Microbiology (2006), **152**, 2757–2765 DOI 10.1099/mic.0.28986-0

Metabolic fingerprints of *Mycobacterium bovis* cluster with molecular type: implications for genotype–phenotype links

Catherine L. Winder,[1] Stephen V. Gordon,[2] James Dale,[2] R. Glyn Hewinson[2] and Royston Goodacre[1]

[1]School of Chemistry, Faculty of Engineering and Physical Sciences, The University of Manchester, PO Box 88, Sackville Street, Manchester M60 1QD, UK

[2]TB Research Group, Veterinary Laboratories Agency (VLA), Weybridge, New Haw, Addlestone, Surrey KT15 3NB, UK

Correspondence
Catherine L. Winder
C.Winder@manchester.ac.uk

Received 10 March 2006
Revised 1 June 2006
Accepted 5 June 2006

Mycobacterium bovis is the causative agent of bovine tuberculosis. Various genetic typing techniques have been used to trace the reservoirs of infection; however, they have limited success in population genetics and outbreak studies. Fourier-transform infrared spectroscopy (FT-IR) is a rapid phenotypic typing technique, which may be used to generate a metabolic fingerprinting and is increasingly used to characterize bacteria. When coupled with multivariate cluster analysis, this powerful combination has sufficient resolving power to discriminate bacteria down to subspecies level; however, to date this method has not been used in the differentiation of mycobacteria. Multiple isolates of the ten major spoligotypes in the UK, recovered from different geographical locations, were analysed using FT-IR. Hierarchical cluster analysis of the spectra showed that the isolates could be differentiated according to their spoligotypes. Six of the spoligotype FT-IR clusters were very homogeneous and all isolates were recovered together. However, the remaining four groups displayed a more heterogeneous phenotype, which may reflect greater variation than previously suspected within these groups. Included in the ten spoligotypes are the two most dominant isolates in the UK, designated types 9 and 17. Whilst type 17 showed a highly conserved phenotype as judged by FT-IR, type 9 showed a very heterogeneous metabolic profile and isolates were recovered throughout the dendrogram. This variation in type 9 is reflected in the high degree of diversity observed by variable number tandem repeats (VNTR) analysis, underlining the exquisite resolving power of FT-IR.

INTRODUCTION

Mycobacterium bovis is the causative agent of tuberculosis in a range of animal species and man, with worldwide annual losses to agriculture estimated at US$3 billion. Many animals, including badgers, foxes, ferrets and deer, are believed to act as vectors in the transmission to cattle (Aranaz *et al.*, 1996; Clifton-Hadley *et al.*, 1995; Delahay *et al.*, 2001; Griffin *et al.*, 2005), with the badger (*Meles meles*) implicated as providing the most significant reservoir of infection in the UK. In the face of complex interactions between wildlife reservoirs and domesticated animals, it is essential that robust epidemiological techniques are applied to determine sources of infection and explore the extant *M. bovis* population. An *M. bovis* typing technique that is highly discriminatory, reproducible and high-throughput is therefore a major research goal.

The ideal microbial typing technique would be easy to use, automated, with intuitive nomenclature and have good inter-laboratory reproducibility (Roring *et al.*, 2004). Restriction fragment length polymorphisms (RFLP) (Goyal *et al.*, 1997) of the IS*6110* element are generally used in discriminating members of the *M. tuberculosis* complex; however, this technique has limited use in *M. bovis* due to the low copy number of IS*6110*. Therefore, alternative genetic typing methods have been used in the analysis of *M. bovis*, with spacer oligonucleotide typing, spoligotyping (Kamerbeek *et al.*, 1997), the current method of choice. Spoligotyping focuses on identifying repeat variation in the direct repeat (DR) locus, which is composed of multiple, virtually identical 36 bp repeats interspersed with unique

Abbreviations: DFA, discriminant function analysis; DR, direct repeat; DTGS, deuterated triglycine sulphate; DVR, direct variant repeat; ETR, exact tandem repeat; FT-IR, Fourier-transform infrared spectroscopy; HCA, hierarchical cluster analysis; MCT, mercury-cadmium-telluride; MIRU, mycobacterial interspersed repetitive unit; OADC, oleic acid, albumin, dextrose (glucose) and catalase; PCA, principal components analysis; VNTR, variable number tandem repeat.

DNA spacer sequences (direct variant repeats; DVRs) of similar size. Different spoligotype patterns are identified by the presence or absence of DVRs. In the UK ten major spoligotypes account for ~90 % of all isolates, with the two dominant spoligotypes 9 and 17 representing 65 % of all isolates (Inwald *et al.*, 2002). The dominance of two main types and the tendency of the spoligotype pattern to alter slowly with time (Smith *et al.*, 2003) reduce the usefulness of spoligotyping as an epidemiological tool.

Alternative genotypic typing techniques identifying sets of tandem repeats, such as exact tandem repeats (ETRs), mycobacterial interspersed repetitive units (MIRUs) and variable number tandem repeats (VNTRs) have been described as independent tools for the discrimination of members of the *M. tuberculosis* complex (Roring *et al.*, 2004). VNTRs have been successfully used in the discrimination of *M. bovis* isolates, and the combination of spoligotyping with VNTR analysis vastly improves the levels of discrimination (Gibson *et al.*, 2004). Indeed, our previous work has shown that VNTR analysis can successfully subdivide spoligotype 9 into clusters of related strains; however, the VNTR technique requires optimization and the level of discrimination depends upon the loci applied and test panel chosen for the study. Different degrees of discrimination may be appropriate for different studies, with some loci useful for population-based studies whereas more discriminating loci may be required to monitor outbreak analysis (Roring *et al.*, 2004).

In this study we investigated the ability of the vibrational spectroscopic technique Fourier-transform infrared spectroscopy (FT-IR) to produce robust biochemical 'fingerprints' of *M. bovis* isolates. Numerous advantages of using this method of discrimination over traditional microbiological and molecular techniques are documented; these include speed, minimal sample preparation, automation and relatively low expense. The use of multivariate statistical techniques to analyse the infrared spectra has proved successful in the discrimination of the bacteria to subspecies levels (Maquelin *et al.*, 2003; Naumann *et al.*, 1991; Timmins *et al.*, 1998; Winder *et al.*, 2004). To date, the discrimination of members of the *Mycobacterium* genus by FT-IR has not been reported in the literature. Moreover, we show that the FT-IR method successfully clustered strains of the same molecular type, suggesting that molecular types share distinct phenotypic characteristics that were discriminated by FT-IR. The implications of this result are discussed.

METHODS

Cultivation of *M. bovis*. Representative isolates of the ten major spoligotypes in the UK were selected. Details of the origins of the isolates are given in Table 1. Spectra were excluded from the study when the signal-to-noise ratio was below 80; this was calculated using the wavenumber range 1900–1800 cm^{-1} for noise and the 1700–1600 cm^{-1} region for signal. The strains were cultured in triplicate for 4 weeks at 37 °C on Middlebrook 7H10 medium [supplemented with 10 % oleic acid, albumin, glucose and catalase (OADC); Becton Dickinson USA] and with antifungal and selective agents fungizone (5 mg ml^{-1}), polymixin B (100 000 IU ml^{-1}), amoxil (200 mg ml^{-1}) added. The biomass from the slopes was harvested in 100 µl sterile Millipure water. The cells were pasteurized at 80 °C for 1 h and the viability of the cells was checked to ensure no growth occurred.

FT-IR spectroscopy of mycobacteria. Two methods of FT-IR were investigated, using two different sample presentation techniques – transmission and reflection – on two different spectrometers.

Reflectance-based FT-IR. A 10 × 10 cm aluminium plate was rinsed with acetone and dried at 50 °C for 10 min prior to use. FT-IR analysis was initially performed on the plate without samples to provide a reference reading for each well. The plate was loaded onto the motorized stage of a reflectance TLC accessory (Winder *et al.*, 2004) attached to a Bruker IFS28 FT-IR spectrometer (Bruker Spectrospin). This was equipped with a mercury-cadmium-telluride (MCT) detector, which was cooled with liquid N_2. A 5 µl aliquot of each sample was evenly applied in triplicate onto the 400-well plate (the so-called 'machine' replicates). The plate was then dried at 50 °C for 30 min before FT-IR analysis was performed. Spectra were collected over the wavelength range of 4000–600 cm^{-1} under the control of an IBM-compatible computer programmed with Opus 2.1 running under IBM OS/2 Warp, which was provided by the manufacturers. Spectra were acquired with a resolution of 4 cm^{-1} and 256 spectra were co-added and averaged. The collection time for the spectra was approximately 10 s per sample. The spectra are displayed in terms of absorbance, which was calculated from the reflectance-absorbance spectra using Opus software.

Transmission-based FT-IR. Prior to use a 96-well zinc selenide plate was rinsed with 2-propanol and deionized water and allowed to dry at room temperature. Aliquots (15 µl) of each bacterial sample were evenly applied in triplicate onto the plate and dried at 50 °C for 30 min. The plate was loaded onto a motorized microplate module HTS-XT attached to an Equinox 55 module (Bruker Optics). The motorized module of this instrument introduces the plate into the airtight optics of the instrument, in which tubes of desiccant are contained to remove moisture (Harrigan *et al.*, 2004). A deuterated triglycine sulphate (DTGS) detector was employed for transmission measurements of the samples to be acquired. Spectra were collected over the wavelength range 4000–900 cm^{-1} (a reduced range is used in comparison to the reflectance measurements due to the zinc selenide focusing element) under the control of a computer programmed with Opus 4, operated under MS Windows 2000. Spectra were acquired at a resolution of 4 cm^{-1} and 64 spectra were co-added and averaged to improve the signal-to-noise ratio. The collection time for each spectrum was approximately 1 min and the spectra were displayed in terms of absorbance.

Preprocessing. The ASCII data were imported into Matlab version 6 (The MathWorks). To minimize problems arising from baseline shifts, Matlab was used to correct for CO_2 vibrations [the CO_2 peaks at 2403–2272 cm^{-1} and 683–656 cm^{-1} (when present) were removed and filled with a trend] and windows of the spectra likely containing H_2O vibrations were smoothed with a window of 55 cm^{-1} to reduce noise (4000–3615 cm^{-1} and 1998–1763 cm^{-1}). The spectra were normalized such that the smallest recorded absorbance was set to 0 and the highest was set to 1 for each spectrum and then the second derivatives (Savitzky & Golay, 1964), with a window of 9), were used for cluster analysis.

Cluster analyses. The unsupervised clustering method principal components analysis (PCA; Jolliffe, 1986) was performed on the spectra to reduce the dimensionality of the multivariate data whilst preserving the variance, prior to discriminant function analysis (DFA). DFA is a supervised technique that discriminates between

Table 1. Details of *M. bovis* spoligotypes studied

Spoligotype	VLA reference	VNTR	Source	Origin	Prevalence in analysis		Designation in Fig. 4
					Ref	Trans	
9	61/6855/98	7555*32.1	Cow	West Glamorgan	*	*	9a
9	61/3558/00	7524*33.1	Cow	Gwent	*	*	9b
9	61/5344/98	7555*32.1	Cow	Dyfed	*	*	9c
9	61/0501/01	7554*33.1	Cow	Wiltshire		*	9d
9	61/1198/01	6554*33.1	Cow	Devon	*	*	9e
9	61/0038/01	6554*33.1	Cow	Cornwall	*		9f
9	61/0293/01	7555*32.1	Cow	Dyfed	*		9g
9	61/5523/90	7554*33.1	Cow	Aberdeenshire	*		9h
9	61/0393/01	7554*33.1	Cow	Northumberland	*		9i
9	61/0276/92	6554*33.1	Cow	Devon	*		9k
9	61/1286/99	7555*32.1	Cow	West Glamorgan	*		9m
10	61/0024/90	7554*33.1	Cow	Kirkcudbright	*	*	
10	61/0430/89	7554*33.1	Roe deer	West Sussex	*	*	
10	61/0658/89	7554*33.1	Red deer	West Sussex	*	*	
10	61/1855/89	7554*33.1	Cow	Somerset	*	*	
10	61/6460/89	7554*33.1	Red deer	West Sussex	*	*	
10	61/7432/88	7554*33.1	deer	Gwynedd	*	*	
10	61/1304/01	6554*33.1	Cow	Devon	*		
10	61/0272/01	7554*33.1	Cow	Hereford and Worcester	*		
10	61/2463/01	7554*33.1	Cow	Oxfordshire	*		
10	61/5548/00	7554*33.1	Cow	Oxfordshire	*		
10	61/6427/89	7554*33.1	Red deer	West Sussex	*		
11	61/0180/01	7554*33.1	Cow	Devon	*	*	
11	61/0209/01	7554*33.1	Cow	Devon	*	*	
11	61/0468/01	7554*33.1	Cow	Dyfed	*	*	
11	61/0800/01	7554*33.1	Cow	Dorset	*	*	
11	61/1249/01	7554*33.1	Cow	Cambridgeshire	*	*	
11	61/1685/96	7554*33.1	Cow	Devon	*	*	
11	61/3295/99	7554*33.1	Cow	Cornwall	*	*	
11	61/4925/90	9554*33.1	Cow	Gloucestershire	*	*	
11	61/5415/00	7554*33.1	Cow	Hereford and Worcester	*	*	
11	61/894/01	7554*33.1	Cow	Somerset		*	
11	61/0894/01	7554*33.1	Cow	Somerset	*		
12	61/0054/01	7454*33.1	Cow	Cornwall	*	*	
12	61/1975/90	7454*33.1	Cow	Lancashire		*	
12	61/2208/01	7454*33.1	Cow	Cornwall	*	*	
12	61/2368/01	7454*33.1	Cow	Cornwall	*	*	
12	61/2433/01	7454*33.1	Cow	Cornwall	*	*	
12	61/5696/99	7454*33.1	Cow	Somerset	*	*	
13	61/0223/00	7353*33.1	Cow	East Sussex	*	*	
13	61/0954/92	7353*33.1	Cow	East Sussex	*	*	
13	61/1659/92	7353*33.1	Badger	West Sussex	*	*	
13	61/1935/89	7356*33.1	Cow	Wiltshire	*	*	
13	61/2265/89	7353*33.1	Cow	East Sussex	*	*	
13	61/3979/00	7353*33.1	Cow	East Sussex	*	*	
13	61/1932/89	7356*33.1	Cow	Cornwall	*		
17	61/0128/01	7555*33.1	Cow	Hereford and Worcester	*	*	17a
17	61/0319/01	7555*33.1	Cow	Avon	*	*	17b
17	61/0359/01	7555*33.1	Cow	Gloucestershire	*	*	17c
17	61/0484/01	7555*33.1	Cow	Shropshire	*	*	17d
17	61/0810/01	7555*33.1	Cow	Dyfed	*	*	17e
17	61/1020/01	7555*33.1	Cow	Staffordshire	*	*	17f
17	61/1943/89	7555*33.1	Cow	Cornwall	*	*	17g

Table 1. cont.

Spoligotype	VLA reference	VNTR	Source	Origin	Prevalence in analysis		Designation in Fig. 4
					Ref	Trans	
17	61/3459/90	7555*33.1	deer	Dorset	*	*	17h
17	61/0351/99	7555*33.1	Cow	Hereford and Worcester	*		17i
17	61/1121/01	7555*33.1	Cow	Wiltshire	*		17k
17	61/0214/01	7555*33.1	Cow	Warwickshire	*		17m
17	61/0522/90	7555*33.1	Badger	Gloucester	*		17n
17	61/0519/90	7555*33.1	Badger	Hereford and Worcester	*		17o
20	61/1637/99	7354*33.1	Cow	Dumfrieshire	*	*	
20	61/2084/01	7554*33.1	Cow	Cornwall	*	*	
20	61/2646/96	7554*33.1	Cow	Gwent	*	*	
20	61/5768/99	7554*33.1	Cow	Kirkcudbright	*	*	
20	61/3781/92	7554*33.1	Cow	Cornwall	*		
20	21/0140/01	7554*33.1	Cow	Cornwall	*		
22	61/0207/01	7524*33.1	Cow	Gwent	*	*	
22	61/0288/01	7524*33.1	Cow	Powys	*	*	
22	61/0477/01	7524*33.1	Cow	Powys	*	*	
22	61/0927/01	7524*33.1	Cow	Hereford and Worcester	*	*	
22	61/0928/91	7524*33.1	Red deer	Gloucestershire	*	*	
22	61/1535/01	7524*33.1	Cow	Avon	*	*	
22	61/1867/96	7524*33.1	Cow	Gwent	*	*	
22	61/4438/98	7524*33.1	Cow	Hereford & Worcester		*	
22	61/3048/98	7524*33.1	Cow	Shropshire	*		
25	61/0157/01	6554*23.1	Cow	Staffordshire	*	*	
25	61/0384/01	6554*23.1	Unknown	Oxfordshire	*	*	
25	61/0681/01	6554*23.1	Cow	Derbyshire	*	*	
25	61/0888/92	6554*23.1	Cow	Leicestershire	*	*	
25	61/1258/01	6554*23.1	Cow	Powys	*	*	
25	61/2145/01	6554*23.1	Cow	Shropshire	*	*	
25	61/2396/01	6554*23.1	Cow	Staffordshire	*	*	
25	61/2880/99	6554*23.1	Cow	Devon	*	*	
25	61/6875/98	6554*23.1	Cow	Staffordshire	*		
25	61/1495/01	6454*23.1	Cow	Staffordshire	*		
25	61/2451/01	6554*23.1	Cow	Clwyd	*		
35	21/0643/01	3354*33.1	Cow	Somerset	*	*	
35	21/0284/01	3354*33.1	Cow	Shropshire	*	*	
35	21/9176/00	3354*33.1	Cow	Devon	*	*	
35	61/0186/01	3354*33.1	Cow	Shropshire	*	*	
35	61/0507/01	3354*33.1	Cow	Gloucestershire	*	*	
35	61/1207/01	3354*33.1	Cow	Hereford & Worcester		*	
35	61/1307/01	3354*33.1	Cow	Shropshire	*	*	
35	61/3274/97	3354*33.1	Cow	Dyfed	*	*	
35	61/4119/96	3354*33.1	Cat	North Yorkshire	*	*	
35	61/2255/90	3534*33.1	Cow	Cumbria	*		

groups on the basis of the retained principal components with *a priori* knowledge of which spectra were replicates. As this process uses information on the machine replicates for each isolate it does not bias the results in any way (MacFie *et al.*, 1978; Windig *et al.*, 1983). Hence the cluster analysis would not bias the clustering of the isolates according to their respective spoligotypes because the *a priori* knowledge is based on the isolate replicates. DFA was programmed to minimize 'within group' variance and maximize 'between group' variance. Finally, hierarchical cluster analysis (HCA) was used to construct a dendrogram from the *a priori* group centres in the PC-DFA space, using scaled Euclidean distances as described by Goodacre *et al.* (1998), and dendrograms were produced using the mean linkage clustering algorithm (Manly, 1994).

RESULTS AND DISCUSSION

Typical transmission-generated FT-IR absorbance spectra of *M. bovis* illustrating examples of spoligotypes 9 and 17 are shown in Fig. 1; the baseline of spoligotype 9 is slightly offset

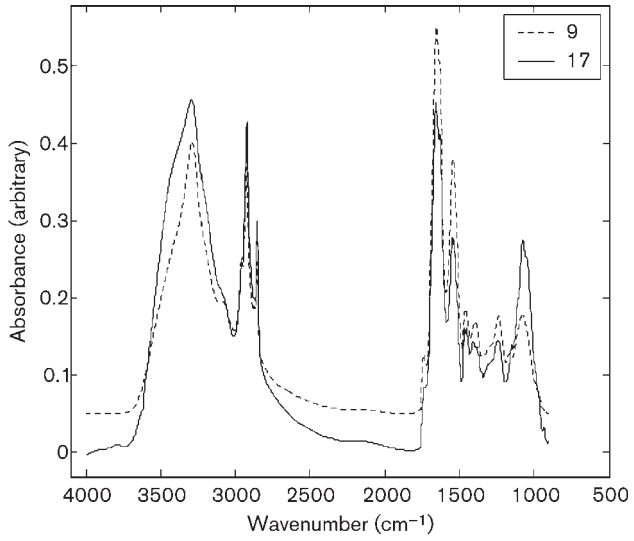

Fig. 1. Typical example of FT-IR spectra of *M. bovis* (spoligotypes 9 and 17; the spectrum representing spoligotype 9 is offset).

to demonstrate the contours of each sample. Due to the complexity and qualitative similarity of these spectra, and indeed all data collected, multivariate statistical analysis was used to discriminate the strains of *M. bovis*.

Due to the different methods of FT-IR used, the data could not be analysed together. This is expected since the transmission-based sampling produces spectra from the absorbance of infrared light during single transmission (i.e. the sample is held on an IR-transparent window and light passes directly through it) and detected using a DTGS detector. By contrast, the reflection-based protocol involves the IR beam bouncing through the bacteria held on an aluminium plate and being reflected off the bottom of the wells back through the sample. This results in absorbance via reflection (two beam passes). In addition, because the sample is not 'infinitely' thin, some diffuse-reflectance also occurs, which is recorded by a MCT detector together with the reflected light. Note that the DTGS and MCT detectors are different: the DTGS operates at room temperature whilst the MCT is liquid N_2 cooled; the MCT has a quicker response time so that spectra are collected in 10 s compared to 60 s for DTGS; finally, the DTGS tends to have a more stable baseline than MCT. While these data could not be analysed together, it is important to assess which produces more reproducible data and which is better at bacterial discrimination.

PC-DFA and HCA were used to produce ordination (scatter) plots and dendrograms respectively. Whilst both the ordination plots and dendrograms were used to investigate the discriminatory ability of FT-IR, for brevity and clarity the dendrograms are predominantly shown for reflection-based FT-IR (Fig. 2) and transmission-based measurements (see Fig. 4).

It is clear from the two analyses that the discrimination of the various spoligotypes is different for reflection (Fig. 2) versus transmission (Fig. 4) measurements. These two figures reflect a series of analyses in which certain bacteria are sequentially removed because they cluster according to spoligotype and are clearly differentiated from the others. The reanalysis of the remaining spectra allows finer (more subtle) discrimination to be observed, and is commonplace when analysing spectroscopic data (Jarvis & Goodacre, 2004; Lopez-Diez & Goodacre, 2004). This process is briefly discussed below.

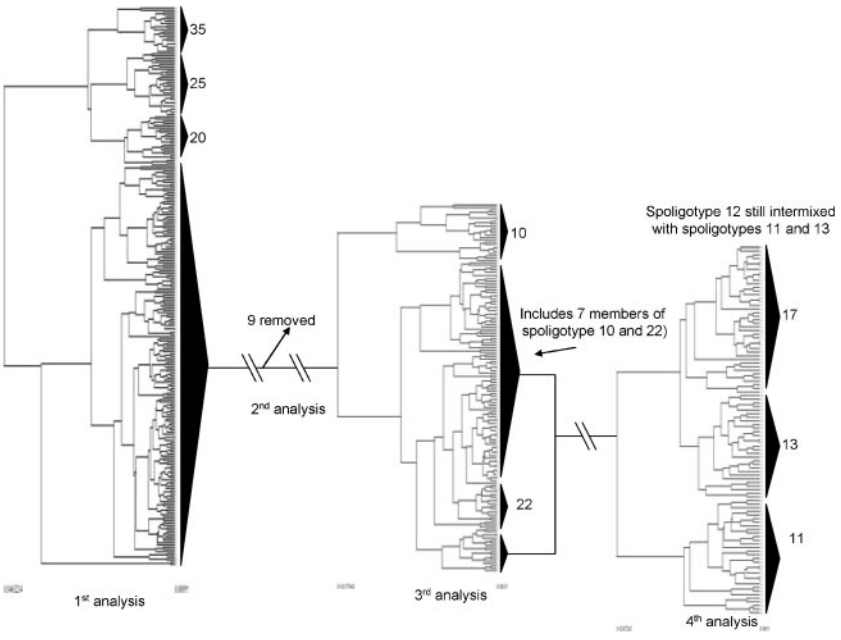

Fig. 2. Composite dendrogram illustrating the hierarchial cluster analysis of the data collected by diffuse reflectance FT-IR.

Reflectance-based FT-IR

The resulting dendrogram for the reflectance FT-IR illustrates the clear discrimination of spoligotypes 20, 25 and 35 (Fig. 2). The PC-DFA ordination plot of the first analysis of the composite dendrogram is shown in Fig. 3(a), to illustrate the clustering of spoligotypes 20, 25 and 35. Fig. 3(b) indicates the mean centres of the *a priori* groups and the circles indicate a 95 % χ^2 confidence region around the point. The overlapping circles are shown for spoligotypes 20, 25 and 35; the circles representing the other spoligotypes are not shown for clarity of the figure. The analysis shows the isolates representing the chosen spoligotypes cluster within the confidence regions. Removal of the three discriminated spoligotypes and the subsequent reanalysis of the data demonstrated the dispersal of spoligotype 9 throughout the HCA space (data not shown). The successive removal of spoligotype 9 and reanalysis of the data illustrates the recovery of spoligotypes 10 and 22. A fourth analysis was then performed on the remaining data following the removal of spoligotypes 10 and 22. In this final dendrogram, the representative samples of spoligotype 17 were recovered together; however, the isolates of spoligotypes 12 were recovered with *M. bovis* spoligotypes 11 and 13 isolates.

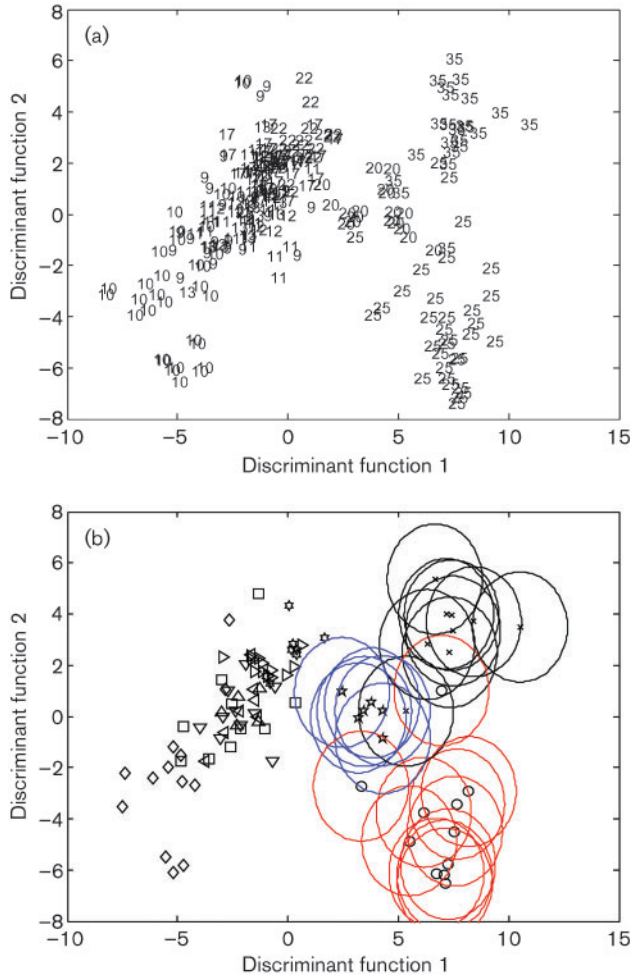

Fig. 3. (a) PC-DFA performed on the data illustrated in the first analysis in Fig. 2. PCs 1–100 (total cumulative variance is 99·94 %) were used by the DFA algorithm with *a priori* knowledge of isolate biological replicates and DFs 1–6 used for HCA. (b) The points indicate the mean centre of each *a priori* group used in the DFA. The circles show the boundary of 95 % confidence region for that particular point. Each point would have the same area representing the 95 % χ^2 confidence region; however, only the circles representing spoligotypes 35 (black), 25 (red) and 20 (blue) are shown, for clarity. The spoligotypes are represented as follows: square, 9; diamond, 10; down triangle, 11; up triangle, 12; left triangle, 13; right triangle, 17; pentagram, 20; hexagram, 22; circle, 25; ×-mark, 35.

Transmission-based FT-IR

The initial analysis of all the data collected by transmission FT-IR clearly shows the discrimination of spoligotypes 10, 25 and 35 (Fig. 4). Following their removal and subsequent reanalysis of the data, it was again evident that the isolates representing spoligotype 9 were recovered throughout the HCA space (data not shown). Spoligotype 9 was therefore removed from the dataset and the resulting dendrogram (third analysis, Fig. 4) illustrated the clear discrimination of spoligotypes 17, 20 and 22. The samples representing spoligotypes 11, 12 and 13 were slightly intermixed, while isolates from 12 and 13 were recovered together and a second branch of spoligotype 13 was observed. Isolates of spoligotype 11 were recovered in two separate branches within the dendrogram. Further analysis of only spoligotypes 12 and 13 illustrates clustering of the isolates from spoligotype 12. Although the representatives of spoligotype 13 were grouped together on the dendrogram, the branching pattern illustrates diffuse clustering in the HCA space (fourth analysis, Fig. 4).

Comparison of the two FT-IR methods

In the initial analysis of both the reflectance and transmission data, three distinct branches were observed according to three spoligotypes. Spoligotypes 25 and 35 were recovered in both analyses but the third branch varied according to the different FT-IR techniques. The relationship between the spoligotypes analysed in this study illustrates spoligotypes 25 and 35 to be closer to the ancestor, in which all the spacers are present (Fig. 5). Spoligotype 10 was resolved for the transmission data whereas spoligotype 20 was initially resolved in the analysis of the reflectance data. The isolates representing spoligotype 9 were intermixed with the samples representing the other spoligotypes in the analysis for both the reflectance and transmission data. The evolutionary history indicates that a number of spoligotypes (10, 11, 12, 13, 17 and 22) evolved after spoligotype 9 (Fig. 5). The remaining spoligotypes (with the exceptions of 12 and 13) were resolved in the third analysis for the transmission data. A fourth analysis was necessary to recover the majority of the reflectance-collected samples according to spoligotype. In both analyses, the isolates representing

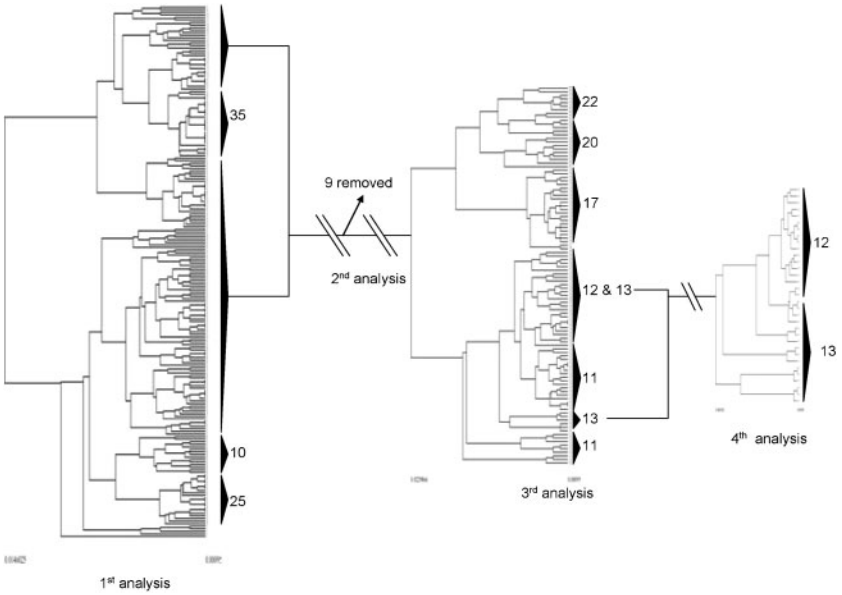

Fig. 4. Composite dendrogram illustrating the hierarchial cluster analysis of the data collected by transmission FT-IR.

spoligotypes 11, 12 and 13 were poorly resolved in the analysis. The representative samples of spoligotypes 12 and 13 (transmission data) were analysed separately; the isolates of spoligotype 12 were recovered in a cluster, but those from spoligotype 13 appeared diffuse in the analysis.

Based on these observations and the complexity of the sequential cluster analysis, we believe that the transmission-generated data are more appropriate for this analysis. Indeed, the lack of appreciable baseline shifts and the simpler optical interrogation of the sample by transmission rather than a reflective measurement would indicate greater spectral reproducibility and support this conclusion.

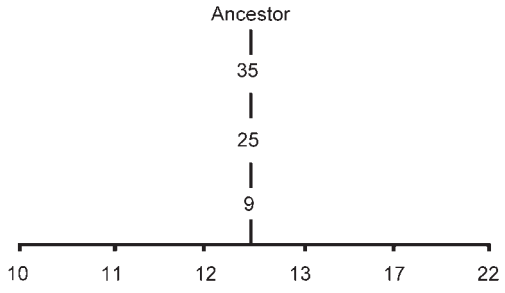

Fig. 5. Shows the relationships between the spoligotype patterns identified by a simple algorithm that assumes spacers are lost and never regained and that change involves the minimum number of deletion events at a time (VLA, unpublished data). A deletion event is any loss of spacers (one or more contiguous spacers). Ancestor refers to a spoligotype pattern with all the spacers present. The relationship of spoligotype pattern 20 is ambiguous and is therefore not shown. This is not a phylogenetic tree of the strains with these spoligotype patterns; however, for this set of strains further studies of informative single nucleotide mutations have shown that the relationships shown here do represent the evolutionary history of these strains.

Analysis of UK dominant *M. bovis* isolates

As reported above, spoligotypes 9 and 17 are the two dominant spoligotypes in the UK, accounting for 65 % of all isolates from cattle (Inwald *et al.*, 2002). We therefore focused the final analysis on just type 9 and 17 isolates. The resulting PC-DFA scores plot is shown in Fig. 6, illustrating the differentiation of five clusters; this was confirmed by the 95 % χ^2 confidence region. It can be clearly seen that a tight cluster of the isolates representing spoligotype 17 was

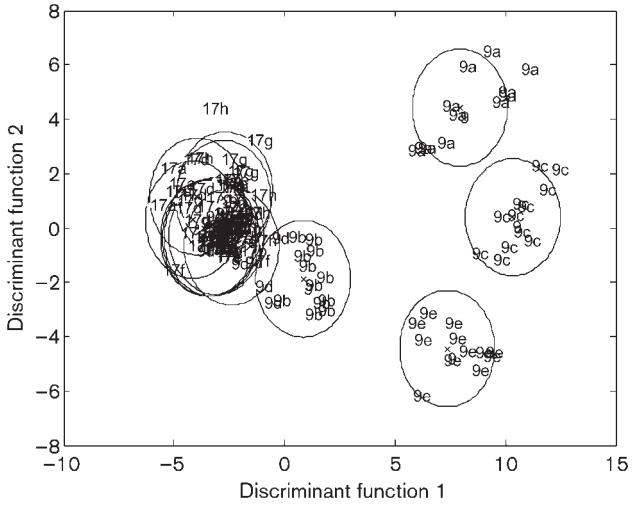

Fig. 6. PC-DFA performed on spoligotypes 9 and 17. PCs 1–20 (total cumulative variance is 99·93 %) were used by the DFA algorithm with *a priori* knowledge of isolate biological replicates and DFs 1–3 used for HCA. The ×-marks indicate the mean centres of each *a priori* group and the circles illustrate the 95 % χ^2 confidence regions. Details of sample isolation are provided in Table 1.

observed, which is perhaps not surprising as these are a single VNTR group (7555*33.1), whilst four of the five isolates representing spoligotype 9 in the analysis were dispersed throughout the PC-DFA space [note that each of the spoligotype 9 isolates was of a different VNTR group (Table 1)]. This clearly reflects what was observed in the dendrograms (Figs 2 and 4) discussed above. A similar pattern was observed in the analysis of the reflectance data (data not shown). The remaining representative of spoligotype 9 (9d, VNTR group 7554*33.1) clustered with the isolates of spoligotype 17 (VNTR group 7555*33.1). This may possibly result from the close phylogenetic relationship of these isolates according to the VNTR phylogeny as seen by us (to be reported elsewhere).

The phylogeny of UK spoligotype 9 strains is complex, with the identification of 22 VNTR types for strains with this spoligotype pattern (Smith *et al.*, 2003). As can be seen in Table 1, the VNTR diversity in the spoligotype 9 strains is greater than for any other grouping. Indeed, the majority of distinct VNTR types of spoligotype 9 are representative of one geographical location, indicating recent clonal expansion of the strains in these locations (Smith *et al.*, 2003). The wide distribution of spoligotype 9 through the PC-DFA space indicates phenotypic differences in these isolates, which reflects genotypic differences observed by VNTR analysis. In comparison, the representatives of spoligotypes 17 are tightly clustered in the PC-DFA space (Fig. 6), reflecting their limited VNTR variation (Table 1). The heterogeneity of spoligotype 9 is highlighted by the inclusion of nine other spoligotypes in this analysis (Figs 2 and 4).

Hence, the analysis of the strains using FT-IR allowed us to generate clusters that showed excellent agreement with clustering performed using genetic markers; this suggests that strains of the same molecular type have conserved phenotypic traits. This is intriguing when considered in the light of our knowledge of the population structure of *M. bovis* in Great Britain, which found that the frequency of recovery of spoligotypes 9 and 17 isolated in the UK cannot be explained by random mutation and drift (Smith *et al.*, 2003). Instead these strains have undergone a clonal expansion, rising to a high frequency in the population either through a favourable mutation or by colonization of a new geographical location (Smith *et al.*, 2003). The results reported here suggest that the success of *M. bovis* clonal groupings is due to favourable mutations in these strains that may affect virulence, transmissibility, or survival outside the host. Our FT-IR data therefore provide us with a unique opportunity to examine the relationship between clonal groups and phenotype.

Conclusions

The data collected by transmission FT-IR illustrated better differentiation of the *M. bovis* isolates according to spoligotypes than that collected by reflectance FT-IR. The phenotypic analysis in this study indicates homogeneity within spoligotypes 10, 17, 20, 22, 25 and 35, while spoligotypes 11, 12 and 13 appeared diffuse and difficult to cluster in the analysis. Tight clustering of molecular types detected by metabolic fingerprinting suggests that these clonal groupings share distinct phenotypic traits that may have a significant role in their success as pathogens. Isolates of spoligotype 9 appear very heterogeneous relative to other spoligotypes, an observation that was supported by VNTR analysis and reflects the wide variation within this group. Therefore this study describes the first use of FT-IR for the discrimination of *M. bovis* strains. As well as being a robust and high-throughput method that produces strain clusters reflecting those generated by genetic methods, FT-IR has the added benefit of providing the first insights into phenotype–genotype links in *M. bovis* clones. This study therefore sets the baseline for a more in-depth phenotypic analysis of *M. bovis* molecular types.

ACKNOWLEDGEMENTS

We acknowledge the Department of Environment Food and Rural Affairs (Defra) for funding this work. We are grateful to Noel Smith for kindly providing Fig. 5 and Roger Jarvis for his assistance and useful comments with the data analysis.

REFERENCES

Aranaz, A., Liebana, E., Mateos, A. & 8 other authors (1996). Spacer oligonucleotide typing of *Mycobacterium bovis* strains from cattle and other animals: a tool for studying epidemiology of tuberculosis. *J Clin Microbiol* **34**, 2734–2740.

Clifton-Hadley, R. S., Wilesmith, J. W., Richards, M. S., Upton, A. & Johnston, S. (1995). The occurance of *Mycobacterium bovis* infection in cattle in and around an area subject to extensive badger (*Meles meles*) control. *Epidemiol Infect* **114**, 179–193.

Delahay, R. J., Cheeseman, C. L. & Clifton-Hadley, R. S. (2001). Wildlife disease reservoirs: the epidemiology of *Mycobacterium bovis* infection in the European badger (*Meles meles*) and other British mammals. *Tuberculosis* **81**, 43–49.

Gibson, A. L., Hewinson, G., Goodchild, T., Watt, B., Story, A., Inwald, J. & Drobniewski, F. A. (2004). Molecular epidemiology of disease due to *Mycobacterium bovis* in humans in the United Kingdom. *J Clin Microbiol* **42**, 431–434.

Goodacre, R., Timmins, E. M., Burton, R., Kaderbhai, N., Woodward, A. M., Kell, D. B. & Rooney, P. J. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology* **144**, 1157–1170.

Goyal, M., Saunders, N. A., VanEmbden, J. D. A., Young, D. B. & Shaw, R. J. (1997). Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS*6110* restriction fragment length polymorphism. *J Clin Microbiol* **35**, 647–651.

Griffin, J. M., Williams, D. H., Kelly, G. E., Clegg, T. A., O'Boyle, I., Collins, J. D. & More, S. J. (2005). The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Prev Vet Med* **67**, 237–266.

Harrigan, G. G., LaPlante, R. H., Cosma, G. N., Cockerell, G., Goodacre, R., Maddox, J. F., Luyendyk, J. P., Ganey, P. E. & Roth, R. A. (2004). Applications of high-throughput Fourier-transform infrared spectroscopy in toxicology studies: contribution to a study

on the development of an animal model for idiosyncratic toxicity. *Toxicol Lett* **146**, 197–205.

Inwald, A., Hinds, J., Dale, J., Palmer, S., Butcher, P., Hewinson, R. G. & Gordon, S. V. (2002). Microarray-based comparative genomics: genome plasticity in *Mycobacterium bovis*. *Comp Funct Genomics* **3**, 342–344.

Jarvis, R. M. & Goodacre, R. (2004). Discrimination of bacteria using surface-enhanced Raman spectroscopy. *Anal Chem* **76**, 40–47.

Jolliffe, I. (1986). *Principal Component Analysis*. New York: Springer-Verlag.

Kamerbeek, J., Schouls, L., Kolk, A. & 8 other authors (1997). Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol* **35**, 907–914.

Lopez-Diez, E. C. & Goodacre, R. (2004). Characterization of microorganisms using UV resonance Raman spectroscopy and chemometrics. *Anal Chem* **76**, 585–591.

MacFie, H. J. H., Gutteridge, C. S. & Norris, J. R. (1978). Use of canonical variates in differentiation of bacteria by pyrolysis gas-liquid chromatography. *J Gen Microbiol* **104**, 67–74.

Manly, B. F. J. (1994). *Multivariate Statistical Methods*. London: Chapman & Hall.

Maquelin, K., Kirschner, C., Choo-Smith, L. P. & 7 other authors (2003). Prospective study of the performance of vibrational spectroscopies for rapid identification of bacterial and fungal pathogens recovered from blood cultures. *J Clin Microbiol* **41**, 324–329.

Naumann, D., Helm, D. & Labischinski, H. (1991). Microbiological characterizations by FT-IR spectroscopy. *Nature* **351**, 81–82.

Roring, S., Scott, A. N., Hewinson, R. G., Neill, S. D. & Skuce, R. A. (2004). Evaluation of variable number tandem repeat (VNTR) loci in molecular typing of *Mycobacterium bovis* isolates from Ireland. *Vet Microbiol* **101**, 65–73.

Savitzky, A. & Golay, M. J. E. (1964). Smoothing and differentiation of data by simpli®ed least squares procedures. *Anal Chem* **36**, 1627–1633.

Smith, N. H., Dale, J., Inwald, J., Palmer, S., Gordon, S. V., Hewinson, R. G. & Smith, J. M. (2003). The population structure of *Mycobacterium bovis* in Great Britain: clonal expansion. *Proc Natl Acad Sci U S A* **100**, 15271–15275.

Timmins, E. M., Howell, S. A., Alsberg, B. K., Noble, W. C. & Goodacre, R. (1998). Rapid differentiation of closely related *Candida* species and strains by pyrolysis mass spectrometry and Fourier transform-infrared spectroscopy. *J Clin Microbiol* **36**, 367–374.

Winder, C. L., Carr, E., Goodacre, R. & Seviour, R. (2004). The rapid identification of *Acinetobacter* species using Fourier transform infrared spectroscopy. *J Appl Microbiol* **96**, 328–339.

Windig, W., Haverkamp, J. & Kistemaker, P. G. (1983). Interpretation of sets of pyrolysis mass spectra by discriminant analysis and graphical rotation. *Anal Chem* **55**, 81–88.