The population structure of *Mycobacterium bovis* in Great Britain: Clonal expansion

Noel H. Smith*, James Dale†, Jacqueline Inwald†, Si Palmer†, Stephen V. Gordon†, R. Glyn Hewinson†, and John Maynard Smith*‡

*Centre for the Study of Evolution, University of Sussex, Falmer, Brighton BN1 9QL, United Kingdom; and †TB Research Group, Veterinary Laboratories Agency (VLA), Weybridge, New Haw, Addlestone, Surrey KT15 3NB, United Kingdom

Contributed by John Maynard Smith, October 10, 2003

We have analyzed 11,500 isolates of *Mycobacterium bovis* (the cause of tuberculosis in cattle and other mammals) isolated in Great Britain (England, Wales and Scotland)] and characterized by spoligotype. Genetic exchange between cells is rare or absent in strains of the *Mycobacterium tuberculosis* complex so that, by using spoligotypes, it is possible to recognize "clones" with a recent common ancestor. The distribution of variable numbers of tandem repeats types in the most common clone in the data set is incompatible with random mutation and drift. The most plausible explanation is a series of "clonal expansions," and this interpretation is supported by the geographical distribution of different genotypes. We suggest that the clonal expansion of a genotype is caused either by the spread of a favorable mutation, together with all other genes present in the ancestral cell in which the mutation occurred, or by the invasion of a novel geographical region by a limited number of genotypes. A similar pattern is observed in *M. tuberculosis* (the main cause of tuberculosis in humans). The significance of clonal expansion in other bacteria that have recombination is discussed.

variable numbers of tandem repeats | spoligotype | bacterial evolution | bovine tuberculosis

This article analyzes patterns of migration and selection in strains of *Mycobacterium bovis* from Great Britain collected from cattle and badgers and characterized by two molecular typing methods, spoligotyping and polymorphism at variable numbers of tandem repeats (VNTR) loci (1, 2).

M. bovis, the causative agent of bovine tuberculosis, is responsible for severe worldwide economic losses and can cause disease in both humans and a wide variety of domesticated and wild animals. In contrast to most other European countries, the control of *M. bovis* in cattle has proved problematic in Great Britain and Ireland (3). The European badger has been implicated as an important wildlife reservoir of *M. bovis* and a significant reason for the failure of the British cattle testing and control program that began in the 1950s to eradicate bovine tuberculosis (4).

The most common molecular typing method applied to isolates of *M. bovis* in Great Britain is spoligotyping (5). This method identifies polymorphism in the presence or absence of spacer units in the direct repeat (DR) region (2). The DR region is composed of multiple, virtually identical 36-bp repeats interspersed with unique DNA spacer sequences of similar size (direct variant repeats or DVR units). The positions of each unique spacer sequence within the DR region have been shown to be highly conserved (6, 7). In spoligotyping, the DR region is amplified by PCR, followed by hybridization of the labeled PCR product to immobilized spacer oligonucleotides (2). Spoligotype patterns are recognized by the presence or absence of hybridization signal from DVRs.

Members of the *Mycobacterium tuberculosis* complex are unusual in that there seems to be no exchange of genetic material between individuals (8–11). Spoligotype patterns should therefore contain sufficient phylogenetic signal to reconstruct recent evolutionary events with some confidence. Strains bearing the same spoligotype pattern are assumed to be a set of individuals derived relatively recently by clonal replication from a single ancestral cell. The ability to recognize such clones is a powerful tool for analyzing population structure and history. However, spoligotype patterns change rather slowly. We have therefore generated data for a second genetic system, polymorphism at VNTR loci, which change rapidly, and so can be highly variable within a group of cells bearing the same spoligotype pattern. The VNTR typing system identifies polymorphism in the number of repeats at tandemly arranged repetitive DNA sequences, similar to eukaryotic minisatellites (12). Changes in the number of repeats at a given locus occur frequently, in both directions, so that phylogenetic trees are hard to construct and are not easily rooted. However, strains with the same spoligotype are typically descended from a single cell whose VNTR type can often be deduced with some confidence.

The questions we attempt to answer are as follows: (*i*) Has natural selection affected the pattern of genetic variation? (*ii*) What can we say about the geographical spread of *M. bovis* in Great Britain?

We report simulations of the evolution of VNTR types in strains with spoligotype pattern SB0140, the most common in our sample. We show that the distribution of VNTR types is not compatible with a simple neutral drift model; the number of different genotypes in the simulation is much higher and the number of individuals with the commonest mutant genotypes is much lower. We conclude that the observed data can be explained by a "clonal expansion" model. This model is supported by the geographical localization of spoligotypes and VNTR types. Finally, we discuss the relevance of clonal expansion to the evolution of other bacterial species in which recombination does occur.

Materials and Methods

Spoligotype Patterns. As part of the bovine tuberculosis control program, all cultured isolates of *M. bovis* from Great Britain are routinely submitted to the Veterinary Laboratories Agency (VLA, Weybridge, U.K.) for spoligotype analysis according to the method of Kamerbeek *et al.* (2). The VLA spoligotype database presently holds typing information on over 20,000 *M. bovis* strains isolated from 1975 to 2003 (with 95% of data for strains isolated since 1997). We analyzed a subset of this database containing information on the frequency and geographical origin of >11,500 strains isolated between 1975 and 2001. The most frequently recovered spoligotype patterns in this subset are SB0140 and SB0263, representing 34% and 29% of the isolates, respectively. Authoritative names for spoligotype patterns were obtained from www.mbovis.org.

Abbreviation: VNTR, variable numbers of tandem repeats.

‡To whom correspondence should be addressed at: The John Maynard Smith Building, University of Sussex, Falmer, East Sussex BN1 9RH, United Kingdom. E-mail: bafo2@central.susx.ac.uk.

POPULATION BIOLOGY

Table 1. VNTR types and frequency of recovery for spoligotypes SB0140 and SB0263

Spoligotype SB0140 (n = 1169)		Spoligotype SB0263 (n = 640)	
Frequency	VNTR	Frequency	VNTR
358	6-5-5-4	595	7-5-5-5
197	7-5-5-4	17	7-4-5-5
198	7-5-5-5	10	7-5-5-4
145	8-5-5-5	5	7-6-5-5
114	7-5-2-4	4	6-5-5-5
85	7-5-4-5	4	8-5-5-5
35	3-3-5-5	3	4-5-5-5
13	5-5-5-4	1	5-5-5-5
6	7-4-5-4	1	7-3-5-5
3	6-5-2-4		
3	7-4-5-5		
2	6-4-5-4		
1	3-5-2-4		
1	3-5-5-4		
1	5-6-5-4		
1	6-3-5-4		
1	6-5-5-5		
1	6-5-6-4		
1	7-3-5-4		
1	7-3-5-5		
1	7-5-4-6		
1	8-3-5-5		

VNTR Analysis. We have analyzed 1,169 strains with spoligotype pattern SB0140 and 640 strains with spoligotype pattern SB0263 for the six VNTR loci originally identified by Frothingham and Meeker-O'Connell (1). However, we have limited our analysis to the exact tandem repeat ETR-A to -D loci; the ETR-E and -F loci were virtually monomorphic within the dataset. The VNTR genotype of a strain, representing the number of repeat elements at each locus, is presented as a series of four integers, between 1 and 8, separated by hyphens.

Results

Simulation of VNTR Evolution. The first step toward simulating the evolution of strains with a specific spoligotype pattern is to estimate the mean number of mutations, *dbar*, between the common ancestor of the clone and isolates in the sample. This presents no difficulty for most spoligotypes. For example, Table 1 shows the frequency of VNTR types identified in 640 isolates of strains with pattern SB0263. Of these, 595 are of a single VNTR type, and all others differ from this type at one locus only. If we make the natural assumption that this is the ancestral type, all other VNTR types can be generated by change at only one locus. Thus, $dbar = 45/640 = 0.07$.

However, difficulties arise with the VNTR types identified in strains bearing pattern SB0140 (Table 1). The phylogeny is complex, and there are many more VNTR types identified: 22 in total (Table 1). We therefore choose as ancestor that VNTR type that minimizes Σnd, summed over all isolates, where n is the number of isolates with a given genotype and d is the genetic distance to the potential root being tested. Rather surprisingly, for strains with spoligotype pattern SB0140, this method suggests that the most common VNTR type is not the ancestor. Thus, for the most common VNTR genotype, 6-5-5-4 (358 isolates), Σnd = 1,736 and $dbar = 1.48$; whereas, if we take genotype 7-5-5-4, with only 197 isolates, as the root, then Σnd = 1,501, and $dbar = 1.28$. The identification of VNTR type 7-5-5-4 as ancestral, rather than 6-5-5-4, is supported by the fact that strains of type 7-5-5-4 are widely distributed whereas strains with VNTR type 6-5-5-4 are almost exclusively located in Cornwall (see below).

In Table 2, the observed frequency of isolates of different VNTR types is compared with the frequencies expected in the absence of selection. The observed data are for the 1,169 isolates with spoligotype pattern SB0140 that have been characterized for VNTR type. The first row is based on the assumption that the ancestral VNTR type is that which minimizes the mean genetic distance to other isolates (7-5-5-4). We regard this as the most likely assumption, but for comparison we give values assuming that the most frequently recovered VNTR type (6-5-5-4) is ancestral. These data are compared with the results of two simulations, for a "star" and a "branching" phylogeny. In both cases, a population of ≈10,000 individuals was simulated, from which a random sample of 1,169 was drawn after g generations. The mutation rate per individual per generation (m) was chosen by trial and error to give approximately the observed mean genetic distance (1.28) between ancestor and members of the sample. Varying m while maintaining mg constant showed no great differences. The mutation process in the simulations was chosen so as to resemble the observed pattern; the relative frequencies of changes at each locus and the change in the number of repeats (from +1 to −4) were estimated by analyzing all single-locus changes in VNTR genotype for strains with spoligotype pattern SB0140 and SB0263.

Star Phylogeny. For the star phylogeny, each individual is derived independently from the original ancestor. This is easy to simulate: in each generation, a fraction m of the population is mutated and is then transmitted to the next generation; there is no replication.

The differences shown in Table *2* between the simulated and observed results are striking: (*i*) The number of different genotypes in the simulation is much higher. (*ii*) The number of

Table 2. Observed and simulated VNTR variation for spoligotype SB0140

			No. of isolates in		No. of genotypes	
	Root*	*dbar*	Ancestral genotype	Five largest mutant genotypes	1–2 isolates	>2 isolates
Observed†	7-5-5-4	1.28	197	900	11	11
	6-5-5-4	1.48	358	739	11	11
Simulation‡						
Star phylogeny		1.3	252	218	95	81
Branching phylogeny		1.28	164	268	79	78

*Sample of 1,169 isolates with spoligotype SB0140 from Great Britain.

†7-5-5-4 is the preferred root because it minimizes the mean genetic distance, *dbar*, to all other isolates. 6-5-5-4 is the most common VNTR type.

‡Simulations are a sample of 1,169 from a simulated population of 8,000–10,000, with the number of generations and mutation rate chosen to give the observed value of *dbar*.

Table 3. Observed and simulated VNTR variation for spoligotype SB0263

	No. of genotypes with		No. of individuals with		
	1–2 isolates	>2 isolates	No mutations	Two most common mutant classes	
Observed	2	7	595	17	10
Simulated*	6	8	597	8	6

*Analysis of a sample of 640 from a simulated population of 10,000: branching phylogeny, 100 generations, 7 mutations per generation.

individuals with the most common mutant genotypes is much lower.

It seemed possible that the discrepancies arose from the implausible assumption of a star phylogeny. Thus, if, typically, two individuals in the sample have a common ancestor more recent than the clonal ancestor, we might expect the same mutant genotypes to recur more frequently in the final population and fewer different genotypes to be present. We therefore repeated the simulation, assuming that, in each generation, after mutation, each individual had an equal chance of dying without descendants, or of leaving two descendants identical to itself.

Branching Phylogeny. In simulating a branching phylogeny, we started with a single cell. In each generation, each cell was replicated, and the daughter cells were mutated. This continued for 13 generations, giving a final population of 8,192; as before, the mutation rate was chosen to give the observed value of *dbar*. To our surprise, introducing death and replication, and hence a branching phylogeny, made rather little difference. The two differences between the observed and simulated results listed above remain the same (Table 2).

A second difficulty arises because we have assumed that a group of cells with the same spoligotype are monophyletic, but this need not be the case. The same spacer or linked group of spacers can be lost more than once (7). Recombination is also a possibility. We are satisfied that it is rare, but we cannot prove that it never occurs. However, this finding does not affect the argument for clonal expansion. The assumption that the 1,169 isolates are descended from a common ancestor with spoligotype pattern SB0140 is not necessary for our conclusions. It is sufficient that their common ancestor (and they certainly had one) was recent enough that *dbar* is of the order 1.5 or less. What the existence of spoligotypes does is to make it easy to pick out a group of closely related isolates. For completeness, we assigned each VNTR profile to one of two subsets, assuming two separate roots, 7-5-5-4 or 6-5-5-4. Both subsets showed a discrepancy between observed and simulated distributions as great as that for the complete sample.

We conclude that the observed frequency of VNTR types recovered for strains with spoligotype SB0140 cannot be explained by random mutation and drift. It could be suggested that certain configurations of repeat units at VNTR loci are resistant to change, and this finding may explain the reduced number of genotypes observed. However, the VNTR profiles of strains of both *M. bovis* or of strains from the *M. tuberculosis* complex show no evidence of overrepresentation of the VNTR profiles seen here (13–15). Furthermore, the stability of VNTR profile cannot explain the geographical localization of VNTR types discussed below.

The apparent explanation for the differences between the observed and simulated data are that, from time to time, a particular VNTR genotype increases dramatically in frequency, relative to others with the same spoligotype. We refer to this phenomenon as "clonal expansion."

Simulation of the SB0263 Clone. Table 3 shows the results of simulating the evolution of isolates bearing spoligotype pattern SB0263. The agreement between observation and simulation is close. The reason is that this clone is younger than the strains with spoligotype pattern SB0140. Assuming that the rate of mutation in each lineage is roughly constant, *dbar* can be used as an estimate of the relative age of the clone. Thus, the clone with spoligotype SB0263 (*dbar* = 0.07) is ≈1/20th the age of strains bearing spoligotype pattern SB0140 (*dbar* = 1.28). If there have been any cases of clonal expansion, they have been rare and would have originated in an individual with the ancestral genotype, and so would be hard to detect.

Geographical Distribution of Clones. The geographical distribution of the most frequently recovered spoligotypes in Great Britain is shown in Table 4. Most spoligotypes show geographical localization. For example, >90% of isolates bearing spoligotype SB0145 or spoligotype SB0275 are recovered from Cornwall.

In Table 5, we show the geographical distribution of the seven most common VNTR types of strains with spoligotype pattern SB0140. Most of the VNTR types are localized to one area. For

Table 5. Number of isolates and primary county of origin of the seven most common VNTR types of spoligotype SB0140

VNTR type	No. of isolates*	Principal county	Percentage recovered from principal county
6-5-5-4	352	Cornwall	72
7-5-5-4	185	Wiltshire	41
7-5-5-5	169	Dyfed	88
8-5-5-5	143	Cornwall	68
7-5-2-4	114	Powys	80
7-5-4-5	83	Dorset	95
3-3-5-5	33	Avon	94

*Number of isolates for which county of origin is recorded.

Table 4. Distribution by county of the most common spoligotypes of *M. bovis* in Great Britain

Spoligotype	Cornwall	Devon and Somerset	Wiltshire	Dorset	Gloucestershire	Hereford and Worcestershire	Dyfed	Gwent	Powys
SB0140	531	464	203	81	56	57	345	25	113
SB0263	18	49	92	7	502	473	47	12	71
SB0673	2	6	2	0	121	124	1	207	50
SB0274	17	633	3	7	4	3	3	2	1
SB0134	6	1	0	0	57	59	1	1	1
SB0271	93	5	0	0	1	1	0	0	1
SB0145	117	3	0	1	0	0	0	0	0
SB0275	169	11	0	0	2	0	0	0	0
SB0130	3	40	1	7	2	1	0	0	1

example, the most common VNTR type of strains with pattern 6-5-5-4 is almost exclusively located in Devon and Cornwall (98% of isolates). We suggest that geographical localization of VNTR types and spoligotypes provides evidence of the recent clonal expansion of these strains.

It is notable that the most likely ancestral VNTR type (7-5-5-4) of strains with spoligotype pattern SB0140 has been identified in 21 counties in Great Britain and is more geographically dispersed than VNTR type 6-5-5-4 (data not shown). If we assume that an ancestral VNTR type will be more geographically dispersed than a more recently derived VNTR type, then this observation supports the suggestion that the VNTR type of the ancestral cell bearing pattern SB0140 was 7-5-5-4 rather than the more frequently recovered 6-5-5-4.

Discussion

Clonal Expansion of *M. bovis*. We conclude that the observed frequency of VNTR types recovered for strains of *M. bovis* isolated in Britain and bearing spoligotype SB0140 cannot be explained by random mutation and drift without selection. The apparent explanation is that, from time to time, a particular VNTR genotype increases dramatically in frequency. We refer to this phenomenon as clonal expansion. There are two possible mechanisms, both of which probably occur: (*i*) First is selection, which need not be selection on either the spoligotype or the VNTR genotype itself. If a selectively favorable mutation occurs anywhere in the genome, it will increase the frequency of the spoligotype and the VNTR type of the strain bearing the selectively favorable mutation. (*ii*) Second is ecological opportunity. Suppose that a new host species, or a new geographical region, is invaded. Invasion may be by one or a few cells, and hence by one or a few genotypes, which, in the absence of competition, will become common. It is possible that the severe bottleneck in the population of *M. bovis* in cattle in the 1950s and 1960s caused by the cattle testing and eradication program may have been followed by the clonal expansion of spoligotypes and VNTR types in regions where the incidence of the disease was severely reduced but not eliminated (16). After the event, it may be difficult to distinguish between these two mechanisms although invasion of a new host species is unlikely to be the explanation for the distribution of VNTR types in *M. bovis* in Great Britain.

Geographical Localization of Spoligotypes. The geographical localization of spoligotypes and VNTR types is supporting evidence for the clonal expansion model of *M. bovis* in Great Britain. We suggest that the clone bearing spoligotype pattern SB0263, for example, originated somewhere in Gloucestershire and has recently undergone a clonal expansion in the adjacent counties of Herefordshire and Worcestershire (Table 4). The data also suggest that strains with pattern SB0134 have undergone a recent clonal expansion in the same region as strains with pattern SB0263; that strains with pattern SB0274 have done so in Devon and Somerset; and that strains with spoligotype patterns SB0271, SB0275, and SB0145 have undergone selective clonal expansion in Cornwall. If this interpretation of the present distributions is correct, it does not imply that the change in spoligotype itself was selectively favored: it seems more likely that a new favorable mutation sometimes occurs in a rare spoligotype, resulting in a local increase in frequency. If so, there are presumably many favorable mutants that do not happen to occur in a rare spoligotype, and many new spoligotypes that arise, but are lost before they become common: such events will not show up in the data. It may be that we are overinterpreting the geographic data and that the distributions in Table 4 can be explained by genetic drift. However, this result seems unlikely if we bear in mind the data in Table 2, and the difficulty in interpreting these data without assuming frequent clonal expansions.

If data on VNTR types are taken into account in conjunction with geographical distribution data, this strengthens the case for clonal expansion. In particular, there are cases in which a genotype almost unknown elsewhere has become the dominant type in a particular region. For example, (*i*) in Dorset, 81/103 isolates with spoligotype pattern SB0140 have VNTR type 7-5-4-5, a VNTR type recorded in only two other isolates in the whole data set. Only 29 isolates belonging to other spoligotypes have been reported from Dorset. (*ii*) In Powys, 113/227 isolates belong to pattern SB0140 and VNTR type 7-5-2-4, a type recorded only once in the rest of the data set. Isolates belonging to other spoligotypes are not uncommon.

The interpretation of such distributions depends on how *M. bovis* spreads geographically. A difficulty arises because there are at least two common hosts, cattle and badgers (4, 17). The range of genotypes infecting the two hosts in a given region are similar: it is likely that transmission between the two host species occurs in both directions. Two mechanisms, not mutually exclusive, of geographic spread of a particular genotype are plausible: (*i*) The first is "diffusion" in the local badger or cattle population. Individual animals do not travel far, but members of a group infect one another so that the spread of a particular genotype is likely to occur slowly, on a wide front. (*ii*) The second is long distance transfer in an infected cow. Cattle are not infrequently transported long distances. A clear picture will emerge only when genetic and geographic data for isolates from both cattle and badgers are considered together, alongside data on the movement of cattle.

***M. tuberculosis*.** The very closely related human pathogen *M. tuberculosis* also shows evidence of clonal expansion. Among major clades of *M. tuberculosis*, identified by several molecular techniques (13), four clonal families represent 35% of isolates (18). Clonal families of *M. tuberculosis* also exhibit geographical localization: in the Beijing area of China, 92% of isolates of *M. tuberculosis* belong to the "Beijing" clonal family although the mobility of the human host has tended to disperse these strains worldwide (19). It has been suggested that the implementation of bacillus Calmette–Guérin (BCG) vaccination and use of anti-TB drugs has selected certain genotypes (20). Furthermore, the selective advantage of the Beijing clonal family may be an elevated mutation rate that facilitates the rapid accumulation of antibiotic resistance (21).

Clonal Expansion in Other Bacteria. Does a similar phenomenon occur in other bacteria, or does it require the complete absence of recombination? There is no reason why frequent favorable mutations should be peculiar to *M. bovis*, although these mutations are easier to recognize because of the absence of recombination. In pathogenic bacteria, favorable mutations may be frequent because of the "arms race" between host and parasite (22). If so, we would expect clonal expansion to be common in all pathogens. However, a complete absence of recombination is not necessary. Given the highly localized nature of recombination in prokaryotes, a favorable mutation will cause a parallel increase in frequency of all alleles present in the cell in which it occurred (23). Given recombination, the favorable mutation can transfer to other cell lineages, thus limiting the duration of the selective sweep.

Selective sweeps in bacteria, and their role in speciation, have recently been discussed by Cohan (24, 25). In eukaryotes, a species is a population of interbreeding individuals whose cohesion and distinctness is ensured because each new individual has two parents and is genetically intermediate between them. Two distinct species cannot exist unless there is some mating barrier, or they are geographically isolated. If they are to coexist in the same region, they must occupy different ecological niches (i.e., their populations must be regulated by different resources).

In bacteria, "species" of this kind do not exist: bacterial recombination does not ensure cohesion into distinct groups. Cohan argues that bacterial species are in effect ecotypes, which often do not correspond to named species. Bacterial species typically consist of many ecotypes (24, 25).

Essentially, we agree with this picture but would emphasize that there are clusters of similar bacteria at different scales, from stable ecotypes like the host-adapted races of pathogens to recently arisen and short-lived clones, with no unique ecological niche. There is no "right" way of classifying bacteria into species although the temptation to do so is hard to resist.

After the discovery of bacterial recombination in the laboratory, those working on populations of free-living bacteria emphasized that such populations are "clonal" (26, 27). Two kinds of evidence were quoted. First, polymorphic loci are usually in linkage disequilibrium. Second, multilocus "electrophoretic types" are widely distributed in space and time (e.g., refs. 26 and 28). These observations are correct, but it was not fully appreciated that they are compatible with bacterial-type recombination (29).

Recently, a more detailed picture of the population structure of bacteria has emerged from multilocus sequence typing (MLST), the sequencing of a number of loci spaced round the chromosome, from large samples, usually of pathogenic bacteria (30). Such data are relevant to the present discussion for two reasons. First, they provide a measure of r/m, the relative probabilities per generation that a locus will be changed by recombination (r) or by mutation (m). The ratio turns out to be highly variable: r/m is in the range 5–10 in *meningococcus* and *pneumococcus*, and $\approx 1/10$ in *Staphyllococcus aureus* (31). More immediately relevant, MLST has revealed a clonal structure [e.g., Feil *et al*. (32)]. Samples usually fall into a number of "clonal complexes," consisting of a set of isolates identical at all sequenced loci, or differing by only a few genetic changes (mutations or recombinatons). Such clonal complexes are thought to arise from the hitch-hiking effects of a single favorable mutation and gradually merge into the general population by recombination. In *S. aureus*, the sample consisted of three subsamples, collected, respectively, from individuals without symptoms, with disease acquired in the community, and with hospital-acquired disease (32). Members of a given clonal complex were equally likely to occur in any one of the three samples. Thus, there is no evidence in this case for an ecotypic difference between the groups. However, it is possible that the clonal complexes were generated by selection, for example favoring novel antigenic types. There is evidence in other pathogenic bacteria for the spread of particular ecotypes; for *Neisseria meningitidis* serogroup A, the clonal complexes (genoclouds) associated with global pandemics in humans may represent ecotypes selected by the human immune system (33).

As yet, we do not know in *M. bovis* how often clonal complexes arise in response to particular geographical opportunities and how often they are caused by selection for specific mutations. But geographical and other data are accumulating that should make it possible to answer this question.

We thank Kristin Kremer and Peter Durr for critically reading the manuscript. We acknowledge and thank Paul Upton and the SVS for the work involved in generating and maintaining the spoligotype database. N.H.S. was supported by Grant SE3020 from the Department for Enviroment, Food, and Rural Affairs (United Kingdom).

1. Frothingham, R. & Meeker-O'Connell, W. A. (1998) *Microbiology* **144,** 1189–1196.
2. Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M. & van Embden, J. (1997) *J. Clin. Microbiol.* **35,** 907–914.
3. Phillips, C. (2003) *Res. Vet. Sci.* **74,** 1–15.
4. Delahay, R. J., De Leeuw, A. N., Barlow, A. M., Clifton-Hadley, R. S. & Cheeseman, C. L. (2002) *Vet. J.* **164,** 90–105.
5. Durr, P. A., Clifton-Hadley, R. S. & Hewinson, R. G. (2000) *Rev. Sci. Tech.* **19,** 689–701.
6. van Embden, J. D., van Gorkom, T., Kremer, K., Jansen, R., van Der Zeijst, B. A. & Schouls, L. M. (2000) *J. Bacteriol.* **182,** 2393–2401.
7. Warren, R. M., Streicher, E. M., Sampson, S. L., van der Spuy, G. D., Richardson, M., Nguyen, D., Behr, M. A., Victor, T. C. & van Helden, P. D. (2002) *J. Clin. Microbiol.* **40,** 4457–4465.
8. Supply, P., Warren, R. M., Baänuls, A. L., Lesjean, S., Van Der Spuy, G. D., Lewis, L. A., Tibayrenc, M., Van Helden, P. D. & Locht, C. (2003) *Mol. Microbiol.* **47,** 529–538.
9. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., III, *et al*. (1998) *Nature* **393,** 537–544.
10. Gutacker, M. M., Smoot, J. C., Migliaccio, C. A., Ricklefs, S. M., Hua, S., Cousins, D. V., Graviss, E. A., Shashkina, E., Kreiswirth, B. N. & Musser, J. M. (2002) *Genetics* **162,** 1533–1543.
11. Sreevatsan, S., Pan, X., Stockbauer, K. E., Connell, N. D., Kreiswirth, B. N., Whittam, T. S. & Musser, J. M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 9869–9874.
12. Frothingham, R. (1995) *J. Clin. Microbiol.* **33,** 840–844.
13. Kremer, K., van Soolingen, D., Frothingham, R., Haas, W. H., Hermans, P. W., Martâin, C., Palittapongarnpim, P., Plikaytis, B. B., Riley, L. W., Yakrus, M. A., *et al*. (1999) *J. Clin. Microbiol.* **37,** 2607–2618.
14. Skuce, R. A., McCorry, T. P., McCarroll, J. F., Roring, S. M., Scott, A. N., Brittain, D., Hughes, S. L., Hewinson, R. G. & Neill, S. D. (2002) *Microbiology* **148,** 519–528.
15. Supply, P., Lesjean, S., Savine, E., Kremer, K., van Soolingen, D. & Locht, C. (2001) *J. Clin. Microbiol.* **39,** 3563–3571.
16. Krebs, J. R., Anderson, R. M., Clutton-Brock, T., Donnelly, C. A., Frost, S., Morrison, W. I., Woodroffe, R. & Young, D. (1998) *Science* **279,** 817–818.
17. Gallagher, J. & Clifton-Hadley, R. S. (2000) *Res. Vet. Sci.* **69,** 203–217.
18. Filliol, I., Driscoll, J. R., Van Soolingen, D., Kreiswirth, B. N., Kremer, K., Valâetudie, G., Anh, D. D., Barlow, R., Banerjee, D., Bifani, P. J., *et al*. (2002) *Emerg. Infect. Dis.* **8,** 1347–1349.
19. Glynn, J. R., Whiteley, J., Bifani, P. J., Kremer, K. & van Soolingen, D. (2002) *Emerg. Infect. Dis.* **8,** 843–849.
20. van Soolingen, D., Qian, L., de Haas, P. E., Douglas, J. T., Traore, H., Portaels, F., Qing, H. Z., Enkhsaikan, D., Nymadawa, P. & van Embden, J. D. (1995) *J. Clin. Microbiol.* **33,** 3234–3238.
21. Rad, M. E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D. & Gicquel, B. (2003) *Emerg. Infect. Dis.* **9,** 838–845.
22. Hamilton, W. D., Axelrod, R. & Tanese, R. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 3566–3573.
23. Atwood, K. C. & Ryan, F. J. (1951) *Proc. Natl. Acad. Sci. USA* **37,** 146–155.
24. Cohan, F. M. (2002) *Genetica* **116,** 359–370.
25. Cohan, F. M. (2002) *Annu. Rev. Microbiol.* **56,** 457–487.
26. Selander, R. K. & Levin, B. R. (1980) *Science* **210,** 545–547.
27. Whittam, T. S., Ochman, H. & Selander, R. K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1751–1755.
28. Ochman, H., Whittam, T. S., Caugant, D. A. & Selander, R. K. (1983) *J. Gen. Microbiol.* **129,** 2715–2726.
29. Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 4384–4388.
30. Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., *et al*. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 3140–3145.
31. Feil, E. J., Holmes, E. C., Bessen, D. E., Chan, M. S., Day, N. P., Enright, M. C., Goldstein, R., Hood, D. W., Kalia, A., Moore, C. E., *et al*. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 182–187.
32. Feil, E. J., Cooper, J. E., Grundmann, H., Robinson, D. A., Enright, M. C., Berendt, T., Peacock, S. J., Smith, J. M., Murphy, M., Spratt, B. G., *et al*. (2003) *J. Bacteriol.* **185,** 3307–3316.
33. Zhu, P., van der Ende, A., Falush, D., Brieske, N., Morelli, G., Linz, B., Popovic, T., Schuurman, I. G., Adegbola, R. A., Zurth, K., *et al*. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5234–5239.