

Spread of Mycobacterium Tuberculosis Lineage 4 in Sichuan-chongqing Region of China Influenced by Maritime Silk Road and “huguang Filling Sichuan”

Wei Wu

Key Laboratory of Bio-resource and Eco-environment of the Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan, P. R. China

Peng-kuan Liang

Key Laboratory of Bio-resources and Eco-environment of the Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan, P. R. China

Hai-Bing Yuan

School of Archaeology and Museology & National Demonstration Center for Experimental Archaeology Education, Sichuan University, Chengdu, Sichuan, P. R. China

Yu-Ang Tian

College of Life Sciences, Sichuan University, Chengdu, Sichuan, P. R. China

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, human migration, Huguang filling Sichuan, Maritime Silk Road, interpolation analysis

Posted Date: December 1st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-114225/v1>

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

[Read Full License](#)

Abstract

Background

The spread of *Mycobacterium tuberculosis* complex (MTBC) was closely influenced by human migration. Lineage 4, a lineage of MTBC mainly being epidemic in Europe and America, was not the major one in China. Its proportion in Sichuan-Chongqing region, however, was significantly higher than that at the national level. Previous study by the most recent common ancestor calculation based on Bayes model suggested that lineage 4 epidemic in South China today probably entered China around the 13th century, and the most important human migration events associating with this region during and after this period of time included Maritime Silk Road and “Huguang filling Sichuan”. The purpose of this study was to explore whether the high proportion of lineage 4 in Sichuan-Chongqing region was impacted by these two events.

Methods

The genotyping data of 17,269 MTBC isolates, the proportion of targeted surname populations related to “Huguang filling Sichuan”, and the information of main ports’ location around Maritime Silk Road were collected. Interpolation analysis was used to estimate the geographical distribution of lineage 4 and its three sub-lineages (L4.2, L4.4, L4.5), as well as the targeted surname populations. And the correlation between the geographical distribution in the main ports on Maritime Silk Road and the regions with the proportion of L4.2 and L4.4 was analyzed.

Results

Interpolation and correlation analysis showed that lineage 4’s distribution in China could be mapped to the regions affected by “Huguang filling Sichuan”, while L4.2 and L4.4’s distributing in Asia, Europe, and Africa could be those by Maritime Silk Road,

Conclusions

Maritime Silk Road and “Huguang Filling Sichuan” potentially posed an essential combined influence on the greater occurring of lineage 4 in Sichuan-Chongqing region. This study may contribute to the tracing and epidemiological investigation of tuberculosis and other world-wide infectious diseases.

Background

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 into TB cases [2]. And there were about 9.96 million new cases and about 1.21 million HIV negative deaths of TB worldwide in 2019; among them, there were about 833,000 new cases of TB in China, ranking third in the world, and about 31,000 HIV negative people died of TB [2]. According to the phylogenetic analysis based on whole-genome sequencing (WGS), MTBC was divided

into 7 lineages: Indo-Oceanic lineage (lineage 1), East Asian lineage (lineage 2), East African–Indian lineage (lineage 3), Euro-American lineage (lineage 4), West African 1 lineage (lineage 5), West African 2 lineage (lineage 6) and Ethiopia lineage (lineage 7) [3–6]. Previous study suggested that different lineages of MTBC had obvious regional distribution characteristics or were highly correlated with specific infected people; and lineage 4 was mainly epidemic in Europe and America and might be associated with European and American people [7, 8].

In China, more than 80% of MTBC strains belonged to lineage 2 [9], and lineage 4 accounted for about 15% [10]. In Sichuan-Chongqing region (Fig. 2a), including provinces of Sichuan and Chongqing in China, lineage 2 accounted for about only 60%–70% of all strains in MTBC, which was lower than the national level [11–13] while lineage 4 accounted for about 25%–35%, significantly higher than that of the whole country (Supplementary Fig. 1) [10, 13, 14]. Because the expansion and evolution of MTBC were mainly caused by human activities, such as wars, adventures, trade and immigration [10, 15–17], historical human activities might be an important cause of the special distribution of lineage 4 in Sichuan-Chongqing region. However, China's historical documents did not contain a description of any TB pandemic in South China (Supplementary Fig. 1) [18, 19], so it is not clear what historical events caused this special distribution.

Liu et al. suggested that an external incoming event of lineage 4 occurred in South China during A.D. 1150–1268 [10]. And Li et al. found that the most recent common ancestor of the largest strain complex (in the same evolutionary branch of the Neighbor-Joining tree) of lineage 4 strains in Sichuan-Chongqing region appeared during A.D. 1069–1498 [14]. The most important human activities close to this period and closely related to Sichuan-Chongqing region were “Huguang filling Sichuan” and Maritime Silk Road. “Huguang Filling Sichuan” mainly refers to many immigration incidents in Chinese history from the 14th century to the end of the 18th century, in which immigrants from Huguang region (Fig. 2a) spontaneously or were forced by the government to settle in Sichuan-Chongqing region [20, 21]. It took more than 400 years, and the number of immigrants in each stage can reach several hundred thousand to one or two million [22]. Maritime Silk Road was the maritime passage for foreign exchanges, trade and cultural exchanges in ancient China, which lasted from the 1st century to the middle and late 19th century [23]. This maritime trade line was the main channel of communication between ancient China and Arabia, Africa, and Europe. The combined influence of “Huguang Filling Sichuan” and Maritime Silk Road might be an important factor contributing to the special distribution of MTBC in Sichuan-Chongqing region.

In order to explore whether these two historical events were important reasons for the formation of special lineage 4 distribution in Sichuan-Chongqing region, genotyping data of MTBC isolates from 32 provinces in China ($n = 11,171$) and 74 countries ($n = 6,098$) were collected and analyzed. Moreover, because the geographical distribution of the surname populations could be related to human migration [24, 25], the distribution of four surnames was used to represent the migration caused by “Huguang Filling Sichuan” to analyze the potential relationship between this event and the distribution of lineage 4. In addition, the correlation between the geographical distribution of the main ports associated with Maritime Silk Road and the distribution of the areas with high proportion of lineage 4 was also analyzed.

Methods

Collection of 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 (no suitable data were collected from provinces of Hong Kong and Macao). According to the relationship between spoligotyping and lineage classifications results based on WGS data [26], each strain was mapped to lineage 1-4.

At the same time, lineage classification results of 6098 MTBC isolates from 74 countries were collected, of which 805 strains belonging to 3 sub-lineages of lineage 4 (L4.2, L4.4, L4.5) (Supplementary Table 2). Among them, we further distributed the collected strains from China to corresponding provinces (no suitable data were collected from provinces of Hong Kong, Macao and Taiwan) (Supplementary Table 3).

Collection of the 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 4 surname populations (Zeng, Tang, Deng and Zhong) in major provinces of China was collected (Supplementary Table 4). No suitable data have been collected from provinces of Hainan, Taiwan, Tibet, Hong Kong and Macao.

Analysis of the Geographical Distribution of MTBC

Using ArcMap (version 10.1) to match the valid typing results of 11171 MTBC clinical isolates collected from 32 provinces (no suitable data were collected from provinces of Hong Kong and Macao) to the corresponding provinces in China. The relative proportion of MTBC lineages 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.

Using ArcMap (version 10.1) for empirical Bayesian Kriging interpolation prediction, the number of overlaps was 1 and the number of simulations was 100. Based on the valid typing results of 11171 MTBC clinical isolates collected, the variation of population proportion of lineage 4 with geographical location in various parts of China was predicted, and the distribution of lineage 4 in various parts of China was deduced.

Using ArcMap (version 10.1) for IDW (Inverse Distance Weighted) interpolation prediction; number of points:12. Based on the statistical data of L4.2, L4.4 and L4.5 collected and identified from 74 countries, the proportion of these three sub-lineages in the total number of local MTBC strains was calculated. In order to improve the accuracy of the results, the proportion of three sub-lineages of China's major provinces in MTBC strains was used to replace the overall data of China.

Moreover, using ArcMap (version 10.1) to obtain the coordinates of the main ports along Maritime Silk Road and the coordinates of regions with high proportion of L4.2 and L4.4. The information of main ports' location were from the Old World Trade Routes Project (www.ciolek.com/owtrad.html), and

corresponding coordinate data was 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 as dependent variable to draw 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 4 Surname Populations (Zeng, Tang, Deng and Zhong)

Using ArcMap (version 10.1) to carry out IDW interpolation analysis, based on the collected number of 4 surname populations (Zeng, Tang, Deng and Zhong) in 30 provinces of China. The proportion of the 4 surname populations in the local total population in different parts of China was predicted respectively with the change of geographical location, thus inferring the distribution of the 4 surname populations in different parts of China.

Results

MTBC lineages in China

Among MTBC clinical isolates collected from 32 provinces in China, lineage 2 occupied the great majority (81.3%, 9,085/11,171), followed by lineage 4 which accounted for 15.9% (1,778/11,171), which were consistent with the conclusions of previous studies [9-14]. Lineage 2 was widely distributed throughout China while lineage 4 was mainly distributed in few provinces of South China (Figure 1). In addition, in China, lineage 1 (1.9%, 211/11,171) was only distributed in Taiwan, and lineage 3 (0.9%, 97/11,171) was only distributed 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% (449/1,519), 44.2% (114/258), 25.7% (188/732), 25.0% (10/40) and 30.2% (39/129). Except Xinjiang, which had a relatively high proportion of lineage 4 (20.2%, 23/114), the whole proportion of lineage 4 in North China (Supplementary Figure 1) was relatively low (9.0%, 490/5,445). Particularly, 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%) while the latter provinces were closer in spatial distance to the area with highest proportion of lineage 4 in China. Moreover, Bayesian kriging interpolation analysis showed that lineage 4 was gathering in South China, and had 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 larger than those in other regions of China ($p < 0.01$). The proportion of Tang population in Sichuan-Chongqing region and Huguang region was significantly larger than those in other regions of China ($p < 0.01$). And 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 estimates 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 2b), which simulated the historical route of “Huguang filling Sichuan”.

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

L4.2 in Asia, Africa and Europe was mainly distributed in eastern coastal region of Africa, the Mediterranean region of Europe and Chinese southeast coastal region, Sichuan-Chongqing region, eastern coastal region (Figure 3b). 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 Maritime Silk Road in Asia, Europe, and Africa, including Sichuan-Chongqing region. The distribution of L4.5 did not show obvious correlation with the route of Maritime Silk Road.

Moreover, there was a significant correlation between the relative distance from the main ports along 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).

Discussion

The proportion of lineage 4 population in Sichuan-Chongqing region was high (29.6%). On the contrary, the proportion of lineage 4 population in North China was low (9.0%). Some studies suggested that temperature and humidity could affect the incidence rate of TB and make it seasonal change in the same area [27, 28]. However, lineage 4 of MTBC was widely distributed in the world, and high burden countries spanned different latitudes and different climatic regions [2, 29]. Therefore, temperature and humidity may not be the important factors leading to significant differences in distribution of lineage 4 in China. In addition, although human activities brought about by the convenience of transportation and the deepening of globalization in modern times may bring interference to the geographical distribution of MTBC, its distribution in different countries and continents was still significant different [6, 29]. Therefore, modern human activities may also not be the main factors affecting the distribution of lineage 4. As

large-scale historical human activities can affect the spread of MTBC [30, 31], special 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 route of East Asians (about 15,000 to 18,000 years ago [32]), and lineage 4 was introduced into China from Central Asia or Siberia and spread from north to south [33]. If China's lineage 4 distribution pattern today was mainly derived from this way, the proportion of lineage 4 in North China should be higher than that in Sichuan-Chongqing region and other areas of South China. However, the actual proportion of lineage 4 in China was higher in the south than in the north. This shows that there were other factors have significantly influenced the formation of lineage 4 population in China. Moreover, in Sichuan-Chongqing region, the genetic diversity of lineage 4 was high but the genetic differentiation among the main strains within this region was not significant [14]. This seems to imply that there have been many times in history when lineage 4 was spread to Sichuan-Chongqing region, but the lineage 4 mainly epidemic in this region today could be traced back to limited incoming events. In addition, Liu et al. 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 the three main sub-lineages of lineage 4 (L4.2, L4.4 and L4.5) epidemic in South China, appeared in about A.D. 1208, 1268 and 1160; the number of MTBC in China might have increased rapidly between A.D. 1300 and 1400, and the proportion of lineage 4 also increased greatly [10]. Therefore, the lineage 4 that originated from this period and appeared in South China, may be the main factor affecting the distribution of lineage 4 in China today.

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

It is necessary to study the transmission path of the lineage 4 spread to China during A.D. 1150–1268. It was seen that the geographic distribution of L4.5, as a China's endemic species, which might originate from South China [10, 29], was continuous. This situation may be related to the expansion of the Mongol Empire [34]. On the contrary, distribution of regions with high proportions of L4.2 and L4.4 in Asia, Africa and Europe was discontinuous. This suggested that L4.2 and L4.4's geographical distribution might not primarily be caused by the Mongol Empire's expansion. In addition, 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. And 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 [10], which means that L4.2 and L4.4 isolates in these areas were not earlier than Chinese samples in evolutionary status, and that L4.2 and L4.4 were spread and diversified in China before being exported to these areas. Therefore, it is more likely that the lineage 4 introduced in China during A.D. 1150–1268 went from Europe to China through the sea route.

In about the 13-14th century, the main maritime migration route was Maritime Silk Road. Previous studies have shown that the pathogens of infectious diseases could be spread through Maritime Silk Road in

ancient times. For example, *Yersinia pestis*, which caused the most terrible plague in European history, was likely to arrive in Europe from China through Maritime Silk Road [35, 36]. Then it is feasible for lineage 4 to reach southeast coastal region of China from Europe through Maritime Silk Road. In addition, 31 main ports associated with Maritime Silk Road were geographically related to regions with a high proportion of L4.2 and L4.4, which indicates that the geographical distribution of L4.2 and L4.4 in Asia, Africa and Europe may be related to Maritime Silk Road. Moreover, the time of the external introduction of lineage 4 (A.D. 1150–1268) is in the Southern Song Dynasty of China (A.D. 1127–1279 [37]), when the foreign trade was highly developed. This dynasty mainly relied on ports such as Guangzhou and Quanzhou in southeast coastal region to conduct trade activities and population movements with Europe through Maritime Silk Road [23, 38–40]. At that time, some foreigners who arrived through Maritime Silk Road lived in southeast coastal region [41, 42]. This also provided suitable conditions for the incoming and local spread of lineage 4.

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

Sichuan-Chongqing region is mainly located in Sichuan Basin, where the geographical conditions were closed, the traffic was underdeveloped, and the population mobility was difficult in ancient times. It was not easy for spontaneous people-to-people MTBC to enter Sichuan-Chongqing region. However, the proportion of lineage 4 in some provinces that are closer in geographical distance to the possible first place where the external inflow of lineage 4 occurred (southeast coastal region) and that are more convenient in population mobility was lower than that in Sichuan-Chongqing region. Moreover, 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 [14]. This time point was close to the lineage 4 external incoming event discovered by Liu et al. [9]. Therefore, in Sichuan-Chongqing region, the appearance of the lineage 4 entering China in A.D. 1150–1268, was possibly not caused by simple population diffusion. It may be that there was a special population flow which led to lineage 4 entering Sichuan-Chongqing region. In history, the migration event that occurred in Sichuan-Chongqing region at the right time was “Huguang Filling Sichuan”.

There were records of families with the same surname migrated from Huguang region to Sichuan-Chongqing region in “Huguang Filling Sichuan” [22, 43]. And the geographical distribution of the surname populations might represent the population migration [24, 25]. Therefore, the distribution of 4 surname populations (Zeng, Tang, Deng and Zhong) that in the historical data migrated from Huguang region to Sichuan-Chongqing region in the period of “Huguang Filling Sichuan” [43], could represent the population migration in “Huguang filling Sichuan”. It is found that the 4 surname populations were mainly distributed in Sichuan-Chongqing region, Huguang region and southeast coastal region, which is basically consistent with the areas that had a high proportion of lineage 4 in China. And the key distribution areas of the 4 surname populations were all on the “Huguang Filling Sichuan” immigration route. Moreover, there was a significant correlation between the proportion of surname populations of Zeng, Tang, Deng, and Zhong and that of lineage 4 in each province included in the statistics ($p < 0.05$). It proves that a large-scale population migration might have taken place among these areas, and that the entry of the lineage 4 that

spread to China during A.D. 1150–1268 into Sichuan-Chongqing region may be related to “Huguang Filling Sichuan” (Fig. 4). Moreover, for more than 400 years of “Huguang Filling Sichuan” [22], lineage 4 was able to continue to flow and spread in the above areas, resulting in the current distribution and high proportion of lineage 4 in South China, including Sichuan-Chongqing region.

Conclusion

The lineage 4 that was mainly epidemic in Sichuan-Chongqing region, Huguang region and southeast coastal region at present, might first enter the southeast coastal region through Maritime Silk Road in the 12th-13th century, then spread to the neighboring Huguang region, and finally enter Sichuan-Chongqing region through the large population migration of “Huguang Filling Sichuan”. Maritime Silk Road and “Huguang Filling Sichuan” were likely to make great contributions to the historical dissemination of the lineage 4 epidemic in Sichuan-Chongqing region today. Therefore, this study may provide new evidence for the long-distance transmission of infectious diseases between the East and the West in ancient times, and shows that the cross-border and regional population migration in history may have a profound impact on the current distribution of TB and other world-wide infectious diseases. It may contribute to the tracing and epidemiological investigation of TB and other world-wide infectious diseases.

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 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 National Key Research and Development Projects of China (SQ2019YFE010495). 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 and Peng-Kuan Liang 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 and Hai-Bing Yuan collected the data; Wei Wu and Peng-Kuan Liang analyzed the data. Wei Wu, Peng-Kuan Liang and Yu-Ang Tian wrote the manuscript; Hai-Bing Yuan and Qun Sun read and approved the manuscript.

Acknowledgements

We thank Dr. Ming-Fei Li, the Institute of Ancient History, Chinese Academy of Social Sciences, and Dr. Xiao-Shan Chen, the Institute for the History of Natural Sciences, Chinese Academy of Sciences, for their highly constructive suggestions.

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 2019*. World Health Organization; 2019.
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. Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nature Reviews Microbiology*. 2018;16(4):202-13.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. Barbier M, Wirth T. The Evolutionary History, Demography, and Spread of the *Mycobacterium tuberculosis* Complex. *Microbiology Spectrum*. 2016;4:453-73.
16. 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.
17. 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.
18. 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.
19. Zhang Z. Epidemic chronology of epidemic diseases in ancient China [in Chinese]: Fujian science and Technology Press; 2007.

20. 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.
21. 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.
22. 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):P:193-211.
23. Shangji S, Guiling XU. Historical Evolution of China's Maritime Silk Road. Tropical Geography. 2015.
24. Piazza A, Rendine S, Zei G, Moroni A, Cavallisforza LL. Migration rates of human populations from surname distributions. Nature. 1987;329(6141):714-6.
25. Zei G, Guglielmino C, Siri E, Moroni A, Cavalli-Sforza L. Surnames as neutral alleles: Observations in Sardinia. Human biology. 1983;55:357-65.
26. Barbier M, Wirth T. The Evolutionary History, Demography, and Spread of the Mycobacterium tuberculosis Complex. Microbiology Spectrum. 2016;4:453-73.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. Lin N. Genome analysis of Mycobacterium tuberculosis in China [in Chinese]: Fujian agriculture and Forestry University; 2014.
34. 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.
35. 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.
36. 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.
 37. 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.
 38. Li R, Wang Q, Cheong K. Quanzhou: Reclaiming a glorious past. *Cities*. 2016;50:168-79.
 39. Guiling X, Shangji S. *The Development of the Maritime Silk Road in Guangzhou and the Construction of World Historical Cities*. Urban Insight. 2011.
 40. 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.
 41. Wang GP, Yong MA. *Islamism in Guangzhou at the time of Tang and Song*. Journal of Guangzhou University. 2006.
 42. Zhou X. *A Study on the Origins of Guangzhou Maritime Silk Road and its Development Process*. China Ancient City. 2018.
 43. Gong Y. A textual research on the origin of the immigrants from Huguang to Sichuan in Qing Dynasty [in Chinese]. *Lingnan literature and history*. 2009(4):27-34.

Figures

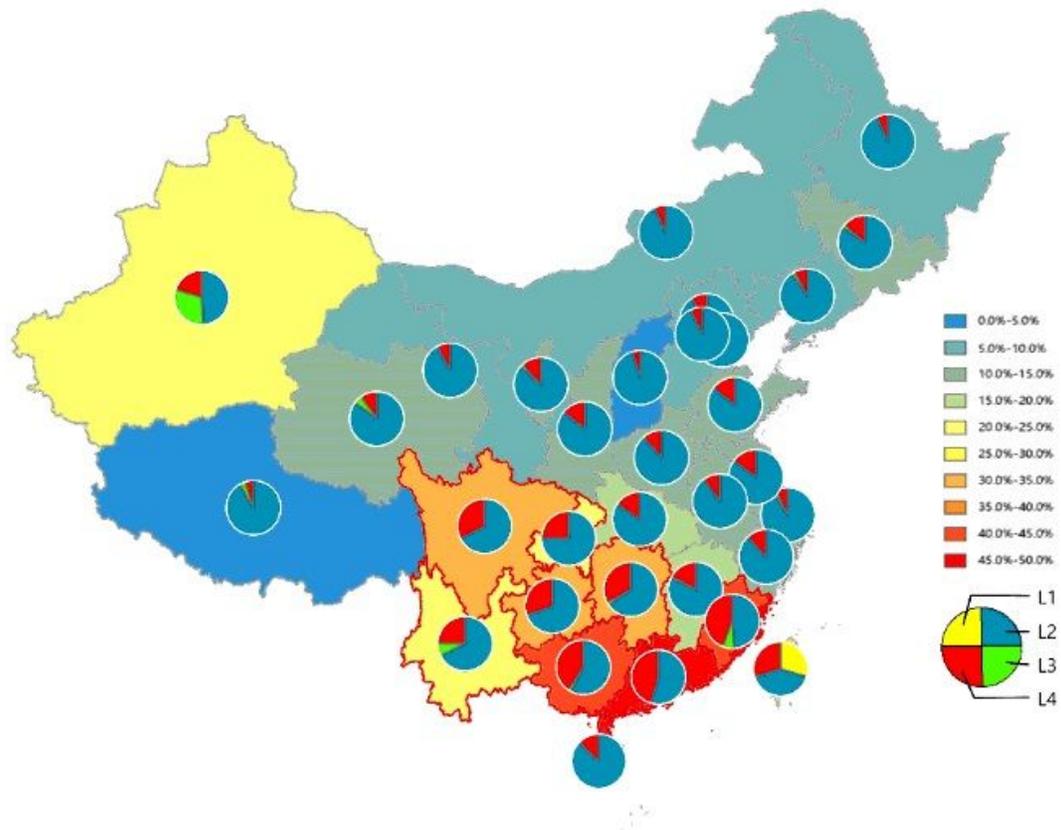


Figure 1

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.

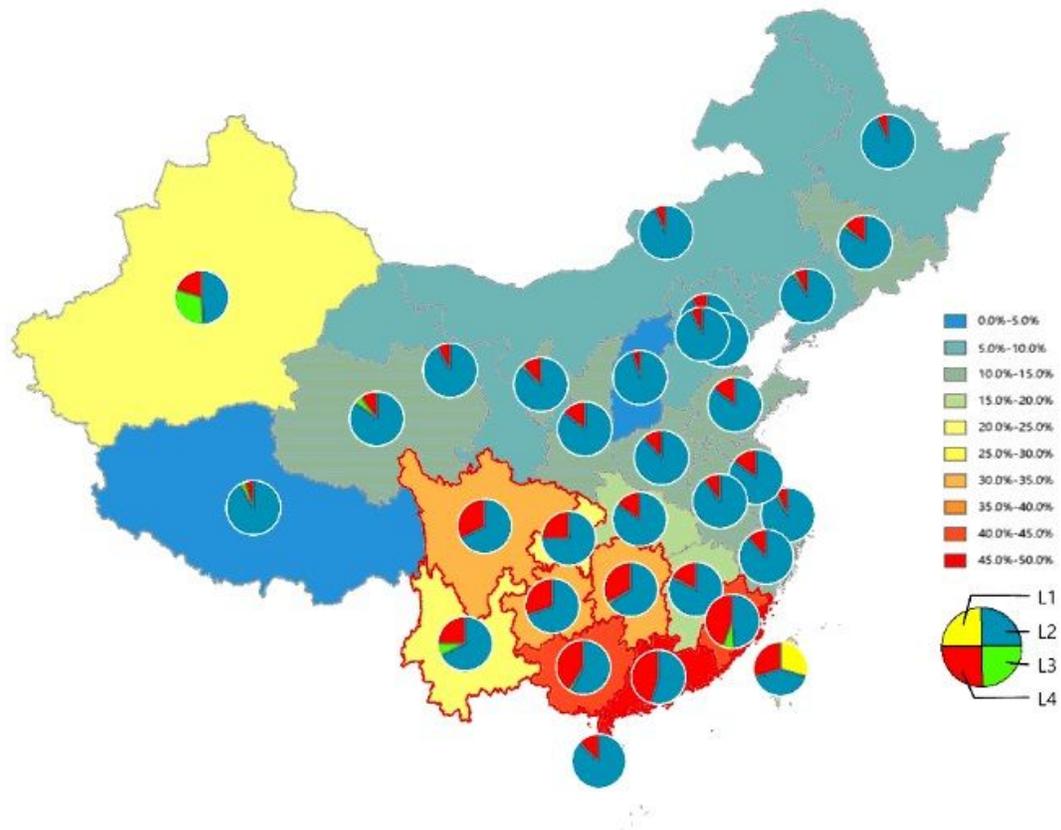


Figure 1

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.

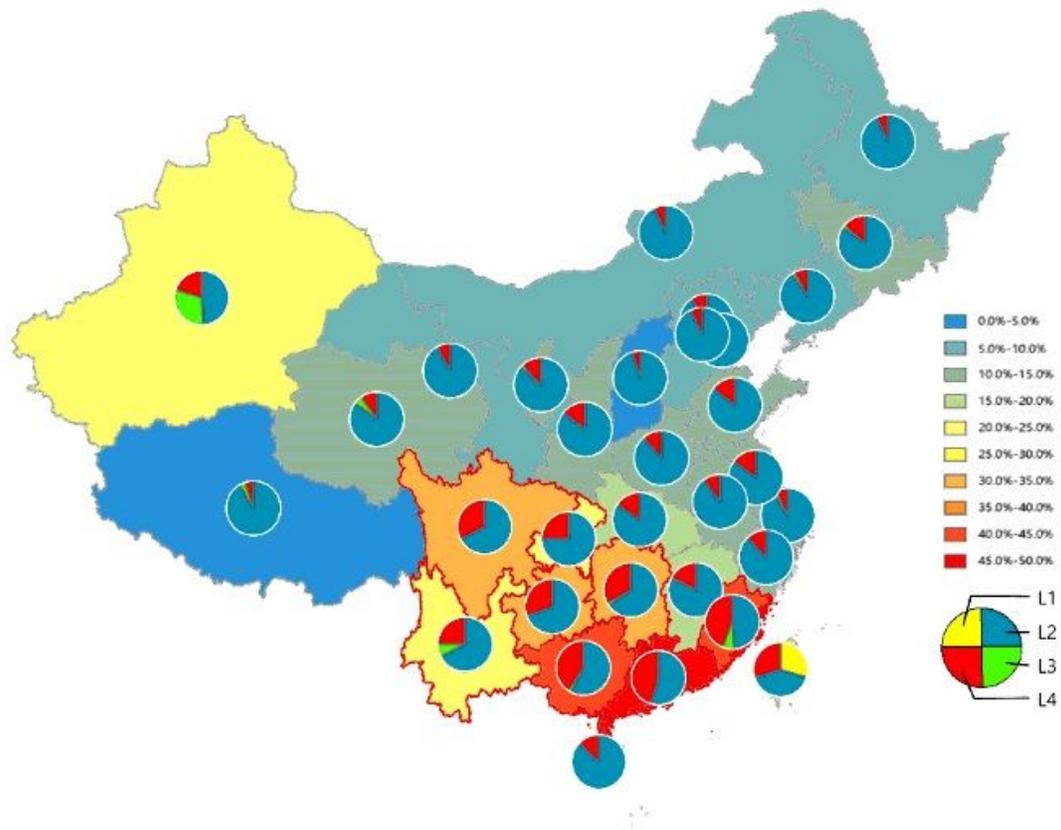


Figure 1

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.

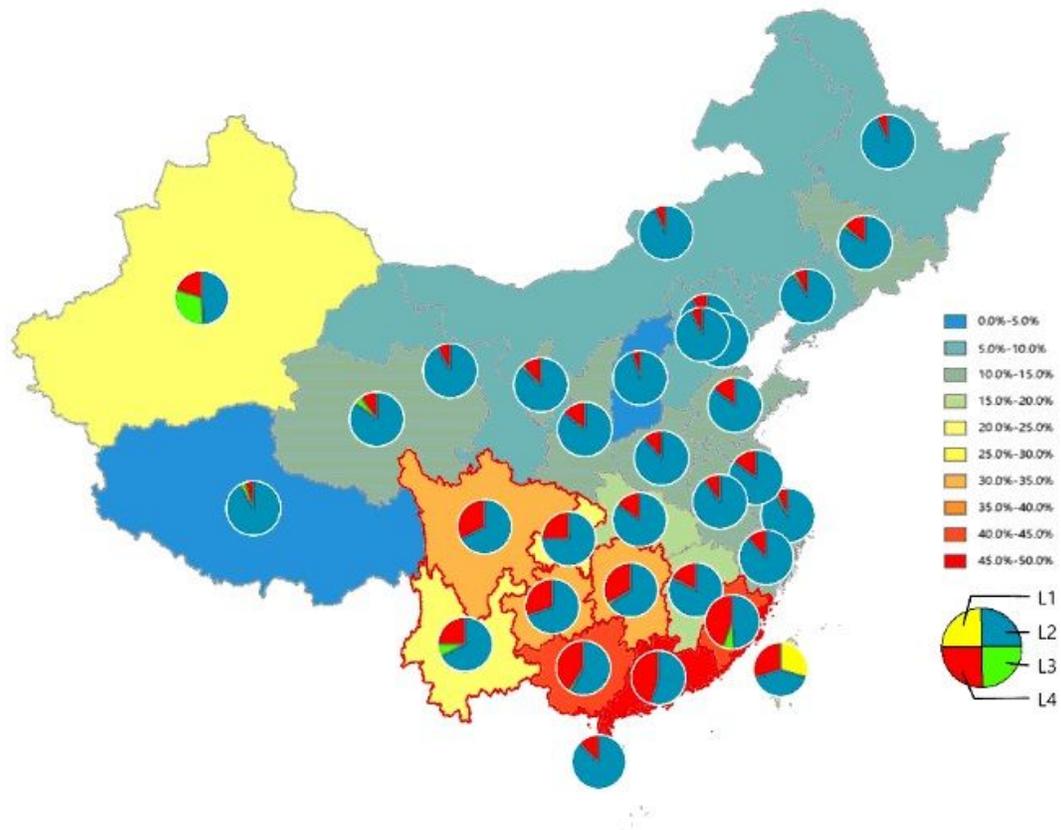


Figure 1

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.

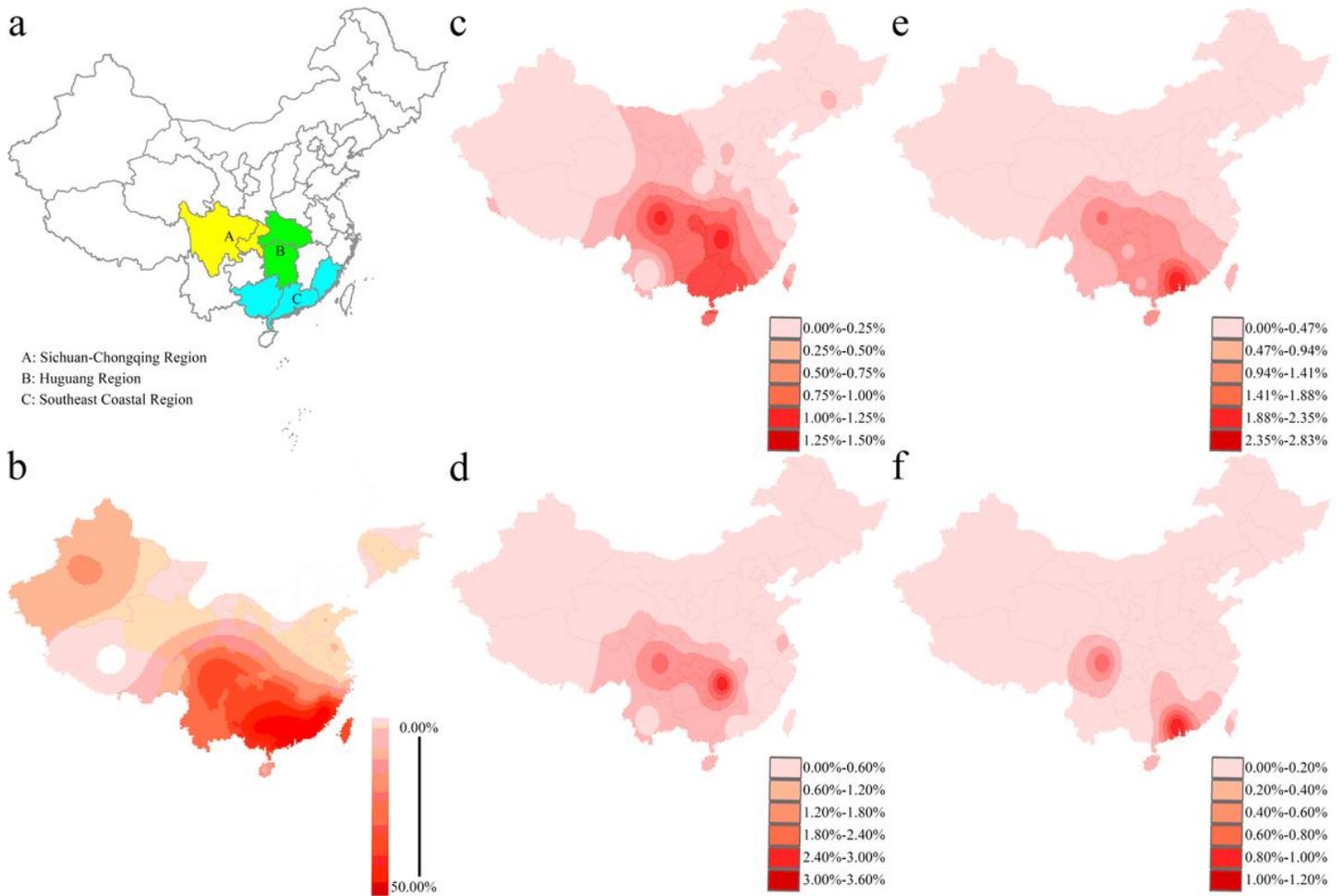


Figure 2

Distribution of lineage 4 and four surname populations (Zeng, Tang, Deng and Zhong) in China. a, Some of the Chinese regions mentioned. “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. It should be noted that Sichuan-Chongqing region, Huguang region were originally administrative regions in history. This map uses modern administrative divisions to show the main scope of these two regions, but they are still a little different from the real scope 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 4 surnames “Zeng, Tang, Deng and Zhong” ’s distribution in China. Color ranges showing the percentage prevalence of (c) Zeng, (d) Tang, (e) Deng, (f) Zhong based on collected data. 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.

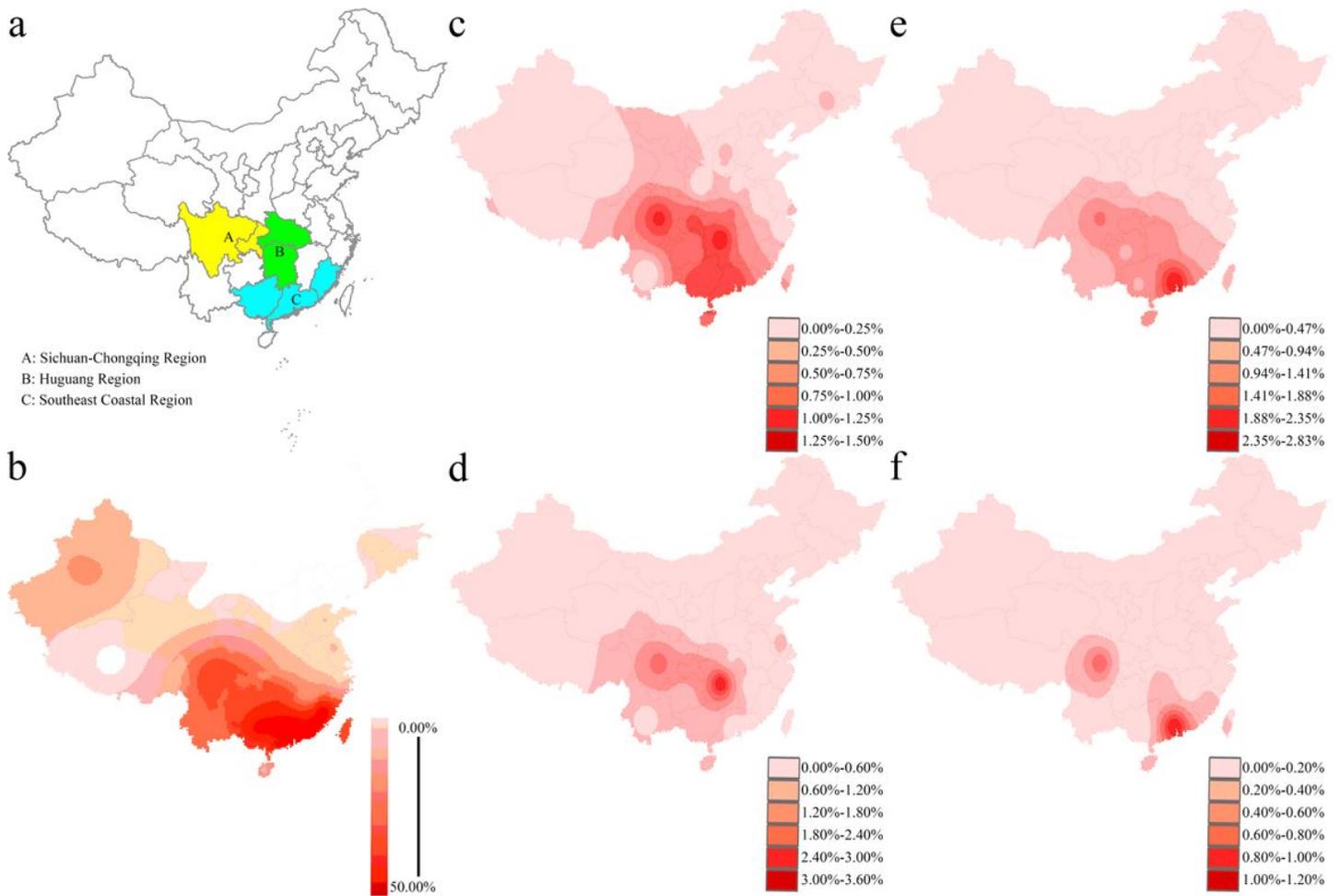


Figure 2

Distribution of lineage 4 and four surname populations (Zeng, Tang, Deng and Zhong) in China. a, Some of the Chinese regions mentioned. “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. It should be noted that Sichuan-Chongqing region, Huguang region were originally administrative regions in history. This map uses modern administrative divisions to show the main scope of these two regions, but they are still a little different from the real scope 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 4 surnames “Zeng, Tang, Deng and Zhong” ’s distribution in China. Color ranges showing the percentage prevalence of (c) Zeng, (d) Tang, (e) Deng, (f) Zhong based on collected data. 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.

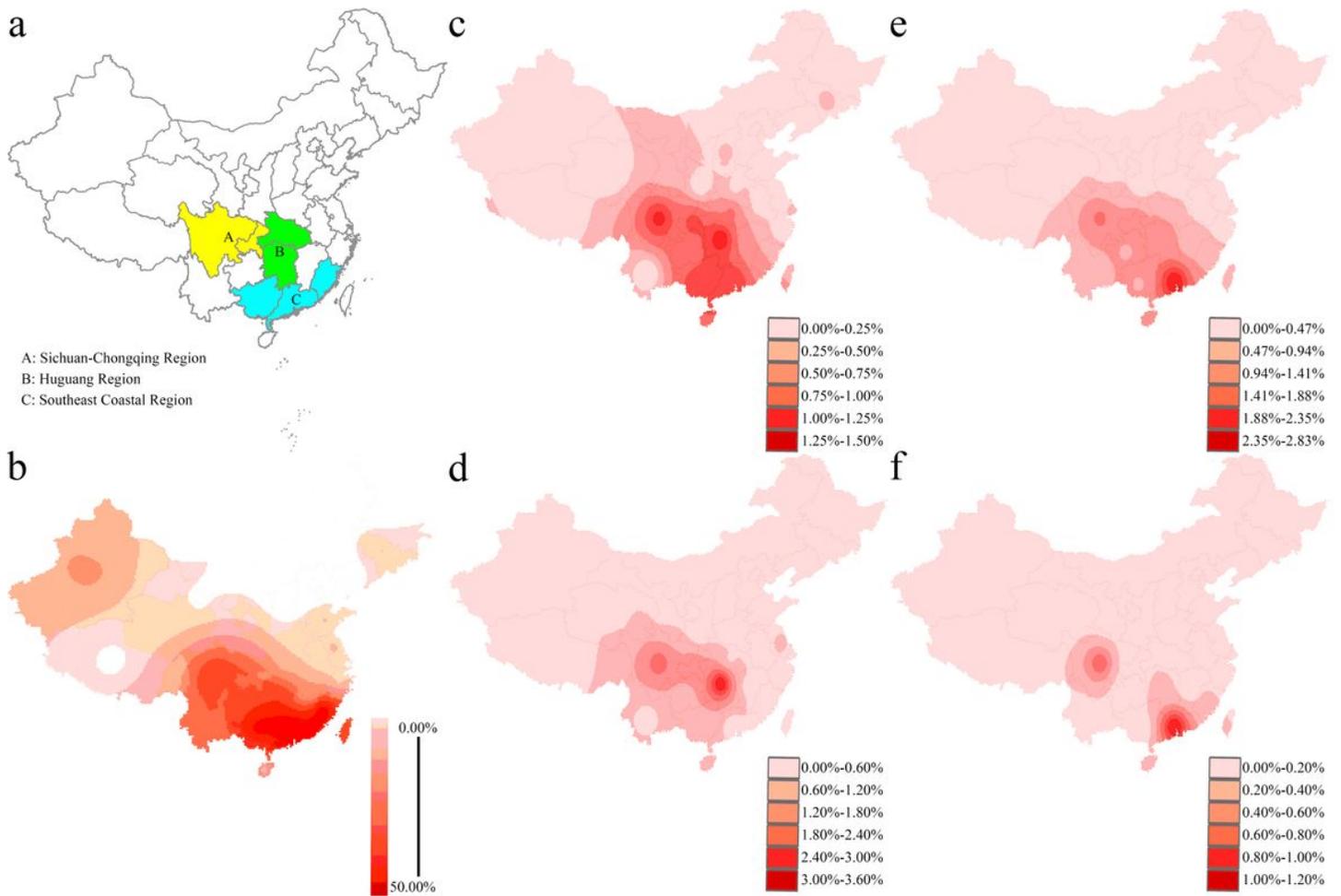


Figure 2

Distribution of lineage 4 and four surname populations (Zeng, Tang, Deng and Zhong) in China. a, Some of the Chinese regions mentioned. “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. It should be noted that Sichuan-Chongqing region, Huguang region were originally administrative regions in history. This map uses modern administrative divisions to show the main scope of these two regions, but they are still a little different from the real scope 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 4 surnames “Zeng, Tang, Deng and Zhong” ’s distribution in China. Color ranges showing the percentage prevalence of (c) Zeng, (d) Tang, (e) Deng, (f) Zhong based on collected data. 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.

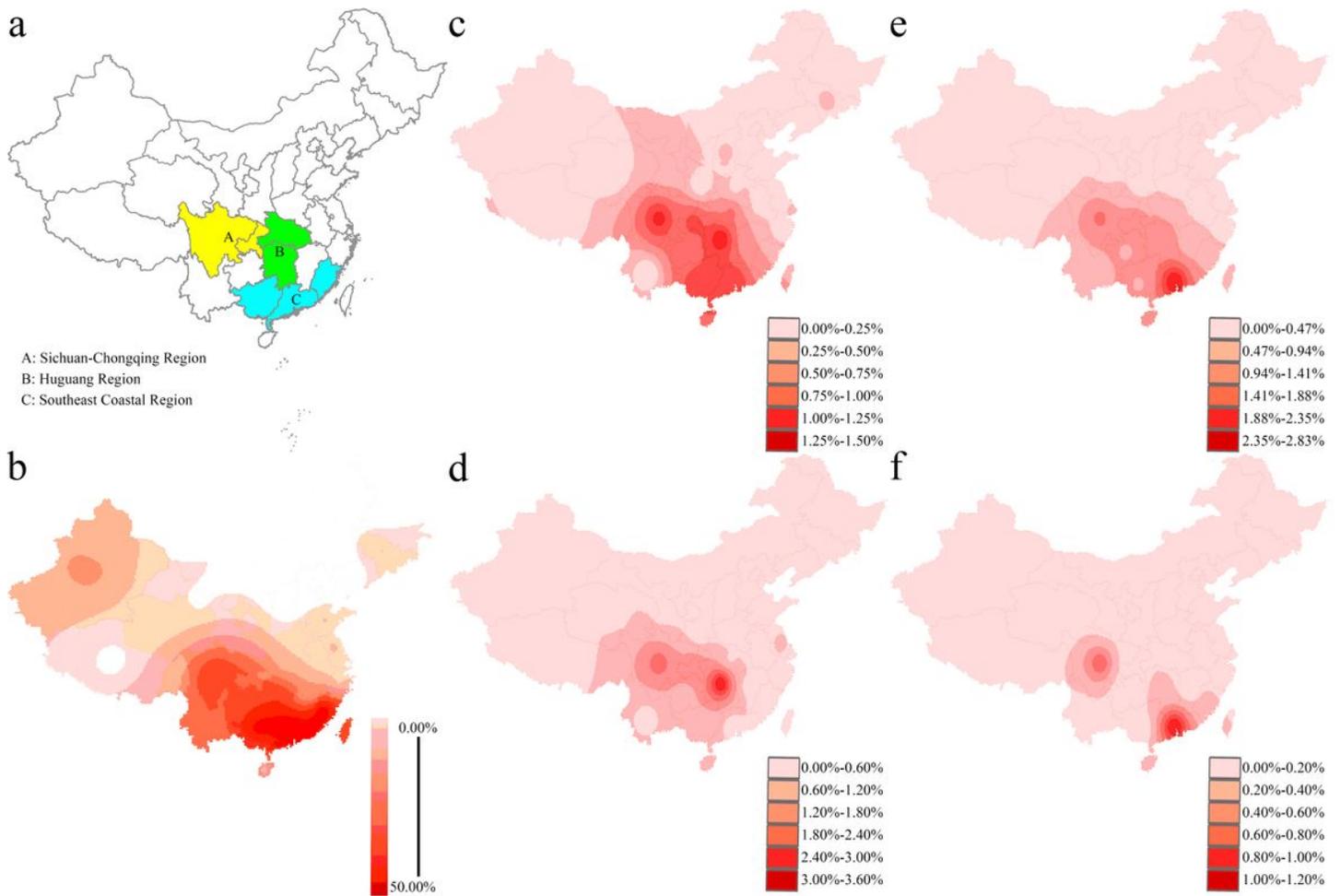


Figure 2

Distribution of lineage 4 and four surname populations (Zeng, Tang, Deng and Zhong) in China. a, Some of the Chinese regions mentioned. “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. It should be noted that Sichuan-Chongqing region, Huguang region were originally administrative regions in history. This map uses modern administrative divisions to show the main scope of these two regions, but they are still a little different from the real scope 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 4 surnames “Zeng, Tang, Deng and Zhong” ’s distribution in China. Color ranges showing the percentage prevalence of (c) Zeng, (d) Tang, (e) Deng, (f) Zhong based on collected data. 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.

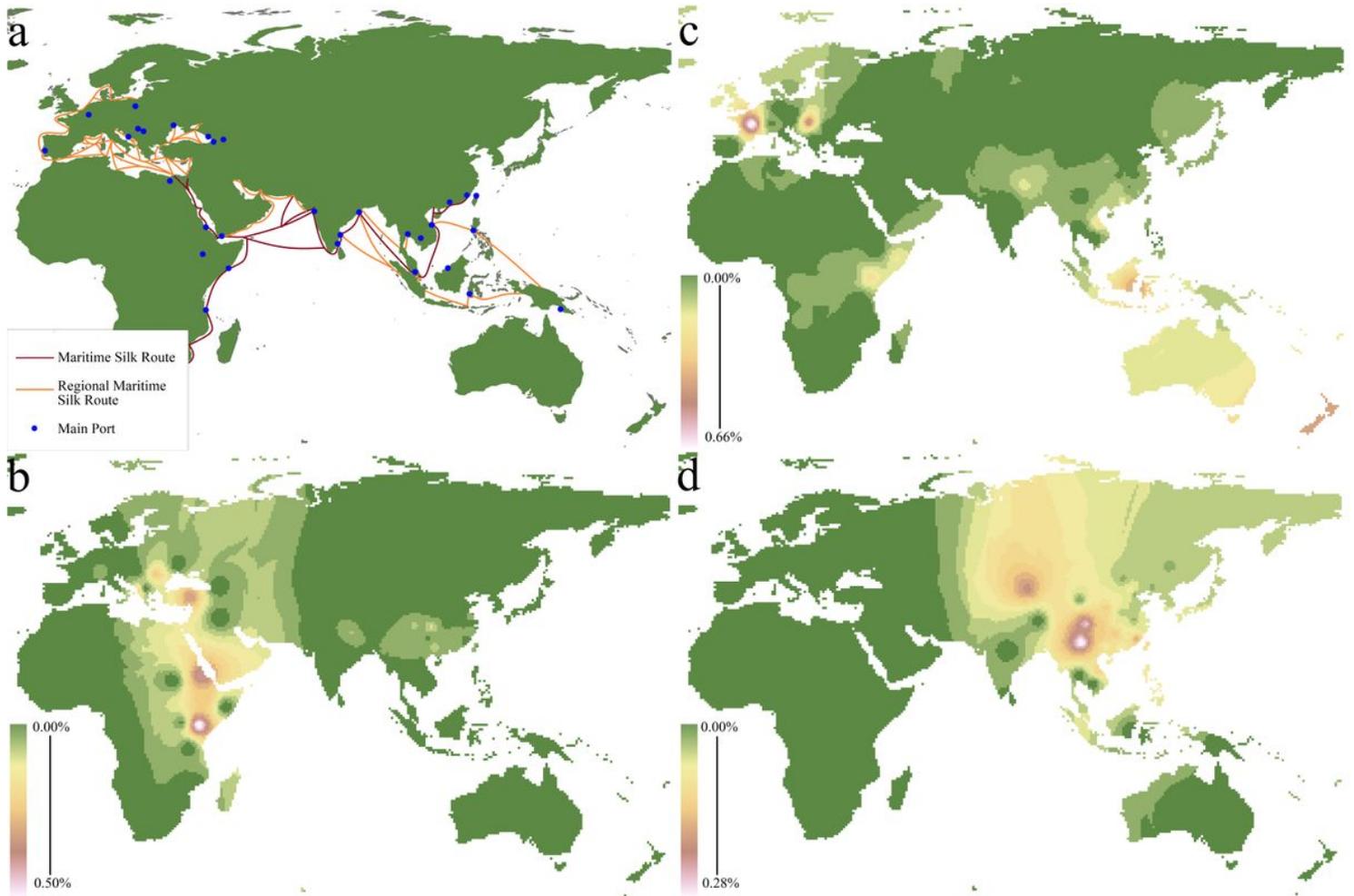


Figure 3

Distribution of L4.2, L4.4, L4.5 and Maritime Silk Road's main ports. 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 and Europe. 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.

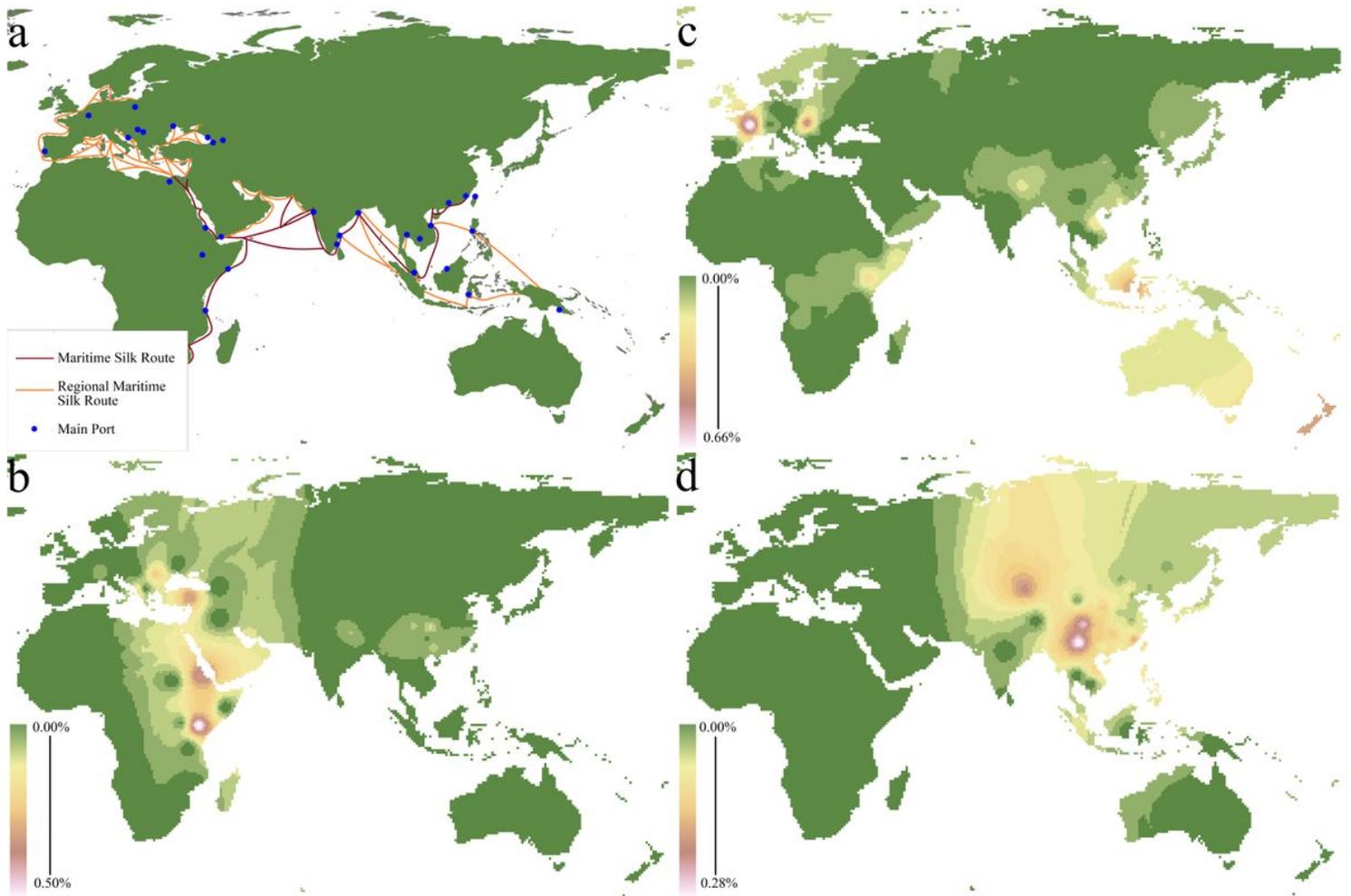


Figure 3

Distribution of L4.2, L4.4, L4.5 and Maritime Silk Road's main ports. 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 and Europe. 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.

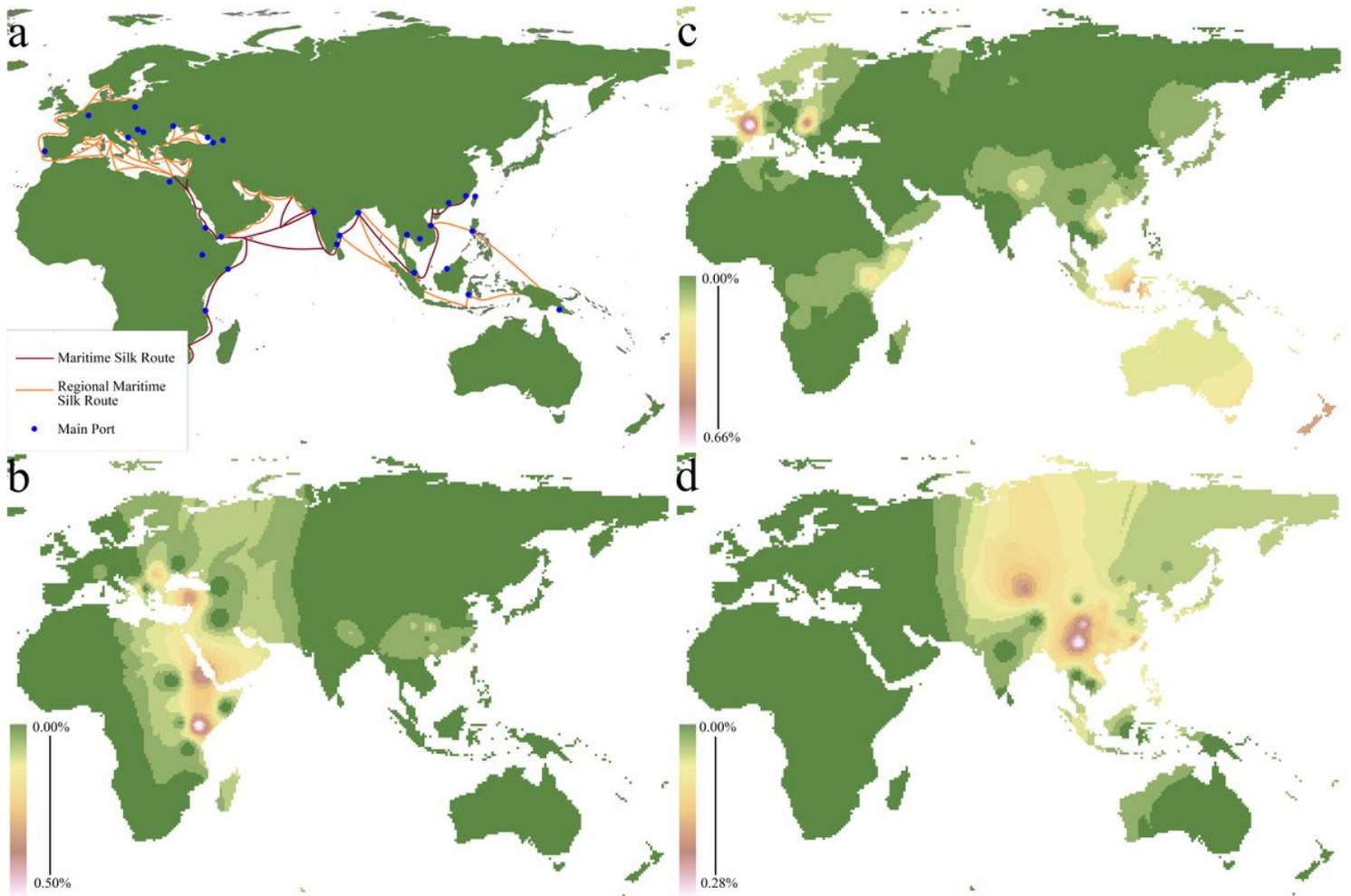


Figure 3

Distribution of L4.2, L4.4, L4.5 and Maritime Silk Road's main ports. 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 and Europe. 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.

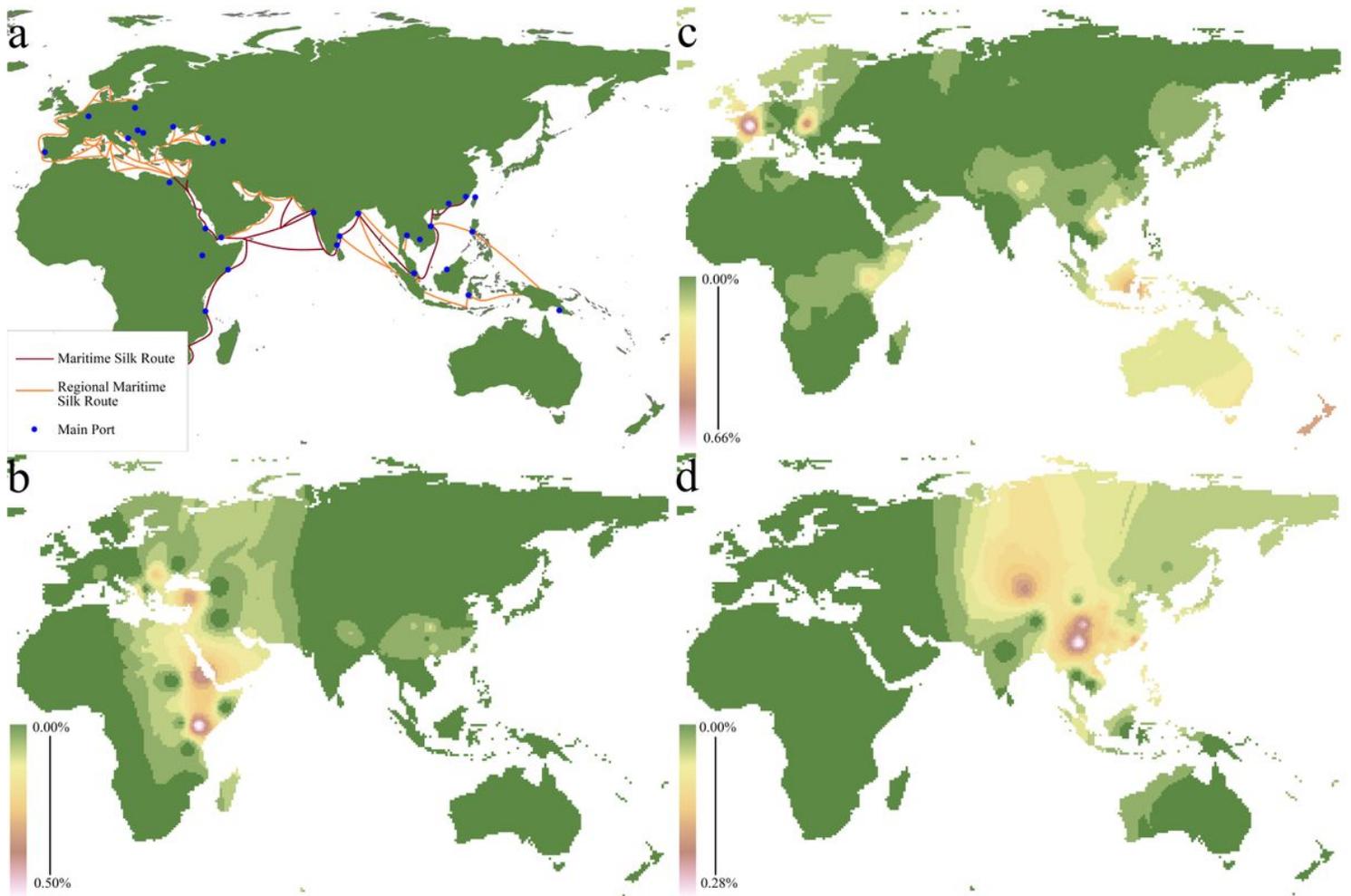


Figure 3

Distribution of L4.2, L4.4, L4.5 and Maritime Silk Road's main ports. 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 and Europe. 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.

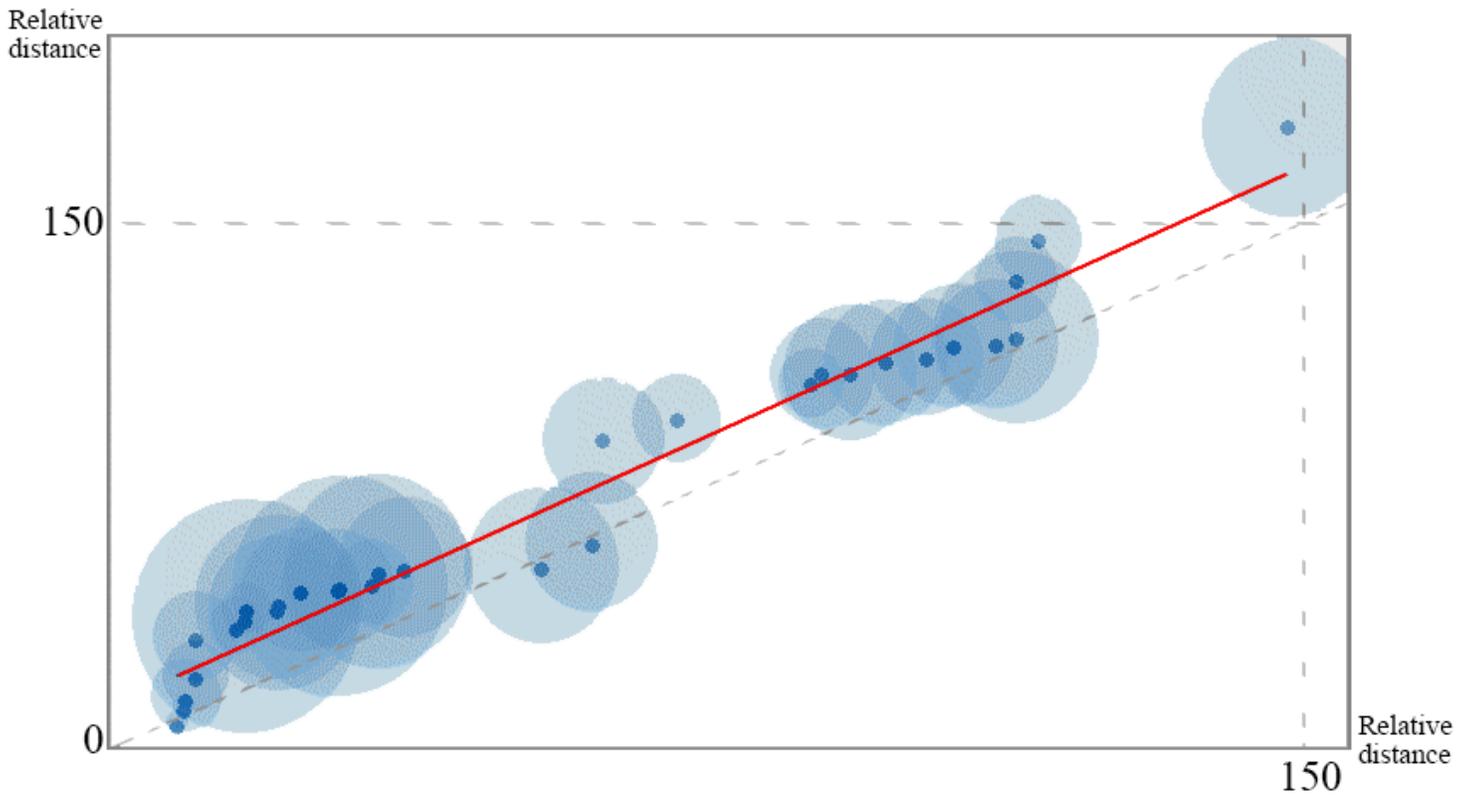


Figure 4

Scatter diagram of the 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.

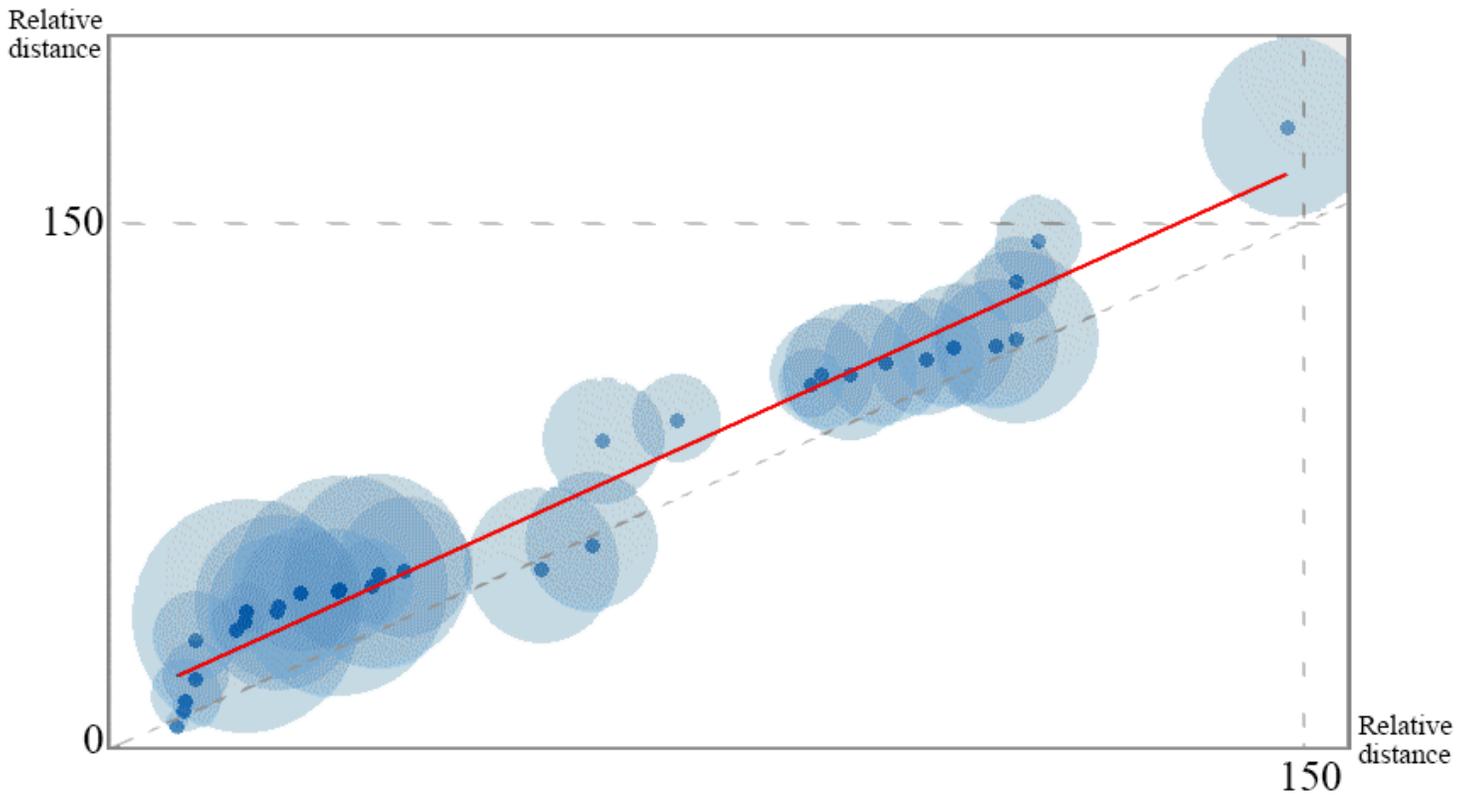


Figure 4

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.

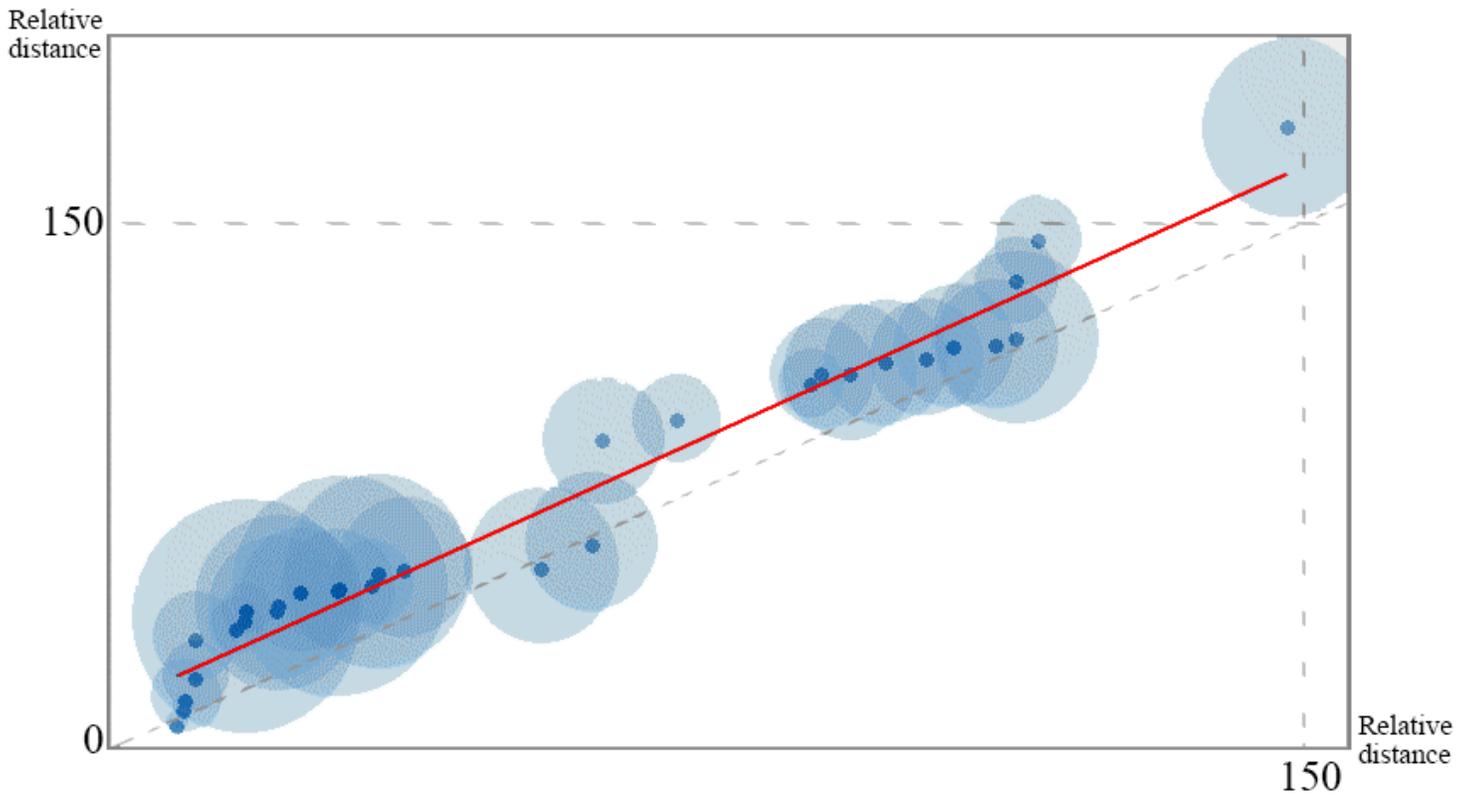


Figure 4

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.

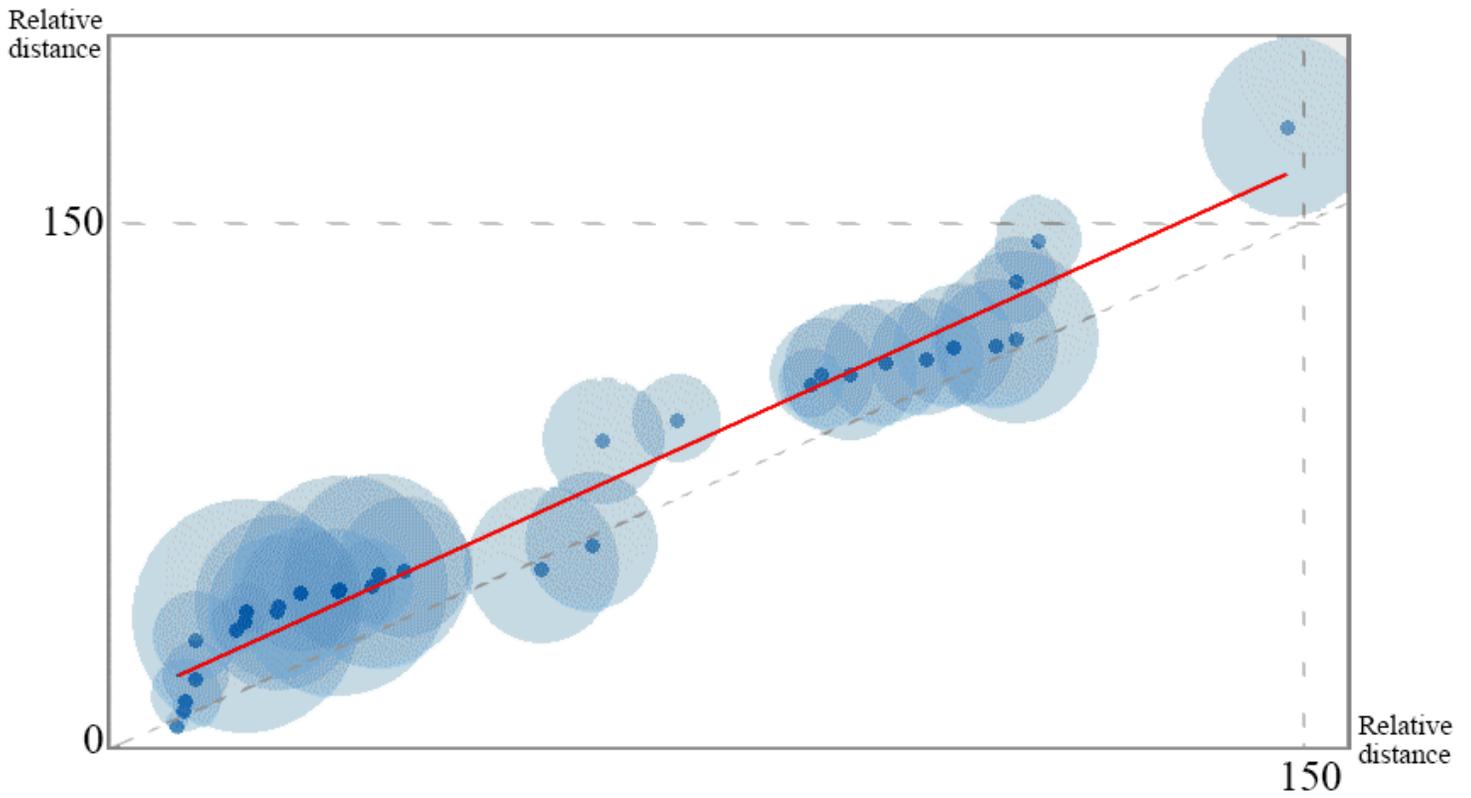


Figure 4

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 possible 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 yellow area represents the Sichuan-Chongqing region, including the provinces of Sichuan-Chongqing region, the green area represents the Huguan region, including provinces of Hunan and Hubei, the blue one for the southeast coastal region, including provinces of Guangxi, Guangdong, and Fujian, and the black area represents the city Guangzhou and Quanzhou. It should be noted that Sichuan-Chongqing region, Huguan region and southeast coastal

region were originally administrative regions in history. This map uses the current administrative divisions to show the main scope of these three regions, but they are different from the scope in history. 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 5

The possible 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 yellow area represents the Sichuan-Chongqing region, including the provinces of Sichuan-Chongqing region, the green area represents the Huguang region, including provinces of Hunan and Hubei, the blue one for the southeast coastal region, including provinces of Guangxi, Guangdong, and Fujian, and the black area represents the city Guangzhou and Quanzhou. It should be noted that Sichuan-Chongqing region, Huguang region and southeast coastal region were originally administrative regions in history. This map uses the current administrative divisions to show the main scope of these three regions, but they are different from the scope in history. 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 5

The possible 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 yellow area represents the Sichuan-Chongqing region, including the provinces of Sichuan-Chongqing region, the green area represents the Huguang region, including provinces of Hunan and Hubei, the blue one for the southeast coastal region, including provinces of Guangxi, Guangdong, and Fujian, and the black area represents the city Guangzhou and Quanzhou. It should be noted that Sichuan-Chongqing region, Huguang region and southeast coastal

region were originally administrative regions in history. This map uses the current administrative divisions to show the main scope of these three regions, but they are different from the scope in history. 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 5

The possible 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 yellow area represents the Sichuan-Chongqing region, including the provinces of Sichuan-Chongqing region, the green area represents the Huguang region, including provinces of Hunan and Hubei, the blue one for the southeast coastal region, including provinces of Guangxi, Guangdong, and Fujian, and the black area represents the city Guangzhou and Quanzhou. It should be noted that Sichuan-Chongqing region, Huguang region and southeast coastal region were originally administrative regions in history. This map uses the current administrative divisions to show the main scope of these three regions, but they are different from the scope in history. 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.

- [MHTBQsupplementalfileforreview.doc](#)
- [MHTBQsupplementalfileforreview.doc](#)
- [MHTBQsupplementalfileforreview.doc](#)
- [MHTBQsupplementalfileforreview.doc](#)