JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 1999, p. 3872–3878 Vol. 37, No. 12
0095-1137/99/$04.00+0

Evolution and Clonal Traits of *Mycobacterium tuberculosis* Complex in Guinea-Bissau

GUNILLA KÄLLENIUS,[1]* TUIJA KOIVULA,[1,2] SOLOMON GHEBREMICHAEL,[1] SVEN E. HOFFNER,[1] RENÉE NORBERG,[1] ERIKA SVENSSON,[1] FRANCISCO DIAS,[2] BRITT-INGER MARKLUND,[3] AND STEFAN B. SVENSON[1,4]

Department of Bacteriology, Swedish Institute for Infectious Disease Control, S-17182 Solna,[1] Department of Microbiology, Umeå University, 90 187 Umeå,[3] Department of Bacteriology, Biomedicum, Swedish University for Agricultural Sciences, S-75123 Uppsala, Sweden,[4] and Laboratorio Nacional de Saude Publica, Bissau, Guinea-Bissau[2]

Received 25 November 1998/Returned for modification 12 January 1999/Accepted 23 July 1999

Two hundred twenty-nine consecutive isolates of *Mycobacterium tuberculosis* complex from patients with pulmonary tuberculosis in Guinea-Bissau, which is located in West Africa, were analyzed for clonal origin by biochemical typing and DNA fingerprinting. By using four biochemical tests (resistance to thiophene-2-carboxylic acid hydrazide, niacin production, nitrate reductase test, and pyrazinamidase test), the isolates could be assigned to five different biovars. The characteristics of four strains conformed fully with the biochemical criteria for *M. bovis*, while those of 85 isolates agreed with the biochemical criteria for *M. tuberculosis*. The remaining 140 isolates could be allocated into one of three biovars (biovars 2 to 4) representing a spectrum between the classical bovine (biovar 1) and human (biovar 5) tubercle bacilli. By using two genotyping methods, restriction fragment length polymorphism analysis with IS*6110* (IS*6110* RFLP analysis) and spoligotyping, the isolates could be separated into three groups (groups A to C) of the *M. tuberculosis* complex. Group A ($n = 95$), which contained the majority of classical human *M. tuberculosis* isolates, had large numbers of copies of IS*6110* elements (mean number of copies, 9) and a distinctive spoligotyping pattern that lacked spacers 33 to 36. Isolates of the major group, group B ($n = 119$), had fewer IS*6110* copies (mean copy number, 5) and a spoligotyping pattern that lacked spacers 7 to 9 and 39 and mainly comprised isolates of biovars 1 to 4. Group C isolates ($n = 15$) had one to three IS*6110* copies, had a spoligotyping pattern that lacked spacers 29 to 34, and represented biovar 3 to 5 isolates. Four isolates whose biochemical characteristics conformed with those of *M. bovis* clustered with the group B isolates and had spoligotype patterns that differed from those previously reported for *M. bovis*, in that they possessed spacers 40 to 43. Interestingly, isolates of group B and, to a certain extent, also isolates of group C showed a high degree of variability in biochemical traits, despite genotypic identity in terms of IS*6110* RFLP and spoligotype patterns. We hypothesize that isolates of groups B and C have their evolutionary origin in West Africa, while group A isolates are of European descent.

Tuberculosis (TB) is globally a major cause of morbidity and mortality, with most cases occurring in developing countries (20, 27). According to the World Health Organization, one-third of the world's population is infected with organisms of the *Mycobacterium tuberculosis* complex, with about 10 million cases of active TB disease reported each year, leading to 3 million deaths annually (6). An increased morbidity and mortality in TB patients is associated with the ongoing human immunodeficiency virus (HIV) pandemic. Guinea-Bissau has one of the highest incidences of TB in the world, with the annual incidence estimated to be 150/100,000 population. The current epidemic of HIV type 2 (HIV-2) infections (21) and the now emerging HIV-1 epidemic in Guinea-Bissau may contribute to a still higher incidence of TB.

It has long been observed that West African *M. tuberculosis* strains are biochemically more heterogeneous than European strains. Prat and colleagues (23) speculate that the biochemical heterogeneity observed among African strains represents a continuous spectrum linking the classical human and bovine variants of *M. tuberculosis*. David and coworkers (4), on the other hand, found that *M. africanum* strains tended to form two clusters on the basis of biochemical traits: one related to the human type and one related to the bovine type of *M. tuberculosis* complex. In a previous epidemiological study (16) we subtyped 56 *M. tuberculosis* complex strains isolated from patients with pulmonary TB in Guinea-Bissau into five biovars. On the basis of our biochemical methods, we identified five biovars representing a spectrum of strains ranging from classical human *M. tuberculosis* strains to classical *M. bovis* strains. Two biovars corresponded to the so-called African I variant, and one biovar corresponded to the African II variant described by Collins et al. (2).

To study the population structure of *M. tuberculosis* complex isolates in Guinea-Bissau, we have investigated 229 *M. tuberculosis* isolates by the same biochemical methods used in our previous study, as well as by genetic molecular fingerprinting, using the insertion sequence IS*6110* and the spacer regions within the direct repeat (DR) locus of *M. tuberculosis* as targets. The insertion sequence IS*6110* belongs to the larger IS*3* family of insertion elements and is regarded to be specific for the *M. tuberculosis* complex. The DR locus of the *M. tuberculosis* complex genome contains multiple, highly conserved 36-bp DRs separated by 35- to 41-bp variable spacer sequences (9). On the basis of the results presented here, we hypothesize

* Corresponding author. Mailing address: Swedish Institute for Infectious Disease Control, S-17182 Solna, Sweden. Phone: 46 8 4572430. Fax: 46 8 301797. E-mail: gunilla.kallenius@smi.ki.se.

that the majority of *M. tuberculosis* complex isolates belong to a unique family of strains that originated and evolved in West Africa.

MATERIALS AND METHODS

Patients. Approximately 900 consecutive patients with suspected TB (and without known previous treatment for TB) were examined from 1989 to 1994 as part of an epidemiological study of TB in Guinea-Bissau (5, 15, 16, 19, 21). The ages and sexes of the patients were recorded, and during 1992 and 1993, testing for HIV-1 and HIV-2 was performed as described previously (21). Patients with positive cultures for *M. tuberculosis* complex (n = 229) were included in this study.

***M. tuberculosis* complex isolates.** Sputum samples were collected from all patients by a registered nurse and were transported at 4°C to the laboratory. Acid-fast microscopy was performed after staining by the Ziehl-Neelsen method. Before culture, the sputum samples were decontaminated of nonmycobacterial microorganisms by the sodium lauryl sulfate method (10). A 0.5-ml aliquot of the homogenized specimen was then inoculated into both conventional Löwenstein-Jensen egg medium (LJ) and LJ supplemented with 0.6% pyruvate. The samples were incubated at 37°C and were examined weekly for 7 weeks. Growth of mycobacteria was confirmed by microscopic observation of acid-fast bacilli. In total, 229 strains of *M. tuberculosis* complex were isolated from these patients. Fifty-six of these isolates were also part of a previous study (16).

Characterization of isolates. All isolates were further characterized by macroscopic and microscopic appearance and growth characteristics and were biochemically characterized as described earlier (15, 18). Resistance to thiophene-2-carboxylic acid hydrazide (TCH; 5 mg/liter) was determined by radiometric respirometry (BACTEC system; Becton Dickinson, Sparks, Md.). Detection of niacin was performed as described by Wayne (33), and the nitrate reductase test and pyrazinamidase test were performed as described by Kent and Kubica (18). Testing for susceptibility to the drugs ethambutol, isoniazid, rifampin, and streptomycin was performed by using the BACTEC system (15, 25). All isolates were also tested with a nucleic acid probe specific for *M. tuberculosis* complex (Accuprobe system; Gen-Probe, San Diego, Calif.).

IS*6110* RFLP analyses. A standardized method for restriction fragment length polymorphism (RFLP) analysis with IS*6110* (IS*6110* RFLP analysis) (29) was used. Genomic *M. tuberculosis* DNA was extracted and digested with the restriction endonuclease *Pvu*II, and Southern blotting was performed after separation of the restriction fragments by electrophoresis. Hybridization was performed with a 245-bp PCR fragment of the IS*6110* sequence as a probe. The probe was nonradioactively labelled with peroxidase and was subsequently visualized by the use of an enhanced chemiluminescence kit (Amersham International plc, Little Chalfont, United Kingdom).

The IS*6110* DNA patterns were analyzed with Gelcompar software, version 3.1b (Applied Math, Kortrijk, Belgium) after radiographs had been scanned at 74.8 dots/cm (190 dots/in; HP Scanjet IIcx/T; Hewlett-Packard, Camas, Wash.). On the basis of the molecular sizes of the hybridizing fragments and the number of IS*6110* copies for each isolate, the fingerprint patterns were compared by the unweighted pair-group method of arithmetic averaging by using the Dice coefficient. Dendrograms were constructed to show the degree of relatedness among strains by using a previously described algorithm (31), and similarity matrices were generated to visualize the relatedness between the banding patterns of all isolates.

Spoligotyping. Spoligotyping relied on the amplification of the polymorphic DR region (17) to obtain hybridizations patterns of the amplified DNA by using multiple synthetic spacer oligonucleotides which are covalently bound to a membrane (Isogen Bioscience BV, Maarssen, The Netherlands).

RESULTS

M. tuberculosis complex isolates from 229 consecutive patients with bacteriologically verified pulmonary TB were characterized by biochemical tests, drug susceptibility testing, IS*6110* RFLP analysis, and spoligotyping.

Patient characteristics. All except eight of the patients were native to Guinea-Bissau. The eight patients from other countries came from the neighboring West African countries of Guinea Conakry (five patients), Senegal (two patients), and The Gambia (one patient). The mean age of the patients was 39 years, with 58% being male (mean age, 41 years) and 42% being female (mean age, 37 years). Of the 229 patients, 154 were tested for HIV; 118 were HIV negative, 3 (1.9%) were HIV-1 positive, 29 (18.8%) were HIV-2 positive, and 4 (2.6%) were infected with both HIV-1 and HIV-2.

Characterization of isolates. All isolates had typical macro- and microscopic appearances upon acid-fast staining and were positive by the *M. tuberculosis* complex-specific nucleic acid probe assay. Fifteen isolates were resistant to any of four major antituberculous drugs (six were resistant to streptomycin, eight were resistant to isoniazid, and one was resistant to streptomycin and isoniazid) (5). By using four biochemical tests (resistance to TCH, niacin production, nitrate reductase test, and pyrazinamidase test) the *M. tuberculosis* complex isolates could be assigned to five different biovars (Table 1). The characteristics of 4 strains conformed fully with the biochemical criteria for *M. bovis* (sensitivity to TCH and negativity by the other tests), while those of 85 isolates agreed with the biochemical criteria for *M. tuberculosis* (resistance to TCH and positivity by the other tests). The remaining 140 isolates could be allocated to one of three biovars representing a spectrum between the classical human and bovine tubercle bacilli (Table 1).

TABLE 1. Biovars of 229 isolates of *M. tuberculosis* complex

Biovar	Result of the following tests[a]:				No. of isolates
	TCH	NO_3	Niacin	PZ	
1[b]	S	−	−	−	4
2	S	−	−	+	20
3	S	−	+	+	69
4	S	+	+	+	51
5[c]	R	+	+	+	85

[a] TCH, resistance to TCH (5 mg/liter); S, susceptible; R, resistant; NO_3, presence of nitrate reductase; Niacin, production of niacin; PZ, presence of pyrazinamidase.
[b] Classical *M. bovis*.
[c] Classical human *M. tuberculosis*.

Characteristics of IS*6110* RFLP patterns. The 229 *M. tuberculosis* complex isolates were analyzed by an RFLP assay with the insertion sequence IS*6110* as the target. All isolates had RFLP patterns that contained one or more bands that hybridized with the IS*6110* probe (Fig. 1), with the number of IS*6110* insertions ranging between 1 and 17 copies. The Dice coefficient of similarity for all pairwise comparisons of the isolates in the whole collection ranged from 0 to 100%; 147 (64%) of the isolates belonged to clusters of strains with identical and/or closely related banding patterns (defined as a coefficient of similarity of >90%), while 36% had less similar patterns. Altogether, 119 distinct banding patterns were found; 33 patterns were shared by two or more isolates, constituting 37 different clusters (defined as at least two isolates with identical RFLP patterns). The different clusters each comprised 2 to 33 isolates (Fig. 1). The largest cluster comprised isolates with a four-band pattern (Fig. 1). Twelve isolates, which showed only a single band, could be separated into two groups of four and eight isolates, respectively, on the basis of slightly different band migration patterns.

As visualized in Fig. 1, the different IS*6110* RFLP patterns divided the isolates into three distinct groups (designated groups A to C). Within group A, there was one major family of 35 strains (on the basis of related banding patterns, defined as a similarity coefficient of >65%), which could be further separated into seven clusters (clusters A:2 to A:8). Another 10 clusters (clusters A:1 and A:9 to A:17), each containing between 2 and 10 strains, could also be defined.

The other major group that was identified (group B), comprised 120 (52%) isolates which contained small numbers of IS*6110* insertions (three to eight copies; mean, five copies) and which had IS*6110* RFLP patterns which were highly distinct (similarity coefficient, <20%) from those for the isolates of group A. Group B contained a family of 87 strains (38% of all isolates) which were further separated into 12 related clusters

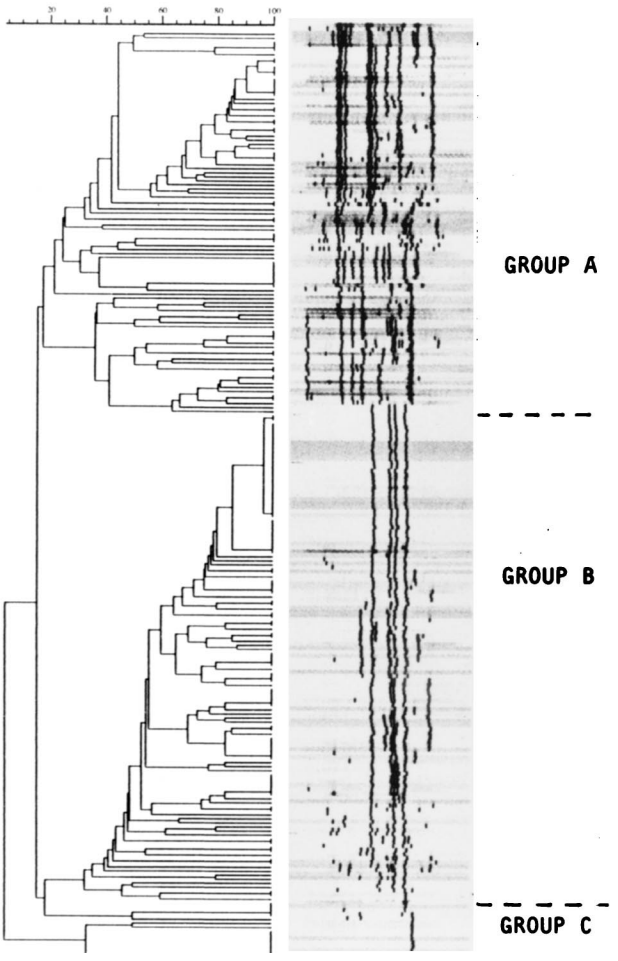

FIG. 1. IS*6110* banding patterns and similarity matrices for 229 *M. tuberculosis* complex isolates from Guinea-Bissau. Banding patterns are ordered by similarity. The corresponding dendrograms are to the left of the panels. The positions of the bands in each lane are adjusted (normalized) so that band positions for all strains are comparable. Scale depicts similarity coefficients (which are defined elsewhere [29]). Groups A to C are as defined in the Results.

(clusters B:1 to B:12). Cluster B:5, which was the largest single cluster (32 isolates), showed a distinctive four-band IS*6110* RFLP pattern.

Group C consisted of 15 isolates, each containing one to three copies of IS*6110*, including two clusters (clusters C:1 and C:2) of isolates which each contained a single copy of IS*6110*. Cluster C:1 consisted of four isolates in which the IS*6110* element was located on an approximately 1.35-kb *Pvu*II fragment, while cluster C:2 consisted of eight isolates in which the IS*6110* element was located on a slightly larger *Pvu*II fragment of approximately 1.46 kb. Three additional isolates separated into group C: one isolate contained two copies of IS*6110*, with one located on a fragment common to that for the C:1 cluster, while the other two isolates contained three copies of IS*6110*, one common to that for the C:2 cluster, and an additional two copies. Interestingly, the IS*6110* insertion in cluster C:1 isolates was located at the same position as one of the insertions within group B isolates.

There was a greater conservation of IS*6110* insertion patterns within group B than within group A. Thus, 84 (69%) of group B isolates were in clusters, whereas 51 (54%) of isolates in group A were in clusters. Additionally, the mean cluster size was larger in group B than in group A (mean, 4.6 versus 3.0 isolates per cluster, respectively).

TABLE 2. Correlation between spoligotype, IS*6110* RFLP pattern, and biovar classification for the 229 *M. tuberculosis* complex isolates

Spoligotype	No. of isolates by:							
	IS*6110* RFLP group			Biovar				
	A	B	C	1	2	3	4	5
sptA[a] (n = 88)	88					1	13	74
sptB[b] (n = 117)	1	116		4	19	62	28	4
sptC[c] (n = 15)			15			4	9	2
Others (n = 9)	6	3			1	2	1	5
Total (n = 229)	95	119	15	4	20	69	51	85

[a] Lacking spacer 33 to 36.
[b] Lacking spacers 7 to 9 and 39.
[c] Lacking spacers 34 and 29 to 32.

Spoligotyping. Spoligotyping was performed with all 229 isolates. We identified 73 different spoligotype patterns, and these were used to classify the isolates into three main groups, designated sptA, sptB, and sptC (Table 2). The sptA group, consisting of 88 isolates, lacked spacer 33 to 36, with 38 of these isolates also lacking spacer 21 to 24. The largest group (117 isolates), sptB, typically lacked spacers 7 to 9 and 39, with 30% of these strains also lacking between 1 and 16 additional spacers. The smallest group, sptC, consisting of 15 isolates, typically lacked spacer 34, with either spacers 29 to 33 or spacers 29 to 32 also absent. In addition to the three baseline patterns described above, many individual isolates further lacked one or more spacers, forming clusters of strains with identical spoligotype patterns.

Correlation between IS*6110* RFLP pattern and spoligotype. Although the fingerprints obtained by spoligotyping were less polymorphic than those obtained by probing with the IS*6110* probe, there was an overall correlation between spoligotype and IS*6110* RFLP pattern, even for strains with small numbers of IS*6110* copies, confirming the genetic relatedness of the isolates identified by RFLP analysis.

All but five of the RFLP group A isolates exhibited the basic sptA spoligotype pattern (Table 2). Only one strain (strain IH-167), which fell into group A, had an sptB spoligotype pattern. This was a strain with a unique six-band IS*6110* RFLP pattern and the only biovar 3 isolate that was sorted into group A on the basis of its RFLP pattern. All except four of the group B strains had the basic sptB spoligotype pattern, which lacked spacers 7 to 9 and 39. The four exceptional isolates possessed one or three spacers 7 to 9, while one of these isolates (isolate IH-133) additionally possessed spacer 39. The 15 group C isolates, which comprised the 12 isolates with one IS*6110* copy, all had the same basic spoligotype pattern (pattern sptC).

Correlation between IS*6110* copy number and biovar. When the RFLP patterns of the individual isolates were compared to the results of the biochemical tests, a significant correlation was observed between IS*6110* copy number and biovar classification (Fig. 2). Thus, the group of classical human *M. tuberculosis* isolates (biovar 5) had a significantly ($P < 0.01$ by the Wilcoxon rank sum test) larger number of IS*6110* copies (mean copy number, 9; median copy number, 10), than the isolates that belonged to the intermediary biovars, biovars 2 to 4 (mean copy number, 5; median copy number, 5). The four *M. bovis* isolates carried three (one isolate), four (two isolates), and five (one isolate) copies of IS*6110* (data not shown).

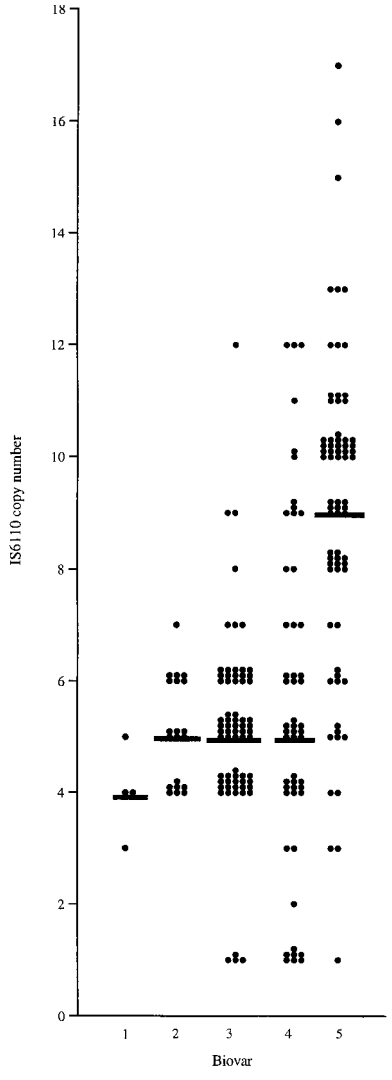

FIG. 2. Number of IS*6110* copies present in isolates of five different biovars of the *M. tuberculosis* complex. Differences in numbers of IS*6110* copies between biovar 5 (classical human *M. tuberculosis*) and biovars 2, 3, and 4 were significant ($P < 0.01$ by the Wilcoxon rank sum test).

Correlation between IS*6110* RFLP pattern and biovar. With a few exceptions, the biovar 5 isolates had the group A RFLP pattern and formed a distinct group. A total of 78 of the 85 biovar 5 isolates were in group A, while only 14 of the 51 biovar 4 isolates, 2 of the biovar 3 isolates, and none of biovar 1 and 2 isolates were in group A (Table 2). Additionally, many of the biovar 5 isolates in group A were highly related; 59 biovar 5 isolates belonged to two families of isolates with highly related banding patterns (similarity coefficient, >65%). In comparison to biovar 5, most of the isolates in biovars 1 to 4 had the main group B RFLP pattern.

Interestingly, however, isolates of the different biovars were more or less randomly distributed at the cluster level (similarity coefficient, >90%). Thus, isolates of the individual biovars did not segregate into specific individual RFLP clusters. For example, the largest cluster, cluster B:5 (32 isolates), with a typical four-band pattern, comprised isolates of all biovars, including one isolate of biovar 5 (Fig. 3). The isolates in group C, which comprised the isolates with two one-band clusters, were of biovars 3, 4, and 5 (Table 2).

***M. bovis* isolates.** The four strains that were biochemically classified as *M. bovis* all belonged to group B. They had between three and five IS*6110* insertions, but intriguingly, none of the insertions was located on the 1.9-kb *Pvu*II restriction fragment that is typical for most *M. bovis* and *M. bovis* BCG strains. Interestingly, one IS*6110* insertion location was common to all four *M. bovis* isolates: the IS*6110* sequence was located on a fragment identical in size to the single band of cluster C:1. Additionally, all *M. bovis* isolates possessed spacers 40 to 43, which, together with spacer 39, have been reported to be lacking in *M. bovis* strains (17).

Correlation between type of strain and patient characteristics. There was no correlation between age, sex, or positivity for HIV with any particular type of strain. Of the eight isolates from patients from neighboring countries, one biovar 2 isolate from a patient from The Gambia and one biovar 3 isolate from a patient from Guinea Conakry clustered with the group B isolates. The six other isolates were of biovar 5 and clustered with the group A isolates.

DISCUSSION

As in a previous study (16), we have found a high degree of biochemical heterogeneity within strains of the *M. tuberculosis* complex isolated from Guinea-Bissau. In order to characterize this heterogeneity further at a genetic level, we subtyped isolates by DNA fingerprinting using the insertion element IS*6110* and spoligotyping. Using these techniques, we were able to classify *M. tuberculosis* complex isolates into three groups. Group A isolates contained large numbers of IS*6110* elements and a basic spoligotype pattern typical of classical human *M. tuberculosis* strains. Group B mainly comprised the intermediate biovars and the classical *M. bovis* strains, which are recognizable by the presence of few IS*6110* copies per genome and a distinctive spoligotype pattern. In contrast, group C comprised a few isolates that mainly possessed a single IS*6110* element. These DNA fingerprint results conform with recent observations from a study of *M. africanum* strains from Sierra Leone and Uganda (12). This study demonstrated that strains which had biochemical traits resembling those of *M. bovis* (Africanum I type) had IS*6110* fingerprints that exhibited fewer bands, resulting in large clusters of strains with identical banding patterns, whereas strains that were biochemically more closely related to *M. tuberculosis* (Africanum II subtype) had a large number of fragments, with many individual banding patterns being observed.

European strains of *M. tuberculosis* are usually described as classical, are distinguishable by a large number of copies of IS*6110* elements, and frequently lack the spoligotyping spacers 33 to 36 (8, 17). Interestingly, isolates of group A from Guinea-Bissau resembled European *M. tuberculosis* isolates in terms of biochemical traits, RFLP patterns, and spoligotypes when they were compared with isolates in an international database of the IS*6110* patterns of *M. tuberculosis* complex isolates and also with strains from South and Central America (28a). Guinea-Bissau is a small country whose population has a rather high degree of mobility within the country but little contact with the populations of neighboring African countries. Historically, it is a former colony of Portugal, which is a country with a high incidence of TB caused by the classical human *M. tuberculosis* type. Even today, contacts between the populations of Guinea-Bissau and Portugal are probably still more numerous than contacts between the populations of Guinea-Bissau and other African countries. We therefore speculate that the group A *M. tuberculosis* strains, to which the majority of the biovar 5 isolates (i.e., the classical human variants) belonged, may have

Isolate	Spoligotype	IS6110 cluster	No of IS6110 copies	Biovar
G-3/92	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	5
G-14/92	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-69	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-168	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-302	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-14	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
IH-179	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX··XXX	B:5	4	4
IH-101	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
IH-174	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	4
IH-365	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX····X	B:5	4	3
G-1/89	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
G-28/89	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
G-30/89	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-203	XXXX·····XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	4
IH-319	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-34	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-38	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX··XXX	B:5	4	3
IH-229	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
G-99/91	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	4
G-115/91	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
G-23/90	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
G-34/90	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
G-38/90	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
G-126/90	X·XXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	4
SN-359	X··XXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	4
IH-128	XXXXXX···XXXXXXXXXXXXXXXXXXXXXX·XXXXX·XXXX	B:5	4	3
IH-59	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
SN-124	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
IH-70	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	2
IH-79	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
SN-11	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	3
SN-68	XXXXXX···XXXXXXXXXXXXXXXXXXXXXXXXXXXX·XXXX	B:5	4	1

FIG. 3. Schematic representation of spoligotypes of 32 *M. tuberculosis* isolates belonging to cluster B:5. The isolates are sorted in the same order in which they were sorted by their IS*6110* RFLP patterns in Fig. 1. X, positive hybridization signals; dots, lack of hybridization.

their origin within Europe, in particular, Portugal. Indeed, some of the isolates of group A had RFLP patterns that were identical to those of isolates from TB patients in Portugal (22).

Isolates of group B had few copies of IS*6110* (mean copy number, 5) and highly conserved RFLP patterns as well as a homogeneous basic spoligotype pattern, which typically lacked spacers 7 to 9 and 39. This spoligotype pattern forms a baseline for all group B isolates. These common features indicate that the group B isolates are highly related to each other. Despite the genotypic similarities among these strains, the group B isolates showed a high degree of variability in terms of biochemical traits. Most of the larger clusters (on the basis of the IS*6110* patterns) within group B contained isolates of different biovars. For example, the largest cluster (cluster B:5) of 32 isolates, presenting a typical four-band pattern, comprised isolates of all biovars, even including one isolate of biovar 5. Thus, despite the genetic stability in terms of the IS*6110* fingerprints and the DR locus, the isolates varied in several phenotypic traits, which therefore could be used to further subdivide the different clusters determined by the DNA fingerprinting methods. One could speculate that the diversity in biochemical traits among these isolates reflects the ongoing evolution of *M. tuberculosis* complex in Guinea-Bissau or West Africa from classical *M. bovis*-like variants to more classical *M. tuberculosis* variants. This would mean, however, that the evolution of these properties is more rapid than the genetic evolution of IS*6110* and the DR locus. Alternatively, the strains of biovars 1 to 4 may more readily switch on and off genes for, for example, niacin production during the adaptation to hosts of several species (and hence broaden their host tropism). Since only certain combinations of biochemical traits were found, such on-off switches for genes do not appear to occur randomly but, rather, occur as a stepwise turning on and off of genes in a defined sequence. For example, 69 isolates were positive for the production of niacin and negative for the production of nitrate reductase, but no isolate produced nitrate reductase but did not produce niacin. Recently, the pyrazinamidase gene (*pncA*) was cloned and a single point mutation in the gene was found to be unique to *M. bovis* (24). We sequenced the *pncA* genes of five cluster B:5 isolates, each representing one the five different biovars. Interestingly, none of them contained the *M. bovis*-specific mutation (data not shown).

Virtually all group B isolates contained spacers 33 to 36. These spacers were derived from the sequence of the BCG genome (17), supporting the relatedness of the group B isolates to *M. bovis*. It is generally supposed that *M. bovis* strains are more virulent for cattle, while classical *M. tuberculosis* strains are thought to be more virulent for humans. In Guinea-Bissau there is a high incidence of tuberculosis among cattle and goats. So far we have analyzed only one strain from a goat, with this isolate being of biovar 4 and showing an RFLP pattern and spoligotype identical to those of the four human biovar 4 isolates.

The four isolates in this study that were phenotypically characterized as *M. bovis* clustered with the group B isolates in terms of IS*6110* RFLP and spoligotype patterns. Hence, they

all had three to five copies of the IS*6110* fragment, they all lacked the spacers 7 to 9 and 39, and they all contained spacers 33 to 36. *M. bovis* strains typically have small numbers of copies of the IS*6110* insertion sequence (31, 32), while human and cattle *M. bovis* isolates in general harbor only a single copy of IS*6110* (3, 11, 28), and it is usually located in the same position in the genome, on a characteristic 1.9-kb *Pvu*II restriction fragment. In a report on the IS*6110* RFLP patterns of 24 human *M. bovis* strains from The Netherlands, all strains had less than six copies, and one-third of the strains had only a single copy (31). By contrast, isolates from other animals appear to carry more IS*6110* copies. In a study by van Soolingen et al. (30), the majority of isolates from cattle harbored a single IS*6110* element, whereas the vast majority (29 of 34) of *M. bovis* isolates from a wide variety of other animals contained multiple IS*6110* elements. During a study of Swedish *M. bovis* isolates from deer, we found that all isolates contained seven copies of the IS*6110* element, yielding a highly specific RFLP pattern (28). Additionally, a study from Spain (11) demonstrated that *M. bovis* strains isolated from 23 goats carried six to eight copies of IS*6110*, while most isolates from cattle carried only a single copy of this element. Interestingly, the four *M. bovis* isolates in this study all had the sptB spoligotype pattern typical of the group B strains, which means that they all possessed spacers 40 to 43, which together with spacer 39 are reported to be lacking in all *M. bovis* and *M. bovis* BCG strains (17).

Cave and coworkers (1) suggested that the IS*6110* insertion sequence originally appeared as a single copy in a progenitor of the *M. tuberculosis* complex and was retained as a single copy in *M. bovis* and then replicated to produce the multiple copies present in *M. tuberculosis*. They also suggested that these multiple insertional events may account for some of the phenotypic differences between *M. tuberculosis* and *M. bovis*. Our finding of a correlation between the phenotypic variants of strains of the *M. tuberculosis* complex and the number of IS*6110* insertion elements supports such a hypothesis. Fomukong and coworkers (7) analyzed selected *M. tuberculosis* complex strains from Malaysia, Tanzania, and Oman as well as *M. bovis* isolates and *M. bovis* BCG strains, all of which carried a single copy of the insertion sequence, and suggested that in these organisms the IS*6110* element is defective in transposition and that the loss of transposability may have occurred at an early stage in the evolution of the *M. tuberculosis* complex. The progenitor would therefore have had a single copy of the insertion sequence; the multicopy *M. tuberculosis* strains would therefore have evolved by replicative transposition of the element. It may be speculated that the small number of copies of the IS*6110* fragment in the group B and C strains in this study reflects a similar defect in transposition of IS*6110*. Hermans and coworkers (14), however, found that the insertion sequence sequence element of *M. bovis* BCG was not defective in transposition but, rather, that it was located in a hot-spot integration region in the DR locus.

It has been reported earlier that *M. tuberculosis* strains isolated from patients in countries with a high prevalence of TB exhibit less DNA polymorphism than strains from patients in countries that have a low prevalence of TB, such as The Netherlands (31). In countries with a high prevalence of TB, such as Ethiopia and Tunisia, the majority of circulating *M. tuberculosis* strains belong to a limited number of strain families (13) and supposedly descend from a few clones that expanded in the recent past. In the present study there was a high proportion of clustering of the isolates by IS*6110* RFLP patterns, resulting in a limited number of different RFLP banding patterns. This clustering was highest in group B, with clusters significantly larger than those in group A. On the basis of the large size of the major group B clusters, it appears that these strains have some selective advantage in this geographical area. In this study two of the eight isolates from patients from neighboring countries clustered with group B isolates; one was a biovar 2 isolate from a patient from The Gambia and one was a biovar 3 isolate from a patient from Guinea Conakry.

From an evolutionary point of view, we suggest that the group B and C isolates, including the four *M. bovis* isolates, belong to a unique branch of the *M. tuberculosis* tree, with their origin in West African countries. We suggest that these genetically similar strains have a recent common ancestral "*M. bovis*-like" origin and that they are better adapted for infection and survival in the range of hosts present in this geographical area. This genotype may have a short history of adaptation to humans, since TB in sub-Saharan Africa has been suggested to be a recently occurring disease, in contrast to TB in northern Europe and the Americas (26). We also propose that the *M. bovis* strains of the "European" type (which possess a single IS*6110* fragment and which lack DR spacer sequences 39 to 43) branched off at an earlier stage than the *M. bovis* strains of the type found in this study.

ACKNOWLEDGMENTS

We thank H. Heersma for helpful advice and J. van Embden and Dan Anderson for fruitful discussions.

The study was supported by the Swedish Medical Research Council (project no. 13027); the Agency for Research Cooperation with Developing Countries; the Swedish Heart-Lung Foundation; the Commission of the European Communities, Directorate General XII, Biomedical and Health Research, Biomed 1 (contract BMH1-CT93-1614); and the World Health Organization Programme for Vaccine Development.

REFERENCES

1. **Cave, M. D., K. D. Eisenach, P. F. McDermott, J. H. Bates, and J. T. Crawford.** 1991. IS*6110*: conservation of sequence in the *Mycobacterium tuberculosis* complex and its utilization in DNA fingerprinting. Mol. Cell. Probes **5:**73–80.
2. **Collins, C. H., M. D. Yates, and J. M. Grange.** 1982. Subdivision of *Mycobacterium tuberculosis* into five variants for epidemiological purposes: methods and nomenclature. J. Hyg. Camb. **89:**235–242.
3. **Collins, D. M., S. Erasmuson, D. M. Stephens, G. F. Yates, and G. W. De Lisle.** 1993. DNA fingerprinting of *Mycobacterium bovis* strains by restriction fragment analysis and hybridization with insertion elements IS*1081* and IS*6110*. J. Clin. Microbiol. **31:**1143–1147.
4. **David, H. L., M. T. Jahan, A. Jumin, J. Grandry, and E. H. Lehman.** 1978. Numerical taxonomy analysis of *Mycobacterium africanum*. Int. J. Syst. Bacteriol. **28:**467–472.
5. **Dias, F., S. Ghebremichael, S. E. Hoffner, L. Martins, R. Norberg, and G. Källenius.** 1993. Drug susceptibility in *Mycobacterium tuberculosis* of a sample of patients in Guinea Bissau. Tubercle **74:**129–130.
6. **Dolin, P. J., M. C. Raviglione and A. Kochi.** 1994. Global tuberculosis incidence and mortality during 1990–2000. Bull W. H. O. **72:**213–220.
7. **Fomukong, N. G., J. W. Dale, T. W. Osborn and J. M. Grange.** 1992. Use of gene probes based on the insertion sequence IS*986* to differentiate between BCG vaccine strains. J. Appl. Bacteriol. **72:**126–133.
8. **Goguet de la Salmoniere, Y.-O., H. Minh Li, G. Torrea, A. Buschoten, A. J. van Embden, and B. Gicquel.** 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. **35:**2210–2214.
9. **Groenen, P. M. A., A. E. Bunschoten, D. van Soolingen, and J. D. A. van Embden.** 1991. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel method. Mol. Microbiol. **105:**1057–1065.
10. **Groothuis, D. G., and M. D. Yates (ed.).** 1991. Manual of diagnostic and public health mycobacteriology, 2nd ed., Bureau of Hygiene and Tropical Diseases, European Society for Mycobacteriology, London, United Kingdom.
11. **Gutierrez, M., S. Samper, J.-A. Gavigan, J. F. G. Marin, and C. Martin.** 1995. Differentiation by molecular typing of *Mycobacterium bovis* strains causing tuberculosis in cattle and goats. J. Clin. Microbiol. **33:**2953–2956.
12. **Haas, W. H., G. Bretzel, B. Amthor, K. Schilke, G. Krommes, S. Rusch-Gerdes, V. Sticht-Groh, and H. J. Bremer.** 1997. Comparison of DNA fin-

gerprint patterns of isolates of Mycobacterium africanum from East and West Africa. J. Clin. Microbiol. **35:**663–666.
13. **Hermans, P. W. M., F. Messadi, H. Guebrexabher, D. van Soolingen, P. E. W. de Haas, H. Heersma, H. de Neeling, A. Ayoub, F. Portaels, D. Frommel, M. Zribi, and J. van Embden.** 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J. Infect. Dis. **171:**1504–1513.
14. **Hermans, P. W. M., D. van Soolingen, E.-M. Bik, P. E. W. de Haas, J. W. Dale, and J. D. A. van Embden.** 1991. The insertion element IS*987* from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. Infect. Immun. **59:**2695–2705.
15. **Hoffner, S. E., and G. Källenius.** 1998. Susceptibility of streptomycin-resistant *Mycobacterium tuberculosis* strains to amikacin. Eur. J. Clin. Microbiol. Infect. Dis. **7:**188–190.
16. **Hoffner, S. E., S. B. Svenson, R. Norberg, F. Dias, S. Ghebremichael, and G. Källenius.** 1993. Biochemical heterogeneity of *Mycobacterium tuberculosis* complex isolates in Guinea-Bissau. J. Clin. Microbiol. **31:**2215–2217.
17. **Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. Van Soolingen, S. Kuijper, A. Bunschoten, H. Molbuizen, R. Shaw, M. Goyal, and J. van Embden.** 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35:**907–914.
18. **Kent, P. T., and G. P. Kubica.** 1985. Public health mycobacteriology. A guide for the level III laboratory. Centers for Disease Control, Atlanta, Ga.
19. **Koivula, T., S. Hoffner, N. Winqvist, A. Nauciér, F. Dias, S. Svenson, R. Norberg, and G. Källenius.** 1996. *Mycobacterium avium* complex sputum isolates from patients with respiratory symptoms in Guinea Bissau. J. Infect. Dis. **173:**263–265.
20. **Murray, C. J. L., K. Styblo, and A. Rouillon.** 1990. Tuberculosis in the developing countries: burden, intervention and cost. Tuber. Lung Dis. **65:**6–24.
21. **Naucler, A., N. Winquist, F. Dias, T. Koivula, L. De Lacerda, S. B. Svenson, G. Biberfeld, R. Norberg, and G. Källenius.** 1996. Pulmonary tuberculosis in Guinea Bissau. Clinical and bacteriological findings, HIV-status and short term survival of hospitalized patients. Tuber. Lung Dis. **77:**197–292.
22. **Portugal, G., L. Brum, M. Viveiros, J. Moniz Pereira, and H. David.** 1998. Tipificacao genetica de estirpes multiresistentes de Mycobacterium tuberculosis isoladas na regiao de Lisboa. Rev. Port. Doencas Infect. **21:**54–59.
23. **Prat, R., N. Rist, N. Dumitrescu, A. Mugabushaka, S. Clavel, and C. Duponchel.** 1974. Special characteristics of the cultures of tubercle bacilli isolated in Ruanda. Bull. Int. Union Tuber. **49:**53–62.
24. **Scorpio, A., and Y. Zhang.** 1996. Mutations in *pncA*, a gene encoding pyrazinamidase/nicotinamidase, cause resistance to antituberculous drug pyrazinamide in tubercle bacillus. Nat. Med. **2:**662–667.
25. **Snider, D. E., R. C. Good, J. O. Kilburn, L. F. Laskowski, Jr., R. H. Lusk, J. J. Marr, Z. Reggiardo, and G. Middlebrook.** 1981. Rapid drug-susceptibility testing of *M. tuberculosis*. Am. Rev. Respir. Dis. **123:**402–406.
26. **Stead, W. W.** 1997. The origin and erratic global spread of tuberculosis. Clin. Chest Med. **18:**65–77.
27. **Sudre, P., H. G. ten Dam, and A. Kochi.** 1992. Tuberculosis: a global overview of the situation today. Bull. W. H. O. **70:**149–159.
28. **Szewzyk, R., S. E. Hoffner, G. Bölske, S. B. Svenson, and G. Källenius.** 1995. Molecular epidemiological studies of *Mycobacterium bovis* infections in humans and animals in Sweden. J. Clin. Microbiol. **33:**3183–3185.
28a. **van Embden, J.** Personal communication.
29. **van Embden, J., M. D. Cave, J. T. Crawford, J. W. Dale, K. D. Eisenach, B. Gicquel, P. Hermans, C. Martin, R. McAdam, T. M. Shinnik, and P. M. Small.** 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. **31:**406–409.
30. **van Soolingen, D., P. E. W. de Haas, J. Haagsma, T. Eger, P. W. Hermans, V. Ritacco, A. Alito, and J. D. van Embden.** 1994. Use of various genetic markers in differentiation of *Mycobacterium bovis* strains from animals and humans and for studying epidemiology of bovine tuberculosis. J. Clin. Microbiol. **32:**2425–2433.
31. **van Soolingen, D., P. W. M. Hermans, P. E. W. de Haas, D. R. Soli, and J. D. A. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. **29:**2578–2586.
32. **van Soolingen, D., P. W. M. Hermans, P. E. W. de Haas, and J. D. A. van Embden.** 1992. Insertion element IS*1081*-associated restriction fragment length polymorphism in *Mycobacterium tuberculosis* complex species: a reliable tool for recognizing *Mycobacterium bovis* BCG. J. Clin. Microbiol. **30:**1772–1777.
33. **Wayne, L. G.** 1985. The "atypical" mycobacteria: recognition and disease association. Crit. Rev. Microbiol. **12:**185–222.