

UNIVERSIDAD NACIONAL AUTONOMA
DE MÉXICO

INSTITUTO DE INVESTIGACIONES BIOMÉDICAS

DEPARTAMENTO DE INMUNOLOGIA

EXPRESIÓN DIFERENCIAL DE LOS GENES DE LA FAMILIA
PE EN AISLADOS CLÍNICOS DE *Mycobacterium tuberculosis*

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE
DOCTOR EN CIENCIAS

P R E S E N T A :

M. En C. JOSE JAVIER FLORES ESTRADA

DIRECTOR DE TESIS: DRA. CLARA INÉS ESPITIA PINZÓN

COMITÉ TUTORAL: DR. RAUL MANCILLA JIMENEZ
DR. ROBERTO CORIA
DRA. CLARA INÉS ESPITIA PINZÓN

MÉXICO, D.F.

2007



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Este trabajo fue realizado en el Departamento de Inmunología del Instituto de Investigaciones Biomédicas de la Universidad Nacional Autónoma de México

El presente trabajo se realizó bajo la dirección de la DRA. CLARA INÉS ESPITIA PINZÓN del departamento de Inmunología del Instituto de Investigaciones Biomédicas de la Universidad Nacional Autónoma de México.

Este trabajo fue financiado por los donativos: CONACYT (33580-M) y DGPA (IN221599) Universidad Nacional Autónoma de México.

Durante el desarrollo del trabajo el autor fue becario del Consejo Nacional de Ciencias y Tecnología (CONACyT).

AGRADECIMIENTOS

A la que además de ser una gran asesora, me brindo su confianza, su apoyo y amistad en las condiciones más difíciles de mi vida.

A mi comité tutorial por su objetividad y sabiduría.

A los doctores que participaron como jurado: Laura Camarena, Luz María López, Yolanda López Vidal, Clara Espitia, Gonzalo Castillo, Rogelio Hernández y Jorge Gonzales por sus críticas y tiempo dedicado en revisar este trabajo.

A los técnicos académicos: Rafael Sánchez Roldan y a la Dra. Martha Lara por su asesoría técnica.

A amigos del laboratorio: Rafael, Angélica, Jaime, Mariano, Wendy, Marcos, Martha, Pedro, Cristina, Ingrid y Toño, a todos ellos que por sus acertadas críticas y sugerencias que permitieron llevar con éxito mi trabajo, sin menospreciar su amistad entrañable.

Al siempre apreciable Dr. Jorge Gonzales y Merchand que por su invaluable ayuda incondicional formó una parte muy importante para finalizar mis ensayos.

DEDICATORIA

A mí siempre amada Sonia

Y a mis chaparritos Alan y Yael,

Que son y serán el pilar de mi vida.

INDICE

Summary	1
Resumen	2
1. Introducción	4
1.1 Historia de la tuberculosis	4
1.2 Epidemiología	5
1.2.1 La tuberculosis en el mundo	5
1.2.3 La tuberculosis en México	6
1.3 Tuberculosis: agente causal	7
1.4 Tuberculosis: la enfermedad	8
1.5 El genoma de <i>Mycobacterium tuberculosis</i>	8
2. Justificación	14
3. Hipótesis	15
4. Objetivo general	15
5. Objetivos específicos	15
6. Materiales y métodos	16
6.1 Bacterias y condiciones de cultivo	16
6.2 Diseño de oligonucleótidos	16
6.3 Southern blot	17
6.4 Extracción del DNA genómico	19
6.5 Extracción del RNA	20
6.6 RT-PCR	20
6.6.1 Polimorfismo conformacional de cadena sencilla (SSCP)	21
6.6.2 Extracción, purificación y secuenciación de los amplificados resueltos por RT-PCR_SSCP.	22
6.7 RT-PCR y RFLP-PCR	22
7 Resultados	24
7.1 Diseño de los oligonucleótidos y condiciones de amplificación	24

7.2 Identificación de los genes de la familia PE	28
7.3 Identificación de los genes de la familia PE expresados diferencialmente	42
7.4 Identificación de los genes de la familia PE no expresados	46
8. Discusión	52
9. Conclusiones	58
10. Bibliografía	59
Anexos	67

INDICE DE TABLAS

Tabla 1. Patrones de restricción generados por RFLP-PCR del fragmento amplificado de 235 pb de las cepas estudiadas.	35
Tabla 2. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 046.	42
Tabla 3. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 1000.	43
Tabla 4. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 1463.	43
Tabla 5. Análisis de restricción teórico de los genes de la familia PE con <i>Pvu I</i> .	44
Tabla 6. Análisis de restricción teórico de los genes de la familia PE con <i>Sal I</i> .	48
Tabla 7. Análisis de restricción teórico de los genes de la familia PE con <i>Stu I</i> .	50

INDICE DE FIGURAS

Figura 1. Polimorfismo de Longitud de los Fragmentos de Restricción (RFLP).	17
Figura 2. Alineamiento de las secuencias de la familia PE.	24
Figura 3. PCR.	25
Figura 4. Estandarización de la PCR.	26
Figura 5. RT-PCR.	27
Figura 6. PCR-SSCP.	28
Figura 7. Southern-blot.	29
Figura 8. PCR-SSCP.	30
Figura 9. Alineamiento de secuencias.	31
Figura 11. SSCP-PCR.	33
Figura 12. RFLP-PCR de <i>M. tuberculosis</i> H37Rv.	34
Figura 13. RFLP-PCR de <i>M. tuberculosis</i> H37Rv.	36
Figura 14. RFLP-PCR de <i>M. canettii</i> 046.	37
Figura 15. RFLP-PCR de <i>M. tuberculosis</i> 1000.	38
Figura 16. RFLP-PCR de <i>M. tuberculosis</i> 1463.	39
Figura 17. PCR-RFLP.	40
Figura 18. PCR-RFLP.	41
Figura 19. PCR-RFLP.	45
Figura 20. PCR.	46
Figura 21. RFLP-PCR.	47
Figura 22. PCR-RFLP.	49
Figura 23. PCR-RFLP.	51

SUMMARY

The sequencing of the genome of *M. tuberculosis* complex and other mycobacteria is a very important tool to understand the Biology of these microorganisms as well as the mechanisms that allow the bacilli to invade, multiply and persist in their host. A gene family restricted to mycobacterial species called PE has been identified in the genomes of slow growing pathogenic mycobacteria. This family contains the PE and PE_PGRS subfamilies and in *M. tuberculosis* encompasses 99 members widely polymorphic. Recent experimental evidence shows that PE/PE_PGRS proteins participate in host-bacteria recognition and are involved in pathogenesis. It has been also observed that expression of PE genes occur under different stimulus, indicating that the sequences could be expressed under the stress conditions generated during the bacteria infection.

The identification and characterization of PE/PE_PGRS sequences which are expressing in mycobacteria that infect the human as well the study of the potential role these sequences are playing in the immune response could be very important for the knowledge of mycobacteria Biology and could have a strong impact in the design of new tools in the fight against tuberculosis.

In this study, the expression of the *M. tuberculosis* PE/PE_PGRS gene family was assayed by PCR-SSCP y RT-PCR-RFLP in *M. tuberculosis* H37Rv, *M. canetti*, and two clinical isolates of *M. tuberculosis*. Restriction analyses and sequencing of the RT-PCR products showed that all the strains expressed the PE Rv1172c gene while the PE_PGRS Rv3652 gene was only expressed by one of the *M. tuberculosis* clinical isolates and the PE_PGRS Rv0578c was not expressed by *M. canetti*. It was also determined that the PE_PGRS Rv0278c and Rv0279c were not expressed by any of the studied strains. The genes could be amplified from DNA but not from cDNA of all the strains. These findings suggest that PE/PE_PGRS may play a role in protein variation between *M. tuberculosis* strains.

RESUMEN

La secuencia de los genomas de cepas de *M. tuberculosis* H37Rv así como de otras micobacterias, ha constituido una herramienta muy valiosa para el estudio de la biología de estos microorganismos y para el entendimiento de los mecanismos que permiten a las micobacterias patógenas invadir, multiplicarse y persistir en su hospedero. En el genoma de *M. tuberculosis* se ha identificado una familia de 99 genes conocidas como PE que se distribuyen en dos subfamilias. Una de ellas corresponde a la subfamilia PE y la otra subfamilia se le conoce como, PE_PGRS. Las proteínas codificadas por esta familia se han restringido principalmente a las micobacterias patógenas de crecimiento lento. Su gran abundancia, amplio polimorfismo y algunas evidencias experimentales hacen pensar que estos genes podrían estar involucrados en la patogénesis de la bacteria, así como ser como moléculas blanco de reconocimiento en la interacción patógeno-hospedero. Se ha observado que la expresión de los genes PE/PE_PGRS ocurre ante diferentes y variados estímulos, lo cual sugiere que estos genes podrían expresarse preferencialmente bajo condiciones de estrés a las cuales se ve sometida la bacteria durante la infección. La identificación y caracterización de las secuencias PE/PE_PGRS que se están expresando en las micobacterias que infectan al humano así como el papel que ellas juegan en la respuesta del huésped es de gran importancia al conocimiento biológico así como al diseño de nuevos métodos inmunoproliféricos en tuberculosis.

Por lo tanto en este estudio determinamos la expresión de los genes de la familia PE/PE_PGRS entre los aislados clínicos del complejo *M. tuberculosis* 1000, 1463, H37Rv y *M. canetti*. La expresión de los genes se estudio por técnicas como el PCR-SSCP y PCR-RFLP. El análisis comparativo de ambas técnicas nos permitió identificar que el gen Rv1172c fue expresado por todas las cepas estudiadas. Mientras que el gen PE_PGRS Rv0578c fue expresado por las cepas H37Rv, 1000 y 1463 pero no por la cepa *M. canetti* 046 y el gen Rv3652 fue expresado únicamente por la cepa 1463. Se demostró que el gen PE_PGRS Rv0279c y el PE_PGRS Rv0278c estaban

presentes en el DNA de los fragmentos amplificados, pero ausentes en el cDNA de las cepas estudiadas. Estos hallazgos sugieren que las PE/PE_PGRS pudieran jugar un papel en la variación de proteínas entre las cepas de *M. tuberculosis*

1. INTRODUCCIÓN

1.1 HISTORIA DE LA TUBERCULOSIS

La tuberculosis es una enfermedad que ha afectado al ser humano desde hace miles de años. Las primeras referencias sobre ésta aparecen en escritos del Antiguo Egipto, Babilonia y China. El DNA de *Mycobacterium tuberculosis*, su agente causal ha sido detectado en fragmentos de columnas vertebrales de momias egipcias con más de 5000 años de antigüedad. Por otra parte, las descripciones patológicas y anatómicas de la tuberculosis comenzaron a aparecer en el siglo XVII y es así como en 1679, Franciscus de la Boe describió por primera vez los cambios en los pulmones que caracterizan la patología de la enfermedad, como son los tubérculos y el progreso de las lesiones hacia abscesos y cavernas (Bloom, 1994).

En 1720, Marten sugirió que la tuberculosis podría ser causada por criaturas diminutas responsables de las lesiones y síntomas de la enfermedad. El término tuberculosis, derivado del latín *tubercula* (protuberancia o nódulo pequeño) se empezó a utilizar alrededor del año 1839 (Schlossberg, 2000), pero no fue sino hasta el año de 1882 que Robert Koch aisló por primera vez al bacilo de la tuberculosis demostrando su infectividad y transmisibilidad. Es por eso, que a *M. tuberculosis* se le conoce como "bacilo de Koch". Años después, Koch desarrolló la prueba de la tuberculina que actualmente se usa como parte del diagnóstico de la enfermedad y está basada en la reacción de hipersensibilidad tardía que se presenta a las proteínas de la bacteria en los individuos infectados (Bloom, 1994).

Las primeras epidemias causadas por esta bacteria, aparecen en Europa y en los Estados Unidos en el siglo XIX. El mayor número de muertes causadas por la tuberculosis se registraron durante la Revolución Industrial; en esta época, Bayle y Laëneg establecieron las formas y estadios de la enfermedad (Schlossberg, 2000). En 1924, Albert Calmette y Alphonse Guérin desarrollaron una vacuna denominada Bacilo de Calmette-Guérin (BCG) a partir de la atenuación por los pases en medio de cultivo de la cepa de *Mycobacterium bovis* que produce la tuberculosis en el ganado. A

mediados de la Segunda Guerra Mundial, se empezó a utilizar la quimioterapia para eliminar a la bacteria que había amenazado a la humanidad por tantos años. El primer agente quimioterapéutico específico para la tuberculosis fue la estreptomina, descubierta por Selman Waksman en 1944. A partir de este avance, los índices de mortalidad disminuyeron significativamente. En 1948, se empezó a utilizar el ácido para-amino salicílico (PAS) y más tarde la isoniacida. Posteriormente, en los años sesenta se descubrió el etambutol y en los setenta, la rifampicina (Sudre *et al.*, 1992).

1.2 EPIDEMIOLOGÍA

1.2.1 La tuberculosis en el Mundo

La tuberculosis continúa siendo un problema de salud pública en el mundo. La Organización Mundial de la Salud (OMS) estima que durante el periodo 2000 a 2020 podría haber 1000 millones de personas infectadas, 200 millones de los cuales enfermarán y 35 millones morirán si la enfermedad no es controlada. Las incidencias más altas se encuentran en los países subdesarrollados de África, Asia y América Latina (Cauten *et al.*, 2002). Así mismo, la OMS estimó que un tercio de la población mundial está infectada por *M. tuberculosis*, siendo la incidencia anual de 8 millones de casos nuevos, con una mortalidad global de aproximadamente de 2-3 millones de personas por año (Frieden., 2002). Los 22 países más afectados son de los más pobres del planeta, donde la prevalencia de la enfermedad se encuentra condicionada por la pobreza, la coinfección con el Virus de la Inmunodeficiencia Humana (VIH), la falta de control y de un diagnóstico certero (Gazzard., 2001). Otra diferencia existente entre países es la distribución de los casos con respecto a la edad. En países desarrollados el 80% de los casos se presenta en individuos mayores de 50 años. En los países en vías de desarrollo, 75% de los infectados corresponde a individuos menores de 50 años (Rajagopalan, 2001).

La incidencia mundial de tuberculosis aumenta aproximadamente un 0.4% al año. La mayoría de los individuos infectados controlan la infección inicial debido a que

montan una respuesta inmune celular que previene el desarrollo de la enfermedad pero permite la persistencia de micobacterias viables, según WHO REPORT 2003 Global Tuberculosis Control: Entre el 5-10% de los individuos infectados podrían desarrollar la enfermedad (Davey, 2001).

1.2.3 La tuberculosis en México

La incidencia de tuberculosis en México es de 18 por cada 100,000 habitantes siendo inferior al promedio latinoamericano, que a mediados de los años noventa se ubicaba en 32 casos nuevos por cada 100,000 habitantes (Pérez-Padilla, 2001). Valdespino y García analizaron las tendencias de la morbilidad por tuberculosis las cuales mostraron un descenso desde 1941 hasta 1976, año en que se estabilizan (García y Valdespino, 1994). Sin embargo, para 1993 la tasa de morbilidad fue de 17.1 casos por cada 100,000 habitantes y de haberse conservado la tendencia descendente, la tasa para 1993 debería haber sido de 12.5 casos por 100,000 habitantes (Valdespino *et al.*, 1989). Se calcula que en los últimos 10 años han ocurrido 27,000 casos en exceso. Al aplicar el modelo de historia natural de la enfermedad desarrollado por Miller y publicado por la OMS en 1982, se puede estimar la tasa real de casos de tuberculosis que ocurren en México. Este modelo calcula que existen de 39 a 59 casos por cada 100,000 habitantes, y sugiriendo estas observaciones que la tasa de tuberculosis en México pudiera estar subestimada.

En México la tuberculosis constituye un problema que va en ascenso y que presenta nuevos retos para su control, entre ellos su asociación con la infección por VIH/SIDA (Sepúlveda and Bronfman, 1988). Esta asociación es la causa de 2.4 a 2.9 millones de muertes anuales entre personas de 15 a 49 años de edad según varias estimaciones (Soberón-Acevedo *et al.*, 1996).

Otro problema asociado a este padecimiento es el surgimiento de cepas resistentes a los fármacos utilizados para el tratamiento de la enfermedad. Un estudio realizado en tres estados mexicanos mostró que mientras la fármaco-

resistencia adquirida a isoniazida y rifampina estaba dentro de los límites esperados (22.4%), la resistencia primaria era muy alta (2.4%). Esta información ha motivado el uso de un tratamiento corto y la administración de cuatro medicamentos, el programa se conoce como TAES (Tratamiento acortado estrictamente supervisado) y se está aplicando en diferentes países (Flisser *et al.*, 2002).

1.3 TUBERCULOSIS: AGENTE CAUSAL

La tuberculosis es una enfermedad infecciosa crónica causada por bacterias del género *Mycobacterium*. El ser humano puede infectarse por tres especies distintas de micobacterias: *M. tuberculosis*, *M. bovis* y *Mycobacterium africanum*. *M. tuberculosis* es el patógeno que mas comúnmente se asocia al desarrollo de la enfermedad en el hombre (Bloom, 1994).

M. tuberculosis es un bacilo aerobio estricto, inmóvil, no esporulado, no encapsulado, que presenta un tiempo de generación muy prolongado (15-20 horas) y que requiere de 2 a 6 semanas para crecer en medios sólidos. La temperatura corporal (35-37°) y una atmósfera enriquecida con CO₂ estimula su desarrollo. Microscópicamente es un bacilo alargado de aproximadamente 0.4 x 3 µm y de forma arrozada, que pertenece a la clase Actinobacteridae; Orden Actinomycetales; Familia Mycobacteriaceae; Género *Mycobacterium* y al complejo *Mycobacterium tuberculosis* (Bloom, 1994).

Como todas las micobacterias, *M. tuberculosis*, se caracterizan por poseer una pared celular resistente constituida por ácidos micólicos, ocasionando que ciertos colorantes derivados de las anilinas (fucsina fenicada) sean retenidos a pesar de ser tratadas con un ácido y un alcohol, por esto se les ha denominado bacilos ácido-alcohol-resistentes. Estos organismos sobreviven en un clima seco durante largos periodos aunque pueden ser eliminados, por ejemplo, mediante rayos de sol directos o por la radiación ultravioleta. Las micobacterias poseen una pared celular con un contenido lipídico extraordinariamente alto, como consecuencia, la pared constituye

una verdadera barrera de defensa contra el ataque del hospedero (Schlossberg, Schlossberg, 2000).

1.4 TUBERCULOSIS: LA ENFERMEDAD

La enfermedad causada por *M. tuberculosis*, se caracteriza por la formación de estructuras celulares denominadas tuberculomas donde los bacilos quedan aislados, afectando principalmente los pulmones, pero la bacteria puede diseminarse a otros órganos, enfermedad que se denomina tuberculosis extrapulmonar o sistémica. Los bacilos pueden además permanecer latentes en el organismo durante un largo periodo, hasta que una disminución en las defensas del hospedero le confiere la oportunidad de multiplicarse y producir los síntomas de la enfermedad (tuberculosis de reactivación) (Schlossberg, 2000).

Existen otras especies de micobacterias que pueden causar infecciones en el hombre, patológica y clínicamente similares a la tuberculosis. Por ejemplo, el complejo *Mycobacterium avium* (MAC) y *Mycobacterium intracellulare* produce una infección oportunista. En las personas con sistemas inmunológicos competentes, estos organismos se mantienen bajo control y muy pocas veces producen enfermedad. Sin embargo, en sujetos con infección por el VIH pueden presentarse como infecciones oportunistas (Bloom, 1994).

1.5 EL GENOMA DE *Mycobacterium tuberculosis*

A pesar que en los últimos años se ha incrementado considerablemente la investigación en la tuberculosis y su agente causal, aun no se conocen con certeza los factores de virulencia y patogenicidad que caracterizan este microorganismo. La secuencia de los genomas de *M. tuberculosis* H37Rv así como el de algunos aislados clínicos, y otras cepas del complejo tuberculosis (Cole *et al.*, 1998; Cole *et al.*, 2001; Fleischmann *et al.*, 2002; Li *et al.*, 2005), ha constituido una herramienta muy valiosa para el estudio de la biología de estos microorganismos y por ende del entendimiento

de los mecanismos que permiten a las micobacterias patógenas invadir, multiplicarse y persistir en su hospedero.

El genoma de *M. tuberculosis* contiene 4 411,529 pb que codifican alrededor de 3,924 proteínas, 40% de las cuales tienen una función predicha, 44% presentan similitud con otras proteínas y un 16% son proteínas hipotéticas. Interesantemente, dentro de este genoma se han identificado dos familias de genes conocidas como PE y PPE que ocupan aproximadamente el 5% de la capacidad codificadora (ver anexo 1). Las proteínas codificadas por estas dos familias presentan una región amino-terminal conservada seguida por una región carboxilo-terminal que varía en longitud y secuencia. La presencia del motivo Pro-Glu o Pro-Pro-Glu en la posición 8 y 9 o la posición 8, 9 y 10 del amino-terminal determinó los nombres PE y PPE, respectivamente (Cole *et al.*, 1998) (anexo 2).

La familia PE contiene 100 miembros que se distribuyen principalmente en dos subfamilias. Una de ellas corresponde a la subfamilia PE caracterizada por contener solo la región amino-terminal conservada de 110 aminoácidos. Alrededor de 37 de estos genes poseen un marco de lectura abierto con un promedio de 110 aminoácidos, pero puede variar desde 28 (Rv3018A) hasta 588 aminoácidos (Rv0151c). La otra subfamilia se le conoce como PE_PGRS (PGRS del inglés Polymorphic GC Rich Sequences) y comprende 67 miembros, caracterizados por presentar la región amino terminal conservada PE, seguida por una serie de repeticiones en tándem con motivo Gly-Gly-A-Gly-Gly-X (Cole *et al.*, 1998). El gen más grande, el Rv3508, codifica para 1901 aminoácidos conteniendo 960 residuos de glicina y 88 repeticiones - Gly-Gly-A-Gly-Gly-X (anexo 3).

El polimorfismo genético característico de las secuencias PE_PGRS podría ser consecuencia de la naturaleza altamente repetitiva de estos genes, lo cual facilitaría eventos de inserciones o pérdidas, ocasionados por errores en la replicación. Esta variabilidad genética a su vez podría tener efectos sobre la función y la antigenicidad de las proteínas, ya que las inserciones, deleciones y/o mutaciones puntuales podrían generar variaciones en las proteínas, lo cual podría reflejarse en el reconocimiento

inmunológico en el caso de que estas funcionasen como antígenos (Mizrahi *et al.*, 1998).

Hasta la fecha, las proteínas PE_PGRS han sido encontradas únicamente en las micobacterias y dentro de estas se han restringido principalmente en los miembros del complejo de *M. tuberculosis* y en algunas otras especies cercanas. Su gran abundancia, el amplio polimorfismo y algunas evidencias experimentales hacen pensar que estos genes podrían estar involucrados en la patogénesis de la bacteria (Ramakrishnan *et al.*, 2000).

Estudios de hibridación molecular usando como sondas la región PGRS de algunas PE_PGRS han indicado que los genes PE_PGRS existen en *M. marinum*, *M. goodnae*, *M. kansasii*, *M. microti*, *M. bovis* (Poulet and Cole *et al.*, 1995; Ross *et al.*, 1992). Por otra parte, los estudios de genómica comparativa revelan que en el genoma de *Mycobacterium leprae*, las secuencias PE_PGRS son pseudogenes. Además, es de resaltar que la falta de genes intactos PE /PE_PGRS así como de los genes PPE en el genoma de esta micobacteria contribuyen significativamente al reducido tamaño del genoma (Cole *et al.*, 2001). La comparación de las secuencias de los genes PE/PE_PGRS de *M. tuberculosis*, muestran una gran homología entre los dominios PE y PGRS, pero el polimorfismo particularmente ocurre dentro del dominio PGRS el cual es rico en GC y las proteínas codificadas por estos genes contienen regiones repetidas ricas en glicina. Por ejemplo, la comparación entre la PE_PGRS Rv0746 con la proteína correspondiente en *M. bovis* BCG muestra que estos genes difieren por una inserción de 29 codones y una delección de 46 codones en el dominio PGRS de *M. tuberculosis* (Cole *et al.*, 1998). Las cepas de la vacuna BCG carecen completamente de ciertos genes PE/PE_PGRS y PPE encontrados en *M. tuberculosis* H37Rv (Behr *et al.*, 1999).

El análisis comparativo de los genomas de *M. tuberculosis* H37Rv y del aislado clínico CDC1551, mostró que las diferencias entre los dos está determinada principalmente por la variación entre los genes PE/PE_PGRS (Fleischman *et al.*, 2002), y estas variaciones se deben a inserciones y/o delecciones y mutaciones puntuales. En

un estudio reciente, el gen homólogo a la PE_PGRS Rv1818c de *M. tuberculosis* H37Rv fue amplificado y secuenciado en 123 aislados clínicos de *M. tuberculosis*. El 68.3 % de los aislados mostraron variaciones, del tipo inserciones, deleciones y cambios en un único nucleótido en la región PGRS con respecto al gen Rv1818c. Estas observaciones muestran que las variaciones en la región PGRS interesantemente ocurren tanto interespecie como intracepa. La amplia diversidad genética apoya la hipótesis de que este gen y probablemente otros genes PE_PGRS sirvan como una fuente de variación antigénica en las micobacterias (Talarico *et al.*, 2005).

Aunado a esto, se ha visto que ciertas proteínas PE_PGRS se encuentran en la superficie de las micobacterias y podrían tener funciones antigénicas y estructurales (Banu *et al.*, 2002; Brennan *et al.*, 2002; Brennan *et al.*, 2001). Igualmente, estudios por microscopia de fluorescencia de *M. smegmatis* que expresa el gen PE_PGRS Rv1818c fusionado a la proteína verde fluorescente muestran que esta proteína se localiza en la superficie de la micobacteria (Delogu *et al.*, 2001). Interesantemente, la mutación del gen PE_PGRS Rv1818c en BCG altera significativamente la interacción bacteria-bacteria y la interacción bacteria-macrófago, sugiriendo que esta proteína pudiera funcionar como una adhesina de superficie (Brennan *et al.*, 2001). De hecho hay evidencia que muestra que algunas proteínas PE_PGRS unen también fibronectina, sugiriendo una interacción de la micobacteria con las células del huésped y la matriz extracelular (Abou-Zeid *et al.*, 1991; Espitia *et al.*, 1999).

La similitud de los dominios PGRS ricos en glicina con las proteínas fibrosas encontradas en el gusano de seda (Hayashi and Lewis, 1998) y en la pared celular de las plantas (Sachetto-Martins *et al.*, 2000; Ye *et al.*, 1991) que se caracterizan por su elasticidad y fuerza ténsil sugiere que las proteínas PE_PGRS pudieran jugar un papel importante en la estructura de la pared celular de las micobacterias.

La expresión de los genes PE_PGRS *in vivo* se infirió inicialmente por la reacción de los sueros de animales o humanos infectados por *M. tuberculosis* con las proteínas PE_PGRS recombinantes (Delogu *et al.*, 2001; Espitia *et al.*, 1999; Singh *et*

al., 2001). Sin embargo, la presencia de reactividad cruzada de epitopes con el dominio repetitivo de las PGRS, impide concluir que la positividad de un antisuero esté dirigida contra una proteína PE_PGRS específica.

Recientemente, la expresión de los genes PE_PGRS así como PE y PPE, se ha evidenciado con el uso del RT-PCR. Banu *et al.*, (2002) encontraron 10 diferentes genes PE_PGRS expresados en *M. tuberculosis* H37Rv crecido *in vitro*. Otros estudios sugieren que la expresión de algunos genes PE_PGRS ocurren únicamente *in vivo* (Dheenadhayalan *et al.*, 2006), tal es el caso de 2 genes de *M. marinum* que corresponden a las PE_PGRS *Rv1651c* y *Rv0381* de *M. tuberculosis*. Estos genes se han asociado con virulencia, ya que la delección de los mismos lleva a una menor sobrevivencia de *M. marinum* en tejido granulomatoso en el hospedero batracio de esta micobacteria (Ramakrishnan *et al.*, 2000).

Las investigaciones recientes utilizando la técnica de microarreglos, muestran que la expresión de los genes PE, *Rv0285*, *Rv0916c* y *Rv1195*, pudieran estar bajo el control de un sistema regulador dependiente de hierro (Rodríguez *et al.*, 2002), o depender de las condiciones de nutrientes (Betts *et al.*, 2002) o de ambientes acídicos (Fisher *et al.*, 2002); el uso de la tecnología de la expresión *in vivo* basado en la recombinasa (RIVET) ha indicado que el gen PE_PGRS *Rv0834c* está sobre regulado bajo condiciones de bajo pH (Saviola *et al.*, 2003).

Utilizando también la tecnología de microarreglos Voskuil *et al.*, en 2004 demostraron que tanto las secuencias PE/PE_PGRS como las PPE se expresan en respuesta a estímulos tan variados como la disminución de oxígeno, la limitación de nutrientes, los cambios de pH, los cambios en temperatura, etc. Estos trabajos apoyan la hipótesis de que las proteínas PE/PE_PGRS y PPE podrían ayudar a las micobacterias a sobrevivir a los cambios dramáticos en el microambiente del hospedero. Sin embargo, no existen estudios sobre las proteínas PE_PGRS nativas de *M. tuberculosis* u otras micobacterias, por ejemplo si éstas poseen modificaciones post-traduccionales, ya que hasta la fecha no se han podido aislar proteínas PE_PGRS

nativas. Recientemente, por primera vez se identificó una proteína de la familia PE/PE_PGRS la Rv0532, por espectrometría de masas. La expresión del gen que codifica para esta proteína se indujo en condiciones de falta de oxígeno y en presencia de AMP cíclico. Con este trabajo una vez mas se demuestra la gran versatilidad de estas proteínas y la importancia potencial que pudieran tener en la sobrevivencia de las micobacterias especialmente cuando son sometidas a condiciones extremas (Gazdik and McDonough, 2005).

2. JUSTIFICACIÓN

La familia PE de *M. tuberculosis* está compuesta por aproximadamente 100 genes con un alto porcentaje de identidad presentes únicamente en el género micobacteria. La importancia potencial de estas secuencias, como factores de virulencia, moléculas de reconocimiento o como fuentes de variación antigénica las hacen blanco de estudio. De forma indirecta, se ha demostrado que algunos miembros de la subfamilia PE/PE_PGRS son expresados en *M. tuberculosis*; sin embargo, hasta la fecha no existen datos que demuestren la expresión diferencial entre cepas de la misma especie o en diferentes especies.

Por otra parte, el polimorfismo genético existente en las secuencias PE_PGRS observado por RFLP en aislados clínicos de *M. tuberculosis* así como las diferencias observadas en estas secuencias después de la comparación de los genomas de *M. tuberculosis* H37Rv y CDC1551, sugieren que estos genes podrían ser relevantes para evadir la respuesta inmune generando una elevada diversidad de posibles antígenos. Además, dadas a las variaciones observadas en la expresión de las mismas sugiere que estas podrían expresarse diferencialmente entre las cepas. La identificación y caracterización de las secuencias PE/PE_PGRS que se están expresando en las micobacterias que infectan al humano así como el papel que ellas juegan en la respuesta del huésped será de gran importancia tanto al conocimiento básico, como al diseño de nuevos métodos inmunoprolácticos en tuberculosis.

3. HIPÓTESIS

Dada la diversidad entre los miembros de la familia PE en los aislados clínicos de *M. tuberculosis* es factible suponer que estos genes se expresan diferencialmente entre las cepas.

4. OBJETIVO GENERAL

Determinar si existe expresión diferencial de los genes de la familia PE/PE_PGRS entre los aislados clínicos de *M. tuberculosis*.

5. OBJETIVOS ESPECÍFICOS

1. Diseñar oligonucleótidos que permitan amplificar simultáneamente por RT-PCR varios genes de la familia génica PE.
2. Analizar e identificar las secuencias amplificadas por RT-PCR de los aislados clínicos de *M. tuberculosis*.
3. Analizar y comparar por patrones de restricción de los productos amplificados por RT-PCR de los aislados clínicos de *M. tuberculosis*.

6. MATERIALES Y MÉTODOS

6.1 BACTERIAS Y CONDICIONES DE CULTIVO

En el presente estudio se utilizaron 4 cepas de *M. tuberculosis*: la cepa de referencia H37Rv que fue obtenida de American Type Culture Collection (ATCC, Rockville, MD, USA), *Mycobacterium canetti* NLA009600046 (cepa 046) y los aislados clínicos NLA009501000 (cepa 1000, perteneciente al genotipo Beijing) y la NLA009501463 (cepa 1463). Las tres últimas fueron donadas por Dr. Dick van Soolingen (National Institute of Public Health and Environment, The Netherlands), incluyendo a una cepa de *M. fortuitum* (028) que fue usada como control negativo ya que esta carece de secuencias PE demostrado por RFLP.

Las cepas 046, 1000 y 1463 son representantes de un grupo de aislados que fueron clasificadas epidemiológica y molecularmente como miembros de una misma familia basados en el patrón del polimorfismo de la longitud de los fragmentos de restricción (del inglés: RFLP Restriction Fragment Length Polymorphism) generados por la secuencia de inserción IS6110 (Fig. 1).

6.2 DISEÑO DE OLIGONUCLEÓTIDOS

Se seleccionaron 31 secuencias PE/PE_PGRS al azar y se alinearon las primeras 700 pb a partir del codón de inicio ATG, con el programa GenWork (anexo 4). Las regiones más conservadas fueron localizadas en las posiciones 1-20 y 214-235 y fueron seleccionadas para generar una secuencia consenso de oligonucleótidos basado en la frecuencia de cada base. Los oligonucleótidos PGRS1V 5´-ATG TCG TTT GTG ATC GCG GC 3´ y el antisentido, PGRS2RU 3´TAC TGG TCA AGC ACG TCC GGA-5´ fueron sintetizados para amplificar un fragmento de 235 pb.

También fueron diseñados oligonucleótidos específicos para la amplificación de un fragmento de 170 pb del gen PE_PGRS Rv3652 usando el programa GeneFisher

(<http://bibiserv.techfak.uni-bielefeld.de/genefisher/>) Rv3652A 5´- GGC CAC GGA TTT GGC TACTCT C3´ y Rv3652B 5´-CAA GGT ACT CGC CAA GCA CGT-3.

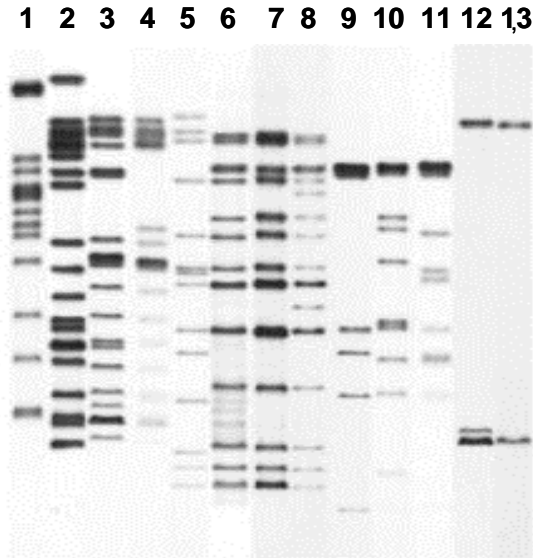


Figura 1. Polimorfismo de Longitud de los Fragmentos de Restricción (RFLP). Patrones de restricción generados con *Pvu*I en el DNA de las diferentes cepas de *M. tuberculosis* e hibridados con la sonda IS6110. Los datos y cepas fueron proporcionados por Dr. Dick van Soolingen (National Institute of Public Health and Environment. The Netherlands). (1) H37Rv, (2) 9501000, (3) 17914, (4) 940170, (5) 9501911, (6) 9501291, (7) 2000-216, (8) 2000-367, (9) 9600368, (10) 9400104, (11) 940431, (12) 9600046 y (13) S093. En este estudio se utilizaron las cepas (1) H37Rv, (2) 9501000 y la (12) 9600046. López B, et al. Clin Exp Immunol. 2003; 133:30–37.

6.3 SOUTHERN BLOT

Para demostrar que los oligonucleótidos PGRS1V y PGRS2RU amplificaban un fragmento de 235 pb de dos o más secuencias de la familia PE_PGRS, se utilizaron como controles positivos dos genes de la familia PE_PGRS (RV 1759c y PGRS 81), clonados en el vector pUC18 y pUC19, respectivamente, los cuales se encuentran transformando a la cepa XL1-blue de *E. coli*.

Los plásmidos se aislaron y se purificaron con el kit Qiagen Plasmid Mini. El DNA fue precipitado con etanol al 70%, resuspendido en agua y cuantificado por espectrofotometría.

Para demostrar la especificidad de los fragmentos de 235 pb producidos por PCR y RT-PCR de las cepas estudiadas se llevó a cabo un Southern blot, usando como sonda el gen completo de 3.6 Kb de la PGRS81 insertado en pUC19, el cual se liberó con la enzima *Kpn* I por digestión toda la noche a 37°C. El fragmento fue separado por electroforesis en gel de agarosa al 1% y purificado. Esta sonda fue marcada con el sistema de marcaje y detección de ácidos nucleico de ECL Direct Nucleic Acid Labelling and Detection Systems (Amersham Life Science). 100 ng de DNA (10 ng/μl) del gen PGRS81 fueron desnaturalizados por ebullición durante 5 min en un tubo de 500 μl, inmediatamente el tubo fue colocado en baño de hielo por 5 min y se le adicionó un volumen igual de reactivo de marcaje. Posteriormente, se agregó un volumen igual al inicial, de solución de glutaraldehído y fue incubado por 20 min. a 37°C.

Los fragmentos de 235 pb producido por el PCR del DNA y el DNA complementario (cDNA) de *M. tuberculosis*, Rv1759c y PGRS81 fueron separadas en geles de poliacrilamida al 12%, en condiciones desnaturalizantes y no desnaturalizantes. Después de la electroforesis, el DNA fue transferido a una membrana nylon (Hybond-N+ amershams, Life science) a 30 volts por 2 hrs a 4°C, usando Tris-Glicina al 1X (50 mM Tris-base, 380 mM glicina pH 8.3) como solución amortiguador de transferencia, posteriormente la membrana fue lavada por 1 min con Tris-Glicina al 0.1X en agitación suave y fijada por 5 min con la emisión de luz UV emitida por un transiluminador con lámpara de 254 nm de longitud de onda. La hibridación se llevó a cabo en un horno de hibridación con tubos rotatorio, la membrana fue pre-hibridada a 42°C por 1 hr, con 5 ml de solución de prehibridación (0.5 M de NaCl, 5% de bloqueador). Después de descartar la solución, se agregó nueva solución de prehibridación con los 100 ng de la sonda marcada y fue incubada toda la

noche a 42°C. Al día siguiente, la membrana fue lavada con solución primaria (0.4% de SDS, 0.5X de SSC) por 20 min a 42°C. Posteriormente, se hizo un segundo lavado con SSC al 2X por 5 min. La detección se realizó incubando la membrana por 1 min. a temperatura ambiente (TA), cubierta con la mezcla de reactivo de detección (350 µl de reactivo 1 y 350 µl de reactivo 2) y colocada sobre un papel tipo envolpack para evitar el secado de la membrana. Posteriormente, las bandas de hibridación fueron reveladas usando películas autoradiográficas expuestas sobre la membrana, dentro de un cassette para películas de rayos X.

6.4 EXTRACCIÓN DEL DNA GENÓMICO

Las cepas de micobacterias fueron cultivadas en medio de cultivo 7H9 enriquecido con ADC (Albúmina, Dextrosa y Cisteína) en agitación a 37°C de 2-4 semanas hasta obtener la fase logarítmica de crecimiento. Se cosecharon 3 ml de cultivo por centrifugación a 2500xg durante 10 min a TA. Se descartó el sobrenadante y la pastilla fue resuspendida en 400 µl de solución amortiguadora TE 1X (tris 10 mM, EDTA 1 mM pH 8.0), la muestra fue transferida a un tubo de 1.5 ml, e incubada a 80°C por 30 min, Posteriormente, cuando la muestra bacterias alcanzó la TA, se le adicionó 50 µl de lisozima (10 mg/ml) y se incubó a 37°C toda la noche. Al día siguiente, la muestra se incubó a 65°C con una mezcla de 5 µl de proteinasa K (10 mg/ml) y 70 µl de SDS al 10%. Al finalizar la incubación se le adicionó 100 µl de NaCl 5 M y 100 µl de solución NaCl/CTAB) (Bromuro de Cetil Trimetil Amonio) pre-calentado a 65°C; la muestra se incubó por 10 min a 65°C. El DNA fue purificado con la adición de 750 µl de cloroformo:alcohol isoamílico 24:1 y centrifugado a 12,000xg por 5 min a TA. La fase superior acuosa fue transferida a un tubo nuevo de 1.5 ml. La precipitación del DNA se realizó con 0.6 volúmenes de isopropanol a -20°C por 30 min. El DNA fue recuperado por centrifugación a 10,000xg por 5 min a TA. Posteriormente, la pastilla de DNA fue lavada con 1 ml de etanol al 70% y

centrifugado a 12,000xg por 5 min a TA. Finalmente, el DNA fue resuspendido en agua estéril y almacenado a -70°C hasta su uso.

6.5 EXTRACCIÓN DEL RNA (KINGER *et al.*, 1993)

Las cepas fueron cultivadas en medio 7H9 enriquecido con ADC en agitación a 37°C durante 2-4 semanas. Una vez que los cultivos alcanzaron la fase logarítmica de crecimiento, se cosecharon 50 ml por centrifugación a 2500xg por 10 min a TA. La pastilla fue resuspendida en 2 ml de amortiguador de lisis Tripure™ (Roche Applied Science) y la suspensión celular fue inmediatamente transferida a tubos con rosca de 1.5 ml conteniendo de 300-400 µl de perlas de vidrio de 106 micrómetros de diámetro (Sigma, St. Louis, MO). Las células fueron mecánicamente lisadas en un homogenizador Fast-Prep (Bio101, Savant) por dos pulsos de 45s a 4.5 m/s, seguidas por cuatro pulsos de 15s a 6.5 m/s. Entre cada pulso, las muestras fueron mantenidas en hielo por 2-5 min. Los ácidos nucleicos fueron entonces extraídos con 200 µl de cloroformo, seguido por una centrifugación a 4°C por 10 min a 10,000xg. La fase superior acuosa fue transferida a un nuevo tubo de 1.5 ml y el RNA total fue precipitado por 30 min a TA con 0.13 volúmenes de isopropanol. La pastilla fue lavada con etanol al 75% y resuspendido el RNA total en 30-50 µl con agua tratada con DEPC (di-etil pirocarbonato). La calidad del RNA total purificado fue analizada mediante la electroforesis en gel de agarosa. El RNA fue almacenado a -70°C.

6.6 RT-PCR

Previa a la reacción de la transcriptasa reversa, 2 µg de RNA fueron tratados con DNasa I grado amplificación (Gibco BRL) durante 30 min a TA. Posteriormente la enzima fue inactivada con 2.5 mM de EDTA a 37°C por 15 min.

La síntesis de la primera cadena a cDNA fue llevada a cabo usando el protocolo descrito para secuencias ricas en GC (SuperScript™ Preamplification System for

First Strand cDNA Síntesis, Gibco BRL). La reacción fue realizada con 0.5 µg de oligo dT, 20 mM de Tris HCl (pH 8.4), 50 mM de KCl, 2.5 mM MgCl₂, 0.5 mM de mezcla de dNTPs, 10 mM de DTT y 1 U de superscript II reversa transcriptasa. Las muestras fueron incubadas a 50°C por una hora. Subsecuentemente, el PCR fue realizado con 10 pmol de cada oligonucleótido (PGRS1V y PGRS2RU) en 1X de amortiguador, 1 mM de MgCl₂, 0.2 mM de cada dNTPs, 2-5 µl de muestra de cDNA y 1.25 U de amplitaq Gold Taq polimerasa (Applied biosystems). Los ciclos de temperatura fueron los siguientes: Un ciclo de 8 min a 94°C; seguidos por 30 ciclos con 1 min a 94°C, 1.5 min a 50°C y 1.5 min a 72°C; y un ciclo de 7 min a 72°C. Los productos amplificados fueron analizados por electroforesis en un gel de poliacrilamida al 8%. El mismo protocolo de PCR fue usado para amplificar el DNA de las cepas analizadas.

La PCR usando los oligonucleótidos Rv3652A y Rv3652B fue llevada a cabo con el fragmento producido por RT-PCR de 235 pb y el DNA total, obtenido de todas las cepas estudiadas. Se utilizaron las mismas condiciones de reacción mencionadas arriba, a excepción de la temperatura de alineamiento a 60°C.

Para la identificación de los genes amplificados por RT-PCR se utilizaron diferentes estrategias.

6.6.1 Polimorfismo Conformacional de Cadena Sencilla (SSCP) (Fujita and Silver, 1995; Orita *et al.*, 1989)

El fragmento purificado de 235 pb amplificado por RT-PCR y PCR obtenido de los diferentes aislados fue analizado por SSCP (RT-PCR_SSCP). De 100 a 200 ng del producto amplificado fueron desnaturalizados con amortiguador de carga desnaturalizante (20 mM de EDTA pH 8.0, 0.5% de azul de bromofenol, 0.5% de xilen-cianol en 95% de formamida) en una proporción de 1:6 y mantenidos en baño maría a 94°C por 8 min, y posteriormente las muestras fueron mantenidas en hielo por 5 min o hasta su uso.

El SSCP se analizó en geles de poliacrilamida no desnaturalizantes al 8 y al 12% (poliacrilamida al 30%, Tris-glicina al 5%, 10% de Persulfato de amonio y 7 μ l de TEMED), la electroforesis se realizó en una cámara de tipo vertical de 8 X 10 cm². a 10 mA a 4 °C por 5 hrs, en el amortiguador de corrida Tris-Glicina 1X. Las bandas en los geles fueron visualizadas con tinción de plata (se utilizó el protocolo descrito por Silver Stain Plus (Bio-Rad), omitiendo la solución incrementadora de revelado), bromuro de etidio y/o marcaje con digoxigenina dependiendo del tipo de análisis.

6.6.2 Extracción, purificación y secuenciación de los amplificadores resueltos por RT-PCR_SSCP

Para determinar e identificar los genes amplificados dentro del fragmento de 235 pb, las bandas resueltas por RT-PCR_SSCP fueron teñidas con bromuro de etidio. Una vez identificadas, las bandas visibles se seccionaron cuidadosamente y cada una de ellas fueron transferidas a un tubo de polipropileno para PCR de 500 μ l. El pedazo de gel fue fragmentado mecánicamente con una punta de polipropileno de 10 μ l. Se realizó un lavado de 1 min. con 200 μ l de agua estéril y la muestra se centrifugó a 10,000xg por 3 min y cuidadosamente se retiró la solución de lavado. Posteriormente, el material genético incluido en el gel fragmentado, fue directamente sometido a la amplificación por PCR usando los primers 2RU y 1V bajo las mismas condiciones de reacción descritas anteriormente. El producto amplificado fue cargado en un gel de agarosa al 1.5% y posteriormente, fue purificado y enviado para determinar su secuencia.

6.7 RT-PCR y RFLP-PCR

Los fragmentos de 235 pb obtenidos de la amplificación fueron extraídos y purificados mediante el procedimiento de QIAquick Gel Extraction Kit. Una vez purificados fueron digeridos con diferentes enzimas de restricción. El fragmento y sus productos fueron entonces separados en geles de agarosa al 1.5 y 2.0 %,

respectivamente. Se utilizó la solución TAE al 1X como amortiguador de corrida. Finalizada la corrida el gel fue sumergido en 250 ml de solución teñidora (Bromuro de Etidio 10 mg por mililitro) y se visualizaron las bandas en un transiluminador. La banda de interés fue extraída y colocada en un tubo de 1.5 ml. De acuerdo al peso del gel extraído le fueron adicionados 3 volúmenes de amortiguador QG (100 mg correspondería a 100 μ l). Para disolver el gel, el tubo fue incubado a 50°C por 10 min. Posteriormente, el DNA fue precipitado con la adición de un volumen de isopropanol y la suspensión fue cargada en una columna de purificación. El DNA se unió a la columna a través de 1 min. de centrifugación a 8000xg a TA. El DNA fue lavado con 0.5 ml de amortiguador QG y luego con 0.75 ml de amortiguador PE. En cada lavado fue descartada la solución amortiguadora por centrifugación por 1 min. a 8000xg a TA y 5 min, respectivamente.

El DNA fue liberado de la columna mediante la adición de 30 μ l de agua y centrifugación a 13000xg. Posteriormente fue enviado para la secuenciación automatizada usando el Secuenciador ABI PRISM 310 Genetic Analyzer PE (Applied Biosystems).

7 RESULTADOS

7.1 DISEÑO DE LOS OLIGONUCLEÓTIDOS Y CONDICIONES DE AMPLIFICACIÓN

Mediante el alineamiento de las secuencias correspondientes a 31 genes PE seleccionados al azar (Anexo 4), se eligieron las 2 regiones de mayor consenso y tomando en consideración la frecuencia de cada base en dichas posiciones, se definió la identidad del nucleótidos (Fig. 2). Así, se obtuvieron los oligonucleótidos PGRSIV y PGRS2RU, los cuales fueron usados para amplificar el fragmento esperado de 235 pb en el DNA y DNA complementario (cDNA) de las cepas estudiadas.

	1	45	227	270				
1	SEQ27	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	AGGCGTTTCATGCC	CAGTTCGTGCAGGCA	TTGACCTCCGG-TGG	251
2	SEQ29	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	AGGCGTTTCATGCC	CAGTTCGTGCAGGCA	TTGACCTCCGG-TGG	251
3	SEQ31	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	AGGCGTTTCATGCC	CAGTTCGTGCAGGCG	TTGACCTCCGG-TGG	251
4	SEQ14	-----ATG-TC	AT-GGG-TGATGTT	TCGCCGAGCTGGTG	CGGCGTTTACGCC	CAGTTCGTGCAGGCC	CTGAGCGCCGG-CGG	251
5	SEQ28	-----ATG-TC	GT-TTG-TGATCGCG	GTGCCGAGACTATC	CGGCGTTTCATGGT	CGGTTTCGTGCAGGCG	CTGACCCCGG-GGG	251
6	SEQ12	-----ATA-TC	GT-TTG-TGATCGCG	GCGCCGAGGCCATC	CGACGTTTCATATC	CAGTTTGTGCAGGCC	TTGACCCCGGG-CGC	251
7	SEQ11	-----GTG-TC	CT-TTG-TGATCGCA	CAACCGGAGATGATC	CCGCTTTCACGCC	CAGTTCGTGCAGGCC	CTGACCTGGA-CGC	251
8	SEQ7	-----ATG-TC	AT-TTG-TGATCGCG	GCACCGGAGTTTTTA	CGGCGTTTCATCAG	CAATTTGTGCAGGCC	TTGACCCCGGG-CGC	251
9	SEQ13	-----ATG-TC	GT-TCG-TGCTGGCG	ATGCCGAGGTTGTTG	CGGCGTTTCACGCC	CAGTTCGTGCAGGCC	TTGACCCCGGG-TGC	251
10	SEQ22	-----ATG-TC	GT-TTG-TAATCGCC	GCTCCGAGACGCTG	CGGCGTTTCATGCC	CAGTTCGTGCAGGCC	TTGACCCCGGG-CGC	251
11	SEQ23	-----ATG-TC	GA-ACG-TGATGGTA	GTCCCGGAATGTTG	CCAGATTTCAATA	CAGTTCGTGCAGGCC	TTAACCAGGAG-CGT	251
12	SEQ30	-----ATG-TC	GT-TTG-TCAGCGTA	GCCCGGAGATTGTTG	CGGCGTTTCATCAG	CAGTTCGTGCAGGCC	CTTGCCGGTGG-CGC	251
13	SEQ10	-----ATG-TC	GT-TTG-TGATTGCT	GCGCCGAGGCGTTG	CGGCGTTTCATGCC	CAGTTCGTGCAGGCC	TTGACCTCCGG-CGG	251
14	SEQ8	-----ATG-TC	GT-TTG-TAGTCGTG	GCGCCGAGGTTGTTG	CGGCGTTTCACGCC	CAGTTCGTGCAGGCC	TTGACCCGGTGC-CGG	251
15	SEQ18	-----ATG-TC	GT-TGG-TGATCGTG	GCCCGGAGACGGTG	CGGCGTTTCATGAC	CAGTTCGTGCAGGCC	TTAACCAGCCG-GGT	251
16	SEQ19	-----ATG-TC	GT-TGG-TGATCGTA	ACCCCGGAGACGGTG	CAGCGTTTCATGAC	CAGTTCGTGCAGGCC	TTAACCAGCCG-GGT	251
17	SEQ20	-----ATG-TC	CT-ACA-TGATTGCG	GTGCCAGACATGCTC	CAGCGTTTCATGAC	CAGTTCGTGCAGGCC	CTGACGTCAAG-TGC	251
18	SEQ16	-----GTG-TC	GT-TTG-TGTTGGTG	TCTCCGTCGACGCTG	CGCTGTTTCACGCC	CAGTTCGTGCAGGCC	TTGACCCGGGG-GGC	251
19	SEQ17	GAGGAACCGTG-TC	GT-TTG-TGTTGGTG	TCTCCGTCGACGCTG	CGCTGTTTCACGCC	CAGTTCGTGCAGGCC	TTGACCCGGGG-GGC	251
20	SEQ25	-----ATG-TC	GT-TTG-TGATTGCG	GTTCCGAGTTCCTG	CGACGTTTCATGCC	CAGTTCGTGCAGGCC	TTGAGTCCGGG-TGC	251
21	SEQ6	-----ATG-TC	AT-ATG-TGATCGCG	GCGCCGAGGCGCTG	CGGCGTTCCATGAG	CGGTTTCGTGCAGGCC	TTGACCCAGGG-TGG	250
22	SEQ26	-----ATGGTC	ATGTCGCTGATGGTG	GCGCCGAGGCTGGTG	CGACGTTTCATGAG	CAGTTCGTGCAGGCC	CTGACCCGGGG-TGG	254
23	SEQ1	-----ATG-TC	GT-TCG-TTTCGCG	GTGCCAGAGATGGTG	CGGCGTTTCATGAC	CGGTTTCGTGCAGGCC	CTAAGCCCGCC-AGC	251
24	SEQ3	-----ATG-TC	GT-TTG-TGTTGGTT	TCGCCGAGACCGTG	CAGCGTTTCATAGC	CGGTTTCGTGCAGGCC	CTGACCCCGGG-TGC	251
25	SEQ4	-----ATG-TC	GT-TCG-TGTTGATC	GCACCGAATTCGTTG	CGGCTTACCACCG	CGGTTTCGTGCAGGCC	TTGAGTACCGG-CGC	251
26	SEQ5	-----ATG-TC	GT-TCG-TGTTGATC	GCACCGAATTCGTTG	CGGCTTACCACCG	CGGTTTCGTGCAGGCC	TTGAGTACCGG-CGC	251
27	SEQ21	-----ATG-TC	GT-TTG-TGTTCCGG	GCGCCAGAGGCACTG	ACAGCCTTCCACGACC	CAGTTCGTGCAGGCC	CTGAGAGCAAG-TGC	254
28	SEQ2	-----ATG-TC	GT-TCG-TGTTGACA	GTGCCGAGGCGCGTG	CGGCGTTTCATACC	GAGTTCGTGCAGGCC	TTGACCCCGGG-CGC	251
29	SEQ9	-----ATG-TC	AT-ACA-TGATTGCG	ACACCGAGGCGGTTG	CGGCGTTTCACGCC	CTGTTTCGTGCAGGCC	TTGACTGGGGGTTGC	252
30	SEQ15	-----ATG-TC	AT-TTG-TGATCGTG	GCGCCGAGGCGGTTG	CGGCGTATCACGAG	CAGTTCGTGCAGGCC	TTGACCCCGGG-AGC	251
31	SEQ24	-----ATG-TC	AT-TCC-TGACGACA	CAGCCTGAAGGTTG	CTGCGGTATACGAC	CTGTTTCGTGCAGGCC	CTGCGGGTGCAG-CGC	254
	consenso	-----ATG TC	GT-TTG-TGATCGCG	GC-----	-----ATGAC	CAGTTCGTGCAGGCC	T	

Figura 2. ALINEAMIENTO DE SECUENCIAS DE LA FAMILIA PE. Se seleccionaron 31 secuencias de la familia PE obtenidas de la base de datos del genoma de *M. tuberculosis* H37Rv. Se alinearon las primeras 251 pb desde el ATG. Se seleccionaron las secuencias mas conservadas para diseñar un par de oligonucleótidos que amplifican por PCR un fragmento de 235 pb.

Para determinar las condiciones de reacción y especificidad de la amplificación del fragmento esperado de 235 pb se utilizó el DNA de la PGRS81 y del Rv1759c clonados en vectores y el DNA genómico de la cepa de *M. tuberculosis* H37Rv. Con el DNA de la PGRS81 y el Rv1759c se estableció que las condiciones idóneas de

amplificación eran a una temperatura de alineamiento de 50°C (Fig. 3). Con las condiciones de la mezcla de reacción de la PCR en el DNA del gen PE_PGRS Rv1759c y el DNA genómico de la cepa de *M. tuberculosis* H37Rv se demostró la especificidad de los oligonucleótidos al amplificar el fragmento esperado de 235 pb sólo cuando dicho par de oligonucleótidos es utilizado (Fig. 4).

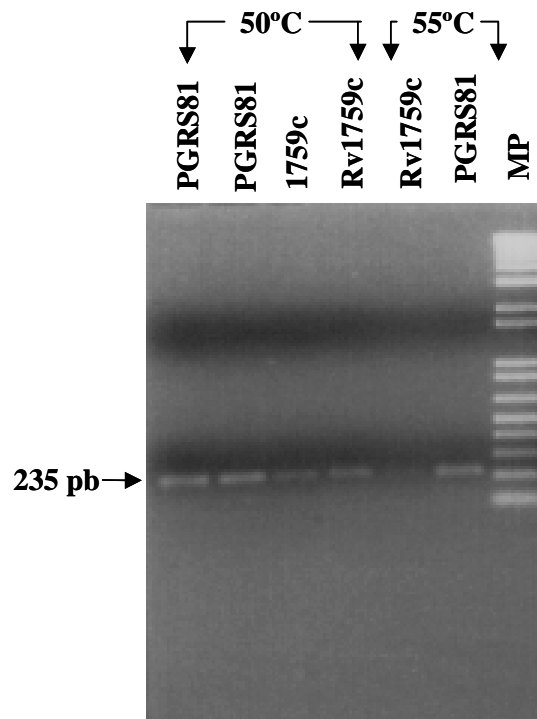


Figura 3. PCR. Amplificación del fragmento de 235 pb del DNA de la PGRS81 y Rv1759c usando una temperatura de alineamiento de 50°C y 55°C, con 1.5 mM de MgCl₂. MP marcador de peso molecular en escalera de 100 pb. Gel de agarosa al 1.5%.

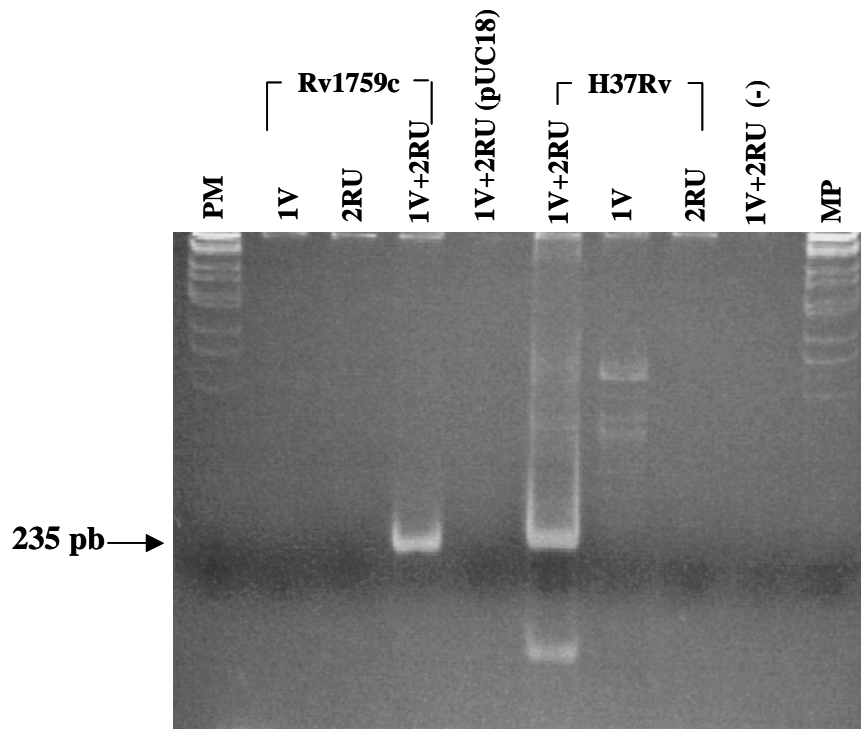


Figura 4. Estandarización de la PCR. Se utilizaron como controles el DNA del gen Rv1759c clonado en pUC18, el vector pUC18 sin el gen Rv1759c y el DNA genómico de *M. tuberculosis*, H37Rv. Se incluyó en la reacción un solo oligonucleótido (1V o 2RU) y en par (1V+2RU). 1V+2RU (-) Control negativo sin oligonucleótidos y MP: marcador de peso molecular. Electroforesis en gel de poliacrilamida al 8%.

Posteriormente, se establecieron las condiciones de aislamiento del RNA y su transcripción inversa. El rompimiento de las micobacterias con un homogenizador para tejidos, usando perlas de vidrio de 0.16mm de diámetro demostró ser la técnica de mayor utilidad. En cuanto a la purificación del RNA, la precipitación con 0.13 volúmenes de isopropanol fue la que produjo un material de mayor calidad con alto rendimiento. Para la transcripción inversa no hubo cambios significativos en el procedimiento descrito por la casa comercial. Sin embargo, para determinar cuales oligonucleótidos eran los más adecuados para sintetizar el cDNA de las cepas

estudiadas, se llevó a cabo un RT-PCR con el RNA total de la cepa de *M. tuberculosis* H37Rv, usando los oligonucleótidos azarosos y el oligo dT. En el Fig. 5 se muestra que ambos oligonucleótidos amplificaron el producto de 235 pb. Por lo tanto, se seleccionó el oligo dT para los siguientes estudios.

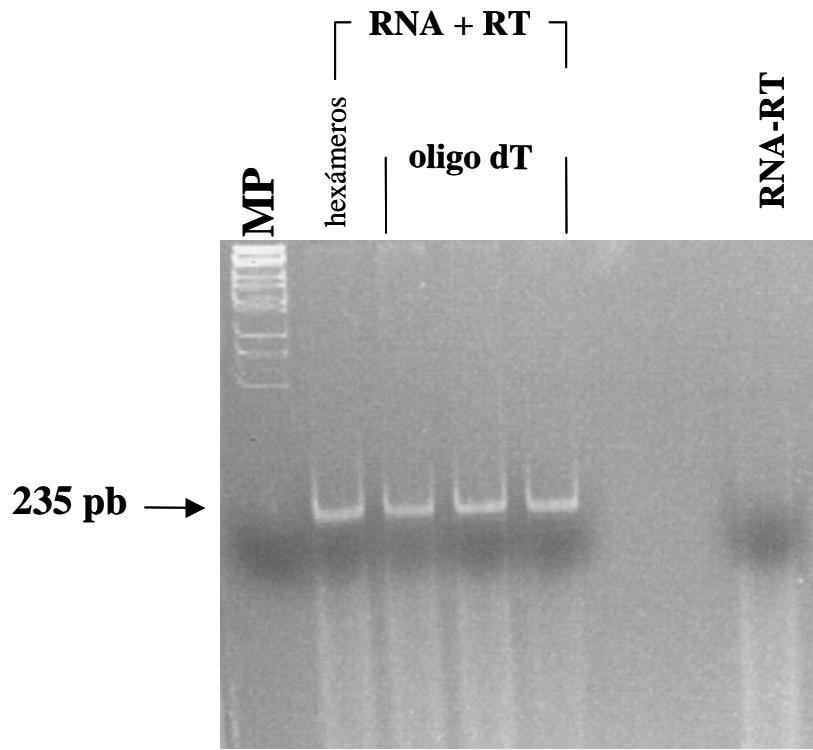


Figura 5. RT-PCR. La reacción de RT-PCR se llevó a cabo utilizando 4 μg de RNA total de *M. tuberculosis* H37Rv, los hexámeros y oligonucleótidos dT. Para el PCR se emplearon los oligonucleótidos PGRS1V y PGRS2RU. RNA+RT y RNA-RT con y sin enzima reverso transcriptasa, respectivamente. MP: Marcador de peso molecular. Electroforesis en gel de poliacrilamida al 8% teñido con bromuro de etidio.

7.2 IDENTIFICACIÓN DE LOS GENES DE LA FAMILIA PE

Con el fin de determinar si los oligonucleótidos diseñados en este estudio eran capaces de amplificar dos o más genes de la familia PE, se utilizó la técnica de SSCP (del inglés: Single Strand Conformational Polymorphic). Los fragmentos de 235 pb obtenidos tanto del DNA PE_PGRS81 como del cDNA de la cepa H37Rv se mantuvieron en doble cadena y otros fueron desnaturalizados a cadena sencilla. Las bandas fueron separadas en un gel de poliacrilamida al 12% y teñidas con nitrato de plata.

En la Figura 6, se muestra que en el fragmento desnaturalizado de 235 pb producido por el DNA de la PE_PGRS81 produce las dos bandas esperadas; mientras que la desnaturalización del fragmento de 235 pb producido por RT-PCR de la cepa H37Rv generó al menos 10 bandas. En los DNAs controles (no desnaturalizados) no se observaron productos desnaturalizados o bandas extras.

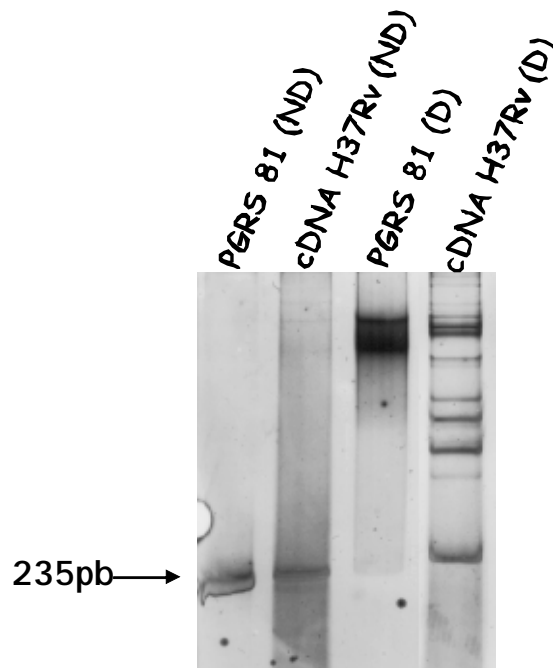


Figura 6. PCR-SSCP. Fragmentos de 235 pb amplificados del DNA de la PGRS 81 y del cDNA de la cepa H37Rv, fueron corridos en condiciones desnaturalizantes (D) y no desnaturalizantes (ND) en un gel de poliacrilamida al 12 % y teñido con tinción de plata.

Para demostrar que los fragmentos generados por PCR-SSCP correspondían específicamente a secuencias de la familia PE, se llevó a cabo un Southern Blot usando como sonda el gen completo de 3.9 Kb de la PGRS81. Los resultados muestran que la sonda hibridizó con los fragmentos amplificados de 235 pb, producidos por el PCR del DNA Rv1759c, PGRS81, H37Rv y el producto del RT-PCR de la cepa H37Rv. Interesantemente, la mayoría de las bandas generadas por la desnaturalización del producto del RT-PCR dieron señal positiva de hibridización (Fig. 7 panel A y B).

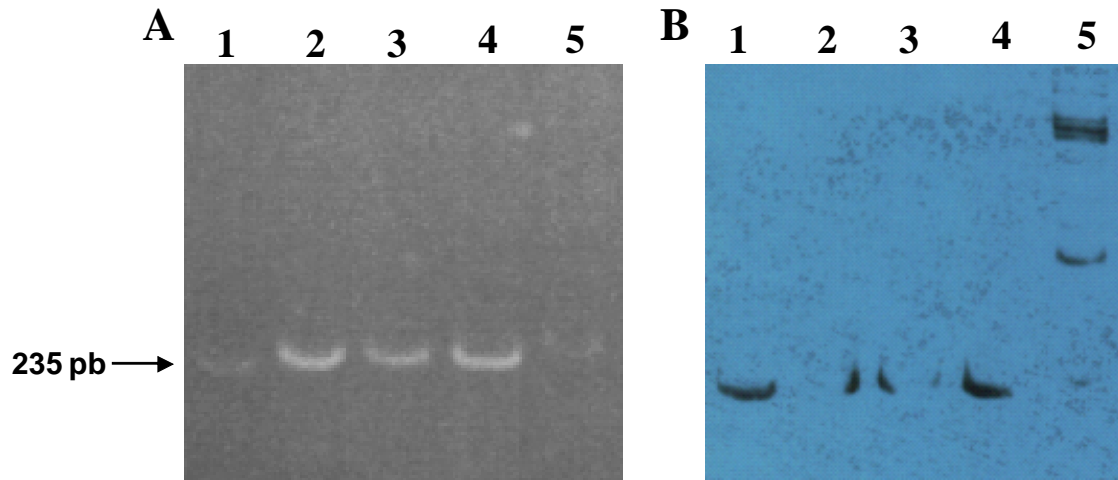


Figura 7. Southern-blot. Demostración de la amplificación específica de las secuencias PE por hibridación de los fragmentos de 235 pb producido por PCR de los DNAs de la 1759c (1), PGRS81 (2), H37RV (3), el cDNA (4) y el SSCP del RT-PCR de la cepa H37Rv de *M. tuberculosis* (5). Se utilizó como sonda el gen completo de 3.9 Kb de la PGRS81. En el panel A). RT-PCR_SSCP en un gel de poliacrilamida al 12% teñido con bromuro de etidio. B). Su correspondiente autoradiografía.

Para identificar los genes contenidos en el fragmento de 235 pb producido por RT-PCR de la cepa H37Rv, las bandas fueron separadas por el ensayo SSCP y se visualizaron con tinción de nitrato de plata. Posteriormente, para reamplificar y secuenciar cada banda, el SSCP fue teñido con la tinción de bromuro de etidio. En la Figura 8 panel A, se observan las bandas generadas por el SSCP del producto total del PCR del DNA de H37Rv, En el panel B se muestran las bandas del fragmento purificado del RT-PCR de la cepa H37Rv. Las bandas B1, B2, B3 y B4 fueron las únicas detectadas en el gel de poliacrilamida teñido con bromuro de etidio, estas fueron reamplificadas y secuenciadas. El alineamiento de las secuencias obtenidas se muestra en la Figura 9. La comparación de la secuencia con la base de datos mostró un alto porcentaje de similitud con el gen Rv1172c, considerado como un miembro de la subfamilia PE.

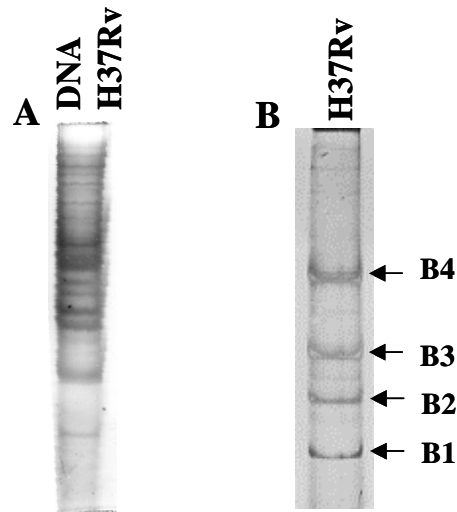


Figura 8. PCR-SSCP. Separación a cadena sencillas de los fragmentos producidos por PCR del DNA y cDNA de las cepa de *M. tuberculosis* H37Rv. En el panel A se muestra el SSCP del DNA de la H37Rv como control positivo. En el panel B se muestran los fragmentos purificados (B1-B4) obtenidos mediante RT-PCR y SSCP a partir del RNA de la cepa H37Rv. B1-B4 fueron los fragmentos purificados y secuenciados. Gel de poliacrilamida al 8% teñido con nitrato de plata.

BANDA 3	-----	-----	---CCAGAGN	CTGGGNGCGG	CCGCTGCGGN	27
BANDA 1	-----	-----	-CCAGAG-C-	CTGGGCGCGG	CCGCTGCGGA	27
BANDA 4	-----	-----	-CCAGAG-C-	CTGGGCGCGG	CCGCTGCGGN	27
1172c-c	ATGTCGTTTG	TGTTCCGCGGC	GCCAGAGGCA	CTGGGCGCGG	CCGCTGCGGA	50
BANDA 2	-----	-----	CTTNCANAGC	CTGGGCGCGG	CCG-TGCGGA	29
BANDA 3	-ATGGCCNGN	ATCGGTTCA	CTCTNAACGC	CGCCAATGTG	GTTGCGGGCG	76
BANDA 1	-ATGGCCGGT	ATCGGTTCA	CTCTTAACGC	CGCCAATGTG	GTTGCGG-CG	75
BANDA 4	-ATGGCCGGT	ATCGGTTCA	CTCTTAACGC	CGCCAATGTG	GTTGCGG-CG	75
1172c-c	CATGGCCGGT	ATCGGTTCA	CTCTTAACGC	CGCCAATGTG	GTTGCGG-CG	99
BANDA 2	-ATGGCCGGT	ATCGGTTCA	CTCTTAACGC	CGCCAATGTG	GTTGCGG-CG	77
BANDA 3	GTTCCACCA	NCGNANTNCT	GGCCGCAGCC	GNNGAACGAG	GTCTCGNCTN	126
BANDA 1	GTTCCACCA	CCGGAGTCCT	GGCCGCAGCC	G-CGGACGAG	GTCTCGACTC	124
BANDA 4	GTTCCACCA	CCGGAGTCCT	GGCCGCAGCC	G-CGGACGAG	GTCTCGACTC	124
1172c-c	GTTCCACCA	CCGGAGTCCT	GGCCGCAGCC	G-CGGACGAG	GTCTCGACTC	148
BANDA 2	GTTCCACCA	CCGGAGTCCT	GGCCGCAGCC	G-CGGACGAG	GTCTCGACTC	126
BANDA 3	ATGTCGCCNN	GCTGCTTCC	NNCCATGCTN	ANGGTATAA	CNAGCTCACC	176
BANDA 1	AGGTCGCCGC	GCTGCTTCC	GCGCATGCTC	AGGGGTATCA	GCAGCTCAGC	174
BANDA 4	AGGTCGCCGC	GCTGCTTCC	GCGCATGCTC	AGGGGTATCA	GCAGCTCAGC	174
1172c-c	AGGTCGCCGC	GCTGCTTCC	GCGCATGCTC	AGGGGTATCA	GCAGCTCAGC	198
BANDA 2	AGGTCGCCGC	GCTGCTTCC	GCGCATGCTC	AGGGGTATCA	GCAGCTCAGC	176
BANDA 3	CAAGCAGGAT	GATGACAGNT	CTAECATGAC	CAGNTEGTGC	AGGCCTA	223
BANDA 1	C-GGCA-GAT	GATGACAG-C	CTTCCATGAC	CAGTTCGTGC	AGGCCTA	218
BANDA 4	C-GGCA-GAT	GATGACAG-C	CTTCCATGAC	CAGTTCGTGC	AGGCCTA	218
1172c-c	C-GGCA-GAT	GATGACAG-C	CTTCCACGAC	CAGTTCGTGC	AGGCG--	240
BANDA 2	C-GGCA-GAT	GATGACAG-C	CTTCCATGAC	CAGTTCGTGC	AGGCCTA	220

Figura 9. Alineamiento de secuencias. El fragmento purificado de 235 pb producido por RT-PCR de la cepa H37Rv fue separado a cadena sencilla por SSCP. Las bandas B1-B4 fueron secuenciadas y alineadas con la PE Rv1172c obtenida de la base de datos.

Para determinar si existían diferencias de expresión en los genes de la familia PE entre los aislados clínicos (046, 028, 1000 y 1463) y en la cepa de referencia H37Rv, se realizó el aislamiento y purificación del RNA. En la Figura 10, panel A, se muestran los RNA purificados con y sin el tratamiento con DNasa. En el panel B, se muestra el RT-PCR de las cepas estudiadas. En todas las cepas se obtuvo el fragmento esperado de 235 pb, a excepción de la cepa 028, que se usó como control negativo.

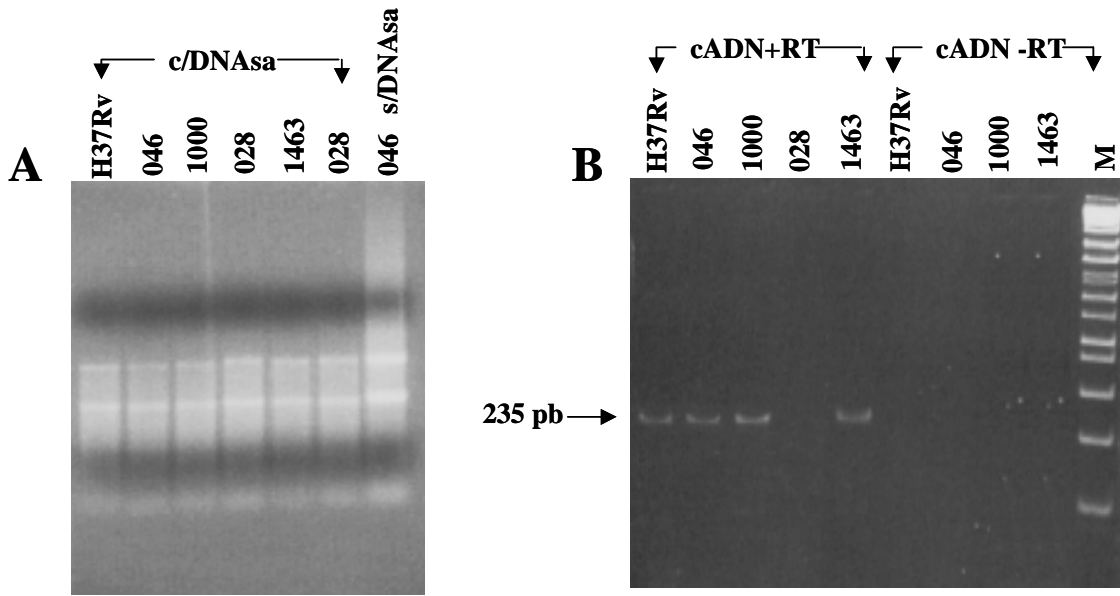


FIGURA 10. - RT-PCR. - Síntesis del ADNc y PCR de las cepas estudiadas de *M. tuberculosis*. En el panel A (gel de agarosa no desnaturalizante al 1.5%) se muestra el RNA tratado con DNasa de las cepas H37Rv, 046, 1000, 028 y 1463. En el panel B (gel de acrilamida al 8%) se muestra el RT-PCR de las respectivas cepas. +RT y -RT (con y sin la reversa transcriptasa, respectivamente). MP marcador de peso molecular en escalera de 100 pb. Geles teñidos con bromuro de etidio.

El SSCP fue usado como un primer paso para identificar los genes expresados en cada una de las cepas estudiadas. El patrón de bandas polimórficas determinó la heterogeneidad entre las cepas (Fig. 11). El SSCP permitió visualizar las 4 bandas vistas anteriormente con la cepa H37Rv (Fig. 8 panel B), estas bandas fueron similarmente compartidas entre las cepas estudiadas. Sin embargo, puesto que la tinción con bromuro de etidio es mucho menos sensible que la de plata, no fue posible visualizar las bandas que especialmente podrían establecer diferencias entre las cepas estudiadas. Por lo que no fue posible secuenciar e identificar los genes expresados.

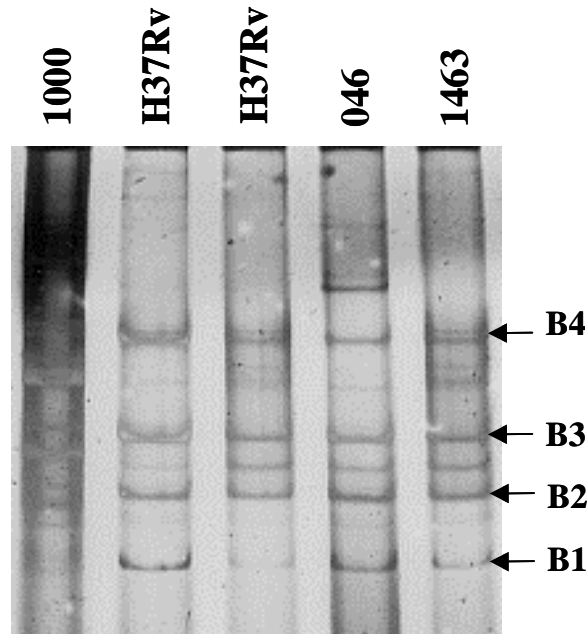


Figura 11. SSCP-PCR. La heterogeneidad entre las cepas es mostrado con el producto total del RT-PCR analizado por SSCP. En las cepas de *M. tuberculosis* 1000, H37Rv, 046 y 1463 se observan las bandas compartidas y secuenciadas (B1-B4). Gel de poliacrilamida al 12 % teñido con nitrato de plata.

Debido a las limitaciones del SSCP, se decidió utilizar otra estrategia a fin de identificar los genes presentes en el fragmento de 235 bp. Mediante la predicción de los fragmentos de restricción teóricamente generados con las diferentes enzimas presentes en la región de 235 pb de las 31 secuencias previamente seleccionadas (anexo 5) y el análisis por RFLP fue posible comprobar la presencia o ausencia de los mensajeros (RNAm) de la familia PE en las cepas estudiadas.

El análisis determinó que los sitios de restricción mas comúnmente presentes en las 99 secuencias de la familia PE fueron para las enzimas *Alu I*, *Eco RV*, *Not I*, *Pvu I*, *Sal I* y *Sph I*, esto teóricamente permitiría discriminar secuencias y facilitar el análisis en los geles de poliacrilamida (anexo 5). Para determinar qué secuencias estaban presentes dentro del fragmento de 235 pb, se utilizó el fragmento purificado

de 235 pb obtenido del DNA de H37Rv. La digestión con cada enzima se llevó a cabo en reacciones separadas. A excepción de *Eco RV*, todas las enzimas presentaron sitios de corte. La enzima *Alu I* produjo dos fragmentos de 192 y 43 pb; *Not I*, 4 fragmentos de 200, 140, 95 y 35 pb; *Pvu I* de 160 y 75 pb; *Sal I* de 144 y 91 y *Sph I* de 175 y 60 pb (Fig. 12).

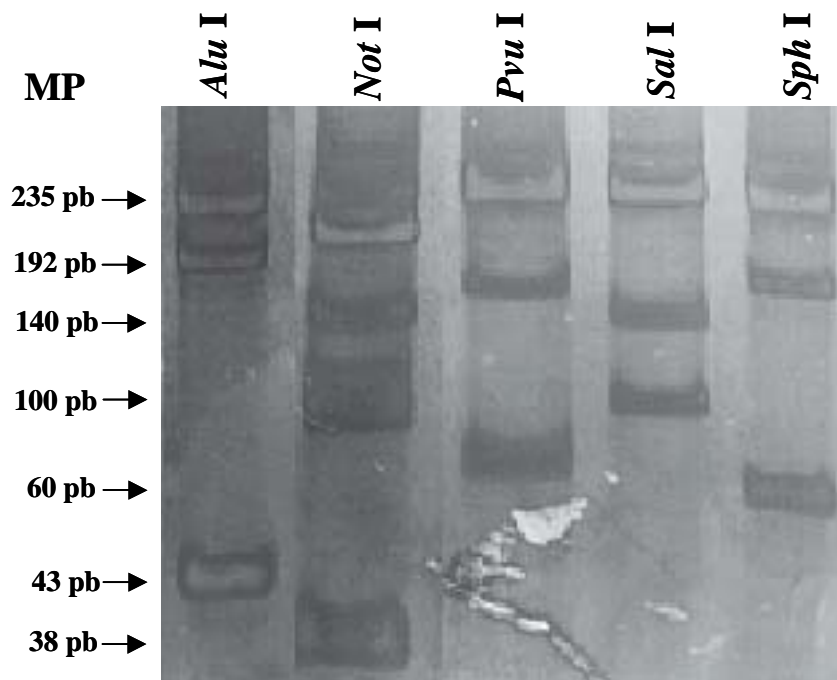


Figura 12. RFLP-PCR de *M. tuberculosis* H37Rv. polimorfismo de la longitud de los fragmentos de restricción del fragmento de 235 pb producido por PCR del DNA de H37Rv. Se emplearon las enzimas *Alu I*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento. Gel de poliacrilamida al 12% teñido con plata. MP (marcador de peso molecular en escalera de 100 pb).

Posteriormente, el producto de 235 pb purificado del RT-PCR de cada una de las cepas estudiadas fue analizado por el PCR-RFLP. Se utilizaron las enzimas *Alu I*, *Eco RV*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir los fragmentos de 235 pb. En las Figuras 13, 14, 15 y 16 se muestran los resultados de los patrones obtenidos con las

cepas H37Rv, 046, 1000 y 1463, respectivamente. Todas las cepas mostraron el mismo patrón de digestión con las enzimas *Alu* I (192 y 43 pb) y *Sph* I (175 y 60 pb). Con *Not* I, todas las cepas presentaron los 4 fragmentos de 200, 140, 95 y 35, a excepción de la cepa 046 que solo presentó los fragmentos de 35 y 200 pb (Fig. 14 y 17) y en la cepa 1463 que presentó un fragmento adicional de aproximadamente 90 pb (Fig. 16 y 17). La digestión con *Pvu* I solo fue observada en la cepa 1463, correspondiendo a los mismos fragmentos de 160 y 75 pb generados con el DNA de H37Rv (Fig. 16 y 18). No se generaron fragmentos con las digestiones de *Sal* I y *Eco* RV en el producto del RT-PCR de todas las cepas estudiadas. Un fragmento de 100 pb es coamplificado en el RT-PCR de todas las cepas estudiadas, debido a la baja resolución del gel de agarosa este fragmento fue un contaminante constante de la banda de 235 pb durante el proceso de purificación. Sin embargo, este fragmento fue secuenciado e identificado como parte del gen Rv1172c. En la tabla 1 se resume los patrones de restricción producidos a partir del fragmentos de 235 pb amplificado por PCR y RT-PCR observado en el PCR-RFLP de las cepas estudiadas.

Tabla 1. Patrones de restricción generados por RFLP-PCR del fragmento amplificado de 235 pb de las cepas estudiadas.

FIGURA	CEPA	MATERIAL DE ESTUDIO	ENZIMAS DE RESTRICCIÓN Y FRAGMENTOS PRODUCIDOS EN pb					
			<i>Alu</i> I	<i>Eco</i> RV	<i>Not</i> I	<i>Pvu</i> I	<i>Sal</i> I	<i>Sph</i> I
12	H37 Rv	DNA	192 y 43		200 y35, 140 y 95	160 y 75	144 y 91	175 y 60
13	H37Rv	cDNA	192 y 43, 100*	100	200 y35, 140 y 95, 100*	160 y 75, 100*	144 y 91, 100*	175 y 60, 100*
14	046	cDNA	192 y 43, 100*	100*	200 y35, 100*	160 y 75, 100*	144 y 91, 100*	175 y 60, 100*
15	1000	cDNA	192 y 43, 100*	100*	200 y35, 140 y 95, 100*	160 y 75, 100*	144 y 91, 100*	175 y 60, 100*
16	1463	cDNA	192 y 43, 100*	100*	200 y35, 140 y 95, 100*	160 y 75, 100*	144 y 91, 100*	175 y 60, 100*

El 100* indica el fragmento no digerido coamplificado con el de 235 pb.

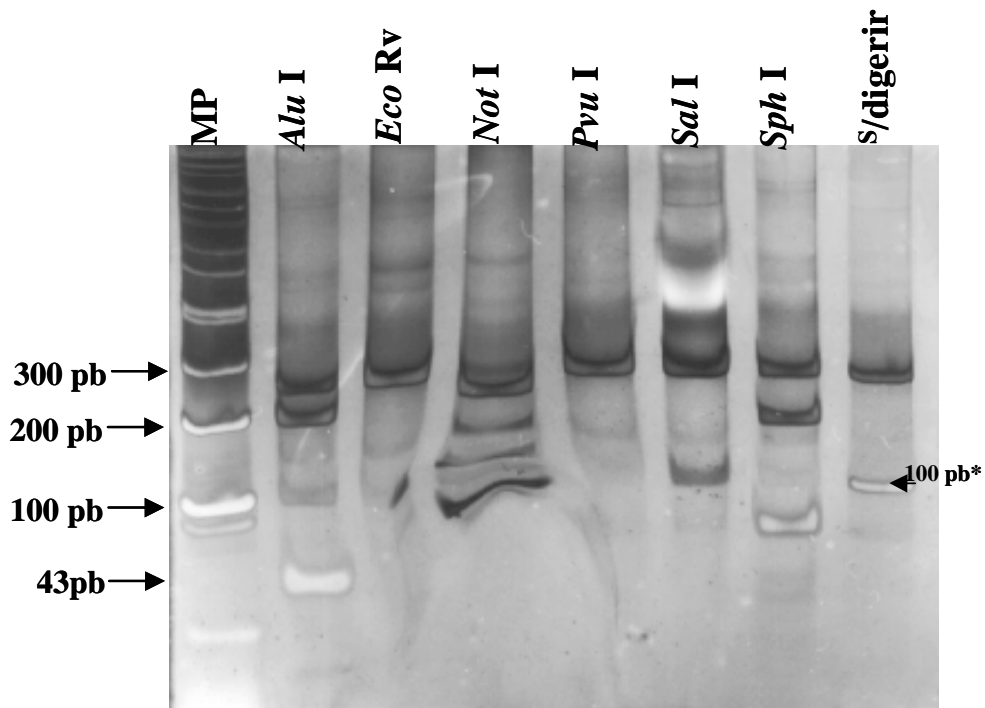


Figura 13. RFLP-PCR de *M. tuberculosis* H37Rv. polimorfismo de la longitud de los fragmentos de restricción del fragmento amplificado. Se emplearon las enzimas *Alu I*, *Eco Rv*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento de 235 pb producido por RT-PCR del cDNA de la cepa H37Rv. Gel de poliacrilamida al 12% teñido con tinción de plata. S/digerir (fragmento de 235 pb sin digerir) y MP (marcador de peso molecular en escalera de 100 pb). * Fragmento coamplificado de 100 pb.

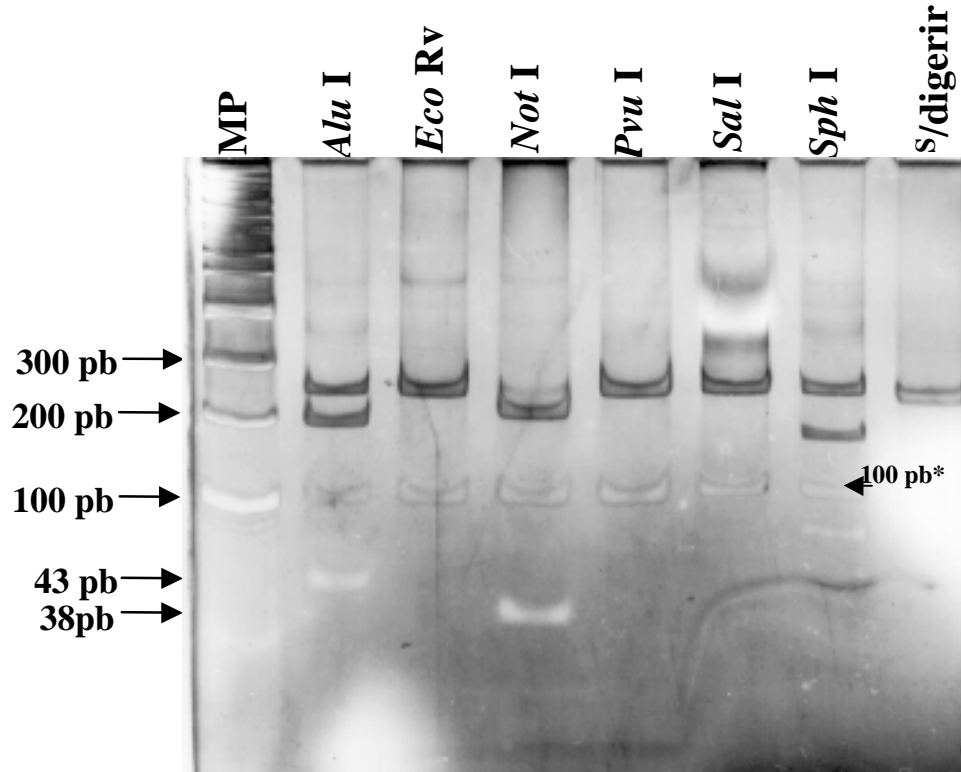


Figura 14. RFLP-PCR de *M. canetti* 046. polimorfismo de la longitud de los fragmentos de restricción del fragmento amplificado. Se emplearon las enzimas *Alu I*, *Eco Rv*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento de 235 pb producido por RT-PCR del cDNA de la cepa 046. Gel de poliacrilamida al 12% teñido con tinción de plata. S/digerir (fragmento de 235 pb sin digerir) y MP (marcador de peso molecular en escalera de 100 pb). * Fragmento coamplificado de 100 pb.

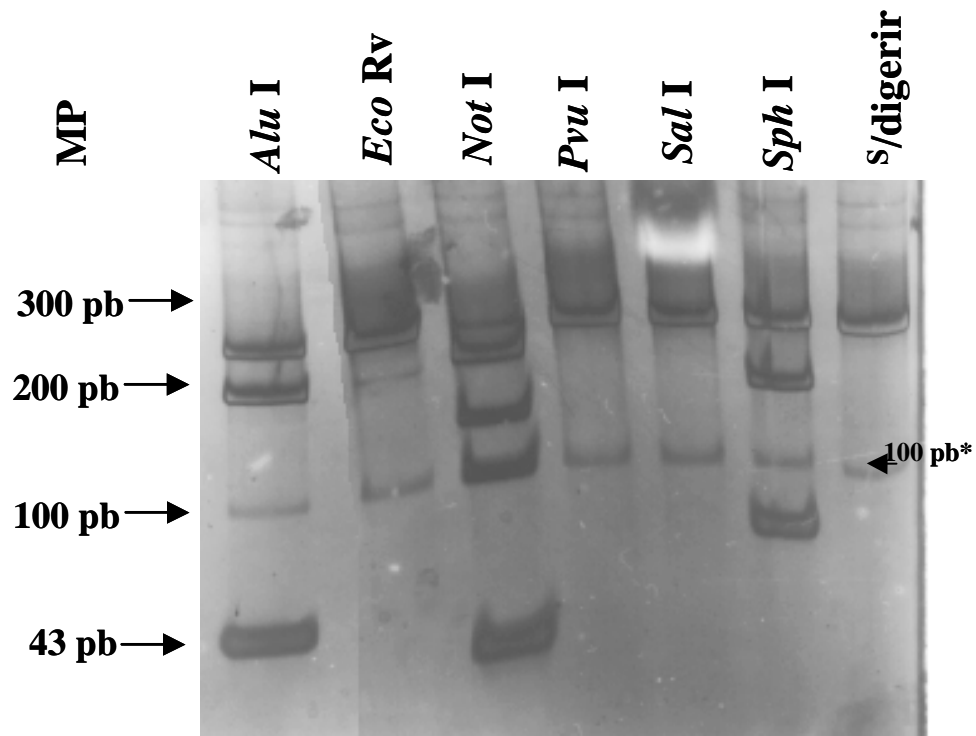


Figura 15. RFLP-PCR de *M. tuberculosis* 1000. Se emplearon las enzimas *Alu I*, *Eco Rv*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento de 235 pb producido por RT-PCR del cDNA de la cepa 1000. Gel de poliacrilamida al 12% teñido con tinción de plata. S/digerir (fragmento de 235 pb sin digerir) y MP (marcador de peso molecular en escalera de 100 pb). * Fragmento coamplificado de 100 pb.

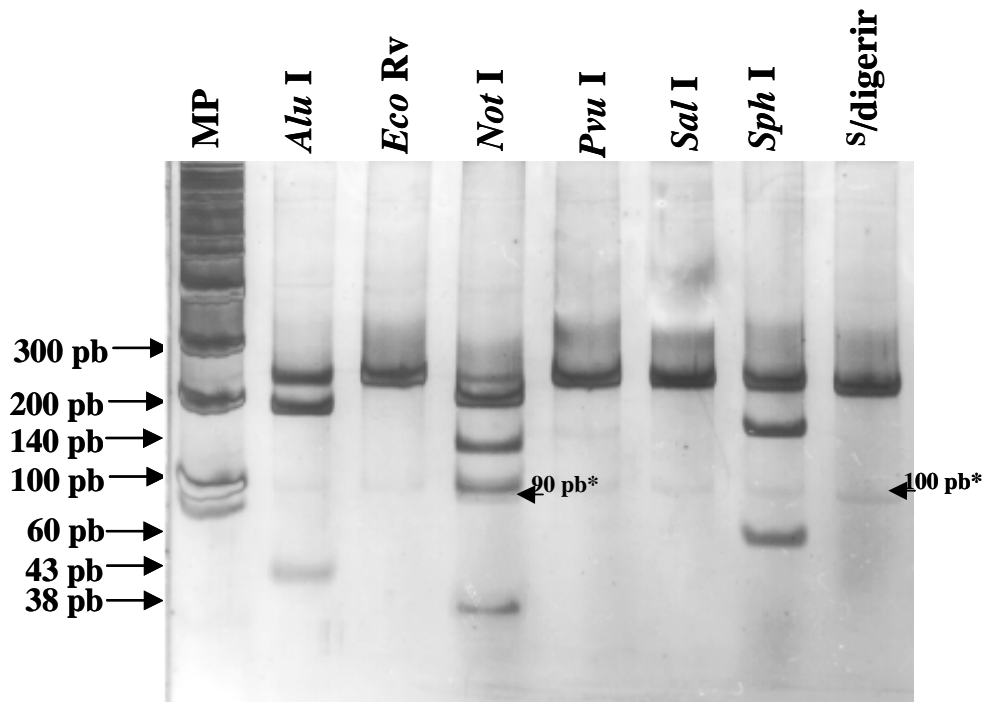


Figura 16. RFLP-PCR de *M. tuberculosis* 1463. Se emplearon las enzimas *Alu I*, *Eco Rv*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento de 235 pb producido por RT-PCR del cDNA de la cepa 1463. Gel de poliacrilamida al 12% teñido con tinción de plata. S/digerir (fragmento de 235 pb sin digerir) y MP (marcador de peso molecular en escalera de 100 pb).
* Fragmento coamplificado de 100 pb.

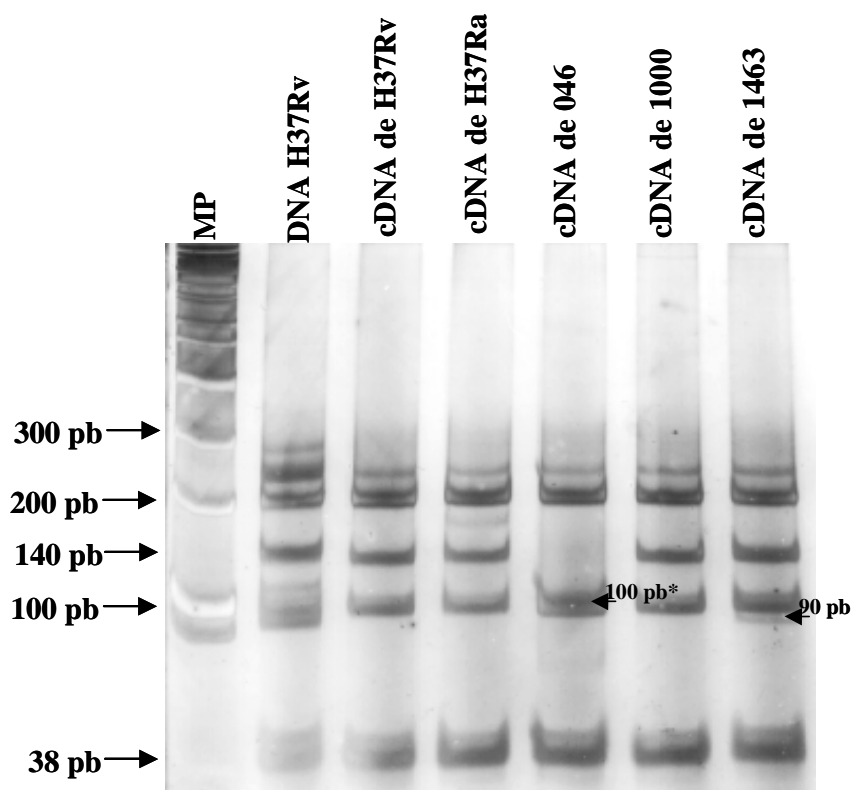


Figura 17. PCR-RFLP. Se empleó la enzima *Not* I para digerir los fragmento de 235 pb producidos por RT-PCR del cDNA de las cepas H37Rv, H37Ra, 046, 1000 y 1463; y del PCR del DNA de la cepa H37Rv. Gel de poliacrilamida al 12% teñido con tinción de plata. MP (marcador de peso molecular en escalera de 100 pb). * Fragmento coamplificado de 100 pb.

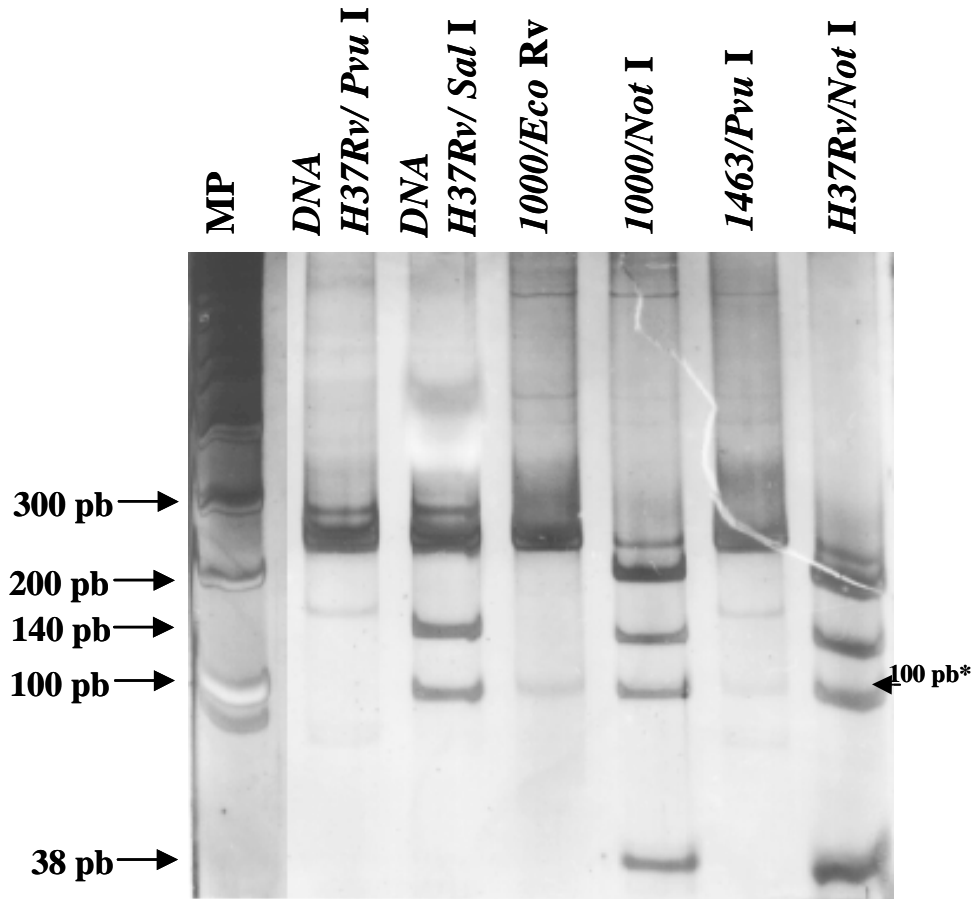


Figura 18. PCR-RFLP. Se emplearon las enzimas *Eco Rv*, *Not I* y *Pvu I*, para digerir los fragmento de 235 pb producidos por RT-PCR de las cepas H37Rv, 1000 y 1463; y del PCR del DNA de la cepa H37Rv digerido con *Pvu I* y *Sal I*. Gel de poliacrilamida al 12% teñido con tinción de plata. MP (marcador de peso molecular en escalera de 100 pb). * Fragmento coamplificado de 100 pb.

7.3 IDENTIFICACIÓN DE LOS GENES DE LA FAMILIA PE EXPRESADOS DIFERENCIALMENTE

Puesto que en la cepa 046 solo se observó el patrón de digestión característico del gen Rv1172c (anexo 5), se decidió utilizar una mezcla de enzimas *Alu I*, *Not I* y *Sph I* para digerir el fragmento de 235 pb (Tabla 2). Las bandas no digeridas después de la reacción y visualizadas con la tinción en bromuro de etidio se secuenciaron, así como la del fragmento obtenido sin el proceso de digestión. El resultado de las secuencias confirmó que en la cepa 046 bajo las condiciones utilizadas en este estudio únicamente se expresa el gen Rv1172c.

Tabla 2. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 046.

Reacción enzimática	gen identificado
1. - Sin enzimas	Rv1172c
2. - <i>Not I</i>	ND
3. - <i>Not I</i> , <i>Alu I</i> , <i>Sph I</i>	ND
4.- <i>Alu I</i> , <i>Sph I</i>	Rv1172c

*ND. No determinado porque el fragmento fue totalmente digerido.

El mismo procedimiento fue llevado a cabo con las demás cepas. Así, se pudo identificar el gen representado por los fragmentos producidos por *Not I* de 140 y 95 pb.

El análisis enzimático realizado en la cepa 1000 permitió la identificación de dos secuencias, que al compararlas con la base de datos correspondieron a la Rv1172c y la secuencia RV0578c (Tabla 3), esta última pertenece a un miembro de la subfamilia de la PE_PGRS.

Tabla 3. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 1000.

Reacción enzimática	gen identificado
1. - <i>Not</i> I	1172c
2. - <i>Alu</i> I, <i>Sph</i> I y <i>Eco</i> RV	Rv0578c
3. - <i>Alu</i> I y <i>Sph</i> I	Rv0578c

Con el propósito de verificar si los genes Rv1172c y Rv0578c estaban presentes en la cepa 1463, se llevó a cabo el mismo procedimiento utilizando la mezcla enzimática. Sin embargo, esto solo permitió identificar el gen Rv1172c (Tabla 4). La identificación del gen Rv0578c se logró por la secuenciación del fragmento de 140 pb producido por *Not* I.

Tabla 4. Análisis de restricción y secuencias del fragmento de 235 pb de la cepa 1463.

Reacción enzimática	gen identificado
1.- <i>Not</i> I	1172c
2.- <i>Alu</i> I, <i>Sph</i> I y <i>Eco</i> RV	ND
3.- <i>Not</i> I, <i>Alu</i> I y <i>Sph</i> I	ND

*ND. No determinado porque el fragmento fue totalmente digerido.

Por otro lado, no fue posible secuenciar los fragmentos de 160 y 74 pb producidos por *Pvu* I, quizás debido a la baja concentración de los productos amplificados que fueron difícilmente detectados con la tinción en bromuro de etidio.

Para identificar este gen, se realizó un análisis teórico de los sitios de restricción y de los fragmentos que podrían ser generados en los 99 genes de la

familia PE. En 6 de los 99 genes se identificó el sitio *Pvu* I que generaría los fragmentos de 160 y 74 pb (Rv0742, Rv1759, Rv1818, Rv1840, Rv2371, Rv3652). El mismo análisis mostró que los seis genes podrían ser discriminados al emplearse las enzimas *Kpn* I y *Bgl* II o bien la utilización de una mezcla de dos enzimas como *Pvu* I/*Not* I y *Pvu* I/*Stu* I (Tabla 5).

Tabla 5. Análisis de restricción teórico de los genes de la familia PE con *Pvu*

I

GEN	<i>Pvu</i> I	<i>Pvu</i> I/ <i>Not</i> I	<i>Pvu</i> I/ <i>Stu</i> I	<i>Kpn</i> I	<i>Bgl</i> II	RESULTADO
Rv0742	74	18, 74, 143	74, 87 , 54			No producido
Rv1759	74				137	No producido
Rv1818	74	15, 74, 146	74, 87 , 54			No producido
Rv1840	74		74, 87 , 54			No producido
Rv2371	74			46		No producido
Rv3652	74	15, 74, 146				Esperado

Los fragmentos marcados en **negrita** muestran los fragmentos no observados en un gel de poliacrilamida al 12% teñido con nitrato de plata.

En la práctica, usando el fragmento de 235 pb producido por el DNA de *M. tuberculosis* H37Rv y el uso de una o dos enzimas de restricción, se demostró el poder discriminativo de la técnica ya que fue posible determinar el gen expresado en la cepa 1463. Aunado a esto, principalmente mediante la ausencia de fragmentos, se pudo identificar a cada uno de los 6 genes candidatos. La combinación de *Pvu* I/*Stu* I se descartó a los genes Rv0742 Rv1818 y Rv1840 debido a la ausencia de las patrones esperados (Fig. 19 A); Así mismo con *Bgl* II se descartó a la Rv1759c y con *Kpn* I a la Rv2371 (datos no mostrados). Con *Pvu* I/*Not* I no fue posible diferenciar a los genes

debido a la baja resolución en el corrimiento electroforético en el gel de poliacrilamida (Fig. 19 B). Por lo tanto, el gen que podría estar presente en los fragmentos de 160 y 74 pb producidos por *Pvu I*, en la cepa 1463 sería el gen Rv3652,

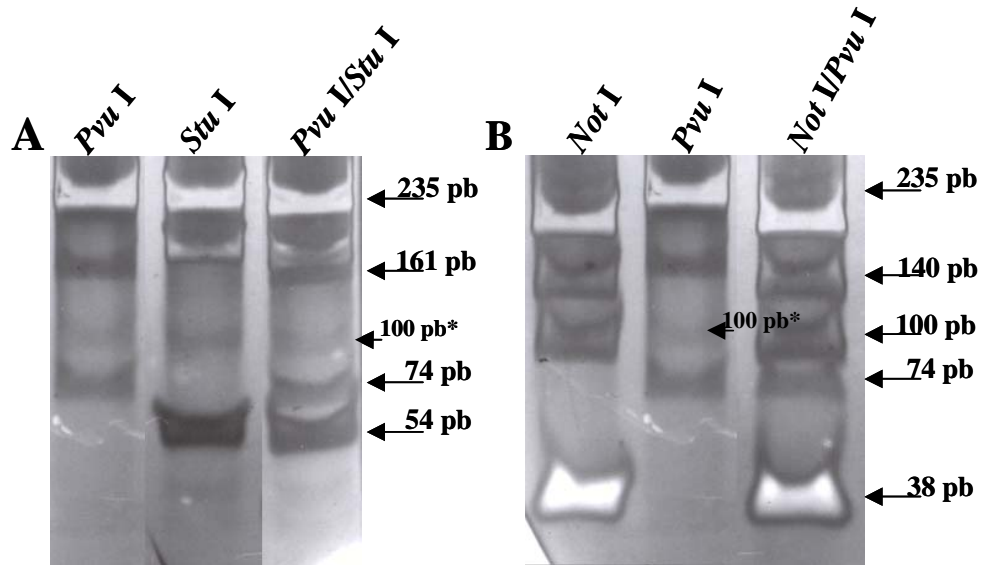


Figura 19. PCR-RFLP. Análisis experimental para la discriminación o identificación del gen que genera con *Pvu I* los fragmentos de 160 y 74 pb. Se usaron las enzimas *Pvu I*, *Stu I* y *Not I* para digerir el fragmento de 235 pb producidos por PCR del DNA de la cepa H37Rv. Gel de poliacrilamida al 12% teñido con tinción de plata. * Fragmento coamplificado de 100 pb.

sugerido también por la presencia del fragmento de 90 pb generado por la digestión de *Not I* en el DNA de la cepa 1463 (Fig. 16).

Para confirmar la presencia del gen Rv3652 expresado únicamente en la cepa 1463, se amplificó un fragmento específico de 170 pb, usando la región de 235 pb previamente amplificada. El producto de 170 pb fue producido en el cDNA de la cepa 1463 así como en el DNA de todas las cepas estudiadas (Fig. 20). Lo anterior fue confirmado a través de la digestión con *Pvu I* de los fragmentos amplificados de 235

pb de los DNAs. A excepción de la cepa *M. canetti* (cepa 046), todas presentaron los fragmentos de 160 y 74 pb. Sugiriendo que el gen Rv3652 es diferencialmente expresado entre las cepas y bajo las condiciones de este estudio.

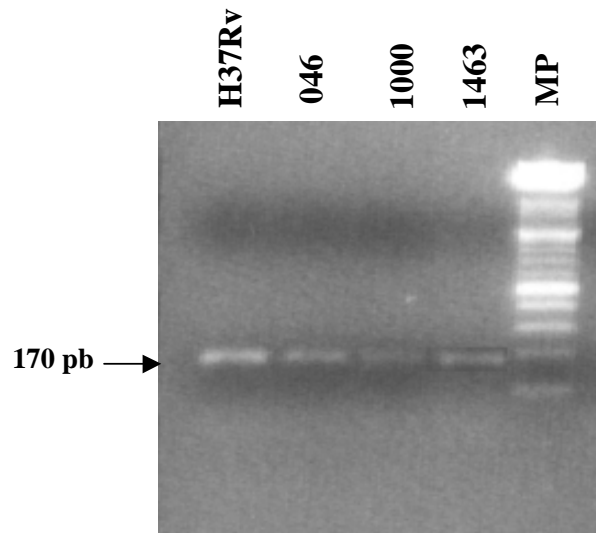


Figura 20. PCR. Amplificación del fragmento específico para el gen RV3652 de 170 pb en los DNAs de las cepas H37Rv, 046, 1000 y 1462. gel de agarosa al 2%, teñido con bromuro de etidio. MP (marcador de peso molecular en escalera de 100 pb).

7.4 IDENTIFICACIÓN DE LOS GENES DE LA FAMILIA PE NO EXPRESADOS

Se ha observado que en el amplificado de 235 pb del DNA de las cepas analizadas se generan los fragmentos de 141 y 94 pb producidos por la enzima *Sal* I (Fig. 21). Estos fragmentos no fueron visualizados en el cDNA de estas cepas.

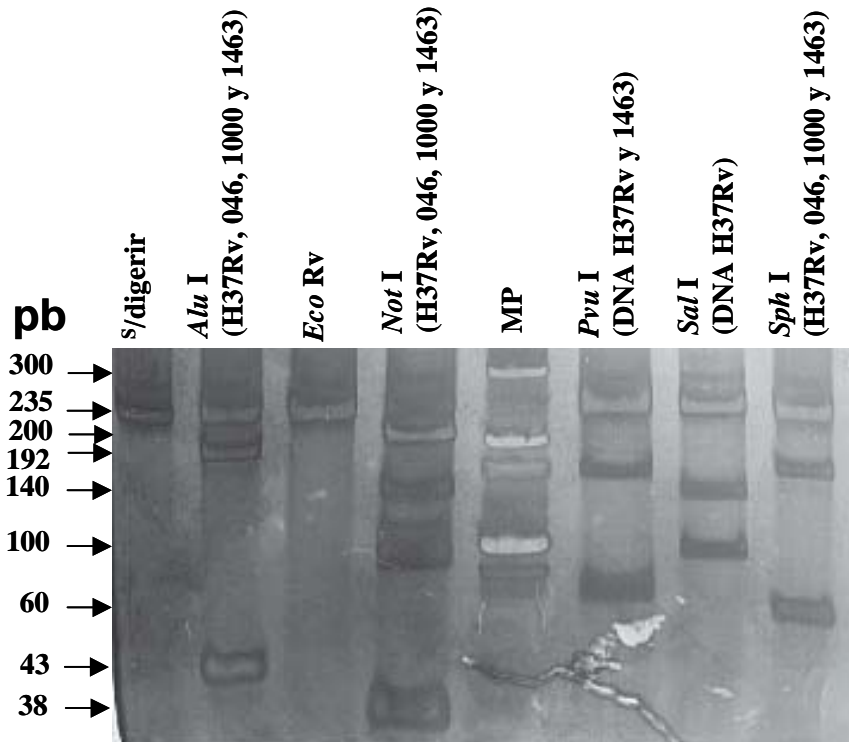


Figura 21. RFLP-PCR. polimorfismo de la longitud de los fragmentos de restricción del fragmento amplificado. Se emplearon las enzimas *Alu I*, *Eco Rv*, *Not I*, *Pvu I*, *Sal I* y *Sph I* para digerir el fragmento de 235 pb producido por PCR del DNA de las cepas analizadas. Gel de poliacrilamida al 12% teñido con tinción de plata. *S*/digerir (fragmento de 235 pb sin digerir) y MP (marcador de peso molecular en escalera de 100 pb).

Para identificar este gen se llevó a cabo el análisis teórico de restricción. Se pronosticaron 3 genes con el mismo patrón producido con *Sal I* (Rv0151, Rv0278 y Rv0747). En el mismo análisis también se observó que *Stu I* genera los fragmentos de 54 y 181 pb en los genes Rv0278 y Rv0747 y *Alu I* genera los fragmentos de 136, 27 y 77 en Rv0747. El gen Rv0151 no presentó sitios para estas dos enzimas (Tabla 6), se pensó que el uso de una enzima o en combinación podría identificar el o los genes no expresados en las cepas estudiadas.

Tabla 6. Análisis de restricción teórico de los genes de la familia PE con *Sal* I

Gen	<i>Sal</i> I	<i>Stu</i> I	<i>Alu</i> I	<i>Sal</i> I / <i>Stu</i> I	resultado
Rv0151	94/141				No producido
Rv0278	141/94	181		40, 141, 54	Producido
Rv0747	141/94	181	27, 77, 136	40, 141, 54	No producido

Los fragmentos marcados en **negrita** muestran los patrones no observados en un gel de poliacrilamida al 12% teñido con nitrato de plata.

En el análisis experimental se descartó que el gen Rv0747 se estuviera amplificando debido a que el DNA de las cepas estudiadas digerido con *Alu* I, solo produce los fragmentos de 140 y 95 pb, característicos del gen Rv1172c. Por la presencia de los fragmentos de 54 y 181 pb generados por *Stu* I, fue posible descartar al gen RV0151 ya que este carece de sitios de restricción dentro de la región amplificada de 235 pb. Por lo tanto los resultados sugieren que el gen Rv0278 es el que se amplifica en los DNAs pero no es expresado en las cepas estudiadas. El resultado es también apoyado por los patrones de restricción generados por *Sal* I / *Stu* I de 141, 54 y 40 pb. Tabla 6 y Figura 22.

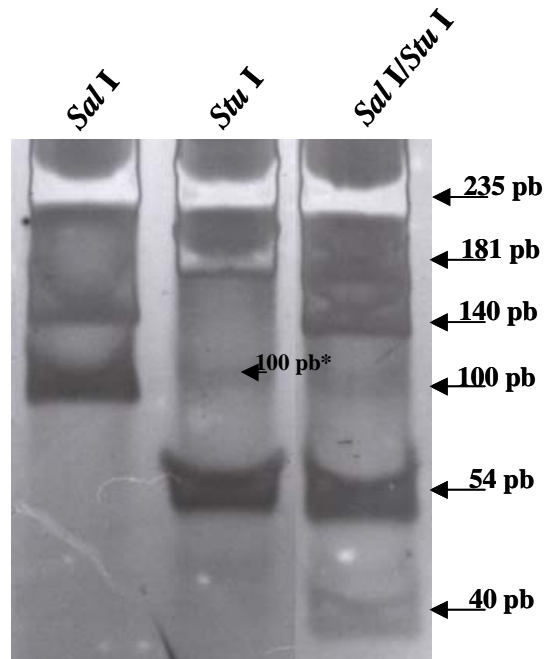


Figura 22. PCR-RFLP. Análisis experimental para la discriminación o identificación del gen que genera con *Sal* I los fragmentos de 141 y 94 pb. Se usaron las enzimas *Sal* I y *Stu* I, para digerir el fragmento de 235 pb producidos por PCR del DNA de la cepa H37Rv. Gel de poliacrilamida al 12% teñido con plata. * Fragmento coamplificado de 100 pb.

Interesantemente, después de haber generado los fragmentos de 40, 141, 54 pb con *Sal* I/*Stu* I correspondientes al gen Rv0278, aún se conservaba el fragmento de 181 pb, sugiriendo una digestión parcial o posiblemente la presencia de un gen adicional con sitios de corte para la enzima *Stu* I (Fig. 22).

Para determinar si dentro del fragmento de 181 pb existía un gen adicional al Rv0278c se realizó el análisis teórico con *Stu* I, encontrando que 14 de los 99 genes de la familia PE presentan el sitio de en la posición de 181 pb y otro más en la posición de 76 pb (Tabla 7). La mayoría de ellos fueron descartados con los análisis previos realizados con las enzimas *Pvu* I, *Not* I, *Alu* I y *Sal* I. Los genes Rv0278, Rv0742.

Rv1818, Rv1840, Rv3388 fueron discriminados con las enzimas *Ava I*, *Ava I / Stu I* y *Stu I / Pvu I* (Tabla 4 y 6).

Tabla 7. Análisis de restricción teórico de los genes de la familia PE con *Stu I*

Gen	<i>Stu I</i>	<i>Pvu I</i>	<i>Not I</i>	<i>Alu I</i>	<i>Sa I</i>	<i>Ava I</i>	<i>Stu I / Pvu I</i>	<i>Ava I / Stu I</i>	<i>Stu I / Not I</i>
Rv0124	<u>181</u>	<u>152</u>					<u>152, 29, 54</u>		
Rv0278	181				141				
Rv0279	181		89			61,174		120. 61, 54	89,92,54
Rv0297	181		<u>92, 125, 206</u>						92,89,33, 81
Rv0747	181			<u>27, 77</u>	141				
Rv0742	181	74	92				74,107,54		92,89,54
Rv0872	181		<u>206</u>						25.29,181
Rv1243	181		41	<u>82</u>					41,140, 54
Rv1325	181				<u>66</u>	<u>22,213</u>		22,159,54	
Rv1818	181	74	89		<u>66</u>		74,107,54		89,92,54
Rv1840	181	74					74,107,54		
Rv2340	181	<u>62</u>	95	<u>51</u>			62,119,54		95,86,54
Rv3388	181		92			22,213		22,159,54	92,89,54
Rv3650	181	<u>152</u>	89				<u>29, 152, 54</u>		89,92,54
Rv3653	<u>76</u>		104			<u>206,29</u>		29,76,130	28,76,131

Los fragmentos subrayados muestran los patrones no observados en un gel de poliacrilamida al 12% teñido con nitrato de plata.

El resultado de las digestiones nos hace sugerir la presencia del gen Rv0279 en el DNA de las cepas estudiadas, demostrado por la digestión con la enzima *Ava I* que produce los fragmentos 174 y 61 pb y de *Ava I / Stu I* al digerir el fragmento de 181 pb para generar los fragmentos de 120 y 61 pb (Fig. 23).

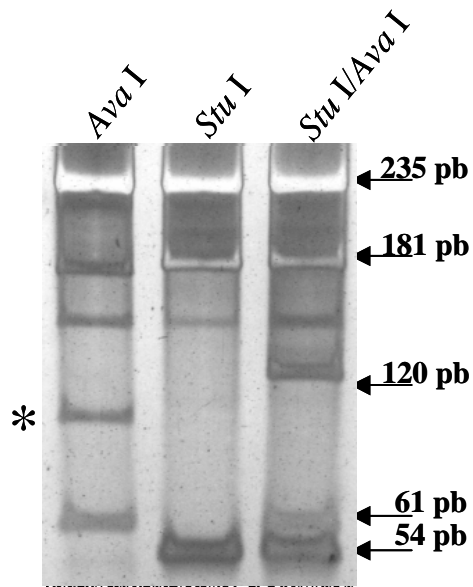


Figura 23. PCR-RFLP. Análisis experimental para la identificación del gen adicional, generado con *Stu I*. Se usaron las enzimas *Stu I* y *Ava I* para digerir el fragmento de 235 pb producidos por PCR del DNA de la cepa H37Rv. Gel de poliacrilamida al 12% teñido con nitrato de plata. * artefacto en la amplificación.

8. DISCUSIÓN

En el presente trabajo se estudió la expresión diferencial de los genes que codifican para la familia PE/PE_PGRS en aislados clínicos de *M. tuberculosis*. Para este estudio se utilizó la cepa de referencia de *M. tuberculosis* H37Rv, que constituyó una herramienta importante puesto que el genoma de esta cepa se encuentra completamente secuenciado. Además, se incluyeron la cepa 046 que representa a una variante lisa de *M. canetti*, esta es considerada como el ancestro común de las cepas modernas de *M. tuberculosis* (Brosch *et al.*, 2002) y dos cepas de *M. tuberculosis* de origen humano (1000 y 1463). Todas ellas presentaban un patrón polimórfico diferente, determinado por RFLP y usando la secuencia de inserción 6110 (IS6110) o la región PGRS como marcadores genéticos.

La cepa 1000, pertenece al genotipo Beijing que predomina en el continente asiático y ha adquirido una gran importancia debido su alta transmisibilidad hacia otros continentes. Se ha demostrado la presencia de esta cepa asociada con brotes epidémicos y de fármaco-resistencia (Kremer, 2005).

Interesantemente, en un estudio reciente se demostró que en el modelo de infección murino, existían diferencias en el grado de patogenicidad entre las cepas utilizadas en este estudio. Los ratones infectados con la cepa 1000 mostraron un mayor índice de mortalidad, comparado a los infectados con *M. canetti* 046, mientras que la cepa H37Rv mostró un índice de mortalidad intermedia. Estos resultados sugieren que en la patogenicidad de las diferentes cepas de micobacterias se encuentran involucrados genes diferencialmente expresados, que podrían estar sujetos a diferente regulación, y por lo tanto presentar variaciones en su expresión entre las cepas (López *et al.*, 2003).

En el presente trabajo se demostró la expresión diferencial de algunos miembros de la familia PE en aislados clínicos de *M. tuberculosis* durante su crecimiento *in vitro*. Los genes expresados fueron amplificados simultáneamente a

partir de un par de oligonucleótidos diseñados cuidadosamente a fin de que pudieran producir el mayor número de secuencias posibles. La secuencia consenso de los oligonucleótidos se determinó a través de la máxima frecuencia de cada nucleótido.

Con esta herramienta, pudo confirmarse que al resolver los amplificadores por el método del PCR-SSCP, existían al menos 20 bandas en el fragmento de 235 pb amplificado del DNA de *M. tuberculosis* H37Rv, lo cual sugería la presencia de al menos 10 genes diferentes. En el cDNA de todas las cepas se visualizaron 6 bandas, cuatro de ellas correspondieron al gen PE Rv1172c, una indicación de que este es expresado por todas las cepas. La presencia de 4 bandas con la misma secuencia podría explicarse por el alto contenido de GC de las secuencias amplificadas, que da origen a diferentes estructuras secundarias palindrómicas. Estas estructuras de tipo tallo-lazo se han descrito comúnmente en secuencias ricas en GC (Viswanathan *et al.*, 2000; Fujita *et al.*, 1995; Forsdyke, 1995). Debido a las limitaciones de la técnica, así como su baja reproducibilidad, fue necesario emprender otro abordaje, aunque consideramos que el PCR-SSCP fue un paso inicial muy importante en nuestro estudio ya que nos permitió identificar al gen Rv1172c y a partir del cual pudieron identificarse los demás genes.

El análisis teórico de los sitios de restricción presentes en la región de 235 pb de los 99 genes de la familia PE, permitió predecir los genes presentes en el fragmento amplificado de 235 pb. La digestión del fragmento de 235 pb con las enzimas seleccionadas y el posterior análisis de los patrones de digestión generados por PCR-RFLP permitió corroborar, la expresión del gen PE Rv1172c en todas las cepas, así como identificar a los demás genes expresados.

El gen Rv1172c es un miembro de la subfamilia PE que codifica para una proteína de 31 kDa. Los miembros esta subfamilia son altamente conservados. En el genoma de *M. tuberculosis* existen 37 genes de la subfamilia PE y 67 asociados a una PGRS (Banu *et al.*, 2002). Un análisis reciente comparativo de la Rv1818c en diferentes aislados clínicos de *M. tuberculosis* mostró que a pesar de la conservación de esta región,

también en ella se presentan polimorfismos de un solo nucleótido (Talarico *et al.*, 2005). Es importante resaltar que mientras las regiones PGRS parecen estar ausentes en el genoma de *M. leprae* y *M. avium* (Cole *et al.*, 2001), los genes PE se encuentran presentes en estas especies aunque en menor cantidad. En un estudio reciente usando microarreglos se estableció que la expresión del gen Rv1172 se reprime cuando la bacteria es sometida a diferentes estímulos como Diethylenetriamine/Nitric Oxide Adduct (DETA/NO o DNO); persistencia no replicativa (NRP). Cianuro de Potasio (KCN) y Carbonyl-cyanide 3-Chlorophenylhydrazone (CCCP), los cuales exponen a la bacteria a un estrés oxidativo, óxido nítrico y bajos niveles de oxígeno (Voskuil *et al.*, 2004).

En contraste con el gen Rv1172 cuya expresión se confirmó en todas las cepas estudiadas, encontramos que el gen PE_PGRS Rv0578c fue expresado por las cepas H37Rv, 1000 y 1463 pero no por la cepa *M. canetti* 046. No obstante, el análisis de restricción del fragmento de 235 pb del DNA de la cepa 046 y la secuencia correspondiente del fragmento de 140 pb producido por *Not* I mostraron que el gen está presente en el genoma de esta cepa. Estas observaciones sugieren que el gen Rv0578c se está regulando de manera diferente en *M. canetti*, o bien, que el gen presenta cambios en su secuencia, mas allá de la región amplificada afectando su transcripción. Será necesario secuenciar el gen completo en *M. canetti* 046 a fin de verificar su integridad ya que las principales diferencias entre los genes PE_PGRS ortólogos radica principalmente en las variaciones en la región PGRS debido a pérdida o ganancia de DNA como se ha observado por genómica comparativa ([Gey van Pittius et al., 2006](#)). Uno de los ejemplos más demostrativos es el del gen PE_PGRS MT1866 de *M. tuberculosis* CDC1551 que tiene una delección 30 codones y una inserción de 3 codones dentro del dominio PGRS, comparado al correspondiente gen Rv1818c encontrado en *M. tuberculosis* H37Rv. Poulet y Cole en 1995 propusieron que durante la replicación de estos genes el mecanismo de deslizamiento de la cadena ("strand

slippage") puede ocurrir frecuentemente en el dominio PGRS rico en regiones repetidas de GCs las cuales codifican a glicina-alanina.

El gen PE_PGRS Rv0578c codifica una proteína de 105.9 kDa y se ha visto por análisis de microarreglos que se induce en condiciones de hipoxia y en presencia de ácido palmítico (Voskuil *et al.*, 2004).

Por otro lado, los fragmentos producido por *Pvu* I de 74 y 164 pb estuvieron presentes en el DNA de todas las cepas y únicamente en el cDNA de la cepa de *M. tuberculosis* 1463. Desafortunadamente, no fue posible secuenciar estos fragmentos debido a la baja concentración del producto amplificado y aún después de ser reamplificado. Sin embargo, el análisis de restricción del fragmento de DNA de 235 pb con diferentes enzimas permitió descartar a 5 de los 9 genes que pudieron haberse generado con los mismos patrones producidos por *Pvu* I. Los datos apuntaban a que el gen expresado en la cepa 1463 era la PE_PGRS Rv3652. Esta observación fue confirmada por la secuencia del fragmento de 170 pb, amplificado con oligonucleótidos específicos en el fragmento de 235 pb del cDNA de la cepa *M. tuberculosis* 1463.

Interesantemente, encontramos que el marco de lectura abierto del gen PE_PGRS Rv3652 es mayor a los 159 pb, descrito en el genoma de *M. tuberculosis* H37Rv, esto se debe que hay un cambio en el marco de lectura que daría origen a una secuencia más larga que presenta un codón de paro a los 312 pb. Dado que nosotros amplificamos en la cepa 1463 un cDNA de 235 pb es de suponer que transcripción del gen Rv3652 incluye el marco de lectura rio abajo del gen Rv3652 reportado para H37Rv y estos cambios por lo tanto pueden ser una fuente adicional de variabilidad de la proteína. La proteína putativa reportada tiene un peso molecular de 4.8 kDa, en contraste con la proteína que se genera con el nuevo marco de lectura la cual tendría una masa molecular de 10.5 kDa. Se ha reportado por microarreglos que este gen se induce principalmente en presencia de Carbonyl-cyanide 3-Chlorophenylhydrazone (CCCP) (Voskuil *et al.*, 2004).

El análisis comparativo de los patrones de restricción entre el DNA y el cDNA nos permitió identificar también algunos genes que no son expresados. Se demostró que el gen PE_PGRS Rv0279c y el PE_PGRS Rv0278c estaban presentes en el DNA de los fragmentos *Sal* I de 141/94 pb y *Stu* I de 181/54 pb, respectivamente, pero ausentes en el cDNA de las cepas estudiadas. El uso de las enzimas *Sal* I / *Stu* I y *Ava* I / *Stu* I nos permitió confirmar que el gen PE_PGRS Rv0278c se encuentra presente en el DNA. Por lo tanto, interpretamos que en nuestras condiciones de crecimiento *in vitro* los genes PE_PGRS Rv0278c y Rv0279c no son expresados en las cepas estudiadas.

Estudios recientes con microarreglos muestran que el gen PE_PGRS Rv0279c, parece estar regulado por hierro. Adicionalmente al gen Rv0279c, otros genes PE y PE_PGRS parecen estar regulados por hierro como son los genes PE Rv1169c, Rv1195 y PE_PGRS Rv2741 que fueron inducido por hierro, mientras que el gene PE Rv0285 fue reprimido por hierro e IdeR (del inglés: Iron-dependent regulatory protein) (Rodríguez *et al.*, 2002).

Por RT-PCR usando oligonucleótidos específicos se ha demostrado la transcripción de 10 genes PE_PGRS en la cepa de *M. tuberculosis* H37Rv. Sin embargo, hasta el momento no se han revelado las proteínas codificadas por análisis de proteoma o por la caracterización convencional de proteínas (Banu *et al.*, 2002).

Usando microarreglos se ha podido determinar que la expresión de los genes PE/PE_PGRS es regulada por diversos estímulos como hipoxia, choque térmico, cambios de pH etc. Estas observaciones han llevado a plantearse varias hipótesis sobre la posible función de estas moléculas en la interfase hospedero-bacteria. El hecho de que se encuentren dispersas por el genoma y que respondan a tan variados estímulos, sugiere que estas moléculas podrían estar reguladas en situaciones de estímulos ambientales severos para el microorganismo en el caso de ser proteínas estructurales, como una forma de conservar la integridad de la pared celular (Brennan *et al.*, 2001).

Interesantemente, también se les ha atribuido a estas proteínas la capacidad de permitir la interacción bacteria- bacteria (Delogu *et al.*, 2004) o huésped-bacteria (Brennan *et al.*, 2001). Se propone que el polipéptido PE pudiera anclar a la proteína PGRS en la pared celular permitiendo que el dominio variable PGRS acceda al ambiente extracelular. Es también posible que junto con muchas proteínas PE-PGRS, las proteínas PE formen un complejo de proteínas en la superficie o dentro de la micobacteria similar a las observadas en la fimbria. De hecho, las investigaciones usando fusiones de la proteína con GFP sugieren que las proteínas PE pueden estar localizadas en regiones polares definidas en la pared celular. Esta observación es compatible con la formación de complejos de estructura de superficie. Otra posibilidad es que los polipéptidos PE funcionan similarmente a las secuencias PEST encontradas en algunas proteínas eucarióticas y en la proteína listeriolisina O formadora de poro del patógeno intracelular *Listeria monocytogenes*. En un análisis estructural se ha inferido que algunos complejos PE-PPE son formados pudiendo estar involucrados en transducción de señales adheridas a la membrana o como proteínas solubles (Strong *et al.*, 2006).

Aunado a esto, la genómica estructural ha permitido identificar un número de genes importantes que son regulados *in vivo* en situaciones de estrés que favorecen la sobrevivencia de la micobacteria. Se ha visto por ejemplo que en la mutante en el factor sigma D (SigD), se favorece la expresión de un grupo de genes PE_PGRS los cuales aumentan su expresión (Raman *et al.*, 2004; Wu *et al.*, 1997) estable durante la fase logarítmica y estacionaria pero declina rápidamente cuando el oxígeno decae. Estos resultados indican que al menos las proteínas PE_PGRS reguladas por Sig D podrían estar jugando un papel importante durante la latencia (Raman *et al.*, 2004).

En este trabajo demostramos la expresión 3 genes PE/PE-PGRS. Aunque el número de secuencias detectadas en nuestros experimentos fue bajo comparado con otros estudios en condiciones *in vitro* (Banu *et al.*, 2002), hemos podido detectar un

perfil de expresión de secuencias PE y PE_PGRS en las cepas estudiadas bajo las condiciones de crecimiento de la bacteria *in vitro*.

9. CONCLUSIONES

1. Bajo las condiciones empleadas en el presente trabajo se pudo establecer que algunas secuencias PE/PE_PGRS son diferencialmente expresadas en los aislados clínicos de *M. tuberculosis* durante su crecimiento *in vitro*.
2. El DNA que codifica para las secuencias PE/PE_PGRS identificadas en este estudio pudo ser detectado en todas las cepas, sin embargo no todas las secuencias fueron expresadas.
3. El uso de la técnica combinada PCR-RFLP resulto ser de mucha utilidad ya que permitió la identificación y diferenciación de los genes PE/PE_PGRS.
4. El par de oligonucleótidos utilizados en este estudio fueron capaces de amplificar al menos 5 miembros de la familia génica PE. El gen Rv1172c que pertenece a la subfamilia PE se encuentra expresado en todas las cepas; los genes PE_PGRS Rv0578c y Rv3652 muestran estar expresados diferencialmente; En la cepa 046 no fue expresado el gen Rv0578c, mientras que el gen Rv3652 únicamente fue expresado en la cepa 1463. En los DNAs de las cepas estos genes estuvieron presentes al amplificar el fragmento esperado de 235 pb.
5. Los genes PE_PGRS Rv0578c y Rv0579c fueron amplificados en los DNAs de las cepas estudiadas. Sin embargo, no fue posible amplificarlos en el cDNA. Por lo tanto, es posible que no sean expresados en nuestras condiciones experimentales.

10. BIBLIOGRAFÍA

Abou-Zeid C, Garbe T, Lathigra R, Wiker HG, Harboe M, Rook GA, Young DB. Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. Infect Immun. 1991 Aug; 59(8):2712-8.

Banu S, Honore N, Saint-Joanis B, Philpott D, Prevost MC, Cole ST. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens?. Mol Microbiol. 2002 Apr; 44(1):9-19.

Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science. 1999; 284(5419):1520-3.

Betts JC, Lukey PT, Robb LC, McAdam RA, Duncan K. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. Mol Microbiol. 2002 Feb; 43(3):717-31.

Bloom BR, 1994. Tuberculosis: Pathogenesis, Protection, and Control. American Society for Microbiology Press, USA

Brennan MJ, Delogu G, Chen Y, Bardarov S, Kriakov J, Alavi M, Jacobs WR Jr. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect Immun. 2001 Dec; 69(12):7326-33.

Brennan MJ, Delogu G. The PE multigene family: a 'molecular mantra' for mycobacteria. Trends Microbiol. 2002 May; 10(5):246-9.

Brennan MJ, Nico C, Van Pittius G, Espitia C. The PE and PPE multigene families of the mycobacteria. PAG 513-525. ED COLE ST *et al.* ASM Press tuberculosis and the tubercle bacillus Washintong D.C. I SBN 1-55581-295-3.

Brosch R, Gordont S. V, Marmiesse M., Brodin P, Buchrieser C, Eiglmeier K, Garnier T, Gutierrez C, Hewinson G, Kremer K, Parsons LM, Pym AS, Samper S, van Soolingen D and. Cole ST. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. PNAS; 2002 March; (99):3684-89.

Cauten GM, Pio A, Tem Dam HG. Annual Risk of Tuberculosis Infection. Bull World Health Organ. 2002; 80(6):503-11.

Cole ST, Barrell BG. In Genetics and Tuberculosis (eds. chadwick DJ. and Cardew G. Novartis Foundation Symp. 217). 1998; 160-72 (Wiley, Chichester, 1998).

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. 1998; 393: 537-44.

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R,

Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG. Massive gene decay in the leprosy bacillus. *Nature*. 2001 Feb; 409(6823):1007-11.

Davey S. New global plan to health TB. *Bull World Health Organ*. 2001; 79(12):1172-3.

Delogu G, Brennan MJ. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. *Infect Immun*. 2001 Sep; 69(9):5606-11.

Delogu G, Pusceddu C, Bua A, Fadda G, Brennan M J, Zanetti S. Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. *Mol Microbiol*. 2004; 52:725-733.

Dheenadhayalan V, Delogu G, Sanguinetti M, Fadda G, Brennan M. Variable expression patterns of *Mycobacterium tuberculosis* PE_PGRS genes: Evidence that PE_PGRS16 and PE_PGRS26 are inversely regulated in vivo. *J Bacteriol*. 2006 May; 188(10): 3721-5.

Espitia C, Lacleste JP, Mondragon-Palomino M, Amador A, Campuzano J, Martens A, Singh M, Singh M, Cicero R, Zhang Y, Moreno C. The PE-PGRS-glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins?. *J Microbiol*. 1999; 145:3487-95.

Fisher MA, Plikaytis BB, Shinnick TM. Microarray analysis of the *Mycobacterium tuberculosis* transcriptional response to the acidic conditions found in phagosomes. *J Bacteriol*. 2002 Jul; 184(14):4025-32.

Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D, Hickey E, Kolonay JF, Nelson WC, Umayam LA, Ermolaeva M, Salzberg SL, Delcher A, Utterback T, Weidman J, Khouri H, Gill J, Mikula A, Bishai W,

Jacobs Jr WR Jr, Venter JC, Fraser CM. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol.* 2002 Oct; 184(19):5479-90.

Flisser A, Velasco-Villa A, Martínez-Campos C, González-Domínguez F, Briseño-García B, García-Suárez R, Caballero-Servín A, Hernández-Monroy I, García-Lozano H, Gutiérrez-Cogco L, Rodríguez-Angeles G, López-Martínez I, Galindo-Virgen S, Vázquez-Campuzano R, Balandrano-Campos S, Guzmán-Bracho C, Olivo-Díaz A, De la Rosa JL, Magos C, Escobar-Gutiérrez A, Correa D. Infectious Diseases in Mexico. A Survey from 1995-2000. *Arch Med Res.* 2002; 33(4):343-50.

Forsdyke DR. A stem-loop "Kissing" model for the initiation of recombination and the origin of intron. *Mol.Bio.Evol.* 1995; 12(5):949-58.

Frieden T. Can tuberculosis be controlled? *Int J Epidemiol.* 2002; 31:894-9.

Fujita K, Silver J. Single-strand conformational polymorphism. En: Dieffenbach CW, Dveksler GS. *PCR primer a laboratory manual.* CSHL Press. 1995; 143-55.

García ML, Valdespino JL. Tuberculosis. En: Valdespino et al, ed. *Enfermedades Tropicales en México.* México, D.F: Secretaría de Salud. 1994; 215-228.

Gazdik Michaela A, McDonough Kathleen A. Identification of Cyclic AMP-Regulated Genes in *Mycobacterium tuberculosis* Complex Bacteria under Low-Oxygen Conditions *J Bacteriol.* 2005 Apr; 187(8): 2681-92.

Gazzard B. Tuberculosis, HIV and the developing world. *Clin Med.* 2001; 1(1): 62-8.

[Gey van Pittius NC](#), [Sampson SL](#), [Lee H](#), [Kim Y](#), [van Helden PD](#), [Warren RM](#). Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. [BMC Evol Biol](#). 2006 Nov; 15:6-95.

Hayashi CY, Lewis RV. Evidence from flagelliform silk cDNA for the structural basis of elasticity and modular nature of spider silks. *J Mol Biol*. 1998 Feb 6; 275(5):773-84.

Kinger AK, Verma A, Tyagi JS. A method for the isolation of pure intact RNA from mycobacteria. *Biotechniques*. 1993; 14:424-5.

Kremer K. Genetic markers for *Mycobacterium tuberculosis*; caracterización and spread of the Beijing genotype. Tesis. 2005.

Li Lingling, Bannantine John P, Zhang Qing, Amonsin Alongkorn, May Barbara J, Alt David, Banerji Nilanjana, Kanjilal Sagarika, and Kapur Vivek. The complete genome sequence of *Mycobacterium avium* subspecies *paratuberculosis*. *Proc Natl Acad Sci U S A*. 2005 August 30; 102(35):12344-9.

López B, Aguilar D, Orozco H, Burger M, Espitia C, Ritacco V, Barrera L, Kremer K, Hernandez-Pando R, K. Huygen K, Van Soolingen D. A marked difference in pathogenesis and immune response induced by different *Mycobacterium tuberculosis* genotypes. *Clin Exp Immunol*. 2003; 133:30-37.

Mizrahi V, Andersen SJ. DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence?. *Mol Microbiol*. 1998; 29; 1331-9.

Orita M, Iwahana H, Kanazawa H, Hayashi K, Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci U S A.* 1989 Apr;86(8):2766-70.

Pérez-Padilla R. La tuberculosis en México, deuda añeja de salud pública. *Gac Med Mex.* 2001; 137(1): 93-4.

Poulet S, Cole ST. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch Microbiol.* 1995 Feb; 163(2):87-95.

Rajagopalan S. Tuberculosis and Aging: A Global Health Problem. *Aging and Infectious diseases. Clin Infect Dis.* 2001.;33(7):1034-9.

Ramakrishnan L, Federspiel N, Falkow S. Granuloma-Specific expression of mycobacterium virulence proteins from the glycine-rich PE-PGRS family. *Science.* 2000; 288:1436-9.

Raman Sahadevan, Rohan Hazra, Dascher Christopher C, Husson Robert N. Transcription Regulation by the *Mycobacterium tuberculosis* Alternative Sigma Factor SigD and Its Role in Virulence. *J Bacteriol.* 2004 Oct; 186 (19):6605-16.

Rodriguez GM, Voskuil MI, Gold B, Schoolnik GK, Smith I. IdeR, An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. *Infect Immun.* 2002 Jul; 70(7):3371-81.

Ross BC, Raios K, Jackson K, Dwyer B. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. *J. Clin: Microbiol.* 1992; 30:942-6.

Sachetto-Martins G, Franco LO, de Oliveira DE. Plant glycine-rich protein: a family or just proteins with a common motif. *Biochimica et Biophysica Acta*. 2000; 1492:1-14.

Saviola B, Woolwine SC, Bishai WR. Isolation of acid-inducible genes of *Mycobacterium tuberculosis* with the use of recombinase-based in vivo expression technology. *Infect Immun*. 2003 Mar; 71(3):1379-88.

Schlossberg D. 2000. Tuberculosis e infecciones por micobacterias no tuberculosas. *Mc Graw Hill*. Cuarta edición. USA.

Sepúlveda J, Bronfman M. AIDS in Mexico: characteristics and perspectives. *Salud Pública Mex*. 1988; 30(4):501-3.

Singh KK, Zhang X, Patibandla AS, Chien P Jr, Laal S. Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. *Infect Immun*. 2001 Jun; 69(6):4185-91.

Soberón-Acevedo G, Vilar-Puig P, Ramon de la Fuente J, Izazola-Licea JA, Mann JM, Cowal SG, Zacarias F, Liguori AL, Romero-Keith J, Tapia-Conyer R, Sepulveda J, Bronfman M, Rico B, Mohar A, Langer A, Ponce de Leon-Rosales S, Rangel-Frausto MS, Vazquez de-la-Serna A, Huertas M, Martinez-Abaroa C, Munoz-Hernandez O, Zarate-Aguilar A, Garduno-Espinosa J, Zuniga-Avila J, Mendoza-Zepeda R. AIDS in Mexico and the world: a global view. *Gac Med Mex*. 1996; 132(1):1-138.

Strong M, Sawaya MR, Wang, Phillips, Cascio D, Eisenberg D. Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *PNAS*. 2006 May; 103(21):8060-5.

Sudre P, Ten Dam G, Kochi A. Tuberculosis in the present time: A global overview of the tuberculosis situation. *Bull World Health Organ.* 1992; 70(2):149-59.

Talarico Sarah, Cave M. Donald, Marrs Carl F, Foxman Betsy, Zhang Lixin and Yang Zhenhua. Variation of the *Mycobacterium tuberculosis* PE_PGRS33 Gene among Clinical Isolates. *J. Clin. Microbiol.* 2005 Oct; 43(10):4954-60.

Valdespino JL, I zazola JA, Rico B. AIDS in Mexico: trends and projections. *Bull Pan Am Health Organ.* 1989, 23(1-2):20-3.

Viswanathan M, Lacirignola JJ, Hurley RL. Lovett ST. A novel Mutational hotspot in a natural Quasipalindrome in *Escherichia coli*. *J. Mol. Biol.* 2000; 302:553-64.

Voskuil MI, Schnappinger D, Rutherford R, Liub Y. Schoolnik G.K. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis.* 2004; 84:256-62.

World Health Organization (WHO). AIDS epidemic update, 2004 Dec. Available at: http://www.unaids.org/wad2004/EPI_1204_pdf_en/EpiUpdate04_en.pdf.

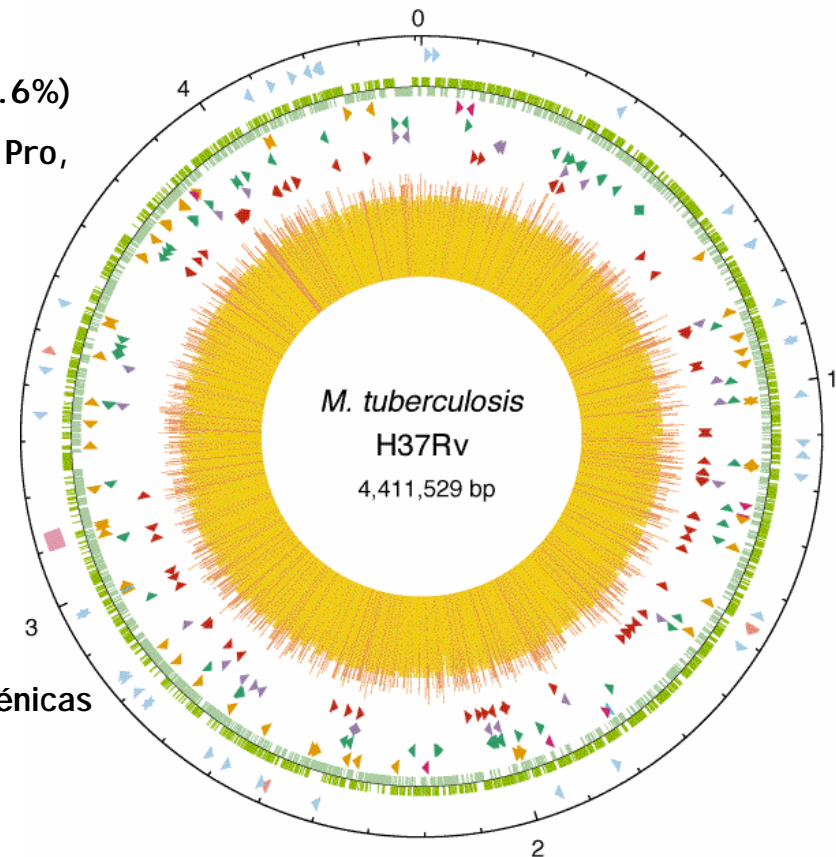
World Health Organization (WHO). REPORT 2003 Global Tuberculosis Control: SURVEILLANCE, PLANNING, FINANCING. *Bull World Health Organ.* 2003; <http://www.who.int/gtb>.

Wu QL, Kong D, Lam K, Husson RN. A mycobacterial extracytoplasmic function sigma factor involved in survival following stress. *J Bacteriol.* 1997 May;179(9):2922-9.

Ye ZH, Song YR, Marcus A, Varner JE. Comparative localization of three classes of cell wall proteins. *Plant J.* 1991 Sep; 1(2):175-83.

ANEXOS

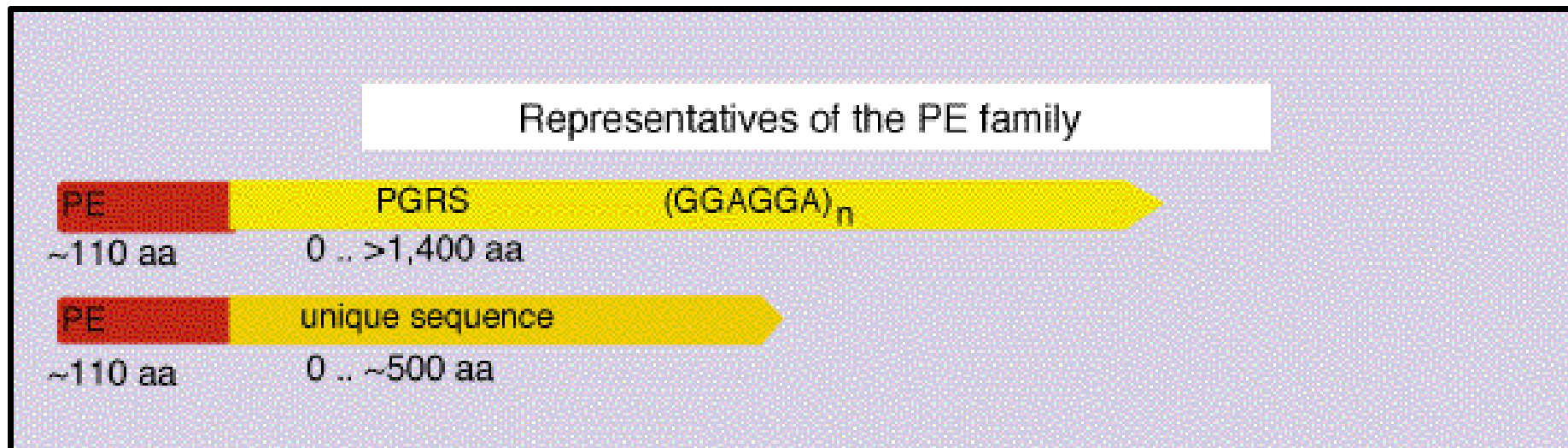
- 4.411 mega bases
- Presenta un alto contenido de GC (aprox 65.6%)
- Preferencia por aminoácidos como Ala, Gly, Pro, Arg y Trp.
- 3924 marcos de lectura abierto
- 91% con capacidad codificadora
 - 40% para funciones conocidas
 - 44% por predicción
 - 16% con funciones desconocidas
- El 30% del genoma codifica para enzimas involucradas en el metabolismo de los lípidos
- 9% del genoma comprende a dos familias génicas que codifican para las proteínas denominadas:
 - PPE
 - PE



ANEXO 1. Composición estructural del cromosoma de *M. tuberculosis*, H37RV

ANEXO 3

REPRESENTACIÓN ESQUEMÁTICA DE LA ESTRUCTURA DE LA SUBFAMILIA PE/PE_PGRS IDENTIFICADAS EN EL GENOMA DE *M. tuberculosis* H37Rv



ANEXO 4

ALINEAMIENTO DE NUCLEOTIDES DE 31 SECUENCIAS DE LOS GENES DE LA FAMILIA PE

PAGINA 1/10

	1	15	16	30	31	45	46	60	61	75	76	90
1	SEQ27	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	GCGGCAGCGGCAACG	GATTTGGCCAGTCTC	GGATCGAGCATCAGC					78
2	SEQ29	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	GCGGCAGCGGCAACG	GATTTGGCCAGTCTC	GGATCGAGCATCAGC					78
3	SEQ31	-----ATG-TC	GT-TTG-TGATTGCG	GCGCCAGAGGTTATC	GCGGCAGCGGCAACG	GATTTAGCCAGTCTC	GAGTCGAGCATCGCC					78
4	SEQ14	-----ATG-TC	AT-GGG-TGATGGTT	TCGCCGGAGCTGGTG	GTGGCGGCGGCAGCG	GATTTGGCGGGGATC	GGGTCGGCGATTAGC					78
5	SEQ28	-----ATG-TC	GT-TTG-TGATCGCG	GTGCCGGAGACTATC	GCGGCAGCGGCGACG	GATCTAGCCGATCTC	GGCTCGACGATCGCT					78
6	SEQ12	-----ATA-TC	GT-TTG-TGATCGCG	GCGCCGGAGGCCATC	GCGGCAGCGGCCACG	GATTTGGCAAGCATC	GGTTCGACGATCGGG					78
7	SEQ11	-----GTG-TC	CT-TTG-TGATCGCA	CAACCGGAGATGATC	GCGGCAGCGGCCGGT	GAGTTGGCCAGCATC	AGATCGGCGATCAAC					78
8	SEQ7	-----ATG-TC	AT-TTG-TGATCGCG	GCACCGGAGTTTTTA	ACGGCGGCAGCAATG	GACTTGGCGAGCATC	GGCTCGACAGTGAGC					78
9	SEQ13	-----ATG-TC	GT-TCG-TGCTGGCG	ATGCCGGAGGTGTTG	GGGTCGGCGGCAACG	GATCTGGCCGCTCTG	GGCTCGGTGCTGGGG					78
10	SEQ22	-----ATG-TC	GT-TTG-TAATCGCC	GCTCCCGAGACGCTG	GTGAGGGCCGCTTCG	GATTTGGCGAACATC	GGGTCGACGCTGGGG					78
11	SEQ23	-----ATG-TC	GA-ACG-TGATGGTA	GTCCCGGAAATGTTG	TCAGCGGCGGCACG	GATGTGGCGAGCATC	GGTGCAGCGCTTAGC					78
12	SEQ30	-----ATG-TC	GT-TTG-TCAGCGTA	GCCCCGGAGATTGTTG	GTGGCCGCGGCAACA	GACCTGGCGGGTATC	GGATCGGCGCATCAGC					78
13	SEQ10	-----ATG-TC	GT-TTG-TGATTGCT	GCGCCGGAGGCGTTG	GTGCGGTCGCTTTCG	GATCTGGCGGGCATT	GGGTCGGCGCTGGCG					78
14	SEQ8	-----ATG-TC	GT-TTG-TAGTCGTG	GCGCCGGAGGTGTTG	GCGGCAGCGGCTTTCG	GATCTAGCGGGCATC	GGGTCGACACTGGCG					78
15	SEQ18	-----ATG-TC	GT-TGG-TGATCGTG	GCCCCGGAGACGGTG	GCGGCCGCAGCCTTA	GATGTGGCGCGCATC	GGGTCATCGATCGGC					78
16	SEQ19	-----ATG-TC	GT-TGG-TGATCGTA	ACCCCGGAGACGGTG	GCCGCGGCGGCCCTCG	GATGTGGCGCGCATC	GGGTCATCGATCGGT					78
17	SEQ20	-----ATG-TC	CT-ACA-TGATTGCG	GTGCCAGACATGCTC	AGCTCGGCGGCTGGG	GATTTGGCCAGCATC	GGATCGTCGATCAAT					78
18	SEQ16	-----GTG-TC	GT-TTG-TGTTGGTG	TCTCCGTGCGAGCTG	ATGGCCGCGGCGGCC	GATGTGGCCGGTATC	GGTTCGGCAATCAGC					78
19	SEQ17	GAGGAAACCGTG-TC	GT-TTG-TGTTGGTG	TCTCCGTGCGAGCTG	ATGGCCGCGGCGGCC	GATGTGGCCGGTATC	GGTTCGGCAATCAGC					87
20	SEQ25	-----ATG-TC	GT-TTG-TGATTGCC	GTTCCGGAGTTCCTG	TCGGCGGCGGCCACG	GATTTGGCCAATCTC	GGTTCGACGATCAGC					78
21	SEQ6	-----ATG-TC	AT-ATG-TGATCGCG	GCGCCGGAGGCGCTG	GTGGCGGCGGCCACG	GATTTGGCTACTCTC	GGCTCGACGATCGGC					78
22	SEQ26	-----ATGGTC	ATGTCGCTGATGGTG	GCGCCGGAGCTGGTG	GCGGCAGCGGCGGCG	GACTTGACCGGGATT	GGGCAGGCCATCAGC					81
23	SEQ1	-----ATG-TC	GT-TCG-TTTTCGCG	GTGCCAGAGATGGTG	GCGGCAACCGCTTCC	GATTTGGCCAGCCTC	GGAGCGGCGCTGAGC					78
24	SEQ3	-----ATG-TC	GT-TTG-TGTTGGTT	TCGCCGGAGACCGTG	GCGGCGGTGGCCACG	GATCTCAAGCGCATC	GGCGCCTCGCTGGCC					78
25	SEQ4	-----ATG-TC	GT-TCG-TGTTGATC	GCACCGGAATTCGTG	ACAGCAGCCGCGGGG	GATCTGACGAATCTG	GGTTCGTCGATTAGC					78
26	SEQ5	-----ATG-TC	GT-TCG-TGTTGATC	GCACCGGAATTCGTG	ACAGCAGCCGCGGGG	GATCTGACGAATCTG	GGTTCGTCGATTAGC					78
27	SEQ21	-----ATG-TC	GT-TTG-TGTTGCGG	GCGCCAGAGGCACTG	GCGGCGGCGGCTGCG	GACATGGCCGGTATC	GGTTCGACTCTTAAC					78
28	SEQ2	-----GTG-TC	GT-TCG-TGGTCACA	GTGCCGGAGGCCGCTG	GCGGCTGCGGCGGGG	GATTTGGCGGCCATC	GGCTCGACGCTTCGG					78
29	SEQ9	-----ATG-TC	AT-ACA-TGATTGCC	ACACCAGCGGCGTTG	ACGGCGGCGGCAACG	GATATCGACGGGATT	GGCTCGGCGGTTAGC					78
30	SEQ15	-----ATG-TC	AT-TTG-TGATCGTG	GCGCGGGACGCGTTG	GCGGCGGCGGCGGCG	GATCTAGCGCAGATC	GGTTCGGCAGTGAAT					78
31	SEQ24	-----ATG-TC	AT-TCC-TGACGACA	CAGCCTGAAGAGTTG	GCGGCTGCGGCCGGC	AAGCTCGAGACCATC	GGCTCGGCGATGGTC					78

	91	105	106	120	121	135	136	150	151	165	166	180
1	SEQ27	GCGGCAAACGCAGCC	GCGGCGGCCAACACC	AC---GGCACTGATG	GCCGCGGGCGCCGAC	GAGGTGTCGACAGCC	ATCGCCGCGCTGTTC	165				
2	SEQ29	GCGGCAAACGCAGCC	GCGGCGGCCAACACC	AC---GGCACTGATG	GCCGCGGGCGCCGAC	GAGGTGTCGACAGCC	ATCGCCGCGCTGTTC	165				
3	SEQ31	GCGGCCAACGCGGCC	GCGGCGGCCAACACC	AC---AGCACTGCTG	GCCGCGGGTGCCGAT	GAAGTCTCGACGGCG	GTTGCGGCGCTGTTC	165				
4	SEQ14	TCGGCTAATGCGGCG	GCGGCCGTCAACACG	AC---GGGATTGTTG	ACCGCGGGTGCCGAT	GAGGTGTCGACAGCG	ATTGCGGCGTTGTTC	165				
5	SEQ28	GGGGCAAACGCGGCT	GCGGCGGCCAACACG	AC---GAGCCTGCTG	GCCGCCGGTGCCGAT	GAGATCTCGGCGGCA	ATCGCTGCGTTGTTC	165				
6	SEQ12	GCGGCCAACGCGCGG	GCCGCGGCCAACACG	AC---GGCGGTGCTG	GCCGCGGGCGCCGAT	CAGGTGTCGGTGGCC	ATCGCGGCGGCTTTT	165				
7	SEQ11	GCGGCCAATGCGGCG	GCCGCGGCCAGACC	AC---CGGAGTCATG	TCGGCGGCCGCCGAC	GAGGTGTCTACGGCG	GTTGCCGCGCTGTTT	165				
8	SEQ7	GCGGCCAGTGCCGCC	GCATCAGCCCCACG	GT---CGCGATCCTG	GCCGCGGGCGCCGAT	GAGGTGTCGATAGCC	GTCGCGGCGCTGTTC	165				
9	SEQ13	GCGGCCGATGCGGCC	GCGGCGGCTACGACG	AC---GGGCATCGTG	GCCGCGGCCAGGAT	GAGGTGTCGGCGGGCG	ATCGCGGCGTTGTTT	165				
10	SEQ22	GCCGCCAACGCGCGG	GCATTTGGGCCCCGACG	AC---GGAACCTTCTG	GCAGCCGGTGCTGAT	GAGGTGTCGGCGGGCC	ATCGCGTCGCTGTTC	165				
11	SEQ23	GCGGCCAACGGAGCC	GCGGCACCCACGACG	GC---TGGTGTCCCTG	GCCGCCGGGGCCGAT	GAGGTGTCGGCGGGCC	ATCGCGTCGTTGTTT	165				
12	SEQ30	GCGGCCAATGCCGCC	GCGGCTGCGCCGACC	AC---CGCCGTGCTG	GCCGCGGGTGCCGAT	GAGGTGTCGGCGGGCG	ATCGCGGCGCTGTTT	165				
13	SEQ10	GAGGCCAACGCGCGG	GCGTTGGCCCCGACG	AC---GGCGTTGTTG	GCCGCGGGTGCCGAT	GAGGTGTCGGCGGGCG	ATCGCGGCGCTGTTT	165				
14	SEQ8	CAGGCCAACGCGCGG	GCGTTGGCCCCGACC	AC---CGCGGTGTTG	GCCGCGGGTGCTGAT	GAGGTTTCCGCGGCA	ATCGCGTCGCTGTTT	165				
15	SEQ18	GCGGCCAATGCGGCG	GCGGCGGGTTCGACC	AC---CAGCGTGCTG	GCCGCGGGCGCCGAT	GAGGTGTCGGCGGGCG	ATCGCGACGCTGTTT	165				
16	SEQ19	GTAGCCAACAGCGCG	GCGGCGGGTCAACC	AC---CAGCGTGCTG	GCCGCGGGCGCCGAT	GAGGTGTCGGCGGGCG	ATCGCGACGCTGTTT	165				
17	SEQ20	GCCCTCAACGAGGGCG	GCCGCGGCCGCGACG	AC---GAGGTTGTTG	CCGGCCGCCGCCGAC	GAAGTGTGCGGCCAC	ATTGCGGCACTGTTT	165				
18	SEQ16	GCGGCCAACGCGGCC	GCCCTGGCCCCGACG	TC---GGTGCTGGCC	GCGGCCGGTGCCGAT	GAGGTGTCCGCGGGCG	GTCGCGGCACTGTTT	165				
19	SEQ17	GCGGCCAACGCGGCC	GCCCTGGCCCCGACG	TC---GGTGCTGGCC	GCGGCCGGTGCCGAT	GAGGTGTCCGCGGGCG	GTCGCGGCACTGTTT	174				
20	SEQ25	GCGGCTAACGCGGCT	GCCAGCATCCCGACC	AC---CGGCGTGCTG	GCCGCGGGCGCAGAT	GACGTATCGGCGGGCG	ATTGCGGCGCTCTTC	165				
21	SEQ6	GCCGCCAACGCGGCC	GCTGCGGGCTCGACA	AC---GGCGTTGCTG	ACCGCCGGCGCCGAC	GAAGTGTGCGGCGGGCG	ATAGCGGC-CTATTC	164				
22	SEQ26	GCGGCGAATGCGGCG	GCAGCGGGCCCCGACG	AC---GCAGGTGTTG	GCGGCCGCCGGTGAT	GAGGTGTCGGCGGGCG	ATCGCGGCGTTGTTT	168				
23	SEQ1	GAGGCCACCGCGGCG	GCGGCTATCCCCACC	AC---ACAAGTACTG	GCCGCGGCCCGCCGAT	GAGGTGTCGGCGGGCC	ATCGCGGAGTTGTTC	165				
24	SEQ3	CACGAAAACGCGTTCG	GCGGCCGCTTCGACG	AC---GGCGGTGGTC	TCCGCGGCCCGCCGAC	GAGGTATCGACGGCG	GTCGCCGCTCTGTTC	165				
25	SEQ4	GCGGCCAACGCGTTCG	GCAGCCAGTGCGACC	AC---GCAGGTGCTG	GCTGCGGGCGCCGAT	GAGGTGTCTGCCCGT	ATTGCGGCGCTGTTC	165				
26	SEQ5	GCGGCCAACGCGTTCG	GCAGCCAGTGCGACC	AC---GCAGGTGCTG	GCTGCGGGCGCCGAT	GAGGTGTCTGCCCGT	ATTGCGGCGCTGTTC	165				
27	SEQ21	GCCGCCAATGTGGTT	GCGGCGGTTCCCACC	AC---CGGAGTCCTG	GCCGCAGCCGCGGAC	GAGGTCTCGACTCAG	GTCGCCGCGCTGCTT	165				
28	SEQ2	GAAGCGACCGCTGCG	GCGGCGGGCCCCACG	AC---CGGGCTGGCG	GCCGCGGCCCGCCGAC	GACGTGTCGATCGCT	GTCCTCGCAGCTGTTC	165				
29	SEQ9	GTTCGGAACGCGCGG	GCGGTGCGCCGCGACA	AC---CGGAGTGCTG	GCCGCCGGTGCCGAT	GAAGTGTGGCGGGCC	ATCGCTAGGCTGTTC	165				
30	SEQ15	GCGGGCAATCTGGCC	GCAGCCAATCCGACG	AC---CGCTGTGGCG	GCGGCGGCCCGCCGAC	GAGGTATCGGCGGCA	CTCGCGGCGCTGTTC	165				
31	SEQ24	GCCCAGAACGCGGCC	GCGGCCGCCCCGACC	ACCACGGGCGTGATT	CCCGCCGCCGCTGAT	GAGATATCAGTGCTA	CAGGCGCGGTTGTTC	168				

	181	195	196	210	211	225	226	240	241	255	256	270
1	SEQ27	GGCGCCACGGCCAG	GCCTATCAGGCGCTC	AGCGCCCAAGCGC--	-AGGCGTTTCATGCC	CAGTTCGTGCAGGCA	TTGACCTCCGG-TGG	251				
2	SEQ29	GGCGCCACGGCCAG	GCCTATCAGGCGCTC	AGCGCCCAAGCGC--	-AGGCGTTTCATGCC	CAGTTCGTGCAGGCA	TTGACCTCCGG-TGG	251				
3	SEQ31	GGCGCCACGGCCAG	GCCTATCAGGCGCTC	AGCGCCCAAGCGC--	-AGGCGTTTCATGCC	CAGTTCGTGCAGGCG	TTGACCTCCGG-TGG	251				
4	SEQ14	GGTGCCCAAGGCCAG	GCCTACCAGGCGGCG	AGCGCACAGGCGG--	-CGGCGTTTTACGCC	CAGTTCGTGCAGGCC	CTGAGCGCCGG-CGG	251				
5	SEQ28	GGCGCGCACGGCCGG	GCCTATCAGGCGGCG	AGCGCCGAGGCGG--	-CGGCGTTTCATGGT	CGTTCGTGCAGGCG	CTGACCACCGG-GGG	251				
6	SEQ12	GGGGCGCACGGCCAG	GCCTATCAGGCGCTC	AGCGCGCAGGCGG--	-CGACGTTTCATATC	CAGTTTGTGCAGGCC	TTGACC GCGGG-CGC	251				
7	SEQ11	TCCTCGCATGCCCAG	GCCTATCAGGCCGCC	AGCGCGCAAGCGG--	-CCGCCTTTCACGCC	CAGGTGGTGCGGACC	CTGACC GTGGA-CGC	251				
8	SEQ7	GGAATGCATGGCCAG	GCATATCAGGCCCTC	AGCGTGCAGGCAT--	-CGGCGTTTCATCAG	CAATTTGTGCAGGCC	TTGACC GCGGG-CGC	251				
9	SEQ13	TCCGCCCACGGCCGG	GCCTATCAGGTGGCC	AGTGCGCAGGCGG--	-CGGCGTTTCACGCC	CAGTTCGTGGAGGCG	TTGAGCGCGGG-TGC	251				
10	SEQ22	GCCGCGCATGGCCAG	GCCTATCAGGCGGTG	AGCGCTCAGATGT--	-CGGCGTTTCATGCC	CAGTTCGTTCAAACC	TTTACTGCGGG-CGC	251				
11	SEQ23	TCTGGCTACGCCCGC	GACTATCAGGCGCTC	AGTGCCTCAAATGG--	-CCAGATTTTCATCAA	CAGTTCGTGCAGGCC	TTAACC GCGAG-CGT	251				
12	SEQ30	TCCGGCCACGCTCAG	GCCTATCAGGCGCTC	AGCGCCCAGGCGG--	-CGGCGTTTCATCAG	CAGTTCGTGCAGACG	CTTGCCGGTGG-CGC	251				
13	SEQ10	GGCGCGCACGGGCAG	GCGTATCAGACGGTT	AGCGCCCAGGCGT--	-CGGCGTTTCATGCC	CAGTTTGTGCAGGCG	TTGACTGGCGG-CGG	251				
14	SEQ8	GGGGCGCATGGTTCAG	GCGTATCAGGCGGTG	AGCGCCCAAATGT--	-CGGCGTTTCACGCC	CAGTTCATGCAGGCG	TTGACGGGTGC-CGG	251				
15	SEQ18	GGCAGCCATGCTCGG	GAGTATCAGGCGATC	AGCACGCAGGTGG--	-CGGCGTTTCATGAC	CGATTTGCGCAGACG	TTAAGCGCCG-GGT	251				
16	SEQ19	GGCAGCCATGCTCGG	GAGTATCAGGCGATC	AGCACGCAGGTGG--	-CAGCGTTTCATGAC	CGATTTGCGCAGACG	TTAAGCGCCG-GGT	251				
17	SEQ20	AGCGGACATGGCGAG	GGCTATCAGGCGATC	GCCC GG CAGATGG--	-CAGCGTTTCATGAC	CAGTTCACGCTGGCC	CTGACGTCAAG-TGC	251				
18	SEQ16	TCCGCCCACGCCGGC	CAGTATCAACAGCTG	GGCGCGCGTGCAG--	-CGCTGTTTCACGAG	CAGTTCGTGCAGGCA	TTGACC GGGGG-GGC	251				
19	SEQ17	TCCGCCCACGCCGGC	CAGTATCAACAGCTG	GGCGCGCGTGCAG--	-CGCTGTTTCACGAG	CAGTTCGTGCAGGCA	TTGACC GGGGG-GGC	260				
20	SEQ25	GGCGCGCACGCCAG	GCGTATCAGACGATC	AGCGCGCAGGCGG--	-CGACGTTTCATGCC	CAGTTCGTGCAGACC	TTGAGT GCGGG-TGC	251				
21	SEQ6	GGAATGCACGGCCAG	ACCTATCAGGCACCTC	AGTGC GCGGGCGG--	-CGGCGTTCCATGAG	CGTTCGTGCAGGCC	TTGGCCACAGG-TGG	250				
22	SEQ26	GGTACCCACGCGCAG	GAGTACCAGGCGTTG	AGCGCCCGGTGG--	-CGACGTTTCATGAG	CAGTTTGTGCGCTCG	CTGACC GCGGC-TGG	254				
23	SEQ1	GGTGCGCACGGCCAA	GAATTTCAAGCGCTC	AGCGCCCAGGCAT--	-CGGCGTTTCATGAC	CGTTCGTGCGGGCC	CTAAGCGCCG-AGC	251				
24	SEQ3	TCCCAACACGCCAG	GGCTACCAAGCGGCG	GCCGCTCAGGTAG--	-CAGCGTTTCATAGC	CGTTCGTGCAAGCC	CTGACGGCCGG-TGC	251				
25	SEQ4	GGCGGGTTTGGCCTG	GAGTACCAGGCGATT	AGTGC GCGAGGTGG--	-CGGCCTACCACCAG	CGTTCGTGCAGGCC	TTGAGT ACCGG-CGC	251				
26	SEQ5	GGCGGGTTTGGCCTG	GAGTACCAGGCGATT	AGTGC GCGAGGTGG--	-CGGCCTACCACCAG	CGTTCGTGCAGGCC	TTGAGT ACCGG-CGC	251				
27	SEQ21	TCCGCGCATGCTCAG	GGGTATCAGCAGCTC	AGCCGGCAGATGATG	ACAGCCTTCCACGAC	CAGTTCGTGCAGGCG	CTGAGAGCAAG-TGC	254				
28	SEQ2	GGCAGGTACGGCCAG	GAATTTCAAACCGTG	AGCAACCAACTGG--	-CCGCGTTTCATACC	GAGTTCGTACGCACG	TTGAAC GCGGG-CGC	251				
29	SEQ9	AACGCAAACGCCAG	GAATATCAGCCCTC	AGCGCGCAGGTGG--	-CGGCGTTTCAAACC	CTGTTTGTGCGCACC	TTGACTGGGGGGTGC	252				
30	SEQ15	GGCGCGCATGCCCAG	GAGTATCAGGCGGCG	GCGGCGCAGGCG--	-CGGCGTATCACGAG	CAGTTTGTGCACCGA	TTGAGCGGGC-AGC	251				
31	SEQ24	ACCGCGTACGGCACT	CTCTACCAGCAGGTC	AGTGC TGAAGCAG--	-CTGCGGTATACGAC	CTGTTTGTGAAGACC	CTGGGGGT CAG-CGC	254				

	271	285	286	300	301	315	316	330	331	345	346	360
1	SEQ27	CGGCGCCTATGCGGC	CGCCGAGGCCGCC--	-GCGGTGTCG-----	-----CCG-----	-----C	TGCTGGATCCGATCA	307				
2	SEQ29	CGGCGCCTATGCGGC	CGCCGAGGCCGCC--	-GCGGTGTCG-----	-----CCG-----	-----C	TGCTGGAT-----	300				
3	SEQ31	CGGCGCGTACGCCGC	CGCCGAGGCCGCC--	-GCCACCTCG-----	-----CCG-----	-----C	TGCTCGCCCCGATCA	307				
4	SEQ14	AGGCGCGTATGCGGC	CGCCGAGGCCGCC--	-GCCGTGTCG-----	-----CCG-----	-----C	TGCTGGCCCCGATCA	307				
5	SEQ28	GGGCGCCTATGCGGC	CGCCGAGGCCGCC--	-GCCGTGACG-----	-----CCG-----	-----C	TGCTCAACTCGATCA	307				
6	SEQ12	GGGCTCGTATGCGGC	CGCCGAGGCCGCCAG	CGCCGCGTCCATAAC	CAGT-CCG-----	-----C	TGCTCGACGCGATCA	319				
7	SEQ11	GGGAGCGTATGCCAG	CGCCGAGGCCGCCAA	CGCCGGGCCG-----	-----AAC-----	-----A	TGCTGGCCGCGGTCA	310				
8	SEQ7	GTACTCGTATGCCTC	CGCTGAAGCCGCC--	-GCCGTGACACCGCT	TCAG-CAA-----	-----C	TAGTCGATGTGATAA	316				
9	SEQ13	GGGGGCTACGCCAG	CGCGGAGGCCGCCGG	CGCGGCGGTGCTGGC	CAAC-CCGGCGCAGA	GCGTGCAGCAGGACC	TGCTGGCCGCCGTCA	340				
10	SEQ22	TGGCGCGTACGCGAG	CGCTGAGGCCGCC--	-GCCGCGGC-----	-----CCGTTGGA--	-----GGGCC	TGCTCAACATTGTGA	316				
11	SEQ23	GGGCTCGTATGCCGC	CGCCGAGGCCGCCAA	CGCTCCCCCTTG--	CAGG-CCCTCGAA--	-----CAACAGG	TGCTCGCGGCCATCA	328				
12	SEQ30	TGGAGCATATGCGGC	CGCCGAGGCCCG--	-GTCGAGCAG-----	-----CAG-----	-----C	TGCTGGCCGCGATCA	307				
13	SEQ10	CGGGGCGTATGCGGC	TGCCGAGGCCGCCAA	CGTCTCGGCGGCACA	GAGCACCGAC-----	-----CAGCGGC	TGCTCGATCTGATCA	328				
14	SEQ8	CGGGGCTTATGCGGC	TGCCGAGGCCGTCAA	CGTCTCGGCGGCACA	GA---GCGTGGAA--	-----CAAGACC	TGTTGGCCGCGATCA	328				
15	SEQ18	CGGCTCGTATGTCAG	CGCGGAGGCCGACCAA	CGCCGCACCGTTGGC	GACG-CTGGAG----	-----CACAAACG	TGCTCAATGCCCTCA	328				
16	SEQ19	CGGCTCGTATGTCAG	CGCGGAGGCCGACCAA	CGCCGCACCGTTGGC	GACG-CTGGAG----	-----CACAAACG	TGCTCAATGCCCTCA	328				
17	SEQ20	GGGCGCGTATGCCAG	CGCCGAGGCCGACCAA	CGTCAACAG-----	-----CAG-----	-----G	TGCTGGGTTTGATCA	310				
18	SEQ16	GAGCGCGTATGCCAG	CGCCGAGGCCGACCAA	TGTCGAGCAG-----	-----CAG-----	-----G	TGCTGGGTTTGATCA	310				
19	SEQ17	GAGCGCGTATGCCAG	CGCCGAGGCCGACCAA	TGTCGAGCAG-----	-----CAG-----	-----G	TGCTGGGTTTGATCA	319				
20	SEQ25	GGGGGCGTATGCCAA	CGCCGAGGCCGCCAA	CGTTCAGCAG-----	-----AGT-----	-----T	TGCTGAACGCGATCA	310				
21	SEQ6	GGGCGCCTATGCGGC	CGCCGAGGCCGCCAG	CGTCTCGCCGCTG--	-----CAG-----	-----AG-CG	CGCTCGATTTGCTGA	315				
22	SEQ26	CAGCGCGTATGCGAC	TGCCGAGGCCGCCAA	TGCATCACCGCTG--	-----CAGGCGCTG-	G-----AGCAGCAAG	TGTTGGGTGCGATCA	331				
23	SEQ1	GGGCTGGTATGTCGA	CGCCGAGGCCGCCAA	CGCCGCGCTGGTGGG	CACCGCGGCCAC---	-----CGGCGCG	TCGGAGTTGGGGTCA	330				
24	SEQ3	CGGGGCGTACGCATT	TGCCGAGGCCGCCAA	CGCGTCGCCGCT---	-A---CAGTCA-----	-----G	CCATGGGTGCGGT-A	315				
25	SEQ4	GGGCGCATATGCCTC	GGCCGAGGCCGCC--	-GCCGCTGAG-----	-----CAGAT-----	-----CG	TGCTGGGCGTGATCA	310				
26	SEQ5	GGGCGCATATGCCTC	GGCCGAGGCCGCC--	-GCCGCTGAG-----	-----CAGAT-----	-----CG	TGCTGGGCGTGATCA	310				
27	SEQ21	AGACGCGTATGCAAC	CGCCGAGGCCGAGC--	-GCCGCGCAGA-----	-----CCA-----	-----	TGGTGAACGCCGTGA	310				
28	SEQ2	GGCGGCGTATCTCAA	CACCGAAAGCGCTAA	CGGCGGGCAG-----	-----	-----C	TGTTCCGGTCAGATCG	307				
29	SEQ9	GGAGTCTTTCGCCGG	CGCCGAGGCCGCCAA	TGCGTACACAGCTG--	-----CAGAGCA---	-----TCGCG	CGGCAGGTGCGGG--	320				
30	SEQ15	GACATCGTATGCGGT	TACCGAGGTGACCAT	CGCGACGTCGCT---	-----CCG-----	-----GGGGG	CGCTGGGCTCG----	312				
31	SEQ24	TGGCACCTATGCGGC	CACCGAGGCCGCCAA	CTCATCCGCGGCCGC	TTCG-CCGTTGTCGG	GT-----ATCGCG	AGCATCTTGGGTTCC	336				

	361	375	376	390	391	405	406	420	421	435	436	450	
1	SEQ27	A-CGAGTTCTTCCTG	GCGAATACC-GG--G	CGCCCCGTTGATCGGC	AACGGCGCCAACGGC	GCCCCCGGC-A-CCG	GGG-CCAATG-GCGG						389
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	300
3	SEQ31	A-CGAGTTCTTCCTG	GCGAATACC-GG--G	CGCCCCGCTGATCGGC	AACGGCACCAACGGC	GCCCCCGGC-A-CCG	GGG-CCAATG-GCGG						389
4	SEQ14	A-CGCGCAATTCGTG	GCGGCCACC-GG--G	CGCCCCGCTGATCGGC	AACGGCGCCAACGGC	GCCCCCGGG-A-CCG	GAG-CCAACG-GCGG						389
5	SEQ28	A-CGCGCCCGTCCCTG	GCCGCTACC-GG--C	CGTCCGCTGATCGGT	AACGGGGCTAACGGT	GCTCCCGGC-A-CCG	GGG-CCAACG-GAGG						389
6	SEQ12	A-CGCGCCCTTCCTG	GCGGCGTTG-GG--G	CGCCCCGCTGATCGGT	AACGGCGCCGACGGG	GCGCCGGGG-A-CCG	GGG-CCGCCG-GCGG						401
7	SEQ11	A-CGCCCCCGCCAG	GCGCTGTTG-GG--G	CGCCCCACTGATCGGC	AACGGTGCCAACGGG	GCGCCGGGG-A-CCG	GGC-AGGCCG-GCGG						392
8	SEQ7	A-TGCGCCCTTCAGA	AGCGCGCTC-GG--C	CGCCCCCTGATCGGC	AACGGCGCCAACGGT	AAACCGGGG-A-CCG	GAC-AAGACG-GCGG						398
9	SEQ13	A-TGCGCAAAGTGTC	GCGCTCACG-GG--G	CGCCCCGTTGATCGGC	AACGGCGCCAACGGG	GCCCCGGGG-A-CGG	GGG-CCAATG-GTGC						422
10	SEQ22	A-CACGCCCACCCAG	CTGTTGCTG-GG--G	CGTCCGCTGATCGGC	AACGGCGCCAACGGC	GCACCGGGG-A-CCG	GGC-AAGCCG-GCGG						398
11	SEQ23	A-CGCGCCCACCCAG	ACACTGCTG-GG--G	CGTCCGCTGATCGGC	AACGGCGCCGACGG-	GCTGC-----CCG	GTC-AAAACG-GTGG						404
12	SEQ30	A-CGCGCCCACCCAG	GCGCTGCTG-GG--G	CGCCCCCTGATCGGC	AACGGTGCCGATGGG	GCGCCGGGG-A-CTG	GGC-AGGCCG-GCGG						389
13	SEQ10	A-TGGGCCACCCAG	GCGTTGTTG-GG--G	CGTCCACTGATCGGT	GATGGCGCCAACGGC	GGGCC--G-----G	GGC-AAGACG-GCGG						404
14	SEQ8	A-CGCTCGCTTCGAG	CGGATTTTT-GG--G	CGCCCCGCTGATCGGT	GATGGCGCCAACGGG	GGGCC--G-----G	GAC-AAGACG-GCGG						404
15	SEQ18	A-TGCGCCCACCCAG	GCGTTGCTG-GG--T	CGCCCCGTTGATCGGT	GATGGGGCGGCTGGA	GCACCCGGC-A-CCG	GGC-AGGCCG-GCGG						410
16	SEQ19	A-TGCGCCCACCCAG	GCGTTGCTG-GG--T	CGCCCCGTTGATCGGT	GATGGGGCGGCTGGA	GCACCCGGC-A-CCG	GGC-AGGCCG-GCGG						410
17	SEQ20	A-CGCGCCCACCCAG	GCGCTGTTG-GG--G	CGTCCGCTGATCGGC	AACGGCGCCGATGGG	ACCGC-GG----CCA	ATC-CCAACG-GTGG						389
18	SEQ16	A-CGCGCCCACCCAG	GCGCTGTTG-GG--G	CGTCCGCTGATCGGC	AACGGCGCCGATGGG	ACCGCGG-----CCA	ATC-CCAACG-GTGG						389
19	SEQ17	A-CGCGCCCACCCAG	GCGCTGTTG-GG--G	CGTCCGCTGATCGGC	AACGGCGCCGATGGG	ACCGCGG-----CCA	ATC-CCAACG-GTGG						398
20	SEQ25	A-TGCGCCCACCCAG	GCACTGTTG-GG--A	CGTCCGTTGATCGGT	GATGGTGCCGACGGG	ACGGCGC-----CCG	GGC-AAAACG-GCGG						389
21	SEQ6	A-TGCGCCCACCTCAG	GCGCTGTTG-GG--G	CGTCCGTTGGTGGGC	AATGGCGCCAATGGG	GCCCCGGGG-A-CTG	GGG-CAAACG-GCGG						397
22	SEQ26	A-CGCGCCCACACAG	CTGTGGTTG-GG--G	CGCCCCGCTGATCGGT	GATGGCGTTCACGGG	GCGCCGGGG-A-CCG	GGC-AGCCCG-GTGG						413
23	SEQ1	GGTGGGCGCACGGCG	CTGATTTCT--GG--G	CTCCACCGGAACCC	GCGACCGCCCTTCG-	ACTACATGC-AGCAG	GTCTACGACC-GCTA						413
24	SEQ3	AGCGGCTCTGCGCAG	ACGCTGTTG-TC--G	CGCCCCGTTGATCGGC	AATGGCGCCAATGCG	ACGAC-G-----CCG	GGCGGTAACG-GCGG						395
25	SEQ4	A-TGCGCCCACCCAG	GCGCTGCTG-GG--G	CGCCCCGTTGATCGGT	GACGGCGCCAATGCG	ACGAC-----TCCC	GGCGGGGCCG-GCGG						389
26	SEQ5	A-TGCGCCCACCCAG	GCGCTGCTG-GG--G	CGCCCCGTTGATCGGT	GACGGCGCCAATGCG	ACGAC-----TCCC	GGCGGGGCCG-GCGG						389
27	SEQ21	A-TGCGCCCGCAAGA	GCGTTGCTG-GG--G	CATCCACTGAT----	--TAGCGCCGACGCC	TCGACGGGT-G--GG	GGCTCGAACGCGCTG						387
28	SEQ2	AGGCGGGACAGCGCG	CCGTTTCCGCGGGCCG	CGGCCGCGCTCCGG	G-CGGCGCATAACGG-	CCAAC---T---CGT	TGCCAACACGGCCAC						389
29	SEQ9	---GCGCCGTCAACG	CCGTCGCGG-----	-GTCAGGTGACGGGC	AATGGCGGCTCCGG-	-CAACAG-----C--	GGC-ACTTCG-GCTG						389
30	SEQ15	---GCGCCG--CGT	CCGTTTCCGACGG-G	TTCCAAGCGTTTCGTC	TATGGTCCGATTCAC	GCGACCGGCCAGCAA	TGGATCAACA-GCCC						395
31	SEQ24	A---CGCCCCGGAAG	GTGCCCTCCTGG--A	TTTCAGAT-ATTGCC	AATATCTTCAACATT	GGGGCTGGG-AACTG	GGC--CTCCG--CCG						415

	451	465	466	480	481	495	496	510	511	525	526	540	
1	SEQ27	GGACGGCG-----GC	TGGTT--AATCGTCA	--ACGGCG-GCGCCG	-GAGGAT-CCGGCGC	GGCCGGCG-----	-----GC-						447
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	300
3	SEQ31	GGACGGCG-----GC	TGGTT--AATCGGCA	--ACGGCG-GCGCCG	-GAGGAT-CCGGCGC	GGCCGGCGTCAACGG	CGGGGCCGGC-GGC-						465
4	SEQ14	GCCCGGCG-----GG	TGGTT--GATCGGCA	--ACGGCG-GCGCCG	-GCGGGT-CTGGCGC	CCCCGGCGCT-----	-GGGGCCGGC-GGT-						459
5	SEQ28	GGATGCCG-----GC	TGGTT--GATCGGCA	--ACGGTG-GCGCCG	-GCGGAT-CCGGTGC	AAAGGGCGCCAACGG	CGGGGCTGGT-GGC-						465
6	SEQ12	GGCCGGCG-----GA	TTGTT--GTTCGGCA	--ACGGCG-GCGCGG	-GCGGGT-CCGGCGC	GCCCGGCG-----	-----						457
7	SEQ11	CGACGGTG-----GG	CTGTT--GTTCGGCA	--ACGGCG-GCAACG	-GCGGGT-CCGGCGC	ACCCGGAC---A---	---GGCCGGC-GGG-						459
8	SEQ7	GGCCGGCG-----GA	CTCTT--GTACGGCA	--GCGGCG-GTAACG	-GGGGAT-CAGGGCT	GGCCGGCTCC---GG	CCAGAAGGGC-GGT-						471
9	SEQ13	GCCGGGCG-----GG	TGGTT--GCTCGGTA	--ATGGTG-GGGCCG	-GCGGGT-CCGCCG	CGCTGGCTCG---GG	CCTGCCCGGC-GGG-						495
10	SEQ22	GGCCGGCG-----GG	CTGCT--GTACGGCA	--ACGGCG-GGGCCG	-GCGGGT-CCGGGGC	GCCCGGCC-----	--AGGCCGGC-GGC-						465
11	SEQ23	GGCCGGCG-----GG	TTGCT--GTGGGGCA	--ACGGCG-GGAACG	-GCGGAG-CCGGCGA	CGCCGCCAC-----	-CCCAACGGC-GGC-						474
12	SEQ30	GGCTGGGG-----GG	ATCTT--GTACGGCA	--ATGGCG-GCAATG	-GCGGCT-CCGGGGC	GGCTGGGC-----	--AGGCCGGG-GGT-						456
13	SEQ10	GCCCGGGG-----GG	TTGCT--GTACGGCA	--ACGGCG-GCAACG	-GCGGCA-CTAGTAC	CACCGCCG-----GG	-GTGGCCGGC-GGC-						474
14	SEQ8	GCCCGGCG-----GG	TTGCT--GTACGGCA	--ACGGTG-GCAACG	-GCGGCA-CCAGCAC	GACCGTGG-----GG	-ATGGCCGGC-GGC-						474
15	SEQ18	GGCCGGCG-----GG	ATCTT--GTGGGGCA	--ACGGTG-GGGCCG	-GCGGCT-CCGGCGC	GCCCGCC-----	--AAGTCGGC-GGG-						477
16	SEQ19	GGCCGGCG-----GG	ATCTT--GTGGGGCA	--ACGGTG-GGGCCG	-GCGGCT-CCGGCGC	GCCCGGCC-----	--AAGTCGGC-GGG-						477
17	SEQ20	CGCCGGTG-----GG	TTGTT--GTACGGCA	--ACGGCG-GCAACG	-GGTTTT-CCCAAAC	CACCGCCG-----GG	-CTCACCGGG-GGT-						459
18	SEQ16	CGCCGGTG-----GG	TTGTT--GTACGGCA	--ACGGCG-GCAACG	-GGTTTT-CCCAAAC	CACCGCCG-----G	GCTCACCGGG-GGT-						459
19	SEQ17	CGCCGGTG-----GG	TTGTT--GTACGGCA	--ACGGCG-GCAACG	-GGTTTT-CCCAAAC	CACCGCCG-----G	GCTCACCGGG-GGT-						468
20	SEQ25	GGCGGGCG-----GG	CTGCT--GTATGGCA	--ATGGCG-GTAATG	-GCGCGG-CCGGGGT	GAACGCGG-----G	GATTGCCGGC-GGC-						459
21	SEQ6	CGATGGCG-----GG	ATTTT--GTTCGGGT	--CCGGGG-GGGCCG	-GCGGGT-CCGGAGC	GGCCGGCA-----	--TGGCGGGT-GGC-						464
22	SEQ26	GGCCGGGG-----GG	TTGTT--GTGGGGTA	--ATGGCG-GTAACG	-GCGGTT-CCGGGGC	GGCCGGTC-----	--AAGTCGGT-GGG-						480
23	SEQ1	CATCGCAC-----CC	CACTA--CTTGGGCT	--ATGCGT-TTTCCG	-GCCTGT-AC-ACGC	CCGCGCAGTT-----	--TCAGCCGT-GG--						480
24	SEQ3	CGACGGCG-----GA	TGGCT--ATTCGGCA	--GCGGCG-GCAACG	-GCGCGC-CCGGCGC	GGCGGGCC-----	--AGTCCGGC-GGT-						462
25	SEQ4	GGCCGGCG-----GT	CTGCT--GTTCGGCA	--ACGGCG-GGGCCG	-GGGAG-CCGGGGC	GCCCGGCC-----	--AGGCCGGC-GGG-						456
26	SEQ5	GGCCGGCG-----GT	CTGCT--GTTCGGCA	--ACGGCG-GGGCCG	-GGGAG-CCGGGGC	GCCCGGCC-----	--AGGCCGGC-GGG-						456
27	SEQ21	AGCCGGGTCCAAAGC	ATGTT--CCTCGGCA	--CTGGCG-GCTCCA	-GTGCAC-TTGGCGG	TAGCGCCGCT---GC	AAATGCCGCT-G---						463
28	SEQ2	CAACCTGG-----AA	TCCCT--CTACGGCG	C-ATGGTC-GGCCAA	-CCCGTT-CCCATT	CTCCGCCAGATCATC	GCCAACCAGCAGGTT						468
29	SEQ9	CGGCGGCC--A--AC	CCGAA--TTCCGACA	ACACAGCGAGCATCG	-CCGATA-GGGGCAC	AAGCGCCATC-----	-ATGACC-----						457
30	SEQ15	GGTCGGCGA-GGCGC	TCGCCCCGATGTCA	ATGCGCCGACAAACG	TGCTGCT-CCGGCCG	GATCTGAT-----	-----CGGC-----						465
31	SEQ24	CGTCGGAC-----	TTGCTAGGGCTGGCC	A-GCGGTGGTCTTCT	TCCGGCTGCCGAGGA	AGCCGCC-----	--T---CG--AGG--						481

	541	555	556	570	571	585	586	600	601	615	616	630	
1	SEQ27	-AACGGCGG--CGCC	GGCGG--GCTCATCG	GCAACGGCGGGCGCCG	GCGG---CGCT----	GG---CGGGGTCGCC	AGCAGT-----	GGG	516				
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	300				
3	SEQ31	-AACGGCGG--CGCC	GGCGG--GCTCATCG	GCAACGGCGGGCGCCG	GCGG---CGCC----	GG---CGGGAGGGCC	AGCACG-----	GGG	534				
4	SEQ14	-AACGGCGG--GGCC	GGCGG--GCTGTTTCG	GCAGCGGGCGGGGCCG	GCGG---GGCC----	TC---CACCGACGTC	GCCGGC-----	GGG	528				
5	SEQ28	-CCTGGTGG--GGCC	GCCGG--GCTGTTTCG	GCAACGGCGGGGCCG	GCGG---TGCC----	GG---CGGAACCGCC	ACCGCCAACAACGGG	540					
6	SEQ12	---GGGC----CGGC	GGATT--GTTGTTTCG	GCAACGGCGGGCGCCG	GCGG---CCCC----	GG---CGCGTCCGGC	GGCGCGC----	TGGG	524				
7	SEQ11	-GCCGGCGG--GGCG	GCCGG--GTTCTTTCG	GCAACGGTGGCAACG	GCGG---GGAC----	GG---C-----	-----	GGG	513				
8	SEQ7	-AACGGAGG--AGCT	GCCGG--ATTGTTTG	GCAACGGCGGGGCCG	GCGG---TGCC----	GG---CGCGTCCAAC	CAAGCC-----	GGC	540				
9	SEQ13	-GCCGGCGG--GGCC	GCCGG--GTTGTTTCG	GCACCGGCGGGGCTG	GTGG---GGCC----	GG---CGGGAGTTCC	A-CGGTA-----	GGT	564				
10	SEQ22	-CCAGGTGG--AGCC	GCGGG--GCTATTTCG	GCAACGGCGGGGCCG	GTGG---GGCC----	GG---GGGTGACGGC	--CCG-----	GGC	531				
11	SEQ23	-AACGGCGG--GGAT	GCCGG--AATGTTTCG	GTAACGGCGGGGCCG	GCGG---TGCC----	GTTACAGCCCGGCG	GCCGGAACCGGCGC	552					
12	SEQ30	-GCCGGCGG--GCCG	GCCGG--GCTGATCG	GCCATGGCGGGTCCG	GCGG---GGCC----	GG---C-----	-----	GGC	510				
13	SEQ10	-AACGGTGG--CGCC	GCCGG--GCTGATCG	GCAACGGCGGGGCCG	GGG---CGGC----	GG---C-----	-----	GGG	528				
14	SEQ8	-AACGGTGG--TGCC	GCCGG--GCTGATCG	GCAACGGTGGGTTTCG	GGG---CGGC----	GG---C-----	-----	GGG	528				
15	SEQ18	-GCCGGCGG--GGCC	GCCGG--GTTGTTTCG	GCACCGGCGGGGCCG	GTGG---GGCC----	GG-----	-----	TGGG	531				
16	SEQ19	-GCCGGCGG--GGCC	GCCGG--GTTGTTTCG	GCACCGGCGGGGCCG	GTGG---GGCC----	GG-----	-----	TGGG	531				
17	SEQ20	-ACCGGCGG--GTCG	GCCGG--GTTGATCG	GCAACGGCGGCAACG	GAGG---GGCC----	GG-----	-----	CGGA	513				
18	SEQ16	-ACCGGCGG--GTCG	GCCGG--GTTGATCG	GCAACGGCGGCAACG	GAGG---GGCC----	GG-----	-----	CGGA	513				
19	SEQ17	-ACCGGCGG--GTCG	GCCGG--GTTGATCG	GCAACGGCGGCAACG	GAGG---GGCC----	GG-----	-----	CGGA	522				
20	SEQ25	-AGCGGCGG--GGCG	GCCGG--GTTGATCG	GCAACGGCGGATCCG	GTGG---TGCC----	GG-----	-----	CGGA	513				
21	SEQ6	-AACGGCGG--GGCC	GCCGG--GCTGTTTCG	GCAACGGCGGAGCCG	GCGG---AGCC----	GG-----	CGGC	AGCGCGACGGCCGCT	533				
22	SEQ26	-CCCGGCGG--CGCG	GCCGG--GTTGTTTCG	GCAACGGCGGGTCCG	GCGG---GTCC----	GG-----	-----	CGGG	534				
23	SEQ1	-ACCGGCAT--CCCC	AGCCT--GACCTACG	ACCAATCGGTCGCCG	AAGG---CGCC----	GG---CTATCTTCAC	ACCGCGATCATGCAG	555					
24	SEQ3	-AACGGCGG--GTCA	GCCGG--ACTGTGGG	GTAACGGCGGGCGCGG	GTGG---CGCC----	GG-----	CGGC	AGC	519				
25	SEQ4	-CCTGGCGG--GCCC	GCCGG--ATTGTGGG	GCAACGGCGGGCCCCG	GCGG---GGCC----	GGC---GGC-----	-----	AGC	513				
26	SEQ5	-CCTGGCGG--GCCC	GCCGG--ATTGTGGG	GCAACGGCGGGCCCCG	GCGG---GGCC----	GGC---GGC-----	-----	AGC	513				
27	SEQ21	--CCAGCGG--TGCA	CTGCA--GCTCCAAC	CCACCGGTGGGGCCA	GCGGTTTGTCC----	GCCGTGCGGCGCCCTG	CTGCCGCGCGCCGGA	543					
28	SEQ2	TACTGGCAGCAGATC	GCCGC--GGCGCTCG	CCAACGCCGTCCAGA	ACTTCCCCGCCCT--	GGTGGCGAATTTGCC	AGCG-----	GCC	546				
29	SEQ9	--ACGGCAAG--CGCG	ACCGC--G-TCTTC-	-CACGGGCGTCGATG	GCGGAATAGC-----	GGCGACGTAT----	G	-----	CGGT	519			
30	SEQ15	-AACGGCGTACCCGG	GACGGCGGCAGCTC-	CCAACGGTGGCCCCG	GCGGT-TTGCATTTC	GGTGACGG-----	-----	TGGG	534				
31	SEQ24	--AGGGCTT--GGAG	GGCGCGGGCCTATCG	GAGCTGGGGGCGGCG	GAGGC--GGCCGT-C	GGCG-----	-----	AGGC	542				

	631	645	646	660	661	675	676	690	691	705	706	720	
1	SEQ27	ATCG-GCGGGTCCGG	-CGGAGCCGGCGGCA	ACGCC---AT--GCT	GTTTCGGCGC-CGGCG	GCGCCGGCGGGCGCCG	GCGGGCGGCGTAGTAG						598
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	300
3	SEQ31	ACCG-GCGGCGCCGG	-CGGCGCCGGCGGGG	CCGCCGGCAT--GCT	GTTTCGGGGC-CGCCG	GGGTCGGCGGTCCCG	GCGGATTCGCAGCCG						619
4	SEQ14	GCCG-GTGGGGCCGG	-CGGGGCCGGCGGAA	ACGCCGGCAT--GCT	GTTTCGGCGC-CGCCG	GGGTCGGCGGCGTTCG	GCGGATTCCTGAACG						613
5	SEQ28	ATCG-GCGGGGCCGG	-TGGCGCTGGCGGGT	CCGCC---AT--GCT	GTTTGGGGC-CGGCG	GCGCCGGCGGGCGCCG	GCGGGGCTGCGACGT						622
6	SEQ12	CTGA-TCGGCAACGG	-CGGTAACGGCGGTA	AGGG-----	-----	-----	-----	-----	-----	-----	-----	-----	556
7	SEQ11	GCCG-GAGCGAACGG	-CGGCGCCGGCGGCA	CCGCCGGCTG--GTT	CTTTCGGCTT-CGGCG	GCAACGGCGGGGGCCG	GCGGGATCGGTGTTG						598
8	SEQ7	AACG-GCGGCGCCGG	-CGGAAACGGCGGCG	CCGGTGGGCT--GAT	CTGGGGCAC-CGCCG	GGACCGGTGGCAACG	GCGGGTTACCACCT						625
9	SEQ13	GATG-GCAGGGCCGG	-GGGTGCCGGTGGAT	CAGGTGGCTG--GTT	GTTGGGCAC-CGGTG	GGGTCGGCGGGGTTCG	GCGGGCTCGGGGC--						647
10	SEQ22	AACG-GCGCGGCTGG	-CGGGGCCGGCGGGG	CCGGCGGGCT--GCT	GTTTCGGCTC-CGGCG	GCGCCGGTGGGCCCCG	GCGGAGTCGGGAACA						616
11	SEQ23	GCCG-GTGGAGCCGG	-CGGCGCCGGTGGAG	CCGGTGGCTG--GCT	AAGCGGCAA-CGGCG	GCGCCGGCGGCAACG	GCGGTACTGGCGCA						637
12	SEQ30	TCCG-GCGCGGCCGG	-CGGGGCCGGCGGGC	ACGGCGGATG--GCT	GTGGGGCAA-CGGCG	GCGTCGGCGGATCCG	GCGGGCGGGTGTTCG						595
13	SEQ10	GCCG-GCGCGGCCGG	-CGGCAATGGCGGTTG	CGGGCGGGTG--GCT	GTATGGCAA-CGGCG	GCGCCGGCGGGGCCG	GTGGGACATCGGTGA						613
14	SEQ8	CCC-GCGCGGCCGG	-CGGCAACGGCGGCG	CCGGCGGGTG--GCT	ATTCGGCAA-CGGCG	GCGCCGGCGGTTGCCG	GCGGCCTCGGCGTAG						613
15	SEQ18	CCC-GCGCGGCCGG	-TGGGGCCGGGGGTA	GCGGCGGCTG--GCT	GCTGGGCAA-TGGTG	GAGTCGGCGGGGCCG	GCGGG--CAGAGCCT						614
16	SEQ19	GCCG-GCGCCGCCGG	-TGGGGCCGGGGGTA	GCGGCGGCTG--GCT	GCTGGGCAA-TGGTG	GAGTCGGCGGGGCCG	GCGGG--CAGAGCCT						614
17	SEQ20	GCCG-GCGCCAACGG	-CGGCGCGGGCGGCA	ACGGCGGCTG--GCT	CTACGGCAG-CGGCG	GCAATGGCGGGCGCCG	GCGGCGCCGGCCCCG						598
18	SEQ16	GCCG-GCGCCAACGG	-CGGCGCGGGCGGCA	ACGGCGGCTG--GCT	CTACGGCAG-CGGCG	GCAATGGCGGGCGCCG	GCGGCGCCGGCCCCG						598
19	SEQ17	GCCG-GCGCCAACGG	-CGGCGCGGGCGGCA	ACGGCGGCTG--GCT	CTACGGCAG-CGGCG	GCAATGGCGGGCGCCG	GCGGCGCCGGCCCCG						607
20	SEQ25	GCCG-GTGGCGCCGG	-CGGGTCCGGCGGAC	AAGGCGGGCT--GCT	GTACGGCAA-CGGCG	GCGCGGGCGGCAACG	GCGGGG-CGGCGACT						597
21	SEQ6	GCGG-CCGGGGCCGG	-CGGGAACGGCGGGG	CCGGCGGGCT--GCT	GTTTCGGTAC-CGCCG	GGGCGGGCGGCAACG	GCGGGTTAAGCCTCG						618
22	SEQ26	GCCG-GCGCTGCCGG	-CGGTGTTCGGCGGAT	CCGGCGGGTG--GTT	GAACGGCAA-CGGCG	GGGCGGGCGGGGCCG	GCGGG-----						609
23	SEQ1	CAAG-TCGCGGCCGG	-CAATGACGTTGT--	--GGTGT-TG--GGT	TTCTCGCA---GGGC	GCGTCGGTCGCCACC	CTGGAATGCGCCAT						633
24	SEQ3	GGCG-GCGCCGCCGG	-CGGCAACGGCGGTA	ACGGCGGGTG--GCT	GTTTCGGCGC-CGGCG	GCACCGGCGGTATCG	GCGGCACCGGTGCTC						604
25	SEQ4	GGTG-GGGGCACCGG	-CGGTGCCGGCGGCG	CCGGTGGGTG--GCT	GTTTCGGGGT-TGGCG	GCGCCGGCGGTGTTCG	GTGGGGCCGGTGGCG						598
26	SEQ5	GGTG-GGGGCACCGG	-CGGTGCCGGCGGCG	CCGGTGGGTG--GCT	GTTTCGGGGT-TGGCG	GCGCCGGCGGTGTTCG	GTGGGGCCGGTGGCG						598
27	SEQ21	GCGG-CCGCCGCCG-	-CGGCGCTGCCGGCT	CTGGCCGCCGA-GTC	GATCGGCAA-CGCAA	TCAAGAATCTCTACA	ACGCCGTTCG-AACCG						627
28	SEQ2	ATCG-ACGCGGCCGT	CCAGCAATTCCTG-G	CCTTCA-ACGC-GGC	GTACTIONAT-CCAAC	AGATTATTAGTCTC--	GCAGA-TCGGCTTCG						628
29	SEQ9	CGCC-TCGCAATGGG	ATGGTGGCTACGTGG	CCAATTACAC--GAT	CACCCAATT-CGGGC	GCGACTTCGATGACC	GATTGGCGGTTGCAA						605
30	SEQ15	GCCG-GCTATACCGG	-CGGTAACGGTGGGA	GTGCCGG-----GTT	AATCGGCAA-CGGGC	GTACCGGTGGCGCCG	GCTTTGCCGGCGGAG						616
31	SEQ24	GCCGATCGCGGCCGG	TCTGGG--TGCGGCT	CCGCTAGCCGCCGGC	CTTAGTCGGGCGTTCG	TCAATCGGCGCATTG	TCGGTGCCGCC-CAG						629

	721	735	736	750	751	765	766	780	781	795	796	810	
1	SEQ27	---CCC---TCACGG	GCGGGGCCCGCGGCG	C---CGGCGG--GGC	TGCCGGCAA-----C	GCCGGCC-TGC-----	---TCTT--CGGCGCC						663
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	300
3	SEQ31	---CTT---TC---G	GCGCCACCGGCGGCG	C---CGGCGG--GGC	CGGCGGAAA-----C	GGCGG-----GC-----	--TGTT--CGCCGAC						678
4	SEQ14	---GC-----G	GTGCCACCGGCGGGG	C---AGGCGG--GGC	CGGCGGGGC-----G	GGCGG-----GC-----	--TGTT--TGGCGCC						669
5	SEQ28	---CTC---TTGTCTG	GTGGCATCGGCGGTA	C---CGGCGG--AAC	CGGCGGCAA-----C	GCCGGTA-TGC-----	--TCGC--CGGCGCC						687
6	SEQ12	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	556
7	SEQ11	---CCGGCATCAACG	GCGGTCTCGGCGGCG	C---CGGCGG--CGA	CGGCGGCAA-----C	GCCGG-----GT-----	--TCTT--CGGCAAC						663
8	SEQ7	---TTC----T--TG	ATGCCGCTGGGGGTG	C---CGGCGG--GGC	CGGCGGCGC-----C	GGTGG-----GC-----	--TGTT--CGGCGCG						684
9	SEQ13	-----	-CGGCGCCGGTGGGG	C---CGGCGG--GGT	TGGTGGGGC-----C	GGCGG-----GC-----	--TGTT--GGGTGCT						699
10	SEQ22	---C-----	-CGGCACAGGCGGGC	T---CGGCGG--CGA	CGGCGGGGC-----G	GCCGG-----GC-----	--TGTT--CGGAGCC						669
11	SEQ23	---GC-----G	GCGCCGACGGCGGCG	G---CGGCCT--CCC	CCCCGTCCC-----C	GCCAGC---CC-----	--TGG---CGGTAAC						693
12	SEQ30	---GCG-----CAG	GCGTGGCTGGCGGTC	A---CGGCGG--TGC	GGGCGGTGC-----C	GCCGG-----GC-----	--TGTG--GGGCGCC						654
13	SEQ10	---TAC-----CCG	GTGTCCCGGCGGCA	A---TGGCGG--GGC	TGGCGGGTC-----C	GCGGG---AC-----	--TGTG--GGGTACC						672
14	SEQ8	---CGC-----CCG	GCGTGCCCGGCGGCG	C---CGGCGG--TGC	CGGCGGCGC-----C	GGCGG---TG-----	--TCGG--CGGACCC						672
15	SEQ18	---GCTG----GGCG	GGGCAACCGGCGGGG	C---CGGCGG--CAA	CGCCGACT-----G	TTCGG---GG-----	--TCGG--CGGAACC						675
16	SEQ19	---GCTG----GGCG	GGGCAACCGGCGGGG	C---CGGCGG--CAA	CGCCGACT-----G	TTCGG---GG-----	--TCGG--CGGAACC						675
17	SEQ20	---CCGAGCCATCG	GCGCGCCGGGCGTGC	C---GGGTGG--TGC	CGGCGGGGC-----C	GGGGGAA-CGGCCGG	ATTGTT--CGGCAAC						672
18	SEQ16	---CCGAGCCATCG	GCGCGCCGGGCGTGC	C---GGGTGG--TGC	CGGCGGGGC-----C	GGGGGAA-CGGCCGG	ATTGTT--CGGCAAC						672
19	SEQ17	---CCGAGCCATCG	GCGCGCCGGGCGTGC	C---GGGTGG--TGC	CGGCGGGGC-----C	GGGGGAA-CGGCCGG	ATTGTT--CGGCAAC						681
20	SEQ25	---ATCC-----CCG	GCGGCAACGGCGGCG	C---CGGTGG--TGC	CGGCGGCAA-----C	GCGTG---GC-----	--TATT--CGGCAAC						657
21	SEQ6	---GTT-----TGG	GCGTCGCCGGCGGCG	C---CGGCGG--CGC	CGGCGGGTC-----G	GGCGGTA--GT--GA	CACCGC--CGGACAC						683
22	SEQ26	-----	-----ACCGGCGCTA	A---CGGTGG--TGC	CGGCGGCAA-----C	GCCTG---GT-----	--TGTT--CGGGGCC						657
23	SEQ1	---CTGGCAAGCCTG	CCG--GCCGGCGTGC	---CGCCGA--GTC	CGGATCAGC-----T	CTCGTT--CGT-----	ATTGCT--GGGCAAC						699
24	SEQ3	---CC-----G	GCGCCATGGGCGGCA	C---CGGCGG--CAA	CGGCGGCAA-----C	GGCGCGC-TGC-----	--TGAT--CGGCGGC						663
25	SEQ4	GCACCGGCGGGGCGG	GCGGGCCCGGTGGTT	TGATCTGGGG--CGG	CGGCGGGGC-----C	GGCGGTGTCGGTGGG	GCCGGTGGCGGCACC						681
26	SEQ5	GCACCGGCGGGGCGG	GTGGGCCCGGTGGTT	TGATCTGGGG--CGG	CGGCGGGGC-----C	GGCGGTGTCGGTGGG	GCCGGTGGCGGCACC						681
27	SEQ21	TGGGTGCAGT--ACG	GCTTCAACCTCACCG	C-ATGGGCGGTGGGA	TGGCTGCCCTACATC	GGCATACTGGCACCG	CAGATCAACTTCTTC						714
28	SEQ2	---CCCAGCTATTCTG	CCACGACGGTCGGTC	A----GGGGGT-CAC	CAGCGTCAT-----T	GCCGGGT--GGCCAA	CCTTGCGGCGGAG-C						703
29	SEQ9	TTCACT-----TTG	CCTGAAAATGCCTCT	ATTTCGAACG----C	GTGCTGCGCT-----	-----	-----						650
30	SEQ15	TGGGCGG---CATGG	GCGGCACCGGCGGCT	GG-TTGATGGG-CAA	CGGCGGCAT-----G	GGTGG---CGCGGGC	GGTGTCCGGCGTAAC						693
31	SEQ24	----CTG-----G	GCGGGACAGGCCAAC	C---TGATATC-GAG	CACCAGCAC---ACT	GCAGGGCGCTGGCTG	GACCACCGCGGCACC						701

	811	825	826	840	841	855	856	870	871	885	886	900	
1	SEQ27	GCCGGAGTCGGCGGG	GCCGGCGG	---ATTC	ACGAACGGCA	-----	-----	-----	-----	-----	-----	-----	700
2	SEQ29	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	300
3	SEQ31	GGCGGGGTCGGTGGC	GCCGGCG	-----	-----	-----	-----	-----	-----	-----	-----	-----	700
4	SEQ14	GGAAGGGAACGCGGC	AGCGGCGG	---GTCC	GGCAACCTCACTGGC	GGGGCCGGCGGGGCC	-----	-----	-----	-----	-----	-----	726
5	SEQ28	GCCGGGGCCGGCG	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	700
6	SEQ12	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	556
7	SEQ11	GGCGGCAACGGCGGC	ATGGGCGG	---GGCC	GGGGCGGCCGGCGTG	AACGCCGTCAATCCC	GGCCTGGCCACCCCG	GTCACCCCGGCGGCC	-----	-----	-----	-----	750
8	SEQ7	GGCGGGGCCGGCGGC	GTAGGCGG	---CGCC	GCCCTCGGCGGCGGC	GCCCAGGCCGCGGTT	GGCAACGGCGGTGCG	GGCGGGGTCG	-----	-----	-----	-----	766
9	SEQ13	GGCGGGCACGGCGGC	GCCGGCGG	---GCTA	GGCGCCGTCACCGGT	GGGGTCGGGGGAAC	GGCGGAGCCGGTGGG	CTGCT	-----	-----	-----	-----	776
10	SEQ22	GGTGGCATCGGCGGG	GCCGGCGG	---GCCA	GGCTTCAACGGTGGC	GCCGGCGGGGCC	-----	-----	-----	-----	-----	-----	723
11	SEQ23	GGCGGTGGCGGTGAT	GCCGGTGG	---TGCT	GCTGGGATGTTCCGC	ACCGGCGGCGCCGTT	GGGACCGGCGG	-----	-----	-----	-----	-----	761
12	SEQ30	GGCGGCGGCGGTGGC	AATGGCGG	---GAAC	GGCGCCGATGCCAAC	ATCG	-----	-----	-----	-----	-----	-----	700
13	SEQ10	GGCGGGGCCGGTGGC	GACGGCGG	---CAAC	GGCCGGTCCGGGGCCA	GTCAACGTCGCCGGC	AGCGC	-----	-----	-----	-----	-----	734
14	SEQ8	GCCGGGTTGTGGGG	CACGGGGG	---TGCC	GGCGGGGTCGGCGGT	GCCGGCGTG	-----	-----	-----	-----	-----	-----	723
15	SEQ18	GGCGGGCCCCGGCGGG	CCCGGCGG	---GCC	GGCGGGTCCGGCGGT	ACCGGTGGTGCCGTT	GGCCTGGGCGGGACC	CTCTACGGGGCCGGC	-----	-----	-----	-----	762
16	SEQ19	GGCGGGGCCGGCGGG	CCCGGCGG	---GCC	GGCGGGTCCGGCGGT	ACCGGTGGTGCCGTT	GGCCTGGGCGGGACC	CTCTACGGGGCCGGC	-----	-----	-----	-----	762
17	SEQ20	GGCGGTGCCGGCGGT	GCCGGCGG	---AGCG	GGCGGCGCCGGCGGC	CGCGGTGGCGACGGC	GGGTCGGCCGGCTGG	CT	-----	-----	-----	-----	746
18	SEQ16	GGCGGTGTCGGCGGC	GTCGGCGG	---TGAC	GGCGGACAAGGCGGC	AACGGTGCCGGCGCC	GGGGCCAGCGGCACC	AAGGGCGGTGACGCC	-----	-----	-----	-----	759
19	SEQ17	GGCGGTGTCGGCGGC	GTCGGCGG	---TGAC	GGCGGACAAGGCGGC	AACGGTGCCGGCGCC	GGGGCCAGCGGCACC	AAGGGCGGTGACGCC	-----	-----	-----	-----	768
20	SEQ25	GGTGGGGCCGGGGGG	CTCGGCGC	---GGCC	GGCGGCGCGGGTGGC	GCCGGGGTCAACCCC	CTGACCG	-----	-----	-----	-----	-----	721
21	SEQ6	GGGGGACCGGTGGT	GCCGGCGG	CCTGCTA	TTCGGGCGCCGGCGAG	GACGGCACAACGCCC	GGTGGCAACGGTGGG	GCGGGCGGTGTCGC	-----	-----	-----	-----	772
22	SEQ26	GGCGGTCCGGCGGC	GCCGGCAC	---CAAT	GGTGGCGTCGGCGGG	T	-----	-----	-----	-----	-----	-----	700
23	SEQ1	CCCAACAACCCAAAC	GGGG	-----	-----	-----	-----	-----	-----	-----	-----	-----	718
24	SEQ3	GGCGGCTCGGCGGC	GCCGGCGGCATGGGT	-----	GGCACCGGCGGCGGC	ACCGGCGGCACCGGC	GGCAACGGCGGCAAC	GGCGCGCTGCTGATC	-----	-----	-----	-----	753
25	SEQ4	GGCGGGGCCGGCGGC	CGCGCCGAGCTGCTG	-----	TTCGGGCGCCGGCGGT	GCGGGTGGGGCGGGC	ACCGACGGCGGGCCC	GGTGCTACCGGCGG	-----	-----	-----	-----	770
26	SEQ5	GGCGGGGCCGGCGGC	CGCGCCGAGCTGCTG	-----	TTCGGGCGCCGGCGGT	GCGGGTGGGGCGGGC	-----	-----	-----	-----	-----	-----	741
27	SEQ21	TATTACCTCGGCGAG	CCCATCGTGCAGGCA	-----	GTCCTGTTCAATGCG	ATCGACTTCGTGGAC	GGGAC	-----	-----	-----	-----	-----	779
28	SEQ2	TTCAGCTAGCGTTTC	AACAGCTT	-CTGGT	GGTGACTIONAACGCC	G	-CGGTGGCGAACCT	GGGTAAGGCCAT	-----	-----	-----	-----	773
29	SEQ9	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	650
30	SEQ15	GGCGGCGCCGGGGGC	-CAGGCGC	--TGTTG	TTCGGCAACGGCGGC	CTGGGCGGAGCCGGC	GGGGCTGG	-----	-----	-----	-----	-----	758
31	SEQ24	GCATGGTGCCGCGGG	GACGGTCATT	-----	-----	-----	-----	-----	-----	-----	-----	-----	726

Orden secuencial del alineamiento de los genes de la familia PE

1 Rv1430	14 Rv747	27 PGRS81
2 Rv3812	15 Rv754	28 Rv1759c
3 Rv3507	16 Rv1067	29 ORF3´
4 Rv3508	17 Rv1068	30 Rv124
5 Rv3514	18 Rv1450	31 Rv278
6 Rv3652	19 Rv1452	
7 Rv2741	20 Rv1068	
8 Rv1087	21 Rv1172	
9 Rv1088	22 Rv1325	
10 Rv1091	23 Rv1441	
11 Rv297	24 Rv2769	
12 Rv742	25 Rv3595	
13 Rv746	26 Rv3345	



ANEXO 5

ANALISIS TEORICO DE LOS POSIBLES SITIOS DE RESTRICCIÓN ENZIMATICA DE LAS SECUENCIAS DE LA FAMILIA PE DE *M. tuberculosis*

SECUENCIA	<i>Alu</i> I	<i>Eco</i> RV	<i>Not</i> I	<i>Pvu</i> I	<i>Sal</i> I	<i>Sph</i> I	<i>Stu</i> I
Rv0109			95				
Rv0124				152			181
Rv0151					94		
Rv0152	125, 171		206				
Rv0159			125				
Rv0160	43						
Rv0278					141		181
Rv0279			89				181
Rv0285	213					85	
Rv0297			92, 125, 206			175	181
Rv0335	47					32	
Rv0532	27						
Rv0746			89	152			
Rv0747	27, 77				141		181
Rv0742			92	74			181, 235
Rv0754			38, 125			175	
Rv0832			125	152			
Rv0833	220						
Rv0834			92				
Rv0872			206				181, 235
Rv0916				62			
Rv0977	170			152			
Rv0978				149			
Rv0980				149			
Rv1040	205		41				
Rv1067	30, 192		89				
Rv1068	35		92	194			
Rv1087			38		66		
Rv1088		51					
Rv1089	36						
Rv1091				152			
Rv1169	180			194			
Rv1195							
Rv1214	170						254
Rv1243	82		41				181
Rv1325					66		181
Rv1386	213						

SECUENCIA	Alu I	Eco RV	Not I	Pvu I	Sal I	Sph I	Stu I
Rv1396			92				
Rv1430			125				
Rv1441							235
Rv1450			35	74, 152	102		
Rv1452				74, 152			
Rv1468			119				
Rv1646			38				
Rv1651	55, 233		41				
Rv1768			89			175	
Rv1788			124				
Rv1791	37		89				
Rv1803	198						
Rv1806	145		125	209			
Rv1818			89	74	66		181
Rv1840				74		235	181
Rv1983			119				
Rv2098	8						
Rv2099							
Rv2107	43						
Rv2126							
Rv2162	204		38				
Rv2328			89		142, 176		205
Rv2340	51		95	62			181
Rv2371				74			
Rv2396			95				
Rv2408	210	21					
Rv2431				69			
Rv2487	27		89				
Rv2490							
Rv2519			92, 149				
Rv2591	19, 224						
Rv2615				149			
Rv2634	182, 204				66		
Rv2741							235
Rv2769	51, 211	143	89				
Rv2853	182		125			175	
Rv3020		210	182	86		166	
Rv3097	193, 207			155			
Rv3344							
Rv3345	30		41, 125	155			
Rv3367	182		38				235
Rv3388			92				
Rv3477			32			172	
Rv3507			95, 125, 194				

SECUENCIA	Alu I	Eco RV	Not I	Pvu I	Sal I	Sph I	Stu I
Rv3508							235
Rv3511							
Rv3512							
Rv3514							235
Rv3539			80				
Rv3590			95, 122				
Rv3595							
Rv3622	180					175	
Rv3650			89	152			181
Rv3652			89	74			234
Rv3653			104				76
Rv3746	107, 213						
Rv3812	159		119	146			
Rv3872	213			26			
Rv3893	43, 179		80				
Rv0578			95/140				
Rv1172	192/43		38/197			175/60	
Rv1759				74/161			

Differential expression of PE and PE_PGRS genes in *Mycobacterium tuberculosis* strains

Javier Flores, Clara Espitia*

Departamento de Inmunología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México,
Apartado Postal 70228, México, D.F. 04510, Mexico

Received 27 November 2002; received in revised form 30 May 2003; accepted 13 June 2003

Received by S. Salzberg

Abstract

In this study, the expression of the *Mycobacterium tuberculosis* PE, PE_polimorphic GC-rich sequences (PGRS) gene family encoding approximately 99 glycine-rich proteins was assayed by reverse-transcriptase polymerase chain reaction (RT-PCR) in *M. tuberculosis* H37Rv, *Mycobacterium canettii* and two clinical isolates of *M. tuberculosis*. Restriction analyses and sequencing of the RT-PCR products showed that all the strains expressed the PE Rv1172c gene while the PE_PGRS Rv3652 gene was only expressed by one of the *M. tuberculosis* clinical isolates, and the PE_PGRS Rv0578c was not expressed by *M. canettii*. It was also determined that the PE_PGRS Rv0278c and Rv0279c were not expressed by any of the studied strains. The data presented in this report show that the PE, PE_PGRS genes are differentially expressed in *M. tuberculosis* strains during in vitro growth. These findings suggest that PE, PE_PGRS genes may play a role in protein variation between *M. tuberculosis* strains.

© 2003 Elsevier B.V. All rights reserved.

Keywords: RT-PCR; PE, PE_PGRS family; Expression; Mycobacteria

1. Introduction

The existence of the PE, PE_polimorphic GC-rich sequences (PGRS) family genes identified in the *Mycobacterium tuberculosis* genome has been one of the most relevant findings in the genome sequence of these bacteria. These intriguing sequences encode proteins with a 110 amino acid conserved N-terminal domain and with a large glycine-rich domain containing many repetitions of the consensus GGxGG motif (Cole et al., 1998).

Information has been recently generated about the biological role of PE, PE_PGRS in host–bacteria interactions. They are T and B cell immune response inducers (Espitia et al., 1999; Delogu and Brennan, 2001; Singh et al., 2001) and have been considered as possible virulence factors in *Mycobacterium marinum* (Ramakrishnan et al., 2000). It has also been observed that they are involved in

the cell surface interaction among mycobacteria as well as in the interaction of mycobacteria with macrophages (Brennan et al., 2001). However, direct evidence of PE, PE_PGRS protein expression is not available yet. On the other hand, mRNA transcripts for some of these genes have been identified with microarray and reverse-transcriptase polymerase chain reaction (RT-PCR) techniques (Rodriguez et al., 2002; Banu et al., 2002; Betts et al., 2002).

The PE_PGRS genes have a conserved repetitive structure that can be the target of homologous recombination, which could explain the high polymorphism observed in these sequences and demonstrated by RFLP analysis with *M. tuberculosis* and *Mycobacterium bovis* clinical isolates (Ross et al., 1992; Bigi et al., 1995). Differences in PE, PE_PGRS sequences were also observed by comparison of *M. tuberculosis* H37Rv and CDC1551 genomes (Fleischmann et al., 2002). More recently, diversity of PE_PGRS proteins among *M. tuberculosis* clinical isolates was observed by immunoblot, where size variants were detected with antibodies (Banu et al., 2002).

These observations suggest that the PE, PE_PGRS genes could be expressed differentially between strains. In the

Abbreviations: PGRS, Polimorphic GC-rich sequences; RT-PCR, Reverse-transcriptase polymerase chain reaction; RFLP, Restriction fragment length polymorphism.

* Corresponding author. Tel.: +52-5-6223818; fax: +52-5-6223369.

E-mail address: espitia@servidor.unam.mx (C. Espitia).

present work, we found differential expression of PE, PE_PGRS in *M. tuberculosis* strains under in vitro conditions. Present results contribute to support the hypothesis that PE, PE_PGRS could be a source of variable antigens in mycobacteria.

2. Materials and methods

2.1. Microorganisms and culture conditions

The laboratory strain H37Rv was obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). *Mycobacterium canettii* NLA009600046 and two *M. tuberculosis* clinical isolates NLA009501000 (1000) (belonging to Beijing genotype) and NLA009501463 (1463) were from the National Institute of Public Health and Environment, Bilthoven, The Netherlands.

2.2. RNA extraction

Mycobacterial strains were grown to mid-logarithmic phase in Middlebrook 7H9 broth (Difco, Detroit, MI) enriched with ADC (Becton Dickinson) with shaking at 37 °C.

Fifty milliliters of mycobacterial cultures were harvested by centrifugation at 3000 rpm for 10 min at room temperature. Pellets were resuspended in 2 ml of Tripure (Boehringer, Mannheim), and cell suspensions were immediately transferred to tubes containing 300–400 µl of 106 micron-glass beads (Sigma, St. Louis, MO). Bacteria were mechanically broken in a Fastprep cell disruptor (Bio101, Savant) by two pulses of 45 s at 4.5 m/s, followed by four pulses of 15 s at 6.5 m/s. Between pulses, samples were maintained on ice for 2–5 min. Nucleic acids were then extracted with 200 µl of chloroform, and after 10 min of centrifugation at 13000 rpm and 4 °C, the top aqueous layer was transferred to a fresh tube and total RNA was precipitated for 30 min at room temperature with 0.13 volumes of isopropanol. The pellet was washed with 75% ethanol and dissolved in 30–50 µl diethylpyrocarbonate-treated dH₂O. The quality of purified total RNA was assessed by agarose denaturing gel electrophoresis. Total RNA suspensions were stored at –70 °C.

2.3. Primer design

Thirty one PE, PE_PGRS sequences were randomly chosen, and their first 700 bp from the ATG start codon were aligned. The more conserved regions, located in positions 1–20 and 214–235 were chosen to generate a consensus primer sequence based on the frequency of each base in the sequence. The oligonucleotides PGRS1V 5'-ATGTCGTTTGTGATCGCGGC-3' and the anti-sense PGRS2RU 5'-TACTGGTCAAGCACGTCCTCGGA-3' were synthesized to amplify a 235-bp fragment. Oligonucleotides

specific to PE_PGRS Rv3652 were also designed to amplify a 170-bp fragment Rv3652A 5'-GGCCACGGATTGGC-TACTCTC-3' and Rv3652C 5'-CAAGGTACTCGCCAAG-CACGT-3'.

2.3.1. DNase I digestion and RT-PCR assay

Total RNA (2–4 µg) was digested with 2 units DNase I (Gibco BRL) for 30 min at room temperature. The enzyme was inactivated with 2.5 mM EDTA at 37 °C for 15 min. First strand cDNA synthesis of the transcript with high GC content was performed using Superscript preamplification System (Gibco BRL). The reaction contained 0.5 µg Oligo(dT), 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM MgCl₂, 0.5 mM dNTPs mix, 10 mM DTT and 1 U Superscript II RT reverse transcriptase (Gibco BRL). Samples were incubated at 50 °C for 1 h. Subsequent PCR reactions were carried out with 10 pmol of each primer (PGRSIV and PGRS2RU) in 1 × PCR buffer, 1 mM MgCl₂, 0.2 mM of each dNTPs, 2–5 ml first strand cDNA and 1.25 U Gold Taq polymerase (Perkin-Elmer, Applied Biosystems). Samples were denatured for 8 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1.5 min at 50 °C, 1.5 min at 72 °C and 1 cycle of 7 min at 72 °C. Amplification products were analyzed by electrophoresis in an 8% acrylamide gel. The same PCR protocol was used for amplifications of DNA of the assayed strains.

PCR reactions with Rv3652A and Rv3652C primers were also carried out with the cDNA 235-bp fragment obtained from all the studied strains. Sample was treated as above and denatured for 8 min at 94 °C, followed by 30 cycles of 1 min at 94 °C, 1 min at 58 °C, 1 min at 72 °C and 1 cycle of 7 min at 72 °C.

2.3.2. RT-PCR restriction fragment length polymorphism (RFLP)

Restriction enzyme analyses of the first 235 bp from PE, PE_PGRS sequences were carried out with the rebase software (DNAMAN V.3). Based on the presence of at least one restriction site for any of the restriction enzymes in the 99 PE, PE_PGRS sequences, 9 enzymes, *Ava*I, *Aha*I, *Bgl*II, *Eco*RV, *Not*I, *Pvu*I, *Sal*I, *Sph*I and *Stu*I, were chosen to digest the 235-bp fragment amplified from DNA and cDNA of all the studied strains. Some enzymes were used in double digestions (*Pvu*I/*Stu*I), (*Sal*I/*Stu*I) and (*Stu*I/*Ava*I) to discriminate the sequences with similar cut sites.

The restricted fragments were analyzed by electrophoresis in 8% and 10% acrylamide gels. Bands were visualized by ethidium bromide or silver stain (Bio-Rad).

2.3.3. DNA sequence

The 235 cDNA fragment, the restricted 235-bp cDNA fragments and the 170-bp fragment were separated in 3% agarose gel and subsequently stained with ethidium bromide; bands were excised from the gel and purified with the QIAquick gel extraction kit (Qiagen). DNA sequence of the

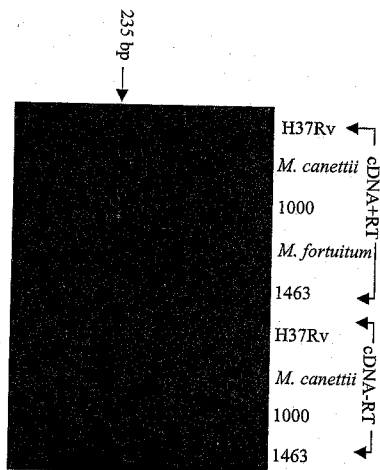


Fig. 1. RT-PCR. cDNA fragments obtained from different *M. tuberculosis* strains (cDNA + RT). Primers were designed to amplify a fragment of 235 bp. cDNA-RT indicates the same PCR mixture without reverse transcriptase (negative control).

fragments was carried out in a Genetic Analyzer ABI PRISM 310 PE (Perkin-Elmer, Applied Biosystems).

3. Results and discussion

In this study, we found that some PE, PE_PGRS genes are differentially expressed in *M. tuberculosis* strains. Although the number of sequences detected by our approach was low compared with other *in vitro* study (Banu et al., 2002), it allowed to detect heterogeneity in PE, PE_PGRS

profile expression among the strains cultured under the same conditions. All the strains studied in this work were from human origin; the reference strain H37Rv and the clinical *M. tuberculosis* isolates belong to a group considered as "modern *M. tuberculosis* strains" in contrast with *M. canettii*, a smooth variant of *M. tuberculosis*, which could have diverged from the common ancestor of the *M. tuberculosis* complex before the loss of RD, RvD and TbD1 regions, since these strains lack none of these deleted regions (Brosch et al., 2002).

Differences in growth rate curves were observed among the different mycobacteria strains. After about 9 days, *M. canettii* reached the mid-logarithmic phase followed by the laboratory strain H37Rv (11 days) and *M. tuberculosis* 1463 (12 days). In contrast, *M. tuberculosis* 1000 showed the slowest growth rate (about 21 days).

3.1. RT-PCR/RFLP of the 235-bp fragment

Due to the extensive conservation of the N-terminal region of the PE, PE_PGRS family, a consensus pair of primers was designed with the potentiality to amplify more than one PE, PE_PGRS sequence in the DNA of *M. tuberculosis* studied strains. A 235-bp DNA and cDNA fragments were amplified from all the strains but not from *Mycobacterium fortuitum*, which was used as a negative control (Poulet and Cole, 1995) (Fig. 1).

In order to identify the genes amplified from DNA and cDNA 235-bp fragments, an RFLP analysis was performed. Seven out of the nine used enzymes digested the 235-bp DNA-amplified fragment producing different restriction fragments: *AluI* (192/43 bp), *NotI* (197/38, 95/140, 86/146 bp), *SphI* (175/60 bp), *PvuI* (74/161 bp), *StuI* (181/54 bp), *SalI* (141/94 bp) and *AvaI* (61/174 bp). PCR-RFLP profile of *M. tuberculosis* H37Rv DNA fragment digested with the chosen enzymes is shown in Fig. 2. *BgIII* and *EcoRV* did

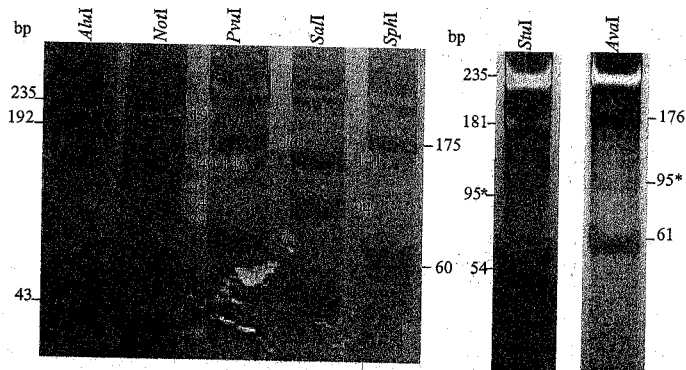


Fig. 2. PCR-RFLP. Restriction pattern of 235-bp fragment amplified from DNA of *M. tuberculosis* H37Rv. The same pattern was observed with the DNA of all studied strains except *M. canettii*, which was not cut with *PvuI*. The 95-bp bands labeled with asterisks are unspecific PCR amplification products. Size of the generated fragments is indicated in bp.

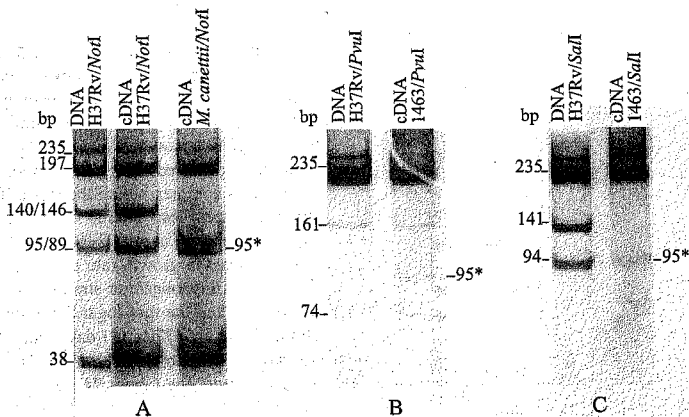


Fig. 3. RT PCR-RFLP. Representative restriction patterns generated with DNA and cDNA 235-bp fragments amplified from different *M. tuberculosis* strains. (A) The DNA and cDNA patterns obtained with *NotI* were the same for all strains as shown for *M. tuberculosis* H37Rv. However, the cDNA pattern obtained with *M. canettii* was different from the rest of the strains. (B) *M. tuberculosis* H37Rv 235-bp DNA *PvuII* digestion produced bands of 161 and 74 bp. This pattern was seen in the DNA of all the studied strains except in *M. canettii*, but at cDNA level it was only observed in *M. tuberculosis* 1463. (C) *M. tuberculosis* H37Rv 235-bp DNA *SmaII* digestion produced bands of 141 and 94 bp. The restriction patterns generated were the same for all strains. No digestion was observed with the cDNA from all strains as shown for *M. tuberculosis* 1463. Asterisks indicate unspecific PCR product. Size of the generated fragments is indicated in bp.

not digest the 235-bp DNA-amplified fragment (results not shown). Unspecific ~ 95-bp PCR products were frequently amplified from both cDNA and DNA 235-bp fragments. This was determined because reamplification of the purified 235-bp fragment produces both the 235- and 95-bp fragments. Supporting these observations is the fact that the 95-bp product is present in the undigested 235-bp fragment (Fig. 3C, cDNA 1463/*SmaII*). This and other artifactual bands that appeared in the gels are indicated with an asterisk in the gel figures.

The restriction fragments obtained with the chosen enzymes were compared with the restriction pattern of the 99 PE, PE_PGRS reported in the *M. tuberculosis* genome database. Twenty-nine sequences were selected to be po-

tentially digested with these enzymes, producing fragments of similar sizes as those mentioned above (Table 1).

The same restriction analyses were then carried out with the cDNA 235-bp fragments from all the studied strains. Results are shown in Fig. 3 and Table 2. cDNA restriction analysis showed that *NotI* 197/38-bp fragments were present in all examined strains, but *NotI* 95/140-bp fragments were lacking from *M. canettii* (Fig. 3A). In order to define the sequences present in the amplified fragments, the cDNA 235-bp fragment and their respective restricted fragments were sequenced.

The cDNA sequence of *NotI* 197-bp fragment corresponded to PE Rv1172c that codes for a 30.9-kDa protein. The sequence of *NotI* 140 bp corresponded to PE_PGRS

Table 1
Restriction fragments generated after digestion of 235-bp DNA fragment from *M. tuberculosis* H37Rv

Enzymes	<i>AhaI</i>	<i>NotI</i>	<i>SphI</i>	<i>NotI</i>	<i>NotI</i>	<i>PvuII</i>	<i>SnaI</i>	<i>SaII</i>	<i>SaII/SnaI</i>	<i>AvaI</i>	<i>AvaI/SnaI</i>
Fragments ^a (bp)	192/43	197/38	175/60	95/140	89/146	74/161	181/54	141/94	181, 141, 40, 54	61/174	181, 120, 61, 54
Sequences ^b	Rv1172c Rv0160c Rv2107	Rv1172c Rv1646 Rv3367 Rv3477	Rv1172c Rv1768 Rv3622c Rv3477	Rv0578c Rv0109 Rv2396	Rv3652 Rv0742 Rv1818c	Rv0742 Rv1759c Rv1818c Rv1840c Rv3652	Rv0278c Rv0279c Rv0742 Rv1818c Rv1840c Rv3388	Rv0151c Rv0278c Rv0747	Rv0278c	Rv0279c Rv0754 Rv1040c Rv1450c Rv1452c Rv2769c Rv1768 Rv1646 Rv3539	Rv279c

^a The first number indicates the position where the enzyme cut.

^b PE, PE_PGRS sequences that could be cut by the enzymes.

Table 2

PE, PE_PGRS sequences identified by RFLP in the 235-bp DNA and cDNA fragment from the studied strains

Sequence	Rv1172c	Rv0578c	Rv3652	Rv0278c	Rv0279c
Strain	DNA/ cDNA	DNA/ cDNA	DNA/ cDNA	DNA/ cDNA	DNA/ cDNA
H37Rv	++	++	+–	+–	+–
<i>M. canettii</i>	++	+–	+–	+–	+–
1463	++	++	++	+–	+–
1000	++	++	+–	+–	+–

Rv0578c that encodes a 105.9-kDa protein. The sequence obtained from the 235-bp cDNA of *M. canettii* corresponded to PE Rv1172c. PE Rv1172c and PE_PGRS Rv0578c were also present in the DNA 235-bp fragment amplified from all strains (Fig. 3A).

On the other hand, *PvuI* 74/161-bp fragments were present only in the cDNA 235-bp fragment amplified from *M. tuberculosis* 1463 strain (Fig. 3B), however, sequencing of the cDNA 74/161-bp fragments was not possible due to the low cDNA amount. Double digestions of the DNA 235-bp fragment with *PvuI/StuI* (Fig. 4A) enzymes allowed us to identify the PE_PGRS Rv3652 as the sequence present in these fragments; 5 out of the 99 PE and PE_PGRS genes (Rv1759c, Rv1818c, Rv1840c and Rv0742 Rv3652) could be digested with *PvuI* to produce 74- and 161-bp fragments and 3 of them (Rv1818c, Rv1840c and Rv0742) could be digested with *StuI* to produce 181- and 54-bp fragments. Double digestions of these sequences with *PvuI/StuI* would generate fragments of 74, 107 and 54 bp. In this way, Rv1818c, Rv1840c and PE Rv0742 were eliminated because the double digestion did not produce the 107-bp expected fragment. The origin of the PCR unspecific amplified ~95-bp band present in Fig. 4A was described above. It appears at the same position in both single (*PvuI*

and *StuI*) and double digestions (*PvuI/StuI*). PE_PGRS Rv1759c was eliminated because digestion with *BglII* did not produce the 137- and 98-bp expected fragments (data not shown). Taken together, these results show that the most probable sequence would be Rv3652. This observation was later confirmed by sequencing of the 170-bp fragment amplified with specific primers from the cDNA 235-bp fragment of *M. tuberculosis* 1463 strain (from the 63 to the 211 position). Interestingly, as reported in the *M. tuberculosis* genome database (<http://genolist.pasteur.fr/TubercuList/>), the PE_PGRS Rv3652 ORF of 159 bp appears to continue in the subsequent ORF but frameshifted. In addition, the frameshift gives origin to a longer sequence that stops at 312 bp. Since we amplified a cDNA of 235 bp from *M. tuberculosis* 1463 containing the frameshift mutation described for PE_PGRS Rv3652 gene, it is clear that the longer ORF is being transcribed in this strain and therefore, the frameshift mutations could be an additional source of PE_PGRS protein variability.

The comparative restriction analysis of cDNA and DNA of the 235-bp amplified fragments allowed to identify some PE_PGRS genes that were not expressed in our *in vitro* conditions. These results are important because it is possible that the not-expressed genes could be subjected to specific regulation signals. In fact, the PE_PGRS Rv0279c, one of the not-expressed genes, proved to be downregulated by iron (Rodríguez et al., 2002). These genes were demonstrated because *SalI* 141/94-bp fragments (Fig. 3C) and *StuI* 181/54-bp fragments (not shown) were absent from the cDNA of all studied strains but not from the DNA. Double digestion with *SalI/StuI* (Fig. 4B) and *AvaI/StuI* (Fig. 4C) of the DNA 235-bp fragment allowed us to confirm that PE_PGRS Rv0278c and PE_PGRS Rv0279c were the sequences present in 235-bp DNA but not in the cDNA. Three sequences can be potentially cut with *SalI* to produce

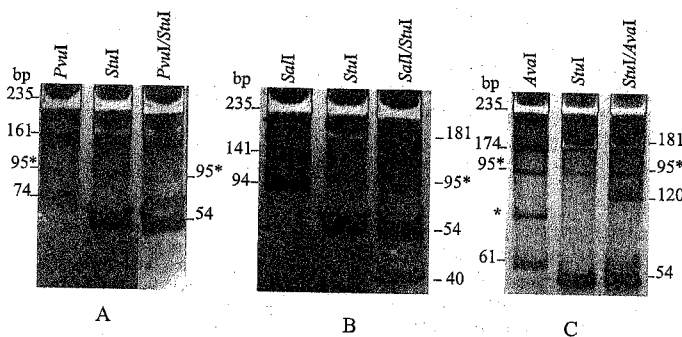


Fig. 4. PCR-RFLP. Comparative single and double digestions to define the PE, PE_PGRS sequences present in the restricted fragments. (A) Restriction pattern of the *M. tuberculosis* DNA 235-bp fragment from 1463 strain digested with *PvuI* (74/161 bp), *StuI* (181/54 bp) and with both enzymes (no additional bands were observed). (B) Restriction pattern of *M. tuberculosis* H37Rv DNA 235-bp fragment digested with *SalI* (141/94 bp) and *StuI* (181/54 bp) and with both enzymes (a new band of 40 bp defines Rv0278c). (C) Restriction pattern of DNA 235-bp fragment amplified from *M. tuberculosis* H37Rv digested with *AvaI* (61/174 bp) and *StuI* (181/54 bp) and with both enzymes (a new band of 120 bp defines Rv0279c). Bands labeled with an asterisk indicate unspecific PCR products. Size of the generated fragments is indicated in bp.

141 y 94 bp fragments: PE Rv0151c, PE_PGRS Rv0747 and PE_PGRS Rv0278c (Table 1). PE_PGRS Rv0747 was eliminated because *A*luI digestion of the DNA 235-bp fragment did not produce the 158-, 50- and 27-bp pattern bands expected for this sequence. On the other hand, PE Rv0151c was eliminated because double digestion with *S*alI/*S*tuI of the DNA 235-bp fragment generated the expected pattern for Rv0278c and not for PE Rv0151c, the 94-bp fragment disappears to give origin to a 40-bp fragment (Fig. 4B). Nine PE, PE_PGRS genes have *A*vaI sites that could generate fragments with sizes of about 61 and 174 bp: PE_PGRS Rv0754, Rv1450c Rv0279c, Rv1452c, PE Rv1040c, Rv1646, Rv1768, Rv2769c and Rv3539 (Table 1). However, among them, only the PE_PGRS Rv0279c could be digested with *S*tuI in 181-bp position, while double digestion with *A*vaI/*S*tuI produced an expected 120-bp fragment, confirming that the gene present was the PE_PGRS Rv0279c (Fig. 4C). PE_PGRS Rv0278c and Rv0279c genes code for an 83.8- and a 68.9-kDa protein, respectively. Since fragments generated by these enzymes were only seen in the DNA of the 235-bp fragment but not in the cDNA, we assumed that these sequences were not expressed in the culture conditions.

In addition to Rv0279c, other PE, PE_PGRS seem to be regulated by iron-like PE Rv1169c and Rv1195. It was also found that the PE_PGRS Rv2741 gene was induced by iron, while PE Rv0285 was repressed by iron and IdeR in a microarray assay (Rodriguez et al., 2002).

By using DNA microarrays to search for gene expression profiles of bacteria submitted to nutrient starvation, several upregulated (PE Rv3477 and Rv0335c, PE_PGRS Rv0872c and Rv3367) and downregulated (PE Rv1195, Rv1386, Rv2431c, Rv3872, Rv3020c Rv0285) PE, PE_PGRS were found (Betts et al., 2002).

By RT-PCR with specific oligonucleotides, Banu et al. (2002) demonstrated mRNA transcripts for 10 PE_PGRS in *M. tuberculosis* H37Rv (Rv0980c, Rv1067c, Rv1068c, Rv1243c, Rv1325c, Rv1369c, Rv1441c, Rv1818c, Rv1983 and Rv2162c). These results demonstrate that several PE_PGRS are being expressed. However, until now, no protein products have been revealed by proteome analysis or by traditional protein characterization protocols.

3.2. Conclusions

M. tuberculosis strain variability has proved to play a pivotal role in the outcome of the exposure to *M. tuberculosis*. One of the most important sources of genetic variability, intra- and interstrains, in these pathogenic bacteria is given by the presence of PE families. This has ranked these molecules as important targets of investigation.

Although evidence of the importance of these proteins in host–bacteria relationships is growing, direct evidence of PE protein family expression is not available yet.

The challenge to validate the antigenic variation hypothesis associated with PE family members implies the dem-

onstration that the genes encoding these proteins are expressed differentially between strains, and/or variation in pattern expression profile could be occurring during the course of the infection.

In the present work, we used a pair of consensus primers from the conserved region of PE, PE_PGRS to search for their expression in different *M. tuberculosis* strains during culture growth. With this approach, three PE, PE_PGRS genes were found to be expressed; one of them, the PE Rv1172c, was detected in all studied strains. Another, the PE_PGRS Rv0578c, was not detected in *M. canettii*, and PE_PGRS Rv3652 was detected only in the *M. tuberculosis* 1463 clinical isolate. The PE_PGRS Rv0278c and Rv0279c were not expressed in any of the studied strains. *M. canettii* showed the expression profile with most differences (Table 2), which is not surprising in view that these bacteria have shown important differences with respect to the other members of *M. tuberculosis* complex (Brosch et al., 2002).

Present results indicate that at least under the conditions used in this study, PE, PE_PGRS sequences are differentially expressed in *M. tuberculosis* strains. Although biological events have been attributed to PE family members, nothing is known about gene expression regulation. Comparing clinical isolates of *M. tuberculosis* grown in different culture conditions as well as in vivo, fundamental observations can be made on the expression of these molecules.

Acknowledgements

We are grateful to Jorge Gonzalez-y-Merchand for his help with mycobacteria RNA isolation, Dick van Soolingen for *M. tuberculosis* strains and Martha Lara for technical assistance. Isabel Perez Montfort corrected the English version of the manuscript.

This work was supported by grants from Conacyt (33580-M) and DGPA (IN221599), Universidad Nacional Autonoma de México.

References

- Banu, S., Honore, N., Saint-Joanis, B., Philpott, D., Prevost, M.C., Cole, S.T., 2002. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? *Mol. Microbiol.* 44, 9–19.
- Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., Duncan, K., 2002. Evaluation of a nutrient starvation model of *Mycobacterium tuberculosis* persistence by gene and protein expression profiling. *Mol. Microbiol.* 43, 717–731.
- Bigi, F., Romano, M.I., Alito, A., Cataldi, A., 1995. Cloning of a novel polymorphic GC-rich repetitive DNA from *Mycobacterium bovis*. *Res. Microbiol.* 146, 341–348.
- Brennan, M.J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M., Jacobs Jr., W.R., 2001. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. *Infect. Immun.* 69, 7326–7333.

- Brosch, R., et al., 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. 19, 3684–3689.
- Cole, S.T., et al., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537–544.
- Delogu, G., Brennan, M.J., 2001. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. Infect. Immun. 69, 5606–5611.
- Espitia, C., Lacllette, J.P., Mondragon-Palomino, M., Amador, A., Campuzano, J., Martens, A., Singh, M., Cicero, R., Zhang, Y., Moreno, C., 1999. The PE-PGRS-glycine-rich proteins *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? Microbiology 145, 3487–3895.
- Fleischmann, R.D., et al., 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. 184, 5479–5490.
- Poulet, S., Cole, S.T., 1995. Characterization of the highly abundant polymorphic GC-rich-repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. Arch. Microbiol. 163, 87–95.
- Ramakrishnan, L., Federspiel, N., Falkow, S., 2000. Ganuloma-specific expression of mycobacterium virulence proteins from the glycine-rich PE_PGRS family. Science 288, 1436–1439.
- Rodriguez, G.M., Voskuil, M.I., Gold, B., Schoolnik, G.K., Smith, I., et al., 2002. An essential gene in *Mycobacterium tuberculosis*: role of IdeR in iron-dependent gene expression, iron metabolism, and oxidative stress response. Infect. Immun. 70, 3371–3381.
- Ross, B.C., Raios, K., Jackson, K., Dwyer, B., 1992. Molecular cloning of a highly repeated DNA element from *Mycobacterium tuberculosis* and its use as an epidemiological tool. J. Clin. Microbiol. 30, 942–946.
- Singh, K.K., Zhang, X., Patibandla, A.S., Chien Jr., P., Laal, S., 2000. Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect. Immun. 69, 4185–4191.