



HAL
open science

Characterization and cartography of the global genetic diversity, evolution and spread of *Yersinia pestis*

Guillem Mas Fiol

► **To cite this version:**

Guillem Mas Fiol. Characterization and cartography of the global genetic diversity, evolution and spread of *Yersinia pestis*. Microbiology and Parasitology. Université Paris Cité, 2022. English. NNT : 2022UNIP5245 . tel-04689528

HAL Id: tel-04689528

<https://theses.hal.science/tel-04689528v1>

Submitted on 5 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**INSTITUT
PASTEUR**



**Université
Paris Cité**

Université Paris Cité

Ecole doctorale Bio Sorbonne Paris Cité – ED562

**Institut Pasteur, UMR6047 Microbiologie Intégrative et Moléculaire,
Unité de Recherche Yersinia**

**Characterization and cartography of the
global genetic diversity, evolution and spread
of *Yersinia pestis***

Par Guillem Mas Fiol

Thèse de doctorat de Microbiologie

Dirigée par Javier Pizarro-Cerdá, HDR, PhD

Présentée et soutenue publiquement le 14 octobre 2022

Devant un jury composé de :

M. Julian Parkhill, Full Professor, University of Cambridge, Rapporteur

Mme. Sandra Reuter, Group Leader, University of Freiburg Medical Center, Rapportrice

Mme. Kate Baker, Full Professor, University of Liverpool, Examinatrice

M. François-Xavier Weill, Professeur HDR, Institut Pasteur, Université Paris Cité,
Examineur et Président du jury

M. Javier Pizarro-Cerdá, DR, Institut Pasteur, Université Paris Cité, Directeur de thèse



**INSTITUT
PASTEUR**



**Université
Paris Cité**

Université Paris Cité

Ecole doctorale Bio Sorbonne Paris Cité – ED562

**Institut Pasteur, UMR6047 Microbiologie Intégrative et Moléculaire,
Unité de Recherche Yersinia**

**Characterization and cartography of the
global genetic diversity, evolution and spread
of *Yersinia pestis***

By Guillem Mas Fiol

Doctoral thesis in Microbiology

Supervised by Javier Pizarro-Cerdá, HDR, PhD

Publicly presented and defended on the 14th of October 2022

In front of a jury composed by:

M. Julian Parkhill, Full Professor, University of Cambridge, Reviewer

Mme. Sandra Reuter, Group Leader, University of Freiburg Medical Center, Reviewer

Mme. Kate Baker, Full Professor, University of Liverpool, Examiner

M. François-Xavier Weill, Professeur HDR, Institut Pasteur, Université Paris Cité, Examiner
and President of the jury

M. Javier Pizarro-Cerdá, DR, Institut Pasteur, Université Paris Cité, Supervisor

This work is licensed under a [Creative Commons 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license

Ce travail est distribué sous licence [Creative Commons 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)



PH.D. THESIS IN MICROBIOLOGY

Characterization and cartography
of the global genetic diversity,
evolution and spread of *Yersinia*
pestis

by Guillem Mas Fiol

Ecole Doctorale BioSPC, ED562
Université Paris Cité
Paris, France

at Institut Pasteur
and Université Paris Cité

publicly presented and defended on the 14th of October 2022

Examining Committee:

Julian Parkhill, reviewer

Sandra Reuter, reviewer

Kate Baker, examiner

François-Xavier Weill, examiner

Javier Pizarro-Cerdá, thesis director

Title: Characterization and cartography of the global genetic diversity, evolution and spread of *Yersinia pestis*

Abstract

Plague is a severe zoonotic infectious disease caused by the gram-negative bacterium *Yersinia pestis*. Mostly known by its pandemic past, *Y. pestis* has spread from Central Asia in multiple occasions and it has successfully established in persistent endemic areas present in Asia, Africa, and the Americas, where plague represents a threat for public health. The patterns of global genetic diversity and spread of *Y. pestis* have not been characterized in detail yet, as some lineages and important endemic areas were missed in previous genomic studies. Here, we generate whole-genome sequences of more than 1000 *Y. pestis* isolates collected over more than 100 years from major endemic areas of plague. Together with most of the available modern and ancient genomes in public repositories, we conduct the largest phylogenomic analysis of *Y. pestis* to date. We refine the species population structure by identifying new clades and infer routes of spread of plague in regions that were overlooked in previous studies, notably in Africa. We also characterize new signatures of expansion and functional diversification within lineages, identifying functional variants in lineages that had been poorly sampled before. By using different genomic approaches, we identify specific gene repertoires in different lineages and characterize the expansion of insertion sequences (IS) during *Y. pestis* evolution, particularly within the 1.ANT lineage where we describe an enrichment in IS-driven gene interruptions. In the present study, we also apply phylodynamic and phylogeographic approaches to uncover the evolution and spread of *Y. pestis* in Vietnam, using hundreds of isolates recovered across four decades in the context of extensive

human plague outbreaks and animal surveillance efforts. We identify a major clone that circulated during major outbreaks in the country and describe contrasting patterns of *Y. pestis* spread between the period of maximal incidence of plague, compared to the decades following the issue of the Vietnam War. Our results support the hypothesis that the rapid large-scale spread of *Y. pestis* in Vietnam was human-mediated and that it determined the long-term persistence of plague in the last decades of the past century. This work provides unprecedented insights into the evolutionary mechanisms that have shaped the current diversity within the *Y. pestis* species and describes the use of new frameworks for the study of *Y. pestis* maintenance and spread in epidemic and endemic contexts.

Keywords

Plague, *Yersinia pestis*, genomics, evolution, phylogenetics, molecular epidemiology, whole-genome sequencing

Titre: Caractérisation et cartographie de la diversité génétique globale, évolution et dispersion de *Yersinia pestis*

Résumé

La peste est une maladie infectieuse zoonotique grave causée par la bactérie gram-négative *Yersinia pestis*. Notamment célèbre en raison de son passé pandémique, *Y. pestis* s'est propagée depuis l'Asie centrale à plusieurs reprises et s'est établi avec succès dans des zones endémiques présentes en Asie, en Afrique et dans les Amériques, où la peste représente une menace pour la santé publique. La diversité génétique et la propagation de *Y. pestis* n'ont pas encore été caractérisées en profondeur, car certaines lignées et zones endémiques importantes ont été omises dans les études génomiques précédentes. Dans cette étude, nous avons généré des génomes complets de plus de 1000 souches de *Y. pestis* collectées au long de 100 ans dans les principales zones endémiques de peste. En combinant ces génomes avec la plupart des génomes modernes et anciens disponibles dans les bases de données publiques, nous effectuons la plus grande analyse phylogénomique de *Y. pestis* à ce jour. Une mise à jour de la structure de la population de l'espèce a été effectuée en identifiant de nouveaux clades, et des nouvelles routes de propagation de la peste ont été identifiées dans des régions qui ont été négligées dans les études précédentes, notamment en Afrique. Nous caractérisons également des nouveaux signaux d'expansion et de diversification fonctionnelle au sein des lignées, en identifiant des variantes fonctionnelles dans des lignées qui avaient été peu échantillonnées auparavant. En utilisant différentes approches génomiques, nous identifions des répertoires de gènes spécifiques dans différentes lignées et caractérisons l'expansion de séquences d'insertion (IS) au cours de l'évolution de *Y. pestis*, en particulier au sein de la lignée 1.ANT où nous décrivons un enrichissement en interruptions de gènes induites par IS.

Dans la présente étude, nous appliquons également des approches phylodynamiques et phylogéographiques pour analyser l'évolution et la propagation de *Y. pestis* au Vietnam, en utilisant des centaines de souches collectées sur quatre décennies dans le contexte d'importantes épidémies de peste et durant des travaux de surveillance chez l'animale. Nous identifions un clone principal qui a circulé lors des épidémies majeures dans le pays et décrivons des différentes dynamiques de propagation de *Y. pestis* entre la période d'incidence maximale de la peste et les décennies qui ont suivi l'issue de la guerre du Vietnam. Nos résultats soutiennent l'hypothèse selon laquelle la propagation rapide et à grande échelle de *Y. pestis* au Vietnam était d'origine humaine et qu'elle a déterminé la persistance à long terme de la peste au cours des dernières décennies du siècle dernier. Ce travail fournit des informations sans précédent sur les mécanismes et les événements évolutifs qui ont façonné la diversité actuelle au sein de l'espèce *Y. pestis* et décrit l'utilisation de nouvelles méthodes pour l'étude de la dispersion et du maintien de *Y. pestis* dans des contextes épidémiques et endémiques.

Mots clés

Peste, *Yersinia pestis*, génomique, évolution, phylogénie, épidémiologie moléculaire, séquençage

Résumé long

La peste est une maladie infectieuse zoonotique grave causée par la bactérie gram-négative *Yersinia pestis* et transmise par des ectoparasites, notamment différentes espèces de puces. La peste est l'une des maladies connues les plus anciennes pour l'homme, tant qu'elle est considérée la seule maladie infectieuse épidémique caractérisée par la présence de ganglions inflammés, appelés "bubos", ce qui donne le nom à la forme principale de la maladie, connue sous le nom de peste bubonique. Notamment célèbre en raison de son passé pandémique, la peste a en effet été la cause d'au moins trois pandémies historiques. La première, ou peste du Justinien, a touché la région méditerranéenne aux 6e et 8e siècles, tandis que la deuxième pandémie (qui démarre par la Peste Noire) causa la mort d'entre 30-50% de la population européenne au Moyen Age (14e-18e siècles). Enfin, la troisième pandémie démarre en Chine vers la fin du 19e siècle et se propage à échelle mondiale grâce aux routes de transport maritime en plein développement. Aujourd'hui, la peste représente toujours un risque pour la santé publique dans plusieurs pays où la maladie s'est établie de façon endémique. La peste existe surtout en tant que maladie des rongeurs qui servent de réservoir principal du bacille *Y. pestis*, à partir desquels la transmission zoonotique vers les humains a lieu de manière accidentelle. En effet, la peste circule de façon active dans le réservoir animal dans les régions du monde communément appelées foyers, qui touchent une trentaine de pays dans le globe qui ont notifié des cas de peste humaine dans les dernières 30 années. Les puces représentent des éléments clés pour l'écologie et la transmission de la peste, car elles servent de vecteurs majeurs pour la transmission de *Y. pestis* chez les rongeurs, ainsi que

pour la transmission zoonotique de ceux derniers vers les populations humaines qui habitent dans les zones endémiques. En raison de son réservoir animal et de sa nature zoonotique, la peste reste une maladie qui ne peut probablement pas être éradiquée, ce qui justifie la nécessité de surveiller de façon continue la présence et distribution de *Y. pestis* dans les foyers endémiques, ainsi que de poursuivre la recherche sur son risque épidémique et transmission, le risque d'antibiorésistance ainsi que des stratégies prophylactiques. L'épidémie de peste à Madagascar en 2017 sert d'exemple de ce risque de réémergence épidémique de la peste dans un pays endémique et de sa capacité de propagation jusqu'aux zones urbaines, ce qui peut éventuellement porter à la propagation de la maladie à l'échelle internationale dans un contexte de mouvements humains globalisés. Les pays qui ont notifié le plus de cas de peste humaine dans les dernières 25 années sont en effet Madagascar, la République Démocratique du Congo, la Tanzanie et l'Ouganda, mais des cas sporadiques de peste chez les hommes ont eu aussi lieu dans les dernières années en Chine, Mongolie, Pérou ou encore les Etats Unis, ce qui montre l'ampleur de la distribution globale de la maladie aujourd'hui. La peste demeure aujourd'hui une maladie liée à la pauvreté, car c'est dans des régions isolées et parfois politiquement instables où les populations humaines continuent à habiter dans des cabanes, exposées au contact avec des rongeurs sylvaux et domestiques et leurs parasites qui peuvent être vecteurs de maladies comme la peste. Le contrôle de la peste humaine dans les régions endémiques nécessite donc des approches de santé globale, y compris la surveillance du pathogène et sa distribution, le développement d'outils de diagnostic rapide, des méthodes prophylactiques efficaces, des connaissances sur les relations entre hôte, vecteur et pathogène, ainsi que d'une maîtrise pratiques culturelles humaines.

La bactérie *Y. pestis* est une gammaprotéobactérie anaérobie facultative et non-motile qui appartient au genre *Yersinia*. Le genre *Yersinia* est composée de 26 espèces identifiées et comprends trois lignées qui sont des pathogènes majeurs chez l'homme et d'autres mammifères : *Y. pestis*, l'agent de la peste, et *Y. pseudotuberculosis* et *Y. enterocolitica*, qui causent des

yersiniose entérique. Le genre *Yersinia* reste un exemple intéressant d'acquisition parallèle des mécanismes de pathogénicité dans ces lignées qui sont pathogènes chez l'homme. Cette acquisition des capacités de pathogénicité est associée à l'acquisition du plasmide pYV, qui contient les gènes responsables de l'expression du système de sécrétion de type 3 (T3SS) ainsi qu'à des protéines effectrices, et de l'îlot de pathogénicité qui contient le sidérophore Yersiniabactin. Le séquençage du génome complet de la souche de référence CO92 révéla d'importants détails sur les caractéristiques génétiques et évolutives de *Y. pestis*, à savoir, des évidences de sa plasticité génétique, l'accumulation de pseudogènes, l'enrichissement en séquences d'insertion, ainsi que la présence de trois plasmides de différentes origines. Ce génome de référence consiste d'un chromosome de 4,65 Mbp et de trois plasmides (pYV : 70,3 kbp ; pMT1 : 96,2 kbp ; pPla : 9,6 kbp). La comparaison des séquences géniques entre des souches de *Y. pestis* et de *Y. pseudotuberculosis* montra que *Y. pestis* ne serait donc qu'un clone de *Y. pseudotuberculosis* qui aurait acquis des mécanismes de virulence spécifiques lui conférant un phénotype d'infection et de transmission qui permettent d'expliquer son extrême pathogénicité et signes cliniques différents de ceux des infections par *Y. pseudotuberculosis*. L'émergence de *Y. pestis* à partir de son ancêtre *Y. pseudotuberculosis* est effectivement caractérisée par des événements d'acquisition de nouveaux gènes par transfert horizontal (notamment les plasmides pMT1 et pPla), ainsi que par la réduction génomique menant à l'inactivation et la perte d'expression de plusieurs facteurs (environ 10% des gènes ont été inactivés chez *Y. pestis* par rapport à *Y. pseudotuberculosis*). Ces événements moléculaires marquent la transition d'un mode de vie et de transmission typique d'un pathogène gastro-intestinal vers celui d'un pathogène adapté à la colonisation d'une nouvelle niche, celui du tracte digestif des puces qui deviennent son vecteur principal de transmission.

L'entrée dans l'ère génomique a permis la génération de données de séquençage à haut débit pour plusieurs centaines de souches de *Y. pestis* au cours de la dernière décennie. L'usage de marqueurs moléculaires stables et distribués au long du chromosome tels que

des polymorphismes par mutations ponctuelles (SNP) a permis pour la toute première fois d'identifier les principales lignées évolutives et de caractériser leur distribution géographique et dissémination en dehors de l'Asie à plusieurs reprises. Ces analyses précédentes se sont basées sur un nombre réduit de souches de *Y. pestis* collectées notamment dans l'Asie, et qui ont montré l'évolution clonale du pathogène au niveau de l'espèce, avec des lignées représentées par des branches qui émergent d'une branche évolutive centrale dans la phylogénie de *Y. pestis*. La nomenclature couramment utilisée pour classifier les génomes dans les lignées identifiées s'est basée sur ces analyses, et prend la branche principale d'appartenance du génome (0-4) ainsi que la biovar basée à priori sur la capacité de certaines lignées à dégrader des sucres particuliers (biovars ORI, ANT, MED, PE, IN). De même, ces analyses suggèrent un possible origine de l'émergence de *Y. pestis* quelque part en Chine ou en Asie Centrale, ainsi que la dissémination d'une lignée clonale, 1.ORI, au niveau global durant la 3e Pandémie (1855-jusqu'à présent). Les analyses phylogénomiques représentent par conséquent des approches particulièrement intéressantes pour analyser la diversité génétique du bacille de la peste, pour comprendre sa maintenance et dispersion, ce qui peut éventuellement apporter des informations de valeur pour comprendre l'épidémiologie de la peste. De plus, le développement des analyses paléogénomiques dans les années récentes a permis la reconstruction des premiers génomes et métagénomes anciens de *Y. pestis* à partir des ADN isolés de victimes associées à des pandémies historiques de peste en Europe, et sporadiquement dans des échantillons d'ADN ancien du Néolithique. Ces analyses ont permis non seulement de prouver la présence de *Y. pestis* lors de ces pandémies historiques et d'autres périodes historiques, mais aussi d'utiliser ces analyses génomiques pour inférer les lignées évolutives de *Y. pestis* qui ont causé ces pandémies historiques, notamment la peste du Justinien et les pestes du Moyen Age (deuxième pandémie).

Malgré ces avancements récents, notre compréhension de la diversité génétique et de l'évolution de l'espèce *Y. pestis* s'avère incomplète car des régions endémiques importantes ont été omises

ou peu couvertes dans des analyses génomiques précédentes, ce qui limite une connaissance plus approfondie des dynamiques de maintien et dispersion de ce pathogène. En outre, les études ayant analysé les variations génomiques altérant la fonction des gènes (et donc causant des potentielles différences entre souches menant à des changements d'adaptation ou de virulence) se sont basées sur l'analyse d'un faible nombre de génomes, ce qui a pu limiter l'identification de mutations présentes dans des lignées qui étaient absentes dans ces analyses. L'objectif principal de cette thèse était donc de générer un ensemble de données génomiques nous permettant de développer des comparaisons de génomes de *Y. pestis* à l'échelle de l'espèce, y compris la diversité qui est disponible actuellement avec les ADN anciens. Dans ce but, nous avons séquencé le génome complet de plus de 1000 souches de *Y. pestis* d'une collection globale de souches collectées au long du dernier siècle jusqu'à nos jours (1899-2017), et qui comprend des importantes régions endémiques de peste présentes parmi 29 pays représentés dans la collection. Tout prenant une perspective génomique et d'évolution moléculaire, nous développons d'abord dans un premier manuscrit une caractérisation de la diversité génétique, de la distribution géographique et de l'histoire évolutive de *Y. pestis*, afin de répondre à des questions qui font partie de la compréhension de l'évolution globale de *Y. pestis* et de l'épidémiologie de la peste. Nous avons donc reconstruit l'analyse phylogénomique la plus complète de l'espèce *Y. pestis* (plus de 2800 génomes), en ajoutant les données générées dans ce projet avec celles disponibles dans les bases de données publiques. En se concentrant sur des centaines de génomes de souches isolées lors de la Troisième Pandémie, nous retraçons l'évolution et dispersion d'une sous-lignée qui représenterait une des vagues de dispersion de cette pandémie et qui serait introduite parallèlement en Afrique, en Amérique du Sud et sporadiquement en Europe vers la fin du 19e siècle, d'après nos estimations issues de phylogénies datées. Nous avons apporté pour la toute première fois une estimation de l'émergence et de l'évolution de la lignée 1.ANT qui circule notamment en Afrique Centrale et de l'Est, ce qui servira de base pour répondre à des questions concernant

le possible origine et route d'introduction de cette lignée dans le continent Africain. En effet, l'évolution globale des lignées de *Y. pestis* est marquée par leur histoire de dispersion dans des nouvelles régions. Nous avons montré que les différentes lignées présentent aussi des différentes longueurs de branches terminales dans les phylogénies, ce qui pourrait être interprété comme un nouvel indicateur du degré d'expansion des lignées de *Y. pestis*. D'autre part, nous développons une analyse intégrée de différents mécanismes de variation génomique sur des centaines de génomes de *Y. pestis* afin de révéler des possibles signaux de diversification et d'adaptation chez *Y. pestis*. Nous avons observé un enrichissement en mutations dans des gènes codant des facteurs de réponse aux stressés parmi des génomes des lignées 1.ORI et 1.ANT et nous identifions plusieurs gènes associés à la virulence des *Yersinia* parmi les gènes ayant accumulé le plus de mutations parmi les génomes analysés. De même, l'analyse pangénomique de plus de mil génomes apporte des nouvelles évidences que les différentes lignées se caractérisent par la présence de répertoires géniques spécifiques. Nous montrons enfin que le nombre de copies d'éléments d'insertion (IS) n'a pas cessé d'augmenter au long de l'évolution de l'espèce *Y. pestis*, contribuant à l'inactivation de certaines fonctions géniques, notamment pour la lignée 1.ANT. Les effets spécifiques de ces variations génomiques sur l'adaptation et virulence de *Y. pestis* seront explorés dans des prochains études.

D'autre part, des analyses phylodynamiques et phylogéographiques peuvent servir pour comprendre la maintenance et la dispersion de *Y. pestis* dans des pays où la peste est endémique. Dans le deuxième manuscrit, nous explorons les dynamiques évolutives et la dispersion de *Y. pestis* à partir de centaines de souches isolées au siècle dernier au Vietnam, qui fut l'une des régions de peste les plus actives dans le siècle passé. Nous démontrons l'émergence et circulation d'un clone majeur dans le pays lors des importantes épidémies de peste dans la décennie de 1960, ce qui suggère des causes anthropogènes derrière la dissémination de ce clone (appelé sc1). La dispersion rapide de ce clone majeur se déroula dans un contexte de conflit armé prolongé dans le pays, et sa transmission et établissement dans des nou-

velles aires géographiques mena à l'émergence de plusieurs lignées de *Y. pestis* (appelées sc3-sc13) qui seraient responsables de la persistance de la peste au Vietnam jusqu'à la fin du siècle dernier. Cette étude montre également la nécessité de maintenir la surveillance de la circulation de *Y. pestis* dans les pays endémiques au long de plusieurs décennies afin de mieux comprendre les dynamiques épidémiologiques et évolutives de *Y. pestis* dans les pays endémiques.

En définitive, cette thèse montre que l'intégration des données génomiques de centaines ou milliers de génomes comprenant des régions de peste historiquement actives reste essentiel pour découvrir des évènements de variation génomique et des possibles signaux d'évolution adaptative chez *Y. pestis*, ainsi que de caractériser la diversité génétique de ce pathogène de la distribution de ses lignées. Ces efforts doivent être menés davantage grâce à la collaboration entre plusieurs acteurs, de ceux chargés de maintenir la surveillance dans les pays endémiques jusqu'à ceux chargés de développer des approches d'épidémiologie génomique, ce qui permettrait de développer une connaissance plus approfondie des dynamiques de maintenance et dispersion de la peste et éventuellement d'informer des mesures de contrôle de la maladie plus efficaces.

Acknowledgments

The arbitrary combination of decisions and events that allowed me to complete this work begins by Javier Pizarro Cerdá, who gave me the opportunity to join his (re)emerging laboratory and allowed me to make my own this unique project he envisioned four years ago. Throughout these years, he has provided me with the independence to propose my views at every stage of the study and has always encouraged me to go further when I could. Besides his honest supervision, he has always given me the time to try to learn by myself with the advent of the different challenges that this work represented and taught me that seeking for help is also part of the scientific process. This is to me one of the things that make a true PhD experience. His qualities as a thesis director extend to his role as an inspiring leader of the Yersinia unit, without whom the lab would have a completely different color. *Pura vida.*

I take a step back and thank professors Roland Brosch and Lluís Quintana Murci for giving me the chance to discover the Institut Pasteur when this was my true ambition during my Master studies. I recall the early days when I joined the project of Dr. Jan Madacki in Brosch's lab on Mycobacteria for my internship, who encouraged me to continue in research and stay in the singular scientific environment that the Institute offers to all pasteurians.

I would like to thank Dr. Julian Parkhill and Dr. Sandra Reuter, reviewers of this thesis, for their time in the critical reading of the manuscript. I would also like to thank Dr. Kate Baker and Dr. François-Xavier Weill for having accepted, together with the reviewers, to take part in the defense jury. Their own work and experience have been a source of inspiration for the goodness of this project. My thanks to Professor Weill extend to his

mentoring, critical comments and valuable ideas given during our regular discussions from his role as tutor in the University committee all these last years.

This work would not have been reached till this point if it was not by no other than Dr. Nicolás Rascovan. Even before he started his new and dynamic lab in Pasteur, he was already convinced that this project deserved his help and coordination of critical analyses and methodology. I am aware that all the discussions and technical implementations developed on his side occupied a relevant amount of time that he could have otherwise dedicated to other projects of his own, but he decided to give us this chance and I truly thank the decision he made. His playlist of great Argentinian rock and punk anthems that he shared to me will continue to hammer the walls of my room for long time. My sincere thanks also go to Aida Andrades Valtueña, Frédéric Lemoine and Damien Mornico for their implication in the inception and development of important analyses of this work, without whom some of the most intriguing observations presented here would not have been reached on time.

To our colleagues in the CIBU-ERI Unit, Charlotte Balière, Aurelia Kwasiborski and Valérie Caro for making the sequencing of the historical *Yersinia pestis* collection a reality thanks to their expertise and sequencing platform. I truly appreciated the time invested in preparing what looked like an endless number of libraries to generate, as well as for their kind supervision and time in training me for genome libraries preparation and shotgun sequencing.

I am grateful to Dr. Simon Dellicour for giving me the chance to share with him my views on the project and helping us developing analyses for which his pioneering work has been determinant. I would like to extend my thanks to Dr. Eduardo Rocha and Dr. Eugen Pfeifer for all the interactions and letting me join their intriguing projects, that I wish I can help developing in the following months. I also thank Professor Hendrik Poinar and Ravneet Sidhu for sharing with us their current research and letting us help in what I am sure will be a beautiful story.

Perhaps the person who is responsible for my very first steps into microbial genomics

and phylogenies is my colleague and deputy director of the NRL for plague & Yersinioses, Cyril Savin. Besides his decisive implication in the conceiving of this project, he has always been eager to help in any request I had and invested his time in training me at the very beginning and continued throughout all these years. His sincere and contagious good mood and regular references to French culture, food, beer, and football are one of the good reasons I would join again the lab if I had the chance to do so.

A mes chers collègues et amis de l'Unité Yersinia et CNR Peste & Yersinioses, pour avoir aidé à créer une ambiance de recherche unique où la discussion et l'entraide sont la norme. Ces dynamiques du quotidien ainsi que toutes les soirées et apéros ont fait cette expérience de thèse très agréable et beaucoup plus facile à porter: Mara, Hebert, Jazmin, Julien, Clarisse, Remi, Marion, Pierre Le-Bury, Lou, Christian, Anne, Pierre Goossens, Olivier, Anne-Sophie, Elisabeth, Sylvie, Nadira et Laura.

To my Pasteurian colleagues in the StaPa association, for constantly bringing fresh air and making the "normal" experience at Pasteur a singular one: Laura, Virginia, Valerio, Dariusz, Francesca, Carys, Viviana, Eugen, Florian, Selma and Alexandra.

Les souvenirs coulent aussi avec toutes les personnes que j'ai eu la chance de croiser vos chemins à Paris. A tous mes camarades de Master, avec qui j'ai pu partager mes premiers jours dans cette ville et avec qui nous avons combattu ensemble cette petite bataille qui est le doctorat: Sara, Andrea, Bertrand, Valeria, Yohann, Marina, Oghuzan, Caterina, Lisa, Charlie, Serena, Grégoire and Sofia.

A Maylis Layan, pour les longues discussions et soutien face à l'inexorable dynamique de nos projets et intérêts scientifiques.

A François Simon, Egill Richard et David Routhieau pour être les responsables de la plupart de mes connaissances sur la culture populaire et la langue française, et pour avoir été le pilier qui soutenait moralement les jours sombres de confinement. To my colleagues in Pasteur & Paris with whom we have shared many good moments and supported each other:

Fabiana, Lena, Flo, Remi, Nathan, Kyrie, Fernando, Sol, Lucia, Mireia and many more!

Στις φίλες μου Νεφέλη και Εβίτα, που με εκαναν να ανακαλυψω οτι κουβαλουσα ενα κομματι της Ελλαδας μεσα στο πνευμα μου. Με πολλη αγαπη και αλκοολ.

To my music friends in the “Fronton” band, for our good old days of concerts in Paris and the dinners with Italian pasta that largely justified my presence in all the rehearsals at the Cité Universitaire: Adonnis, Samantha, Josemi, Pablo, Carlos and Jonathan.

A Leire Ugarte, per compartir amb el món un camí de poesia i política i amb mi uns anys d'experiències comunes en aquesta ciutat. Maitasun asko eta irabazi arte. A na Cristina Costa, pels records més divertits dels vespres i despertars a Paris. A en Xals de Berga, per totes les llargues marxegues dels dissabtes d'hivern, els records més bojós que tenc d'aquests anys, i per haver-me fet descobrir la connexió espiritual entre els nostres pobles, mediada per diferents beuratges i costums ancestrals.

A mes camarades et sympathisant.e.s de l'Alternative International Movement - Section Paris pour leur infatigable travail pour maintenir vive la flamme de la musique et des concerts alternatifs sur Paname, tout se débarrassant de toute ambiguïté politique: Théo, Bapt, Mathilde, Goui, Fred, Marcel, Sol, Luana, Lucas, Riko et Alison. A mes camarades du créneau de boxe populaire, pour tout l'apprentissage et pour m'avoir découvert une forme saine de dégager le stress et de se défendre de manière collective.

Als meus amics de sempre, de Felanitx, que des d'aquells dies on sèiem a un banc del pati d'institut no hem abandonat mai aquest humor i ganes de viure que nos varen juntar. Lo que ha faltat més en aquests quatre anys han estat tots els moments en els que no he pogut compartir amb vosaltres com solem fer.

A la meua familia, que m'ha donat el privilegi de poder decidir el meu propi camí i donat suport en totes les decisions: el que m'ha duit a avui escriure això no deixa d'estar impregnat de tota la vostra inspiració.

Contents

Abstract	ii
Résumé	iv
Résumé long	vi
Acknowledgments	xiii
I Introduction	1
1 Plague	2
1.1 Plague epidemiology and distribution	4
1.2 Reservoirs, vectors and transmission cycle	12
1.2.1 Hosts & animal reservoirs of plague	12
1.2.2 Vectors and vector-borne transmission of plague	13
1.2.3 Other sources of <i>Y. pestis</i> in nature	16
1.3 Transmission cycle and plague ecology	18
1.3.1 Transmission within the animal reservoir	18
1.3.2 Transmission of plague in human populations	19
1.4 Clinical signs and pathology of plague	22
1.4.1 Bubonic plague	22
1.4.2 Septicemic plague	23
1.4.3 Pneumonic or pulmonary plague	24

1.5	Plague diagnostics	26
1.6	Plague treatment & antibiotic resistance	29
1.7	Vaccination against plague	32
2	<i>Yersinia pestis</i>	34
2.1	Evolution	35
2.1.1	<i>Yersinia</i> genus: taxonomy, ecology, and evolution	35
2.1.2	<i>Y. pestis</i> genome and molecular microbiology	38
2.1.3	Emergence of <i>Y. pestis</i> from <i>Y. pseudotuberculosis</i>	46
2.2	Population genetics of <i>Y. pestis</i>	50
2.2.1	Genetic monomorphism, selection, and neutral evolution in <i>Y. pestis</i>	50
2.2.2	Genetic diversity of <i>Y. pestis</i> , genotyping and distribution	52
2.2.3	Microevolution of <i>Y. pestis</i>	57
2.2.4	Diverse mutation rates and the cryptic molecular clock of <i>Y. pestis</i> .	59
2.3	Spread of plague inferred from phylogenies	63
2.3.1	Spread of plague during the 3 rd Plague Pandemic and contemporary <i>Y. pestis</i> lineages	64
2.3.2	Spread of plague inferred from ancient DNA and paleogenomics . . .	66
2.4	Molecular typing and genomic epidemiology of plague	73
II	Objectives	79
III	Results	83
3	Results	84
3.1	Manuscript A - Genetic diversity, evolution and spread of plague revealed through the analysis of global <i>Yersinia pestis</i> genomes	84

3.2	Manuscript B - Outbreak dynamics and spread of <i>Yersinia pestis</i> in an endemic country using phylodynamic and phylogeographic approaches	122
IV	General discussion	144
V	Concluding remarks	153
VI	References	157

Part I

Introduction

Chapter 1

Plague

Plague is a vector-borne zoonotic disease which is transmitted by ectoparasites, mostly different species of fleas, to a variety of wildlife rodents. Transmission within rodent communities creates natural reservoirs for plague, which are present widely around the world and ensure the maintenance of the disease (1). Caused by the gram-negative bacterium *Yersinia pestis*, plague is one of the oldest known diseases in human history (2), likely being the only epidemic-causing disease characterized by the presence of swollen lymph nodes known as *buboes* that appear in the groin or armpit (figure 1.1). This feature gives its name to the main form of the disease known as Bubonic plague. However, plague has diverse clinical manifestations, including septicemic infection (septicemic plague), and pneumonic infection (pneumonic or pulmonary plague). The word plague, derived from its Latin form *plaga*, meaning “malignant disease”, “destruction” or “wound”, and has been coined since the ancient times to refer to a variety of scourges or diseases in humans and animals (see Horobin S.). Alternatively, the term *pestilence*, derived from Latin *pestis*, has also been used to refer to the disease and therefore gives its name to its etiological agent, the plague bacillus *Yersinia pestis*. Indeed, any of the terms *plague*, *pestilence* and *pest* have been generally used to describe catastrophic events in the past (3). This portrayal of plague in our common language can be understood by the devastating effects of plague pandemics in history (see

section 2.3.2.).

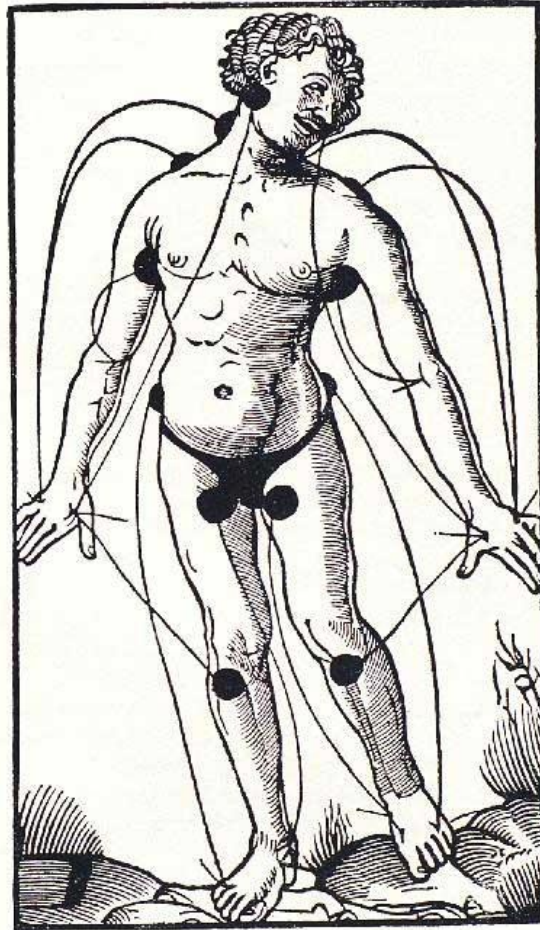


FIGURE I. — *Siège des Bubons.*

Estampe de l'édition française du *Règime contre la peste*, fait et composé par Messieurs les Médecins de la cité de Bâle, imprimée par Claude Nourry, Lyon, après 1519 (Klebs et Droz).

Figure 1.1. Location of the *buboes* in human plague. As indicated, the print was obtained from the 1519 French version of the "Regime against pestilence, done and composed by the doctors from the city of Bale", extracted from (4).

Section 1.1

Plague epidemiology and distribution

Plague still represents a threat in public health in numerous countries in which the disease is endemic. Existing first as a disease of wild rodents, plague is maintained within rodent populations that serve as the main reservoir of the disease, from which zoonosis can occur (1). Fleas represent key elements in plague ecology and transmission, acting as major vectors of *Y. pestis* transmission within rodent reservoirs (5). Plague actively circulates within the animal reservoir in regions of the world frequently called in the literature as *natural plague foci* (6). Therefore, countries considered to endemic for plague are those in which natural foci are present and co-exist with human populations, creating thus a risk of human plague (see the [WHO site](#)). At the present day, about 33 countries may be considered endemic for plague, since they have at least one natural plague focus (having shown signs of activity in the last 30 years)(7). Despite not being a WHO-notifiable disease since 2005, plague is subject to International Health Regulations and the occurrence of suspect cases of pulmonary plague in any country and suspect cases of plague in areas not considered to be endemic for plague must be notified to the WHO (8). Concerning the clinical manifestations of plague in outbreaks, declared cases of plague altern between bubonic form and less frequent pneumonic form of the disease (9). However, pneumonic plague has been responsible of worrisome urban plague outbreaks, since it allows interhuman transmission from pneumonic cases to close contacts and can therefore lead to a rapid expansion of the disease (10, 11), whereas bubonic forms are mostly associated to sporadic plague cases (7).

Overall, human plague incidence around the world in the recent decades appears to decline. This may be explained by the occurrence of plague cases mostly in remote areas, which makes epidemic events less likely, as well as the deployment of surveillance and response efforts in endemic countries such as Madagascar (7, 11). It is however important to note

that the distribution of human plague incidence is far less noticeable than the distribution of the plague endemic foci: if about 33 countries appear to host endemic areas of plague, only 5 countries declared human plague cases in 2018. Indeed, areas at risk of plague are diverse and vast in extension, yet some characteristics are common to most known plague foci (2). Endemic foci are frequently found I) in low-populated or remote rural areas, II) at higher altitudes relative to the rest of the country and III) within arid or semiarid areas with limited average precipitation (12–18). It has been recently pointed out that high soil salinity co-localize with plague outbreaks in the US (19), in agreement with the observed resistance of *Y. pestis* to high salinity (20). Most importantly, active plague foci require indeed the presence and survival of *Y. pestis* bacteria, one or more rodent reservoir species and one or more competent ectoparasite vector. Vector competence represents the ability of a vector to transmit a disease and hence determines its transmission potential (21).

Plague in Africa. In the last 25 years (data from 1994-2018), 45 262 cases of human plague have been notified to the WHO by 25 different countries in Africa, Asia and the Americas (7, 8, 22, 24) (**figure 1.2A**). The African continent alone has experienced most cases worldwide (91%), being Madagascar and in the Democratic Republic of Congo (DRC) at the head of plague incidence (16, 23). Madagascar has experienced plague cases every year, with plague foci distributed within the central highlands of the island, mainly distributed within the Tsiroanomandidy and the Ambositra districts (25). This *triangle of plague* endemic areas is linked to the presence of two flea species that are vectors of plague, *Xenopsylla cheopis* and *Synopsyllus fonquerniei*, which are present at high elevations but not in lower and in the coast of Madagascar (26) Competent species of fleas are considered responsible for the maintenance and spread of plague among rodent populations (*Rattus rattus* and *Rattus norvegicus*) in Madagascar endemic areas, contributing to the risk of zoonosis to humans (27). Consequently, most cases of plague in Madagascar are bubonic (transmitted from rodents mostly by flea bites), that in some cases may progress to pulmonary forms

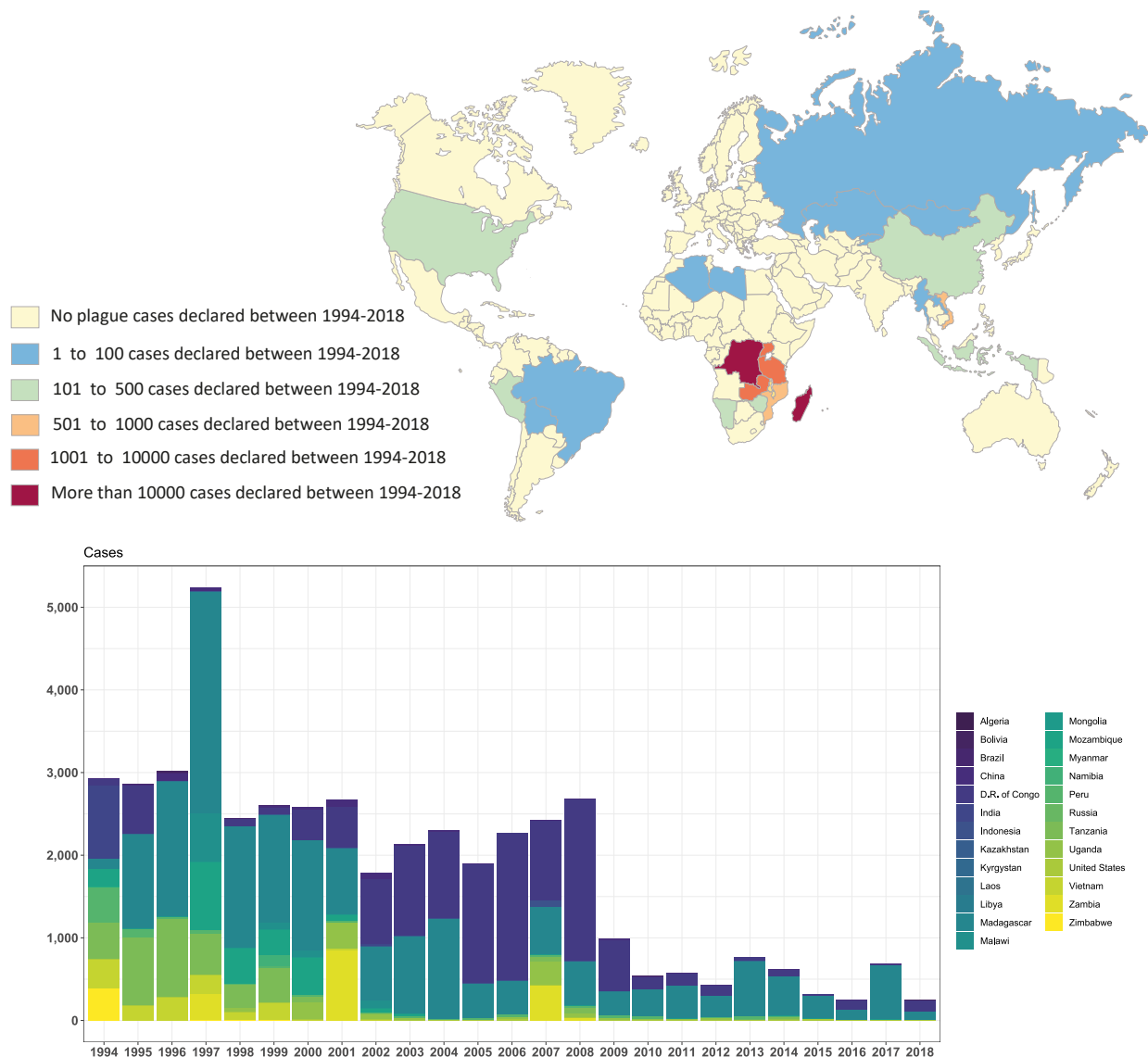


Figure 1.2. A: World map with countries colored based on cases of human plague declared to the World Health Organization (WHO) in the last 25 years of available reports (1994-2018). **B:** Incidence of plague in the last 25 years. The bars summarize the total recorded cases of human plague reported to the WHO (1994-2018). Based on available data from the WHO (7, 22, 23) (see the [Weekly Epidemiological Records](#)).

(28). In 2017, however, Madagascar experienced a plague outbreak of unusual size consisting mostly of pulmonary plague cases, which affected urban areas including the capital city Antananarivo (11). This outbreak raised concerns regarding the confirmed human-to-human transmission of pulmonary plague occurring in urban areas (7), especially in populated cities near ports and airports where international spillover may occur. The last concerning episodes

of urban plague in Madagascar had been observed in the 1990s within and near the seaport city of Mahajanga in the north of the island, where the pathogen appeared to have been imported in multiple occasions from the endemic areas in the central highlands (29).

After Madagascar, the DRC has notified most cases of plague in Africa (**figure 1.2B**), mainly present in remote areas of the Orientale province of the country, with sporadic outbreaks occurring in Ituri (14, 30, 31). Plague cases also recurrently occur on the other side of the DRC border, in the West Nile region in Uganda (12, 32, 33). These dispersed and localized, yet stable plague endemic foci in DRC and neighboring Uganda appear to be within the oldest known focus in the African continent and probably have a common origin shared with other Central and Eastern African plague endemic areas in Kenya, Tanzania, and Zambia (see **section 2.2.2**). In the 1990s, human plague cases rose in several African countries including Madagascar, DRC, Tanzania, Mozambique, Uganda, Zambia and Zimbabwe (34–38), which may be explained by either increase in plague activity and improvements in notifications to the WHO in areas that were less monitored in the past despite being endemic for plague (16). In the last years, plague has also been reported in Northern Africa. In 2003, Algeria notified a small outbreak (23, 39), as it occurred later in 2008 (40). In 2009, neighboring Libya also declared a limited plague outbreak near Tobruk, raising the question of a possible expansion of the plague focus in neighboring Algeria, which was active the previous year. Molecular typing of *Y. pestis* isolates from Algerian and Libyan outbreaks revealed that two different genotypes circulated respectively in each of the countries (41), suggesting that plague foci in these two countries of the Maghreb region are independent.

Plague in Asia. Plague endemic foci are also active in several Asian countries. China has experienced human plague cases almost every year in the last decades. The Inner Mongolia Autonomous Region natural focus has been active in the recent years, with bubonic and pulmonary plague cases occurring in 2020 and 2019 (42). This raised concerns when the first confirmed case of pulmonary plague was imported from this area to the major city of Beijing

(43). The small outbreak, with 4 cases, was related to the contact with an infected hare in the Inner Mongolian plague focus, mostly characterized for the presence of gerbils *Meriones unguiculatus* serving as a major reservoir of animal plague (44, 45). The Yunnan province was the birthplace of the expansion of plague around the world by the second half of the 19th century (46). Outbreaks of human plague have been reported in Yunnan until 2016 related to the exposure to infected house rats (*Rattus flavipectus*), that are main reservoirs of animal plague in Yunnan (47). It is however the Qinghai-Tibet plateau the largest animal plague focus in terms of extension present in China (17). This vast area is the one at highest risk of human plague in China, where *Marmota himalayana* marmots are considered major reservoirs of animal plague. Several outbreaks in human have been reported in the area, that were associated to the contact with infected Tibetan sheeps (*Ovis aries*) that appear to be important intermediate hosts between reservoir marmots and humans in this focus (48). Marmots are indeed important animal reservoirs of plague in Asia. In Mongolia, two fatal cases of plague occurred in 2019 associated with the consumption of raw marmot meat and organs, being one of few cases recorded of *Y. pestis* infection via the oral route (49). Rodent surveillance in Mongolia revealed a remarkable prevalence of *Y. pestis*, notably in the western regions such as Bayan Ulgii (49), whereas most of recorded cases of human plague appear to be traced to contact with infected marmots (15, 50). A similar situation occurs in neighboring Russia (Gorno-Altai and Tuva mountains) and Kyrgyzstan (Tien Shan, Alai and Talas mountains), where cases of human plague have also been traced to marmots (51, 52).

Despite the decline observed since the entry into the 21st century, plague foci were still very active in South-East Asia by the decade of 1990s (**figure 1.2B**) (8). The southern provinces of Vietnam have experienced important epidemics in the past and they are considered endemic regions of the disease (notably the Central Highlands region). In the 1960s, human plague cases in Vietnam accounted for around 90% of plague incidence worldwide (22). No cases of human plague have been recorded in Vietnam since 2002, and no *Y. pestis*

could be isolated between 2001 and 2006, suggesting that the plague focus in Vietnam may have entered a period of silence (18). Like the Madagascar endemic foci, plague in Vietnam is driven by the dynamics of rat populations (*R. rattus*), that act as reservoirs, and that of the oriental flea (*X. cheopis*), that act as major vector of the disease (18). After important epidemics early in the 20th century, plague re-emerged in India in 1994 causing 876 cases and 54 deaths (22). Affecting mostly the city of Surat in the Gujarat state, the outbreak of pneumonic and bubonic plague caused panic into the local population when nearly 500 000 residents of Surat fled the city trying to escape the disease (53). An earthquake and important floods that occurred in the Gujarat region the previous year are believed to have disturbed the equilibrium of rodent populations in this plague focus, which may then have created the conditions for zoonoses to neighboring human populations in the 1994 outbreak (22).

Plague in the Americas. In the Americas, a total 1073 human plague cases have been notified in the last 25 years by Peru, the United States, Brazil and Bolivia (8, 22). Peru dominates the plague picture in the Americas, recording plague in humans almost every year. Reservoirs of animal plague in Peru include guinea pigs (*Cavia porcellus*) and black rats (*R. rattus*) (54). In 2010, an outbreak of four pneumonic plague cases (two of them of nosocomial origin) were recorded in La Libertad region (55), where a plague outbreak had already occurred the previous year (56). Vast areas in the Western part of the United States (US) are considered endemic for plague, including the states of New Mexico, Arizona, Colorado, California, Oregon and Nevada, where most plague cases in humans are sporadic and appear in the bubonic form (57) (see the [CDC Maps and statistics](#)). Although the maintenance of plague in animal reservoirs in North America is not fully understood, plague activity is usually documented in mammals such as ground squirrels (*Spermophilus spp.*) or prairie dogs (*Cynomys spp.*), mediated by fleas (*Oropsylla spp.*) as main vectors in the US (58). The north-eastern coast of Brazil (regions of Alagoas, Bahia, Ceara, Minas Gerais,

Paraíba, Piauí, Pernambuco and Rio Grande do Norte) is endemic for plague and experienced human plague cases until the very end of the 20th century (58). Plague in Brazil is limited to foci in rural areas, where despite the restricted human plague cases declared in the recent years, serological surveys indicate that these foci are still active (59). These foci would be sustained by endemic circulation of *Y. pestis* within black rats (*R. rattus*) and local wildlife rodents such as *Necromys lasiurus*, *Calomys expulsus* and *Monodelphis domestica*. (60).

Plague in Europe. Europe has not experienced human plague cases since the 1940s, that corresponded mostly to outbreaks associated with introductions of the disease in major ports during the 3rd plague pandemic (61). If plague disappeared from Western and Central Europe, with no known sylvatic animal reservoirs to date, human and animal plague had a different story in Eastern Europe, specifically in the Caucasus (62). The Caucasus and Northern Caspian regions in present day Russia have a long-standing plague epidemic history due to the establishment of stable plague foci in the area. Therefore, Russia has recurrently experienced outbreaks early in the 20th century, and live *Y. pestis* bacteria have been isolated from infected animals up to 2015, suggesting that animal plague foci in these areas in the Caucasus are still active (63, 64). These foci appear to be related to a diverse range of animal species including gerbils (*Meriones meridianus*, *M. tamariscinus*, *Rhombomys opimus*) and the small souslik (*Spermophilus pygmaeus*), and main vectors are fleas *Neopsylla setosa* and *Citellophilus tesquorum* (64).

Plague reemergence. A remarkable feature of plague is the observed pattern of reemergence of the disease in humans after long periods of silence in endemic countries. Plague is indeed considered a reemerging infectious disease since 1996 (65, 66) because of its natural animal reservoirs and epidemic history. Known endemic areas of plague around the world are therefore at risk for reemergence of the disease, where it represents an important threat for public health (24). Plague reemergence in endemic areas also implies the risk of spread

of the disease into areas that had never suffered from plague in the past. Several countries have experienced a reemergence of human plague in the recent years: India in 1994 (after 30 years without declared cases), Algeria in 2003 (after > 50 years), Libya in 2009 (after 25 years) and Yunnan province in China in 2016 (after 10 years). Some of these events, such as the 1994 outbreak in Surat, India, led to important social impact due to episodes of panic, as well as an estimated 260 million dollars loss related to measures imposed to control the outbreak (see [the Surat Plague and its Aftermath](#)). This highlights the important social and economic cost associated to plague epidemics, notably in developing countries, which calls for efficient and regular efforts to monitor and control the disease. The factors driving the dynamics of *silent* periods and subsequent reemergence with spillover to human populations are still major questions in plague epidemiology and highlight the complex ecological cycle of plague (**see section 1.2.**). This cycle brings together the dynamics of rodent reservoir and vector populations, which in their turn are influenced by climate and other ecological variations, and finally human factors (1, 67–69).

Human socio-cultural and economic factors shall also be considered, as they can influence the re-emergence of plague and its spread at local levels. Human plague represents at the present-day a disease related to poverty. Present-day bubonic plague outbreaks occur mostly in rural and remote areas, where ancient traditions and beliefs are still maintained (70). It is indeed in those remote and sometimes politically disrupted regions in endemic countries such as the Democratic Republic of Congo, Uganda, Tanzania or Madagascar where human communities still live in simple mud houses, exposed to the contact with sylvatic and domestic rodents, their parasites and pathogens they carry. The control of human plague in endemic areas requires so global health approaches, including improved diagnostics tools and case management, the development of prophylactic treatments (i.e.: vaccines) and a thorough understanding of both human cultural practices and of the host-vector-pathogen interplay (71).

Section 1.2

Reservoirs, vectors and transmission cycle**1.2.1. Hosts & animal reservoirs of plague**

The plague bacillus, *Y. pestis*, can be isolated in nature from different sources. More than 300 mammal species and even three bird species can serve as hosts and act as reservoirs for plague (72)(figure 1.3). Most of these species are of the order of the rodents, whereas several species of carnivores, lagomorphs, insectivores, and ungulates can also be infected with *Y. pestis*. Regarding birds, known hosts are wheatears of the *Oenanthe spp.* genus (73–75).

This incredibly diverse set of animal species that can act as hosts of plague also represents a wide range of susceptibility of the different species to *Y. pestis* infection, hence contributing differently to the dynamics of maintenance and transmission of plague around the world in reservoirs. If some animals appear to be highly sensitive (73), many mammals appear to have some degree of resistance to *Y. pestis* infection, such as particular species of marmots, rats, gerbils, grasshopper mice, voles, dogs, raccoons and badgers (2, 73). The degree of resistance of these species to *Y. pestis* infection may reflect their long-term exposure and species coexistence with *Y. pestis* in wildlife (76). For instance, it has been reported that in Madagascar, black rats (*R.rattus*) from plague endemic areas exhibit high resistance to *Y. pestis* infection compared to rats captured from non-endemic areas (26) and that this resistance could be transferred to the offspring (77). Dogs appear to be relatively plague-resistant and have been shown to develop antibodies against *Y. pestis*, although they can eventually develop primary pneumonic plague and hence contaminate humans (78, 79). In contrast, some carnivores such as cats appear to be highly susceptible to *Y. pestis* infection, which may help explaining the important pneumonic plague transmission from ill cats to humans (80, 81). Gunnison’s prairie dogs are also highly susceptible to plague,

being frequently found naturally infected in the United States (82, 83). Carnivores may be infected by *Y. pestis* when ingesting contaminated rodents or by the bites of fleas laying on their preys (84). In Northern Africa, Middle East and other Asian areas, camels have been reported to be hosts of *Y. pestis* after exposure to contaminated rodent carcasses or their excrement (3), although flea-borne transmission to camels has also been proved experimentally (85). Consumption of raw camel organs of infected individuals can therefore lead to the development of plague in humans (86), comparable to what has recently been observed with consumption of raw organs of plague-infected marmots in Asia (49).

Rodent species		Other animal hosts		Flea vectors	
<i>Acomys</i> spp.	<i>Dipus sagitta</i>	<i>Mus</i> spp.	<i>Tachyoryctes</i> spp.	<i>Monodelphis</i> spp.	<i>Odontopsyllus</i> spp.
<i>Aethomys</i> spp.	<i>Echimys</i> spp.	<i>Mylomys</i> spp.	<i>Tamias</i> spp.	<i>Mustela</i> spp.	<i>Opisocroslis</i> spp.
<i>Akodon</i> spp.	<i>Eligmodon</i> spp.	<i>Mystromys albicaudatus</i>	<i>Tamiasciurus</i> spp.	<i>Odocoileus</i> spp.	<i>Opisodasys</i> spp.
<i>Allactaga</i> spp.	<i>Ellobius</i> spp.	<i>Neotamias</i> spp.	<i>Tartera</i> spp.	<i>Oenante oenante</i>	<i>Ochopeas</i> spp.
<i>Alticola</i> spp.	<i>Funambulus</i> spp.	<i>Neotoma</i> spp.	<i>Thallomys</i> spp.	<i>Oryctolagus</i> spp.	<i>Oropsylla</i> spp.
<i>Ammospermophilus</i> spp.	<i>Galea</i> spp.	<i>Ochotona</i> spp.	<i>Thomomys</i> spp.	<i>Ovis</i> spp.	<i>Athyphloceras</i> spp.
<i>Apodemus</i> spp.	<i>Gerbilliscus</i> spp.	<i>Oenomys</i> spp.	<i>Thrichomys</i> spp.	<i>Pekaniapennanti</i> spp.	<i>Callopsylla</i> spp.
<i>Arvicola</i> spp.	<i>Gerbillus</i> spp.	<i>Olygoryzomys</i> spp.	<i>Xerus</i> spp.	<i>Presbytis</i> spp.	<i>Catallagia</i> spp.
<i>Atlantorexus getulus</i>	<i>Glaucomyus</i> spp.	<i>Oryzomys</i> spp.		<i>Procyon</i> spp.	<i>Ceratophyllus</i> spp.
<i>Bandidocota</i> spp.	<i>Golunda</i> spp.	<i>Otomys</i> spp.		<i>Puma</i> spp.	<i>Chiastopsylla</i> spp.
<i>Bathyergus</i> spp.	<i>Grammomys</i> spp.	<i>Otospermophilus</i> spp.	<i>Antilocapra americana</i>	<i>Spilogale</i> spp.	<i>Citellophilus</i> spp.
<i>Bolomys</i> spp.	<i>Graomys</i> spp.	<i>Parotomys</i> spp.	<i>Bassariscus astutus</i>	<i>Suncus</i> spp.	<i>Coptopsylla</i> spp.
<i>Brachytarsomys</i> spp.	<i>Graphiurus</i> spp.	<i>Pedetes</i> spp.	<i>Camelus</i> spp.	<i>Suricata suricatta</i>	<i>Craneopsylla</i> spp.
<i>Callospermophilus</i> spp.	<i>Hesperomys</i> spp.	<i>Pelomys</i> spp.	<i>Canis</i> spp.	<i>Sus</i> spp.	<i>Ctenocephalides</i> spp.
<i>Calomys</i> spp.	<i>Holochilus</i> spp.	<i>Perognallms</i> spp.	<i>Capra</i> spp.	<i>Sylvilagus</i> spp.	<i>Diamanus montanus</i>
<i>Cardiocranius</i> spp.	<i>Jaculus</i> spp.	<i>Peromyscus</i> spp.	<i>Cervidae</i>	<i>Sylvisorex</i> spp.	<i>Dinopsylla</i> spp.
<i>Cavia</i> spp.	<i>Kerodon</i> spp.	<i>Phyllotis</i> spp.	<i>Crociodura</i> spp.	<i>Syncerus</i> spp.	<i>Eumolpianus eumolpi</i>
<i>Caviella</i> spp.	<i>Lagostomus</i> spp.	<i>Praomys</i> spp.	<i>Cynictis penicillata</i>	<i>Urocyon</i> spp.	<i>Xenopsylla</i> spp.
<i>Cerradomys</i> spp.	<i>Lagurus</i> spp.	<i>Psammomys</i> spp.	<i>Didelphis</i> spp.	<i>Ursus</i> spp.	<i>Foxella</i> spp.
<i>Cricetomys</i> spp.	<i>Lasiopodomys</i> spp.	<i>Reithrodontomys</i> spp.	<i>Equus</i> spp.	<i>Vulpes</i> spp.	<i>Frontopsylla</i> spp.
<i>Cryptomys</i> spp.	<i>Lemniscomys</i> spp.	<i>Rhabdomys</i> spp.	<i>Erinaceus</i> spp.		<i>Hectopsylla</i> spp.
<i>Ctenomys</i> spp.	<i>Lophuromys</i> spp.	<i>Rhipidomys</i> spp.	<i>Felis</i> spp.		<i>Hoplopsyllus</i> spp.
<i>Cynomys</i> spp.	<i>Marmota</i> spp.	<i>Rhombomys</i> spp.	<i>Lama glama</i>		<i>Hystrichopsylla</i> spp.
<i>Dasymys</i> spp.	<i>Mastomys</i> spp.	<i>Saccostomus</i> spp.	<i>Lepus</i> spp.		<i>Listropsylla</i> spp.
<i>Dendromys</i> spp.	<i>Meriones</i> spp.	<i>Scandentia</i> spp.	<i>Loxodonta</i> spp.		<i>Malaraeus</i> spp.
<i>Desmodillus</i> spp.	<i>Mesocricetus</i> spp.	<i>Sciurus</i> spp.	<i>Lynx</i> spp.		<i>Megabothris</i> spp.
<i>Dipodomys</i> spp.	<i>Microcavia</i> spp.	<i>Sigmodon</i> spp.	<i>Macaca</i> spp.		<i>Megathroglossus</i> spp.
	<i>Micromys</i> spp.	<i>Spermophilus</i> spp.	<i>Martes</i> spp.		<i>Meringis</i> spp.
	<i>Milardia</i> spp.	<i>Steatomys</i> spp.	<i>Mephitidae</i> spp.		<i>Monopsyllus</i> spp.
			<i>Mephitis</i> spp.		<i>Neopsylla</i> spp.
					<i>Neothiphloceras</i> spp.
					<i>Megathroglossus</i> spp.
					<i>Opisocroslis</i> spp.
					<i>Opisodasys</i> spp.
					<i>Ochopeas</i> spp.
					<i>Oropsylla</i> spp.
					<i>Parapsyllus</i> spp.
					<i>Pleochaetis</i> spp.
					<i>Polygenis</i> spp.
					<i>Psocopsylla</i> spp.
					<i>Pulex irritans</i>
					<i>Rhadinopsylla</i> spp.
					<i>Stivalius</i> spp.
					<i>Synosternus</i> spp.
					<i>Tiamastus</i> spp.
					<i>Tritopsylla</i> spp.
					<i>Xenopsylla</i> spp.

Figure 1.3. Schematic list of documented animal reservoirs, intermediate hosts and ectoparasite vectors described in the literature.(2, 72)

1.2.2. Vectors and vector-borne transmission of plague

The transmission of plague within animal reservoirs can be achieved through their associated ectoparasites that act as vectors of *Y. pestis* transmission. About 80 different species of fleas are known vectors of plague in rodents (87–89) (figure 1.3). Indeed, fleas are considered to play a pivotal role in the transmission and maintenance of plague in nature (90). The discovery of the role of fleas in the transmission of plague came slightly after the discovery of

Y. pestis as its etiological agent and lead for the first time to a model for the understanding of plague vector-borne transmission. Paul-Louis Simond, a medical doctor sent by the Institut Pasteur, observed plague bacilli in rat flea digestive tract in 1898 India (91). Almost simultaneously, Masanori Ogata (colleague of Kitasato), suspected the same role of fleas for plague transmission and could successfully infect mice using fleas that had been feeding on an infected rat (92). Simond pursued his research and performed an experiment in which he placed a naïve rat in a cage close to a dying infected rat covered with fleas, that he had found in the house of a plague victim. It appeared that the naïve rat died 5 days after the death of the infected rat, all without allowing any direct contact between both animals. He therefore concluded that fleas were the missing link that conciliated the transmission of plague between infected rats, from rats to humans and between humans (91), a discovery that made him feel, in his own words: “a baffling emotion when thinking that I had just violated a secret that has been worrying humanity since the appearance of plague in the world”.

Similar to what seems to occur for hosts of *Y. pestis* infection, different species fleas and other ectoparasites also contribute differently to the maintenance and circulation of plague in nature. Consequently, some species may be considered as *truly* vectors of plague if they are i) parasites of major plague hosts ii) present stably and abundantly in populations, even along different periods of the year iii) highly mobile, migrating within individuals and host populations iv) frequently feeding from plague hosts allowing a stable level of transmission and if v) *Y. pestis* has been highly adapted to colonizing the digestive tract of the particular insect species (72). This definition may however be too limited to completely describe the interaction between fleas and the highly flea-adapted plague bacillus. Soviet scientist Vladimir Stepanovich Petrov argued indeed that considering fleas *only* a vector could not fully explain their role on the plague transmission cycle (93). He highlighted that the total time that *Y. pestis* resides inside fleas was actually much higher than the time the bacillus

resides inside the bodies of infected mammals. Moreover, a long-term maintenance of the pathogen in fleas would be necessary to preserve bacteria during periods when transmission chances decrease dramatically such as during hibernation months in mammals such as marmots. This interesting view would consequently attribute to fleas not only a role as vectors of *Y. pestis* transmission, but they could be interpreted as true reservoirs of plague too.

Y. pestis appear to be highly adapted to the colonization of the flea digestive tract, where it will remain confined during the entire course of infection. This tract is divided into three segments: the esophagus with its suction pump and the proventriculus valve, the midgut or stomach, and the intestines and rectum of the insect (5). Two major modes of flea-borne transmission have been characterized. The first one is known as early-phase transmission and occurs within the first hours to the first few days after an infectious blood meal (94), when fleas are immediately infectious and hence can transmit *Y. pestis* to another animal during its following blood meal (95). This mode of transmission appears rarely unless several fleas are feeding from the same individual and transmission efficiencies are low, depending on the flea species (96) being from ~0-10% depending on the bacterial load in each infectious blood meal for fleas such as *X. cheopis*(97), *Oropsylla hirsuta* (98) and *O. montana* (94). The second mode of transmission is known as “biofilm-dependent transmission” and occurs after the formation of a cohesive bacterial biofilm by *Y. pestis* in the proventricular valve of the flea midgut within 1 to 3 days after ingestion (5, 96). As *Y. pestis* divides, the biofilm grows until occluding the proventricular valve blocking it, preventing the flea to feed and ultimately leading to starvation. In this situation, the aspirated blood comes into contact with *Y. pestis* present in the biofilm blocking the proventriculus, contaminating the blood meal. The blocked flea will not be able to swallow the meal and therefore will regurgitate it at the same biting site, transmitting *Y. pestis* bacteria present within the contaminated aspirated blood. The blockage of the flea by *Y. pestis* biofilm dramatically changes the feeding behavior of the fleas in the benefit of bacilli to enhance their transmission: a blocked

flea will then spend the following days trying desperately to feed in repeated occasions, each of them representing a potentially new transmission of *Y. pestis*. The efficiency of transmission of the biofilm-dependent mode appears to be much higher than the one of early-phase transmission. Blocked *X. cheopis* may transmit *Y. pestis* in 25-50% of their bites (5), therefore representing increased chances of transmitting the disease to a healthy individual. For transmission to occur, a minimum infection threshold in the mammal host is required. A bacteremia level in the infected host of about 10^7 bacteria/mL would be necessary to infect 30 to 50% of fleas (97, 99). This important bacteremia, causing high mortality, appears to be necessary for *Y. pestis* transmission back to fleas, that is why it is thought that the hypervirulence characterizing *Y. pestis* may actually serve to compensate for the poor vector efficiencies of most flea species (100). The evolutionary adaptation of *Y. pestis* to the colonization of the flea midgut has been the source of extensive research efforts (see **section 2.1**).

1.2.3. Other sources of *Y. pestis* in nature

Besides its circulation within the mammal-ectoparasite transmission cycle in plague endemic areas, *Y. pestis* has also been isolated from other sources in nature. Several observations and experiments suggest that *Y. pestis* can persist in the soil for a certain period and under certain conditions. This was hypothesized as early as 1894 in Hong Kong, when Alexandre Yersin isolated plague bacilli from the soil in a house of plague victims (101). The hypothesis of contaminated soils as a source of *Y. pestis* was also explored by Henri Mollaret, who observed in an experiment the survival of *Y. pestis* up to 16 months in a contaminated soil (102). In a more recent experiment, *Y. pestis* was found to persist in soil up to 280 days while maintaining full virulence fitness in subsequent mice infections using the contaminated soil (103). On the field, several *Y. pestis* isolates have been recovered from soil in different natural plague endemic areas such as Iran, from a burrow that hosted animals that died from plague 11 months before (104). In Arizona, a *Y. pestis* isolate was recovered from the

soil under the carcass of a mountain lion 3 weeks after its death (95). More recently, another *Y. pestis* isolate (Algeria3) was obtained when analyzing saline soil samples collected in an endemic area in Algeria (20). Related to the persistence of *Y. pestis* in the soil, it has been hypothesized that plague bacilli may be maintained in the soil by surviving intracellularly in protozoa such as amoebas (105). The survival of *Y. pestis* and its intracellular replication for 2-5 days in *Dictyostelium discoideum* and *Acanthamoeba castellanii* trophozoites has been recently described *in vitro* (106, 107). The persistence of *Y. pestis* in the soil, either as free organisms or living inside protozoan cells, may allow the infection of burrowing rodents via the aerosol route or through cutaneous wounds. This hypothesis would provide new perspectives for the understanding of the sustainment of plague in endemic areas as well as during periods in which the activity of animal reservoirs and their vectors is limited (103). However, the possible contribution of soil and protozoa for the maintenance of *Y. pestis* in nature to serve as a reservoir for the cycle of plague still remains subject of debate, requiring field work and long-term surveillance of soils in endemic areas of plague.

Section 1.3

Transmission cycle and plague ecology**1.3.1. Transmission within the animal reservoir**

The understanding of the ecological cycle of plague is decisive for the development and application of effective plague control measures in endemic areas. In endemic and non-endemic areas, plague can be related with the animal reservoir in two types of state: enzootic and epizootic (**figure 1.4**). The *enzootic* state refers to the *Y. pestis* circulation in endemic areas, where the disease persists stably in communities of partially resistant rodents that act as reservoirs in the rodent-flea infection cycle. On the other hand, in the *epizootic* state *Y. pestis* circulates into moderately to highly susceptible mammals. For the development of epizootics, it is therefore necessary that more susceptible hosts are present, so that important bacteremia occur in infected mammals, and that an important number of fleas eventually become infected to sustain the epizootic (70, 72). Epizootics can occur when *Y. pestis* is introduced into communities of susceptible rodents in regions not endemic for plague, leading to important rodent mortality (73). In endemic areas, epizootic periods may alternate with enzootic periods. This alternance dynamics between enzootic and epizootic circulation of *Y. pestis* is influenced by multiple factors and is not yet fully understood (67, 70, 108). Perhaps the most commonly accepted hypothesis to explain the alternance between enzootic and epizootic cycles remains the *trophic cascade hypothesis*, which is based on a typical climatic model. According to the trophic cascade hypothesis, periods of exceptional precipitation induce a subsequent increase in plant growth, representing in a more abundant food source for rodents that can be epizootic hosts for plague. Population density of these rodents increases in consequence and so do the populations of their fleas and other parasites. When this period of abundance is followed by a period of severe drought, food becomes scarce and rodent populations may starve and debilitate, hence be more susceptible to infection

by *Y. pestis*. Therefore, a much higher number of rodents may develop septicemic plague infection and die. Fleas from these individuals, infected with *Y. pestis*, will then need to switch onto the survivors seeking for new sources of food, thereby perpetuating the process of epizootic and susceptible rodent death (72). Besides the vector-borne transmission of plague between rodent communities, the capacity of *Y. pestis*-infected soil and flea feces for infecting rodents or fleas through the oral or aerosol route has not been fully explored (2). In conclusion, the series of factors driving the dynamics of persistence and ecological cycling of plague in natural foci remain fundamental questions to be answered in plague epidemiology.

1.3.2. Transmission of plague in human populations

Vector-borne transmission. Due to the increased *Y. pestis* transmission during epizootics, the risk of zoonotic transmission of *Y. pestis* to neighboring human populations is higher during epizootic periods (**figure 1.4**). In humans, plague is a zoonotic disease that may be contracted directly through animal bites, their carcasses, the consumption of raw infected meat and organs, or indirectly mediated by ectoparasites acting as vectors (73, 74). Handling of infected carcasses of marmots, rats, mountain lions, camels, Tibetan sheeps, guinea pigs and dogs has been reported as sources of plague infection (48, 91, 109–112). Infection through contact with live or dead infected animals may occur via the inhalation of infectious aerosols, leading to pneumonic infection, or through breaches on the skin, leading to bubonic plague (111, 113, 114).

In the recent year, plague transmission through the consumption of contaminated meat and organs has also been reported in Asia and Northern Africa (86, 115). When epizootics affect rodents living in urban areas associated to human communities (peridomestic rodents), transmission from their associated fleas to humans becomes more frequent, leading to the development of bubonic plague outbreaks that may remain limited. The oriental rat flea, *X. cheopis*, is considered a major vector of plague epidemics in humans since the late 19th century, notably in tropical and subtropical areas (2). In addition to fleas, some authors point

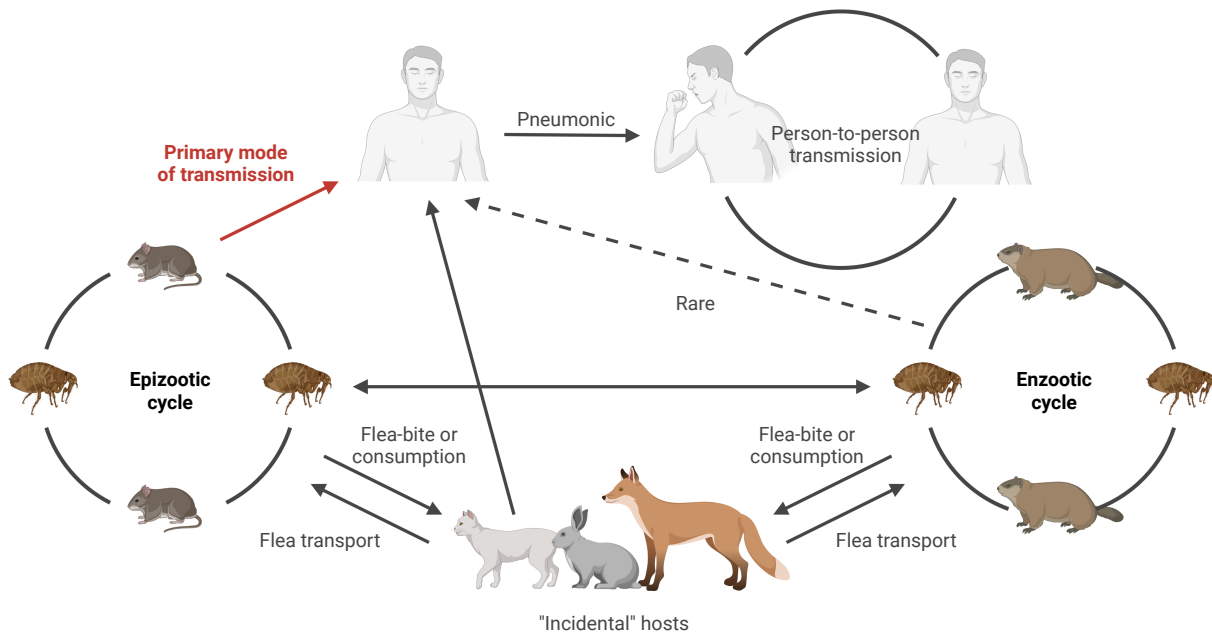


Figure 1.4. Natural cycle of plague and pathways of transmission of *Y. pestis* within the animal reservoir, intermediate hosts and spillover to humans. This figure was created by my fellow Mara E. Carloni.

to the possible contribution of human ectoparasites to the human-to-human transmission of plague, although this hypothesis remains disputed. The capacity of body lice (*Pediculus humanus*) to be infected by *Y. pestis* has occasionally been described on the field (30), and their ability to transmit *Y. pestis* to animal hosts has been observed experimentally (116, 117). Human fleas (*Pulex irritans*) have also been studied as a possible vector of *Y. pestis* infection in humans, after finding *P. irritans* in houses of plague victims of several endemic countries for plague (118), and the transmission capacity of *P. irritans* evaluated in animals with poor efficiency (119). The possible contribution of human ectoparasites on plague transmission in historical pandemics has been subject of discussion (2, 120).

Aerosol transmission. *Y. pestis* is highly infectious via the aerosol route, and inhalation of infectious droplets leads to the development of primary pneumonic plague, the most severe form of plague (see section 1.4.) (121). Due to its nature, this form of plague can

be transmitted between humans through the aerosol route. Survival of *Y. pestis* on surfaces such as steel or glass appears to be limited to less than 72h (122). Interhuman transmission of pneumonic plague thus appears to require close contact with a pneumonic plague patient (≤ 1 m) (10, 123). This situation is likely to occur in familiar context or in shared transports, for instance (11, 27). Airborne transmission of plague has been historically associated with major pneumonic plague epidemics in Manchuria (present-day China) in 1910-11 (124, 125). More recently, a pneumonic plague outbreak occurred in Madagascar affecting major urban areas (11). It has been estimated that a pneumonic plague patient may infect 1.3 healthy individuals on average (126). The impact of pneumonic plague in the interhuman plague transmission during historical pandemics remains difficult to assess (2, 127), but it may be one of the factors influencing the reemergence capacity of plague in endemic countries such as Madagascar (121).

Section 1.4

Clinical signs and pathology of plague

The clinical manifestations of plague in humans vary according to the infectious route. It can be generally considered that symptoms develop after an incubation period of 2-7 days, with patients initially experiencing unspecific, flu-like symptoms such as sudden fever, headache, chills, weakness, nausea and vomiting (70, 128). Humans are particularly sensitive to *Y. pestis* infection, as individuals of any age and sex remain susceptible (129). Besides these general symptoms, specific symptoms take on several forms that lead to the main clinical presentations of the disease, being bubonic, septicemic, and pneumonic plague.

1.4.1. Bubonic plague

Bubonic plague is the most prevalent clinical form of plague since flea-borne transmission of *Y. pestis* to humans is the most frequent route of transmission of plague to humans (**figure 1.5**). This form develops after the bite of one or several infected fleas and has 40-60% mortality rates if left untreated (73). Classical symptoms of bubonic plague develop after an incubation period of 2-7 days and include the adenopathy of lymph nodes draining the site of inoculation. The swollen, enlarged, painful lymph nodes are known as *buboes*, hence giving the name to this commonly known form of plague. Buboes develop most frequently in femoral and inguinal lymph nodes, but they can also develop in the axillary and cervical nodes (74). Buboes developing in the neck (cervical nodes) appear to correlate with the development of severe symptoms with poor outcome for the patient. What distinguishes bubonic plague from adenopathy due to other causes is the fact that plague is the only disease characterized by the development of buboes that is associated with important epidemics. Moreover, the rapid onset of plague and swelling buboes is another specific feature of bubonic plague (128). If bubonic infection progresses without treatment, a patient may then develop a systemic infection (septicemic plague) that can lead to pulmonary infection and development

of secondary pneumonic plague (80).

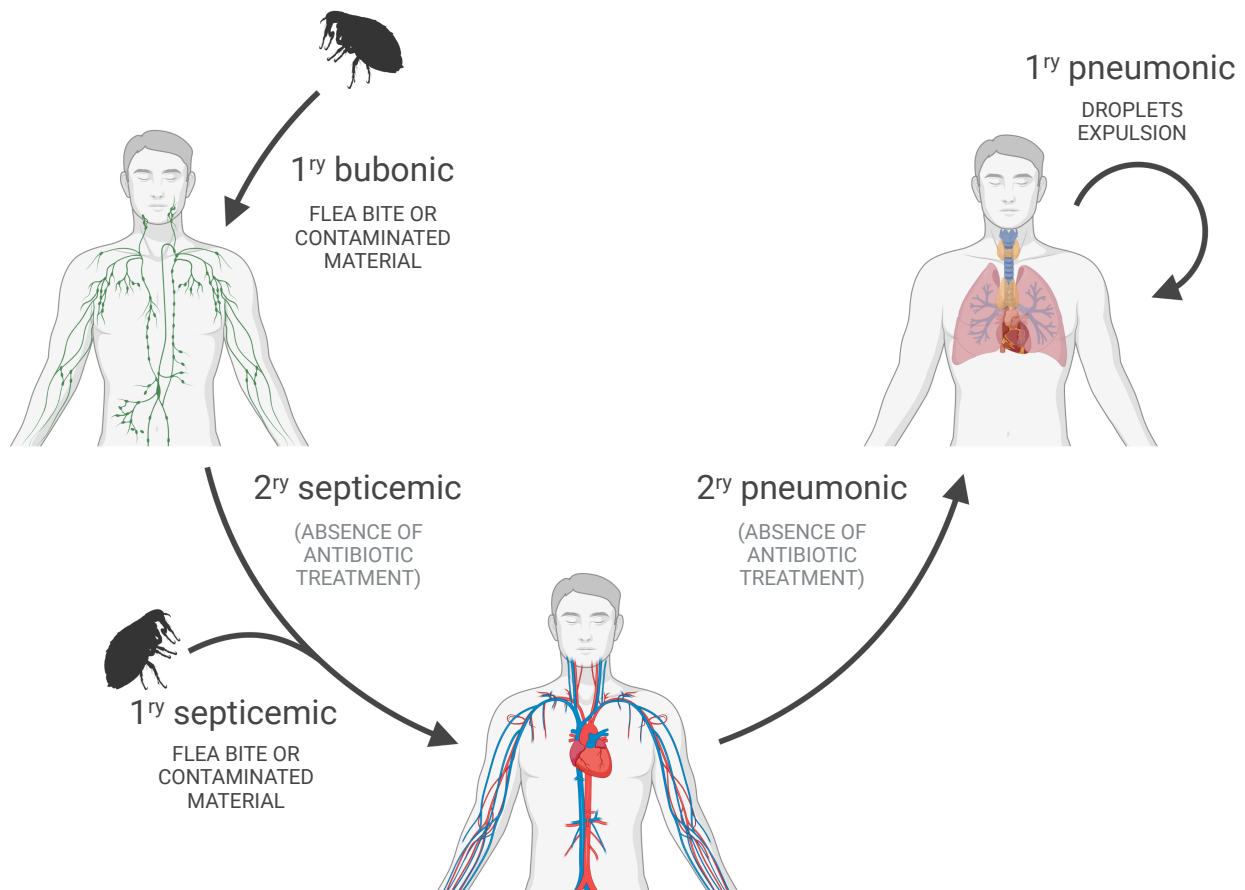


Figure 1.5. Schematic representation of the three clinical manifestations of plague and their relationships. This figure was created by my fellow Pierre L -Bury.

1.4.2. Septicemic plague

Septicemic plague occurs when *Y. pestis* reaches the bloodstream directly, resulting either from a flea bite, direct infection through breaks on the skin, or through the progression of the adenopathic infection in a bubonic plague patient. This form of plague is characterized by a septic shock and patients may develop bleeding from the skin and mucous membranes, internal hemorrhages, and peripheral gangrene due to blood coagulation. Positive culture of *Y. pestis* from blood samples can help diagnose septicemic plague. The onset of septicemic infection by *Y. pestis* is fast and patients may die in less than 24 h if not given adequate antibiotic treatment (70, 128).

1.4.3. Pneumonic or pulmonary plague

Pneumonic plague is the most virulent and lethal form of plague, although it is also less frequent. Primary pneumonic plague develops after exposure to infected droplets of a coughing patient of pneumonic plague, and secondary pneumonic plague occurs as an outcome of the systemic dissemination of *Y. pestis* to the lungs after bubonic or septicemic plague (2). The incubation period in primary pneumonic plague is shorter than in bubonic plague, generally 2 to 4 days after exposure. Pneumonic plague can also result from laboratory-acquired infections due to exposure to high bacterial loads in aerosols, which may only show an incubation period of few hours. Pneumonic plague manifests as a bronchopneumonic illness and includes cough, chest pain, fever, scant sputum and hemoptysis. It is estimated that 8-10% of bubonic plague patients develop pneumonic plague if left untreated (128). This form of plague shows the highest fatality rates of *Y. pestis* infections, of virtually 100% but may be reduced to 25% if antibiotic treatment is given within the first 24 h after the onset of respiratory symptoms (74, 121). As mentioned earlier, pneumonic plague is the only form of plague that is human-transmissible and can therefore lead to the emergence of localized outbreaks, notably in urban environments (130). One of the most intriguing aspects of *Y. pestis* infection is the ferocious transition from the absence of clinical symptoms in the incubation period, to the burst of inflammation and eventual fatal septicemic infection caused by this dreadful pathogen (131). A *pre-inflammatory phase* characterizes this early lag period in which *Y. pestis* travels towards the lymph node from the flea bite site (in bubonic plague) or infects the lungs (in pneumonic plague). This phase is characterized by the modulation of host's innate immune response by *Y. pestis* (132). Indeed, *Y. pestis* can survive within macrophages and can inhibit phagosomal maturation (133, 134). An important distinction between bubonic and pneumonic plague that may explain their differences in incubation period and disease onset is the environmental temperature in which bacteria multiplied before its transmission to a new host: for aerosol transmission acquired from a pneumonic patient, *Y. pestis* grew

around 37 °C within the lungs, while for bubonic plague *Y. pestis* grew < 30 °C in fleas that are the transmission vectors (121). Growth temperature regulates the expression of major *Y. pestis* virulence factors that promote the progression towards the inflammatory stage of infection. The *inflammatory phase* is characterized by systemic dissemination of *Y. pestis*, inducing tissue destruction and the consequent onset of clinical signs of disease in the host that follows the innate immune system response (131).

Section 1.5

Plague diagnostics

Early diagnostic of human plague is of critical relevance to begin antibiotic treatment as quickly as possible to prevent the risk of fatality associated to any of the three forms of plague. However, plague diagnostic is challenging in endemic areas of plague due to the need of trained personnel and biosafety equipment for the processing of samples. Diagnostic of plague can be established from clinical samples including bubo aspirates, respiratory samples (sputum), blood, lymph node biopsies, pharyngeal swabs and urine (70). Handling of clinical samples suspected of containing *Y. pestis* can be performed under biosafety level 2 environments (BSL-2), but all cultures and manipulation of *Y. pestis* cultures should be performed in a biosafety level 3 (BSL-3) environment (6). The gold standard for plague diagnostic requires the confirmation by *in vitro* culture and isolation of *Y. pestis* from the clinical sample. In 2021, the WHO published a revision of the diagnostic criteria for plague case definition (135). According to the WHO definition, a confirmed plague case requires either a positive culture of *Y. pestis*, or the detection of *Y. pestis* DNA using conventional polymerase chain reaction (PCR) targeting specific targets on the chromosome and plasmids of *Y. pestis* genome. It is important to note that for a plague case to be confirmed, the patient should show clinical features suggestive of plague, as well as the placement of the case in a wider epidemiological context (i.e.: in case of an outbreak, recent travel to an endemic area, or the exposure to infected animals).

Assessment of *Y. pestis* presence in diagnostic samples can be performed directly by microscopic observation, where adequately stained bacteria have a “safety pin” appearance. *Y. pestis* grows relatively easily *in vitro* when cultured in brain heart infusion, Luria-Bertani or MacConkey agars after 24 to 72 h incubation in aerobic conditions at 28 °C, which is its optimum temperature. The use of selective agar for *Yersinia spp.*, which is supplemented

with cefsulodin, irgasan and novobiocin (CIN) limits the growth of other bacteria that may be present in the complex clinical samples (70). After 48-72 h incubation, suspected colonies may be identified using biochemical tests or PCR diagnostic (131).

Of great interest to facilitate the quick diagnostic of plague in case of outbreak is the recent development of rapid diagnostic tests (RDT) that is based on the *Y. pestis* specific F1 capsular antigen. These tests can be read in just 15 minutes and has been validated for the diagnostic of bubonic plague with specificities and sensibilities reaching 100% (136). The tests, based on a solid immunochromatographic band, are produced at the Institut Pasteur de Madagascar and appear to be significantly more economic than other available molecular tests. However, the use of these F1 RDT assays has yet limited efficacy for the diagnostic of pneumonic plague due to the observed rates of false-positive and negative lectures from sputum samples collected during the 2017 pneumonic plague outbreak in Madagascar (131). Sample quality and processing, as well as the availability of trained personnel is critical to evaluate the efficacy of the use of RDT on the field in an epidemic situation. The WHO has emitted strong recommendations for the use of F1 RDT for the detection of bubonic plague in endemic areas, whereas its use for pneumonic plague detection in case of outbreak is only conditionally recommended (135). Several PCR-based protocols have been developed and allow for the detection of *Y. pestis* directly on bubo aspirates, blood, rodent biopsies and fleas (137–139). The amplification of *Y. pestis* DNA also permits the confirmation of isolated colonies from cultured samples. The amplification of both plasmid-encoded *pla* and *caf1* loci allows for the positive identification of *Y. pestis*, but in case of uncertain results, a third reaction targeting the chromosomal *inv* (1100 bp amplicon size specific to *Y. pestis*) locus can help providing positive identification of the pathogen (131). The *caf1* gene is considered specific to *Y. pestis*, whereas *pla* gene is also present in *Citrobacter spp.* and *Escherichia spp.* (140, 141). Portable, real-time PCR instruments have been developed and may be helpful for their use in remote endemic areas (142). Interestingly, rapid detection of

Y. pestis DNA using loop-mediated isothermal amplification (LAMP) assays have recently been developed, showing high sensitivity and specificity in human plague biological samples, including sputum from pneumonic plague patients ([143](#)).

Section 1.6

Plague treatment & antibiotic resistance

The extreme virulence of *Y. pestis* and the quick development of severe symptoms either in bubonic, septicemic or pneumonic plague is detrimental for the treatment of the disease. Owing to the fast, life-threatening evolution of plague, the administration of adequate antimicrobial treatment as early as possible is fundamental for the recovery of plague patients (73). The earliest efforts to treat plague patients in the era of microbiology date to Alexandre Yersin, who developed antiserum together with Calmette, Roux and Borrel at the Institut Pasteur in 1895 (2). When Yersin established in Vietnam, he dedicated himself to the obtention of plague antiserum to be used in local outbreaks occurring in the coast where he established a Pasteur institute (144). Antiserums were developed until antimicrobials were introduced for the treatment of plague, as for other infections caused by bacteria. For more than four decades, streptomycin was the antibiotic recommended against *Y. pestis* infection and is still being used in endemic countries such as Madagascar (70). The use of gentamicin alone or combined with tetracycline is an adequate substitute to streptomycin (145). The current antibiotics endorsed by the WHO and Centers for Disease Control (CDC) as being effective for the treatment of plague include common antibiotics against *Enterobacteriaceae*: doxycycline, streptomycin, gentamicin, levofloxacin, ciprofloxacin, moxifloxacin and chloramphenicol (146–149). In 2021, the WHO also emitted conditional recommendations on the use of fluoroquinolones as substitute to streptomycin, chloramphenicol and doxycycline for the treatment of bubonic, pneumonic and septicemic plague (150). Suspected bubonic plague patients should be placed in isolation until 2 days since the start of the antimicrobial treatment to prevent the possibility of airborne transmission if the patient develops pneumonic plague (73). Treatment should continue uninterrupted for 10 to 14 days and supportive therapy may be necessary in case of septic shock or gangrenes occurring in septicemic

plague (2).

The use of the above mentioned available antimicrobial treatments is generally effective in absence of infection with resistant *Y. pestis* strains. Antimicrobial resistance (AMR) does not yet represent a threat for plague treatment in endemic countries. This is illustrated by the absence of known resistance plasmids in more than 700 *Y. pestis* strains isolated in the United States (151), as well as by the absence of resistance phenotype against 8 common antibiotics in almost 400 *Y. pestis* strains collected from 17 countries (152). All the *Y. pestis* isolates associated with the 2017 unusual plague outbreak in Madagascar were susceptible to antibiotics commonly used to treat plague (11). However, the rapid expansion of AMR in bacterial species other than *Y. pestis* may represent a risk for the acquisition of resistances in the plague bacillus through its interaction with the available resistome increasingly present in the environment (153). In 1995, a *Y. pestis* strain was isolated from a patient in Madagascar that appeared to be resistant to eight antimicrobials, including those recommended for plague treatment (tetracycline, streptomycin, chloramphenicol), its prophylaxis (sulfonamide and tetracycline) as well as ampicillin, kanamycin and spectinomycin. The multiresistance was mediated by a transferable plasmid (154). A second and third isolates from Madagascar appeared to be resistant to streptomycin and doxycycline (155, 156), respectively. In all cases antimicrobial resistance was mediated by transferrable plasmids of 40kb, 150kb and 170 kb that were unrelated one to each other, suggesting independent acquisitions in *Y. pestis* strains from Madagascar. These reported horizontal acquisitions of AMR in *Y. pestis* may have occurred via conjugation with some enteric bacterial species within an infected host (rodent or human) or flea vector (154). Recently, a streptomycin-resistant *Y. pestis* strain was isolated from a patient in Tibet and showed a novel mechanism of streptomycin resistance based on a single amino acid substitution in the ribosomal gene *rpsL* (157). A second multi-drug resistant strain was isolated in Mongolia that showed resistance to seven antibiotics (chloramphenicol, tetracycline, sulfomethoxacole, gentamicin and doxycycline),

although the genetic basis conferring the resistance phenotype was not uncovered in the study (158). Although antibiotic resistant *Y. pestis* are rare and do not yet seem to represent an important threat for the treatment of plague, their sporadic presence strengthens the need for surveillance of AMR in *Y. pestis* isolated regularly in endemic areas of plague.

Section 1.7

Vaccination against plague

Despite the existence of effective antimicrobial treatments against plague, the endemic nature of plague exposes even some of the most remote areas in Africa, Asia, and the Americas to the risk of plague. In this context, prophylactic vaccination would be beneficial to communities inhabiting in endemic areas of plague, notably when the risk of antibiotic resistance is present (131). At present, there is yet no vaccine recommended by the WHO for the prevention of plague in endemic countries (2). The use of attenuated *Y. pestis* strains to confer immunity against plague dates back to the times of Alexandre Yersin, when he developed attenuated strains that protected mice (159). The first vaccine against plague that was widely used in humans is the live attenuated *Y. pestis* EV76 vaccine strain (160). EV76 vaccine was developed from a strain isolated by Girard and Robic in Madagascar in 1926 and appeared to protect against bubonic and pneumonic plague but also displayed important side effects, such as headache, malaise and pyrexia (161). A subculture of EV76 was established in the former Soviet Union in 1940, designated EV NIIEG vaccine, and has been used to protect plague researchers and inhabitants of plague endemic areas in the former USSR until present day (159). A killed whole-cell *Y. pestis* vaccine was produced in the United States and Australia but appeared to have limited efficacy and did not protect against pneumonic plague (131). Some of the most advanced candidates are subunit vaccines based on *Y.pestis* F1 (main component of its bacterial pseudocapsule) and V (essential component of a secretion system) antigens. Two doses of these F1-V vaccine candidates provide up to 100% protection against bubonic and pneumonic plague (162, 163). However, F1-V type vaccines appeared unable to protect African green monkeys against aerosolized *Y. pestis*, raising questions regarding its possible efficacy in humans (Smiley ST 2008). Recently, several more subunit vaccine candidates are being developed and may be administered through the intra-nasal route to

protect against pneumonic plague (164–167). A promising strategy for the development of plague vaccines is the use of live *Y. pseudotuberculosis*. A vaccine candidate based on a live attenuated *Y. pseudotuberculosis* producing *Y. pestis* F1 pseudocapsule and other antigens has been developed by the Yersinia Unit in the Institut Pasteur. The candidate, named VTnF1, can be administered through the oral route and provided protection against both bubonic and pneumonic plague in mice (168, 169).

Chapter 2

Yersinia pestis

Y. pestis is a pleomorphic, Gram-negative bacillus of 0.5-0.8 μm diameter and 1-3 μm long. It exhibits a bipolar, “hairpin” staining with Wright’s stain. The bacillus is a facultative anaerobe, nonmotile, nonsporulating and nonlactose fermenter, growing within 24 to 72 h between 4 to 40 °C, although its optimum growth is reached around 25-28 °C. On solid media, *Y. pestis* grows gray-white, translucent colonies of 1-2 mm diameter after 48 h incubation. Between 48 and 72 h incubation, colonies usually appear with a raised and irregular “fried egg” appearance (70, 73). Like other *Enterobacteriales* species, *Y. pestis* has typical cell wall but its lipopolysaccharide is defined as rough, carrying the core components but lacking the extended O-group side chains. Therefore, no O-antigen serogroups can be defined for *Y. pestis*. *Y. pestis* also possesses a pseudocapsule (bacterial envelope) often termed capsular antigen or fraction 1 (F1) antigen, made up by proteins and sugars, (73, 170, 171). At 37 °C, the bacillus loses its capacity to ferment glycerol, and requires biotin, thiamine, pantothenate and glutamic acid. These metabolic requirements have several consequences regarding the type of colonization and multiplication in both the flea vector and potential hosts of *Y. pestis*, precluding a saprophytic existence of the organism within the animal reservoir (70, 172).

Section 2.1

Evolution**2.1.1. *Yersinia* genus: taxonomy, ecology, and evolution**

The plague bacillus, *Y. pestis*, was one of the first species to be described within what would become later the genus *Yersinia*. The genus belongs to the *Yersiniaceae* family and is a member of the larger *Enterobacterales* order, together with other gram-negative, facultatively anaerobic gammaproteobacterial genera such as *Escherichia*, *Enterobacter*, *Klebsiella*, *Salmonella*, *Serratia* and *Shigella* (173). The genus *Yersinia* is composed by 26 species and includes three major human and animal pathogens that have been extensively characterized: *Y. pestis*, the agent of plague, and *Y. pseudotuberculosis* and *Y. enterocolitica* that cause enteric yersinioses (131). *Y. pseudotuberculosis* and *Y. enterocolitica* are transmitted through the fecal-oral route and have their main reservoirs in farm and wild animals, diverse species of birds and rodents, for *Y. pseudotuberculosis*, and mostly pigs for pathogenic *Y. enterocolitica* (174). Enteritis caused by *Yersinia* are considered the 4th most prevalent foodborne zoonotic diseases in the European Union (175), justifying the need for a thorough understanding of *Yersinia* genus diversity and distribution. A multilocus sequence typing (MLST) analysis permitted the characterization of the “*Y. pseudotuberculosis* complex”, a lineage composed by *Y. pseudotuberculosis*, *Y. pestis*, *Y. wautersii* and *Y. similis* (176). *Y. wautersii* is considered putatively pathogenic since harbors known pathogenicity factors but the specific mechanisms driving its pathogenicity remain to be investigated (177). Two other *Yersinia* species are known animal pathogens: *Y. ruckeri*, a fish pathogen, and *Y. entomophaga*, an insect pathogen (178, 179). In addition, 14 non-pathogenic *Yersinia* species have been described in the last decades: *Y. aldovae*, *Y. aleksiciae*, *Y. bercovieri*, *Y. frederiksenii*, *Y. intermedia*, *Y. kristensenii*, *Y. massiliensis*, *Y. mollaretii*, *Y. nurmii*, *Y. pekkanenii*, *Y. rohdei*, *Y. similis*, *Y. hibernica* and *Y. canariae*. In the recent years,

the advent of genome sequencing technology and its implementation into the framework of surveillance and reference laboratories has facilitated the automatization of species identification and strain characterization for epidemiological purposes. Recently, Savin and colleagues developed a core-genome MLST (cgMLST) scheme based on sequences of 500 genes shared in all *Yersinia* (180). This allowed to describe 6 novel, non-pathogenic species in the last years: *Y. artesianiana*, *Y. proxima*, *Y. alsatica*, *Y. occitanica*, *Y. vastinensis*, and *Y. thracica* (181). The updated phylogenetic structure of the genus *Yersinia* is detailed in **figure 2.1A**.

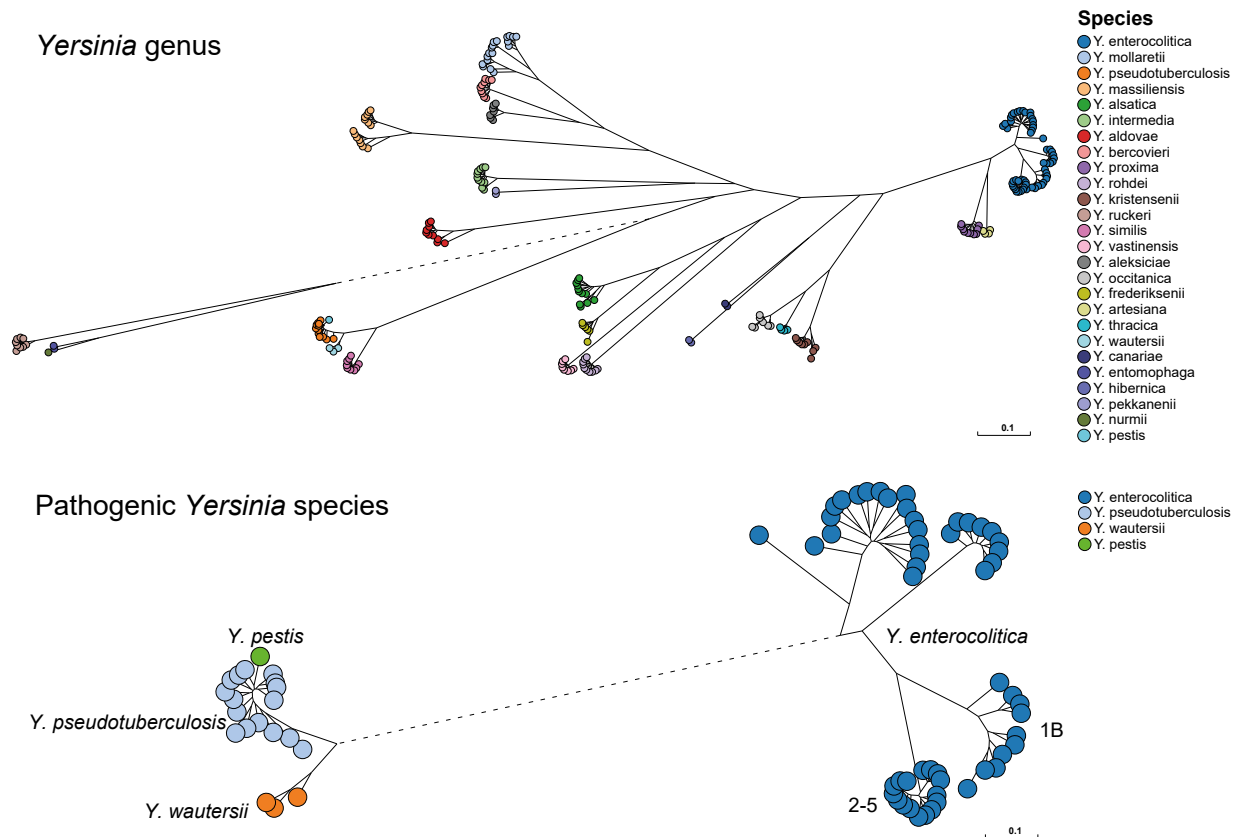


Figure 2.1. A: Phylogeny of most recently delineated species within the *Yersinia* genus. The tree was adapted from (181) and (180) with permission from the authors. **B:** A focus on the described pathogenic species for humans, from the same phylogenetic tree: *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica* biotypes 1B and 2-5. *Y. wautersii* is considered probably pathogenic to humans.

One notable feature of the *Yersinia* genus is the speciation trajectory that led to the emergence of pathogenic species in humans and other mammals (*Y. pestis*, *Y. pseudotuberculosis* and *Y. enterocolitica*), while their related species within the genus appear to be environmental, non-pathogenic bacteria. It was thus thought that human pathogenic *Yersinia* species had evolved from a common ancestor that acquired the shared pathogenicity and virulence features (182, 183). However, pathogenic *Yersinia* lineages (the *Y. pseudotuberculosis*, *Y. pestis* and *Y. wautersii* cluster on the one hand, and a subgroup of *Y. enterocolitica* on the other) are found in opposed branches in the *Yersinia* genus phylogeny (figure 2.1B). Besides the phylogenetic reconstruction, the access to WGS data of diverse *Yersinia* species allowed the comparative analysis of specific genes between lineages of the genus to uncover markers of the acquisition of pathogenicity in known pathogenic *Yersinia*. The factors that were early acquired independently in the two lineages are the intimin Ail and the pYV plasmid encoding the T3SS, which is considered a requirement for the capacity of pathogenic *Yersinia* to invade host cells and cause inflammation and hence determining their pathogenic potential (184, 185). Other important examples of gene gain in *Yersinia* evolution include the high pathogenicity island (HPI) encoding the siderophore yersiniabactin, present in *Y. pseudotuberculosis*, *Y. pestis* and a virulent lineage of *Y. pseudotuberculosis* (186, 187) and the *hms* locus, present only in the *Y. pseudotuberculosis*-*Y. pestis* complex and important for the *Y. pestis* flea-borne transmission (188). These cornerstone events mark the acquisition of virulence in pathogenic *Yersinia* lineages and altogether this is interpreted as a hallmark of the independent and parallel acquisition of pathogenicity in different *Yersinia* species (189). Indeed, the horizontal acquisition of virulence factors in environmental organisms to become highly adapted pathogens appear to be a common path in the emergence of pathogenicity within several *Enterobacteriales* genera (189, 190). For the ancestors of pathogenic *Yersinia*, this allowed them to colonize similar ecologic niches (i.e.: the mammal gut) and to independently specialize for its colonization and establishment into the new environment.

2.1.2. *Y. pestis* genome and molecular microbiology

In the recent decades, the genomes of numerous major human pathogens have been sequenced, providing unprecedented opportunities to understand their evolution and pathogenicity. Early entering the genomic era, sequence comparison of five housekeeping genes between tenths of diverse *Yersinia* isolates led to strong evidence proving that *Y. pestis* is an evolved clone of the more diverse *Y. pseudotuberculosis* bacillus and that therefore, all known *Y. pestis* share a common ancestor with *Y. pseudotuberculosis* (191). The first completed genome of *Y. pestis* was decoded few years later, in 2001 by Parkhill and colleagues in the Wellcome Sanger Institute (192). The genome corresponded to strain CO92, a clinical isolate from a fatal human case of pneumonic plague acquired from an infected cat in Colorado (USA), 1992 (193). CO92 is thereafter considered the reference genome for *Y. pestis* and is accessible under the RefSeq reference ID NC_003143. It consists of four replicons (figure 2.2): the 4.65-Mb chromosome as well as three plasmids, namely pYV/pCD1 (70.3 kb), pMT1 or pFra (96.2 kb) and pPCP1 or pPla (9.6 kb).

Interestingly, *Y. pestis* chromosome harbors a 102-kb unstable chromosomal region named *pgm* (pigmentation) locus carrying important genes for virulence and transmission. The locus is flanked by copies of insertion sequences of the type IS100, explaining its transferability through recombination (194). The *pgm* locus (figure 2.2) is composed of two parts: a 68-kb segment called the “pigmentation” segment coding for the *hms* gene (haemin storage), a factor required for the formation of biofilms that allow *Y. pestis* to block the flea foregut and enhance the vectorial transmission (188), and a 35-kb segment composed by an integrative and conjugative element (ICE) that is known as the *Yersinia* high pathogenicity island (HPI). The presence of the HPI distinguishes pathogenic *Yersinia* able to cause severe systemic infections, often fatal in humans, from other pathogenic *Yersinia* that do not carry this genomic island (195). The HPI carries genes (i.e.: *fyuA*, *irp1*, *irp2*, *ybtP*) necessary for the expression and transport of the siderophore yersiniabactin, a high-affinity iron-chelating

system that allows *Y. pestis* to sequester iron from the host, to import zinc as well as to limit the respiratory oxidative burst response of immune cells during infection (186, 196, 197). The HPI is found in several bacterial genera including enteric pathogens such as extraintestinal pathogenic (ExPEC) *Escherichia coli* (198), suggesting the horizontal transferability is common in other pathogens sharing similar ecological niches. Another described genomic island is the filamentous prophage Ypf Phi (**figure 2.2**), similar to an *E. coli* 018:K1:H7 prophage called CUS-1 (199). The prophage was shown to be infectious and conserved in diverse *Y. pestis* strains and could be either maintained as an episome or integrated within the chromosome such as in the CO92 reference genome. The depletion of the prophage appeared to slightly decrease *Y. pestis* pathogenicity in mice, suggesting its acquisition conferred a selective advantage for the pathogen (200). *Y. pestis* genome also harbors traces of HGT from insect pathogens, such as insecticidal toxin complexes (Tcs) from *Photorhabdus luminescens*, *Serratia entomophila* and *Xenorhabdus nematophilus* (201).

Perhaps the most striking feature highlighted when the CO92 genome was first analyzed was the surprising GC skew variation (G-C/G+C), which was suggestive for the very recent horizontal acquisition of DNA (i.e.: prophages, integrons), or rather the recombination of large blocks of DNA that may occur when growing *Y. pestis* in vitro. PCR confirmation of flanking regions in CO92 subcultures confirmed the observed translocations, suggesting an important plasticity of *Y. pestis* genome. Importantly, *Y. pestis* genome appears to be rich in repetitive sequences, mostly interspersed repetitive sequences and insertion sites (IS) elements (**figure 2.3**), which are thought to be major driving forces of *Y. pestis* genome plasticity (192). IS are mobile DNA sequences of less than 2.5 kb composed by repeat sequences flanking a gene cassette coding for a recombinase enzyme (transposase) that ensures its own translocation into the same or other DNA replicons (202). CO92 genome includes at least 4 different families of IS: IS200/IS1541, IS100, IS285 and IS1661 (203). *Y. pestis* appears to have accumulated up to 10-fold more IS copies compared to the ancestor species

Y. pseudotuberculosis and to other species of *Yersinia* (192). The accumulation of repetitive sequences, including IS, is a hallmark of pathogens having undergone evolution to adapt to particular hosts or environments. Gene inactivation due to IS insertion results in metabolic streamlining, which is characteristic of host-adapted pathogens (204). Recently, it has been shown that IS expansion shaped the evolution of *Shigella flexneri* and *S. sonnei*, which are human-adapted clades of *E. coli*, by driving their diversification and convergent loss of metabolic functions via the IS insertion in genes affecting similar functional pathways (205). *Y. pestis* genome is rich in pseudogenes, and genome reduction is considered to have played a decisive role during *Y. pestis* speciation (see **section 2.1.3**). In CO92 genome, out of 149 identified pseudogenes, 51 appeared to be consequence of an IS insertion within their reading frame (192). Other closed *Y. pestis* genomes have been released since CO92, including used laboratory strains such as KIM10+ and 91001, which have both different genetic backgrounds compared to the reference genome, but that are overall closely related one to each other (206, 207).

In addition to the chromosome, most *Y. pestis* isolates also carry three plasmids of different size that include genes important for *Y. pestis* virulence and vector transmission:

pCD1/pYV. The 70-kd plasmid called pYV (for *Yersinia* virulence) or pCD1 (for calcium dependence) is a plasmid that is specific to pathogenic *Yersinia* species and is critical for their virulence (131). The *Y. pestis* pYV was inherited from is conserved with the one of its ancestor *Y. pseudotuberculosis*. This plasmid encodes the *Yersinia* type 3 secretion system (T3SS) (**figure 2.2**) composed of 29 proteins (*ysc* genes) with a strong induction of their expression at 37 °C but includes also secreted effector proteins called Yops (*Yersinia* outer proteins). The main role of pYV is to prevent phagocytosis as well as to overcome the immune response of the mammalian host, ensuring the survival and multiplication of bacteria during infection (195). Once induced its expression, the T3SS delivers the Yops effector inside target hosts cells on contact, which will trigger the inhibition of actin polymerization (208). This

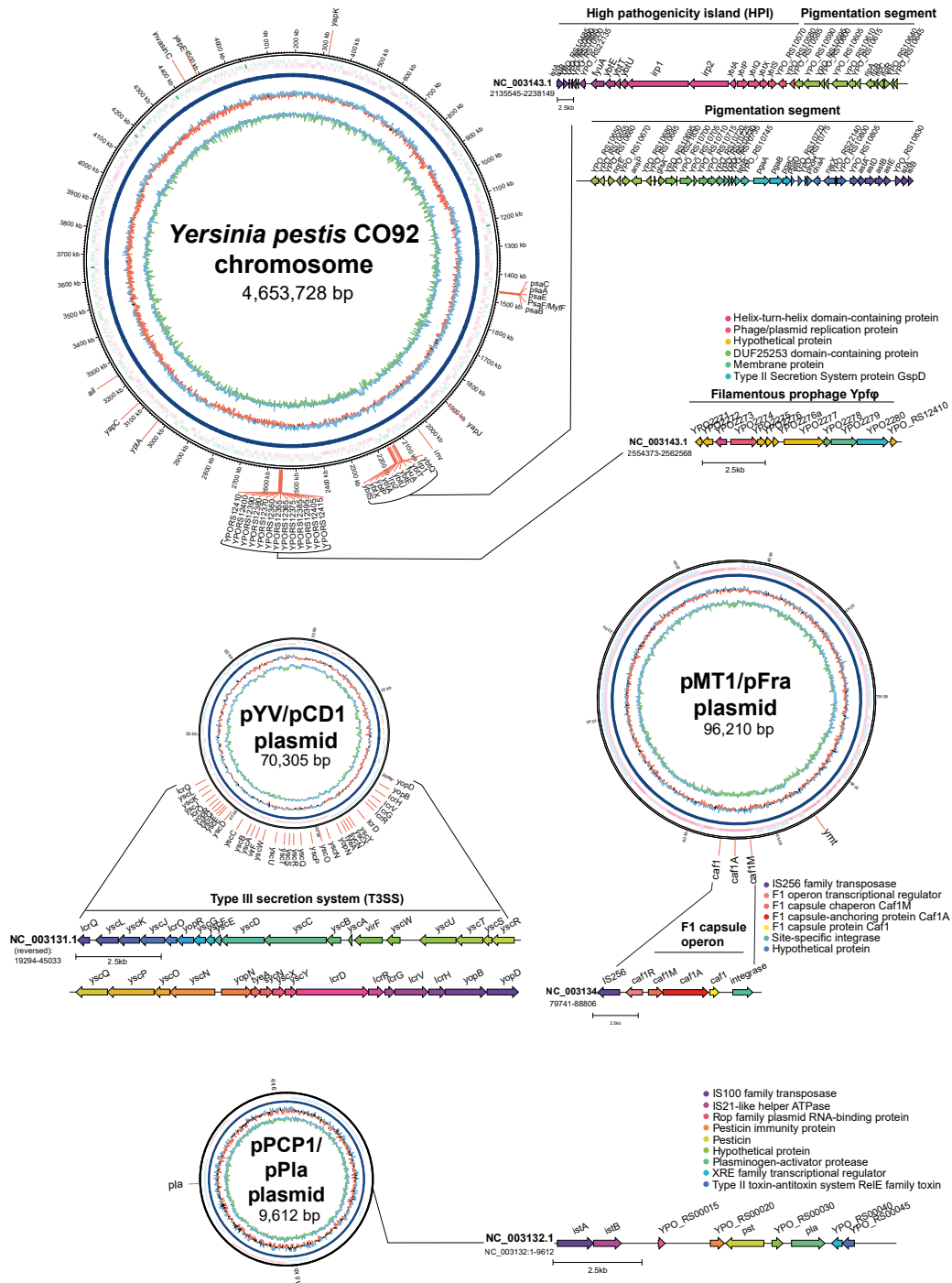


Figure 2.2. Circular representation of the replicons in the *Y. pestis* CO92 reference genome (NC_003143, NC_003131, NC_003132, NC_003134) (192). Relevant operons and loci associated to pathogenicity and vector transmission (discussed through the main text) are displayed.

mechanism prevents the phagocytosis of *Yersinia* by phagocytosing cells and suppresses the transcription of genes for the innate immune response, allowing bacterial proliferation in the host (209, 210). pYV also encodes the YadA adhesin, which is important for cell invasion in *Y. enterocolitica* and *Y. pseudotuberculosis*, but *yadA* is a pseudogene in *Y. pestis* due to a 1 bp deletion causing a frameshift (211). Few copies of pYV are found in each *Y. pestis* cell, although the copy-number may depend on the growth status and environment of the cell. Indeed, it has been shown that *Yersinia* up-regulate the pYV copy number during infection in order to ensure the dose of plasmid-encoded genes, a mechanism that is necessary for the full virulence phenotype in pathogenic *Yersinia* (212).

pPCP1/pPla. The smallest, *Y. pestis*-specific multi-copy (>100 copies) plasmid of 9.6 kb called pPla (for plasminogen activator) contains only 9 ORFs (figure 2.2) including a plasminogen activator protease (Pla) and a bacteriocin system (*pst* and *pim*). Pla is an outer membrane protease with multiple targets either on the same bacterial cell surface and on host cell and extracellular proteins. Its proteolytic activities on host complement components like C3 may contribute to inhibit cell apoptosis and suppress the immune reaction in phagocytes (195). Most importantly, the capacity of Pla of cleaving host plasminogen into plasmin inhibits host coagulation cascade inducing the disruption of blood clots. This activity may facilitate the dissemination of *Y. pestis* from the site of inoculation by the flea to distant sites such as the lymph node and spleen (214). Moreover, Pla has been shown to be required for *Y. pestis* to cause the fulminant respiratory infection that characterizes pneumonic plague (215). These described roles credit the protease Pla as one of the major virulence factors of *Y. pestis* both for bubonic and pneumonic routes of infection. The origin of pPla remains unknown, although its almost universal distribution in diverse *Y. pestis* suggests it was horizontally acquired early during *Y. pestis* emergence (131).

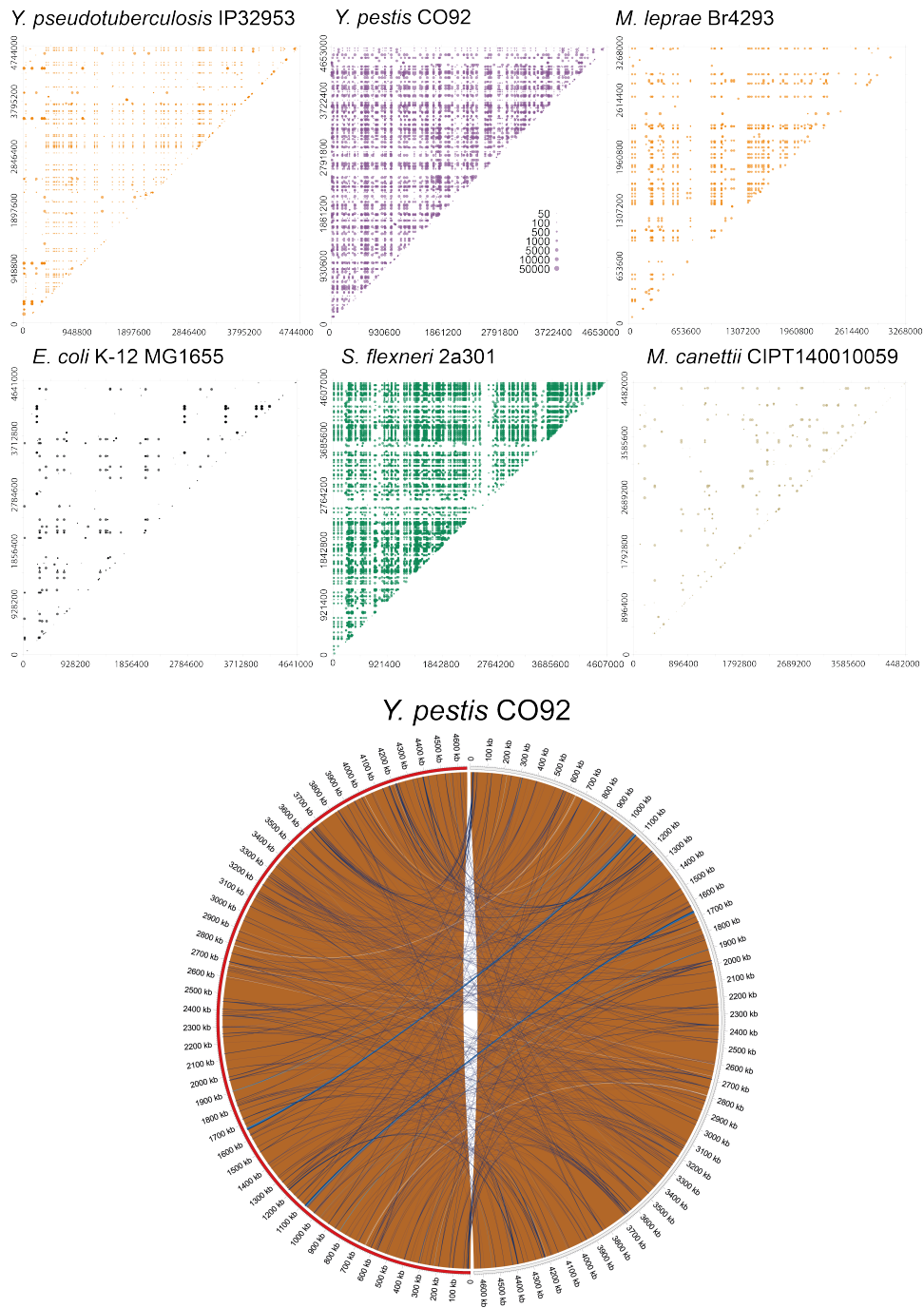


Figure 2.3. Up: Distribution of long repeat regions along selected bacterial chromosomes. The different plots represent the symmetric matrix of every pair of k-mer sizes repeats within each chromosome. These plots were generated using RepeatPlot (A. Criscuolo: <https://gitlab.pasteur.fr/GIPhy/RepeatPlot>, based on the graphical representation of the pioneering work by E. Rocha (213). **Down:** Circular representation of links between repetitive sequences within *Y. pestis* CO92 chromosome.

pMT1/pFra. The second specific plasmid that differentiates *Y. pestis* from *Y. pseudotuberculosis* is pMT1 (for murine toxin) or pFra (for fraction 1). First full sequence of pMT1 revealed its mosaic structure with important variations in G+C composition, suggesting the plasmid has undergone multiple rearrangements and events of HGT (216). The pMT1 plasmid is present in 1-2 copies per cell, shows variability in size within different *Y. pestis* strains (90-100 kb) and harbors loci that have been extensively characterized: the F1 antigen or capsule operon and the murine toxin *ymt*. The F1 capsule operon (figure 2.2) is composed by *caf1* (capsular structural subunit, also called F1 antigen), *caf1M* (chaperone), *caf1A* (capsule anchoring protein) and *caf1R* (AraC transcriptional activator). The F1 capsular protein is an important *Y. pestis* surface antigen which has also served as basis of several vaccine candidates (195). A polymer of F1 antigen forms an envelope or pseudocapsule around the bacterial cell when multiplying at 37°C, which appears to be important to provide some resistance to phagocytosis by monocytes (217). However, *caf1*/F1 or other capsular operon genes may not be present in all *Y. pestis* isolates found in wildlife, suggesting their presence is not essential for *Y. pestis* virulence (218, 219). The second well-studied pMT1-encoded gene is the *Yersinia* murine toxin Ymt. The gene encodes a phospholipase D enzyme which may be toxic to rats but not for other mammals (195). The *ymt* gene appears to have been acquired horizontally in pMT1 and it was initially thought to be essential for the colonization of the flea digestive tract by *Y. pestis*, thus playing a critical role in the flea-borne transmission of plague and pathogen ecological cycling in the animal reservoir (220). Nonetheless, it has been recently shown that the protective function of *ymt* rather depends on the type of blood in which the flea feeds and hence where *Y. pestis* will multiply, since Ymt-deficient *Y. pestis* could efficiently survive and be transmitted when fleas fed on brown rats. Therefore, the critical role of Ymt would be to greatly expand the range of hosts in which *Y. pestis* can establish a flea-borne circulation, such as in black rats, mice, and humans (221). Interestingly, both the F1 operon and the *ymt* gene showed divergent G+C composition compared to the

rest of pMT1 sequence, indicating that they had been possibly incorporated into pMT1 by HGT (216).

Despite the significant role played by pMT1 as a replicon harboring relevant genes for *Y. pestis* transmission and cell envelope characteristics, little is known about the origin of pMT1 and its acquisition by the common ancestor of the species. Sequencing of pMT1 provided some insights, such as the homology of around 50% of its sequence with a plasmid of *Salmonella enterica* serovar Typhi, as well as deletions blocks of 2kb observed when comparing pMT1 sequences from different strains (222). Since no conjugation system was originally found in CO92 pMT1 sequence, it was hypothesized that the plasmid would have been acquired through transformation. Recently, Pfeifer and colleagues made a captivating observation when analyzing the presence and distribution of phage-plasmids among diverse bacterial genera. Phage-plasmids (P-P) are elements known to be both plasmids and phages (i.e.: P1 phage, SSU5...) since they encode the necessary genetic components for their replication in line with the cell cycle, but also a full set of genes required for lysogeny and particle formation (223). P-Ps are therefore also temperate phages that can infect bacteria, produce virions, and lyse host cells (224). The study on the pangenome and gene homologies within putative P-Ps found that *Y. pestis* pMT1 has all the characteristic elements of a P-P and that most of its genes have homology with the SSU5 supercommunity of P-Ps, which includes SSU5/pHCM2 from *S. enterica* sv Typhi, as well as other *Enterobacteriales* putative P-Ps (i.e.: pKpn from *Klebsiella*). This recent description of pMT1 as a putative P-P provides an interesting hypothesis on the possible origin and relationship of pMT1 with replicons found in other bacterial pathogens. The possible P-P nature of pMT1 may suggest it arrived to *Y. pestis* common ancestor as an infectious phage that was then maintained in the genome. The described capacity of pMT1 to recombine and integrate *Y. pestis* chromosome (225) goes in the sense that pMT1 may be better understood as a P-P, although this hypothesis remains to be explored experimentally.

Altogether, *Y. pestis* genome is characterized by its genetic plasticity, the accumulation of pseudogenes through different mechanisms including frequent disruption by IS, as well as by the presence of 3 extrachromosomal replicons of different origins. These features of genome plasticity, horizontal acquisition of plasmids and gene inactivation probably resulted from the recent evolutionary adaptation of *Y. pestis* ancestor to a new ecologic niche through its transition from an enteric life (that of *Y. pseudotuberculosis*), towards that of a vector-borne, systemic pathogen.

2.1.3. Emergence of *Y. pestis* from *Y. pseudotuberculosis*

Since the identification of *Y. pestis* by Alexandre Yersin in 1894, and the description of other bacteria from the genus *Yersinia* through the past century, it became clear that *Y. pestis* and *Y. pseudotuberculosis* are closely related as they shared many characteristics (73). *Y. pestis* and *Y. pseudotuberculosis* have more than 95% average nucleotide identity in their sequences (figure 2.4), which has been proposed as a cut-off for bacterial species delineation (226, 227). Therefore, *Y. pestis* and *Y. pseudotuberculosis* form a single species at the molecular level but have been classified in different species because of their strikingly different pathogenicity, clinical features during infection and ecological niche (189, 228). To unravel more details on the evolutionary relationship between these two closely related pathogens before their genome sequences were available, Achtman and colleagues compared the sequences of five housekeeping genes in different *Yersinia*. Since the genetic distances between alleles in *Y. pestis* and *Y. pseudotuberculosis* appeared only within the range of *Y. pseudotuberculosis*, it became clear that *Y. pestis* is a clone that emerged from *Y. pseudotuberculosis* (191).

One of the central and most fascinating questions regarding *Y. pestis* evolution was the understanding of the molecular changes that led to the emergence of an extremely virulent, vector-borne pathogen like *Y. pestis* from a gastrointestinal pathogen ancestor like *Y. pseudotuberculosis*. The sequencing of first genomes of both *Y. pestis* (CO92 strain) and *Y. pseudotuberculosis* (IP32953 strain) permitted to shed some light on the divergence process

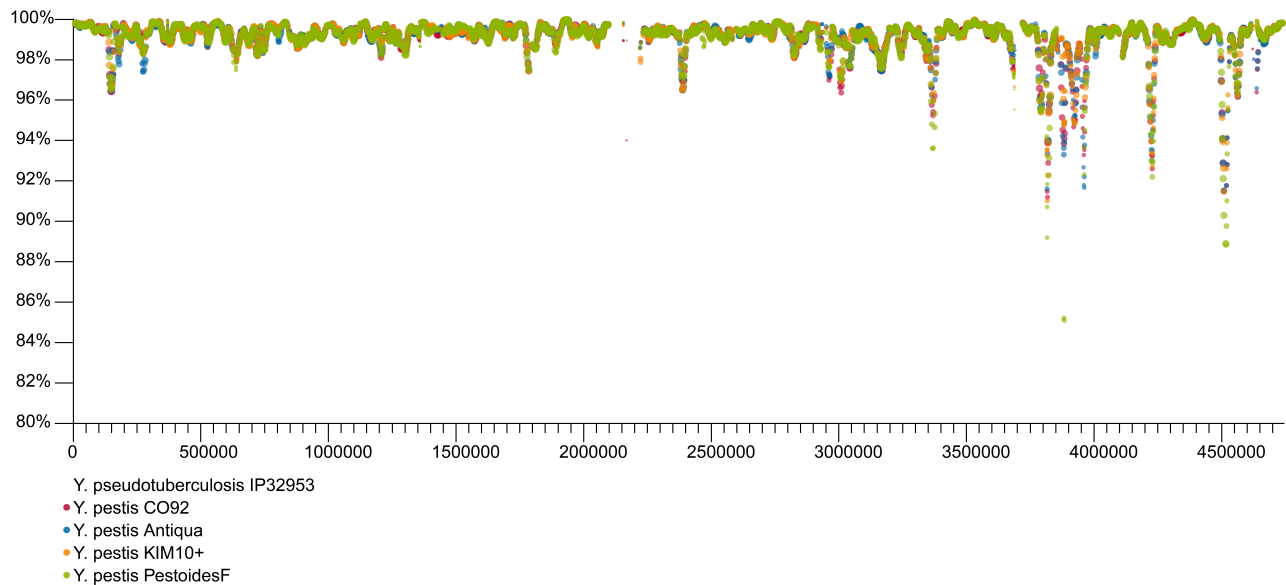


Figure 2.4. A: Similarity plot displaying the average nucleotide identities of several *Y. pestis* chromosomes over *Y. pseudotuberculosis* chromosome sequence. The plot was generated using SimiPlot (Criscuolo A., <https://gitlab.pasteur.fr/GIPhy/SimiPlot>).

that led to *Y. pestis* speciation by comparing their full sequences (229). The evolutionary events that guided the emergence of *Y. pestis* from *Y. pseudotuberculosis* have been gradually described through the last decades. These can be summarized as the combination of gene acquisition via HGT and genome reduction via gene loss or inactivation, both processes impacting the gain of virulence and the adaptation to a new environment by the ancestor of *Y. pestis* (figure 2.5) (230).

Major gene gain during *Y. pestis* speciation is associated with the horizontal acquisition of the two plasmids that are specific to the species, namely the pPla/pPCP1 and pFra/pMT1. As mentioned previously, the acquisition of these plasmids (or phage-plasmid for pMT1) were determinant for *Y. pestis* evolution as they encode factors allowing to expand the mammal host range of infection (Ymt phospholipase), and for the fatal infection and tissue destruction associated to *Y. pestis* infection (Pla protease) (215, 221). Although the impact of gene gain has been extensively explored in *Y. pestis*, major changes are associated with the loss of function of genes that are active in *Y. pseudotuberculosis*. Genome decay or reduction,

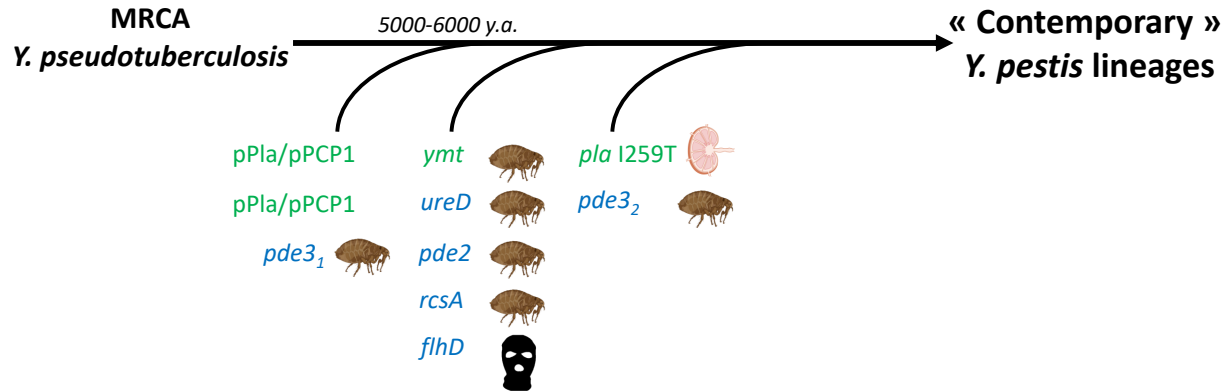


Figure 2.5. Schematic representation of relevant evolutionary events during the emergence and early evolution of *Y. pestis* from *Y. pseudotuberculosis*. Gene gain events and gene loss are displayed in green and blue, respectively. The flea represents an effect in vector-borne transmission by fleas, the lymph node represents an effect in bubonic plague, and the balaclava represents an effect in immune system evasion. The scheme is inspired from (131).

meaning gene inactivation and loss, often occurs when genes or gene clusters appear no longer useful for the survival of a microbe or when the microbe undertakes adaptation to a new ecological niche (231). Gene loss has been shown to play a central role in the emergence of some bacterial pathogens, as removing unnecessary genetic elements may confer pathogens a better fitness when adapting to a novel niche (232). This mechanism is a common signature in the emergence of major pathogens such as *Mycobacterium tuberculosis*, *M. leprae* and *Shigella* spp., (233–235). When comparing the genomes of *Y. pestis* and *Y. pseudotuberculosis*, it was observed that around 10% of *Y. pestis* genes were pseudogenes (lost their function) compared to their homologs in the *Y. pseudotuberculosis* ancestor. Many of the genes that were inactivated during *Y. pestis* emergence encode virulence factors (Inv and YadA invasins and adhesins, respectively), metabolic and transporter genes (*pheA*, *iucA*), and motility systems (*flhD* regulator), factors that were important for the life of *Y. pseudotuberculosis* as a gastrointestinal pathogen (182, 192, 229). This niche was no longer typically inhabited by *Y. pestis*, which adapted to colonize the flea gut and to become a pathogen infecting the mammalian lymphatic system. Therefore, the inactivation of several genes served

also to the survival and adaptation of *Y. pestis* to the flea, notably, the *pde2*, *pde3*, *rcaA* and *ureD* loci (**figure 2.5**). The inactivation of these first three loci led to an enhanced ability of *Y. pestis* to form biofilms in the flea foregut, which is important for a mechanism of flea-borne transmission of plague (159). The loss of function of the urease gene *ureD* was critical to allow a reduced toxicity in the flea tract (236). These changes through pseudogenization marked the switch in the pathogenic lifestyle of *Y. pestis* ancestor towards that of a vector-borne pathogen (230). Another feature that characterizes *Y. pestis* compared to *Y. pseudotuberculosis* is its cell surface, as *Y. pestis* expresses a rough lipopolysaccharide (LPS) that lacks the O antigen. The inactivation of five genes necessary to build a functional O antigen appeared to have facilitated the function of the Pla surface protease and to contribute to the tissue invasion by *Y. pestis* (237). *Y. pestis* also expresses a modified (tetra-acylated) lipid A on its LPS due to the loss of the *lpxL* gene. The hexa-acylated LPS is commonly present in many Gram-negative bacteria and is efficiently recognized by host toll-like receptor 4 (TLR4) signaling, in contrast to the *Y. pestis* modified LPS which has low affinity for TLR4 (203). These mutations impacting the cell surface are thought to have contributed to the gain of virulence and immune subversion by *Y. pestis* (132).

Overall, the genomic and functional analyses of both *Y. pestis* and its ancestor *Y. pseudotuberculosis* have revealed precious insights into the process of gene gain and genome decay that led to the emergence of *Y. pestis* as an extremely virulent, flea-borne pathogen causing plague. Still, several specific mechanisms of virulence, transmission, and immune evasion by *Y. pestis* are not fully understood. It is probable that the broadening of *Y. pestis* genomic data, advances in *in vitro* and *in vivo* microscopy and infection models may allow to reveal more secrets that make *Y. pestis* a remarkable model to understand bacterial evolution and pathogenicity.

Section 2.2

Population genetics of *Y. pestis*

The entering into the genomic era has represented an exponentially increasing amounts of available bacterial and viral sequencing data (238). Available genomic data for *Y. pestis* has raised from the first closed genome in 2001 (192), to more than 1,500 samples included in integrated bacterial genomics databases such as EnteroBase (consulted in June 2022) (239). Within these two decades, the field of paleogenomics emerged and has provided for the first time with genomic data from *Y. pestis* present in infected human beings in the past millennia (240). Population genomics methods have permitted the unravelling of key evolutionary events marking the emergence of *Y. pestis*, the identification of the main lineages defining the taxonomy of the species, and an improved understanding of the dynamics of spread and maintenance of plague both during historical pandemics and within endemic countries. Phylogenetic approaches are particularly attractive to analyze genomic data to illuminate the genetic diversity of the plague bacillus, and the understanding of its distribution and dynamics can ultimately provide valuable information for the understanding of the epidemiology of plague. In these last introductory sections, an overview will be provided on *Y. pestis* diversity, evolution and spread through the lens of population genomics.

2.2.1. Genetic monomorphism, selection, and neutral evolution in *Y. pestis*

The genetic diversity within bacterial species is usually large enough that the comparison of several housekeeping genes is usually enough to discriminate between unrelated strains. The analysis of few conserved genes was therefore a powerful and rapid strategy to understand the population genetic structure of pathogen isolates (241). However, some bacterial pathogens appeared to have a small to no polymorphism when comparing few conserved genes between diverse isolates. These pathogens with such a reduced level of genetic diversity have sometimes been called as genetically monomorphic organisms (242). Interestingly, most relevant

microbes described as being genetically monomorphic include highly pathogenic microorganisms such as *Bacillus anthracis*, *Mycobacterium tuberculosis*, *Burkholderia mallei*, *Salmonella enterica* serovar Typhi, *Shigella sonnei* and *Yersinia pestis* (191, 243–246). Many of these genetically monomorphic pathogens appear to have clonal population structures in which homologous recombination and homoplasies are rare compared to other bacteria (242). The observed clonal evolution is consistent with the idea that they represent successful, adapted virulent clones that emerged out from a pool of more diverse bacteria (233). In the case of *Y. pestis*, the reduced levels of genetic diversity may be explained by the evolutionary bottleneck that resulted from its emergence and change of ecological niche compared to its ancestor *Y. pseudotuberculosis* (242). Further evidence supporting this hypothesis came for the first time with the whole-genome sequencing of more than a hundred diverse *Y. pestis*, mainly from Asia, allowed to identify single nucleotide polymorphisms (SNP) that were used to build a species phylogeny that confirmed the clonality characterizing *Y. pestis* evolution (247). The analysis of the distribution in the phylogeny of the identified SNPs within core-genes was used to explore possible traces of natural selection having influenced *Y. pestis* evolution. The dN/dS ratio was estimated for individual nodes of the phylogeny, and ranged values near to 1, which was interpreted as a signal of neutral evolution. This represented that most of the observed diversity in *Y. pestis* species evolution would result from genetic drift rather than selection (203). The search for homoplastic SNPs as markers of diversifying and parallel selection revealed only 28 homoplasies within the core-genes in the dataset studied. Finally, the analysis of clusters of non-synonymous, protein-altering, SNPs (nsSNPs) revealed only 22 outlier loci that accumulated more nsSNPs and/or homoplasies compared to the rest of the core-genes. No obvious unifying trends were found within the list of genes showing these markers of diversifying selection (247). Still, despite the diverse genomes included in the study, these analyses were based in more than a hundred samples, in absence of a long-term and more exhaustive sampling for several plague endemic areas outside of Asia. The iden-

tification of genes that may be negatively selected or hyperconserved is another aspect that still needs to be explored regarding the functional evolution in *Y. pestis* species.

2.2.2. Genetic diversity of *Y. pestis*, genotyping and distribution

Since the discovery of *Y. pestis* as the agent of plague, efforts were put on characterizing its diversity in order to establish possible relationships between some strains and historical pandemics as well as variants circulating in endemic areas of plague. *Y. pestis* is a homogeneous species with a single O serogroup and phage type (203), and hence other typing were developed compared to other bacterial pathogens characterized by pathovars. The most widely used typing method for *Y. pestis* is biotyping, based on the ability of some *Y. pestis* strains to ferment glycerol and reduce nitrate. Based on the distribution of the strains biotyped and historical data, three biovars were established: Antiqua, Medievalis and Orientalis (248). This nomenclature assumed that these biovars caused, respectively, the 1st, 2nd and 3rd plague pandemics, although only biovar Orientalis is correctly associated with the 3rd plague pandemic, whereas the other two biovars cannot be retraced back to any pandemic. *Y. pestis* biovar Antiqua can reduce nitrate compared to biovar Medievalis and Orientalis, while *Y. pestis* biovar Orientalis lost the ability to ferment glycerol. Additional biochemical reactions were added to differentiate between *Y. pestis* strains, including the ability to ferment arabinose, rhamnose and melibiose (87). With the advent of molecular typing techniques such as the analysis of variable number of tandem repeats in multiple locus (MLVA), further differences between *Y. pestis* strains could be established, and differences in virulence in humans and host distribution were also added to the classification (249, 250). Accordingly, *Y. pestis* may be classified into 5 different biovars: Antiqua, Medievalis, Orientalis, Intermedium and Pestoides (or Microtus). Pestoides are thought to be diverse *Y. pestis* which are strongly associated to several sylvatic animal plague areas (mainly voles of the *Microtus* genus), notably in the former Soviet Union countries (251). The biotyping classification had important limitations and incongruencies resulting from the possibility that some

strains may revert their phenotype, and molecular typing methods based on repetitive sequences analysis had limited resolution and suffered from recurrent homoplasies (242). Once the first *Y. pestis* genomes were sequenced, the definition of *Y. pestis* groups according to phylogenetic groups based on SNPs identified within the sequenced genomes appeared to be an adequate strategy to define the intra-species taxonomy of the pathogen (252).

The use of SNPs for phylogenetic and genotyping purposes is particularly useful in the case of genetically monomorphic bacteria such as *Y. pestis* as they may provide more stable phylogenetic relationships across samples and, if identified in a whole-genome fashion, they can provide greater resolute power (253). Whole-genome SNPs (wgSNPs) have therefore been widely used to reconstruct the phylogenies of different human pathogens in the recent years (254–256). To conciliate the classical *Y. pestis* biotypes nomenclature (based on biochemical reactions) and the phylogenetic groups (based on marker SNPs initially identified from the first *Y. pestis* genomes), Achtman and colleagues defined a nomenclature system that designated the *Y. pestis* populations according to their phylogenetic branch (i.e.: 0-4), followed by the abbreviation for the known biovars: PE, Pestoides (also called Microtus); MED, Medievalis; ANT, Antiqua; IN, Intermedium; ORI, Orientalis (252). As more *Y. pestis* isolates were classified using this nomenclature, a number was added after the biovar abbreviation to define the monophyletic sublineages within the main population. For instance, the monophyletic lineage of *Y. pestis* from the biovar Orientalis fall in the branch that was called branch 1, and therefore the lineage is called 1.ORI; accordingly, 1.ORI1, 1.ORI2 and 1.ORI3 would be different sublineages of the 1.ORI population. This nomenclature system has been widely used in recent studies that sought to provide more genomes and classify the new samples into main *Y. pestis* lineages (62, 64, 89, 247). The main drawback of this nomenclature system is that is based on phylogenetic relationships established using the *Y. pestis* genetic diversity inherent to the isolates being included in each new analysis. Consequently, when isolates forming novel branches in the *Y. pestis* phylogeny are added, if the

names of the previously called lineages or sublineages is maintained, the real phylogenetic order does not appear obvious from the strict nomenclature fashion (i.e.: 0.PE7 lineage is ancestral to 0.PE2 and 0.PE4 in *Y. pestis* phylogeny).

The first attempt to unravel the patterns of *Y. pestis* global genetic diversity, resulting from a collective and coordinated effort from reference laboratories around the world, compared 17 genomes allowing identify 1364 SNPs within core genes sequences (conserved in all isolates). This SNP discovery approach was then used to perform SNP-typing on 288 global isolates to call the states of selected SNPs and use these states to reconstruct a *Y. pestis* tree based on a minimum spanning tree approach (46). Few years later, the first evolutionary insights into *Y. pestis* diversity fully based on WGS came from the sequencing of 118 diverse *Y. pestis* from China and Mongolia, that were combined to the few available genomes to build a core-genome SNP phylogeny that revealed new branches in the *Y. pestis* species tree (247). These two cornerstone studies provided for the first time a global depiction of the diversity, evolution and spread of *Y. pestis* and plague. First, the reconstructed phylogenies confirmed the clonal population structure of *Y. pestis* species, in which several branches emerged from a central evolutionary branch that derives directly from the *Y. pestis* common ancestor with *Y. pseudotuberculosis* (**figure 2.6**). The second relevant finding was that the newly identified lineages and sublineages revealed patterns of strong geographical clustering. Several lineages or populations from the groups 0.PE and 0.ANT emerge early from the central branch and were mainly isolated from China, Mongolia and countries from Central Asia and the Caucasus. Lineage 0.PE3 is comprised only by the particular Angola strain, which according to its name receives an alleged African origin, although it is considered unknown (257). At the end of branch 0, a polytomy appears in most *Y. pestis* phylogenies (also called “big bang” polytomy), where branches 1, 2, 3 and 4 emerged. Branch 2 was split into two lineages: 2.ANT, represented by samples from China, Nepal and Russia, and lineage 2.MED, including *Y. pestis* isolates distributed within China and the Middle East.

Branch 3 included isolates from China and Mongolia (3.ANT lineage), while branch 4 is represented by few isolates from Mongolia and the former Soviet Union (4.ANT lineage). Branch 1 includes isolates from diverse geographic origins and comprises three lineages: 1.ANT, 1.IN and 1.ORI. 1.ANT appeared to be restricted to Central and Eastern Africa, while 1.IN are exclusively found in the eastern regions of China. The most evolved lineage in branch 1 is composed by 1.ORI isolates, which are associated to the 3rd Plague Pandemic that emerged by the end of the 19th century in China and are distributed through all the continents (46, 247). Even if these first attempts to characterize the global distribution of *Y. pestis* isolates could identify most of its diversity, these studies were performed before or early in the advent of next-generation sequencing and hence an in-depth genomic analysis was not always possible. The study by Morelli and colleagues could attribute the lineage of few hundreds of selected *Y. pestis* isolates around the world, but their complete genome sequence was not generated but for a tenth of them, which could be a source of phylogenetic discovery bias (46, 242). On the other hand, the work by Cui and colleagues was based on WGS but only on few representative isolates from each lineage and sublineages present mainly in China (247).

The development of paleogenomics and paleomicrobiology (**see later**) fields was marked by the sequencing of the first ancient *Y. pestis* genomes and metagenomes that were recovered from victims of historical plague pandemics in Europe (240). Ancient *Y. pestis* genomes reconstructed from victims of the 1st Plague Pandemic or Justinianic plague in Europe (6th-8th centuries) form a monophyletic lineage emerging from the central branch 0 after modern 0.ANT1 isolates (258, 259). Recent ancient DNA (aDNA) studies have provided with tenths of *Y. pestis* recovered from victims of the 2nd Plague Pandemic, including the 14th century Black Death outbreak. These studies placed the 2nd Pandemic ancient samples branching early within branch 1 (forming the also called 1.PRE lineage), being the most ancestral *Y. pestis* genomes in this branch (239, 260–263). The oldest *Y. pestis* aDNA samples that

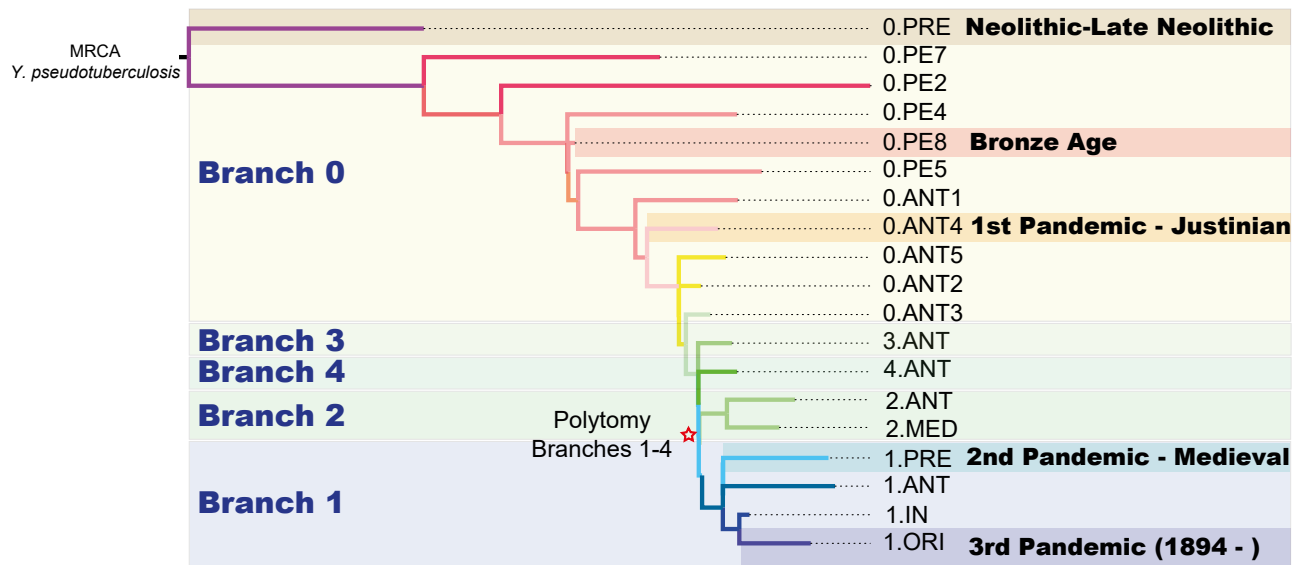


Figure 2.6. A: Schematic representation of *Y. pestis* phylogeny, where major identified lineages are displayed. The Angola genome (0.PE3) is not shown for clarity of the visualization due to its long branch. Major branches in the phylogeny and the polytomy leading to the emergence of branches 1-4 are indicated at left. Lineages associated to historical plague pandemics or to ancient DNA metagenomes from prehistory are highlighted (right).

have been analyzed contemporary to this period date to 1338-39 in present-day Kyrgyzstan. Strikingly, these samples fall in the middle of the polytomy in *Y. pestis* phylogeny (**figure 2.6**), preceding all the 2nd Pandemic samples and hence representing the direct ancestor of the lineage that caused this pandemic in Europe (264). The sequencing of aDNA from prehistoric human samples has also allowed the reconstruction of *Y. pestis* genomes from around 5,000 years ago Eurasia (265–269). These ancient genomes fall at the most ancestral part of *Y. pestis* phylogeny, emerging from branch 0 such as a lineage associated to the late Neolithic and early Bronze Age plague (also called LNBA)(270). These prehistoric *Y. pestis* represent the most ancestral described genomes in the species phylogeny, and branch early to the contemporary isolates of the 0.PE7 lineage. It is important to note that the branches associated to the prehistoric plague, as well as those associated to the 1st Pandemic, only include genomes recovered from aDNA and therefore represent lineages that probably have become extinct, with no contemporary representative *Y. pestis* persisting until the present day.

The increase in next-generation sequencing implementation in the recent years has provided with a general outline of *Y. pestis* diversity, including lineages represented by aDNA metagenomes recovered from victims of historical plague pandemics. Besides these works characterizing new main lineages in *Y. pestis* phylogeny, several studies reporting novel genomes have permitted to expand the diversity found within lineages already described in the past, giving a dynamic, step-wise and sometimes scattered overview of *Y. pestis* genetic groups (48, 62, 64, 271–273). This situation gives the opportunity to recap the species global phylogenetic history, mutation spectrum and gene content using the hundreds of genomes that have been published in the recent years. Thus, the sequencing and share of WGS data of novel *Y. pestis* isolates remains critical for our understanding of the evolutionary relationships of the plague bacillus.

2.2.3. Microevolution of *Y. pestis*

The recent increase in *Y. pestis* genomic data, both from the sequencing of isolates collected in the contemporary era and ancient genomes from victims of historical plague epidemics has revealed some remarkable traces of evolutionary adaptation of the plague bacillus having occurred since the divergence of the most ancestral lineages. The identification of genetic variations that may have contributed to *Y. pestis* adaptation remains particularly difficult due to its reduced levels of genetic diversity in terms of SNPs (242, 247). Nevertheless, the combination of genomic analysis of ancient and modern genomes has allowed the identification of several genomic variants that have further been investigated in experimental setups. Such variants may be point mutations, small insertions and deletions, or gene content variation. These genomic alterations helped to retrace the microevolution of *Y. pestis* and has uncovered some patterns of adaptation that may have contributed to the evolutionary success of the plague bacillus once it had already diverged from the *Y. pseudotuberculosis* ancestor. One of the investigated factors important for *Y. pestis* virulence is the Pla protease, encoded in the *Y. pestis*-specific pPla/pPCP1 plasmid (214). It has been observed that the punctual

mutation from isoleucine to threonine at position 259 (I259T) differentiate the most ancestral *Y. pestis* genomes (the prehistoric samples from aDNA, as well as contemporary lineages 0.PE7, 0.PE2 and 0.PE4) from the rest of *Y. pestis* (207, 240, 274, 275). If the presence of the derived threonine in Pla protease did not appear to be required to cause pneumonic plague, this mutation was associated with the capacity of *Y. pestis* to cause a more invasive infection when studied in mice models of bubonic plague (215). *Y. pestis* would thus have gained Pla early in its emergence by horizontally acquiring the pPla/pPCP1 plasmid, allowing to cause pneumonic plague, and then critically selected the I259T Pla variant to facilitate tissue invasion and its dissemination within the mammal host during bubonic plague. Intriguingly, another genomic alteration implicated the *pla* locus later in *Y. pestis* evolution. The analysis of ancient *Y. pestis* genomes from human remains of the late Middle Age in Latvia led to the identification of a decay in *pla* locus coverage (276). If the locus was still present in the recovered genomes, its decay highlighted the presence of *pla*-deleted pPCP1 plasmid mixed with *pla*-positive pPCP1 plasmids, suggesting the coexistence of both forms within the same *Y. pestis* genomes. The partial deletion of a ~2kb region in pPCP1 plasmid was also found in several other late 2nd and 1st pandemic *Y. pestis* ancient genomes, indicative of the convergent evolution of *Y. pestis* to reduce the *pla* gene dose at the end of two historical plague pandemics (277). This decay in *pla* gene as a factor relevant for *Y. pestis* virulence in bubonic plague infection was hypothesized to be a possible cause for the extinction of plague in Europe, as the branches where the deletion has been described include only ancient DNA genomes (276). However, the virulence potential of *Y. pestis* carrying the mixed *pla*+/*pla*- pPCP1 plasmids can just be evoked in a theoretical framework, as the impact of the partial *pla* deletion has never been investigated in experimental plague models of infection. On the other hand, *Y. pestis* from the 0.PE2 lineage isolated in natural animal plague foci in Asia appear to lack the complete pPCP1 plasmid (131, 278). These isolates fully lacking the Pla protease encoded by the plasmid are still fully capable of causing septicemic infection in

mice (214). The convergent partial depletion in pPCP1 implicating the *pla* locus, as well as the loss of the whole plasmid in 0.PE2 lineage suggest that the reduction or loss of Pla activity may be beneficial for *Y. pestis* to adapt to particular environments. Intriguingly, a 49-kb chromosomal deletion including *mtgB* and *mtgC* genes appear in parallel in several late 1st and 2nd pandemic ancient *Y. pestis* genomes. The implicated loci encode for Mg²⁺ transporters that may be important for *Y. pestis* survival within host cells such as macrophages (262, 279). Again, the impact of this convergent deletion has not been addressed experimentally, and thus it is difficult to estimate the impact of such alterations in *Y. pestis* virulence.

2.2.4. Diverse mutation rates and the cryptic molecular clock of *Y. pestis*

Pathogen phylogenies are attractive tools to understand the genealogy, diversity and spread of infectious diseases. When the time of sampling of the microbial isolate or the age of the reconstructed genome is known, this information can be incorporated into phylogenetic and evolutionary models. This allows the calibration of the phylogenetic tree, building a dated or time-calibrated tree in which: i) branch lengths are converted to time units, ii) the positions of branch tips correspond to the indicated sampling times, and iii) the ages of internal nodes are inferred and placed in the most likely divergence time according to a distribution (280). Powerful methods to reconstruct dated trees are based on Bayesian phylogenetic inference, which is implemented in popular software such as BEAST (281, 282). Bayesian phylogenetic analysis is based on Bayesian statistical methodology allowing to make direct probabilistic statements about the model and parameters of the phylogenetic inference (283). Briefly, the tree topology and substitution model indicate the statistical model for the input data, that usually are DNA sequence alignments and their sampling times. Molecular phylogenies that use Bayesian inference will therefore generate a dated tree to which a posterior probability is associated, that is, the probability that the reconstructed evolutionary model is correct given the data (284). Dated molecular phylogenies also use molecular clock assumptions to model

the divergence between sequences, calibrated by their sampling times. This can be used to estimate the mutation rates of a pathogen throughout the tree and how these mutation rates change between taxa (285). The reconstruction of dated phylogenies can then be used to infer the time of emergence of a particular lineage, to estimate the time a pathogen spread into a particular area (i.e.: in phylogeographic analysis), as well as to provide insights on the mutation rates associated to microbial lineages (286).

The first insights on the evolution and mutation rates of *Y. pestis* lineages were obtained from the analysis of a dataset of 134 diverse isolates (247). This study described the overdispersion of branch lengths between different *Y. pestis* lineages, which could not be directly explained by the progressive accumulation of diversity during the sampling period of the isolates. The reconstruction of a model for the temporal evolution using a Bayesian phylogenetic inference found that a relaxed molecular clock performed better than the assumption of a strict molecular clock to model *Y. pestis* evolution. Intriguingly, this inference showed that the mutation rate had changed within branches and throughout time in the phylogeny in up to 40-fold between the fastest and slowest-evolving branches. The possible drivers of the inferred historical variations in *Y. pestis* mutation rates are still not fully understood. As homoplasies are relatively rare in *Y. pestis* and the analysis of mutation spectrum did not provide with strong patterns of selective pressure, the hypothesis that seems more plausible is that rate variation would be a mirror of the demographic changes in *Y. pestis* evolutionary history (247). Indeed, *Y. pestis* phylogeny is defined by the presence of star-like branching events (polytomies), which may reflect the clonal expansion during the spread of the pathogen in epidemics. Such examples of radiation are observed in the 1.ORI lineage associated to the 3rd Pandemic, or the polytomy preceding the emergence of branches 1, 2, 3 and 4 (also known as “Big Bang”)(287). According to this hypothesis, *Y. pestis* mutation rate variation would change according to the different bacterial replication rates when plague bacilli circulate within their sylvatic animal reservoirs (*enzootic cycle*), compared to

their spread to other intermediate animal hosts (*epizootic* cycle) and in human plague epidemics. Therefore, mutation rates would accelerate in epidemics and epizootics because of a quick increase in the number of bacterial generations in a given period of time, and hence that fastest-evolving branches in *Y. pestis* phylogeny may be associated to epidemic events (247). This hypothesis may be complemented by recent observations of the mutation rates in emerging viruses such as SARS-Cov-2. It has been described that the emergence of new SARS-Cov-2 lineages or variants, particularly variants of concern (VOC), is characterized by the fixation of a certain number of mutations in a small period of time. Interestingly, the episodic increase of mutation rates up to 4-fold at the stem of the branch leading to an epidemic VOC lineage has been observed for SARS-Cov-2 (288). In the case of SARS-Cov-2, the emergence of VOCs is followed by their fast epidemic expansion, thanks to beneficial mutations acquired during its emergence, that may help the virus escape host immunity (289). Therefore, the increased mutation rate during the emergence of new variants would be also related to the selection of mutations that increase transmissibility between hosts, which has not yet been described in the case of *Y. pestis* pandemic lineages. The precise mechanism of emergence of VOCs in SARS-Cov-2 is still unclear.

On the light of the generation of genomic data for several hundreds modern and ancient *Y. pestis* genomes in the recent years, Eaton and colleagues attempted to further explore the mysterious variation in clock rates of the plague bacillus. When attempting to perform statistical tests on the temporal signal in the dataset, they could not fit a single clock model to the whole dataset comprising all diverse *Y. pestis* lineages, in line with the previous observations of rate variation in the plague bacillus (247, 287). Indeed, it was observed that some lineages of *Y. pestis* were more variable in terms of genetic to temporal distance regression, while others had lower levels of variation (evolving more “clock-like”). Accordingly, temporal signal was found in most lineages when their evolution was assessed in a lineage-specific fashion, although the lack of temporal signal in some lineages may be due to the absence

of isolates covering a sufficiently large period (i.e.: several decades) to provide enough signal. This study proposed a strategy to model the *Y. pestis* temporal evolution by targeting specifically each of the monophyletic lineages, with the aim of bypassing the conflicting rate variation observed along *Y. pestis* phylogeny. It has been recently discussed that mutation and speciation rate variations across species phylogenies may lead to inference errors of divergence times (290). This strategy may thus allow to provide more informative divergence dates and help reconstruct the evolution of individual lineages. On the other hand, when attempting to study the evolution and divergence of *Y. pestis* lineages at the species level, this strategy may not fully benefit from the function that aDNA genomes can provide for calibrating the molecular clock at precise positions of the phylogeny (287).

Reconstructing the temporal evolution of *Y. pestis* seems problematic due to the complexity of its patterns of genetic diversity (i.e.: multiple polytomies), demographic history (interspersed pandemics and natural circulation in endemic areas), and the disparities in mutation rates that are inferred when attempting to analyze its evolution at species level. It is likely that a combination of different approaches, together with the development of new evolutionary models, may provide a deeper understanding of the evolutionary dynamics of *Y. pestis* through the last millennia.

Section 2.3

Spread of plague inferred from phylogenies

The advent of next-generation sequencing has been followed by the development and implementation of methods allowing to reconstruct with relative confidence the temporal and evolutionary relationships between pathogen genomes. The combination of genomic data and metadata associated to the different pathogen isolates may be used to uncover spatial patterns of transmission of infectious diseases (291). This is particularly attractive in the case of *Y. pestis* as a pathogen having caused pandemics through human history, claiming the lives of million human beings almost all over the world. *Y. pestis* is a clonal pathogen, for which all its lineages emerged from a central evolutionary branch that evolved from the common ancestor with *Y. pseudotuberculosis* ancestor. This feature is a mirror of the unidirectionality guiding *Y. pestis* evolution from the root to the tips in its phylogeny, allowing to directly infer the historical routes of plague transmission out of the branching of different lineages in the tree (46). Therefore, phylogenomics methods, combined with historical record, have been used to infer the historical spread of plague at different scales of time, including the periods associated to the three documented plague pandemics (203). Historical sources detailing plague epidemics include archives, personal diaries, paintings, records, and human remains (i.e.: mass graves) (2). According to the available records, plague has caused at least three major pandemics in human history. The 1st Plague Pandemic or Plague of Justinian, named after the emperor of the Roman Empire Justinian I, spanned between the 6th to the 8th century, mostly affecting the Mediterranean area. The 2nd Plague Pandemic started in 1346 and continued until the 18th century in Europe. Finally, the 3rd Plague Pandemic started in the late 18th century in China, and since 1894-98 it spread to all continents to persist until present day (2, 46, 73). The combination of phylogenetics and population genomics approaches with a historical perspective, has revealed interesting details concerning

each of the different pandemics.

2.3.1. Spread of plague during the 3rd Plague Pandemic and contemporary *Y. pestis* lineages

The 3rd Plague Pandemic is thought to have started in the Chinese province of Yunnan by the second half of the 19th century, when plague epidemics reemerged in the region. Plague was already endemic in Yunnan, and by 1894 the epidemic had spread to the cities of Canton and Hong Kong causing thousands of deaths (73, 101). Between 1898 and the first years of the 20th century, plague was imported to all continents travelling with the expanding steamships trading routes connecting colonial empires at the time. More than 100 countries recorded plague cases during the 3rd Pandemic, affecting mostly port cities and coastline provinces (61, 108). Plague invaded territories that had never been previously documented plague before the 3rd Pandemic, including North and South America, the South and West of Africa and Australia. Subsequent outbreaks caused more than 10 million deaths only in India and China alone (203). The timeline of the 3rd Plague Pandemic parallels the one of the developments of modern microbiology techniques and the discovery of the plague agent by Alexandre Yersin in Hong Kong in 1894. Since the late 19th and all along the 20th centuries, numerous *Y. pestis* isolates have been recovered by laboratories around the globe either in an epidemic context or for surveillance purposes (mostly isolates recovered from infected mammals and fleas). Morelli and colleagues called the genotypes of few hundred isolates from a global *Y. pestis* collection, based on conserved SNPs identified within few available genomes. It was observed that *Y. pestis* from the biovar *Orientalis* (1.ORI lineage), associated with the 3rd Pandemic, suffered multiple events of radiation in its phylogeny, that they interpreted as multiple waves of plague transmission in the broader pandemic context (46). The emergence of the 1.ORI lineage as the lineage causing the 3rd pandemic probably occurred in the Yunnan province, as the isolates sharing a common ancestor to the lineage were found in China (1.IN). A first wave of *Y. pestis* transmission defined by the 1.ORI1

sublineage spread from China to the neighboring India (where a first outbreak occurred in 1896), Vietnam, Indonesia (Java), as well as to the western coast of USA. The 1.ORI2 sublineage exhibits a radial expansion of *Y. pestis* isolates that were recovered from South America, Africa, Southeast Asia and Europe, where plague outbreaks were described in historical records although the precise origin of the 1.ORI2 ancestor is rather unclear. Due to the extensive sampling in the country, a third monophyletic sublineage (1.ORI3) was defined by the *Y. pestis* isolated from Madagascar. *Y. pestis* may have been imported to Madagascar by an infected ship from India (292) and was permanently established in the island, as isolates from the 1.ORI3 have been subsequently isolated since. 1.ORI3 appeared to have been also sporadically transmitted to the Middle East, although these transmissions could not create stable foci such as those present in Madagascar (203). After the radial expansion associated with 1.ORI lineage during the 3rd Pandemic, *Y. pestis* permanently established in several countries, creating plague endemic areas that persisted for decades, even until the present day. *Y. pestis* from different clades of the 1.ORI lineage have since been isolated along different decades in endemic countries such as Madagascar, Brazil, and the USA (46, 293, 294). Since *Y. pestis* from the 1.ORI lineage form monophyletic clades in these countries, it is considered that all extant *Y. pestis* circulating in these endemic areas of plague derived from a single successful introduction in each of these countries (203). The maintenance of plague foci that were established early during the 3rd plague pandemic has led to regular human plague activity in these endemic areas and this it is considered that the pandemic has not concluded, as we still live in it (2).

In addition to the inference of routes of plague transmission linked to the 3rd Plague Pandemic, whole-genome SNP-based phylogenies including *Y. pestis* isolates from the last century allowed to describe the transmission of other *Y. pestis* lineages not associated to the 3rd Pandemic lineage (1.ORI). Despite having emerged previous to 1.ORI, these lineages spread in some cases at the same time period as the 3rd Pandemic. The different sublineages

of *Y. pestis* from the 2.MED lineage appeared to be distributed along Asia, following the points corresponding to ancient human trade routes like the Silk Road. Consistent with the inferred emergence of 2.MED around the 15th century, this lineage may have travelled westwards to the Middle East following ancient trade routes (46, 247). However, the posterior finding of *Y. pestis* composing the most ancestral clade within 2.MED in the Caucasus region may indicate that the lineage was already widespread in Asia before this transmission occurred (62). The recent sequencing of multiple isolates from the 2.MED lineage in the Caspian Sea and Caucasus areas revealed the circulation of *Y. pestis* in the region, that suffered plague epidemic episodes with important mortality in the last century (64). The spread of *Y. pestis* from the 1.ANT lineage, restricted to Eastern and Central Africa, has also been subject of discussion. Sublineages of 1.ANT are distributed within Kenya, the DRC and Uganda, but the origin of the ancestor of the lineage remains unknown. It was hypothesized that infected ships from China could have carried the ancestor of this lineage to Africa early in the Middle Ages, although this hypothesis appeared less probable after inferring the emergence of 1.ANT around the 16th century (247). Plague may have been imported from somewhere in the Middle East to Eastern Africa by a route that is still unknown, as the origin of the emergence of this lineage remains obscure (295).

2.3.2. Spread of plague inferred from ancient DNA and paleogenomics

The understanding of past disease has benefitted from the use of diverse sources of information to reconstruct epidemic events and describe the health of past populations. Besides historical sources, archaeologists and anthropologists have studied skeletons and preserved bodies to study past diseases (240). For instance, the analysis of skeletal remains may reveal particular bone modifications or malformations that may indicate traces of infectious diseases such as tuberculosis (*M. tuberculosis*), leprosy (*M. leprae*) and syphilis (*T. pallidum*)(296). Concerning plague, in the last decades archaeologists and anthropologists have become more interested in plague burials, often studying mass graves and pits that were

in most cases chronologically associated with pandemics such as the 1st or the 2nd Plague pandemics (297, 298). The retrieval of DNA from skeletons of putative victims of ancient diseases may be used to identify past pathogens that allow to establish their cause of death. The recovery of such nucleic acids from skeletons, or ancient DNA (aDNA) research, can be interpreted as molecular fossils for past pathogens and microbiomes, permitting to understand their diversity, date their evolution and spread as well as determine their virulence potential (267). The sequencing of aDNA metagenomic samples and the reconstruction of ancient pathogen genomes has given birth in the last decade to the emerging field of paleogenomics and paleomicrobiology (240). aDNA may allow to reconstruct almost complete pathogen genomes using a combination of library enrichment and shotgun sequencing, followed by reads mapping to a reference genome (258, 261, 299). However, paleogenomics research faces several challenges due to the intrinsic characteristics of aDNA. Environmental DNA represents the majority of DNA present in the sample, and aDNA appears to be usually very fragmented and damaged due to deaminations of cytosines (C) to uracil (U), interpreted as thymines (T) by DNA polymerases during library generation (240). These drawbacks impose the need to generate huge sequencing libraries to recover enough genome depth, as well as the implementation of authentication methods allowing to unambiguously validate the ancient origin of the DNA sample (300). In 1998, ancient *Y. pestis* was detected for the first time by PCR targeting several loci, from DNA isolated from the dental pulp of victims found in a mass grave associated to the 2nd pandemic (301). After the development of NGS technology and authentication methods, the first ancient *Y. pestis* genome was reconstructed from skeletons recovered in London that had been associated to the Black Death epidemics (299). Both the detection of *Y. pestis* DNA and the reconstruction of the first ancient genome provided for the first-time unambiguous evidence proving the implication of *Y. pestis* in historical plague pandemics such as the 2nd pandemic. Since these early works, paleogenomics and paleomicrobiology studies have provided multiple and interesting insights

regarding the evolution and historical spread of *Y. pestis* in the past.

Paleogenomics of the 1st Plague Pandemic/Justinian plague (541 – 8th century). The 1st Pandemic of Plague of Justinian, named after the emperor of the Roman Empire, Justinian I, began in 541, spreading around the Mediterranean area during three years in the first wave of the pandemic. In the further centuries, multiple outbreaks succeeded in consecutive waves lasting until the 8th century, affecting first port cities and then penetrating deep inland (**figure 2.7**) (2). Although the impact this pandemic had on civilizations present at the time remains controversial, some authors estimate the death toll of the 1st Pandemic in between 15 to 100 million victims (302, 303). About a dozen *Y. pestis* ancient genomes associated to the plague of Justinian have been recovered in the recent years in Germany, England, France, Spain and Kyrgyzstan (258, 259, 304, 305). All 1st Pandemic ancient genomes fall in the same clade in *Y. pestis* phylogeny, branching between contemporary lineages called 0.ANT1, 0.ANT2 and 0.ANT5. The finding of samples from the same period of time and associated pandemic context in diverse locations in Europe suggested that plague was widespread across the continent at the period, contrary to the historical sources that suggested that the pandemic was restricted to the Mediterranean (73). A 45-kb and 49-kb deletions were described in few derived strains within the clade of the 1st pandemic (0.ANT4). These deletions affected the *mgtB* and *mgtC* loci encoding transporters that may be important for *Y. pestis* virulence (304). The origin of the Justinian Plague remains subject of discussion, as historical sources described the outbreak origins in Egypt, further spreading to the Eastern Mediterranean and then to Europe (306, 307). However, the placement of the lineage associated to the 1st pandemic between different 0.ANT clades, composed by contemporary isolates present in China and Kyrgyzstan, suggest a Central Asian origin for the pandemic (51, 247). The recovery of ancient *Y. pestis* genomes from Africa and the Eastern Mediterranean from the 6th century may help to solve the route of entry of the 1st pandemic in Europe.

Paleogenomics of the 2nd Plague Pandemic, including the Black Death (1346-18th century). It is considered that the 2nd Pandemic began somewhere in central Asia by the 1330s (247, 262). After the first outbreaks in Western Asia, plague spread to Northern Africa and Europe and within 7 years of a first epidemic known as “Black Death” (1346-1353), plague killed between 30-50% of the European population at that time (308). From the 14th to the 17th century, plague became a common disease in Western Europe (figure 2.7), to become more rare in the 18th, when it extinguished from Europe for reasons that are yet not fully understood (2). More than 50 ancient *Y. pestis* genomes have been recovered from victims of the 2nd Plague Pandemic. The phylogenetic placement of these samples has revealed interesting details regarding the evolution and spread of this pandemic. First, the few genomes associated to the Black Death period fall together early in the branch 1 of *Y. pestis* tree, suggesting that the Black Death outbreak was caused by the spread of a major clone (261). Recently, the probable source for the Black Death outbreak has been described thanks to the recovery of ancient genomes from plague victims in present-day Kyrgyzstan in 1338-39. This suggests that the *Y. pestis* that caused the Black Death few years later had emerged from the plague foci present in Central Asia, foci that are still active until the last decades represented by isolates of the O.ANT lineages (264).

Second, the ancient genomes recovered from plague victims in Europe in the following 16-18th centuries fall together in the same branch immediately after the Black Death samples, suggesting that a common lineage of *Y. pestis* circulating in Europe had been evolving for centuries and appears now to be extinct, as no contemporary representative isolates have been identified (262). The interpretation of the observed phylogenetic patterns during the 2nd Pandemic regarding the spread of plague in Europe still causes controversy, as two hypotheses have been proposed. Some scholars have proposed that plague was introduced into Europe in multiple occasions from a reservoir near in Asia, given the level of genetic diversity and the support of climatic indicators (260, 309). This hypothesis would aim to explain the

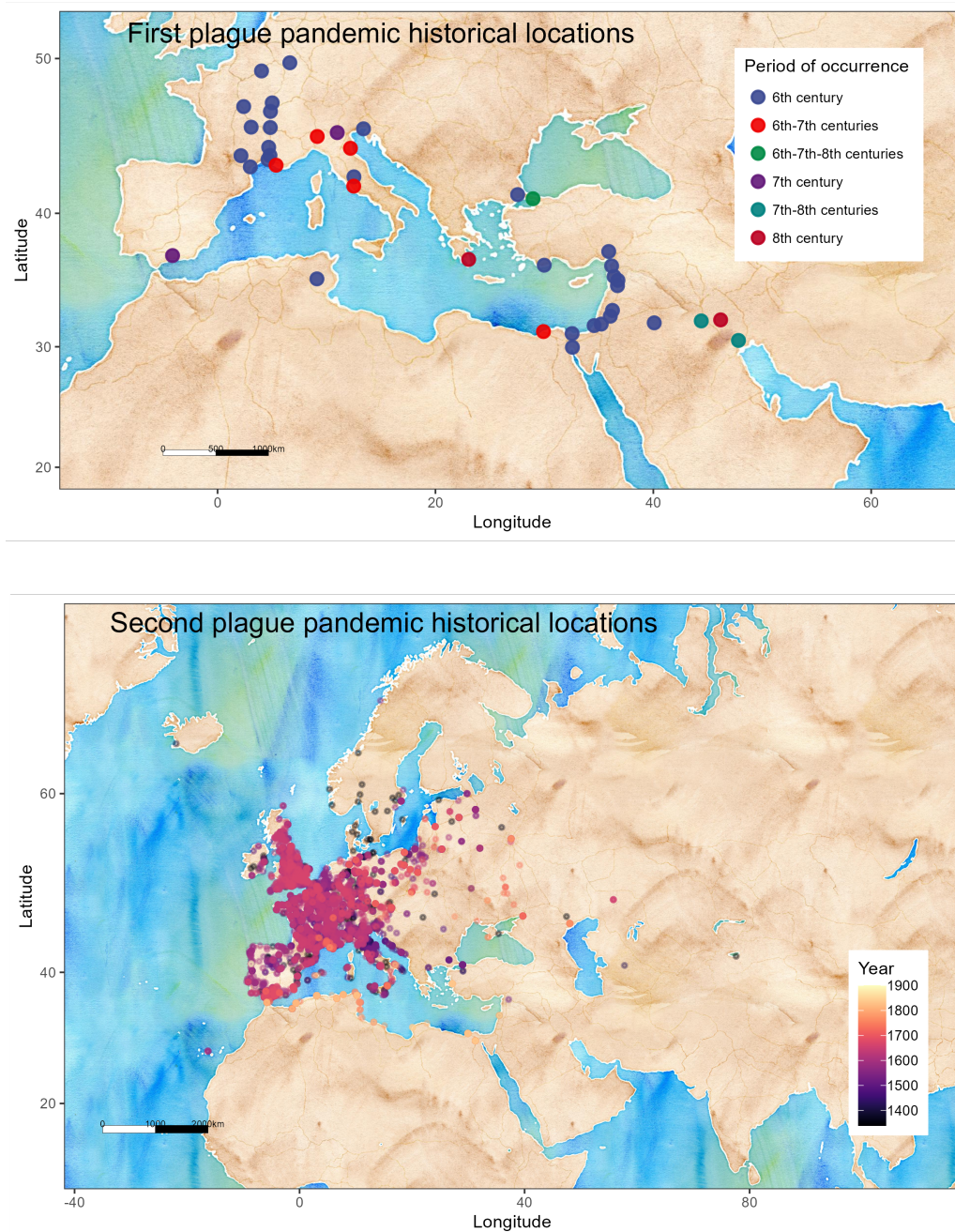


Figure 2.7. World maps of the First and Second plague pandemics. Up: Locations where there are historical mentions of the First plague pandemic or plague of Justinian (6th to 8th centuries). **Down:** Map of the Second plague pandemic (including the Black Death, 1346 to the 18th century) in Europe. Each point represents a putative plague outbreak based on historical records. Coordinates of the locations from historical sources used to make the figure were kindly provided by Rémi Barbieri, previously published in a complete review (2).

multiple epidemic waves described along the centuries. On the other hand, the hypothesis of a single introduction of plague in Europe has been proposed, which would represent the existence of stable plague animal foci within or nearby Europe explaining the persistence of the disease through four centuries. This hypothesis is supported by aDNA data, as Black Death genomes fall together at the same point in the phylogeny, and subsequent genomes form a derived branch after this first outbreak samples (261, 262). It is interesting to note that the Black Death genomes fall early in Branch 1, representing the ancestors of the derived *Y. pestis* that caused the 3rd Pandemic in the late 19th century in China (260, 262). Still, the precise location of plague foci that could explain the maintenance of plague in Europe during the 2nd Pandemic, as well as the reasons why plague disappeared from Europe after the 18th century are still open questions.

Paleogenomics of prehistoric plague and the origin of *Y. pestis* emergence.

Ancient *Y. pestis* genomes have been mostly recovered from locations and time associated to historical plague pandemics, as historical accounts are available to support the molecular evidence of the epidemics given by aDNA. However, aDNA can also provide evidence for the existence of *Y. pestis* infection thousands of years before the first documented plague pandemic. The most ancient *Y. pestis* genomes recovered date back to around 5,000 years ago in Latvia and Sweden, confirming the existence of probably life-threatening *Y. pestis* infections in a period and context without indicators of epidemics (i.e.: mass graves)(266, 269). In the recent years, several prehistoric *Y. pestis* ancient genome data have been recovered, leading to the hypothesis of an emergence of prehistoric plague epidemics. In a recent work, Andrades Valtueña and colleagues described the emergence and persistence of a Late Neolithic/Early Bronze Age (LNBA) *Y. pestis* lineage that was widespread in Europe, but likely became extinct. The study also showed the presence of divergent prehistoric *Y. pestis* genomes contemporary to the LNBA lineage that led to the emergence of lineages that persist up to the present day (270). The genomic analysis of prehistoric genomes led

to the discovery of a fragment in pMT1 plasmid including the *ymt* locus that was missing in most ancestral *Y. pestis* genomes and the LNBA lineage. The absence of *ymt*, as well as the active form of the *ureD* urease gene and several biofilm regulator loci led to the hypothesis that these prehistoric *Y. pestis* were not yet fully adapted to colonize fleas, and therefore, that bubonic plague may not have been the form of the disease caused by these strains (265, 266). Since Neolithic *Y. pestis* genomes represent the most ancestral branches in the species phylogenies, their sources have also been considered for the understanding of the possible origin of the emergence of *Y. pestis* from *Y. pseudotuberculosis*. Some of the *Y. pestis* lineages that emerged early and persisted up to the present are only found in China and/or in some foci in Central Asia. This observation, together with the remarkable genetic diversity and widespread presence of Chinese isolates in the phylogeny led to the believe that *Y. pestis* may have emerged somewhere in China, and later spread abroad following ancient human trade routes (247). On the other hand, aDNA studies found the most ancestral *Y. pestis* emerging in Europe around 6,000-7,000 years ago and the quick diversification of *Y. pestis* in the Neolithic (269, 270). The recovery of more prehistoric ancient *Y. pestis* from Central and Eastern Asia may provide critical clues for the understanding of the possible source of speciation of the pathogen from its common ancestor with *Y. pseudotuberculosis*.

Section 2.4

**Molecular typing and genomic epidemiology of
plague**

Microbiologists have for long sought for markers that allow to subdivide pathogens into groups, as means to discriminate between strains. Discriminatory makers can be used for epidemiological purposes to i) test whether outbreaks share a common source, ii) track epidemic spread of particular strains, and iii) understand the long-term outline of geographic and epidemiological patterns of infectious diseases (242, 310). Tracking the source of a contamination is especially relevant when considering the potential use of bioterror agents such as *Y. pestis*, which is considered a category A select agent (131, 311). Molecular markers, mainly based on DNA variations, have provided with a myriad of typing methods that have been implemented in many diverse bacterial species (312). However, some of these typing methods are not suitable for subdividing genetically monomorphic microbes such as *Y. pestis* due to their low diversity (242, 252). Despite this challenge, due to the historical significance, extreme virulence, and potential threat of *Y. pestis* infections, its typing has been subject of many efforts. Some of the techniques that have been used for typing *Y. pestis* isolates include pulse-field gel electrophoresis (PFGE), the analysis of the presence of insertion sequences (IS fingerprinting), the analysis of CRISPR sequences, and the analysis of microsatellite repetitive sequences in MLVA (252, 313–316). However, these methods face several problems, including standardization between laboratories, phylogenetic consistency, and frequency of homoplasies (242). The advent of WGS, its implementation into many research and surveillance laboratories, as well as the development of different genomics methods have gradually pushed to abandon the molecular typing methods used in the past to study *Y. pestis*. SNPs identified from whole-genome sequences (wgSNPs) are considered the most resolutive and consistent markers of the molecular evolution of *Y. pestis*, allowing to track the evolutionary

and transmission history of this pathogen (310). Whole-genome sequence data of new isolates can help identifying novel SNPs compared to a reference, as well as characterizing other genetic features that may be relevant to understand their diversity and virulence potential (i.e.: indels, gene presence/absence, rearrangements. . .) (317, 318).

Several SNPs defining major *Y. pestis* lineages have been identified and are useful to attribute novel isolates to known populations (46, 317). However, to date no genotyping scheme based on wgSNPs has been developed for the subdivision of *Y. pestis* into genotypes to provide direct informative data on the genetic and geographic relationships between samples. The absence of a standardized genotyping method represents that currently the only way of classifying *Y. pestis* genomes into lineages is through the reconstruction of whole-genome phylogenies incorporating samples from different prior studies (62, 287). Most research and public health laboratories have hence directly relied on phylogenetic analyses in order to attribute novel *Y. pestis* genomes to a known lineage and identify their SNPs (294, 319). Recently, a novel core-genome MLST (cgMLST) typing method has been implemented for typing *Y. pseudotuberculosis* isolates in outbreak context (320), and a cgMLST clustering approach (HierCC) has been implemented into the Enterobase database for the genotyping of *Yersinia* species (321). Given reduced diversity described in *Y. pestis*, it may be difficult to extend of these schemes for the resolutive typing of *Y. pestis*. On the other hand, genotyping frameworks based on wgSNPs have been developed for major human pathogens such as *Mycobacterium tuberculosis* (322), *Shigella sonnei* (256) and *Salmonella enterica* serovar Typhi (323). These bacterial species are also genetically monomorphic or display low levels of genetic diversity comparable to *Y. pestis* (242). Such frameworks permit a fast and accurate typing of clinical or surveillance isolates directly from WGS without the need for time-consuming and not always user-friendly comparative genomics and phylogenetics methods (256). As WGS becomes available in an increasing number of laboratories for the sequencing of *Y. pestis* isolates, the development of a hierarchical wgSNP-based framework

for the typing of *Y. pestis* would facilitate a straightforward identification of lineages and sublineages from WGS data, helping the fast tracking of *Y. pestis* strains in outbreaks and emergency contexts.

Indeed, many epidemiologically relevant questions can be addressed using pathogen genomic data. What is the origin of an outbreak and the disease mode of transmission? Is an outbreak related to another previous or distant one? How many introductions of the pathogen into human communities there have been? What may be the drivers of (re)emergence of outbreaks? Providing answers to these questions remains critical for the development of intervention strategies to control infectious diseases (324). The integration of epidemiological data with pathogen genomic data to investigate infectious diseases has given name to the emerging field known as genomic epidemiology (325). Microbial WGS data has hence provided with unique opportunities to track the possible source of pathogen strains during outbreaks (190) to provide insights into the emergence and development of epidemics (326, 327), to characterize genetic features of circulating pathogen groups (328), and to better understand the emergence and dissemination of antibiotic-resistant lineages (329). Such evolutionary and genomic epidemiology studies are usually based in popular workflows for finding SNPs within bacterial genomes, building phylogenetic trees and other evolutionary reconstructions, comparative genomics and the integration of temporal and geographic data of the isolates. In most of the recent works using WGS for the genomic epidemiology and evolution of *Y. pestis*, similar strategies have been used. Several studies have attempted to characterize the diversity of *Y. pestis* based on isolates from historical collections from different geographic sources (46, 62, 247), while others have focused on isolates recovered from surveillance efforts in the same plague endemic areas along several decades (271, 293, 294). Shotgun sequencing is usually performed on genomic DNA to generate short reads in Illumina-based libraries, and reads are assembled into contigs using *de novo* assembly approaches. Then, wgSNPs and indels are then called by mapping the reads to

the reference chromosome *Y. pestis* strain CO92 (330), and positions that have low depth or fall within repetitive sequences and non-core regions are filtered out. The alignment of concatenated wgSNPs can be used to build a phylogenetic tree using maximum-likelihood algorithms (331, 332), or to infer population structure relationships by minimum spanning distance trees (333). When the isolation time of the bacterial genomes is known, different phylogenetic dating approaches can be implemented to infer divergence dates for the different branches in the phylogeny (281, 282, 334). Such analyses may allow to define different genetic groups present in the dataset, as well as to infer their dynamics of emergence and maintenance through time. The study of the relationships between genome sequence evolution and geography of infectious diseases has been addressed in the emerging research field known as phylogeography (335). If both the sampling time and location coordinates of isolation of the different microbial isolates are known, the dissemination of pathogens from one location to another can be deduced from dated phylogenies. Some phylogeographic models have been developed in order to explicitly assign geographic locations to unsampled ancestral isolates represented by internal nodes in the tree, that represents the collection of pathogen movements that are phylogenetically linked (336, 337). Phylogeographic approaches can hence be used to draw the patterns of pathogen spread, providing unique insights into the spatial dynamics of diseases during epidemics (335, 338). This information can then be applied for evaluating the possible association of external factors on the maintenance and spread of pathogens, which is critical to inform possible counteracting measures to control and prevent the spread of ongoing or future outbreaks (339, 340).

Although many of these approaches have become popular to study the temporal and spatial dynamics of different viral infections, their implementation for the study of *Y. pestis* dynamics has been less recurrent. Some studies have focused on the molecular typing of isolates in plague endemic areas to track the possible source of *Y. pestis* during local plague outbreaks (319). This has been usually achieved by calling wgSNPs to both determine the

main phylogenetic lineage of the isolates as well as to describe more refined subgroups within them. For instance, this strategy has permitted the identification of new sublineages circulating in known endemic areas (272, 341), and to trace human plague cases back to different animal reservoirs and intermediate hosts associated to plague in the region (47, 48, 342). Other works have focused on understanding the spatial, temporal and evolutionary dynamics of *Y. pestis* in endemic countries. Phylogeographic analyses have illustrated the dynamic nature of natural plague foci, describing changes in the abundance of local genetic groups and fluctuations in plague epidemics (343). The sequencing and genomic analysis of historical isolates in the Caucasus and Caspian Sea plague foci revealed the expansion and reactivation of plague foci in the area as a result of the reintroduction and long-term establishment of a 2.MED1 sublineage in the second half of the 20th century (64). Interestingly, phylogenomic studies have also revealed common evolutionary patterns in endemic countries in which plague was established during the 3rd Plague Pandemic in the late 19th century/early 20th century. In such countries, *Y. pestis* was introduced during the pandemic spread and successfully established in permanent in sylvatic rodent populations, creating stable endemic foci and recurrent risk of human plague (46). After this successful introduction in the 3rd pandemic, a shared feature within several plague foci studied is the quick spread within the country (invasion phase), which appears as a radial expansion (i.e.: polytomies) and a low degree of population structure in phylogenies (294). This phase is followed by the formation of genetically and geographically restrained groups that display spatial and temporal stability, emerging as clades or sublineages in the phylogeny (344). The genomic investigation of *Y. pestis* isolates from the same plague endemic areas covering several decades has been developed in Madagascar, one of the countries where plague is most active. In Madagascar, higher levels of *Y. pestis* genetic diversity were observed in areas with higher elevation, suggesting environmental factors play an important role on the maintenance of plague in this country (293). The analysis of isolates from long-term plague surveillance may also

reveal the transient spread of *Y. pestis* from active endemic areas into areas normally not affected by plague, as it was observed in the plague outbreaks in the port city of Mahajanga in Madagascar along the 1990s decade ([29](#)).

Part II

Objectives

1 – Large-scale characterization of the patterns of global genetic and geographic diversity of *Yersinia pestis* populations and their functional diversification

Genomic analyses represent powerful tools to study infectious diseases, allowing to discriminate between circulating strains, infer the origins of dispersion and identify genetic determinants of their virulence and adaptation mechanisms. *Yersinia pestis* is a clonal pathogen with reduced levels of genetic diversity, and hence it is necessary to use whole-genome sequencing data to reveal finer patterns of genetic change between isolates. A few phylogenomic analyses in the past decade have described the major lineages in *Y. pestis* species evolution and given a preliminary cartography of their geographic distribution. In parallel, within the recent years the field of microbial paleogenomics has emerged and provided for the first time with genomic data from *Y. pestis* present in infected human remains over the past millennia, including the strains that circulated in major plague pandemics in human history. However, the extent of the global genetic diversity of *Y. pestis* has not been fully explored, as some important endemic areas that remain active in the past decades remain poorly covered in past studies, and some identified lineages include only few representative genomes. Moreover, there have been few studies attempting to characterize the patterns of genomic variation affecting gene function in the microevolution of *Y. pestis* species. These previous attempts have been based in genomic comparisons of some tenths of genomes, which may have obscured the identification of variants present in lineages and geographic areas absent from the analyses.

The first general goal of the present thesis was to generate a large genomic dataset allowing to perform comparisons of *Y. pestis* genomes at the species level, including the diversity that has been made available in the recent years with the inclusion of ancient DNA datasets. For this purpose, we envisioned the whole-genome sequencing of a global collection of more than 1000 *Y. pestis* isolates maintained by the Yersinia Research Unit and the National Ref-

erence Laboratory for Plague & other Yersinioses at the Institut Pasteur. The initial phase of the project requiring the sequencing of such a large collection of isolates was developed together with our colleagues in the Environment and Infectious Risk – CIBU unit at the Institut Pasteur, who manage the sequencing platform. To integrate the tenths of available ancient DNA metagenomes of *Y. pestis* and generate the most complete phylogenetic and genomic analysis of *Y. pestis* to date, we established a close collaboration with our colleagues in the Microbial Paleogenomics Unit (MPU) in the Institut Pasteur, headed by Nicolas Rascovan. The development of different bioinformatic workflows to process the complex genomic datasets of *Y. pestis* (ancient and modern sequencing data) was determinant for this first objective of this thesis. This collaborative effort allowed the generation of whole-genome sequences of more than 1100 *Y. pestis* isolates collected from different sources, which was required for this project and for setting up the basis of new tools for the genotyping of *Y. pestis* isolates in future works. We used population genomics methods to develop the most extensive analysis to date of *Y. pestis* diversity, evolutionary history, and global spread. We developed an integrative approach to interrogate different mechanisms in *Y. pestis* evolution with the aim of deciphering important factors for the emergence of lineages and their functional diversification.

2 – Study of the dynamics of *Yersinia pestis* in the contexts of epidemic and endemic circulation using advanced phylogeographic approaches

The development of genomic and phylogeographic methods in the recent years has permitted the reconstruction of pathogen dissemination during outbreaks, revealing the dynamics behind the spread of infectious diseases. Such information inferred from models based on

the combination of pathogen evolution and temporal-geographical data can be used to inform about the complex patterns of maintenance and spread of diseases, which in turn may contribute to the response to emerging and reemerging infections such as plague. Concerning plague, few endemic areas have been studied in detail in the recent years, including the ones present in Madagascar, China or Brazil, where phylogenomic analyses have in some cases helped to reveal local factors associated to the dynamics of plague in these areas. In addition, recently developed landscape phylogeographic approaches, allowing to retrace pathogen spread in a continuous time and space, have never been applied to study *Yersinia pestis* dispersion in epidemic and endemic context.

In the second part of this thesis, we apply a phylodynamic and phylogeographic approach to investigate the evolution and spread of *Y. pestis* in Vietnam, using hundreds of isolates recovered across four decades in the context of extensive human plague outbreaks and animal surveillance efforts. To develop this research, we worked together with Simon Dellicour (Université Libre de Bruxelles, KU Leuven), whose pioneering work on the phylogeography of infectious diseases and critical advices were important to fulfill our study. Despite having accounted for up to 90% of the worldwide burden of plague during several years of the last century, there are currently no available molecular studies that characterize the spread of *Y. pestis* in Vietnam. Major outbreaks of plague were recorded in the country at the height of the long-standing Vietnam War of 1955-1975. The aim of this part was to place the recovered *Y. pestis* isolates in a spatio-temporal context, providing information on the linkage of infections through time and space in the epidemic and endemic circulation of *Y. pestis* in this country. The inference of dispersal events would then be a valuable resource for future analyses aiming to illuminate possible factors that influenced the spread of plague in Vietnam.

Part III

Results

Chapter 3

Results

Section 3.1

Manuscript A - Genetic diversity, evolution and spread of plague revealed through the analysis of global *Yersinia pestis* genomes

Next will be presented the manuscript detailing the results of the first part of this thesis, which explores the evolution and diversity of *Y. pestis* at the species level. This investigation is currently ongoing, and an updated version of the manuscript (including further results) is intended to be submitted for publication in the following months.

Genetic diversity, evolution and spread of plague revealed through the analysis of global *Yersinia pestis* genomes

Guillem Mas Fiol¹, Aida Andrades Valtueña^{2,3}, Frédéric Lemoine³, Damien Mornico³, Cyril Savin^{1,4,5}, Charlotte Balière^{6,7}, Aurélia Kwasiborski^{6,7}, Valérie Caro^{6,7}, Rémi Beau^{1,2,3,6}, Sebastian Duchene⁸, Manuel Céspedes⁹, Minoarisoa Rajerison¹⁰, Jean-Christophe Shako¹¹, Elisabeth Carniel¹, Nicolás Rascován³, Javier Pizarro-Cerdá^{1,4,5}

¹Yersinia Research Unit, Institut Pasteur, Université Paris Cité, CNRS UMR 6047, F-75015 Paris, France; ²Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ³Microbial Paleogenomics Unit, Institut Pasteur, Université Paris Cité, Paris, France; ⁴World Health Organization Collaborating Center for Plague FRA-140, Institut Pasteur, F-75015 Paris, France; ⁵Yersinia National Reference Laboratory, Institut Pasteur, Université Paris Cité, CNRS UMR 6047, F-75015 Paris, France; ⁶Environment and Infectious Risk Unit, Institut Pasteur, Université Paris Cité, Paris, France; ⁷Cellule d'Intervention Biologique d'Urgence (CIBU), Institut Pasteur, Université Paris Cité, Paris, France; ⁸Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, University of Melbourne, Melbourne, Victoria, Australia; ⁹Bacterial Zoonoses Laboratory, National Institute of Health, Lima, Peru; ¹⁰Plague Unit, Institut Pasteur de Madagascar, Antananarivo, Madagascar; ¹¹Plague Reference Laboratory, Bunia, Congo

Abstract

The zoonotic pathogen *Yersinia pestis*, the agent of human plague, has caused three major pandemics and established in endemic areas present worldwide that are at risk of plague reemergence. We used whole-genome sequences of 1549 *Y. pestis* isolates from 29 countries collected over more than 100 years, and together with 1257 publicly available genomes from modern and ancient genomes, we conduct the largest phylogenomic analysis of *Y. pestis* to date. We refine the species population structure and infer routes of spread of plague in regions that were overlooked in previous studies, notably in Africa, where multiple introductions are retraced to the Third Pandemic. We also characterize the expansion rates and functional diversification within lineages, identifying unsuspected trends within the previously poorly described 1.ANT population in Africa. By using different genomic approaches, we identify specific gene repertoires in different lineages and characterize the expansion of insertion sequences (IS) during *Y. pestis* evolution, particularly within the 1.ANT lineage where an enrichment in IS-driven gene interruptions is described. Our work provides unprecedented insights into the evolutionary mechanisms and patterns that have shaped the current diversity within the *Y. pestis* species.

Introduction

Plague is a fatal zoonotic disease caused by the gram-negative bacterium *Yersinia pestis*, responsible of three major pandemics that have profoundly impacted human history (1). *Y. pestis* common ancestor evolved 5000-6000 years ago (2) and since, *Y. pestis* has spread from Asia in multiple occasions. During the third plague pandemic, plague reemerged in China and spread worldwide (3), and today plague is maintained in endemic areas that are

currently present in Africa, Asia, Eastern Europe, and the Americas (4, 5). Major plague endemic areas include Madagascar (6, 7), the Democratic Republic of Congo (8, 9) and China (10, 11), where outbreaks of human plague have occurred in the last decades. Plague is a vector-borne disease, transmitted to humans by fleas from rodent reservoirs (12). Bubonic plague is the major form of the disease, produced upon bacterial unrestricted growth in local draining lymph nodes (1). Pneumonic plague is a less frequent and more aggressive clinical manifestation in which *Y. pestis* proliferates in the lungs, potentially leading to inter-human transmission of the disease (13). Septicemic plague is also a highly severe form that arises when bacteria reach the blood system (14).

Considered as one of the most virulent bacterial pathogens to humans (1), *Y. pestis* is a recently emerged clone of its enteropathogenic ancestor *Y. pseudotuberculosis* (15) that evolved through the combination of events of gene gain (including two virulence-associated plasmids) and genome reduction (16, 17). These changes allowed *Y. pestis* to colonize an insect vector and to establish new transmission routes to infect mammal hosts (18–20). In the recent years, genomic analyses have started to retrace the clonal evolution of *Y. pestis*, leading to the identification of major lineages from the sequencing of bacterial isolates circulating in the past century (21–24), as well as from the analysis of ancient DNA metagenomes recovered from victims of historical plague pandemics (25, 26) and until prehistoric times (27, 28). The steady expansion of *Y. pestis* genomic data has led to a stepwise increase in the genetic diversity present in *Y. pestis* phylogenies, in which lineages that have been circulating in the past decades are interspersed between the emergence of extinct lineages that caused pandemics and outbreaks centuries ago (29). However, previous attempts to characterize the global genetic diversity of *Y. pestis* were based on the analysis of a reduced number of genomes from isolates that were recovered mainly in areas within Asia (3, 21–23), while some regions in the world where plague actively circulates remained underrepresented. Moreover, the analysis of *Y. pestis* genetic diversity has mainly focused on the characterization of point mutations, ignoring other mechanisms that may be important drivers of genomic variation within *Y. pestis* isolates.

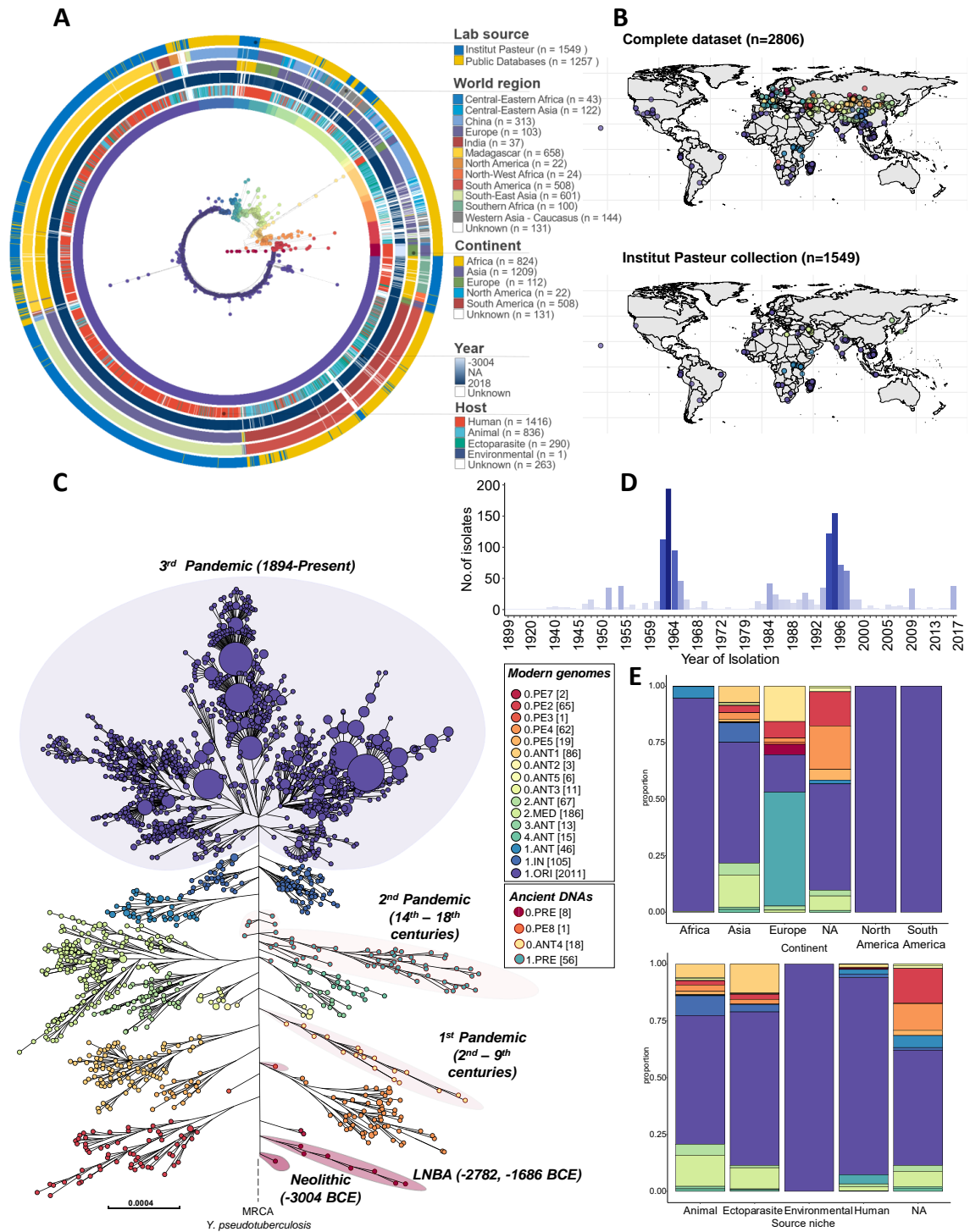
Here, we take steps to overcome these limitations by sequencing 1549 novel *Y. pestis* genomes of isolates from several important endemic regions worldwide that were collected all along the past and current century. We use population genomics methods to develop the most extensive analysis to date of *Y. pestis* diversity, evolutionary history, and global spread. We take an integrative approach to interrogate different mechanisms impacting *Y. pestis* evolution, deciphering important factors for the emergence of lineages and their functional diversification.

Results

Population structure and diversity of *Y. pestis* lineages

We sequenced 1549 novel *Y. pestis* genomes originating from 29 different countries that cover important plague endemic areas in the world, including Africa, the Middle East, South-East Asia and South America (fig. 1AB, fig.S1 dataset S1). Isolates were mostly recovered from human clinical cases of plague (n=1087/1549, 70.2%), from which a minority were isolated from pulmonary plague cases (56/1087, 5.2%). The rest of novel isolates analyzed in this study were recovered in the context of animal plague surveillance, including a range of rodent species (239/1549, 15.4%) and fleas (88/1549, 5.7%) (fig. S1). Animal hosts include different species of rats (*Rattus* sp. 168/239, 70.3%) and gerbils (*Meriones* sp. 29/239, 12.1%), while the oriental rat flea is the most common ectoparasite vector present in the dataset (*X. cheopis* 57/88, 64.7%) (dataset S1). The collection covers a time span of 118 years (1899-2017), with the oldest samples closely following the first isolation of *Y. pestis* in 1894 (31). Most were isolates sampled all along the second half of the 20th century (fig. 1D, fig. S1).

To provide an updated overview of the *Y. pestis* phylogenetic diversity and population structure, we built a maximum likelihood phylogeny including most of the publicly available modern genomes (n=1174), together with the 1549 novel sequences. We implemented a workflow (see Methods) allowing to include ancient DNA (aDNA) *Y. pestis* metagenomes (n=83) recovered from victims past *Y. pestis* infections, including historical pandemics, to the genomic analysis of modern isolates. This allowed the reconstruction of the largest phylogeny of *Y. pestis* species to date, using a total of 2806 modern and ancient genomes (fig. 1AC, Dataset S2). The tree is divided into 20 different lineages following a previously established nomenclature (3, 21, 23, 29) and confirmed the overall clonal evolution of *Y. pestis*, with lineages emerging from a central branch (fig. 1C, fig. S2). These lineages have uneven frequencies, as 1.ORI (n=2036) represents the 72.5% of the sampled isolates, while some lineages



are represented by less than 20 isolates: 0.PE5 (n=19), 4.ANT (n=15), 3.ANT (n=13), 0.ANT3 (n=11), 0.ANT5 (n=6), 0.ANT2 (n=3), 0.PE7 (n=2), 0.PE3 (n=1) (fig. 1, fig. S2). Congruent with previous studies (2, 25, 26, 32), the phylogeny also placed branches corresponding to ancient *Y. pestis* genomes from the Neolithic and Late Neolithic as the most ancestral in *Y. pestis* evolution, while branches corresponding to samples associated to the Justinian Plague and to the Second Pandemic fall at the trunk of branch 0 and of branch 1, respectively.

The different lineages of contemporary *Y. pestis* genomes display contrasting levels of average nucleotide diversity index (π) (Kruskal-Wallis test, $p < 2.2e-16$, fig. S3). We found 0.PE lineages to be the most diverse, consistent with the placement of 0.PE lineages within the most ancestral clades of the phylogeny, while more recently evolved lineages (i.e.: 1.ORI, 1.IN) have accumulated less diversity. Surprisingly, we found 1.ANT to be within the lineages presenting the highest diversity, despite its more recent emergence as part of branch 1 (fig. S2, fig. S3).

The global *Y. pestis* phylogeny also displays a strong geographical structure of lineages, in line with previous observations (3). 1.ORI lineage is found in all continents except Australia; 1.ANT lineage is restricted to Central and Eastern Africa, while all the other lineages (0.PE, 0.ANT, 1.IN, 2.ANT, 2.MED, 3.ANT, 4.ANT) are almost exclusively present over Eurasia (fig. 1, fig. S4). The extensive sampling and widespread distribution of 1.ORI is consistent with its dissemination during the Third Pandemic (fig. 1C)(3). The comparative analysis of thousands of 1.ORI genomes provided additional evolutionary substructure within this lineage. 1.ORI lineage presents a star-like topology, defined by the radiation of different sublineages that were interpreted as reflecting different waves of plague expansion during the Third Pandemic (fig. 1AC)(3). The most ancestral sublineage, 1.ORI1, is present in China, South-East Asia and North America. A polytomy following 1.ORI1 gave rise to multiple sublineages of diverse abundances. 1.ORI2 (n=12) is found only in India and South Africa while 1.ORI3 (n=668) covers all genomes from Madagascar and four from the Middle East. We identified two extensively sampled monophyletic sublineages that we named 1.ORI4 and 1.ORI5. 1.ORI4 (n=593) is present in South-East Asia, and 1.ORI5 (n=675) is present in Africa, South America and Europe (fig. 1, fig. S5). We sequenced 43 new genomes of the 1.ANT lineage (n=46) that were distributed within three sublineages following previous descriptions (3). The early-branching 1.ANT3 (n=11) is found in Yemen, Kenya and the Democratic Republic of Congo (DRC). 1.ANT2 (n=14) is present in Kenya and Tanzania, while 1.ANT1 (n=21) circulates within the plague endemic area of Ituri (DRC) and neighboring Uganda (fig. S6). The phylogenetic analysis also confirms the subdivision of the 2.MED lineage (n=186, widely distributed over Asia and the Middle East) into four sublineages: 2.MED0 (n=8), 2.MED1 (n=147), 2.MED2 (n=6) and 2.MED3 (n=25). We generated 60 novel genomes of the 2.MED1 sublineage that were recovered from the Middle East and Libya, in the context of animal plague surveillance (Kurdistan in Iran) and of human plague outbreaks (2009 plague outbreak in Libya (33), (fig. S6).

Retracing the routes of plague spread

The reconstructed phylogenies, including the new and previously published isolates, provide new insights into the evolution and spreading history of the *Y. pestis* lineages, particularly 1.ORI (Third Pandemic), 1.ANT, and 2.MED1 (fig. 2). We selected a subset of representative ancient and modern *Y. pestis* genomes (table S1) and used the Bayesian framework implemented in BEAST (34) to date the divergence of different clades (fig. S7). The Third Plague Pandemic is associated with the emergence of 1.ORI lineage in China, followed by its global dissemination starting in 1894 (3, 35). Our estimations of the evolution of 1.ORI lineage (tMRCA 1759 AD; 95% highest posterior density (HPD) 1695-1814) slightly predate previous estimations (21), but the interval of both estimations overlap. As it has been previously described (3), 1.ORI1 was introduced early in the Third Pandemic into South-East Asia and North America, and 1.ORI2 spread to India and South Africa. These transmissions were followed by the introduction of 1.ORI3 into Madagascar and the Middle East, and of 1.ORI4 into Vietnam and Myanmar (fig. 2A, fig. S5). These observations support the proposed view of different waves of plague spreading during the Third Pandemic (3)(fig. 1, fig. S5).

The extensive sampling of Third Pandemic isolates revealed new patterns of spread within the newly described 1.ORI5 sublineage. 1.ORI5 emerged more recently and radially spread to Africa, South America, and Europe (fig. 2BC). Plague may have invaded the Southern region of Africa from the coast in South Africa because two descendant clades of 1.ORI5 include isolates from South Africa, Zimbabwe and Namibia (fig. 2BC). We es-

estimate this transmission early in the 20th century (Zimbabwe-Namibian clade: 1904 AD; 95% HPD 1857- 1942), consistent with historical records describing the first plague outbreaks in the South African coast in 1900 and further inland expansion in the context of the Anglo-Boer war of 1899-1902 (36). 1.ORI5 may have also invaded Western Africa on multiple occasions, since we find two independent clades containing isolates from Senegal. Plague probably spread from the Western African coast into Northern Africa because a descendant clade of 1.ORI5 consists of isolates from Senegal as well as from Morocco and Algeria (fig. 2B). This clade diverged in the late 19th century (1866 AD; 95% HPD 1816-1899), coinciding with the recorded plague invasions of the Northern African coast at the turn of the 20th century (37, 38). Multiple independent radiations of 1.ORI5 were responsible for the transmission of plague into South America (Argentina, Bolivia, Brazil, Peru) and Europe (Germany). A clade of 1.ORI5 including isolates from Bolivia, Germany and Argentina may illustrate the recurrent introduction of plague into European ports by infected ships from former colonies such as Buenos Aires (39) (fig. 2B). Overall, the patterns of 1.ORI5 spread follow the described behavior of the early Third Pandemic, repeatedly infecting major ports and causing subsequent urban outbreaks (40, 41).

We also report the spread of 1.ANT, the first lineage emerging after the ancient *Y. pestis* genomes that caused the Second Pandemic (fig. 1A, fig. S2, fig. S6). We estimate the emergence of 1.ANT around the 15th century (1437 AD; 95% HPD 1357- 1545), and its early diversification in the late 16th century (1595 AD; 95% HPD 1476-1712), giving rise to 1.ANT3 (fig. 2A). Interestingly, the finding of an isolate from Yemen within the most ancestral positions of the 1.ANT3 sublineage supports the idea of the introduction of 1.ANT into Africa from Central Asia via the Middle East, following ancient trade routes (fig. 2C) (40, 42, 43). 1.ANT2-1.ANT1 diverged as early as the 17th century (1726 AD, 95% CI 1615 -1819) and represent the further spread of plague from Eastern (Tanzania, Kenya) into Central Africa (Uganda, DRC). The 1.ANT1 sublineage may have evolved early in the 19th century (1815 AD; 95% HPD 1725- 1888). We found two distinct clades of 1.ANT1 circulating during the human plague outbreaks in the Ituri region (DRC) between 2004-2010 (fig. S6).

The 2.MED lineage emerged in the 15th century (median 1404.18, 95% HPD 1249.87 - 1560.15) and further spread through Asia (fig. S6). The estimation dates indicate that the most recently derived sublineage, 2.MED1, evolved in the 18th century (1722 AD; 95% HPD 1631 -1802) establishing in vast areas in the Caucasus, the Middle East and Central Asia (24). The 2.MED1 clade that currently circulates in Iran evolved at the end of the 19th century (1871 AD; 95% HPD 1827- 1905), spreading from the Kurdistan area into the neighboring Caucasus (Russia, Azerbaijan), Turkey and Northern Africa (Libya)(fig. 2AD). The estimated transmission of plague into Libya from or nearby Western Iran early in the 20th century (1938 AD; 95% HPD 1914-1948) closely follows the last decades of rule on present-day Libya by the Ottoman Empire (1864-1912), which extended from South of the Caucasus up to the Maghreb (44). It is tempting to speculate that human movements and trade within such areas under the same rule were responsible for the establishment of 2.MED1 into Libya, where it has caused human plague outbreaks until the last decade (33). We notice the frequent presence of human isolates interspersed between *Y. pestis* isolates recovered from animal reservoirs and their ectoparasite vectors both in 1.ANT and 2.MED1 phylogenies (fig. 2AB), indicating that zoonotic spillovers are common in these lineages despite not being associated to historical pandemics.

Different signatures of expansion within *Y. pestis* lineages

The different patterns of historical spread of plague have probably influenced the evolution of the diverse *Y. pestis* lineages. The presence of multiple polytomies (star-like radiations) across the phylogeny has been stressed in previous studies as being a recurrent feature in *Y. pestis* evolution (21). Such polytomies include the node known as the “Big Bang” that marks the split of branch 0 into branches 2, 3, 4 and 1, but polytomic nodes are also present in the 1.IN and the Third Pandemic 1.ORI lineages (fig. 1, fig. S2, fig. S5). These radial topologies have been interpreted as hallmarks of rapid effective viral or bacterial population expansion events in a context of outbreak spread (45). To further investigate possible phylogenetic signatures of expansion in *Y. pestis* evolution, we compared the distributions of terminal branch lengths for major groups of modern *Y. pestis* lineages as a measure of their expansion, expecting that shorter terminal branch length distributions may indicate higher expansion rates (46, 47). We found branch 1 lineages (1.ORI, 1.IN, 1.ANT) to have the shortest median terminal branch lengths compared to the rest of lineages, except 4.ANT (pairwise two-sided Wilcoxon rank-sum tests significant p-value

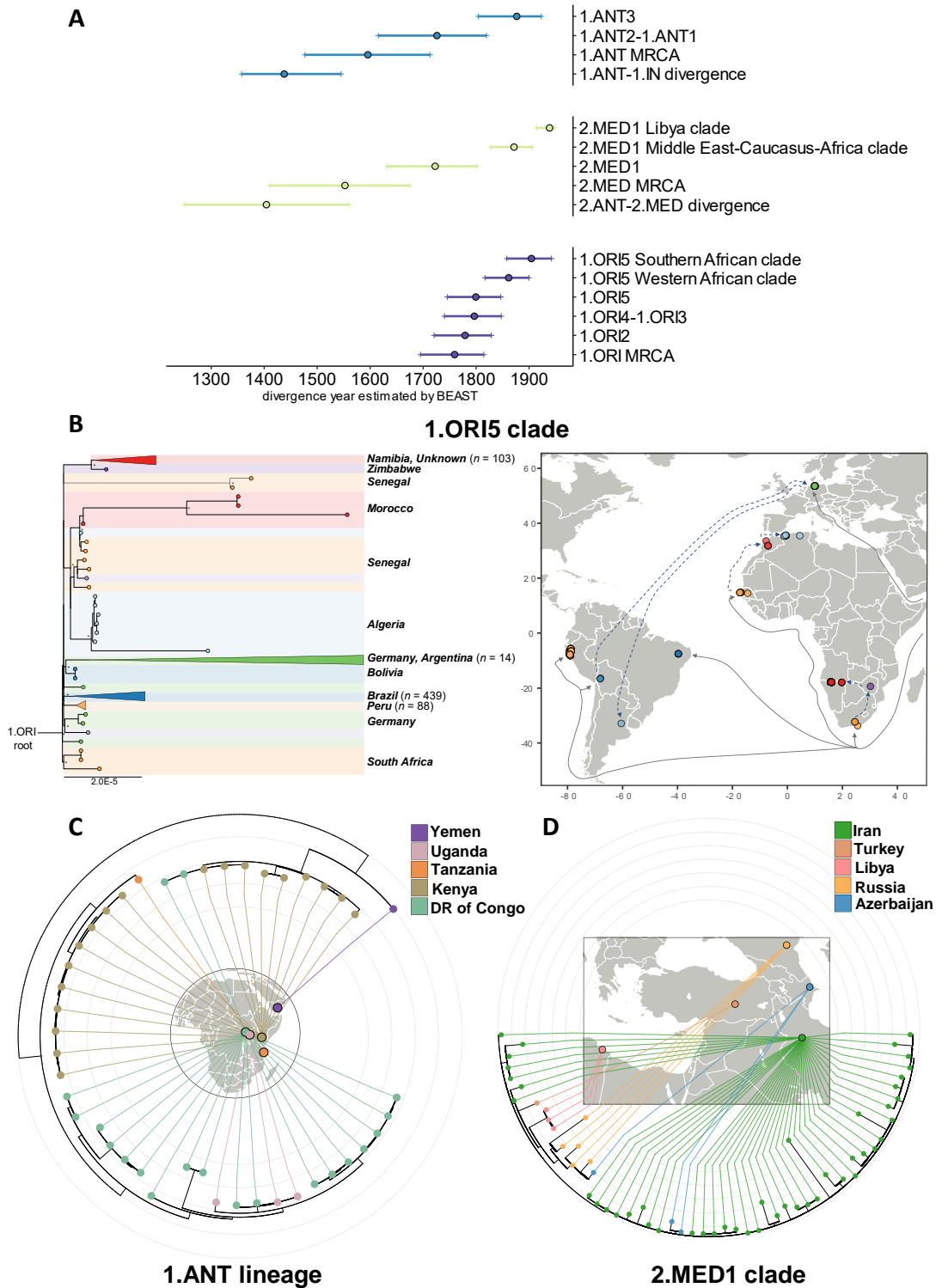


Figure 2. Inference of transmission events of different *Y. pestis* lineages. **A:** Dates of relevant nodes of the time-scaled phylogeny inferred using BEAST. The plotted date intervals correspond to the 95% highest posterior density (HPD) intervals, while the points indicate the median value of the inferred ages for selected nodes within the 1.ANT, 2.MED and 1.ORI lineages, respectively. The complete dated tree is shown in fig. S7. **B:** Focus on the phylogeography of the 1.ORI5 identified in this study (see Methods). Subtree of the 1.ORI5 clade (left), with tips colored with the colors displayed in the right map. The nodes marked with asterisks have bootstrap supports >85%. The closer map (right) summarizes the inferred spreading routes of the 1.ORI5, where transmissions indicated by continuous lines represent introductions inferred from clades that directly emerge from the 1.ORI5 common ancestor, while discontinuous lines represent transmissions that are inferred from monophyletic clades within 1.ORI5. **C:** Focus on the Central-East African 1.ANT lineage phylogeography. Each genome in the circular subtree is connected to the geographic location at the country level. The 1.ANT lineage is divided into three sublineages: 1.ANT3 (upper clade), 1.ANT2 (left clade) and 1.ANT1 (down clade), which describe an eastwards expansion of the lineage within this region in Africa. **D:** Focus on the clade within the 2.MED1 sublineage that includes novel genomes from the Kurdistan in Iran (*n*=54, in green), that are the most ancestral within the clade. Transmissions to neighboring Turkey, Azerbaijan, Russia and Libya are inferred from the circular subtree, represented as in fig. S6. The three isolates from Libya were collected in the context of reemergence of plague in Libya in 2009.

< 0.05, fig. S8). However, we note that lineages 4.ANT (n=15) and 3.ANT (n=13) are barely sampled, limiting their comparison. As 1.ORI isolates are overrepresented in the phylogeny (n=2023), we also estimated the medians of the terminal branch lengths on a random subset (n=200) of 1.ORI isolates using nonparametric bootstrapping (see Methods), together with all the rest of lineages. Bootstrapping revealed a similar hierarchy, with 1.ORI, 1.IN and 1.ANT presenting the shortest median terminal branch lengths, followed by 4.ANT, 0.ANT, 3.ANT, 2.MED, 2.ANT and 0.PE (fig. S9). The finding that branch 1 lineages have the shortest terminal branch length suggests that these lineages might be under expansion. This signature of increased dissemination is supported by the global spread of 1.ORI during the Third Pandemic in the last century, as well as the circulation of 1.IN and 1.ANT causing human plague outbreaks in the recent decades (48–51). In contrast, 0.PE lineages present nearly ten-fold longer median terminal branch lengths compared to 1.ORI, suggesting their maintenance in more restrained cycles of transmission (fig. S9). 0.PE (also called “Pestoides”) lineages rarely cause outbreaks in humans and their circulation is considered to remain mostly confined to the animal reservoir (52, 53).

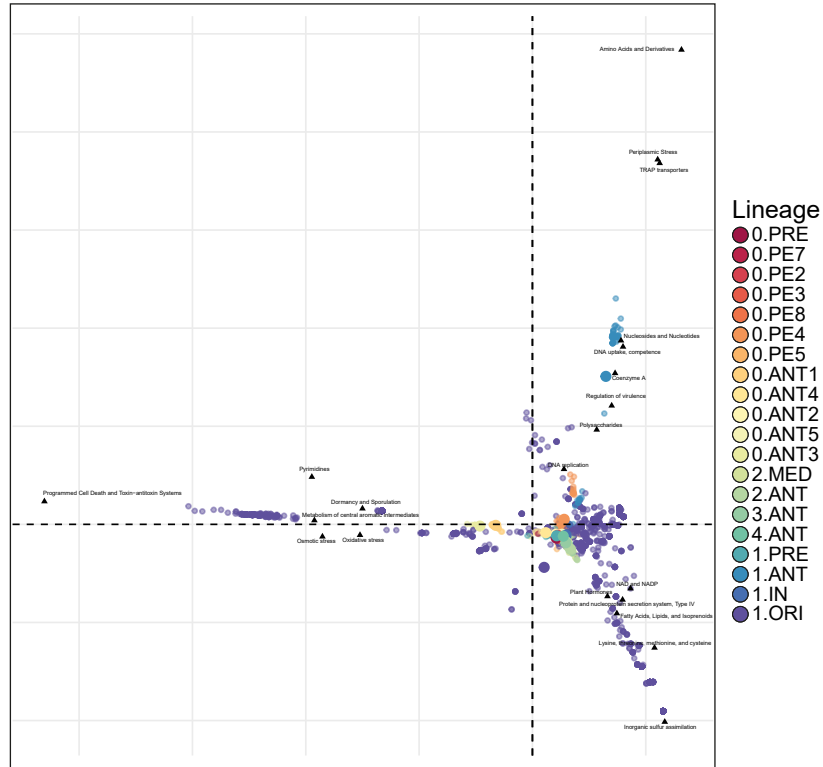
Functional diversification in *Y. pestis* lineages

The particular evolutionary paths and spreading histories of different *Y. pestis* lineages may have resulted in point mutations contributing to their virulence and adaptation. To better understand the relationship between lineage diversification and changes in gene function in *Y. pestis*, we analyzed the protein-altering (PA) mutations (i.e.: nonsynonymous or nonsense mutation events) identified within the 2806 genomes included in the global phylogeny. Different depth thresholds were applied for ancient and modern *Y. pestis* genomes, and both singletons and shared variants (shared by 90% of genomes) were removed, allowing to identify 1615 PA mutations. To characterize functional variants that may be enriched in particular *Y. pestis* lineages, genes were annotated and grouped into different 95 functional categories (see Materials and Methods). To allow comparisons between functional categories, counts of mutation events were normalized based on the sum of sequence lengths of genes present in each category (fig. S10). We then used correspondence analysis to investigate possible relationships between *Y. pestis* lineages and mutation enrichment for some functional categories. For instance, we found that isolates from the 1.ANT group specifically acquired PA mutations affecting genes related to stress sensing (*barA*, *rseP*, *skp*, *surA*) and membrane traffic (YPO1579). Missense variants in genes related to response to different types of stress (*proV*, *cdr*) also appear to be enriched in some 1.ORI clades (fig. 3A). As both 1.ANT and 1.ORI lineages have spread to different locations since their emergence, such variants may have contributed to their adaptation to the conditions encountered at the areas where they have established. On the other hand, genes that exhibit an excess of PA variants may be indicative of diversifying selection. To uncover genes showing high levels of variation, we estimated the ratio of PA mutation events per gene, normalized to the count of synonymous mutations per gene, as a measure of diversifying selection. We identified thirteen genes presenting a high PA ratio (>3) and three genes showing clustered mutations (>10 sites altered) (fig. 3B). Most of these variants are scattered along the phylogeny or shared by some lineages (fig. S11). Interestingly, within these loci we find genes that have been shown to be important for virulence or adaptation in pathogenic *Yersinia* spp. (*yapF*, *ail*, *yopT*) or other bacteria (*aspA*, *ftsK*, *cysJ*, *cdiA*) (table S2). Two of these genes (*aspA*, *ftsK*) had been identified in a previous work as genes presenting hallmarks of diversifying selection in *Y. pestis* (21). We note that two of the variants present in *YopT* effector locus (YPCD1.20, 437G>A, 103C>T) cause the formation of premature stop codons that probably abrogate its function. These nonsense variants are exclusively present in isolates of the 1.ANT3 and 1.ANT2 sublineages, while 1.ANT1 maintained the functional genotype.

Genomic content variation shaped the evolution of *Y. pestis*

Besides coding sequence diversification, variation in gene content may have also contributed to the emergence of different *Y. pestis* lineages. The acquisition of two plasmids and the loss of several gene functions are considered as cornerstone evolutionary events that led to the speciation of *Y. pestis* from its ancestor *Y. pseudotuberculosis* (18, 54). However, little is known about the genomic differences that characterize the different lineages as they emerged in the evolution of *Y. pestis*. A previous study identified the specific loss of 18 loci in branch 2 lineages (2.ANT and 2.MED), suggesting that genomic variation has also occurred during *Y. pestis* microevolution

A Correspondence analysis



B Highly variable genes

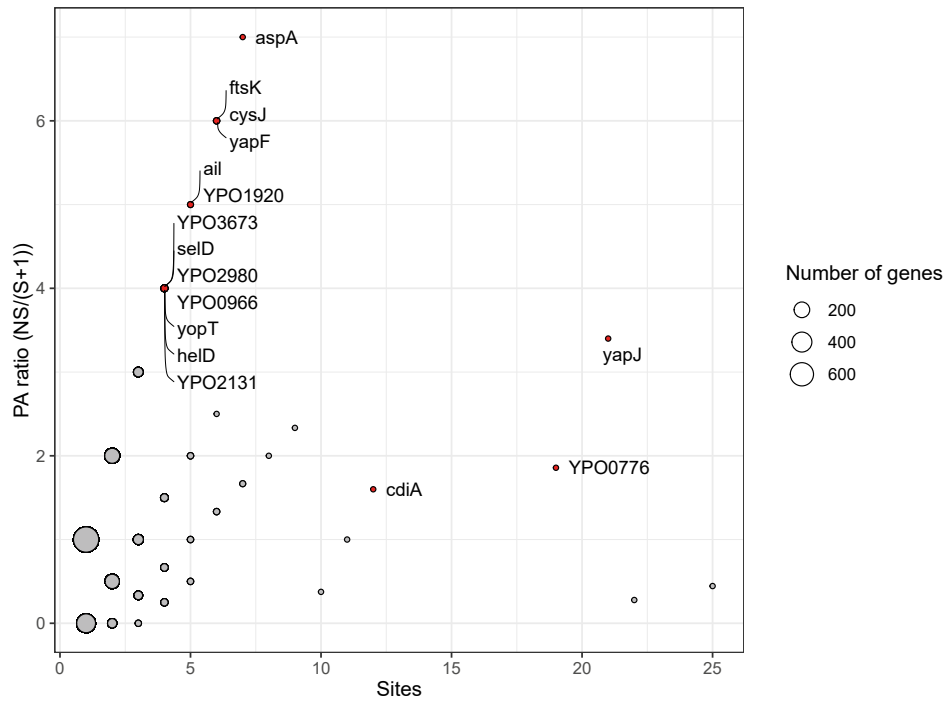


Figure 3. Patterns of protein-altering mutations in *Y. pestis* lineages. **A:** Biplot of the correspondence analysis describing the enrichment in protein-altering (PA) mutations identified within ancient and modern *Y. pestis* lineages. The total identified PA mutations ($n=1615$) are conserved in >5 genomes and were classified into 95 functional categories following the annotations of the SEED server (subcategory level, fig. S10). Total PA counts within each subcategory were normalized on the total sequence lengths of genes within the subcategory (see Methods). The correspondence analysis describes associations between groups of *Y. pestis* genomes and the functional subcategories enriched in PA mutations for these genomes. **B:** Identification of highly variable genes. The plot shows groups of genes according to their ratio of PA mutations per gene normalized to the number of synonymous mutations per gene (y axis), and the number of nucleotide sites affected by PA mutations in each gene (x axis). Genes presenting a high PA ratio (>3) or mutations in multiple sites (>10) were marked as presenting possible signals of diversifying selection (S11).

(55). We extend these observations by comparing the genomes of 1939 modern isolates, including most of the novel genomes sequenced in this study. We reconstructed the *Y. pestis* pan-genome from *de novo* assemblies and identified a total of 23523 orthologous gene families (with a minimum 80% identity cut-off), 3717 of which were present in >90% of isolates, that we defined as the core-genome. We identified 10583 gene families that were variably present in some isolates, defining the accessory genome. To further understand the changes in coding capacities between different lineages, we focused on the distribution of accessory genes that were likely not part of the three common plasmids. For this purpose, we filtered only gene families that were present within contigs that did not contain sequences of any of the three known plasmids (see Materials and Methods). Our analysis reveals a mosaic pattern of gene repertoires distributed across the phylogeny (fig. S12), indicating that the emergence of different *Y. pestis* lineages was characterized by events of gene gain and gene loss besides their diversification through point mutations in core-genes. We could identify clusters of 4-106 genes that are lineage-specific, while some other clusters are shared by two or more lineages (fig. 4). The presence of distinctive gene repertoires in *Y. pestis* lineages raises questions regarding their impact on the adaptation to different environments and host niches in which the different lineages circulate. Further analyses are needed to better understand the interplay of phylogeny, geography, and host niche in shaping *Y. pestis* pangenome.

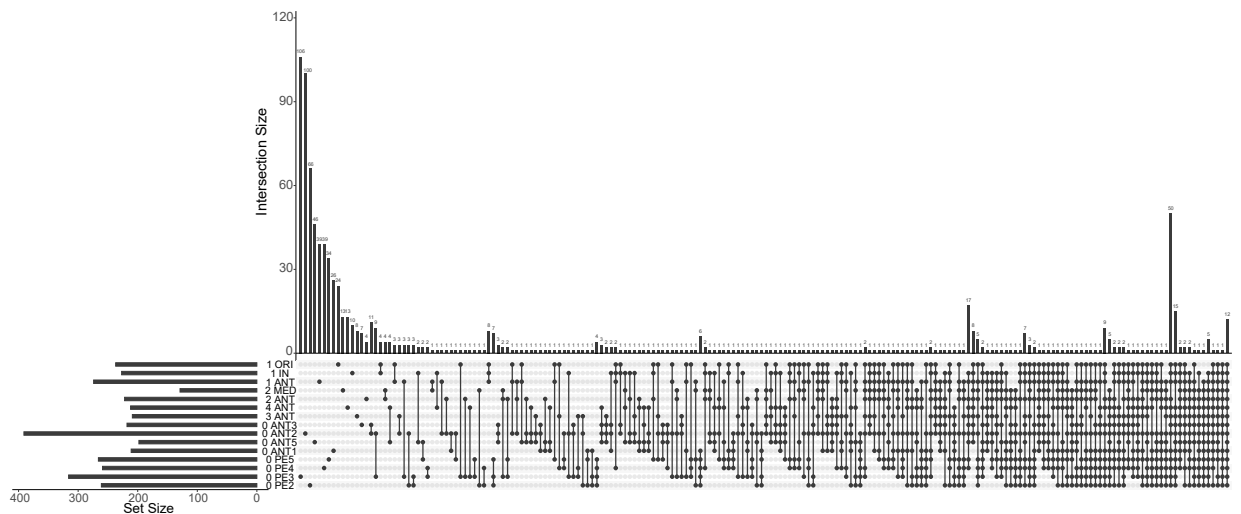


Figure 4. *Y. pestis* lineages are defined by differential gene repertoires. Gene families from the pangenome analysis of 1939 modern genome assemblies which were shared by at least 50% of genomes were included in the plot. These correspond to chromosomal non-core and non-singleton gene families. The combination matrix (down) displays combinations of intersection of modern *Y. pestis* lineages (vertical lines) according to the shared gene repertoires within groups of genomes. Single black points indicate gene families that are specific to the lineage indicated at left. The intersection size represents the total number of gene families shared by the lineages indicated in the matrix, while the set size panel (left) describes the total number of gene families present in each lineage.

Insertion sequence expansion during *Y. pestis* microevolution

Genome reduction or gene inactivation is another major process that guided the adaptative emergence of *Y. pestis* as a vector-borne pathogen, evolving from an enteropathogenic ancestor (18). A hallmark of this adaptation has been the accumulation of insertion sequences (IS), which are thought to be major driving forces of *Y. pestis* genomic plasticity, as they can interrupt genes and cause genomic rearrangements. IS are highly prevalent in *Y. pestis*, and about a third of the total pseudogenes in *Y. pestis* reference genome are consequence of IS insertions (54). Previous population genomics studies in *Y. pestis* mostly focused on point mutations and indels, ignoring the possible contribution of IS in genomic variation between *Y. pestis* isolates. We identified six types of IS that are present in most available *Y. pestis* closed genomes (n=62): IS200 (also called IS1541), IS100, IS285, IS1661, ISYpe1 and ISYps7. As most of these IS are also present in *Y. pseudotuberculosis*, we compared the number of IS copies in closed genomes from both species (table S3). As previously reported (18, 54), *Y. pestis* chromosomes present more IS than *Y. pseudotuberculosis* (median 122 vs 20), which account for 2.6-4.5% of bases in *Y. pestis* compared to <1% in *Y. pseudotuberculosis* (fig. S13). The most prevalent IS types in *Y. pestis* closed genomes are IS200, IS100 and IS285, in line with previous findings (54). To better understand the evolutionary trajectories

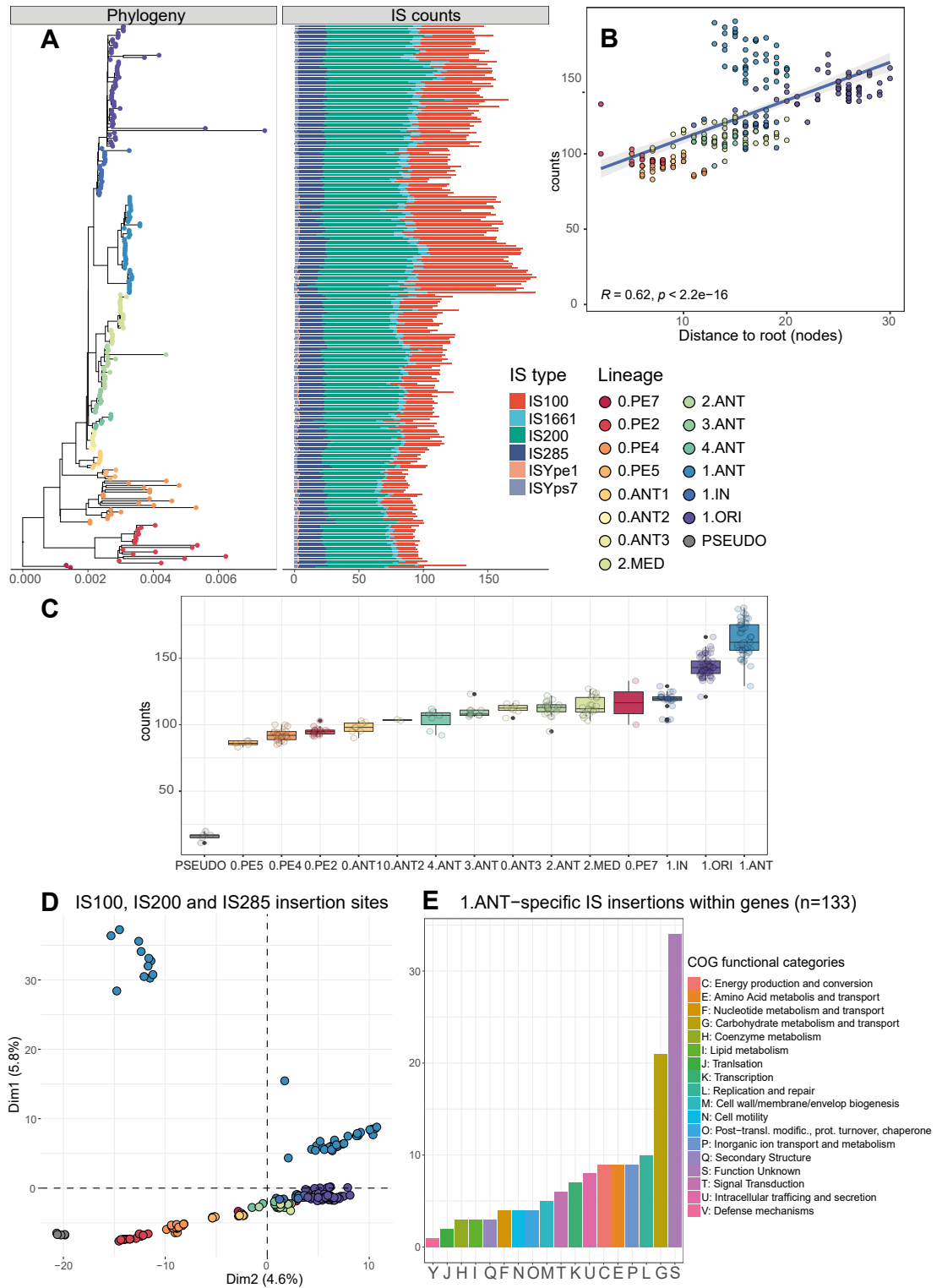


Figure 5. Expansion of insertion sequences (IS) during *Y. pestis* evolution and impact on gene loss within 1.ANT lineage genomes. **A:** IS dynamics according to *Y. pestis* population structure. Total counts of the 6 identified IS types per genome are shown next to a ML tree describing their phylogenetic relationships. The tree was rooted on *Y. pseudotuberculosis* and is ordered from most ancestral (down) to more recently evolved lineages (up). Chromosomal IS insertions were identified using ISMapper (56) on 245 *Y. pestis* and 5 *Y. pseudotuberculosis* (not shown) by mapping short-reads data onto an IS-free reference genome of *Y. pestis* CO92 (see Methods). **B:** Linear correlation between the total IS counts and the patristic distance in number of nodes to root per each genome shown in **A**. Points are colored according to lineages correspondences of the tree. The number of nodes to the root is considered a marker of lineage diversification in *Y. pestis* (21). **C:** Distribution of total IS counts between the different lineages present in the IS analysis dataset, compared to the counts in 5 *Y. pseudotuberculosis* genomes as comparison. **D:** PCA of the matrix of IS insertion sites positions on the chromosome, with points coloured by lineage as indicated by the tips of the tree in **A**. **E:** Functional annotation of 133 loci impacted by IS insertions which are specific to at least one 1.ANT lineage genome. Proteins were annotated based on clusters of orthologous genes (COG) categories, and the sum of proteins within each category is shown.

of IS in *Y. pestis* lineages and its potential contribution to functional gene loss, we explored the IS variation in a dataset of modern *Y. pestis* genomes (n=245) representative of the genetic and geographic diversity of the pathogen, together with five *Y. pseudotuberculosis* genomes as comparison. We identified chromosomal insertion sites for the six common IS types in short-read data (see Materials and methods) and detected a similar number (median 119) of insertions counts found in closed genomes. The analysis also confirmed the higher prevalence of IS200, IS100, and IS285 insertions within diverse *Y. pestis* isolates (fig. S14, table S3). To infer the dynamics of IS in relation to the population structure, we analyzed the distribution of IS counts between lineages across the *Y. pestis* phylogeny (fig. 5A). Interestingly, we found a strong positive correlation between the total IS counts and the number of nodes from each genome to the most recent common ancestor (MRCA) of the genealogy ($R^2 = 0.62$, fig. 5B), indicating that IS have been also expanding since the divergence of most ancestral *Y. pestis* lineages. Since the number of branching events across the *Y. pestis* phylogeny are considered indicative of lineage emergence and expansion (21), the accumulation of IS insertions may be a secondary marker of lineage diversification. Accordingly, the distribution of IS insertions within *Y. pestis* chromosome clustered the genomes mostly following the lineages that are defined by their specific point mutation variation (fig. 5CD). We found 1.ANT to have the highest number of IS insertions compared to the rest of lineages, which appears to be mainly caused by the expansion of IS100 in 1.ANT isolates (fig. 5C). To investigate the possible impact of this signature of IS expansion on functional gene loss in 1.ANT lineage, we focused on IS insertions that were specific to 1.ANT isolates. We detected 133 specific gene interruptions caused by IS insertion that are present in at least one isolate. Most of the genes interrupted by IS insertions are related to membrane transport and carbohydrate metabolism (fig. 5E), while the interruption of several genes coding for Tc toxins (*tcaC*, *YPO3681*) and structural genes of type VI secretion systems (*vgrG*, *tssH*) is common in most isolates of the 1.ANT3 and 1.ANT2 sublineages. The impact of the observed gene-truncating IS insertions on the 1.ANT virulence and ecology remains to be determined, together with the specific gene loss in other lineages that may be attributed to IS.

Discussion

In the last decade, genomic analyses have provided the first insights into the evolution and diversity of *Y. pestis* species. These analyses have retraced the clonal distribution of *Y. pestis* lineages, both from the sequencing of modern bacterial isolates (3, 21, 23, 24, 57) and ancient DNA metagenomes from victims of past *Y. pestis* infections (2, 25–28). The phylogeny of *Y. pestis* species is defined by the alternance of *Y. pestis* lineages that have caused historical plague pandemics and outbreaks in the past, interspersed between the emergence of groups that have been evolving for long, which are responsible for the persistence of plague over vast locations in Asia. In this study, we used phylogenomic analyses to uncover missing pieces in the evolutionary history and global genetic diversity of *Y. pestis*. Through the sequencing of a large collection of isolates from the past and current century, we inferred new patterns of dissemination of *Y. pestis* during the Third Pandemic, notably for the introductions of the newly described 1.ORI5 sublineage into Africa. During the period in which the ongoing Third Pandemic has developed, *Y. pestis* from lineages other than 1.ORI have also circulated during outbreaks and several collected isolates of these lineages are part of the studied dataset. We have expanded the available genetic diversity of the East African 1.ANT lineage, and we retraced its evolution and distribution up to the outbreaks in the last decade caused by this *Y. pestis* population (8). The origin of 1.ANT and its establishment into the African continent remain subjects of debate (42, 58), as it is the first lineage emerging after the ancient DNA genomes recovered from victims of the Second Pandemic in Europe, suggesting a possible link between both events (fig. 1). Our results support the hypothesis of the introduction of 1.ANT from Asia into Africa from the East, possibly through the Red Sea (42). The 2.MED lineage distribution is particularly large, appearing widespread across longitudes in Asia. We further retrace the spread of the 2.MED1 sublineage through the Caucasus area and westwards into Africa, suggesting its historical spread was human-mediated.

The phylogeny of *Y. pestis* thus appears to mirror the spreading *Y.* events of plague that are responsible for the geographical distribution of the disease. Such historic of spread have probably impacted the evolution of *Y. pestis*, and hallmarks of expansion events may become apparent in the phylogenies, such as the recurrent presence of polytomies in the tree (21). We provide new insights into this question by analyzing the distribution of terminal

branch lengths among isolates, which has been used to inform about lineages that show rapid dissemination in other pandemic pathogens such as *Mycobacterium tuberculosis* (46, 47). We showed that isolates from lineages of branch 1 (1.ORI, 1.ANT and 1.IN) presented shorter terminal branch lengths, suggesting that these lineages had undergone recent expansion. If these results are expected for 1.ORI due to its spread and establishment into new areas during the Third Pandemic, it is more interesting for 1.ANT and 1.IN, which appear to be more geographically restricted (fig. 1, fig. S4). Terminal branch lengths may be used as indicators of a higher recent transmission history for these lineages, compared to other lineages that rather appear to remain restricted to transmission cycles within rodent communities, such as 0.PE lineages (53). These findings may illuminate groups of *Y. pestis* that have circulated more actively, hence representing an increased risk of spillover to human communities living nearby. The distributions of terminal branch lengths might result from the interaction of several factors including transmission rates, mutation rates and dynamics of latency (59). Significant variations in mutation rates have been observed to occur along the *Y. pestis* phylogeny, which have also been related to different branching patterns (i.e.: polytomies)(21, 60). In turn, it was hypothesized that this heterogeneity in clock rates was caused by episodic events of increased epidemic transmission occurring throughout the evolution of *Y. pestis* (21). The comparative analysis of terminal branch lengths and mutation rates may therefore help to clarify the relationships between these factors in *Y. pestis* evolution.

In the second part of our study, we go beyond historical reconstructions and explore the functional diversity of *Y. pestis* lineages from different angles, starting by protein-altering (PA) variants. *Y. pestis* is considered a genetically monomorphic pathogen (61), as it accumulated reduced levels of genetic diversity since the divergence of the most ancestral lineages. Previous works have analyzed the ratio of PA (nonsynonymous) to synonymous mutation (dN/dS) by estimating its value across the nodes on the *Y. pestis* phylogeny. The distribution of dN/dS closely remained close to parity, ruling out possible hypotheses of strong diversifying selective pressures guiding *Y. pestis* diversification and in favor of its global evolution tending to neutrality and genetic drift (21). However, we argue that the sole exploration of ratios values does not inform about the distribution and functional impact of PA mutations occurring in specific *Y. pestis* populations, which may hide possible signatures of adaptation or diversifying selection. For instance, it has been recently shown that PA variants affecting the *rpoZ* gene influence biofilm formation in *Y. pestis*, and the occurrence of these variants has been placed into the context of climatic variations through decades in the region where isolates carrying these variants circulate (62). By stepping on the considerable increase in genomic data generated in this and other recent studies, we have explored the functional diversification in *Y. pestis* genomes and identified enrichment trends in genetic groups from the 1.ANT and 1.ORI lineages. The presence of specific PA mutations in genes coding for proteins related to stress sensing and response in these lineages may point to factors that contributed to the adaptation of *Y. pestis* to different environmental conditions. Moreover, the finding that multiple loci associated to virulence mechanisms are part of the genes presenting highest rates of nonsynonymous variants is remarkable and deserve further analyses to characterize their possible impact in gene function using complementary approaches. On the other hand, our previous understanding of *Y. pestis* diversity in terms of gene content remains limited, as precedent analyses of the *Y. pestis* pangenome were based in a reduced set of genomes (21, 55). By comparing the genomes of thousands of *Y. pestis* isolates, we provide evidence that *Y. pestis* lineages are characterized by specific gene repertoires, suggesting that variation in gene content may have played a role in the emergence of different *Y. pestis* lineages.

We finally analyzed the prevalence and distribution of IS within *Y. pestis* lineages to explore possible patterns of functional gene loss during *Y. pestis* microevolution. Genome reduction through functional gene loss is one of the hallmarks of the emergence of *Y. pestis* from its ancestor species *Y. pseudotuberculosis*, which is thought to have contributed to its adaptation to a different ecologic niche and transmission route (16). For instance, the inactivation of *Y. pestis* Invasin (*inv*) by an IS200 insertion is considered to have played a role in this transition from an enteropathogenic to a vector-borne lifestyle (16, 63). IS are considered key determinants of the adaptation and genome reduction of other relevant pathogens, including *Bordetella pertussis*, *Mycobacterium leprae* and *Shigella* spp. (64–67). A recent study found that the parallel expansion of IS in different *Shigella* spp. led to convergent patterns of gene inactivation and loss of specific metabolic pathways (67). We demonstrate that IS have been expanding during *Y. pestis* species microevolution, notably due an expansion in IS100 copies. As IS100 are also present in all three plasmids common in most *Y. pestis* isolates (54), the expansion of this IS type in *Y.*

pestis chromosomes may be plasmid-driven, as it has been suggested for the described expansion of IS in *Shigella* species (67). Intriguingly, we found a specific signature of IS expansion in isolates of the 1.ANT lineage, leading to the interruption of tenths of loci in genomes of this lineage. The impact of IS expansion in the functional gene loss in different *Y. pestis* lineages and across the evolutionary path specific to *Y. pestis* remains to be carefully investigated. Moreover, the finding that *Y. pestis* genomes cluster in similar groups according to the positions of IS insertion in the chromosome and according to phylogenetic lineages is particularly interesting. These patterns of insertion, together with the observed increase IS copies throughout the phylogeny, indicate that IS expansion may represent a signature of diversification in *Y. pestis* lineages. Our results complement recent studies that have shown that the patterns of small indels in *Y. pestis* lineages also mirror their phylogenetic relationships defined by point mutations (68). These findings altogether point towards a more complete understanding of *Y. pestis* evolution, in which multiple mechanisms (point mutations, gene gain, and gene inactivation through IS insertions and indels) act as driving forces in the emergence of *Y. pestis* lineages and their functional diversification. Future analyses should focus on elucidating possible selective signals in *Y. pestis* genomes by investigating in parallel these three mechanisms that appear now as possible drivers of genetic diversity in *Y. pestis* species. The different patterns of diversity caused by these mechanisms need to be placed into the specific ecologic and epidemic backgrounds and the evolutionary history of *Y. pestis* lineages. Their analysis, together with functional assays, may help revealing factors that are important for the adaptation of *Y. pestis* during its spread and successful establishment into new areas. The sequencing and analysis of comprehensive genomic datasets is essential to identify new evolutionary patterns and provide enough statistical power to contrast different hypotheses. We generated the largest genomic dataset for *Y. pestis* to date, which has doubled the available number of genomes for this pathogen. Recent studies have mostly been focused on the diversity of isolates present in Asia and few other endemic countries or analyzed ancient *Y. pestis* genomes that have been recovered mainly in Europe, while other regions in the world remained under-represented. The genomic data from diverse regions in the African continent provided in our work may help filling this missing gap in our understanding of *Y. pestis* diversity in Africa. The growing availability of more genetically and geographically diverse *Y. pestis* datasets will also be of help for the typing and genomic epidemiology of plague. In the present study, we develop an integrative approach to interrogate different mechanisms in *Y. pestis* evolution. Taking further steps into these aspects may facilitate our understanding of the mechanisms that have allowed one the most successful pandemic pathogens to persist and reemerge on several occasions throughout the last centuries.

Acknowledgments

We thank François-Xavier Weill for critical comments during the development of this project. We also thank Sylvie Brémont for the assistance in the recovery of pure cultures of old *Y. pestis* collections. This work stems from the collaborative heritage of the historical collection of isolates that have been recovered and sent from diverse sources over the past century by many different actors, to which we are grateful. This work was supported by the French government Agence Nationale de la Recherche - ASTRID program (project ANR-18-ASTR-0004). GMF received a PhD fellowship by the Institut Pasteur the Direction Générale de l'Armement, as well as from the Université Paris Cité, through the EUR G.E.N.E. Doctoral School (project ANR-17-EURE-0013). This work was also supported by the Pasteur Network (Pasteur International Unit 'Plague: Evolution, Spread and Maintenance) and LabEx IBEID (grant ANR-10-LABX-62-IBEID).

Data availability

All supplementary tables and the XML file used for the inference of time-scaled trees in BEAST2 have been deposited in a repository: <https://github.com/guillemmasfiol/GlobalPlague>.

The full ML phylogeny with available metadata have been imported into a project in MicroReact: <https://microreact.org/project/jx9yf1sDRf3Svc3PBi7XLL-global-phylogeny-of-2806-genomes>.

Methods

Bacterial isolates and genome sequencing

A global collection of 1549 *Y. pestis* isolates was collected from different teams and Institut Pasteur around the world between 1899 and 2017, in the context of human outbreaks and of plague surveillance efforts, notably in Vietnam (n=575) and Madagascar (n=586), where exhaustive sampling was obtained within the 1960s and 1990s decade, respectively. The total isolates were obtained from 29 present-day countries (fig. 1, dataset S1), mostly from Africa (n=747, 48%), and Asia (n=686, 44%). Most of the isolates were recovered from human clinical cases of plague (n=1087, 70%), while the rest of new isolates included in this study were recovered in the context of animal plague surveillance, in which different species of 16 different animal hosts and flea vectors were sampled.

All manipulations of live *Y. pestis* isolates were performed in biosafety level 3 laboratory conditions at the Institut Pasteur. Whole-genome sequencing reads of the isolates were obtained using an Illumina MiSeq platform, generating pair-end sets of 2x300 bp lengths. Sequencing libraries were prepared following the protocol of the Nextera DNA Flex library preparation kit (Illumina).

Sequence analysis and de novo assembly

The sequencing reads of the 1549 *Y. pestis* isolates from the Pasteur collection were added to a larger genomic dataset that included most of the available *Y. pestis* samples available in public repositories (n= 1363, as of July 2021). Raw sequencing reads were used when possible, while for 404 samples only genomic assemblies were available. This initial dataset included ancient DNA metagenomes from different centuries that have been analyzed in recent studies (n=136).

Genomic assemblies present in public repositories (NCBI, Enterobase (29) for the samples included in the dataset were obtained and kept for downstream analyses. For all modern *Y. pestis* genomes for which only raw sequencing reads were available (n=2076), including the 1549 newly sequenced isolates, de novo assemblies were generated using the fd2dna workflow (<https://gitlab.pasteur.fr/GIPhy/fq2dna>), which develops the steps that are described next. Correction of reads, clipping and trimming from adapter sequences and nonconfident calls was performed using AlienTrimmer v.0.4.0 (69), and redundant or over-represented reads were identified and reduced using the khmer tool package v1.3 (70). A minimum threshold of 12x was used in the pre-processing steps, and sequencing errors were identified and corrected by Musket v1.1 (71). De novo assemblies of pre-processed reads into contigs and scaffold sequences were finally generated using SPAdes v3.1.2.0 (72).

Phylogenetic analyses

To process the large dataset of samples in this study, we implemented a workflow based on Nextflow (73) workflow manager, that is available on a GitLab repository: <https://gitlab.pasteur.fr/metapaleo/amphy>. This workflow allowed to integrate raw sequencing data from modern *Y. pestis* genomes and ancient DNA metagenomes, together with samples for which only contigs were available. For contigs, reads were simulated using wgsim function in samtools v1.13 (74). Reads for all samples were trimmed for adapter sequences using Adapter-Removal v2.3.1. (75) and only bases with a quality of 2 or less were removed and reads larger than 30 bp were retained. Trimmed reads for paired-end data were collapsed and merged and mapped against *Y. pestis* CO92 reference genome (NC_003143.1, NC_003131.1, NC_003134.1, NC_003132.1) using BWA mem v0.7.17 (76). Aligned reads were processed and using samtools v1.13 (74), discarding reads with an overall mapping quality lower than 30. Then, duplicates were detected and removed using MarkDuplicates from Picard-Tools v2.23.8 <https://broadinstitute.github.io/picard/> and merged for each sample before realignment using GATK v4.1.9 (77). For ancient DNA samples, further steps were necessary, and pre-processed reads were therefore also rescaled using mapDamage v2.2.1 (78) for base damage (i.e.: masking the C > T and G > A mutations predicted resulting from ancient DNA degradation). Genotypes were called from clean aligned reads on *Y. pestis* CO92 genome using GATK HaplotypeCaller. Sites were filtered for a minimum 4x depth for ancient DNA samples or 10x for modern ones, and maximum 1000x depth. A minimum base quality score of 20 was also used, and calls were filtered for allelic balance of 0.1. Indels were insertions were removed, while deletions were replaced by "N", and resulting

VCFs were transformed into consensus fasta sequences. Coding sequences were extracted from the consensus using gene annotations from the reference genome and only genes with less than 10% of missing calls (“N”) for each sample were kept in a first step, while only genes kept in at least 50% of samples were retained for further steps. Gene consensus sequences passing the filters mentioned above for each file were then aligned and used for downstream phylogenetic analyses. A total of 2806 *Y. pestis* genomes (dataset S2) passing the previously mentioned filters were retained, including the 1549 novel genomes and 1257 modern and ancient genomes present in public repositories. To analyze the population structure of *Y. pestis* species including all the retained samples, a phylogeny was reconstructed. Variants present within the sequences were partitioned in 3 positions based on codon positions and a maximum-likelihood phylogeny was computed using RAxML-NG v1.0.2 (79) using a GTR+Gamma substitution model. Finally, 100 Bootstrap distance trees were generated using fastME v2.1.6 (80). The analysis of terminal branch length (TBL) for the 2806 taxons in the maximum-likelihood phylogeny was developed by extracting the estimated branch lengths from the tree using gotree v1.18.3 (81), and plotting their distributions using the R-package ggplots2 (82). Non-parametric bootstrapping (NPR) was performed in R v4.1.2 on the estimated TBLs by taking the full taxon sets in all groups of *Y. pestis* lineages except for 1.ORI. For this lineage, the values of TBL for 200 random samples were taken and then NPR was performed using 1000 bootstrap replicates for TBL of each lineage group as shown in fig. S9. The inferred phylogeny and available metadata for the 2806 *Y. pestis* genomes have been integrated into a project in MicroReact, which is accessible via the following link: <https://microreact.org/project/jx9yf1sDRf3Svc3PBi7XLL-global-phylogeny-of-2806-genomes>.

To gain resolution in the phylogenetic relationships of lineages of interest (shown in fig. 2), a parallel inference was developed by restricting the analysis to modern *Y. pestis* genomes and including single nucleotide variants (SNV) that are intergenic, together with variants falling within coding sequences. For this analysis, that we called the “focus” analysis, 1010 representative genomes of all lineages with contemporary genomes and including all available countries were used. Samples kept in this analysis are detailed in dataset S2. In short, trimmed, pre-processed reads generated in the previous analysis were used for mapping on *Y. pestis* CO92 chromosome (NC_003143.1) using the RedDog-nf pipeline <https://github.com/katholt/reddog-nf> with default parameters (a minimum 5x read depth to call variants, 50% minimum coverage and minimum 10x average depth for the reads mapping on the replicon). All the 1010 samples covered >90% of the reference chromosome with an average depth >10x. Variants falling within repetitive regions and variable sequences identified in previous works (3) were removed from the variant tables. Only variable positions that were conserved in at least 95% of genomes were kept. A maximum-likelihood phylogeny was generated with the recovered SNVs using IQ-TREE 2 v2.2 (83) using ModelFinder (84) and 1000 bootstrap replicates.

The time-scaled phylogenetic tree (fig. S7) was reconstructed in BEAST2 v2.6.0 (34) using the BEAGLE library v3.1.2 (85) for parallel acceleration in GPU nodes. We used a subset of 198 *Y. pestis* genomes representative of all known major lineages and sublineages, including new genomes obtained in this study to date their divergence, together with the sequence of *Y. pseudotuberculosis* IP32953 used as outgroup to root the tree. The aligned data used to build the maximum-likelihood tree mentioned above was partitioned based on codon positions and based on previous experiments, a constant population size coalescent and an unlinked GTR substitution model with four gamma rate categories were used (2, 27). A relaxed log normal clock was assumed for modelling the mutation rates within branches, as previous analyses have found strong evidence for the adequacy of loose clocks for modelling *Y. pestis* evolution (21, 60). For this analysis, 7 independent Markov Chain Monte Carlo (MCMC) were run with 200 million states each, sampling every 1000 states and first 10% states of each run were removed as burn-in. The estimated effective sample sizes (ESS) of all estimated parameters were all > 200. Trees of the combined runs were summarized using TreeAnnotator (86) to generate a maximum clade credibility (MCC) tree annotated based on median heights.

Functional analysis of variants

To characterize variation in gene function due to point mutations in different *Y. pestis* lineages and detect highly variable genes, we recovered variant calls from VCFs files generated in the previously described phylogenetic workflow (AMPHY). In order for a variant to be called, a minimum cut-off 10 reads supporting the call for modern genomes and 5 reads for ancient genomes was set up. We merged in the same table all selected calls for the

2806 *Y. pestis* genomes included in the phylogeny. We excluded variants (including singletons) that were not conserved in at least 5 genomes, or that they shared the same call in >2500 genomes, identifying a total of 3451 variants. The functional effect of each variant was determined using SnpEff v.4.3 (87), and variants falling within coding regions (n=2407) were recovered. To identify highly variable genes, we computed the ratio of protein-altering (nonsynonymous and nonsense) mutations per gene, normalized to the synonymous mutations per gene and plotted the distribution of the ratio using the ggplots package (82). To compare the gene functions impacted by variants in different *Y. pestis* lineages, we only kept protein-altering variants (nonsynonymous and nonsense, n=1615) for downstream analyses. The metabolic function categories of the genes where the filtered variants fall were determined using the SEED subcategories annotation available in the RAST server (88) for the *Y. pestis* CO92 genome (NC_003143.1). To allow comparisons of variant counts between the different functional subcategories, the total counts of variants in each subcategory was normalized to the sum of sequence lengths of genes present in the corresponding subcategory using the formula:

$$counts = \frac{v_i}{\sum_{i=1}^n l_i} 10^6$$

Where v_i represents the total number of variants present in each functional subcategory and l_i represents the length in bp of each gene present in the subcategory.

Heatmaps of recovered protein-altering variants and of total variant counts within different functional subcategories were generated using R package ggplots. Genomes were sorted in heatmaps using the maximum-likelihood phylogeny reconstructed above, imported using the DECIPHER (89) and Dendextend (90) R-packages. To investigate relationships between lineages defined by the phylogeny and variant counts present in different functional subcategories, correspondence analysis was developed on normalized variant counts using the FactoMineR and FactoExtra (91) R-packages. The output biplots of the correspondence analysis are defined by axes that cover the maximum inertia of the data, and allow graphical associations between groups of genomes and functional subcategories enriched in mutations for these genomes.

Pangenome analysis

In order to explore the overall patterns of gene presence and absence in *Y. pestis*, we computed the pangenome of 2666 *Yersinia pestis* genomes using PanACoTA v. 1.3.1 (92). PanACoTA is a modular tool to build pangenomes from assemblies and during the first step performs a quality assessment of the data. In the quality assessment, a total of 720 genomes were excluded due to low quality (511), being identical or having a distance higher than 0.6 to other genomes in the dataset (209). The rest of the genomes (1949) were then annotated, and the gene clustered (-i 0.8) into gene families. We additionally removed 11 genomes due to an unusually high number of genes (>5000 genes) and a genomes of *Y. pseudotuberculosis* that was included by mistake in the workflow, being 1937 the final number of genomes included in the pangenome reconstruction. The resulting pangenome contains a total of 23523 gene families, which contain both chromosomal and plasmid encoded genes. In order to identify if a gene family was chromosomal or plasmid encoded, we extracted all the genes present in contigs labelled as complete plasmid in our dataset, resulting in the plasmids found in 42 assemblies. The gene families containing any of those genes were labelled as plasmid since it is likely that those genes were at some point in their evolution plasmid encoded and they could have a different evolutionary history that those encoded in the chromosome. By this procedure we detected 1172 plasmid-encoded gene families.

The output of PanACoTA was transformed in a matrix of presence (1) and absence (0) where gene families were located in the columns and the genomes in the rows. We defined the core genome of *Y. pestis* by selecting genes present in more than 90% of the genomes (3717 gene families). We excluded the core genes and the genes present in a single genome (singletons, n= 9001), resulting in 10583 gene families that are part of the accessory genome. We analyzed gene families that were chromosomally encoded (n= 10112) by creating lists of all gene families grouped by lineage. For a gene family to be included in a lineage list, this should be present in at least one genome of the lineage. We plotted the intersections between the lineage gene families lists using the UpSetR package (93). The matrix of presence and absence for the non-core and non-singleton chromosomally encoded gene families were also visualized as heatmaps using the heatmaply (94), where the genomes were

ordered with a dendrogram based on the phylogenetic tree (see section Phylogenetic Analyses) using the ape (95) and phylogram (96) R packages.

Insertion sequence analysis

IS present in *Y. pestis* CO92 reference genome were detected using ISSaga (97) for IS sequences that are present in the curated database ISFinder (98). IS types that were present in at least one complete copy with 80% nucleotide identity to an IS in the database were kept (IS100, IS200/IS1541, IS285, IS1661, ISYpe1 and ISYps7). To obtain a preliminary estimation of the IS burden in the general *Y. pestis* and *Y. pseudotuberculosis* populations and compare both species in terms of IS, all genomes of both species marked as 'completed' in PATRIC database (n=62 *Y. pestis*, n=26 *Y. pseudotuberculosis*, as of March 2022) were downloaded. BLAST+ v.2.2.3 was used to search the six previously identified IS types within the closed genomes. Nucleotide BLAST+ hits with >95% identity and >90% coverage were kept and counts per each genome are reported in table S3.

To identify IS insertions within paired-end data, we employed the strategy developed by Hawkey et al (67). First, to allow more precise detection of IS insertion sites on *Y. pestis* chromosome, an IS-free version of *Y. pestis* CO92 reference chromosome was generated as follows. BLAST+ hits positions on the reference chromosome for copies of the six previously identified IS types (ISSaga) were annotated and inspected. The regions where the identified IS types copies fell were removed from the reference chromosome to generate the IS-free reference. Annotations of the CDS and gene features in the original, complete, chromosome were then transferred to the IS-free reference using RATT (99). The IS-free reference chromosome has been deposited in GitHub: <https://github.com/guillemmasfiol/GlobalPlague>. Then, we used ISMapper v2 (56) with default parameters to detect IS insertions for the six IS types onto the IS-free CO92 reference, using raw reads of a dataset of 245 modern *Y. pestis* genomes representative of all lineages and sublineages, together with 5 *Y. pseudotuberculosis* genomes as comparison (table S3). To place detected IS copies with ISMapper together with the phylogenetic context of the genomes analyzed in the dataset, reads were also mapped onto the IS-free reference using RedDog-nf with default parameters, and a maximum-likelihood phylogeny was built on the variants conserved in 95% of samples using IQ-TREE2. This phylogeny was used to generate the plot in fig. 5A, and the patristic distance in number of nodes for each taxon to the root of the tree was calculated using gotree. Finally, insertions of most prevalent IS types (IS100, IS200 and IS285) that were specific to at least one 1.ANT lineage genome were identified, and insertions falling within coding sequences (n=133) were extracted from the ISMapper output tables. Functional annotations for the list of genes that were putatively interrupted by IS in 1.ANT genomes were generated using eggNOG-mapper v2 (100) and counts for each COG category were plotted in ggplots2 R-package.

Bibliography

1. Robert D Perry and Jacqueline D Fetherston. *Yersinia pestis*—etiologic agent of plague. *Clinical microbiology reviews*, 10(1):35–66, 1997.
2. Nicolas Rascovan, Karl-Goran Sjogren, Kristian Kristiansen, Rasmus Nielsen, Eske Willerslev, Christelle Desnues, and Simon Rasmussen. Emergence and spread of basal lineages of *Yersinia pestis* during the neolithic decline. *CELL*, 176(1-2):295+, JAN 10 2019. ISSN 0092-8674. doi: 10.1016/j.cell.2018.11.005.
3. Giovanna Morelli, Yajun Song, Camila J. Mazzoni, Mark Eppinger, Philippe Roumagnac, David M. Wagner, Mirjam Feldkamp, Barica Kusecek, Amy J. Vogler, Yanjun Li, Yujun Cui, Nicholas R. Thomson, Thibaut Jombart, Raphael Leblois, Peter Lichtner, Lila Rahalison, Jeannine M. Petersen, Francois Balloux, Paul Keim, Thierry Wirth, Jacques Ravel, Ruifu Yang, Elisabeth Carniel, and Mark Achtman. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *NATURE GENETICS*, 42(12):1140+, DEC 2010. ISSN 1061-4036. doi: 10.1038/ng.705.
4. Michael B Prentice and Lila Rahalison. Plague. *The Lancet*, 369(9568):1196–1207, 2007.
5. Eugene Tikhomirov. Epidemiology and distribution of plague. *Plague manual: epidemiology, distribution, surveillance and control*, pages 11–41, 1999.
6. Eric Bertherat and É Bertherat. Plague around the world in 2019. *Weekly epidemiological record*, 4, 2019.
7. Rindra Randremanana, Voahangy Andrianaivoarimanana, Birgit Nikolay, Beza Ramasindrazana, Juliette Paireau, Quirine Astrid ten Bosch, Jean Marius Rakotondramanga, Soloandry Rahajandraibe, Soanandrasana Rahelinirina, Fanjasoa Rakotomanana, Feno M. Rakotoarimanana, Lea Bricette Randriamampionona, Vaoary Razafimbia, Mamy Jean De Dieu Randria, Mihaja Raberahona, Guillain Mikaty, Anne-Sophie Le Guern, Lamina Arthur Rakotonjanabelo, Charlotte Faty Ndiaye, Voahangy Rasolofo, Eric Bertherat, Maherisoa Ratsitorahina, Simon Cauchemez, Laurence Baril, Andre Spiegel, and Minoarisoa Rajerison. Epidemiological characteristics of an urban plague epidemic in madagascar, august–november, 2017: an outbreak report. *LANCET INFECTIOUS DISEASES*, 19(5):537–545, MAY 2019. ISSN 1473-3099. doi: 10.1016/S1473-3099(18)30730-8.
8. Aaron Aruna Abedi, Jean-Christophe Shako, Jean Gaudart, Bertrand Sudre, Benoit Kebela Ilunga, Stomy Karhemere Bi Shamamba, Georges Diatta, Bernard Davoust, Jean-Jacques Muyembe, Tamfum Renaud Piarroux, and Martine Piarroux. Ecologic features of plague outbreak areas, democratic republic of the congo, 2004–2014. *EMERGING INFECTIOUS DISEASES*, 24(2):210–220, FEB 2018. ISSN 1080-6040. doi: 10.3201/eid2402.160122.
9. Renaud Piarroux, Aaron Aruna Abedi, Jean-Christophe Shako, Benoit Kebela, Stomy Karhemere, Georges Diatta, Bernard Davoust, Didier Raoult, and Michel Drancourt. Plague epidemics and lice, democratic republic of the congo. *EMERGING INFECTIOUS DISEASES*, 19(3):505–506, MAR 2013. ISSN 1080-6040. doi: 10.3201/eid1903.121542.

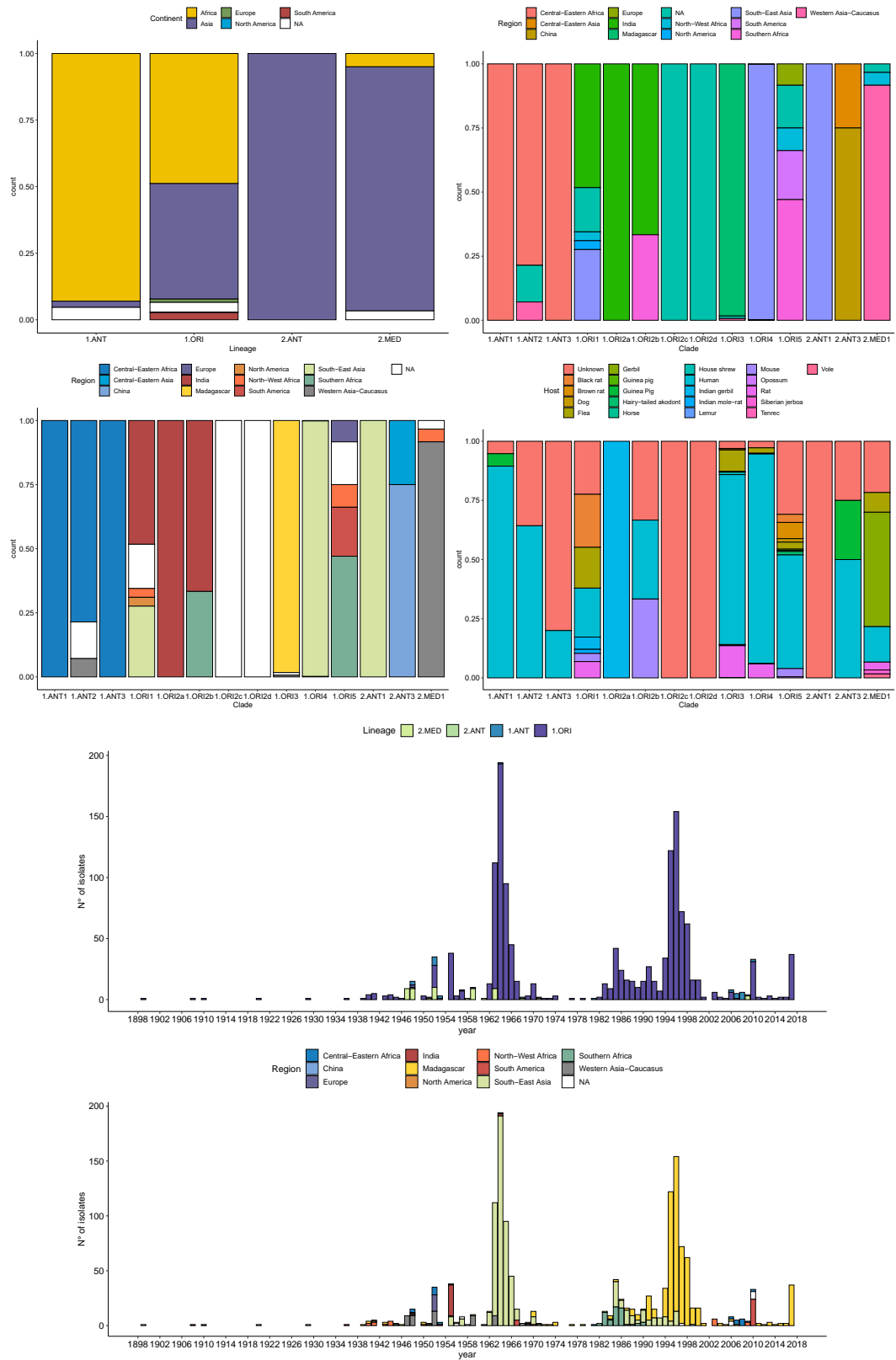
10. Xiaona Shen, Jianyun Li, Mengguang Fan, Lianxu Xia, and Wei Li. A reemergent case of bubonic plague-inner mongolia autonomous region, china, july, 2020. *China CDC Weekly*, 2(29):549–550, 2020.
11. Tamara Ben-Ari, Simon Neerincx, Lydiane Agier, Bernard Cazelles, Lei Xu, Zhibin Zhang, Xiye Fang, Shuchun Wang, Qiyong Liu, and Nils C. Stenseth. Identification of chinese plague foci from long-term epidemiological data. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 109(21):8196–8201, MAY 22 2012. ISSN 0027-8424. doi: 10.1073/pnas.1110585109.
12. BJ Hinnebusch and DL Erickson. *Yersinia pestis* biofilm in the flea vector and its role in the transmission of plague. *Bacterial biofilms*, pages 229–248, 2008.
13. Roger D. Pechous, Vijay Sivaraman, Nikolas M. Stasulli, and William E. Goldman. Pneumonic plague: The darker side of *Yersinia pestis*. *TRENDS IN MICROBIOLOGY*, 24(3):190–197, MAR 2016. ISSN 0966-842X. doi: 10.1016/j.tim.2015.11.008.
14. Christian E Demeure, Olivier Dussurget, Guillem Mas Fiol, Anne-Sophie Le Guern, Cyril Savin, and Javier Pizarro-Cerdá. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes & Immunity*, 20(5):357–370, 2019.
15. M Achtman, K Zurth, C Morelli, G Torrea, A Guiyoule, and E Carniel. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 96(24):14043–14048, NOV 23 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.24.14043.
16. Alan McNally, Nicholas R. Thomson, Sandra Reuter, and Brendan W. Wren. 'add, stir and reduce': *Yersinia* spp. as model bacteria for pathogen evolution. *NATURE REVIEWS MICROBIOLOGY*, 14(3):177–190, MAR 2016. ISSN 1740-1526. doi: 10.1038/nrmicro.2015.29.
17. Dongsheng Zhou and Ruifu Yang. Molecular darwinian evolution of virulence in *Yersinia pestis*. *Infection and immunity*, 77(6):2242–2250, 2009.
18. PSG Chain, E Carniel, FW Larimer, J Lamerdin, PO Stoutland, WM Regala, AM Georgescu, LM Vergez, ML Land, VL Motin, RR Brubaker, J Fowler, J Hinnebusch, M Marceau, C Medigue, M Simonet, V Chenal-Francisque, B Souza, D Dacheux, JM Elliott, A Derbise, LJ Hauser, and E Garcia. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 101(38):13826–13831, SEP 21 2004. ISSN 0027-8424. doi: 10.1073/pnas.0404012101.
19. Daniel L. Zimble, Jay A. Schroeder, Justin L. Eddy, and Wyndham W. Lathem. Early emergence of *Yersinia pestis* as a severe respiratory pathogen. *NATURE COMMUNICATIONS*, 6, JUN 2015. ISSN 2041-1723. doi: 10.1038/ncomms8487.
20. David M Bland, Adélaïde Miarinjara, Christopher F Bosio, Jeanette Calarco, and B Joseph Hinnebusch. Acquisition of *Yersinia murine* toxin enabled *Yersinia pestis* to expand the range of mammalian hosts that sustain flea-borne plague. *PLoS Pathogens*, 17(10):e1009995, 2021.
21. Yujun Cui, Chang Yu, Yanfeng Yan, Dongfang Li, Yanjun Li, Thibaut Jombart, Lucy A. Weinert, Zuyun Wang, Zhaobiao Guo, Lizhi Xu, Yujiang Zhang, Hancheng Zheng, Nan Qin, Xiao Xiao, Mingshou Wu, Xiaoyi Wang, Dongsheng Zhou, Zhizhen Qi, Zongmin Du, Honglong Wu, Xianwei Yang, Hongzhi Cao, Hu Wang, Jing Wang, Shusen Yao, Alexander Rakin, Yingrui Li, Daniel Falush, Francois Balloux, Mark Achtman, Yajun Song, Jun Wang, and Ruifu Yang. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 110(2):577–582, JAN 8 2013. ISSN 0027-8424. doi: 10.1073/pnas.1205750110.
22. Angelina A Kislichkina, Aleksandr G Bogun, Lidiya A Kadnikova, Nadezhda V Maiskaya, Mikhail E Platonov, Nikolai V Anisimov, Elena V Galkina, Svetlana V Dentovskaya, and Andrey P Anisimov. Nineteen whole-genome assemblies of *Yersinia pestis* subsp. *microtus*, including representatives of biovars caucasica, talassica, hissarica, altaica, xilingolensis, and uleigeica. *Genome announcements*, 3(6):e01342–15, 2015.
23. Vladimir V. Kutyrev, Galina A. Eroshenko, Vladimir L. Motin, Nikita Y. Nosov, Jaroslav M. Krasnov, Lyubov M. Kukleva, Konstantin A. Nikiforov, Zhanna V. Al'khova, Eugene G. Oglodin, and Natalia P. Guseva. Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of commonwealth of independent states. *FRONTIERS IN MICROBIOLOGY*, 9, MAY 25 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.01106.
24. Galina A. Eroshenko, Nikolay V. Popov, Zhanna V. Al'khova, Lyubov M. Kukleva, Alina N. Balykova, Nadezhda S. Chervyakova, Ekaterina A. Naryshkina, and Vladimir V. Kutyrev. Evolution and circulation of *Yersinia pestis* in the northern caspian and northern aral sea regions in the 20(th)-21(st) centuries. *PLOS ONE*, 16(2), FEB 11 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0244615.
25. Marcel Keller, Maria A. Spyrou, Christiana L. Scheib, Gunnar U. Neumann, Andreas Kropelin, Brigitte Haas-Gebhard, Bernd Paffgen, Jochen Haberstroh, Albert Ribera i Lacomba, Claude Raynaud, Craig Cessford, Raphael Durand, Peter Stadler, Kathrin Nagele, Jessica S. Bates, Bernd Trautmann, Sarah A. Inskip, Joris Peters, John E. Robb, Toomas Kivisild, Dominique Castex, Michael McCormick, Kirsten I. Bos, Michaela Harbeck, Alexander Herbig, and Johannes Krause. Ancient *Yersinia pestis* genomes from across western europe reveal early diversification during the first pandemic (541-750). *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 116(25):12363–12372, JUN 18 2019. ISSN 0027-8424. doi: 10.1073/pnas.1820447116.
26. Maria A. Spyrou, Marcel Keller, Rezeda I. Tukhbatova, Christiana L. Scheib, Elizabeth A. Nelson, Aida Andrades Valtuena, Gunnar U. Neumann, Don Walker, Amelie Alterauge, Niamh Carty, Craig Cessford, Hermann Fetz, Michael Gourvennech, Robert Hartle, Michael Henderson, Kristin von Heyking, Sarah A. Inskip, Sacha Kacki, Felix M. Key, Elizabeth L. Knox, Christian Later, Prishita Maheshwari-Aplin, Joris Peters, John E. Robb, Juergen Schreiber, Toomas Kivisild, Dominique Castex, Sandra Losch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos, and Johannes Krause. Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *NATURE COMMUNICATIONS*, 10, OCT 2 2019. doi: 10.1038/s41467-019-12154-0.
27. Simon Rasmussen, Morten Erik Allentoft, Kasper Nielsen, Ludovic Orlando, Martin Sikora, Karl-Goeran Sjoegren, Anders Gorm Pedersen, Mikkel Schubert, Alex Van Dam, Christian Mollin Outzen Kapel, Henrik Bjorn Nielsen, Soren Brunak, Pavel Avetisyan, Andrey Epimakhov, Mikhail Viktorovich Khalyapin, Artak Gnuni, Aivar Kriiska, Irena Lasak, Mait Metspalu, Vyacheslav Moiseyev, Andrei Gromov, Dalia Pokutta, Lehti Saag, Liivi Varul, Levon Yepiskoposyan, Thomas Sicheritz-Ponten, Robert A. Foley, Marta Mirazon Lahr, Rasmus Nielsen, Kristian Kristiansen, and Eske Willerslev. Early divergent strains of *Yersinia pestis* in eurasia 5,000 years ago. *CELL*, 163(3):571–582, OCT 22 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2015.10.009.
28. Aida Andrades Valtueña, Gunnar U Neumann, Maria A Spyrou, Lyazzat Musralina, Franziska Aron, Arman Beisenov, Andrey B Belinskiy, Kirsten I Bos, Alexandra Buzhilova, Matthias Conrad, et al. Stone age *Yersinia pestis* genomes shed light on the early evolution, diversity, and ecology of plague. *Proceedings of the National Academy of Sciences*, 119(17):e2116722119, 2022.
29. Zhemin Zhou, Nabil-Fareed Alikhan, Khaled Mohamed, Yulei Fan, Mark Achtman, Derek Brown, Marie Chattaway, Tim Dallman, Richard Delahay, Christian Kornschober, et al. The enterobase user's guide, with case studies on salmonella transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* core genomic diversity. *Genome research*, 30(1):138–152, 2020.
30. Zhemin Zhou, Nabil-Fareed Alikhan, Martin J Sergeant, Nina Luhmann, Cátia Vaz, Alexandre P Francisco, João André Carriço, and Mark Achtman. Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome research*, 28(9):1395–1404, 2018.
31. A Yersin. Sur la peste de hongkong (lundi 30 juillet 1894). *CR Hebd Séances Acad Sci*, 119:356, 1894.
32. Aida Andrades Valtueña, Alissa Mittnik, Felix M Key, Wolfgang Haak, Raili Allmäe, Andrej Belinskiy, Mantas Daubaras, Michal Feldman, Rimantas Jankauskas, Ivor Janković, et al. The stone age plague and its persistence in eurasia. *Current biology*, 27(23):3683–3691, 2017.
33. Nicolas Cabanel, Alexandre Leclercq, Viviane Chenal-Francisque, Badereddin Annajar, Minoarisoa Rajerison, Souad Bekkhoucha, Eric Bertherat, and Elisabeth Carniel. Plague outbreak in libya, 2009, unrelated to plague in algeria. *Emerging Infectious Diseases*, 19(2):230, 2013.

34. Remco Bouckaert, Timothy G Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, et al. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS computational biology*, 15(4):e1006650, 2019.
35. R. Pollitzer. Plague studies*. *Bulletin of the World Health Organization*, 4(4):475 – 533, 1951.
36. RKK Molefi. Of rats, fleas, and peoples: towards a history of bubonic plague in southern africa, 1890-1950. 2001.
37. Eric Bertherat, Souad Bekhoucha, Saada Chougrani, Fathia Razik, Jean B Duchemin, Leila Houti, Larbi Deharib, Corinne Fayolle, Banaouda Makrrougrass, Radia Dali-Yahia, et al. Plague reappearance in algeria after 50 years, 2003. *Emerging infectious diseases*, 13(10):1459, 2007.
38. Ricardo Jorge. *La peste africaine...* Office International d'Hygiène Publique, 1935.
39. Barbara Bramanti, Katharine R. Dean, Lars Walloe, and Nils Chr Stenseth. The third plague pandemic in europe. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 286(1901), APR 17 2019. ISSN 0962-8452. doi: 10.1098/rspb.2018.2429.
40. Simon Neerincx, Eric Bertherat, and Herwig Leirs. Human plague occurrences in africa: an overview from 1877 to 2008. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(2):97–103, 2010.
41. Myron Echenberg. Plague ports. In *Plague Ports*. New York University Press, 2007.
42. Monica H. Green. Putting africa on the black death map: Narratives from genetics and history. *AFRIQUES-DEBATS METHODES ET TERRAINS D HISTOIRE*, 9, 2018. ISSN 2108-6796. doi: 10.4000/afriques.2125.
43. Ingrid Wiechmann, Ole Jørgen Benedictow, Raffaella Bianucci, and Sacha Kacki. History of the plague. *RCC Perspectives*, (3):63–74, 2012.
44. Orhan Koloğlu. Libya, from the ottoman perspective (1835-1918). *Africa: Rivista trimestrale di studi e documentazione dell'Istituto italiano per l'Africa e l'Oriente*, 63(2):275–282, 2008.
45. Caroline Colijn and Jennifer Gardy. Phylogenetic tree shapes resolve disease transmission patterns. *Evolution, medicine, and public health*, 2014(1): 96–108, 2014.
46. Kathryn E Holt, Paul McAdam, Phan Vuong Khac Thai, Nguyen Thuy Thuong, Dang Thi Minh Ha, Nguyen Ngoc Lan, Nguyen Huu Lan, Nguyen Thi Quynh Nhu, Hoang Thanh Hai, Vu Thi Ngoc Ha, et al. Frequent transmission of the mycobacterium tuberculosis beijing lineage and positive selection for the esw beijing variant in vietnam. *Nature genetics*, 50(6):849–856, 2018.
47. Luca Freschi, Roger Vargas, Ashaque Husain, SM Kamal, Alena Skrahina, Sabira Tahseen, Nazir Ismail, Anna Barbova, Stefan Niemann, Daniela Maria Cirillo, et al. Population structure, biogeography and transmissibility of mycobacterium tuberculosis. *Nature communications*, 12(1):1–11, 2021.
48. Yanfeng Yan, Hu Wang, Dongfang Li, Xianwei Yang, Zuyun Wang, Zhizhen Qi, Qingwen Zhang, Baizhong Cui, Zhaobiao Guo, Chang Yu, Jun Wang, Jian Wang, Guangming Liu, Yajun Song, Yingrui Li, Yujun Cui, and Ruifu Yang. Two-step source tracing strategy of yersinia pestis and its historical epidemiology in a specific region. *PLOS ONE*, 9(1), JAN 9 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0085374.
49. Xiaoqing Xu, Yujun Cui, Youquan Xin, Xiaoyan Yang, Qingwen Zhang, Yong Jin, Haihong Zhao, Jian He, Xing Jin, Cunxiang Li, Juan Jin, Xiang Li, Haisheng Wu, and Zhizhen Qi. Genetic diversity and spatial-temporal distribution of yersinia pestis in qinghai plateau, china. *PLOS NEGLECTED TROPICAL DISEASES*, 12(6), JUN 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006579.
50. Laurel B. Respicio-Kingry, Brook M. Yockey, Sarah Acayo, John Kaggwa, Titus Apangu, Kiersten J. Kugeler, Rebecca J. Eisen, Kevin S. Griffith, Paul S. Mead, Martin E. Schriefer, and Jeannine M. Petersen. Two distinct yersinia pestis populations causing plague among humans in the west Nile region of Uganda. *PLOS NEGLECTED TROPICAL DISEASES*, 10(2), FEB 2016. ISSN 1935-2735. doi: 10.1371/journal.pntd.0004360.
51. Russell E. Ensore, Nackson Babi, Gerald Amate, Linda Atiku, Rebecca J. Eisen, Kimberly M. Pepin, Rommelle Vera-Tudela, Christopher Sexton, and Kenneth L. Gage. The changing triad of plague in Uganda: invasive black rats (*Rattus rattus*), indigenous small mammals, and their fleas. *JOURNAL OF VECTOR ECOLOGY*, 45(2):333–355, DEC 2020. ISSN 1081-1710.
52. Scott W Bearden and Robert R Brubaker. Recent findings regarding maintenance of enzootic variants of yersinia pestis in sylvatic reservoirs and their significance in the evolution of epidemic plague. *Vector-Borne and Zoonotic Diseases*, 10(1):85–92, 2010.
53. AP Anisimov, LE Lindler, and GB Pier. Intraspecific diversity of yersinia pestis. *CLINICAL MICROBIOLOGY REVIEWS*, 17(2):434+, APR 2004. ISSN 0893-8512. doi: 10.1128/CMR.17.2.434-464.2004.
54. J Parkhill, BW Wren, NR Thomson, RW Titball, MTG Holden, MB Prentice, M Sebahia, KD James, C Churcher, KL Mungall, S Baker, D Basham, SD Bentley, K Brooks, AM Cerdeno-Tarraga, T Chillingworth, A Cronin, RM Davies, P Davis, G Dougan, T Feltwell, N Hamlin, S Holroyd, K Jagels, AV Karlyshev, S Leather, S Moule, PCF Oyston, M Quail, K Rutherford, M Simmonds, J Skelton, K Stevens, S Whitehead, and BG Barrell. Genome sequence of yersinia pestis, the causative agent of plague. *NATURE*, 413(6855):523–527, OCT 4 2001. ISSN 0028-0836. doi: 10.1038/35097083.
55. Katy J Califf, Paul S Keim, David M Wagner, and Jason W Sahl. Redefining the differences in gene content between yersinia pestis and yersinia pseudotuberculosis using large-scale comparative genomics. *Microbial genomics*, 1(2), 2015.
56. Jane Hawkey, Mohammad Hamidian, Ryan R Wick, David J Edwards, Helen Billman-Jacobe, Ruth M Hall, and Kathryn E Holt. IsMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data. *BMC genomics*, 16(1):1–11, 2015.
57. Liyuan Shi, Guirong Yang, Zhikai Zhang, Lianxu Xia, Ying Liang, Hongli Tan, Jinrong He, Jianguo Xu, Zhizhong Song, Wei Li, and Peng Wang. Reemergence of human plague in yunnan, china in 2016. *PLOS ONE*, 13(6), JUN 13 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0198067.
58. George D. Sussman. Scientists doing history: Central africa and the origins of the first plague pandemic. *JOURNAL OF WORLD HISTORY*, 26(2): 325–354, JUN 2015. ISSN 1045-6007. doi: 10.1353/jwh.2016.0030.
59. Fabrizio Menardo. Understanding drivers of phylogenetic clustering and terminal branch lengths distribution in epidemics of *Mycobacterium tuberculosis*. *eLife*, 11:e76780, jun 2022. ISSN 2050-084X. doi: 10.7554/eLife.76780.
60. Katherine Eaton, Leo Featherstone, Sebastian Duchene, Ann Carmichael, Nükhet Varlik, Brian Golding, Edward Holmes, and Hendrik Poinar. Plagued by a cryptic clock: Insight and issues from the global phylogeny of yersinia pestis, 2021.
61. Mark Achtman. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual review of microbiology*, 62 (1):53–70, 2008.
62. Yujun Cui, Boris V Schmid, Hanli Cao, Xiang Dai, Zongmin Du, W Ryan Easterday, Haihong Fang, Chenyi Guo, Shanqian Huang, Wanbing Liu, et al. Evolutionary selection of biofilm-mediated extended phenotypes in yersinia pestis in response to a fluctuating environment. *Nature communications*, 11 (1):1–8, 2020.
63. M Simonet, B Riot, N Fortineau, and P Berche. Invasin production by yersinia pestis is abolished by insertion of an is200-like element within the inv gene. *INFECTIO AND IMMUNITY*, 64(1):375–379, JAN 1996. ISSN 0019-9567. doi: 10.1128/IAI.64.1.375-379.1996.
64. Julian Parkhill, Mohammed Sebahia, Andrew Preston, Lee D Murphy, Nicholas Thomson, David E Harris, Matthew TG Holden, Carol M Churcher, Stephen D Bentley, Karen L Mungall, et al. Comparative analysis of the genome sequences of bordetella pertussis, bordetella parapertussis and bordetella bronchiseptica. *Nature genetics*, 35(1):32–40, 2003.
65. ST Cole, K Eiglmeier, J Parkhill, KD James, NR Thomson, PR Wheeler, N Honore, T Garnier, C Churcher, D Harris, et al. Massive gene decay in the leprosy bacillus. *Nature*, 409(6823):1007–1011, 2001.

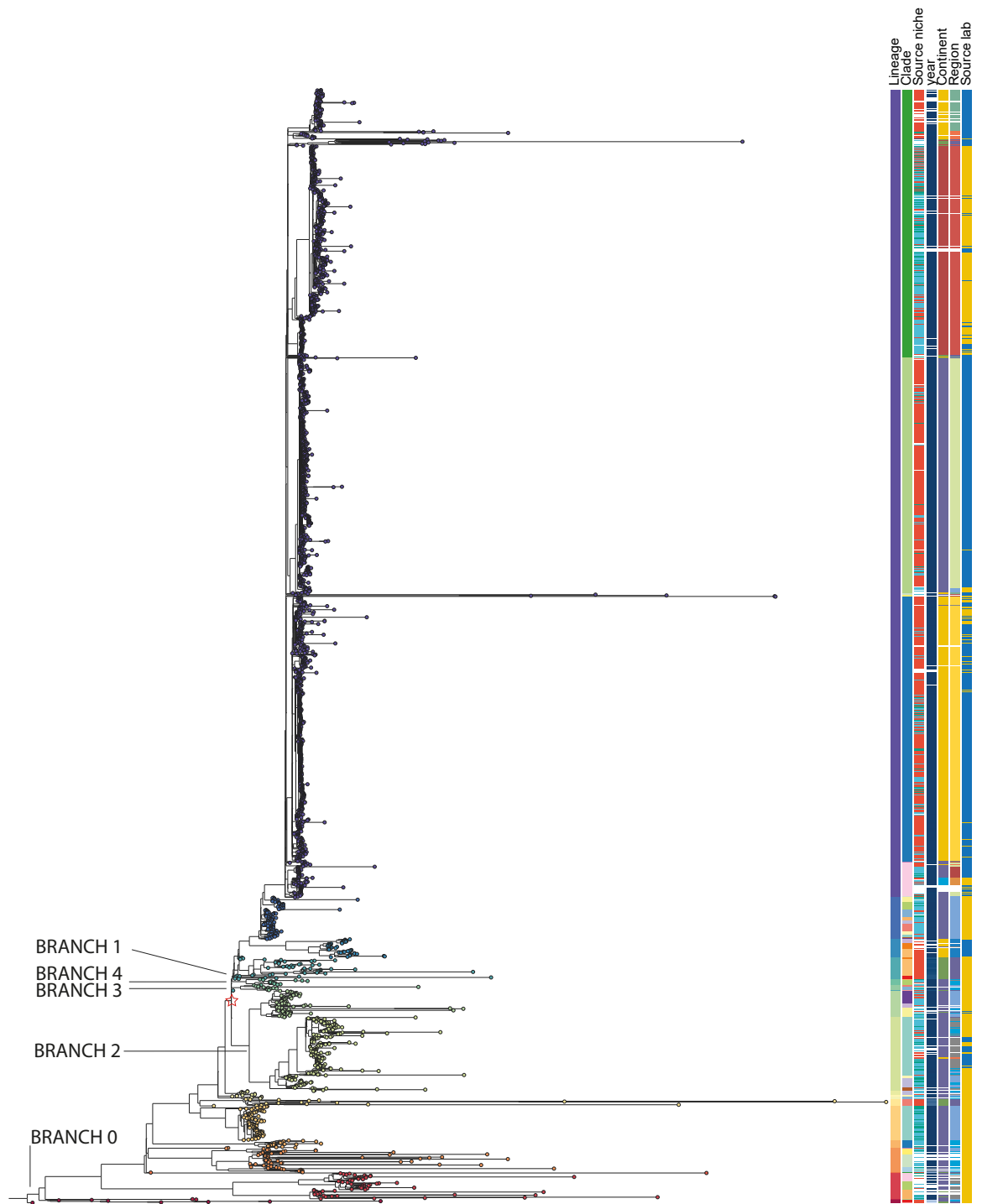
66. Stewart T Cole, Philippe Supply, and Nadine Honoré. Repetitive sequences in mycobacterium leprae and their impact on genome plasticity. *Leprosy review*, 72(4):449–461, 2001.
67. Jane Hawkey, Jonathan M Monk, Helen Billman-Jacobe, Bernhard Palsson, and Kathryn E Holt. Impact of insertion sequences on convergent evolution of shigella species. *PLoS genetics*, 16(7):e1008931, 2020.
68. Yarong Wu, Tongyu Hao, Xiuwei Qian, Xianglilan Zhang, Yajun Song, Ruifu Yang, and Yujun Cui. Small insertions and deletions drive genomic plasticity during adaptive evolution of yersinia pestis. *Microbiology spectrum*, 10(3):e02242–21, 2022.
69. Alexis Criscuolo and Sylvain Brisse. Alientrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics*, 102(5-6):500–506, 2013.
70. Michael R Crusoe, Hussien F Alameldin, Sherine Awad, Elmar Boucher, Adam Caldwell, Reed Cartwright, Amanda Charbonneau, Bede Constantinides, Greg Edverson, Scott Fay, et al. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research*, 4, 2015.
71. Yongchao Liu, Jan Schröder, and Bertil Schmidt. Musket: a multistage k-mer spectrum-based error corrector for illumina sequence data. *Bioinformatics*, 29(3):308–315, 2013.
72. Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, Alexander S Kulikov, Valery M Lesin, Sergey I Nikolenko, Son Pham, Andrey D Prjibelski, et al. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5):455–477, 2012.
73. Paolo Di Tommaso, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, and Cedric Notredame. Nextflow enables reproducible computational workflows. *Nature biotechnology*, 35(4):316–319, 2017.
74. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.
75. Mikkel Schubert, Stinus Lindgreen, and Ludovic Orlando. Adapterremoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes*, 9:88, Feb 2016. doi: 10.1186/s13104-016-1900-2.
76. Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *bioinformatics*, 25(14):1754–1760, 2009.
77. Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo Del Angel, Manuel A Rivas, Matt Hanna, et al. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5):491–498, 2011.
78. Hakon Jonsson, Aurelien Ginolhac, Mikkel Schubert, Philip L. F. Johnson, and Ludovic Orlando. mapdamage2.0: fast approximate bayesian estimates of ancient dna damage parameters. *BIOINFORMATICS*, 29(13):1682–1684, JUL 1 2013. ISSN 1367-4803. doi: 10.1093/bioinformatics/btt193.
79. Alexey M Kozlov, Diego Darriba, Tomás Flouri, Benoit Morel, and Alexandros Stamatakis. Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21):4453–4455, 2019.
80. Vincent Lefort, Richard Desper, and Olivier Gascuel. Fastme 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular biology and evolution*, 32(10):2798–2800, 2015.
81. Frédéric Lemoine and Olivier Gascuel. Gootree/goalign: toolkit and go api to facilitate the development of phylogenetic workflows. *NAR Genomics and Bioinformatics*, 3(3):lqab075, 2021.
82. Randle Aaron M Villanueva and Zhuo Job Chen. ggplot2: elegant graphics for data analysis, 2019.
83. Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *MOLECULAR BIOLOGY AND EVOLUTION*, 37(5):1530–1534, MAY 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.
84. Subha Kalyaanamoorthy, Bui Quang Minh, Thomas KF Wong, Arndt Von Haeseler, and Lars S Jermin. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6):587–589, 2017.
85. Daniel Ayres, Michael Cummings, Guy Baele, Aaron Darling, Paul Lewis, David Swofford, John Huelsenbeck, Philippe Lemey, Andrew Rambaut, and Marc Suchard. Beagle 3: Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *SYSTEMATIC BIOLOGY*, 68(6):1052–1061, NOV 2019. ISSN 1063-5157. doi: 10.1093/sysbio/syz020.
86. Marc A. Suchard, Philippe Lemey, Guy Baele, Daniel L. Ayres, Alexei J. Drummond, and Andrew Rambaut. Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *VIRUS EVOLUTION*, 4(1), JAN 2018. doi: 10.1093/ve/vey016.
87. Pablo Cingolani, Adrian Platts, Le Lily Wang, Melissa Coon, Tung Nguyen, Luan Wang, Susan J. Land, Xiangyi Lu, and Douglas M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w(1118); iso-2; iso-3. *FLY*, 6(2):80–92, APR-JUN 2012. ISSN 1933-6934. doi: 10.4161/fly.19695.
88. Ramy K Aziz, Daniela Bartels, Aaron A Best, Matthew DeJongh, Terrence Disz, Robert A Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M Glass, Michael Kubal, et al. The rast server: rapid annotations using subsystems technology. *BMC genomics*, 9(1):1–15, 2008.
89. Erik S Wright et al. Using decipher v2. 0 to analyze big biological sequence data in r. *R J.*, 8(1):352, 2016.
90. Tal Galili. dendextend: an r package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22):3718–3720, 2015.
91. Sébastien Lê, Julie Josse, and François Husson. Factominer: an r package for multivariate analysis. *Journal of statistical software*, 25:1–18, 2008.
92. Amandine Perrin and Eduardo PC Rocha. Panacota: a modular tool for massive microbial comparative genomics. *NAR genomics and bioinformatics*, 3(1):lqaa106, 2021.
93. Jake R Conway, Alexander Lex, and Nils Gehlenborg. Upsetr: an r package for the visualization of intersecting sets and their properties. *Bioinformatics*, 2017.
94. Tal Galili, Alan O’Callaghan, Jonathan Sidi, and Carson Sievert. heatmaply: an r package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*, 34(9):1600–1602, 2018.
95. Emmanuel Paradis and Klaus Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in r. *Bioinformatics*, 35(3):526–528, 2019.
96. Shaun P Wilkinson and Simon K Davy. phylogram: an r package for phylogenetic analysis with nested lists. *Journal of Open Source Software*, 3(26):790, 2018.
97. Alessandro M Varani, Patricia Siguier, Edith Gourbeyre, Vincent Charneau, and Mick Chandler. Issaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome biology*, 12(3):1–9, 2011.
98. Patricia Siguier, Jocelyne Pérochon, L Lestrade, Jacques Mahillon, and Michael Chandler. Isfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research*, 34(suppl_1):D32–D36, 2006.
99. Thomas D Otto, Gary P Dillon, Wim S Degraeve, and Matthew Berriman. Ratt: rapid annotation transfer tool. *Nucleic acids research*, 39(9):e57–e57, 2011.
100. Carlos P Cantalapiedra, Ana Hernández-Plaza, Ivica Letunic, Peer Bork, and Jaime Huerta-Cepas. eggnoG-mapper v2: functional annotation, orthology

assignments, and domain prediction at the metagenomic scale. *Molecular biology and evolution*, 38(12):5825–5829, 2021.

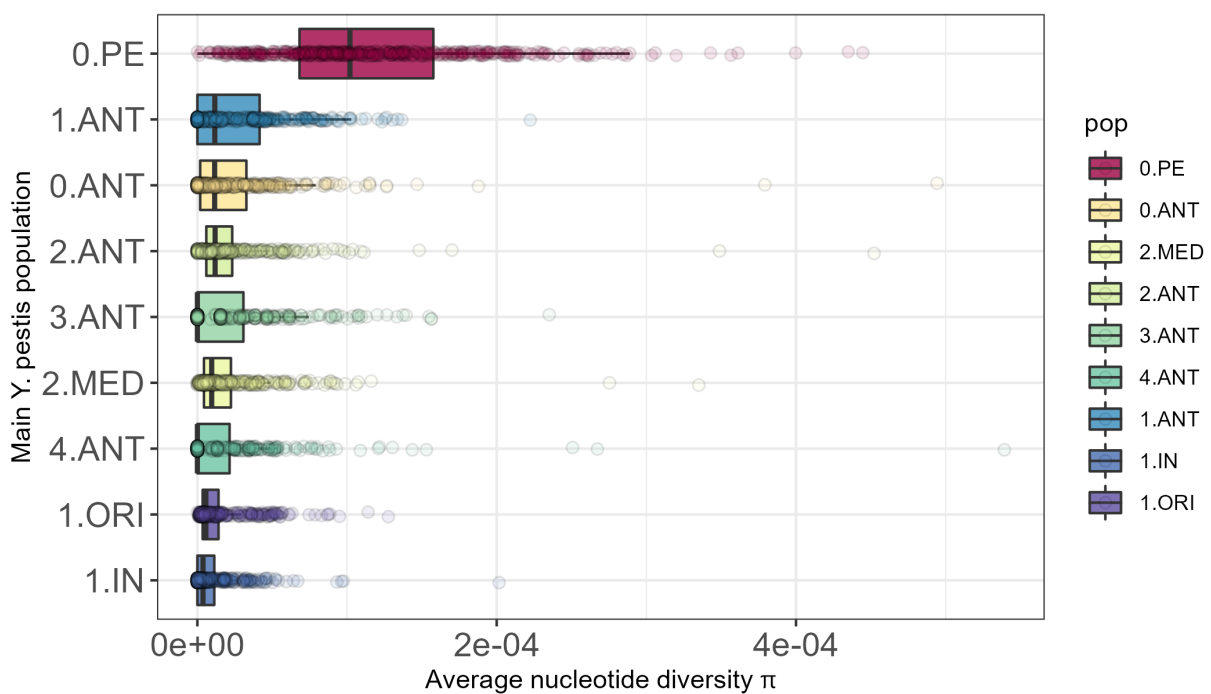
Supplementary Figures



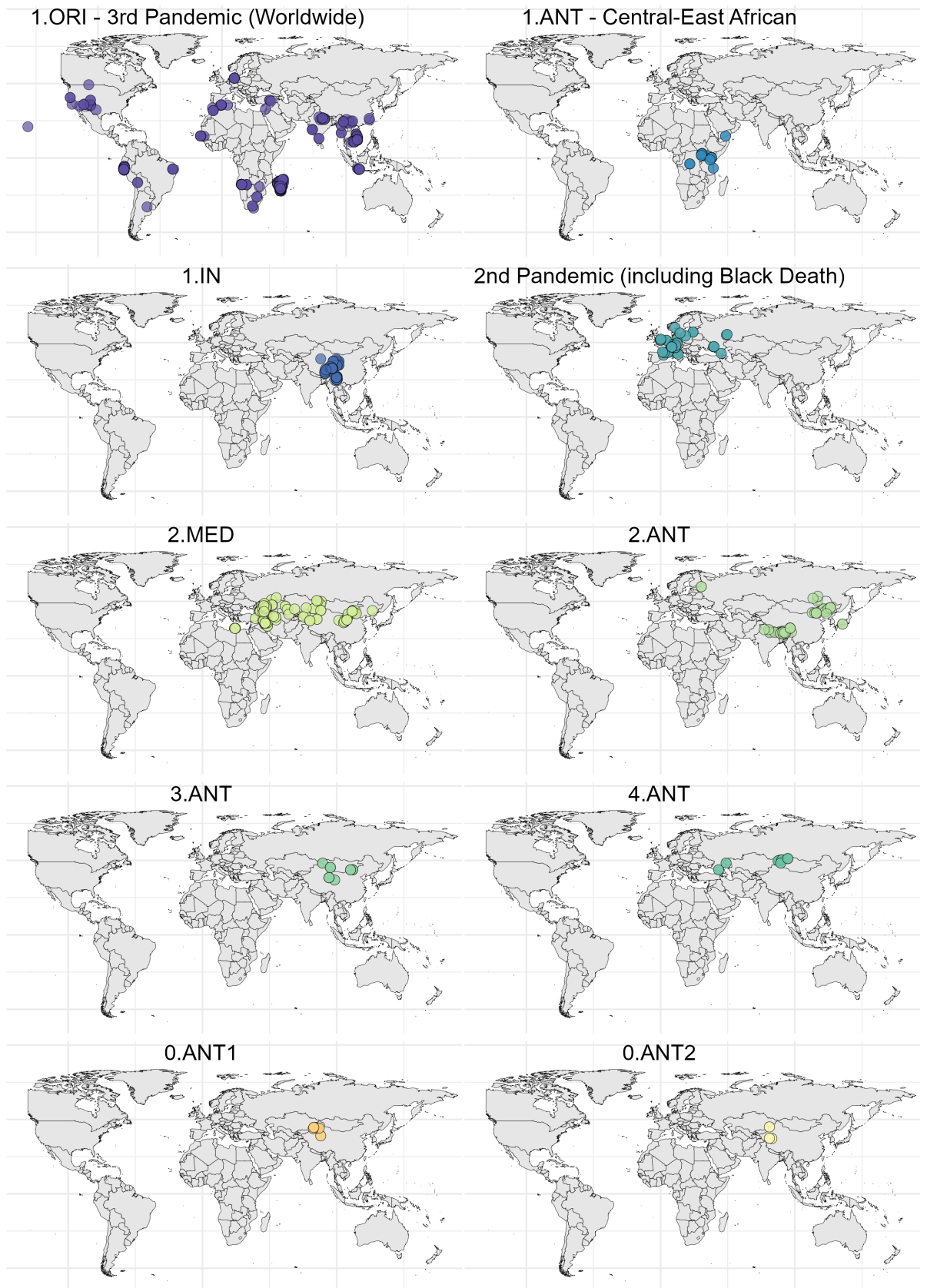
Supplementary Figure 1. Geographical and temporal distribution of the genomes sequenced in this study from the Institut Pasteur collection (n=1549).



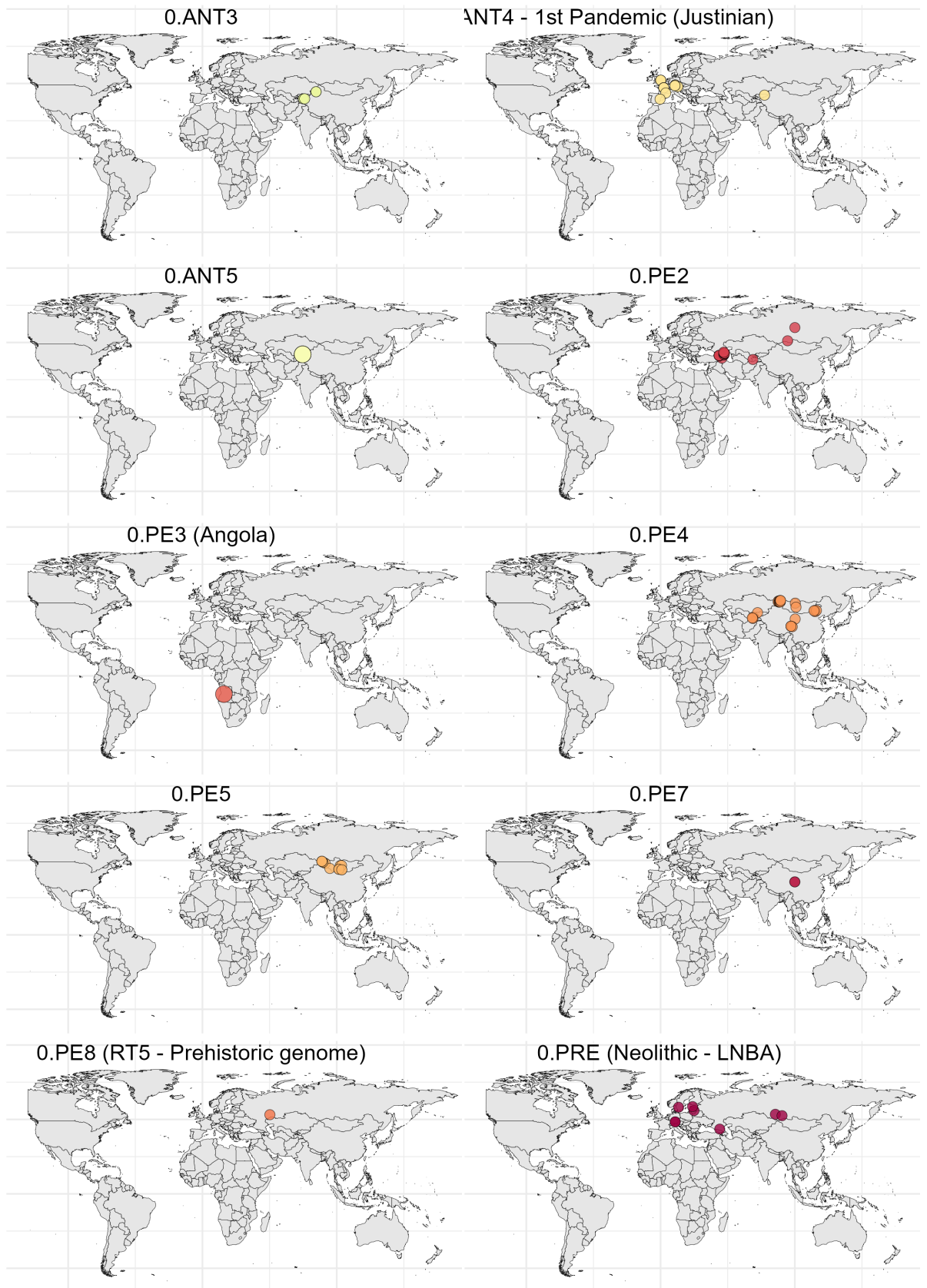
Supplementary Figure 2. Maximum-likelihood phylogeny of the 2806 modern and ancient *Y. pestis* genomes analyzed in this study. The nodes defining the start of major branches (0-4) as defined in previous studies are indicated. The star marks the node of emergence of branches 1-4, also known as "Big Bang". Color strips include annotations for the correspondence of genomes to major lineages, sublineages or clades, isolation niche, year, continent, geographic region and source laboratory, respectively. Color legend corresponds to the tree shown in figure 1A.



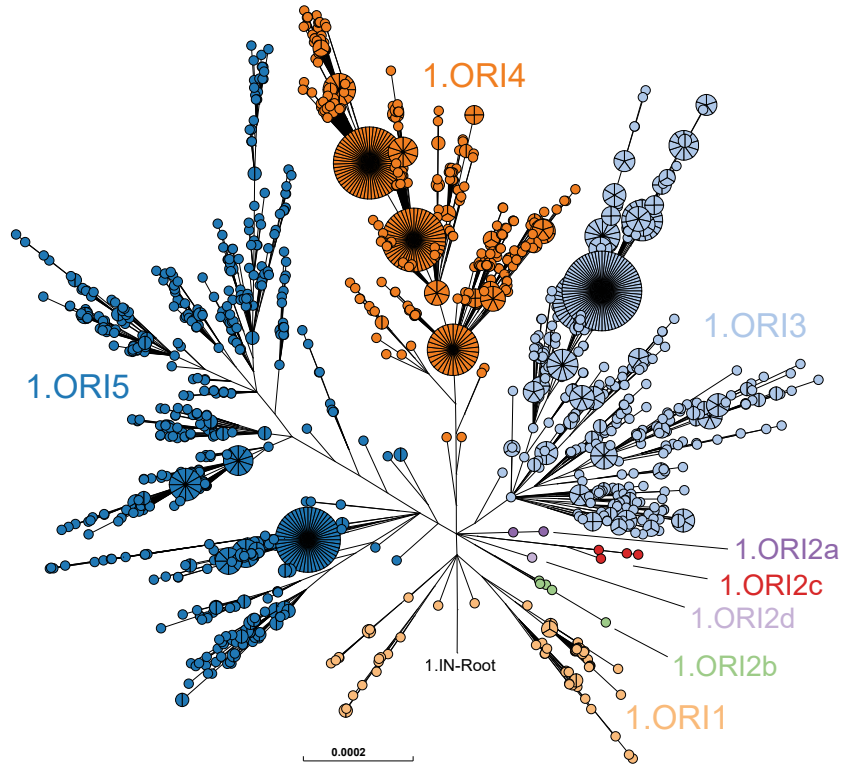
Supplementary Figure 3. Average nucleotide diversity within genomes from groups of modern *Y. pestis* lineages (n=2723 genomes). Average nucleotide diversity estimates were computed by windows of 10000 bp from the variant tables (VCF), each one representing a point in the displayed distributions.



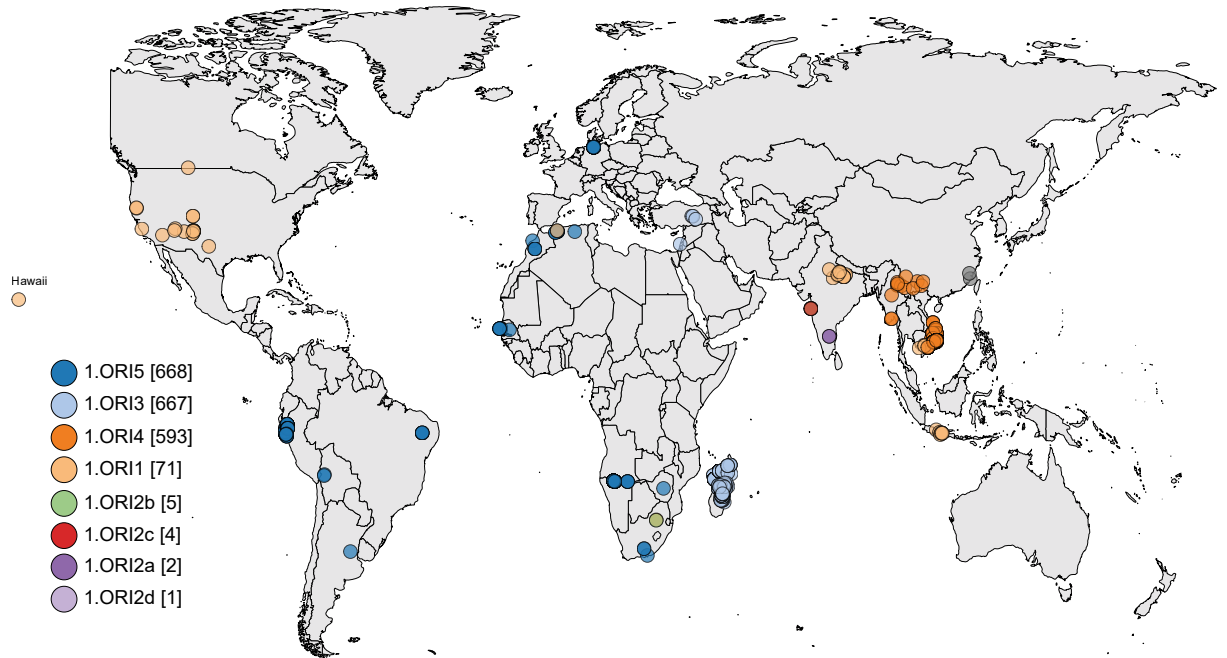
Supplementary Figure 4. Part 1 (continues in next page). Cartography of the locations of genomes for each of the 20 major lineages included in the 2806 phylogeny. Only genomes for which at least the location at country level is known are displayed (n=2675). Coordinates correspond to the maximum level of precision when available (commune), whereas centroid points are shown when only the country of isolation of the genome was available.



Supplementary Figure 4. Part 2. Cartography of the locations of genomes for each of the 20 major lineages included in the 2806 phylogeny. : Only genomes for which at least the location at country level is known are displayed (n=2675). Coordinates correspond to the maximum level of precision when available (commune), whereas centroid points are shown when only the country of isolation of the genome was available.

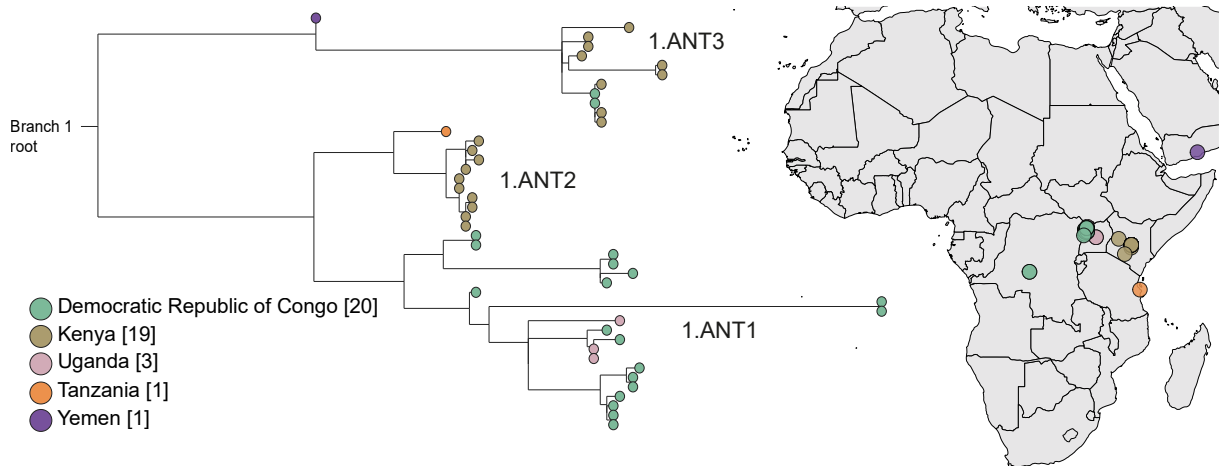


3rd Pandemic 1.ORI - Sublineages

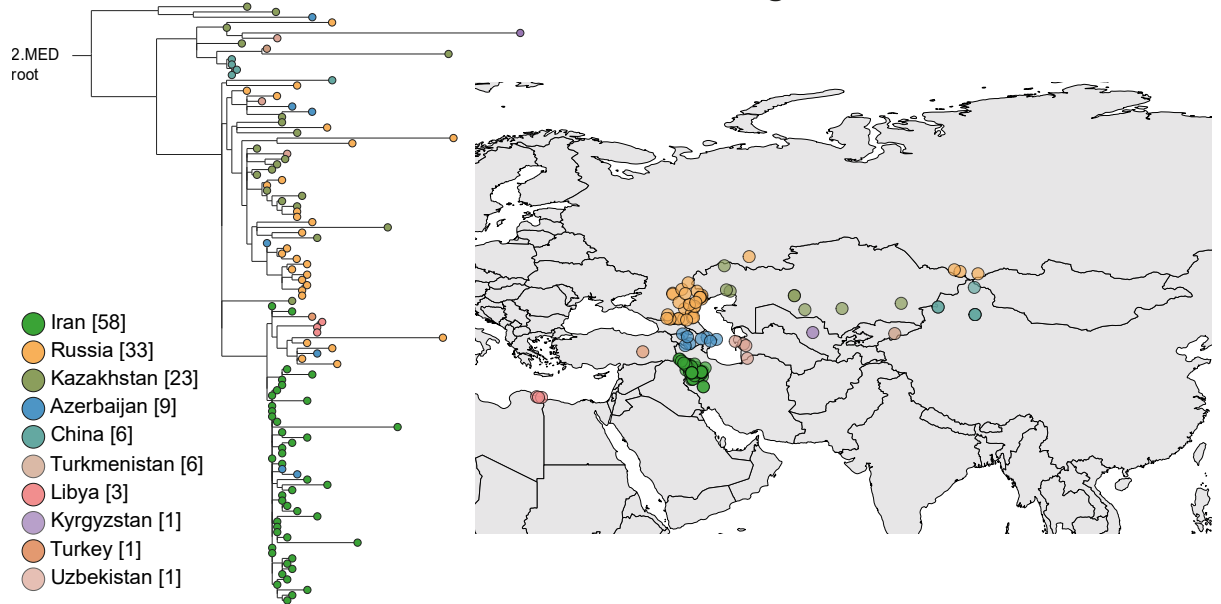


Supplementary Figure 5. Phylogeny and geographic distribution of the 2036 Third Pandemic 1.ORI lineage genomes. Up: A zoom on the 1.ORI lineage from the phylogenetic tree shown in figure 1C. The identified sublineages or clades within 1.ORI are indicated. **Down:** Geographic distribution of 1.ORI isolates. Points are colored according to the identified sublineages.

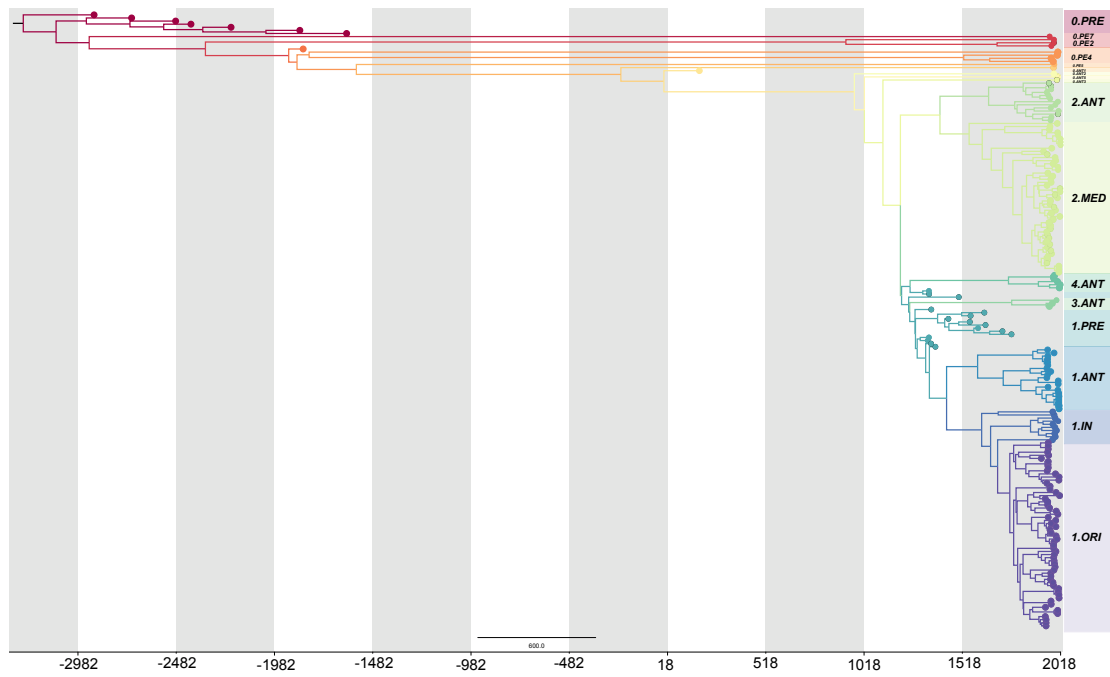
Central-East African 1.ANT lineage



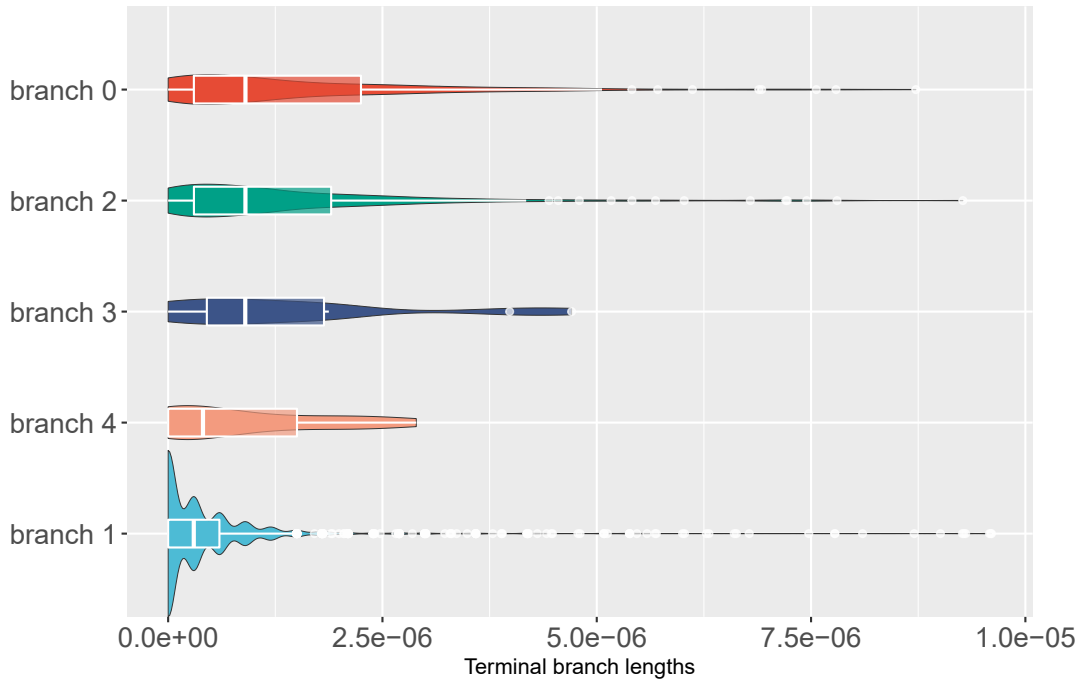
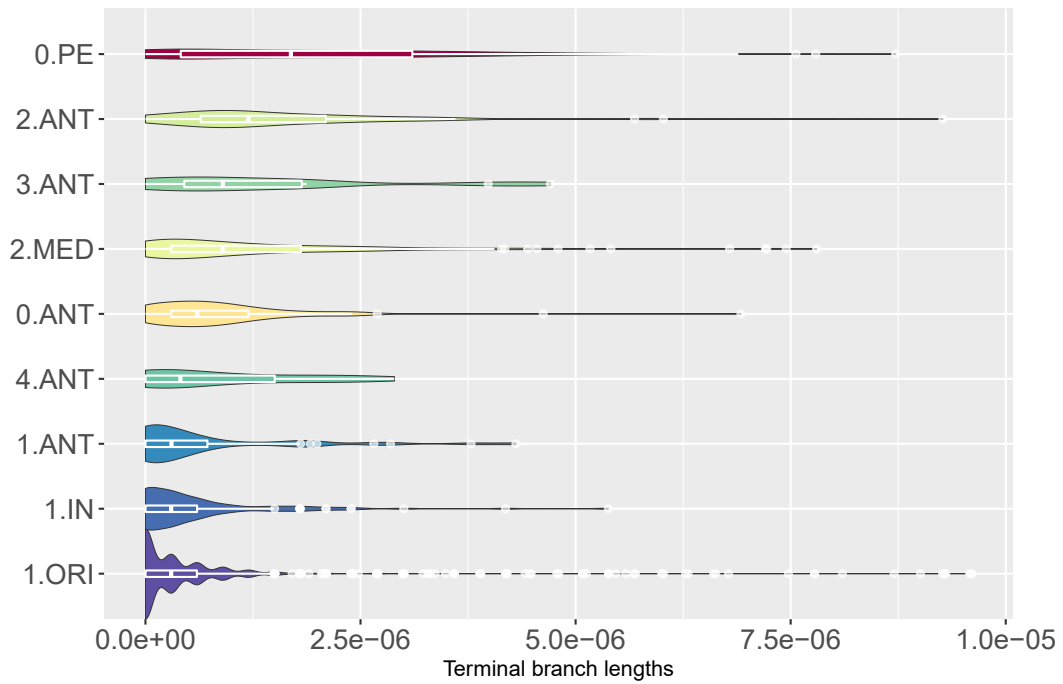
Asian 2.MED1 sublineage



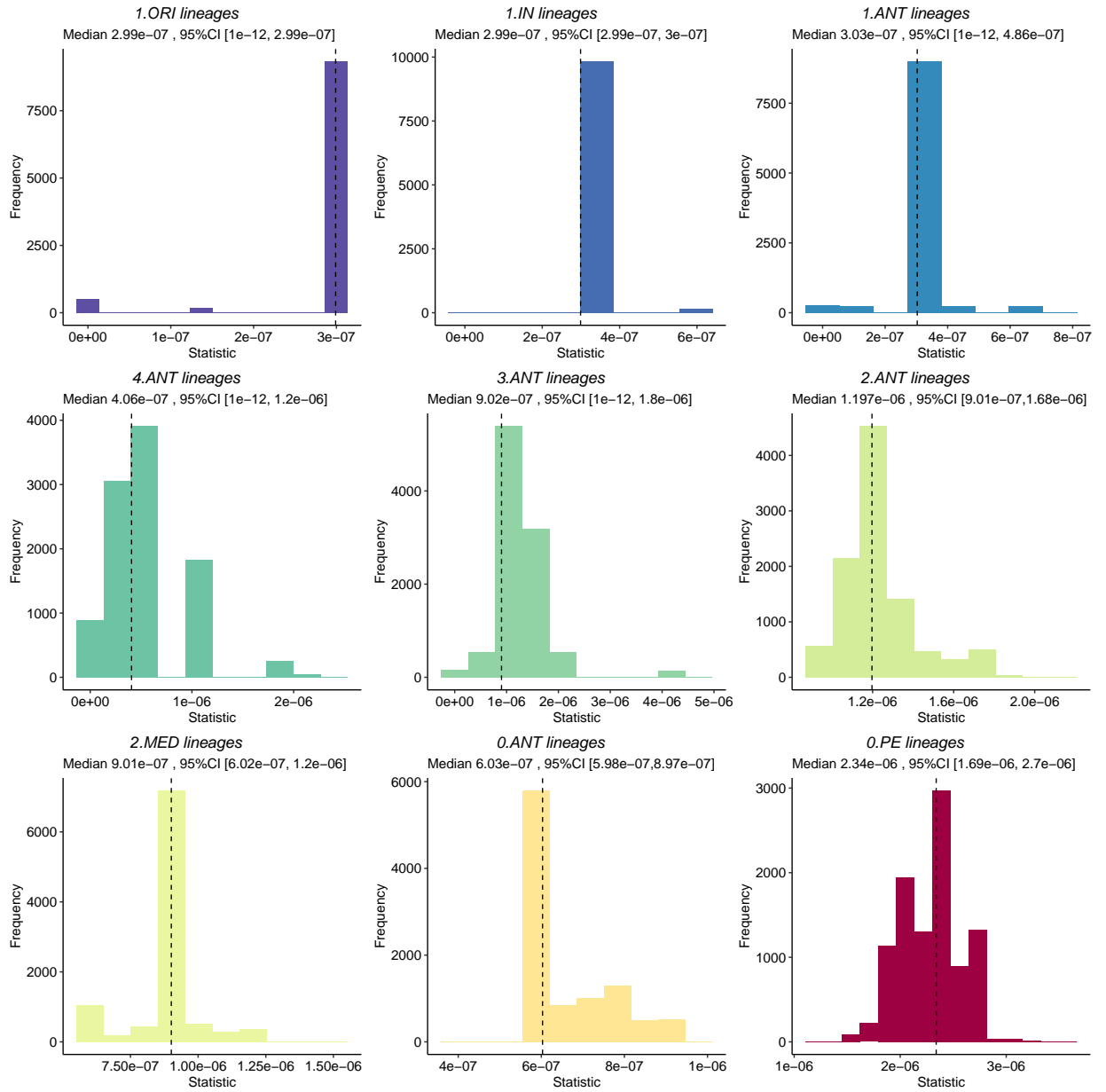
Supplementary Figure 6. Phylogeny and geographic distribution of the 1.ANT and 2.MED lineages. Up: Focus on the subtree corresponding to the 1.ANT lineage, from the focus phylogeny (see **Methods**) shown in figure 2. **Down:** Focus on the subtree of the complete 2.MED1 sublineage from the focus phylogeny. Tips are colored according to the country correspondence of the isolates.



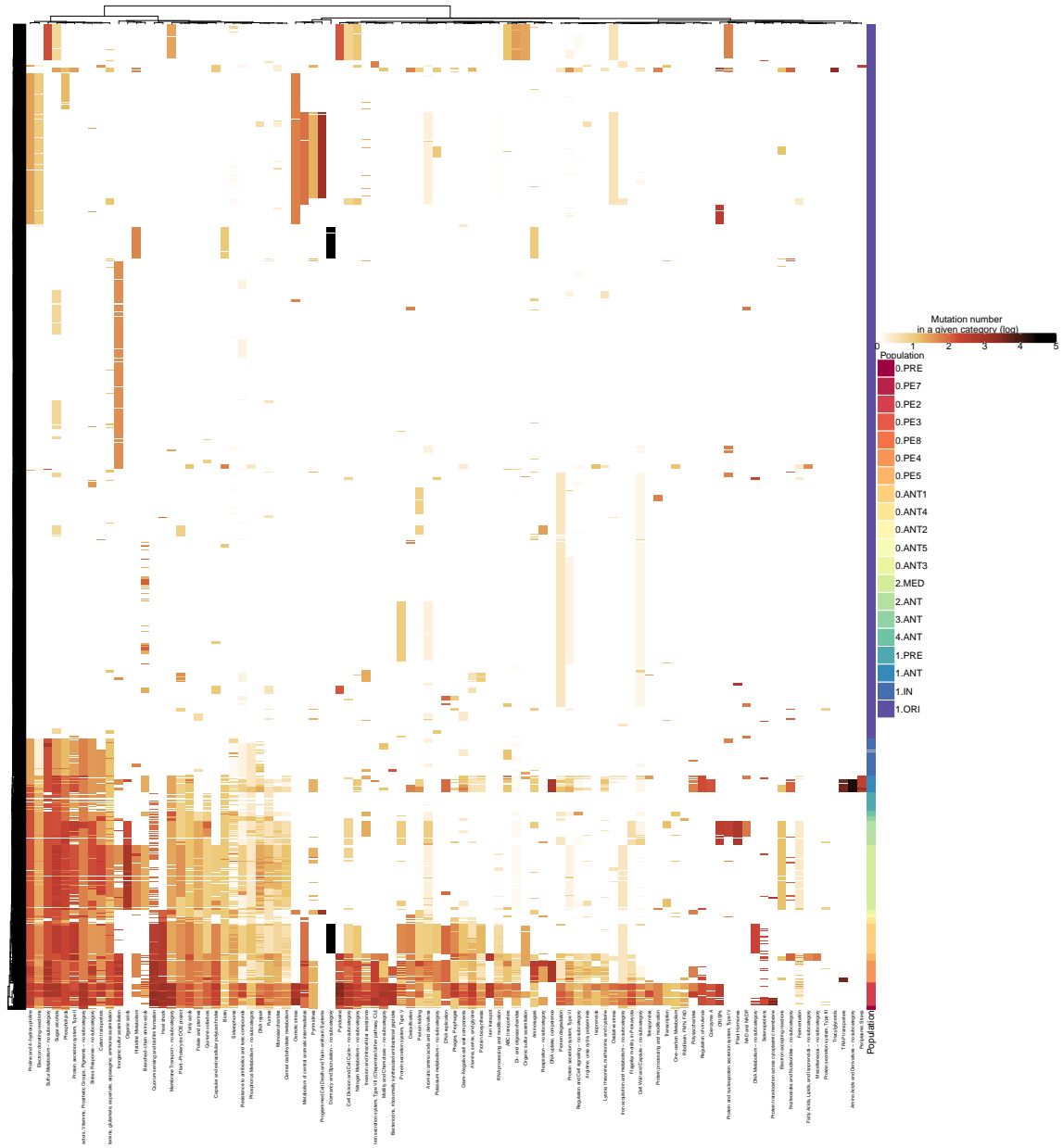
Supplementary Figure 7. Maximum Clade Credibility (MCC) tree from the molecular dating inference using BEAST. The dating dataset includes 198 modern and ancient *Y. pestis* genomes, and *Y. pseudotuberculosis* IP32953 to root the tree (not shown to facilitate visualization of the divergence dates within *Y. pestis*). The tree is scaled to inferred median divergence dates, from the most recently sampled isolates (2018) to the MRCA of all *Y. pestis* (median: 5278 years before present (yBP), 95: 4934-5816 yBP). Divergence dates for selected lineages from this tree are shown in figure 2A. The different lineages included in the phylogeny are indicated at the right panel.



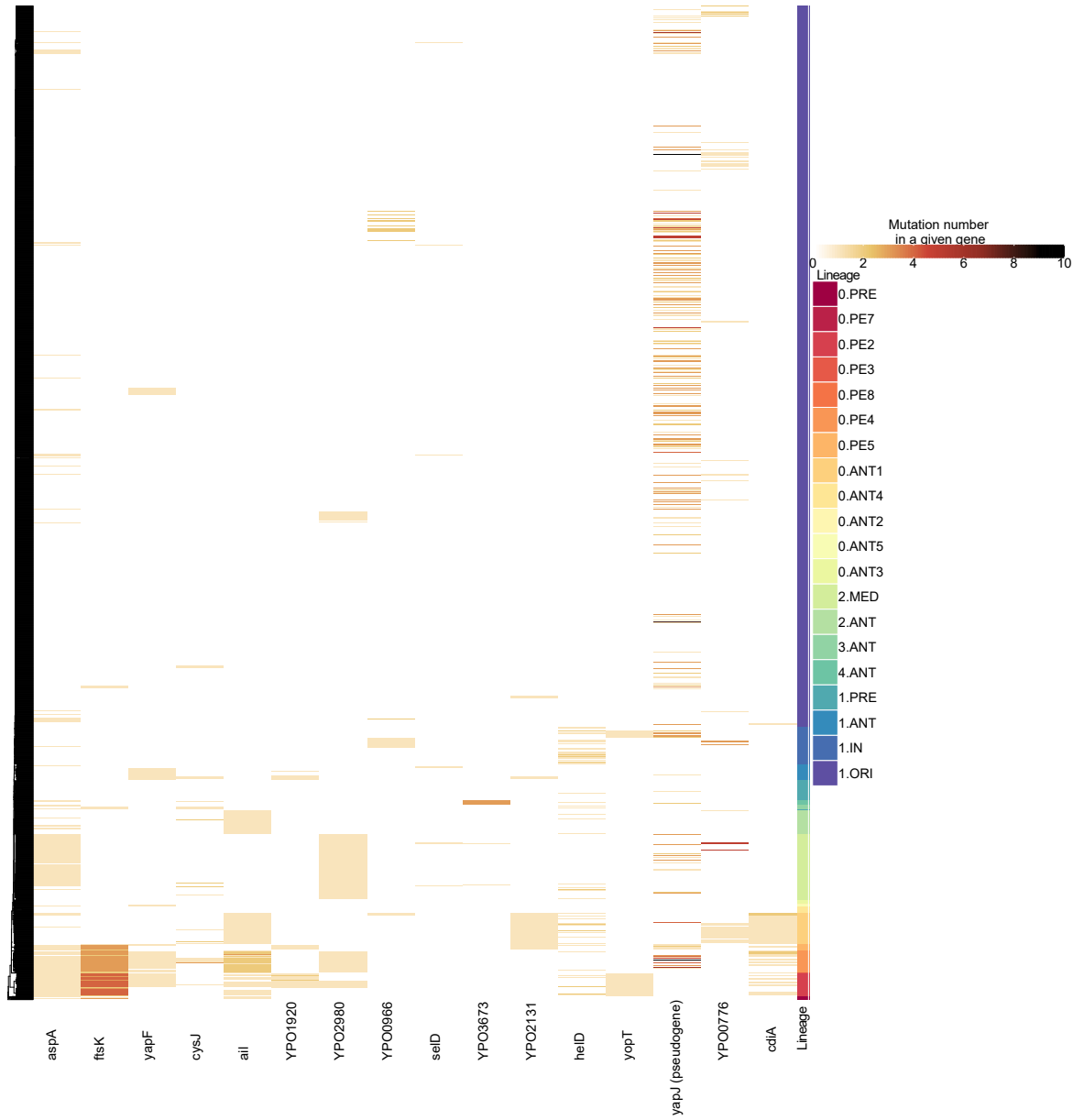
Supplementary Figure 8. Distributions of terminal branch lengths for groups of *Y. pestis* lineages (up) and major phylogenetic branches (down). Pairwise two-sided Wilcoxon rank-sum tests with Benjamini-Hochberg correction were performed to test the significance of comparisons. Significant comparisons: 0.ANT vs 0.PE (p-value = $4.65e-10$), 0.ANT vs 1.ANT (p-value = 0.042), 0.ANT vs 1.IN (p-value = 0.000261), 0.ANT vs 1.ORI (p-value = $2.19e-13$), 0.ANT vs 2.ANT (p-value = $2.31e-05$), 0.ANT vs 2.MED (p-value = 0.014), 0.PE vs 1.ANT (p-value = $6.72e-08$), 0.PE vs 1.IN (p-value = $4.16e-14$), 0.PE vs 1.ORI (p-value = $1.87e-41$), 0.PE vs 2.ANT (p-value = 0.035), 0.PE vs 2.MED (p-value = $9.54e-06$), 0.PE vs 4.ANT (p-value = 0.005), 1.ANT vs 1.ORI (p-value = 0.042), 1.ANT vs 2.ANT (p-value = $1.18e-05$), 1.ANT vs 2.MED (p-value = 0.001), 1.ANT vs 3.ANT (p-value = 0.042), 1.IN vs 1.ORI (p-value = 0.042), 1.IN vs 2.ANT (p-value = $8.02e-10$), 1.IN vs 2.MED (p-value = $3.99e-08$), 1.IN vs 3.ANT (p-value = 0.027), 1.ORI vs 2.ANT (p-value = $3.23e-21$), 1.ORI vs 2.MED, (p-value = $6.97e-32$), 1.ORI vs 3.ANT (p-value = 0.002), 2.ANT vs 4.ANT (p-value = 0.042). Comparisons below significance thresholds: 0.ANT vs 3.ANT (p-value = 0.169), 0.ANT vs 4.ANT (p-value = 0.696), 0.PE vs 3.ANT (p-value = 0.222), 1.ANT vs 1.IN (p-value = 0.659), 1.IN vs 4.ANT (p-value = 0.366), 1.ORI vs 4.ANT (p-value = 0.084), 2.ANT vs 2.MED (p-value = 0.073), 2.ANT vs 3.ANT (p-value = 0.678), 2.MED vs 4.ANT (p-value = 0.201), 3.ANT vs 4.ANT (p-value = 0.305). For major phylogenetic branches (0-4), significant comparisons are: branch 1 vs branch 2 (p-value = $6.2e-48$), branch 1 vs branch 3 (p-value = 0.004), branch 1 vs branch 0 (p-value = $2.2e-48$). Comparisons below significance threshold are: branch 1 vs branch 4 (p-value = 0.14), branch 0 vs branch 2 (p-value = 0.332), branch 0 vs branch 3 (p-value = 0.884), branch 0 vs branch 4 (p-value = 0.111), branch 2 vs branch 3 (p-value = 0.884), branch 2 vs branch 4 (p-value = 0.151), branch 3 vs branch 4 (p-value = 0.332).



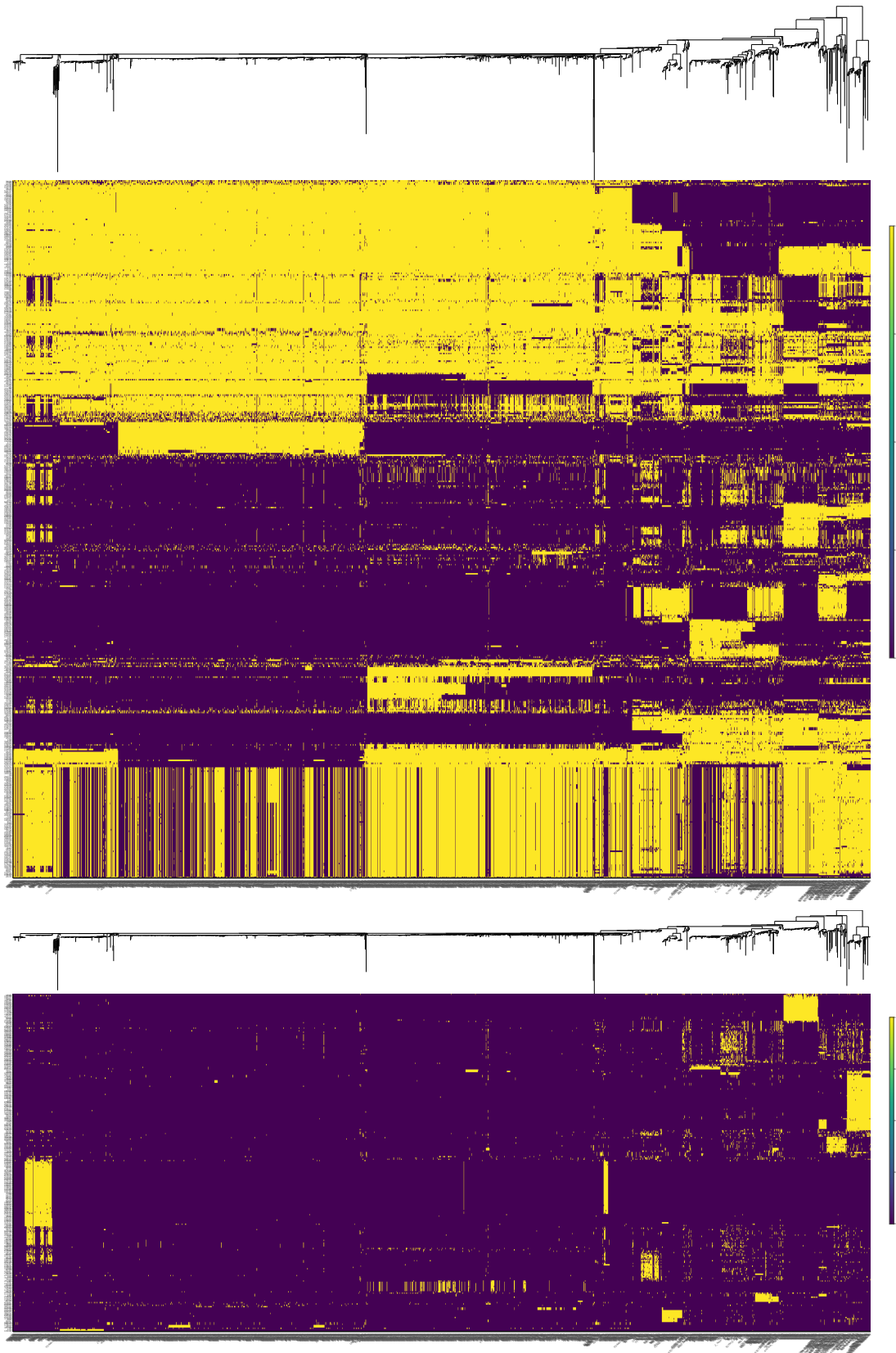
Supplementary Figure 9. Nonparametrical bootstrapping computed on the terminal branch lengths from the global *Y. pestis* phylogeny. To avoid sampling bias, 200 random terminal branch lengths (TBL) of 1.ORI genomes were used as representatives for the lineage, while all genomes for lineages other than 1.ORI were used. Non-parametric bootstrapping was performed on 1000 permutations. The median values and confidence intervals for the TBLs within each group of lineages.



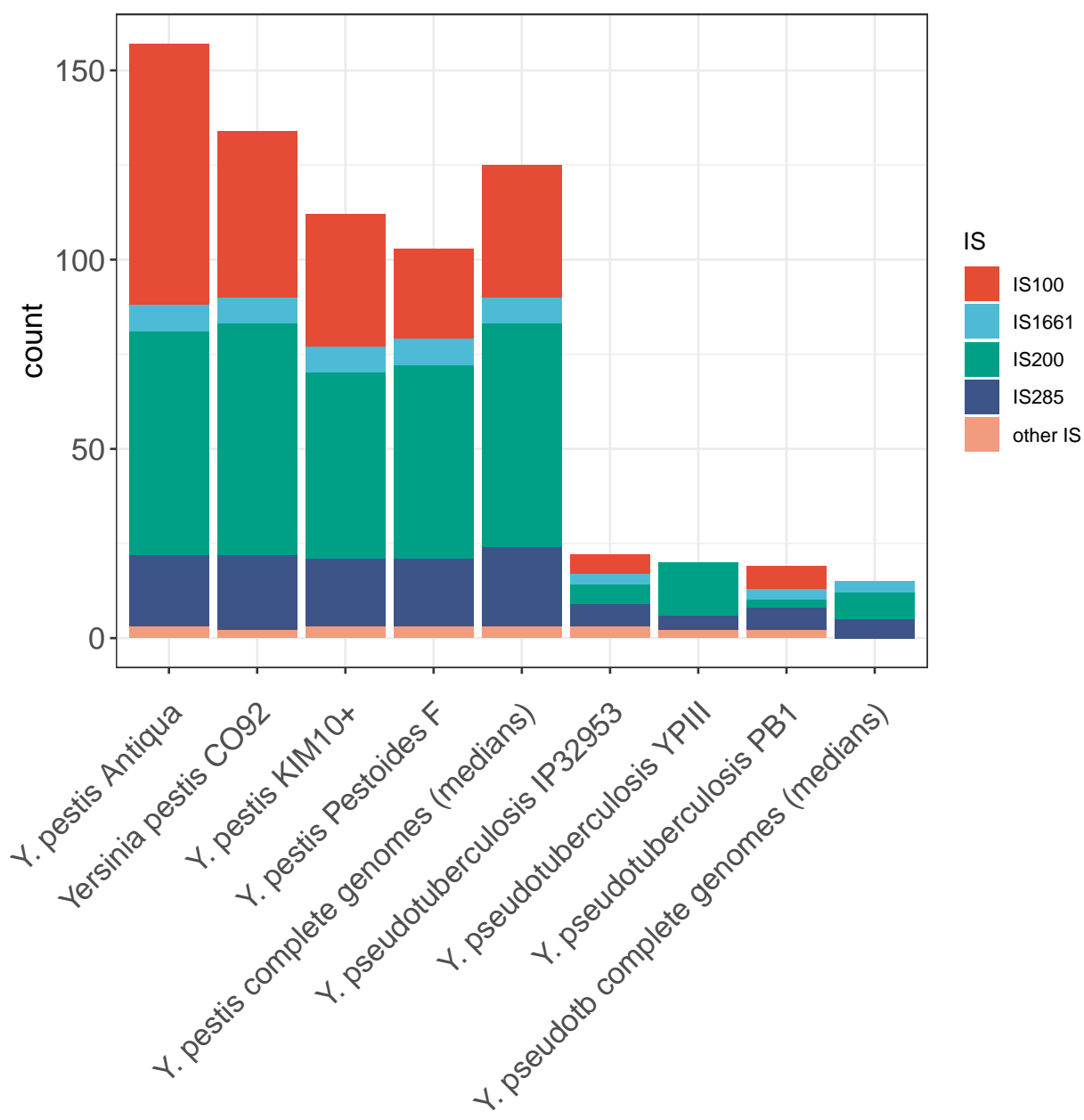
Supplementary Figure 10. Nonsynonymous mutation counts between functional categories in 2806 *Y. pestis* genomes. The heatmap includes the 1615 protein-altering (PA) mutations identified within 2806 ancient and modern *Y. pestis* genomes. These are variants selected after removing singletons and variants shared in >90% of genomes. PA counts in each subcategory were normalized to the total length of gene sequences present within each subcategory (see Methods).



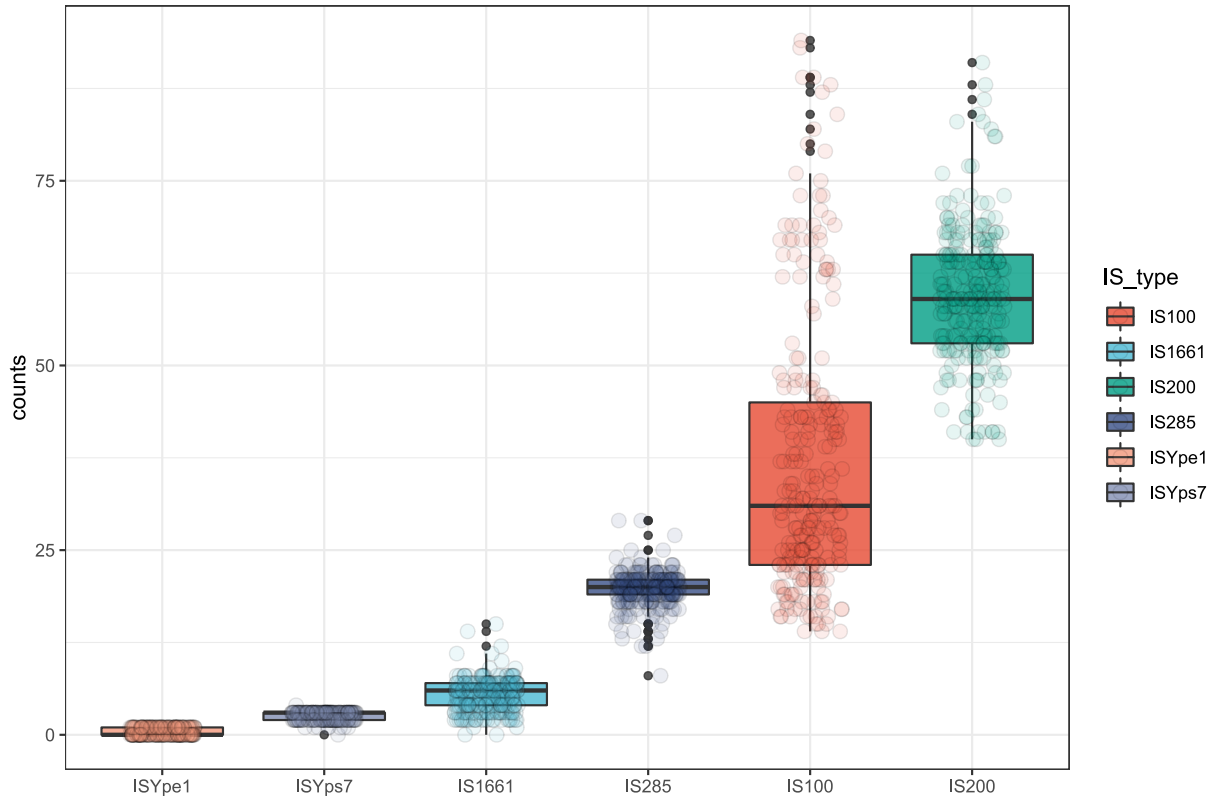
Supplementary Figure 11. Distribution of protein-altering mutations for 16 highly variable genes. The heatmap displays the normalized protein-altering (PA) mutation counts in 2806 genomes for 16 highly variable genes identified in figure 4. Rows in the heatmap are ordered following the phylogenetic tree shown in figure 1.



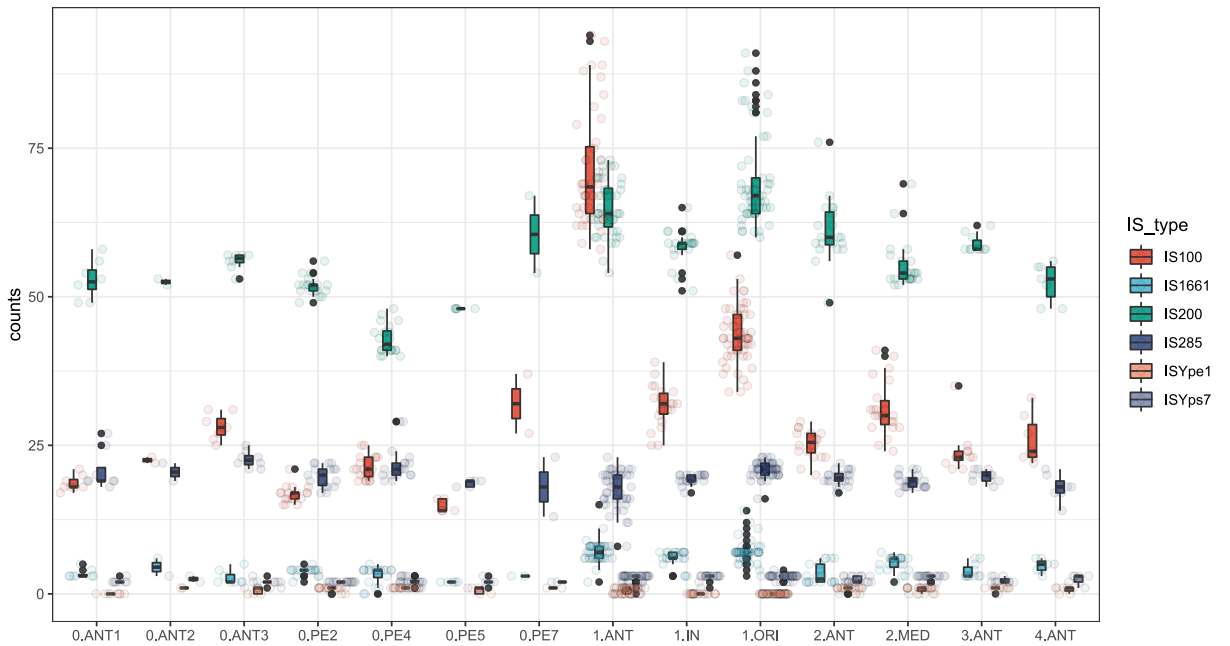
Supplementary Figure 12. Differences in presence and absence of genes between *Y. pestis* genomes. Heatmaps based on presence and absence of genes from the *Y. pestis* pangenome analysis of 1939 modern genomes. Each column represents the identified gene content of a *Y. pestis* genome, and columns follow the hierarchy defined by the phylogenetic tree shown in figure 1. **Upper plot:** gene families shared in at least 100 genomes. **Lower plot:** gene families shared between 50 and 100 genomes. These different conservation cut-offs allow to highlight groups of gene families that are either exclusive of particular lineages, shared between some of them, or genes that were lost during the evolutionary path (from right to left).



Supplementary Figure 13. Counts of 6 IS types within *Y. pestis* and *Y. pseudotuberculosis* closed genomes. Barplots showing the total number of IS in selected close genomes available on the PATRIC database, as well as the median values within the populations of the two species (table S3).



Supplementary Figure 14. Prevalence of the 6 IS types commonly found in *Y. pestis* genomes. Boxplots showing the distribution of number of IS in the dataset of 245 *Y. pestis* genomes, identified using ISMapper on short-read data.



Supplementary Figure 15. Comparison of the IS burden in *Y. pestis* genomes. Boxplots showing the distribution of counts for the 6 IS types in each modern lineage included in the analyzed dataset of 245 genomes in order to identify the IS types that have expanded the most.

Section 3.2

**Manuscript B - Outbreak dynamics and spread of
Yersinia pestis in an endemic country using
phylodynamic and phylogeographic approaches**

Next will be presented the manuscript detailing the results of the second part of this thesis, which explores the dynamics of spread of *Y. pestis* within an endemic country. This investigation is currently ongoing, and an updated version of the manuscript (including further results) is intended to be submitted for publication in the following months.

Outbreak dynamics and spread of *Yersinia pestis* in an endemic country using genomic and phylogeographic approaches

Guillem Mas Fiol^{1,✉}, Simon Dellicour^{2,3}, Mathias Vandebogaert^{4,5}, Charlotte Balière^{4,5}, Aurélia Kwasiborski^{4,5}, Valérie Caro^{4,5}, and Javier Pizarro-Cerdà^{1,6,✉}

¹Yersinia Research Unit, Institut Pasteur, Université Paris Cité, Paris, France

²Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Brussels, Belgium

³Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium

⁴Environment and Infectious Risk Unit, Institut Pasteur, Université Paris Cité, Paris, France

⁵Cellule d'Intervention Biologique d'Urgence (CIBU), Institut Pasteur, Université Paris Cité, Paris, France

⁶WHO Plague Research Collaboration Center, Institut Pasteur, Université Paris Cité, Paris, France

In the recent years, whole-genome sequencing has provided valuable insights into the epidemiology of plague in endemic countries by analyzing the distribution and genetic diversity of *Yersinia pestis* populations in relation to the animal reservoir. However, genomic data has rarely been used to explore the dynamics of transmission and spread of *Y. pestis* during outbreaks of human plague in endemic countries. Here, we apply a phylogenetic and phylogeographic approach to uncover the evolution and spread of *Yersinia pestis* in Vietnam, using hundreds of isolates recovered across four decades in the context of extensive human plague outbreaks and animal surveillance efforts. We identify a single clone that circulated during major outbreaks in the country and describe contrasting patterns of *Y. pestis* spread between the period of maximal incidence of plague, compared to the decades following the issue of the Vietnam War. Our results support the hypothesis that the rapid large-scale spread of *Y. pestis* in Vietnam was human-mediated and that it determined the long-term persistence of plague in the last decades of the past century.

Introduction

Plague is a severe zoonotic infectious disease caused by the gram-negative bacterium *Yersinia pestis*. Mostly known by its pandemic past, *Y. pestis* has spread from Central Asia in multiple occasions and it has successfully established in persistent natural foci present in Asia, Africa, and the Americas, where it is mostly maintained as a vector-borne disease in rodents (1, 2). Because of its persistence in natural animal reservoirs and its epidemic potential, plague represents a threat for public health in areas where the disease is endemic (3). This risk is illustrated by recent outbreaks of human plague in Madagascar, China and Mongolia (4–6). *Y. pestis* circulates in endemic areas by an enzootic transmission cycle primarily involving fleas and rodents (7). In such areas, zoonotic spillover to humans mainly results from the bites of infected fleas, and subsequent interhuman transmission is possible through the aerosol route (4). Plague is fatal in 40–100% of cases if proper antimicrobial treatment is not delivered in the first days of infection (8). *Y. pestis* disseminated globally during the third pandemic that started in China in the late 19th century, leading to the current world-

wide distribution of plague (1). In this study, we focus on the evolution and spread of plague in Vietnam, a country that recorded nearly 90% of worldwide plague incidence in the 1960s decade (9). The plague epidemic in Vietnam likely resulted from its successful introduction during the third pandemic. Historical records date the first outbreak in Vietnam in Nha Trang in 1898, carried by an infested ship coming from Hong Kong (10). The city of Saigon (present-day Ho Chi Minh City) was infected in 1906 and became an endemic focus during the posterior decades (11). From the beginning of the 20th century, until the 1950s, plague expanded from the south to the coastal provinces, while from 1954 to 1997, Vietnam reported cases of human plague nearly every year (12). Starting in 1962, plague manifested in areas not considered to be endemic in the previous years (10). Major outbreaks with thousands of cases were recorded between 1966 and 1972 in a context of recrudescence of the prolonged armed conflict developing in the former South Vietnam (1955–1975), which disrupted its healthcare capabilities. Living conditions created by war brought into close contact humans and an extremely abundant rodent population, which may have triggered zoonotic infections such as plague (10). After the conclusion of the armed conflict, plague cases continued to occur along the central coasts of Vietnam during the 1980s, although the incidence importantly decreased (13). In the last decades, four provinces in the Central Highlands (Tay Nguyen) became the major focus of plague and represented a source for the persistence of the disease in the country, which reported the last cases of human plague in 2003 (14). The exceptional incidence of plague in Vietnam along the second half of the 20th century received attention from local public health authorities aiming to control the disease (14). This led to the recovery of hundreds of *Y. pestis* isolates collected at various locations during outbreaks and surveillance campaigns in rodents across four decades. Despite the important plague activity in this country, the war context in which major epidemics developed limited the recovery of detailed epidemiological information that could have allowed a thorough understanding of the spread of plague in Vietnam. Here, we used genomics and spatially explicit phylogeographic

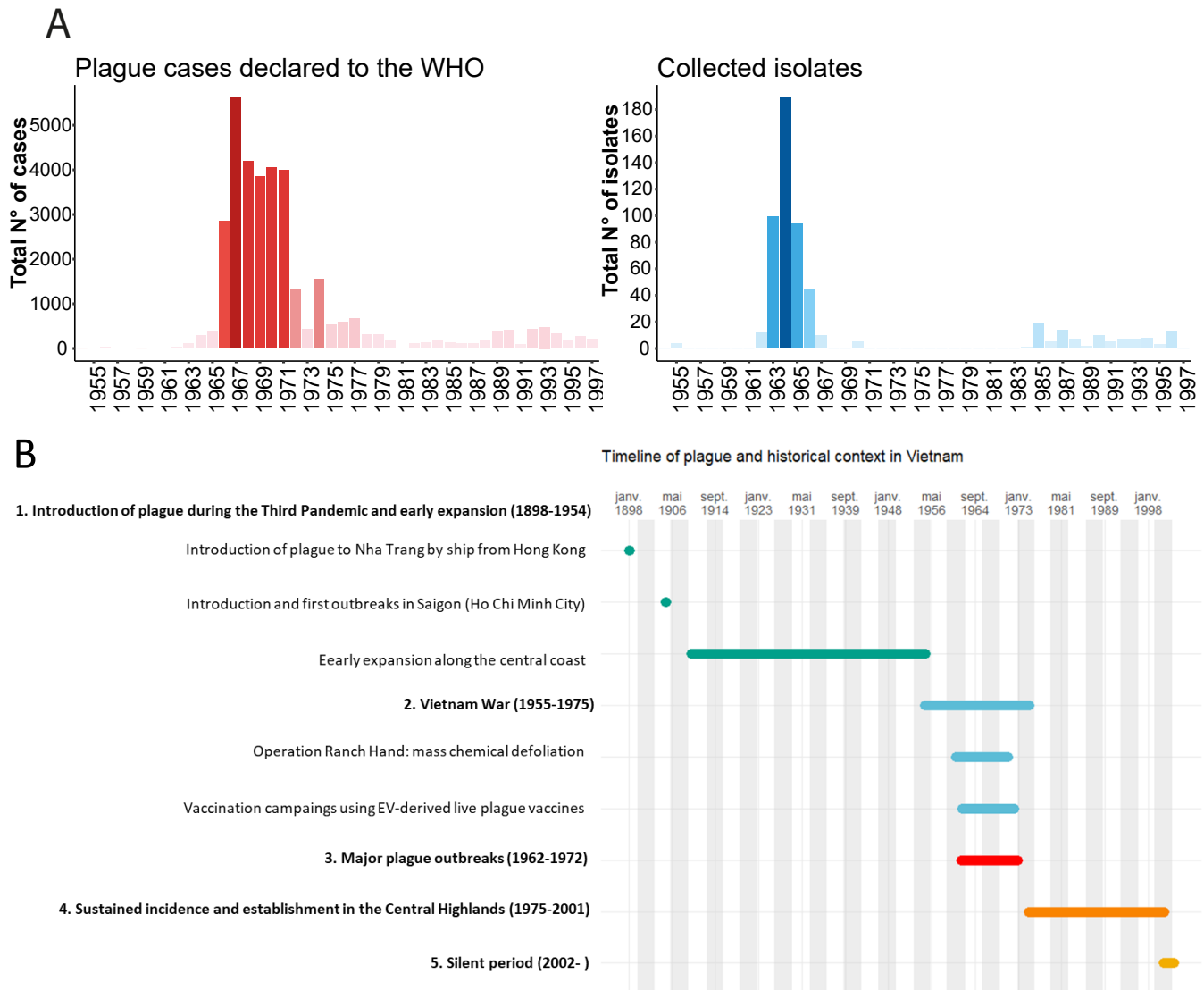


Figure 1. Epidemiological reports on human plague incidence and historical context in the second half of the 20th century Vietnam. **A:** Timeline of human plague cases declared to the WHO between 1954 and 1996 (left graph), including the period of maximum incidence (1966-1972), marked by the development of large outbreaks throughout the former South Vietnam. Timeline of total the number of *Y. pestis* isolates sequenced in this study (right graph). These isolates were collected from Vietnam and neighboring Cambodia by three different centers (Institut Pasteur of Nha Trang, Institut Pasteur of Da Lat and Institut Pasteur of Ho Chi Minh City (former IP of Saigon)) and sent to the WHO Research & Collaboration Center for Plague in the Institut Pasteur in Paris. **B:** Schematic timeline of plague and major historical events over the 20th century Vietnam. After the first recorded plague outbreaks at the turn of the 20th century, the first decades were marked by an initial expansion of plague along the central coast, eventually setting into stable foci. The War of Vietnam developed throughout two decades at the middle of the century. The airborne mass spray of chemical defoliants over large areas in Vietnam (operation Ranch Hand) was perpetuated for almost a decade (1961-72) by the US and South Vietnam air forces. The context of war overlapped with the development of major plague outbreaks and the expansion of plague into areas that did not record plague in the previous decades. After the conclusion of the conflict and reunification of Vietnam, plague activity was sustained along the central coasts of Vietnam although the incidence importantly decreased. In the last decades of the century, four provinces in the Central Highlands (Tay Nguyen) became the major focus of plague persistence in the country. By 2002, the last cases of human plague were recorded, indicating that the disease have entered a silent period. The timeline summarizes previous epidemiological descriptions (10, 11, 13, 14).

approaches to unveil the dispersal history and dynamics of *Y. pestis* in Vietnam. By analyzing the genomes of 574 of isolates, we link the epidemic in Vietnam to the global radiation of the third plague pandemic and its subsequent endemic establishment in Vietnam and other areas in South-East Asia. We find that a single clone of *Y. pestis* circulated during the outbreaks in the 1960s and further diverged in different restricted groups that were responsible for the maintenance of plague in the following decades. Using state-of-

the-art phylodynamic and continuous space phylogeographic models, we retrace the demographic dynamics and dispersal history of *Y. pestis* in Vietnam, revealing contrasting patterns of spread between the period of maximal incidence of plague and the decades following the issue of the war in Vietnam. Our study highlights the pertinence of phylodynamic and spatially explicit phylogeographic approaches to uncover the transmission of plague in an endemic country, notably when precise epidemiological data is absent or limited.

Results

Endemic circulation of plague in South-East Asia and introductions of *Y. pestis* into Vietnam.

Between 1955 and 1996, a total of 578 *Y. pestis* isolates were recovered and sent by clinical and public laboratories of three health centers in present-day Vietnam to the WHO reference laboratory for plague of the Institut Pasteur in Paris. The isolates originate from Vietnam (n=574) and neighboring Cambodia (n=4) and samples with traceable sources consist of clinical isolates of human plague (87%), but also from surveillance activities in rodents (10%) and their associated ectoparasite vectors (3%) (fig.1, dataset S1). *De novo* assemblies were generated for all the isolates from the generated short reads sequencing data. Quality control and genome statistics for the sequenced genomes is detailed in table S1. To place the recovered isolates into its broader evolutionary background and to identify the potential introductions of *Y. pestis* in Vietnam, a core-genome maximum-likelihood phylogeny was built including all the isolates together with a diverse dataset (n=142) representative of all the known *Y. pestis* lineages available in public repositories, as well as *Y. pseudotuberculosis* as the closest ancestor to *Y. pestis* to root the tree (dataset S2). The isolates fell into two separate monophyletic groups within the 1.ORI lineage, representing two independent introductions of *Y. pestis* into Vietnam (fig. 2A). This finding is consistent with previous descriptions of two distinct groups of 1.ORI having spread into Vietnam from China, where the 1.ORI lineage emerged (2). Both transmissions can be placed in the context of the global dissemination of *Y. pestis* from the 1.ORI lineage, which caused the Third Plague Pandemic starting by the late 19th century (15). The first introduction affected Cambodia and the South of Vietnam and is associated to *Y. pestis* of the 1.ORI1 clade (n=8 isolates). Plague may have entered Cambodia first, as the most ancestral isolates in the 1.ORI1 clade are from the South of Cambodia, followed by its spread to the neighboring Vietnam. This suggests an early circulation of plague around the Mekong Delta region, near the present-day border between Cambodia and Vietnam (fig. 2B). However, this 1.ORI1 clade present in Cambodia and Vietnam likely became extinct, since all isolates were collected in 1955 and no other isolates belonging to this clade have been identified past this date. The second monophyletic group identified includes the rest of Vietnamese isolates (n=570) and is part of the 1.ORI4 clade together with another group of isolates from Myanmar, as well as from the Yunnan and Guanxi provinces in China (fig. 2AB). The 1.ORI4 clade represents a second introduction of *Y. pestis* in South-East Asia that led to the endemic establishment of plague in Vietnam, as *Y. pestis* from this clade have been isolated along several decades (1962 to 1996)(dataset S1). Therefore, this endemic clade was critical for plague epidemiology in Vietnam in the second half of the 20th century (fig. 1). Altogether, these results indicate that the activity of plague in mainland South-East Asia in the past century is mainly the result of the endemic circulation of the *Y. pestis* present within each of the territories where plague established long-term foci, rather than repeated cycles of in-

roduction from external sources.

A single *Y. pestis* clone circulated in major plague outbreaks in Vietnam.

To characterize the genetic groups of *Y. pestis* that caused plague in Vietnam, we explored the relationships among the 570 *Y. pestis* genomes which are part of the South-East Asian 1.ORI4 clade. We generated a maximum-likelihood phylogeny based on variant sites identified in the core-genome of the isolates, together with 2 Chinese 1.ORI4 clade isolates that were used as outgroups to root the tree. From this tree, we identified 13 different monophyletic lineages or subclades (sc1 - sc13) of isolates with strong support in the phylogenetic tree (i.e.: >85% bootstrap supports at their common ancestor node) (fig. 3A). A tree-independent hierarchical Bayesian clustering approach was run on the same alignment to compare its population structure to the identified lineages, inferring at the subcluster hierarchy level the same groups of isolates that are monophyletic in the previous phylogeny, except for one subcluster (2) that was paraphyletic (fig. S1). All the 13 subclades contained isolates recovered from infected human patients, while isolates from 7 different lineages were recovered from Polynesian rats (*Rattus exulans*). Only sc12 was isolated from brown rats (*Rattus norvegicus*) and sc10 from black rats (*Rattus rattus*). Finally, several isolates from the sc1, sc6, sc10 and sc12 groups were collected from oriental rat fleas (*Xenopsylla cheopis*) (fig. 3A). We distinguished two main groups of lineages that are in sharp contrast in terms of genetic diversity and spatiotemporal distribution: sc1 on one side and the rest of subclades (sc2 to sc13) on the other. These latter subclades (sc2 to sc13) are each composed by less than 25 *Y. pestis* isolates that were all collected between 1984-1996 except for sc2 which was collected in 1963 (fig.3B, fig. S2). We found sc4 (n=19), sc10 (n=10), sc8 (n=12), sc9 (n=5), sc2 (n=2) and sc3 (n=2) to be mainly present in provinces of the central coast and the south of Vietnam (Khanh Hoa, Binh Dinh, Phu Yen, Ho Chi Minh City). In contrast, sc5 (n=5), sc6 (n=9), sc7 (n=7), sc11, (n=2) and sc13 (n=1) were almost exclusively collected in provinces of the Central Highlands (Lam Dong, Dak Lak, Gia Lai, Dak Nong). Interestingly, sc12 (n=21) presents a mixed pattern as it is present in six different provinces scattered along the central coast, the Central Highlands, and Ho Chi Minh City (fig. 3B, fig. S2). On the other hand, we found sc1 (n=463) to be the major *Y. pestis* group causing plague along the 1962-1970 period. Sc1 is present in all locations sampled in this period, which are distributed along a North-South axis linking the central coast (Khanh Hoa province) to the major city in the south (Ho Chi Minh City) via the central highlands (Lam Dong province) (fig. 3B, fig. S2). Moreover, sc1 lacks population structure in the phylogeny, as most of sc1 isolates (62%) are separated by <2 single nucleotide variants, compared to the rest of lineages together (38%) (fig. 3A, fig. S3). According to these observations, sc1 may represent a single clone that caused the remarkable plague outbreaks in the 1960s decade in Vietnam. Its wide distribution across the center and south of the country suggests that sc1

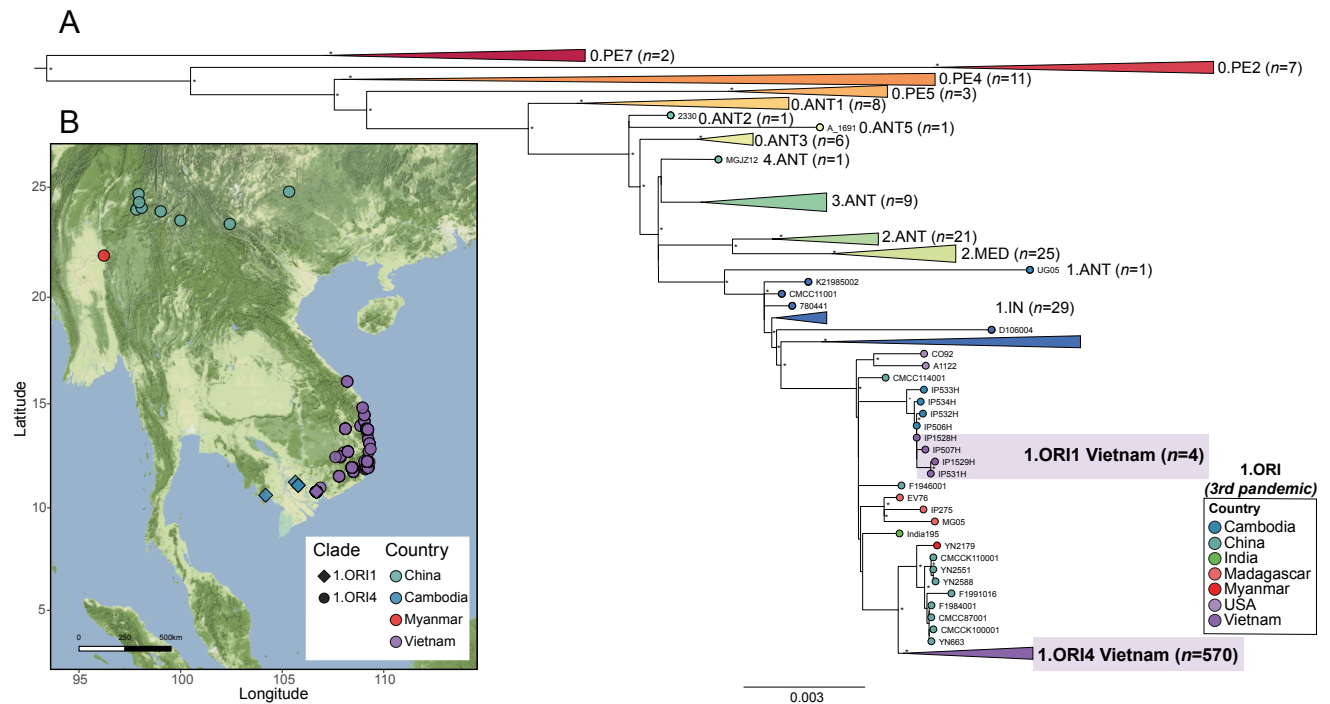


Figure 2. Endemic circulation of *Y. pestis* in South-East Asia in the context of the Third Pandemic and introduction events into Vietnam. **A:** Maximum-likelihood phylogeny computed from whole-genome variants ($n=16127$) identified within the 578 isolates from Vietnam ($n=574$) and Cambodia ($n=4$) sequenced in this study (table S1), together with 142 *Y. pestis* representative genomes for all the known lineages available in public repositories (table S2). Nodes marked with asterisks have a bootstrap support $>80\%$. The tree places the isolates collected from Vietnam into two distinct clades (1.ORI1, 1.ORI4, in purple) of the Third Pandemic 1.ORI lineage, a lineage that emerged in China and reached mainland South-East Asia in at least two occasions, but was also introduced into Madagascar, India and the United States. The strong geographical clustering within these clades highlight the endemic circulation of *Y. pestis* in the countries where they have been isolated. **B:** Sampling locations of the isolates from the two clades, each representing an independent introduction of *Y. pestis* into South-East Asia. The 1.ORI1 clade includes only 8 isolates from Vietnam and Cambodia. The 1.ORI4 clade ($n=570$ Vietnamese isolates), widespread over the center and South of Vietnam, established in long-term foci in Vietnam.

likely affected most provinces recording plague in this period (10), although no *Y. pestis* isolates from other provinces are present in the dataset to prove this hypothesis (dataset S1). The finding that sc1 dominated the global picture of plague outbreaks in the 1960s decade raises questions regarding an enhanced virulence fitness of this *Y. pestis* subclade. To investigate if the extensive spread of sc1 during outbreaks was caused by a genetic background suggestive of possible virulence or adaptive advantages, we screened the presence of known virulence determinants in sc1 assembled genomes and compared them to the rest of lineages identified in Vietnam. Most of the isolates carried the major known virulence determinants of *Y. pestis*, and no clear virulence gene composition appeared to be specific to any of the identified subclades (fig. S4). Two of the known *Y. pestis*-specific plasmids (pFra and pPla) and an integron including the Yersiniabactin siderophore (*pgm* locus) were not detected in several hundred isolates, but these replicons may be cured when isolates are stored or when they multiply *in vitro* to extract their DNA (16, 17). We complemented this analysis by generating the pangenome of the same set of isolates ($n=570$) and found no particular differences in the sizes of gene repertoires between the different identified subclades (fig. S5). The absence of obvious differences in virulence gene repertoires between *Y. pestis* isolates from Vietnam suggests that the expansion of sc1 likely resulted from other factors rather than acquiring

any particular virulence fitness during its emergence.

Evolutionary dynamics of *Y. pestis* genetic groups.

The expansion of sc1 and subsequent evolution of the rest of lineages may have been driven by demographic variations in the *Y. pestis* populations circulating in Vietnam. We employed a phylodynamic approach using the Skygrid model implemented in BEAST v1.10 (18, 19) and a relaxed molecular clock model (20) to infer the dynamics of *Y. pestis* genetic diversity through the last four decades of the past century. Before implementing time-scaled phylodynamic inference, we identified isolates (table S3, $n=16$, including the only two sc2 genomes) which deviated from the expected regression of a relaxed clock (see Materials and Methods). After excluding these outlier samples, the model allowed the estimation of a time-scaled phylogeny with 554 isolates (fig. 4A, fig. S6), which showed that the epidemic in Vietnam in the 1960s originated from the recent emergence and expansion of sc1. We estimated the date of the most recent common ancestor of the sc1 isolates to the end of the 1950s (median: 1959.97, Bayesian 95% highest posterior density (HPD) interval 1955.4–1960.28), few years before the period of maximum incidence of plague in Vietnam (1966–1972, fig. 4A). Interestingly, the dated phylogeny also inferred the parallel emergence of the different lineages (sc3 to sc13) from the sc1 that circulated preceding their emergence. All these lin-

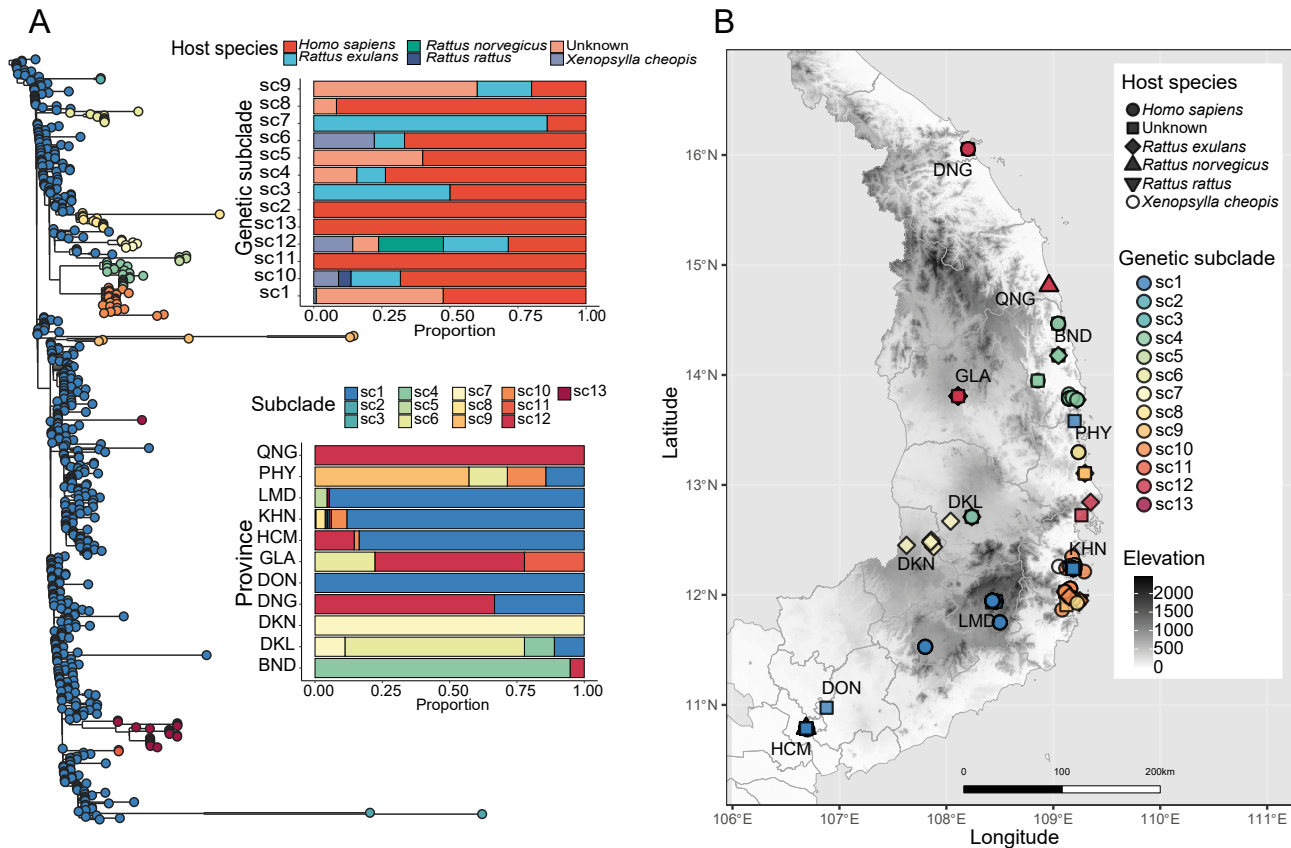


Figure 3. Population structure and distribution of *Y. pestis* genetic groups in Vietnam. **A:** Maximum-likelihood phylogeny based on the $n=570$ isolates part of the 1.ORI4 clade that established into Vietnam. Tips are colored according to the 13 identified monophyletic lineages or subclades (sc). Sc2 to sc13 emerge from sc1. The plots next to the tree show the proportion of isolates per subclade and host species where they were isolated (up) and per province (down; QNG:Quang Ngai, PHY:Phu Yen, LMD:Lam Dong, KHN:Khanh Hoa, HCM:Ho Chi Minh City, GLA:Gia Lai, DON:Dong Nai, DNG:Da Nang, DKN:Dak Nong, DKL:Dak Lak, BND:Binh Dinh). **B:** Geographic distribution of the isolates analysed in the phylogenetic tree. Points are colored according to the genetic group attributed to the isolate, and point shape corresponds to the host species from which they were isolated. The temporal and geographic distribution of the different genetic groups is shown in fig. S2.

eages emerged between 1961 and 1966, during the temporal window of diversification of sc1 (fig. 4A), which is supported by their populations measurably evolving after the isolates of sc1 due to the strong temporal signal detected in the dataset ($R^2 = 0.66$) (fig. 4B). The Skygrid coalescent model used allowed the estimation of the bacterial effective population size (N_e), which has been interpreted as an indicator of the genetic diversity of the evolving population (19, 21). The inferred *Y. pestis* effective population size resulted in estimates that globally followed the trends of human plague recorded in Vietnam (fig. 4C). The estimates show the raise in the effective population size starting in 1960 and peaks around 1964, coinciding with the inferred emergence and expansion of sc1. This increase in the estimated population sizes seems to stabilize towards the end of the 1960s decade until 1980. We note that this plateau is likely a consequence of the lack of sampling in the whole 1970s decade, since it is expected that the absence of sequences to coalesce may lead to an apparent plateauing in the estimates (19). The population size estimates between 1970 and 1980 are in consequence uncertain, as shown by the wider 95% posterior density ranges in this period (fig. 4C). Estimations of the *Y. pestis* genetic di-

versity raise again after 1980, and the peaks in the effective population size estimates remarkably mirror the distribution of recorded cases in the last decades of the 20th century until January 1996 (time of the most recent isolate). The inferred phylodynamic model suggests the emergence and rapid radial expansion of sc1 early in the 1960s, causing the outbreaks recorded in this decade. The epidemic expansion of sc1 also led to the emergence and evolution of lineages that would cause plague outbreaks in the following decades.

Contrasting patterns of *Y. pestis* spread throughout decades.

The emergence of *Y. pestis* lineages and their dissemination during epidemics illustrates the potential of plague to travel long distances and effectively establish into areas where it had never been present before (22). These dynamics resemble the ones observed in viral zoonotic diseases such as Ebola (23, 24) and West Nile (25, 26). However, the dynamics beyond these spreading events remain poorly understood within plague endemic countries. In the last part of this study, we used a spatially explicit phylogeographic approach to model the *Y. pestis* spread in Vietnam throughout decades. These

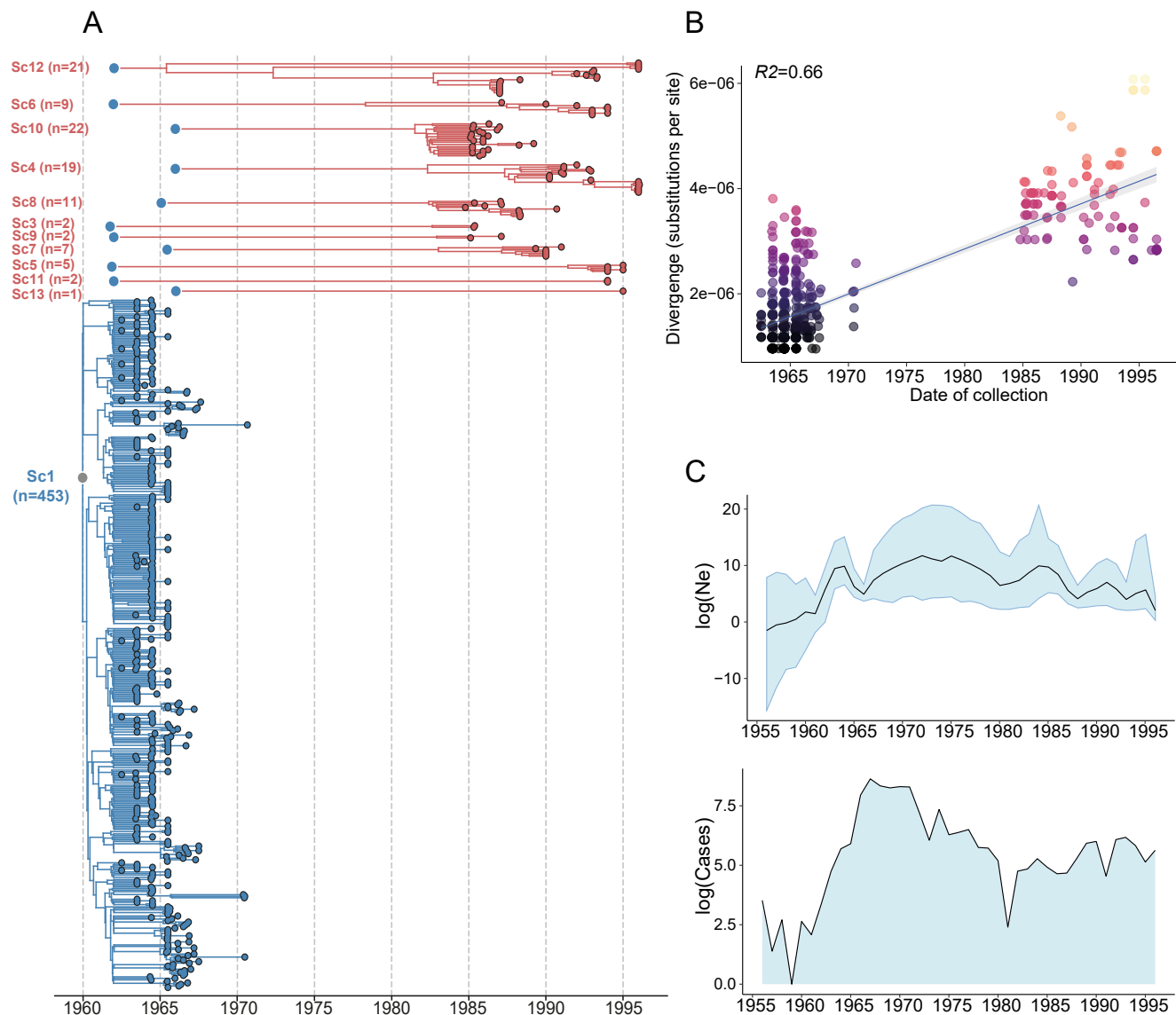


Figure 4. Phylodynamics of *Y. pestis* in Vietnam. **A:** Maximum clade credibility (MCC) tree of the Skygrid analysis inferred using BEAST (18), including 554 genomes after excluding outlier sequences. The exploded time-scaled phylogeny shows the emergence of different monophyletic lineages (sc3 - sc13) compared to the evolution of sc1 lineage. These lineages emerged during the period of diversification of sc1 in the 1960s decade, and later diversified in the following decades. The complete MCC tree is shown in fig. S6. **B:** Root-to-tip regression of the genetic and temporal distances within the genomes included in the Skygrid and subsequent phylogeographic analyses, suggestive of a strong molecular signal in the dataset. **C:** Dynamics of bacterial effective population sizes (N_e) through time inferred using the flexible Skygrid model (up), and the temporal distribution of human plague cases declared to the WHO between 1955 and 1996.

approaches integrate sampling times and spatial coordinates of pathogen isolates to the inference of molecular clock phylogenies, allowing the reconstruction of the probable location of the root and internal nodes and hence the inference of epidemic spread using pathogen genetic data (18). We employed a relaxed random walk model that accommodates diffusion rate heterogeneity among *Y. pestis* lineages (27). The phylogeographic reconstruction was represented on a map, revealing two contrasting patterns of spread of plague between decades. First, the recurrent dissemination of *Y. pestis* along an axis that links the central coast (Khanh Hoa province) and the major city of Ho Chi Minh via the Central Highlands (Lam Dong province) since the end of the 1950s until the

early 1970s decade (fig. 5, fig. S7). This spreading axis follows one of the major transport routes in the country (fig. S8), although the statistical association between major roads and the inferred events of *Y. pestis* spread was not formally tested. On the other hand, the inferred spreading events through the last twenty years of the 20th century unveil the circulation of *Y. pestis* within provinces in the central coast and the Central Highlands, consistent with the recorded establishment of the Central Highlands as the major focus of endemic maintenance of plague in these last decades (13, 14). The dispersal events inferred through the 1980s and 1990s correspond to the lineages (sc3 to sc13) that evolved after the period of epidemic spread of plague in Vietnam, and that are more ge-

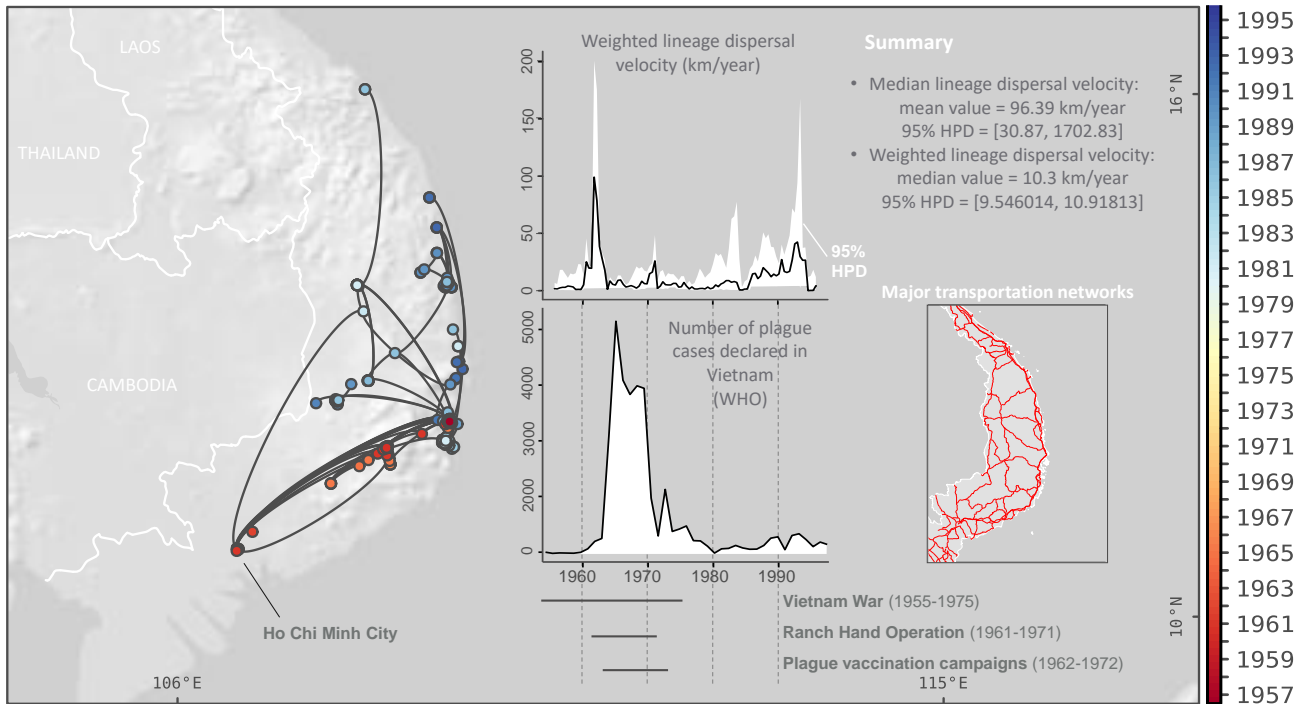


Figure 5. Spatio-temporal diffusion of *Y. pestis* in Vietnam. The Maximum clade credibility (MCC) tree represented on the map was obtained by a continuous phylogeographic inference using a relaxed random walk (RRW) model, based on 100 posterior trees (see Methods). Nodes of the tree are colored according to the inferred divergence dates, from red (the time to the most recent common ancestor, TMRCA), to blue (most recent sampling time). The locations of the 95% highest posterior density (HPD) estimates polygons of the intermediate and terminal nodes are shown in fig. S7. This figure also details the dispersal statistics inferred using the phylogeographic model and averaged over the entire bacterial spread. The evolution of the weighted dispersal velocity through time is highlighted, together the temporal distribution of plague cases in Vietnam declared to the WHO, and the timeline of relevant social-historical events in the country. The distribution of major transportation networks (train, major roads) is also displayed. A complete version of this map is shown in fig. S8. The geospatial dataset containing the displayed transport networks was obtained from the [OpenDevelopment site](#).

ographically restricted (fig. 3B). This general pattern of *Y. pestis* maintenance in geographically restricted groups is illustrated by the presence of isolates from the same lineage in the same province over several consecutive years (i.e.: sc4 in Binh Dinh province, or sc8 and sc10 in Khanh Hoa province). However, long-distance *Y. pestis* dissemination also appears to have occurred in this second time period, as illustrated by the inferred spread of sc12 across five different provinces in the central coast after its emergence in the Central Highlands (Gia Lai province).

To estimate how fast *Y. pestis* circulated across decades, we calculated several key statistics of spatial dynamics by extracting the spatio-temporal information embedded in the phylogenies from the Bayesian phylogeographic analysis (table S4). We estimated that *Y. pestis* circulated in Vietnam with a mean lineage dispersal velocity of ~ 96 km/year throughout the four decades covered by the phylogeographic inference. We also calculated the weighted lineage dispersal velocities, which represent the total distance covered by the dispersal events estimated in the trees divided by the sum of the branch duration (28) (see Materials and Methods). Weighted statistics provide with lower variance estimates that may be more useful to compare the evolution of dispersal velocities among particular time periods and datasets (29). We

analyzed how the weighted lineage dispersal velocity evolved through time and found that dispersal velocity remarkably changed across decades, presenting peaks that closely follow the epidemiological records in the country (fig. 5). The highest inferred dispersal velocity is observed around 1962, corresponding to the expansion of sc1, which diversified in the same time period as previously observed in the phylodynamic reconstruction (fig.4AC). This early peak of *Y. pestis* dispersal velocity overlaps with the beginning of airborne military operations in Vietnam (1961-1971) that resulted in the mass spraying of herbicides, including Agent Orange, by the United States and Republic of Vietnam air forces over large surfaces along the central coast and the south of the country (30, 31). It has been hypothesized that this massive spray of chemical defoliants (over seven million liters within ten years) may have disrupted the environment of rodent populations that are reservoirs of plague (9), leading to the eventual expansion of plague areas and its subsequent risk of spillover to neighboring human populations. The rapid spread of *Y. pestis* inferred in the first half of the 1960s decade also occurs in the context of important human migration from rural areas into cities escaping the war situation. Population movements and hygiene conditions in refugee communities have been pointed as possible drivers of the spread of plague in

the time of the war in Vietnam (32). We also note that this peak in *Y. pestis* dispersion immediately precedes the start of mass vaccination campaigns using live plague vaccine in the country that lasted from 1962 to 1972, although it was estimated that the immunized population (over 25 million people) only represented 43% of the total population in infected areas (10, 33). Moreover, as plague is a zoonotic infection, it is likely that immunization may have only reduced mortality rates in human infections, without affecting pathogen circulation in the animal reservoir. The increase in lineage dispersal velocities observed in the 1960s outbreaks was followed by a reduced and more sustained pattern of *Y. pestis* spread, notably in the 1980s-90s decade (fig. 5). These differential spreading dynamics represent the endemic circulation of *Y. pestis* lineages that locally evolved during previous decades (sc3 to sc13), as spreading events in these decades also seem to implicate generally smaller distances between the central coast and the Central Highlands. These results altogether indicate that *Y. pestis* rapidly spread between provinces early in the period of maximum outbreak incidence in Vietnam, followed by its endemic circulation in the following decades, which present reduced dispersal rates. The analysis of the inferred dispersal velocities also revealed non-negligible variability among lineages, as estimated by the variation of diffusion coefficients (mean CV = 21.13). This heterogeneity in dispersal velocity suggests that external factors (i.e.: environmental, human) may have influenced the spreading dynamics of *Y. pestis* in Vietnam, although the statistical association between phylogenetic branch durations (determinants of velocities) and external factors remain to be explored in future analyses.

Discussion

Plague remains a public health concern in endemic countries due to its capacity to reemerge and rapidly spread into new areas during outbreaks (12). The case of Vietnam in the second half of the last century remains a remarkable example of this risk, being one of the countries reporting the highest incidence of plague (10, 13). The occurrence of human plague cases in Vietnam had been shown to peak during the hot and dry season (February to April) (11, 34). This seasonality is thought to be due to the population dynamics of rats and their fleas that act as vectors of plague transmission, both influenced by climatic conditions that determine the availability of food (14). However, variations in these factors alone can hardly explain the exceptional occurrence of plague outbreaks recorded in the last century, which developed in the context of long-term conflict involving large areas in the center and south of the country (9). This context limited the recovery of information that may have allowed a deeper understanding of plague in Vietnam (35). Moreover, the process that led from the widespread occurrence of plague across several provinces to the eventual establishment of the Central Highlands as the major areas of long-term persistence of plague is not fully understood (10, 13). In this study, we generated genomic data of 574 isolates from Vietnam recovered over 40 years from 11 provinces that recorded a high

occurrence of plague outbreaks, and use this data to infer the dynamics of plague in the country.

We first linked the plague epidemic in Vietnam to the global expansion of the third plague pandemic, which resulted in the endemic establishment of *Y. pestis* in Vietnam and other areas of South-East Asia (fig. 2). By integrating genomic data into spatially explicit phylogeographic and phylodynamic inference, we retrace the dispersal history and the dynamics of plague outbreaks in Vietnam. Through the investigation of the genetic relationships of *Y. pestis* lineages and the comparative analysis of their dispersal statistics, we highlight two contrasting trends in the spread of *Y. pestis* in this country. We demonstrate the wide scale spread of *Y. pestis* in plague outbreaks for almost a decade (1962-1970), followed by a second period presenting a reduced and sustained spread that established the endemic persistence of *Y. pestis* in the last decades of the 20th century. The first period is characterized by a raise in the bacterial effective population in the first half of the 1960s decade, followed by a second period which is associated with a more steady and cyclic variation of effective population size (fig. 4C). We provide evidence for the emergence and large-scale dissemination of a clone (sc1) in this first period. This clone was probably responsible of the plague outbreaks in the 1960s decade, which had a markedly high incidence representing up to the 90% of worldwide incidence at the time (9). In few years, sc1 rapidly spread following a latitudinal gradient between the central coast, the Central Highlands, and the main urban area of Ho Chi Minh City (fig. 5). Since we could not find evidence indicating that sc1 had a particular virulence background suggestive of a selective advantage, its emergence and broad dissemination likely resulted from sc1 encountering the driving factors and conditions permitting the expansion of its geographic range. The inferred extensive *Y. pestis* spread developed in the context of prolonged conflict and foreign military intervention in Vietnam. It has been suggested that the transportation of food and population displacements from rural areas towards cities trying to escape the war may have driven the spread of plague in the 1960s Vietnam (10, 34, 36, 37). The phylogeographic reconstruction revealed the spread of *Y. pestis* in the 1960s followed some of the main transportation networks in the south of Vietnam (fig. 5, fig. S8), consistent with a possible role of human-mediated transport in the spread of plague in this decade. This scenario is also supported by the presence of the same clone (sc1) in diverse landscapes across multiple consecutive years: from the central coast (Nha Trang town) to the higher altitudes in the Lam Dong province until the urban areas in Ho Chi Minh city. While transportation routes have been shown to positively associate with the spreading velocity of plague in China (38), their association to the inferred spreading events and lineage dispersal velocities in Vietnam remains to be formally tested. The recovery of historical data accounting for population movements in Vietnam may also contribute to clarify this question. On the other hand, the large-scale use of herbicides and bombing campaigns during the war have also been pointed as possible drivers of

the occurrence of plague in Vietnam by disturbing the environment of rodent communities that are reservoirs of plague (9, 39). Mass spraying of chemical defoliant agents deeply altered the forest and crops in affected areas (30, 31), and agricultural crops influence food availability and hence the dynamics of rodent populations that are reservoirs of *Y. pestis* (40). If the disruption of the ecological cycle of plague in rodent populations possibly increased the risk of spillover of *Y. pestis* to human communities nearby, the possible impact of mass defoliation on rodent dynamics appears unlikely to explain the rapid dissemination of *Y. pestis* across hundreds of kilometers estimated by our phylogeographic reconstruction (~103 km/year at the turn of 1962). Considering the limited geographic extent of the ecology and transmission cycle of plague in rodents (41), these rapid and long-distance dispersal events are likely human-mediated.

Another relevant finding is that the expansion and spread of sc1 settled the basis for the parallel emergence of different *Y. pestis* genetic groups that would be responsible for its maintenance and circulation in the 1980-90s decades. The emergence of these lineages (sc3-sc13) overlaps with the inferred period of expansion of sc1 in the first half of the 1960s decade (fig. 4A), suggesting that, during its spread into different areas, sc1 found a favorable niche for its establishment and further diversification. The presence and abundance of susceptible hosts is considered a factor that likely determine the successful long-term establishment of plague after its spread into a new location (41, 42). The rodent species most frequently found infected by *Y. pestis* in this period were rats (*R. norvegicus*, *R. rattus*, *R. exulans*) and the house shrew *Suncus murinus*, and *X. cheopis* fleas were considered the main vector of plague in Vietnam (34). Such species of peridomestic rats and fleas were found in large amounts in the 1960s inhabiting close to urban areas (43) and are thought to be responsible for the long-term maintenance of plague in Vietnam (13). The observed maintenance of *Y. pestis* in genetic groups that are more geographically restricted in the 1980s-90s decade suggests a switch towards sustained cycles of local transmission that may have driven the dynamics of plague until the end of the last century. This latter pattern of endemic maintenance and diversification of *Y. pestis* has been better studied in Madagascar, which harbors some of the most active endemic areas of plague (4). In Madagascar, different *Y. pestis* subpopulations have been found to persist in the same geographic locations over decades, and the sizes of such populations appear to fluctuate between expansion and contraction events (44, 45). We observe a similar pattern in the evolution of *Y. pestis* lineages in the last decades of the 20th century in Vietnam, suggesting that related mechanisms may have guided the endemic circulation of plague in this country. If several ecological components influencing the dynamics of *Y. pestis* populations established in Madagascar have been identified (i.e.: elevation, temperature fluctuations, crop harvests)(40, 46), the local determinants of *Y. pestis* maintenance in Vietnam are less clear, as the disease seems to have entered in a period of silence without human cases since 2002 (14). Therefore, the maintenance of surveil-

lance activities in endemic areas will be critical to assess the risk of plague reemergence in Vietnam.

Our study also points out the versatility of spatially explicit phylogeographic and phylodynamic reconstructions for the understanding of the dynamics of plague in an endemic country. These analytical approaches allowed to uncover the extension of rapid and long-distance spread of *Y. pestis* during its largest outbreaks in Vietnam, but also the subsequent local establishment and evolution of *Y. pestis* in more restricted groups. The inferred weighted dispersal velocity (average ~ 10 km/year) of *Y. pestis* lineages in Vietnam over four decades is comparable to the velocities inferred for other zoonotic diseases using similar analytical frameworks, such as animal Rabies (9-35 km/year) (47–50). The highest inferred weighted lineage dispersal velocity for *Y. pestis* in Vietnam at a particular time point is of ~103 km/year (at the end of 1961), which is proportionate to the inferred velocities of the invasion phase of West Nile Virus in the USA (~165 km/year)(26). If the application of landscape phylogeographic and phylodynamic reconstructions appears particularly useful to reconstruct the epidemic history of infectious diseases, such investigations require the recovery of pathogen samples with their detailed temporal and geographic coordinates information. The collection of isolates needs to cover a sufficiently long time frame in order to capture enough temporal signal, a concept also referred as “phylodynamic threshold” (51, 52). This may hold particularly challenging for *Y. pestis*, which is a slowly-evolving pathogen with remarkable clock variation at the species level (53, 54). However, we provide evidence for a strong molecular clock signal for *Y. pestis* evolving for over forty years in Vietnam, suggesting that it may be necessary to collect isolates over several decades from the same *Y. pestis* populations in order to provide enough signal to reliably calibrate the molecular clock. Further, the extension of the conclusions drawn from spatially explicit phylogeographic inference also rely on the sampling distribution of the isolates (55). Although the great number of isolates collected in the 1960s outbreaks is probably a good indicator of the prevalence of sc1 lineage that we find widespread in this period, the recovery of isolates of this group from only six locations available in our dataset suggests they may not represent the true extent of its spread during the decade of maximum incidence. It would be thus useful to perform analyses based on larger and more geographically diverse datasets that cover regions under-sampled in the current study, particularly for the 1960s decade. However, the intensification of the war at the end of this decade made the recovery of isolates from many areas a particularly dangerous and difficult task, which may have prevented the collection of samples from many locations. For instance, in 1972, no *Y. pestis* isolation was made to confirm the reported cases from 11 out of 30 administrative areas that recorded plague outbreaks (32). Even if the resolution of phylogeographic analyses relies on sampling, they still represent powerful tools to elucidate the dispersion of pathogen lineages available in the dataset (26). By reconstructing the dispersal and evolutionary dynamics of *Y. pestis*

in Vietnam, we shed light on the epidemic history of plague in a country that had never been explored using genomic epidemiology approaches, despite holding some of the highest incidences of human plague in the 20th century (12). Using phylogeographic inference, we place the recovered *Y. pestis* isolates in a spatio-temporal context, providing information on the linkage of infections through time and space in an endemic country for plague. The inferred dispersal events may be a valuable resource for future analyses aiming to illuminate the environmental or anthropogenic factors that influenced the spread of plague in Vietnam. Recent works have provided with statistical frameworks to move beyond historical reconstructions of outbreak spread and formally test for epidemiological hypotheses using pathogen dispersal information (26). The analysis of large collections of *Y. pestis* isolates from other endemic countries with sustained plague activity in the past may also benefit from the implementation of these frameworks to unravel the dynamics of plague in such areas. Such information may eventually be useful to inform strategies to better control and monitor plague in endemic countries.

Acknowledgments

We thank Maylis Layan for support in raster preparation for landscape phylogeography and critical comments in this work. We also thank Cyril Savin for the assistance in the implementation of different analysis. This work was supported by the French government Agence Nationale de la Recherche - ASTRID program (project ANR-18-ASTR-0004), the Pasteur International Network (Pasteur International Unit 'Plague: Evolution, Spread and Maintenance) and LabEx IBEID (ANR-10-LABX-62-IBEID). GMF received a PhD fellowship by the Institut Pasteur the Direction Générale de l'Armement, as well as from the Université Paris Cité, through the EUR G.E.N.E. Doctoral School (project ANR-17-EURE-0013).

Methods

Dataset collection.

Between 1955 and 1996, a total of 578 *Y. pestis* isolates were recovered from Vietnam and Cambodia by three centers in present-day Vietnam (the Institut Pasteur of Nha Trang, the Institut Pasteur of Ho Chi Minh City and the Institut Pasteur of Da Lat). Bacterial isolates were sent to the WHO Collaborative Reference Center and French National Reference Laboratory for Plague in the Institut Pasteur in Paris. For clinical isolates, available patient data were delinked from all the samples and any culture or derived DNA from them, and samples were analyzed anonymously by attributing a barcode (IP plus digits) to each one of the isolates. The associated data for the complete dataset of *Y. pestis* isolates (n=578) with sequencing data generated in this study is detailed in (table S1). The geographic coordinates of isolation of each isolate were recovered using the provided metadata of the samples when the most precise level of location was available (commune level, n=537 isolates). When the sampling locations were

known only at the province level (n=41 isolates), we considered centroid points of their corresponding provinces as their sampling coordinates.

Whole-genome sequencing and *de novo* assembly.

Y. pestis isolates were grown in vitro and their total DNA was isolated using PureLink Genomic DNA kit (Invitrogen). All manipulations of live *Y. pestis* isolates were performed in biosafety level 3 laboratory conditions at the Institut Pasteur. Sequencing libraries were prepared following the protocol of the Nextera DNA Flex library preparation kit (Illumina). Sequencing of the pooled libraries was performed using an Illumina MiSeq platform rendering 2x300-bp pair-end reads. *De novo* assemblies were built for all the generated read data using the fq2dna pipeline (<https://gitlab.pasteur.fr/GIPhy/fq2dna>). In short, fq2dna performs pre- and post-processing steps for treating high-throughput sequencing reads, their filtering and enhancing, *de novo* assembly and scaffold sequence accuracy assessment. We used a minimum sequencing depth of 12x was used in the pre-processing steps. Trimming and clipping of reads is performed using AlienTrimmer v.0.4.0 (56), and redundant or over-represented reads are identified and reduced using the khmer tool package v1.3 (57). Sequencing errors are finally corrected using Musket v1.1 (58). Finally, *de novo* assembly was performed using the pre-processed reads using SPAdes v3.1.2.0 (59).

Phylogenetic analysis of dataset and publicly available genomes.

Paired-end raw sequencing reads were mapped onto the chromosome of the reference genome *Yersinia pestis* strain CO92 (RefSeq ID NC_003143) (60) using the snippy suite (Seeman T., <https://github.com/tseemann/snippy>), setting a minimum depth of 10 reads covering each particular position and a minimum quality of 100 in the resulting VCF.

To place the newly sequenced isolates into the *Y. pestis* species phylogenetic context, a maximum-likelihood (ML) tree was generated based on the 578 novel isolates sequences, together with 142 genomes from modern *Y. pestis* isolates available in public repositories (table S2) that represent the global genetic (all identified major lineages in previous publications (54, 61), and *Y. pseudotuberculosis* IP32953 genome was used as outgroup to root the phylogeny. For this purpose, whole-genome alignments of the reference-based consensus sequences generated using the snippy suite were used. Putative recombinogenic regions within the alignment were identified and masked using Gubbins v1.4.10 (62). Polymorphic sites present within the recombination-free alignment were then used to build a ML phylogeny using IQ-TREE v2.0.6 (63), including ModelFinder (64) for selecting the best substitution model fitting to the dataset (GTR+F+R2 model), and 1000 ultra-fast bootstrap iterations (65). The resulting phylogeny was visualized using FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Phylogenetic analysis and population structure of the main Vietnamese clade.

To reconstruct the phylogenetic relationships and infer the population structure of the $n=570$ *Y. pestis* isolates identified as part of the 1.ORI4 clade (fig. 2A), the same described steps were followed but including only the 570 Vietnamese isolates together with two other 1.ORI4 isolates (F1955003 and F1953005) part of the same clade to root the phylogeny of *Y. pestis* isolates from Vietnam. Outgroup branch was removed from the final tree, leaving 570 isolates. The phylogeny was based on 1001 identified core-genome chromosomal variants within the isolates. Monophyletic lineages were identified from the resulting ML tree (clades presenting >85% of bootstrap support at their common node, termed “subclades” or “lineages” in the main text). To compare the population structure inferred by the ML reconstruction, we used the same alignment to inform a tree-independent hierarchical Bayesian clustering algorithm (66, 67). This analysis resulted in 11 different groups at subclusters level that followed the same monophyletic clades detected before in all cases except for hierBAPS subcluster 2 which was paraphyletic. The annotated ML phylogeny with the identified subclades and hierarchical Bayesian subclusters can be consulted in the MicroReact project: <https://microreact.org/project/ofJq3XjWUNmxn9uc14EUWC-plague-in-vietnam-1962-1996>.

Virulence gene analysis and pangenome reconstruction.

To explore the presence of virulence determinants within *Y. pestis* isolates from Vietnam, we queried all the *de novo* assemblies of genomes in this study against the gene sequences available in the Virulence Factor DataBase (<http://www.mgc.ac.cn/VFs/>, (68) using Abricate (Seeman T, GitHub, <https://github.com/tseemann/abricate>.). The coverages of the best hits were extracted and plotted on a heatmap ordered by the phylogenetic tree of the 570 *Y. pestis* isolates generated in the previous step (fig. S4). To estimate the gene families present within the newly sequenced genomes, we annotated the *de novo* assemblies using Prokka v.1.14.5 (69) and estimated the pangenome in the present isolates using Panaroo v1.2.3 (70) using the --clean-mode strict flag.

Time-scaled phylogeny and phylodynamic analysis.

Before reconstructing time-scaled phylogenies and phylodynamic inference, we explored the temporal signal in the $n=570$ genomes the major clade of *Y. pestis* in Vietnam present in our dataset. We used *treeTime* (71) to detect sequences that deviated more than 3 inter-quartile distances from the regression of the genetic and temporal distance. We found $n=16$ isolates that did not follow this criterion for a loose molecular clock and were considered as outliers. Therefore, the resulting set of $n=554$ *Y. pestis* genomes was used for subsequent inference of time-scale phylogenies. The list of genomes kept and excluded from this analysis is detailed in table S1. A maximum-likelihood phylogeny

was built for the 554 sequences and the root-to-tip regression of their sampling dates and genetic distances was consistent with the populations measurably evolving through time assuming a molecular clock ($R=0.66$) (fig. 4). We inferred time-scaled phylogenetic trees using BEAST 1.10.4 (18) and BEAGLE 3 library (72) to accelerate the inference using GPU nodes. Based on previous works inferring the time-scaled evolution of *Y. pestis*, we used a GTR+ Γ substitution model and assumed a relaxed molecular clock, with independent evolutionary rates for each branch (20). We specified the flexible skygrid model as the tree prior to model changes in population effective sample size through time (19). Precise dates (resolution to isolation day) were used when available and assumed a mid-point (June) for the isolation year when only the year of recovery of the sample was known.

We ran two independent Markov Chain Monte-Carlo (MCMC) of 2×10^6 generations each. Tracer (73) was used ensure the mixing and convergence of both runs as well as to confirm that the estimated sampling sizes (ESS) values for all estimated parameters were >200. After discarding the initial 10% of trees as burn-in, we summarized the inferred trees to generate a maximum clade credibility (MCC) tree using TreeAnnotator 1.10.4 (18). This dated tree was used to infer divergence dates of the different *Y. pestis* lineages in the dataset (referred as the “time-scaled phylodynamic inference” in the main text). The estimates of the effective bacterial population size through time were extracted from the skygrid analysis using Tracer and plotted using the R-package ggplots2 (74).

Spatially explicit phylogeographic analysis.

The continuous space phylogeographic analysis was developed using the relaxed random walk (RRW) diffusion model implemented in BEAST (v.1.10.4)(27, 75). This model allows the reconstruction of temporally and spatially referenced phylogenies while considering the variation in dispersal velocity among branches (76). Despite we launched parallel GPU nodes to speed-up the analyses, the poor MCMC mixing did not allow the recovery of enough sampling from the posterior in reasonable amount of time, an issue that has already been pointed out in other works (26). To circumvent this challenging issue, we implemented the same approach as in Dellicour et al.(26) and performed 100 independent phylogeographic analyses, each based on a different fixed tree sampled from the posterior distribution a previous skygrid analysis. This approach may provide a reasonable strategy to account for phylogenetic uncertainty, as each analysis is based on a different fixed topology. Each analysis was run until ESS values were all > 200, and the last spatially annotated tree from each independent run was extracted. All the subsequent phylogeographic analyses (dispersal history reconstruction and estimation of dispersal statistics) were based on this collection of 100 spatially annotated trees. We used TreeAnnotator 1.10.4 (18) to generate a MCC tree with the annotations of the spatially explicit phylogeographic reconstruction (displayed in fig. 5). The inferred location of internal nodes and dispersal history were plotted in R using

custom scripts and the R-package “seraphim” (28, 77). The locations of the 95% HPD polygons for the internal nodes and tips are displayed in fig. S7.

Estimation and comparison of lineage dispersal statistics.

Each phylogenetic branch (called “lineage” in the main text) in the phylogeographic reconstruction can be considered as conditionally independent movement vectors (75). We extracted the spatio-temporal information for the branches in the 100 independent spatially-annotated trees using the R-package “seraphim”, allowing to summarize the branches as movement vectors over geographic coordinates (77). The R-package “seraphim” was also used to estimate and summarize the dispersal statistics based on the collection of vectors extracted from the trees. We estimated two statistics: the main lineage dispersal velocity (v_{mean}) and the weighted lineage dispersal velocity (v_{weighted}) (77). These metrics measure the dispersal velocity associated with phylogenetic branches, and particularly the weighted statistic implicates normalizing by branch time duration (48), and are computed as follows:

$$v_{\text{mean}} = \frac{1}{n} \sum_{i=1}^n \frac{d_i}{t_i}$$

and

$$v_{\text{weighted}} = \frac{\sum_{i=1}^n d_i}{\sum_{i=1}^n t_i}$$

Where d_i and t_i represent the traveled geographic distance vectors (great-circle distance in km) and the time elapsed (in years) for each phylogeny branch, respectively. The interest of weighted metric is that it produces lower variance estimates, useful to compare branch dispersal velocities between different subsets of the same dataset. This is because phylogeny branches are with short branch duration will have a lower impact on the weighted lineage dispersal velocity compared to the mean lineage dispersal velocity (29).

Data availability.

Supplementary tables and the XML files for the phylodynamic and phylogeographic analyses can be found on a dedicated repository in GitHub: <https://github.com/guillemmasfiol/PestisPhylogeo>.

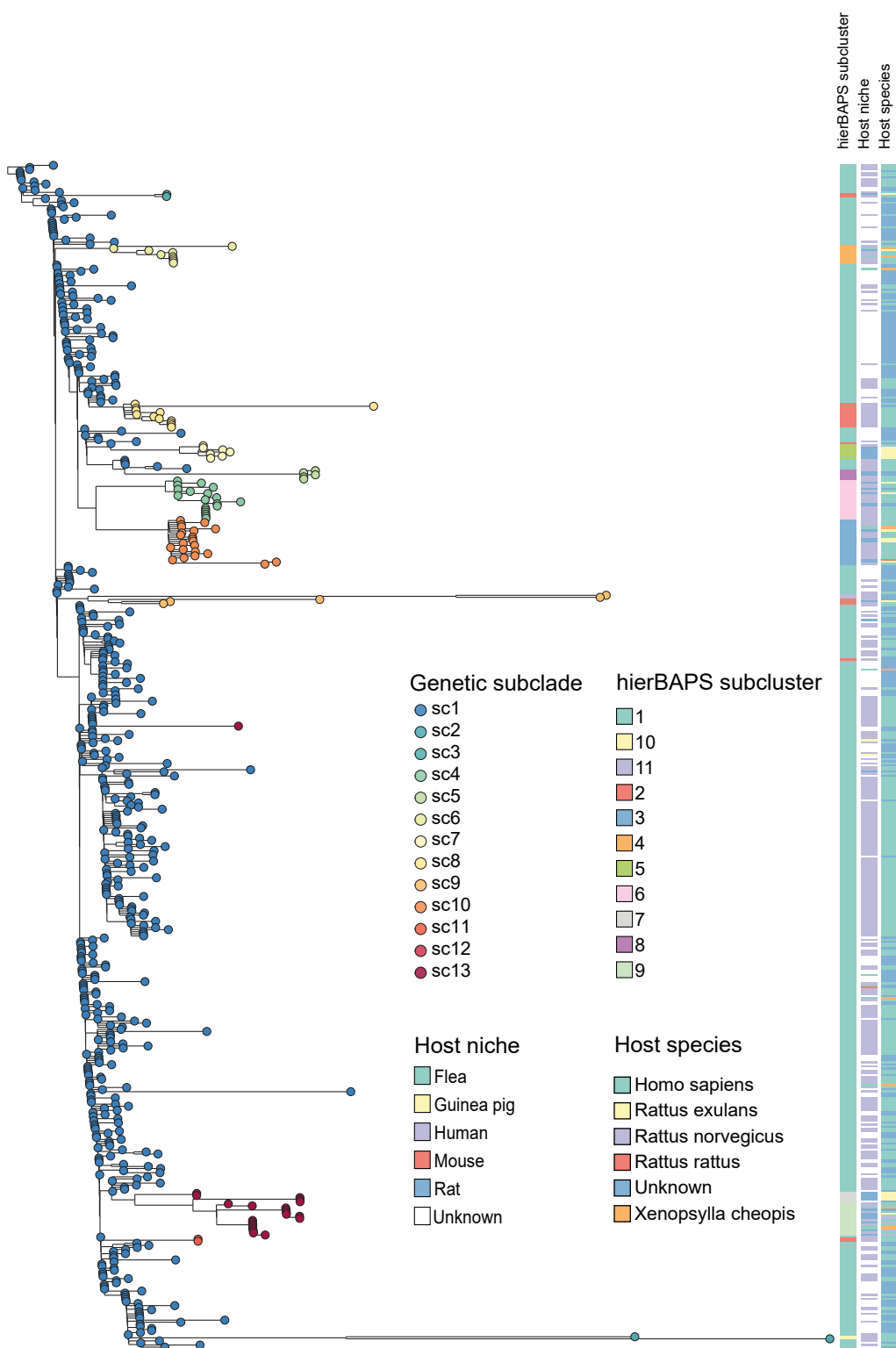
Bibliography

1. Robert D Perry and Jacqueline D Fetherston. *Yersinia pestis*—etiologic agent of plague. *Clinical microbiology reviews*, 10(1):35–66, 1997.
2. Giovanna Morelli, Yajun Song, Camila J. Mazzoni, Mark Eppinger, Philippe Roumagnac, David M. Wagner, Mirjam Feldkamp, Barica Kusecek, Amy J. Vogler, Yanjun Li, Yujun Cui, Nicholas R. Thomson, Thibaut Jombart, Raphael Leblois, Peter Lichtner, Lila Rahalison, Jeannine M. Petersen, Francois Balloux, Paul Keim, Thierry Wirth, Jacques Ravel, Ruiyu Yang, Elisabeth Carniel, and Mark Achtman. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *NATURE GENETICS*, 42(12):1140+, DEC 2010. ISSN 1061-4036. doi: 10.1038/ng.705.
3. World Health Organization = Organisation mondiale de la Santé. Plague around the world, 2010–2015 = la peste à travers le monde: 2010-2015. *Weekly Epidemiological Record = Relevé épidémiologique hebdomadaire*, 91(8):89–93, 2016-02-26.
4. Rindra Rendremanana, Voahangy Andrianaivoarimanana, Birgit Nikolay, Beza Ramasin-drzana, Juliette Paireau, Quirine Astrid ten Bosch, Jean Marius Rakotondramanga, Soloandry Rahajandraibe, Soanandrasana Rahelinirina, Fanjasoa Rakotomanana, Feno M. Rakotoarimanana, Lea Bricette Randriamampionona, Vaory Razafimbina, Mamy Jean

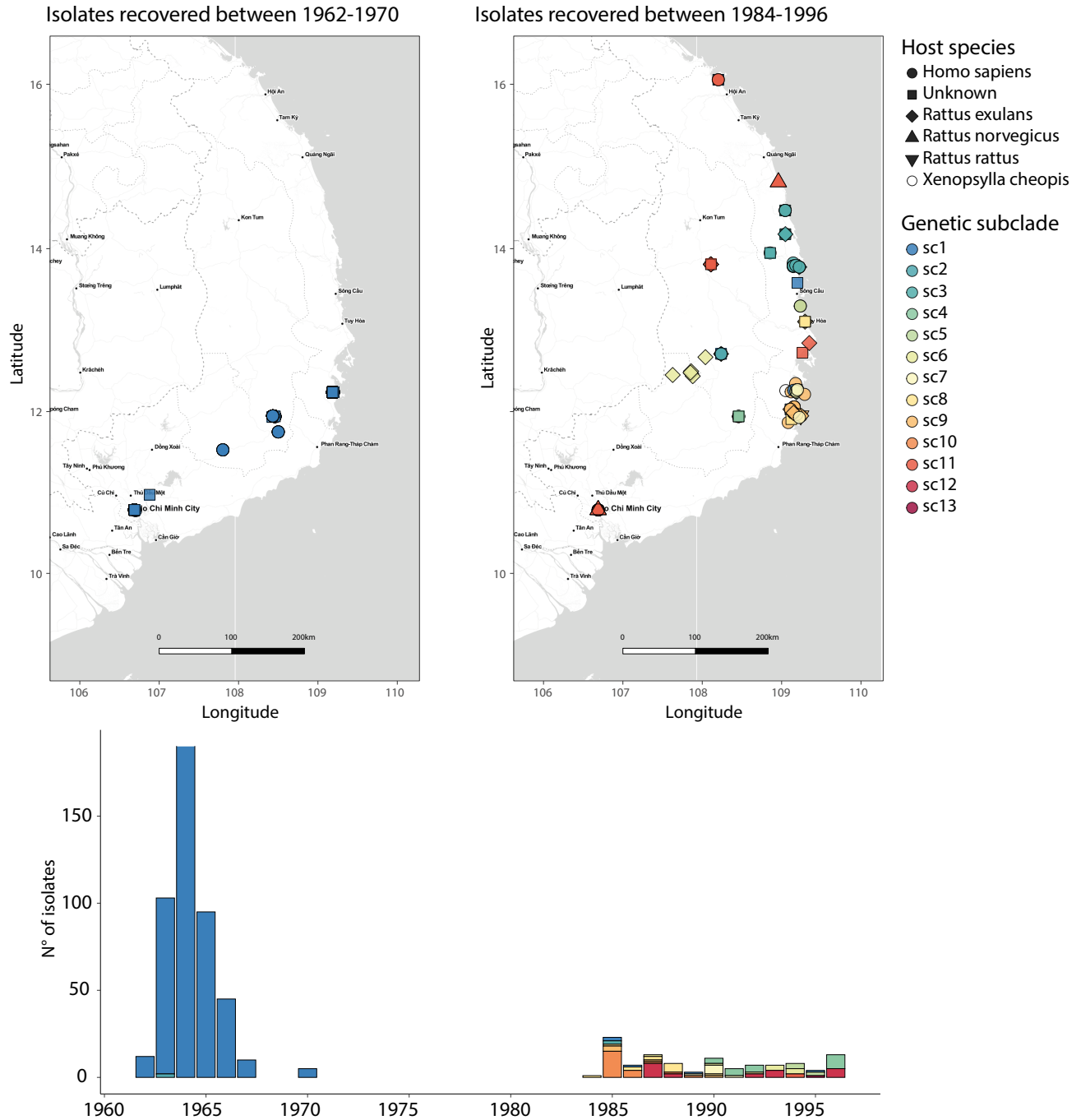
- De Dieu Randria, Mihaja Raberahona, Guillaïn Mikaty, Anne-Sophie Le Guern, Lamina Arthur Rakotonjanabelo, Charlotte Faty Ndiaye, Voahangy Rasolofo, Eric Berthart, Maherisoa Ratsitorahina, Simon Cauchemez, Laurence Baril, Andre Spiegel, and Minoarisoa Rajerison. Epidemiological characteristics of an urban plague epidemic in madagascar, august-november, 2017: an outbreak report. *LANCET INFECTIOUS DISEASES*, 19(5): 537–545, MAY 2019. ISSN 1473-3099. doi: 10.1016/S1473-3099(18)30730-8.
5. Xiaona Shen, Jianyun Li, Mengguang Fan, Lianxu Xia, and Wei Li. A reemergent case of bubonic plague-inner mongolia autonomous region, china, july, 2020. *China CDC Weekly*, 2(29):549–550, 2020.
6. Yumeng Wang, Lei Zhou, Mengguang Fan, Quanyi Wang, Jianyun Li, Qun Li, Zijian Feng, George F Gao, Cheng Xu, Lijuan Chen, et al. Isolated cases of plague—inner mongolia-beijing, 2019. *China CDC weekly*, 1(1):13, 2019.
7. Nils Chr. Stenseth, Noelle I. Samia, Hildegunn Viljugrein, Kyrre Linne Kausrud, Mike Begon, Stephen Davis, Herwig Leirs, V. M. Dubyanskiy, Jan Esper, Vladimir S. Ageyev, Nikolay L. Klassovskiy, Sergey B. Pole, and Kung-Sik Chan. Plague dynamics are driven by climate variation. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 103(35):13110–13115, AUG 29 2006. ISSN 0027-8424. doi: 10.1073/pnas.0602447103.
8. R. Pollitzer. Plague studies*. *Bulletin of the World Health Organization*, 4(4):475–533, 1951.
9. AK AKIEV. Epidemiology and incidence of plague in the world, 1958-79. *BULLETIN OF THE WORLD HEALTH ORGANIZATION*, 60(2):165–169, 1982.
10. B VELIMIROVIC. Plague in southeast asia - brief historical summary and present geographical distribution. *TRANSACTIONS OF THE ROYAL SOCIETY OF TROPICAL MEDICINE AND HYGIENE*, 66(3):479+, 1972. ISSN 0035-9203. doi: 10.1016/0035-9203(72)90280-5.
11. Dan C Cavanaugh, HG Dangerfield, DH Hunter, RJ Joy, JD Marshall Jr, DV Quy, S Vivona, and PE Winter. Some observations on the current plague outbreak in the republic of vietnam. *American Journal of Public Health and the Nations Health*, 58(4):742–752, 1968.
12. World Health Organization et al. Who report on global surveillance of epidemic-prone infectious diseases. Technical report, World Health Organization, 2000.
13. V. V. Suntsov. Ecological aspects of the origin of yersinia pestis, causative agent of the plague: Concept of intermediate environment. *CONTEMPORARY PROBLEMS OF ECOLOGY*, 7(1):1–11, FEB 2014. ISSN 1995-4255. doi: 10.1134/S1995425514010144.
14. Hau V. Pham, Dat T. Dang, Nguyen N. Tran Minh, D. Nguyen, and Tuan V. Nguyen. Correlates of environmental factors and human plague: an ecological study in vietnam. *INTERNATIONAL JOURNAL OF EPIDEMIOLOGY*, 38(6):1634–1641, DEC 2009. ISSN 0300-5771. doi: 10.1093/ije/dyp244.
15. M Achtman, K Zurth, C Morelli, G Torrea, A Guiyoule, and E Carniel. *Yersinia pestis*, the cause of plague, is a recently emerged clone of yersinia pseudotuberculosis. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 96(24):14043–14048, NOV 23 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.24.14043.
16. Carmen Buchrieser, Michael Prentice, and Elisabeth Carniel. The 102-kilobase unstable region of yersinia pestis comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *Journal of Bacteriology*, 180(9):2321–2329, 1998.
17. TC Leal-Balbino, NC Leal, CV Lopes, and AMP de Almeida. Differences in the stability of the plasmids of yersinia pestis cultures in vitro: impact on virulence. *Memórias do Instituto Oswaldo Cruz*, 99:727–732, 2004.
18. Marc A. Suchard, Philippe Lemey, Guy Baele, Daniel L. Ayres, Alexei J. Drummond, and Andrew Rambaut. Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *VIRUS EVOLUTION*, 4(1), JAN 2018. doi: 10.1093/ve/vey016.
19. Verity Hill and Guy Baele. Bayesian estimation of past population dynamics in beast 1.10 using the skygrid coalescent model. *Molecular biology and evolution*, 36(11):2620–2628, 2019.
20. Alexei J Drummond, Simon Y. W Ho, Matthew J Phillips, and Andrew Rambaut. Relaxed phylogenetics and dating with confidence. *PLOS Biology*, 4(5):null, 03 2006. doi: 10.1371/journal.pbio.0040088.
21. Edward C Holmes and Bryan T Grenfell. Discovering the phylodynamics of rna viruses. *PLoS computational biology*, 5(10):e1000505, 2009.
22. Lei Xu, Leif C. Stige, Herwig Leirs, Simon Neerinx, Kenneth L. Gage, Ruiyu Yang, Qiyong Liu, Barbara Bramanti, Katharine R. Dean, Hui Tang, Zhe Sun, Nils Chr. Stenseth, and Zhibin Zhang. Historical and genomic data reveal the influencing factors on global transmission velocity of plague during the third pandemic. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 116(24): 11833–11838, JUN 11 2019. ISSN 0027-8424. doi: 10.1073/pnas.1901366116.
23. Edward C Holmes, Gytis Dudas, Andrew Rambaut, and Kristian G Andersen. The evolution of ebola virus: Insights from the 2013–2016 epidemic. *Nature*, 538(7624):193–200, 2016.
24. Jessica R Spengler, Elizabeth D Ervin, Jonathan S Towner, Pierre E Rollin, and Stuart T Nichol. Perspectives on west africa ebola virus disease outbreak, 2013–2016. *Emerging infectious diseases*, 22(6):956, 2016.
25. RS Lanciotti, JT Roehrig, V Deubel, J Smith, M Parker, K Steele, B Crise, KE Volpe, MB Crabtree, JH Scherret, et al. Origin of the west Nile virus responsible for an outbreak of encephalitis in the northeastern united states. *Science*, 286(5448):2333–2337, 1999.
26. Simon Dellicour, Sebastian Lequime, Bram Vrancken, Mandev S Gill, Paul Bastide, Karthik Gangavarapu, Nathaniel L Matteson, Yi Tan, Louis Du Plessis, Alexander A Fisher, et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nature communications*, 11(1):1–11, 2020.
27. Philippe Lemey, Andrew Rambaut, John J. Welch, and Marc A. Suchard. Phylogeography takes a relaxed random walk in continuous space and time. *MOLECULAR BIOLOGY AND EVOLUTION*, 27(8):1877–1885, AUG 2010. ISSN 0737-4038. doi: 10.1093/molbev/msq067.
28. Simon Dellicour, Rebecca Rose, and Oliver G. Pybus. Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC BIOINFORMATICS*, 16, FEB 11 2016. ISSN 1471-2105. doi: 10.1186/s12859-016-0924-x.
29. Simon Dellicour, Rebecca Rose, Nuno Rodrigues Faria, Luiz Fernando Pereira Vieira, Herve Bourhy, Marius Gilbert, Philippe Lemey, and Oliver G. Pybus. Using viral gene se-

- quences to compare and explain the heterogeneous spatial dynamics of virus epidemics. *MOLECULAR BIOLOGY AND EVOLUTION*, 34(10):2563–2571, OCT 2017. ISSN 0737-4038. doi: 10.1093/molbev/msx176.
30. Colin Norman. Vietnam's herbicide legacy: Recent studies in vietnam have found long-term environmental damage and a possible rise in birth defects from us spraying. *Science*, 219(4589):1196–1197, 1983.
 31. Jeanne Mager Stellman, Steven D Stellman, Richard Christian, Tracy Weber, and Carrie Tomasallo. The extent and patterns of usage of agent orange and other herbicides in vietnam. *Nature*, 422(6933):681–687, 2003.
 32. B Velimerovic. Review of the global epidemiological situation of plague since the last conference of tropical medicine in 1973. *CRC handbook series in zoonoses*, 1:560–596, 1979.
 33. Nguyen Dang Que. Plague in vietnam. symposium on plague, saigon. *Unpublished*, 4(4), 1970.
 34. JD Marshall Jr, RJT Joy, NV Ai, DV Quy, JL Stockard, FL Gibson, et al. Plague in vietnam 1965-1966. *American journal of epidemiology*, 86(3):603–16, 1967.
 35. D C Cavanaugh and J D Jr Marshall. The influence of climate on the seasonal prevalence of plague in the republic of vietnam. *Journal of wildlife diseases*, 8(1):85–94, Jan 1972. ISSN 0090-3558.
 36. John S Whitehall. Plague in a time of war: an experience in south vietnam. *Medical Journal of Australia*, 191(11):671, 2009.
 37. Michael E Staub. Making sense of plague in the vietnam war. *Environmental History*, 27(2): 242–268, 2022.
 38. Lei Xu, Leif Chr. Stige, Kyrre Linne Kausrud, Tamara Ben Ari, Shuchun Wang, Xiye Fang, Boris V. Schmid, Qiyong Liu, Nils Chr. Stenseth, and Zhibin Zhang. Wet climate and transportation routes accelerate spread of human plague. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 281(1780), APR 7 2014. ISSN 0962-8452. doi: 10.1098/rspb.2013.3159.
 39. JB Neillands. Vietnam: progress of the chemical war. *Asian Survey*, 10(3):209–229, 1970.
 40. Soanandrasana Rahelinirina, Jean Marc Duplantier, Jocelyn Ratovonjato, Olga Ramilisoana, Mammy Ratsimba, and Lila Rahalison. Study on the movement of rattus rattus and evaluation of the plague dispersion in madagascar. *Vector-borne and zoonotic diseases*, 10(1):77–84, 2010.
 41. Amy J. Vogler, Voahangy Andrianavaoimanana, Sandra Telfer, Carina M. Hall, Jason W. Sahl, Crystal M. Hepp, Heather Centner, Genevieve Andersen, Dawn N. Birdsell, Lila Rahalison, Roxanne Nottingham, Paul Keim, David M. Wagner, and Minoarisoa Rajerison. Temporal phylogeography of yersinia pestis in madagascar: Insights into the long-term maintenance of plague. *PLOS NEGLECTED TROPICAL DISEASES*, 11(9), SEP 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005887.
 42. Paul S Keim and David M Wagner. Humans and evolutionary and ecological forces shaped the phylogeography of recently emerged diseases. *Nature Reviews Microbiology*, 7(11): 813–821, 2009.
 43. David T Dennis, Kenneth L Gage, Norman G Gratz, Jack D Poland, Evgueni Tikhomirov, World Health Organization, et al. Plague manual: epidemiology, distribution, surveillance and control. Technical report, World Health Organization, 1999.
 44. Amy J. Vogler, Fabien Chan, David M. Wagner, Philippe Roumagnac, Judy Lee, Roxanne Nera, Mark Eppinger, Jacques Ravel, Lila Rahalison, Bruno W. Rasoamanana, Stephen M. Beckstrom-Sternberg, Mark Achtman, Suzanne Chanteau, and Paul Keim. Phylogeography and molecular epidemiology of yersinia pestis in madagascar. *PLOS NEGLECTED TROPICAL DISEASES*, 5(9), SEP 2011. ISSN 1935-2735. doi: 10.1371/journal.pntd.0001319.
 45. Amy J. Vogler, Jason W. Sahl, Nilma C. Leal, Marise Sobreira, Charles H. D. Williamson, Molly C. Bollig, Dawn N. Birdsell, Andrew Rivera, Brian Thompson, Roxanne Nottingham, Antonio M. Rezende, Paul Keim, Alzira M. P. Almeida, and David M. Wagner. A single introduction of yersinia pestis to brazil during the 3rd plague pandemic. *PLOS ONE*, 14(1), JAN 9 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0209478.
 46. René Migliani, Suzanne Chanteau, Lila Rahalison, Maherisoa Ratsitorahina, Jean Paul Boutin, Lala Ratsifasoamanana, and Jean Roux. Epidemiological trends for human plague in madagascar during the second half of the 20th century: a survey of 20 900 notified cases. *Tropical Medicine & International Health*, 11(8):1228–1237, 2006.
 47. Natalia A Kuzmina, Philippe Lemey, Ivan V Kuzmin, Bonny C Mayes, James A Ellison, Lillian A Orciari, Dillon Hightower, Steven T Taylor, and Charles E Rupprecht. The phylogeography and spatiotemporal spread of south-central skunk rabies virus. *PLoS one*, 8(12): e82348, 2013.
 48. Simon Dellicour, Cecile Troupin, Fatemeh Jahanbakhsh, Akram Salama, Siamak Massoudi, Majid K. Moghaddam, Guy Baele, Philippe Lemey, Alireza Gholami, and Herve Bourhy. Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages - application to rabies virus spread in iran. *MOLECULAR ECOLOGY*, 28(18):4335–4350, SEP 2019. ISSN 0962-1083. doi: 10.1111/mec.15222.
 49. Raphaëlle Klitting, Liana E Kafetzopoulou, Wim Thiery, Gytis Dudas, Sophie Griseels, Anjali Kotamarthi, Bram Vrancken, Karthik Gangavarapu, Mambu Momoh, John Demby Sandy, et al. Predicting the evolution of lassa virus endemic area and population at risk over the next decades. *bioRxiv*, 2021.
 50. Carolina Torres, C Lema, F Gury Dohmen, F Beltran, L Novaro, S Russo, MC Freire, A Velasco-Villa, Viviana Andrea Mbayed, and DM Cisterna. Phylodynamics of vampire bat-transmitted rabies in a argentina. *Molecular ecology*, 23(9):2340–2352, 2014.
 51. Sebastian Duchene, Leo Featherstone, Melina Haritopoulou-Sinanidou, Andrew Rambaut, Philippe Lemey, and Guy Baele. Temporal signal and the phylodynamic threshold of sars-cov-2. *Virus evolution*, 6(2):veaa061, 2020.
 52. Anthony Lam and Sebastian Duchene. The impacts of low diversity sequence data on phylodynamic inference during an emerging epidemic. *Viruses*, 13(1):79, 2021.
 53. Sebastian Duchène, Kathryn E Holt, François-Xavier Weill, Simon Le Hello, Jane Hawkey, David J Edwards, Mathieu Fourment, and Edward C Holmes. Genome-scale rates of evolutionary change in bacteria. *Microbial genomics*, 2(11), 2016.
 54. Yujun Cui, Chang Yu, Yanfeng Yan, Dongfang Li, Yanjun Li, Thibaut Jobart, Lucy A. Weichert, Zuyun Wang, Zhaobiao Guo, Lizhi Xu, Yujiang Zhang, Hancheng Zheng, Nan Qin, Xiao Xiao, Mingshou Wu, Xiaoyi Wang, Dongsheng Zhou, Zhizhen Qi, Zongmin Du, Honglong Wu, Xianwei Yang, Hongzhi Cao, Hu Wang, Jing Wang, Shusen Yao, Alexander Rakin, Yin-grui Li, Daniel Falush, Francois Balloux, Mark Achtman, Yajun Song, Jun Wang, and Ruifu Yang. Historical variations in mutation rate in an epidemic pathogen, yersinia pestis. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 110(2):577–582, JAN 8 2013. ISSN 0027-8424. doi: 10.1073/pnas.1205750110.
 55. Simon Dellicour, Philippe Lemey, Marc A Suchard, Marius Gilbert, and Guy Baele. Accommodating sampling location uncertainty in continuous phylogeography. *Virus Evolution*, 2022.
 56. Alexis Criscuolo and Sylvain Brisse. Alientrimmer: a tool to quickly and accurately trim off multiple short contaminant sequences from high-throughput sequencing reads. *Genomics*, 102(5-6):500–506, 2013.
 57. Michael R Cruseo, Hussien F Alameldin, Sherine Awad, Elmar Boucher, Adam Caldwell, Reed Cartwright, Amanda Charbonneau, Bede Constantinides, Greg Edvenson, Scott Fay, et al. The khmer software package: enabling efficient nucleotide sequence analysis. *F1000Research*, 4, 2015.
 58. Yongchao Liu, Jan Schröder, and Bertil Schmidt. Musket: a multistage k-mer spectrum-based error corrector for illumina sequence data. *Bioinformatics*, 29(3):308–315, 2013.
 59. Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A Gurevich, Mikhail Dvorkin, Alexander S Kulikov, Valery M Lesin, Sergey I Nikolenko, Son Pham, Andrey D Pribelski, et al. Spades: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*, 19(5):455–477, 2012.
 60. J Parkhill, BW Wren, NR Thomson, RW Titball, MTG Holden, MB Prentice, M Sebahia, KD James, C Churcher, KL Mungall, S Baker, D Basham, SD Bentley, K Brooks, AM Cerdeno-Tarraga, T Chillingworth, A Cronin, RM Davies, P Davis, G Dougan, T Feltwell, N Hamlin, S Holroyd, K Jagels, AV Karlyshev, S Leather, S Moule, PCF Oyston, M Quail, K Rutherford, M Simmonds, J Skelton, K Stevens, S Whitehead, and BG Barrell. Genome sequence of yersinia pestis, the causative agent of plague. *NATURE*, 413(6855):523–527, OCT 4 2001. ISSN 0028-0836. doi: 10.1038/35097083.
 61. Vladimir V. Kutyrev, Galina A. Eroshenko, Vladimir L. Motin, Nikita Y. Nosov, Jaroslav M. Krasnov, Lyubov M. Kukleva, Konstantin A. Nikiforov, Zhanna V. Al'khova, Eugene G. Oglodin, and Natalia P. Guseva. Phylogeny and classification of yersinia pestis through the lens of strains from the plague foci of commonwealth of independent states. *FRONTIERS IN MICROBIOLOGY*, 9, MAY 25 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.01106.
 62. Nicholas J Croucher, Andrew J Page, Thomas R Connor, Aidan J Delaney, Jacqueline A Keane, Stephen D Bentley, Julian Parkhill, and Simon R Harris. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using gubbins. *Nucleic acids research*, 43(3):e15–e15, 2015.
 63. Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *MOLECULAR BIOLOGY AND EVOLUTION*, 37(5):1530–1534, MAY 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.
 64. Subha Kalyaanamoorthy, Bui Quang Minh, Thomas KF Wong, Arndt Von Haeseler, and Lars S Jermini. Modelfinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6):587–589, 2017.
 65. Diep Thi Hoang, Olga Chernomor, Arndt Von Haeseler, Bui Quang Minh, and Le Sy Vinh. Ufboot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35(2):518–522, 2018.
 66. Lu Cheng, Thomas R Connor, Jukka Sirén, David M Aanensen, and Jukka Corander. Hierarchical and spatially explicit clustering of dna sequences with baps software. *Molecular biology and evolution*, 30(5):1224–1228, 2013.
 67. Gerry Tonkin-Hill, John A Lees, Stephen D Bentley, Simon DW Frost, and Jukka Corander. Rhierbaps: An r implementation of the population clustering algorithm hierbaps. *Wellcome open research*, 3, 2018.
 68. Bo Liu, Dandan Zheng, Qi Jin, Lihong Chen, and Jian Yang. Vfdp 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic acids research*, 47(D1): D687–D692, 2019.
 69. Torsten Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14): 2068–2069, 2014.
 70. Gerry Tonkin-Hill, Neil MacAlasdair, Christopher Ruis, Aaron Weimann, Gal Horesh, John A Lees, Rebecca A Gladstone, Stephanie Lo, Christopher Beaudoin, R Andres Floto, et al. Producing polished prokaryotic pangenomes with the panaroo pipeline. *Genome biology*, 21(1):1–21, 2020.
 71. Pavel Sagulenko, Vadim Puller, and Richard A Neher. Treetime: Maximum-likelihood phylogenetic analysis. *Virus evolution*, 4(1):vex042, 2018.
 72. Daniel Ayres, Michael Cummings, Guy Baele, Aaron Darling, Paul Lewis, David Swofford, John Huelsenbeck, Philippe Lemey, Andrew Rambaut, and Marc Suchard. Beagle 3: Improved performance, scaling, and usability for a high-performance computing library for statistical phylogenetics. *SYSTEMATIC BIOLOGY*, 68(6):1052–1061, NOV 2019. ISSN 1063-5157. doi: 10.1093/sysbio/syz020.
 73. Andrew Rambaut, Alexei J Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. Posterior summarization in bayesian phylogenetics using tracer 1.7. *SYSTEMATIC BIOLOGY*, 67(5):901–904, SEP 2018. ISSN 1063-5157. doi: 10.1093/sysbio/syy032.
 74. Randle Aaron M Villanueva and Zhuo Job Chen. ggplot2: elegant graphics for data analysis, 2019.
 75. Oliver G. Pybus, Marc A. Suchard, Philippe Lemey, Flavien J. Bernard, Andrew Rambaut, Forrest W. Crawford, Rebecca R. Gray, Nimalan Arinaminpathy, Susan L. Stramer, Michael P. Busch, and Eric L. Delwart. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 109(37):15066–15071, SEP 11 2012. ISSN 0027-8424. doi: 10.1073/pnas.1206598109.
 76. Guy Baele, Simon Dellicour, Marc A. Suchard, Philippe Lemey, and Bram Vrancken. Recent advances in computational phylodynamics. *CURRENT OPINION IN VIROLOGY*, 31:24–32, AUG 2018. ISSN 1879-6257. doi: 10.1016/j.coviro.2018.08.009.
 77. Simon Dellicour, Rebecca Rose, Nuno R. Faria, Philippe Lemey, and Oliver G. Pybus. Seraphim: studying environmental rasters and phylogenetically informed movements. *BIOINFORMATICS*, 32(20):3204–3206, OCT 15 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw384.

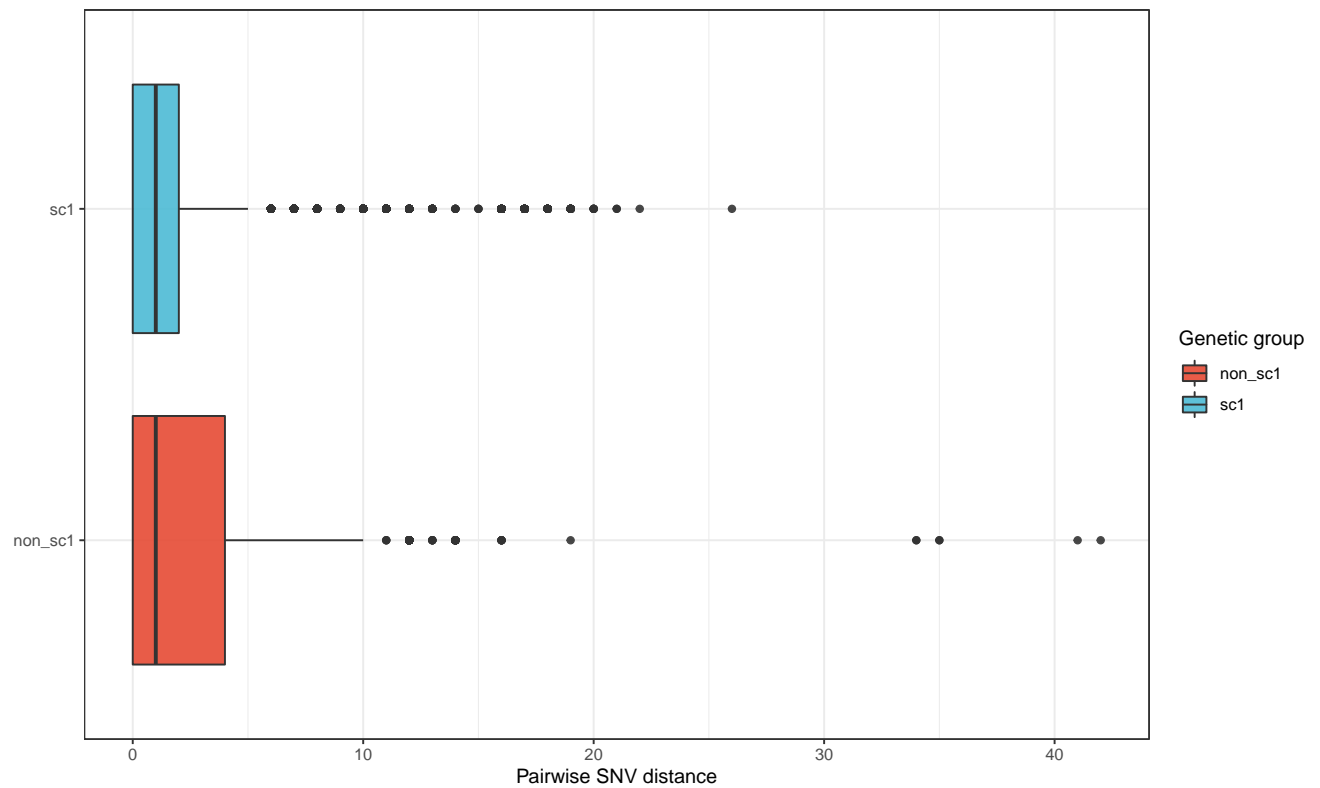
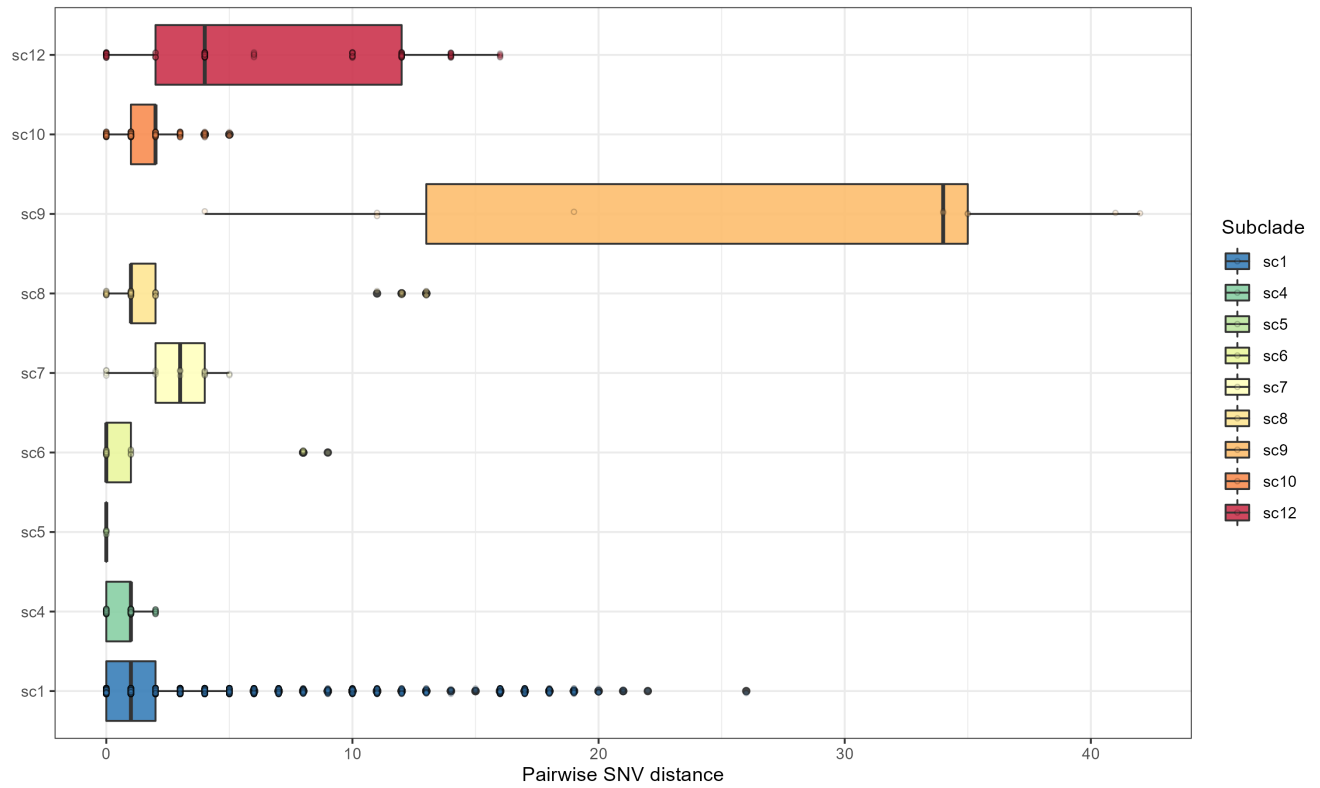
Supplementary Figures



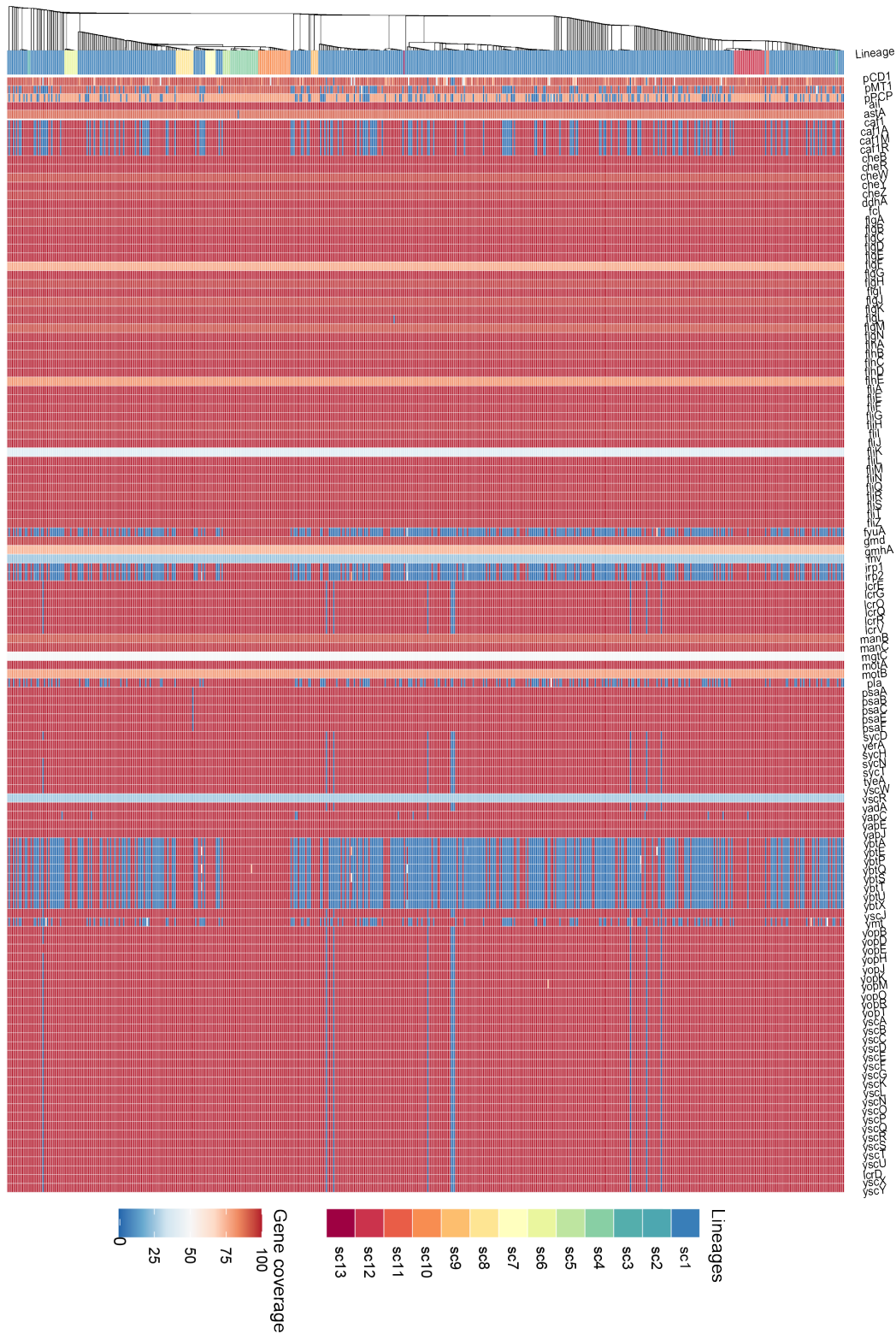
Supplementary Figure 1. ML phylogeny of the 570 *Y. pestis* isolates of the clade circulating between 1962-1996 (1.ORI4 clade). The ML tree corresponds to the phylogeny shown in Fig. 3. Tips are colored according to the identified monophyletic subclades. The color strip annotations show the genetic group at the subcluster level estimated by the Bayesian analysis of population structure approach (hierBAPS). hierBAPS subclusters follow the same distribution as the identified monophyletic subclades except for the hierBAPS subcluster 2 which is paraphyletic. The annotation of host niche and animal species where each isolate was collected is displayed



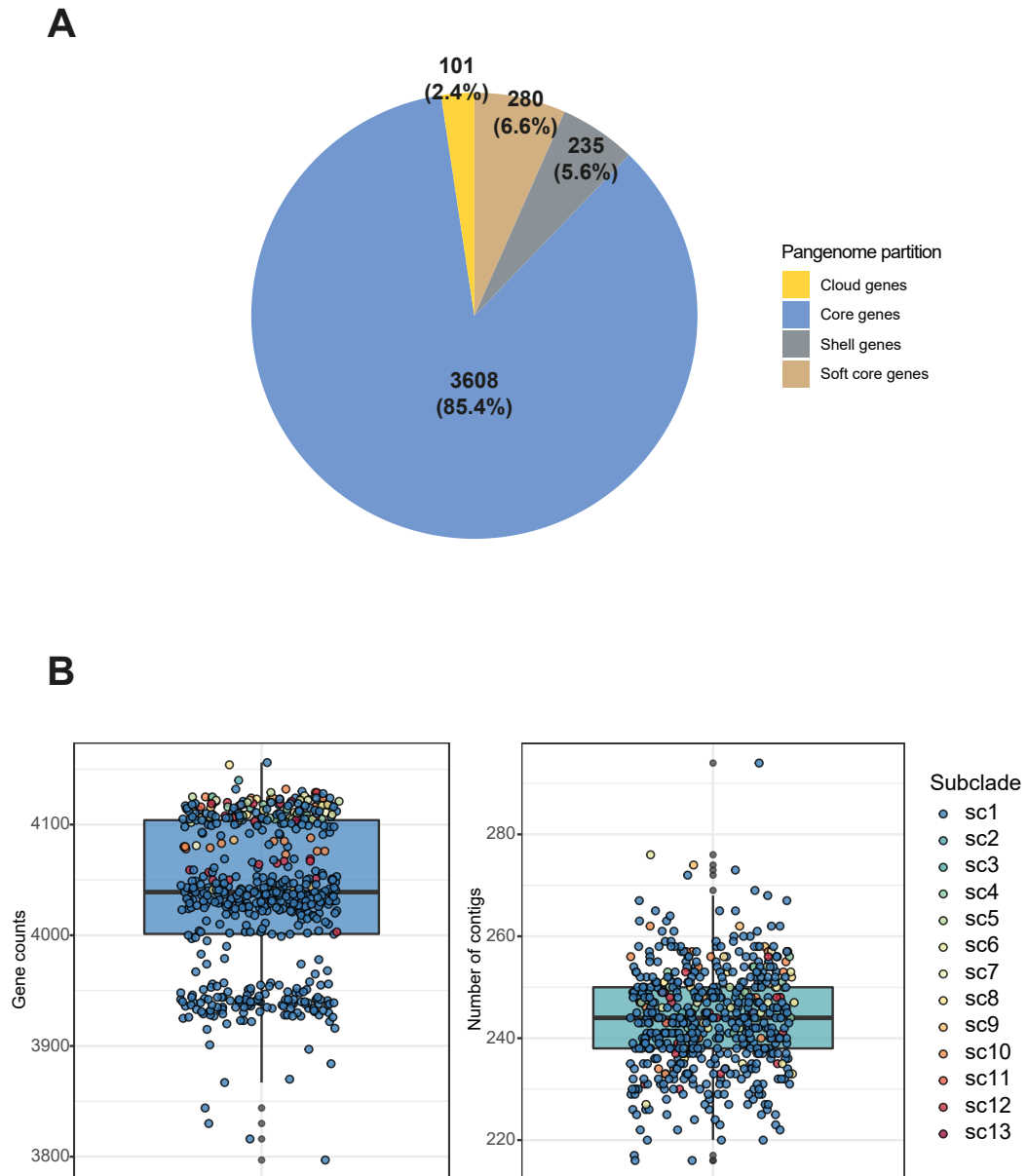
Supplementary Figure 2. Temporal and geographic distribution of the *Y. pestis* isolates in Vietnam from two major periods of sampling. The map shows the locations of the 570 *Y. pestis* isolates used in the inference of the ML tree, divided in the two major decades of sampling in the dataset (1960s decade and 1980s-90s decade). Colors and shapes are defined as in Fig. 3. The timeline of sampling of the isolates is colored according to the identified lineages of the isolates.



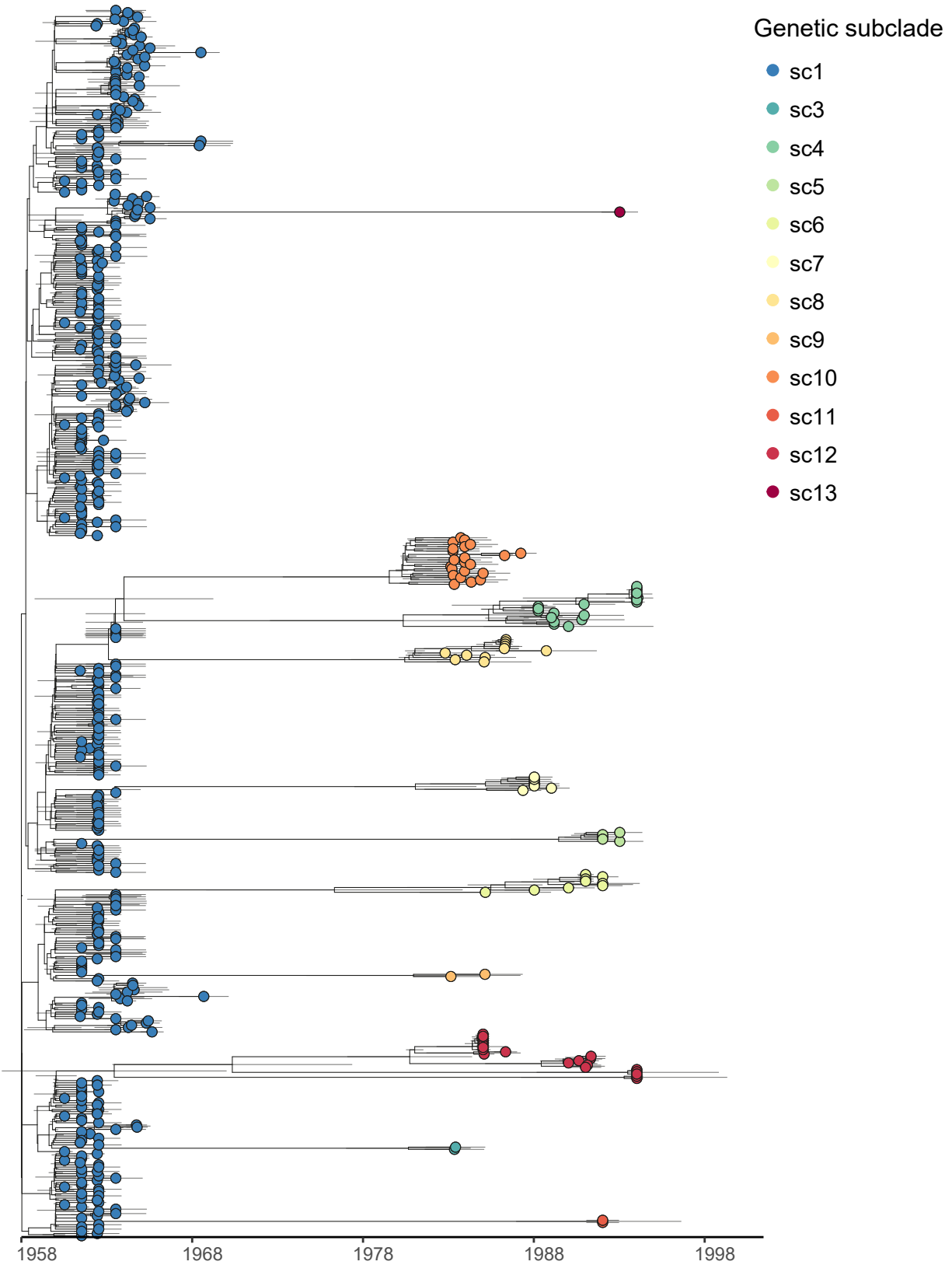
Supplementary Figure 3. Distribution of pairwise SNV distances within isolates from the different identified lineages. Distances were calculated for all monophyletic lineages that are represented by at least two genomes (up). The second plot compares the distribution of pairwise SNV distances within isolates of sc1 compared to the other lineages together.



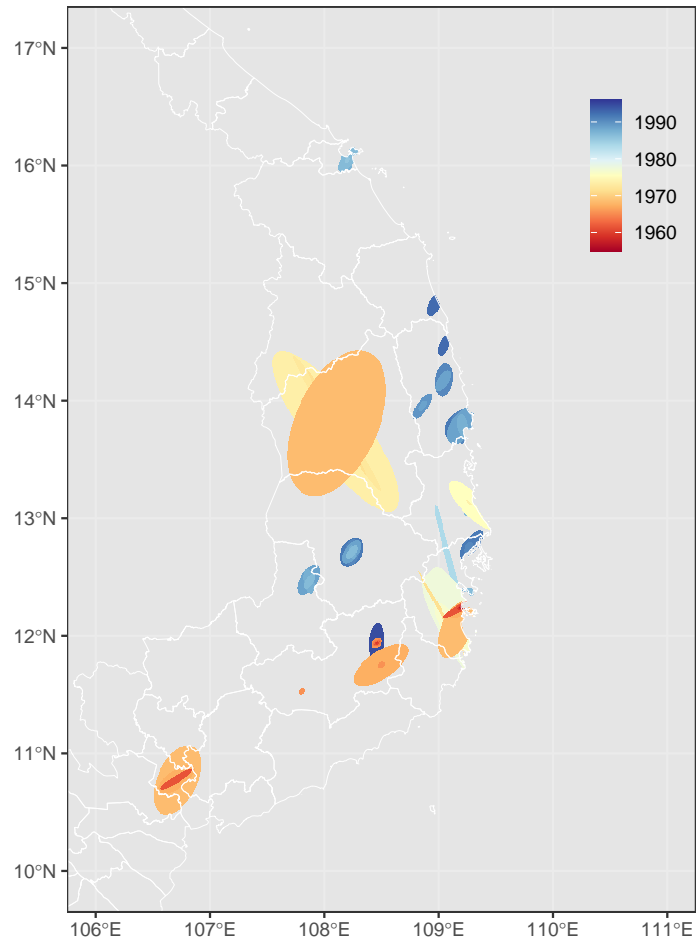
Supplementary Figure 4. Virulence gene analysis in the 570 *Y. pestis* genomes part of the major population circulating in Vietnam. Heatmap of hits for virulence genes screened in the *Y. pestis* genomes of the 1.ORI4 clade. Each column corresponds to a virulence-associated gene ($n=123$) present in the genomes of the dataset resulting from BLAST hits against the Virulence Factor Database, as well as for the sequences of the three major plasmids in *Y. pestis* reference CO92 (pYV, pMT1, pPla). The heatmap colors are scaled to the percentage of coverage on the hit sequence, and the y axis is ordered according to the phylogenetic tree shown in Fig. 3. The lineage correspondence for each genome is also indicated.



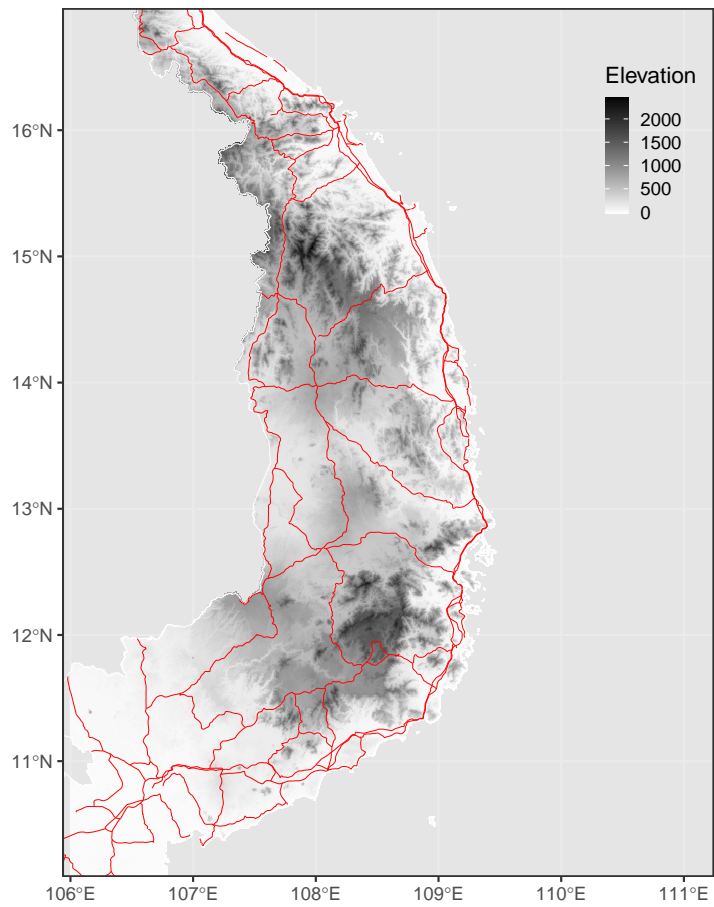
Supplementary Figure 5. Pangenome analysis of the 570 Vietnamese genomes **A:** Pangenome partitions as identified by Panaroo from the detection of 4224 orthologous gene families. The total number of gene families in each partition and their percentage are shown. Core genes correspond to gene families conserved in at least 99% of isolates. The accessory genome was partitioned between the soft core genes (conserved in >95%), shell genes (conserved in between 15% and 95 % of isolates) and cloud genes (conserved in up to 15% of isolates). **B:** Distributions of total gene counts (left) per genome and of the total number of contigs (right) in the de novo assemblies used as input for the pangenome reconstruction. Dots are colored by the corresponding lineage of the genome. The discrete groups of genomes according to their total gene count is due to the presence or absence of replicons that are frequently lost during in vitro culture of the isolates: the pMT1 plasmid and the pgm locus (as explained in the main text).



Supplementary Figure 6. Maximum clade credibility (MCC) tree from the Skygrid phylodynamic inference. The time-scaled tree includes the dataset of 554 genomes after excluding outlier sequences. The tree was inferred using the flexible Skygrid model and corresponds to the one displayed in Fig. 4. Tips of each leaf are colored according to the monophyletic lineage, and horizontal bars indicate at each node indicate the 95% highest posterior density (HPD) intervals for their inferred age.



Supplementary Figure 7. 95% HPD polygons for the locations of nodes in the spatially explicit phylogeographic reconstruction. The polygons correspond to the geographically-annotated MCC tree based on 100 posterior trees, as shown in Fig. 5, and are colored according to their inferred median age.



Supplementary Figure 8. Major transportation networks in Vietnam. The distribution of major transportation routes in the country as shown in the small plot in Fig. 5, including railways, principal roads and secondary roads. The geospatial dataset was obtained from the [OpenDevelopment site](#).

Part IV

General discussion

The presented thesis explored the diversity, geographic distribution, and evolutionary history of *Y. pestis* species through the lens of genomics and molecular evolution. The first manuscript developed the analysis of a global collection of *Y. pestis* isolates that were collected over the last century. From this analysis, we aimed at retracing the recent evolutionary paths of plague that were missed in previous works, and to integrate the new genomes into a global overview of *Y. pestis* species evolution and genetic diversity. In the second manuscript, we studied the dynamics of *Y. pestis* spread during outbreaks in an endemic country with relevant epidemic history.

Genomic analyses have emerged in the recent years as powerful tools to study infectious diseases. With the technological advances to generate, manipulate and analyze whole-genome sequencing data, new horizons have appeared in the study of human pathogens. Pathogen phylogenies represent the reconstruction of microbe genealogies and allow to place the historic of infections into a temporal and spatial context (335). In the last decade, a few phylogenomic analyses have allowed to identify the major lineages in *Y. pestis* species evolution, mainly through the sequencing of isolates collected over the past century (46, 62, 64, 89). In parallel, within this last decade the field of microbial paleogenomics has emerged and provided for the first time with genomic data from *Y. pestis* present in infected human remains over the past millennia (240), including the strains that circulated in major plague pandemics in history (262, 264, 304). Pathogen phylogenies represent however just a glimpse, a snapshot, of the diversity and evolutionary relationships of the samples that are present in the analysis. Therefore, the hypotheses that are formulated from phylogenetic inference may be complemented, if not reformulated, when new diversity is added through the obtention of more samples (345). The recent increase in available *Y. pestis* genomes has expanded the diversity found within lineages described in the past, giving a stepwise and usually partial outline of *Y. pestis* evolution, as most studies have limited to the interpretation of the placement of the novel genomes, but rarely on analyzing the general patterns of *Y. pestis*

diversity at species level. In this work, we aimed to fill this gap and take the opportunity to recapitulate the species global phylogenetic distribution and patterns of genomic variation using thousands of *Y. pestis* genomes from global sources. By taking a genomic-driven approach, in the presented work we addressed some questions that are part of the global understanding of *Y. pestis* evolution and epidemiology.

When and where do *Y. pestis* lineages emerge and spread?

Phylogenetic analyses have been extensively used to retrace the emergence of different lineages of *Y. pestis* and in relation to the development of events which are relevant to human history. Reconstructing the emergence and spread of lineages can give valuable information to better understand plague epidemiology and develop strategies to control the disease. In the first manuscript, we used the large collection of *Y. pestis* isolates mainly collected in the context of the Third Pandemic to retrace the pandemic spread of the 1.ORI lineage, confirming its dissemination in successive waves that introduced *Y. pestis* into different continents (46). This led us to focus on one of this waves, that we related to the emergence of the 1.ORI5 sublineage, which spread into Europe, Africa and South America. We have shown that this sublineage was introduced into different coastal areas in the South and North-West of the African continent at the end of the 19th century, that we could associate to historical records. Some of these introductions were responsible for the long-term establishment of plague, where it has persisted to cause outbreaks until the recent decades like in Algeria (23). We have also provided for the first time with a detailed outline of the emergence and evolution of 1.ANT lineage in Central and Eastern Africa. The possible origin and transmission route into Africa of this lineage aroused great interest from the historians of plague because 1.ANT directly branches after the 2nd pandemic ancient genomes from Europe, a lineage that become extinct with the disappearing of plague from Europe. We hope that the provided 1.ANT genomes contribute to solve ongoing discussions regarding

the emergence and spread of this lineage.

An intriguing question for which *Y. pestis* phylogenetic analyses have been particularly informative has to do with the endemic establishment of plague into a new area during epidemic or pandemic expansion. It can be pointed out that for several countries that became endemic for plague during the Third Pandemic (i.e.: Brazil, Madagascar, Peru), all the *Y. pestis* diversity present at the long term resulted from a single and successful introduction event. Many countries experienced plague outbreaks at the early years of the Third Pandemic (16, 61), when the major 1.ORI clone was spreading worldwide, and hence these countries were exposed in multiple occasions to the long-term establishment of plague within local rodent populations. This conclusion is relevant for the epidemiology of plague since it indicates that even in the context of pandemic circulation of *Y. pestis*, the successful introduction and establishment of plague is a relatively rare event. The reasons beyond the dynamics of maintenance of plague are complex and require the understanding of disease ecology (1). It has been suggested that *Y. pestis* may sometimes only become transiently introduced into local rodent communities, dissipating after few years (293, 344). This interplay between the animal reservoir of plague, and the (re)emergence and spread of *Y. pestis* in outbreaks or pandemics has also been approached from the genomic analysis of *Y. pestis*. For instance, it has been possible to infer that major outbreaks of human plague were likely caused by the emergence and rapid expansion of genetically similar *Y. pestis* strains, such as during the Black Death in the 14th century (262, 299), or the early Third Pandemic (46). The analyses of *Y. pestis* genomes associated to these outbreaks could not deduce any evidence suggesting that the emergence of these clones circulating in widespread outbreaks could be explained through the acquisition of genetic factors favoring a possible increase in virulence (299). A plausible explanation for the rapid spread of epidemic clones may have to do with the meeting of multiple conditions, including the abundance of susceptible rodent populations (346), their ecological relation with humans nearby (72) and the movements of

human populations, that influence the spread of infections (347). Our analysis of terminal branch lengths may represent a new perspective in the interpretation of *Y. pestis* phylogenies and the emergence and spread of lineages, as shorter terminal branch lengths may illuminate lineages that have undergone recent expansion.

The sequencing of *Y. pestis* genomes from the same endemic area during different outbreaks separated in time can provide relevant insights into the maintenance of *Y. pestis* between outbreaks and its epidemic spread. Phylodynamic approaches can be used to address the question of whether a second outbreak stems from the persistence of the pathogen circulating in a previous outbreak or if it results from a new introduction event from an external source. In the second manuscript, we had the opportunity to explore these dynamics of evolution and spread of *Y. pestis* from the analysis of isolates that were collected over decades from Vietnam, which was one of the most active endemic areas of plague in the past century (348). Our analyses suggested that the *Y. pestis* strains that were responsible for the human plague activity in the 1980s and 90s decade evolved from the clone that had become widespread in Vietnam during major outbreaks in the 1960s decade, and that these strains had been evolving between the outbreaks separated through decades. The analysis of *Y. pestis* isolates collected over decades in the same endemic country contributed to give a better approximation of the minimum temporal threshold that may be necessary to calculate robust estimates of the evolutionary rates of *Y. pestis* in endemic countries. Moreover, we estimated the dynamics of spread of *Y. pestis* during outbreaks in Vietnam using continuous space phylogeographic approaches, which had never been used to explore past plague outbreaks. In this analysis it was possible to link infections through time and space and infer the spreading velocities of plague in Vietnam. The context of development of the fastest *Y. pestis* transmissions overlapped with the development of a long-standing war in Vietnam, which may have triggered human displacements that would explain the rapid spread of plague that we inferred. This context reminds us the distressing relationship between wars

and infectious diseases, which extends to the present times with the recent plague outbreaks in the Ituri region of the Democratic Republic of Congo, as well as for the outbreaks of cholera in Yemen (2016-17)(326), or of measles in Syria (2017-18)(349).

Patterns of genomic variation between *Y. pestis* lineages

In another main axis of this thesis, that we began to develop in the first manuscript, we go beyond historical and evolutionary reconstructions and explore the functional diversity of *Y. pestis* lineages from different standpoints. The identification of genetic variations that contribute to *Y. pestis* adaptation may be challenging due to the reduced levels of genetic diversity that were initially found when comparing the point mutations between most of the known *Y. pestis* lineages (242, 247). Genomic comparisons have yet identified few genomic variants that appeared after the emergence of *Y. pestis* and may have been important for its virulence and dissemination. A single amino acid substitution in the Pla protease has been associated with the capacity of *Y. pestis* to cause a more invasive infection in animals, as revealed when studied in mice models of bubonic plague (215). The acquisition of novel factors also appears to have played a role in the evolutionary success of *Y. pestis* (230). The analysis of ancient DNA metagenomes of *Y. pestis* from the Late Neolithic has permitted to refine the estimations for the acquisition of the Ymt phospholipase (265, 270), which would have allowed *Y. pestis* to greatly expand the range of hosts in which it can establish a flea-borne circulation (221). These molecular changes occurred at the most ancestral lineages in *Y. pestis* phylogeny, indicating that most of the contemporary *Y. pestis* lineages emerged once the bacillus already had become a flea-borne, extremely virulent pathogen. Therefore, we can hypothesize that many of the adaptative changes in *Y. pestis* lineages that circulate nowadays may rather be related to ensuring its survival and suitable transmission in the endemic cycles where they circulate. An example of such adaptation is the recent finding that *Y. pestis* can use variants in the *rpoZ* gene to alter its levels of biofilm formation

according to fluctuating climatic conditions (271). There may be also genomic changes in recently emerged lineages implicating differential virulence in mammal hosts, but such variants have not yet been identified.

Genomic variation may also have helped *Y. pestis* to establish permanently into new geographic areas after epidemic or pandemic expansions such as the Third Pandemic, during which *Y. pestis* established in many different countries, with their specific ecologic conditions and rodent and vector species diversities. This hypothesis, however, has not been addressed in previous *Y. pestis* genomic analyses. Our work is, to the best of our knowledge, the first large-scale genomic analysis of *Y. pestis* species that explores in parallel multiple sources of genomic variation to address possible hallmarks of adaptative evolution in *Y. pestis* lineages. We identified an enrichment of missense variants in genes coding for factors associated to stress sensing in isolates from the 1.ORI and 1.ANT lineages, and we found that multiple loci associated to virulence mechanisms are part of the genes presenting highest rates of missense variants. The pangenomic analysis of hundreds of assemblies provided evidence of the variation in gene repertoires associated with different lineages. Finally, we demonstrate the expansion of IS during *Y. pestis* microevolution, notably through the expansion of IS100 and IS200 copies. The pattern of IS insertions in the *Y. pestis* chromosome was also informative of the lineage to which the genome belongs to, which confirms the basis of previous efforts for using the specific patterns of IS insertions for the typing of *Y. pestis* isolates (315). The expansion of IS was remarkably higher in 1.ANT lineage, which has led to the specific inactivation of more than a hundred genes in genomes of this lineage. The characterization of the potential functional impact of these inactivations is an exciting perspective that we would like to adress in short term.

Even if these results are still preliminary, the uncovering of these novel patterns of diversity is particularly exciting and altogether points towards a more complex understanding of *Y. pestis* evolution, in which multiple mechanisms including point mutations, gene gain,

gene inactivation through IS insertions (and probably through indels too (350)) act together as driving forces in the emergence of *Y. pestis* lineages and their functional diversification. The combination of these different types of genomic variation has possibly contributed to the adaptation of *Y. pestis* during its historical spread and long-term establishment into diverse geographies and ecological niches. This work highlights the need for developing large genomic analyses of diverse *Y. pestis* isolates to be able to effectively detect possible signals of adaptation in the evolution of *Y. pestis* lineages. Future steps will be directed towards characterizing the impact of the variants identified in different isolates, using experimental approaches. The pursuing of the search of signatures of adaptative selection in *Y. pestis* will require the analysis of other signals that have not been explored yet in this work, such as homoplasies. Large datasets of global *Y. pestis* genomes may also be used to infer patterns of genetic variation present within major epitopes that are targets of vaccine candidates currently under development (351). This may help identifying possible pitfalls within future efforts in plague immunization and contribute to the evaluation of ongoing vaccine candidates. Genome sequencing of large collections of isolates can also inform about antibioresistance risks in the era of widespread antimicrobial resistance, by detecting the prevalence of resistance gene sequences within the bacterial population. In this study, no new antimicrobial resistant *Y. pestis* isolate was identified within the isolates sequenced (data not shown), which suggests that antimicrobial resistance has not yet been a selective factor for *Y. pestis* populations causing the spread of resistant clones. However the recent description of *Y. pestis* isolates resistant to streptomycin in Madagascar and China (157, 352) strengthens the need for surveillance of antimicrobial resistances in the context of countries where plague is endemic.

In all, analyses of *Y. pestis* genomic diversity have provided relevant details on the complex dynamics of *Y. pestis* in plague endemic areas, which are affected by diverse ecological and epidemiological factors. The study of the genomic epidemiology of plague requires the

recovery of *Y. pestis* isolates over decades of surveillance in infected animals and human patients, together with a detailed knowledge of the epidemiological history of plague in the area. The regular monitoring of *Y. pestis* circulation and isolation of live bacilli from rodent populations and during outbreaks of human plague is therefore key to infer the dynamics of the disease in endemic countries. This planned surveillance is not only informative for epidemiological trends of plague in such countries, but it has been also decisive for the success of plague control in important endemic areas for decades, like in the former Soviet Union (109). At a broader scale, the integration of genomic data from large collections covering different historically active plague regions is essential to uncover novel patterns of genomic variation and possible adaptive evolution signatures, as well as to characterize the diversity and specific markers of the global genetic and geographic groups of *Y. pestis*. This highlights the need for the open sharing of pathogen genomic data within countries and institutions, which has been stressed during the COVID-19 pandemic as a key to outbreak preparedness (353). The development of a resolutive scheme for the genotyping of *Y. pestis* isolates and its implementation for research and public health laboratories would represent an opportunity for retracing the possible sources of future plague outbreaks in the era of unprecedented international travel and communication, which will be decisive for taking early moves on the control of future outbreaks.

Part V

Concluding remarks

With this thesis, I hope to have contributed to the better understanding of the diversity, evolution, and epidemiology of the plague bacillus. The genomic resources generated with this work will be valuable for both plague research and public health strategies, and may I realize when summarizing these results that we have probably just scratched the surface of the patterns that were hiding behind the large amount of sequencing data generated here. This project was presented itself as challenging and ambitious since its conception. Nevertheless, it altogether represented the extraordinary chance of developing a project on a unique collection of strains in the world, in one of the oldest laboratories within the Institut Pasteur. I could feel the legacy of previous fellows and their pioneering work, implicating the research on different aspects of plague through many international partnerships, since the very early days working in the Yersinia Research Unit. I was fortunate enough to join a team of passionate and valuable researchers both at the personal and scientific levels, which regularly fueled the discussions and interactions on the different projects. Coming from a molecular biology and microbiology background, this project has also introduced me to an almost full-genomics work, which has provided me with a surprisingly balanced combination of constant learning experiences and frustrations. These early steps into academic research have highlighted to me an essential aspect of scientific research, which may as well be the only possible solution to some of the deep, structural problems that the scientific community will have to address in the following decades. Indeed, collaborative research is this key factor for the development of both this project and a myriad of ongoing scientific research efforts. Collaboration represents necessarily the sharing of work, resources, expertise, and infrastructure applied at all steps of a research project. The work hereby presented thus represents a collaborative effort at all these different levels. First, the collection of *Y. pestis* isolates stems from the previous and ongoing interactions and share of strains between different laboratories of the Pasteur international network and others, who developed the decisive field work, and the Institut Pasteur in Paris. The generation of sequencing data of all the collection

also involved the laboratory hosting the sequencing platform and the technical expertise on sequencing methods. It would not be possible to put all these samples into their genetic and evolutionary context without the publication of genomic data of the different strains analyzed in previous studies by other teams. The open sharing of pathogen genomic data has been critical for this and many other genomic studies; for *Y. pestis*, which is a select agent with bioterrorism potential, sharing genomic data of local strains is way more convenient than the difficult transfer of live strains between laboratories. Collaboration extends to a third level, as collaborative and transdisciplinary efforts are determinant to analyze the data and provide valuable interpretation. The implication of different skillful actors able to develop bioinformatic approaches, analytical frameworks and historical context was thus another meaningful aspect of this project. Last, the open sharing of analytical tools and methods is decisive for the analysis of genomic data. This project, like many others, has benefited from the myriad of bioinformatic tools and approaches that are widely and openly distributed to the community for the prosperity of future scientific works. This example of fruitful development of collaborative research may perhaps appear as one of the oases within the desert of generalized, frenetic competition into which scientific research, and mostly biomedical research, appears to be navigating since some time in the past. Competition has been imposed into all levels of scientific production, from funding attribution to results communication into the system of scientific journals that may appear outdated in the digital era. Ironically enough, some of the most revolutionary findings in biological research have not come from a constant sake of competition but from fortuitous and basic scientific observations. The finding of penicillin or the discovery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) in bacterial genomes are two powerful reminders that competition may not be the most effective engine for scientific advancement. The dynamics of competition and ever-growing seek for justification of practical or commercial prospects of research appear to be detrimental for research ([354](#), [355](#)), and may be a source disap-

pointment to many young researchers taking the early steps into what could otherwise be an exciting journey in research. This challenging horizon for many graduates and early-stage scientists deserves the implication of all actors, and particularly the young generations, to decide how should science be done and towards which goals. The open sharing of information and resources and the generalization of collaboration as a new paradigm of scientific research may be a legitimate way to advance in our understanding of infectious diseases and any other biological research field. Altogether, cooperation dynamics, together with a reliable funding may encourage the development of liberty and creativity that will provide us with original and high-quality science, for the goodness of all scientists and that of the community.

Part VI

References

Bibliography

1. Nils Chr. Stenseth, Bakyt B. Atshabar, Mike Begon, Steven R. Belmain, Eric Bertherat, Elisabeth Carniel, Kenneth L. Gage, Herwig Leirs, and Lila Rahalison. Plague: Past, present, and future. *PLOS MEDICINE*, 5(1):9–13, JAN 2008. ISSN 1549-1277. doi: 10.1371/journal.pmed.0050003.
2. R Barbieri, M Signoli, D Chevé, C Costedoat, S Tzortzis, G Aboudharam, D Raoult, and M Drancourt. *Yersinia pestis*: the natural history of plague. *Clinical microbiology reviews*, 34(1):e00044–19, 2020.
3. Barbara Bramanti, Nils Chr. Stenseth, Lars Walloe, and Xu Lei. Plague: A disease which changed the path of human civilization. In R Yang and A Anisimov, editors, *YERSINIA PESTIS: RETROSPECTIVE AND PERSPECTIVE*, volume 918 of *Advances in Experimental Medicine and Biology*, pages 1–26. 2016. ISBN 978-94-024-0890-4; 978-94-024-0888-1. doi: 10.1007/978-94-024-0890-4_1.
4. Ricardo Jorge. *Summa epidemiologica de la peste. Epidemias antiguas y modernas*. Office International d'Hygiene Publique, T.XXV,Annee1933, fasc.n3, 1933.
5. B. Joseph Hinnebusch, Clayton O. Jarrett, and David M. Bland. “fleaing” the plague: Adaptations of *Yersinia pestis* to its insect vector that lead to transmission. In S Gottesman, editor, *ANNUAL REVIEW OF MICROBIOLOGY, VOL 71*, volume 71 of *Annual Review of Microbiology*, pages 215–232. 2017. ISBN 978-0-8243-1171-1. doi: 10.1146/annurev-micro-090816-093521.
6. David T Dennis, Kenneth L Gage, Norman G Gratz, Jack D Poland, Evgeni Tikhomirov, World Health Organization, et al. Plague manual: epidemiology, distribution, surveillance and control. Technical report, World Health Organization, 1999.
7. Eric Bertherat and É Bertherat. Plague around the world in 2019. *Weekly epidemiological record*, 4, 2019.
8. World Health Organization et al. International meeting on preventing and controlling plague: the old calamity still has a future. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, 81(28):278–284, 2006.

-
9. Charles Evans. Pneumonic plague: Incidence, transmissibility and future risks. *Hygiene*, 2 (1):14–27, 2022.
 10. JL Kool. Risk of person-to-person transmission of pneumonic plague. *CLINICAL INFECTIOUS DISEASES*, 40(8):1166–1172, APR 15 2005. ISSN 1058-4838. doi: 10.1086/428617.
 11. Rindra Rendremanana, Voahangy Andrianaivoarimanana, Birgit Nikolay, Beza Ramasindrazana, Juliette Paireau, Quirine Astrid ten Bosch, Jean Marius Rakotondramanga, Soloandry Rahajandraibe, Soanandrasana Rahelinirina, Fanjasoa Rakotomanana, Feno M. Rakotoarimanana, Lea Bricette Randriamampionona, Vaoary Razafimbiana, Mamy Jean De Dieu Randria, Mihaja Raberahona, Guillain Mikaty, Anne-Sophie Le Guern, Lamina Arthur Rakotonjanabelo, Charlotte Faty Ndiaye, Voahangy Rasolofo, Eric Bertherat, Maherisoa Ratsitorahina, Simon Cauchemez, Laurence Baril, Andre Spiegel, and Minoarisoa Rajerison. Epidemiological characteristics of an urban plague epidemic in madagascar, august-november, 2017: an outbreak report. *LANCET INFECTIOUS DISEASES*, 19(5): 537–545, MAY 2019. ISSN 1473-3099. doi: 10.1016/S1473-3099(18)30730-8.
 12. Rebecca J. Eisen, Kevin S. Griffith, Jeff N. Borchert, Katherine MacMillan, Titus Apangu, Nicholas Owor, Sara Acayo, Rogers Acidri, Emily Zielinski-Gutierrez, Anna M. Winters, Russell E. Ensore, Martin E. Schriefer, Charles B. Beard, Kenneth L. Gage, and Paul S. Mead. Assessing human risk of exposure to plague bacteria in northwestern uganda based on remotely sensed predictors. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 82(5):904–911, MAY 2010. ISSN 0002-9637. doi: 10.4269/ajtmh.2010.09-0737.
 13. Mary Danforth, Mark Novak, Jeannine Petersen, Paul Mead, Luke Kingry, Matthew Weinburke, Danielle Buttke, Gregory Hacker, James Tucker, Michael Niemela, Bryan Jackson, Kerry Padgett, Kelly Liebman, Duc Vugia, and Vicki Kramer. Investigation of and response to 2 plague cases, yosemite national park, california, usa, 2015. *EMERGING INFECTIOUS DISEASES*, 22(12):2045–2053, DEC 2016. ISSN 1080-6040. doi: 10.3201/eid2212.160560.
 14. Aaron Aruna Abedi, Jean-Christophe Shako, Jean Gaudart, Bertrand Sudre, Benoit Kebela Ilunga, Stomy Karhemere Bi Shamamba, Georges Diatta, Bernard Davoust, Jean-Jacques Muyembe, Tamfum Renaud Piarroux, and Martine Piarroux. Ecologic features of plague outbreak areas, democratic republic of the congo, 2004-2014. *EMERGING INFECTIOUS*

-
- DISEASES*, 24(2):210–220, FEB 2018. ISSN 1080-6040. doi: 10.3201/eid2402.160122.
15. Bolormaa Galdan, Undraa Baatar, Baigalmaa Molotov, and Otgonbaatar Dashdavaa. Plague in mongolia. *VECTOR-BORNE AND ZOOLOGICAL DISEASES*, 10(1, SI):69–75, FEB 2010. ISSN 1530-3667. doi: 10.1089/vbz.2009.0047.
 16. Simon Neerinckx, Eric Bertherat, and Herwig Leirs. Human plague occurrences in africa: an overview from 1877 to 2008. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 104(2):97–103, 2010.
 17. Tamara Ben-Ari, Simon Neerinckx, Lydiane Agier, Bernard Cazelles, Lei Xu, Zhibin Zhang, Xiye Fang, Shuchun Wang, Qiyong Liu, and Nils C. Stenseth. Identification of chinese plague foci from long-term epidemiological data. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 109(21):8196–8201, MAY 22 2012. ISSN 0027-8424. doi: 10.1073/pnas.1110585109.
 18. Hau V. Pham, Dat T. Dang, Nguyen N. Tran Minh, D. Nguyen, and Tuan V. Nguyen. Correlates of environmental factors and human plague: an ecological study in vietnam. *INTERNATIONAL JOURNAL OF EPIDEMIOLOGY*, 38(6):1634–1641, DEC 2009. ISSN 0300-5771. doi: 10.1093/ije/dyp244.
 19. R. Barbieri, B. H. A. Mai, T. Chenal, M-L. Bassi, D. Gandia, L. Camoin-Jau, H. Lepidi, G. Aboudharam, and M. Drancourt. A 2,000-year-old specimen with intraerythrocytic bartonella quintana. *SCIENTIFIC REPORTS*, 10(1), JUN 22 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-66917-7.
 20. Maliya Alia Malek, Idir Bitam, Anthony Levasseur, Jerome Terras, Jean Gaudart, Said Azza, Christophe Flaudrops, Catherine Robert, Didier Raoult, and Michel Drancourt. *Yersinia pestis* halotolerance illuminates plague reservoirs. *SCIENTIFIC REPORTS*, 7, JAN 5 2017. ISSN 2045-2322. doi: 10.1038/srep40022.
 21. Maria Anice M Sallum, Jan E Conn, Eduardo S Bergo, Gabriel Z Laporta, Leonardo SM Chaves, Sara A Bickersmith, Tatiane MP de Oliveira, Elder Augusto G Figueira, Gilberto Moresco, Lêuda Olivêr, et al. Vector competence, vectorial capacity of *nyssorhynchus darlingi* and the basic reproduction number of *plasmodium vivax* in agricultural settlements in the amazonian region of brazil. *Malaria journal*, 18(1):1–15, 2019.

-
22. World Health Organization et al. Who report on global surveillance of epidemic-prone infectious diseases. Technical report, World Health Organization, 2000.
 23. Eric Bertherat, Souad Bekhoucha, Saada Chougrani, Fathia Razik, Jean B Duchemin, Leila Houti, Larbi Deharib, Corinne Fayolle, Banaouda Makrerougrass, Radia Dali-Yahia, et al. Plague reappearance in algeria after 50 years, 2003. *Emerging infectious diseases*, 13(10): 1459, 2007.
 24. Eric Bertherat. Plague around the world, 2010-2015/la peste a travers le monde: 2010-2015. *Weekly epidemiological record*, 91(8):89–94, 2016.
 25. Sitraka Rakotosamimanana, Daouda Kassie, François Taglioni, Josélyne Ramamonjisoa, Fanjasoa Rakotomanana, and Minoarisoa Rajerison. A decade of plague in madagascar: a description of two hotspot districts. *BMC Public Health*, 21(1):1–9, 2021.
 26. Voahangy Andrianaivoarimanana, Katharina Kreppel, Nohal Elissa, Jean-Marc Duplantier, Elisabeth Carniel, Minoarisoa Rajerison, and Ronan Jambou. Understanding the persistence of plague foci in madagascar. *PLOS NEGLECTED TROPICAL DISEASES*, 7(11), NOV 2013. ISSN 1935-2735. doi: 10.1371/journal.pntd.0002382.
 27. JM Duplantier, JB Duchemin, S Chanteau, and E Carniel. From the recent lessons of the malagasy foci towards a global understanding of the factors involved in plague reemergence. *VETERINARY RESEARCH*, 36(3):437–453, MAY-JUN 2005. ISSN 0928-4249. doi: 10.1051/vetres:2005007.
 28. Shinya Tsuzuki, Hyojung Lee, Fuminari Miura, Yat Hin Chan, Sung-mok Jung, Andrei R. Akhmetzhanov, and Hiroshi Nishiura. Dynamics of the pneumonic plague epidemic in madagascar, august to october 2017. *EUROSURVEILLANCE*, 22(46):2–7, NOV 16 2017. ISSN 1560-7917. doi: 10.2807/1560-7917.ES.2017.22.46.17-00710.
 29. Amy J. Vogler, Fabien Chan, Roxanne Nottingham, Genevieve Andersen, Kevin Drees, Stephen M. Beckstrom-Sternberg, David M. Wagner, Suzanne Chanteau, and Paul Keim. A decade of plague in mahajanga, madagascar: Insights into the global maritime spread of pandemic plague. *MBIO*, 4(1), JAN-FEB 2013. ISSN 2150-7511. doi: 10.1128/mBio.00623-12.
 30. Renaud Piarroux, Aaron Aruna Abedi, Jean-Christophe Shako, Benoit Kebela, Stomy

-
- Karhemere, Georges Diatta, Bernard Davoust, Didier Raoult, and Michel Drancourt. Plague epidemics and lice, democratic republic of the congo. *EMERGING INFECTIOUS DISEASES*, 19(3):505–506, MAR 2013. ISSN 1080-6040. doi: 10.3201/eid1903.121542.
31. Eric Bertherat, Melissa J. Mueller, Jean-Christophe Shako, and Mathieu Picardeau. Discovery of a leptospirosis cluster amidst a pneumonic plague outbreak in a miners' camp in the democratic republic of the congo. *INTERNATIONAL JOURNAL OF ENVIRONMENTAL RESEARCH AND PUBLIC HEALTH*, 11(2):1824–1833, FEB 2014. doi: 10.3390/ijerph110201824.
32. Laurel B. Respcio-Kingry, Brook M. Yockey, Sarah Acayo, John Kaggwa, Titus Apangu, Kiersten J. Kugeler, Rebecca J. Eisen, Kevin S. Griffith, Paul S. Mead, Martin E. Schriefer, and Jeannine M. Petersen. Two distinct yersinia pestis populations causing plague among humans in the west Nile region of Uganda. *PLOS NEGLECTED TROPICAL DISEASES*, 10(2), FEB 2016. ISSN 1935-2735. doi: 10.1371/journal.pntd.0004360.
33. Gerald Amatre, Nackson Babi, Russell E. Ensore, Asaph Ogen-Odoi, Linda A. Atiku, Anne Akol, Kenneth L. Gage, and Rebecca J. Eisen. Flea diversity and infestation prevalence on rodents in a plague-endemic region of Uganda. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 81(4):718–724, OCT 2009. ISSN 0002-9637. doi: 10.4269/ajtmh.2009.09-0104.
34. Michael H Ziwa, Mecky I Matee, Bernard M Hang'ombe, Eligius F Lyamuya, and Bukheti S Kilonzo. Plague in Tanzania: an overview. *Tanzania journal of health research*, 15(4):252–8, Oct 2013. ISSN 1821-6404.
35. Anne Laudisoit, Herwig Leirs, Rhodes H. Makundi, Stefan Van Dongen, Stephen Davis, Simon Neerinckx, Jozef Deckers, and Roland Libois. Plague and the human flea, Tanzania. *EMERGING INFECTIOUS DISEASES*, 13(5):687–693, MAY 2007. ISSN 1080-6040. doi: 10.3201/eid1305.061084.
36. Rhodes H. Makundi, Apia W. Massawe, Loth S. Mulungu, Abdul Katakweba, Thomas J. Mbise, and Georgies Mgone. Potential mammalian reservoirs in a bubonic plague outbreak focus in Mbulu district, northern Tanzania, in 2007. *MAMMALIA*, 72(3):253–257, 2008. ISSN 0025-1461. doi: 10.1515/MAMM.2008.038. 10th International African Small Mammals Symposium, Abomey Calavi, BENIN, AUG 20-25, 2007.

-
37. Stanley S. Nyirenda, Bernard M. Hang'ombe, Robert Machang'u, Jackson Mwanza, and Bukheti S. Kilonzo. Identification of risk factors associated with transmission of plague disease in eastern zambia. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 97(3):826–830, 2017. ISSN 0002-9637. doi: 10.4269/ajtmh.16-0990.
 38. Bernard M Hang'ombe, Ichiro Nakamura, Kenny L Samui, Davy Kaile, Aaron S Mweene, Bukheti S Kilonzo, Hirofumi Sawa, Chihiro Sugimoto, and Brendan W Wren. Evidence of yersinia pestis dna from fleas in an endemic plague area of zambia. *BMC research notes*, 5:72, Jan 2012. doi: 10.1186/1756-0500-5-72.
 39. Maliya Alia Malek, Idir Bitam, and Michel Drancourt. Plague in arab maghreb, 1940-2015: A review. *FRONTIERS IN PUBLIC HEALTH*, 4, JUN 3 2016. doi: 10.3389/fpubh.2016.00112.
 40. Idir Bitam, Saravanan Ayyadurai, Tahar Kernif, Mohammed Chetta, Nabil Boulaghman, Didier Raoult, and Michel Drancourt. New rural focus of plague, algeria. *EMERGING INFECTIOUS DISEASES*, 16(10):1639–1640, OCT 2010. ISSN 1080-6040. doi: 10.3201/eid1610.091854.
 41. Nicolas Cabanel, Alexandre Leclercq, Viviane Chenal-Francisque, Badereddin Annajar, Minoarisoa Rajerison, Souad Bekkhoucha, Eric Bertherat, and Elisabeth Carniel. Plague outbreak in libya, 2009, unrelated to plague in algeria. *Emerging Infectious Diseases*, 19 (2):230, 2013.
 42. Xiaona Shen, Jianyun Li, Mengguang Fan, Lianxu Xia, and Wei Li. A reemergent case of bubonic plague-inner mongolia autonomous region, china, july, 2020. *China CDC Weekly*, 2(29):549–550, 2020.
 43. Yumeng Wang, Lei Zhou, Mengguang Fan, Quanyi Wang, Jianyun Li, Qun Li, Zijian Feng, George F Gao, Cheng Xu, Lijuan Chen, et al. Isolated cases of plague—inner mongolia-beijing, 2019. *China CDC weekly*, 1(1):13, 2019.
 44. Yilan Feng, Mengguang Fan, Yulong Gao, Jianyun Li, Dayu Zhang, Shuyi Wang, and Wenrui Wang. Epidemiological features of four human plague cases in the inner mongolia autonomous region, china in 2019. *Biosafety and Health*, 2(1):44–48, 2020.
 45. Hai-Wen Du, Yong Wang, Da-Fang Zhuang, and Xiao-San Jiang. Temporal and spatial

-
- distribution characteristics in the natural plague foci of chinese mongolian gerbils based on spatial autocorrelation. *INFECTIOUS DISEASES OF POVERTY*, 6, AUG 7 2017. ISSN 2095-5162. doi: 10.1186/s40249-017-0338-7.
46. Giovanna Morelli, Yajun Song, Camila J. Mazzoni, Mark Eppinger, Philippe Roumagnac, David M. Wagner, Mirjam Feldkamp, Barica Kusecek, Amy J. Vogler, Yanjun Li, Yujun Cui, Nicholas R. Thomson, Thibaut Jombart, Raphael Leblois, Peter Lichtner, Lila Rahalison, Jeannine M. Petersen, Francois Balloux, Paul Keim, Thierry Wirth, Jacques Ravel, Ruifu Yang, Elisabeth Carniel, and Mark Achtman. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *NATURE GENETICS*, 42(12):1140+, DEC 2010. ISSN 1061-4036. doi: 10.1038/ng.705.
 47. Liyuan Shi, Guirong Yang, Zhikai Zhang, Lianxu Xia, Ying Liang, Hongli Tan, Jinrong He, Jianguo Xu, Zhizhong Song, Wei Li, and Peng Wang. Reemergence of human plague in yunnan, china in 2016. *PLOS ONE*, 13(6), JUN 13 2018. ISSN 1932-6203. doi: 10.1371/journal.pone.0198067.
 48. Ruixia Dai, Baiqing Wei, Haoming Xiong, Xiaoyan Yang, Yao Peng, Jian He, Juan Jin, Yumeng Wang, Xi Zha, Zhikai Zhang, Ying Liang, Qingwen Zhang, Jianguo Xu, Zuyun Wang, and Wei Li. Human plague associated with tibetan sheep originates in marmots. *PLOS NEGLECTED TROPICAL DISEASES*, 12(8), AUG 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006635.
 49. Jan Kehrmann, Walter Popp, Battumur Delgermaa, Damdin Otgonbayar, Tsagaan Gantumur, Jan Buer, and Nyamdorj Tsogbadrakh. Two fatal cases of plague after consumption of raw marmot organs. *Emerging Microbes & Infections*, 9(1):1878–1880, 2020.
 50. Julia M. Riehm, Gilles Vergnaud, Daniel Kiefer, Tserennorov Damdindorj, Otgonbaatar Dashdavaa, Tungalag Khurelsukh, Lothar Zoeller, Roman Woelfel, Philippe Le Fleche, and Holger C. Scholz. *Yersinia pestis* lineages in mongolia. *PLOS ONE*, 7(2), FEB 17 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0030624.
 51. Galina A. Eroshenko, Nikita Yu Nosov, Yaroslav M. Krasnov, Yevgeny G. Oglodin, Lyubov M. Kukleva, Natalia P. Guseva, Alexander A. Kuznetsov, Sabyrzhan T. Abdikarimov, Aigul K. Dzharparova, and Vladimir V. Kutyrev. *Yersinia pestis* strains of ancient phylogenetic branch 0. ant are widely spread in the high-mountain plague foci of kyrgyzstan. *PLOS ONE*, 12(10), OCT 26 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0187230.

-
52. NV Popov, IG Karnaukhov, ND Pakskina, GA Eroshenko, AA Kuznetsov, AN Matrosov, AM Porshakov, EV Kouklev, AV Ivanova, VM Korzun, et al. Analysis of the current epidemiological situation in natural plague foci around the world. enhancement of the effectiveness of epidemiological surveillance in natural plague foci of the russian federation and forecast of their epizootic activity for 2019. *Problems of Particularly Dangerous Infections*, (1): 81–88, 2019.
 53. Joseph J Kingston, Urmil Tuteja, Minakshi Kapil, Harishchandra S Murali, and Harsh V Batra. Genotyping of indian yersinia pestis strains by mlva and repetitive dna sequence based pcrs. *Antonie van Leeuwenhoek*, 96(3):303–312, 2009.
 54. A Ruiz. Plague in the americas. *EMERGING INFECTIOUS DISEASES*, 7(3, S):539–540, 2001. ISSN 1080-6040.
 55. Luis F Donaires, Manuel Céspedes, Pedro Valencia, Juan Carlos Salas, María E Luna, Alex Castañeda, Víctor Peralta, César Cabezas, and Paul E Pachas. Peste neumónica primaria con transmisión intrahospitalaria en la libertad, Perú 2010. *Revista Peruana de Medicina Experimental y Salud Pública*, 27:326–336, 2010.
 56. Maria Cristina Schneider, Patricia Najera, Sylvain Aldighieri, Deise I. Galan, Eric Bertherat, Alfonso Ruiz, Elsy Dumit, Jean Marc Gabastou, and Marcos A. Espinal. Where does human plague still persist in latin america? *PLOS NEGLECTED TROPICAL DISEASES*, 8(2), FEB 2014. ISSN 1935-2735. doi: 10.1371/journal.pntd.0002680.
 57. Kiersten J. Kugeler, J. Erin Staples, Alison F. Hinckley, Kenneth L. Gage, and Paul S. Mead. Epidemiology of human plague in the united states, 1900–2012. *EMERGING INFECTIOUS DISEASES*, 21(1):16–22, JAN 2015. ISSN 1080-6040. doi: 10.3201/eid2101.140564.
 58. Jessica M Girard, David M Wagner, Amy J Vogler, Christine Keys, Christopher J Allender, Lee C Drickamer, and Paul Keim. Differential plague-transmission dynamics determine yersinia pestis population genetic structure on local, regional, and global scales. *Proceedings of the National Academy of Sciences*, 101(22):8408–8413, 2004.
 59. Celso Tavares, Antônia Ivoneida Aragão, Nilma Cintra Leal, Tereza Cristina Leal-Balbino, Maria Betânia Melo de Oliveira, Geane Maria de Oliveira Gonçalves Ferreira, and Alzira Maria Paiva de Almeida. Plague in brazil: from now and then. In *Advances in Yersinia Research*, pages 69–77. Springer, 2012.

-
60. Erika de Cássia Vieira Da Costa, Marise Sobreira, Nilma Cintra Leal, and Alzira Maria Paiva De Almeida. Rodents and other small mammal reservoirs in plague foci in northeastern Brazil. *The Journal of Infection in Developing Countries*, 11(05):426–430, 2017.
 61. Barbara Bramanti, Katharine R. Dean, Lars Walloe, and Nils Chr Stenseth. The third plague pandemic in Europe. *PROCEEDINGS OF THE ROYAL SOCIETY B-BIOLOGICAL SCIENCES*, 286(1901), APR 17 2019. ISSN 0962-8452. doi: 10.1098/rspb.2018.2429.
 62. Vladimir V. Kutyrev, Galina A. Eroshenko, Vladimir L. Motin, Nikita Y. Nosov, Jaroslav M. Krasnov, Lyubov M. Kukleva, Konstantin A. Nikiforov, Zhanna V. Al'khova, Eugene G. Oglodin, and Natalia P. Guseva. Phylogeny and classification of *Yersinia pestis* through the lens of strains from the plague foci of commonwealth of independent states. *FRONTIERS IN MICROBIOLOGY*, 9, MAY 25 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.01106.
 63. GA Eroshenko, NV Popov, Zh V Al'khova, AN Balykova, LM Kukleva, and VV Kutyrev. Phylogenetic analysis of *Yersinia pestis* strains of medieval biovar, isolated in precaspian north-western steppe plague focus in the XX century. *Problems of Particularly Dangerous Infections*, (2):55–61, 2019.
 64. Galina A. Eroshenko, Nikolay V. Popov, Zhanna V. Al'khova, Lyubov M. Kukleva, Alina N. Balykova, Nadezhda S. Chervyakova, Ekaterina A. Naryshkina, and Vladimir V. Kutyrev. Evolution and circulation of *Yersinia pestis* in the northern Caspian and northern Aral Sea regions in the 20(th)-21(st) centuries. *PLOS ONE*, 16(2), FEB 11 2021. ISSN 1932-6203. doi: 10.1371/journal.pone.0244615.
 65. S Chanteau, L Rahalison, M Ratsitorahina, Mahafaly, M Rasolomaharo, P Boisier, T O'Brien, J Aldrich, A Keleher, C Morgan, and J Burans. Early diagnosis of bubonic plague using F1 antigen capture ELISA assay and rapid immunogold dipstick. *INTERNATIONAL JOURNAL OF MEDICAL MICROBIOLOGY*, 290(3):279–283, JUL 2000. ISSN 1438-4221.
 66. Han Wang and Andres Rodriguez-Pose. Local institutions and pandemics: City autonomy and the black death. *APPLIED GEOGRAPHY*, 136, NOV 2021. ISSN 0143-6228. doi: 10.1016/j.apgeog.2021.102582.
 67. Nils Chr. Stenseth, Noelle I. Samia, Hildegunn Viljugrein, Kyrre Linne Kausrud, Mike Begon, Stephen Davis, Herwig Leirs, V. M. Dubyanskiy, Jan Esper, Vladimir S. Ageyev,

-
- Nikolay L. Klassovskiy, Sergey B. Pole, and Kung-Sik Chan. Plague dynamics are driven by climate variation. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 103(35):13110–13115, AUG 29 2006. ISSN 0027-8424. doi: 10.1073/pnas.0602447103.
68. Stephen Davis, Mike Begon, Luc De Bruyn, Vladimir S Ageyev, Nikolay L Klassovskiy, Sergey B Pole, Hildegunn Viljugrein, Nils Chr Stenseth, and Herwig Leirs. Predictive thresholds for plague in kazakhstan. *Science*, 304(5671):736–738, 2004.
 69. Jennifer Alderson, Max Quastel, Emily Wilson, and Duncan Bellamy. Factors influencing the re-emergence of plague in madagascar. *Emerging Topics in Life Sciences*, 4(4):423–433, 2020.
 70. Mandy Elschner, Sally Cutler, Manfred Weidmann, and Patrick Butaye. *BSL3 and BSL4 agents: epidemiology, microbiology and practical guidelines*. John Wiley & Sons, 2012.
 71. Xavier Valles, Nils Chr Stenseth, Christian Demeure, Peter Horby, Paul S. Mead, Oswaldo Cabanillas, Mahery Ratsitorahina, Minoarisoa Rajerison, Voahangy Andrianaivoarimanana, Beza Ramasindrazana, Javier Pizarro-Cerda, Holger C. Scholz, Romain Girod, B. Joseph Hinnebusch, Ines Vigan-Womas, Arnaud Fontanet, David M. Wagner, Sandra Telfer, Yazdan Yazdanpanah, Pablo Tortosa, Guia Carrara, Jane Deuve, Steven R. Belmain, Eric D’Ortenzio, and Laurence Baril. Human plague: An old scourge that needs new answers. *PLOS NEGLECTED TROPICAL DISEASES*, 14(8), AUG 2020. ISSN 1935-2735. doi: 10.1371/journal.pntd.0008251.
 72. Vladimir M Dubyanskiy and Aidyn B Yeszhanov. Ecology of yersinia pestis and the epidemiology of plague. *Yersinia pestis: Retrospective and Perspective*, pages 101–170, 2016.
 73. Robert D Perry and Jacqueline D Fetherston. Yersinia pestis—etiologic agent of plague. *Clinical microbiology reviews*, 10(1):35–66, 1997.
 74. R. Pollitzer. Plague studies*. *Bulletin of the World Health Organization*, 4(4):475 – 533, 1951.
 75. Rachel C Abbott and Tonie E Rocke. Plague. Technical report, US Geological Survey, 2012.
 76. Joseph A. Lewnard and Jeffrey P. Townsend. Climatic and evolutionary drivers of phase

-
- shifts in the plague epidemics of colonial india. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 113(51):14601–14608, DEC 20 2016. ISSN 0027-8424. doi: 10.1073/pnas.1604985113.
77. Voahangy Andrianaivoarimanana, Patrice Piola, David M. Wagner, Fanjasoa Rakotomanana, Viviane Maheriniaina, Samuel Andrianalimanana, Suzanne Chanteau, Lila Ralalison, Maherisoa Ratsitorahina, and Minoarisoa Rajerison. Trends of human plague, madagascar, 1998-2016. *EMERGING INFECTIOUS DISEASES*, 25(2):220–228, FEB 2019. ISSN 1080-6040. doi: 10.3201/eid2502.171974.
78. James H Rust Jr, Dan C Cavanaugh, Roy O'Shita, and John D Marshall Jr. The role of domestic animals in the epidemiology of plague. i. experimental infection of dogs and cats. *Journal of Infectious Diseases*, 124(5):522–526, 1971.
79. Stefanie B. Campbell, Christina A. Nelson, Alison F. Hinckley, and Kiersten J. Kugeler. Animal exposure and human plague, united states, 1970-2017. *EMERGING INFECTIOUS DISEASES*, 25(12):2270–2273, DEC 2019. ISSN 1080-6040. doi: 10.3201/eid2512.191081.
80. RD Perry and JD Fetherston. *Yersinia pestis* - etiologic agent of plague. *CLINICAL MICROBIOLOGY REVIEWS*, 10(1):35+, JAN 1997. ISSN 0893-8512. doi: 10.1128/CMR.10.1.35.
81. Peter W Gasper, Allan M Barnes, Thomas J Quan, Jacqueline P Benziger, Leon G Carter, Mala L Beard, and Gary O Maupin. Plague (*Yersinia pestis*) in cats: description of experimentally induced disease. *Journal of Medical Entomology*, 30(1):20–26, 1993.
82. JF Cully, AM Barnes, TJ Quan, and G Maupin. Dynamics of plague in a Gunnison's prairie dog colony complex from New Mexico. *JOURNAL OF WILDLIFE DISEASES*, 33(4):706–719, OCT 1997. ISSN 0090-3558. doi: 10.7589/0090-3558-33.4.706.
83. Jonathan N. Pauli, Steven W. Buskirk, Elizabeth S. Williams, and William H. Edwards. A plague epizootic in the black-tailed prairie dog (*Cynomys ludovicianus*). *JOURNAL OF WILDLIFE DISEASES*, 42(1):74–80, JAN 2006. ISSN 0090-3558. doi: 10.7589/0090-3558-42.1.74.
84. RE THOMAS, ML BEARD, TJ QUAN, LG CARTER, AM BARNES, and CE HOPLA. Experimentally induced plague infection in the northern grasshopper mouse (*Onychomys*

-
- leucogaster) acquired by consumption of infected prey. *JOURNAL OF WILDLIFE DISEASES*, 25(4):477–480, OCT 1989. ISSN 0090-3558. doi: 10.7589/0090-3558-25.4.477.
85. V N FEDOROV. Plague in camels and its prevention in the ussr. *Bulletin of the World Health Organization*, 23:275–81, 1960. ISSN 0042-9686.
86. AA Bin Saeed, NA Al-Hamdan, and RE Fontaine. Plague from eating raw camel liver. *EMERGING INFECTIOUS DISEASES*, 11(9):1456–1457, SEP 2005. ISSN 1080-6040. doi: 10.3201/eid1109.050081.
87. AP Anisimov, LE Lindler, and GB Pier. Intraspecific diversity of yersinia pestis. *CLINICAL MICROBIOLOGY REVIEWS*, 17(2):434+, APR 2004. ISSN 0893-8512. doi: 10.1128/CMR.17.2.434-464.2004.
88. BJ Hinnebusch. The evolution of flea-borne transmission in yersinia pestis. *CURRENT ISSUES IN MOLECULAR BIOLOGY*, 7:197–212, 2005. ISSN 1467-3037.
89. Angelina A Kislichkina, Aleksandr G Bogun, Lidiya A Kadnikova, Nadezhda V Maiskaya, Mikhail E Platonov, Nikolai V Anisimov, Elena V Galkina, Svetlana V Dentovskaya, and Andrey P Anisimov. Nineteen whole-genome assemblies of yersinia pestis subsp. microtus, including representatives of biovars caucasica, talassica, hissarica, altaica, xilingolensis, and ulegeica. *Genome announcements*, 3(6):e01342–15, 2015.
90. BJ Hinnebusch and DL Erickson. Yersinia pestis biofilm in the flea vector and its role in the transmission of plague. *Bacterial biofilms*, pages 229–248, 2008.
91. Paul-Louis Simond. La propagation de la peste. *Annales de l'Institut Pasteur*, 1898.
92. Kenneth L Gage and Michael Y Kosoy. Natural history of plague: Perspectives from. *Annu. Rev. Entomol*, 50:505–28, 2005.
93. PA Petrov, AI Goncharov, NF Labunets, MA Akhundov, PI Osyko, et al. The duration of the life and migration of fleas of the red-tailed gerbil in the reservoir of plague on the transcaucasian plains. *Zoologicheskii Zhurnal*, 47(9):1354–1358, 1968.
94. Rebecca J. Eisen, Scott W. Bearden, Aryn P. Wilder, John A. Monteneri, Michael F. Antolin, and Kenneth L. Gage. Early-phase transmission of yersinia pestis by unblocked fleas as a mechanism explaining rapidly spreading plague epizootics. *PROCEEDINGS OF*

THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA, 103(42):15380–15385, OCT 17 2006. ISSN 0027-8424. doi: 10.1073/pnas.0606831103.

95. Rebecca J. Eisen, Jennifer L. Holmes, Anna M. Schotthoefer, Sara M. Vetter, John A. Monteneri, and Kenneth L. Gage. Demonstration of early-phase transmission of yersinia pestis by the mouse flea, *aetheca wagneri* (siphonaptera: Ceratophyllidae), and implications for the role of deer mice as enzootic reservoirs. *JOURNAL OF MEDICAL ENTOMOLOGY*, 45(6):1160–1164, NOV 2008. ISSN 0022-2585. doi: 10.1603/0022-2585(2008)45{[]1160:DOETOY}2.0.CO;2.
96. AL BURROUGHS. Sylvatic plague studies - the vector efficiency of 9 species of fleas compared with *xenopsylla-cheopis*. *JOURNAL OF HYGIENE*, 45(3):371–&, 1947. ISSN 0950-2688. doi: 10.1017/S0022172400014042.
97. Rebecca J. Eisen, Pamela J. Reynolds, Paul Ettestad, Ted Brown, Russell E. Enscoe, Brad J. Biggerstaff, James Cheek, Rudy Bueno, Joseph Targhetta, John A. Monteneri, and Kenneth L. Gage. Residence-linked human plague in new mexico: A habitat-suitability model. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 77(1):121–125, JUL 2007. ISSN 0002-9637. doi: 10.4269/ajtmh.2007.77.121.
98. Aryn P. Wilder, Rebecca J. Eisen, Scott W. Bearden, John A. Monteneri, Daniel W. Tripp, R. Jory Brinkerhoff, Kenneth L. Gage, and Michael F. Antolin. Transmission efficiency of two flea species (*ropsylla tuberculata cynomuris* and *ropsylla hirsuta*) involved in plague epizootics among prairie dogs. *ECOHEALTH*, 5(2):205–212, JUN 2008. ISSN 1612-9202. doi: 10.1007/s10393-008-0165-1.
99. DM Engelthaler and KL Gage. Quantities of yersinia pestis in fleas (siphonaptera : Pulicidae, ceratophyllidae, and hystrichopsyllidae) collected from areas of known or suspected plague activity. *JOURNAL OF MEDICAL ENTOMOLOGY*, 37(3):422–426, MAY 2000. ISSN 0022-2585. doi: 10.1603/0022-2585(2000)037{[]0422:QOYPIF}2.0.CO;2.
100. EA Lorange, BL Race, F Sebbane, and BJ Hinnebusch. Poor vector competence of fleas and the evolution of hypervirulence in yersinia pestis. *JOURNAL OF INFECTIOUS DISEASES*, 191(11):1907–1912, JUN 1 2005. ISSN 0022-1899. doi: 10.1086/429931.
101. A Yersin. Sur la peste de hongkong (lundi 30 juillet 1894). *CR Hebd Séances Acad Sci*, 119:356, 1894.

-
102. MOLLARET HH. Experimental preservation of plague in soil. *Bulletin de la Societe de Pathologie Exotique et de ses Filiales*, 56:1168–1182, 1963.
103. Saravanan Ayyadurai, Linda Houhamdi, Hubert Lepidi, Claude Nappes, Didier Raoult, and Michel Drancourt. Long-term persistence of virulent yersinia pestis in soil. *MICROBIOLOGY-SGM*, 154(9):2865–2871, SEP 2008. ISSN 1350-0872. doi: 10.1099/mic.0.2007/016154-0.
104. Y. KARIML. *BULL SOC PATHOL EXOTIQUE*, 56((6)):1183–1186, 1963.
105. I V Domaradskii. *Meditssinskaia parazitologiya i parazitarnye bolezni*, (2):10–3, 1999. ISSN 0025-8326.
106. Javier A. Benavides-Montano and Viveka Vadyvaloo. Yersinia pestis resists predation by acanthamoeba castellanii and exhibits prolonged intracellular survival. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 83(13), JUL 2017. ISSN 0099-2240. doi: 10.1128/AEM.00593-17.
107. David W. Markman, Michael F. Antolin, Richard A. Bowen, William H. Wheat, Michael Woods, Mercedes Gonzalez-Juarrero, and Mary Jackson. Yersinia pestis survival and replication in potential ameba reservoir. *EMERGING INFECTIOUS DISEASES*, 24(2):294–302, FEB 2018. ISSN 1080-6040. doi: 10.3201/eid2402.171065.
108. Lei Xu, Leif C. Stige, Herwig Leirs, Simon Neerinckx, Kenneth L. Gage, Ruifu Yang, Qiyong Liu, Barbara Bramanti, Katharine R. Dean, Hui Tang, Zhe Sun, Nils Chr. Stenseth, and Zhibin Zhang. Historical and genomic data reveal the influencing factors on global transmission velocity of plague during the third pandemic. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 116(24):11833–11838, JUN 11 2019. ISSN 0027-8424. doi: 10.1073/pnas.1901366116.
109. Susan D Jones, Bakyt Atshabar, Boris V Schmid, Marlene Zuk, Anna Amramina, and Nils Chr Stenseth. Living with plague: Lessons from the soviet union's antiplague system. *Proceedings of the National Academy of Sciences*, 116(19):9155–9163, 2019.
110. A Arbaji, S Kharabsheh, S Al-Azab, M Al-Kayed, ZS Amr, M Abu Baker, and MC Chu. A 12-case outbreak of pharyngeal plague following the consumption of camel meat, in north-eastern jordan. *ANNALS OF TROPICAL MEDICINE AND PARASITOLOGY*, 99

-
- (8):789–793, DEC 2005. ISSN 0003-4983. doi: 10.1179/136485905X65161.
111. JM Gabastou, J Proano, A Vimos, G Jaramillo, E Hayes, K Gage, M Chu, J Guarner, S Zaki, J Bowers, C Guillemard, H Tamayo, and A Ruiz. An outbreak of plague including cases with probable pneumonic infection, ecuador, 1998. *TRANSACTIONS OF THE ROYAL SOCIETY OF TROPICAL MEDICINE AND HYGIENE*, 94(4):387–391, JUL-AUG 2000. ISSN 0035-9203. doi: 10.1016/S0035-9203(00)90114-7.
112. Janine K. Runfola, Jennifer House, Lisa Miller, Leah Colton, Donna Hite, Alex Hawley, Paul Mead, Martin Schriefer, Jeannine Petersen, Colleen Casaceli, Kristine M. Erlandson, Clayton Foster, Kristy L. Pabilonia, Gary Mason, and John M. Douglas, Jr. Untitled. *MMWR-MORBIDITY AND MORTALITY WEEKLY REPORT*, 64(16):429–434, MAY 1 2015. ISSN 0149-2195.
113. David Wong, Margaret A. Wild, Matthew A. Walburger, Charles L. Higgins, Michael Callahan, Lawrence A. Czarnecki, Elisabeth W. Lawaczeck, Craig E. Levy, J. Gage Patterson, Rebecca Sunenshine, Patricia Adem, Christopher D. Paddock, Sherif R. Zaki, Jeannine M. Petersen, Martin E. Schriefer, Rebecca J. Eisen, Kenneth L. Gage, Kevin S. Griffith, Ingrid B. Weber, Terry R. Spraker, and Paul S. Mead. Primary pneumonic plague contracted from a mountain lion carcass. *CLINICAL INFECTIOUS DISEASES*, 49(3):E33–E38, AUG 1 2009. ISSN 1058-4838. doi: 10.1086/600818.
114. CF VONREYN, AM BARNES, NS WEBER, T QUAN, and WJ DEAN. Bubonic plague from direct exposure to a naturally infected wild coyote. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 25(4):626–629, 1976. ISSN 0002-9637. doi: 10.4269/ajtmh.1976.25.626.
115. AB CHRISTIE, TH CHEN, and SS ELBERG. Plague in camels and goats - their role in human epidemics. *JOURNAL OF INFECTIOUS DISEASES*, 141(6):724–726, 1980. ISSN 0022-1899. doi: 10.1093/infdis/141.6.724.
116. G BLANC. *Revue d'hygiene et de medecine sociale*, 4(6):535–62, Sep 1956. ISSN 0484-8454.
117. Linda Houhamdi, Hubert Lepidi, Michel Drancourt, and Didier Raoult. Experimental model to evaluate the human body louse as a vector of plague. *JOURNAL OF INFECTIOUS DISEASES*, 194(11):1589–1596, DEC 1 2006. ISSN 0022-1899. doi: 10.1086/508995.

-
118. M BALTAZARD. [decline and fate of an infectious disease: the plague]. *Bulletin of the World Health Organization*, 23:247–62, 1960. ISSN 0042-9686.
119. G Blanc, M Baltazard, et al. Plague infection of *p. irritans*. *Compte Rendu de l'Academie des Sciences*, 213(22):813–16, 1941.
120. Katharine R. Dean, Fabienne Krauer, Lars Walloe, Ole Christian Lingjaerde, Barbara Bramanti, Nils Chr. Stenseth, and Boris V. Schmid. Human ectoparasites and the spread of plague in europe during the second pandemic. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 115(6):1304–1309, FEB 6 2018. ISSN 0027-8424. doi: 10.1073/pnas.1715640115.
121. Roger D. Pechous, Vijay Sivaraman, Nikolas M. Stasulli, and William E. Goldman. Pneumonic plague: The darker side of *yersinia pestis*. *TRENDS IN MICROBIOLOGY*, 24(3): 190–197, MAR 2016. ISSN 0966-842X. doi: 10.1016/j.tim.2015.11.008.
122. LJ Rose, R Donlan, SN Banerjee, and MJ Arduino. Survival of *yersinia pestis* on environmental surfaces. *APPLIED AND ENVIRONMENTAL MICROBIOLOGY*, 69(4):2166–2171, APR 2003. ISSN 0099-2240. doi: 10.1128/AEM.69.4.2166-2171.2003.
123. EM Begier, G Asiki, Z Anywaine, B Yockey, ME Schriefer, P Aleti, A Ogen-Odoi, JE Staples, C Sexton, SW Bearden, and JL Kool. Pneumonic plague cluster, uganda, 2004. *EMERGING INFECTIOUS DISEASES*, 12(3):460–467, MAR 2006. ISSN 1080-6040. doi: 10.3201/eid1203.051051.
124. Hiroshi Nishiura. Epidemiology of a primary pneumonic plague in kantoshu, manchuria, from 1910 to 1911: statistical analysis of individual records collected by the japanese empire. *INTERNATIONAL JOURNAL OF EPIDEMIOLOGY*, 35(4):1059–1065, AUG 2006. ISSN 0300-5771. doi: 10.1093/ije/dyl091.
125. Mark Gamsa. The epidemic of pneumonic plague in manchuria 1910–1911. *Past & Present*, 190(1):147–183, 2006.
126. R Gani and S Leach. Epidemiologic determinants for modeling pneumonic plague outbreaks. *EMERGING INFECTIOUS DISEASES*, 10(4):608–614, APR 2004. ISSN 1080-6040. doi: 10.3201/eid1004.030509.
127. Katharine Rose Dean. Modeling plague transmission in medieval european cities. Master's

thesis, 2015.

128. Vladimir V. Nikiforov, He Gao, Lei Zhou, and Andrey Anisimov. Plague: Clinics, diagnosis and treatment. In R Yang and A Anisimov, editors, *YERSINIA PESTIS: RETROSPECTIVE AND PERSPECTIVE*, volume 918 of *Advances in Experimental Medicine and Biology*, pages 293–312. 2016. ISBN 978-94-024-0890-4; 978-94-024-0888-1. doi: 10.1007/978-94-024-0890-4_11.
129. Thomas Butler. Plague into the 21st century. *CLINICAL INFECTIOUS DISEASES*, 49(5): 736–742, SEP 1 2009. ISSN 1058-4838. doi: 10.1086/604718.
130. Ali A. Rabaan, Shamsah H. Al-Ahmed, Shahab A. Alsuliman, Fatimah A. Aldrazi, Wadha A. Alfouzan, and Shafiul Haque. The rise of pneumonic plague in madagascar: current plague outbreak breaks usual seasonal mould. *JOURNAL OF MEDICAL MICROBIOLOGY*, 68 (3):292–302, MAR 2019. ISSN 0022-2615. doi: 10.1099/jmm.0.000915.
131. Christian E Demeure, Olivier Dussurget, Guillem Mas Fiol, Anne-Sophie Le Guern, Cyril Savin, and Javier Pizarro-Cerdá. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes & Immunity*, 20(5):357–370, 2019.
132. Sara W. Montminy, Naseema Khan, Sara McGrath, Mitchell J. Walkowicz, Fiona Sharp, Joseph E. Conlon, Koichi Fukase, Shoichi Kusumoto, Charles Sweet, Kensuke Miyake, Shizuo Akira, Robert J. Cotter, Jon D. Goguen, and Egil Lien. Virulence factors of *Yersinia pestis* are overcome by a strong lipopolysaccharide response. *NATURE IMMUNOLOGY*, 7 (10):1066–1073, OCT 2006. ISSN 1529-2908. doi: 10.1038/ni1386.
133. Celine Pujol, Kathryn A. Klein, Galina A. Romanov, Lance E. Palmer, Carol Ciota, Zijiang Zhao, and James B. Bliska. *Yersinia pestis* can reside in autophagosomes and avoid xenophagy in murine macrophages by preventing vacuole acidification. *INFECTION AND IMMUNITY*, 77(6):2251–2261, JUN 2009. ISSN 0019-9567. doi: 10.1128/IAI.00068-09.
134. Michael G. Connor, Amanda R. Pulsifer, Christopher T. Price, Yousef Abu Kwaik, and Matthew B. Lawrenz. *Yersinia pestis* requires host rab1b for survival in macrophages. *PLOS PATHOGENS*, 11(10), OCT 2015. ISSN 1553-7366. doi: 10.1371/journal.ppat.1005241.
135. Organisation mondiale de la Santé, World Health Organization, et al. Revision of the

-
- international definition of plague cases—révision de la définition internationale des cas de peste. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire*, 96(24): 238–240, 2021.
136. S Chanteau, L Rahalison, L Ralafiarisoa, J Foulon, M Ratsitorahina, L Ratsifasoamanana, E Carniel, and A Nato. Development and testing of a rapid diagnostic test for bubonic and pneumonic plague. *LANCET*, 361(9353):211–216, JAN 18 2003. ISSN 0140-6736. doi: 10.1016/S0140-6736(03)12270-2.
 137. Julia M. Riehm, Lila Rahalison, Holger C. Scholz, Bryan Thoma, Martin Pfeffer, Lea Mamiharisoa Razanakoto, Sascha Al Dahouk, Heinrich Neubauer, and Herbert Tomaso. Detection of yersinia pestis using real-time pcr in patients with suspected bubonic plague. *MOLECULAR AND CELLULAR PROBES*, 25(1):8–12, FEB 2011. ISSN 0890-8508. doi: 10.1016/j.mcp.2010.09.002.
 138. Nilma C Leal and Alzira MP de ALMEIDA. Diagnosis of plague and identification of virulence markers in yersinia pestis by multiplex-pcr. *Revista do Instituto de Medicina Tropical de São Paulo*, 41:339–342, 1999.
 139. Herbert Tomaso, Emil C Reisinger, Sascha Al Dahouk, Dimitrios Frangoulidis, Alexander Rakin, Olfert Landt, and Heinrich Neubauer. Rapid detection of yersinia pestis with multiplex real-time pcr assays using fluorescent hybridisation probes. *FEMS Immunology & Medical Microbiology*, 38(2):117–126, 2003.
 140. Fabrice Armougom, Idir Bitam, Olivier Croce, Vicky Merhej, Lina Barassi, Ti-Thien Nguyen, Bernard La Scola, and Didier Raoult. Genomic insights into a new citrobacter koseri strain revealed gene exchanges with the virulence-associated yersinia pestis ppcp1 plasmid. *FRONTIERS IN MICROBIOLOGY*, 7, MAR 16 2016. doi: 10.3389/fmicb.2016.00340.
 141. Stephanie Hansch, Elisabetta Cilli, Giulio Catalano, Giorgio Gruppioni, Raffaella Bianucci, Nils C Stenseth, Barbara Bramanti, and Mark J Pallen. The pla gene, encoding plasminogen activator, is not specific to yersinia pestis. *BMC research notes*, 8:535, Oct 2015. doi: 10.1186/s13104-015-1525-x.
 142. Markos Molsa, Heidi Hemmila, Anna Katz, Jukka Niemimaa, Kristian M. Forbes, Otso Huitu, Peter Stuart, Heikki Henttonen, and Simo Nikkari. Monitoring biothreat agents (francisella tularensis, bacillus anthracis and yersinia pestis) with a portable real-time pcr

-
- instrument. *JOURNAL OF MICROBIOLOGICAL METHODS*, 115:89–93, AUG 2015. ISSN 0167-7012. doi: 10.1016/j.mimet.2015.05.026.
143. Lovasoa N Randriantseheno, Anjanirina Rahantamalala, Ando L Randrianierenana, Minoarisoa Rajerison, and Voahangy Andrianaivoarimanana. Development and evaluation of loop-mediated isothermal amplification for detection of yersinia pestis in plague biological samples. *PloS one*, 15(8):e0237655, 2020.
144. T BUTLER, YS FU, L FURMAN, C ALMEIDA, and A ALMEIDA. Experimental yersinia-pestis infection in rodents after intra-gastric inoculation and ingestion of bacteria. *INFECTION AND IMMUNITY*, 36(3):1160–1167, 1982. ISSN 0019-9567. doi: 10.1128/IAI.36.3.1160-1167.1982.
145. LL Boulanger, P Ettestad, JD Fogarty, DT Dennis, D Romig, and G Mertz. Gentamicin and tetracyclines for the treatment of human plague: Review of 75 cases in new mexico, 1985-1999. *CLINICAL INFECTIOUS DISEASES*, 38(5):663–669, MAR 1 2004. ISSN 1058-4838. doi: 10.1086/381545.
146. W Mwengee, T Butler, S Mgema, G Mhina, Y Almasi, C Bradley, JB Formanik, and CG Rochester. Treatment of plague with gentamicin or doxycycline in a randomized clinical trial in tanzania. *CLINICAL INFECTIOUS DISEASES*, 42(5):614–621, MAR 1 2006. ISSN 1058-4838. doi: 10.1086/500137.
147. Natalie Kwit, Christina Nelson, Kiersten Kugeler, Jeannine Petersen, Lydia Plante, Hayley Yaglom, Vicki Kramer, Benjamin Schwartz, Jennifer House, Leah Colton, Amanda Feldpausch, Cherie Drenzek, Joan Baumbach, Mark DiMenna, Emily Fisher, Emilio Debess, Danielle Buttke, Matthew Weinburke, Christopher Percy, Martin Schriefer, Ken Gage, and Paul Mead. Human plague - united states, 2015. *MMWR-MORBIDITY AND MORTALITY WEEKLY REPORT*, 64(33):918–919, AUG 28 2015. ISSN 0149-2195. doi: 10.15585/mmwr.mm6433a6.
148. Titus Apangu, Kevin Griffith, Janet Abaru, Gordian Candini, Harriet Apio, Felix Okoth, Robert Okello, John Kaggwa, Sarah Acayo, Geoffrey Ezama, Brook Yockey, Christopher Sexton, Martin Schriefer, Edward Katongole Mbidde, and Paul Mead. Successful treatment of human plague with oral ciprofloxacin. *EMERGING INFECTIOUS DISEASES*, 23(3):553–555, MAR 2017. ISSN 1080-6040. doi: 10.3201/eid2303.161212.

-
149. Christina A Nelson, Dana Meaney-Delman, Shannon Fleck-Derderian, Katharine M Cooley, A Yu Patricia, and Paul S Mead. Antimicrobial treatment and prophylaxis of plague: recommendations for naturally acquired infections and bioterrorism response. *MMWR Recommendations and Reports*, 70(3):1, 2021.
 150. Sophie Jullien, Paul Garner, and Martha Chaplin. Who guidelines for plague management: revised recommendations for the use of rapid diagnostic tests, fluoroquinolones for case management and personal protective equipment for prevention of post-mortem transmission. 2021.
 151. David M Wagner, Janelle Runberg, Amy J Vogler, Judy Lee, Elizabeth Driebe, Lance B Price, David M Engelthaler, W Florian Fricke, Jacques Ravel, and Paul Keim. No resistance plasmid in yersinia pestis, north america. *Emerging infectious diseases*, 16(5):885, 2010.
 152. Sandra K Urich, Linda Chalcraft, Martin E Schriefer, Brook M Yockey, and Jeannine M Petersen. Lack of antimicrobial resistance in yersinia pestis isolates from 17 countries in the americas, africa, and asia. *Antimicrobial agents and chemotherapy*, 56(1):555–558, 2012.
 153. Michael J Bottery, Jonathan W Pitchford, and Ville-Petri Friman. Ecology and evolution of antimicrobial resistance in bacterial communities. *The ISME Journal*, 15(4):939–948, 2021.
 154. M Galimand, A Guiyoule, G Gerbaud, B Rasoamanana, S Chanteau, E Carniel, and P Courvalin. Multidrug resistance in yersinia pestis mediated by a transferable plasmid. *NEW ENGLAND JOURNAL OF MEDICINE*, 337(10):677–680, SEP 4 1997. ISSN 0028-4793. doi: 10.1056/NEJM199709043371004.
 155. Nicolas Cabanel, Christiane Bouchier, Minoarisoa Rajerison, and Elisabeth Carniel. Plasmid-mediated doxycycline resistance in a yersinia pestis strain isolated from a rat. *International Journal of Antimicrobial Agents*, 51(2):249–254, 2018.
 156. A Guiyoule, G Gerbaud, C Buchrieser, M Galimand, L Rahalison, S Chanteau, P Courvalin, and E Carniel. Transferable plasmid-mediated resistance to streptomycin in a clinical isolate of yersinia pestis. *EMERGING INFECTIOUS DISEASES*, 7(1):43–48, JAN-FEB 2001. ISSN 1080-6040. doi: 10.3201/eid0701.010106.

-
157. Ruixia Dai, Jian He, Xi Zha, Yiting Wang, Xuefei Zhang, He Gao, Xiaoyan Yang, Juan Li, Youquan Xin, Yumeng Wang, et al. A novel mechanism of streptomycin resistance in yersinia pestis: Mutation in the rpsl gene. *PLoS neglected tropical diseases*, 15(4): e0009324, 2021.
158. Daniel Kiefer, Ganbold Dalantai, Tserennorov Damdindorj, Julia M. Riehm, Herbert Tomaso, Lothar Zoeller, Otgonbaatar Dashdavaa, Kurt Pfister, and Holger C. Scholz. Phenotypical characterization of mongolian yersinia pestis strains. *VECTOR-BORNE AND ZOOLOGICAL DISEASES*, 12(3):183–188, MAR 2012. ISSN 1530-3667. doi: 10.1089/vbz.2011.0748.
159. Wei Sun, Shilpa Sanapala, Jeremy C. Henderson, Shandiin Sam, Joseph Olinzock, Stephen Trent, and Roy Curtiss, III. Lcrv delivered via type iii secretion system of live attenuated yersinia pseudotuberculosis enhances immunogenicity against pneumonic plague. *INFECTION AND IMMUNITY*, 82(10):4390–4404, OCT 2014. ISSN 0019-9567. doi: 10.1128/IAI.02173-14.
160. KF MEYER. Effectiveness of live or killed plague vaccines in man. *BULLETIN OF THE WORLD HEALTH ORGANIZATION*, 42(5):653–&, 1970. ISSN 0042-9686.
161. E. D. Williamson and P. C. F. Oyston. The natural history and incidence of yersinia pestis and prospects for vaccination. *JOURNAL OF MEDICAL MICROBIOLOGY*, 61(7): 911–918, JUL 2012. ISSN 0022-2615. doi: 10.1099/jmm.0.037960-0.
162. Maria J. Chiuchiolo, Julie L. Boyer, Anja Krause, Svetlana Senina, Neil R. Hackett, and Ronald G. Crystal. Protective immunity against respiratory tract challenge with yersinia pestis in mice immunized with an adenovirus-based vaccine vector expressing v antigen. *JOURNAL OF INFECTIOUS DISEASES*, 194(9):1249–1257, NOV 1 2006. ISSN 0022-1899. doi: 10.1086/507644.
163. JE Osorio, TD Powell, RS Frank, K Moss, EJ Haanes, SR Smith, TE Rocke, and DT Stinchcomb. Recombinant raccoon pox vaccine protects mice against lethal plague. *VACCINE*, 21(11-12):1232–1238, MAR 7 2003. ISSN 0264-410X. doi: 10.1016/S0264-410X(02)00557-1.
164. Christina D’Arco, Alison A. McCormick, and Paul M. Arnaboldi. Single-dose intranasal subunit vaccine rapidly clears secondary sepsis in a high-dose pneumonic plague infection.

-
- VACCINE*, 39(9):1435–1444, MAR 1 2021. ISSN 0264-410X. doi: 10.1016/j.vaccine.2021.01.040.
165. Ankit Gupta, Bineet Narayan, Subodh Kumar, and Shailendra Kumar Verma. Vaccine potential of a recombinant bivalent fusion protein lcrv-hsp70 against plague and yersiniosis. *Frontiers in immunology*, 11:988, 2020.
 166. Paul B. Kilgore, Jian Sha, Emily K. Hendrix, Vladimir L. Motin, and Ashok K. Chopra. Combinatorial viral vector-based and live attenuated vaccines without an adjuvant to generate broader immune responses to effectively combat pneumonic plague. *MBIO*, 12(6), NOV-DEC 2021. ISSN 2150-7511. doi: 10.1128/mBio.03223-21.
 167. Fahreta Hamzabegovic, Johannes B Goll, William F Hooper, Sharon Frey, Casey E Gelber, and Getahun Abate. Flagellin adjuvanted f1/v subunit plague vaccine induces t cell and functional antibody responses with unique gene signatures. *NPJ vaccines*, 5(1):1–10, 2020.
 168. Anne Derbise, Yuri Hanada, Manal Khalife, Elisabeth Carniel, and Christian E. Demeure. Complete protection against pneumonic and bubonic plague after a single oral vaccination. *PLOS NEGLECTED TROPICAL DISEASES*, 9(10), OCT 2015. ISSN 1935-2735. doi: 10.1371/journal.pntd.0004162.
 169. Anne Derbise, Chloé Guillas, Christiane Gerke, Elisabeth Carniel, Javier Pizarro-Cerdà, and Christian E Demeure. Subcutaneous vaccination with a live attenuated yersinia pseudotuberculosis plague vaccine. *Vaccine*, 38(8):1888–1892, 2020.
 170. Robert R Brubaker. The genus yersinia: Biochemistry and genetics of virulence with 3 figures. *Current topics in microbiology and immunology*, pages 111–158, 1972.
 171. AM Barnes and TJ Quan. Plague. *Infectious Diseases. Philadelphia, Pa: WB Saunders Co*, pages 1285–1291, 1992.
 172. Robert R Brubaker. Factors promoting acute and chronic diseases caused by yersiniae. *Clinical microbiology reviews*, 4(3):309–324, 1991.
 173. Mobolaji Adeolu, Seema Alnajar, Sohail Naushad, and Radhey S. Gupta. Genome-based phylogeny and taxonomy of the ‘enterobacteriales’: proposal for enterobacterales ord. nov divided into the families enterobacteriaceae, erwiniaceae fam. nov., pectobacteriaceae fam. nov., yersiniaceae fam. nov., hafniaceae fam. nov., morganellaceae fam. nov., and

-
- budviciaceae fam. nov. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, 66(12):5575–5599, DEC 2016. ISSN 1466-5026. doi: 10.1099/ijsem.0.001485.
174. JA Hammerl, N Vom Ort, A Barac, C Jäckel, L Grund, S Dreyer, C Heydel, A Kuczka, H Peters, and S Hertwig. Analysis of yersinia pseudotuberculosis isolates recovered from deceased mammals of a german zoo animal collection. *Journal of Clinical Microbiology*, 59(6):e03125–20, 2021.
175. European Center for Disease Control. Yersiniosis - annual epidemiological report for 2019. Technical report, eCDC, 2021.
176. Cyril Savin, Liliane Martin, Christiane Bouchier, Sofia Filali, Jerome Chenau, Zhemin Zhou, Francois Becher, Hiroshi Fukushima, Nicholas R. Thomson, Holger C. Scholz, and Elisabeth Carniel. The yersinia pseudotuberculosis complex: Characterization and delineation of a new species, yersinia wautersii. *INTERNATIONAL JOURNAL OF MEDICAL MICROBIOLOGY*, 304(3-4):452–463, MAY 2014. ISSN 1438-4221. doi: 10.1016/j.ijmm.2014.02.002.
177. Philip Slavin. Death by the lake: Mortality crisis in early fourteenth-century central asia. *JOURNAL OF INTERDISCIPLINARY HISTORY*, 50(1, SI):59–90, MAY 2019. ISSN 0022-1953. doi: 10.1162/jinh\._a\._01376.
178. Gokhlesh Kumar, Simon Menanteau-Ledouble, Mona Saleh, and Mansour El-Matbouli. Yersinia ruckeri, the causative agent of enteric redmouth disease in fish. *Veterinary research*, 46(1):1–10, 2015.
179. Mark RH Hurst, Amy K Beattie, Sandra A Jones, Pei-Chun Hsu, Joanne Calder, and Chikako van Koten. Temperature-dependent galleria mellonella mortality as a result of yersinia entomophaga infection. *Applied and environmental microbiology*, 81(18):6404–6414, 2015.
180. Cyril Savin, Alexis Criscuolo, Julien Guglielmini, Anne-Sophie Le Guern, Elisabeth Carniel, Javier Pizarro-Cerdá, and Sylvain Brisse. Genus-wide yersinia core-genome multilocus sequence typing for species identification and strain characterization. *Microbial genomics*, 5(10), 2019.
181. Anne-Sophie Le Guern, Cyril Savin, Hilde Angermeier, Sylvie Brémont, Dominique Cler-

-
- mont, Estelle Mühle, Petya Orozova, Hristo Najdenski, and Javier Pizarro-Cerdá. *Yersinia artesiana* sp. nov., *Yersinia proxima* sp. nov., *Yersinia alsatica* sp. nov., *Yersinia vastinensis* sp. nov., *Yersinia thracica* sp. nov. and *Yersinia occitanica* sp. nov., isolated from humans and animals. *International journal of systematic and evolutionary microbiology*, 70(10): 5363, 2020.
182. Brendan W Wren. The yersiniae—a model genus to study the rapid evolution of bacterial pathogens. *Nature Reviews Microbiology*, 1(1):55–64, 2003.
183. Elisabeth Carniel. Evolution of pathogenic yersinia, some lights in the dark. *The Genus Yersinia*, pages 3–12, 2004.
184. C-J Lian, WS Hwang, JK Kelly, and CH Pai. Invasiveness of yersinia enterocolitica lacking the virulence plasmid: an in-vivo study. *Journal of medical microbiology*, 24(3):219–226, 1987.
185. Yvonne Schmid, Guntram A Grassl, Oliver T Buhler, Mikael Skurnik, Ingo B Autenrieth, and Erwin Bohn. *Yersinia enterocolitica* adhesin a induces production of interleukin-8 in epithelial cells. *Infection and immunity*, 72(12):6780–6789, 2004.
186. Soren Schubert, Bertrand Picard, Stéphanie Gouriou, Jurgen Heesemann, and Erick Denamur. *Yersinia* high-pathogenicity island contributes to virulence in *Escherichia coli* causing extraintestinal infections. *Infection and immunity*, 70(9):5335–5337, 2002.
187. Elisabeth Carniel, Ingrid Guilvout, and Michael Prentice. Characterization of a large chromosomal high-pathogenicity island in biotype 1b *Yersinia enterocolitica*. *Journal of bacteriology*, 178(23):6743–6751, 1996.
188. B Joseph Hinnebusch, Robert D Perry, and Tom G Schwan. Role of the *Yersinia pestis* hemin storage (hms) locus in the transmission of plague by fleas. *Science*, 273(5273): 367–370, 1996.
189. Sandra Reuter, Thomas R. Connor, Lars Barquist, Danielle Walker, Theresa Feltwell, Simon R. Harris, Maria Fookes, Miquette E. Hall, Nicola K. Petty, Thilo M. Fuchs, Jukka Corander, Muriel Dufour, Tamara Ringwood, Cyril Savin, Christiane Bouchier, Liliane Martin, Minna Miettinen, Mikhail Shubin, Julia M. Riehm, Riikka Laukkanen-Ninios, Leila M. Sihvonen, Anja Siitonen, Mikael Skurnik, Juliana Pfrimer Falcao, Hiroshi Fukushima, Hol-

-
- ger C. Scholz, Michael B. Prentice, Brendan W. Wren, Julian Parkhill, Elisabeth Carniel, Mark Achtman, Alan McNally, and Nicholas R. Thomson. Parallel independent evolution of pathogenicity within the genus *Yersinia*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 111(18):6768–6773, MAY 6 2014. ISSN 0027-8424. doi: 10.1073/pnas.1317161111.
190. David A Rasko, Dale R Webster, Jason W Sahl, Ali Bashir, Nadia Boisen, Flemming Scheutz, Ellen E Paxinos, Robert Sebra, Chen-Shan Chin, Dimitris Iliopoulos, et al. Origins of the *E. coli* strain causing an outbreak of hemolytic–uremic syndrome in Germany. *New England Journal of Medicine*, 365(8):709–717, 2011.
191. M Achtman, K Zurth, C Morelli, G Torrea, A Guiry, and E Carniel. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 96(24):14043–14048, NOV 23 1999. ISSN 0027-8424. doi: 10.1073/pnas.96.24.14043.
192. J Parkhill, BW Wren, NR Thomson, RW Titball, MTG Holden, MB Prentice, M Sebahia, KD James, C Churcher, KL Mungall, S Baker, D Basham, SD Bentley, K Brooks, AM Cerdeno-Tarraga, T Chillingworth, A Cronin, RM Davies, P Davis, G Dougan, T Feltwell, N Hamlin, S Holroyd, K Jagels, AV Karlyshev, S Leather, S Moule, PCF Oyston, M Quail, K Rutherford, M Simmonds, J Skelton, K Stevens, S Whitehead, and BG Barrell. Genome sequence of *Yersinia pestis*, the causative agent of plague. *NATURE*, 413(6855):523–527, OCT 4 2001. ISSN 0028-0836. doi: 10.1038/35097083.
193. John M Doll, Paul S Zeitz, Paul Ettestad, Ann L Bucholtz, Ted Davis, and Kenneth Gage. Cat-transmitted fatal pneumonic plague in a person who traveled from Colorado to Arizona. *The American journal of tropical medicine and hygiene*, 51(1):109–114, 1994.
194. Carmen Buchrieser, Michael Prentice, and Elisabeth Carniel. The 102-kilobase unstable region of *Yersinia pestis* comprises a high-pathogenicity island linked to a pigmentation segment which undergoes internal rearrangement. *Journal of Bacteriology*, 180(9):2321–2329, 1998.
195. E Carniel. Plasmids and pathogenicity islands of *Yersinia*. *Pathogenicity Islands and the Evolution of Pathogenic Microbes*, pages 89–108, 2002.

-
196. Alexander G. Bobrov, Olga Kirillina, Jacqueline D. Fetherston, M. Clarke Miller, Joseph A. Burlison, and Robert D. Perry. The yersinia pestis siderophore, yersiniabactin, and the znuabc system both contribute to zinc acquisition and the development of lethal septicaemic plague in mice. *MOLECULAR MICROBIOLOGY*, 93(4):759–775, AUG 2014. ISSN 0950-382X. doi: 10.1111/mmi.12693.
197. Armand Paauw, Maurine A Leverstein-van Hall, Kok PM van Kessel, Jan Verhoef, and Ad C Fluit. Yersiniabactin reduces the respiratory oxidative stress response of innate immune cells. *PLoS one*, 4(12):e8240, 2009.
198. Biliana Lesic, Sandrine Bach, Jean-Marc Ghigo, Ulrich Dobrindt, Jörg Hacker, and Elisabeth Carniel. Excision of the high-pathogenicity island of yersinia pseudotuberculosis requires the combined actions of its cognate integrase and hef, a new recombination directionality factor. *Molecular microbiology*, 52(5):1337–1348, 2004.
199. Iman Chouikha, Lucie Charrier, Sofia Filali, Anne Derbise, and Elisabeth Carniel. Insights into the infective properties of ypf ϕ , the yersinia pestis filamentous phage. *Virology*, 407(1):43–52, 2010.
200. Anne Derbise, Viviane Chenal-Francisque, Flavie Pouillot, Corinne Fayolle, Marie-Christine Prevost, Claudine Medigue, Bernard Joseph Hinnebusch, and Elisabeth Carniel. A horizontally acquired filamentous phage contributes to the pathogenicity of the plague bacillus. *MOLECULAR MICROBIOLOGY*, 63(4):1145–1157, FEB 2007. ISSN 0950-382X. doi: 10.1111/j.1365-2958.2006.05570.x.
201. Nicholas R Waterfield, David J Bowen, Jacqueline D Fetherston, Robert D Perry, et al. The tc genes of photorhabdus: a growing family. *Trends in microbiology*, 9(4):185–191, 2001.
202. M Chandler and P Siguier. Insertion sequences. 2013.
203. Yujun Cui and Yajun Song. Genome and evolution of yersinia pestis. *Yersinia pestis: Retrospective and Perspective*, pages 171–192, 2016.
204. Fan Yang, Jian Yang, Xiaobing Zhang, Lihong Chen, Yan Jiang, Yongliang Yan, Xudong Tang, Jing Wang, Zhaohui Xiong, Jie Dong, et al. Genome dynamics and diversity of shigella species, the etiologic agents of bacillary dysentery. *Nucleic acids research*, 33(19):

-
- 6445–6458, 2005.
205. Jane Hawkey, Jonathan M Monk, Helen Billman-Jacobe, Bernhard Palsson, and Kathryn E Holt. Impact of insertion sequences on convergent evolution of shigella species. *PLoS genetics*, 16(7):e1008931, 2020.
206. Wen Deng, Valerie Burland, Guy Plunkett III, Adam Boutin, George F Mayhew, Paul Liss, Nicole T Perna, Debra J Rose, Bob Mau, Shiguo Zhou, et al. Genome sequence of yersinia pestis kim, 2002.
207. Yajun Song, Zongzhong Tong, Jin Wang, Li Wang, Zhaobiao Guo, Yanpin Han, Jianguo Zhang, Decui Pei, Dongsheng Zhou, Haiou Qin, et al. Complete genome sequence of yersinia pestis strain 91001, an isolate avirulent to humans. *DNA research*, 11(3):179–197, 2004.
208. Guy R Cornelis and Hans Wolf-Watz. The yersinia yop virulon: a bacterial system for subverting eukaryotic cells. *Molecular microbiology*, 23(5):861–867, 1997.
209. Nadine Grosdent, Isabelle Maridonneau-Parini, Marie-Paule Sory, and Guy R Cornelis. Role of yops and adhesins in resistance of yersinia enterocolitica to phagocytosis. *Infection and immunity*, 70(8):4165–4176, 2002.
210. E Bohn, S Müller, J Lauber, R Geffers, N Speer, C Spieth, J Krejci, B Manncke, J Buer, A Zell, et al. Gene expression patterns of epithelial cells modulated by pathogenicity factors of yersinia enterocolitica. *Cellular microbiology*, 6(2):129–141, 2004.
211. Robert D Perry, Susan C Straley, Jacqueline D Fetherston, Debra J Rose, Jason Gregor, and Frederick R Blattner. Dna sequencing and analysis of the low-ca²⁺-response plasmid pcd1 of yersinia pestis kim5. *Infection and Immunity*, 66(10):4611–4623, 1998.
212. Peng Wang, Wei Li, Zhikai Zhang, Ying Guo, Liyuan Shi, Rui Ye, Zhigang Cui, Guangcan Yang, Shanshan Dong, and Zhizhong Song. Characters of yulong yersinia pestis strains from yunnan province, china. *INTERNATIONAL JOURNAL OF CLINICAL AND EXPERIMENTAL MEDICINE*, 9(3):6394–6402, 2016. ISSN 1940-5901.
213. EP Rocha, Antoine Danchin, and Alain Viari. Analysis of long repeats in bacterial genomes reveals alternative evolutionary mechanisms in bacillus subtilis and other competent prokaryotes. *Molecular biology and evolution*, 16(9):1219–1230, 1999.

-
214. Florent Sebbane, Vladimir N Uversky, and Andrey P Anisimov. Yersinia pestis plasminogen activator. *Biomolecules*, 10(11):1554, 2020.
215. Daniel L. Zimble, Jay A. Schroeder, Justin L. Eddy, and Wyndham W. Lathem. Early emergence of yersinia pestis as a severe respiratory pathogen. *NATURE COMMUNICATIONS*, 6, JUN 2015. ISSN 2041-1723. doi: 10.1038/ncomms8487.
216. Luther E Lindler, Gregory V Plano, Valerie Burland, George F Mayhew, and Frederick R Blattner. Complete dna sequence and detailed analysis of the yersinia pestis kim5 plasmid encoding murine toxin and capsular antigen. *Infection and immunity*, 66(12):5731–5742, 1998.
217. Kelly J Davis, David L Fritz, Margaret L Pitt, Susan L Welkos, Patricia L Worsham, and Arthur M Friedlander. Pathology of experimental pneumonic plague produced by fraction 1-positive and fraction 1-negative yersinia pestis in african green monkeys (cercopithecus aethiops). *Archives of pathology & laboratory medicine*, 120(2):156–163, 1996.
218. Tatyana V Meka-Mechenko. F1-negative natural y. pestis strains. In *The Genus Yersinia*, pages 379–381. Springer, 2004.
219. Stephen T. Smiley. Current challenges in the development of vaccines for pneumonic plague. *EXPERT REVIEW OF VACCINES*, 7(2):209–221, MAR 2008. ISSN 1476-0584. doi: 10.1586/14760584.7.2.209.
220. BJ Hinnebusch, AE Rudolph, P Cherepanov, JE Dixon, TG Schwan, and A Forsberg. Role of yersinia murine toxin in survival of yersinia pestis in the midgut of the flea vector. *SCIENCE*, 296(5568):733–735, APR 26 2002. ISSN 0036-8075. doi: 10.1126/science.1069972.
221. David M Bland, Adélaïde Miarinjara, Christopher F Bosio, Jeanette Calarco, and B Joseph Hinnebusch. Acquisition of yersinia murine toxin enabled yersinia pestis to expand the range of mammalian hosts that sustain flea-borne plague. *PLoS Pathogens*, 17(10):e1009995, 2021.
222. Michael B Prentice, Keith D James, Julian Parkhill, Stephen G Baker, Kim Stevens, Mark N Simmonds, Karen L Mungall, Carol Churcher, Petra CF Oyston, Richard W Titball, et al. Yersinia pestis p_{fra} shows biovar-specific differences and recent common ancestry with a salmonella enterica serovar typhi plasmid. *Journal of bacteriology*, 183(8):2586–2594,

-
- 2001.
223. Eugen Pfeifer, Jorge A Moura de Sousa, Marie Touchon, and Eduardo PC Rocha. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. *Nucleic acids research*, 49(5):2655–2673, 2021.
224. Victor Ravin, Nikolai Ravin, Sherwood Casjens, Michael E Ford, Graham F Hatfull, and Roger W Hendrix. Genomic sequence and analysis of the atypical temperate bacteriophage n15. *Journal of molecular biology*, 299(1):53–73, 2000.
225. TC Leal-Balbino, NC Leal, CV Lopes, and AMP de Almeida. Differences in the stability of the plasmids of yersinia pestis cultures in vitro: Impact on virulence. *MEMORIAS DO INSTITUTO OSWALDO CRUZ*, 99(7):727–732, NOV 2004. ISSN 0074-0276. doi: 10.1590/S0074-02762004000700011.
226. Konstantinos T Konstantinidis and James M Tiedje. Towards a genome-based taxonomy for prokaryotes. *Journal of bacteriology*, 187(18):6258–6264, 2005.
227. Michael Richter, Ramon Rosselló-Móra, Frank Oliver Glöckner, and Jörg Peplies. Jspeciesws: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, 32(6):929–931, 2016.
228. Hervé Bercovier, Jan Ursing, J Don Brenner, Arnold G Steigerwalt, G Richard Fanning, Geraldine P Carter, and H Hr Mollaret. Yersinia kristensenii: A new species of enterobacteriaceae composed of sucrose-negative strains (formerly called atypical yersinia enterocolitica or yersinia enterocolitica-like). *Current Microbiology*, 4(4):219–224, 1980.
229. PSG Chain, E Carniel, FW Larimer, J Lamerdin, PO Stoutland, WM Regala, AM Georgescu, LM Vergez, ML Land, VL Motin, RR Brubaker, J Fowler, J Hinnebusch, M Marceau, C Medigue, M Simonet, V Chenal-Francisque, B Souza, D Dacheux, JM Elliott, A Derbise, LJ Hauser, and E Garcia. Insights into the evolution of yersinia pestis through whole-genome comparison with yersinia pseudotuberculosis. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 101(38):13826–13831, SEP 21 2004. ISSN 0027-8424. doi: 10.1073/pnas.0404012101.
230. Alan McNally, Nicholas R. Thomson, Sandra Reuter, and Brendan W. Wren. ‘add, stir and reduce’: Yersinia spp. as model bacteria for pathogen evolution. *NATURE REVIEWS*

-
- MICROBIOLOGY*, 14(3):177–190, MAR 2016. ISSN 1740-1526. doi: 10.1038/nrmicro.2015.29.
231. Ana Cerdeno-Tarraga, Nicholas Thomson, and Julian Parkhill. Pathogens in decay. *Nature Reviews Microbiology*, 2(10):774–775, 2004.
232. Mark J Pallen and Brendan W Wren. Bacterial pathogenomics. *Nature*, 449(7164):835–842, 2007.
233. Philip Supply, Michael Marceau, Sophie Mangenot, David Roche, Carine Rouanet, Varun Khanna, Laleh Majlessi, Alexis Criscuolo, Julien Tap, Alexandre Pawlik, et al. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of mycobacterium tuberculosis. *Nature genetics*, 45(2):172–179, 2013.
234. Hao Chung The, Duy Pham Thanh, Kathryn E Holt, Nicholas R Thomson, and Stephen Baker. The genomic signatures of shigella evolution, adaptation and geographical spread. *Nature Reviews Microbiology*, 14(4):235–250, 2016.
235. ST Cole, K Eiglmeier, J Parkhill, KD James, NR Thomson, PR Wheeler, N Honore, T Garnier, C Churcher, D Harris, et al. Massive gene decay in the leprosy bacillus. *Nature*, 409(6823):1007–1011, 2001.
236. Christian E. Demeure, Olivier Dussurget, Guillem Mas Fiol, Anne-Sophie Le Guern, Cyril Savin, and Javier Pizarro-Cerda. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination and diagnostics. *MICROBES AND INFECTION*, 21(5-6, SI):202–212, JUN-JUL 2019. ISSN 1286-4579. doi: 10.1016/j.micinf.2019.06.007.
237. Maini Kukkonen and Timo K Korhonen. The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in escherichia coli to systemic spread of yersinia pestis. *International Journal of Medical Microbiology*, 294(1):7–14, 2004.
238. Phelim Bradley, Henk C Den Bakker, Eduardo PC Rocha, Gil McVean, and Zamin Iqbal. Ultrafast search of all deposited bacterial and viral genomic data. *Nature biotechnology*, 37(2):152–159, 2019.
239. Zhemin Zhou, Nabil-Fareed Alikhan, Khaled Mohamed, Yulei Fan, Mark Achtman, Derek Brown, Marie Chattaway, Tim Dallman, Richard Delahay, Christian Kornschober, et al.

-
- The enterobase user's guide, with case studies on salmonella transmissions, yersinia pestis phylogeny, and escherichia core genomic diversity. *Genome research*, 30(1):138–152, 2020.
240. Maria A. Spyrou, Kirsten Bos, I, Alexander Herbig, and Johannes Krause. Ancient pathogen genomics as an emerging tool for infectious disease research. *NATURE REVIEWS GENETICS*, 20(6):323–340, JUN 2019. ISSN 1471-0056. doi: 10.1038/s41576-019-0119-1.
241. Martin CJ Maiden. Multilocus sequence typing of bacteria. *Annu. Rev. Microbiol.*, 60: 561–588, 2006.
242. Mark Achtman. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual review of microbiology*, 62(1):53–70, 2008.
243. Matthew N Van Ert, W Ryan Easterday, Lynn Y Huynh, Richard T Okinaka, Martin E Hugh-Jones, Jacques Ravel, Shaylan R Zanecki, Talima Pearson, Tatum S Simonson, Jana M U'Ren, et al. Global genetic population structure of bacillus anthracis. *PloS one*, 2(5): e461, 2007.
244. Srinand Sreevatsan, XI Pan, Kathryn E Stockbauer, Nancy D Connell, Barry N Kreiswirth, Thomas S Whittam, and James M Musser. Restricted structural gene polymorphism in the mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. *Proceedings of the National Academy of Sciences*, 94(18):9869–9874, 1997.
245. Claire Kidgell, Ulrike Reichard, John Wain, Bodo Linz, Mia Torpdahl, Gordon Dougan, and Mark Achtman. Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old. *Infection, Genetics and Evolution*, 2(1):39–45, 2002.
246. Daniel Godoy, Gaynor Randle, Andrew J Simpson, David M Aanensen, Tyrone L Pitt, Reimi Kinoshita, and Brian G Spratt. Multilocus sequence typing and evolutionary relationships among the causative agents of melioidosis and glanders, burkholderia pseudomallei and burkholderia mallei. *Journal of clinical microbiology*, 41(5):2068–2079, 2003.
247. Yujun Cui, Chang Yu, Yanfeng Yan, Dongfang Li, Yanjun Li, Thibaut Jombart, Lucy A. Weinert, Zuyun Wang, Zhaobiao Guo, Lizhi Xu, Yujiang Zhang, Hancheng Zheng, Nan Qin, Xiao Xiao, Mingshou Wu, Xiaoyi Wang, Dongsheng Zhou, Zhizhen Qi, Zongmin Du, Honglong Wu, Xianwei Yang, Hongzhi Cao, Hu Wang, Jing Wang, Shusen Yao, Alexander Rakin, Yingrui Li, Daniel Falush, Francois Balloux, Mark Achtman, Yajun Song, Jun

-
- Wang, and Ruifu Yang. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 110(2):577–582, JAN 8 2013. ISSN 0027-8424. doi: 10.1073/pnas.1205750110.
248. Rene Deignat. Varietes de l'espece *Pasteurella pestis*: nouvelle hypothese. *Bulletin of the World Health Organization*, 4(2):247, 1951.
249. Dongsheng Zhou, Zongzhong Tong, Yajun Song, Yanping Han, Decui Pei, Xin Pang, Junhui Zhai, Min Li, Baizhong Cui, Zhizhen Qi, et al. Genetics of metabolic variations between *Yersinia pestis* biovars and the proposal of a new biovar, *microtus*. *Journal of bacteriology*, 186(15):5147–5152, 2004.
250. Yanjun Li, Yujun Cui, Yolande Hauck, Mikhail E Platonov, Erhei Dai, Yajun Song, Zhaobiao Guo, Christine Pourcel, Svetlana V Dentovskaya, Andrey P Anisimov, et al. Genotyping and phylogenetic analysis of *Yersinia pestis* by mlva: insights into the worldwide expansion of central asia plague foci. *PloS one*, 4(6):e6000, 2009.
251. Zhizhen Qi, Yujun Cui, Qingwen Zhang, and Ruifu Yang. Taxonomy of *Yersinia pestis*. *Yersinia pestis: Retrospective and Perspective*, pages 35–78, 2016.
252. M Achtman, G Morelli, PX Zhu, T Wirth, I Diehl, B Kusecek, AJ Vogler, DM Wagner, CJ Allender, WR Easterday, V Chenal-Francoise, P Worsham, NR Thomson, J Parkhill, LE Lindler, E Carniel, and P Keim. Microevolution and history of the plague bacillus, *Yersinia pestis*. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 101(51):17837–17842, DEC 21 2004. ISSN 0027-8424. doi: 10.1073/pnas.0408026101.
253. Inaki Comas, Susanne Homolka, Stefan Niemann, and Sebastien Gagneux. Genotyping of genetically monomorphic bacteria: Dna sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. *PloS one*, 4(11):e7815, 2009.
254. Claire Chewapreecha, Matthew TG Holden, Minna Vehkala, Niko Välimäki, Zhirong Yang, Simon R Harris, Alison E Mather, Apichai Tuanyok, Birgit De Smet, Simon Le Hello, et al. Global and regional dissemination and evolution of *Burkholderia pseudomallei*. *Nature microbiology*, 2(4):1–8, 2017.

-
255. Alexandra Moura, Noémie Lefrancq, Thierry Wirth, Alexandre Leclercq, Vítor Borges, Brent Gilpin, Timothy J Dallman, Joachim Frey, Eelco Franz, Eva M Nielsen, et al. Emergence and global spread of listeria monocytogenes main clinical clonal complex. *Science advances*, 7(49):eabj9805, 2021.
256. Jane Hawkey, Kalani Paranagama, Kate S Baker, Rebecca J Bengtsson, François-Xavier Weill, Nicholas R Thomson, Stephen Baker, Louise Cerdeira, Zamin Iqbal, Martin Hunt, et al. Global population structure and genotyping framework for genomic surveillance of the major dysentery pathogen, shigella sonnei. *Nature communications*, 12(1):1–12, 2021.
257. Zhizhen Qi, Yujun Cui, Qingwen Zhang, and Ruifu Yang. Taxonomy of yersinia pestis. In R Yang and A Anisimov, editors, *YERSINIA PESTIS: RETROSPECTIVE AND PERSPECTIVE*, volume 918 of *Advances in Experimental Medicine and Biology*, pages 35–78. 2016. ISBN 978-94-024-0890-4; 978-94-024-0888-1. doi: 10.1007/978-94-024-0890-4_3.
258. David M. Wagner, Paul S. Keim, Holger C. Scholz, Edward C. Holmes, and Hendrik Poinar. Yersinia pestis and the three plague pandemics reply. *LANCET INFECTIOUS DISEASES*, 14(10):919, OCT 2014. ISSN 1473-3099. doi: 10.1016/S1473-3099(14)70923-5.
259. Michal Feldman, Michaela Harbeck, Marcel Keller, Maria A. Spyrou, Andreas Rott, Bernd Trautmann, Holger C. Scholz, Bernd Paeffgen, Joris Peters, Michael McCormick, Kirsten Bos, Alexander Herbig, and Johannes Krause. A high-coverage yersinia pestis genome from a sixth-century justinianic plague victim. *MOLECULAR BIOLOGY AND EVOLUTION*, 33(11):2911–2923, NOV 2016. ISSN 0737-4038. doi: 10.1093/molbev/msw170.
260. Amine Namouchi, Meriam Guellil, Oliver Kersten, Stephanie Hansch, Claudio Ottoni, Boris V. Schmid, Elsa Pacciani, Luisa Quaglia, Marco Vermunt, Egil L. Bauer, Michael Derrick, Anne O. Jensen, Sacha Kacki, Samuel K. Cohn, Jr., Nils C. Stenseth, and Barbara Bramanti. Integrative approach using yersinia pestis genomes to revisit the historical landscape of plague during the medieval period. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 115(50):E11790–E11797, DEC 11 2018. ISSN 0027-8424. doi: 10.1073/pnas.1812865115.
261. Maria A. Spyrou, Rezeda I. Tukhbatova, Michal Feldman, Joanna Drath, Sacha Kacki, Julia Beltran de Heredia, Susanne Arnold, Airat G. Sitdikov, Dominique Castex, Joachim Wahl, Ilgizar R. Gazimzyanov, Danis K. Nurgaliev, Alexander Herbig, Kirsten I. Bos, and Johannes Krause. Historical y. pestis genomes reveal the european black death as the source

-
- of ancient and modern plague pandemics. *CELL HOST & MICROBE*, 19(6):874–881, JUN 8 2016. ISSN 1931-3128. doi: 10.1016/j.chom.2016.05.012.
262. Maria A. Spyrou, Marcel Keller, Rezeda I. Tukhbatova, Christiana L. Scheib, Elizabeth A. Nelson, Aida Andrades Valtuena, Gunnar U. Neumann, Don Walker, Amelie Alterauge, Niamh Carty, Craig Cessford, Hermann Fetz, Michael Gourvennec, Robert Hartle, Michael Henderson, Kristin von Heyking, Sarah A. Inskip, Sacha Kacki, Felix M. Key, Elizabeth L. Knox, Christian Later, Prishita Maheshwari-Aplin, Joris Peters, John E. Robb, Juergen Schreiber, Toomas Kivisild, Dominique Castex, Sandra Losch, Michaela Harbeck, Alexander Herbig, Kirsten I. Bos, and Johannes Krause. Phylogeography of the second plague pandemic revealed through analysis of historical yersinia pestis genomes. *NATURE COMMUNICATIONS*, 10, OCT 2 2019. doi: 10.1038/s41467-019-12154-0.
263. Irina Morozova, Artem Kasianov, Sergey Bruskin, Judith Neukamm, Martyna Molak, Elena Batiieva, Aleksandra Pudło, Frank J Rühli, and Verena J Schuenemann. New ancient eastern european yersinia pestis genomes illuminate the dispersal of plague in europe. *Philosophical transactions of the Royal Society B*, 375(1812):20190569, 2020.
264. Maria A Spyrou, Lyazzat Musralina, Guido A Gneccchi Ruscone, Arthur Kocher, Pier-Giorgio Borbone, Valeri I Khartanovich, Alexandra Buzhilova, Leyla Djansugurova, Kirsten I Bos, Denise Kühnert, et al. The source of the black death in fourteenth-century central eurasia. *Nature*, 606(7915):718–724, 2022.
265. Simon Rasmussen, Morten Erik Allentoft, Kasper Nielsen, Ludovic Orlando, Martin Sikora, Karl-Goeran Sjoegren, Anders Gorm Pedersen, Mikkel Schubert, Alex Van Dam, Christian Moliin Outzen Kapel, Henrik Bjorn Nielsen, Soren Brunak, Pavel Avetisyan, Andrey Epimakhov, Mikhail Viktorovich Khalyapin, Artak Gnuni, Aivar Kriiska, Irena Lasak, Mait Metspalu, Vyacheslav Moiseyev, Andrei Gromov, Dalia Pokutta, Lehti Saag, Liivi Varul, Levon Yepiskoposyan, Thomas Sicheritz-Ponten, Robert A. Foley, Marta Mirazon Lahr, Rasmus Nielsen, Kristian Kristiansen, and Eske Willerslev. Early divergent strains of yersinia pestis in eurasia 5,000 years ago. *CELL*, 163(3):571–582, OCT 22 2015. ISSN 0092-8674. doi: 10.1016/j.cell.2015.10.009.
266. Nicolas Rascovan, Karl-Goran Sjoegren, Kristian Kristiansen, Rasmus Nielsen, Eske Willerslev, Christelle Desnues, and Simon Rasmussen. Emergence and spread of basal lineages of yersinia pestis during the neolithic decline. *CELL*, 176(1-2):295+, JAN 10 2019. ISSN

0092-8674. doi: 10.1016/j.cell.2018.11.005.

267. Aida Andrades Valtueña, Alissa Mittnik, Felix M Key, Wolfgang Haak, Raili Allmäe, Andrej Belinskij, Mantas Daubaras, Michal Feldman, Rimantas Jankauskas, Ivor Janković, et al. The stone age plague and its persistence in eurasia. *Current biology*, 27(23):3683–3691, 2017.
268. He Yu, Maria A. Spyrou, Marina Karapetian, Svetlana Shnaider, Rita Radzeviciute, Kathrin Naegele, Gunnar U. Neumann, Sandra Penske, Jana Zech, Mary Lucas, Petrus LeRoux, Patrick Roberts, Galina Pavlenok, Alexandra Buzhilova, Cosimo Posth, Choongwon Jeong, and Johannes Krause. Paleolithic to bronze age siberians reveal connections with first americans and across eurasia. *CELL*, 181(6):1232+, JUN 11 2020. ISSN 0092-8674. doi: 10.1016/j.cell.2020.04.037.
269. Julian Susat, Harald Lubke, Alexander Immel, Ute Brinker, Aija Macane, John Meadows, Britta Steer, Andreas Tholey, Ilga Zagorska, Guntis Gerhards, Ulrich Schmolcke, Marcis Kalnins, Andre Franke, Elina Petersone-Gordina, Barbara Tessman, Mari Torv, Stefan Schreiber, Christian Andree, Valdis Berzins, Almut Nebel, and Ben Krause-Kyora. A 5,000-year-old hunter-gatherer already plagued by yersinia pestis. *CELL REPORTS*, 35(13), JUN 2021. ISSN 2211-1247. doi: 10.1016/j.celrep.2021.109278.
270. Aida Andrades Valtueña, Gunnar U Neumann, Maria A Spyrou, Lyazzat Musralina, Franziska Aron, Arman Beisenov, Andrey B Belinskiy, Kirsten I Bos, Alexandra Buzhilova, Matthias Conrad, et al. Stone age yersinia pestis genomes shed light on the early evolution, diversity, and ecology of plague. *Proceedings of the National Academy of Sciences*, 119(17):e2116722119, 2022.
271. Yujun Cui, Boris V Schmid, Hanli Cao, Xiang Dai, Zongmin Du, W Ryan Easterday, Haihong Fang, Chenyi Guo, Shanqian Huang, Wanbing Liu, et al. Evolutionary selection of biofilm-mediated extended phenotypes in yersinia pestis in response to a fluctuating environment. *Nature communications*, 11(1):1–8, 2020.
272. Liyuan Shi, Jingliang Qin, Hongyuan Zheng, Ying Guo, Haipeng Zhang, Youhong Zhong, Chao Yang, Shanshan Dong, Fengyi Yang, Yarong Wu, Guangyu Zhao, Yajun Song, Ruifu Yang, Peng Wang, and Yujun Cui. New genotype of yersinia pestis found in live rodents in yunnan province, china. *FRONTIERS IN MICROBIOLOGY*, 12, APR 15 2021. doi: 10.3389/fmicb.2021.628335.

-
273. Alina N Balykova, Lyubov M Kukleva, Ekaterina A Naryshkina, Galina A Eroshenko, and Vladimir V Kuttyrev. Five draft genome sequences of historical yersinia pestis strains of phylogroups 2. med4 and 2. med1 of the medieval biovar. *Microbiology Resource Announcements*, pages e00044–22, 2022.
274. Mark Eppinger, Patricia L Worsham, Mikeljon P Nikolich, David R Riley, Yinong Sebastian, Sherry Mou, Mark Achtman, Luther E Lindler, and Jacques Ravel. Genome sequence of the deep-rooted yersinia pestis strain angola reveals new insights into the evolution and pangenome of the plague bacterium. *Journal of bacteriology*, 192(6):1685–1699, 2010.
275. Timo K Korhonen, Johanna Haiko, Liisa Laakkonen, Hanna M Järvinen, and Benita Westerlund-Wikström. Fibrinolytic and coagulative activities of yersinia pestis. *Frontiers in cellular and infection microbiology*, 3:35, 2013.
276. Julian Susat, Joanna H. Bonczarowska, Elina Petersone-Gordina, Alexander Immel, Almut Nebel, Guntis Gerhards, and Ben Krause-Kyora. Yersinia pestis strains from latvia show depletion of the pla virulence gene at the end of the second plague pandemic. *SCIENTIFIC REPORTS*, 10(1), SEP 3 2020. ISSN 2045-2322. doi: 10.1038/s41598-020-71530-9.
277. Barbara Bramanti, Yarong Wu, Ruifu Yang, Yujun Cui, and Nils Chr Stenseth. Assessing the origins of the european plagues following the black death: A synthesis of genomic, historical, and ecological information. *Proceedings of the National Academy of Sciences*, 118(36):e2101940118, 2021.
278. Andrew A Filippov, Nikolai S Solodovnikov, Liubov M Kookleva, and Olga A Protsenko. Plasmid content in yersinia pestis strains of different origin. *FEMS Microbiology Letters*, 67(1-2):45–48, 1990.
279. Meriam Guellil, Oliver Kersten, Amine Namouchi, Stefania Luciani, Isolina Marota, Caroline A Arcini, Elisabeth Iregren, Robert A Lindemann, Gunnar Warfvinge, Lela Bakanidze, et al. A genomic and historical synthesis of plague in 18th century eurasia. *Proceedings of the National Academy of Sciences*, 117(45):28328–28335, 2020.
280. Guy Baele, Marc A Suchard, Andrew Rambaut, and Philippe Lemey. Emerging concepts of data integration in pathogen phylodynamics. *Systematic biology*, 66(1):e47–e65, 2017.
281. Marc A. Suchard, Philippe Lemey, Guy Baele, Daniel L. Ayres, Alexei J. Drummond, and

-
- Andrew Rambaut. Bayesian phylogenetic and phylodynamic data integration using beast 1.10. *VIRUS EVOLUTION*, 4(1), JAN 2018. doi: 10.1093/ve/vey016.
282. Remco Bouckaert, Timothy G Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, et al. Beast 2.5: An advanced software platform for bayesian evolutionary analysis. *PLoS computational biology*, 15(4):e1006650, 2019.
283. Fabrícia F Nascimento, Mario dos Reis, and Ziheng Yang. A biologist's guide to bayesian phylogenetic analysis. *Nature ecology & evolution*, 1(10):1446–1454, 2017.
284. Verity Hill and Guy Baele. Bayesian estimation of past population dynamics in beast 1.10 using the skygrid coalescent model. *Molecular biology and evolution*, 36(11):2620–2628, 2019.
285. Simon YW Ho, Urmas Saarma, Ross Barnett, James Haile, and Beth Shapiro. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One*, 3(2):e1615, 2008.
286. Mandev S Gill, Philippe Lemey, Marc A Suchard, Andrew Rambaut, and Guy Baele. On-line bayesian phylodynamic inference in beast with application to epidemic reconstruction. *Molecular biology and evolution*, 37(6):1832–1842, 2020.
287. Katherine Eaton, Leo Featherstone, Sebastian Duchene, Ann Carmichael, Nükhet Varlık, Brian Golding, Edward Holmes, and Hendrik Poinar. Plagued by a cryptic clock: Insight and issues from the global phylogeny of yersinia pestis, 2021.
288. John H Tay, Ashleigh F Porter, Wytamma Wirth, and Sebastian Duchene. The emergence of sars-cov-2 variants of concern is driven by acceleration of the substitution rate. *Molecular Biology and Evolution*, 39(2):msac013, 2022.
289. John R Mascola, Barney S Graham, and Anthony S Fauci. Sars-cov-2 viral variants—tackling a moving target. *Jama*, 325(13):1261–1262, 2021.
290. Andrew M Ritchie, Xia Hua, and Lindell Bromham. Investigating the reliability of molecular estimates of evolutionary time when substitution rates and speciation rates vary. *BMC ecology and evolution*, 22(1):1–19, 2022.

-
291. Verity Hill, Christopher Ruis, Sumali Bajaj, Oliver G Pybus, and Moritz UG Kraemer. Progress and challenges in virus genomic epidemiology. *Trends in Parasitology*, 37(12): 1038–1049, 2021.
292. Edouard-Raoul Brygoo. Epidémiologie de la peste à madagascar. *Archives de l'Institut Pasteur de Madagascar*, 1966.
293. Amy J. Vogler, Voahangy Andrianaivoarimanana, Sandra Telfer, Carina M. Hall, Jason W. Sahl, Crystal M. Hepp, Heather Centner, Genevieve Andersen, Dawn N. Birdsell, Lila Rahalison, Roxanne Nottingham, Paul Keim, David M. Wagner, and Minoarisoa Rajerison. Temporal phylogeography of yersinia pestis in madagascar: Insights into the long-term maintenance of plague. *PLOS NEGLECTED TROPICAL DISEASES*, 11(9), SEP 2017. ISSN 1935-2735. doi: 10.1371/journal.pntd.0005887.
294. Amy J. Vogler, Jason W. Sahl, Nilma C. Leal, Marise Sobreira, Charles H. D. Williamson, Molly C. Bollig, Dawn N. Birdsell, Andrew Rivera, Brian Thompson, Roxanne Nottingham, Antonio M. Rezende, Paul Keim, Alzira M. P. Almeida, and David M. Wagner. A single introduction of yersinia pestis to brazil during the 3rd plague pandemic. *PLOS ONE*, 14 (1), JAN 9 2019. ISSN 1932-6203. doi: 10.1371/journal.pone.0209478.
295. Monica H. Green. Putting africa on the black death map: Narratives from genetics and history. *AFRIQUES-DEBATS METHODES ET TERRAINS D HISTOIRE*, 9, 2018. ISSN 2108-6796. doi: 10.4000/afriques.2125.
296. Brenda J Baker, Gillian Crane-Kramer, Michael W Dee, Lesley A Gregoricka, Maciej Henneberg, Christine Lee, Sheila A Lukehart, David C Mabey, Charlotte A Roberts, Ann LW Stodder, et al. Advancing the understanding of treponemal disease in the past and present. *American Journal of Physical Anthropology*, 171:5–41, 2020.
297. Daniel Antoine. The archaeology of "plague". *Medical history. Supplement*, (27):101–14, 2008. ISSN 0950-5571.
298. Olivier Dutour and Michel Signoli. Etude anthropologique d'un charnier de la grande peste de marseille (1720-1722) premiers résultats. *Anthropologie et préhistoire*, 108:147–158, 1997.
299. Kirsten I. Bos, Verena J. Schuenemann, G. Brian Golding, Hernan A. Burbano, Nicholas

-
- Waglechner, Brian K. Coombes, Joseph B. McPhee, Sharon N. DeWitte, Matthias Meyer, Sarah Schmedes, James Wood, David J. D. Earn, D. Ann Herring, Peter Bauer, Hendrik N. Poinar, and Johannes Krause. A draft genome of yersinia pestis from victims of the black death. *NATURE*, 478(7370):506–510, OCT 27 2011. ISSN 0028-0836. doi: 10.1038/nature10549.
300. Aida Andrades Valtueña. *Beyond phylogenies: advancing analytical approaches for the field of ancient pathogenomics*. PhD thesis, Dissertation, Jena, Friedrich-Schiller-Universität Jena, 2021, 2021.
301. M Drancourt, G Aboudharam, M Signoli, O Dutour, and D Raoult. Detection of 400-year-old yersinia pestis dna in human dental pulp: An approach to the diagnosis of ancient septicemia. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 95(21):12637–12640, OCT 13 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.21.12637.
302. Kyle Harper. The fate of rome. In *The Fate of Rome*. Princeton University Press, 2017.
303. Stephen Mitchell. *A history of the later Roman Empire, AD 284-641*. John Wiley & Sons, 2014.
304. Marcel Keller, Maria A. Spyrou, Christiana L. Scheib, Gunnar U. Neumann, Andreas Kropelin, Brigitte Haas-Gebhard, Bernd Paffgen, Jochen Haberstroh, Albert Ribera i Lacomba, Claude Raynaud, Craig Cessford, Raphael Durand, Peter Stadler, Kathrin Nagele, Jessica S. Bates, Bernd Trautmann, Sarah A. Inskip, Joris Peters, John E. Robb, Toomas Kivisild, Dominique Castex, Michael McCormick, Kirsten I. Bos, Michaela Harbeck, Alexander Herbig, and Johannes Krause. Ancient yersinia pestis genomes from across western europe reveal early diversification during the first pandemic (541-750). *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 116(25):12363–12372, JUN 18 2019. ISSN 0027-8424. doi: 10.1073/pnas.1820447116.
305. Peter de Barros Damgaard, Nina Marchi, Simon Rasmussen, Michael Peyrot, Gabriel Renaud, Thorfinn Korneliussen, J. Victor Moreno-Mayar, Mikkel Winther Pedersen, Amy Goldberg, Emma Usmanova, Nurbol Baimukhanov, Valeriy Loman, Lotte Hedeager, Anders Gorm Pedersen, Kasper Nielsen, Gennady Afanasiev, Kunbolot Akmatov, Almaz Al-dashev, Ashyk Alpaslan, Gabit Baimbetov, Vladimir I. Bazaliiskii, Arman Beisenov, Bazartseren Boldbaatar, Bazartseren Boldgiv, Choduraa Dorzhu, Sturla Ellingvag, Diimaajav Er-

-
- denebaatar, Rana Dajani, Evgeniy Dmitriev, Valeriy Evdokimov, Karin M. Frei, Andrey Gromov, Alexander Goryachev, Hakon Hakonarson, Tatyana Hegay, Zaruhi Khachatryan, Ruslan Khaskhanov, Egor Kitov', Alina Kolbina, Tabaldiev Kubatbek, Alexey Kukushkin, Igor Kukushkin, Nina Lau, Ashot Margaryan, Inga Merkyte, Ilya V. Mertz, Viktor K. Mertz, Enkhbayar Mijiddorj, Vyacheslav Moiyesev, Gulmira Mukhtarova, Bekmukhanbet Nurmukhanbetov, Z. Orozbekova, Irina Panyushkina, Karol Pieta, Vaclav Smrcka, Irina Shevnina, Andrey Logvin, Karl-Goran Sjogren, Tereza Stolcova, Kadic Ha Tashbaeva, Alexander Tkachev, Turaly Tulegenov, Dmitriy Voyakin, Levon Yepiskoposyan, Sainbileg Undrakhbold, Victor Varfolomeev, Andrzej Weber, Nikolay Kradin, Morten E. Allentoft, Ludovic Orlando, Rasmus Nielsen, Martin Sikora, Evelyne Heyer, Kristian Kristiansen, and Eske Willerslev. 137 ancient human genomes from across the eurasian steppes. *NATURE*, 557(7705):369+, MAY 17 2018. ISSN 0028-0836. doi: 10.1038/s41586-018-0094-2.
306. P Sarris. The justinianic plague: origins and effects. *CONTINUITY AND CHANGE*, 17 (2):169–182, AUG 2002. ISSN 0268-4160. doi: 10.1017/S0268416002004137.
307. P Allen. The “justinianic” plague. *Byzantion; revue internationale des etudes byzantines*, 49:5–20, 1979. ISSN 0378-2506.
308. Ole Jørgen Benedictow, Ole L Benedictow, et al. *The Black Death, 1346-1353: the complete history*. Boydell & Brewer, 2004.
309. Boris V. Schmid, Ulf Buentgen, W. Ryan Easterday, Christian Ginzler, Lars Walloe, Barbara Bramanti, and Nils Chr Stenseth. Climate-driven introduction of the black death and successive plague reintroductions into europe. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 112(10):3020–3025, MAR 10 2015. ISSN 0027-8424. doi: 10.1073/pnas.1412887112.
310. Amy J Vogler, Paul Keim, and David M Wagner. A review of methods for subtyping yersinia pestis: from phenotypes to whole genome sequencing. *Infection, Genetics and Evolution*, 37:21–36, 2016.
311. S Schmedes and B Budowle. Microbial forensics. *Encyclopedia of Microbiology*, page 134, 2019.
312. Wenjun Li, Didier Raoult, and Pierre-Edouard Fournier. Bacterial strain typing in the genomic era. *FEMS microbiology reviews*, 33(5):892–916, 2009.

-
313. Lewis M Graves and Bala Swaminathan. Pulsenet standardized protocol for subtyping *listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. *International journal of food microbiology*, 65(1-2):55–62, 2001.
314. P Keim, LB Price, AM Klevytska, KL Smith, JM Schupp, R Okinaka, PJ Jackson, and ME Hugh-Jones. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *bacillus anthracis*. *Journal of bacteriology*, 182(10):2928–2936, 2000.
315. Gabriela Torrea, Viviane Chenal-Francisque, Alexandre Leclercq, and Elisabeth Carniel. Efficient tracing of global isolates of *yersinia pestis* by restriction fragment length polymorphism analysis using three insertion sequences as probes. *Journal of clinical microbiology*, 44(6):2084–2092, 2006.
316. Ibtissem Grissa, Gilles Vergnaud, and Christine Pourcel. The crisprdb database and tools to display crispns and to generate dictionaries of spacers and repeats. *BMC bioinformatics*, 8(1):1–10, 2007.
317. Yanjun Li and Yujun Cui. Genotyping of *yersinia pestis*. In *Yersinia Pestis Protocols*, pages 35–59. Springer, 2018.
318. Chao Yang and Yujun Cui. Genome-wide variation analysis of *yersinia pestis*. In *Yersinia Pestis Protocols*, pages 61–66. Springer, 2018.
319. Yanfeng Yan, Hu Wang, Dongfang Li, Xianwei Yang, Zuyun Wang, Zhizhen Qi, Qingwen Zhang, Baizhong Cui, Zhaobiao Guo, Chang Yu, Jun Wang, Jian Wang, Guangming Liu, Yajun Song, Yingrui Li, Yujun Cui, and Ruifu Yang. Two-step source tracing strategy of *yersinia pestis* and its historical epidemiology in a specific region. *PLOS ONE*, 9(1), JAN 9 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0085374.
320. Cyril Savin, Anne-Sophie Le Guern, Fanny Chereau, Julien Guglielmini, Guillaume Heuzé, Christian Demeure, and Javier Pizarro-Cerdá. First description of a *yersinia pseudotuberculosis* clonal outbreak in france, confirmed using a new core genome multilocus sequence typing method. *bioRxiv*, 2022.
321. Zhemin Zhou, Jane Charlesworth, and Mark Achtman. Hiercc: A multi-level clustering scheme for population assignments based on core genome mlst. *Bioinformatics*, 37(20): 3645–3646, 2021.

-
322. Francesc Coll, Ruth McNERney, José Afonso Guerra-Assunção, Judith R Glynn, João Perdigão, Miguel Viveiros, Isabel Portugal, Arnab Pain, Nigel Martin, and Taane G Clark. A robust snp barcode for typing mycobacterium tuberculosis complex strains. *Nature communications*, 5(1):1–5, 2014.
323. Vanessa K Wong, Stephen Baker, Thomas R Connor, Derek Pickard, Andrew J Page, Jayshree Dave, Niamh Murphy, Richard Holliman, Armine Sefton, Michael Millar, et al. An extended genotyping framework for salmonella enterica serovar typhi, the cause of human typhoid. *Nature communications*, 7(1):1–11, 2016.
324. Nathan D Grubaugh, Jason T Ladner, Philippe Lemey, Oliver G Pybus, Andrew Rambaut, Edward C Holmes, and Kristian G Andersen. Tracking virus outbreaks in the twenty-first century. *Nature microbiology*, 4(1):10–19, 2019.
325. S Pollett, JR Fauver, Irina Maljkovic Berry, M Melendrez, A Morrison, LD Gillis, MA Johansson, RG Jarman, and ND Grubaugh. Genomic epidemiology as a public health tool to combat mosquito-borne virus outbreaks. *The Journal of Infectious Diseases*, 221 (Supplement_3):S308–S318, 2020.
326. François-Xavier Weill, Daryl Domman, Elisabeth Njamkepo, Abdullrahman A Almesbahi, Mona Naji, Samar Saeed Nasher, Ankur Rakesh, Abdullah M Assiri, Naresh Chand Sharma, Samuel Kariuki, et al. Genomic insights into the 2016–2017 cholera epidemic in yemen. *Nature*, 565(7738):230–233, 2019.
327. Mihaela Oprea, Elisabeth Njamkepo, Daniela Cristea, Anna Zhukova, Clifford G Clark, Anatoly N Kravetz, Elena Monakhova, Adriana S Ciontea, Radu Cojocaru, Jean Rauzier, et al. The seventh pandemic of cholera in europe revisited by microbial genomics. *Nature communications*, 11(1):1–10, 2020.
328. Caisey V Pulford, Blanca M Perez-Sepulveda, Rocío Canals, Jessica A Bevington, Rebecca J Bengtsson, Nicolas Wenner, Ella V Rodwell, Benjamin Kumwenda, Xiaojun Zhu, Rebecca J Bennett, et al. Stepwise evolution of salmonella typhimurium st313 causing bloodstream infection in africa. *Nature microbiology*, 6(3):327–338, 2021.
329. Vanessa K Wong, Stephen Baker, Derek J Pickard, Julian Parkhill, Andrew J Page, Nicholas A Feasey, Robert A Kingsley, Nicholas R Thomson, Jacqueline A Keane, François-Xavier Weill, et al. Phylogeographical analysis of the dominant multidrug-resistant h58

-
- clade of salmonella typhi identifies inter-and intracontinental transmission events. *Nature genetics*, 47(6):632–639, 2015.
330. Julian Parkhill and Brendan W. Wren. Bacterial epidemiology and biology - lessons from genome sequencing. *GENOME BIOLOGY*, 12(10), 2011. ISSN 1474-760X. doi: 10.1186/gb-2011-12-10-230.
331. Alexey M Kozlov, Diego Darriba, Tomáš Flouri, Benoit Morel, and Alexandros Stamatakis. Raxml-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21):4453–4455, 2019.
332. Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. Iq-tree 2: New models and efficient methods for phylogenetic inference in the genomic era. *MOLECULAR BIOLOGY AND EVOLUTION*, 37(5):1530–1534, MAY 2020. ISSN 0737-4038. doi: 10.1093/molbev/msaa015.
333. Zhemin Zhou, Nabil-Fareed Alikhan, Martin J Sergeant, Nina Luhmann, Cátia Vaz, Alexandre P Francisco, João André Carriço, and Mark Achtman. Grapetree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome research*, 28(9):1395–1404, 2018.
334. Thu-Hien To, Matthieu Jung, Samantha Lycett, and Olivier Gascuel. Fast dating using least-squares criteria and algorithms. *Systematic biology*, 65(1):82–97, 2016.
335. Nuno Rodrigues Faria, Marc A Suchard, Andrew Rambaut, and Philippe Lemey. Toward a quantitative understanding of viral phylogeography. *Current opinion in virology*, 1(5):423–429, 2011.
336. Oliver G. Pybus, Marc A. Suchard, Philippe Lemey, Flavien J. Bernardin, Andrew Rambaut, Forrest W. Crawford, Rebecca R. Gray, Nimalan Arinaminpathy, Susan L. Stramer, Michael P. Busch, and Eric L. Delwart. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 109(37):15066–15071, SEP 11 2012. ISSN 0027-8424. doi: 10.1073/pnas.1206598109.
337. Simon Dellicour, Keith Durkin, Samuel L Hong, Bert Vanmechelen, Joan Martí-Carreras,

-
- Mandev S Gill, Cécile Meex, Sébastien Bontems, Emmanuel André, Marius Gilbert, et al. A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of sars-cov-2 lineages. *Molecular biology and evolution*, 38(4):1608–1613, 2021.
338. Tanja Stadler, Denise Kühnert, David A Rasmussen, and Louis du Plessis. Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data. *PLoS currents*, 6, 2014.
339. Philippe Lemey, Andrew Rambaut, Trevor Bedford, Nuno Faria, Filip Bielejec, Guy Baele, Colin A. Russell, Derek J. Smith, Oliver G. Pybus, Dirk Brockmann, and Marc A. Suchard. Unifying viral genetics and human transportation data to predict the global transmission dynamics of human influenza h3n2. *PLOS PATHOGENS*, 10(2), FEB 2014. ISSN 1553-7366. doi: 10.1371/journal.ppat.1003932.
340. Simon Dellicour, Sebastian Lequime, Bram Vrancken, Mandev S Gill, Paul Bastide, Karthik Gangavarapu, Nathaniel L Matteson, Yi Tan, Louis Du Plessis, Alexander A Fisher, et al. Epidemiological hypothesis testing using a phylogeographic and phylodynamic framework. *Nature communications*, 11(1):1–11, 2020.
341. Yujiang Zhang, Tao Luo, Chao Yang, Xihong Yue, Rong Guo, Xinhui Wang, Mingde Buren, Yuqin Song, Ruifu Yang, Hanli Cao, Yujun Cui, and Xiang Dai. Phenotypic and molecular genetic characteristics of yersinia pestis at an emerging natural plague focus, junggar basin, china. *AMERICAN JOURNAL OF TROPICAL MEDICINE AND HYGIENE*, 98(1):231–237, 2018. ISSN 0002-9637. doi: 10.4269/ajtmh.17-0195.
342. Gulmira Sariyeva, Gulnara Bazarkanova, Ravshambek Maimulov, Sabirzhan Abdikarimov, Berzhan Kurmanov, Aigul Abdirassilova, Anton Shabunin, Zaurbek Sagiyev, Aigul Dzhaparova, Ziyat Abdel, Raikhan Mussagaliyeva, Serge Morand, Vladimir Motin, and Michael Kosoy. Marmots and yersinia pestis strains in two plague endemic areas of tien shan mountains. *FRONTIERS IN VETERINARY SCIENCE*, 6, JUL 4 2019. doi: 10.3389/fvets.2019.00207.
343. Xiaoqing Xu, Yujun Cui, Youquan Xin, Xiaoyan Yang, Qingwen Zhang, Yong Jin, Haihong Zhao, Jian He, Xing Jin, Cunxiang Li, Juan Jin, Xiang Li, Haisheng Wu, and Zhizhen Qi. Genetic diversity and spatial-temporal distribution of yersinia pestis in qinghai plateau, china. *PLOS NEGLECTED TROPICAL DISEASES*, 12(6), JUN 2018. ISSN 1935-2735. doi: 10.1371/journal.pntd.0006579.

-
344. Amy J. Vogler, Fabien Chan, David M. Wagner, Philippe Roumagnac, Judy Lee, Roxanne Nera, Mark Eppinger, Jacques Ravel, Lila Rahalison, Bruno W. Rasoamanana, Stephen M. Beckstrom-Sternberg, Mark Achtman, Suzanne Chanteau, and Paul Keim. Phylogeography and molecular epidemiology of yersinia pestis in madagascar. *PLOS NEGLECTED TROPICAL DISEASES*, 5(9), SEP 2011. ISSN 1935-2735. doi: 10.1371/journal.pntd.0001319.
345. Ch Villabona-Arenas, William P Hanage, Damien C Tully, et al. Phylogenetic interpretation during outbreaks requires caution. *Nature microbiology*, 5(7):876–877, 2020.
346. Soanandrasana Rahelinirina, Jean Marc Duplantier, Jocelyn Ratovonjato, Olga Ramilijaona, Mamy Ratsimba, and Lila Rahalison. Study on the movement of rattus rattus and evaluation of the plague dispersion in madagascar. *Vector-borne and zoonotic diseases*, 10(1):77–84, 2010.
347. Ying Zhou, Renzhe Xu, Dongsheng Hu, Yang Yue, Qingquan Li, and Jizhe Xia. Effects of human mobility restrictions on the spread of covid-19 in shenzhen, china: a modelling study using mobile phone data. *The Lancet Digital Health*, 2(8):e417–e424, 2020.
348. AK AKIEV. Epidemiology and incidence of plague in the world, 1958-79. *BULLETIN OF THE WORLD HEALTH ORGANIZATION*, 60(2):165–169, 1982.
349. The Lancet Infectious Diseases. War and infectious diseases: brothers in arms, 2022.
350. Yarong Wu, Tongyu Hao, Xiuwei Qian, Xianglilan Zhang, Yajun Song, Ruifu Yang, and Yujun Cui. Small insertions and deletions drive genomic plasticity during adaptive evolution of yersinia pestis. *Microbiology spectrum*, 10(3):e02242–21, 2022.
351. Wei Sun and Amit K Singh. Plague vaccine: recent progress and prospects. *npj Vaccines*, 4(1):1–9, 2019.
352. Voahangy Andrianaivoarimanana, David M Wagner, Dawn N Birdsell, Birgit Nikolay, Faniry Rakotoarimanana, Lovasoa N Randriantseheno, Amy J Vogler, Jason W Sahl, Carina M Hall, Nawarat Somprasong, et al. Transmission of antimicrobial resistant yersinia pestis during a pneumonic plague outbreak. *Clinical Infectious Diseases*, 74(4):695–702, 2022.
353. Allison Black, Duncan R MacCannell, Thomas R Sibley, and Trevor Bedford. Ten recommendations for supporting open pathogen genomic analysis in public health. *Nature medicine*, 26(6):832–841, 2020.

-
354. Ferric C Fang and Arturo Casadevall. Competitive science: is competition ruining science?, 2015.
355. Joeri K Tijdink, Reinout Verbeke, and Yvo M Smulders. Publication pressure and scientific misconduct in medical scientists. *Journal of Empirical Research on Human Research Ethics*, 9(5):64–71, 2014.