

Accepted Manuscript

Turkish and Japanese *Mycobacterium tuberculosis* sublineages share a remote common ancestor

Guislaine Refregier, Edgar Abadia, Tomoshige Matsumoto, Hiromi Ano, Tetsuya Takashima, Izuo Tsuyuguchi, Elif Aktas, Füsün Cömert, Michel Kireopori Gomgnimbou, Stefan Panaiotov, Jody Phelan, Francesc Coll, Ruth Mcnerney, Arnab Pain, Taane G. Clark, Christophe Sola

PII: S1567-1348(16)30418-X
DOI: doi: [10.1016/j.meegid.2016.10.009](https://doi.org/10.1016/j.meegid.2016.10.009)
Reference: MEEGID 2949

To appear in:

Received date: 14 June 2016
Revised date: 23 September 2016
Accepted date: 12 October 2016

Please cite this article as: Refregier, Guislaine, Abadia, Edgar, Matsumoto, Tomoshige, Ano, Hiromi, Takashima, Tetsuya, Tsuyuguchi, Izuo, Aktas, Elif, Cömert, Füsün, Gomgnimbou, Michel Kireopori, Panaiotov, Stefan, Phelan, Jody, Coll, Francesc, Mcnerney, Ruth, Pain, Arnab, Clark, Taane G., Sola, Christophe, Turkish and Japanese *Mycobacterium tuberculosis* sublineages share a remote common ancestor, (2016), doi: [10.1016/j.meegid.2016.10.009](https://doi.org/10.1016/j.meegid.2016.10.009)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Turkish and Japanese *Mycobacterium tuberculosis* sublineages share a remote common ancestor

Guislaine REFREGIER^{a*}, Edgar ABADIA^b, Tomoshige MATSUMOTO^c, Hiromi ANO^c, Tetsuya TAKASHIMA^c, Izuo TSUYUGUCHI^c, Elif AKTAS^d, Füsün CÖMERT^e, Michel Kireopori GOMGNIMBOU^{a#}, Stefan PANAIOTOV^f, Jody PHELAN^g, Francesc COLL^g, Ruth MCNERNEY^{g,h}, Arnab PAINⁱ, Taane G. CLARK^g, Christophe SOLA^a.

^aInstitute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ Paris-Sud, Université Paris-Saclay, 91198, Gif-sur-Yvette cedex, France; ^bInstituto Venezolano de Investigaciones Científicas (IVIC), Caracas, Venezuela; ^cDepartment of Clinical Research and Development, Osaka Prefectural Hospital Organization, Osaka Prefectural Medical Center for Respiratory and Allergic Diseases, Habikino-city, Japan^d; Şişli Etfal Research and Training Hospital, Istanbul, Turkey; ^eFaculty of Medicine, Bülent Ecevit University, Zonguldak, Turkey; ^fNational Center of Parasitic and Infectious Diseases, Sofia, Bulgaria; ^gFaculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK; ^hInfection and Immunity Unit, UCT Lung Institute, University of Cape Town, Old Main Building, Groote Schuur Hospital, Cape Town, South Africa; ⁱPathogen Genomics Group, Biological, Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

Running title: Linking Turkish and Japanese tuberculosis

#Present address: Centre MURAZ, Bobo Dioulasso, Burkina Faso

*corresponding author : Guislaine REFREGIER, Institut de Biologie Intégrative de la Cellule, UMR9198, CEA-CNRS-Université Paris-Saclay, F-91405, Orsay-Cedex, Ph: +33 (0) 1 69 15 46 48, Fax : +33 (0) 1 69 15 66 78, guislaine.refregier@u-psud.fr

Abstract

Two geographically distant *M. tuberculosis* sublineages, Tur from Turkey and T3-Osaka from Japan, exhibit partially identical genotypic signatures (identical 12-loci MIRU-VNTR profiles, distinct spoligotyping patterns).

We investigated T3-Osaka and Tur sublineages characteristics and potential genetic relatedness, first using MIRU-VNTR locus analysis on 21 and 25 samples of each sublineage respectively, and second comparing Whole Genome Sequences of 8 new samples to public data from 45 samples uncovering human tuberculosis diversity. We then tried to date their Most Recent Common Ancestor (MRCA) using three calibrations of SNP accumulation rate (long-term=0.03 SNP/genome/year, derived from a tuberculosis ancestor of around 70,000 years old; intermediate=0.2 SNP/genome/year derived from a Peruvian mummy; short-term=0.5 SNP/genome/year). To disentangle between these scenarios, we confronted the corresponding divergence times with major human history events and knowledge on human genetic divergence.

We identified relatively high intrasublineage diversity for both T3-Osaka and Tur. We definitively proved their monophyly; the corresponding super-sublineage (referred to as “T3-Osa-Tur”) shares a common ancestor with T3-Ethiopia and Ural sublineages but is only remotely related to other Euro-American sublineages such as X, LAM, Haarlem and S.

The evolutionary scenario based on long-term evolution rate being valid until T3-Osa-Tur MRCA was not supported by Japanese fossil data. The evolutionary scenario relying on short-term evolution rate since T3-Osa-Tur MRCA was contradicted by human history and potential traces of past epidemics. T3-Osaka and Tur sublineages were found likely to have diverged between 800y and 2,000 years ago, potentially at the time of Mongol Empire.

Altogether, this study definitively proves a strong genetic link between Turkish and Japanese tuberculosis. It provides a first hypothesis for calibrating TB Euro-American lineage

molecular clock; additional studies are needed to reliably date events corresponding to intermediate depths in tuberculosis phylogeny.

283 words.

ACCEPTED MANUSCRIPT

1. Introduction

Tuberculosis (TB), caused by bacteria in the *Mycobacterium tuberculosis* complex (MTBC), is the second most prevalent infectious disease, with about 8.6 million cases every year. It is an ancient airborne infectious disease affecting both humans and domestic animals (Grmek, 1994; Palfi et al., 2015; WHO, 2013). Seven lineages have been identified up to now: three ‘ancient’ *i.e.* having likely diverged rapidly after the Most Recent Common Ancestor (MRCA): lineage 1-EAI and two *M. africanum* lineages termed 5 and 6; three ‘modern’ (lineages 2, 3, 4), and one intermediate (lineage 7) recently described in Ethiopia (Blouin et al., 2012; Firdessa et al., 2013). Dating the emergence of the MTBC and main diversification events is both of fundamental and applied interest: it is a necessary step for identifying correlation between ecological conditions and tuberculosis evolution, which can in turn provide better predictions of outbreak risks.

1.1. Long-term substitution rate differs from short-term mutation rate

Dating past evolutionary events is a difficult task. For macro-organisms, a comfortable but rarely existing situation is to have a series of dated fossils sharing distinctive phenotypes (Ho et al., 2011). For all other situations, common practice is to infer the evolution rate of one molecule and apply it to phylogenies linking present samples (Perler et al., 1980). Inferring DNA evolution rate can be achieved *indirectly* using a dated fossil, or *directly* in pedigrees, laboratory lines, or through heterochronous sampling (Ho et al., 2011). The two methods in fact give different results both in macro- and micro-organisms: the indirect method integrates evolutionary mechanisms related to demography and environment such as random sorting and selection, which impacts depend on the gene and on the position of the mutation in the codon (Kuo and Ochman, 2009; Parsons et al., 1997). The long-term evolutionary rate is classically referred to as “substitution rate”. Substitution rate can be as much as 10 times lower than the

short-term evolution rate depending on the species and its population dynamics (Ho et al., 2011; Howell et al., 2003; Soares et al., 2009).

1.2. Current knowledge on tuberculosis evolution rate

Short-term evolutionary rate in *M. tuberculosis* complex has been estimated to be between 0.3 and 0.5 punctual mutations per genome per year (also referred to as Single Nucleotide Polymorphism or SNP per genome per year) which corresponds to an annual rate per site of 6.8×10^{-8} to 1.1×10^{-7} (Bryant et al., 2013; Ford et al., 2011; Guerra-Assuncao et al., 2015; Roetzer et al., 2013; Walker et al., 2013). Mutation rate was also assessed directly using whole genome sequence of a tuberculosis sample extracted from a 1,000-year old Peruvian mummy: it was measured as 4.8×10^{-8} SNP/site/year *i.e.* 0.22 SNP/genome/year (Bos et al., 2014).

To infer long-term evolutionary rate of MTBC, several fossils may be considered. Turkish *Homo erectus* remains showing bone lesions partly similar to those caused by MTBC suggest that a TB ancestor could have infected human lineage at least 0.5 million years ago. No molecular data however are available to support presence of MTBC DNA in these remains. The corresponding mutation rate, in the range of 10^{-10} per site per year, would be surprisingly low as compared to that of other bacteria (Comas et al., 2013; Kappelman et al., 2008; Roberts et al., 2009). Detection of *M. tuberculosis* ancient DNA has occurred in more recent fossils: molecular analyses of palaeolithic bison bone remains showed a case of infection 17,000 years ago in North America; several fossils dating from more than 7,000 years ago are known since early 20th century in Eastern Europe and Middle East: the most ancient are 9,000 years old and were found in Atlit Yam in Israël; they harbor genetic features suggesting that they belong to a “modern” lineage (Hershkovitz et al., 2008; Masson et al., 2015); others in Italy, Hungary date back to 7,000 years ago (Bartels, 1907).

Notwithstanding direct estimation of mutation rate using fossils, long-term *M. tuberculosis* evolution rate has been estimated using the assumption of host codivergence (Comas et al., 2013; Kuo and Ochman, 2009). From the three hypotheses tested by Comas et al. (2013), namely 1) TB divergence upon emergence of anatomically modern humans circa -185,000 years ago, 2) divergence coincident with the out-of-Africa expansion circa -70,000 years ago, and 3) divergence during the Neolithic transition circa -12,000 years ago, the -70,000 years ago hypothesis provided the best match with human paths of migration: the first modern human migration towards East Asia coincides with the emergence of the East-African-Indian (EAI) clade (lineage 1). In addition, the increases of the effective population size derived from present diversity were found contemporaneous to important events in human history: a first tuberculosis expansion around -30/-20,000 years ago, when human populations first rose to an effective population size close to 1,000 individuals, and a second pronounced increase between -8 and -4 kya in a period of great human expansion associated with agriculture onset (Comas et al., 2013). The corresponding substitution rate was described to be 2.58×10^{-9} substitutions per site per year; it matches that derived from enterobacteria divergence as used for first inferences on the age of MTBC MRCA (Gutierrez et al., 2005; Hughes et al., 2002).

Accepted estimates of *M. tuberculosis* annual mutation rate therefore range from 2.58×10^{-9} for the long-term to approximately 10^{-7} for the short term. If these estimates are true, the question becomes how and when evolution rate switches from one rate to the other.

Two studies have tried to infer intermediate evolutionary events using either long-term or short-term mutation rates. Using the long-term evolution rate that they had derived (-70,000 years hypothesis), two expansions have been described in the Beijing lineage (lineage 2) as dating back to respectively 1) between -10,000 and -6,000 years ago and 2) between -5,000 to -3,000 years ago, i.e. at the time of human Neolithic expansions. This matches

known human history and potentially further supports expansion -70,000 years ago (Comas et al., 2013). Accordingly, long-term substitution rate seems valid until the divergence of the so-called “modern lineages” (lineages 2, 3 and 4) expansions. Another study focusing on the Beijing lineage argues that the implementation of a short-term mutation rate of 10^{-7} SNP/site/year (i.e. 38.8 times faster mutation rate) leads to a convincing evolutionary history from the demographic explosion of Beijing lineage on, as Beijing expansion would coincide with human demography (Merker et al., 2015): first expansion upon the onset of industrial revolution in the 19th century (1822-1843) i.e. around 180 years ago, stasis linked with increased hygiene on the turn of 20th century, second burst associated with First World War or slightly before (1896-1916), around 110 years ago, decline associated with antibiotic era in 1966, and reactivation linked with HIV starting around 1987. This implies that the short-term evolutionary rate would be valid at least for the last 180 years. However when applying the short-term rate for a period of 180 years, the number of mutations is the same as that derived from the long-term substitution rate for ~7,000 years ago, i.e. 80/90 SNPs. This observation means that the same first molecular signal has been interpreted as an expansion dating back to 7,000 or 180 years ago. Conversely, the same second signal (55 SNPs divergence) was interpreted dating back to either ~3/4,000 or ~110 years ago. As the same genetic divergence cannot account for expansions at different time points, the contradiction between these two analyses can indicate either that one of two applied rate does not apply for a range of 50/100 SNPs or that none of these rates can apply. Correspondence between inferred tuberculosis demography and known human demography could have arisen solely by chance as both *M. tuberculosis* and human demography seem to have followed a recurrent pattern of expansions and stases. Additional studies are needed to propose new calibration points and thereby build a more complete and accurate model of tuberculosis history.

1.3. Definition and general characteristics of Tur and T3-Osaka sublineages

Our attention was caught by the potential relatedness between two MTC sublineages, the Turkish and the T3-Osaka. These two sublineages were first defined based on standard genotyping methods: Insertion sequence-based polymorphism (IS6110-RFLP), combined mini-satellites analysis (12-loci MIRU-VNTR¹), profile at the CRISPR² locus (spoligotype) with specific deletions (Fig. 1).

The Turkish (Tur) sublineage was first described in Turkey where it represents 10 to 30% of all clinical isolates depending on regions, and named as “LAM7-Tur” in 2005 (Durmaz et al., 2007; Kisa et al., 2012; Oral Zeytinli and Koksak, 2012; Otlu et al., 2009; Zozio et al., 2005). The absence of phylogenetic relationship with other LAM sublineages such as the absence of *ligB* synonymous mutation C1212G led to its progressive renaming as Tur (Abadia et al., 2010; Dos Vultos et al., 2008). This lineage has been identified in several neighboring countries such as Bulgaria (Valcheva et al., 2008; Vasileva, 1992), and is recently spreading to Middle-East and Europe as evidenced by SIT41 map generated by SITVITWEB (Fig. S1A). Tur prototypic genotype carries two IS6110 copies (2.1 and 4.8 kb), and a characteristic 777777404760771 spoligotype signature lacking spacers 20-24, 26-27, 33-36 (SIT41 in SpolDB4) (Fig. 1A-B). Most isolates carry a 215125113322 12-loci MIRU signature defined as MIRU-international-type 310 (MIT310) in SITVIT2 database (Zozio et al., 2005) (Fig. 1C). Isolates from the Tur sublineage have recently acquired Multi-Drug-Resistance (MDR) and this sublineage is one of the most prevalent among MDR-TB isolates in Bulgaria (Gomgnimbou et al., 2013).

The T3-Osaka sublineage was reported for the first time in Okayama district and nearby in Osaka in South-Japan in 2004 (Ohata and Tada, 2004). It was fully described by Iwamoto and coworkers as carrying one to two IS6110 copies (Takashima and Iwamoto,

¹ Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats

² Clustered Regularly Interspersed Short Palindromic Repeats

2006); its prototypic spoligotype is characterized by the absence of spacers 5-8, 13, 33-36 (such as in 741777777760771 spoligotype, SIT627 in SpolDB4), and most isolates also carry a MIT310 signature (Fig. 1C). Variants of the SIT627 spoligotype pattern in the SITVIT2 database (Demay et al., 2012) are almost only isolated from Japanese patients as shown in Fig. S1B. The SIT627 was also found in Kobe in an independent study (C. Sola & T. Iwamoto, unpublished). In Japan, the SIT627 is the most prevalent non-Beijing spoligotype pattern; it represents around 5% of clinical isolates around Okayama (Ohata and Tada, 2004), and recent epidemiological studies show that it is even more frequent in Yamagata (Fig. S1C). T3-Osaka VNTR diversity showing high frequency of MIT310 was studied in Okinawa in 2007 (Millet et al., 2007).

Both Tur and T3-Osaka sublineages are “low IS6110 copy” (1 or 2 copies), harbor most frequently a MIT310 genotype (215125113322), but divergent spoligotypes (although both lack spacers 33 to 36, which characterizes the Lineage 4 also known as “Euro-American”). Two isolates of each sublineage were later found to harbour the same specific G809T mutation in the DNA repair replication and recombination (3R) gene *ligC* (Abadia et al., 2010; Dos Vultos et al., 2008). In contrast, five other 3R SNPs (on *recC*, *end*, *recN*, *ruvB*, *dinF*) distinguished between T3-Osaka and Tur (Dos Vultos et al., 2008). In a method development study, *ligC* G809T mutation was detected in 18 additional Tur isolates and 25 additional SIT627 isolates and in none of the other isolates included in that study, suggesting a phylogenetic link between Tur and T3-Osaka sublineages (Abadia et al., 2010).

1.4. Objectives of the study

We aimed here at proving the relatedness of these two sublineages, and if confirmed, at using it to calibrate in *M. tuberculosis* divergence rate. We first explored the intra sublineage diversity using 46 samples and investigating VNTR profiles on markers outside

the MIRU-VNTR12 panel. We then obtained whole genome sequences for eight isolates (n=3 for T3-Osaka; n=3 for Tur; n=2 for outgroups) and analyzed them together with other publicly available sequences. Our results definitively confirm that these sublineages are highly related and emerged relatively long-time ago. We then explored different hypotheses regarding the age of their MRCA and compared them to data on tuberculous fossils and major events regarding human history and demography. Regarding *M. tuberculosis* molecular clock, we aimed at answering the following questions: 1) may long-term substitution rate be valid from 70,000 years ago until Tur and T3-Osaka MRCA? ; 2) may short-term substitution rate be valid since Tur and T3-Osaka MRCA? ; 3) if an intermediate substitution rate is needed to account for Tur and T3-Osaka MRCA likely age, what is its range *i.e.* what is the shape of the function describing substitution rate decrease with longer time periods?

2. Results

2.1. MIRU-VNTR diversity among Tur and T3-Osaka sublineages

In order to characterize intra and inter-sublineages diversity, we genotyped a total of 46 isolates characterized as Tur (n=21) or T3-Osaka (n=25). We used 9 VNTRs out of MIRU₁₂ panel and known to be variable between Euro-American isolates: Mtub04, Mtub21, Mtub29, Mtub30, Mtub39, Qub11b, Qub26, Qub4156 (Table 1). Only Mtub04 presented no common allele between the T3-Osaka and the Tur strains and thus might serve as a surrogate marker to discriminate between the two sublineages (4 or 5 repetitions for Tur; 2 or 3 repetitions for T3-Osaka). Variants were relatively frequent, with up to four locus variations (4LV) among the Turkish Tur isolates at the 9 targeted VNTRs, and up to 5LV when comparing them with Bulgarian Tur. T3-Osaka isolates also harboured up to 5LV for the 9 targeted VNTRs and 6

when including others VNTRs (Table S1). Bulgarian Tur isolates harboured specific genotypes suggesting that they could form a sublineage inside Tur (Table 1). For the 25 T3-Osaka samples and the 3 Tur samples from Bulgaria, for which we checked the whole 24 MIRU-VNTR patterns, only 13 samples (46%) carried the expected MIT310 genotype and six of the 12 MIRUs carried at least one different allele other than that present in MIT310 (Table S1). MIRU10 was the most variable locus in this sample, five out of the 25 T3-Osaka samples having from 1 to 4 repetitions instead of 5. Consensus 12-loci MIRU signature is 21X1251X3X2X, with ETRA-B-C signature of “324” and Mtub(04-21-29-30-34-39) signature of X3443X. We searched for such signature along recent data gathered from China and found no such isolate (data not shown).

2.2. WGS sequencing and phylogenetic analysis

To confirm relatedness and position these two sublineages in *M. tuberculosis* phylogeny, we proceeded to whole genome sequencing of prototypic specimen. Three isolates of each sublineage (Tur and T3-Osaka) were selected, respectively from a Bulgarian and a Japanese collection. One other isolate in each of the two populations was included as representative of the local diversity. These eight genomes were compared to 45 additional genomes representing the diversity of standard MTBC excluding *M. canettii* (Benavente et al., 2015; Coll et al., 2014a; Coll et al., 2014b; Firdessa et al., 2013) (Table S2). We specifically included an Ural strain to test the hypothesis of relatedness between Tur and Ural sublineages proposed by Mokrousov (2015). A total of 14,071 variable positions were identified among which 14,013 SNPs (Table S3).

Phylogenetic analysis clustered the “Tur” and “T3-Osaka” isolates respectively (Fig. 2). Interestingly, the three “T3-Osaka” isolates diverged relatively long ago (long terminal branches) and at the same time. Such feature is rare in phylogenies, especially when sampling

is random. It may indicate a time of intense proliferation. In the case of tuberculosis, it might correspond to an epidemic.

“Tur” and “T3-Osaka” sublineages clustered together with high confidence (100% as estimated by 500 bootstrap replications; sublineage denoted as “T3-Osa-Tur”). The phylogenetic analysis shows that the “T3-Osa-Tur” sublineage is related to a “T3-Ethiopia” sample collected in the Netherlands and an “Ural” sample collected in Germany (Fig. 2). We identified 74 SNPs potentially specific of “T3-Osaka” sublineage, 96 SNPs potentially specific of the Tur lineage and 7 SNPs specific of the “T3-Osa-Tur” sublineage (Table 2). In addition, a 2bp small insertion in Rv0577 was found in all “T3-Osa-Tur” isolates (Table S4). The lineage 4.2.2.1-specific SNP in *prfA* gene (Rv1299, T→C in position 1,455,780) described by Coll et al. (Coll et al., 2014a; Coll et al., 2014b) is one of the SNPs found to be common to the Tur isolates: this 4.2.2.1 lineage may thus be considered as a synonym of “Tur” sublineage. Within the SNPs list, we confirmed 2 out of 5 candidate SNPs previously identified (Dos Vultos et al., 2008): *ruvB* was found specific of T3-Osaka isolates and *recN* specific of Tur isolates. In contrast, SNPs in *recC*, *end*, *dinF* were absent from any isolate included in this study. Among the SNPs common to the three “T3-Osaka” isolates, as the divergence between them is quite large (see branch length in Fig. 2) most of them may be characteristic of “T3-Osaka” lineage. We propose C→G in genome position 7,121 in *gyrB* gene to define this sublineage.

2.3. Dating the split between Turkish and T3-Osaka sub-lineages according to different available mutation rates

We performed BEAST analyses to infer divergence dates using three mutation rates: one corresponding to the MRCA of *M. tuberculosis* tree of 67,000 years old according to congruence with human evolution (Comas et al., 2013), the second one corresponding to the

mutation rate derived from the Peruvian mummy *i.e.* 4.8×10^{-8} mutation/site/year (17), the third being in the range of what was observed in clinical studies (10^{-7} mutation/site/year). Under the first scenario, the timing of lineages 1, 2, 3 and 4 divergence derived from our tree correlated well with the previously reported ages reported (Fig. S5 in Comas *et al.* 2013). The MRCA reported for the Japanese (T3-Osaka) and Turkish (Tur) was estimated to be approximately 12,000 years old (95% CI 10,907-12,753 years) and the divergence of T3-Osaka isolates potentially corresponding to an epidemic around 7,500 years old (Fig. 3A and Fig. S2). Under the second scenario coined as “intermediate”, we came out with a T3-Osa-Tur MRCA of around 800 years old, *i.e.* dated to 1200 A.D and potential Japanese epidemic 500 years ago (Fig. 3B and Fig. S3). The last scenario (“fast”) proposes a T3-Osa-Tur MRCA of around 370 years old and Japanese epidemic 230 years ago *i.e.* around 1780 AD (Fig. 3C and Fig. S4).

2.4. Confronting human and tuberculosis information to assess the time of transition between slow and rapid mutation rate

Fossil, human genetic and linguistic evidence was used to investigate the three scenarios of divergence. We searched for events: A) in Middle East and over the world possibly corresponding to emergence of “modern” lineages (lineages 2-4); B) in Central Asia possibly corresponding to colonization of Asia by TB; C) in Japan possibly corresponding to colonization of the country by T3-Osaka TB sublineage and later potential epidemic; D) in Turkey possibly corresponding to colonization of the country by Tur TB sublineage. As intermediate evolution rate differs from only a factor 2 from the fast (short-term) evolution rate, Table 3 only reports the slow and intermediate rate scenarios. We found clues compatible with any scenario at all time scales, however, two lines of evidence suggested that some scenarios are unlikely for certain time scales (Table 3). First, intermediate and fast

evolution rates were found poorly supported for the timeframe of Asia colonization by TB-infected humans. Indeed only so-called “modern lineages” are found in this part of the world suggesting that they colonized Asia after “modern lineage” emergence; under the fast and intermediate evolution rate scenarios, “modern lineages” emerged around 4,000 years ago, which would mean that Asia would have been colonized by tuberculosis afterwards. This is not compatible with human remains with scars of TB found in Baikal region and dating back to 4,500 years ago (Lieverse et al., 2014). Second, slow evolution rate scenario does not match field data linked to TB colonization of Japan: this scenario infers that MRCA of T3-Osaka in Japan and the epidemics likely responsible of almost contemporaneous divergence of all 3 present isolates, would be approximately 7,500 years old (CI [8,300-6,600 years old], Fig. S2), however no human remains from the Jomon period (12,000-1,700 years ago) were found to have TB scars despite examination of more than 10,000 skeletons (Suzuki et al., 2008). We suggest that TB proliferation having left no traces among this large sample of remains is unlikely. War during Sengoku period (15th – 16th century) may in contrast have favored an epidemic (possibly 540-440 years old *i.e.* dating back to 1470-1570 years AD) as per intermediate mutation rate scenario (Table 3, Fig. S3). Such timing seems also more likely than that according to fast-evolution rate scenario (Fig. 3) as the potential epidemic is inferred to have occurred between 260 and 210 years ago *i.e.* between 1750-1800 AD period, in the middle of Edo-Tokugawa period renowned for prosperity and peace (Table 3, Fig. S4). Altogether, possible history of tuberculosis at the different time scales is: divergence of the 7 main lineages more than 20,000 years ago in Africa and in the Middle East according to slow mutation rate scenario; divergence between T3-Osaka and Tur and subsequent dispersal towards Japan and Turkey, and Japanese epidemic, respectively less than 2,000 and 1,000 years ago, possibly at the time corresponding to intermediate evolution rate scenario (*i.e.* ~800 and ~400 years ago) (Fig. 4).

4. Discussion

Using SNP, MIRU-VNTR data and whole genome sequencing, we investigated one of the rare cases of incongruence between MIRU-VNTR and spoligotyping results in *M. tuberculosis* diversity (Azé et al., 2015) and we tried to estimate their time of divergence. These sublineages are Turkish (Tur) and Japanese T3-Osaka (referred to as T3-Osaka) Euro-American tuberculosis isolates carrying a low number of IS6110 copies, and belonging to the sublineage 4.2 (Coll et al., 2014a). We definitively proved a strong phylogenetic link between Tur and T3-Osaka isolates and name the corresponding lineage “T3-Osa-Tur”. We identified that Tur and T3-Osaka sublineages show a relatively high level of diversity, suggesting that their divergence is truly ancient and not the trace of contemporaneous migrations. The Tur sublineage diversity may even have been underestimated as included Tur samples came from a limited number of countries where it is present, and in Turkey, from a single city (Zonguldak). The MRCA of the T3-Osa-Tur lineage likely harboured a 12-loci MIRU pattern concordant with that of the consensus found for both sublineages: 21X1251X3X2X, and potentially the most frequent genotype *i.e.* MIT310, 215125113322. Its spoligotype pattern must have exhibited no deletion specific to one of these sublineages, *i. e.* to have simply carried 33-36 spacers deletion (SIT53).

We inferred potential ages for all MRCAs for T3-Osa and Tur sublineages along *M. tuberculosis* phylogeny using the different molecular clocks described in the literature (Bos et al., 2014; Bryant et al., 2013; Comas et al., 2013; Ford et al., 2011; Merker et al., 2015). Specific human fossils and events in human history allowed us to consider that slow evolution rate may not apply for the time of T3-Osaka and Tur divergence. Possible contemporaneous divergence between the three Japanese samples (as derived from the inferred phylogeny)

suggests they can have diverged at a time of intense proliferation of the sublineage. If this signal truly is related to a TB epidemic, the timing of this epidemic under a fast-evolution rate scenario is unlikely: epidemic would have occurred at a time of renowned peace and prosperity. Fast evolution rate thus does not seem to apply for 330-200 SNPs divergence. According to the evolution rate measured in the 1,000 years old Peruvian Mummy (Bos et al., 2014), we hypothesize that T3-Osaka and Tur sublineages diverged around the beginning of second Millenary AD (1,000 years ago), and that T3-Osaka proliferated in Japan during the 15th or 16th century (400-500 years ago). However, larger sampling would be necessary to confirm T3-Osaka proliferation of samples with ~200 SNPs divergence.

We will first discuss the additional historical data compatible with tuberculosis phylogeny of provided by the history of Japan, of Turkey for the time of divergence between T3-Osaka and Tur sublineages. We then discuss others data concordant with a genetic link between Turkish and Japanese human populations. Last, we discuss the likely mutation rate corresponding to T3-Osaka and Tur divergence point and global evolution of mutation rate through time.

4.1. Japanese history and TB colonization of Japan

Anthropological data suggest that the Japanese population structure lies on two pillars: The Jomon and Yayoi cultures (Hammer et al., 2006; Nakagome et al., 2015). Jomon people would have remained isolated in the Japanese archipelago since circa -16,000 years ago when they diverged from Tibet, Korean or North-East Asian populations whereas Yayoi are assumed to share a common ancestor with Chinese dating back to 5,000 years ago (Hammer et al., 2006; Nakagome et al., 2015; Tajima et al., 2004). In the extreme North (Ainu) and South (Okinawa and Tyuku), current Japanese populations are mostly from Jomon ancestry while in central regions, they are mostly from Yayoi ancestry (Hammer et al., 2006).

Subsequent migrations from diverse regions, including Korea, have occurred; potential admixture from Mongol populations during Mongol empire has been detected (Hellenthal et al., 2014). Beijing tuberculosis lineage has been proposed to have entered Japan during Yayoi period (Mokrousov, 2008; Suzuki and Inoue, 2007). T3-Osaka lineage could have been imported by a small North-East Asian population afterwards. The Sengoku period (15th-late 16th century, 400-500 years ago) characterized by frequent wars with Korea, China and inside Japan may have favoured some tuberculosis epidemics. However, earlier periods also have experienced wars. T3-Osaka sublineage proliferation and settlement in the country may altogether date from Sengoku or an earlier period.

4.2. Turkish history and emergence of TB Tur sublineage

Turkey is a mosaic of people descending mainly from nomadic tribes, some of which still have some representatives in Anatolian mountains. A megalithic at Urfa, East of Turkey, called Göbekli Tepe, confirms hunter-gatherer settlements from 13,000 to 10,000 years old but colonization by modern humans is inferred to have occurred 50,000 years ago from Middle East (Henn et al., 2012). Present Turkish populations show a strong component of “Eastern components” such as so-called D-cluster in mitochondrial diversity analyses (Comas et al., 1996). This admixture may have occurred at several time points, a stronger signal dating from 1,300 A.D. *i.e.* 700 years ago (Hellenthal et al., 2014). Historical events possibly having caused admixture from Asian populations into current Turkish ones are: 1) migration towards Mesopotamia 5,000 years ago of Yamanyan herders, descendants of Caucasus Hunter gatherers inhabiting North Asia since 25,000 years ago (Jones et al., 2015); 2) intense exchanges associated with first organized Mesopotamian civilization having set up first law code (Hammurabi code, 1,754 BCE *i.e.* 2,750 years ago) inducing migrations between Mesopotamia and all neighboring countries; 3) intra-empire migrations in the Hittite empire

set up circa 2,200 years ago, the empire encompassing present Turkey, Syria and Northern Mesopotamia (Kaniewski et al., 2013); 4) intra-empire migrations in the Roman Empire, associated with wars and colonization from 2,200 to 1,500 years ago; 5) Expansion of Turkic empire arisen in Mongolia around the 5th century AD *i.e.* 1,500 years ago; 6) migrations caused by wars led by Mongol Empire around 600-800 years ago (Sinor, 1990). All these migrations events may have brought different lineages of tuberculosis to Turkey. Import of Tur lineage may thus in theory be due to migrations other than that of Mongols. Still, scenario of import by Mongols both matches the intermediate mutation rate found in *M. pinnipedii* and events in both Japanese and in Turkish history. Under this hypothesis, T3-Osa-Tur lineage would have emerged in Mongolia around 1,200 AD (800 years ago) almost concomitantly with T3_Ethiopia (Fig. 3A), and diverged around 1,250 AD (750 years ago) into T3-Osa and Tur sublineages (Fig. 4).

4.3. Japanese-Turkish genetic and cultural relatedness

Direct relatedness between tuberculosis sublineages may be due to host genetic relatedness and vertical transmission of the pathogen. Direct relatedness between Japanese and Turkish human populations is scarce. To our knowledge, only one common genetic pattern associated with schizophrenia has been described (Ozcan, 2006). Some indirect genetic relationship has been identified, chromosomal tracts being shared between Turkish and Mongol populations on one hand (Hellenthal et al., 2014; Yunusbayev et al., 2015) and some being shared by Japanese and Mongols on the other hand (Hellenthal et al., 2014). Common ancestry of Turkish and Japanese human subpopulations is thus possible. This common host ancestry would have transmitted the same tuberculosis sublineage, in that case, T3-Osa-Tur descendants.

Another possible cause for pathogen common ancestry is horizontal transmission due to cultural relationship between populations having later migrated to Japan and Turkey respectively. Japanese and Turkic languages show similarities grouping them into “Ural-altaic languages” (Cavalli-Sforza, 2000; Robbeets, 2004; Robbeets, 2008). Ural-altaic languages supposedly originated in South Siberia around 10,000 years ago, whereas Indo-European languages are believed to have arisen in Anatolia (Present Turkey) around 7,116 to 10,410 years ago, contemporaneous to the spread of agriculture. Indo-European language was replaced by Turkic in Turkey starting 1600 years ago, during the 5th century AD (Yunusbayev et al., 2015). Turkic culture expansion could have promoted the dispersal to the West of tuberculosis sublineages such as Tur.

On the Japanese side, linguistics suggests that Okhotsk people (hunter-gatherer) may have migrated from the Eastern China Amur river basin to northern Hokkaido around 900-1,600 years ago bringing significant contributions to the preexisting Jomon culture (Lee and Hasegawa, 2013). These migrations may have helped tuberculosis sublineages spread, and possibly that of T3-Osaka ancestor. These links, which are anterior to Mongol Empire invoked above, slightly predate T3-Osaka-Tur MRCA derived from intermediate evolutionary rate. They thus present an alternative hypothesis linking Japanese and Turkish tuberculosis during first Millenary AD.

Altogether, Central Asia, and more precisely Mongolia, has been connected to Japan and Turkey at several points of human history. Mongolia could be the place of birth of T3-Osa-Tur lineage. Further investigations on MIT310 variants in this part of the world should shed more light on this question. The fact that no sublineage related to T3-Osaka has been reported in neighboring Asian countries is surprising. However, islands are known to provide isolation that can help rare genotypes to thrive.

4.4. Molecular evolution rate of tuberculosis

We have presented arguments suggesting that a slow molecular clock should be applied to infer ancient TB history, such as emergence of modern lineages more than 30,000 years ago (Comas et al., 2013). As our samples exhibited around 1,900 SNPs divergence, the corresponding mutation rate is 0.03 SNP/genome/year *i.e.* 6.2×10^{-9} SNP/site/year. This is slightly higher than the 2.9×10^{-9} SNP/site/year rate reported by Comas *et al.* (Comas et al., 2013). This difference likely is due to differences in SNPs filtering when handling raw sequencing data.

At the time scale of T3-Osa-Tur emergence (400 SNPs divergence, possibly corresponding ~1,000 years ago), our preferred scenario is characterized by an “intermediate” mutation rate of 4.8×10^{-8} SNP/site/year *i.e.* around 20-times faster. The “intermediate” mutation rate was derived from the animal sublineage *M. pinnipedii* that had infected a 1,000 y-old Peruvian man preserved by mummification (Bos et al., 2014). The fact that it provided a convincing scenario for T3-Osa-Tur divergence may be due: 1) either to true conservation of the substitution rate across tuberculosis lineages whatever the host, 2) or to local convergence in the evolution rates in T3-Osa-Tur human sublineage and animal lineages (human tuberculosis populations evolved under limited selective pressure may accumulate similar mutations as animal bacilli evolving under positive selection associated to host switches); 3) or to a mere artefact: the sublineage of interest could have diverged at a different time point and we simply did not find evidence contradicting this scenario.

We concluded that the same substitution rate cannot be applied at all time scales of tuberculosis phylogeny and that 1) Short mutation rate applies at least for the last ten years BP *i.e.* for a distance of around 10 SNPs, 2) Slow evolution rate applies at least around 70ky BP *i.e.* for a distance of around 1900 SNPs, and 3) intermediate mutation rate of 10^{-7} SNP/site/year may apply for T3-Osa-Tur divergence which corresponds to a mean difference

of 330 SNPs in our sample. When plotting these mutation rates as a function of SNP divergence, we observed a very good exponential correlation ($R^2=0.99$; Fig. 5). We used the regression curve to predict the emergence of Ural sublineage that is characterized by a divergence of 470 SNPs. We inferred that it diverged ~1800 years ago *i.e.* 3th century AD. This prediction matches previously proposed scenario locating its origin in Samartia during Scythia empire, between 5th century BC and 4th century AD (Mokrousov, 2012). Current progress in understanding sublineages phylogeny will help subsequent calibrations of *M. tuberculosis* molecular clock (Mokrousov et al., 2016).

Several parameters may be responsible for the large impact of divergence (*i.e.* time) on mutation rate (Ho et al., 2011). Selection can increase mutation rate and its effect can be emphasized by bottlenecks, however little selection signals have been detected in tuberculosis genome (Dos Vultos et al., 2008; Hershberg et al., 2008; McEvoy et al., 2012). For pathogens, transmission may also affect mutation rate by decreasing generation time and fostering selection associated to host switch. Indeed latency in tuberculosis could theoretically be associated with longer generation time. No difference in mutation rate between latent and active tuberculosis has by now been detected (Ford et al., 2011). Still, resistance acquisition rate was found higher in Beijing lineage (Ford et al., 2013), suggesting that global mutation rate may vary inside tuberculosis phylogeny.

5. Conclusion

We showed here that Tur and T3-Osaka *M. tuberculosis* genotypes are closely related and exhibit ~ 330 SNPs divergence. The specific human history of Japan suggests that T3-Osaka penetrated Japan between 500 to 1,000 years ago and potentially participated to an epidemic during the Sengoku period (15th-late 16th century).

Only intermediate mutation rate derived from the Peruvian mummy isolate could account for this scenario. This mutation rate is ~10 times higher than that corresponding to emergence of human tuberculosis lineages (1,900 SNPs and 70,000 years ago time-scale), and twice lower than that derived from contemporaneous epidemics (10 SNPs and ~10 years time-scale). Altogether this study supports an exponential decrease of tuberculosis mutation rate with divergence. Whether tuberculosis mutation rate is similar in all sublineages at each time scale remains to be tested. When doing so, an important issue will be the standardization of whole genome sequences processing.

Funding information.

This study was financed by UMR9198, CEA-CNRS-UPS, Institut de Biologie Intégrative de la Cellule, Gif-sur-Yvette, France. GR and CS hold respectively an assistant professorship and a full professorship position at University Paris-Sud. No competitive interest is to be declared. JP is supported by a BBSRC UK PhD studentship. TGC is funded by the Medical Research Council UK (Grant no. MR/K000551/1, MR/M01360X/1, MR/N010469/1).

Acknowledgments

This work was initiated in 2005 in the Institut Pasteur of Guadeloupe, and Dr. Nalin Rastogi and Dr. Riza Durmaz are acknowledged for material support, stimulating discussions, and/or for having initiated long-term collaborative studies on MTBC DNA from Turkey. During the 2007-2008 period, T. Dos Vultos, J. Rauzier and B. Gicquel through the extensive 3R MTBC gene diversity study, provided an important advance to the T3-Osa-Tur hypothesis by providing confirmatory preliminary results. Technical support was provided by the Luminex Corp. (Austin, TX) for SNP genotyping, and by King Abdullah University of Science and

Technology (KAUST) for Whole Genome Sequencing. GR and CS acknowledge recurrent support from CNRS-Univ. Paris-Sud.

ACCEPTED MANUSCRIPT

Material and Methods

Clinical isolates and DNA

Two sets of DNA were used. A first set of 43 DNAs (25+18) was provided by Dr. T. Matsumoto and Dr. E. Aktas respectively. DNA was extracted from as many clinical isolates of *Mycobacterium tuberculosis*, from patients living in Japan (Osaka region) and having previously shown to harbour the SIT627 spoligotype for T3-Osaka sublineage, and from patients living in Turkey (Zonguldak region) and having previously shown to harbour the SIT41 spoligotype (Aktas et al., 2008) for Tur sublineage.

The second set is composed by DNAs from 4 Japan clinical isolates, all exhibiting a low *IS6110* copy number, suggesting that they could belong to the T3-Osaka genomic family, and DNAs from 4 Bulgarian isolates, 2 carrying SIT41, one carrying spoligotype patterns likely derived from SIT41, and an outgroup chosen among T1 family. They were respectively prepared by T. Matsumoto (Osaka, Japan) and by S. Panaiotov (Bulgaria). DNAs were sent to U.K. for WGS via France. No ethical committee agreement was asked for this study, since DNAs were obtained locally within the frame of classical diagnostics or molecular epidemiological investigations. No electronic file with patient name was ever created and all isolates are anonymous and cannot be linked to a patient name.

Spoligotyping, MIRU-VNTR

Spoligotyping was performed for all isolates using the microbead-based method (Luminex Corp. Austin, Texas) as previously published (Zhang et al., 2010).

MIRU-VNTR analyses were performed for 9 loci (Mtub04, Mtub21, Mtub29, Mtub30, Mtub34, Mtub39, Qub11b, Qub26, Qub4156) on agarose gels in simplex as described previously in the IGEPE team in Orsay (Zhang et al., 2011).

Sanger Sequencing and Whole Genome Sequencing

The *LigB* and *LigC* genes of the 43 DNA mentioned above were sequenced on a 4 capillary ABI3100 sequencer and using the same primers as those described in Dos Vultos et al (Dos Vultos et al., 2008). Whole genome sequencing was performed at the KAUST genomic facility and coordinated by the London School of Hygiene & Tropical Medicine (LSHTM, Pathogenseq consortium, <http://pathogenseq.lshtm.ac.uk/>). The genomes are part of large drug-resistance and strain diversity studies (Benavente et al., 2015; Coll et al., 2014a; Coll et al., 2014b).

SNPs extraction and phylogenetical analysis

SNPs were extracted using the following filters: exclusion of reads mapping to several loci; 70% allelic frequency; 10 reads minimal coverage; exclusion of missing calls at the point of inflexion among full dataset of samples. Because reads mapping to several loci were excluded, we missed most of PE-PPE sequences. We did not further exclude SNPs in PE-PPE genes remaining sequences but they represented a minority of the SNPs (n=669; 4.8%). The phylogenetic tree based on WGS data was built using MEGA v.6 based on a string of concatenated SNPs deprived of information concerning position in codon (n=14,013) and was midpoint rooted (Kumar et al., 2004).

Dating the split between Turkish and T3-Osaka sub-lineages

BEAST software was used to estimate a split date between the Japanese and Bulgarian (of the Tur family) samples (Drummond and Rambaut, 2007). The software uses a Bayesian framework to estimate parameters from the sequence alignments and selected priors; it can implement either a given age of MRCA or a specific mutation rate to estimate the ages of the

internal nodes. Publically available samples data were included to represent the lineages 1-4 (lineage 1-EAI red, lineage 2-Beijing brown, lineage 3-CAS green, lineage 4-Euro American cyan, lineage 7 purple, lineage 5-Afri2 and lineage 6-Afri1 light and dark blue) (Benavente et al., 2015; Coll et al., 2014a; Coll et al., 2014b). In the first analysis, the root of the tree (i.e. the MRCA of lineages1-4) was fixed to 67,000 years as advised by Gagneux and coworkers, which corresponds to a mutation rate close to 6.2×10^{-9} mutation/site/year as per the mean number of SNPs between our samples (Comas et al., 2013). In the second analysis, the mutation rate was fixed at 4.9×10^{-8} mutation/site/year based on compute mutation rate of Peruvian body of 1,000 years old (Bos et al., 2014). In the third analysis, 10^{-7} mutation/site/year based on several recent estimations was used (Bryant et al., 2013; Ford et al., 2011; Walker et al., 2013). All models used strict molecular clocks with the HKY substitution model. Lineage 1,2,3,4 and 7 were set as a monophyletic taxon set. Note that MRCA ages were given in number of years ago (*i.e.* Before Present), while some historical data were also given as compared to Anno Domini (AD, or BC for Before Christ as in the Christian calendar).

Legend to Tables :

Table 1: MLVA (Mtub and Qub) data on the investigated clinical isolates.

Table 2 : SNPs characteristic of T3-Osa-Tur lineage.

Table 3: Human and TB palaeontology supporting different evolutionary rate scenarios

ACCEPTED MANUSCRIPT

Legend to Figures

Figure 1- Typical and related genotypes used to define Tur and T3-Osaka families (IS6110-RFLP, spoligotypes and VNTR profiles). A. LAM7-Tur isolates described by Zozio et al. 2005; B. T3-Osaka isolates VNTR, IS6110-RFLP and Spoligotypes found by Tomoshige Matsumoto *et al.*, unpublished results

Figure 2. SNP-based tree NJ built featuring T3-Osaka and Tur relatedness. The phylogeny resulting from these data shows very high robustness of almost all branches (see bootstrap values) due to the absence of recombination in the selected data. We observe a high relatedness between all isolates belonging to the same sublineage and detect that T3-Osaka and LAM7-Tur sublineages are forming a monophyletic group.

Figure 3. Dating of *M. tuberculosis* major evolutionary events according to the different evolutionary rates: A. slow mutation rate scenario (calibrated with 67,000 years as root). B. intermediate mutation rate scenario (4.8×10^{-8} SNP/genome/year). C. fast mutation rate scenario (10^{-7} SNP/genome/year).

Figure 4. A possible scenario for tuberculosis phylogeography in Europe and Asia featuring T3-Osaka and Tur emergence.

Figure 5. Mutation rate as a function of pairwise divergence.

References

- Abadia, E., Zhang, J., Vultros, T., Ritacco, V., Kremer, K., Aktas, E., Matsumoto, T., Refregier, G., Soolingen, D.V., Gicquel, B., Sola, C., 2010. Resolving lineage assignment on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. *Infect Genet Evol*.
- Aktas, E., Zozio, T., Comert, F.B., Kulah, C., Aydin, O., Rastogi, N., Sola, C., 2008. A first insight into the genetic diversity and population structure of *Mycobacterium tuberculosis* in Zonguldak, Turkey. *Clin Microbiol Infect*, 55-59.
- Azé, J., Sola, C., Zhang, J., Lafosse-Marin, F., Yasmin, M., Siddiqi, R., Kremer, K., van Soolingen, D., Refrégier, G., 2015. Genomics and machine learning for taxonomy consensus: the *Mycobacterium tuberculosis* complex paradigm. *PLoS one* 10, e0130912.
- Bartels, P., 1907. Tuberkulose (Wirbelkaries) in der jüngeren Steinzeit. *Arch. Anthropol*. 6, 243-255.
- Benavente, E.D., Coll, F., Furnham, N., McNerney, R., Glynn, J.R., Campino, S., Pain, A., Mohareb, F.R., Clark, T.G., 2015. PhyTB: Phylogenetic tree visualisation and sample positioning for *M. tuberculosis*. *BMC Bioinformatics* 16, 155.
- Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., Dehan, C., Cazajous, G., Massoure, P.L., Kraemer, P., Jenkins, A., Garnotel, E., Pourcel, C., Vergnaud, G., 2012. Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching *Mycobacterium tuberculosis* Clade. *PLoS One* 7, e52841.
- Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*.
- Bryant, J.M., Schurch, A.C., van Deutekom, H., Harris, S.R., de Beer, J.L., de Jager, V., Kremer, K., van Hijum, S.A., Siezen, R.J., Borgdorff, M., Bentley, S.D., Parkhill, J., van Soolingen, D., 2013. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. *BMC Infect Dis* 13, 110.
- Cavalli-Sforza, L.L., 2000. *Genes, peoples and languages*. Farrar, Straus and Giroux, New York.
- Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigo, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014a. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat Commun* 5, 4812.
- Coll, F., Preston, M., Guerra-Assuncao, J.A., Hill-Cawthorn, G., Harris, D., Perdigo, J., Viveiros, M., Portugal, I., Drobniewski, F., Gagneux, S., Glynn, J.R., Pain, A., Parkhill, J., McNerney, R., Martin, N., Clark, T.G., 2014b. PolyTB: A genomic variation map for *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)*.
- Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bertranpetit, J., 1996. Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13, 1067-1077.
- Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 45, 1176-1182.

Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., Mokrousov, I., Sola, C., Zozio, T., Rastogi, N., 2012. SITVITWEB – A publicly available international multimer database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. *Infect Genet Evol* 12, 755-766.

Dos Vultos, T., Mestre, O., Rauzier, J., Golec, M., Rastogi, N., Rasolofo, V., Tonjum, T., Sola, C., Matic, I., Gicquel, B., 2008. Evolution and Diversity of Clonal Bacteria: The Paradigm of *Mycobacterium tuberculosis*. *PLoS ONE* 3, e1538.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.

Durmaz, R., Zozio, T., Gunal, S., Yaman, A., Cavusoglu, C., Guney, C., Sola, C., Rastogi, N., 2007. Genetic diversity and major spoligotype families of drug-resistant *Mycobacterium tuberculosis* clinical isolates from different regions of Turkey. *Infect Genet Evol* 7, 513-519.

Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., Gadisa, E., Kiros, T., Habtamu, M., Hussein, J., Zinsstag, J., Robertson, B.D., Ameni, G., Lohan, A.J., Loftus, B., Comas, I., Gagneux, S., Tschopp, R., Yamuah, L., Hewinson, G., Gordon, S.V., Young, D.B., Aseffa, A., 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerg Infect Dis* 19, 460-463.

Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T.R., Sacchettini, J.C., Lipsitch, M., Flynn, J.L., Fortune, S.M., 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. *Nat Genet* 43, 482-486.

Ford, C.B., Shah, R.R., Maeda, M.K., Gagneux, S., Murray, M.B., Cohen, T., Johnston, J.C., Gardy, J., Lipsitch, M., Fortune, S.M., 2013. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat Genet* 45, 784-790.

Gomgnimbou, M.K., Hernandez-Neuta, I., Panaiotov, S., Bachiyska, E., Palomino, J.C., Martin, A., Del Portillo, P., Refrégier, G., Sola, C., 2013. “TB-SPRINT: TuBerculosis-SPoligo-Rifampin-IsoNiazid Typing”; an All-in-One assay technique for surveillance and control of multi-drug resistant tuberculosis on Luminex® devices. *J Clin Microbiol* 51, 3527-3534.

Grmek, M., 1994. Chapitre VII: Une grande Tueuse, la tuberculose, Les maladies à l'aube de la civilisation occidentale. Payot, Paris, pp. 261-290.

Guerra-Assuncao, J.A., Crampin, A.C., Houben, R.M., Mzembe, T., Mallard, K., Coll, F., Khan, P., Banda, L., Chiwaya, A., Pereira, R.P., McNerney, R., Fine, P.E., Parkhill, J., Clark, T.G., Glynn, J.R., 2015. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *Elife* 4.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P., Vincent, V., 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. *PLoS Pathog* 1, e5.

Hammer, M.F., Karafet, T.M., Park, H., Omoto, K., Harihara, S., Stoneking, M., Horai, S., 2006. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51, 47-58.

Hellenthal, G., Busby, G.B., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. *Science* 343, 747-751.

Henn, B.M., Cavalli-Sforza, L.L., Feldman, M.W., 2012. The great human expansion. *Proc Natl Acad Sci U S A* 109, 17758-17764.

Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., Gagneux, S., 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. *PLoS Biol* 6, e311.

- Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. PLoS ONE 3, e3426.
- Ho, S.Y., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A., 2011. Time-dependent rates of molecular evolution. Mol Ecol 20, 3087-3101.
- Howell, N., Smejkal, C.B., Mackey, D.A., Chinnery, P.F., Turnbull, D.M., Herrnstadt, C., 2003. The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. Am J Hum Genet 72, 659-670.
- Hughes, A.L., Friedman, R., Murray, M., 2002. Genomewide Pattern of Synonymous Nucleotide Substitution in Two Complete Genomes of *Mycobacterium tuberculosis*. Emerg Infect Dis 8, 1342-1346.
- Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., Meshveliani, T., Bar-Yosef, O., Muller, W., Belfer-Cohen, A., Matskevich, Z., Jakeli, N., Higham, T.F., Currat, M., Lordkipanidze, D., Hofreiter, M., Manica, A., Pinhasi, R., Bradley, D.G., 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nat Commun 6, 8912.
- Kaniewski, D., Van Campo, E., Guiot, J., Le Burel, S., Otto, T., Baeteman, C., 2013. Environmental roots of the late bronze age crisis. PLoS One 8, e71004.
- Kappelman, J., Alcicek, M.C., Kazanci, N., Schultz, M., Ozkul, M., Sen, S., 2008. First *Homo erectus* from Turkey and implications for migrations into temperate Eurasia. Am J Phys Anthropol 135, 110-116.
- Kisa, O., Tarhan, G., Gunal, S., Albay, A., Durmaz, R., Saribas, Z., Zozio, T., Alp, A., Ceyhan, I., Tombak, A., Rastogi, N., 2012. Distribution of spoligotyping defined genotypic lineages among drug-resistant *Mycobacterium tuberculosis* complex clinical isolates in Ankara, Turkey. PLoS One 7, e30331.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5, 150-163.
- Kuo, C.H., Ochman, H., 2009. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. Biol Direct 4, 35.
- Lee, S., Hasegawa, T., 2013. Evolution of the Ainu language in space and time. PLoS One 8, e62243.
- Lieverse, A.R., Temple, D.H., Bazaliiskii, V.I., 2014. Paleopathological description and diagnosis of metastatic carcinoma in an Early Bronze Age (4588±34 Cal. BP) forager from the Cis-Baikal region of Eastern Siberia. PLoS One 9, e113919.
- Masson, M., Bereczki, Z., Molnar, E., Donoghue, H.D., Minnikin, D.E., Lee, O.Y., Wu, H.H., Besra, G.S., Bull, I.D., Palfi, G., 2015. 7000 year-old tuberculosis cases from Hungary - Osteological and biomolecular evidence. Tuberculosis (Edinb) 95 Suppl 1, S13-17.
- McEvoy, C.R., Cloete, R., Muller, B., Schurch, A.C., van Helden, P.D., Gagneux, S., Warren, R.M., Gey van Pittius, N.C., 2012. Comparative analysis of *Mycobacterium tuberculosis* *pe* and *ppe* genes reveals high sequence variation and an apparent absence of selective constraints. PLoS One 7, e30593.
- Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M.G., Rusch-Gerdes, S., Mokrousov, I., Aleksic, E., Allix-Beguec, C., Antierens, A., Augustynowicz-Kopec, E., Ballif, M., Barletta, F., Beck, H.P., Barry, C.E., 3rd, Bonnet, M., Borroni, E., Campos-Herrero, I., Cirillo, D., Cox, H., Crowe, S., Crudu, V., Diel, R., Drobniewski, F., Fauville-Dufaux, M., Gagneux, S., Ghebremichael, S., Hanekom, M., Hoffner, S., Jiao, W.W., Kalon, S., Kohl, T.A., Kontsevaya, I., Lillebaek, T., Maeda, S., Nikolayevskyy, V., Rasmussen, M., Rastogi, N., Samper, S., Sanchez-Padilla, E., Savic, B., Shamputa, I.C., Shen, A., Sng, L.H., Stakenas, P., Toit, K., Varaine, F., Vukovic, D., Wahl,

- C., Warren, R., Supply, P., Niemann, S., Wirth, T., 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 47, 242-249.
- Millet, J., Miyagi-Shiohira, C., Yamane, N., Sola, C., Rastogi, N., 2007. Assessment of mycobacterial interspersed repetitive unit-QUB markers to further discriminate the Beijing genotype in a population-based study of the genetic diversity of *Mycobacterium tuberculosis* clinical isolates from Okinawa, Ryukyu Islands, Japan. *J Clin Microbiol* 45, 3606-3615.
- Mokrousov, I., 2008. Genetic geography of *Mycobacterium tuberculosis* Beijing genotype: A multifacet mirror of human history? *Infect Genet Evol*.
- Mokrousov, I., 2012. The quiet and controversial: Ural family of *Mycobacterium tuberculosis*. *Infect Genet Evol* 12, 619-629.
- Mokrousov, I., Vyazovaya, A., Iwamoto, T., Skiba, Y., Pole, I., Zhdanova, S., Arikawa, K., Sinkov, V., Umpeleva, T., Valcheva, V., Alvarez Figueroa, M., Ranka, R., Jansone, I., Ogarkov, O., Zhuravlev, V., Narvskaya, O., 2016. Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: Human traces across pathogen's phylogeography. *Mol Phylogenet Evol* 99, 133-143.
- Nakagome, S., Sato, T., Ishida, H., Hanihara, T., Yamaguchi, T., Kimura, R., Mano, S., Oota, H., Asian, D.N.A.R.C., 2015. Model-based verification of hypotheses on the origin of modern Japanese revisited by bayesian inference based on genome-wide SNP data. *Mol Biol Evol* 32, 1533-1543.
- Ohata, R., Tada, A., 2004. [Beijing family and other genotypes of *Mycobacterium tuberculosis* isolates in Okayama district]. *Kekkaku* 79, 47-53.
- Oral Zeytinli, U., Koksal, F., 2012. [Genotyping of *Mycobacterium tuberculosis* strains isolated from pulmonary tuberculosis patients in Cukurova Region, Turkey by spoligotyping and MIRU-VNTR methods]. *Mikrobiyol Bul* 46, 202-210.
- Otlu, B., Durmaz, R., Gunal, S., Sola, C., Zozio, T., Rastogi, N., 2009. Beijing/W and major spoligotype families of *Mycobacterium tuberculosis* strains isolated from tuberculosis patients in Eastern Turkey. *New Microbiol* 32, 255-263.
- Ozcan, E., 2006. Human leukocyte antigen DR1 in Japanese and Turkish patients with schizophrenia. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 30, 423-428.
- Palfi, G., Dutour, O., Perrin, P., Sola, C., Zink, A., 2015. Tuberculosis in evolution. *Tuberculosis (Edinb)* 95 Suppl 1, S1-3.
- Parsons, T.J., Muniec, D.S., Sullivan, K., Woodyatt, N., Alliston-Greiner, R., Wilson, M.R., Berry, D.L., Holland, K.A., Weedn, V.W., Gill, P., Holland, M.M., 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15, 363-368.
- Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., Dodgson, J., 1980. The evolution of genes: the chicken preproinsulin gene. *Cell* 20, 555-566.
- Robbeets, M., 2004. Swadesh 100 on Japanese, Korean and Altaic. *Tokyo University Linguistic Papers TULIP* 23, 99-118.
- Robbeets, M., 2008. Is Japanese Alaic, how can it be so simple, in: Lubotsky, A., Schaeken, J., Wiedenhof, J. (Eds.), *Evidence and Counter-Evidence: Essays in Honour of Frederik Kortlandt*. *Rodopi* (September 30, 2008), Amsterdam-New York, pp. 337-367.
- Roberts, C.A., Pfister, L.A., Mays, S., 2009. Letter to the editor: was tuberculosis present in *Homo erectus* in Turkey? *Am J Phys Anthropol* 139, 442-444.
- Roetzer, A., Diel, R., Kohl, T.A., Ruckert, C., Nubel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rusch-Gerdes, S., Supply, P., Kalinowski, J., Niemann, S., 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. *PLoS Med* 10, e1001387.

- Sinor, D., 1990. The establishment and dissolution of the Turk empire, *The Cambridge History of Early Inner Asia*. Cambridge University Press, Cambridge, pp. 285–316.
- Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A., Salas, A., Oppenheimer, S., Macaulay, V., Richards, M.B., 2009. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 84, 740-759.
- Suzuki, T., Fujita, H., Choi, J.G., 2008. Brief communication: new evidence of tuberculosis from prehistoric Korea—Population movement and early evidence of tuberculosis in far East Asia. *Am J Phys Anthropol* 136, 357-360.
- Suzuki, T., Inoue, T., 2007. Earliest Evidence of Spinal Tuberculosis from the Aneolithic Yayoi Period in Japan. *International Journal of Osteoarchaeology* 17, 392-492.
- Tajima, A., Hayami, M., Tokunaga, K., Juji, T., Matsuo, M., Marzuki, S., Omoto, K., Horai, S., 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49, 187-193.
- Takashima, T., Iwamoto, T., 2006. [New era in molecular epidemiology of tuberculosis in Japan]. *Kekkaku* 81, 693-707.
- Valcheva, V., Mokrousov, I., Narvskaya, O., Rastogi, N., Markova, N., 2008. Molecular snapshot of drug-resistant and drug-susceptible *Mycobacterium tuberculosis* strains circulating in Bulgaria. *Infect Genet Evol* 8, 657-663.
- Vasileva, D., 1992. Bulgarian Turkish emigration and return. *Int Migr Rev* 26, 342-352.
- Walker, T.M., Ip, C.L., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 13, 137-146.
- WHO, 2013. *Global Tuberculosis Report 2013*. WHO, Geneva, Switzerland.
- Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S., Dalimova, D., Nymadawa, P., Bahmanimehr, A., Sahakyan, H., Tambets, K., Fedorova, S., Barashkov, N., Khidiyatova, I., Mihailov, E., Khusainova, R., Damba, L., Derenko, M., Malyarchuk, B., Osipova, L., Voevoda, M., Yepiskoposyan, L., Kivisild, T., Khusnutdinova, E., Villems, R., 2015. The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet* 11, e1005068.
- Zhang, J., Abadia, E., Refregier, G., Tafaj, S., Boschioli, M.L., Guillard, B., Andremont, A., Ruimy, R., Sola, C., 2010. *Mycobacterium tuberculosis* complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. *J Med Microbiol* 59, 285-294.
- Zhang, J., Heng, S., Le Moullec, S., Refregier, G., Gicquel, B., Sola, C., Guillard, B., 2011. A first assessment of the genetic diversity of *Mycobacterium tuberculosis* complex in Cambodia. *BMC Infect Dis* 11, 42.
- Zozio, T., Allix, C., Gunal, S., Saribas, Z., Alp, A., Durmaz, R., Fauville-Dufaux, M., Rastogi, N., Sola, C., 2005. Genotyping of *Mycobacterium tuberculosis* clinical isolates in two cities of Turkey: description of a new family of genotypes that is phylogeographically specific for Asia Minor. *BMC Microbiol* 5, 44.

Table 1: MLVA (Mtb and Qub) data on the investigated clinical isolates. Bds=bands.

| | sample n° | ID | Mtub04 | Mtub21 | Mtub29 | Mtub30 | Mtub34 | Mtub39 | Qub11b | Qub26 | Qub4156 | code MLVA |
|--------------|-----------|-------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|-----------|
| H37Rv | | | 2 | 2 | 4 | 2 | 3 | 5 | 5 | 5 | 2 | 224235552 |
| Turkey | Tur-1 | 1 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 434433673 |
| Turkey | Tur-2 | 2 | 4 | 3 | 4 | 4 | 3 | 3 | 5 | 7 | 2 | 434433572 |
| Turkey | Tur-3 | 3 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 434433683 |
| Turkey | Tur-4 | 4 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 434433683 |
| Turkey | Tur-5 | 5 | 4 | 3 | 2 | 4 | 3 | 3 | 5 | 6 | 2 | 432433562 |
| Turkey | Tur-6 | 6 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 434433683 |
| Turkey | Tur-7 | 7 | 4 | 3 | 4 | 4 | 2 | 3 | 6 | 7 | 3 | 434423673 |
| Turkey | Tur-8 | 8 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 434433673 |
| Turkey | Tur-9 | 9 | 4 | 3 | 4 | 4 | 3 | 3 | 5 | 7 | 2 | 434433572 |
| Turkey | Tur-10 | 10 | 4 | 3 | 4 | 4 | 3 | 3 | 5 | 7 | 2 | 434433572 |
| Turkey | Tur-11 | 11 | 4 | 3 | 4 | 4 | 3 | 3 | 5 | 7 | 2 | 434433572 |
| Turkey | Tur-12 | 12 | 4 | 4 | 4 | 4 | 3 | 3 | 4 | 7 | 1 | 444433471 |
| Turkey | Tur-13 | 13 | 4 | 3 | 4 | 4 | 3 | 4 | 6 | 7 | 3 | 434434673 |
| Turkey | Tur-14 | 14 | 4 | 3bds | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 4X4433673 |
| Turkey | Tur-15 | 15 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 434433673 |
| Turkey | Tur-16 | 16 | 4 | 3 | 4 | 4 | 3 | 3 | 5 | ND | 2 | 4344335N2 |
| Turkey | Tur-17 | 17 | 4 | 3 | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 434433673 |
| Turkey | Tur-18 | 18 | 5 | 3 | 4 | 4 | 3 | 3 | 6 | 7 | 3 | 534433673 |
| Bulg | Tur-19 | 20_10 | 4 | 3 | 4 | 2 | 3 | 3 | 1 | 11 | 3 | 4342331a3 |
| Bulg | Tur-20 | 24_10 | 4 | 3 | 4 | 4 | 3 | 3 | 1 | 11 | 3 | 4344331a3 |
| Bulg | Tur-21 | 78 | 4 | 2 | 4 | 4 | 3 | 3 | 1 | 11 | 3 | 4244331a3 |
| Japan | T30-1 | R1-1 | 1 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 134433683 |
| Japan | T30-2 | R1-2 | 2 | 2bds | 4 | 4 | 3 | 3 | 5 | 6 | 2 | 2X4433562 |
| Japan | T30-3 | R1-3 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 6 | 3 | 234433663 |
| Japan | T30-4 | R1-4 | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| Japan | T30-5 | R1-5 | 2 | 4 | 4 | 4 | 3 | 2 | 6 | 4 | 3 | 244432643 |
| Japan | T30-6 | R1-6 | 2 | 3 | 4 | 4 | 3 | 5 | 6 | 8 | 3 | 234435683 |
| Japan | T30-7 | R1-7 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 234433683 |
| Japan | T30-8 | R1-8 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 234433683 |
| Japan | T30-9 | R1-9 | 2 | 3 | 4 | 4 | 3 | 3 | 2 | 9 | 4 | 234433294 |
| Japan | T30-10 | R1-11 | 2 | 3 | 4 | 4 | 3 | 3 | 7 | 8 | 4 | 234433784 |
| Japan | T30-11 | R1-12 | 2 | 3 | 5 | 4 | 3 | 3 | 6 | 6 | 3 | 235433663 |
| Japan | T30-12 | R1-13 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 5 | 3 | 234433653 |
| Japan | T30-13 | R1-14 | 2 | 3 | 4 | 4 | 3 | 4 | 6 | 8 | 3 | 234434683 |
| Japan | T30-14 | R1-15 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 5 | 3 | 234433653 |
| Japan | T30-15 | R4-1 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 234433683 |
| Japan | T30-16 | R6-1 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 3 | 3 | 234433633 |
| Japan | T30-17 | R6-2 | 3 | 3 | 3 | 4 | 3 | 3 | 6 | 8 | 3 | 333433683 |
| Japan | T30-18 | R6-3 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 234433683 |
| Japan | T30-19 | R6-4 | 2 | 3 | 4 | 4 | 3 | 5 | 6 | ND | 3 | 2344356N3 |
| Japan | T30-20 | R6-5 | 2 | 3 | 4 | 4 | 3 | ND | 6 | ND | ND | 23443N6NN |
| Japan | T30-21 | R6-7 | 2 | 3 | 4 | 4 | 3 | 5 | 6 | 8 | 3 | 234435683 |
| Japan | T30-22 | R6-8 | 2 | 3 | 4 | 4 | 3 | 3 | 6 | 8 | 3 | 234433683 |
| Japan | T30-23 | R6-9 | 2 | 3 | 4 | 4 | 3 | 3 | 7 | 9 | 4 | 234433794 |
| Japan | T30-24 | R6-10 | 3 | 3 | 4 | 4 | 3 | 2 | 6 | 8 | 3 | 334432683 |
| Japan | T30-25 | R6-11 | 3 | 3 | 4 | 4 | 3 | 3 | 6 | 6 | 3 | 334433663 |

Table 2:
SNPs
characteristic
of T3-Osa-
Tur lineage.

ref=reference; alt=alternative; aa=amino acid; CDS= coding Sequence; NS= non-synonymous; S=synonymous. Positions are given according to H37Rv genome numbering.

| position | ref | alt | ref_codon | alt_codon | ref_aa | alt_aa | locus_tag | gene | coding? | type |
|----------|-----|-----|-----------|-----------|--------|--------|---------------|----------------------|---------|------|
| 401678 | C | A | CCC | CAC | P | H | <i>Rv0336</i> | Hypothetical protein | CDS | NS |
| 1562830 | G | C | TTG | TTC | L | F | <i>Rv1387</i> | <i>PPE20</i> | CDS | NS |
| 2208627 | C | G | CGG | GGG | R | G | <i>Rv1965</i> | <i>yrbE3B</i> | CDS | NS |
| 2563611 | G | A | GGG | GGA | G | G | <i>Rv2291</i> | <i>sseB</i> | CDS | S |
| 3985860 | C | T | CTG | TTG | L | L | <i>Rv3546</i> | <i>fadA5</i> | CDS | S |
| 4182566 | G | T | TGG | TTG | W | L | <i>Rv3731</i> | <i>ligC</i> | CDS | NS |
| 4239334 | T | C | GTG | GCG | V | A | <i>Rv3792</i> | <i>aftA</i> | CDS | NS |

Table 3: Human and TB genetics, history and palaeontology supporting different evolutionary rate scenarios. Mitoch : data derived from human mitochondrial DNA. WGS: data derived from Whole Genome Sequencing. Data supporting the corresponding timing are written in bold. Data contradicting timing are underlined. Supported timing is written in bold.

| | SLOW evolution rate scenario | | | | INTERMEDIATE evolution rate scenario | | | |
|---|--|---|---|---|---|--|---|--|
| | Inferred MRCA ages (timing) | TB paleontology | human genetics | culture and language | Inferred MRCA ages | TB paleontology | Human genetics | culture and language |
| A- Middle East and world – ancestral populations | MRCA of Lineages 2-4: ~ 40 ,000y ago | Human modern TB, Israel 9,000 y ago (Hershkovitz, 2008); bison with modern TB in America 17,000 y ago (Rotschild, 2001) | Various genetic and paleoanthropological data indicate an Out-of-Africa demographic expansion ca. 60,000 y ago. WGS: Caucasus Hunter Gatherers (CHG) isolated from ancestors of farmers around 25,000 y ago (Jones, 2015) | | MRCA of lineages 2-4 and 7: 4,000 y ago | | Mitoch: various lineages present in Altai during Iron Age 2,500 y ago (Gonzalez-Ruiz, 2012) | |
| B- Central Asia | T3-Osa-Tur-Ural MRCA: 16,000 y ago; T3-Osa-Tur MRCA: 12,000y ago | TB in human remains of 4,500y ago in Baikal region (Lieverse, 2014) | CHG contributed to the gene pool of East Europe (including Turkey) at least stretching to Kyrgystan (Jones, 2015) | Relatedness between Tungusic languages (Japanese, Ryuku and Ainu) and Turkic languages ; time of divergence: at least 6,000 y ago (Yunusbaiev, 2015) | TB since less than 4,000 y ago | 4,500y-old remains with TB in Baikal region (Lieverse, 2014) | Specific chromosomal tracts in mongolic populations likely herited from Mongolic empire (Yunusbaiev, 2015) | Türk empire 5th-7th century AD (1,600-1,400 y ago); early Islam empires 7th-12 th (1,400-900 y ago); Mongol empire emerging around 1250 AD (750 y ago) (Sinor, 1990) |
| C- Japan colonization and potential epidemic | MRCA T3-Osaka: at least 7,500 y ago | <u>Absence of tuberculous skeletons</u> from Jomon period in Japan despite extensive search (Suzuki, 2007) | Y-chromosom haplogroups D and C mostly present in Tibet and Mongolia colonized Japan around 20,000 and 7,000 y ago (Hammer, 2006; Tajima, 2004; Wang, 2014) | Jomon culture = Hunter Gatherers thrived in Japan and built Jomon culture more than 12,000 y ago (Hammer, 2006) | MRCA Japanese T3O of 500y ago | first known cases of tuberculosis in Japan (between 2,300 and 1,700 y ago; Suzuki 2007) | Chromosome painting: Uncertain admixture in Japanese population during Mongol empire (Hellenthal, 2014) | 5th-16th centuries AD (1,600 to 500 y ago) Turkic then Mongol military movements towards East (Sinor, 1990); tentative invasion of Japan by Genghis Khan (1274 AD, 720 y ago) (Cheow, 2008); wars during Sengoku period (15th-16th cent.), peace and prosperity during Edo-Tokugawa period (1603-1868 AD <i>i.e.</i> 400-100 y ago) |
| D- Turkey colonization | MRCA Tur: 3,500-10,000 y ago | | Several migration events from Central Asia towards Turkey, among which Yamanyan step herders 5kya (Jones, 2015) and Scythian 2.5kya (Gonzalez-Ruiz, 2012); Y-chromosome C haplogroups in Turkey and Chechenia (Nasidze, 2004) | Proto-Turkic language seems to have been used from 4,000 to 2,000 y ago and influence seems to have given rise to many languages among which Turkish (Yunusbaiev, 2015; Dybo, 2008) | Tur MRCA: 800 to 250y ago | | WGS: admixture of Mongolian chromosomal tracts in Turkish population dated back to 9th to 17th century <i>i.e.</i> 1,200 to 400 y ago (Yunusbaiev, 2015); chromosome painting: gene influx from NE Asia to Uzbekistan and further West around 700 y ago (Hellenthal, 2014) | 5th-16th centuries (1,600 to 500 y ago) Turkic then Mongol military movements towards East and West with transient conquests (Sinor, 1990) |

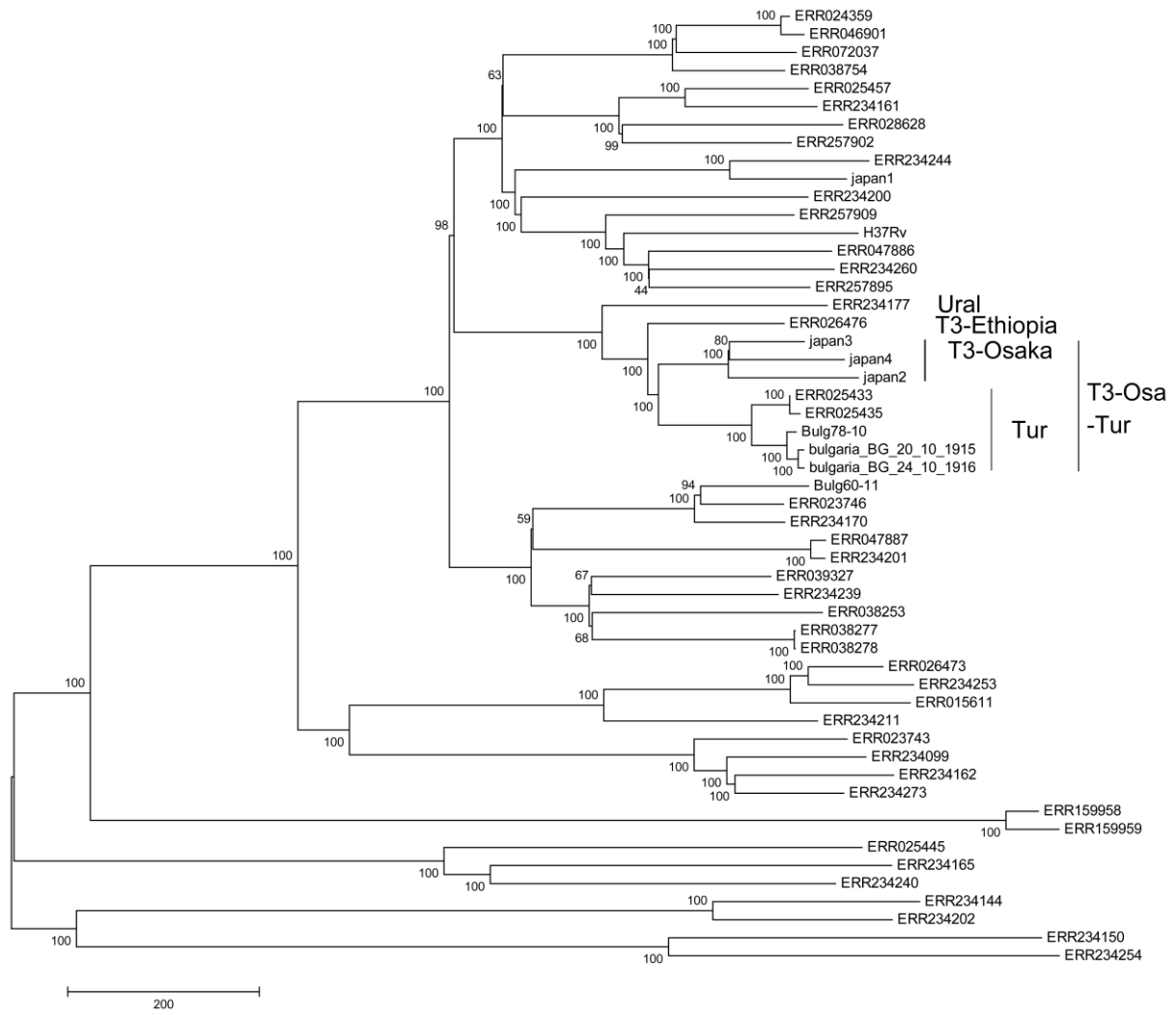
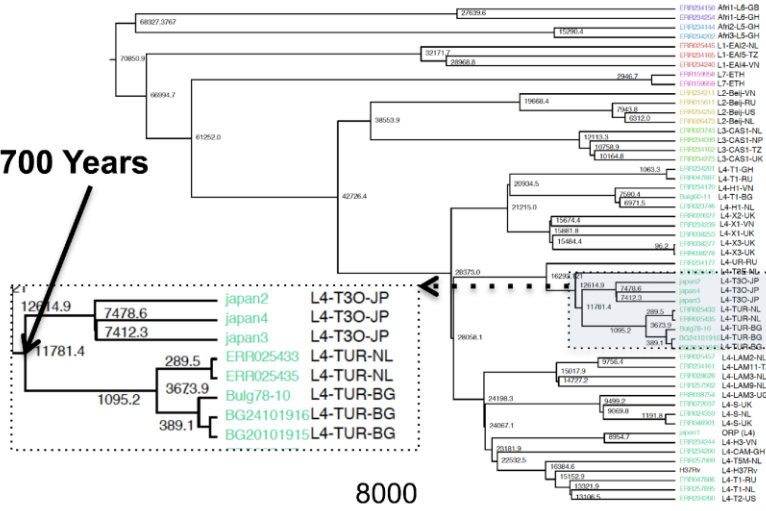


Figure 2

ACCE

3A

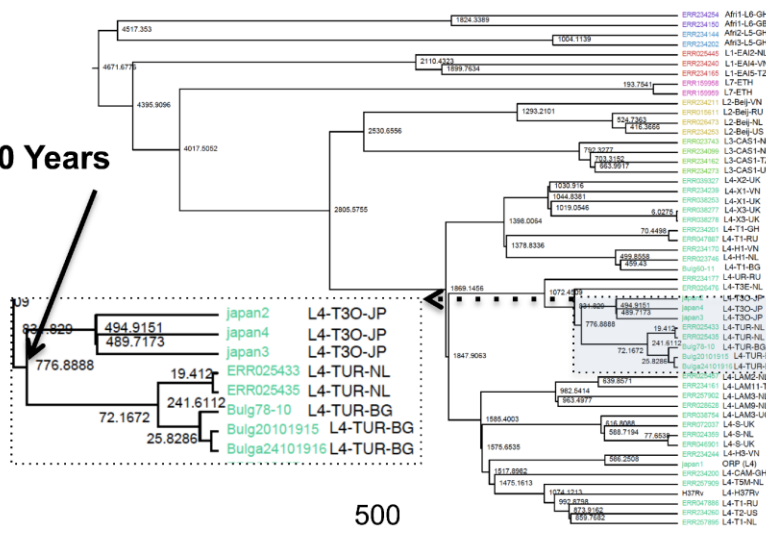
11700 Years



MRCa ~70,000 y BP
 6×10^{-8} SNP/bp/y
 0.03 SNP/genome/y

3B

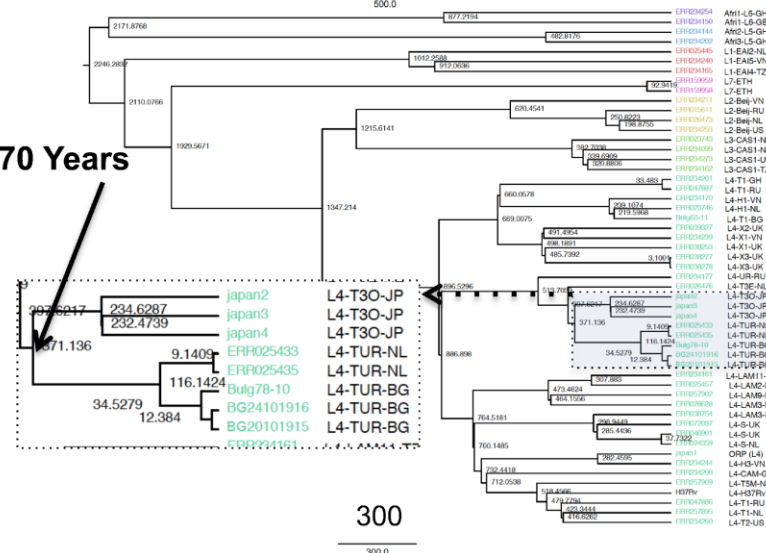
780 Years



4.8×10^{-8} SNP/bp/y
 0.22 SNP/genome/y

3C

370 Years



10^{-7} SNP/bp/y
 0.46 SNP/genome/y

Figure 3

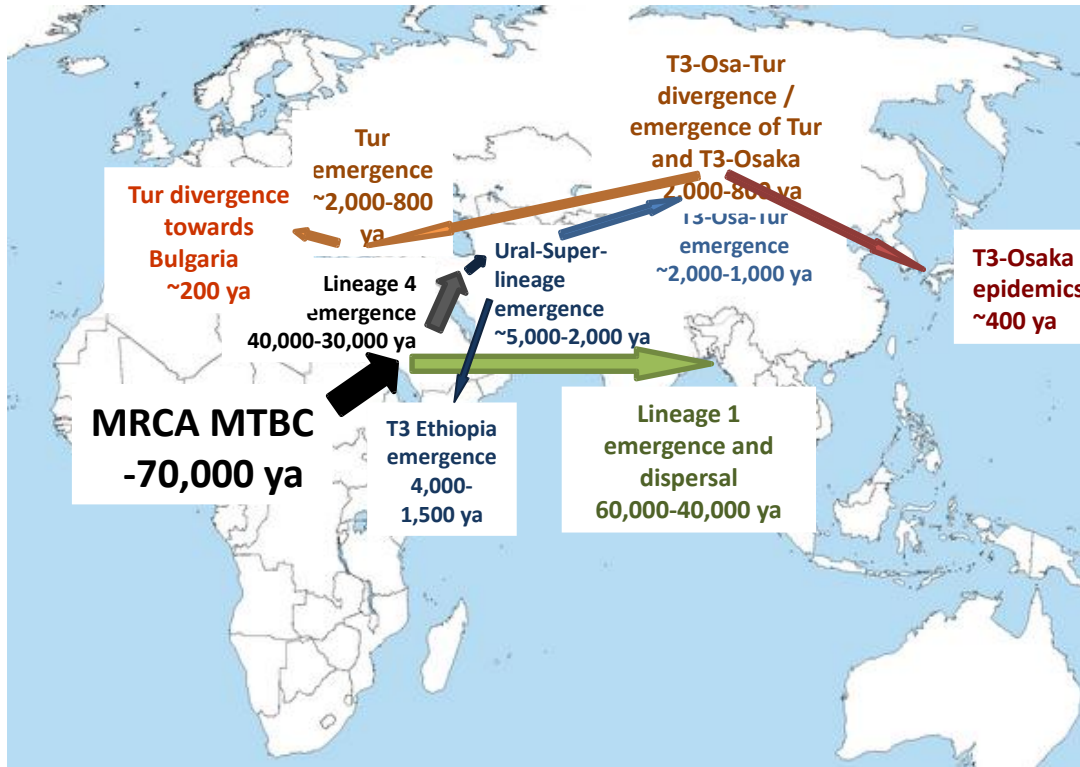
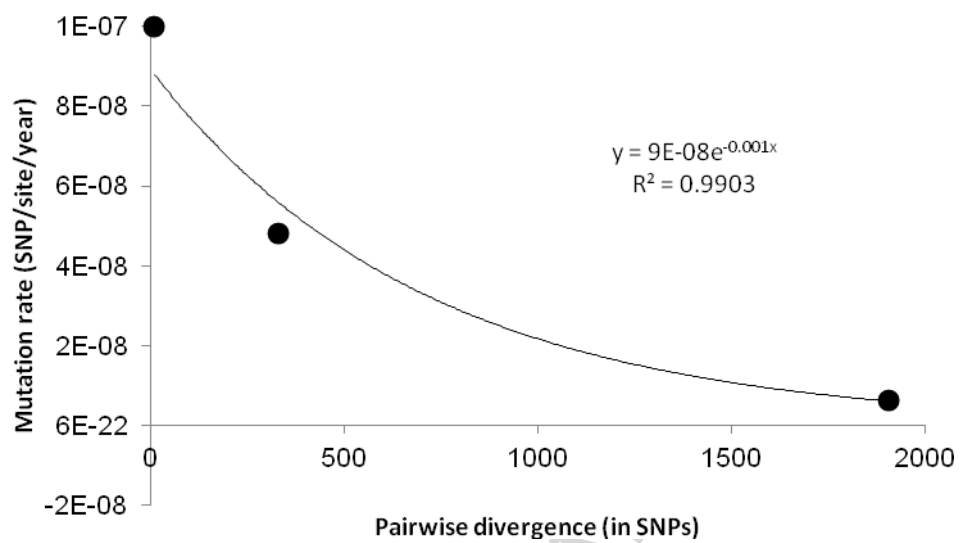


Figure 4

ACCEPTED

Figure 5- Mutation rate as a function of pairwise divergence



ACCEPTED MANUSCRIPT

Highlights

- Whole Genome data prove that T3-Osaka and Tur lineages are highly related
- T3-Osa-Tur Most Recent Common Ancestor (MRCA) may date back to Mongol Empire
- This MRCA may serve as first calibration point for Euro-American TB molecular clock

ACCEPTED MANUSCRIPT