ERJ Express. Published on May 10, 2018 as doi: 10.1183/13993003.02313-2017

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

The use of whole-genome sequencing in cluster investigation of an MDR-TB outbreak

Maeve K. Lalor, Nicola Casali, Timothy M. Walker, Laura F. Anderson, Jennifer A. Davidson, Natasha Ratna, Cathy Mullarkey, Mike Gent, Kirsty Foster, Tim Brown, John Magee, Anne Barrett, Derrick W. Crook, Francis Drobniewski, H. Lucy Thomas, Ibrahim Abubakar

Please cite this article as: Lalor MK, Casali N, Walker TM, *et al*. The use of whole-genome sequencing in cluster investigation of an MDR-TB outbreak. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.02313-2017).

The use of whole-genome sequencing in cluster investigation of an MDR-TB outbreak

Maeve K Lalor[1,2], Nicola Casali[3,4], Timothy M Walker[5], Laura F Anderson[1], Jennifer A Davidson[1], Natasha Ratna[1], Cathy Mullarkey[6], Mike Gent[7], Kirsty Foster[8], Tim Brown[3], John Magee[9,10], Anne Barrett[9], Derrick W Crook[5,11], Francis Drobniewski[3,4], H Lucy Thomas[1], Ibrahim Abubakar[1,2]

[1]Tuberculosis section, National Infection Service, Public Health England, [2]Institute for Global Health, University College London, [3]PHE National Mycobacterium Reference Service South, London, Public Health England, [4]Department Infectious Diseases, Imperial College, [5]Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, [6]TB Health Visiting Service, Leeds Community Healthcare, [7]Yorkshire and the Humber Public Health England Centre, Blenheim House, Duncombe Street, [8]North East Public Health England Centre, Public Health England, [9]PHE North of England Mycobacterium Reference Centre, Level 2, Freeman Hospital, [10]School of Biology, Ridley Building 2, Newcastle University, United Kingdom, [11]National Infection Service, Public Health England

Corresponding author: Maeve K Lalor, Tel: +44 (0) 20 8327 7154, maeve.lalor@phe.gov.uk

Take home message

Investigation of MDR-TB TB outbreak with WGS was useful but whether transmission occurred was often inconclusive.

Plain language summary

We identified an outbreak of tuberculosis in the UK where the bacteria were resistant to the most effective antibiotics used to treat tuberculosis. In order to understand in whom and where transmission of tuberculosis was occurring, we sequenced the bacteria to compare the genetic code of the tuberculosis in different people in the outbreak. We found that a third of people who were in the outbreak had no genetic similarities in their tuberculosis. After investigating the outbreak, with detailed information about the patients and the bacteria, we identified that transmission occurred in the UK, but we could not be conclusive about whether transmission occurred in 40% of patients, even though the bacteria were genetically related. This paper illustrates the opportunities and limitations of this new technology in a drug-resistant tuberculosis outbreak.

ABSTRACT

We used whole-genome sequencing to delineate transmission networks and investigate the benefits of whole-genome sequencing during cluster investigation.

We included clustered cases of M/XDR-TB linked by MIRU-VNTR, or epidemiological information in the national cluster B1006, notified between 2007-2013 in the UK. We excluded cases whose isolates differed by >12 SNPs from further investigation. Data relating to patients' social networks were collected.

Twenty-seven cases were investigated, 22 had whole-genome sequencing; 8 (36%) of which were excluded as their isolates differed by >12 SNPs to other cases. Eighteen cases were ruled into the transmission network based on genomic and epidemiological information. Evidence of transmission was inconclusive in 39% (7/18) of cases in the transmission network following whole-genome sequencing and epidemiological investigation.

This investigation of a drug resistant TB cluster illustrates the opportunities and limitations of whole-genome sequencing in understanding transmission in a setting with a high proportion of migrant cases. The use of WGS should be combined with classical epidemiological methods. However not every cluster will be solvable, regardless of the quality of genomic data.

Key words: Tuberculosis, drug resistance, outbreak, transmission, whole-genome sequencing

INTRODUCTION

Multi-drug resistant tuberculosis (MDR-TB) is a major public health concern globally, with particularly high rates in Europe.[1] In the UK, the number and proportion of MDR/Rifampicin resistant-TB cases more than doubled from 38 cases (1.1%) in 2001 to 95 cases (1.8%) in 2011, and since decreased to 63 cases (1.7%) in 2016.[2] A UK study of MDR-TB cases between 2004 and 2007, combining epidemiological information from cluster investigation with 24-loci mycobacterial interspersed repetitive unit variable tandem repeat (MIRU-VNTR) strain typing, estimated that up to 8.5% of UK MDR-TB cases arose from recent transmission.[3]

Since January 2010, 24-loci MIRU-VNTR typing has been carried out on all culture positive *Mycobacterium tuberculosis* Complex (MtbC) isolates in the UK, and clusters of cases with indistinguishable strain types, which fulfil requisite criteria have been investigated to inform public health action.[4]

As well as providing diagnostic capability to identify MtbC and determine genotypic drug resistance, whole-genome sequencing (WGS) can determine the genetic relatedness between strains with greater resolution than MIRU-VNTR.[5–8] WGS has mainly been used retrospectively to assess transmission networks in outbreaks, including timing and direction of transmission, rather than prospectively during active cluster investigations.[9–16] In England, the roll-out of prospective WGS in routine TB diagnostics began in December 2016 and is expected to cover the whole of England by the end of 2017, replacing MIRU-VNTR.[17]

In 2010, routine cluster review[4] led to the investigation of a possible TB outbreak of the Beijing strain (B1006: MIRU-VNTR 424352332517333456443372) among UK residents. Of the 231 MDR-

TB cases notified between 2010 and 2012, 27% (62) clustered with at least one other MDR-TB case of whom 23% (14/62) were in B1006. As this was the largest MDR-TB cluster and was only the second in this time period to have more than 5 cases, it was considered to be of public health importance. The B1006 strain accounts for up to 25% of MDR-TB isolates tested in the Former Soviet Union.[18] In November 2012, the cluster investigation status was raised to an incident, requiring a national level incident control team to be convened to consider what action to take[4], due to an increase in the number of cases, including six with extensively drug resistant TB (XDR-TB), and the suspicion that transmission of XDR-TB was occurring in the UK. As part of this intensified investigation, more detailed epidemiological information was collected, and WGS was performed on isolates from patients.

Here we describe the impact WGS had on the on-going cluster investigation by ascertaining whether the greater discrimination of WGS reduced the number of cases initially identified by MIRU-VNTR that required further investigation, whether a clearer understanding of the transmission chain within this cluster was possible combining the epidemiological links identified during the investigation and the WGS data, and whether WGS helped to elucidate the direction and timing of transmission.

METHODS

TB cases in the UK notified to the Enhanced Tuberculosis Surveillance system (ETS) were matched probablistically[19] to laboratory results from culture positive isolates, including data on drug susceptibility testing and MIRU-VNTR. ETS collects demographic, clinical, social risk factor (SRF) and treatment outcome data. National TB clusters (defined as at least two cases with indistinguishable MIRU-VNTR strain types in more than one region) were identified using bespoke

software and reviewed monthly in accordance with national guidance, with cluster investigations launched where appropriate.[4]

Cases of M/XDR-TB notified in the UK between 2007 and 2013 with the B1006 MIRU-VNTR profile (424352332517333456443372, with at least 23 complete loci), were included and referred to as 'B1006 clustered cases'. Culture negative cases with a clinical diagnosis of TB and treated for MDR-TB who were epidemiologically linked to cluster B1006 were considered 'probable' clustered cases. MDR-TB isolates with the B1006 strain from cases prior to 2010 that had been typed retrospectively were also included.

Data relating to lifestyle and social networks were collected through questionnaires case managers completed with their patients (**Online appendix**).

A 'confirmed' epidemiological link was defined between two cases where either volunteered the name of the other as a contact, or where cases shared time in the same setting during the period when one of the cases were potentially infectious. A 'probable' epidemiological link was defined where both cases had spent time in the same setting, but the timing was uncertain. A 'possible' epidemiological link was defined where two cases in the same geographical area shared social or behavioural traits (e.g. drug use) but a specific shared setting could not be established.

Cases were considered still 'clustered' following WGS if sequencing data differed by 12[9] or less Single Nucleotide Polymorphisms (SNPs) from another B1006 case, or if they had an epidemiological link to a WGS clustered case. Several SNP cut-offs previously used in tuberculosis investigations of 0-5[11,15] and 0-9 were explored as alternative thresholds to assess if these better captured transmission.[15]

For methods on DNA preparation, whole genome sequencing and phylogenetic analysis see **Online Appendix**.

RESULTS

Twenty-seven cases notified in the UK between 2007 and 2013 were included in cluster B1006; 24 of which were B1006 clustered cases and 3 were probable clustered cases **(Figure 1)**. The majority were non-UK born (20/27) including 14 who were born in Lithuania, and many had a history of SRFs (15/27).

Twenty-two patient isolates were sequenced. Of these, 8/22 (36%) isolates differed by >12 SNPs from all other isolates in the cluster (range 30-179 SNPs difference) **(Figure 2A**). As none were epidemiologically linked within the cluster, they were excluded from further consideration. 14 cases were thus linked by WGS **(Figure 2B)** with a further four, with no sequencing data (3 had no culture and 1 culture could not be re-grown) linked epidemiologically, resulting in a final cluster of 18 cases.

13/18 (72%) cases had at least one SRF (drug misuse (6), alcohol misuse (5), homelessness (5), imprisonment (9), previous TB treatment (4)). 11/18 cases were born in Lithuania, six in the UK and one in South Eastern Europe. Three were children under the age of five, all born in the UK, two to Lithuanian parents. 12/18 (66%) cases were notified between August 2012 and August 2013, including 5 UK born patients, suggesting possible transmission within the UK **(Figure 3)**.

15/18 (83%) had pulmonary disease, 11 (61%) of whom were sputum smear positive. Five of the pulmonary cases were symptomatic for at least 6 months before starting treatment, one of whom was symptomatic for 2 years (LIT12), and all five had chaotic lifestyles with multiple SRFs. Two cases remained culture positive for more than a year after starting treatment (LIT13 and LIT24).

All 15 culture confirmed cases in the cluster had isolates with phenotypic resistance to isoniazid, rifampicin, ethambutol, streptomycin and kanamycin. There were differences in drug susceptibility testing (DST) for pyrazinamide, prothionomide, ethionomide, moxifloxacin and ofloxacin **(Figure 3)**. The 8 cases in Region A and Region C shared the same DST profile and had pyrazinamide resistance; the 3 cases in Region D shared the same profile and were sensitive to pyrazinamide. The four XDR-TB cases in Region B had resistance to moxifloxacin and ofloxacin, 3 of whom were also resistant to prothionamide and ethionamide.

Potential transmission networks

Amongst the 14 cases with isolates within 12 SNPs of each other, each was genomically linked to between 6 and 13 people, but there was clear sub-clustering within each geographic region **(Figure 2B and 2C)**.

Using the 5 SNP threshold suggested there were three unlinked local outbreaks **(Figure 4A)**. Increasing the threshold to 9 SNPs **(Figure 4B)** suggested transmission may have been more widespread between Region A, Region B and Region C, and increasing the threshold to 12 SNPs suggested transmission may have occurred across all geographical areas **(Figure 2C)**.

By combining the epidemiological information, DST profiles, and genomic data, potential transmission networks were identified in four regions of England **(Figure 5)**.

In Region A **(Figure 5A)**, despite the low number of SNPs between cases, epidemiological links could not be identified; it was therefore inconclusive as to whether recent transmission had occurred. The first case was UK born, had extra-pulmonary TB, and shared no distinctive characteristics with,

or epidemiological links to, the other cases in Region A. Her isolate was 2 and 4 SNPs from that of two Lithuanian cases.

In Region B **(Figure 5B)**, there was one probable epidemiological link between two cases who had lived together (timing and location unknown) but their isolates had 12 SNPs difference and different DST profiles, making direct transmission unlikely. Evidence for transmission between the four XDR-TB cases was inconclusive as the epidemiological, microbiological and genomic data were not supportive.

In Region C **(Figure 5C)**, all cases with WGS were 5 or fewer SNPs from at least one other case. The epidemiological data linking patients by household, workplace, and hospital were suggestive of transmission. The most parsimonious interpretation of the phylogeny suggested LIT08 may have transmitted to LIT20 and LIT21, and that LIT20 may have transmitted to LIT09 **(Figure 2A,B)**. However, closer inspection of the sequence data suggests LIT08 is more likely to have been the source of both LIT20 and LIT09, as the variant distinguishing these two sequences from LIT08 is present as a minority allele in LIT08. The epidemiological information did not support LIT20 transmitting to LIT09 but supported the interpretation of the variants that LIT09 was likely to have been infected by LIT08 (the child's parent).

In Region D **(Figure 5D)**, both the WGS and epidemiological data were suggestive of transmission. Isolates were 0 SNPs apart, and the cases were linked through their households. The direction of transmission was inferred from the epidemiological data, but as there were 0 SNP differences between isolates, no additional information relating to the direction of transmission could be determined by the WGS results.

Identification of epidemiological links and transmission settings: added value of WGS in this cluster investigation

A total of 24 epidemiological links between cases were identified **(Table 1)**, 8 (33%) through routine contact tracing (7 household contacts, 1 relative) and an additional 16 epidemiological links (1 prison, 2 work, 2 neighbours, 1 probable previous household (HH), 9 possible work/drug use contact) following MIRU-VNTR cluster investigation. Two of the 8 links that were identified through routine contact tracing and 5/16 links identified following MIRU-VNTR results were confirmed by WGS. The remaining 7 links identified through MIRU-VNTR investigation had more than 5 SNPs between them and were thus unclear and two had more than 12 SNPs and were thus refuted by WGS. Two new settings were identified (a work place and a hospital). Public health actions were undertaken at the work place, but no new cases were identified.

Table 1: Epidemiological links identified following action taken (contact tracing, MIRU-VNTR cluster investigation and WGS). Red shading refers to "Link identifed following", orange "Link confirmed by", blue "Link queried by" and Green "Link refuted by"

	Potential link identified by			Cluster investigation epidemiological link		
CASES LINKED	**CONTACT INVESTIGATION**	**MIRU-VNTR CLUSTER INVESTIGATION**	**WGS**	**STRENGTH OF LINK**	**TYPE OF CONTACT**	**SETTING OF CONTACT**
LIT10-LIT26	Yes	Confirmed	Confirmed 0 SNPs	Known	Relative	Household
LIT08-LIT09	Yes	Confirmed	Confirmed 3 SNPs	Known	Relative (Parent-child)	Household
LIT08-LIT23	Yes	No culture for either	No culture for either	Known	Relative (Husband-Wife)	Household
LIT08-LIT22	Yes	No culture for either	No culture for either	Known		Household
LIT23-LIT09	Yes	No culture for one	No culture for one	Known	Relative (Parent-chilld)	Household
LIT23-LIT22	Yes	No culture for one	No culture for one	Known		Household
LIT09-LIT22	Yes	No culture for one	No culture for one	Known		Household
LIT24-LIT27	Yes	No culture for one	No culture for one	Known	Relative (Parent-child)	Do not live together
LIT06-LIT07	No	Yes	No growth for one	Known		Prison
LIT08-LIT20	No	Yes	Confirmed 1 SNP	Probable	Work contact	Work
LIT23-LIT20	No	No culture for one	No culture for one	Probable	Work contact	Work
LIT24-LIT26	No	Yes	Confirmed 0 SNPs	Probable		Neighbours
LIT10-LIT24	No	Yes	Confirmed 0 SNPs	Probable		Neighbours
LIT09-LIT21	No	Yes (query contamination)	Confirmed 5 SNPs	Probable		Hospital
LIT12-LIT02	No	Yes	Queried 12 SNPs	Probable	Friends	Previous household (timing not known)
LIT06-LIT12	No	Yes	Confirmed 4 SNPs	Possible		Construction in Region A
LIT02-LIT13	No	Yes	Queried 10 SNPs	Possible		Agricultural work
LIT02-LIT14	No	Yes	Queried 12 SNPs	Possible		Cannabis use/agricultural work
LIT14-LIT13	No	Yes	Queried 12 SNPs	Possible		Agricultural work
LIT06-LIT08	No	Yes	Queried 6 SNPs	Possible		Construction in Region A
LIT12-LIT08	No	Yes	Queried 6 SNPs	Possible		Construction in Region A
LIT14-LIT01	No	Yes	Queried 8 SNPs	Possible		Cannabis use/agricultural work
LIT02-LIT01	No	Yes	Refuted 16 SNPs	Possible		Cannabis use/agricultural work
LIT01-LIT13	No	Yes	Refuted 16 SNPs	Possible		Agricultural work

DISCUSSION

A large M/XDR-TB cluster was identified in the UK by MIRU-VNTR strain typing, enabling prompt cluster investigation. Detailed analyses of epidemiological and genomic data provided strong evidence transmission had occurred in the UK. Consistent with recognised limitations of MIRU-VNTR for the Beijing lineage [20], the use of WGS allowed discrimination between cases clustered by MIRU-VNTR and the exclusion of one-third of cases from the investigation. This allowed resources to be focused on the investigation of cases that were more likely to have been part of the same transmission network.

As well as the evidence WGS provided to refute transmission MIRU-VNTR had identified, WGS also provided corroborative evidence that transmission was likely to have occurred in a small number of cases, who had either confirmed or probable epidemiological links. In most cases the WGS data were consistent with the combined findings from MIRU-VNTR and epidemiological investigation. In one instance WGS suggested transmission could have occurred (separated by 2 and 4 SNPs to other cases), yet this seemed epidemiologically implausible as the case with earliest symptom onset had extra-pulmonary TB, and no known links with two subsequent cases. One possible explanation may be the presence of unknown intermediary cases.

Whilst there are anecdotal data where phylogeny have indicated the direction of transmission in TB outbreaks,[15] this analysis also underscores the importance of exploring the raw sequence data at variant sites. During the active investigation of this cluster, and in the routine use of WGS data in cluster investigation, only the phylogenetic tree was used to direct public health action. Only after the incident, when trying to understand why the tree and epidemiological data were inconsistent did we find the presence of a variant allele seen in secondary strains, as a minority allele in an ancestral strain. This deeper analysis of the sequence data concorded with the epidemiological data to understand the possible direction of transmission between cases in this cluster as was recently

demonstrated by Worby et al in other pathogens.[21] The addition of WGS did not facilitate the identification of additional links or missing cases identified in this setting.

The most effective SNP threshold to apply in practice to most efficiently identify transmission networks in tuberculosis has previously been discussed.[16–18,22] While our results suggest that identifying large numbers of SNPs between isolates is extremely helpful for refuting transmission, we found that the 12 SNP cut-off is likely to over-estimate recent transmission, for example significant inter-regional transmission in this cluster looked didn't look plausible and small SNP differences between isolates did not always lead to the identification of epidemiological links, as was also recently shown in a large isoniazid cluster in London.[22] Although distances of a 0-2, or even 0-5 SNPs would normally suggest a high probability of recent transmission, this outbreak highlights the need to remain aware of exceptions to this, particularly when a large number of cases occur in migrants whose disease may be due to reactivation of distantly acquired infection. Other factors may also affect the assessment of transmission through SNP differences such as clinical disease manifestation, duration of infection, patient bacterial load, antibiotic therapy, acquisition of drug resistance and actual infective dose which would need building into any model using SNP number to predict transmission events.

Many of the patients in this cluster had chaotic lifestyles and multiple risk factors associated with delayed diagnosis and poor adherence to treatment, plausibly contributing to further transmission of this strain. Cluster investigation in populations with a high number of risk factors and drug resistant TB requires considerable resources. The lifestyles of some of the cases in this cluster presented a challenge to TB control, as it was difficult to collect epidemiological information, and indeed epidemiological links between cases who were genotypically linked may not have been identified in this cluster. Such investigations can be resource intensive, but to prevent transmission of TB,

including of drug resistant TB, in high risk groups, cluster investigation remains a key component of disease control.

Routine cluster investigation based on MIRU-VNTR was recently scaled back in the UK, in part due to a lack of evidence on its cost effectiveness in preventing further transmission, and in part due to a lack of available resources.[23] Frequently considerable time and resources were invested without identifying additional epidemiological links, transmission settings or cases. This may be in part due to MIRU-VNTR not being able to adequately distinguish clusters in some TB lineages or due to the high proportion of imported TB cases who reactivate in the UK.[24] Other low incidence settings have found similar issues relating to the specificity of MIRU-VNTR.[25] To address this concern, a more targeted approach for flagging high risk national MIRU-VNTR clusters by running a red flag algorithm to identify priority clusters, including clusters with cases with MDR-TB for review, was developed in England. Furthermore, due to the poor resolution of MIRU-VNTR for some TB lineages[20] and high levels of clustering in non-UK born patients with no identified epidemiological links, many clusters with predominantly non-UK born populations in the UK have been assumed to represent common imported strain types and investigation has not been prioritised. The results from this analysis show transmission may be occurring in the UK in a subset of these clusters. Recent analysis suggests WGS is cost effective due to the parallel identification of drug resistant strains.[5] Due to the ability of WGS to predict drug resistance and its greater resolution, replacement of MIRU-VNTR typing may be cost effective. Modelling the cost per quality adjusted life year gained with this technology would be useful.

WGS is now being used routinely in the TB diagnostic pathway in England[26]; potential benefits will include faster results for speciation of mycobacteria, prediction of drug susceptibility and relatedness of cases in a single process[5]. The long term systematic use of WGS should also enable better analysis of transmission dynamics at a population level in England than was possible with MIRU-VNTR, in

order to monitor the impact of policy changes on transmission[27,28]. As the roll-out of WGS is underway, clinical and public health teams have begun to make use of this new technology describing both the benefits in an XDR-TB cluster in London and the limitations in a large isoniazid resistant cluster in London.[22,29]

Bayesian inference methods, which combine information on SNP differences between isolates, time to nearest common ancestor, and epidemiological data, are likely to be of benefit for helping to understand possible transmission networks, and to inform public health action.[30] If automated algorithms could be applied, combining WGS data and epidemiological data informed by highly predictive models, cluster investigation may be considered in middle and other high income settings. In resource poor settings where TB transmission is likely to be more common, little use of real-time cluster investigation has occurred largely due to lack of available resources.

The introduction of universal WGS in England will undoubtedly revolutionise testing for antibiotic resistance[26]. In addition, using WGS in this cluster investigation provided evidence that a third of the cases identified on the basis of this MIRU-VNTR type were not plausibly part of the same transmission network, thus enabling us to focus additional investigative resources on a smaller number of cases, and provided supportive evidence transmission had occurred in a small number of cases with confirmed epidemiological links. Despite the obvious increase in granularity of WGS data, the evidence of whether some cases in this cluster were part of the transmission chain was inconclusive, especially in non-UK born cases. These data suggest that, as was seen using MIRU-VNTR, WGS in combination with epidemiological investigations may not enable the determination of whether recent, and hence UK-based, transmission has occurred in all cases. This study emphasises the importance of the use of classical epidemiological methods; a significant proportion of the epidemiological links we identified were as a result of routine contact tracing. The use of

genotyping data alone, whether that be MIRU-VNTR or WGS, likely over-estimates transmission, resulting in inconclusive determination of networks. Whilst WGS is best viewed as a tool that directs epidemiological investigations with optimal precision, future research should evaluate the impact of WGS use on subsequent public health action and detection of previously unrecognised cases in cluster investigation. This may allow the re-initiation of routine prospective cluster investigations in resource rich, low incidence settings approaching tuberculosis elimination.

Declarations of interest

FD and NC were supported by the Imperial BRC. All other authors declare they have no financial or other conflict of interest.

Ethics approval and consent to participate

Public Health England has authority under the Health and Social Care Act 2012 to hold and analyse national surveillance data for public health and research purposes.

Author's contributions

MKL, HLT and IA conceived and designed the study. MKL, LFA, JAD, NR, CM, MG, KF and FD collected data for the study. TW, NC conducted sequencing and sequencing analysis for the study. MKL carried out the analysis and writing of the manuscript. TB, JM, AB and FD provided isolates for the study. All authors were involved in the incident, contributed to the design of the analysis and commented on manuscript versions. All authors have read and approved the final version of this manuscript.

Funding source

This study was supported by Public Health England, no external funding was received.

In Region A **(Figure 5A)**, the first case (LIT99) was UK born, had extra-pulmonary TB, and no SRFs. She shared no distinctive characteristics with, or epidemiological links to, the other cases in Region A, despite her isolate being separated by only 2 and 4 SNPs from isolates from two of the Lithuanian cases. The low number of SNPs is consistent with presumed recent transmission, however, given that the first case had extra-pulmonary TB and there was no supporting epidemiological evidence it is unlikely she was the source of infection for the subsequent cases. There was one known epidemiological link between two prisoners, but only one had WGS results. The three Lithuanian cases whose isolates were separated by 4-6 SNPs had all worked in the construction industry, although it is unknown whether they ever encountered each other. All cases had the same phenotypic DST profile.

In Region B **(Figure 5B)**, there was one probable epidemiological link between two cases who had lived together at some point (timing and location of shared residency unknown); these cases had 12 SNPs difference between isolates, and they had different DST profiles. All four XDR-TB cases in this location worked in the agricultural industry for agencies with frequent job changes; three of them reported smoking cannabis and three had been in prison in Lithuania, although at different times to each other. The XDR cases had all previously been treated for TB, 3 on more than one occasion in Lithuania (strain type and sensitivities of initial treatment unknown). The four XDR-TB cases had between 8 and 12 SNPs between their isolates and two had different resistance profiles. No confirmed epidemiological links between them could be identified. Evidence for transmission of XDR-TB was therefore inconclusive as the epidemiological, microbiological and genomic data were not supportive when combined.

In Region C **(Figure 5C)**, the first four cases notified were within two families who shared a house. Another case was linked to the workplace where the adults in the household worked and a further

UK born case was treated in the same hospital that the other cases were treated at, but no definitive setting or timing was identified. All cases with WGS results in Region C were 5 or fewer SNPs apart from at least one other case in Region C, confirming the epidemiological evidence for transmission. The most parsimonious interpretation of the phylogenetic tree suggested that LIT08 is likely to have transmitted to LIT20 and LIT21 and that LIT20 is likely to have transmitted to LIT09 **(Figure 2A,B)**. However, closer inspection of the sequence data suggests LIT08 is more likely to have been the source of both LIT20 and LIT09, as the variant distinguishing these two sequences from LIT08 is present as a minority allele in LIT08. The epidemiological information did not support LIT20 transmitting to LIT09 but supported the interpretation of the variants that LIT09 was likely to have been infected by LIT08 (the child's parent).

In Region D **(Figure 5D)**, the first two cases notified lived close to each other with indistinguishable isolates. However, despite intensive investigations, no common setting or evidence that they knew each other was identified. General awareness raising of signs and symptoms of TB and information about local TB services was undertaken with services for vulnerable groups but no active case finding was undertaken. Two further cases were household contacts of the first and second cases, and the isolate of the household contact of the first case was indistinguishable to the first case. No sample was taken from the other household contact. The WGS results supported the epidemiological data that transmission had occurred. The direction of transmission was inferred from the epidemiological data, but as there were 0 SNP differences between isolates, no additional information relating to timing or the direction of transmission could be determined by the WGS results.

References

1 WHO | Global tuberculosis report 2015. WHO. http://www.who.int/tb/publications/global_report/en/ (accessed Jan 29, 2016).

2 Reports of cases of TB to UK enhanced tuberculosis surveillance systems - GOV.UK. https://www.gov.uk/government/statistics/reports-of-cases-of-tb-to-uk-enhanced-tuberculosis-surveillance-systems (accessed Nov 2, 2017).

3 Anderson LF, Tamne S, Brown T, *et al.* Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. *Lancet Infect Dis* 2014; **14**: 406–15.

4 TB strain typing and cluster investigation: handbook - Publications - GOV.UK. https://www.gov.uk/government/publications/tb-strain-typing-and-cluster-investigation-handbook (accessed Jan 29, 2016).

5 Pankhurst LJ, Del Ojo Elias C, Votintseva AA, *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. *Lancet Respir Med* 2016; **4**: 49–58.

6 Takiff HE, Feo O. Clinical value of whole-genome sequencing of Mycobacterium tuberculosis. *Lancet Infect Dis* 2015; **15**: 1077–90.

7 Borgdorff MW, van Soolingen D. The re-emergence of tuberculosis: what have we learnt from molecular epidemiology? *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis* 2013; **19**: 889–901.

8 Le VTM, Diep BA. Selected insights from application of whole-genome sequencing for outbreak investigations. *Curr Opin Crit Care* 2013; **19**: 432–9.

9 Walker TM, Lalor MK, Broda A, *et al.* Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med* 2014; **2**: 285–92.

10 Glynn JR, Guerra-Assunção JA, Houben RMGJ, *et al.* Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. *PloS One* 2015; **10**: e0132840.

11 Guerra-Assunção JA, Crampin AC, Houben RMGJ, *et al.* Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. *eLife* 2015; **4**. DOI:10.7554/eLife.05166.

12 Clark TG, Mallard K, Coll F, *et al.* Elucidating emergence and transmission of multidrug-resistant tuberculosis in treatment experienced patients by whole genome sequencing. *PloS One* 2013; **8**: e83012.

13 Saelens JW, Lau-Bonilla D, Moller A, *et al.* Whole genome sequencing identifies circulating Beijing-lineage Mycobacterium tuberculosis strains in Guatemala and an associated urban outbreak. *Tuberc Edinb Scotl* 2015; **95**: 810–6.

14 Gardy JL, Johnston JC, Ho Sui SJ, *et al.* Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. *N Engl J Med* 2011; **364**: 730–9.

15 Walker TM, Ip CL, Harrell RH, *et al.* Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. *Lancet Infect Dis* 2013; **13**: 137–46.

16 Bryant JM, Schürch AC, van Deutekom H, *et al.* Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. *BMC Infect Dis* 2013; **13**: 110.

17 Hatherell H-A, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. *BMC Med* 2016; **14**: 21.

18 Casali N, Nikolayevskyy V, Balabanova Y, *et al.* Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* 2014; **46**: 279–86.

19 Aldridge RW, Shaji K, Hayward AC, Abubakar I. Accuracy of Probabilistic Linkage Using the Enhanced Matching System for Public Health and Epidemiological Studies. *PLoS ONE* 2015; **10**: e0136179.

20 Allix-Béguec C, Wahl C, Hanekom M, *et al.* Proposal of a consensus set of hypervariable mycobacterial interspersed repetitive-unit-variable-number tandem-repeat loci for subtyping of Mycobacterium tuberculosis Beijing isolates. *J Clin Microbiol* 2014; **52**: 164–72.

21 Worby CJ, Lipsitch M, Hanage WP. Shared Genomic Variants: Identification of Transmission Routes Using Pathogen Deep-Sequence Data. *Am J Epidemiol* 2017; **186**: 1209–16.

22 Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. *PLOS Med* 2016; **13**: e1002137.

23 Mears J, Vynnycky E, Lord J, *et al.* The prospective evaluation of the TB strain typing service in England: a mixed methods study. *Thorax* 2015; published online April 16. DOI:10.1136/thoraxjnl-2014-206480.

24 Mears J, Abubakar I, Crisp D, *et al.* Prospective evaluation of a complex public health intervention: lessons from an initial and follow-up cross-sectional survey of the tuberculosis strain typing service in England. *BMC Public Health* 2014; **14**: 1023.

25 Verma A, Schwartzman K, Behr MA, *et al.* Accuracy of prospective space-time surveillance in detecting tuberculosis transmission. *Spat Spatio-Temporal Epidemiol* 2014; **8**: 47–54.

26 Walker TM, Cruz ALG, Peto TE, Smith EG, Esmail H, Crook DW. Tuberculosis is changing. *Lancet Infect Dis* 2017; **17**: 359–61.

27 Hamblion EL, Menach AL, Anderson LF, *et al.* Recent TB transmission, clustering and predictors of large clusters in London, 2010–2012: results from first 3 years of universal MIRU-VNTR strain typing. *Thorax* 2016; **71**: 749–56.

28 Dedicoat M, Cooke GS. Can routine genetic testing help to end TB transmission? *Thorax* 2016; **71**: 681–2.

29 Arnold A, Witney AA, Vergnano S, *et al.* XDR-TB transmission in London: Case management and contact tracing investigation assisted by early whole genome sequencing. *J Infect* 2016; **73**: 210–8.

30 Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014; **31**: 1869–79.

Figure Legends

Figure 1: Flow chart of cases in cluster B1006

Initial intensified investigation (n=27). 18 cases remained under investigation (highlighted in red) following WGS results, which included 14 clustered cases with <12 SNPs between each case and another case, 3 probable clustered cases with no culture (or WGS) who were epidemiologically linked to cases in the cluster, and 1 clustered case with MIRU-VNTR B1006, but no WGS epi linked to another case in the cluster.

Figure 2: A) Phylogenetic tree showing genetic diversity between isolates sequenced from 22 cases in cluster B1006. Red selection shows isolates that are a maximum of 12 SNP differences from at least one other isolate in the cluster. **B) Phylogenetic tree of isolates, sequenced from 14 cases, with 12 or less SNP differences from at least one other isolate in the cluster.** Black numbers represent the number of SNPs, red numbering represents order of cases in terms of onset of symptoms. Green place names show geographical location of cases. **C) Table of SNP differences between isolates, sequenced from 14 cases, with 12 or less SNP differences from at least one other isolate in the cluster.** Red highlighting represents 0-5 SNPs, yellow represents 6-12 SNPs and white represents 13-16 SNPs. The final two columns show how many cases the case is linked to using 5 or 12 SNP cut offs.

Figure 3: Time-line showing characteristics cases in B1006 with onset of symptoms until the start of MDR-TB treatment

Figure 4: Cluster diagram showing epidemiological links and WGS data nationally, A] showing SNPs <=5, B] showing SNPs <=9 (legend as in Figure 5)

Figure 5: Cluster diagram showing epidemiological links and WGS data within four areas

Footnotes

Foot note for Figure 5:

Cases who spent time in more than one geographical area are presented in each region.

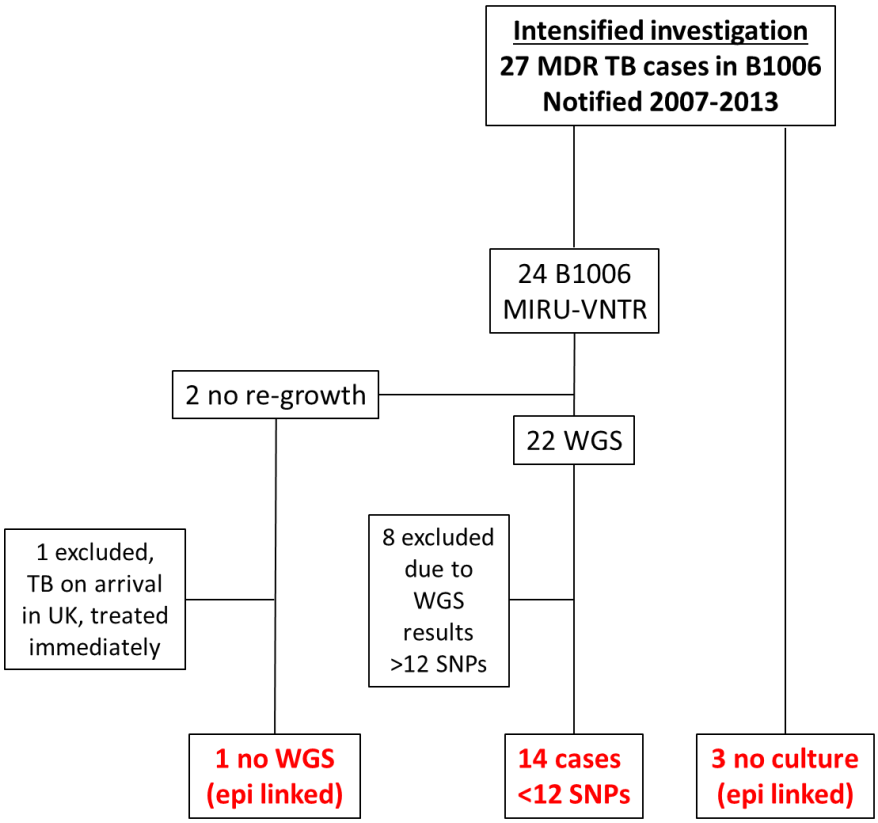
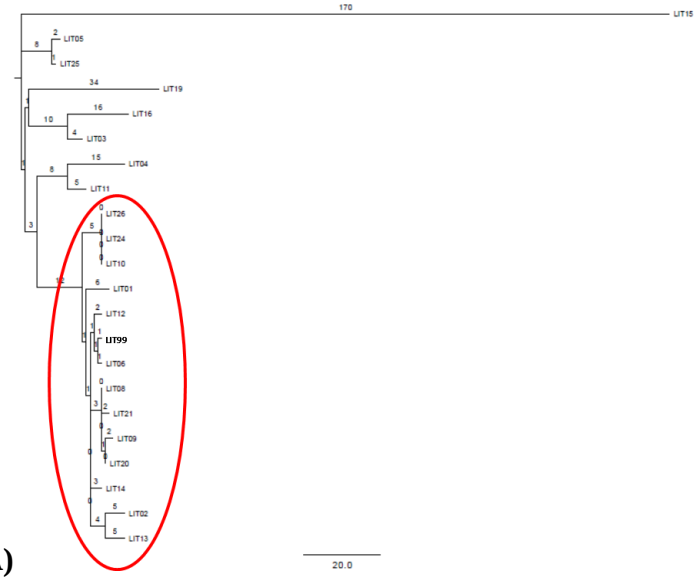
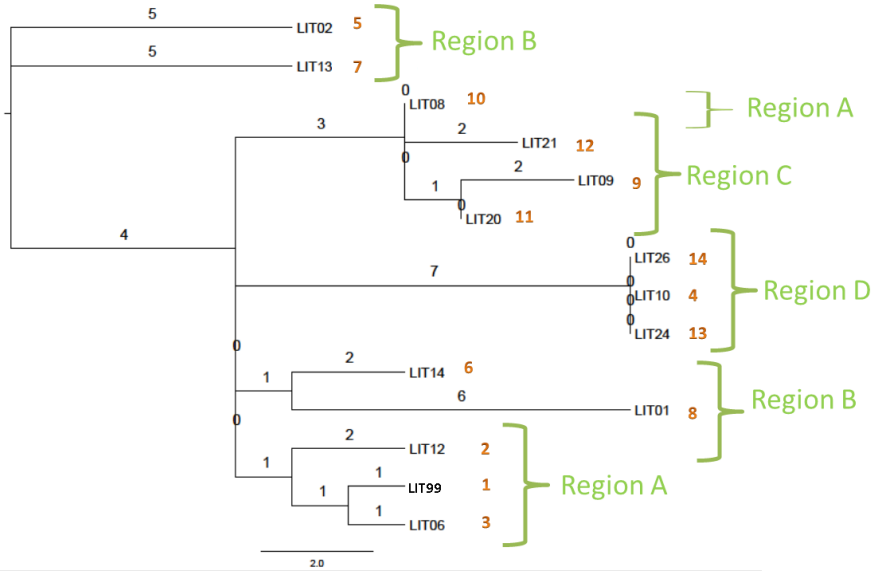

Intensified investigation
27 MDR TB cases in B1006
Notified 2007-2013
24 B1006
MIRU-VNTR
2 no re-growth
22 WGS
1 excluded,
TB on arrival
in UK, treated
immediately
8 excluded
due to
WGS
results
>12 SNPs
1 no WGS
(epi linked)
14 cases
<12 SNPs
3 no culture
(epi linked)

Figure 1

A)



B)



C)

	ID	LIT99	LIT12	LIT06	LIT10	LIT02	LIT14	LIT13	LIT01	LIT09	LIT08	LIT20	LIT21	LIT24	LIT26	≤ 5 SNPs	≤ 12 SNPs
1	LIT99	na	4	2	10	12	6	12	10	9	6	7	8	10	10	2	13
2	LIT12	4	na	4	10	12	6	12	10	9	6	7	8	10	10	2	13
3	LIT06	2	4	na	10	12	6	12	10	9	6	7	8	10	10	2	13
4	LIT10	10	10	10	na	16	10	16	14	13	10	11	12	0	0	2	9
5	LIT02	12	12	12	16	na	12	10	16	15	12	13	14	16	16	0	6
6	LIT14	6	6	6	10	12	na	12	8	9	6	7	8	10	10	0	13
7	LIT13	12	12	12	16	10	12	na	16	15	12	13	14	16	16	0	6
8	LIT01	10	10	10	14	16	8	16	na	13	10	11	12	14	14	0	7
9	LIT09	9	9	9	13	15	9	15	13	na	3	2	5	13	13	3	7
10	LIT08	6	6	6	10	12	6	12	10	3	na	1	2	10	10	3	13
11	LIT20	7	7	7	11	13	7	13	11	2	1	na	3	11	11	3	11
12	LIT21	8	8	8	12	14	8	14	12	5	2	3	na	12	12	3	11
13	LIT24	10	10	10	0	16	10	16	14	13	10	11	12	na	0	2	9
14	LIT26	10	10	10	0	16	10	16	14	13	10	11	12	0	na	2	9

Figure 2

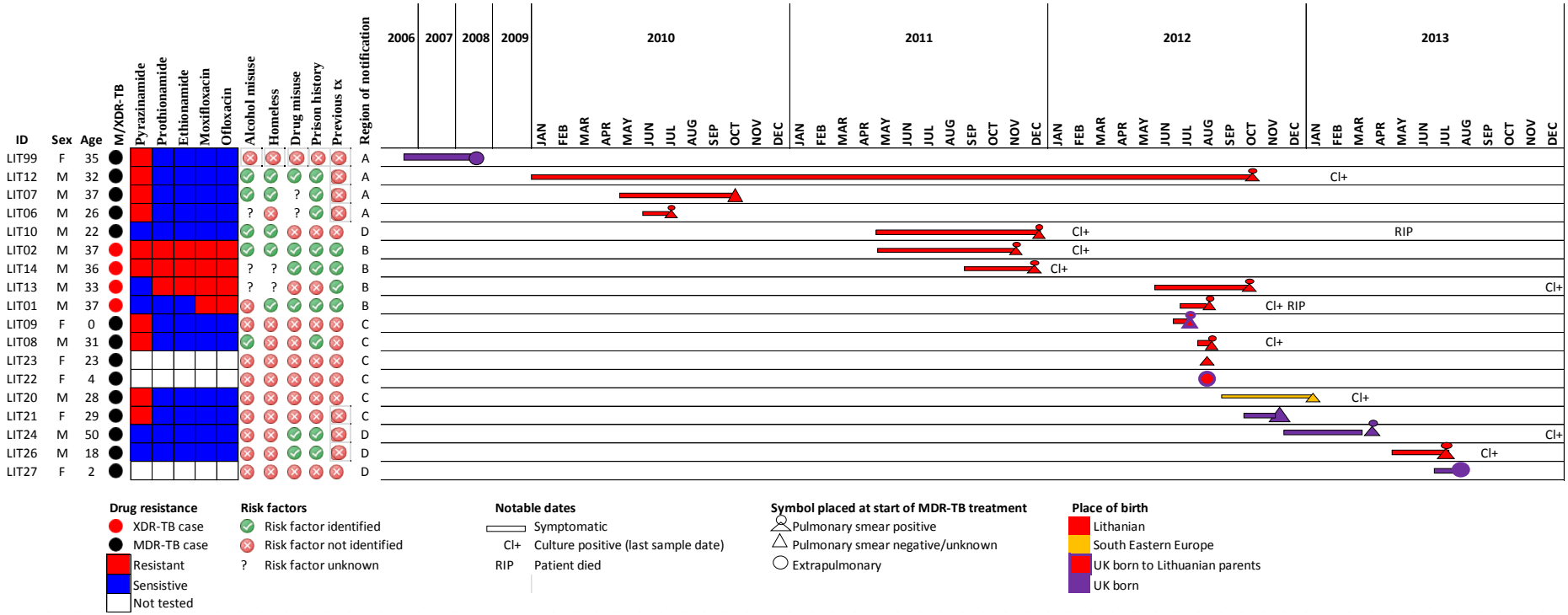

ID
Sex
Age
M/XDR-TB
Pyrazinamide
Prothionamide
Ethionamide
Moxifloxacin
Ofloxacin
Alcohol misuse
Homeless
Drug misuse
Prison history
Previous tx
Region of notification
2006
2007
2008
2009
2010
2011
2012
2013
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
LIT99 F 35 A
LIT12 M 32 A Cl+
LIT07 M 37 A
LIT06 M 26 A
LIT10 M 22 D Cl+ RIP
LIT02 M 37 B Cl+
LIT14 M 36 B Cl+
LIT13 M 33 B Cl+
LIT01 M 37 B Cl+ RIP
LIT09 F 0 C
LIT08 M 31 C Cl+
LIT23 F 23 C
LIT22 F 4 C
LIT20 M 28 C Cl+
LIT21 F 29 C
LIT24 M 50 D Cl+
LIT26 M 18 D Cl+
LIT27 F 2 D
Drug resistance
XDR-TB case
MDR-TB case
Resistant
Sensitive
Not tested
Risk factors
Risk factor identified
Risk factor not identified
? Risk factor unknown
Notable dates
Symptomatic
Cl+ Culture positive (last sample date)
RIP Patient died
Symbol placed at start of MDR-TB treatment
Pulmonary smear positive
Pulmonary smear negative/unknown
Extrapulmonary
Place of birth
Lithanian
South Eastern Europe
UK born to Lithuanian parents
UK born

Figure 3

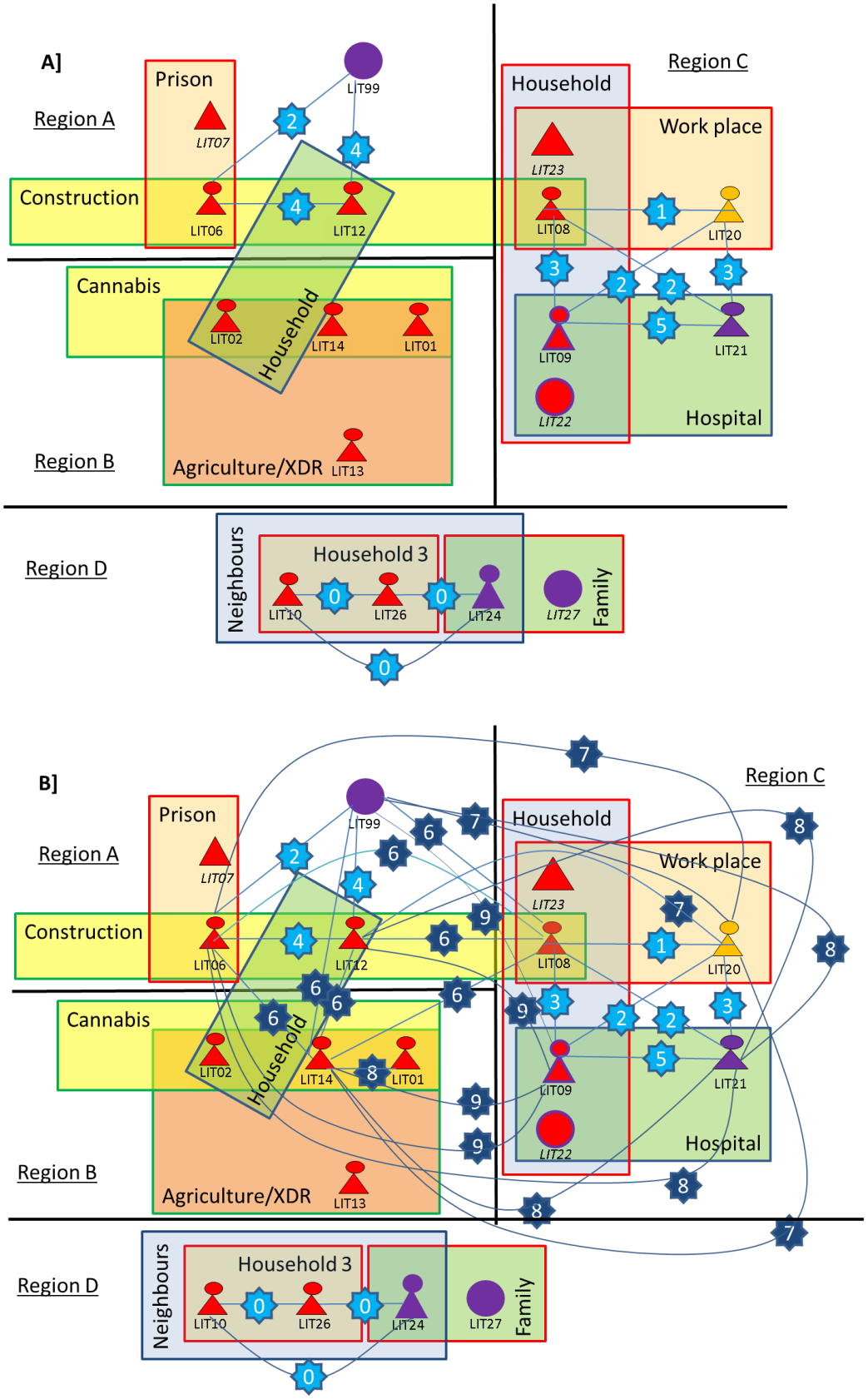

A]
Region A
Prison
LIT07
LIT99
Region C
Household
Work place
LIT23
LIT08
LIT20
Construction
LIT06
LIT12
Cannabis
Household
LIT02
LIT14
LIT01
LIT09
LIT21
LIT22
Hospital
Region B
Agriculture/XDR
LIT13
Region D
Neighbours
Household 3
LIT10
LIT26
LIT24
LIT27
Family
B]
Region A
Prison
Region C
Household
Work place
Construction
Cannabis
Household
Hospital
Region B
Agriculture/XDR
Region D
Neighbours
Household 3
Family

Figure 4

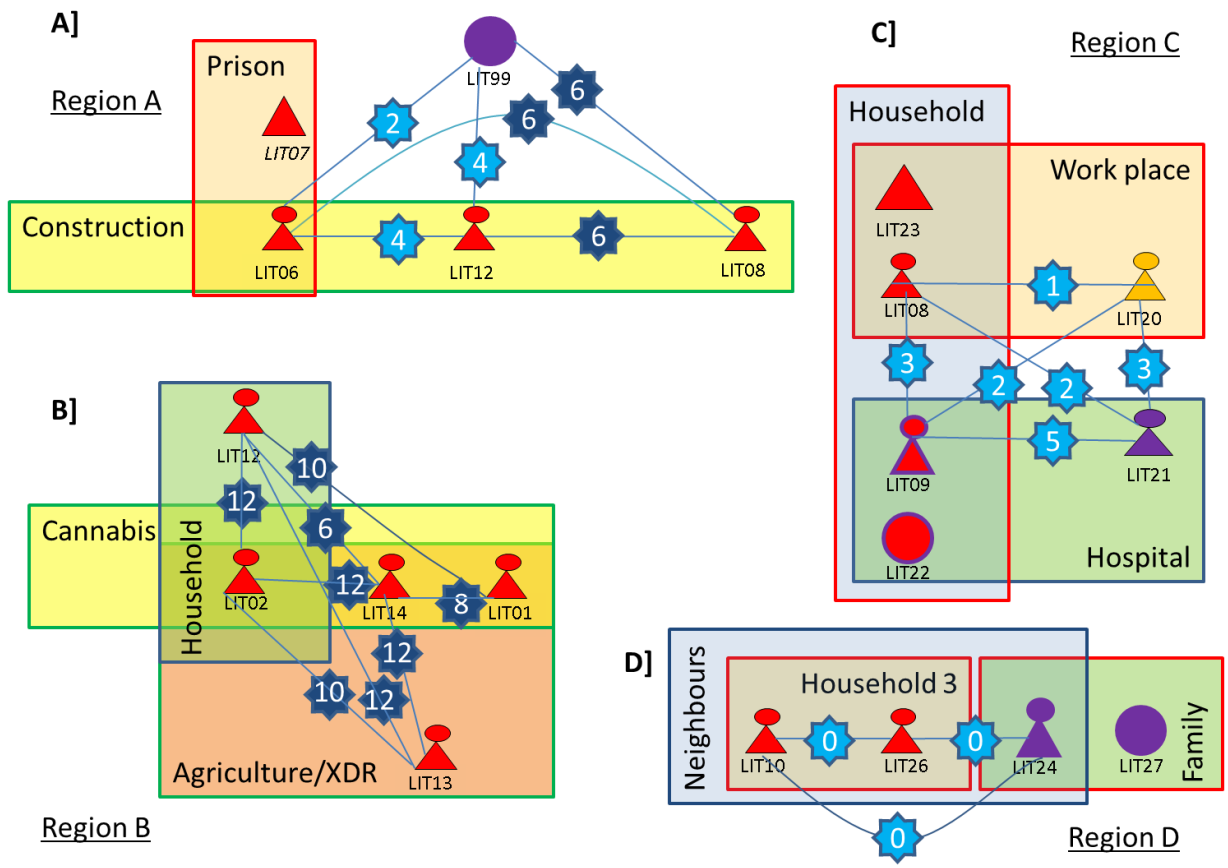

A]
Region A
Prison
LIT07
Construction
LIT06
LIT12
LIT08
LIT99
2
6
6
4
4
6
B]
Cannabis
Household
Agriculture/XDR
LIT12
LIT02
LIT14
LIT01
LIT13
10
12
6
12
8
12
10
12
Region B
C]
Region C
Household
Work place
LIT23
LIT08
LIT20
LIT09
LIT21
LIT22
Hospital
1
3
2
2
3
5
D]
Neighbours
Household 3
LIT10
LIT26
LIT24
LIT27
Family
0
0
0
Region D
Legend
Pulmonary smear positive
Pulmonary smear negative/unknown
Extra-pulmonary
Lithuania
South eastern Europe
UK born, Lithuanian parents
UK born
Known Epidemiological Link
Probable Epidemiological Link
Possible Epidemiological Link
≤5 SNP
6-12 SNP

Figure 5

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 described[9]. 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).