Differences among sublineages of the East-Asian lineage of *Mycobacterium tuberculosis* in genotypic clustering

M. Kato-Maeda[*], **E. Y. Kim**[*], **L. Flores**[*], **L. G. Jarlsberg**[*], **D. Osmond**[†], and **P. C. Hopewell**[*]

[*]Francis J Curry National Tuberculosis Center, Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, Department of Medicine, University of California San Francisco, San Francisco, California,

[†]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

SUMMARY

SETTING—The East-Asian lineage of *Mycobacterium tuberculosis* is composed of five sublineages, and includes the strains from the Beijing spoligotype family. In some studies these strains were highly pathogenic, although other studies did not support this finding.

OBJECTIVE—To determine if the sublineages of the East-Asian lineage of *M. tuberculosis* differ in their capacity to cause secondary cases, as assessed by genotypic clustering of isolates.

DESIGN—In a population-based study of 545 patients with *M. tuberculosis* from the East-Asian lineage in San Francisco, we used DNA-based fingerprinting to identify genotypic clustering, which was compared among the different sublineages defined by large sequence polymorphism.

RESULTS—Strains from sublineage 207 had the highest frequency of genotypic clustering. In the multivariate analysis, only patients born in the United States were associated with clustering.

CONCLUSIONS—We found evidence in a univariate analysis that the different East-Asian sublineages of *M. tuberculosis* have different frequencies of genotypic clustering. The effect size for this difference was unchanged in multivariate analysis, although loss of observations due to missing data resulted in a non-significant *P* value. It is tantalizing to hypothesize that the different East-Asian sublineages may differ in their capacity to cause secondary cases.

Keywords

Mycobacterium tuberculosis; genotypic clustering; East-Asian lineage

THE EAST-ASIAN LINEAGE is one of the six *Mycobacterium tuberculosis* lineages defined by DNA large sequence polymorphisms (LSP), identified as deletions of segments of DNA when compared with the reference strain H37Rv.[1] This lineage includes the Beijing family, which is defined by a spoligotype in which direct variable repeats (DVR) 1 to 36 are missing.[2] The East-Asian lineage characterized by the LSP, also known as region of difference (RD)105,[3] is monophyletic and can be divided into five sublineages by specific RD (Figure 1).

Correspondence to: M Kato-Maeda, San Francisco General Hospital 1001 Potrero Ave, Box 0841 San Francisco, CA 94110, USA. Tel: (+1) 415 206 8121. Fax: (+1) 415 695 1551. midori.kato-maeda@ucsf.edu.

Strains with the characteristic Beijing spoligotype have been identified throughout the world, with the exception of South America,[4–6] and have been associated with outbreaks.[7,8] Several hypotheses have been proposed to explain the widespread dissemination of these strains: 1) dissemination may be related to the large global migrations of persons from Asia in the twentieth century;[4] 2) the bacille Calmette-Guérin (BCG) vaccine may have positively selected for the Beijing genotype by having less protective efficacy against this lineage;[2,9] 3) Beijing strains may be more pathogenic (defined as the ability to cause disease) or more virulent (defined as the ability to cause severe disease) or both, compared to other strains of *M. tuberculosis.*[10,11] These characteristics have been attributed, in part, to the production of polyketide synthase-derived phenolic glycolipid (PGL) by *M. tuberculosis* which inhibits the innate immune response in murine disease models.[12] This lipid is synthesized by the *pks* 1–15 gene, which is present in all the strains from the East-Asian lineage, but is truncated in strains from other lineages.[13] However, it was recently shown that despite having an intact gene, the proportion of strains that produced PGL was respectively 10%, 60% and 80% among strains from sublineages 181, 150 and 142.[14]

In this study, we examined the hypothesis that strains from different sublineages of the East-Asian lineage vary in their capacity to cause secondary cases (as a proxy for pathogenicity) of tuberculosis (TB) as determined by genotypic clustering of *M. tuberculosis.* Because LSPs are unique event polymorphisms and have been shown to be robust markers for constructing phylogenetic trees, we included in the study all isolates from the East-Asian lineage.[3]

STUDY POPULATION AND METHODS

The study comprised all patients with TB caused by the East-Asian lineage from the San Francisco population-based molecular epidemiology study,[15] in which all patients with culture-positive TB reported in San Francisco between January 1991 and December 2006 who met the Centers for Disease Control and Prevention criteria were included as incident cases. The protocols and procedures for the protection of human participants were approved by the University of California, San Francisco.

The lineage of the *M. tuberculosis* isolates was determined using multiplex real-time polymerase chain reaction (PCR) and PCR, as reported previously.[1,16] Sublineages were determined in isolates from the East-Asian lineage using real-time PCR for RD 207, RD 150 and RD 142, and PCR for RD 181, as reported previously.[3]

Restriction fragment length polymorphisms (RFLP) of the insertion sequence (IS) *6110* and the polymorphic guanine-cytosine rich sequence (PGRS) were used to identify genotypic clustering. RFLP typing was performed following standardized procedures,[17,18] and the images were analyzed using Bio-Image Whole Band Analyzer software (version 9 3.3, Bio-Image Corporation, Ann Arbor, MI, USA). The band assignment was edited by two independent readers and the genetic clustering assignment was confirmed visually.

Genotypic clustered cases were defined as cases with isolates that had the same IS*6110* RFLP genotype (and the same PGRS RFLP if fewer than six IS*6110* bands) as at least one other case. Patients within the cluster were considered to have an epidemiologic link and thus to be part of a chain of transmission. The initial case identified was considered to be the index case, and subsequent cases were considered secondary cases. Cases with isolates with no matching RFLPs (unique cases) were considered a result of reactivation of latent infection.

Statistical analysis

To explore the relationship between the sublineages and demographic and clinical characteristics of patients, univariate analyses were performed using the χ^2 test of proportions and the two-tailed Fisher's exact test. We calculated univariate and multivariate odds ratios (ORs) using logistic regression with secondary case status as the dependent variable. The inde pendent variable of primary interest was the specific East-Asian sublineage, but we included and controlled for other characteristics shown to be predictive of developing secondary cases in previous publications (place of birth, smear positivity and cavitary disease). Statistical analyses were performed using SAS version 9.2 (SAS Institute Inc, Cary, NC, USA).

RESULTS

From January 1991 to December 2006, 3585 cases of TB were reported in San Francisco, of whom 3044 (85%) had a positive culture for *M. tuberculosis*. RFLPs and lineage data were available for 2181 (72%) of all culture-positive cases. Isolates of the Euro-American lineage were most common ($n = 1064$, 48.8%), followed by the East-Asian ($n = 590$, 27.1%), Indo-Oceanic ($n = 517$, 23.7%) and East-African-Indian ($n = 10$, 0.5%) lineages (Figure 2).

Among the 590 East-Asian isolates, 545 (92.3%) had sublineage data. Sex, place of birth (US-born vs. foreign-born), sputum smear examination and chest radiographic abnormalities were similar among patients with and those without sublineage data. The distribution of sublineages among the East-Asian isolates is shown in Figure 2. More than half of the isolates belonged to sublineage 181.

The clinical and demographic characteristics of the patients grouped by sublineage are described in Table 1. The proportion of patients born in the United States was different among the sublineages. Sublineage 150 was more frequent among foreign-born patients, and sublineage 142 among the patients born in the United States. There were 477 foreign-born patients from 24 different countries, the most frequent being China (Hong Kong), with 315 patients. There was no association between the individual countries of origin with any of the sublineages.

The number of patients tested for human immunodeficiency virus (HIV) infection was low, and testing was performed mainly in patients born in the United States. A positive HIV test was most common among those with sublineage 142, 105 and 207. The proportion of cases with extra-pulmonary TB varied with sublineage, with the lowest proportion (5%) among those with sublineage 207. Chest radiographic findings among the different sublineages were not statistically different.

Based on RFLP genotyping, 113 secondary cases were associated with 63 index cases. However, we excluded seven index cases with solely extra-pulmonary disease, as their likelihood of transmitting *M. tuberculosis* is low, and assigned index case status to the next pulmonary case in sequence. Our final sample included 106 secondary cases associated with 62 index cases and 370 cases with unique isolates (Table 2). The median size of the clusters was 3.5 patients for sublineage 105, and 2 for the other sublineages. This cluster size is similar to that observed for the other lineages of *M. tuberculosis* in San Francisco (data not shown). Univariate analysis demonstrated that patients born in the United States were more likely to be secondary cases (OR 7.45, $P < 0.001$) as well as patients with isolates from sublineage 207 when compared with the other sublineages (OR 2.01, $P = 0.04$; Table 3). Neither sputum smear positivity nor the presence of cavitation on chest X-ray was associated with genotypic clustering.

The multivariate analysis was based on just 85 clustered cases (of 106) and 437 observations (sputum smear status was not available in several cases). It showed that the only risk factor for being a secondary case was being born in the United States. The OR of sublineage 207 being a secondary case was 1.93 ($P > 0.05$). The other factors known to be associated with being a secondary case, such as presence of cavities or positive smear examination, were also non-significant in the multivariate analysis (Table 3).

Drug susceptibility testing results were available for 542 isolates (Table 1). There were 57 (10.5%) isolates resistant to isoniazid (INH) and 14 (2.6%) isolates resistant to at least both INH and rifampicin (RMP). Prior treatment for TB was associated with resistance to INH and to INH+RMP ($P < 0.0001$). The proportion of resistance to any drug was not statistically different among the sublineages, but the number in each cell was low (Table 1). Sublineage 207 had the highest proportion of INH-resistant isolates, but this was not statistically significant when compared with the other sublineages. Because studies have reported that drug-resistant *M. tuberculosis* is less associated with clustering,[19] we analyzed the clustering among the different sublineages, excluding drug-resistant cases (data not shown). The univariate and multivariate analyses showed no difference in the associations.

DISCUSSION

M. tuberculosis is a successful pathogen that has infected approximately one third of the world's population. However, only about 10% of those infected develop active TB.[20] Although a number of host and environmental factors play important roles in the transmission of *M. tuberculosis* and the pathogenesis of TB, there is also evidence that the genetic variation of *M. tuberculosis* may influence the likelihood of infection and disease among exposed persons.[21,22] Genotyping methods, such as the use of LSPs, have enabled identification of different lineages and sublineages of *M. tuberculosis*.[1] LSPs represent unique event polymorphisms, and as recombination is a very rare event in *M. tuberculosis*, all the progeny of a strain will have the same LSP. In this study, we sought to determine if different sublineages of the East-Asian lineage of *M. tuberculosis* differed in their ability to cause secondary cases of TB measured by genotypic clustering. We demonstrated that, in the univariate analysis, sublineage 207 was significantly associated with having a higher proportion of secondary cases (being in a genotypic cluster) when compared with other sublineages. However, although the effect size for this difference was unchanged in the multivariate analysis, the loss of observations due to missing sputum smear examination results led to a non-significant *P* value. Therefore, it is tantalizing to speculate that features of the organisms in sublineage 207 may play a role in the higher proportion of secondary cases for this sublineage.

Recently, Hanekom et al. used 40 loci to analyze the genomic structure of the Beijing family in South Africa including single nucleotide polymorphisms, IS*6110* insertion sites and LSPs (RDs 105, 181, 150 and 142).[23] Their phylogenetic tree was rooted in *M. bovis* and had seven lineages, three of which are the same as those described by Tsolaki et al.[3] and used in this paper (sublineages 105, 181 and 150). They found that strains in sublineage 7 (which corresponds to sublineage 150) were more frequently clustered by IS*6110* and spoligotyping. They did not report any cases caused by *M. tuberculosis* sublineage 142, and did not identify any sublineage that corresponded to 207. Furthermore, their study site had a high incidence of HIV infection, a risk factor for developing active TB.[24]

There were only 11 isolates in sublineage 142 (2% of the East-Asian strains in San Francisco), and just one secondary case, suggesting that sublineage 142 is less pathogenic. However, there are no direct data to support this contention, and the available evidence is somewhat contradictory. Supporting the possibility that 142 is less pathogenic is the fact that

it was not found among 595 isolates with the Beijing spoligotype in Western Cape, South Africa,[23] nor in 73 isolates from Tuscany, Italy,[25] 85 isolates from Vietnam[22] or 285 in Japan.[26] In contrast, a study from Arkansas, USA, found that 13 (33%) of 39 strains of the East-Asian lineage belonged to sublineage 142 and that the sublineage was associated with extra-pulmonary TB.[27] Moreover, a recent study showed that 80% of the strains from sublineage 142 (8/10 tested) synthesized detectable quantities of PGL, which suppresses pro-inflammatory cytokines (tumor necrosis factor-alpha, interleukin [IL] 6 and IL12) compared with 0% (0/7 tested) from sublineage 105, 10% (1/10 tested) from sublineage 181, and 60% (6/10 tested) from sublineage 150.[14] The authors tested just one strain from sublineage 207, which did not produce PGL. These findings suggest that sublineage 142 may be more pathogenic.[12] Another explanation for the low number of cases of TB with sublineage 142 is that this sublineage is the most recent of the sublineages in this monophylectic group (RD 142 occurs only in the presence of RD 105, RD 207 and RD 181).[3] As the Beijing family is considered an emerging genotype,[28] it is possible that the lower number of cases of sublineage 142 may be due to the fact that it has not yet had time to proliferate and disseminate. Finally, it is possible that the rate of change of IS*6110* may differ among the sublineages. It is possible that sublineage 142 may have a faster rate compared with sublineage 207, and therefore clustering may be under estimated in 142 and overestimated in 207.

Because there is a low frequency of drug-resistant TB in San Francisco, we could not look for a statistical association between drug resistance and the different sublineages. A previous study did not find associations between drug resistance and sublineages 105, 181 and 150; however, this study did not include strains from sublineages 207 and 142.[23] Another study observed a higher proportion of drug resistance in strains from sublineages 105, 207 and 181; however, no information was provided about other factors associated with drug resistance.[26]

Our study has several limitations: 1) because we used the most conservative definition of secondary cases (identical genotype), it is possible that we are underestimating the clustering across sublineages; and 2) genotyping, which requires isolation of *M. tuberculosis*, cannot be used as a proxy for transmission. The pathogenesis of TB involves three phases: acquisition of infection, (generally) a period of latency, and development of active disease (which occurs in only about 10% of persons). Only transmission events that result in active disease during the observation period will be identified. Consequently, the association of the *M. tuberculosis* sublineages with the generation of secondary cases could be related to a greater propensity for transmission, a relative increase in the ability to cause active TB, or both. We cannot determine which of these processes, if any, is affected by strain factors. The ideal measurement to determine which strain has a higher possibility of causing secondary cases is to measure the rate of appearance of active TB for each sublineage from among those who have been infected. Unfortunately, current methods do not allow us to determine the number of individuals latently infected by each sublineage. 3) The *M. tuberculosis* population structure in the patient's country of origin is unknown. It is possible that a sublineage predominates in a particular part of the world where transmission of *M. tuberculosis* is high, resulting in an apparent difference in the proportion of secondary cases.

In summary, we found evidence in a univariate analysis that the different East-Asian sublineages of *M. tuberculosis* have different propensities for being associated with secondary cases, as indicated by the frequency of genotypic clustering. The effect size for this difference was unchanged in multivariate analysis, although loss of observations due to missing data resulted in a non-significant *P* value. In light of these results, it is tantalizing to hypothesize that the different East-Asian sublineages of *M. tuberculosis* may differ in their capacity to cause secondary cases.

Acknowledgments

References

1. Gagneux S, DeRiemer K, Van T, et al. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc Natl Acad Sci USA. 2006; 103:2869–2873. [PubMed: 16477032]

2. van Soolingen D, Qian L, de Haas PE, et al. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol. 1995; 33:3234–3238. [PubMed: 8586708]

3. Tsolaki AG, Gagneux S, Pym AS, et al. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J Clin Microbiol. 2005; 43:3185–3191. [PubMed: 16000433]

4. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. 2002; 10:45–52. [PubMed: 11755085]

5. Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg Infect Dis. 2002; 8:843–849. [PubMed: 12141971]

6. Ritacco V, Lopez B, Cafrune PI, et al. *Mycobacterium tuberculosis* strains of the Beijing genotype are rarely observed in tuberculosis patients in South America. Mem Inst Oswaldo Cruz. 2008; 103:489–492. [PubMed: 18797764]

7. Frieden TR, Sherman LF, Maw KL, et al. A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA. 1996; 276:1229–1235. [PubMed: 8849750]

8. Johnson R, Warren R, Strauss OJ, et al. An outbreak of drug-resistant tuberculosis caused by a Beijing strain in the Western Cape, South Africa. Int J Tuberc Lung Dis. 2006; 10:1412–1414. [PubMed: 17167961]

9. Kremer K, van-der-Werf MJ, Au BK, et al. Vaccine-induced immunity circumvented by typical *Mycobacterium tuberculosis* Beijing strains. Emerg Infect Dis. 2009; 15:335–339. [PubMed: 19193289]

10. Lopez B, Aguilar D, Orozco H, et al. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. Clin Exp Immunol. 2003; 133:30–37. [PubMed: 12823275]

11. Tsenova L, Ellison E, Harbacheuski R, et al. Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. J Infect Dis. 2005; 192:98–106. [PubMed: 15942899]

12. Reed MB, Domenech P, Manca C, et al. A glycolipid of hyper-virulent tuberculosis strains that inhibits the innate immune response. Nature. 2004; 431:84–87. [PubMed: 15343336]

13. Constant P, Perez E, Malaga W, et al. Role of the pks15/1 gene in the biosynthesis of phenolglycolipids in the *Mycobacterium tuberculosis* complex. Evidence that all strains synthesize glycosylated p-hydroxybenzoic methly esters and that strains devoid of phenolglycolipids harbor a frameshift mutation in the pks15/1 gene. J Biol Chem. 2002; 277:38148–38158. [PubMed: 12138124]

14. Reed MB, Gagneux S, Deriemer K, Small PM, Barry CE III. The W-Beijing lineage of *Mycobacterium tuberculosis* over-produces triglycerides and has the DosR dormancy regulon constitutively upregulated. J Bacteriol. 2007; 189:2583–2589. [PubMed: 17237171]

15. Jasmer RM, Hahn JA, Small PM, et al. A molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991– 1997. Ann Intern Med. 1999; 130:971–978. [PubMed: 10383367]

16. Tsolaki AG, Hirsh AE, DeRiemer K, et al. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc Natl Acad Sci USA. 2004; 101:4865–4870. [PubMed: 15024109]

17. Rhee JT, Tanaka MM, Behr MA, et al. Use of multiple markers in population-based molecular epidemiologic studies of tuberculosis. Int J Tuberc Lung Dis. 2000; 4:1111–1119. [PubMed: 11144452]

18. van Embden JD, Cave MD, Crawford JT, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol. 1993; 31:406–409. [PubMed: 8381814]

19. Burgos M, DeRiemer K, Small PM, Hopewell PC, Daley CL. Effect of drug resistance on the generation of secondary cases of tuberculosis. J Infect Dis. 2003; 188:1878–1884. [PubMed: 14673768]

20. Small PM, Fujiwara PI. Management of tuberculosis in the United States. N Engl J Med. 2001; 345:189–200. [PubMed: 11463015]

21. Nicol MP, Wilkinson RJ. The clinical consequences of strain diversity in *Mycobacterium tuberculosis*. Trans R Soc Trop Med Hyg. 2008; 102:955–965. [PubMed: 18513773]

22. Thwaites G, Caws M, Chau TT, et al. Relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J Clin Microbiol. 2008; 46:1363–1368. [PubMed: 18287322]

23. Hanekom M, van der Spuy GD, Streicher E, et al. A recently evolved sublineage of the *Mycobacterium tuberculosis* Beijing strain family is associated with an increased ability to spread and cause disease. J Clin Microbiol. 2007; 45:1483–1490. [PubMed: 17360841]

24. Selwyn PA, Hartel D, Lewis VA, et al. A prospective study of the risk of tuberculosis among intravenous drug users with human immunodeficiency virus infection. N Engl J Med. 1989; 320:545–550. [PubMed: 2915665]

25. Rindi L, Lari N, Cuccu B, Garzelli C. Evolutionary pathway of the Beijing lineage of *Mycobacterium tuberculosis* based on genomic deletions and *mutT* genes polymorphisms. Infect Genet Evol. 2009; 9:48–53. [PubMed: 18977316]

26. Iwamoto T, Yoshida S, Suzuki K, Wada T. Population structure analysis of the *Mycobacterium tuberculosis* Beijing family indicates an association between certain sublineages and multidrug resistance. Antimicrob Agents Chemother. 2008; 52:3805–3809. [PubMed: 18694954]

27. Kong Y, Cave MD, Zhang L, et al. Population-based study of deletions in five different genomic regions of *Mycobacterium tuberculosis* and possible clinical relevance of the deletions. J Clin Microbiol. 2006; 44:3940–3946. [PubMed: 16957037]

28. Cowley D, Govender D, February B, et al. Recent and rapid emergence of W-Beijing strains of *Mycobacterium tuberculosis* in Cape Town, South Africa. Clin Infect Dis. 2008; 47:1252–1259. [PubMed: 18834315]
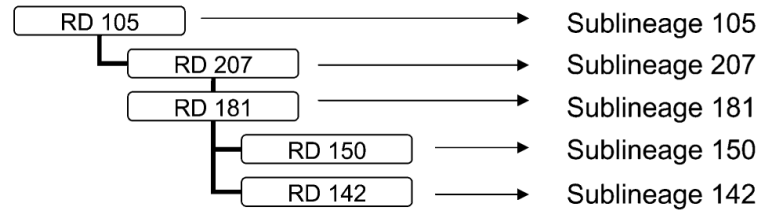
Figure 1.
RD that defines each sublineage of the East-Asian lineage. RD = region of difference.

```
┌─────────────────────────────────┐
│        3585 cases reported       │
└─────────────────────────────────┘
┌─────────────────────────────────┐
│    3044 (85%) culture-positive   │
└─────────────────────────────────┘
┌─────────────────────────────────┐
│  2181 (72%) with RFLP and lineage│
└─────────────────────────────────┘
        ┌──────────────────────────────────┐
        │    1064 (49%) Euro-American       │
        └──────────────────────────────────┘
        ┌──────────────────────────────────┐
        │      590 (27%) East-Asian         │
        └──────────────────────────────────┘
          ┌──────────────────────────────────────┐
          │ 545 (92%) East-Asian with sublineage  │
          └──────────────────────────────────────┘
                    ┌──────────────────────────────┐
                    │   28 (5.3%) sublineage 105    │
                    └──────────────────────────────┘
                    ┌──────────────────────────────┐
                    │   41 (7.3%) sublineage 207    │
                    └──────────────────────────────┘
                    ┌──────────────────────────────┐
                    │  351 (64.4%) sublineage 181   │
                    └──────────────────────────────┘
                    ┌──────────────────────────────┐
                    │  114 (20.9%) sublineage 150   │
                    └──────────────────────────────┘
                    ┌──────────────────────────────┐
                    │   11 (2.0%) sublineage 142    │
                    └──────────────────────────────┘
        ┌──────────────────────────────────┐
        │     517 (24%) Indo-Oceanic        │
        └──────────────────────────────────┘
        ┌──────────────────────────────────┐
        │  10 (<1%) East-African-Indian     │
        └──────────────────────────────────┘
```
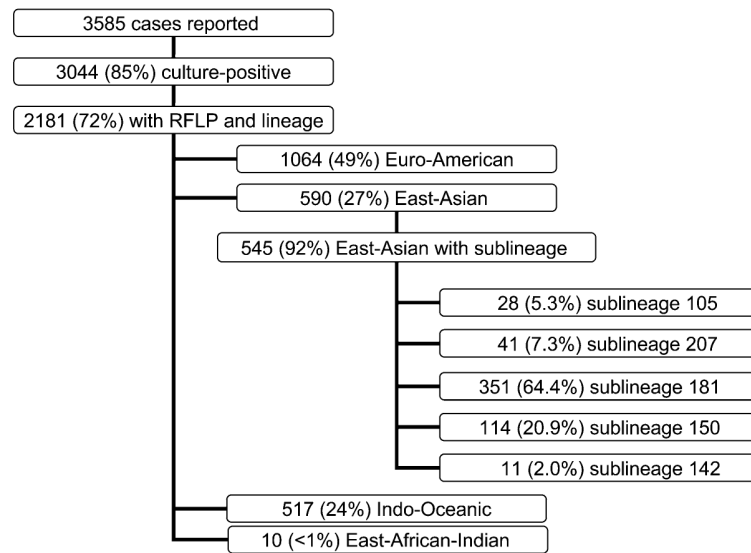
Figure 2.
Frequency of the different lineages and sublineages in San Francisco.

Table 1

Clinical characteristics of patients in the different sublineages

Characteristic	105 (n = 28) n (%)	207 (n = 41) n (%)	181 (n = 351) n (%)	150 (n = 114) n (%)	142 (n = 11) n (%)	P value[*]
Sex						
Female (n = 217)	12 (43)	10 (24)	148 (42)	42 (37)	5 (45)	
Male (n = 328)	16 (57)	31 (76)	203 (58)	72 (63)	6 (55)	0.235
Place of birth						
US-born (n = 68)	5 (18)	10 (24)	42 (12)	7 (6)	4 (36)	
Foreign-born (n = 477)	23 (82)	31 (76)	309 (88)	107 (94)	7 (64)	0.003
Site of TB						
Extra-pulmonary TB (n = 98)[†]	7 (25)	2 (5)	63 (18)	23 (20)	3 (27)	
Pulmonary TB (n = 447)	21 (75)	39 (95)	288 (82)	91 (80)	8 (73)	0.151
Chest radiograph[‡]						
Cavitary (n = 55)	2 (7)	4 (10)	37 (11)	12 (11)	0	
Abnormal non-cavitary (n = 423)	20 (71)	34 (83)	275 (79)	85 (75)	9 (82)	
Normal (n = 63)	6 (21)	3 (7)	36 (10)	16 (14)	2 (18)	0.633
HIV status						
HIV-positive (n = 27)[§]	4 (36)	5 (31)	13 (11)	3 (7)	2 (50)	
HIV-negative (n = 168)	7 (64)	11 (69)	108 (89)	40 (93)	2 (50)	0.004
Smear status						
Sputum smear-negative (n = 278)[¶]	12 (63)	24 (65)	183 (63)	55 (63)	4 (57)	
Sputum smear-positive (n = 162)	7 (37)	13 (35)	106 (37)	33 (38)	3 (43)	0.996
Resistant to						
Isoniazid (n = 57)[#]	1	7	37	12	0	0.4180
Rifampicin (n = 16)[#]	0	3	10	3	0	0.516
Ethambutol (n = 8)[**]	0	0	8	0	0	0.507
Pyrazinamide (n = 6)[††]	0	0	5	1	0	1.000
Multidrug-resistant (n = 14)[#]	0	3	8	3	0	0.353

[*]P value for the distribution of variables among the different sublineages.

[†]Extra-pulmonary TB with no pulmonary involvement.

[‡]Data available for 541.

[§]Data available for 195.

[¶]Data available for 440.

[#]Data available for 542.

[**]Data available for 541.

[††]Data available for 437.

TB = tuberculosis; HIV = human immunodeficiency virus.

Table 2

Frequency of secondary cases among the different sublineages

Sublineage	105 (*n* = 28)	207 (*n* = 41)	181 (*n* = 351)	150 (*n* = 114)	142 (*n* = 11)
Secondary cases, *n* (%) [*]	5 (17.9)	13 (31.7)	68 (19.5)	19 (17.4)	1 (9.1)
OR (95%CI) [†]	0.88 (0.33–2.37)	2.02 (1.01–4.04)	0.96 (0.62–1.49)	0.83 (0.48–1.44)	0.40 (0.05–3.17)
P value [†]	0.801	0.048	0.862	0.504	0.387
Number of clusters	2	7	43	10	1
Cluster size (median) [‡]	3–4 (3.5)	2–6 (2)	2–9 (2)	2–5 (2)	2 (2)

[*] The index case of each cluster is not counted in the number of secondary cases.

[†] Each sublineage compared to all others.

[‡] Cluster size includes the index case.

OR = odds ratio; CI = confidence interval.

Table 3

Univariate and multivariate odds of being a secondary case

Risk factor	Univariate			Multivariate		
	Secondary case *n* (%)	Odds ratio (95%CI)[*]	*P* value	Secondary case *n* (%)	Odds ratio (95%CI)[†]	*P* value
Sublineage						
RD 207	13 (32)	2.01 (1.01–4.04)	0.048	13 (35)	1.93 (0.88–4.25)	0.098
Other	93 (19)			72 (18)		
Birth place						
US	37 (56)	7.45 (4.30–12.9)	<0.001	29 (53)	5.93 (3.22–10.9)	<0.001
Foreign	69 (15)			56 (15)		
Cavities[‡]						
Yes	10 (18)	0.89 (0.43–1.84)	0.770	9 (17)	0.76 (0.33–1.73)	0.515
No	95 (20)			76 (17)		
Sputum smear status[§]						
Positive	37 (23)	1.37 (0.84–2.21)	0.196	37 (23)	1.33 (0.78–2.26)	0.288
Negative	49 (18)			48 (17)		

[*]Excluding 7 extra-pulmonary index cases: 106 secondary cases in 538 observations.

[†]Excluding 7 extra-pulmonary index cases and 101 with missing data: multivariate model composed of 85 secondary cases in 437 observations.

[‡]Excluding 4 missing data on cavitary status: 105 secondary cases in 534 observations.

[§]Excluding 100 missing data on smear status: 86 secondary cases in 438 observations.

CI = confi dence interval; RD = region of difference.