

# Análisis comparativo del gen P de Rinderpest Virus (RPV)

VAGNOZZI, A; CARRILLO, C.

## RESUMEN

Rinderpest Virus (RPV) es el agente causal de una severa enfermedad del ganado doméstico. Es un virus envuelto, cuyo genoma RNA no segmentado, de cadena simple y negativa (-ssRNA), está organizado en seis genes no superpuestos. El gen P es el más complejo de dichos genes, codificando una proteína estructural (P) y dos no estructurales (C y V). La proteína P es esencial para la replicación viral y está involucrada en todos los aspectos del ciclo viral. En cambio, las proteínas C y V son consideradas no esenciales para la replicación, al menos *in vitro*. En este trabajo hemos secuenciado el gen P de 16 aislamientos diferentes de RPV, y realizado el análisis comparativo de las secuencias obtenidas, poniendo el énfasis en las regiones conservadas en relación a su importancia funcional. Rinderpest ha sido una de las amenazas más importantes para la salud del ganado doméstico. En la actualidad la enfermedad se presenta en ciertas regiones de Asia y está en vías de ser erradicada del planeta. La secuenciación y el análisis comparativo realizado en este trabajo es una herramienta sumamente útil y eficiente tanto para seguimiento epidemiológico del virus como para la evaluación de su biología.

Foreign Animal Disease Research Unit, United States Department of Agriculture, Agricultural Research Service, Plum Island Animal Disease Center, Greenport, NY, USA.

USDA, ARS, Plum Island Animal Disease Center, PO Box 848. Greenport, NY 11944-0848

Email: aevagnozzi@piadc.ars.usda.gov

Este trabajo mereció el premio Estímulo a la Investigación Científica, 2005, otorgado por la Facultad de Ciencias Veterinarias, UBA.

InVet. 2006, 8(1): 67-91  
ISSN(papel): 1514-6634  
ISSN (on line) 1668-3498

67

## Comparative analysis of Rinderpest Virus (RPV) P gene

### SUMMARY

Rinderpest Virus (RPV) is the causative agent of a serious disease of domestic cattle. RPV is an enveloped virus with a single stranded, non-segmented, negative RNA genome (-ssRNA), that is organized in six non-overlapping genes. P gene is the most complex of such genes. Its sequence codifies for a structural (P) and two non-structural proteins (C and V). The P protein is essential for viral replication and is involved in every aspect of the viral cycle, whereas, C and V proteins are non essential for viral replication, at least *in vitro*. Here we have made a comparative analysis of sequences of the P gene and its codified proteins from 16 different isolations of RPV, focusing in conserved regions and its functional role. Rinderpest has been one of the most important threats to the cattle health. Actually, the disease is present just in some regions of Asia. The present comparative genomic analysis will be useful as epidemiological work and an important contribution to the current knowledge of RPV biology.

### INTRODUCCIÓN

Rinderpest Virus (RPV) es el agente causante de una enfermedad aguda, febril y altamente contagiosa, que en su presentación clásica puede alcanzar valores de morbilidad y mortalidad tan altos como 90%. Durante los siglos XVIII y XIX Rinderpest se expandió sobre Europa y África generando enormes perjuicios económicos<sup>64</sup>. Basta decir que en Europa entre los años 1711 y 1769, cerca de 10000 animales murieron por día debido a RPV<sup>14</sup>. Actualmente, la enfermedad no tiene la presentación espectacular señalada por las crónicas históricas, sin embargo el virus posee el potencial para generar un impacto económico muy grande<sup>63</sup>. La enfermedad afecta a bovinos, búfalos, y yaks, también puede afectar a ovinos, caprinos, ciertas razas de cerdos, y un gran número de animales de la fauna silvestre, aunque la infección en esos animales usualmente cursa en forma subclínica<sup>2; 3; 63</sup>.

RPV pertenece al genero morbillivirus, familia *Paramyxoviridae*<sup>11</sup>. Es un virus envuelto y su genoma es RNA de cadena simple y negativa

(-ssRNA), que codifica seis proteínas estructurales; **N** (*Nucleocapsid protein*), **P** (*Phosphoprotein*), **M** (*Matrix protein*), **F** (*Fusion protein*), **H** (*Haemagglutinin protein*), y **L** (*Large protein*)<sup>49</sup>, y dos no estructurales, **C** y **V**. El genoma, de 15882 bases, presenta dos regiones no codificantes en los extremos; la región *leader* (55 nucleótidos) en el extremo 3', y la región *trailer* (37 nucleótidos) en el 5', cuya importancia funcional no ha sido completamente esclarecida<sup>49</sup>.

El gen P de RPV, al igual que en el resto de los virus pertenecientes a la familia Paramyxoviridae, es policistrónico<sup>28</sup>. El inicio de este gen se encuentra en la posición 1748 del genoma viral y la señal de poliadenilación en 3402. El primer codon de inicio (AUG) señala el marco de lectura para la proteína estructural P (507 aa), y 20 nucleótidos adelante en sentido 3' a 5' se encuentra el segundo AUG, el cual abre el marco de lectura alternativo para la proteína C (177 aa). La proteína V (299 aa) es producto de la modificación post-transcripcional del RNA mensajero, el cual consiste en la adición de una guanidina en un

sitio específico (sitio de edición), de modo que el nuevo mRNA traduce una proteína cuyo extremo N-terminal es similar a P pero con un C-terminal propio de V y caracterizado por ser rico en residuos de cisteína<sup>49; 58; 71</sup>. P representa un excelente ejemplo de cómo un virus puede compactar tanta información como sea posible en un único gen<sup>49</sup>. Tal potencial de codificación diferencial ha sido utilizado como criterio de clasificación en Paramyxoviridae<sup>71</sup>. Por otra parte, se ha sugerido que el principal determinante de patogenicidad específica de RPV reside en este gen<sup>85</sup>.

La proteína P es el único producto del gen P esencial para la replicación del virus y está involucrado en todos los aspectos del ciclo viral<sup>49</sup>. Dicha proteína es sintetizada en exceso en las células infectadas, y muestra una tasa de renovación alta, presentando funciones transitorias durante el ensamblado de la nucleocápside y la síntesis de RNA<sup>71</sup>. La proteína P unida a L, integra el complejo de polimerasa (RNA polimerasa dependiente de RNA) encargado de la síntesis de RNA (transcripción y replicación)<sup>21; 71</sup>. La proteína L se asocia al RNA «encapsidado» (N-RNA) a través de su interacción con P, que actúa como enlace o adaptador, para formar el complejo de transcripción de ribonucleoproteína (RNP)<sup>49</sup>. La proteína P cumple múltiples roles durante la infección viral: (i) Actúa como trans-activador transcripcional al reclutar proteína L sobre el templado formado por proteína N y RNA viral<sup>22</sup>; (ii) Se une al N-RNA, independientemente de su rol como parte del complejo de polimerasa, interactuando con regiones no-codificantes del genoma viral<sup>22</sup>; (iii) Previene la asociación espuria de N con RNA inespecífico, actuando como «chaperona», mediante la formación del complejo P-N soluble, que es además el precursor para la encapsidación de RNA viral recientemente sintetizado en la replicación<sup>21</sup>.

Estudios realizados con virus RNA negativos, (familias Rhabdoviridae y Paramyxoviridae) reportaron diferentes status de oligomerización para la proteína P, dímeros, trímeros y tetrámeros<sup>22</sup>. En el caso de RPV se ha demostrado que P existe como tetrámero en solución, y que el dominio de tetramerización asienta en el extremo C-terminal, el cual es variable en cuanto a secuencia de nucleótidos pero conservado en función de la estructura<sup>67</sup>. A su vez, se ha sugerido que P tiene diferentes requerimientos de fosforilación para la multimerización<sup>23</sup>, sin embargo se ha demostrado recientemente que la oligomerización de la proteína P de RPV ocurre independientemente de la fosforilación<sup>67</sup>.

Respecto de las proteínas no estructurales, se ha postulado que la proteína C participa en el proceso de transcripción de mRNA, cumpliendo un rol en la iniciación, extensión y terminación del mismo<sup>5</sup>. Numerosos trabajos, utilizando virus recombinantes donde esta anulada la expresión de C y/o V, muestran que la viabilidad viral en cultivo celular esta en menor o mayor medida conservada, sugiriendo que ambas proteínas son «no esenciales» para la replicación viral<sup>5; 82; 50; 44</sup>. Sin embargo, mutantes de Sendai virus (SeV) que no expresan C (y/o C') son prácticamente incapaces de replicar en ratones, y virtualmente no patogénicos para esa especie, demostrando que C es crítica para la replicación viral de SeV *in vivo*<sup>48</sup>.

La importancia funcional de la proteína V dista de estar completamente revelada. Se ha probado que la proteína V interfiere con la respuesta de interferón (IFN), particularmente la respuesta de IFN tipo I (IFN- $\alpha/\beta$ )<sup>4; 24; 49; 73; 78</sup>. Además, la sobre-expresión de esta proteína genera un efecto negativo sobre la replicación pero positivo sobre la transcripción<sup>44; 82</sup>.

Ciertos dominios de P interactúan con la proteína N, tanto monomérica (soluble) como

ensamblada, y dicha interacción con N monomérica es función de una secuencia que está en el extremo N-terminal, el cual es común con V. De este modo, al igual que P, también V está capacitada para unirse a la proteína N<sup>35</sup>. Esta interacción ha sido documentada en diferentes miembros de la familia Paramyxoviridae tales como; Respiravirus, Rubulavirus y Morbillivirus<sup>82</sup>. Se ha sugerido que V puede actuar como «chaperona» en el ensamblado de la nucleocápside<sup>21</sup>, y/o regular el cambio de la transcripción de mRNA (respetando las señales de inicio/parada entre los genes) hacia la síntesis del RNA antigénico (ignorando las señales Intergénicas), el cual servirá como templatado para la producción del RNA genómico<sup>35</sup>.

Se ha mencionado que P interactúa con L a través de un área que se encuentra en C-terminal de P, el cual no está presente en V. en<sup>36; 75</sup>. En adición, se demostró que la interacción entre V y L de RPV existe, pero a través de un motivo diferente al identificado para P<sup>77</sup>.

El dominio N-terminal (común a P), ha sido identificado como un regulador de la replicación del genoma<sup>5; 19; 82</sup>. La diferencia fundamental de V se encuentra en el extremo C-terminal, que presenta un típico motivo de unión a zinc (*zinc-finger*)<sup>62</sup>, que es responsable de la unión a proteínas celulares del hospedador tal como la proteína de Unión Específica a DNA Dañado (DDB)<sup>51</sup>. La proteína V de SV5 produce una disminución en la velocidad del ciclo celular mediante la interacción con esta Proteína<sup>50; 51</sup>. Dado que durante la mitosis el aparato de Golgi presenta una fragmentación que inhibe la fusión de las vesículas y el transporte de las glicoproteínas<sup>54</sup>, el ensamblado viral y por ende la patogénesis se verían beneficiados por la menor velocidad del ciclo celular provocada por V<sup>50</sup>. Se ha postulado que un importante número de proteínas celulares interactúan con C y V

además de DDB, sin embargo el rol que cumplen las proteínas celulares en la biología de morbillivirus debe ser explorado aún<sup>5</sup>. El conocimiento de secuencias, motivos o estructuras conservadas entre diferentes variantes virales de RPV es un elemento importante para el desarrollo de nuevas investigaciones que permitan dilucidar tales interacciones.

RPV ha sido dividido en tres linajes diferentes pertenecientes a un único serotipo que está antigénica y genéticamente relacionado con los otros miembros del género morbillivirus, incluyendo Sarampión virus (MV), Peste de los pequeños rumiantes virus (PPRV), y Distemper canino virus (CDV)<sup>61</sup>. En este trabajo hemos llevado a cabo la secuenciación y análisis comparativo del gen P de 16 aislamientos de RPV (tabla 1a), los cuales representan los tres linajes reconocidos.

## MATERIALES Y MÉTODOS

**Virus:** Todos los aislamientos (Tabla 1a) fueron suministrados por el Servicio de Sanidad Animal y Vegetal (APHIS según sus siglas en inglés), del Departamento de Agricultura de los Estados Unidos (USDA). Para algunos análisis comparativos las secuencias disponibles en el GenBank fueron consideradas (Tabla 1b).

**Extracción de RNA, RT PCR y secuenciación:** Las secuencias del gen P de todos los aislamientos fueron obtenidas mediante la transcripción reversa del RNA genómico viral, seguido de amplificación redundante y secuenciación de fragmentos superpuestos que cubrieron todo el gen P. El RNA total fue extraído a partir de 140 µl de sobrenadante proveniente de cultivo celular infectado utilizando *RNeasy™ mini kit* (Qiagen, Valencia, CA), y resuspendido in 40µl de ddH<sub>2</sub>O. El genoma viral fue transcrito a cDNA usando

Tabla 1a: lista de los aislamientos secuenciados

| Nombre del aislamiento | Origen geográfico   | Fuente                 | Historia del aislamiento |
|------------------------|---------------------|------------------------|--------------------------|
| Kabete O               | Kabete, Kenya       | <i>Syncerus caffer</i> | Virus de referencia      |
| RBok                   | Deriva de Kabete O  | <i>Cultivo Celular</i> | Cepa Vacunal             |
| Pak Chong              | Pak Chong, Thailand | <i>Bubalus bubalis</i> | 2 pasajes CC             |
| Pendik                 | Pendik, Turquía     | -                      | 35 pasajes CC            |
| RBT                    | Tanzania            | <i>Bos taurus</i>      | 5 pasajes CC             |
| Saudita                | Arabia Saudita      | -                      | 2 pasajes en VERO        |
| Kuwait 21              | Kuwait              | -                      | 3 pasajes en VERO        |
| Nigeria Buffalo        | Nigeria             | <i>Syncerus caffer</i> | 2 pasajes en VERO        |
| Yemen                  | Yemen               | -                      | 1 pasaje en VERO         |
| India                  | India               | -                      | 2 pasajes en VERO        |
| Nigeria Sokoto         | Sokoto, Nigeria     | -                      | 1 pasaje en VERO         |
| Kuwait                 | Kuwait              | -                      | 2 pasajes en VERO        |
| Sokoto                 | Sokoto, Nigeria     | <i>Bos spp.</i>        | 2 pasajes en BK          |
| Pakistán               | Pakistán            | -                      | 1 pasaje en VERO         |
| Nigeria vero           | Sokoto, Nigeria     | -                      | 6 pasajes en VERO        |
| Egipto                 | Egipto              | -                      | 3 pasajes en VERO        |

Nota: CC: cultivo celular

Tabla 1b: Secuencias tomadas del Genbank

| Nombre   | Numero de acceso al Genbank | Variante  |
|----------|-----------------------------|-----------|
| RVPCMFH  | Z30697                      | RBOK      |
| RVNPVCA  | X68311 S54798               | RBOK      |
| AB021977 | AB021977                    | Lapinized |
| S44819   | S44819                      | Kabete O  |
| RVNPVCA  | X68311 S54798               | RBOK      |

**Invitrogen Super Script III retrotranscriptase<sup>TM</sup>** (Invitrogen. Carlsbad. CA) y **Super Smart<sup>TM</sup> cDNA synthesis kit** (Clontech. Mountain View, CA) siguiendo las indicaciones del productor pero utilizando oligonucleótidos del tipo **Random Hexamers** (Invitrogen. Carlsbad. CA) y otros diseñados especialmente para ser usado como *primers*. Ciertas regiones inter-cistrónicas presentaron dificultad para la amplificación probablemente debido a una estructura secundaria compleja.

Para superar este inconveniente se realizó un paso adicional de desnaturalización previo a la retro-transcripción. Dicho paso consistió en calentar el RNA junto a la mezcla de dNTPs y los *primers* a 95° C por 5 minutos. Todos los cDNA fueron purificados mediante sistema de columna previo a la amplificación por PCR. Para las PCRs se utilizaron diferentes pares de *primers* degenerados (Tabla 2). Las reacciones de PCR se llevaron a cabo utilizando *taq*Polimerasa (Clontech. Mountain View, CA)

mediante 35 ciclos del siguiente programa: Desnaturalización: 30 segundos a 95° C. *Annealing*: 5 segundos a 65° C. Extensión: 3 minutos a 68° C. O bien mediante el denominado «*touch down*», el cual consistió del mismo programa precedido de 5 ciclos de: Desnaturalización: 15 segundos a 95° C. Extensión: 3 minutos a 70° C. De este modo se lograron productos (amplicones) de 900 a 1500 bases aproximadamente, que superponiéndose cubrieron la totalidad del gen P. Esos amplicones fueron purificados con *StrataPrep® PCR purification kit* (Stratagene, Cedar Creek, TX), y secuenciados (2.5 µl de templado por reacción) utilizando *Big Dye Terminator Cycle Sequencing kits* (Applied Biosystems, Foster City, CA) con *primers* (sense y anti-sense), específicos de posición mediante secuenciador automatizado *PRISM 3700* (Applied Biosystems, Foster City, CA). Con el fin de superar los errores que pudieran ser introducidos en el análisis como consecuencia de la variabilidad viral y el error propio de la reacción de PCR y secuenciación, cada amplificación fue repetida al menos dos veces, y se realizaron múltiples secuenciaciones de cada amplicón a través de la lectura en el sentido positivo y negativo del producto. Adicionalmente se realizó el clonado y posterior

secuenciación de los productos de PCR, para lo cual se utilizó el sistema *TOPO TA Cloning® kit* (Invitrogen, Carlsbad, CA). Aún cuando se detectaron polimorfismos en algunas posiciones de nucleótidos, cada una de las mismas fue asignada con exactitud (evitando cualquier tipo de ambigüedad) a un único nucleótido debido al alto grado de redundancia generado por la estrategia de secuenciación.

**Análisis de las secuencias:** Las bases fueron determinadas desde un cromatograma mediante *PHRED*<sup>26</sup>, el cual también permitió conocer la probabilidad de error predicha para cada posición. Las secuencias virales fueron ensambladas con los programas de ensamblado *PHRAP* y con *CAP 3*<sup>37</sup>. Para el cierre de las áreas sin definir (gaps) se procedió siguiendo pautas previamente descriptas<sup>1</sup>. Los alineamientos múltiples fueron realizados mediante los programas *CLUSTALW 1.7* y *DIALIGN 2.2*<sup>55, 81</sup>. El análisis de nucleótidos por sustitución fue llevado a cabo usando los programas *DISTREE 1.2*<sup>70</sup>, *DNARATES 1.1*<sup>79</sup>, *ALISTAT*<sup>25</sup>, y *PRETTY* (incluido en el paquete *GCG 10*)<sup>25</sup>. Para el análisis de codones, mediante el cálculo del índice de sustitución de sinónimos / no-sinónimos se usó *SNAP*<sup>59</sup>, *CODONW* (<http://www.molbiol.ox.ac.uk/cu/>), y *CODEML* (paquete de software *PAML 3.14*) el

Tabla 2: Primers usados en las amplificaciones

| Área del genoma | Sense                           | Antisense                      |
|-----------------|---------------------------------|--------------------------------|
| 1080-2175       | AGTGCAGGAGCATAYCCCYTGTTGTGGAGC  | TCTTCGACTCCCTCAACCTTTTCACCGCTG |
| 1605-2571       | GCYATGGCCRCATCYTAGRGGAYTCAACC   | AGGYYGGGTGCACCACCTGTTCGATGAA   |
| 1875-2984       | CCTYGAYYCMCWGGTCGTTGARGAAGCCC   | TCAGGGTTGATRTCACATCKGCAGTGGGR  |
| 2423-3535       | CRGGRAAGACHCTTRTGGTGCCGCMGAT    | TGRGGAATCAATCGGCCGTCAGTGTATGTT |
| 2686-4220       | TCKCRGATRCCCAGARTAACAYWRRMMT    | YCTTATGTGAAGGCTTGTTCCCCCTATCCC |
| 3418-4444       | YCTTATGTGAAGGCTTGTTCCCCCTATCCC  | GCGRKRKGGRTGATGYTRBYGAGTRTT    |
| 4059-5145       | RRSRKTGRGSCCTTYCCGARGCGACATTYAT | GGGGTGCGTGGGGGGTGCTGCTGTTG     |
| 4355-5555       | GGCKGTCYTGAACCRTCAGTCCCTCA      | YTTCGAAAGRTTSCCCAGTGGATCTGGCC  |

cual fue utilizado también para la evaluación estadística de la presión de selección heterogénea presente en los residuos de aminoácidos<sup>84</sup>. Para el análisis de variabilidad de nucleótidos y aminoácidos se utilizó el programa **PRETTY**. La búsqueda de motivos y señales fue realizada mediante los programas **MOTIFS** (incluido en el paquete **GCG**), **HMM**<sup>47</sup>, **PFSCAN** (<http://www.isrec.isb-sib.ch/ftp-server/pftools/pft2.2.tar.Z>), y **BLOCKS**<sup>33</sup> provisto por **PROSITE** y la base de datos del GenBank. La predicción de la polaridad de grupos de aminoácidos, segmentos hidrofóbicos y segmentos trans-membrana fue realizada mediante los programas; **MEMSAT**<sup>42</sup>, **TMPRED**<sup>34</sup>, **TOPPED**<sup>83</sup>, **PSORT**<sup>57</sup>, y **SAPS**<sup>15</sup>. La predicción de la estructura secundaria de las proteínas fue realizada mediante los programas **GOR SECONDARY STRUCTURE PREDICTION**<sup>30</sup> y **PRATT**<sup>40;41</sup>.

**Análisis filogenético:** El análisis de filogenia se llevó a cabo sobre un alineamiento de secuencias del gen P, utilizando los métodos de *neighbor-joining* como ha sido implementado en **PHYLO-WIN**<sup>29</sup> ([http://www.ab.informatic.uni-tuebingen.de/software/jsplits/welcome\\_en.html](http://www.ab.informatic.uni-tuebingen.de/software/jsplits/welcome_en.html))<sup>38</sup>. También se utilizó el método de *maximun likelihood* mediante **PUZZLE**<sup>76</sup> y **DNAML** en el paquete **PHYLIP** (<http://evolution.genetics.washington.edu/phylip.html>)<sup>27</sup>.

## RESULTADOS

**Análisis de Gen P:** En la posición 1744 del genoma viral identificamos los nucleótidos **CTT**, los cuales han sido reconocidos previamente como secuencia intergénica<sup>11</sup> e inmediatamente después la secuencia **AGGACCCAGGTC**, que representa el inicio del gen P<sup>11</sup>, conservado en los 16 aislamientos. Este gen está constituido por 1657 nucleótidos,

de los cuales 1275 permanecen invariables en todos los aislamientos, ello representa un 77.04 % de conservación (Tabla 3). Previsiblemente, el área codificante es más conservada (78.08%) que las regiones no-codificantes de los extremos 5' y 3' (66.67 y 61.64 % respectivamente). Dada la naturaleza policistónica del gen P la predicción de variabilidad permite esperar una significativa conservación en la secuencia, y de hecho la proporción de transiciones sobre transversiones es relativamente elevada (Tabla 3), lo que sugiere que la mayoría de los cambios son conservadores. Sin embargo, 2 de cada 3 modificaciones dan lugar a la sustitución del aminoácido codificado (ver la proporción de «no sinónimos» por «sinónimos» de la Tabla 3).

Al analizar las identidades por pares (mediante **ALISTAT**), hemos observado que el promedio de las mismas para el gen P de los 16 aislamientos es de 93.00%. Siendo Nigeria Búfalo y Kuwait las secuencias menos relacionadas, con un porcentaje de identidad de 90.00%, lo cual demuestra una gran similitud entre los aislamientos analizados. Notablemente, las diferencias genéticas permiten separar a los aislamientos en grupos con mínima variabilidad intra-grupo. Mediante el método de *Neighbor-Joining* hemos construido un árbol filogenético que muestra gráficamente dicho agrupamiento (Fig. 3). De este modo, Kuwait y Kuwait 21 resultaron los aislamientos más alejados del grupo constituido por Nigeria Búfalo, Nigeria Vero, Nigeria Sokoto, Sokoto y Egipto. Mientras que Egipto mostró la menor identidad con Kuwait, Kabete O, R Bok, Saudi, India, Pendik, y Pak Chong. Notablemente, Egipto tuvo un 99.60 % de identidad con Nigeria Vero y 99.50 % con Nigeria Búfalo, a pesar de la disparidad geográfica de su origen. Kuwait presentó un 99.60 % de identidad con Saudi, 97,40 % con Pendik y 97,20 % con Pak Chong.

La cercana relación observada entre Nigeria Vero y Nigeria Búfalo (0.0007 KG), entre Kuwait y Kuwait 21, y entre Sokoto y Nigeria Sokoto (menor a 0.0001 KG), sugiere que se trata del mismo aislamiento producto de diferentes historias de pasajes. Lo mismo se puede decir de la corta distancia genética (0.0033 KG) entre el altamente virulento Kabete O (aislado de Búfalo en 1911, actualmente cepa referencia de la FAO), y la cepa vacunal R Bok que deriva de aquel luego de sucesivos pasajes por células BK (riñón bovino) y VERO (riñón de mono verde africano). También hemos observado una cercanía muy estrecha entre Nigeria Búfalo y Nigeria Vero con Egipto (0.0046 y 0.0040 KG respectivamente), y entre Saudi con Kuwait y Kuwait 21 (0.0040 y 0.0046 KG respectivamente). Por otro lado, los pares más distantes fueron Egipto con Kuwait 21 y Pakistán (0.1139 y 0.1130 KG respectivamente) sugiriendo una diferencia real en cuanto a linaje genético. Fue realmente notable la similitud entre R Bok e India (más de 99.99% de

identidad nucleotídica, y menos de 0.0001 KG de distancia genética), y entre Kabete O y R Bok (99.70% de identidad y 0.0033 KG de distancia genética). Estos resultados no contradicen descripciones previas <sup>7</sup>.

**Análisis de la proteína P:** La secuencia de aminoácidos de la proteína P derivada de la secuencia nucleotídica mostró que un 74.56 % de las posiciones resultaron invariantes (Tabla 4), lo que sugiere que al menos el 25.44 % de las posiciones pueden aceptar cambios sin afectar la funcionalidad. Esos cambios son más abundantes en el extremo N-terminal (PNT) que ha sido establecido desde la posición 1 hasta 291. En este trabajo, la conservación del dominio PNT ha sido estimada en 65.98 %, significativamente inferior al 84.26 % del extremo C-terminal (PCT), que abarca desde el residuo 291 al 507 (ver Tabla 4). El rol desempeñado por la proteína P requiere de importantes interacciones con el RNA viral, con las proteínas N, L, y consigo misma, mediante sitios que han sido localizados predominantemente en el dominio PCT (Fig.2)

Tabla 3: Análisis de nucleótidos del gen P de RPV comparando la secuencia de 16 aislamientos

| Producto  | Número de nucleótidos comparados | Número de nucleótidos invariantes | % de nucleótidos invariantes | ts/tv | dn/ds  |
|-----------|----------------------------------|-----------------------------------|------------------------------|-------|--------|
| Gen P     | 1655                             | 1275                              | 77.04                        | -     | -      |
| 5' P (NC) | 61                               | 40                                | 66.67                        | -     | -      |
| P         | 1524                             | 1190                              | 78.08                        | 8.22  | 0.4461 |
| C         | 534                              | 423                               | 79.21                        | 8.46  | 0.4633 |
| V         | 900                              | 707                               | 78.56                        | 9.46  | 0.5913 |
| 3' P (NC) | 73                               | 45                                | 61.64                        | -     | -      |

**Nota:** **5' P (NC):** Región 5' no-codificante del gen P. **P:** Región codificante para la proteína P. **C:** Región codificante para la proteína C. **V:** Región codificante para la proteína V. **3' P (NC):** Región 3' no-codificante del gen P. **ts/tv:** Proporción de transiciones sobre transversiones. **dn/ds:** Proporción de cambios «no-sinónimos» sobre «sinónimos». Programa **PRETTY GCG, DIALIGN, y ALISTAT.**



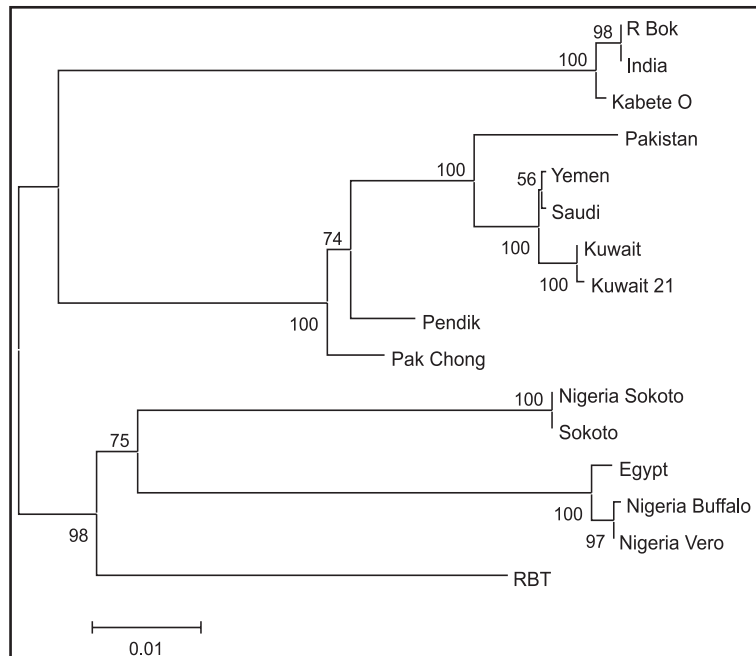


Figura 3: Filogenia de los aislamientos de RPV

Nota: Dendograma construido mediante el método de **Neighbor-Joining** usando **Kimura 2** parámetros con un bootstrap de 1000, provisto por **SPLITSTREE 4.0**

<sup>49</sup>. El sitio de multimerización ubicado entre los residuos 316 y 347 (Fig. 2), ha sido localizado en una región de la proteína caracterizada por el super-enrollamiento, la cual se extiende desde la posición 316 hasta la 382 <sup>67; 73</sup>. Esta legión, que es importante para la función del complejo de polimerasa <sup>67</sup>, muestra un 84.37% de posiciones invariantes. El dominio de unión a proteína L que ha sido «mapeado» entre los aminoácidos 347 y 390 <sup>18</sup> presenta un 90.91% de conservación. Esta área se superpone con aquella que ha sido propuesta como sitio de unión a la región *leader* del genoma (*leRNA*), la cual ha sido localizada entre los residuos 357 y 388 <sup>66</sup> y cuyo porcentaje de invariabilidad resultó 90.62 para los aislamientos secuenciados en este trabajo (Tabla 4). El dominio PNT, que contiene el sitio de unión a

N, es más variable (34.02 %) que el dominio PCT (15.74% de variabilidad), el cual soporta el resto de los sitios de interacción de la proteína P. El sitio de unión a proteína N se encuentra entre las posiciones 1 a 60 de la proteína P <sup>73</sup>, y tiene un 80.00% de posiciones conservadas en las secuencias analizadas, con la particularidad que los primeros 21 aminoácidos están conservados en un 100.00 %, indicando la importancia funcional del extremo N-terminal de este dominio, que implica restricción a la variabilidad de la estructura primaria del área. Notablemente, todo el sitio involucrado en la interacción con N tiene 12 posiciones variables, pero 7 de esas posiciones son cambios extremadamente conservativos que involucran el reemplazo de un aminoácido por otro de similares características.

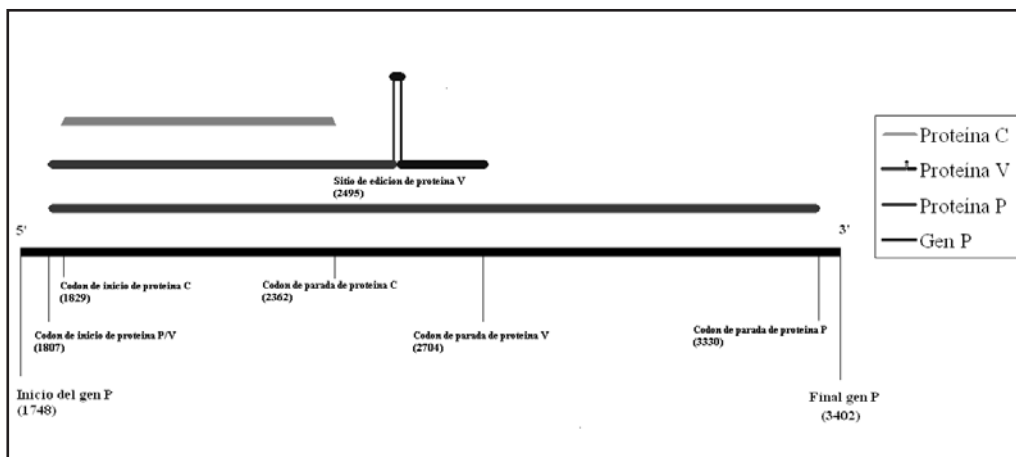


Figura 1: Esquema del gen P de RPV indicando marcos de lectura y sitio de edición.

Nota: Los números indican la posición en relación al genoma completo

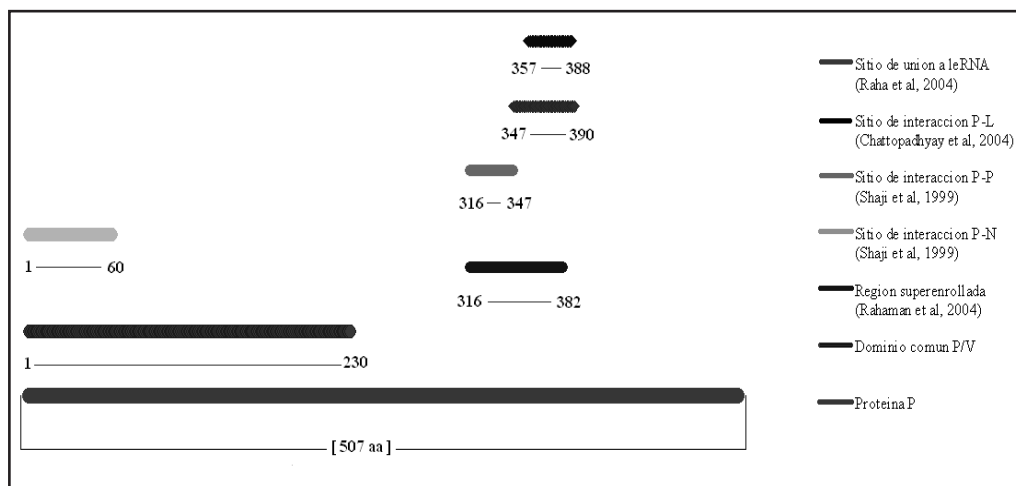


Figura 2: Proteína P de RPV y sus principales sitios de interacción

Nota: Los números indican la posiciones de los residuos de la proteína P de RPV que limitan cada dominio

Sin embargo, los cambios en las posiciones 22, 42, 46, 52, y 59, no respetan esta regla, y ello parece ser un atributo de los aislamientos provenientes del este africano (más el egipcio). Hemos observado (mediante *MEMSAT 1.5*) que la carga eléctrica del dominio PNT tiene un patrón de distribución en forma de «clusters» esparcidos en todo el dominio, y que esa

distribución es similar en todos los aislamientos. Cuando analizamos las características químicas de los cambios de aminoácidos del dominio N-Terminal excluyendo la región de interacción con N (PNT-N), observamos que el 40.26% de las posiciones son variables, pero solo un 10.8% de los reemplazos son por un aminoácido de distinta característica físico-química, lo cual

Tabla 4: Análisis de las proteínas P, V y C (Conservación entre diferentes aislamientos).

|   | Región   | Posición | Número de aminoácidos | Posiciones conservadas | % de conservación de aminoácidos |
|---|----------|----------|-----------------------|------------------------|----------------------------------|
| P | Completa | 1-507    | 507                   | 378                    | 74.56                            |
|   | PNT      | 1-291    | 291                   | 192                    | 65.98                            |
|   | PCT      | 292-507  | 216                   | 182                    | 84.26                            |
|   | DCPV     | 1-229    | 229                   | 151                    | 65.94                            |
|   | DP       | 230-507  | 278                   | 216                    | 77.70                            |
|   | P-N      | 1-60     | 60                    | 48                     | 80.00                            |
|   | DNT-N    | 61-291   | 231                   | 138                    | 59.74                            |
|   | SupEnr   | 316-382  | 67                    | 59                     | 88.06                            |
|   | P-P      | 316-347  | 32                    | 27                     | 84.37                            |
|   | P-L      | 347-390  | 44                    | 40                     | 90.91                            |
|   | P-leRNA  | 357-388  | 32                    | 29                     | 90.62                            |
| V | Completa | 1-299    | 299                   | 214                    | 71.57                            |
|   | VCT      | 230-299  | 70                    | 62                     | 88.57                            |
| C | Completa | 1-177    | 177                   | 130                    | 73.45                            |

Nota: P: Proteína P. PNT: Dominio N-terminal de P. PCT: Dominio C-terminal de P. DCPV: Dominio común entre P y V. DP: Dominio exclusivo de P. P-N: Sitio de unión con proteína N. DNT-N: Dominio N-terminal de P excluyendo sitio de unión a N. SupEnr: Súper enrollamiento. P-P: Sitio de multimerización. P-L: Sitio de unión a proteína L. P-leRNA: Sitio de unión a región leader del genoma viral. V: Proteína V. VCT: Dominio C-terminal de proteína V. C: Proteína C.

indica que el 73.2% de los cambios en el dominio PNT-N se llevan a cabo conservando la naturaleza físico-química del residuo reemplazado. Se ha sugerido que una función de la proteína P efectuada mediante el dominio PNT, es la de trans-activador transcripcional <sup>20</sup>. La existencia de «clusters» cargados negativamente indica la presencia de un dominio similar a los existentes en los trans-activadores celulares <sup>65</sup>. Estos «clusters» se encuentran dispersos en todo el dominio, pero presentan similar localización en todos los aislamientos evaluados, a pesar de la variabilidad de aminoácidos hallada en dicha región.

**Análisis de la proteína V:** El análisis de la proteína V fue realizado de modo similar que para P. La proteína V es traducida a partir de un

mRNA que ha sido modificado al transcribirse, tal modificación, consiste en la adición de una guanina a un sitio aceptor (sitio de edición), generando una proteína que comparte el extremo N-terminal con la original, pero cuyo C-terminal es completamente diferente como consecuencia de haber derivado una posición el marco de lectura <sup>71</sup>. Este dominio C-terminal de V tiene la particularidad de ser rico en residuos de cisteína. El sitio de edición de mRNA requerido para producir proteína V se encuentra en la posición 2489-2503 del genoma (5'TTAAAAAGGGCACAG3') <sup>12</sup>, y está completamente conservado en los aislamientos evaluados. Al igual que para la proteína P, el dominio C-terminal de V está altamente conservado (88.51%). La región rica en cisteína

Tabla 5: Características eléctricas de las proteínas codificadas en el gen P de los aislamientos secuenciados.

| Aislamientos    | Carga neta |            |            | Punto isoeléctrico |            |            |
|-----------------|------------|------------|------------|--------------------|------------|------------|
|                 | Proteína P | Proteína C | Proteína V | Proteína P         | Proteína C | Proteína V |
| RBT             | -22        | 13         | -26        | 4.89               | 11.31      | 4.49       |
| Egipto          | -23        | 13         | -26        | 4.84               | 11.43      | 4.47       |
| Nigeria Buffalo | -23        | 13         | -26        | 4.84               | 11.43      | 4.47       |
| Nigeria Vero    | -23        | 13         | -26        | 4.84               | 11.43      | 4.47       |
| Sokoto          | -24        | 14         | -27        | 4.81               | 12.05      | 4.46       |
| Nigeria Sokoto  | -24        | 14         | -27        | 4.81               | 12.05      | 4.46       |
| Pak Chong       | -26        | 13         | -27        | 4.75               | 11.43      | 4.46       |
| Kuwait          | -26        | 14         | -28        | 4.72               | 11.71      | 4.41       |
| Kuwait 21       | -26        | 14         | -27        | 4.72               | 11.71      | 4.43       |
| Pakistán        | -26        | 13         | -27        | 4.66               | 11.43      | 4.43       |
| Saudi           | -26        | 14         | -28        | 4.72               | 11.59      | 4.41       |
| Yemen           | -26        | 14         | -28        | 4.72               | 11.59      | 4.41       |
| Pendik          | -28        | 13         | -29        | 4.65               | 11.47      | 4.36       |
| India           | -28        | 12         | -28        | 4.70               | 11.04      | 4.42       |
| R Bok           | -28        | 12         | -28        | 4.70               | 11.04      | 4.42       |
| Kabete O        | -29        | 12         | -29        | 4.67               | 11.04      | 4.39       |

Nota: Datos obtenidos mediante el programa *PSORT*

(posición 251 a 279) es invariable en todos los aislamientos con la excepción de la cepa lapinizada (AB021977), que presenta diferencias en las posiciones 262 (A!T) y 268 (T!I). Se ha propuesto que esta región tiene una función de unión a zinc (*zinc-motif finger*)<sup>31</sup>, que explicaría la llamativa conservación de todos los aminoácidos de dicho motivo.

**Análisis de Proteína C:** La proteína C es el producto más pequeño del gen P, obtenido a partir del mismo mensajero que la proteína P, pero iniciando la traducción en un AUG alternativo, que deriva una posición el marco de lectura<sup>49</sup>. Esta estrategia le permite al virus obtener una proteína con características significativamente diferentes, no solo en la

composición de aminoácidos, sino también en sus propiedades (carga eléctrica). Notablemente, la proteína C, tiene una carga neta de entre 12 y 14 en los aislados secuenciados, mientras que P y V son proteínas negativas. (Tabla 5). El análisis cuantitativo de aminoácidos de las proteínas P, V, y C, muestra que las diferencias en carga neta y punto isoeléctrico, es debido a la baja proporción de aminoácidos de carga negativa (ácido aspártico y ácido glutámico) presente en la proteína C en relación a P y V, junto a mayores niveles de aminoácidos básicos, especialmente arginina (tabla 6).

**Sitios potenciales para la fosforilación y glicosilación:** La fosforilación de la proteína P es importante para la modulación de la

Tabla 6: Composición de aminoácidos Proteínas P, V, y C.

| Aminoácido        | Proteína  |           |           | Promedio  |
|-------------------|-----------|-----------|-----------|-----------|
|                   | P         | C         | V         |           |
| A (alanina)       | 7.31±0.26 | 7.41±1.14 | 7.08±0.31 | 6.71±1.14 |
| C (cisteína)      | 0.66±0.14 | 1.12±0.35 | 3.44±0.25 | 1.64±1.01 |
| D (ác. aspártico) | 8.26±0.20 | 2.11±0.35 | 8.80±0.25 | 6.05±2.29 |
| E (ác. glutámico) | 7.95±0.28 | 4.09±0.20 | 10.6±0.36 | 5.94±2.41 |
| F (fenilalanina)  | 1.4±0.12  | 1.14±0.15 | 1.15±0.15 | 2.44±1.36 |
| G (glicina)       | 8.43±0.49 | 3.55±0.46 | 8.04±0.55 | 6.85±1.62 |
| H (histidina)     | 2.11±0.12 | 1.81±0.24 | 2.88±0.15 | 1.89±0.88 |
| I (isoleucina)    | 6.86±0.32 | 5.01±0.35 | 5.08±0.63 | 6.94±1.34 |
| K (lisina)        | 7.41±0.19 | 5.49±0.68 | 4.98±0.22 | 5.65±1.03 |
| L (leucina)       | 7.16±0.35 | 12.5±0.68 | 4.98±0.38 | 9.83±2.49 |
| M (metionina)     | 1.91±0.10 | 6.16±0.41 | 1.62±0.16 | 2.65±1.54 |
| N (asparagina)    | 3.16±0.38 | 2.64±0.57 | 2.66±0.20 | 3.55±0.7  |
| P (prolina)       | 5.60±0.31 | 8.36±0.55 | 6.55±0.42 | 5.73±1.3  |
| Q (glutamina)     | 4.50±0.18 | 5.75±0.62 | 3.28±0.08 | 3.7±1.11  |
| R (arginina)      | 3.73±0.14 | 8.08±0.69 | 5.34±0.30 | 5.8±1.48  |
| S (serina)        | 12.1±0.56 | 10.2±0.53 | 8.69±0.60 | 8.83±1.88 |
| T (treonina)      | 4.64±0.24 | 6.72±0.88 | 4.41±0.34 | 6.11±1.26 |
| V (valina)        | 4.84±0.47 | 4.13±0.43 | 6.90±0.32 | 5.75±1.06 |
| W (triptofano)    | 0.76±0.08 | 3.03±0.30 | 1.94±0.12 | 1.31±0.81 |
| Y (tirosina)      | 1.1±0.12  | 0.63±0.13 | 1.64±0.65 | 2.72±1.45 |

Nota: Datos obtenidos mediante el programa *PSORT*

transcripción y la replicación viral y es llevada a cabo por medio de Kinasas Celulares<sup>23; 45; 52</sup>. La predicción de sitios de fosforilación mediante Casein Kinasa II (CKII), realizada en este trabajo (*PROSITE*), confirmó resultados previamente publicados para RBok<sup>45</sup>. También hemos observado que Ser49, en un contexto adecuado para ser fosforilado, está presente en todos los aislamientos, con la particularidad de que tal posición esta cambiada a Thr48 en Sokoto y Nigeria Sokoto (Tabla 7). El sitio de fosforilación Ser88, previamente señalado como fundamental en la biología de RPV<sup>45</sup>, solo ha sido hallado en RBok y los relacionados

Kabete O e India. El residuo Ser361, presenta un contexto óptimo para la fosforilación mediante CKII, ha sido hallado inserto en el motivo invariante «KQNISISTIEGHLSS», y no ha sido reportado previamente. Señalamos aquí cuatro sitios de fosforilación mediante Protein Kinasa II (PKII) y dos sitios de fosforilación con Protein Kinasa dependiente de cAMP/cGMP, cuya conservación en todos los aislamientos sugieren una importancia funcional pero que no han sido descritos previamente (Tabla 7). En adición, hemos hallado dos sitios de glicosilación invariantes en todas las secuencias analizadas (Tabla 7),

Tabla 7: Sitios de fosforilación y glicosilación conservados en la proteína P de todos los aislamientos

| Protein kinase II |         | Casein Kinase II |              | Protein Kinasa cAMP/cGMP dependiente |         | Sitio de Glicosilación |
|-------------------|---------|------------------|--------------|--------------------------------------|---------|------------------------|
| Posición          | Residuo | Posición         | Residuo      | Posición                             | Residuo | Posición               |
| 224               | S       | 49<br>(48 Sok)   | S<br>(T Sok) | 229                                  | K       | 266                    |
| 348               | S       | 151              | S            | <b>444</b>                           | K       | 357                    |
| <b>402</b>        | S       | 180              | S            |                                      |         |                        |
| 415               | S       | <b>361</b>       | S            |                                      |         |                        |

Nota: Aquellos sitios en negrita son los que tienen un contexto completamente conservado. Los datos para la confección de esta tabla fueron obtenidos mediante *PROSITE*

sugiriendo estar involucrados en alguna función de la proteína, tal como interacción con otras moléculas.

Respecto a la proteína V, hemos observado que el dominio C-terminal presenta dos sitios de fosforilación por PKII (Ser224 y Ser262) invariantes en todas las secuencias con excepción de la cepa lapinizada AB021977 (tomada del Genbank). En la proteína C, hemos encontrado cuatro sitios adecuados para la fosforilación mediante PKII (Ser2, Ser21, Ser49 y Ser72), y un quinto residuo (Ser134) cuyo contexto es ideal para ser fosforilado por CKII (tabla 8), ninguno de los cuales han sido reportado previamente pero completamente conservados en todas las secuencias.

### Comparación de aislamientos de diferente patogenicidad

Se ha sugerido recientemente que en el gen P reside el principal determinante de patogenicidad específica del RPV<sup>85</sup>. Para verificar la presencia de sitios involucrados en patogenicidad, nosotros comparamos la secuencia de aminoácidos de los productos de este gen, en dos aislamientos de diferente patogenicidad, la cepa R Bok (cepa vacunal) y

Kabete O (variante altamente patogénica del cual se originó el R Bok). Llamativamente, solo hemos encontrado dos residuos de diferencia entre la secuencia de la proteína P de estos dos virus, en la posición 84 (G en Kabete O / R en R Bok) y en la posición 370 (V en Kabete O / F en R Bok). No hemos hallado ninguna diferencia al comparar la secuencia de proteína C y V de los virus mencionados.

Tabla 8: Sitios de fosforilación de la proteína C conservados en todos los aislamientos

| Protein kinase II |         | Casein Kinase II |           |
|-------------------|---------|------------------|-----------|
| Posición          | Residuo | Posición         | Residuo   |
| 2                 | S       | 134              | S (T Sok) |
| 21                | S       |                  |           |
| 49                | S       |                  |           |
| 72                | S       |                  |           |

Nota: Datos obtenidos mediante el programa *PROSITE*

## DISCUSIÓN

El gen P es el único gen policistrónico de la familia Paramyxoviridae y es indudablemente el más complejo del genoma de RPV<sup>49</sup>. A pesar de la restricción biológica que representa codificar tres diferentes proteínas en marcos de lectura alternativos en un único gen, hemos observado una importante variabilidad (cerca al 23%) en la secuencia de nucleótidos de este gen (Tabla 3). Mediante el uso de anticuerpos monoclonales se demostró que la proteína P es la más variable entre morbillivirus<sup>11</sup>. Sin embargo, P es el único gen que hibridiza entre todos los miembros del género, debido a secuencias específicas altamente conservadas esparcidas a lo largo del gen<sup>71</sup>. En este trabajo hemos hallado que la proporción de transiciones sobre transversiones (ts/tv) es superior a 8 (Tabla 3), lo que indica que los cambios de nucleótidos observados son mayoritariamente transiciones (aproximadamente 89% de los cambios), por otro lado, el análisis de la variabilidad por codones demuestra una supremacía de cambios de tipo no-sinónimo sobre los del tipo sinónimo (Tabla 3). Estos parámetros de variabilidad sugieren que el gen P es mucho más conservado desde el punto de vista de la naturaleza química de las bases de nucleótidos substituidos, que desde el punto de vista de la secuencia de aminoácidos de los productos codificados. En otras palabras, existe mayor restricción desde el punto de vista del RNA que de la proteína. Debe tenerse en cuenta que el genoma de los virus Paramyxoviridae se «enrolla» alrededor de la proteína N durante la encapsidación, fenómeno que involucra seis nucleótidos por vuelta<sup>17</sup>. Entonces, es razonable que las interacciones de índole físico-químicas entre el RNA y N expliquen el alto índice de **ts/tv**, mientras que a nivel de proteína las restricciones se reducen a dominios específicos y permitan

un relativamente bajo valor de **dn/ds**, como se ha observado en este trabajo (Tabla 3).

Mediante el análisis de identidades por pares, los aislamientos fueron ubicados en tres grupos distintos con mínima diferencia intra-grupo, coincidiendo con el postulado de los tres linajes de RPV. La figura 3 muestra el árbol filogenético (*Neighbor-Joining*), donde claramente se observan tales grupos. Uno formado por Kabete O, R Bok e India, otro formado por los aislamientos provenientes del sudeste asiático y medio oriente (Pak Chong, Pendik, Pakistán, Yemen, Saudi, y ambos Kuwait), y un tercer grupo formado por los aislamientos provenientes de África occidental (Nigeria Vero, Nigeria Buffalo, Nigeria Sokoto y Sokoto) más Egipto y RBT (originario de Tanzania). En este análisis hemos observado una cercanía muy estrecha entre algunos pares de aislamientos: Nigeria Vero con Nigeria Buffalo, Sokoto con Nigeria Sokoto, y Kuwait con Kuwait-21. Ello es debido a que los mismos representan al mismo aislamiento con diferente historia de pasajes en cultivos celulares. Lo mismo sucede con R Bok (cepa vacunal) y el virus parental del cual deriva, Kabete O (altamente patógeno y originariamente aislado en 1911 en Kenya). Notablemente, no hay diferencia entre el aislamiento India y R Bok. Es evidente que se trata del mismo virus, la distancia genética entre ellos es menor aún que la que existe entre R Bok y Kabete O. Es sabido que muchas enfermedades del ganado presentan un cuadro clínico similar al producido por la infección con cepas de RPV de moderada patogenicidad<sup>63</sup>. El diagnóstico de laboratorio (tal como el aislamiento viral) es el que confirma si el animal sospechoso estaba realmente infectado con RPV. Pero si el animal afectado proviene de un rodeo recientemente vacunado contra RP, es factible que al intentar aislar al agente etiológico, lo que se obtenga

sea el virus usado como vacuna (R Bok), identificando erróneamente a RPV como el agente causal <sup>9</sup>. Quizás esa sea la razón de la semejanza genética hallada entre los aislamientos R Bok e India, desafortunadamente no contamos con la información adicional del aislamiento de RPV India que nos permita confirmar esta presunción. Otro punto interesante que surge del análisis, es la cercanía filogenética observada entre los aislamientos de África occidental, Egipto, y RBT, los cuales provienen de regiones geográficas muy alejadas, incluso RBT (Tanzania), fue aislado en 1966, más de 25 años de diferencia con los aislamientos nigerianos, lo cual representa un hallazgo muy interesante desde el punto de vista epidemiológico.

La proteína P ha sido descrita como el único producto del gen P esencial para la síntesis viral <sup>49</sup>. Desde la perspectiva vista estructural esta proteína es la mejor caracterizada de las que componen el complejo polimerasa viral, con dos dominios estructural y funcionalmente distintos, el dominio N-terminal (PTN) y el dominio C-terminal (PCT) <sup>43</sup>. El dominio PCT es el más conservado y tiene todas las regiones requeridas para la transcripción viral, mientras que PNT, el cual es pobremente conservado provee algunas funciones adicionales requeridas para la replicación <sup>22</sup>. Aún cuando la proteína L contiene (aparentemente) todas las funciones catalíticas de polimerasa viral <sup>49</sup>, P es necesaria para la unión de L con el complejo N-RNA <sup>16</sup>. Esta interacción entre P y L requiere de un motivo de P que ha sido localizado en el dominio PCT. Se ha sugerido que para formar parte de ese complejo la proteína P debe estar en forma de oligómero <sup>39; 80</sup>. La oligomerización de P depende de un área caracterizada por el súper-enrollamiento localizado en PCT. En este dominio se ha localizado además el sitio de unión de P a *le*RNA <sup>66</sup>. Es indudable entonces,

que la elevada conservación observada en este trabajo, 84.26% para el dominio PCT (que supera el 90% en los sitios de unión a L y a *le*RNA), está relacionado con la complejidad funcional de esa región, que soporta la mayoría de los sitios de interacción de la proteína P incluyendo el sitio de oligomerización. Tales sitios se superponen entre sí, en menor o mayor medida (Fig. 2), siendo esta característica un factor muy importante en la restricción a la variabilidad del dominio. En contraste, el dominio PNT presenta menor porcentaje de posiciones conservadas (65.98% en nuestro trabajo), pero con una región altamente invariante (80% de conservación) en el extremo N-terminal del dominio, entre los residuos 1 a 60, región que ha sido identificada previamente como sitio de unión a N <sup>73</sup>.

Se ha demostrado *in vitro* que el extremo N-terminal de la proteína P del Virus de Estomatitis Vesicular (VSV), es funcionalmente necesaria para la transcripción de mRNA, aún cuando esta región de P es altamente variable entre los diferentes aislamientos de VSV, lo cual resulta llamativo dada la importancia de su función. Se ha demostrado que la secuencia referida puede ser funcionalmente sustituida por una de la proteína tubulina, cuya única semejanza con la secuencia de P es la naturaleza ácida de su extremo C-terminal <sup>20</sup>. Esta conservación de la función en ausencia de la conservación de la secuencia (a excepción de mantener la carga negativa), sugiere que esta región es reminiscencia de los dominios ácidos de activación (AAD) de los factores de transcripción celular presentes en eucariotas <sup>20</sup>, que parecen requerir solo de residuos cargados negativamente para llevar a cabo la función de activación <sup>65</sup>. Hay evidencia que demuestra que no se requiere una estructura muy elaborada para que las cargas negativas en la región de activación de los factores de transcripción



(RAFT) sean expuestos funcionalmente <sup>65</sup>. En este trabajo hemos observado que el área N-PNT presenta la mayor variabilidad de la proteína P (40.26%), sin embargo la variabilidad baja drásticamente cuando se comparan las cargas eléctricas de los aminoácidos en cada posición. Notablemente, la distribución de la carga eléctrica negativa forma «clusters» a lo largo del dominio que están conservadas en todas las secuencias. Este hallazgo apoya la evidencia de que parte del dominio PNT actúa como RAFT. También se ha sugerido que la fosforilación puede jugar un rol importante al incrementar la potencia de RAFT (quizás simplemente por contribuir a la carga negativa), y de este modo constituir un medio de controlar a los activadores en respuesta a señales ambientales <sup>65</sup>. Coincidentemente, existe la presunción de que la fosforilación de la proteína P, mediante Protein Kinasas, es un evento necesario para la activación del complejo polimerasa <sup>74</sup>. La importancia de la fosforilación de P como mecanismo de regulación ha quedado demostrada en diferentes trabajos <sup>23; 66; 74; 80</sup>. La proteína P puede unirse en forma específica a la región *le*RNA, y esa unión que modula el fenómeno de encapsidación, esta regulada por la fosforilación <sup>66</sup>. Debe tenerse en cuenta, que la fosforilación de P no es indispensable para la transcripción y/o replicación, solo cumple un rol en la modulación del proceso, que es de suma importancia al principio de la infección, cuando la concentración de proteína P es baja <sup>80</sup>. Es interesante notar que la secuencia de aminoácidos entre las proteínas P de toda la familia Paramyxoviridae no está particularmente conservada, sin embargo la fosforilación de la proteína P esta siempre presente (la fosforilación esta conservada, no lo sitios de fosforilación) <sup>74</sup>. Muchos autores enfatizan que la significancia funcional de la fosforilación de P en la familia Paramyxoviridae

no esta suficientemente clara, y que ello se puede deber en parte, a la falta de información sobre la localización precisa de los sitios de fosforilación de la proteína P, por lo cual este trabajo puede representar un aporte al respecto.

Se ha sugerido que diferentes tipos de Kinasas pueden ser efectivas en la fosforilación de la proteína P de Paramyxoviridae *in vivo* <sup>74</sup>. Ha sido demostrado para MV que CKII es la responsable la fosforilación de P, identificándose tres potenciales sitios de fosforilación en la cepa Edmosnton («*vacunal*»). Se ha establecido que serina es el residuo principalmente fosforilado en la proteína P de la mayoría de los virus de la familia Paramyxoviridae <sup>68</sup>, incluyendo los emergentes Nipah Virus (NiV) y Hendra Virus (HeV) <sup>74</sup>. Mediante estudios *in vitro* se ha demostrado que la isoforma ζ de la Protein Kinasa C (PKCζ) es particularmente activa en la fosforilación de la proteína P de CDV y de MV <sup>52</sup>. La efectividad de la fosforilación demostrada *in vitro*, sumado al hecho de que las líneas celulares que soportan la replicación de CDV poseen la isoforma ζ de la enzima PKC, son una fuerte evidencia a favor de la acción de aquella como responsable de la fosforilación en CDV <sup>52</sup>. De este modo, más de un tipo de Kinasa puede estar involucrada en la fosforilación de la proteína P. En RPV se ha observado que CKII es la Kinasa preferentemente involucrada en la fosforilación de la proteína P, y que serina es mejor aceptor de fosfatos que treonina <sup>45</sup>. Se ha sugerido que mientras la fosforilación se lleva a cabo en las posiciones Ser49, Ser88 y Ser151, solo los residuos Ser49 y Ser88 son críticos en la modulación de la transcripción y replicación viral <sup>45</sup>. En otro trabajo (del mismo grupo) se afirma que los mismos residuos están involucrados en la regulación de la unión específica de P a la región *le*RNA y que dicha interacción modularía la encapsidación del genoma <sup>66</sup>.

Interesantemente, se ha observado en este trabajo que los residuos mencionados como cruciales en tales regulaciones no están completamente conservados, y dado que aquellos trabajos fueron realizados usando proteína P de R Bok, es que sugerimos aquí, que diferentes sitios pueden estar involucrados en la regulación de la función de P en aquellas cepas diferentes al linaje R Bok y Kabete O. Adicionalmente, hemos encontrado cuatro residuos serina (224, 348, 402, y 415) y dos residuos lisina (229 y 444) en un contexto adecuado para la fosforilación por medio de PKII y PK dependiente de cAMP/cGMP respectivamente. Los mismos están presentes en todos los aislamientos, inclusive hemos hallado dos posiciones serina (266 y 357) cuyos contextos, completamente conservados, son adecuados para la glicosilación de la proteína. Hasta el momento, no hay en la literatura ninguna referencia a los sitios mencionados, sin embargo la perfecta conservación de los mismos sugiere algún tipo de participación en la biología viral.

La proteína V es una proteína no estructural, producto de la expresión de un mensajero, cuyo marco de lectura está derivado una posición respecto del que codifica para la proteína P, por la inserción programada de un residuo guanidina extra. Esta inserción se lleva a cabo durante la síntesis del RNA mensajero en un proceso llamado edición co-transcripcional<sup>22</sup>. La inserción se produce en cualquier posición entre los nucleótidos 2494 y 2497 (referido al genoma viral), siendo fundamental la presencia de una secuencia específica (5'-TTAAAAAGGGCACAG-3')<sup>71</sup>. En todos los aislamientos hemos observado que la secuencia de nucleótidos del sitio de edición y el área alrededor del mismo se encuentra conservada en un 100 %. Del mismo modo hemos encontrado que la región rica en cisteína se

encuentra entre los residuos 251 y 279. Esta región esta completamente conservada en todos los aislamientos, con la excepción de la cepa lapinizada, la cual presenta cambios en las posiciones 262 (A'→T) y 268 (T'→I). La secuencia de la proteína P de la cepa lapinizada (tomada del Genbank), contiene mutaciones exclusivas que sugieren que la secuencia no es representativa del RPV y que pueden ser artefactos generados por el tipo de manipulación a que fuera sometida esa secuencia<sup>58</sup>. Un elemento importante en la estructura de la proteína V, es la posición de los residuos de cisteína, los cuales se encuentran en las posiciones 251, 255, 267, 269, 272, 276, y 279 en todas las secuencias. Dada la proporción de residuos de cisteína, esta región tendría función de proteína de unión a zinc (*zinc-motif finger*), lo cual indica que mediante este dominio la proteína V puede estar involucrada en interacciones del tipo proteína-proteína o proteína-ácido nucleico<sup>31</sup>. De hecho hay evidencia de la interacción de V con proteínas del hospedador<sup>44: 82</sup>, lo cual ha sido bien documentado en el caso de la interacción de V con proteínas DDB en SV5<sup>51</sup>. Ello explicaría la invariabilidad del mencionado motivo. Al igual que para la proteína P, se ha observado que V puede ser sustrato de Kinasas celulares por lo que la fosforilación de esta proteína podría ser un elemento importante en la regulación de su función. En este trabajo hemos encontrado potenciales sitios de fosforilación (PKII), completamente conservados en las posiciones 224 y 262. La importancia funcional de estos sitios de fosforilación deberán ser demostrados experimentalmente.

La proteína C es codificada por el mismo mensajero que la proteína P, pero su traducción comienza en el segundo AUG, 20 nucleótidos distal del AUG de inicio de la proteína P. De esta forma, el marco de lectura se desplaza una

posición respecto del original, lo cual genera una proteína de características totalmente diferentes de P (Tabla 5 y 6). La particularidad de esta proteína es la carga neta altamente positiva y el punto isoeléctrico definitivamente elevado (Tabla 5), se piensa que la carga eléctrica de la proteína es un elemento importante en relación a sus funciones, las cuales poco a poco están empezando a esclarecerse<sup>4; 24; 71; 72</sup>. Se ha reportado que la proteína C de MV inhibe la respuesta de IFN en células humanas por bloqueo de la vía de transducción de señales<sup>72</sup>. Se ha observado que mutantes de MV cuya capacidad para producir proteína C esta inhibida (MV C-), no pueden replicar en células de origen humano cuyo sistema IFN- $\alpha/\beta$  es funcional (IFN- $\alpha/\beta$  competente). Sin embargo, esa restricción a la replicación no ocurre en células IFN- $\alpha/\beta$  competentes pero de origen no humano<sup>72</sup>. En el mismo trabajo se observó que la replicación de MV C- en células humanas, se lograba mediante el rescate con anticuerpos anti-IFN- $\alpha/\beta$ , lo cual indica que el crecimiento reducido de MV C- en células humanas fue debido al efecto antiviral de IFN de tipo I<sup>72</sup>. Estos resultados sugieren que la replicación de MV C- es inhibida por IFN- $\alpha/\beta$ , y que la pérdida de la habilidad de bloquear la respuesta de IFN- $\alpha/\beta$  en células estimuladas se debe a la falta de la proteína C. En aquel trabajo, los mutantes MV C- expresaban la proteína V, por lo tanto la contribución relativa de las proteínas V y C a la inhibición de la respuesta de IFN *in vivo* no ha sido evaluada. Por otro lado, otro equipo falló en demostrar actividad de interferencia de la proteína C de MV sobre los mecanismos antivirales de la célula<sup>78</sup>. También se ha sugerido que la proteína C tiene responsabilidad en la modulación de la transcripción y replicación de MV<sup>24</sup>. Basado en la evidencia experimental mencionada, se ha sugerido que la proteína C debe considerarse como factor de virulencia,

cumpliendo un rol excluyente en la atenuación del virus. Se ha visto que las cepas de MV que han sido aisladas usando PBMC o células B95a indujeron 10 veces mas IFN- $\alpha/\beta$  que aquellas que fueron atenuadas previamente por pasaje por VERO que es una línea celular carente del gen estructural de IFN  $\beta$ <sup>56</sup>. Es interesante especular que el pasaje por célula puede ejercer como presión de selección para virus que difieren en su capacidad de inducir IFN- $\alpha/\beta$  y que esa diferencia sea la base de la atenuación. En el caso de RPV esto encuentra respaldo en el hecho de que la vacuna (R Bok) ha sido obtenida por pasajes sucesivos del virus salvaje (Kabete O) en células VERO. Entonces la evidencia de este mecanismo de atenuación debería demostrarse, mediante la observación de diferencias, en las secuencias de la proteína C de cepas virulentas y atenuadas. Sin embargo, cuando comparamos las secuencias de C de las cepas RBok (atenuada) con Kabete O (de altísima virulencia), no hemos hallado ninguna diferencia. Coincidente con nuestra observación, recientemente se llevó a cabo un análisis comparativo entre las secuencias de la proteína C de tres cepas vacunales de MV con tres virus salvajes, sin resultados que expliquen la diferencia en la patogenicidad de esas variantes<sup>4</sup>.

Las cepas de campo son altamente variables en virulencia, transmisibilidad y especificidad<sup>10</sup>. La habilidad del virus de cambiar la virulencia, el tropismo o incluso el rango de hospedador (cepas lapinizadas) han sido utilizados para la preparación de vacunas, sin embargo, las bases moleculares que expliquen la atenuación permanecen desconocidas<sup>46; 60</sup>. De hecho la comparación de la secuencia de genoma completo entre la cepa vacunal (RBok) y la cepa virulenta de la cual deriva (Kabete O), solo ha mostrado variabilidad de nucleótidos en 87 posiciones (0.55%), y no ha sido posible

reconocer la diferencia en los aminoácidos que conducen a la atenuación<sup>7</sup>. Se ha sugerido que en los productos del gen P reside el determinante de patogenicidad específica<sup>82</sup>, sin embargo aún quedan muchos puntos por esclarecer respecto de la función de las proteínas C y V. La participación excluyente de C y V en la patogenicidad demostrada en SeV<sup>44; 48</sup> y en MV<sup>82</sup>, junto con la co-localización de C con N, P y L en las células infectadas<sup>5</sup>, sugieren la posibilidad de que C juegue un papel en la atenuación del virus en asociación con otras proteínas virales. Es posible que la virulencia (y atenuación) no dependa de una única proteína, sino de un complejo de virulencia viral, del cual la proteína C y / o V son solo componentes. Podemos especular con ello y plantear que quizás allí resida la dificultad en determinar la base genética de la atenuación. El análisis de la secuencia completa de varios aislamientos ofrecerá la oportunidad de conocer más acerca de las diferencias entre variantes virulentas y atenuadas.

Este trabajo representa el primer análisis genómico comparativo de variantes de RPV de patogenicidad, origen geográfico e historia diferentes. Los elementos conservados entre estos aislamientos sugieren una importancia funcional, que en muchos casos deberán ser probadas mediante correspondientes experimentos. En otros casos la conservación por sí misma es un elemento importante y coincidente con otros estudios. La descripción de elementos potencialmente importantes, tales como sitios de fosforilación que no han sido descritos previamente, y que están conservados en todas las secuencias, puede servir para futuros análisis. RPV ha sido uno de los más importantes agentes infecciosos para los animales domésticos, produciendo un impacto enorme en la economía Europea del

siglo XVIII<sup>14</sup>, y diezmando poblaciones de animales tanto domésticos como salvajes. En la actualidad la enfermedad está localizada en focos en Asia y prácticamente ha sido erradicada de África, esto ha sido debido a una intensa campaña de erradicación de la enfermedad llevada a cabo por la Organización para la Agricultura y Alimentación (FAO según sus siglas en inglés)<sup>3</sup>. Se ha previsto la erradicación completa de RPV para el 2010<sup>63</sup>, sin embargo la aparición de algunos brotes en áreas que se pensaban libres de la enfermedad ha sembrado algunas dudas al respecto. En la actualidad, con la globalización del comercio y la rapidez de intercambio entre regiones geográficas que antes resultaban remotas e inaccesibles entre sí, la importación de agentes infecciosos exóticos se supone como realidad inapelable. Así mismo la amenaza del bioterrorismo y liberación intencionada de microorganismos de alto impacto económico ha reactivado actividades de investigación en enfermedades que, como RP, parecían controladas y en vías de erradicación. En países desarrollados se ha recuperado el interés en profundizar el conocimiento del virus en todos sus aspectos, lo que permitirá un mejor uso de los recursos disponibles para la erradicación total. La secuenciación y análisis del genoma de RPV, como el realizado en este trabajo, es una herramienta sumamente útil y eficiente, tanto desde el punto de vista epidemiológico como en la evaluación de la biología del virus.

## AGRADECIMIENTOS

Este trabajo fue realizado con fondos del departamento de Agricultura de los Estados Unidos de América (USDA), bajo la dirección del Dr. Dan Rock.

## BIBLIOGRAFÍA

1. AFONSO, C., TULMAN, E., LU, Z., OMA, E., KUTISH, G., ROCK, D. 1999. The Genome of *Melanoplus Sanguinipes* Entomopoxvirus. *J. Virol.* 73: 533-552.
2. ANDERSON, E. C. 1995. Morbillivirus Infections in Wildlife (in relation to their population biology and disease control in domestic animals). *Vet. Microbiol.* 44: 319-332.
3. ANDERSON, J., BARRET, T., SCOTT, G. 1996. *FAO Animal Health Manual – Manual on Diagnosis of Rinderpest*, 2<sup>nd</sup> Ed. Rome. Food and Agriculture Organization of the United Nations.
4. BANKAMP, B., WILSON, J., BELLINI, W., ROTA, P. 2005. Identification of Naturally Occurring Amino Acid variation that Affect the Ability of the Measles Virus C Protein to Regulate Genome Replication and Transcription. *Virology.* 336: 120-129.
5. BARON, M., BARRET, T. 2000. Rinderpest Viruses Lacking the C and V Proteins Show Specific Defects in Growth and Transcription of Viral RNAs. *J. Virology.* 74: 2603-2611.
6. BARON, M., BARRET, T., 1997. Rescue of Rinderpest Virus from Cloned cDNA. *Journal of Virology.* 71: 1265-1271.
7. BARON, M., KAMATA, Y., BARRAS, V., GOATLEY, L., BARRET, T. 1996. The Genome Sequence of the Virulent Kabete «O» Strain of Rinderpest Virus: Comparison with the Derived Vaccine. *J. Gen. Virol.* 77: 3041-3046.
8. BARRET, T. 1999. Rinderpest and Distemper Viruses (Paramyxoviridae). *Encyclopedia of Virology.* 2nd E. Academia press. Pp. 1559-1569.
9. BARRET, T., AMAREL-DOEL, C., KITCHING, R., GUSEV, A. 1993. Use of the Polymerase Chain Reaction in Differentiating Rinderpest Field Virus and Vaccine Virus in the Same Animals. *Rev. Sci. Tech. Off. Int. Epiz.* 12: 865-872.
10. BARRET, T., BARON, M. 2000. Rinderpest Viruses Lacking the C and the V Proteins Show Specific Defects in Growth and Transcription of Viral RNAs. *J. Virol.* 74: 2603-2611.
11. BARRET, T., ROSSITER, P. 1999. Rinderpest: The Disease and its Impact on Humans and Animals. *Advances in Virus Research.* 53: 89-110.
12. BARRET, T., SUBBARAO, S., BELSHAM, G., MAHY, B. 1991. *The Molecular Biology of the Morbillivirus. The Paramyxoviruses.* Kingsbury David. Plenum Press, New York Chapter 3. pp 83-102.
13. BASAK, S., RAHA, T., CHATTOPADHYAY, D., MAJUMDER, A., SHAILA, M. 2003. Leader RNA Binding Ability of Chandipura Virus P Protein is Regulated by its Phosphorylation Status: a Possible Role in Genome Transcription-Replication Switch. *Virology.* 307: 372-385.
14. BLANCOU, J. 2003. Rinderpest. History of the Surveillance and Control of Transmissible Animal Diseases. *Office Interantional des Epizooties.* pp. 171-190.
15. BRENDEL, V., BUCHER, P., NOURBAKHSB, B., BLAISDELL, B., KARLIN, S. 1992. Methods and Algorithms for Statistical Analysis of Protein Sequences. *PNAS.* 89: 2002-2006.
16. BROWN, D., COLLINS, F., DUPREX, W., BARON, M., BARRET, T., RIMA, B. 2005. Rescue of Mini-Genomic Constructs and Combinations of Morbillivirus N, P, and L Proteins. *J. Gen. Virology.* 86: 1077-1081.
17. CALAIN, P., ROUX, L., 1993. The Rule of Six, a Basic Feature for Efficient Replication of Sendai Virus Defective Interfering RNA. *Journal of Virology.* 67: 4822-4830.
18. CHATTOPADHYAY, A., SHAILA. M. 2004. Rinderpest Virus RNA Polymerase Subunits: Mapping of Mutual Interacting Domains on the Large Protein L and Phosphoprotein P. *Virus Genes.* 28: 169-178.
19. CURRAN et al 1991, *EMBO J.* 10: 3079-3085
20. CURRAN, J. 1996. Reexamination of the Sendai Virus p Protein Domains Required for RNA Synthesis: A Possible Supplemental Role for the P Protein. *Virology.* 221: 130-140.

21. CURRAN, J., KOLAKOFSKY, D. 1999. Replication of Paramyxovirus. *Adv. Virus Research.* 54: 403-422.
22. CURRAN, J., MARQ, J-B., KOLAKOFSKY, D. 1995. An N-Terminal Domain of the Sendai Virus P Protein acts as a Chaperone for the NP Protein During the Nascent Chain Assembly Step of Genome Replication. *J. Virology.* 69: 849-855.
23. DAS, T., SCHUSTER, A., SCHNEIDER-SCHAULIES, S., BANERJEE, A. 1995. Involvement of Cellular Casein Kinase II in the Phosphorylation of Measles Virus P Protein: Identification of Phosphorylation Sites. *Virology.* 211: 218-226.
24. DEVAUX, P., CATTANEO, R. 2004. Measles Virus Phosphoprotein Gene Products: Conformational Flexibility of the P/V Protein Amino-Terminal Domain and C Protein Infectivity Factor Function. *J. Virology.* 78: 11632-11640.
25. DEVEREUX, J., HAEBERLI, P., SMITHIES, O. 1984. A Comprehensive set of Sequence Analysis Programs for the VAX. *Nucleic. Acids. Res.* 12: 387-395.
26. EWING, B., HILLIER, M., WENDL, C., GREEN, P. 1998. Base-Calling of Automated Sequencer Traces Using PHRED. I. Accuracy Assessment. *Genome Res.* 8: 175-185.
27. FELSESTEIN, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c Distributed by the Author. Department of Genetics, University of Washington, Seattle.
28. GALINSKI, M., WECHSLER, S. The Molecular Biology of the Paramyxovirus Genus. The Paramyxoviruses. 1991. Kingsbury David. Plenum Press, New York Chapter 3. pp 41-82
29. GALTIER, N., GOUY, M., GAUTIER, C. 1996. SeaView and PHYLO\_win, Two Graphic Tools for Sequence Alignment and Molecular Phylogeny. *Comput. Applic. Biosci.* 12: 543-548.
30. GARNIER, J., OSGUTHORPE, D., ROBSON, B. 1978. Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. *J. Mol. Biol.* 120: 97-120.
31. HAAS, L., BARON, M., LIESS, B., BARRET, T. 1995. Editing of Morbillivirus P Gene Transcripts in Infected Animals. *Veterinary Microbiology.* 44: 299-306.
32. HARTY, R., PALESE, P. 1995. Measles Virus Phosphoprotein (P) Requires the NH<sub>2</sub>- and COOH-terminal Domains for Interaction with the Nucleoprotein (N) but only the COOH-terminus for Interaction with itself. *J. Gen. Virol.* 76: 2863-2867.
33. HENIKOFF, S., HENIKOFF, J. 1994. Protein Family Classification Based on Searching a Database of Blocks. *Genomics* 19:97-107.
34. HOFMANN, K., STOFFEL, W. 1993. TMBase, a Database of Membrane Spanning Protein Segments. *Biol. Chem. Hoppe-Seyler.* 347: 166.
35. HORIKAMI, S., CURRAN, J., KOLAKOFSKY, D., and MOYER, S. 1992. Complexes of Sendai Virus NP-P and P-L Proteins are Required for Defective Interfering Particle Genome Replication in vitro. *J. Virology.* 66: 4901-4908.
36. HORIKAMI, S., SMALLWOOD, S., BANKAMP, B., MOYER S. 1994. An Amino Proximal Domain of the L Protein Binds to the P Protein in the Measles Virus RNA Polymerase Complex. *Virology.* 205: 540-545.
37. HUANG, X., MADAM, A. 1999. CAP3: a DNA Sequence Assembly Program. *Genome Res.* 9: 868-877.
38. HUSON, D. 1998. Splittree: a Program for Analyzing and Visualizing Evolutionary Data. *Bioinformatics.* 14: 68-73.
39. JOHANSSON, K., BOURHIS, J., CAMPANACCI, V., CABBILLAU, C., CANARD, B., LONGHI, S. 2003. Crystal Structure of the Measles Virus Phosphoprotein Domain Responsible for the induced Folding of the C-Terminal Domain of the Nucleoprotein. *J. Biol. Chemistry.* 278: 44567-44573.
40. JONASSEN, I. 1997. Efficient Discovery of Conserved Patterns Using a Pattern Graph. *CABIOS* 13: 509-522.

41. JONASSEN, I., COLLINS, J., HIGGINS, D. 1995. Finding Flexible Patterns in Unaligned Protein Sequences. *Protein Science*. 4: 1587-1595.
42. JONES, D., TAYLOR, W., THORNTON, J. 1994. A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology. *Biochemistry*. 33: 3038-3049.
43. KARLIN, D., FERRON, F., CANARD, B., LONGHI, S. 2003. Structural Disorder and Modular Organization in Paramyxovirinae N and P. *J. Gen. Virol.* 84: 3239-3252.
44. KATO, A., KIYOTANI, K., SAKAI, Y., YOSHIDA, T., NAGAI, Y. 1997. The Paramyxovirus, Sendai Virus V Protein Encodes a Luxury Function Required for Viral Pathogenesis. *EMBO*. 16: 578-587.
45. KAUSHIK, R., SHAILA, M. 2004. Cellular Casein Kinase II-Mediated Phosphorylation of Rinderpest Virus P Protein is a Prerequisite for its Role in Replication/Transcription of the genome. *J. Gen. Virol.* 85: 687-691.
46. KOBUNE, F., SAKATA, M., SUGIYAMA, M., SUGIURA, A. 1991. B95a, Marmoset Lymphoblastoid Cell Line, as a Sensitive Host for Rinderpest Virus. *J. Gen. Virol.* 72: 687-692.
47. KROGH, A., BROWN, M., MIAN, I., SJOLANDER, K., HAUSSLER, D. 1994. Hidden Markov Models in Computational Biology: Applications to Protein Modeling. *J. Mol. Biol.* 235: 1501-1531.
48. KUROTAMI, A., KIYOTANI, K., KATO, A., SHIODA, T., SAKAI, Y., MIZUMOTO, K., YOSHIDA, T., NAGAI, Y. 1998. Sendai Virus C Proteins are Categorically non-Essential Gene Products but silencing their Expression Severely Impairs Viral Replication and Pathogenesis. *Genes to Cell*. 3: 111-124.
49. LAMB, R., KOLAKOFOSKY, D. Paramyxoviridae: The Viruses and their Replication. *Fields Virology*. 4<sup>th</sup> E. Vol. I. 2001. Lippincott, Williams & Wilkins. Chapter 41. 1305-1335.
50. LIN, G., LAMB, R. 2000. The Paramyxovirus Simian Virus 5 V Protein Slows Progression of the Cell Cycle. *J. Virology*. 74: 9152-9166.
51. LIN, G., PATERSON, R., RICHARDSON, C., LAMB, R. 1998. The V Protein of the Paramyxovirus SV5 Interact with Damage-Specific DNA Binding Protein. *Virology*. 249: 189-200.
52. LIU, Z., HUNTLEY, C., DE, B., DAS, T., BANERJEE, A., OGLESBEE, M., 1997. Phosphorylation of Canine Distemper Virus P Protein by Protein Kinase C- $\alpha$  and Casein Kinase II. *Virology*. 232: 198-206.
53. LOLE, K., BOLLINGER, R., PARANJAPPE, R., GADKARI, D., KULKARNI, S., NOVAK, N., INGERSOLL, R., SHEPPARD, H., RAY, S. 1999. Full-Length Human Immunodeficiency Virus Type 1 Genomes from Subtype C-infected Seroconverters in India, with Evidence of Inter-subtype Recombination. *J. Virol.* 73: 152-160.
54. LOWE, M., NAKAMURA, N., WARREN, G. 1998. Golgi Division and Membrane Traffic. *Trends Cell Biol.* 8: 40-44.
55. MORGENSTERN, B. 1999. DIALIGN 2: Improvement of the Segment-to-Segment Approach to Multiple Sequences Alignment. *Bioinformatics* 15: 211-218.
56. MOSCA, J., PITHA, P. 1986. Transcriptional and Posttranscriptional Regulation of Exogenous Human Beta Interferon Gene in Simian Cells Defective in Interferon Synthesis. *Mol. Cel. Biol.* 6: 2279-2283.
57. NAKAI, K., HORTON, P. 1999. PSORT: A Program for Detecting Sorting Signals in Proteins and Predicting Their Subcellular Localizations. *Trends in Biochem. Sci.* 24: 34-35.
58. NAKAMURA, K., IWATSUKI, K., IZUMIYA, Y., NISHIMURA, Y., OHASHI, K., MIKAMI, T., KAI, C. 1999. Molecular Cloning and Sequencing of the Phosphoprotein (P) Gene of the Lapinized Rinderpest Virus. *Virus Genes*. 18: 175-178.
59. NEL, M., GJOBORI, T. 1986. Simple Methods for Estimating the Numbers of Synonymous and Non-Synonymous Nucleotide Substitutions. *Mol. Biol. Evol.* 3: 418-426.

60. NORBY, E., KOVAMEES, J., BLIXENKRONE-MOLLER, M., SHARMA, B., ORVELL, C. 1992. Humanized animal Viruses with Special Reference to Primate Adaptation of Morbillivirus. *Vet Microbiol.* 33: 275-286.
61. NORBY, E., SHESHBERDARAN, H., MCCULLOUGH, K., CARPETER, W., ORVELL, C., 1985. Is Rinderpest Virus the Archevirus of the Morbillivirus Genus? *Intervirol.* 23: 228-232
62. O'HALLORAN, T. 1993. Transition metals in control of gene expression. *Science.* 261: 715-725.
63. OIE. 2004. Rinderpest. Manual of Diagnostic Tests and Vaccines for Terrestrial Animals. Fifth Ed. Pp 142-152.
64. PLOWRIGHT, W., 1962. Rinderpest Virus. *Ann. N.Y. Acad. Sci.* 101: 548-563.
65. PTASHNE, M. 1988. How Eukaryotic Transcriptional Activators Work. *Nature.* 335: 683-689.
66. RAHA, T., KAUSHIK, R., SHAILA, M. 2004. Phosphoprotein P of Rinderpest virus binds to plus sense leader RNA: regulation by phosphorylation. *Virus Res.* 104: 191-200
67. RAHAMAN, A., SRINIVASAN, N., SHAMALA, N., SHAILA, M. 2004. Phosphoprotein of The Rinderpest Virus Forms a Tetramer through a Coiled Coil region Important for Biological Function. *J. Biol. Chemistry.* 279: 23606-23614
68. ROBBINS, S., FENIMORE, J., BUSSEL, R. 1980. Structural Phosphoproteins Associated with Measles Virus Nucleocapsids from Persistently Infected Cells. *J. Gen. Virol.* 48: 445-449.
69. SALMINEN, M., CARR, J., BURKE, D., MCCUTCHAN, F. 1995. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Res. Hum. Retroviruses.* 11: 1423-1425.
70. SCHAEFER, J., SCHONIGER, M. 1997. DISTREE: a Tool for Estimating Genetic Distances Between Aligned DNA Sequences. *Comput. Appl. Biosc.* 13: 445-451.
71. SELDMEIERS, R., NEUBERT, W. 1998. The Replicative Complex of Paramyxoviruses: Structure and Function. *Adv. Virus Research.* 50: 101-139.
72. SHAFFER, J., BELLINI, W., ROTA, P. 2003. The C Protein of Measles Virus Inhibits the Type I Interferon Response. *Virology.* 315: 389-397.
73. SHAJI, D., SHAILA, M. 1999. Domains of Rinderpest Virus Phosphoprotein Involved in Interaction with itself and the Nucleocapsid Protein. *Virology.* 258: 415-424.
74. SHIELL, B., GARDNER, D., CRAMERI, G., EATON, B., MICHALSKI, W. 2002. Sites of Phosphorylation of P and V Proteins from Hendra and Nipah Viruses: Newly Emerged Members of Paramyxoviridae. *Virus Research.* 92: 55-65.
75. SMALLWOOD, S., RYAN, K., MOYER, S. 1994. Deletion Analysis Defines a Carboxi-Terminal region of the Sendai Virus P Protein that Binds to the Polymerase L Protein. *Virology.* 202: 154-163.
76. STRIMMER, K., VON HAESELER, A. 1996. Quartet Puzzling: a Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol. Biol. Evol.* 13: 964-969.
77. SWEETMAN, D., MISKIN, J., BARON, M. 2001. Rinderpest Virus C and V Proteins Interact with the Major (L) Component of the Polymerase. *Virology.* 281: 193-204.
78. TAKEUCHI, K., KADOTA, S., TAKEDA, M., MIYAJIMA, N., NAGATA. 2003. Measles Virus V Protein Blocks Interferon (IFN)- $\alpha$  but not IFN- $\beta$  Signaling by Inhibiting STAT1 and STAT2 Phosphorylation. *FEBS.* 254: 177-182.
79. TAMURA, K., NEI, M. 1993. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Mol. Biol. Evol.* 10: 512-526.
80. TARBOURIECH, N., CURRAN, J., EBEL, C., RUIGROK, R., BURMEISTER, W. 2000. On the Domain Structure and the Polymerization State of the Sendai Virus P Protein. *Virology.* 266: 99-109.
81. THOMPSON, J., HIGGINS, D., GIBSON, T. 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.
82. TOBER, C., SEUFERT, M., SCHNEIDER, H., BILLETER, M., JOHNSTON, I., NIEWIESK, S., TERMEULEN, V., SCHNEIDER-SCHAULIES, S. 1998. Expression of Measles Virus V Protein is Associated with Pathogenicity and Control of Viral RNA Synthesis. *J. Virology.* 72: 8124-8132.
83. VON HEIJNE, G. 1992. Membrane Protein Structure Prediction: Hydrophobicity Analysis and the 'Positive Inside' Rule. *J. Mol. Biol.* 225: 487-494.
84. YANG, N., NIELSEN, N., GOLDMAN, N., PEDERSON, A. 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. *Genetics* 155: 431-449.
85. YONEDA, M., MIURA, R., BARRET, T., TSUKIYAMA-KOHARA, K., CHIEKO, K. 2004. Rinderpest Virus Phosphoprotein Gene is a Major Determinant of Species-Specific Pathogenicity. *J. Virology.* 78: 6676-6681.