



UNIVERSIDAD NACIONAL DE COLOMBIA

**ESTUDIO DEL POLIMORFISMO DE LOS GENES QUE CODIFICAN PARA LAS
PROTEÍNAS DE SUPERFICIE DEL MEROZOITO (MSP)-7C, -7H Y -7I DE *Plasmodium*
vivax A PARTIR DE AISLADOS COLOMBIANOS**

Diego Edison Garzón Ospina

Universidad Nacional de Colombia
Posgrado Interfacultades en Microbiología
Bogotá D.C., Colombia

2012

**ESTUDIO DEL POLIMORFISMO DE LOS GENES QUE CODIFICAN PARA LAS
PROTEÍNAS DE SUPERFICIE DEL MEROZOITO (MSP)-7C, -7H Y -7I DE *Plasmodium*
vivax A PARTIR DE AISLADOS COLOMBIANOS**

Diego Edison Garzón Ospina

Trabajo de investigación presentado como requisito para optar al título de:

Magister en Ciencias- Microbiología

Director:

Manuel Alfonso Patarroyo, MD, DrSc.

Coordinador Grupos Funcionales Biología Molecular e Inmunología

Fundación Instituto de Inmunología de Colombia (FIDIC)

Línea de Investigación:

Identificación de Genes Relevantes para el Desarrollo de Vacunas Contra Agentes

Patógenos Parasitarios y Bacterianos

Grupo de Investigación:

Biología Molecular – Malaria

Universidad Nacional de Colombia

Posgrado Interfacultades en Microbiología

Bogotá D.C., Colombia

2012

Agradecimientos

Quiero expresar mis más sinceros agradecimientos a aquellas personas que de una u otra forma estuvieron vinculadas al desarrollo de este trabajo. Gracias a mis compañeros de maestría, con los que compartí durante estos 18 meses dentro y fuera del salón de clase, y a mis compañeros de trabajo en el laboratorio, quienes siempre me han brindaron todo su apoyo y colaboración.

A mi madre, abuela, tíos, primos que me acompañaron de forma incondicional, a Johanna quien me ha brindado su apoyo y cariño, a mi hermano quien entendió mis ausencias a sus encuentros deportivos, a mis amigos quienes siempre me acompañaron a la distancia.

También deseo expresar mi agradecimiento a profesores y personal vinculado al posgrado de microbiología, de quienes me llevo un grato recuerdo y valiosas enseñanzas.

Finalmente un especial agradecimiento al Doctor Manuel Alfonso Patarroyo, por haber confiado en mi persona y por la dirección de este trabajo, a la Fundación Instituto de Inmunología – FIDIC, por brindarme la oportunidad de formarme, no sólo profesionalmente, sino también como persona.

A todos muchas gracias!

**ESTUDIO DEL POLIMORFISMO DE LOS GENES QUE CODIFICAN PARA LAS
PROTEÍNAS DE SUPERFICIE DEL MEROZOITO (MSP)-7C, -7H Y -7I DE *Plasmodium*
vivax A PARTIR DE AISLADOS COLOMBIANOS**

Diego Edison Garzón Ospina

Filiación: Maestría en Ciencias-Microbiología

Director: Manuel Alfonso Patarroyo Gutiérrez

Filiación: Posgrado Interfacultades en Microbiología, Facultad de Ciencias

15/11/2012

Palabras clave: *P. vivax*, *msp-7*, variabilidad genética, selección, vacuna

Resumen: Se evaluó el nivel y la distribución del polimorfismo de los genes *msp-7C*, *msp-7H* y *msp-7I* en aislados naturales de *Plasmodium vivax* y se detectó las fuerzas evolutivas que determinaban este patrón de variación genética en la población parasitaria de individuos infectados en varias regiones de Colombia.

Keywords: *P. vivax*, *msp-7*, genetic diversity, selection, vaccine.

Abstract: The level and distribution of genetic polymorphism in *msp-7C*, *msp-7H* and *msp-7I* genes from a natural population of *Plasmodium vivax* was assessed. The evolutionary forces that determined this pattern of genetic variation in parasite samples from Colombian *P. vivax*-infected individuals were also detected.

Descripción del problema estudiado y resultados más relevantes: La malaria es una de las enfermedades infecciosas que afecta las regiones tropicales y subtropicales del mundo y, a pesar de los múltiples esfuerzos en el desarrollo de una vacuna, no hay hasta el momento una totalmente efectiva. El conocimiento sobre la estructura genética de las poblaciones parasitarias y su dinámica demográfica, puede ser utilizado para la selección de los alelos más frecuentes o para la selección de las regiones conservadas de los antígenos parasitarios, con el fin de evitar las repuestas inmunes alelo-específicas que limiten la eficacia de cualquier vacuna. El gen *msp-7* (del inglés *Merozoite Surface Protein 7*) pertenece a una familia multigénica que en *Plasmodium vivax* tiene 11 miembros funcionales. Algunos de los productos proteicos de estos genes han sido caracterizados como blancos de anticuerpos en infecciones naturales. El presente estudio analizó la variabilidad genética de tres miembros de la familia *Pvmsp-7* en la población parasitaria Colombiana. Las secuencias obtenidas de estos genes fueron comparadas con las

secuencias de referencia de la cepa Sal-I y algunos aislados Coreanos. A diferencia de lo previamente descrito en la familia *msp-7* de *P. falciparum* y los genes ancestrales de *P. vivax*, las duplicaciones especie-específicas (*Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I*) de esta última especie, mostraron tener un alto polimorfismo generado por mutaciones puntuales, regiones repetitivas y recombinación. La región central de estos genes parece estar bajo una presión selectiva positiva y permite distinguir diferentes tipos alélicos. Aunque los tres genes *msp-7* evaluados mostraron una alta variabilidad genética en la región central, el extremo 3' de estos (el cual codifica para el dominio Pfam número: PF12948) es conservado, y podría ser una región importante para ser evaluada como candidato durante el desarrollo de una vacuna antimalárica.

Conclusiones de la Tesis:

1. Existen diferencias en el polimorfismo entre los miembros de la familia *msp-7*, en las dos principales especies de parásito causantes de malaria en el ser humano.
2. La diversidad nucleotídica de *Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I* los coloca entre los genes más polimórficos descritos hasta la fecha.
3. El polimorfismo no está distribuido a lo largo de toda la molécula; mientras la región central es altamente variable, los extremos 5' y 3' son conservados.
4. Diferentes fuerzas evolutivas, entre las que se incluyen la selección, la recombinación y la migración, mantienen el patrón de polimorfismo de los genes *Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I*.
5. La región C-terminal, que es codificada por el extremo 3' es conservada y probablemente sea el dominio funcional de este grupo de proteínas, por lo tanto, puede ser tomada en cuenta para su evaluación durante el desarrollo de una vacuna antimalárica.

Sugerencias:

1. Evaluar el polimorfismo de otros miembros de esta familia multigénica.
2. Evaluar la antigenicidad e inmunogenicidad de la región C-terminal de la proteína.

Agradecimientos:

Quiero expresar mis más sinceros agradecimientos a aquellas personas que de una u otra forma estuvieron vinculadas en el desarrollo de este trabajo. También deseo expresar mi

agradecimiento a profesores y personal vinculado al posgrado de microbiología, de quienes me llevo un grato recuerdo y valiosas enseñanzas.

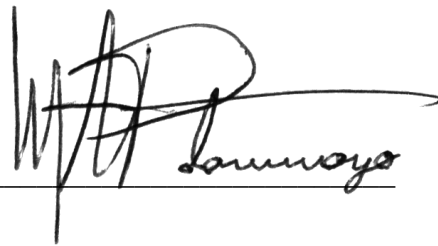
Productos académicos:

Garzón-Ospina D, López C, Forero-Rodríguez J, Patarroyo MA.

Genetic Diversity and Selection in Three *Plasmodium vivax* Merozoite Surface Protein 7 (*PvmSP-7*) Genes in a Colombian Population.

PLoS ONE. 2012; 7(9): e45962. doi:10.1371/journal.pone.0045962

Vo.Bo. Director de Tesis _____



M.A. Patarroyo

Resumen

La malaria, es una de las enfermedades infecciosas que afecta las regiones tropicales y subtropicales del mundo y, a pesar de los múltiples esfuerzos en el desarrollo de una vacuna, no hay hasta el momento una totalmente efectiva. El conocimiento sobre la estructura genética de la población parasitaria y la dinámica demográfica de la enfermedad, puede ser utilizado para la selección de los alelos más frecuentes o para la selección de las regiones conservadas de los antígenos parasitarios, con el fin de evitar las repuestas inmunes alelo específicas que limiten la eficacia de cualquier vacuna. El gen *msp-7* (del inglés *Merozoite Surface Protein 7*) pertenece a una familia multigénica que en *Plasmodium vivax* tiene 11 miembros funcionales. Algunas de las proteínas PvMSP-7 han sido caracterizadas como blancos de anticuerpos en infecciones naturales. El presente estudio analizó la variabilidad genética de tres miembros de la familia *Pvmsp-7* en la población parasitaria Colombiana. Las secuencias obtenidas de estos genes fueron comparadas con las secuencias de referencia de la cepa Sal-I y algunos aislados Coreanos. A diferencia de lo previamente descrito en la familia *msp-7* de *P. falciparum* y los genes ancestrales de *P. vivax*, las duplicaciones especie-específicas (*Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I*) de *P. vivax* mostraron tener un alto polimorfismo generado por mutaciones puntuales, regiones repetitivas y recombinación. La región central de estos genes parece estar bajo una presión selectiva positiva y permite distinguir diferentes tipos alélicos. Aunque los tres genes *msp-7* evaluados mostraron una alta variabilidad genética en la región central, el extremo 3' de estos (el cual codifica para el dominio Pfam número: PF12948) es conservado y podría ser una región importante para ser evaluada como candidato a una vacuna antimalárica basada en subunidades.

Palabras clave: Malaria, *P. vivax*, *msp-7*, variabilidad genética, selección, vacuna.

Abstract

Malaria is one of the major infectious diseases in tropical and subtropical countries around the world. However, a completely effective vaccine is not yet available. Extensive knowledge regarding parasite population structure and demographic dynamics in endemic areas might be useful for selecting the most frequent alleles or conserved regions for vaccine candidate design to avoid allele-specific immune responses that could decrease its protective efficacy. *msh-7* (gene that codifies for the Merozoite Surface Protein 7) belongs to a multigene family where *Plasmodium vivax* has the largest gene number so far. Some PvMSP-7 proteins have been found to be antibody targets in natural infections; thus we analyzed the nucleotide diversity of three *Pvmsh-7* genes in a Colombian parasite population and compared it with that of the Sal-I reference sequences and some Korean isolates. By contrast with what has been described for *P. falciparum* *msh-7* loci and ancestral *P. vivax* *msh-7* genes, specie-specific duplicates (*Pvmsh-7C*, *Pvmsh-7H* and *Pvmsh-7I*) of the latter specie display high genetic diversity, generated by single nucleotide polymorphisms, repeat regions and recombination. At least three major allele types are present in *Pvmsh-7C*, *Pvmsh-7H* and *Pvmsh-7I* and positive selection seems to be acting on the *msh-7*s central region. Although *msh-7* genes have high genetic polymorphism, the 3'-end region (codifying for the Pfam domain ID: PF12948) is conserved and could be an important region to be evaluated when designing a subunit-based antimalarial vaccine.

Keywords: Malaria, *P. vivax*, *msh-7*, genetic diversity, selection, vaccine.

Contenido

	Pág.
Resumen	VII
Lista de figuras	XI
Lista de tablas	XII
Introducción	13
1 PROBLEMA	15
2 JUSTIFICACIÓN	16
3 OBJETIVOS	18
3.1 Objetivo general	18
3.2 Objetivo Específicos	18
4 HIPOTESIS	19
5 MARCO DE REFERENCIA	20
5.1 Agente etiológico	21
5.2 Ciclo de vida	22
5.3 Reconocimiento de las células rojas	23
5.4 Vacunas antimaláricas	26
5.5 Candidatos a vacuna	26
5.6 Polimorfismo genético	27
5.7 Marco Legal	28
6 METODOLOGÍA	29
6.1 Población de estudio	29
6.2 Muestra	29
6.3 Extracción de ADN y confirmación de infección por <i>P. vivax</i>	29
6.4 Genotipificación de las muestras	30
6.5 Amplificación por PCR, purificación y secuenciación de los genes <i>m_{sp}-7</i>	30
6.6 Análisis de Polimorfismo, Neutralidad y Selección	31
6.7 Desequilibrio de ligamiento y recombinación	33
6.8 Análisis de agrupamiento	33
7 RESULTADOS	35
7.1 Polimorfismo en los genes <i>Pvmsp-7</i>	35
7.2 Pruebas de neutralidad y selección en los genes <i>Pvmsp-7</i>	38
7.3 Desequilibrio de ligamiento (LD) y recombinación	45
7.4 Análisis de agrupamiento	46
8 DISCUSIÓN	51
9 CONCLUSIONES	55
10 PERSPECTIVAS	57
11 BIBLIOGRAFÍA	58

12 ANEXOS	65
12.1 Anexo 1: Alineamiento del gen <i>msp-7C</i>	65
12.2 Anexo 2: Alineamiento del gen <i>msp-7H</i>	68
12.3 Anexo 3: Alineamiento del gen <i>msp-7I</i>	72
12.4 Anexo 4: Codones seleccionados positiva o negativamente dentro de cada gen. 76	
12.5 Anexo 5: Topologías inferidas para cada gen teniendo en cuenta la recombinación.	77

Lista de figuras

	Pág.
Figura 1. Distribución mundial de los casos de malaria.	20
Figura 2. Estructura interna del merozoito.	22
Figura 3. Ciclo de vida de <i>Plasmodium sp.</i>	23
Figura 4. Complejo MSP-1/-6/-7.	24
Figura 5. Representación esquemática de la familia <i>msp-7</i> en el género <i>Plasmodium</i> ...	25
Figura 6. Localización geográfica de las zonas de donde provienen las muestras de pacientes infectados por <i>P. vivax</i>	36
Figura 7. ADN genómico (ADNg) extraído a partir de sangre total.....	38
Figura 8. Confirmación de la presencia de ADNg parasitario de <i>P. vivax</i>	38
Figura 9. Genotipificación de las muestras.....	39
Figura 10. Productos amplificados de los genes <i>Pvmsp-7</i>	39
Figura 11. Alineamiento de los tipos alélicos encontrados para <i>Pvmsp-7C</i>	40
Figura 12. Alineamiento de los tipos alélicos encontrados para <i>Pvmsp-7H</i>	40
Figura 13. Alineamiento de los tipos alélicos encontrados para <i>Pvmsp-7I</i>	42
Figura 14. Ventana deslizante para los estadísticos D de Tajima, D* y F* de Fu & Li....	44
Figura 15. Ventana deslizante para ω (d_N/d_S)	44
Figura 16. Regresión lineal entre LD y la distancia nucleotídica.	45
Figura 17. Representación esquemática de los segmentos de recombinación identificados mediante RDP3 v.3.....	46
Figura 18. Filogenias inferidas por ML para <i>Pvmsp-7</i> (A), <i>Pvmsp-7H</i> (B) y <i>Pvmsp-7I</i> (C).	48
Figura 19. Filogenias inferidas por ML para la familia <i>msp7</i> de <i>P. vivax</i>	49
Figura 20. Representación esquemática de los segmentos de recombinación identificados mediante RDP3 v.3.....	50

Lista de tablas

	Pág.
Tabla 1. Análisis de la diversidad genética de <i>msp-7</i> dentro de la población Colombiana.	35
Tabla 2. Diversidad nucleotídica (π) de los antígenos de <i>P. vivax</i>	37
Tabla 3. Valores de π de los tres genes <i>Pvmsp-7</i> en cada región geográfica dentro de Colombia.....	41
Tabla 4. Pruebas de neutralidad, desequilibrio de ligamiento y recombinación.	42
Tabla 5. Número promedio de sustituciones sinónimas (d_S) y no sinónimas (d_N) para los genes <i>Pvmsp-7</i>	43

Introducción

La malaria es una infección parasitaria causada por protozoarios del genero *Plasmodium*, los cuales son capaces de parasitar reptiles, aves y mamíferos [1, 2]. Se han identificado cinco especies de estos parásitos capaces de producir la enfermedad en el ser humano: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* y *P. knowlesi*; siendo *P. falciparum* la especie causante de la forma más letal de la enfermedad, mientras *P. vivax* es la especie con la mayor distribución alrededor del mundo [3-6]. Aunque la enfermedad causada por *P. vivax* era considerada como benigna, recientes estudios han mostrado que esta especie puede causar complicaciones clínicas severas [7-10]. Para el 2009, cerca de 2.85 billones de personas estuvieron en riesgo de ser infectadas por *P. vivax* [5]. Aunque varios grupos han enfocado sus esfuerzos en el desarrollo de una vacuna contra este patógeno, son relativamente pocas las moléculas que han sido evaluadas en ensayos clínicos como candidatas a vacuna [11, 12]. Algunas de las proteínas que hacen parte de la superficie del parásito han sido descritas como candidatos promisorios [13-21], ya que parecen ser esenciales en el reconocimiento e invasión a la célula huésped y por lo tanto, respuestas inmunes dirigidas hacia éstas, evitarían la entrada del parásito [22]. La familia de proteínas de superficie del merozoito MSP (del inglés *Merozoite Surface Protein*) es un grupo de moléculas que están involucradas en la interacción inicial ente el parasito y la célula hospedera [23-26]. Sin embargo, la alta variabilidad genética de algunas de estas proteínas es el mayor problema en el desarrollo de una vacuna [27-31], ya que proteínas con un alto polimorfismo desencadenan respuestas inmunes alelo (o cepa) específicas, mostrando una baja protección después de la vacunación, mientras proteínas con un bajo polimorfismo son candidatas ideales, debido a que las respuestas inmunes no se restringirían a un alelo determinado. No obstante, en el caso de proteínas altamente variables, la investigación podría enfocarse en sus regiones funcionales conservadas [32].

El complejo proteico mayoritario sobre la superficie del merozoito involucrado en la interacción inicial entre el parásito y el eritrocito está formado por MSP-1, MSP-6 y MSP-7 [33, 34]. Las proteínas MSP-1 y MSP-6 son codificadas por genes (*msh-1* y *msh-6*) de copia única. A diferencia de éstas, MSP-7 es codificada por un gen (*msh-7*) que

pertenece a una familia multigénica; el número de copias de este gen difiere entre las diferentes especies del género *Plasmodium*, sin embargo, las proteínas codificadas parecen no haber divergido funcionalmente [35, 36]. Todos los genes de la familia *mSP-7* en *P. vivax* tiene un único exón y fueron nombrados en orden alfabético con respecto al gen adyacente PVX_082640 [35]. En *Plasmodium sp.*, diferentes MSP-7 son expresadas simultáneamente [37-39] y varias de ellas han sido localizadas sobre la superficie del parásito [17, 36, 38, 39], aparentemente no solo formando parte del complejo, sino también interactuando con los glóbulos rojos [23]. Adicionalmente, ensayos de noqueo de genes y de inhibición de la invasión, han mostrado la importancia de estas proteínas en la interacción con la célula hospedera [23, 36, 40].

Por otra parte, el grupo de Wang en 2003 [41] mostró que la MSP-7I (número de acceso en *GenBank*: XM_001350045.1) de *P. falciparum* recombinante fue reconocida por los sueros de pacientes infectados, sugiriendo que esta proteína está accesible al sistema inmune. Inmunizaciones con miembros de esta familia en un modelo animal, han mostrado que algunos de ellos son capaces de generar protección después del reto experimental [38].

P. falciparum y *P. vivax* son las especies que presentan el mayor número de miembros de *mSP-7* [35, 42]. En *P. falciparum*, la familia tiene una baja variabilidad genética [43, 44] y esta misma característica ha sido reportada para los genes ancestrales de *mSP-7* (*mSP-7A*, número de acceso en *GenBank*: XM_001614080.1 y *mSP-7K*, número de acceso en *GenBank*: XM_001614090.1) en *P. vivax* [45]. Recientemente, se reportó que dos duplicaciones especie-específicas de *P. vivax*, MSP-7C (número de acceso en *GenBank*: XM_001614082.1) y MSP-7H (número de acceso en *GenBank*: XM_001614087.1), son reconocidas por anticuerpos IgG de sueros de paciente infectados con este parásito [46]. Estas dos proteínas están filogenéticamente cercanas a PvMSP-7I (número de acceso en *GenBank*: XM_001614088.1) [35, 42] y debido a las implicaciones biológicas que la redundancia funcional tendría en el desarrollo de una vacuna, fragmentos de las diferentes proteínas de MSP-7 deben ser incluidas para bloquear esta ruta de invasión. Teniendo en cuenta lo anteriormente mencionado, en el presente trabajo se evaluó la variabilidad genética de *mSP-7C*, *mSP-7H* y *mSP-7I*, para determinar el potencial uso de sus proteínas codificadas, como candidatas a vacuna.

1 PROBLEMA

La malaria sigue siendo una amenaza para cerca de 106 países ubicados en regiones tropicales y subtropicales del mundo [47]. Cada año, más de 500 millones de personas son infectadas y más de 1 millón de éstas mueren como consecuencia de la enfermedad [48, 49]. Del total de los casos reportados de malaria alrededor del mundo, el 90% son producidos por *P. falciparum* y *P. vivax* [50], siendo el primero el más letal y el segundo el más ampliamente distribuido [2, 51]. La habilidad del parásito en desarrollar resistencia a los fármacos antimaláricos y la ausencia de una vacuna completamente efectiva para el control de la enfermedad, convierten este patógeno en un problema aún mayor.

El desarrollo de una vacuna efectiva contra la malaria se ha visto obstaculizada, entre otras cosas, por la existencia de múltiples variantes en secuencia de las proteínas que circulan dentro de la población parasitaria y las respuestas inmunes alelo-específicas que esto genera. Sin embargo, la utilización de regiones conservadas de estas proteínas, parece ser la alternativa que puede resolver este problema [27, 32].

Entre las proteínas que parecen estar involucradas en la interacción inicial entre el parásito y el hospedero, se encuentra la MSP-7 [23, 36, 40]. Esta proteína pertenece a una familia multigénica y debido a las implicaciones biológicas que la redundancia funcional de esta familia podría tener en el desarrollo de una vacuna, la inclusión de los diferentes miembros se haría necesaria [35]. A la fecha, solo se conoce que *Pvmsp-7A* y *PvMSP7K* presentan un bajo polimorfismo [45], lo cual es ideal para un candidato a vacuna, sin embargo, no se conoce cuál es la variabilidad genética de los demás miembros de la familia *mSP-7* en esta especie.

2 JUSTIFICACIÓN

La malaria representa uno de los mayores retos de control en salud pública en las regiones endémicas de esta enfermedad. Para el 2009, la Organización Mundial de la Salud reportó 225 millones de casos de malaria y 781.000 muertes por esta enfermedad alrededor del mundo [47]. En Colombia, el 14% de la población está en alto riesgo de contraer esta enfermedad [52, 53]. A pesar del desarrollo de antimaláricos para contrarrestar el parásito y de insecticidas para el control de su vector transmisor, esta enfermedad sigue siendo uno de los principales problemas de salud, debido a que estas medidas son cada vez menos eficaces, tanto por el aumento de cepas resistentes a los fármacos, como por la resistencia por parte del vector a los insecticidas [2].

En las últimas décadas, el diseño de una vacuna contra los parásitos que causan malaria, ha sido una de las estrategias propuestas para el control de la enfermedad más atractiva, desde el punto de vista costo-beneficio. A pesar de los múltiples esfuerzos dirigidos al desarrollo de inmunógenos, hasta el momento no existe una vacuna totalmente efectiva contra estos protozoarios para uso en humanos. El desarrollo de una vacuna efectiva contra los parásitos que causan malaria, está obstaculizada en gran medida por la variabilidad genética que presentan los antígenos candidatos [27, 28, 32]. Los análisis en zonas endémicas sobre la estructura de la población parasitaria y la dinámica demográfica de esas poblaciones, es una de las alternativas que permitirá seleccionar los alelos más frecuentes [28] o bien elegir dominios conservados [32] que eviten respuestas inmunes alelo-específicas, características de esta enfermedad [28, 54].

Durante la infección con *Plasmodium sp.* las proteínas de superficie de los estadios parasitarios juegan un papel importante, debido a que estas moléculas permiten el reconocimiento de la célula hospedera [23-26] y, por lo tanto, respuestas inmunes contra estas proteínas podrían inhibir la invasión [22]. En *P. falciparum*, el complejo proteico mayoritario sobre la superficie del merozoito está formado por las proteínas MSP-1, MSP-6 y MSP-7; este complejo interviene en la interacción inicial entre el parásito y el glóbulo rojo. El gen *msp-7* hace parte de una familia multigénica y se ha sugerido que las

proteínas codificadas por las diferentes copias este gen en *Plasmodium sp.* presentan una redundancia funcional [35].

P. vivax es la especie que presenta el mayor número de copias de *msp-7* [35, 42]; un reciente estudio demostró que las proteínas PvMSP-7C y PvMSP-7H son reconocidas durante infecciones naturales [46]. Por otra parte, PvMSP-7I está filogenéticamente relacionada a PvMSP-7H [35, 42] y debido a las implicaciones que la redundancia funcional podría tener, una vacuna basada en MSP-7 tendría que incluir los diferentes miembros para bloquear la interacción inicial entre el parásito y la célula hospedera. Ya que estas proteínas son reconocidas por el sistema inmune, la diversidad genética de éstas debe ser evaluada, con el fin de determinar su potencial uso en una vacuna.

3 OBJETIVOS

3.1 Objetivo general

Evaluar el polimorfismo de los genes *msh-7C*, *msh-7H* y *msh-7I* de *Plasmodium vivax* a partir de ADN parasitario aislado de pacientes procedentes de áreas endémicas de Colombia entre 2007 y 2010.

3.2 Objetivo Específicos

- Identificar los genes que codifican para las proteínas MSP-7C, MSP-7H y MSP-7I de *P. vivax* en aislados naturales.
- Determinar el polimorfismo de estos genes mediante un análisis de genética poblacional con las secuencias obtenidas de los aislados naturales.

4 HIPOTESIS

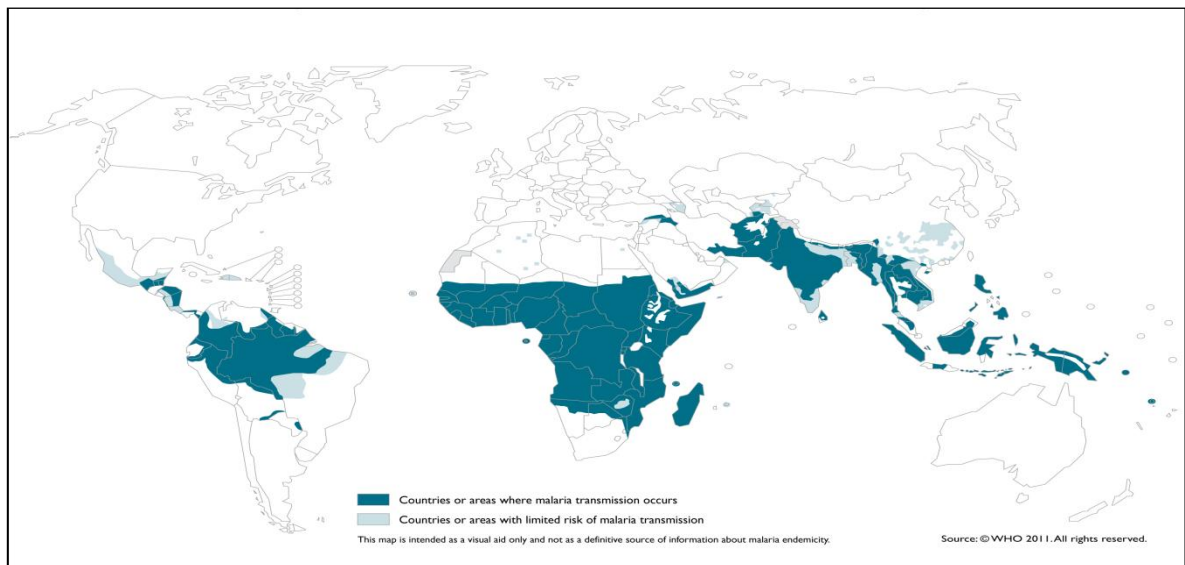
Los genes *msp-7C*, *msp-7H* y *msp-7I* de *P. vivax* presentan una baja variabilidad genética, similar a lo reportado para los genes *msp-7A* y *msp-7K* de esta misma especie, y para la familia completa en *P. falciparum*.

5 MARCO DE REFERENCIA

La malaria es uno de los principales problemas de salud alrededor del mundo [49]; esta enfermedad es producida por un grupo de *Apicomplexos* pertenecientes al género *Plasmodium*, los cuales son capaces de infectar reptiles, aves y mamíferos. En el ser humano, cinco especies son las responsables de causar la infección: *P. falciparum*, *P. vivax*, *P. ovale*, *P. malariae* [4] y *P. knowlesi* [55], los cuales son transmitidos por la picadura de mosquitos pertenecientes al género *Anopheles*.

Se estima que cerca de 106 países son endémicos para esta enfermedad [47] (figura 1) y que 3.2 billones de personas están en riesgo de contraerla. En el mundo, entre 350 y 500 millones de casos clínicos anuales son reportados, de los cuales, aproximadamente un millón terminan en muerte, siendo los niños menores de 5 años y las mujeres embarazadas los más vulnerables [48].

Figura 1. Distribución mundial de los casos de malaria.



Tomado de <http://www.malaria.com/questions/where-malaria-found>

De las cinco especies de *Plasmodium* que afectan al ser humano, *P. falciparum* es la especie que presenta mayor mortalidad y morbilidad, siendo más frecuente en el África subsahariano y algunas zonas de América del sur [49], mientras *P. vivax* es la especie

con la mayor distribución [2], cubriendo zonas de Asia, América y el norte del continente Africano. Aunque se pensaba que la infección por esta especie era benigna, estudios recientes han mostrado que *P. vivax* es capaz de causar anemia severa, complicaciones respiratorias y coma [9]. Las infecciones por *P. ovale* y *P. malariae* son relativamente raras y no suelen ser superiores al 1% [48]; por último, las infecciones por *P. knowlesi* se presentan en regiones donde cohabitan los vectores de malaria humana y los hospederos naturales de este parásito (*Macaca fascicularis* y *M. nemestrina*) [55].

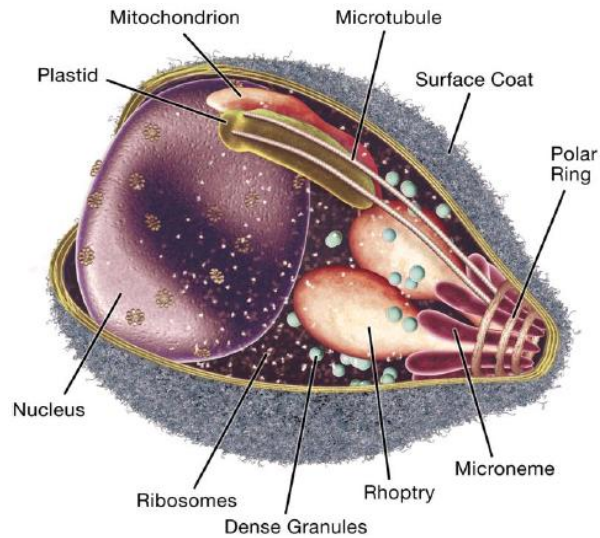
La Organización Mundial de la Salud reportó 225 millones de casos de malaria y 781.000 muertes por esta enfermedad para el 2009 [47]. En América, 23 países son endémicos de malaria, 11 de ellos ubicados en América del sur: Argentina, Bolivia, Brasil, Colombia, Ecuador, Guyana Francesa, Guyana Inglesa, Paraguay, Perú, Surinam y Venezuela. Se estima que en el continente, el 20% de la población total se encuentra en riesgo de contraer la enfermedad, presentándose entre 500.000 a un millón de casos clínicos anuales, de los cuales el 80% son causados por *P. vivax* [47].

En Colombia, el 23% de la población está en riesgo de contraer la enfermedad [47]. Para el 2011, el sistema de vigilancia en salud pública (SIVIGILA) notificó 62.716 casos, de los cuales 73,7% (46.188 casos) correspondieron a *P. vivax*, 25% (15.662 casos) a *P. falciparum*, 1,3% (841) a infecciones mixtas y 0,04% (25) a *P. malariae*; se presentaron 18 muertes distribuidas en los departamentos de Chocó, Córdoba, Nariño, Antioquia, Bolívar, Cesar y Valle del Cauca [52].

5.1 Agente etiológico

La malaria es causada por parásitos intracelulares obligados que pertenecen al filo *Apicomplexa*, orden *Haemosporidia* y género *Plasmodium*. Todos los miembros de este filo se caracterizan por un conjunto de organelos en el extremo apical (roptrias, micronemas y gránulos densos) (figura 2) fundamentales durante el proceso de invasión [56, 57]. Cerca de 200 especies de *Plasmodium* han sido descritas, infectando hospederos diversos [4].

Figura 2. Estructura interna del merozoito.

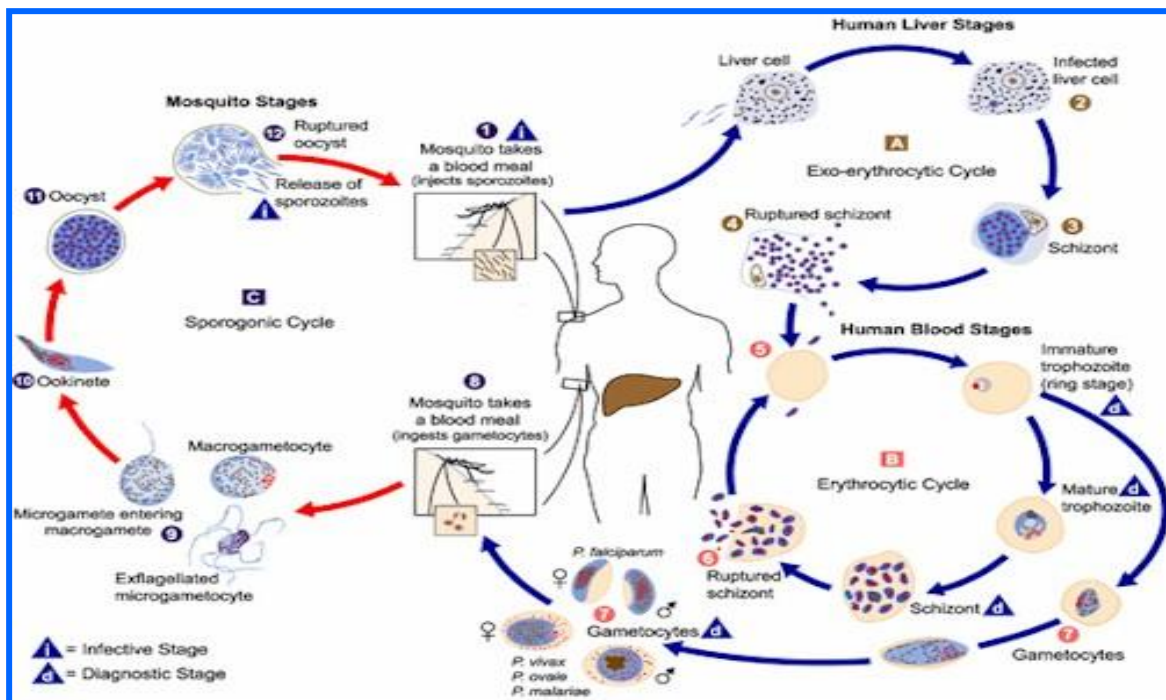


Fuente: Cowman AF, Crabb BS. *Cell*. 2006 Feb 24;124(4):755-66.

5.2 Ciclo de vida

Plasmodium sp. es un organismo digenético, con un hospedero invertebrado y uno vertebrado [4]. Este grupo de parásitos tienen un ciclo de vida complejo (figura 3), presentando en cada uno de los hospederos varios estadios, modificando tanto su expresión proteica como su forma [58]. En el hospedero intermediario (vertebrado), se lleva a cabo la fase exo-eritrocítica y la fase intra-eritrocítica, fases asexuales donde el parásito es haploide. En el hospedero definitivo (invertebrado), ocurre la fase sexual del parásito, única fase donde éste es diploide. Durante su ciclo dentro del hospedero vertebrado, el parásito permanece de forma intracelular, a excepción de dos lapsos cortos de tiempo; el primero ocurre justo después de la inoculación del parásito por el mosquito y dura alrededor de 15 minutos, en éste, los esporozoitos se encuentran extracelularmente antes de invadir las células hepáticas del hospedero; el segundo se presenta recién se liberan los nuevos estadios parásitos (los merozoitos) a partir de los hepatocitos infectados o de los glóbulos rojos infectados. Siendo los esporozoitos y los merozoitos las dos formas en donde el parásito es extracelular, son blanco fácil de la respuesta inmunitaria del hospedero, por lo cual algunas proteínas allí expresadas son consideradas posibles candidatos a vacuna.

Figura 3. Ciclo de vida de *Plasmodium sp.*



Fuente: WHO. *Malaria elimination: a field manual for low and moderate endemic countries*. 2007

5.3 Reconocimiento de las células rojas

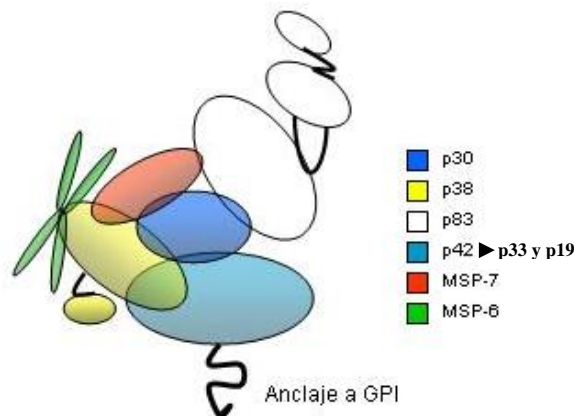
El reconocimiento de las células rojas por parte de los merozoitos está mediado por interacciones proteína-proteína [23-26, 57]. Algunas de las moléculas de superficie involucradas en el reconocimiento están agrupadas en una familia conocida como Proteínas de Superficie del Merozoito (MSP, por sus siglas en inglés), las cuales pueden estar ancladas a la membrana o estar asociadas a ésta de manera no covalente [38, 59, 60]

En *P. falciparum*, la superficie de los merozoitos se encuentra recubierta mayoritariamente por un complejo proteico formado por productos no covalentemente asociados de al menos tres diferentes genes: *msh-1*, *msh-6* y *msh-7* (figura 4) [60, 61]. MSP-1 presenta un anclaje GPI que le permite ubicarse en la membrana parasitaria; en el momento que el parásito sale del eritrocito, MSP-1 sufre sucesivos procesos proteolíticos, cuyos productos se mantienen asociados no covalentemente a la superficie del parásito. Al momento de la invasión, un fragmento de 42 kDa, derivado del precursor

de MSP-1 es procesado nuevamente, resultando dos fragmentos, uno de 33 kDa y otro de 19 kDa, liberando la mayoría del complejo.

MSP-6 es exclusiva de *P. falciparum* y no existen ortólogos en otras especies de *Plasmodium*; se sintetiza como un precursor proteico de 42 kDa. Esta proteína es hidrolizada generando un fragmento de 36 kDa, el cual interactúa con MSP-1 [34].

Figura 4. Complejo MSP-1/-6/-7.



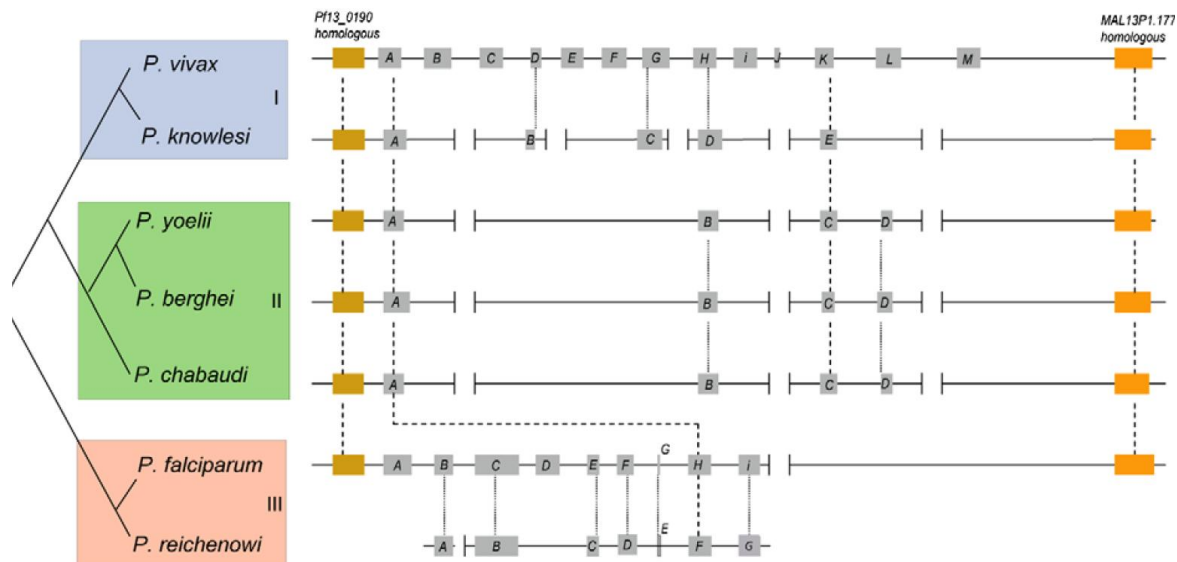
Modificado: Kauth CW, Woehlbier U, Kern M, Mekonnen Z, Lutz R, Mucke N, et al. *The Journal of biological chemistry*. 2006 Oct 20;281(42):31517-27.

El último componente estructural básico del complejo es la MSP-7 [57]. El gen codificante para esta proteína pertenece a una familia multigénica que se ha expandido diferencialmente dentro del género *Plasmodium* (figura 5), y las proteínas allí codificadas presentan una redundancia funcional según un análisis *in silico* realizado previamente [35]. Estudios de los perfiles de expresión de varios miembros de esta familia, sugieren que estos genes son regulados independientemente y son expresados al mismo tiempo. En *P. falciparum*, ARN mensajeros codificantes para estas proteínas han sido detectados, no obstante, las proteínas sintetizadas parecen no colocalizar con MSP-1 [61]. En *P. vivax*, son 11 los genes relacionados a *mSP-7*, de los cuales ocho son expresados durante el estadio de esquizonte [37], pero solo uno de ellos con evidencia experimental de su localización en membrana [17].

En *P. berghei* se ha propuesto que los miembros de MSP-7 tienen mayor o menor afinidad por MSP-1 y que la disrupción del gen *Pbmsp-7*, genera en los merozoitos una

reducción en la capacidad de infectar glóbulos rojos durante los primeros días de la infección, sin embargo, después de pocos días, la eficiencia para infectar células rojas es similar a líneas silvestres [36], esto posiblemente como consecuencia de la redundancia funcional de la familia. Estudios similares han sido llevados a cabo en *P. falciparum*, mostrando que el bloqueo del gen *Pfmsp-7* afecta el proceso de invasión [40].

Figura 5. Representación esquemática de la familia *m*sp-7 en el género *Plasmodium*.



Tomado de Garzon-Ospina D, Cadavid LF, Patarroyo MA. *Molecular phylogenetics and evolution*. 2010 May;55(2):399-408.

La capacidad de generar protección después del reto experimental difiere entre los miembros de esta familia. En *P. yoelii*, se demostró que la proteína denominada PyMSRP-2 (proteína homóloga a MSP-7) genera protección en ratones inmunizados, mientras PyMSP-7 no lo hace [38]. Por otra parte, estudios con sueros de pacientes infectados por *P. falciparum* muestran que las proteínas del complejo MSP-1/-6/-7 son antigénicas durante infecciones naturales [41], además, anticuerpos contra PfMSP-7 son capaces de inhibir el crecimiento del parásito [62]. En *P. vivax*, un reciente estudio mostró que tanto PvMSP-7C como PvMSP-7H son reconocidas por sueros de pacientes infectados por esta especie [46].

5.4 Vacunas antimaláricas

Las vacunas son las intervenciones más costo-efectivas en salud pública, con el mayor potencial de implementación a nivel poblacional para el control de las enfermedades infecciosas. A pesar de los múltiples esfuerzos dirigidos al desarrollo de una vacuna antimalárica, no existe hasta el momento una totalmente efectiva.

Entre los problemas para el diseño de una vacuna efectiva, se incluyen: el complejo ciclo de vida del parásito, la expresión de proteínas estadio-específicas y las diferentes estrategias de evasión al sistema inmune [27, 32, 58]. Entre las estrategias de evasión, se incluye la variabilidad genética de proteínas involucradas en el proceso de invasión [27, 32], debido al fuerte impacto que puede tener ésta sobre la inmunogenicidad y el potencial efecto protector de estas proteínas [27]. El alto polimorfismo que presentan algunos de estos antígenos es el resultado de la presión selectiva ejercida por el sistema inmune del hospedero [63], que permite la aparición de variantes alélicas dentro de la población parasitaria, disminuyendo así la potencial eficacia de cualquier vacuna [12]; como resultado, las diferentes variantes alélicas deberían ser incluidas para lograr una protección completa. Sin embargo, a pesar de ser la variabilidad genética un mecanismo de evasión bastante utilizado por los microorganismos, no toda la proteína puede sufrir cambios en su secuencia, debido a que ciertas regiones están restringidas funcional o estructuralmente.

Los análisis en zonas endémicas sobre la estructura de la población parasitaria y la dinámica demográfica de estas poblaciones, permitirán seleccionar los alelos más frecuentes [28] o bien elegir dominios conservados [32] que eviten respuestas inmunes restringidas a un alelo en particular.

5.5 Candidatos a vacuna

El esporozoito y el merozoito son las formas extracelulares infectivas de *Plasmodium sp.*, por lo tanto, son células importantes en el reconocimiento por parte del sistema inmune. Numerosas proteínas, tanto de superficie como de los organelos apicales del parásito, participan activamente durante la invasión a la célula hospedera. Estas moléculas son blancos primordiales, ya que respuestas inmunes dirigidas hacia ellas podría inhibir la invasión [22, 64, 65]. Sin embargo, estos antígenos deben cumplir con ciertas

características: 1) deben ser accesibles al sistema inmune del hospedero, 2) inducir respuestas inmunes protectoras y 3) carecer de variabilidad genética o por lo menos tener un bajo polimorfismo [32].

A pesar de que ya se cuenta con un número significativo de antígenos en *P. vivax* descritos [13-21], la estructura y la dinámica de la diversidad genética de solo un reducido grupo de estas proteínas ha sido evaluada [29-31, 45, 66-71], lo que se convierte en una limitante para el desarrollo de una vacuna.

5.6 Polimorfismo genético

El polimorfismo o variabilidad genética hace referencia a la existencia de más de una variante de un gen dentro de la población. Esta variabilidad es el resultado de procesos de mutación, los cuales son mantenidos o eliminados de la población por procesos como la selección natural o procesos demográficos aleatorios (como la deriva genética).

En malaria, el estudio de la variabilidad genética se ha enfocado en genes específicos que codifican para proteínas candidatas a vacuna y de genes involucrados en resistencia a antimaláricos. Estos estudios han permitido establecer cómo se han originado los diferentes alelos y cómo se han diseminado dentro de la población. Esta información es esencial para el monitoreo de intervenciones como los tratamientos antimaláricos, así como para determinar el origen de genotipos que pueden evadir la respuesta inmune [72].

Estudios previos en diferentes regiones del mundo han mostrado que algunas de las proteínas de membrana candidatas a vacuna contra *P. vivax*, como la MSP-1, MSP-3 alfa, MSP-5, DBP (del inglés *Duffy Binding Protein*) y AMA-1 (del inglés *Apical Merozoite Antigen-1*), muestran una alta variabilidad genética [30, 31, 67, 69, 70, 73], siendo éste un mecanismo de evasión utilizado por el parásito, que limita su uso en el desarrollo de una vacuna. Sin embargo, otras proteínas de superficie muestran un comportamiento diferente: tanto *Pvmsp-4*, *Pvmsp-7A*, *Pvmsp-7K* y *Pvmsp-10* son altamente conservadas, sugiriendo presiones selectivas diferentes o restricciones funcionales/estructurales en estas proteínas [45, 68]. Este mismo comportamiento se presenta en miembros del complejo de bajo peso molecular de las roptrias, cuyos miembros mostraron una baja

diversidad genética [66, 74]. Estos datos realzan la necesidad de efectuar estudios similares en otros antígenos descritos.

5.7 Marco Legal

Este proyecto fue desarrollado teniendo en cuenta la resolución colombiana N° 008430 del 4 de Octubre de 1993 del Ministerio de Salud, por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud.

6 METODOLOGÍA

6.1 Población de estudio

Aislados naturales de *P. vivax* obtenidos a partir de muestras de sangre total de pacientes infectados de diferentes zonas endémicas de Colombia, previo consentimiento informado.

6.2 Muestra

Para la obtención de ADN parasitario, se tomaron 48 muestras de sangre periférica (5 mililitros de sangre por flebotomía en tubos con EDTA) de pacientes infectados por *P. vivax* provenientes de diferentes municipios del país. Previo a la toma de la muestra, los pacientes fueron informados del propósito de la investigación y dieron su consentimiento por escrito.

6.3 Extracción de ADN y confirmación de infección por *P. vivax*

La extracción de ADN se realizó con el kit comercial *WIZARD™ Genomic DNA Purification* (Promega), a partir de 300 μ L de sangre total, siguiendo las recomendaciones del fabricante. Brevemente, por cada 300 μ L de sangre total se agregó 900 μ L de la solución de lisis celular, las mezclas fueron homogenizadas e incubadas a temperatura ambiente por 10 minutos, para su posterior centrifugación a 16.000 gravedades (X g). Los sobrenadantes fueron removidos y los *pellet* resuspendidos en 300 μ L de la solución de lisis nuclear. Las proteínas fueron separadas mediante la solución de precipitación de proteínas y por centrifugación a 16.000 X g. Los sobrenadantes fueron transferidos a tubos de 1,5 mL, el ADN precipitado con isopropanol al 100% y centrifugado a 16.000 X g para su posterior lavado con 100 mL de etanol al 70%. El ADN fue resuspendido en 100 μ L de una solución que contenía Tris-HCL y EDTA incubada por 60 minutos a sesenta y cinco grados Celsius ($^{\circ}$ C). La integridad del ADN se verificó por electroforesis en gel de agarosa al 1% en TAE 1X. Finalmente, el ADN genómico fue almacenado a -20° C para su posterior uso.

Debido a que la extracción se realizó a partir de sangre total, tanto el ADN del huésped como del parásito están presentes. Para confirmar la presencia de ADN de *P. vivax* de cada una de las extracciones, se realizó una amplificación por PCR con iniciadores específicos, dirigidos hacia la región genómica que codifica para la subunidad ribosomal 18s (SSU ARNr) de *P. vivax* (SSU-F: 5' ATGAACGAGATCTTAACCTGC 3', SSU-R: 5' CATCACGATATGTA5TGATAAAGATTACC 3'), con el siguiente perfil térmico: Una denaturación inicial a 95 °C por 5 minutos, seguida de una temperatura de 65 °C (anillamiento) por 30 segundos y una extensión a 72 °C por 45 segundos durante 10 ciclos. Durante este primer paso, la temperatura de anillamiento fue reducida en 1 grado cada ciclo (65 °C - 55 °C). Luego la PCR fue corrida por 35 ciclos (denaturación 94°C/20 segundos, anillamiento 55°C/30 segundos y extensión 72°C/45 segundos). Los productos de PCR fueron visualizados por electroforesis en gel de agarosa al 2%. La presencia de los dos tipos, A (400 pares de bases) y S (454 o 480 pares de bases) de la SSU ARNr [75] confirmaba la presencia de ADN parasitario de *P. vivax*.

6.4 Genotipificación de las muestras

Para seleccionar muestras con infección única (por un solo clon del parásito), éstas fueron genotipificadas utilizando el marcador polimórfico *Pvm*sp-1 como se ha descrito previamente [76]. Brevemente, se amplificó un fragmento del gen *m*sp-1 que incluía los bloques 6-8, el cual presenta dos diferentes longitudes: 1.080 pares de bases (pb) o 1.150 pb; si en la misma muestra se presentan ambos fragmentos, se asume que es una infección múltiple (por más de un clon de *P. vivax*), pero si se presenta sólo uno (cualquiera que sea) se asume que es una infección por un solo clon. Posteriormente, los fragmentos fueron digeridos con las enzimas de restricción Alu I y Mnl I en reacciones independientes. Si la infección es por un único clon, la suma de los fragmento debe ser igual a la del producto sin digerir; si por el contrario, la suma de los fragmentos es mayor a la del fragmento sin digerir, se trata de una infección múltiple.

6.5 Amplificación por PCR, purificación y secuenciación de los genes *m*sp-7

La amplificación de los genes *Pvm*sp-7C, *Pvm*sp-7H y *Pvm*sp-7I se realizó con aquellas muestras que presentaron infección por un solo clon. Los iniciadores específicos para cada gen fueron diseñados basados en las secuencias de la cepa de referencia Sal-I (Números de acceso en *GenBank* XM_001614082.1, XM_001614087.1 y

XM_001614088.1, para *Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I*, respectivamente). El fragmento del gen *Pvmsp-7C* fue amplificado con los siguientes iniciadores: 7Cdto 5' ACCACAAAGATGAATAAAACG 3' y 7Crev 5' CACCTCAATCGTGTTTCAGC 3'. El gen *Pvmsp-7H* fue amplificado con los iniciadores 7Hdto 5' GTGTGCATCAGTATAGCGAC 3' y 7Hrev 5' AAGAAGGTTAGCCATAAATGC 3', y *Pvmsp-7I* fue amplificado con los iniciadores 7Idto 5' ACAATGAGGGGCAAGTACG 3' y 7Irev 5' TTCATTCGTTGCTCACTTCG 3'. La amplificación de los tres genes fue realizada de forma independiente con la mezcla de la enzima *KAPA HiFi HotStart Readymix* con 0,3 µM de cada iniciador en un volumen final de 25 µL. El perfil térmico fue el siguiente: un ciclo de 5 minutos a 95 °C, seguido de 28 ciclos: 20 segundos a 98 °C, 15 segundos a 62 °C para *Pvmsp-7C* y *Pvmsp-7I* y a 60 °C para *Pvmsp-7H*, 30 segundos a 72 °C, seguido de un ciclo de extensión final por 5 minutos a 72 °C. Los productos de PCR fueron purificados con el *kit* comercial *Wizard PCR preps* (Promega), para luego ser secuenciados bidireccionalmente utilizando el método BigDye con electroforesis capilar usando ABI-3730 XL (MACROGEN, Seúl, Corea del sur) con los iniciadores de amplificación. Adicionalmente, iniciadores internos fueron usados para la secuenciación: para *Pvmsp-7C*: intdc 5' CTGTTGGACCCGGTGGAG3' e intrc 5' CTTGTTGATTCGCTCCTGG 3'; para *Pvmsp-7H*: intdh 5' TCAAATACAGCACAGACTTCC e intrh 5' CCTCAGGACAACCCGAAAG 3' y para *Pvmsp-7I*: intdi 5' TCACAAACGCACAACCCAGG 3' e intri 5' GCTCCATTACCACAACCGG 3'). Dos productos obtenidos de amplificaciones independientes fueron secuenciados para descartar errores. De ser necesario, un tercer producto era secuenciado.

6.6 Análisis de Polimorfismo, Neutralidad y Selección

Los electroferogramas obtenidos de la secuenciación fueron ensamblados y corregidos con el programa *CLC DNA workbench* v.6 (CLC bio, Cambridge, MA, USA). Las secuencias consenso fueron traducidas para obtener las secuencias de aminoácidos hipotéticas y luego éstas fueron alineadas usando ClustalW [77]. Los alineamientos de ADN fueron obtenidos a partir de los alineamientos de aminoácidos con la herramienta PAL2NAL [78]. Esta herramienta toma el alineamiento de proteínas y las secuencias de ADN para generar un alineamiento por codones de las secuencias de interés. Las secuencias Colombianas de *msp-7C*, *msp-7H* y *msp-7I* de *P. vivax* fueron comparadas

con las secuencias respectivas, previamente reportadas de la cepa Sal-I (Números de acceso en *GenBank*: XM_001614082.1, XM_001614087.1, XM_001614088.1) y otras secuencias disponibles en esta base de datos (GU476538.1, GU476534.1, GU476518.1 and ACY66906-26).

El polimorfismo genético fue evaluado con el programa DnaSP v.5 [79] mediante el cálculo de una serie de medidas de variación en el ADN. Los parámetros calculados incluyen: el número de sitios segregantes (SS), el número de sitios parsimoniosos (SP), el número de sitios singleton (S), el número de haplotipos (H), la diversidad haplotípica (dH), la diversidad nucleotídica (π) y el estimador de Waterson (θ^W). Este mismo programa fue usado para evaluar desviaciones del modelo neutral de evolución molecular mediante el cálculo de los estadísticos D de Tajima [80] y D^* y F^* de Fu & Li [81]. El estadístico D compara la diferencia entre el número total de sitios segregantes y el número promedio de diferencias nucleotídicas entre pares de secuencias. La prueba de Fu & Li calcula los estadísticos D^* y F^* ; el primero calcula la diferencia entre el número de sitios singleton y el total de mutaciones, mientras F^* calcula la diferencia entre el número de sitios singleton y el número promedio de diferencias nucleotídicas entre pares de secuencias. Valores estadísticamente significativos diferentes a cero (0) corresponden a desviaciones del modelo neutral. Adicionalmente, fue calculado el estadístico F_s [81] el cual está basado en la distribución de las frecuencias génicas. Para todas las pruebas se realizaron simulaciones considerando la coalescencia, para obtener los intervalos de confianza [79].

La selección natural fue evaluada usando el programa MEGA v.5 [82] con el método modificado de Nei-Gojobori [83] calculando el número de sustituciones no sinónimas por sitio no sinónimo (d_N) y el número de sustituciones sinónimas por sitio sinónimo (d_S). Diferencias entre d_N y d_S fueron determinadas mediante una prueba Z. Adicionalmente, sitios (codones) seleccionados positiva o negativamente fueron identificados con el servidor web *Datamonkey* [84] usando los métodos de máxima verosimilitud: IFEL [85], FEL, SLAC y REL [86]. Todos estos algoritmos estiman omega ($\omega = d_N/d_S$) sobre cada codón del alineamiento. Para IFEL, FEL and SLAC un valor $p \leq 0,1$ fue considerado como significativo, mientras para REL un factor Bayes ≥ 50 fue considerado como significativo. Las posiciones dentro del alineamiento con *gaps* (inserciones / deleciones) o

con información perdida no fueron tenidas en cuenta para los diferentes análisis realizados.

6.7 Desequilibrio de ligamiento y recombinación

El desequilibrio de ligamiento (LD, por sus siglas en inglés) fue evaluado al calcular el estadístico Z_{ns} [87] el cual es el promedio de R^2 sobre todas las posibles comparaciones de pares de secuencias. Una regresión lineal entre LD (R^2) y la distancia en nucleótidos junto con el cálculo del estadístico ZZ [88] fueron implementados para evaluar si la recombinación afecta a los genes *Pvm_{msp-7}*. Los posibles eventos de recombinación fueron determinados mediante el uso del parámetro RM (número mínimo de eventos de recombinación) [89], el cual incorpora el tamaño efectivo de la población y la probabilidad de recombinación entre nucleótidos adyacentes por generación. Todos estos análisis fueron realizados con el programa DnaSP v.5 [79]. Finalmente, el programa RDP3 v3.4 [90] fue utilizado para detectar las regiones de recombinación dentro de cada uno de los genes *msp-7*. Este programa busca evidencia de recombinación entre las secuencias alineadas, al examinar todas las posibles combinaciones por tripletas, utilizando para esto diferentes algoritmos.

6.8 Análisis de agrupamiento

Mediante el programa MEGA v.5 [46] fueron construidos árboles filogenéticos usando los métodos de máxima verosimilitud (ML, por sus siglas en inglés) y *Neighbour-Joining* (NJ) con el fin de evaluar si existe alguna relación entre el polimorfismo observado y la distribución geográfica de los aislados parasitarios. Para cada uno de los alineamientos, el mejor modelo de sustitución nucleotídica fue seleccionado por criterio de información Bayesiano (BIC, por sus siglas en inglés) usando el algoritmo ModelTest [55]. Los modelos para la construcción del árbol usando el método ML fueron: Para *Pvm_{msp-7C}* Hasegawa-Kishino-Yano (HKY) con una distribución Gamma (+G). Para *Pvm_{msp-7H}* fue usado el modelo de Tamura-Nei (TN93) asumiendo que cierta fracción de sitios son invariantes (+I) y para *Pvm_{msp-7I}* el modelo usado fue HKY+I. El modelo TN93+G fue usado para la construcción de árboles con el método NJ para los tres genes. Árboles basados en ML, los cuales describen las consecuencias filogenéticas de la recombinación, fueron inferidos usando el programa RDP3 v3.4 [90] con el método HKY. La fiabilidad de las topologías generadas fue evaluada por el método *Bootstrap* usando

1.000 repeticiones. Posiciones en las cuales estuvieran presentes inserciones/delecciones (*gaps*) fueron eliminadas durante la inferencia filogenética.

7 RESULTADOS

Cuarenta y ocho muestras de sangre periférica de pacientes con sintomatología de malaria provenientes de diferentes municipios del país: Amazonas (Calamar, Guaviare), Andes (Apartado y El Bagre, Antioquia), Caribe (Tierra Alta y Puerto Libertador, Córdoba), Orinoquia (Mapiripán, Meta y Tauramena, Casanare) y Pacífico (Istmina, Chocó y Tumaco, Nariño) fueron recolectadas entre los años 2008 y 2011 (figura 6). Todas las muestras fueron procesadas y el ADN obtenido de cada muestra fue verificado por electroforesis en gel de agarosa (figura 7); todas las muestras fueron positivas para ADN parasitario (figura 8). El análisis de RFLP realizado a las 48 muestras (figura 9) mostró que 42 de éstas correspondían a infecciones por un único clon de *P. vivax* y fueron tenidas en cuenta para las amplificaciones por PCR de los genes *msp-7*. Para algunas de las 42 muestras, no fue posible la obtención de un producto de amplificación; 37 muestras mostraron un amplicón de 1.167 pb para *Pvmsp-7C* y de 1.232 pb para *Pvmsp-7H*, mientras las 42 muestras mostraron un amplicón de 1.182 bp para *Pvmsp-7I* (figura 10).

Tabla 1. Análisis de la diversidad genética de *msp-7* dentro de la población Colombiana.

N	Gen	Sitios	Ss	S	P	H	Hd	θ^w (es)	π (es)
37	<i>msp-7C</i>	1.098	148	1	147	23	0,93 (0,03)	0,0323 (0,01)	0,0548 (0,003)
37	<i>msp-7H</i>	1.137	147	19	128	29	0,96 (0,02)	0,0309 (0,01)	0,0374 (0,003)
42	<i>msp-7I</i>	1.109	155	10	145	29	0,97 (0,01)	0,0325 (0,01)	0,0472 (0,004)

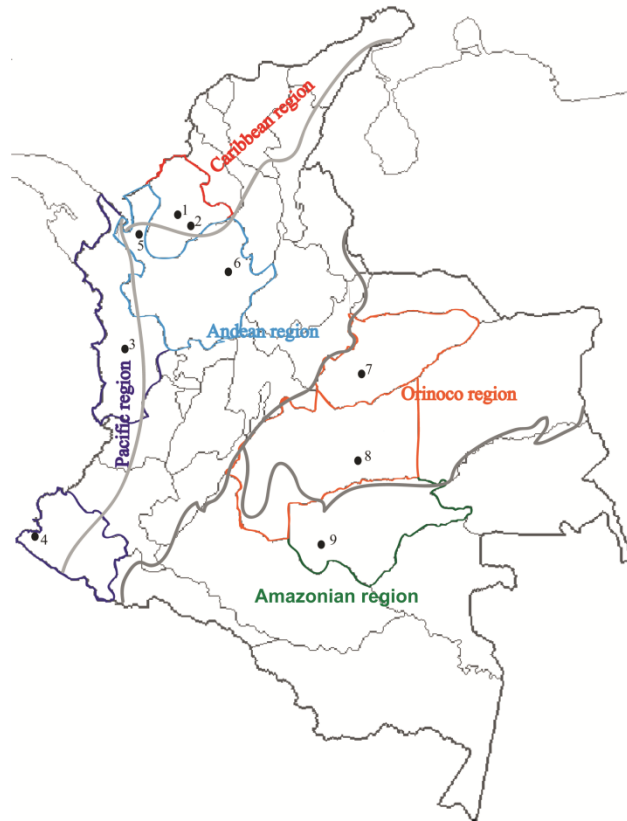
n: número de aislados, *Ss*: sitios segregantes, *S*: sitios singleton, *P*: sitios parsimoniosos, *H*: número de haplotipos, *Hd*: diversidad haplotípica, θ^w : estimador de Waterson, π : diversidad nucleotídica, es: error de estimación.

7.1 Polimorfismo en los genes *Pvmsp-7*

Fragmentos obtenidos por PCR fueron purificados y secuenciados directamente. 23 haplotipos fueron encontrados para *Pvmsp-7C*, con una diversidad haplotípica (*Hd*) de $0,93 \pm 0,03$, mientras 28 haplotipos fueron encontrados para *Pvmsp-7H* y *Pvmsp-7I*, con una *Hd* ($0,96 \pm 0,02$, $0,97 \pm 0,01$, respectivamente) más alta que la reportada para *Pvmsp-7C* (tabla 1). En *Pvmsp-7C* y *Pvmsp-7H*, de los 1.098 y 1.137 sitios analizados

respectivamente, 148 fueron sitios segregantes, mientras en *Pvm*sp-7I, 147 sitios de los 1.109 analizados fueron segregantes (tabla 1).

Figura 6. Localización geográfica de las zonas de donde provienen las muestras de pacientes infectados por *P. vivax*.



1: Puerto Libertador, 2: Tierra Alta, 3: Istmina, 4: Tumaco, 5: Apartadó, 6: El Bagre, 7: Tauramena, 8: Mapiripán, 9: Calamar.

La diversidad nucleotídica (π) de los tres genes (*Pvm*sp-7C $\pi = 0,0548$, *Pvm*sp-7H $\pi = 0,0374$ y *Pvm*sp-7I $\pi = 0,0472$) fue más alta que la π reportada previamente para otros miembros de la familia *m*sp-7 de *P. vivax* (*Pvm*sp-7A $\pi = 0$ y *Pvm*sp-7K $\pi = 0,0022$) [45]. Estos valores sitúan a *Pvm*sp-7C, *Pvm*sp-7H y a *Pvm*sp-7I entre los genes más polimórficos descritos hasta la fecha para este parásito (tabla 2). Cuando se analizó la diversidad nucleotídica (promedio de π) entre las regiones colombianas, se encontró que la región Pacífica fue la más polimórfica, seguida por las regiones Andina, Caribe y Orinoco, mientras la región Amazónica presentó los valores más bajos (tabla 3).

De la secuencia deducida de aminoácidos de PvMSP-7C, entre los residuos 134 al 238 (numeración de acuerdo al anexo 1) se observan tres tipos alélicos mayoritarios (figura 11 y anexo 1). El alineamiento de PvMSP-7H, muestra una la inserción/delección (AEAFG) desde el residuo 146 al 150 y una repetición GGT(n)V desde los residuos 199 al 213 (numeración de acuerdo al anexo 2). Tres regiones con diferentes secuencias peptídicas dentro de la proteína permite describir varios tipos alélicos (figura 12 y anexo 2). La primera región, desde el residuo 153 al 170 (numeración de acuerdo al anexo 2) presenta 4 secuencias peptídicas diferentes. En la segunda región (del residuo 172 al 194), se observan otras 5 secuencias peptídicas, mientras en la tercera región entre los aminoácidos 233 al 241 se observan otras 3 secuencias peptídicas. Las diferentes regiones se encuentran asociadas entre sí de forma aleatoria (figura 12 y anexo 2).

Tabla 2. Diversidad nucleotídica (π) de los antígenos de *P. vivax*.

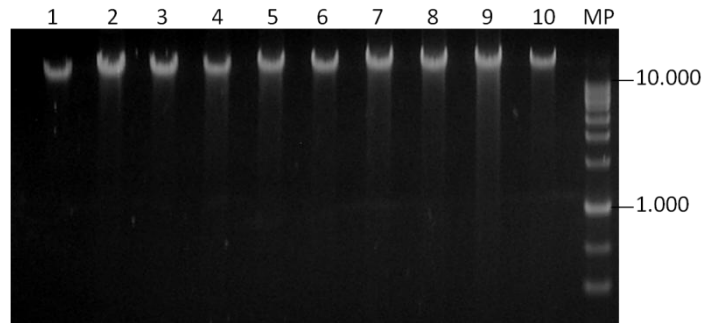
Gen	n	Sitios	π	Referencias
<i>msp-7C</i>	39	1.043	0,0582	-
<i>msp-3α</i>	35	2.577	0,0496	[69]
<i>msp-7I</i>	44	1.053	0,0493	-
<i>msp-7H</i>	59	1.068	0,0480	-
<i>msp-1</i>	175	417	0,0451	[29]
<i>msp-5</i>	22	1.524	0,0375	[30]
<i>ama-1</i>	72	900	0,0130	[91]
<i>Dbp</i>	23	1.388	0,0122	[67]
<i>Trap</i>	39	1.470	0,0059	[92]
<i>msp-7K</i>	27	886	0,0022	[45]
<i>rap-2</i>	35	1.161	0,0013	[66]
<i>msp-4</i>	30	846	0,0011	[68]
<i>rap-1</i>	29	2.413	0,0008	[66]
<i>msp-7A</i>	29	1.100	0,0002	[45]
<i>msp-10</i>	23	1.336	0,0002	[45]

n: Número de aislados, π : diversidad nucleotídica.

En el alineamiento de las secuencias deducidas de aminoácidos del gen *Pvmsp-7I*, se puede observar una repetición caracterizada por los aminoácidos EA (anexo 3). Similar a los otros dos genes, varios tipos alélicos pueden ser observados. Desde el residuo 131 hasta el 161 (numeración de acuerdo a la anexo 3) se distinguen 3 secuencias peptídicas diferentes (figura 13 y anexo 3). Desde el aminoácido 163 hasta el 219, otros cuatro péptidos fueron observados asociados a los primeros de forma no aleatoria (anexo 3).

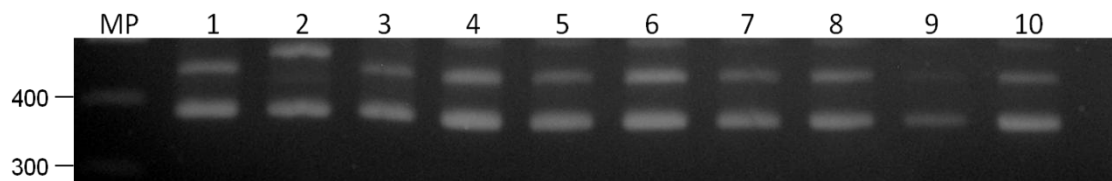
Finalmente, desde el residuo 221 al 234 cuatro péptidos fueron observados y desde el residuo 236 al 264 dos nuevos péptidos fueron encontrados, pero a diferencia de los primeros, estos péptidos están asociados de forma aleatoria (figura 13 y anexo 3).

Figura 7. ADN genómico (ADNg) extraído a partir de sangre total.



Carriles 1-10: ADNg de algunas muestras utilizadas en este estudio. MP: Marcador de peso molecular.

Figura 8. Confirmación de la presencia de ADNg parasitario de *P. vivax*.



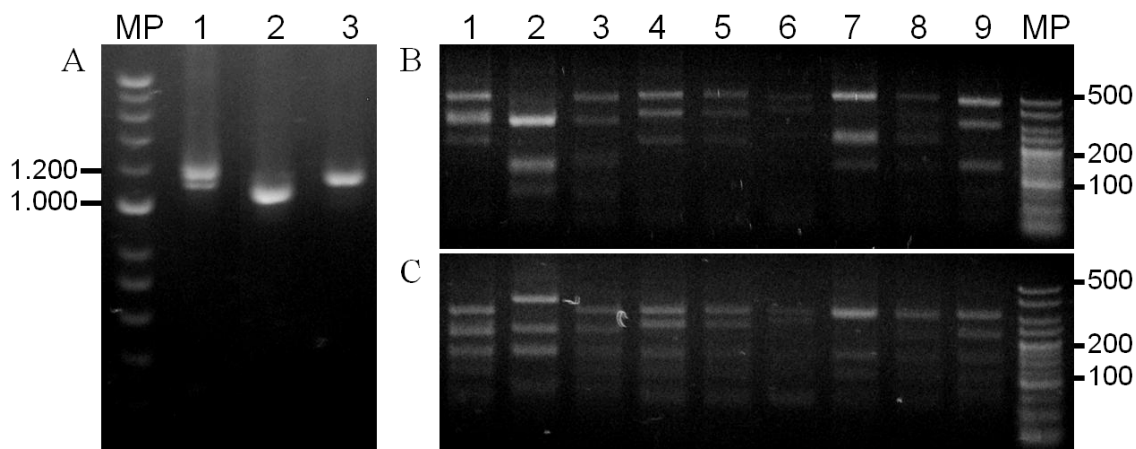
MP: Marcador de peso molecular. Carriles 1-10: Fragmentos de la SSU ARNr de algunas muestras utilizadas en este estudio.

7.2 Pruebas de neutralidad y selección en los genes *Pvmsp-7*

Varias pruebas basadas en el modelo neutral de evolución molecular fueron aplicadas a cada uno de los genes. En *Pvmsp-7C* valores estadísticamente significativos mayores a cero (0) fueron obtenidos para los estadísticos de Tajima (D $p < 0,05$) y Fu & Li (D^* y F^* $p < 0,02$) (tabla 4) indicando un exceso de alelos a frecuencias intermedias. Por otra parte, para *Pvmsp-7H* y *Pvmsp-7I*, las pruebas de neutralidad no mostraron valores estadísticamente significativos (tabla 4), por lo tanto, estos genes a diferencia de *Pvmsp-7C* parecen estar evolucionando de manera neutral. Sin embargo, un análisis de ventana deslizante para los estadísticos D , D^* y F^* mostró que diferentes presiones selectivas podrían estar afectando la evolución de los tres genes analizados aquí. Las curvas en la ventana deslizante (figura 14) sugieren que la selección balanceante podría estar

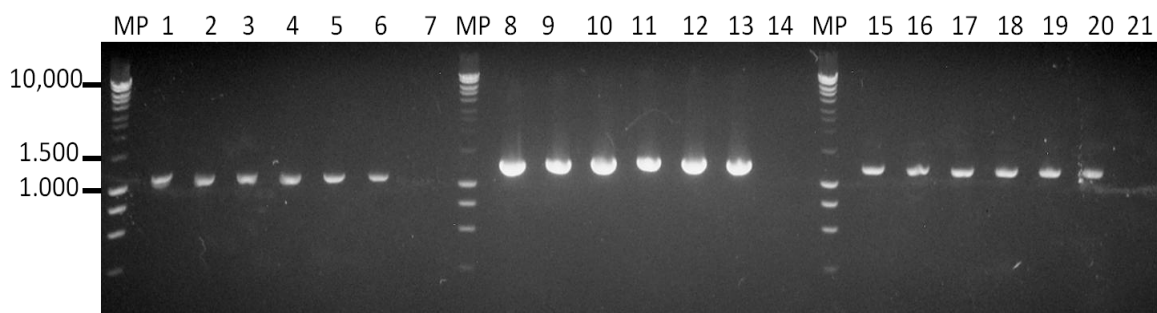
actuando en la región central de los tres genes, mientras la selección negativa estaría actuando en las regiones cinco prima (5') y tres prima (3') (figura 14). Por otra parte, el estadístico F_s de F_u no mostró valores significativos para ninguno de los genes.

Figura 9. Genotipificación de las muestras



A: Amplificación de los bloques 6-8 de *Pvmosp1*. MP: Marcador de peso molecular; Carril 1: Muestra con infección múltiple, la cual muestra dos productos de amplificación; Carril 2: Producto de amplificación de 1.080 pb; Carril 3: Producto de amplificación de 1.115 pb. B. Patrones de restricción por RFLP generados por *Alu I*. MP: Marcador de peso molecular; Carriles 1-9: Algunas de las muestras utilizadas en este estudio. C. Patrones de restricción por RFLP generados por *Mnl I*. MP: Marcador de peso molecular; Carriles 1-9: Algunas de las muestras utilizadas en este estudio.

Figura 10. Productos amplificados de los genes *Pvmosp-7*



Productos de amplificación de tres genes *msp-7* de *P. vivax*. MP: Marcador de peso molecular; Carriles 1-6: Productos de PCR del gen *Pvmosp-7C*; Carriles 8-13: Productos de PCR del gen *Pvmosp-7H*; Carriles 15-20: Productos de PCR del gen *Pvmosp-7I*; Carriles 7, 14 y 21: Control negativo.

Figura 11. Alineamiento de los tipos alélicos encontrados para *Pvm*sp-7C.

```

129                                     244
Sa1-I  KGQADTAPSV KGDVSPPPNL PAAAASSPKE TVPAGTSNGL VEADYVVLNT PDGNPRPVGP GGGSRPSASG PDAASNL-QN -Q--ATAEAG G---STN--T QGSQTGGVST TPGANQ
AND5   .....G.....-P-.....-...-...-...
PAC2   ....AGQT. ESN...R.AS S..DN.L..K .T....S.V .VR..NP.S .S-.SDALS .....SQ. .GTSP.V-. .------S. ----NSQ--- AAAEA .STN .....
CAR9   ....ASQ. . .ADT.GSK. ...D.P.RG .AAD.RNSHV .IG.INR.S A.SS.LAA.S .D.TL.... .GS.QIT.P SPSSPGGVP. NTLTNVQPQ. PVGAG.-TN .....

```

Alineamiento de las secuencias deducidas de aminoácidos de PvMSP-7C. Los puntos indican aminoácidos conservados. Las líneas representan gaps introducidos en el alineamiento.

Para determinar si la selección natural actúa sobre los genes *m*sp-7, el número promedio de sustituciones d_S y d_N fue estimado. Para los tres genes, la tasa de sustituciones d_N fue más alta que la tasa de sustituciones d_S , sin embargo, no hay diferencias estadísticamente significativas entre ambos valores (tabla 5). Una ventana deslizante para la relación d_N/d_S (omega [ω]), mostró que ω fue mayor a 1 (señal de selección positiva) en la región central de los genes *m*sp-7 (figura 15), mientras que para los extremos 5' y 3' el valor de ω fue menor a 1 (señal de selección negativa). Debido a estos resultados, los genes fueron divididos en tres regiones: el extremo 5', la región central y el extremo 3', y luego las tasas d_N y d_S fueron calculadas nuevamente. Para las regiones 5' y 3' d_S fue más alto que d_N , pero con diferencias estadísticamente significativas sólo para el extremo 5'. Por otra parte, la región central mostró una tasa d_N significativamente más alta que la tasa d_S en todos los genes (tabla 5).

Figura 12. Alineamiento de los tipos alélicos encontrados para *Pvm*sp-7H.

```

151                                     255
Sa1-I  GGVFVPTGNSASNSQSTGGSGSQNASPPQGSPPSDSAQGSQVTNSTGSTV---TLNAPSSSHSTGQPQQSAGVSLPTGTAETVASNTAQ TSPPAG
PAC2   ...T.....Q.E.N.GGNP..T....A...G---.STS...Q...S...N.AEP.A..TQE.NT.AG. PP...
CAR2   .....
ORI3   ...T.....T-.PPS..G..KSP.Q..DN.....PG.GG..G.GGTE..STS.....
AMA2   ...T.....T-.PPS..G..ESP.Q..DN.GGG...PS.GG..G.GGTV..STS..GP...GS.P...APSV.N..A.VT..EH S....
VCG-I  ...T.....-...DLN...P.Q.E.NNGGG...PG.GG..G.GGTV..STS..P...GS.P.....
AND3   .....E.NNGG.E..P...EG.V-----T...GP...GS.P.....
AMA3   .....T...GGNP..T....A...G---.STS...Q...S.P...APSV.N..A.VT..EH S....
AND11  ..A.....-..PA.VW...SD.....GGD...T...G.A...G---.STS...Q...GS.P...APSV.N..A.VT..EH S....

```

Alineamiento de las secuencias deducidas de aminoácidos de PvMSP-7H. Los puntos indican aminoácidos conservados. Las líneas representan gaps introducidos en el alineamiento.

Adicional a las tasas globales (d_N y d_S) para cada gen, varios métodos basados en máxima verosimilitud fueron aplicados para identificar codones bajo presión positiva o negativa. De acuerdo con los resultados de los métodos SCAL, FEL, IFEL y REL, *Pvmmsp-7C* tiene 19 codones bajo presión selectiva positiva y 21 sitios bajo presión selectiva negativa (anexo 4). Para *Pvmmsp-7H*, señales de selección positiva fueron identificadas en 50 codones, mientras otros 32 fueron negativamente seleccionados (anexo 4). *Pvmmsp-7I* tiene 10 codones seleccionados positivamente, mientras 34 son seleccionados negativamente (anexo 4). En todos los genes, los codones bajo selección positiva se ubican preferencialmente en la región central, mientras en los extremos 5' y 3' se encuentran los codones seleccionados negativamente.

Tabla 3. Valores de π de los tres genes *Pvmmsp-7* en cada región geográfica dentro de Colombia

n	Gen	Origen	Sitios	π (es)
Amazonas				
4	<i>msp-7C</i>		1.101	0,0059 (0,002)
4	<i>msp-7H</i>		1.146	0,0460 (0,009)
4	<i>msp-7I</i>		1.109	0,0658 (0,014)
Andes				
8	<i>msp-7C</i>		1.098	0,0571 (0,009)
14	<i>msp-7H</i>		1.137	0,0336 (0,004)
13	<i>msp-7I</i>		1.109	0,0392 (0,008)
Caribe				
15	<i>msp-7C</i>		1.098	0,0506 (0,008)
10	<i>msp-7H</i>		1.149	0,0398 (0,006)
12	<i>msp-7I</i>		1.109	0,0425 (0,008)
Orinoco				
4	<i>msp-7C</i>		1.119	0,0531 (0,019)
2	<i>msp-7H</i>		1.149	0,0252 (0,013)
5	<i>msp-7I</i>		1.133	0,0525 (0,014)
Pacífico				
6	<i>msp-7C</i>		1.098	0,0525 (0,010)
7	<i>msp-7H</i>		1.146	0,0443 (0,010)
8	<i>msp-7I</i>		1.109	0,0620 (0,008)

π : Diversidad nucleotídica, es: error de estimación.

Figura 13. Alineamiento de los tipos alélicos encontrados para *PvmSP-7I*.

	125									202
Sal-I	KIITKGAADK	EEEEEDDEDE	EAVEGD-EEA	GERDEAEAEA	VRGDTSHQSG	SGDLTPSPPG	SSERAPAV			
AND11			
AND13			
CAR8V.S	N..DGG....EA..	..EG--.E..	TNEVA.....A	.G.L...T			
AND3VV.S	N..DGG....EA..	..EG--.E..	TNEVA....D	.EASGE.D.A	.G.L...T			
PAC7V.S	N..DGG....AEA	GEE--G.E..	TNEVA....D	.AVSPR.VSE	.R.P.T..			
AMA4SS	..G....---	.DD.D-G..	QET---DKGS	T----.D.PD	.AVSPQRA.E	.P.LTT.T			
	203									270
Sal-I	APGETPRVTA	TSQGPSQTHN	PGSSGNAEPN	SASARTPPAA	AEAPLPQPAS	PDVSTSGEES	PNGRVPNV			
AND11D..			
AND13R.EGQS	.T.SIN..P.	PA.A...S.P	QV...PEG.P	ES.....			
CAR8	.R...ATDSP	..E....T	T...R.EGQS	.T.SIN..P.	PA.A...S.P	QV...PEG.P	ES.....			
AND3	.R..NATDSP	..E....T	T...R.EGQS	.T.SIN..P.	PA.A...S.P	QV...PEG.P	ES.....			
PAC7	V.A..A.DNP	P..ETAP.NP	..P...V...N....	.A.A.....	.V..APEG.P	ES.....			
AMA4	.R.A.V.DSP	..P.L.P.NP	S.Q...VGQS	.T.SIN..P.	PA.A...S.P	QV...PEG.P	ES.....			

Alineamiento de las secuencias deducidas de aminoácidos de *PvMSP-7I*. Los puntos indican aminoácidos conservados. Las líneas representan gaps introducidos en el alineamiento.

Tabla 4. Pruebas de neutralidad, desequilibrio de ligamiento y recombinación.

n	Gen	Tajima	Fu & Li		Fu	Zns	ZZ	RM
		D	D*	F*	Fs			
37	<i>mSP-7C</i>	2,094**	1,810*	2,287*	7,64	0,339	0,173*	22
37	<i>mSP-7H</i>	0,484	0,981	0,959	-0,37	0,130	0,343*	28
42	<i>mSP-7I</i>	1,115	1,385	1,535	2,95	0,234	0,332*	14

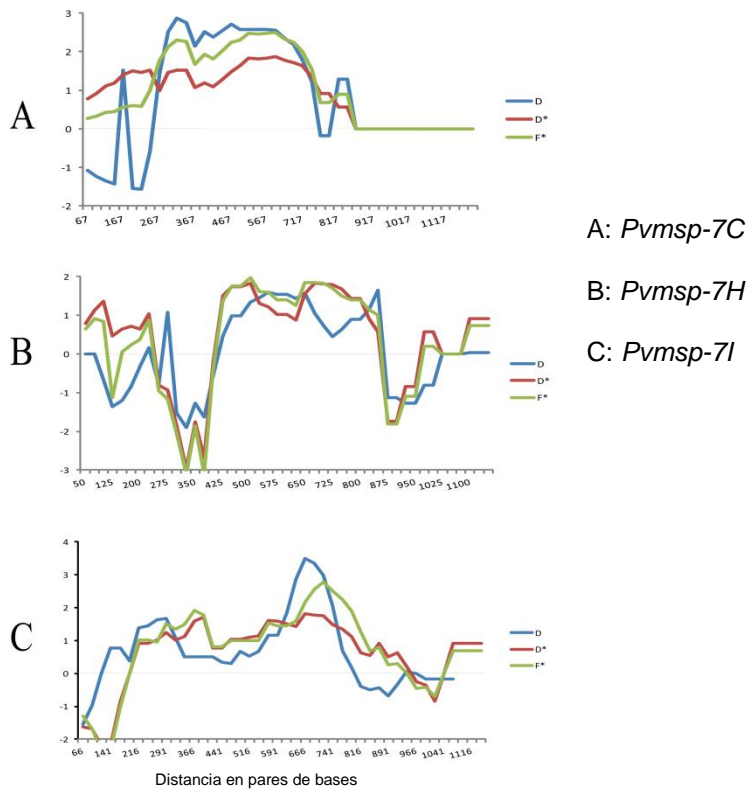
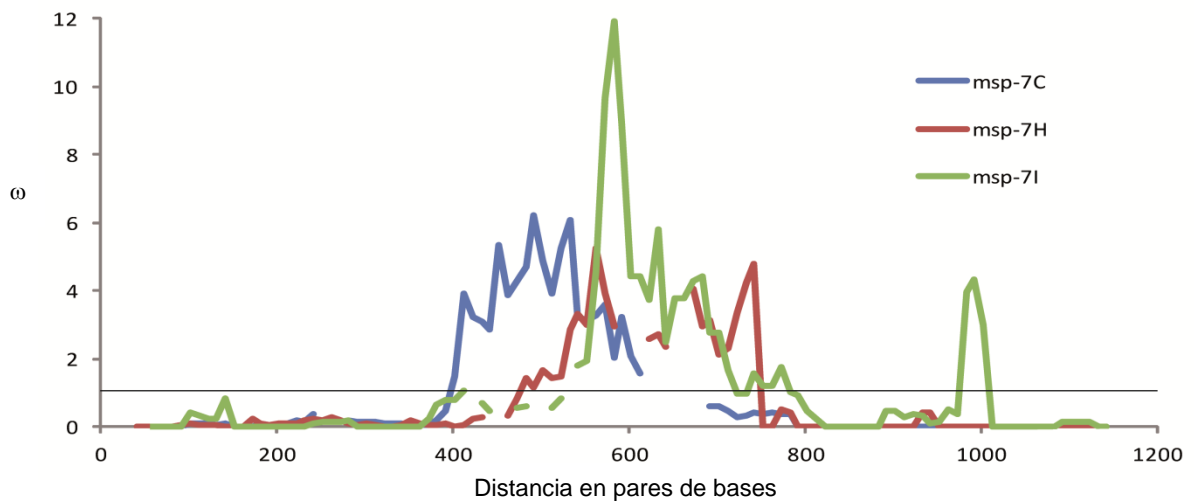
*: $p < 0,02$, **: $p < 0,05$.

Tabla 5. Número promedio de sustituciones sinónimas (d_S) y no sinónimas (d_N) para los genes *Pvm**msp-7*

	Extremo 5'		Región central		Extremo 3'		Longitud total del gen	
	d_S (es)	d_N (es)	d_S (es)	d_N (es)	d_S (es)	d_N (es)	d_S (es)	d_N (es)
<i>msp-7C</i>	0,068 (0,016)**	0,011 (0,004)	0,120 (0,026)	0,293 (0,033)•	0,004 (0,003)	0,000 (0,000)	0,051 (0,008)	0,063 (0,008)
<i>msp-7H</i>	0,055 (0,010)•	0,009 (0,003)	0,060 (0,017)	0,187 (0,024)•	0,009 (0,005)	0,007 (0,003)	0,037 (0,006)	0,040 (0,005)
<i>msp-7I</i>	0,036 (0,013)*	0,012 (0,005)	0,085 (0,020)	0,181 (0,018)**	0,017 (0,007)	0,010 (0,004)	0,045 (0,008)	0,054 (0,006)

*: $p < 0,04$, **: $p < 0,001$, •: $p < 0,0001$, ••: $p < 0,0004$. es: error de estimación

Figura 14. Ventana deslizante para los estadísticos D de Tajima, D* y F* de Fu & Li.

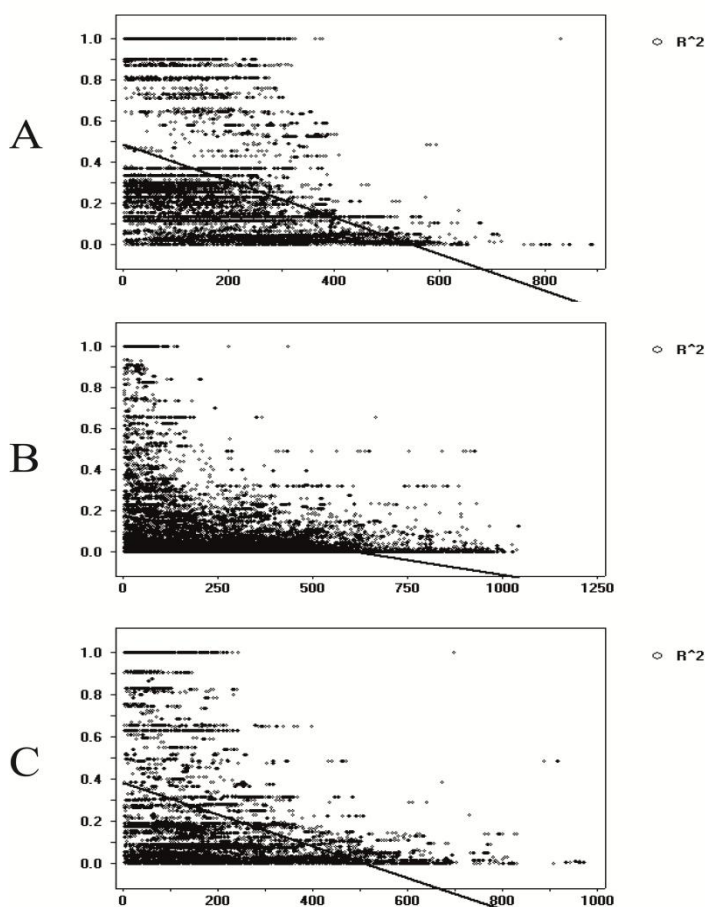
Figura 15. Ventana deslizante para ω (d_N/d_S)

La discontinuidad de las curvas es debida a la presencia de gaps dentro de los alineamientos.

7.3 Desequilibrio de ligamiento (LD) y recombinación

Asociaciones aleatorias fueron observadas entre los tipos alélicos de los genes *Pvmmsp-7*, sugiriendo su segregación independiente. El análisis del LD medido por el estadístico Z_{ns} (promedio de R^2) no mostró valores significativos para ninguno de los tres genes (tabla 4). La relación entre el LD con la distancia nucleotídica por una regresión lineal mostró que los valores de LD disminuyen conforme aumenta la distancia nucleotídica (figura 16). Por otra parte, el estadístico ZZ mostró valores estadísticamente significativos (tabla 4); ambos análisis sugieren que la recombinación intragénica ocurre dentro de los tres genes. Un análisis para detectar los eventos mínimos de recombinación (tabla 4) confirmó la presencia de recombinación dentro de cada gen y el programa RDP v3.4 fue usado para identificar las regiones recombinantes (con un $p < 0,03$), las cuales son mostradas en la figura 17.

Figura 16. Regresión lineal entre LD y la distancia nucleotídica.

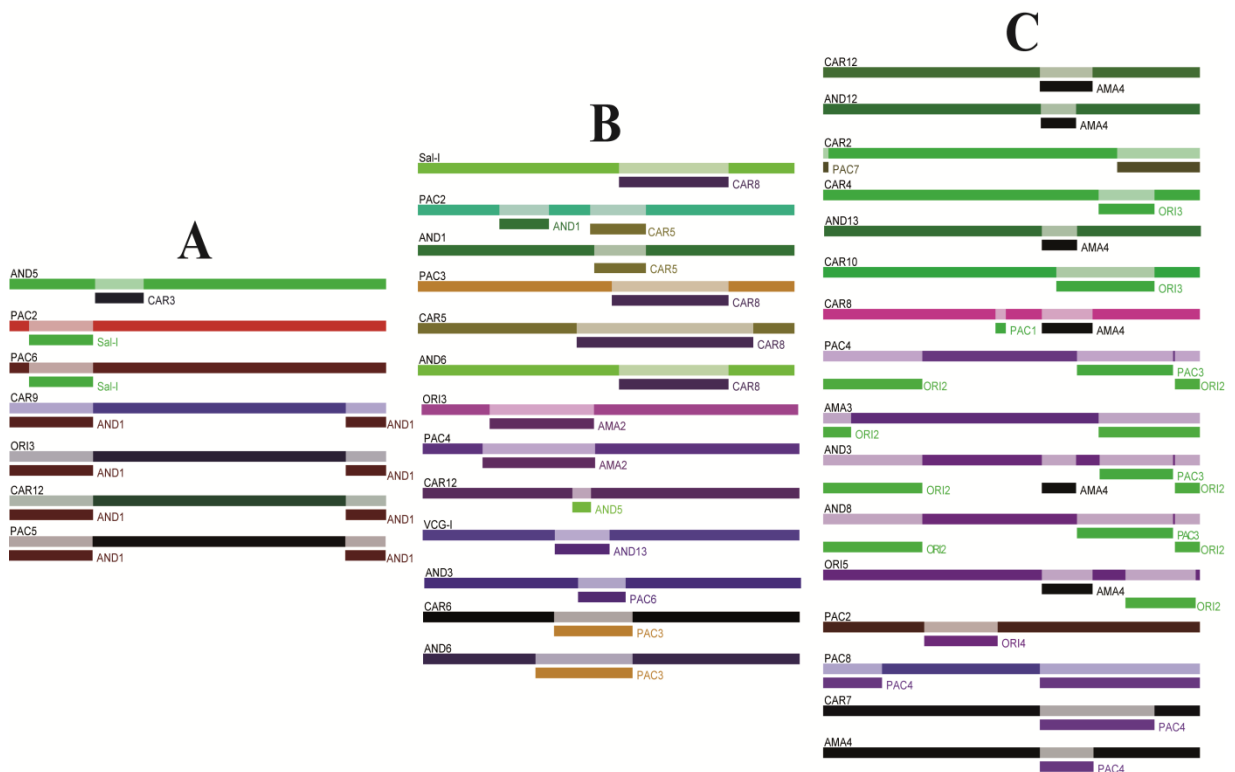


A: *Pvmmsp-7C*, B: *Pvmmsp-7H* y C: *Pvmmsp-7I*

7.4 Análisis de agrupamiento

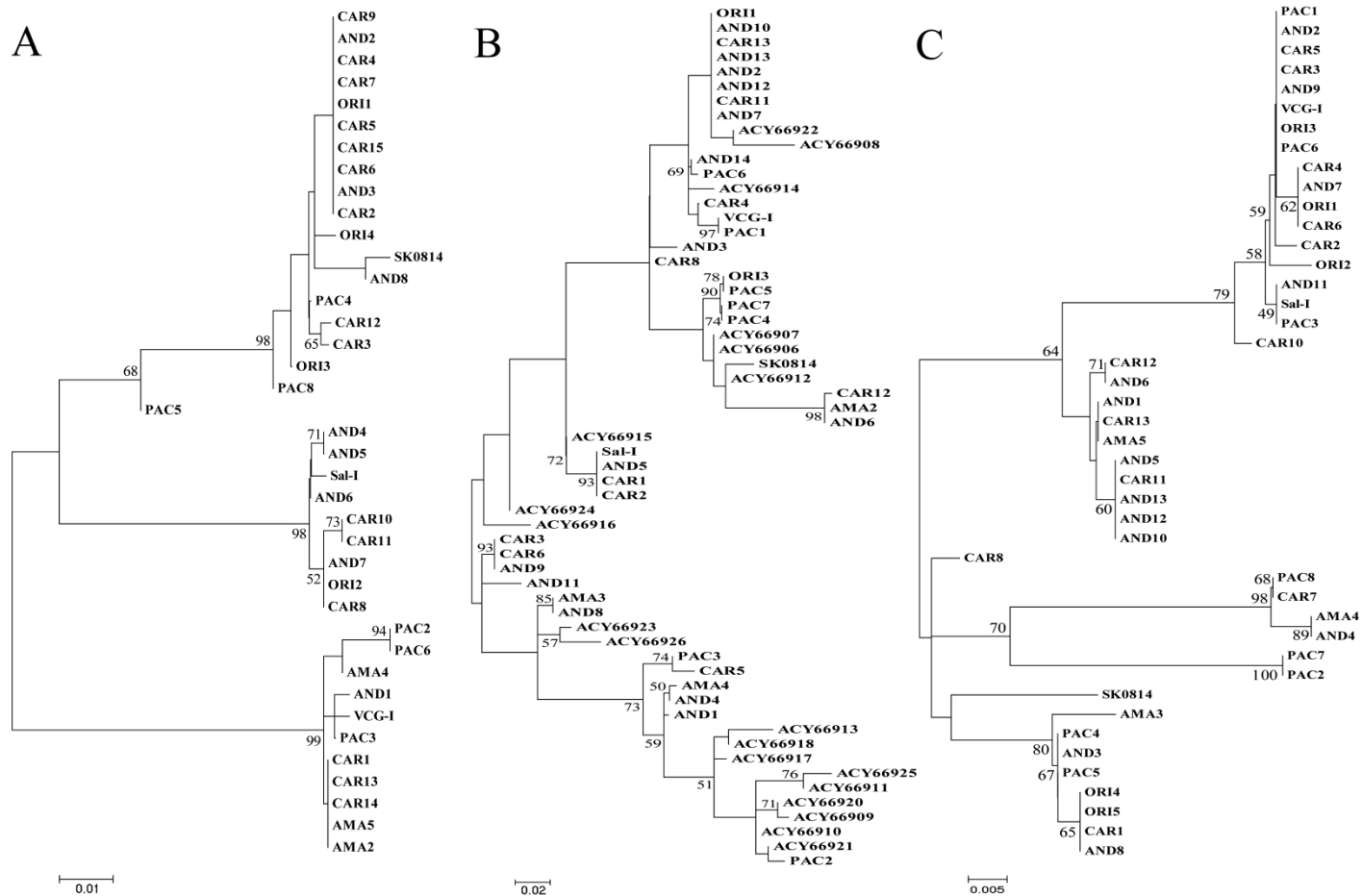
Las topologías inferidas por el método ML se muestran en la figura 18. En ésta, se observan tres clados para *Pvm*sp-7*C* (figura 18A) donde cada agrupación representa uno de los tres tipos alélicos observados. Para los otros genes, se observaron 4 clados, pero a diferencia de *Pvm*sp-7*C*, los tipos alélicos de *Pvm*sp-7*H* y *Pvm*sp-7*I* (figuras 18B y 18C) están agrupados aleatoriamente. Independientemente del método usado (ML o NJ), no se observó asociación geográfica entre los diferentes aislados e inclusive, las secuencias de referencia de la cepa Sal-I y secuencias no publicadas de aislados Coreanos, se agrupan junto con las secuencias Colombianas (figura 18). Sin embargo, estas topologías ignoran la recombinación. Cuando la recombinación fue tomada en cuenta, las diferentes topologías inferidas para cada fragmento recombinante no mostraron ninguna asociación geográfica (anexo 5).

Figura 17. Representación esquemática de los segmentos de recombinación identificados mediante RDP3 v.3.

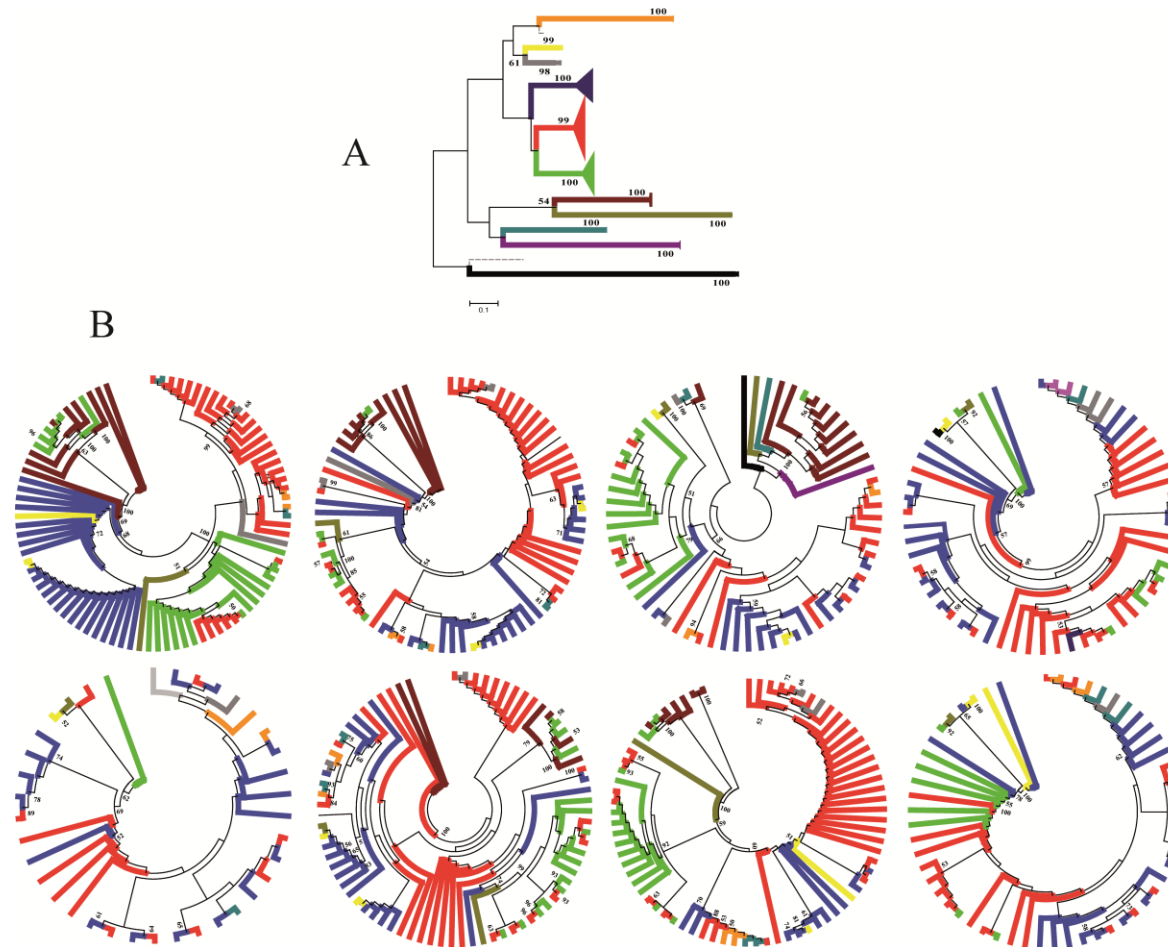


Sólo eventos de recombinación con una $p < 0,03$ fueron tenidos en cuenta. A: *Pvm*sp-7*C*, B: *Pvm*sp-7*H* y C: *Pvm*sp-7*I*.

Por otra parte, todas las secuencias de los diferentes miembros de *Pvm*sp-7 disponibles en *GenBank* fueron alineadas y posteriormente se construyeron árboles usando los métodos NJ y ML con el modelo TN93+G. El árbol muestra varios grupos monofiléticos, los cuales agrupan a cada uno de los miembros de la familia por separado (figura 19A). Esta topología se asemeja a la previamente reportada para esta familia [35, 42]. Cuando fue evaluada la recombinación dentro de la familia, varios fragmentos recombinantes entre estos genes fueron observados en los extremos 5' y 3', pero no en la región central de los genes (figura 20). Teniendo en cuenta que la recombinación parece estar ocurriendo entre los miembros de *m*sp-7, nuevas filogenias fueron inferidas, mostrando que los genes *Pvm*sp-7 son agrupados ahora en clados parafiléticos (figura 19B).

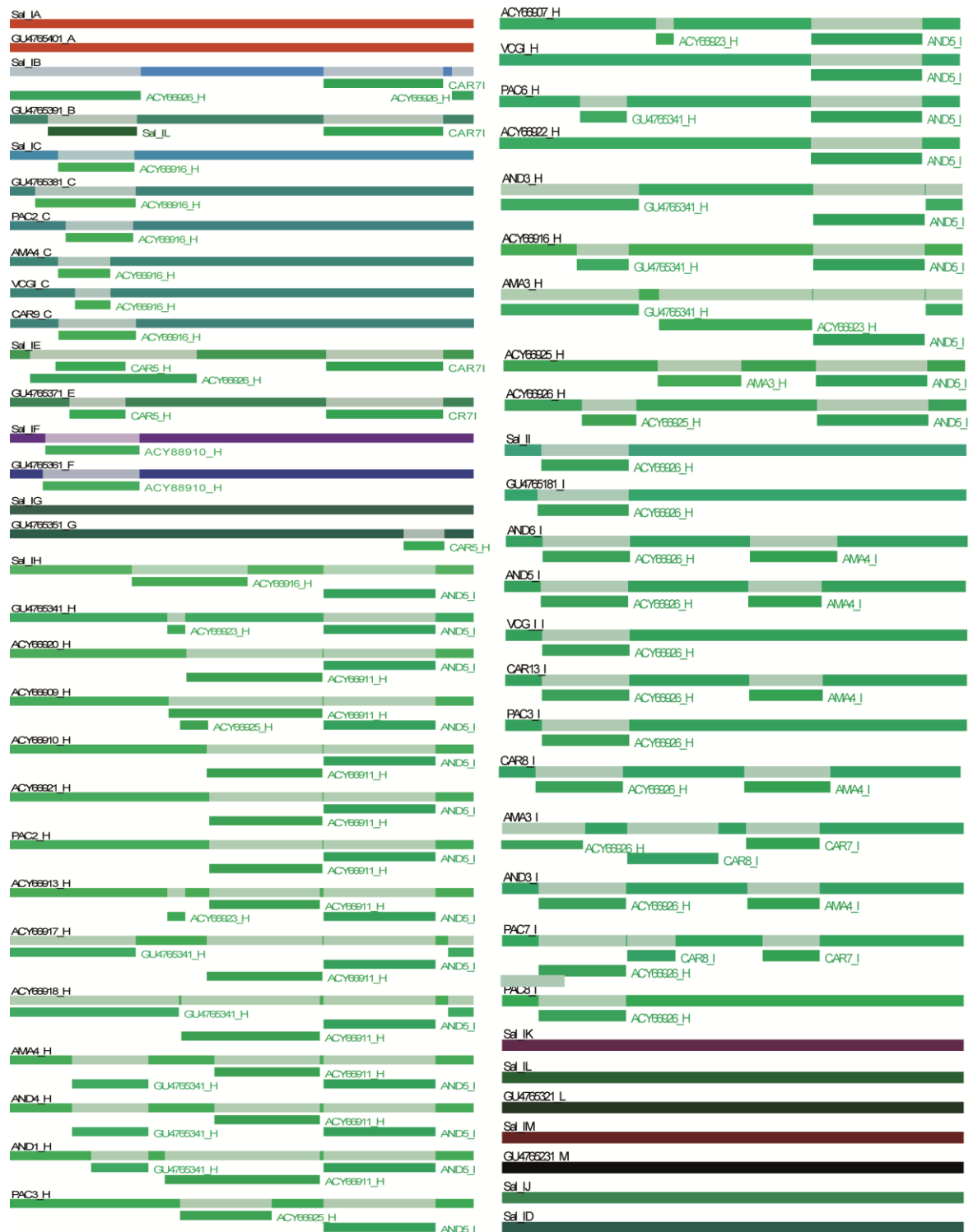
Figura 18. Filogenias inferidas por ML para *Pvmsp-7* (A), *Pvmsp-7H* (B) y *Pvmsp-7I* (C).

Los arboles inferidos por ML y NJ muestran topologías similares, por lo tanto se muestra sólo la topología del método ML. Los números sobre las ramas corresponden a los valores de Bootstrap.

Figura 19. Filogenias inferidas por ML para la familia *msp7* de *P. vivax*

A: Relaciones filogenéticas de la familia *msp-7* de *P. vivax* ignorando la recombinación. B: Árboles que describen las consecuencias filogenéticas de la recombinación. Por cada región recombinante una topología diferente es generada. *msp-7A*: verde oscuro, *msp-7B*: amarillo, *msp-7C*: azul, *msp-7E*: gris, *msp-7F*: verde azulado, *msp-7G*: naranja, *msp-7H*: rojo, *msp-7I*: verde, *msp-7K*: café, *msp-7L*: morado, *msp-7M*: negro, *msp-7D* y *7J*: gris punteado.

Figura 20. Representación esquemática de los segmentos de recombinación identificados mediante RDP3 v.3.



Sólo eventos de recombinación con una $p < 0,03$ fueron tenidos en cuenta.

8 DISCUSIÓN

Once diferentes MSPs han sido descritas en *P. falciparum* y al menos nueve genes ortólogos ha sido reportados en *P. vivax*. En esta última especie, los genes *msh-3* y *msh-7* se han expandido, dando origen a familias multigénicas [35, 93]. De particular interés, la familia *msh-7* presenta 11 genes funcionales [35] de los cuales 8 son transcritos durante el estadio de esquizonte en *P. vivax* [37]. Previamente se demostró que MSP-7 (MSP-7I de *P. falciparum*) hace parte del principal complejo proteico involucrado en la interacción inicial con la célula hospedera [33, 39]. Debido a que diferentes proteínas de la familia han sido localizadas sobre la superficie del merozoito [17, 36, 38, 39] y se ha visto que algunas son blanco para el sistema inmune, podrían ser tenidas en cuenta para el desarrollo de una vacuna.

Además de su reconocimiento por el sistema inmune, los candidatos a vacuna deben carecer o al menos tener una baja variabilidad genética. Los resultados descritos en este estudio muestran una marcada diferencia en el polimorfismo observado entre los miembros de la familia *msh-7* de las dos principales especies que causan malaria en el ser humano. Mientras los genes *msh-7* de *P. falciparum* muestran un bajo polimorfismo [43, 44], los genes *msh-7C*, *msh-7H* y *msh-7I* de *P. vivax* tiene una alta variabilidad genética. Sin embargo, debe notarse que no todos los miembros de esta familia en *P. vivax* tienen el mismo comportamiento. *Pvmsh-7A* y *Pvmsh-7K*, similar a lo que ocurre con *Pfmsh-7A*, *Pfmsh-7B*, *Pfmsh-7C*, *Pfmsh-7D*, *Pfmsh-7E*, *Pfmsh-7F* and *Pfmsh-7H*, (números de acceso en *GenBank*: XM_001350038.1 - 44.1) muestran un bajo polimorfismo [43-45], sugiriendo que los miembros de *msh-7* podrían estar expuestos a diferentes presiones selectivas (como la ejercida por el sistema inmune del hospedero) o a diferentes restricciones funcionales/estructurales. Aunque PvMSP-7C, PvMSP-7H y PvMSP-7I no cumplen con la condición de un bajo polimorfismo, el potencial desarrollo de una vacuna podría enfocarse en el dominio presente en la región C-terminal (número de acceso en Pfam: PF12948). Este dominio presenta una baja (si no ausente) variabilidad genética y está altamente conservado entre los tres genes, por lo tanto es probable que similar a lo que pasa en *P. falciparum*, la región C-terminal podría estar involucrada en la interacción con los glóbulos rojos [23].

La variabilidad genética encontrada en *Pvm*sp-7C, *Pvm*sp-7H y *Pvm*sp-7I los coloca entre los genes de *P. vivax* más polimórficos descritos hasta la fecha. Cuando las secuencias Colombianas fueron comparadas con las secuencias de la cepa de referencia Sal-I (Números de acceso en *GenBank*: XM_001614082.1, XM_001614087.1, XM_001614088.1) y las de aislados Coreanos (Números de acceso en *GenBank* GU476538.1, GU476534.1, GU476518.1 y ACY66906-26), diferentes tipos alélicos fueron encontrados, por lo tanto la variabilidad genética de estos genes parece estar distribuida alrededor del mundo.

La selección balanceante parece estar jugando un rol importante en la evolución del gen *Pvm*sp-7C (valores significativos > 0 de los estadísticos D, D* y F*). Se ha sugerido que este tipo de selección aumenta la variabilidad dentro de las poblaciones [94] manteniendo los diferentes alelos a frecuencias intermedias. Sin embargo, las pruebas de Tajima y Fu & Li no sólo son influenciadas por la selección, sino también por procesos demográficos dentro de la población, los cuales puede alterar las frecuencias alélicas esperadas. Por lo tanto, los valores positivos de las pruebas podrían ser el resultado de una disminución de la población. Sin embargo, no se encontraron valores estadísticamente significativos para la prueba de Fu (estadístico F_s), descartando la posibilidad de tener una población impactada por deriva genética (mecanismo que puede disminuir la población). Adicionalmente a esto, los valores de Hd y π reportados para *Pvm*sp-7C sugieren que la población tiene un tamaño efectivo grande, por lo tanto las frecuencias alélicas no parecen estar influenciadas por procesos demográficos.

*Pvm*sp-7H y *Pvm*sp-7I (a diferencia de *Pvm*sp-7C), parecen estar bajo el modelo neutral de evolución molecular, debido a que no se encontraron valores significativos en las pruebas de Tajima y Fu & Li, y a que no se encontraron diferencias entre las tasas d_N y d_S . Por consiguiente, se espera que las regiones sin restricción estructural/funcional tengan un alto polimorfismo [95]. Sin embargo, desviaciones de la neutralidad son detectadas sólo si el valor promedio del total del gen es significativamente mayor o menor a 0. Un comportamiento similar tiene el método de Nei-Gojobori, la selección positiva (o negativa) puede ser detectada sólo si el promedio d_N (o d_S) de todo el gen es significativamente más alto. Diferentes presiones selectivas podrían estar actuando al mismo tiempo en una proteína, de esta manera, algunos de los residuos de la proteína estarían bajo restricciones funcionales/estructurales y la selección positiva actuaría sólo

en los residuos restantes [67, 68, 70, 96]. Los análisis de ventana deslizante (para las pruebas de neutralidad y para ω) realizados, sugieren que diferentes presiones selectivas están actuando a lo largo de los tres genes. Las pruebas mostraron señales de selección purificante (valores negativos de las pruebas de neutralidad y $\omega > 1$) en los extremos 5' y 3', mientras en la región central de los tres genes, la selección balanceante o la selección positiva (valores positivos de las pruebas de neutralidad y $\omega > 1$) parecen estar actuando, así, a pesar de tener evidencia de selección balanceante (*Pvmmsp-7C*) o de neutralidad (*Pvmmsp-7H* y *Pvmmsp-7I*), la selección natural podría estar variando entre regiones y entre codones. Para comprobar esta hipótesis, dos diferentes pruebas fueron realizadas. La primera, estimo la tasa d_N y d_S en el extremo 5', la región central y extremo 3' de cada gen. La región central parece estar bajo selección positiva, mientras las regiones 5' y 3' parecen estar bajo selección purificante o negativa (d_S más alto que d_N). El segundo enfoque arrojó resultados similares, al determinar codones seleccionados positiva o negativamente usando métodos basados en máxima verosimilitud. Para *Pvmmsp-7C*, la región central presentó codones seleccionados positivamente, mientras a lo largo del gen, existen sitios seleccionados negativamente. Codones seleccionados positiva y negativamente fueron encontrados a lo largo de los genes *Pvmmsp-7H* y *Pvmmsp-7I*, pero los sitios bajo selección negativa se presentaron preferentemente en los extremos 5' y 3', los cuales están relativamente bien conservados en los genes *Pvmmsp-7*, mientras los codones bajo selección positiva se localizaron sobre la región central de ambos genes. Por lo tanto, el bajo polimorfismo y la presencia de una alta tasa d_S , sumados a la presencia de sitios seleccionados negativamente en el extremo 3', pueden ser el resultado de restricciones funcionales y de la selección purificante, posiblemente debido a que el dominio funcional (numero de acceso en Pfam PF12948) se encuentra dentro de esta región. Por consiguiente, la región central (la más polimórfica) podría estar bajo una presión selectiva ejercida por el sistema inmune y los codones seleccionados positivamente aumentan la variabilidad genética, generando diferentes alelos como un mecanismo de evasión.

Además de presiones selectivas y restricciones funcionales, la evolución de los antígenos del parásito puede verse afectada por la recombinación [30, 31, 67, 69, 71]. Varios estadísticos y algoritmos mostraron que la recombinación intragénica juega un rol importante en la generación de nuevas variantes alélicas dentro de los genes *mSP-7*. Así como la recombinación es importante durante la evolución de los genes individuales, las

familias multigénicas pueden verse afectadas por este mecanismo, al intercambiar fragmentos génicos entre genes parálogos y contribuir a la variabilidad genética o a la homogenización de la familia por un evento de entrecruzamiento desigual o por conversión génica [97]. El programa RDP v.3 y los posteriores análisis filogenéticos mostraron que los genes *mSP-7* no evolucionan independientemente; varias regiones de recombinación en 5' y 3' (pero no en la región central) fueron detectadas ($p < 0.03$) entre miembros de esta familia, sugiriendo que la familia multigénica *mSP-7* evoluciona mediante un tipo de conversión génica (posiblemente conversión sesgada), la cual homogeniza los extremos 5' y 3', pero no está involucrada en la generación de nuevas variantes (debido a que no hay evidencia de recombinación entre genes parálogos en la región central). Por lo tanto, la recombinación parece afectar, no solo la evolución de cada gen (en este caso *PvmSP-C*, *PvmSP-7H* y *PvmSP-7I*), sino también la evolución de la familia *mSP-7*.

Los análisis filogenéticos mostraron que con respecto al origen de cada aislado (provenientes de las diferentes regiones de Colombia), estos se agrupan sin tener ningún tipo de asociación geográfica; este comportamiento indica que no existe un aislamiento poblacional, lo cual podría atribuirse a un constante flujo génico de las poblaciones parasitarias. Cuando las secuencias de otros orígenes geográficos (Salvador y Corea) fueron incluidas para las inferencias filogenéticas, tampoco se encontró una asociación geográfica, lo cual podría ser el resultado de un polimorfismo ancestral o una convergencia génica, comportamiento que ha sido reportado para otros genes de *P. vivax* [30, 31, 67].

Estudios previos han mostrado el potencial rol de MSP-7 en la invasión a los glóbulos rojos [23, 36, 40, 62], así como sus características antigénicas [41, 46]; siguiendo la reglas para el desarrollo de vacunas basadas en subunidades [98-100], el dominio conservado de MSP-7 podría ser un atractivo candidato a evaluar durante el desarrollo de una vacuna antimalárica.

9 CONCLUSIONES

La caracterización de la variabilidad genética de tres miembros de la familia *msp-7* de *P. vivax* permitió establecer diferencias marcadas entre los miembros de esta familia, en las dos principales especies de parásito causantes de malaria en el ser humano; mientras en *P. falciparum* esta familia presenta un bajo polimorfismo, en *P. vivax* los genes *msp-7C*, *msp-7H* y *msp-7I* mostraron tener una alta variabilidad. Sin embargo, este polimorfismo no está presente en todos los miembros de *msp-7* de *P. vivax*. Por lo tanto, deben existir presiones selectivas o restricciones funcionales/estructurales diferentes para cada miembro de la familia.

La variabilidad genética de estos tres genes los coloca entre los más polimórficos descritos hasta la fecha. Esta variabilidad es generada a través de mutaciones puntuales, regiones repetitivas y recombinación. Estos mecanismos han permitido la aparición de al menos 3 tipos alélicos, posiblemente como un mecanismo de evasión al sistema inmune. Esta variabilidad no está distribuida a lo largo de toda la molécula, se puede distinguir una región central muy variable, mientras los extremos 5' y 3' son conservados, no solo dentro de cada gen, sino también entre genes.

La evolución de estos tres antígenos está dirigida por diferentes presiones selectivas a lo largo de la molécula. Los extremos 5' y 3' de cada gen, parecen estar bajo una presión selectiva negativa, posiblemente debido a que el extremo 3' codifica para el dominio funcional; mientras la región central, está sometida a una presión selectiva positiva, lo que aumenta la variabilidad.

Los análisis filogenéticos realizados mostraron que los aislados se agrupan independientemente de su origen geográfico, sugiriendo que no hay una subdivisión poblacional, por lo tanto, es posible que exista un flujo génico entre las poblaciones parasitarias en Colombia. Por otra parte, cuando las secuencias de diferentes lugares del mundo son analizadas, tampoco se observa una asociación geográfica. Esto podría ser explicado si el polimorfismo encontrado es ancestral y apareció antes de la dispersión del

parásito, o si las poblaciones han evolucionado por un fenómeno de convergencia, dando lugar a soluciones moleculares similares.

Aunque *Pvmsp-7C*, *Pvmsp-7H* y *Pvmsp-7I* están entre los antígenos más polimórficos descritos, la región C-terminal de estas proteínas es conservada y probablemente sea una región funcional, la cual puede ser tenida en cuenta para su evaluación durante el desarrollo de una vacuna antimalárica, siguiendo las reglas previamente establecidas para el desarrollo de vacunas basadas en subunidades químicamente sintetizadas.

10 PERSPECTIVAS

Al menos 8 genes de la familia *m*sp-7 de *P. vivax* son transcritos en el estadio de esquizonte, 2 de ellos (*m*sp-7A y *m*sp-7K) presentan un bajo polimorfismo, mientras otros 3 (*m*sp-7C, *m*sp-7H y *m*sp-7I) tienen una alta variabilidad genética. Aun no se conoce la estructura genética de los 3 miembros restantes, y debido a la redundancia funcional de la familia, el conocimiento de la estructura genética de la familia completa podría ser útil durante el desarrollo de una vacuna antimalárica que tenga como blanco bloquear el reconocimiento inicial del parásito a la célula hospedera.

A la fecha, sólo se tiene evidencia de la localización de *m*sp-7K sobre la membrana del parásito, motivo por el cual valdría la pena realizar ensayos inmunoquímicos para establecer la localización de los otros miembros de esta familia.

Los candidatos a vacuna deben cumplir tres criterios: (1) ser reconocidos por el sistema inmune, (2) tener una baja variabilidad genética o poseer dominios funcionales conservados y (3) inducir protección después de la vacunación. Los dos primeros criterios son soportados por previos estudios, así como por los resultados presentados aquí. Estudios futuros tendientes a evaluar la antigenicidad e inmunogenicidad de estos antígenos en modelos animales, así como el efecto protector conferido por los anticuerpos generados frente a ellos, son el siguiente paso a seguir.

11 BIBLIOGRAFÍA

1. de Souza-Neiras WC, de Melo LM, Machado RL: **The genetic diversity of *Plasmodium vivax*--a review.** *Memorias do Instituto Oswaldo Cruz* 2007, **102**(3):245-254.
2. Mendis K, Sina BJ, Marchesini P, Carter R: **The neglected burden of *Plasmodium vivax* malaria.** *The American journal of tropical medicine and hygiene* 2001, **64**(1-2 Suppl):97-106.
3. Escalante AA, Ayala FJ: **Phylogeny of the malarial genus *Plasmodium*, derived from rRNA gene sequences.** *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**(24):11373-11377.
4. Rich SM, Ayala FJ: **Progress in malaria research: the case for phylogenetics.** *Advances in parasitology* 2003, **54**:255-280.
5. Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, Temperley WH, Kabaria CW, Tatem AJ, Manh BH, Elyazar IR *et al*: **The international limits and population at risk of *Plasmodium vivax* transmission in 2009.** *PLoS Negl Trop Dis* 2010, **4**(8):e774.
6. Lysenko AJ, Semashko IN: **Medical geography: a medical-geographical study of an ancient disease.** WHO 1968, **Publisher's Code: 911.3:616.936(100).**
7. Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, Warikar N, Sugiarto P, Tjitra E, Anstey NM, Price RN: ***Vivax* malaria: a major cause of morbidity in early infancy.** *Clin Infect Dis* 2009, **48**(12):1704-1712.
8. Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, Karyana M, Lampah DA, Price RN: **Multidrug-resistant *Plasmodium vivax* associated with severe and fatal malaria: a prospective study in Papua, Indonesia.** *PLoS Med* 2008, **5**(6):e128.
9. Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, Anstey NM: ***Vivax* malaria: neglected and not benign.** *The American journal of tropical medicine and hygiene* 2007, **77**(6 Suppl):79-87.
10. Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, Garg S, Kochar A, Khatri MP, Gupta V: **Severe *Plasmodium vivax* malaria: a report on serial cases from Bikaner in northwestern India.** *The American journal of tropical medicine and hygiene* 2009, **80**(2):194-198.
11. Carvalho LJ, Daniel-Ribeiro CT, Goto H: **Malaria vaccine: candidate antigens, mechanisms, constraints and prospects.** *Scand J Immunol* 2002, **56**(4):327-343.
12. Galinski MR, Barnwell JW: ***Plasmodium vivax*: who cares?** *Malar J* 2008, **7** Suppl 1:S9.
13. Angel DI, Mongui A, Ardila J, Vanegas M, Patarroyo MA: **The *Plasmodium vivax* Pv41 surface protein: identification and characterization.** *Biochemical and biophysical research communications* 2008, **377**(4):1113-1117.
14. Mongui A, Angel DI, Gallego G, Reyes C, Martinez P, Guhl F, Patarroyo MA: **Characterization and antigenicity of the promising vaccine candidate *Plasmodium vivax* 34kDa rhoptry antigen (Pv34).** *Vaccine* 2009, **28**(2):415-421.
15. Mongui A, Angel DI, Moreno-Perez DA, Villarreal-Gonzalez S, Almonacid H, Vanegas M, Patarroyo MA: **Identification and characterization of the *Plasmodium vivax* thrombospondin-related apical merozoite protein.** *Malar J* 2010, **9**:283.

16. Mongui A, Perez-Leal O, Rojas-Caraballo J, Angel DI, Cortes J, Patarroyo MA: **Identifying and characterising the Plasmodium falciparum RhopH3 Plasmodium vivax homologue.** *Biochemical and biophysical research communications* 2007, **358**(3):861-866.
17. Mongui A, Perez-Leal O, Soto SC, Cortes J, Patarroyo MA: **Cloning, expression, and characterisation of a Plasmodium vivax MSP7 family merozoite surface protein.** *Biochemical and biophysical research communications* 2006, **351**(3):639-644.
18. Moreno-Perez DA, Mongui A, Soler LN, Sanchez-Ladino M, Patarroyo MA: **Identifying and characterizing a member of the RhopH1/Clag family in Plasmodium vivax.** *Gene* 2011, **481**(1):17-23.
19. Perez-Leal O, Mongui A, Cortes J, Yepes G, Leiton J, Patarroyo MA: **The Plasmodium vivax rhoptry-associated protein 1.** *Biochemical and biophysical research communications* 2006, **341**(4):1053-1058.
20. Perez-Leal O, Sierra AY, Barrero CA, Moncada C, Martinez P, Cortes J, Lopez Y, Salazar LM, Hoebeke J, Patarroyo MA: **Identifying and characterising the Plasmodium falciparum merozoite surface protein 10 Plasmodium vivax homologue.** *Biochemical and biophysical research communications* 2005, **331**(4):1178-1184.
21. Perez-Leal O, Sierra AY, Barrero CA, Moncada C, Martinez P, Cortes J, Lopez Y, Torres E, Salazar LM, Patarroyo MA: **Plasmodium vivax merozoite surface protein 8 cloning, expression, and characterisation.** *Biochemical and biophysical research communications* 2004, **324**(4):1393-1399.
22. O'Donnell RA, de Koning-Ward TF, Burt RA, Bockarie M, Reeder JC, Cowman AF, Crabb BS: **Antibodies against merozoite surface protein (MSP)-1(19) are a major component of the invasion-inhibitory response in individuals immune to malaria.** *The Journal of experimental medicine* 2001, **193**(12):1403-1412.
23. Garcia Y, Puentes A, Curtidor H, Cifuentes G, Reyes C, Barreto J, Moreno A, Patarroyo ME: **Identifying merozoite surface protein 4 and merozoite surface protein 7 Plasmodium falciparum protein family members specifically binding to human erythrocytes suggests a new malarial parasite-redundant survival mechanism.** *J Med Chem* 2007, **50**(23):5665-5675.
24. Ocampo M, Urquiza M, Guzman F, Rodriguez LE, Suarez J, Curtidor H, Rosas J, Diaz M, Patarroyo ME: **Two MSA 2 peptides that bind to human red blood cells are relevant to Plasmodium falciparum merozoite invasion.** *J Pept Res* 2000, **55**(3):216-223.
25. Puentes A, Garcia J, Ocampo M, Rodriguez L, Vera R, Curtidor H, Lopez R, Suarez J, Valbuena J, Vanegas M *et al*: **P. falciparum: merozoite surface protein-8 peptides bind specifically to human erythrocytes.** *Peptides* 2003, **24**(7):1015-1023.
26. Puentes A, Ocampo M, Rodriguez LE, Vera R, Valbuena J, Curtidor H, Garcia J, Lopez R, Tovar D, Cortes J *et al*: **Identifying Plasmodium falciparum merozoite surface protein-10 human erythrocyte specific binding regions.** *Biochimie* 2005, **87**(5):461-472.
27. Genton B, Reed ZH: **Asexual blood-stage malaria vaccine development: facing the challenges.** *Current opinion in infectious diseases* 2007, **20**(5):467-475.
28. Takala SL, Plowe CV: **Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming 'vaccine resistant malaria'.** *Parasite immunology* 2009, **31**(9):560-573.

29. Figtree M, Pasay CJ, Slade R, Cheng Q, Cloonan N, Walker J, Saul A: **Plasmodium vivax synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes.** *Molecular and biochemical parasitology* 2000, **108**(1):53-66.
30. Gomez A, Suarez CF, Martinez P, Saravia C, Patarroyo MA: **High polymorphism in Plasmodium vivax merozoite surface protein-5 (MSP5).** *Parasitology* 2006, **133**(Pt 6):661-672.
31. Putaporntip C, Udomsangpetch R, Pattanawong U, Cui L, Jongwutiwes S: **Genetic diversity of the Plasmodium vivax merozoite surface protein-5 locus from diverse geographic origins.** *Gene* 2010, **456**(1-2):24-35.
32. Richie TL, Saul A: **Progress and challenges for malaria vaccines.** *Nature* 2002, **415**(6872):694-701.
33. Pachebat JA, Ling IT, Grainger M, Trucco C, Howell S, Fernandez-Reyes D, Gunaratne R, Holder AA: **The 22 kDa component of the protein complex on the surface of Plasmodium falciparum merozoites is derived from a larger precursor, merozoite surface protein 7.** *Molecular and biochemical parasitology* 2001, **117**(1):83-89.
34. Trucco C, Fernandez-Reyes D, Howell S, Stafford WH, Scott-Finnigan TJ, Grainger M, Ogun SA, Taylor WR, Holder AA: **The merozoite surface protein 6 gene codes for a 36 kDa protein associated with the Plasmodium falciparum merozoite surface protein-1 complex.** *Molecular and biochemical parasitology* 2001, **112**(1):91-101.
35. Garzon-Ospina D, Cadavid LF, Patarroyo MA: **Differential expansion of the merozoite surface protein (msp)-7 gene family in Plasmodium species under a birth-and-death model of evolution.** *Molecular phylogenetics and evolution* 2010, **55**(2):399-408.
36. Tewari R, Ogun SA, Gunaratne RS, Crisanti A, Holder AA: **Disruption of Plasmodium berghei merozoite surface protein 7 gene modulates parasite growth in vivo.** *Blood* 2005, **105**(1):394-396.
37. Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, Ginsburg H, Nosten F, Day NP, White NJ *et al*: **The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites.** *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(42):16290-16295.
38. Mello K, Daly TM, Long CA, Burns JM, Bergman LW: **Members of the merozoite surface protein 7 family with similar expression patterns differ in ability to protect against Plasmodium yoelii malaria.** *Infection and immunity* 2004, **72**(2):1010-1018.
39. Mello K, Daly TM, Morrisey J, Vaidya AB, Long CA, Bergman LW: **A multigene family that interacts with the amino terminus of plasmodium MSP-1 identified using the yeast two-hybrid system.** *Eukaryotic cell* 2002, **1**(6):915-925.
40. Kadekoppala M, O'Donnell RA, Grainger M, Crabb BS, Holder AA: **Deletion of the Plasmodium falciparum merozoite surface protein 7 gene impairs parasite invasion of erythrocytes.** *Eukaryotic cell* 2008, **7**(12):2123-2132.
41. Wang L, Crouch L, Richie TL, Nhan DH, Coppel RL: **Naturally acquired antibody responses to the components of the Plasmodium falciparum merozoite surface protein 1 complex.** *Parasite immunology* 2003, **25**(8-9):403-412.
42. Kadekoppala M, Holder AA: **Merozoite surface proteins of the malaria parasite: the MSP1 complex and the MSP7 family.** *Int J Parasitol* 2010, **40**(10):1155-1161.

43. Tetteh KK, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, Marsh K, Weedall GD, Conway DJ: **Prospective identification of malaria parasite genes under balancing selection.** *PloS one* 2009, **4**(5):e5568.
44. Roy SW, Weedall GD, da Silva RL, Polley SD, Ferreira MU: **Sequence diversity and evolutionary dynamics of the dimorphic antigen merozoite surface protein-6 and other Msp genes of Plasmodium falciparum.** *Gene* 2009, **443**(1-2):12-21.
45. Garzon-Ospina D, Romero-Murillo L, Tobon LF, Patarroyo MA: **Low genetic polymorphism of merozoite surface proteins 7 and 10 in Colombian Plasmodium vivax isolates.** *Infect Genet Evol* 2011, **11**(2):528-531.
46. Chen JH, Jung JW, Wang Y, Ha KS, Lu F, Lim CS, Takeo S, Tsuboi T, Han ET: **Immunoproteomics profiling of blood stage Plasmodium vivax infection by high-throughput screening assays.** *J Proteome Res* 2010, **9**(12):6479-6489.
47. WHO (ed.): **World malaria report : 2010.** Geneva: WHO; 2010.
48. Suh KN, Kain KC, Keystone JS: **Malaria.** *Cmaj* 2004, **170**(11):1693-1702.
49. WHO (ed.): **Malaria elimination: a field manual for low and moderate endemic countries.** Geneva: WHO; 2007.
50. Carter R, Mendis KN: **Evolutionary and historical aspects of the burden of malaria.** *Clinical microbiology reviews* 2002, **15**(4):564-594.
51. Sachs J, Malaney P: **The economic and social burden of malaria.** *Nature* 2002, **415**(6872):680-685.
52. **Boletín vigilancia de la malaria en Colombia 2-2012** [http://www2.paho.org/COL/index.php?option=com_docman&task=doc_download&gid=1175&Itemid=]
53. **World Malaria Report 2010** [http://www.who.int/malaria/publications/country-profiles/profile_col_en.pdf]
54. Polley SD, Tetteh KK, Lloyd JM, Akpogheneta OJ, Greenwood BM, Bojang KA, Conway DJ: **Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection.** *The Journal of infectious diseases* 2007, **195**(2):279-287.
55. White NJ: **Plasmodium knowlesi: the fifth human malaria parasite.** *Clin Infect Dis* 2008, **46**(2):172-173.
56. Chitnis CE, Blackman MJ: **Host cell invasion by malaria parasites.** *Parasitology today (Personal ed)* 2000, **16**(10):411-415.
57. Cowman AF, Crabb BS: **Invasion of red blood cells by malaria parasites.** *Cell* 2006, **124**(4):755-766.
58. Ferreira MU, da Silva Nunes M, Wunderlich G: **Antigenic diversity and immune evasion by malaria parasites.** *Clinical and diagnostic laboratory immunology* 2004, **11**(6):987-995.
59. Cowman AF, Baldi DL, Duraisingh M, Healer J, Mills KE, O'Donnell RA, Thompson J, Triglia T, Wickham ME, Crabb BS: **Functional analysis of Plasmodium falciparum merozoite antigens: implications for erythrocyte invasion and vaccine development.** *Philosophical transactions of the Royal Society of London* 2002, **357**(1417):25-33.
60. Kauth CW, Woehlbier U, Kern M, Mekonnen Z, Lutz R, Mucke N, Langowski J, Bujard H: **Interactions between merozoite surface proteins 1, 6, and 7 of the malaria parasite Plasmodium falciparum.** *The Journal of biological chemistry* 2006, **281**(42):31517-31527.
61. Pachebat JA, Kadekoppala M, Grainger M, Dluzewski AR, Gunaratne RS, Scott-Finnigan TJ, Ogun SA, Ling IT, Bannister LH, Taylor HM *et al*: **Extensive proteolytic processing of the malaria parasite merozoite surface protein 7 during**

- biosynthesis and parasite release from erythrocytes.** *Molecular and biochemical parasitology* 2007, **151**(1):59-69.
62. Woehlbier U, Epp C, Hackett F, Blackman MJ, Bujard H: **Antibodies against multiple merozoite surface antigens of the human malaria parasite *Plasmodium falciparum* inhibit parasite maturation and red blood cell invasion.** *Malar J* 2010, **9**:77.
 63. Escalante AA, Lal AA, Ayala FJ: **Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*.** *Genetics* 1998, **149**(1):189-202.
 64. Good MF: **Towards a blood-stage vaccine for malaria: are we following all the leads?** *Nature reviews* 2001, **1**(2):117-125.
 65. Rojas-Caraballo J, Mongui A, Giraldo MA, Delgado G, Granados D, Millan-Cortes D, Martinez P, Rodriguez R, Patarroyo MA: **Immunogenicity and protection-inducing ability of recombinant *Plasmodium vivax* rhoptry-associated protein 2 in Aotus monkeys: a potential vaccine candidate.** *Vaccine* 2009, **27**(21):2870-2876.
 66. Garzon-Ospina D, Romero-Murillo L, Patarroyo MA: **Limited genetic polymorphism of the *Plasmodium vivax* low molecular weight rhoptry protein complex in the Colombian population.** *Infect Genet Evol* 2010, **10**(2):261-267.
 67. Martinez P, Suarez CF, Cardenas PP, Patarroyo MA: ***Plasmodium vivax* Duffy binding protein: a modular evolutionary proposal.** *Parasitology* 2004, **128**(Pt 4):353-366.
 68. Martinez P, Suarez CF, Gomez A, Cardenas PP, Guerrero JE, Patarroyo MA: **High level of conservation in *Plasmodium vivax* merozoite surface protein 4 (PvMSP4).** *Infect Genet Evol* 2005, **5**(4):354-361.
 69. Mascorro CN, Zhao K, Khuntirat B, Sattabongkot J, Yan G, Escalante AA, Cui L: **Molecular evolution and intragenic recombination of the merozoite surface protein MSP-3alpha from the malaria parasite *Plasmodium vivax* in Thailand.** *Parasitology* 2005, **131**(Pt 1):25-35.
 70. Putaporntip C, Jongwutiwes S, Ferreira MU, Kanbara H, Udomsangpetch R, Cui L: **Limited global diversity of the *Plasmodium vivax* merozoite surface protein 4 gene.** *Infect Genet Evol* 2009, **9**(5):821-826.
 71. Putaporntip C, Jongwutiwes S, Seethamchai S, Kanbara H, Tanabe K: **Intragenic recombination in the 3' portion of the merozoite surface protein 1 gene of *Plasmodium vivax*.** *Molecular and biochemical parasitology* 2000, **109**(2):111-119.
 72. Cui L, Escalante AA, Imwong M, Snounou G: **The genetic diversity of *Plasmodium vivax* populations.** *Trends in parasitology* 2003, **19**(5):220-226.
 73. Gutierrez A, Vicini J, Patarroyo ME, Murillo LA, Patarroyo MA: ***Plasmodium vivax*: polymorphism in the merozoite surface protein 1 gene from wild Colombian isolates.** *Experimental parasitology* 2000, **95**(3):215-219.
 74. Pacheco MA, Ryan EM, Poe AC, Basco L, Udhayakumar V, Collins WE, Escalante AA: **Evidence for negative selection on the gene encoding rhoptry-associated protein 1 (RAP-1) in *Plasmodium* spp.** *Infect Genet Evol* 2010, **10**(5):655-661.
 75. Prajapati SK, Joshi H, Shalini S, Patarroyo MA, Suwanarusk R, Kumar A, Sharma SK, Eapen A, Dev V, Bhatt RM *et al*: ***Plasmodium vivax* lineages: geographical distribution, tandem repeat polymorphism, and phylogenetic relationship.** *Malar J* 2011, **10**:374.
 76. Imwong M, Pukrittayakamee S, Gruner AC, Renia L, Letourneur F, Looareesuwan S, White NJ, Snounou G: **Practical PCR genotyping protocols for *Plasmodium vivax* using Pvcs and Pvmsp1.** *Malar J* 2005, **4**(1):20.
 77. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,**

- position-specific gap penalties and weight matrix choice.** *Nucleic acids research* 1994, **22**(22):4673-4680.
78. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic acids research* 2006, **34**(Web Server issue):W609-612.
 79. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics (Oxford, England)* 2009, **25**(11):1451-1452.
 80. Tajima F: **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.** *Genetics* 1989, **123**(3):585-595.
 81. Fu YX: **Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.** *Genetics* 1997, **147**(2):915-925.
 82. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28**(10):2731-2739.
 83. Zhang J, Rosenberg HF, Nei M: **Positive Darwinian selection after gene duplication in primate ribonuclease genes.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(7):3708-3713.
 84. Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL: **Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.** *Bioinformatics (Oxford, England)* 2010, **26**(19):2455-2457.
 85. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ: **Adaptation to different human populations by HIV-1 revealed by codon-based analyses.** *PLoS Comput Biol* 2006, **2**(6):e62.
 86. Kosakovsky Pond SL, Frost SD: **Not so different after all: a comparison of methods for detecting amino acid sites under selection.** *Mol Biol Evol* 2005, **22**(5):1208-1222.
 87. Kelly JK: **A test of neutrality based on interlocus associations.** *Genetics* 1997, **146**(3):1197-1206.
 88. Rozas J, Gullaud M, Blandin G, Aguade M: **DNA variation at the rp49 gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure.** *Genetics* 2001, **158**(3):1147-1155.
 89. Hudson RR, Kaplan NL: **Statistical properties of the number of recombination events in the history of a sample of DNA sequences.** *Genetics* 1985, **111**(1):147-164.
 90. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P: **RDP3: a flexible and fast computer program for analyzing recombination.** *Bioinformatics (Oxford, England)* 2010, **26**(19):2462-2463.
 91. Thakur A, Alam MT, Bora H, Kaur P, Sharma YD: **Plasmodium vivax: sequence polymorphism and effect of natural selection at apical membrane antigen 1 (PvAMA1) among Indian population.** *Gene* 2008, **419**(1-2):35-42.
 92. Putaporntip C, Jongwutiwes S, Tia T, Ferreira MU, Kanbara H, Tanabe K: **Diversity in the thrombospondin-related adhesive protein gene (TRAP) of Plasmodium vivax.** *Gene* 2001, **268**(1-2):97-104.
 93. Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, Caler E, Crabtree J, Angiuoli SV, Merino EF, Amedeo P *et al*: **Comparative genomics of the neglected human malaria parasite Plasmodium vivax.** *Nature* 2008, **455**(7214):757-763.
 94. Nielsen R: **Molecular signatures of natural selection.** *Annu Rev Genet* 2005, **39**:197-218.

95. Kimura M: **The neutral theory of molecular evolution**. Cambridge Cambridgeshire ; New York: Cambridge University Press; 1983.
96. Suarez CF, Patarroyo ME, Trujillo E, Estupinan M, Baquero JE, Parra C, Rodriguez R: **Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages**. *Immunogenetics* 2006, **58**(7):542-558.
97. Nei M, Rooney AP: **Concerted and birth-and-death evolution of multigene families**. *Annu Rev Genet* 2005, **39**:121-152.
98. Cifuentes G, Bermudez A, Rodriguez R, Patarroyo MA, Patarroyo ME: **Shifting the polarity of some critical residues in malarial peptides' binding to host cells is a key factor in breaking conserved antigens' code of silence**. *Med Chem* 2008, **4**(3):278-292.
99. Patarroyo ME, Bermudez A, Patarroyo MA: **Structural and immunological principles leading to chemically synthesized, multiantigenic, multistage, minimal subunit-based vaccine development**. *Chemical reviews* 2011, **111**(5):3459-3507.
100. Patarroyo ME, Patarroyo MA: **Emerging rules for subunit-based, multiantigenic, multistage chemically synthesized vaccines**. *Acc Chem Res* 2008, **41**(3):377-386.

	340	*	360	*	380	*	
Sal-I	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVNLLNTIEVV : 387
AND6	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
ORI2	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
CAR8	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
AND7	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
CAR10	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
CAR11	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
AND4	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
AND5	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 373
PAC2	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
PAC6	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
AMA4	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 367
PAC3	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
AND1	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
CAR13	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
AMA5	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
CAR14	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
VCG-I	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
AMA2	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
CAR1	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 366
CAR9	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
AND8	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
ORI1	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
AND2	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
AND3	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR15	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR2	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR4	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR7	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR6	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR5	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
ORI3	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR12	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
CAR3	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
PAC4	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
PAC8	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
ORI4	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
PAC5	: AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN----- : 381
	AACLVNVFKKVL	DDEHLQKEFDNFVHG	FYGF	AKRHNYL	RG	MANENLYKDI	FKNVVN

	220	*	240	*	260	*	280	*	300	*	320																												
Sal-I	: APSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
PAC2	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	TNAGQ	PPAG	ST-G	PAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
AMA4	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	PNAGQ	PPAG	ST-G	PAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
AND4	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	PNAGQ	PPAG	ST-G	PAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
AND1	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	PNAGQ	PPAG	ST-G	PAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
PAC3	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	PNNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304
CAR5	: TSSSSQ	STGQS	QSSNGA	EPPAG	T	QEVN	PNAGQ	TS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 309
AND5	: APSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
CAR1	: APSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
CAR2	: APSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
ORI3	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
PAC7	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
PAC4	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 309	
PAC5	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 309	
AMA2	: TSSSGF	STGGS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAAPP	GGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND6	: TSSSGF	STGGS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAAPP	GGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
CAR12	: TSSSGF	STGGS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAAPP	GGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
VCG-I	: TSSSF	STGGS	QPSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
PAC1	: TSSSF	STGGS	QPSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
ORI1	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
CAR4	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND14	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	
PAC6	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND2	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND7	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND12	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
CAR13	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
CAR11	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND13	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
AND10	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 306	
CAR8	: TSSSSH	STGQP	QSSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 307	
AND3	: TSSGF	STGGS	QPSAGV	SLPTG	TAE	TVASNTA	QTS	PPAG	SPGG	QAATS	SGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 301	
AMA3	: TSSSQ	STGQS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAATS	GGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
AND8	: TSSSQ	STGQS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAATS	GGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 304	
AND9	: TSSSQ	STGGS	QPSAGV	ALPTG	TAE	TVASNTA	QTS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 308	
CAR3	: TSSSQ	STGGS	QPSAGV	ALPTG	TAE	TVASNTA	QTS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 308	
CAR6	: TSSSQ	STGGS	QPSAGV	ALPTG	TAE	TVASNTA	QTS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 308	
AND11	: TSSSQ	STGGS	QPSAGV	APSVG	NAE	AVVTN	TEHSS	PPAG	LPGG	QAAT	PGQPE	SRV	PNVKY	LDKLY	DEVLK	TTDAK	DE	IHV	PPF	H	SKY	NDFR	K	YE	E	F	T	M	N	E	R	E	Y	Q	I	V	K	: 303	

tsSsS STGq Q S g p Gt 2 V n q PPAG G Aat GQPESGRVpNVkYLDKLYDEVLKTTDAKleIHVPPFHsKYNDFRkKYEFTMNEREYQIVK

```

*           340           *           360           *           380           *           400
Sal-1      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
PAC2       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTNLYDKLLKNVVNLLNTIEVK : 383
AMA4       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDKLLKNVVNLLNTIEVK : 383
AND4       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 383
AND1       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDKLLKNVVNLLNTIEVK : 383
PAC3       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
CAR5       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 389
AND5       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
CAR1       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
CAR2       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
ORI3       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 386
PAC7       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 386
PAC4       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 389
PAC5       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 389
AMA2       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTNLYDELLKNVVNLLNTIEVK : 386
AND6       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTNLYDELLKNVVNLLNTIEVK : 386
CAR12      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDKLLKNVVNLLNTIEVK : 386
VCG-I      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 386
PAC1       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 386
ORI1       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 383
CAR4       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTQLYDELLKNVVNLLNTIEVK : 386
AND14      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 383
PAC6       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
AND2       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
AND7       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
AND12      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
CAR13      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
CAR11      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
AND13      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
AND10      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 386
CAR8       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDKLLKNVVNLLNTIEVK : 387
AND3       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 381
AMA3       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
AND8       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 384
AND9       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 388
CAR3       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 388
CAR6       : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 388
AND11      : NLFDAFFKKDGNPSPADAVSFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDTKLYDELLKNVVNLLNTIEVK : 383
nLfdAfFkKdGNpsPADAVsFFKKMLNDPNVQKEFDNFVHGGLYGFAKRHNYLRGERMTDT LYDeLLKNVVNLLNTIEVK

```


	120	*	140	*	160	*	180	*	200	*	220		
Sal-I	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 209		
AND11	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
CAR12	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERDEA	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 205		
AND6	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERDEA	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 205		
AMA5	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERDEA	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 207		
AND1	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERDEA	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 205		
AND5	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR11	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
AND12	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
ORI3	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
AND9	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
PAC6	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
VCG-I	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
PAC1	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
ORI2	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERDEA	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 205		
CAR2	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR4	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
ORI1	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
AND7	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
AND2	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR5	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR6	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR3	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRDTATSQ	GPSQT	: 203		
AND13	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
AND10	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR13	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
PAC3	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	-----	EEAGERD	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 203		
CAR10	: DLDNYDADFIGQGKKIITKGAADKEEEEEED	DEDEEAVEGD	EEAVEGD	EEAGEED	-----	EAAEAEAVRGDTS	HQSGSGDLTPSP	PPGSSERAPAVAP	GETPRVTATSQ	GPSQT	: 210		
CAR8	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGELAPAT	ARGETATDSPT	SQEPSQT	: 202	
PAC4	: DLDNYDADFIGQGKKIITKGVVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGELAPAT	ARGENATDSPT	SQEPSQT	: 202	
PAC5	: DLDNYDADFIGQGKKIITKGVVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGELAPAT	ARGENATDSPT	SQEPSQT	: 202	
AMA3	: DLDNYDADFIGQGKKIITKGVVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGELAPAT	ARGENATDSPT	SQEPSQT	: 202	
AND3	: DLDNYDADFIGQGKKIITKGVVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGELAPAT	ARGENATDSPT	SQEPSQT	: 202	
ORI4	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGEVAPAT	ARGENATDSPT	SQEPSQT	: 202	
CAR1	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGEVAPAT	ARGENATDSPT	SQEPSQT	: 202	
AND8	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGEVAPAT	ARGENATDSPT	SQEPSQT	: 202	
ORI5	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDSEASGEDP	PASGEVAPAT	ARGENATDSPT	SQEPSQT	: 202	
PAC7	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDS	AVSPRSVSE	REPATATVPA	ETARDNPPSQ	ETAFT	: 202
PAC2	: DLDNYDADFIGQGKKIITKGAVDSNEEDGG	DEDEEAVEGD	EA	-----	EAGEEG	-----	EEEATNEVASHQ	SDS	AVSPRSVSE	REPATATAP	GETARDNPPSQ	ETAFT	: 202
PAC8	: DLDNYDADFIGQGKKIITKGAASSEEGEED	EEDEDED	-----	GEAGQET	-----	DKGST	-----	SDQPDSAVSP	QRAPESPEL	TTATARGATV	RDSPTS	PGLSPT	: 193
CAR7	: DLDNYDADFIGQGKKIITKGAASSEEGEED	EEDEDED	-----	GEAGQET	-----	DKGST	-----	SDQPDSAVSP	QRAPESPEL	TTATARGATV	RDSPTS	PGLSPT	: 193
AMA4	: DLDNYDADFIGQGKKIITKGAASSEEGEED	EEDEDED	-----	GEAGQET	-----	DKGST	-----	SDQPDSAVSP	QRAPESPEL	TTATARGATV	RDSPTS	PGLSPT	: 193
AND4	: DLDNYDADFIGQGKKIITKGAASSEEGEED	EEDEDED	-----	GEAGQET	-----	DKGST	-----	SDQPDSAVSP	QRAPESPEL	TTATARGATV	RDSPTS	PGLSPT	: 193
	dLDNYDADFIGQGKKIITKGA	d Ee	dedEEavEgd		eag	e ea	ShQs S	s p S E	apA a get		tSq psqT		

340 * 360 * 380 *

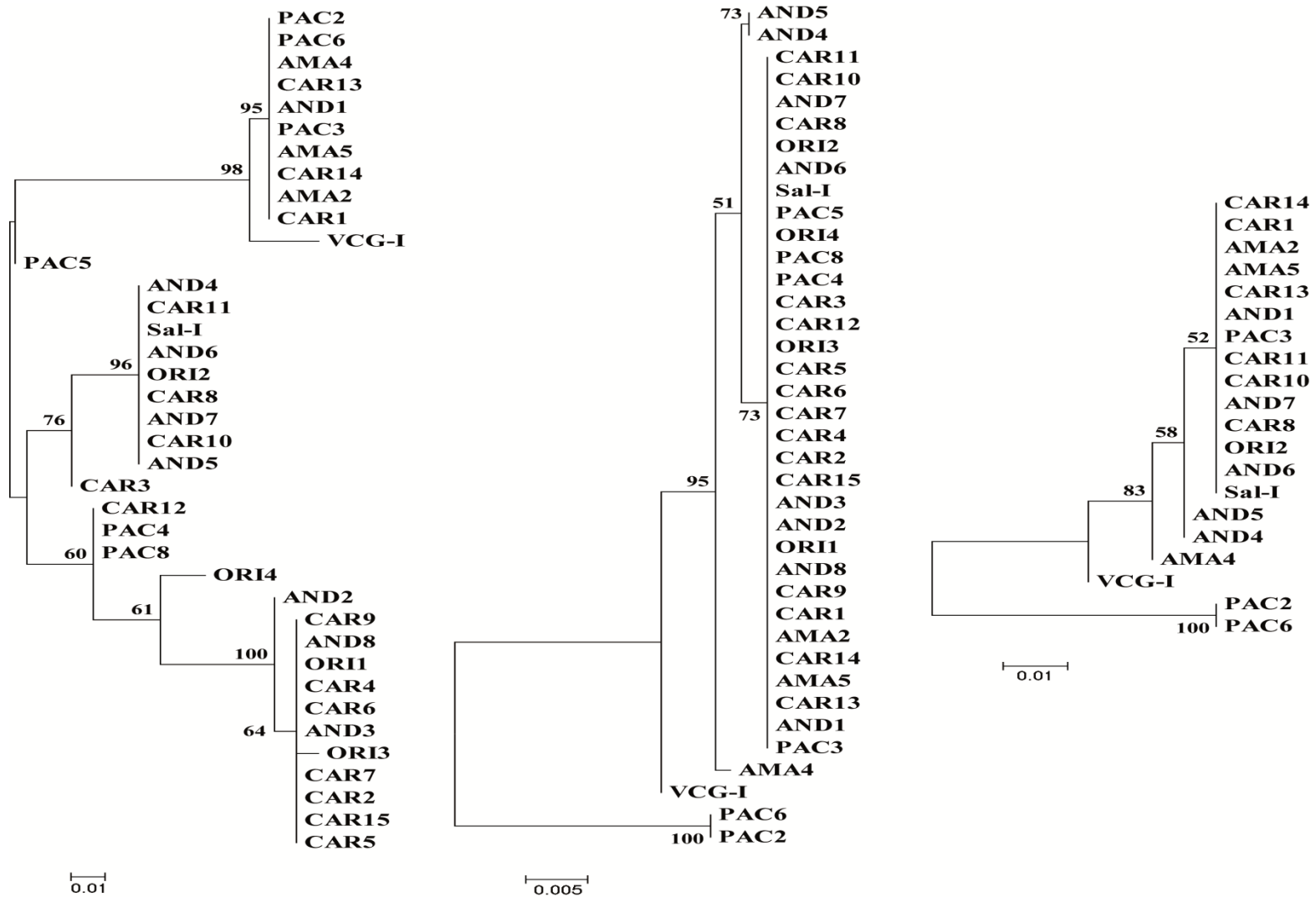
Sal-I : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 385
AND11 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR12 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 381
AND6 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 381
AMA5 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 383
AND1 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 381
AND5 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR11 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
AND12 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
ORI3 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRGERMNLTAHKS L FVNTLSLLDTL : 379
AND9 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRGERMNLTAHKS L FVNTLSLLDTL : 379
PAC6 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
VCG-I : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
PAC1 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
ORI2 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 381
CAR2 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR4 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
ORI1 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
AND7 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
AND2 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR5 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR6 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR3 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRGERMNLTAHKS L FVNTLSLLDTL : 379
AND13 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
AND10 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR13 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
PAC3 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 379
CAR10 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 386
CAR8 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRGERMNLTAHKS L FVNTLSLLDTL : 378
PAC4 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
PAC5 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
AMA3 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
AND3 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
ORI4 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
CAR1 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
AND8 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
ORI5 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
PAC7 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
PAC2 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 378
PAC8 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 369
CAR7 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRGERMNLTAHKS L FVNTLSLLDTL : 369
AMA4 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 369
AND4 : NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL : 369
NPSPADAVSFFKKMLNDPNVQKEFDNFVHGlyGFAKRHNYLRDERMNLTAHKS L FVNTLSLLDTL

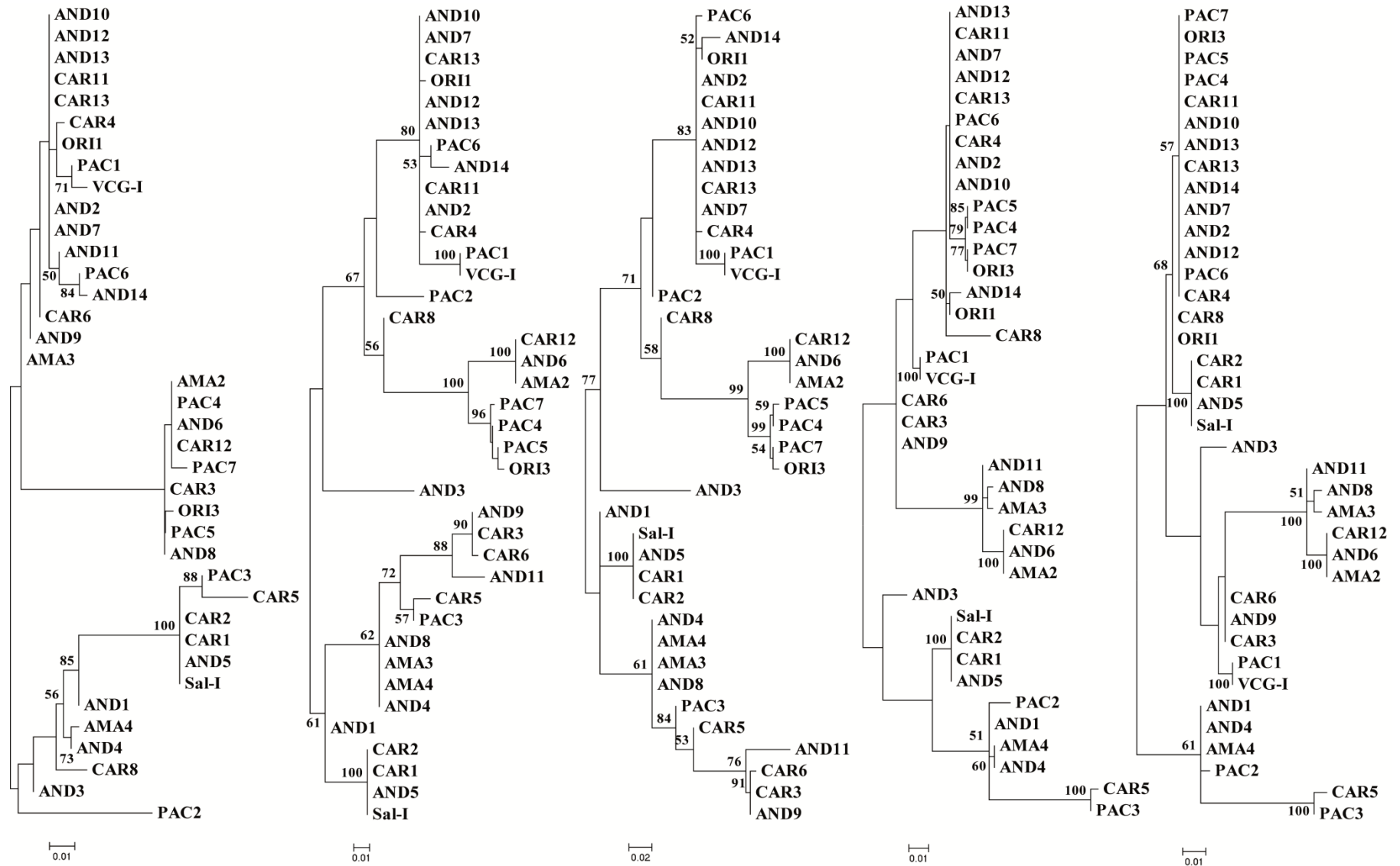
12.4 Anexo 4: Codones seleccionados positiva o negativamente dentro de cada gen.

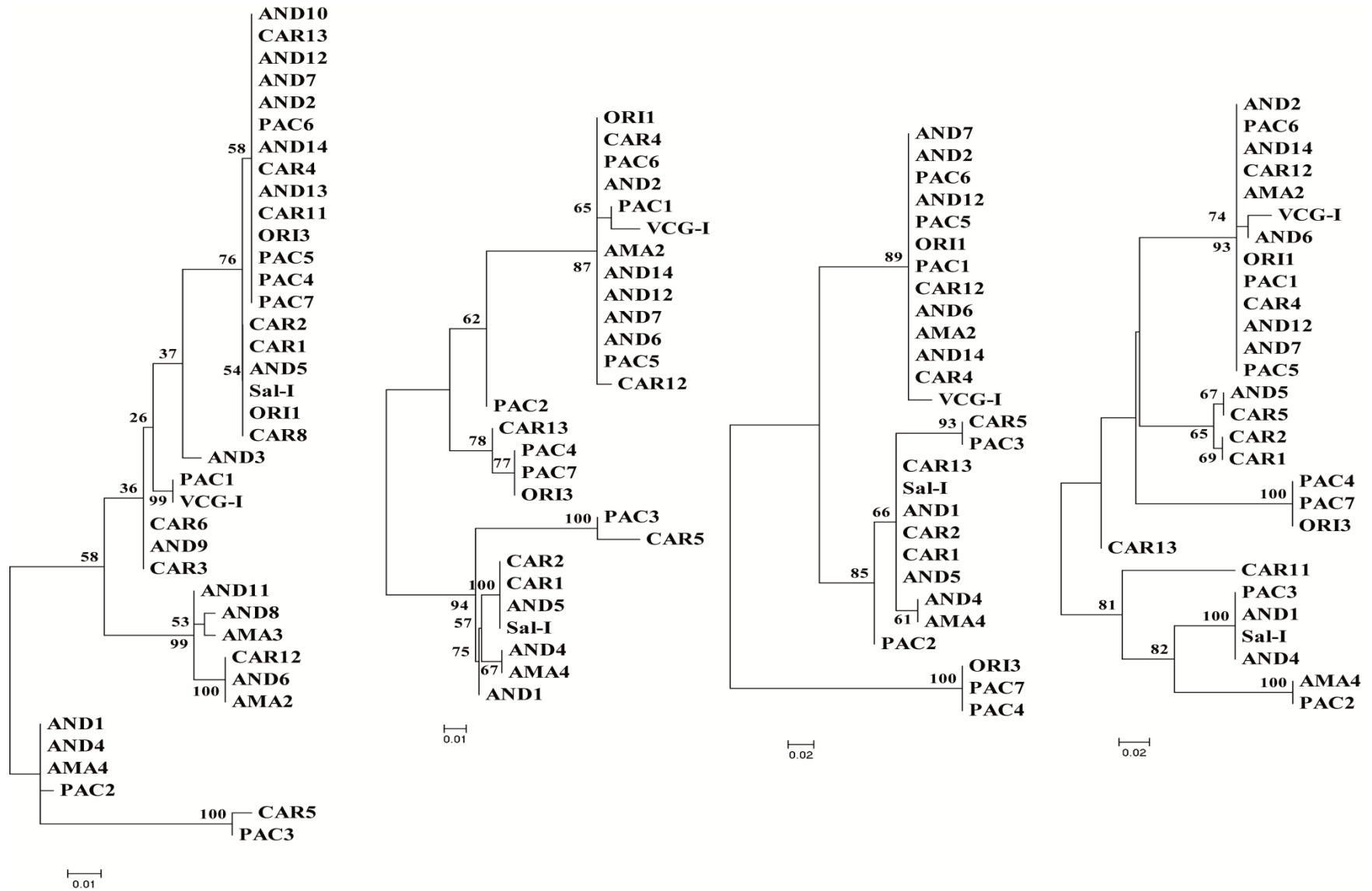
Gen	Codones seleccionados positivamente	Codones seleccionados negativamente
<i>Pvmsp-7C</i>	135, 145, 146, 147, 152, 155, 160, 161, 162, 167, 172, 179, 184, 185, 186, 187, 194, 201, 229	38, 2, 57, 65, 70, 71, 96, 97, 102, 103, 112, 122, 123, 125, 195, 225, 230, 247, 248, 254, 263
<i>Pvmsp-7H</i>	76, 81, 84, 100, 127, 145, 149, 160, 161, 165, 169, 168, 169, 171, 172, 173, 175, 176, 177, 178, 179, 180, 183, 185, 187, 189, 191, 193, 196, 198, 201, 202, 206, 207, 209, 211, 214, 216, 217, 219, 222, 224, 225, 227, 229, 230, 236, 244, 365, 369	25, 37, 38, 44, 47, 53, 56, 57, 69, 73, 75, 79, 85, 88, 91, 97, 101, 105, 106, 111, 122, 124, 129, 144, 146, 147, 182, 204, 253, 255, 261, 296
<i>Pvmsp-7I</i>	178, 179, 181, 185, 187, 193, 211, 214, 218, 294	14, 37, 38, 47, 85, 87, 100, 111, 112, 127, 132, 148, 149, 153, 154, 155, 162, 167, 172, 191, 198, 212, 228, 234, 237, 246, 251, 257, 265, 292, 302, 307, 358, 359

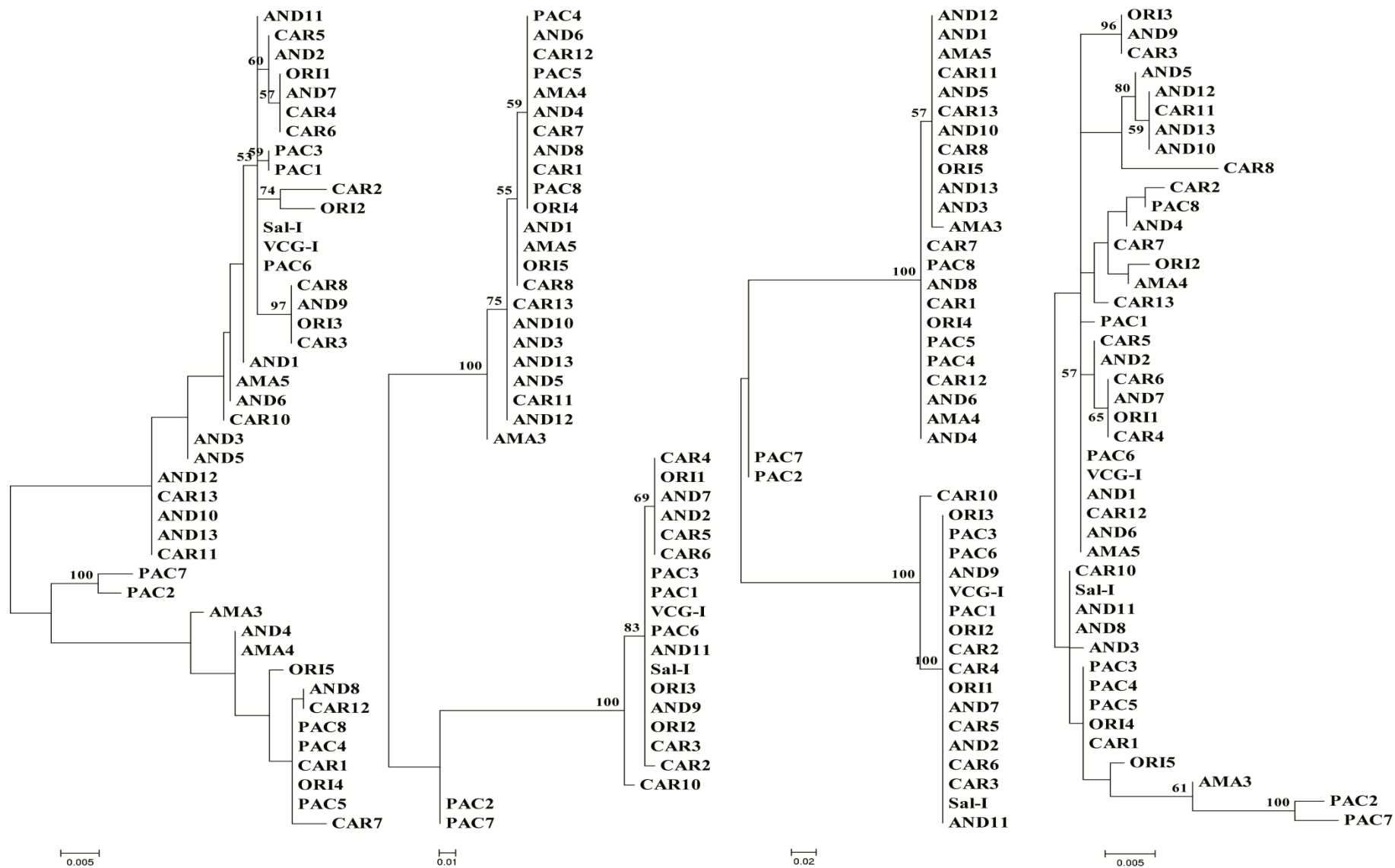
12.5 Anexo 5: Topologías inferidas para cada gen teniendo en cuenta la recombinación.

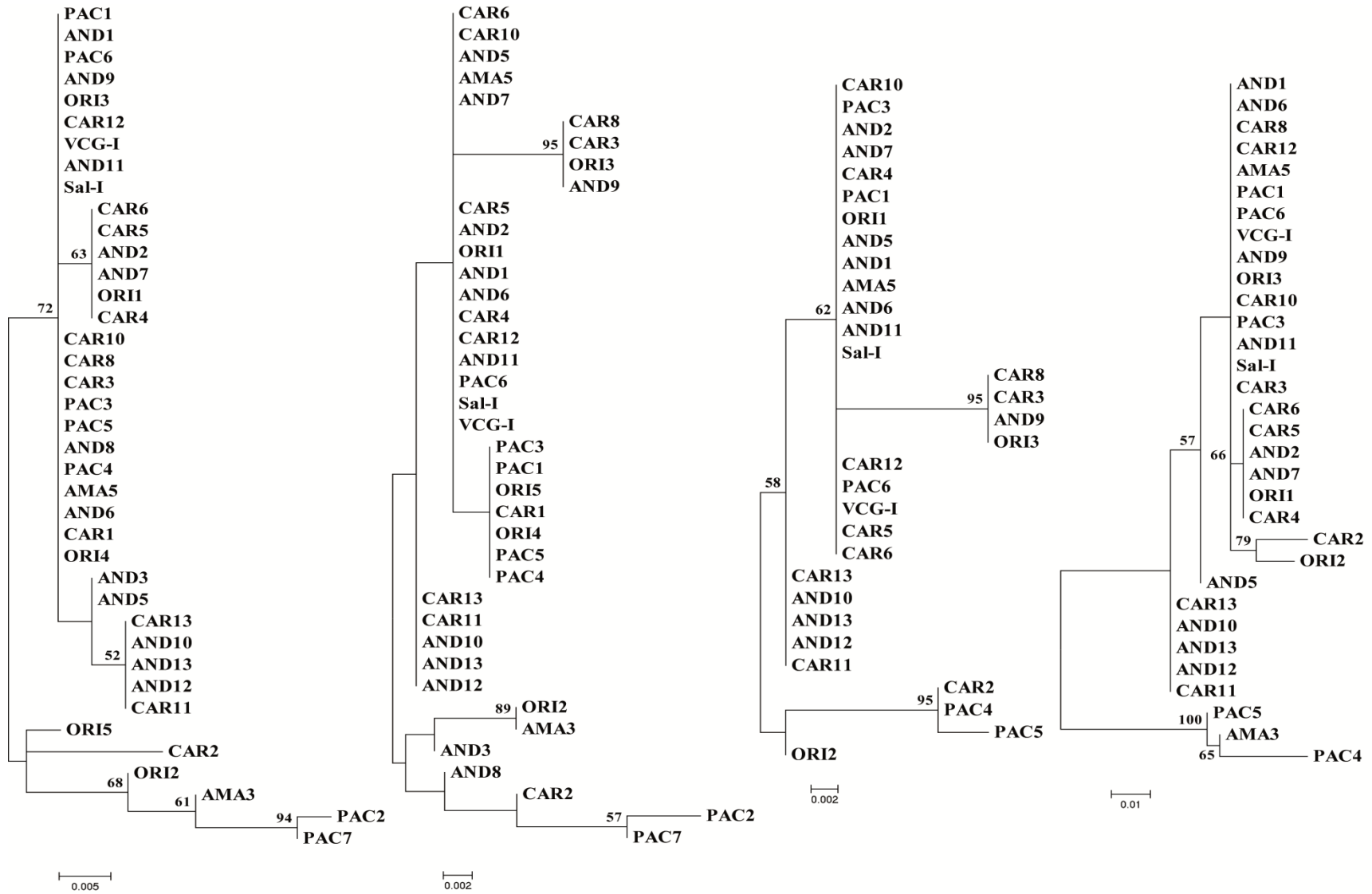
PvMSP7C



***Pvm*msp-7H**

Continuación de *Pvm*sp-7H

*Pvm*msp-7I

Continuación de *Pvmsp-71*

Genetic Diversity and Selection in Three *Plasmodium vivax* Merozoite Surface Protein 7 (*Pvmsp-7*) Genes in a Colombian Population

Diego Garzón-Ospina^{1,2,3}, Carolina López^{1,2,3}, Johanna Forero-Rodríguez¹, Manuel A. Patarroyo^{1,3*}

1 Fundación Instituto de Inmunología de Colombia – FIDIC, Bogotá DC, Colombia, **2** Microbiology postgraduate program, Universidad Nacional de Colombia, Bogotá DC, Colombia, **3** School of Medicine and Health Sciences, Universidad del Rosario, Bogotá DC, Colombia

Abstract

A completely effective vaccine for malaria (one of the major infectious diseases worldwide) is not yet available; different membrane proteins involved in parasite-host interactions have been proposed as candidates for designing it. It has been found that proteins encoded by the merozoite surface protein (*msp*)-7 multigene family are antibody targets in natural infection; the nucleotide diversity of three *Pvmsp-7* genes was thus analyzed in a Colombian parasite population. By contrast with *P. falciparum msp-7* loci and ancestral *P. vivax msp-7* genes, specie-specific duplicates of the latter specie display high genetic variability, generated by single nucleotide polymorphisms, repeat regions, and recombination. At least three major allele types are present in *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I* and positive selection seems to be operating on the central region of these *msp-7* genes. Although this region has high genetic polymorphism, the C-terminus (Pfam domain ID: PF12948) is conserved and could be an important candidate when designing a subunit-based antimalarial vaccine.

Citation: Garzón-Ospina D, López C, Forero-Rodríguez J, Patarroyo MA (2012) Genetic Diversity and Selection in Three *Plasmodium vivax* Merozoite Surface Protein 7 (*Pvmsp-7*) Genes in a Colombian Population. PLoS ONE 7(9): e45962. doi:10.1371/journal.pone.0045962

Editor: Erika Martins Braga, Universidade Federal de Minas Gerais, Brazil

Received: February 27, 2012; **Accepted:** August 28, 2012; **Published:** September 25, 2012

Copyright: © 2012 Garzón-Ospina et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mapatarr.fidic@gmail.com

Introduction

Plasmodium vivax malaria should not be considered a benign disease anymore due to its wide geographical distribution [1,2] and ability to cause severe clinical syndromes [3,4,5,6]. It was estimated that 2.85 billion people were at risk of being infected by this parasite in 2009 [1]. Although several research groups have focused on developing a vaccine against *P. vivax*, relatively few vaccine candidate molecules have been thoroughly characterized [7,8]; some of them are membrane proteins making attractive targets for vaccine development since they are essential for the recognition and invasion of erythrocytes, hence immune responses directed against them could inhibit parasite entry [9]. The merozoite surface protein (MSP) family is a group of surface proteins which are involved in the initial interaction between the parasite and the host cell. However, the genetic diversity of these proteins has been the main problem for developing a vaccine [10,11,12,13,14]. Proteins displaying high antigenic diversity elicit strain-specific immune responses leading to low protective efficacy upon vaccination, while proteins displaying limited variability are attractive targets for testing as vaccine candidates. Research involving proteins displaying high overall polymorphism should be focused on the conserved functional domains [15] so that strain-specific responses may be avoided.

MSP-1, MSP-6 and MSP-7 form the most abundant protein complex on the *P. falciparum* merozoite surface [16,17]. *msp-1* and *msp-6* are both single copy genes while *msp-7* has a large number of genes which seem not to have diverged functionally. Furthermore, the *Plasmodium* genus has dissimilar copy numbers of this gene

[18,19]. The *msp-7* genes have a single exon and have been named in alphabetical order according to their position regarding the PVX_082640 flanking gene [18]. Different *msp-7* genes are expressed simultaneously in the schizont stage of *Plasmodium* species [20,21,22] and several of them have been localized on the membrane surface [19,21,22,23], not only forming part of the main merozoite protein complex but also binding to erythrocytes [24]. Knockout and invasion inhibition assays have also shown that MSP-7 is involved in the invasion of erythrocytes [19,24,25]. Moreover, Wang's group has shown that a recombinant MSP-7 is recognized by sera from malaria-infected individuals where IgG3 subclass antibodies prevail [26]. Immunization with members of this family has been shown to protect vaccinated mice against experimental challenge [21].

P. falciparum has eight *msp-7* genes and *P. vivax* eleven, [18,27]; the members of this family have low genetic variability in the former specie, [28,29] as do the *P. vivax msp-7A* (GenBank ID: XM_001614080.1) and *msp-7K* (GenBank ID: XM_001614090.1) genes [30]. A recent study has shown that *P. vivax* specie-specific duplicates MSP-7C (GenBank ID: XM_001614082.1) and MSP-7H (GenBank ID: XM_001614087.1) are recognized by IgG antibodies from *P. vivax*-infected patient sera [31]. These proteins are phylogenetically related to PvMSP-7I (GenBank ID: XM_001614088.1) [18,27] and fragments from different MSP-7 proteins may have to be included to block this invasion route due to the biological implication that functional redundancy might have on vaccine development. Taking into account that some of these proteins are recognized by the immune system, their genetic

diversity should be evaluated to determine their potential use as potent anti-malarial vaccine candidates.

Results

Forty-eight parasite samples were collected from symptomatic patients living in representative regions of Colombia for this study: Amazonian (Calamar, Guaviare department), Andean (Apartadó and El Bagre, Antioquia department), Caribbean (Tierra Alta and Puerto Libertador, Córdoba department), Orinoquia (Mapiripan, Meta department and Tauramena, Casanare department) and Pacific (Istmina, Chocó department and Tumaco, Nariño department) (Fig. S1). Forty-two samples corresponding to single *P. vivax msp-1* allele infections were considered for PCR amplification even though amplicons were not detected in a few samples (*Pvmsp-7C* n = 37, *Pvmsp-7H* n = 37 and *Pvmsp-7I* n = 42).

Polymorphism in *Pvmsp-7* Loci

1,167-bp *Pvmsp-7C*, 1,232-bp *Pvmsp-7H* and 1,182-bp *Pvmsp-7I* gene fragments were amplified and direct sequenced. Nucleotide sequence data here reported are available in GenBank: accession numbers JQ423957-JQ424037. Twenty-three haplotypes were found for *Pvmsp-7C* while twenty-eight haplotypes were found for both *Pvmsp-7H* and *Pvmsp-7I*. Haplotype diversity (Hd) was lower in the *Pvmsp-7C* gene than in *Pvmsp-7H* and *Pvmsp-7I* (Table 1). 148 sites from the total nucleotide sequence length analyzed were polymorphic in *Pvmsp-7C* and 142 sites in *Pvmsp-7H*, while 121 polymorphic sites were found in *Pvmsp-7I* (Table 1).

Nucleotide diversity for these three genes (*Pvmsp-7C* $\pi = 0.0548$, *Pvmsp-7H* $\pi = 0.0357$ and *Pvmsp-7I* $\pi = 0.0430$) was higher than that previously reported for other *P. vivax msp-7* family members (*Pvmsp-7A* $\pi = 0$ and *Pvmsp-7K* $\pi = 0.0022$) [30]. Furthermore, the *msp-7* genes from this study were among the most polymorphic *P. vivax* genes reported to date (Table S1). Regarding *Pvmsp-7* diversity (π value average) in the different Colombian regions, the Pacific area was the most diverse followed by the Andean, Caribbean and Orinoco regions while the Amazonian was the least polymorphic (Table S2).

Three major allele types were found in deduced PvMSP-7C amino acid sequences (Fig. 1 and Fig. S2) between residues 134 and 238 (numbered according to the alignment in Fig. S2). Allele types had similar frequency in the Colombian parasite population.

An AEAFG insertion-deletion from residue 146 to 150 and a GTG_(n)GT[V/E] repeat region from residue 195 to 209 (numbered according to the alignment in Fig. S3) were revealed by protein sequence alignment in PvMSP-7H. Three regions

throughout the protein sequences led to discriminating several allele types (Fig. 2 and Fig. S3). The first region from residue 153 to 170 (numbered according to the alignment given in the Fig. S3) had four different peptide sequences. Five different sequences were found in the second region (from residue 172 to 194) and three different ones were found from residue 229 to 247. These three regions were randomly associated (Fig. 2 and Fig. S3).

EEAVEGD and EA repeats could be observed in the deduced PvMSP-7I amino acid sequences. Similar to the genes mentioned above, several major allele types were observed between residues 131 and 219 (numbered according to the alignment in Fig. S4) characterized by two regions. The first region (between amino acids 131 and 140) had three different peptide sequences (Fig. 3 and Fig. S4). The second region ran from residue 163 to residue 219, having an extra four different peptide sequences. These regions were not randomly associated (Fig. 3 and Fig. S4). A further four peptide sequences were found from residue 221 to residue 234 and two more from residue 236 to residue 264 and, contrary to those mentioned above, these fragments were found to be associated with either of the alleles described above (Fig. 3 and Fig. S4).

Neutral Evolutionary Test and Selection in the *Pvmsp-7* Loci

Applying neutral evolution tests (Tajima, Fu & Li) to the Colombian *P. vivax* population gave significant values above 0 for *Pvmsp-7C* and *Pvmsp-7I* ($p < 0.05$ Tajima D, $p < 0.02$ Fu & Li D* and F*) (Table 1); these values indicated an excess of intermediate frequency alleles. The neutral evolution tests for *Pvmsp-7H* revealed no statistically significant differences; this gene therefore seems to be evolving under neutral expectations. However, a sliding window analysis for D, D* and F* statistics showed that different selective forces could be acting throughout the *Pvmsp-7* genes. It was found that balancing selection acted on the central gene region while negative selection could be acting at the 5'-ends (significant values for *Pvmsp-7C* and *Pvmsp-7H*) and 3'-ends (no significant values) (Fig. S5).

The average number of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) was estimated to determine whether natural selection was affecting the *msp-7* loci. Although d_N was higher than d_S in *Pvmsp-7C* and *Pvmsp-7I*, d_N was lower than d_S in the *Pvmsp-7H* gene even though these values were not statistically significant (Table 2). Taking into account that a sliding window for ω (d_N/d_S) (Fig. S6) showed that the ω rate was higher than 1 (signal of positive selection) in the central region of the *msp-7* genes analyzed

Table 1. Estimates of DNA diversity and neutrality test for *Pvmsp-7* genes from a Colombian population.

n	Gene	Sites	Ss	S	Ps	H	Hd (se)	θ^W (se)	π (se)	Tajima			Fu & Li		Fu	Z _{ns}	ZZ	RM
										D	D*	F*	D*	F*				
37	<i>msp-7C</i>	1,098	148	1	147	23	0.93 (0.03)	0.0323 (0.001)	0.0548 (0.003)	2.094**	1.810*	2.287*	7.64	0.339	0.173*	22		
37	<i>msp-7H</i>	1,131	142	19	123	28	0.96 (0.02)	0.0301 (0.002)	0.0357 (0.003)	0.388	0.952	0.987	-0.57	0.126	0.343*	27		
42	<i>msp-7I</i>	1,058	127	6	121	28	0.97 (0.01)	0.0280 (0.002)	0.0430 (0.003)	1.420	1.584**	1.818*	2.12	0.232	0.334*	13		

Ss: Number of segregating sites, S: number of singleton sites, Ps: number of parsimony-informative sites, H: number of haplotypes, Hd: haplotype diversity, θ^W : Watterson estimator, π : nucleotide diversity. (se): Standard deviation. Sites excluded from the analysis: *Pvmsp-7C*: nucleotides 1 to 17 (amino acids 1 to 6), 544 to 546 (amino acid 182), 616 to 618 (amino acid 206), 625 to 627 (amino acid 209), 631 to 648 (amino acids 211 to 216), 655 to 666 (amino acids 219 to 222), 676 to 684 (amino acids 226 to 228), 706 to 708 (amino acid 236), 1,105 to 1,107 (amino acid 369) and nucleotides 1,168 to 1,191 (amino acids 390 to 397); *Pvmsp-7H*: nucleotides 436 to 450 (amino acids 146 to 150), 448 to 486 (amino acid 162), 568 to 627 (amino acids 190 to 200) and nucleotides 772 to 774 (amino acid 258); *Pvmsp-7I*: nucleotides 1 to 18 (amino acids 1 to 6), 421 to 522 (amino acids 141 to 174) and nucleotides 526 to 537 (amino acids 176 to 179).

*: $p < 0.02$, **: $p < 0.05$.

doi:10.1371/journal.pone.0045962.t001

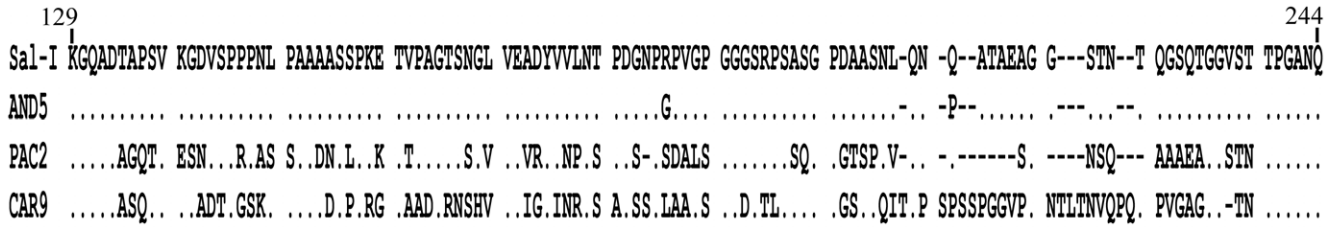


Figure 1. Alignment of deduced PvMSP-7C amino acid sequences, including the Sal-I sequence in which the three main allele types can be observed. Dots indicate conserved residues and dashes represent gaps introduced for alignment.
doi:10.1371/journal.pone.0045962.g001

here, the genes were thus split into 3 regions: 5'-end (*msp-7C*: between nucleotides 1 and 390, *msp-7H*: between nucleotides 1 and 471, *msp-7I*: between nucleotides 1 and 525), central (*msp-7C*: between nucleotides 391 and 717, *msp-7H*: between nucleotides 472 and 771, *msp-7I*: between nucleotides 526 and 789) and 3'-end (*msp-7C*: between nucleotide 718 and 1,191, *msp-7H*: between nucleotide 772 and 1,200, *msp-7I*: between nucleotide 790 and 1,188) (Table S3); d_N and d_S were then computed. d_S was higher than d_N for both 5'- and 3'-ends, just the 5'-end having significant statistical difference. On the other hand, the d_N substitutions in the central region were significantly higher than d_S substitutions for the three genes (Table 2).

Furthermore, several maximum likelihood-based methods were used for identifying which codon sites were under positive or negative selection. *Pvmsp-7C* had 19 sites under positive selection and 21 negatively selected sites according to SCAL, FEL, IFEL and REL methods. Positive selection signatures were found for *Pvmsp-7H* in 50 sites while another 32 sites were negatively selected and *Pvmsp-7I* had 10 sites under positive selection and 30 negatively selected sites (predicted by at least one method, Tables S4 and S5). Taking into account that recombination (see below) can affect the reliability of identifying sites under selection [32], the analysis was performed again, this time considering recombination. Two sites were thus positively selected and 43 were under negative selection for *Pvmsp-7C*. Ten sites were positively selected and 24 were negatively selected for *Pvmsp-7H*, and 4 sites were under positive selection while 13 sites were under negative selection for *Pvmsp-7I* (Tables S6 and S7).

Linkage Disequilibrium (LD) and Recombination

Random association was observed among *Pvmsp-7* haplotypes, therefore suggesting their independent segregation. LD analysis measured by Z_{NS} (average of R^2) for whole data revealed no statistically significant values for the three *msp-7* genes (Table 1). The relationship of R^2 (linkage disequilibrium) with physical distance in regression analysis showed that LD declined as

nucleotide distance increased (Fig. S7). The ZZ statistic had significant values (Table 1). Therefore, both the regression analysis and ZZ statistic suggested that intragenic recombination was taking place in *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I*. Recombination events were detected by using DnaSP which revealed several RM events in all genes (Table 1). Figure 4 shows the recombinant regions detected with RDP v3.4 software.

Clustering Analysis

Since recombination was detected for *msp-7* genes, phylogeny had to be inferred taking it into account. A phylogenetic tree was inferred for each recombinant region detected (Fig. S8, S9, S10). All topologies showed no geographical clustering among the Colombian isolates; moreover, Colombian and Salvadorian sequences clustered together (Fig. S8, S9, S10).

All *Pvmsp-7* sequences were aligned and trees were constructed by using ML with the TN93+G model. The tree showed three monophyletic groups (Fig. S11A); the first group clustered *Pvmsp-7H* sequences, close to the second group which clustered *Pvmsp-7I* sequences, while the third group clustered *Pvmsp-7C* sequences. This topology agreed with the phylogenetic relationship previously reported for *Pvmsp-7* [18,27]. However, *Pvmsp-7* sequences showed recombinant fragments possibly produced by gene conversion between paralogous genes at the 5' and 3'-ends but not in the central region (Fig. S12 and Table S8). The *Pvmsp-7* genes became clustered into paraphyletic groups when trees were inferred taking gene conversion into account (Fig. S11B–E). Codons 38 and 112 in the three genes, as well as the 57, 70, 102 and 125 for *Pvmsp-7C* and *Pvmsp-7H*, and codons 48 and 86 for *Pvmsp-7H* and *Pvmsp-7I* (numbered according to the alignment in Fig. S13) were predicted as negatively selected sites but, taking into account that they lay within the gene conversion tracks detected, purifying selection at these sites could have been the result of gene conversion between paralogous gene loci.

The d_S and d_N rates between paralogous gene pairs were estimated to determine whether homogeneity at the *Pvmsp-7* ends

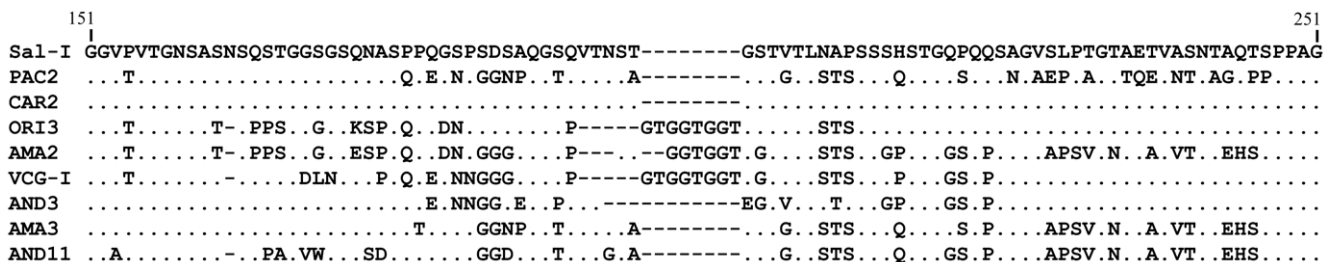


Figure 2. Alignment of deduced PvMSP-7H amino acid sequences, including the Sal-I sequence. The alignment shows major allele types found in Colombian populations. Dots indicate conserved residues and dashes represent gaps introduced for alignment.
doi:10.1371/journal.pone.0045962.g002

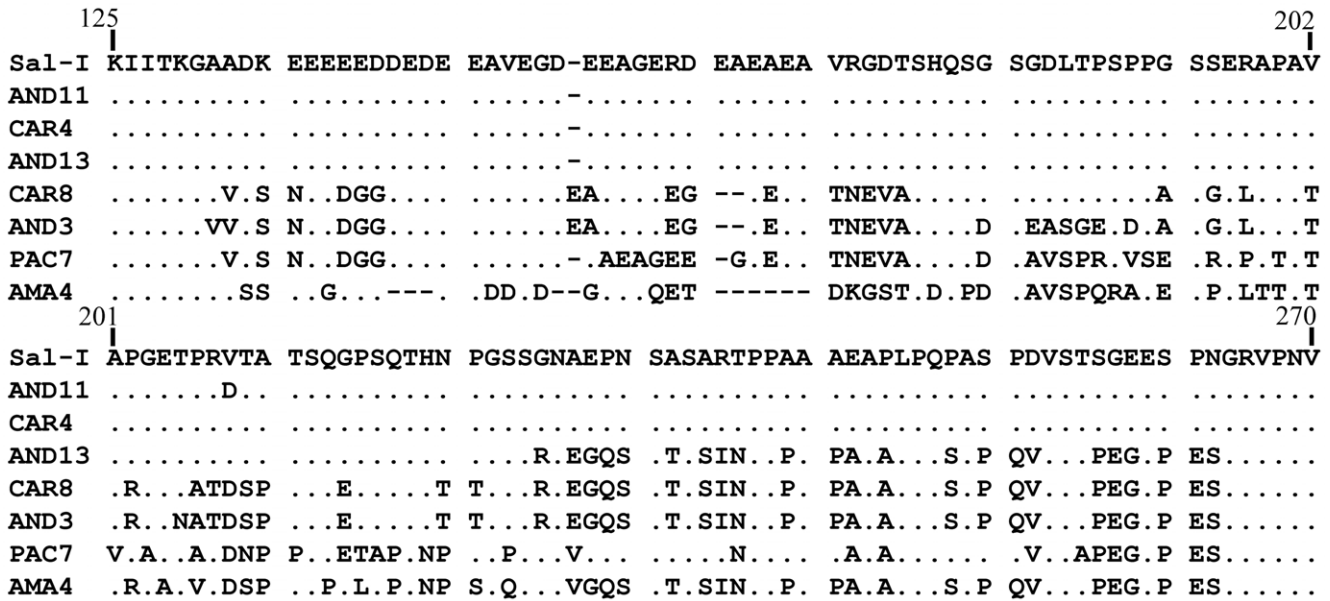


Figure 3. Alignment of deduced PvMSP-7I amino acid sequences, including the Sal-I sequence in which the protein sequences differentiating major allele types can be observed. Dots indicate conserved residues and dashes represent gaps introduced for alignment. doi:10.1371/journal.pone.0045962.g003

was caused by concerted evolution (Table 3); the d_S was higher than d_N at the 5'-end for the three genes. The d_S rate at the 3'-end between *Pvmsp-7C* and the other two genes was higher than d_N rate but the d_S was similar to d_N between *Pvmsp-7H* and *Pvmsp-7I* (Table 3).

Discussion

At least eleven MSPs have been reported in *P. falciparum* and nine *P. vivax* orthologous genes have been described so far. Two *P. vivax* MSP families (*msp-3* and *msp-7*) are particularly interesting since they have expanded differentially [18,33]. The *P. vivax msp-7* gene family has 11 functional genes [18] and at least 8 of them are transcribed during the *P. vivax* schizont stage [20]. MSP-7 forms part of the main protein complex interacting with host cells [16,22]. Taking into account that this protein is localized on the parasite membrane [19,21,22,23], it can be targeted by the host immune system and thus becomes an attractive vaccine candidate.

Besides being recognized by the immune system, vaccine candidates must have limited polymorphism. Results found in this study highlighted differences in allele diversity among *msp-7* gene family members, at least in the two main human malaria

species. By contrast with the very little polymorphism displayed by *Pfmsp-7* genes [28,29], *Pvmsp-7* genes (*Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I*) have high genetic diversity. It is worth noting, however, that not all the *Pvmsp-7* genes display a similar pattern. *Pvmsp-7A* and *Pvmsp-7K*, like *Pfmsp-7A*, *Pfmsp-7B*, *Pfmsp-7C*, *Pfmsp-7D*, *Pfmsp-7E*, *Pfmsp-7F* and *Pfmsp-7H* (GenBank ID numbers: XM_001350038.1–44.1) display very low polymorphism [28,29,30] suggesting that *msp-7* genes might be exposed to different selective pressure (such as that exerted by the immune system) or have different biological constraints. Although low genetic polymorphism does not hold for *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I*, vaccine development using these proteins could be focused on their conserved C-terminal domain (Pfam ID number PF12948) due to the high level of conservation it displays and, like in *P. falciparum*, the C-terminal domain could be involved in interaction with erythrocytes [24].

The genetic diversity found in these three genes places them among the most diverse *P. vivax* genes described so far (Table S1). Different allele types were found when Colombian samples were compared with Sal-I (GenBank ID numbers: XM_001614082.1, XM_001614087.1, XM_001614088.1) and unpublished Korean sequences, (GenBank ID numbers: GU476538.1, GU476534.1,

Table 2. Average number of synonymous substitutions per synonymous site (d_S) and non-synonymous substitutions per non-synonymous site (d_N) covering all sequence pairs at the 5'-end, central region, 3'-end and complete gene.

	5'-end		Central region		3'-end		Full length gene	
	d_S (se)	d_N (se)	d_S (se)	d_N (se)	d_S (se)	d_N (se)	d_S (se)	d_N (se)
<i>msp-7C</i>	0.070 (0.016)**	0.011 (0.004)	0.122 (0.023)	0.287 (0.031)•	0.004 (0.003)	0.000 (0.000)	0.053 (0.008)	0.059 (0.007)
<i>msp-7H</i>	0.059 (0.010)•	0.010 (0.003)	0.052 (0.016)	0.162 (0.019)•	0.010 (0.005)	0.005 (0.002)	0.038 (0.006)	0.037 (0.005)
<i>msp-7I</i>	0.041 (0.014)*	0.020 (0.007)	0.060 (0.019)	0.186 (0.021)•	0.017 (0.007)	0.007 (0.003)	0.036 (0.007)	0.048 (0.006)

se: standard error. 5'-end (*Pvmsp-7C*: nucleotides 1–390, *Pvmsp-7H*: nucleotides 1–471, *Pvmsp-7I*: nucleotides 1–525), central (*Pvmsp-7C*: nucleotides 391–717, *Pvmsp-7H*: nucleotides 472–771, *Pvmsp-7I*: nucleotides 526–789) and 3'-end (*Pvmsp-7C*: nucleotides 718–1,191, *Pvmsp-7H*: nucleotides 772–1,200, *Pvmsp-7I*: nucleotides 790–1,188).

*: $p < 0.04$, **: $p < 0.001$, •: $p < 0.0001$.

doi:10.1371/journal.pone.0045962.t002

variability within a population [34] maintaining alleles at intermediate frequencies, as can be observed in the alignment. Nevertheless, Tajima and Fu & Li tests are influenced not only by selection but also by the population history that can alter neutral allele frequency expectations. Therefore, the positive values in the tests could have been the result of a decrease in the population. However, a population affected by genetic drift (a mechanism that decreases the population) was not found for the Colombian population (Fs not statically significant). Moreover, the Hd and π values for both genes suggested a stable population having a long-term effective population size; allele frequency distribution would not therefore be influenced by a demographic process.

By contrast with the *Pvmsp-7C* and *Pvmsp-7I* results discussed above, *Pvmsp-7H* appeared to be under a standard neutral model of molecular evolution since there were no significant values in neutrality tests and no difference between d_N and d_S was found; consequently, high polymorphism would be expected in regions lacking functional constraints [35]. However, a deviation from neutral expectation would only be detected if the average from the whole gene was significantly greater or smaller than 0. The Nei-Gojobori method behaves similarly; positive selection can be detected only if the d_N average from the whole gene is significantly greater than d_S (the opposite occurs for negative selection). The sliding window test (for neutral statistics and ω) for different selective pressures appeared to be acting throughout *Pvmsp-7* gene sequences, as has been described for others proteins [36,37,38,39]. The tests showed that the 5'- and 3'-ends had purifying selection signals (negative values in the neutrality test and $\omega < 1$) while the central region of the three genes had balancing or positive selection signals (positive values in the neutrality test and $\omega > 1$). So, despite the loci displayed being under balancing selection or under neutrality, natural selection may have varied across codons. Two different approaches were thus followed to investigate this hypothesis. One estimated the d_N and d_S rates at the 5'- and 3'-ends and in the central region of the three genes. The central region was under positive selection while the 5'- and 3'-ends seemed to be under negative selection (d_S higher than d_N , only significant for the 5'-end). Similar results were observed when using a second approach in which several maximum likelihood methods were performed to determine positively and negatively selected codons. The central region of the *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I* genes had codons under positive selection while sites under negative selection were preferentially located at the 5'- and 3'-ends which were relatively fully conserved in *Pvmsp-7* genes.

Besides selective pressure and functional constraints, evolution of malarial antigens might be affected by recombination [13,14,39,40,41]. Several statistics and algorithms showed that intragenic recombination played an important role in generating new allele variants in individual *msp-7* genes. These events affect the accuracy of detecting selected sites, increasing type I errors [32]. When sequences were screened for recombination, the positively selected sites decreased, suggesting that several of the positively selected sites were false positive. However, some were true positively selected sites (Table S6). These sites did not seem to be originated by the stochastic nature of the mutation process, since it would have been expected that d_N substitutions (and/or positive selection) would have been randomly found across the genes; instead, these sites were just observed in the *msp-7* central region and the non-synonymous substitutions in the positively selected sites were found as parsimonious sites and not as singleton sites. Recombination and selection thus seem to drive *Pvmsp-7* genes' antigenic variation, similar to

what occurs in other pathogen antigens [42,43,44]. Likewise, it has been shown that both diversification by recombination and purifying selection take place in different regions of the *E. coli fimA* gene [45], and this also occurs in the *Pvmsp-7* genes (Fig. 4 and Table S7).

On the other hand, multigene families might be evolving by recombination among paralogous genes, thereby contributing to allele diversity or the homogenization of the multigene family in the event of unequal crossover or gene conversion [46]. The gene conversion and subsequent phylogenetic analysis revealed that these genes had not evolved independently; several conversion tracks at the 5'- and 3'-ends (but not in the central region) were detected among *msp-7* genes, suggesting that at least these three genes could be evolving by concerted evolution mediated by a biased gene conversion, homogenizing only the 5' and 3'-ends. Thus, recombination would seem to affect the evolution of the *msp-7* family as a whole, and individual *msp-7* genes. However, protein homogeneity can also be attained by purifying selection. Under the assumption of gene conversion, it would be expected that d_S between duplicated genes would be similar to d_N , whereas if purifying selection is the major evolutionary force, then d_S would be much higher than d_N [47]. According to d_S and d_N rate comparison, the homogeneity at the *Pvmsp-7* genes' 5'-end was apparently caused by functional constraints rather than by concerted evolution, since *Pvmsp-7* genes have diverged extensively by silent nucleotide substitution at these ends. The conservation of the 3'-end in *Pvmsp-7H* and *Pvmsp-7I* but not in *Pvmsp-7C* seems to be maintained by gene conversion. This behavior might be a consequence of closely spaced gene duplicates being more likely to undergo gene conversion [48,49]. *Pvmsp-7H* and *Pvmsp-7I* genes are neighbors separated by 1,086 base pairs while the *Pvmsp-7C* gene is separated from the others by 9,619 base pairs. Accordingly, the high *Pvmsp-7s'* C-terminal conservation may have been the result of functional constraints and purifying selection, or the result of gene conversion (between *Pvmsp-7H* and *Pvmsp-7I*) possibly due to the presence of a functional domain (Pfam ID number PF12948) within this region. Therefore, the central region (the most diverse) could have been under selective pressure exerted by the immune system; consequently, the intragenic recombination and, to a lesser extent, the positively selected sites increased genetic diversity, generating different allele variants as an evasion mechanism.

Phylogenetic analysis showed that, regardless of an isolate's origin, sequences tended to cluster without having clear geographical distribution. This pattern might indicate a constant *P. vivax* gene flow in Colombian regions. This result agreed with previous reports of other highly polymorphic *P. vivax* genes not involved in geographical clustering [13,14,39].

Previous studies have shown the potential role of MSP-7 in parasite invasion of erythrocytes [19,24,25,50] which, added to its immunogenicity [26,31], and following the rules for subunit-based vaccine development [51,52], make the MSP-7 conserved domain an attractive candidate to be evaluated when designing an antimalarial vaccine.

Materials and Methods

Ethics Statement

All *P. vivax*-infected patients who provided us with the blood samples signed an informed consent and the purpose of the study was carefully explained to them. All procedures carried out in this study were approved by our institute's ethics committee.

Source of Parasite DNA and Field Isolate Genotyping

Peripheral blood samples from patients proving positive for *P. vivax* malaria by microscope examination were collected from geographical regions of Colombia from 2007 to 2010. DNA was obtained using a Wizard Genomic DNA Purification kit (Promega) following manufacturer's instructions.

All parasite samples were genotyped by PCR-RFLP of the *Pvmsp-1* gene's blocks 6, 7 and 8 as previously described [53] for selecting only samples having a single *P. vivax msp-1* allele infection.

Amplification and Sequencing

Primers were designed to amplify *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I* DNA fragments based on the Sal-I reference sequences (GenBank IDs: XM_001614082.1, XM_001614087.1 and XM_001614088.1, respectively). The DNA fragment from *Pvmsp-7C* was amplified with 7Cdto 5' ACCACAAAGATGAATAAAACG 3' and 7Crev 5' CACCTCAATCGTGTTCAGC 3' primers. *Pvmsp-7H* was amplified by using 7Hdto 5' GTGTGCATCAGTATAGCGAC 3' and 7Hrev 5' AAGAAGGTTAGCCATAAATGC 3' primers and *Pvmsp-7I* was amplified with 7Idto 5' ACAATGAGGGGCAAGTACG 3' and 7Irev 5' TTCATTCTGCTCACTTCG 3' primers. All PCR reactions were performed using KAPA HiFi HotStart Readymix containing 0.3 μ M of each primer in a final 25 μ L volume. Thermal conditions were set as follows: one cycle of 5 min at 95°C, 25 cycles of 20 sec at 98°C, 15 sec at 62°C for *Pvmsp-7C* and *Pvmsp-7I* and 60°C for *Pvmsp-7H*, 30 sec at 72°C, followed by a 5 min final extension at 72°C. PCR products were purified using the Wizard PCR preps kit (Promega), and then sequenced with a BigDye Terminator kit (MACROGEN, Seoul, South Korea) in both directions. Internal primers were used for sequencing (intdc 5' CTGTTGGACCCGGTGGAG3', intrc 5' CTTGTGTGATTCGCTCCTGG 3'; intdh 5' TCAAATACAGCACAGACTTCC, intrh 5' CCTCAGGACAACCCGAAAG 3'; intdi 5' TCAAACGCACAACCCAGG 3' and 5' GCTCCATTACCAACAACCGG 3'). Two PCR products obtained from independent PCR amplifications were sequenced per isolate to discard errors.

Statistical Analysis for the *msp-7* Sequences

Electropherograms were assembled using CLC DNA workbench 6 (CLC bio, Cambridge, MA, USA) and Clustal W [54] was used for aligning the deduced amino acid sequences, followed by manual editing. Additionally, repeats with 90% similarity in the deduced *msp-7* amino acid sequences were detected by using the T-REKS algorithm [55]. The PAL2NAL program [56] was used for constructing codon alignments from the corresponding aligned amino acid sequences. Colombian *P. vivax msp-7C*, *msp-7H* and *msp-7I* sequences were compared against the previously described Sal-I strain sequences available in GenBank (ID numbers: XM_001614082.1, XM_001614087.1, and XM_001614088.1).

DNA polymorphism was calculated with DnaSP v.5 software [57]. Tests to assess departure from the neutral model were applied using Tajima's D and Fu & Li's D* and F* statistics. The former statistic compares the differences between the total number of segregating sites and the average number of nucleotide differences between sequence pairs. The Fu & Li test calculates the D* statistic which is based on the difference between the number of singletons and the total number of mutations, as well as the F* statistic which is based on the difference between the number of singletons and the average number of nucleotide differences between sequence pairs. Positive and negative values from both tests correspond to departures from neutrality. The Fs statistic [58] was used; it is based on gene frequency distribution. All tests were applied using DnaSP v.5 software, considering coalescent simulations for obtaining confidence intervals [57].

Natural selection was estimated using the modified Nei-Gojobori method [59] to calculate the average number of non-synonymous (d_N) and synonymous (d_S) substitutions. Differences between d_N and d_S were assessed by applying the Z-test using MEGA software v.5 [60]. Additionally, codon sites under positive or negative selection at the population level were assessed by using Datamonkey web server [61] with IFEL, a codon-based maximum likelihood method [62]. This method infers selective pressure at population level; positive or negatively selected sites were assessed by FEL, SLAC and REL methods [63]. All algorithms estimated the ω (d_N/d_S ratio) at every codon in the alignment. A ≤ 0.1 p-value was considered significant for IFEL, FEL and SLAC methods and ≥ 50 Bayes factor for REL. Only Colombian sequences were considered for all analyses performed; positions containing gaps or repeats in the alignment were not taken into account (Table S9).

Linkage disequilibrium (LD) was evaluated by calculating the Z_{NS} statistic [64] which is the average of R^2 over all pairwise comparisons. A lineal regression between LD (R^2) and nucleotide distances was performed to evaluate whether recombination was taking place in *Pvmsp-7* genes. Recombination events were assessed using DnaSP v.5 software [57] applying the ZZ statistic [65] and RM parameter [66]. The latter statistic incorporates the effective population size and probability of recombination between adjacent nucleotides per generation. RDP3 v3.4 software [67] was used for detecting recombination regions in *msp-7* genes. This tool looks for evidence of recombination among aligned sequences by examining all possible triplet combinations following a scanning approach with a range of different recombination detection algorithms. Additionally, the algorithm developed by Betrán *et al.* (1997) [68] incorporated in DnaSP [57] and the GENECONV program [69] incorporated in RDP3 v3.4 software [67] were used to detect gene conversion among paralogous genes.

Geographical Clustering

Maximum Likelihood (ML) trees describing the phylogenetic consequences of the recombination events were constructed using RDP3 v3.4 software [67] with the HKY model (selected by the ModelTest algorithm [70]) to evaluate relationships between polymorphism and the geographical distribution of the isolates. Additionally, *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I* sequences were aligned and trees were constructed by using ML with the TN93+G model selected by the ModelTest algorithm [70]. Bootstrap analysis (with 1,000 replicates each) was used for assigning confidence levels to branch nodes. Positions containing gaps as well as regions in the alignment that contained repeats were not taken into account in the phylogenetic analysis (Table S9).

Supporting Information

Figure S1 Geographical location of the study regions within Colombia. Amazonian region (samples from Calamar in the Guaviare department), Andean region (samples from Apartadó and El Bagre in the Antioquia department), Caribbean region (samples from Tierra Alta and Puerto Libertador in the Córdoba department), Orinoco region (samples from Tauramena in the Casanare department and Mapiripán in the Meta department), and Pacific region (samples from Istmina in the Chocó department and Tumaco in the Nariño department). Black dots on the map represent the areas from which patients came who donated the infected blood samples. 1: Puerto Libertador, 2: Tierra Alta, 3: Istmina, 4: Tumaco, 5: Apartado, 6: El Bagre, 7: Tauramena, 8: Mapiripán, 9: Calamar. (TIF)

Figure S2 Sliding window analysis of Tajima D and Fu & Li D* and F* statistics along the *Pvmsp-7C* (A), *Pvmsp-7H* (B) and *Pvmsp-7I* (C) genes. Bars [Tajima's D (blue), Fu & Li's D* (red) and F* (green)] below each figure represent the regions where the tests showed a significant deviation from the neutral expectation. 5'-end (*Pvmsp-7C*: nucleotides 1–390, *Pvmsp-7H*: nucleotides 1–471, *Pvmsp-7I*: nucleotides 1–525), central (*Pvmsp-7C*: nucleotides 391–717, *Pvmsp-7H*: nucleotides 472–771, *Pvmsp-7I*: nucleotides 526–789) and 3'-end (*Pvmsp-7C*: nucleotides 718–1,191, *Pvmsp-7H*: nucleotides 772–1,200, *Pvmsp-7I*: nucleotides 790–1,188). (PDF)

Figure S3 Sliding window analysis for ω rates (d_N/d_S) throughout the *Pvmsp-7C* (Blue), *Pvmsp-7H* (Red) and *Pvmsp-7I* (Green) genes. Discontinuity of the curves is due to gaps within the alignments which were not considered for the analysis. 5'-end (*Pvmsp-7C*: nucleotides 1–390, *Pvmsp-7H*: nucleotides 1–471, *Pvmsp-7I*: nucleotides 1–525), central (*Pvmsp-7C*: nucleotides 391–717, *Pvmsp-7H*: nucleotides 472–771, *Pvmsp-7I*: nucleotides 526–789) and 3'-end (*Pvmsp-7C*: nucleotides 718–1,191, *Pvmsp-7H*: nucleotides 772–1,200, *Pvmsp-7I*: nucleotides 790–1,188). (PDF)

Figure S4 The linkage disequilibrium (LD) plot for *P. vivax Pvmsp-7C* (A), *Pvmsp-7H* (B) and *Pvmsp-7I* (C). Trace line represents the regression line which declined as nucleotide distance increased suggesting that intragenic recombination was taking place in *msp-7* genes. (PDF)

Figure S5 ML trees describing the phylogenetic consequences of the intragenic recombination in *Pvmsp-7C*. A topology is inferred for each recombinant fragment, (A) from nucleotides 685 to 855, (B) from nucleotides 1,060 to 1,167 (excluding nucleotides 1,105 to 1,107) and from 18 to 266, and (C) from nucleotides 62 to 266. Isolates clustered without a clear geographical distribution. Sal-I: Salvador strain, AMA: Amazon, AND: Andean, CAR: Caribbean, ORI: Orinoco, PAC: Pacific. (TIF)

Figure S6 ML trees describing the phylogenetic consequences of the intragenic recombination in *Pvmsp-7H*. A topology is inferred for each recombinant fragment, (A) nucleotides 189 to 543 (excluding nucleotides 436 to 450 and nucleotides 484 to 486), (B) nucleotides 353 to 661 (excluding nucleotides 436 to 450 and 568 to 627), (C) nucleotides 425 to 655 (excluding nucleotides 436 to 450, 484 to 486 and 568 to 627), (D) nucleotides 500 to 1,057 (excluding nucleotides 568 to 627 and 772 to 774), (E) nucleotides 628 to 977 (excluding nucleotides 772 to 774), (F) nucleotides 630 to 997, (G) nucleotides 416 to 590 (excluding nucleotides 436 to 450 and 484 to 486), (H) nucleotides 472 to 531 (excluding nucleotides 484 to 486), and (I) nucleotides 482 to 646 (excluding nucleotides 568 to 627). Isolates clustered without a clear geographical distribution. Sal-I: Salvador strain, AMA: Amazon, AND: Andean, CAR: Caribbean, ORI: Orinoco, PAC: Pacific. (TIF)

Figure S7 ML trees describing the phylogenetic consequences of the intragenic recombination in *Pvmsp-7I*. A topology is inferred for each recombinant fragment identified, (A) nucleotides 19 to 187 and nucleotides 683 to 1,188, (B) nucleotides 683 to 849, (C) nucleotides 684 to 797, (D) nucleotides 19 to 358 and nucleotides 796 to 1,188, (E) nucleotides 798 to 1,105, (F) nucleotides 868 to 1,188, (G)

nucleotides 920 to 1,188, (H) nucleotides 733 to 1,044, (I) nucleotides 688 to 797, and (J) nucleotides 683 to 849. Isolates clustered without a clear geographical distribution. SAL-I: Salvador strain, AMA: Amazon, AND: Andean, CAR: Caribbean, ORI: Orinoco, PAC: Pacific. (TIF)

Figure S8 (A) Phylogenetic tree obtained by ML for *Pvmsp-7* sequences based on the TN93+G model, ignoring recombination. Three monophyletic groups are shown; the first groups clustered *Pvmsp-7H* (red) sequences, the second group clustered *Pvmsp-7I* (green) sequences and the third group clustered *Pvmsp-7C* (blue) sequences. (B-E) Trees describing phylogenetic consequences of some gene conversion tracks identified. (B) nucleotides 71 to 265 (amino acids 24 to 89), (C) nucleotides 913 to 1,090 (amino acids 305 to 390), (D) nucleotides 913 to 1,090 (amino acids 305 to 354), (E) nucleotides 913 to 1,175 (amino acids 305 to 392). Positions are numbered according to the alignment in Fig. S13. These topologies suggest that at least *Pvmsp-7C*, *Pvmsp-7H* and *Pvmsp-7I* genes did not evolve independently. Numbers represent bootstrap values with 1,000 replicates. (PDF)

Figure S9 Schematic representation of gene conversion tracks identified by DnaSP and GENECONV for *Pvmsp-7C* (blue), *Pvmsp-7H* (red) and *Pvmsp-7I* (green) as a combined data set. Rectangles in a different color are graphical representations of sequence fragments that have potentially been originated by gene conversion and are localized at the conserved 5'- and 3'-ends. The black bars delimit the 5'-end, the central region and the 3'-end, [5'-end (*Pvmsp-7C*: nucleotides 1–390, *Pvmsp-7H*: nucleotides 1–471, *Pvmsp-7I*: nucleotides 1–525), central (*Pvmsp-7C*: nucleotides 391–717, *Pvmsp-7H*: nucleotides 472–771, *Pvmsp-7I*: nucleotides 526–789) and 3'-end (*Pvmsp-7C*: nucleotides 718–1,191, *Pvmsp-7H*: nucleotides 772–1,200, *Pvmsp-7I*: nucleotides 790–1,188)]. (PDF)

Figure S10 Alignment of deduced PvMSP-7C amino acid sequences. (PDF)

Figure S11 Alignment of deduced PvMSP-7H amino acid sequences. (TIF)

Figure S12 Alignment of deduced PvMSP-7C amino acid sequences. (TIF)

Figure S13 Alignment of the paralogous PvMSP-7 deduced amino acid sequences. (PDF)

Table S1 Nucleotide diversity for *P. vivax* antigens. n: number of isolates, π : nucleotide diversity. (PDF)

Table S2 Nucleotide diversity (π) values for subpopulations within Colombia. (PDF)

Table S3 Nucleotide and amino acid positions within the 5'- end, central region and 3'-end. (PDF)

Table S4 Positively selected sites detected for *Pvmsp-7* genes without taking recombination into account. Numbers according to the reference Sal-I protein sequence *Pvmsp-7C*:

XP_001614132.1, *Pvmsp-7H*: XP_001614137.1 and *Pvmsp-7I*: XP_001614138.1. (PDF)

Table S5 Negatively selected sites detected for *Pvmsp-7* genes without taking recombination into account. Numbers according to the reference Sal-I protein sequence *Pvmsp-7C*: XP_001614132.1, *Pvmsp-7H*: XP_001614137.1 and *Pvmsp-7I*: XP_001614138.1. (PDF)

Table S6 Positively selected sites detected for *Pvmsp-7* genes taking recombination into account. Numbers according to the reference Sal-I protein sequence *Pvmsp-7C*: XP_001614132.1, *Pvmsp-7H*: XP_001614137.1 and *Pvmsp-7I*: XP_001614138.1. (PDF)

Table S7 Negatively selected sites detected for *Pvmsp-7* genes taking recombination into account. Numbers according to the reference Sal-I protein sequence *Pvmsp-7C*: XP_001614132.1, *Pvmsp-7H*: XP_001614137.1 and *Pvmsp-7I*: XP_001614138.1. (PDF)

References

- Guerra CA, Howes RE, Patil AP, Gething PW, Van Boeckel TP, et al. (2010) The international limits and population at risk of Plasmodium vivax transmission in 2009. *PLoS Negl Trop Dis* 4: e774.
- Lysenko AJ, Semashko IN (1968) Medical geography: a medical-geographical study of an ancient disease. WHO Publisher's Code: 911.3: 616.936(100).
- Poespoprodjo JR, Fobia W, Kenangalem E, Lampah DA, Hasanuddin A, et al. (2009) Vivax malaria: a major cause of morbidity in early infancy. *Clin Infect Dis* 48: 1704–1712.
- Tjitra E, Anstey NM, Sugiarto P, Warikar N, Kenangalem E, et al. (2008) Multidrug-resistant Plasmodium vivax associated with severe and fatal malaria: a prospective study in Papua, Indonesia. *PLoS Med* 5: e128.
- Price RN, Tjitra E, Guerra CA, Yeung S, White NJ, et al. (2007) Vivax malaria: neglected and not benign. *Am J Trop Med Hyg* 77: 79–87.
- Kochar DK, Das A, Kochar SK, Saxena V, Sirohi P, et al. (2009) Severe Plasmodium vivax malaria: a report on serial cases from Bikaner in northwestern India. *Am J Trop Med Hyg* 80: 194–198.
- Carvalho LJ, Daniel-Ribeiro CT, Goto H (2002) Malaria vaccine: candidate antigens, mechanisms, constraints and prospects. *Scand J Immunol* 56: 327–343.
- Galinski MR, Barnwell JW (2008) Plasmodium vivax: who cares? *Malar J* 7 Suppl 1: S9.
- O'Donnell RA, de Koning-Ward TF, Burt RA, Bockarie M, Reeder JC, et al. (2001) Antibodies against merozoite surface protein (MSP)-1(19) are a major component of the invasion-inhibitory response in individuals immune to malaria. *J Exp Med* 193: 1403–1412.
- Genton B, Reed ZH (2007) Asexual blood-stage malaria vaccine development: facing the challenges. *Curr Opin Infect Dis* 20: 467–475.
- Takala SL, Plowe CV (2009) Genetic diversity and malaria vaccine design, testing and efficacy: preventing and overcoming 'vaccine resistant malaria'. *Parasite Immunol* 31: 560–573.
- Figtree M, Pasay CJ, Slade R, Cheng Q, Cloonan N, et al. (2000) Plasmodium vivax synonymous substitution frequencies, evolution and population structure deduced from diversity in AMA 1 and MSP 1 genes. *Mol Biochem Parasitol* 108: 53–66.
- Gomez A, Suarez CF, Martinez P, Saravia C, Patarroyo MA (2006) High polymorphism in Plasmodium vivax merozoite surface protein-5 (MSP5). *Parasitology* 133: 661–672.
- Putapornitip C, Udomsangpet R, Pattanawong U, Cui L, Jongwutiwes S (2010) Genetic diversity of the Plasmodium vivax merozoite surface protein-5 locus from diverse geographic origins. *Gene* 456: 24–35.
- Richie TL, Saul A (2002) Progress and challenges for malaria vaccines. *Nature* 415: 694–701.
- Pachebat JA, Ling IT, Grainger M, Trucco C, Howell S, et al. (2001) The 22 kDa component of the protein complex on the surface of Plasmodium falciparum merozoites is derived from a larger precursor, merozoite surface protein 7. *Mol Biochem Parasitol* 117: 83–89.
- Trucco C, Fernandez-Reyes D, Howell S, Stafford WH, Scott-Finnigan TJ, et al. (2001) The merozoite surface protein 6 gene codes for a 36 kDa protein associated with the Plasmodium falciparum merozoite surface protein-1 complex. *Mol Biochem Parasitol* 112: 91–101.

Table S8 Conversion tracks identified by DnaSP and GENECONV, between *Pvmsp-7* genes. Nucleotides and amino acids based in alignment of the Fig. S13. (PDF)

Table S9 Nucleotides and amino acid positions excluded from the analysis. (PDF)

Acknowledgments

We would like to thank the Instituto Nacional de Salud (INS), Bogotá, Colombia, for supplying the samples. We would like to thank Jason Garry for translating the manuscript. Our special gratitude goes to Francy Homez Homez and Pamela Jory Rios for their technical assistance. We would also like to thank Professor Manuel Elkin Patarroyo for his invaluable comments and suggestions. In loving memory of Doctor Carlos Fernando Cortés, who passed away recently.

Author Contributions

Conceived and designed the experiments: DG-O MAP. Performed the experiments: DG-O JF-R CL. Analyzed the data: DG-O JF-R CL. Wrote the paper: DG-O MAP.

- Garzon-Ospina D, Cadavid LF, Patarroyo MA (2010) Differential expansion of the merozoite surface protein (msp)-7 gene family in Plasmodium species under a birth-and-death model of evolution. *Mol Phylogenet Evol* 55: 399–408.
- Tewari R, Ogun SA, Gunaratne RS, Crisanti A, Holder AA (2005) Disruption of Plasmodium berghei merozoite surface protein 7 gene modulates parasite growth in vivo. *Blood* 105: 394–396.
- Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, et al. (2008) The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. *Proc Natl Acad Sci U S A* 105: 16290–16295.
- Mello K, Daly TM, Long CA, Burns JM, Bergman LW (2004) Members of the merozoite surface protein 7 family with similar expression patterns differ in ability to protect against Plasmodium yoelii malaria. *Infect Immun* 72: 1010–1018.
- Mello K, Daly TM, Morrissey J, Vaidya AB, Long CA, et al. (2002) A multigene family that interacts with the amino terminus of plasmodium MSP-1 identified using the yeast two-hybrid system. *Eukaryot Cell* 1: 915–925.
- Mongui A, Perez-Leal O, Soto SC, Cortes J, Patarroyo MA (2006) Cloning, expression, and characterisation of a Plasmodium vivax MSP7 family merozoite surface protein. *Biochem Biophys Res Commun* 351: 639–644.
- Garcia Y, Puentes A, Curtidor H, Cifuentes G, Reyes C, et al. (2007) Identifying merozoite surface protein 4 and merozoite surface protein 7 Plasmodium falciparum protein family members specifically binding to human erythrocytes suggests a new malarial parasite-redundant survival mechanism. *J Med Chem* 50: 5665–5675.
- Kadekoppala M, O'Donnell RA, Grainger M, Crabb BS, Holder AA (2008) Deletion of the Plasmodium falciparum merozoite surface protein 7 gene impairs parasite invasion of erythrocytes. *Eukaryot Cell* 7: 2123–2132.
- Wang L, Crouch L, Richie TL, Nhan DH, Coppel RL (2003) Naturally acquired antibody responses to the components of the Plasmodium falciparum merozoite surface protein 1 complex. *Parasite Immunol* 25: 403–412.
- Kadekoppala M, Holder AA (2010) Merozoite surface proteins of the malaria parasite: the MSP1 complex and the MSP7 family. *Int J Parasitol* 40: 1155–1161.
- Tetteh KK, Stewart LB, Ochola LI, Amambua-Ngwa A, Thomas AW, et al. (2009) Prospective identification of malaria parasite genes under balancing selection. *PLoS One* 4: e5568.
- Roy SW, Weedall GD, da Silva RL, Polley SD, Ferreira MU (2009) Sequence diversity and evolutionary dynamics of the dimorphic antigen merozoite surface protein-6 and other Msp genes of Plasmodium falciparum. *Gene* 443: 12–21.
- Garzon-Ospina D, Romero-Murillo L, Tobon LF, Patarroyo MA (2011) Low genetic polymorphism of merozoite surface proteins 7 and 10 in Colombian Plasmodium vivax isolates. *Infect Genet Evol* 11: 528–531.
- Chen JH, Jung JW, Wang Y, Ha KS, Lu F, et al. (2010) Immunoproteomics profiling of blood stage Plasmodium vivax infection by high-throughput screening assays. *J Proteome Res* 9: 6479–6489.
- Anisimova M, Nielsen R, Yang Z (2003) Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164: 1229–1236.
- Carlton JM, Adams JH, Silva JC, Bidwell SL, Lorenzi H, et al. (2008) Comparative genomics of the neglected human malaria parasite Plasmodium vivax. *Nature* 455: 757–763.

34. Nielsen R (2005) Molecular signatures of natural selection. *Annu Rev Genet* 39: 197–218.
35. Kimura M (1983) The neutral theory of molecular evolution. Cambridge Cambridge; New York: Cambridge University Press. xv, 367 p.
36. Suarez CF, Patarroyo ME, Trujillo E, Estupinan M, Baquero JE, et al. (2006) Owl monkey MHC-DRB exon 2 reveals high similarity with several HLA-DRB lineages. *Immunogenetics* 58: 542–558.
37. Martinez P, Suarez CF, Gomez A, Cardenas PP, Guerrero JE, et al. (2005) High level of conservation in *Plasmodium vivax* merozoite surface protein 4 (PvMSP4). *Infect Genet Evol* 5: 354–361.
38. Putaporntip C, Jongwutiwes S, Ferreira MU, Kanbara H, Udomsangpetch R, et al. (2009) Limited global diversity of the *Plasmodium vivax* merozoite surface protein 4 gene. *Infect Genet Evol* 9: 821–826.
39. Martinez P, Suarez CF, Cardenas PP, Patarroyo MA (2004) *Plasmodium vivax* Duffy binding protein: a modular evolutionary proposal. *Parasitology* 128: 353–366.
40. Mascorro CN, Zhao K, Khuntirat B, Sattabongkot J, Yan G, et al. (2005) Molecular evolution and intragenic recombination of the merozoite surface protein MSP-3alpha from the malaria parasite *Plasmodium vivax* in Thailand. *Parasitology* 131: 25–35.
41. Putaporntip C, Jongwutiwes S, Seethamchai S, Kanbara H, Tanabe K (2000) Intragenic recombination in the 3' portion of the merozoite surface protein 1 gene of *Plasmodium vivax*. *Mol Biochem Parasitol* 109: 111–119.
42. Orsi RH, Ripoll DR, Yeung M, Nightingale KK, Wiedmann M (2007) Recombination and positive selection contribute to evolution of *Listeria monocytogenes* inLA. *Microbiology* 153: 2666–2678.
43. Andrews TD, Gojabori T (2004) Strong positive selection and recombination drive the antigenic variation of the PilE protein of the human pathogen *Neisseria meningitidis*. *Genetics* 166: 25–32.
44. Polley SD, Conway DJ (2001) Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. *Genetics* 158: 1505–1512.
45. Peek AS, Souza V, Eguiarte LE, Gaut BS (2001) The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (fimA) from *Escherichia coli*. *J Mol Evol* 52: 193–204.
46. Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39: 121–152.
47. Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci U S A* 97: 10866–10871.
48. Semple C, Wolfe KH (1999) Gene duplication and gene conversion in the *Caenorhabditis elegans* genome. *J Mol Evol* 48: 555–564.
49. Katju V, Lynch M (2003) The structure and early evolution of recently arisen gene duplicates in the *Caenorhabditis elegans* genome. *Genetics* 165: 1793–1803.
50. Wochlbier U, Epp C, Hackett F, Blackman MJ, Bujard H (2010) Antibodies against multiple merozoite surface antigens of the human malaria parasite *Plasmodium falciparum* inhibit parasite maturation and red blood cell invasion. *Malar J* 9: 77.
51. Cifuentes G, Bermudez A, Rodriguez R, Patarroyo MA, Patarroyo ME (2008) Shifting the polarity of some critical residues in malarial peptides' binding to host cells is a key factor in breaking conserved antigens' code of silence. *Med Chem* 4: 278–292.
52. Patarroyo ME, Patarroyo MA (2008) Emerging rules for subunit-based, multiantigenic, multistage chemically synthesized vaccines. *Acc Chem Res* 41: 377–386.
53. Imwong M, Pukrittayakamee S, Gruner AC, Renia L, Letourneur F, et al. (2005) Practical PCR genotyping protocols for *Plasmodium vivax* using Pvcs and Pvmsp1. *Malar J* 4: 20.
54. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
55. Jorda J, Kajava AV (2009) T-REKS: identification of Tandem REpeats in sequences with a K-means based algorithm. *Bioinformatics* 25: 2632–2638.
56. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34: W609–612.
57. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.
58. Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915–925.
59. Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* 95: 3708–3713.
60. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
61. Delpont W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* 26: 2455–2457.
62. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, et al. (2006) Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol* 2: e62.
63. Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22: 1208–1222.
64. Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics* 146: 1197–1206.
65. Rozas J, Gullaud M, Blandin G, Aguade M (2001) DNA variation at the rp49 gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. *Genetics* 158: 1147–1155.
66. Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111: 147–164.
67. Martin DP, Lemey P, Lott M, Moulton V, Posada D, et al. (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26: 2462–2463.
68. Betran E, Rozas J, Navarro A, Barbadilla A (1997) The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* 146: 89–99.
69. Sawyer S (1989) Statistical tests for detecting gene conversion. *Mol Biol Evol* 6: 526–538.
70. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.