

## ORIGINAL ARTICLE

## Molecular epidemiology of tuberculosis in London 1995–7 showing low rate of active transmission

H Maguire, J W Dale, T D McHugh, P D Butcher, S H Gillespie, A Costetsos, H Al-Ghusein, R Holland, A Dickens, L Marston, P Wilson, R Pitman, D Strachan, F A Drobniewski, D K Banerjee

*Thorax* 2002;57:617–622

**Background:** Tuberculosis notification rates for London have risen dramatically in recent years. Molecular typing of *Mycobacterium tuberculosis* has contributed to our understanding of the epidemiology of tuberculosis throughout the world. This study aimed to assess the degree of recent transmission of *M tuberculosis* in London and subpopulations of the community with high rates of recent transmission.

**Methods:** *M tuberculosis* isolates from all persons from Greater London diagnosed with culture positive tuberculosis between 1 July 1995 and 31 December 1997 were genetically fingerprinted using IS6110 restriction fragment length polymorphism (RFLP) typing. A structured proforma was used during record review of cases of culture confirmed tuberculosis. Cluster analysis was performed and risk factors for clustering were examined in a univariate analysis followed by a logistic regression analysis with membership of a cluster as the outcome variable.

**Results:** RFLP patterns were obtained for 2042 isolates with more than four copies of IS6110; 463 (22.7%) belonged to 169 molecular clusters, which ranged in size from two (65% of clusters) to 12 persons. The estimated rate of recent transmission was 14.4%. Young age (0–19 years) (odds ratio (OR) 2.65, 95% confidence interval (CI) 1.59 to 4.44), birth in the UK (OR 1.55, 95% CI 1.04 to 2.03), black Caribbean ethnic group (OR 2.19, 95% CI 1.15 to 4.16), alcohol dependence (OR 2.33, 95% CI 1.46 to 3.72), and streptomycin resistance (OR 1.82, 95% CI 1.15 to 2.88) were independently associated with an increased risk of clustering.

**Conclusions:** Tuberculosis in London is largely caused by reactivation or importation of infection by recent immigrants. Newly acquired infection is also common among people with recognised risk factors. Preventative interventions and early diagnosis of immigrants from areas with a high incidence of tuberculosis, together with thorough contact tracing and monitoring of treatment outcome among all cases of tuberculosis (especially in groups at higher risk of recent infection), remains most important.

See end of article for authors' affiliations

Correspondence to:  
Dr T D McHugh,  
Department of Medical  
Microbiology, Royal Free &  
University College Medical  
School, Royal Free  
Campus, London  
NW3 2PF, UK;  
tmchugh@rfc.ucl.ac.uk

Revised version received  
6 December 2001  
Accepted for publication  
1 February 2002

Epidemiological studies of tuberculosis have been greatly assisted by the advent of restriction fragment length polymorphism (RFLP) typing using the insertion sequence IS6110,<sup>1</sup> since epidemiologically related isolates will have identical banding patterns. Strains that are indistinguishable are therefore generally assumed to represent recently transmitted infection (rather than reactivation). However, they can reflect reactivation of strains common in the past. Furthermore, identical banding patterns may occur as a result of preferred insertion sites.<sup>2,3</sup> Nevertheless, the technique has proved useful in many parts of the world for exploring the epidemiology of tuberculosis.<sup>4–10</sup>

Between 1992 and 1998 the tuberculosis notification rate in London rose from 23 to 35 per 100 000, representing 41% of notifications in England and Wales in 1998.<sup>11</sup> Previous studies in England and Wales have suggested poverty<sup>12,13</sup> or ethnic origin<sup>14</sup> as important reasons for increased incidence of tuberculosis. Tuberculosis and HIV co-infection has been increasing in parts of London and recent reports have suggested that HIV may be contributing to this rise.<sup>15–17</sup>

## METHODS

### Study population

A multicentre study was conducted in London to describe the RFLP patterns of isolates of *Mycobacterium tuberculosis* from July 1995 to the end of 1997 to assess the degree of apparent recent transmission, to identify subpopulations with high rates of recent transmission, and to determine the magnitude of potential risk factors associated with recent transmission.

Ethical approval for the study was obtained from all district and trust ethics committees in London. Isolates of *M tuberculosis* from persons diagnosed with tuberculosis between 1 July 1995 and 31 December 1997 in any participating laboratory in Greater London (54 NHS laboratories inside the London orbital motorway, M25) were included. Isolates archived at the Public Health Laboratory Service (PHLS) Mycobacterium Reference Unit (MRU) at Dulwich Public Health Laboratory and at the reference facility of the Royal Brompton Hospital were obtained. Isolates of mycobacteria other than tuberculosis were excluded.

### Molecular analysis

Isolates of *M tuberculosis* were genetically fingerprinted using IS6110 RFLP typing using the international standard protocol<sup>1</sup> at St George's Hospital Medical School (SGHMS) and the Royal Free & University College Medical School (RF&UCMS). All patterns were entered by one researcher (AD) onto a database using GelCompar software (Version 4.0, Applied Maths, Koutrai, Belgium) at RF&UCMS and then were analysed independently with the aid of GelCompar at the University of Surrey. Cluster analysis was performed using the Dice coefficient. Similarity defined by this coefficient was calculated using the parameter settings at 1.2% band position tolerance with optimisation. A molecular cluster was defined as a series of isolates with identical banding patterns (100% identity), subject to visual verification. Strains that differed by one band were regarded as not belonging to the same molecular cluster. Isolates with only 1–4 copies of IS6110 were

**Table 1** Epidemiological and risk factor data for 2495 culture positive tuberculosis patients from Greater London, 1 July 1995 to 31 December 1997

Variable	% (n/N)
Age	
0–19	7.1 (178/2495)
20–34	39 (972/2495)
35–59	32.5 (812/2495)
60+	17.6 (437/2495)
Unknown	3.9 (96/2495)
Sex	
Male	57.4 (1433/2495)
Female	40.1 (1001/2495)
Unknown	2.4 (61/2495)
Birth in UK	
UK born	14 (273/1951)
Not UK born	57.3 (1117/1951)
Unknown	28.8 (561/1951)
Ethnic origin	
White	17.9 (348/1947)
ISC	22.1 (431/1947)
Black Caribbean	2.9 (56/1947)
Black African	25.2 (490/1947)
Black other	1.7 (33/1947)
Other	9.4 (182/1947)
Unknown	20.9 (407/1947)
Sputum smear	
Positive	39.5 (770/1951)
Negative	26.5 (516/1951)
Unknown	34.1 (665/1951)
Site of infection lung	64.2 (1196/1864)
Previous treatment	
Yes	7.4 (145/1951)
No	52.3 (1020/1951)
Unknown	40.3 (786/1951)
Immigrant screen	
Yes	3.2 (63/1951)
No	70.7 (1379/1951)
Unknown	26.1 (509/1951)
Contact trace	
Yes	2.7 (53/1951)
No	72.4 (1412/1951)
Unknown	24.9 (486/1951)
Ever homeless?	
Yes	4.4 (110/2495)
No	22.6 (564/2495)
Unknown	73 (1821/2495)
Residential institution	
Yes	0.8 (16/1951)
No	5.6 (110/1951)
Unknown	93.5 (1825/1951)
Prison	
Yes	1.2 (23/1951)
No	25.9 (505/1951)
Unknown	72.9 (1423/1951)
Alcohol abuse	
Yes	6.5 (126/1951)
No	40.9 (797/1951)
Unknown	52.7 (1028/1951)
Intravenous drug use	
Yes	1.0 (20/1951)
No	33.8 (659/1951)
Unknown	65.2 (1272/1951)
Other substance abuse	
Yes	1.3 (26/1951)
No	27 (526/1951)
Unknown	71.7 (1399/1951)
Immunosuppressed	
Yes	7.9 (155/1951)
No	24.2 (472/1951)
Unknown	67.9 (1324/1951)
Retroviral	
Yes	5.8 (114/1951)
No	3.5 (69/1951)
Unknown	90.6 (1768/1951)
Multidrug resistant tuberculosis	
Yes	1.9 (47/2495)
No	98.1 (2448/2495)

ISC=Indian subcontinent.

**Table 1 Continued** Epidemiological and risk factor data for 2495 culture positive tuberculosis patients from Greater London, 1 July 1995 to 31 December 1997

Variable	% (n/N)
Antibiotic	% resistant
Rifampicin	2 (49/2479)
Isoniazid	7.6 (189/2479)
Ethambutol	0.5 (13/2478)
Pyrazinamide	1.1 (26/2468)
Streptomycin	5.8 (144/2475)
Rifabutin	40 (6/15)
Ciprofloxacin	1.5 (3/206)
Clofazimine	56.2 (9/17)
Clarithromycin	18.1 (4/22)
Amikacin	0
Ethanalamine	12.8 (11/86)
Capreomycin	4.6 (3/65)
PAS	0
Cycloserine	9.5 (6/63)

ISC=Indian subcontinent.

considered unevaluable and were excluded from the analysis. These isolates were submitted to secondary typing analysis and the data will be presented elsewhere.

Isolates were also excluded from the study if they were the result of a laboratory contamination event, fulfilling all of the following criteria: (1) only one culture was obtained from the patient, (2) there was no evidence of smear positivity on any sample, (3) the date of specimen processing in the source laboratory was within five working days of an indistinguishable isolate from another patient, and (4) the clinical course (in retrospect) was not consistent with tuberculosis infection.

The percentage clustering was calculated by subtracting the number of clusters from the number of isolates in the clusters and dividing by the total number of isolates. This procedure avoids double counting of clustered isolates by reducing the size of each cluster by one.<sup>8</sup>

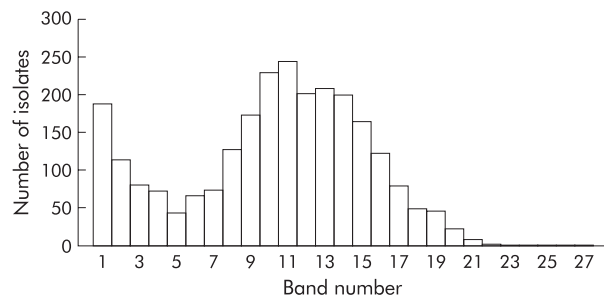
### Epidemiological data collection

A structured proforma was used during record review to collect information about cases of culture confirmed tuberculosis. Over 150 clinicians and clinic staff at more than 50 hospitals (or clinics) assisted with data collection unaware of cluster status. Two health authorities (in east and south east London) were able to provide data that were stored in local databases. Additional microbiological data from the Mycobacterial Reference Unit and the Royal Brompton Hospital were obtained from the PHLS UK antimicrobial resistance surveillance network (Mycobnet) database. Ethnic groups were those used in the 1991 census. Potential links between members of the clusters were determined by reference to these proformas after molecular cluster results were known.

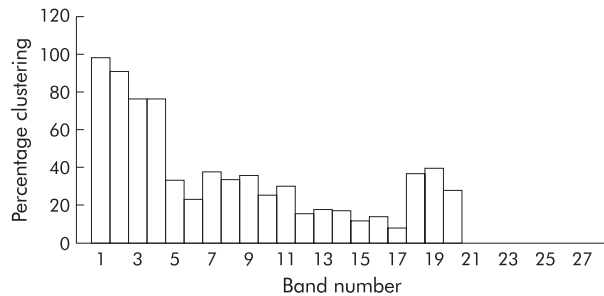
### Cluster analysis

Established links between members of a cluster were considered to exist if members were known to each other, were named contacts or family members, or had close social or institutional contact. Possible links were judged to exist if two or more members lived in the same or an adjacent four digit postcode area and either (a) one member had a risk factor (HIV, alcohol misuse, homelessness) or (b) two members had the same ethnic group or country of birth.

Risk factors for "clustering" were examined initially in a univariate analysis. Two models were then constructed in a logistic regression analysis with membership of a cluster as the outcome variable. Both included demographic variables (age, sex, ethnic group); one also included epidemiological variables and the other microbiological variables. A further



**Figure 1** Frequency distribution of number of copies of IS6110 for 2490 clinical isolates of *Mycobacterium tuberculosis*.



**Figure 2** Percentage of clustered isolates against IS6110 copy number demonstrating the poor differentiation of strains with  $\leq 4$  copies.

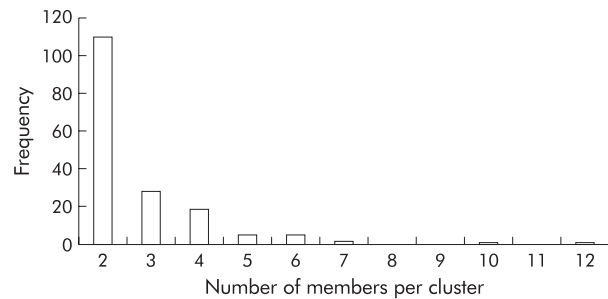
model, which will be discussed further in this paper, was constructed using the demographic variables and those which were significant ( $p=0.05$ ) from either the epidemiological or the microbiological models. Epidemiological analyses were carried out using Stata Statistical Software, Release 5.0, 1997 (StataCorp, College Station, Texas, USA).

## RESULTS

Epidemiological and risk factor data were collected on 2495 patients (table 1). RFLP patterns were obtained for 2779 isolates of *M. tuberculosis*. After eliminating multiple isolates from the same patient, 2500 individual patients were included in the study comprising 77% of the total number of isolates from individuals in London ( $n=3260$ ) reported as culture confirmed cases in the relevant time frame to Mycobnet. Ten further isolates were excluded as showing evidence of originating through cross contamination, making a final study population of 2490 individuals. Missed isolates were not systematically different in respect of age, sex, ethnic group, or source hospital.

There was a bimodal distribution of IS6110 copy number (fig 1): 448 isolates (17.9%) had 1–4 copies and 2042 isolates had five or more copies (modal value=11). The low copy number isolates are generally regarded as not fully evaluable with IS6110 due to the low degree of discrimination shown, although there is no consistent definition of the cut off. In this study, analysis of clustering by band number showed that the isolates with five copies were well differentiated (fig 2) and we therefore included these in the multicopy group.

Of the 2042 isolates with more than four copies of IS6110 there were 463 individuals (22.7%) whose isolates fell in 169 clusters which ranged in size from 2 to 12 persons (fig 3). Most of these clusters were small: 110 (65%) contained only two members and only 14 (8.3%) had more than four members. Assuming that clustering represents recent transmission, the estimated rate of active transmission of tuberculosis in London was 14.4%.



**Figure 3** Frequency distribution of cluster size for isolates with  $\geq 5$  copies of IS6110.

## Description of clusters

The largest cluster (designated cluster 3908) involved 12 people; eight were known to have been born in Somalia, all of whom had arrived in the UK since 1990 (five during or since 1995). All 12 members lived in widely separated areas of London (table 2) in seven different districts and they were diagnosed at 11 different hospitals. There were no known connections between them apart from their ethnic group. Three had pulmonary disease and one was known to have been smear positive. Seven had organisms that were sensitive to all the main antimicrobial agents (pyrazinamide, isoniazid, rifampicin, ethambutol, streptomycin); four had streptomycin resistance and the second case with a specimen date in September 1995 who had arrived from Somalia earlier that year had multidrug resistant (MDR) tuberculosis (resistant to isoniazid and rifampicin).

The next largest cluster involved 10 patients (cluster 2066). Members were of mixed ethnic group (four white, one black African, one black Caribbean, and one Chinese). Three were known to have been born outside the UK (one China, one Ireland, one Nigeria), arriving in 1983 from Ireland and in 1992 from Nigeria. Members mainly lived in different parts of London (seven districts, diagnosed at eight different hospitals; table 2), although two were in adjacent four digit postcode areas. Eight had pulmonary disease and six were sputum smear or culture positive. No connections were found between them except that three were alcohol dependent.

There were two clusters involving seven people. The first (cluster 905) involved seven men aged 29–64 years; four of five whose ethnicity was known were white and the other was black Caribbean, and four of these five were UK born including the black Caribbean. The country of birth for the remainder of this cluster was unknown. Four lived in south east London, two of whom were in adjacent four digit postcode areas. Three were alcohol dependent including one with HIV infection and other substance misuse who was non-compliant with treatment, and a second patient who also misused other recreational drugs. An additional 42 year old white patient was homeless. The second cluster of seven (cluster 4350) comprised people of mixed ethnicity of whom four were known to have been born in the UK.

One cluster of six people (cluster 1171) included a husband (with retroviral infection noted to be poorly compliant with treatment) and wife who were both black Caribbean with MDR tuberculosis. Four of the six had adjacent postcodes. The other cluster of six (cluster 4143) involved five white men and one white woman aged 33–66 years (three alcoholic and one with retroviral infection). Three were contacts and poor compliance with treatment was noted.

## Transmission links

For 14 clusters (8.3%), including clusters 1171 and 4143 (table 2), transmission links were established. In a further 27 clusters (15.9%) links were possible (including clusters 905, 2066, and 3908; table 2), but 128 clusters (79.8%), including

**Table 2** Composition and characteristics of the large clusters

Cluster identity	No of members (men)	Age range	No. of diagnosing hospitals	No. of health authority districts	Known ethnic groups	Country of birth	Antimicrobial resistance	Sputum smear positive	Pulmonary disease	Immunosuppressive conditions	Known risk factors	Transmission category
3908	12 (7)	15–76	11	7	8 BA, 1 ISC	8 Somalia, 1 India	1 MDR, 4 STR	3	3	1 renal disease, 1 diabetes, 1 Down's syndrome	1 refugee, 1 homeless hostel	B
2066	10 (7)	33–90	8	7	4W, 1 BC, 1 BA, 1 C	1 UK, 1 China, 1 Nigeria, 1 Ireland	All sensitive	6	8	1 lung cancer, 1 diabetes, 1 sarcoidosis (treated with steroids)	3 alcoholic (inc 1 prison, 1 homeless), 2 difficult to treat	B
905	7 (7)	29–64	5	3	4W, 1 BC	4 UK	All sensitive	3	5	1 diabetes, 1 HIV +ve	3 alcoholic (inc 1 HIV +ve), 1 homeless, 2 substance abuse, 1 non-compliant	B
4350	7 (5)	22–62	5	3	2W, 1 BA, 1 BC, 1 ISC	4 UK, 1 Pakistan, 1 Nigeria	All sensitive	3	5	2 diabetes		C
1171	6 (3)	25–72	4	1	2 W, 2 BC	4 UK	1 Rif, INH, PZA, STR; 1 Rif, INH, PZA, ETH	2	4	1 non-Hodgkin's lymphoma	1 HIV +ve	A
4143	6 (5)	33–66	4	4	6 W	2 UK, 1 Ireland	All sensitive	2	6		3 alcoholic, 1 HIV +ve	A

BA=black African; BC=black Caribbean; ISC=Indian subcontinent; C=Chinese; W=white; Rif=rifampicin; INH=isoniazid; PZA=pyrazinamide; STR=streptomycin; ETH=ethambutol. Transmission categories: A=confirmed link; B=possible link; C=no link identified. Confirmed links: 1171 (husband and wife and four adjacent postcodes); 1143 (family and social contacts).

cluster 4350, remained unexplained. Transmission routes that were identified included household spread, family links, and close social contact or other close contact within an institution.

### Statistical analysis of clustering

Detailed epidemiological and risk factor information was obtained for 1578 (77%) of the 2042 individuals with a high copy number (>4 IS6110 bands) strains; age, sex, and antimicrobial sensitivity data were derived from Mycobnet for all 2042 cases.

In the multivariate statistical analysis (table 3) young age (0–19 years, OR 2.65, 95% CI 1.59 to 4.44), birth in the UK (OR 1.55, 95% CI 1.04 to 2.03), black Caribbean ethnic group (OR 2.19, 95% CI 1.15 to 4.16), alcohol dependence (OR 2.33, 95% CI 1.46 to 3.72), and streptomycin resistance (OR 1.82, 95% CI 1.15 to 2.88) were independently associated with increased clustering. Factors not independently associated with clustering included HIV infection, MDR tuberculosis, black African or Indian subcontinent ethnicity, and drug misuse.

### DISCUSSION

The overall rate of apparently recent transmission in London from 1995 to 1997 estimated in this study (14.4%) is substantially lower than that reported in other population based studies (21–35%).<sup>6, 8, 18, 19</sup> This comparison needs to be treated with caution because of considerable differences in study design and analysis. We considered the possibility that a longer time period might be necessary to disclose the full extent of clustering. However, analysis of our data by time periods showed that the rate of clustering had reached 11% after only 1 year and thereafter increased only slightly, suggesting that a longer time period would have made very little difference. On the other hand, the maximum cluster size did continue to increase throughout the study, as would be expected, which may contribute to the difference between this and other studies where larger cluster sizes have been reported. However, it seems unlikely, even in a much longer study, that recruitment to clusters would achieve the levels reported in the Netherlands and Denmark as these relate to specific and known chains of transmission.<sup>9, 20</sup>

In this study we have excluded isolates with few copies of IS6110, although in some comparable studies these isolates were subjected to secondary typing—for example, by PGRS typing.<sup>21, 22</sup> There has been no consensus cut off definition for low copy strains, so that isolates with fewer than six IS copies<sup>23</sup> or fewer than five IS copies<sup>9</sup> and identical by both IS6110 and PGRS typing are also counted as clustered. We have chosen to consider these low copy isolates separately as there is at present no certainty that the two definitions are equivalent, and merging the data may produce unreliable results. However, it seems highly unlikely that this could contribute significantly to the lower level of clustering found in London in this study.

It therefore seems likely that the relatively low rate of clustering and the absence of really large clusters are a genuine reflection of differences between the epidemiology of tuberculosis in London at the time of this study and that in the other populations. Our results are comparable to those from a previous but smaller study (n=555) in inner London in 1993 which found a rate of clustering of 19%.<sup>24</sup> Recent infection, as indicated by clustering, was found to be more common in our study in people with recognised risk factors such as alcohol misuse, as well as in younger patients. Similar associations have been found in some other studies.<sup>18, 25, 26</sup> HIV infection was not associated with clustering, which is similar to the findings of some previous studies<sup>9, 26, 27</sup> but not others.<sup>8, 18</sup> Previous studies have shown an association of multiple drug resistance with apparent recent transmission.<sup>5</sup> We did not find such an association and these results are similar to those of van Soolingen *et*

**Table 3** Multivariate analysis: final model variables with odds ratios (OR) and 95% confidence intervals (CI)

Variable	No	% clustering	Unadjusted odds ratios (OR)		Adjusted odds ratios (OR)	
			OR	95% CI	OR	95% CI
<b>Age</b>						
0–19	150	33.3	2.49	1.59 to 3.89	2.65	1.59 to 4.44
20–34	831	24.2	1.54	1.11 to 2.14	1.51	1.02 to 2.22
35–59	648	23.6	1.59	1.14 to 2.24	1.43	0.97 to 2.11
60+	335	16.7	1.00	Reference	1.00	Reference
<b>Sex</b>						
Male	1169	24.5	1.00	Reference	1.00	Reference
Female	827	21	0.82	0.66 to 1.01	0.88	0.69 to 1.12
Unknown*	50	14	0.50	0.22 to 1.12	0.46	0.10 to 2.15
<b>Ethnic group</b>						
White	274	37.2	1.00	Reference	1.00	Reference
ISC	340	13.2	0.26	0.17 to 0.38	0.34	0.22 to 0.52
BC	48	54.2	1.99	1.07 to 3.70	2.19	1.15 to 4.16
BA	409	33.5	0.85	0.62 to 1.17	1.02	0.68 to 1.53
Black other	24	29.2	0.69	0.28 to 1.73	0.76	0.30 to 1.96
Other	156	19.9	0.42	0.26 to 0.66	0.52	0.31 to 0.86
Unknown*	328	23.8	0.53	0.37 to 0.75	0.62	0.41 to 0.92
<b>Birth in UK</b>						
UK born	221	40.3	2.12	1.56 to 2.88	1.55	1.04 to 2.03
Not-UK born	928	24.1	1.00	Reference	1.00	Reference
Unknown*	431	26.4	1.13	0.87 to 1.47	1.17	0.87 to 1.57
<b>Other factors</b>						
Alcohol dependence	94	48.9	2.78	1.82 to 4.23	2.33	1.46 to 3.72
Streptomycin resistant	115	32.2	1.66	1.11 to 2.49	1.82	1.15 to 2.88

BA=black African; BC=black Caribbean; ISC=Indian subcontinent.  
\*Information not known following record review.

*al.*<sup>9</sup> An association was noted with birth in the UK. We suggest that in those not born in the UK a high proportion will have acquired tuberculosis outside the UK and thus will be part of clusters not seen in this study. A further association with clustering is observed with streptomycin resistance but not with any other antimicrobial agent. We conclude that this observation is confounded by an association between Somali born and both clustering and streptomycin resistance, noted in cluster 3908; country of birth was not a factor in the multivariate analysis.

Our study suggests that recent transmission in London plays a lesser role in the occurrence of tuberculosis than in the other cities studied, and that most tuberculosis in London is due to reactivation of previous infections or importation of infection by recent immigrants. It is notable that the largest cluster identified was composed largely of Somalis, and may represent transmission before arrival in London. On further analysis it is clear that cluster 3908 forms part of a closely related family (80% similarity), of which 73% (36/49) of the isolates of known country of birth are Somali. We conclude that cluster 3908 belongs to a family of isolates common to Somalis in London and, by inference, common in Somalia, although we have not been able to confirm this. There is evidence that newly acquired infection is also common among people with recognised risk factors such as alcohol dependence. The association with the black Caribbean ethnic group independent of country of birth may indicate a population in whom contact tracing is particularly important. The results suggest that tuberculosis control in London would be improved by maximising opportunities for preventive intervention and early diagnosis among immigrants from areas with a high incidence of tuberculosis. Ensuring thorough contact tracing and monitoring of treatment and outcome among all cases of tuberculosis, especially in groups at higher risk of recent infection, remains of paramount importance.

## ACKNOWLEDGEMENTS

We would like especially to thank the following: H Atkinson, P Atkinson, O J Billington, G Bothamley, J Carless, M Chadwick, R Coker, M Connolly, S Crossland, N Crowcroft, G French, H Gaya, Y Gileece, S Handford, J Herbert, P Horby, K Lau, D Lamprecht, H Milburn, L Newport, J O'Sullivan, F Palmer, B Pankhania, M Perkin, A Pozniak, I Roddick, R Shaw, J Watson, M Yates. We also thank consultant microbiologists, chest physicians, TB nurses, consultants in communicable disease control, and other staff right across London. We are indebted to all colleagues who gave time and resources to the study.

## Authors' affiliations

**H Maguire**, Public Health Laboratory Service (PHLS) Communicable Disease Surveillance Centre, London W2 3QR, UK  
**J W Dale**, Molecular Microbiology Group, School of Biomedical & Life Sciences, University of Surrey, Guildford, Surrey GU2 7XH, UK  
**T D McHugh**, **S H Gillespie**, **A Dickens**, Department of Medical Microbiology, Royal Free & University College Medical School, Royal Free Campus, London NW3 2PF, UK  
**P D Butcher**, **A Costetsos**, **H Al-Ghusein**, **R Holland**, **D K Banerjee**, Department of Medical Microbiology, St George's Hospital Medical School, London SW17 0RE, UK  
**L Marston**, **D Strachan**, Public Health Sciences, St George's Hospital Medical School, London SW17 0RE, UK  
**P Wilson**, Microbiology Department, The Royal London Hospital, London E1 1BB, UK  
**R Pitman**, Respiratory Division, PHLS Communicable Disease Surveillance Centre, London NW9 5EQ, UK  
**F A Drobniewski**, PHLS Mycobacterial Reference Unit, Public Health Laboratory and Medical Microbiology, Guys, Kings & St Thomas' School of Medicine, London SE22 8QF, UK

DKB is Chairman and Project Leader of the Molecular Epidemiology of Tuberculosis in London Steering Committee and responsible for overall coordination of the project. The project was conceived of and instigated by DKB, PDB, JWD, SHG, HM, TDMcH, and PW. All authors contributed to the study design and interpretation of results. PDB, JWD, FAD, SHG, and TDMcH were responsible for the design and supervision of molecular microbiological methods. HA-G, AD, and TDMcH performed the typing methods. AD and JWD undertook the typing data entry and analysis. Epidemiological methodology was developed by HM. AC and RH were responsible for collation of the epidemiological data, supervised by HM,

JWD, and DS. Statistical analysis and interpretation was performed by HM, LM, RP, and DS. Linkage of datasets was performed by RP and JWD. HM produced the first draft of this paper and revised drafts with TDMcH and JWD with feedback from all authors. All authors have seen and approved the final version for submission. DKB is guarantor.

Funding: This work was undertaken by DKB who received funding from the NHS Executive London, Research & Development Programme. The views expressed in the publication are those of the authors and are not necessarily those of the NHS Executive or the Department of Health. The authors are also grateful for support from the European Union under grants BMH4-CT97-91202 and SMT4-CT96-2097 (provision of GelCompar software).

## REFERENCES

- 1 **van Embden JDA**, Cave MD, Crawford JT, *et al*. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. *J Clin Microbiol* 1993;**31**:406–9.
- 2 **McHugh TD**, Dickens A, Gillespie SH. False molecular clusters due to non-random association of IS6110 with Mycobacterium tuberculosis. *J Clin Microbiol* 2000;**38**:2081–6.
- 3 **Fang ZG**, Forbes KJ. A Mycobacterium tuberculosis IS6110 preferential locus (ipl) for insertion into the genome. *J Clin Microbiol* 1997;**35**:479–81.
- 4 **Diaz R**, Kremer K, de Haas PE, *et al*. Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994 to June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. *Int J Tuberc Lung Dis* 1998;**2**:743–50.
- 5 **Frieden TR**, Woodley CL, Crawford JT, *et al*. The molecular epidemiology of tuberculosis in New York City: the importance of nosocomial transmission and laboratory error. *Tuberc Lung Dis* 1996;**77**:407–13.
- 6 **Gutiérrez C**, Torrea G, Aubert D, *et al*. Molecular strain typing for the evaluation of tuberculosis transmission in the Paris region. *Tuberc Lung Dis* 1996;**77**(Suppl 2):116.
- 7 **Gillespie SH**, Kennedy N, Ngowi FI, *et al*. Restriction fragment length polymorphism analysis of Mycobacterium tuberculosis isolated from patients with pulmonary tuberculosis in northern Tanzania. *Trans R Soc Trop Med Hyg* 1995;**89**:335–8.
- 8 **Small PM**, Hopewell PC, Singh SP, *et al*. The epidemiology of tuberculosis in San Francisco. *N Engl J Med* 1994;**330**:1703–9.
- 9 **Van Soolingen D**, Borgdorff MW, De Haas PEW, *et al*. Molecular epidemiology of tuberculosis in the Netherlands: a nationwide study from 1993 through 1997. *J Infect Dis* 1999;**180**:726–36.
- 10 **Yang ZH**, de Haas PE, Wachmann CH, *et al*. Molecular epidemiology of tuberculosis in Denmark in 1992. *J Clin Microbiol* 1995;**33**:2077–81.
- 11 **Atkinson P**, Maguire H, Duckworth G. *CDSC London Annual Report on communicable disease in London 1999*. London: Public Health Laboratory Service, 2000.
- 12 **Spence DPS**, Hotchkiss J, Williams CSD, *et al*. Tuberculosis and poverty. *BMJ* 1993;**307**:7759–61.
- 13 **Mangtani P**, Jolley DJ, Watson JM, *et al*. Socioeconomic deprivation and notification rates for tuberculosis in London. *BMJ* 1995;**310**:963–6.
- 14 **Doherty MJ**, Davies PDO, Ballis MA, *et al*. Tuberculosis in England and Wales. Ethnic origin is more important than social deprivation. *BMJ* 1995;**311**:187.
- 15 **Foley NM**, Miller RF. Tuberculosis and AIDS: is the White Plague up and coming? *J Infect* 1993;**26**:39–43.
- 16 **Marshall BG**, Mitchell DM, Shaw RJ, *et al*. HIV and tuberculosis co-infection in an inner London hospital: a prospective anonymised seroprevalence study. *J Infect* 1999;**38**:162–6.
- 17 **Bowen EF**, Rice PS, Cooke NT, *et al*. HIV seroprevalence by anonymous testing in patients with Mycobacterium tuberculosis and in tuberculosis contacts. *Lancet* 2000;**356**:1488–9.
- 18 **Alland D**, Kalkut GE, Moss AR, *et al*. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med* 1994;**330**:1710–6.
- 19 **Van Deutekom H**, Gerritsen JJJ, Van Soolingen D, *et al*. Molecular epidemiological approach to studying the transmission of tuberculosis in Amsterdam. *Clin Infect Dis* 1997;**25**:1071–7.
- 20 **Bauer J**, Yang ZH, Poulsen S, *et al*. Results from 5 years of nationwide DNA fingerprinting of Mycobacterium tuberculosis complex isolates in a country with a low incidence of M tuberculosis infection. *J Clin Microbiol* 1998;**36**:305–8.
- 21 **Ross BC**, Raios K, Jackson K, *et al*. Molecular cloning of a highly repeated DNA element from Mycobacterium tuberculosis and its use as an epidemiological tool. *J Clin Microbiol* 1992;**30**:942–6.
- 22 **Yang ZH**, Chaves F, Barnes PF, *et al*. Evaluation of method for secondary DNA typing of Mycobacterium tuberculosis with pTBN12 in epidemiologic study of tuberculosis. *J Clin Microbiol* 1996;**34**:3044–8.
- 23 **Jasmer RM**, Hahn JA, Small PM, *et al*. Molecular epidemiologic analysis of tuberculosis trends in San Francisco, 1991–1997. *Ann Intern Med* 1999;**130**:971–8.
- 24 **Hayward AC**, Goss S, Drobniowski F, *et al*. Molecular epidemiology of tuberculosis in London using PCR-based typing systems. Third Meeting of the EC Concerted Action on Genetic Markers and TB Epidemiology, Cordoba, June 1997.
- 25 **Gutiérrez MC**, Vincent V, Aubert D, *et al*. Molecular fingerprinting of Mycobacterium tuberculosis and risk factors for tuberculosis transmission in Paris, France, and surrounding area. *J Clin Microbiol* 1998;**36**:486–92.
- 26 **Hayward AC**, Herbert J, Bennett DE, *et al*. A case-control study of risk factors for drug resistance in Mycobacterium tuberculosis isolates in England and Wales: 1993–1994. *Tuberc Lung Dis* 1996;**77**(Suppl 2):62.
- 27 **Samper S**, Martin C, Pinedo A, *et al*. Transmission between HIV-infected patients of multidrug-resistant tuberculosis caused by Mycobacterium bovis. *AIDS* 1997;**11**:1237–42.