Identification of Chinese plague foci from long-term epidemiological data

Tamara Ben-Ari[a,b,1], Simon Neerinckx[a], Lydiane Agier[a], Bernard Cazelles[b,c], Lei Xu[d], Zhibin Zhang[d], Xiye Fang[e], Shuchun Wang[e], Qiyong Liu[e,2], and Nils C. Stenseth[a,2]

[a]Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, N-0316 Oslo, Norway; [b]Centre National de la Recherche Scientifique, Ecole Normale Supérieure, Unité Mixte de Recherche 7625, Université Pierre et Marie Curie, 75230 Paris Cedex 05, France; [c]Unité de Modélisation Mathématique et Informatique des Systèmes Complexes, UMI 209, Institut de Recherche pour le Développement et Université Pierre et Marie Curie, 93142 Bondy Cedex, France; [d]State Key Laboratory of Integrated Management on Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; and [e]State Key Laboratory for Infectious Diseases Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

Edited* by Hans R. Herren, Millennium Institute, Arlington, VA, and approved April 2, 2012 (received for review July 1, 2011)

Carrying out statistical analysis over an extensive dataset of human plague reports in Chinese villages from 1772 to 1964, we identified plague endemic territories in China (i.e., plague foci). Analyses rely on (*i*) a clustering method that groups time series based on their time-frequency resemblances and (*ii*) an ecological niche model that helps identify plague suitable territories characterized by value ranges for a set of predefined environmental variables. Results from both statistical tools indicate the existence of two disconnected plague territories corresponding to Northern and Southern China. Altogether, at least four well defined independent foci are identified. Their contours compare favorably with field observations. Potential and limitations of inferring plague foci and dynamics using epidemiological data is discussed.

Yersinia pestis | *Ecological Niche Modeling* | wavelet analysis | plague dynamics

Plague is an ancient disease caused by the bacterium *Yersinia pestis*. It is hosted by small mammals (mainly rodents) and vectored through bites from infected fleas (1). In the natural cycle of plague, humans are secondary hosts (i.e., they do not participate in plague's maintenance cycle). Human plague reports are nonetheless the main available source of worldwide data on plague (2), with the notable exception in Asia of Kazakhstan and other previous Soviet Union countries (3). It is therefore important that we find ways to efficiently use human data to infer where and how plague is endemically maintained.

Plague has a long epidemic history in China (4). It has been claimed (5, 6) that the first two plague pandemics both originated in China. To our knowledge, the oldest recorded evidence of plague can be found around 1353 in the northeastern Hebei province (4) (Fig. S1 provides cited provinces). The third pandemic most likely started in China is responsible for the death of 12 million people in China and India alone (7). It appeared in Hong Kong in 1894 and spread to major world ports through merchant ship transports (8). The plague source for the third pandemic may have been northwestern Yunnan province, where it remained confined since the earliest confirmed records in the city of Dali (in 1772 or earlier according to different sources) (9). Expansion of plague to the southeast occurred during the first part of the twentieth century, while human plague incidence also increased in northern China (10). In modern times, plague remains an important public health concern to Chinese authorities as exemplified by the 2009 pulmonary outbreak in the city of Ziketan, Qinghai province (11).

There have been few studies focusing on occurrences of human plague in China. The most comprehensive study (12) revealed that plague intensity shows a positive correlation with a wetness index up to a threshold level dependent upon location in northern or southern China. Another study also linked plague intensity in China to climate at the province scale (13): Their results emphasized the existence of different, separated plague environments within these predetermined regions. Here we relax the administrative constraint on the data and strive to objectively establish the contours of plague foci.

Our dataset covers the period 1772–1964 over all of China. It is potentially hindered by some shortcomings, which we discuss in detail. The time coverage includes epidemic (respectively endemic) periods, that is, periods when above (respectively. below) number of villages reported plague. Endemic periods, when outbreaks are of smaller amplitude and localized in space, are of great interest to us because they provide the opportunity to identify territories where plague is maintained without regular external input. Indeed, the main focus of this paper is to identify plague foci with a typical resolution of hundreds of kilometers. We thus go beyond simple data aggregation (*Preamble on the Data*) and to define foci by a combination of similarity in the local time series (TS) (*Clustering*) and similarity of environments [*Ecological Niche Modeling (ENM)*].

First, local TS are clustered on the basis of a decomposition using wavelet decompositions. Subsequently, local time series are classified according to their degree of wavelet spectrum resemblance (14). This allows us to identify common patterns without any a priori assumption on time series stationarity or on the consistency of plague dynamics within a defined geographical unit. We show that local TS tend to display spatial coherency. Three regions robustly emerge from our analysis, namely a southwest, a northwest, and an eastern area.

Second, we use ecological niche modeling (15–18) [see *Ecological Niche Modeling (ENM)*] to characterize and segregate different plague niches in China. Results from the heuristic, spectral, and ecological approaches are reconciled in *Discussion* to propose a map of plague enzootic foci in China.

Preamble on the Data

The dataset is a compilation of human plague cases recorded in local gazetteers (or difangzhi) (19). We calculated a yearly number of plague locations within a regular grid. Fig. S2 presents the location of these in 0.5*0.5° grid cells separated in three time windows corresponding to below (period 1: 1772–1893 and period 3: 1953–1964) or above (period 2: 1894–1952) average number of village reports for all of China. Separation years (vertical lines in Fig. S2 *B–D*), 1894 and 1952, correspond to some outstanding dates for plague. The former is the start of the third pandemic and the latter the end of the Chinese civil war and

Author contributions: T.B.-A., S.N., and N.C.S. designed research; T.B.-A., S.N., and L.A. performed research; T.B.-A., S.N., L.A., B.C., L.X., Z.Z., X.F., and S.W. analyzed data; and T.B.-A. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]Present address: Institut National de la Recherche Agronomique (INRA), 147 rue de l'université 75007, Paris, France.

[2]To whom correspondence may be addressed. E-mail: liuqiyong@icdc.cn or n.c.stenseth@bio.uio.no.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1110585109/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1110585109

the start of the plague control program around 1950 (20). We will henceforth name periods 1 and 3 (Fig. S2 *B* and *D*) endemic as opposed to the epidemic period 2 (Fig. S2*C*). During the first period, plague only occurs in southern China, mainly Yunnan and to a lesser extent in the Guangdong and Fujian provinces; there are only sporadic reports of plague in the northern provinces of Hebei and Inner Mongolia. The beginning of the third plague pandemic in 1894 led to high plague incidence over the most extended area. Finally, the short period 3 extends from 1952 to 1964 and corresponds to a contraction of the plague area to the western part of Qinghai province and the eastern part of the Inner Mongolia province. Noticeably, this period is characterized by a drastic reduction of plague reports in the entire south. Overall, endemic periods suggest the existence of five separate portions of territory subsequently called northwest (NW, a few dispersed occurrences), north central (NC), northeast (NE), southwest (SW), and southeast (SE), in which there are more than one report per grid cell. Note that grid cells with rarer plague occurrences (1 or 2 to up to 10 villages) tend to surround the region with the highest plague levels, suggesting that these plague territories are composed of "endemic cores." A visual determination of endemic areas indeed requires the exclusion of epidemic period 2 and of all grid cells with only one plague occurrence in periods 1 and 3. The former period defines territories likely to have been infected as a result of a spillover and the latter corresponds to a high probability of an accidental one-time importation. We are thus left with four portions of territory (excluding NC, which was plague-free outside the epidemic period): a possible subdivision of SE into two subregions and a large contrast in activity between northern and southern territories, southwest and to a lesser extent the east being very active before 1894 and nearly inactive after 1952. Conversely, most of the plague activity in the north was concentrated after 1894, becoming the dominant plague area after 1950. The data themselves thus suggest four endemic territories in China throughout 1772–1964. In *Results*, we present objective ways to delineate contours of endemic plague territories and elaborate on these aspects in *Discussion*.

Results

Clustering. Here, the entire dataset in its gridded form is used to identify large-scale patterns of plague occurrence dynamics. Our method (*SI Methods* contains descriptions) reveals spatial coherency of local TS (Fig. 1*A*). The distribution of these groups is mostly independent from the number of groups, N. Precisely, for $n \geq 3$, groups emerge that gather a large number of grid cells: SW, the western part of SE and NW for group 1 TS, NC for group 2, and the eastern part of NE plus SE for group 3 TS. Groups 1 and 3 thus cover the four suspected nonepidemic plague territories revealed by Fig. S2, whereas group 2 TS are associated with the 1894–1952 epidemic period. The robustness of the aggregation is demonstrated by the comparison between grouping results for $n = 3$ and $n = 6$. Adding more groups only removes a very limited number of outliers from each group (not shown for clarity). Other parameters than N could also influence our results. To check their robustness, we carried out a series of sensitivity tests. A small number of TS (<10%) may belong to one group or the other depending on changes in the parameters (e.g., N), which does not invalidate the main results above or the interpretations made in the discussion.

TS for each of the three groups are presented in Fig. 1*B* and reflect the importance of time patterns in the grouping. Main characteristics of each group TS are as follows: reports distributed throughout the whole time period with a long period of above-normal (epidemic) conditions for group 1, a vast majority of plague reports between 1937 and 1952 for group 2 and finally, two to three bursts in 1917–1918, 1928, and 1932 with recurrent low incidence reports after 1900 (no cases before) for group 3. Noticeably, the two coherent and geographically distinct sets that form group 3 remain aggregated for up to $n = 15$. Also, rerunning the clustering for that subset does not indicate major dissimilarities within it (this is also true for groups 1 and 2, not shown for clarity). Connection between southeast and northeast territories is an important result of this method that we downplay in *Discussion*.

Ecological Niche Modeling (ENM). ENM allows the characterization of environmental differences between plague territories. To avoid including accidental/nonsignificant occurrences and imported cases, we use plague occurrences having more than one plague occurrence over the nonepidemic periods (henceforth labeled "valid") as model inputs. Here we present results obtained using the Maxent software but we also used GARP. Maxent is known for overfitting the data and GARP for overpredicting the extension of suitable areas (21, 22) (Fig. S3 provides a visual comparison of the outputs of two runs with identical inputs). However, the fact that both models predict essentially similar suitable territories (not shown for clarity) when run over regional input subsets (black dots in Fig. 2) gives us confidence in the robustness of the conclusions.

We first investigate environmental differences between north and south territories by dividing the valid set of plague occurrence points into northern and southern subsets (Fig. 2, *Middle* black dots). Two well-circumscribed niches are predicted by ENM that each roughly coincide with their respective occurrence points. Most noticeably, none of the areas predicted by any

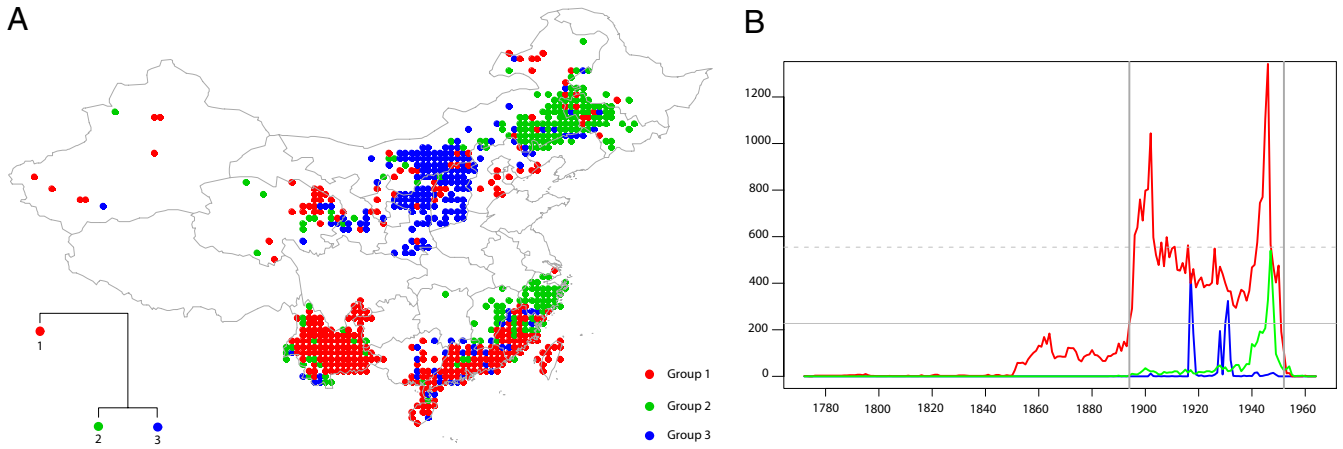

Fig. 1. Cluster analysis output. (*A*) Colors indicate the cluster group to which the local time series belongs. Note the coherent spatial distribution of time series with similar time-frequency patterns. (*B*) Aggregated time series for each cluster group (same color coding). In this and in the following figures, only plague infected or related regions are represented.

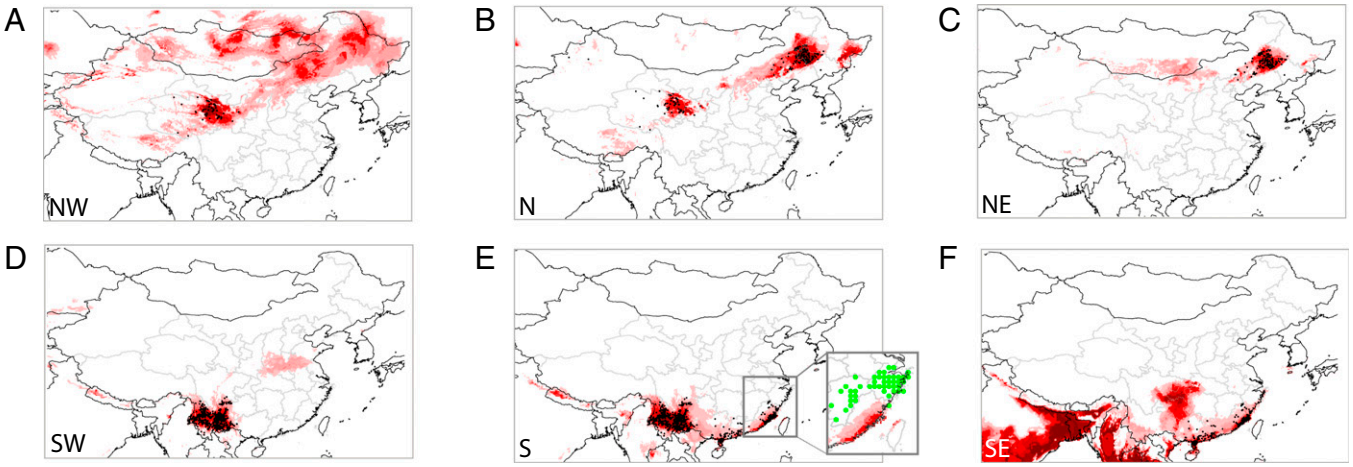

Fig. 2. Ecological niches associated with each of the six subsets corresponding to NW (*A*), N (*B*), NE (*C*), SW (*D*), S (*E*), and SE (*F*). Intensity of color indicates level of suitability index (darker stands for higher suitability). The central north territory is not considered because it contains no plague report outside the epidemic period of 1894–1951.

individual subset encompasses another subset's realm. They are well separated from each other by the central China area that our ENM model successfully predicts as plague-free. The area predicted by the southern subset encompasses northern Viet Nam where plague is known to have occurred (23, 24) and parts of Nepal and northeastern India (which is not inconsistent with ref. 25). Other areas are predicted as suitable for plague but cannot be completely confirmed due to the scarcity of data (e.g., Mongolia or Russian borders). In *Discussion* we elaborate on why a territory could be suited for plague but have no known expression in humans. For several retained environmental variables, little or no overlap is observed between northern and southern niches. Plague is predicted in very dry to arid areas in the north. The southern niche is generally wetter (annual precipitation ranging from ~750–1,700 $mm.y^{-1}$ in the south versus 0–700 $mm.y^{-1}$ in the north) and warmer (8–27 °C versus −5–8 °C). This niche has significantly fewer marked temperature seasonal fluctuations as revealed by the annual temperature range (20–25 °C with outliers around 30 °C compared with 35–50 °C in the north). Interestingly, the total predicted niche (north + south) covers roughly the entire available value range for most of the input chosen WorldClim (26) variables. One exception is that endemic plague is not predicted for the coldest temperatures (i.e., areas with an average annual temperature below −5 °C that also have the highest annual temperature range).

To confirm the regional coherence found with the cluster analysis (*Clustering*) and the visual pooling presented in *Preamble on the Data*, we now subdivide the set of valid plague occurrence points into the four patches that geographically stand out in Fig. S2 *B* and *D*. ENM-predicted plague niches for each regional subset are represented in Fig. 2. The SW niche predicts an area corresponding to the Yunnan province. The SE niche covers Fujian, Guangdong, Guangxi, and Guizhou, the latter being outside the epidemiological plague area. The NE niche almost exactly matches the region delimited by its associated plague occurrence points. The NW niche extends over a much broader territory that covers most of northern China. The predicted area spans the NE territory and successfully includes the NC area where plague is observed only during the epidemic period and was thus excluded from the ENM input subsets (but see finer scale investigation in Fig. S4).

Distribution ranges for the seven input environmental variables across subsets (Fig. S5) show that differences are comparatively smaller within the northern/southern niches than between them. This is not unexpected because most relevant ENM variables vary primarily with latitude. We also find well-defined and separated ranges for climate variables associated with each subset. In particular, the ranges for the NW niche are surprisingly restricted in regard to its geographical coverage. This niche corresponds to the coldest and driest plague territory (Fig. S6). Conversely, the warmest and wettest territory is the SE, which is noticeably different from NE with respect to average temperature and precipitation.

Discussion

Using epidemiological data to investigate the distribution of plague foci in China requires caution due to the complex and poorly understood relationship between human plague cases and enzootic/epizootic cycles of the disease (1, 27). Humans do not belong to the maintenance cycle of plague; they are accidental hosts most likely to get infected when an epizooty is already underway (i.e., when plague has already spread outside of its strict enzootic area). In the present study we attempt to neutralize the limitations imposed by the use of epidemiological data by carefully selecting predicting points. For a stationary human population (i.e., with fixed size, spatial distribution, commerce, and migration), human plague cases and locations would provide a partial, possibly unbiased view of the natural cycle of plague. Obviously, some rescaling (e.g., by population size) would be necessary to allow quantitative conclusions on plague dynamics. During our study period, dramatic sociodemographic changes and events including urbanization, warfare, and transit system development, could have altered both plague dynamics and data collecting (19). Hence the stationarity condition is not satisfied. Regardless we argue that this unique long-term dataset contains important information on plague. We are convinced that these data can help reveal qualitative aspects of plague dynamics in China and identify underlying spatial patterns, namely plague foci (or at least endemic plague territories, see *Conclusion*) and the connection between them.

North/South Distinction. The most evident characteristic of Chinese plague is the existence of at least two distinct plague territories; one in the north and one in the south. Support for this can first be found in the ENM analysis that predicts two well-separated niches corresponding to northern and southern subsets and distinct ranges for their environmental variables. This is also corroborated by the fact that our data contain no plague reports in the in-between area (from about 30–35° N with small variations depending on the longitude) although it is densely populated. Field observations also confirmed that no suitable territory exists in central China (20). This indicates that there is no evidence of a plague pathway joining northern and southern plague territories. In addition, ENM prediction results show that none of the four subsets has environmental conditions that project into central China (Fig. 2 *A*, *C*, *D*, and *F*). Biomes and WorldClim data are consistent with

the existence of two environmentally distinct territories: an herbaceous cover with increasing elevation toward the west in the north versus evergreen tree or shrub cover in the south (Fig. 3). Therefore, a conclusion from our findings is that long-term coexistence of all three plague components (i.e., pathogens, rodents, and fleas) is possible under distinct environmental conditions.

Northeast/Southeast Puzzle. As group 3 (green dots in Fig. 1*A*) encompasses both northeast and southeast territories, the clustering analysis may seem more ambiguous about a north/south disconnect. However, the procedure through which TS have been aggregated needs to be considered. In fact, visual inspection of the local TS (Fig. 1*B*) indicates that TS composing group 3 share the following characteristic: a pronounced reported incidence of plague in the short period 1937–1952 not seen in the other groups (fraction of gridded counts of infected villages during 1937–1952 over the entire period are 0.65 for this group versus 0.25 for group 1 and 0.23 for group 2). The grouping of northeast/southeast territories together in group 3 may thus occur for circumstantial reasons that should not be overinterpreted. These areas simultaneously underwent Japanese invasion during 1937–1945. In fact, according to Liu (20), plague was repeatedly spread by the Japanese army in the southeast during 1940–1948. The complete absence of overlap between SE and NE ENM niches also provides evidence against a plague focus spanning the entire group 3; they do not share similar environmental traits (Figs. S5 and S6). In fact, ENM analysis further suggests that the southern subset of group 3 may not even be part of a plague enzootic territory. Indeed, the corresponding infected gridded incidences only occurring during the war period sit outside the niche predicted by southern and SE occurrence points (green dots in Fig. 2 *E*, *Inset* and *F*). This is generally consistent with the reconstruction of plague spread proposed by Benedict (4) that involves contact diffusion (i.e., spread outward from a central focus) from Fujian to Zhejiang during the epidemic period of 1884–1949. This is also consistent with the idea that, during epidemics, plague spills over from favorable (i.e., where plague is enzootic or at least can maintain itself over relatively long periods of times without being regularly imported) to unfavorable environments such as the area where the southern green dots are located (Fig. 2*E*, *Inset*). The southern points that make up group 3 do not fulfill the conditions to qualify as an enzootic plague territory and do not refute the existence of a north/south plague disconnect.

Southern Plague Region. Plague habitats in the south are quite similar (Fig. 3): tropic/subtropic forest zone with homogeneous annual temperature range and average and, although lower in Yunnan, comparable annual precipitation (Fig. S5 *A*, *D*, and *E* and also ref. 20). Despite this relative coherency, SW is characterized by higher elevations probably responsible for observed differences in mean diurnal range and precipitation seasonality (Fig. S5 *B* and *G*).

Our epidemiological data suggest that at least two plague regions exist in southern China with some degree of independence between them and most of the Guanxi province (plague-free during the entire study period) as a physical separation. The first region corresponds to Yunnan (SW) and the second one to

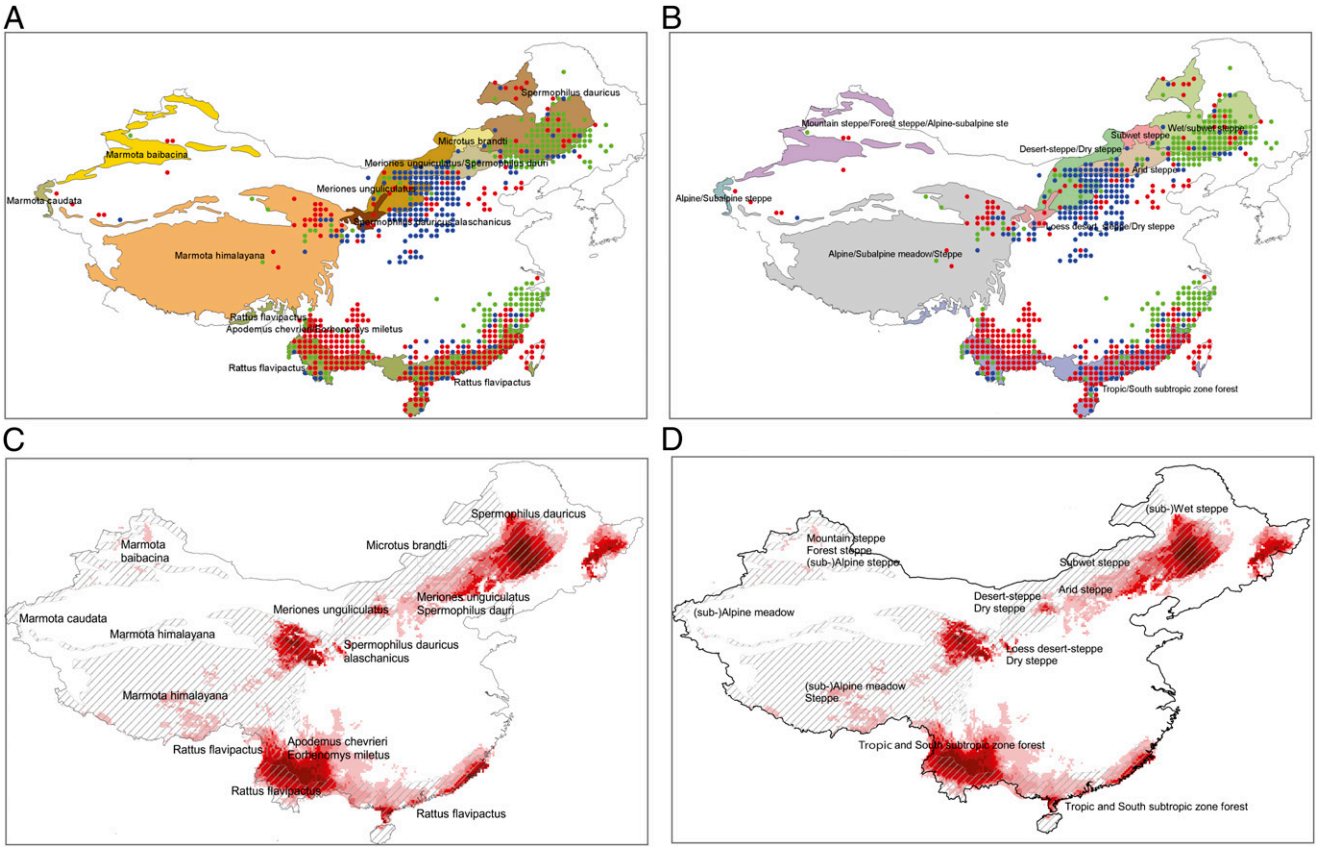

Fig. 3. Correspondence between inferred and sampled plague foci. (*A*) Human plague locations (color-filled circles; circles of a given color correspond to gridded occurrence time series that have similar time/frequency behavior) superimposed on the suspected dominant plague host territories (obtained by georeferencing the map of rodents tested positively to plague antibodies as reported in ref. 20). (*B*) Similarly with a georeferenced biome map from the same source. (*C*) Map cumulating plague niches predicted with northern and southern occurrence points as input subsets using ENM superimposed on the suspected dominant plague host territories. (*D*) Similarly with a georeferenced biome map from the same source. (*A* and *B*) Occurrences tend to extend outside the contours of host territories; this is simply because the classification is based on all data, whereas ENM is based only the valid set of inputs (i.e., nonepidemic occurrences). (*C* and *D*) Note good correspondence between niches predicted with climate variables and experimentally determined enzootic foci.

Guangdong and Fujian (S and SE). Support for the existence of two separate regions can be found in the relative number of plague reports as a function of time: the first plague record outside Yunnan is in Beihai (Juangxi province) in 1867 (i.e., plague remained within the boundaries of Yunnan for at least a hundred years after the first confirmed record in its borders) and the first occurrence in Fujian is in 1884 at a time of mild epidemic conditions in Yunnan. On the other hand, no plague cases were reported in Yunnan from 1956 until 1964, whereas concomitant sporadic occurrences were reported in SE. ENM analysis would also be consistent with the existence of two subregions in the south. The SW predicted niche does not include eastern territories, whereas the SE subset of occurrence points predict a wider niche that spans into southeast Asia and some areas in Yunnan, albeit with a lower suitability index. Interestingly, field studies conclude that *Rattus flavipectus* is the unique dominant plague host for the whole south (Fig. 3) but with variation: *Apodemus chevrieri* as a secondary host in some areas in the south and SE and *Eothenomys miletus* in the highland of western Yunnan (20). It is difficult, at this resolution to strongly argue in favor of either a unique or two plague foci in the south.

Physiographic macroregions (28) are self-contained regions with major geomorphologic features (constraining travel and exchanges, thus coherent in terms of economic resources and development) and drainage basins. The totality of what we have called southern China is separated into four macroregions: Yungui, roughly corresponding to the Yunnan province; Lignan, to Guangxi and Guandong; southeast coast, roughly to Fujian and Zhejiang; and part of lower Yangzi, roughly to Jiangsu. According to Benedict's work (4), plague first slowly spread within Yungui. During the second part of the 19th century, coincident with the Muslim rebellion (1855–1873), it spread outside of its source province and spread to Lignan, the southeast coast and lower Yangzi. The reconstruction work by Benedict shows how plague spread along trade routes and within the boundaries of the regional trading system, explaining the link between plague dynamics and Chinese macroregions. Cities and ports served as distribution centers where infected rodents or fleas were passively transported in merchandise (4). According to Benedict, a secondary spread then would have occurred from newly infected cities inland. This is consistent with a spread from the west to the east we have described and a subsequent settlement of plague within the contours of a favorable plague niche found in our ENM analysis (Fig. 2). This is also consistent with the similarity observed for the south plague system (similar dominant hosts and fleas, interannual plague dynamics, and environmental characteristics).

Northern Plague Territories. In contrast to the south, northern plague territories exhibit important heterogeneities and a wider variety of biomes (Fig. 3). Fig. S2 reveals that there could be three distinct patches of plague occurrence in the north: one in the east (NE as captured by group 3 in the cluster analysis; Fig. 1*A*) and two other, northwest (NW) and north central (NC, essentially group 2) that are somewhat geographically isolated from each other. Only two (NE and NW) would qualify as endemic Fig. S2*D*) because NC has no reporting outside the epidemic period.

NE local TS mostly belong to cluster group 3 because they have a concentration of reports in the 1937–1952 period (although less so than for the southern part of this group, as we have verified). We suspect this concentration should be even less marked had the 1910–1911 and 1920–1921 epidemics been fully reported (10, 29, 30) (partial control of the Russians and Japanese over the region may be responsible for underreporting or loss of information). Historical reports suggest an endemic focus corresponding to NE with no or little expression in indigenous humans; pneumonic outbreaks later occurred in relation to an increased hunting of marmots (31). Infection apparently followed rail routes throughout Heilongjiang and northeastern Inner Mongolia. There are several clues suggesting that the NE plague territory is isolated from the rest of the northern plague region. Most importantly, the ENM analysis using the NE subset predicts a very restricted territory corresponding to a cultivated wet/subwet steppe biome (Fig. 3 *B* and *D*) that differs from NC and NW in this regard.

In contrast, the NW/NC plague region is more heterogeneous. Although plague reports are essentially contiguous, our analyses suggest it is made of at least two distinct plague territories. Village reports for NC (corresponding to the Inner Mongolia, Shanxi, Shaanxi, and small parts of Gansu provinces) mostly belong to group 2, whereas those in the far west mostly belong to group 1 (Qinghai province). NW group 1 TS have plague reports over the entire study period, including at the very beginning of the time series, an indication of the enzootic nature of this plague territory. On the other hand, group 2 reports occur during the epidemic period with two major peaks in 1917–1918 and 1931, indicative of epidemic conditions (>180 villages for each peak). However, low levels of plague occurrences are also found throughout the 1900–1950 period (first and last reported cases in the region); in 4–10 y per decade, 1–17 villages reported plague. This is compatible with the existence of an enzootic area (ancestral as it has been suggested elsewhere (10) or more recently established due to successful and sustainable importation of plague) whose precise location can be inferred from the positions of villages infected during these low plague level years (Fig. S4). A complementary ENM analysis with NC group 2 locations as input predicts a niche that is roughly consistent with this set of points. The Mongolian gerbil (*Meriones unguiculatus*) and the Daurian ground squirrel (*Spermophilus dauricus*) could be the dominant plague hosts in most of the NC plague territory (Fig. 3). However, because NC occurrences are reported during the epidemic period, the geographical range of NC points is undoubtedly wider than that of the NC actual endemic core. In terms of biomes, NC occurrences occupy a region straddling several environments (desert, arid to dry steppes; herbaceous cover to bare areas). On the other hand, NW points that belong to group 1 are associated with an alpine-type biome with marmots (*Marmota himalayana* and *M. baibacina*) as main hosts.

All these elements (behavior of the time series and hence grouping; ENM niches, hosts, and biomes) suggest the existence of three plague foci in the north with an important degree of independence between them. Interestingly, there are three physiographic macroregions in the north: NE corresponds well to the northeast macroregion; both NW and parts of NC are included in the northwestern macroregion; to the contrary, the north macroregion as defined in ref. 28 is plague-free in our dataset, except for a few single occurrences that we deem insufficient to support the existence of an endemic plague focus there.

Note however that according to different sources (9, 29–31), plague was endemically present in northern territories and went unnoticed before the third pandemic when it spread as a result of population increase and transit system development (29, 30). Similarly, it is probable that plague presence in the NW was underreported because this territory is scarcely populated. Overall, we conclude the possibility of at least three independent foci in the north.

Conclusion

In this study, we have analyzed long-term and large-scale epidemiological data in an attempt to identify plague foci in China. There is one major conceptual difficulty when doing so: providing an operative definition of a plague focus is not straightforward. In practice we have relied on a definition that combines similarities in the local time series and in underlying environments. A definition consistent with the notion of enzootic focus: an area where plague can maintain itself without importation, is generally associated with a specific combination of host and vector communities.

Altogether, our analysis shows that plague was found in China in diverse environments associated with a variety of climates. Behind this diversity, we find five coherent plague territories that are likely to be distinct enzootic foci. A sixth possible plague foci was discarded despite reporting numerous plague cases because these are strongly concentrated toward the worst part of the

epidemic period. This territory is also environmentally dissimilar to the neighboring regions and therefore not a likely extension of nearby enzootic areas. Both types of analyses were helpful to achieve this separation. Analysis of the spatiotemporal distribution of human plague cases helps identify possible coherent territories in terms of plague dynamics and indicate where plague is robustly settled. Carrying ENM over the identified regions provides additional information by indicating the degree of independence between them. Importantly, our results are strikingly consistent with plague foci classification on the basis of host identification (Fig. 3).

A number of caveats must be mentioned. Our methods unavoidably tie human activity and plague dynamics, whereas human hosts are not essential for plague. We may also have missed enzootic foci that have little or no expression in terms of human plague cases. More generally, large-scale spatiotemporal analyses may not be enough to disentangle plague enzootic foci, diffusion patterns, the variations around these two concepts (temporary versus ancestral foci), and their possible links (diffusion–relocation, diffusion–activation) (4). An avenue for improvement may lie in more elaborate and finer-scale analyses that would help confirm our findings and presumably unravel the complex dynamics of plague at scales below the one we have considered.

Methods

Classification Method. The 744 time series, corresponding to the aggregated yearly number of villages reporting plague in a $0.5 \times 0.5°$ grid, are, as most often the case with epidemiological time series, nonstationary (i.e., their mean, variance, or frequency components change in time). We thus apply wavelet analysis where the original time series are decomposed into oscillatory components. In contrast to Fourier analysis, this time–frequency analysis captures the amplitude of frequency components as a function of time (32, 33). For further details see *SI Methods*.

Ecological Niche Modeling. We characterized spatial patterns of the landscape that are necessary for plague maintenance or transmission. Our approach is based on ecological niches defined as the set of environmental conditions under which a species is able to maintain populations without immigration (34, 35) (or, in this case, the conditions that prevail in plague endemic territories). Known occurrences of species (here plague occurrence points) can be related to raster coverage summarizing environmental variation (36). We used available bioclimatic layers at resolution ~10–14 km from the WorldClim (11) data archive (1950–2000). For further details see *SI Methods*.

To avoid fitting models in too many environmental dimensions, we chose the 7 least correlated ($P < 0.6$) of the 19 available variables (Fig. S5). These are average and range of annual temperature, average diurnal temperature range, isothermality, annual precipitation, precipitation of the driest month, and precipitation seasonality. We used two algorithms for ENM development based on presence-only data (no real absence points): the genetic algorithm for rule-set prediction (GARP) using the software DesktopGarp (38, 39) and via maximum entropy optimization approaches using the program Maxent (40–42). For further details see *SI Methods*.

ACKNOWLEDGMENTS. We thank Xavier Capet for comments and help in elaborating the structure of this manuscript; Carol Benedict, Mike Begon, and Life Christian Stige for many detailed comments; and the following individuals for compiling the historical plague data: Naiwu Chen, Jiefan Zhang, Guangming Wang, Chengzhong Yan, Yongling Zhao, Zongxiao Mao, Chongxi Lei, Kechang Gong, Yan Yang, Zizhong Yu, Gengxing Wang, Yuanmin Chen, Shengchang Tong, Di Zhang, Wu Ma, Zuyin Pan, Zhenhua Liu, Biaocheng Zeng, Bohong Pan, Jiabao Huang, Xiyi Fu, Hua Zhang, Daxi Lin, Renguang Hu, Rongxuan Shen, Liusheng Liu, Guanwen Wei, Wuyi Yang, Wenyuan Wong, Ligong Zhou, Aihua Huang, Zhenpan Hong, Hongqing Lu, Jingya Xiao, and Tianmin Sun. We also thank two anonymous reviewers. T.B.A. acknowledges support from Early Training Site Marie Curie and Center for Ecological and Evolutionary Synthesis postdoctoral funds. Z.Z. and L.X. acknowledge support from the National Basic Research Program (2007CB109101) of Ministry of Science & Technology (MOST); Q.L., X.F., and S. W. acknowledge support from the National Basic Research Program (2012CB95500) of MOST.

1. Gage KL, Kosoy MY (2005) Natural history of plague: Perspectives from more than a century of research. *Annu Rev Entomol* 50:505–528.
2. Stenseth NC, et al. (2008) Plague: Past, present, and future. *PLoS Med* 5:e3.
3. Stenseth NC, et al. (2006) Plague dynamics are driven by climate variation. *Proc Natl Acad Sci USA* 103:13110–13115.
4. Benedict CA (1996) *Bubonic Plague in Nineteenth-Century China* (Stanford Univ Press, Stanford, CA).
5. Little LK (2007) *Plague and the End of Antiquity. The Pandemic of 541–750* (Cambridge Univ Press in association with the American Academy in Rome, Cambridge, UK), p 380.
6. Cohn SKJ (2008) Epidemiology of the Black Death and successive waves of plague, ed Nutton V (Medical History, London), pp 74–100.
7. Perry RD, Fetherston JD (1997) *Yersinia pestis*—Etiologic agent of plague. *Clin Microbiol Rev* 10:35–66.
8. Link VB (1955) A History of Plague in the United States of America. Public Health Monographs, NIH, Bethesda, MD (no. 26).
9. Kraminskii VA (1964) *History and Geography of Plague in China* (US Army Biological Center, Fort Detrick, MD), pp 1–28.
10. Gray GD, Edin MD (1911) The septicaemic and pneumonic plague outbreak in Manchuria and Northern China (Autumn 1910-Spring 1911). *Lancet* 1152–1158.
11. Global Alert and Response (GAR) (2009) Plague in China (WHO, Geneva), Available at http://www.who.int/csr/don/2009_08_11/en/. Accessed April 20, 2012.
12. Xu L, et al. (2011) Nonlinear effect of climate on plague during the third pandemic in China. *Proc Natl Acad Sci USA* 108:10214–10219.
13. Zhang Z, et al. (2007) Relationship between increase rate of human plague in China and global climate index as revealed by cross-spectral and cross-wavelet analyses. *Integr Zool* 2:144–153.
14. Rouyer T, Fromentin JM, Stenseth NC, Cazelles B (2008) Analysing multiple time series and extending significance testing in wavelet analysis. *Mar Ecol Prog Ser* 359:11–23.
15. Peterson AT, Bauer JT, Mills JN (2004) Ecologic and geographic distribution of filovirus disease. *Emerg Infect Dis* 10:40–47.
16. Peterson AT, Lash RR, Carroll DS, Johnson KM (2006) Geographic potential for outbreaks of Marburg hemorrhagic fever. *Am J Trop Med Hyg* 75:9–15.
17. Peterson AT, Sánchez-Cordero V, Beard CB, Ramsey JM (2002) Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico. *Emerg Infect Dis* 8:662–667.
18. Giles J, Peterson AT, Almeida A (2011) Ecology and geography of plague transmission areas in northeastern Brazil. *PLoS Negl Trop Dis* 5:e925.
19. Vermeer E (1992) New county histories: A research note on their compilation and value. *Modern China* 18:438–467.
20. Liu Y, Tan J, Shen E (2000) *The Atlas of Plague and Its Environment in the People's Republic of China*, ed Inst. Geographical Sci. Nat. (Chinese Academy of Sciences, Bejing).
21. Elith J, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151.
22. Peterson AT, Pape M, Eaton M (2007) Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography* 30:550–560.
23. Cavanaugh DC, et al. (1968) Some observations on the current plague outbreak in the Republic of Vietnam. *Am J Public Health Nations Health* 58:742–752.
24. Cavanaugh DC, Marshall JD, Jr. (1972) The influence of climate on the seasonal prevalence of plague in the Republic of Vietnam. *J Wildl Dis* 8:85–94.
25. WHO (2000) *WHO Report on Global Surveillance of Epidemic-Prone Infectious Disease: Plague*, ed Response D of CDS (WHO, Geneva).
26. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25:1965–1978.
27. Ben-Ari T, et al. (2011) Plague and climate: Scales matter. *PLoS Pathog* 7:e1002160.
28. Skinner GW (1977) *The City in Late Imperial China* (Stanford Univ Press, Stanford, CA).
29. Teh WL (1923) The second pneumonic plague epidemic in Manchuria, 1920–21: I. A general survey of the outbreak and its course. *J Hyg (Lond)* 21:262–288.
30. Gamsa M (2006) The epidemic of pneumonic plague in Manchuria 1910–1911. *Past Present* 190:147–183.
31. Chernin E (1989) Richard Pearson Strong and the Manchurian epidemic of pneumonic plague, 1910–1911. *J Hist Med Allied Sci* 44:296–319.
32. Cazelles B, Chavez M, Magny GC, Guégan J-F, Hales S (2007) Time-dependent spectral analysis of epidemiological time-series with wavelets. *J R Soc Interface* 4:625–636.
33. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bull Am Meteorol Soc* 79:61–78.
34. Soberón J (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecol Lett* 10:1115–1123.
35. Soberón J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodivers Inf* 2:1–10.
36. Peterson AT (2006) Ecologic niche modeling and spatial patterns of disease transmission. *Emerg Infect Dis* 12:1822–1826.
37. Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.
38. DesktopGarp (2008) Available at http://www.nhm.ku.edu/desktopgarp/. Accessed April 20, 2012.
39. Stockwell D (1999) The GARP modelling system: Problems and solutions to automated spatial prediction. *Int J Geogr Inf Sci* 13:143–158.
40. Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Modell* 190:231–259.
41. Maxent (2009) Available at http://www.cs.princeton.edu/~schapire/maxent. Accessed April 20, 2012.
42. Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. *Ecography* 31:161–175.