Acta Academica Supplementum 2005(1): 82-109

Anneke van der Spoel-van Dijk, Zac Mokhethi, Macala Khumalo, Choala Shamputa, Zacheus Matebesi, Dingie van Rensburg, Françoise Portaels & Leen Rigouts

DNA fingerprinting analyses of *M tuberculosis*-complex isolates from the Free State, South Africa, as part of a multidisciplinary study

The objective of this study was to serve as a complement to socio-economic analyses of TB patients in a DOTS system, providing microbiological data and documenting the TB population dynamics. Sputum samples were collected from smear-positive TB patients in the Goldfields, Thaba Nchu and Qwaqwa areas. Laboratory analyses comprised the culturing of *Mycobacterium tuberculosis* isolates and DNA fingerprinting. The primary aim was hampered by problems encountered during specimen sampling, inadequate resources, and a low culture-positivity rate. Nevertheless, the fingerprinting data of a random sample showed a heterogenous TB population, suggesting that reactivation might be an important factor in the area studied. Clustering was the highest in the mining area. Preliminary data from serial isolates also detected possible re-infection during treatment or initial mixed infections in five of the eleven patients.

DNS-vingerafdrukanalise van *M tuberculosis*-kompleks isolate van die Vrystaat, Suid-Afrika, as deel van 'n multidissiplinêre studie

Die doel was om die sosio-ekonomiese analise van TB-pasiënte in 'n DOTS-sisteem met mikrobiologiese data te komplementeer en om die TB-populasiedinamika te dokumenteer. Sputummonsters is versamel van smeer-positiewe TB pasiënte in die Goudveld, Thaba Nchu en Qwaqwa. Laboratoriumanalise het bestaan uit die kweking van *Mycobacterium tuberculosis* isolate en DNS-vingerafdrukke. Die primêre doelstelling is gekortwiek deur probleme wat ondervind is met monsterversameling,

Ms A van der Spoel-van Dijk, Mr S Z Mokhethi & Mr M J Khumalo, Dept of Medical Microbiology, University of the Free State, P O Box 339, Bloemfontein 9300; Prof H C J van Rensburg & Dr S Z Matebesi, Centre for Health Systems Research & Development, University of the Free State, P O Box 339, Bloemfontein 9300; Mr I C Shamputa, Prof F Portaels & Dr L Rigouts, Mycobacteriology Unit, Dept of Microbiology, Institute of Tropical Medicine, Nationalestraat 155, B-2000 Antwerp, Belgium; E-mail: gnmbavds@mail.uovs.ac.za & lrigouts@microbiol.itg.be

gebrek aan fondse en 'n lae positiewe kwekingswaarde. Nogtans het vingerafdrukdata van 'n ewekansige monster 'n heterogene TB-populasie getoon wat suggereer dat heraktivering van vorige infeksies 'n belangrike faktor in die studiepopulasie mag wees. Die grootste saambundeling was teenwoordig in die myngebied. Verder toon voorlopige data van opeenvolgende isolate moontlike herinfeksie gedurende behandeling of aanvanklike gemengde infeksies in vyf van die elf pasiënte.

Since tuberculosis (TB) was declared a "global emergency" in 1993, the Directly Observed Treatment short course chemotherapy strategy (DOTS) and the later DOTS-plus for multi-drug-resistant tuberculosis (MDRTB) have been implemented in many countries and their impact on treatment adherence and outcome has been studied intensively.[1] An estimated two-thirds of the South African population is infected by latent TB bacilli and 60.0% of TB patients are co-infected with the Human Immunodeficiency Virus (DoH 2000). Recent statistics reveal an increase of 37.0% in reported cases from the inception of DOTS in 1996 to 2001. An incidence of 526/100 000 cases in 2002 placed South Africa at number 9 in the world rankings. Nationally 144 910 new cases of pulmonary TB (PTB) were reported in 2001; 9 978 (352/100 000) of them in the Free State (Kironde & Bamford 2002). New cases represented 78.15% and re-treatment cases 21.85%. 91.55% of the cases were treated under the DOT system (Dept of Health 2001).

DNA fingerprinting is used world-wide to monitor the transmission of TB strains, even across geographical borders (Casper *et al* 1996) — to confirm suspected cases of transmission, to detect laboratory con-

1 This kind of research would not be possible without the assistance of many dedicated people. We would like to acknowledge the Free State Department of Health, especially Annatjie Peters, Annette Furter, Leona Smith, and Me Bolofo and the staff at the clinics for their tremendous efforts and their support for the project. Sonja van der Merwe assisted in extracting data from the Government TB database, while students and assistants performed laboratory work and time-consuming data capturing. We thank patients for their willingness to take part in the study and Jeanette Buys for developing and maintaining the patient access database. We would also like to acknowledge the Medical Research Council of South Africa and the Dean's fund for funding, as well as the Flemish Government (Grant number BIL00/74) for supporting the bilateral collaboration between the University of the Free State and the Institute of Tropical Medicine.

tamination, to study outbreaks of disease and to characterise the clones that are circulating in a particular geographical region. They can also shed light on the degree of recent transmission versus reactivation among cases of recurrence (Warren *et al* 1996; Van Rie *et al* 1999).

IS*6110*-RFLP, a standardised restriction fragment length polymorphism (RFLP) method based on the insertion sequence IS*6110*, uses the polymorphism in the number and location of IS elements in the genome of *M tuberculosis*-complex isolates to differentiate individual strains (Van Embden *et al* 1993). This DNA-fingerprinting technique has been internationally agreed upon for the provision of information on TB populations (Saunders 1999; Van Soolingen 2001). Furthermore, the direct repeat (DR) sequence is used in a more rapid method called spoligotyping (Kamerbeek *et al* 1997), while the Mycobacterium Interspersed Repetitive Units (MIRUs) are used in a typing method named Variable Number of Tandem Repeats (VNTR) (Supply *et al* 1997). These latter methods comprise a DNA-amplifying step, permitting analysis from a small number of bacilli, or even directly from sputum specimens (Van der Zanden *et al* 2003).

Active TB may be caused by the reactivation of a previously acquired infection (endogenous origin) or by recent infection from a new source (exogenous origin). In developed countries, more than 90.0% of cases are thought to result from the reactivation of an earlier infection. However, studies on outbreaks among patients in crowded settings in the USA have suggested that recent transmission may also constitute a significant cause (Mc Luaghlin *et al* 2003). The phenomenon of re-infection has also been observed in community-based studies conducted in San Franscisco and New York City where respectively more than 31% and 37.5% of cases were found to be due to recent transmission (Alland *et al* 1994; Small *et al* 2002).

A study in Cape Town, South Africa, involving 16 patients with recurrent TB after a previous cure suggests that re-infection may be a major factor of recurrent disease (Van Rie *et al* 1999). Further information from South Africa comes from a study conducted on a goldmine, where in 25 out of 39 patients re-treatment was required due to reactivation (Sonnenberg *et al* 2001). This was confirmed by the findings of the study by Godfrey-Faussett and colleagues, which estimated

recent transmission at between 30.0 and 40.0% (Godfrey-Faussett *et al* 2000).

In this study we have attempted to complement socio-economic analyses of TB patients in a DOTS system with microbiological data from the same study population in the Free State Province of the Republic of South Africa. The TB population dynamics were documented on a representative random sample and the microbiological data of patients was gathered during treatment follow-up.

1. Materials and methods

1.1 Sample collection and isolation of *M tuberculosis*

The aim was to complement the findings of a socio-economic survey of 220 clinic-based PTB patients with microbiological data. Sputum samples were therefore collected from the same three selected high-burden clinics in each of the Thaba Nchu and Qwaqwa areas of the province. In the Goldfields area, however, only the Thabong clinic was included in the microbiological analysis. The other two clinics had to be omitted from the study due to lack of funding at one clinic and a bad response from the other. All smear-positive PTB patients over the age of 12 were meant to be enrolled, subject to written consent. As the culturing of mycobacteria in new cases is not routinely performed for the diagnosis of TB in the Free State, an additional sputum specimen was requested after TB had been confirmed but before treatment was initiated. Samples were collected from June 2001 to April 2003 in Thaba Nchu and from October 2001 to June 2003 in the Goldfields and Qwaqwa, due to practical and organisational difficulties encountered. In addition, sputum specimens were collected from all patients with a smear-positive result after completion of the initial phase of treatment (2 months) or at any further point during the course of treatment. Samples were transported daily to the microbiology laboratory of the University of the Free State and stored at -20°C pending further analysis.

The *Mycobacterium tuberculosis* complex was isolated by culturing on Löwenstein Jensen (LJ) media, after digestion and decontamination of the sputum by the N-acetyl-L-cysteine sodium hydroxide method

(Metchock *et al* 1995). To minimise the risk of cross-contamination, sputum specimens were processed in batches of 16; only one tube at a time was uncapped for the addition of solutions; buffer solutions were prepared as individual aliquots in single use tubes, and tubes were only opened five minutes after centrifugation — as recommended by Small and colleagues (1993).

1.2 DNA extraction and DNA fingerprinting

DNA was extracted from heat-killed (at 80°C for 1 hour) bacterial suspensions harvested from LJ slants using a phenol-chloroform method as described by Warren *et al* (1996).

IS*6110*-RFLP fingerprinting was performed using the standardised method (Van Embden *et al* 1993). Briefly, extracted DNA was restricted with *Pvu* II (Amersham Biosciences, UK) and electrophoresed on a 0.8% (w/v) Seakem® ME agarose gel (BioWhittaker molecular applications, USA). The DNA was transferred to a nitrocellulose membrane (Hybond–N+, Amersham Biosciences, UK) using a vacuum blotter (BioRad Laboratories, Hercules, CA, USA) according to the manufacturer's instructions. The enhanced chemiluminescence method was used to label the IS*6110* probe and detect restriction fragments containing IS*6110* insertions (Gene Images CDP star labelling and detection kit, Amersham Biosciences, UK). On the basis of the molecular sizes of the hybridising fragments and the number of IS*6110* copies in each isolate, similarity between isolates was determined using the GelCompar II version 2.5 Dice unweighted pair group method with arithmetic averages (Applied Maths, St-Maartens Latem, Belgium).

1.3 MIRU-VNTR typing

MIRU-VNTR typing for 12 loci containing MIRU was carried out according to Mazars *et al* (2001). A 3% NuSieve agarose gel was used instead of the 2% in the method described.

1.4 Collection of patients' demographic data

The disease histories and demographic data of patients were collected from patient files at the various clinics at intervals during treatment to ensure that correct information was obtained about treatment ad-

herence and outcomes. This data was captured in an Access database with limited access to ensure the confidentiality of patient identities. All specimens received by the laboratory were renumbered and handled anonymously.

2. Results

During the study period 565 smear-positive PTB patients were registered at the three selected clinics in Thaba Nchu, 461 at the three clinics in Qwaqwa, and 657 at the Thabong clinic in the Goldfields area (Table 1). A total of 647 sputum specimens were collected from 530 of these patients and sent to the microbiology laboratory for further analysis (Tables 1 and 6). In order to document the TB population dynamics, IS*6110*-RFLP typing was performed on a random sample of patients. All *M tuberculosis* isolates from the Thaba Nchu clinics and the isolates from the first 50 samples received from the Qwaqwa and the Thabong clinics were typed. In addition, DNA fingerprints were available for 11 patients who had additional specimens after two or three months of treatment (three from Thaba Nchu, three from Qwaqwa and five from Thabong). This resulted in good fingerprints for a total of 174 patients (Table 1). A comparison of the TB history data (Table 2) from all registered patients, all patients from whom a sputum specimen had been collected for the study, and all patients fingerprinted did not reveal significant differences. Demographic data (sex and age) were only available for new cases (Figure 1) and for the period from July 2001 to December 2002, thus totalling only 889 cases. No difference was observed among the sample groups. Since there was no significant difference in the distribution of new and re-treatment cases among the various sample groups, and the age distribution for new cases was not significantly different for both groups, it may be assumed that the same probably holds good for the re-treatment cases, and that our sample with fingerprint data can therefore be considered representative.

Slightly more than 93% (163/174) of patients with fingerprinted isolates were between the ages of 20 and 59, and 104 (59.8%) were men (Figure 1). 118 (67.8%) were diagnosed with tuberculosis for the first time (new cases) while 56 had previously been treated for tuberculosis for at least one month (re-treatment cases) (WHO, 2003) (Table 2). Among the re-treatment cases 24 of the 56 (42.9%) had

Table 1: Overview of registered patients and sputum specimens received from the selected clinics in the various study areas during the study period, with culture and DNA-fingerprinting analyses performed.

Study area	Number of smear-positive pulmonary TB patients				
	Registered	With sputum specimen collected for study			
		Total	Culture performed	Culture positive	IS*6110*-RFLP fingerprint
Thaba Nchu	565	218	205	92	74
Qwaqwa	461	193	131	86	44
Goldfields (Thabong clinic)	657	119	99	102	56
Total	1683	530	435	280	174

Table 2: Distribution of pulmonary TB patients according to treatment history

Patient categories	Number of smear-positive PTB patients		
	Registered (%)	With sputum collected for study (%)	With DNA fingerprint (%)
History unknown		14 (2.6)	
New cases	1271(75.5)	387 (73.0)	118 (67.8)
Re-treatment cases	412(24.5)	129 (24.3)	56 (32.2)
Cure		42	24
Completed		47	22
Interrupted		13	7
Failure		10	3
Unknown		17	0
Total	1683	530	174

Re-treatment cases have been specified as follows: Cure = patient declared cured after previous treatment, *i e* smear-negative at microscopy and no clinical signs; Completed = previous treatment completed without proven cure or failure; Interrupted = previous treatment was interrupted; Failure = previous treatment failed, *i e* no smear conversion occurred during treatment.

successfully completed a previous treatment and been declared cured, and therefore were probably reactivation or re-infection cases. 22 of the 56 (44.9%) re-treatment patients had completed a previous treatment whose success could not be documented, whereas seven (12.5%) had interrupted their previous treatment and three had undergone failed treatment.

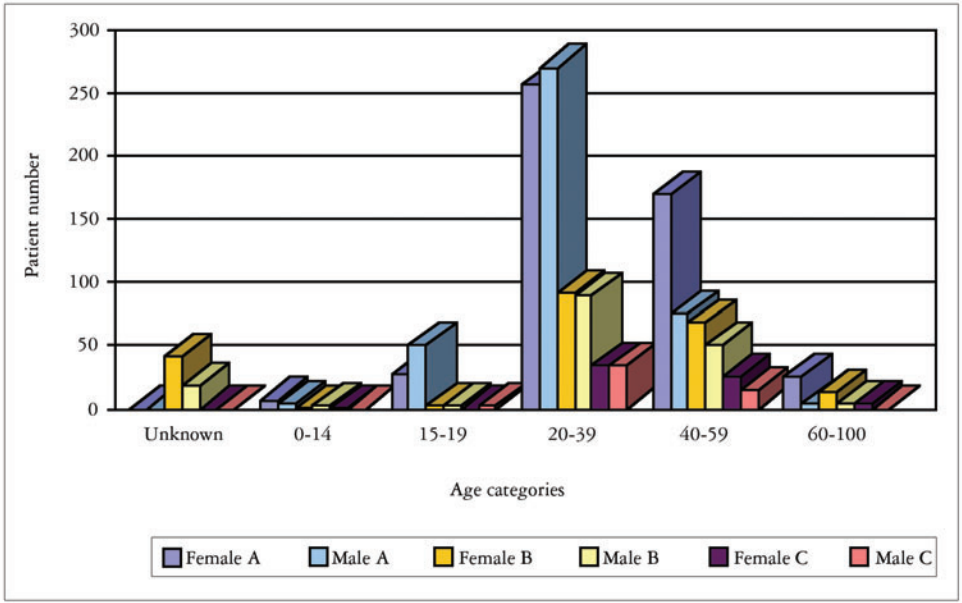
The calculation of similarity levels between IS*6110*-RFLP profiles of the 174 *M tuberculosis* isolates analysed yielded 17 clusters each containing 2-7 isolates with identical fingerprinting patterns (Figure 2). Information of the patients with clustered isolates is given in Table 3. Thirteen clusters comprised only two identical isolates; two clusters (10 and 11) included three, and another two (8 and 15) included four isolates. Eight of the 17 clusters included patients from the same area, for example, the strain from cluster 1 was isolated in two patients from the Goldfields area. In addition, the strains from clusters 1, 2, 5, 8 and 14 were not only collected in the same area but also from the same clinic. All the clusters from the same areas included both new cases and re-treatment cases except for cluster 14, which had two re-treatment cases, and cluster 11, which had three new cases. Similarly, among the eight clusters with patients from different areas only two (9 and 12) included only new cases. However, no evidence of index cases could be determined for any of the clusters since these patients could not be reached for interviews after they had completed treatment.

The strains from clusters 15, 16 and 17 had five RFLP bands or fewer, and therefore could not be definitely regarded as a cluster without further sub-typing by another genotyping method such as the MIRU-VNTR. However, these results were not available in this study.

Overall, there was 10.9% clustering within the geographical areas of our study, whereas 12.1% were cross-border clusters including patients residing in two of the three geographical areas (Table 4). The intra-regional clustering was significantly higher in the mining area than in the Thaba Nchu area ($p = 0.021$).

Apart from the clusters with 100% identical profiles, we also looked for the presence of genotype families with strains showing similar but not identical profiles. At a similarity level of 65%, 23 family groups could be distinguished as well as 19 individual strains that did not belong to any group (Figure 2). Twenty isolates had fewer than six

Figure 1: Age distribution and gender of new smear-positive pulmonary TB patients from the three study areas during the study period



A = from all new smear-positive registered pulmonary patients in the selected clinics from July 2001 to December 2002 (n = 889);
B = from those new smear-positive patients who had a sputum specimen collected for the study (n = 387);
C = from those new smear-positive patients whose *M tuberculosis* isolates were DNA-typed (n =118).

Table 3: Patient and isolate characteristics of clustered *M tuberculosis* isolates: isolates showing identical IS*6110*-RFLP patterns

Isolate characteristics				Patient characteristics			
Isolate number	IS*61110* copies	IS*6110* cluster	Resistance profile	Gender	Age	Clinic (area and clinic number)	Patient category
GF 16	10	1	S	M	40	Goldfields 1	CURE
GF 15	10	1	ND	M	41	Goldfields 1	NEWC
GF 05	13	2	ND	M	46	Goldfields 1	COMP
GF 03	13	2	ND	M	51	Goldfields 1	NEWC
GF 11	15	3	ND	M	64	Goldfields 1	COMP
ZT 47	15	3	S	F	20	Thaba Nchu 3	NEWC
Q 36	14	4	ND	M	29	Qwaqwa 2	NEWC
Q 35	14	4	ND	M	55	Qwaqwa 3	INTR
GF 21	15	5	ND	F	29	Goldfields 1	COMP
GF 20	15	5	ND	M	23	Goldfields 1	NEWC
Q 24	10	6	ND	F	24	Qwaqwa 2	COMP
ZT 34	10	6	S	F	34	Thaba Nchu 2	NEWC
GF 40	9	7	ND	M	29	Goldfields 1	NEWC
ZT 35	9	7	ND	M	51	Thaba Nchu 2	CURE
GF 30	11	8	ND	M	38	Goldfields 1	NEWC
GF 63	11	8	ND	M	43	Goldfields 1	NEWC
GF 32	11	8	R	M	37	Goldfields 1	NEWC
GF 31	11	8	ND	M	32	Goldfields 1	INTR
ZT 59	8	9	ND	F	35	Thaba Nchu 1	NEWC
GF 47	8	9	R	M	57	Goldfields 1	NEWC
Q 32	10	10	ND	F	32	Qwaqwa 2	NEWC
ZT 61	10	10	ND	M	27	Thaba Nchu 2	CURE
ZT 60	10	10	S	F	41	Thaba Nchu 2	NEWC
ZT 12	9	11	ND	F	37	Thaba Nchu 3	NEWC
ZT 10	9	11	S	F	21	Thaba Nchu 2	NEWC
ZT 09	9	11	S	M	41	Thaba Nchu 3	NEWC
Q 18	14	12	ND	F	49	Qwaqwa 2	NEWC
ZT 17	14	12	S	M	41	Thaba Nchu 1	NEWC
Q 22	9	13	ND	M	50	Qwaqwa 1	CURE
Q 21	9	13	ND	F	35	Qwaqwa 3	NEWC
Q 14	12	14	ND	F	30	Qwaqwa 2	CURE
Q 13	12	14	ND	M	70	Qwaqwa 2	CURE
Q43	4	15	ND	M	55	Qwaqwa 2	COMP
Q44	4	15	ND	M	30	Qwaqwa 2	NEWC
Q45	4	15	ND	M	25	Qwaqwa 2	COMP
GF57	4	15	R	M	47	Goldfields 1	CURE
Q05	5	16	ND	F	43	Qwaqwa 1	NEWC
ZT30	5	16	S	F	39	Thaba Nchu 2	CURE
Q04	4	17	ND	F	42	Qwaqwa 3	NEWC
ZT74	4	17	S	M	46	Thaba Nchu 3	NEWC

ZT = Thaba Nchu; GF = Goldfields; Q = Qwaqwa; F = female; M = male; IS*6110* clusters comprising isolates showing 100% similarity after IS*6110*-RFLP analysis; S = susceptible to 0.2µg/ml H; R = resistant to H at 0.2µg/ml; ND = susceptibility to H was not determined; NEWC = patient smear-positive for tuberculosis for the first time; CURE = patient declared cured after previous treatment, smear-negative at microscopy and no clinical signs; Completed = previous treatment completed without proven cure or failure; INTR = previous treatment interrupted; FAIL = previous treatment failed, no smear conversion occurred during treatment. Alternating the colour of different clusters using white and yellow highlights isolates with clustered IS*6110*-RFLP patterns.

Figure 2: IS*6110*-based dendrogram of initial *M tuberculosis* isolates from 174 patients in the three study areas

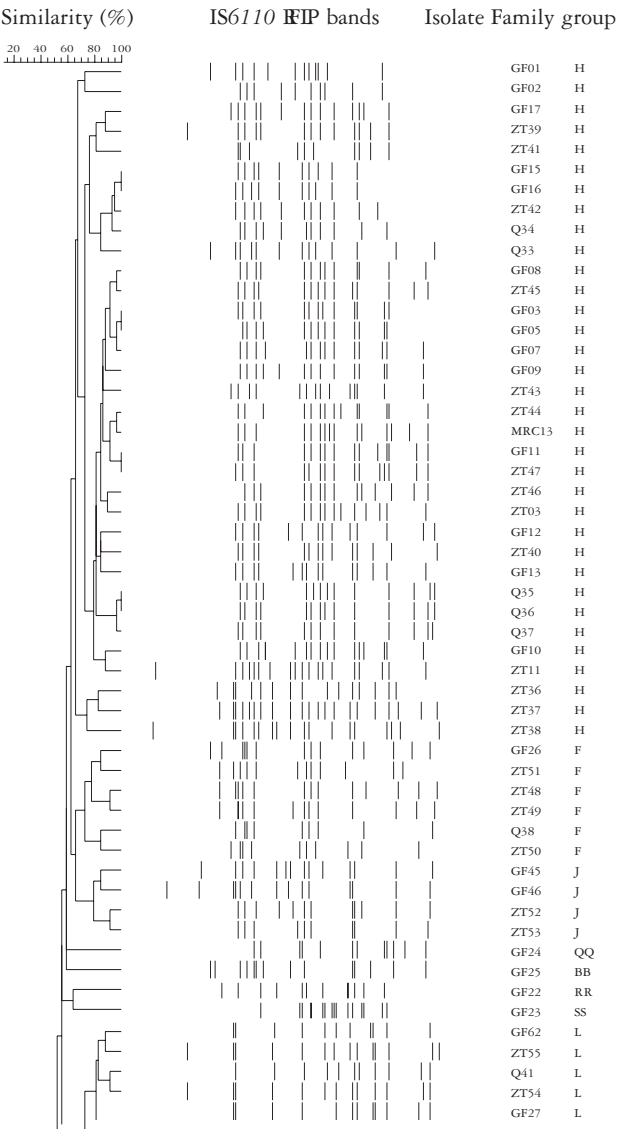


Figure 2: IS*6110*-based dendrogram of initial *M tuberculosis* isolates from 174 patients in the three study areas (continued)

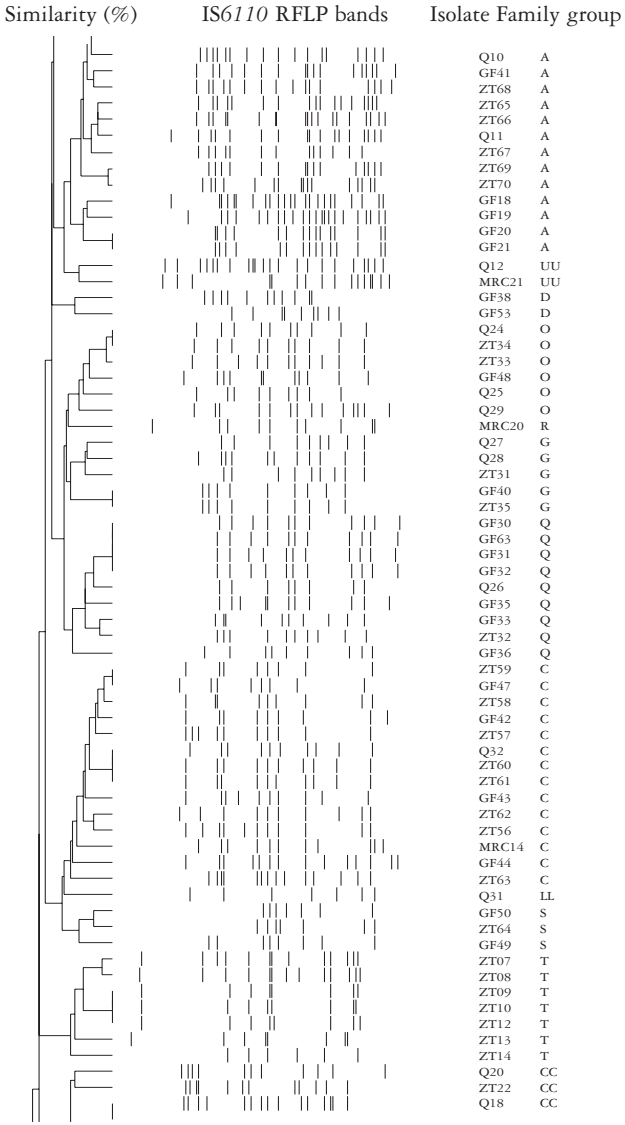
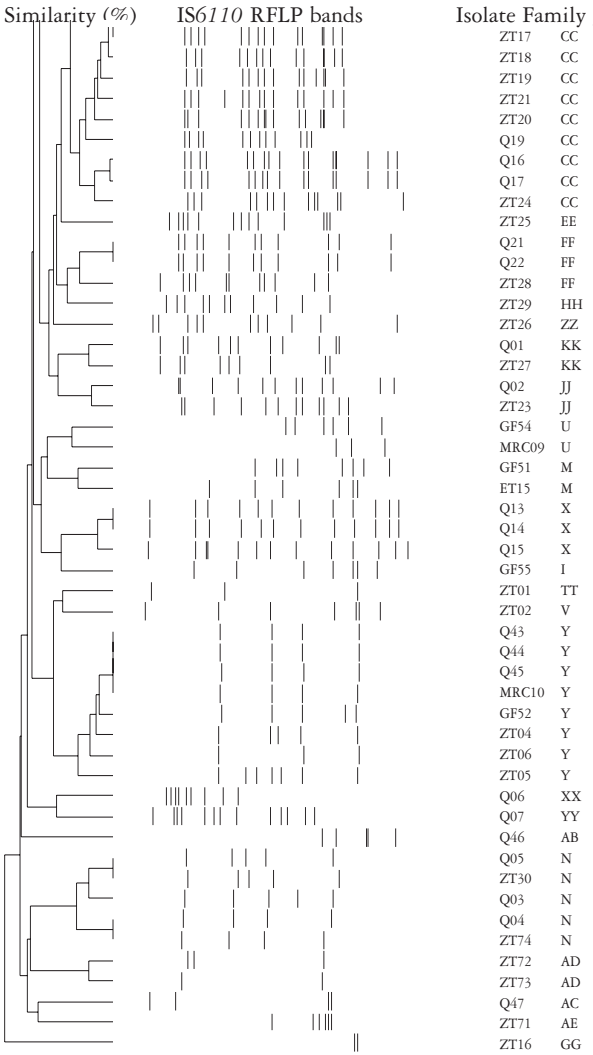


Figure 2: IS*6110*-based dendrogram of initial *M tuberculosis* isolates from 174 patients in the three study areas (continued)



The fingerprinting patterns were analysed for similarity using the Dice coefficient and the unweighted pair group method to calculate the dendrogram (Gelcompar software). ZT = Thaba Nchu; GF = Goldfields; Q = Qwaqwa.

IS*6110* insertions, including the strains from the three clusters mentioned above. The distribution of the family groups for the 154 isolates with more than 5 IS*6110* copies is presented in Figure 3. Four families, A, C, CC and H, contained more than 10 isolates. The largest group was H, with 34 isolates, representing 22.1% of the 154 isolates. This family was more prevalent among patients from the Goldfields (15 out of 53, 28.3%) compared to Thaba Nchu (14 out of 65, 21.5%) and Qwaqwa (5 out of 36, 13.9%).

None of the strains from the three regions had an IS*6110* profile typical of the Beijing family. One strain (GF52) with a pattern similar to the super-strain DRF150 reported in the Western Cape and Mpumalanga was found in the Goldfields area. However, the strain in our study was susceptible to isoniazid and rifampicin, unlike the previously reported strain, which was multi-drug-resistant (Victor 2003).

For 98 patients in our study, an additional sputum specimen was provided after two (T2) and/or three (T3) months of treatment. IS*6110*-RFLP typing yielded good results for both isolates for only 11 of these patients, of whom nine had two isolates and two, three. Results are presented in Figure 4 and summarised in Table 5.

Only two of the 11 patients with two successive samples showed identical fingerprints for both isolates (patient 1 from Goldfields and patient 8 from Qwaqwa). Follow-up specimens from these patients were sampled after two months, treatment and both patients became negative for microscopy after three months' treatment. Patient 8 was declared cured after treatment but patient 1 died of a non-TB disease.

Consecutive isolates from the remaining nine patients yielded different RFLP profiles. For two of them (patients 2 and 17) none of the other isolates fingerprinted showed a similar profile, making laboratory cross-contamination unlikely, and suggesting either initial mixed infection or re-infection during treatment. Patient 2 was a new case and was transferred out during treatment, whereas patient 17 was a re-treatment case after treatment interruption, who was finally declared cured. Both isolates of patient 17 were susceptible to first-line drugs.

On the other hand, the initial or follow-up isolates of seven patients (3, 5, 6, 10, 11, 13 and 15) produced RFLP patterns identical to the initial isolates of patients 4, 7, 9, 12, 14 and 16, respectively. Sub-