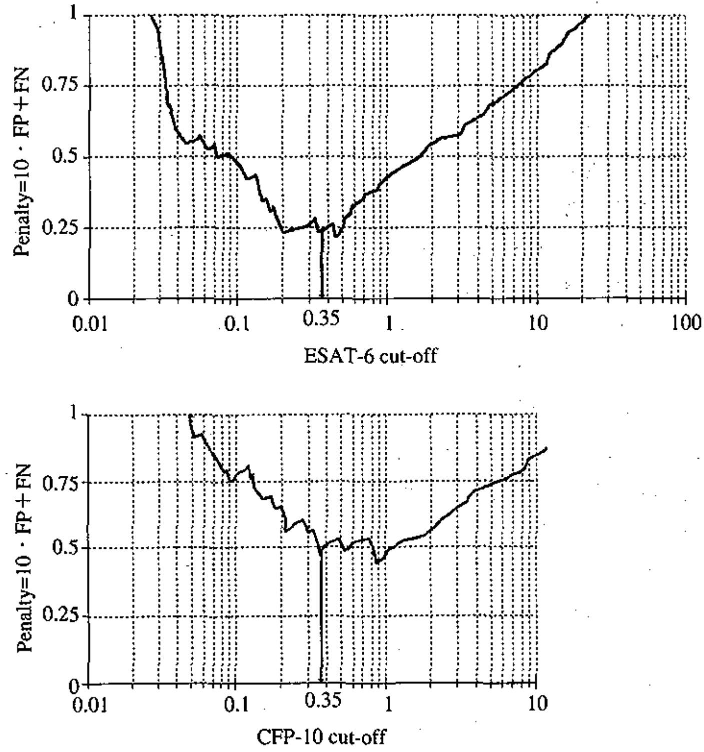

Fig. 7 Analysis for determination of the optimal cut-off value. Penalty values were plotted for each ESAT-6 (top) or CFP-10 (bottom) cut-off values. The optimal cut-off values were chosen when penalty value becomes minimum.

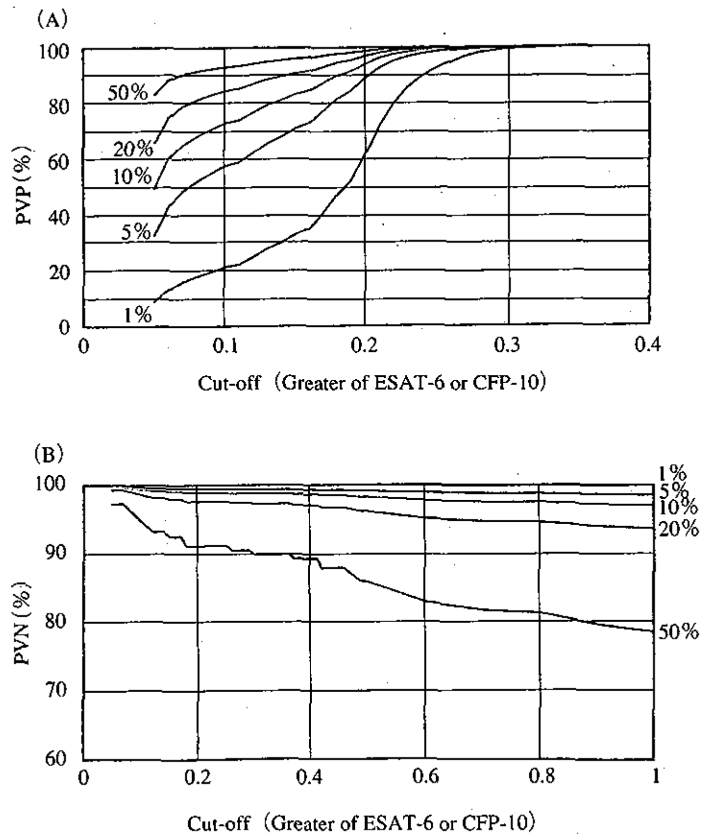

Fig. 8 Predictive value of positive (A) or negative (B) by cut-offs for selected levels of prevalence.

特異抗原に対するIFN-γ応答の定量的分析は，すべて測定値に対数変換を施して行った。これによって，患者においては正規型の，また未感染者においては指数型の分布が観察され，接触者においては両者の混合体としての複合分布が見られた。これらはツベルクリン反応の反応径と全く同様である。

患者におけるIFN-γ応答には，ツベルクリン反応の場合[18]と違って有意の年齢差は見られなかった。年齢のほか病状との関連などは冒頭に述べた他の条件との関連と同様，今後の研究課題である。

ESAT-6とCFP-10に対するIFN-γの関連を結核患者について見ると，0.303という弱い正の相関が見られる。これに関しては，ESAT-6とCFP-10は同じ遺伝子ファミリーに属し，低いながらも類似性を持つこと[7]が関係している可能性が考えられるが，いずれにせよ，この低い相関からして，設定したカットオフによる陽性区分はESAT-6とCFP-10で一致性は偶然以上のものでないという結果となり，両者への反応を組み合わせて診断に用いることの有効性が出てくることになった。なお，BCG既接種者においても両者の間にはごく弱い相関が見られた。本来ならばここでは無相関のはずであるが，このようなことになった原因としては，わずかにあったかもしれない結核菌や抗酸菌感染など非特異的な反応要因の影響が考えられる。

未感染健常者においてツベルクリン反応とIFN-γ応答の関連については，ツベルクリン反応発赤・硬結径の分布とIFN-γ応答の分布が全く異なることから，無相関は自明である。事実ピアソン相関もそれを裏付けていた。このようにこの抗原がBCG接種によって作られた細胞免疫ないしツベルクリン・アレルギーとはほぼ完全に独立していることが確認されるのである。

結核感染によるツベルクリン反応とIFN-γ応答の間には，結核患者における観察から，弱いが有意の相関があることが知られる。しかしその弱さからツベルクリン反応にはIFN-γだけでない多くの要因が絡んで反応の強弱が決定されることを改めて知らされる。

カットオフ設定の問題は，ツベルクリン反応を含む多くの意思決定問題とまったく同様に損失関数と事前確率の処理に帰する。これを一般的に解決することは不可能なので，多くの実用的なケースにならって，便宜的な仮定と戦略を採り，その上でそのカットオフの有用性を検討した。このようにしてカットオフはESAT-6，CFP-10双方に対して0.35 IU/*ml*と設定し，実際的にはそのいずれかに対してこの値を超える応答が見られたケースを陽性とすることとした。（仮定的）絶対基準に照らして，その感度は89.0%，特異度は98.1%であったが，このパフォーマンスはカットオフ値を0.2～0.4 IU/*ml*程度の範囲で変化させても大きくは変わらない。

これに対して，別の患者集団にこれを適用した経験から0.35 IU/*ml*の感度は結核菌塗抹陽性肺結核患者において73%と低く，たとえば，0.2 IU/*ml*（感度は87%）0.1 IU/*ml*（同93%）程度にしてはどうかという意見[19]もあった。このような観察されたパフォーマンスの差については研究に参加した患者の免疫能に影響する要因の分布の違いも考えなければならないが，感度を上げるためにカットオフ値を下げることが有効なことは言うまでもない。しかし同時にそれは特異度を下げるというtrade-offを考えなければならない。ちなみに，カットオフを0.1 IU/*ml*に下げることで感度は95%に上げられるが，特異度は92%にまで低下する。この特異度の低下は，Fig. 8で見るように，有病率の低い集団にこの検査を適用した場合に深刻な影響を及ぼす。たとえば有病率が1%の集団では，カットオフが0.35 IU/*ml*ならば陽性的中率はほぼ100%だが，0.1 IU/*ml*になると21%に下がる。つまりこの判定基準で「陽性」とされた者のうち真に感染を受けた者は21%しかいないのである。同様に設定したカットオフ値が高い場合には，高い有病率（既感染率）の場合に「陰性的中率」（陰性と判定された者のうち，真の未感染者の割合）に大きな影響を与える。

このように最善のカットオフの設定は方法で見たように対象集団の有病率水準に依存するので，ツベルクリン反応検査においては日本では濃厚接触者集団では発赤10 mm，一般小児集団では30 mmという「二重基準」を推奨している[20]。米国でも同様である[2]。これにならってQFTにおいても，感染の危険の高い集団では0.1 IU/*ml*を，一般の集団では0.35 IU/*ml*をそれぞれ陽性の基準とするような戦略をとることも考えるべきである。この考え方に基づき，すでにわれわれ（結核研究所抗酸菌レファレンスセンター）は，濃厚接触があり，0.35 IU/*ml*で陽性者が高頻度で発生している集団においては0.1 IU/*ml*～0.35 IU/*ml*の者（仮に「疑陽性者」と区分）に対しても化学予防を推奨している。もちろん上記「疑陽性者」に対しては感染に関する事前確率の推定のみによって化学予防の適否を決定するだけでなく，免疫能の関連要因，感染からの時間経過等も考慮するほか，再検査や追跡のような措置の上での妥協的な対応を考えることも意味がある。

実用的な立場から，われわれはもとの測定値のカットオフによる陽性・陰性の判定を行う方法をこの検査法の判定戦略とした。一方，数学的には判別関数を用いて2個の独立変量に重みづけを行い，これにより基準変量（結核・健康）を判別する戦略も考えられる。試みにこれを行ってみると正準相関係数（重みづけされた各測定値と診断〔結核＝1，健康＝0〕との間の相関係数）は0.911

と高く，結核の事前確率を0.5とした場合の正しい判別率は96.4%となった（われわれの方法では93.5%）。しかしわれわれの簡便な方法の有用性はすでに十分に高いので，判別関数もこれに代わるものではない。

カットオフ設定にあたってわれわれは偽陽性と偽陰性の損失を10：1と仮定した。前者には不必要な精密検査，薬剤投与，薬剤副作用，追跡検査，それに苦悩など，後者には結核発病とそれによる周囲への感染という損失がそれぞれある。これを10：1（w＝10）と評価したことになるが，もちろん議論の余地は大いにある。ただしw値を2～10の範囲で変化させてもPenaltyを最小にするx値はあまり変化しなかったことから見て，設定そのものは無理な仮定ではなかったと考える。ただし損失についてはより精密な検討が必要であり，さらには検査の経費も含めた，検査体系全体の経費・効果分析が必要であろう（予備的な分析を森[21]が行っている）。

まとめ

①活動性結核患者，未感染健常者，患者接触者の3集団について，結核菌特異抗原に対する末梢血のIFN-γ応答を定量し，それぞれにおいて結核感染を表す特徴的な分布を観察し得た。とくに未感染健常者においてはそのIFN-γ応答はBCG接種に全く影響されなかった。

②上記の分布のROC曲線と陽性・陰性の誤分類による損失の検討から，IFN-γ応答による結核感染の判定のためのカットオフ値を設定した。

③2種の特異抗原に対するIFN-γ応答の間の相関は弱く，判定は両者を独立に用いることにより感度を上げることができた。

④ツベルクリン反応検査と同様に，適用する集団の既感染率が高い集団では，カットオフ値はより低いレベルに設定することにより陽性的中率をあまり下げずに見落としを減らすことができる。

謝　辞

健常者，患者における知見は以下に掲げる研究者の協力の下に得られた。坂谷光則・岡田全司・鈴木克洋・井上義一・露口一成（独立行政法人国立病院機構近畿中央胸部疾患センター），山岸文雄・佐々木結花（独立行政法人国立病院機構千葉東病院），高嶋哲也・露口泉夫（大阪府立呼吸器・アレルギーセンター），川辺芳子（独立行政法人国立病院機構東京病院），長尾啓二（千葉大学医学部保健管理センター），重藤えり子（独立行政法人国立病院機構東広島医療センター），松本憲子（宮崎県立看護大学），御手洗聡（結核予防会結核研究所）。接触者については関係保健所担当者諸氏の援助を受けた。統計処理については内村和広氏（結核予防会結核研究所）の支援を受けた。心から感謝の念を捧げたい。

文　献

1）World Health Organization: Global Tuberculosis Control. Surveillance, Planning, Financing. WHO Report 2004. Geneva, Switzerland, 2004.
2）*ATS/CDC Statement* Committee on Latent Tuberculosis Infection: Targeted Tuberculin Testing and Treatment of Latent Tuberculosis Infection. MMWR 2000/49 (RR06) 2000: 1–54.（邦訳　資料と展望. 2001; 36: 25–68).
3）von Reyn CF, Horsburgh CR, Olivier KN, et al.: Skin test reactions to *Mycobacterium tuberculosis* purified protein derivative and *Mycobacterium avium* sensitin among health care workers and medical students in the United States. Int J Tuberc Lung Dis. 2001; 5: 1122–1128.
4）Mahairas GG, Sabo PJ, Hickey MJ, et al.: Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J Bacteriol. 1996; 178: 1274–1282.
5）Andersen P, Andersen AB, Sorensen AL, et al.: Recall of long-lived immunity to *Mycobacterium tuberculosis* infection in mice. J Immunol. 1995; 154: 3359–3372.
6）Sorensen AL, Nagai S, Houen G, et al.: Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. Infect Immun. 1995; 63: 1710–1717.
7）Berthet FX, Rasmussen PB, Rosenkrands I, et al.: A *Mycobacterium tuberculosis* operon encoding ESAT-6 and a novel low-molecular-mass culture filtrate protein (CFP-10). Microbiology. 1998; 144: 3195–3203.
8）Arend SM, Andersen P, van Meijgaarden KE, et al.: Detection of active tuberculosis infection by T cell responses to early-secreted antigenic target 6-kDa protein and culture filtrate protein 10. J Infect Dis. 2000; 181: 1850–1854.
9）Food and Drug Administration, Center for Devices and Radiological Health. QuantiFERON®-TB-P010033 [Letter]. www. fda. gov/cdrh/pdf/P010033b. pdf. 2002, 5–13.
10）Mori T, Sakatani M, Yamagishi F, et al.: Detection of tuberculosis infection with an interferon-gamma based assay using new antigens. Am J Respir Crit Care Med. 2004; 170: 59–64.
11）原田登之，森　亨，宍戸眞司，他：集団感染事例における新しい結核感染診断法QuantiFERON®TB-2Gの有効性の検討．結核．2004; 79: 637–643.
12）Weinstein MC, Fineberg HV: *Clinical decision analysis*. W.B. Saunders Co, Philadelphia, 1980, 75–130.
13）水野哲夫：臨床・公衆衛生のための意思決定．南光堂，東京，1968，219–230.
14）Hirsch CS, Toossi Z, Othieno C, et al.: Depressed T-cell interferon-gamma responses in pulmonary tuberculosis: analysis of underlying mechanisms and modulation with therapy. J Infect Dis. 1999; 180: 2069–2073.
15）Pathan AA, Wilkinson KA, Klenerman P, et al.: Direct ex

vivo analysis of antigen-specific IFN-gamma-secreting CD4 T cells in *Mycobacterium tuberculosis*-infected individuals: associations with clinical disease state and effect of treatment. J Immunol. 2001 ; 167 : 5217–5225.
16) Vekemans J, Lienhardt C, Sillah JS, et al.: Tuberculosis contacts but not patients have higher gamma interferon responses to ESAT-6 than do community controls in The Gambia. Infect Immun. 2001 ; 69 : 6554–6557.
17) 森　亨：結核の現状と対策. 日本内科学会雑誌. 2001 ; 90 : 111–115.
18) 森　亨：「ツベルクリン反応検査」. 結核予防会, 東京, 1995, 34–36.
19) 川辺芳子, 鈴木純子, 益田公彦, 他：新しい結核感染診断キット QuantiFERON®-TB の臨床評価—とくに判定基準の検討. 結核. 2004 ; 79 : 201.
20) 厚生省保健医療局疾病対策課結核・感染症対策室長通知：初感染結核に対する INH の投与について. 平成元年2月28日健医感発第20号.
21) 森　亨：新たな結核対策の技術と展望. 結核. 2004 ; 79 : 587–604.

Original Article

BASIC CHARACTERISTICS OF A NOVEL DIAGNOSTIC METHOD (QuantiFERON®TB-2G) OF LATENT TUBERCULOSIS INFECTION WITH A USE OF *MYCOBACTERIUM TUBERCULOSIS*-SPECIFIC ANTIGENS, ESAT-6 AND CFP-10

[1]Nobuyuki HARADA, [1]Kazue HIGUCHI, [1]Yukie SEKIYA, [2]Jim ROTHEL, [3]Takashi KITOH, and [1]Toru MORI

Abstract [Purposes] To determine the optimum cut-off level of a newly developed method for diagnosing tuberculosis infection based on whole-blood interferon-gamma measurement, and to study the basic characteristics of the method.

[Study Subjects] 1) A total of 220 young, healthy individuals having no apparent exposure to tuberculosis infection, most of whom have had a vaccination with BCG vaccine. 2) One hundred eighteen tuberculosis patients who were diagnosed by positive *Mycobacterium tuberculosis* on culture. 3) A group of 75 youngsters exposed to an infectious tuberculosis patient and who showed a strong tuberculin reaction (with erythema diameter of 30 mm or more).

[Method] Whole-blood specimens of donors were stimulated with antigens, i.e., ESAT-6 and CFP-10, and then cultured. Plasma concentrations of interferon-gamma discharged were then determined with QuantiFERON®-CMI. Correlation between interferon-gamma concentrations in response to ESAT-6 and CFP-10, and their correlation with Mantoux test results were analyzed for various categories of donors. The Receiver Operating Characteristics analysis was performed considering the loss due to misclassification.

[Results and Discussion] The optimum cut-off level was determined as 0.35 IU/*ml* for both ESAT-6 and CFP-10. This gave the test a sensitivity of 89.0% and specificity of 98.1% in detecting tuberculosis infection. The correlation of interferon-gamma response with tuberculin tests among BCG-vaccinated individuals was low, which suggested that the test was not influenced by previous BCG vaccination. The low correlation between ESAT-6 and CFP-10 tests suggested that the simultaneous use of the two tests was beneficial. As in the case of clinical tests in general, the cut-off should be set at a lower level when the test is applied to high prevalence situation and vice versa.

Key words: Latent tuberculosis infection, Diagnostics, ESAT-6, CFP-10, IFN-γ, ELISA

[1]Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association (JATA), [2]Cellestis Ltd, Melbourne, Australia, [3]Nichirei Corporation

Correspondence to: Nobuyuki Harada, Immunology Division, Research Institute of Tuberculosis, JATA, 3–1–24, Matsuyama, Kiyose-shi, Tokyo 204–8533 Japan.
(E-mail: harada@jata.or.jp)

集団感染事例における新しい結核感染診断法 QuantiFERON®TB-2Gの有効性の検討

原田 登之　森 亨　宍戸 眞司　樋口 一恵
関谷 幸江

要旨：〔目的〕結核接触者健診における結核感染診断の目的のために，結核菌特異抗原（ESAT-6およびCFP-10）を用いた全血インターフェロンγ応答を定量する方法（QuantiFERON®TB-2G）がどのように有効であるかを検討するために本研究を行った。〔対象と方法〕ある若年者集団（専門学校生）における結核患者の発生に際して，ツ反とともに本法を適用し，その知見をツ反と比較分析した。〔結果〕初発患者との接触が濃厚な群とそれ以外の接触者とでは本法の陽性率に大きな差が見られたが，ツ反では違いはわずかで，本法がBCG接種の影響を受けずに結核感染診断が正確に行えることを示唆していた。〔結論〕ツベルクリン反応検査との比較から，この方法は従来のツベルクリン反応検査とその便宜的な診断基準による方法で回避できない不必要に過剰な化学予防の指示を大幅に節減し，また適用方法を工夫することによって必要な対象者を最大限補足することが可能になると考えられる。いっそう信頼性のある方法として広範に利用できるものとなるために克服すべきいくつかの課題が残されているものの，この方法は今後の結核対策のなかで重要な手段のひとつとなるであろう。
キーワーズ：集団感染，結核感染，ツベルクリン反応，全血インターフェロンγ応答測定法，予防内服

緒　言

結核感染の頻度が低下しているのに逆比例して，若年者を中心に結核の集団感染（集団発生）のリスクは大きくなっている。厚生労働省の資料[1]によれば2000年から2004年の初めまでの期間に同省に報告された集団感染事例170件のうち60%は学校および若者の多い事業所で起こっていた。このような事例では感染を受け未だ結核を発病していない者が多かれ少なかれ発見されるが，そのような者に対しては化学予防を行い，将来の発病の危険性を下げることが必要である。しかし日本のようにBCG接種が広範に行われている国においては，ツベルクリン反応検査（以下，ツ反）による正確な結核感染の診断は困難である。これに対して日本では，「BCG接種によるツ反応は，確率的に結核感染によるものよりも弱い」ということを前提に，かなり便宜的な判定基準を設け[2]，さらに過去の反応との比較やBCG瘢痕の状況，被験者が集団構成員であれば集団全体のツ反の状況など，といった「状況証拠」も考慮に入れた総合的な判断を行ってきた[3]。しかしこの方法はかなりの過不足（不必要な化学予防指示と必要な人の看過）を免れない。BCG接種の予防効果と引き替えに，われわれはこのような不都合に甘んじてきたが，同時にBCG既接種の影響を受けない正確な結核診断技術の開発はわれわれが長年希求してきたところである。

近年の技術革新により末梢血免疫細胞の特異抗原刺激に対するインターフェロンγ応答を定量することが可能になり，その簡易キット（QuantiFERON®-TB, Cellestis社，オーストラリア）が開発された[4]。さらにこれに結核菌特異抗原としてESAT-6，CFP-10というタンパクを使用することにより上記の課題がようやく解決に向かいつつある[5]。われわれはこの方法（QuantiFERON®TB-2G，同

財団法人結核予防会結核研究所

連絡先：原田登之，結核予防会結核研究所抗酸菌レファレンスセンター免疫検査科，〒204-8533 東京都清瀬市松山3-1-24
(E-mail: harada@jata.or.jp)
(Received 25 Aug. 2004/Accepted 14 Oct. 2004)

上；以下QFTと略記）をある集団感染事例に応用する機会を得たので，その知見について検討を加えて報告する。

対象と方法

（1）集団発生事例の概要

某中都市でそろばん塾・カラオケ教室を経営し，さらに経理関係の専門学校で講師をしていた男性A氏（57歳）が発端患者である。A氏は2002年11月初旬より風邪様症状を発症していたが，12月24日受診して結核（ガフキー9号，学会病型bⅢ2）と診断され，入院した。A氏は結核の既往はなく，1999年4月に近医で胸部X線撮影を受け，異常なしといわれたが，その後は健診等を受けていない。家族以外の接触者は専門学校生300人（年齢18歳～25歳），職員38人のほか，そろばん教室生徒57人（5歳～12歳），カラオケ教室生徒などであった。とくに専門学校生のうち，73人は2002年11月30日から12月5日にかけてA氏と共にハワイへの修学旅行に参加していた。

1月27日専門学校生のB氏（18歳，女性）が気胸を発症，つづいて喀痰結核菌PCR陽性と診断された。ついで接触者健診の結果，ハワイ旅行に参加した専門学校生C氏（20歳，男性）が2月3日に活動性結核（rⅢ1）と診断された。

接触者に対して行った健診（ツ反，X線検査，およびツ反歴）および本研究で行ったQFTの成績を参考として合計23人（専門学校生18人，そろばん教室生徒5人）に化学予防が指示された。その後2004年7月現在まで患者発生は報告されていない。

（2）QFTによる検査法

QFTの原理や方法の詳細は他[5][6]に譲り，ここでは要点のみを記述する。被験者から静脈血を5 mlヘパリン採血し，採血後概ね12時間以内にその一定量に抗原（ESAT-6，CFP-10；実際はそれぞれの合成ペプチドのプール），および陰性対照用生理食塩水，陽性対照としてPHAをそれぞれ作用させて，免疫細胞からインターフェロンγ（以下IFN-γ）を産生させる。37℃で16～24時間静置培養した上清をとり，含まれたIFN-γを，QFTキットを用いて，規定の使用法に従って測定した。測定されたそれぞれの抗原刺激によるIFN-γ計測値から陰性対照での計測値を差し引いた値を，対応する実効放出量として，いずれかの抗原での放出量が0.35 IU/ml以上のものを「陽性」と定義した。このカットオフについては別に論じた[7]。

（3）ツ反・QFTの対象・方法

ツ反は日本の現行の制度上化学予防の対象となる30歳未満の接触者である，そろばん教室および専門学校生と定め，通常の方法で実施した。判定は医師が発赤長径および硬結横径を測定した。二重発赤の有無，水疱等のいわゆる副反応の有無も記載した。本研究では専門学校生をQFT試験の観察対象としたので，以下この群のみについて記述する。

QFTは原則としてツ反で発赤長径が30 mmを超える者を対象とした。QFTの実施に先立ち，定期外健診を担当する職員から本人または保護者に対して説明を行い，承諾をとった。

（4）分析

ESAT-6，CFP-10による実効IFN-γ放出量にもとづいて各被験者は陽性と陰性に判別されるが，各放出量の分布を検討する場合はその値を対数変換したものを用いた。この際，元の値が0または負の場合には，仮に元の値が仮定的に0.005として集計・解析を進めた。

成　　績

（1）ツ反の結果

専門学校生270人について行われたツ反発赤径の分布はFig. 1に示すとおりである。やはり右方に長く裾を引く正規分布に近似した分布で，明らかな二峰性は認めない。平均値は25.8 mm（標準偏差16.1 mm）であり，発赤

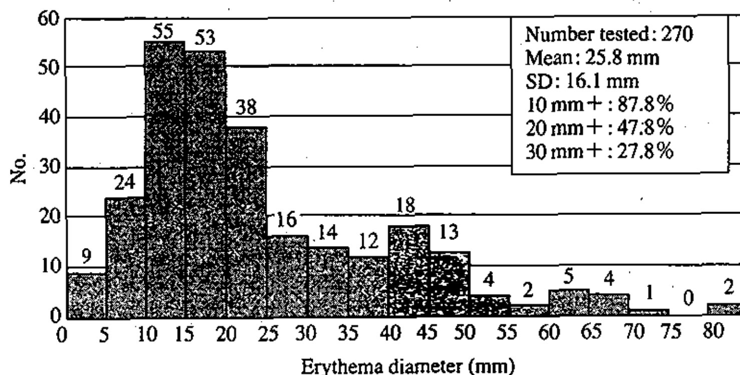

Fig. 1 Distribution of contacts by erythema size of tuberculin skin test (college students)

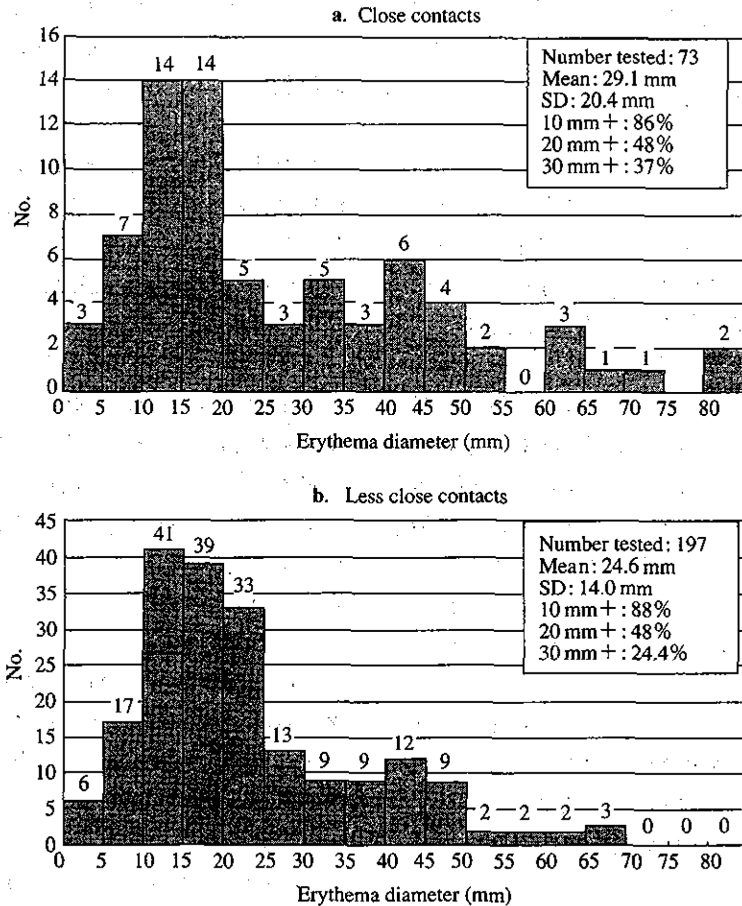

Fig. 2 Distribution of close and less close contacts according to erythema size of tuberculin skin test (college sudents)

Note: "Close contacts" include those who joined the excursion to Hawaii islands with the index case during 5 days, and thus considered to have been in closer contact with the infection source than "Less close contacts".

10 mm以上が87.8%，20 mm以上47.8%，30 mm以上が27.8%であった。平均年齢は19.7歳（標準偏差は1.8歳）であった。

この群をハワイ修学旅行への参加の有無によって二分し（それぞれ濃厚接触群，非濃厚接触群と呼ぶ），ツ反発赤分布を比較したのがFig. 2である。ヒストグラムを比較すると，両群はよく似ているが，ハワイ群（a. Close contacts）では非ハワイ群（b. Less close contacts）に比して右方の裾が厚くみえ，かすかながら二峰性とみることもできる。平均値は有意に前者で大きい（両側検定の$t=0.0407$，$p=0.041$）。30 mm以上に達する反応の頻度も37.0%対24.4%で，有意に前者で多い（非濃厚接触群を基準としたOdds比＝1.82，カイ自乗の$p=0.040$）。

（2）QFTの結果

QFT分析は75人に対して行われた。いずれもツ反被検査のうち発赤径30 mm以上の者である。これら両群のIFN-γ産生量の分布をESAT-6，CFP-10について見たのがFig. 3である。IFN-γの分布は対数正規分布をするので，元の値の自然対数変換を行っている。濃厚接触群の分布は明らかに二峰性であり，対数値－2（真数で0.14 IU/ml）近傍がantimodeとなっている。一方非濃厚接触者では分布はやや不規則な指数関数型の分布で右肩下がりである。0.35 IU/mlを基準として陽性率を見るとTableのようになる。

（3）IFN-γとツ反の相関

Fig. 4はIFN-γ（対数変換済み）とツ反発赤径の相関を示す。ツ反発赤径30 mm以上の者のみについてのQFT応答であるが，両者の間にはESAT-6，CFP-10のいずれも弱い相関がみられる（相関係数はそれぞれ0.387，0.207；無相関の検定のp値は0.000，0.074）。相関係数の自乗値（決定係数）はそれぞれ0.15，0.04で，ツ反発赤の変動の15%，4%だけがこれらの抗原への応答に依存するということになる。

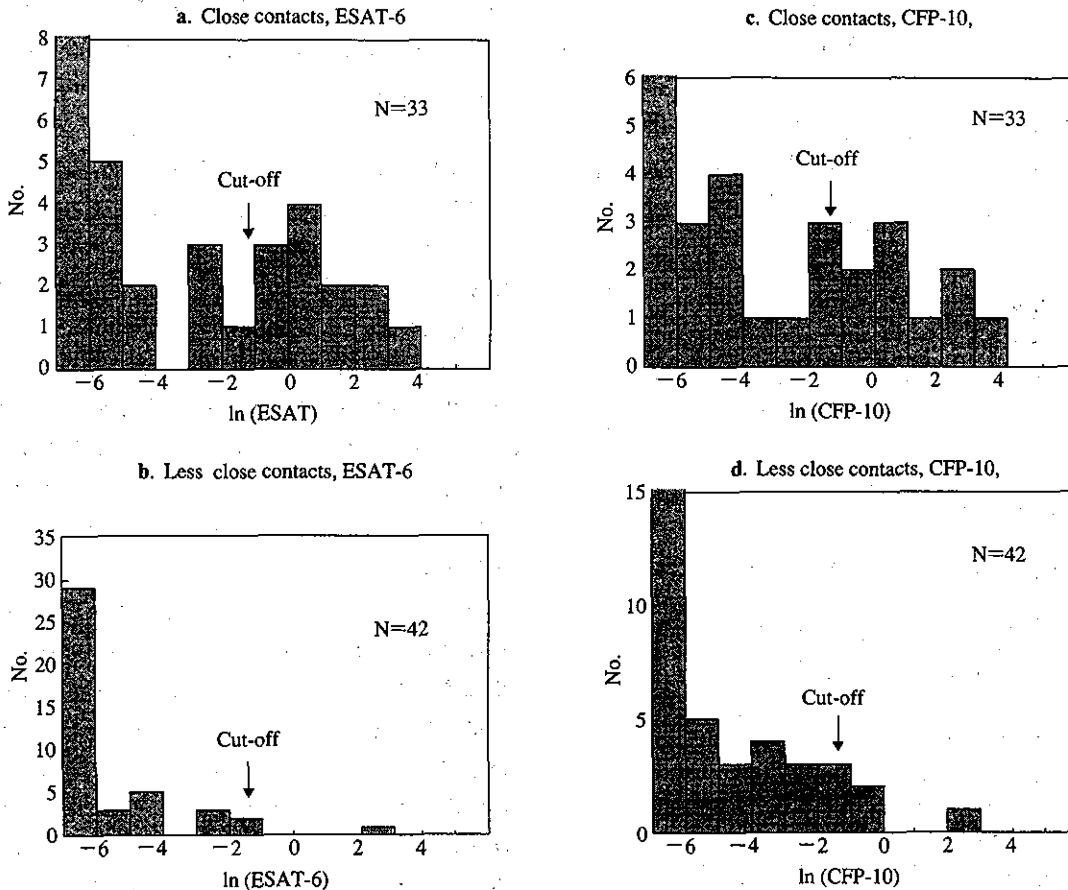

Fig. 3 Distribution of subjects according to log-transformed values of IFN-γ (close vs less close contacts)

Table QFT positivity among close contacts and less close contacts, Cut-off = 0.35

	Close contacts			Less close contacts			Comparison*	
	Tested	Positive	(%)	Tested	Positive	(%)	OR	p-value
ESAT-6		12	36.4		1	2.4	24.0	0.000
CFP-10	33	9	27.3	42	3	7.1	4.9	0.041
Either		15	45.5		3	7.1	12.8	0.000

*OR: Odds ratio. P-values are based on Chi-square test with Yates correction.

考　察

著者らは先にQFTの結核感染の診断における特性について感度89.0%，特異度98.1%という知見を得た[6]。ただその感度については，目標とする（潜在的）結核感染の代理指標として菌陽性肺結核（治療開始前の）を用いたもので，感染して未だ発病していない状態においてQFTがどのようなパフォーマンスを示すかを見たものではない。今回は比較的規模の大きい集団感染，しかも集団を感染曝露の濃淡の差で二分しうる状況でこの検査の所見を吟味したもので，間接的ながら潜在感染における特性についてより詳細に検討することができた。

対象となった専門学校生の集団は年齢が比較的均質な若者で構成されており，初発患者の症状出現からほぼ2カ月間接触があり，その一部は特に濃厚な接触関係（1週間のハワイ旅行に同行）を持っていた（ただし続発例がこの申告による症状出現時期から2カ月で発病していることから，初発患者の発病時期はこれよりさらに以前だった可能性もあり，真の接触期間はもっと長かったかもしれない）。これによって分けられた濃厚接触，非濃厚接触両群から活動性結核患者が続発しており，集団感染は何らかの程度で起こっていることは十分考えられた。そして，従来の接触者健診で行われるツ反検査の成績（濃厚接触群と非濃厚接触群の比較）において，前者でやや強い反応が見られることから，集団感染の存在がさらに疑われた。従来であれば，便宜的に設定されている「30 mm以上（いわゆる強反応）」を「潜在感染」の基準として用い，全接触者中から75人（28%）に対して化

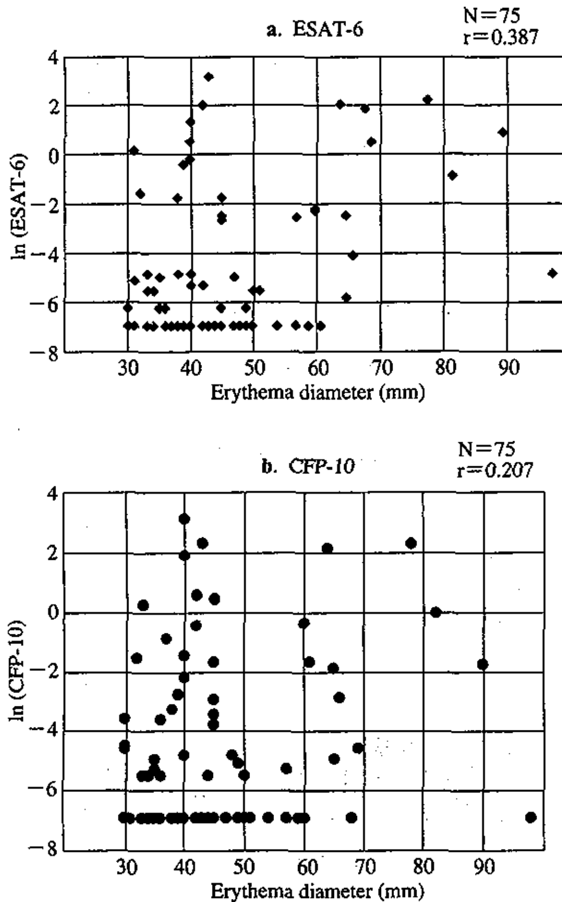

Fig. 4 Correlation between tuberculin reaction and IFN-γ response

学予防を指示することになったはずである。

QFTをこれらの者に適用したところ，2群の反応は感染曝露に対応した明確な違いを示した。濃厚接触群ではIFN-γ応答は感染の有無を示唆する2個の亜群からなる截然とした二峰性の分布を示し，非濃厚接触群では低頻度の被感染者の存在は未感染者の応答に覆われてはっきりしない分布となっていた。先に設定されたQFTの診断基準[7]から，濃厚接触群の45.5%，非濃厚接触群7.1%，あわせて24.0%が潜在感染の状態にあるとされた。これはもとの接触者全体からみると，濃厚接触群の20.5%，非濃厚接触群の1.5%，全体では6.7%（95%信頼区間4.3%～10.3%，実人数18人）に相当する。集団感染の対応としては，これに基づきこれら18人に化学予防が指示された。QFTを用いることによって化学予防の指示は従来法の75人からその4分の1に節減することができたことになる。

非濃厚接触者におけるQFT陽性率1.5%（95%信頼区間0.5%～4.3%）は，1998年の20～29歳に対して森[8]が推定した既感染率1.6%と類似しており，ここでは感染はそれほど大量に起こっていたわけではないことがうかがわれる。その一方で濃厚接触者群での陽性率20.5%（95%信頼区間12.9%～31.2%）がいかに濃密な感染伝播であったかを物語る。

ただし，ここには問題が少なくとも2点残されている。まずQFTの潜在感染に対する感度である。上に述べた結核患者におけるQFT陽性率は89%であり，潜在感染ではIFN-γ応答は患者よりも高いことがあっても低くはないとする論拠は少なくない[9)～11)]。それが正しければ，QFTでは（ツ反で30 mm以上の者の中から）最大10%程度の感染が見落とされることになる。本事例における化学予防の対象者は，抗原特異的IFN-γ産生量が0.35 IU/ml以上の者であったが，別に論じたようにこのような既感染率が高い集団に適応する際にはカットオフ値を下げることにより陽性的中率をあまり下げずに見落としを減らすことができる[7]。

第2の問題は，QFTの被験者をツ反30 mm以上に限定したことによる感染者の見落としの可能性である。結核患者におけるツ反は平均値を発赤30 mm近傍にもつ正規分布である[12]。もし潜在感染者のツ反がこれと同じ分布に従うとすれば，発赤30 mm以上に感染の疑いをかけることにより，分布の半数にあたる30 mm未満の反応の者を見落とすことになる。上の議論により潜在感染は患者よりも強いツ反を示す可能性があり，さらにわれわれが問題にしている既接種集団では，接種後の自然感染によるツ反は強くなることが考えられる。Fig. 1に見るように感染によって生じたと考えられる二峰性の右方のモードは40～45 mm付近にあることからもうなずける。このことはこれまでに報告された学校等での集団感染でも繰り返し観察されている[13]。これからすれば，30 mmの基準で見落とされるのは被感染者の半数には達しないとしても，いずれにせよかなりの見落としは免れないことを忘れることはできない。これは今後QFTを限定的に適用する場合に，考えなければならない問題である。ただし現行の化学予防基準と比較する場合には，ツ反発赤30 mm以上とすることは見落としを増やすことにはならない。いきなり接触者全員に本法を適用するまではしなくても，たとえば発赤20 mm以上に対し適用するなどの妥協を行うだけでもツ反による「見落とし」を飛躍的に回復することが可能になるであろう。

BCG既接種者を含む集団での接触者健診にQFTが有用なことを示す事例は最近デンマークからも報告されている[14]。長期間有症状だった高校生患者の接触者700人の健診で大半が未接種の高校生および多くの既接種者を含む成人について検査が行われ，未接種者ではQFTとツ反はよく一致し，また既接種者では濃厚接触者と非濃厚接種者の間でQFT陽性率の明らかな対比が見られた（陽性率は50%対6%）。既接種集団にはツ反皮膚反応で

はなく，PPDを抗原にしたIFN-γ応答（QFT-PPD）を見ているが，これでの陽性率は濃厚接触群，非濃厚接触群でそれぞれ38%，44%であり，皮膚反応と同様に曝露の影響を浮き彫りにすることはできなかった。

この報告もわれわれの事例と同様，QFTがBCG既接種の影響を受けずに結核感染をよく反映することを示しており，さらにわれわれが観察し得なかった未接種集団でのツ反との高い一致性をも示している。同じ特異抗原を用い，IFN-γ応答ではなく，感作リンパ球を検出する方法（ELISPOT）による感染診断も同様な特性を示している。Ewerら[15]は英国で発生した中学校での集団感染で535人の接触者を接触程度に5段階に分け，ツ反とELISPOT法を行い，後者による陽性率のほうが接触程度によく相関する成績をあげたことを見ている。

このようにQFTは日本のようにBCG既接種者の多い集団（さらには環境中抗酸菌感染の多い集団[16]）での結核感染の正確な診断に非常に有用な方法となると思われる。ただし，感染源への曝露の時期から反応が陽性になるまでの時間経過やさらに長期の応答の消長，化学予防や化学療法の影響など，まだその診断特性についても知見が十分ではない。また小児，特に幼児の場合の特性についても同様であり，さらに現行の最低採血量（5 ml）からくる制約の克服も課題である。さらに一般には採血してから12時間以内の処理の必要性という操作上の制約もある。これについては最近，採血管に抗原を固定しておき，採血と同時に抗原刺激が自動的に起こる第3世代QFT（http://www.cellestis.com）が開発されつつあり，いちおうの解決の可能性が見えてきた。

一方，新たな結核対策への転換の時期にあって，化学予防への積極的な取り組みは，米国[17]のような国に限らず日本でもますます重大な課題である。接触者健診に限らず，一般臨床の場でも結核発病のリスクにある者には化学予防を年齢に限らず行うべき時期に来ていると思われる。このような適用を決めるための感染診断は今後ますます重要な意味を持ってくると思われ，そのためにもQFTのような方法が確立され，広く利用されるようになることが望まれる。

まとめ

若年者集団での結核発生に際して結核菌特異抗原（ESAT-6およびCFP-10）を用いた全血インターフェロンγ応答を定量する方法（QuantiFERON®TB-2G）を試行したところ，感染源との接触の濃さの違いに応じた明確な陽性率の差が得られ，この方法はBCG接種の影響を受けない，特異度の高い結核感染診断の方法として有用であることが知られた。

謝辞

接触者健診の機会を利用させて頂き，諸般にわたり研究にご協力頂いた保健所職員の皆様，ならびに特異抗原を提供されたCellestis社（オーストラリア）に深謝する。

本研究は平成14年厚生労働科学研究補助金による新興再興感染症研究事業研究課題（主任研究者　森　亨），および平成15年度厚生労働科学研究補助金による新興再興感染症研究事業研究課題（主任研究者　森　亨）のそれぞれ一部として行われた。

本研究の成績の一部は第79回日本結核病学会総会（2004年，名古屋）で樋口が発表した。

文献

1）厚生労働省結核感染症課：「結核の統計2004」. 結核予防会, 東京, 2004.
2）厚生省保健医療局疾病対策課結核・感染症対策室長通知：初感染結核に対するINHの投与について. 平成元年2月28日健医感発第20号.
3）森　亨：ツベルクリン反応検査による予防投薬対象の決定.「命令入所及び初感染結核の取扱いとその説明」. 厚生省監修, 結核予防会, 1989, 117.
4）Streeton JA, Desem N, Jones SL: Sensitivity and specificity of a gamma interferon blood test for tuberculosis infection. Int J Tuberc Lung Dis. 1998 ; 2 : 443–450.
5）Andersen P, Munk ME, Pollock JM, et al.: Specific immune-based diagnosis of tuberculosis. Lancet. 2000 ; 356 : 1099–1104.
6）Mori T, Sakatani M, Yamagishi F, et al.: Specific detection of tuberculosis infection with an interferon-gamma based assay using new antigens. Am J Respir Crit Care Med. 2004 ; 170 : 59–64.
7）原田登之, 樋口一恵, 関谷幸江, 他：結核菌抗原ESAT-6およびCFP-10を用いた結核感染診断法QuantiFERON®TB-2Gの基礎的検討. 結核.（印刷中）
8）森　亨：結核の現状と対策. 日本内科学会雑誌. 2001 ; 90 : 111–115.
9）Hirsch CS, Toossi Z, Othieno C, et al.: Depressed T-cell interferon-gamma responses in pulmonary tuberculosis: analysis of underlying mechanisms and modulation with therapy. J Infect Dis. 1999 ; 180 : 2069–2073.
10）Pathan AA, Wilkinson KA, Klenerman P, et al.: Direct ex vivo analysis of antigen-specific IFN-gamma-secreting CD4T cells in *Mycobacterium tuberculosis*-infected individuals: associations with clinical disease state and effect of treatment. J Immunol. 2001 ; 167 : 5217–5225.
11）Vekemans J, Lienhardt C, Sillah JS, et al.: Tuberculosis contacts but not patients have higher gamma interferon responses to ESAT-6 than do community controls in The Gambia. Infect Immun. 2001 ; 69 : 6554–6557.
12）森　亨：「ツベルクリン反応検査」. 結核予防会, 東京, 1995, 27–29.

13) 森 亨：「結核の統計1984」. 厚生省保健医療局結核難病課編. 結核予防会, 東京, 1984, 11.
14) Brock I, Weldingh K, Lillebaek T, et al.: Comparison of tuberculin skin test and new specific blood test in tuberculosis contacts. Am J Respir Crit Care Med. 2004 ; 170 : 65–69.
15) Ewer K, Deeks J, Alvarez L, et al.: Comparison of T-cell-based assay with tuberculin skin test for diagnosis of *Mycobacterium tuberculosis* infection in a school tuberculosis outbreak. Lancet. 2003 ; 361 : 1168–1173.
16) Arend SM, van Meijgaarden KE, de Boer K, et al.: Tuberculin skin testing and *in vitro* T cell responses to ESAT-6 and culture filtrate protein 10 after infection with *Mycobacterium marinum* or *M. kansasii*. J Infect Dis. 2002 ; 186 : 1797–1807.
17) ATS/CDC Statement Committee on Latent Tuberculosis Infection: Targeted tuberculin testing and treatment of latent tuberculosis infection. MMWR 2000/49 (RR06) ; 1–54, 2000（邦訳 資料と展望. 2001 ; 36 : 25–68）.

Original Article

USEFULNESS OF A NOVEL DIAGNOSTIC METHOD OF TUBERCULOSIS INFECTION, QuantiFERON®TB-2G, IN AN OUTBREAK OF TUBERCULOSIS

Nobuyuki HARADA, Toru MORI, Shinji SHISHIDO, Kazue HIGUCHI, and Yukie SEKIYA

Abstract [Objective] The purpose of this study was to evaluate QuantiFERON®TB-2G (QFT), a novel method of detecting tuberculosis infection among contacts of a tuberculosis patient by determining the whole-blood interferon-gamma response to the specific antigens.

[Subjects and Methods] A teacher of a college who had been coughing for the preceding two months was diagnosed with smear-positive tuberculosis. About 270 students of the college were considered to have been exposed to tuberculosis infection, of whom 73 were in closer contact with the index case because they participated in a one-week group excursion attended by the teacher. Two of the contact students developed active tuberculosis shortly thereafter. Tuberculin tests were conducted to almost all students, and QFT was performed for only those with tuberculin reactions having erythema diameters of 30 mm or larger.

[Results] Tuberculin tests of students, all of whom had been vaccinated with BCG at least once, revealed that the distribution of the close contact group was slightly shifted to right (larger side) than those with less close contacts. The QFT positive rate for close contacts was 45.5%, while that for less close contacts was only 7.1%, which obviously indicates that QFT is hardly affected by the tuberculin allergy due to past BCG vaccination. The distribution of interferon-gamma measurements (log-transformed) of the close contacts showed typical bimodality, one mode representing the infected, another the non-infected. This was not clear for the less close contacts. The correlation of interferon-gamma measurements (log-transformed) with tuberculin reaction erythema size was weak, if not non-significant.

[Conclusion] It was concluded that QFT was a useful method for diagnosing tuberculosis infection and was unaffected by the BCG-caused tuberculin allergy. In the case of the outbreak mentioned above, QFT greatly reduced the indication of chemoprophylaxis, from 28% of all the contacts solely based on tuberculin test to only 7%.

Although there remains some problems to be overcome for QFT to be widely used with high confidence, this technology will provide a high possibility for wider and more accurate indication of chemoprophylaxis and will be one of the essential tools of tuberculosis control of the 21st century in Japan.

Key words: Tuberculosis outbreak, Latent tuberculosis infection, Tuberculin skin test, Whole blood interferon-gamma assay, Chemoprophylaxis

Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association (JATA)

Correspondence to: Nobuyuki Harada, Immunology Division, Research Institute of Tuberculosis, JATA, 3-1-24, Matsuyama, Kiyose-shi, Tokyo 204-8533 Japan.
(E-mail: harada@jata.or.jp)

JOURNAL OF CLINICAL MICROBIOLOGY, Aug. 2004, p. 3538–3548 Vol. 42, No. 8
0095-1137/04/$08.00+0 DOI: 10.1128/JCM.42.8.3538–3548.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

Characterization of a Trinucleotide Repeat Sequence $(CGG)_5$ and Potential Use in Restriction Fragment Length Polymorphism Typing of *Mycobacterium tuberculosis*

Yayoi Otsuka,[1] Pawel Parniewski,[2,3] Zofia Zwolska,[4] Masatake Kai,[1]† Tomoko Fujino,[1] Fumiko Kirikae,[1] Emiko Toyota,[1] Koichiro Kudo,[1] Tadatoshi Kuratsuji,[1] and Teruo Kirikae[1]*

International Medical Center of Japan, Shinjuku, Tokyo 162-8655, Japan,[1] and Centre for Medical Biology and Microbiology, Polish Academy of Science, Lodz 93-232,[2] Centre for Medical Biology and Microbiology, Swietokrzyska Academy, Kielce 25-369,[3] and National Research Institute of Tuberculosis and Lung Diseases, Warsaw 01-138,[4] Poland

Received 16 November 2003/Returned for modification 10 February 2004/Accepted 13 April 2004

The genomes of 28 bacterial strains, including mycobacterial species *Mycobacterium tuberculosis* and *Mycobacterium bovis*, were analyzed for the presence of a special class of microsatellite, that of trinucleotide repeat sequences (TRS). Results of a search of all 10 possible TRS motifs (i.e., CCT, CGG, CTG, GAA, GAT, GTA, GTC, GTG, GTT, and TAT) with five or more repeating units showed that $(CGG)_5$ was highly represented within the genomic DNA of *M. tuberculosis* and *M. bovis*. Most of the $(CGG)_5$ repeats in the genome were within the open reading frames of two large gene families encoding PE_PGRS and PPE proteins that have the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE). $(CGG)_5$-probed Southern hybridization showed that some mycobacterial species, such as *Mycobacterium marinum*, *Mycobacterium kansasii*, and *Mycobacterium szulgai*, possess many copies of $(CGG)_5$ in their genomes. Analysis of clinical isolates obtained from Tokyo and Warsaw with both IS*6110* and $(CGG)_5$ probes showed that there is an association between the fingerprinting patterns and the geographic origin of the isolates and that $(CGG)_5$ fingerprinting patterns were relatively more stable than IS*6110* patterns. The $(CGG)_5$ repeat is a unique sequence for some mycobacterial species, and $(CGG)_5$ fingerprinting can be used as an epidemiologic method for these species as well as IS*6110* fingerprinting can. If these two fingerprinting methods are used together, the precise analysis of *M. tuberculosis* isolates will be accomplished. $(CGG)_5$-based fingerprinting is particularly useful for *M. tuberculosis* isolates with few or no insertion elements and for the identification of other mycobacterial species when informative probes are lacking.

DNA fingerprinting of the inserted IS*6110* element specific for the *Mycobacterium tuberculosis* complex is a powerful epidemiological tool for visualizing DNA restriction fragment length polymorphisms (RFLP) of *M. tuberculosis* (26). The major limitation of IS*6110*-based RFLP typing is the difficulty of discriminating genetic polymorphisms of *M. tuberculosis* isolates with only a few copies of the element. In addition, there are two reports (1, 29) that described IS*6110*-based RFLP as unstable, although other studies have confirmed a high degree of stability (5, 15). Yeh and colleagues (29) indicated that genotypes with IS*6110* were relatively unstable because they changed rapidly compared with those based on another marker. Alito et al. (1) reported that a multidrug-resistant outbreak strain changed rapidly, according to IS*6110* RFLP, over a period of a few years.

A number of alternative typing methods for *M. tuberculosis* isolates that use genetic markers, such as polymorphic GC-rich repetitive sequences (PGRS) (19), tandem repeat sequences of 10 bp found in PPE family proteins (10), the direct repeat (9), a $(GTG)_5$ repeat (28), IS*1547* (6), *katG* (30), and tandem repeats of 40 to 100 bp (14, 24), have been reported.

Trinucleotide repeat sequences (TRS) comprise a class of microsatellites that are involved in human neurodegenerative diseases (27). Studies in *Escherichia coli* showed that these TRS, such as $(CTG)_n$ and $(CGG)_n$, may effect genetic instability during DNA replication, transcription, and repair processes (17). $(GAA)_{12}$ has been found in a plasmid of *Mycoplasma gallisepticum* (12, 13), and it positively regulates gene expression in this plasmid. It is not as well known whether bacterial genomes possess tandem repeat sequences. The types, lengths, and distribution of such sequences may serve as valuable markers for phylogenetic or epidemiologic studies of various bacteria.

In the present study, we searched for all possible TRS in various bacterial strains and found that *M. tuberculosis* and *Mycobacterium bovis* possess many $(CGG)_5$ repeats. We also analyzed *M. tuberculosis* clinical isolates obtained from Japan and Poland with $(CGG)_5$-based DNA fingerprinting and show that this method is useful for the genetic analysis of clinical isolates of *M. tuberculosis*.

* Corresponding author. Mailing address: Department of Infectious Diseases and Tropical Medicine, International Medical Center of Japan, Toyama 1-21-1, Shinjuku, Tokyo 162-8655, Japan. Phone: 81 3 3202 7181, ext. 2838. Fax: 81 3 3202 7364. E-mail: tkirikae@ri.imcj.go.jp.

† Present address: Department of Anatomy and Developmental Biology, University College London, London WC1E 6BT, United Kingdom.

MATERIALS AND METHODS

Bacterial strains. The sources of mycobacterial strains used in this study are listed in Table 1. Clinical isolates were obtained from the International Medical

TABLE 1. Mycobacterial strains used in this study

Strain	Property or origin[a]	Source or reference
M. tuberculosis H37Rv	ATCC 27294	ATCC
M. tuberculosis H37Ra	ATCC 25177	ATCC
M. abscessus	Clinical isolate (IMCJ 268)	IMCJ
M. avium	ATCC 25291	ATCC
M. bovis BCG	Japanese strain 172	Japan BCG Laboratory
M. chelonae	JCM 6390 (ATCC 14472)	JCM
M. fortuitum	Clinical isolate (IMCJ 531)	IMCJ
M. gastri	GTC 610 (ATCC 15754)	GTC
M. intracellulare	JCM 6384 (ATCC 13950)	JCM
M. kansasii	JCM 6379 (ATCC 12478)	JCM
M. marinum	GTC 616 (ATCC 927)	GTC
M. nonchromogenicum	JCM 6364 (ATCC 19530)	JCM
M. peregrinum	Clinical isolate (IMCJ 460)	IMCJ
M. scrofulaceum	JCM 6381 (ATCC 19981)	JCM
M. simiae	GTC 620 (ATCC 25275)	GTC
M. smegmatis	ATCC 19420	ATCC
M. szulgai	JCM 6383 (ATCC 35799)	JCM
M. terrae	GTC 623 (ATCC 15755)	GTC
M. xenopi	Clinical isolate (IMCJ 788)	IMCJ

[a] JCM, Japan Collection of Microorganisms, The Institute of Physical and Chemical Research (RIKEN), Saitama, Japan; GTC, Gifu Type Culture Collection, Department of Microbiology-Bioinformatics, Regeneration and Advanced Medical Science, Gifu University, Graduate School of Medicine, Bacterial Genetic Resources, Gifu, Japan.

Center of Japan (IMCJ) in Tokyo, Japan, in 2001 and from the National Research Institute of Tuberculosis and Lung Diseases in Warsaw, Poland, in 2000. These clinical isolates were obtained from different patients. Drug susceptibility testing was performed by conventional culture on solid media with a proportion method (Wellpack; Japan BCG Laboratory, Tokyo, Japan) or by a microdilution method with Vit spectrum SR (Kyokuto Pharmaceutical Co., Ltd., Tokyo, Japan). The antituberculosis drugs tested and the concentrations used were as follows: isoniazid, 0.2 and 1.0 μg/ml; rifampin, 40 μg/ml; ethambutol, 2.5 μg/ml; streptomycin, 10 μg/ml; *para*-aminosalicylic acid, 0.5 μg/ml; cycloserine, 30 μg/ml; ethionamide, 20 μg/ml; kanamycin, 20 μg/ml; enviomycin, 20 μg/ml; and levofloxacin, 1.0 μg/ml. Drug resistance is defined as resistance to at least one drug. Serial cultures were made from *M. tuberculosis* strain H37Rv and a clinical isolate from Japan (IMCJ 541) and were passaged weekly over 9 weeks.

Genome sequence. The genome sequences of 28 bacterial strains were downloaded from the National Center for Biotechnology Information GenBank database (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/Complete.html), The Institute for Genomic Research website (http://www.tigr.org/CMR), the Sanger Center (http://www.sanger.ac.uk), and the DNA Data Bank of Japan (http://www.ddbj.nig.ac.jp).

Isolation and restriction enzyme digestion of mycobacterial DNA. Chromosomal DNA of the mycobacterial strains and *M. tuberculosis* clinical isolates were prepared as described previously (16, 26) with slight modifications. Briefly, for isolation of genomic DNA, *M. tuberculosis* strains were grown on egg-based Ogawa solid medium (Kyokuto Pharmaceutical Co., Ltd.) for 3 to 5 weeks. All bacterial cells from one slant were transferred to 400 μl of TE buffer (0.01 M Tris-HCl, 0.001 M EDTA [pH 8.0]), and the solution was heated at 80°C for 20 min to kill the bacteria. Fifty microliters of lysozyme (10 mg/ml) was added, and the tube was incubated overnight at 37°C. Seventy microliters of sodium dodecyl sulfate (10%) and 5 μl of proteinase K (10 mg/ml) were added, and the mixture was incubated for 10 min at 65°C. A 100-μl volume of 5 M NaCl and the same volume of an *N*-cetyl-*N*,*N*,*N*-trimethylammonium bromide (CTAB)-NaCl solution (4.1 g of NaCl and 10 g of CTAB per 100 ml) were added together. The tubes were vortexed and incubated for 10 min at 65°C. An equal volume of chloroform-isoamylalcohol (24:1) was added, the mixture was centrifuged for 5 min at 12,000 × *g*, and the aqueous supernatant was carefully transferred to a fresh tube. The total DNA was precipitated in isopropanol and was redissolved in 20 μl of 0.1× TE buffer. All restriction enzymes used in this study, AatII, AfaI, AluI, EcoRI, HinfI, MluI, NruI, NsbI, PstI, PvuII, SacI, Sau3AI, SalI, SmaI, XhoI, and XspI, were purchased from Takara Bio Inc. (Shiga, Japan). Chromosomal DNA was digested overnight with each restriction enzyme (1 U/μg of DNA) under the conditions specified by the manufacturer. The digested fragments were separated by electrophoresis on horizontal 1% agarose gel at 15 V for 20 h (14-cm gel) in 1× TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA). A 1-kb DNA ladder and λ DNA restricted with HindIII (Promega Corp., Madison, Wis.) were used as size markers. The gels were then stained with ethidium bromide, and the results were recorded photographically.

Southern blotting. Gels were depurinated in 0.25 M HCl for 30 min and then denatured in 0.5 M NaOH and 1.5 M NaCl for 30 min. DNA fragments were transferred to an N$^+$ Hybond membrane (Amersham Biosciences, Little Chalfont, Buckinghamshire, United Kingdom) overnight, and the DNA was fixed to the membrane by UV irradiation.

The IS*6110* probe used in this study was a 245-bp DNA fragment amplified by PCR as described previously (26). Briefly, oligonucleotides INS1 (5′-CGTGAGGGCATCGAGGTGGC-3′) and INS2 (5′-GCGTAGGCGTCGGTGACAAA-3′) were used to amplify a 245-bp fragment from purified chromosomal *M. bovis* BCG DNA by PCR. The 15-mer oligonucleotide (CGG)$_5$, 5′-CGGCGGCGGCGGCGG-3′, was synthesized (Nippn TechnoCluster, Inc., Tokyo, Japan). These probes were labeled with horseradish peroxidase by the ECL direct system (Amersham Biosciences). Hybridization and detection were performed according to the recommendations of the manufacturer. Autoradiographs were obtained by exposing the membrane to X-ray film.

Analysis. IS*6110*- and (CGG)$_5$-based fingerprinting patterns were analyzed with Molecular Analyst Fingerprinting Plus software, version 1.6 (Bio-Rad Laboratories, Inc., Hercules, Calif.). To facilitate the comparison of the fingerprinting patterns, normalization was carried out with the use of molecular weight standards and the IS*6110*- or (CGG)$_5$-fingerprinting patterns of two clinical isolates, IMCJ 541 and a Poland-derived isolate, no. 28 (P 28), on each gel. Each dendrogram was calculated with the unweighted pair group method with average linkage according to the supplier's instructions.

RESULTS

Presence of TRS in mycobacterial strains and other bacterial species. To detect TRS among bacterial genomes and to determine the types of TRS and their repeat sizes, we searched for all 10 possible TRS motifs (i.e., CCT, CGG, CTG, GAA, GAT, GTA, GTC, GTG, GTT, and TAT) of five or more repeating units with the BLASTN algorithm (2). Among 28 bacterial strains, the numbers of TRS displayed large variation, with values ranging from zero to 38 (shown in the extreme right column in Table 2). *M. tuberculosis* strains H37Rv and CDC1551 and *M. bovis* possessed markedly more TRS copies than other species examined. The majority of the other species possessed fewer than 10 copies. Five strains, *Listeria innocua*, *Listeria monocytogenes*, *Staphylococcus aureus* N315, *Thermoplasma acidophilum*, and *Thermoplasma volcanium*, did not possess any TRS. The types of TRS varied (Table 2). (CCT)$_5$ did not exist in any of the bacteria examined in this study. CGG repeats, predominantly (CGG)$_5$, existed with high frequency in the genomes of *M. tuberculosis* strains H37Rv and CDC1551 and *M. bovis*; the frequencies of the appearance of CGG with five or more repeats were one per 150 to 200 kb. *Neisseria meningitidis* MC58 and *Pseudomonas aeruginosa* possessed six copies of (CGG)$_5$ with a frequency of one copy per 380 kb and five copies with a frequency of one copy per 1,250 kb, respectively. Few (CGG)$_5$ repeats were found in *E. coli* K12-MG1655, *E. coli* O157:H7 EDL933, *E. coli* O157:H7 VT2-Sakai, *N. meningitidis* serogroup A Z2491, *Salmonella enterica*, and *S. enterica* serovar Typhimurium. There were no (CGG)$_5$ repeats in *Clostridium acetobutylicum*, *Clostridium perfringens*, *Helicobacter pylori* 26695, *H. pylori* J99, *L. innocua*, *L. monocytogenes*, *Mycobacterium leprae*, *Mycoplasma genitalium*, *Mycoplasma pneumoniae*, *Mycoplasma pulmonis*, *Rickettsia conorii*, *Rickettsia prowazekii*, *S. aureus* Mu50, *S. aureus* N315, *T. acidophilum*, *T. volcanium*, and *Yersinia pestis*. Other possible repeats of CTG, GAA, GAT, GTA, GTC, GTG, GTT, and TAT were found sporadically among various bacterial strains.

TABLE 2. Distribution among bacterial genomes of TRS with five or more repeats

Microorganism (genome size [bp]) (GenBank accession no.)	No. of triplet repeats (no. of TRS copies per genome)										Total no. of TRS copies
	CCT	CGG	CTG	GAA	GAT	GTA	GTC	GTG	GTT	TAT	
C. acetobutylicum ATCC 824 (3940880) (NC_003030)			5 (1)	5 (1)					5 (1)	5 (2) 6 (2)	7
C. perfringens 13 (3031430) (NC_003366)										5 (3) 6 (1)	4
E. coli K12-MG1655 (4639221) (NC_000913)		5 (2)									2
E. coli O157:H7 EDL933 (5528445) (NC_002655)		5 (1)							5 (1)		2
E. coli O157:H7 VT2-Sakai (5498450) (NC_002695)		5 (1)							5 (1)		2
H. pylori 26695 (1667867) (NC_000915)						8 (1)			5 (1)		2
H. pylori J99 (1643831) (NC_000921)				5 (1)		8 (1)		5 (1)	5 (1)		4
L. innocua CLIP 11262 (3011208) (NC_003212)											0
L. monocytogenes EGD-e (2944528) (NC_003210)											0
M. bovis AF2122/97 (4345492) (NC_002945)		5 (22)			5 (1)		5 (1)	5 (4)			28
M. leprae TN (3268203) (NC_002677)				21 (1)	5 (2)	5 (3) 9 (1)		5 (1) 6 (1)			9
M. tuberculosis CDC1551 (4403836) (NC_002755)		5 (32) 6 (1)			5 (1)		5 (1)	5 (3)			38
M. tuberculosis H37Rv lab strain (4411529) (NC_000962)		5 (27) 6 (1) 7 (1)			5 (1)		5 (1)	5 (3)			34
M. genitalium G-37 (580074) (NC_000908)				5 (1) 6 (1) 16 (1)		5 (1) 7 (1) 8 (1) 9 (1) 10 (1) 11 (1) 16 (1)			11 (1)		11
M. pneumoniae M129 (816394) (NC_000912)						5 (1) 7 (1)					2
M. pulmonis UAB CTIP (963879) (NC_002771)									5 (2)	6 (1)	3
N. meningitidis MC58 (2272351) (NC_003112)		5 (6)									6
N. meningitidis serogroup A Z2491 (2184406) (NC_003116)		5 (2) 6 (1)									3
P. aeruginosa PA01 (6264403) (NC_002516)		5 (5)						5 (2) 14 (1)			8
R. conorii Malish 7 (1268755) (NC_003103)										5 (1)	1
R. prowazekii Madrid E (1111523) (NC_000963)								5 (1)		5 (1)	2
S. enterica serovar Typhi CT18 (4809037) (NC_003198)		5 (2) 6 (1)	10 (1)					5 (1)			5
S. enterica serovar Typhimurium LT2 SGSC1412 (4857432) (NC_003197)		5 (4)	5 (2)								6
S. aureus Mu50 (2878040) (NC_002758)					5 (1)						1
S. aureus N315 (2160837) (NC_002745)											0
T. acidophilum DSM 1728 (1564906) (NC_002578)											0
T. volcanium GSS1 (1584804) (NC_002689)											0
Y. pestis CO92 (4653728) (NC_003143)			5 (1) 6 (1)					5 (2)			4

However, only a few copies of these TRS were found. For example, one copy of $(CTG)_5$ was found in *C. acetobutylicum*, one $(CTG)_{10}$ was found in *S. enterica* serovar Typhi, two $(CTG)_5$ repeats were found in *S. enterica* serovar Typhimurium, and one $(CTG)_5$ and one $(CTG)_6$ repeat were found in *Y. pestis*. Relatively large TRS with 21 or 16 repeats were detected in *M. leprae* and *Mycoplasma genitalium*, respectively. *M. genitalium* possessed three types of TRS repeats (GAA, GTA, and GTT) and different numbers of repeats [$(GAA)_5$, $(GAA)_6$, and $(GAA)_{16}$; $(GTA)_5$, $(GTA)_7$, $(GTA)_8$, $(GTA)_9$, $(GTA)_{10}$, $(GTA)_{11}$, and $(GTA)_{16}$; and $(GTT)_{11}$].

Positions of $(CGG)_5$, $(CGG)_6$, and $(CGG)_7$ in the genome. The *M. tuberculosis* and *M. bovis* genomes consist of 4.4 and 4.3 Mb, respectively. All $(CGG)_5$, $(CGG)_6$, and $(CGG)_7$ repeats in both *M. tuberculosis* strains H37Rv and CDC1551 were located between 0.05 and 4.0 Mb (Table 3). These repeats appeared to be distributed randomly. In strain H37Rv, one $(CGG)_7$ was located at 0.05 Mb, and one $(CGG)_6$ was located at 2.4 Mb. Five $(CGG)_5$ repeats were between 0.1 and 1.0 Mb, six were between 1.0 and 2.0 Mb, eight were between 2.0 and 3.0 Mb, and eight were between 3.0 and 4.4 Mb. In strain CDC1551, one $(CGG)_6$ repeat was located at 0.05 Mb. Six $(CGG)_5$ repeats were between 0.1 and 1.0 Mb, 6 were between 1.0 and 2.0 Mb, 11 were between 2.0 and 3.0 Mb, and 9 were between 3.0 and 4.4 Mb. In *M. bovis*, four $(CGG)_5$ repeats were located between 0.26 and 1.0 Mb, five were between 1.0 and 2.0 Mb, seven were between 2.0 and 3.0 Mb, and six were between 3.0 and 4.3 Mb (Table 3). Almost all of the $(CGG)_5$, $(CGG)_6$, and $(CGG)_7$ repeats in *M. tuberculosis* and *M. bovis* were located within the open reading frame (ORF), with the exception of six $(CGG)_5$ repeats that were located between 1.1 and 3.96 Mb in strain CDC1551. Among these, the four $(CGG)_5$ repeats at 1.09, 3.74, 3.76, and 3.96 Mb were in the putative ORF with authentic frameshift or point mutation (Table 3).

In strain H37Rv, the genes containing $(CGG)_5$ and $(CGG)_6$ encoded the PPE and PE_PGRS families of proteins. A gene

TABLE 3. Position of $(CGG)_5$, $(CGG)_6$, and $(CGG)_7$ within the genome in three mycobacterial strains

Strain and position (bp)	No. of repeats	Gene no.	Product	Domain	Translation
M. tuberculosis					
55532	7	Rv0050	Probable penicillin-binding protein, PonA	ORF	poly(Pro)
261808	5	Rv0218	Hypothetical protein	ORF	poly(Ala)
340616	5	Rv0280	PPE family protein	ORF	poly(Ala)
362891	5	Rv0297	PE_PGRS family protein	ORF	poly(Gly)
672720	5	Rv0578c	PE_PGRS family protein	ORF	poly(Gly)
968964	5	Rv0872c	PE_PGRS family protein	ORF	poly(Gly)
1091589	5	Rv0977	PE_PGRS family protein	ORF	poly(Gly)
1189183	5	Rv1067c	PE_PGRS family protein	ORF	poly(Gly)
1189430	5	Rv1068c	PE_PGRS family protein	ORF	poly(Gly)
1191358	5	Rv1068c	PE_PGRS family protein	ORF	poly(Gly)
1213387	5	Rv1087	PE_PGRS family protein	ORF	poly(Gly)
1631645	5	Rv1450c	PE_PGRS family protein	ORF	poly(Gly)
2357161	5	Rv2098c	PE_PGRS family protein	ORF	poly(Gly)
2357267	6	Rv2098c	PE_PGRS family protein	ORF	poly(Gly)
2387312	5	Rv2126c	PE_PGRS family protein	ORF	poly(Gly)
2423539	5	Rv2126c	PE_PGRS family protein	ORF	poly(Gly)
2639030	5	Rv2356c	PPE family protein	ORF	poly(Ala)
2639330	5	Rv2356c	PPE family protein	ORF	poly(Ala)
2639442	5	Rv2356	PPE family protein	ORF	poly(Ala)
2802267	5	Rv2490c	PE_PGRS family protein	ORF	poly(Gly)
2922778	5	Rv2591	PE_PGRS family protein	ORF	poly(Gly)
3528969	5	Rv3159c	PPE family protein	ORF	poly(Ala)
3752989	5	Rv3347c	PPE family protein	ORF	poly(Ala)
3766907	5	Rv3350c	PPE family protein	ORF	poly(Ala)
3802146	5	Rv3388	PE_PGRS family protein	ORF	poly(Gly)
3803514	5	Rv3388	PE_PGRS family protein	ORF	poly(Gly)
3969420	5	Rv3532	PPE family protein	ORF	poly(Ala)
3972241	5	Rv3533c	PPE family protein	ORF	poly(Ala)
4029032	5	Rv3587c	Hypothetical protein	ORF	poly(Pro)
M. tuberculosis CDC1551					
55478	6	MT0056	Penicillin-binding protein	ORF	poly(Pro)
261924	5	MT0228	Hypothetical protein	ORF	poly(Ala)
340680	5	MT0292	PPE family protein	ORF	poly(Ala)
362955	5	MT0311	PE_PGRS family protein	ORF	poly(Gly)
674173	5	MT0607	PE_PGRS family protein	ORF	poly(Gly)
927976	5	MT0855	PE_PGRS family protein	ORF	poly(Gly)
968979	5	MT0894	PE_PGRS family protein	ORF	poly(Gly)
1091604	5	MT1004	Putative; PE_PGRS family protein, authentic frame shift	ORF	poly(Gly)
1189231	5	MT1096.1	PE_PGRS family protein	ORF	poly(Gly)
1189478	5	MT1096.1	PE_PGRS family protein	ORF	poly(Gly)
1191406	5	MT1097	PE_PGRS family protein	ORF	poly(Gly)
1213545	5	MT1118.1		UTR	poly(Gly)
1631528	5	MT1497.1	PE_PGRS family protein	ORF	poly(Gly)
2359430	5	MT2159	PE family-related protein	ORF	poly(Gly)
2359536	5	MT2159	PE family-related protein	ORF	poly(Gly)
2385890	5	MT2184 ?	Conserved hypothetical protein ?	Terminator ?	poly(Gly)
2422232	5	MT2220	PE_PGRS family protein	ORF	poly(Gly)
2633780	5	MT2423	PPE family protein	ORF	poly(Ala)
2634080	5	MT2423	PPE family protein	ORF	poly(Ala)
2636362	5	MT2425	PPE family protein	ORF	poly(Ala)
2636662	5	MT2425	PPE family protein	ORF	poly(Ala)
2636774	5	MT2425	PPE family protein	ORF	poly(Ala)
2797756	5	MT2564	PE_PGRS family protein	ORF	poly(Gly)
2918923	5	MT2668.1	PE_PGRS family protein	ORF	poly(Gly)
3524456	5	MT3247	PPE family protein	ORF	poly(Ala)
3526605	5	MT3248	PPE family protein	ORF	poly(Ala)
3745224	5	MT3453	Putative; PPE family protein, authentic frame shift	ORF	poly(Ala)
3759134	5	MT3458	Putative; PPE family protein, authentic frame shift	ORF	poly(Ala)
3793024	5	MT3495	PE_PGRS family protein	ORF	poly(Gly)
3794392	5	MT3495	PE_PGRS family protein	ORF	poly(Gly)
3961566	5	MT3636	Putative; PPE family protein, authentic point mutation	ORF	poly(Gly)
3964387	5	MT3637	PPE family protein	ORF	poly(Ala)
4021174	5	MT3693	Hypothetical protein	ORF	poly(Pro)
M. bovis					
262035	5	Mb0224	Probable conserved transmembrane protein	ORF	poly(Ala)
341620	5	Mb0288	PPE family protein	ORF	poly(Ala)
363940	5	Mb0305	PE_PGRS family protein	ORF	poly(Gly)
673964	5	Mb0593c	PE_PGRS family protein	ORF	poly(Gly)
1092029	5	Mb1002	PE_PGRS family protein	ORF	poly(Gly)
1189891	5	Mb1096c	PE_PGRS family protein	ORF	poly(Gly)
1192527	5	Mb1097c	PE_PGRS family protein	ORF	poly(Gly)

Continued on following page

TABLE 3—*Continued*

Strain and position (bp)	No. of repeats	Gene no.	Product	Domain	Translation
1214721	5	Mb1116	PE_PGRS family protein	ORF	Poly(Gly)
1627966	5	Mb1485c	PE_PGRS family protein	ORF	Poly(Gly)
2339003	5	Mb2125c	Conserved hypothetical protein PE_PGRS family protein	ORF	Poly(Gly)
2339109	5	Mb2125c	Conserved hypothetical protein PE_PGRS family protein	ORF	Poly(Gly)
2367710	5	Mb2150c	Conserved hypothetical protein PE_PGRS family protein	ORF	Poly(Gly)
2604931	5	Mb2376c	PPE family protein	ORF	Poly(Ala)
2607401	5	Mb2377c	PPE family protein	ORF	Poly(Ala)
2607513	5	Mb2377c	PPE family protein	ORF	Poly(Ala)
2769065	5	Mb2517c	PE_PGRS family protein	ORF	Poly(Gly)
3706526	5	Mb3380c	PPE family protein	ORF	Poly(Ala)
3720437	5	Mb3385c	PPE family protein	ORF	Poly(Ala)
3755777	5	Mb3420	PE_PGRS family protein	ORF	Poly(Gly)
3912639	5	Mb3562	PPE family protein	ORF	Poly(Ala)
3915460	5	Mb3563c	PPE family protein	ORF	Poly(Ala)
3972250	5	Mb3618c	Probable conserved membrane protein	ORF	Poly(Pro)

containing $(CGG)_7$, PonA, encoded a penicillin-binding protein (Table 3). In strain CDC1551, the genes containing $(CGG)_5$ encoded the PPE, PE_PGRS, and PE families of proteins. A gene containing $(CGG)_6$ encoded a penicillin-binding protein (Table 3). In *M. bovis*, all genes containing $(CGG)_5$ encoded PPE and PE_PGRS family proteins, with the exception of two genes that encoded probable conserved membrane proteins (Table 3). In all three strains, the $(CGG)_5$ in the PPE genes translated to poly(Ala), and the $(CGG)_5$ and $(CGG)_6$ in the PE_PGRS and PE genes translated to poly(Gly). In both *M. tuberculosis* strains, the $(CGG)_6$ and $(CGG)_7$ in genes encoding penicillin-binding proteins translated to poly(Pro) (Table 3). In *M. bovis*, the two $(CGG)_5$ repeats in genes encoding probable conserved membrane proteins translated to poly(Ala) and poly(Pro) (Table 3). Most of the $(CGG)_5$ repeats within the PPE genes were located in the N-terminal PPE domain of the genes (data not shown). All $(CGG)_5$ and $(CGG)_6$ repeats within the PE_PGRS genes consisting of PE and PGRS domains were located in the PGRS domain (data not shown). Two $(CGG)_5$ repeats within the PE family-related gene (MT2159) in strain CDC1551 were located in the C-terminal domain of the genes (data not shown).

Genomic stability. To examine whether $(CGG)_5$ repeats in the genome are stable, two *M. tuberculosis* strains (H37Rv and IMCJ 541) were analyzed for $(CGG)_5$- and IS*6110*-probed fingerprints. The fingerprint patterns among culture periods were identical for strain H37Rv (Fig. 1A). These findings were confirmed with strain IMCJ 541 (Fig. 1B). The data indicate that $(CGG)_5$ repeats are stable in the genome for at least a few months. In the IS*6110*-probed fingerprints, the patterns did not change during the 9 weeks of culture of strain H37Rv or strain IMCJ 541 (data not shown), indicating that IS*6110* inserts are also stable over a few months.

Comparison of fingerprints between *M. tuberculosis* strains H37Rv and H37Ra. The virulent *M. tuberculosis* strain H37Rv and its avirulent derivative strain H37Ra were originally derived from the same strain, H37 (22, 23). It was reported that there are distinct differences between these strains with respect to IS*6110*-probed fingerprint patterns (3, 11). We investigated whether differences exist between these strains with respect to $(CGG)_5$-probed fingerprint patterns. DNA derived from the H37Rv and H37Ra strains were digested with 16 restriction enzymes as described in Materials and Methods. Unexpectedly, the patterns of $(CGG)_5$-based hybridization showed no differences between the H37Rv and H37Ra strains (Fig. 2A). For example, the $(CGG)_5$-based RFLP patterns of PvuII-digested fragments of H37Rv were identical to those of H37Ra (Fig. 2A, PvuII). However, the IS*6110*-based RFLP patterns of H37Rv were markedly different from those of H37Ra, which were analyzed with the use of the same blot of PvuII-digested fragments used in the $(CGG)_5$-based RFLP analysis (Fig. 2B). In the IS*6110*-based RFLP patterns, H37Rv showed 9 bands, and H37Ra showed 11 bands. Strain H37Rv but not H37Ra showed one band of 5.1 kb. Strain H37Ra but not H37Rv showed three bands of 1.1, 2.3, and 3.0 kb.

IS*6110*- and $(CGG)_5$-probed DNA fingerprinting of *M. tuberculosis* clinical isolates. To assess the potential usefulness of $(CGG)_5$ as an epidemiologic marker for *M. tuberculosis*, 109 clinical isolates obtained from Tokyo (76 isolates) and Warsaw (33 isolates) and the H37Rv and H37Ra strains were analyzed by the IS*6110*- and $(CGG)_5$-probed fingerprint methods. For IS*6110*-probed hybridization, DNA of these isolates was digested with PvuII according to a standardized protocol (26). For $(CGG)_5$-probed hybridization, DNA of the isolates was digested with AluI. When DNA of the H37Rv and H37Ra strains was digested with AatII, EcoRI, MluI, NruI, NsbI, PstI, PvuII, SacI, SalI, or XhoI, relatively higher-molecular-weight DNA fragments were visualized by the probe with a minimum size of 1 to 3.5 kb and a maximum size of more than 10 kb (Fig. 2A). When digested with AfaI, AluI, HinfI, Sau3AI, SmaI, or XspI, DNA fragments of sizes of 0.5 to 8 kb were visualized. When DNA of five clinical isolates selected at random were digested with AluI, clear $(CGG)_5$ fingerprint patterns with 10 to 14 copies of DNA fragments of 0.75 to 8 kb were detected (data not shown). Although we used AluI for this fingerprinting method, other enzymes may also be used.

IS*6110* fingerprint patterns obtained from clinical isolates and the corresponding dendrogram are shown in Fig. 3A. IS*6110* copies were detected in 110 of 111 isolates. One isolate from Japan had no copy. As indicated in Fig. 3A, 10 of 111 isolates (9.0% of tested isolates), including 8 isolates from Japan and 2 from Poland, possessed fewer than 6 copies of IS*6110*, which was insufficient to distinguish polymorphisms. Except for these 10 isolates with fewer than 6 copies of IS*6110*, the IS*6110* fingerprint patterns of 101 isolates showed ≥28% similarity; 98 patterns were found (Fig. 3A). Five clusters with

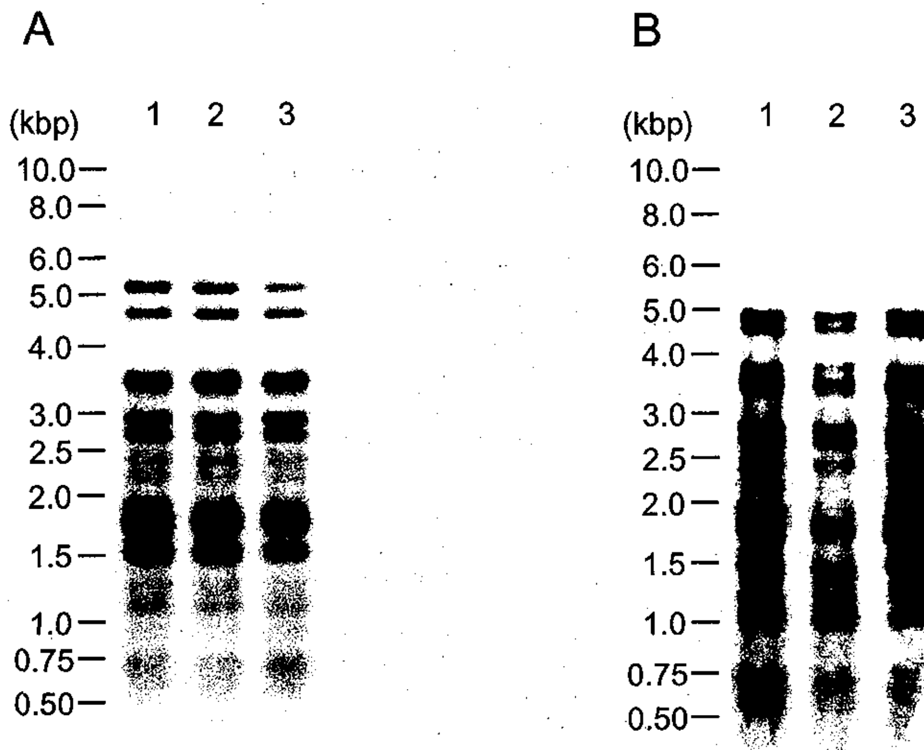

FIG. 1. (CGG)$_5$ fingerprinting of *M. tuberculosis* H37Rv (A) and clinical isolate IMCJ 541 (B), which were cultured and serially passaged weekly. The bacteria were harvested at 0 (lane 1), 3 (lane 2), and 9 (lane 3) weeks after culture.

≥44% similarity, including clusters Ia, IIa, IIIa, IVa, and Va, were detected (Fig. 3A). Cluster Ia was composed of seven Poland-derived isolates. Cluster IIa was composed of two H37 variants and 11 Japan- and 6 Poland-derived isolates. Cluster IIIa was composed of three Japan- and seven Poland-derived isolates. Cluster IVa was composed of four Japan- and five Poland-derived isolates. Cluster Va was composed predominantly of Japan-derived isolates (46 isolates from Japan and 2 from Poland). The majority of Japan-derived isolates (61%) and Poland-derived isolates (76%) belonged to cluster Va and to clusters Ia to IVa, respectively.

(CGG)$_5$ fingerprint patterns and the corresponding dendrogram are shown in Fig. 3B. (CGG)$_5$ copies were detected in all clinical isolates tested. The copy number ranged from 8 to 16, with a mean of 13.0 ± 1.5 per isolate. The number of (CGG)$_5$ copies of Japan- and Poland-derived isolates ranged from 8 to 16, with a mean of 12.9 ± 1.5 per isolate and from 11 to 15, with a mean of 13.2 ± 1.3 per isolate, respectively. A total of

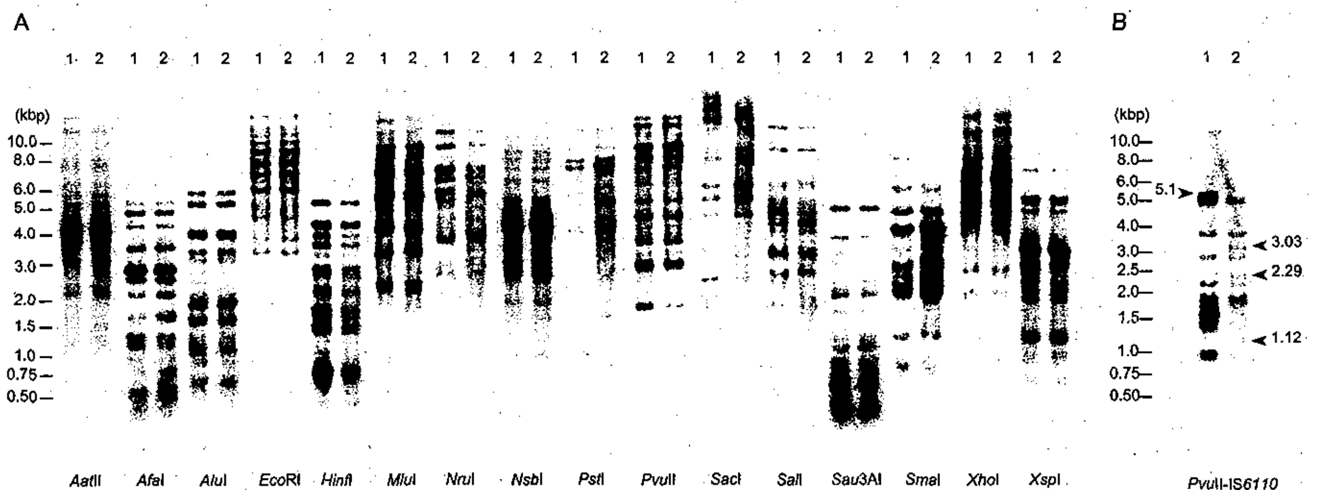

FIG. 2. (CGG)$_5$ (A) and IS*6110* (B) fingerprinting of *M. tuberculosis* strains H37Rv (lane 1) and H37Ra (lane 2). Genomic DNA was digested with 16 restriction enzymes. The digested fragments were separated by electrophoresis.

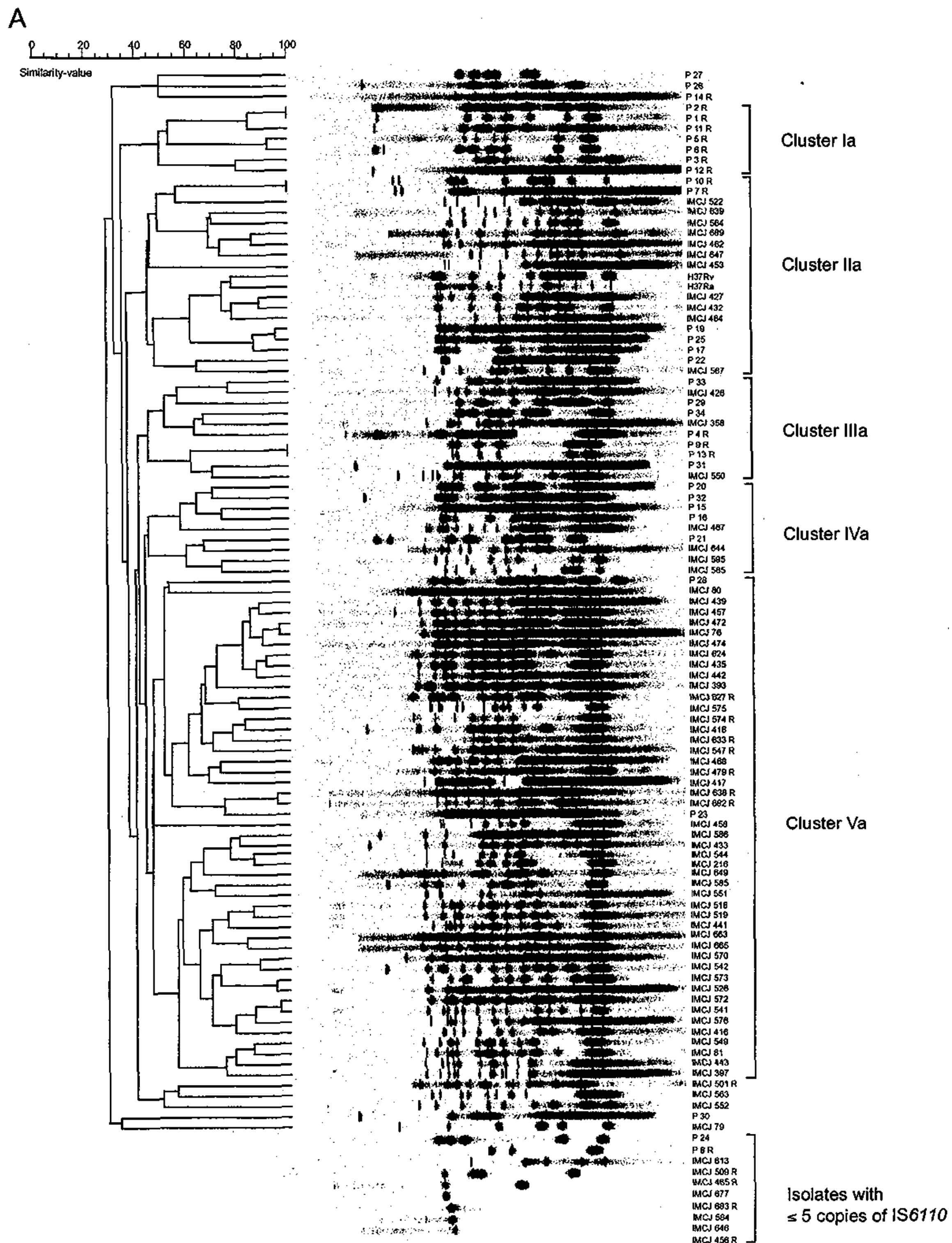

FIG. 3. IS*6110*- and $(CGG)_5$-probed DNA fingerprinting patterns of *M. tuberculosis* clinical isolates from Japan and Poland and the respective corresponding dendrograms. The fingerprint patterns are ordered by similarity. The corresponding dendrograms are to the left of the patterns. The position of each IS*6110* (A) or $(CGG)_5$ (B) band is normalized so that the patterns for all strains are comparable. The scale depicts the similarity of patterns calculated as described in Materials and Methods. In IS*6110*-probed DNA fingerprint patterns, five clusters showing a similarity of more than 44% were designated clusters Ia, IIa, IIIa, IVa, and Va. Isolates with five or less than five copies are indicated in panel A. In $(CGG)_5$-probed DNA fingerprint patterns, four clusters showing a similarity of more than 70% were designated clusters Ib, IIb, IIIb, and IVb. The isolates are named according to their origin as IMCJ (Japan) or P (Poland); the suffix R indicates drug resistance For example, IMCJ 627 R is a Japan-derived drug-resistant isolate.

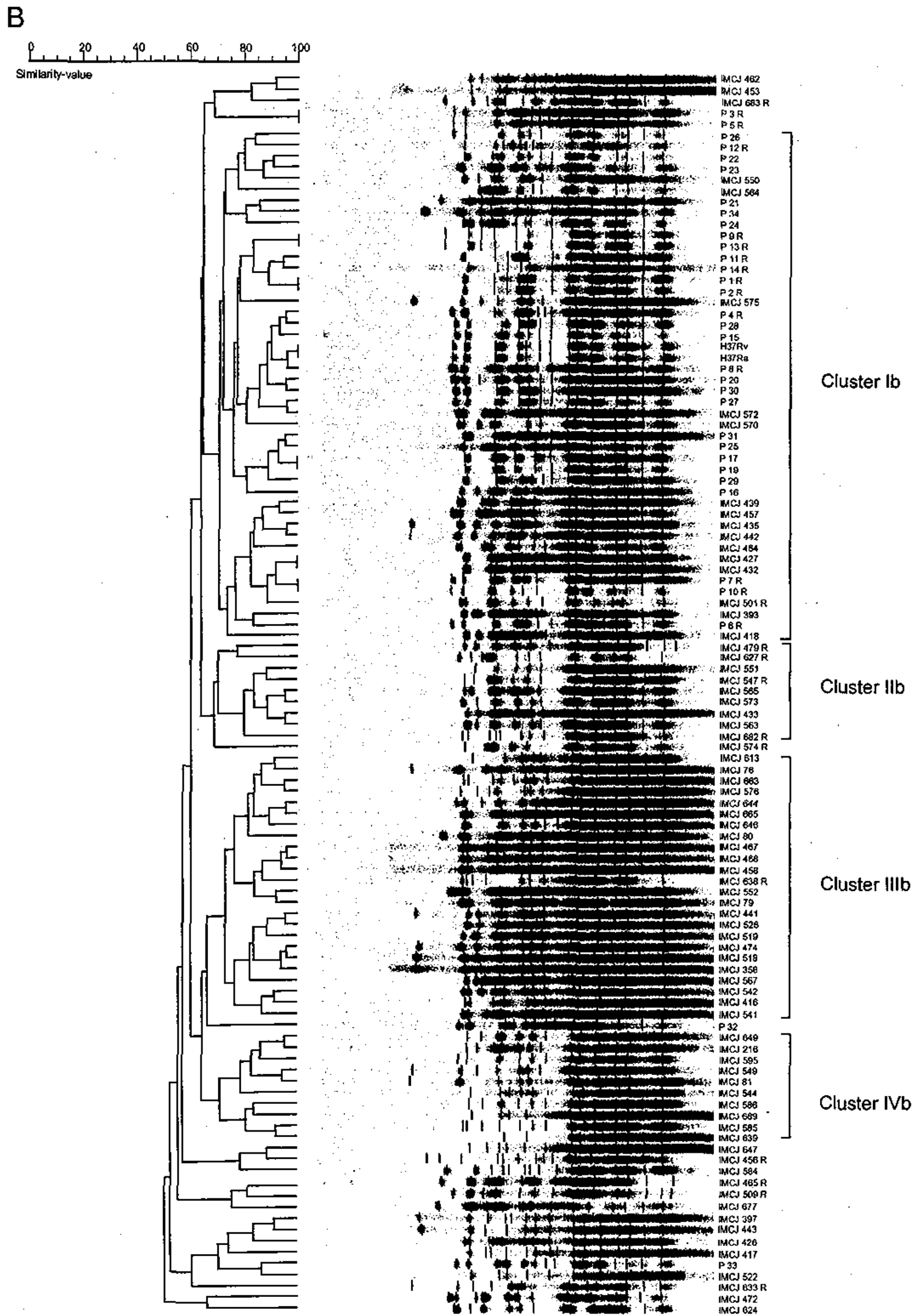

FIG. 3—*Continued.*

104 (CGG)$_5$ fingerprint patterns were found with ≥50% similarity (Fig. 3B). Four clusters with ≥70% similarity, including clusters Ib to IVb, were detected (Fig. 3B). Cluster Ib was composed of two H37 variants and 15 Japan- and 29 Poland-derived isolates. Clusters IIb, IIIb, and IVb were composed of 9, 24, and 10 Japan-derived isolates, respectively. Over half of the Japan-derived isolates (57%) and the majority of the Poland-derived isolates (88%) belonged to clusters IIb to IVb and to cluster Ib, respectively (Fig. 3B).

Both the IS*6110* and (CGG)$_5$ fingerprint analyses showed an association between fingerprint pattern and geographic origin, indicating a correlation between them. Ten isolates that were

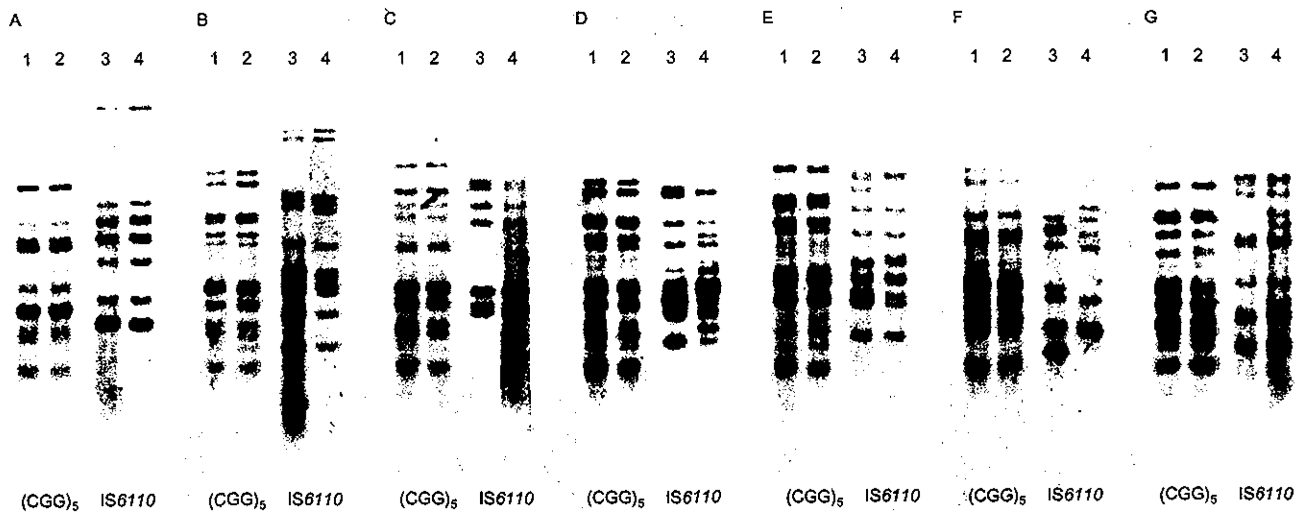

FIG. 4. $(CGG)_5$- and IS*6110*-probed DNA fingerprinting patterns of *M. tuberculosis* isolates that shared identical $(CGG)_5$ fingerprinting. (A) Lanes 1 and 3, P1; lanes 2 and 4, P2. (B) Lanes 1 and 3, P 7; lanes 2 and 4, P 10. (C) Lanes 1 and 3, P 9; lanes 2 and 4, P 13. (D) Lanes 1 and 3, H37Rv; lanes 2 and 4, H37Ra. (E) Lanes 1 and 3, IMCJ 427; lanes 2 and 4, IMCJ 432. (F) Lanes 1 and 3, P 3; lanes 2 and 4, P 5. (G) Lanes 1 and 3, P 17; lanes 2 and 4, P 19.

indistinguishable by IS*6110* RFLP because of the presence of few copies of the marker could be analyzed by $(CGG)_5$ marker. Three and seven pairs of isolates were identical to each other in the IS*6110* and $(CGG)_5$ fingerprint patterns, respectively (Fig. 4). The three pairs P 1 and P 2, P 7 and P 10, and P 9 and P 13 were identical to each other in the IS*6110* and $(CGG)_5$ fingerprint patterns (Fig. 4A to C, respectively). The four pairs H37Rv and H37Ra, IMCJ 427 and IMCJ 432, P 3 and P 5, and P 17 and P 19 were identical to each other in the $(CGG)_5$ fingerprint pattern but different in the IS*6110* fingerprint pattern (Fig. 4D, E, F, and G, respectively). The data suggest that the $(CGG)_5$ fingerprint patterns are more stable than the IS*6110* patterns.

Occurrence of $(CGG)_5$ among various mycobacterial strains. We investigated the presence of $(CGG)_5$ repeat sequences in mycobacterial species. $(CGG)_5$ hybridization patterns from various mycobacterial species are shown in Fig. 5. Bands ranging from 0 to 20 in number were seen. *Mycobacterium szulgai* possessed 20 bands. *M. bovis* BCG, *Mycobacterium marinum*, and *Mycobacterium kansasii* possessed 16 bands. *Mycobacterium nonchromogenicum*, *Mycobacterium terrae*, *Mycobacterium gastri*, *Mycobacterium simiae*, *Mycobacterium smegmatis*, and *Mycobacterium intracellulare* possessed 14, 12, 8, 5, 5, and 3 bands, respectively. *Mycobacterium peregrinum* possessed two bands. *Mycobacterium fortuitum* and *Mycobacterium chelonae* possessed one band. *Mycobacterium scrofulaceum*, *Mycobacterium avium*, *Mycobacterium xenopi*, and *Mycobacterium abscessus* showed no bands.

DISCUSSION

In this study, we found that various bacterial strains contain TRS in their genomes. In humans, TRS are associated with hereditary neurologic and neuromuscular disorders, including myotonic dystrophy, Huntington's disease, Fragile X syndrome, and Friedreich's ataxia (27). These diseases result from TRS expansion such as $(CTG)_n$, $(CGG)_n$, and $(GAA)_n$ (27). The TRS sizes associated with these diseases are usually quite large. For example, 80 to 3,000 repeats of CTG have been found in myotonic dystrophy, 230 to 2,000 repeats of CGG have been found in Fragile X syndrome, and 200 to 900 repeats of GAA have been found in Friedreich's ataxia (21). These expanded TRS can form hairpin structures or intramolecular triplex structures that result in genetic instability (21). The TRS sizes found in bacteria were relatively small. The largest size TRS identified was 21 repeats of GAA in *M. leprae*. The most frequently identified TRS was five repeats of CGG in *M. tuberculosis* and *M. bovis*. TRS found in bacteria are not likely to be linked to genetic instability because of the lower repeat number.

The $(CGG)_5$ TRS found in two strains of *M. tuberculosis* (H37Rv and CDC1551) and in one strain of *M. bovis* existed in genes encoding PE protein families, including a PE_PGRS subfamily and PPE protein families comprising 88 to 101 and 61 to 69 kinds of proteins, respectively, which occupy approximately 8% of the genome (4, 7, 8). The functional properties of $(CGG)_5$ in these genes are unknown, but $(CGG)_5$ should not play an important role in the development of the variations among different strains. $(CGG)_5$ in the PPE genes was located in the conserved N-terminal domain PPE but not in the C-terminal variable domain containing the major polymorphic tandem repeats with the consensus sequence of GCCGGT GTTG (10, 18). $(CGG)_5$ in the PE_PGRS genes was within the C-terminal variable domain containing the PGRS with the consensus sequence of CGGCGGCAA (18, 19). $(CGG)_5$ in the PE_PGRS genes did not comprise part of the consensus sequence of PGRS. $(CGG)_5$ was contained in 13 and 12 PE_PGRS genes in H37Rv and CDC1551, respectively. Among these genes, deletion or insertion was detected at one site of Rv1068c, two sites of Rv1087, and two sites of Rv1450c compared with their orthologs, MT1097, MT1118.1, and MT1497.1, respectively (data not shown). However, $(CGG)_5$ was not near these sites, indicating that it did not directly affect the deletion and insertion of PE_PGRS genes. $(CGG)_5$ in