

La fièvre paratyphoïde et la fièvre récurrente détectée dans l'armée dévastée de Napoléon en 1812.

Résumé :

Pendant la retraite de Napoléon de Russie en 1812¹, d'innombrables soldats de l'armée française sont morts de maladies infectieuses, mais l'agent ou les agents pathogènes responsables restent débattus²⁻⁵. Nous avons extrait et séquencé l'ADN ancien provenant des dents de treize soldats qui, selon les sources historiques, seraient morts de maladies infectieuses, afin d'identifier les agents responsables de leurs décès⁶. Nos résultats ont confirmé la présence de *Salmonella enterica* subsp. *enterica*, appartenant à la lignée Para C, agent de la fièvre paratyphoïde⁷, ainsi que de *Borrelia recurrentis*, responsable de la fièvre récurrente transmise par les poux de corps⁸. En revanche, nous n'avons pas détecté *Rickettsia prowazekii* (agent du typhus) ni *Bartonella quintana* (agent de la fièvre des tranchées), précédemment associés à cet épisode meurtrier sur la base de résultats PCR et de descriptions symptomatiques historiques³. La présence de ces agents pathogènes jusque-là insoupçonnés chez ces soldats suggère qu'ils ont pu contribuer à la dévastation de la Grande Armée de Napoléon lors de sa désastreuse retraite de 1812.

Résumé en une phrase : La fièvre paratyphoïde et la fièvre récurrente transmise par les poux ont contribué à la mort des soldats de Napoléon lors de la retraite de Russie en 1812.

Main Text

En juin 1812, Napoléon Ier, l'empereur Français, rassembla une force militaire d'environ 500 000 à 600 000 soldats pour envahir la Russie⁹. Après être arrivé à Moscou sans avoir réussi à vaincre décisivement l'armée russe, les troupes napoléoniennes, se retrouvant isolées dans une ville en ruines, décidèrent d'entamer la retraite et d'établir des camps d'hiver le long de la frontière polonaise en octobre de la même année. La retraite de Russie s'étendit du 19 octobre au 14 décembre 1812¹⁰ et entraîna la perte de presque toute l'armée napoléonienne. Selon les historiens, ce ne fut pas les interventions de l'armée russe qui causa la mort d'environ 300 000 hommes², mais plutôt le froid rigoureux de l'hiver russe, conjugué à la faim et aux maladies. Un médecin de la campagne de Russie, J. R. L. de Kirckhoff, rédigea un ouvrage décrivant les maladies ayant touché les soldats en 1812. Il y mentionne notamment la fréquence du typhus, de la diarrhée, de la dysenterie, des fièvres, de la pneumonie et de l'ictère⁵. D'autres médecins¹¹, ainsi que des officiers¹², firent des observations similaires sur les affections qui touchaient les soldats.

Différentes maladies infectieuses, telles que le typhus, ont été décrites dans les régiments français avant même le début de la campagne de Russie¹³. Le typhus, en particulier, souvent appelé « fièvre des camps » en raison de son association fréquente avec les armées, a longtemps été soupçonné d'être la principale cause infectieuse de la disparition de la Grande Armée en 1812. Cette hypothèse a

La fièvre paratyphoïde et la fièvre récurrente détectée dans l'armée dévastée de Napoléon en 1812.

Texte simplifié pour un public non scientifique

Résumé :

Lors de la retraite de Russie de Napoléon en 1812, de nombreux soldats français sont morts de maladies infectieuses, mais les bactéries responsables restent débattues. Pour enquêter, nous avons extrait et séquencé de l'ADN ancien à partir des dents de 13 soldats qui, selon les archives historiques, sont probablement morts de maladies infectieuses. Nous avons détecté *Salmonella enterica* Paratyphi C, une branche des bactéries de type *Salmonella* responsable de la fièvre paratyphoïde, et *Borrelia recurrentis*, une bactérie transmise par les poux du corps qui cause la fièvre récurrente. En revanche, nous n'avons pas trouvé *Rickettsia prowazekii* (l'agent du typhus) ni *Bartonella quintana* (l'agent de la fièvre des tranchées), ien que ces bactéries aient déjà été associées à la retraite de Russie dans des travaux antérieurs portant sur l'ADN ancien et/ou sur des sources historiques. La découverte inattendue de ces bactéries chez ces soldats suggère qu'elles ont pu contribuer à la dévastation de la Grande Armée de Napoléon pendant la retraite de Russie.

Résumé en une phrase: Paratyphoid fever and louse-borne relapsing fever contributed to the deaths of Napoleon's soldiers during the 1812 retreat from Russia.

Texte principal

Contexte historique

En juin 1812, l'empereur Napoléon Ier rassembla une immense armée de 500 000 à 600 000 soldats pour envahir la Russie. La campagne ne se déroula pas comme prévu : après avoir atteint Moscou sans décisivement vaincre les forces russes, les troupes de Napoléon se retrouvèrent bloquées dans une ville en ruines et vide. En octobre, les troupes Napoléoniennes commencèrent une longue retraite depuis Moscou, cherchant à atteindre la Pologne pour y établir leur camp d'hiver. Cette retraite, qui dura du 19 octobre au 14 décembre 1812, détruisit quasiment entièrement l'armée française. En effet, presque tous les soldats de Napoléon furent perdus, les historiens estiment qu'environ 300 000 hommes moururent. Selon la plupart des sources historiques, ce ne furent pas seulement les attaques russes qui causèrent ces morts, mais les conditions hivernales glaciales, la famine et les maladies. Un médecin de l'armée française, J.R.L. de Kirckhoff, décrivit plus tard les maladies qui frappèrent les soldats pendant la campagne. Il rapporta de nombreux cas de typhus, de diarrhée, de dysenterie, de fièvres, de pneumonie et de jaunisse. D'autres médecins et officiers militaires écrivirent également sur les mêmes maladies sévissant dans l'armée.

Les études précédentes mentionnant des maladies infectieuses sur ce site.

Avant même le début de la campagne de Russie, plusieurs maladies, dont le typhus, avaient déjà frappé les régiments français. Le typhus, souvent appelé « fièvre des camps » parce qu'il se propage facilement dans les armées, est depuis longtemps suspecté d'avoir été un acteur majeur du désastre de 1812. Cette hypothèse a été renforcée par deux découvertes faites sur le site de Vilnius, en Lituanie : premièrement, des poux de corps (les insectes qui transmettent le

été renforcée par la découverte de poux du corps, principaux vecteurs du typhus, parmi les restes de soldats napoléoniens morts lors de la Grande Retraite de Russie en décembre 1812 à Vilnius, en Lituanie, ainsi que par l'identification supposée de séquences de *R. prowazekii* et *B. quintana* amplifiées par PCR nichée (nested PCR) chez certains de ces individus³. Cependant, cette première étude était limitée par les technologies disponibles à l'époque et reposait uniquement sur l'amplification de deux courts fragments d'ADN (de 192 et 429 paires de bases), ce qui ne permettait pas d'obtenir une preuve non ambiguë de la présence de ces agents pathogènes dans l'armée napoléonienne. Plusieurs années plus tard, une autre étude a réussi à détecter de l'ADN ancien (aDNA) viral appartenant à la famille des Anelloviridae chez des soldats napoléoniens retrouvés à Kaliningrad datant de 1812⁴. Toutefois, ces virus sont omniprésents et asymptomatiques dans les populations humaines, et ne peuvent donc être associés au décès de ces soldats. En utilisant des méthodologies d'ADN ancien de pointe, nous avons réanalysé des échantillons de soldats napoléoniens morts à Vilnius et identifié du matériel génétique spécifique à certains agents pathogènes, fournissant ainsi une preuve directe de la présence d'agents infectieux ayant pu contribuer à l'effondrement de l'armée.

Les données de séquençage obtenues à partir de treize dents provenant de Vilnius (Figure 1), comportant chacune environ 20 millions de lectures, ont d'abord été classées taxonomiquement à l'aide de KrakenUniq¹⁴, en rapport avec l'ensemble de la base de données microbienne du NCBI. Nous avons ensuite recherché les agents pathogènes humains connus en comparant les TaxIDs identifiés à une liste de référence de 535 TaxIDs correspondant à 185 espèces bactériennes pathogènes pour l'humain, issue de la base de données PATRIC (voir Document S2)¹⁵. À partir de cette première analyse, nous avons établi une liste de taxons présentant plus de 200 lectures uniques attribuées à des agents pathogènes connus dans au moins un échantillon, un seuil défini comme optimal pour la détection métagénomique d'agents pathogènes anciens selon Pochon et al., 2023¹⁶. Ce critère a permis d'identifier quatorze

typhus) ont été retrouvés sur les restes de soldats ; deuxièmement, Ensuite, une étude ADN sur ces soldats avait détecté des traces des bactéries responsables du typhus (*Rickettsia prowazekii*) et de la fièvre des tranchées (*Bartonella quintana*), grâce à un test très sensible appelé PCR nichée (nested PCR). Ce test initial est puissant mais délicat. Il permet d'amplifier de très petites quantités d'ADN (en produisant de nombreuses copies identiques d'une séquence spécifique), mais il est aussi sujet à capter de l'ADN étranger (contaminations ou ADN non ciblé). Dans cette étude, les conditions expérimentales n'étaient pas très strictes (la « température » à laquelle les brins d'ADN étaient autorisés à se fixer pour générer les copies était basse, ce qui permet une hybridation non spécifique), de sorte que des séquences similaires, mais pas exactement identiques à celles recherchées, pouvaient être copiées. Cette étude a identifié quatre fragments amplifiés de *R. prowazekii* et 10 fragments de *B. quintana*, lesquels, comparés à une base de données exhaustive (NCBI nt) regroupant tous les organismes séquencés à ce jour montre que les séquences amplifiées restent plus proches des espèces pathogènes ciblées que de toute autre présente dans la base (ce qui reste un bon critère d'identification). Toutefois, une limite potentielle supplémentaire réside dans le fait que les fragments d'ADN rapportés dans cette étude étaient assez longs pour du matériel ancien (192 et 429 « lettres », paires de bases), alors que l'ADN ancien survit généralement sous forme de fragments beaucoup plus courts (souvent inférieurs à ~100 paires de bases). Idéalement, plusieurs régions du génome devraient être testées ou couvertes pour accroître la fiabilité de la détection. Dans l'ensemble, ces remarques signifient que les signaux précédemment détectés restent valides, mais doivent être interprétés avec prudence. Plus tard, une autre équipe a trouvé de l'ADN provenant d'un groupe de virus appelés Anelloviridae dans les restes de soldats Napoléoniens retrouvés à Kaliningrad. Mais comme ces virus sont extrêmement courants chez les humains et ne sont pas pathogènes, ils ne peuvent pas être tenus pour responsables de la mort des soldats. Dans notre étude, nous avons appliqué des techniques modernes et avancées d'analyse de l'ADN ancien conçues pour des restes anciens, endommagés et dégradés. Ces méthodes nous ont permis de rechercher de manière non-biaisée la présence de matériel génétique de pathogènes (au lieu de cibler quelques pathogènes d'intérêts, comme avec la PCR) et d'obtenir de nouvelles preuves directes, claires et non ambiguës de la présence de maladies infectieuses ayant pu contribuer à l'effondrement de l'armée Napoléonienne.

Comment les agents pathogènes ont-ils été identifiés dans cette étude ? Pourquoi les agents pathogènes précédemment signalés n'ont-ils pas été détectés ?

Les chercheurs ont analysé l'ADN provenant des dents de 13 soldats enterrés à Vilnius (Figure 1). Chaque dent a produit environ 20 millions de courts fragments d'ADN de 31 à 150 paires de bases/lettres, appelés « lectures » (*reads*). Ces fragments pouvaient provenir de sources environnementales, de l'individu (ADN humain) et, éventuellement, d'un agent pathogène ayant infecté l'individu (s'il y en avait un). Ces fragments ont ensuite été comparés à une base de données contenant les génomes de tous les microbes connus afin de voir à quels organismes ils correspondaient. Les résultats ont ensuite été filtrés en recherchant si une correspondance figurait dans une liste de 185 bactéries connues pour provoquer des maladies chez l'homme. Pour réduire le risque de faux positifs, les chercheurs n'ont conservé que les cas où au moins 200 lectures pouvaient être attribuées de manière unique à une espèce particulière. Ce seuil s'est avéré, dans des études antérieures, être une limite fiable pour la détection de microbes pathogènes anciens, car les fragments d'ADN endommagés

taxons candidats (Tableau S1). Ces derniers ont ensuite été évalués sur la base du nombre de *k*-mers, des taux de duplication et de la couverture estimée par KrakenUniq, puis hiérarchisés selon leur pertinence épidémiologique connue et leur plausibilité au regard du contexte militaire européen du début du XIX^e siècle et de la symptomatologie historique documentée (par ex. fièvre, diarrhée, ictère).

Ces étapes ont permis d'identifier deux agents pathogènes candidats présentant des signaux cohérents : *Salmonella enterica* dans l'échantillon YYY087A (268 lectures) et *Borrelia recurrentis* dans l'échantillon YYY093A (239 lectures) (Tableaux S1–S3). Il faut noter qu'après l'ensemble des étapes de filtrage et d'authentification (voir STAR Methods), aucune lecture fiable n'a été attribuée à *R. prowazekii* (agent du typhus) ni à *B. quintana* (agent de la fièvre des tranchées), deux agents pathogènes précédemment rapportés chez des individus de ce site par des méthodes fondées sur la PCR³. Cependant, l'absence de lectures authentifiées pour ces deux espèces dans notre jeu de données ne permet pas d'exclure leur présence sur le site ou lors de cet événement historique, compte tenu notamment de la dégradation de l'ADN ancien et de la variabilité potentielle de la charge pathogène entre individus.

peuvent parfois donner des signaux trompeurs et faux si trop peu sont trouvés. En utilisant cette approche, ils ont d'abord identifié 14 pathogènes probables. Ces candidats ont ensuite été examinés de plus près. L'équipe a évalué leurs nombres de *k*-mer, une façon de mesurer combien de courts « mots » d'ADN d'une longueur fixe *k* (31 paires de bases/lettres dans ce cas) apparaissent de manière cohérente dans les données, ce qui aide à évaluer si la correspondance est réelle ou aléatoire. Ils ont également examiné les taux de duplication (pour vérifier que le signal ne provenait pas simplement de copies répétées de la même région génomique) et la couverture (la répartition des séquences d'ADN sur le génome de référence). Enfin, seuls les microbes pouvant être cohérent avec l'époque, la géographie et les sources historiques décrivant les symptômes rapportés à l'époque tels que la fièvre, la diarrhée et la jaunisse ont été conservés. Grâce à ce processus minutieux, deux bactéries ont produit des signaux cohérents : *Salmonella enterica* dans l'échantillon YYY087A (268 lectures correspondantes) et *Borrelia recurrentis* dans l'échantillon YYY093A (239 lectures correspondantes). En revanche, aucune lecture fiable et spécifique n'a été trouvée pour *Rickettsia prowazekii* (typhus) ou *Bartonella quintana* (fièvre des tranchées), les deux bactéries précédemment signalées sur ce site, mais chez des individus différents, à l'aide de méthodes basées sur la PCR (comme expliqué ci-dessus). Cependant, l'absence de signaux dans notre étude ne signifie pas que ces infections étaient absentes. L'ADN ancien est souvent très dégradé, et la quantité d'ADN pathogène peut varier considérablement d'un individu à l'autre. De plus, la PCR est une technique beaucoup plus sensible pour « capturer » un ADN spécifique, même lorsqu'il représente une proportion insignifiante de l'ADN total. Cependant, puisque le séquençage « shotgun » utilisé dans cette nouvelle étude lit au hasard tout fragment d'ADN présent dans l'échantillon, il captera toujours les séquences d'ADN les plus abondantes (souvent dominées par des bactéries environnementales). Il est donc encore possible que le typhus ou la fièvre des tranchées aient affecté certains des soldats testés dans cette étude, mais que l'ADN n'ait pas été conservé ou ait été présent en proportion inférieure au seuil de détection.

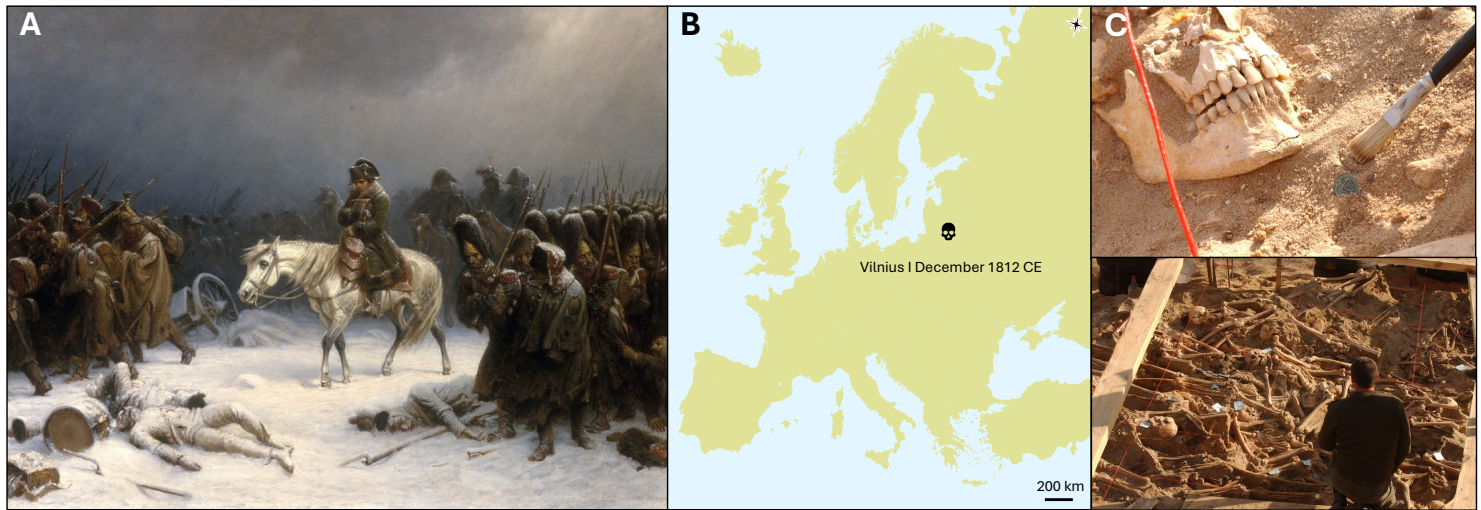


Figure 1. Contexte historique, géographique et archéologique de l'étude.

(A) Peinture datée de 1851 intitulée « *Napoleon's retreat from Moscow* » par Adolph Northen, représentant les conditions de retraite de l'armée napoléonienne. (B) Carte géographique de l'Europe indiquant la localisation et la datation du site archéologique de Vilnius, en Lituanie, d'où proviennent les échantillons étudiés. (C) Photographies *in situ* prises lors de la fouille des tranchées contenant les corps des soldats napoléoniens. La photographie du haut montre la découverte d'un bouton d'uniforme de type impérial dans la fosse commune. Celle du bas présente une vue d'ensemble de la fosse commune (crédit photo : Michel Signoli).

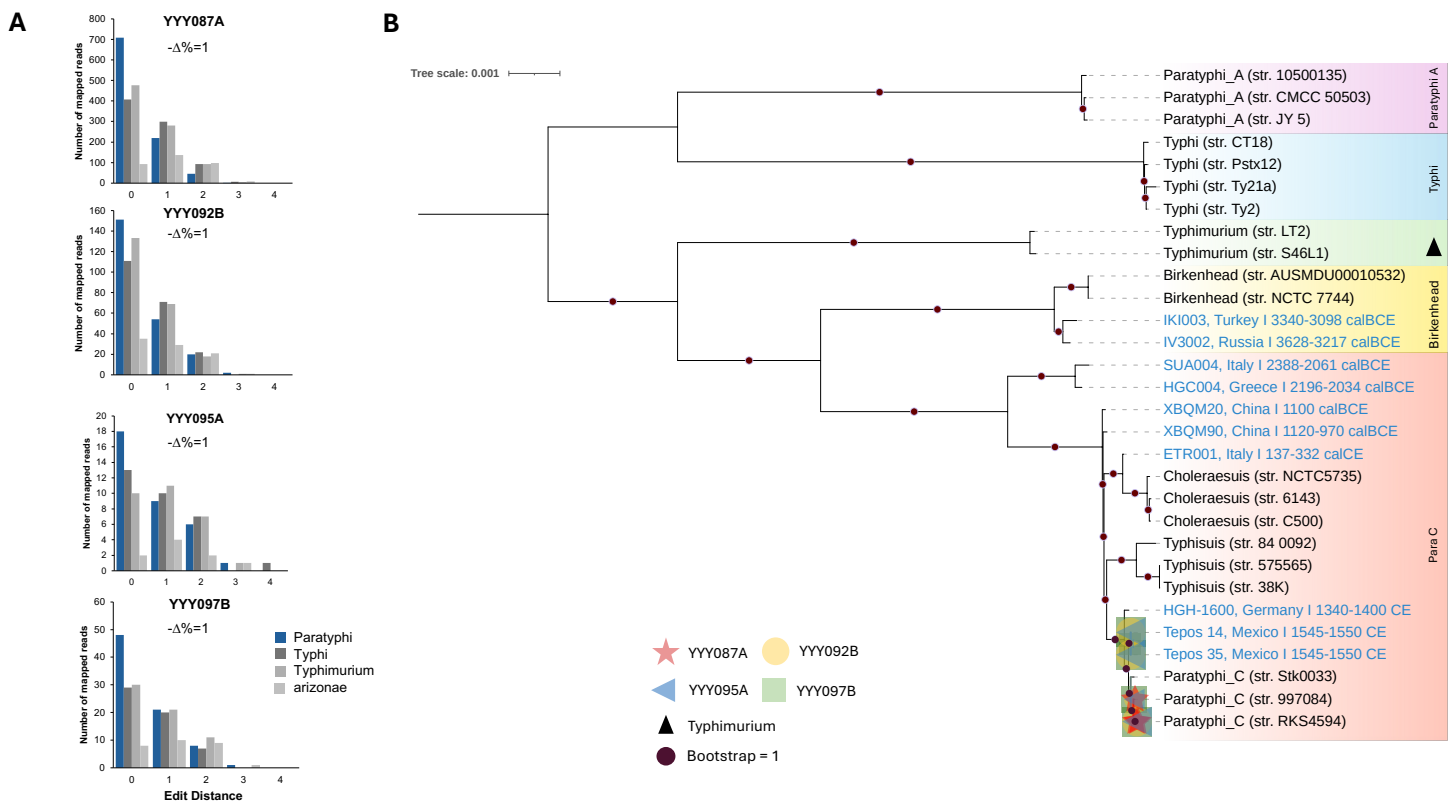


Figure 2. Authentification des données d'ADN ancien de *S. enterica*. (A) Distribution des distances d'édition pour les lectures des échantillons YYY087A, YYY092B, YYY095A et YYY097B, accompagnée de la proportion de différence négative (-Δ%)¹⁷. (B) Phylogénie en vraisemblance maximale de 31 génomes de *S. enterica* représentatifs de la diversité connue de l'espèce, calculée sur 4 083 601 positions alignées du génome (dont 91 636 sites variables). L'échantillon YYY087A est représenté par des étoiles rouges, YYY092B par des cercles jaunes, YYY095A par des triangles bleus et YYY097B par des carrés verts. Les positions phylogénétiques de ces échantillons ont été déterminées à l'aide du logiciel *epa-ng*³⁰ et reposent sur les séquences partielles du génome reconstruites à partir des lectures attribuées à l'espèce *S. enterica* par *MEGAN* (voir *STAR Methods*). Les génomes anciens publiés précédemment apparaissent en bleu, et les génomes modernes en noir. Les lignées sont codées par couleur comme suit : le violet représente la lignée *paratyphi A*, le bleu correspond à la lignée *typhi*, le vert à la lignée *typhimurium*, le jaune à la lignée *Birkenhead*, et le rouge à la lignée *para C*. Les points violets indiquent une valeur de bootstrap égale à 1. Les symboles de bootstrap sont affichés sur les branches menant à chaque nœud soutenu, selon les paramètres par défaut d'iTOL. La présence de plusieurs symboles indique les différents placements possibles d'un même échantillon sur l'arbre phylogénétique. La position de chaque symbole correspond à un placement probable, la taille variant en fonction du *like weight ratio* et des valeurs de *distal length*. Voir également les Figures S1, S2, S4 et le Tableau S4.

Figure 2. A) Ces fragments proviennent-ils réellement d'une souche ancienne de *Salmonella* ? Chaque graphique montre à quel point les fragments d'ADN de quatre soldats correspondent à *Salmonella enterica*. Sur l'axe des abscisses figure le nombre de différences (« edit distance ») entre chaque fragment et le génome de référence. Pour un ADN ancien authentique, on s'attend à une courbe descendante : beaucoup de correspondances parfaites ou quasi parfaites, puis de moins en moins de mauvaises correspondances. La valeur notée $-A\%$ résume cette pente : des valeurs proches de 1 indiquent un profil typique d'ADN ancien ; des valeurs faibles suggèrent du bruit ou une contamination. B) **Où se situent ces souches dans l'arbre évolutif de *Salmonella* ?** Cet arbre regroupe les génomes modernes (en noir) et anciens publiés précédemment (en bleu) de *S. enterica*. Les bandes colorées indiquent les principales lignées (par exemple, rouge = Para C). Les quatre échantillons napoléoniens sont représentés par différents symboles : étoile rouge ★, cercle jaune ●, triangle bleu ▲, carré vert ■. Bien que seules de petites portions de leurs génomes aient été récupérées, une méthode appelée **placement phylogénétique** permet de les positionner dans l'arbre en se basant sur les sites génomiques effectivement couverts. Leurs symboles se situent à l'intérieur du groupe **Para C (Paratyphi C)**. Les points violets sur les branches indiquent un soutien bootstrap maximal (= 1), ce qui garantit la fiabilité de la structure de l'arbre dans lequel les séquences ont été placées. Lorsqu'un même symbole apparaît à plusieurs endroits proches, cela reflète des placements alternatifs mais similaires ; les symboles plus grands indiquent les positions les plus probables. **Message clé :** Le panneau A montre que l'ADN se comporte comme un ADN ancien authentique de *Salmonella enterica*, génétiquement plus proche de *Paratyphi C* que des autres sérovars de *S. enterica*. Le panneau B montre que, malgré la faible couverture génomique, les fragments récupérés placent de manière cohérente les infections des soldats au sein de la lignée *Paratyphi C*.

Pour évaluer la présence de *S. enterica*, nous avons examiné à la fois le nombre de lectures et leurs profils d'alignement entre différents sérovars. Chez quatre individus (YYY087A, YYY092B, YYY095A et YYY097B), entre 34 et 968 lectures uniques (c'est-à-dire après suppression des doublons) se sont alignées majoritairement et en plus grand nombre sur le sérovar *Paratyphi C* (souche RKS4594), plutôt que sur *Typhi* ou *Typhimurium* (Tableau 1, Figure 2A). Les distributions correspondantes des distances d'édition présentaient le profil décroissant attendu, cohérent avec un signal authentique d'agent pathogène ancien, ainsi qu'une proportion de différence négative (PDN) de 1,0 (Tableau S2)¹⁷. À l'inverse, les autres échantillons présentant un faible nombre de lectures alignées sur *S. enterica* affichaient des profils de distances d'édition plus plats ou irréguliers, des valeurs de PDN faibles, et ont été considérés comme des résultats négatifs, probablement liés à des espèces environnementales ou à des régions génomiques conservées présentant un faible pouvoir discriminant entre taxons (Tableau S2). Une analyse similaire a été réalisée pour les espèces de *Borrelia*. Chez l'individu YYY093A, 4 062 lectures uniques se sont alignées sur *B. recurrentis*, contre 1 556 et 1 441 lectures alignées respectivement sur *B. duttonii* et *B. crocidurae* (Tableau S1, Figure 3A). Les lectures s'alignaient sur l'ensemble des huit fragments du génome de *B. recurrentis* (le chromosome et sept plasmides), et les profils de distance d'édition présentaient à nouveau un schéma décroissant cohérent, avec une PDN = 1 (Figure 3A, Tableau S3)¹⁷. Un second individu (YYY092B), également positif pour *S. enterica*, présentait 322 lectures uniques alignées sur *B. recurrentis*, avec une distribution des distances d'édition également compatible avec un signal authentique (PDN = 1 dans le Tableau S3, Figure 3A). Tous les profils de distance d'édition et les analyses subséquentes ont été réalisés à partir de lectures dédoublées et de haute qualité d'alignement (MAPQ > 30).

Stratégie utilisée pour définir les spécificités de *Salmonella enterica* et de *Borrelia recurrentis*.

Pour confirmer la présence de *Salmonella enterica*, les chercheurs ont examiné non seulement le nombre de fragments d'ADN correspondant à cette bactérie, mais aussi la façon dont ces fragments s'alignaient avec différents sous-groupes, appelés sérovars. Un sérovar est comme une « branche familiale » au sein d'une espèce bactérienne, différenciée par de petites variations dans ses molécules de surface. Ces différences peuvent influencer la manière dont la bactérie se propage ou la sévérité de la maladie qu'elle provoque. Par exemple, *Salmonella enterica* possède de nombreux sérovars, dont *Typhi* (qui cause la fièvre typhoïde), *Typhimurium* (qui provoque généralement des intoxications alimentaires) et *Paratyphi C* (qui cause la fièvre paratyphoïde). Chez quatre individus (échantillons YYY087A, YYY092B, YYY095A et YYY097B), entre 34 et 968 fragments d'ADN uniques (après suppression des doublons) étaient plus proches du sérovar *Paratyphi C* (souche RKS4594), que *Typhi* ou *Typhimurium* (Tableau 1 en bas du fichier, Figure 2A). Pour vérifier ces correspondances, l'équipe a examiné la distribution des distances d'édition. La distance d'édition mesure le nombre de « fautes d'orthographe » ou de désaccords entre un fragment d'ADN ancien et un génome de référence moderne. Avec de l'ADN ancien authentique, la plupart des fragments présentent une correspondance proche, ce qui produit une courbe descendante régulière (c'est-à-dire, la plupart sans aucune discordance, suivies de beaucoup moins avec une seule discordance, et ainsi de suite). C'est exactement ce qu'ils ont observé dans ces cas (dans la Figure 2A ci-dessus). Ils ont également utilisé un test statistique appelé proportion de différence négative (*Negative Difference Proportion*, NDP), qui vérifie si la courbe suit réellement le schéma attendu. Une NDP de 1,0 signifie que la correspondance est totalement cohérente avec une infection ancienne authentique. En revanche, d'autres échantillons contenant seulement quelques lectures correspondant à *S. enterica* ont montré des courbes de distance d'édition plates ou irrégulières et affichaient de faibles valeurs de NDP. Ces signaux ont été considérés comme des faux positifs, probablement issus de bactéries environnementales du sol ou provenant de régions d'ADN commune à de nombreux micro-organismes. Le même type d'analyse a ensuite été effectué pour les espèces de *Borrelia*. Chez un soldat (échantillon YYY093A), 4 062 fragments uniques d'ADN correspondaient à *Borrelia recurrentis*, contre 1 556 pour *B. duttonii* et 1 441 pour *B. crocidurae* (Figure 3A). Là encore, la courbe de distance d'édition a diminué comme prévu, confirmant que l'infection était authentique (voir Figure 3A ci-dessous). Un deuxième soldat (échantillon YYY092B), qui avait également été testé positif pour *S. enterica*, présentait 322 lectures uniques correspondant à *B. recurrentis*, avec le même schéma.

Compte tenu du faible nombre de lectures, la couverture génomique variait de 0,0003 à 0,009 pour *S. enterica* et de 0,002 à 0,14 pour *B. recurrentis*. La distribution de la longueur des lectures, avec une moyenne de 56 pb pour *B. recurrentis* et de 43 pb pour *S. enterica*, indiquait une fragmentation importante de l'ADN, conforme aux profils de dégradation caractéristiques de l'ADN ancien (aDNA)¹⁸. L'analyse des altérations post-mortem dans les 4 062 lectures de l'échantillon YYY093A alignées sur le génome de référence moderne de *B. recurrentis* (souche A1) a révélé qu'environ 5 % présentaient une désamination des cytosines aux extrémités des lectures (Figure S1C). Ce taux est légèrement inférieur aux ~8 % observés dans l'ADN humain provenant du même échantillon (Figure S1D), mais il reste dans la même plage que celle des autres individus du site (Figures S1E–H) ou des échantillons d'ADN ancien datés de la même période¹⁹, ce qui est donc cohérent avec la présence d'un ADN bactérien authentique. Il est à noter que des taux de désamination plus faibles pour l'ADN bactérien que pour l'ADN humain provenant d'un même reste ont déjà été signalés, suggérant que l'ADN bactérien peut se dégrader à un rythme différent, souvent plus lentement, que l'ADN humain^{20–22}. Malheureusement, pour *S. enterica*, le nombre de lectures alignées était trop faible pour permettre une estimation fiable des profils de désamination (Tableau 1), plusieurs centaines de lectures étant généralement nécessaires pour des évaluations précises.

Pour authentifier davantage les lectures attribuées aux deux agents pathogènes, nous avons réanalysé les lectures déduplicées alignées en les comparant à la base de données complète NCBI nt à l'aide de BLASTN, suivie d'une classification taxonomique selon l'algorithme LCA implémenté dans MEGAN²³ (Figures S1A–B ; Tableaux S2 et S3). Cette analyse a confirmé les attributions initiales, une large proportion des lectures étant bien assignée aux agents pathogènes attendus, tout en éliminant celles qui ne pouvaient pas être attribuées de manière non ambiguë aux taxons correspondants. Ces résultats ont renforcé l'identification de quatre échantillons positifs pour *S. enterica* (YYY087A, YYY092B, YYY095A, YYY097B) et d'un échantillon (YYY093A), voire possiblement deux (YYY092B, avec seulement 18 lectures spécifiques sur 322 lectures alignées), pour *B. recurrentis* (Tableaux S2 et S3).

Première étape d'authentification : rechercher des signes de dégradation de l'ADN.

Comme seulement un nombre relativement faible de lectures a été récupéré, la couverture génomique globale était très faible : entre 0,0003 et 0,009 pour *Salmonella enterica*, et entre 0,002 et 0,14 pour *Borrelia recurrentis*. La couverture génomique correspond à la partie du génome total d'un organisme qui a pu être reconstituée à partir des fragments d'ADN retrouvés. Ces valeurs signifient que seules de minuscules portions des génomes bactériens ont pu être reconstruites.

La longueur des fragments d'ADN a également montré à quel point le matériel était dégradé. En moyenne, les fragments de *B. recurrentis* mesuraient 56 paires de bases, et ceux de *S. enterica* 43 paires de bases en moyenne. Ces longueurs très courtes sont conformes à la fragmentation typique de l'ADN ancien. Les chercheurs ont aussi recherché un type de dommage post-mortem de l'ADN appelé désamination de la cytosine, considéré comme une marque caractéristique de l'ADN ancien. Ce processus se produit lorsqu'une des lettres de l'ADN, la cytosine (C), subit une modification chimique après la mort (se transformant en U, lequel est ensuite converti en T lors de la préparation de la bibliothèque de séquençage), et il est le plus souvent visible aux extrémités des fragments d'ADN.

Dans les 4 062 lectures de *B. recurrentis* provenant de l'échantillon YYY093A, environ 5 % présentaient ce type de dommage. À titre de comparaison, l'ADN humain du même échantillon montrait environ 8 % de dommage. Ces deux valeurs se situent dans la plage attendue pour des restes archéologiques de cet âge, et sont donc compatibles avec un ADN ancien authentique. Des études ont également observé que l'ADN bactérien présente souvent moins de désamination que l'ADN humain provenant du même squelette, ce qui signifie que l'ADN bactérien peut se dégrader à un rythme légèrement différent. Pour *S. enterica*, cependant, il y avait trop peu de fragments d'ADN (Table 1) pour évaluer la proportion de lectures arborant une désamination de manière fiable. Ce type d'analyse nécessite généralement plusieurs milliers de fragments, et ici, les nombres étaient en dessous de ce seuil.

Deuxième étape d'authentification : s'assurer que toutes les séquences s'alignant avec les génomes de référence proviennent réellement de l'espèce présumée.

Pour confirmer de manière plus précise que les fragments d'ADN alignés avec les génomes de références des deux espèces suspectées appartenaient réellement à ces pathogènes, les chercheurs ont ré-analysé les lectures déduplicées en les comparant à la base de données complète de nucléotides du NCBI (une base de données qui contient les génomes de la plupart des espèces séquencées à ce jour) à l'aide de BLASTN, un programme qui compare des séquences d'ADN à tout le matériel génétique connu dans la database. Ils ont ensuite appliqué un algorithme du plus bas ancêtre commun (LCA) dans le programme MEGAN, lequel évalue toutes les espèces/génomes correspondants à chaque lecture, puis attribue chaque fragment au niveau taxonomique le plus spécifique possible (espèce, genre, famille, etc.), ce qui permet d'éliminer les correspondances incertaines ou erronées. Cette seconde analyse a confirmé les résultats précédents : la plupart des fragments correspondaient aux agents pathogènes attendus, tandis que les fragments ambigus ont été exclus. Les résultats ont renforcé

l'identification de quatre cas positifs de *Salmonella enterica* (échantillons YYY087A, YYY092B, YYY095A, YYY097B) et d'un cas de *Borrelia recurrentis* (échantillon YYY093A), avec un second cas possible chez YYY092B, bien que, chez cet individu, seulement 18 des 322 fragments cartographiés étaient suffisamment spécifiques pour être considérés comme des correspondances véritables.

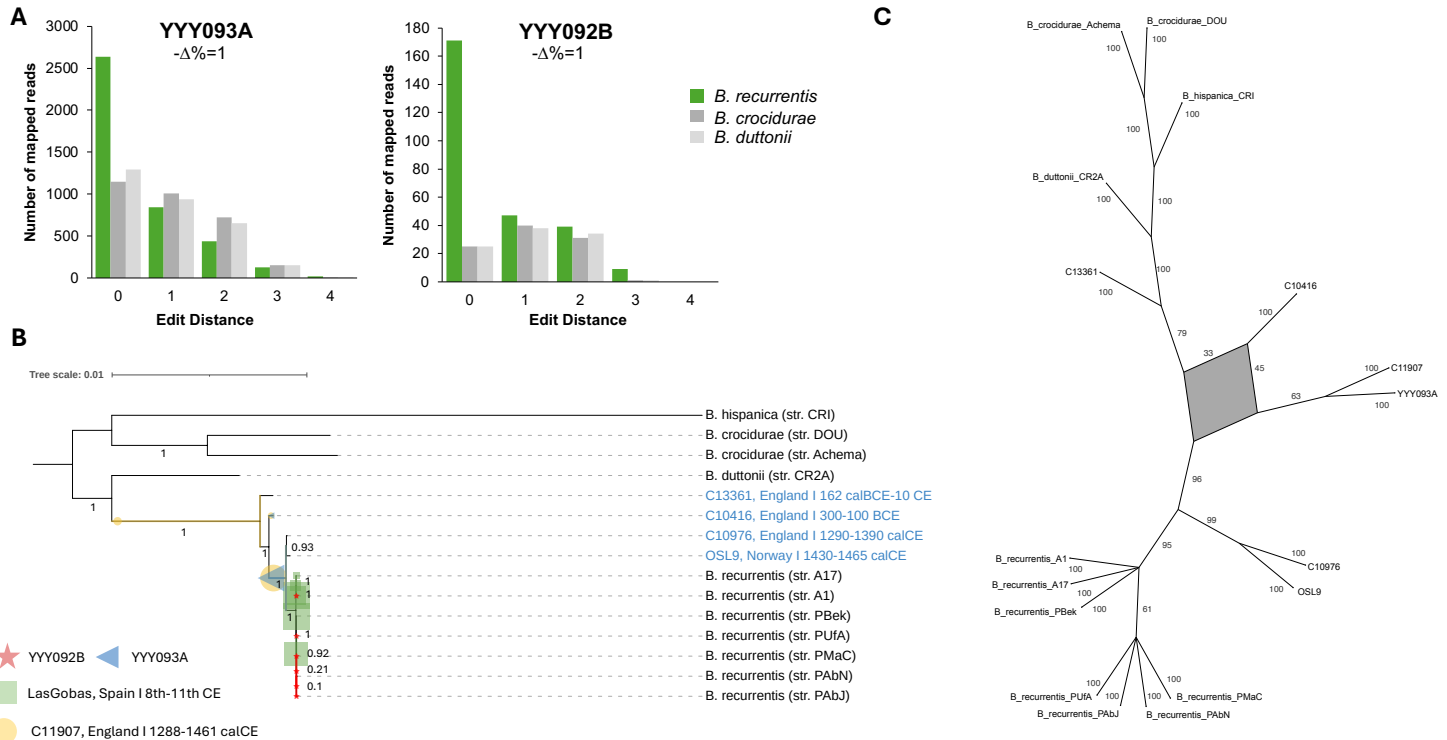


Figure 3. Authentification des données d'ADN ancien de *B. recurrentis*. (A) Distribution des distances d'édition des lectures alignées des échantillons YYY093A et YYY092B sur trois espèces de *Borrelia* étroitement apparentées, accompagnée de la proportion de différence négative (-Δ%)¹⁷. (B) Phylogénie en vraisemblance maximale de 15 génomes de *Borrelia* spp., calculée sur 1 032 378 positions alignées du génome (dont 39 217 sites variables). Le placement phylogénétique, réalisé avec *epa-ng*³⁰, repose sur les lectures classées dans le genre *Borrelia* par *MEGAN*. Les échantillons YYY093A et YYY092B sont représentés respectivement par un triangle bleu et une étoile rouge. En raison de leur faible couverture, les deux génomes anciens précédemment publiés, C11907 et Las Gobas, ont été intégrés à ce graphique par placement phylogénétique, de la même manière que les échantillons de cette étude, et sont indiqués par un cercle jaune et un carré vert. La taille des symboles est déterminée par le rapport de poids de vraisemblance et les valeurs de distance distale. Le génome ancien précédemment publié est indiqué en bleu, et les génomes modernes en noir. Les valeurs numériques correspondent aux soutiens des branches (valeurs de bootstrap), affichées sur les branches menant à chaque nœud soutenu selon les paramètres par défaut d'ITOL. (C) Réseau de consensus construit à l'aide de *SplitsTree*³⁹ à partir de 100 arbres bootstrappés issus d'un alignement multiple (MSA) de 144 470 sites couverts par les lectures spécifiques de YYY093A (5 854 sites variables), afin d'évaluer la nature de l'incertitude topologique du placement de YYY093A (Figure S3). Le réseau montre que, bien que les valeurs de bootstrap soient faibles pour le clade YYY093A-C11907 (Figures S3E-F), ces deux souches, ainsi que C10416, sont systématiquement regroupées dans un cluster distinct à travers les 100 arbres bootstrappés. Voir également les Figures S1, S4 et le Tableau S4.

Figure 3. A) Ces fragments proviennent-ils réellement d'une souche ancienne de *Borrelia* (et de quelle espèce) ? Pour deux soldats (YYY093A et YYY092B), les graphiques montrent à quel point leurs fragments d'ADN correspondent à trois espèces de *Borrelia* étroitement apparentées. Sur l'axe des abscisses figure le nombre de différences (« edit distance ») par rapport à chaque génome de référence. Pour un ADN ancien authentique, on s'attend à une courbe descendante : de nombreuses correspondances parfaites ou presque parfaites, puis de moins en moins de correspondances médiocres. Le nombre résumé -Δ% indique la régularité de cette pente : des valeurs proches de 1 confirment un signal d'ADN ancien authentique plutôt qu'un bruit aléatoire ou une contamination. **B) Où se situent ces souches dans l'arbre évolutif des *Borrelia* ?** Cet arbre inclut des génomes de *Borrelia* modernes (étiquettes noires) et anciens publiés précédemment (étiquettes bleues). Parce que seules de petites portions d'ADN pathogène aient été récupérées, une méthode appelée placement phylogénétique permet de positionner les fragments génomiques issus de chaque échantillon sur l'arbre évolutif de l'espèce ou du genre, en se basant sur les sites réellement couverts. Les deux échantillons napoléoniens sont indiqués (YYY093A = triangle bleu; YYY092B = étoile rouge). Deux génomes anciens publiés (cercle jaune et carré vert) sont représentés de la même manière à titre de comparaison. Lorsqu'un même symbole apparaît à plusieurs endroits proches, cela reflète des placements alternatifs mais similaires ; les symboles plus grands correspondent aux positions les plus probables. Les nombres figurant sur les branches représentent les soutiens bootstrap : plus ils sont élevés, plus la séparation entre les branches est fiable. **C) Quelle est la fiabilité du placement de YYY093A ?** Comme YYY093A couvre un plus grand nombre de sites, les auteurs ont également construit un réseau de consensus à partir de 100 arbres bootstrappés, en utilisant uniquement les positions génomiques couvertes par cet échantillon, même pour les souches disposant de génomes entiers (afin de comparer sur la même base d'information). Ce réseau montre que, bien que certaines branches internes présentent une incertitude, YYY093A se regroupe systématiquement avec le même ensemble de génomes anciens de *B. recurrentis* dans tous les arbres bootstrappés, indiquant un voisinage évolutif stable, proche des souches de l'âge du Fer et du Moyen Âge en Angleterre.

Message clé : Le panneau A montre que l'ADN se comporte comme un ADN ancien authentique de *Borrelia*. Le panneau B situe les infections des soldats dans la lignée de *B. recurrentis*, malgré une faible couverture génomique. Le panneau C confirme que le placement plus profond de YYY093A est solidement soutenu, même lorsque l'incertitude phylogénétique est explicitement prise en compte.

Comme dernière étape d'authentification, nous avons analysé les lectures spécifiques identifiées par **BLASTN** et **MEGAN** (B+M) à l'aide d'une stratégie de placement phylogénétique précédemment appliquée à des génomes anciens de faible couverture^{24, 25}. Afin d'établir un cadre de référence robuste, nous avons d'abord reconstruit des phylogénies génomiques complètes de *S. enterica* spp. et *B. recurrentis*, incluant à la fois des souches anciennes et modernes représentatives de la diversité connue de chaque espèce^{26–29}. Les méthodologies employées étaient similaires à celles des études originales et suivaient les pratiques standards en génomique des agents pathogènes anciens (voir *STAR Methods*), reproduisant avec succès les topologies d'arbres précédemment publiées avec un fort soutien bootstrap. Nous avons ensuite appliqué **epa-ng**³⁰ pour évaluer le placement le plus probable des souches anciennes, en se basant sur les positions couvertes par les lectures spécifiques B+M dans l'alignement multiple de séquences (MSA) utilisé pour construire les arbres de référence (Tableaux 1 et S4). **EPA-ng** place les génomes de faible couverture sur une phylogénie et un alignement multiple de séquences (MSA) fixes, en calculant la position la plus probable pour chaque échantillon selon les sites génomiques couverts. Cette approche est particulièrement adaptée à l'ADN ancien, dont seule une fraction réduite du génome est généralement récupérée, et permet souvent une attribution de lignée fiable même en cas de couverture limitée.

Pour *B. recurrentis*, les lectures spécifiques (on-target) de l'échantillon YYY093A se sont alignées sur 125 514 positions de l'alignement multiple de séquences (MSA), dont 5 854 sites variables, tandis que celles de l'échantillon YYY092B se sont alignées sur 584 positions couvrant 21 sites variables, sur un total de 1 032 378 positions dans le MSA. Pour *S. enterica*, les lectures spécifiques issues des échantillons YYY087A, YYY092B, YYY095A et YYY097B se sont alignées respectivement sur 26 116, 4 864, 1 236 et 1 382 positions, couvrant 800, 140, 44 et 42 sites variables sur une longueur totale du MSA de 4 083 601 positions (voir Tableau 1). Notre phylogénie finale de *S. enterica* comprenait 10 génomes anciens et 20 génomes modernes publiés, représentatifs de la diversité connue de l'espèce, au sein de laquelle nous avons pu placer les séquences des quatre individus positifs de Vilnius (YYY087A, YYY092B, YYY095A, YYY097B) dans le sérovar Paratyphi C, un clade qui inclut également trois génomes anciens datés entre le XIV^e et le XVI^e siècle^{31–36} (Figures 2B et S2). Dans l'ensemble, les placements phylogénétiques et les génotypes observés aux positions variables (Figures S4A–D) renforcent nettement nos estimations initiales de sérovar. Cependant, malgré le placement cohérent de toutes les souches au sein du même sérovar, la résolution de nos données n'était pas suffisante pour déterminer une affiliation plus précise à une branche ancienne ou moderne spécifique de Paratyphi C.

Troisième et dernière étape d'authentification : positionnement phylogénétique des fragments de génome récupérés.

Comme dernière étape d'authentification, les chercheurs ont utilisé une méthode appelée positionnement phylogénétique, qui permet de déterminer où un échantillon d'ADN se situe sur « l'arbre généalogique » évolutif d'une espèce (appelé arbre phylogénétique). Ils ont d'abord construit des arbres de référence pour *Salmonella enterica* et *Borrelia recurrentis*, en utilisant des génomes complets de souches modernes et anciennes afin de représenter toute la diversité connue de chaque bactérie. Ces arbres ont reproduit les mêmes schémas de ramification observés dans des études antérieures et ont été confirmés par des tests statistiques, prouvant que leurs structures étaient robustes. Les fragments d'ADN ancien de chaque soldat Napoléoniens et de chaque espèce ont ensuite été analysés avec un programme appelé epa-ng. Cet outil prend les fragments d'ADN courts et incomplets et calcule où ils s'insèrent le mieux dans l'arbre de référence en se basant sur les quelques positions en commun avec les génomes de référence. En d'autres termes, même si seule une petite partie d'un génome pathogène a survécu, epa-ng peut le placer à l'endroit le plus probable, un peu comme si l'on déterminait où quelques pièces de puzzle s'emboîtent le mieux dans une image déjà complétée. Cela en fait un outil particulièrement utile pour l'ADN ancien où une assignation fiable à une lignée reste possible même quand il ne reste que quelques fragments du pathogène.

Pour *Borrelia recurrentis*, un soldat (l'échantillon YYY093A) possédait des fragments d'ADN correspondant à 125 514 positions du génome, dont 5 854 positions où les souches diffèrent les unes des autres. Ces positions variantes servent de base pour construire des arbres, ou pour déterminer à quel endroit d'un arbre une séquence appartient. Un autre soldat (YYY092B) présentait un signal beaucoup plus faible, avec seulement 584 positions et 21 sites variables correspondants, sur un total de plus d'un million de positions possibles dans l'alignement de référence. Pour *Salmonella enterica*, les quatre soldats positifs présentaient une couverture globale plus faible. Un soldat (YYY087A) présentait 26 116 positions avec 800 variantes, un autre (YYY092B) 4 864 positions avec 140 variantes, tandis que deux autres (YYY095A et YYY097B) présentaient à un peu plus de 1 200 et 1 300 positions, avec respectivement 44 et 42 variantes. Ce sont des chiffres très faibles comparés à l'ensemble du génome de référence, qui comprend plus de 4 millions de positions. Lorsque les chercheurs ont placé ces résultats sur un « arbre généalogique » de *S. enterica* construit à partir de 10 génomes anciens et 20 génomes modernes, les quatre échantillons Napoléoniens se sont tous regroupés dans le groupe Paratyphi C, la même branche qui contient trois génomes anciens datant du XIV^e au XVI^e siècles. Leur position sur l'arbre et les différences génétiques observées ont fourni de solides preuves que les infections probables étaient causées par Paratyphi C. Cependant, comme les fragments d'ADN étaient très limités, les chercheurs n'ont pas pu déterminer si ces

souches étaient plus proches des branches anciennes ou modernes au sein de *Paratyphi C*.

Une souche de *Borrelia recurrentis* appartenant à une lignée ayant persisté pendant des millénaires en Europe.

Pour *Borrelia recurrentis*, l'équipe a construit un arbre évolutif de référence à partir de quatre génomes anciens et onze modernes. La souche provenant du soldat Napoléonien YYY093A a été placée avec une forte confiance (LWR = 0,89) dans une position basale, ce qui signifie qu'elle s'est séparée plus tôt que les souches médiévales mieux connues des XIV^e–XV^e siècles (C10976 et OSL9) ainsi que des de toutes les souches modernes de *B. recurrentis* (Figure 3B). Cela indique que YYY093A appartenait à une lignée plus ancestrale que celle regroupant ces souches médiévales. Comme il existe très peu de génomes de cette espèce, et afin de mieux comprendre sa diversité, les chercheurs ont réanalysé d'autres données d'ADN ancien publiées qui n'avaient pas été étudiées selon une approche de placement phylogénétique, en appliquant la même méthode que celle utilisée pour les souches de cette étude. L'un d'eux, connu sous le nom de C11907 et découvert en Angleterre médiévale (1288–1461 CE), avait environ un tiers de son génome conservé. Lui aussi a été placé dans une position basale par rapport aux souches médiévales plus tardives (LWR = 0,81), confirmant des résultats antérieurs qui l'avaient relié à un génome de l'âge du fer encore plus ancien (C10416, 300–100 av. n. è.). Ce qui est particulièrement frappant, c'est que la souche Napoléonienne YYY093A et la souche anglaise C11907 se sont retrouvées presque au même endroit dans l'arbre. Lorsqu'elles ont été analysées ensemble, YYY093A s'est regroupée à la base de la branche de C11907 avec un score de confiance élevé (LWR = 0,99). Cette proximité suggère que les deux souches provenaient d'une même lignée ancienne de *B. recurrentis*, une lignée distincte des souches plus récentes qui circulent encore aujourd'hui.

Pour approfondir l'analyse du positionnement phylogénétique de YYY093A, nous avons construit un arbre phylogénétique *de novo* restreint aux 144 470 positions génomiques couvertes par les lectures classées dans le genre *Borrelia* par B+M (Figures S3E–F). Dans cette analyse, YYY093A et C11907 formaient un clade monophylétique beaucoup plus proche du génome de l'âge du Fer C10416 que de C10976 ou OSL9. Étant donné que le soutien bootstrap pour cette topologie était relativement faible (58 %), nous avons cherché à identifier la source de cette incertitude à l'aide de **SplitsTree**³⁷. En visualisant les 100 arbres bootstrappés sous forme de réseau de consensus, nous avons observé que l'ambiguïté topologique était limitée au sous-arbre englobant C10416, C11907 et YYY093A, plutôt que répartie sur l'ensemble de la phylogénie (Figure 3C). Ce résultat suggère que YYY093A et C11907 se situent de manière constante dans le même voisinage évolutif que la souche de l'âge du Fer, ce qui soutient l'hypothèse selon laquelle ces génomes représentent une lignée distincte et relativement ancienne, ou un groupe différencié de *B. recurrentis*, ayant persisté en Europe pendant au moins deux millénaires et demeuré en circulation jusqu'au début du XIX^e siècle.

Pour explorer plus en détail la position de la souche Napoléonienne YYY093A, les chercheurs ont construit un nouvel arbre phylogénétique en utilisant uniquement les 144 470 sites du génome de référence couvert par des lectures assignées avec certitude au genre *Borrelia*. Dans cette analyse, YYY093A et la souche anglaise médiévale C11907 se sont regroupées dans un clade monophylétique sur une branche commune, située beaucoup plus près de la souche de l'âge du fer C10416 que des souches médiévales ultérieures C10976 ou OSL9. Le soutien statistique à ce placement, appelé *bootstrap support*, était de 58%, soit relativement faible. Cela signifie que l'arbre ne pouvait pas être considéré comme entièrement stable dans la zone où ces souches étaient positionnées.. Pour comprendre ce manque de stabilité, les chercheurs ont utilisé un programme appelé SplitsTree pour visualiser 100 arbres avec des tests répétés. Cela a montré que l'incertitude régissant l'organisation de l'arbre était limitée à une autre section de l'arbre contenant C10416, C11907 et YYY093A, tandis que le reste de l'arbre restait stable (Figure 3C). Le regroupement constant de YYY093A avec C11907, toutes deux proches du génome de l'âge du fer, suggère que ces souches appartenaient à une lignée distincte et très ancienne de *B. recurrentis*. Cette lignée semble avoir persisté en Europe pendant au moins deux mille ans, continuant à circuler depuis l'âge du fer jusqu'au début du XIX^e siècle.

En contraste, deux souches à plus faible couverture, YYY092B (cette étude) et Las Gobas (VIII^e–XI^e siècle EC, Espagne²⁸), ont été placées au sein du clade regroupant les génomes modernes presque identiques de *B. recurrentis* (Figure 3B, Figures S3A, B, D, S4G). Alors que le placement du génome de Las Gobas semble bien étayé (50 657 positions alignées ; 2 357 sites variables, Tableau S4), YYY092B ne s'aligne que sur 584 positions couvrant 21 variants (Tableau 1), ce qui limite la résolution du placement. Cependant, la cohérence du positionnement entre les deux souches et le signal relativement robuste provenant de Las Gobas suggèrent que ces génomes pourraient effectivement appartenir au clade regroupant les génomes modernes disponibles. Il est important de noter que, bien que la faible couverture de YYY092B ne permette pas une attribution définitive à une sous-lignée, elle soutient néanmoins sa classification au niveau de l'espèce. Dans l'ensemble, nos résultats appuient l'existence de plusieurs lignées distinctes de *B. recurrentis* ayant circulé en Europe au cours des derniers siècles.

Les travaux récents en ADN ancien ont révélé que les fièvres paratyphoïdes étaient présentes en Europe depuis plusieurs millénaires^{26, 31–36}, et elles étaient déjà bien connues et documentées en 1812. La maladie se transmet à l'être humain par la consommation d'aliments ou d'eau contaminés par des matières fécales infectées. Les symptômes comprennent de la fièvre, des maux de tête, une éruption cutanée, une faiblesse générale, une perte d'appétit, de la diarrhée, de la constipation, des douleurs abdominales et des vomissements³⁸. Au cours de la campagne de Russie de Napoléon, ni la fièvre paratyphoïde ni la fièvre typhoïde ne sont mentionnées dans les sources historiques connues, probablement en raison de la nature non spécifique et variée de leurs symptômes. Cependant, un rapport de 1812 rédigé par J. R. L. de Kirckhoff, médecin au sein de l'armée napoléonienne, contient des informations essentielles sur les événements pouvant expliquer l'origine d'une épidémie. Dans ce document, il précise que les soldats souffraient de typhus, de dysenterie et de diarrhée à leur arrivée à Vilnius. En insistant sur ce dernier point, il écrit : « La diarrhée était commune parmi nous en Lituanie. Une cause puissante de cette maladie provenait du fait que nous rencontrions dans presque chaque maison, d'Orcha à Wilna, de grands tonneaux de betteraves salées (*buraki kwaszone*), que nous mangions et dont nous buvions le jus lorsque nous avions soif, ce qui nous incommodait fortement et irritait vivement le tube intestinal »⁵. Cette description pourrait correspondre aux caractéristiques d'une infection par fièvre paratyphoïde transmise par des aliments contaminés, ainsi qu'aux symptômes digestifs typiquement associés à cette maladie, bien que nous reconnaissons qu'ils pourraient également correspondre à d'autres pathologies fréquentes en Europe au XIX^e siècle. De plus, même aujourd'hui, deux siècles plus tard, il serait encore impossible d'établir un diagnostic différentiel entre le typhus, la fièvre typhoïde ou la

Contrairement à la lignée plus ancienne représentée par YYY093A, deux autres souches se situaient beaucoup plus près des formes modernes de *Borrelia recurrentis*. L'une provenait d'un autre échantillon de soldat Napoléonien, YYY092B, et l'autre du site de Las Gobas en Espagne, datée des VIII^e–XI^e siècles CE. Toutes deux ont été placées dans le clade contenant les génomes de *B. recurrentis* actuels. La souche de Las Gobas était bien soutenue, avec plus de 50 000 positions du génome couvertes et plus de 2 000 points de variation identifiés. YYY092B, en revanche, présentait beaucoup moins de correspondances : seulement 584 positions et 21 variantes, ce qui limitait la confiance dans son placement phylogénétique. Néanmoins, le fait que YYY092B et Las Gobas s'alignent systématiquement avec la branche de type moderne suggère qu'elles appartenaient à la même lignée que les souches actuelles de *B. recurrentis*, bien que cela ne soit pour l'instant qu'une hypothèse. Conceptuellement, cela signifie qu'au Moyen Âge, et jusqu'au début du XIX^e siècle, l'Europe abritait au moins deux souches différentes de la bactérie : une lignée ancienne, représentée par YYY093A et C11907, remontant à l'âge du fer, et une autre déjà très proche de la forme moderne de la maladie. Cela montre que la fièvre récurrente en Europe n'était pas causée par une seule souche, mais par plusieurs lignées coexistantes ayant des histoires évolutives distinctes.

Discussion des résultats et lien avec les sources historiques.

Des recherches récentes sur l'ADN ancien ont montré que les fièvres paratyphoïdes, causées par *Salmonella enterica* Paratyphi C, étaient présentes en Europe depuis des millénaires, et qu'en 1812 la maladie était déjà connue et documentée. Elle se propage lorsque des aliments ou de l'eau sont contaminés par des excréments humains contenant la bactérie. Les symptômes sont variés et non spécifiques : fièvre, maux de tête, éruptions cutanées, faiblesse, perte d'appétit, douleurs abdominales, diarrhée, constipation et vomissements. Cependant, pendant la campagne de Russie de Napoléon, ni la fièvre paratyphoïde ni la fièvre typhoïde ne sont mentionnées spécifiquement dans les documents historiques que nous connaissons. Cela est probablement dû au fait que les symptômes étaient très variés et pouvaient facilement être confondus avec d'autres maladies courantes de l'époque. Pourtant, un rapport de 1812 rédigé par J.R.L. de Kirckhoff, médecin au service de l'armée Napoléonienne, en donne une description révélatrice. Il écrit que les soldats arrivant à Vilnius souffraient de typhus, de dysenterie et surtout de diarrhée. Il souligna qu'une cause probable était leur habitude de consommer des tonneaux de betteraves salées, trouvées dans presque toutes les maisons entre Orcha et Vilnius. Les soldats ne se contentaient pas de manger les betteraves, ils en buvaient aussi le jus salé lorsqu'ils avaient soif, ce qui, selon lui, leur irritait l'estomac et les intestins. Ce témoignage pourrait correspondre à une possible infection par la fièvre paratyphoïde, car les aliments et l'eau contaminés sont des sources typiques d'infection et la diarrhée en est un symptôme classique. En même temps, ces signes pourraient aussi correspondre à d'autres maladies courantes du XIX^e siècle, comme la dysenterie ou le typhus. En fait, même aujourd'hui, avec la médecine moderne, il serait toujours impossible de distinguer entre typhus, typhoïde et paratyphoïde uniquement sur la base des symptômes et des témoignages, ce qui montre à quel point de telles épidémies pouvaient être mal identifiées dans le passé.

fièvre paratyphoïde sur la seule base des symptômes ou des témoignages des survivants.

D'un point de vue moléculaire, nos résultats apportent un solide appui à l'hypothèse selon laquelle les soldats étaient infectés par une fièvre paratyphoïde causée par *S. enterica Paratyphi C*. Bien que nous n'ayons pas obtenu une couverture génomique suffisante pour déterminer le positionnement phylogénétique précis au sein de la diversité connue de cette lignée, notre protocole rigoureux d'authentification constitue une preuve solide de sa présence. Notre étude fournit ainsi la première preuve directe que la fièvre paratyphoïde a contribué aux décès de soldats napoléoniens lors de leur retraite catastrophique de Russie. Cependant, le nombre limité d'échantillons analysés ($n = 13$), rapporté au nombre total de corps identifiés sur le site (plus de 3 000), ne permet pas de conclure que ce seul agent pathogène soit responsable de l'ensemble des décès. Compte tenu des conditions extrêmes et particulièrement dures de cette retraite, la présence d'infections multiples et concomitantes est hautement plausible. Le typhus a longtemps été rapporté comme ayant touché la Russie durant cette période, mais les données disponibles restent insuffisantes pour démontrer de manière concluante le rôle de cette maladie dans la dévastation de l'armée napoléonienne.

À la lumière de nos résultats, un scénario plausible pour expliquer la mort de ces soldats serait une combinaison de fatigue, de froid et de plusieurs maladies, parmi lesquelles la fièvre paratyphoïde et la fièvre récurrente transmise par les poux. Bien que cette dernière ne soit pas nécessairement mortelle, elle aurait pu considérablement affaiblir des individus déjà épuisés. Notre étude confirme la présence de deux agents pathogènes jusque-là non documentés, mais l'analyse d'un plus grand nombre d'échantillons sera nécessaire pour comprendre pleinement le spectre des maladies épidémiques ayant touché l'armée napoléonienne durant la retraite de Russie. Nos travaux démontrent que le séquençage à haut débit de l'ADN ancien constitue une approche puissante pour explorer la dynamique historique des maladies, et mettent en évidence sa capacité à identifier avec précision des agents pathogènes anciens, même lorsque seules des données génomiques limitées sont disponibles.

Limites de l'étude et portée de l'interprétation des résultats.

En ce qui concerne les évidences génétiques, l'étude apporte de solides éléments indiquant que certains soldats de l'armée Napoléonienne étaient infectés par la fièvre paratyphoïde, causée par *Salmonella enterica Paratyphi C*. Même si l'équipe n'a pas pu reconstituer le génome complet de la maladie pour le situer précisément au sein de la diversité globale de cette bactérie, l'authentification rigoureuse des fragments d'ADN ne laisse guère de doute quant à la présence du pathogène. Cela fait de cette étude la première à offrir une preuve directe que la fièvre paratyphoïde aurait pu jouer un rôle dans la mort de soldats Napoléoniens pendant la retraite de Russie. Cependant, ces résultats doivent être replacés dans leur contexte : seuls 13 soldats ont été analysés sur plus de 3 000 corps signalés sur le site. C'est un échantillon bien trop petit pour conclure que la fièvre paratyphoïde à elle seule était responsable de la mort de milliers de soldats. La retraite de Russie fut marquée par l'épuisement, les températures glaciales et la faim, autant de conditions qui auraient pu en elles-mêmes tuer ces soldats ou les rendre vulnérables à diverses maladies infectieuses. Le typhus, par exemple, est depuis longtemps suspecté d'avoir été un facteur majeur à cette période, mais les preuves disponibles restent aujourd'hui trop limitées pour confirmer sa présence et son impact.

Dans l'ensemble, les résultats suggèrent que la mort des soldats n'a pas été causée par une seule maladie, mais par une combinaison d'épuisement extrême, de froid intense et de multiples infections bactériennes. La fièvre paratyphoïde et la fièvre récurrente transmise par les poux de corps, toutes deux identifiées dans cette étude, ont probablement joué un rôle important dans l'épisode infectieux de Vilnius. La fièvre récurrente, à elle seule, n'est pas toujours mortelle, mais chez des soldats déjà affamés, transis et affaiblis, elle pouvait avoir de graves conséquences. Cette étude confirme la présence de deux agents pathogènes qui n'avaient jamais été documentés auparavant dans l'armée Napoléonienne. Cependant, comme un petit nombre seulement de soldats ont été analysés, beaucoup d'autres échantillons devront être étudiés pour obtenir un tableau complet des maladies épidémiques qui ont frappé les soldats pendant la retraite de Russie. Au-delà de ce cas précis, ces travaux montrent également comment les technologies modernes de séquençage de l'ADN peuvent nous permettre de mieux comprendre les épidémies du passé. Même lorsque seuls de petits fragments d'ADN subsistent, ces méthodes permettent d'identifier les agents pathogènes avec une précision remarquable et de reconstituer le rôle que les maladies ont joué dans les grandes crises sanitaires de l'histoire.

Pathogène	Échantillon ID	Lectures uniques alignées	Lectures uniques authentifiées	Longueur moyenne des lectures (bp)	Positions totales alignées	Positions non-N dans epa-ng MSA	Sites variables couverts
<i>S. enterica</i>	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
<i>B. recurrentis</i>	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. Résumé des statistiques de séquençage, d'authentification et de placement phylogénétique pour les échantillons anciens positifs à *Salmonella enterica* et *Borrelia recurrentis*. Ce tableau présente, en un seul endroit, à la fois la quantité d'ADN ancien récupérée et son utilité pour l'identification et le positionnement phylogénétique de chaque infection. Les « lectures uniques alignées » correspondent aux fragments dédupliqués qui s'alignent sur le génome d'un agent pathogène ; les « lectures authentifiées » représentent la partie de ces fragments confirmée de manière indépendante comme appartenant réellement à cet agent (et non à de l'ADN environnemental) grâce à une vérification taxonomique plus stricte ; la « longueur moyenne des lectures » (en paires de bases) indique le degré de fragmentation de l'ADN ancien (plus les fragments sont courts, plus l'ADN est dégradé) ; les « positions totales alignées » correspondent au nombre de sites du génome de l'agent pathogène couverts par au moins un fragment ; les « positions non-N dans l'alignement » sont les sites exploitables issus de ces fragments qui peuvent être comparés à un alignement multiple de référence pour le placement dans l'arbre phylogénétique ; enfin, les « sites variables couverts » représentent la partie de ces sites exploitables qui diffèrent effectivement entre les souches et portent donc le signal évolutif le plus informatif. Ces colonnes doivent être lues conjointement : un plus grand nombre de fragments authentifiés, de positions couvertes et de sites variables renforce la confiance dans la détection et dans la position d'un échantillon au sein de l'arbre phylogénétique de l'agent pathogène. Ainsi, pour *B. recurrentis*, l'échantillon YYY093A (4 062 lectures uniques ; 3 239 authentifiées ; 125 514 positions exploitables ; 5 854 sites variables) constitue une preuve très solide, tandis que YYY092B (322 ; 21 ; 584 ; 21) reste beaucoup plus limité. Le même schéma s'observe pour *S. enterica* : YYY087A (968 ; 874 ; 26 116 ; 800) est bien plus informatif que YYY095A (34 ; 29 ; 1 236 ; 44). Les deux génomes médiévaux de *Borrelia* publiés précédemment, en particulier C11907 (14 027 ; 13 235 ; 270 532 ; 13 392), présentent une couverture et un nombre de sites variables bien supérieurs, constituant ainsi des points de référence robustes. Dans l'ensemble, le tableau illustre la conservation inégale de l'ADN entre les individus, comme on peut s'y attendre pour des restes anciens, tout en apportant un soutien quantitatif solide à la présence de *S. enterica* et *B. recurrentis* chez certains soldats napoléoniens, et à la fiabilité du signal phylogénétique permettant leur positionnement évolutif.

Table 1. Résumé des paramètres de séquençage, d'authentification et de placement phylogénétique pour les échantillons anciens positifs à un agent pathogène. Les « lectures uniques alignées » correspondent aux fragments filtrés des doublons qui s'alignent sur le génome de l'agent pathogène ; les « lectures authentifiées » sont celles confirmées, à travers les trois étapes d'authentification, comme appartenant réellement à cet agent (et non à de l'ADN de fond) ; la « longueur moyenne des lectures (pb) » reflète le degré de fragmentation de l'ADN ancien ; les « positions totales alignées » indiquent le nombre de sites du génome couverts par au moins un fragment ; les « positions non-N dans l'alignement » sont les sites exploitables utilisés pour le placement phylogénétique ; enfin, les « sites variables couverts » représentent les différences informatives entre souches effectivement utilisées pour les inférences phylogénétiques. Des valeurs élevées dans ces colonnes renforcent à la fois la fiabilité de la détection et celle du placement (par exemple : *B. recurrentis* YYY093A >> YYY092B ; *S. enterica* YYY087A >> YYY095A). Voir également les Figures 2–3 et le Tableau 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., Kuncevicus, 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., Kirdök, E., Vicente, M., Naidoo, T., Van Der Valk, T., Altunışık, 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. Hübner, 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., Górski, 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., Holtmark 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., Borriani, 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., Lundström, 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., 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. <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](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., Kernysky, 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. (Librairie 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.