Ecologic, Geoclimatic, and Genomic Factors Modulating Plague Epidemics in Primary Natural Focus, Brazil

Article *in* Emerging Infectious Diseases · September 2024

DOI: 10.3201/eid3009.240468

13 authors, including:

Matheus Filgueira Bezerra
Oswaldo Cruz Foundation
57 PUBLICATIONS **384** CITATIONS
SEE PROFILE

João Luiz de Lemos Padilha Pitta
Oswaldo Cruz Foundation
5 PUBLICATIONS **13** CITATIONS
SEE PROFILE

Natan Diego Alves de Freitas
Federal University of Paraíba
8 PUBLICATIONS **14** CITATIONS
SEE PROFILE

Elainne Christine Souza Gomes
Fundação Oswaldo Cruz - Recife, Brazil - Instituto Aggeu Magalhães
63 PUBLICATIONS **553** CITATIONS
SEE PROFILE

Ecologic, Geoclimatic, and Genomic Factors Modulating Plague Epidemics in Primary Natural Focus, Brazil

Matheus F. Bezerra, Diego L.R.S. Fernandes, Igor V. Rocha, João L.L.P. Pitta, Natan D.A. Freitas, André L.S. Oliveira, Ricardo J.P.S. Guimarães, Elainne C.S. Gomes, Cecília Siliansky de Andreazzi, Marise Sobreira, Antonio M. Rezende, Pedro Cordeiro-Estrela, Alzira M.P. Almeida

Plague is a deadly zoonosis that still poses a threat in many regions of the world. We combined epidemiologic, host, and vector surveillance data collected during 1961–1980 from the Araripe Plateau focus in northeastern Brazil with ecologic, geoclimatic, and *Yersinia pestis* genomic information to elucidate how these factors interplay in plague activity. We identified well-delimited plague hotspots showing elevated plague risk in low-altitude areas near the foothills of the plateau's concave sectors. Those locations exhibited distinct precipitation and vegetation coverage patterns compared with the surrounding areas. We noted a seasonal effect on plague activity, and human cases linearly correlated with precipitation and rodent and flea *Y. pestis* positivity rates. Genomic characterization of *Y. pestis* strains revealed a foundational strain capable of evolving into distinct genetic variants, each linked to temporally and spatially constrained plague outbreaks. These data could identify risk areas and improve surveillance in other plague foci within the Caatinga biome.

Plague is a deadly bacterial disease caused by *Yersinia pestis* and is implicated in major pandemics throughout the course of human history. Despite the decline in global incidence, plague outbreaks still occur in regions where the bacterium maintains a sylvatic cycle (*1*,*2*). In addition, plague resurgence has been reported after long periods of quiescence, making active animal surveillance in its natural foci critical to prevent outbreaks in human populations (*3*).

Plague incidence typically undergoes accentuated variations over time (*2*). Evidence from studies performed in central Asia and North America demonstrate that such variations are triggered by the combined effect of geoclimatic and ecologic factors. Temperature, precipitation, landscape, vegetation, soil composition, and host–vector densities and diversity have been reported to interplay in a trophic web that leads to plague dissemination in both wild fauna and humans (*3*–*5*). However, information regarding the role the geoclimatic and ecologic aspects in the modulation of plague systems in foci of Brazil is lacking.

The current demarcation of plague focal areas in Brazil lacks granularity and encompasses extensive regions solely based on geologic formations with reported plague cases (*6*). Thus, identification of specific geoclimatic and environmental patterns indicating hotspots for increased plague risk could provide valuable information for implementing more assertive and targeted animal surveillance.

Author affiliations: Instituto Aggeu Magalhães, Fiocruz, Brazil (M.F. Bezerra, D.L.R.S Fernandes, I.V. Rocha, J.L.L.P. Pitta, E.C.S. Gomes, M. Sobreira, A.M. Rezende, A.M.P. Almeida); Laboratório de Mamíferos, Pós Graduação em Ciências Biológicas (Zoologia), Universidade Federal da Paraíba, João Pessoa, Brazil (N.D.A. Freitas, P. Cordeiro-Estrela); Núcleo de Geoprocessamento, Instituto Aggeu Magalhães, Fiocruz (A.L.S. Oliveira); Laboratório de Geoprocessamento, Instituto Evandro Chagas, Belém, Brazil (R.J.P.S. Guimarães); Laboratório de Biologia e Parasitologia de Mamíferos Silvestres Reservatórios, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, Brazil (C.S. de Andreazzi); Universidad Complutense de Madrid, Madrid, Spain (C.S. de Andreazzi); International Platform for Science, Technology and Innovation in Health, PICTIS, Ílhavo, Portugal (C.S. de Andreazzi); Group of Biotechnology Applied to Pathogens, René Rachou Institute, Fiocruz (A.M. Rezende)

DOI: https://doi.org/10.3201/eid3009.240468

The Araripe Plateau (Chapada do Araripe) constitutes a geologic landmark situated along the border of Pernambuco, Ceará, and Piauí states of Brazil. Although the region is in the core of the semiarid biome known as Caatinga, the plateau is topped by savanna, and its slopes have lush forest vegetation because of orographic precipitation and a permeable tabletop (*7*,*8*). The Araripe Plateau was considered the epicenter of plague in Brazil until 1980, when the last confirmed human case in the region was notified (*9*).

We took advantage of the geographically well-delimited features of the Araripe Plateau and the robust amount of data obtained from the Pilot Plague Program (*10*) regarding human and animal surveillance to perform a case study of a plague focus in Brazil through a One Health perspective. By combining 20 years of epidemiologic data with ecologic, geoclimatic, and *Y. pestis* genomic variables, we aimed to unravel the intricate dynamics of plague in a natural focus in Brazil.

Methods

Data Collection

In brief, we retrieved information on *Y. pestis* reservoirs, vector surveillance, and notification of human plague cases from the national Brazilian Plague Reference Service (https://www.cpqam.fiocruz.br/sr/peste) document repository. We obtained metadata from the Fiocruz *Y. pestis* collection of the World Data Centre for Microorganisms (https://www.wdcm.org; collection no. 1040) and the CONCEPAS database (http://cyp.fiocruz.br). We collected demographic and climatic variables from multiple public databases (Appendix 1 Figure 1, https://wwwnc.cdc.gov/EID/article/30/9/24-0468-App1.pdf).

Geospatial Analysis

We collected coordinates from human cases where public health conducted site visits. To identify plague risk hotspots, we calculated the kernel density estimation (KDE) and space-time scan (SaTScan, https://www.satscan.org) statistics. We interpolated rainfall data from meteorologic stations by using inverse weighted distance and used the normalized difference vegetation index (NDVI) as a proxy for vegetation coverage (Appendix 1).

Laboratory Testing for Plague

Human plague diagnosis was performed through a combination of clinical and epidemiologic assessment and laboratory testing, such as culturing bubo aspirates or blood cultures. Rodent diagnoses were made by direct microscopy observation of spleen imprints and blood smears, followed by conventional bacteriology. Fleas also were tested via bacterial culturing (Appendix 1).

Statistical Analysis and Ecologic Networks

We calculated linear regression, nonlinear correlations, principal component analysis (PCA) and receiver operating characteristic curves by using GraphPad Prism version 10.1 (GraphPad, https://www.graphpad.com). We analyzed the fundamental properties of each host–vector plague network by using the bipartite package in R (The R Project for Statistical Computing, https://www.r-project.org). We considered $p<0.05$ statistically significant.

Whole-Genome Sequencing and Bioinformatic Analysis

We performed DNA extraction by using the DNeasy Blood & Tissue Kit (QIAGEN, https://www.qiagen.com). We conducted genomic library preparation using the Nextera XT Library Preparation Protocol (Illumina, https://www.illumina.com) and used the MiSeq sequencer and 600-cycle version 3 cartridge (Illumina) for sequencing. We assembled genomes by using VelvetOptimiser and annotated by using the Prokka pipeline, as previously described (*11*). We conducted core single-nucleotide variant (SNV) calling by using Snippy (https://github.com/tseemann/snippy) and used IQ-TREE2 (http://www.iqtree.org) to perform phylogenetic analyses based on the core SNV. We then visualized results by using the iTOL (https://itol.embl.de) platform.

Results

Spatiotemporal Features of Plague Outbreaks in the Araripe Plateau Region

During 1961–1980 in the Araripe Plateau focus, 551 human plague cases were reported; 292 were in the state of Pernambuco, 240 in the state of Ceará, and only 1 in the state of Piauí (Appendix 1 Figure 2, panels A, B). However, we excluded 18 cases in Pernambuco from the spatial analysis because case locations were unknown.

Human plague cases were concentrated mostly at the foothills of the Araripe Plateau, in regions of lower altitude, but close to the plateau's slope (Figure 1). KDE analysis concentrated the plague risk more intensely in 3 areas of lower altitude in the municipalities of Exu, Bodocó, and Araripina, and those areas

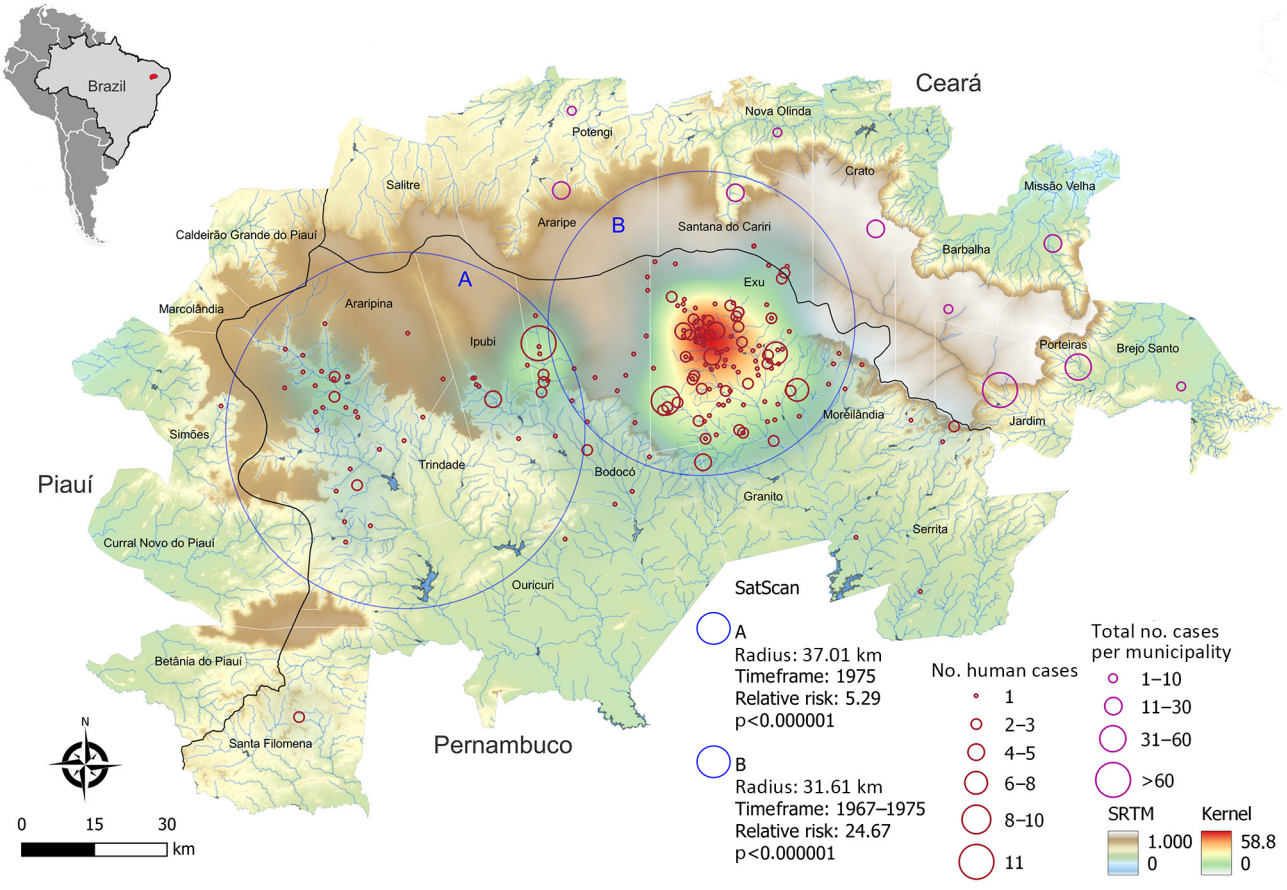


Figure 1. Spatial distribution and risk analysis of human plague cases in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. Background colors show the altimetry (m) from SRTM. The black line shows the tri-state boundaries between Pernambuco, Ceará, and Piauí. Red circles identify plague risk areas by application of KDE in human cases in Pernambuco. Blue circles A and B indicate plague risk clusters calculated by SaTScan for 1975 (A) and 1967–1975 (B). Pink circles indicate spatial distribution of human plague cases by number of occurrences per municipality in Ceará. Inset map shows Brazil with the Araripe Plateau focus in red. KDE, kernel density estimation; SaTScan, space-time scan (https://www.satscan.org) statistics; SRMT, Shuttle Radar Topography Mission (https://www.earthdata.nasa.gov).

are closed in by concave cliffs (Figure 1). Conversely, only a few cases were reported on the top of the plateau. To overcome bias of heterogeneous population density, we performed a SaTScan analysis using the census tracts as geographic areas, which confirmed the findings from the KDE analysis. Furthermore, the estimated relative risk for the population living within the radius of 2 areas were much higher than for the surrounding areas; 5.29-fold higher in area A and 24.67-fold higher in area B (Figure 1). The year-by-year dynamics of plague cases in the region was clearly visible (Video, https://wwwnc.cdc.gov/EID/article/30/9/24-0468-V1.htm).

Effect of Eco-Epidemiologic and Climatic Variables on the Dynamics of Human Plague Cases

By analyzing precipitation levels and data from 40,972 rodents and 39,150 fleas tested for *Y. pestis* in the Araripe Plateau region during 1966–1980, we were able to assess the seasonal and annual effect of those variables on the number of human cases. Overall, the years with higher rodent and flea *Y. pestis* positivity rates overlapped the years with more human cases (Figure 2, panels B, D, F, H, J). In addition, using rodent capture success as a proxy for animal abundance, we found that years with more human cases and *Y. pestis*–positive animals also had reduced rodent abundance, indicating that *Y. pestis* circulation had a major effect on the rodent population. Of note, the annual dynamics of average pluviosity (rainfall amount) showed a 1-year delay effect when compared with human cases until 1976 (Figure 3, panel A), after which plague became quiescent in the region. Thus, we considered the pluviosity from the previous year in further analyses.

Next, we evaluated the seasonal aspects of those variables by measuring the average monthly rates for several additional variables during 1966–1980. Along with the human cases, rodent and flea *Y. pestis* positivity rates, flea index (number of fleas per host), and rodent abundance all showed a strong seasonal component, peaking during the end of winter and beginning of spring (August–October) (Figure 2). We noted that rainfall occurred mostly in the first semester, peaking during late summer and early fall (February–April) (Figure 3, panel B).

The occurrence of human cases was proportionally related to rodent (R^2 = 0.94) and flea (R^2 = 0.93) *Y. pestis* positivity rates, and human cases related more moderately (R^2 = 0.65) to the previous year's pluviosity (Figure 4, panel A). We also ran those

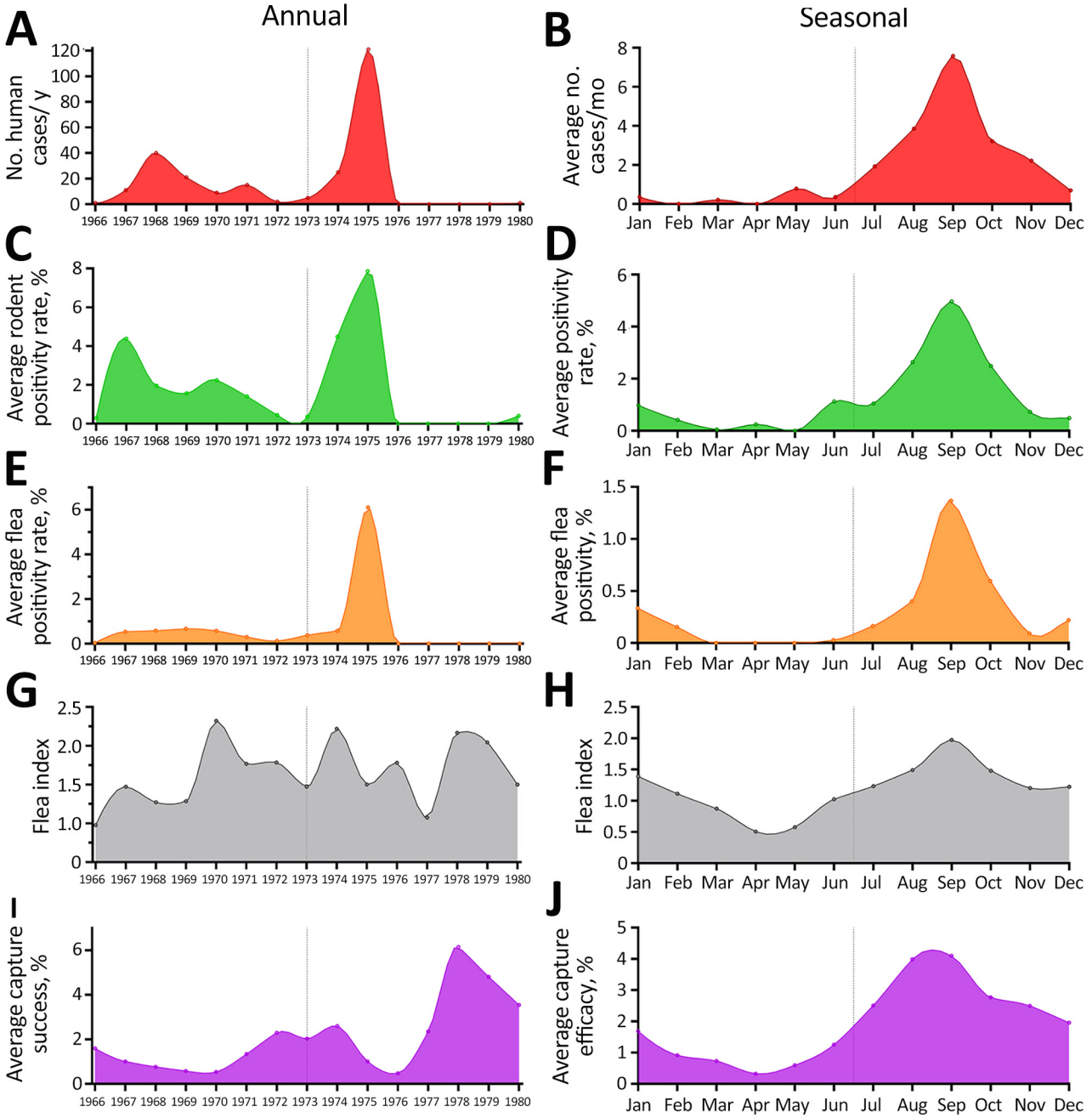


Figure 2. Annual (left column) and seasonal (right column) dynamics of plague occurrence and flea and rodent capture and abundance rates in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A, B) Human cases; C, D) rodent positivity; E, F) flea positivity; G, H) flea index (number of fleas per host); I, J) rodent capture success. Vertical lines provide midpoints for comparison between measured indices.

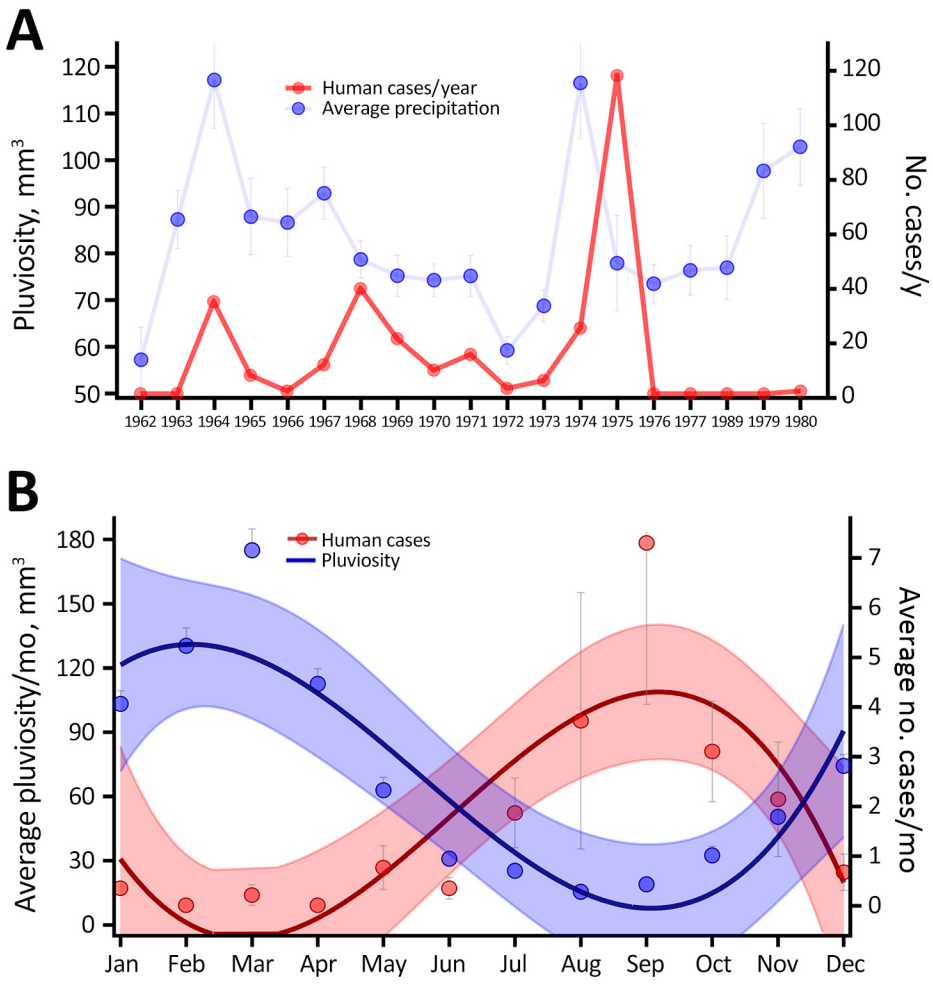


Figure 3. Effects of climatic variables on human cases in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A) Yearly average pluviosity (rainfall amount); B) monthly average pluviosity. Dots indicate averages; whiskers indicate upper and lower limits. Average pluviosity (mm^3) was measured in municipalities in the state of Pernambuco in the Araripe Plateau region during 1962–1980. The curves in panel B represent the third order polynomial interpolation of cases and pluviosity averages; shaded areas indicate 95% CIs.

variables in a multivariable regression model; however, rodent and flea positivity were also mutually dependent and therefore redundant ($R^2 = 0.94$) (Appendix 1 Figure 2, panels C, D). For that reason, the model was better explained by a single variable linear regression. Linear regression of flea index showed no impact of this variable on human cases or rodent infection. By comparing monthly averages during 1966–1980, we found that the rodent abundance (capture success), flea index, and rodent and flea *Y. pestis* positivity rates had a stark linear proportionality with human cases at a seasonal level (Figure 4, panel B). Of note, although the effect of rodent and flea *Y. pestis* positivity rates on human cases showed a linear pattern, rodent abundance had a negative exponential correlation with plague activity in humans, and *Y. pestis* positivity in animal hosts and vectors (Figure 5).

The PCA of multiple eco-epidemiologic and climatic features revealed that distinct years clustered together according to the intensity of human cases, suggesting a synergic effect among those variables. Of note, 1975, which had the highest number of human cases, showed unique eco-climatic features. Those features included lower abundance of *Necromys lasiurus* mice; distinct ecologic network parameters, such as host–vector robustness and modularity; high precipitation (considering a 1-year lagged effect); and higher *Y. pestis*–positivity rates in rodents and vectors (Figure 6; Appendix 1 Figure 2, panel E). We designated years as epidemic or low-activity according to the PCA analysis.

Spatiotemporal Dynamic of Human Plague Cases by Rainfall and Vegetation Coverage

On the basis of our preliminary findings that human plague cases correlated with the average pluviosity from the previous years in the Araripe Plateau region (Figure 3, panel A), we expanded our analysis to evaluate whether that effect would also demonstrate a spatial pattern. By interpolating pluviometry data from meteorologic stations in the region, we observed

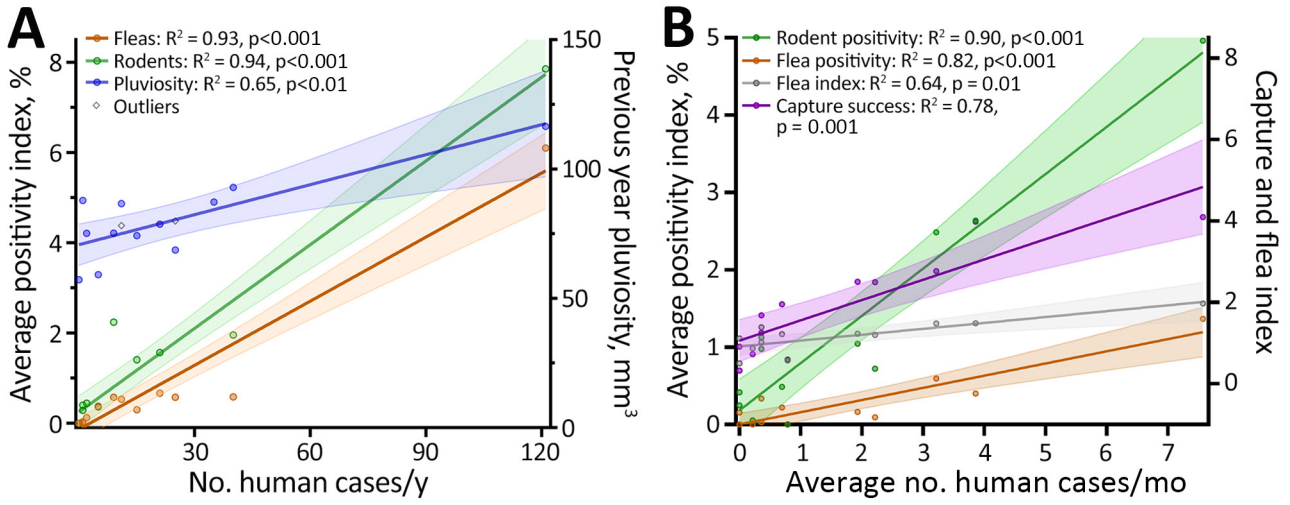


Figure 4. Linear regression of human cases, rodent and flea positivity, and pluviosity (rainfall amount) in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. Annual (A) and monthly (B) average number of human cases compared with *Yersinia pestis* positivity among rodents and fleas and average pluviosity are shown. The previous year pluviosity data only included years from the plague outbreaks, 1966–1976. Solid lines indicate averages, shaded areas indicate 95% CIs, and dots indicate outliers.

a marked overlap between clusters of plague cases and rainfall volume during the study period (Figure 7, panel A). Similar to the plague cases, rain was more intense in lower altitude areas neighboring the elevated plateau. Next, we also investigated the spatial distribution of rainfall in the years that anticipated major outbreaks (1967 and 1974) or epidemiologic silence (1972) and observed a 1-year delay in spatial overlap between rainfall and plague cases (Appendix 1 Figure 3).

Considering the semiarid climate in the region, we hypothesized that one of the mechanisms through which rainfall could modulate plague would be effects on the availability of vegetation as a food source for wild rodents. NDVI data showed that vegetation was concentrated in the slopes of the Araripe Plateau, overlapping the main rainfall and plague case areas (Figure 7, panel B).

Ecologic Aspects of Plague in Araripe Plateau and Ecologic Networks

N. lasiurus mice were the most representative rodent species during the study period. Nevertheless, *N. lasiurus* mice representation varied through time and was greatly reduced in epidemic years (Figure 8, panel A). Moreover, reduction of *N. lasiurus* mice populations had a dramatic effect on the overall capture success rate. Of note, after the abrupt shift from a major plague outbreak in 1975 to quiescence, the wild rodent population quickly reestablished.

Among flea vector species, the most common in captured rodents was *Polygenis* spp. fleas, which were the predominant and almost exclusive species in wild rodents, including *N. lasiurus* mice. On the other hand, *Xenopsylla* spp. fleas were mostly found in the synanthropic *Rattus rattus* rats. Less commonly,

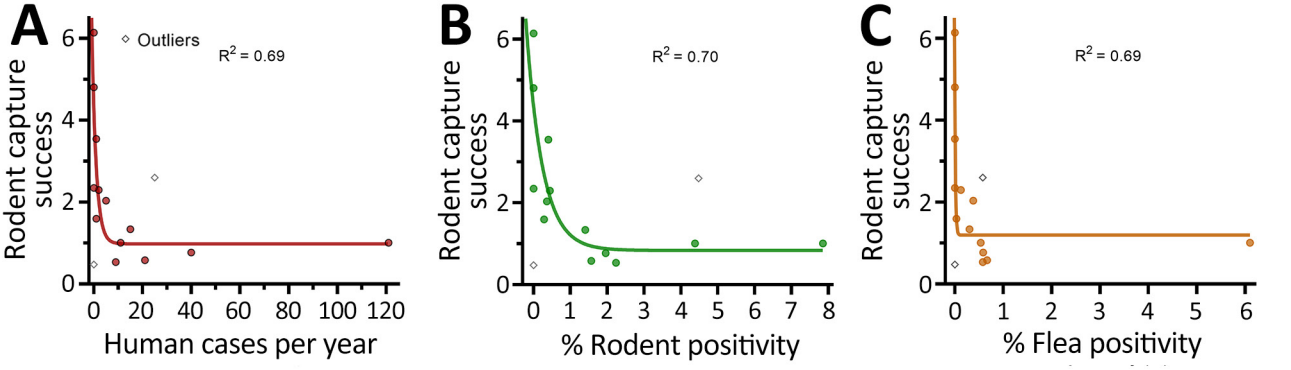


Figure 5. Exponential correlation between rodent capture success and human and animal *Yersinia pestis* positivity in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A) Human cases; B) rodent positivity; C) flea positivity. Capture success serves as a proxy for rodent abundance.

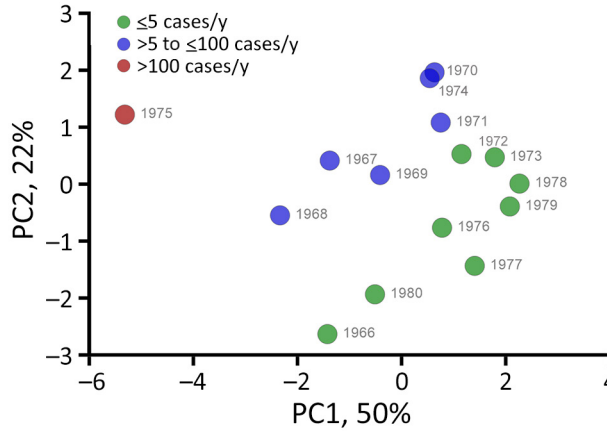


Figure 6. Year-based principal component analysis in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. PC based on eco-epidemiologic and climatic features. The weight of each included variable is provided in Appendix 1 Figure 2, panel E (https://wwwnc.cdc.gov/EID/30/9/24-0468-App1.pdf). PC, principal component.

Adorapsylla spp. and *Ctenocephalides* spp. fleas were also identified (Figure 8, panel B). Because the method of flea species classification changed over time, we referred to fleas only at the genus level.

The structural properties of the host–vector network exhibited significantly higher robustness in epidemic years than low-activity epidemiologic years (Figure 6), and effects were seen in both small mammal (Mann-Whitney p = 0.014) and flea (Mann-Whitney p = 0.029) groups (Figure 9). Modularity (quantitative modularity) and nestedness (weighted nested overlap and decreasing fill [NODF]) values surpassed those predicted by null models, and appeared similar across epidemiologic years as the connectance, a term used to describe the ratio of observed ecological interactions to the total potential for such interactions. By sorting years according to their epidemiologic status, we observed distinct patterns of biologic interactions between small mammals and flea species (Figure 10, panel A). We also calculated annual values for the ecologic network metrics (Appendix 1 Figure 4, panel A).

Using only epidemiologic variables, 11 of 14 plague years (1977 singularity resulted in an invalid value) were correctly classified by the linear discriminant analysis; 1969 and 1975 were misclassified as not epidemic, and 1976 was misclassified as epidemic. When using 5 different network variable combinations with epidemiologic variables, we found the correct classification of those 14 years was obtained by adding only modularity and nestedness. Those findings showed that the network variables increased the correct discrimination by 21.43% compared with the eco-epidemiologic variables alone. We summarized probabilities of classification for all models by using epidemiologic and network variables, and found that, in terms of predictive probability, 1969 and 1976 were less prone to correct prediction by different models (Appendix 1 Figure 4, panel B).

Monthly-Based Eco-Epidemiologic Surveillance Parameters as Predictors of Human Plague Risk

To estimate how data from animal surveillance can provide informative alerts of risk for human plague, we next analyzed monthly data. The receiver operating characteristic curve analysis revealed that 0.86% rodent positivity and 0.34% flea positivity are the optimal cutoff values of risk for human cases. We found that rodent and flea *Y. pestis* positivity rates were strong predictors of plague activity in humans in the same month: for fleas, area under the curve was 0.77, sensitivity was 66.7% and specificity was 88.3%; for rodents, area under the curve was 0.86, sensitivity was 83.3%, and specificity was 81.8%. The other ecologic variables, such as capture success and flea index, on the other hand, were poor predictors (Figure 11).

Genomic Features from *Y. pestis* Isolates

We analyzed 913 *Y. pestis* isolates from human cases, animal reservoirs, and vectors, including 439 isolates from Brazil and 474 from elsewhere. The core genome analysis resulted in a conserved sequence that was equivalent to 46% of the CO92 reference strain (GenBank accession no. GCA_000009065.1). The core SNV contained a total of 1,867 SNV sites that diverged from the CO92 genome.

The core SNV approach identified discernible patterns of geographic distribution of *Y. pestis* lineages at the global level. Strains from Brazil were a relatively homogeneous genetic group, assigned within the 1.ORI phylogenetic population (Figure 12). Isolates from Brazil constituted a monophyletic clade that derived directly from strains from South America and North America, which in turn branched from strains from Yunnan and Southeast Asia. That genetic connection sheds light on the historical dissemination of *Y. pestis* in Brazil. Among *Y. pestis* genomes from the NextStrain dataset (https://nextstrain.org), strains from the same geographic plague foci grouped together within the phylogenetic tree (Figure 12).

The plague spikes in Brazil during 1966–1972 were constituted by a basal occurrence of the main, undifferentiated, genetic group that we termed Araripe ancestor (Figure 12), which derived some transitory groups that were unsuccessful in replacing the basal lineage over time. One of those

genetic groups, group D, was isolated in August 1969 and quickly expanded, representing 76% (28/37) of *Y. pestis* isolates during October 1969–March 1970, when the cases in the region ceased. In the second semester of 1970, when plague started to disseminate again, we did not identify any group D strains, and we assigned the new isolates to the Araripe ancestor group.

The *Y. pestis* strains isolated during the 1974–1975 outbreak formed a monophyletic and discernible clade, group E (Figure 12). Of note, the only 2 isolates during the quiescent year of 1973 were group E, which was uncommon in previous years. After 1973, however, that clade promptly replaced the previous strains circulating in the region, which overlapped with an unprecedented upsurge of plague cases (Figure 13). In addition, the lineage replacement caused by group E strains was observed temporally and spatially (Figure 14). We noted the apomorphic mutations that defined each of those derived groups, along with their predicted effects on protein function (Appendix 2 Table 1, https://wwwnc.cdc.gov/EID/article/30/9/24-0468-App2.xlsx).

Discussion

We aimed to dissect plague epidemics that occurred in the Araripe Plateau focus of Brazil. By combining epidemiologic, ecologic, climatic, and genomic information, we were able to learn relevant aspects of the intricate spatiotemporal dynamics of plague in the region. Analysis from the temporal series showed that eco-epidemiologic and climatic factors had a stark and linear influence on the risk for human plague. Monthly rodent and flea positivity rates strongly

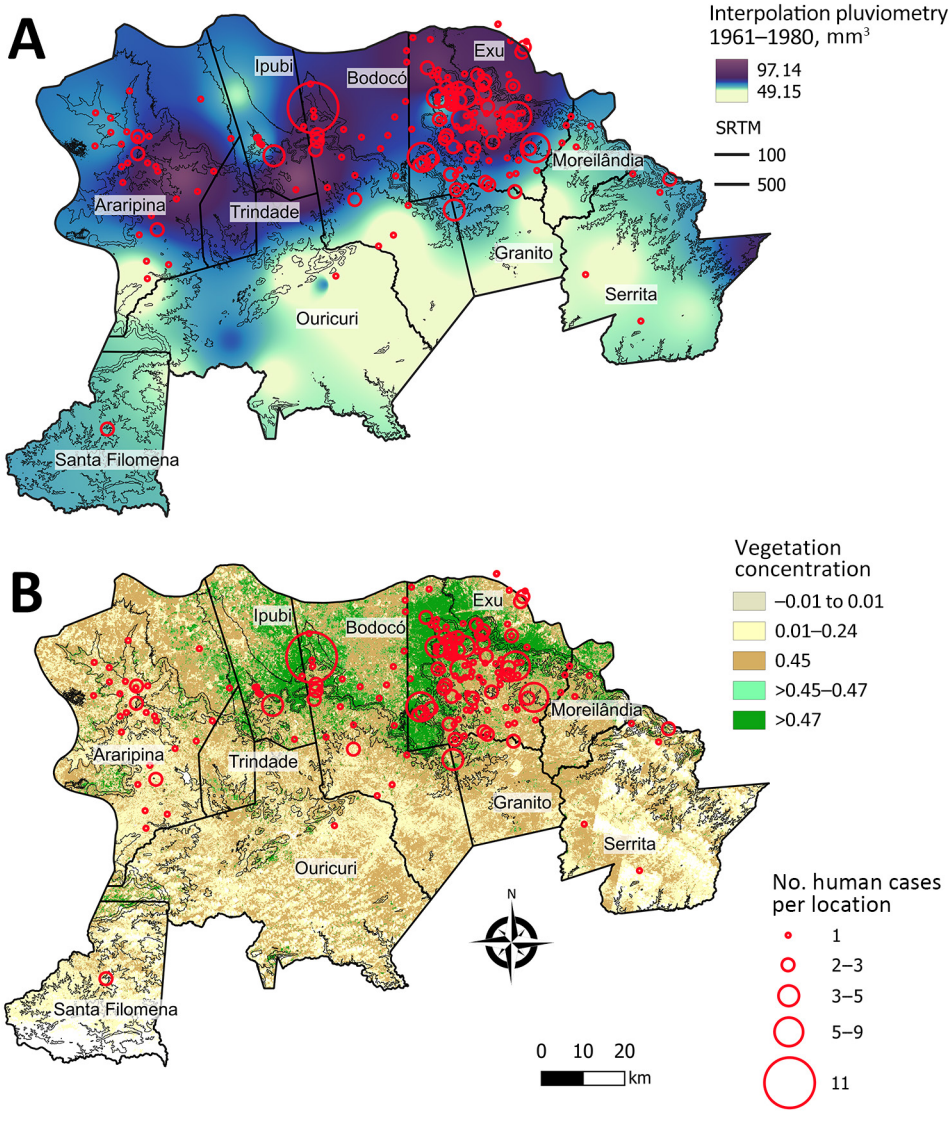


Figure 7. Spatial distribution of rainfall (pluviometry) and vegetation coverage in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A) Locations of human cases and interpolation of average pluviometry (1961–1980) in the Araripe Plateau municipalities of Pernambuco state. B) Locations of human cases and normalized difference vegetation index model showing vegetation concentrated in the slopes of the Araripe Plateau, overlapping the main rainfall and plague case areas. Images obtained by the Landsat 4 satellite in 1984.

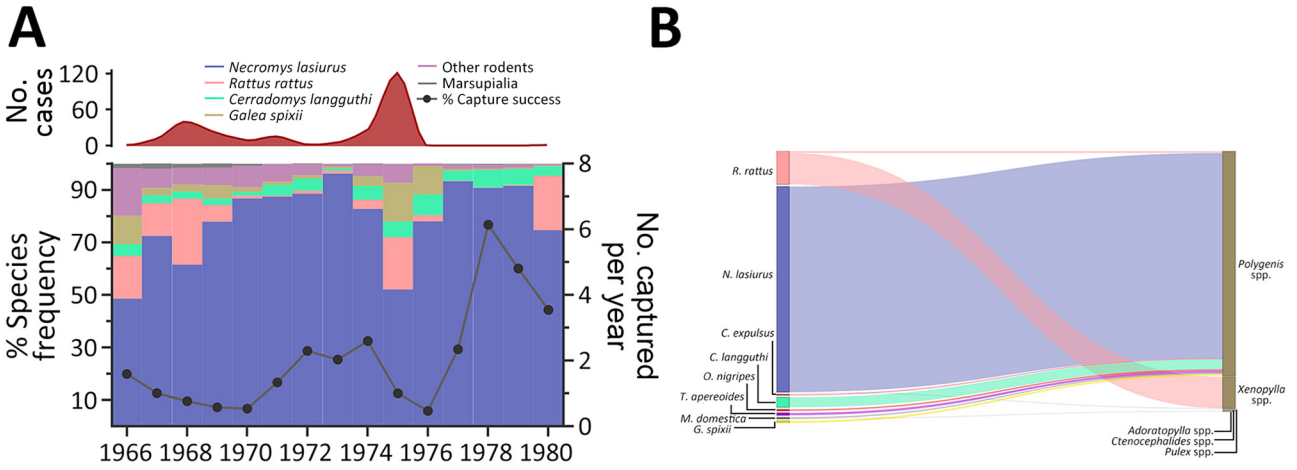


Figure 8. Diversity of rodent and flea species during distinct epidemiologic scenarios of plague in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A) Frequency of each rodent species on capture traps per year (left axis), yearly capture success (right axis) and human plague cases (red curve on top). B) Sankey diagram representing the distribution of rodents and other small mammal hosts according to the flea species and frequency.

indicated the risk for human cases and cutoff values were <1% positivity, meaning any basal detection of *Y. pestis* bacteria in wildlife signaled a risk for human plague cases. Although these indicators have high sensitivity and specificity, detecting early-stage risk requires a large number of sampled animals. Those findings support previous studies suggesting that serosurveys in sentinel species, such as dogs, are more efficient and require fewer tests (*12*).

Human plague in the Araripe Plateau region typically started at the end of the rainy season and peaked in the driest months, August–September. Our results showed that rainfall affected plague dynamics in this semiarid region by modulating the wild rodent and flea abundance. Sylvatic rodent and flea populations expand after the rainy season until the peak of the dry season, and once food and water sources become scarce, rodents carrying vector fleas

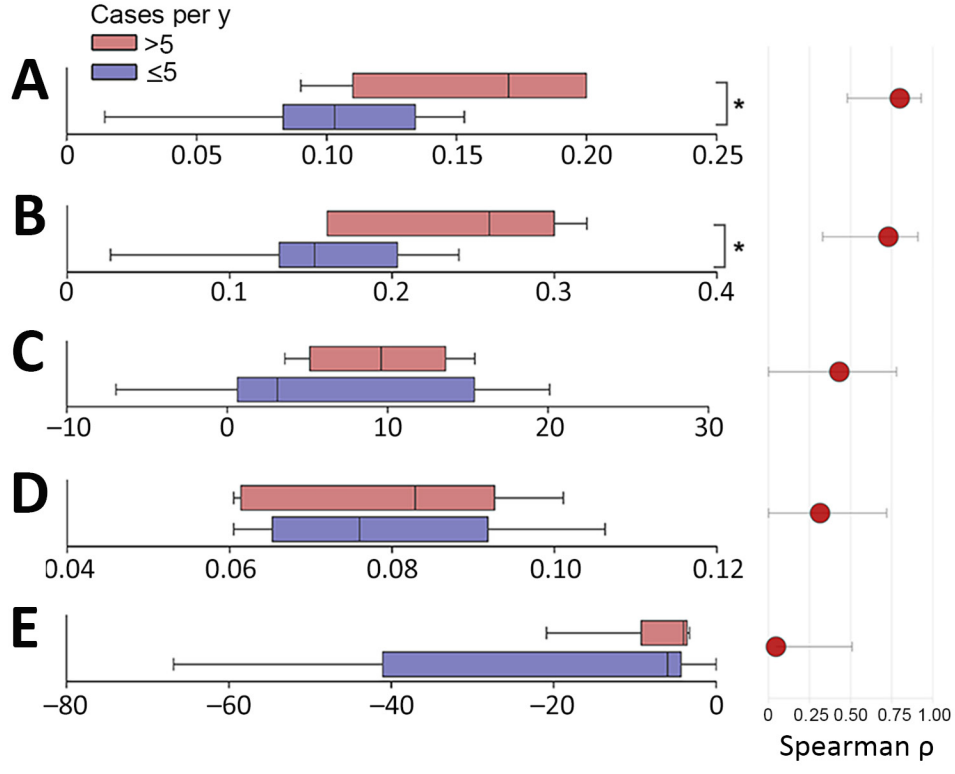


Figure 9. Comparison between ecologic networks parameters in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. Graphs compare nonepidemic years (cases ≤5) with epidemic years (cases >5). A) Host robustness; B) vector robustness; C) modularity; D) connectance, which is used to characterize community-wide specialization; E) nestedness based on overlap and decreasing fill. Cutoff value for epidemic status was defined according to the cluster at the principal component analysis. Mann-Whitney test was used to compare groups. Asterisks (*) indicate statistically significant difference ($p<0.05$). The lower and upper error bars correspond to the first and third quartiles (the 25th and 75th percentiles). Vertical lines within boxes represent medians.

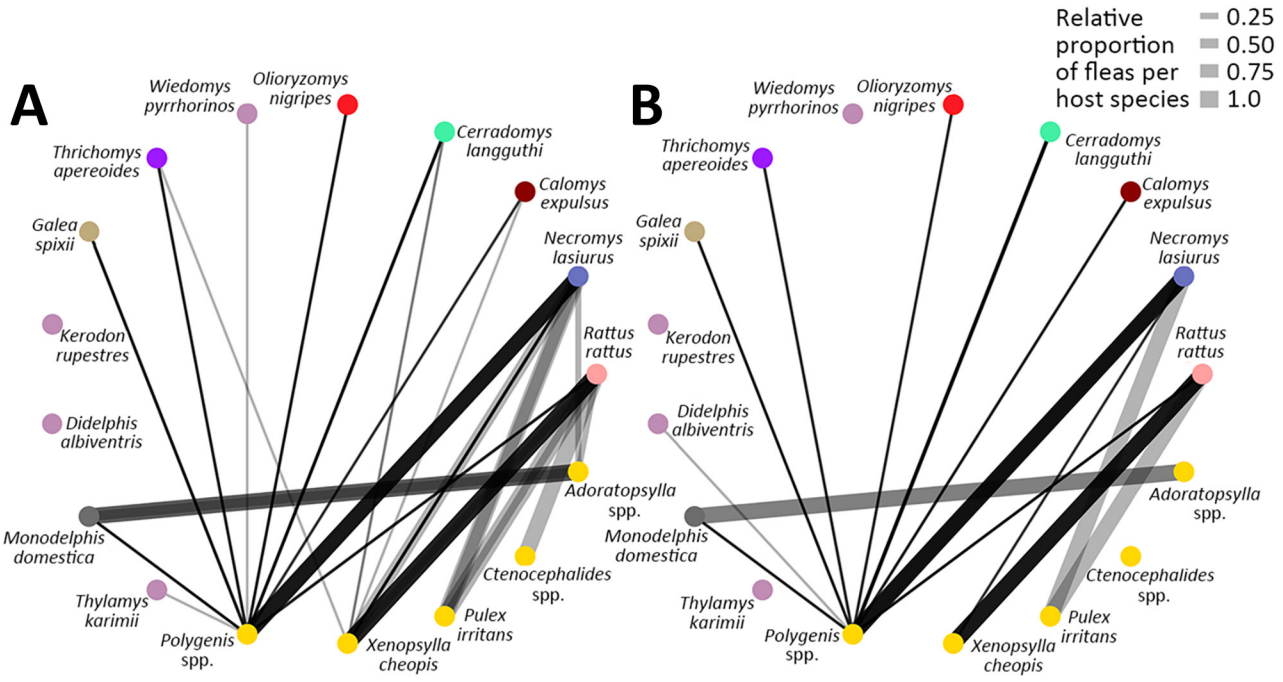
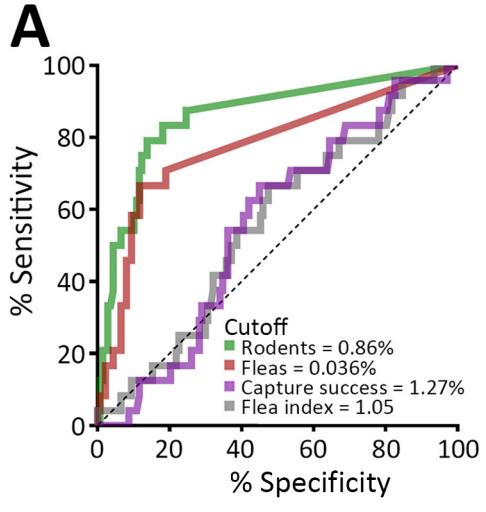


Figure 10. Potential *Yersinia pestis* transmission networks in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. A) Epidemic years, ≥5 human cases; B) nonepidemic years, <5 human cases. Transmission networks were based on biologic interactions between host and vector species. Weight of links represent relative proportion of flea species per small mammal.

are attracted to human domestic perimeters by the grain crops grown there. Those events overlap the quick dissemination of plague in the rodent communities, leading to rodent dieoffs and the search for new hosts by the fleas.

Precipitation patterns have been implicated in plague risk in distinct ecosystems (*4*,*13*–*15*). Our findings suggest that rainfall could set the ground conditions for plague dissemination in the following year. Similarly, studies from other plague systems also reported a lagged effect of precipitation on plague risk (*15*–*17*). Of note, the plague epidemic described here overlapped with a period of cool and negative Pacific decadal oscillation teleconnection that lasted until 1976 (*18*–*20*), which is supported by previous studies performed in other ecosystems (*17*).

Geospatial analysis revealed that locations at lower altitudes and closer to the concave sectors from the plateau were at higher risk for plague occurrence. We hypothesized that those locations provided specific conditions that led to the dissemination of plague in wildlife and, eventually,



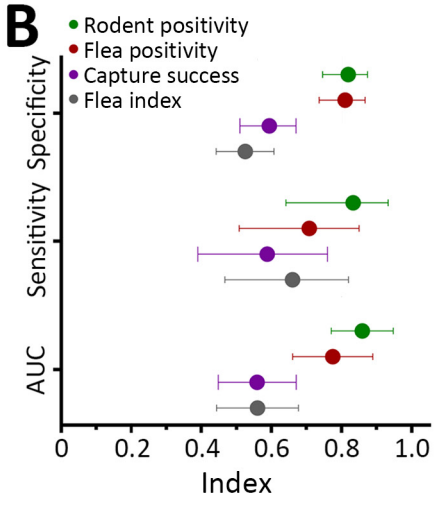


Figure 11. Human risk prediction in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. Prediction used ecologic variables at a monthly level. A) Receiver operating characteristic curves and cutoff values of the ecologic variables for the prediction of >2 human cases within the same month. B) Sensitivity, specificity, and AUC for each variable. Error bars indicate 95% CIs. AUC, area under the curve.

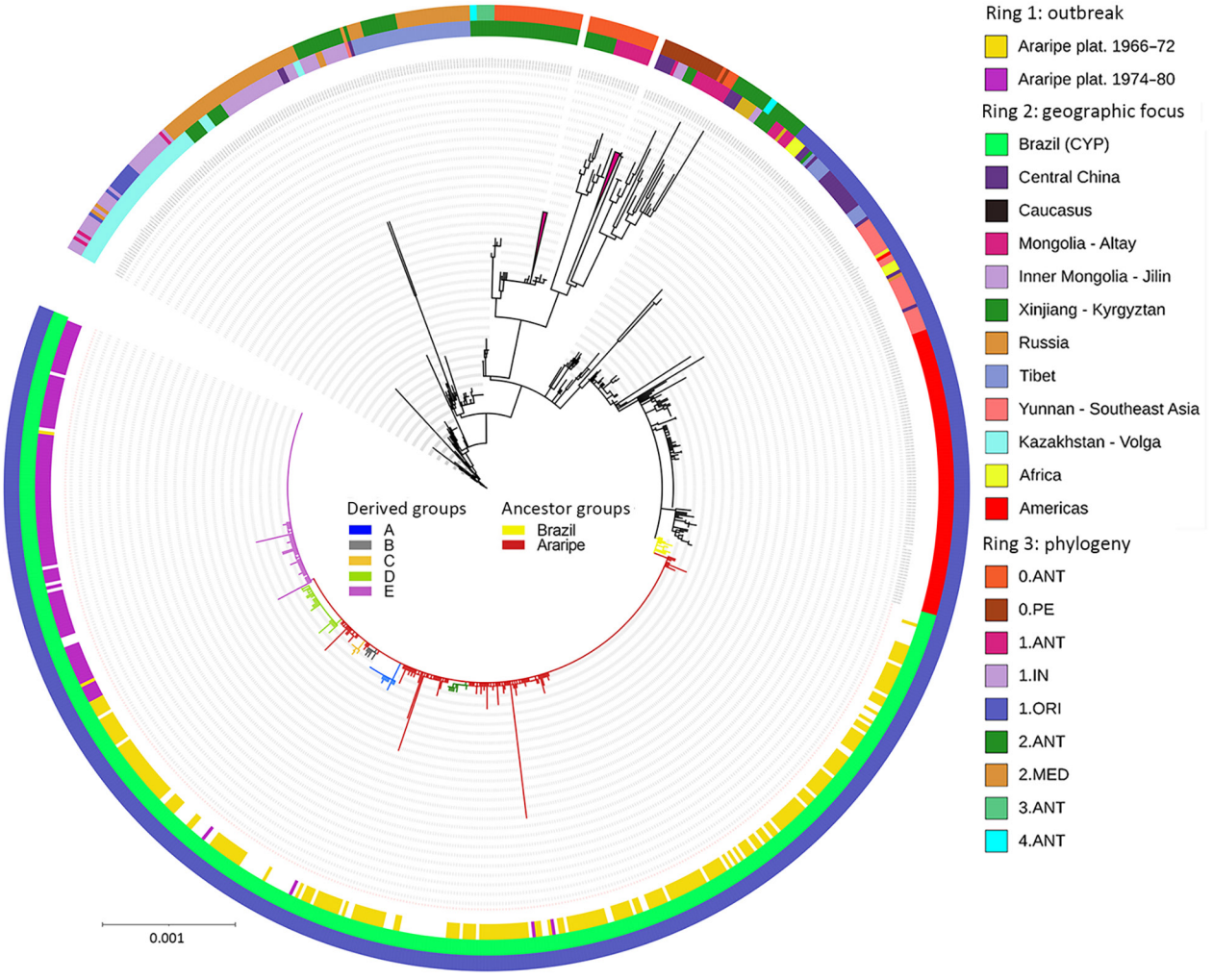


Figure 12. Genomic characterization of *Yersinia pestis* strains in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. Phylogenetic tree was based on the 1,867 single-nucleotide variants identified in the core genome from 913 strains isolated in the Araripe Plateau and included in the analysis. The rings contain metadata regarding the epidemiologic features from the Araripe Plateau outbreaks (ring 1), the attributed geographic foci (ring 2), and genetic group provided in the NextStrain dataset (https://nextstrain.org) (ring 3). Brazil branches are colored according to their genetic subgroups. Scale bar indicates nucleotide substitutions per site.

in humans. That hypothesis was further supported by the marked spatial overlap between plague hotspots with increased rainfall and vegetation coverage areas. Of note, although the top of the plateau has savanna-like vegetation, the concave slopes are described to encompass unique characteristics,

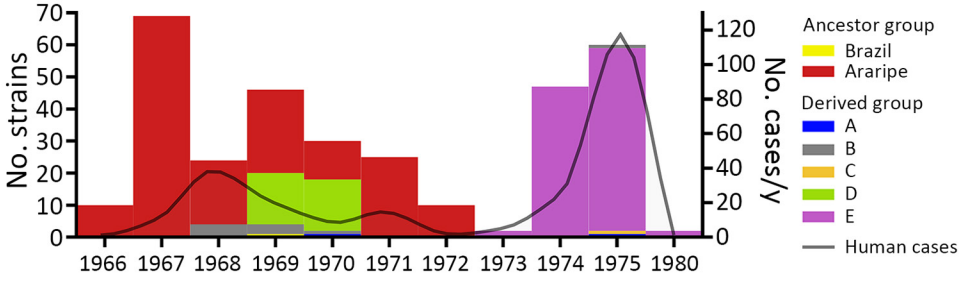


Figure 13. Temporal distribution of *Yersinia pestis* strains in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. The graph shows genomic characterization of *Y. pestis* strains from this study compared with the number of human plague cases per year. Scales for the y-axes differ substantially to underscore patterns, but do not permit direct comparisons.

including forest vegetation and increased water availability due to orographic rain (*7*,*21*). Those findings support the trophic cascade hypothesis and underscore the influence of climate factors, such as water availability, vegetation coverage, and host–vector abundance, on *Y. pestis* dissemination. Similar patterns have been observed in other systems, indicating a broader ecologic context where climate is implicated in the spread of plague (*14*,*17*,*22*). Identifying well-defined hotspots with conditions optimal for plague outbreaks is crucial for maximizing an assertive epidemiologic surveillance of this neglected yet reemerging disease, especially in resource-constrained settings.

The scarcity of satellite-derived data throughout the period studied restricted our access to variables essential for modeling the suitability of plague resurgence. Moreover, the abrupt disappearance of plague

Figure 14. Spatial distribution of *Yersinia pestis* strains in a study of ecologic, geoclimatic, and genomic factors modulating plague epidemics in primary natural focus, Brazil. The sequence strains from the Araripe Plateau are shown during various timeframes: A) August 1966–October 1969; B) November 1969–March 1970; C) April 1970–May 1973; and D) June 1973–December 1980. Detailed spatiotemporal dynamics of *Y. pestis* lineages in the Araripe Plateau are available at https://microreact.org/project/fNz1zKcNyCTmKQrtke33Gm-yersinia-pestis-strains-in-the-araripe-plateau-brazil.

after a massive rodent dieoff in 1975 potentially could be attributed to the depletion of sensitive hosts or vectors and the expansion of resistant ones (*22*,*23*). Consequently, further research encompassing later timeframes and integrating satellite-derived climatic, topographic, and land-use data, alongside assessments of host–vector suitability for *Y. pestis* infection, is imperative for accurately modeling potential scenarios of plague resurgence.

Beyond its effects on human health, sylvatic plague raises major conservation concerns. Our findings show that plague activity correlated with substantial declines in abundance of *N. lasiurus* mice. The drastic effect on wild rodent abundance has been reported to cascade through the ecosystem, affecting prey–predator balance and consequently, biodiversity and ecosystem stability (*24*). Furthermore, plague can disturb ecologic systems by directly infecting a wide range of mammal species (*25*,*26*). In this study, the presence of generalist species with extensive interactions contributed substantially to the network's robustness during epidemic years; however, a sharp population decline in those species impacts most of their interaction partners, rendering generalist species surveillance crucial (*27*,*28*). We found the potential for plague maintenance via multiple species interactions, akin to patch dynamics (*29*), is especially evident during epidemic years. Such periods exhibited greater stability in maintaining transmission pathways despite local species extinctions, which was evidenced by higher network robustness in epidemic years.

The amalgamation of phylogenetic and spatiotemporal characteristics in *Y. pestis* isolates from the Araripe focus highlighted the incidence of a foundational ancestral strain. That strain exhibited the capability to evolve into distinct genetic groups, each linked to temporally and spatially constrained outbreaks. Of note, the large outbreak of 1974–1975 was observed in immediate succession to the displacement of previously prevalent *Y. pestis* strains by a singular clade. The expansion of specific *Y. pestis* clades temporarily overlapped with 2 atypical years, 1969 and 1976, in the eco-epidemiologic linear model. Phylogroup D appeared in 1969, and 1976 marked the aftermath of the largest plague outbreak period in the region; strains from that outbreak were genetically characterized by displacement of prior lineages by phylogroup E.

The results from this study rely on records generated during epidemiologic surveillance of human and animal plague. The first limitation of this study is the lack of case-level data in the state of Ceará in the northern range of the Araripe Plateau, which limited most data analysis to the southern part in Pernambuco state. The second limitation is that, although plague surveillance in rodents and fleas was uninterrupted during the study period, outbreaks in specific sites might have biased the rodent capture effort toward specific locations. Finally, because satellite images were available only after the study period, a 4-year delay occurs between surveillance and NDVI representation.

In conclusion, this study identified well-delimited pockets of plague activity spanning areas just a few kilometers wide in the Araripe Plateau region of Brazil. Through the lens of One Health, we determined plague hotspots by using a specific combination of rainfall, vegetation, and landscape features. We also measured the impact of plague activity on wild rodent populations and spillover events in humans. Our research provides a holistic understanding of the ecologic implications of plague in an ecosystem of Brazil. Our findings from Araripe Plateau data could help refine plague risk areas and improve surveillance in other plague foci within the Caatinga biome.

Acknowledgments

We thank Cássia Docena and Gustavo Barbosa Lima and the Núcleo de Plataforma Tecnológica Fiocruz for the assistance during the whole-genome sequencing experiments. We are also thankful to the Coordenação de Coleções Biológicas and Coordenação de Vigilância em saúde e Laboratórios de Referência at Fiocruz for their support. In addition, we thank the Pathogen and Microbiome Institute at Northern Arizona University, Flagstaff, Arizona, USA, for supporting the sequencing of *Yersinia pestis* strains.

The spatial coordinates of meteorology stations, the NextStrain dataset metadata, the sequencing data of the newly sequenced strains and MicroReact input dataset with the Araripe Plateau strains are available in Appendix 2 Tables 3–6 and Appendix 3 (https://wwwnc.cdc.gov/EID/article/30/9/24-0468-App2.xlsx). Fiocruz-CYP strains are also deposited at World Data Centre for Microorganisms (https://www.wdcm.org; collection ID 1040). Individualized data on human cases and surveillance may be accessed upon request to the authors. Data from satellite images and precipitation were obtained from publicly available sources described in Appendix 1. The sequences used in the work are freely available in GenBank; accession numbers are provided in Appendix 2 Table 6.

This work was supported by the Inova Novos Talentos–Fiocruz/Fundação Oswaldo Cruz (grant no. VPPCB-008-FIO-18-2-73 to M.F.B.) and Inova PROEP (grant no. IAM-005-FIO-22-2-2-36 to M.S.).

Author contributions: M.F.B, D.L.S.R., and A.M.P.A. conceived the project; D.L.S.R., E.C.S.G., A.L.S.O, and R.J.P.S. performed the geoprocessing analysis; I.V.R. and M.S. performed the *Yersinia pestis* culture-related experiments and curated the Fiocruz-CYP collection data; I.V.R. performed the genomic sequencing; J.L.L.P.P. and A.M.R. performed the bioinformatic analysis; N.D.A.F., C.S.A., and P.C.E. performed the ecological networks analysis; M.F.B. and A.M.P.A. performed the epidemiological and animal surveillance analysis; A.M.P.A. and D.L.S.R. accessed the PPP documentary collection to acquire epidemiologic and animal surveillance data; A.M.P.A., M.S., and M.F.B. acquired the funding; M.F.B. wrote the original draft of the manuscript, and all authors revised and approved the submitted version.

About the Author

Dr. Bezerra is a biomedical technologist at the Plague National Reference Laboratory, Aggeu Magalhães Institute–Oswaldo Cruz Foundation–Recife, Brazil. His research interests include laboratory diagnosis, genomic surveillance, and epidemiology of plague and other zoonotic diseases.

References

1. Vallès X, Stenseth NC, Demeure C, Horby P, Mead PS, Cabanillas O, et al. Human plague: an old scourge that needs new answers. PLoS Negl Trop Dis. 2020;14:e0008251. https://doi.org/10.1371/journal.pntd.0008251
2. World Health Organization. Plague around the world in 2019. Wkly Epidemiol Rec. 2019;94:289–92.
3. Bramanti B, Stenseth NC, Walløe L, Lei X. Plague: a disease that changed the path of human civilization. Adv Exp Med Biol. 2016;918:1–26. https://doi.org/10.1007/978-94-024-0890-4_1
4. Xu L, Wang Q, Yang R, Ganbold D, Tsogbadrakh N, Dong K, et al. Climate-driven marmot-plague dynamics in Mongolia and China. Sci Rep. 2023;13:11906. https://doi.org/10.1038/s41598-023-38966-1
5. Carlson CJ, Bevins SN, Schmid BV. Plague risk in the western United States over seven decades of environmental change. Glob Change Biol. 2022;28:753–69. https://doi.org/10.1111/gcb.15966
6. Faria ACM, Almeida AMP, Lima Júnior FEF, Fonseca LX, Sobreira M, Nunes ML, et al. Plague [in Portuguese]. In: Werneck GL, editor. Health Surveillance Guide, 3rd edition [in Portuguese]. Brasília: Ministry of Health; 2023. p. 1077–90.
7. Guerra M, Silva MJSE. Palm swamps of the Araripe Plateau: subspaces of exception in the semiarid of the state of Ceará, Brazil. Ateliê Geográfico. 2020;14:51–66.
8. Menezes Silva MT, Falcão Sobrinho J, Bezerra de Souza E. Floristic composition in slope areas with different land uses in Chapada do Araripe–CE [in Portuguese]. Geosul. 2022;37:117–40. https://doi.org/10.5007/2177-5230.2022.e82020
9. Fernandes DLRDS, Gomes ECS, Bezerra MF, E Guimarães RJPS, de Almeida AMP. Spatiotemporal analysis of bubonic plague in Pernambuco, northeast of Brazil: case study in the municipality of Exu. PLoS One. 2021;16:e0249464. https://doi.org/10.1371/journal.pone.0249464
10. Baltazard M. The exemplary approach of a field epidemiologist M. Baltazard and the plague foyers of the Brazilian Northeast [in Portuguese]. Bull Soc Pathol Exot. 2004;97:87–117.
11. Pitta JLLP, Bezerra MF, Fernandes DLRDS, Block T, Novaes AS, Almeida AMP, et al. Genomic analysis of *Yersinia pestis* strains from Brazil: search for virulence factors and association with epidemiological data. Pathogens. 2023;12:991. https://doi.org/10.3390/pathogens12080991
12. Gage K. Plague surveillance. In: Dennis DT, Gage KL, Gratz NG, Poland JD, Tikhomirov E, editors. Plague manual: epidemiology, distribution, surveillance and control. Geneva: World Health Organization; 1999.
13. Yuan X, Yang L, Li H, Wang L. Spatiotemporal variations of plague risk in the Tibetan Plateau from 1954–2016. Biology (Basel). 2022;11:304. https://doi.org/10.3390/biology11020304
14. Xu L, Schmid BV, Liu J, Si X, Stenseth NC, Zhang Z. The trophic responses of two different rodent-vector-plague systems to climate change. Proc Biol Sci. 2015;282:20141846. https://doi.org/10.1098/rspb.2014.1846
15. Parmenter RR, Yadav EP, Parmenter CA, Ettestad P, Gage KL. Incidence of plague associated with increased winter-spring precipitation in New Mexico. Am J Trop Med Hyg. 1999;61:814–21. https://doi.org/10.4269/ajtmh.1999.61.814
16. Brown HE, Ettestad P, Reynolds PJ, Brown TL, Hatton ES, Holmes JL, et al. Climatic predictors of the intra- and inter-annual distributions of plague cases in New Mexico based on 29 years of animal-based surveillance data. Am J Trop Med Hyg. 2010;82:95–102. https://doi.org/10.4269/ajtmh.2010.09-0247
17. Enscore RE, Biggerstaff BJ, Brown TL, Fulgham RE, Reynolds PJ, Engelthaler DM, et al. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960–1997. Am J Trop Med Hyg. 2002;66:186–96. https://doi.org/10.4269/ajtmh.2002.66.186
18. Andreoli R, Kayano MT. ENSO-related rainfall anomalies in South America and associated circulation features during warm and cold Pacific decadal oscillation regimes. Int J Climatol. 2005;25:2017–30. https://doi.org/10.1002/joc.1222
19. Rolim LZR, Oliveira da Silva SM, de Souza Filho F. Analysis of precipitation dynamics at different timescales based on entropy theory: an application to the State of Ceará, Brazil. Stochastic Environ Res Risk Assess. 2022;36:2285–301. https://doi.org/10.1007/s00477-021-02112-y
20. Garreaud RD, Vuille M, Compagnucci R, Marengo J. Present-day South American climate. Palaeogeogr Palaeoclimatol Palaeoecol. 2009;281:180–95. https://doi.org/10.1016/j.palaeo.2007.10.032
21. Lima FJ, Corrêa ACB, Lima GG, Marçal MS, Paisani JC, Pontelli ME. Late quaternary geomorphological evolutionary dynamics of the Araripe sedimentary plateau, northeast of Brazil. J S Am Earth Sci. 2023;124:1–13.
22. Ben Ari T, Gershunov A, Gage KL, Snäll T, Ettestad P, Kausrud KL, et al. Human plague in the USA: the importance of regional and local climate. Biol Lett. 2008;4:737–40. https://doi.org/10.1098/rsbl.2008.0363
23. Andrianaivoarimanana V, Rajerison M, Jambou R. Exposure to *Yersinia pestis* increases resistance to plague in black rats and modulates transmission in Madagascar.

BMC Res Notes. 2018;11:898. https://doi.org/10.1186/s13104-018-3984-3
24. Biggins ED, Kosoy MY. Influence of introduced plague on North American mammals: implications from ecology of plague in Asia. J Mammal. 2001;82:906–16. https://doi.org/10.1644/1545-1542(2001)082<0906:IOIPON>2.0.CO;2
25. Eads DA, Livieri TM, Dobesh P, Hughes JP, Fly J, Redmond H, et al. Plague mitigation for prairie dog and black-footed ferret conservation: degree and duration of flea control with 0.005% fipronil grain bait. Curr Res Parasitol Vector Borne Dis. 2023;3:100124. https://doi.org/10.1016/j.crpvbd.2023.100124
26. Bevins SN, Chandler JC, Barrett N, Schmit BS, Wiscomb GW, Shriner SA. Plague exposure in mammalian wildlife across the western United States. Vector Borne Zoonotic Dis. 2021;21:667–74. https://doi.org/10.1089/vbz.2020.2765
27. Dallas T, Cornelius E. Co-extinction in a host-parasite network: identifying key hosts for network stability. Sci Rep. 2015;5:13185. https://doi.org/10.1038/srep13185
28. Reis Da Silva Fernandes DL, Filgueira Bezerra M, Sobreira Bezerra DA Silva M, Leal NC, DE Souza Reis CR, DE Almeida AMP. Rodent hosts and flea vectors in Brazilian plague foci: a review. Integr Zool. 2021;16:810–9. https://doi.org/10.1111/1749-4877.12480
29. Strona G, Lafferty KD. Environmental change makes robust ecological networks fragile. Nat Commun. 2016;7:12462. https://doi.org/10.1038/ncomms12462

Address for correspondence: Matheus Bezerra, Departamento de Microbiologia, Instituto Aggeu Magalhães, Brazil Av. Professor Moraes Rego–Campus UFPE, Fiocruz 50740-465, Brazil; email: matheus.bezerra@fiocruz.br

Article DOI: https://doi.org/10.3201/eid3009.240468

Ecologic, Geoclimatic, and Genomic Factors Modulating Plague Epidemics in Primary Natural Focus, Brazil

Appendix 1

Additional Methods

Data Collection

We combined multiple data sources, including reservoirs and vector laboratorial surveillance, notification of human cases, the *Yersinia pestis* Fiocruz/CYP biologic collection (WDCM accession number: 1040), CONCEPAS database, and climatic/environment variables from public databases (Appendix 1 Figure 1). The details from each dataset are provided below. Information regarding human plague cases in the Araripe plateau region from 1961 to 1980 was acquired through the examination of Human Plague Occurrence notification forms, which are maintained at the National Plague Control Program's documentary collection. Notably, municipalities in Pernambuco and Piauí provided comprehensive details, including individual-level data such as the date of case notification/disease onset and geographic coordinates. In contrast, municipalities in the state of Ceará were aggregated annually for each municipality.

Data from the reservoir/vector epidemiologic surveillance (1966–1980) included rodent capture success (a proxy of rodent abundance), rodent species relative abundance and their ectoparasites, rodents and flea infection positivity index, average number of fleas per individual (flea index). These data were obtained by consulting the original laboratory registry books. Information regarding the *Y. pestis* strains isolated in the region was extracted from the CONCEPAS database (http://cyp.fiocruz.br/index?services).

The precipitation data were extracted from the Water and Climate agency from Pernambuco state (Agência Pernambucana de Águas e Climas - APAC,

https://www.apac.pe.gov.br). The yearly averages were calculated as the average of monthly averages from the given year. Monthly average was determined by the average values of the measured water volume from all the meteorological stations available in the region. For the seasonal analysis, we calculated the average values for each month considering the study timeframe (1961–1980). The list of all meteorological stations used in the study is provided in Appendix 2 Table 3. Importantly, the pluviosity-related analyses are limited to the portion of the Araripe plateau located in the Pernambuco state because there was poor data availability for Ceará and Pauí states during the study timeframe.

Laboratory Testing for Plague

The captured rodents were maintained in quarantine cages and were observed on a daily basis. The individuals that died during quarantine were immediately tested for plague and the survivors were tested 4 weeks after capturing. The laboratorial diagnosis in rodents during the PPP program was performed through *Y. pestis* detection by direct microscopy observation of spleen imprints and blood smears, followed by conventional bacteriology as previously described (*1*). Suspected colonies were confirmed with the bacteriophage test (*2*). The fleas were also tested in bacterial culturing. The diagnosis in humans was performed through the combination of clinical and epidemiologic assessment and laboratory testing (bubo aspirates and blood cultures).

Geoprocessing

Study area: The geographic scope encompasses the plateau, characterized by an expansive summit plateau situated at an elevation of 950 m above sea level. The plateau is surrounded by scalloped escarpments undergoing noticeable erosive retreat, featuring a topographical disparity of 250 m in the extreme west (Araripina county) and 500 m in the east (Exu and Moreilândia counties).

For the production of the georeferenced maps, coordinates were collected with on-site visits. Locations of interest were georeferenced using the global positioning system (GPS) technology, using a GPS receiver model eTrex Vista Cx, Garmin (Kansas City, USA), UTM projection, Datum SIRGAS 2000, with a minimum accuracy of 10 m. The Cartographic Base (limit of municipalities, districts, census tracts, subnormal agglomerations), data on the estimated population and hydrography information were obtained from the Brazilian Institute of Geography and Statistics (IBGE - http://www.ibge.gov.br).

Kernel Density Estimation (KDE) was constructed, using the shapefile from the locations where human cases of plague occurred, the radius was calculated from the matrix of distance of the points and the mean and standard deviation (SD) values, applied to the following equation: $R = \bar{x} \pm \bar{x}\sigma$. The calculated value was weighted by the number of cases per location. Incidence map was calculated as follows: the total number of plague cases in the period from 1961 to 1980 per municipality divided by the population in the middle of the period (1970), adjusted for 100,000 inhabitants/year. The space-time scan statistics (https://www.satscan.org) used the discrete Poisson model. To increase the granularity of the analysis, the municipalities were subdivided into smaller polygons, called census tracts (n = 268). The reference population size was that of the 2000 census (first year in which the population was recorded at a census tracts level). For the construction of pluviosity maps, the rainfall data described above was interpolated using the Inverse Weighted Distance - IDW method, in the linear interpolation and quartile mode.

The Normalized Difference Vegetation Index (NDVI) map was constructed using the mosaic of satellite images that represents the extension of the territory of Araripe Plateau. The formula NDVI = (NIR - red) / (NIR + red) was used, where NIR is near infrared and red: visible red light spectrum. LandSat4 images were downloaded from the GloVis-USGS database (https://earthexplorer.usgs.gov). The hydrography was obtained from the Mineral Resources Research Company (CPRM) (https://www.cprm.gov.br/en/Hydrology-83) and the Digital Elevation Model (DEM) data was obtained from the Shuttle Radar Topography Mission (SRTM) refined for the Brazilian territory from the original resolution of 3 arc seconds to 1 arc second using a geostatistical approach (http://www.dsr.inpe.br/topodata).

Ecologic Networks

We constructed *Y. pestis* potential transmission networks based on interactions between small mammal species (reservoirs) and the fleas (vectors) captured during plague surveillance activities between 1966–1980 in the Araripe plateau. A network was structured for each year, with nodes representing small mammals and fleas species, and links denoted the relative proportion of fleas species per small mammals species, resulting in 15 networks.

We analyzed the fundamental properties of each host-vector plague network using the bipartite R package (*3*). The structural properties of each network were analyzed temporally

through the examination of connectance, which realized the proportion of possible interactions between fleas and small mammals species (*4*). We tested for a nested arrangement of the networks using the quantitative method weighted NODF (*5*). Network robustness (*6*) were estimated by quantifying the weighted loss of interactions between fleas after randomly removing species of small mammals, and vice versa. We searched for groups of species that interact more inside the group than outside using the Q modularity metrics (*7*). Due to variations in network size, modularity and nestedness were standardized to z-scores. This involved comparing observed values minus the mean value estimated by null models, divided by the standard deviation of null models, using the *vaznull* null model method (*8*).

We anticipated higher connectance and link density during epidemic years, potentially amplifying transmission probabilities. Furthermore, the presence of generalist and abundant species was expected to boost nestedness values during epidemics, thereby promoting plague transmission within the network. We hypothesized that more modular networks, by limiting connections into distinct groups, would reduce transmission flow during low activity years. Additionally, we expected higher network robustness in epidemic years, as they maintain more connections even after species loss. When a species was removed from epidemic networks, it was expected that other generalist species would compensate for the interactions without losing connections within the subset.

The network's structural metrics were correlated with the epidemiologic status of the year (Epidemic >5 human plague cases; non-epidemic ≤5 human plague cases) to test for the impact of network properties on the outbreaks.

Statistical Analysis of Eco-Epidemiologic Data

Comparison between the epidemiologic status of the year with continuous variables was performed by linear regression. The R-squared for the rodents' capture success were calculated using an exponential (one-phase decay) nonlinear regression. Principal component analysis (PCA) was calculated selecting PCs based on eigenvalues method. Regressions, PCA and Receiver Operating Characteristic (ROC) curves were calculated using GraphPad Prism v10.1 (Dotmatics, Boston, USA). The Sankey diagram was designed using the Sankey Diagram Generator (http://sankey-diagram-generator.acquireprocure.com). Statistic tests were run using a 95% confidence interval and considered significant when $p<0.05$.

Linear Discriminant Analysis

To test the contribution of variables to the prediction of plague epidemics by year, the years were assigned to groups according to the results from the PCA analysis: epidemic years (when human cases >5) and non-epidemic years (when human cases ≤5). We carried a linear discriminant analysis (LDA) with cross validation with different combinations of the epidemiologic variables (Rodent capture success, Flea positivity, Rodent positivity, Flea index) and network variables (modularity, vector robustness, host robustness). The discriminant analyses were carried only with the epidemiologic variables and with all combinations of network variables. A contingency table was obtained based on the different combination of variables which years were classified as epidemic and not epidemic as well as the probability of classification of each year.

Whole-Genome Sequencing

The *Y. pestis* cultures isolated in the Araripe Plateau are maintained in the cultures collection (Fiocruz/CYP) of the SRP from the IAM. In previous studies, 407 cultures were recovered and their genomes sequenced and analyzed (*9*,*10*). Here we revived and sequenced 33 additional strains that were originally contaminated, using the CYP broth medium (*11*).

DNA extraction was carried out using the DNeasy Blood & Tissue Kit (Qiagen) in accordance with the manufacturer's recommendations. Subsequently, the extracted DNA was quantified using the Qubit® 4 Fluorometer and the Qubit dsDNA High Sensitivity Kit (Thermo Fisher Scientific Inc.). Genomic library preparation was conducted using the Nextera XT Library Preparation Protocol (Illumina Inc.). Subsequently, library quantification and normalization were conducted using the ProNex® NGS Library Quant Kit (Promega) through Quantitative Real-Time PCR (qPCR) for standardized sample input using the 7500 Fast Real-Time PCR System (Applied Biosystems). After quantification, the samples were diluted to 19 pmol in ultrapure water and transferred to a single microtube containing Hybridization Buffer HT1 (Illumina Inc.). Subsequently, they were denatured at 96°C for 2 minutes in a thermocycler and a total of 600 µL of the denatured samples was loaded into a MiSeq Reagent Kits v3 (Illumina Inc.).

Genome Assembly, Annotation, and Core Genome

The quality of the sequencing data of the 439 Brazilian strains of *Y. pestis* was evaluated using FastQC 0.11.8 and the results were grouped by MultiQC v1.10 as previously described

(*12*,*13*). The sequencing metrics and metadata for the 33 newly sequenced Brazilian strains are available in Appendix 2 Tables 4, 5. Sequencing data was filtered using Trimmomatic v. 0.38 (*14*) and the assembly was performed using the VelvetOptimiser. To perform gene predictions and functional annotations of the newly assembled genomes, the Prokka pipeline (*15*) was used as described by Pitta *et al.* (*9*). GenomeTools 1.5.8 (*16*) was used to evaluate the annotations performed.

SNV Calling and SNV Core Genome

We conducted SNV calling with the Snippy software (*17*) and the CO92 strain was used as a reference genome. The results obtained were subsequently employed to generate a core SNV alignment, which served as the basis for the subsequent steps involved in constructing a phylogenetic analysis. Given the substantial volume of genomes, the 'snippy-multi' script (an integral component of the Snippy package) was employed for the automated analysis of all genomes in comparison to the reference genome.

Phylogenetic Analyses

The alignments were used as input for IQ-TREE2 (*18*) to perform phylogenetic analyses based on the core SNV profile with a bootstrap value of 1,000 and the ultrabootstrap approximation technique to enhance computational efficiency and reduce analysis time. Additionally, the ModelFinder tool (*19*), integrated into IQ-TREE2, was employed to automatically assess the best substitution model for the phylogenetic analysis. The results were visualized with the iTOL (Interactive Tree of Life) web platform (*20*).

NextStrain Dataset

We retrieved from the NextStrain dataset 538 *Y. pestis* genomes (filters: samples dated later than 1900 and not from Brazil). The metadata from the web service was subsequently downloaded, and based on the bioproject associated with each sample, a list for the download of these genomes was generated through the NCBI web service. Out of the 538 genomes, 64 are no longer available on NCBI; therefore, a total of 474 *Y. pestis* genomes were successfully downloaded (Appendix 3). A variable regarding the geographic focus was created for the phylogenetic analysis.

Characterization of the Phylogenetic Subgroups SNVs

Based on the variant call file (VCF) generated by Snippy during the core SNP analysis and the phylogeny constructed from this analysis, we identified which the apomorphic SNVs for each of the phylogenetic subgroups deriving from the "Araripe ancestor" group. Subsequently, SnpEff was used to predict the functional impacts of these variants in the highlighted samples, focusing on changes such as amino acid substitutions. The mutation list is available in Appendix 2 Table 1.

References

1. Bahmanyar M, Cavanaugh DC. Plague manual. Geneva: World Health Organization; 1976.

2. Karimi Y. Rapid laboratory diagnosis of plague [in French]. Bull Soc Pathol Exot. 1978;71:45–8.

3. Dormann CF, Fründ J, Blüthgen N, Gruber B. Indices, graphs and null models: analyzing bipartite ecological networks. The Open Ecology Journal. 2009;2:12917–22. https://doi.org/10.2174/1874213000902010007

4. Dunne JA, Williams RJ, Martinez ND. Food-web structure and network theory: The role of connectance and size. Proc Natl Acad Sci U S A. 2002;99:12917–22. PubMed https://doi.org/10.1073/pnas.192407699

5. Almeida-Neto M, Guimarães P, Guimarães PR Jr, Loyola RD, Ulrich W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos. 2008;117:1227–39. https://doi.org/10.1111/j.0030-1299.2008.16644.x

6. Memmott J, Waser NM, Price MV. Tolerance of pollination networks to species extinctions. Proc Biol Sci. 2004;271:2605–11. PubMed https://doi.org/10.1098/rspb.2004.2909

7. Beckett SJ. Improved community detection in weighted bipartite networks. R Soc Open Sci. 2016;3:140536. PubMed https://doi.org/10.1098/rsos.140536

8. Vázquez D, Melian CJ, Williams NM, Blüthgen N, Krasnov BR, Poulin R. Species abundance and asymmetric interaction strength in ecological networks. Oikos. 2007;116:1120–7. https://doi.org/10.1111/j.0030-1299.2007.15828.x

9. Pitta JLLP, Bezerra MF, Fernandes DLRDS, Block T, Novaes AS, Almeida AMP, et al. Genomic analysis of *Yersinia pestis* strains from Brazil: search for virulence factors and association with epidemiological data. Pathogens. 2023;12:991. PubMed https://doi.org/10.3390/pathogens12080991

10. Vogler AJ, Sahl JW, Leal NC, et al. A single introduction of *Yersinia pestis* to Brazil during the 3rd plague pandemic. PLoS One. 2019;14:e0209478. PubMed https://doi.org/10.1371/journal.pone.0209478

11. Rocha IV, Andrade CAN, Sobreira M, Leal NC, Almeida AMP, Bezerra MF. CYP broth: a tool for *Yersinia pestis* isolation in ancient culture collections and field samples. Appl Microbiol Biotechnol. 2023;107:2653–60. PubMed https://doi.org/10.1007/s00253-023-12452-0

12. Andrews S. FastQC: a quality control tool for high throughput sequence data [cited 2020 Apr 21]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc

13. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8. PubMed https://doi.org/10.1093/bioinformatics/btw354

14. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. PubMed https://doi.org/10.1093/bioinformatics/btu170

15. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. PubMed https://doi.org/10.1093/bioinformatics/btu153

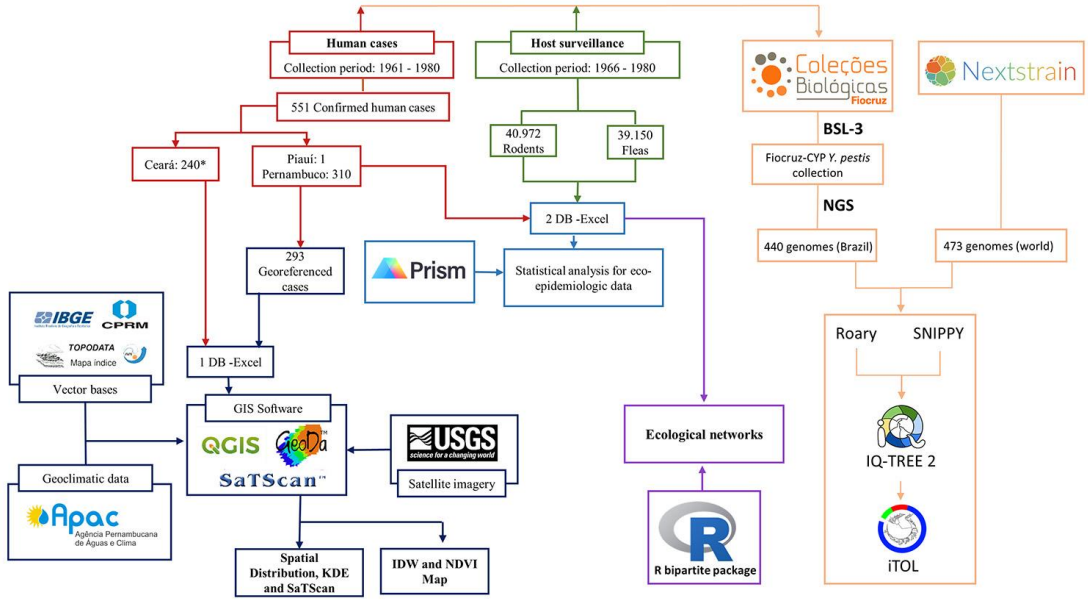
16. Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans Comput Biol Bioinformatics. 2013;10:645–56. PubMed https://doi.org/10.1109/TCBB.2013.68

17. Seemann T. snippy: fast bacterial variant calling from NGS reads [cited 2023 Sep 5]. https://github.com/tseemann/snippy

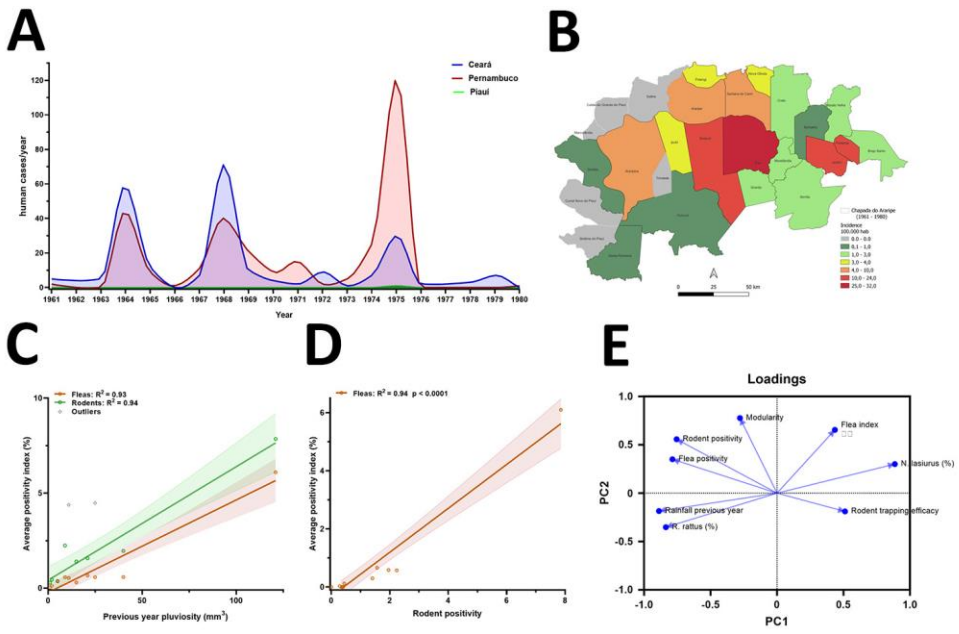
18. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4. PubMed https://doi.org/10.1093/molbev/msaa015

19. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. PubMed https://doi.org/10.1038/nmeth.4285

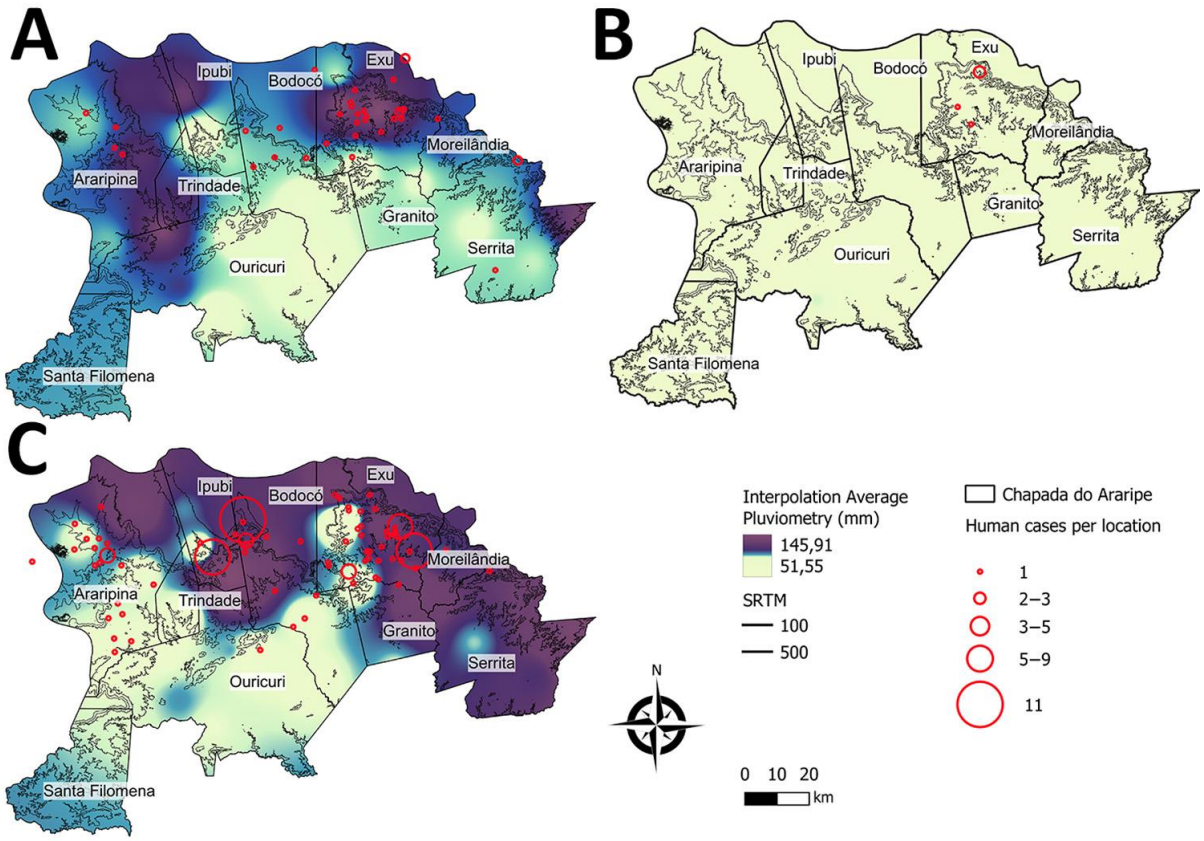
20. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–9. PubMed https://doi.org/10.1093/nar/gkz239



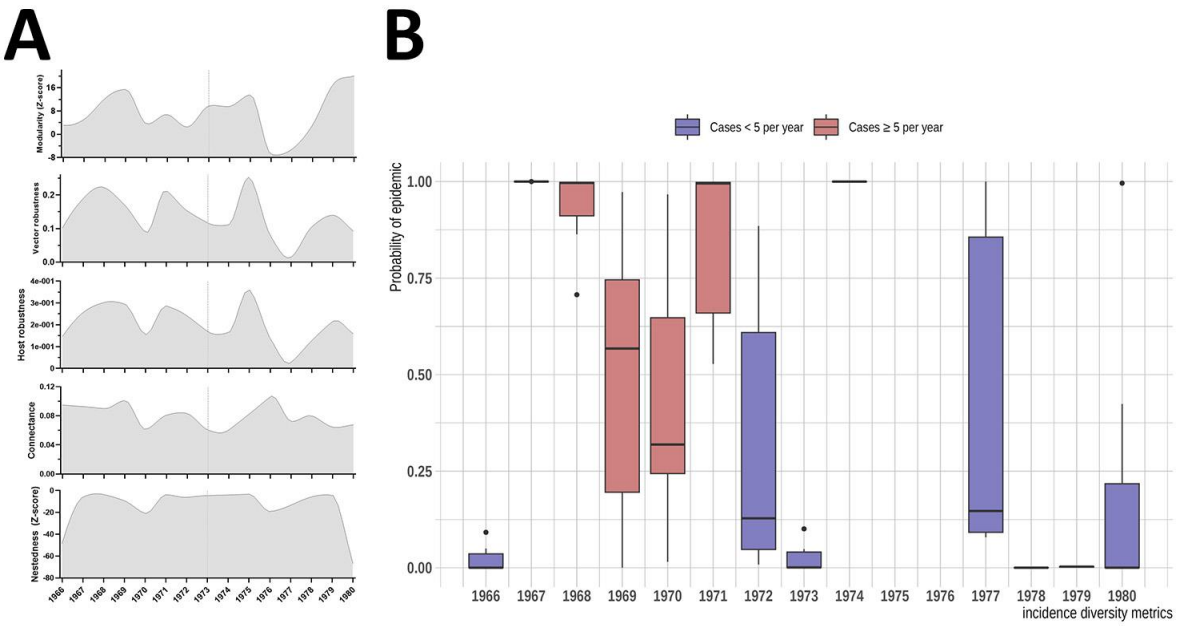
Appendix 1 Figure 1. Summary of the study workflow. The study uses multiple data sources including epidemiologic, ecologic, geospatial, climatic and genomic data. *Data from human cases from Ceará is limited to the total amount per municipality per year.



Appendix 1 Figure 2. Human cases of plague in the Araripe Plateau and support information for Figure 2. A) Plague cases according to the state (Pernambuco, Ceará, and Piauí). B) Colormap showing the average plague incidence per municipality (1961–1980). C) Correlation between flea and rodent positivity rate with previous year pluviosity. D) Correlation between flea and rodent *Yersinia pestis* positivity rates. E) Loading plot of the PC1 and PC2, complementary to the PCA plot in Figure 5.



Appendix 1 Figure 3. Pluviometry (IDW Interpolation) and human cases of plague in Araripe Plateau focus in Pernambuco. A) Spatial distribution of the 1967 average rainfall and 1968 plague cases. B) 1972 average rainfall and 1973 plague cases and C) 1974 average rainfall and 1975 plague cases.



Appendix 1 Figure 4. Annual variation of ecologic networks metrics and LDA. A) Each plot describes one aspect of the ecologic networks in the Araripe Plateau during the study period. The values were calculated based on the relative proportion of interactions of fleas in small mammal species. B) Linear discriminant analysis with cross validation with different combinations of the epidemiologic variables. The bars show the probabilities of classification for all models. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The whiskers extend from the hinge to the largest/smallest values no further than 1.5 interquartile ranges and outliers are plotted individually.