

## RESEARCH ARTICLE

## Phylogenetic analysis of the origin and spread of plague in Madagascar

Luis Roger Esquivel Gomez<sup>1\*</sup>, Cyril Savin<sup>2,3,4</sup>, Voahangy Andrianaivoarimanana<sup>5</sup>, Soloandry Rahajandraibe<sup>5</sup>, Lovasoa Nomena Randriantseho<sup>5</sup>, Zheming Zhou<sup>6,7</sup>, Arthur Kocher<sup>1,8</sup>, Xavier Didelot<sup>9,10</sup>, Minoarisoa Rajerison<sup>5</sup>, Denise Kühnert<sup>1\*</sup>

**1** Transmission, Infection, Diversification & Evolution Group (*tide*), Max Planck Institute for Geoanthropology (formerly MPI for the Science of Human History), Jena, Germany, **2** Institut Pasteur, Université de Paris, Yersinia Research Unit, Paris, France, **3** Institut Pasteur, Université de Paris, Yersinia National Reference Laboratory, Paris, France, **4** Institut Pasteur, Université de Paris, WHO Collaborative Reference & Research Center for Plague FRA-140, Paris, France, **5** Plague Unit, Central Laboratory for Plague, Institut Pasteur de Madagascar, Antananarivo, Madagascar, **6** Pasteurien College, Medical school of Soochow University, Soochow University, Suzhou, China, **7** Warwick Medical School, University of Warwick, Coventry, United Kingdom, **8** Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **9** School of Life Sciences, University of Warwick, Coventry, United Kingdom, **10** Department of Statistics, University of Warwick, Coventry, United Kingdom

\* [esquivelgomez@shh.mpg.de](mailto:esquivelgomez@shh.mpg.de) (LREG); [kuehnert@shh.mpg.de](mailto:kuehnert@shh.mpg.de) (DK)



## OPEN ACCESS

**Citation:** Esquivel Gomez LR, Savin C, Andrianaivoarimanana V, Rahajandraibe S, Randriantseho LN, Zhou Z, et al. (2023) Phylogenetic analysis of the origin and spread of plague in Madagascar. *PLoS Negl Trop Dis* 17(5): e0010362. <https://doi.org/10.1371/journal.pntd.0010362>

**Editor:** Abdallah M. Samy, Faculty of Science, Ain Shams University (ASU), EGYPT

**Received:** March 28, 2022

**Accepted:** April 3, 2023

**Published:** May 1, 2023

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

**Data Availability Statement:** Xml files with all the analyzed sequences and output files are included in the [Supporting information](#). The assemblies of all sequences can be downloaded from EnteroBase after creating an account (<https://enterobase.warwick.ac.uk/species/index/yersinia>). All sequence reads are publicly available in the Sequence Read Archive (SRA) of the NCBI. [S1 Table](#) contains additional information about the samples, including the EnteroBase IDs and NCBI accession numbers.

## Abstract

## Background

Plague is a zoonotic disease caused by the bacterium *Yersinia pestis*, highly prevalent in the Central Highlands, a mountainous region in the center of Madagascar. After a plague-free period of over 60 years in the northwestern coast city of Mahajanga, the disease reappeared in 1991 and caused several outbreaks until 1999. Previous research indicates that the disease was reintroduced to the city of Mahajanga from the Central Highlands instead of reemerging from a local reservoir. However, it is not clear how many reintroductions occurred and when they took place.

## Methodology/Principal findings

In this study we applied a Bayesian phylogeographic model to detect and date migrations of *Y. pestis* between the two locations that could be linked to the re-emergence of plague in Mahajanga. Genome sequences of 300 *Y. pestis* strains sampled between 1964 and 2012 were analyzed. Four migrations from the Central Highlands to Mahajanga were detected. Two resulted in persistent transmission in humans, one was responsible for most of the human cases recorded between 1995 and 1999, while the other produced plague cases in 1991 and 1992. We dated the emergence of the *Y. pestis* sub-branch 1.ORI3, which is only present in Madagascar and Turkey, to the beginning of the 20<sup>th</sup> century, using a Bayesian molecular dating analysis. The split between 1.ORI3 and its ancestor lineage 1.ORI2 was dated to the second half of the 19<sup>th</sup> century.

## Conclusions/Significance

Our results indicate that two independent migrations from the Central Highlands caused the plague outbreaks in Mahajanga during the 1990s, with both introductions occurring during

**Funding:** LREG and DK are funded by the Max Planck Society through the Max Planck Research Groups program, <https://www.mpg.de/max-planck-research-groups>. XD is funded by the NIHR Health Protection Research Unit program, grant number NIHR200892, <https://www.nihr.ac.uk/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

the early 1980s. They happened over a decade before the detection of human cases, thus the pathogen likely survived in wild reservoirs until the spillover to humans was possible. This study demonstrates the value of Bayesian phylogenetics in elucidating the re-emergence of infectious diseases.

## Author summary

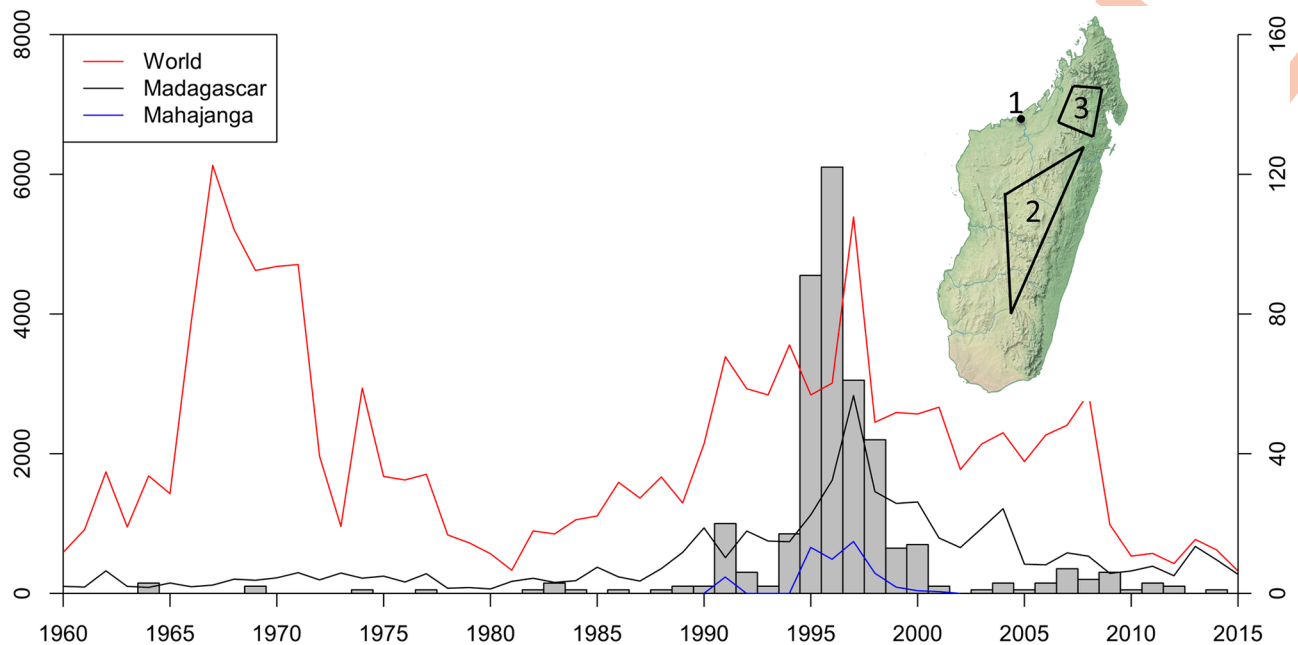
In 1991 human cases of plague were reported in the city of Mahajanga, located in the west coast of Madagascar, after 60 years without human cases. Existing evidence suggests that *Yersinia pestis*, the causal agent of the disease, was reintroduced to the city from a mountainous region known as the Central Highlands. We performed a phylogeographic analysis on 300 *Y. pestis* genome sequences to determine how many migrations of the pathogen between the two locations were related to the reappearance of the disease in Mahajanga. The results revealed that two migrations from the Central Highlands were the cause of the outbreaks of plague in the west coast of the country during the 1990s. We also aimed to date the emergence of the *Y. pestis* variant that circulates in Madagascar and is also present in Turkey. To do this, we conducted a molecular dating analysis using an extended data set of 445 sequences, which contained sequences from Turkey and India (the country from which the pathogen was exported to Madagascar for the first time). The analysis indicated that this particular variant emerged in the first decade of the 20<sup>th</sup> century.

## Introduction

The gram-negative bacterium *Yersinia pestis* is the causal agent of plague, a zoonotic disease transmitted by hematophagous vectors (fleas). Despite being primarily a rodent disease, plague has infected humans for thousands of years [1,2], and has been responsible for at least three severe pandemics throughout human history: the first pandemic or Justinian plague (541–750 CE), the second pandemic that started with the Black Death (14<sup>th</sup>–18<sup>th</sup> Century CE) and the third pandemic that originated in China in 1855 and spread globally from 1894 [3]. Altogether, plague claimed millions of lives during the course of these pandemics.

Currently, over 90% of the plague cases occur in the African continent [4]. The most affected country is Madagascar, an island located on the eastern coast of southern Africa, which accounts for 75% of all infections worldwide [5]. The strains of *Y. pestis* circulating in Madagascar belong to the 1.ORI3 phylogenetic sub-branch, which is part of the *Y. pestis* lineage responsible for the third plague pandemic [3]. According to historical records, plague was introduced in Madagascar when steamboats coming from India arrived at the port city of Toamasina in November 1898 [6,7]. Phylogenetic evidence suggests that plague was exported from Madagascar to the Anatolian peninsula (modern day Turkey) [3].

In decades following its introduction to Madagascar, plague spread through the country establishing three infection foci. Two are located in the Northern and Central Highlands, regions with an altitude of 800 meters above sea level, while the third one is found in the city of Mahajanga, located on the west coast of the country [8,9]. After reaching the capital city Antananarivo in 1921, *Y. pestis* became established in the Central Highlands [10]. In this region, human cases occur every year, most of them reported during the rainy season, from October to April [8,11]. This period coincides with a peak in the population levels of *Y. pestis* vectors (the flea species *Xenopsylla cheopis* and *Synopsyllus fonquerniei*) and with the lowest



**Fig 1. Reported plague cases worldwide, number of sequences used in the study and location of plague foci in Madagascar.** Number of plague cases reported worldwide, for Madagascar and Mahajanga from 1960 to 2014 (left y-axis) [10]. The histogram represents the number of sequences used in the study (right y-axis). Upper right: Location of the three infection foci of plague in Madagascar. 1) Mahajanga, 2) Central Highlands, 3) Northern Highlands. The map was created using QGIS 3.22.2 [17] and uses Natural Earth vector map data from <https://www.naturalearthdata.com/downloads/>.

<https://doi.org/10.1371/journal.pntd.0010362.g001>

abundance of black rats (*Rattus rattus*), the main host of *Y. pestis* in the country [9,12]. In Mahajanga, the first cases of plague were reported in 1902, with outbreaks occurring in 1907 and during the 1920s [7]. After 1928, no human cases were recorded in the city for over six decades until the disease reappeared in August 1991 (Fig 1) [13]. Plague continued to have a strong presence in the city for the rest of the 1990s and caused annual outbreaks until 1999, with the exception of the years of 1993 and 1994 [14]. The high transmission season in humans during these outbreaks ranged from July to November, coinciding with a high abundance of fleas on Asian Shrews (*Suncus murinus*), an important reservoir of *Y. pestis* in Mahajanga [9,15,16].

The sudden reappearance of plague in Mahajanga in 1991 raised the question of whether *Y. pestis* remained circulating in local wild reservoir populations for decades or if it was reintroduced to the city, presumably from the Central Highlands. Previous research based on the discovery of new single nucleotide polymorphisms (SNPs) found evidence of multiple transmissions of *Y. pestis* between the two locations, supporting the role of the Central Highlands as the source of the 1990s outbreaks [9,18]. While it is believed that a single genotype was responsible for the outbreaks in Mahajanga [9], it is not clear how many re-introductions of the bacterium were linked to the re-emergence of plague during the early 1990s, and how much time passed between the re-introduction of *Y. pestis* to Mahajanga and the detection of the first human case in 1991.

In this study, we aimed to better understand the migration dynamics of *Y. pestis* within Madagascar by performing a Bayesian phylogeographic analysis to estimate migration rates and the number and date of migration events of *Y. pestis* lineages between Mahajanga and the Central Highlands. We also performed a molecular dating analysis to estimate the time to the most recent common ancestor (tMRCA) of the 1.ORI3 strains present in Madagascar and Turkey, and to date the split between 1.ORI3 and its ancestral lineage 1.ORI2.a found in India.

This allowed us to generate a timeline for the evolutionary epidemiology of *Y. pestis* starting with the split between 1.ORI2.a and 1.ORI3, followed by the introduction of the bacterium to India, Madagascar and subsequently to Turkey.

## Methods

### Data set

A total of 445 publicly available genomes of *Y. pestis* sampled between 1908 and 2014 from humans, rodents, fleas and shrews, were recovered from the online database EnteroBase (S1 Table) [19]. 440 genomes were sampled from Madagascar, three from Turkey and two from India. The sequences were searched using the Fields “Species” (*Yersinia pestis*), AND “Country” (Madagascar) OR “Country (India) OR “Country (Turkey). Indian sequences that did not belong to the 1.ORI2.a sub-branch were removed, as well as sequences coming from samples that were cultured in laboratories for extended periods of time as they likely present an excess of mutations. All the assembled sequences can be downloaded from EnteroBase after creating an account. All sequence reads are publicly available in the Sequence Read Archive (SRA) of the NCBI (S1 Table). We provide in the next sections background information about the collection of samples and the generation of the sequences.

### Ethics statement

Of the *Y. pestis* strains used in this study 411 were originally isolated from culture of human biological samples at the Central Laboratory for Plague (CLP) and the Plague Unit of the Institut Pasteur of Madagascar (IPM) as part of the Plague National Control Program (PNCP) of the Malagasy Ministry of Public Health. This PNCP requires declaration of all suspected human plague cases and collection of biological samples from those cases for confirmation. These samples were collected under this mandatory reporting system and thus exempt from ethical approval. In addition, all culture isolates and biological samples were de-linked from the patients’ identifiable information and analyzed anonymously. Therefore, no approval from the Malagasy Ethical Committee was required for this study.

For animal sampling, protocols for the care and use of animals were approved by the Institut Pasteur of Madagascar Ad hoc committee in accordance with guidelines of the American Society of Mammalogists [20] and the directive 2010/63/EU of the European Parliament [21].

### Sample collection and *Y. pestis* strain isolation

Biological samples (bubo, sputum and/ or organ punctures) from patients clinically suspected for plague were sent to the CLP at the Plague Unit—IPM for biological confirmation along with a notification form. Rodents and shrews’ spleen were collected from plague active and passive surveillance, stored in a Cary-Blair medium and transported to the CLP. Spleen samples were removed from the transportation medium and crushed using 1 ml of sterile PBS. Fleas were collected in 1 ml of Brain Heart Infusion (BHI) broth and then crushed before analysis.

A rapid diagnostic test (RDT) against F1-antigen (produced by IPM) was performed on each sample [22] and then streaked on selective medium Cefsulodin Irgasan Novobiocin (CIN, Oxoid) agar plate and cultured in peptone broth before incubation at 26°C for more than 48h. In parallel, amplification by intraperitoneal injection of the sample was also performed in two mice which were subsequently observed for 10 days. If the mouse died between 2 and 10 days following injection, the spleen was removed, crushed and cultured on CIN agar plate for isolation of *Y. pestis* strain [23]. Suspicious *Y. pestis* colonies were then enriched on

BHI slant agar media for 24h before biochemical identification using API20E (Biomérieux) strip and bacteriophage lysis test. Isolated *Y. pestis* strains were stored in agar media at room temperature and sent to the Yersinia Research Unit (Institut Pasteur, Paris) with dry ice for genomic sequencing and analysis.

### Sequence generation

The 411 *Y. pestis* isolates from Madagascar together with epidemiological data were sent to the Institut Pasteur (Paris, France) for genomic sequencing. Isolates were grown in LB broth at 28°C for 24 h. Genomic DNA was extracted using PureLink Genomic DNA Mini Kit (ThermoFisher Scientific, Waltham, MA, USA) according to manufacturer's instructions. 100 nucleotides paired-end library were prepared and sequenced on an Illumina platform HiSeq2500 (Illumina, San Diego, CA, USA). The sequences were uploaded to Enterobase between the years 2015 and 2017. The remaining 34 sequences were generated in previous studies [3,9,24].

### SNP analysis

The Enterobase Toolkit (EToKi) was used to align the sequences against the *Y. pestis* reference genome CO92 (accession number AL590842.1), to produce a SNP matrix. In brief, EToKi align module adopted minimap2 to align each genome against the CO92 genome, and extracted SNPs in the aligned regions. The same module also identified repeat regions including CRISPR regions, tandem repeats and interspersed repeats in the CO92 genome using PILER-CR, TRF and a self-comparison by minimap2, respectively. Any SNPs in the repeat regions, or in a region that was shared by  $\leq 95\%$  of genomes were discarded. The resulting core genomic, non-repetitive SNP matrix contained 783 sites representing 143 different genotypes (S2 Table).

While the genotype of the Indian sequences had been determined by a previous study [3], the presence in the matrix of 226 previously described SNPs related to 1.ORI3 (S3 Table) confirmed the genotype of the remaining sequences. This entire matrix was used to generate a time calibrated phylogeny of *Y. pestis*. For the phylogeographic analysis only sequences sampled from humans were used to avoid possible bias introduced by having samples from hosts with different travel ranges and capacities. This human-specific data set consisted of 300 sequences collected between 1964 and 2012, including 183 from the Central Highlands and 117 from Mahajanga. Sequences with duplicated genotypes were kept in the datasets as they inform the evolutionary and demographic processes that have occurred through time. However, identical sequences do not contain information regarding the topology and branch lengths of a subtree of identical sequences, those are the result of the underlying stochastic process of tree generation combined with the branch rate variation estimated from the entire data set [25].

### Test for phylogenetic signal

All Bayesian phylogenetic analyses were done with BEAST 2 [26], a platform developed for the inference of time calibrated binary trees. The evaluation of temporal signal for the two data sets was done with the BETS approach [27]. In this method, the marginal likelihoods of a heterochronous model (M1), with the real sampling dates, and an isochronous model (M2), with all dates set to zero and a fixed molecular clock rate, are estimated in a Path Sampling analysis [28]. A positive log Bayes Factor value (log marginal likelihood M1—log marginal likelihood M2) indicates the presence of a temporal signal [27,29]. The Path Sampling analysis ran for 50 steps with a Markov Chain Monte Carlo (MCMC) length of 8,000,000 each.

### Phylogeographic analysis of *Y. pestis*

The phylogeographic analysis was done with the BEAST 2 package *bdmm* [30], which uses a structured (also called “multi-type”) birth–death model for the tree-generating process. Under this approach each sequence is assigned to a “type” or subpopulation and the type changes are estimated, following a type change rate  $m$ , jointly with the phylogenetic reconstruction. This analysis produces a multi-type tree, in which lineages of different types are annotated by different colors, and the migrations (type changes) are indicated by color changes occurring along the branches. Migration rates per lineage per year and the diversification-extinction ratio, which represents the number of new lineages generated per year and is also known as the reproductive number ( $R$ ) in epidemiological contexts, are also estimated for each type.

We performed three replicate analyses with a MCMC of 300 million steps using a relaxed lognormal molecular clock and the HKY model of nucleotide substitutions. Sequences from Mahajanga were assigned to the type “0” and sequences from the Central Highlands to the type “1”. Given the evidence supporting the absence of plague in Mahajanga before 1991, we fixed the migration rates to 0 before 1980. This time point was selected since an important increment in the number of plague cases in the Central Highlands started to occur during the early 1980s [31]. The same setup was applied for the estimation of the diversification-extinction ratio for Mahajanga (since there are no sequences sampled before 1991), however the value before 1980 was allowed to differ from 0 for the Central Highlands. In order to account for the uneven number of sequences per year observed in the data set, we estimated a different sampling proportion for each year and type. A Beta distribution with a mean equal to the number of samples divided by the number of reported cases for each year, was chosen as the mean value of the Beta prior distribution used for each sampling proportion value. This allowed us to use plague incidence data to model the sampling process through time. Tip dates were sampled using a uniform distribution over the sampling year since it is very unlikely that all the sequences from one year were sampled on the same day. To improve MCMC mixing, the clock rate was fixed to a value of  $2.85 \times 10^{-8}$ , obtained in a previous study for the branch 1 of *Y. pestis* [32]. As only SNP alignments were used, we corrected for ascertainment bias by specifying the number of invariable sites for each of the four bases, as found in the CO92 reference genome. The rate parameter for becoming uninfected ( $\delta$ ) was fixed to a value of 1 to avoid identifiability issues with the other parameters [33].

The output files were combined with the software LogCombiner [34], after removing a 10% burn-in. This yielded effective sample sizes above the standard threshold of 200 for all parameters, as assessed by the software Tracer v.1.7 [35]. The maximum a posteriori tree (MAP), which represents the tree with the greatest posterior probability in the combined runs, was used for the visual representation of the migrations between foci, as it can display the type changes along a branch. The tree was visualized and edited with FigTree 1.4.4 [36]. As migration times on the MAP tree represent only point estimates, the median and 95% posterior intervals of migration times were calculated for each migration using a custom R script.

### Molecular dating analysis

The estimation of divergence times was done using a relaxed lognormal molecular clock, the HKY model of nucleotide substitutions and the Bayesian birth-death skyline tree prior implemented in the BEAST 2 package *BDSKY* [33]. The birth-death skyline model follows a forward in time birth-death process with only one lineage existing at time 0 in the past. As we move closer to the present, the lineages can split (birth event) at a rate  $\lambda$ , go extinct (death event) at a rate  $\mu$ , and be sampled at a rate  $\Psi$ . These rates are reparametrized to estimate parameters like

the sampling proportion ( $s$ ), the become uninfected rate ( $\delta$ ) and the diversification-extinction ratio [33]. The sampling proportion prior was specified per year, and tip date sampling and the correction for ascertainment bias were also applied in this analysis. The rate parameter for becoming uninfected ( $\delta$ ) was also fixed to a value of 1.

Three replicate analyses were done with a MCMC length of 480 million states, and output files were later combined with LogCombiner, after removing a 10% burn-in. All parameters obtained effective sample sizes above 200. A maximum clade credibility tree was obtained for the combined runs using the program TreeAnnotator [37] with a 10% burn-in. The tree was edited with FigTree 1.4.4. The BEAST 2 XML files, output log files and the MAP and MCC tree files are provided as supporting information (S1–S6 Appendices).

## Results

### Migration dynamics of *Y. pestis*

We conducted a phylogeographic analysis based on the 300 human sequences from Madagascar. The temporal signal was significant for this subset with a BETS analysis giving a Bayes Factor of 52. According to the results of the bdm analysis, the migration rate from the Central Highlands to Mahajanga after 1980 had a median value of  $1.26 \times 10^{-3}$  per lineage per year (95% HPD =  $2.64 \times 10^{-4}$ – $2.78 \times 10^{-3}$ ), and the rate from Mahajanga to the Central Highlands was  $4.68 \times 10^{-4}$  per lineage per year (95% HPD =  $7.71 \times 10^{-6}$ – $1.8 \times 10^{-3}$ ). The diversification-extinction ratio was greater for Mahajanga with a median estimate of 5.37 (95% HPD = 4.72–6). For the Central Highlands the median values were 1.18 (95% HPD = 1.04–1.37) before 1980, and 2.98 (95% HPD = 2.65–3.27) afterwards (Table 1).

The MAP tree used to summarize the migration history displayed five migration events, four of which were from the Central Highlands to Mahajanga (Fig 2). The first migration occurred around 1981 (95% HPD = 1980–1983) and originated a clade of sequences sampled in Mahajanga between 1995 and 1999. The second migration happened a year later (95% HPD = 1980–1985) and gave rise to a clade containing sequences sampled in 1991 and 1992. The additional migrations from the Central Highlands to Mahajanga had median estimates of 1989 (95% HPD = 1982–1996) and 1990 (95% HPD = 1985–1995) but each led to only a single sequence sampled in 1995 and 1997, respectively. Therefore, it seems that these introductions did not establish transmission chains in humans. The only migration from Mahajanga to the Central Highlands took place around 1996 (95% HPD = 1992–1997) leading to a single genome sampled in the capital Antananarivo in 1997.

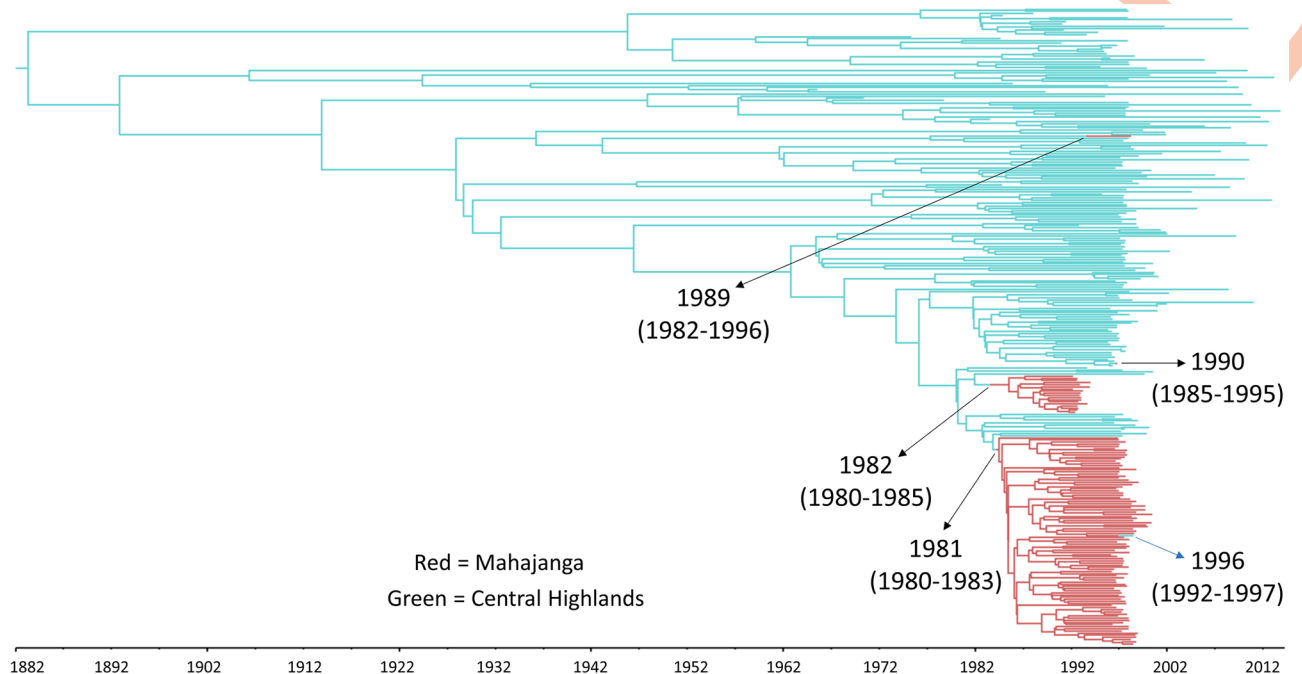
### Molecular dating of *Y. pestis*

The BETS analysis for the full data set of 445 genomes of *Y. pestis* resulted in a log Bayes Factor of 203, which indicates a significant level of temporal signal. The BEAST 2 molecular dating

**Table 1. Diversification-extinction ratio estimated for each infection focus, and migration rates of *Y. pestis* lineages per year from Mahajanga to the Central Highlands and vice-versa with the 95% highest posterior density intervals shown in brackets.**

	Diversification-extinction ratio
Mahajanga	Post-1980: 5.37 [4.72–6]
Central Highlands	Pre-1980: 1.18 [1.04–1.37]
	Post-1980: 2.98 [2.65–3.27]
	Migration rates per lineage per year
Mahajanga to Central Highlands	$4.68 \times 10^{-4}$ [ $7.71 \times 10^{-6}$ – $1.8 \times 10^{-3}$ ]
Central Highlands to Mahajanga	$1.26 \times 10^{-3}$ [ $2.64 \times 10^{-4}$ – $2.78 \times 10^{-3}$ ]

<https://doi.org/10.1371/journal.pntd.0010362.t001>



**Fig 2. Maximum a posteriori typed tree of 300 *Y. pestis* sequences sampled in Mahajanga and the Central Highlands.** Posterior median estimates for the dates of the migration events are shown with their respective 95% highest posterior density intervals (95% HPD). The four black arrows indicate migrations from the Central highlands to Mahajanga. The single blue arrow indicates a migration from Mahajanga to the Central Highlands. This is a binary rooted time tree, with branch lengths proportional to time.

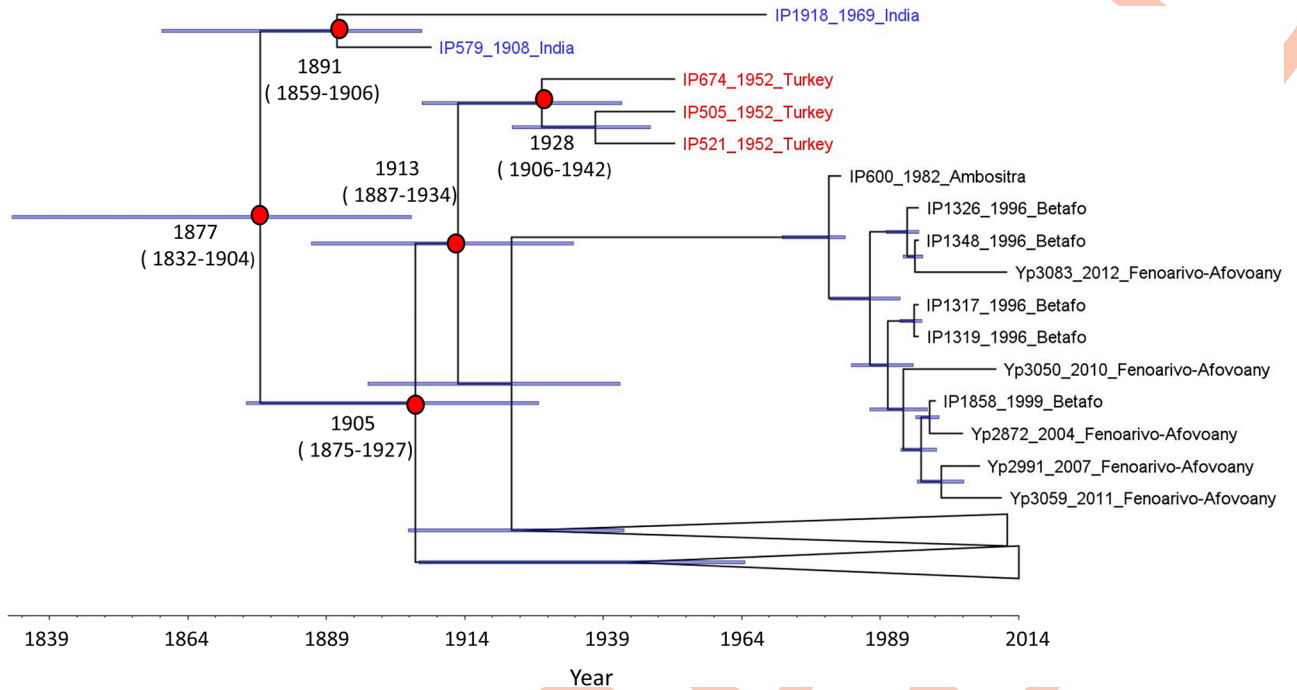
<https://doi.org/10.1371/journal.pntd.0010362.g002>

analysis estimated a median evolutionary rate of  $4.924 \times 10^{-8}$  substitutions per site per year (95% HPD =  $3.9 \times 10^{-8}$ – $6.11 \times 10^{-8}$ ). The coefficient of variation of the clock rate had a median of 1.35 (95% HPD = 1.02–1.76), which indicates an important deviation from the strict clock model and a high degree of rate variation among the sequences. The tMRCA of all the sequences was estimated in 1877 (95% HPD = 1832–1904). The two Indian sequences shared a common ancestor in 1891 (95% HPD = 1859–1906). The most recent common ancestor of all sequences from Madagascar and Turkey (i.e., the L.ORI3 sub-branch) was dated to 1905 (95% HPD = 1875–1927), while the divergence of the Turkish branch was dated to 1913 (95% HPD = 1887–1934). Lastly the three sequences from Turkey shared a most recent common ancestor in 1928 (95% HPD = 1906–1942) (Fig 3).

## Discussion

In this study we conducted a phylogeographic analysis of 300 *Y. pestis* genomes. The results revealed a higher migration rate of *Y. pestis* from the Highlands to Mahajanga, which supports the role of the former as the “source” of the *Y. pestis* populations that caused the outbreaks in the coastal city. Additional evidence in favor of a reintroduction rather than a reemergence of *Y. pestis* in Mahajanga include the lack of reports of massive deaths among rat populations during the period in which no human cases were reported in Mahajanga [9]. Also, while rodent populations of the black and brown rat (*Rattus norvegicus*) from the Central Highlands exhibit some degree of plague resistance, this is not the case for rats living in the west coast [38]. On the other hand, Asian shrews, currently the predominant plague reservoirs in Mahajanga exhibit a high resistance to *Y. pestis* [16].





**Fig 3. Maximum clade credibility tree of 445 sequences of *Yersinia pestis*.** Sequences from India indicated with blue, sequences from Turkey with red. Posterior median estimates for the ages of the relevant nodes are shown with their respective 95% highest posterior density intervals (95% HPD). Error bars for the node heights are shown. Lower part of the phylogeny is collapsed to help with the visualization of the tree. This is a binary rooted time tree, with branch lengths proportional to time.

<https://doi.org/10.1371/journal.pntd.0010362.g003>

Interestingly, most of the sequences from Mahajanga formed two distinctive clusters in the MAP tree, and each originated from different reintroductions of the pathogen that occurred in the early 1980s. This result suggests that two independent migrations were responsible for the establishment of the *Y. pestis* lineages that successfully started chains of transmission in humans.

The early 1980s was a period with difficult climatic conditions in Madagascar, with a handful of tropical cyclones and intense storms hitting the Central Highlands in 1981 and 1982. Cyclone Justine also affected Mahajanga in March of 1982 [39]. The floods caused the destruction of roads and crop fields, and forced the evacuation of thousands of people. The constant movement of a high number of people across the country, potentially carrying infected fleas and/or rats, could have facilitated the re-introduction of *Y. pestis* to Mahajanga.

Curiously, it was the second migration event who originated the first outbreak of the decade, which started in August 1991 and ended seven months later [14], while the first migration of *Y. pestis* from the Central Highlands was the most successful, even if the introduced lineages jumped to human populations later. This result suggests that rodents infected by *Y. pestis* lineages introduced in the second migration were the first to reach, maybe just by chance, urban zones of Mahajanga that provided a good environment for rodent proliferation (e.g. abundant food, bad hygienic and economic conditions) and greater chances of close contact with humans.

Both migrations occurred over a decade before the sampling of the oldest sequence in each clade, so it is very likely that the bacterium circulated in wild rodent or shrew populations before it jumped to humans, first in 1991 and a second time in 1995. After 1999, no human cases have been reported in Mahajanga, but unlike the disappearance of the disease after 1928,

it is believed that *Y. pestis* continues to be present in the city, circulating in shrews with occasional spillover events to rats [16].

A previous study determined that there was a cycle of generation and loss of diversity in Mahajanga during the 1990s outbreaks with many lineages disappearing at the end of an epidemic year and new ones emerging during the next year [9]. This continued generation of lineages in Mahajanga could explain the greater diversification-extinction ratio found for this location compared to the Central Highlands. The greater diversification-extinction ratio of *Y. pestis* in Mahajanga, can also be related to the different environmental conditions of the two locations. The outbreaks in Mahajanga were centered around big markets with bad hygienic conditions, a good habitat to support a large number of rodents [9]. In the Highlands, plague is still considered a rural disease and the transmission of *Y. pestis* from rats to humans is linked in most cases to agricultural activities with low and high transmission seasons in humans present every year. The high transmission season in humans coincides with a peak in the flea population size and a low abundance of rats, as a result of the scarcity of food and the low temperatures that follow the harvest period, and the rat deaths caused by the disease itself [13]. Overall, it is believed that the maintenance of plague in Madagascar is the result of endemic cycles of transmission within the different districts of the country, mostly located in the Central Highlands [24].

While the results found here support previous claims of a reintroduction of *Y. pestis* to Mahajanga from the Central Highlands, it is also possible that additional migrations of this pathogen occurred from the Northern Highlands, the third plague focus in the country, and from other regions of Madagascar. In this regard, a recent study detected positive samples of *Y. pestis* in fleas sampled in the southwest part of the country [40]. While most of the samples were isolated from a flea species (*Echidnophaga gallinacea*) not suitable to be an efficient vector, the presence of the pathogen in an area far away from the plague foci suggests that the transmission of *Y. pestis* among non-human hosts is more widespread than previously thought. However, hypotheses of *Y. pestis* migrations to or from regions other than the Central Highlands couldn't be tested with our dataset given the limited number of available sequences (<10) sampled outside the two main foci discussed here.

In this study we also generated a dated phylogeny of 445 *Y. pestis* sequences belonging to the ORI (Orientalis) biovar, collected from India (1.ORI2.a), Madagascar and Turkey (1.ORI3). The 1.ORI2.a sub-branch was exported from Hong Kong to India in 1896 [3,41] and represents the ancestor node of 1.ORI3, which is only found in Madagascar and Turkey. We estimated a median tMRCA for the 1.ORI2.a and 1.ORI3 sub-branches of 1877. This result indicates that the diversification of ORI likely occurred after the beginning of the third pandemic in 1855, as previously proposed [3], and was fueled by continuous exportations of the pathogen from Asia via sea trade.

The common ancestor of 1.ORI2.a was dated to 1891, 5 years before its arrival to India from Hong Kong. However, the 95% HPD intervals did contain the historical introduction year of 1896, supporting the arrival of plague to India at the end of the 19<sup>th</sup> century. The estimate of 1905 for the most recent common ancestor of the sequences from Madagascar and Turkey suggests that the 1.ORI3 emerged around 7 years after the introduction of *Y. pestis* to the island in 1898. The appearance of this sub-branch within the first 10 years after the arrival of the pathogen to Madagascar could be the result of *Y. pestis* rapidly spreading and adapting to local reservoirs. The fact that no 1.ORI3 sequences have been detected outside Madagascar and Turkey, fit with a scenario in which 1.ORI2.a was introduced to Madagascar in 1898 with 1.ORI3 emerging within the next decade. A previous study estimated a median tMRCA of 1954 for Madagascan strains, although using only 33 genomes [24].

After reaching Madagascar, 1.ORI3 spread to Turkey probably in the early 1930s, as historical records mention the presence of two plague-infected individuals, coming from Madagascar, in the Middle East around 1931[3]. The results presented here support this claim, with the three sequences from Turkey having a median tMRCA of 1928.

The evolutionary rate obtained in this study for the sub-branches 1.ORI2.a and 1.ORI3 is faster than the estimate for the Branch 1 of *Y. pestis* used in the phylogeographic analysis, which was obtained using sequences from the biovars 1.ANT (Antiqua), 1.IN (Intermediate) and a set of ancient genomes from the Second Pandemic [32], in addition to 1.ORI. This difference could be caused by the different time spans covered by the sequences in both studies. The sampling dates of the data set used in this study span 116 years, while some of the Second Pandemic genomes are over 600 years old. In this regard, a slower clock rate is expected in longer time spans as the effect of purifying selection becomes stronger [42]. This indicates that the slower rate, used to estimate the dates of the migrations, is likely closer to the true evolutionary rate of the branch 1 of *Y. pestis*. Moreover, assuming a higher evolutionary rate would not change the results of the phylogeographic analysis qualitatively.

Taking into account the results presented here, our study provides a timeline of the diversification of *Y. pestis* from the arrival of the 1.ORI2.a sub-branch to India in the late 19<sup>th</sup> century to the spread of 1.ORI3 to Turkey a few decades later. This study, also presents the first application of a structured birth-death migration model to the study of *Y. pestis*, which allowed the dating of migration events and estimation of migration rates, while enabling the detailed modeling of the sampling process with the use of incidence data. Furthermore, if more *Y. pestis* sequences sampled from rodent and other small mammal species become available, a similar analysis could be done using a more complex model to account for the circulation of the bacteria in non-human hosts to assess their role in the spread and re-emergence of plague.

## Supporting information

**S1 Table. *Yersinia pestis* samples used in this study**  
(CSV)

**S2 Table. SNP matrix generated with the 445 *Yersinia pestis* genomes**  
(CSV)

**S3 Table. SNPs present in the 1.ORI3 subbranch of *Yersinia pestis***  
(CSV)

**S1 Appendix. BEAST2 xml file used for the bdm analysis**  
(XML)

**S2 Appendix. Combined log file obtained with bdm**  
(LOG)

**S3 Appendix. MAP tree obtained with bdm.** File can be open with a phylogenetic tree visualization tool, like Figtree or IcyTree.  
(TREE)

**S4 Appendix. BEAST2 xml file used for the BDSKY analysis**  
(XML)

**S5 Appendix. Combined log file obtained with BDSKY**  
(LOG)

**S6 Appendix. MCC tree obtained with BDSKY.** File can be open with a phylogenetic tree visualization tool, like Figtree or IcyTree.  
(TREE)

## Acknowledgments

We would like to thank Mark Achtman and Elisabeth Carniel for supervising the acquisition of the sequences and for their feedback during the project conceptualization. We thank Michelle O'Reilly for providing the Madagascar map for Fig 1. We also thank Johannes Krause, Kirsten Bos, Alexander Herbig and Maria Spyrou for their valuable inputs and suggestions for the analyses, and Sanni Översti and Ariane Weber for their comments on the manuscript.

## Author Contributions

**Conceptualization:** Luis Roger Esquivel Gomez, Denise Kühnert.

**Data curation:** Cyril Savin, Voahangy Andrianaivoarimanana, Soloandry Rahajandraibe, Lovasoa Nomena Randriantseheno, Zhemín Zhou, Minoarisoa Rajerison.

**Formal analysis:** Luis Roger Esquivel Gomez, Arthur Kocher, Denise Kühnert.

**Funding acquisition:** Xavier Didelot, Denise Kühnert.

**Investigation:** Luis Roger Esquivel Gomez, Cyril Savin, Voahangy Andrianaivoarimanana, Soloandry Rahajandraibe, Lovasoa Nomena Randriantseheno, Arthur Kocher, Xavier Didelot, Minoarisoa Rajerison, Denise Kühnert.

**Methodology:** Zhemín Zhou, Denise Kühnert.

**Project administration:** Luis Roger Esquivel Gomez, Voahangy Andrianaivoarimanana, Soloandry Rahajandraibe, Lovasoa Nomena Randriantseheno, Minoarisoa Rajerison, Denise Kühnert.

**Resources:** Denise Kühnert.

**Supervision:** Denise Kühnert.

**Validation:** Luis Roger Esquivel Gomez.

**Visualization:** Luis Roger Esquivel Gomez.

**Writing – original draft:** Luis Roger Esquivel Gomez.

**Writing – review & editing:** Cyril Savin, Voahangy Andrianaivoarimanana, Soloandry Rahajandraibe, Lovasoa Nomena Randriantseheno, Zhemín Zhou, Arthur Kocher, Xavier Didelot, Minoarisoa Rajerison, Denise Kühnert.

## References

1. Andrades Valtueña A, Mitnik A, Key FM, Haak W, Allmæe R, Belinskij A, et al. The Stone Age Plague and Its Persistence in Eurasia. *Curr Biol*. 2017; 27:3683–3691.e8. <https://doi.org/10.1016/j.cub.2017.10.025> PMID: 29174893
2. Spyrou MA, Tukhbatova RI, Wang CC, Valtueña AA, Lankapalli AK, Kondrashin VV. et al. Analysis of 3800-year-old *Yersinia pestis* genomes suggests Bronze Age origin for bubonic plague. *Nat Commun*. 2018; 9, 2234. <https://doi.org/10.1038/s41467-018-04550-9> PMID: 29884871
3. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Wagner DM, Feldkamp M, et al. Phylogenetic diversity and historical patterns of pandemic spread of *Yersinia pestis*. *Nat Genet*. 2010; 42(12):1140–3.

4. Andrianaivoarimanana V, Piola P, Wagner DM, Rakotomanana F, Maheriniaina V, Andrianalimanana S, et al. Trends of Human Plague, Madagascar, 1998–2016. *Emerg Infect Dis*. 2019; 25(2):220–8. <https://doi.org/10.3201/eid2502.171974> PMID: 30666930
5. World Health Organization. Plague Around the World. *Weekly Epidemiological Record*. 2019; 94 (25):289–292.
6. Pollitzer R. Plague studies. 1. A summary of the history and survey of the present distribution of the disease. *Bull World Health Organ*. 1951; 4(4):475–533. PMID: 14925809
7. Brygoo ER. Epidémiologie de la peste à Madagascar. *Arch Inst Pasteur Madagascar*. 1966; 35: 9–147.
8. Chanteau S, Ratsifasoamanana L, Rasoamanana B, Rahalison L, Randriambelosoa J, Roux J, et al. Plague, a reemerging disease in Madagascar. *Emerg Infect Dis*. 1998; 4(1):101–4. <https://doi.org/10.3201/eid0401.980114> PMID: 9452403
9. Vogler AJ, Chan F, Nottingham R, Andersen G, Drees K, Beckstrom-Sternberg SM, et al. A decade of plague in Mahajanga, Madagascar: Insights into the global maritime spread of pandemic plague. *MBio*. 2013; 4(1):e00623–12. <https://doi.org/10.1128/mBio.00623-12> PMID: 23404402
10. Chanteau S. Atlas de la peste à Madagascar. Marseille: IRD Éditions; 2006.
11. Chanteau S, Ratsitorahina M, Rahalison L, Rasoamanana B, Chan F, Boisier P, et al. Current epidemiology of human plague in Madagascar. *Microbes Infect*. 2000; 2(1):25–31. [https://doi.org/10.1016/s1286-4579\(00\)00289-6](https://doi.org/10.1016/s1286-4579(00)00289-6) PMID: 10717537
12. Migliani R, Chanteau S, Rahalison L, Ratsitorahina M, Boutin JP, Ratsifasoamanana L, et al. Epidemiological trends for human plague in Madagascar during the second half of the 20th century: A survey of 20 900 notified cases. *Trop Med Int Heal*. 2006; 11(8):1228–37. <https://doi.org/10.1111/j.1365-3156.2006.01677.x> PMID: 16903886
13. Andrianaivoarimanana V, Kreppel K, Elissa N, Duplantier JM, Carniel E, Rajerison M, et al. Understanding the Persistence of Plague Foci in Madagascar. *PLoS Negl Trop Dis*. 2013; 7(11): e2382. <https://doi.org/10.1371/journal.pntd.0002382> PMID: 24244760
14. Boisier P, Rahalison L, Rasolomaharo M, Ratsitorahina M, Mahafaly, Razafimahefa M, et al. Epidemiologic features of four successive annual outbreaks of bubonic plague in Mahajanga, Madagascar. *Emerg Infect Dis*. 2002; 8(3):311–6. <https://doi.org/10.3201/eid0803.010250> PMID: 11927030
15. Duplantier JM, Duchemin JB, Chanteau S, Carniel E. From the recent lessons of the Malagasy foci towards a global understanding of the factors involved in plague reemergence. *Vet Res*. 2005; 36(3):437–53. <https://doi.org/10.1051/vetres:2005007> PMID: 15845233
16. Rahelinirina S, Rajerison M, Telfer S, Savin C, Carniel E, Duplantier JM. The Asian house shrew *Suncus murinus* as a reservoir and source of human outbreaks of plague in Madagascar. *PLoS Negl Trop Dis*. 2017; 11(11): e0006072. <https://doi.org/10.1371/journal.pntd.0006072> PMID: 29155827
17. QGIS.org, 2022. QGIS Geographic Information System. QGIS Association. <http://www.qgis.org>
18. Vogler AJ, Chan F, Wagner DM, Roumagnac P, Lee J, Nera R, et al. Phylogeography and molecular epidemiology of *Yersinia pestis* in Madagascar. *PLoS Negl Trop Dis*. 2011; 5(9):e1319. <https://doi.org/10.1371/journal.pntd.0001319> PMID: 21931876
19. Zhou Z, Alikhan NF, Mohamed K, Fan Y, Achtman M. The Enterobase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia coli* genomic diversity. *Genome Res*. 2020; 30(1):138–52. <https://doi.org/10.1101/gr.251678.119> PMID: 31809257
20. Sikes RS. Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy*. 2016; 97(3):663–88.
21. European Parliament and the Council of the European Union. Directive 2010/63/EU of the European Parliament and of the Council of 22 september 2010 on the protection of animals used for scientific purposes. Official Journal of the European Union. [Internet]. [cited 2022 Jun 21]. <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:EN:PDF>.
22. Chanteau S, Rahalison L, Ralafiarisoa L, Foulon J, Ratsitorahina M, Ratsifasoamanana L, et al. Development and testing of a rapid diagnostic test for bubonic and pneumonic plague. *Lancet*. 2003; 361:211–21 [https://doi.org/10.1016/S0140-6736\(03\)12270-2](https://doi.org/10.1016/S0140-6736(03)12270-2) PMID: 12547544
23. Rasoamanana B, Rahalison L, Raharimanana C, Chanteau S. Comparison of *Yersinia* CIN agar and mouse inoculation assay for the diagnosis of plague. *Trans R Soc Trop Med Hyg*. 1996; 90(6):651. [https://doi.org/10.1016/S0035-9203\(96\)90420-4](https://doi.org/10.1016/S0035-9203(96)90420-4) PMID: 9015505
24. Vogler AJ, Andrianaivoarimanana V, Telfer S, Hall CM, Sahl JW, Hepp CM, et al. Temporal phylogeography of *Yersinia pestis* in Madagascar: Insights into the long-term maintenance of plague. *PLoS Negl Trop Dis*. 2017; 11(9):e0005887. <https://doi.org/10.1371/journal.pntd.0005887> PMID: 28873412
25. Boskova V, Stadler T. PIQME: Bayesian Phylodynamic Method for Analysis of Large Data Sets with Duplicate Sequences. *Mol Biol Evol*, 2020; 37(10):3061. <https://doi.org/10.1093/molbev/msaa136> PMID: 32492139

26. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*. 2019; 15(4):e1006650. <https://doi.org/10.1371/journal.pcbi.1006650> PMID: 30958812
27. Duchene S, Lemey P, Stadler T, Ho SYW, Duchene DA, Dhanasekaran V, et al. Bayesian evaluation of temporal signal in measurably evolving populations. *Mol Biol Evol*. 2020; 37(11):3363–79. <https://doi.org/10.1093/molbev/msaa163> PMID: 32895707
28. Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. *Syst Biol*. 2006; 55(2):195–207. <https://doi.org/10.1080/10635150500433722> PMID: 16522570
29. Kass RE, Raftery AE. Bayes Factors. *J Am Stat Assoc*. 1995; 90(430):773.
30. Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Phylodynamics with Migration: A Computational Framework to Quantify Population Structure from Genomic Data. *Mol Biol Evol*. 2016; 33(8):2102–16. <https://doi.org/10.1093/molbev/msw064> PMID: 27189573
31. Ratsitorahina M, Migliani R, Ratsifasoamanana L, Ratsimba M, Chan Ho Thin F, Rahalison L, et al. Recrudescence et extension géographique de la peste à Madagascar de 1980 à 1999 *Arch Inst Pasteur Madagascar*. 2001; 67(1–2):14–8.
32. Spyrou MA, Keller M, Tukhbatova RI, Scheib CL, Nelson EA, Andrades Valtueña A, et al. Phylogeography of the second plague pandemic revealed through analysis of historical *Yersinia pestis* genomes. *Nat Commun*. 2019; 10, 4470. <https://doi.org/10.1038/s41467-019-12154-0> PMID: 31578321
33. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A*. 2013; 110(1):228–33. <https://doi.org/10.1073/pnas.1207965110> PMID: 23248286
34. LogCombiner. BEAST Documentation [Internet]. [cited 2021 Dec 9]. <https://beast.community/logcombiner>.
35. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol*. 2018; 67(5):901–4. <https://doi.org/10.1093/sysbio/syy032> PMID: 29718447
36. FigTree. BEAST Documentation [Internet]. [cited 2021 Dec 9]. <https://beast.community/figtree>.
37. TreeAnnotator. BEAST 2 [Internet]. [cited 2021 Dec 9]. <https://www.beast2.org/treeannotator/>.
38. Tollenaere C, Rahalison L, Ranjalahy M, Duplantier JM, Rahelinirina S, Telfer S, et al. Susceptibility to *Yersinia pestis* experimental infection in wild *Rattus rattus*, reservoir of plague in Madagascar. *Eco-health*. 2010; 7(2):242–7. <https://doi.org/10.1007/s10393-010-0312-3> PMID: 20443044
39. Donque G. La saison cyclonique 1981–1982. *Mad. Rev. de Geo*. 1982; 41:121–125.
40. Ehlers J, Krüger A, Rakotondrany SJ, Ratvonamana RY, Poppert S, Ganzhorn JU, et al. Molecular detection of *Rickettsia* spp., *Borrelia* spp., *Bartonella* spp. and *Yersinia pestis* in ectoparasites of endemic and domestic animals in southwest Madagascar. *Acta Trop*. 2020; 205:105339. <https://doi.org/10.1016/j.actatropica.2020.105339> PMID: 31935354
41. Yu HL, Christakos G. Spatiotemporal modelling and mapping of the bubonic plague epidemic in India. *Int J Health Geogr*. 2006; 5:12. <https://doi.org/10.1186/1476-072X-5-12> PMID: 16545128
42. Ho SY, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. Evidence for time dependency of molecular rate estimates. *Syst Biol*. 2007; 56(3):515–22. <https://doi.org/10.1080/10635150701435401> PMID: 17562475