

Spread of *Mycobacterium tuberculosis* lineage 4 in South China influenced by Maritime Silk Road and “Huguang Filling Sichuan” population migration

Wei Wu

Sichuan University School of Life Sciences

Peng-Kuan Liang

Sichuan University School of Life Sciences

Dong-Xia Luo

Public Health and Clinical Center of Chengdu

Yu-Ang Tian

Sichuan University School of Life Sciences

Zhong-Hua Jiang

Sichuan University School of Life Sciences

Zhuo-Chong Liu

Sichuan University School of Life Sciences

Mei Luo

Public Health and Clinical Center of Chengdu

Hai-Bing Yuan

Sichuan University School of Archaeology and Museology ☒ National Demonstration Center for Experimental Archaeology Education

Nalin Rastogi

WHO Supranational TB Reference Laboratory,

Qun Sun (✉ qunsun@scu.edu.cn)

Key Laboratory of Bio-resources and Eco-environment of the Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan, P.R. China. <https://orcid.org/0000-0002-4372-8865>

Research Article

Keywords: Mycobacterium tuberculosis, lineage 4, the Maritime Silk Road, “Huguang Filling Sichuan” population migration, interpolation analysis

Posted Date: April 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-388094/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Lineage 4 of *Mycobacterium tuberculosis* complex (MTBC), mainly epidemic in Europe and Americas, presents in high proportions in South China and is believed to enter China around the 13th century, when the important human migratory events of the Maritime Silk Road and “Huguang Filling Sichuan” population migration happened in China. This study was to explain the coincidence of these two events with lineage 4’s high proportion in South China.

Methods

Based on the spatial interpolation analysis of the genotyping data of 25,575 MTBC isolates, the distribution of lineage 4 was compared with that of targeted surname populations and the Maritime Silk Road’s main ports.

Results

The results showed that lineage 4 distribution in China could be mapped to the regions affected by “Huguang Filling Sichuan” population migration; while the distribution of lineage 4’s two sub-lineages in Asia, Europe, Africa and Oceania could be best explained by the Maritime Silk Road.

Conclusion

Our results suggest that these two events might pose a crucial shared influence, leading to the greater incidence of lineage 4 in South China. And this may contribute to our better understanding of the prevailing tuberculosis landscape in China and facilitate the epidemiological investigations and tracking of emerging MTBC clones.

Introduction

Mycobacterium tuberculosis (MTB) and other members of *Mycobacterium tuberculosis* complex (MTBC) lead to human tuberculosis (TB) [1]. In 2019, about 2 billion people were MTBC carriers in the world, of which 5%-10% might develop active disease [2]. Out of 9.96 million new TB cases and 1.21 million deaths observed among the HIV negative patients worldwide in 2019, China accounted for 833,000 new cases of TB and about 31,000 deaths, ranking third in the world [2]. According to the phylogenetic analysis based on whole-genome sequencing (WGS), MTBC was divided into seven major TB lineages [3–6]: Lineage 1 (Indo-Oceanic), Lineage 2 (East-Asian), Lineage 3 [East-African-Indian (EAI)], Lineage 4 (Euro-American), Lineage 5 (West-Africa 1), Lineage 6 (West-Africa 2) and Lineage 7 (Ethiopian or *Aethiops vetus* lineage). Stable associations between MTBC strains and their human host populations led to a distinct gradient of

phylogeographical distribution; e.g., lineage 4 is mainly epidemic in Europe and Americas, apparently associated with European and American people [7, 8]. In addition, two newly discovered lineages (Lineage 8 and Lineage 9) seemingly restricted to Africa were recently described [9, 10].

In China, more than 80% of MTBC strains belonged to lineage 2 [11], and lineage 4 accounted for about 15% [12]. The distribution of lineage 4 was uneven, which was significantly higher in some provinces of South China (Supplementary Fig. 1) than in other regions [12]. In particular, in Sichuan-Chongqing region (Supplementary Fig. 2), including provinces of Sichuan and Chongqing of South China, an area with underdeveloped traffic and reduced population mobility in ancient times, lineage 2 accounted for about only 60%-70% of all strains in MTBC, which was lower than the national level [13–15] while lineage 4 accounted for about 25%-35%, significantly higher than that of the whole country [12, 15, 16]. Because the expansion and evolution of MTBC are mainly caused by human doings, such as migration, trade, wars, etc. [12, 17–19], historical human activities might be an important cause of the distinctive distribution of lineage 4 in Sichuan-Chongqing region even the whole South China. Unfortunately, China's historical documents did not contain a description of any TB pandemic in South China [20, 21], making it unclear which historical events may have caused this singular distribution pattern observed.

Liu et al. suggested that an external incoming event of lineage 4 occurred in South China during A.D. 1150–1268 [12]. Further, Li et al. reported that the most recent common ancestor of the largest strain complex of lineage 4 strains (within the same evolutionary branch of the Neighbor-Joining tree) in Sichuan-Chongqing region appeared during A.D. 1069–1498 [16]. The most important human activities close to this period and possibly bringing lineage 4 from Europe to South China and making it spread in a particular direction, were the Maritime Silk Road and “Huguang Filling Sichuan” population migration. As regards to the Maritime Silk Road, which represents the first official international sea-trading route in Chinese history from the 1st century to the middle of 19th century [22], it constituted a key passage for foreign exchanges, trade and cultural communication between ancient China and Southeast Asia, the Indian subcontinent, Arabian peninsula, Somalia, Egypt and Europe [23]. Further, as an important part of ancient world maritime trade, the Maritime Silk Road has been proved that infectious diseases could spread across continents through it [24, 25]. “Huguang Filling Sichuan” population migration refers to the multiple immigration episodes in Chinese history from the 14th century to the end of the 18th century, in which immigrants from Huguang region (Supplementary Fig. 2) and other provinces of South China were forced by the government, or settled spontaneously, in Sichuan-Chongqing region [26, 27]; at least millions of migrants, including those from southeast coastal region (Supplementary Fig. 2), had made a purposeful move in South China during this event [26].

We hypothesized that the combined influence of the Maritime Silk Road and “Huguang Filling Sichuan” population migration might be important factors contributing to the singular lineage 4 distribution pattern of MTBC in South China. Since the distribution of surnames can supply quantitative information on the structure of human populations and migration rates [28, 29], we decided to explore available MTBC genotyping data from 32 provinces in China ($n = 11,171$) and 80 countries ($n = 14,404$), in conjunction with family names potentially representing migratory movements. The distribution of four representative

surnames was used to characterize the migration. In addition, the correlation between the geographical distribution of the main ports associated with the Maritime Silk Road and the distribution of the areas with high proportion of lineage 4 was analyzed likewise.

Methods

MTBC Genotyping Data

Spoligotyping (spacer oligonucleotide typing) results of 11171 MTBC clinical strains were obtained from 23 published literatures (Supplementary Table 1), covering 32 provinces in China (excluding Hong Kong and Macao). According to the relationship between spoligotyping and lineage classification results based on WGS data [30], strains were mapped to corresponding lineages. Simultaneously, lineage classification results of 14,404 MTBC strains from 80 countries were collected [including 6,098 results from published literatures and 8,306 results obtained by using TB-profiler (version: 2.3.0) to analyze the collected whole genome isolates (Supplementary Table 7)], of which 1,546 strains belonged to 3 sub-lineages of lineage 4 (L4.2, L4.4, and L4.5; Supplementary Table 2). Among these, strains collected from China were labeled to corresponding provinces (excluding Hong Kong, Macao and Taiwan; Supplementary Table 3).

Data on proportion of four surname populations (Zeng, Tang, Deng and Zhong) in each province

From public data from the Internet and published literature, the proportion of four representative surname populations (Zeng, Tang, Deng and Zhong) in major provinces of China was collected (excluding Hainan, Hong Kong, Macao, Tibet, and Taiwan; Supplementary Table 4).

Analysis of the Geographical Distribution of MTBC

The following work was mainly completed by using ArcMap (version 10.1). Firstly, the typing results of 11171 MTBC clinical isolates (Supplementary Table 1) was matched to the corresponding provinces in China; the relative proportion of MTBC lineages (lineage 1-4) in each province was shown by pie chart, and the proportion of lineage 4 in all isolates was expressed by the depth of map background color. Further, based on the typing results from most provinces of China (Supplementary Table 3), empirical Bayesian Kriging interpolation prediction was used to predict the variation of lineage 4's distribution with geographical location in various parts of China. The number of overlaps was 1 and the number of simulations was 100. Moreover, based on 14,404 lineage classification results from 80 countries (Supplementary Table 2), IDW (Inverse Distance Weighted) interpolation prediction was used to predict the distribution of lineage 4's three sub-lineages (L4.2, L4.4 and L4.5) in Asia, Europe, Africa and Oceania. In order to improve the accuracy of this prediction, the typing results from most provinces was used to replace the overall data of China.

In addition, based on the prediction of L4.2, L4.4 and L4.5's distribution in Asia, Europe, Africa and Oceania, we chose the coordinates of regions with high proportion of L4.2 and L4.4. And we obtained the coordinates of the main ports along the Maritime Silk Road. The information of main ports' location was

from the Old World Trade Routes Project (www.ciolek.com/owtrad.html), and all corresponding coordinates were from GCS: WGS 1984 (Global Coordinate System: World Geodetic System 1984). Origin was the point at coordinate (0,0) in GCS: WGS 1984. The relative distance from the coordinates of main ports and high proportion regions to the origin were calculated respectively (Supplementary Table 5, 6). Minitab (version 17.1.0) was used to analyze the correlation between the relative distance from the main ports to the origin and the relative distance from the high proportion regions to the origin. The former was regarded as independent variable and the latter dependent one to draw the scatter plot (the two sets of relative distance data were mapped one by one in order from large to small).

Analysis of Geographical Distribution of four Surname Populations (Zeng, Tang, Deng and Zhong)

Based on the collected number of 4 surname populations (Zeng, Tang, Deng and Zhong) in 30 provinces of China, ArcMap (version 10.1) was used to carry out IDW interpolation prediction to predict the 4 surname populations' distribution in China.

Results

MTBC lineages in China

Among MTBC clinical isolates collected from 32 provinces in China, lineage 2 largely predominated (81.3%, $n=9,085/11,171$), followed by lineage 4 which accounted for 15.9% of strains ($n=1,778/11,171$); these observations were consistent with the conclusions of previous studies [11-16]. Lineage 2 was widely distributed throughout China while lineage 4 was mostly distributed in few provinces of South China (Figure 1). Additionally, lineage 1 (1.9%, $n=211/11,171$) was limited to Taiwan, while lineage 3 (0.9%, $n=97/11,171$) was exclusively found in Xinjiang, Qinghai, Tibet and Fujian.

Distribution of lineage 4 in China

Lineage 4 had a significantly higher proportion in Sichuan-Chongqing region, southeast coastal region, Huguang region, Yunnan and Guizhou in South China than in other areas of Chinese mainland ($P < 0.01$) (Figure 1), accounting for 29.6% ($n=449/1,519$), 44.2% ($n=114/258$), 25.7% ($n=188/732$), 25.0% ($n=10/40$) and 30.2% ($n=39/129$). Except Xinjiang, which had a relatively high proportion of lineage 4 (20.2%, $n=23/114$), its proportion was relatively low in North China (9.0%, $n=490/5,445$; Supplementary Figure 1). Above all, the proportion of lineage 4 in Sichuan-Chongqing region (29.6%) was significantly higher than that in Jiangxi (16.9%), Zhejiang (11.9%), Anhui (10.1%) and Jiangsu (14.8%); even though the latter provinces were closer in spatial distance to the area with highest proportion of lineage 4 in China. Additionally, Bayesian kriging interpolation prediction showed that lineage 4 congregated in South China, with the highest proportion in southeast coastal region. The regions with high lineage 4 proportion extended directionally from southeast coastal region to Sichuan-Chongqing region (Figure 2b).

Distribution of four Surname Populations in China

The distribution of four surname populations (Zeng, Tang, Deng and Zhong) in China showed obvious regional characteristics (Figure 2c-f). Among them, the proportions of Zeng and Deng in Sichuan-Chongqing region, Huguang region and southeast coastal region were significantly greater than those in other regions of China ($p < 0.01$). The proportion of Tang population in Sichuan-Chongqing region and Huguang region was significantly higher than those in other regions of China ($p < 0.01$), while the proportion of Zhong population in Sichuan-Chongqing region and southeast coastal region was significantly larger than those in other regions of China ($p < 0.01$). These four surname populations were rarely distributed in North China. Interpolation prediction showed that the distribution of the population of these four surnames extended directionally from southeast coastal region and Huguang region to Sichuan-Chongqing region (Figure 2c-f), which simulated the historical route of “Huguang Filling Sichuan” population migration.

Distribution of three sub-lineages of lineage 4 in Asia, Africa, Europe and Oceania

L4.2 in Asia, Africa, Europe and Oceania was mainly distributed in eastern coastal region of Africa, the Mediterranean region of Europe and Chinese southeast coastal region, Sichuan-Chongqing region (Figure 3b), while L4.4 was mainly distributed in Western Europe, the Mediterranean region, Northern Europe, Southeast Asia and Chinese eastern coastal region, southeast coastal region, Sichuan-Chongqing region of China (Figure 3c). L4.5 was mainly distributed in China, Mongolia, Central Asia, Russia (Figure 3d). L4.2 and L4.4 tended to be distributed along the Maritime Silk Road in Asia, Africa, Europe and Oceania, including some areas of South China. The distribution of L4.5 did not show obvious correlation with the route of the Maritime Silk Road.

Moreover, there was a significant correlation between the relative distance from the main ports along the Maritime Silk Road to the origin (coordinate is (0,0)) and that from the regions with a high proportion of L4.2 and L4.4 to the origin ($r = 0.980$, $P < 0.01$). The scatter plot constructed by the two sets of relative distance showed significant aggregation (Moran's I index = 0.772861, z score = 7.479802, $P < 0.01$), and the slope of regression line was close to 1.00 (Figure 4). The results suggested that there was indeed a geographical correlation between main ports along the Maritime Silk Road and the regions with a high proportion of L4.2 and L4.4.

Discussion

The results obtained revealed that the proportion of MTBC lineage 4 population in some provinces of South China was significantly higher (> 25%) than in other provinces, while it was much lower in North China (9.0%). Although seasonal changes in TB incidence rate tentatively associated with temperature and humidity variations have been observed within a given area [31, 32], this may not be considered a determining factor for the particular distribution pattern of lineage 4 observed in our study. Indeed, lineage 4 of MTBC is widely distributed in the world, and high burden countries span different latitudes and climatic conditions [2, 33]. Additionally, despite modern era's increased human activities, transportation and expanding globalization, the geographical distribution of MTBC has not yet been

effusively affected; i.e., its population structure in different regions of the world has remained significantly different [6, 33], probably arising from longer duration of historical human activities that shaped its worldwide geographical distribution landscape [34, 35]. This could explain why lineage 4 comprises both globally distributed and geographically restricted sub-lineages [33]; an observation that led us presume that modern human activities could not be the focal elements affecting the observed distribution gradient of lineage 4 in China. We subsequently concentrated on how singular historical events may have contributed to the particular distribution of lineage 4 in China.

Previous studies suggested that the early transmission of lineage 4 was related to the “Northern Route” migration of East Asians about 15,000–18,000 years ago [36], and lineage 4 was introduced into China from Central Asia or Siberia, spreading from north to south [37]. However, the current proportion of lineage 4 in China being higher in the south than in the north, we looked into other factors that might have significantly influenced the lineage 4 population structure in China.

Bearing in mind higher genetic diversity of lineage 4 but not genetic differentiation among the circulating strains in Sichuan-Chongqing region [16], and despite the likelihood of successive events of lineage 4 strains arriving to this region historically, it seems probable that only limited incoming events have led to the pool of lineage 4 strains epidemic today. Besides, Liu et al. [12] found that during the period of A.D. 1150–1268, an important external incoming event of lineage 4 occurred in China; the most recent common ancestors of its three main sub-lineages (L4.2, L4.4 and L4.5) epidemic in South China, appearing around A.D. 1208, 1268 and 1160. Apparently, China's TB epidemic stems from historical expansion of a limited number of MTBC genotypes, whose deeply rooted phylogenetic positions and geographic restriction are indicative of their expansion in situ following a small number of introductions to China; a finding that explains why the number of MTBC (including lineage 4 strains) increased rapidly between A.D. 1300 and 1400 in China [12]. Therefore, one may presume that lineage 4 did originate in this period in South China, an event that constitutes the main factor affecting its current distribution in China today.

The potential spread of lineage 4 to South China through the Maritime Silk Road

It seemed crucial to study the transmission path of the lineage 4 spread to China during A.D. 1150–1268. Our results showed that the geographic distribution of L4.5 was continuous in Asia and Europe, which is consistent with the conclusion that lineage 4 might have its origin in South China [12, 33], and its spread might be related to the expansion of the Mongol Empire [38]. On the contrary, distribution of regions with high proportions of L4.2 and L4.4 in Asia, Africa, Europe and Oceania was intermittent, suggesting that their geographical distribution primarily did not relate to the expansion of Mongol Empire.

Besides, the possibility of L4.2 and L4.4's spread by land route could be basically ruled out, because if they were spread by land route, the countries along the route would have higher proportion of them, and the lineage 4 isolates of these countries would be ahead of the lineage 4 isolates of China in evolutionary status. However, the proportion of L4.2 and L4.4 was low in Central Asia, Siberia and other regions where the land route connecting Europe and China had to pass. Additionally, L4.2 and L4.4 isolates from some

countries in the Middle East and West Asia were nested in the samples from China on the evolutionary tree [12], signifying that L4.2 and L4.4 isolates in these areas were not earlier than Chinese samples in evolutionary status. Therefore, it seems likely that the lineage 4 was introduced in China earlier (around A.D. 1150–1268) from Europe through the sea route, and that L4.2 and L4.4 spread and diversified in China before being exported to the above areas in the Middle East and West Asia.

Around 13-14th century, the main maritime migration route was the Maritime Silk Road, and some infectious diseases did probably spread through it in ancient times, e.g., *Yersinia pestis*, which caused one of the most terrible plague episodes in European history, most likely arrived in Europe from China through the Maritime Silk Road [24, 25]. It is therefore feasible that MTBC lineage 4 reached the southeast coastal region of China from Europe through the Maritime Silk Road in a similar manner. In addition, 31 main ports associated with the Maritime Silk Road were geographically related to regions with a high proportion of L4.2 and L4.4, which suggests that the geographical distribution of L4.2 and L4.4 in Asia, Africa, Europe and Oceania may be related to the Maritime Silk Road. Remarkably, the time of introduction of lineage 4 (A.D. 1150–1268) falls within the Southern Song Dynasty of China (A.D. 1127–1279 [39]), when the foreign trade was highly developed. The Song dynasty mainly relied on ports such as Guangzhou and Quanzhou in southeast coastal region to conduct trade activities and population movements with Europe through the Maritime Silk Road [22, 40–42], and numerous arriving immigrants through this route settled in southeast coastal region of China [43, 44], thus providing suitable conditions for arrival and local spread of lineage 4.

The potential spread of lineage 4 in South China caused by “Huguang Filling Sichuan” population migration

Sichuan-Chongqing region is mainly located in Sichuan Basin, a lowland region in the west of South China surrounded by upland regions and mountains on all sides, with underdeveloped traffic and reduced population mobility in ancient times till A. D. 1950s. A difficult access certainly made spontaneous spread of MTBC to Sichuan-Chongqing region from outside territories rather difficult. Considering that the external inflow of lineage 4 occurred in the southeast coastal region, it is worth mentioning that in some provinces nearby that are more prone to easier population mobility, the proportion of lineage 4 was lower than that in Sichuan-Chongqing region. Most likely, a large number of lineage 4 carriers migrated from coastal areas to inland in a specific direction, and eventually reached Sichuan-Chongqing region.

Furthermore, the most recent common ancestor of the largest strain complex of lineage 4 collected by Li et al. in Sichuan-Chongqing region appeared during A.D. 1069–1498 [16], a time point close to the lineage 4 external incoming event discovered by Liu et al. [11]. Therefore, the arrival of the ancestor entering Sichuan-Chongqing region, was most probably triggered by exceptional population influxes. Since the largest population migration related to Sichuan-Chongqing region is “Huguang Filling Sichuan” population migration [27], and the distribution of surnames can supply quantitative information on the structure of human populations and migration rates [28, 29], we analyzed the distribution of four surname populations (Zeng, Tang, Deng and Zhong) in China (it was clearly recorded in the ancient literature that these four surnames populations migrated to Sichuan-Chongqing region in “Huguang

Filling Sichuan” population migration [26]). The result showed that these populations were very concentrated, and the key distribution areas of the four surname populations were on the “Huguang Filling Sichuan” immigration route (Fig. 2c-f). Interestingly, the four surname populations were mainly distributed in Sichuan-Chongqing region, Huguang region and southeast coastal region, which was basically consistent with the areas that had a high proportion of lineage 4 in South China. It supports a large-scale population migration in these areas, and suggests that the entry of the lineage 4 (that spread to South China in A.D. 1150–1268) into Sichuan-Chongqing region (in A.D. 1069–1498) might be related to “Huguang Filling Sichuan” population migration (Fig. 4), that covered a time span of 400 years [26]. Lineage 4 was then inexorably able to flow and spread ensuing its current distribution and high proportion in South China.

Conclusion

This study suggested that human activity on Maritime Silk Road and “Huguang Filling Sichuan” population migration might have contributed to the historic spread of the lineage 4 epidemic in today's South China. Therefore, our work may provide new evidence for the long-distance transmission of infectious diseases between the East and the West in ancient times, and show that cross-border and regional population migration in history may have a profound impact on the current distribution of TB, which may be also true to other infectious diseases around the world. Such studies may help us have a better understanding of the current TB landscape in China and facilitate the follow-up emerging MTBC clones. Further, this study may have a guiding significance for the global epidemiological investigations and traceability, suggesting that a global perspective in the epidemiological study of infectious diseases such as TB and COVID-19, etc. is necessary.

Abbreviations

MTB: *Mycobacterium tuberculosis*;

MTBC: *Mycobacterium tuberculosis* complex;

TB: tuberculosis;

WGS: whole-genome sequencing;

Spoligotyping: spacer oligonucleotide typing;

IDW: Inverse Distance Weighted;

GCS: WGS 1984: Global Coordinate System: World Geodetic System 1984

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Genotyping data of MTBC isolates, the proportion of targeted surname populations related to “Huguang Filling Sichuan”, and the information of main ports’ location around the Maritime Silk Road were listed in the supplementary files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Projects (2019YFE0103800); Sichuan Science and Technology Program (2019YFH0113, 2021YFH0060, 2021YFH0072, 2021ZHFP0045, 2021YFN0092). This funding source had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors’ contributions

Wei Wu, Peng-Kuan Liang and Dong-Xia Luo contributed to the work equally and should be regarded as co-first authors. Wei Wu, Peng-Kuan Liang and Qun Sun conceived the study; Wei Wu, Peng-Kuan Liang, Dong-Xia Luo, Yu-Ang Tian, Hai-Bing Yuan, Mei Luo, Zhong-Hua Jiang and Zhuo-Chong Liu collected the data; Wei Wu, Peng-Kuan Liang, Dong-Xia Luo, Zhong-Hua Jiang and Zhuo-Chong Liu analyzed the data. Wei Wu, Peng-Kuan Liang and Yu-Ang Tian wrote the manuscript; Dong-Xia Luo, Nalin Rastogi and Qun Sun reviewed and provided critical comments; Hai-Bing Yuan, Nalin Rastogi and Qun Sun read and approved the final manuscript.

Acknowledgements

We thank Dr. Ming-Fei Li from the Institute of Ancient History, Chinese Academy of Social Sciences, and Dr. Xiao-Shan Chen from the Institute for the History of Natural Sciences, Chinese Academy of Sciences, for their highly constructive suggestions. We also thank National Cryosphere Desert Data Center (<http://www.ncdc.ac.cn>) for providing data set of 1 km resolution DEM in China.

References

1. Brosch R, Gordon SV, Marmiesse M, Brodin P, Buchrieser C, Eiglmeier K, et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proceedings of the National Academy of

- Sciences of the United States of America. 2002;99(6):3684-9.
2. WHO. Global Tuberculosis Report 2020. World Health Organization; 2020.
 3. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Katomaeda M, et al. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nature Genetics*. 2013;45(10):1176-82.
 4. De Jong BC, Antonio M, Gagneux S. *Mycobacterium africanum*—Review of an Important Cause of Human Tuberculosis in West Africa. *PLOS Neglected Tropical Diseases*. 2010;4(9):e744.
 5. Gagneux S, Deriemer K, Van T, Kato-Maeda M, De Jong B, Narayanan S, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(8):2869-73.
 6. Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infectious Diseases*. 2007;7(5):328-37.
 7. Hirsh A, Tsolaki A, DeRiemer K, Feldman M, Small P. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proceedings of the National Academy of Sciences of the United States of America*. 2004;101:4871-6.
 8. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*. 2018;16(4):202-13.
 9. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. *Nature Communications*. 2020;11(1):2917.
 10. Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. *Microb Genom*. 2021;7(2).
 11. Lu W, Lu B, Liu Q, Dong H, Shao Y, Jiang Y, et al. Genotypes of *Mycobacterium tuberculosis* isolates in rural China: using MIRU-VNTR and spoligotyping methods. *Scand J Infect Dis*. 2014;46(2):98-106.
 12. Liu Q, Ma A, Wei L, Pang Y, Wu B, Luo T, et al. China's tuberculosis epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. *Nature Ecology & Evolution*. 2018;2(12):1982-92.
 13. Chen H, He L, Cai C, Liu J, Jia J, Ma L, et al. Characteristics of distribution of *Mycobacterium tuberculosis* lineages in China. *Science China-life Sciences*. 2018;61(6):651-9.
 14. Pang Y, Zhou Y, Zhao B, Liu G, Jiang G, Xia H, et al. Spoligotyping and Drug Resistance Analysis of *Mycobacterium tuberculosis* Strains from National Survey in China. *PLOS ONE*. 2012;7(3):e32976.
 15. Zhao Y, Feng Q, Tang K, Zhang C, Sun H, Luo T, et al. The population structure of drug-resistant *Mycobacterium tuberculosis* clinical isolates from Sichuan in China. *Infection, Genetics and Evolution*. 2012;12(4):718-24.
 16. Li Y, Fu Y, Yuan M, Dong L, Huang H, Li W, et al. Study on the population-genetics of *Mycobacterium tuberculosis* from Sichuan Basin in China. *Zhonghua liu xing bing xue za zhi = Zhonghua*

- liuxingbingxue zazhi. 2015;36:374-8.
17. Barbier M, Wirth T. The Evolutionary History, Demography, and Spread of the *Mycobacterium tuberculosis* Complex. *Microbiology Spectrum*. 2016;4:453-73.
 18. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet*. 2015;47(3):242-9.
 19. Wirth T, Hildebrand F, Allixbeguec C, Wolbeling F, Kubica T, Kremer K, et al. Origin, Spread and Demography of the *Mycobacterium tuberculosis* Complex. *PLOS Pathogens*. 2008;4(9):e1000160.
 20. Wang Y. Chronology of epidemic situation in ancient China (2) (674 BC-1911 AD) [in Chinese]. *Journal of Tianjin University of traditional Chinese Medicine*. 2003(4):33-6.
 21. Zhang Z. Epidemic chronology of epidemic diseases in ancient China [in Chinese]: Fujian science and Technology Press; 2007.
 22. Shangji S, Guiling XU. Historical Evolution of China's Maritime Silk Road. *Tropical Geography*. 2015.
 23. Qin D, Yuan J. Ancient Silk Trade Routes: Selected Works from Symposium on Cross Cultural Exchanges and Their Legacies in Asia: WORLD SCIENTIFIC; 2015.
 24. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nature Genetics*. 2010;42(12):1140-3.
 25. Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, et al. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(2):577-82.
 26. Li Y. An investigation of immigrants from Huguang to Sichuan from the end of Yuan Dynasty to the end of Ming Dynasty [in Chinese]. *A collection of studies on local culture*. 2017;000(1):193-211.
 27. Zhang M, Ke L, Sun S. The migration and influence of "Huguang filling Sichuan" in the late Ming and early Qing Dynasties [in Chinese]. *Changshu Institute of Technology*. 2008;22(5):80-2.
 28. Piazza A, Rendine S, Zei G, Moroni A, Cavallisforza LL. Migration rates of human populations from surname distributions. *Nature*. 1987;329(6141):714-6.
 29. Zei G, Guglielmino C, Siri E, Moroni A, Cavalli-Sforza L. Surnames as neutral alleles: Observations in Sardinia. *Human biology*. 1983;55:357-65.
 30. Barbier M, Wirth T. The Evolutionary History, Demography, and Spread of the *Mycobacterium tuberculosis* Complex. *Microbiology Spectrum*. 2016;4:TBTB2-0008.
 31. Naranbat N, Nymadawa P, Schopfer K, Rieder HL. Seasonality of Tuberculosis in an Eastern–Asian Country With an Extreme Continental Climate. *The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology*. 2009;34:921-5.
 32. Willis M, Winston C, Heilig C, Cain K, Walter N, Kenzie W. Seasonality Of Tuberculosis In The United States, 1993-2008. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012;54:1553-60.
 33. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, et al. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature Genetics*.

- 2016;48(12):1535-43.
34. Pepperell CS, Casto AM, Kitchen A, Granka JM, Feldman MW. The Role of Selection in Shaping Diversity of Natural *M. tuberculosis* Populations. *PLOS Pathogens*,9,8(2013-8-15). 2013;9(8):e1003543.
 35. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature*. 2007;445(7130):915-8.
 36. Zhong H, Shi H, Qi X-B, Duan Z-Y, Tan P-P, Jin L, et al. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Molecular biology and evolution*. 2011;28(1):717-27.
 37. Lin N. Genome analysis of *Mycobacterium tuberculosis* in China [in Chinese]: Fujian agriculture and Forestry University; 2014.
 38. Mokrousov I, Shitikov E, Skiba Y, Kolchenko S, Chernyaeva E, Vyazovaya A. Emerging peak on the phylogeographic landscape of *Mycobacterium tuberculosis* in West Asia: Definitely smoke, likely fire. *Molecular Phylogenetics and Evolution*. 2017;116:202-12.
 39. Wang Y, Zhu T, Yang G, Tan X, Ye D, Chen H. The method to soften the concretions of ceramics in the "Nanhai I" Shipwreck of China Southern Song Dynasty (1127–1279AD). *Heritage Ence*. 2018;6(1):4.
 40. Li R, Wang Q, Cheong K. Quanzhou: Reclaiming a glorious past. *Cities*. 2016;50:168-79.
 41. Guiling X, Shangji S. The Development of the Maritime Silk Road in Guangzhou and the Construction of World Historical Cities. *Urban Insight*. 2011.
 42. Lam PYK. A ceramic legacy of Asia's maritime trade : Song dynasty Guangdong wares and other 11th to 19th century trade ceramics found on Tioman Island, Malaysia. *Journal of Fluid Mechanics*. 2013;393(393):11.
 43. Wang GP, Yong MA. Islamism in Guangzhou at the time of Tang and Song. *Journal of Guangzhou University*. 2006.
 44. Zhou X. A Study on the Origins of Guangzhou Maritime Silk Road and its Development Process. *China Ancient City*. 2018.

Figures

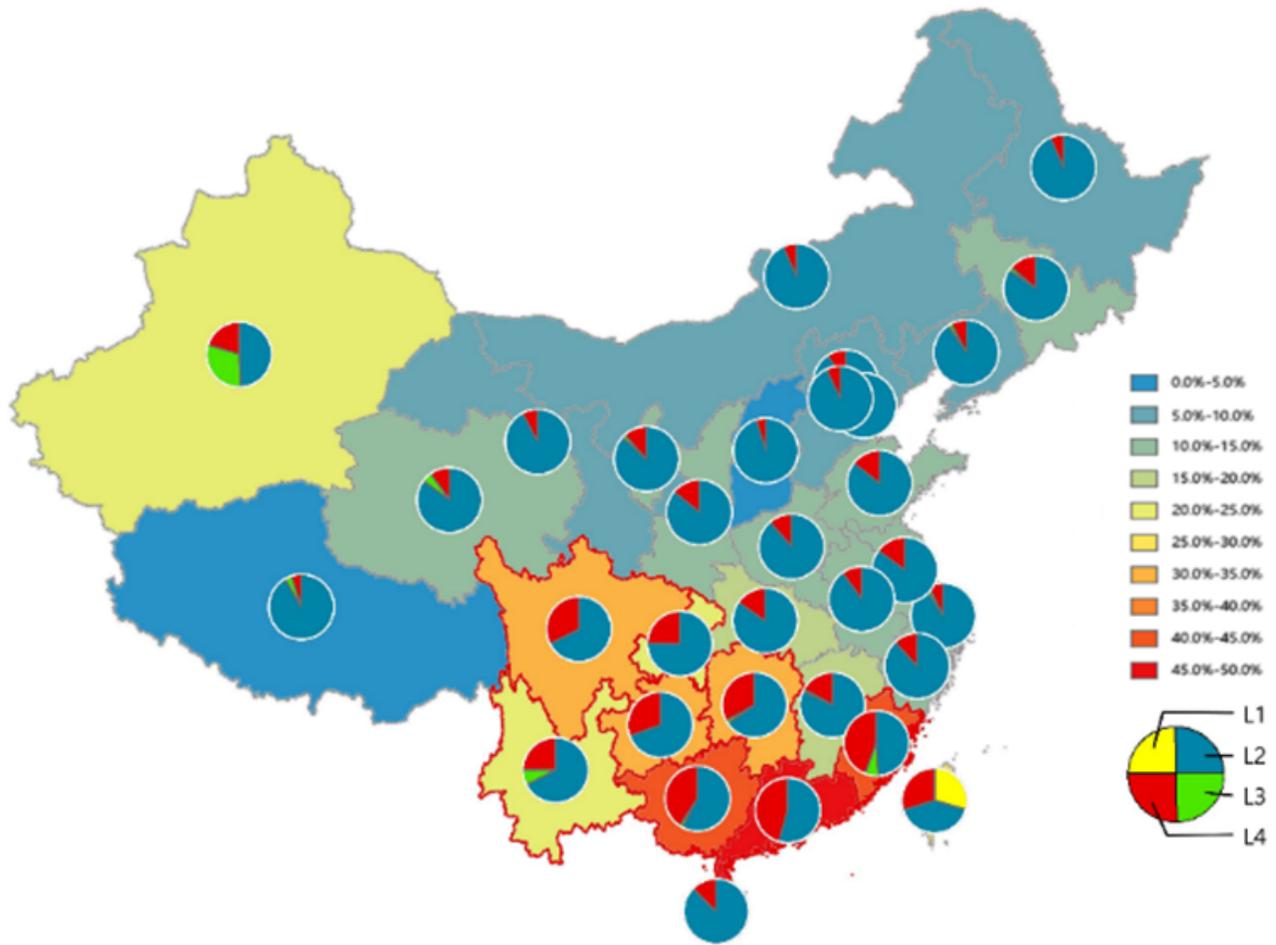


Figure 1

Distribution of four MTBC lineages in China. Each pie chart represents the proportion of four lineages of MTBC in the corresponding province, and the different colors of the province represent the proportion of lineage 4 to the total number of MTBC. The provinces with high lineage 4 proportion and concentrated distribution are circled in red. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

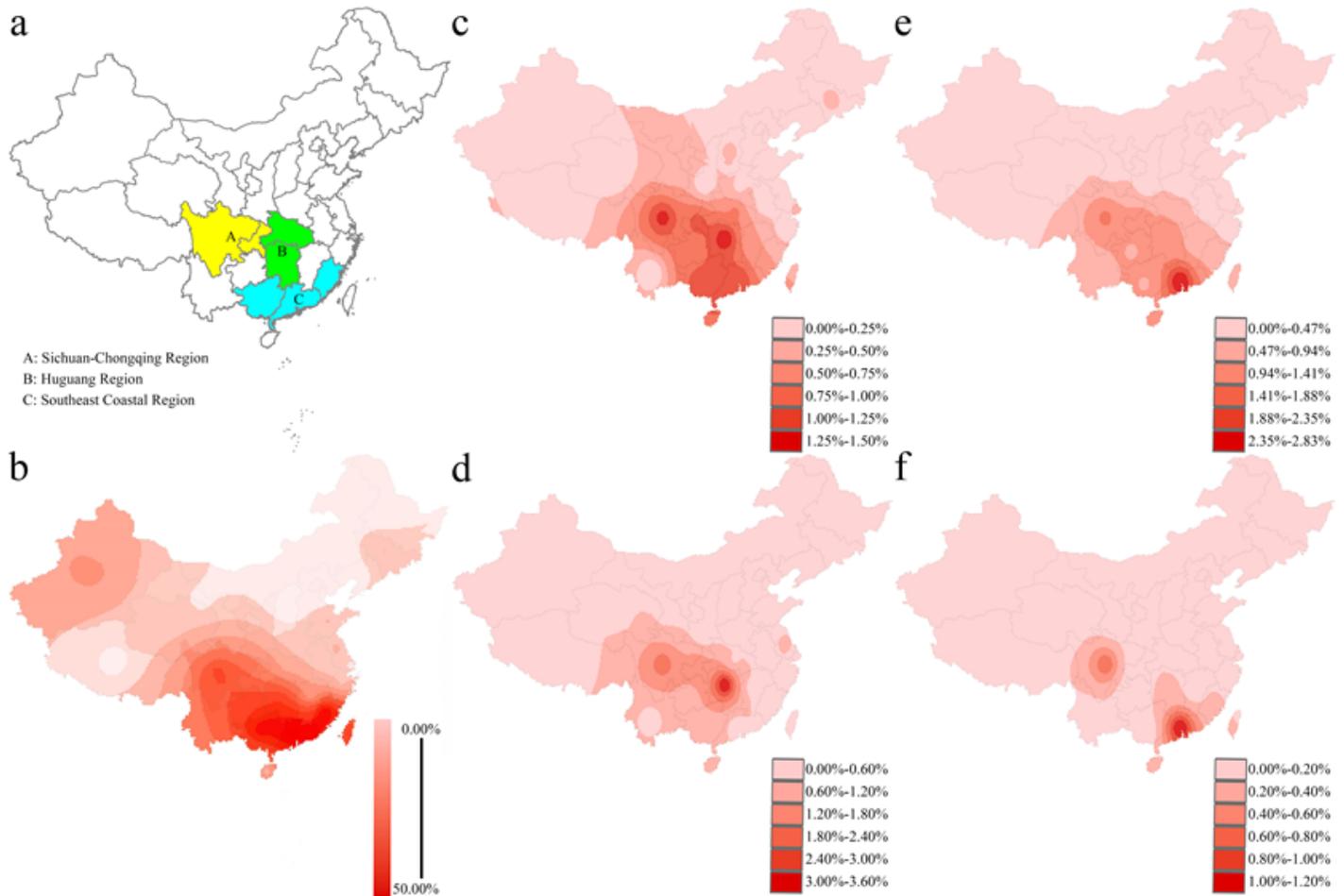


Figure 2

Distribution of lineage 4 and four surname populations in China. a, Partial Chinese regions mentioned in this article. “A” stands for Sichuan-Chongqing region, including provinces of Sichuan and Chongqing; “B” stands for Huguang region, including provinces of Hunan and Hubei; “C” stands for southeast coastal region, including provinces of Guangxi, Guangdong, and Fujian. Note that Sichuan-Chongqing, and Huguang regions were originally administrative regions in history. This map uses modern administrative divisions to show the main boundaries of these two regions, but they are still a little different from the real boundaries in history. b, Bayes Kriging interpolation analysis chart, of the proportion of lineage 4 in all strains in each province changing with the region. c-f, IDW interpolation analysis of the four surname populations’ distribution in China. Color ranges showing the percentage prevalence of (c) Zeng, (d) Tang, (e) Deng, (f) Zhong. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

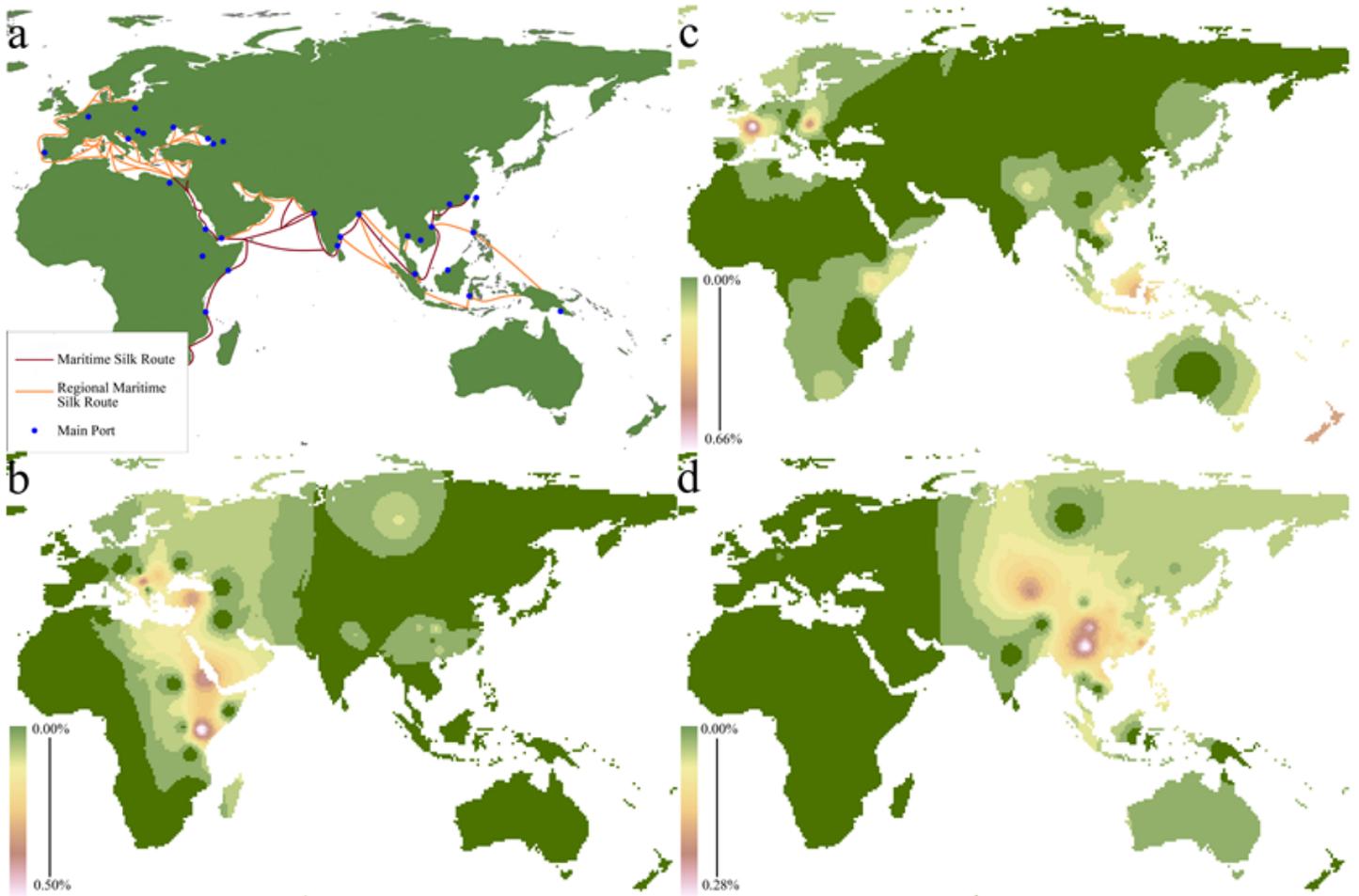


Figure 3

Distribution of Maritime Silk Road's ports and L4.2, L4.4, L4.5. a, Schematic diagram of Maritime Silk Road's route and its related main ports. b-d, The distribution of (b) L4.2, (c) L4.4, (d) L4.5 in Asia, Africa, Europe and Oceania. The proportion in the figure refers to the proportion of (b) L4.2, (c) L4.4, (d) L4.5 to the total number of local MTBC strains. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

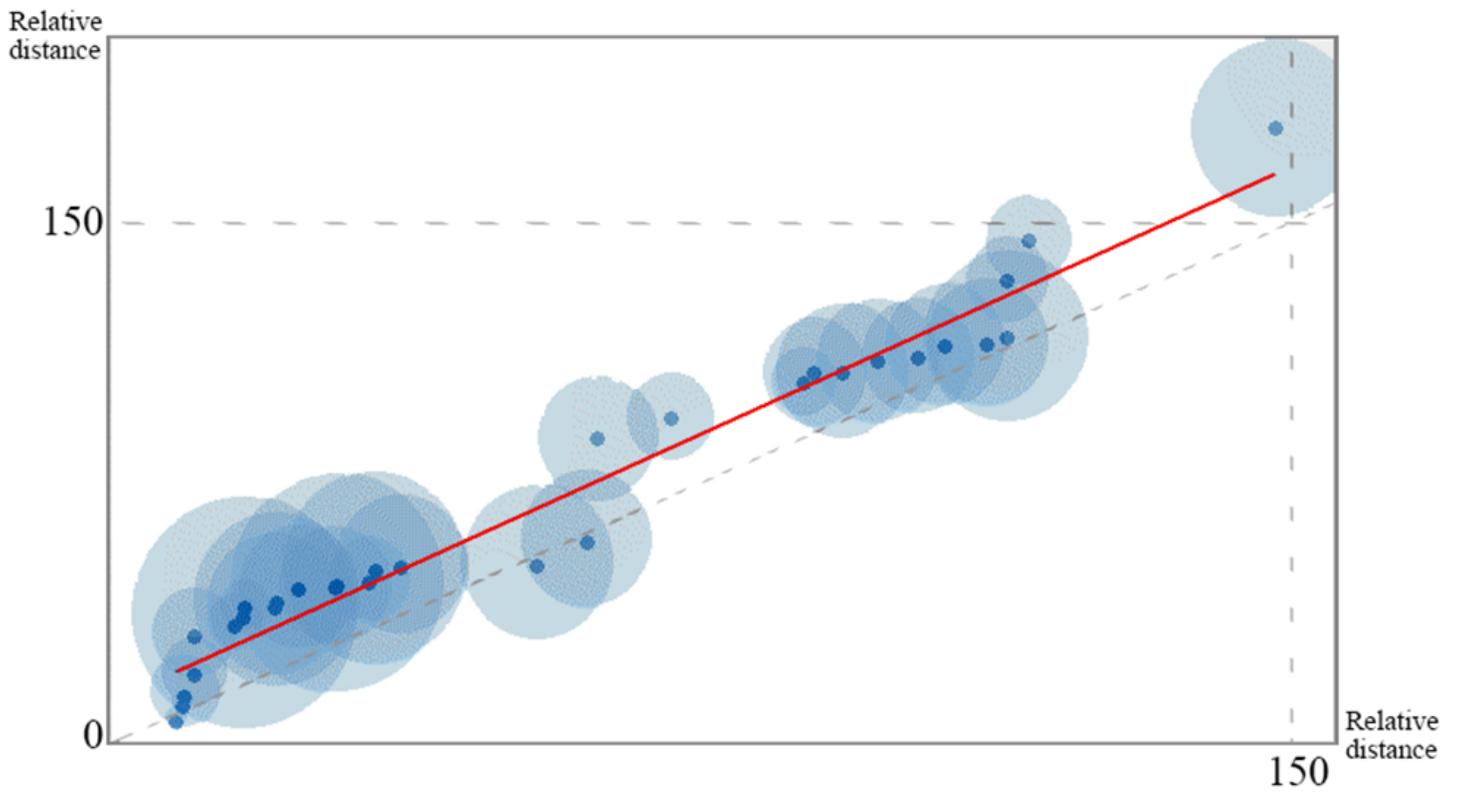


Figure 4

Scatter diagram of the relative distance. Scatter diagram of the distance from the origin to the main ports of Maritime Silk Road (X coordinate) and L4.2 or L4.4 high-scale areas (Y coordinate). Red line is the regression line, and the bubble size represents the average of L4.2 and L4.4 proportion.

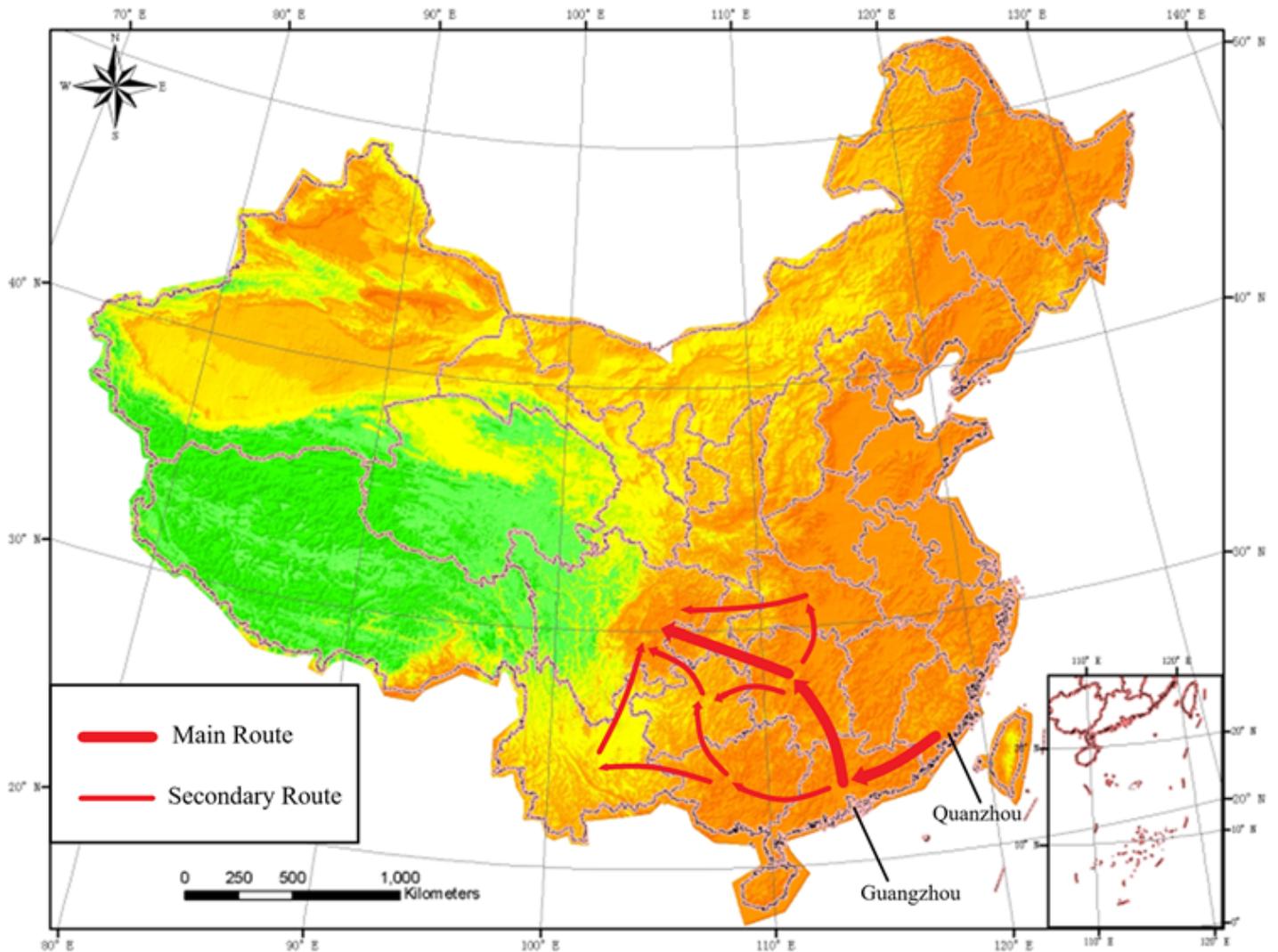


Figure 5

The proposed transmission route of lineage 4 in South China. The thicker red arrow represents the main route, and the thinner one for the secondary route. The background map represents the digital elevation model of China, and it was provided by National Cryosphere Desert Data Center (<http://www.ncdc.ac.cn>). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [MHTBIDPOsupportfile.doc](#)