Paratyphoid Fever and Relapsing Fever in 1812 Napoleon's Devastated Army

Abstract:
During Napoleon's retreat from Russia in 1812[1], countless soldiers of the French army succumbed to infectious diseases, but the responsible pathogen or pathogens remain debated[2–5]. We recovered and sequenced ancient DNA from the teeth of 13 soldiers who, based on historical records, likely died from infectious diseases, aiming to identify the pathogens responsible for their deaths[6]. Our results confirmed the presence of *Salmonella enterica* subsp *enterica* belonging to the lineage Para C, the causative agent of paratyphoid fever[7] ; and *Borrelia recurrentis*, responsible for relapsing fever transmitted by body lice[8]. We were not able to detect *Rickettsia prowazekii* (the agent of typhus) and *Bartonella quintana* (the cause of trench fever), which had previously been associated with this deadly event, based on PCR results and historical symptom descriptions[3]. The presence of these previously unsuspected pathogens in these soldiers reveals that they could have contributed to the devastation of Napoleon's Grande Armée during its disastrous retreat in 1812.

One-Sentence Summary: Paratyphoid fever and louse-borne relapsing fever contributed to the deaths of Napoleon's soldiers during the 1812 retreat from Russia.

Main Text

In June of 1812, Napoleon I, the French emperor, assembled a military force of about 500,000 to 600,000 soldiers to invade Russia[9]. After arriving in Moscow without decisively defeating the Russian army, the Napoleonic forces, finding themselves isolated in a ruined city, opted to initiate a retreat and to establish winter encampments along the border with Poland in October that year. The retreat from Russia spanned from October 19th to December 14th 1812[10] and resulted in the loss of nearly the entire Napoleonic army. According to historians, it wasn't the harassment from the Russian army that claimed the lives of about 300,000 men[2], but rather the harsh cold of the Russian winter, coupled with hunger and diseases. A doctor during the Russian campaign, J.R.L. de Kirckhoff, authored a book detailing the illnesses that afflicted soldiers in 1812. Specifically, he documented the prevalence of typhus, diarrhea, dysentery, fevers, pneumonia, and jaundice[5]. Other physicians[11], as well as officers[12], made similar observations about the illnesses affecting soldiers.

Different infectious diseases, such as typhus, have been described in French regiments even before the start of the Russian Campaign[13]. Typhus, in particular, which is commonly referred to as camp fever due to its frequent association with armies, has long been suspected of being the main infectious cause of the demise of the Grande Armée in 1812. This assumption was fueled by the discovery of body lice, the main vector of typhus, among the remains of Napoleonic soldiers who perished during the Great Retreat from Russia in December 1812 in Vilnius, Lithuania, as well

Paratyphoid Fever and Relapsing Fever in 1812 Napoleon's Devastated Army

Simplified text for a wide non-scientific audience

Abstract:
During Napoleon's 1812 retreat from Russia, many French soldiers died from infectious diseases, but which specific bacteria were responsible has remained debated. To investigate, we extracted and sequenced ancient DNA from the teeth of 13 soldiers who, according to historical records, likely died from infections. We detected *Salmonella enterica* Paratyphi C, a branch of *Salmonella* bacteria that causes paratyphoid fever, and *Borrelia recurrentis*, a bacterium spread by body lice that causes relapsing fever. In contrast, we did not find *Rickettsia prowazekii* (the cause of typhus) or *Bartonella quintana* (the cause of trench fever), even though these had been previously linked to the event through earlier DNA studies and historical accounts. The discovery of these unexpected bacteria in the soldiers suggests they may have contributed to the devastation of Napoleon's Grande Armée during the retreat.

One-Sentence Summary: Paratyphoid fever and louse-borne relapsing fever contributed to the deaths of Napoleon's soldiers during the 1812 retreat from Russia.

Main Text

Context of the study

In June 1812, Emperor Napoleon I gathered a massive army of between 500,000 and 600,000 soldiers to invade Russia. The campaign did not go as planned: after reaching Moscow without defeating the Russian forces, Napoleon's troops found themselves stranded in a ruined and empty city. In October, they began a long retreat, aiming to reach Poland and set up winter camps. This retreat, which lasted from October 19th to December 14th, 1812, destroyed the army. Nearly all of Napoleon's soldiers were lost, with historians estimating that around 300,000 men died. According to most accounts, it was not primarily Russian attacks that caused these deaths, but the freezing winter conditions, starvation, and widespread disease. A French army doctor, J.R.L. de Kirckhoff, later described the illnesses that struck soldiers during the campaign. He reported many cases of typhus, diarrhea, dysentery, fevers, pneumonia, and jaundice. Other doctors and military officers also wrote about the same kinds of diseases devastating the army.

Previous reports of infectious diseases in this site

French regiments were already affected by several diseases before the Russian campaign even began, and typhus was among them. Typhus—often called "camp fever" because it spreads easily in armies—has long been suspected as a major contributor to the disaster of 1812. This idea was strengthened by two findings from Vilnius, Lithuania: first, body lice (the insects that transmit typhus) were found on the remains of soldiers; second, an earlier DNA study reported signals of the bacteria that cause typhus (*Rickettsia prowazekii*) and trench fever (*Bartonella quintana*) using a very sensitive test called nested PCR. That initial test is powerful but

as by the alleged identification of *R. prowazekii* and *B. quintana* sequences amplified by nested PCR in some of these individuals[3]. However, this earlier study was limited by the technologies available at the time and relied solely on the amplification of two short DNA fragments (192 and 429 base pairs long), which did not offer sufficient resolution to provide unambiguous evidence for the presence of these pathogens in Napoleon's army. Several years later, another study successfully detected *Anelloviridae* viral ancient DNA (aDNA) in Napoleonic soldiers recovered in Kaliningrad in 1812[4], but these viruses are ubiquitous and asymptomatic in human populations, and therefore unrelated to the fatal fate of these soldiers. Using *state-of-the-art* aDNA methodologies, we reanalyzed samples from Napoleonic soldiers who died in Vilnius and identified pathogen-specific genetic material, providing direct evidence of infectious agents that may have contributed to the army's collapse.

Sequencing data from 13 teeth from Vilnius (Figure 1), each comprising approximately 20 million reads, were first taxonomically classified using KrakenUniq[14] against the full microbial NCBI database. We then screened for known human pathogens by comparing the identified TaxIDs against a curated list of 535 TaxIDs representing 185 bacterial species known to be pathogenic to humans, retrieved from the PATRIC database (see Document S2)[15]. From this initial analysis, we generated a list of taxa with more than 200 uniquely assigned reads to known pathogens in at least one sample, a threshold established as optimal for ancient metagenomic pathogen detection in Pochon et al. 2023[16]. This criterion yielded 14 candidate taxa (Table S1). These were further evaluated based on their *k-mer* counts, duplication rates and coverage estimated by KrakenUniq, and prioritized according to their known epidemiological relevance and their plausibility given the early 19th-century European military context and historically documented symptomatology (e.g., fever, diarrhea, jaundice). These steps identified two candidate pathogens with consistent signals: *Salmonella enterica* in sample YYY087A (268 hits) and *Borrelia recurrentis* in sample YYY093A (239 hits) (Tables S1-S3). Notably, after all screening and authentication steps (see STAR Methods), we did not identify any reliable reads attributable to *R. prowazekii* (the agent of typhus) or *B. quintana* (the cause of trench fever), two pathogens previously reported in individuals from this site using PCR-based methods[3]. However, the absence of authenticated reads

delicate. It allows to amplify tiny amounts of DNA (generate identical copies of a specific DNA sequence), but it is also prone to picking up stray DNA (contaminations or off-target DNA). In that study, the experimental conditions used were not very strict (the "temperature" at which DNA strands were allowed to stick together to generate the copies was low, allowing inespecific DNA "hybridization"), so sequences that are similar, but not exactly the target ones, could be copied. However, this may not have been an issue in this study, because the analysis of the 4 *R. prowazekii* PCR fragments and the 10 of *B. quitana* against a comprehensive database (NCBI nt) of all organisms sequenced to date, indicate that the amplified sequenced are still closest to the targeted pathogen species than any other in the database, which is a good authentication criterion. However, a potential flaw is that the reported DNA pieces were quite long for ancient material—192 and 429 "letters" (base pairs)—whereas ancient DNA usually survives as much shorter fragments (often under ~100 base pairs), and that, ideally, several genomic regions should be tested/covered to increase certainty. Taken together, these points mean that the earlier signals remain valid, but should be interpreted with some caution.

Later on, another team found DNA from a group of viruses called *Anelloviridae* in the remains of soldiers from Kaliningrad. But since these viruses are extremely common in people and do not make us sick, they cannot explain why the soldiers died. In our study, we applied advanced techniques for ancient DNA, designed to deal well with ancient, damaged, and degraded material. These methods allowed us to screen for the presence of genetic material of ALL known pathogen (instead of looking specifically for specific ones, as with the case of PCR), and get new, unambigous, and direct evidence of infections that may have contributed to the collapse of Napoleon's army.

How were pathogens identified in this study? Why previously reported pathogens were not detected?

The researchers analyzed DNA from the teeth of 13 soldiers buried in Vilnius (Figure 1). Each tooth yielded around 20 million short DNA pieces of 31 to 150 base pairs/letters, known as "reads", but they could originate from environmental sources, from the individual (human DNA) and, perhaps, from a pathogen infecting the individual (if there was any). These fragments were then compared against a database containing the genomes of all sequenced microbes to see which organisms they matched. The results were then filtered by searching if any hit matched a list of 185 bacteria known to cause disease in humans. To reduce the risk of false positives, the researchers only kept cases where at least 200 reads could be uniquely assigned to a particular species. This threshold has been shown in earlier studies to be a reliable cut-off for detecting ancient disease-causing microbes, since damaged DNA fragments can sometimes give misleading and false signals if too few are found. Using this approach, they initially identified 14 possible matches. These candidates were then examined more closely. The team evaluated their *k-mer* counts, a way of measuring how many short DNA "words" of a fixed length *k* (31 base-pairs/letters in this case) consistently appear in the data, which helps to further test whether the match is real or random. They also looked at duplication rates (to check that the signal wasn't simply repeated copies of the identical genomic region) and coverage (how evenly the DNA sequences are distributed across the reference genome). Finally, they considered which microbes made sense historically, given the symptoms reported at the time such as fever, diarrhea, and jaundice. Through this careful process, two bacteria produced consistent signals: *Salmonella enterica* in sample YYY087A (268 matching reads) and

from these two species in our dataset does not preclude their presence at this site or during this historical event, especially given the degraded nature of ancient DNA and potential variability in pathogen load among individuals.

Borrelia recurrentis in sample YYY093A (239 matching reads). By contrast, no reliable and specific reads were found for *Rickettsia prowazekii* (typhus) or *Bartonella quintana* (trench fever), the two bacteria previously reported at this site, but in different individuals, using PCR-based methods (as explained above). However, the absence of confirmed signals here does not mean these infections were absent altogether. Ancient DNA is often very degraded, and the amount of pathogen DNA can vary greatly between individuals. Also, PCR is a much more sensitive technique to "catch" specific DNA even when it represent an insignificant proportion of the total DNA. However, since *shotgun* sequencing used in this new study randomly reads any DNA fragment in the sample, it will always catch the DNA sequences that are in higher propotion (which is often dominated by environmental bacteria). So it is still possible that typhus or trench fever affected some of the tested soldiers in this study, but that the DNA was not conserved or was present in a proportion under the detection limit.

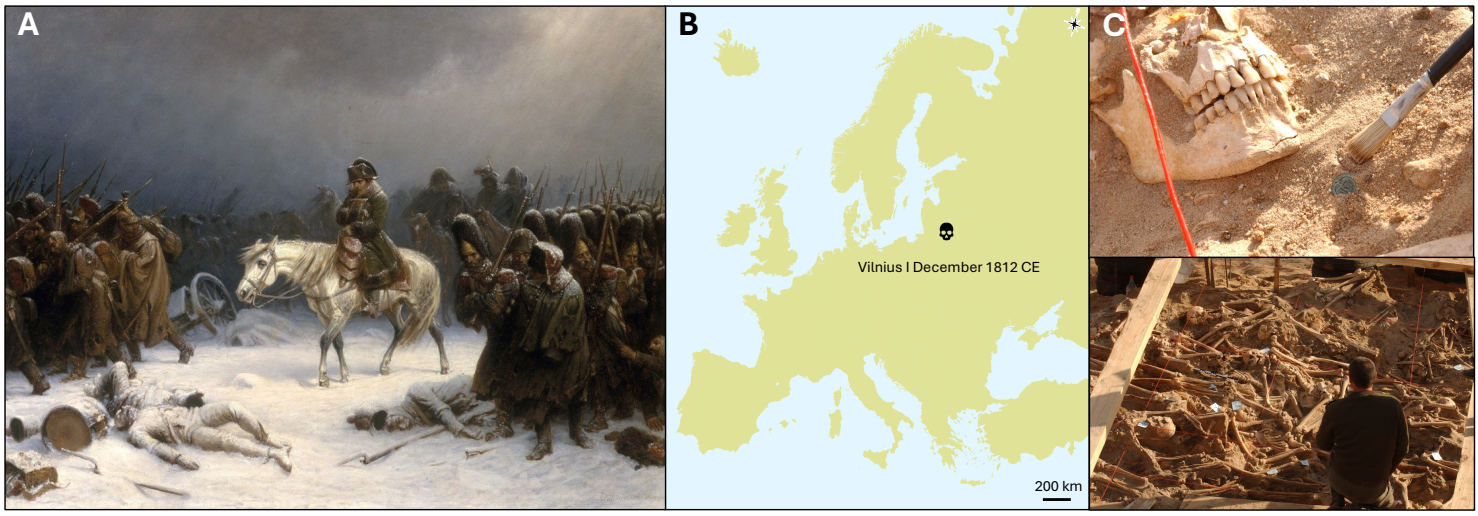

Figure 1. Historical, geographical and archeological context of the study. (A) Painting dating from 1851 entitled "Napoleon's retreat from Moscow" by Adolph Northen, depicting the conditions of the retreat of Napoleon's army. (B) Geographical map of Europe showing the location and dating of the archaeological site in Vilnius, Lithuania, from which the samples in our study were collected. (C) In situ photographs taken during the excavation of the trenches containing the bodies of Napoleonic soldiers. The top photograph shows the discovery of an imperial-type button of uniform in the mass grave. The lower photograph shows a general view of the mass grave (credit photo, Michel Signoli).

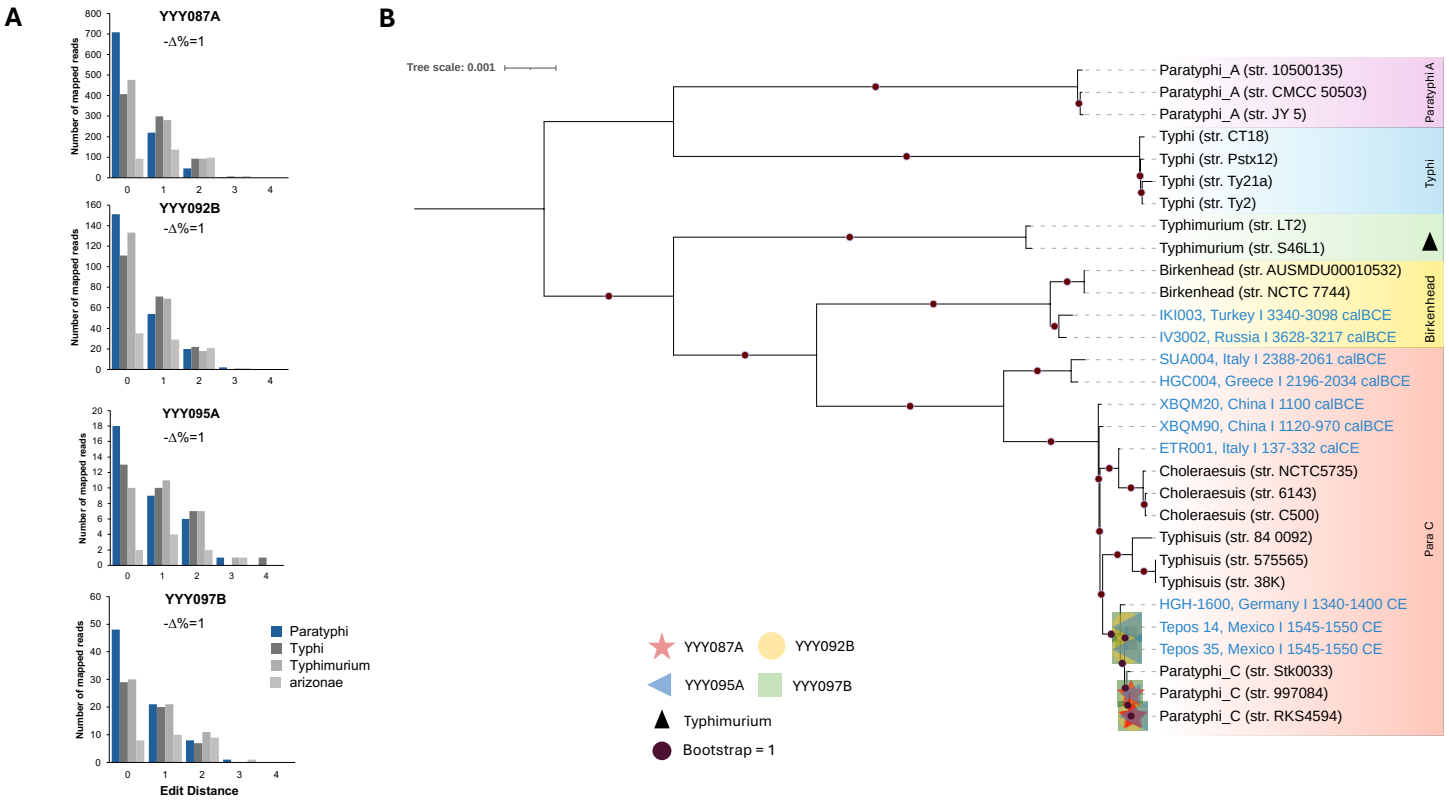

Figure 2. Authentication of *S. enterica* aDNA data. (A) Distribution of edit distances of the YYY087A, YYY092B, YYY095A and YYY097B reads, with the associated *negative difference proportion* (-Δ%)[17] (B) Maximum likelihood phylogeny of 31 *S. enterica* genomes representative of the known diversity of the species, calculated on 4,083,601 aligned genome positions (including 91,636 variant sites). The YYY087A sample is represented by red stars, YYY092B by yellow circles, YYY095A by blue triangles and YYY097B by green squares. The phylogenetic placements of these samples were done using the epa-ng software[30] and are based on the partial genome sequence reconstructed from reads assigned to the *S. enterica* species by MEGAN (see STAR methods). The previously published ancient genomes are in blue, and modern genomes in black. The lineages are color-coded as follows: purple represents the paratyphi A lineage, blue corresponds to the typhi lineage, green stands for the typhimurium lineage, yellow represents the Birkenhead lineage, and red indicates the para C lineage. The purple dots indicate a bootstrap value of 1. Bootstrap symbols are displayed on the branch leading to each supported node, following iTOL's default parameters. The multiple symbol size represents different possible placements of a single sample on the phylogenetic tree. Each symbol's location indicates a probable placement, with varying sizes determined by the like weight ratio and distal length values. See also Figure S1, Figure S2, Figure S4 and Table S4.

Figure 2. A) Are these really ancient Salmonella reads? Each panel shows how well the DNA fragments from four soldiers match *Salmonella enterica*. On the x-axis is the number of mismatches ("edit distance") between each fragment and the reference genome. With genuine ancient DNA, you expect a downward-sloping curve (many perfect or very good matches, fewer poor ones). The number called –Δ% summarizes this slope: values close to 1 mean a clear ancient-DNA pattern; low values would suggest noise or contamination. B) **Where do these strains sit on the Salmonella family tree?** This tree gathers modern (**black**) and previously published ancient (**blue**) *S. enterica* genomes. Colored bands mark major lineages (e.g., **red = Para C**). The four Napoleonic samples are shown with different symbols (red star ★, yellow circle ●, blue triangle ▲, green square ■). Even though only small parts of their genomes were recovered, a method called "phylogenetic placement" lets us position them on the tree using the genome sites they do cover. Their symbols are located within the Para C (Paratyphi C) group. Purple dots on branches mark very strong support of all the branches of the tree (bootstrap = 1), so we can trust the strcture of the tree in which the sequences were placed. If you see the same symbol drawn in slightly different nearby spots, that reflects alternative but similar placements; larger symbols indicate more likely positioning. **Take-home message:** Panel A shows the DNA behaves like authentic ancient *Salmonella enterica* that is genetically closer to Paratyphi than to other *S. enterica* serovars; Panel B shows that, despite low coverage, the fragment of the genome covered by the recovered reads consistently place the soldiers' infections within the Paratyphi C lineage.

To assess the presence of *S. enterica*, we examined both the number of reads and their alignment profiles across different serovars. In four individuals (YYY087A, YYY092B, YYY095A, YYY097B), between 34 and 968 unique reads (i.e., after duplicates removal) mapped most closely and in greater numbers to the *Paratyphi C* serovar (strain RKS4594), as opposed to *Typhi* or *Typhimurium* (Table 1, Figure 2A). The corresponding edit distance distributions displayed the expected declining pattern consistent with a genuine ancient pathogen signal, and a negative difference proportion of 1.0 (NDP, Table S2)[17]. In contrast, other

Initial strategy used to define the specific types of *Salmonella enterica* and *Borrelia recurrentis*

To confirm the presence of *Salmonella enterica*, the researchers looked not only at how many DNA fragments matched this bacterium, but also at how well these fragments aligned with different sub-groups, known as serovars. A serovar is like a "family branch" within a bacterial species, distinguished by small differences in its surface molecules. These differences can change how the bacterium spreads or which disease it causes. For example, *Salmonella enterica* has many serovars, including Typhi (which causes typhoid fever), Typhimurium (which usually causes food poisoning), and Paratyphi C (which causes paratyphoid fever). In

samples with low read counts mapping to *S. enterica* exhibited flatter or irregular edit distance profiles, and low NDP values and were considered negative hits, likely originating from environmental related species or conserved genomic regions with limited discriminatory power between taxa (Table S2). A similar analysis was performed for *Borrelia* species. In individual YYY093A, 4,062 unique reads mapped to *B. recurrentis*, compared to 1,556 and 1,441 reads mapping to *B. duttonii* and *B. crocidurae*, respectively (Table S1, Figure 3A). Reads mapped across all eight fragments of *B. recurrentis* genome (the chromosome and seven plasmids), and the edit distance profiles again showed a consistent declining pattern and a NDP=1 (Figure 3A , Table S3)[17]. A second individual (YYY092B), who also tested positive for *S. enterica*, yielded 322 unique reads mapping to *B. recurrentis* and an edit distance distribution similarly consistent with a true hit (NDP=1 in Table S3, Figure 3A). All edit distance profiles and downstream analyses were conducted using deduplicated and high-mapping-quality reads (MAPQ > 30).

Given the low number of reads, the breadth of coverage ranged from 0.0003 to 0.009 for *S. enterica* and from 0.002 to 0.14 for *B. recurrentis*. The read-length distribution, with an average of 56 bp for *B. recurrentis* and 43 bp for *S. enterica*, indicated extensive DNA fragmentation, consistent with aDNA degradation patterns[18]. Analysis of post-mortem damage in the 4,062 reads from sample YYY093A that aligned to the modern *B. recurrentis* reference genome (strain A1) revealed that approximately 5% displayed cytosine deamination at read termini (Figure S1C). This is slightly lower than the ~8% observed in human DNA from the same sample (Figure S1D), but within the same range as other individuals from the site (Figure S1E-H) or aDNA samples from the same period[19], and is thus consistent with authentic bacterial aDNA. It is worth noting that lower deamination rates in bacterial compared to human aDNA within the same host remains have been reported previously, suggesting that bacterial DNA can degrade at a different rate and often less extensively than human DNA[20–22]. Unfortunately, for *S. enterica*, the number of mapping reads was too low to reliably assess deamination patterns (Table 1), as several hundred aligned reads are typically needed for accurate estimations.

four individuals (samples YYY087A, YYY092B, YYY095A, and YYY097B), between 34 and 968 unique DNA fragments (after removing duplicates) matched most closely to the Paratyphi C serovar (strain RKS4594), rather than to Typhi or Typhimurium (Table 1 at the bottom of the file, Figure 2A). To check these matches, the team examined the edit distance distribution. Edit distance measures how many “spelling mistakes” or mismatches exist between an ancient DNA fragment and a modern reference genome. With authentic ancient DNA, most fragments are a close match (i.e., most have 0 missmatches, followed much fewer with just 1 missmatch, and so on) which ends up producing a steady downward curve. That is exactly what they observed in these cases, in Figure 2A above. They also used a statistical test called the negative difference proportion (NDP), which checks whether the curve really follows this expected pattern. An NDP of 1.0 means the match is fully consistent with a genuine ancient infection. By contrast, other samples with only a handful of reads mapping to *S. enterica* showed flat or irregular edit distance patterns and had low NDP values. These were considered false signals, likely coming from environmental bacteria in the soil or from shared DNA regions that are too general to prove infection. The same type of analysis was then done for Borrelia species. In one soldier (sample YYY093A), 4,062 unique fragments mapped to *Borrelia recurrentis*, compared to 1,556 for *B. duttonii* and 1,441 for *B. crocidurae* (Figure 3A). Once again, the edit distance curve declined as expected confirming the infection as authentic (see Figure 3A below). A second soldier (sample YYY092B), who had also tested positive for *S. enterica*, had 322 unique reads mapping to *B. recurrentis*, with the same authentic pattern.

First authentication step: find sings of DNA degradation

Because only a relatively small number of reads were recovered, the overall genome coverage was very low: between 0.0003 and 0.009 for *Salmonella enterica*, and between 0.002 and 0.14 for *Borrelia recurrentis*. Genome coverage refers to the fraction of the organism’s full genetic code that is represented by the recovered fragments. These values mean that only tiny portions of the bacterial genomes could be reconstructed. The length of the DNA fragments also showed how damaged the material was. On average, fragments from *B. recurrentis* were 56 base pairs long on average, and those from *S. enterica* were 43 base pairs long. These very short lengths are consistent with the heavy fragmentation that typically occurs in ancient DNA. The researchers also looked for a type of post-mortem DNA damage called cytosine deamination, which is considered a hallmark of ancient DNA. This process happens when one of the DNA letters, cytosine (C), chemically changes after death (changing to U, that is then transformed into T during sequencing library preparation), and is most often visible at the ends of DNA fragments. In the 4,062 *B. recurrentis* reads from sample YYY093A, about 5% showed this damage. For comparison, the human DNA in the same sample showed about 8% damage. Both values fall within the expected range for remains of this age, and are therefore consistent with authentic ancient DNA. Previous studies have also noted that bacterial DNA often shows less deamination than human DNA from the same skeleton, meaning that bacterial DNA can degrade at a slightly different pace. For *S. enterica*, however, there were too few DNA fragments (Table 1) to reliably assess the proportion of reads with deamination. This type of analysis usually requires thousands of fragments, and here the numbers were below that threshold.

To further authenticate the reads assigned to both pathogens, we re-analyzed the deduplicated mapping reads against the full NCBI nt database using BLASTN, followed by taxonomic classification with the LCA algorithm implemented in MEGAN[23] (Figures S1A-B; Tables S2 and S3). This analysis confirmed the initial assignments, with a large proportion of reads attributed to the expected pathogens, while also filtering out those that could not be unambiguously assigned to the corresponding taxa. These results reinforced the identification of four positive samples for *S. enterica* (YYY087A, YYY092B, YYY095A, YYY097B) and one (YYY093A), possibly two (YYY092B, with only 18 on-target hits out of 322 mapped reads), for *B. recurrentis* (Tables S2 and S3).

Second authentication step: Be sure that all the sequences aligning with the reference genomes are truly originating from the presumed species

To further confirm that the DNA fragments that aligned to the reference genome of each pathogen species really belonged to these pathogens, the researchers re-analyzed the deduplicated reads against the complete NCBI nucleotide database (a database that contains the genomes of most sequenced species to date) using BLASTN, a program that compares DNA sequences to all the genetic sequences in the database. They then applied a lowest common ancestor (LCA) algorithm in the program MEGAN, which evaluates all the species/genomes that were "hit" by each read, and then assign each fragment to the most specific taxonomic level possible (species, genus, family, etc.), allowing to filter out uncertain and wrong matches. This second round of analysis confirmed the earlier findings: most of the fragments matched the expected pathogens, while ambiguous ones were excluded, allowing to clean much further the genetic signal of the infecting strains. The results strengthened the identification of four positive cases of *Salmonella enterica* (samples YYY087A, YYY092B, YYY095A, YYY097B) and one clear case of *Borrelia recurrentis* (sample YYY093A), with a possible second in YYY092B, although in this individual only 18 of the 322 mapped fragments were specific enough to count as true hits.

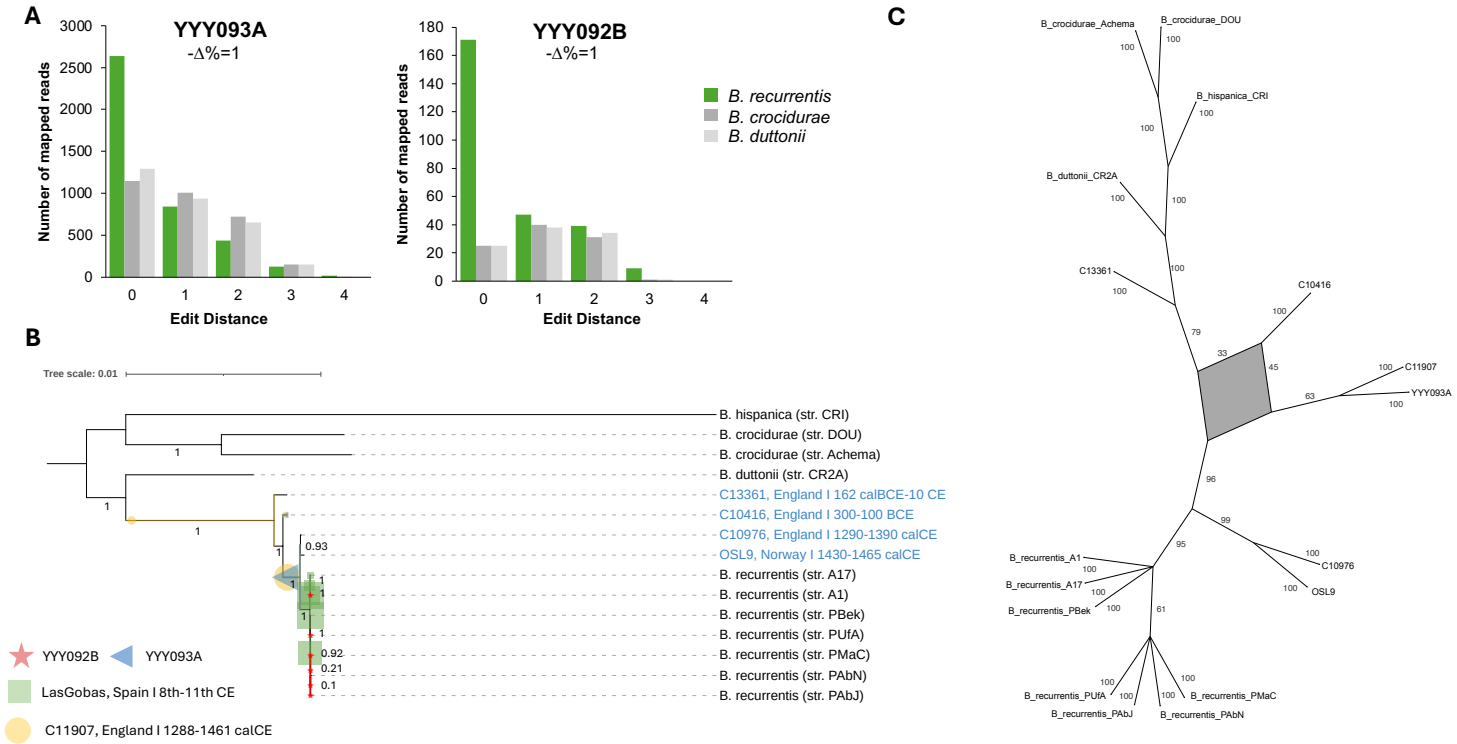

Figure 3. Authentication of *B. recurrentis* aDNA data. (A) Distribution of edit distances of YYY093A and YYY092B mapping reads on three closely related *Borrelia* species, with the associated *negative difference proportion* (-Δ%)[17] (B) Maximum likelihood phylogeny of 15 *Borrelia* sp. genomes calculated on 1,032,378 aligned genome positions (including 39,217 variant sites). The phylogenetic placement done by epa-ng[30] is based on reads classified under the genus *Borrelia* by MEGAN. The YYY093A and YYY092B are represented by a blue triangle and a red star respectively. Due to their low coverage, the two previously published ancient genomes, C11907 and Las Gobas, have been incorporated in this plot by phylogenetic placement, similarly to our study samples, and are indicated by a yellow circle and a green square. The symbol sizes are determined by the like weight ratio and distal length values. The previously published ancient genome is written in blue font, and modern genomes in black. Numerical values correspond to branch bootstrap supports. Bootstrap values are displayed on the branch leading to each supported node, following iTOL's default parameters C) Consensus network constructed using SplitsTree[39] based on 100 bootstrapped trees built from an MSA of 144,470 sites that are covered by YYY093A on-target reads (5,854 variant sites), to assess the nature of the topological uncertainty in the placement of YYY093A (Figure S3). The network reveals that although bootstrap values are low for the YYY093A–C11907 clade (Figure S3E-F), together with C10416 the three strains are consistently grouped within a distinct cluster across the 100 bootstrapped trees. See also Figure S1, Figure S4 and Table S4.

Figure 3. A) **Are these really ancient *Borrelia* reads (and which *Borrelia*)?** For two soldiers (YYY093A and YYY092B), the plots show how well their DNA fragments match three closely related *Borrelia* species. On the x-axis is the number of mismatches ("edit distance") to each reference genome. With genuine ancient DNA, you expect a downward-sloping curve (many perfect or very good matches, fewer poor ones). The summary number –Δ% tells you how clean that slope is: values near 1 support an authentic ancient-DNA signal rather than random noise or contamination. B) **Where do these strains sit on the *Borrelia* family tree?** This tree includes modern (black labels) and previously published ancient (**blue labels**) *Borrelia* genomes. Because only small pieces of pathogen DNA were recovered, a method called phylogenetic placement can still position the overall genomic fragments recovered from each sample on the phylogenetic tree of the species/genus using the sites it does cover. The two Napoleonic samples are marked (YYY093A = **blue triangle**; YYY092B = **red star**). Two published ancient genomes (**yellow** circle and **green** square) are shown the same way for comparison. If you see a symbol drawn in slightly different nearby spots, that reflects alternative but similar placements; larger symbols indicate more likely positions. Numbers on branches are bootstrap supports—higher values mean stronger support for that split in the tree. C) **How certain is the placement of YYY093A?** Because YYY093A covers more sites, the authors also built a consensus network from 100 bootstrap trees using only the genomic positions covered by this sample, even for the samples of the tree with whole genome data (i.e., using the same information from all strains to build the tree). The network shows that—even if some internal branching order is fuzzy, YYY093A consistently clusters with the same group of ancient *B. recurrentis* genomes across the bootstraps, indicating a stable evolutionary neighborhood, close to Iron-Age and Medieval times in England. **Take-home message:** Panel A shows the DNA behaves like authentic ancient Borrelia; Panel B places the soldiers' infections within the B. recurrentis part of the tree despite low coverage; Panel C shows that the deeper placement of YYY093A is consistently supported when uncertainty is examined explicitly.

As a final authentication step, we analyzed the on-target reads identified via BLASTN and MEGAN (B+M) using a phylogenetic placement strategy previously applied to low-coverage ancient genomes[24,25]. To establish a robust reference framework, we first reconstructed whole-genome phylogenies of *S. enterica spp.* and *B. recurrentis*, including both ancient and modern strains representative of the known diversity for each species[26–29]. The methodologies used were similar to those of the original studies, and followed standard practices in ancient pathogen genomics (see STAR Methods), successfully reproducing previously published tree topologies with high bootstrap support. We then applied epa-ng[30] to evaluate the most likely placement of the ancient strains, based on the positions covered by B+M on-target reads across the multiple sequence alignment (MSAs) used to build the reference trees (Tables 1 and S4). EPA-ng places low-coverage genomes onto a fixed reference phylogeny and multiple sequence alignment (MSA) by computing the most likely position for each sample based on its covered genomic sites. This approach is well suited to ancient DNA, where only a small fraction of the genome is typically recovered, and often enables confident lineage assignment even under sparse coverage.

For *B. recurrentis*, on-target reads from sample YYY093A aligned to 125,514 positions in the MSA (of which 5,854 were variant sites), while sample YYY092B aligned to 584 positions covering 21 variant sites, out of a total of 1,032,378 positions in the MSA. For *S. enterica*, the on-target reads from YYY087A, YYY092B, YYY095A, and YYY097B aligned to 26,116, 4,864, 1,236 and 1,382 positions, respectively, and covered 800, 140, 44 and 42 variant sites out of a total MSA length of 4,083,601 positions (see Table 1). Our final *S. enterica* phylogeny consisted of 10 ancient and 20 modern published genomes representative of the known diversity of the species, in which we were able to place the sequences from the four positive Vilnius individuals (YYY087A, YYY092B, YYY095A, YYY097B) within the paratyphi C serovar, a clade that also includes three ancient genomes dating from the 14th to 16th centuries[31–36] (Figures 2B and Figure S2). Together, phylogenetic placements and the genotypes at variant positions (Figure S4A-D) add a much stronger support to our initial serovar estimations. However, despite the consistent placement of all strains within the same serovar, the resolution of our data was not sufficient to determine

Third and final authentication step: Phylogenetic placement of recovered genome fragments

As a final authentication step, the researchers used a method called phylogenetic placement, which helps determine where a DNA sample belongs on the evolutionary "family tree" of a species (called phylogenetic trees). They first built reference trees of *Salmonella enterica* and *Borrelia recurrentis*, using the whole genomes of both modern and ancient strains to capture the known diversity of each bacterium. These trees reproduced the branching patterns seen in earlier studies and were strongly supported by statistical tests, confirming the framework was sound. The ancient DNA fragments recovered from each soldier for each species were then analyzed with a program called epa-ng. This tool takes short and incomplete DNA sequences and calculates where they best fit on the reference tree, based on the few genomic positions that they cover. In other words, even if only a small part of a genome survives, epa-ng can place it in the most likely spot, much like figuring out where a few puzzle pieces belong in an already completed picture. This makes it especially useful for ancient DNA, where only fragments remain, but a reliable lineage assignment is still possible.

For *Borrelia recurrentis*, one soldier (sample YYY093A) had DNA fragments that lined up with 125,514 positions in the genome, including 5,854 places where strains differ from each other. These variant positions are the basis to build trees, or to know where in a tree a sequence belong. Another soldier (YYY092B) showed a much weaker signal, with only 584 positions and 21 variable sites matched, out of a total of over one million possible positions in the reference alignment. For *Salmonella enterica*, the four positive soldiers gave lower coverage overall. One soldier (YYY087A) matched 26,116 positions with 800 variants, another (YYY092B) matched 4,864 positions with 140 variants, while two others (YYY095A and YYY097B) matched just over 1,200 and 1,300 positions, with 44 and 42 variants respectively. These are very small numbers compared to the entire reference genome, which includes over 4 million positions. When the researchers placed these results onto a "family tree" of *S. enterica* built from 10 ancient and 20 modern genomes, all four Napoleonic samples were positioned within the Paratyphi C group, the same branch that contains three ancient genomes from the 14th to 16th centuries (Figure 2B). Both the placement on the

a stronger affiliation with a specific ancient or modern branch within *Paratyphi C*.

For *B. recurrentis*, we reconstructed a reference phylogenetic tree using four ancient and eleven modern publicly available genomes from across the *Borrelia* genus. Our 19th-century strain YYY093A was confidently placed (LWR = 0.89) in a basal position relative to two other ancient strains from the 14th–15th centuries (C10976 and OSL9)[27,29] as well as all modern *B. recurrentis* genomes (Figures 3B, Figure S3A-B, Figure S4E-F). To further explore *B. recurrentis* diversity, we applied our phylogenetic placement approach to two previously published ancient strains that had not been analyzed with this technique. Sample C11907, recovered from England, dated to 1288–1461 calCE and covering ~32% of the genome (Table S4), was also placed in a basal position relative to C10976 and OSL9 (LWR = 0.81), consistent with its original placement by Swali et al.[29] where it clustered near the Iron Age genome C10416 (300–100 BCE). Strikingly, YYY093A exhibited an almost identical phylogenetic placement to C11907 (Figures 3B, Figure S3A-B), and when placed on a tree that included the C11907 partial genome, YYY093A clustered at the base of this strain's branch with very high confidence (LWR = 0.99, Figure S3C-D).

To further investigate the phylogenetic positioning of YYY093A, we built a *de novo* phylogenetic tree restricted to the 144,470 genome positions covered by reads classified as *Borrelia* genus by B+M (Figure S3E-F). In this analysis, YYY093A and C11907 formed a monophyletic clade much closer to the Iron Age genome C10416 than to C10976 or OSL9. Since the bootstrap support for this topology was relatively low (58%), we decided to examine the source of this uncertainty using SplitsTree[37]. By visualizing the 100 bootstrapped trees as a consensus network, we observed that the topological ambiguity was constrained to the subtree encompassing C10416, C11907, and YYY093A, rather than being spread across the entire phylogeny (Figure 3C). This result suggests that YYY093A and C11907 consistently fall within the same evolutionary neighborhood as the Iron Age strain, supporting that these genomes represent a distinct and relatively ancient lineage or differentiated group of *B. recurrentis* that persisted in Europe for at least two millennia and remained in circulation into at least the early 1800s.

By contrast, two lower-coverage strains (YYY092B (this study) and Las Gobas (8th–11th century CE, Spain[28]) were placed within the clade of the near-identical modern *B. recurrentis*

tree and the observed genetic differences provided strong support that the infections were caused by Paratyphi C. However, because the Napoleonic DNA fragments were so limited, the researchers could not tell whether these strains were closer to ancient or to modern branches within Paratyphi C.

A *Borrelia recurrentis* strain from a lineage that persisted for millennia in Europe

For *Borrelia recurrentis*, the team built a reference evolutionary tree using four ancient and eleven modern genomes. The strain from the Napoleonic soldier YYY093A was placed with high confidence (LWR = 0.89) in a basal position, meaning it branched off earlier than both the better-known medieval strains from the 14th–15th centuries (C10976 and OSL9) and all modern *B. recurrentis* strains. (Figure 3B) This indicates that YYY093A belonged to a lineage that is more ancestral than the one grouping these medieval strains. Because there are very few genomes of this species, and to better understand its diversity, the researchers reanalyzed other published ancient DNA data that had not been studied with phylogenetic placement, and applied the same approach used with the strains in this study. One, known as C11907 and recovered from medieval England (1288–1461 CE), had about one-third of its genome preserved. It too was placed in a basal position relative to the later medieval strains (LWR = 0.81), echoing earlier results that had linked it to an even older Iron Age genome (C10416, 300–100 BCE). What is especially striking is that the Napoleonic strain YYY093A and the English strain C11907 ended up in nearly the same spot on the tree. When analyzed together, YYY093A clustered at the base of C11907's branch with very high confidence (LWR = 0.99). This close placement suggests that both strains came from the same ancestral lineage of *B. recurrentis*, a lineage distinct from the more recent strains that circulated in later centuries and survive today.

To explore the position of the Napoleonic strain YYY093A in more detail, the researchers built a new phylogenetic tree using only the 144,470 genome sites of the reference genome that had been confidently covered by reads assigned to the *Borrelia* genus (by the BLASTN/MEGAN analysis). In this analysis, YYY093A and the medieval English strain C11907 grouped together in a monophyletic clade in a shared branch, positioned much closer to the Iron Age strain C10416 than to the later medieval strains C10976 or OSL9. The statistical support for this arrangement, known as bootstrap support, was relatively low at 58%. This means the tree could not be considered fully stable in the area in which these strains were positioned. To understand why, the researchers used a program called SplitsTree to visualize 100 repeated test trees. This showed that the uncertainty was limited to the small portion of the tree containing C10416, C11907, and YYY093A, while the rest of the tree remained stable (Figure 3C). The consistent grouping of YYY093A with C11907, both close to the Iron Age genome, indicated that these strains belonged to a distinct and very ancestral lineage of *B. recurrentis*. This lineage appears to have persisted in Europe for at least two thousand years, continuing to circulate from the Iron Age all the way into the early 19th century.

In contrast to the older lineage represented by YYY093A, two other strains fell much closer to the modern forms of *Borrelia recurrentis*. One was another Napoleonic sample, YYY092B, and

genomes (Figure 3B, Figure S3A,B,D, Figure S4G). While the placement of Las Gobas genome appears well supported (50,657 mapped positions; 2,357 variant sites, Table S4), YYY092B aligned to only 584 positions and covered 21 variants (Table 1), limiting placement resolution. However, the consistency in placement across both strains and the relatively robust signal from Las Gobas suggest that these genomes may indeed represent members of the clade grouping the available modern genomes. Importantly, while the low coverage of YYY092B precludes definitive sub-lineage assignment, it nonetheless supports its classification at the species level. Altogether, our findings support the existence of multiple distinct *B. recurrentis* lineages circulating in Europe during the past centuries.

Recent aDNA work has revealed that paratyphoid fevers have been present in Europe for millennia[26,31–36], and it was already well known and documented by 1812. The disease is transmitted to man through food or water contaminated with infected feces, and symptoms include fever, headache, rash, weakness, loss of appetite, diarrhea, constipation, stomach pain and vomiting[38]. Throughout Napoleon's Russian campaign, paratyphoid or typhoid fever was not mentioned in any historical sources of our knowledge, likely due to these nonspecific and varied symptoms. Yet, an 1812 report from J.R.L. de Kirckhoff, a physician serving in Napoleon's army, contains key information about the events that could potentially explain the origins of an epidemic. In this document, he specifies that soldiers suffered from typhus, dysentery, and diarrhea on their arrival in Vilnius. Insisting on this last aspect, he wrote "Diarrhea was common among us in Lithuania. One powerful contributing factor to this illness was that we encountered in almost every house, from Orcha to Wilna, large barrels of salted beets (*buraki kwaszone*), which we ate and drank the juice of when we were thirsty, greatly upsetting us and strongly irritating the intestinal tract"[5]. This description could be consistent with both the characteristics of a paratyphoid fever infection caused by contaminated food, and the digestive symptoms typically associated with the disease, although we acknowledge that they can also match various other diseases that were common in the 19th-century in Europe. Furthermore, even today, two centuries later, it would still be impossible to perform a differential diagnosis between typhus, typhoid or paratyphoid fever based solely on the symptoms or the testimonies of survivors.

From a molecular perspective, our findings provide strong support that the soldiers were infected with paratyphoid fever caused by *S. enterica* Paratyphi C. Although we did not recover sufficient genome coverage to determine the specific phylogenetic positioning within the known diversity of this lineage, our thorough

the other came from Las Gobas in Spain, dated to the 8th–11th centuries CE. Both were placed in the clade that contains the nearly identical genomes of present-day *B. recurrentis*. The Las Gobas strain was well supported, with over 50,000 genome positions covered and more than 2,000 points of variation identified. YYY092B, however, had far fewer matches, only 584 positions and 21 variants, which limited the confidence of its exact placement. Even so, the fact that both YYY092B and Las Gobas consistently aligned with the modern-like branch suggests that they belonged to the same lineage as today's *B. recurrentis*, although this remains just a hypothesis. Conceptually, this means that by the Middle Ages, and continuing into the early 1800s, Europe was home to at least two different versions of the bacterium: one related to the ancestral lineage, represented by YYY093A and C11907, that stretched back to the Iron Age, and another related to a lineage that was already very similar to the modern form of the disease. This highlights that relapsing fever in Europe was not caused by a single strain, but by multiple coexisting lineages with different histories.

Discussion of the results, in relation to what is known from historical records

Recent research on ancient DNA has shown that paratyphoid fevers, caused by *Salmonella enterica* Paratyphi C, have been present in Europe for thousands of years, and by 1812 the disease was already familiar and documented. It spreads when food or water is contaminated with human waste carrying the bacteria. The symptoms are wide-ranging and non-specific, including fever, headache, rash, weakness, loss of appetite, stomach pain, diarrhea, constipation, and vomiting. During Napoleon's Russian campaign, however, neither paratyphoid nor typhoid fever was specifically mentioned in the records we know of. This lack of reference to this disease is likely because the symptoms were so varied and could easily be mistaken for other common illnesses of the time. Still, one 1812 report by J.R.L. de Kirckhoff, a physician serving in Napoleon's army, gives a telling description. He wrote that the soldiers arriving in Vilnius suffered from typhus, dysentery, and especially diarrhea. He emphasized that a likely cause was their habit of consuming barrels of salted beets, found in almost every house from Orcha to Vilnius. Soldiers not only ate the beets but also drank the salty juice when thirsty, which, according to him, upset their stomachs and irritated their intestines. This account fits well with the possibility of paratyphoid fever infection, since contaminated food and water are typical sources of the disease and diarrhea is a classic symptom. At the same time, the same signs could also match other common 19th-century illnesses, such as dysentery or typhus. In fact, even today, with modern medicine, it would still be impossible to distinguish between typhus, typhoid, and paratyphoid fever based on symptoms and testimonies alone, showing how easily such epidemics could be misidentified in the past.

Limitations of the study, and how far we can interpret the results

From the genetic evidence, the study provides strong support that some of Napoleon's soldiers were infected with paratyphoid fever, caused by *Salmonella enterica* Paratyphi C. Even though the team could not recover enough of the genome to place these strains precisely within the broader diversity of this

authentication workflow resulted in a solid proof of its presence. Our study thus provides the first direct evidence that paratyphoid fever contributed to the deaths of Napoleonic soldiers during their catastrophic retreat from Russia. However, the limited number of samples that were processed (n=13), in relation to the large number of reported bodies in this site (over 3000), is not sufficient evidence to conclude that this pathogen alone contributed to all the deaths at the site. Considering the extreme and harsh conditions of this retreat, the presence of multiple overlapping infections is highly plausible. Typhus has long been reported to have affected Russia in this period, but the existing evidence remains inconclusive to support a role of this disease in the devastation of Napoleon's Army.

In light of our results, a reasonable scenario for the deaths of these soldiers would be a combination of fatigue, cold, and several diseases, including paratyphoid fever and louse-borne relapsing fever. While not necessarily fatal, the louse borne relapsing fever could significantly weaken an already exhausted individual. Our study confirms the presence of two previously undocumented pathogens, but the analysis of a larger number of samples will be necessary to fully understand the spectrum of epidemic diseases that impacted the Napoleonic army during the Russian retreat. Our work demonstrates that high-throughput sequencing of ancient DNA is a powerful approach for investigating historical disease dynamics and underscores its capacity to accurately identify ancient pathogens, even when only limited genomic data are available.

bacterium, the careful authentication of the DNA fragments leaves little doubt about its presence. This makes the study the first to offer direct proof that paratyphoid fever could have contributed to the deaths of soldiers during the retreat from Russia. At the same time, the results need to be kept in perspective. Only 13 soldiers were analyzed, compared with more than 3,000 bodies reported at the site. This is far too small a sample to conclude that paratyphoid fever was a major cause of the mass deaths. The retreat was marked by exhaustion, freezing temperatures and starvation, all conditions that could have killed the soldiers by themselves, or that would have made soldiers vulnerable to diverse infections. Typhus, for example, has long been suspected as a major factor in this period, but the available evidence remains too inconclusive to be fully sure of its presence and the impact it may have had.

Taken together, the results suggest that the soldiers' deaths were not caused by a single disease, but by a combination of extreme exhaustion, freezing conditions, and multiple infections. Paratyphoid fever and louse-borne relapsing fever, both identified in this study, likely contributed to the deaths in this site. Relapsing fever on its own is not always fatal, but in soldiers who were already starving, freezing, and weakened, it could have made survival more difficult. This study confirms the presence of two pathogens that had not previously been documented in Napoleon's army. However, since only a small number of soldiers were analyzed, many more samples will need to be studied to gain a complete picture of the epidemic diseases that struck during the Napoleonic's retreat from Russia. Beyond this specific case, the work also shows how modern DNA sequencing technologies can open a window onto past epidemics. Even when only small and damaged fragments survive, these methods allow us to identify pathogens with remarkable accuracy and reconstruct the role disease played in history.

Pathogen	Sample ID	Unique mapped reads	Unique authenticated reads	Avg. read length (bp)	Total mapped positions	Non-N positions in epa-ng MSA	Variant sites covered in epa-ng MSA
S. enterica	YYY087A	968	874	45	29,725	26,116	800
	YYY092B	225	209	42	6,471	4,864	140
	YYY095A	34	29	58	1,374	1,236	44
	YYY097B	74	54	40	1,568	1,382	42
B. recurrentis	YYY093A	4,062	3,239	60	144,470	125,514	5,854
	YYY092B	322	21	45	718	584	21
	Las Gobas (*)	999	760	96	60,432	50,657	2,357
	C11907 (*)	14,027	13,235	44	312,604	270,532	13,392

Table 1. Summary of sequencing, authentication, and phylogenetic placement statistics for ancient samples positive for *Salmonella enterica* and *Borrelia recurrentis*. This table summarizes, in one place, both how much ancient DNA the team recovered and how useful it was for identifying and placing each infection. "Unique mapped reads" are the deduplicated fragments that matched a pathogen's genome; "authenticated reads" are the portion of those fragments independently confirmed as truly belonging to that pathogen (not environmental DNA) by a stricter taxonomic check; the "average read

length" in base pairs indicates how broken the ancient DNA is (shorter means more degraded); "total mapped positions" counts how many spots in the pathogen genome were covered by at least one fragment; "non-N positions in the alignment" are the usable sites from those fragments that could be compared against a reference multiple-sequence alignment for tree placement; and "variant sites covered" are the subset of those usable sites that actually differ between strains and therefore carry the most evolutionary signal. Read these columns together: higher counts of authenticated fragments, more covered positions, and more variant sites all strengthen confidence in detection and in where a sample falls on the pathogen's family tree. That's why, for *B. recurrentis*, YYY093A (4,062 unique mapped reads; 3,239 authenticated; 125,514 usable alignment positions; 5,854 variants) provides very strong evidence, whereas YYY092B (322; 21; 584; 21) is much weaker. The same pattern appears for *S. enterica*: YYY087A (968; 874; 26,116; 800) is far more informative than YYY095A (34; 29; 1,236; 44). The two previously published medieval *Borrelia* genomes, especially C11907 (14,027; 13,235; 270,532; 13,392), have far higher coverage and variant counts and thus serve as robust reference points. Overall, the table shows uneven preservation across individuals, as expected with ancient remains, but also provides thorough, quantitative support that *S. enterica* and *B. recurrentis* were present in some of Napoleon's soldiers and that several samples carry enough signal to place them confidently in the evolutionary analyses.

Table 1. Summary of sequencing, authentication, and phylogenetic placement metrics for ancient pathogen-positive samples. "**Unique mapped reads**" are duplicate-filtered fragments matching the pathogen genome; "**authenticated reads**" are those confirmed through the three authentication steps to belong to that pathogen (not background DNA); "**average read length (bp)**" reflects how fragmented the ancient DNA is; "**total mapped positions**" counts genome sites covered by ≥1 fragment; "**non-N positions in the alignment**" are the usable sites for phylogenetic placement; "**variant sites covered**" are the informative differences between strains that are effectively used in phylogenetic inferences. Higher values across these columns strengthen both detection and placement (e.g., *B. recurrentis* YYY093A >> YYY092B; *S. enterica* YYY087A >> YYY095A). See also Figures 2–3 and Table S4.

References and Notes

REFERENCES

1. Lieven, D. (2010). Russia against Napoleon: The true story of the campaigns of war and peace (Penguin).

2. Talty, S. (2009). The Illustrious Dead: The Terrifying Story of How Typhus Killed Napoleon's Greatest Army (Crown).

3. Raoult, D., Dutour, 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.

4. Bédarida, S., Dutour, O., Buzhilova, A.P., de Micco, P., and Biagini, P. (2011). Identification of viral DNA (Anelloviridae) in a 200-year-old dental pulp sample (Napoleon's Great Army, Kaliningrad, 1812). Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. *11*, 358–362. https://doi.org/10.1016/j.meegid.2010.11.007.

5. De Kirckhoff, J.R.L.D. (1836). Histoire des maladies observées à la Grande Armée française: pendant les campagnes de Russie en 1812 et d'Allemagne en 1813 (Janssens).

6. Signoli, M., Ardagna, Y., Adalian, P., Devriendt, W., Lalys, L., Rigeade, C., Vette, T., Kuncevicius, A., Poskiene, J., Barkus, A., et al. (2004). Discovery of a mass grave of Napoleonic period in Lithuania (1812, Vilnius). Comptes Rendus Palevol *3*, 219–227. https://doi.org/10.1016/j.crpv.2004.02.003.

7. Paratyphoid fever https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/paratyphoid-fever.

8. Warrell, D.A. (2019). Louse-borne relapsing fever (Borrelia recurrentis infection). Epidemiol. Infect. *147*, e106.

9. Chandler, D.G. (2009). The Campaigns of Napoleon (Simon and Schuster).

10. Austin, P.B. (2012). 1812: The Great Retreat (Frontline Books).

11. Larrey, D.J.B. (1812). Mémoires de chirurgie militaire et campagne (chez J. Smith).

12. Lejeune, L.-F. (1896). Mémoires du général Lejeune (Librairie de Firmin-Didot et Cie, imprimeurs de l'Institut).

13. Boulangier, V.A. (1820). Dissertation sur le typhus contagieux, observé à Auxerre à la fin de 1811, et au commencement de 1812, etc.

14. Breitwieser, F.P., Baker, D.N., and Salzberg, S.L. (2018). KrakenUniq: confident and fast metagenomics classification using unique k-mer counts. Genome Biol. *19*, 198. https://doi.org/10.1186/s13059-018-1568-0.

15. Wattam, A.R., Davis, J.J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E.M., Disz, T., and Gabbard, J.L. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. Nucleic Acids Res. *45*, D535–D542.

16. Pochon, Z., Bergfeldt, N., Kırdök, E., Vicente, M., Naidoo, T., Van Der Valk, T., Altınışık, N.E., Krzewińska, M., Dalén, L., Götherström, A., et al. (2023). aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. Genome Biol. *24*, 242. https://doi.org/10.1186/s13059-023-03083-9.

17. Hübler, R., Key, F.M., Warinner, C., Bos, K.I., Krause, J., and Herbig, A. (2019). HOPS: automated detection and authentication of pathogen DNA in archaeological remains. Genome Biol. *20*, 280. https://doi.org/10.1186/s13059-019-1903-0.

18. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., De Filippo, C., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science *338*, 222–226. https://doi.org/10.1126/science.1224344.

19. Sandoval-Velasco, M., Jagadeesan, A., Ramos-Madrigal, J., Ávila-Arcos, M.C., Fortes-Lima, C.A., Watson, J., Johannesdóttir, E., Cruz-Dávalos, D.I., Gopalakrishnan, S., and Moreno-Mayar, J.V. (2023). The ancestry and geographical origins of St Helena's liberated Africans. Am. J. Hum. Genet. *110*, 1590–1599.

20. Lopopolo, M., Avanzi, C., Duchene, S., Luisi, P., De Flamingh, A., Ponce-Soto, G.Y., Tressieres, G., Neumeyer, S., Lemoine, F., Nelson, E.A., et al. (2025). Pre-European contact leprosy in the Americas and its current persistence. Science, eadu7144. https://doi.org/10.1126/science.adu7144.

21. Nguyen-Hieu, T., Aboudharam, G., and Drancourt, M. (2012). Heat degradation of eukaryotic and bacterial DNA: an experimental model for paleomicrobiology. BMC Res. Notes *5*, 528. https://doi.org/10.1186/1756-0500-5-528.

22. 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. https://doi.org/10.1126/science.1238286.

23. Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., and Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLOS Comput. Biol. *12*, e1004957. https://doi.org/10.1371/journal.pcbi.1004957.

24. Seersholm, F.V., Sjögren, K.-G., Koelman, J., Blank, M., Svensson, E.M., Staring, J., Fraser, M., Pinotti, T., McColl, H., Gaunitz, C., et al. (2024). Repeated plague infections across six generations of Neolithic Farmers. Nature *632*, 114–121. https://doi.org/10.1038/s41586-024-07651-2.

25. Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., de la Fuente Castro, C., Allentoft, M.E., de Barros Damgaard, P., Hansen, A.J., Holtsmark Nielsen, S., Strand, L.M., et al. (2020). Diverse variola virus (smallpox) strains were widespread in northern Europe in the Viking Age. Science *369*, eaaw8977. https://doi.org/10.1126/science.aaw8977.

26. Neumann, G.U., Skourtanioti, E., Burri, M., Nelson, E.A., Michel, M., Hiss, A.N., McGeorge, P.J.P., Betancourt, P.P., Spyrou, M.A., Krause, J., et al. (2022). Ancient Yersinia pestis and Salmonella enterica genomes from Bronze Age Crete. Curr. Biol. CB *32*, 3641-3649.e8. https://doi.org/10.1016/j.cub.2022.06.094.

27. Guellil, M., Kersten, O., Namouchi, A., Bauer, E.L., Derrick, M., Jensen, A.Ø., Stenseth, N.C., and Bramanti, B. (2018). Genomic blueprint of a relapsing fever pathogen in 15th century Scandinavia. Proc. Natl. Acad. Sci. U. S. A. *115*, 10422–10427. https://doi.org/10.1073/pnas.1807266115.

28. Rodríguez-Varela, R., Yaka, R., Pochon, Z., Sanchez-Pinto, I., Solaun, J.L., Naidoo, T., Guinet, B., Pérez-Ramallo, P., Lagerholm, V.K., de Anca Prado, V., et al. (2024). Five centuries of consanguinity, isolation, health, and conflict in Las Gobas: A Northern Medieval Iberian necropolis. Sci. Adv. *10*, eadp8625. https://doi.org/10.1126/sciadv.adp8625.

29. Swali, P., Booth, T., Tan, C.C.S., McCabe, J., Anastasiadou, K., Barrington, C., Borrini, M., Bricking, A., Buckberry, J., Büster, L., et al. (2024). Ancient Borrelia genomes document the evolutionary history of louse-borne relapsing fever. Preprint at bioRxiv, https://doi.org/10.1101/2024.07.18.603748 https://doi.org/10.1101/2024.07.18.603748.

30. Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019). EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. Syst. Biol. *68*, 365–369. https://doi.org/10.1093/sysbio/syy054.

31. 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. https://doi.org/10.1038/s41559-017-0446-6.

32. Haller, M., Callan, K., Susat, J., Flux, A.L., Immel, A., Franke, A., Herbig, A., Krause, J., Kupczok, A., Fouquet, G., et al. (2021). Mass burial genomics reveals outbreak of enteric paratyphoid fever in the Late Medieval trade city Lübeck. iScience *24*, 102419. https://doi.org/10.1016/j.isci.2021.102419.

33. 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. CB *28*, 2420-2428.e10. https://doi.org/10.1016/j.cub.2018.05.058.

34. Key, F.M., Posth, C., Esquivel-Gomez, L.R., Hübler, 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. https://doi.org/10.1038/s41559-020-1106-9.

35. Wu, X., Ning, C., Key, F.M., Andrades Valtueña, A., Lankapalli, A.K., Gao, S., Yang, X., Zhang, F., Liu, L., Nie, Z., et al. (2021). A 3,000-year-old, basal S. enterica lineage from Bronze Age Xinjiang suggests spread along the Proto-Silk Road. PLoS Pathog. *17*, e1009886. https://doi.org/10.1371/journal.ppat.1009886.

36. de-Dios, T., Carrión, P., Olalde, I., Llovera Nadal, L., Lizano, E., Pàmies, D., Marques-Bonet, T., Balloux, F., van Dorp, L., and Lalueza-Fox, C. (2021). Salmonella enterica from a soldier from the 1652 siege of Barcelona (Spain) supports historical transatlantic epidemic contacts. iScience *24*, 103021. https://doi.org/10.1016/j.isci.2021.103021.

37. Huson, D.H., and Bryant, D. (2024). The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. Nat. Methods *21*, 1773–1774.

38. Bhan, M.K., Bahl, R., and Bhatnagar, S. (2005). Typhoid and paratyphoid fever. Lancet Lond. Engl. *366*, 749–762. https://doi.org/10.1016/S0140-6736(05)67181-4.

39. Huson, D.H., and Bryant, D. (2024). The SplitsTree App: interactive analysis and visualization using phylogenetic trees and networks. Nat. Methods *21*, 1773–1774.

40. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.

41. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88. https://doi.org/10.1186/s13104-016-1900-2.

42. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

43. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

44. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421. https://doi.org/10.1186/1471-2105-10-421.

45. Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007). MEGAN analysis of metagenomic data. Genome Res. *17*, 377–386. https://doi.org/10.1101/gr.5969107.

46. Picard Tools - By Broad Institute https://broadinstitute.github.io/picard/.

47. 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. https://doi.org/10.1093/bioinformatics/btt193.

48. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

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

50. Lemoine, F., and Gascuel, O. (2021). Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. NAR Genomics Bioinforma. *3*, lqab075. https://doi.org/10.1093/nargab/lqab075.

51. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

52. Team, R.C. (2020). R A language and environment for statistical computing, R Foundation for Statistical. Computing.

53. Racine, J.S. (2012). Rstudio: A Platform-Independent Ide for R and Sweave. J. Appl. Econom. *27*, 167–172.

54. Tranié, J., and Carmigniani, J.C. (1997). Napoléon : 1812, la campagne de Russie (Pygmalion/G. Watelet).

55. Coignet, J.-R. (2022). Les cahiers du Capitaine Coignet (DigiCat).

56. Marbot, baron de, Jean-Baptiste-Antoine-Marcelin, Dutourd, J., and Garnier, J. (1983). Mémoires du général Baron de Marbot (Mercure de France).

57. Signoli, M., and Vette, T. (2013). Mémoires d'os : la fouille et l'étude d'un charnier de soldats de la Grande Armée. In (Musée de l'Armée/Somogy Editions d'Art), p. 60.

58. Signoli, M., Vette, T., Dutour, O., and Ardagna, Y. (2008). Les Oubliés de la Retraite de Russie; Vilna 1812-Vilnius 2002. (Libraiire Historique Teissedre Paris).

59. Palubeckaitė, Ž., Jankauskas, R., Ardagna, Y., Macia, Y., Rigeade, C., Signoli, M., and Dutour, O. (2006). Dental status of Napoleon's Great Army's (1812) mass burial of soldiers in Vilnius: childhood peculiarities and adult dietary habits. Int. J. Osteoarchaeol. *16*, 355–365. https://doi.org/10.1002/oa.846.

60. Keller, M., and Scheib, C.L. (2023). Decontamination of tooth roots/petrous bone cores for ancient DNA extraction.

61. Keller, M., and Scheib, C.L. (2023). Ancient DNA extract purification (chunk samples/high volume).

62. Keller, M., Scheib, C.L., and Bonucci, B. (2023). Library preparation (dsDNA double indexing, non-UDG, 2x split).

63. Spyrou, M.A., Bos, K.I., Herbig, A., and Krause, J. (2019). Ancient pathogen genomics as an emerging tool for infectious disease research. Nat. Rev. Genet. *20*, 323–340.

64. Bettisworth, B., Psonis, N., Poulakakis, N., Pavlidis, P., and Stamatakis, A. (2025). Read Length Dominates Phylogenetic Placement Accuracy of Ancient DNA Reads. Mol. Biol. Evol., msaf006.

65. Martiniano, R., De Sanctis, B., Hallast, P., and Durbin, R. (2022). Placing ancient DNA sequences into reference phylogenies. Mol. Biol. Evol. *39*, msac017.

Resource availability

Lead contact
Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Nicolás Rascovan.

Material availability
This study did not generate any new reagents

Data and code availability
- Raw sequencing data from the 13 sequenced individuals have been deposited at the SRA Archive under the bioproject PRJNA1188378.
- The complete source code used in this study is available from GitHub (https://github.com/Metapaleo/Napoleon1812).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgments
We want to thank the HPC Core Facility from Institut Pasteur for their support for computational analyses and Marc Monot and Laurence Motreff from the Biomics Platform (supported by France Génomique (ANR-10-INBS-09-09 and IBISA) for their assistance in sample sequencing. This project was made possible through the following funding sources: ERC-2020-STG - PaleoMetAmerica – 948800 (NR), Institut Pasteur and CNRS UMR funding (NR) and INCEPTION program (Investissement d'Avenir grant ANR-16-CONV-0005, NR).

Author's contribution
RB: conceptualization, methodology, validation, investigation, data curation, original draft preparation, writing, review and editing, visualization, project administration; JF: methodology, validation, investigation, data curation, writing, review and editing, visualization; HK: methodology, validation, investigation, writing, review and editing, visualization; CLS: methodology, validation, investigation, writing, review and editing, visualization; MS: conceptualization, methodology, validation, investigation, resources, data curation, original draft preparation, writing, review and editing, visualization, supervision, project administration, funding acquisition; CC: conceptualization, methodology, validation, investigation, resources, data curation, original draft preparation, writing, review and editing, visualization, supervision, project administration; NR: conceptualization, methodology, validation, investigation, writing, review and editing, visualization, supervision, project administration, funding acquisition;

All authors have read and agreed to the published version of the manuscript.

Declaration of interests: The authors declare no competing interests.