

Online appendix:

Lifestyle and social network data collection

Interpreters were used when required, and patients were re-interviewed when further information was required. Questions included details of settings visited in the previous two years, including educational institutions, workplaces, places of worship, social settings (e.g. pubs/bars/clubs), prisons, rehabilitation centres, hostels/homeless shelters, and hospitals. Further questions were asked on any previous exposure to TB, travel within the UK or abroad, having visitors from abroad or elsewhere in the UK, and previous addresses of residence. Additional information relating to previous treatment abroad was collected where possible by contacting clinical or public health colleagues in the countries of origin. Following the review of all available data, cases were asked additional questions about specific possible transmission settings that had been identified.

DNA preparation, Whole Genome Sequencing and phylogenetic analysis

Isolates from 22 cases that met the confirmed MIRU-VNTR clustered case definition were available for sequencing. Results from the epidemiological investigation and those from WGS were held separately and compiled once the phylogenetic trees had been produced. Methods were as previously described⁹. Briefly, cultures were grown in Becton-Dickinson Mycobacterial Growth Indicator Tubes containing modified Middlebrooks 7H9 liquid medium and on Löwenstein-Jensen agar. DNA was extracted and purified using the Fuji Quickgene kit (Fuji-Sciences, France) with an added mechanical disruption step using the MP Biomedicals Fastprep homogeniser and Lysing Matrix B. Isolates were sequenced on the Illumina MiSeq platform at the Wellcome Trust Centre for Human Genetics in Oxford and on the Illumina Genome Analyzer GAII or HiSeq 2000 at the Wellcome Trust Sanger Institute. Sequence reads were processed against the rv37 *M.tuberculosis* reference strain in both

Oxford and the Public Health England National Mycobacterial Reference Laboratory for comparison. Results presented in this paper were those obtained from the Oxford university pipeline. The phylogenetic tree was visualized with FigTree (tree.bio.ed.ac.uk/software/figtree).