reads were trimmed of sequencing adapters, filtered for reads of less than 30bp and merged using *AdapterRemoval* with default parameters (Schubert et al., 2016). Clipped reads were then mapped against the human reference genome hg37/19 and the Revised Cambridge mitochondrial DNA reference sequence using *BWA aln/samse* with the seeding option disabled (Andrews et al., 1999; Li and Durbin, 2009; Li et al., 2009; Church et al., 2011; Schubert et al., 2012). Next, duplicated reads were removed using *Picard* and only reads with a mapping quality equal or above 30 were considered for downstream analysis (Broad Institute, n.d.). Mapping statistics were calculated using *SAMtools* and *Qualimap2* (Li et al., 2009; Okonechnikov et al., 2016). Finally, to assess the ancient authenticity of the DNA reads, *post-mortem* associated DNA damage was estimated using *mapDamage2* and *PMDtools* (Jónsson et al., 2013; Skoglund et al., 2014).

### Sex determination and uniparental markers analysis

Molecular sex was assigned using *Ry\_compute*, a script designed to determine the sex of individuals sequenced at low coverage based on the ratio of reads mapping to each sex chromosome (Skoglund

et al., 2013). Mitochondrial haplogroups were determined using haplogrep2 (Weissensteiner et al., 2016). Y chromosome haplogroup determination was performed by manually annotating variants from the International Society of Genetic Genealogy (<http://www.isogg.org>).

### Contamination estimates

Modern human contamination was estimated using two approaches. For mitochondrial contamination, we applied *Schmutzi*, which calculates the degree of modern contamination by considering the profile of aDNA associated deamination in the sample (Renaud et al., 2015). For assessing nuclear DNA contamination, and considering that both individuals are compatible with being males, we estimated the exogenous DNA contamination based on the heterozygosity of the X chromosome sites using *angsd* (Korneliusson et al., 2014). To discard contamination in the reagents, two extraction blanks generated along the project were also sequenced.

### Human population genetics analysis

To analyze these 17th-century individuals in the context of present-day human genetic diversity, their genomic data was merged with 1134 West-Eurasian individuals genotyped in the Human origins (HO) array (Patterson et al., 2012; Lazaridis et al., 2014, 2016; Biagini et al., 2019). Principal Component Analysis (PCA) was computed using the modern HO individuals, and the ancient samples were projected onto the first two components (PC1 and PC2) using options 'lsqproject: YES' and 'shrinkmode: YES' of *smartpca* built-in module of EIGENSOFT (Patterson et al., 2006; Price et al., 2006).

### Pathogens' screening and Salmonella sequences mapping

To explore the presence of relevant microbial organisms in the samples we collapsed unique reads from the human-free sequences and removed from the dataset low complexity sequences using *Prinseq* (Schmieder and Edwards, 2011). Afterward, we applied *kraken2* to assign reads against a standard database (bacteria, archaea, fungi, protozoa and viral) (Wood et al., 2019). We found almost no evidence of *Yersinia pestis* DNA reads, but one of the samples indicated a significant presence of *Salmonella enterica*.

To validate this signal, and to identify the closest modern representative to our sample amongst a diverse set of *S. enterica*, we downloaded 839 *Salmonella* published assemblies from NCBI (as for 03/04/2020) and created a custom database. Human-free reads were then mapped against this database using the local alignment algorithm *mem* of BWA (Li and Durbin, 2009). We mapped the sequences resultant of the local alignment of the free DNA reads against each *Salmonella* assembly downloaded independently using BWA's global alignment algorithm *backtrack*. The settings used were then optimized for mapping ancient genomes by disabling the seeding option, setting an edit distance value of 0.01 and a gap open penalty value of 2. After that, duplicated sequenced were removed using *Picard* tools and DNA reads with mapping quality >25 were retained (Schubert et al., 2012). Likewise, the absence of *Y. pestis* sequences in the samples was checked by mapping against the *Y. pestis*, *Y. pseudotuberculosis*, *Y. similis* and *Y. enterocolitica* reference genomes, and afterward, possible candidate sequences were explored for taxonomic origin using blast to further assess their authenticity (Altschul et al., 1990). The mapping statistics and presence of post-mortem aDNA damage were determined as described for the human sequences.

We determined that the most suitable assembly to map against was the *Salmonella enterica* subsp. *enterica* ser. Paratyphi C reference genome (Liu et al., 2009) (NCBI accession GCF\_000018385.1). The number of mapped reads, the mean coverage, the fraction of the genome recovered, and the mean edit distance are provided in Table S4. The presence of *Salmonella* related virulence plasmids was searched by mapping the raw sequences against a comprehensive plasmid database (Acman et al., 2020).

### Coverage and mappability of Salmonella enterica Paratyphi C

We determined the mappability of the reads to the Paratyphi C reference genome by mapping *k*-mers of length 40 to 100 base pairs. The coverage of the sample was determined using *bedtools* specifying windows of 1kb for the chromosome and 100 bases for the plasmid (Quinlan and Hall, 2010). To ascertain if the fraction of the reference recovered with our sample matches the expected coverage given the sequences' characteristics we assumed a random distribution matching the actual coverage, assuming sample authenticity, following a previously described approximation (Rasmussen et al., 2015) which calculates the probability of a position being covered given the presence of *r* reads of length *l* using the following

formula:  $c = 1 - \prod_{i=1}^N (1 - l_i/g)^{r_i}$ , in which  $N$  are the different  $l_i$  to  $l_N$ , read lengths with counts  $r_1$  to  $r_N$ . This value must be corrected by multiplying  $c$  for the mappability of the reference genome, otherwise the true coverage may not match the expected. An additional scenario, leading to the expected and actual coverage not matching, is when the reference has a region with no coverage in the ancient sample (Rasmussen et al., 2015).

### Virulence gene presence

The sequences of a set of previously described virulence-related genes in Paratyphi C were concatenated to create a reference panel (Connor et al., 2016; Vågene et al., 2018). We mapped our sequences against this reference using the parameters described in the previous section. The percentage of recovered positions for each gene was calculated using *bedtools* and visualized using *ggplot2* (Quinlan and Hall, 2010; Wickham, 2016).

### Variant calling and phylogenetics dataset creation

Additionally, a dataset of 407 modern *Salmonella enterica* representative of the AESB (Alikhan et al., 2018; Key et al., 2020; Zhou et al., 2020), as well as five historical samples (Vågene et al., 2018; Zhou et al., 2018; Key et al., 2020) were curated. For all modern samples raw read data processing was as described for F1691-1810 except for employing the *BWA mem* algorithm with default settings instead of the algorithm *backtrack*, the latter being advised for ancient DNA mapping pipelines. Historical samples were processed as described for the ancient samples. Variants were called from processed published sequences using GATK algorithm *UnifiedGenotyper* (Depristo et al., 2011; Van der Auwera et al., 2013). For each sample, positions were filtered for minimum coverage of 5X and indels were excluded. This resulted in a high-quality dataset containing 230,597 SNPs.

The Colonial strains from Mexico, Tepos-20 and Tepos-37 were not considered due to their low coverage; and Tepos-10 was also discarded due to its high heterozygosity levels. Although it did not reach the coverage thresholds, the la Sagrera genome was used in the subsequent analysis after undergoing strict filtering. We compared different filtering parameters to find a balance between covered positions and possible erroneous calling (Figure S8). Accordingly, we decided to trim 10 bases from the ends of each of the DNA reads using *TrimBam* (Jun et al., 2015). Following this step all positions in the la Sagrera sample were called with the reference allele considered as valid independent of their coverage. For variant positions, only SNPs found in the high-quality dataset were considered independently again of their coverage; all other positions were discarded.

### Maximum likelihood tree

All positions satisfying the filtering criteria were used to create a consensus genome *fasta* using the reference genome as template. Those positions that were filtered out in the previous steps were masked in the resultant *fasta* file. This resulted in a phylogenetic dataset comprising 414 *Salmonella enterica* AESB and an outgroup (ERS217420).

An initial Maximum Likelihood (ML) tree was built using *RAxML* with the nucleotide substitution model GTRCAT and a strain of *Salmonella enterica* ser. Enteritidis as an outgroup. The resultant tree was visualized using the R package *ggtree* and *iTOL* (Stamatakis, 2014; Yu et al., 2017; Letunic and Bork, 2019). Following this, we applied *ClonalFrameML*, with default parameters, to correct the subsequent tree for the presence of homologous recombination events (Didelot et al., 2015). *ClonalFrameML* can infer the location of a recombination event in each branch of a ML tree, which can impact the resultant terminal branch lengths. Additionally, *ClonalFrameML* imputes missing data from the input sequences. We particularly analyzed the presence of recombination events in the Paratyphi C subtree. The transition / transversion ratio used ( $\kappa$ ) was 2.097.

After removing the recombination events present in the Paratyphi C lineage, we proceeded to remove all remaining homoplasies from the dataset using *HomoplasyFinder* (Crispell et al., 2019). A homoplasy is defined as a substitution event which has arisen independently in different separate lineages, and which can alter the tree topology, hence these may arise from undetected recombination or spurious variant calls which may appear in low coverage samples. Homoplasies are determined by their consistency index value which ranges between 0 and 1, with values closer to zero denoting more homoplastic variants (Fitch, 1971). A final ML tree, based on the filtered alignment of 3680 variant sites, was generated using *RAxML* with the

parameters described above. As before, the Paratyphi C phylogeny was visualized using *ggtree* (Yu et al., 2017).

### Temporal signal exploration

In order to perform phylogenetic tip-calibration it is necessary to first confirm the presence of significant temporal signal in the alignment. Specifically, we tested the Paratyphi C branch of the phylogeny. A temporal regression of sampling date (years) against root-to-tip distance was performed using the R package *BactDating*, which additionally conducts a date randomization test of significance (Didelot et al., 2018). Samples without an assigned collection date were treated as missing. We established the most suitable model by running all models and comparing their *DIC* and *ESS* value (Plummer et al., 2006; Didelot et al., 2018) (Table S8). We performed the analysis using the model *arc* inference model and  $1 \cdot 10^7$  iterations. The resulting time calibrated phylogeny was visualized using *ggtree* (Table S9).

### QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of model statistics were performed in R v4.1 using the packages CODA and *BactDating*. Lowest *DIC* values and highest *ESS* values were prioritized. A *p-value* of  $\leq 0.05$  was considered statistically significant.