Letter

Origin and primary dispersal of the Mycobacterium tuberculosis Beijing genotype: Clues from human phylogeography

Igor Mokrousov,¹,⁶ Ho Minh Ly,³ Tatiana Otten,² Nguyen Ngoc Lan,⁴ Boris Vyshnevskyi,² Sven Hoffner,⁵ and Olga Narvskaya¹

¹Laboratory of Molecular Microbiology, St. Petersburg Pasteur Institute, St. Petersburg, 197101, Russia; ²Research Institute of Phtisiopulmonology, St. Petersburg, 193063, Russia; ³National Institute of Hygiene and Epidemiology, Hanoi 10000, Vietnam; ⁴Pham Ngoc Thach Tuberculosis and Lung Diseases Centre, Ho Chi Minh City, Vietnam; ⁵Swedish Institute for Infectious Disease Control, Solna 17182, Sweden

We suggest that the evolution of the population structure of microbial pathogens is influenced by that of modern humans. Consequently, the timing of hallmark changes in bacterial genomes within the last 100,000 yr may be attempted by comparison with relevant human migrations. Here, we used a lineage within Mycobacterium tuberculosis, a Beijing genotype, as a model and compared its phylogeography with human demography and Y chromosome-based phylogeography. We hypothesize that two events shaped the history of the Beijing genotype: (1) its Upper Palaeolithic origin in the Homo sapiens sapiens K-M9 cluster in Central Asia, and (2) primary Neolithic dispersal of the secondary Beijing NTF: IS6110 lineage by Proto-Sino-Tibetan farmers within east Asia (human O-M214/M122 haplogroup). The independent introductions of the Beijing strains from east Asia to northern Eurasia and South Africa were likely historically recent, whereas their differential dissemination within these areas has been influenced by demographic and climatic factors.

[Supplemental material is available online at www.genome.org.]

Intriguing clues about the history of a biological species can be derived from the study of the geographical distribution of the phylogenetic/genealogical lineages, in the approach known as “phylogeography” (Avise et al. 1987). The underlying assumption of human phylogeography is that there is a correspondence between the overall distribution of haplotypes and haplogroups and past human movements. The uniparentally inherited non-recombining haploid mtDNA and the Y chromosome loci are particularly sensitive to the influences of drift, especially founder effect. Consequently, these loci are suitable for assessing the origins of contemporary population diversity and provide context for paleontological hypothesis testing (Foley 1998). The mutation rate of the maternally transmitted mitochondrial genome is ∼10 times higher than that of nuclear DNA, which provides abundance of polymorphic sites but creates difficulties in reconstructing genealogies owing to repeated and reverse mutations. By contrast, the mutation rate of the paternally inherited non-recombining portion of the Y chromosome (NRY) is comparable to that of nuclear DNA, which means that polymorphisms are more difficult to find but genealogies are easier to reconstruct. In addition, the greater length of the NRY DNA compared with mtDNA compensates in data analysis for its lower mutation rate (Cavalli-Sforza and Feldman 2003).

The rarity of back and recurrent mutations in NRY contributes to the property of displaying the strongest geographic correlation and greatest diversity among, rather than within, populations. To date, NRY binary polymorphisms have been widely used to trace the origin and migration events of modern humans (Underhill et al. 2001). Here, we propose the hypothesis that NRY-based phylogeography of H. sapiens sapiens, offers a convenient spatiotemporal framework for inferring early primary dispersals of those human pathogens that are essentially (1) devoid of horizontal gene transfer and (2) family/household-transmitted. We believe that Mycobacterium tuberculosis makes an ideal model to test this hypothesis.

A 70% excess of male over female tuberculosis (TB) cases are reported globally each year (Uplekar et al. 2001). Although the reasons for this difference are unclear, there exists a strong opinion that claims that there is a gender bias in TB infection (Thomas 2000). Therefore, assuming (1) a predominantly family/household mode of transmission of TB in a preindustrialized time and (2) a likely bias in TB infectivity toward males, the M. tuberculosis transmission seems to resemble the unidirectional inheritance of the paternally transmitted human NRY DNA.

Whereas the M. tuberculosis complex includes mycobacteria widely differing in terms of their host specificities, M. tuberculosis sensu stricto (M. tuberculosis) is exclusively a human pathogen. The tubercle bacillus has the remarkable ability to persist in the human host in the form of a long-term asymptomatic infection referred to as latent TB. In latent TB, the host immune response is capable of controlling the infection and yet falls short of eradicating the pathogen. One third of the world population is estimated to have a latent or dormant TB infection, which was perhaps the predominant mode of M. tuberculosis coexistence with human host in a preindustrialized time when transmission of the pathogen was mostly family-linked. M. tuberculosis has a clonal population structure, and the circulating strains show...
strong geographical affinities. Among several genetic families identified within *M. tuberculosis* (Sola et al. 2001; Baker et al. 2004), the Beijing genotype is marked by genetic homogeneity and geographical omnipresence (Bifani et al. 2002; Glynn et al. 2002). Taken together, these data likely reflect its rapid global spread during the past century, if not the past few decades. For the first time, it was identified in the *M. tuberculosis* strains isolated in the Beijing area in China, which coined its name (van Soolingen et al. 1995). Subsequent studies have shown that these strains are endemically prevalent in east Asia, South Africa, and northern Eurasia (Bifani et al. 2002; Glynn et al. 2002).

Three chromosomal regions are sufficiently, although not exceedingly, polymorphic to be relevant to the evolutionary scale studies of this genotype (Figs. 1, 2). The first one is the direct repeat (DR) locus which consists of minisatellite alternating exact DRs and variable spacers (Kamerbeek et al. 1997) and represents a large polyphyletic family of DNA repeats found in many bacterial lineages (Jansen et al. 2002). Van Embden et al. (2000) hypothesized that such a locus in *M. tuberculosis* might have initially presented a region consisting of hundreds of short (36-bp) tandem repeats. Variable spacers emerged and accumulated further during evolution, and subsequent neutral changes in the DR locus in *M. tuberculosis* have occurred and are still occurring via consecutive deletions of either single units or contiguous blocks, occasionally including IS6110-mediated disruption and recombination (Beggs et al. 2000; van Embden et al. 2000). In the Beijing genotype, most of the DR units were deleted during evolution, perhaps by a single event mediated by IS6110 recombination (Beggs et al. 2000). Thus the characteristic structure of the DR locus is a marker that defines the Beijing genotype and distinguishes it from other families within *M. tuberculosis*.

Second, a genetic marker specific to the Beijing genotype is the so-called NTF region that may harbor IS6110 insertion(s) (Plikaytis et al. 1994; Bifani et al. 2002). Three NTF variants of Beijing strains are distinguished based on the presence/absence of IS6110 sequence, thus providing a rough subdivision within this genotype (Fig. 2). A strictly clonal population structure of *M. tuberculosis* species as a whole further implies a single founder population of one of its lineages (the Beijing strain), thus justifying a straightforward evolutionary scenario and unidirectional evolution of the NTF locus. The W-branch prevalent in the United States has no IS6110 insertion during evolution, and subsequent neutral changes in the DR locus in *M. tuberculosis* have occurred and are still occurring via consecutive deletions of either single units or contiguous blocks, occasionally including IS6110-mediated disruption and recombination (Beggs et al. 2000). Van Embden et al. (2000) hypothesized that such a locus in *M. tuberculosis* might have initially presented a region consisting of hundreds of short (36-bp) tandem repeats. Variable spacers emerged and accumulated further during evolution, and subsequent neutral changes in the DR locus in *M. tuberculosis* have occurred and are still occurring via consecutive deletions of either single units or contiguous blocks, occasionally including IS6110-mediated disruption and recombination (Beggs et al. 2000). Thus the characteristic structure of the DR locus is a marker that defines the Beijing genotype and distinguishes it from other families within *M. tuberculosis*.

![Figure 1](Image)

Figure 1. Position of the MIRU-VNTR and DR loci on the *M. tuberculosis* H37Rv chromosome and their structure. In bold are eight polymorphic *MIRU* loci used to build the Beijing genotype network (Fig. 3). NTF region is specific for Beijing strains and not shown on the map.
We therefore designated the Beijing genotype strains from Singapore as those first- or second-generation immigrants from mainland China. Singapore suggested that the majority of the Singaporean population is of Chinese descent. In 2004, Sun et al. who reported Beijing strains from Singapore also obtained their relatives. Dotted branches depict possible alternatives. Question mark means intermediate node in which no strain was found in our database. In 2004, Sun et al. who reported Beijing strains from Singapore suggested that the majority of the Singaporean population is of Chinese descent and that tuberculosis mainly affects the elderly, many of whom are first- or second-generation immigrants from mainland China. We therefore designated the Beijing genotype strains from Singapore as Chinese.

All 12 loci taken together, the allelic diversity of the local samples varied from high (China, 0.95; Vietnam, 0.88) to moderate (South Africa, 0.75) and low (Russia, 0.65). The estimate for Bangladesh (0.78) is probably biased (underestimated) due to the small sample size (n = 15). At first glance, two principal core types are present at the MIRU-based most parsimonious Beijing network, MT2 and, especially, MT11 to which most other types are directly linked (Fig. 3). These two types are likely to be the initial and most ancient Beijing variants. A closer look at the type distribution in the particular samples reveals that the Chinese strains are located all over the network, at both inner and many terminal nodes. Together with highest diversity, this suggests the Chinese sample to be the most ancient. In contrast, the South African strains are found only in the terminal positions and are probably the youngest sample, that is, comparatively most recently introduced. Finally, Russian strains are located in both core nodes but show the least diversity, a pattern that may reflect their relatively ancient introduction but only a recent dissemination in this area.

To estimate a geographical component of the variation in the Beijing genotype and, ultimately, to partly reconstruct its dispersals, we further compared the four local populations of the Beijing strains (Russia, Vietnam, China, and South Africa). The Bangladesh sample was excluded from this analysis because of its small sample size; in addition, these strains were not confirmed by IS6110-RFLP typing to be truly unlinked strains (Banu et al. 2004).

The Neighbor-Joining (NJ) tree showed a topology that strikingly resembled geographical distances (Fig. 4A). It may be noted that using the NJ method, mixing between populations shortens the mixed branch and moves it toward the tree’s origin (Cavalli-Sforza 2001). Therefore, a short distance between Vietnamese and Chinese populations (Fig. 4A) may result from their geographical proximity. Conversely, a long South African branch (Fig. 4A) may be due to a very small number of founders.

A representation in the form of phylogenetic tree may oversimplify relations between complex populations, and we, therefore, also estimated them by principal component (PC) analysis by means of multidimensional scaling (MDS) (Fig. 4B). The Chinese population is placed in the central position in both the NJ tree and the MDS graph and, thus, may be interpreted as the most ancient and ancestral compared with the other three populations corroborating the above network analysis. The PC method has been shown in many studies to be very useful in resolving superimposed human migrations since PCs are statistically independent from one another and so can isolate independent expansions (Cavalli-Sforza 2001). At the same time, PCs cannot distinguish two expansions having the same area of origin and taking place at different times. In this view, although the first PC of MIRU-VNTR loci explains virtually all observed geographical variation of the Beijing genotype (i.e., four populations from the major high-prevalence endemic areas), this result should be interpreted with caution since it does not determine how many expansions emanated from east Asia, bringing the Beijing genotype to Russia and South Africa.

**NTF locus analysis**

Multiplex PCR analysis of the NTF region revealed that 42 of 44 Russian Beijing strains harbored one IS6110 insertion. The two remaining Russian Beijing strains had no IS6110 insertions in this region and hence belonged to the ancestral N-branch. These two strains (IS6110-RFLP profiles) were previously shown to represent 7% of the Beijing strains circulating in Russia and were also defined as ancient by use of other genomic markers, such as the IS1547 and Rv3135-PPE sequences (Mokrousov et al. 2002).

It should be noted that the NTF region was not analyzed in the previously published MIRU studies that we used for our analysis, and conversely, published data on the NTF variation were not accompanied by MIRU profiles. In particular, MIRU-VNTR data are not yet available for the W-branch strains and therefore are not included in our MIRU network analysis. As a consequence, we
could not directly superpose NTF alleles onto the whole Beijing network based on MIRU’s, except for the Russian M. tuberculosis populations (Fig. 3). This comparison placed one of two ancient Beijing strains from Russia into the core MIRU type MT11. Consequently, this type MT11, to which many other types are directly linked, may indeed be considered as the ancestral one to which all the network may be rooted (Fig. 3). However, further studies on both NTF and MIRU loci in more Beijing strains from diverse geographical areas are undoubtedly required to confirm this hypothesis.

Discussion

Thirty years ago, William H. McNeill (1976) reviewed how human history appeared influenced by various local and global epidemics, the Black Death being the most notorious example. Here, we propose a reverse hypothesis that (1) the population structure of microorganisms may be partly shaped by history of their human carriers, and consequently, (2) the timing of hallmark changes in bacterial genomes within the past 100,000 yr may be attempted by comparison with relevant human migrations. Here, we used as a model one globally dispersed genetic lineage within Mycobacterium tuberculosis, the Beijing genotype. Currently, these strains attract great attention worldwide, because they demonstrate some important pathogenic features: increased virulence in the BCG-vaccinated mice (Lopez et al. 2003), the ability to more rapidly multiply in human macrophages (Zhang et al. 1999), and a presumably easier adaptation to changing environments due to mutator alleles of the mutT genes (Rad et al. 2003).

In an attempt to confirm and date the routes of expansion of the Beijing genotype suggested from its phylogeography and to trace back its origin, we compared our results with known prehistoric migrations of modern humans, since primary “out-of-Africa” dispersals supported by NRY binary markers (Underhill et al. 2001). In brief, a single most parsimonious phylogeny was recently constructed for the 153 human NRY haplotypes, and a hierarchical nomenclature system, which superseded and unified past nomenclatures, was suggested (Y Chromosome Consortium 2002). Two complementary nomenclatures of NRY were proposed. The first one is hierarchical, and capital letters (A-R) are used to identify the major clades constituting front symbols of all subsequent subclades. Alternatively, or rather complementarily, haplogroups can be named by the “mutations” that define lineages rather than by the “lineages” themselves. Thus, a second nomenclature retains the major haplogroup information (i.e., 19 capital letters) followed by the name of the terminal mutation that defines a given haplogroup.

We sought to define a human host population in which the most recent common ancestor of the Beijing primordial N-branch (currently, endemic North American) and the ancient MIRU types MT11 and MT2 (radiating through a presumably Chinese primary expansion) appeared. This most recent ancestor could not be in the initial group of humans migrating from Africa, nor it could be in next step Levantine populations since this genotype is not endemic in Africa as a whole or in the Middle East and Europe (Bifani et al. 2002; Glynn et al. 2002). It could not have arisen in China (east Asia), since Chinese isolates already had one IS6110 insertion in the NTF locus (Bifani et al. 2002) and presented a second step in the Beijing evolution. It seems unlikely that the North American N-branch (i.e., initial Beijing variant) emerged in situ, since we can hardly imagine any significant human gene flow from there to East Asia or Eurasia as a whole. Rather, the low-level endemcity of the most ancient Beijing N-branch in North America demonstrates that it was brought to this continent from Eurasia with a small human group, an event that corresponds to the first entry of humans to this continent.

Summing up, it appears that most parsimoniously the Beijing genotype (specific deletion in the DR locus) (van Soolingen et al. 1995) might have originated in the human population of NRY K-M9 haplogroup that emerged in central Asia in humans migrating from the Middle East during a second out-of-Africa migration in the Upper Palaeolithic (Fig. 5A; Underhill et al. 2001; Y Chromosome Consortium 2002). This K-M9 human cluster diversified into several lineages that generally moved eastward to Siberia and east Asia (Underhill et al. 2001). The M45 mutation founded haplogroup P in the northeast direction, while the M214 mutation founded haplogroup O in the southeast direction in northern China (Fig. 5A; Underhill et al. 2001; Deng et al. 2004). The first major split within the Beijing genotype likely involved alteration in the NTF genome region and took place in these small founding (human) subpopulations. The ancestral Beijing lineage (intact NTF) must have remained in the human P-M45 population, which then moved further through Siberia, eventually reaching Beringia and entering North America prior to the Last Glacial Maximum (LGM), a relatively brief period at the end of the Upper Palaeolithic, 17,000–20,000 yr ago (Fig. 5B). Native Americans are known to have experienced two episodes of reduced population size: one with the peopling of the

Figure 5. Hypothesized chronology of the early evolution of the M. tuberculosis Beijing genotype superimposed onto human prehistoric migrations and NRY based phylogeography (Underhill et al. 2001). Arrows indicate human migrations. In bold are human haplogroups and mutations according to the NRY hierarchical nomenclature system (Y Chromosome Consortium 2002). Bf indicates Beijing genotype defined by specific signature of the DR locus (absence of spoligosnaps 1–34) (Bifani et al. 2002; Glynn et al. 2002); Kya, thousand years ago.
Americas and the other with European contact (Mulligan et al. 2004). Either or both might have contributed to the currently low-level endemicity of the initial Beijing N-branch in North America, and further studies are needed to elucidate this issue. It is noteworthy that two N-branch strains with intact NTF region were found in our collection of Russian Beijing strains and that Beijing strains, defined as ancestral by other markers, have previously been described, although as low-endemic, in modern Russia and the United States (Mokrousov et al. 2002; Rad et al. 2003). Thus, they may represent relic strains left on the first passage of the Beijing primordial sublineage (intact NTF) through Siberia to North America 20,000–30,000 yr ago.

The history of the second Beijing lineage marked by the first IS6110 insertion in the NTF locus (Fig. 2) may be related to the Neolithic agriculture revolution that occurred independently in different continents following rapid global warming and the retreat of glaciers at the end of LGM (Mithen 2004). One of the main derivatives of the K-M9 Asian cluster is haplogroup O, which achieved very high frequency in east Asia (Underhill et al. 2001). The haplogroup O-M214/M175 lineage seems to have originated in Northern China. Its distribution may reflect the impact of the millet and rice agriculture on east Asian demographic history (Cohen 1998; Shelach 2000; Underhill et al. 2001; Bettinger et al. 2005), displacing to a great extent all other NRY variants with a clinal frequency from the expected Chinese area of origin (Underhill et al. 2001). The Beijing NTF::IS6110 subpopulation was likely brought by K-M9/O-M214 humans to northern China ~20,000–30,000 yr ago. After the LGM, this ancient population (O-M214/M175/M122) who lived in the upper-middle Yellow River basin, expanded mirroring the radiation of the Proto-Sino-Tibetan languages (Su et al. 2000; Deng et al. 2004). This population expansion brought the Beijing NTF::IS6110 lineage to all areas of east Asia (Fig. 5C), whereas the already achieved divergence within the Beijing genotype was differentially transferred to new areas by distinct human subpopulations. The absence of the MIRU types shared by the Beijing strains from Vietnam and Bangladesh (Fig. 3) suggests that these areas may have been contaminated in the early Neolithic when human subpopulations were small enough to exert a founder effect of the particular Beijing variants (MIRU types) differentially carried by the host populations.

As we mentioned above, the \textit{M. tuberculosis} Beijing genotype is neither frequent nor endemic variant in Africa as a whole. The exception is South Africa, namely, the Cape Town area. The introduction of the Beijing genotype (NTF::IS6110 branch) to South Africa appears to have occurred relatively recently. Strictly speaking, migrations from China itself are not likely to have contributed to the genotype’s importation here. Currently, the Chinese human population constitutes only 30,000 individuals in South Africa (http://www.nationmaster.com/encyclopedia/South-Africa). Significant, but extremely transient, Chinese gene flow to South Africa occurred only 100 yr ago, when ~50,000 Chinese workers were imported for the Rand gold mines in 1903–1907 (http://61.191.encyclopedia.org/S/SO/SOUTH_AFRICA.htm). However, by 1910 they were all repatriated (http://61.191.encyclopedia.org/S/SO/SOUTH_AFRICA.htm), and their impact on the health situation in South Africa was likely to be negligible. Recently, van Helden et al. (2002) suggested that Beijing strains might have been introduced to South Africa following the sea trade route from east Asia to Europe that started 400 yr ago. Indeed, in the 17th and 18th centuries, Dutch colonists at the Cape of Good Hope largely imported slaves from Indonesia, Madagascar, Mozambique, and India. Descendants of these slaves, who often married with Dutch settlers, later became known as “Cape Coloreds” or “Cape Malays,” and present form the majority of the 4.7 million population of the Western Cape Province (http://www.nationmaster.com/encyclopedia/South-Africa). Therefore, it is likely that the Beijing strains were historically recently brought to South Africa, not directly from its primary focus of origin (China), but from the secondary one (Indonesia). The historical evidence is supported by genetic data: In the Beijing genetic network all main South African (Indonesian?) types are found at distant nodes and hence appear to be the youngest.

Unlike for South Africa, the exact timing of the first entry and primary dispersals of the Beijing strains in Northern Eurasia (modern Russia) is elusive. The Russian Beijing genotype population is the least diverse of all local Beijing populations. Assuming that more diversity results from longer clonal expansion, this may reflect a most recent dissemination of the currently circulating and locally predominant Beijing strains in Russia compared with the other areas studied. This is readily explained by the cold climate and, until very recently, the extremely low population density in a vast area of Russia and Siberia (Christian 1998) compared with the warmer conditions and the fast growing and denser population in east Asia and South Africa. Both network (Fig. 3) and PC (Fig. 4B) analyses suggest that Russia and South Africa were infected with distinct subsets of Beijing strains. However, a strong founder effect may have played a crucial role in the evolution of the Beijing genotype’s population structure in South Africa and would consequently generate the unordinately long branch in the NJ tree (Fig. 4A). Since the Beijing genotype is not a European endemic variant, the published PC analysis of European human populations allows us to rule out those migrations that equally concerned both Russia and Europe as sources of the Beijing strains. These are defined by Finno-Ugric (PC2) (Cavalli-Sforza 2000), Scythe (PC3) (Cavalli-Sforza 2001), and Hun (Christian 1998) expansions. We may further speculate that trade contacts as such, even long-lasting ones, are not sufficient for an effective dissemination of the \textit{M. tuberculosis} strains if they are not supported by a kind of demic diffusion of the strains’ carriers, manifested as population growth and migration. The Silk Road connected China with Europe for almost two millennia, 2 BC–1600 AD (Christian 1998), and this route may have been opened much earlier, based on the transfer of the first ceramics technology from Japan to the Middle East and Europe at the beginning of agricultural practice (Cavalli-Sforza 2001). However, it is appropriate to reiterate that Beijing strains are not identified as a European endemic variant.

Finally, we suggest the TB spread related to the Genghis (or Chinggis) Khan invasion to be more plausible. The Mongol empire of the 13th century brought the different parts of Eurasia closer than they had ever been before and created an economic and cultural system embracing much of the Eurasian land mass (Christian 1998). It was also a period of remarkable ethnic mixing since the Mongol army grew by incorporating the armies of many different nations that it had defeated, including Han Chinese (Christian 1998). McNeill (1976) suggested that Mongol invasions also unified Eurasia epidemiologically, allowing the exchange of the disease vectors throughout Eurasia. Genghis Khan did eventually come in the center of Europe, but for a short time. This was sufficient for the dissemination of \textit{Yersinia pestis} to occur, but not for that of the far less contagious \textit{M. tuberculosis}. Even if some \textit{M. tuberculosis} Beijing genotype strains had been
brought to Europe in this way, this may not have manifested rapidly. Subsequently, the Black Death that decimated European human populations could have efficiently eliminated rare carriers of the M. tuberculosis Beijing genotype. By contrast, further close interaction between Rus’ and Orda was prolonged for three centuries, and it may be possible that the Mongol invasion and the subsequent yoke/cohabitation were indeed the vehicle that brought M. tuberculosis Beijing genotype strains to Russia.

The identity of the immediate ancestor of the Beijing genotype remains open. The peculiar DR locus structure, actually, the Beijing genotype identification, is “abridged” and thereby prevents an unambiguous reconstruction of the DR locus of a “pre-Beijing” strain. A comparison of major extant genetic families within M. tuberculosis (such as Latin-American-Mediterranean, Haarlem, East African-Indian, Delhi, and Beijing) produced an SNP-based star-like tree (Baker et al. 2004), where the Beijing genotype is not derived from any of these types but is instead the subsequent yoke/cohabitation that were indeed the vehicle that brought M. tuberculosis Beijing genotype strains to Russia.

A comparison of major extant genetic families within M. tuberculosis (such as Latin-American-Mediterranean, Haarlem, East African-Indian, Delhi, and Beijing) produced an SNP-based star-like tree (Baker et al. 2004), where the Beijing genotype is not derived from any of these types but is instead the subsequent yoke/cohabitation that were indeed the vehicle that brought M. tuberculosis Beijing genotype strains to Russia.

To conclude, the implications from such comparative studies appear to be, at least, twofold: (1) timing specific events in the evolution of bacterial genomes as attempted here with one lineage of M. tuberculosis, and (2) tracing hidden patterns of human migrations of which recent studies on Helicobacter pylori (Falush et al. 2003) and polyomavirus JC (Pavesi 2004) are exciting examples. In the case of M. tuberculosis, a more detailed information on many genotypes within this species from many locations is still needed to address this latter issue. Perhaps, a comparison of phylogenies between M. tuberculosis strains and the Y chromosome haplotypes directly sampled from TB patients from non-urban isolated areas (less influenced by recent human migrations) could also give clues to better understanding of our coevolution.

Methods

Bacterial strains

M. tuberculosis strains from Russia (mainly northwest) and Vietnam were recovered from unlinked adult patients with pulmonary TB, 1997–2003. A total of 520 Russian and 125 Vietnamese strains were studied.

DNA fingerprinting

The DNA was isolated following the recommended method and subjected to IS6110-RFLP typing as described previously (van Embden et al. 1993). The variation in the DR locus (absence/presence of 43 different spacers) was studied by the spoligotyping macroarray-based method as described previously (Kamerbeek et al. 1997). In short, the PCR-amplified biotin-labeled DR locus is hybridized against an array of immobilized 43 different DR spacers. Resulting hybridization signals are revealed by chemiluminescence and are visualized as a profile of discrete dots (Fig. 1).

The 44 Russian and 43 Vietnamese strains with different IS6110-RFLP profiles and identified as Beijing genotype by spoligotyping (absence of signals 1–34) were selected for MIRU analysis performed as described by Supply et al. (2001) and Mokrousov et al. (2004). Search of the MIRU profiles of the Beijing strains from other studies was done by using the Entrez-PubMed and Google engines, followed by inspection of the retrieved articles for the presence of information on Beijing isolates. It yielded 226 more strains (Supply et al. 2003; Banu et al. 2004; Sun et al. 2004), the largest samples from South Africa (38 strains), Singapore (160 strains), and Bangladesh (15 strains). Information on strain sources and their MIRU profiles is provided in the Supplemental Table. We designated the Beijing genotype strains from Singapore as Chinese because the majority of the Singaporean population is of Chinese descent and TB mainly affects the elderly, many of whom are first- or second-generation immigrants from mainland China (Sun et al. 2004).

NTF locus analysis

A multiplex PCR approach was used to determine possible IS6110 insertion(s) in the NTF region of the M. tuberculosis strains. Primers located within the NTF region and IS6110 sequence as well as PCR parameters were previously described by Plikaytis et al. (1994).

Statistical analysis

The MIRU data were entered into an Excel spreadsheet, and the strains were subdivided into MIRU types with unique 12-loci profiles (see the Supplemental Table). A comparison was done for five areas: Russia, Singapore (i.e., China), South Africa, Vietnam, and Bangladesh. The Hunter-Gaston (Hunter and Gaston 1988) index was used as an estimate of the allelic diversity for specific areas, for the 12 loci taken together. HGDI is a probability that two strains consecutively taken from a given population would be placed into different types by the typing method; in other words, the lower the index value is, the less discriminative is the typing method (and less diverse is the population in a given locus/loci). The HGDI was calculated by using the following formula:

\[
\text{HGDI} = 1 - \left[ \frac{1}{N(N-1)} \sum_{j=1}^{s} n_j (n_j - 1) \right]
\]

where \( N \) is the total number of strains in the typing scheme, \( s \) is the total number of distinct patterns discriminated by the typing method, and \( n_j \) is the number of strains belonging to the \( j \)th pattern.

The most parsimonious network of the Beijing MIRU types was built by using the PARS routine of PHYLIP 3.6 package (Felsenstein 2004). PARS is a general parsimony program that carries out the Wagner parsimony method with multiple states. It assumes that different characters and different lineages evolve independently and changes to all other states are equally probable. This is applicable to evolution of MIRU-VNTR loci treated as categorical variables. The following conditions were applied to our analysis: (1) only types with two and more strains from the target areas were included, (2) only loci polymorphic in these shared types were included (eight loci), and (3) MIRU alleles were treated as categorical variables (i.e., any changes in copy number were assumed as equivalent). The bootstrapping procedure was not applied to test the robustness of the tree since there were too few number of variable characters (only eight MIRU loci).

Genetic distances between the geographical populations of the Beijing strains (Russia, Singapore [China], Vietnam, and South Africa) were calculated based on the MIRU shared types and eight polymorphic MIRU-VNTR loci as coancestry coefficients (F statistics or \( \Theta \) values) (Weir 1990) using formula imple-
mented in GDA software (Lewis and Zaykin 2001). The input Nexus file is available upon request. The resulting distance matrix was used to construct a NJ tree with GDA software and for PC analysis by means of multidimensional scaling with Permap software (Heady and Lucas 1997).

Acknowledgments

We thank Elena Limeschenko and Anna Vyazovaya for technical assistance and Lidia Steklova for providing us with some of the clinical isolates. We are grateful to the three anonymous reviewers whose invaluable comments and critiques helped us to significantly improve the manuscript.

Note added in proof

While this manuscript was in review, Kremer (2005) reported the atypical Beijing strains from Southeast Asia. This finding appears to disagree with the previous estimation of Bifani et al. (2002) and, in our opinion, supports the Chinese and more recent origin of the M. tuberculosis Beijing genotype.

References


Web site references

http://www.nationmaster.com/encyclopedia/South-Africa; Encyclopedia: South Africa
http://61.1911encyclopedia.org/S/SO/SOUTH_AFRICA.htm; South Africa

Received February 17, 2005; accepted in revised form July 12, 2005.
Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: Clues from human phylogeography

Igor Mokrousov, Ho Minh Ly, Tatiana Otten, et al.

*Genome Res.* 2005 15: 1357-1364
Access the most recent version at doi:10.1101/gr.3840605

---

**Supplemental Material**

http://genome.cshlp.org/content/suppl/2005/09/16/gr.3840605.DC1

**References**

This article cites 37 articles, 15 of which can be accessed free at:
http://genome.cshlp.org/content/15/10/1357.full.html#ref-list-1

**License**

---

**Email Alerting Service**

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

---

To subscribe to *Genome Research* go to:
http://genome.cshlp.org/subscriptions

---

Cold Spring Harbor Laboratory Press