JCM
Journals.ASM.org

Rapid Whole-Genome Sequencing for Investigation of a Suspected Tuberculosis Outbreak

M. Estée Török,[a,b,c] Sandra Reuter,[d] Josephine Bryant,[d] Claudio U. Köser,[a,c] Sian V. Stinchcombe,[b] Bernadette Nazareth,[e] Matthew J. Ellington,[c] Stephen D. Bentley,[d] Geoffrey P. Smith,[f] Julian Parkhill,[d] Sharon J. Peacock[a,b,c,d]

Department of Medicine, University of Cambridge, Cambridge, United Kingdom[a]; Cambridge University Hospitals NHS Foundation Trust, Cambridge, United Kingdom[b]; Health Protection Agency, Cambridge, United Kingdom[c]; Wellcome Trust Sanger Institute, Hinxton, United Kingdom[d]; Norfolk, Suffolk and Cambridgeshire Health Protection Unit, Thetford, United Kingdom[e]; Illumina, Ltd., Chesterford, United Kingdom[f]

Two Southeast Asian students attending the same school in the United Kingdom presented with pulmonary tuberculosis. An epidemiological investigation failed to link the two cases, and drug resistance profiles of the *Mycobacterium tuberculosis* isolates were discrepant. Whole-genome sequencing of the isolates found them to be genetically identical, suggesting a missed transmission event.

Tuberculosis is a major global public health challenge, with an incidence of 8.8 million new cases and 1.4 million deaths in 2010 (1). The United Kingdom is a low-prevalence country, with 8,963 new cases reported in 2011, giving an overall incidence of 14.4 cases per 100,000 population per year (2). The majority of cases (74%) occur in persons born outside the United Kingdom, and the rate of tuberculosis is almost 21 times higher in non-United Kingdom-born than United Kingdom-born individuals. The United Kingdom Health Protection Agency collects clinical and epidemiological information about all tuberculosis cases through the National Enhanced Tuberculosis Surveillance system. In 2010, a national strain typing service was introduced in order to aid in the investigation of outbreaks. This uses the 24-locus mycobacterial interspersed repetitive unit-variable number tandem repeats (MIRU-VNTRs) to type all *Mycobacterium tuberculosis* isolates that are submitted to the reference laboratory (2). *M. tuberculosis* is a clonal organism, and accumulating evidence suggests that conventional typing may be insufficient to discriminate between closely related strains within a given lineage. Whole-genome sequencing provides the ultimate resolution of genomic data and has been shown to be superior to MIRU-VNTR typing in the investigation of a large outbreak in Canada (3). Here, we describe the use of whole-genome sequencing to investigate a suspected tuberculosis outbreak in a language school in Cambridge, United Kingdom.

Case histories. On 5 November 2010, a 22-year-old South Korean man presented to our hospital with smear-positive pulmonary tuberculosis. He had no medical history and no known contact with tuberculosis and was studying at a local English language school. Treatment was commenced with standard antituberculosis therapy (rifampin, isoniazid, pyrazinamide, and ethambutol), and he was discharged from the hospital on 17 November 2010. Antibiotic susceptibility results subsequently revealed resistance to isoniazid, prothionamide, and streptomycin, and moxifloxacin was substituted for isoniazid.

On 22 November 2010, a 24-year-old Japanese man presented to our hospital with smear-positive pulmonary tuberculosis. He had no medical history and no known contact with tuberculosis and was also studying at the same language school. Treatment was commenced with standard antituberculosis therapy. Progress was complicated by hepatitis (which necessitated temporary interruption of rifampin, isoniazid, and pyrazinamide and the addition of streptomycin to ethambutol), a spontaneous pneumothorax (which resolved), and prolonged fever. Antimicrobial susceptibility results revealed resistance to isoniazid and prothionamide, and moxifloxacin was substituted for isoniazid. He was discharged from the hospital on 31 December 2010.

Outbreak investigation. In view of the geographical and temporal association of the two cases, an outbreak investigation was performed by the local Health Protection Unit. Although the two cases attended the same language school, they lived in different households and did not know each other or attend the same classes, other than for a brief period during April 2010. Contact tracing and screening for tuberculosis were performed initially for household contacts of the first case and then extended to all staff and students at the language school after the second case was diagnosed. A total of 609 persons (576 of whom were foreign students) were estimated to be in the at-risk group. A total of 468 students had left the United Kingdom before the start of the screening program. Of the 141 individuals in the United Kingdom who were eligible for screening, 96 were screened and 45 declined screening. Thirteen cases of latent tuberculosis were identified, and there were no further cases of active tuberculosis.

Laboratory methods. Sputum specimens were processed at the Cambridge Health Protection Agency Microbiology Laboratory using standard methods for microscopy and culture of mycobacteria. Samples that were culture positive were sent to the National Mycobacterial Reference Laboratory for identification, antimicrobial susceptibility testing, and 24-locus MIRU-VNTR typing. Typing of the two *M. tuberculosis* isolates revealed that they were both Beijing strains with the same

Received 24 August 2012 Returned for modification 16 October 2012
Accepted 15 November 2012
Published ahead of print 21 November 2012
Address correspondence to M. Estée Török, estee.torok@addenbrookes.nhs.uk.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/JCM.02279-12.

doi:10.1128/JCM.02279-12
The authors have paid a fee to allow immediate free access to this article.

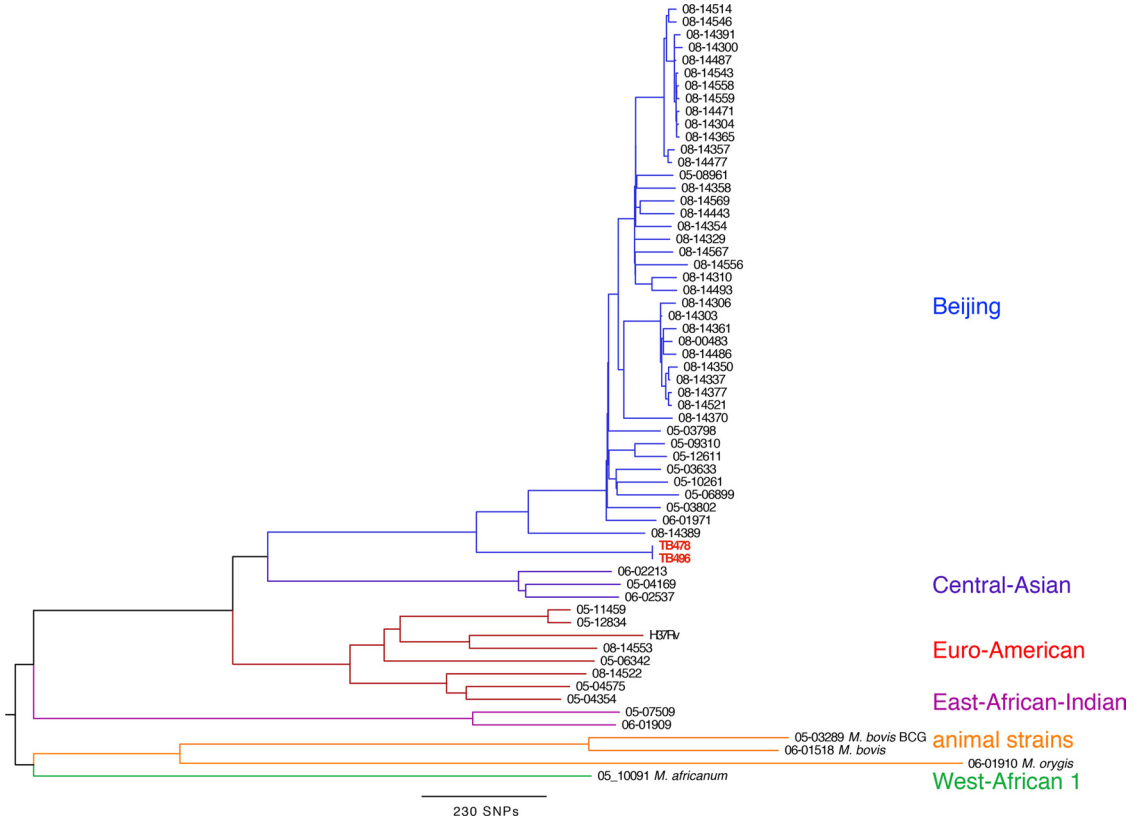

FIG 1 Phylogenetic analysis of *M. tuberculosis* isolates. A phylogenetic tree based on whole-genome sequence data comparing the suspected outbreak strains with previously published representatives of the major lineages of the *M. tuberculosis* complex (5). This includes the various *M. tuberculosis* lineages (Beijing, Central-Asian, Euro-American, and East-African-Indian), as well as one *Mycobacterium africanum* West-African 1 strain and animal strains (*Mycobacterium orygis*, *Mycobacterium bovis*, and the *Mycobacterium bovis* BCG vaccine). Both outbreak strains of patient 1 (TB478) and patient 2 (TB496) belonged to the Beijing lineage, and there were no SNP differences in the mapped genome between the two strains. This suggests transmission between the two patients or from a common, unidentified source.

MIRU-VNTR type (see Table S1 in the supplemental material). Although this supported the possibility of a transmission event, this could not be concluded with confidence, given the endemicity of these strains in Southeast Asia and the inability of current typing methods to discriminate between closely related strains within a given lineage.

Genomic analysis. DNA was extracted from each *M. tuberculosis* isolate (50 ng) and prepared for 150-bp sequencing on a rapid sequencing platform (Illumina MiSeq) as described previously (4). Reads were deposited in the European Nucleotide Archive under the accession numbers ERS128255 (TB478) and ERS128265 (TB496). Sequence reads for the two isolates were mapped against the corrected reference genome *M. tuberculosis* strain H37Rv (5). The average coverage was 82× for TB478 and 102× for TB496 (2,414,734 and 3,048,330 mapped reads, respectively). *M. tuberculosis* isolates described by Casali and colleagues (5) were included in the analysis to place the two outbreak isolates in a broader phylogenetic context. Single nucleotide polymorphisms (SNPs) were identified (6), with SNPs with low quality scores removed and SNPs filtered for presence in at least 75% of mapped reads. A maximum likelihood phylogeny was estimated with RAxML, using the general time-reversible model with gamma correction for among-site variation.

We analyzed the two clinical isolates in the context of a global collection of *M. tuberculosis* isolates (5) and found that both were ancestral Beijing strains (Fig. 1), which agreed with the MIRU-VNTR results (see Table S1 in the supplemental material). The two isolates were genetically indistinguishable, with no SNP differences in the mapped genome, which excluded the 5% of the genome that is not represented due to poor coverage of the repetitive sequence. This strongly suggested that they were either transmitted from one patient to the other or transmitted to both patients from a common but unidentified source.

We then analyzed the 35 genes and, where applicable, their promoter regions that are known to be associated with resistance to antituberculosis agents to test whether we could account for the reported phenotypic resistance results (see Table S2 in the supplemental material). Both isolates shared a promoter-up mutation (C-15T) in *inhA*, the shared target of isoniazid and prothionamide. This mutation is known to generally confer low-level resistance to isoniazid, which can be overcome with a high dose of isoniazid, and full resistance to prothionamide (7). Of note, both isolates had been reported to have high-level resistance to isoniazid. With regard to streptomycin resistance, both isolates shared a novel mutation (V77G) in *gidB*, which encodes a putative 16S rRNA methyltransferase. Mutations in this gene have been shown

to confer low-level resistance to streptomycin resistance (8), and the novel V77G mutation could therefore contribute to the observed phenotype.

Concluding remarks. The primary objective of this study was to determine whether we could use rapid whole-genome sequencing to confirm or refute transmission of *M. tuberculosis* in a suspected outbreak. Despite the presentation of two cases that appeared to be temporally and geographically related, an epidemiological investigation failed to establish a direct link between them. One possible explanation for this is that the outbreak occurred in an educational establishment with a transient, high-turnover student population. Although there were a large number of potential contacts, the majority had left the United Kingdom prior to the start of the screening program. There may have been a common, unidentified index case among this group. Furthermore, only two-thirds of those who were still available for screening consented to be screened. Thus, overall, only 16% of the at-risk population were screened, raising the possibility that there may have been additional cases that were not detected locally and presented elsewhere.

Both *M. tuberculosis* isolates belonged to the Beijing lineage and had the same MIRU-VNTR pattern. The Beijing lineage is common in Southeast Asia, and this finding was consistent with the patients' countries of origin. Identical VNTR types are routinely used to identify clusters of tuberculosis cases that may be epidemiologically linked. Yet, traditional typing methods have a lower resolution than whole-genome sequencing, given that they rely on a small number of molecular surrogates for the complete, underlying evolution of the bacterium (3, 9, 10). In this case, the isolates were indistinguishable even using whole-genome sequencing. Given the low mutation rate of *M. tuberculosis*, which has been estimated to be about 0.39 SNPs per genome per year (11), we could not unequivocally infer a recent common ancestry. Nonetheless, these results afforded us greater certainty than MIRU-VNTR typing that a transmission event had occurred at the language school, either directly between the two patients or from a common but unidentified source.

In clinical practice, phenotypic differences in antibiotic resistance profiles are often used to distinguish bacterial isolates. The observed difference in streptomycin resistance suggested that the two strains were different, providing evidence against a transmission event. However, whole-genome sequencing revealed that the genotypic resistance profiles were identical, thus highlighting the weakness of this inference. The apparent phenotypic difference in susceptibility to streptomycin may have resulted from a lack of reproducibility of phenotypic testing, if the shared *gidB* mutation increased the MICs close to the breakpoint. However, there is some uncertainty whether the breakpoint for streptomycin is set appropriately, highlighting the need for further study of the influence of *gidB* mutations on MICs and treatment outcome (8).

In summary, this study highlights both the potential uses and limitations of rapid whole-genome sequencing for the investigation of *M. tuberculosis*. The low rate of evolution of *M. tuberculosis* at a whole-genome level relative to the rate of transmission represents a significant limitation for epidemiological studies of tuberculosis outbreaks compared to other more rapidly growing organisms, such as methicillin-resistant *Staphylococcus aureus* (MRSA) (4). Moreover, a better understanding of the relationship between genotype and phenotype is required to facilitate genotypic susceptibility testing. Nevertheless, we have demonstrated the utility of whole-genome sequencing for the investigation of a suspected outbreak of tuberculosis and also for the analysis of discrepant drug resistance profiles. The increase in speed accompanied by the falling costs of whole-genome sequencing compared with current phenotypic methods means that this technology is likely to be adopted for routine use in diagnostic microbiology laboratories in well-resourced countries. For this to be realized, the sequence analysis will have to be fully automated, rather than relying on skilled bioinformaticians, as was the case for this study. Such a software tool will require that a continuously updated database of resistance mutations akin to the Stanford HIV Drug Resistance Database (12) is maintained, which is not currently the case (13). Moreover, a well-curated database to share clinical sequence data is called for to track cross-border transmission (13). Indeed, it is plausible that some of the students that had returned to their home countries before we could screen them were exposed to tuberculosis. Should they present with active disease in the future, their isolate could be rapidly linked to the language school, thereby simplifying the outbreak investigations by the health authorities in question.

ACKNOWLEDGMENTS

We thank Carol Steggles (Tuberculosis Specialist Nurse), Jenny Butcher (Norfolk, Suffolk and Cambridgeshire Health Protection Unit), and Claire Jenkins (Cambridge HPA Microbiology Laboratory) for their assistance with this study.

This work was supported by grants from the United Kingdom Clinical Research Collaboration (UKCRC) Translational Infection Research Initiative (TIRI); the Medical Research Council (G1000803), with contributions from the Biotechnology and Biological Sciences Research Council, the National Institute for Health Research (NIHR) on behalf of the United Kingdom Department of Health, and the Chief Scientist of the Scottish Government Health Directorate; the Health Protection Agency; the NIHR Cambridge Biomedical Research Centre; and the Wellcome Trust.

The following authors have potential conflicts of interest to declare: M.E.T. (speaker's honoraria and book royalties from Oxford University Press), G.P.S. (employee of Illumina, Ltd.), J.P. (travel, accommodation, and meeting expenses from Pacific Biosciences and Illumina, Ltd.), and S.J.P. (consultancy fees from Pfizer).

REFERENCES

1. **World Health Organization.** 2011. Global tuberculosis control: WHO report 2011. World Health Organization, Geneva, Switzerland.
2. **Health Protection Agency.** 2012. Tuberculosis in the UK: annual report on tuberculosis surveillance in the UK, 2012. Health Protection Agency, London, United Kingdom.
3. **Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJ, Brinkman FS, Brunham RC, Tang P.** 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Engl. J. Med. **364**:730–739.
4. **Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N. Engl. J. Med. **366**:2267–2275.
5. **Casali N, Nikolayevskyy V, Balabanova Y, Ignatyeva O, Kontsevaya I, Harris SR, Bentley SD, Parkhill J, Nejentsev S, Hoffner SE, Horstmann RD, Brown T, Drobniewski F.** 2012. Microevolution of extensively drug-resistant tuberculosis in Russia. Genome Res. **22**:735–745.
6. **Harris SR, Feil EJ, Holden MT, Quail MA, Nickerson EK, Chantratita N, Gardete S, Tavares A, Day N, Lindsay JA, Edgeworth JD, de Lencastre H, Parkhill J, Peacock SJ, Bentley SD.** 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science **327**: 469–474.

7. **Warren RM, Streicher EM, Gey van Pittius NC, Marais BJ, van der Spuy GD, Victor TC, Sirgel F, Donald PR, van Helden PD.** 2009. The clinical relevance of Mycobacterial pharmacogenetics. Tuberculosis (Edinb.) **89**:199–202.
8. **Wong SY, Lee JS, Kwak HK, Via LE, Boshoff HI, Barry CE.** 2011. Mutations in *gidB* confer low-level streptomycin resistance in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. **55**:2515–2522.
9. **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. doi:10.1371/journal.pone.0007815.
10. **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**:e24737. doi:10.1371/journal.pone.0024737.
11. **Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, Mohaideen N, Ioerger TR, Sacchettini JC, Lipsitch M, Flynn JL, Fortune SM.** 2011. Use of whole genome sequencing to estimate the mutation rate of *Mycobacterium tuberculosis* during latent infection. Nat. Genet. **43**:482–486.
12. **Shafer RW.** 2006. Rationale and uses of a public HIV drug-resistance database. J. Infect. Dis. **194**(Suppl 1):S51–S58.
13. **Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J, Peacock SJ.** 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. **8**:e1002824. doi:10.1371/journal.ppat.1002824.