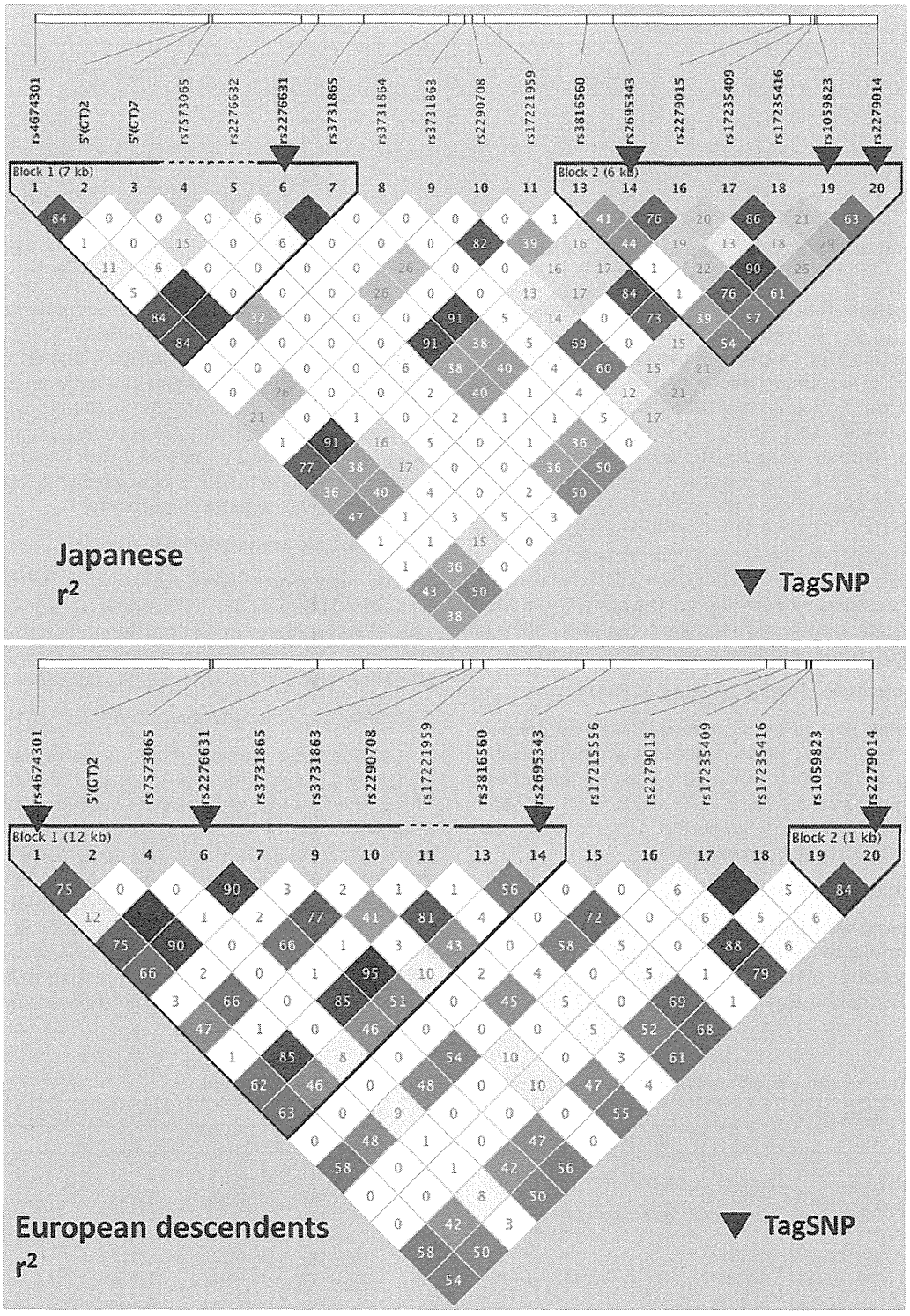


**Fig. 1.** The pair-wise linkage disequilibrium (LD; $r^2$) blocks of the analyzed variants of *SLC11A1* in the controls samples from Japanese and European descendents. Tag single nucleotide polymorphisms (SNPs) are indicated by arrowheads. The strength of LD between any pair of SNPs is represented by the standard color scheme. Shades of gray are in proportion to the $r^2$ value, expressing the strength of the LD.

### 3.2. Association of SLC11A1 SNP with MAC disease

Four tag SNPs of 8 candidates were identified from our data set of the Japanese population, as illustrated in Fig. 1. Three tag SNPs were located in the 3′ LD block and 1 was in the 5′ LD block. Two frequently reported polymorphisms in Asians, rs17235409 (D543N) and rs17235416 (3′ UTR insertion/deletion of TGTG), were also added, despite their rarity. In total, 6 variants (rs2276631, rs2695343, rs17235409, rs17235416, rs1059823, and rs2279014) were analyzed for possible association with the disease and clinical phenotypes using allelic, genotypic, recessive, and dominant models in the Japanese case and control groups. The genotype frequencies in the control group did not deviate significantly from Hardy–Weinberg equilibrium using the $\chi^2$ test with $p < 0.05$ as the cutoff calculated for the level of significance. The minor T allele of rs2279014 was associated with protection from MAC when comparing allele frequencies with an OR of 0.582 (95% CI 0.379 – 0.894,

**Table 2**
Haplotype frequencies in Japanese and European descendents

| Marker variants | | | | | | | | | | | | Haplotype frequency (SD) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs4674301 | rs34448891 | rs2276631 | rs3731865 | rs2290708 | rs3816560 | rs2695343 | rs2279015 | rs17235409 | rs17235416 | rs1059823 | rs2279014 | JP | ED |
| T | (3) | C | G | C | T | G | G | G | (ins) | A | C | **0.6676** (0.0075) | **0.5100** (0.0092) |
| C | (2) | T | C | C | C | A | A | G | (ins) | G | T | 0.0803 (0.0017) | 0.0300 (0.0018) |
| T | (3) | C | G | C | T | A | A | A | (del) | G | T | 0.0540 (0.0061) | 0.0189 (0.0078) |
| T | (7) | C | G | C | T | A | A | G | (ins) | G | C | 0.0400 (0.0000) | (Absent) |
| C | (2) | T | C | T | C | A | A | G | (ins) | G | T | 0.0390 (0.0032) | **0.2126** (0.0065) |

Haplotype structure for the *SLC11A1* gene. Major haplotypes differences are shown in boldface.
SD, estimated standard deviation of haplotype frequency; JP, Japanese population; ED, European descendents.

$p = 0.013$) and this association remained significant in permutation analysis ($perm_{(100,000)}$ $p = 0.044$). The genotype frequency of this SNP also differed in the MAC cases when compared with controls ($p = 0.028$; Table 3). In a dominant allelic model (comparing genotype AA vs Aa/aa), the T allele of rs2279014 was also associated with protection from MAC (OR [95% CI] = 0.511 [0.311–0.839], $p = 0.008$; Table 4). In addition, using logistic regression models, we confirmed that the T allele demonstrated a significant negative association with MAC disease, even after adjustment for sex in the first control panel (OR = 0.523 [0.312–0.877], $p = 0.014$) and for sex and age at diagnosis in the second control panel to assess reproducibility (OR = 0.494 [0.288–0.848], $p = 0.010$). None of the other tag SNPs was associated with altered susceptibility to MAC infection. Taken together, these results suggest that the T allele of rs2279014 is associated with protection against MAC infection.

### 3.3. Genetic polymorphisms in strong LD with rs2279014

To identity possible new variants in strong LD with the disease-associated tag SNP rs2279014, we sequenced the promoter region, coding region, and 3′ UTR of *SLC11A1* in 35 Japanese individuals. Three polymorphisms located in the 3′ UTR, rs17229009 (CAAA/-), rs17229016, and rs13062, were in perfect LD ($r^2 = 1$) with rs2279014 among the 35 Japanese samples.

### 3.4. Subgroup analysis and MAC disease

The subsets of cases were analyzed to determine whether association differs according to age at onset. The cases were stratified considering 50 years old as the average menopausal age of Japanese women and the association was compared. SNP rs2279014 was significantly associated with late-onset patients ($p = 0.031$; Table 5). Further, patients were categorized into 2 groups: with and without comorbidity. In this analysis, higher frequency of the CC genotype was observed in patients without comorbidity (74.6%) compared with healthy controls (56.9%; $p = 0.021$). However, the subgroup with comorbidity did not exhibit significant associations compared with healthy controls. Taken together, these results indicate that rs2279014 SNP is associated with late onset (> 50 years old) and patients without comorbidity.

### 3.5. Haplotypic analysis and MAC disease

The haplotypes were constructed with the 6 variants (rs2276631, rs2695343, rs17235409, rs17235416, rs1059823, and rs2279014) analyzed in the association study and compared, but no significant associations were observed in these haplotype frequencies between cases and controls (data not shown).

### 3.6. Allelic expression imbalance of the SLC11A1 gene

Considering a possible modification of mRNA expression induced by 5′(GT)n or disease-associated polymorphisms in the 3′ UTR of the *SLC11A1* gene, an allele-specific mRNA imbalance assay was performed in 12 heterozygous individuals. Pairwise haplotypes were estimated with 7 polymorphisms: rs34448891, rs2276631, rs2695343, rs17235409, rs17235416, rs1059823, and rs2279014. Among them, 3 polymorphisms (rs17235409, rs17235416, and rs1059823) were included in the PCR amplification for SSCP. First, to elucidate the functional role of the associated SNP marker rs2279014, allelic expression imbalance was evaluated; however, the RT-PCR amplification was not efficient and the

**Table 3**
Associations of SLC11A1 polymorphisms with *Mycobacterium avium* complex (MAC) disease in allelic and genotypic analyses

| Polymorphisms | Allelic model | | | | | Genotypic model | | | | | HWP |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | A (%) | a (%) | OR (95% CI) | $\chi^2$ | $p^a$ | AA (%) | Aa (%) | aa (%) | $\chi^2$ | $p^a$ | |
| rs2276631: C>T | | | | | | | | | | | |
| Control | 356 (84.4) | 66 (15.6) | | | | 150 (71.1) | 56 (26.5) | 5 (2.4) | | | 0.932 |
| MAC | 196 (88.3) | 26 (11.7) | 0.72 (0.44–1.16) | 1.83 | 0.176 | 86 (77.5) | 24 (21.6) | 1 (0.9) | 1.96 | 0.409[b] | |
| rs2695343: G>A | | | | | | | | | | | |
| Control | 283 (67.1) | 139 (32.9) | | | | 97 (46.0) | 89 (42.2) | 25 (11.8) | | | 0.511 |
| MAC | 156 (70.3) | 66 (29.7) | 0.86 (0.61–1.23) | 0.69 | 0.406 | 52 (46.8) | 52 (46.8) | 7 (6.3) | 2.62 | 0.268 | |
| rs17235409: G>A | | | | | | | | | | | |
| Control | 388 (91.9) | 34 (8.1) | | | | 178 (84.4) | 32 (15.2) | 1 (0.5) | | | 0.731 |
| MAC | 211 (95.0) | 11 (5.0) | 0.59 (0.30–1.20) | 2.15 | 0.142 | 101 (91.0) | 9 (8.1) | 1 (0.9) | 3.43 | 0.127[b] | |
| rs17235416: Ins/Del | | | | | | | | | | | |
| Control | 386 (91.5) | 36 (8.5) | | | | 176 (83.4) | 34 (16.1) | 1 (0.5) | | | 0.636 |
| MAC | 211 (95.0) | 11 (5.0) | 0.56 (0.28–1.12) | 2.75 | 0.097 | 101 (91.0) | 9 (8.1) | 1 (0.9) | 4.19 | 0.086[b] | |
| rs1059823: A>G | | | | | | | | | | | |
| Control | 285 (67.5) | 137 (32.5) | | | | 94 (44.5) | 97 (46.0) | 20 (9.5) | | | 0.482 |
| MAC | 163 (73.4) | 59 (26.6) | 0.75 (0.52–1.08) | 2.38 | 0.123 | 59 (53.2) | 45 (40.5) | 7 (6.3) | 2.49 | 0.307[b] | |
| rs2279014: C>T | | | | | | | | | | | |
| Control | 322 (76.3) | 100 (23.7) | | | | 120 (56.9) | 82 (38.9) | 9 (4.3) | | | 0.278 |
| MAC | 188 (84.7) | 34 (15.3) | 0.58 (0.38–0.89) | 6.2 | **0.013**[c] | 80 (72.1) | 28 (25.2) | 3 (2.7) | 7.14 | **0.028**[b] | |

HWP, Hardy–Weinberg equilibrium analyzed in the control group with a cutoff value of $p < 0.05$.
[a]Values in boldface are less than 0.05 for the allelic and genotypic analyses between MAC disease ($n = 111$; 11 samples were not available) and the first control panel ($n = 211$).
[b]Fisher's exact $p$ value.
[c]Permutation test ($n = 100,000$), $p$ value = 0.044.

**Table 4**
Associations of SLC11A1 polymorphisms with MAC disease in dominant and recessive models

| Polymorphisms | Recessive model | | | | | Dominant model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | AA + Aa | aa | OR (95% CI) | $\chi^2$ | $p^a$ | AA | Aa + aa | OR (95% CI) | $\chi^2$ | $p^a$ |
| rs2276631: C>T | | | | | | | | | | |
| Control | 206 (97.6) | 5 (2.4) | | | | 150 (71.1) | 61 (28.9) | | | |
| MAC | 110 (99.1) | 1 (0.9) | 0.37 (0.04–3.25) | 0.86 | 0.669[b] | 86 (77.5) | 25 (22.5) | 0.71 (0.42–1.22) | 1.52 | 0.218 |
| rs2695343: G>A | | | | | | | | | | |
| Control | 186 (88.2) | 25 (11.8) | | | | 97 (46.0) | 114 (54.0) | | | |
| MAC | 104 (93.7) | 7 (6.3) | 0.50 (0.21–1.20) | 2.5 | 0.114 | 52 (46.8) | 59 (53.2) | 0.97 (0.61–1.53) | 0.02 | 0.881 |
| rs17235409: G>A | | | | | | | | | | |
| Control | 210 (99.5) | 1 (0.5) | | | | 178 (84.4) | 33 (15.6) | | | |
| MAC | 110 (99.1) | 1 (0.9) | 1.91 (0.12–30.82) | 0.21 | 1.000[b] | 101 (91.0) | 10 (9.0) | 0.53 (0.25–1.13) | 2.76 | 0.096 |
| rs17235416: Ins/Del | | | | | | | | | | |
| Control | 210 (99.5) | 1 (0.5) | | | | 176 (83.4) | 35 (16.6) | | | |
| MAC | 110 (99.1) | 1 (0.9) | 1.91 (0.12–30.82) | 0.21 | 1.000[b] | 101 (91.0) | 10 (9.0) | 0.50 (0.24–1.05) | 3.48 | 0.062 |
| rs1059823: A>G | | | | | | | | | | |
| Control | 191 (90.5) | 20 (9.5) | | | | 94 (44.5) | 117 (55.5) | | | |
| MAC | 104 (93.7) | 7 (6.3) | 0.64 (0.26–1.57) | 0.95 | 0.401[b] | 59 (53.2) | 52 (46.8) | 0.71 (0.45–1.12) | 2.16 | 0.142 |
| rs2279014: C>T | | | | | | | | | | |
| Control | 202 (95.7) | 9 (4.3) | | | | 120 (56.9) | 91 (43.1) | | | |
| MAC | 108 (97.3) | 3 (2.7) | 0.62 (0.17–2.35) | 0.5 | 0.555[b] | 80 (72.1) | 31 (27.9) | 0.51 (0.31–0.84) | 7.14 | **0.008** |

[a]Values in boldface are less than 0.05 for dominant and recessive models between MAC disease ($n$ = 111; 11 samples were not available) and the first control panel ($n$ = 211).
[b]Fisher's exact $p$ value.

amplicon was not sufficient for further analysis. Therefore, we selected surrogate SNPs (a) rs17235409, rs17235416, and rs1059823 and (b) rs17235409 because they were located near the 5′ end of exon 15. These polymorphisms were amplified efficiently with a sense primer in exon 14 and an antisense primer in exon 15, which generated the optimal length of PCR product for (a) SSCP analysis and (b) RFLP. Importantly, the haplotypes obtained by these surrogate SNPs (GIA/GIG) could easily distinguish C/T alleles of rs2279014 because GIA was observed as the major haplotype (3-CGGIAC) and its expression was compared with that of GIG containing minor haplotypes (2-TAGIGT, 3-CAGIGC, or 7-CAGIGC), but no difference in allele-specific expression of mRNA level was observed. The expression level exhibited by ADG (3-CAADGT, 7-CAADGT) was hard to interpret with the observed dual bands (Suppl. Fig. 1). Therefore, a different assay system was tested, targeting rs17235409 using the RT-PCR-RFLP method. However, the expression level did not differ. Together, these results indicate that the haplotypes tested in this study did not appear to exhibit strong effects on mRNA expression levels in peripheral white blood cells at baseline levels.

## 4. Discussion

In this study, first, to determine the reason behind inconsistent results of disease associations with frequently studied polymorphisms within *SLC11A1*, we investigated the genomic structure of the *SLC11A1* gene and identified variations in LD structures and haplotype frequencies between different ethnic populations. Next, we searched for SNPs exhibiting a stronger association with MAC disease than historical polymorphisms and the newly identified rs2279014 T allele as an SNP exhibiting protection against MAC infection in the Japanese population. However, the functional significance of the polymorphism remains unknown because the allele-specific mRNA imbalance assay did not detect different levels of *SLC11A1* mRNA expression derived from blood cells in our analysis.

Patterns of pairwise LD have indicated considerable variation between different populations. The TGTG insertion/deletion polymorphism, rs17235416, in the 3′ UTR was repeatedly associated with tuberculosis in Asians, but not in Europeans [25,26]. The functional difference between the TGTG insertion allele and deletion allele is totally unknown and may be in LD with a true disease-causing variant. Our result indicated that 3′ LD block structures differed between Japanese and European descendents, and we propose that in Europeans, a tag SNP rs2279014, more representative of the 3′ LD block, should be tested for association studies rather than rs17235416. In addition, the second major haplotype carrying the T allele of rs2290708 in intron 7 as well as allele 2 of GT repeat in the promoter region, with a frequency of 21% in European descendents, was rare in the Japanese population in this study. The 4th frequent haplotype in the Japanese population had allele 7 of 5′(GT)n, which is thought to be present only in Asian populations [33]. Although Bayele et al. demonstrated that allele 3 drove the highest expression level, whereas the other alleles containing alleles 2 and 7 were associated with lower expression [34], the results of association studies with 5′(GT)n were inconsistent [25,26]. The differences in the haplotype structure between populations might influence the disease association with 5′(GT)n alleles. The combination of 5′(GT)n with other potentially functional variations located downstream within the gene may confer compli-

**Table 5**
Association of rs2279014 polymorphism and subgroups of patients with *Mycobacterium avium* complex (MAC) disease

| SNP | Outcome | Genotype frequency (%) | | | $\chi^2$ | $p^a$ |
|---|---|---|---|---|---|---|
| | | CC | CT | TT | | |
| rs2279014: C>T | Control | 120 (56.9) | 82 (38.9) | 9 (4.3) | | |
| | MAC infection, early onset (≤50 years)[b] | 11 (64.7) | 5 (29.4) | 1 (5.9) | 0.63 | 0.595 |
| | MAC infection, late onset (>50 years)[b] | 71 (72.4) | 25 (25.5) | 2 (2.0) | 7.00 | **0.031** |
| | MAC infection with comorbidity | 40 (67.8) | 19 (32.2) | 0 (0.0) | 3.99 | 0.141 |
| | MAC infection without comorbidity | 47 (74.6) | 13 (20.6) | 3 (4.8) | 7.18 | **0.021** |

SNP, single nucleotide polymorphism.
[a]Values calculated by Fisher's exact test and values less than 0.05 are shown in boldface.
[b]Information about age at onset was not available in 7 patients.

cated effects on gene expression and may partly explain discrepancies in the previously reported association studies.

An SNP, rs2279014, located in the 3′ UTR of the reference sequence (NM_000578.3), indicated moderate associations with MAC disease under several analytic models. However, the mechanism by which this polymorphism itself affects the gene expression is unclear because it lies ~680 bp downstream of the putative poly-A signal. It may be possible that this polymorphism is in strong LD with other polymorphisms and may serve as a marker for functional difference. We screened genetic polymorphisms in 35 Japanese individuals and observed 3 polymorphisms in perfect LD with rs2279014. No novel polymorphism in strong LD ($r^2 \geq 0.8$) with rs2279014 was identified. The tag SNP rs2279014 was in perfect LD with rs17229009 (CAAA/-), which is a candidate polymorphism for functional difference, because it was previously reported to be associated with tuberculosis or disease outcome of HIV infection [20,35]. In addition, the association of rs2279014 with MAC disease cannot exclude a possible role of other polymorphisms of nearby genes [36,37].

In the subgroup analysis, the *SLC11A1* polymorphism indicated a relatively strong association with MAC infection without comorbidity. This finding indicates that a genetic predisposition may be predominant in this subgroup, whereas patients with comorbidity may develop the disease, being more affected by extrinsic factors. A significant association with MAC patients was also observed in a subgroup with late onset of the disease. Gene–hormonal interaction may be investigated in the future because menopause, changes in hormonal balance accompanied with aging, has been postulated to play a role in the development of the disease in elderly women [38,39].

In this study, we could not confirm previously reported associations of rs3731865 (INT4) captured by rs2276631 (tag SNP), rs17235409 (D543N), and rs17235416 (3′ UTR). These polymorphisms have previously been reported to be associated with susceptibility to MAC diseases in some studies [10,11], but not in others [40,41]. No significant associations were observed in haplotype frequencies constructed with the 6 variants analyzed in our association study. These negative findings could be caused by low power resulting from the sample size used to detect weak or complicated effects of the variations, like many case–control association studies for rare diseases. Furthermore, the significance of the association was modest. However, the inclusion of 122 well-characterized patients makes this an important study for genetic susceptibility to non-HIV MAC disease.

It was repeatedly reported that allele 3 of 5′(GT)n had higher transcriptional activity than allele 2 [34,42,43]. Considering the reported difference in promoter activity and a hypothesis that 3′ UTR sequences also affect mRNA stability [44], we examined *SLC11A1* mRNA expression levels and possible allelic expression imbalance in the single SNP rs17235409 as well as in the haplotype containing 3 polymorphisms in the 3′ UTR. Detecting allele-specific mRNA expression is a well-established method to measure relative amounts of mRNA generated from each of the 2 alleles within each subject, using a genetic marker in the transcribed region of the gene [45–47]. In our study, no differences in allelic expression imbalance were observed in tested haplotypes and in the single marker. In human tissue, *SLC11A1* is expressed in monocytes/macrophages and polymorphonuclear leukocytes [48–50]. Although the expression of the *SLC11A1* gene may be maximized in tissue macrophages, whereas it may be suboptimal in circulating monocytes and polymorphonuclear leukocytes, we used the peripheral blood in this study to demonstrate the *in situ* allele-specific imbalance because of difficulties in accessing actual pulmonary lesions of MAC infection. Although the present experimental design was intended to detect constitutive levels of gene expression in circulating blood cells, the difference in *SLC11A1* gene expression might be more relevant to the function of activated macrophages at disease sites. Alternatively, the true causative variant, if any, which is in LD with rs2279014, must be sought further and analysis of its functional significance might be required.

In summary, a significant association of *SLC11A1* polymorphisms with pulmonary MAC diseases was observed in the Japanese population. Although these data are suggestive of an association between *SLC11A1* and MAC disease, there are certain limitations of this study, and replication in independent samples is required to verify these findings. Furthermore, gene expression analysis in relation to the associated variant and possibly in the pre- and posttreated samples would further clarify the importance of this variant. Overall, there is a clear requirement of functional analysis targeting MAC lesions to validate these findings.

## Acknowledgments

We thank Ms. Keiko Wakabayashi for her technical assistances. We also thank Dr. Ryo Usui for his experimental arrangement (Tokyo Metropolitan Bokutoh Hospital). BR Sapkota was supported by a fellowship from the Asian Development Bank, Japanese Scholarship Program, during the study period. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Appendix. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.humimm.2012.02.008.

## References

[1] Field SK, Fisher D, Cowie RL. *Mycobacterium avium* complex pulmonary disease in patients without HIV infection. Chest 2004;126:566–81.
[2] Billinger ME, Olivier KN, Viboud C, de Oca RM, Steiner C, Holland SM, et al. Nontuberculous mycobacteria-associated lung disease in hospitalized persons, United States, 1998–2005. Emerging Infect Dis 2009;15:1562–9.
[3] Sakatani M. [Nontuberculous mycobacteriosis; the present status of epidemiology and clinical studies]. Kekkaku 1999;74:377–84.
[4] Vannberg FO, Chapman SJ, Hill AV. Human genetic susceptibility to intracellular pathogens. Immunol Rev 2011;240:105–16.
[5] Sapkota BR, Macdonald M, Berrington WR, Misch EA, Ranjit C, Siddiqui MR, et al. Association of TNF, MBL, and VDR polymorphisms with leprosy phenotypes. Hum Immunol 2010;71:992–8.
[6] Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 2002;20:581–620.
[7] Takahashi M, Ishizaka A, Nakamura H, Kobayashi K, Nakamura M, Namiki M, et al. Specific HLA in pulmonary MAC infection in a Japanese population. Am J Respir Crit Care Med 2000;162:316–8.
[8] Kubo K, Yamazaki Y, Hanaoka M, Nomura H, Fujimoto K, Honda T, et al. Analysis of HLA antigens in *Mycobacterium avium-intracellulare* pulmonary infection. Am J Respir Crit Care Med 2000;161:1368–71.
[9] Um SW, Ki CS, Kwon OJ, Koh WJ. HLA antigens and nontuberculous mycobacterial lung disease in Korean patients. Lung 2009;187:136–40.
[10] Tanaka G, Shojima J, Matsushita I, Nagai H, Kurashima A, Nakata K, et al. Pulmonary *Mycobacterium avium* complex infection: association with NRAMP1 polymorphisms. Eur Respir J 2007;30:90–6.
[11] Koh WJ, Kwon OJ, Kim EJ, Lee KS, Ki CS, Kim JW. NRAMP1 gene polymorphism and susceptibility to nontuberculous mycobacterial lung diseases. Chest 2005;128:94–101.
[12] Shojima J, Tanaka G, Keicho N, Tamiya G, Ando S, Oka A, et al. Identification of MICA as a susceptibility gene for pulmonary *Mycobacterium avium* complex infection. J Infect Dis 2009;199:1707–15.
[13] Yim JJ, Kim HJ, Kwon OJ, Koh WJ. Association between microsatellite polymorphisms in intron II of the human Toll-like receptor 2 gene and nontuberculous mycobacterial lung disease in a Korean population. Hum Immunol 2008;69:572–6.
[14] Marquet S, Lepage P, Hudson TJ, Musser JM, Schurr E. Complete nucleotide sequence and genomic structure of the human NRAMP1 gene region on chromosome region 2q35. Mamm Genome 2000;11:755–62.
[15] Jabado N, Jankowski A, Dougaparsad S, Picard V, Grinstein S, Gros P. Natural resistance to intracellular infections: natural resistance-associated macrophage protein 1 (Nramp1) functions as a pH-dependent manganese transporter at the phagosomal membrane. J Exp Med 2000;192:1237–48.
[16] Blackwell JM, Searle S, Goswami T, Miller EN. Understanding the multiple functions of Nramp1. Microbes Infect 2000;2:317–21.
[17] Govoni G, Vidal S, Cellier M, Lepage P, Malo D, Gros P. Genomic structure, promoter sequence, and induction of expression of the mouse Nramp1 gene in macrophages. Genomics 1995;27:9–19.

[18] Vidal S, Gros P, Skamene E. Natural resistance to infection with intracellular parasites: molecular genetics identifies Nramp1 as the Bcg/Ity/Lsh locus. J Leukoc Biol 1995;58:382–90.
[19] Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med 1998;338:640–4.
[20] Fitness J, Floyd S, Warndorff DK, Sichali L, Malema S, Crampin AC, et al. Large-scale candidate gene study of tuberculosis susceptibility in the Karonga district of northern Malawi. Am J Trop Med Hyg 2004;71:341–9.
[21] Ma X, Wright J, Dou S, Olsen P, Teeter L, Adams G, et al. Ethnic divergence and linkage disequilibrium of novel SNPs in the human NLI-IF gene: evidence of human origin and lack of association with tuberculosis susceptibility. J Hum Genet 2002;47:140–5.
[22] Søborg C, Andersen AB, Madsen HO, Kok-Jensen A, Skinhøj P, Garred P. Natural resistance–associated macrophage protein 1 polymorphisms are associated with microscopy-positive tuberculosis. J Infect Dis 2002;186:517–21.
[23] Gao PS, Fujishima S, Mao XQ, Remus N, Kanda M, Enomoto T, et al. Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International Tuberculosis Genetics Team. Clin Genet 2000;58:74–6.
[24] Delgado JC, Baena A, Thim S, Goldfeld AE. Ethnic-specific genetic associations with pulmonary tuberculosis. J Infect Dis 2002;186:1463–8.
[25] Li HT, Zhang TT, Zhou YQ, Huang QH, Huang J. SLC11A1 (formerly NRAMP1) gene polymorphisms and tuberculosis susceptibility: a meta-analysis. Int J Tuberc Lung Dis 2006;10:3–12.
[26] Li X, Yang Y, Zhou F, Zhang Y, Lu H, Jin Q, et al. SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PLoS ONE 2011;6:e15831.
[27] Tokunaga K, Imanishi T, Takahashi K, Juji T. On the origin and dispersal of East Asian populations as viewed from HLA haplotypes. In: Akazawa T, Szathmary EJE, editors Prehistoric mongoloid dispersals, Oxford: Oxford University Press; 1996.
[28] Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, et al. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 2008;83:445–56.
[29] ATS. Diagnosis and treatment of disease caused by nontuberculous mycobacteria. Medical section of the American Lung Association. Am J Respir Crit Care Med 1997;156:S1–25.
[30] Kishi F. Isolation and characterization of human Nramp cDNA. Biochem Biophys Res Commun 1994;204:1074–80.
[31] Liu J, Fujiwara TM, Buu NT, Sánchez FO, Cellier M, Paradis AJ, et al. Identification of polymorphisms and sequence variants in the human homologue of the mouse natural resistance-associated macrophage protein gene. Am J Hum Genet 1995;56:845–53.
[32] Yen JH, Lin CH, Tsai WC, Ou TT, Wu CC, Hu CJ, et al. Natural resistance-associated macrophage protein 1 gene polymorphisms in rheumatoid arthritis. Immunol Lett 2006;102:91–7.
[33] Kojima Y, Kinouchi Y, Takahashi S, Negoro K, Hiwatashi N, Shimosegawa T. Inflammatory bowel disease is associated with a novel promoter polymorphism of natural resistance-associated macrophage protein 1 (NRAMP1) gene. Tissue Antigens 2001;58:379–84.
[34] Bayele HK, Peyssonnaux C, Giatromanolaki A, Arrais-Silva WW, Mohamed HS, Collins H, et al. HIF-1 regulates heritable variation and allele expression phenotypes of the macrophage immune response gene SLC11A1 from a Z-DNA forming microsatellite. Blood 2007;110:3039–48.
[35] McDermid JM, van der Loeff MF, Jaye A, Hennig BJ, Bates C, Todd J, et al. Mortality in HIV infection is independently predicted by host iron status and SLC11A1 and HP genotypes, with new evidence of a gene–nutrient interaction. Am J Clin Nutr 2009;90:225–33.
[36] Yip SP, Leung KH, Lin CK. Extent and distribution of linkage disequilibrium around the SLC11A1 locus. Genes Immun 2003;4:212–21.
[37] Velez DR, Hulme WF, Myers JL, Stryjewski ME, Abbate E, Estevan R, et al. Association of SLC11A1 with tuberculosis and interactions with NOS2A and TLR2 in African-Americans and Caucasians. Int J Tuberc Lung Dis 2009;13:1068–76.
[38] Han XY, Tarrand JJ, Infante R, Jacobson KL, Truong M. Clinical significance and epidemiologic analyses of *Mycobacterium avium* and *Mycobacterium intracellulare* among patients without AIDS. J Clin Microbiol 2005;43:4407–12.
[39] Tsuyuguchi K, Suzuki K, Matsumoto H, Tanaka E, Amitani R, Kuze F. Effect of oestrogen on *Mycobacterium avium* complex pulmonary infection in mice. Clin Exp Immunol 2001;123:428–34.
[40] Huang JH, Oefner PJ, Adi V, Ratnam K, Ruoss SJ, Trako E, et al. Analyses of the NRAMP1 and IFN-gammaR1 genes in women with *Mycobacterium avium-intracellulare* pulmonary disease. Am J Respir Crit Care Med 1998;157:377–81.
[41] Tanaka E, Kimoto T, Matsumoto H, Tsuyuguchi K, Suzuki K, Nagai S, et al. Familial pulmonary *Mycobacterium avium* complex disease. Am J Respir Crit Care Med 2000;161:1643–7.
[42] Searle S, Blackwell JM. Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. J Med Genet 1999;36:295–9.
[43] Zaahl MG, Robson KJ, Warnich L, Kotze MJ. Expression of the SLC11A1 (NRAMP1) 5-(GT)n repeat: opposite effect in the presence of -237C→T. Blood Cells Mol Dis 2004;33:45–50.
[44] Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. Gene 2001;276:73–81.
[45] Caffrey TM, Joachim C, Paracchini S, Esiri MM, Wade-Martins R. Haplotype-specific expression of exon 10 at the human MAPT locus. Hum Mol Genet 2006;15:3529–37.
[46] Heighway J, Bowers NL, Smith S, Betticher DC, Koref MF. The use of allelic expression differences to ascertain functional polymorphisms acting in cis: analysis of MMP1 transcripts in normal lung tissue. Ann Hum Genet 2005;69:127–33.
[47] Hirota T, Ieiri I, Takane H, Maegawa S, Hosokawa M, Kobayashi K, et al. Allelic expression imbalance of the human CYP3A4 gene and individual phenotypic status. Hum Mol Genet 2004;13:2959–69.
[48] Cellier M, Govoni G, Vidal S, Kwan T, Groulx N, Liu J, et al. Human natural resistance-associated macrophage protein: cDNA cloning, chromosomal mapping, genomic organization, and tissue-specific expression. J Exp Med 1994;180:1741–52.
[49] Cellier M, Shustik C, Dalton W, Rich E, Hu J, Malo D, et al. Expression of the human NRAMP1 gene in professional primary phagocytes: studies in blood cells and in HL-60 promyelocytic leukemia. J Leukoc Biol 1997;61:96–105.
[50] Yoshida T, Kishi F. Expression of NRAMP1 molecule in human peripheral blood leukocytes. Immunol Lett 1997;55:105–8.

# Genetic Diversity and Transmission Characteristics of Beijing Family Strains of *Mycobacterium tuberculosis* in Peru

**Tomotada Iwamoto[1]*, Louis Grandjean[2,3,4], Kentaro Arikawa[1], Noriko Nakanishi[1], Luz Caviedes[4], Jorge Coronel[4], Patricia Sheen[4], Takayuki Wada[5], Carmen A. Taype[6], Marie-Anne Shaw[6], David A. J. Moore[3,4], Robert H. Gilman[4,7]**

**1** Department of Microbiology, Kobe Institute of Health, Kobe, Japan, **2** Wellcome Centre for Clinical Tropical Medicine, St. Mary's Campus, Imperial College, London, United Kingdom, **3** Clinical Research Department and TB Centre, London School of Hygiene and Tropical Medicine, London, United Kingdom, **4** Laboratorio de Investigación de Enfermedades Infecciosas, Universidad Peruana Cayetano Heredia, Lima, Peru, **5** Department of Microbiology, Osaka City Institute of Public Health and Environmental Sciences, Osaka, Japan, **6** Faculty of Biological Science, University of Leeds, Leeds, United Kingdom, **7** Department of International Health, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, Maryland, United States of America

## Abstract

Beijing family strains of *Mycobacterium tuberculosis* have attracted worldwide attention because of their wide geographical distribution and global emergence. Peru, which has a historical relationship with East Asia, is considered to be a hotspot for Beijing family strains in South America. We aimed to unveil the genetic diversity and transmission characteristics of the Beijing strains in Peru. A total of 200 Beijing family strains were identified from 2140 *M. tuberculosis* isolates obtained in Lima, Peru, between December 2008 and January 2010. Of them, 198 strains were classified into sublineages, on the basis of 10 sets of single nucleotide polymorphisms (SNPs). They were also subjected to variable number tandem-repeat (VNTR) typing using an international standard set of 15 loci (15-MIRU-VNTR) plus 9 additional loci optimized for Beijing strains. An additional 70 Beijing family strains, isolated between 1999 and 2006 in Lima, were also analyzed in order to make a longitudinal comparison. The Beijing family was the third largest spoligotyping clade in Peru. Its population structure, by SNP typing, was characterized by a high frequency of Sequence Type 10 (ST10), which belongs to a modern subfamily of Beijing strains (178/198, 89.9%). Twelve strains belonged to the ancient subfamily (ST3 [n = 3], ST25 [n = 1], ST19 [n = 8]). Overall, the polymorphic information content for each of the 24 loci values was low. The 24 loci VNTR showed a high clustering rate (80.3%) and a high recent transmission index ($RTI_{n-1} = 0.707$). These strongly suggest the active and ongoing transmission of Beijing family strains in the survey area. Notably, 1 VNTR genotype was found to account for 43.9% of the strains. Comparisons with data from East Asia suggested the genotype emerged as a uniquely endemic clone in Peru. A longitudinal comparison revealed the genotype was present in Lima by 1999.

**Citation:** Iwamoto T, Grandjean L, Arikawa K, Nakanishi N, Caviedes L, et al. (2012) Genetic Diversity and Transmission Characteristics of Beijing Family Strains of *Mycobacterium tuberculosis* in Peru. PLoS ONE 7(11): e49651. doi:10.1371/journal.pone.0049651

**Editor:** Igor Mokrousov, St. Petersburg Pasteur Institute, Russian Federation

**Received** August 5, 2012; **Accepted** October 11, 2012; **Published** November 21, 2012



**Funding:** This work was supported by Health Science Research grants(H24-SHINKO-IPPAN-011) from Ministry of Health, Labour and Welfare of Japan, MEXT/JSPS KAKENHI Grant Number 24590845, R24 TW007988 NIH FIC, D43 TW006581 NIH FIC, RO1 HD059005 NIH NICHD, WT088559MA with the Wellcome Trust, Wellcome Trust grant 078067/Z/05, a research grant of the Institute for Fermentation, Osaka, Japan, and the US-Japan Cooperative Medical Science Program (TB and Leprosy panel). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

* E-mail: kx2t-iwmt@asahi-net.or.jp

## Introduction

Strains of the Beijing family of *Mycobacterium tuberculosis* (*M. tuberculosis*), first described in 1995 [1], have attracted worldwide attention because of their wide geographical distribution and global emergence. These strains have also been shown to have an endemic prevalence in certain regions, including Asia, except for the Indian subcontinent; northern Eurasia; and South Africa [2,3,4,5,6,7]. The lineage has been reported to have caused major outbreaks worldwide, some involving drug-resistant variants [2,8,9]. These characteristics suggest that strains belonging to this family might have selective advantages (higher virulence or transmissibility) over other *M. tuberculosis* strains [10,11,12,13]. Therefore, a better understanding of the contribution of the Beijing family to the tuberculosis (TB) pandemic is vital to improve global TB control.

The Beijing family is reported to be phylogenetically divisible into 2 main subfamilies: the modern (typical) and ancient (atypical) subfamilies, [14,15]. The modern subfamily is highly prevalent in China, western Russia, South Africa, and Thailand [3,15,16,17], whereas the ancient subfamily, with a deleted region of difference (RD) RD181[-] (late ancient type), is endemic in Japan and the ancient subfamily with RD181[+] (early ancient type) is endemic in Korea [4,18,19,20,21,22,23,24]. Although the reasons for the phylogeographical differences remain elusive, the above trends can be used to assess the influence of the Beijing family strains from East Asian countries, where the prevalence is very high [2,3], on the prevalence of Beijing family strains in other regions. In fact, a

large number of Chinese and Japanese immigrants settled in Peru in 19th century. Korea also has a history of migration to Peru, but this migration has occurred more recently. This historical relationship with East Asia encouraged us to characterize the Beijing family strains in Peru.

Unlike in other South American countries, the proportion of Beijing family strains of *M. tuberculosis* in Peru is exceptionally high [25]. In Brazil, Colombia, Paraguay, Venezuela, Argentina, Chile, and Ecuador the prevalence of Beijing family strains was reported to be less than 1% [25,26,27,28,29]. However, the prevalence in Peru was 5.9% (11/185) in 1999 [25] and 9.3% (30/323) for samples obtained between 2004 and 2006 [30]. Ritacco et al. [25] speculated that the Beijing family strains were first introduced into Peru, and eventually into other South American countries, when Peru received a significant number of Chinese immigrants in the mid-19th century. This same study also showed considerable diversity in the insertion sequence IS*6110* restriction fragment length polymorphism (RFLP) patterns, supporting the concept of earlier introduction(s) of different ancestral strains during the past 150 years. In addition to the importation of Beijing family strains from Asia to Peru, there is also evidence that the Beijing strains were imported into Europe through a South American route, specifically through Peru [31,32]. Therefore, Peru can be considered as the South American country that has been most strongly affected by the introduction of Beijing family and is also most commonly associated with the spread of these strains to other South American and European countries.

An in-depth analysis of Beijing family strains in Peru may have a significant impact on the understanding of global epidemics involving the *M. tuberculosis* Beijing family strains. The aims of the current study were to unveil the genetic diversity and transmission characteristics of Beijing family strains of *M. tuberculosis* in Lima, Peru, and to elucidate the probable impact of past immigration from East Asian countries.

## Materials and Methods

### Ethics Statement

Prior to the start of the study, ethical approval was obtained from both Universidad Peruana Cayetano Heredia and Imperial College London and institutional approval was obtained from the Peruvian Ministry of Health. Samples for this study were anonymized.

### Study Samples

The Beijing family strains used in this study were identified on the basis of deletion of the spacers 1–34 assessed by the spoligotyping assay [33]. In order to negate the inclusion of "Pseudo-Beijing strains" [34], all of the strains classified as the early ancient type of Beijing family (RD 181[+]) were subjected to the RD 105 analysis [35].

A total of 200 Beijing family strains were identified from 2140 *M. tuberculosis* isolates obtained through a population-level implementation of a new diagnostic test (MODS, Microscopic Observation Drug Susceptibility) [36,37], which was conducted in Callao and South Lima between December 2008 and January 2010. Strains from all culture-positive patients with respiratory symptoms in the study area were included and none of the samples were duplicated from a single patient. Of the 200 Beijing family strains, 2 were excluded from this study because of insufficient DNA samples. For the remaining 198 Beijing strains, detailed information is provided in Tables 1 and S1.

An additional 70 Beijing family strains isolated in Lima between 1999 and 2006 were also analyzed in order to make a longitudinal comparison. In detail, 26 were obtained from a previous study in North Lima between 2004 and 2006 [30] and 44 were obtained from 4 distinct areas in Lima (North, South, East, and Central Lima) through the hospital based studies between 1999 and 2004. The details of the strains are described in Tables 1 and S2.

### VNTR Typing

Genotypic data for the 24 loci that comprised the international standard set of 15 loci of variable number of tandem repeat(s) of mycobacterial interspersed repetitive units (15-MIRU-VNTR) [38], and 9 additional loci (2074, 2372, 3155, 3336, 3232, 3820, 4120, QUB11a, QUB18) were analyzed. The 9 additional loci were selected as Beijing-type optimized loci because of their highly discriminatory values in different studies focused on Beijing family strains [4,24,39,40,41,42,43]. This 24-loci VNTR was called the $24_{\text{Beijing}}$-VNTR. The polymerase chain reaction (PCR) primers and the number of repeats for each locus, based on the *M. tuberculosis* H37Rv strain, are described in Table S3; the PCR conditions were as described previously [4]. The amplicon samples were diluted 20-fold with ultrapure water and analyzed on an AB3500 genetic analyzer system (Applied Biosystems, Foster City, CA) at a constant room temperature of 25°C following the manufacturers' instruction, i.e., capillary temperature, 60°C; electrophoresis voltage, 8.5 kV; and separation time, 5800 s. A GeneScan 1200 LIZ Size Standard (Applied Biosystems) was used to provide internal size markers. Fragment sizes were measured using GeneMapper Ver. 4 (Applied Biosystems). The numbers of repeats at each locus were calculated using the offset values of the size, which correct differences in relative migration between the size standard and the amplicons depending on the locus. The reproducibility and accuracy of size calling and the size offsets were checked by including *M. tuberculosis* H37Rv and 1 Beijing family strain (reference for quality control) into every batch of the analysis (one 96-well plate was used as a batch). For the large alleles (specifically, for the one larger than 1000 bases) of a locus, we used stutter peak counting, as shown in Figure S1, to obtain unambiguous results with high reproducibility [4]. Moreover, we confirmed the reproducibility of our assay by blindly re-testing 22 selected samples (Table S1). The allelic diversity of each VNTR locus was evaluated using Nei's diversity index [44], i.e., the polymorphic information content (PIC) (Table S4). Genotypic discrimination of the 198 Beijing strains, based on 15-MIRU-VNTR and $24_{\text{Beijing}}$-VNTR, were calculated using the Hunter-Gaston discriminatory index (HGDI) [45]. A recent transmission index ($RTI_{n-1}$) [46,47] was also calculated using the VNTR profiles. A minimum spanning tree (MST), based on VNTR types, was constructed using Bionumerics software (Bionumerics ver. 4.2; Applied Math., Sint-Martens-Latem, Belgium), as previously described [23].

### Single Nucleotide Polymorphisms at 10 Loci

The sequence types (ST) were determined based on the 10 synonymous SNPs, which were sufficient to divide the Beijing strains obtained from the global population [48]. Each chromosomal position in the whole-genome sequence of H37Rv [49] was as follows: 797736, 909166, 1477596, 1548149, 1692069, 1892017, 2376135, 2532616, 2825581, and 4137829. The SNP at position 1477596 is the same as *ogt*12, which discriminates between the ancient and modern type Beijing strains [16]. The SNP at 2532616 is the same as *adhE2* (codon 124), which further discriminates the modern type [50]. SNPs at 797736 and 2825581

**Table 1.** Patient demographics for the Beijing family strains in this study.

| Characteristics | No. (%) of isolates | |
|---|---|---|
| | Between 2008 and 2010 | Between 1999 and 2006 |
| Total | 198 (100) | 70 (100) |
| Sex | | |
| Male | 133 (67) | 45 (64) |
| Female | 63 (32) | 25 (36) |
| Unknown | 2 (1) | 0 (0) |
| Age group | | |
| <25 | 85 (43) | 19 (27) |
| 25–34 | 45 (23) | 17 (24) |
| 35–44 | 32 (16) | 8 (11) |
| 45–54 | 10 (5) | 3 (4) |
| 55–64 | 5 (3) | 0 (0) |
| 65+ | 8 (4) | 2 (3) |
| Unknown | 13 (7) | 21 (30) |
| Previous TB | | |
| Yes | 55 (28) | ND |
| No | 141 (71) | ND |
| Unknown | 2 (1) | 70 (100) |
| HIV status | | |
| Positive | 9 (5) | 9 (13) |
| Negative | 188 (95) | 35 (50) |
| Unknown | 1 (1) | 26 (37) |
| *M. tuberculosis* resistance | | |
| MDR | 17 (9) | 10 (14) |
| not MDR | 177 (89) | 47 (67) |
| Unknown | 4 (2) | 13 (19) |

doi:10.1371/journal.pone.0049651.t001

can discriminate between the early ancient (RD181[+]) and late ancient (RD181[-]) Beijing strains [19]. Polymorphic nucleotides of the respective isolates were determined according to Hanekom et al. [17]. STs were designated according to Filliol et al. [48] and Iwamoto et al. [19].

### Data Retrieved from Previous Reports

15-MIRU-VNTR data and the results of subfamily classifications were retrieved from previous publications for the isolates in China [50], Japan [18], and Korea [20]. In order to compare the population structure, based on Filliol's STs, further sublineage classifications for the Chinese and Korean isolates were determined according to the phylogenetically informative VNTR loci data proposed by Wada and Iwamoto [51]. The VNTR data from Peru and these 3 East Asian countries were compared by constructing MST, based on 15-VNTR-MIRU, as previously described [23].

### Statistical Analysis

Fisher's exact test (SPSS 17.0; IBM, NY, USA) was used to determine the association of large cluster-forming isolates with patient gender, history of TB treatment, HIV status, and strain drug susceptibility; the association with the median age of patients was evaluated by Mann-Whitney *U*-test using PASW statistics 18 (IBM).

## Results

### Proportion of Beijing Family Strains in Peru

The Beijing family strain, with a prevalence of 9.3% (200/2140), was the third largest spoligotyping clade after the H3 and T1 clades in the study population sampled between 2008 and 2010. This ratio is the same as for the previous study (9.3%, 30/323) conducted in 2004 and 2006 [30] but higher than in 1999 (5.9%, 11/185) [25]. In another study setting, which collected Beijing strains between 1999 and 2004 from 4 distinct areas in Lima, the ratio was 5% (46/912, 2 Beijing strains were excluded in this study because of the lack of DNA). This would imply an increasing trend of Beijing family strains in Lima, Peru. We negated the existence of "Pseudo-Beijing strains" [34] in our sample set by the detection of RD 105 [-] [35] for all of the strains classified as early ancient type by SNP typing.

### Classification of Beijing Family Strains by SNP Typing and MIRU-VNTR

Further subdivision of the Beijing family strains in Peru, according to the 10 loci SNP panel [17,19,48], revealed similar population structures between the 2 sample sets (2008–2010 and 1999–2006) (Table 2). The high prevalence of isolates associated with the modern subfamily in Peru is consistent with the worldwide trend, except in the case of Japan and Korea where

the ancient subfamily predominates [20,50,51,52,53,54]. When we compared the population structure of the Beijing family strains in Peru with China, Japan, and Korea (Table 2), the characteristics of the Peruvian population structure highlighted a high frequency of the ST10 sublineage. ST19/ST25, which evolved before the modern subfamily [16,51], predominated in the ancient subfamily, as has been seen in China. Recently, Wada et al. proposed the use of SNP at 1576481 instead of 909166 due to the probable homoplastic behavior of SNP 909166 [55]. This is especially critical for discriminating between STK and ST3, which are highly prevalent in Japan. We, therefore, applied it for 4 ST3 stains in our sample set and confirmed them as true ST3 sublineage. Interestingly, the phylogenetic informativity of certain VNTR loci reported in Asian strains [16,51] was also retained for the strains in Peru (Table 3). Specifically, VNTR 4156, 1955, and 3155 demonstrated high sensitivity and specificity for the sublineages classification.

The 15-MIRU-VNTR data from the modern subfamily of Beijing strains in China, Korea, and Japan were retrieved from previous reports [18,20,50] and compared with the data of the population-based study in Peru (2008–2010) (Fig. 1). Two extremely large clusters were found in the MST, one (n = 103) composed of almost exclusively Peruvian isolates (101 Peruvian isolates and 2 Japanese isolates) (C1 in Fig. 1) and the other (n = 88) composed of isolates from all of the 4 countries (C2 in Fig. 1). None of the large clusters were composed exclusively of Japanese, Chinese, or Korean isolates. The biggest cluster, C1, did not show a star-like network with the others; rather, it had a nearly terminal topological position. In addition, the VNTR profile could not be found in the MIRU-VNTR*plus* database [56] or in other reports from Russia and China [57,58]. Together, these observations suggested that this genotype emerged as a uniquely endemic clone in Peru. The other large cluster (C2) was found at the core position and was connected with many isolates, regardless of their geographical origin. Because the Peruvian isolates belonging to the C2 cluster were further subdivided into many different genotypes by the $24_{Beijing}$-VNTR analysis (data not shown), it is clear that the large cluster was formed because of the convergence of VNTR profiles. Other than these 2 large clusters, all of the other isolates dispersed equally (Fig. 1), i.e., no other branches consisted solely of Peruvian isolates.

## Allelic Diversity of 24 VNTR Loci in Peru

The allelic diversity of each of the 24 VNTR loci is listed in Table S4, including the data from a previous report for Russia, Japan, and China [43]. Overall, the allelic diversity of the VNTR loci in Peru was much lower than in China or Japan. This low diversity was similar to that observed in Russia, where the Beijing family strains are considered to have recently emerged [57,59]. Only 11 out of the 24 loci had a PIC value greater than 0.1. Of these loci, 7 were in the 9 additional VNTR loci described in the Methods section. The PIC for the 3 hypervariable loci [4] were high: VNTR 3232 (0.710), VNTR 3820 (0.592), and VNTR 4120 (0.495). These results clearly supported the necessity of the use of the 9 additional loci for improving the discriminating power of VNTR genotyping for these strains. As expected, the use of $24_{Beijing}$-VNTR improved discrimination compared to that of 15-MIRU-VNTR (Table 4).

## Transmission Characteristics of Beijing Family Strains

Although $24_{Beijing}$-VNTR improved the power of strain discrimination, the results showed high levels of clustering (80.3%) and $RTI_{n-1}$ (0.707) (Table 4). This strongly suggests the active and on-going transmission of Beijing family strains in the survey area. Notably, a large size clustering (43.9%, 87/198) was identified, which was named Peru Cluster Type 001 (PCT001) (Fig. 2). The PCT001 strains belong to the largest cluster (C1) in 15 MIRU-VNTR (Fig. 1), therefore, they are considered a singular genotype in Peru. In the MST (Fig. 2), it formed a star-like VNTR-based network with the probable derivatives from it, mainly resulting from single locus changes. The results suggest the continuous evolution of the clone through active transmission and thus being a "currently successful clone" of the Beijing family strains in Peru. The second and third largest clusters were much smaller than PCT001:15 isolates (7.6%) and 10 isolates (5.1%), respectively. Although the number of multidrug-resistant *M. tuberculosis* (MDR-TB) belonging to the PCT001 cluster is small at the moment (n = 6), this should be an addressed as a potential public health threat. The comparison between patients harboring the PCT001 strains and those of other patients harboring Beijing strains revealed that only gender is significantly different ($P = 0.02$) (Table 5).

To ascertain whether the PCT001 genotype strains were present before the surveillance periods (in 2008–2010), an MST

**Table 2.** Distribution of Beijing sublineage strains.

| SNP type | No. (%) of isolates | | | | | Definition[2] |
|---|---|---|---|---|---|---|
| | Peru | | Japan | China[1] | Korea[1] | |
| | 2008–2010 (n = 198) | 1999–2006 (n = 70) | ref [18] (n = 714) | ref [50] (n = 187) | ref [20] (n = 62) | |
| ST11 | 0 (0) | 0 (0) | 3 (0.4) | 9 (4.8) | 29 (46.8) | Early ancient |
| ST26 | 0 (0) | 4 (5.7) | 50 (7.0) | | | |
| STK | 0 (0) | 0 (0) | 111 (15.5) | 0 | 2 (3.2) | Late ancient |
| ST3 | 3 (1.5) | 1 (1.4) | 182 (25.5) | 3 (1.6) | 3 (4.8) | |
| ST25 | 1 (0.5) | 0 (0) | 6 (0.8) | 52 (27.8) | 10 (16.1) | |
| ST19 | 8 (4.0) | 5 (7.1) | 195 (27.3) | | | |
| ST10 | 178 (89.9) | 53 (75.7) | 135 (18.9) | 93 (49.7) | 18 (29.0) | Modern |
| ST22 | 8 (4.0) | 7 (10.0) | 32 (4.5) | 30 (16.0) | | |

[1]Two isolates in Korea and 4 isolates in China could not be assigned sublineages and excluded from the analysis.
[2]Early ancient, RD181 [+]; Late ancient, RD181 [−].
doi:10.1371/journal.pone.0049651.t002

**Table 3.** Specific VNTR allele for Beijing sublineages *M. tuberculosis* in Peru (n = 268).

| VNTR locus | Specific allele | Corresponding sublineage(s) | No. of isolates | Sensitivity (%) | Specificity (%) |
|---|---|---|---|---|---|
| 4156 | 3 | Modern | 246 | 246/246 (100) | 246/248 (99.2) |
| | 4 | Early Ancient | 4 | 4/4 (100) | 4/4 (100) |
| | 5 | Late Ancient | 16 | 16/16 (100) | 16/18 (88.9) |
| 1955 | 4 | Modern | 246 | 235/246 (95.5) | 235/235 (100) |
| 3155 | 2 | Early Ancient | 4 | 4/4 (100) | 4/4 (100) |

doi:10.1371/journal.pone.0049651.t003

was constructed based on the $24_{Beijing}$-VNTR profiles of the 268 Beijing family strains, comprised of the 198 strains (2008–2010) and the additional 70 strains isolated between 1999 and 2006 (Fig. 3). Twelve strains, which formed the second largest cluster among the 70 Beijing family strains, belonged to PCT001 (1 in 1999, 2 in 2000, 4 in 2001, 2 in 2004, and 3 in 2005), indicating that the genotype had existed since at least 1999. A striking difference between the 2 sample sets was found in the ratio of PCT001 genotype strains to total Beijing strains, i.e., a larger cluster (13 strains) than PCT001 was found in the 70 Beijing strains (PCT002 in Fig. 3). When we just focused on the strains isolated in South Lima (24/70 [34.3%]), the overlapping area of these 2 sample sets, PCT002 was still identified as the largest cluster (7 strains) followed by PCT001 (6 strains). These results suggested that the prevalence of PCT001 in the survey area increased over a short period.

## Discussion

The high prevalence of Beijing family strains in Peru (5.9%, 11/185) compared to other South American countries was first reported as the result of a survey of 7 countries on the continent [25]. Very recently, Taype et al. [30] reported that the proportion of TB patients with the Beijing family strains in Peru was 9.3% (30/323). The report nicely described the genetic diversity of all the *M. tuberculosis* in Peru. However, the population size was too

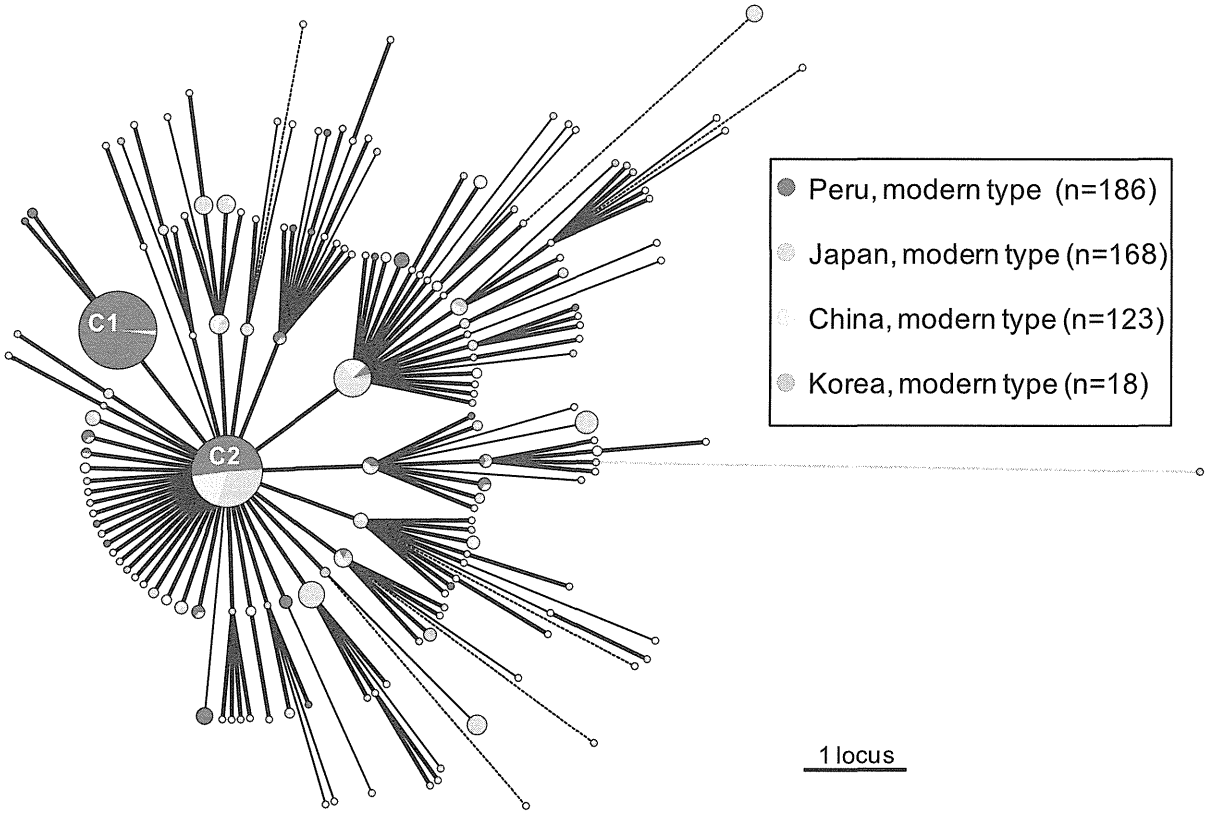


**Figure 1. A minimum spanning tree based on 15 loci of a variable number of tandem repeat(s) of mycobacterial interspersed repetitive units (15-MIRU-VNTR) genotyping of the modern subfamily of *M. tuberculosis* Beijing strains from Peru (n = 186), Japan (n = 168), China (n = 123), and Korea (n = 18).** Circles correspond to the different types discriminated by 15-MIRU-VNTR genotypes. Their sizes are proportional to the numbers of isolates sharing an identical pattern. The origin of each isolate is represented by different colors. Heavy lines connecting 2 types denote single-locus variants; thin lines connect double-locus variants; and dotted lines (black), triple-locus variants. The gray dotted lines indicate the most likely connection between 2 types differing by more than 3 VNTR loci.
doi:10.1371/journal.pone.0049651.g001

**Table 4.** Clustering analysis of 198 Beijing family strains (2008–2010) in Peru.

| | No. of patterns | No. of clusters | No. clustered isolates | Clustering rate (%) | $RTI_{n-1}$ | HGDI |
|---|---|---|---|---|---|---|
| 15 VNTR | 31 | 14 | 181 | 91.4 | 0.843 | 0.688 |
| 24 VNTR | 58 | 19 | 159 | 80.3 | 0.707 | 0.797 |

doi:10.1371/journal.pone.0049651.t004

small to investigate the Beijing family strains in detail. In the current study, with a much larger population size and a well-designed sample collection scheme between 2008 and 2010, the genetic diversity and transmission dynamics of Beijing strains in Peru was clarified. The results highlighted the following: 1) an increasing prevalence of Beijing family strains during a past decade in Peru, 2) high clonality of the Beijing family strains, suggesting active and ongoing transmission, 3) the successful clone "PCT001 genotype," which has existed since at least 1999 as a singular clone in Peru, and 4) China as the greatest contributor of imported Beijing family strains into Peru. This study is limited by being unable to determine whether the highly active, ongoing transmission of Beijing family strains is specific for the family or a common trend for all strains in Peru. However, we could find a similarly high clustering rate of Beijing family strains (24/30 [80%]), compared with that of the non-Beijing family strains (174/293 [59.4%]) in the previous study using 12-MIRU-VNTR and spoligotyping [30]. This higher clustering in Beijing family strains implies a high transmission of it in Peru (more successful than non-Beijing strains). This trend could be confirmed later on by further VNTR analysis of all of the 2140 isolates.

A longitudinal comparison with the 70-strain sample set collected between 1999 and 2006 suggested that the successful clone, PCT001, was already present in Lima, by 1999, but at a lower prevalence (n = 12 [17.1%]) (Fig. 3). This difference in the ratio of PCT001 genotype strains to the total Beijing strains between the 2 sample sets suggests that the increase in PCT001 prevalence occurred recently, over a relatively short period. HIV infection, one of the high-risk characteristics for a large-scale outbreak [2], did not explain this high prevalence of PCT001 strains in the survey area (Table 5). One of the possible explanations for the high prevalence of PCT001 strains could be that it is highly transmissible and/or has increased virulence. The highly prevalent strains from large clusters have been previously reported to be more virulent than the sporadic strains of lower prevalence [59,60]. An attractive hypothesis is that PCT001 strains gained a selective advantage that allowed them to have spread more easily between the 2 sample collection periods (1999–2006 and 2008–2010).

The 15-MIRU-VNTR data from the modern subfamily of Beijing strains in China, Japan, and Korea were retrieved from previous reports [18,20,50] for comparison with the Peruvian isolates. The topology of the MST, based on these data (Fig. 1),

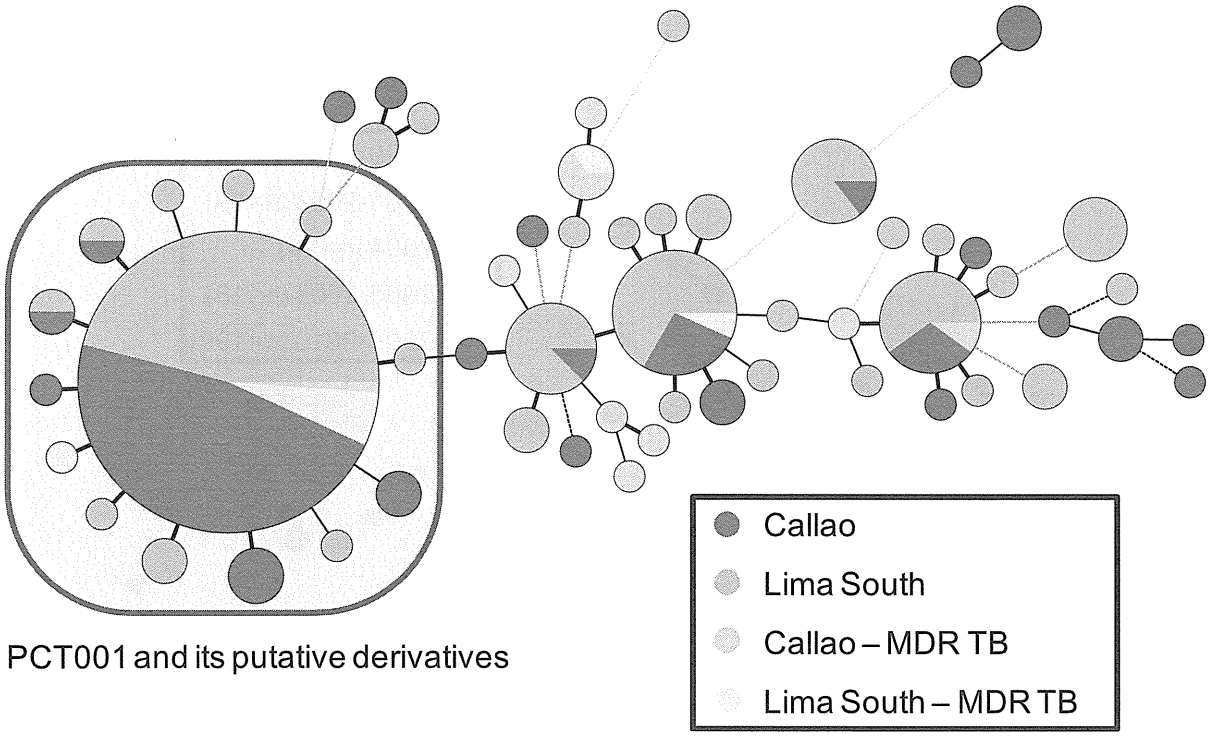


**Figure 2. A minimum spanning tree of 198 Beijing family strains from Peru based on the 24-loci variable number of tandem repeats (VNTR).** The colors of the circles represent the areas where the strain was isolated and its multidrug-resistant status. The designations for each circle and line in the tree are the same as in Fig. 1.
doi:10.1371/journal.pone.0049651.g002

**Table 5.** Demographics of successful clone "PCT001" and other strains.

| Variable | No. of strains (%) | | *P* value |
|---|---|---|---|
| | PCT001 (n=87) | Others (n=111) | |
| Gender | | | |
| Male | 67 (77) | 66 (59) | 0.02 |
| Female | 20 (23) | 43 (39) | |
| Unknown | 0 (0) | 2 (2) | |
| Median age (range) | 23.5 (12–69) | 27 (12–83) | 0.07 |
| Unknown | 4 (5) | 9 (8) | |
| Previous TB | | | 1 |
| Yes | 24 (28) | 31 (28) | |
| No | 63 (72) | 78 (70) | |
| Unknown | 0 (0) | 2 (2) | |
| HIV status | | | 0.30 |
| Positive | 2 (2) | 7 (6) | |
| Negative | 85 (98) | 103 (93) | |
| Unknown | 0 (0) | 1 (1) | |
| MDR TB | | | 0.45 |
| Yes | 6 (7) | 11 (10) | |
| No | 81 (93) | 96 (86) | |
| Unknown | 0 (0) | 4 (4) | |

doi:10.1371/journal.pone.0049651.t005

suggested that the PCT001 genotype was a uniquely endemic clone in Peru. Except for the 2 largest clusters, the VNTR patterns from these countries were similarly diverse and none of the branches was composed exclusively of Peruvian isolates. This result would further support the earlier introduction of different ancestral strains into Peru from Asia during the past 150 years, an idea that was proposed on the basis of the diversity in IS*6110* RFLP patterns in a previous study [25].

The 10 loci SNPs could identify the modern Beijing lineage ST10 as the predominant sublineage in Peru. Very recently, this sublineage was reported as the most common in Taiwan and Thailand [16,61]. Allelic distribution of VNTR loci in each sequence type of Peruvian samples revealed the phylogenetic informativity of 3 MIRU-VNTR loci (4156, 1955, 3155) (Table 3). This is consistent with the results for East Asian strains [51] and could be an evidence for sharing the common ancestors of the Beijing family strains in Peru with those in East Asian countries. With the tremendous increase in whole-genome sequencing data, an increasing number of SNP typing systems have been developed [54,62,63]. The 10 loci SNPs typing is useful for a classification of Beijing family strains into sublineages level but apparently needed more discriminatory power to compensate for the VNTR homoplasy effect (VNTR-based clusters with mixed SNP sublineages), which was demonstrated with optical sets of 8 SNPs in Shanghai [50]. Moreover, inclusion of the lineage specific SNPs for the clades other than Beijing family [62,63] would expand the potential for phylogenetic studies. Further elaboration and optimization of the SNP sets would facilitate future molecular epidemiology and phylogenetic studies on *M. tuberculosis* in Peru.

In conclusion, the current results revealed the predominance of the modern subfamily and active transmission of Beijing family strains within Peru. Moreover, the emergence of the highly prevalent strains with the PCT001 genotype was also detected. The importation of Beijing family strains into European countries from Peru has already been reported [31,32], raising concern over transnational transmission. Future trends regarding the prevalence of PCT001 strains and the changes in population structure need to be carefully monitored from both the local and global epidemiological standpoints.

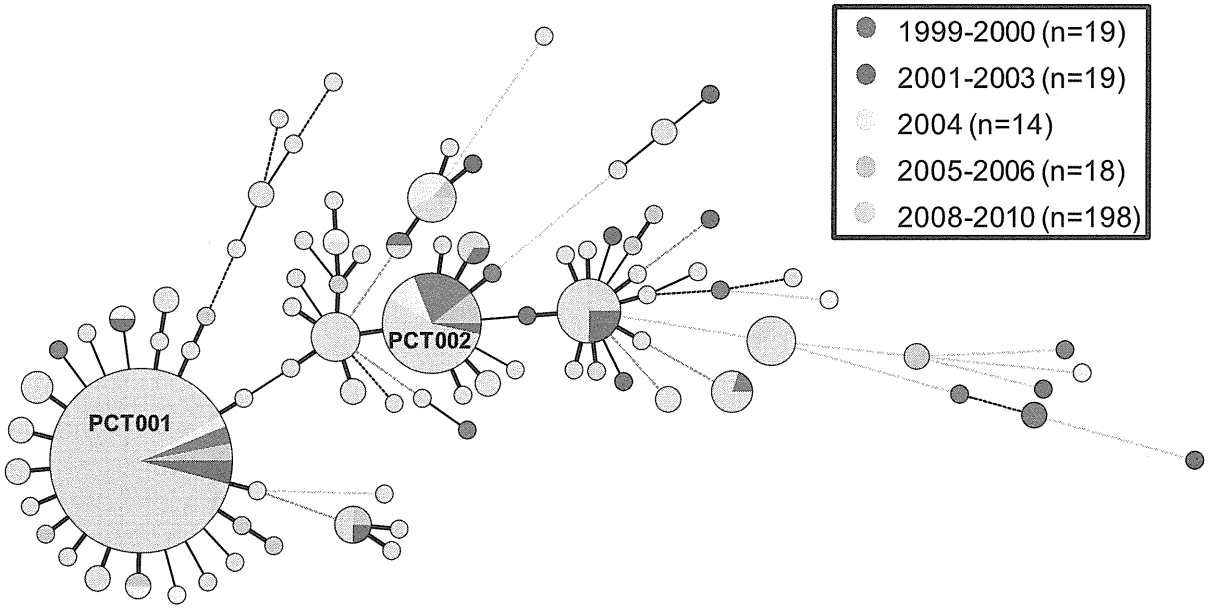


**Figure 3. A minimum spanning tree of 268 Beijing family strains comprised of the 198 strains from the population-based study between 2008 and 2010 and an additional 70 strains isolated between 1999 and 2006.** The colors of the circles represent the years of isolation. The designations for the circles and lines in the tree are the same as in Fig. 1.
doi:10.1371/journal.pone.0049651.g003

## Supporting Information

**Figure S1 Quality control approach for VNTR.**
(PDF)

**Table S1 Description of *M. tuberculosis* Beijing family strains obtained between December 2008 and January 2010 in Lima.**
(XLS)

**Table S2 Description of the additional 70 *M. tuberculosis* Beijing family strains in Lima, Peru.**
(XLS)

**Table S3 Locus designations and PCR primer sequences of the VNTR locus.**
(XLS)

**Table S4 Allelic diversity of VNTR loci.**
(XLS)

## Acknowledgments

We thank D. Sara, J.B. Phu, P. Fuentes, W. Solano, and T. Ichijo for excellent technical assistance.

## Author Contributions

Conceived and designed the experiments: TI LG DM RG. Performed the experiments: TI LG KA NN JC LC PS. Analyzed the data: TI LG KA NN RG. Contributed reagents/materials/analysis tools: JC LC PS TW CT MS DM RG. Wrote the paper: TI LG CT MS RG.

## References

1. van Soolingen D, Qian L, de Haas PE, Douglas JT, Traore H, et al. (1995) Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J Clin Microbiol 33: 3234–3238.
2. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN (2002) Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol 10: 45–52.
3. Glynn JR, Whiteley J, Bifani PJ, Kremer K, van Soolingen D (2002) Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg Infect Dis 8: 843–849.
4. Iwamoto T, Yoshida S, Suzuki K, Tomita M, Fujiyama R, et al. (2007) Hypervariable loci that enhance the discriminatory ability of newly proposed 15-loci and 24-loci variable-number tandem repeat typing method on *Mycobacterium tuberculosis* strains predominated by the Beijing family. FEMS Microbiol Lett 270: 67–74.
5. Kremer K, Glynn JR, Lillebaek T, Niemann S, Kurepina NE, et al. (2004) Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. J Clin Microbiol 42: 4040–4049.
6. Mokrousov I (2008) Genetic geography of *Mycobacterium tuberculosis* Beijing genotype: a multifacet mirror of human history? Infect Genet Evol 8: 777–785.
7. Wan K, Liu J, Hauck Y, Zhang Y, Zhao X, et al. (2011) Investigation on *Mycobacterium tuberculosis* diversity in China and the origin of the Beijing clade. PLoS One 6: e29190.
8. Devaux I, Kremer K, Heersma H, Van Soolingen D (2009) Clusters of multidrug-resistant *Mycobacterium tuberculosis* cases, Europe. Emerg Infect Dis 15: 1052–1060.
9. Affolabi D, Faihun F, Sanoussi N, Anyo G, Shamputa IC, et al. (2009) Possible outbreak of streptomycin-resistant *Mycobacterium tuberculosis* Beijing in Benin. Emerg Infect Dis 15: 1123–1125.
10. Parwati I, van Crevel R, van Soolingen D (2010) Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. Lancet Infect Dis 10: 103–111.
11. Parwati I, Alisjahbana B, Apriani L, Soetikno RD, Ottenhoff TH, et al. (2010) *Mycobacterium tuberculosis* Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia. J Infect Dis 201: 553–557.
12. Gagneux S, Small PM (2007) Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect Dis 7: 328–337.
13. Reed MB, Gagneux S, Deriemer K, Small PM, Barry CE 3rd (2007) The W-Beijing lineage of *Mycobacterium tuberculosis* overproduces triglycerides and has the DosR dormancy regulon constitutively upregulated. J Bacteriol 189: 2583–2589.
14. Mokrousov I, Ly HM, Otten T, Lan NN, Vyshnevskyi B, et al. (2005) Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res 15: 1357–1364.
15. Mokrousov I, Narvskaya O, Otten T, Vyazovaya A, Limeschenko E, et al. (2002) Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res Microbiol 153: 629–637.
16. Faksri K, Drobniewski F, Nikolayevskyy V, Brown T, Prammananan T, et al. (2011) Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on IS6110, SNP, LSP and VNTR profiles from Thailand. Infect Genet Evol 11: 1142–1149.
17. Hanekom M, van der Spuy GD, Streicher E, Ndabambi SL, McEvoy CR, et al. (2007) 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 45: 1483–1490.
18. Iwamoto T, Fujiyama R, Yoshida S, Wada T, Shirai C, et al. (2009) Population structure dynamics of *Mycobacterium tuberculosis* Beijing strains during past decades in Japan. J Clin Microbiol 47: 3340–3343.
19. Iwamoto T, Yoshida S, Suzuki K, Wada T (2008) Population structure analysis of the *Mycobacterium tuberculosis* Beijing family indicates an association between certain sublineages and multidrug resistance. Antimicrob Agents Chemother 52: 3805–3809.
20. Kang HY, Wada T, Iwamoto T, Maeda S, Murase Y, et al. (2010) Phylogeographical particularity of the *Mycobacterium tuberculosis* Beijing family in South Korea based on international comparison with surrounding countries. J Med Microbiol 59: 1191–1197.
21. Maeda S, Wada T, Iwamoto T, Murase Y, Mitarai S, et al. (2010) Beijing family *Mycobacterium tuberculosis* isolated from throughout Japan: phylogeny and genetic features. Int J Tuberc Lung Dis 14: 1201–1204.
22. Millet J, Miyagi-Shiohira C, Yamane N, Mokrousov I, Rastogi N (2012) High-resolution MIRU-VNTRs typing reveals the unique nature of *Mycobacterium tuberculosis* Beijing genotype in Okinawa, Japan. Infect Genet Evol 12: 637–641.
23. Wada T, Iwamoto T, Maeda S (2009) Genetic diversity of the *Mycobacterium tuberculosis* Beijing family in East Asia revealed through refined population structure analysis. FEMS Microbiol Lett 291: 35–43.
24. Yokoyama E, Hachisu Y, Hashimoto R, Kishida K (2012) Population genetic analysis of *Mycobacterium tuberculosis* Beijing subgroup strains. Infect Genet Evol 12: 630–636.
25. Ritacco V, Lopez B, Cafrune PI, Ferrazoli L, Suffys PN, et al. (2008) *Mycobacterium tuberculosis* strains of the Beijing genotype are rarely observed in tuberculosis patients in South America. Mem Inst Oswaldo Cruz 103: 489–492.
26. Gomes HM, Elias AR, Oelemann MA, Pereira MA, Montes FF, et al. (2012) Spoligotypes of *Mycobacterium tuberculosis* complex isolates from patients residents of 11 states of Brazil. Infect Genet Evol 12: 649–656.
27. Cerezo I, Jimenez Y, Hernandez J, Zozio T, Murcia MI, et al. (2012) A first insight on the population structure of *Mycobacterium tuberculosis* complex as studied by spoligotyping and MIRU-VNTRs in Bogota, Colombia. Infect Genet Evol 12: 657–663.
28. Candia N, Lopez B, Zozio T, Carrivale M, Diaz C, et al. (2007) First insight into *Mycobacterium tuberculosis* genetic diversity in Paraguay. BMC Microbiol 7: 75.
29. Abadia E, Sequera M, Ortega D, Mendez MV, Escalona A, et al. (2009) *Mycobacterium tuberculosis* ecology in Venezuela: epidemiologic correlates of common spoligotypes and a large clonal cluster defined by MIRU-VNTR-24. BMC Infect Dis 9: 122.
30. Taype CA, Agapito JC, Accinelli RA, Espinoza JR, Godreuil S, et al. (2012) Genetic diversity, population structure and drug resistance of *Mycobacterium tuberculosis* in Peru. Infect Genet Evol 12: 577–585.
31. Garcia de Viedma D, Chaves F, Inigo J (2006) New route of importation of *Mycobacterium tuberculosis* Beijing genotype. Emerg Infect Dis 12: 169–170.
32. Lari N, Rindi L, Bonanni D, Rastogi N, Sola C, et al. (2007) Three-year longitudinal study of genotypes of *Mycobacterium tuberculosis* isolates in Tuscany, Italy. J Clin Microbiol 45: 1851–1857.
33. Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D, et al. (1997) Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J Clin Microbiol 35: 907–914.
34. Fenner L, Malla B, Ninet B, Dubuis O, Stucki D, et al. (2011) "Pseudo-Beijing": evidence for convergent evolution in the direct repeat region of *Mycobacterium tuberculosis*. PLoS One 6: e24737.
35. Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, et al. (2004) Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc Natl Acad Sci U S A 101: 4865–4870.
36. Caviedes L, Lee TS, Gilman RH, Sheen P, Spellman E, et al. (2000) Rapid, efficient detection and drug susceptibility testing of *Mycobacterium tuberculosis* in sputum by microscopic observation of broth cultures. The Tuberculosis Working Group in Peru. J Clin Microbiol 38: 1203–1208.
37. Moore DA, Mendoza D, Gilman RH, Evans CA, Hollm Delgado MG, et al. (2004) Microscopic observation drug susceptibility assay, a rapid, reliable diagnostic test for multidrug-resistant tuberculosis suitable for use in resource-poor settings. J Clin Microbiol 42: 4432–4437.
38. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, et al. (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44: 4498–4510.

39. Murase Y, Mitarai S, Sugawara I, Kato S, Maeda S (2008) Promising loci of variable numbers of tandem repeats for typing Beijing family *Mycobacterium tuberculosis*. J Med Microbiol 57: 873–880.
40. Zhang L, Chen J, Shen X, Gui X, Mei J, et al. (2008) Highly polymorphic variable-number tandem repeats loci for differentiating Beijing genotype strains of *Mycobacterium tuberculosis* in Shanghai, China. FEMS Microbiol Lett 282: 22–31.
41. Velji P, Nikolayevskyy V, Brown T, Drobniewski F (2009) Discriminatory ability of hypervariable variable number tandem repeat loci in population-based analysis of *Mycobacterium tuberculosis* strains, London, UK. Emerg Infect Dis 15: 1609–1616.
42. Smittipat N, Billamas P, Palittapongarnpim M, Thong-On A, Temu MM, et al. (2005) Polymorphism of variable-number tandem repeats at multiple loci in *Mycobacterium tuberculosis*. J Clin Microbiol 43: 5034–5043.
43. Wang J, Liu Y, Zhang CL, Ji BY, Zhang LZ, et al. (2011) Genotypes and characteristics of clustering and drug susceptibility of *Mycobacterium tuberculosis* isolates collected in Heilongjiang Province, China. J Clin Microbiol 49: 1354–1362.
44. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, et al. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J Bacteriol 182: 2928–2936.
45. Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26: 2465–2466.
46. Small PM, Hopewell PC, Singh SP, Paz A, Parsonnet J, et al. (1994) The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med 330: 1703–1709.
47. Durmaz R, Zozio T, Gunal S, Allix C, Fauville-Dufaux M, et al. (2007) Population-based molecular epidemiological study of tuberculosis in Malatya, Turkey. J Clin Microbiol 45: 4027–4035.
48. Filliol I, Motiwala AS, Cavatore M, Qi W, Hazbon MH, et al. (2006) Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188: 759–772.
49. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393: 537–544.
50. Luo T, Yang C, Gagneux S, Gicquel B, Mei J, et al. (2012) Combination of single nucleotide polymorphism and variable-number tandem repeats for genotyping a homogenous population of *Mycobacterium tuberculosis* Beijing strains in China. J Clin Microbiol 50: 633–639.
51. Wada T, Iwamoto T (2009) Allelic diversity of variable number of tandem repeats provides phylogenetic clues regarding the *Mycobacterium tuberculosis* Beijing family. Infect Genet Evol 9: 921–926.
52. Mokrousov I, Jiao WW, Sun GZ, Liu JW, Valcheva V, et al. (2006) Evolution of drug resistance in different sublineages of *Mycobacterium tuberculosis* Beijing genotype. Antimicrob Agents Chemother 50: 2820–2823.
53. Dou HY, Tseng FC, Lu JJ, Jou R, Tsai SF, et al. (2008) Associations of *Mycobacterium tuberculosis* genotypes with different ethnic and migratory populations in Taiwan. Infect Genet Evol 8: 323–330.
54. Mestre O, Luo T, Dos Vultos T, Kremer K, Murray A, et al. (2011) Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. PLoS One 6: e16020.
55. Wada T, Iwamoto T, Hase A, Maeda S (2012) Scanning of genetic diversity of evolutionarily sequential *Mycobacterium tuberculosis* Beijing family strains based on genome wide analysis. Infect Genet Evol 12: 1392–1396.
56. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D (2010) MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res 38: W326–331.
57. Mokrousov I, Narvskaya O, Vyazovaya A, Millet J, Otten T, et al. (2008) *Mycobacterium tuberculosis* Beijing genotype in Russia: in search of informative variable-number tandem-repeat loci. J Clin Microbiol 46: 3576–3584.
58. Jiao WW, Mokrousov I, Sun GZ, Guo YJ, Vyazovaya A, et al. (2008) Evaluation of new variable-number tandem-repeat systems for typing *Mycobacterium tuberculosis* with Beijing genotype isolates from Beijing, China. J Clin Microbiol 46: 1045–1049.
59. Lasunskaia E, Ribeiro SC, Manicheva O, Gomes LL, Suffys PN, et al. (2010) Emerging multidrug resistant *Mycobacterium tuberculosis* strains of the Beijing genotype circulating in Russia express a pattern of biological properties associated with enhanced virulence. Microbes Infect 12: 467–475.
60. Hernandez-Pando R, Marquina-Castillo B, Barrios-Payan J, Mata-Espinosa D (2012) Use of mouse models to study the variability in virulence associated with specific genotypic lineages of *Mycobacterium tuberculosis*. Infect Genet Evol 12: 725–731.
61. Chen YY, Chang JR, Huang WF, Kuo SC, Su IJ, et al. (2012) Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on SNP and VNTR typing profiles in Asian countries. PLoS One 7: e39792.
62. Homolka S, Projahn M, Feuerriegel S, Ubben T, Diel R, et al. (2012) High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. PLoS One 7: e39855.
63. Stucki D, Malla B, Hostettler S, Huna T, Feldmann J, et al. (2012) Two New Rapid SNP-Typing Methods for Classifying *Mycobacterium tuberculosis* Complex into the Main Phylogenetic Lineages. PLoS One 7: e41253.

Infection, Genetics and Evolution 16 (2013) 174–177

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Short communication

# Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages

Noriko Nakanishi [a], Takayuki Wada [b], Kentaro Arikawa [a], Julie Millet [c], Nalin Rastogi [c], Tomotada Iwamoto [a,*]

[a] Department of Microbiology, Kobe Institute of Health, Kobe, Japan
[b] Department of International Health, Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan
[c] WHO Supranational TB Reference Laboratory, Unite de la Tuberculose et des Mycobacteries, Institut Pasteur de Guadeloupe, Abymes Cedex, Guadeloupe, France

ARTICLE INFO

*Article history:*
Received 22 October 2012
Received in revised form 12 February 2013
Accepted 13 February 2013
Available online 22 February 2013

*Keywords:*
Beijing family
Genotyping
IS*6110*
Molecular epidemiology
*Mycobacterium tuberculosis*
Single nucleotide polymorphism

ABSTRACT

Genotypic classification in *Mycobacterium tuberculosis* has greatly contributed to the comprehension of phylogenetic and population genetic relationships. It is, therefore, necessary to verify the robustness of the genetic markers for phylogenetic classification. In this study, we report some examples of homoplasy for two molecular markers, the IS*6110* insertion at the NTF region, and a single nucleotide polymorphism (SNP) at locus 909166, through genotyping of 1054 Beijing family strains. Our data revealed that a small fraction of strains traditionally classified into modern sublineages by IS*6110* insertion at NTF actually belong to an ancient sublineage. We also proved that the robustness of branches in the evolutionary tree established using the putative homoplasious SNP 909166 is relatively low. Our findings highlight the importance of validating genetic markers used to establish phylogeny, evolution, and phenotypic characteristics.



## 1. Introduction

Recent progress in research on genetic diversity of *Mycobacterium tuberculosis* has brought new evidence that the outcome of infection and disease may differ depending on the genotype, lineage, and sublineage of the strain involved (Coscolla and Gagneux, 2010; Malik and Godfrey-Faussett, 2005). Both large sequence polymorphisms (LSPs), and single nucleotide polymorphisms (SNPs) have facilitated comprehension of phylogenetic and population genetic relationships (Comas and Gagneux, 2009; Comas et al., 2009), because they exhibit minimal rates of homoplasy. Moreover, highly discriminatory markers, including multilocus variable number of tandem repeats (VNTRs), are widely used to track particular strains in the community (Comas et al., 2009).

The Beijing family of *M. tuberculosis* strains is the most frequently analyzed lineage (Coscolla and Gagneux, 2010), and various genetic markers have been developed for subdividing the species into the Beijing family and its sublineages (Dou et al., 2008; Faksri et al., 2011; Filliol et al., 2006; Hanekom et al., 2007; Kremer et al., 2004; Luo et al., 2012; Mokrousov et al., 2002, 2005; Tsolaki et al., 2005; Wada et al., 2009b). On the other hand, some recent reports have indicated the risk of misclassification of Beijing family strains based on traditional molecular markers due to their homoplasy (Faksri et al., 2011; Fenner et al., 2011). We therefore decided to investigate the robustness of sublineage categorization using two molecular markers that have already shown their usefulness for evolutionary studies, namely: (i) the IS*6110* insertion at the NTF region, which is traditionally used to divide the family into modern (typical) and ancient (atypical) sublineages (Mokrousov et al., 2005), and (ii) SNP locus 909166, which is 1 of the 10 SNPs in a set used to classify the family into Filliol's sequence types (STs) (Filliol et al., 2006).

## 2. Materials and methods

We obtained 1340 *M. tuberculosis* isolates from newly diagnosed tuberculosis (TB) in Kobe, Japan, between 2002 and 2009. Of these, 909 isolates overlapped with the isolates which were previously used for population structure analysis (Iwamoto et al., 2009). In total, 1054 isolates were identified as Beijing family based on spoligotyping (Kamerbeek et al., 1997), and subsequent confirmation by LSPs (RD181 and/or RD105) and SNPs described below. These isolates included 67 Beijing variant spoligotyping patterns that lack some of the spacers between 35 and 43. These 67 isolates were confirmed by RD181 and/or RD105 analysis as belonging to the Beijing family (Tsolaki et al., 2004). All the Beijing family strains were classified into ancient and modern Beijing sublineages based on the presence of IS*6110* in the NTF region (Mok-

* Corresponding author. Address: Department of Microbiology, Kobe Institute of Health, 4-6 Minatojima-nakamachi, Chuo-ku, Kobe 650-0046, Japan. Tel.: +81 78 302 6251; fax: +81 78 302 0894.
*E-mail address:* kx2t-iwmt@asahi-net.or.jp (T. Iwamoto).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.02.007

**Table 1**
Distribution of 1054 Beijing family strains in each sublineage according to SNPs.

| Beijing sublineages | SNPs | | | | | | | | | | No. (%) of strains | IS*6110* insertion in the NTF region | Subgroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1548149 | 1692069 | 797736 | 2825581 | 1892017 | 4137829 | 1477596 | 2376135 | 2532616 | 909166 | | | |
| *a. Classification based on phylogeny using published SNPs* | | | | | | | | | | | | | |
| ST11 | G | A | C | T | T | C | C | A | G | C | 4(0.4%) | None | Ancient |
| ST26 | G | A | C | T | T | C | C | A | G | T | 68(6.5%) | None | Ancient |
| STK | G | A | T | G | T | C | C | A | G | T | 167(15.8%) | None[d] | Ancient |
| ST3 | G | A | T | G | T | C | C | A | G | C | 271(25.7%) | None[d] | Ancient |
| ST25 | G | A | T | G | C | T | C | A | G | C | 10(0.9%) | None | Ancient |
| ST19 | G | A | T | G | C | T | C | A | G | T | 288(27.3%) | None[d] | Ancient |
| ST10 | G | A | T | G | C | T | T | A | G | T | 196(18.6%) | NTF:: IS6110 | Modern |
| ST22 | G | A | T | G | C | T | T | G | A | T | 50(4.7%) | NTF:: IS6110 | Modern |
| | 1548149 | 1692069 | 797736 | 2825581 | 1576481 | 1892017 | 4137829 | 1477596 | 2376135 | 2532616 | | | |
| *b. Classification based on phylogeny using published SNPs and the additional SNP at position 1576481* | | | | | | | | | | | | | |
| ST11/ST26 | G | A | C | T | T | T | C | C | A | G | 72(6.8%) | None | Ancient |
| STKr[c] | G | A | T | G | T | T | C | C | A | G | 175(16.6%)[a] | None[d] | Ancient |
| ST3r[c] | G | A | T | G | G | T | C | C | A | G | 263(25.0%)[b] | None[d] | Ancient |
| ST25/ST19 | G | A | T | G | G | C | T | C | A | G | 298(28.2%) | None[d] | Ancient |
| ST10 | G | A | T | G | G | C | T | T | A | G | 196(18.6%) | NTF:: IS6110 | Modern |
| ST22 | G | A | T | G | G | C | T | T | G | A | 50(4.7%) | NTF:: IS6110 | Modern |

[a] Twelve isolates originally classified as ST3 were re-classified as STKr by using SNP 1576481 instead of SNP 909166.
[b] Four isolates originally classified as STK were re-classified as ST3r by using SNP 1576481 instead of SNP 909166.
[c] Previously STK and ST3 sublineages were renamed STKr and STK3r according to the definition of the new 10-SNP set, respectively.
[d] Three isolates horboring IS*6110* insertion in the NTF region belonged to the STKr, ST3r and ST19 sublineages, respectively.

rousov et al., 2005; Wada et al., 2009b). They were also classified into ST sublineages using Filliol's 10 SNPs (Filliol et al., 2006; Hanekom et al., 2007; Iwamoto et al., 2008). In addition to the 10-SNP set, SNP locus 1576481 recently introduced in the panel as the replacement with SNP 909166 (Wada et al., 2012), was also analyzed to assess the homoplasious behavior of SNP 909166. We negated the existence of "Pseudo-Beijing strains" (Fenner et al., 2011) in our sample set by the detection of RD105 (Tsolaki et al., 2004) for all of the strains classified as ST11 and ST26 sublineages using SNP genotyping (Iwamoto et al., 2008; Iwamoto et al., 2012). All strains were subjected to Supply's optimized 15-locus variable number of tandem repeats (15-MIRU-VNTR) analysis (Supply et al., 2006; Iwamoto et al., 2012).

## 3. Results and discussion

Recently, Faksri et al. (2011) pointed out that three of the traditional 10-SNP set (SNPs 909166, 1548149, and 4137829) did not represent unique, irreversible events. Although they noted that the robustness of branches harboring these reversible SNPs is relatively low, this has not yet been formally evaluated using large number of sample sets. Of these three SNPs, in particular, SNP 909166 has played a key role in facilitating division of strains into the STK and ST3 sublineages that are predominant in Japan (Iwamoto et al., 2008, 2009; Wada et al., 2009b). Therefore, evaluation of the robustness of the sublineage categorization for evolutionary studies is imperative.

The results obtained on the classification of 1054 Beijing family strains using published SNPs and the additional SNP at position 1576481 are summarized in Tables 1a and b. When we applied SNP 1576481, which is expected to compensate for the probable homoplasious behavior of SNP 909166 (Wada et al., 2012), to our sample set, all of the 1054 Beijing family isolates could be classified into sublineages without considering the reversibility of the SNP (Table 1b).The striking difference of this classification from that using SNP 909166 (Table 1a) confirmed that SNP 1576481 (T to G) is evolutionarily a more informative SNP than the former SNP.

Given this result, we aimed to verify the robustness of classification of sublineages using SNP 909166 (a reversible SNP), in comparison with classification using SNP 1576481 (an irreversible SNP). The results revealed a discrepancy in the classification of 16 isolates (16/438 [3.7%]) between these two SNPs (Tables 1a and b). Specifically, 12 isolates originally classified as ST3 isolates were re-classified as STKr, and four isolates originally classified as STK isolates were re-classified as ST3r, when SNP 1576481 was used (Table 1b). A minimum spanning tree (MST), based on 15-MIRU-VNTR which included phylogenetically informative loci for Beijing family strains (Wada and Iwamoto, 2009a; Faksri et al., 2011), was consistent with the classification using SNP 1576481 (Fig. 1). These results strongly suggest that the robustness of the branches identified using SNP 909166 is relatively low, while an evolutionarily more reliable classification for STKr and ST3r was possible using SNP 1576481. Although we could not verify the robustness of classification using the other two reversible SNPs, 1548149 and 4137829, in our sample set, their utilization as genetic markers may be reconsidered using the same approach.

The IS*6110* insertion in the NTF region is well established as a genetic marker for grouping Beijing family strains into "modern" and "ancient" sublineages (Mokrousov et al., 2002, 2005). Although all of the modern-type strains in our study (ST10 and ST22) as classified by SNPs, certainly possessed an IS*6110* insertion in the NTF region, we also found three ancient strains (3/1,054, 0.3%) harboring an IS*6110* insertion in the NTF region (Table 2). Since there was no double peak in the VNTR profiles, mixed infection was eliminated (Table 2). For these three ancient strains, IS*6110* insertion in the NTF region was identified by PCR amplification and its sequence analysis. IS*6110* insertion sites in the NTF region in three strains exhibited distinctive patterns (data not shown). When we investigated the IS*6110* insertion sites at NTF region in the 10 modern-type strains, they were identified in the same position but different from that of the three ancient strains (data not shown). These results suggested that IS*6110* insertion in the NTF region occurs independently in these ancient strains. They belonged to the ST3r, ST19, and STKr sublineages, based on SNP data. These isolates presented sublineage-specific VNTR allele profiles, as previously described (Wada and Iwamoto, 2009a) (Table 2), which supports their assignment as ancient types. This result provided clear evidence that some of the ancient type isolates could be misclassified as modern type under the current definition, i.e., presence of

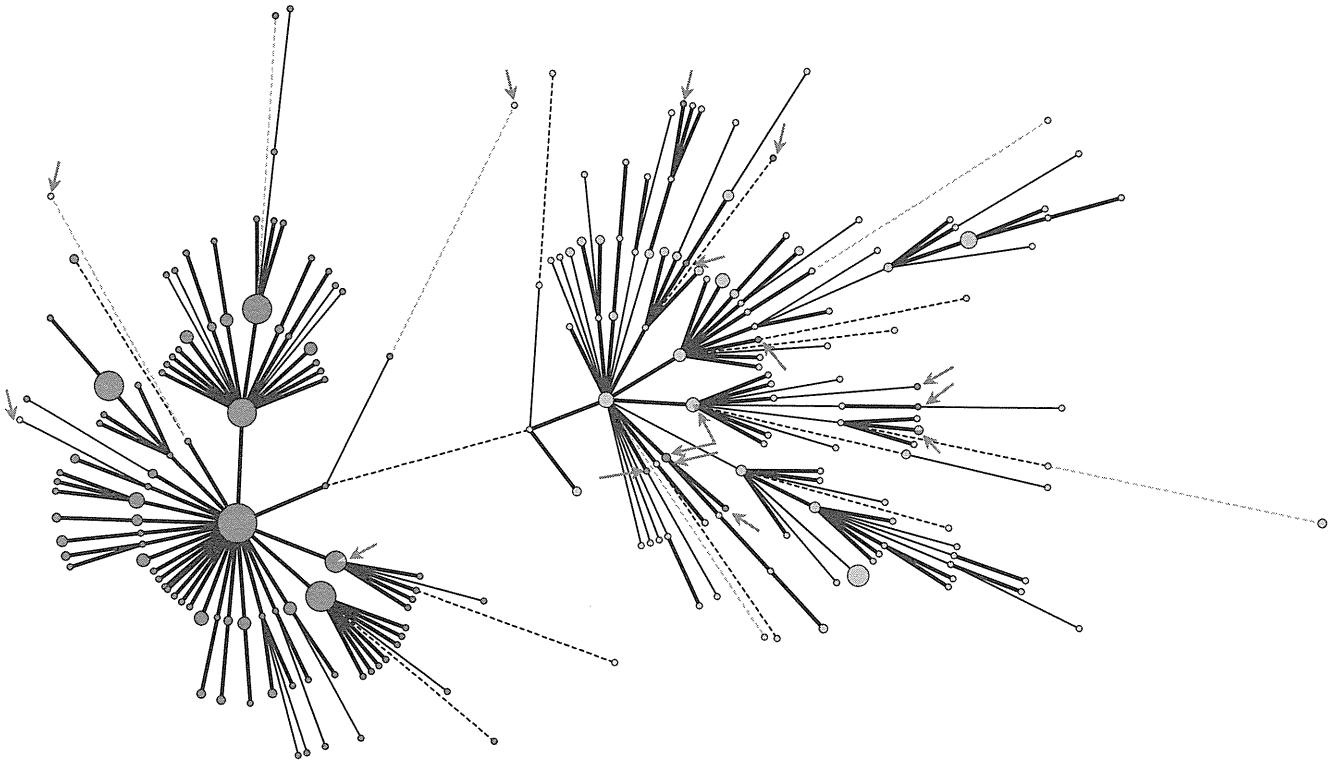

**Fig. 1.** A minimum spanning tree based on 15-MIRU-VNTR genotyping of 271 ST3 (red) and 167 STK (green) sublineage isolates, which were categorized using SNP 909166. Red arrows show 16 misclassified isolates. Twelve ST3 and 4 STK sublineages, as defined by SNP 909166, belonged to the STK and ST3 branches, respectively.(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**
15-MIRU-VNTR alleles in the three isolates harboring IS*6110* insertion in NTF region.

| No. | Years | Beijing sublineages | IS*6110* insertion in the NTF region | 15-MIRU-VNTR loci allele profile* | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | MIRU 4 | MIRU 10 | MIRU16 | MIRU 26 | MIRU 31 | MIRU 40 | ETR A | ETR C | QUB-11b | QUB-26 | QUB-4156 | Mtub04 | Mtub21 | Mtub30 | Mtub39 |
| 1 | 2007 | ST3r | NTF:: IS*6110* | 2 | **1** | 3 | 7 | 4 | 3 | 4 | 4 | **7** | 8 | **5** | 4 | 3 | 4 | 3 |
| 2 | 2009 | ST19 | NTF:: IS*6110* | 2 | 3 | 4 | 7 | 5 | 3 | 4 | 4 | 7 | 2 | **5** | 3 | 3 | 4 | 3 |
| 3 | 2009 | STKr | NTF:: IS*6110* | 2 | 3 | 3 | 7 | 5 | 3 | 1 | 4 | **3** | 7 | **4** | 4 | 3 | **2** | 3 |

* Boldface data indicate specific VNTR alleles observed in the ancient phylogenetic sublineage of Beijng family as previously reported (Wada and Iwamoto, 2009a).

IS*6110* in the NTF region. These rare exceptions can be negligible in population-based studies. However, when exploring strain-specific differences in experimental or clinical phenotypes, we need to be careful of these rare but important exceptions to the classification of modern type Beijing.

In conclusion, we demonstrated homoplasious events in two molecular markers: the IS*6110* insertion in the NTF region and SNP 909166. Our findings highlight the importance of validating such genetic markers used to establish phylogeny, evolution, and phenotypic characteristics. With the tremendous increase in whole-genome sequence data, verification of the robustness of polymorphic loci has become very important for establishing the valid SNP-based classification scheme for *M. tuberculosis* strains.

## Acknowledgments

This work was supported by Health Science Research grants (H24-SHINKO-IPPAN-011) from Ministry of Health, Labor and Welfare of Japan, MEXT/JSPS KAKENHI (Grant Numbers 24590845 and 24689034), and the US-Japan Cooperative Medical Science Program (TB and Leprosy panel).

## References

Comas, I., Gagneux, S., 2009. The past and future of tuberculosis research. PLoS Pathog. 5, e1000600.

Comas, I., Homolka, S., Niemann, S., Gagneux, S., 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4, e7815.

Coscolla, M., Gagneux, S., 2010. Does *M. tuberculosis* genomic diversity explain disease diversity? Drug Discov. Today Dis. Mech. 7, e43–e59.

Dou, H.Y., Tseng, F.C., Lin, C.W., Chang, J.R., Sun, J.R., Tsai, W.S., Lee, S.Y., Su, I.J., Lu, J.J., 2008. Molecular epidemiology and evolutionary genetics of *Mycobacterium tuberculosis* in Taipei. BMC Infect. Dis. 8, 170.

Faksri, K., Drobniewski, F., Nikolayevskyy, V., Brown, T., Prammananan, T., Palittapongarnpim, P., Prayoonwiwat, N., Chaiprasert, A., 2011. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family based on IS*6110*, SNP, LSP and VNTR profiles from Thailand. Infect. Genet. Evol. 11, 1142–1149.

Fenner, L., Malla, B., Ninet, B., Dubuis, O., Stucki, D., Borrell, S., Huna, T., Bodmer, T., Egger, M., Gagneux, S., 2011. "Pseudo-Beijing": evidence for convergent evolution in the direct repeat region of *Mycobacterium tuberculosis*. PLoS One 6 (9), e24737.

Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbón, M.H., Bobadilla del Valle, M., Fyfe, J., García-García, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., León, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendón, A., Sifuentes-Osornio, J., Ponce de León, A., Cave, M.D., Fleischmann, R., Whittam, T.S., Alland, D., 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J. Bacteriol. 188, 759–772.

Hanekom, M., van der Spuy, G.D., Streicher, E., Ndabambi, S.L., McEvoy, C.R., Kidd, M., Beyers, N., Victor, T.C., van Helden, P.D., Warren, R.M., 2007. 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. 45, 1483–1490.

Iwamoto, T., Grandjean, L., Arikawa, K., Nakanishi, N., Caviedes, L., Coronel, J., Sheen, P., Wada, T., Taype, C.A., Shaw, M.A., Moore, D.A., Gilman, R.H., 2012. Genetic diversity and transmission characteristics of Beijing family strains of *Mycobacterium tuberculosis* in Peru. PLoS One 7, e49651.

Iwamoto, T., Fujiyama, R., Yoshida, S., Wada, T., Shirai, C., Kawakami, Y., 2009. Population structure dynamics of *Mycobacterium tuberculosis* Beijing strains during past decades in Japan. J. Clin. Microbiol. 47, 3340–3343.

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

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35, 907–914.

Kremer, K., Glynn, J.R., Lillebaek, T., Niemann, S., Kurepina, N.E., Kreiswirth, B.N., Bifani, P.J., van Soolingen, D., 2004. Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. J. Clin. Microbiol. 42, 4040–4049.

Luo, T., Yang, C., Gagneux, S., Gicquel, B., Mei, J., Gao, Q., 2012. Combination of single nucleotide polymorphism and variable number tandem repeats for genotyping a homogenous population of *Mycobacterium tuberculosis* Beijing strains in China. J. Clin. Microbiol. 50, 633–639.

Malik, A.N.J., Godfrey-Faussett, P., 2005. Effects of genetic variability of *Mycobacterium tuberculosis* strains on the presentation of disease. Lancet Infect. Dis. 5, 174–183.

Mokrousov, I., Ly, H.M., Otten, T., Lan, N.N., Vyshnevskyi, B., Hoffner, S., Narvskaya, O., 2005. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res. 15, 1357–1364.

Mokrousov, I., Narvskaya, O., Otten, T., Vyazovaya, A., Limeschenko, E., Steklova, L., Vyshnevskyi, B., 2002. Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res. Microbiol. 153, 629–637.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rüsch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H., Roring, S., Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M.C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C., van Soolingen, D., 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 44, 4498–4510.

Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M., Goguet de la Salmoniere, Y.O., Aman, K., Kato-Maeda, M., Small, P.M., 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc. Natl. Acad. Sci. USA 101, 4865–4870.

Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.O., Kreiswirth, B.N., van Soolingen, D., Small, P.M., 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 43, 3185–3191.

Wada, T., Iwamoto, T., 2009a. Allelic diversity of variable number of tandem repeats provides phylogenetic clues regarding the *Mycobacterium tuberculosis* Beijing family. Infect. Genet. Evol. 9, 921–926.

Wada, T., Iwamoto, T., Maeda, S., 2009b. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family in East Asia revealed through refined population structure analysis. FEMS Microbiol. Lett. 291, 35–43.

Wada, T., Iwamoto, T., Hase, A., Maeda, S., 2012. Scanning of genetic diversity of evolutionarily sequential *Mycobacterium tuberculosis* Beijing family strains based on genome wide analysis. Infect. Genet. Evol. 12, 1392–1396.

Tuberculosis 92 (2012) 351–357

Contents lists available at SciVerse ScienceDirect

Tuberculosis

journal homepage: http://intl.elsevierhealth.com/journals/tube

MECHANISMS OF PATHOGENESIS

# Non-acid-fastness in *Mycobacterium tuberculosis* Δ*kasB* mutant correlates with the cell envelope electron density

Hiroyuki Yamada [a,*], Apoorva Bhatt [b], Radostin Danev [c,f], Nagatoshi Fujiwara [d], Shinji Maeda [a], Satoshi Mitarai [a], Kinuyo Chikamatsu [a], Akio Aono [a], Koji Nitta [c,g], William R. Jacobs Jr. [e], Kuniaki Nagayama [c,h]

[a] *Department of Mycobacterium Reference and Research, The Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, 3-1-24, Matsuyama, Kiyose, Tokyo 204-8533, Japan*
[b] *School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*
[c] *Okazaki Institute for Integrative Bioscience and National Institute for Physiological Sciences, National Institutes of Natural Sciences, Okazaki, Aichi 444-8787, Japan*
[d] *Department of Bacteriology, Osaka City University Graduate School of Medicine, Osaka 545-8585, Japan*
[e] *Howard Hughes Medical Institute, Department of Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA*

ARTICLE INFO

*Article history:*
Received 9 December 2011
Received in revised form
8 February 2012
Accepted 11 February 2012

*Keywords:*
*Mycobacterium tuberculosis*
Virulent strain
Acid-fastness
Cell envelope
Transmission electron microscopy

SUMMARY

The acid-fastness is the most important and the most specific characteristics in mycobacteria, the mechanism of which is not clear but may be attributed to the lipid rich cell wall of this bacterium. While the exact component(s) responsible for this staining method remained unidentified, a *Mycobacterium tuberculosis* mutant, attenuated strain that produced shorter mycolic acids with defects in trans-cyclopropanation was shown to be acid fast negative. In this study, we examined the ultrastructure of the cell envelope (CE) of the mutant strain Δ*kasB* (missing a beta-ketoacyl-ACP synthase involved in mycolic acid biosynthesis), the parental CDC1551 (wild type strain) and *kasB* complemented strain, and compared ultrastructural differences among them with conventional transmission electron microscopy (TEM) and cryo-transmission electron microscopy (CEM). Conventional TEM revealed that there were no detectable differences in the thickness of the cell envelope among three strains (wild-type: 43.35 ± 6.13 nm; Δ*kasB*: 45.98 ± 11.32 nm; complement: 40.71 ± 6.3 nm). However, CEM data demonstrated that the region between the inner and outer membranes of the mutant strain, which is composed mainly of cell wall anchored mycolic acids (MA), showed a significant decrease in electron density as compared to the wild type and *kasB* complement strain (567.1 ± 372.7 vs. 301.4 ± 262.1, or vs. 235.2 ± 174.9, $p < 0.02$ or $p < 0.001$, respectively). These results suggested that altered MA patterns in the *kasB* mutant may have affected the packing of the lipid rich layer of the *M. tuberculosis* cell envelope, resulting in a reduced electron density of this layer as seen by CEM and loss of acid-fastness in light microscopical observation, and we propose a novel model of the cell envelope structure in tubercle bacilli.


## 1. Introduction

The World Health Organization estimates that there were 8.8 million incidences of tuberculosis, 12 million prevalent cases and 1.1 million deaths among HIV-negative cases in 2010. The estimated global incidence rate is 178 cases per 100,000 population, and the rate continues to fall from 2004, but too slowly. Furthermore, there were 650.000 new multidrug resistance tuberculosis (MDR-TB) cases in 2010. Extensively drug-resistant TB (XDR-TB) cases have been confirmed in 77 countries.[1]

In the era of molecular detection method, acid-fast staining has been sustaining its role as conventional but simple, rapid, and inexpensive detection method for tubercle bacilli, especially in developing countries because this method provides not only quantitative, but also qualitative information, such as the shape of bacilli. Yamada et al. have recently established an efficient preparation method of artificial sputum for training of microscopists in

* Corresponding author. Tel.: +81 42 493 5072; fax: +81 42 492 4600.
*E-mail address:* hyamada@jata.or.jp (H. Yamada).
[f] Present addrerss: Max Planck Institute of Biochemistry, Am Klopferspitz, 1882152, Martinsried, Germany.
[g] Present addrerss: Analysis Technology Center, Recearch and Development Management Headquarters, Fuji Photo Film Co., Ltd. Nakanuma, Minamiashigara, Kanagawa 250-0193, Japan.
[h] Present addrerss: National Institutes of Natural Sciences, Okazaki, Aichi 444-8787, Japan.

1472-9792/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tube.2012.02.006

sputum smear microscopy stained by Ziehl-Neelsen and fluorescent staining methods.[2–4]

Mycobacteria have a distinctive cell envelope (CE) that is principally composed of a lipid bilayer plasma membrane (PM), peptidoglycan, arabinogalactan, mycolic acid (MA) and an outer membrane (OM), where these fundamental components are covalently linked.[5–7] The structure of the mycobacterial CE provides these organisms with various properties, such as acid-fastness, prevention of phagosome-lysosome fusion, and inhibition of cytokine-mediated host responses.[8–13]

Bhatt et al. described a mutant of *Mycobacterium tuberculosis* that did not retain the basic fuchsin stain after decolorization with 3% hydrochloric acid-ethanol (i.e., acid-fast negative) and also failed to cause disease in mice.[5,6] The mutant strain, which lacked *kasB* (Δ*kasB*), one of the two β-ketoacyl-ACP synthases involved in the biosynthesis of MA, produced shorter MA chains and was defective in the trans-cyclopropanation of oxygenated MA.[5,6,14–17]

However, the mechanism underlying acid-fastness remains to be elucidated, although the mechanism underlying Gram stain, another conventional staining method widely used in microbiology, has already been illustrated by the combination of electron microscopy and element analysis.[18–20] Despite the fact that a large number of molecular biological studies have assessed the function of KasA and KasB, several CEM-based studies of mycobacteria, with a resolving power of less than 10 nm (i.e., equivalent to the thickness of the PM), have not discussed the mechanism underlying the correlation between acid-fastness and ultrastructural properties of CE.[12,13,21,22] Recently, Yamada et al. described a novel rapid freeze-substitution (RFS) electron microscopy method for studying the ultrastructure of the pathogenic virulent bacteria, such as *M. tuberculosis*.[23] In the present study, we examined the acid-fast-negative Δ*kasB* mutant prepared by RFS as well as CEM after fixation with glutaraldehyde and performed comparative microscopic analysis of the CE among acid-fast-positive (wild-type and complement strains) and acid-fast-negative (Δ*kasB*) strains, thus probing the potential relationship between CE structure and acid-fastness. We describe the differences in the structure of the CE among these 3 strains with data from conventional transmission electron microscopy (TEM) and CEM.

## 2. Materials and methods

### 2.1. Bacteria and the construction of the deletion mutant

A specialized transducing phage, ph*ΔkasB*, containing an allelic exchange substrate designed to replace *kasB* with a hygromycin resistance cassette (hyg), was transduced into the virulent CDC1551 *M. tuberculosis* strain. The construction of the Δ*kasB* mutant and complemented strains was described previously.[5] The bacterial strains were cultured in 7H9 broth (Difco, Sparks, MD, USA) containing 10% Middlebrook OADC enrichment; in addition, 75 μg/mL hygromycin was supplemented for Δ*kasB* mutant culture and 75 μg/mL hygromycin and 20 μg/mL kanamycin were supplemented for complement strain culture. All strains were subjected to Ziehl-Neelsen staining to confirm acid-fastness.

### 2.2. Rapid freeze substitution (RFS) for transmission electron microscopy

The sandwich method was performed as described previously.[23–26] Briefly, the single hole copper grids (hole diameter, 0.1 mm) used in this method were subjected to glow-discharge in order to be hydrophilized with an ion bombarder (Model PIB-10 Plasma Ion Bombarder; Vacuum Device Inc., Mito, Ibaraki, Japan) just prior to use. Less than 1 μL of the highly concentrated bacillary pellet was prepared by centrifugation at 9000×*g* for 30 s and transferred to the surface of a single hole copper grid, then, the pellet was sandwiched with another grid. The sandwiched grids were picked up with tweezers and quickly immersed in a slush of liquid propane for 20 s (Model VFZ-101 Multi-Purpose Quick Freezing Device; Vacuum Device Inc., Mito, Ibaraki, Japan). The sandwiched grids were detached in liquid nitrogen and transferred into a cooled 2% osmium tetroxide acetone solution in the freezing device described above. The samples were transferred from the biosafety facility and placed in a −85 °C freezer for >12 h; thereafter, they were moved to a −20 °C freezer and to refrigerator to adjust the sample temperature to room temperature gradually. Once the sample grids were at room temperature, the osmium solution was discarded and the grids were washed 3 times with absolute acetone and transferred to a flat-bottomed BEEM capsule (TAAB Laboratories Equipment Ltd., Aldermaston, Berkshire, UK). The capsule was filled with Spurr's resin and allowed to polymerize for 16 h at 70 °C.[23] The sample grid was then detached from the polymerized epoxy block. The flat area containing bacilli clumps was trimmed and ultra-thin sections were cut with a Reichert Ultracut E microtome (Reichert-Jung, Vienna, Austria), stained with uranyl acetate and lead citrate, and examined with a JEM-1230 electron microscope (JEOL, Tokyo, Japan).

### 2.3. Cryo-electron microscopy (CEM)

As the CEM facility was not inside a biosafety level 3 area, the pathogenic tubercle bacilli were fixed with 2.5% glutaraldehyde in phosphate buffer (0.1 N, pH 7.4) overnight at 4 °C, and rinsed 3 times with distilled water. Two μL of the suspension was applied to a glow-discharged carbon grid with holes (Quantifoil copper grids R1.2/1.3; Quantifoil MicroTools, Jena, Germany) and mounted in an environmentally controlled chamber at 100% humidity, excess water was removed by blotting, and the grids were frozen in vitreous ice by plunging them into liquid ethane cooled with liquid nitrogen using a Vitrobot (FEI, Hillsboro, OR, USA). The grid was transferred into a JEM-3100FFC (JEOL) cryo-EM equipped with a field emission gun. The microscope was operated at 300 kV acceleration voltage. The specimen temperature was maintained at 55 K, as described previously.[27,28] Data were collected at a nominal magnification of 10,000× and recorded with a 2 k × 2 k CCD camera (Megascan 795; Gatan, Pleasanton, CA, USA) at a pixel size of 1.968 nm at the specimen through Digital Micrograph software (Gatan, Pleasanton, CA, USA).[27,29,30]

### 2.4. Image analysis

Printed electron micrographs taken at 30,000× original magnification for RFS samples were scanned through Adobe Photoshop Elements version 5 with a CanoScan 8800F (Canon, Tokyo, Japan). The images obtained through CEM were recorded on a CCD camera with Digital Micrograph software (Gatan). Stored images were analyzed by ImageJ software[31] in order to measure the length and width of the bacterial cells as well as the thickness. Electron density of the CEs were measured. Electron density is the term used to express the degree of darkness derived from differences in atomic masses in the samples. This density can be enhanced by the result of both osmium tetroxide fixation and uranyl acetate/lead citrate staining in conventional TEM samples. However, because all materials are composed from atoms, electron density occurs in every biological sample without osmium fixation or staining in CEM. The differences in atomic density result in the differences of the electron density in CEM, which provide images. In our study, we quantified the electron density of the CE as the relative gray value (RGV) with the plot profile option menu of ImageJ software. The RGV was inversely proportional to the electron density, and the