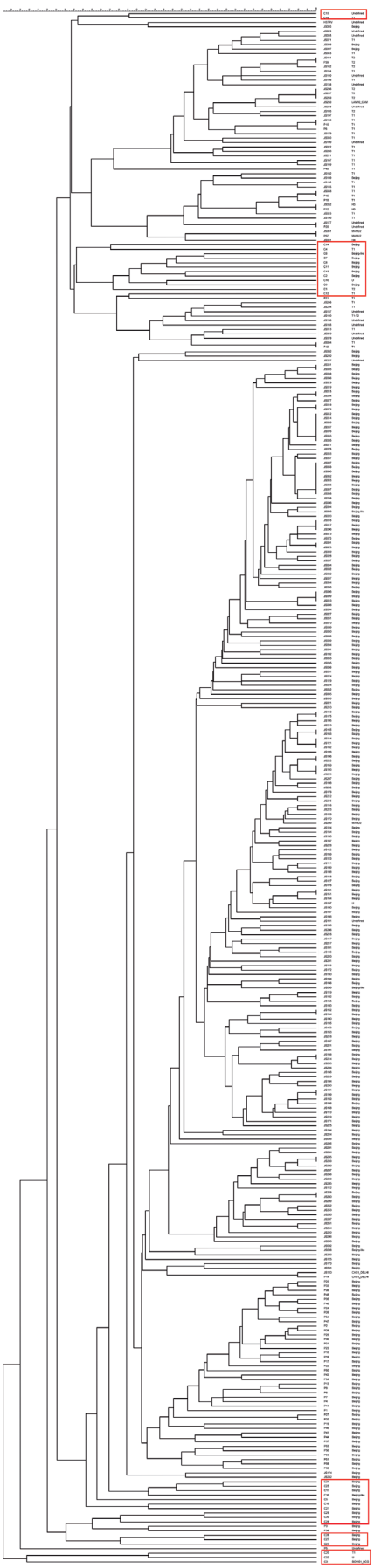
Appendix S2

A comparison of the molecular characteristics of PTB and EBTB

Knowledge of the genetics of *M. tuberculosis* strains found in EPTB is limited (1, 2, 35, 36). The successful culture of CTB strains from skin specimens offers an excellent opportunity to gain insight into their genomic properties, especially in the context of PTB strains involved in endemic and epidemic infections. Specimens from cutaneous patients are collected at our hospital. Thus far, no CTB patients with symptoms of PTB by chest X-ray have been seen. On the other hand, at Shanghai pulmonary hospital and Chinese Center for Disease Control and Prevention, where the focus is on PTB, clinicians have not detected any PTB patients with CTB. In addition, there is no research into the relationship between CTB and PTB. This study begins to address this gap by genotyping CTB and PTB strains. This molecular study is an extension of the fundamental microbiological, diagnostic and therapeutic work up on CTB patients detected in China. It is based on methodologies that have been standardized and universally implemented for PTB molecular epidemiology, such as RD deletion analysis, spoligotyping, and MIRU-VNTR typing, and provides the first multi-level genomic description of CTB strains. Such techniques have provided powerful means to analyse the frequency of various *M. tuberculosis* phylogenetic lineages to understand their evolutionary history and of regional genotypes to trace transmission networks. At the lineage level, *M. tuberculosis* strains cluster into 7 lineages, each associated with specific global geographical locations (37). Lineage 2, as shown by the deletion of the RD105 marker, consisting of mainly Beijing genotype, is prevalent in East Asia and Russia. The present study found that CTB of China, like their PTB counterparts, are derived from endemic long-standing lineages, i.e Lineage 2 (38).

Spoligotyping is a rapid and easy method to screen for Beijing family; however, pseudo-Beijing strains that have the typical spoligotype, but are not phylogenetically related, have been discovered (39). In this study, 85 strains (29 CTB and 56 PTB) could be divided into Beijing and Beijing-Like, T1, T2, H3, U, CAS, MANU2, BOVIS and CAS1-DELHI families. Only 3 spoligotypes (1 CTB and 2 PTB) could not be assigned to a family (i.e. are undefined). The percentage of CTB Beijing (70%) genotype strains was nearly consistent with that of the PTBs (77.6%) analysed in this study, and within the range reported for PTB across China: Jiangsu (80.38%) (40), Anhui (85.35%) (41), Jiangsu and Zhejiang (69.23%) (42), Shanghai (77.14%) (43, 44), and Gansu (86.23%) (45). However, spoligotyping does not have sufficient ability to discriminate *M. tuberculosis* strains for molecular epidemiology applications, thus it should be performed in combination with another independent marker system, such as the MIRU-VNTRs. Multiple MIRU-VNTR loci increase the resolution, and so a 24-locus set has been standardized (46). In this study, a high discriminatory power was obtained by MIRU-VNTR typing. When a dendrogram of the 30 CTB strains was generated in MIRU-VNTRPLUS database, all the strains were distributed in 3 large clades. The majority of the CTB which were of Beijing genotype grouped with other Beijing strains in the database; the rest of the genotypes also partitioned consistent with spoligotype families of strains in the database (**SFig. 1**). This indicates that, despite the higher resolution of strain types by MIRU-VNTRs, the broader classification according to spoligotypes was retained.

The HGDI for loci in CTB were higher than for PTB, indicating that CTB are more dispersed in space and time than PTB (Table SIII[1]). The lack of clustering (defined as identical genotypes across all 24 MIRU-VNTR loci), among CTB or PTB genotypes, or all of them, is not unexpected, as CTB and PTB have been collected passively over a wide time-period from patients with no known epidemiological links. Neither the 30 CTB nor 58 PTB represent linked recent infections. Nevertheless, the combination of spoligotyping and MIRU-VNTR genotypes, analysed by metric MDS, revealed that all the strains could still be divided into 2 larger groups, i.e. Beijing and non-Beijing genotypes. Furthermore, the Beijing genotypes could be divided into CTB and PTB subgroups (Fig. S3[1]), with the CTBs more divergent than the PTBs. The Beijing strain is more readily aerosolized (47) and more infectious than non-Beijing strains (48). The infection progresses more quickly in PTB than in EPTBs. Therefore, PTB strains are more transmissible, resulting, at MIRU-VNTR level, in more closely related genotypes than CTBs. CTB infection develops more slowly with more subtle clinical symptoms, which may explain why the causative strains are divergent compared with those that cause TB of the lungs.



SFig. 1. Neighbor-joining dendogram based on genotype using the web-based MIRU-VNTRplus tool. The cutaneous tuberculosis (CTB) isolates reported here are highlighted in boxes, reference strains are from the MIRU-VNTR*plus* database.