

ORIGINAL ARTICLE

Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype

Manish Pareek,^{1,2,3} Jason Evans,⁴ John Innes,⁵ Grace Smith,⁴ Suzie Hingley-Wilson,¹ Kathryn E Loughheed,⁶ Saranya Sridhar,¹ Martin Dedicoat,⁵ Peter Hawkey,^{4,7} Ajit Lalvani¹

► Additional data are published online only. To view these files please visit the journal online (<http://dx.doi.org/10.1136/thoraxjnl-2012-201824>).

¹Tuberculosis Research Unit, National Heart and Lung Institute, Imperial College London, London, UK

²Department of Infectious Disease Epidemiology, Imperial College London, London, UK

³Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, UK

⁴Health Protection Agency Midlands Laboratory, Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, UK

⁵Department of Infection and Tropical Medicine, Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, UK

⁶Department of Integrative Cell Biology, Division of Cell & Molecular Biology, Imperial College London, London, UK

⁷School of Immunity and Infection, The Medical School, University of Birmingham, Edgbaston, Birmingham, UK

Correspondence to

Prof Ajit Lalvani, Tuberculosis Research Unit, National Heart and Lung Institute, Imperial College London, London W21PG, UK; a.lalvani@imperial.ac.uk

Received 27 February 2012

Revised 11 August 2012

Accepted 15 August 2012

Published Online First

27 September 2012

ABSTRACT

Background Emerging evidence suggests that *Mycobacterium tuberculosis* (*Mtb*) lineage and host ethnicity can determine tuberculosis (TB) clinical disease patterns but their relative importance and interaction are unknown.

Methods We evaluated prospectively collected TB surveillance and *Mtb* strain typing data in an ethnically heterogeneous UK population. Lineage assignment was denoted using 15-loci mycobacterial interspersed repetitive units containing variable numbers of tandem repeats (MIRU-VNTR) and MIRU-VNTRplus. Geographical and ethnic associations of the six global *Mtb* lineages were identified and the influence of lineage and demographic factors on clinical phenotype were assessed using multivariate logistic regression.

Results Data were available for 1070 individuals with active TB which was pulmonary only, extrapulmonary only and concurrent pulmonary–extrapulmonary in 52.1%, 36.9% and 11.0% respectively. The most prevalent lineages were Euro-American (43.7%), East African Indian (30.2%), Indo-Oceanic (13.6%) and East Asian (12.2%) and were geo-ethnically restricted with, for example, Indian subcontinent ethnicity inversely associated with Euro-American lineage (OR 0.23; 95% CI 0.14 to 0.39) and positively associated with the East African-Indian lineage (OR 4.04; 95% CI 2.19 to 7.45). Disease phenotype was most strongly associated with ethnicity (OR for extrathoracic disease 21.14 (95% CI 6.08 to 73.48) for Indian subcontinent and 14.05 (3.97 to 49.65) for Afro-Caribbean), after adjusting for lineage. With East Asian lineage as the reference category, the Euro-American (OR 0.54; 95% CI 0.32 to 0.91) and East-African Indian (OR 0.50; 95% CI 0.29 to 0.86) lineages were negatively associated with extrathoracic disease, compared with pulmonary disease, after adjusting for ethnicity.

Conclusions Ethnicity is a powerful determinant of clinical TB phenotype independently of mycobacterial lineage and the role of ethnicity-associated factors in pathogenesis warrants investigation.

INTRODUCTION

Worldwide, *Mycobacterium tuberculosis* (*Mtb*) continues to cause significant morbidity and mortality with an estimated 8.8 million cases and 1.1 million deaths in 2011.¹ Genetic strain typing has identified six major lineages of *Mtb* which epidemiologically coassociate with distinct geographical regions.^{2–5} Moreover, lineages can differ in the

Key messages

What is the key question?

► *Mycobacterium tuberculosis* (*Mtb*) presents as a spectrum of disease but the relative importance of mycobacterial lineage and host ethnicity in determining clinical phenotype is unknown.

What is the bottom line?

► Ethnicity is a powerful determinant of clinical tuberculosis phenotype independently of mycobacterial lineage.

Why read on?

► This large UK-based study provides the first evidence for the predominance of host ethnicity over mycobacterial lineage in determining clinical disease phenotype and highlights the need to investigate ethnicity-associated factors in the pathogenesis of tuberculosis.

host response that they induce⁶ with, for example, the Beijing (East Asian) lineage inducing lower levels of inflammatory cytokines^{7,8} while exhibiting enhanced intracellular growth.⁹

Although these *in vitro* differences in host response might be expected to result in the well recognised different clinical phenotypes observed in distinct demographic groups,^{10–12} the impact and influence of different lineages on clinical patterns of active tuberculosis (TB) still remains unclear. A notable exception is the Beijing (East Asian) lineage which, in certain ethnic groups, has been associated with a more aggressive clinical course¹³ and extrapulmonary forms of TB,^{14–16} although other studies have failed to document these associations.^{17–20} More recently, European studies^{19,21} have suggested that the Central Asian lineage is associated with extrapulmonary disease whilst two Pakistani^{22,23} studies found the opposite. However, studies conducted to date have, with one recent exception,²⁰ mainly examined homogeneous populations where the relative importance of ethnicity, compared with lineage, on disease phenotype was not independently examined.

Given the recognised association of different lineages and ethnic groups,^{2–5} and the known link between extrapulmonary disease and non-white

To cite: Pareek M, Evans J, Innes J, *et al.* *Thorax* 2013;**68**:221–229.

ethnic groups,^{10–12} we reasoned that to determine whether lineages influence clinical patterns of disease it would be necessary to dissect out the relative influence of the lineage itself from ethnicity. We therefore prospectively collated data on host demographic factors, clinical patterns of disease and mycobacterial genotype for culture-confirmed cases of TB in a single-centre UK setting, with a unique ethnic mix, to assess the global phylogeography of the strains in our dataset, identify the demographic factors associated with different lineages and to establish which, if any, factors, including lineage, were associated with distinct clinical patterns of disease.

METHODS

Study design and data source

We analysed data from a routine, anonymised, centralised clinical surveillance database maintained as part of usual care for all patients diagnosed with active TB in two closely related metropolitan areas in the UK (Birmingham and Solihull—combined population approximately 1.2 million;²⁴ TB incidence varies from 9–88 per 100 000 population per year²⁵).

MIRU-VNTR typing

Since mid-2003 all culture-positive *Mtb* isolates received, and identified, as *Mtb* complex by the Health Protection Agency Midlands Regional Centre for Mycobacteriology have been routinely analysed by 15-loci mycobacterial interspersed repetitive units containing variable numbers of tandem repeats (MIRU-VNTR) typing (ETR-A, B, C, D, E and MIRU-02, 10, 16, 20, 23, 24, 26, 27, 39, 40). MIRU-VNTR typing analyses the number of repetitive DNA sequences at multiple independent genetic loci.^{26–27} The values generated are automatically recorded on the clinical surveillance database.

Study population, variables and assignment to lineage

We included all cases of active TB diagnosed between 1 July 2003 and 28 February 2010 with complete information on age, sex, country of birth (local or foreign born), ethnicity, site of disease and MIRU-VNTR value; patients with missing information for one or more of these variables or culture-negative disease were excluded from analysis.

Information on year of notification, age, sex, local/foreign born, country of birth if foreign born, ethnicity, time since entry (only applies to foreign-born immigrants and calculated as time between arrival in the UK and diagnosis of active TB), employment status, HIV status, previous history of TB, site of disease, specific organs involved, duration of symptoms before treatment commencement, drug sensitivity and MIRU-VNTR value were retrieved from the database.

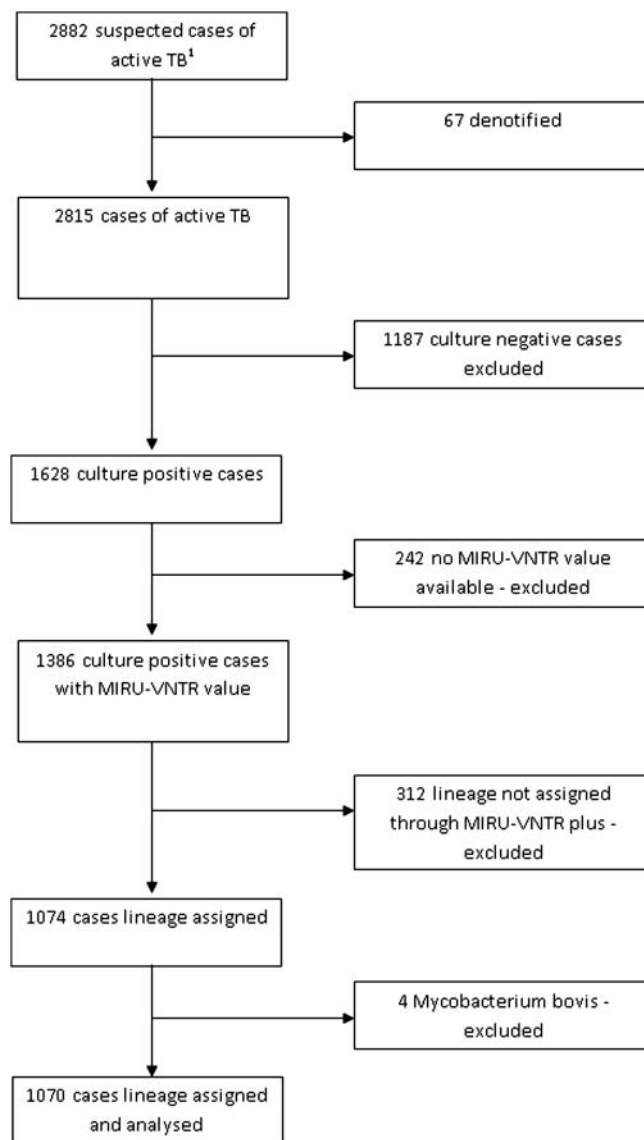
We used the MIRU-VNTRplus database²⁸ (<http://www.miru-vntrplus.org/>) to assign strains with MIRU-VNTR values to one of six global lineages as defined by Gagneux *et al*²: East Asian, Euro-American, East African Indian, Indo-Oceanic, West African-1 and West African-2. Strains which could not be assigned to any lineage and those which were identified as being non *Mtb* (such as *Mycobacterium bovis*) were excluded from analysis.

Ethnicity was defined in line with the Office for National Statistics classification²⁹ (White, Indian subcontinent, Afro-Caribbean, Oriental/Other Asia, Other). We explored the patterns of disease in local-born ethnic groups versus foreign-born ethnic groups by creating a composite variable 'location of birth/ethnicity' with individuals categorised as being local or foreign born non-white ethnicity.

HIV testing of patients diagnosed with active TB is highly inconsistent in the UK and this was reflected by this variable being inconsistently recorded in our dataset. Therefore, we considered patients either as HIV positive (if recorded) or as negative/unknown (for all other individuals).

Definitions of TB phenotypes

We classified all cases of active TB into the following clinical phenotypes: extrathoracic disease only (only extrapulmonary site(s) outside the thoracic cavity); extrapulmonary only (only one or more extrapulmonary site(s) involved but including pleural TB and intrathoracic lymph node TB²⁰); pulmonary only (only lung parenchyma involved); pulmonary and extrapulmonary (pulmonary disease and one or more extrapulmonary sites involved concurrently); and lymph node disease only (intrathoracic or extrathoracic). For comparative analyses we excluded all patients with concurrent pulmonary–extrapulmonary disease, as others^{10–11–30} have done, to minimise misclassification bias as the dominant disease site is unclear.



¹All suspected cases of active TB between 1 July 2003 and 28 February 2010 included

Figure 1 Flow chart of cases included in data analysis.

Data analysis

Details of the data analysis are presented in the online supplementary information.

RESULTS

Description of cohort

The study flowchart is presented in figure 1 and demographic and clinical details of the cohort in table 1. Median time between arrival in the UK and development of active TB was

5 years (IQR 2–21). Individuals with MIRU-VNTR values (n=1070) were not significantly different, in terms of demographic and clinical features, to individuals without MIRU-VNTR values and to individuals with MIRU-VNTR values that could not be assigned to a lineage (table 1).

Distribution, and temporal trends, of lineages in cohort

In our cohort, Euro-American was the commonest lineage (43.7%), followed by East African Indian (30.2%),

Table 1 Demographic characteristics of individuals with active tuberculosis compared with those with no mycobacterial interspersed repetitive units containing variable numbers of tandem repeats (MIRU-VNTR) value and when no mycobacterial lineage could be assigned

Variable	Subjects included in study (n=1070)	Subjects with no MIRU available (n=242)	p Value*	Subjects with no lineage assigned (n=312)	p Value†
Age					
Median (IQR)	34.8 (25.0–49.7)	36.2 (24.0–49.6)	0.11	32 (25.3–52.2)	0.42
Age categories					
Under 16	36 (3.4%)	12 (5.0%)		18 (5.8%)	
16–25	273 (25.5%)	54 (22.3%)		66 (21.2%)	
26–35	261 (24.4%)	53 (21.9%)		98 (31.4%)	
36–45	194 (18.1%)	36 (14.9%)		34 (10.9%)	
46–55	96 (9.0%)	21 (8.7%)		32 (10.3%)	
over 55	210 (19.6%)	66 (27.3%)		64 (20.5%)	
Gender					
Men	579 (54.1%)	145 (59.9%)	0.12	159 (51.0%)	0.33
Ethnicity					
White	148 (13.8%)	30 (12.4%)	0.50	24 (7.7%)	0.024
Indian subcontinent	577 (53.9%)	125 (51.7%)		185 (59.3%)	
Afro-Caribbean	278 (26.0%)	72 (29.8%)		89 (28.5%)	
Oriental/other Asia	23 (2.2%)	8 (3.3%)		3 (1.0%)	
Other	44 (4.1%)	7 (2.9%)		11 (3.5%)	
Place of birth					
Foreign born	752 (70.3%)	181 (74.8%)	0.16	248 (79.5%)	0.001‡
Time since entry to the UK (years)					
Median (IQR)	5 (2–21)§	6 (3–20)	0.54	6 (3–18)	0.76
<1	24 (3.4%)	8 (4.9%)		11 (4.8%)	
1–5	338 (48.0%)	72 (44.2%)		95 (41.1%)	
6–10	95 (13.5%)	24 (14.7%)		47 (20.4%)	
>10	247 (35.1%)	59 (36.2%)		78 (33.8%)	
HIV status					
Negative	70 (84.3%)¶	13 (86.7%)	0.82	23 (95.8%)	0.14
Previous TB					
No	1030 (96.3%)	228 (94.2%)	0.14	299 (95.8%)	0.72
Type of disease					
Pulmonary only	557 (52.1%)	115 (47.5%)	0.23	155 (49.7%)	0.48
Pulmonary and extrapulmonary	118 (11.0%)	37 (15.3%)	0.08	40 (12.8%)	0.42
Extrapulmonary only	395 (36.9%)	90 (37.2%)	0.94	117 (37.5%)	0.89
Extrathoracic only	318 (29.7%)	73 (30.2)	0.88	96 (30.8)	0.73
Lymph node only	204 (19.1%)	44 (18.2)	0.79	63 (20.2)	0.68
Smear result					
Positive	545 (50.9%)	108 (44.6%)		150 (48.1%)	0.40
Drug sensitivity**					
Fully sensitive	976 (95.2%)	117 (88.0%)	0.007	288 (96.6%)	0.43
Mono-drug resistant	37 (3.6%)	12 (9.0%)		9 (3.0%)	
Multi-drug resistant	12 (1.2%)	4 (3.0%)		1 (0.3%)	

*p Value for subjects with no MIRU-VNTR available versus subjects included in study; continuous variables compared using the Mann–Whitney U test and categorical variables (percentages/proportions) using Fisher's exact test.

†p Value for subjects with no lineage assigned versus subjects included in study; continuous variables compared using the Mann–Whitney U test and categorical variables (percentages/proportions) using Fisher's exact test.

‡Subjects with no lineage assigned (compared with those included in the study) were significantly more likely to be foreign-born individuals (p=0.001).

§Data available for 704 individuals.

¶Data available for 83 individuals.

**Data available for 1025 individuals.

Table 2 Distribution of *Mycobacterium tuberculosis* lineages by different world regions of birth*

World region of birth	Lineage				
	East Asian (n=130)	Euro-American† (n=459)	East African Indian‡ (n=321)	Indo-Oceanic§ (n=144)	West African 1/2 (n=4)
Americas and Caribbean	0 (0.0)	14 (87.5)	2 (12.5)	0 (0.0)	0 (0.0)
Europe	34 (10.2)	209 (62.8)	74 (22.2)	15 (4.5)	1 (0.3)
Middle East/North Africa	0 (0.0)	11 (84.6)	1 (7.69)	1 (7.7)	0 (0.0)
South Africa	4 (8.7)	36 (78.3)	0 (0.0)	6 (13.0)	0 (0.0)
West and Central Africa	2 (11.8)	12 (70.6)	0 (0.0)	2 (11.8)	1 (5.9)
East Africa	25 (16.8)	74 (49.7)	16 (10.7)	32 (21.5)	2 (1.3)
Indian subcontinent and Central Asia	59 (12.6)	102 (21.7)	227 (48.3)	82 (17.4)	0 (0.0)
Southeast and East Asia	6 (42.9)	1 (7.1)	1 (7.1)	6 (42.9)	0 (0.0)

*Data presented are numbers and percentages of each lineage in each world region of birth (ie, the data should be read across).

†Data missing for nine individuals.

‡Data missing for two individuals.

§Data missing for one individual.

Indo-Oceanic (13.6%) and Beijing (12.2%). West African-1 (0.0%) and West African-2 (0.7%) lineages were rarely found in this cohort. The proportion of cases caused by any specific lineage in any given year did not change significantly over the course of the study period (see online supplementary figure S1).

Distribution of lineages by geographical origin and ethnic group

Table 2 and online supplementary figure S2 present the global distribution of lineages stratified by region, and country, of origin respectively. In general, lineages were geographically restricted and associated with specific regions. For example, the Euro-American lineage was predominantly found in Europe, the Americas and Africa, which together accounted for 74.7% of all cases caused by this lineage. East African Indian and Indo-Oceanic lineages were mainly found in the Indian subcontinent and East Africa (75.7%) and Indian subcontinent, East Africa and Asia (83.3%) respectively. By contrast, the East Asian strain appeared more geographically widespread, being found in the Indian subcontinent and East Africa and accounted for 55.6% of cases in the small number of individuals (nine) from East Asia.

Stratification of lineages by ethnic group (see online supplementary table S1) illustrated that, within specific regions, individuals of certain ethnicity were preferentially associated with particular lineages. Individuals of White (OR 4.56; 95% CI 3.08 to 6.75) and Afro-Caribbean (OR 3.21; 95% CI 2.42 to 4.28) ethnicity were significantly more likely to present with disease caused by the Euro-American lineage. By contrast, individuals in the Indian subcontinent group were significantly associated with the East African Indian (OR 5.89; 95% CI 4.30 to 8.06) and Indo-Oceanic lineages (OR 1.69; 95% CI 1.17 to 2.43) whereas the Oriental/Other Asia group was associated with the East Asian lineage (OR 3.29; 95% CI 1.33 to 8.15).

Differences in lineage were also evident when comparing UK-born and foreign-born individuals of non-white ethnicity (see online supplementary table S1). The Euro-American lineage was significantly less common among foreign-born (compared with the UK-born) individuals of Indian subcontinent (OR 0.38; 95% CI 0.25 to 0.58) and Afro-Caribbean (OR 0.40; 95% CI 0.20 to 0.82) ethnicity.

Association of time since arrival and development of active TB for different lineages

Online supplementary figure S3 illustrates, for the foreign-born cohorts, the time between arrival in the UK and the development of active TB for the lineages in our cohort. Patients infected with the East African Indian lineage presented later (median 7 years, IQR 3–29) than the Euro-American lineage (median 5 years, IQR 2–18) ($p=0.001$).

Demographic factors associated with mycobacterial lineage

Table 3 depicts results of the univariate and multivariate analysis of the demographic and clinical factors associated with the four main lineages in our cohort (East Asian, Euro-American, East African Indian and Indo-Oceanic). On multivariate analysis, foreign birth and, in comparison to White ethnicity, certain non-white ethnic groups were inversely associated with Euro-American lineage (table 3). By contrast, Afro-Caribbean individuals were equally likely as White individuals to harbour the Euro-American lineage. Certain ethnic groups were either associated (Indian subcontinent) or inversely associated (Afro-Caribbean) with the East African Indian lineage. Foreign-birth and non-white ethnicity were significantly associated with the Indo-Oceanic lineage (table 3).

Factors associated with disease patterns on univariate analysis

Table 4 illustrates the clinical phenotypes of active TB (data for pulmonary only and extrathoracic disease only presented in table 4) stratified by lineage and ethnic group. In general, extrapulmonary only, extrathoracic only and lymph node only disease were more frequent in individuals who were foreign born ($p<0.0001$ for all three) and of non-white ethnicity ($p<0.0001$ for all three). UK-born individuals of White ethnicity had a significantly lower proportion of extrapulmonary only (7.7%), extrathoracic only (2.4%) and lymph node only (0.8%) disease compared with individuals of UK-born non-white ethnicity (31.9%, 25.9% and 20.4% respectively) ($p<0.0001$ for all three) and both these groups had significantly lower proportions of all three extrapulmonary phenotypes compared with individuals of foreign-born non-white ethnicity (51.1%, 46.5% and 35.4%; $p<0.0001$ for UK-born White and UK-born non-white).

Table 3 Univariate and multivariate association of demographic, clinical and temporal factors with different *Mycobacterium tuberculosis* lineages

Variable	East Asian (n=130)	Adjusted OR† (95% CI)	Euro-American (n=468)	Adjusted OR† (95% CI)	East African Indian (n=323)	Adjusted OR‡ (95% CI)	Indo-Oceanic (n=145)	Adjusted OR‡ (95% CI)
Period								
2003–2004	25/227 (11)	1‡	113/227 (49.8)	1	64/227 (28.2)	1‡	24/227 (10.6)	1
2005–2006	48/342 (14)	1.27 (0.75 to 2.14)	133/342 (38.9)	0.65 (0.44 to 0.96)	104/342 (30.4)	1.08 (0.72 to 1.63)	55/342 (16.1)	1.47 (0.87 to 2.50)
2007–2008	35/336 (10.4)	0.89 (0.51 to 1.57)	153/336 (45.5)	0.86 (0.58 to 1.26)	105/336 (31.3)	1.18 (0.79 to 1.78)	42/336 (12.5)	1.14 (0.65 to 2.00)
2009–2010	22/165 (13.3)	1.32 (0.69 to 2.50)	69/165 (41.8)	0.71 (0.44 to 1.15)	50/165 (30.3)	1.03 (0.63 to 1.70)	24/165 (14.5)	1.46 (0.76 to 2.78)
Age categories n (%)								
Under 16	4/36 (11.1)	1	19/36 (52.8)	1	9/36 (25)	1	4/36 (11.1)	1
16–25	32/273 (11.7)	0.88 (0.28 to 2.74)	128/273 (46.9)	1.10 (0.50 to 2.44)	75/273 (27.5)	0.90 (0.37 to 2.20)	35/273 (12.8)	0.95 (0.30 to 2.96)
26–35	31/261 (11.9)	0.86 (0.27 to 2.75)	104/261 (39.8)	1.01 (0.45 to 2.29)	82/261 (31.4)	0.90 (0.36 to 2.24)	43/261 (16.5)	1.09 (0.35 to 3.46)
36–45	31/194 (16)	1.25 (0.39 to 4.03)	87/194 (44.8)	0.88 (0.38 to 2.04)	46/194 (23.7)	0.75 (0.29 to 1.93)	30/194 (15.5)	1.28 (0.39 to 4.13)
46–55	11/96 (11.5)	0.73 (0.20 to 2.69)	40/96 (41.7)	1.19 (0.47 to 3.01)	29/96 (30.2)	0.83 (0.30 to 2.28)	16/96 (16.7)	1.11 (0.32 to 3.87)
Over 55	21/210 (10)	0.71 (0.21 to 2.42)	90/210 (42.9)	1.50 (0.63 to 3.56)	82/210 (39)	0.99 (0.38 to 2.54)	17/210 (8.1)	0.53 (0.15 to 1.82)
Gender								
Men	71/579 (12.3)	1	263/579 (45.4)	1	162/579 (28.0)	1	81/579 (14.0)	1
Women	59/491 (12.0)	0.94 (0.64 to 1.4)	205/491 (41.8)	1.14 (0.85 to 1.52)	161/491 (32.8)	0.99 (0.73 to 1.33)	64/491 (13.0)	0.85 (0.58 to 1.24)
Place of birth								
UK	30/318 (9.4)	1	200/318 (62.9)	1*	72/318 (22.6)	1	15/318 (4.7)	1***
Foreign born	100/752 (13.3)	1.52 (0.84 to 2.75)	268/752 (35.6)	0.40 (0.27 to 0.60)	251/752 (33.4)	1.34 (0.87 to 2.08)	130/752 (17.3)	2.44 (1.31 to 4.53)
Ethnicity								
White	16/148 (10.8)	1	110/148 (74.3)	1*	21/148 (14.2)	1*	1/148 (0.7)	1****
Indian subcontinent	69/577 (12)	0.79 (0.36 to 1.71)	151/577 (26.2)	0.23 (0.14 to 0.39)	262/577 (45.4)	4.04 (2.19 to 7.45)	94/577 (16.3)	14.02 (1.81 to 108.52)
Afro-Caribbean	35/278 (12.6)	0.77 (0.33 to 1.81)	180/278 (64.7)	1.54 (0.85 to 2.82)	21/278 (7.6)	0.38 (0.18 to 0.84)	39/278 (14)	9.21 (1.15 to 73.58)
Oriental/other Asia	7/23 (30.4)	2.63 (0.78 to 8.84)	4/23 (17.4)	0.12 (0.03 to 0.47)	6/23 (26.1)	1.78 (0.56 to 5.62)	6/23 (26.1)	21.99 (2.28 to 212.42)
Other	3/44 (6.8)	0.43 (0.10 to 1.75)	23/44 (52.3)	0.98 (0.43 to 2.23)	13/44 (29.5)	1.84 (0.73 to 4.66)	5/44 (11.4)	6.38 (0.64 to 63.52)
HIV status								
Negative/unknown	128/1057 (12.1)	1	460/1057 (43.5)	1	322/1057 (30.5)	1	143/1057 (13.5)	1
Positive	2/13 (15.4)	1.21 (0.25 to 5.91)	8/13 (61.5)	1.06 (0.31 to 3.61)	1/13 (7.7)	0.78 (0.09 to 6.90)	2/13 (15.4)	0.97 (0.20 to 4.69)
Previous TB								
No	130/1030 (12.6)	1	447/1030 (43.4)	1	306/1030 (29.7)	1**	143/1030 (13.9)	1
Yes	0/40 (0.0)	1.35 (0.66 to 2.76)	21/40 (52.5)	1.35 (0.66 to 2.76)	17/40 (42.5)	2.07 (1.00 to 4.31)	2/40 (5.0)	0.37 (0.09 to 1.58)
Drug sensitivity								
Fully sensitive	119/976 (12.2)	1	431/976 (44.2)	1	293/976 (30.0)	1	130/976 (13.3)	1
Drug resistant§	7/49 (14.3)	1.21 (0.52 to 2.83)	15/49 (30.6)	0.65 (0.33 to 1.29)	17/49 (34.7)	1.07 (0.55 to 2.07)	9/49 (18.4)	1.27 (0.58 to 2.78)
Duration of symptoms before treatment								
0–15 days	5/59 (8.5)	1	29/59 (49.2)	1	18/59 (30.5)	1	7/59 (11.9)	1
16–30 days	6/56 (10.7)	1.10 (0.32 to 3.61)	25/56 (44.6)	1.10 (0.32 to 3.61)	17/56 (30.4)	1.10 (0.32 to 3.61)	8/56 (14.3)	1.10 (0.32 to 3.61)
31–45 days	7/61 (11.5)	1.10 (0.32 to 3.61)	25/61 (41)	1.10 (0.32 to 3.61)	21/61 (34.4)	1.10 (0.32 to 3.61)	8/61 (13.1)	1.10 (0.32 to 3.61)
46–60 days	5/43 (11.6)	1.10 (0.32 to 3.61)	20/43 (46.5)	1.10 (0.32 to 3.61)	10/43 (23.3)	1.10 (0.32 to 3.61)	8/43 (18.6)	1.10 (0.32 to 3.61)
61 days and over	31/243 (12.8)	1.10 (0.32 to 3.61)	99/243 (40.7)	1.10 (0.32 to 3.61)	83/243 (34.2)	1.10 (0.32 to 3.61)	30/243 (12.3)	1.10 (0.32 to 3.61)
Smear positive								
No	74/525 (14.1)	1	211/525 (40.2)	1	162/525 (30.9)	1	76/525 (14.5)	1
Yes	56/545 (10.3)	0.73 (0.49 to 1.08)	257/545 (47.2)	0.98 (0.74 to 1.31)	161/545 (29.5)	1.18 (0.88 to 1.59)	69/545 (12.7)	1.04 (0.71 to 1.53)

*p<0.0001; **p=0.05; ***p=0.005; ****p=0.015.

†Model mutually adjusted for year of notification, age, gender, ethnicity, country of birth, HIV status, drug sensitivity and smear positivity; data analysed using logistic regression.

‡Model mutually adjusted for year of notification, age, gender, ethnicity, country of birth, HIV status, previous TB, drug sensitivity and smear positivity; data analysed using logistic regression.

§Drug-resistant category includes individuals with mono-resistant TB and multi-drug resistant TB.

TB, tuberculosis.

Table 4 Different clinical phenotypes of active tuberculosis stratified by lineage and demographic group

Mycobacterial lineage	Clinical phenotype and demographic group								p Value§¶
	Pulmonary only (n=545)*				Extrathoracic only (n=317)*				
	UK born white	UK born non-white	Foreign born non-white	Total†¶	UK born white	UK born non-white	Foreign born non-white	Total†¶	
East Asian	12 (100.0)	5 (62.5)	34 (43.0)	51	0 (0.0)	3 (37.5)	45 (57.0)	48	<0.0001
Euro-American	88 (97.8)	62 (77.5)	118 (58.7)	268	2 (2.2)	18 (22.5)	83 (41.3)	103	<0.0001
East African Indian	19 (95.0)	33 (73.3)	118 (58.1)	170	1 (5.0)	12 (26.7)	85 (41.9)	98	1.00
Indo-Oceanic	1 (100.0)	8 (61.5)	47 (41.9)	56	0 (0.0)	5 (38.5)	63 (57.3)	68	<0.0001

*The figures presented in each column refer to the number of cases and percentage. Percentages of cases that were pulmonary only and extrathoracic only are calculated for each lineage and demographic group separately. Data for extrapulmonary TB and lymph node TB not shown but results are essentially the same as those shown in table 4 for extrathoracic TB disease.

†Total refers to the total number of cases that were pulmonary only for each lineage (ie should be read across).

‡Total refers to the total number of case that were extrathoracic only for each lineage (ie should be read across).

§p Value refers to the comparison between the three demographic groups (UK-born white versus UK-born non-white versus foreign-born non-white) of the proportion of cases that are extrathoracic only for each lineage in turn; data analysed using Fisher's exact test.

¶¶For the total cohort the East Asian (51/545, 9.4% vs 48/317, 15.1%; p=0.014) and Indo-Oceanic lineages (51/545, 9.4% vs 48/317, 15.1%; p=0.014) were significantly more likely to cause extrathoracic only disease compared with pulmonary only disease whereas the Euro-American lineage was significantly more likely to cause pulmonary only disease than extrathoracic only disease (268/545, 49.2% vs 103/317, 32.5%; p<0.0001); data analysed using Fisher's exact test.

TB, tuberculosis.

East Asian and Indo-Oceanic lineages were significantly more likely to cause extrapulmonary only, extrathoracic only and lymph node only disease compared with pulmonary only disease (table 4). By contrast, the Euro-American lineage was significantly more likely to cause pulmonary disease and was inversely associated with extrapulmonary only, extrathoracic only and lymph node only disease (table 4).

Factors associated with disease patterns on multivariate analysis

Multivariate associations of factors associated with extrathoracic disease are presented in table 5. Certain lineages (with East Asian as the referent category) were negatively associated with extrathoracic only disease (East African Indian and Euro-American). With Euro-American lineage (the most frequent lineage in this cohort) as the reference category, the East Asian lineage (OR 1.86; 95% CI 1.10 to 3.15) was positively associated with extrathoracic disease but the East African Indian lineage was not (OR 0.94; 95% CI 0.62 to 1.42) (data not shown).

However, ethnicity was most strongly associated with clinical phenotype with Afro-Caribbean and Indian subcontinent ethnicities conferring adjusted ORs of 14 to 21, respectively, for extrathoracic disease.

Additional factors independently associated with the different extrapulmonary phenotypes were increasing age, female gender, foreign birth and time since UK entry (table 5).

DISCUSSION

TB can present as a disease spectrum usually involving the lungs but not uncommonly purely extrapulmonary or extrathoracic sites. Although these clinical disease patterns must represent the end result of an interaction between host, environment and mycobacterial lineage the relative importance of these three factors is unknown and can only be dissected in an ethnically heterogeneous population with detailed demographic, clinical and strain typing data, as presented here.

We found that *Mtb* lineages mapped with patients' geographical origins and were associated with distinct ethnic groups, as previously described.²⁻⁵ After adjustment for demographics, certain lineages, such as Euro-American and East African Indian, were significantly less likely to cause extrapulmonary TB

whereas others, such as East Asian, were associated with extrapulmonary disease. However, ethnicity was the factor most strongly associated with clinical phenotype. This was exemplified by the large population of Indian subcontinent ethnicity who, as previously described,²⁻³ harboured predominantly the East African Indian lineage which was preferentially associated with pulmonary disease; however, ethnicity overrode this association resulting in a predominantly extrathoracic and extrapulmonary pattern of disease. Indeed, with mutual multivariate adjustment, ethnicity in this ethnically diverse UK setting was more strongly associated with clinical phenotype than *Mtb* lineage, suggesting that the host is a relatively more powerful determinant of disease phenotype than mycobacterial lineage. Although non-white ethnicity has previously been recognised to be associated with extrapulmonary disease, the impact of *Mtb* lineage was not taken into account. Interpretation of such associations has therefore been difficult, given the association of different lineages with different ethnic groups and patients' geographical origins.

Only Click and colleagues, in the USA, assessed the association between *Mtb* lineage and disease phenotype taking ethnicity and geographical origins into account.²⁰ They found that while lineage was clearly associated with disease phenotype only weak associations with ethnicity were observed.²⁰ The reason for the weaker influence of ethnicity compared with our study is unclear but may lie in differences in the ethnicities of non-white ethnic groups in the USA compared with the UK. Specifically, Indian subcontinent ethnicity was not included as an ethnic grouping in the US study whereas in our study it was the largest subgroup and the most strongly associated with extrathoracic disease. This strong association of Indian subcontinent ethnicity with extrathoracic disease, independently of lineage, has broad epidemiological relevance because not only is this ethnic group the most important in terms of national TB burden in the UK,³¹ but it also accounts for more than a quarter of TB cases globally.¹ Although our findings highlight that the host is dominant in shaping clinical disease phenotype, we cannot currently distinguish whether the key determinant is the host per se (ie, host genotype) or whether ethnicity serves as a proxy for another host-related factor.³²

Previous studies of different lineages with clinical disease pattern have generated conflicting results. Associations identified

Table 5 Univariate and multivariate association of demographic, clinical and temporal factors with extrathoracic tuberculosis for four main lineages (excluding West African-1 and 2)

Variable	Extrathoracic TB only* (n=318)	Unadjusted OR (95% CI)	Adjusted OR† (95% CI)	p Value
Period				
2003–2004	54/192 (28.1)	1	1	0.53
2005–2006	114/280 (40.7)	1.76 (1.18 to 2.60)	1.39 (0.89 to 2.17)	
2007–2008	100/269 (37.2)	1.51 (1.01 to 2.26)	1.21 (0.77 to 1.92)	
2009–2010	50/134 (37.3)	1.52 (0.95 to 2.44)	1.16 (0.67 to 2.00)	
Age categories n (%)				
Under 16	9/27 (33.3)	1	1	0.005
16–25	69/234 (29.5)	0.84 (0.36 to 1.95)	0.72 (0.29 to 1.80)	
26–35	88/208 (42.3)	1.47 (0.63 to 3.42)	1.10 (0.43 to 2.82)	
36–45	71/163 (43.6)	1.54 (0.65 to 3.64)	1.74 (0.66 to 4.59)	
46–55	33/75 (44)	1.57 (0.63 to 3.95)	1.91 (0.64 to 5.65)	
Over 55	48/168 (28.6)	0.80 (0.34 to 1.90)	0.87 (0.32 to 2.42)	
Gender				
Men	147/461 (31.9)	1	1	0.016
Women	171/414 (41.3)	1.50 (1.14 to 1.98)	1.49 (1.08 to 2.06)	
Country of birth/time since entry				
UK born	41/270 (15.2)	1	1	0.0018
Foreign born/<1 year	5/20 (25)	1.86 (0.64 to 5.40)	0.74 (0.22 to 2.58)	
Foreign born/1–5 years	130/278 (46.8)	4.91 (3.26 to 7.37)	2.26 (1.38 to 3.69)	
Foreign born/6–10 years	42/75 (56)	7.11 (4.04 to 12.5)	3.08 (1.60 to 5.92)	
Foreign born/>10 years	82/195 (42.1)	4.05 (2.62 to 6.28)	1.37 (0.75 to 2.51)	
Foreign born/time not known	18/37 (48.6)	5.29 (2.56 to 10.93)	2.28 (1.00 to 5.22)	
Ethnicity				
White	3/132 (2.3)	1	1	<0.0001
Indian subcontinent	206/466 (44.2)	34.07 (10.69 to 108.59)	21.14 (6.08 to 73.48)	
Afro-Caribbean	88/222 (39.6)	28.24 (8.71 to 91.52)	14.05 (3.97 to 49.65)	
Oriental/other Asia	4/20 (20)	10.75 (2.20 to 52.43)	4.56 (0.82 to 25.27)	
Other	17/35 (48.6)	40.61 (10.82 to 152.44)	23.62 (5.68 to 98.17)	
Lineage				
East Asian	48/100 (48)	1	1	0.01
Euro-American	103/378 (27.2)	0.41 (0.26 to 0.64)	0.54 (0.32 to 0.91)	
East African Indian	98/269 (36.4)	0.62 (0.39 to 0.99)	0.50 (0.29 to 0.86)	
Indo-Oceanic	68/124 (54.8)	1.32 (0.78 to 2.23)	0.91 (0.51 to 1.63)	
Employment status				
Unemployed	34/98 (34.7)	1	1	
Retired	12/54 (22.2)	0.54 (0.25 to 1.16)		
Employed/housewife	140/324 (43.2)	1.43 (0.89 to 2.29)		
Student/child	24/85 (28.2)	0.74 (0.39 to 1.39)		
HIV status				
Negative/unknown	314/867 (36.2)	1	1	0.88
Positive	4/8 (50.0)	1.76 (0.44 to 7.09)	0.9 (0.21 to 3.84)	
Previous TB				
No	308/841 (36.6)	1	1	0.79
Yes	10/34 (29.4)	0.72 (0.34 to 1.53)	0.89 (0.39 to 2.04)	
Drug sensitivity				
Fully sensitive	288/800 (36.0)	1	1	0.95
Drug-resistant‡	16/41 (39.0)	1.14 (0.60 to 2.17)	1.02 (0.5 to 2.1)	

*Comparison of extrathoracic TB cases only versus pulmonary TB only (concurrent pulmonary–extrapulmonary cases excluded); data for extrapulmonary TB and lymph node TB not shown but results are essentially the same as those shown in table 5 for extrathoracic TB disease.

†Model mutually adjusted for year of notification, age, gender, location of birth, time since arrival in the UK, ethnicity, TB lineage, HIV status, past history of TB and drug sensitivity; data analysed using logistic regression.

‡Drug-resistant category includes individuals with mono-resistant TB and multi-drug resistant TB, tuberculosis.

in a given ethnically homogeneous population have frequently not been replicable in other populations while studies in ethnically heterogeneous populations have hitherto failed to control for ethnicity, with the exception of Click *et al.*²⁰

We found that the Euro-American and East African Indian lineages were associated with pulmonary disease whereas Click

and colleagues found these lineages were associated with extra-pulmonary disease.²⁰ Previous authors have come to conflicting conclusions about the phenotypic associations of the Euro-American^{19 33} and East African Indian lineages.^{19 21–23} With the Euro-American lineage as the reference, we found the East Asian lineage was significantly associated, on multivariate

analysis, with extrapulmonary disease. The mechanism underlying this propensity to disseminate is not known but may involve the ability of the Beijing (East Asian) lineage to produce a bioactive phenolic glycolipid which inhibits the innate immune system.^{33–35} As with other lineages, earlier studies have come to conflicting conclusions about the phenotypic association of the East Asian lineage.^{14–21}

The predominant lineages in our study were geographically associated (table 2): Euro-American with Europe/Americas/Africa, East African Indian with the Indian subcontinent/East Africa, Indo-Oceanic with the Indian subcontinent/East Africa/Asia and East Asian with the Indian subcontinent/East Africa/Asia. Our findings are in line with previous studies,^{2 3 5} although the high proportion of individuals in our cohort from the Indian subcontinent, compared with the small proportion from East, and Southeast, Asia may have masked the association of these latter regions with the Indo-Oceanic and East Asian lineages.^{4 5}

Lineages were also associated with distinct ethnic groups, with the Euro-American lineage associated with White and Afro-Caribbean individuals, East African Indian and Indo-Oceanic with Indian subcontinent ethnicity and East Asian with Oriental and Other Asia ethnicity. These findings are consistent with those of Evans *et al*⁴ who assigned ethnicity by name and Gagneux *et al*² who found that US-born individuals of non-white ethnicity (mainly Chinese and Filipino) had the same lineages as individuals born in China and the Philippines. These close associations of lineages with self-reported ethnicity, an accurate marker of ancestry,³⁶ are consistent with coevolution of mycobacteria and their human hosts. Alternatively, these associations could reflect infection with similar lineages following transmission within ethnic communities in the host countries through assortative mixing, after travel to countries of origin or visits from family based overseas.³⁷

Our study had several limitations. One of these was lineage assignment. Although we used 15 loci MIRU-VNTR (the standard DNA fingerprinting method at the time) an online database (MIRU-VNTRplus) and published literature to assign lineages to isolates, a proportion could not be definitively classified into any of the six major lineages either due to the lower resolution of 15-MIRU-VNTR or because the online database did not contain data on enough lineages. However, the demographic and clinical profile of these individuals did not differ from the cohort that was analysed; sensitivity analyses including the non-lineage-assigned isolates did not provide any evidence that these isolates formed a distinct lineage with a separate clinical phenotype. In addition, the 15-MIRU-VNTR method was recently effectively used for lineage assignment by Click *et al*.²⁰ An improved version of 15-MIRU-VNTR typing has subsequently been described that analyses 24-loci and provides greater resolution; an alternative would be to base lineage assignment on large sequence polymorphisms or whole-genome sequencing.

Although national guidelines now state that patients with TB should be tested for HIV,³⁸ this was inconsistently recorded in our dataset as in many national surveillance systems.^{10 11 21 39} While untreated HIV infection is well recognised to be an important factor in determining an extrapulmonary clinical phenotype,⁴⁰ its influence on clinical phenotype in our predominantly Indian subcontinent study population is probably very small as TB-HIV coinfection in this segment of the UK population is very low.⁴¹ Moreover, there is no association between HIV and clinical disease pattern in the post-antiretroviral therapy era.¹² Other factors which may affect clinical

presentation include age, iatrogenic immunosuppression and diabetes mellitus.^{42–44} To ensure unambiguous clinical phenotypes for our regression analyses, we classified cases as exclusively pulmonary and exclusively extrathoracic; the resulting exclusion of individuals with concurrent disease might therefore have resulted in some associations being missed. Future work should explore the factors (including ethnicity) that are associated with different extrathoracic disease subtypes. Whilst our study focused primarily on host factors and mycobacterial lineage, an alternative area of potential relevance, and future study, is the impact of environment on clinical phenotype with, for example, the recency of *Mtb* transmission potentially playing a role in determining clinical phenotype.

In conclusion, although *Mtb* lineages, which are geo-ethnically restricted, can be associated with specific clinical phenotypes, host ethnicity appears to be more important in determining the clinical pattern of tuberculous disease, at least in the case of Indian subcontinent ethnicity and extrathoracic and extrapulmonary disease. Future work should integrate host demographic and genotypic data with environmental data for factors associated with TB, such as smoking and vitamin D levels.

Acknowledgements This publication made use of the MIRU-VNTRplus database website (<http://www.miru-vntrplus.org/>) developed by D Harmsen, S Nieman, P Supply and T Weniger.

Contributors AL, MP, JI and GS conceived of the idea for the study; JI: collected the clinical data as part of routine care; JE, EGS, PH: undertook the MIRU-VNTR typing and maintained the genotyping database; MP: undertook the data analysis; MP, AL: wrote the first draft of the manuscript with subsequent revisions made by all other coauthors. All coauthors had sight of the submitted paper. AL: guarantor for the paper.

Competing interests AL is inventor for patents underpinning T-cell-based diagnosis. The ESAT-6/CFP-10 ELISpot was commercialised by an Oxford University spin-out company (Oxford Immunotec Ltd, Abingdon, UK) in which Oxford University and Professor Lalvani have a minority share of equity. MP, JE, JI, EGS, SHW, KL, SS, MD, PH have no conflict of interest.

Funding MP is currently a NIHR Academic Clinical Lecturer in Infectious Diseases; his PhD was funded by a Medical Research Council Capacity Building Studentship. AL is a Wellcome Senior Research Fellow in Clinical Science and NIHR Senior Investigator.

Ethics approval No patient-specific data or personal identifiers were used in the preparation of this report which was an analysis of routine data collected as part of service evaluation.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 World Health Organisation. *Global tuberculosis control*. Geneva: WHO, 2011.
- 2 Gagneux S, DeRiemer K, Van T, *et al*. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* 2006;103:2869–73.
- 3 Reed MB, Pichler VK, McIntosh F, *et al*. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J Clin Microbiol* 2009;47:1119–28.
- 4 Evans JT, Gardiner S, Smith EG, *et al*. Global origin of *Mycobacterium tuberculosis* in the Midlands, UK. *Emerg Infect Dis* 2010;16:542–5.
- 5 Brown T, Nikolayevskyy V, Velji P, *et al*. Associations between *Mycobacterium tuberculosis* strains and phenotypes. *Emerg Infect Dis* 2010;16:272–80.
- 6 Coscolla M, Gagneux S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov Today Dis Mech* 2010;7:e43–59.
- 7 Portevin D, Gagneux S, Comas I, *et al*. Human macrophage responses to clinical isolates from the *Mycobacterium tuberculosis* complex discriminate between ancient and modern lineages. *PLoS Pathog* 2011 7:e1001307.
- 8 Wang C, Peyron P, Mestre O, *et al*. Innate immune response to *Mycobacterium tuberculosis* Beijing and other genotypes. *PLoS ONE* 2010;5:e13594.
- 9 Li Q, Whalen CC, Albert JM, *et al*. Differences in rate and variability of intracellular growth of a panel of *Mycobacterium tuberculosis* clinical isolates within a human monocyte model. *Infect Immun* 2002;70:6489–93.
- 10 Peto HM, Pratt RH, Harrington TA, *et al*. Epidemiology of extrapulmonary tuberculosis in the United States, 1993–2006. *Clin Infect Dis* 2009;49:1350–57.
- 11 te Beek LA, van der Werf MJ, Richter C, *et al*. Extrapulmonary tuberculosis by nationality, The Netherlands, 1993–2001. *Emerg Infect Dis* 2006;12:1375–82.

- 12 Kruijshaar ME, Abubakar I. Increase in extrapulmonary tuberculosis in England and Wales 1999–2006. *Thorax* 2009;64:1090–95.
- 13 Thwaites G, Caws M, Chau TTH, et al. The relationship between *Mycobacterium tuberculosis* genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. *J Clin Microbiol* 2008;46:1363–8.
- 14 Kong Y, Cave MD, Zhang L, et al. Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: insights from epidemiologic and clinical characterization of the three principal genetic groups of *M. tuberculosis* clinical isolates. *J Clin Microbiol* 2007;45:409–14.
- 15 Kong Y, Cave MD, Yang D, et al. Distribution of insertion- and deletion-associated genetic polymorphisms among four *Mycobacterium tuberculosis* phospholipase C genes and associations with extrathoracic tuberculosis: a population-based study. *J Clin Microbiol* 2005;43:6048–53.
- 16 Hesselting AC, Marais BJ, Kirchner HL, et al. Mycobacterial genotype is associated with disease phenotype in children. *Int J Tuberc Lung Dis* 2010;14:1252–58.
- 17 Nicol MP, Sola C, February B, et al. Distribution of strain families of *Mycobacterium tuberculosis* causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. *J Clin Microbiol* 2005;43:5779–81.
- 18 Borgdorff MW, van Deutekom H, de Haas PEW, et al. *Mycobacterium tuberculosis*, Beijing genotype strains not associated with radiological presentation of pulmonary tuberculosis. *Tuberculosis (Edinb)* 2004;84:337–40.
- 19 Lari N, Rindi L, Cristofani R, et al. Association of *Mycobacterium tuberculosis* complex isolates of Bovis and Central Asian (CAS) genotypic lineages with extrapulmonary disease. *Clin Microbiol Infect* 2009;15:538–43.
- 20 Click ES, Moonan PK, Winston CA, et al. Relationship between *Mycobacterium tuberculosis* phylogenetic lineage and clinical site of tuberculosis. *Clin Infect Dis* 2012;54:211–19.
- 21 Svensson E, Millet J, Lindqvist A, et al. Impact of immigration on tuberculosis epidemiology in a low-incidence country. *Clin Microbiol Infect* 2011;17:881–7.
- 22 Hasan Z, Tanveer M, Kanji A, et al. Spoligotyping of *Mycobacterium tuberculosis* isolates from Pakistan reveals predominance of Central Asian Strain 1 and Beijing isolates. *J Clin Microbiol* 2006;44:1763–68.
- 23 Tanveer M, Hasan Z, Siddiqui A, et al. Genotyping and drug resistance patterns of *M. tuberculosis* strains in Pakistan. *BMC Infect Dis* 2008;8:171.
- 24 Office for National Statistics. *Annual Population Survey: Population by country of birth and nationality October 2009 to September 2010*. London: Office for National Statistics, 2010.
- 25 Health Protection Agency. *Three-year Average Tuberculosis Case Reports and Rates by Primary Care Trust, England, 2006–2008*. London: Health Protection Agency, 2009.
- 26 Evans JT, Hawkey PM, Smith EG, et al. Automated high-throughput mycobacterial interspersed repetitive unit typing of *Mycobacterium tuberculosis* strains by a combination of PCR and nondenaturing high-performance liquid chromatography. *J Clin Microbiol* 2004;42:4175–80.
- 27 Supply P, Allix C, Lesjean S, et al. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J Clin Microbiol* 2006;44:4498–510.
- 28 Allix-Beguec C, Harmsen D, Weniger T, et al. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. *J Clin Microbiol* 2008;46:2692–99.
- 29 Office of National Statistics. *Ethnic Group Statistics: A Guide for the Collection and Classification of Ethnicity Data*. London: ONS, 2009.
- 30 García-Rodríguez JF, Álvarez-Díaz H, Lorenzo-García MV, et al. Extrapulmonary tuberculosis: epidemiology and risk factors. *Enferm Infect Microbiol Clin* 2011;29:502–9.
- 31 Health Protection Agency. *Tuberculosis in the UK: Annual Report on Tuberculosis Surveillance and Control in the UK 2010*. London: Health Protection Agency Centre for Infections, 2010.
- 32 Wilkinson RJ, Llewelyn M, Toossi Z, et al. Influence of vitamin D deficiency and vitamin D receptor polymorphisms on tuberculosis among Gujarati Asians in west London: a case-control study. *Lancet* 2000;355:618–21.
- 33 Caws M, Thwaites G, Dunstan S, et al. The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog* 2008;4:e1000034.
- 34 Reed MB, Domenech P, Manca C, et al. A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* 2004;431:84–7.
- 35 Tsenova L, Ellison E, Harbacheuski R, et al. Virulence of selected *Mycobacterium tuberculosis* clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. *J Infect Dis* 2005;192:98–106.
- 36 Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic structure of human populations. *Science* 2002;298:2381–85.
- 37 Cobelens FGJ, van Deutekom H, Draayer-Jansen IWE, et al. Risk of infection with *Mycobacterium tuberculosis* in travellers to areas of high tuberculosis endemicity. *Lancet* 2000;356:461–65.
- 38 National Collaborating Centre for Chronic Conditions. *Tuberculosis: Clinical Diagnosis and Management of Tuberculosis, and Measures for its Prevention and Control*. London: Royal College of Physicians, 2011.
- 39 Rodger AJ, Story A, Fox Z, et al. HIV prevalence and testing practices among tuberculosis cases in London: a missed opportunity for HIV diagnosis? *Thorax* 2010;65:63–9.
- 40 Yang Z, Kong Y, Wilson F, et al. Identification of risk factors for extrapulmonary tuberculosis. *Clin Infect Dis* 2004;38:199–205.
- 41 Ahmed AB, Abubakar I, Delpech V, et al. The growing impact of HIV infection on the epidemiology of tuberculosis in England and Wales: 1999–2003. *Thorax* 2007;62:672–76.
- 42 Ong A, Creasman J, Hopewell PC, et al. A molecular epidemiological assessment of extrapulmonary tuberculosis in San Francisco. *Clin Infect Dis* 2004;38:25–31.
- 43 Keane J, Gershon S, Wise RP, et al. Tuberculosis associated with infliximab, a tumor necrosis factor α -neutralizing agent. *N Engl J Med* 2001;345:1098–104.
- 44 Dooley KE, Chaisson RE. Tuberculosis and diabetes mellitus: convergence of two epidemics. *Lancet Infect Dis* 2009;9:737–46.

Online Supplementary Information

Ethnicity and mycobacterial lineage as determinants of tuberculosis disease phenotype

Manish Pareek PhD MRCP *Clinical Lecturer*^{a,b,c}

Jason Evans MSc PhD *Clinical Scientist*^d

John Innes FRCP *Honorary Consultant*^e

E Grace Smith FRCP FRCPath *Consultant*^d

Suzie Hingley-Wilson PhD *Research Associate*^a

Kathryn E Lougheed PhD *Research Associate*^f

Saranya Sridhar DPhil *Research Associate*^a

Martin Dedicoat PhD MRCP *Consultant*^e

Peter Hawkey PhD *Professor*^{d,g}

Ajit Lalvani DM FRCP *Chair of Infectious Diseases and Director*^a

^aTuberculosis Research Unit, Department of Respiratory Medicine, National Heart and Lung Institute, Imperial College London, UK

^bDepartment of Infectious Disease Epidemiology, Imperial College London, UK

^cDepartment of Infection, Immunity and Inflammation, University of Leicester, UK

^dHealth Protection Agency Midlands Laboratory, Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, UK

^eDepartment of Infection and Tropical Medicine, Heart of England NHS Foundation Trust, Bordesley Green East, Birmingham, UK

^fDepartment of Integrative Cell Biology, Division of Cell & Molecular Biology, Imperial College London, UK

^gSchool of Immunity and Infection, The Medical School, University of Birmingham, Edgbaston, B15 2TT

Number of supplementary tables: 1 Number of supplementary figures: 3

Supplementary data analysis

Demographic characteristics of the cohort were described using median and interquartile range (IQR) (for continuous variables) and proportions/percentages (categorical variables); comparisons were made using the non-parametric Mann-Whitney U-test and Pearson's chi-square test (or Fisher's exact test if appropriate) respectively.

We initially described the temporal dynamics and distribution of the global lineages in our cohort separately by the individuals' country of origin, region of origin (subdivided into Americas and Caribbean, Europe, Middle East/North Africa, South Africa, West and Central Africa, East Africa, Indian Subcontinent and Central Asia, Southeast and East Asia) and ethnic group. Comparisons of which lineages caused disease in UK- and foreign-born individuals was made using Fisher's exact test.

Factors associated with *Mtb* lineage were assessed by constructing separate regression models for each of the main lineages in our study - East Asian, Euro-America, East African Indian and Indo-Oceanic. We did not undertake regression analyses for two lineages (West African-1 and West African-2) due to the small number of cases caused by these lineages in our dataset. Univariate association of clinical and demographic factors individually associated with each lineage was assessed using logistic regression and reported as crude odds ratios (OR) and 95% confidence intervals (CI). To calculate adjusted odds ratios (and 95% CI) we mutually adjusted for year of notification, age, gender, ethnicity, country of birth, HIV status, previous history of TB, drug sensitivity and smear positivity.

For the next stage of the analysis we described the broad clinical phenotypes (pulmonary only, extrapulmonary only and concurrent pulmonary-extrapulmonary) associated with each of the lineages. We then compared, for each lineage, the proportion of cases that were pulmonary only as compared to extrapulmonary only, extrathoracic only and lymph node only respectively; comparisons were made using univariate logistic regression.

We then assessed the factors (especially *Mtb* lineage) associated with extrathoracic only, extrapulmonary only and lymph node only disease respectively versus exclusively pulmonary disease with three separate logistic regression models (for each clinical phenotype). Univariate, and multivariate, associations of factors individually associated with each of the three clinical phenotypes detailed above (compared against pulmonary cases only) was reported as crude, and adjusted, odds ratios (OR) respectively (and 95% confidence intervals). Adjusted odds ratios (and 95% CI) were obtained by mutually adjusting for the following factors in all three models: year of notification, age, gender, ethnicity, country of birth, time since arrival, HIV status, previous history of TB, drug sensitivity and lineage. In the descriptive analysis we categorised individuals as having fully sensitive disease, mono-resistant disease (in other words resistance to one/more chemotherapeutic agent but not multi-drug resistant) and multi-drug resistant disease (in other words resistance to rifampicin and isoniazid). For the logistic modelling due to small numbers of drug resistant TB we combined mono-resistant TB and multi-drug resistant TB into one category “drug-resistant”. In the regression models, we concentrated on the four main lineages present in our cohort. We used the East Asian lineage as the reference category as this strain has been previously shown to be associated with extrapulmonary disease although,¹⁴⁻¹⁶ as a sensitivity analysis, we reran the multivariate model using the Euro-American lineage (which was the commonest) as the reference lineage against which the other lineages were compared.

Analyses used STATA 9.2 (StataCorp, College Station, TX). All tests were two tailed; p-value of ≤ 0.05 was considered significant.

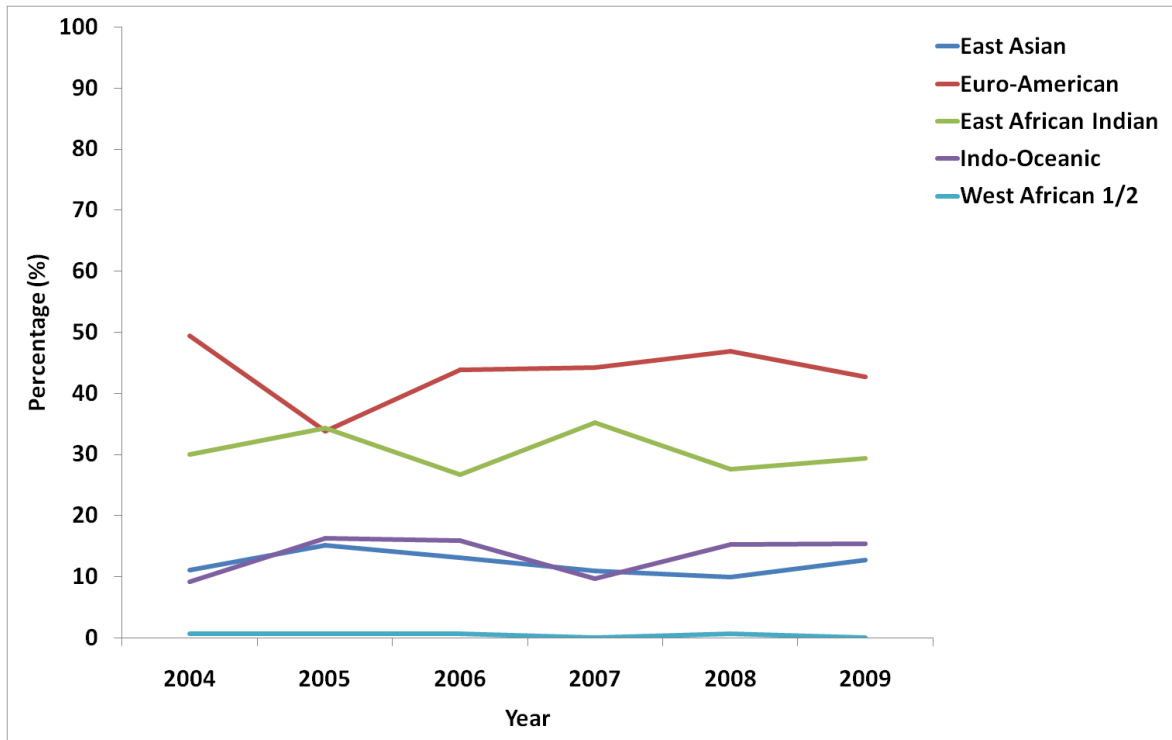
Supplementary tables

Supplementary table 1. Distribution of *M. tuberculosis* lineages by different demographic groups (stratified by ethnicity and location of birth) (percentages reflect the percentage of each lineage in each of the demographic groups – ie. Should be read across for each row)

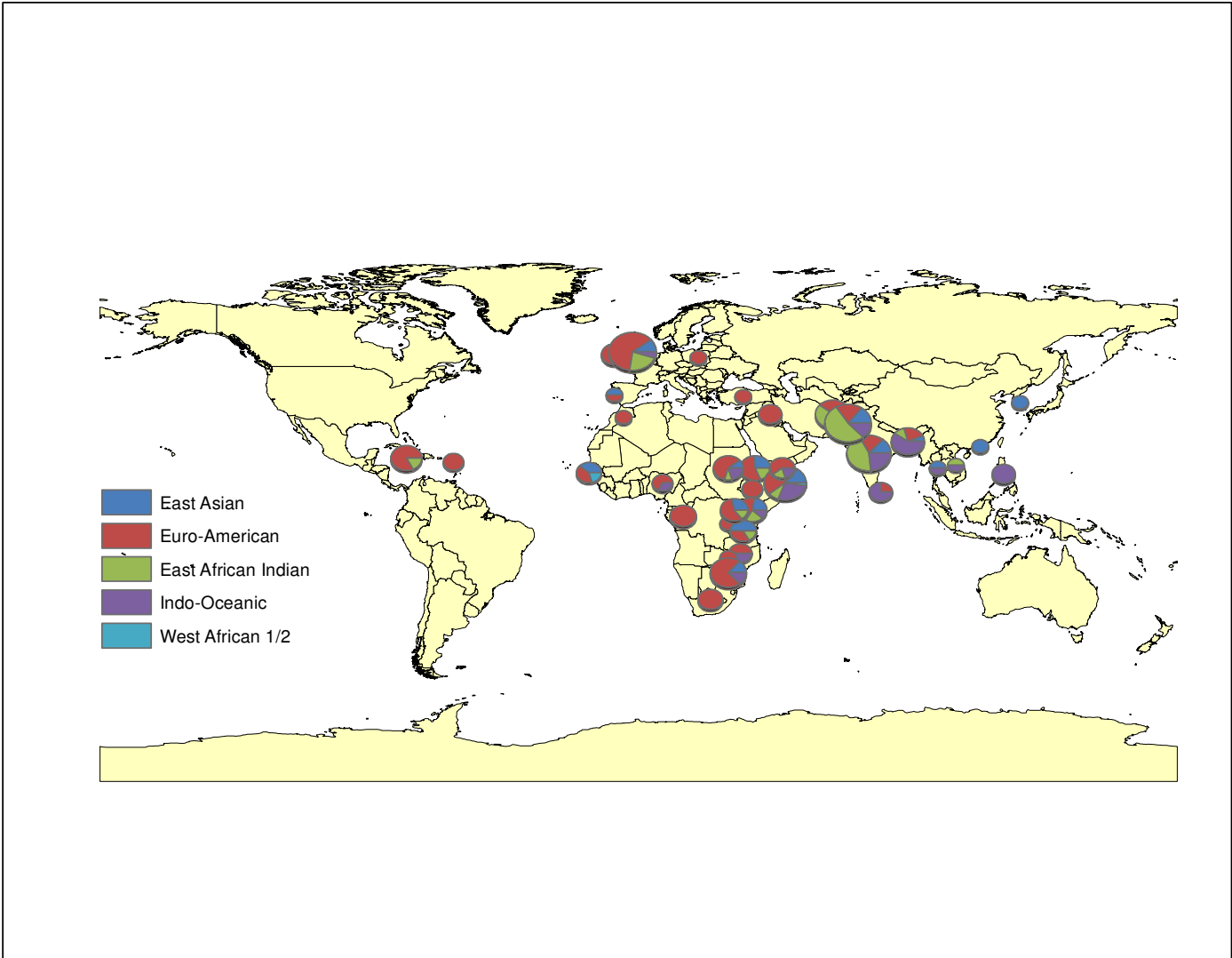
Location of birth/ethnicity	Lineage				
	East Asian (n=130)	Euro-American (n=468)	East African Indian (n=323)	Indo-Oceanic (n=145)	West African 1/2 (n=4)
Total White	16 (10.8)	110 (74.3)	21 (14.2)	1 (0.7)	0 (0.0)
UK born White	14 (10.2)	102 (74.5)	20 (14.6)	1 (0.7)	0 (0.0)
Foreign-born White	2 (18.2)	8 (72.7)	1 (9.1)	0 (0)	0 (0.0)
Total Indian Subcontinent	69 (12.0)	151 (26.2)	262 (45.4)	94 (16.3)	1 (0.2)
UK born Indian Subcontinent	8 (6.8)	50 (42.7)	45 (38.5)	13 (11.1)	1 (0.9)
Foreign-born Indian Subcontinent	61 (13.3)	101 (22.0)	217 (47.2)	81 (17.6)	0 (0.0)
Total Afro-Caribbean	35 (12.6)	180 (64.8)	21 (7.6)	39 (14.0)	3 (1.1)
UK born Afro-Caribbean	6 (11.1)	43 (79.6)	5 (9.3)	0 (0.0)	0 (0.0)
Foreign-born Afro-Caribbean	29 (13.0)	137 (61.2)	16 (7.1)	39 (17.4)	3 (1.4)
Total Oriental/Other Asia	7 (30.4)	4 (17.4)	6 (26.1)	6 (26.1)	0 (0.0)
UK born Oriental/Other Asia	1 (33.3)	1 (33.3)	0 (0.0)	1 (33.3)	0 (0.0)
Foreign-born Oriental/Other Asia	6 (30.0)	3 (15.0)	6 (30.0)	5 (25.0)	0 (0.0)
Total Other	3 (6.8)	23 (52.3)	13 (29.6)	5 (11.4)	0 (0.0)
UK born Other	1 (14.3)	4 (57.1)	2 (28.6)	0 (0.0)	0 (0.0)
Foreign-born Other	2 (5.4)	19 (51.4)	11 (29.7)	5 (13.5)	0 (0.0)

Supplementary figures

Supplementary figure 1. Temporal trends in lineages causing active tuberculosis between 2004 and 2009



Supplementary figure 2. Distribution of *M. tuberculosis* lineages by country of origin amongst the individuals in the cohort (size of pie charts is scaled relative to the number of cases from that country and the proportion of cases from each lineage) for all lineages



Supplementary figure 3. Distribution of mycobacterial lineages by time since arrival for cases amongst the foreign-born

