

# Análisis de la dinámica poblacional del virus distemper canino mediante secuenciación masiva



MSc. Nicolás Sarute

Tesis de Doctorado PEDECIBA Biología

Subárea Genética

Orientadora: Dra. Yanina Panzera

Co-Orientador: Dr. Marco Vignuzzi

Sección Genética Evolutiva – Laboratorio de Poblaciones Virales y Patogénesis

Diciembre, 2014



## **Agradecimientos**

A la Sección Genética Evolutiva por permitirme realizar mis estudios de doctorado, especialmente a mi directora de tesis Yanina

A Marco Vignuzzi por haber co-dirigido este trabajo, y a los integrantes del laboratorio Poblaciones Virales y Patogénesis del Institut Pasteur Paris

A Alejandro y Lourdes por facilitarnos los casos clínicos y realizar valiosos aportes en el área veterinaria

A los integrantes de la Sección, compañeros de trabajo y amigos de la vida

Al Laboratorio de Virología Molecular del Centro de Investigaciones Nucleares por brindarnos sus instalaciones para el desarrollo de nuestras actividades experimentales

A María Victoria por su apoyo y amor incondicional

A la familia y amig@s que siempre han estado presentes

A los integrantes del Tribunal por su disposición e interés en nuestro trabajo

Al Programa de Desarrollo de las Ciencias Básicas (PEDECIBA), Comisión Académica de Posgrado (CAP) y Comisión Sectorial de Investigación Científica (CSIC) por su apoyo a través de becas para el desarrollo de mis estudios doctorales, para la realización de pasantías y asistencia a cursos y congresos en el exterior, así como para la difusión de nuestros resultados

Y muy especialmente al abuelo Félix quien siempre se preocupó y me acompañó durante toda la carrera, y que esperaba con ansias el momento de la defensa porque quería anunciar orgullosamente en el diario del pueblo que su nieto es Doctor

Gracias

## ÍNDICE

<b>Resumen</b>	<b>I</b>
<b>1. INTRODUCCIÓN</b>	<b>1</b>
1. 1 Genética y evolución de virus RNA	1
1.1.1 Generalidades	1
1.1.2 Mutación y tasas de mutación	2
1.1.3 Teoría de cuasiespecies	5
1.1.4 Compartimentalización tejido-específica y flujo genético	10
1.1.5 Análisis de la diversidad poblacional de virus RNA	12
1.2 Virus distemper canino	14
1.2.1 Clasificación taxonómica	14
1.2.2 Estructura, genoma y proteínas virales	14
1.2.2.1 Estructura viral	14
1.2.2.2 Genoma viral	15
1.2.2.3 Proteínas virales	16
1.2.2.3.1 Nucleoproteína (N)	16
1.2.2.3.2 Fosfoproteína P y proteínas V y C	16
1.2.2.3.3 Proteína de matriz (M)	17
1.2.2.3.4 Proteína de fusión (F)	17
1.2.2.3.5 Proteína hemaglutinina (H)	18
1.2.2.3.6 Proteína <i>large</i> (L)	18
1.2.3 Ciclo replicativo	19
1.2.3.1 Unión y fusión virus-célula	19
1.2.3.2 Transcripción y replicación del genoma viral	20
1.2.3.3 Formación del complejo ribonucleoproteico y brotación	20
1.2.3.4 Propagación entre células del hospedero	21
1.3 Distemper: patogénesis, sintomatología clínica y epidemiología	21
1.3.1 Patogénesis	21
1.3.2 Sintomatología Clínica	22
1.3.3 Epidemiología	23
1.4 Antecedentes del equipo de trabajo	25
1.5 Hipótesis de trabajo	27
1.6 Objetivos	27

1.6.1 Objetivo General	27
1.6.2 Objetivos Específicos	27
<b>2. MATERIALES Y MÉTODOS</b>	<b>28</b>
2.1 Muestras	28
2.2 Metodologías moleculares	29
2.2.1 Extracción de RNA	29
2.2.2 Retrotranscripción	30
2.2.3 Reacción de PCR	30
2.2.3.1 Detección del genoma viral	30
2.2.3.2 Amplificación del genoma completo	30
2.2.4 Electroforesis y purificación de DNA	32
2.2.5 Secuenciación por Sanger	32
2.3. Aislamiento viral en cultivo celular	32
2.3.1 Propagación de la línea celular Vero-DST	32
2.3.2 Preparación de inóculos	32
2.3.3 Titulación de los inóculos por dosis infectiva al 50% (TCID <sub>50</sub> )	34
2.3.4 Infección de la línea celular Vero-DST	34
2.3.5 Extracción del genoma viral	34
2.3.6 Ensayos de RT-PCR, electroforesis y purificación de amplicones	35
2.4 Secuenciación masiva (NGS)	35
2.4.1 Cuantificación del DNA purificado	35
2.4.2 Fragmentación y purificación de DNA por beads	35
2.4.3 Construcción de librerías	35
2.4.4 Secuenciación por síntesis	37
2.5 Análisis bioinformático	39
2.5.1 Análisis de secuencias nucleotídicas	39
2.5.2 Implementación del programa Viral Variance Analysis (ViVan)	39
2.5.3 Estimación de frecuencias de mutación y reconstrucción de secuencias consenso	40
2.5.4 Análisis filodinámico	41
<b>3. RESULTADOS</b>	<b>42</b>
3.1 Metodologías moleculares	42

3.1.1 Estandarización de la amplificación del genoma completo y secuenciación por Sanger	42
3.2 Aislamiento viral en cultivo celular	42
3.2.1 Infección de la línea Vero-DST con la cepa vacunal Onderstepoort	42
3.2.2 Titulación de los inóculos por dosis infectiva al 50% (TCID <sub>50</sub> )	43
3.2.3 Infección de la línea celular Vero-DST con inóculos de los casos clínicos y detección del genoma viral	44
3.2.4 Amplificación del genoma completo	45
3.3 Secuenciación masiva (NGS)	46
3.3.1 Construcción de librerías y secuenciación por síntesis	46
3.4 Análisis informático	47
3.4.1. Análisis de secuencias nucleotídicas	47
3.4.2 Implementación del programa Viral Variance Analysis (ViVan)	48
3.4.2.1 Identificación de variantes mayoritarias y minoritarias	49
3.4.2.2 Identificación de cambios sinónimos y no sinónimos	53
3.4.2.3 Identificación de cambios aminoacídicos debidos a variantes mayoritarias	53
3.4.2.4 Tasas de transición y transversión	56
3.4.3 Estimación de frecuencias de mutación y reconstrucción de secuencias consenso	57
3.4.3.1 Frecuencias de mutación	57
3.4.3.2 Reconstrucción de secuencias genómicas consenso	65
3.4.4 Análisis filodinámico	65
3.4.4.1 Análisis por <i>likelihood mapping</i>	65
3.4.4.2 Análisis de compartimentalización y migración	65
<b>4. DISCUSIÓN</b>	<b>74</b>
4.1 Conclusiones	89
4.2 Perspectivas	89
<b>5. REFERENCIAS BIBLIOGRÁFICAS</b>	<b>90</b>
<b>Anexo I:</b> Análisis de secuencias obtenidas por secuenciación masiva	II
<b>Anexo II:</b> Publicaciones	III
<b>Anexo III:</b> Avances de secuenciación masiva en el Instituto Pasteur Montevideo IV.	

## Resumen

El virus distemper canino (CDV) pertenece a la familia *Paramyxoviridae*, género *Morbillivirus*, el cual incluye diversos patógenos de importancia a nivel mundial que afectan al hombre (sarampión) y animales (peste bovina). CDV es el agente etiológico de una severa enfermedad sistémica denominada Distemper canino, caracterizada por la presencia de manifestaciones clínicas asociadas al sistema respiratorio, gastrointestinal y nervioso. Distemper afecta a todas las familias de carnívoros terrestres, varias especies de carnívoros marinos, e incluso a primates no homínidos. Los brotes ocurridos en primates y la adaptación de CDV al uso de receptores humanos *in vitro*, destacan el potencial zoonótico del virus. CDV es un virus envuelto y pleomórfico con un diámetro de 150-300 nanómetros, y presenta un genoma de RNA simple hebra, polaridad negativa, no segmentado y de aproximadamente 15,7 kilobases.

Los virus con genoma de RNA presentan tasas de mutación en el orden de  $10^{-3}$ - $10^{-5}$  por nucleótido copiado por ciclo de replicación, debido principalmente a la baja fidelidad de copia de las RNA polimerasas virales. Las elevadas tasas de mutación, junto a los enormes tamaños poblacionales y cortos tiempos generacionales de los virus RNA, determinan que éstos presenten una estructura y diversidad poblacional particular referida como cuasiespecies: distribución de variantes genéticas organizadas en torno a un genoma principal, las cuales forman una compleja “nube” de variantes a través de los sucesivos ciclos de replicación. Las variantes genéticas interactúan continuamente mediante los procesos de cooperación y complementación determinando las características de la población.

La dinámica poblacional de virus RNA también incluye a la diversificación de variantes dentro de los tejidos de un huésped (subpoblaciones) durante el transcurso de una infección, proceso conocido como compartimentalización tejido-específica. La migración de variantes entre compartimientos y la subsecuente colonización de diferentes tejidos determina, en última instancia, el curso patogénico de la infección.

El presente trabajo tuvo como objetivos analizar la diversidad poblacional de CDV, establecer su eventual organización en cuasiespecies, estimar patrones de compartimentalización y posibles eventos de migración de variantes entre tejidos/fluidos dentro de un huésped canino mediante la amplificación del genoma completo y secuenciación masiva. El genoma de CDV fue amplificado a partir de muestras de cerebelo, pulmón, médula espinal y orina/vejiga (matrices) provenientes de tres necropsias de perros domésticos con diagnóstico clínico y molecular de Distemper.

Las diferentes matrices fueron utilizadas como inóculos para la infección de la línea celular Vero-DogSLAMTag y se recuperaron los viriones para el aislamiento del genoma viral. El RNA viral fue utilizado para las reacciones de RT-PCR para amplificar al genoma completo en nueve fragmentos solapantes. Los nueve amplicones provenientes de las cuatro matrices de los tres casos clínicos fueron purificados y fragmentados para generar librerías genómicas para la secuenciación masiva.

El análisis de las secuencias reveló la presencia de miles de variantes genéticas características de cada matriz, demostrando la organización de CDV en un sistema de cuasiespecies. Se detectaron variantes genéticas representadas en distinta frecuencia dentro de la población, la frecuencia de las variantes minoritarias correspondió a 0.1-1%, mientras que las variantes mayoritarias presentaron una frecuencia >1%. La ocurrencia de cambios aminoacídicos en la proteína hemaglutinina, determinante antigénico del virus, debidos a variantes mayoritarias en dos casos clínicos, podría indicar una adaptación al neurotropismo. Los datos de secuenciación también permitieron estimar la frecuencia de mutación para cada una de las matrices analizadas. Los valores de frecuencia se ubicaron en el rango de  $0,24-3,08 \times 10^{-4}$ , los cuales son semejantes a los valores descritos para virus RNA.

El análisis de la diversidad de variantes reveló la existencia de subpoblaciones características (compartimentalización tejido-específica) en cada una de las matrices. Mediante aproximaciones heurísticas se realizó un análisis filodinámico con el objetivo de establecer posibles patrones de flujo génico de variantes genéticas entre las diferentes matrices dentro del huésped. Se determinó la matriz ancestral a partir de la cual se diseminó la infección para los tres casos clínicos. En un caso clínico se estableció que la matriz ancestral fue el pulmón, en un segundo caso fue el cerebelo, mientras que en el tercero dos matrices (cerebelo y pulmón) presentaron la misma probabilidad de ser la matriz ancestral. El análisis filodinámico junto con los datos clínicos de cada animal permitió estimar el curso patogénico de la infección de cada caso. Mediante este análisis proponemos que una cepa de los casos clínicos es neurovirulenta.

Nuestro trabajo constituye el primer estudio de diversidad poblacional para un miembro del género *Morbillivirus*, determinando la organización en cuasiespecies de CDV en tres casos de infecciones naturales, y sus posibles implicancias en la patogénesis viral.

## INTRODUCCIÓN

### 1. 1 Genética y evolución de virus RNA

#### 1.1.1 Generalidades

Los virus con genoma de RNA son los parásitos intracelulares más abundantes y exitosos del planeta siendo capaces de infectar huéspedes procariotas (bacterias y arqueas) y eucariotas (Elena, 2012). De acuerdo al sistema de clasificación viral de Baltimore, existen cuatro categorías correspondientes a virus RNA (Figura 1) (Baltimore, 1971). Grupo I) Virus RNA doble cadena: presentan genomas segmentados y cada segmento genómico es transcrito independientemente. Grupo II) Virus RNA simple cadena de polaridad positiva: el genoma viral tiene la misma polaridad que el RNA mensajero (mRNA). Para la generación de nuevos genomas se requiere de una molécula de RNA molde de polaridad negativa. Grupo III) Virus RNA simple cadena de polaridad negativa: la producción de mRNAs es el primer paso en la replicación y la transcripción viral, estas moléculas sirven de molde para la replicación genómica y son empleadas en la expresión del genoma. Los virus pertenecientes a los tres grupos descritos anteriormente se caracterizan por utilizar RNA polimerasas dependientes de RNA (vRdRP) para la producción de moléculas de RNA complementarias durante la replicación y expresión genómica. Grupo IV) Virus RNA simple cadena de polaridad positiva con intermediario de DNA: el genoma es de polaridad positiva pero no es utilizado como mRNA, sino que sirve de molde para la transcripción reversa mediada por una DNA polimerasa dependiente de RNA.

Los virus RNA presentan mecanismos de variación genética que incluyen la mutación, la recombinación y el *reassortment*. Las mutaciones ocurren de forma aleatoria en cada ciclo de replicación, mientras que los otros dos mecanismos de variación genética definen al “sexo viral” y requieren de la co-infección de dos o más partículas virales (Simone-Loriere & Holmes, 2011).

La mutación es de particular importancia en los virus RNA y conlleva numerosas implicancias genéticas y biológicas, y es el mecanismo de variación en el cual nos centraremos en esta Tesis.

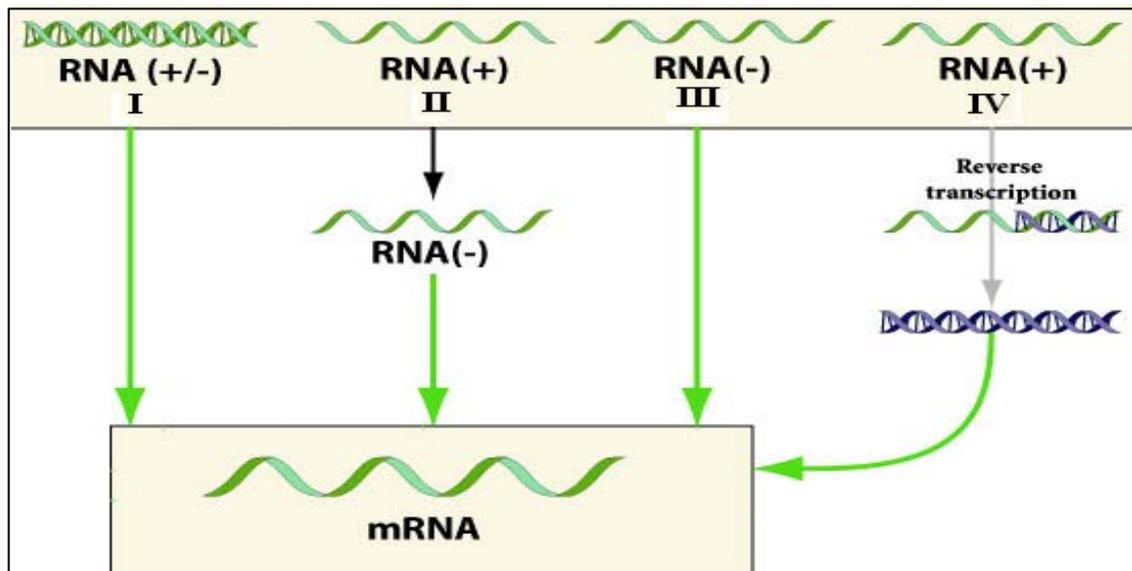


Figura 1. Clasificación de Baltimore para los virus RNA. Se detalla la arquitectura genómica y el modo de generación del RNA mensajero (mRNA) para los cuatro grupos (I, II, III, IV). Modificada del sitio web <http://viralzone.expasy.org/>.

### 1.1.2 Mutación y tasas de mutación

Una mutación se define como cualquier cambio transmisible en la información genética. Las mutaciones pueden clasificarse en base a su naturaleza molecular, siendo las sustituciones, las inserciones y las deleciones las más comunes. Las sustituciones son el tipo de mutación más simple ya que implican el cambio de un único nucleótido en el material hereditario. Hay dos tipos de sustituciones, las transiciones que implican el cambio de una purina (adenina, guanina) por otra purina y de una pirimidina (citosina, timina) por otra pirimidina, y las transversiones que son cambios de purina por una pirimidina y viceversa. El número de transversiones posibles es el doble que de transiciones pero estas últimas suelen ser más frecuentes (Figura 2). Las mutaciones también pueden clasificarse según su agente causal en espontáneas e inducidas. Las mutaciones espontáneas ocurren fundamentalmente por errores en la replicación, mientras que las inducidas son consecuencia de la exposición a mutágenos químicos o biológicos, o a radiaciones ionizantes (Watson et al., 2006).

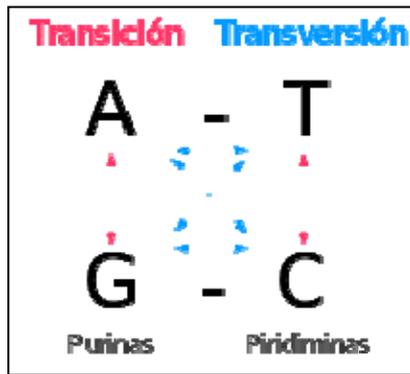


Figura 2. Representación de sustituciones nucleotídicas. En color rojo se detallan las transiciones y en azul las transversiones.

En virus RNA, las mutaciones son resultado de tres tipos de factores: intrínsecos (actividad de las RNA polimerasas, tamaño y polaridad del genoma, modo de replicación viral), dependientes del huésped

(patrones de expresión génica) y externos (mutágenos) (Combe & Sanjuán, 2014a). El error intrínseco está determinado principalmente por la actividad de las enzimas vRDRP, responsables de copiar el genoma de RNA en una molécula de RNA complementario. Las vRDRP carecen generalmente de actividad correctora (*proofreading*) lo cual influye directamente sobre su fidelidad de copia y el número de errores introducidos durante la replicación viral. La gran mayoría de los virus RNA no presentan mecanismos *proofreading*, y a la fecha solamente se ha demostrado la presencia de esta actividad en la familia *Coronaviridae* (Smith et al., 2013).

Otros factores intrínsecos que influyen en la tasa de mutación son el tamaño y la polaridad del genoma. Se ha postulado que existe una correlación negativa entre el tamaño del genoma y la ocurrencia de mutaciones, es decir, los virus con genomas más pequeños tienden a acumular un número mayor de mutaciones en menor tiempo. Sin embargo, esta correlación podría estar sesgada debido a que los virus con genomas de mayor longitud (*Coronaviridae*) son justamente los que presentan actividad *proofreading* (Combe & Sanjuán, 2014a). La polaridad genómica influye directamente sobre la ocurrencia de mutaciones. En virus simple hebra de polaridad positiva (Grupo II) el RNA genómico se encuentra desnudo y tiende a formar estructuras secundarias que favorecen la ocurrencia de mutaciones por cambios en el marco de lectura. Por el contrario, los virus de simple hebra de polaridad negativa (Grupo III) portan sus propias vRdRP en el virión, y el RNA genómico está protegido por nucleoproteínas disminuyendo la posibilidad de daño en el genoma (Combe & Sanjuán, 2014a).

Otro factor intrínseco a considerar es el modo de replicación viral. Durante la replicación viral se sintetizan RNAs intermediarios (antigenomas) que se emplean como molde para la producción de nuevos genomas. El número de genomas utilizados en la

producción de antigenomas influirá sobre la abundancia y distribución de las mutaciones en estas moléculas (Sardanyés et al., 2009).

Respecto a los factores dependientes del huésped, se han descrito proteínas desaminasas en eucariotas que generan mutaciones (transiciones) involucradas en procesos de hipermutación viral. La proteína APOBEC3 es una citidina-desaminasa capaz de editar el DNA complementario (cDNA) de retrovirus y producir hipermutaciones G-A, mientras que la proteína ADAR, una adenosina-desaminasa, produce hipermutaciones A-G en paramyxovirus y rhabdovirus mediante la edición del mRNA por un mecanismo post-transcripcional (Combe & Sanjuán, 2014b).

Por último, el daño inducido por agentes externos como mutágenos químicos es de especial importancia ya que determina la ocurrencia de un gran número de mutaciones que impactan negativamente sobre la adaptabilidad y eficiencia replicativa del virus, pudiendo determinar su extinción (Sanjuán et al., 2010).

El número de errores por nucleótido incorporado por ciclo de replicación/infección se define como tasa de mutación. En virus RNA, los factores intrínsecos, asociados al huésped y externos determinan tasas de mutación en el rango de  $10^{-3}$ - $10^{-5}$ , siendo las tasas más elevadas descritas. En virus DNA, las tasas de mutación son menores ubicándose en el orden de  $10^{-8}$  al igual que en bacterias y eucariotas inferiores (Figura 3) (Duffy et al., 2008).

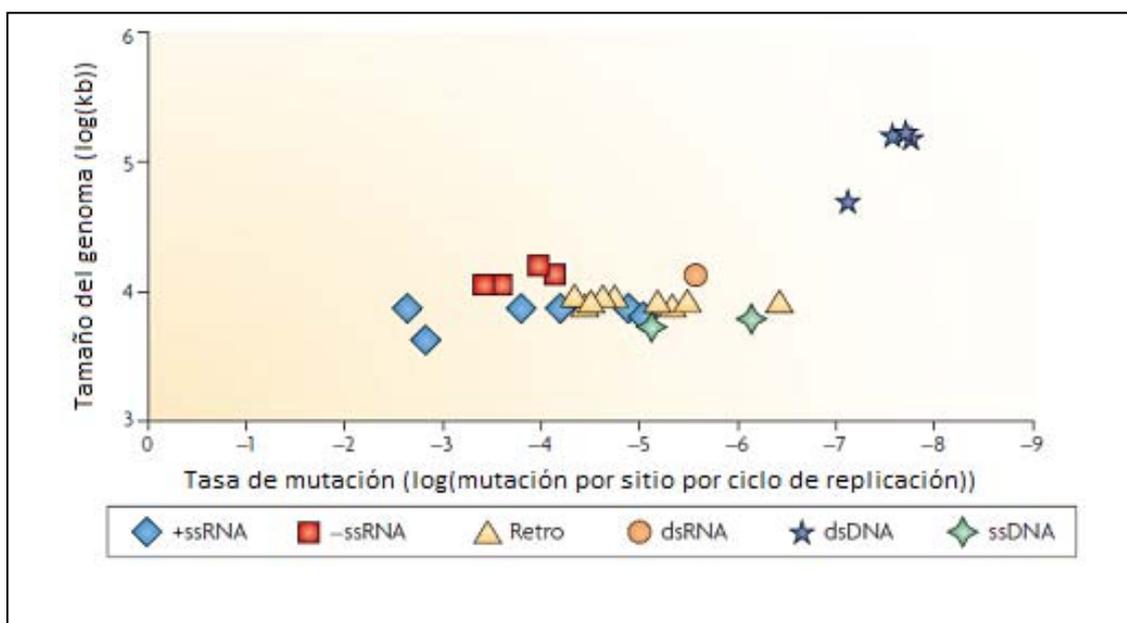


Figura 3. Tasas de mutación espontáneas en virus con distinta organización genómica ajustadas por la tasa de replicación. Modificado de Duffy et al., 2008.

Las elevadas tasas de mutación, sumado a tiempos generacionales cortos y enormes tamaños poblacionales determinan el éxito de los virus RNA como parásitos (Elena, 2012). La influencia del ciclo vital y del tamaño poblacional en la biología de los virus RNA puede ejemplificarse con el virus de la inmunodeficiencia humana tipo 1 (HIV-1). El tiempo generacional *in vivo* del virus se ha estimado en 30 horas, durante el cual se producen  $10^{10}$  partículas virales. Durante una infección, la replicación viral ocurre en el huésped de forma constante con tasas de mutación de  $10^{-4}$ - $10^{-6}$  (Perelson et al., 1996). La combinación de tasas de mutación elevadas y la generación constante de prole determina la formación de un enorme repertorio de variantes (mutantes) relacionadas genéticamente, las cuales constituyen en su conjunto una población referida como cuasiespecie viral (Elena, 2012).

### 1.1.3 Teoría de cuasiespecies

El término cuasiespecies fue adoptado en virología en base a la teoría de Manfred Eigen para describir el comportamiento y evolución de macromoléculas auto-replicativas (Eigen, 1971). De acuerdo a esta teoría, a partir de una copia maestra de una molécula se producen diversos mutantes con distinta probabilidad en cada ciclo de replicación. Eigen acuñó el término *comet tail* para referirse a la distribución y ordenamiento de las copias mutantes, con el fin de explicar la replicación propensa al error desde un punto de vista molecular (Eigen & Schuster, 1978).

Esta teoría inicial fue desarrollada posteriormente junto a Peter Schuster, definiendo a las cuasiespecies como una distribución estacionaria de mutantes, organizadas en torno a una secuencia maestra que domina el espectro de mutantes (Eigen & Schuster, 1979). La teoría de cuasiespecies establece una condición necesaria para la conservación de la información genética en un sistema replicativo con elevadas tasas mutacionales referida como umbral de error. En un sistema replicativo con cierta información genética no redundante existe una tasa máxima de error compatible con su mantenimiento y en los casos donde se supera este umbral el número de variantes viables se reducirá drásticamente (Eigen & Schuster, 1979). La primera evidencia experimental del comportamiento en cuasiespecies de un virus RNA se obtuvo analizando la replicación del bacteriófago Q $\beta$  en su huésped natural *Escherichia coli*, donde se identificaron un

gran número de secuencias individuales que diferían entre si, en lugar de una estructura genómica única y definida (Domingo et al., 1978).

Actualmente, la teoría de cuasiespecies analiza las consecuencias de la replicación propensa al error de los virus RNA y su importancia en la evolución genómica (Lauring & Andino, 2010). La replicación característica de estos virus y los enormes tamaños poblacionales son beneficiosos a largo plazo ya que pueden generarse combinaciones de variantes (genotipos) con mutaciones de carácter adaptativo para la población, sin embargo son perjudiciales a corto término debido a que la mayoría de las mutaciones generadas tienen efectos deletéreos. Los virus RNA han optimizado finamente sus tasas de mutación durante la evolución para que no ocurran demasiadas mutaciones que determinarían la extinción viral, ni para que su ocurrencia sea aislada lo cual impediría al virus afrontar presiones selectivas (Lancaster & Pfeiffer, 2012). Las variantes potencialmente beneficiosas dentro de la población otorgan a la cuasiespecie una probabilidad mayor de adaptarse a nuevos ambientes y evadir respuestas del huésped durante una infección (Vignuzzi et al., 2006).

Para comprender la distribución de las variantes en la cuasiespecie, se representa la replicación del genoma principal y la generación de miles de genomas que difieren en promedio en una única posición nucleotídica. En las sucesivas rondas de replicación, se generará una distribución de mutantes cada vez más compleja y divergente. Este conjunto de mutantes constituye una población heterogénea que se distribuyen formando una “nube” de variantes en torno al genoma principal (Figura 4) (Lauring & Andino, 2010, Wright et al., 2011).

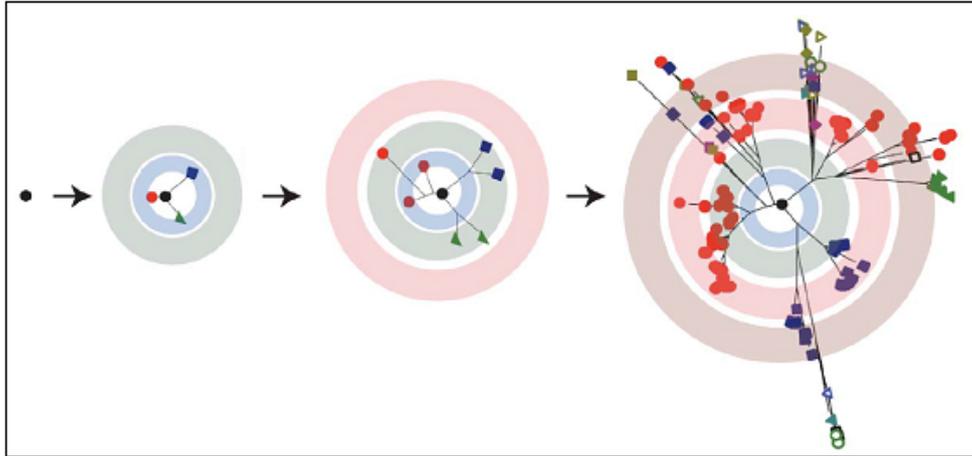


Figura 4. Representación esquemática de la generación de variantes genéticas a partir de un único genoma. Los círculos concéntricos representan ciclos de replicación y la distancia genética entre dos variantes se representa por líneas proporcionales a la divergencia. Tomado de Luring & Andino, 2010.

La organización genética de las poblaciones virales se modela dentro del *sequence space*: representación geométrica de todos los posibles genotipos de la población donde las distancias físicas reflejan distancias genéticas. Conforme aumentan las rondas de replicación, las variantes en la población se encontrarán más alejadas en el *sequence space*. Además, se ha propuesto que las poblaciones virales evolucionan dentro de un *fitness landscape*, donde el nivel basal es la representación de todos los genotipos del *sequence space* y la altura dentro de cada región representa el *fitness* de un genotipo en particular (Figura 5).

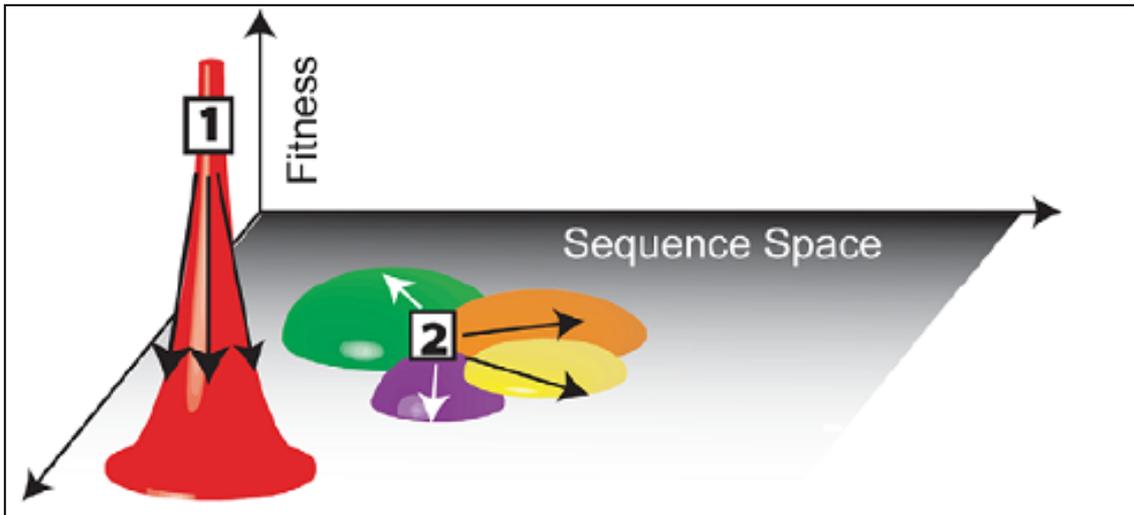


Figura 5. Representación del *fitness landscape* de un virus RNA. El plano horizontal corresponde a todos los genotipos que pueden hallarse en la población (*sequence space*), mientras que el eje vertical representa el *fitness* de cada genotipo. En 1 se representa un genotipo con escasa diversidad genética con un pico alto correspondiente a un *fitness* elevado. En 2 el genotipo tiene una mayor diversidad ocupando una región más vasta del *sequence space* pero con un *fitness* menor. Tomado de Lauring et al., 2012.

El ambiente y las presiones selectivas determinan los límites del *landscape* correspondiente a cada genotipo, y la adaptación a cierto ambiente involucra un movimiento mutacional de un punto a otro (Lauring & Andino, 2010). Una población con una gran diversidad genética puede contener o ser capaz de adquirir rápidamente mutaciones necesarias para sobrevivir en un ambiente cambiante o frente a presiones selectivas (Lauring et al., 2012). La estructura de una población viral replicándose en un ambiente estable se caracteriza por un pico dentro del *fitness landscape* que representa a la secuencia consenso rodeada por un arreglo de variantes en baja frecuencia (Figura 5) (Acevedo et al., 2013). Las secuencias consenso representan la composición promedio del nucleótido más común en cada posición del genoma, y pueden no existir como tal dentro de la población que definen. Esta secuencia puede modificarse por eventos estocásticos (deriva génica) o selectivos a nivel poblacional. El aumento en la frecuencia de ciertos mutantes y el reemplazo de la distribución previa del espectro de mutantes determina la generación de una nueva secuencia consenso. Sin embargo, una secuencia consenso invariante no implica la ausencia de mutaciones en la población, de hecho pueden ocurrir perturbaciones biológicas en el repertorio de mutantes sin modificar esta secuencia (Domingo et al., 2012). La secuencia consenso identifica, por

tanto, las posiciones nucleótidas predominantes dentro de una población pero no es informativa sobre la nube de variantes subyacente (Wright et al., 2011).

La generación continua de variantes genéticas a nivel poblacional cumple un rol fundamental en la adaptación viral ya que los virus deben sortear diversos obstáculos durante su ciclo vital para poder transmitirse de forma eficiente. Entre éstos se destacan la evasión de la respuesta inmune, el incremento del *fitness* replicativo en cierto tipo celular, y/o la modificación del tropismo celular a través del uso de nuevos receptores celulares (Bordería et al., 2011).

Un estudio realizado con poliovirus (*Picornaviridae*) aportó evidencia directa sobre la acción conjunta de las variantes de la cuasiespecie mediante el proceso de complementación. Para ello, inocularon ratones con un clon neurovirulento en el contexto de una población viral genéticamente diversa (*wild type*) y otra de diversidad reducida. En el caso de la población *wild-type* ocurrió infección del sistema nervioso central (SNC), mientras que en la población con menor diversidad el clon no fue capaz de alcanzar el SNC (Figura 6). Este estudio demostró que la diversidad de la cuasiespecie *per-se* determina cambios en la patogénesis viral, y definieron un concepto central para la teoría de cuasiespecies: la colonización exitosa de un ecosistema (tejido/huésped) ocurre por la cooperación de variantes que ocupan distintas regiones del *sequence space* (complementación), y por tanto el repertorio de mutantes es objeto de la selección natural y no las variantes individuales (Vignuzzi et al., 2006).

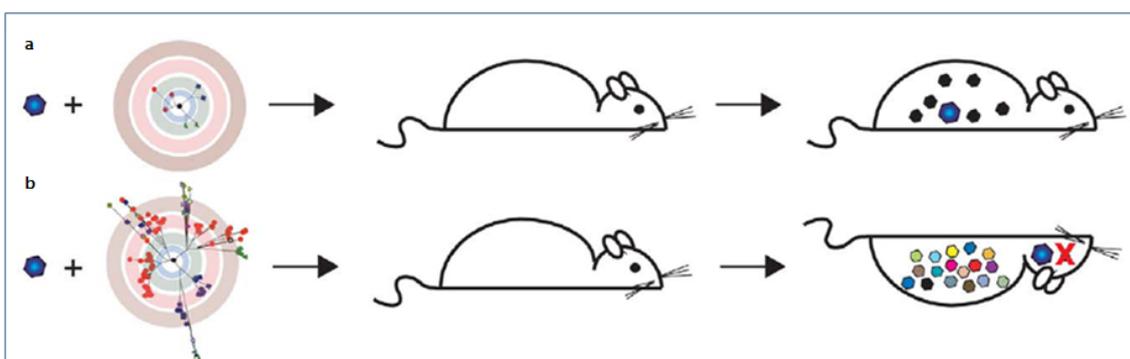


Figura 6. Representación de la complementación entre las variantes de la cuasiespecie. a) un clon neurovirulento es inoculado en el contexto de una población con diversidad genética reducida siendo incapaz de alcanzar el sistema nervioso central (SNC). b) en el contexto de una población heterogénea el clon neurovirulento alcanza el SNC provocando la muerte del animal. Tomado de Luring & Andino, 2010.

Una forma extrema de adaptación que generalmente involucra un gran número de mutaciones y etapas adaptativas lo constituye el cambio de huésped. La población inicial debe generar al menos un mutante que sea capaz de interactuar con el nuevo huésped, si el espectro de mutantes es genéticamente diverso la ocurrencia de este evento será más probable. Una vez que el/los mutante/s alcanzan el nuevo huésped, deben ocurrir cambios adaptativos como el uso de nuevos receptores celulares en un tiempo relativamente corto para asegurar un ciclo infeccioso productivo. Finalmente, y quizás lo más dificultoso, es que el virus sea capaz de evolucionar y transmitirse entre los miembros de la nueva especie una vez que el salto inter-específico ha ocurrido. El virus Influenza (*Orthomyxoviridae*) es un excelente modelo para el estudio del cambio de huésped entre especies no relacionadas. Para este virus se han descrito mutaciones que afectan la actividad de la polimerasa viral modificando el espectro de variantes a nivel poblacional con la capacidad de mediar la adaptación de cepas aviarias a huéspedes mamíferos (Gabriel et al., 2005, 2007).

#### **1.1.4 Compartimentalización tejido-específica y flujo genético**

La compleja organización de las variantes dentro de la cuasiespecie puede relevarse entre individuos de una población y/o entre tejidos de un mismo huésped (Figura 7). Estudios realizados con el virus de la hepatitis C (*Flaviviridae*) mostraron la existencia de subpoblaciones con distinta composición genética en células de sangre periférica, hígado y linfocitos, lo cual dio origen al concepto de compartimentalización tejido-específica (Navas et al., 1998).

Los diferentes patrones de compartimentalización pueden explicarse por diferencias en la susceptibilidad genética del huésped, presencia de barreras anatómicas, y tasas de replicación viral y mutación diferenciales. La compartimentalización puede involucrar eventos de deriva génica y efectos fundadores, y/o ser consecuencia de presiones selectivas. En términos generales, una subpoblación puede presentar un *fitness* elevado en un tejido, y una marcada reducción del mismo cuando se ubica en otro tejido.

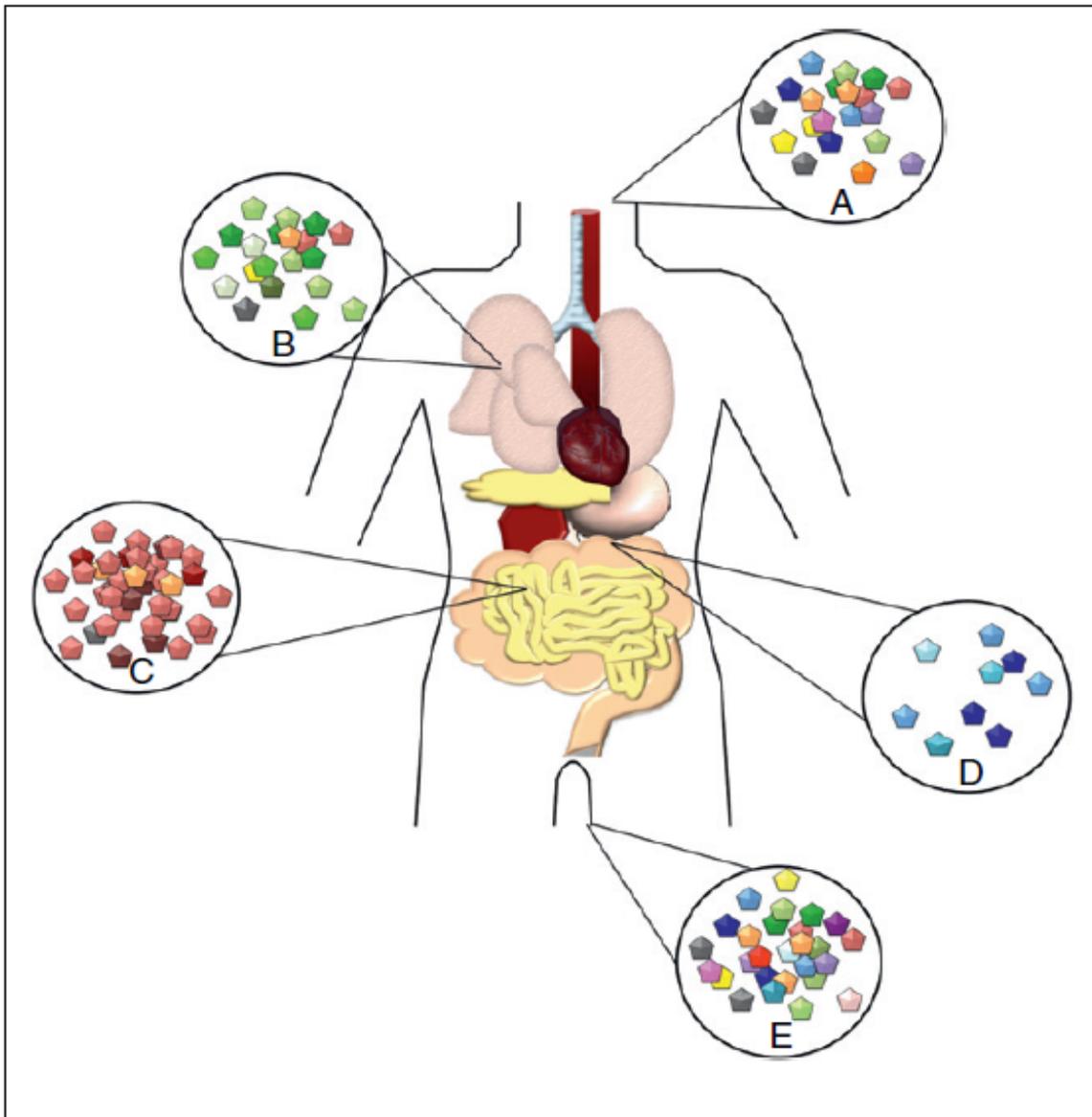


Figura 7. Representación de subpoblaciones virales (A-E) (compartimentalización tejido-específica) dentro del huésped. Tomado de Lancaster & Pfeiffer, 2012.

Por tanto, para que una infección sea exitosa el virus debe ser capaz de sobrevivir en diferentes ambientes celulares y evadir mecanismos antivirales. La resolución de estas

presiones determina el tropismo tisular, la presencia de supoblaciones características entre tejidos, y el curso patogénico de la infección (Vignuzzi et al., 2006). La diversidad poblacional y la compartimentalización están íntimamente relacionadas en tanto que la cooperación de variantes genéticamente diversas (complementación) favorece la colonización de nuevos tejidos mediante eventos de migración de variantes (Navas et al., 1998; Vignuzzi et al., 2006). Debido a esta dinámica poblacional, es difícil analizar el *fitness* replicativo de un virus en el contexto de un modelo de infección y aún más dentro de un huésped (Vignuzzi & Andino, 2012).

Las aproximaciones basadas en análisis filodinámicos permiten establecer patrones de migración combinando análisis filogenéticos y poblacionales para correlacionar el comportamiento evolutivo y patogénico del virus (Figura 8) (Salemi et al., 2005, Ramirez et al., 2012). Los análisis de migración o flujo genético permiten identificar al tejido ancestral y a la subpoblación asociada a la fase inicial de la infección, y estimar la posterior diseminación y diversificación del virus a otros compartimientos (Salemi et al., 2005).

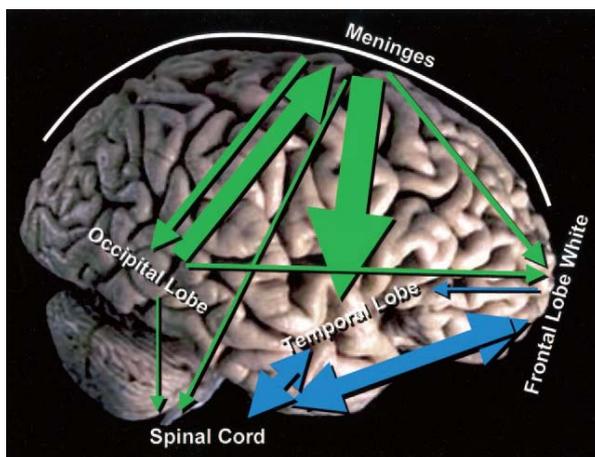


Figura 8. Red de migración de variantes genéticas del virus VIH-1 entre tejidos del sistema nervioso central. Las flechas indican la polaridad de la migración y su tamaño representa la proporción de migrantes de un tejido a otro. Tomado de Salemi et al., 2005.

### 1.1.5 Análisis de la diversidad poblacional de virus RNA

El análisis de la diversidad poblacional se basa en la identificación de la composición nucleotídica de las variantes genéticas presentes en la población viral. El método de secuenciación de Sanger es ampliamente utilizado para la identificación viral y análisis filogenéticos, pero presenta muy baja resolución para estudios poblacionales ya que se basa en una única lectura de secuencia, y es capaz de detectar dos variantes para cada posición nucleotídica solamente si están representadas en elevada frecuencia (30-40%). El método de secuenciación de Sanger no permite determinar la secuencia nucleotídica

de una molécula de RNA genómico, sino de una multitud de genomas que definen a la secuencia consenso de la población (Isakov et al., 2014).

Con el fin de aumentar la sensibilidad en la caracterización de las poblaciones virales, se ha empleado la clonación de productos de PCR y posterior secuenciación de Sanger. La secuenciación de clones individuales permite analizar con mayor resolución la heterogeneidad poblacional debido a que cada clon representa, de forma ideal, a una única variante genética (Gong et al., 2013). Además, la utilización de enzimas *high fidelity* que poseen una fidelidad de copia dos órdenes mayor respecto a las *Taq* polimerasas convencionales disminuyen los errores en la reacción de PCR, que podrían ser interpretados como cambios nucleotídicos en las variantes de la población (Beaucourt et al., 2011). Esta estrategia se ha utilizado para analizar la heterogeneidad poblacional del virus de la aftosa (*Picornavirus*), hepatitis C (*Flavivirus*), norovirus humano (*Calicivirus*), entre otros (Chen & Wang, 2007, Carlson et al., 2009, Cottam et al., 2009). Si bien la resolución de la clonación es mayor que la secuenciación directa por Sanger, este procedimiento aplicado al análisis de la diversidad poblacional demanda tiempo, es costoso y laborioso, ya que se deben analizar como mínimo 50-100 clones por amplicón. Además, el análisis poblacional mediante esta estrategia debe restringirse a regiones genómicas parciales, debido a que los virus RNA poseen genomas que alcanzan los 32 kb, y a que generalmente los fragmentos clonados no superan los 1000 pb (Beaucourt et al., 2011, Gong et al., 2013).

Un gran avance en el análisis de la diversidad poblacional ha sido la utilización de la tecnología de secuenciación masiva (NGS), debido a su alto rendimiento y a la enorme cantidad de información que se obtiene en una sola corrida (Gong et al., 2013). La secuenciación masiva genera miles de *reads* para cada posición secuenciada, obteniéndose una cobertura ultra-profunda de 50.000 *reads* en promedio por nucleótido analizado. Los estudios de diversidad poblacional mediante NGS utilizando enzimas *high fidelity* han mostrado mejoras sustanciales ya que permiten estudiar con mayor profundidad al repertorio de variantes de las poblaciones, analizar genomas completos independientemente de su longitud y reducir el tiempo del protocolo de secuenciación (Isakov et al., 2014). Recientemente, se han desarrollado nuevas estrategias basadas en NGS para analizar con una resolución aún mayor la diversidad poblacional. El método CirSeq se basa en la secuenciación de fragmentos de RNA repetidos en tándem que se

unen físicamente formando un círculo rodante. Esto permite obtener información redundante (por triplicado) para cada posición genómica permitiendo discriminar cambios nucleotídicos de posibles errores asociados a la secuenciación (Acevedo et al., 2013).

La tecnología de NGS se ha empleado para analizar la dinámica poblacional, respuesta a tratamientos antivirales, y la progresión de enfermedades en diversas familias virales. Dentro de las familias más estudiadas se encuentran *Flaviviridae* (Dengue), *Hepadnaviridae* (Hepatitis B), *Orthomyxoviridae* (Influenza), *Picornaviridae* (Poliovirus), *Retroviridae* (HIV-1) y *Togaviridae* (Chikungunya) (Domingo et al., 2012).

Para virus pertenecientes a la familia *Paramyxoviridae* se han realizado estudios de diversidad poblacional en tres géneros mediante clonación de fragmentos genómicos y secuenciación de Sanger (Barbezange & Justin, 2005, Sauder et al., 2006, Deplanche et al., 2007, Kattenbelt et al., 2010), y recientemente por secuenciación masiva mediante el análisis de un genoma completo (Abolnik et al., 2012). Sin embargo, no se cuentan con estudios para el género *Morbillivirus* al cual pertenece el virus distemper canino, objeto de estudio de esta Tesis.

## **1.2 Virus distemper canino**

### **1.2.1 Clasificación taxonómica**

El virus distemper canino (CDV) es un virus envuelto del género *Morbillivirus* perteneciente a la familia *Paramyxoviridae*, orden Mononegavirales. La familia *Paramyxoviridae* se divide en dos subfamilias: *Paramyxovirinae* incluye siete géneros (*Aquaparamyxovirus*, *Avulavirus*, *Ferlavirus*, *Henipavirus*, *Morbillivirus*, *Respirovirus*, *Rubulavirus*), y *Pneumovirinae* incluye dos géneros (*Metapneumovirus*, *Pneumovirus*) (Lamb & Parks, 2007).

### **1.2.2 Estructura, genoma y proteínas virales**

#### **1.2.2.1 Estructura viral**

CDV es un virus envuelto y pleomórfico con un diámetro entre 150 y 300 nm (Zipperle et al., 2010). El RNA genómico se encuentra empaquetado por la nucleoproteína (N) y es replicado por el complejo de la polimerasa viral formado por la proteína *large* (L) y

su cofactor, la fosfoproteína (P). Las proteínas N, P y L junto al RNA viral forman el complejo ribonucleoproteico (RNP), el cual dirige la síntesis secuencial de mRNA a partir de los genes virales y la replicación de los antígenomas (RNAs de polaridad positiva). La envoltura contiene dos proteínas integrales de membrana, la proteína de fusión (F) y la hemaglutinina (H), y una proteína asociada a la membrana que interactúa con el complejo RNP, la proteína de matriz (M) (Figura 9) (von Messling et al., 2001).

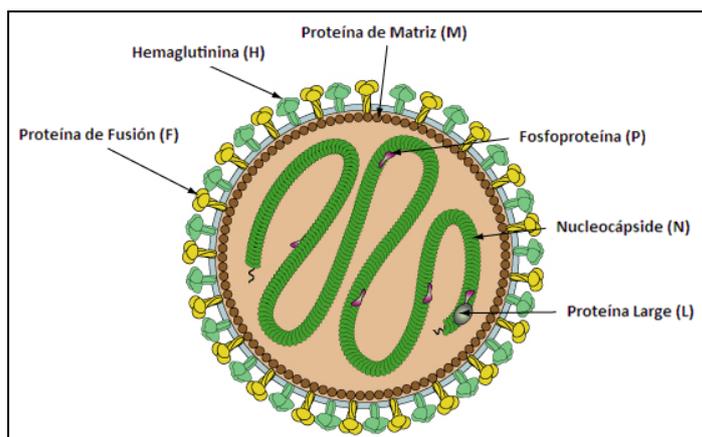


Figura 9. Representación esquemática del virión de un *Morbillivirus*. Modificada del sitio <http://viralzone.expasy.org/>.

### 1.2.2.2 Genoma viral

CDV presenta un genoma de RNA no segmentado, cadena simple, polaridad negativa y aproximadamente 15.7 kb (Lamb & Parks, 2007). El genoma posee seis genes denominados N, P/V/C, M, F, H y L que codifican las seis proteínas estructurales del virión. Cada gen codifica una única proteína, excepto el gen P/V/C que codifica la fosfoproteína P y dos proteínas no estructurales denominadas V y C. En los extremos del RNA genómico existen regiones UTR no codificantes, denominadas *leader* 3' y *trailer* 5', que son esenciales para la replicación y transcripción (Lamb & Parks, 2007) (Figura 10). Entre genes adyacentes existe un triplete intergénico consenso (GAA) que no se transcribe y que separa al sitio de poliadenilación de un gen, del sitio de inicio de la transcripción del gen contiguo (Liermann et al., 1998).

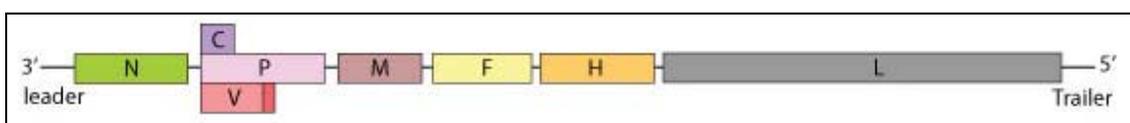


Figura 10. Representación esquemática del genoma de CDV. Imagen tomada del sitio web <http://viralzone.expasy.org/>.

Los genes son transcritos por el complejo RNP a partir de un promotor simple localizado en la región *leader* mediante un mecanismo denominado *start-stop*. Este

mecanismo implica que la transcripción se inicie en el primer gen del extremo 3' (N) y se detenga en cada región intergénica. Esto ocurre debido a que la RNA polimerasa viral puede escindirse del RNA molde, determinando la interrupción del proceso. Este mecanismo, por tanto, conduce a un gradiente transcripcional que se mantiene durante la infección, de tal modo que los genes próximos al extremo 3' del genoma se transcriben en mayor proporción que aquellos cercanos al extremo 5' (Figura 11) (Anderson & von Messling, 2008).

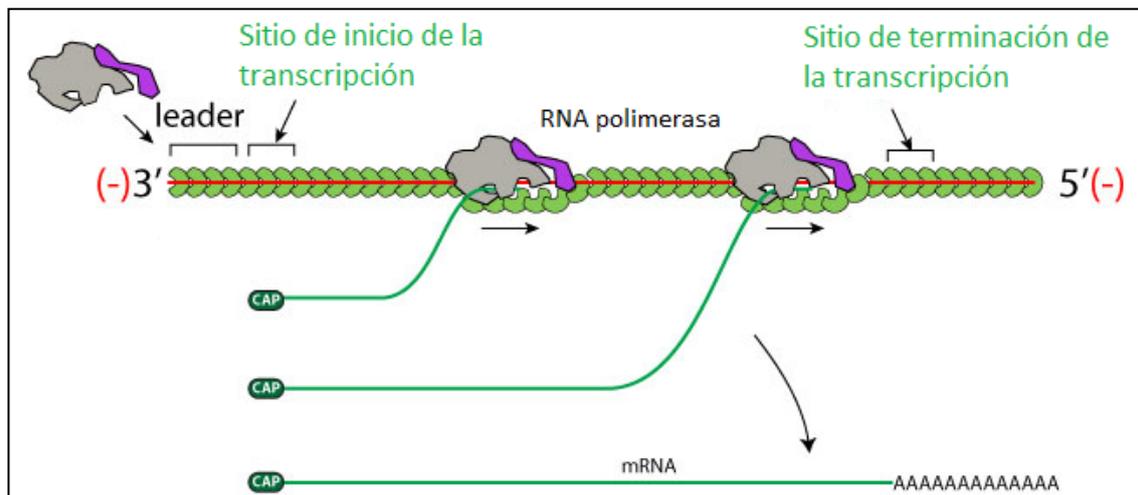


Figura 11. Mecanismo de transcripción de *Morbillivirus*. Imagen modificada del sitio web <http://viralzone.expasy.org/>.

### 1.2.2.3 Proteínas virales

#### 1.2.2.3.1 Nucleoproteína (N)

La nucleoproteína es codificada por el gen N y presenta 525 aminoácidos (aa). Esta proteína se autoensambla sobre el genoma viral y el RNA antisentido para formar el complejo RNP junto a las proteínas P y L (Figura 9) (Lamb & Parks, 2007).

#### 1.2.2.3.2 Fosfoproteína P y proteínas V y C

La fosfoproteína P está constituida por 507 aa y es un co-factor de la polimerasa que se activa por fosforilación y forma parte activa del complejo RNP (Figura 9) (Lamb & Parks, 2007).

La proteína C es traducida a partir del mismo mRNA que P, pero presenta un codón de inicio alternativo situado 22 nucleótidos corriente abajo del codón iniciador de la proteína P. Esto produce un corrimiento en el marco abierto de lectura (ORF) y la

generación de un codón de terminación, determinando que la proteína C tenga una extensión de 174 aa (Bellini et al., 1985).

La proteína V consta de 299 aa y es traducida a partir del mismo codón de inicio que la proteína P. Sin embargo, mediante la edición del mRNA se añade una guanina que determina un corrimiento en el ORF y la formación de un codón de terminación prematuro. Ambas proteínas presentan la misma secuencia aminoacídica en los primeros 231 aa, pero los 68 aa del extremo carboxi-terminal de la proteína V difieren respecto a la proteína P (Cattaneo et al., 1989).

Se ha sugerido que las proteínas accesorias V y C están involucradas en la evasión de la respuesta inmune innata. Específicamente, la proteína V interfiere con las proteínas STAT2 y mad5 implicadas en las vías de señalización celular inhibiendo la respuesta del interferón  $\beta$  (Svitek et al., 2014).

#### **1.2.2.3.3 Proteína de matriz (M)**

La proteína de matriz es codificada por el gen M y está constituida por 335 aa. Esta proteína se posiciona debajo de la envoltura, e interactúa con el complejo RNP y con las colas citoplasmáticas de las glicoproteínas de membrana H y F (Figura 9). La proteína M tiene un papel fundamental en la morfología y el ensamblaje del virión, debido a las interacciones que establece (Lamb & Parks, 2007).

#### **1.2.2.3.4 Proteína de fusión (F)**

La proteína de fusión es una glicoproteína transmembrana tipo I de 662 aa que participa en la fusión de la envoltura viral con la membrana plasmática de la célula hospedero (Lamb & Parks, 2007). La proteína F es codificada por el gen F bajo la forma de un precursor inactivo denominado pre-F0. El procesamiento de la proteína implica el reconocimiento del péptido señal (Fsp) por una molécula específica y el subsecuente transporte hacia el retículo endoplásmico, donde el precursor pre-F0 es procesado co-traduccionalmente (von Messling & Cattaneo, 2002). El procesamiento ocurre entre los aminoácidos 135-136 por la acción de una peptidasa celular (SPase), generando al Fsp de 135 aa y al precursor inmaduro F0 de 527 aa. El precursor F0 es glicosilado y clivado por una furina celular que actúa en el compartimiento del Golgi, formando las subunidades F1 y F2. Estas subunidades forman un heterodímero que constituye la

forma activa de la proteína F (Plattet et al., 2007). A nivel de la envoltura, la proteína forma un trímero que interactúa con la hemaglutinina durante el proceso de fusión virus-célula (Figura 12) (von Messling et al., 2004).

#### 1.2.2.3.5 Proteína hemaglutinina (H)

La hemaglutinina es una glicoproteína transmembrana tipo II de 607 aa codificada por el gen H. Esta proteína media la unión de los viriones a los receptores celulares, y posee la capacidad de promover la fusión virus-célula (Lamb & Parks, 2007). La proteína H presenta un dominio citoplasmático corto en su extremo amino-terminal, un dominio hidrofóbico transmembrana con funciones de localización y anclaje a la membrana, y un ectodominio carboxilo-terminal (Zipperle et al., 2010). Dicha proteína forma un tetramero a nivel de la envoltura que interactúa físicamente con la proteína F (Figura 12) (von Messling et al., 2004). La hemaglutinina es el determinante principal del tropismo celular del virus. En macrófagos, células dendríticas y linfocitos B y T reconoce y se une al receptor celular CD150/SLAM (Signaling Lymphocytic Activation Molecule) (Stern et al., 1995, von Messling et al., 2001), mientras que en células epiteliales del tracto respiratorio y queratinocitos interactúa con el receptor Nectina-4 (Poliovirus Receptor-like-4) (Muhlebach et al., 2011).

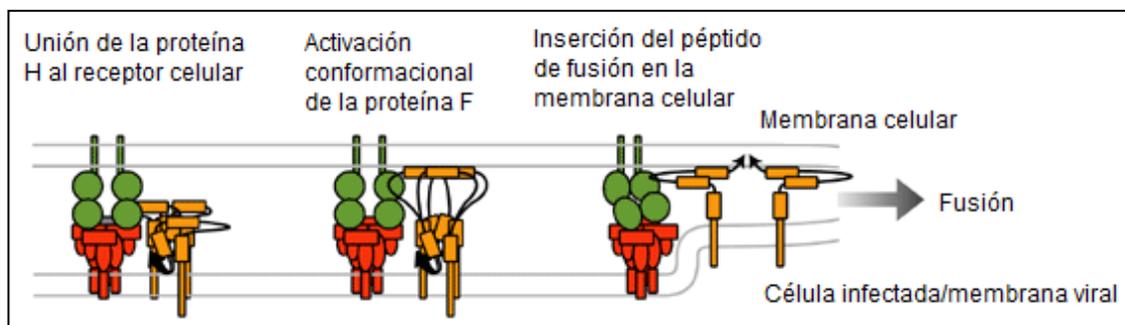


Figura 12. Representación del mecanismo de fusión virus-célula mediado por la hemaglutinina y la proteína de fusión. Modificado de Expert Reviews in Molecular Medicine. Cambridge University Press (2002).

#### 1.2.2.3.6 Proteína *large* (L)

La proteína *large*, codificada por el gen L, consta de 2184 aa y es la subunidad fundamental del complejo de la RNA polimerasa debido a su rol catalítico en la síntesis del RNA viral. La proteína L interactúa con la fosfoproteína P para formar el complejo

activo de la polimerasa y es constituyente integral de la nucleocápside (Figura 9) (Lamb y Parks, 2007).

### **1.2.3 Ciclo replicativo**

#### **1.2.3.1 Unión y fusión virus-célula**

La fusión de la envoltura viral con la membrana plasmática de la célula huésped ocurre a pH neutro y permite el ingreso del complejo RNP a la célula. La actividad de fusión se realiza mediante la acción concertada de las glicoproteínas de membrana H y F (Stern et al., 1995, Von Mesling et al., 2001, Zipperle et al., 2010). El proceso de fusión comienza por la unión de la hemaglutinina al receptor CD150/SLAM. La proteína H induce cambios conformacionales sobre la proteína F que favorecen su actividad fusogénica desencadenando la fusión de la membrana celular y la envoltura viral (Sawatsky & von Messling, 2010) (Figura 12, 13). Dicho proceso ocurre por la acción específica del péptido de fusión localizado en la subunidad F1 (Von Messling et al., 2004).

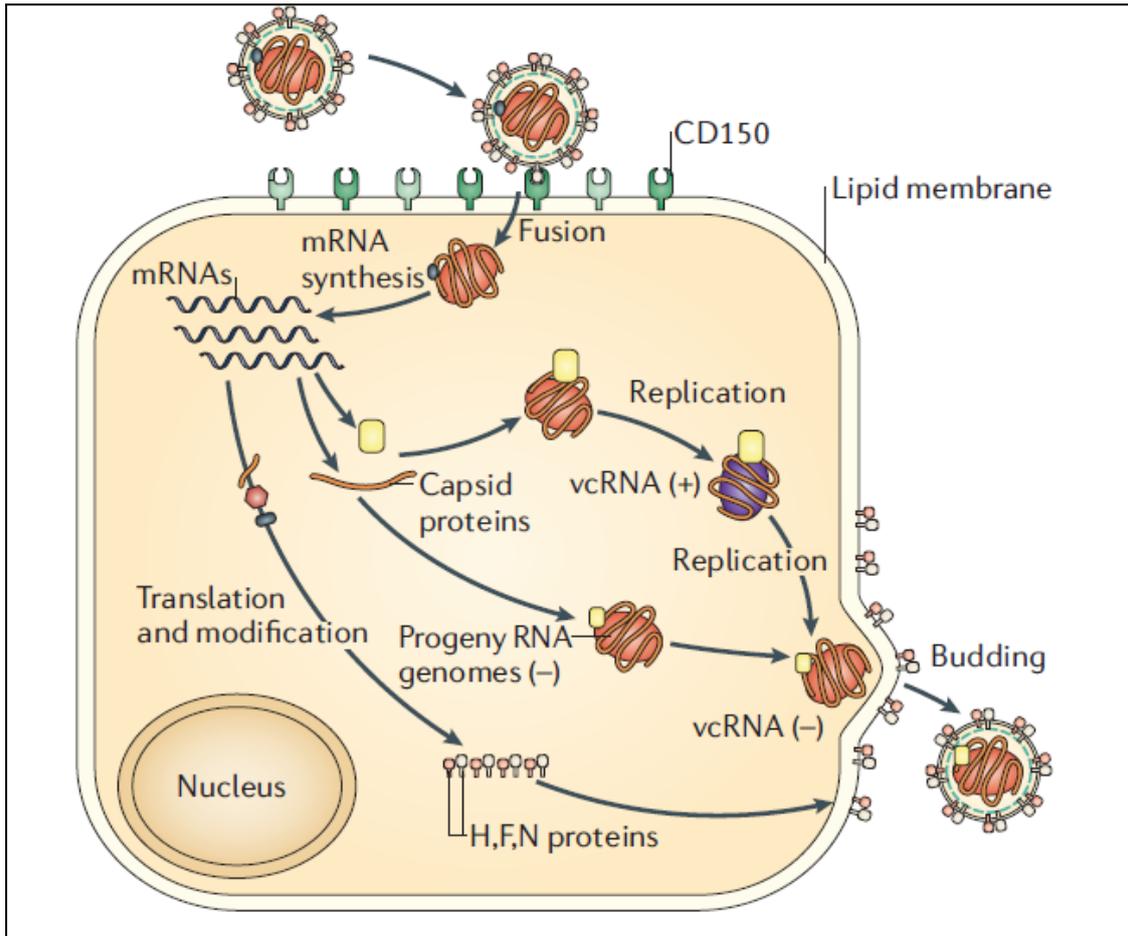


Figura 13. Representación del ciclo de vida de un *Morbillivirus* en la célula del hospedero. Tomado de Moss & Griffin (2006).

### 1.2.3.2 Transcripción y replicación del genoma viral

Tras el ingreso del complejo RNP al citoplasma, ocurre la transcripción y traducción del genoma desde el extremo 3'. Luego de la transcripción, la RNA polimerasa sintetiza en forma continua los antigenomas que serán utilizados como molde para la producción de nuevos RNA genómicos. Los genomas virales neo-sintetizados son encapsidados y transportados hacia la membrana celular para formar nuevas partículas virales (Figura 13) (Lamb & Parks, 2007).

### 1.2.3.3 Formación del complejo ribonucleoproteico y brotación

El RNA genómico encapsidado por las proteínas N, P y L forma el complejo RNP. Dicho complejo se asocia con la proteína M localizada en la membrana celular, donde yacen las glicoproteínas H y F previamente exportadas. Las partículas virales neo-

sintetizadas son liberadas mediante brotación de la membrana y poseen la capacidad de infectar nuevas células susceptibles (Figura 13) (Lamb & Parks, 2007).

#### **1.2.3.4 Propagación entre células del hospedero**

El virus se propaga entre las células del hospedero mediante la formación de células multinucleadas (sincitios). Los sincitios son el resultado de la fusión de células infectadas que expresan las glicoproteínas virales en su superficie con células susceptibles que presentan el receptor CD150/SLAM (Stern et al., 1995). La proteína H es el determinante principal de la fusogenicidad debido a los cambios conformacionales que induce en la proteína F, regulando la eficiencia de la fusión célula-célula (Von Messling et al., 2001).

### **1.3 Distemper: patogénesis, sintomatología clínica y epidemiología**

#### **1.3.1 Patogénesis**

El modo principal de transmisión del virus es a través de aerosoles de secreciones respiratorias. En la fase inicial de la infección, CDV infecta macrófagos y células dendríticas de las vías respiratorias mediante el receptor CD150/SLAM, una inmunoglobulina con cuatro motivos que se une a la hemaglutinina a través de sus ectodominios IgV e IgD2 (Zipperle et al., 2010). Dichas células atraviesan el epitelio respiratorio alcanzando órganos linfáticos donde el virus se replica vigorosamente. Posteriormente, el virus reingresa a las vías respiratorias mediante el receptor nectina-4 que actúa como receptor de salida. La nectina-4 es una inmunoglobulina de unión célula-célula que interactúa con la hemaglutinina a través de su dominio distal (Muhlebach et al., 2011).

La replicación del virus en los ganglios linfáticos ocurre dentro de las primeras 24 horas y posteriormente se disemina hacia las amígdalas y los nódulos linfáticos bronquiales. La replicación en estos órganos ocurre entre 2-4 días post-infección (PI). Posteriormente, el virus prolifera en los órganos linfoides entre los 4-6 días PI, y se disemina vía sanguínea al aparato respiratorio, digestivo y urinario, y al sistema nervioso central (SNC) entre los 8-9 días PI (Appel & Summers, 1999).

La patogénesis entre los días 9 y 14 PI depende de la respuesta inmune humoral y celular del hospedero. Animales con títulos adecuados de anticuerpos y de citotoxicidad

celular pueden eliminar al virus de la mayoría de los tejidos sin que se evidencien síntomas clínicos, mientras que animales con una respuesta inmune inadecuada sufren la diseminación del virus a diversos tejidos. En términos generales, la diseminación viral puede determinar infección por un período de 60-90 días (Figura 14) (Deem et al., 2000).

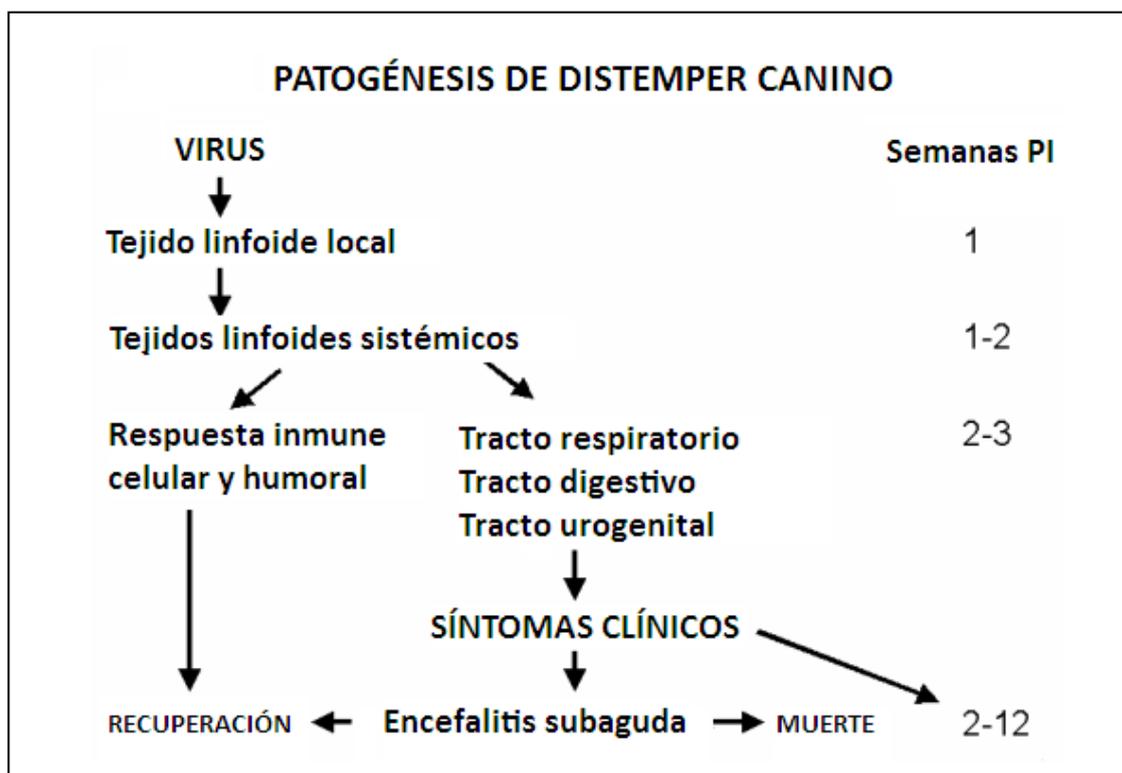


Figura 14. Esquema de la patogénesis de Distemper. PI: post infección. Modificado de Appel & Summers (1999).

### 1.3.2 Sintomatología Clínica

La sintomatología asociada a la enfermedad depende de la virulencia de la cepa viral, y de la edad, estado inmune y sanitario del animal. En especies susceptibles, los sistemas respiratorios, gastrointestinal y el SNC son los más comprometidos. Hasta el 70% de las infecciones en canes domésticos son subclínicas, manifestándose la enfermedad de forma leve con ocurrencia de languidez, anorexia, fiebre e infección del tracto respiratorio superior. Sin embargo, la forma aguda de la enfermedad presenta una elevada mortalidad con ocurrencia de síntomas clínicos asociados al sistema respiratorio y gastrointestinal, incluyendo conjuntivitis, descargas óculo-nasales, neumonía, diarrea (muchas veces hemorrágica) y deshidratación severa (Deem et al., 2000). Los animales

infectados eliminan al virus en todas las excreciones corporales independientemente de los síntomas clínicos que presenten (Frolich et al., 2000).

La primera viremia ocurre entre los días 4-6 PI y determina la infección de los tejidos linfáticos con aparición de fiebre y el comienzo de la linfopenia. La segunda viremia se acompaña de pirexia y ocurre la infección de todos los tejidos epiteliales. Durante esta etapa comienza la fase sintomática, y se observan manifestaciones secundarias como hipoplasia dental e hiperqueratosis plantar (Figura 15). Dependiendo de la cepa viral, puede ocurrir encefalomiелitis aguda en asociación o inmediatamente después de la manifestación sistémica (von Messling et al., 2003). Dentro de las numerosas manifestaciones neurológicas asociadas a la infección por CDV (rigidez cervical, convulsiones, síntomas vestibulares y cerebrales y ataxia sensorial), la mioclonia es el único síntoma neurológico sugestivo de Distemper (Deem et al., 2000). La inmunosupresión severa y duradera asociada a la infección, potencia la susceptibilidad a infecciones secundarias que contribuyen a las elevadas tasas de mortalidad de la enfermedad (von Messling et al., 2005).



Figura 15. Manifestaciones clínicas secundarias características de infección por CDV. A – hipoplasia dental. B – descarga óculo-nasal. C – hiperqueratosis plantar. Tomado del sitio web [www.adoptgl.com](http://www.adoptgl.com).

### 1.3.3 Epidemiología

CDV representa una importante amenaza para carnívoros domésticos y silvestres en todo el mundo. En cánidos domésticos Distemper es una de las enfermedades virales de mayor incidencia con elevadas tasas de morbilidad y mortalidad (Appel & Summers, 1999). En las últimas décadas, se han registrado infecciones en todas las familias de carnívoros terrestres (Deem et al., 2000, Frolich et al., 2000, Origgi et al., 2012) y en diversos carnívoros marinos (de Swart et al., 1995, Stanton et al., 2004, Butina et al., 2010). En ciertas regiones geográficas los carnívoros silvestres representan el reservorio del virus favoreciendo su transmisión a canes domésticos. Por ejemplo, se han

registrado epidemias regulares de Distemper en mapaches (*Procyon lotor*) de Norteamérica y en civetas de cola en Japón (*Paguma larvata*), los cuales han tenido un rol fundamental en la transmisión del virus a canes y a otras especies susceptibles (Hashimoto et al., 2001, Lednicky et al., 2004) (Figura 16). Alternativamente, en regiones como África donde no se implementan planes de vacunación adecuados, los canes son el reservorio del virus siendo capaces de transmitirlo a especies silvestres (Deem et al., 2000). Recientemente se han reportado brotes de CDV en colonias de macacos en Asia, destacando el potencial zoonótico del virus (Bieringer et al., 2013, Sakai et al., 2013). En la actualidad, la protección cruzada contra CDV en humanos es provista por la vacunación o infecciones naturales con el virus del sarampión, relacionado con CDV dentro del género *Morbillivirus*. Sin embargo, las campañas exitosas de erradicación del sarampión pueden determinar que las poblaciones no estén eficientemente protegidas, proveyendo un nicho para que los *Morbillivirus* animales puedan adaptarse al uso de receptores celulares humanos (Ludlow et al., 2014).

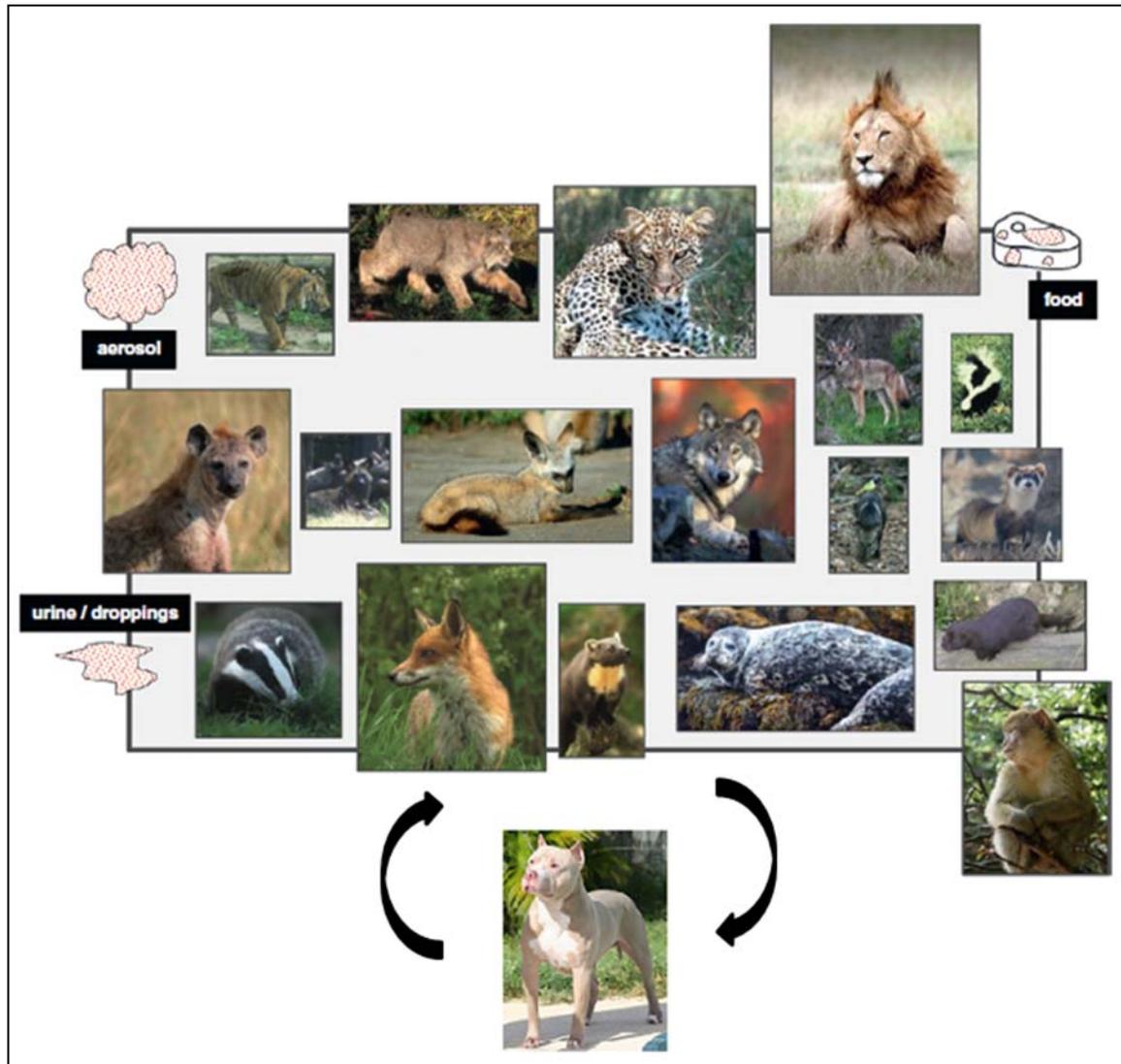


Figura 16. Transmisión interespecífica de CDV entre huéspedes silvestres y canes domésticos a través de diversas vías (aerosoles, orina, alimentación). Modificada de Ludlow et al., 2014.

#### 1.4 Antecedentes del equipo de trabajo

Nuestra línea de investigación en *Paramyxoviridae* comenzó en el año 2006 con el virus distemper canino (CDV), y tiene como principal objetivo el relevamiento y análisis de la evolución genómica del virus. En el contexto de mi trabajo de grado, implementamos por primera vez en Uruguay una metodología de diagnóstico molecular mediante RT-PCR, y un método de diferenciación basado en RFLP (Polimorfismo en el Largo de los Fragmentos de Restricción) que nos permitió discriminar entre cepas de campo y vacunales (Sarute et al., 2011). Durante mis estudios de Maestría relevamos la variabilidad genética de las cepas circulantes en el Río de la Plata estableciendo la

presencia de dos linajes divergentes en Sudamérica con distinta distribución, prevalencia y origen (Panzera et al., 2012).

En el marco de mi tesis doctoral nos planteamos analizar la diversidad poblacional de CDV mediante una colaboración con el Dr. Marco Vignuzzi del Institut Pasteur París, una autoridad mundial en genética poblacional y evolución de virus RNA.

De este modo, en el proyecto de tesis original nos propusimos los siguientes objetivos específicos:

1. Analizar la diversidad genética del virus en distintas matrices (orina, secreción óculo-nasal, sangre, cerebelo) obtenidas de dos casos clínicos.
2. Establecer la frecuencia de mutación en cada matriz.
3. Estimar los efectos sobre la diversidad poblacional mediante un modelo celular *in vitro*.
4. Realizar un análisis comparativo (filogeografía *intra-host*) entre las variantes detectadas en cada matriz.

Dichos objetivos se llevarían a cabo analizando la variabilidad de un fragmento de 1033 pb del gen H mediante dos abordajes experimentales: amplificación, clonación y secuenciación de Sanger, lo cual sería realizado en la Sección Genética Evolutiva, y análisis del fragmento por secuenciación masiva (NGS) a realizarse en el laboratorio del Dr. Vignuzzi utilizando el equipo Illumina Genome Analyzer IIX. Durante una primera pasantía en el laboratorio del Dr. Vignuzzi, se analizó el fragmento de 1033 pb del gen H a partir de muestras de orina y sangre de un caso clínico. El análisis de las secuencias nucleotídicas de 144 clones (72 clones por muestra) reveló una elevada variabilidad genética en las variantes de ambos fluidos. Estos resultados preliminares impulsaron que nuestro proyecto inicial de tesis fuera ampliado y mejorado sustancialmente, ya que decidimos analizar la dinámica poblacional de CDV mediante la amplificación del genoma completo (15700 pb) y posterior secuenciación masiva a partir de siete matrices de tres casos clínicos. Estos cambios en nuestro plan de trabajo fueron posibles gracias a la excelente disposición de los médicos veterinarios quienes nos proveyeron de los casos clínicos y realizaron las necropsias de los tres animales, y al laboratorio del Dr. Vignuzzi que financió las corridas de secuenciación masiva. De este modo, partiendo de las mismas hipótesis de trabajo nos propusimos objetivos más ambiciosos para estudiar con mayor profundidad aspectos poblacionales y evolutivos de CDV.

## **1.5 Hipótesis de trabajo**

- El virus Distemper canino está organizado en un sistema de causiespecies.
- Existe compartimentalización (subpoblaciones) del virus dentro de los tejidos del huésped.

## **1.6. Objetivos**

### **1.6.1 Objetivo General**

Analizar la diversidad poblacional y la compartimentalización del virus distemper canino.

### **1.6.2. Objetivos Específicos**

- Establecer la diversidad genética del virus en diversos tejidos/fluidos (matrices) mediante la amplificación del genoma completo y secuenciación masiva (NGS).
- Analizar la compartimentalización tejido-específica del virus.
- Establecer el flujo genético y la proporción de variantes migrantes entre las distintas matrices.

## **2. MATERIALES y MÉTODOS**

En el presente trabajo se analizaron tres casos clínicos correspondientes a canes domésticos. Los tres casos fueron diagnosticados como positivos para CDV en base a su sintomatología clínica, y por metodologías moleculares estandarizadas en nuestro laboratorio, las cuales implican la amplificación de un fragmento del gen N de 287 pb (Sarute et al., 2011). Los tres animales incluidos en el estudio perecieron y se realizó la necropsia de cada caso obteniéndose muestras de diversos tejidos/fluidos (matrices). Posteriormente, se procedió a diseñar nueve juegos de cebadores capaces de amplificar los 15,7 kb del genoma completo de CDV. Las condiciones de amplificación con cada par de cebadores fueron estandarizadas en una única matriz (pulmón/251). Los amplicones solapantes fueron secuenciados mediante el método de Sanger y se ensambló la secuencia genómica.

Paralelamente, se implementó la propagación de la línea celular Vero-DogSLAMTag (DST) utilizada para realizar la titulación por dosis infectiva al 50% en cultivo celular (TCID<sub>50</sub>), e infecciones a partir de homogeneizados (inóculos) de siete matrices de los tres canes y de una dilución de una vacuna comercial utilizada como control. A partir de los viriones recuperados de cada infección se realizó extracción de RNA y posterior RT-PCR para la detección del genoma viral y la posterior amplificación de los nueve fragmentos solapantes del genoma completo. Los amplicones obtenidos fueron fragmentados para la construcción de librerías genómicas y se secuenciaron mediante un protocolo de secuenciación masiva (NGS). Las secuencias nucleotídicas fueron analizadas mediante programas informáticos desarrollados para estudiar la diversidad poblacional en virus RNA, y fueron sometidas a estudios filodinámicos para realizar análisis de compartimentalización tejido-específica y de flujo génico entre las matrices de cada caso clínico.

### **2.1 Muestras**

Los tres canes incluidos en el estudio (245, 246 y 251) presentaron distintas manifestaciones clínicas de la enfermedad (cuadros respiratorios, gastrointestinales y/o nerviosos) (Tabla 1). Cuando los animales perecieron se realizaron las necropsias por parte del Dr. Alejandro Benech y la Dra. Lourdes Francia en el Hospital de la Facultad de Veterinaria, gracias a una colaboración que mantenemos con ambos profesionales veterinarios. Para cada uno de los casos clínicos se obtuvieron muestras de cerebelo,

pulmón, médula espinal, estómago, intestino grueso, hígado y vejiga/orina. El material obtenido de las necropsias fue almacenado a -80 °C hasta su procesamiento.

Tabla 1. Casos clínicos analizados en el estudio. Se detalla denominación, sintomatología clínica, raza y sexo, edad, año de colecta y estado de vacunación de los individuos. ND: no determinado.

Individuo	Sintomatología clínica	Raza/Sexo	Edad	Año de colecta	Vacunación
245	Pirexia, diarrea sanguinolenta, tics, convulsiones, piodermia abdominal, infección oídos	Cruza/hembra	4 meses	2012	No
246	Síntomas neurológicos agudos	Galgo/macho	6 meses	2012	No
251	Conjuntivitis, descargas óculo-nasales, piodermia abdominal, diarrea, depresión	Dogo/macho	6 meses	2013	ND

## 2.2 Metodologías moleculares

### 2.2.1 Extracción de RNA

La extracción del RNA viral se realizó con el kit Quick-RNA Miniprep (Zymo Research) a partir de 50 miligramos (mg) de tejido pulmonar del caso 251 (pulmón/251) mediante el siguiente protocolo:

- Agregar 800 µl de *buffer* de lisis ZR a la muestra de tejido y homogeneizar mecánicamente
- Transferir 800 µl de cada mezcla a una columna Zymo-Spin IIIC en un tubo colector, y centrifugar a 12.000 g por 1 min. Descartar el filtrado
- Adicionar 400 µl de *buffer* de pre-lavado a la columna y centrifugar a 12.000 g por 1 min. Descartar el filtrado
- Adicionar 700 µl de *buffer* de lavado y centrifugar a 12.000 g por 30 seg. Descartar el filtrado. Repetir el paso anterior utilizando 400 µl de *buffer* de lavado
- Centrifugar la columna Zymo-Spin IIIC a 12.000 g por 2 min para remover completamente el *buffer* de lavado
- Colocar la columna Zymo-Spin IIIC en un tubo de 1,5 ml. Adicionar 50 µl de H<sub>2</sub>O libre de DNA/RNA nucleasas directo a la matriz de la columna e incubar por 1 min
- Centrifugar a máxima velocidad por 30 segundos (seg) para eluir el RNA. Utilizar inmediatamente o almacenar a -80 °C

### 2.2.2 Retrotranscripción

La reacción de retrotranscripción se realizó a partir del RNA viral con el kit RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) utilizando cebadores hexaméricos universales (*random primers*) de acuerdo al siguiente protocolo:

- Colocar en un tubo 5-10  $\mu\text{l}$  de RNA y 1  $\mu\text{l}$  de *random primers* (0,2  $\mu\text{g}/\mu\text{l}$ )
- Incubar la mezcla a 70 °C por 5 min, llevar a hielo por 2 min
- Adicionar 4  $\mu\text{l}$  de *Buffer RT*, 2  $\mu\text{l}$  de dNTPs 10mM, 1  $\mu\text{l}$  de Ribolock RNAase Inhibitor y 1  $\mu\text{l}$  de RevertAid Reverse Transcriptase.
- Incubar a 42 °C por 60 min y a 70 °C por 10 min.

### 2.2.3 Reacción de PCR

#### 2.2.3.1 Detección del genoma viral

El cDNA obtenido fue utilizado para amplificar el fragmento de 287 pb del gen N con la enzima Taq DNA polymerase (Thermo Scientific). Los reactivos utilizados y las condiciones de ciclado se detallan en la Tabla 2.

Tabla 2. Reactivos y condiciones de ciclado empleados en las reacción de PCR para la detección del genoma viral

Componente	Reacción 10 $\mu\text{l}$			
H <sub>2</sub> O	6,1 $\mu\text{l}$			
10X Reaction Buffer	1 $\mu\text{l}$			
10mM dNTP's	0,2 $\mu\text{l}$			
Cebador directo	0,3 $\mu\text{l}$ (0.3 $\mu\text{M}$ )			
Cebador reverso	0,3 $\mu\text{l}$ (0.3 $\mu\text{M}$ )			
cDNA	2 $\mu\text{l}$			
Taq Polymerase	0,1 $\mu\text{l}$			

Desn. Inicial	94 °C	3 min	30X
Desnaturalización	94 °C	30 seg	
Hibridación	50 °C	30 seg	
Extensión	72 °C	45 seg	
Extensión final	72 °C	7 min	

#### 2.2.3.2 Amplificación del genoma completo

La amplificación del genoma completo (15,7 kb) se realizó en nueve fragmentos solapantes utilizando los nueve juegos de cebadores diseñados en este estudio (Tabla 3). Los cebadores se diseñaron en base a la cepa 5804 (AY386315) debido a su elevada identidad con las cepas uruguayas descritas a la fecha (Panzera et al., 2012). La temperatura de *melting*, formación de estructuras secundarias (*hairpins* y homodímeros), así como la formación de heterodímeros se calcularon mediante la herramienta OligoAnalyzer 3.1 del sitio web [www.idtdna.com](http://www.idtdna.com).

Tabla 3. Cebadores diseñados para la amplificación del genoma completo discriminados por gen/fragmento genómico. Se detalla el nombre, secuencia nucleotídica y posición en el genoma en base a la cepa de referencia 5804 (AY386315).

Cebadores	Secuencia nucleotídica (5' - 3')	Posición en el genoma
<b>gen N</b>		
FT	ACC AGA CAA AGT TGG CTA AGG	1-22
R1907	CTG ATG CTG CTG ACC TCT TGA	1886-1907
<b>gen P</b>		
F1765	ATC AAT CAT TCA TCC GAC CAC C	1765-1787
R3541	ACT CTG ACC TGA GGT ACT AAC C	3519-3541
<b>gen M</b>		
F2935	GAC ACT GGA GAC CCT ACG GCA	2935-2956
R4957	CAG GAG CGA CAT CGA GAC AGT TG	4936-4957
<b>gen F</b>		
F4854	TCC AGG ACA TAG CAA GCC AAC A	4854-4876
R7102	CAC CTT GTC TTG GTA GGA GAG C	7080-7102
<b>gen H</b>		
F7061	CTT AGG GCT CAG GTA GTC CAA C	7061-7083
R8969	GTC GGT AAG GGA TTT CTC ACC AC	8946-8969
<b>gen L</b>		
<b>fragmento L1</b>		
F8822	GAG GCT GAC ATC ACC AAC TCT	8822-8843
R10138	ATT CTC TGC TGC GGT TAT TGC T	10116-10138
<b>fragmento L2</b>		
F9985	ATG GCT TCT ATA CTG AAG AGA C	9985-10007
R12155	TGG AAC AAT CCT TTC AGC ATA G	12133-12155
<b>fragmento L3</b>		
F12015	AGT GAC CCT TAT TCT GCC AAT C	12015-12037
R14006	GTC ATT CCA TGT CAG ACC AAC A	13984-14006
<b>fragmento L4</b>		
F13842	GAC AAC ATA CAA GCT AGG CAC C	13842-13864
R15690	ACC AGA CAA AGC TGG GTA TGA T	15668-15690

Las reacciones de PCR para amplificar los nueve fragmentos genómicos fueron realizadas con la enzima Phusion High-Fidelity DNA Polymerase (Thermo Scientific). Los reactivos utilizados y las condiciones de ciclado se detallan en la Tabla 4.

Tabla 4. Reactivos y condiciones de ciclado empleados en las reacciones de PCR para la amplificación del genoma completo.

Componente	Reacción 25 $\mu$ l
H <sub>2</sub> O	13,25 $\mu$ l
5X Phusion Buffer	5 $\mu$ l
10mM dNTP's	0,5 $\mu$ l
Cebador directo	1,5 $\mu$ l (0,6 $\mu$ M)
Cebador reverso	1,5 $\mu$ l (0,6 $\mu$ M)
cDNA	3 $\mu$ l
Phusion DNA Polymerase	0,25 $\mu$ l

Desn. Inicial	98 °C	2 min	
Desnaturalización	98 °C	15 seg	<b>35X</b>
Hibridación	58-62 °C	30 seg	
Extensión	72 °C	2 min	
Extensión final	72 °C	10 min	

#### 2.2.4 Electroforesis y purificación de DNA

Los productos de PCR fueron separados mediante electroforesis en gel de agarosa al 1% teñidos con bromuro de etidio (0,5  $\mu$ g/ $\mu$ l) durante 30 min a 90 volts constantes. La corrida se realizó con *buffer* TAE 1X (40 mM Tris base (pH 7.6), 20 mM ácido acético glacial, 1mM EDTA). Los geles fueron visualizados mediante transiluminación con luz ultravioleta.

Los amplicones correspondientes al genoma completo de pulmón/251 fueron escindidos del gel y se purificaron con el kit Illustra GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare) según el siguiente protocolo:

- Adicionar 10  $\mu$ l de *buffer* de captura por 10 mg de gel. Mezclar por inversión
- Cargar la mezcla en una columna GFX MicroSpin column. Centrifugar a 13.000 rpm por 30 seg. Descartar el filtrado
- Adicionar 500  $\mu$ l de *buffer* de lavado a la columna. Centrifugar a 13.000 rpm por 30 seg. Descartar el filtrado
- Secar la columna centrifugando a 13.000 rpm por 30 seg adicionales
- Transferir la columna a un tubo de 1.5 ml, agregar 10 -50  $\mu$ l de *buffer* de elución e incubar a temperatura ambiente (TA) por 1 min
- Recuperar el DNA purificado centrifugando a 13.000 rpm por 1 min
- Almacenar el DNA a -20 °C

### **2.2.5 Secuenciación por Sanger**

El DNA purificado de los nueve fragmentos genómicos se secuenció en ambas direcciones con los cebadores específicos de cada fragmento genómico (Tabla 3) y cebadores diseñados previamente en nuestro laboratorio en la empresa Macrogen (Corea del Sur). Las secuencias obtenidas fueron editadas y ensambladas con el programa SeqMan (Lasergene).

## **2.3 Aislamiento viral en cultivo celular**

### **2.3.1 Propagación de la línea celular Vero-DST**

El trabajo asociado a cultivo celular se realizó en el laboratorio de Virología Molecular del Centro de Investigaciones Nucleares (CIN), el cual se encuentra equipado con una cabina de flujo laminar de bioseguridad tipo II.

Las células Vero-DST fueron cedidas por la Dra. Branka Horvat del International Center for Infectiology Research-CIRI (Lyon, Francia). La línea Vero-DST es la mejor línea disponible para la propagación de CDV, debido a que las células están transfectadas con un plásmido de expresión que codifica la proteína de membrana SLAM/CD150, receptor celular de CDV (Seki et al., 2003). La propagación inicial se realizó a partir de una ampolla almacenada en nitrógeno líquido, la cual fue rápidamente incubada a 37° C en un baño termostático. Los pasajes de la línea se realizaron utilizando medio de cultivo completo compuesto por Dulbecco's Modified Eagle Medium (DMEM) (GE Healthcare), suero fetal bovino (SFB) (GE Healthcare) al 10% y Estreptomicina/Penicilina (S/P) (Sigma) al 1%. Además, cada tres pasajes se adicionó el antibiótico G418 (Sigma) que mantiene seleccionado al plásmido para la expresión del receptor SLAM. Las células se cultivaron en botellas de 25 y 75 cm<sup>2</sup> (T25, T75) en estufa de CO<sub>2</sub> al 5% y 37° C.

### **2.3.2 Preparación de inóculos**

Se tomaron 250 mg/μl de cerebelo, estómago, intestino, hígado, médula espinal, pulmón y vejiga/orina de los tres casos clínicos, y se homogeneizaron en medio de cultivo DMEM y S/P al 10% utilizando pistones plásticos. El homogeneizado fue centrifugado a 1000g por 10 minutos (min) a 4°C, y el sobrenadante fue filtrado con filtros de 0,45 μm (Sartorius) para ser utilizado como inóculo en los ensayos de titulación e infección.

### 2.3.3 Titulación de inóculos por dosis infectiva al 50% (TCID<sub>50</sub>)

A partir de 150 µl de los inóculos de cada matriz de los tres casos clínicos, se realizaron diluciones seriadas (-1 a -8) con medio DMEM libre de SFB. Las diluciones se utilizaron para infectar monocapas confluentes de células Vero-DST en placas de 96 pocillos (10.000 células por pocillo). Las placas se almacenaron en estufa de CO<sub>2</sub> al 5% y 37° C, y se revelaron una semana post-infección (PI) utilizando formalina al 4% y solución de cristal violeta 0,025%. El título de cada inóculo se determinó como el número de partículas virales por mililitro (ml).

### 2.3.4 Infección de la línea celular Vero-DST

Se sembraron placas de seis pocillos (2x10<sup>5</sup> células por pocillo) y se incubaron durante 24 hs en estufa de CO<sub>2</sub> al 5% y 37° C para obtener una confluencia del 70-80%. La cepa vacunal Onderstepoort (OP) fue utilizada como control y para familiarizarnos con el efecto citopático (CPE) del virus en cultivo. Las infecciones se realizaron con 200 µl de una dilución (10<sup>-2</sup>) de la cepa vacunal y con una multiplicidad de infección (MOI) = 1 para los inóculos de los tres casos clínicos, haciendo un total de 22 infecciones. Luego de sembrados los inóculos, se incubaron las placas por 1 hora a TA y posteriormente se incubaron en estufa de CO<sub>2</sub> al 5% y 37° C. En cada placa se incluyeron células sin infectar utilizadas como control negativo (MOCK). Las placas se observaron 48-72 hs PI en un microscopio invertido para determinar la presencia de CPE, se documentaron las imágenes de cada placa (servicio de microscopía del Instituto Pasteur de Montevideo) y se recuperaron los viriones del sobrenadante y los asociados a la monocapa celular, los cuales se almacenaron a -80° C.

### 2.3.5 Extracción del genoma viral

El aislamiento de RNA se realizó a partir del sobrenadante y la monocapa celular de todas las infecciones en paralelo con el MOCK, mediante el kit Quick-RNA Miniprep (Zymo Research) de acuerdo al protocolo descrito en la sección 2.2.1.

Sin embargo, el primer paso del protocolo fue modificado para la extracción a partir del sobrenadante y de la monocapa celular. En el caso de sobrenadante se agregaron 600 µl de *buffer* de lisis ZR por 200 µl de muestra, mientras que para la monocapa se adicionaron 800 µl de *buffer* ZR por pocillo (10<sup>5</sup> células) y se removieron las células de la superficie mediante el uso de rastrillos plásticos (*cell scrapers*).

### 2.3.6 Ensayos de RT-PCR, electroforesis y purificación de amplicones

Para confirmar la presencia del genoma viral en el sobrenadante de cultivo o en la monocapa celular tras las infecciones, se realizó RT-PCR del fragmento de 287 pb del gen N (Tabla 2), utilizando el RNA aislado del MOCK para cada infección como control negativo. Una vez confirmada la presencia del genoma viral, se amplificó el genoma completo utilizando los nueve juegos de cebadores con las condiciones de ciclado estandarizadas para cada amplicón (Tabla 3, 4).

Los productos de PCR fueron separados mediante electroforesis y los amplicones del tamaño esperado fueron escindidos del gel y purificados de acuerdo al protocolo de la sección 2.2.4.

## 2.4 Secuenciación masiva (NGS)

Las actividades experimentales de secuenciación masiva fueron realizadas en el laboratorio del Dr. Vignuzzi, ya que se cuenta con un protocolo estandarizado para la técnica para los diversos modelos virales que se estudian en su laboratorio.

### 2.4.1 Cuantificación del DNA purificado

La cuantificación del DNA se realizó utilizando la molécula fluorescente PicoGreen (Life Technologies) en un fluorómetro MFX (Dynex Technologies) según el siguiente protocolo:

- En una placa de 96 pocillos agregar 50 µl de *buffer* TE-3 (1mM EDTA, 10mM Tris-HCl (pH 7.5)) en un pocillo (blanco de reacción) y en los pocillos restantes mezclar 48 µl de *buffer* TE-3 y 2 µl de DNA
- Agregar 50 µl de PicoGreen previamente diluido (1:200 en *buffer* TE-3) en cada pocillo
- Incubar la placa a TA por 5 min protegida de la luz
- Ingresar la placa al fluorómetro y realizar la medida de concentración

### 2.4.2 Fragmentación y purificación de DNA por *beads*

La reacción de fragmentación del DNA se realizó en un volumen final de 20 µl, incluyendo 10 µl de DNA (1 ng/µl), 2 µl de 10X *buffer* de reacción y 1µl de la fragmentasa NEBnext dsDNA (New England Biolabs) que genera secuencias de 70-80

pb. La mezcla se incubó por 50 min a 37° C y la reacción se detuvo adicionando 5 µl de 0,5M EDTA.

El DNA fragmentado de cada amplicón fue purificado mediante el kit Agencourt AMPure XP (Beckman-Coulter) el cual utiliza cuentas magnéticas (*beads*), según el siguiente protocolo:

- Mezclar por *vortex* la solución de los *beads* y agregar 65 µl en cada pocillo de una placa de 96
- Mezclar el DNA fragmentado de todos los amplicones de una muestra en un mismo pocillo
- Incubar la placa a TA por 15 min
- Colocar la placa en agitador magnético a TA por 15 minutos o hasta que el líquido se torne incoloro
- Descartar el sobrenadante de cada pocillo y agregar 200 µl de etanol 80%
- Incubar la placa por 1 min y realizar el lavado con etanol 80% en agitador magnético por 5 min
- Remover la placa del agitador y almacenarla a TA por 15 min
- Resuspender el *pellet* en 62.5 µl de *buffer* de resuspensión e incubar la placa a TA por 2 min
- Colocar nuevamente la placa en el agitador magnético a TA por 5 min
- Almacenar 60 µl de sobrenadante de cada pocillo correspondiente al DNA purificado

#### **2.4.3 Construcción de librerías**

La construcción de librerías se realizó con el kit NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs) el cual adiciona secuencias adaptadoras en los extremos de los fragmentos de DNA. Los adaptadores hibridan con cebadores para la generación de *clusters* de amplificación (Figura 17).

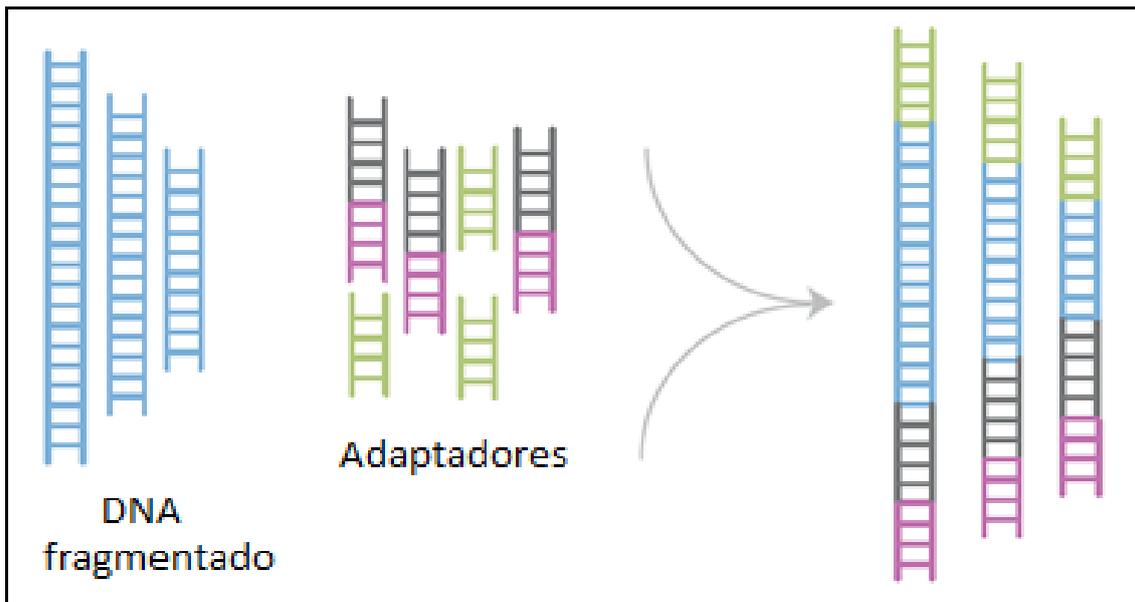


Figura 17. Representación de la construcción de librerías para secuenciación masiva. Se detalla la ligación de los adaptadores a los fragmentos de DNA.

#### 2.4.4 Secuenciación por síntesis

El DNA fue empleado para generar *clusters* clonales de cada fragmento mediante PCR *bridge* utilizando el equipo cBot (Illumina). Los fragmentos de DNA fueron hibridados a un panel de cebadores universales inmovilizados en la superficie de la celda mediante sus adaptadores. El DNA fue sintetizado desde la zona de unión de los cebadores mediante extensión del extremo 3' utilizando una enzima de alta fidelidad. La polimerasa copia el DNA hibridado formando puentes de DNA doble hebra (dsDNA), que son desnaturalizados en forma de dos cadenas simples. Estas hebras hibridan nuevamente con los cebadores y son extendidas para formar *loops* de dsDNA. Este proceso se repite para cada molécula mediante numerosos ciclos de desnaturalización y amplificación, para crear millones de *clusters* clonales, los cuales contienen aproximadamente 2000 copias de la molécula molde. Posteriormente, un cebador de secuenciación se hibrida a su secuencia complementaria en los extremos libres de la secuencia molde de los *clusters* para comenzar el proceso de secuenciación (Figura 18).

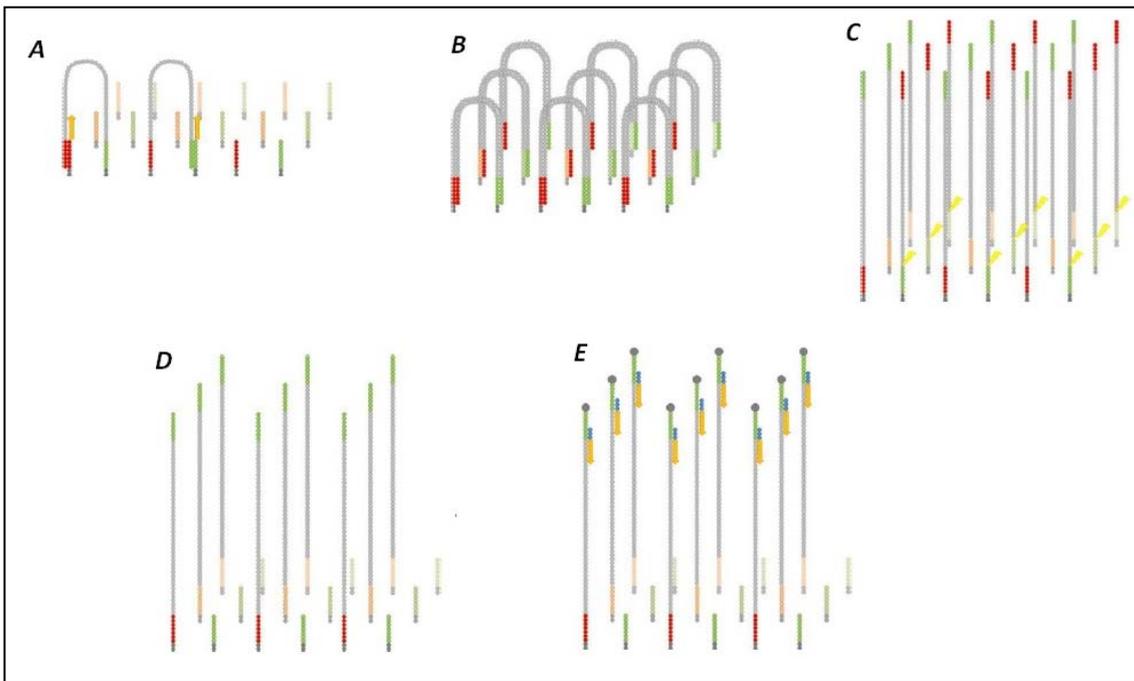


Figura 18. Representación esquemática del PCR *bridge* y las primeras etapas de secuenciación. A. Amplificación por *bridge* para generación de *clusters*. B. Desnaturalización de las moléculas doble hebra de DNA. C. Clivaje y remoción de los fragmentos reversos. D. Unión covalente de las hebras directas del *cluster* a los cebadores de secuenciación. E. Comienzo de la secuenciación por síntesis (SBS). Modificado del sitio web <http://www.slideshare.net/cosentia/illumina-gaiix-for-high-throughput-sequencing>.

El equipo Illumina GAIIX utiliza la química de secuenciación por síntesis (SBS) (TruSeq Technology). La química SBS permite la secuenciación masiva y paralela de millones de fragmentos utilizando un método de terminadores fluorescentes reversibles que detectan a cada base incorporada en la cadena nascente de DNA. Los terminadores son removidos luego de la detección de cada base permitiendo que el proceso de secuenciación se extienda cíclicamente hasta 100 veces (Figura 19).

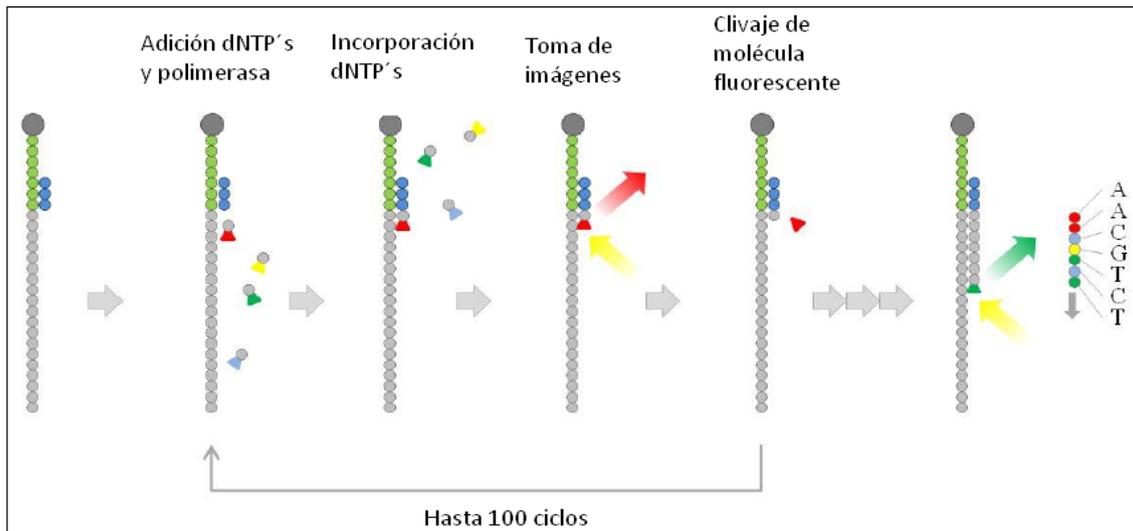


Figura 19. Representación esquemática de la secuenciación por síntesis (SBS) en el equipo Illumina GAIIx. Modificado del sitio web <http://www.slideshare.net/cosentia/illumina-gaiix-for-high-throughput-sequencing>.

## 2.5 Análisis bioinformático

### 2.5.1 Análisis de secuencias nucleotídicas

Los datos de secuenciación primarios (.bcl) se convirtieron en archivos fastq con el programa CASAVA 1.8.2 de la plataforma Illumina. Mediante la herramienta fastq-mcf (ea-utils) se realizó un control de calidad inicial para eliminar las bases de baja calidad en los extremos de cada *read*. Las secuencias con una probabilidad de error mayor a 0.001 (valores *phred*) y con menos de 16 bases de longitud fueron eliminadas para evitar errores en el alineamiento de los *reads* y en la asignación de nucleótidos. Los *reads* que superaron este control fueron alineados y comparados con la secuencia genómica de referencia 5804 (AY385316) mediante la herramienta Burrows Wheeler Aligner (Li & Durbin, 2009).

### 2.5.2 Implementación del programa Viral Variance Analysis (ViVan)

El programa ViVan fue desarrollado por el laboratorio del Dr. Vignuzzi para el análisis de datos de secuenciación masiva en estudios de poblaciones virales, facilitando la identificación, caracterización y comparación de variantes genéticas dentro de las poblaciones. El programa brinda la frecuencia de cada nucleótido para cada posición genómica, identifica variantes en una frecuencia significativa en la población, y las traduce en cambios aminoacídicos. Además, aporta información adicional sobre la

diversidad poblacional, como cambios sinónimos y no sinónimos, y la tasa de transición/transversión para cada muestra analizada (Isakov et al., 2014).

El ViVan analiza las secuencias alineadas en base a la secuencia de referencia mediante *pileup*, el cual estima que los nucleótidos para cada posición genómica estén correctamente representados en los *reads*. Solamente aquellos *reads* que estén alineados en una misma región genómica (alineamientos únicos) serán incorporados en el *pileup*, el cual se realiza mediante la herramienta SAMtools (Li et al., 2009) incorporada al ViVan.

Una vez realizado el *pileup*, el programa analiza cada posición genómica con una cobertura mayor a 100X y presenta los datos de secuenciación en formato .csv (valores separados por comas). El Anexo I detalla la información obtenida para cada matriz organizada en tres hojas de cálculo realizadas en Apache OpenOffice ([www.openoffice.org/es](http://www.openoffice.org/es)). La hoja 1, secuencias, detalla la posición genómica en base al genoma de referencia (AY385316) (columna A-B), la profundidad de cobertura (columna C), el número y tipo de nucleótidos secuenciados (columnas D-G) y la frecuencia relativa de cada nucleótido para cada posición genómica (columnas H-L) y discriminadas por gen (*feature*) (columna L). La hoja 2, cambios aa, detalla los cambios aminoacídicos a nivel poblacional para cada posición genómica en función del nucleótido representado en mayor frecuencia. Se detalla el gen correspondiente y la posición genómica (columnas A-B), la cobertura (columna C), el nucleótido en la secuencia de referencia y el nucleótido alternativo (columnas D-E), la frecuencia del nucleótido alternativo (columna F), la posición dentro del gen y la secuencia aminoacídica (columnas G-H), y el cambio aminoacídico generado (columna I).

En la hoja 3, métrica, se especifican otros parámetros de la población viral. En la parte superior se detalla la tasa de transición y transversión de cada nucleótido para la secuencia genómica, las tasas de transición y transversión totales, y el cociente transición/transversión para la secuencia genómica de cada matriz. En la parte inferior se detallan los nucleótidos secuenciados para cada gen y el total para cada genoma (columnas B-C) y los cambios sinónimos y no sinónimos (columnas D-E).

### **2.5.3 Estimación de frecuencias de mutación y reconstrucción de secuencias consenso**

Con la información obtenida del ViVan se estimó la frecuencia de mutación para cada gen/fragmento genómico de cada matriz analizada de los tres casos clínicos. Además, se

realizó la reconstrucción de la secuencia genómica consenso con las posiciones secuenciadas para cada matriz en base a las frecuencias nucleotídicas de cada posición. Aquellos nucleótidos representados en mayor frecuencia se utilizaron para definir a cada posición dentro de la secuencia consenso (Anexo I, hoja secuencias, columna M).

#### 2.5.4 Análisis filodinámico

En los análisis de compartimentalización y flujo génico (migración) de cada caso clínico se utilizaron las secuencias genómicas consenso, y todas las variantes genéticas que se diferenciaban del consenso en una posición para cada matriz. En nuestro estudio, definimos a las variantes mayoritarias como aquellas presentes en una frecuencia  $>1\%$ , y a las variantes minoritarias como las representadas en una frecuencia entre  $0,1-1\%$  dentro de la población. Para la construcción de los conjuntos de datos se estimó en primer lugar, la señal filogenética de cada gen/fragmento genómico con el programa TreePuzzle (Schmidt & von Haeseler, 2007), implementando un análisis de cuartetos de secuencias por *likelihood mapping* utilizando un *dataset* de 45 secuencias genómicas obtenidas de la base de datos del Genbank. El modelo de sustitución nucleotídica más apropiado se determinó a través de los estadísticos AIC (*Akaike Information Criterion*) y BIC (*Bayesian Information Criterion*), implementados en el programa jModelTest (Darriba et al., 2012).

La reconstrucción filogenética de las variantes obtenidas para cada caso clínico y su migración entre compartimientos se co-estimaron utilizando la aproximación bayesiana del programa BEAST versión 1.7.5 (Drummond et al., 2012), implementando Cadenas de Markov de Tiempo Continuo (CTMCs) para la reconstrucción de caracteres con estados discretos, los cuales fueron asignados dependiendo del origen de la muestra: pulmón (P), médula espinal (M), cerebelo (C), orina (O) o vejiga (V).

Para corroborar la hipótesis de compartimentalización, el conjunto de árboles posteriores obtenidos en el análisis anterior fue utilizado como datos de entrada para el programa BaTS (Parker et al., 2008). Este programa implementa los algoritmos *parsimony score* (PS) de Slatkin & Maddison (1989), *association index* (AI) de Wang et al. (2001) y *maximum monophyletic clade* (MC) de Parker et al. (2007) para determinar si existe una asociación estadísticamente significativa entre la posición de cada secuencia en la filogenia y su estado (cada una de las matrices).

### **3. RESULTADOS**

#### **3.1 Metodologías moleculares**

##### **3.1.1 Estandarización de la amplificación del genoma completo y secuenciación por Sanger**

En la muestra de pulmón del caso 251 (251P) se confirmó la presencia del genoma viral mediante la amplificación del fragmento de 287 pb del gen N (dato no mostrado).

La amplificación del genoma completo (15,7 kb) se realizó en nueve fragmentos solapantes utilizando los juegos de cebadores diseñados en este estudio (Tabla 3, 4).

Los nueve amplicones del tamaño esperado fueron obtenidos, purificados y secuenciados bidireccionalmente por el método de Sanger. Las secuencias fueron editadas y ensambladas, obteniéndose el genoma de la muestra 251P. El genoma codificante (15447 pb) de los seis genes virales (3' N-P-M-F-H-L 5') fue enviado a la base de datos del GenBank (nº acceso KM280689) y se redactó un manuscrito en base a su análisis publicado en la revista *Genome Announcements* (Sarute et al., 2014) (Anexo II).

#### **3.2 Aislamiento viral en cultivo celular**

##### **3.2.1 Infección de la línea Vero-DST con la cepa vacunal Onderstepoort**

La placa de infección de la cepa vacunal Onderstepoort fue observada diariamente, y a las 48 hs post-infección (PI) se detectó la presencia de sincitios multinucleados característicos del efecto citopático (CPE) del virus (Karzon & Bussell, 1959). En los pocillos con células sin infectar (MOCK) no se registró CPE aparente (Figura 20).

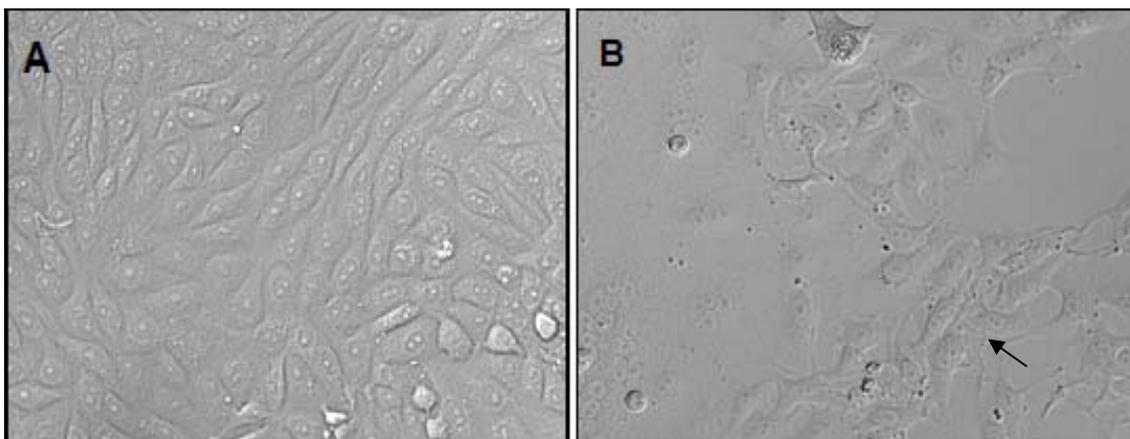


Figura 20. Imágenes (20X) de la infección de la línea Vero-DST. A. MOCK, B. Inóculo cepa vacunal Onderstepoort. En la imagen B se señalan con una flecha numerosos sincitios multinucleados.

### 3.2.2 Titulación de inóculos por dosis infectiva al 50% (TCID<sub>50</sub>)

Los 21 homogeneizados (inóculos) de cerebelo (C), estómago (E), intestino (I), hígado (H), médula espinal (M), pulmón (P) y vejiga (V)/orina (O) de los tres casos clínicos (245, 246 y 251) se titularon mediante TCID<sub>50</sub>. Se obtuvieron títulos para tres matrices, mientras que para las 18 restantes no se registraron títulos mensurables con las diluciones del inóculo inicial de 150 µl (Tabla 5). El título más elevado correspondió a la matriz 246P (1,61x10<sup>4</sup> partículas virales por mililitro) (Tabla 5, Figura 21).

Tabla 5. Títulos virales obtenidos por TCID<sub>50</sub> (partículas virales/ml) para las siete matrices analizadas de los tres casos clínicos. ND: no determinado.

Caso clínico	Matriz	Título
245	Cerebelo	ND
	Estómago	ND
	Intestino	ND
	Hígado	ND
	Médula	ND
	Pulmón	1,68E+02
	Vejiga	ND
246	Cerebelo	ND
	Estómago	ND
	Intestino	ND
	Hígado	ND
	Médula	ND
	Pulmón	1,61E+04
	Orina	ND
251	Cerebelo	1,78E+03
	Estómago	ND
	Intestino	ND
	Hígado	ND
	Médula	ND
	Pulmón	ND
	Orina	ND

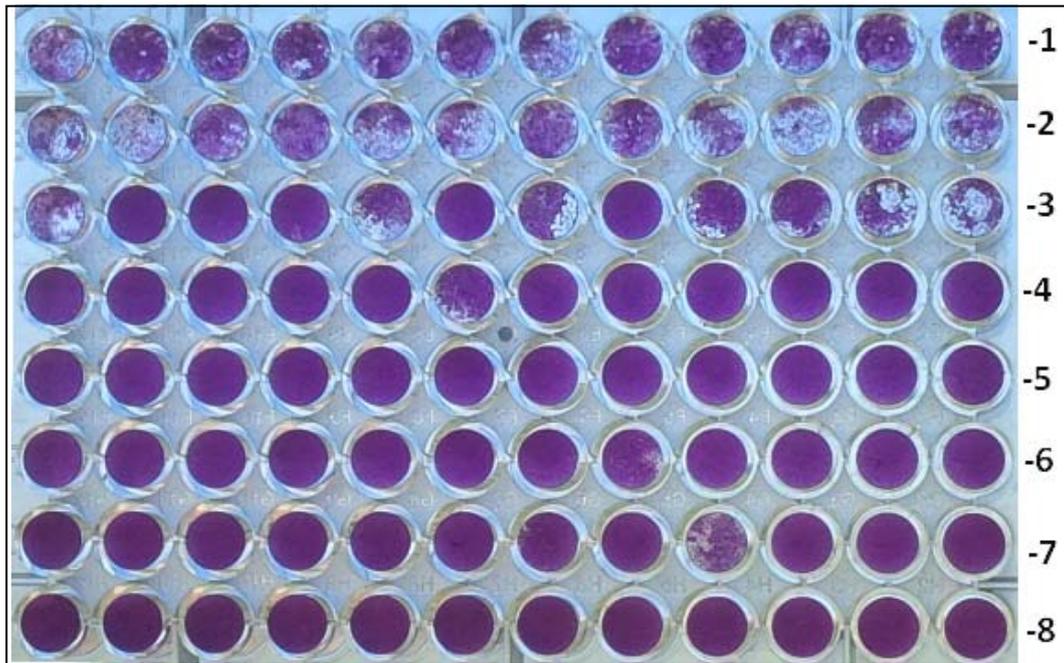


Figura 21. Placa de titulación por TCID<sub>50</sub> utilizando diluciones de -1 a -8 del inóculo de 246P.

### 3.2.3 Infección de la línea Vero-DST con inóculos de los casos clínicos y detección del genoma viral

Las infecciones de la línea Vero-DST se realizaron con un MOI=1 (n° partículas virales/célula) para las tres matrices donde se obtuvo título viral, mientras que en las 18 matrices restantes se utilizaron directamente 200 µl de inóculo. Todas las placas fueron observadas a las 48-72 hs PI, y en las infecciones provenientes de cerebelo y pulmón de los tres casos clínicos y en las de médula de dos casos (246 y 251) se identificaron sincitios multinucleados (Figura 22). En estos casos se recuperó el sobrenadante de cada pocillo y se almacenó la placa de infección. En la matriz 245M, en las matrices estómago, intestino, hígado y vejiga/orina de los tres casos clínicos y en el MOCK de todas las infecciones no se registró CPE aparente a los 6-7 días PI. Para las trece matrices donde no se observó CPE se realizaron tres infecciones independientes a partir de nuevos inóculos, y a pesar de no detectar CPE, se recogió el sobrenadante de cultivo y se conservó la placa de cada infección a las 48-72 hs PI.

Se realizó extracción de RNA viral a partir del sobrenadante para todas las matrices, incluso para aquellas donde no se obtuvieron títulos mensurables por TCID<sub>50</sub> ni se observó CPE. El RNA viral se utilizó en primera instancia para realizar RT-PCR del fragmento de 287 pb del gen N para confirmar la presencia del genoma viral (Tabla 2).

En las matrices cerebro, pulmón, vejiga/orina y médula de los tres casos clínicos se obtuvo el amplicón del tamaño esperado (Figura 23). En el MOCK de las infecciones y en las matrices estómago, hígado e intestino no se detectó la presencia de genoma viral, por lo cual estas matrices no se emplearon en los análisis posteriores.

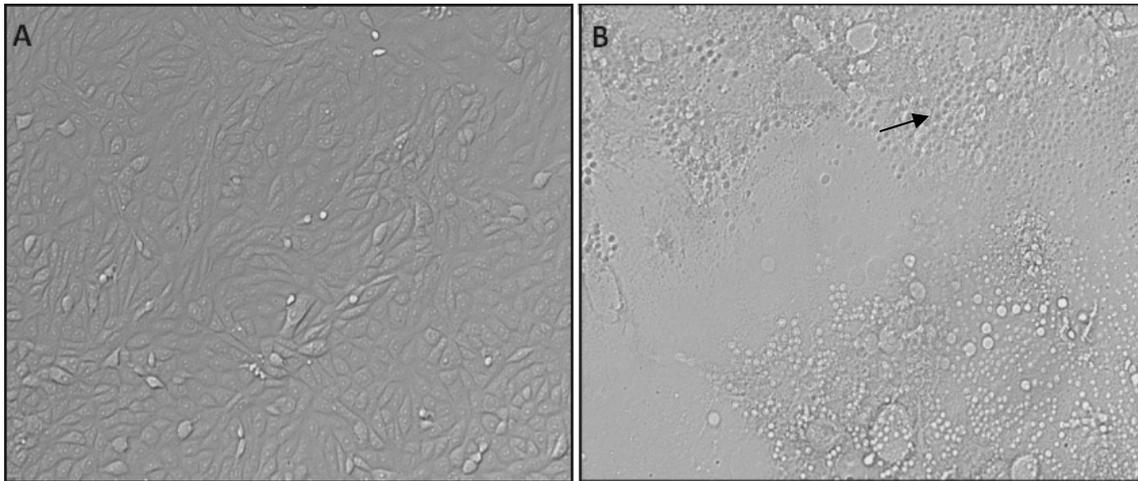


Figura 22. Imágenes (10X) de la infección de la línea Vero-DST. A. MOCK, B. Inóculo 246P. La flecha señala un sincitio multinucleado.

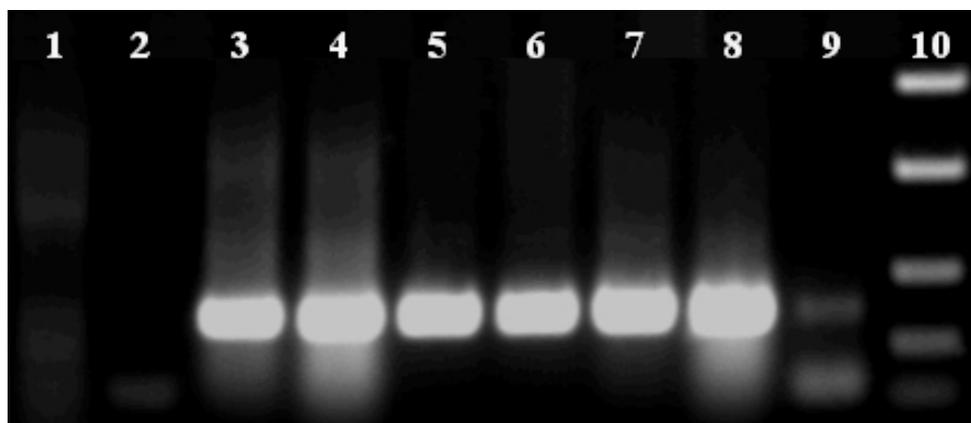


Figura 23. Amplificación por RT-PCR de un fragmento de 287 pb del gen N. 1. MOCK, 2. Blanco de PCR, 3. 245C, 4. 245P, 5. 246C, 6. 246M 7. 251C, 8. 251P, 9. 246O, 10. Marcador de peso molecular *FastRuler Low Range DNA Ladder* (Thermo Scientific) (1500 pb, 850 pb, 400 pb, 200 pb, 50 pb)

### 3.2.5 Amplificación del genoma completo

La obtención de los genomas completos para las doce matrices donde se detectó el genoma viral implicaba la amplificación de los nueve fragmentos solapantes, haciendo un total de 108 amplicones. Se obtuvieron por PCR los nueve fragmentos del tamaño esperado en ocho matrices: cerebro (245C, 246C y 251C), y pulmón (245P, 246P y

251P) de los tres casos clínicos, y médula de dos casos (246M y 251M) (Tabla 6, Figura 24). En orina de dos casos clínicos (246O y 251O), se amplificaron siete fragmentos genómicos, mientras que en la matriz 245V se obtuvieron seis fragmentos totalizando 92 amplicones, un 85% del total inicial (Tabla 6). En la matriz 245M no se obtuvo ninguno de los nueve amplicones, aunque la concentración de reactivos y las condiciones de ciclado de la PCR fueron modificadas (Tabla 6).

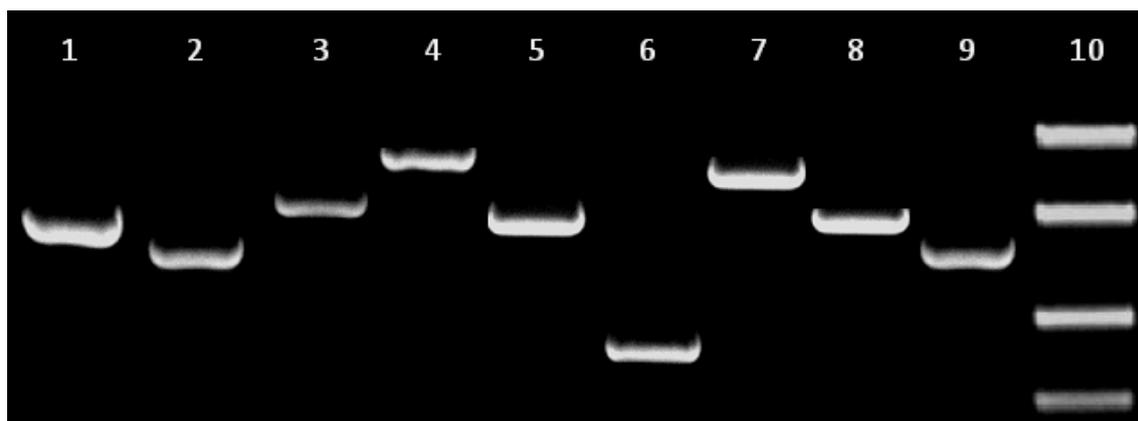


Figura 24. Amplificación por RT-PCR de los nueve fragmentos solapantes del genoma viral de la muestra 246P. 1. Gen N, 2. Gen P, 3. Gen M, 4. Gen F, 5. Gen H, 6. Fragmento L1, 7. Fragmento L2, 8. Fragmento L3, 9. Fragmento L4, 10. Marcador de peso molecular *GeneRuler 1 kb DNA Ladder* (Thermo Scientific) (2500 pb, 2000 pb, 1500 pb, 1000 pb).

### 3.3 Secuenciación masiva (NGS)

#### 3.3.1 Construcción de librerías y secuenciación por síntesis

Los 92 amplicones obtenidos fueron purificados y cuantificados mediante PicoGreen mostrando valores superiores a 1ng/μl (Tabla 6). Todos los amplicones fueron fragmentados en secuencias de ~70 pb para la construcción de las librerías genómicas para cada matriz. El DNA fragmentado fue inmovilizado en la celda de secuenciación para la generación de *clusters* clonales, amplificación por *bridge* PCR y secuenciación por síntesis. Se realizaron once librerías genómicas con los 92 amplicones provenientes de las diferentes matrices de cada caso clínico, las cuales se incluyeron en una primera corrida de secuenciación en el equipo GAIIIX (Illumina).

Tabla 6. Representación de los resultados de la reacción de PCR y concentración del DNA (ng/ $\mu$ l) de los nueve fragmentos genómicos utilizados para la construcción de librerías para la secuenciación masiva de las once matrices. Se detalla entre paréntesis el tamaño de cada fragmento genómico en pares de bases. Los casos donde se obtuvo el amplicón del tamaño esperado se destacan en amarillo y en rojo los casos donde no fue posible amplificar el gen/fragmento genómico.

	N (1907)	P (1776)	M (2022)	F (2248)	H (1908)	L1 (1316)	L2 (2148)	L3 (1991)	L4 (1848)
<b>245</b>									
Cerebelo	5	7,2	29,3	12,2	13	5,8	3,1	4,8	9,1
Pulmón	8,2	5,4	17	9,1	7	6,1	5,5	7,5	8,2
Médula									
Vejiga	4,4	1,1		4,1		3,7		9,1	4
<b>246</b>									
Cerebelo	3,2	9,8	5,4	10,6	9,6	5,2	2,6	5	7,1
Pulmón	6,1	2	12,5	10,7	6	3,3	4,5	3,6	4,3
Médula	4,3	6,3	12,6	6,6	7,6	4	5,5	4,6	4,9
Orina		6	9,6	4,6	3,1	4,9		5,3	8
<b>251</b>									
Cerebelo	7,2	3,5	14,9	8,8	5,3	10,2	2,3	4,2	5,8
Pulmón	8,8	13,9	9,9	7,7	5,2	3,5	16	4,6	5,2
Médula	7,6	6,2	9,7	12,6	8,4	11,5	13,7	5,4	4,5
Orina	5,4	4,6	6	1,1		5,3		12,8	5,4

### 3.4 Análisis bioinformático

#### 3.4.1 Análisis de secuencias nucleotídicas

Las secuencias (fastq) de los 92 amplicones superaron el control de calidad inicial de la herramienta fastq-mcf. En esta corrida se obtuvieron  $2,12 \times 10^9$  bases en total y una cobertura promedio de 22600 *reads* por base secuenciada, registrándose valores de hasta 152000 *reads* por base. La cobertura más baja para las cuatro matrices correspondió al gen M, al fragmento L2, y en menor medida al gen P, haciendo un total de 28 secuencias con valores de cobertura que no alcanzaron el mínimo establecido (100X). (Anexo I, hoja secuencias). Con el objetivo de descartar que el problema de la secuenciación fuera inherente a los amplicones, se diseñaron nuevos cebadores para la amplificación del gen M y el fragmento L2 (Tabla 7). En estos casos, repetimos el protocolo experimental a partir de nueva extracción de RNA viral del sobrenadante y la monocapa celular, RT-PCR, construcción de librerías y secuenciación.

Tabla 7. Nuevos cebadores diseñados para el gen M y el fragmento L2. Se detalla la secuencia de cada cebador y su posición en el genoma respecto a la cepa de referencia 5804 (AY386315).

Cebadores	Secuencia nucleotídica (5´- 3´)	Posición en el genoma
<b>gen M</b>		
F3441	GTG TAC GAC TTC GAT CAG TCT T	3441-3463
R5018	GGC GGA TTT GAT ACT ATG TTG TG	4995-5018
<b>fragmento L2</b>		
F9387	TAT GCA AAG CTC ACA ATG GTT	9387-9408
R12547	AGA CAT GAG TCT TGT TTG CTT TG	12524-12547

Mediante la utilización de los nuevos cebadores se obtuvieron los amplicones del tamaño esperado del gen M y el fragmento L2 en cerebelo (245C, 246C, 251C), pulmón (245P, 246P, 251P), médula (246M y 251M) y orina (246O y 251O). En el caso del gen P, se modificaron las condiciones de PCR y se obtuvo el amplicón en cerebelo (245C), pulmón (245P, 246P, 251P), médula (246M y 251M) y orina (246O y 251O).

Los 28 amplicones correspondientes a los genes P y M, y al fragmento L2 fueron secuenciados en una segunda corrida utilizando el equipo MiSeq (Illumina), adquirido a principios de este año por el laboratorio del Dr. Vignuzzi. En esta segunda corrida se obtuvo una cobertura promedio de 790 *reads* por base secuenciada. Las regiones con menor cobertura correspondieron nuevamente al gen M y al fragmento L2. Por tanto, los datos de secuenciación de esta corrida no se pudieron utilizar en los análisis posteriores (información no detallada).

Actualmente, seguimos trabajando para obtener una mejor cobertura en estas secuencias y hemos realizado una tercera corrida de secuenciación en el Institut Pasteur Montevideo. Los protocolos experimentales y los avances de los resultados son presentados en el Anexo III.

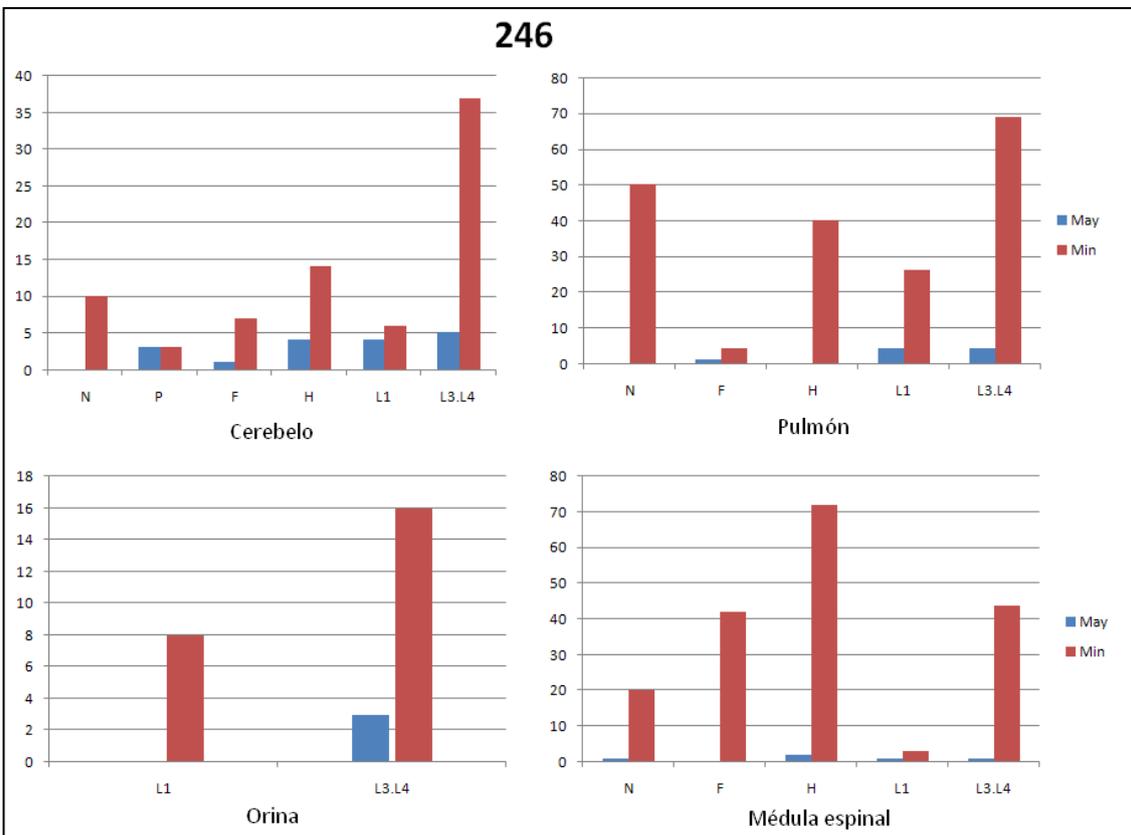
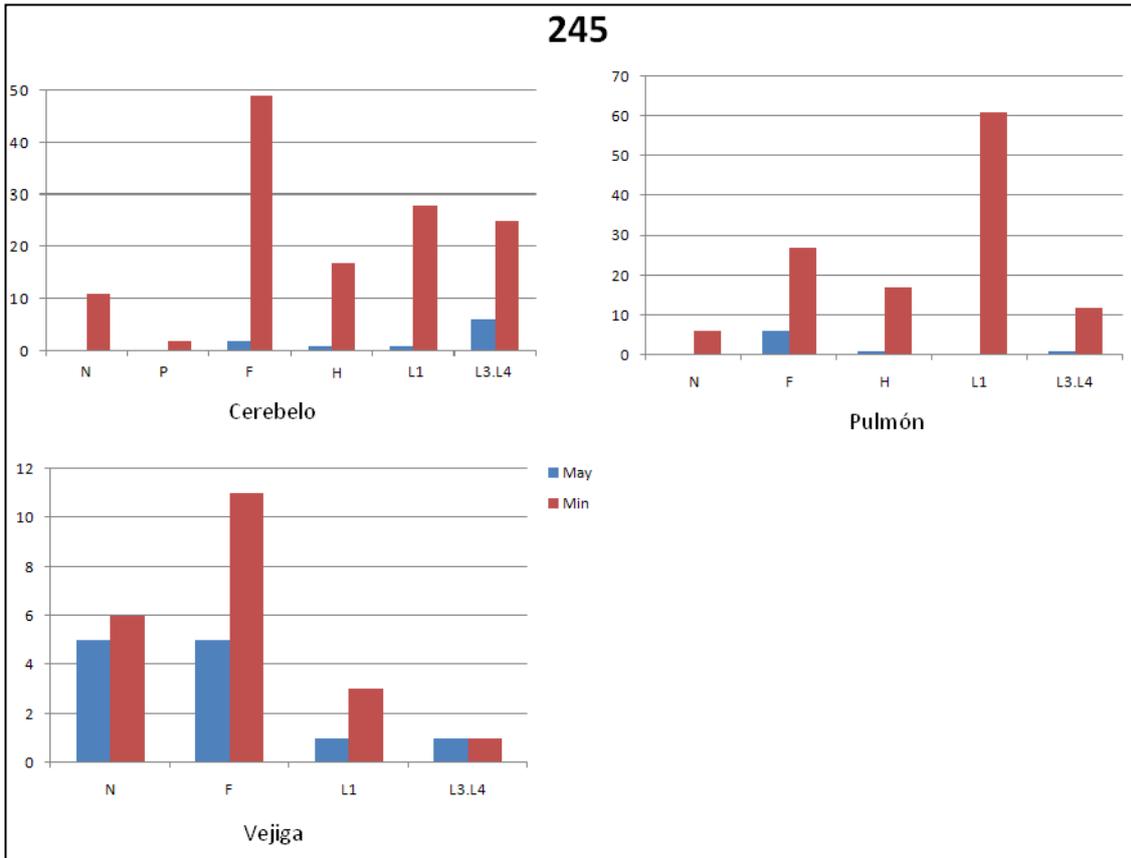
### 3.4.2 Implementación del programa Viral Variance Analysis (ViVan)

Se analizaron mediante el ViVan las 64 secuencias genómicas que superaron el filtro inicial de cobertura (100X), correspondientes a once matrices de los tres casos clínicos. Para una mejor comprensión de los datos de secuenciación, en el Anexo I se detallan los resultados obtenidos mediante el ViVan organizados en tres hojas de cálculo para cada matriz de los tres casos clínicos. La hoja 1 (secuencias) detalla el número de nucleótidos secuenciados y su frecuencia para cada posición genómica, la hoja 2 (cambios aa)

muestra los cambios aminoacídicos dentro de cada gen producto de los nucleótidos representados en mayor frecuencia, y la hoja 3 (métrica) presenta la tasa de transición y transversión, el cociente transición/transversión y los cambios sinónimos y no sinónimos (ver detalle en sección 2.5.2 de Materiales y Métodos).

#### **3.4.2.1 Identificación de variantes mayoritarias y minoritarias**

En las once matrices analizadas, se identificaron variantes presentes en distinta frecuencia dentro de la población para cada gen/fragmento genómico (Anexo I, hoja cambios aa). Estas variantes se clasifican en minoritarias, cuando se encuentran en una frecuencia de  $0.1 < \text{min} < 1$ , y mayoritarias cuando su frecuencia es  $> 1$  a nivel poblacional (Isakov et al., 2014). En la Figura 25 se representa el número total de variantes para todos los genes/fragmentos genómicos de las once matrices, y su análisis en función de la longitud de cada gen/fragmento genómico (Figura 26). Se registraron variantes minoritarias en el rango de 2-92 para las once matrices, el valor más alto correspondió a los fragmentos L3-L4 de 251O y el más bajo al gen P de 245C. Las variantes mayoritarias se ubicaron en el rango de 0-29. Los fragmentos L3-L4 de 251O presentaron el número mayor de estas variantes, mientras que el gen N presentó la menor representación de variantes mayoritarias, e incluso no se registró ninguna para cerebelo (245C, 246C), pulmón (245P, 246P, 251P) y orina (251O) (Figura 25, 26).



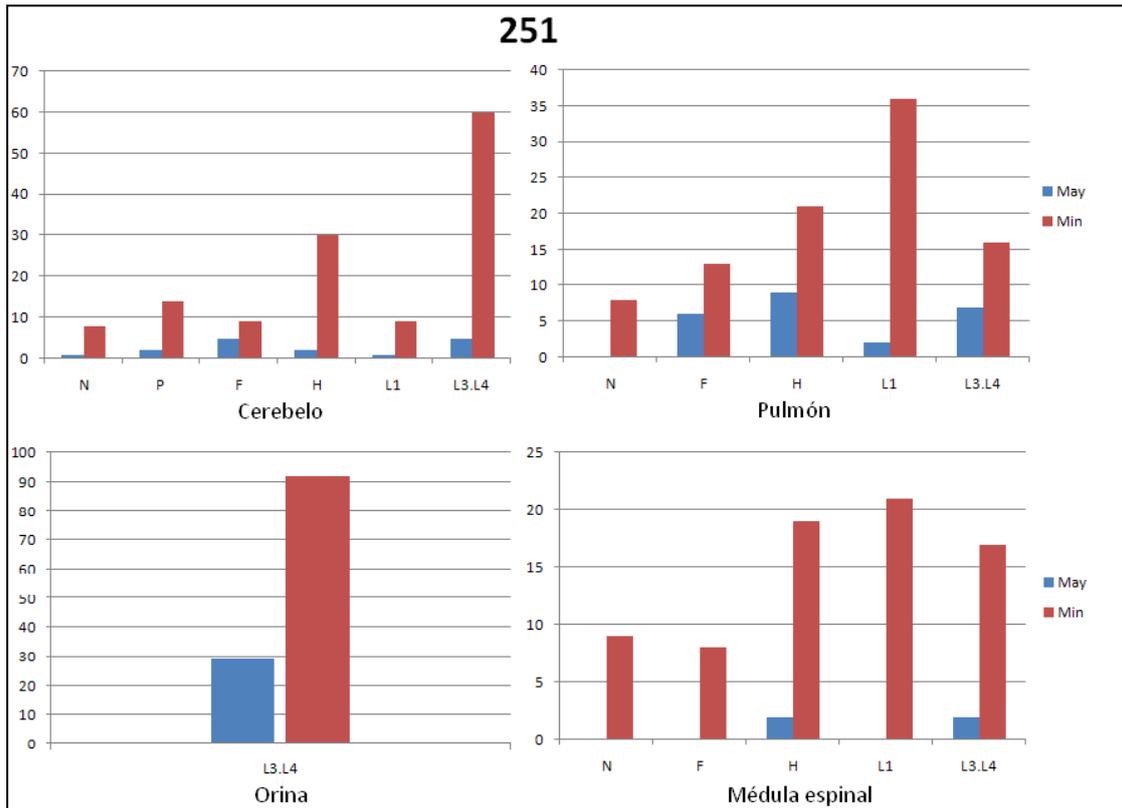


Figura 25. Representación gráfica de número de variantes mayoritarias (may) en color azul, y minoritarias (min) en color rojo discriminadas por gen/fragmento genómico para cada matriz de los tres casos clínicos.

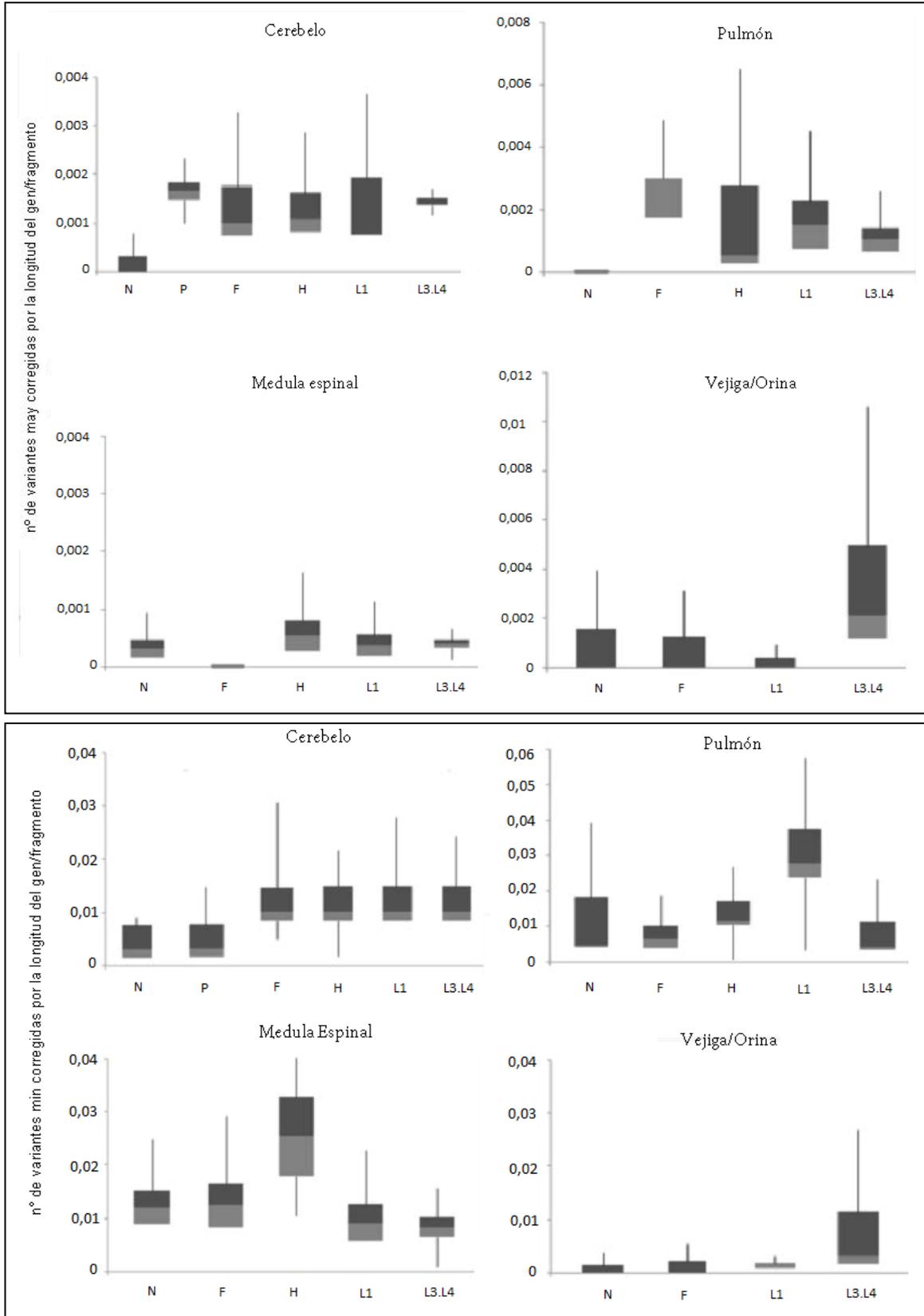


Figura 26. Representación mediante diagramas de caja de proporción de variantes mayoritarias y minoritarias corregidas por la longitud gen/fragmento genómico para cada matriz de los tres casos clínicos.

### 3.4.2.2 Identificación de cambios sinónimos y no sinónimos

Se identificaron los cambios sinónimos y no sinónimos en las regiones codificantes de las secuencias genómicas de las once matrices. En la Tabla 8 se detalla el número total de cambios para cada matriz de los tres casos clínicos, mientras que el análisis detallado de los cambios sinónimos y no sinónimos discriminados por gen para cada matriz se muestra en el Anexo I, hoja métrica.

Los cambios sinónimos se registraron en el rango de 35-343. El valor más alto se registró para la matriz 251O, mientras que el más bajo fue para 246O. Los cambios no sinónimos se registraron en el rango de 35-1260, correspondiendo los valores extremos nuevamente a las matrices 246O y 251O (Tabla 8, Anexo I, hoja métrica).

Tabla 8. Número de cambios sinónimos y no sinónimos para cada matriz. Se detalla el número de nucleótidos secuenciados por genoma.

<b>Matriz</b>	<b>Nucleótidos secuenciados</b>	<b>Sinónimos</b>	<b>No sinónimos</b>
245C	10869	109	166
245P	12953	205	347
245V	7223	41	43
246C	12380	154	233
246P	12666	295	544
246M	10857	165	324
246O	4992	35	26
251C	12676	164	242
251P	13165	164	257
251M	13625	196	383
251O	3675	343	1260

### 3.4.2.3 Identificación de cambios aminoacídicos debidos a variantes mayoritarias

Para las regiones codificantes se identificaron cambios aminoacídicos en todas las matrices de los tres casos clínicos. Estos cambios se asociaron a sustituciones no sinónimas determinadas por variantes mayoritarias (Anexo I, hoja cambios aa). El análisis de cada matriz reveló cambios en la secuencia aminoacídica de los genes N, P, F, H y L. En el gen N y P se registraron cambios solamente en una matriz, 245V y 251C respectivamente. El gen H mostró cambios aminoacídicos para cerebelo (245C, 246C, 251C), pulmón (251P) y médula (246M), a nivel del gen F se observaron cambios en cerebelo (246C, 251C), pulmón (246P, 251P) y en la matriz 245V, mientras que los

cambios en el gen L se observaron en las cuatro matrices: cerebelo (245C, 246C, 251C), pulmón (245P, 246P, 251P), médula (251M), vejiga/orina (245V, 246O, 251O). En términos generales, los genes F, H y L presentaron un número mayor de cambios, mientras que el gen P presentó el número menor de cambios aminoacídicos (Tabla 9).

Algunos de los cambios observados estuvieron presentes en más de una matriz, específicamente en la región codificante de los genes H y F. Las matrices 246M y 251P presentaron el mismo cambio (P>S) en las posiciones 275 y 327 de la hemaglutinina, mientras que las matrices 245C y 251C presentaron el cambio (H>Y) en la posición 549 de esta proteína. Respecto a la proteína de fusión, se registró el mismo cambio (I>V) en la posición 284 en las matrices 245V, 251C y 251P (Tabla 9).

Tabla 9. Cambios aminoacídicos determinados por variantes mayoritarias discriminadas por gen para cada matriz. Se resaltan en color violeta y azul los cambios presentes en más de una matriz correspondientes al gen H y F respectivamente.

Gen	posición	Ref	Alt	frec. Alt	posición cds	posición aa	cambio aa
<i>245C</i>							
H	8723	C	T	0.32	1645	549	H>Y
L	9445	A	G	0.045	416	139	D>G
<i>245P</i>							
L	15079	A	G	0.012	6050	2017	Y>C
<i>245V</i>							
N	181	G	C	0.021	74	25	G>A
N	204	C	A	0.11	97	33	H>N
N	334	T	C	0.024	227	76	I>T
N	982	A	G	0.047	875	292	E>G
F	5617	T	C	0.19	683	228	V>A
F	5784	A	G	0.01	850	284	I>V
F	5974	G	A	0.36	1040	347	S>N
L	9036	T	C	0.023	7	3	S>P
L	14793	A	T	0.059	5764	1922	S>C
<i>246C</i>							
F	6477	A	G	0.012	1543	515	I>V
H	7798	A	C	0.023	720	240	E>D
L	9577	C	T	0.011	548	183	T>I
L	9940	T	C	0.0151	911	304	F>S
<i>246P</i>							
F	6582	C	T	0.018	1648	550	P>S
L	10262	G	T	0.42	1233	411	M>I
L	10566	A	G	0.021	1537	513	M>V
L	11437	A	G	0.035	2408	803	Y>C
L	11623	G	A	0.013	2594	865	R>Q
<i>246O</i>							
L	12675	A	C	0.27	3646	1216	T>P
L	13078	T	A	0.32	4049	1350	L>H
<i>246M</i>							
H	7901	C	T	0.011	823	275	P>S
H	8057	C	T	0.014	979	327	P>S
<i>251C</i>							
P	2291	C	G	0.11	491	164	A>G
P	2597	G	T	0.022	797	266	S>I
F	5193	A	G	0.015	259	87	I>V
F	5784	A	G	0.015	850	284	I>V
H	7809	A	G	0.0138	731	244	D>G
H	8723	C	T	0.17	1645	549	H>Y
L	12464	G	A	0.047	3435	1145	W>*
L	12583	A	C	0.025	3554	1185	H>P
L	13203	A	G	0.058	4174	1392	S>G
L	13302	G	A	0.044	4273	1425	V>I
<i>251P</i>							
F	5581	T	G	0.019	647	216	V>G
F	5784	A	G	0.03	850	284	I>V
H	7476	T	C	0.012	398	133	F>S
H	7901	C	T	0.036	823	275	P>S
H	8057	C	T	0.022	979	327	P>S
H	8537	A	C	0.011	1459	487	S>R
L	9921	T	A	0.017	892	298	F>I
L	9922	T	A	0.017	893	298	F>Y
L	13255	C	T	0.049	4226	1409	A>V
L	13335	C	A	0.12	4306	1436	L>I
L	13678	C	A	0.028	4649	1550	P>H
L	13684	A	G	0.016	4655	1552	H>R
<i>251M</i>							
L	12583	A	C	0.022	3554	1185	H>P
L	13525	T	C	0.024	4496	1499	I>T
<i>251O</i>							
L	13845	A	T	0.04	4816	1606	N>Y
L	13902	T	A	0.017	4873	1625	C>S
L	13938	T	A	0.011	4909	1637	C>S
L	13965	T	G	0.012	4936	1646	S>A
L	13993	T	A	0.016	4964	1655	L>Q

### 3.4.2.4 Tasas de transición y transversión

Se estimó la tasa de transición y transversión para cada nucleótido (A, C, T, G) dentro de la población, para cada posición genómica de las matrices de los tres casos clínicos (Anexo I, hoja métrica). Además, se calculó la tasa de transición y transversión total, y el cociente transición/transversión para cada matriz. Los valores más extremos correspondieron a 246C, donde se registró una tasa de transición de 0,88 y de transversión de 0,12 y un cociente de 7,94 (Tabla 10).

Tabla 10. Tasa de transición y transversión para cada matriz discriminada por nucleótido.

	Tasa de transición	Tasa de tranversión	Matriz
<b>A</b>	0.81	0.19	245C
<b>C</b>	0.89	0.11	
<b>T</b>	0.86	0.14	
<b>G</b>	0.875	0.125	
<b>Total</b>	0.85	0.15	
<b>Transición/Transversión</b>	5.72		
<b>A</b>	0.735	0.265	245P
<b>C</b>	0.8	0.2	
<b>T</b>	0.747	0.253	
<b>G</b>	0.993	0.007	
<b>Total</b>	0.83	0.17	
<b>Transición/Transversión</b>	4.86		
<b>A</b>	0.83	0.17	245V
<b>C</b>	0.83	0.17	
<b>T</b>	0.53	0.47	
<b>G</b>	0.997	0.003	
<b>Total</b>	0.823	0.176	
<b>Transición/Transversión</b>	4.67		
<b>A</b>	0.847	0.153	246C
<b>C</b>	0.91	0.09	
<b>T</b>	0.837	0.163	
<b>G</b>	0.999	0.001	
<b>Total</b>	0.88	0.12	
<b>Transición/Transversión</b>	7.94		
<b>A</b>	0.772	0.228	246P
<b>C</b>	0.996	0.004	
<b>T</b>	0.852	0.148	
<b>G</b>	0.833	0.167	
<b>Total</b>	0.846	0.154	
<b>Transición/Transversión</b>	5.5		
<b>A</b>	0.781	0.219	246M
<b>C</b>	0.89	0.11	
<b>T</b>	0.79	0.21	
<b>G</b>	0.915	0.085	
<b>Total</b>	0.85	0.15	
<b>Transición/Transversión</b>	5.6		
<b>A</b>	0.616	0.384	246O
<b>C</b>	1	0	
<b>T</b>	0.61	0.39	
<b>G</b>	0.999	0.001	
<b>Total</b>	0.766	0.234	
<b>Transición/Transversión</b>	3.27		

<b>A</b>	0.822	0.178	251C
<b>C</b>	0.812	0.188	
<b>T</b>	0.784	0.216	
<b>G</b>	0.988	0.0012	
<b>Total</b>	0.852	0.146	
<b>Transición/Transversión</b>	5.76		
<b>A</b>	0.719	0.281	251P
<b>C</b>	0.815	0.185	
<b>T</b>	0.77	0.23	
<b>G</b>	0.962	0.038	
<b>Total</b>	0.813	0.187	
<b>Transición/Transversión</b>	4.36		
<b>A</b>	0.734	0.266	251M
<b>C</b>	0.74	0.26	
<b>T</b>	0.82	0.18	
<b>G</b>	0.984	0.016	
<b>Total</b>	0.803	0.196	
<b>Transición/Transversión</b>	4.1		
<b>A</b>	0.48	0.52	251O
<b>C</b>	0.88	0.12	
<b>T</b>	0.44	0.56	
<b>G</b>	0.962	0.038	
<b>Total</b>	0.62	0.38	
<b>Transición/Transversión</b>	1.64		

### 3.4.3 Estimación de frecuencias de mutación y reconstrucción de secuencias consenso

#### 3.4.3.1 Frecuencias de mutación

Las 64 secuencias que superaron el control de calidad inicial (cobertura >100X) fueron empleadas para el cálculo de las frecuencias de mutación (Anexo I, hoja secuencias). Los cálculos de frecuencia realizados en 10 secuencias que presentaban una cobertura de 100-2500X por posición, mostraron una sobreestimación de los valores (dato no mostrado). Un ejemplo de ello lo constituyen los genes M y P de la matriz 245C, con valores de cobertura promedio de 100 y 2500X, respectivamente. El valor de frecuencia obtenido para el gen P fue de  $7,5 \times 10^{-4}$ , mientras que para el gen M correspondió  $3,6 \times 10^{-4}$ . Por tanto, se utilizaron 54 secuencias que presentaron una cobertura mayor a 3000X para el cálculo de frecuencias de mutación por gen/matriz/caso clínico.

Mediante este análisis se estableció el número de sitios invariantes, sitios con 1 SNP, sitios con 2 SNPs y sitios con 3 SNPs para todas las posiciones genómicas. Los sitios invariantes se detectaron en el rango de 0-1256, sitios con 1 SNP entre 2-1314, sitios con 2 SNPs entre 30-1478 y sitios con 3 SNPs en el rango de 48-2858 (Tabla 11). En

aquellas posiciones del genoma donde se registró más de un SNP, se contabilizaron los cambios de forma aditiva para estimar las frecuencias, ya que representan a variantes genéticas distintas dentro de la población. La frecuencia de mutación se estimó con la siguiente ecuación: según Beaucourt et al. (2011)  $F = (\text{número de SNPs} / \text{número total de bases secuenciadas}) \times 10.000$

En la Tabla 12 y Figura 27 se detallan las frecuencias de mutación de todos los genes/fragmentos genómicos para las once matrices de los tres casos clínicos.

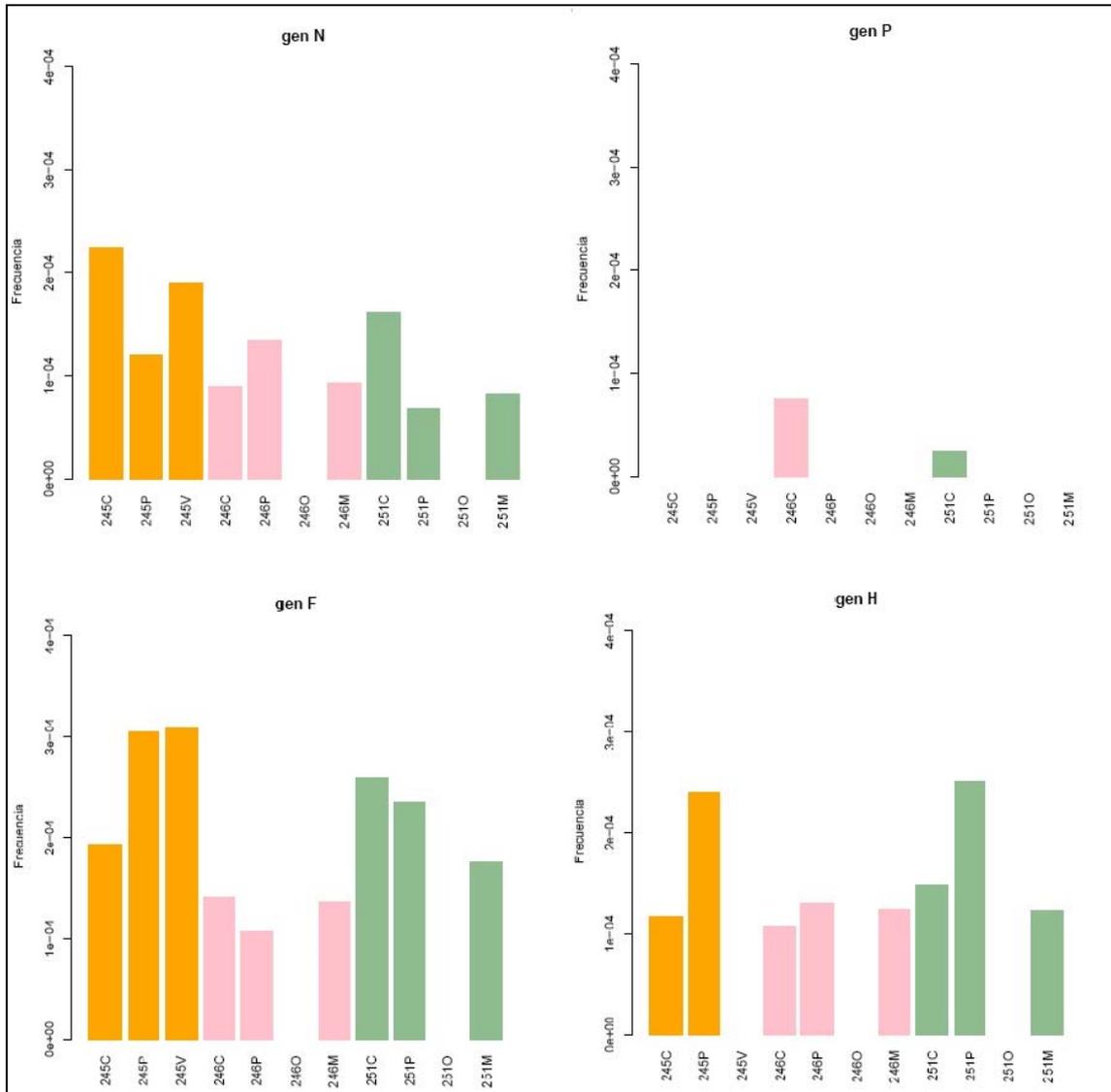
Tabla 11. Número de sitios invariantes y variables para las 54 secuencias con cobertura >3000X. Se detallan las bases secuenciadas y la cobertura promedio para cada gen/fragmento genómico dentro de cada matriz y a nivel global.

	Sitios invariantes	Sitios con 1 SNP	Sitios con 2 SNPs	Sitios con 3 SNPs	Bases secuenciadas	gen/fragmento (pb)	Cobertura promedio
<b>245C</b>							
gen N	98	490	716	268	12149065	1572	7728
gen F	254	553	809	371	17064665	1989	8580
gen H	98	307	661	655	30999276	1824	16995
fragmento L1	0	12	316	778	52937882	1613	32820
fragmentos L3.L4	79	494	250	1841	136515547	3570	38240
<b>245 P</b>							
gen N	0	16	415	1141	35592070	1572	22641
gen F	357	712	725	179	8836303	1989	4443
gen H	200	440	694	385	12436997	1824	6819
fragmento L1	7	64	404	632	22338781	1613	13849
fragmentos L3.L4	6	46	748	2769	133928772	3570	37515
<b>245V</b>							
gen N	41	321	742	468	16810111	1572	10693
gen F	772	809	355	48	5391808	1989	2711
fragmento L1	3	36	341	727	52911817	1613	32803
fragmentos L3.L4	34	194	768	747	68351325	3570	18599
<b>246C</b>							
gen N	1	31	424	1116	47453954	1572	30187
gen P	1256	166	30	72	5854590	1524	3842
gen F	173	266	827	721	28749304	1989	14454
gen H	56	131	563	1074	41823824	1824	22930
fragmento L1	21	183	550	352	15125563	1613	9377
fragmentos L3.L4	510	868	955	1237	56273201	3570	15763
<b>246P</b>							
gen N	6	142	704	720	27677668	1572	17607
gen F	3	159	427	1387	48317942	1989	24293
gen H	13	142	520	1148	35776529	1824	19614
fragmento L1	0	226	646	979	24592483	1613	15246
fragmento L2	0	226	646	979	24592483	1881	13074
fragmentos L3.L4	3	43	666	2858	110435224	3570	30934
<b>246M</b>							
gen N	2	71	500	999	43841168	1572	27889
gen F	188	404	829	567	27397696	1989	13775
gen H	526	733	359	154	15355778	1824	8419
fragmento L1	128	25	378	702	44055216	1613	27313
fragmentos L3.L4	10	409	1345	1760	104537808	3570	29282

<b>246O</b>							
fragmento L1	0	12	182	913	55340876	1613	34309
fragmentos L3.L4	91	1034	1478	838	35070706	3570	9824
<b>251C</b>							
gen N	12	175	664	721	22567348	1572	14356
gen P	1256	166	30	72	5854590	1524	3842
gen F	143	544	805	349	12266662	1989	6167
gen H	118	461	742	490	22868714	1824	12538
fragmento L1	19	7	176	924	59386319	1613	36817
fragmentos L3.L4	6	510	1055	1905	116368417	3570	32596
<b>251P</b>							
gen N	0	6	249	1317	65048712	1572	41380
gen F	223	757	632	197	11121705	1989	5592
gen H	503	744	396	107	7390488	1824	4052
fragmento L1	187	367	486	167	9166273	1613	5683
fragmentos L3.L4	4	378	1060	2094	149197881	3570	41792
<b>251M</b>							
gen N	0	6	135	1431	55932098	1572	35580
gen F	43	221	710	892	24530705	1989	12333
gen H	87	399	744	592	29804012	1824	16340
fragmento L1	23	2	143	962	43327300	1613	26861
fragmentos L3.L4	4	760	779	1725	75264497	3570	21082
<b>251O</b>							
fragmentos L3.L4	502	1314	864	315	11533330	3570	3231
Bases secuenciadas totales	<b>2120165483</b>						
Cobertura promedio	<b>20905</b>						

Tabla 12. Frecuencias de mutación para cada matriz de los casos clínicos discriminados por gen/fragmento genómico. ND: no determinado.

	Gen N	Gen P	Gen M	Gen F	Gen H	Gen L (L1, L2, L3,L4)
<b>245C</b>	2,24E-4	ND	ND	1,92E-4	1,16E-4	0,42E-4 L1 ND L2 0,47E-4 L3, L4
<b>245P</b>	1,2E-4	ND	ND	3,05E-4	2,4E-4	1,24E-4 L1 ND L2 0,74E-4 L3, L4
<b>245V</b>	1,9E-4	ND	ND	3,08E-4	ND	1,55E-4 L1 ND L2 1,5E-4 L3, L4
<b>246C</b>	0,89E-4	0,75E-4	ND	1,42E-4	1,07E-4	1,54E-4 L1 ND L2 1,35E-4 L3, L4
<b>246P</b>	1,34E-4	ND	ND	1,07E-4	1,29E-4	1,39 E-4 L1 1,8 E-4 L2 0,9 E-4 L3,L4
<b>246O</b>	ND	ND	ND	ND	ND	0,56E-4 L1 ND L2 0,85E-4 L3, L4
<b>246M</b>	0,93E-4	ND	ND	1,37E-4	1,24E-4	0,65 E-4 L1 ND L2 0,8 E-4 L3,L4
<b>251C</b>	1,62E-4	0,24E-4	ND	2,6E-4	1,49E-4	0,53E-4 L1 ND L2 0,71E-4 L3, L4
<b>251P</b>	0,68E-4	ND	ND	2,35E-4	2,51E-4	2,0 E-4 L1 ND L2 0,59E-4 L3, L4
<b>251O</b>	ND	ND	ND	ND	ND	1,1E-4 L3, L4
<b>251M</b>	0,82E-4	ND	ND	1,76E-4	1,23E-4	0,77 E-4 L1 ND L2 0,99 E-4 L3, L4



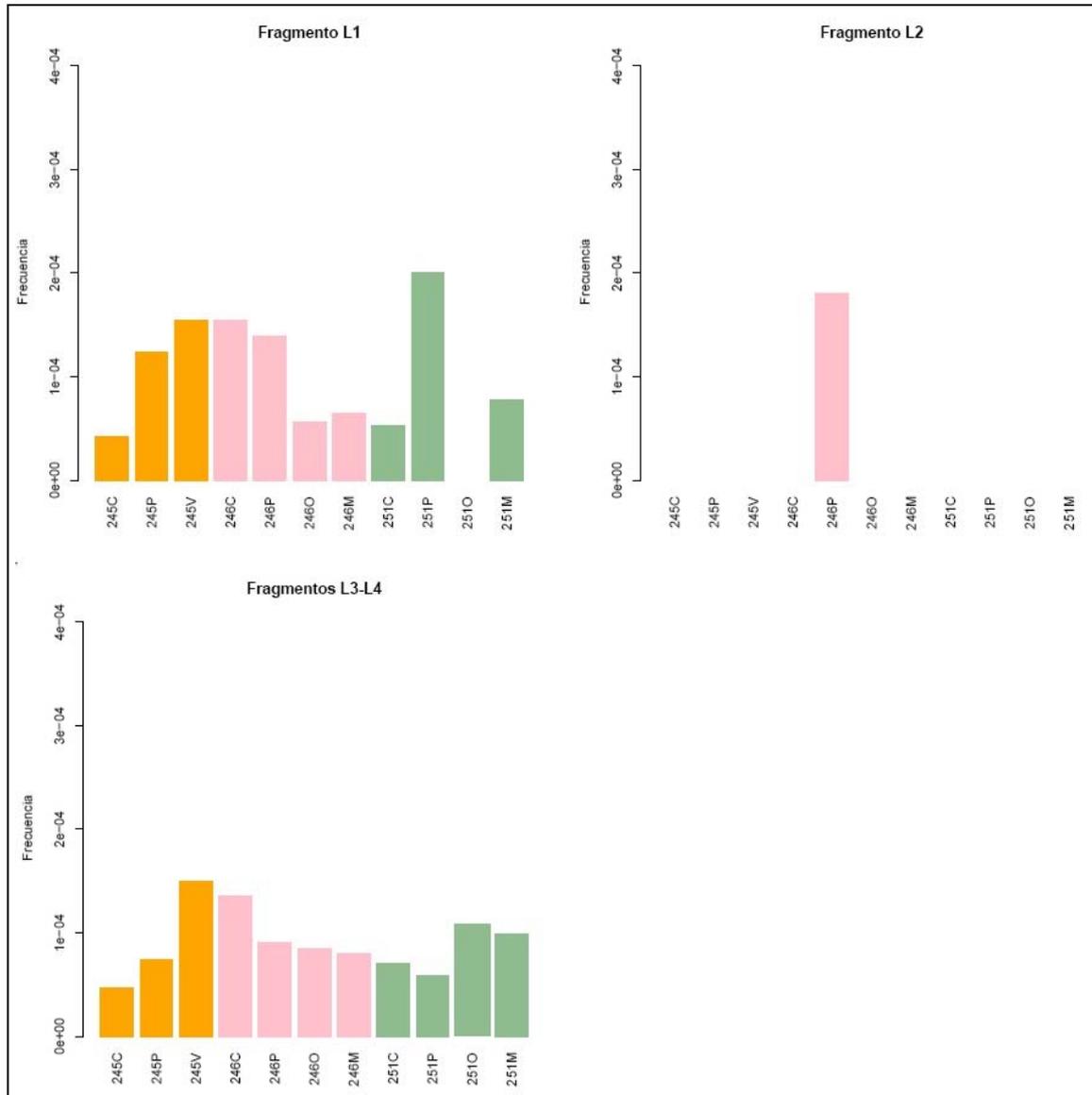


Figura 27. Representación gráfica de las frecuencias de mutación para las matrices de los tres casos clínicos discriminadas por gen/fragmento genómico.

Además, se analizó la distribución de la frecuencia de mutación de cada gen/fragmento genómico corregida por su longitud, discriminadas por matriz para los tres casos clínicos juntos. Los genes de las glicoproteínas de superficie H y F tuvieron la mayor frecuencia de mutación para las cuatro matrices alcanzando valores de  $2,51$  y  $3,08 \times 10^{-4}$  respectivamente, mientras que el gen P mostró los valores de frecuencia más bajos ( $0,24 - 0,74 \times 10^{-4}$ ) (Figura 28).

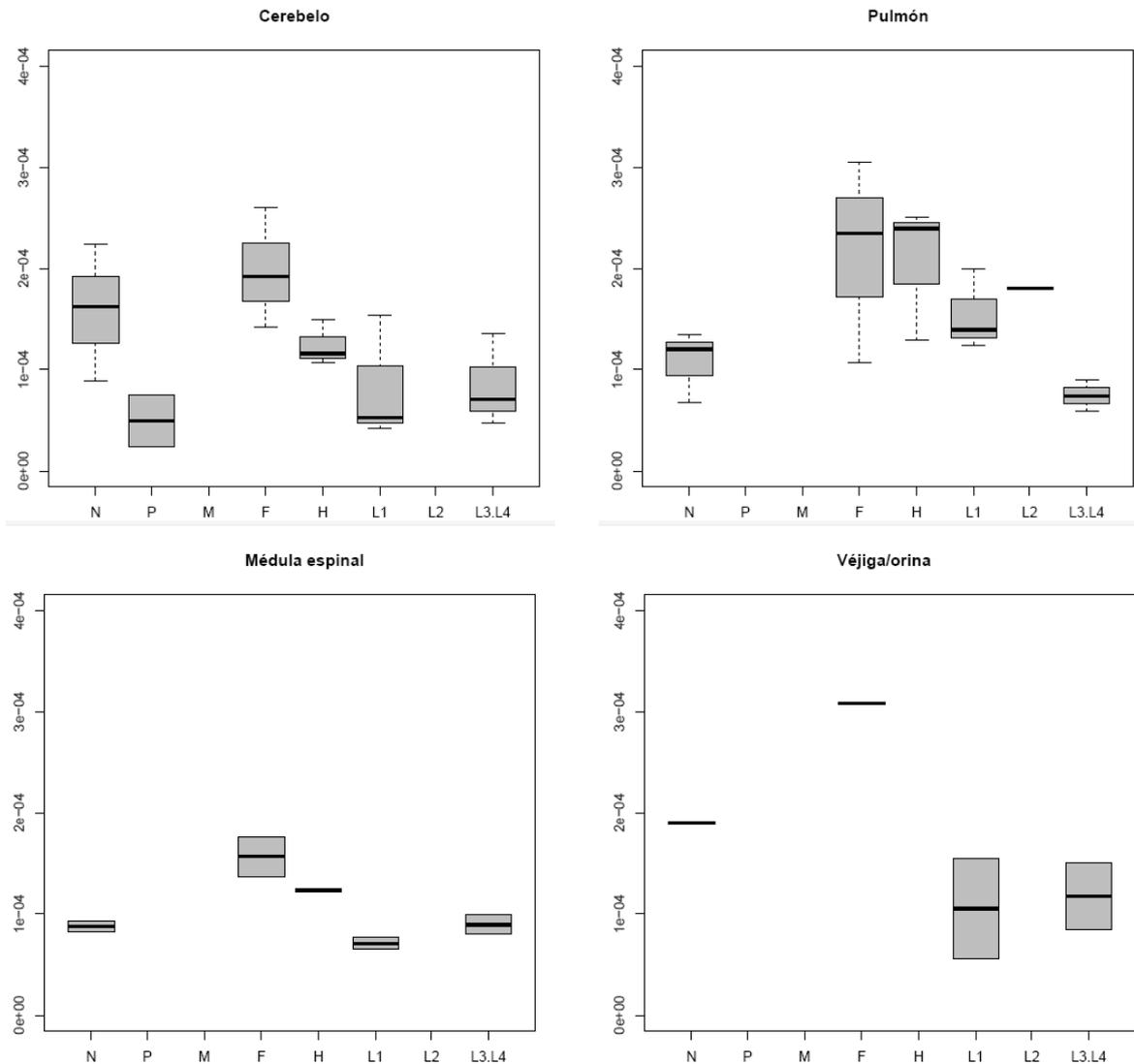


Figura 28. Representación mediante diagramas de caja de las frecuencias de mutación corregidas por la longitud de cada gen/fragmento genómico discriminadas por matriz para los tres casos clínicos.

### 3.4.3.2 Reconstrucción de secuencias genómicas consenso

Las secuencias genómicas consenso de cada matriz fueron inferidas a partir de los nucleótidos representados en mayor frecuencia en cada posición genómica. La longitud de las secuencias fue variable entre las once matrices en función de los genes/fragmentos genómicos secuenciados. La secuencia de mayor longitud correspondió a la matriz 251M con 13625 pb, mientras que la de menor longitud fue la de 251O con 3675 pb (Tabla 13) (Anexo I, hoja secuencias).

Las secuencias genómicas de 251P obtenidas por secuenciación de Sanger y NGS se compararon para estimar las diferencias nucleotídicas generadas por ambos protocolos. La comparación se realizó para 13165 nucleótidos del genoma de 251P secuenciados por ambos métodos, revelando cuatro sustituciones sinónimas C4916T, A5193G, A9329G y C12017T (Anexo I, hoja secuencias 251P).

Tabla 13. Reconstrucción de secuencias genómicas consenso para las once matrices analizadas. Se detalla para cada matriz la longitud de la secuencia consenso.

Matriz	Secuencia genómica (pb)
245C	10869
245P	12953
245V	7223
246C	12380
246P	12666
246M	10857
246O	4992
251C	12676
251P	13165
251M	13625
251O	3675

### 3.4.4 Análisis filodinámico

#### 3.4.4.1 Análisis por *likelihood mapping*

El método *likelihood mapping* (LM) permitió establecer cuál gen/fragmento genómico es el más adecuado para reconstrucciones filogenéticas mediante la estimación de su señal filogenética. El valor de la señal se calcula sumando las probabilidades en los vértices del triángulo que representan al porcentaje de árboles con topología resuelta, mientras que el valor del centro representa el porcentaje de topologías no definidas (Figura 29). Se considera que un conjunto de secuencias es apropiado para

reconstrucciones filogenéticas si el valor de la señal filogenética es superior al 90% (Schmidt & von Haeseler, 2007). Los genes H y F y los fragmentos L2 y L4 presentaron valores de topologías resueltas superiores a 90%, y valores de topologías no definidas en el rango de 2,1-7,4%. La señal filogenética más elevada fue de 95,8% y correspondió al fragmento L4 (Figura 29, 30). Los genes N, P y M y los fragmentos L1 y L3 no alcanzaron el valor de 90% en sus vértices, y a nivel de la región central presentaron valores en el rango de 11-15,5% (Figura 29, 30).

Debido a que el fragmento L4 (1848 pb) presentó valores de cobertura elevados en todas las matrices y la mayor señal filogenética, se escogió para realizar los análisis de migración de variantes entre matrices para cada caso clínico.

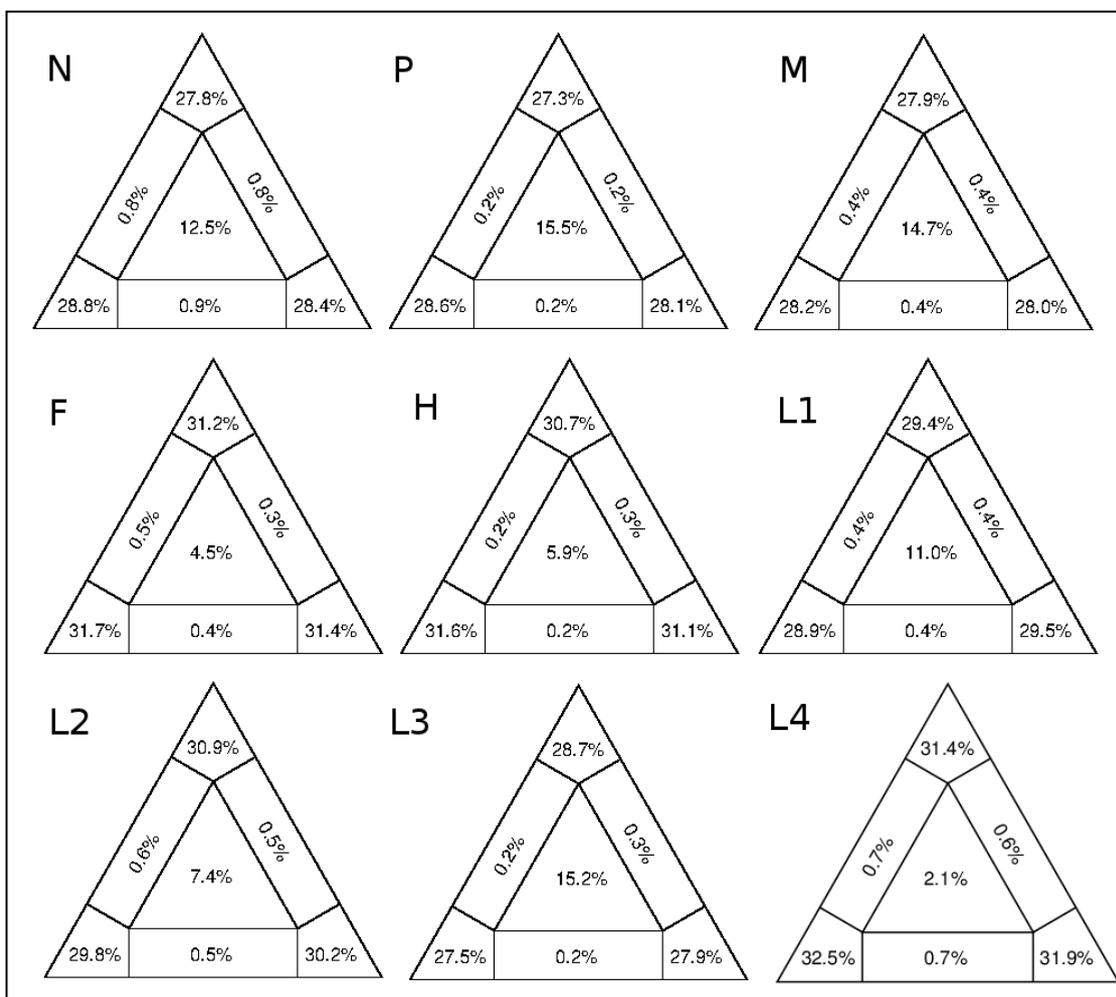


Figura 29. *Likelihood mapping* de los genes/fragmentos genómicos. Las probabilidades en los vértices de los triángulos representan a los árboles de topología definida, las cuales son tomadas en cuenta para estimar la señal filogenética de cada región.

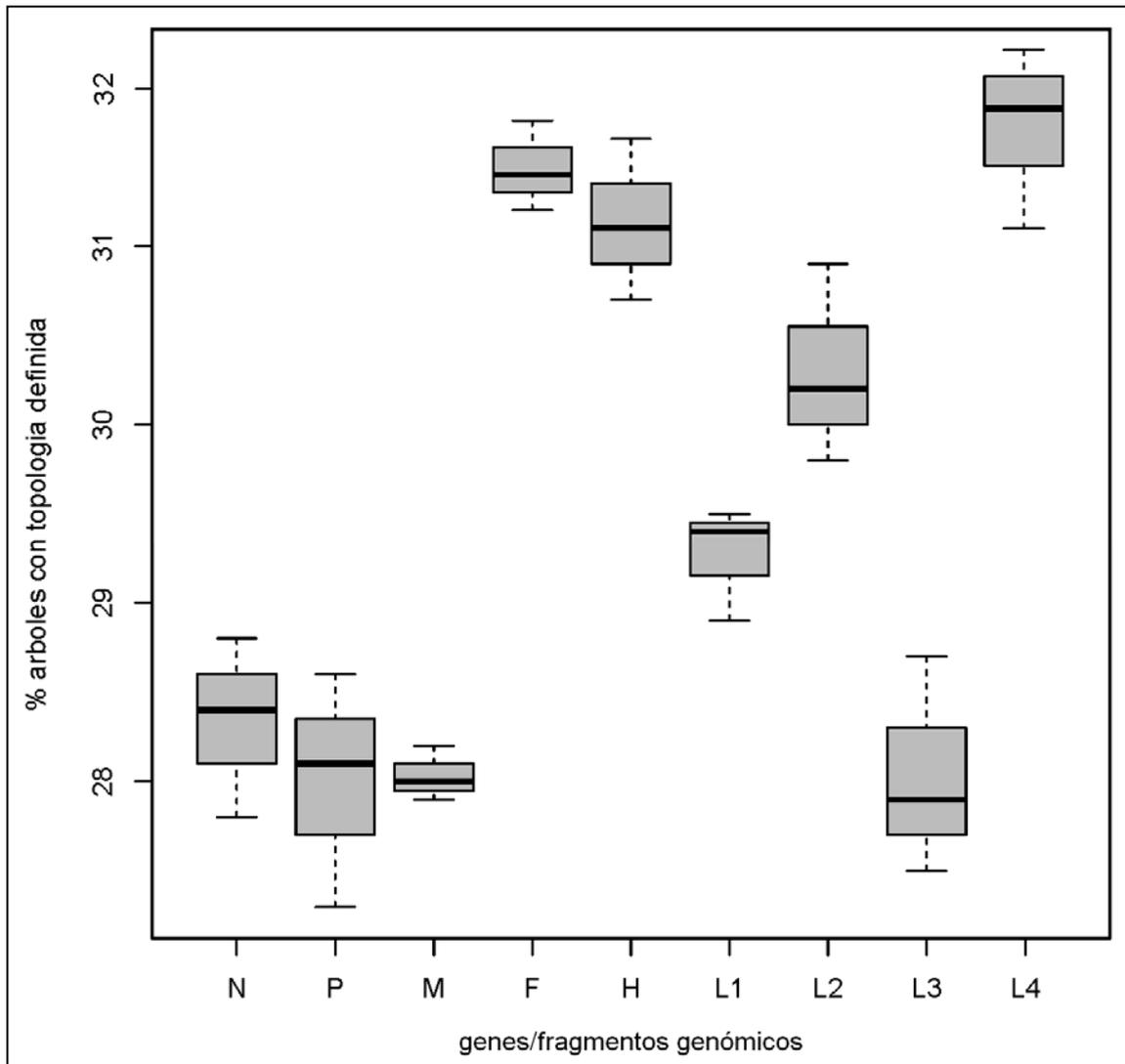


Figura 30. Representación de la distribución porcentual de árboles con topología definida para cada gen/fragmento genómico mediante diagramas de caja.

#### 3.4.4.2 Análisis de compartimentalización y migración

La existencia de variantes genéticas específicas para cada una de las matrices de los tres casos clínicos indica la presencia de subpoblaciones del virus en los distintos compartimientos (matrices) (Anexo I, hoja secuencias y cambios aa). El análisis de migración de variantes entre matrices se realizó utilizando las secuencias consenso y todas las variantes mayoritarias y minoritarias descritas para el fragmento L4 en las matrices de cada caso.

En el caso 245, se estableció que el cerebelo y el pulmón tuvieron la misma probabilidad de ser el tejido ancestral a partir del cual se diseminó la infección (Figura 31). La población ancestral de variantes se diversificó en ambos tejidos en dos clados

diferenciados. Posteriormente, ocurrió migración de variantes desde el cerebelo hacia el aparato urogenital, detectadas en vejiga (Figura 31). La migración de variantes entre matrices (subpoblaciones) se representó, además, mediante una red de migración en la cual se observa la migración de variantes desde el pulmón hacia el cerebelo, y un evento posterior correspondiente a la migración de variantes a la vejiga (Figura 32).

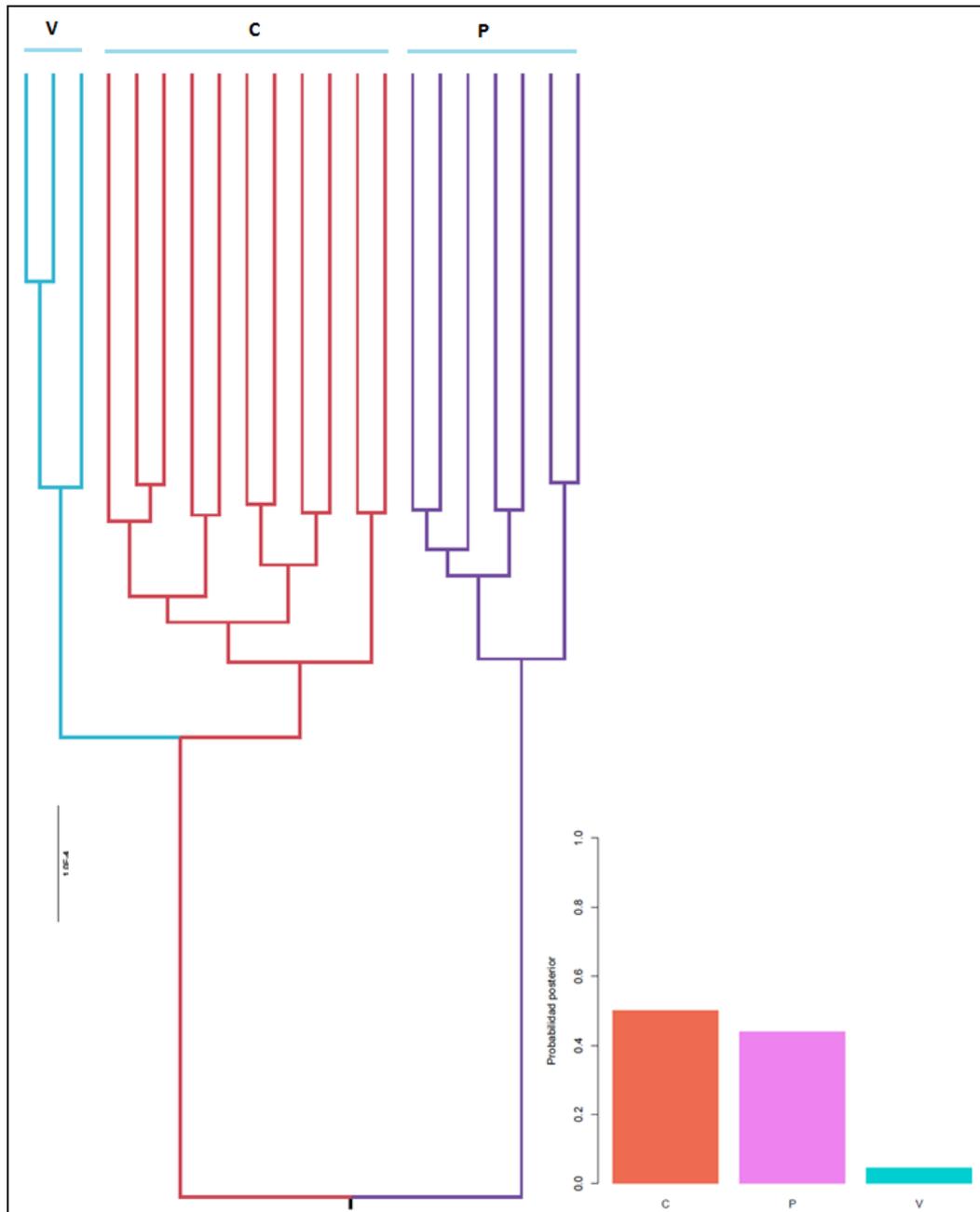


Figura 31. Filograma de variantes genéticas descritas en pulmón (P), cerebelo (C), y vejiga (O) del caso 245. Se detalla la probabilidad de ancestralidad para cada tejido.

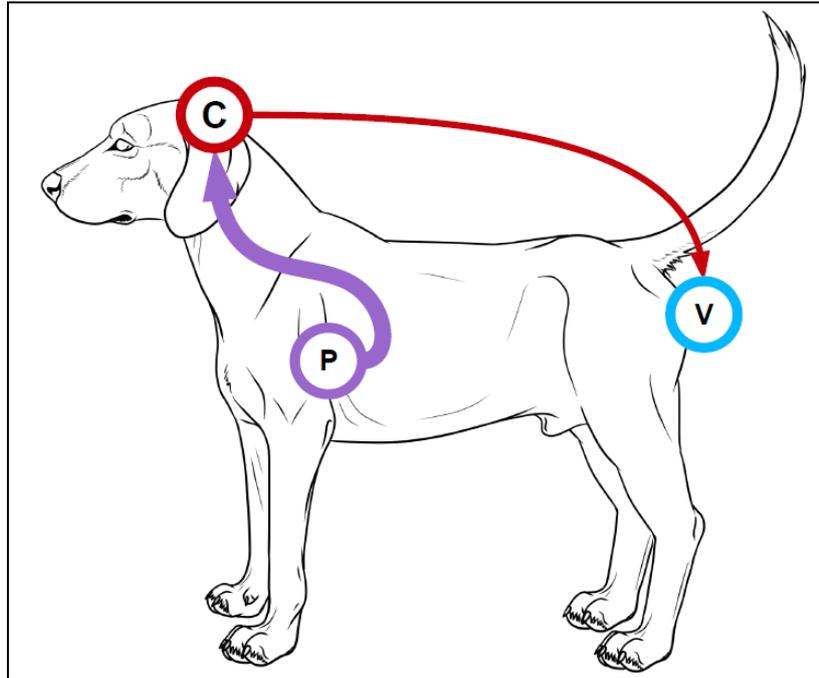


Figura 32. Red de migración de variantes genéticas entre matrices del caso 245. C: cerebelo, P: pulmón, V: vejiga. Las flechas indican la dirección de la migración y su grosor la proporción de migrantes hacia cada matriz.

Para el caso 246 se estableció que el pulmón presentó una mayor probabilidad posterior (0,6) de ser la matriz ancestral (Figura 33). Las variantes de pulmón se diversificaron en dos clados: el primer grupo de variantes se diseminó hacia la médula, y a partir de ésta hubo migración de variantes hacia el cerebelo. El segundo grupo de variantes se estableció en el pulmón, y posteriormente dio lugar a la subpoblación observada en orina (Figura 33). En la red de migración de este caso, se observó como las variantes de pulmón se diversificaron hacia los tejidos del aparato urogenital siendo detectadas en la orina, y en mayor parte se diseminaron hacia la médula espinal, a partir de la cual alcanzaron el cerebelo (Figura 34).

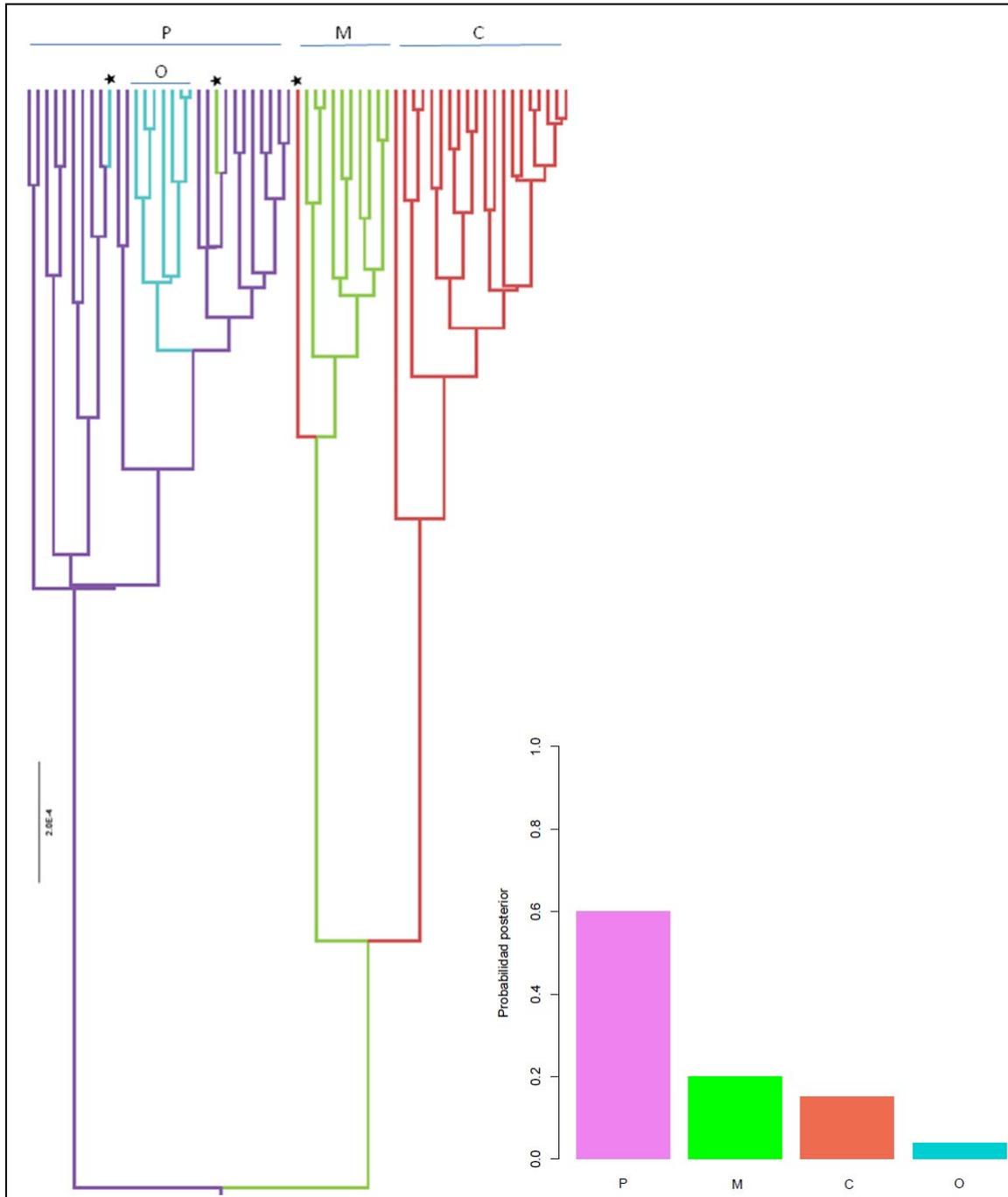


Figura 33. Filograma de variantes genéticas descritas en pulmón (P), cerebelo (C), médula espinal (M) y orina (O) del caso 246. Se detalla la probabilidad de ancestralidad para cada tejido. Los asteriscos marcan variantes que han migrado entre matrices (subpoblaciones).

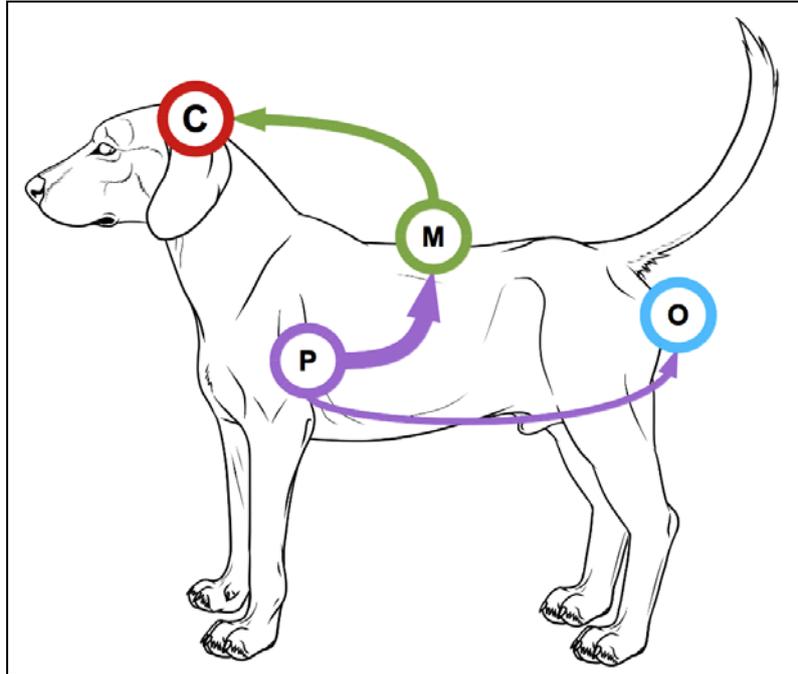


Figura 34. Red de migración de variantes genéticas entre matrices del caso 246. C: cerebelo, P: pulmón, M: médula, O: orina. Las flechas indican la dirección de la migración y su grosor la proporción de migrantes hacia cada matriz.

En el caso 251, el cerebelo presentó una probabilidad posterior de 0,6 de ser la matriz ancestral (Figura 35). Las variantes de cerebelo se diversificaron en dos grupos: el primero se diseminó hacia la médula, y a partir de ésta hubo migración de variantes hacia el aparato urogenital detectadas en orina. El segundo grupo de variantes se diversificó desde el cerebelo hacia el pulmón, y posteriormente ocurrió un segundo movimiento migratorio nuevamente hacia el cerebelo (Figura 35). La red de migración mostró que las variantes de cerebelo se diversificaron hacia la médula, dando origen a la subpoblación de orina, y posteriormente hacia el pulmón donde hubo migración de vuelta hacia el cerebelo. También se observaron migrantes de segundo orden desde el cerebelo al pulmón y a la médula (Figura 36).

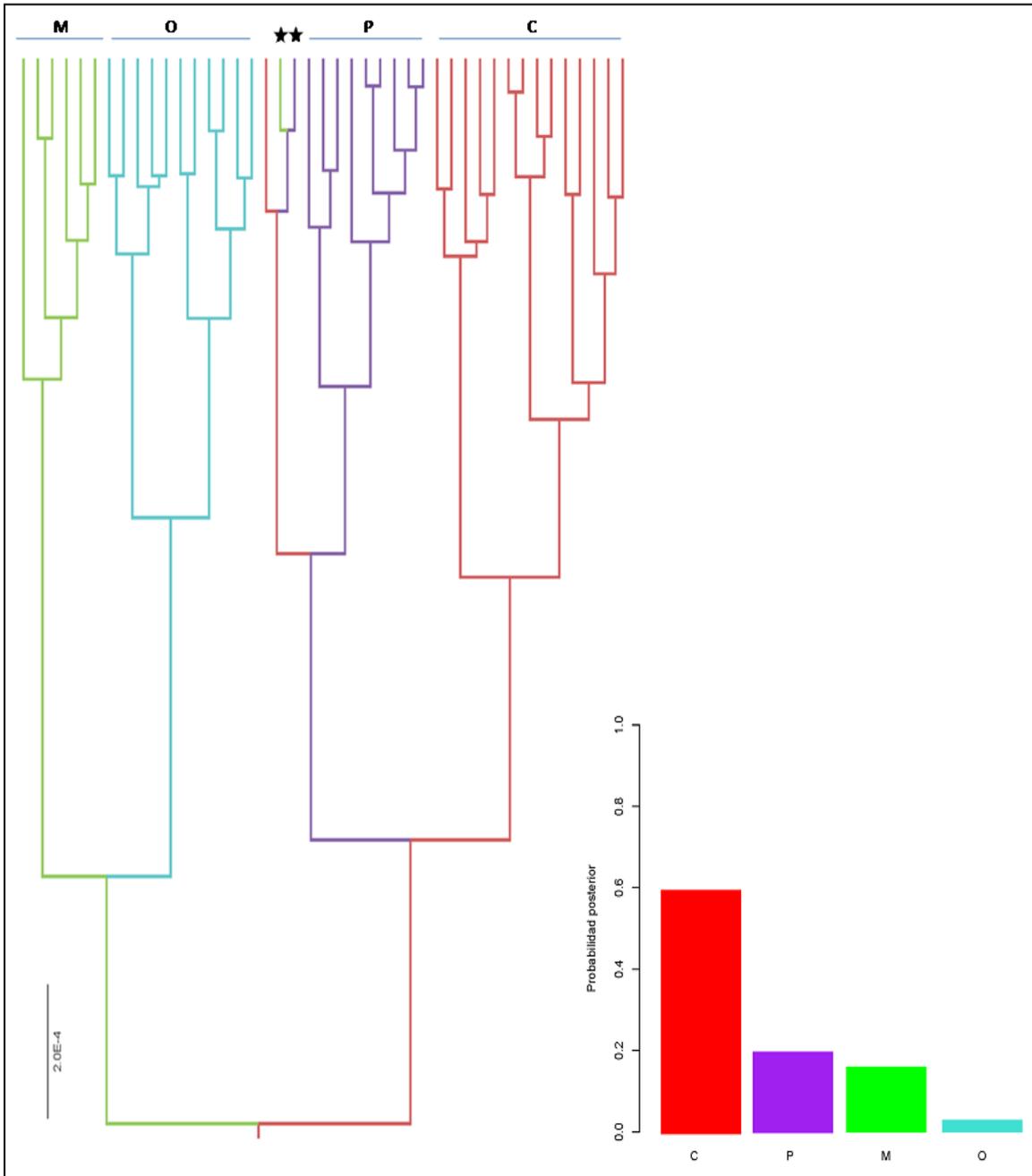


Figura 35. Filograma de variantes genéticas descritas en pulmón (P), cerebelo (C), médula espinal (M) y orina (O) del caso 251. Se detalla la probabilidad de ancestralidad para cada tejido. Los asteriscos marcan variantes que han migrado entre matrices (subpoblaciones).

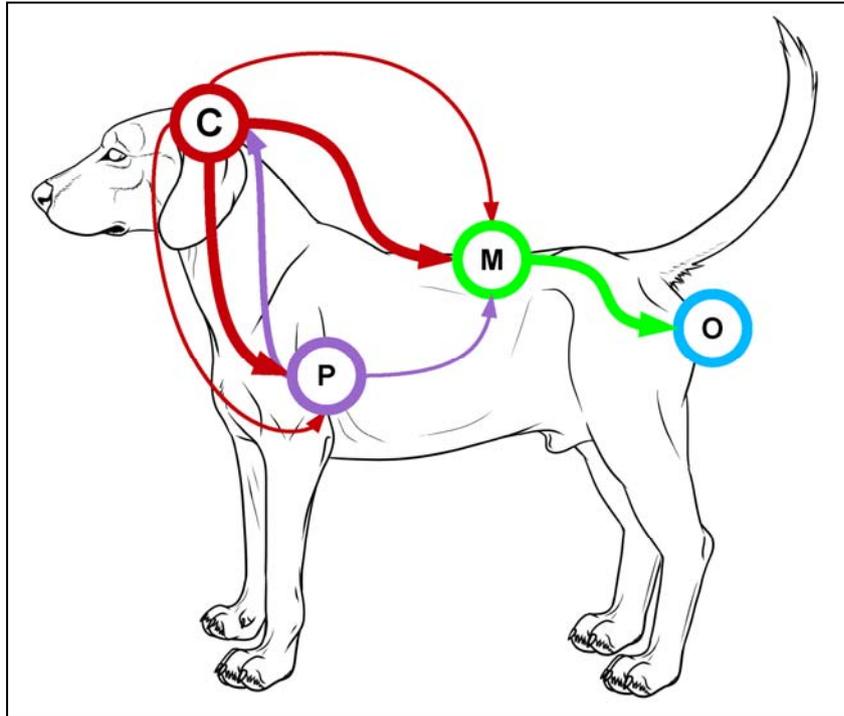


Figura 36. Red de migración de variantes genéticas entre matrices del caso 251. C: cerebelo, P: pulmón, M: médula, O: orina. Las flechas indican la dirección de la migración y su grosor la proporción de migrantes hacia cada matriz.

#### 4. Discusión

Los paramyxovirus son patógenos que afectan a organismos vertebrados y causan enfermedades infecciosas agudas de importancia sanitaria y económica. Dentro de esta familia, se incluyen patógenos humanos como el sarampión, paperas, virus sincitial respiratorio (RSV) y metapneumovirus humano, virus zoonóticos como Hendra y Nipah, y virus de importancia veterinaria que afectan aves de corral (virus Newcastle, metapneumovirus aviar), bovinos (RSV bovino) (Chang & Dutch, 2012) y carnívoros (virus distemper canino) (Deem et al., 2000).

El virus distemper canino (CDV) (*Paramyxoviridae*, *Morbillivirus*) es responsable de una severa enfermedad infecciosa denominada Distemper, con elevada incidencia y mortalidad en carnívoros domésticos y salvajes a nivel mundial (Deem et al., 2000). En las últimas décadas se han registrado numerosos brotes de la enfermedad en distintas regiones geográficas (Lednicky et al., 2004, Lan et al., 2006, Martella et al., 2006, An et al., 2008, Panzera et al., 2012, Woma et al., 2009), así como una marcada expansión en el rango de huéspedes (Roelke-Parker et al., 1996, van de Bildt et al., 2002, Kissui & Parcker, 2004, Goller et al., 2010). Actualmente, Distemper afecta a todas las familias de carnívoros terrestres, varias especies de mamíferos marinos, e incluso primates no homínidos (Ludlow et al., 2014, Sakai et al., 2013, Sun et al., 2010). Incluso, se ha propuesto que CDV ha favorecido la extinción de especies susceptibles como el tigre de Tasmania (*Thylacinus cynocephalus*) en la década de 1980 (Prowse et al., 2013).

Los registros históricos sugieren que los brotes de Distemper han ocurrido desde el siglo XVIII, y que el virus se habría dispersado desde Perú hacia Europa, ocasionando graves epidemias en canes domésticos (Howell, 1965, Blancou, 2004). Sin embargo, el virus fue aislado por primera vez a comienzos del siglo XX por Henri Carré quien estableció que CDV era el agente etiológico de Distemper (Carré, 1905). En la actualidad CDV está distribuido mundialmente y la caracterización de cepas de campo, basada en el análisis del gen de la hemaglutinina, ha permitido su clasificación en linajes genéticos, los cuales están distribuidos principalmente según un patrón geográfico (Martella et al., 2006). En Sudamérica, se han reportado casos de Distemper en canidos domésticos y silvestres en Argentina, Bolivia, Brasil, Chile, Colombia, Ecuador y Uruguay, y se ha determinado la presencia de cuatro linajes circulantes (Panzera et al., 2014). Los estudios de caracterización han permitido conocer la dispersión y evolución de las cepas de campo del virus a nivel mundial. No obstante, los estudios de tipificación viral no son informativos sobre la diversidad poblacional del virus y su dinámica dentro del

huésped. Se ha establecido que la organización en cuasiespecies es una propiedad inherente de virus RNA debido a sus tamaños poblacionales, tiempos generacionales y elevadas tasas de mutación (Domingo et al., 2012). Sin embargo, a la fecha no se ha establecido experimentalmente que CDV se organice en un sistema de cuasiespecies.

En la presente Tesis Doctoral analizamos aspectos genéticos y evolutivos de CDV mediante un enfoque multidisciplinario a partir de dos hipótesis de trabajo: CDV está organizado en un sistema de cuasiespecies, y presenta compartimentalización (subpoblaciones) del virus dentro de los tejidos del huésped. Resultados preliminares prominentes basados en el análisis de secuencias genómicas parciales de un caso clínico, nos llevaron a extender nuestro proyecto de investigación original. La estrategia experimental desarrollada en esta Tesis consistió en la amplificación del genoma completo a partir de diferentes matrices (tejidos/fluidos) de tres casos clínicos y su análisis por secuenciación masiva. En una primera instancia realizamos la puesta a punto de los protocolos de biología molecular a partir de una única matriz de un caso clínico (251P), y su posterior confirmación por secuenciación Sanger. Posteriormente, se utilizaron a las matrices como inóculos para las infecciones de la línea Vero-DST para la obtención del genoma completo. Nuestra estrategia fue apropiada para cumplir con los objetivos planteados y nos permitió analizar por primera vez la dinámica poblacional de CDV.

Los cebadores diseñados permitieron amplificar los nueve fragmentos genómicos solapantes de 251P, los cuales fueron secuenciados por Sanger, editados y ensamblados para obtener su secuencia genómica. Las regiones terminales del genoma (*leader* 3' y *tráiler* 5') no se incluyeron en nuestro análisis debido a los problemas asociados a la secuenciación y la secuencia codificante de 251P fue enviada a la base de datos del Genbank (KM280689). Esta estrategia nos permitió obtener, por primera vez en Uruguay, el genoma de una cepa de campo de CDV aislada de una infección natural de un perro doméstico, constituyendo además el primer registro para Sudamérica (Sarute et al., 2014) (Anexo II).

Si bien obtuvimos exitosamente el genoma completo directamente de una muestra biológica (251P), la amplificación del genoma e incluso de regiones genómicas específicas resulta dificultoso (Demeter et al., 2007, Muller et al., 2011, Panzera et al., 2012). Debido a ello, en nuestro estudio generamos inóculos a partir de las siete matrices de los tres casos clínicos para la infección de la línea celular Vero-DST, con el fin de amplificar el número de partículas virales tras 48-72 hs (~ 2-3 ciclos replicativos)

(Gallina et al., 2011) y facilitar la obtención del genoma completo a partir de todas las matrices. Se ha postulado que la adaptación de un virus a una línea celular representa una presión selectiva que puede determinar cambios genómicos a nivel poblacional (Kato et al., 2009). Sin embargo, estudios previos analizaron los efectos sobre la secuencia genómica de una cepa de CDV luego de 20 pasajes por la línea Vero-DST mediante la amplificación de 4201 pb del genoma viral, detectándose una sola sustitución nucleotídica que determinó un cambio aminoacídico (Lan et al., 2006). En base a este antecedente, realizamos la comparación de la secuencia genómica de 251P obtenida directamente de la muestra biológica y secuenciada por el método de Sanger, con la secuencia consenso obtenida por secuenciación masiva (NGS) luego de un pasaje en cultivo celular, para evaluar si el pasaje del virus en la línea Vero-DST efectivamente introduce mutaciones que puedan sesgar nuestros análisis poblacionales. La comparación se realizó para los 13165 nucleótidos secuenciados por ambos protocolos, y se identificaron tan solo cuatro sustituciones nucleotídicas sinónimas ubicadas en la región intergénica M-F (posición 4916), en el gen F (posición 5193) y en el gen L (posiciones 9329 y 12017). En la sustitución C4916T, la secuenciación NGS reveló una frecuencia de 0,69 para timina (T) y de 0,31 para citosina (C) por lo cual se anotó el nucleótido 4916T en la secuencia consenso, mientras que la secuenciación de Sanger asignó una C a esta posición (4916C). Esta discrepancia entre ambos métodos puede deberse a la elevada frecuencia de C a nivel poblacional, lo cual podría determinar la lectura de este nucleótido por el método de Sanger y la consecuente transición C-T. En el caso de las sustituciones A5193G y A9329G la frecuencia de guanina (G) a nivel poblacional fue mayor al 97% y este nucleótido fue anotado en la secuencia consenso obtenida por NGS. La presencia de adenina (A) en la secuenciación por Sanger no podría ser explicada por su frecuencia ya que ésta es menor al 3% en ambos casos, y este método de secuenciación no puede discriminar variantes en tales proporciones (Isakov et al., 2014). En estos casos, podríamos asumir que ambas sustituciones serían consecuencia de los errores de la secuenciación de Sanger, los cuales se hallan en el rango de 0,001-1% (Hoff, 2009). Un tercer tipo de incongruencia se observó en la sustitución C12017T, esta posición se encuentra dentro de la secuencia de un cebador lo que podría determinar que haya un incorrecto *base-calling* (asignación de nucleótido) en ambos protocolos de secuenciación (Ekblom et al., 2014). El análisis de la posición 12017 en la secuencia consenso de 251P obtenida por NGS muestra una inusual representación de los cuatro nucleótidos. Podemos afirmar que las frecuencias

nucleotídicas para esta posición están asociadas a una incorrecta asignación del nucleótido, y que no reflejan diversidad poblacional. Es más, la presencia de los cuatro nucleótidos en la posición 12017 también se observó para las demás matrices, confirmando que estos errores son propios de la secuenciación NGS (Anexo I, hoja secuencias). Si bien nuestra comparación se basa en dos protocolos de secuenciación distintos, podemos afirmar que las diferencias entre las secuencias de 251P obtenidas por Sanger y NGS se debieron principalmente a errores o limitantes en la secuenciación, y no a errores introducidos durante la replicación del virus en la línea Vero-DST. Estos resultados sustentan que nuestra estrategia experimental fue adecuada para el análisis de diversidad poblacional, debido a que el pasaje por cultivo celular no influyó sobre la estimación de los cambios nucleotídicos detectados en las variantes a nivel poblacional. Las infecciones de la línea Vero-DST se realizaron a partir de homogeneizados de siete matrices de los tres casos clínicos, los cuales fueron previamente titulados por TCID<sub>50</sub> obteniéndose títulos en el rango de  $1,68 \times 10^2$  -  $1,61 \times 10^4$  partículas virales/ml en solo tres matrices (245P, 246P y 251C) (Tabla 5). Nuestros resultados en la titulación no difieren sustancialmente de los obtenidos por otros investigadores. Estudios basados en infecciones de la línea Vero-DST a partir de homogeneizados tisulares de casos clínicos, revelaron títulos virales en el orden de  $1.7 \times 10^3$  -  $7.9 \times 10^5$  en los casos donde éste pudo ser determinado (Lan et al., 2009). Es más, en infecciones realizadas en dos líneas murinas (HmLu y BHK), utilizadas usualmente para la propagación de CDV, también se observaron títulos virales discretos ( $2.1 \times 10^1$  -  $2.9 \times 10^2$ ), incluso luego de un pasaje en cultivo celular (Sultan et al., 2009). En las matrices donde se obtuvo título viral mensurable, se observó efecto citopático (CPE) característico en cultivo con la formación de sincitios multinucleados (Figura 22), al igual que en la cepa Onderstepoort utilizada como control (Figura 20), la cual se caracteriza por presentar un CPE marcado con formación de grandes sincitios (Appel & Gillespie, 1972, Haig, 1956).

Sin embargo, en las 18 matrices restantes donde no se obtuvo título viral, se observaron distintos fenotipos en la línea Vero-DST. En las infecciones provenientes de cerebelo, pulmón y en dos de médula (casos 246 y 251) se observó CPE, y se confirmó la presencia del genoma viral mediante la amplificación del fragmento de 287 pb del gen N (Tabla 2, Figura 23). En las infecciones de 245M y las matrices vejiga/orina de los tres casos clínicos no se observó CPE aparente. Sin embargo, realizamos RT-PCR a partir del RNA viral, detectando al fragmento diagnóstico del gen N (Figura 23). Finalmente, en las infecciones con inóculos de estómago, intestino e hígado de los tres

casos clínicos no se observó CPE, ni fue posible amplificar el fragmento diagnóstico a partir del sobrenadante ni de la monocapa celular, incluso 6-7 días de incubación PI. Debido a que no se detectó al virus en ninguna de las infecciones realizadas, estas matrices fueron descartadas para los análisis poblacionales. El hecho de no haber detectado el genoma viral en tejidos del aparato gastrointestinal podría asociarse a que ciertos productos fueron tóxicos para las células Vero-DST. Mediante el paso de filtración en la preparación de inóculos (sección 2.3.2) se eliminan microorganismos que producen toxinas que reducen drásticamente la viabilidad celular (Osada et al., 2014). Se podría especular que ciertos productos celulares como enzimas o metabolitos podrían afectar la viabilidad de las células en cultivo. Sin embargo, no se ha demostrado que las enzimas hepáticas sean tóxicas para las líneas celulares comerciales, independientemente de la dosis en que sean administradas (Nelson et al., 2013). Podemos asumir, entonces, que en estos tejidos el título viral era discreto o inexistente al momento de realizarse las necropsias.

Nuestros resultados mostraron que la replicación del virus en cultivo no siempre es evidenciada por la formación de sincitios. Sin embargo, este hecho no es llamativo ya que las cepas de CDV se clasifican en tres tipos según su fenotipo en células Vero: 1) cepas que se replican lentamente y no presentan CPE, 2) cepas que se replican y crecen normalmente sin formación aparente de sincitios, 3) cepas de replicación y crecimiento normales con un CPE evidente (Lan et al., 2005).

Del total de 108 amplicones correspondientes a los nueve fragmentos solapantes del genoma completo de las cuatro matrices de los tres casos clínicos, se obtuvieron 92 (85%) provenientes de once matrices, mientras que en la matriz 245M no se obtuvo ningún amplicón (Tabla 6). Para esta matriz repetimos el protocolo experimental exitosamente desde el comienzo y los datos de secuenciación están siendo actualmente analizados (Anexo III). Los 92 fragmentos genómicos fueron secuenciados mediante un protocolo de secuenciación masiva, de los cuales 28 no se emplearon en los análisis poblacionales debido a que no alcanzaron los valores mínimos de cobertura (100X). El análisis de las secuencias obtenidas reveló la ocurrencia de miles de sustituciones nucleotídicas en las once matrices analizadas, las cuales representan de forma ideal a todas las variantes genéticas presentes en las poblaciones de los tres casos clínicos (Anexo I, hoja secuencias). En este sentido, nuestro trabajo es de suma importancia debido a que confirmamos nuestra hipótesis inicial demostrando experimentalmente que CDV se organiza en cuasiespecies. La teoría de cuasiespecies desarrollada por Eigen &

Schuster fue extendida por Esteban Domingo quien estableció que la organización en cuasiespecies es una característica inherente a todos los virus RNA (Domingo et al., 2012). En base a este enunciado, desarrollamos nuestra hipótesis de trabajo logrando establecer por primera vez la organización en cuasiespecies para un miembro del género *Morbillivirus*.

Si bien la organización en cuasiespecies fue sugerida para otros géneros de la familia *Paramyxoviridae*, la mayoría de los estudios se basan en el análisis de un número limitado de clones (~30) por secuenciación de Sanger; para el género *Avulavirus* se analizó la heterogeneidad poblacional del paramyxovirus de paloma-tipo1 (Barbezange & Justin, 2005) y del virus de Newcastle (Kattenbelt et al., 2010), para el género *Pneumovirus* el virus sincitial respiratorio bovino (Deplanche et al., 2007), y para el género *Rubulavirus* el virus Mumps (paperas) (Sauder et al., 2006). A la fecha, solamente se ha realizado un estudio de diversidad mediante la secuenciación masiva de un genoma completo del paramyxovirus aviar-tipo 4 perteneciente al género *Avulavirus*, confirmando su organización en cuasiespecies (Abolnik et al., 2012). Nuestro estudio es, por tanto, el primero en analizar la diversidad poblacional mediante secuenciación masiva para la familia *Paramyxoviridae* analizando diversas matrices provenientes de tres casos clínicos.

Las 28 secuencias que no alcanzaron la cobertura mínima establecida (100X) en el análisis con el programa ViVan, correspondieron principalmente al gen M y al fragmento L2. Los errores en los protocolos de NGS pueden ocasionarse durante el proceso de fragmentación del DNA, por la estructura del gen/fragmento genómico, variaciones a nivel de secuencia, y errores en el mapeo y alineamiento de los *reads*. La consecuencia general de estos errores es una marcada variación en la profundidad de cobertura y en la correcta asignación de nucleótidos para cada posición genómica. Ciertos patrones de secuencia como repetidos invertidos, contenido GC y motivos GGC se asocian con un elevado número de errores en las plataformas de Illumina (Ekblom et al., 2014). Debido a ello, se estimaron los parámetros asociados a errores en la secuenciación. El contenido GC fue determinado mediante la herramienta GC-Profile (<http://tubic.tju.edu.cn/GC-Profile>), obteniéndose valores de 43,05 y 42,88% para el gen M y el fragmento L2, respectivamente. El análisis de los nueve fragmentos genómicos determinó valores de contenido GC en el rango de 41,1- 47,4%, correspondiendo el valor más elevado al gen H. Sin embargo, para este gen no ocurrieron problemas asociados a la cobertura para ninguna de las matrices analizadas, por lo cual podemos

descartar que este parámetro haya influido en la baja profundidad de cobertura del gen M y el fragmento L2. Los motivos GGC y repetidos invertidos se analizaron mediante la implementación del algoritmo REPFIND (<http://zlab.bu.edu/repfind>) y no se registraron valores en proporciones significativas para el gen M y el fragmento L2, desestimando que estos parámetros hayan influido en la baja cobertura. Otros errores de la secuenciación, denominados sistemáticos, ocurren cuando hay errores en la lectura de una misma región genómica obtenida a partir de *reads* diferentes. Estos errores son de especial importancia cuando se registra baja cobertura debido a que ciertas regiones genómicas no tendrán nucleótidos asignados (Meacham et al., 2011). La ocurrencia de errores sistemáticos en los amplicones del gen M (2022 pb) y el fragmento L2 (2148 pb) parece improbable debido a que estos errores deberían haber ocurrido en toda su longitud, de forma tal que no se pudieran asignar nucleótidos para ninguna posición genómica. Por tanto, no podemos establecer en primera instancia los motivos que determinaron una baja profundidad de cobertura en el gen M y el fragmento L2 debido a que ninguno de los parámetros que afectan a la secuenciación fue detectado, y a que los errores sistemáticos parecen improbables. Es más, aún cambiando nuestra estrategia experimental mediante el diseño de nuevos juegos de cebadores que hibridan y amplifican distintos fragmentos (Tabla 2, 7) y realizando una segunda corrida en otro secuenciador de Illumina (MiSeq), nuevamente se obtuvo baja profundidad de cobertura imposibilitando incluir a las secuencias de estas regiones genómicas en nuestros análisis poblacionales. Actualmente, se están analizando los datos de la tercera corrida realizada en el secuenciador MiSeq del Instituto Pasteur Montevideo, los cuales se obtuvieron mediante un protocolo de secuenciación masiva que implementa otro método de fragmentación para la generación de las librerías genómicas (Anexo III).

El análisis de las secuencias de las once matrices mediante el programa ViVan permitió establecer parámetros poblacionales tales como la presencia de variantes mayoritarias y minoritarias, cambios aminoacídicos asociados a las variantes, cambios sinónimos y no sinónimos, y tasas de transición y transversión para las once matrices analizadas (Anexo I). El análisis de la diversidad poblacional mostró la presencia de variantes genéticas (minoritarias y mayoritarias) en distinta proporción en las once matrices. Las variantes dentro de cada matriz de los tres casos clínicos representan subpoblaciones caracterizadas por distintos genotipos que exploran el *sequence space* de manera diferencial. En términos generales, no se observó una mayor diversidad de variantes para una matriz específica. En el caso 245 el mayor número de variantes se registró en

cerebelo, en 246 correspondió a pulmón, y en el caso 251 la matriz con mayor diversidad de variantes fue la orina (Figura 25). Las variantes de cada huésped se diversifican en subpoblaciones dentro de los tejidos y fluidos corporales (compartimientos), y el patrón de diversificación puede asociarse al curso patogénico de la infección y/o al tropismo celular del virus. Un estudio de compartimentalización basado en el análisis de la heterogeneidad poblacional del virus de la hepatitis C en muestras de suero sanguíneo, células mononucleares de sangre periférica y tejido hepático mostró que la mayor diversidad de variantes correspondió al suero, fluido por el cual se transmite el virus, aunque éste presenta un marcado tropismo por los hepatocitos (Navas et al., 1998). De acuerdo a este patrón, en nuestro estudio la mayor diversidad de variantes se debería haber observado en la orina, debido a que constituye una de las vías principales de transmisión de CDV. Como fue mencionado anteriormente este fenómeno ocurrió solamente para el caso clínico 251, descartando en base a nuestros resultados que la mayor diversidad poblacional está asociada a los fluidos de excreción.

Los cambios aminoacídicos asociados a variantes mayoritarias se detectaron en las secuencias aminoacídicas de los cinco genes analizados (N, P, F, H y L) (Tabla 8). Incluso se observaron cambios conservados en los mismos sitios de las glicoproteínas H y F en matrices de distintos casos clínicos, los cuales podrían tener valor adaptativo y servir a distintas funciones dentro de la población, debido al rol de estas proteínas en la invasión y propagación del virus entre células del huésped (Lamb & Parks, 2007). Dentro de los cambios conservados se identificó a la sustitución en la posición 284 de la proteína F, y tres sustituciones en las posiciones 275, 327 y 549 en la proteína H. A la fecha no se cuenta con la estructura cristalográfica de la proteína F para ningún miembro del género *Morbillivirus*, requiriéndose estudios adicionales para poder establecer eventuales interacciones que sirvan a la función de la proteína F. En el caso de la hemaglutinina, se cuenta con la estructura de la proteína del virus del sarampión e incluso se la ha cristalizado unida al receptor celular SLAM (Hashiguchi et al., 2011). Según estos modelos, los residuos 275 y 327 están localizados en hojas  $\beta$  del dominio transmembrana de la proteína, pero no se encuentran asociados a regiones de función conocida, mientras que el residuo 549 se localiza en la región de unión al receptor SLAM (Zipperle et al., 2010). Debido a la localización del residuo 549 se ha discutido su función sobre las interacciones que establece la hemaglutinina con el receptor SLAM. McCarthy et al. (2007) realizaron estudios de máxima verosimilitud con el

objetivo de determinar el rol de la selección sobre la evolución de CDV, empleando 73 secuencias del gen H obtenidas de cánidos domésticos y silvestres. El análisis reveló que una pequeña proporción (0.09%) de los residuos de la hemaglutinina se encuentra bajo selección positiva, dentro de los cuales se incluye al 549. En base al análisis de los residuos seleccionados en las 73 secuencias, propusieron que la sustitución Y549H estaría involucrada en la transmisión inter-específica de CDV debido a que detectaron la presencia del residuo 549H en la mayoría de las secuencias de las especies silvestres analizadas (McCarthy et al., 2007). Un estudio posterior amplió el *dataset* de secuencias analizando 146 secuencias del gen H pertenecientes a huéspedes domésticos y silvestres, mediante la implementación de un test binomial para establecer si existía un sesgo para la presencia del residuo 549H en especies silvestres. Los resultados indicaron que no existe un sesgo significativo para la presencia de la sustitución 549H en carnívoros silvestres, por tanto no se puede concluir que esta sustitución esté seleccionada en huéspedes silvestres ni que sea determinante para el salto inter-específico de CDV (Nikolin et al., 2010). Esta incongruencia sugiere que la sustitución de este residuo en la región de unión al receptor de la hemaglutinina podría tener otras implicancias.

Nuestro análisis reveló que la sustitución Y549H estaba presente en variantes aisladas de cerebelo de los casos 245 y 251. Podríamos especular que esta sustitución estaría involucrada en superar la barrera del sistema nervioso central (SNC) y en el neurotropismo, en lugar de favorecer la transmisión entre especies. Recientemente, se ha establecido que la dispersión de CDV en astrocitos y neuronas es mediada por la proteína H y su unión a un receptor celular de la familia Nectina (Mori et al., 2013). La presencia de esta sustitución en variantes aisladas de cerebelo indicaría que esta mutación está involucrada en la invasión del SNC y en la infección de células nerviosas mediante la unión al receptor. En términos generales, la capacidad de un virus de infectar a una célula no depende exclusivamente del reconocimiento del receptor celular, sino también de factores intracelulares que permitan a la célula ser permisiva o refractaria a la multiplicación viral, y a la formación y liberación de viriones. Si bien la unión virus-receptor puede no ser suficiente para que ocurra la multiplicación del virus a nivel celular, es necesaria para los primeros pasos de la infección (Domingo et al., 2012). Para evaluar si el cambio Y549H es relevante para la unión del virus a células del SNC y etapas posteriores de la infección, se requieren de estudios funcionales *in*

*vitro* o con modelos animales con el fin de establecer el posible rol de esta sustitución sobre el fenotipo neurotrópico.

Los cambios aminoacídicos en las proteínas de nucleocápside, fosfoproteína, matriz y *large* polimerasa podrían estar implicados en modificar interacciones proteína-proteína durante el ciclo replicativo. En poliovirus se han descrito mutantes que determinan una fidelidad de copia más baja (*mutator*) o más elevada (*antimutator*) respecto a virus *wild-type*. Los fenotipos *mutator* y *antimutator* están determinados por sustituciones aminoacídicas puntuales en la polimerasa viral que modulan su actividad. La fidelidad de copia más baja se debe a la sustitución H237R, mientras que la fidelidad más elevada es producto de la sustitución G64S (Vignuzzi et al., 2006, Korboukh et al., 2014). La mutación G64S determina un fenotipo resistente a la ribavirina, un mutágeno químico que incrementa la frecuencia de mutación sobre el umbral de error, y se ha establecido en poliovirus que esta resistencia está asociada al fenotipo *antimutator* (Vignuzzi et al., 2006). Estudios realizados *in vitro* han demostrado que la ribavirina inhibe la replicación de CDV de una forma dependiente del tiempo y la dosis, observándose la mayor eficacia cuando los viriones son liberados al medio extracelular (24 hs) y a bajas concentraciones (Elia et al., 2008, Gallina et al., 2011). El aislamiento de mutantes resistentes a la ribavirina en un sistema *in vitro* nos permitiría evaluar si los cambios identificados en nuestro estudio en el dominio catalítico de la polimerasa viral (fragmento L1) afectan la fidelidad de copia y el fenotipo viral. Si bien en poliovirus los fenotipos *mutator* y *antimutator* determinaron una reducción en la virulencia en modelos *in vivo*, este fenómeno no debería ocurrir necesariamente para CDV, y los mutantes aislados podrían tener un comportamiento distinto influyendo incluso sobre la dinámica poblacional del virus.

Las variantes detectadas a nivel poblacional también permitieron calcular la frecuencia de mutación para cada gen/fragmento genómico de las once matrices. Este valor refleja la proporción de mutantes presentes en la población, resultado de la mutabilidad del virus, tamaño poblacional, duración del ciclo replicativo, ocurrencia de eventos selectivos y/o la presencia de barreras genéticas o anatómicas (Domingo et al., 2012). La frecuencia se determinó en base a los sitios variables dentro de un gen/fragmento genómico en función de las bases secuenciadas. En los casos donde se obtuvo baja cobertura los valores de frecuencia son marcadamente mayores debido al valor del denominador en relación al número de sitios variables. El programa ViVan establece que una cobertura de 100X permite discriminar variantes en la población (*true calls*) de

errores asociados a la secuenciación, y por tanto las secuencias podrían emplearse para describir la heterogeneidad poblacional (Isakov et al., 2014). Sin embargo, nuestros cálculos de frecuencia se realizaron con secuencias que presentaron una cobertura mayor a 3000X debido a que con valores menores la frecuencia se sobreestimaba. Por tanto, en nuestro estudio debimos definir un filtro de calidad adicional para que los cálculos de frecuencia no estén sesgados por la profundidad de cobertura. Implementando este filtro adicional en las secuencias nuestros estudios poblacionales se realizaron en base a 54 secuencias genómicas.

Los valores de frecuencia de mutación para las secuencias de los 54 genes/fragmentos genómicos de las once matrices se ubicaron en el rango de  $0,24 - 3,08 \times 10^{-4}$ . Los valores más elevados correspondieron a los genes de las glicoproteínas H y F, siendo  $2,51$  y  $3,08 \times 10^{-4}$ , respectivamente (Tabla 12). Estos valores podrían estar asociados a que ambas proteínas se hallan en la superficie y están sujetas a diversas presiones selectivas como la respuesta inmune celular y humoral, y la acción de factores antivirales que el virus debe evadir para que se establezca la infección de células susceptibles (Lamb & Parks, 2007).

Estudios realizados en otros paramyxovirus establecieron la frecuencia de mutación de regiones genómicas específicas mediante clonación molecular y secuenciación de Sanger. En el virus sincitial respiratorio bovino (bRSV) se analizaron 774 pb del gen de la glicoproteína G en 26 clones, detectándose valores de frecuencia en el rango de  $6,8 - 10,1 \times 10^{-4}$  (Deplanche et al., 2007), mientras que en el paramyxovirus de paloma-tipo 1 (pPMV-1) se analizaron fragmentos parciales (~600 pb) de los genes F y HN de 18 clones obteniéndose valores en el rango de  $0,5 - 2,7 \times 10^{-4}$  (Barbezange & Justin, 2005). La estimación de frecuencias de mutación a partir de datos de secuenciación masiva no se ha realizado para ningún paramyxovirus, ya que el único trabajo de diversidad poblacional que emplea la tecnología NGS no determina este parámetro (Abolnik et al., 2012). Los valores obtenidos para el bSRV y pPMV-1 se encuentran en el mismo orden ( $10^{-4}$ ) de los valores de nuestro estudio. Sin embargo, se requiere de la amplificación de genes completos para poder establecer una comparación preliminar con nuestros resultados, y de la secuenciación de un número mayor de clones, con el fin de aumentar la resolución del estudio, y evitar la asignación incorrecta de cambios nucleotídicos debidos a errores de la secuenciación.

En términos de diversidad poblacional, la cuasiespecie debe considerarse como una unidad replicativa dentro de una célula infectada que producirá una progenie

genéticamente diversa capaz de invadir a otras células. Este concepto es importante para analizar el flujo génico de variantes dentro del huésped, ya que en la viremia se genera competencia entre genotipos provenientes de diferentes unidades replicativas para invadir diversos órganos y tejidos, originando subpoblaciones características de cada matriz (compartimiento) (Domingo et al., 2012). Los estudios filodinámicos permitieron estimar la matriz ancestral a partir de la cual se diseminó la infección, la proporción de migrantes para cada caso clínico y en última instancia, el curso patogénico de la infección entre las matrices de cada caso. El análisis de migración se realizó utilizando la secuencia consenso y las secuencias de todas las variantes mayoritarias y minoritarias del fragmento L4 de las once matrices. Esta región fue secuenciada para todas las matrices y presentó la señal filogenética más elevada, siendo la más adecuada para realizar reconstrucciones filogenéticas dentro del genoma de CDV.

En el caso 245, dos matrices (cerebelo y pulmón) tuvieron la misma probabilidad de ser la matriz ancestral, lo cual podría deberse a la falta de variantes aisladas de la médula que aportarían una mayor resolución a nuestro análisis. A partir de la población ancestral se detectó migración desde el cerebelo hacia la vejiga sin la ocurrencia de eventos de migración posteriores (Figura 32). La migración de variantes del cerebelo a la vejiga puede haber ocurrido a través de la médula como consecuencia de la respuesta inmune a nivel del SNC frente a la infección por CDV para ser finalmente eliminadas a través de la orina. Cuando ocurre infección en el SNC se desarrolla una intensa respuesta en el foco infeccioso, determinando la migración del virus hacia otras regiones evadiendo la inmunidad generada (Plattet et al., 2005). Sin embargo, no podemos descartar que haya existido migración de variantes de pulmón hacia la médula debido a que no contamos con datos de secuencias de esta matriz. El caso 245 presentaba síntomas en el aparato auditivo, digestivo y síntomas neurológicos como tics y convulsiones (Tabla 1). La presencia de variantes en cerebelo es coherente con los registros clínicos, sin embargo, no siempre se puede correlacionar la sintomatología con el curso de la infección. Mediante los análisis poblacionales y de compartimentalización demostramos que el virus se replicó activamente en pulmón y en vejiga, pero no mostró síntomas clínicos aparentes asociados al sistema respiratorio ni urogenital. Del modo opuesto, el animal presentó síntomas a nivel del aparato digestivo, pero no pudimos recuperar al virus a partir de estómago e intestino. En este caso, podemos especular que la respuesta inmune habría eliminado al virus de estos tejidos al momento del deceso,

aunque el animal sufrió diarreas y hemorragias a nivel digestivo durante el curso de la infección.

Para el caso 246 se determinó que el pulmón tuvo la probabilidad mayor de ser la matriz ancestral, y a partir de ésta ocurrieron eventos complejos de migración y de diversificación de variantes hacia otras matrices. El primer evento migratorio involucró a las variantes ancestrales que alcanzaron la médula espinal y de allí migraron hacia al cerebelo, mientras que en la población original de pulmón hubo un movimiento principal de migrantes hacia el aparato urogenital dando lugar a la subpoblación de orina. Además de estos eventos principales, se detectaron migrantes de segundo orden hacia la médula y la orina a partir de pulmón, y un segundo evento migratorio de la médula al cerebelo (Figura 34). A nivel de la sintomatología, en el caso 246 solamente se presentaron síntomas neurológicos agudos. Nuestros análisis mostraron que el virus se diseminó a las cuatro matrices mediante distintos eventos de migración, pero una vez que el virus alcanzó el cerebelo no se registró flujo génico a otras matrices. Este patrón es característico de la ocurrencia de una infección crónica a nivel del SNC, donde la persistencia viral en astrocitos y neuronas dirige una progresiva demielinización y eventualmente la muerte del animal. La persistencia es un proceso que no requiere de partículas infectivas y es resultado de la transmisión viral célula-célula (Plattet et al., 2005). Así, el virus es capaz de propagarse entre diversas áreas del SNC provocando lesiones severas sin que sea neutralizado por la respuesta inmune (Wyss-Fluehmann et al., 2010). Las cepas de CDV que determinan persistencia se denominan neurovirulentas y presentan cambios en sus proteínas virales que determinan modificaciones en el ensamblaje y la brotación del virus. En las cepas neurovirulentas, la interacción entre la nucleocápside y la proteína de matriz durante el proceso de brotación es menor a nivel de la membrana celular, por lo cual la liberación de partículas virales al medio extracelular estará disminuida (Zurbriggen et al., 1995). Si bien no podemos establecer experimentalmente si la infección del caso 246 fue provocada efectivamente por una cepa neurovirulenta, nuestros análisis poblacionales y de compartimentalización y migración, junto a la sintomatología clínica observada, sugiere que la cepa 246 podría ser neurovirulenta.

En el caso 251, el cerebelo tuvo la probabilidad mayor de ser la matriz ancestral, y a partir de ésta ocurrieron eventos de migración y diversificación de variantes hacia pulmón y médula. Los eventos migratorios a partir de las variantes de médula originaron la subpoblación de orina, mientras que las variantes de pulmón se

diversificaron hacia la médula y nuevamente al cerebelo. Además de estos eventos principales, se detectaron migrantes de segundo orden hacia la médula y el pulmón desde el tejido ancestral (Figura 36). Respecto a la sintomatología, el caso 251 presentó síntomas gastrointestinales y manifestaciones secundarias como conjuntivitis y descargas óculo-nasales (Tabla 1). En este caso clínico es interesante el patrón de migración observado entre cerebelo y pulmón. En una etapa inicial de la infección, el virus habría alcanzado el SNC mediante el transporte por vía sanguínea mediada por células del sistema inmune infectadas. Posteriormente, el virus se diseminó a los epitelios del sistema respiratorio, y una vez allí otra vez invadió el SNC probablemente a través de la vía anterógrada mediada por el nervio olfativo (Figura 36) (Rudd et al., 2006). Si bien ocurrieron complejos eventos migratorios entre el sistema nervioso central y periférico, y el sistema respiratorio, en primera instancia no podríamos relacionar estos patrones migratorios con la sintomatología clínica, la cual estuvo asociada principalmente al aparato digestivo.

Los análisis filodinámicos revelaron aspectos interesantes de la dinámica poblacional del virus entre las distintas matrices de los tres casos clínicos, incluso para el caso 246 sugerimos una correlación entre los eventos migratorios y el curso patogénico de la infección (Figura 34). Sin embargo, es dificultoso establecer correlaciones directas entre migración y sintomatología clínica debido a que no conocemos el tiempo que transcurrió desde que el animal se infectó hasta su deceso para ningún caso clínico, y además nuestra información clínica es parcial, ya que contamos con los síntomas que presentó el animal en el ingreso a la clínica pero no contamos con el registro clínico completo. Contar con dicha información para los casos clínicos analizados habría sido de utilidad para realizar una integración más precisa de los análisis filodinámicos con los síntomas clínicos, ya que durante la viremia el virus alcanza distintos órganos y tejidos de un modo dependiente del tiempo post-infección (Deem et al., 2000). De forma ideal, deberíamos realizar infecciones experimentales en modelos animales para trabajar con condiciones de muestreo controladas. CDV ha sido utilizado como modelo para establecer mecanismos de patogénesis y virulencia en el género *Morbillivirus* (von Messling et al., 2003). La posibilidad de realizar infecciones controladas que nos permitan seguir la progresión de la enfermedad, permitiría vincular precisamente la diversidad poblacional con la patogénesis viral, y extender los resultados a otros virus relacionados con un curso patogénico similar, como el virus del sarampión.

El presente trabajo de Tesis tuvo como objetivos analizar la diversidad poblacional, la compartimentalización tejido/específica y eventos de flujo genético para tres casos de infección natural por CDV. La estrategia experimental diseñada permitió confirmar nuestras hipótesis de trabajo, estableciendo que CDV está organizado en un sistema de cuasiespecies, lo cual representa el primer registro para el género *Morbillivirus*. La estimación de otros parámetros poblacionales como la frecuencia de mutación para cada gen/fragmento genómico de las once matrices y la identificación de cambios aminoacídicos producto de variantes mayoritarias que podrían estar involucrados en la interacción virus-célula, permitió analizar aún en mayor profundidad la composición genética de las poblaciones virales de cada matriz.

El análisis de la diversidad poblacional para cada caso clínico reveló patrones de compartimentalización tejido-específica, ya que se identificaron subpoblaciones de variantes características para cada matriz analizada. Mediante aproximaciones heurísticas se realizó un análisis filodinámico y se estableció la matriz ancestral a partir de la cual se diseminó la infección en cada caso, y se describieron los eventos de migración de variantes entre matrices. El análisis de los datos clínicos de cada animal junto a los datos filodinámicos permitió estimar cual habría sido el curso patogénico de la infección para cada caso. El caso 246 habría desarrollado una infección crónica a nivel del SNC, mientras que en los casos 245 y 251 se establecieron los patrones de migración entre matrices, pero no se pudieron asociar directamente con la sintomatología.

En el presente trabajo de tesis se presenta información novedosa y de importancia respecto a la heterogeneidad y organización poblacional de CDV, sus patrones de compartimentalización y migración, y la eventual relación de la diversidad poblacional con la patogénesis viral.

#### **4.1 Conclusiones**

- El análisis de las secuencias obtenidas por secuenciación masiva reveló que CDV está organizado en cuasiespecies.
- CDV presenta subpoblaciones (compartimentalización) de variantes genéticas características para cada matriz de los tres casos clínicos.
- El análisis filodinámico reveló complejos eventos de migración entre las matrices dentro del huésped, los cuales podrían asociarse al curso patogénico de la infección.
- El caso 246 se trataría de una cepa neurovirulenta en virtud de su dinámica poblacional dentro del huésped y de la sintomatología clínica.

#### **4.2 Perspectivas**

Dentro de las actividades a realizar por nuestro grupo de investigación se destacan:

- Construcción de un sistema de genética reversa para analizar el efecto de las mutaciones detectadas en distintas líneas celulares.
- Realizar infecciones experimentales con CDV en modelos animales para poder analizar el curso de la infección en tiempo real.
- Establecer colaboraciones con especialistas en el área de producción de proteínas recombinantes y cristalografía, para contar con la estructura atómica de las proteínas virales con el objetivo de localizar y analizar residuos aminoacídicos específicos.

## **5. Referencias bibliográficas**

- Abolnik C, de Castro M, Rees J (2012). Full genomic sequence of an African avian paramyxovirus type 4 strain isolated from a wild duck. *Virus Genes*, 45: 537-541.
- Acevedo A, Brodsky L, Andino R (2013). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 505: 686-690.
- An DJ, Yoon SH, Park JY, No IS, Park B (2008). Phylogenetic characterization of Canine Distemper Virus isolates from naturally infected dogs and a marten in Korea. *Vet. Microbiol.*, 132: 389-395.
- Anderson DE, von Messling V (2008). Region between the canine distemper virus M and F genes modulates virulence by controlling fusion protein expression. *J. Virol.*, 82: 10510-10522.
- Appel MJG, Gillespie JH (1972). Canine distemper virus. En Gard S, Hallauer S y Meyer KF (eds.), *Virology monographs* 11: 1-96. Springer-Verlag, New York.
- Appel MJ, Summers BA (1999). Canine Distemper: Current Status. *Recent Advances in Canine Infectious Diseases*. Carmichael LE Ed. International Veterinary Information Service.
- Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res.* doi: 10.1093/nar/gks498.
- Baltimore D (1971). Expression of animal virus genomes. *Bacteriol Rev.*, 35: 235-241.

- Barbezange C, Jestin V (2005). Molecular study of the quasispecies evolution of a typical pigeon paramyxovirus type 1 after serial passages in pigeons by contact. *Avian Pathol.* 34: 111-122.
- Beaucourt S, Bordería AV, Coffey LL, Gnädig NF, Sanz-Ramos M, Beeharry Y, Vignuzzi M (2011). Isolation of fidelity variants of RNA viruses and characterization of virus mutation frequency. *J Vis Exp.*, 16 (52). pii: 2953
- Bellini WJ, Englund G, Richardson CD, Rozenblatt S, Lazzarini RA (1985). Matrix genes of measles virus and canine distemper virus: cloning, nucleotide sequences, and deduced amino acid sequences. *J Virol*, 58: 408-416.
- Bieringer M, Han JW, Kendl S, Khosravi M, Plattet P, Schneider-Schaulies J (2013). Experimental adaptation of wild-type canine distemper virus (CDV) to the human entry receptor CD150. *PLoS ONE*. 8(3):e57488.
- Blancou J (2004). Dog distemper: imported into Europe from South America?. *Historia Medicina Veterinaria*, 29: 35-41.
- Bordería AV, Stapleford KA, Vignuzzi M (2011). RNA virus population diversity: implications for inter-species transmission. *Curr Opin Virol.* 1: 643-648.
- Butina TV, Denikina NN, Belikov SI (2010). Canine distemper virus diversity in Lake Baikal seal (*Phoca sibirica*) population. *Vet Microbiol.* 144: 192-197.
- Carlsson B, Lindberg AM, Rodríguez-Díaz J, Hedlund KO, Persson B, Svensson L (2009). Quasispecies dynamics and molecular evolution of human norovirus capsid P region during chronic infection. *J Gen Virol.*, 90: 432-441.
- Carré H (1905). Sur la maladie des jeunes chiens. *C. R. Acad. Sci. (Paris)*., 140: 689–690, 1489–1491.

- Cattaneo R, Kaelin K, Baczko K, Billeter MA (1989). Measles virus editing provides an additional cysteine-rich protein. *Cell*, 56: 759-764.
- Chang A, Dutch RE (2012). Paramyxovirus fusion and entry: multiple paths to a common end. *Viruses*, 4: 613-636.
- Chen S, Wang YM (2007). Evolutionary study of hepatitis C virus envelope genes during primary infection. *Chin Med J (Engl)*. 120: 2174-2180.
- Combe M, Sanjuán R (2014a). Variability in the mutation rates of RNA viruses. *Future Virol.*, 9: 605-615
- Combe M, Sanjuán R (2014b). Variation in RNA Virus Mutation Rates across Host Cells. *PLoS Pathog.* 10(1): e1003855.
- Cottam EM, King DP, Wilson A, Paton DJ, Haydon DT (2009). Analysis of Foot-and-mouth disease virus nucleotide sequence variation within naturally infected epithelium. *Virus Res.*, 140: 199-204.
- Darriba D, Taboada GL, Doallo R, Posada D (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods*. 9: 772.
- de Swart RL, Harder TC, Ross PS, Vos HW, Osterhaus AD (1995). Morbilliviruses and morbillivirus diseases of marine mammals. *Infect Agents Dis*, 4: 125-130.
- Deem SL, Spelman LH, Yates RA, Montali RJ (2000). Canine distemper in terrestrial carnivores: a review. *J. Zoo. Wildl. Med.*, 31: 441-451.
- Demeter Z, Lakatos B, Palade EA, Kozma T, Forgach P, Rusvai M (2007). Genetic diversity of Hungarian canine distemper virus strains. *Vet. Microbiol.*, 122: 258-269.

- Deplanche M, Lemaire M, Mirandette C, Bonnet M, Schelcher F, Meyer G (2007). In vivo evidence for quasispecies distributions in the bovine respiratory syncytial virus genome. *J Gen Virol.*, 88: 1260-1265.
- Domingo E, Sabo D, Taniguchi T, Weissmann C (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell*, 13: 735–744.
- Domingo E, Sheldon J, Perales C (2012). Viral quasispecies evolution. *Microbiol Mol Biol Rev.*, 76: 159-216.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol.* 29:1969-1973.
- Duffy S, Shackelton LA, Holmes EC (2008). Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.*, 9: 267-276.
- Eigen M (1971). Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58: 465–523.
- Eigen M, Schuster P (1978). The hypercycle-a principle of natural self-organization. C. The realistic hypercycle. *Naturwissenschaften*, 65: 341–369.
- Eigen M, Schuster P (1979). The hypercycle. A principle of natural self-organization. Springer, Berlin, Alemania.
- Ekblom R, Smeds L, Ellegren H (2014). Patterns of sequencing coverage bias revealed by ultra-deep sequencing of vertebrate mitochondria. *BMC Genomics*. 15: 467.
- Elena SF (2012). RNA virus genetic robustness: possible causes and some consequences. *Curr Opin Virol.*, 2: 525-530.

- Fiorello CV, AJ Noss Deem SL (2006). Demography, hunting ecology and pathogen exposure of domestic dogs in the Isoso of Bolivia. *Conservation Biology*, 20: 762-771.
- Frisk AL, König M, Moritz A, Baumgartner W (1999). Detection of Canine Distemper Virus Nucleoprotein RNA by Reverse Transcription-PCR Using Serum, Whole Blood, and Cerebrospinal Fluids from Dogs with Distemper. *J Clin Microbiol.*, 37: 3634-3643.
- Frolich K, Czupalla O, Haas L, Hentschke J, Dedek J, Fickel, J (2000). Epizootiological investigations of canine distemper virus in free-ranging carnivores from Germany. *Vet Microbiol*, 74: 283-292.
- Gabriel G, Dauber B, Wolff T, Planz O, Klenk HD, Stech J (2005). The viral polymerase mediates adaptation of an avian influenza virus to mammalian host. *Proc Natl Acad Sci USA*. 102: 18590-18595.
- Gabriel G, Abram M, Keiner B, Wagner R, Klenk HD, Stech J (2007). Differential polymerase activity in avian and mammalian cells determines host range of influenza virus. *J Virol*. 81: 9601-9604.
- Gallina L, Dal Pozzo F, Galligioni V, Bombardelli E, Scagliarini A (2011). Inhibition of viral RNA synthesis in canine distemper virus infection by proanthocyanidin A2. *Antiviral Res.*, 92: 447-452
- Goller KV, Fyumagwa RD, Nikolin V, East ML, Kilewo M, Speck S, Müller T, Matzke M, Wibbelt G (2010). Fatal canine distemper infection in a pack of African wild dogs in the Serengeti ecosystem, Tanzania. *Vet Microbiol*. 146: 245-252.
- Gong L, Han Y, Chen L, Liu F, Hao P, Sheng J, Li XH, Yu DM, Gong QM, Tian F, Guo XK, Zhang XX (2013). Comparison of next-generation sequencing and

clone-based sequencing in analysis of hepatitis B virus reverse transcriptase quasispecies heterogeneity. *J Clin Microbiol.* 51: 4087-4094.

- Harder TC, Kenter M, Vos H, Siebelink K, Huisman W, van Amerongen G, Orvell C, Barrett T, Appel MJ, Osterhaus AD (1996). Canine distemper virus from diseased large felids: biological properties and phylogenetic relationships. *J Gen Virol*, 77: 397-405.
- Hashimoto M, Une Y, Mochizuki M (2001). Hemagglutinin genotype profiles of canine distemper virus from domestic dogs in Japan. *Arch Virol.* 146: 149-155.
- Hoff KJ (2009). The effect of sequencing errors on metagenomic gene prediction. *BMC Genomics.* 10: 520.
- Howell DG (1965). Bacterial and viral disease. *Brit Vet J.*, 121: 398-401.
- Isakov O, Bordería A, Golan D, Hamenahem A, Celniker G, Yoffe L, Blanc H, Vignuzzi M, Shomron N (2014). Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. Aceptado en *Genome Biology*.
- Karzon DT, Bussel RH (1959). Cytopathic effect of canine distemper virus in tissue culture. *Science.* 130: 1708-1709.
- Kato N, Abe K, Mori K, Ariumi Y, Dansako H, Ikeda M (2009). Genetic variability and diversity of intracellular genome-length hepatitis C virus RNA in long-term cell culture. *Arch Virol.*, 154: 77-85.
- Kattenbelt JA, Stevens MP, Selleck PW, Gould AR (2010). Analysis of Newcastle disease virus quasispecies and factors affecting the emergence of virulent virus. *Arch Virol.*, 155: 1607-1615

- Kissui BM, Packer C (2004). Top-down population regulation of a top predator: lions in the Ngorongoro Crater. *Proc Biol Sci.*, 271: 1867-1874.
- Korboukh VK, Lee CA, Acevedo A, Vignuzzi M, Xiao Y, Arnold JJ, Hemperly S, Graci JD, August A, Andino R, Cameron CE (2014). RNA Virus Population Diversity: An Optimum for Maximal Fitness and Virulence. *J Biol Chem.* pii: jbc.M114.592303.
- Lamb RA, Parks GD (2007). *Paramyxoviridae: The Viruses and Their Replication en Fields Virology*, Lippincott Williams & Wilkins, 5<sup>a</sup> Ed.
- Lan NT, Yamaguchi R, Kai K, Uchida K, Kato A, Tateyama S (2005). The growth profiles of three types of canine distemper virus on Vero cells expressing canine signaling lymphocyte activation molecule. *J Vet Med Sci.* 67: 491-495.
- Lan NT, Yamaguchi R, Kawabata A, Uchida K, Kai K, Sugano S, Tateyama S (2006). Stability of canine distemper virus (CDV) after 20 passages in Vero-DST cells expressing the receptor protein for CDV. *Vet Microbiol.* 118: 177-188.
- Lan NT, Yamaguchi R, Hirai T, Kai K, Morishita K (2009). Relationship between growth behavior in vero cells and the molecular characteristics of recent isolated classified in the Asia 1 and 2 groups of canine distemper virus. *J Vet Med Sci.* 71: 457-461.
- Lancaster KZ, Pfeiffer JK (2012). Viral population dynamics and virulence thresholds. *Curr Opin Microbiol.*, 15: 525-530.
- Lauring AS, Andino R (2010). Quasispecies Theory and the Behavior of RNA Viruses. *PLoS Pathog.* 6(7): e1001005.

- Lauring AS, Acevedo A, Cooper SB, Andino R (2012). Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell Host Microbe*, 12: 623-632.
- Lednicky JA, Dubach J, Kinsel MJ, Meehan TP, Bocchetta M, Hungerford LL, Sarich NA, Witecki KE, Braid MD, Pedrak C, Houde CM (2004). Genetically distant American Canine distemper virus lineages have recently caused epizootics with somewhat different characteristics in raccoons living around a large suburban zoo in the USA. *Viol. J.* 1:2.
- Liermann H, Harder TC, Lochelt M, von Messling V, Baumgartner W, Moennig V, Haas L (1998). Genetic analysis of the central untranslated genome region and the proximal coding part of the F gene of wild-type and vaccine canine Distemper Morbilliviruses. *Virus Genes*. 17: 259-270.
- Ludlow M, Rennick LJ, Nambulli S, de Swart RL, Duprex WP (2014). Using the ferret model to study morbillivirus entry, spread, transmission and cross-species infection. *Curr Opin Virol*. 4: 15-23.
- Martella V, Elia G, Lucente M, Decaro N, Lorusso E, Banyai K, Blixenkrone-Møller M, Lan N, Yamaguchi R, Cirone F, Carmichael L, Buonavoglia C. (2006) Heterogeneity within the hemagglutinin genes of canine distemper virus (CDV) strains detected in Italy. *Vet Microbiol* 116: 301–309.
- McCarthy AJ, Shaw MA, Goodman SJ (2007). Pathogen evolution and disease emergence in carnivores. *Proc Biol Sci*. 274: 3165-3174.
- Meacham F, Boffelli D, Dhahbi J, Martin DI, Singer M, Pachter L (2011). Identification and correction of systematic error in high-throughput sequence data. *BMC Bioinformatics*. 12:451.
- Mori M, Rikitake Y, Mandai K, Takai Y (2013). Roles of Nectins and Nectin-Like Molecules in the Nervous System. *Cell Adhesion Molecules*, Capítulo 6, Springer-Verlag, New York.

- Moss WJ, Griffin DE (2006). Global measles elimination. *Nat Rev Microbiol.*, 4: 900-908.
- Mühlebach MD, Mateo M, Sinn PL, Prüfer S, Uhlig KM, Leonard VH, Navaratnarajah CK, Frenzke M, Wong XX, Sawatsky B, Ramachandran S, McCray PB Jr, Cichutek K, von Messling V, Lopez M, Cattaneo R (2011). Adherens junction protein nectin-4 is the epithelial receptor for measles virus. *Nature*, 480: 530-533.
- Müller A, Silva E, Santos N, Thompson G (2011). Domestic dog origin of canine distemper virus in free-ranging wolves in Portugal as revealed by hemagglutinin gene characterization. *J. Wildl. Dis.*, 47: 725-729.
- Munson L, Terio KA, Kock R, Mlengeya T, Roelke ME, Dubovi E, Summers B, Sinclair AR, Packer C (2008). Climate extremes promote fatal co-infections during canine distemper epidemics in African lions. *PLoS ONE*, 3(6):e2545.
- Nava AF, Cullen L, Sana DA, Nardi MS, Filho JD, Lima TF, Abreu KC, Ferreira F. (2008). First evidence of canine distemper in Brazilian free-ranging felids. *Ecohealth*, 5: 513-518.
- Navas S, Martín J, Quiroga JA, Castillo I, Carreño V (1998). Genetic diversity and tissue compartmentalization of the hepatitis C virus genome in blood mononuclear cells, liver, and serum from chronic hepatitis C patients. *J Virol.*, 72: 1640-1646.
- Nikolin VM, Wibbelt G, Michler FU, Wolf P, East ML (2010). Susceptibility of carnivore hosts to strains of canine distemper virus from distinct genetic lineages. *Vet Microbiol.* 156: 45-53.
- Origi FC, Plattet P, Sattler U, Robert N, Casaubon J, Mavrot F, Pewsner M, Wu N, Giovannini S, Oevermann A, Stoffel MH, Gaschen V, Segner H, Ryser-Degiorgis MP (2012). Emergence of canine distemper virus strains with modified

molecular signature and enhanced neuronal tropism leading to high mortality in wild carnivores. *Vet Pathol.* 49: 913-929.

- Panzera Y, Calderón MG, Sarute N, Guasco S, Cardeillac A, Bonilla B, Hernández M, Francia L, Bedó G, LaTorre J, Pérez R (2012). Evidence of two co-circulating genetic lineages of canine distemper virus in South America. *Virus Res.* 163: 401-404.
- Panzera Y, Sarute N, Carrau L, Aldaz J, Perez R (2014). Genetic diversity of canine distemper virus in South America. *Brit J Virol.* 1: 48-53.
- Parker J, Rambaut A, Pybus OG (2008). Correlating viral phenotypes with phylogeny: accounting for phylogenetic uncertainty. *Infect Genet Evol.*, 8: 239-246.
- Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell lifespan, and viral generation time. *Science*, 271: 1582-1586.
- Plattet P, Rivals JP, Zuber B, Brunner JM, Zurbriggen A, Wittek R (2005). The fusion protein of wild-type canine distemper virus is a major determinant of persistent infection. *Virology.* 337: 312-326.
- Plattet P, Cherpillod P, Wiener D, Zipperle L, Vandeveld M, Wittek R, Zurbriggen A (2007). Signal peptide and helical bundle domains of virulent canine distemper virus fusion protein restrict fusogenicity. *J Virol.* 81: 11413-11425.
- Prowse TA, Johnson CN, Lacy RC, Bradshaw CJ, Pollak JP, Watts MJ, Brook BW (2013). No need for disease: testing extinction hypotheses for the thylacine using multi-species metamodels. *J Anim Ecol.* 82: 355-364.
- Ramírez H, Reina R, Bertolotti L, Cenoz A, Hernández MM, San Román B, Glaría I, de Andrés X, Crespo H, Jáuregui P, Benavides J, Polledo L, Pérez V, García-

- Marín JF, Rosati S, Amorena B, de Andrés D (2012). Study of compartmentalization in the visna clinical form of small ruminant lentivirus infection in sheep. *BMC Vet Res.*, 26: 8-18.
- Roelke-Parker ME, Munson L, Packer C, Kock R, Cleaveland S, Carpenter M, O'Brien SJ, Pospischil A, Hofmann-Lehmann R, Lutz H, Mwamengele GL, Mgasa MN, Machange GA, Summers BA, Appel MJ (1996). A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature*. 379: 441-445.
  - Sakai K, Nagata N, Ami Y, Seki F, Suzaki Y (2013). Lethal canine distemper virus outbreak in cynomolgus monkeys in Japan in 2008. *J Virol.* 87: 1105-1114.
  - Salemi M, Lamers SL, Yu S, de Oliveira T, Fitch WM, McGrath MS (2005). Phylodynamic analysis of human immunodeficiency virus type 1 in distinct brain compartments provides a model for the neuropathogenesis of AIDS. *J Virol.* 79: 11343-11352.
  - Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R (2010). Viral mutation rates. *J Virol.*, 84: 9733-9748.
  - Sardanyés J, Solé RV, Elena SF (2009). Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. *J Virol.*, 83: 12579-12589.
  - Sarute N, Pérez R, Francia L, Hernández M, Bedó G, Bonilla B, Guasco S, Cardeillac A, Panzera Y (2011). Primer diagnóstico molecular y caracterización parcial del gen de la nucleoproteína del Virus Distemper Canino en Uruguay. *Veterinaria*, 47: 7-13.
  - Sarute N, Calderon MG, Perez R, La Torre J, Hernandez M, Francia L, Panzera Y (2013). The Fusion Protein Signal-Peptide-Coding Region of Canine Distemper

Virus: A Useful Tool for Phylogenetic Reconstruction and Lineage Identification. PLoS ONE 8(5): e63595.

- Sarute N, Delgado MV, Carrau L, Benech A, Francia L, Pérez R, Panzera Y (2014). First genome sequence of a canine distemper virus strain from South America. *Genome Announcements*, 2(5): e01009-14.
- Sauder CJ, Vandeburgh KM, Iskow RC, Malik T, Carbone KM, Rubin SA (2006). Changes in mumps virus neurovirulence phenotype associated with quasispecies heterogeneity. *Virology*. 350: 48-57.
- Sawatsky B, von Messling V (2010). Canine distemper viruses expressing a hemagglutinin without N-glycans lose virulence but retain immunosuppression. *J Virol*. 84: 2753-2761.
- Schmidt HA, von Haeseler A (2007). Maximum-likelihood analysis using TREE-PUZZLE. *Curr Protoc Bioinformatics*. Capítulo 6, Unidad 6.6.
- Seki F, Ono N, Yamaguchi R, Yanagi Y (2003). Efficient isolation of wild strains of canine distemper virus in Vero cells expressing canine SLAM (CD150) and their adaptability to marmoset B95a cells. *J Virol*. 77: 9943-9950.
- Simone-Loriere E, Holmes EC (2011). Why do RNA viruses recombine? *Nat Rev Microbiol.*, 9: 617-626.
- Smith EC, Blanc H, Vignuzzi M, Denison MR (2013). Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.*, 9(8):e1003565.
- Stanton JB, Brown CC, Poet S, Lipscomb TP, Saliki J, Frasca S Jr (2004). Retrospective differentiation of canine distemper virus and phocine distemper virus in phocids. *J Wildl Dis*. 40: 53-59.

- Stern LB, Greenberg M, Gershoni JM, Rozenblatt S (1995). The hemagglutinin envelope protein of canine distemper virus (CDV) confers cell tropism as illustrated by CDV and measles virus complementation analysis. *J Virol.* 69: 1661-1668.
- Sultan S, Lan NT, Ueda T, Yamaguchi R, Maeda K, Kai K (2009). Propagation of Asian isolates of canine distemper virus (CDV) in hamster cell lines. *Acta Vet Scand.*, 51:38.
- Sun Z, Li A, Ye H, Shi Y, Hu Z, Zeng L (2010). Natural infection with canine distemper virus in hand-feeding Rhesus monkeys in China. *Vet Microbiol.* 24: 374-378
- Svitek N, Gerhauser I, Goncalves C, Grabski E, Döring M, Kalinke U, Anderson DE, Cattaneo R, von Messling V (2004). Morbillivirus control of the interferon response: relevance of STAT2 and mda5 but not STAT1 for canine distemper virus virulence in ferrets. *J. Virol.*, 88: 2941-2950.
- van de Bildt MW, Kuiken T, Visee AM, Lema S, Fitzjohn TR, Osterhaus AD (2002). Distemper outbreak and its effect on African wild dog conservation. *Emerg Infect Dis.*,8: 211-213.
- Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, 439: 344-348.
- Vignuzzi M, Andino R (2012). Closing the gap: the challenges in converging theoretical, computational, experimental and real-life studies in virus evolution. *Curr Opin Virol.*, 2: 515-518.
- von Messling V, Zimmer G, Herrler G, Haas L, Cattaneo R (2001). The hemagglutinin of canine distemper virus determines tropism and cytopathogenicity. *J Virol.* 75: 6418-6427.

- von Messling V, Cattaneo R (2002). Amino-terminal precursor sequence modulates canine distemper virus fusion protein function. *J Virol.* 76: 4172-4180.
- von Messling V, Springfield C, Devaux P, Cattaneo R (2003). A Ferret Model of Canine Distemper Virus Virulence and Immunosuppression. *J Virol.* 77: 12579-12591.
- von Messling V, Milosevic D, Devaux P, Cattaneo R (2004). Canine distemper virus and measles virus fusion glycoprotein trimers: partial membrane-proximal ectodomain cleavage enhances function. *J Virol.* 78: 7894-7903.
- von Messling V, Oezguen N, Zheng Q, Vongpunsawad S, Braun W, Cattaneo R (2005). Nearby Clusters of Hemagglutinin Residues Sustain SLAM-Dependent Canine Distemper Virus Entry in Peripheral Blood Mononuclear Cells. *J Virol.* 79: 5857-5862.
- Watson JD (2006). *Biología Molecular del Gen. Mutabilidad y reparación del DNA.* Capítulo 9, 5ta Ed, Editorial Médica Panamericana.
- Woma TY, van Vuuren M, Bosman AM, Quan M, Oosthuizen M (2009). Phylogenetic analysis of the haemagglutinin gene of current wild-type canine distemper viruses from South Africa: lineage Africa. *Vet. Microbiol.*, 143: 126-132.
- Wright CF, Morelli MJ, Thébaud G, Knowles NJ, Herzyk P, Paton DJ, Haydon DT, King DP (2011). Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using nextgeneration genome sequencing. *J. Virol.*, 85: 2266 –2275.
- Wyss-Fluehmann G, Zurbriggen A, Vandeveld M, Plattet P (2010). Canine distemper virus persistence in demyelinating encephalitis by swift intracellular cell-to-cell spread in astrocytes is controlled by the viral attachment protein. *Acta Neuropathol.* 119: 617-630.

- Zhang X, Rennick LJ, Duprex WP, Rima BK (2013). Determination of spontaneous mutation frequencies in measles virus under nonselective conditions. *J Virol.*, 87: 2686-2692.
- Zipperle L, Langedijk JP, Orvell C, Vandeveldelde M, Zurbriggen A, Plattet P (2010). Identification of key residues in virulent canine distemper virus hemagglutinin that control CD150/SLAM-binding activity. *J Virol.* 84: 9618-9624.
- Zurbriggen A, Graber HU, Vandeveldelde M (1995). Selective spread and reduced virus release leads to canine distemper virus persistence in the nervous system. *Vet Microbiol.*, 44: 281-288.

## Anexo II: Publicaciones

En el presente anexo se detallan los artículos publicados durante el desarrollo de la Tesis (adjuntos). Para los cinco artículos se presenta el resumen y su vinculación con la tesis doctoral

**1. Sarute N, Calderón MG, Pérez R, La Torre J, Hernández M, Francia L, Panzera Y (2013).** The Fusion Protein Signal-Peptide-Coding Region of Canine Distemper Virus: Useful Tool for Phylogenetic Reconstruction and Lineage Identification. PLoS ONE 8(5): e63595.

Resumen: se analizaron y compararon secuencias del gen de la hemaglutinina (H) y de la región codificante del péptido señal de la proteína de fusión (Fsp) provenientes de las mismas cepas pertenecientes a ocho linajes genéticos de CDV, con el fin de evaluar la robustez de la región Fsp para reconstrucciones filogenéticas e identificación de linajes. Nuestros análisis revelaron que el péptido Fsp presenta una variabilidad aminoacídica cuatro veces mayor que la hemaglutinina, establece las mismas relaciones evolutivas que el gen H y presenta una elevada señal filogenética, haciendo de la región Fsp un excelente marcador para análisis evolutivos de CDV.

**2. Sarute N, Pérez R, Aldaz J, Alfieri AA, Alfieri AF, Name D, Llanes J, Hernández M, Francia L, Panzera Y (2014).** Molecular typing of canine distemper virus strains reveals the presence of a new genetic lineage in South America. Virus Genes, 48: 474-478.

Resumen: se amplificó y analizó la región Fsp de cepas provenientes de Uruguay, Brasil y Ecuador, y se comparó con otras cepas descritas en el mundo. Dicho análisis nos permitió caracterizar por primera vez las cepas circulantes en Ecuador, evidenciando la presencia de un nuevo linaje circulante en Sudamérica (SA3), exclusivo de este país y diferente a los presentes en el resto de los países analizados: Uruguay (EU1/SA1), Brasil (EU1/SA1) y Argentina (EU1/SA1 y SA2).

Vinculación: Las diferentes regiones genómicas obtenidas durante el desarrollo de la tesis fueron analizadas para determinar su señal filogenética por *likelihood mapping*, comenzando por los genes de las glicoproteínas de superficie H y F. Este abordaje nos permitió evaluar por primera vez la robustez del gen H para análisis filogenéticos, el

cual es ampliamente utilizado en estudios filogenéticos y evolutivos de CDV. Sin embargo, este gen es difícil de amplificar directamente a partir de muestras biológicas. Detectamos que una corta región de 405 pb del gen F correspondiente al péptido señal (Fsp) presentó una señal filogenética semejante al gen H, lo cual indica que tiene una excelente resolución para análisis filogenéticos. La región codificante del Fsp presenta una ventaja sobre el gen H, ya que es más sencillo de amplificar y secuenciar por su corta longitud, confirmando su utilidad como marcador para estudios de CDV.

En el artículo 1 demostramos por diversos abordajes la utilidad de este nuevo marcador para caracterización de cepas y análisis evolutivos, mientras que en el artículo 2 utilizamos este marcador para la caracterización de nuevas cepas provenientes de diversos países de Sudamérica, detectando la circulación de un nuevo linaje en nuestro continente.

**3. Sarute N, Delgado MV, Carrau L, Benech A, Francia L, Pérez R, Panzera Y (2014).** First genome sequence of a canine distemper virus strain from South America. *Genome Announcements*, 2(5): e01009-14.

Resumen: en este trabajo se presenta la secuencia genómica de la cepa (Uy251/2012) obtenida de una infección natural de un perro de Uruguay. Su análisis reveló que Uy251/2012 pertenece al linaje Europa1/Sudamérica1 y constituye el primer reporte de una secuencia genómica completa de CDV para Sudamérica.

Vinculación: la estandarización de los protocolos de PCR para la amplificación del genoma completo nos permitió obtener el genoma viral aislado de una única matriz (251P). La implementación de las técnicas moleculares, la secuenciación de los amplicones y su ensamblado para obtener la secuencia genómica, fue de relevancia para el desarrollo de la Tesis debido a que pudimos evaluar la eficiencia de los cebadores diseñados para la amplificación de los nueve fragmentos genómicos por RT-PCR empleados para la obtención del genoma completo en las matrices analizadas. Además, la secuencia genómica obtenida por Sanger se comparó con la secuencia obtenida por secuenciación masiva del mismo caso clínico luego de un pasaje por cultivo celular, confirmando que nuestra estrategia fue exitosa para el análisis de la diversidad

poblacional, ya que no detectamos cambios nucleotídicos significativos debido al pasaje en la línea celular.

4. Panzera Y, **Sarute N**, Carrau L, Aldaz J, Pérez R (2014). Genetic Diversity of Canine Distemper Virus in South America (Review). *British Journal of Virology*, 1(2): 48 – 53.

Resumen: en este trabajo se presenta información actualizada sobre la dinámica de CDV en Sudamérica, mediante un análisis detallado de la diversidad genética, evolución y epidemiología del virus en varios países. Nuestro estudio describe los linajes circulantes y su distribución en Sudamérica, siendo el continente con más linajes descritos a la fecha y destacando la elevada diversidad genética de las cepas circulantes.

Vinculación: nuestros antecedentes con el virus por más de ocho años, llevo a que nuestro grupo sea invitado para resumir la información obtenida para el virus a la fecha en varios países de Sudamérica. En esta revisión no solo se discute información obtenida durante el desarrollo de esta Tesis doctoral, sino también a lo largo de toda mi formación académica acerca de aspectos genéticos y evolutivos de CDV.

5. Pérez R, Calleros L, Marandino A, **Sarute N**, Iraola G, Grecco S, Blanc H, Vignuzzi M, Isakov O, Shomron N, Carrau L, Hernández M, Francia L, Sosa K, Tomás G, Panzera Y (2014). Phylogenetic and Genome-Wide Deep-Sequencing Analyses of Canine Parvovirus Reveal Co-Infection with Field Variants and Emergence of a Recent Recombinant Strain. *PLoS ONE*, 9(11): e111779

Resumen: en este trabajo se analizó la ocurrencia de eventos de co-infección y recombinación en una población natural de Parvovirus canino (CPV) en Uruguay formada por variantes altamente divergentes (CPV-2a y CPV-2c). Mediante la secuenciación masiva del genoma completo de las variantes circulantes, se detectaron dos casos de co-infección (364 y 370). El análisis del caso 370 reveló la existencia de una co-infección de variantes 2a y 2c en proporciones similares, mientras que el caso 364 resultó ser una co-infección de una variante 2a en una proporción del 13,3%, y una cepa recombinante 2a-2c, en una proporción del 86,7%. Nuestro estudio destaca la importancia de la secuenciación masiva para analizar en profundidad la composición nucleotídica de las poblaciones virales, y para comprender en mayor medida los mecanismos que rigen la diversidad molecular de CPV.

Vinculación: la formación recibida en protocolos de secuenciación masiva y análisis de datos durante el desarrollo de mi tesis doctoral, permitió ampliar nuestros estudios a otros modelos virales en los cuales trabaja nuestro grupo de investigación en la Sección Genética Evolutiva. La preparación de las muestras, la secuenciación masiva y el análisis de datos se realizó paralelamente con las muestras de CDV de los tres casos clínicos incluidos en la presente Tesis, durante una pasantía en el laboratorio del Dr. Vignuzzi en el Institut Pasteur Paris.

Actualmente, estamos redactando un manuscrito que incluirá los datos presentados en esta Tesis y el análisis de las secuencias obtenidas de la tercera corrida de secuenciación realizada en el Instituto Pasteur Montevideo.

Sarute N, Vignuzzi M, Iraola G, Delgado MV, Pérez R, Benech A, Francia L, Blanc H, Isakov O, Shomron N, Panzera Y (2014). A comprehensive analysis of *in vivo* population dynamics of canine distemper virus by genome-wide ultra deep sequencing.

Resumen: el presente trabajo se comenzó a redactar con los datos de secuenciación masiva obtenidos a partir de las once matrices de los tres casos clínicos, destacándose la organización de CDV en cuasiespecies, la heterogeneidad poblacional detectada en las matrices, la identificación de variantes en una frecuencia significativa dentro de la población y los cambios aminoacídicos asociados, y la frecuencia de mutación para cada gen/región genómica. Además, se incluye el análisis de compartimentalización y migración mediante el análisis de las variantes genéticas presentes en cada caso clínico, y la estimación del curso patogénico de la infección para cada caso mediante análisis filodinámicos.

Significancia: este estudio establece la organización de CDV en cuasiespecies, lo cual constituye el primer registro para el género *Morbillivirus* y para la familia *Paramyxoviridae* utilizando secuenciación masiva y analizando distintas matrices de tres casos clínicos, las cuales presentaron una diversidad de variantes característica de cada matriz (subpoblaciones). En las muestras de cerebelo de los casos 245 y 251 se identificó una sustitución aminoacídica en la hemaglutinina (Y549H) en la zona de unión al receptor celular SLAM, que podría estar involucrada en el neurotropismo. Los análisis de compartimentalización y migración revelaron la diversidad poblacional de

las subpoblaciones y permitieron estimar el curso de la infección. En los casos 245 y 251 el virus se diseminó sistémicamente y una vez alcanzado el SNC se registraron eventos de migración posteriores ocasionados posiblemente por la inmunidad generada, mientras que en el caso 246 se observó un patrón de infección crónica a nivel del SNC. En base a estos resultados, sugerimos que el caso 246 corresponde a una cepa neurovirulenta.



### **Anexo III: Secuenciación masiva en el Instituto Pasteur Montevideo**

En la tercera corrida de secuenciación se amplificaron nuevamente los 28 genes/fragmentos genómicos que no presentaron la cobertura mínima (100X) en las corridas anteriores, y se incluyeron 26 amplicones adicionales con el objetivo de mejorar los datos de secuenciación obtenidos previamente. El aislamiento del RNA se realizó a partir de sobrenadantes de cultivo almacenados a -80 °C y se siguió el protocolo de RT-PCR para la amplificación del genoma completo.

En esta corrida se utilizó un nuevo kit para la preparación de las muestras Nextera DNA Sample Preparation Kit (Illumina), y la corrida de secuenciación se realizó en el aparato MiSeq de Illumina.

El kit Nextera requiere de una concentración de DNA inicial (*input*) de 1ng/μl. Por tanto, los productos de PCR purificados fueron cuantificados mediante el fluorómetro Qubit 2.0 utilizando el kit Labeling detection Qubit dsDNA HS Assay kit (Life Technologies), el cual incorpora un agente intercalante fluorescente de unión específica a DNA doble hebra. El equipo Qubit 2.0 es calibrado con dos estándares de concentración, y para la medición de cada muestra se utilizó 1μl de DNA y el intercalante en una dilución 1/200. El DNA de todos los amplicones utilizados en la preparación de las librerías se diluyó a una concentración de 0,2 ng/μl y se emplearon 5μl como DNA *input*.

Protocolo del kit Nextera DNA Sample Preparation Kit

#### **1. Tagmentación**

En este paso el transposón Nextera XT fragmenta y añade adaptadores en los extremos del DNA *input* simultáneamente.

- Adicionar 10 ul de *buffer* TD (Tagment DNA Buffer) a 5μl del DNA *input*
- Agregar 5 ul de *buffer* ATM (Amplicon Tagment Mix)
- Incubar la mezcla a 55 °C por 5 minutos (min), y luego incubar en hielo
- Agregar 5 ul del *buffer* NT (Neutralize Tagment Buffer) e incubar por 5 min a temperatura ambiente (TA)

## 2. Amplificación por PCR

En este paso se agregaron los cebadores (*index*) para la formación de *clusters* del DNA fragmentado. A cada tubo con la mezcla de la tagmentación, se le añadieron 15 µl del *buffer* NPM (Nextera PCR Master Mix), 5 µl de un cebador directo y 5 µl de un cebador reverso.

La reacción de PCR se realizó bajo las siguientes condiciones de ciclado:

72 °C por 3 min

95°C por 30 segundos (seg)

95°C por 10 seg

55 °C por 30 seg

72 °C por 30 seg

72 °C por 5 min

} 12 ciclos

## 3. Purificación de productos de PCR

Este paso se realizó con el sistema Agencourt AMPure XP (Beckman-Coulter) el cual está basado en cuentas magnéticas (*beads*)

- Adicionar a cada tubo 30 µl de *beads* y homogeneizar con pipeta
- Incubar la mezcla 5 min a TA
- Transferir la mezcla a tubos de 1,5 ml y colocarlos en el MagnaRack Magnetic Separation Rack (Invitrogen)
- Incubar por 2 min hasta que la mezcla clarifique
- Descartar el sobrenadante y adicionar 200 µl de etanol 80%
- Incubar por 30 seg y remover el sobrenadante
- Repetir el paso de lavado con etanol 80%
- Secar los tubos a TA por 15 min
- Retirar los tubos del MagnaRack y agregar 52,5 µl del *buffer* RSB (Resuspension Buffer)
- Incubar a TA por 2 min
- Colocar los tubos en el MagnaRack por 2 min y transferir 50 µl de cada librería purificada a un nuevo tubo
- Incubar a -20 °C

#### 4. Cuantificación y control de calidad y tamaño de las librerías

El análisis de calidad y cuantificación de las librerías purificadas se realizó con el aparato Bioanalyzer 2100 (Agilent Technologies). El análisis con el Bioanalyzer reveló que las librerías tuvieron una concentración en el rango de 0,8 – 5,2 ng/μl y el tamaño de los fragmentos se ubicó entre 529 -605 pb.

#### 5. Normalización de librerías

Este proceso normaliza la cantidad de cada librería para asegurarse una igual representación al momento de la corrida

- Realizar una mezcla con 550 μl del reagente LNA1 (Library Normalization Additives 1) y 100 μl de LNB1 (Library Normalization Beads 1) para normalizar 12 librerías
- Preparar tubos de 1,5 ml con 45 μl de la mezcla LNA1/LNB1 con 20 μl de cada librería
- Agitar en *shaker* por 30 min a 2000 rpm
- Colocar los tubos en el MagnaRack por 2 min y descartar el sobrenadante
- Remover los tubos del MagnaRack y realizar un lavado con 45 μl del *buffer* LNW1 (Library Normalization Wash 1)
- Agitar en *shaker* por 5 min a 2000 rpm
- Repetir el paso de lavado y agitación
- Colocar los tubos en el MagnaRack e incubar por 2 min. Descartar el sobrenadante
- Remover los tubos del MagnaRack y añadir 30 μl de 0,1 N NaOH
- Agitar en *shaker* por 5 min a 2000 rpm
- Colocar los tubos en el MagnaRack e incubar por 2 min
- Transferir 30 μl del sobrenadante a un nuevo tubo de 1,5 ml con 30 μl del *buffer* LNS1 (Library Normalization Storage Buffer 1)
- Incubar a -20 °C

#### 6. Generación de *clusters* y secuenciación en el aparato MiSeq

En este paso se combinaron volúmenes iguales de las librerías normalizadas, se diluyeron en un *buffer* de hibridación y se desnaturalizaron por calor para ser cargadas

en el cartucho de secuenciación del aparato MiSeq. Esta parte del protocolo fue realizada por el técnico de secuenciación Mag. Gonzalo Greif en el Instituto Pasteur Montevideo.

### **7. Obtención de secuencias nucleotídicas**

Las secuencias correspondientes a cada matriz fueron discriminadas por la combinación de cebadores (*index*) utilizados en la corrida. Los datos de la corrida se analizaron mediante el programa *MiSeq Reporter* integrado al secuenciador MiSeq, y se obtuvieron las secuencias que superaron los filtros de calidad y cobertura del aparato en formato fastq.

### **8. Análisis de secuencias mediante el programa ViVan**

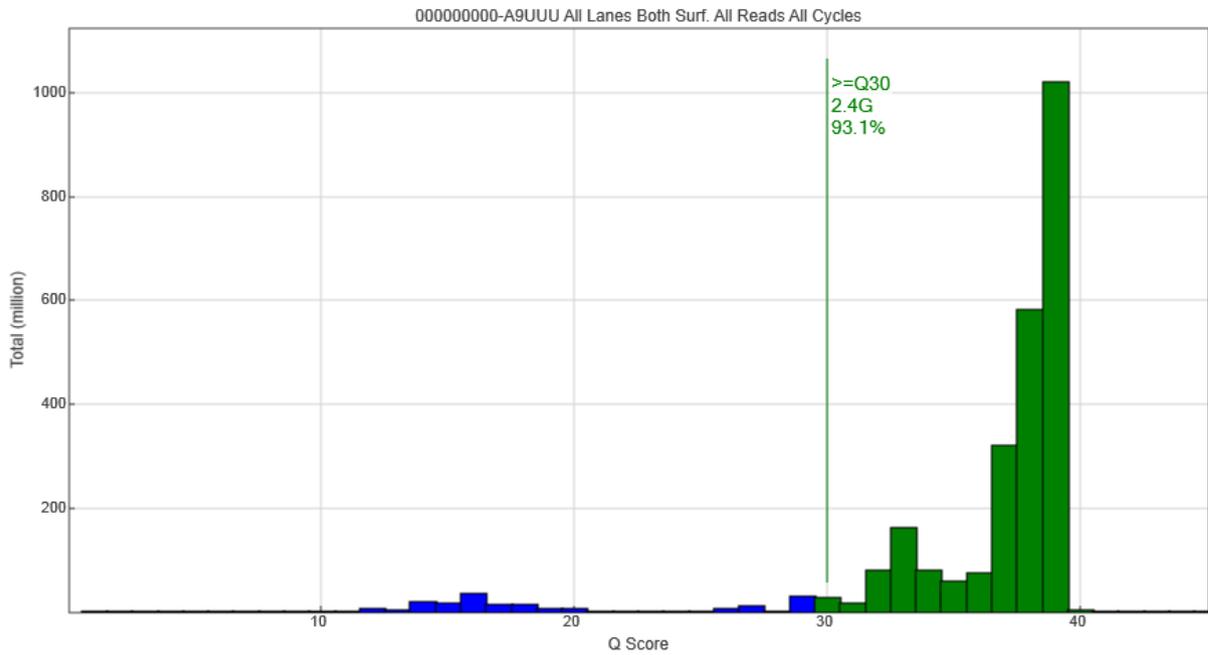
Las secuencias obtenidas de esta tercera corrida están siendo analizadas mediante el ViVan para obtener los datos de diversidad poblacional de los amplicones secuenciados para ser integrados a nuestros análisis.

CORRIDA Paired End 2 x 150

Nicolás Sarute

12 librerías con el kit Nextera XT

13 y 14 de Noviembre de 2014



Illumina Sequencing Analysis Viewer 1.8.20 - 141029\_M02189\_0020\_000000000-A9UUU

## Sequencing Analysis Viewer

Run Folder: D:\Illumina\MiSeqOutput\141029\_M02189\_0020\_000000000-A9UUU

Browse

Refresh

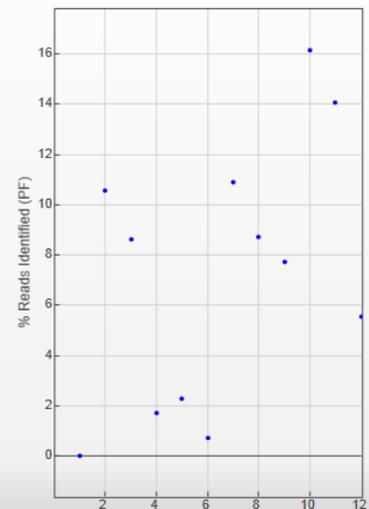
Analysis Imaging Summary Tile Status TruSeq Controls Indexing

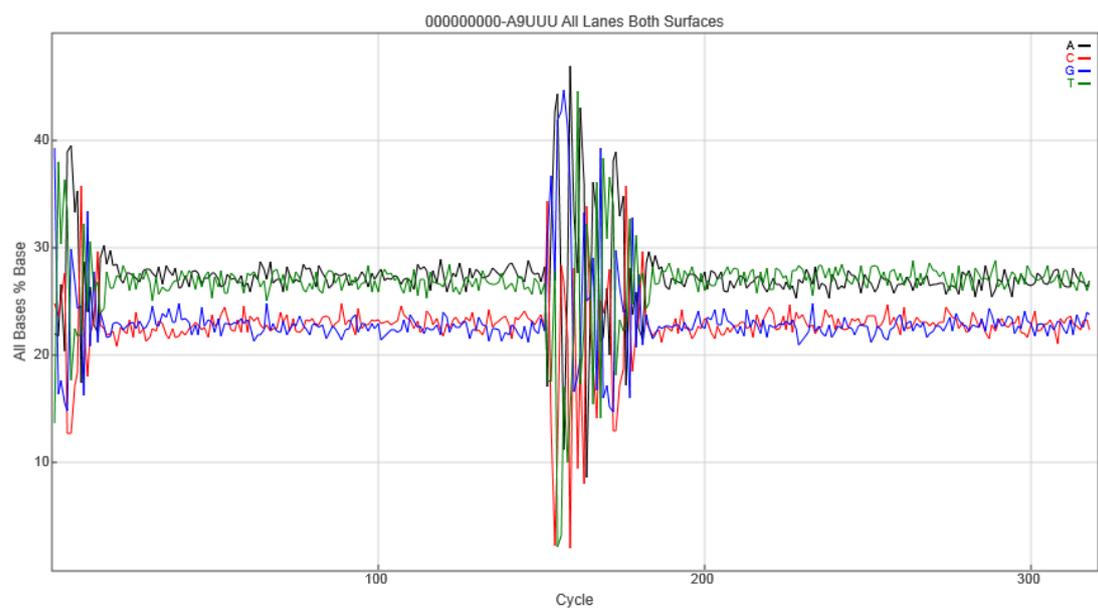
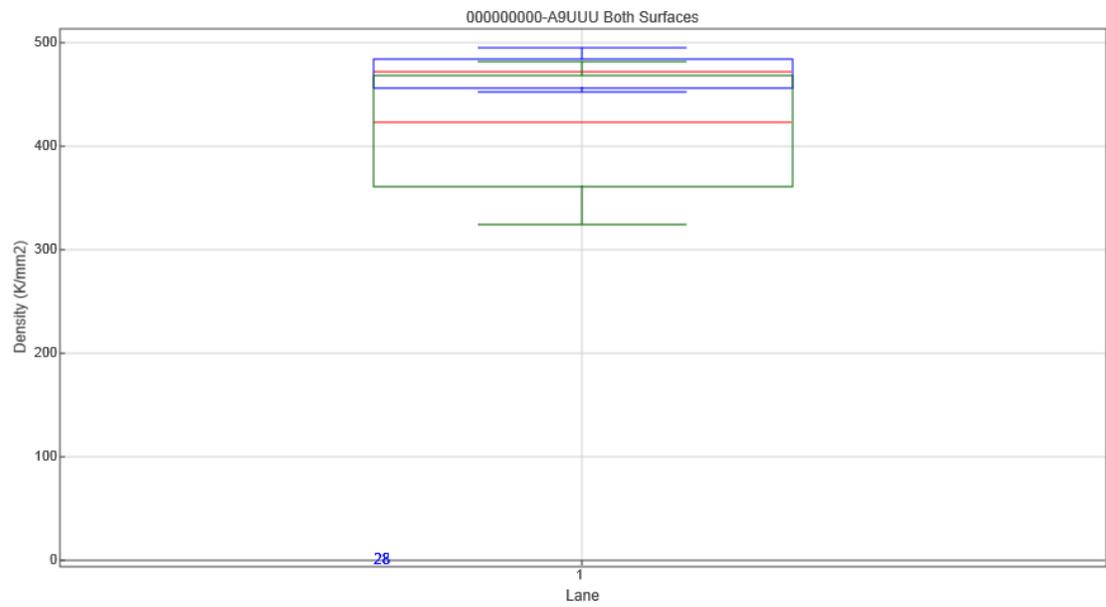
Lane 1

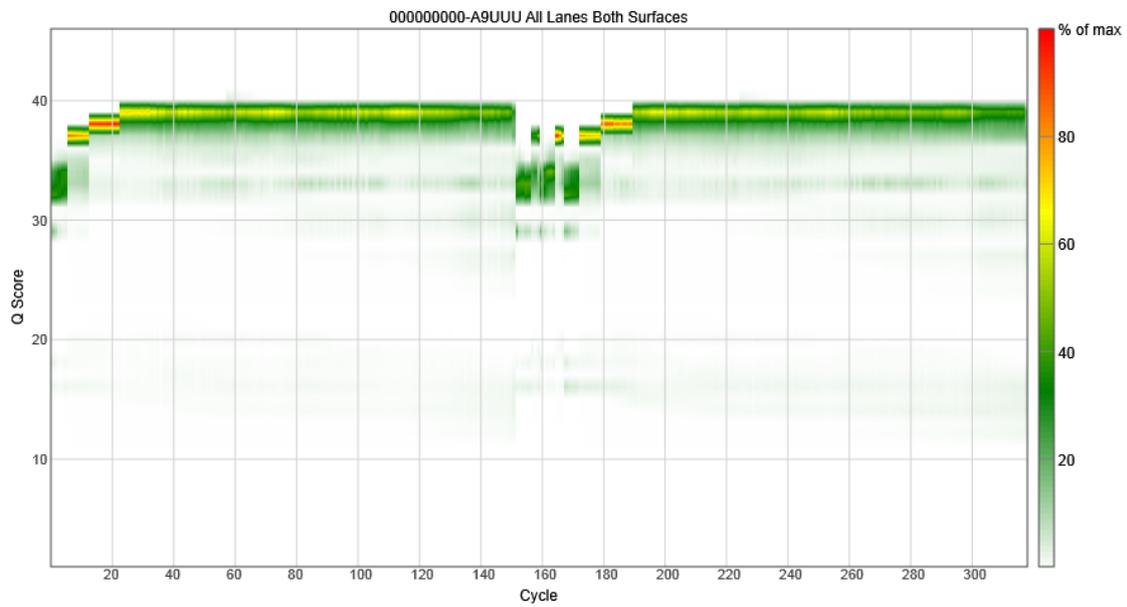
Reads mapped to Index Id

Total Reads	PF Reads	% Reads Identified (PF)	CV	Min	Max
9372124	8204738	87.0497	0.7294	0.001	16.1543

Index Number	Sample Id	Project	Index 1 (I7)	Index 2 (I5)	% Reads Identified (PF)
1	245C	YPanzer	CAGAGAGG	TATCCTCT	0.001
2	245P	YPanzer	CGTACTAG	CTCTCTAT	10.5737
3	245M	YPanzer	AGGCAGAA	AAGGAGTA	8.6214
4	245V	YPanzer	TCCTGAGC	AGAGTAGA	1.7012
5	246C	YPanzer	GGA CTCT	GTAAGGAG	2.2873
6	246P	YPanzer	TAGGCATG	ACTGCATA	0.7261
7	246M	YPanzer	GCTACGCT	TAGATCGC	10.9078
8	246O	YPanzer	CGAGGCTG	CTCTCTAT	8.7275
9	251C	YPanzer	AAGAGGCA	TATCCTCT	7.7041
10	251P	YPanzer	GTAGAGGA	AGAGTAGA	16.1543
11	251M	YPanzer	TAAGGCGA	GTAAGGAG	14.0762
12	251O	YPanzer	CGTACTAG	ACTGCATA	5.5691



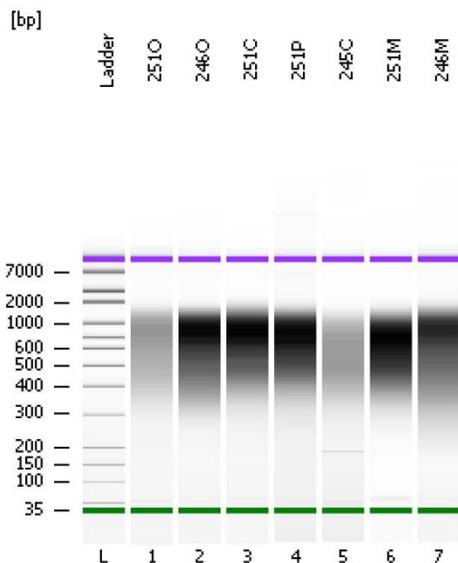




Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

**Electrophoresis File Run Summary**



Instrument Information:

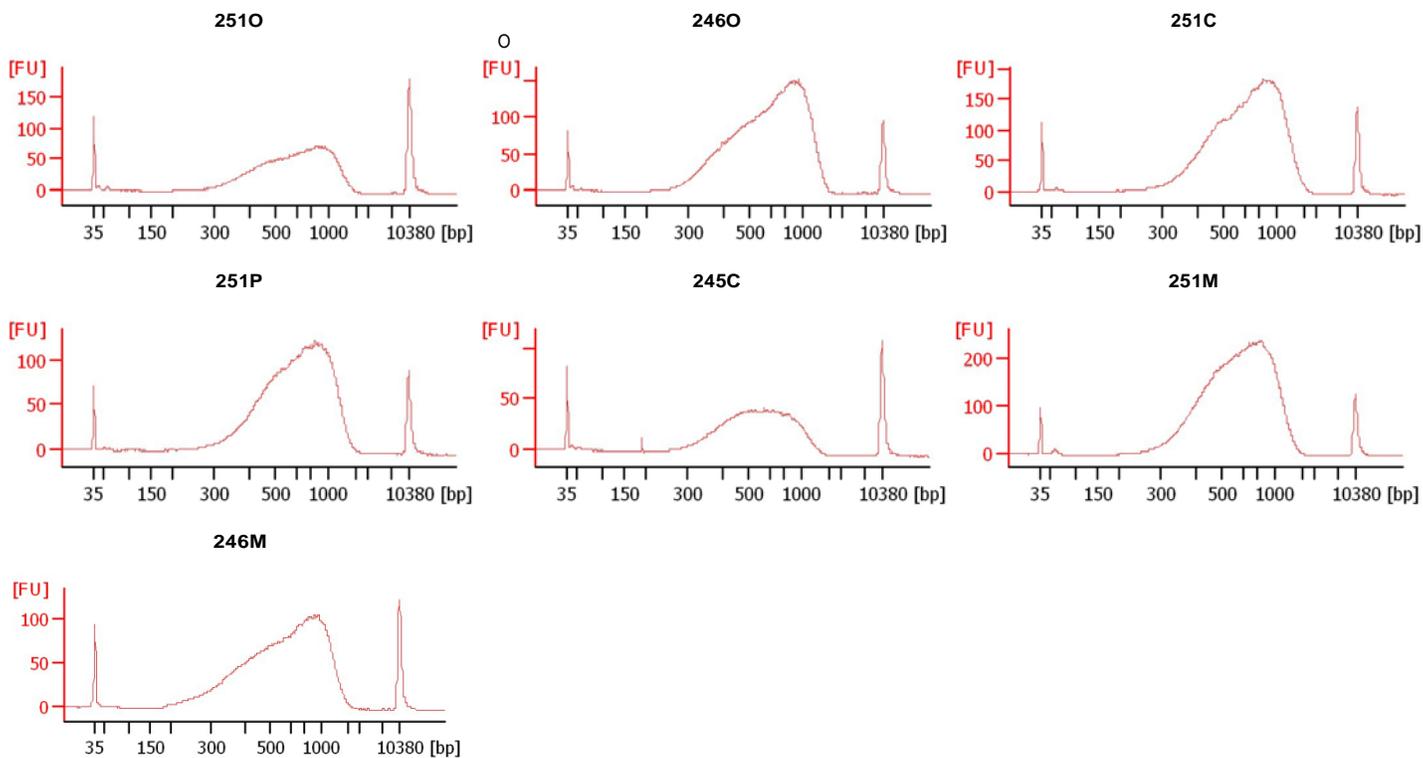
Instrument Name: DE72904909      Firmware: C.01.069  
 Serial#: DE72904909      Type: G2938C

Assay Information:

Assay Class: High Sensitivity DNA Assay  
 Version: 1.03  
 Assay Comments: Copyright © 2003-2010 Agilent Technologies

Chip Information:

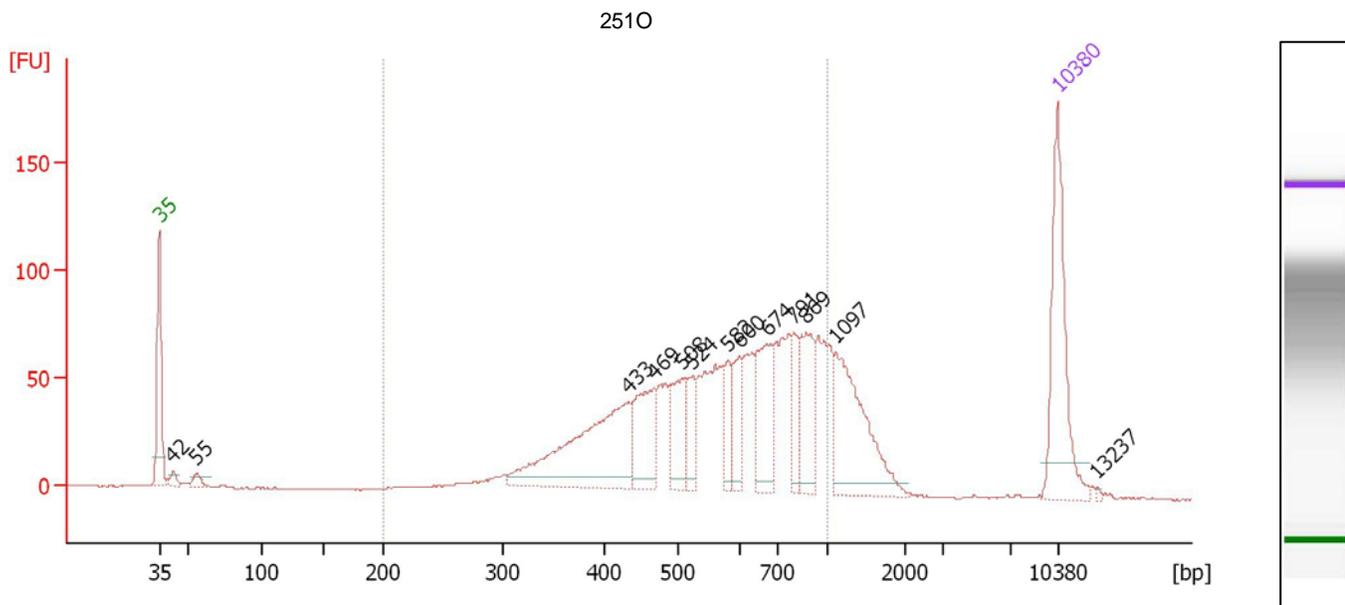
Chip Lot #:  
 Reagent Kit Lot #:  
 Chip Comments:



Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

**Electropherogram Summary**



**Overall Results for sample 1 : 251O**

Number of peaks found: 13                      Corr. Area 1: 1,526.6  
 Noise: 0.2

**Peak table for sample 1 : 251O**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	42	7.91	286.3	
3	55	9.31	254.9	
4	433	148.40	519.8	
5	469	58.37	188.7	
6	508	43.19	128.9	
7	524	27.12	78.4	
8	582	23.58	61.3	
9	600	28.36	71.6	
10	674	57.60	129.5	
11	791	25.38	48.6	
12	869	48.08	83.9	
13	1,097	88.44	122.2	
14	10,380	75.00	10.9	Upper Marker
15	13,237	0.00	0.0	

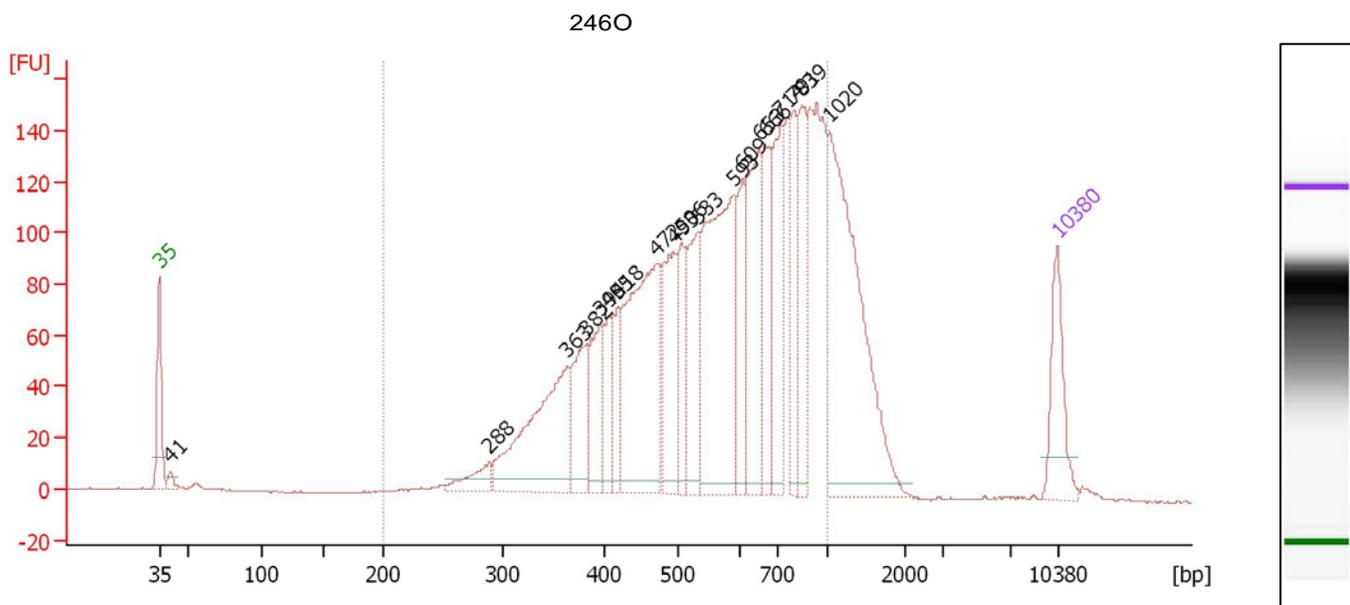
**Region table for sample 1 : 251O**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	1,526.6	83	587	30.6	782.27	2,323.6	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 2 : **246O**

Number of peaks found: 19                      Corr. Area 1: 2,927.5  
 Noise: 0.3

Peak table for sample 2 : **246O**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	41	15.71	586.4	
3	288	40.22	211.5	
4	363	271.69	1,135.0	
5	382	113.75	451.0	
6	395	101.94	391.3	
7	405	61.03	228.1	
8	418	72.42	262.3	
9	472	345.87	1,111.1	
10	495	155.56	476.1	
11	506	80.19	239.9	
12	533	130.32	370.6	
13	593	353.85	904.5	
14	609	105.27	261.7	
15	653	187.12	434.5	
16	666	126.69	288.3	
17	714	137.78	292.3	
18	791	103.00	197.4	
19	839	122.51	221.2	
20	1,020	388.21	576.6	
21	10,380	75.00	10.9	Upper Marker

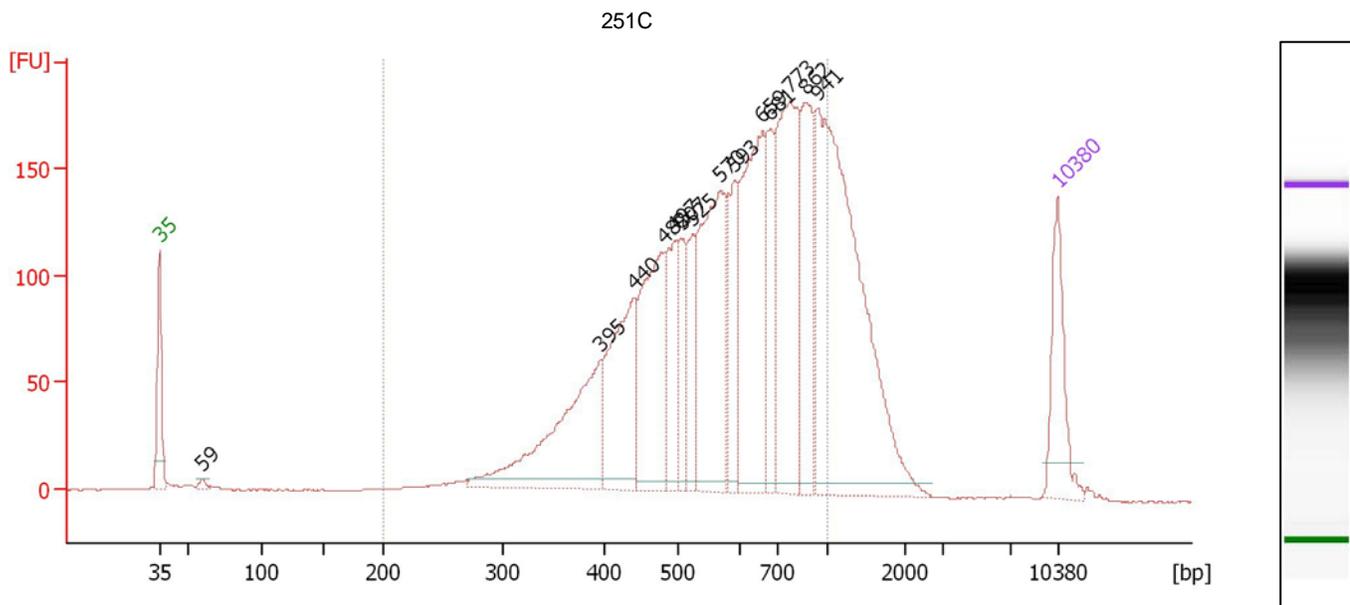
Region table for sample 2 : **246O**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	2,927.5	85	590	30.9	2,945.23	8,713.3	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 3 : **251C**

Number of peaks found: 14      Corr. Area 1: 3,321.4  
 Noise: 0.4

Peak table for sample 3 : **251C**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	59	8.62	220.3	
3	395	251.10	963.0	
4	440	200.50	689.8	
5	482	210.93	663.1	
6	497	99.35	303.0	
7	507	70.34	210.2	
8	525	74.04	213.5	
9	570	250.59	666.4	
10	593	80.69	206.2	
11	659	264.92	608.9	
12	681	101.52	226.0	
13	773	276.03	541.1	
14	862	142.17	249.8	
15	941	515.05	829.5	
16	10,380	75.00	10.9	Upper Marker

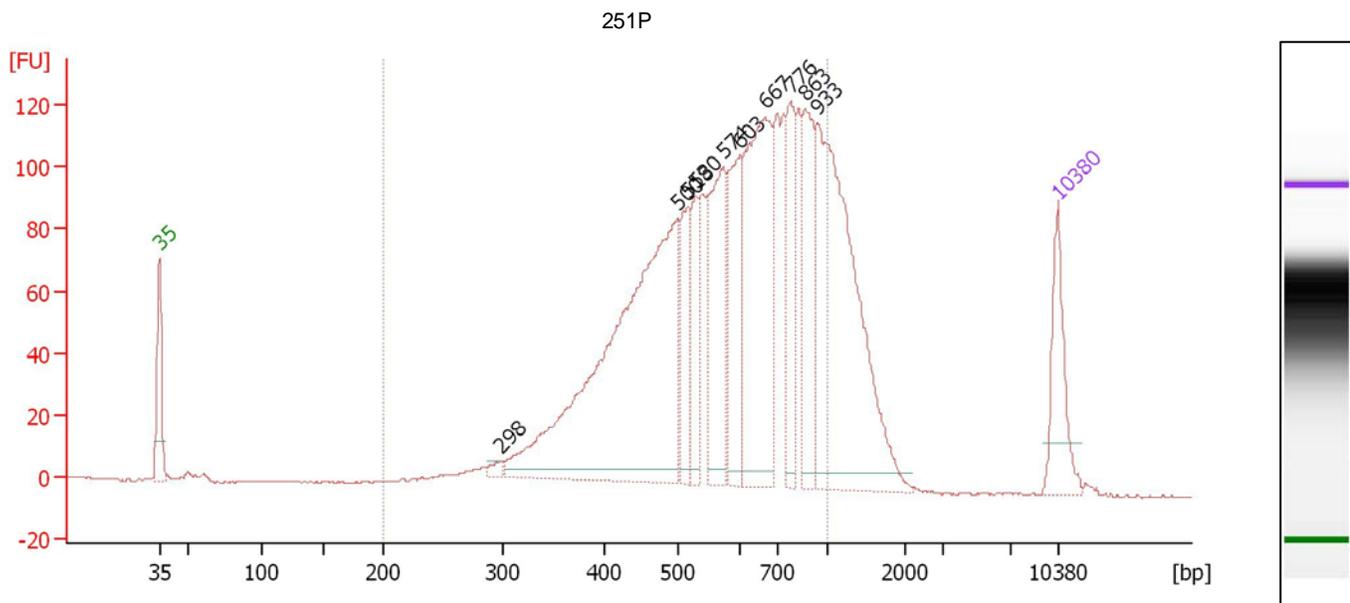
Region table for sample 3 : **251C**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	3,321.4	83	605	29.4	2,339.30	6,672.4	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 4 : **251P**

Number of peaks found: 10                      Corr. Area 1: 2,277.3  
 Noise: 0.3

Peak table for sample 4 : **251P**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	298	11.01	56.0	
3	500	755.80	2,289.8	
4	518	104.34	305.3	
5	530	83.17	237.9	
6	574	189.85	501.1	
7	603	139.05	349.6	
8	667	346.84	788.1	
9	776	103.66	202.5	
10	863	141.11	247.7	
11	933	457.67	743.6	
12	10,380	75.00	10.9	Upper Marker

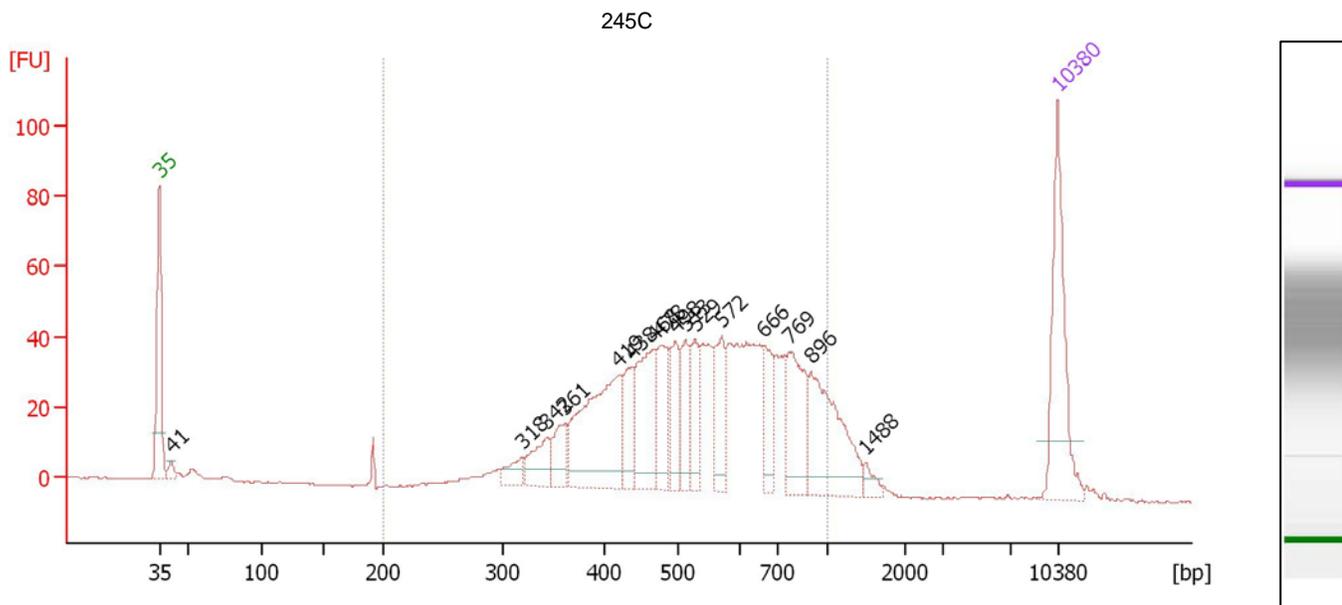
Region table for sample 4 : **251P**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	2,277.3	86	603	28.6	2,470.29	7,013.7	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 5 : **245C**

Number of peaks found: 16      Corr. Area 1: 982.6  
 Noise: 0.3

Peak table for sample 5 : **245C**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	41	10.87	401.6	
3	318	16.34	77.7	
4	342	31.95	141.4	
5	361	29.55	124.1	
6	419	142.95	516.4	
7	438	37.86	130.8	
8	468	79.93	258.8	
9	478	46.46	147.4	
10	498	34.17	104.1	
11	513	36.06	106.5	
12	529	35.42	101.4	
13	572	44.70	118.4	
14	666	37.45	85.2	
15	769	64.36	126.8	
16	896	93.48	158.1	
17	1,488	7.56	7.7	
18	10,380	75.00	10.9	Upper Marker

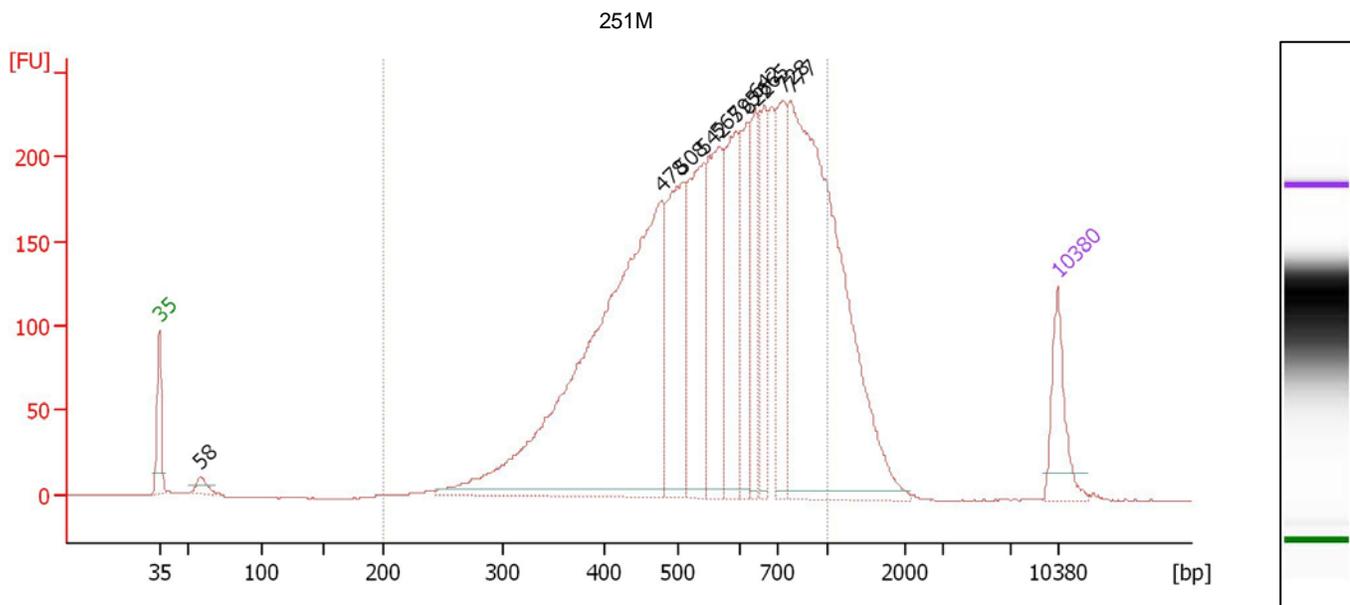
Region table for sample 5 : **245C**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	982.6	92	557	29.2	927.38	2,824.0	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 6 : **251M**

Number of peaks found: 11      Corr. Area 1: 4,724.4  
 Noise: 0.2

Peak table for sample 6 : **251M**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	58	28.41	740.6	
3	478	1,162.07	3,683.4	
4	508	296.88	884.8	
5	542	263.73	736.9	
6	567	260.06	695.1	
7	595	215.04	547.9	
8	622	165.96	404.0	
9	642	124.21	293.1	
10	665	135.92	309.8	
11	728	175.04	364.3	
12	777	918.98	1,790.9	
13	10,380	75.00	10.9	Upper Marker

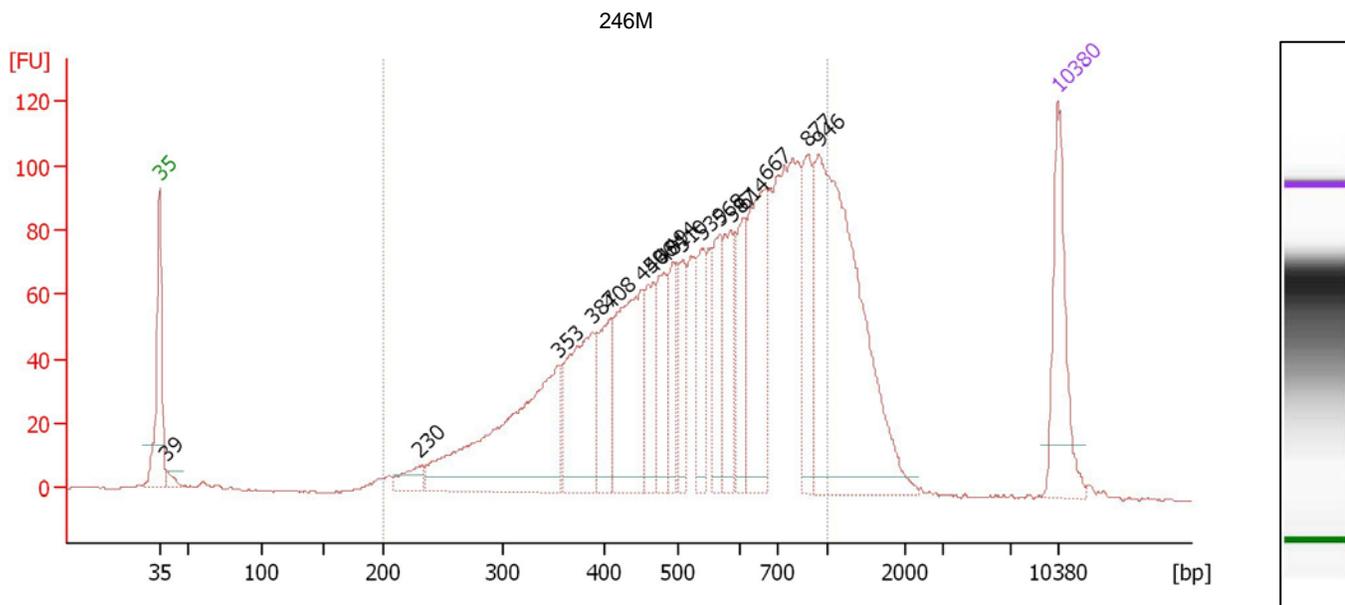
Region table for sample 6 : **251M**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	4,724.4	90	586	29.1	3,597.38	10,500.8	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-28\_08-56-09.xad

Created: 10/28/2014 8:56:08 AM  
 Modified: 10/28/2014 9:37:34 AM

Electropherogram Summary Continued ...



Overall Results for sample 7 : **246M**

Number of peaks found: 17      Corr. Area 1: 2,238.3  
 Noise: 0.2

Peak table for sample 7 : **246M**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	39	10.31	404.9	
3	230	24.25	160.0	
4	353	264.58	1,134.6	
5	387	136.80	535.8	
6	408	68.11	252.8	
7	450	156.39	526.4	
8	466	57.46	186.7	
9	481	62.83	197.7	
10	494	47.01	144.2	
11	510	54.04	160.7	
12	539	57.83	162.7	
13	568	61.81	165.0	
14	587	64.76	167.1	
15	614	66.31	163.5	
16	667	129.04	293.1	
17	877	74.54	128.8	
18	946	338.00	541.5	
19	10,380	75.00	10.9	Upper Marker

Region table for sample 7 : **246M**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	2,238.3	85	565	33.7	1,809.61	5,803.3	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
 Modified: 10/10/2014 3:59:44 PM

**Electrophoresis File\_Run Summary**

Instrument Information:

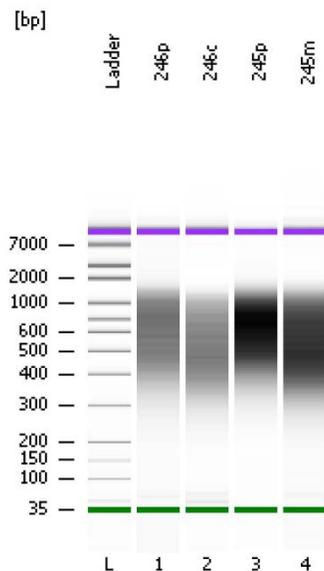
Instrument Name: DE72904909      Firmware: C.01.069  
 Serial#: DE72904909      Type: G2938C

Assay Information:

Assay Origin Path:  
 C:\Program Files\Agilent\2100 bioanalyzer\2100  
 expert\assays\dsDNA\High Sensitivity DNA.xsy  
 Assay Class: High Sensitivity DNA Assay  
 Version: 1.03  
 Assay Comments: Copyright © 2003-2010 Agilent Technologies

Chip Information:

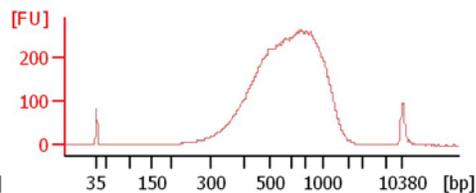
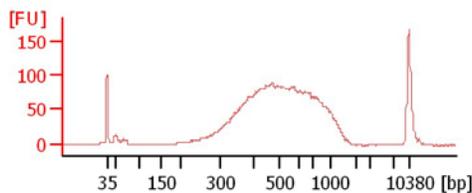
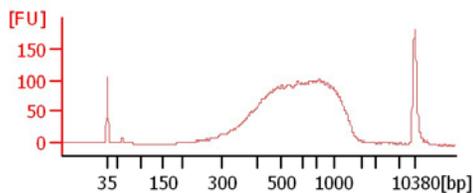
Chip Lot #:  
 Reagent Kit Lot #:  
 Chip Comments:



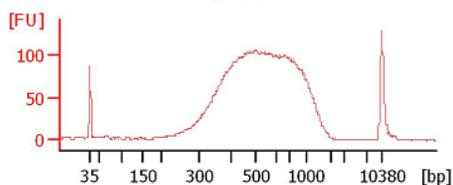
246p

246c

245p



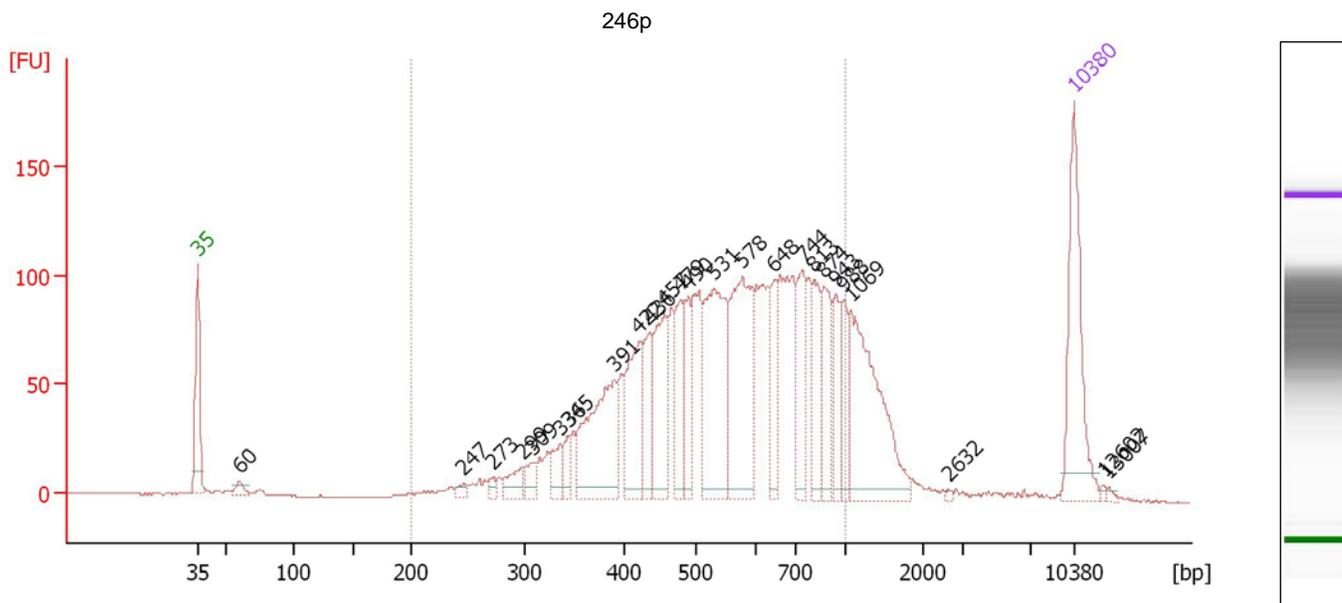
245m



Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
 Modified: 10/10/2014 3:59:44 PM

**Electropherogram Summary**



**Overall Results for sample 1 : 246p**

Number of peaks found: 25                      Corr. Area 1: 2,459.1  
 Noise: 0.2

**Peak table for sample 1 : 246p**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	60	9.54	241.7	
3	247	5.55	34.1	
4	273	6.30	34.9	
5	299	19.44	98.6	
6	309	15.53	76.1	
7	336	20.45	92.2	
8	345	14.38	63.1	
9	391	114.18	442.3	
10	422	75.56	271.5	
11	436	45.80	159.0	
12	457	72.13	239.1	
13	479	47.49	150.1	
14	490	42.42	131.3	
15	531	131.15	374.2	
16	578	124.48	326.2	
17	648	34.43	80.5	
18	744	53.05	108.1	
19	813	41.69	77.7	
20	874	45.67	79.2	
21	943	31.82	51.1	
22	988	30.27	46.4	
23	1,069	125.76	178.2	
24	2,632	1.15	0.7	
25	10,380	75.00	10.9	Upper Marker
26	12,602	0.00	0.0	

Assay Class: High Sensitivity DNA Assay  
Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
Modified: 10/10/2014 3:59:44 PM

**Electropherogram Summary Continued ...**

... Peak table for sample 1 : 246p

Peak	Size [bp]	Conc. [pg/ $\mu$ l]	Molarity [pmol/l]	Observations
27	13,007	0.00	0.0	

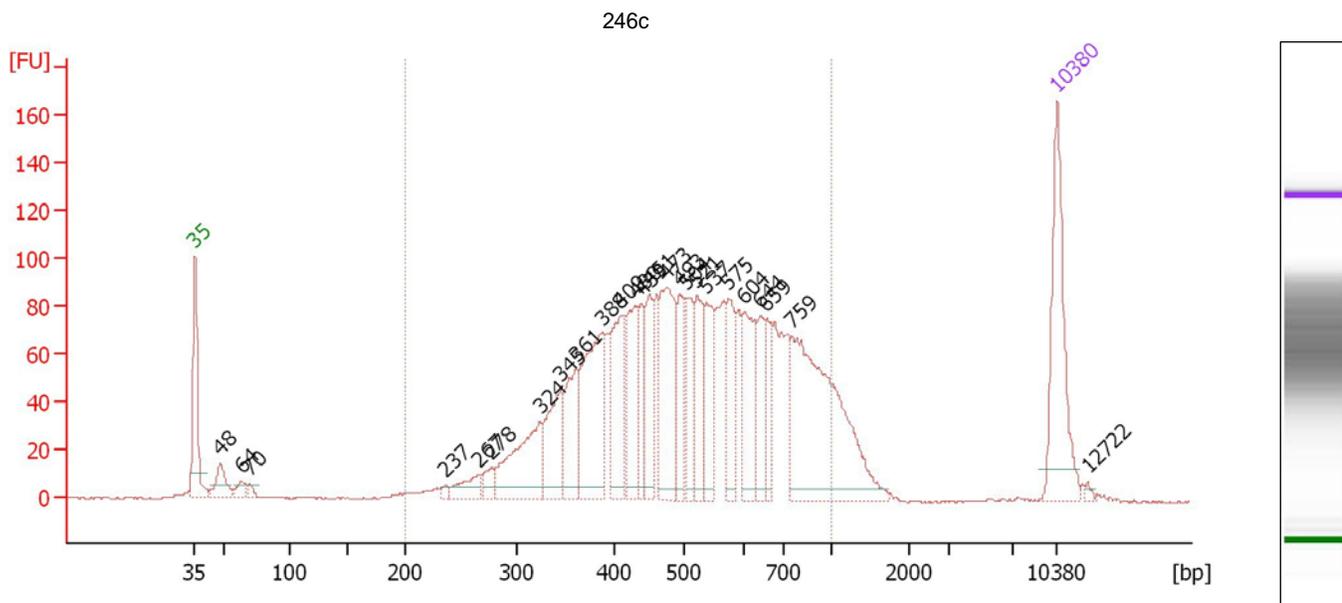
Region table for sample 1 : 246p

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ $\mu$ l]	Molarity [pmol/l]	Color
200	1,000	2,459.1	87	571	30.4	1,341.00	4,071.9	

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
 Modified: 10/10/2014 3:59:44 PM

Electropherogram Summary Continued ...



Overall Results for sample 2 : **246c**

Number of peaks found: 25                      Corr. Area 1: 2,343.9  
 Noise: 0.3

Peak table for sample 2 : **246c**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	48	28.38	889.7	
3	64	11.28	267.6	
4	70	6.15	133.4	
5	237	4.38	28.1	
6	267	22.11	125.5	
7	278	12.78	69.8	
8	324	87.57	409.2	
9	345	61.42	269.8	
10	361	55.61	233.3	
11	388	114.30	446.6	
12	409	74.03	274.0	
13	430	66.99	235.8	
14	439	36.88	127.4	
15	451	46.43	156.1	
16	473	97.01	310.6	
17	493	39.61	121.6	
18	504	41.61	125.1	
19	521	48.31	140.5	
20	537	45.83	129.2	
21	575	43.76	115.3	
22	604	56.85	142.7	
23	644	34.84	82.0	
24	659	28.04	64.5	
25	759	196.39	392.3	
26	10,380	75.00	10.9	Upper Marker

Assay Class: High Sensitivity DNA Assay  
Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
Modified: 10/10/2014 3:59:44 PM

**Electropherogram Summary Continued ...**

... Peak table for sample 2 : 246c

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
27	12,722	0.00	0.0	

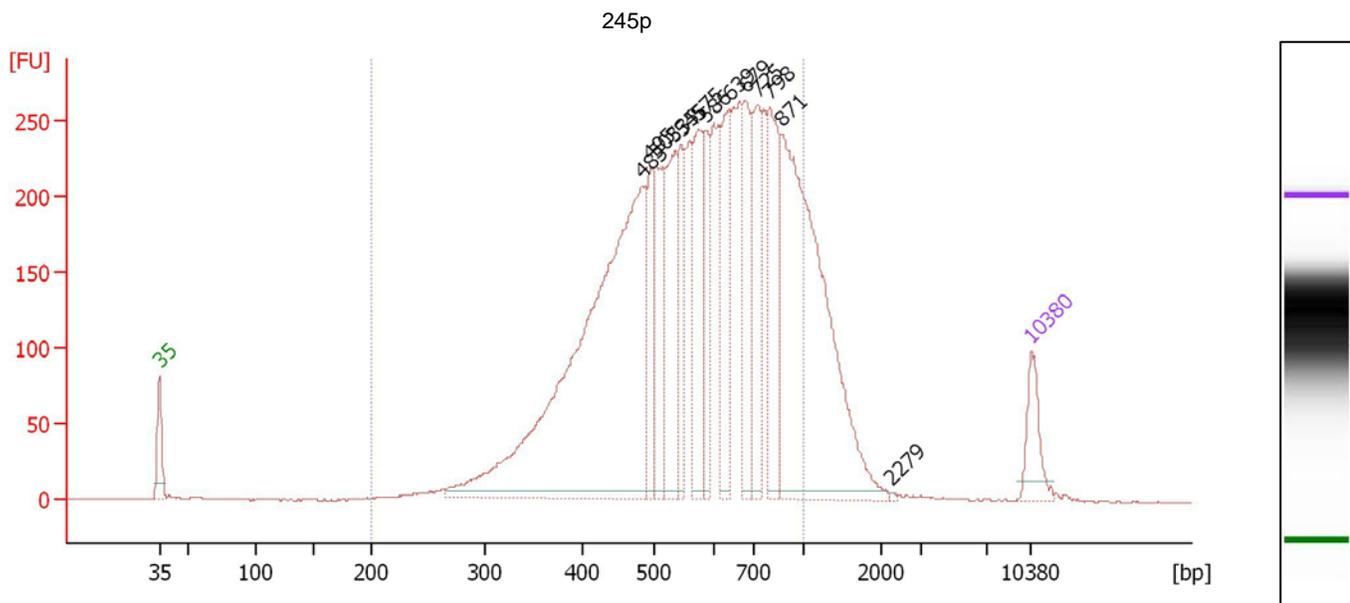
Region table for sample 2 : 246c

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	2,343.9	92	525	31.4	1,397.30	4,624.2	

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
 Modified: 10/10/2014 3:59:44 PM

Electropherogram Summary Continued ...



Overall Results for sample 3 : **245p**

Number of peaks found: 13                      Corr. Area 1: 5,409.6  
 Noise: 0.2

Peak table for sample 3 : **245p**

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	483	1,629.46	5,116.2	
3	495	157.31	481.6	
4	503	221.00	665.5	
5	535	298.75	846.4	
6	545	158.09	439.5	
7	575	243.31	641.4	
8	586	156.09	403.6	
9	639	209.72	497.6	
10	679	203.17	453.6	
11	725	223.84	467.8	
12	798	266.45	505.9	
13	871	980.68	1,705.9	
14	2,279	2.12	1.4	
15	10,380	75.00	10.9	Upper Marker

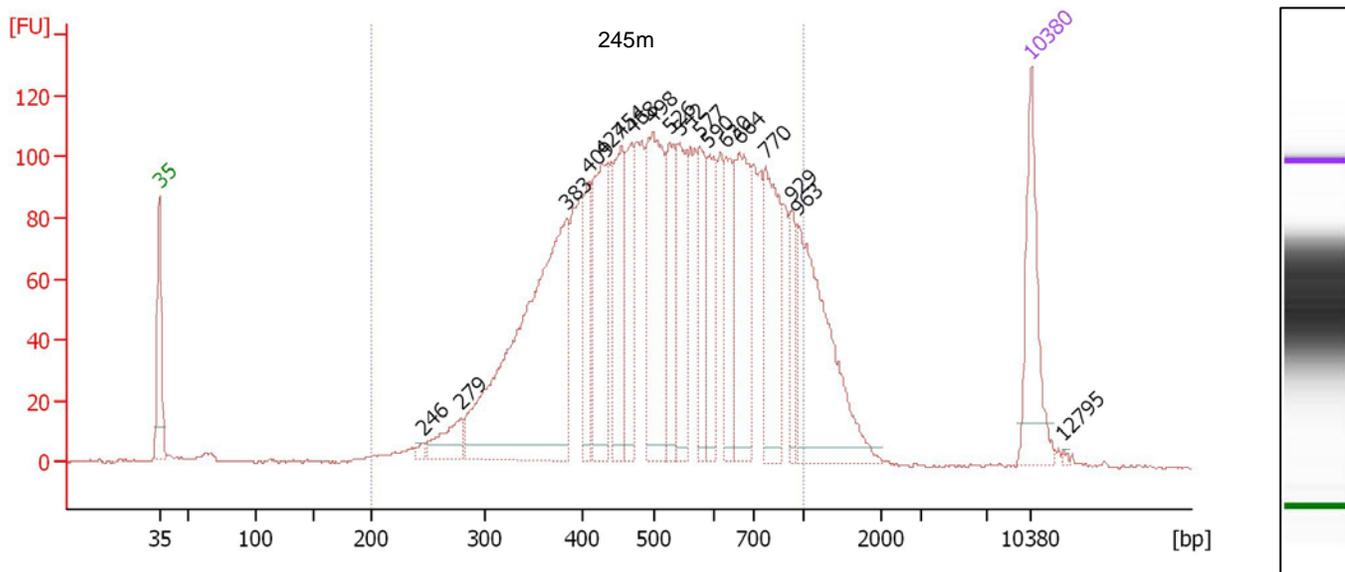
Region table for sample 3 : **245p**

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]	Color
200	1,000	5,409.6	89	591	28.2	5,218.70	14,991.6	Blue

Assay Class: High Sensitivity DNA Assay  
 Data Path: D:\...gh Sensitivity DNA Assay\_DE72904909\_2014-10-10\_15-18-28.xad

Created: 10/10/2014 3:18:27 PM  
 Modified: 10/10/2014 3:59:44 PM

Electropherogram Summary Continued ...



Overall Results for sample 4 : 245m

Number of peaks found: 18                      Corr. Area 1: 2,897.7  
 Noise: 0.3

Peak table for sample 4 : 245m

Peak	Size [bp]	Conc. [pg/μl]	Molarity [pmol/l]	Observations
1	35	125.00	5,411.3	Lower Marker
2	246	6.40	39.4	
3	279	42.61	231.2	
4	383	445.01	1,758.5	
5	409	65.16	241.6	
6	427	141.31	501.7	
7	454	120.31	401.9	
8	468	81.55	264.2	
9	498	181.82	553.6	
10	526	90.90	262.0	
11	542	93.01	260.1	
12	577	74.17	194.8	
13	590	75.56	194.0	
14	630	65.82	158.2	
15	664	132.11	301.3	
16	770	123.38	242.8	
17	929	38.27	62.4	
18	963	185.67	292.2	
19	10,380	75.00	10.9	Upper Marker
20	12,795	0.00	0.0	

Region table for sample 4 : 245m



---

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/ $\mu$ l]	Molarity [pmol/l]	Color
200	1,000	2,897.7	92	537	31.5	2,400.58	7,760.4	