



Programa de
Desarrollo de las
Ciencias Básicas



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



FACULTAD DE
CIENCIAS
UDELAR | fcien.edu.uy



Patrones evolutivos y mecanismos de restricción antiviral en la adaptación de virus ARN zoonóticos.

Lic. Bioq. Mercedes Paz Rijo

Orientador: Dr. Gonzalo Moratorio

Co-Orientadora: Dra. Pilar Moreno

Doctorado en Ciencias Biológicas

Programa de Desarrollo de las Ciencias Básicas

Área Biología

Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo

Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares,
Facultad de Ciencias, Universidad de la República

Abril 2026

Montevideo, Uruguay

Patrones evolutivos y mecanismos de restricción antiviral en la adaptación de virus ARN zoonóticos

Lic. Bioq. Mercedes Paz Rijo

Tesis de Doctorado presentada al Programa de Desarrollo de las Ciencias Básicas (PEDECIBA), Área Biología, Facultad de Ciencias de la Universidad de la República, como parte de los requisitos necesarios para la obtención del título de Doctora en Ciencias Biológicas.

Orientador: Dr. Gonzalo Moratorio

Co-orientadora: Dra. Pilar Moreno

Tribunal: Dra. Adriana Delfraro

Dra. Natalia Olivero

Dr. Juan Pablo Tosar

La tesis fue llevada a cabo en el Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la República y el Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo, entre marzo de 2022 y abril de 2026.

AGRADECIMIENTOS

Mis gracias totales no pueden ir a otras personas que no sean mis orientadores, Pili y Gonza, quienes confiaron en mí y me invitaron a formar parte de su laboratorio desde los comienzos del cive. Fueron años de trabajo que van a quedar marcados y de los que me llevo muy lindos recuerdos. Gracias por la oportunidad de desarrollar este trabajo y de poder realizar un posgrado.

A Gonza, por transmitirme su enorme experiencia y su pasión por la evolución viral, y por apoyarnos a pelear y salir de la zona de confort para conseguir nuestros sueños aún durante la batalla que le toca pelear. A Pili, por ser una científica excelente que me guió y fue soporte en momentos difíciles, y que con sus palabras siempre nos ofrece calma, manteniendo además el grupo humano tan lindo con el que me tocó compartir estos años. A Juan, porque sin él esta tesis tampoco sería posible, y porque desde que compartimos el viaje a la Antártida en 2016, de alguna forma siempre ha sido fundamental en mi formación.

No puedo agradecer a mis compañeros de laboratorio uno por uno porque sería injusta al no nombrar a alguien, ya que tengo la fortuna de haber recibido siempre ayuda y colaboración cuando la necesité, y espero haber hecho lo mismo por ellos. Gracias por la buena onda, la paciencia y todos los momentos compartidos. Sí quiero agradecer especialmente a Ali, con quien tuve la suerte de compartir mesada desde el comienzo; es otra científica que admiro y de la que aprendí muchísimo.

Gracias también a todos los del LGM, con quienes compartimos el civecafé y que siempre estuvieron dispuestos y con la mejor onda.

A mi mamá, gracias a ella he llegado hasta acá. A Fede, por su amor y apoyo incondicional, no existe mejor compañero de vida. A mis amigas de Floresta y a los que me dejó la Fcien.

Finalmente, quiero dedicar esta tesis a mis abuelos.

RESUMEN

Los virus ARN presentan alta plasticidad genética, que les permite adaptarse a nuevos hospederos bajo presiones selectivas que operan a nivel poblacional e intracelular. En esta tesis se investigó cómo estas presiones moldean la arquitectura genómica y la trayectoria adaptativa de dos virus ARN zoonóticos.

En el primer capítulo se utilizó SARS-CoV-2 como modelo de emergencia y diversificación reciente, evaluando su dinámica evolutiva mediante un estudio composicional del gen S en VOCs y de la emergencia de Ómicron. Análisis multivariados (PCA y mapas de calor) evidenciaron que los patrones de uso de codones y la composición nucleotídica se asocian con la diferenciación genotípica entre variantes virales. La divergencia de Ómicron no se limita a sustituciones aminoacídicas puntuales, sino que podría reflejar una reorganización más amplia de la arquitectura genómica del gen S. En particular, Alfa, Beta, Gamma y Delta mostraron mayor proximidad genética entre sí, mientras que Ómicron ocupó un espacio multivariado claramente diferenciado, consistente con una trayectoria evolutiva independiente en el árbol filogenético de SARS-CoV-2. Estos resultados sugieren que el gen S contiene suficiente información evolutiva para asignar VOCs a sus linajes incluso a partir de variables composicionales globales, reforzando su utilidad como marcador evolutivo. Asimismo, análisis filodinámicos sugieren que los ancestros de este linaje habrían circulado durante un período prolongado antes de su detección, evidenciando diversificación temprana.

En el segundo capítulo se estudió el rol de la inmunidad innata como fuerza selectiva sobre arbovirus, con foco en la proteína antiviral de dedos de zinc (ZAP) de vertebrados, que reconoce dinucleótidos CpG en ARN virales. Los virus ARN que infectan vertebrados presentan una marcada supresión de CpG asociada a la actividad restrictiva de ZAP. Sin embargo, los arbovirus deben replicar eficientemente en vertebrados e invertebrados, por lo que están sometidos a presiones evolutivas potencialmente divergentes. Para abordar esta pregunta se utilizó una variante de MAYV enriquecida en CpG (FG+) y se generó una línea celular A549 deficiente en ZAP mediante CRISPR/Cas9. Los resultados evidenciaron que el aumento de CpG produce una marcada atenuación viral dependiente de este factor restrictivo. Además, se observó que la isoforma ZAPL actúa como principal mediador restrictivo durante etapas tempranas de la infección y que la estimulación con IFN α potencia esta restricción de manera dependiente de ZAP. También se evidenció que MAYV silvestre (WT) es susceptible a la inhibición mediada por ZAP, lo que sugiere que la modulación del contenido de CpG constituye un compromiso adaptativo relevante para arbovirus que alternan entre distintos hospedadores.

En conjunto, estos capítulos muestran cómo distintos niveles de presión selectiva, desde procesos evolutivos poblacionales hasta mecanismos de inmunidad innata, moldean la composición genómica y el potencial adaptativo de virus ARN.

Finalmente, en el tercer capítulo se exploró el rediseño del genoma de MAYV como herramienta experimental mediante la generación de variantes sintéticas capaces de expresar un ARN guía del sistema CRISPR/Cas9 dirigido contra ZAP. Aunque no se detectó edición génica eficiente en las condiciones evaluadas, los resultados demostraron la viabilidad de utilizar MAYV como plataforma para la expresión de ARN pequeños y resaltan su potencial para estudios funcionales.

Palabras clave: SARS-CoV-2; virus Mayaro; alfavirus; arbovirus; CpG; ZAP; inmunidad innata; interferón tipo I; evolución viral; composición nucleotídica; filodinámica; biología sintética; CRISPR/Cas9; sgRNA.

ABSTRACT

RNA viruses exhibit high genetic plasticity, enabling them to adapt to new hosts under selective pressures acting at both the population and intracellular levels. In this thesis, we investigated how these pressures shape the genomic architecture and adaptive trajectories of two zoonotic RNA viruses.

In the first chapter, SARS-CoV-2 was used as a model of recent emergence and diversification to examine its evolutionary dynamics through a compositional analysis of the S gene across variants of concern (VOCs), with particular focus on the emergence of Omicron. Multivariate analyses (PCA and heatmaps) showed that codon usage patterns and nucleotide composition are associated with the genotypic differentiation among viral variants. The divergence of Omicron is therefore not limited to point amino acid substitutions but may reflect a broader reorganization of the genomic architecture of the S gene. In particular, Alpha, Beta, Gamma, and Delta showed greater genetic proximity to one another, whereas Omicron occupied a clearly distinct multivariate space, consistent with an independent evolutionary trajectory within the SARS-CoV-2 phylogenetic tree. These findings suggest that the S gene contains sufficient evolutionary signal to assign VOCs to their respective lineages even using global compositional variables, highlighting its value as an evolutionary marker. Phylodynamic analyses further suggest that the ancestors of this lineage circulated for an extended period before its detection, indicating early diversification.

The second chapter investigated how innate immunity shapes arbovirus genome composition, focusing on the vertebrate zinc-finger antiviral protein (ZAP), which recognizes CpG dinucleotides in viral RNA. RNA viruses infecting vertebrates typically show strong CpG suppression, a feature linked to the restrictive activity of ZAP. However, arboviruses must replicate efficiently in both vertebrate and invertebrate hosts and are therefore exposed to potentially divergent evolutionary pressures. To address this, a CpG-enriched MAYV variant (FG+) was used together with a ZAP-deficient A549 cell line generated by CRISPR/Cas9. Increasing CpG content resulted in marked viral attenuation in a ZAP-dependent manner. The ZAPL isoform was identified as the main restrictive mediator during the early stages of infection, and IFN α stimulation further enhanced this restriction in a ZAP-dependent manner. Importantly, wild-type MAYV was also susceptible to ZAP-mediated inhibition, suggesting that modulation of CpG content represents an important adaptive trade-off for arboviruses that alternate between different hosts.

Together, these chapters show how selective pressures operating at different levels—from population-level evolutionary processes to mechanisms of innate immunity—shape the genomic composition and adaptive potential of RNA viruses.

Finally, the third chapter explored the use of genome engineering in MAYV as an experimental tool through the generation of synthetic viral variants capable of expressing a CRISPR/Cas9 guide RNA targeting ZAP. Although efficient gene editing was not detected under the conditions tested, these results demonstrate the feasibility of using MAYV as a platform for the expression of small RNAs and highlight its potential as an experimental system for functional studies.

Keywords: SARS-CoV-2; Mayaro virus; alphavirus; arbovirus; CpG; ZAP; innate immunity; type I interferon; viral evolution; nucleotide composition; phylodynamics; synthetic biology; CRISPR/Cas9; sgRNA.

TABLA DE CONTENIDO

INTRODUCCIÓN GENERAL.....	13
Emergencia zoonótica y capacidad adaptativa de virus ARN.....	14
Interacción virus–hospedero como motor de evolución viral.....	15
Justificación y estructura de la tesis.....	18
CAPÍTULO 1.....	19
Introducción.....	20
Emergencia de SARS-CoV-2 desde el reservorio animal hacia la transmisión humana.....	20
Organización genómica y ciclo replicativo.....	22
Mecanismos moleculares de generación de variabilidad en SARS-CoV-2.....	24
Clasificación y vigilancia genómica de variantes virales.....	25
Emergencia y evolución de variantes durante la pandemia.....	26
Filogenia molecular y filodinámica en el estudio de brotes y epidemias virales.....	30
Análisis composicional de genomas virales: uso de codones y sesgos nucleotídicos.....	31
Hipótesis.....	33
Objetivo general.....	33
Objetivos específicos.....	33
Materiales y métodos.....	34
Secuencias.....	34
Alineamiento de secuencias.....	34
Asignación de genotipo de SARS-CoV-2.....	34
Análisis de composición nucleotídica y uso de codones.....	35
Análisis filodinámicos.....	35
Resultados.....	36
Caracterización de patrones evolutivos en el gen S de VOCs.....	36
Análisis de coalescencia bayesiana de las cepas Ómicron.....	40
Discusión.....	43
Conclusiones.....	46
Perspectivas.....	47
CAPÍTULO 2.....	48
Introducción.....	49
Virus Mayaro.....	49
Clasificación.....	49
Estructura genómica y ciclo replicativo.....	49
Epidemiología y relevancia clínica.....	52
Respuesta inmune innata frente a MAYV.....	53
Proteína antiviral de dedos de zinc (ZAP).....	56
Identificación y especificidad.....	56
Isoformas y regulación de la expresión transcripcional.....	57
Diferencias estructurales y funcionales.....	58

Mecanismos moleculares de restricción viral mediados por ZAP.....	61
Reconocimiento preferencial de dinucleótidos CpG y dependencia del contexto genómico.....	62
Implicancias adaptativas de ZAP en el sesgo de dinucleótidos CpG en genomas de virus ARN.....	64
Relevancia de ZAP en arbovirus y justificación del estudio en MAYV.....	67
Antecedentes y resultados preliminares.....	68
Hipótesis.....	71
Objetivo general.....	71
Objetivos específicos.....	71
Materiales y métodos.....	72
Células.....	72
Clones Infecciosos.....	72
Generación de stocks virales.....	72
Secuenciación del genoma completo de MAYV mediante NGS.....	73
Titulación viral mediante ensayo en placa.....	73
Infecciones virales y tratamiento con IFN α	74
Extracción de ARN y PCR Cuantitativa en Tiempo Real (RT-qPCR) para: ZAP y MAYV.....	74
Preparación de lisados celulares para SDS-PAGE y Western blot de ZAP, nsP1, E1 y Cas9.....	75
Generación de células A549 ZAP KO mediante CRISPR/Cas9.....	75
Silenciamiento génico de ZAPL y ZAPS por DsiRNA.....	76
Estadística.....	76
Resultados.....	77
Análisis de la expresión de ZAP en células A549 infectadas con MAYV.....	77
Generación de células A549 ZAP KO mediante CRISPR/Cas9.....	80
La atenuación de MAYV FG+ en células A549 es dependiente de ZAP.....	83
ZAPL es la principal isoforma responsable de la restricción de MAYV FG+ en células A549.....	86
El tratamiento con IFN α demuestra que la atenuación de MAYV FG+ es dependiente del grado de activación de ZAP.....	87
ZAP regula negativamente la expresión transcripcional de IFN β y de las citoquinas CXCL10 y CCL5, pero no OAS1 en células A549 infectadas con MAYV.....	89
La atenuación de MAYV FG+ por IFN α requiere la expresión de ZAP.....	92
ZAP reduce el ARN viral de MAYV FG+ en las primeras etapas de la infección..	93
Discusión.....	96
Conclusiones.....	102
Perspectivas.....	103
CAPÍTULO 3.....	104
Introducción.....	105
CRISPR/Cas9 para el estudio funcional de interacciones virus–hospedero...	105
Limitaciones de los CRISPR screens convencionales.....	107
Alfavirus como plataforma de expresión de secuencias heterólogas.....	109

Origen del proyecto presentado en este capítulo y justificación del uso de MAYV como vector de sgRNAs.....	110
Hipótesis.....	112
Objetivo general.....	112
Objetivos específicos.....	112
Materiales y Métodos.....	113
Células.....	113
Generación de clones infecciosos.....	113
Secuenciación de clones infecciosos.....	117
Generación de stocks virales.....	117
Cinética de replicación viral.....	117
Evaluación de la estabilidad genómica del inserto en los virus sintéticos MAYV sgRNA2_miR y MAYV sgRNA2_HDV.....	118
Cuantificación relativa del micro-RNA 124 (miR-124) maduro mediante stem-loop RT-qPCR.....	119
Eficiencia de edición génica en células infectadas con MAYV sgRNA2_miR y MAYV sgRNA2_HDV.....	119
Western blot para ZAP y Cas9 en células A549-Cas9.....	120
Resultados.....	121
Diseño in silico de clones infecciosos codificantes para ARN guías.....	121
Generación de clones infecciosos de MAYV codificantes para un sgRNA contra el gen ZC3HAV1.....	122
Obtención de stock virales (P2) a partir de los clones infecciosos desarrollados.	123
Caracterización funcional de los virus desarrollados para dirigir la edición génica mediante CRISPR/Cas9 en células infectadas.....	125
1- Análisis de cinética de replicación en células A549-Cas9.....	125
2- Análisis de expresión de ZAP en células A549-Cas9.....	133
3- Eficiencia de edición génica en células infectadas con MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV.....	134
Expresión del miR-124 en células infectadas con MAYV_sgRNA2_miR.....	136
Discusión.....	137
Conclusiones.....	140
Perspectivas.....	141
CONCLUSIONES FINALES.....	143
REFERENCIAS BIBLIOGRÁFICAS.....	145
INFORMACIÓN SUPLEMENTARIA.....	154
Tablas y figuras suplementarias:.....	155
Producción científica asociada directamente a esta tesis.....	172
Producción científica relacionada a la temática de la tesis.....	172

LISTA DE FIGURAS

Figura 1.1. Estructura del virión y organización genómica de SARS-CoV-2.....	23
Figura 1.2. Emergencia y diversificación de VOCs hasta el año 2022.....	28
Figura 1.3. Dinámica temporal global de la frecuencia relativa de variantes de SARS-CoV-2.....	29
Figura 1.4. PCA del uso de codones en la proteína S de VOCs.....	37
Figura 1.5. Mapa de calor del uso de codones en la proteína S de VOCs.....	38
Figura 1.6. PCA de la composición nucleotídica en la proteína S de VOCs.....	39
Figura 1.7. Mapa de calor de las frecuencias nucleotídicas por posición de codón en la proteína S de VOCs.....	40
Figura 1.8. Inferencia filogenética de Ómicron B.1.1.529 (21M) hasta diciembre de 2021 mediante una aproximación bayesiana basada en MCMC.....	42
Figura 2.1. Representación de la estructura de MAYV.....	50
Figura 2.2. Estructura genómica de los alfavirus.	50
Figura 2.3. Ciclo replicativo de los alfavirus.	52
Figura 2.4. Respuesta inmune innata frente a la infección por alfavirus.....	54
Figura 2.5. Cascada de señalización JAK/STAT desencadenada por IFN-I.....	55
Figura 2.6. Isoformas de ZAP.....	60
Figura 2.7. Dominio de unión a ARN (RBD) de ZAP humana en complejo con ARN.	61
Figura 2.8. Sesgo de CpG en virus ssRNA y sus hospederos metazoos.....	65
Figura 2.9. Sesgo de CpG en virus que infectan vertebrados y/o invertebrados.	68
Figura 2.10. Distribución de CpG presentes en una posición genómica determinada para MAYV.	69
Figura 2.11. Variantes sintéticas de MAYV con alto contenido de CpG son atenuadas en células de vertebrado.....	70
Figura 2.12. MAYV WT induce una mayor expresión de ZAP que la variante FG+ en células A549.....	79
Figura 2.13. Diseño y evaluación de ARN guías contra el gen <i>ZC3HAV1</i> codificante para ZAP.....	81
Figura 2.14. Validación de ZAP KO en clones A549 (E11_3, E12_2 y G9_1).....	83

Figura 2.15. La atenuación de MAYV FG+ en células A549 es dependiente de ZAP.	85
Figura 2.16. ZAPL es la principal isoforma responsable de la restricción de MAYV FG+ en células A549.....	86
Figura 2.17. La activación de ZAP tras la infección con MAYV WT y FG+ se ve incrementada por IFN α	88
Figura 2.18. IFN α inhibe la replicación de MAYV, observándose una mayor inhibición de FG+ que WT.....	89
Figura 2.19. ZAP regula negativamente la expresión de IFN β , CXCL10 y CCL5, pero no de OAS1.....	91
Figura 2.20. ZAP es necesaria para el efecto antiviral del IFN α sobre la variante FG+.	93
Figura 2.21. ZAP reduce los niveles de ARN genómico de MAYV FG+ a tiempos tempranos post-infección.....	94
Figura 2.22. La activación de ZAP ocurre a tiempos tardíos postinfección.....	95
Figura 2.23. Modelo propuesto del papel de ZAP en el control de la infección por MAYV.....	101
Figura 3.1. Esquema general del flujo de trabajo de <i>CRISPR screen</i> basado en CRISPR/Cas9.....	106
Figura 3.2. Esquema del funcionamiento de un <i>CRISPR screen</i> mediado por una librería de virus codificantes para sgRNA.....	108
Figura 3.3. Genoma de alfavirus con doble promotor subgenómico para la expresión de secuencias heterólogas.....	110
Figura 3.4. Generación de MAYV_sgRNA2_miR por el sistema de clonado <i>In-Fusion</i>	114
Figura 3.5. Generación de MAYV_sgRNA2_HDV por PCR de fusión y clonación.	116
Figura 3.6. Representación esquemática del diseño in silico de clones infecciosos MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV y procesamiento intracelular hipotético.	122
Figura 3.7. Efecto citopático en células Vero-E6 transfectadas con distintos clones infecciosos de MAYV.....	124
Figura 3.8. Análisis de la estabilidad del inserto en MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV durante pasajes seriados (P0–P2) mediante RT-PCR.....	125
Figura 3.9. Cinética de replicación viral a baja MOI muestra un fenotipo atenuado de MAYV_sgRNA2_miR.....	126

Figura 3.10. Cinética de replicación viral a alta MOI de MAYV_sgRNA2_miR.....	127
Figura 3.11. Estabilidad del inserto en MAYV_sgRNA2_miR a lo largo del tiempo.	128
Figura 3.12. Perfiles de cobertura de amplicones de MAYV_sgRNA2_miR obtenidos por secuenciación ONT.....	129
Figura 3.13. Cinética de replicación de MAYV_sgRNA2_HDV en células A549 y A549-Cas9.....	130
Figura 3.14. Estabilidad del inserto en MAYV_sgRNA2_HDV.....	131
Figura 3.15. Perfiles de cobertura de amplicones de MAYV_sgRNA2_HDV obtenidos por secuenciación ONT.....	132
Figura 3.16. Expresión de ZAP en células A549-Cas9 infectadas con MAYV_sgRNA2_miR.....	133
Figura 3.17. Expresión de ZAP en células A549-Cas9 infectadas con MAYV_sgRNA2_HDV.....	134
Figura 3.18. Cuantificación relativa de miR-124 maduro en células A549-Cas9 infectadas.....	136
Figura S1.1. Análisis Bayesiano MCMC de las cepas BA.1, BA.2 y BA.3 de SARS-CoV-2 (Ómicron).....	156
Figura S2.1. Cinética de replicación de MAYV WT y FG+ en el pool celular A549 ZAP KD.....	158
Figura S2.2. Inducción de IFN β y OAS1 en células pretratadas con IFN α	159
Figura S3.1. PCR para la amplificación del gBlock de 351 pb y el backbone MAYV WT.....	170
Figura S3.2. PCR de fusión para la construcción del clon infeccioso MAYV_sgRNA2_HDV.....	170
Figura S3.3. Secuenciación por Sanger y análisis ICE en células A549-Cas9 infectadas.....	171

LISTA DE TABLAS

Tabla 1.1 Inferencia coalescente bayesiana de las cepas Ómicron de SARS-CoV-2.....	41
Tabla 3.1 Frecuencia de edición en el locus <i>ZC3HAV1</i> luego de la infección con virus recombinantes MAYV.....	135

Tabla S1.1. Secuencias de Ómicron (BA.2 y BA.3) incluidas en el análisis filodinámico bayesiano suplementario.....	155
Tabla S2.1: Lista de cebadores y su temperatura de fusión (T _m) en grados centígrados (°C).....	157
Tabla S3.1: Lista de cebadores y su temperatura de fusión (T _m) en grados centígrados (°C).....	160
Tabla S3.2 Cobertura de secuenciación por ONT de amplicones realizados para evaluar la estabilidad genómica del inserto en los virus MAYV_sgRNA2_miR (820 nt) y MAYV_sgRNA2_HDV (697 nt).....	161
Tabla S3.3: Métricas de control de calidad de la secuenciación Illumina para las bibliotecas de amplicones secuenciadas en la plataforma NovaSeq X Plus (PE150).....	168

LISTA DE ABREVIATURAS

A549: Línea celular de adenocarcinoma de pulmón humano
 ACE2: enzima convertidora de angiotensina 2
 Cas9: CRISPR-associated protein 9
 CCL5: C-C motif chemokine ligand 5
 CpG: Dinucleótido citosina-fosfato-guanina
 CRISPR: Clustered Regularly Interspaced Short Palindromic Repeats
 CXCL10: C-X-C motif chemokine ligand 10
 DNA: Ácido desoxirribonucleico
 dsRNA: ARN de doble cadena
 E1: Glicoproteína de envoltura E1 de alfavirus
 E2: Glicoproteína de envoltura E2 de alfavirus
 FG+: Variante de MAYV enriquecida en dinucleótidos CpG
 GISAID: Global Initiative on Sharing All Influenza Data
 HDV: ribozima del virus de la hepatitis delta
 IFN: Interferón
 IFN-I: Interferón tipo I
 IFN α : Interferón alfa
 IFN β : Interferón beta
 IFNAR: Receptor de interferón tipo I
 IRF: Interferon Regulatory Factor
 ISG: Interferon-Stimulated Gene
 ISGF3: Interferon-Stimulated Gene Factor 3
 ISRE: Interferon-Stimulated Response Element
 JAK: Janus kinase
 KD: Knockdown
 KO: Knockout
 MAVS: Mitochondrial Antiviral Signaling protein

MAYV: Virus Mayaro
MCMC: Markov Chain Monte Carlo
miR: microRNA
MOI: Multiplicidad de Infección
Mxra8: Matrix Remodeling Associated 8
NF- κ B: Nuclear Factor kappa B
NLRP1: NOD-like receptor family pyrin domain-containing 1
NLRP3: NOD-like receptor family pyrin domain containing 3
NTD: Dominio N-terminal
nsP: Proteínas no estructurales virales
OAS1: 2'-5'-Oligoadenylate synthetase 1
ONT: Oxford Nanopore Technologies
PAMPs: Pathogen-Associated Molecular Patterns
PARP13: Poly(ADP-ribose) polymerase 13 (ZAP)
PCA: Principal Component Analysis
PCR: Polymerase Chain Reaction
PRRs: Pattern Recognition Receptors
qPCR: Quantitative Polymerase Chain Reaction
RBD: Dominio de unión al receptor
RdRp: RNA-dependent RNA polymerase
RIG-I: Retinoic acid-Inducible Gene I
RT-qPCR: Reverse transcription quantitative PCR
sgRNA: single guide RNA
SNP: Single-nucleotide polymorphism
Cas9: Streptococcus pyogenes Cas9
SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2
STAT: Signal Transducer and Activator of Transcription
TBK1: TANK-binding kinase 1
TLR: Toll-like receptor
VERO-E6: Células de riñón de mono verde africano
VOCs: Variants of Concern
WT: Wild type
ZAP: Zinc-finger antiviral protein
ZC3HAV1: Gen que codifica para ZAP

INTRODUCCIÓN GENERAL

Emergencia zoonótica y capacidad adaptativa de virus ARN

Se estima que entre el 60% y el 75% de los patógenos emergentes en humanos, descritos en las últimas décadas, derivan de reservorios animales, lo que refleja la creciente interacción entre poblaciones humanas, fauna silvestre y animales domésticos ¹. Estos patógenos circulan naturalmente en reservorios animales y, bajo determinadas condiciones ecológicas, como la expansión urbana, la deforestación, la intensificación de la producción animal, la movilidad global, o los cambios climáticos que modifican la distribución de vectores pueden generar eventos de transmisión interespecie (*spillover*) ². Sin embargo, la coincidencia ecológica no es suficiente para la emergencia. El establecimiento exitoso en la nueva especie requiere compatibilidad funcional con el entorno celular del hospedero, incluyendo la capacidad de utilizar receptores específicos, replicarse eficientemente y evadir o modular los mecanismos de la inmunidad innata. La mayoría de estos eventos no progresa más allá de infecciones aisladas: el virus puede ingresar y replicar de manera limitada sin alcanzar niveles suficientes para sostener la transmisión. Sólo en circunstancias particulares ciertas variantes logran superar estas barreras iniciales y establecer cadenas de transmisión entre individuos. En consecuencia, la emergencia viral no depende exclusivamente de la exposición sino de la capacidad adaptativa del virus frente al nuevo entorno biológico ^{2,3}.

Los virus de genoma ARN presentan propiedades intrínsecas que favorecen este proceso ⁴. A lo largo de la historia, distintos virus ARN han sido responsables tanto de brotes localizados como de pandemias globales, incluyendo la pandemia de influenza de 1918 causada por el virus influenza A H1N1, la emergencia del coronavirus del síndrome respiratorio agudo severo (SARS-CoV-1) entre 2002 y 2003, la pandemia de gripe A (H1N1) de 2009–2010 y, más recientemente, la pandemia iniciada en 2019 por el coronavirus SARS-CoV-2. Asimismo, numerosos virus ARN, como los transmitidos por vectores (como por ejemplo los virus Zika, Dengue y fiebre amarilla) y otros patógenos humanos persistentes, como los virus hepatitis C y el virus de la inmunodeficiencia humana (VIH), son un claro ejemplo de la capacidad recurrente de los virus ARN para atravesar barreras interespecie, adaptarse a nuevos hospederos y establecer una transmisión sostenida en humanos pudiendo alcanzar una escala global.

La replicación de los virus ARN está mediada por polimerasas dependientes de ARN (RdRp, del inglés *RNA-dependent RNA polymerase*) de baja fidelidad (entre 10^{-4} y 10^{-6} sustituciones por nucleótido por generación), que en la mayoría de las familias virales carecen de mecanismos eficientes de corrección de errores ⁵. Como resultado, dentro

de un mismo hospedero no existe un genoma único, sino una población de variantes estrechamente relacionadas, configurando una estructura poblacional dinámica con distinta aptitud replicativa ⁶. Esta diversidad constituye el sustrato sobre el cual actúa la selección natural ante las presiones selectivas en un hospedero nuevo. A esta variabilidad se suman mecanismos adicionales de generación de diversidad genética, como la recombinación o, en el caso de virus con genomas segmentados, el reordenamiento genómico. Estos procesos pueden producir cambios fenotípicos abruptos y, ocasionalmente, variantes con ventajas replicativas o de transmisión. Durante la infección y la transmisión entre hospederos ocurren además cuellos de botella poblacionales que reducen drásticamente la diversidad viral y favorecen la fijación de ciertas variantes por selección o deriva genética ⁷.

La adaptación viral tras un evento de *spillover* debe entenderse entonces como un proceso continuo. Comienza en el reservorio animal, donde el virus explora la diversidad genética, y continúa tras la infección humana, donde enfrenta nuevas restricciones celulares y fisiológicas. La mayoría de los linajes se extinguen rápidamente, mientras que algunos adquieren combinaciones genéticas compatibles con el nuevo hospedero y logran sostener la transmisión. La emergencia representa, por lo tanto, el resultado de múltiples ciclos de generación de diversidad y filtrado selectivo más que un evento puntual.

En conjunto, la elevada tasa de variación genética, los tiempos replicativos cortos y los grandes tamaños poblacionales convierten a los virus ARN en sistemas altamente eficientes para explorar el espacio evolutivo. Estas características excepcionales pueden explicar su alta participación en la emergencia y reemergencia de patógenos humanos y subrayan la importancia de comprender los determinantes moleculares que condicionan su adaptación a nuevos hospederos.

Interacción virus–hospedero como motor de evolución viral

Tras la infección viral, el ambiente celular del hospedero impone múltiples restricciones al establecimiento de la replicación viral. La entrada del virus a la célula no garantiza una replicación eficiente, ya que durante el cambio de hospedero pueden existir incompatibilidades moleculares relacionadas con la disponibilidad de factores celulares, la estabilidad del ARN o la eficiencia de traducción, que pueden afectar su replicación incluso antes de la activación de la respuesta inmune.

Las células poseen mecanismos constitutivos capaces de detectar material genético exógeno y restringir su expresión. El reconocimiento de patrones moleculares

asociados a patógenos (PAMPs) por receptores de reconocimiento de patrones (PRRs) permite discriminar entre ARN propio y no propio e iniciar cascadas de señalización intracelular que activan programas antivirales, incluyendo la degradación selectiva del ARN viral, la inhibición de la traducción y la restricción de la replicación genómica. Estas respuestas se activan de manera temprana tras la entrada del virus y constituyen una primera barrera de selección intracelular ^{8,9}.

A esta restricción basal se suma la respuesta inducida por interferones, que incrementa la presión selectiva mediante la expresión coordinada de numerosos genes estimulados por interferón (ISGs)¹⁰. Desde una perspectiva evolutiva, estos mecanismos no sólo reducen la carga viral, sino que condicionan qué variantes genómicas pueden mantenerse en el tiempo. En consecuencia, la evolución viral se encuentra fuertemente determinada por la necesidad de evadir el reconocimiento y la restricción celular, de modo que la aptitud viral resulta del balance entre eficiencia replicativa y capacidad de evitar la activación de las defensas del hospedero.

Un aspecto central de esta interacción es que la presión selectiva no actúa únicamente sobre proteínas virales, sino también sobre la secuencia nucleotídica del genoma. La composición de nucleótidos, el uso de codones y la frecuencia de determinados motivos de secuencia no son aleatorios, sino que reflejan la compatibilidad con el entorno intracelular. Además de la información para codificar proteínas virales, determinadas configuraciones genéticas pueden favorecer la estabilidad del ARN o su traducción, mientras que otras aumentan la probabilidad de reconocimiento por mecanismos antivirales. Así, el genoma viral constituye un rasgo fenotípico sujeto a selección. Como resultado, muchos virus presentan sesgos composicionales característicos que se conservan a lo largo de su evolución en hospederos específicos, lo que sugiere la existencia de restricciones dependientes de secuencia que operan de manera sostenida durante la replicación viral. La adaptación viral implica entonces un equilibrio entre maximizar la eficiencia replicativa y minimizar la detección por la célula, lo que conduce a la fijación de configuraciones genómicas compatibles con el hospedero ¹¹.

En algunos casos, la presión selectiva se concentra en regiones funcionales específicas del genoma, particularmente en proteínas involucradas en la entrada a la célula o en el reconocimiento por anticuerpos neutralizantes ^{12,13}. Estas regiones pueden estar sometidas a selección positiva sostenida, favoreciendo mutaciones que incrementan la afinidad por receptores celulares o disminuyen la susceptibilidad a la neutralización inmune. Cuando estas variantes presentan ventajas de transmisión o

escape inmunológico, pueden expandirse poblacionalmente y dar lugar a procesos de adaptación molecular, evidenciados por la aparición y reemplazo sucesivo de linajes virales durante brotes epidémicos o pandemias ¹⁴.

En conjunto, la interacción virus–hospedero no sólo determina el éxito inmediato de la infección, sino que moldea progresivamente la arquitectura del genoma viral. La composición genómica puede interpretarse entonces como el resultado acumulativo de presiones selectivas que operan tanto a nivel intracelular como poblacional, donde la supervivencia viral depende del equilibrio entre la eficiencia replicativa y la capacidad de evadir o tolerar los mecanismos de defensa del hospedero.

Desde una perspectiva funcional, comprender cómo las presiones selectivas moldean la organización y composición de los genomas virales no sólo permite interpretar los procesos de adaptación al hospedero, sino que también proporciona información valiosa para poder manipular experimentalmente estos genomas. El conocimiento de los elementos funcionales que regulan la replicación, traducción y estabilidad del ARN viral permite identificar regiones del genoma que toleran modificaciones o la inserción de secuencias heterólogas sin comprometer la replicación viral, lo que ha posibilitado el desarrollo de variantes virales sintéticas derivadas de genomas modificados con fines experimentales o biotecnológicos. En este contexto, los alfavirus constituyen plataformas particularmente versátiles para la generación de variantes virales sintéticas debido a su eficiente replicación citoplasmática y a la organización modular de su genoma, propiedades que han permitido utilizarlos para la expresión de secuencias heterólogas durante la infección, como proteínas reporteras, facilitando así el estudio funcional de genes del hospedero y de los mecanismos celulares que controlan la replicación viral.

Justificación y estructura de la tesis

La evidencia acumulada a lo largo de la historia indica que los virus ARN zoonóticos constituyen una amenaza persistente para la salud pública debido a su elevada capacidad de adaptación genética y a su potencial de transmisión entre especies ¹⁵. La interacción dinámica entre virus y hospedero genera presiones selectivas que moldean la composición del genoma viral, su eficiencia replicativa y su capacidad de evasión inmune ¹¹. Comprender estos procesos requiere integrar enfoques experimentales y evolutivos que permitan analizar, por un lado, los determinantes moleculares del hospedero que restringen la replicación viral y, por otro, los patrones adaptativos que emergen en los genomas virales a escala poblacional.

En este contexto, la presente tesis enmarca distintos niveles de análisis complementarios organizados en tres capítulos. Estos están vinculados a la adaptación de virus ARN zoonóticos frente a las presiones selectivas impuestas por el hospedero, abordando la interacción virus–hospedero desde distintas escalas de análisis y utilizando dos sistemas virales.

En el **capítulo 1** se caracterizan los patrones evolutivos del gen S en las variantes de preocupación de SARS-CoV-2 (VOCs; Alfa, Beta, Gamma, Delta y Ómicron) mediante el análisis comparativo de la composición nucleotídica y el uso de codones, con el objetivo de evaluar su asociación con la diferenciación genotípica entre linajes virales. Asimismo, se examina la dinámica evolutiva asociada a la emergencia y expansión de la variante Ómicron como ejemplo contemporáneo de adaptación viral a escala poblacional.

En el **capítulo 2**, se analiza experimentalmente la interacción virus–hospedero a escala intracelular, utilizando el arbovirus Mayaro (MAYV) como modelo. Este capítulo se enfoca en el estudio del rol antiviral de la proteína ZAP específica de vertebrados frente a MAYV y su impacto en la adaptación de este virus a un ciclo replicativo dual.

Por último, en el **capítulo 3** se desarrolla una aproximación funcional basada en la modificación del genoma de MAYV para la expresión de ARN guía (sgRNA) del sistema CRISPR/Cas9. Se explora la viabilidad y funcionalidad de variantes sintéticas para llevar a cabo la edición génica en células que expresan Cas9 de manera constitutiva, permitiendo evaluar la funcionalidad de los sgRNA en el contexto de una infección activa.

CAPÍTULO 1

Dinámica evolutiva de SARS-CoV-2: análisis
composicional del gen S y emergencia de
Ómicron

Introducción

Emergencia de SARS-CoV-2 desde el reservorio animal hacia la transmisión humana.

El análisis del genoma completo de un virus detectado en muestras clínicas, procedentes de un brote de neumonía detectado en la ciudad de Wuhan (China) en diciembre de 2019, permitió identificar un nuevo agente viral con capacidad de infectar humanos. La primera anotación del genoma completo de SARS-CoV-2, de 29.903 nucleótidos, reveló que se trataba de un virus de ARN monocatenario de sentido positivo (+ssRNA) perteneciente al género *Betacoronavirus* (β -CoV), dentro de la familia *Coronaviridae*¹⁶. Posteriormente, el Comité Internacional de Taxonomía de Virus (ICTV) denominó oficialmente al virus SARS-CoV-2, mientras que la enfermedad asociada recibió el nombre de COVID-19 por la Organización Mundial de la Salud (OMS)¹⁷.

En pocas semanas tras publicarse la primera secuencia genómica completa, se confirmó la transmisión sostenida entre humanos, evidenciada por brotes familiares y nosocomiales. A finales de enero de 2020 el virus ya se había diseminado dentro y fuera de China, lo que llevó a la OMS a declarar una emergencia de salud pública de importancia internacional el 30 de enero, y posteriormente declararse el estado de pandemia el 11 de marzo de 2020.

Los análisis filogenéticos demostraron que SARS-CoV-2 comparte la organización genómica y características estructurales con otros coronavirus, incluido SARS-CoV, agente causante del brote de síndrome respiratorio agudo grave en Asia en 2003. Asimismo, su linaje evolutivo se relaciona con coronavirus que circulan naturalmente en murciélagos. Sin embargo, el origen de SARS-CoV-2 continúa en debate, ya que ninguno de los coronavirus conocidos representa su progenitor directo y la divergencia genética observada sugiere décadas de evolución no muestreada, posiblemente con la participación de un hospedero intermedio aún no identificado^{18,19}.

El coronavirus más cercano descrito hasta la fecha a nivel de genoma completo es BANAL-20-52 (BANAL-52), identificado en murciélagos en cuevas del norte de Laos, que comparte aproximadamente un 96,8% de identidad nucleotídica y un 98,4% de identidad aminoacídica con SARS-CoV-2^{20,21}. Previamente, el virus de murciélago RaTG13, aislado de *Rhinolophus affinis* (Yunnan, China, 2013), era considerado el pariente más cercano, con 96,1–96,2% de identidad genómica. Estos virus pertenecen al subgénero *Sarbecovirus* (de los *Betacoronavirus*), que incluye a SARS-CoV-2 y a diversos coronavirus relacionados circulantes en murciélagos. Sin embargo, el dominio

de unión al receptor (RBD) de la proteína Spike (S) de estos *Sarbecovirus* animales presenta divergencias relevantes respecto al de SARS-CoV-2, incluyendo diferencias en residuos clave implicados en la interacción con el receptor celular: enzima convertidora de angiotensina 2 (ACE2). En particular, tanto RaTG13 como BANAL-52 carecen de la inserción de cuatro aminoácidos (PRRA) en el sitio de clivaje por furina de la proteína S, característica distintiva de SARS-CoV-2 asociada a una mayor eficiencia en la entrada celular²². Además, se ha demostrado que la afinidad de unión entre el RBD de RaTG13 y el receptor humano ACE2 es decenas de veces inferior a la observada para el RBD de SARS-CoV-2, mientras que los RBD de algunos *Sarbecovirus* aislados de murciélagos en Laos (por ejemplo BANAL-52/103/236) muestran afinidades en el rango nanomolar, comparables a las de SARS-CoV-2²¹.

Análisis filogenéticos adicionales identificaron coronavirus derivados de pangolines (como PCoV-GD y PCoV-GX) que se agrupan dentro de este subgénero de *Sarbecovirus* de murciélago junto con SARS-CoV-2. Estos virus presentan una elevada similitud aminoacídica en el RBD (hasta ~96,8 %) y conservan varios de los residuos críticos implicados en la interacción con ACE2²³. En conjunto estos hallazgos respaldan la hipótesis de que SARS-CoV-2 podría haberse originado a partir de eventos de recombinación entre coronavirus relacionados que circulan en murciélagos y pangolines. Asimismo, otros coronavirus identificados en murciélagos del sudeste asiático muestran una alta similitud en regiones críticas del RBD y capacidad para utilizar ACE2 de distintas especies, incluido el humano²¹.

En conjunto, la evidencia genómica y filogenética disponible respaldan un origen zoonótico de SARS-CoV-2, probablemente asociado a uno o varios eventos de *spillover* desde reservorios animales, con posible participación de un hospedero intermedio aún no identificado. Sin embargo, el proceso preciso de emergencia, así como la secuencia de eventos evolutivos que precedieron a su adaptación y las etapas iniciales que permitieron la transmisión eficiente entre humanos, aún no están completamente esclarecidas. Esta incertidumbre se ve agravada por la disponibilidad limitada de información clave durante los primeros meses de la pandemia, incluyendo secuencias genómicas tempranas, datos sobre los animales comercializados en los mercados de Wuhan y documentación sobre actividades y condiciones de bioseguridad en laboratorios locales, lo que ha dificultado la reconstrucción detallada de los eventos asociados al origen del virus²⁴.

Organización genómica y ciclo replicativo

SARS-CoV-2 posee un genoma de ARN monocatenario de sentido positivo de aproximadamente 30 kb, uno de los de mayor tamaño entre los virus ARN. En el

extremo 5' del genoma se localizan los marcos de lectura abiertos ORF1a y ORF1b, que codifican poliproteínas procesadas en 16 proteínas no estructurales (nsP1–nsP16) involucradas en la replicación y transcripción viral. En la región 3' se encuentran los genes codificantes para las proteínas estructurales: spike (S), envoltura (E), membrana (M) y nucleocápside (N); junto con varias proteínas accesorias que modulan la interacción con el hospedero y la eficiencia replicativa ²⁵ (**Fig. 1.1**).

La proteína S desempeña un rol central en el tropismo y la entrada viral al mediar la unión al receptor celular ACE2 y la posterior fusión de membranas. Está formada por dos subunidades: S1, que contiene el dominio RBD y el dominio amino-terminal (NTD), y S2, responsable de la fusión viral. El RBD incluye el motivo de unión al receptor celular y, junto con el NTD, constituye el principal blanco de los anticuerpos neutralizantes. Tras la unión a ACE2, la proteína S es activada por clivaje proteolítico en el sitio S2', mediado por TMPRSS2 en la superficie celular o por catepsinas endosomales, lo que desencadena la fusión entre la envoltura viral y la membrana celular y la liberación del genoma viral al citoplasma ²⁶.

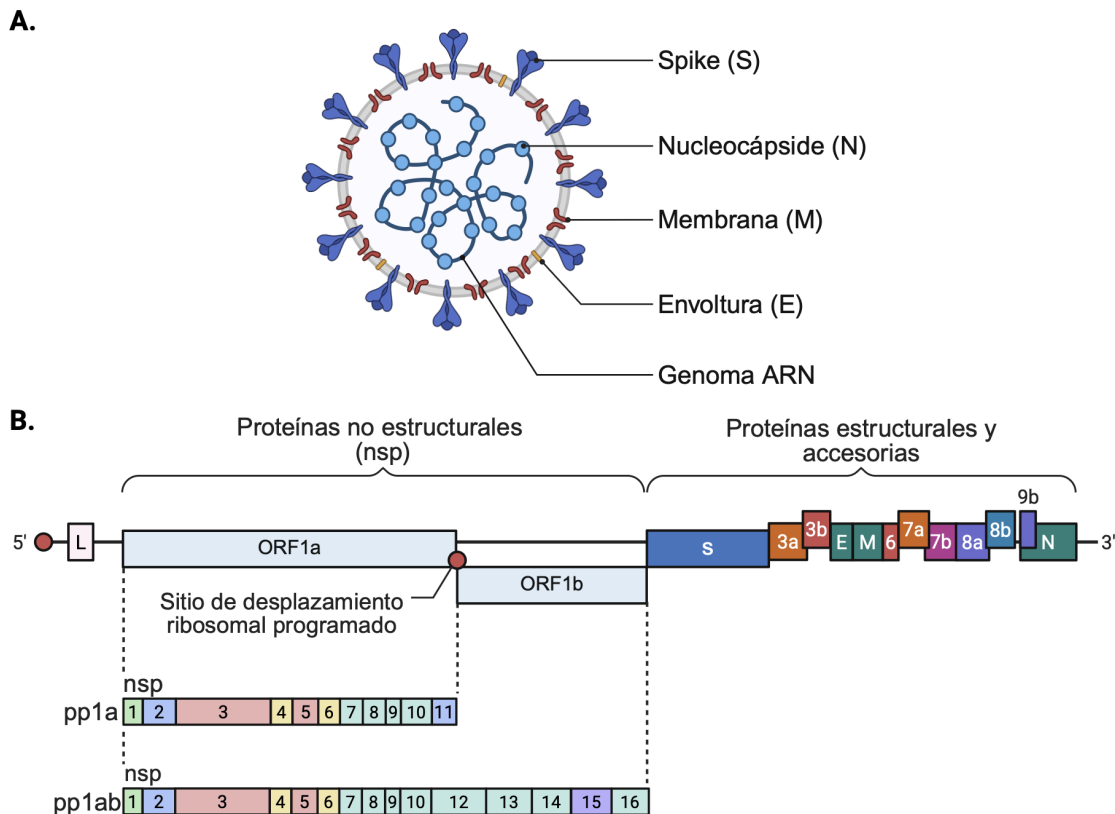


Figura 1.1: Estructura del virión y organización genómica de SARS-CoV-2.

A) Representación esquemática de la partícula viral, mostrando la envoltura lipídica que contiene las proteínas estructurales Spike (S), Membrana (M) y Envoltura (E), y en su interior la nucleocápside (N) asociada al genoma de ARN monocatenario de sentido positivo. **B)** Organización del genoma viral. En el extremo 5' se localizan los marcos de lectura ORF1a y ORF1b, traducidos mediante un desplazamiento ribosomal programado -1 para generar las poliproteínas pp1a y pp1ab, posteriormente procesadas en las proteínas no estructurales (nsP1-16) que conforman el complejo de replicación-transcripción. En la región 3' se encuentran los genes que codifican las proteínas estructurales (S, E, M y N) y las proteínas accesorias virales. Imagen adaptada de ²⁶ y creada en Biorender.com.

Una vez en el citoplasma, el ARN genómico actúa directamente como ARN mensajero, donde los ORF1a y ORF1b se traducen en las poliproteínas precursoras pp1a (de ORF1a) y pp1ab (de ORF1a/ORF1b mediante desplazamiento ribosomal programado -1), que luego son procesadas por las proteasas PLpro (nsP3) y 3CLpro (nsP5, también llamada Mpro) para generar las 16 proteínas nsPs. Este proceso da origen al complejo de replicación-transcripción, cuyo componente central es la RdRp, constituida por la subunidad catalítica nsP12 y los cofactores nsP7 y nsP8. Este complejo es responsable de la síntesis tanto del ARN genómico como de los ARN subgenómicos ²⁶.

Las nsPs se ensamblan en vesículas de doble membrana derivadas del retículo endoplasmático (RE), donde la RdRp sintetiza primero una hebra de ARN negativa complementaria y, a partir de ella, nuevos genomas virales y ARN subgenómicos. Estos últimos se traducen para producir las proteínas estructurales y accesorias. La proteína S se inserta en membranas del RE y del compartimento intermedio RE-Golgi, mientras que el ARN genómico positivo se asocia con la proteína N formando la nucleocápside. El ensamblaje de nuevos viriones ocurre en estas membranas que contienen las proteínas M, E y S, generando partículas virales completas que luego son transportadas en vesículas hacia la membrana plasmática y liberadas por exocitosis, para infectar nuevas células ²⁶.

Desde una perspectiva evolutiva, el ciclo replicativo impone fuertes presiones selectivas sobre la proteína S, determinante de la entrada celular y principal blanco inmunológico ^{14,27}. A su vez, la RdRp se ha caracterizado como la mayor responsable de la generación de diversidad genética como se detalla a continuación ²⁸.

Mecanismos moleculares de generación de variabilidad en SARS-CoV-2

La diversificación genética de SARS-CoV-2 surge principalmente como consecuencia directa de su proceso replicativo intracelular. En general, los virus ARN replican como poblaciones dinámicas o “nubes de mutantes”, compuestas por variantes estrechamente relacionadas genéticamente, debido a la baja fidelidad de las RdRp. En este contexto, el equilibrio entre generación de variabilidad y mantenimiento de la integridad genómica resulta crítico para la viabilidad viral. El gran tamaño del genoma de los coronavirus (~30 kb) los sitúa cerca del umbral de error característico de los virus ARN, límite máximo mutacional compatible con la viabilidad del genoma. Para compensarlo, estos virus han evolucionado generando mecanismos que reducen el número de mutaciones durante la replicación ^{29,30}. En particular, la fidelidad de la RdRp está modulada por un sistema de corrección de errores mediado por la exonucleasa nsP14 (ExoN), que disminuye significativamente la frecuencia de errores respecto a otros virus ARN, aunque no los elimina. En este contexto, la tasa de sustitución estimada para SARS-CoV-2 es del orden de 10^{-3} sustituciones por sitio por año, lo que se traduce, a nivel poblacional, en una acumulación aproximada de dos sustituciones por genoma por mes durante su circulación en humanos ^{28,31}.

La mayoría de las mutaciones generadas durante la replicación del genoma son neutras o deletéreas y no se fijan en la población viral. Sin embargo, algunas alteran propiedades funcionales como la afinidad por receptores celulares, la eficiencia

replicativa o la evasión inmunológica, pudiéndose expandir bajo selección positiva ³². Estas variaciones se han concentrado con mayor frecuencia en la proteína S, donde cambios en el dominio RBD o en regiones antigénicas afectan simultáneamente la infectividad y el reconocimiento inmunológico.

Además de las sustituciones puntuales (SNPs), la RdRp posee capacidad de cambio de molde durante la síntesis (*template switching*), lo que puede generar inserciones, deleciones y recombinaciones entre genomas virales. Este mecanismo puede producir cambios fenotípicos abruptos y contribuir a la aparición de variantes con múltiples mutaciones simultáneas ³³. Estos eventos, sin embargo, no ocurren al azar: ciertos sitios del genoma actúan como “puntos calientes” asociados a la estructura del ARN y a señales reguladoras de transcripción, sugiriendo que la arquitectura genómica condiciona la generación de diversidad ³⁴.

En conjunto, la variabilidad de SARS-CoV-2 surge de tres procesos interrelacionados: errores de replicación de la RdRp, recombinación por cambio de molde y restricciones estructurales del genoma viral. Tras su introducción en la población humana inicialmente *naive*, estos mecanismos generaron rápidamente un amplio repertorio de variantes sobre las cuales actuó la selección natural tanto a nivel intra- como inter-hospedero, favoreciendo la expansión de combinaciones de mutaciones con ventajas de transmisión o evasión inmune, dando lugar a la emergencia recurrente de linajes divergentes que caracterizaron la dinámica evolutiva de la pandemia.

Clasificación y vigilancia genómica de variantes virales

La pandemia de COVID-19 se desarrolló de manera concomitante con la expansión global de las tecnologías de secuenciación de próxima generación (NGS), permitiendo la obtención de millones de genomas virales en un corto período de tiempo. Este volumen de datos impulsó el desarrollo de sistemas estandarizados para clasificar y monitorear la aparición de nuevas variantes de SARS-CoV-2.

Entre ellos, Pango, Nextstrain y GISAID proporcionaron marcos filogenéticos para describir la diversidad viral. El sistema Pango (del inglés, *Phylogenetic Assignment of Named Global Outbreak Lineages*) asigna linajes dinámicos basados en relaciones de descendencia para rastrear cadenas de transmisión ³⁵. Nextstrain agrupa secuencias en clados definidos por mutaciones características y contexto temporal (sistema de año en el que emergió el clado y letra que diferencia el grupo dentro de ese año, ej. 20A, 21K) ³⁶, mientras que GISAID (del inglés, *Global Initiative on Sharing All Influenza Data*) clasifica grandes grupos genéticos de circulación global ³⁷. Dado que estas

nomenclaturas resultaban complejas para la comunicación pública, la OMS introdujo una nomenclatura basada en letras del alfabeto griego para designar variantes con impacto epidemiológico relevante, como las denominadas variantes de preocupación (VOCs, del inglés *Variant of Concern*)³⁸.

El seguimiento de estas variantes fue posible gracias a consorcios interinstitucionales nacionales e internacionales dedicados a la vigilancia genómica, que secuenciaron sistemáticamente muestras positivas y generaron millones de genomas virales depositados en las bases de datos públicas. Este enfoque permitió monitorear el desempeño de los métodos diagnósticos, reconstruir rutas de transmisión, identificar introducciones geográficas y analizar la dinámica temporal de expansión de las variantes³⁹⁻⁴².

En paralelo, se desarrollaron estrategias complementarias de diagnóstico molecular, como el desarrollo de RT-qPCR multiplex dirigidas a mutaciones características de las VOCs circulantes. Así, la combinación de secuenciación masiva y genotipificación molecular rápida permitió detectar tempranamente nuevas variantes y seguir su reemplazo poblacional en tiempo casi real, un escenario sin precedentes en vigilancia epidemiológica^{41,43,44}.

Emergencia y evolución de variantes durante la pandemia

Desde su emergencia a finales de 2019, SARS-CoV-2 acumuló mutaciones de forma progresiva durante su expansión global. Una de las primeras con impacto epidemiológico fue la sustitución A23403G, que produce el cambio de aminoácidos D614G en la proteína S y define el clado G (Nextstrain 20A y descendientes). La emergencia de esta mutación se estimó entre mediados y finales de enero de 2020, y se extendió rápidamente, en Europa y luego al resto del mundo, constituyendo uno de los primeros eventos de reemplazo poblacional asociados a una ventaja de transmisión sin un aumento marcado de la patogenicidad⁴⁵.

A fines de 2020 comenzaron a detectarse linajes con un número inusualmente alto de mutaciones y claras ventajas epidemiológicas, lo que llevó a la OMS a definir el concepto de VOCs, caracterizadas por presentar mayor transmisibilidad y cambios antigénicos con un impacto directo en la efectividad de medidas sanitarias (**Fig. 1.2**).

La primera VOC fue Alfa, identificada en el Reino Unido a finales del 2020, marcando el inicio de una diversificación adaptativa más acelerada⁴⁶. Poco después, surgieron de manera independiente las VOCs Beta en África austral y Gamma en Sudamérica, ambas con mutaciones convergentes en el RBD de la proteína S, indicando presiones

selectivas similares en diferentes poblaciones y regiones geográficas distantes ^{47,48}. Posteriormente, la variante Delta, detectada inicialmente en el sur de Asia, presentó un aumento sustancial de transmisibilidad y desplazó globalmente a las VOCs previas, constituyendo una nueva fase de adaptación viral ⁴⁹.

Finalmente, a fines de 2021 emergió Ómicron, caracterizada por una divergencia genética considerable respecto de variantes anteriores y por la presencia de más de 30 mutaciones en la proteína S, un sitio antigénico clave ⁵⁰. Esta variante dio origen a múltiples sublinajes y reemplazó rápidamente a Delta a escala mundial. En la **figura 1.2** se puede ver la emergencia y la diversificación de las VOCs Alfa (20I), Beta (20H), Gamma (20J), Delta (21I y 21J) y Ómicron (21M) durante el transcurso de la pandemia. En conjunto, todas las variantes dominantes derivaron del genotipo ancestral de SARS-CoV-2 clado 20A (Nextstrain) o B.1 (Pango), y se sucedieron temporalmente mediante procesos de reemplazo poblacional, reflejando episodios repetidos de adaptación viral como se observa en la **figura 1.3**. Este patrón evidencia una evolución no lineal, donde distintas combinaciones de mutaciones confirieron ventajas selectivas en distintos contextos epidemiológicos y de inmunidad poblacional.

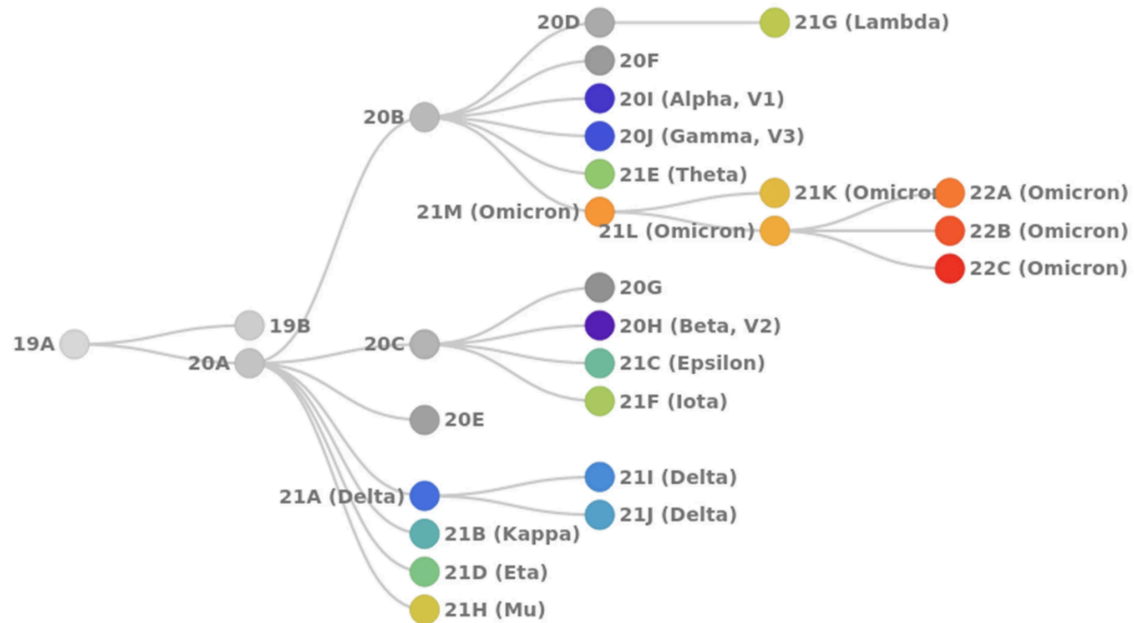


Figura 1.2. Emergencia y diversificación de VOCs hasta el año 2022. Árbol filogenético de los principales linajes de SARS-CoV-2 y VOCs basado en la nomenclatura de clados de Nextstrain, que agrupa los linajes de SARS-CoV-2 en clados evolutivos definidos por mutaciones compartidas y por su año de aparición. El diagrama parte del clado ancestral 19A (Dic. 2019) observándose la rápida emergencia del clado 20A (B.1 según Pango) e ilustra la diversificación progresiva del virus a partir de este clado. Se destacan las trayectorias evolutivas que condujeron a la aparición de VOCs Alfa, Beta, Gamma y Delta, y posteriormente, la expansión del clado 21M correspondiente a la VOC Ómicron (B.1.1.529), del cual derivó un gran número de sublinajes (no mostrado). En esta imagen se observan los sublinajes de Ómicron que emergieron hasta el 2022. Tomado de ⁵¹

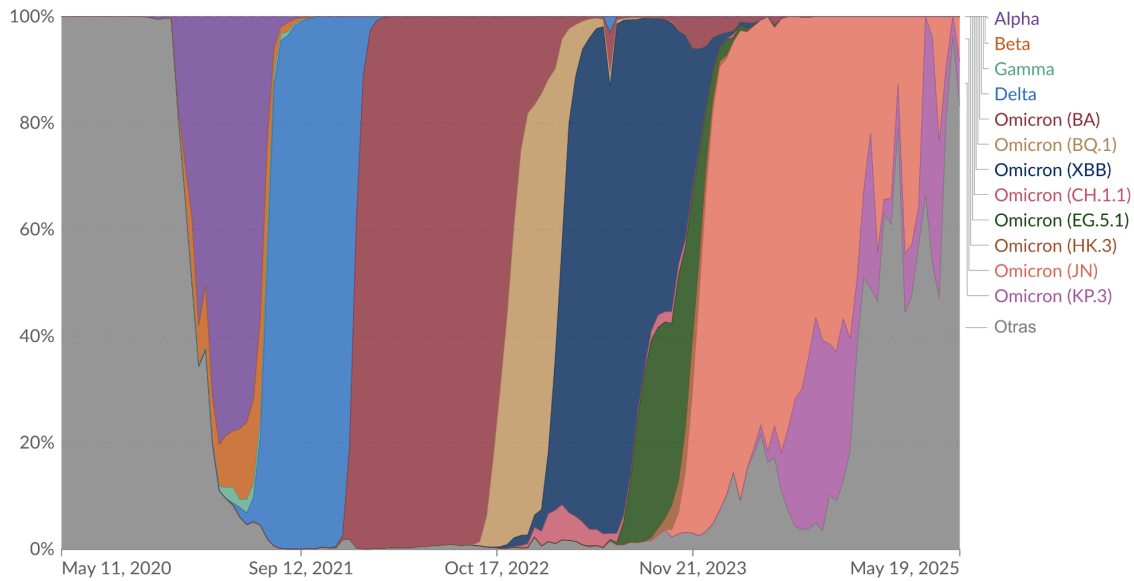


Figura 1.3. Dinámica temporal global de la frecuencia relativa de variantes de SARS-CoV-2. Gráfico de áreas apiladas que muestra la proporción de genomas secuenciados correspondientes a linajes virales de VOCs a nivel global a lo largo del tiempo hasta 2025. Se observa el reemplazo secuencial de VOCs durante la pandemia, desde Alfa, Beta, Gamma a finales de 2020 y comienzo de 2021, seguido por Delta y culminando en la dominancia de Ómicron y la diversificación de sus sublinajes (BA, BQ.1, XBB, CH.1.1, EG.5.1, JN, KP.3, entre otros). Imagen tomada de Our World in Data (consultado en febrero de 2026) a partir de secuencias depositadas en GISAID, vía CoVariants.org.

Es importante destacar que, como mencionamos anteriormente, la designación de “VOCs” fue introducida por la OMS únicamente para facilitar la comunicación pública sobre linajes asociados a cambios epidemiológicos relevantes. Sin embargo, con el avance de la vacunación y el aumento de la inmunidad poblacional producto de sucesivas olas de infección, este contexto cambió sustancialmente. Muchas variantes previamente clasificadas como VOC dejaron de cumplir los criterios de riesgo al disminuir sustancialmente la circulación del virus o perder impacto epidemiológico. Desde 2023 la OMS discontinuó progresivamente esta clasificación para los linajes derivados de Ómicron, y al momento de la redacción de esta tesis no existen variantes de SARS-CoV-2 que cumplan la definición operativa de VOC. Actualmente, la diversidad viral circulante corresponde principalmente a sublinajes evolutivamente relacionados cuya dinámica se explica por adaptación antigénica incremental más que por la aparición de linajes altamente divergentes (por ejemplo, las VOCs Beta y Gamma) como ocurrió al inicio de la pandemia ⁵².

Filogenia molecular y filodinámica en el estudio de brotes y epidemias virales

Los genomas virales constituyen registros históricos de la evolución de una población viral, ya que las mutaciones heredables se acumulan durante la transmisión entre hospederos, permitiendo que las diferencias genéticas entre aislamientos reflejen su parentesco evolutivo. De este modo, la comparación sistemática de secuencias permite reconstruir la historia evolutiva de poblaciones virales, la cual se representa mediante árboles filogenéticos, donde la topología indica las relaciones entre secuencias, los nodos terminales corresponden a los aislamientos observados y los nodos internos a ancestros comunes hipotéticos. En este marco, las variantes no necesariamente surgen de forma secuencial, sino que pueden originarse de manera independiente a partir de un ancestro compartido, dando lugar a trayectorias evolutivas paralelas ⁵³.

La reconstrucción filogenética puede realizarse mediante distintos enfoques, clasificados en métodos basados en distancias y métodos basados en caracteres. Los primeros estiman la distancia genética global entre pares de secuencias y utilizan la matriz resultante para inferir el árbol evolutivo, como en el método *neighbor-joining* (NJ). En contraste, los métodos basados en caracteres analizan directamente los cambios observados en cada posición del alineamiento, considerando cada sitio como un carácter evolutivo con distintos estados posibles (nucleótidos o aminoácidos). Entre estos se incluyen *Maximum parsimony* (MP), *Maximum likelihood* (ML) y la inferencia bayesiana, ampliamente utilizados para estimar el escenario evolutivo más probable ⁵³.

Independientemente del enfoque aplicado, la identificación de clados diferenciados permite distinguir entre descendencia directa y divergencia independiente, así como detectar procesos de diversificación intralínea.

En el contexto de brotes y/o epidemias virales, la reconstrucción filogenética constituye además la base de los análisis filodinámicos. Estos integran evolución viral, epidemiología y dinámica poblacional para inferir procesos de transmisión y diversificación a partir de datos genómicos. Este enfoque resulta particularmente útil en virus de ARN, cuyos genomas acumulan mutaciones a medida que se diseminan en la población.

Al incorporar información adicional, como la fecha y el lugar de muestreo, los análisis filodinámicos permiten calibrar los árboles filogenéticos en tiempo calendario mediante modelos de reloj molecular (para una revisión detallada de estos modelos, ver la bibliografía ⁵⁴). Estas aproximaciones permiten estimar la tasa evolutiva, reconstruir la historia temporal de divergencia entre linajes y aproximar la fecha del ancestro común

más reciente (tMRCA), además de inferir cambios en el tamaño poblacional efectivo y patrones de expansión viral durante una epidemia.

En conjunto, estos enfoques permiten estudiar la dinámica evolutiva de los virus en tiempo real, proporcionando información sobre el origen de los linajes, su diversificación y los procesos de transmisión que moldean la estructura genética de las poblaciones virales.

Análisis composicional de genomas virales: uso de codones y sesgos nucleotídicos

Además de la acumulación de mutaciones puntuales, los genomas virales presentan propiedades composicionales globales que reflejan la acción de distintos procesos evolutivos. Entre estas propiedades se encuentran la composición nucleotídica y los patrones de uso de codones, los cuales pueden variar entre virus, entre linajes dentro de una misma especie viral o incluso entre diferentes regiones del genoma.

El uso de codones se refiere a la frecuencia relativa con la que distintos codones sinónimos son utilizados para codificar un mismo aminoácido. Aunque múltiples codones pueden codificar el mismo aminoácido, estos no siempre se utilizan con la misma frecuencia, fenómeno conocido como sesgo de uso de codones. Este sesgo puede estar influido por diversos factores, incluyendo presiones mutacionales, selección a nivel translacional, adaptación al hospedador y restricciones estructurales del genoma viral ^{11,55}.

De manera relacionada, los sesgos en la composición nucleotídica, definidos como variaciones en el contenido de A, U/T, G y C en posiciones específicas del codón, pueden reflejar procesos evolutivos que actúan sobre escalas más amplias que las sustituciones individuales ⁵⁶⁻⁵⁸. El análisis de estas propiedades composicionales se ha utilizado para investigar procesos de adaptación viral, evolución de linajes y relaciones evolutivas entre virus.

En el caso de SARS-CoV-2, diversos estudios han analizado la composición nucleotídica y los patrones de uso de codones en comparación con otros *betacoronavirus*. En general, el genoma de este virus presenta un sesgo composicional caracterizado por un predominio de nucleótidos A y U y una marcada depleción de G y C. Este patrón se refleja en los codones preferenciales, que con frecuencia terminan en A o U, lo que sugiere que la presión mutacional constituye una de las principales fuerzas que moldean el uso de codones en SARS-CoV-2 ^{59,60}

La mayoría de los estudios evolutivos se han centrado en la identificación de mutaciones específicas asociadas a cambios fenotípicos. Sin embargo, el análisis de la arquitectura composicional del genoma viral puede aportar información adicional sobre las trayectorias evolutivas de distintos linajes. Este enfoque resulta particularmente útil en virus ARN, cuyas poblaciones evolucionan rápidamente, acumulando cambios a lo largo del tiempo. En particular, la variante Ómicron se caracteriza por presentar un número inusualmente elevado de mutaciones en la proteína Spike, lo que plantea interrogantes acerca de los procesos evolutivos que precedieron a su emergencia. En este contexto, el análisis de patrones de uso de codones y sesgos nucleotídicos en este gen podría aportar información adicional para caracterizar la singularidad evolutiva de este linaje.

Un estudio reciente que analizó más de 80.000 genomas recolectados entre 2020 y 2024 evidenció que la evolución del uso de codones no ocurrió de manera uniforme en todo el genoma viral. En los genes no estructurales ORF1a y ORF1ab necesarios para la replicación del genoma viral se observó una tendencia en el transcurso de los años hacia un uso de codones similar al usado por humanos, mientras que los genes estructurales mostraron una tendencia contraria. Estos resultados sugieren que la evolución del uso de codones en SARS-CoV-2 es dinámica y dependiente del gen, y podría estar moldeada por la interacción entre presión mutacional, selección y la dinámica epidemiológica de las variantes virales ⁶¹.

Hipótesis

La divergencia de la variante Ómicron no se explica únicamente por la acumulación de sustituciones puntuales en la proteína S, sino que refleja una trayectoria evolutiva diferenciada caracterizada por cambios en la composición nucleotídica y en los patrones de uso de codones del gen S. Estos patrones composicionales permitirían discriminar genotípicamente a Ómicron respecto a otras VOCs. Asimismo, análisis filodinámicos permitirían estimar su origen, emergencia y evolución a lo largo del tiempo.

Objetivo general

Analizar la evolución de SARS-CoV-2 mediante el análisis composicional del gen S y la emergencia de la variante de preocupación Ómicron.

Objetivos específicos

- 1- Caracterizar patrones evolutivos del gen S en las VOCs (Alfa, Beta, Gamma, Delta y Ómicron) de SARS-CoV-2 mediante el análisis comparativo de composición nucleotídica y uso de codones, con el fin de evaluar su asociación con la diferenciación genotípica entre linajes virales.
- 2- Reconstruir la historia evolutiva de Ómicron estimando su origen temporal y tasa de evolución molecular a partir de análisis filodinámicos.
- 3- Analizar la diversificación genética intralínea de Ómicron y su dispersión geográfica temprana, identificando subclados y patrones de expansión global.

Materiales y métodos

Secuencias

Se descargaron secuencias completas del gen S y genomas completos de SARS-CoV-2 desde la base de datos GISAID, seleccionando únicamente aquellas con: i) secuencia completa o casi completa, ii) alta cobertura, iii) ausencia de regiones ambiguas extensas, iv) fecha de muestreo completa y v) país de origen especificado.

A partir de las secuencias descargadas, se realizó un análisis comparativo de uso de codones y composición nucleotídica. Se incluyeron 159 secuencias del gen S correspondientes a la variante Ómicron (aisladas entre el 13/11/2021 y el 02/12/2021) y 97 secuencias representativas de las variantes Alfa, Beta, Gamma y Delta. Para el análisis filodinámico Bayesiano mediante cadenas de Markov Monte Carlo (MCMC) se utilizaron 89 genomas completos de Ómicron disponibles hasta el 2 de diciembre de 2021 que cumplían criterios de calidad y comparabilidad temporal.

Para cada secuencia obtenida se registraron metadatos: número identificador GISAID (*Accession ID*), fecha de aislamiento, país de origen, linaje Pango asignado, información de clado cuando correspondía. Los números de acceso, fechas de aislamiento y países de origen de todas las secuencias incluidas en los análisis se encuentran en el siguiente [link](#) (información suplementaria).

Alineamiento de secuencias

Las secuencias nucleotídicas fueron alineadas utilizando el programa MAFFT v7, aplicando los parámetros por defecto. Los alineamientos se inspeccionaron manualmente para verificar la correcta conservación del marco de lectura y excluir secuencias con regiones ambiguas o incompletas.

Asignación de genotipo de SARS-CoV-2

La asignación de linajes virales se realizó mediante la herramienta Pangolin, que permite clasificar las secuencias de SARS-CoV-2 dentro del sistema dinámico de nomenclatura Pango ³⁵. Esta aproximación permitió integrar la diversidad genética local y global y asegurar una clasificación coherente de los aislamientos analizados dentro de los linajes Alfa, Beta, Gamma, Delta y Ómicron.

Análisis de composición nucleotídica y uso de codones

Las frecuencias nucleotídicas y el uso de codones del gen S fueron calculados mediante el programa CodonW, implementado a través del servidor Galaxy (versión 1.4.4). Se analizaron las frecuencias nucleotídicas totales, así como las frecuencias por posición en el codón, junto con el uso relativo de codones. Para explorar la relación entre variables composicionales y variantes virales se realizó un Análisis de Componentes Principales (PCA), mediante descomposición en valores singulares (SVD). Se aplicó un escalado por varianza unitaria con el fin de que todas las variables contribuyeran de manera equivalente al análisis. Además, se construyeron mapas de calor (*heatmaps*) para visualizar los patrones de agrupamiento entre secuencias y variables composicionales, agrupando tanto filas como columnas mediante distancia de correlación y método de enlace promedio (*average linkage*). Estos análisis se realizaron con la herramienta ClustVis ⁶².

Análisis filodinámicos

La historia evolutiva de la variante Ómicron se reconstruyó mediante inferencia bayesiana basada en MCMC, implementado en el paquete BEAST v2.5.2 ⁶³. El modelo de sustitución nucleotídica, que mejor ajustó a los datos, se seleccionó con IQ-TREE, utilizando como criterios de selección el *Bayesian Information Criterion* (BIC), el *Akaike Information Criterion* (AIC) y el logaritmo de la verosimilitud (LnL), resultando seleccionado el modelo HKY+I, que considera frecuencias de bases desiguales, diferentes tasas de transiciones y transversiones, y una proporción de sitios invariantes. Se evaluaron tanto un reloj molecular estricto como un reloj relajado, junto con distintos modelos demográficos (tamaño poblacional constante, crecimiento exponencial, *Bayesian Skyline* y *Birth-Death Skyline Serial*). La incertidumbre estadística se estimó mediante intervalos de densidad de probabilidad más alta (HPD) al 95 %. La convergencia de las cadenas MCMC se verificó mediante valores de tamaño efectivo de muestra (ESS) superiores a 200 utilizando el programa Tracer v1.7.2, y la comparación entre modelos se realizó mediante el criterio AICM. Finalmente, los árboles de máxima credibilidad de clado (*Maximum Clade Credibility trees*) se generaron con TreeAnnotator ⁶³ y se visualizaron mediante FigTree v1.4.4 ⁶⁴.

Resultados

Caracterización de patrones evolutivos en el gen S de VOCs

En noviembre de 2021 se reportó la detección de una nueva variante de SARS-CoV-2, posteriormente designada por la OMS como Ómicron (21M, B.1.1.529), la cual mostró una rápida expansión global. Esta variante se caracterizó por presentar un número inusualmente elevado de sustituciones y deleciones en la proteína S, incluyendo mutaciones compartidas con VOCs previas (Alfa, Beta, Gamma y Delta) y otras exclusivas, lo que sugirió una trayectoria evolutiva prolongada previa a su emergencia e independiente.

Si bien numerosos estudios se han centrado en las consecuencias funcionales de estas mutaciones, la magnitud y distribución de los cambios observados plantean que la divergencia de Ómicron podría involucrar modificaciones globales en la arquitectura composicional del gen S. En este contexto, el análisis del uso de codones y de la composición nucleotídica permite evaluar diferencias evolutivas más allá de sustituciones puntuales.

Para ello, se analizaron las frecuencias de uso de codones en 256 secuencias del gen S correspondientes a las VOCs Alfa, Beta, Gamma, Delta y Ómicron. Se calcularon las frecuencias de uso de codones y se integraron mediante PCA y mapas de calor con *clustering*. Los resultados del PCA muestran una clara separación de Ómicron respecto al resto de las variantes en el plano PC1–PC2, indicando un patrón distintivo de uso de codones en el gen S asociado al genotipo (**Fig. 1.4**).

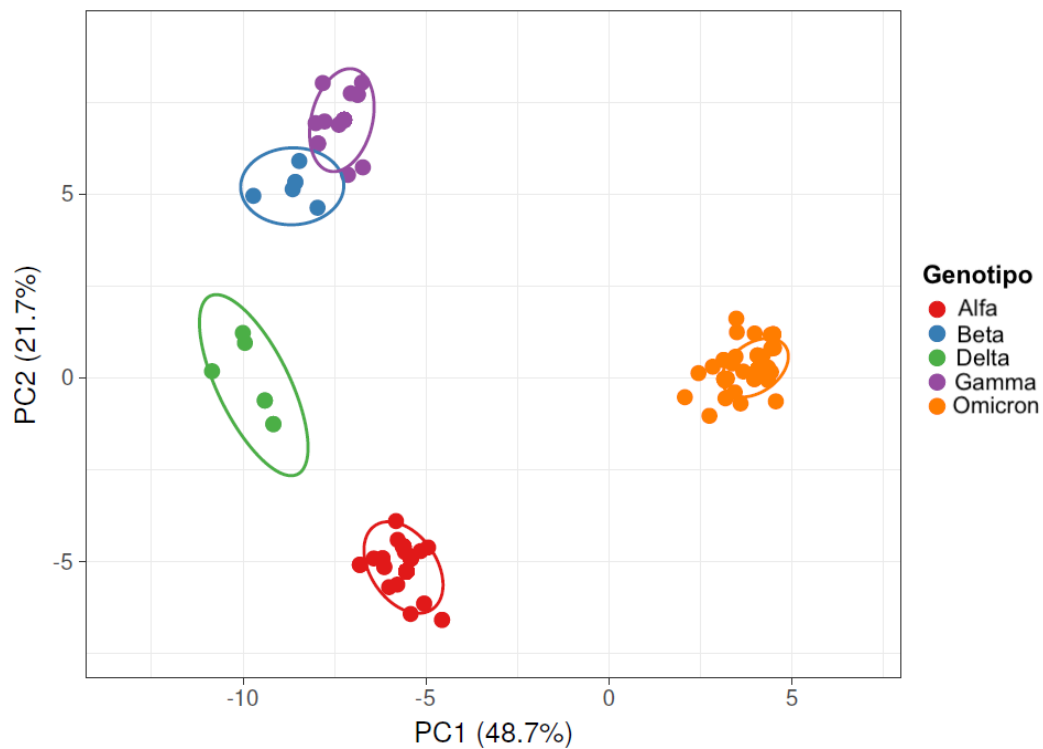


Figura 1.4: PCA del uso de codones en la proteína S de VOCs. Se realizó un Análisis de Componentes Principales (PCA) del uso de codones del gen S en $n=256$ secuencias de SARS-CoV-2. Las frecuencias fueron estimadas con CodonW y el PCA se calculó mediante descomposición en valores singulares (SVD) con escalado a varianza unitaria, método que permite una comparación equitativa entre linajes al normalizar la contribución de cada variable independientemente de la sobrerrepresentación de la variante Ómicron ($n=159$) frente al resto de las VOCs ($n=97$). Los dos primeros componentes explican el 48,7% (PC1) y el 21,7% (PC2) de la variación total. Se observa una separación clara de Ómicron respecto a las demás variantes. Las elipses representan el intervalo de confianza del 95%.

El mapa de calor y el análisis de clustering mostraron resultados consistentes con el PCA, evidenciando que Alfa, Beta, Gamma y Delta presentan mayor proximidad genética entre sí, mientras que Ómicron se posiciona como un grupo claramente diferenciado, sugiriendo una trayectoria evolutiva independiente. Asimismo, se observó cierta heterogeneidad intragrupo dentro de Ómicron, lo que sugiere variabilidad composicional entre los aislados analizados (**Fig. 1.5**).

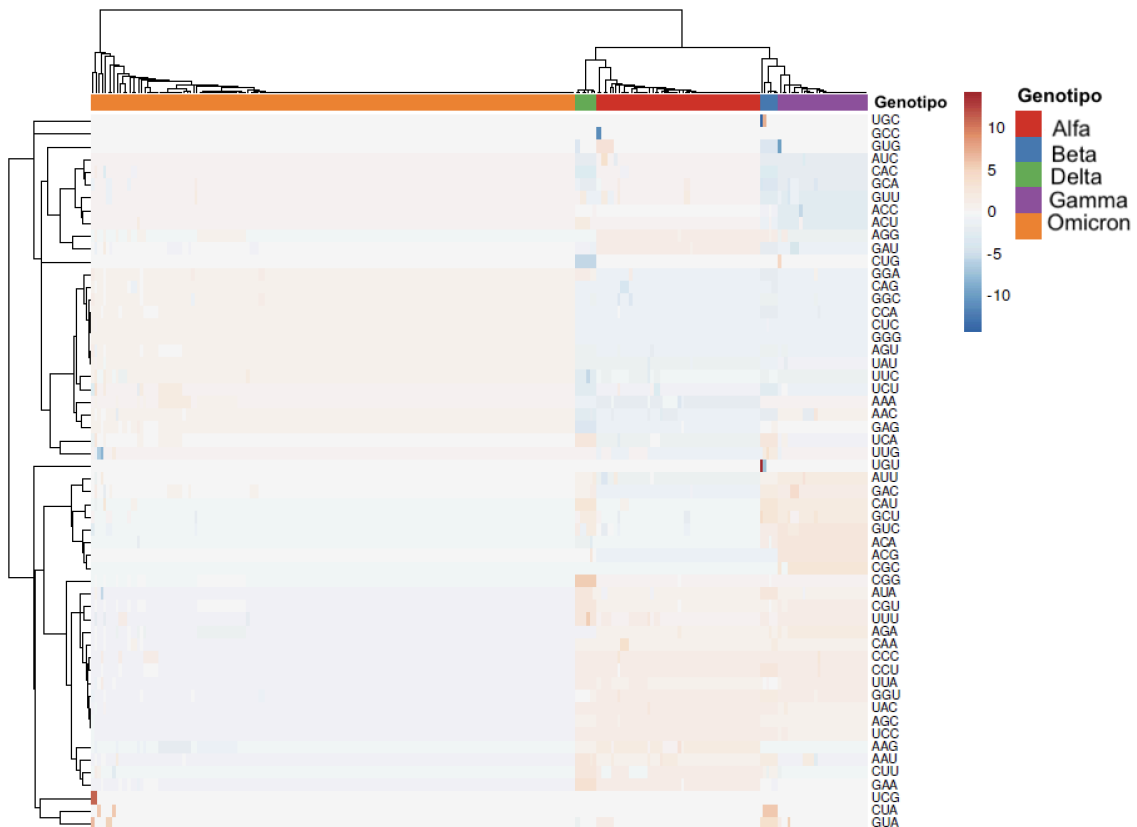


Figura 1.5: Mapa de calor del uso de codones en la proteína S de VOCs. Mapa de calor del uso de codones del gen S en las mismas n=256 secuencias analizadas en la figura 1.4. Cada columna corresponde a una secuencia S diferente proveniente de VOCs, cuyo genotipo se indica en la parte superior de la figura. Los datos fueron escalados a varianza unitaria y agrupados mediante distancia de correlación y método de enlace promedio (*average linkage*) utilizando ClustVis. Esta metodología estadística asegura que la identificación de clústeres se base en patrones de uso de codones, evitando que la mayor proporción de secuencias de Ómicron (n= 159) desvíe el agrupamiento jerárquico.

Con el fin de evaluar si estos patrones evolutivos podrían estar asociados a la composición nucleotídica, se analizaron las frecuencias de A, C, U y G en cada posición del codón en los genes S completos de las variantes estudiadas. Los análisis de PCA y mapas de calor evidenciaron la presencia de un sesgo significativo en la composición nucleotídica entre las proteínas S de las distintas VOCs, consistente con los patrones observados en el uso de codones (**Fig. 1.6 y 1.7**).

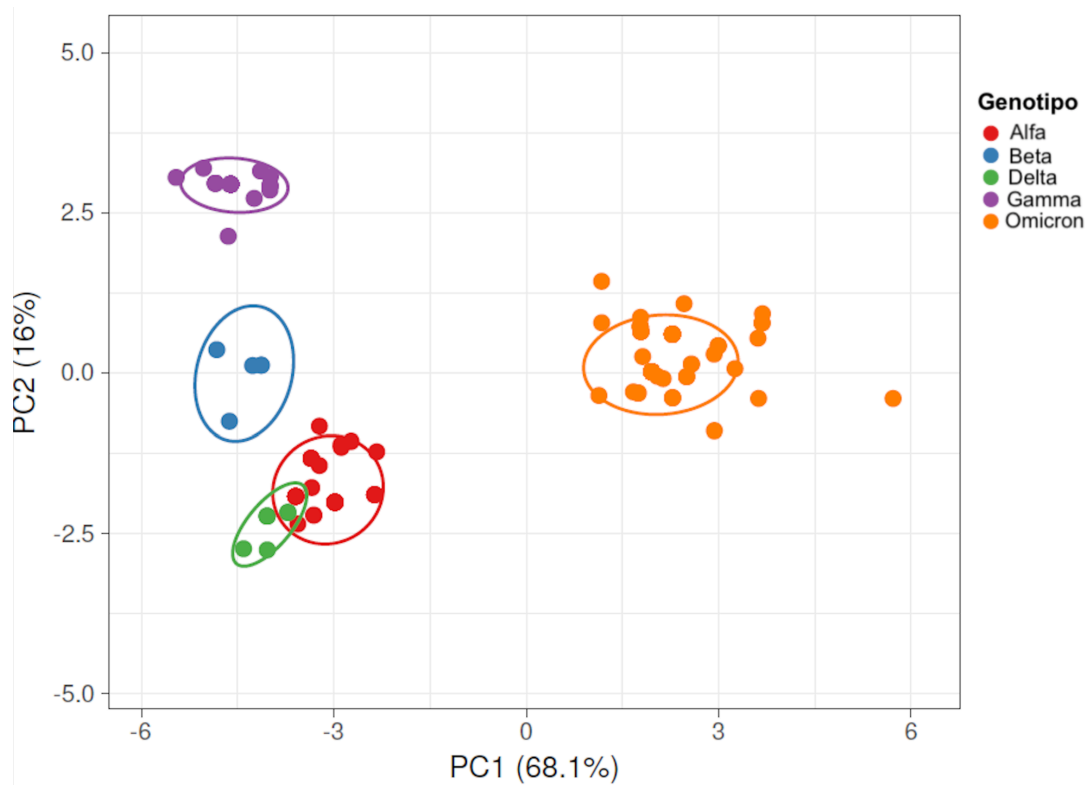


Figura 1.6: PCA de la composición nucleotídica en la proteína S de VOCs. Para analizar diferencias en la composición nucleotídica dentro del gen S entre VOCs, se realizó un PCA de las frecuencias de A, C, U y G calculadas en primera, segunda y tercera posición de codón (N=256; Ómicron n=159; Alpha/Beta/Gamma/Delta n=97). Las frecuencias se estimaron con CodonW y el PCA se calculó mediante SVD con escalado a varianza unitaria, lo que permite una contribución equivalente de cada variable y mitiga el efecto del tamaño desigual de los grupos. Se muestran PC1 (68,1%) y PC2 (16%) y elipses de predicción al 95% para cada genotipo.

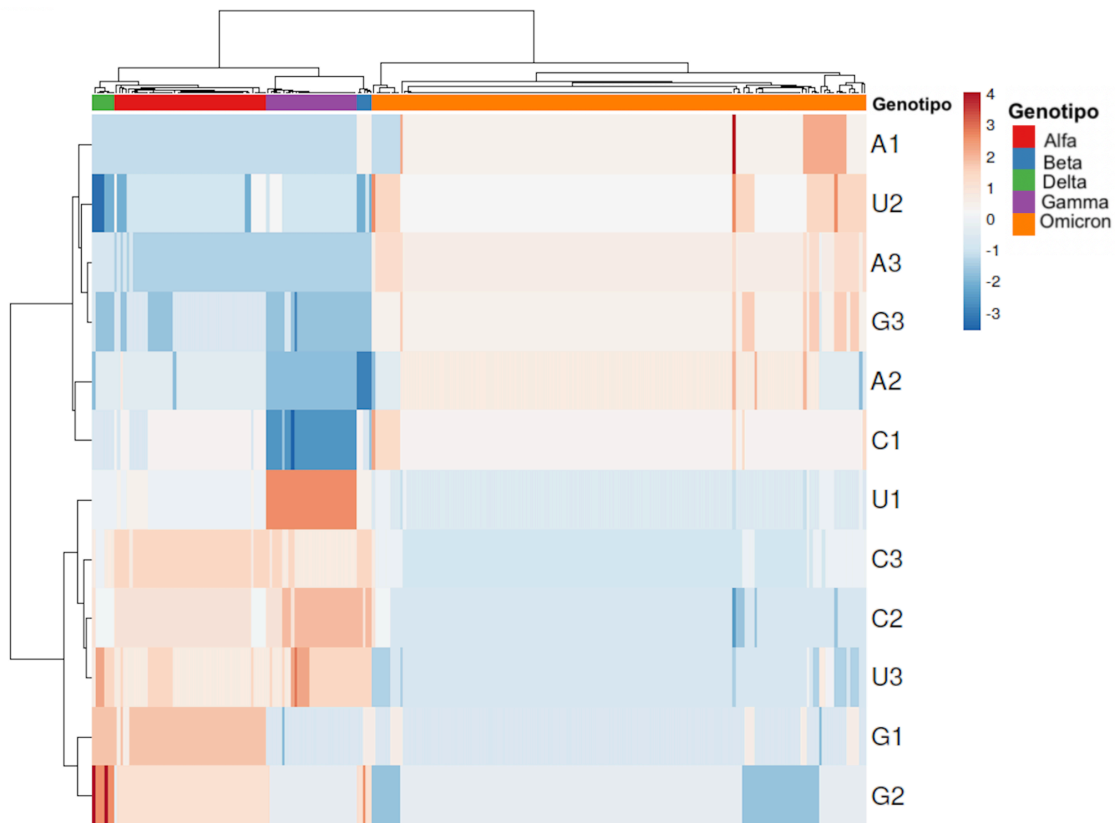


Figura 1.7. Mapa de calor de las frecuencias nucleotídicas por posición de codón en la proteína S de VOCs. Se calcularon las frecuencias de A, C, U y G en la primera, segunda y tercera posición de cada codón del gen S en un total de 256 secuencias ($n = 256$) y se representaron mediante un mapa de calor. Las posiciones de codón se indican como 1, 2 y 3. Los datos fueron escalados a varianza unitaria y agrupados mediante distancia de correlación y método de enlace promedio (*average linkage*) utilizando ClustVis. Esta metodología estadística asegura que la identificación de clústeres se base en patrones de uso de codones, evitando que la mayor proporción de secuencias de Ómicron ($n= 159$) desvíe el agrupamiento jerárquico.

Análisis de coalescencia bayesiana de las cepas Ómicron

Para reconstruir la historia evolutiva de Ómicron, se realizó un análisis coalescente Bayesiano MCMC utilizando 89 genomas completos y comparables disponibles al 2 de diciembre de 2021 correspondientes al linaje B.1.1.529 (clado 21M).

En la **tabla 1.1** se muestran los resultados del análisis de coalescencia bayesiana, producto de 20 millones de pasos del MCMC, utilizando el modelo de sustitución nucleotídica HKY+I, un reloj molecular estricto y el modelo de población *Birth-Death Skyline Serial*.

Los resultados indican que las cepas Ómicron analizadas derivan de un ancestro común que habría existido alrededor del 15 de mayo de 2020, lo que sugiere un periodo prolongado de circulación no detectada previo a su identificación en la población humana.

La tasa media de evolución estimada para las cepas de Omicron fue de 5.61×10^{-4} sustituciones/sitio/año (s/s/y) (valores de densidad de probabilidad alta del 95% de 3.051×10^{-4} a 9.014×10^{-4} s/s/y).

Por otra parte, estos estudios sugieren que los linajes directamente asociados a los brotes iniciales en Sudáfrica se originaron alrededor de octubre de 2021, poco antes de su detección (noviembre de 2021).

Tabla 1.1 Inferencia coalescente bayesiana de las cepas Ómicron de SARS-CoV-2.

Grupo ^a	Parámetro	Valor ^b	HPD ^c	ESS ^d
Omicron SARS-CoV-2	Posterior	-41527.05	-41604.75 to -41432.15	214.50
Secuencias completas	Prior	318.78	255.70 to 397.14	209.40
	Probabilidad del árbol	-41845.83	-41863.00 to -41830.62	314.10
	Tasa de evolucion ^e	5.613×10^{-4}	3.051×10^{-4} to 9.014×10^{-4}	208.10
	tMRCA ^f	1.539	0.806 to 2.251	210.70
		15 Mayo, 2020		
	tMRCA Sudáfrica ^g	0.310	0.120 to 0.570	
		10 Octubre, 2021		

^aLas cepas incluidas en el análisis de inferencia coalescente bayesiana se encuentran en la información suplementaria en el siguiente [link](#). ^bEn todos los casos se muestran los valores medios. ^cHPD, valores de alta densidad de probabilidad. ^dESS, tamaño efectivo de la muestra. ^eLa tasa de evolución se indica en sustituciones/sitio/año. ^ftMRCA, el tiempo del ancestro común más reciente se muestra en años. ^gtMRCA Sudáfrica, tiempo hasta el ancestro común más reciente para todas las cepas aisladas en Sudáfrica. La fecha estimada para los tMRCA se indica en negrita.

El análisis filogenético basado en árboles de máxima credibilidad de clados (MCC) reveló que las secuencias de Ómicron aisladas en Sudáfrica se agrupan en dos clados genéticos principales, evidenciando una diversificación temprana de Ómicron en ese país tras su emergencia (**Fig. 1.8**, mostrados en azul y rojo).

Adicionalmente, se observa un *clúster* diferenciado conformado por secuencias de Malasia e India, lo que podría reflejar introducciones independientes o eventos de dispersión temprana hacia Asia.

Dentro del *clúster* principal sudafricano se agrupan secuencias provenientes de múltiples regiones (Singapur, Hong Kong, Europa, Brasil y EE. UU.), lo que sugiere una rápida dispersión internacional del linaje tras su emergencia.

La topología del árbol, caracterizada por una estructura relativamente poco profunda y con distancias genéticas cortas entre secuencias, es consistente con un evento de emergencia reciente seguido de una rápida dispersión global. En conjunto, estos

resultados apoyan un escenario de diversificación temprana en Sudáfrica, acompañado de múltiples eventos de exportación a distintos continentes en un corto período de tiempo.

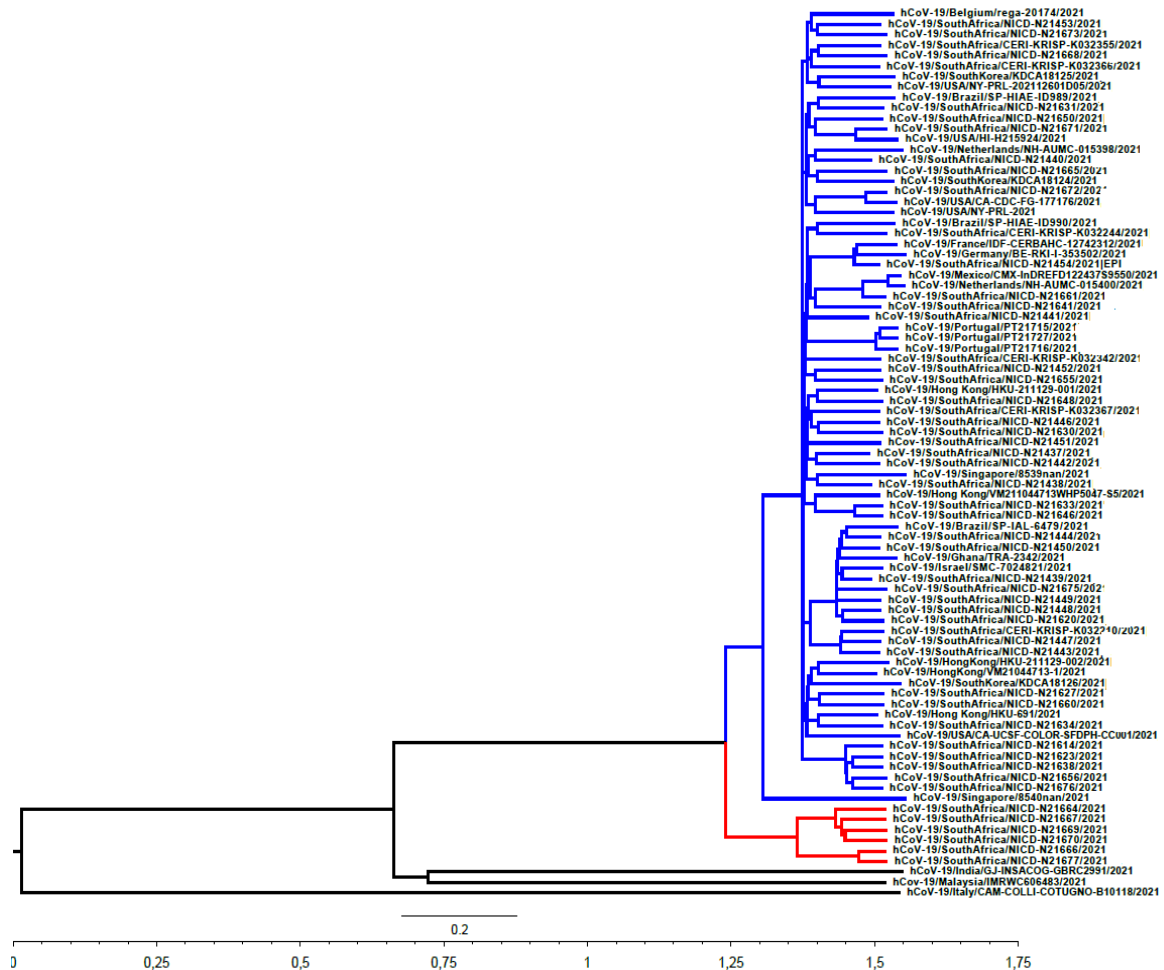


Figura 1.8. Inferencia filogenética de Ómicron B.1.1.529 (21M) hasta diciembre de 2021 mediante una aproximación bayesiana basada en MCMC. Se muestra un árbol de máxima credibilidad de clados (MCC) obtenido utilizando el modelo HKY+I, un reloj molecular estricto y el modelo de población *Birth-Death Skyline Serial*. El árbol está enraizado en el Ancestro Común Más Reciente (MRCA). Los dos clados principales que contienen las cepas aisladas en Sudáfrica se muestran en azul y rojo, respectivamente. El tiempo hasta el MRCA se indica en años en la parte inferior de la figura. La barra en la base del árbol indica el tiempo en años.

Discusión

El 26 de noviembre de 2021, la OMS clasificó a Ómicron (B.1.1.529) como variante de preocupación tras evidenciarse un aumento sostenido de casos en Sudáfrica y su rápida detección en múltiples países. Esta expansión temprana, junto con el elevado número de mutaciones acumuladas en la proteína S, dio lugar a la hipótesis de una trayectoria evolutiva prolongada e independiente, potencialmente asociada a infecciones crónicas en individuos inmunocomprometidos ⁵⁰.

Las primeras secuencias de Ómicron revelaron más de 30 sustituciones y múltiples deleciones en la proteína S, incluyendo mutaciones compartidas con VOCs previas y un conjunto extenso de sustituciones e inserciones exclusivos en la proteína S, como A67V, L212I, ins214EPE, G339D, S371P, S375F, N440K, N466S, S477N, Q493R, Y505H, T547K, N679K, N764K, D796Y, N856K, Q954H, N969K y L981F. Este perfil genético sugiere no sólo la acumulación de mutaciones adaptativas, sino también una divergencia evolutiva más profunda respecto a los linajes previamente circulantes ⁵⁰.

1. Análisis composicionales en el gen S de VOCs

Dada la magnitud de sustituciones presentes en la proteína S y su relevancia para la transmisión y evasión inmune, el análisis composicional de este gen ofrece una dimensión complementaria para comprender los procesos evolutivos asociados a la emergencia de Ómicron.

Los resultados obtenidos demuestran que el uso de codones y la composición nucleotídica contienen información suficiente para discriminar entre las VOCs y revelar la trayectoria independiente asociada a la emergencia Ómicron.

Nuestros estudios indican que los patrones de uso de codones y la composición nucleotídica (en cada posición del codón del gen S) se asocian de forma consistente con la diferenciación genotípica entre linajes virales, evidenciando que la divergencia observada no se limita a sustituciones aminoacídicas puntuales, sino que involucra una reorganización más amplia de la arquitectura genómica del gen S. Este enfoque se alinea con otros trabajos que han descrito sesgos marcados en la composición de nucleótidos y en el uso de codones en genes estructurales de SARS-CoV-2 y otros coronavirus, los cuales se relacionan con adaptación al hospedero y con diferencias en la historia evolutiva de cada linaje ^{60,65}.

Los análisis multivariados de uso de codones (PCA y mapas de calor) evidenciaron una asociación clara entre los patrones composicionales del gen S y los genotipos correspondientes a las distintas VOCs. En particular, las variantes Alfa, Beta, Gamma y Delta se agrupan estrechamente entre sí, mientras que Ómicron ocupa una posición

claramente diferenciada en el espacio composicional. Este patrón es consistente con una trayectoria evolutiva independiente y sugiere que el gen S contiene suficiente información evolutiva como para permitir la asignación de variantes emergentes a sus respectivos linajes, incluso a partir de variables composicionales globales, y refuerza la utilidad de este gen como marcador evolutivo para la caracterización de nuevas variantes.

Asimismo, la separación observada pudo explicarse en parte por sesgos en la composición nucleotídica. Las diferencias observadas en las frecuencias de A, C, U y G por posición de codón indican que los cambios composicionales son sistemáticos y no aleatorios, lo que apoya la hipótesis de que la divergencia del linaje Ómicron no se limita a la acumulación de mutaciones funcionales en sitios críticos del RBD, sino que podría deberse a una reorganización composicional más amplia del gen S.

Estos hallazgos resaltan la diversidad latente de SARS-CoV-2 y sugieren que los procesos evolutivos que conducen a la emergencia de nuevas variantes pueden involucrar tanto cambios adaptativos específicos como trayectorias composicionales diferenciadas. En el caso particular de Ómicron, su arquitectura composicional singular es compatible con escenarios de evolución prolongada bajo presiones selectivas particulares, como infecciones crónicas en individuos inmunocomprometidos, donde se ha documentado una rápida acumulación de mutaciones y una aceleración de la evolución intrahospedero.

En conjunto, los análisis composicionales revelan la presencia de una arquitectura genómica distintiva en el gen S de Ómicron, donde el análisis de uso de codones y sesgos nucleotídicos podrían complementar herramientas como la filogenia clásica para la identificación temprana y la caracterización evolutiva de variantes emergentes de SARS-CoV-2.

2. Análisis filogenéticos y filodinámicos de la variante Ómicron

El análisis coalescente bayesiano permitió estimar el origen temporal del linaje Ómicron y reconstruir su historia evolutiva reciente. Los resultados indican que este linaje evolucionó a partir de ancestros que existían hacia mediados de mayo de 2020, aunque su detección ocurrió recién en noviembre de 2021 en Sudáfrica. Esta estimación es consistente con estudios independientes que proponen una divergencia temprana del linaje previa a su identificación oficial ⁶⁶.

A nivel filogenético, las variantes sudafricanas de Ómicron se distribuyeron en al menos dos clados genéticos distintos, evidenciando una diversificación temprana intralínea. Además, la presencia de variantes aisladas en Singapur y Hong Kong dentro de uno de los principales linajes sudafricanos sugiere que la circulación del

virus pudo no haber estado restringida geográficamente a Sudáfrica, sino que pudo haber circulado también en múltiples regiones con capacidad limitada de vigilancia genómica.

Durante el proceso de revisión del manuscrito asociado al trabajo presentado en este capítulo, se evidenció una rápida diversificación del linaje de Ómicron. Durante ese período, el linaje B.1.1.529 (21M) derivó en los sublinajes BA.1 (compuesto por los primeros aislados en Botswana, Sudáfrica y en los demás países que fueron utilizados en el análisis de la **figura 1.8**), el sublinaje BA.2 (el cual fue ampliamente distribuido mundialmente tras su emergencia) y un tercer sublinaje BA.3 (minoritario y reservado a aislados predominantemente de Polonia). Por este motivo, se realizaron análisis filodinámicos adicionales al análisis realizado en la **figura 1.8**, incorporando secuencias de Ómicron BA.2 y BA.3 (**Tabla S1.1 y Fig. S1.1**). Se confirmó que estos pertenecen a linajes genéticos diferenciados, reforzando la noción de que Ómicron representa un complejo evolutivo heterogéneo más que una variante monolítica.

En conjunto, los resultados filogenéticos y filodinámicos respaldan un escenario en el cual la emergencia de Ómicron fue el resultado de una historia evolutiva caracterizada por una divergencia significativa con respecto a otros linajes de SARS-CoV-2. Este patrón pone de manifiesto la importancia de mantener sistemas de vigilancia genómica robustos, capaces de detectar tempranamente linajes altamente divergentes con potencial impacto epidemiológico.

Los resultados mostrados en este capítulo dieron lugar a la publicación: Paz, Mercedes *et al.* “An evolutionary insight into Severe Acute Respiratory Syndrome Coronavirus 2 Omicron variant of concern.” *Virus research* vol. 314 (2022): 198753. doi:10.1016/j.virusres.2022.198753, de la cual soy primera autora ⁶⁷.

Conclusiones

- El análisis composicional del gen S reveló diferencias significativas en el uso de codones y en la composición nucleotídica entre las VOCs de SARS-CoV-2, asociadas a su diferenciación genotípica.
- Las variantes Alfa, Beta, Gamma y Delta mostraron perfiles composicionales similares, mientras que Ómicron presentó una divergencia composicional consistente con una trayectoria evolutiva independiente.
- El gen S contiene información suficiente para discriminar variantes emergentes mediante análisis multivariados basados en variables composicionales, ampliando las herramientas disponibles para la caracterización de linajes virales.
- El análisis filodinámico bayesiano situó el origen del linaje Ómicron hacia mayo de 2020, lo que respalda la hipótesis de un período prolongado de circulación no detectada previo a su identificación oficial.
- Ómicron se diversificó tempranamente en múltiples clados y dio origen a sublinajes, evidenciando rápida heterogeneidad intralínea.
- La divergencia de Ómicron no se limita a la acumulación de mutaciones puntuales, sino que involucra cambios composicionales globales en el gen S.
- Estos resultados subrayan la relevancia de la vigilancia genómica y resaltan el valor de combinar análisis composicionales y filodinámicos para detectar linajes divergentes y comprender la dinámica evolutiva de virus ARN emergentes a escala poblacional.

Perspectivas

La evidencia de una trayectoria evolutiva diferenciada en la variante Ómicron plantea interrogantes relevantes acerca de los mecanismos que dan origen a linajes altamente divergentes en virus ARN. En este sentido, futuros estudios deberían profundizar en los procesos evolutivos que pueden favorecer la emergencia de variantes con características genómicas y composicionales distintivas, incluyendo el rol de infecciones crónicas, reservorios alternativos y dinámicas intra-hospedero.

En particular, resulta de interés analizar en mayor detalle los sublinajes derivados de Ómicron, evaluando cómo han evolucionado en el tiempo los patrones de composición nucleotídica y uso de codones desde su emergencia. Es esperable que estos sublinajes conserven rasgos composicionales heredados de su ancestro común, pero que también presenten divergencias progresivas asociadas a distintas presiones selectivas, lo que permitiría caracterizar con mayor resolución las trayectorias evolutivas dentro de este linaje.

Por otra parte, el volumen sin precedentes de secuencias generadas durante la pandemia ha permitido estudiar la evolución de SARS-CoV-2 con una resolución temporal y geográfica inédita. Sin embargo, esta abundancia de datos también introduce desafíos, particularmente asociados a la sobrerrepresentación de ciertas regiones geográficas y a las variaciones en la intensidad de muestreo a lo largo del tiempo. Asimismo, es importante considerar la evolución del sesgo de muestreo; mientras que en las primeras fases de la pandemia la vigilancia era comunitaria y masiva, actualmente se ha desplazado hacia un enfoque clínico, priorizando la secuenciación en pacientes hospitalizados o con cuadros graves. Esta transición también puede condicionar la representatividad de las secuencias disponibles en las bases de datos públicas. La evaluación sistemática de estos sesgos y su impacto en los análisis filogenéticos y filodinámicos constituye un aspecto clave para mejorar la interpretación de las trayectorias evolutivas.

Desde una perspectiva metodológica, la integración de variables composicionales con enfoques filogenéticos, filodinámicos y herramientas de aprendizaje automático podría contribuir al desarrollo de modelos capaces de identificar tempranamente linajes altamente divergentes o con potencial de expansión. Este tipo de aproximaciones integradoras podría complementar los sistemas actuales de vigilancia genómica, particularmente en contextos con capacidad limitada de secuenciación.

En conjunto, estas líneas de trabajo fortalecerán la comprensión de la emergencia viral y las estrategias de respuesta frente a futuros virus ARN con potencial pandémico.

CAPÍTULO 2

Rol de la proteína de dedos de zinc ZAP en la
infección por virus Mayaro: mecanismos de
restricción antiviral y presiones selectivas en un ciclo
replicativo dual

Introducción

Virus Mayaro

Clasificación

El virus Mayaro (MAYV) pertenece al género *Alfavirus* dentro de la familia *Togaviridae* e integra el complejo antigénico del virus Semliki Forest (SFV), en conjunto con los virus Chikungunya (CHIKV), Una (UNAV), Bebaru (BEBV), Getah (GETV), O'nyong-nyong (ONNV) y Ross River (RRV) ⁶⁸. Además de la alta similitud antigénica, los miembros de este complejo presentan una alta homología genómica (entre el 70% y 90% de identidad de nucleótidos), poseen genomas +ssRNA de replicación citoplasmática y un ciclo replicativo dual en hospederos vertebrados y vectores artrópodos ⁶⁹.

Dado su ciclo replicativo dual, ecológicamente MAYV se clasifica como un arbovirus (del inglés, *arthropod-borne virus*), término que define a aquellos virus que requieren de la replicación biológica en un artrópodo hematófago para su transmisión entre huéspedes vertebrados. MAYV se transmite a humanos a través de la picadura de mosquitos, manteniendo primordialmente un ciclo selvático o enzoótico donde los vectores del género *Haemagogus* (especialmente *H. janthinomys*) son los principales responsables de la transmisión ⁷⁰. No obstante, existe una creciente preocupación en salud pública debido al potencial de urbanización del virus, dada su capacidad experimental de infección y transmisión por vectores urbanos como el *Aedes aegypti* ⁷⁰.

Estructura genómica y ciclo replicativo

MAYV es un virus envuelto por una membrana lipídica externa que rodea la cápside viral, sobre la cual se proyectan las glicoproteínas de envoltura (E1 y E2) (**Fig. 2.1**). Posee un genoma +ssRNA de aproximadamente 11.5 kb, y su organización es característica de los alfavirus ⁷¹ (**Fig. 2.2**).

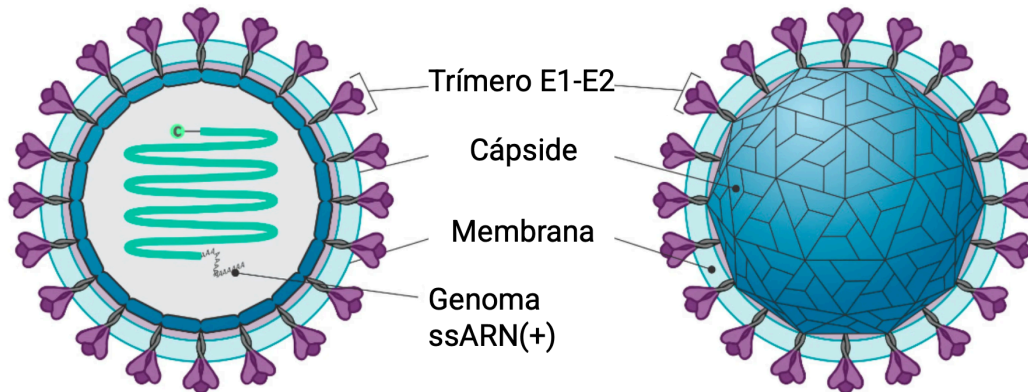


Figura 2.1: Representación de la estructura del virus Mayaro (MAYV). Se observa el virión de MAYV, compuesto por un trímero de glicoproteínas estructurales E1–E2, la proteína de la cápside y un genoma +ssRNA. El diseño de la figura está basado en la estructura del virión de MAYV maduro determinada por criomicroscopía electrónica (cryo-EM) (PDB ID: 7KO8) según ⁷². Imagen tomada y adaptada de ⁷³.

Su genoma posee un extremo 5' con un residuo de 7-metilguanósina (m7G) unido al primer nucleótido del ARN mediante un puente trifosfato (m7GpppN) característico de eucariotas, y una cola 3' poli-A, lo que le permite mimetizar el ARN mensajero celular (ARNm) y ser traducido inmediatamente por los ribosomas una vez dentro de las células del hospedero. El genoma codifica dos marcos abiertos de lectura (ORF, del inglés *open reading frame*). El primer ORF (ORF1) codifica para las proteínas no estructurales (nsP1-4), cuya función principal es la replicación y transcripción del genoma. El segundo ORF (ORF2) codifica las proteínas estructurales que conforman la cápside y la envoltura (C-E3-E2-6K-E1). Esta región se traduce a partir de un ARN subgenómico 26S que es sintetizado durante la infección, a partir de un promotor subgenómico ubicado entre los dos ORFs.

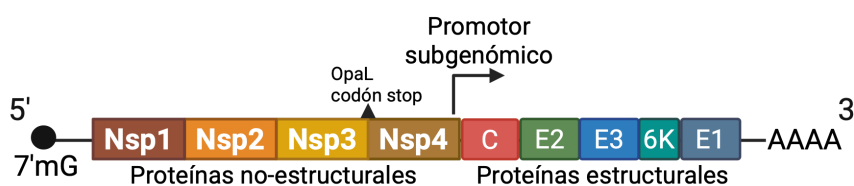


Figura 2.2. Estructura genómica de los alfavirus ⁷⁴. Imagen creada en Biorender.com

El ciclo replicativo de MAYV, similar al de otros alfavirus, comprende una serie de etapas definidas que incluyen entrada, traducción, replicación y liberación de nuevas partículas virales (**Fig. 2.3**).

La infección se inicia con la unión del virión a receptores celulares, principalmente a través de la glicoproteína E2, que reconoce al receptor Mxra8 (del inglés, *Matrix Remodeling Associated 8*), seguido de la internalización por endocitosis mediada principalmente por claritina^{75,76}. También se ha demostrado una posible vía alternativa a través de la endocitosis mediada por caveolinas⁷⁷. Mxra8 (también conocido como DICAM, ASP3 o limitrina) es una proteína de adhesión celular identificada como un mediador crítico de la entrada a la célula para alfavirus⁷⁸.

Una vez internalizada, la partícula viral queda confinada en un endosoma donde la acidificación del lumen permite la fusión de membranas y la liberación de la nucleocápside al citoplasma, donde ocurre el desensamblaje y la liberación del ARN genómico para iniciar la traducción. El genoma viral de ARN de sentido positivo es traducido inmediatamente, generando una poliproteína no estructural que es procesada en las proteínas maduras nsP1, nsP2, nsP3 y nsP4⁷⁹. Estas proteínas forman el complejo de replicación, inicialmente orientado a la síntesis de ARN de sentido negativo y posteriormente a la producción de ARN genómico y subgenómico (26S) de sentido positivo.

La replicación del genoma ocurre en invaginaciones de la membrana plasmática denominadas esférulos⁸⁰. Allí, nsP1 ancla el complejo a la membrana citoplasmática y cataliza la adición de la caperuza 5' al +ssRNA⁸¹. nsP2 actúa como helicasa y proteasa, desempeñando un papel central en la supresión de las respuestas antivirales del hospedero⁸². nsP3 media interacciones con proteínas del hospedero como G3BP y FXR1, contribuyendo a la formación de vacuolas citopáticas (CPVs) que albergan los complejos de replicación. nsP4 funciona como la polimerasa de ARN dependiente de ARN, esencial para la síntesis de nuevos genomas de ARN⁸³ y también posee actividad adenilil transferasa terminal (TATasa)⁸⁴.

El ARN subgenómico se transcribe a partir de un promotor interno y actúa como ARNm para las proteínas estructurales, cuya traducción genera una poliproteína procesada en cápside, E3, E2, 6K y E1⁸³. Estas proteínas son transportadas a través del retículo endoplásmico y el aparato de Golgi, donde las glicoproteínas se procesan e insertan en la membrana plasmática. Las proteínas de la cápside encapsidan el ARN genómico formando nucleocápsides, que interactúan con E1 y E2 en la membrana celular^{72,79,85}. La replicación culmina con la gemación de nuevas partículas virales, liberando viriones maduros e infecciosos⁷⁹.

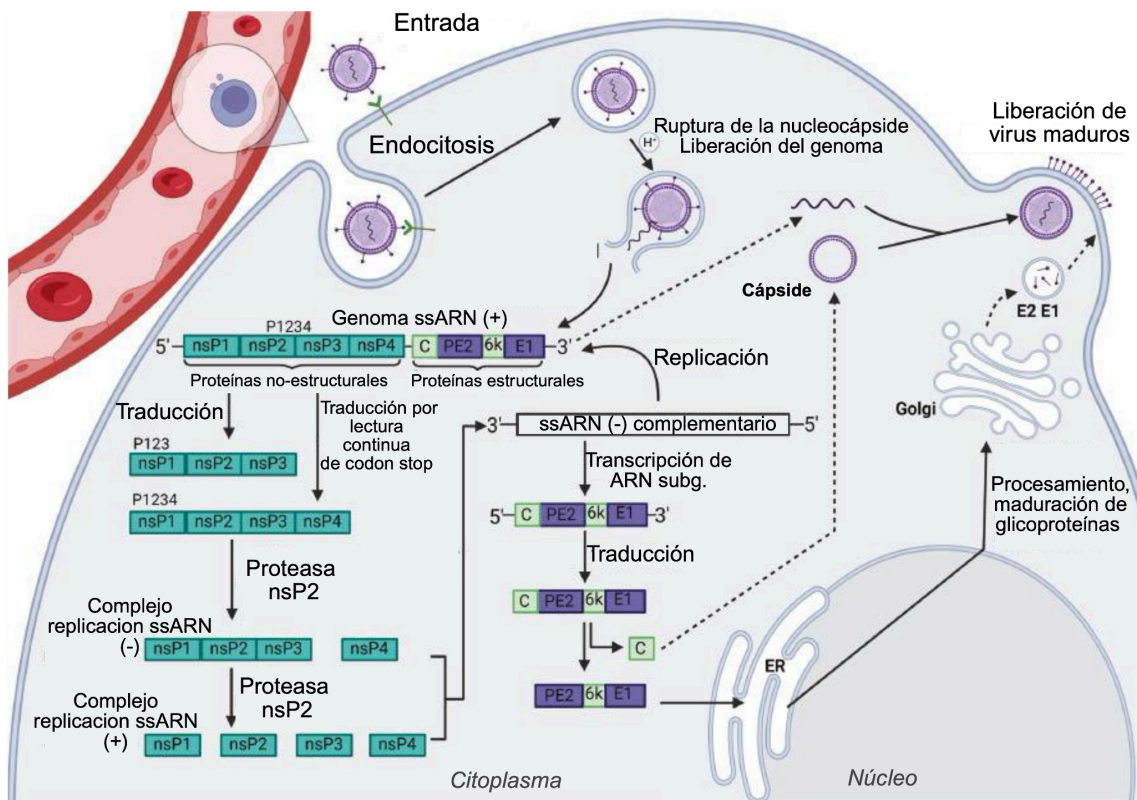


Figura 2.3 Ciclo replicativo de los alfavirus. Esquema de las principales etapas del ciclo replicativo de MAYV (1) unión a receptores celulares y entrada por endocitosis mediada por receptores; (2) traducción de proteínas virales y replicación del genoma viral; (3) ensamblaje del nucleocápside; y (4) gemación y liberación de nuevas partículas virales desde la célula hospedera. Adaptada de ⁷⁴.

Epidemiología y relevancia clínica

MAYV fue identificado por primera vez en 1954 en Trinidad y Tobago en trabajadores forestales con cuadros febriles, siendo posteriormente descrito como el agente etiológico de la fiebre de Mayaro ⁸⁶. La infección en humanos produce una enfermedad similar a otras infecciones por arbovirus como CHIKV y Dengue (DENV), caracterizada por fiebre aguda de 3 a 5 días, acompañada de cefalea, dolor retroorbitario, artralgia, mialgia, vómitos, diarrea y erupciones cutáneas ⁸⁷. Un subconjunto de pacientes desarrolla dolor articular crónico, y en casos severos pueden ocurrir complicaciones neurológicas, hemorragias y miocarditis ⁸⁸.

MAYV circula principalmente en América Central y del Sur, con mayor prevalencia en la Amazonia y regiones cercanas, aunque también se han reportado casos en áreas subtropicales ⁷³. La similitud clínica y la superposición geográfica con otros arbovirus como DENV y CHIKV dificultan el diagnóstico diferencial, contribuyendo al subregistro de casos y a la limitada disponibilidad de secuencias genómicas ^{73,89}.

Su principal vector es el mosquito del género *Haemagogus* ⁹⁰, aunque también se ha detectado en *Aedes aegypti* y *Aedes albopictus*, lo que sugiere una alta capacidad de

adaptación a distintos vectores ⁹¹. Además, factores como la globalización, el cambio climático, la deforestación y el aumento del contacto entre humanos, reservorios y vectores incrementan el riesgo de diseminación.

Aunque la transmisión urbana no está completamente establecida, se ha detectado ARN viral en *Aedes aegypti* en Brasil, y estudios experimentales han demostrado su susceptibilidad a la infección ^{92,93}. Estos hallazgos refuerzan la preocupación sobre el potencial de urbanización del virus y su relevancia como patógeno emergente.

En este contexto, el desarrollo de estrategias preventivas resulta clave. Actualmente no existen vacunas aprobadas contra MAYV, aunque se han evaluado diversos candidatos en fases preclínicas, incluyendo vacunas inactivadas, atenuadas, de ADN, vectores virales y VLPs (del inglés, *Virus-like particles*) ⁶⁹.

Respuesta inmune innata frente a MAYV

La infección por alfavirus activa la respuesta inmune innata mediante el reconocimiento de PAMPs por receptores PRRs (**Fig. 2.4**). El ARN viral es detectado por receptores endosomales tipo Toll (TLR3, TLR7 y TLR8) y por receptores citosólicos tipo RIG-I (RLRs), incluyendo RIG-I (*Retinoic acid-inducible gene 1*) y MDA5 (*Melanoma Differentiation-Associated gene 5*), los cuales inician cascadas de señalización antiviral que conducen a la activación de factores de transcripción como IRF-3/7 (*Interferon Regulatory Factor*) y NF- κ B (*Nuclear Factor kappa B*) ⁹⁴⁻⁹⁶. Más recientemente, un estudio ha reportado que receptores tipo NOD (NLRs), como NLRP3 y NLRP1, pueden formar inflamasomas que activan caspasa-1 y promueven la expresión de interleucinas proinflamatorias como IL-1 β en macrófagos infectados con MAYV ⁹⁷.

1. Entrada viral y reconocimiento de PAMPs virales por PRRs

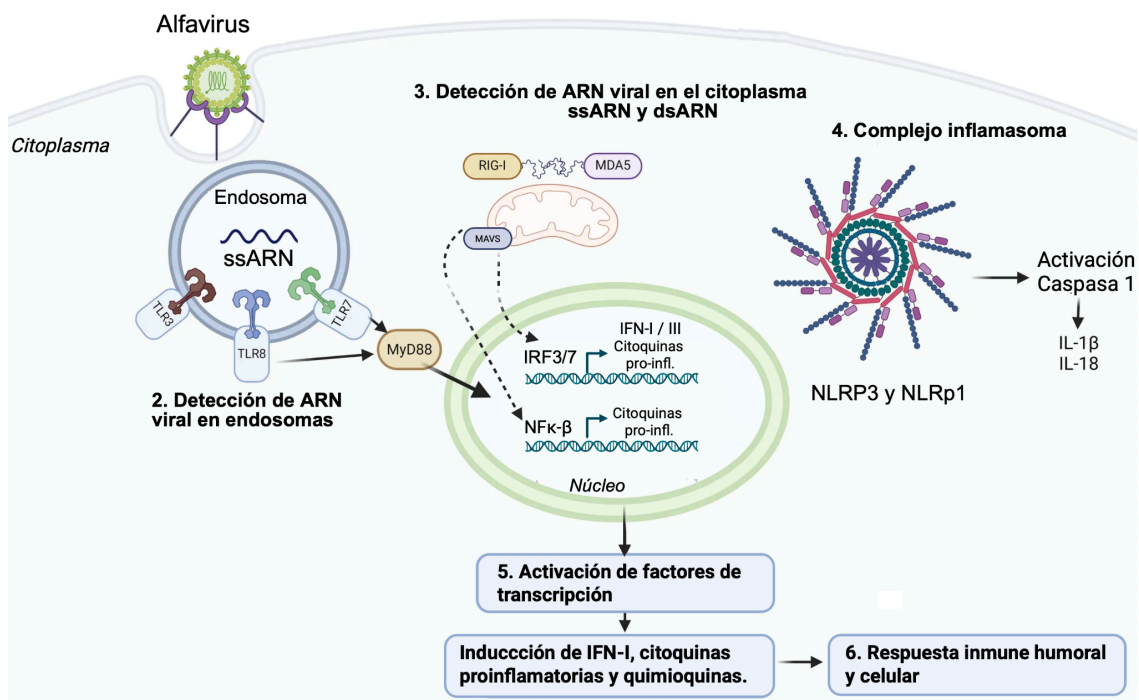


Figura 2.4: Respuesta inmune innata frente a la infección por alfavirus. Los alfavirus desencadenan una respuesta inmune innata mediada por diversos mecanismos moleculares. **1.** El proceso comienza tras la entrada viral, con el reconocimiento de PAMPs a través de PRRs. **2.** Esta detección es coordinada por receptores endosomales como TLR3, TLR7 y TLR8. **3.** Además, los receptores citoplasmáticos tipo RIG-I (RLR), específicamente RIG-I y MDA5, detectan ARN viral en el citoplasma celular que activa MAVS (del inglés, *Mitochondrial Antiviral Signaling protein*) en la mitocondria desencadenando la fosforilación y translocación nuclear de los factores NF- κ B o IRF3 y IRF7. **4.** También los receptores tipo NOD (NLR), como NLRP3 y NLRP1, promueven el ensamblaje de inflamasomas que activan la caspasa 1, facilitando así la liberación de citocinas proinflamatorias como la IL-1 β y la IL-18. **5.** La señalización derivada de los PRR culmina en una respuesta robusta que induce la expresión de IFN-I (IFN α/β), diversas citoquinas y quimioquinas. **6.** En conjunto, estas vías son determinantes para establecer una defensa antiviral eficaz, tanto a nivel humoral como celular. Imagen adaptada de ⁹⁸

La activación coordinada de estas vías culmina en la inducción de interferones tipo I (IFN-I) como IFN α e IFN β e interferones tipo III. Los IFN-I actúan mediante mecanismos autocrinos y paracrinos al unirse al receptor IFNAR (receptor de interferón α/β), lo que desencadena la vía JAK-STAT y promueve la expresión de una amplia batería de genes estimulados por interferón (ISGs) (**Fig. 2.5**). Junto con la producción de citoquinas y quimioquinas proinflamatorias, este proceso establece un estado antiviral crítico que no solo limita la replicación viral en etapas tempranas, sino que también orquesta la transición hacia una respuesta inmune adaptativa, tanto humoral como celular.

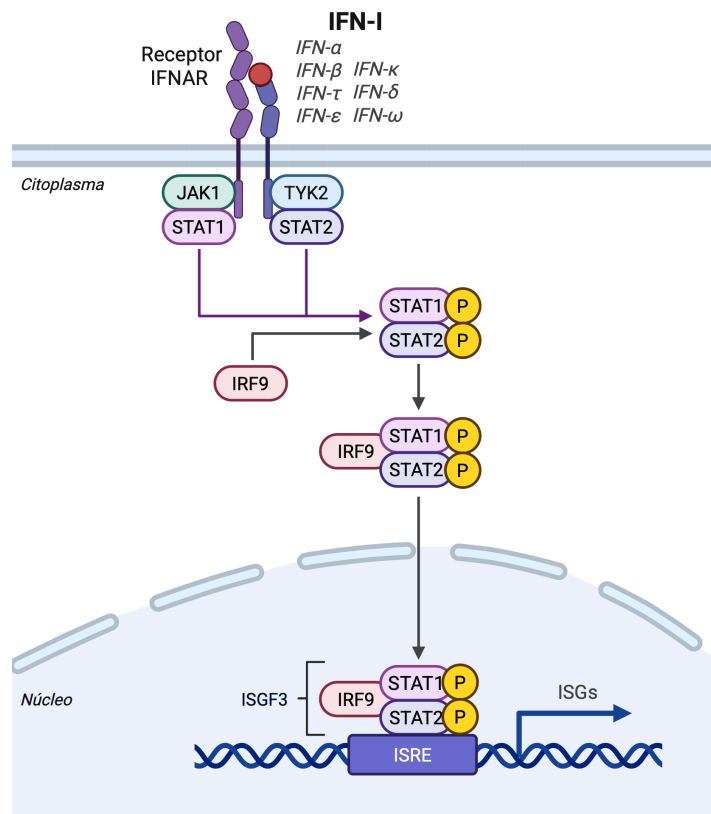


Figura 2.5: Cascada de señalización JAK/STAT desencadenada por IFN-I. El receptor transmembrana IFNAR (heterodímero IFNAR1/2) está acoplado a las enzimas JAK1 y TYK2. Estas quinasas se activan mediante la interacción entre los IFN y sus receptores, lo que les permite fosforilar las proteínas STAT1 y STAT2. Las proteínas STAT fosforiladas interactúan con IRF9 para formar el complejo ISGF3, el cual es transportado al núcleo donde se une a secuencias ISRE (*Interferon-Stimulated Response Element*). Esto induce la transcripción de ISGs, los cuales poseen dicha secuencia ISRE cerca de su promotor. Creado en biorender.com.

Estudios en modelos murinos y celulares han demostrado que la señalización por IFN-I es crítica para el control de MAYV *in vivo*, ya que su disrupción (por ejemplo en ratones IFNAR KO) conduce a viremia elevada, diseminación sistémica y mayor letalidad⁹⁹⁻¹⁰¹.

La infección por MAYV se asocia con una respuesta inflamatoria caracterizada por la producción de citoquinas y quimioquinas proinflamatorias, así como por el reclutamiento de células mieloides a tejidos afectados, contribuyendo tanto al control viral como a la inmunopatogénesis articular^{90,100}.

En conjunto, estos estudios posicionan a MAYV dentro del grupo de alfavirus altamente sensibles a la inmunidad innata dependiente de IFN-I y resaltan la relevancia de ISGs específicos como mediadores directos de la restricción viral. Entre

ellos, la proteína ZAP es un ISG que emerge como un factor central de esta respuesta, dado su reconocimiento preferencial de ARN enriquecido en dinucleótidos CpG y su potencial rol como presión selectiva en la configuración composicional del genoma viral, aspecto que se desarrolla en la sección a continuación.

Proteína antiviral de dedos de zinc (ZAP)

La interacción entre los virus y el sistema inmune innato de sus hospederos se caracteriza por una carrera armamentista evolutiva constante. En este contexto, la proteína antiviral de dedos de zinc (ZAP, del inglés *Zinc Finger Antiviral Protein*) codificada por el gen *ZC3HAV1* (del inglés, Zinc Finger CCCH-Type Containing, Antiviral 1) es un ISG que ha surgido en la última década como un factor antiviral crítico de vertebrados, capaz de sensar el ARN viral y restringir su replicación. A diferencia de otros sensores que detectan estructuras secundarias del ARN, como RIG-I o MDA5, ZAP reconoce ARN de cadena simple en el citoplasma, identificando específicamente motivos de dinucleótidos citosina-guanina (CpG), que actúan como firmas evolutivas del genoma viral. De este modo, ZAP logra diferenciar el ARN viral del ARN celular en un hospedero vertebrado, ya que estos hospederos presentan una muy baja frecuencia genómica de CpG ^{102,103}.

Identificación y especificidad

ZAP fue identificada en 2002 como un factor antiviral contra el virus de la leucemia murina (MLV) mediante un cribado de bibliotecas de expresión de ADN copia (ADNc) diseñado para buscar genes antivirales ¹⁰⁴. Posteriormente, se estableció como un componente clave de la respuesta inmune innata temprana frente a infecciones virales ^{105–107}. ZAP es un ISG, pero se ha reportado que su expresión se puede inducir directamente a través del factor IRF3, el cual se activa en respuesta a una infección viral o por la estimulación de células con análogos de ARN doble hebra (dsRNA) o ADN doble hebra (dsDNA), lo que ocurre río arriba de la señalización por interferón y de manera independiente de NF- κ B ¹⁰⁸.

ZAP ejerce una actividad antiviral de amplio espectro frente a numerosos virus mediante el reconocimiento específico de motivos CpG en el ARN viral, afectando tanto a virus de ARN como de ADN. Clásicamente, esta actividad se describió principalmente frente a virus +ssRNA y -ssRNA, así como virus ADN, mientras que los virus de dsRNA se consideraban en gran medida refractarios a su acción ¹⁰⁹. Sin embargo, estudios recientes han demostrado que ZAP también puede inhibir de manera eficiente algunos virus dsRNA, como el virus de la lengua azul (Bluetongue virus, BTV) y el virus de la enfermedad hemorrágica epizoótica (EHDV), aunque esta

restricción no es universal dentro de este grupo, ya que otros virus como el rotavirus (RV) no son sensibles a su acción ¹¹⁰.

La sensibilidad a la restricción por ZAP es altamente variable y depende de características específicas de cada virus como la composición y estructura del ARN, en particular la distribución de CpG. Esta variabilidad se observa incluso entre virus estrechamente relacionados o entre distintas cepas de un mismo virus, y puede estar modulada tanto por propiedades intrínsecas del genoma viral como por mecanismos virales que antagonizan la expresión o función de ZAP. Se ha reportado que ZAP restringe eficazmente la replicación del virus de la encefalitis japonesa (JEV) ¹¹¹, mientras que los virus DENV, ZIKA (ZIKV) y fiebre amarilla son resistentes a su acción ¹¹². Por otra parte, se ha descrito para citomegalovirus humano (HCMV) que la depleción de ZAP en las células aumenta significativamente la replicación de la cepa de alto pasaje AD169, pero apenas afecta a una cepa de bajo pasaje Merlin. Además, se observaron niveles diferenciales de expresión de ZAP en células infectadas, atribuibles a las distintas capacidades de estas cepas para modular la señalización de IFN-I e inducción de ZAP ¹¹³. De manera similar, en el herpesvirus simple (HSV), la proteína viral UL41 inhibe la expresión de ZAP mediante la degradación de su ARNm, interfiriendo con la restricción por esta proteína ¹¹⁴.

En conjunto, estos antecedentes indican que, aunque ZAP actúa como un factor antiviral de amplio espectro, su actividad es marcadamente dependiente del contexto viral.

Isoformas y regulación de la expresión transcripcional

ZAP presenta dos isoformas principales, ZAPL (902 aa) y ZAPS (699 aa), cuya producción está regulada a nivel del procesamiento del ARN, y predominan en la mayoría de los tipos celulares estudiados ^{104,115}. Más recientemente, mediante una caracterización exhaustiva de variantes de *splicing*, se identificaron isoformas adicionales ZAPXL (1024 aa) y ZAPM (821 aa), las cuales se originan por la combinación de un exón 4 extendido con mecanismos alternativos de terminación ¹⁰⁷ (**Fig. 2.6**).

ZAPL se genera a partir de la inclusión completa de los 13 exones del gen ZC3HAV1, mientras que ZAPS se origina por un proceso de poliadenilación alternativa a partir de una señal no canónica ubicada en el intrón 9 ^{115,116}. Esta poliadenilación alternativa es promovida principalmente por el factor de escisión CSTF2 ¹¹⁵, mientras que diversos factores de *splicing*, como las ribonucleoproteínas nucleares heterogéneas hnRNPA1/A2, las proteínas de unión a trectos polipirimidínicos PTBP1/2 y la

ribonucleoproteína nuclear pequeña U1 (snRNP) la inhiben, favoreciendo la producción de ZAPL ¹¹⁶.

La expresión de ZAP es inducida por la detección de PAMPs virales a través de los PRRs, lo cual activa al factor de transcripción IRF3, el cual se une al promotor de ZAP induciendo su expresión. Adicionalmente, la señalización mediada por interferones amplifica esta respuesta mediante señales autocrinas y paracrinas ¹⁰⁸. En este contexto, ZAPS y en menor medida ZAPM son inducibles por la señalización mediada por IFN-I, mientras que ZAPL presenta una inducción más moderada, ya que constituye la isoforma predominante en condiciones celulares basales ¹⁰⁷.

Además, se ha demostrado que el cofactor de ZAP TRIM25 (del inglés, *Tripartite Motif Containing 25*), una ligasa E3 de ubiquitina con capacidad de unión a ARN, puede favorecer la producción de ZAPS mediante la regulación del *splicing* alternativo hacia la expresión de esta isoforma, en el contexto de una infección por HCMV. En este estudio, se observó que el silenciamiento de TRIM25 reduce los niveles de ZAPS y aumenta concomitantemente ZAPL, evidenciando su rol en la regulación del equilibrio entre isoformas ¹¹⁷.

Diferencias estructurales y funcionales

Todas las isoformas de ZAP comparten un dominio N-terminal de unión a ARN compuesto de cuatro dedos de Zn de tipo CCCH (RNA-binding domain, RBD), responsable de su actividad antiviral mediante el reconocimiento específico de dinucleótidos CpG en el ARN viral ¹¹⁸ (**Fig. 2.6**).

La resolución estructural del dominio RBD de ZAP en complejo con ARN mostró que sus cuatro dedos de zinc generan una superficie rica en residuos básicos en la superficie de RBD, que permite la interacción con el ARN, donde el segundo dedo de zinc forma un bolsillo capaz de reconocer específicamente dinucleótidos CpG (**Fig. 2.7**) ^{119,120}. Además, se ha demostrado que este dominio es suficiente para mediar la restricción de diversos virus, como MLV, HBV y PRRSV ¹²¹⁻¹²³, lo que resalta su papel central en el reconocimiento de CpG y en la actividad antiviral de ZAP.

Las isoformas largas (ZAPXL y ZAPL) contienen además un dominio C-terminal tipo PARP (PARP, del inglés *poly(ADP-ribose) polymerase*), que ubica a ZAP dentro de la familia de las poly (ADP-ribosa) polimerasas, y se la nombra también como PARP13 (**Fig. 2.6**) ¹²⁴. Las proteínas PARP constituyen una familia de enzimas, implicadas en la señalización celular y la respuesta al estrés. Su función principal es llevar a cabo la poli-(ADP-ribosil)ación (PARilación), una modificación postraduccional en la que se añaden polímeros de poli(ADP-ribosa) (PAR) a proteínas diana, utilizando NAD⁺ como

sustrato. Sin embargo, el dominio PARP de ZAP es catalíticamente inactivo, ya que carece del residuo de ácido glutámico catalítico esencial (correspondiente a la posición 988 en PARP-1) y no presenta estructura compatible con la unión de NAD⁺, aunque sigue siendo importante para su actividad antiviral ^{124,125}.

Además, se ha demostrado que las diferencias estructurales entre ZAPS y ZAPL conllevan una distribución subcelular distinta para cada isoforma. ZAPL se asocia preferentemente a compartimentos membranosos, como el retículo endoplásmico y estructuras endolisosomales. Esto es debido a que ZAPL, dentro del dominio PARP, contiene un motivo CaaX S-farnesilado (prenilación S de la proteína), lo que aumenta la hidrofobicidad de las proteínas y mejora su afinidad por las membranas celulares, lo cual es crucial para su direccionamiento a vesículas intracelulares y membranas, así como para su actividad antiviral eficiente ¹⁰⁵. Por su parte, ZAPS, que carece del dominio PARP, se localiza mayoritariamente en el citoplasma. En consecuencia, ambas isoformas acceden a distintos *pools* de ARN viral y participan en mecanismos restrictivos parcialmente diferentes ¹¹⁵.

ZAPS es la isoforma preferencialmente inducida tras la estimulación con IFN-I, mediante el mecanismo de *splicing* alternativo regulado por CSTF2 ¹¹⁵. Además de su actividad antiviral directa, ZAPS modula la vía de señalización de RIG-I: puede potenciarla al unirse a RIG-I y promover su oligomerización y actividad ATPasa, aumentando la producción de interferón ¹⁰⁶; o bien atenuarla al unirse al extremo 3' UTR de los ARNm de IFN-I y promover su degradación ¹¹⁵. En línea con esto, células y modelos murinos deficientes en ZAP (KO) presentan respuestas de interferón más elevadas y sostenidas ^{115,126}. En conjunto, estos resultados evidencian que ZAP cumple un rol dual, dependiendo del contexto celular y experimental. Por ello, resulta fundamental dilucidar en qué contextos y mediante qué interacciones específicas ZAP potencia o restringe la respuesta inmunitaria.

Además de los dominios N- y C-terminales, todas las isoformas de ZAP comparten una región central que contiene un quinto dedo de zinc (ZnF5) y dominios WWE dispuestos en tándem (WWE1 y WWE2) (**Fig. 2.6**). En esta arquitectura, el bolsillo del dominio WWE2 es responsable de la unión a polímeros de poli(ADP-ribosa) (PAR), interacción que contribuye a potenciar su actividad restrictiva frente a ciertos virus ^{127,128}.

Asimismo, entre el dominio N-terminal de unión a ARN y el dominio central de ZAP se encuentra una región amplia que presenta características de desorden estructural (IDR, del inglés *intrinsically disordered region*) (**Fig. 2.6**). Este tipo de regiones

desordenadas en proteínas puede favorecer la separación de fases y la formación de orgánulos sin membrana, como los gránulos de estrés, que participan en la respuesta antiviral y en la supervivencia celular bajo condiciones de estrés. Estudios anteriores muestran que ZAP puede localizarse temporalmente en estos gránulos durante la replicación viral, sugiriendo que esta región desordenada podría estar relacionada con su función antiviral y su reclutamiento a dichos orgánulos^{129,130}.

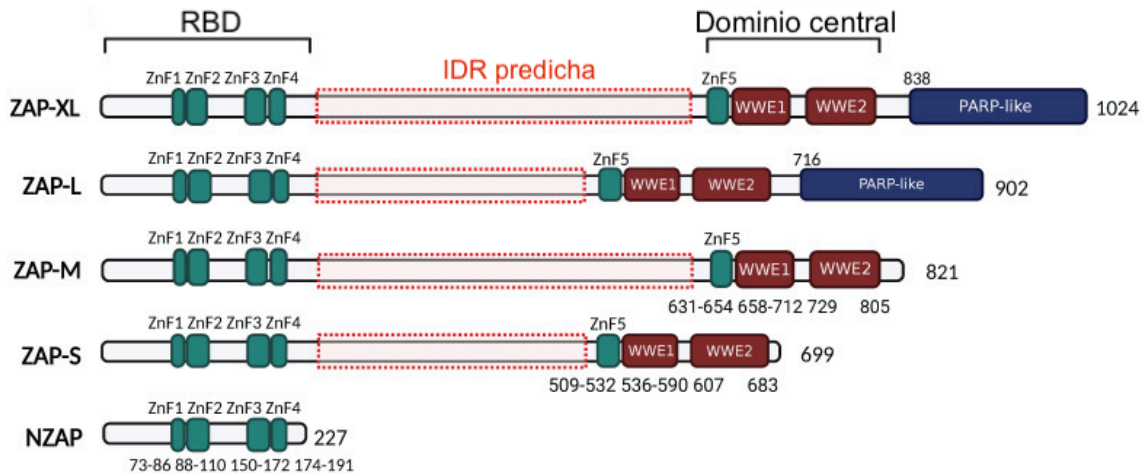


Figura 2.6: Isoformas de ZAP. Las isoformas comparten una región N-terminal dedicada a la unión de ARN (NZAP, aminoácidos 1–191), que contiene un dominio de unión a ARN (RBD, de inglés *RNA-binding domain*) compuesto de cuatro dedos de Zn (ZnF1-4) de tipo CCCH. Además, presentan un dominio central que contiene un quinto dedo de Zn (ZnF5) y dos motivos WWE dispuestos en tándem. Las isoformas ZAP-L y ZAP-XL incluyen adicionalmente un dominio tipo PARP. Las regiones intrínsecamente desordenadas predichas en las isoformas de ZAP (IDR, del inglés *intrinsically disordered region*) se señalan mediante recuadros discontinuos. Figura tomada y adaptada de¹⁰⁹.

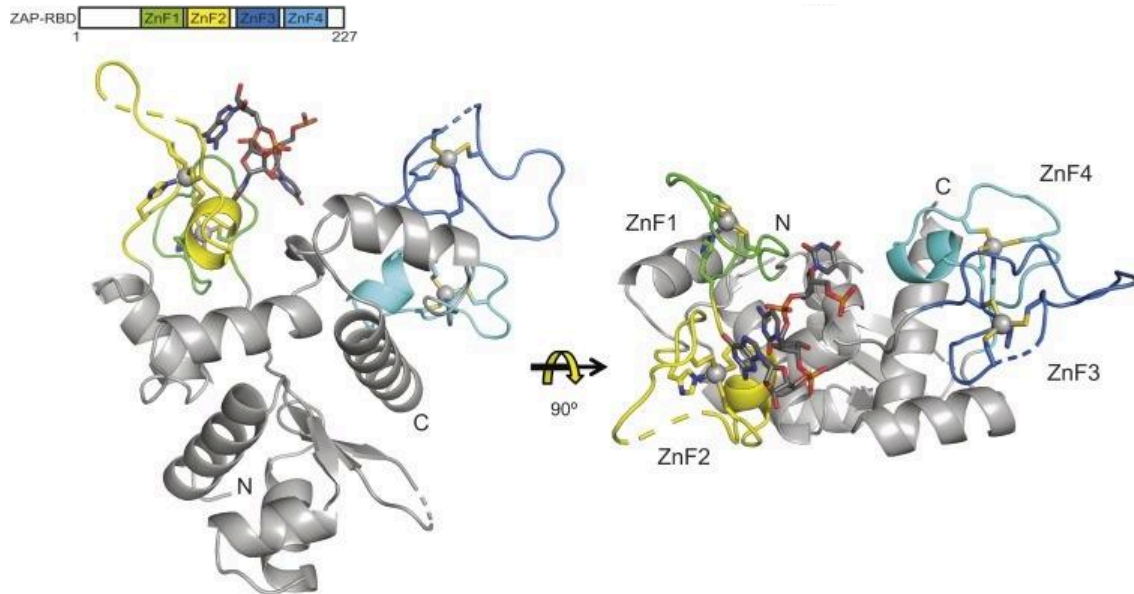


Figura 2.7: Dominio de unión a ARN (RBD) de ZAP humana en complejo con ARN. Representación estructural del complejo entre el dominio RBD de ZAP y el ARN. Los cuatro dedos de zinc (ZnF1–ZnF4) se muestran en colores diferenciados (verde, amarillo, azul y cian, respectivamente) en vistas laterales y superiores del RBD. El motivo CG del ARN se asocia específicamente con ZnF2. El trinucleótido UCG y los residuos que coordinan el zinc se ilustran en representación de varillas con codificación de colores atómicos, mientras que los iones Zn se indican como esferas. Adaptado de ¹²⁰.

Mecanismos moleculares de restricción viral mediados por ZAP

Tras unirse al ARN viral, ZAP inhibe la replicación viral mediante dos mecanismos: la inhibición de la traducción de proteínas virales y la promoción de la degradación del ARN viral, cuya contribución relativa depende del virus, del contexto celular y de la composición nucleotídica del genoma viral.

ZAP inhibe la traducción viral al interactuar con el factor de iniciación eIF4A, e interferir su asociación con la proteína eIF4G, impidiendo la formación del complejo de iniciación de la traducción ¹³¹. Este mecanismo se ha observado para distintos virus como VIH-1 y JEV ^{111,131}. En alfavirus, este mecanismo depende además del cofactor TRIM25, que potencia la actividad de ZAP mediante su ubiquitinación, favoreciendo la unión al ARN viral ¹³².

De forma complementaria, ZAP promueve la degradación del ARN viral a través de múltiples rutas. Por un lado interactúa con la deadenilasa PARN (poly(A)-specific ribonuclease) y con el complejo exosoma con actividad exonucleasa 3'→5' para mediar la deadenilación y la degradación del ARN viral desde el extremo 3' ^{111,121,133}. Este proceso mediado por PARN elimina la cola poli(A), dando lugar a una molécula de ARN menos estable y más susceptible a la degradación exosomal ¹³³. Por otra parte, ZAP puede reclutar el complejo DCP1A–DCP2 que elimina la caperuza 5' del

ARN viral, exponiéndolo a nucleasas 5'→3' como Xrn1^{133,134}. De este modo, ZAP coordina un programa en dos etapas para la degradación del ARN: una degradación 3'→5' mediada por el exosoma tras una deadenilación dependiente de PARN, y una degradación 5'→3' mediada por Xrn1 luego de la decaperuza dependiente del complejo DCP.

Además, la degradación del ARN viral mediada por ZAP requiere la acción de cofactores como TRIM25 y KHNYN, que incrementan significativamente la velocidad y la eficiencia de este proceso. KHNYN es una endonucleasa dependiente de magnesio con actividad antiviral frente a retrovirus como VIH-1 y MLV, cuya acción se asocia a su interacción con ZAP^{135,136}. KHNYN interactúa tanto con ZAPS como con ZAPL, y es independiente de la unión al ARN, siendo sus dominios KH y NYN esenciales para su actividad antiviral. Además, el silenciamiento de TRIM25 en líneas celulares que expresan ZAP reduce la actividad antiviral de KHNYN en la inhibición del VIH-1, lo que sugiere que TRIM25 podría actuar como un regulador funcional entre ZAP y KHNYN¹³⁵.

Sin embargo, la proteína KHNYN no parece estar implicada en la restricción de alfavirus, ya que en líneas celulares deficientes en KHNYN, la replicación del virus Sindbis (SINV) no se vio afectada¹³⁵. Los autores del estudio postulan que esta divergencia en el requerimiento de KHNYN entre retrovirus y alfavirus podría radicar en diferencias en los mecanismos de acción: mientras que el efecto antiviral frente a los retrovirus mediado por ZAP/KHNYN ocurre predominantemente a nivel de la estabilidad del ARN, la inhibición de los alfavirus parece estar regida principalmente por un bloqueo de la traducción, aunque se requieren más estudios para confirmar esta hipótesis.

Reconocimiento preferencial de dinucleótidos CpG y dependencia del contexto genómico

La caracterización estructural de la interacción de ZAP con el ARN viral representó un avance clave en la comprensión de su mecanismo antiviral, al establecer que este factor reconoce de forma preferencial los dinucleótidos CpG. Estos estudios revelaron que ZAP se une selectivamente a estos motivos a través de residuos ubicados principalmente en el segundo dedo de zinc, el cual forma una bolsa hidrofóbica donde puede acomodar exclusivamente dinucleótidos CpG (**Fig. 2.7**)^{119,120}. Sin embargo, la sola presencia de motivos CpG aislados no es suficiente para inducir una restricción antiviral eficiente, ya que la potencia de esta respuesta antiviral depende del efecto acumulativo de múltiples sitios de unión distribuidos a lo largo del genoma.

En concordancia, se ha demostrado que el incremento artificial de la frecuencia de estos dinucleótidos mediante mutaciones sinónimas atenúa significativamente la replicación viral, incluso en virus que han demostrado ser resistentes a ZAP, como el echovirus 7 (E7) y ZIKV ¹³⁷⁻¹⁴¹.

De manera crítica, se ha observado que la actividad restrictiva de ZAP no depende exclusivamente de la abundancia global de dinucleótidos CpG, sino también de su localización y del contexto genómico en el que estos se encuentran. Así, un incremento global en la frecuencia de CpG en el genoma viral no necesariamente se traduce en una mayor sensibilidad a ZAP. En este sentido, Ficarelli *et al.* (2020) demostraron que, en VIH-1 recodificado, los CpGs introducidos en la región *env* son reconocidos con mayor eficiencia por ZAP que aquellos localizados en *gag* o *pol*, aun cuando el número total de CpGs introducidos sea equivalente ¹⁴².

Adicionalmente, estudios basados en CLIP-seq mostraron que la unión de ZAP al ARN viral es más eficiente cuando los CpGs se encuentran espaciados a una distancia media de aproximadamente 32 nucleótidos, mientras que un espaciamiento menor, cercano a 11 nucleótidos, se asocia con una menor afinidad de unión. Estos resultados sugieren que la interferencia entre múltiples moléculas de ZAP o sus cofactores puede afectar negativamente su actividad antiviral ¹⁴⁰. En el mismo estudio, se observó además que una mayor proporción de nucleótidos uridina (U) o adenina (A) entre CpGs potencia la restricción antiviral mediada por ZAP.

En línea con estas observaciones, se ha demostrado que el contexto nucleotídico inmediato también modula la sensibilidad a ZAP. Por ejemplo, el flanqueo de dinucleótidos CpG con dos nt de U o A a los extremos 5' y 3' de una región no codificante de un replicón del enterovirus E7 reduce significativamente la replicación del ARN viral, en comparación con replicones en los que los CpGs están flanqueados por otros nucleótidos ¹³⁸. Estudios posteriores refinaron este análisis, mostrando que el posicionamiento de una única U a cada lado es suficiente para potenciar la restricción mediada por ZAP ¹²⁹.

En conjunto, estos resultados indican que tanto la distribución como el contexto nucleotídico que rodea a los CpG son determinantes clave de la eficiencia de la restricción antiviral mediada por ZAP, aunque los mecanismos moleculares subyacentes aún no se comprenden completamente. En concordancia con estas observaciones, otros estudios han reportado una recuperación parcial de la replicación de virus enriquecidos en UpA en células carentes de ZAP ¹⁴³. No obstante, análisis estructurales de la interacción de ZAP con el dinucleótido CpG sugieren que, desde un

punto de vista estructural, es poco probable que ZAP se una directamente a dinucleótidos UpA ¹⁴⁴.

Implicancias adaptativas de ZAP en el sesgo de dinucleótidos CpG en genomas de virus ARN

Un rasgo característico de la mayoría de los virus de genoma ssRNA que infectan vertebrados es la marcada subrepresentación del dinucleótido CpG en sus genomas, especialmente en aquellos virus cuyos hospederos se encuentran filogenéticamente cercanos a los mamíferos ¹⁴⁵. Este rasgo composicional suele evaluarse mediante el cálculo del sesgo de CpG, definido como la relación entre la frecuencia observada y la esperada de CpG (O/E CpG) en una secuencia genómica dada, en función de la abundancia de citosinas y guaninas. Así, un valor cercano a 1 indica que los CpGs aparecen aproximadamente con una frecuencia esperada por azar, mientras que valores inferiores a 1 reflejan una subrepresentación de este dinucleótido.

Un análisis comparativo donde se analizó el sesgo de CpG en genomas de virus ssRNA y sus hospederos mostró que aquellos virus que infectan vertebrados presentan valores de O/E CpG significativamente menores a uno, mientras que los virus asociados a invertebrados exhiben valores cercanos a uno (**Fig. 2.8a**). Este patrón se replica en los genomas de los hospederos, en donde notablemente los vertebrados muestran una supresión de CpG considerablemente mayor que los invertebrados (**Fig. 2.8b**), lo que sugiere la existencia de procesos de coevolución entre virus y hospedero en relación con este rasgo composicional (Tesis de Doctorado de Diego Simón).

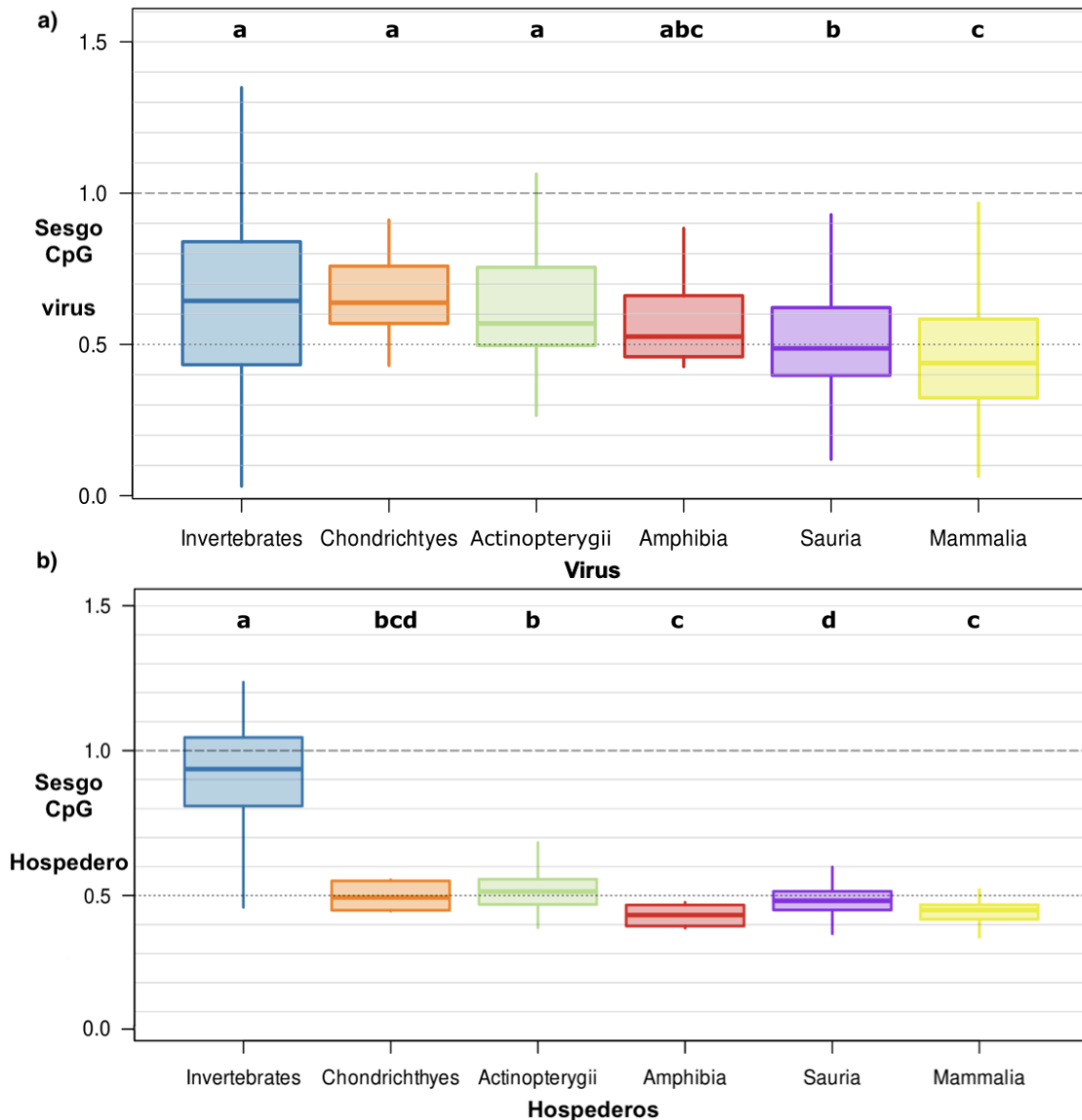


Figura 2.8: Sesgo de CpG en virus ssRNA y sus hospederos metazoos: a) Sesgo de CpG (CpG bias), expresado como la razón de CpG observado/esperado (O/E), en virus de ARN monocatenario (ssRNA) que infectan distintos grupos de vertebrados (box plots en naranja, verde, rojo, violeta y amarillo) e invertebrados (en azul). **b)** Distribución del sesgo de CpG (razón O/E) en hospederos metazoos, con una subdivisión detallada de los vertebrados. La razón O/E se calcula como la frecuencia observada (O) del dinucleótido CpG dividida por la frecuencia esperada (E) en función del número de citosinas y guaninas en los genomas. Las letras sobre los diagramas de caja indican diferencias estadísticamente significativas entre grupos (test de Wilcoxon con corrección de Benjamini–Hochberg), donde los grupos que no comparten la misma letra difieren de manera significativa entre sí ($P \leq 0,05$). Imagen tomada de Simón *et al.*, 2025¹⁴⁵, y parte de su tesis doctoral.

En los vertebrados, la supresión genómica de CpG se ha vinculado a mecanismos de regulación epigenética, particularmente a la metilación de citosinas en la conformación CpG. Las citosinas metiladas son susceptibles a desaminación espontánea convirtiéndose en timina (es decir, TpG), lo que a lo largo del tiempo evolutivo habría

contribuido a la pérdida de CpGs en los genomas de los vertebrados ¹¹. En contraste, la metilación de citosinas en los genomas de invertebrados es limitada, lo que explica la ausencia de supresión de CpG en estos organismos ^{11,146,147}.

En los virus, se ha propuesto que la ausencia de estos mecanismos epigenéticos, junto con la falta de un ortólogo funcional de ZAP en invertebrados, permite la replicación de genomas sin un sesgo marcado de CpG en estos hospedadores ^{148–150}. En contraste, en vertebrados, factores restrictivos de la inmunidad innata, en particular ZAP (diversificada en el último ancestro común de los tetrápodos y los peces pulmonados hace más de 400 millones de años), junto con sesgos mutacionales, constituyen presiones selectivas que moldean la composición de CpG en virus ssRNA de vertebrados ¹⁴⁵. En concordancia, el enriquecimiento experimental de dinucleótidos CpG mediante mutaciones sinónimas en genomas con baja representación de CpG se ha demostrado que conduce a una atenuación de la replicación dependiente de ZAP, reforzando la hipótesis de que esta proteína constituye uno de los principales determinantes de la supresión de CpG en virus ssRNA que infectan vertebrados ^{138, 139,140}.

En el caso de los arbovirus, estas presiones selectivas se complejizan debido a la necesidad de replicarse eficientemente tanto en invertebrados como en vertebrados, dos hospederos separados por más de 570 millones de años de evolución y con sistemas inmunes marcadamente distintos. Mientras que los virus específicos de insectos mantienen genomas relativamente enriquecidos en CpG y son restringidos en células de mamíferos por ZAP ^{151,152}, los arbovirus exhiben distintos grados de supresión de CpG, lo que sugiere estrategias intermedias de adaptación que reflejan compromisos evolutivos entre ambos ambientes celulares ¹⁵³.

En particular, se ha propuesto que los alfavirus conservan genomas relativamente enriquecidos en CpG para así favorecer su replicación en mosquitos, a costa de una mayor restricción mediada por ZAP en hospederos vertebrados ^{154,155}. Además, estudios comparativos entre distintos alfavirus sugieren que la sensibilidad a ZAP no depende exclusivamente del contenido global de CpG, sino también de su distribución localizada en ciertas regiones del genoma viral, especialmente en regiones que codifican proteínas no estructurales como nsP2 ¹⁵⁶. Esta supresión localizada de CpG sugiere un mecanismo de evasión antiviral más específico, que permitiría a los alfavirus modular su susceptibilidad a ZAP y equilibrar las presiones selectivas impuestas por ambos hospederos.

Relevancia de ZAP en arbovirus y justificación del estudio en MAYV

Como se ha descrito hasta ahora, ZAP es un factor clave de la inmunidad innata capaz de restringir la replicación viral mediante el reconocimiento de motivos CpG en ARN viral. Si bien este mecanismo ha sido caracterizado en diversos virus ARN, incluidos algunos arbovirus modelo ¹³⁹, el rol de ZAP en la adaptación genómica de los arbovirus, que deben mantener un equilibrio evolutivo entre hospederos vertebrados e insectos hematófagos, continúa siendo poco explorado.

En este contexto, el virus MAYV, un alfavirus emergente en América Latina, representa un modelo particularmente relevante, ya que su biología molecular y su interacción con la inmunidad innata del hospedero permanecen poco caracterizadas. Hasta el momento, no existen estudios que evalúen la sensibilidad de MAYV a la restricción mediada por ZAP ni el impacto de la introducción de mutaciones sinónimas que incrementen el contenido de CpG sobre su replicación y aptitud viral en distintos hospederos. Asimismo, este tipo de aproximaciones ha sido escasamente explorado en otros arbovirus, lo que limita la comprensión de cómo el contenido y la distribución de motivos CpG influyen en la restricción antiviral y contribuye a definir el rango de hospederos de estos virus ^{154,155}.

Estas lagunas de conocimiento plantean interrogantes fundamentales sobre cómo los arbovirus modulan la composición de sus genomas para compatibilizar presiones evolutivas opuestas entre insectos y vertebrados y cuál es el costo asociado a evadir la restricción antiviral mediada por ZAP. En este sentido, MAYV constituye un modelo idóneo para profundizar en los mecanismos que regulan la restricción viral y la adaptación a distintos hospederos.

Por lo tanto, la presente tesis utiliza variantes sintéticas de MAYV enriquecidas en CpG, las cuales fueron desarrolladas por nuestro grupo, como un modelo experimental para profundizar en los mecanismos mediante los cuales ZAP y la inmunidad innata limitan la replicación viral y condicionan la adaptación de los arbovirus a un ciclo replicativo dual.

Antecedentes y resultados preliminares

Nuestro grupo analizó la composición de CpG en los genomas de arbovirus y la comparó con la de virus ssRNA específicos de insectos o con vector desconocido (**Fig. 2.9**). Se observó que los virus que carecen de vector artrópodo (o se les desconoce) presentan una fuerte supresión de CpG, mientras que los virus específicos de insectos muestran valores elevados de este dinucleótido. Por su parte, los arbovirus exhiben niveles de CpG intermedios, lo que sugiere que su composición refleja un equilibrio evolutivo frente a presiones selectivas divergentes entre vertebrados e invertebrados.

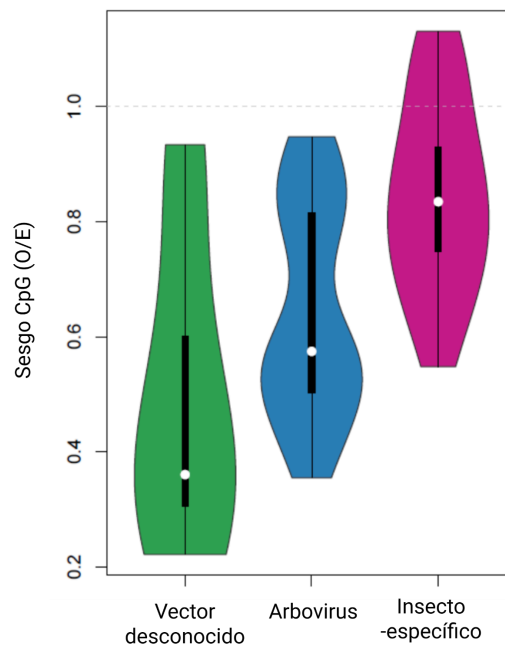


Figura 2.9: Sesgo de dinucleótidos CpG en virus que infectan vertebrados y/o invertebrados según grupo de transmisión. Sesgo de CpG en virus que infectan vertebrados y/o invertebrados, resumido entre grupos de transmisión mediante gráficos de violín. El punto blanco indica la mediana, la barra negra gruesa el rango intercuartílico, y la línea horizontal discontinua indica un sesgo de CpG de 1 (sin depleción ni enriquecimiento). Análisis realizado por el Dr. Diego Simon (LEEV, IP Mon y LVM, UdelaR).

A continuación, con el objetivo de explorar el rol de los CpG en la evolución de los alfavirus y su adaptación a un ciclo replicativo dual, se analizó la distribución de estos dinucleótidos en 71 genomas de MAYV. La mayoría de los sitios CpG resultaron ser extremadamente raros, presentes en una o dos secuencias y ausentes en el clon infeccioso de referencia (MAYV WT), lo que sugiere una fuerte selección negativa que impediría su fijación en el genoma viral (**figura 2.10**).

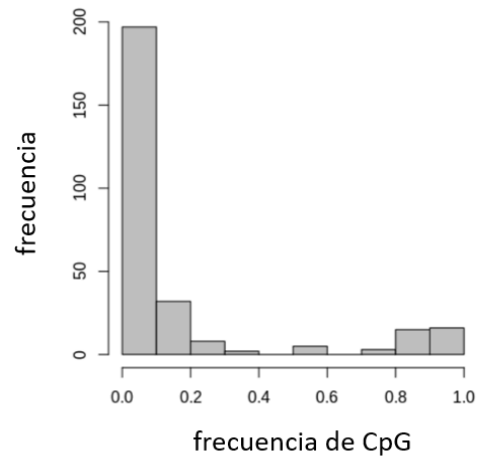


Figura 2.10. Distribución de CpG presentes en una posición genómica determinada para un *dataset* de n= 71 genomas de MAYV. Distribución de CpG presentes en una posición genómica determinada para un *dataset* de n = 71 genomas de MAYV. Cada posición analizada presenta un dinucleótido CpG en al menos un genoma del conjunto de datos, pero dicho dinucleótido está ausente en el clon infeccioso MAYV WT. El eje x representa la proporción de genomas (de un total de 71) en los que se observa un dinucleótido CpG en cada posición genómica, mientras que el eje y indica el número de posiciones genómicas (frecuencia) que presentan cada valor de dicha proporción. Análisis realizado por el Dr. Diego Simon (LEEV, IP Mon y LVM, UdelaR).

A partir de este análisis, en el marco de la tesis de Doctorado de Irene Ferreiro, se generaron en el laboratorio tres variantes de MAYV enriquecidas en CpG, incorporando CpGs observados en al menos 2 de las 71 secuencias disponibles en NCBI mediante mutaciones sinónimas: FG+ (con aumento de CpG en toda la región codificante), NS+ (con aumento de CpG únicamente en la región no estructural) y S+ (con aumento de CpG en la región estructural).

Entre ellas, el virus FG+ mostró la mayor atenuación en células A549, con una reducción de ~10 veces en los títulos virales a las 24 h postinfección con respecto al virus WT (**Fig. 2.11A**), efecto que también se reprodujo *in vivo* en ratones BALB/cJ (**Fig. 2.11C**). En contraste, no se observaron diferencias en células de insecto ni en mosquitos *Aedes aegypti* (**Fig. 2.11B y D**). En conjunto, estos resultados indican que el aumento de CpG atenúa la replicación en hospederos vertebrados, pero no en el vector, sugiriendo una presión selectiva específica del hospedero vertebrado.

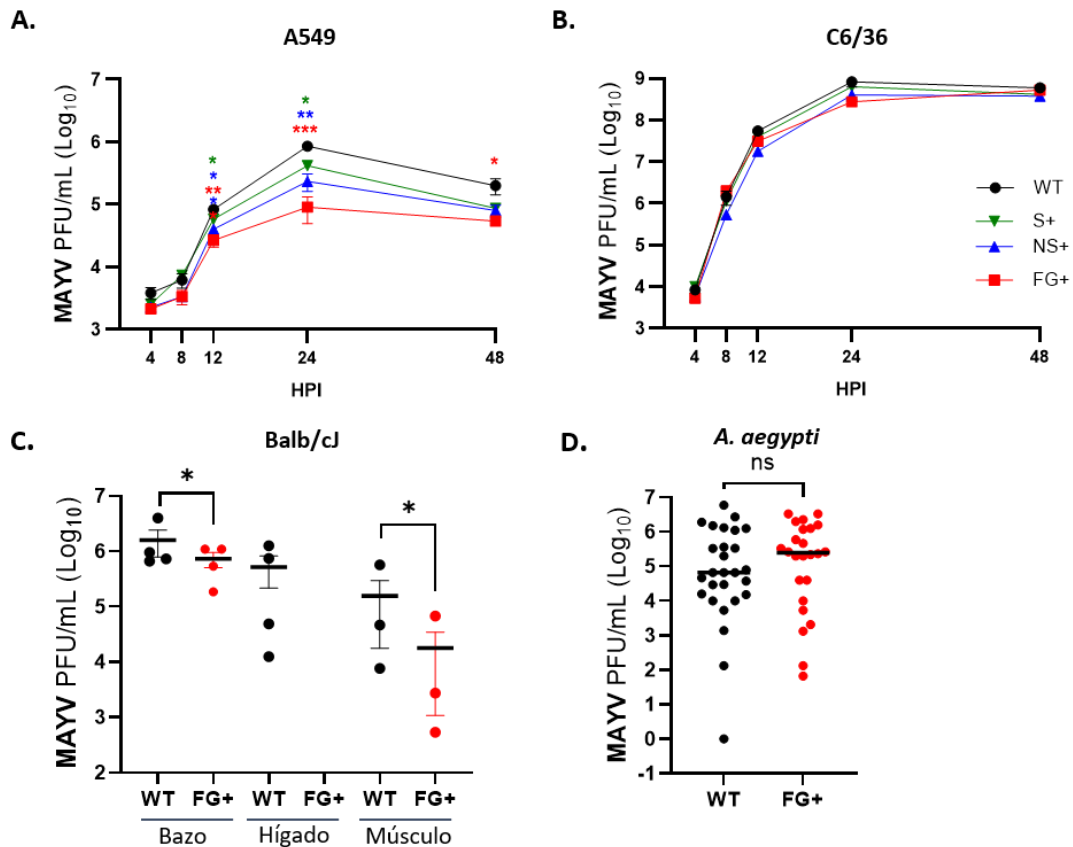


Figura 2.11: Variantes sintéticas de MAYV enriquecidas en CpG son atenuadas en células humanas y en ratones Balbc/cJ, pero no en *A. aegypti* o células C6/36 derivadas de mosquitos. (A-B) Células A549 y C6/36 fueron infectadas con MAYV WT y mutantes NS+, S+ y FG+ a una multiplicidad de infección de 3 (MOI=3). Los títulos virales (PFU/mL) se determinaron por *plaque assay* en células Vero a las 4, 8, 12, 24 y 48 HPI. Se muestra la media de tres réplicas biológicas \pm error estándar. *P < 0.05, **P < 0.01, ***P < 0.001 (ANOVA con post-test de Bonferroni vs. WT en cada tiempo). (Tesis de Doctorado de Irene Ferreiro). **(C)** Títulos virales a los 2 días postinfección en bazo, hígado y músculo. Cada punto indica un animal; la línea muestra la media. *P < 0.05, **P < 0.01, ***P < 0.001 (ANOVA con post-test de Bonferroni). Ensayo realizado por la Dra. Marianoel Pereira y la Dra. Paula Arévalo (IP Montevideo). **(D)** Infección en *Aedes aegypti*. Cada punto representa un mosquito; la línea indica la media. No se detectaron diferencias significativas (ANOVA con post-test de Bonferroni). Ensayo realizado por el Dr. Álvaro Fajardo (NYU).

Hipótesis

El enriquecimiento de dinucleótidos CpG en el genoma de MAYV favorece su reconocimiento por la proteína ZAP en células de vertebrados, promoviendo una restricción antiviral más eficiente y resultando en una replicación viral disminuida. En consecuencia, la atenuación observada en variantes enriquecidas en CpG reflejaría un mecanismo de restricción mediado por la inmunidad innata que contribuye a moldear la adaptación de los arbovirus a sus hospederos.

Objetivo general

Caracterizar el rol de ZAP en la respuesta antiviral celular frente al virus Mayaro (MAYV) y su contribución a la atenuación de una variante atenuada enriquecida en CpG (MAYV FG+).

Objetivos específicos

- 1- Analizar la expresión y activación de ZAP tras la infección de células humanas A549 con MAYV WT y la variante MAYV FG+.
- 2- Desarrollar una línea celular A549 con knockout de ZAP (ZAP KO) mediante CRISPR/Cas9, como modelo experimental para evaluar el impacto directo de ZAP en la atenuación de la variante MAYV FG+.
- 3- Estudiar la replicación de MAYV FG+ vs. MAYV WT en líneas celulares que expresan ZAP (A549) y en la línea ZAP KO (A549 ZAP KO) para determinar el rol de ZAP en la atenuación de MAYV.
- 4- Evaluar el rol diferencial de las isoformas ZAPL y ZAPS en la restricción de la replicación de MAYV mediante el análisis temporal de su inducción tras la infección por MAYV y el silenciamiento selectivo con siRNAs específicos.
- 5- Profundizar en el rol de ZAP en la activación de la respuesta innata y su relación con la señalización por interferón.

Materiales y métodos

Células

A549 (carcinoma pulmonar humano, ATCC CCL-185) y Vero (riñón de mono verde africano, ATCC CCL-81) fueron obtenidas del American Type Culture Collection (ATCC) sin autenticación adicional por nuestro laboratorio, pero se confirmó que estaban libres de micoplasmas. La línea celular A549 ZAP KO fue desarrollada como parte de esta tesis doctoral, cuyo método se detalla a continuación. Las células se mantuvieron en medio DMEM (del inglés Dulbecco's Modified Eagle Medium) suplementado con 10% de suero fetal bovino (FBS, Invitrogen, CA, USA) y 1% de solución de penicilina-estreptomina (Invitrogen).

Clones Infecciosos

El clon infeccioso del virus Mayaro salvaje (MAYV WT) fue proporcionado gentilmente por el Dr. James Weger-Lucarelli (Virginia Tech: Blacksburg, Virginia, USA). El clon infeccioso del virus atenuado MAYV FG+ (**Fig. 2.11**, representado en rojo), fue desarrollado por el Dr. Simón, el Dr. Fajardo y la Lic. Ferreira, integrantes de nuestro grupo de trabajo, en el Laboratorio de Virología Molecular de la Facultad de Ciencias, UdelaR y en el Laboratorio de Evolución Experimental de Virus del Institut Pasteur de Montevideo.

Generación de stocks virales

Para generar la progenie viral inicial (P0), los clones infecciosos WT y FG+ se transfectaron en células Vero. Estas células son deficientes en la producción de interferón tipo I, lo que permite que el virus se replique a títulos muy altos en poco tiempo. Las transfecciones se realizaron en células con un 70 % de confluencia, sembradas en placas de seis pocillos, utilizando el reactivo Lipofectamine™ 3000 (Invitrogen, EE. UU.), de acuerdo con las instrucciones del fabricante. A las 48 horas post-transfección (HPT), cuando el efecto citopático fue evidente, se recuperó el medio celular que contenía la progenie viral P0. Esta progenie se utilizó posteriormente para generar una progenie P1 mediante la infección de células Vero-E6 en frascos T75. Siguiendo el mismo procedimiento, se generó una progenie P2 en un pasaje subsiguiente infectando nuevas células Vero-E6 sembradas en frascos T-175. En cada pasaje se utilizaron 500 µL de la progenie previa y el medio celular se recolectó a las 48 horas postinfección (HPI), cuando se observó un efecto citopático. Todas las progenies se almacenaron en alícuotas de 500 µL a -80 °C.

Secuenciación del genoma completo de MAYV mediante NGS

El ARN viral fue extraído con TRIzol (Invitrogen) o el kit QIAamp Viral RNA Mini (Qiagen) y purificado siguiendo las instrucciones del fabricante. El ARN obtenido (5 µl) fue retrotranscrito utilizando el kit *Maxima H Minus First Strand cDNA Synthesis* (Thermo Fisher Scientific Inc., MA, USA) con oligo dT. A partir del ADNc se amplificó el genoma viral completo de MAYV (11300 pb) en dos amplicones de PCR. Se incluyeron un control de extracción y un control de PCR (sin template), los cuales se mantuvieron a lo largo de la preparación de librerías de secuenciación. Las reacciones de PCR (Vf: 25 µL) se realizaron con la polimerasa de ADN de alta fidelidad Q5® (12,5 µL) (New England Biolabs, MA, USA), 1.25 µL de cebadores directos y reversos (10 µM) y 5 µL de ADNc. Se utilizaron los siguientes cebadores para la amplificación del genoma viral por PCR: (1) JW55-MAYV 1-21F y NotI_Rv, y (2) MAYV 5856-5875F y 11382-11401_Rv. Las secuencias y temperaturas de *annealing* se encuentran en la **tabla S2.1**.

Los productos de PCR se purificaron con AMPure XP *beads* a 0.6X (Beckman Coulter, Brea, CA, USA) siguiendo las instrucciones del fabricante y se cuantificaron con el ensayo fluorométrico *Qubit dsDNA High Sensitivity* (Thermo Fisher Scientific Inc.) en el instrumento Qubit 2.0 (Thermo Fisher Scientific Inc.).

Las bibliotecas de secuenciación se prepararon según un protocolo optimizado a partir del protocolo de secuenciación descrito por Freed N. y Silander O. con el kit *Rapid barcoding* SQL-RBK 110.96 (Oxford Nanopore Technologies, ONT) ¹⁵⁷. Con este fin se utilizaron 100 ng de productos de PCR combinados a una relación molar 1:1 para la preparación de la biblioteca con el kit mencionado anteriormente. Se cargaron 20 fmol en una *flow cell* FLO-FLG001 R9.4.1 (ONT) y se secuenciaron en la plataforma de secuenciación GridION X5 (ONT), la cual se ejecutó durante 12-24 horas. La asignación de bases y la demultiplexación se realizaron con Guppy 6.3.9 (Oxford Nanopore Technologies, 2003a), utilizando el modo de súper precisión. Los genomas consenso se generaron mediante el flujo de trabajo Epi2Me (ONT, 2003b), utilizando los parámetros predeterminados.

Titulación viral mediante ensayo en placa.

Se sembraron células Vero-E6 en placas de seis pocillos y se prepararon diluciones seriadas 1:10 de las muestras virales en medio DMEM sin suplementar. Las células se lavaron dos veces con PBS y se infectaron con 100-150 µl de cada dilución durante 1 hora a 37 °C. Posteriormente, se agregó una capa semisólida que comprendía medio DMEM y 1% en peso/volumen de agarosa ultrapura. A las 48 horas postinfección

(HPI), las células se fijaron con una solución de formalina al 4% y se tiñeron con violeta de cristal al 0.2%. Finalmente, se contaron las placas de lisis para determinar la concentración de partículas virales infectivas, expresada como unidades formadoras de placa por mililitro (PFU/mL).

Infecciones virales y tratamiento con IFN α

Para evaluar el efecto de ZAP en la replicación viral, células A549 o A549 ZAP KO se infectaron con el virus MAYV WT y MAYV FG+ a una multiplicidad de infección (MOI) de 3. A las 2, 4, 8, 16, 24 y 48 HPI, se recolectaron sobrenadantes (200 μ L) y pellets celulares. Los sobrenadantes se utilizaron para analizar la cinética de replicación viral mediante ensayo de placa. Los pellets celulares se emplearon para: a) extracción de proteínas y el análisis por western blot de ZAP, nsP1 y E1; y b) extracción de ARN y cuantificación por PCR cuantitativa en tiempo real (RT-qPCR) de los transcritos intracelulares de ZAP (ZAPL y ZAPS), IFN β , CXCL10, CCL5 y OAS1, así como del ARN genómico (utilizando cebadores dirigidos a la región nsP1) y subgenómico (utilizando cebadores dirigidos al gen E2) del genoma viral. Los cebadores y temperaturas de *annealing* se detallan en la **tabla S2.1**. Para el ensayo con IFN α -2b, las células fueron pretratadas con 500 U/mL de IFN α -2b en DMEM completo durante 16 horas previas a la infección viral.

Extracción de ARN y PCR Cuantitativa en Tiempo Real (RT-qPCR) para: ZAP y MAYV

Las células A549 se sembraron a una densidad de 1.5×10^5 células/pocillo en placas de 12 pocillos en DMEM suplementado con 10% de FBS y 1% de solución de Pen-Strep. Al día siguiente, las células se infectaron con una MOI 3 durante 1 hora a 37°C y 5% de CO₂ con agitación cada 15 minutos. Las células se incubaron en DMEM suplementado con 2% de FBS y 1% de Pen-Strep de 2 a 24 horas. El ARN total se extrajo utilizando el kit RNAeasy Mini (QIAGEN), siguiendo las instrucciones del fabricante. Para la síntesis del ADN copia (ADNc), se emplearon 0.5 μ g de ARN total en un volumen final de 10 μ L con 2 μ L de 5X LunaScript™ RT SuperMix (NEB) y agua libre de nucleasas hasta un volumen total. La mezcla se incubó secuencialmente a 25°C durante 2 minutos, a 55°C durante 10 minutos y a 95°C durante 5 minutos. El ADNc se diluyó 1:5 para todas las reacciones, excepto para la amplificación de la región del genoma viral codificante para la proteína nsP1, donde se utilizó una dilución 1:50. La RT-qPCR se realizó en una reacción de 20 μ L que contenía 1 μ L de ADNc diluido, 2 μ L de 5X Luna® Universal qPCR Master Mix (NEB), 0.5 μ L de cada cebador (10 μ M) y completando el volumen con agua libre de nucleasas. Las

secuencias de los cebadores y sus temperaturas de *annealing* se detallan en la **tabla S2.1**.

Preparación de lisados celulares para SDS-PAGE y Western blot de ZAP, nsP1, E1 y Cas9.

Las células adherentes A549 se lisaron directamente en 200 µL de buffer de muestra 2X (125 mM TRIS-HCl, 4% SDS, 16% glicerol, 10% β-mercaptoetanol, 0.002% azul de bromofenol) y se hirvieron durante 5 minutos a 100 grados centígrados. La concentración de proteínas totales se determinó mediante el kit Pierce 660 nm con suplemento de Reactivo de Compatibilidad de Detergente Iónico (IDCR) (Invitrogen, USA) siguiendo las indicaciones del fabricante.

Para el WB, se separaron cantidades iguales de proteínas de cada muestra, hasta 20 µg por carril, en geles precolados Mini Protean TGX Stain-Free, 4-20 % (Bio-Rad, USA) y se transfirieron a membranas de nitrocelulosa. Los anticuerpos primarios utilizados fueron: anti-ZAP (ab154680, Abcam, Cambridge, MA), 1:15000; proteínas virales anti-nsP1 y anti-E1 (cedidos por el Dr. José González, Instituto GORGAS, Panama), 1:5000; anti-spCas9 (E7M1H, Cell Signalling), 1:1000; anti-alfa tubulina conjugada con HRP (ab40742, Abcam), 1:5000; anti-GAPDH (3683S, Cell Signaling, Danvers, MA), 1:1500. Como anticuerpo secundario se utilizó un anticuerpo de cabra anti-conejo IgG H&L conjugado con peroxidasa. La detección de proteínas se realizó con el sustrato quimioluminiscente SuperSignal West Pico PLUS (Invitrogen). Las intensidades de las bandas se midieron por densitometría utilizando el software ImageJ (Rasband, W.S., ImageJ, Institutos Nacionales de Salud de EE. UU., Bethesda, Maryland, EE. UU., <http://imagej.nih.gov/ij/>, 1997-2016) y se normalizaron a los niveles de expresión de α-tubulina o GAPDH.

Generación de células A549 ZAP KO mediante CRISPR/Cas9

Se emplearon 3 ARN guías (sgRNAs) diseñados para inhibir la expresión del gen ZC3HAV1: sgRNA1 (5'-CGACGTGGTCCAGAACATCC-3')¹⁴³, sgRNA2 (5'-TCTGTGACCACTTCACCCGA-3') diseñado en este trabajo con el software Synthego Knockout Guide Design, y el sgRNA3 (5'-ACTTCCATCTGCCTTACCGG-3')¹³⁵. Las células A549 fueron resuspendidas en buffer de electroporación MaxCyte (Maxcyte, Walkersville, MD) a una concentración de 5×10^6 células/mL. Se mezclaron 45 µL de suspensión celular con partículas ribonucleoproteicas (RNP) que consistían en 272 pmol de ARN guía sintético (sgRNA, Kit CRISPRvolution sgRNA EZ, modificado; Synthego, Redwood City, CA), y 68 pmol de proteína Cas9 2NLS (Synthego, Redwood, CA). Las transfecciones se realizaron con el dispositivo MaxCyte STx™ con el programa para células A549 y las celdas de procesamiento R-50x3™.

Las células transfectadas se cultivaron durante 48 h. en placas de 24 pocillos en DMEM completo a 37°C y 5% de CO₂ y luego se sembraron en placas de 96 pocillos a una densidad de 0.8 células/pocillo para obtener clones individuales. La eficiencia de edición génica se evaluó mediante secuenciación por Sanger (Macrogen Inc, Seúl, Corea del Sur) utilizando los cebadores For_sgRNA2 y Rev_sgRNA2 (**tabla S2.1**) y las secuencias fueron analizadas utilizando la herramienta en línea de Inferencia de Ediciones CRISPR (ICE, Synthego, USA). El *knock out* (KO) para ZAP fue confirmado mediante WB utilizando un anticuerpo contra ZAP humana, el cual reconoce todas sus isoformas.

Silenciamiento génico de ZAPL y ZAPS por DsiRNA

Los Custom Dicer-substrate siRNA (DsiRNA) para el silenciamiento de las isoformas ZAPL o ZAPS y un control negativo (DsiNC) fueron adquiridos en IDT (ver secuencias en **tabla S2.1**)¹¹³. Células A549 fueron transfectadas con 10 o 20 nM de cada DsiRNA utilizando Lipofectamina 3000 (Invitrogen) por 48 h o 72 h, según el protocolo del fabricante, y posteriormente infectadas con MAYV FG+ (MOI=3). El silenciamiento de ZAP se verificó mediante WB.

Estadística

Los gráficos y los análisis estadísticos se realizaron con el software GraphPad (Prism). La significancia estadística de las diferencias entre los promedios de grupos fue evaluada mediante los análisis indicados en cada pie de figura, considerando significativo un valor de $p \leq 0.05$.

Resultados

Análisis de la expresión de ZAP en células A549 infectadas con MAYV.

Dado que no existían antecedentes sobre la inducción de ZAP durante la infección con MAYV, se analizó la expresión de sus isoformas (ZAPL y ZAPS) en células A549 infectadas con MAYV WT o MAYV-FG+ (MOI=3). A las 24 HPI se evaluaron los niveles de ZAPL y ZAPS por WB y RT-qPCR (**Fig. 2.12A.i y Fig. 2.12A.ii**), utilizando células no infectadas como condición basal (mock).

Como se puede observar en la **figura 2.12A.i** (gráfico izquierdo), en ausencia de infección, ZAPL fue la isoforma predominante, mientras que ZAPS fue prácticamente indetectable, en concordancia con lo descrito en la literatura.

La cuantificación proteica por WB reveló que la infección con MAYV WT indujo la expresión de ZAPL, mientras que MAYV FG+ no produjo cambios detectables en su expresión. En contraste, la isoforma ZAPS aumentó significativamente tras la infección con ambos virus, con un incremento de aproximadamente tres veces para MAYV WT y de dos veces para la variante FG+ enriquecida en CpG (**Fig. 2.12A.i**, gráfico derecho).

De manera complementaria, el análisis por RT-qPCR (**Fig. 2.12A.ii**) mostró que MAYV WT incrementó significativamente los niveles de los transcritos de ambas isoformas. En el caso de MAYV FG+, si bien se observó una tendencia al incremento en los niveles de ZAPS, esta no alcanzó a ser estadísticamente significativa, aunque sí se detectó un aumento claro a nivel proteico. Esto sugiere que incluso un aumento transcripcional moderado podría ser suficiente para elevar la expresión proteica. Cabe señalar que la falta de significancia observada para FG+ podría estar influenciada por el reducido número de réplicas biológicas (n=2).

En conjunto, estos resultados indican que la infección por MAYV induce la expresión de ZAP, con perfiles diferenciales según la variante. Mientras que el virus WT incrementó significativamente los niveles de ambas isoformas, la variante FG+ mostró un aumento restringido a ZAPS y de menor magnitud. En este contexto, una mayor densidad de CpG no se correlacionó con una inducción proporcionalmente mayor de ZAP.

Como control de sistema experimental, y dado que ZAP es un ISG, se evaluó su expresión tras el tratamiento con IFN α o la transfección con poly I:C, un análogo intracelular de dsRNA que activa la señalización a través de los receptores citosólicos tipo RIG-I (RLRs).

En concordancia con lo observado tras la infección con reportes previos, la isoforma ZAPS fue predominantemente inducida por ambos estímulos: el tratamiento con IFN α produjo un aumento de dos veces respecto a células no tratadas, mientras que la transfección con poly I:C indujo un incremento de 15 veces respecto a células no tratadas. En contraste, ZAPL no mostró cambios detectables en su expresión bajo ninguno de estos tratamientos (**Fig. 2.12.B, +IFN α** y **Fig. 2.12.C, +poly I:C**).

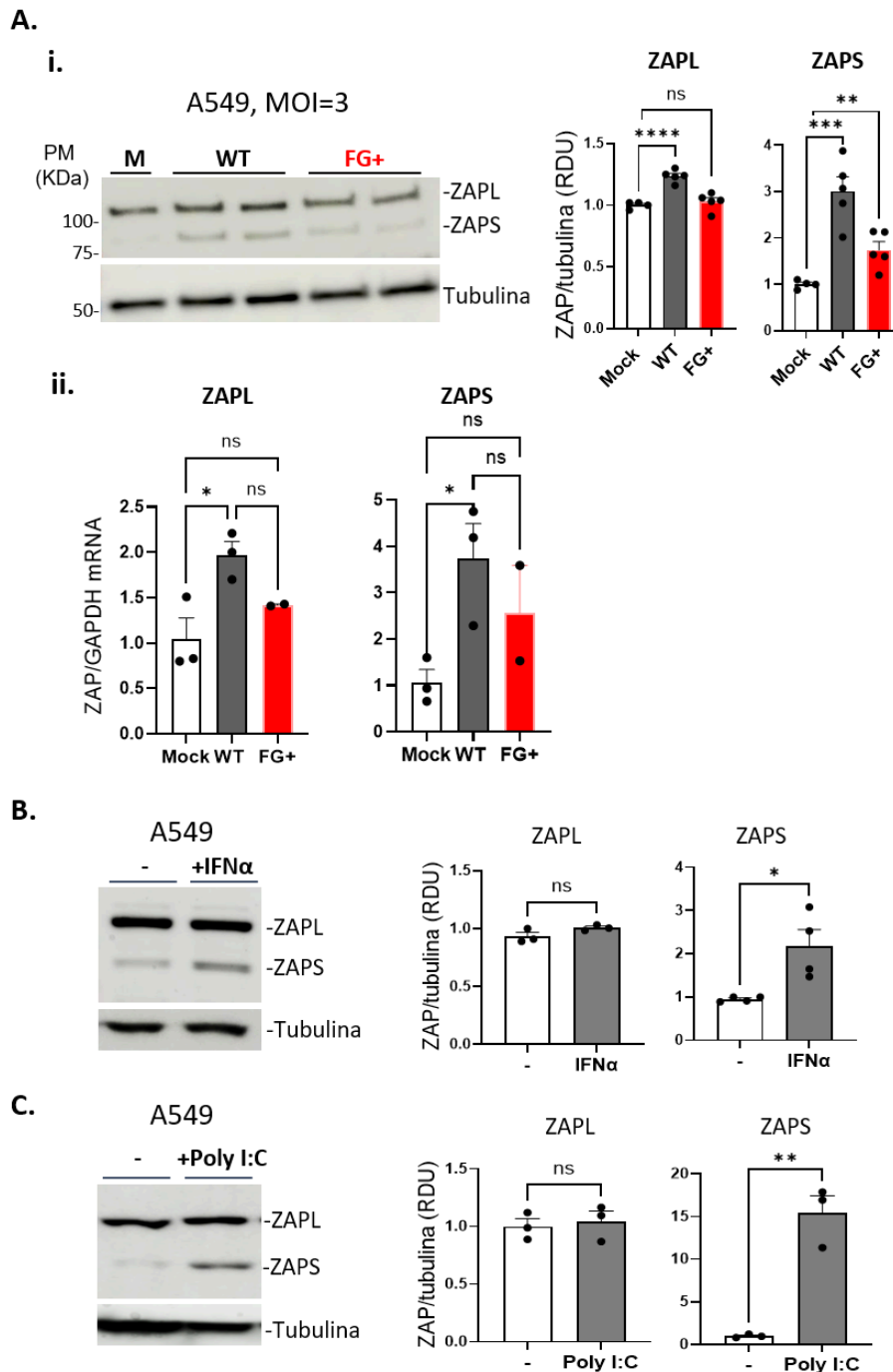


Figura 2.12: MAYV WT induce una mayor expresión de ZAP que la variante FG+ en células A549. **A)** Células A549 fueron infectadas con MAYV WT o FG+ (MOI=3) durante 24 h, o se mantuvieron sin infectar (M, mock). **i:** WB representativo para ZAP y tubulina (panel izquierdo), y análisis de la densitometría de banda (RDU, del inglés *relative densitometry units*) para ZAPL (110 kDa) y ZAPS (90 kDa) con el software ImageJ (panel derecho); **ii:** cuantificación relativa de los niveles del transcrito de ZAPL o ZAPS por RT-qPCR. **B)** Células A549 fueron tratadas durante 8 h con 500 U/mL de IFN α (+IFN α) o se mantuvieron sin tratar (-). **C)** Células A549 transfectadas con poly I:C por 24 h o sin transfectar (-). En B y C se muestra un WB representativo contra ZAP y tubulina (panel izquierdo) y el análisis relativo de ZAPL y ZAPS (gráficos de barras en el panel derecho). Se graficó la media \pm SEM, normalizada a los niveles Mock, de $n = 5$ (A.i), $n = 4$ (B) y $n = 3$ (A.ii y C) réplicas experimentales. ns (no significativo); * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$; **** $p \leq 0,0001$; prueba *t-Student*.

Generación de células A549 ZAP KO mediante CRISPR/Cas9.

Con el fin de investigar el papel de ZAP en la atenuación de MAYV FG+, se generaron células A549 ZAP knockout (KO mediante CRISPR/Cas9). Esta herramienta emplea un ARN guía (sgRNA) para dirigir a Cas9 a secuencias específicas, donde actúa como “tijera molecular” generando cortes que se reparan intracelularmente mediante inserciones/deleciones por unión de extremos no homólogos (NHEJ) o reparación dirigida por homología (HDR).

Dado que el gen ZC3HAV1 da origen a cuatro isoformas por empalme alternativo, se seleccionó el exón 3, común a todas, para abolir la expresión de todas las isoformas de ZAP. Para ello se diseñó un sgRNA específico (sgRNA2) (**Fig. 2.13.A**), y se evaluaron además dos sgRNAs previamente reportados: sgRNA1 (contra el exón 3) y sgRNA3 (contra el exón 6) ^{135,143}.

A continuación, se generaron complejos ribonucleoproteicos (RNP sgRNAs-Cas9) los cuales fueron transfectados en células A549 y posteriormente la eficiencia de edición fue evaluada por WB a las 96 h post-infección. Con el objetivo de maximizar la eficiencia de eventos de edición génica, se evaluaron dos relaciones molares sgRNA:Cas9 (8:1 y 4:1) (**Fig. 2.13B**). Entre los tres ARNs guías analizados, sgRNA2 en una relación molar sgRNA:Cas9 de 4:1 mostró la mayor eficiencia, reduciendo la expresión de ZAPL en aproximadamente un 80% y llevando a niveles prácticamente indetectables de ZAPS.

Cabe destacar que, a pesar de que el sitio blanco del sgRNA3 se localiza en un exón más distal al extremo 5' (**Fig. 2.13A**), este mostró una buena eficiencia, reduciendo la expresión de ZAPL en un ~70% y logrando la desaparición de la banda de ZAPS. Por el contrario, el sgRNA1 resultó ser el menos eficaz, con una reducción de apenas el 30% para ZAPL.

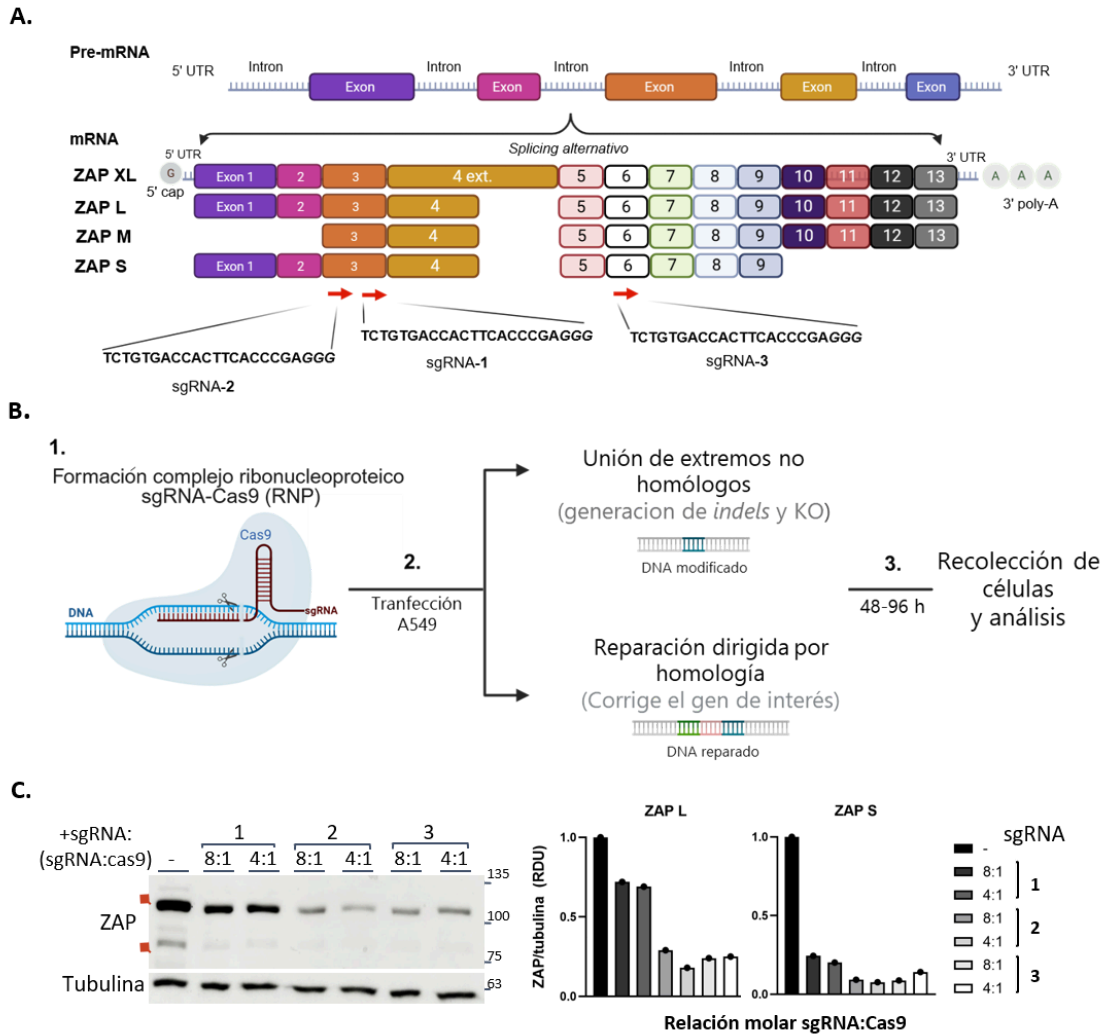


Figura 2.13: Diseño y evaluación de ARN guías contra el gen ZC3HAV1 codificante para ZAP. **A)** Representación de los cuatro transcritos de ZAP (ZAPXL, ZAPL, ZAPM y ZAPS). Los sitios blancos para cada ARN guía, sgRNA1, sgRNA2 y sgRNA3, están indicados por flechas rojas en la parte inferior del esquema. **B)** Representación del mecanismo de edición génica mediante el complejo ribonucleoproteico (RNP) sgRNA-Cas9. Se detalla el reconocimiento del sitio blanco y la inducción del corte de doble cadena, seguido de las dos vías intracelulares de reparación: la generación de indels y knockout (KO) por unión de extremos no homólogos, o la corrección del gen de interés mediante reparación dirigida por homología. **C)** Panel izquierdo, WB a partir de homogeneizados de células A549 electroporadas con complejos ribonucleoprotéicos (RNP) de sgRNA (+sgRNA 1, 2 o 3) y proteína Cas9, a una relación molar de 8:1 o 4:1, o células A549 electroporadas sin RNP (-); se utilizaron anticuerpos contra ZAP y tubulina. Panel derecho, análisis de densitometría de banda (RDU) a partir del WB mostrado en el panel izquierdo, de las isoformas ZAPL (gráfico izquierdo) o ZAPS (gráfico derecho). Se muestran valores de RDU relativos a tubulina, normalizados con respecto a las células sin RNP (-).

Se obtuvieron líneas clonales ZAP KO a partir del pool A549 ZAP KD mediante dilución límite en placas de 96 pocillos. De los 16 clones analizados, 12 no presentaron expresión detectable de ZAP por western blot (**Fig. 2.14A**).

Diez de estos clones sin expresión detectable de ZAP fueron caracterizados mediante PCR de una región flanqueante al sitio blanco del sgRNA2 (566 pb), seguida de secuenciación Sanger (**Fig. 2.14B**). El análisis ICE mostró que los clones E11_1 y E12_2 eran homocigotos, con una delección de un nucleótido, mientras que G9_1 presentó múltiples variantes (una inserción del 70 % y dos delecciones del 19 % y 9 %), sugiriendo heterocigosidad o mezcla de poblaciones (**Fig. 2.14C**).

Finalmente, la ausencia de ZAP en los clones E11_1, E12_2 y G9_1 también se confirmó en condiciones de infección con MAYV, confirmando así el fenotipo ZAP KO en condiciones de infección (**Fig. 2.14D**).

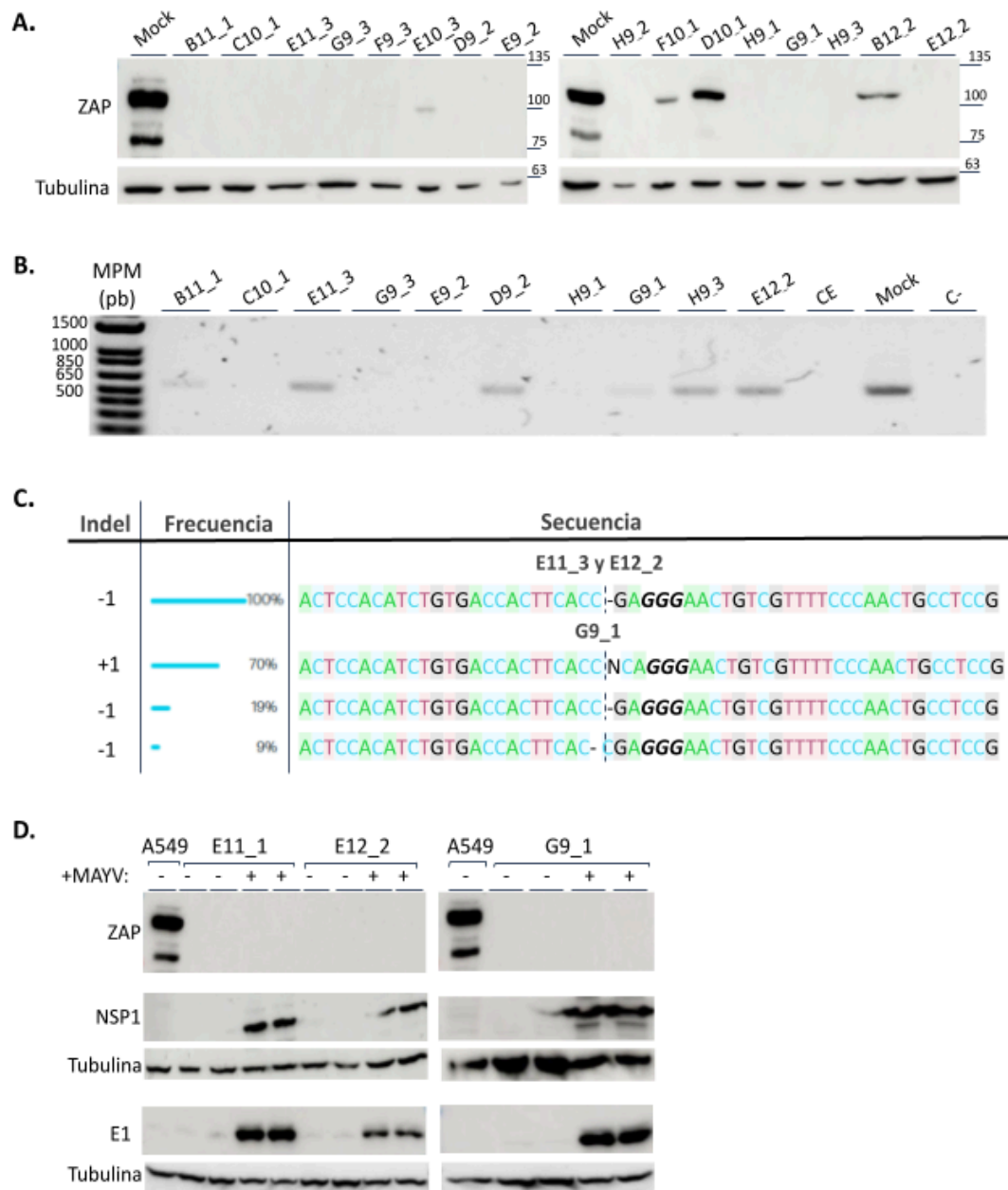


Figura 2.14: Validación de ZAP KO en clones A549 (E11_3, E12_2 y G9_1). **A)** WB para ZAP y tubulina en clones obtenidos por dilución límite a partir del pool A549 ZAP KD. **B)** Gel de agarosa de productos de PCR flanqueantes al sitio blanco del sgRNA2. **C)** Análisis ICE de productos secuenciados por Sanger. Se indican los *indels* (inserciones N, deleciones -), su frecuencia (%) y el sitio de corte por Cas9 (líneas punteadas; PAM en GGG). **D)** WB para ZAP, nsP1 y E1 en células A549 ZAP KO infectadas con MAYV WT (+) o no infectadas (-), 24 HPI (n=2).

La atenuación de MAYV FG+ en células A549 es dependiente de ZAP.

Con el objetivo de evaluar el rol de ZAP en la atenuación de la variante de MAYV enriquecida en CpG (MAYV FG+), se utilizaron células A549 y una de las líneas A549

ZAP KO generadas en esta tesis. Ambas fueron infectadas con MAYV WT y MAYV FG+ a una MOI = 3, y se analizó la replicación viral a las 12 y 24 HPI mediante titulación viral por ensayo de placas (**Fig. 2.15A**). A las 24 HPI, además, se cuantificaron las copias relativas intracelulares de ARN genómico y subgenómico por RT-qPCR (**Fig. 2.15B**) y se analizó la expresión de las proteínas virales nsP1 (traducida a partir de ARN genómico) y E1 (traducida a partir de ARN subgenómico) por WB (**Fig. 2.15C**). De manera complementaria, se analizó la cinética de replicación de MAYV WT y MAYV FG+ en el *pool* celular ZAP KD (con 80% de reducción de la expresión de ZAPL), previamente caracterizado (**Fig. 2.14B**), cuyos resultados se muestran en la información suplementaria (**Fig. S2.1**).

En células A549, la variante FG+ mostró una reducción significativa en el título viral, acompañada de una disminución en los niveles intracelulares de ARN genómico y en la expresión de la proteína nsP1, en comparación con el virus WT a las 12 y 24 HPI. Sin embargo, ni los niveles de ARN subgenómico (**Fig. 2.15B**, gráfico derecho) ni la expresión de la proteína estructural E1 (**Fig. 2.15C**, panel inferior) se vieron reducidos en la variante FG+ con respecto al virus WT en presencia de ZAP.

En conjunto, estos resultados sugieren que la atenuación de MAYV FG+ se asocia específicamente a una disminución del ARN genómico intracelular y a una menor expresión de la proteína nsP1, y no a una alteración en los niveles de ARN subgenómicos o en la expresión de proteínas estructurales.

En células A549 ZAP KO, la variante FG+ alcanzó niveles de replicación comparables a los del virus WT, sin diferencias significativas en títulos virales, genomas o niveles de proteína nsP1 (**Fig. 2.15.A-C**). De manera complementaria, en células A549 ZAP KD también se observó que la variante FG+ no mostró diferencias significativas en títulos virales respecto al virus WT (**Fig. S2.1**). En conjunto, estos resultados identifican a ZAP como el mediador principal de la atenuación de MAYV FG+ y descartan la participación de otros mecanismos independientes de ZAP que afecten la replicación de MAYV tras la introducción sinónima de CpG en células de mamífero.

Asimismo, se observó que tanto MAYV WT como FG+ replican mejor en ausencia de ZAP, evidenciado por un aumento significativo en los títulos virales para ambos virus en células ZAP KO en comparación con células A549, lo que se vio reflejado en mayores copias de ARN genómico o subgenómico y una mayor expresión de proteínas virales a nivel intracelular también para el virus WT (**Fig. 2.15**, A549 vs. A549 ZAP KO en gráficos de barra), sugiriendo que ZAP posee acción antiviral sobre ambos virus.

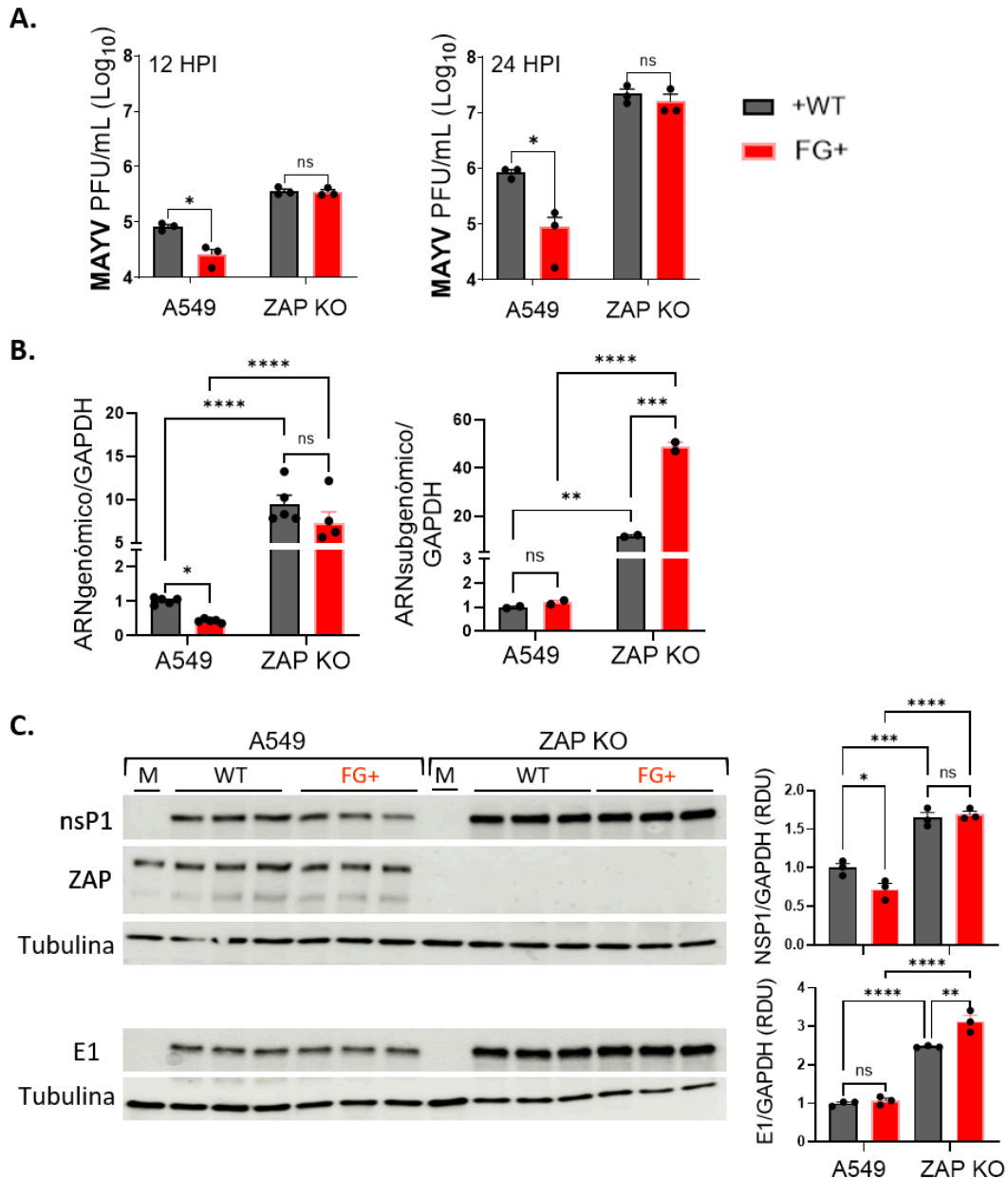


Figura 2.15: La atenuación de MAYV FG+ en células A549 es dependiente de ZAP. A) Titulación viral por ensayo de placa de MAYV WT y FG+ a las 12 o 24 HPI **B)** Cuantificación relativa de ARN viral genómico (gráfico de barras izquierdo) y ARN viral subgenómico (gráfico de barras derecho) por RT-qPCR **C)** Panel izquierdo, WB para las proteínas virales nsP1 y E1, ZAP y tubulina con n=3 réplicas biológicas; panel derecho, análisis de densitometría (RDU) de para los niveles de proteínas virales nsP1 y E1 normalizados a los de tubulina, relativos a células A549 infectadas con MAYV WT. En A-C se infectaron células A549 y A549 ZAP KO, con MAYV WT (barras grises) o MAYV FG+ (barras rojas) a MOI = 3. M, mock sin infectar. Para gráficos de barras: ns, no significativo, *p ≤ 0,05; **p ≤ 0,01; ***p ≤ 0,001; ****p ≤ 0,0001; *two-way* ANOVA (A549 vs. A549 ZAP KO) y prueba *t-Student* múltiple (FG+ vs. WT). Se graficó la media ± SEM (n = 3 experimentos independientes).

ZAPL es la principal isoforma responsable de la restricción de MAYV FG+ en células A549

Con el objetivo de identificar la isoforma de ZAP implicada en la restricción de MAYV FG+, se evaluó el efecto de un silenciamiento específico de ZAPL o ZAPS sobre la replicación viral en células A549 infectadas con este virus (**Fig. 2.16**). El análisis por WB confirmó la eficacia y especificidad de los DsiRNAs dirigidos contra el exón 12 (ZAPL) y la región 3'UTR (ZAPS).

La titulación viral por ensayo de placa mostró que el silenciamiento de la isoforma ZAPL produjo un aumento significativo de los títulos virales y de la expresión de nsP1 (**Fig. 2.16**). En contraste, la inhibición de ZAPS no produjo cambios significativos en ninguno de los parámetros evaluados.

En conjunto, los ensayos de silenciamiento selectivo mostraron que ZAPL es la principal responsable de la restricción de MAYV FG+ en células A549.

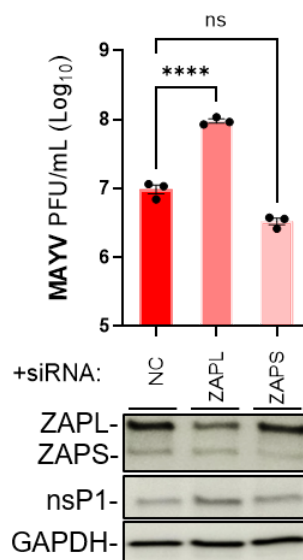


Figura 2.16: ZAPL es la principal isoforma responsable de la restricción de MAYV FG+ en células A549. Titulación viral (PFU/mL, log₁₀) en células A549 infectadas con MAYV FG+ (MOI=3, 24 HPI) tras la transfección con un DsiRNA control no dirigido (NC), DsiRNA específico contra ZAPL o contra ZAPS. WB representativo de lisados celulares mostrando los niveles de ZAPL y ZAPS, la proteína viral nsP1 y GAPDH como control de carga. ns, no significativo; ****p ≤ 0,0001; *one-way* ANOVA con prueba post hoc de comparaciones múltiples. Se muestra la media ± SEM (n = 3 réplicas biológicas independientes).

El tratamiento con IFN α demuestra que la atenuación de MAYV FG+ es dependiente del grado de activación de ZAP.

En análisis previos (**Fig. 2.12A**) observamos que MAYV FG+ es menos eficiente en la inducción de ZAP que MAYV WT a las 24 HPI, con aumento significativo limitado a la isoforma ZAPS. Sin embargo, los ensayos de silenciamiento selectivo (**Fig. 2.16**) indicaron que ZAPL es la principal responsable de la restricción viral, mientras que ZAPS cumple un rol menor. Estas diferencias son consistentes con la mayor actividad antiviral de ZAPL frente a alfavirus, atribuida a su prenilación, que permite su asociación con compartimentos membranosos ^{105,107}, en contraste con el rol predominantemente inmunomodulador de ZAPS ¹¹⁵.

Dado que MAYV FG+ induce ZAP de forma menos robusta, se planteó que un aumento previo de sus niveles podría potenciar la restricción viral. Para ello se realizó un pretratamiento de células A549 con IFN α O/N para inducir su expresión antes de la infección y posteriormente fueron infectadas con MAYV WT o FG+ (MOI = 3). A las 24 HPI, se evaluó la expresión de ZAP mediante WB, incluyendo controles sin IFN α y células tratadas sin infectar (M, mock) (**Fig. 2.17**).

El pretratamiento con IFN α , en combinación con la infección viral, mostró un efecto aditivo sobre la inducción de ZAP. En células infectadas sin IFN α , ZAPS fue indetectable (ND), mientras que en células pretratadas e infectadas su expresión aumentó consistentemente, con un incremento de ~12 veces respecto a células tratadas sin infección. En contraste, ZAPL mostró un aumento más moderado en respuesta al IFN α , tanto en células no infectadas (~0.5 veces) como infectadas (~1.2 veces). No se realizó análisis estadístico para la condición sin infectar debido a la disponibilidad de una única réplica.

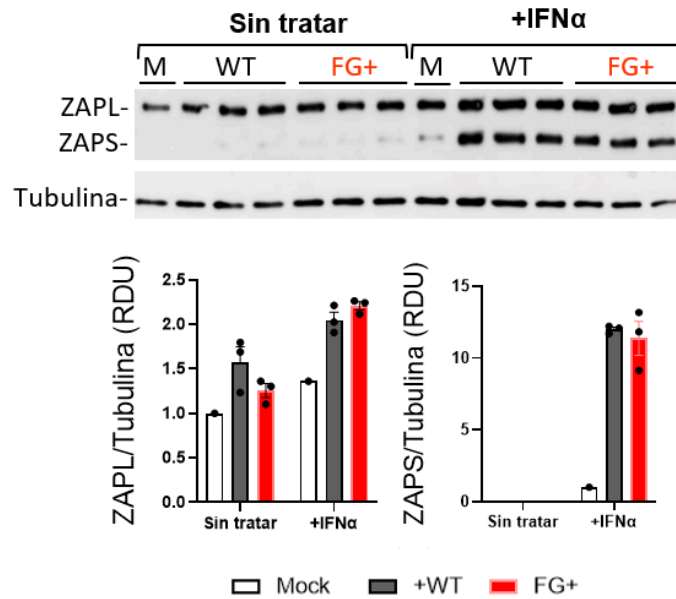


Figura 2.17: La activación de ZAP tras la infección con MAYV WT y FG+ se ve incrementada en células pretratadas con IFN α . Panel superior, WB con anticuerpos contra ZAP y tubulina en tres réplicas independientes; panel inferior, análisis de densitometría de banda (RDU) de los niveles de ZAPL (gráfico izquierdo) o ZAPS (gráfico derecho) normalizados a la expresión de tubulina y relativos a células A549 no infectadas (M, mock) y sin tratamiento. Se infectaron células A549 con MAYV WT y FG+ 24 HPI, MOI=3 sin tratar o con un pretratamiento O/N IFN α 500 u/mL (+IFN α). ND, no detectado. Se graficó la media \pm SEM (n = 3 réplicas biológicas independientes) excepto para el mock, que se muestra sin barras de error (n=1) como referencia. No se realizó análisis estadístico dado el n=1 para la condición sin infectar.

Posteriormente, se evaluó la replicación viral mediante titulación por ensayo de placas, la cuantificación de ARN genómicos por RT-qPCR y la expresión de nsP1 por WB, a distintos tiempos postinfección (**Fig. 2.18**). El tratamiento con IFN α inhibió significativamente la replicación de ambos virus, observándose una reducción en la producción de partículas infectivas tanto a las 12 como a las 24 HPI (**Fig. 2.18A**). El efecto fue más pronunciado en MAYV FG+ con una reducción de \sim 3 log a las 24 HPI, en comparación con \sim 2 log para MAYV WT. De manera consistente, se observó una disminución notable en los niveles de ARN genómico y la proteína nsP1 (**Fig. 2.18B**), siendo prácticamente indetectables para la variante FG+ en células A549 tratadas con IFN α .

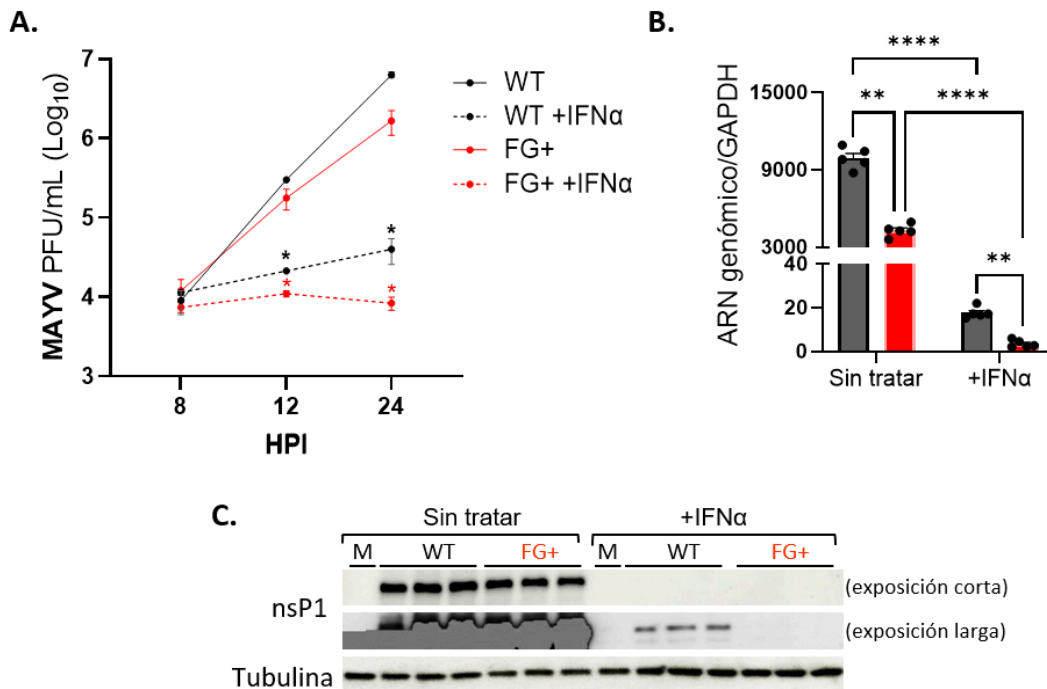


Figura 2.18: IFN α inhibe la replicación de MAYV, observándose una mayor inhibición de la variante FG+. **A)** Titulación viral por ensayo de placa en células Vero-E6 para MAYV WT y FG+ a las 8, 12 y 24 HPI de células A549 infectadas a MOI = 3, con un pretratamiento O/N con IFN α 500 U/mL (línea punteada, +IFN α) o sin tratar (línea continua). **B)** RT-qPCR para la cuantificación de ARN genómico viral relativo a GAPDH. **C)** WB para la proteína viral nsP1 y tubulina. M, mock (células sin infectar). * $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$; **** $p \leq 0,0001$; two-way ANOVA (sin tratar vs. + IFN α) y prueba *t-Student* múltiple (FG+ vs. WT). Se graficó la media \pm SEM, $n = 3$ (A) y $n = 4$ (B) experimentos independientes.

En conjunto, estos resultados indican que el pretratamiento con IFN α potencia la restricción de MAYV, con un efecto significativamente mayor sobre la variante enriquecida en CpG. Considerando que IFN α induce la expresión de ZAP, y que los ensayos de silenciamiento demostraron que ZAPL es la principal isoforma antiviral, estos datos sugieren que el aumento en los niveles intracelulares de ZAP (en particular de ZAPL) constituye un determinante clave de la mayor sensibilidad de MAYV FG+ a la respuesta antiviral inducida por interferón.

ZAP regula negativamente la expresión transcripcional de INF β y de las citoquinas CXCL10 y CCL5, pero no OAS1 en células A549 infectadas con MAYV.

Además de su función antiviral asociada al reconocimiento de motivos CpG en los genomas virales, se ha demostrado que ZAPS, al igual que otras proteínas de la familia PARP, modula la vía de señalización mediada por IFN-I. En este sentido, se han descrito roles opuestos. Por un lado, Hayakawa, S *et al.* (2011) han reportado que ZAPS puede ser un potente estimulador del receptor de reconocimiento de patrones

RIG-I, dado que una mayor expresión de esta isoforma se asoció con una mayor inducción de IFN α tras la activación de células HEK293T humanas con 3p-RNA (ARN con el extremo 5' tri-fosforilado, agonista de RIG-I) ¹⁰⁶. Por otro lado, estudios más recientes han documentado que ZAP inhibe la inducción de IFN α tras la infección viral o la activación de PRRs con análogos sintéticos de dsARN en células Huh7 humanas.

Con el objetivo de evaluar el papel inmunomodulador de ZAP en la señalización de IFN-I durante la infección por MAYV, se cuantificó mediante RT-qPCR la expresión relativa del ARNm de IFN β en células A549 y A549 ZAP KO infectadas con MAYV WT o FG+ (MOI=3). En ausencia de ZAP, se observó una inducción significativamente mayor de IFN β tanto a las 12 como a las 24 HPI, independientemente de la variante viral analizada (**Fig. 2.19**). Estos resultados indican que ZAP limita la expresión de IFN β durante la infección por MAYV, además de su acción antiviral directa.

Para ampliar este análisis, se evaluó la expresión de genes inducidos por PRRs (CXCL10 y CCL5) y de un ISG canónico, la enzima 2'-5'-oligoadenilato sintetasa 1 (OAS1) dependiente de la vía IFNAR-JAK-STAT, en células A549 y A549 ZAP KO infectadas con MAYV (WT o FG+). La ausencia de ZAP provocó un aumento significativo en la expresión de CXCL10 y CCL5 en comparación con células que expresan ZAP, particularmente a las 24 HPI, lo que sugiere una activación transcripcional sostenida de PRRs (**Fig. 2.19**).

En contraste, la expresión de OAS1 no mostró diferencias significativas entre células A549 y A549 ZAP KO ni a las 12 ni a las 24 HPI, a pesar de observarse niveles elevados de IFN β en células ZAP KO. Este desacople entre la producción de IFN-I y la inducción del ISG OAS1 sugiere que ZAP no amplifica la señalización *downstream* del IFNAR, sino que posiblemente actúa principalmente restringiendo la disponibilidad de ARN viral en el interior celular y, por ende, la activación prolongada de PRRs.

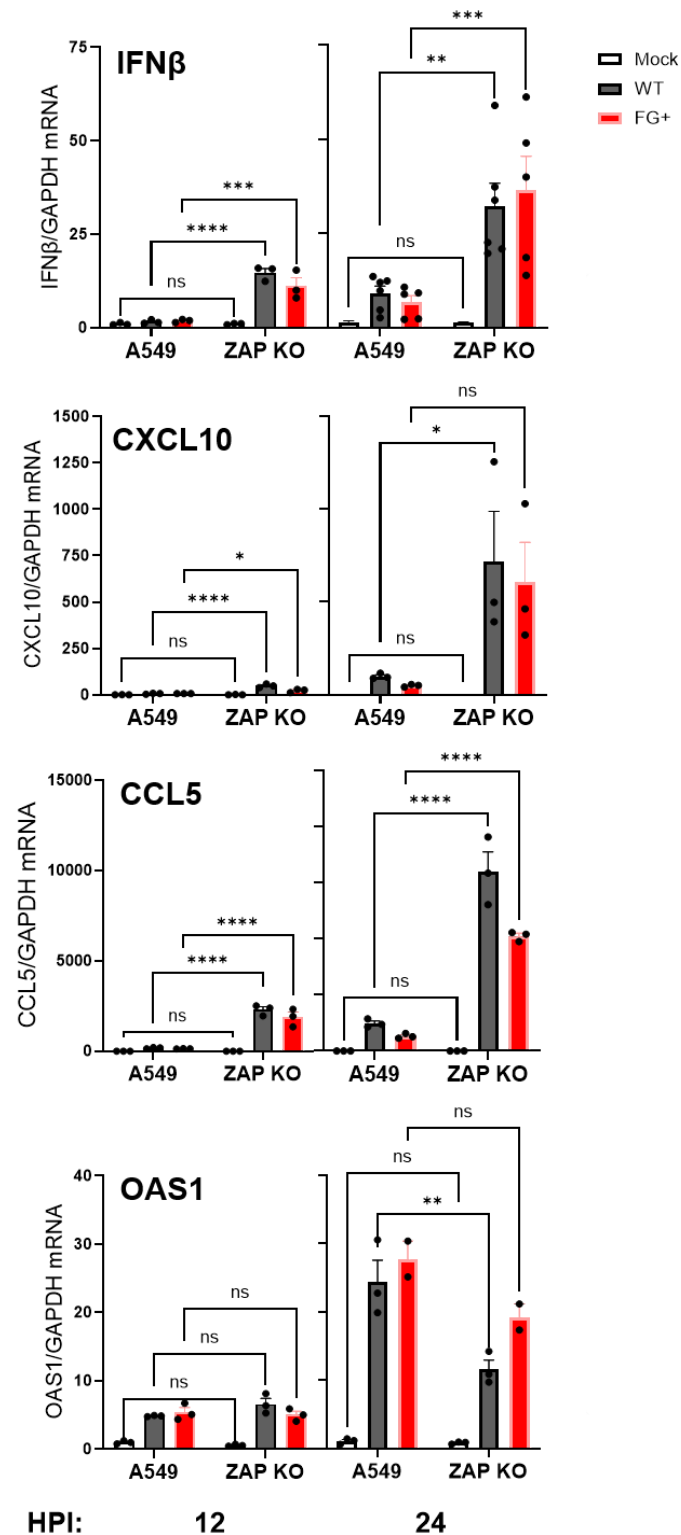


Figura 2.19: ZAP regula negativamente la expresión de IFN β , CXCL10 y CCL5, pero no de OAS1 en células A549 infectadas con MAYV WT y FG+. Cuantificación relativa del ARNm de IFN β , CXCL10, CCL5 y OAS1 por RT-qPCR. Células A549 o ZAP KO infectadas con MAYV WT y FG+, 12 HPI (gráfico izquierdo) y 24 HPI (gráfico derecho), a MOI = 3 o sin infectar (mock). Se graficó la media \pm SEM de n=3 experimentos independientes, la cual fue normalizada frente a células A549 sin infectar (mock) a las 12 HPI. ns, no significativo *p \leq 0,05; **p \leq 0,01; ***p \leq 0,001; ****p \leq 0,0001; two-way ANOVA y post-test Šidák's para comparaciones múltiples.

La atenuación de MAYV FG+ por IFN α requiere la expresión de ZAP.

Si bien el tratamiento con IFN α indujo fuertemente la expresión de ZAP y aumentó la susceptibilidad de MAYV FG+ (**Figs. 2.17 y 2.18**), no se puede descartar la contribución de otros ISGs en la atenuación observada, considerando la alta sensibilidad de los alfavirus a la respuesta de IFN-I y la actividad antiviral reportada de diversos ISGs frente a MAYV ^{100,158}.

En este contexto, se evaluó si el efecto antiviral del IFN α sobre MAYV FG+ podía depender de mecanismos independientes de ZAP.

Para ello, células A549 y A549 ZAP KO fueron pretratadas con IFN α O/N y posteriormente infectadas con MAYV WT o FG+ (MOI = 3). A las 24 HPI, se evaluó la replicación viral mediante titulación por ensayo de placa, se cuantificó el ARN genómico viral intracelular y la expresión de la proteína nsP1 por WB. Notablemente, en ausencia de ZAP, el pretratamiento con IFN α no logró inhibir la replicación de la variante MAYV FG+, no observándose diferencias significativas en los títulos virales entre FG+ y WT en células A549 ZAP KO tratadas con IFN α (**Fig. 2.20A**), ni en los niveles de ARN genómico (**Fig. 2.20B**) ni en la expresión de nsP1 (**Fig. 2.20C**). Estos resultados contrastan con lo observado en células A549, donde el IFN α reduce marcadamente la replicación viral, particularmente en la variante FG+.

Para verificar que ambas líneas celulares respondieron adecuadamente al tratamiento con IFN α , se cuantificó la expresión relativa de IFN β y OAS1. No se detectaron diferencias significativas en la inducción de estos genes, entre células A549 y A549 ZAP KO pretratadas con IFN α (**Fig. S2.2**), lo que indica que la vía de señalización de que el IFN α se encuentra funcional en ambas condiciones.

En conjunto, estos resultados indican que la actividad antiviral inducida por IFN α frente a MAYV FG+ depende críticamente de la presencia de ZAP, sugiriendo que este factor constituye un determinante central de la mayor sensibilidad de esta variante enriquecida en CpG a la respuesta interferón-dependiente.

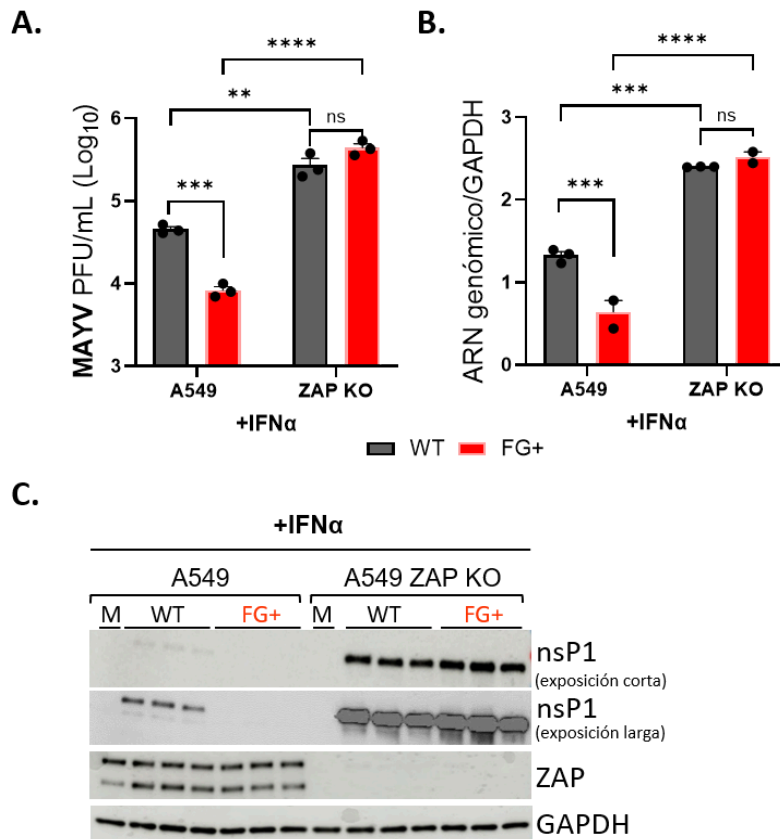


Figura 2.20: ZAP es necesario para el efecto antiviral del IFNα sobre la variante FG+. Células A549 y A549 ZAP KO fueron tratadas con 500 U/mL de IFNα O/N y posteriormente infectadas con MAYV WT o FG+ (MOI = 3) o mantenidas sin infección (mock). Las muestras se recolectaron a las 24 HPI. **A)** Titulación viral por ensayo de placa para MAYV WT y FG+. **B)** Cuantificación relativa de ARN genómico viral mediante RT-qPCR. **C)** Análisis por WB para las proteínas nsP1, ZAP y GAPDH. Se muestran los valores como media ± SEM de n=2 o 3 réplicas experimentales independientes. ns, no significativo *p ≤ 0,05; **p ≤ 0,01; ***p ≤ 0,001; ****p ≤ 0,0001; *two-way* ANOVA y post-test Šídák's para comparaciones múltiples.

ZAP reduce el ARN viral de MAYV FG+ en las primeras etapas de la infección.

Con el objetivo de caracterizar la cinética del efecto antiviral de ZAP, se analizó la dinámica temporal de la replicación viral en células A549 y A549 ZAP KO infectadas con MAYV WT o FG+ (MOI = 3), a partir de las 2 HPI. Para ello, se cuantificó el ARN genómico viral mediante RT-qPCR, se evaluó la expresión de la proteína viral nsP1 por WB (Fig. 2.21).

En células A549, se detectó una reducción en los niveles de ARN genómico de la variante FG+ en comparación con el virus WT a partir de las 2 HPI. Si bien esta diferencia no alcanzó a ser significativa a nivel estadístico, se observó una disminución cercana al 50%. A partir de las 4 HPI, la inhibición mediada por ZAP fue evidente en

todos los tiempos analizados y resultó estadísticamente significativa (**Fig. 2.21A**). La proteína nsP1 fue detectada a partir de las 12 HPI, y su expresión fue menor en células infectadas con MAYV FG+ en comparación con las infectadas con el virus WT (**Fig. 2.21C**, panel izquierdo), en concordancia con los niveles reducidos de ARN genómico observados.

En contraste, en células A549 ZAP KO, tanto los niveles de ARN viral (**Fig. 2.21B**) como la expresión de nsP1 (**Fig. 2.21C**, panel derecho) fueron comparables entre MAYV WT y FG+ en todos los tiempos postinfección evaluados. En conjunto, estos resultados sugieren que ZAP restringe la replicación de MAYV FG+ en células A549 mediante la reducción de las copias genómicas intracelulares desde etapas tempranas del ciclo viral.

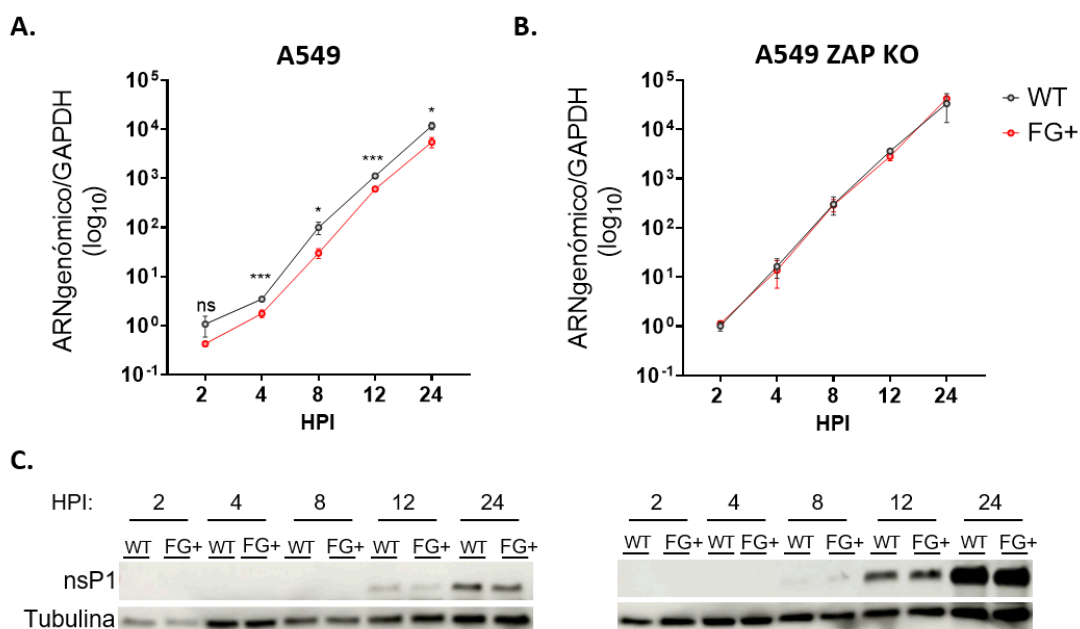


Figura 2.21: ZAP reduce los niveles de ARN genómico de MAYV FG+ a tiempos tempranos postinfección.

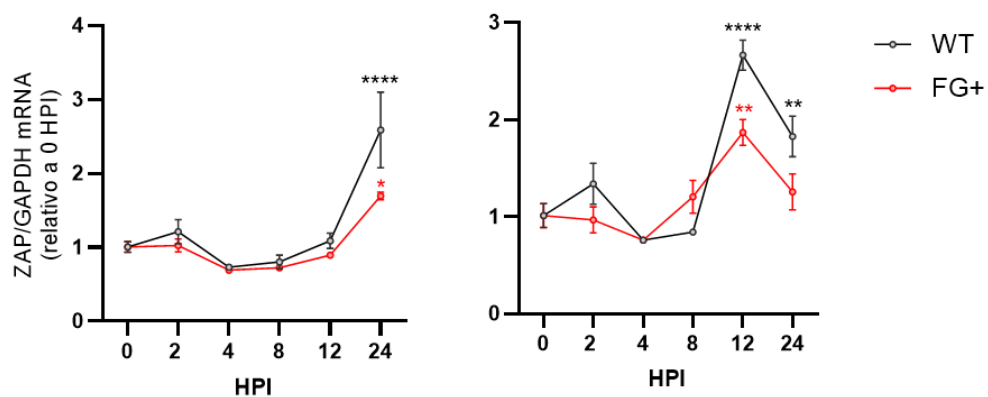
(A-B) Cuantificación relativa de ARN genómico viral mediante RT-qPCR en células A549 (A) y A549 ZAP KO (B) infectadas con MAYV WT o FG+ (MOI = 3). Los valores se normalizaron a GAPDH y se transformaron a escala log₁₀, expresándose de forma relativa a las 2 HPI (media ± SEM de n = 3 réplicas biológicas independientes). La significancia estadística se evaluó mediante prueba *t-Student* no pareada para comparar WT vs. FG+ en cada punto temporal, con ajuste de Holm-Šidák para comparaciones múltiples. *p ≤ 0,05; **p ≤ 0,01; ***p ≤ 0,001; ns: no significativo. **C)** WB para nsP1 en las mismas condiciones experimentales, utilizando tubulina como control de carga. Los paneles de la izquierda y derecha corresponden a células A549 y A549 ZAP KO, respectivamente.

Paralelamente, se caracterizó la cinética de expresión de ZAPL y ZAPS en células A549 infectadas con MAYV WT o FG+ (MOI = 3) a partir de las 2 HPI, evaluando tanto la respuesta transcripcional como la proteica (**Fig. 2.22A y 2.22B**). Ambas isoformas

mostraron una inducción tardía. En el caso de ZAPL, el incremento fue detectable únicamente a las 24 HPI, tanto a nivel de ARNm como de proteínas. Por su parte, para ZAPS mostró un incremento en los niveles de ARNm a partir de las 12 HPI, traduciéndose en una mayor acumulación de proteína hacia las 24 HPI.

A pesar de esta respuesta tardía, la variante FG+ mostró una disminución en la replicación del genoma desde las 2 HPI. Este hallazgo sugiere que los niveles basales de ZAP, particularmente ZAPL, constituyen una barrera antiviral constitutiva que actúa en etapas tempranas del ciclo viral, previa a la inducción mediada por la infección. En este sentido, la mayor restricción observada anteriormente bajo condiciones de pretratamiento con IFN α (Fig. 2.18) refuerza la idea de que la eficacia antiviral contra MAYV FG+ es estrictamente dependiente de la concentración intracelular de ZAP. Asimismo, la capacidad de MAYV para retrasar la inducción de ZAP sugiere un posible mecanismo de evasión viral orientado a proteger las etapas iniciales de la replicación, antes de que la respuesta inmune innata alcance niveles efectivos.

A.



B.

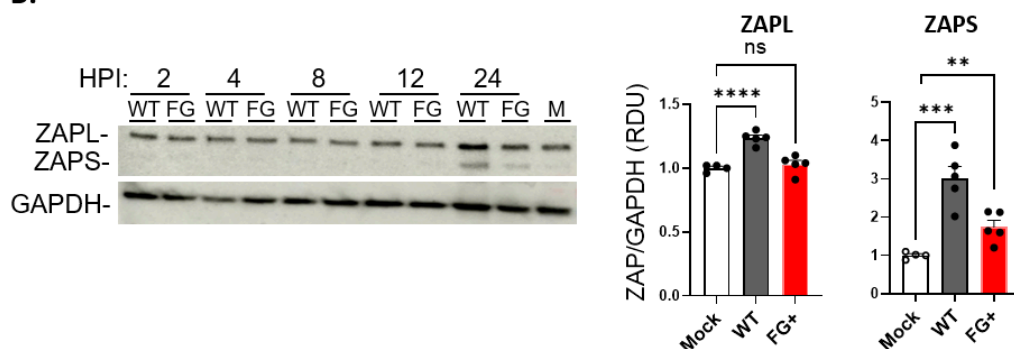


Figura 2.22: La activación de ZAP ocurre a tiempos tardíos post-infección. Células A549 fueron infectadas con MAYV WT o FG+ (MOI=3), las muestras se recolectaron a los tiempos indicados en cada figura. **A)** Cuantificación relativa por RT-qPCR del ARNm de ZAPL (gráfico izquierdo) o ZAPS (gráfico derecho), normalizados contra 0 HPI. **B)** WB representativo con anticuerpos específicos contra ZAP y GAPDH. En panel A, se muestran los valores como media \pm SEM de n=3 réplicas experimentales independientes. *t Student* para comparaciones múltiples (cada tiempo analizado vs. 0 HPI). *p \leq 0,05; **p \leq 0,01; ***p \leq 0,001.

Discusión

Los resultados presentados en este capítulo demuestran que ZAP desempeña un rol clave en la restricción de la variante sintética MAYV FG+ validando el enriquecimiento sinónimo de dinucleótidos CpG como una estrategia eficaz de atenuación viral. Mostramos que la incorporación sinónima de CpG en el genoma viral, basado exclusivamente en la incorporación de CpG presentes en aislados virales, sensibiliza a MAYV a la inhibición por ZAP, lo que respalda la hipótesis de que la subrepresentación de CpG observada en el genoma de MAYV (y otros virus ARN) supone una estrategia evolutiva para evadir la restricción por ZAP en hospederos vertebrados.

A nivel mecanístico, el desarrollo de una línea celular A549 ZAP KO mediante CRISPR/Cas9 permitió demostrar que la ausencia de ZAP revierte el fenotipo atenuado de la variante FG+ hacia niveles de replicación comparables al virus WT (**Fig. 2.15**, células A549 KO y **Fig. S2.1**, células A549 KD). En células que expresan ZAP, la variante FG+ presenta una disminución significativa en las copias genómicas virales y en la expresión de la proteína no estructural nsP1, mientras que sus niveles fueron comparables al WT en células ZAP KO (**Fig. 2.15** paneles B y C). En contraste, no se detectaron diferencias significativas en los niveles de ARNs subgenómicos ni en la expresión de la proteína estructural E1, sugiriendo que la restricción mediada por ZAP afecta principalmente etapas tempranas del ciclo viral o la estabilidad del ARN genómico, más que la síntesis de ARN subgenómico o la traducción de proteínas estructurales. Estos resultados son consistentes con un modelo en el cual ZAP actúa promoviendo la degradación del ARN viral o interfiriendo con la replicación inicial. De todas formas, se necesitan más estudios para explicar por qué el ARN subgenómico no se ve afectado por ZAP y cuál es el mecanismo específico de restricción.

Si bien la variante FG+ mostró una mayor sensibilidad a ZAP, nuestros resultados también indican que MAYV WT es susceptible a esta restricción, evidenciado por el aumento de su replicación en células ZAP KO. Este hallazgo amplía el espectro de alfavirus sensibles a ZAP, previamente descritos para otros miembros del género (incluyendo SINV, Semliki Forest, RRV, CHIKV y VEEV)¹³⁰, y sugiere que podría constituir un mecanismo antiviral conservado. Asimismo, este hallazgo es consistente con el análisis comparativo del sesgo de CpG en genomas virales (**Fig. 2.9**), que muestra que los arbovirus, si bien presentan una baja frecuencia de CpG, mantienen niveles relativamente más altos que los virus que están estrictamente adaptados a vertebrados. Esto podría explicar la sensibilidad de MAYV WT a la restricción ejercida por ZAP. Sería interesante utilizar herramientas basadas en ensayos de cross-linking e inmunoprecipitación de ARN, tales como RIP o CLIP-seq, que permitirían identificar de

manera directa las regiones específicas del ARN viral asociadas a ZAP durante la infección, así como determinar si existe una unión preferencial al ARN genómico en comparación con el ARN subgenómico y las regiones de unión al genoma de MAYV WT.

En relación con el rol de los CpG, estudios previos han sugerido que la susceptibilidad a la restricción mediada por ZAP dentro del género alfavirus no depende exclusivamente del contenido global de CpG en el genoma viral, sino también de su distribución. Nguyen y colaboradores reportaron que una mayor sensibilidad a ZAP no está mediada por una mayor unión de esta proteína al ARN, sino que se correlaciona con una mayor frecuencia de CpG en una ventana de 500 nt en el ORF de nsP2 en los alfavirus estudiados ¹⁵⁶. Sin embargo, los resultados obtenidos por nuestro grupo con la variante FG+ (la cual presentó mayor atenuación dentro de las tres variantes desarrolladas en nuestro laboratorio y contiene el mayor número de CpG, **Fig 2.11A**) sugieren que un aumento y distribución global de CpG también puede potenciar significativamente la restricción antiviral. Nuestros hallazgos son consistentes con estudios en variantes sintéticas de ZIKV, donde un mayor contenido en CpGs se asocia con una mayor sensibilidad a ZAP y una mayor atenuación viral ¹⁵⁹. Asimismo, en otros virus ARN como VIH-1 e Influenza A, se ha demostrado que no solo la cantidad, sino también la distribución y el espaciamiento de los CpGs influyen en la sensibilidad a ZAP ^{160, 161}.

El análisis temporal de la infección reveló que la restricción mediada por ZAP ocurre desde etapas muy tempranas del ciclo viral. En células A549, se observó una reducción en la replicación viral de FG+ respecto al virus WT a partir de las 2 HPI, previo a la inducción transcripcional de ZAP (tanto ZAPL o ZAPS) (**Fig. 2.21A** y **Fig. 2.22A**), lo que sugiere que la expresión constitutiva de ZAP, en particular de la isoforma ZAPL, actúa como una barrera antiviral intrínseca que limita la replicación viral en etapas tempranas. Esta observación está en línea con reportes previos que destacan un rol antiviral basal de ZAP frente a otros virus ARN ^{115,133,160}.

También observamos que la inducción de ZAP, tanto ZAPL como ZAPS, ocurre de manera tardía durante la infección (**Fig. 2.22**). A modo comparativo, en otros estudios que han caracterizado la dinámica temporal en la inducción de ZAP tras la infección viral, como por ejemplo, en células infectadas con HCMV o con echovirus 7 se ha observado una inducción más temprana de ZAP, entre las 2 y 4 HPI ^{143,162}. Estos resultados sugieren que MAYV podría estar interfiriendo con la inducción temprana de ZAP como estrategia de evasión antiviral. Estudios previos han reportado que los alfavirus producen un silenciamiento transcripcional y traduccional generalizado en

células infectadas (conocido como *transcriptional* o *translational shutoff*), como un mecanismo eficiente para contrarrestar la respuesta inmune del hospedero ¹⁶³. Para MAYV se ha demostrado que la proteína viral nsP2 promueve la degradación del factor TFIIIE2 (del inglés, *transcription initiation factor IIE subunit 2*) y de la subunidad Rpb1 de la ARN polimerasa II, bloqueando la transcripción dependiente de esta polimerasa ⁸². Sería interesante investigar en un futuro si esta proteína viral contribuye a la inducción limitada de ZAP en las etapas más tempranas de la infección en nuestras condiciones experimentales como mecanismo de evasión de la respuesta inmune.

Por otro lado observamos que la variante FG+ induce niveles más bajos de ZAP en comparación con el virus WT (**Fig. 2.12**), lo que sugiere que un mayor contenido genómico de CpG no es suficiente por sí solo para desencadenar una activación de ZAP, sino que la expresión de esta proteína podría estar más relacionada a la replicación viral y la activación de PRRs. Una posible explicación es que tras la infección con la variante FG+ haya una menor activación de sensores citosólicos de ARN viral, como RIG-I y MDA5, cuya expresión se ha visto aumentada por la infección con MAYV ¹⁰⁰. Aunque en este trabajo no evaluamos directamente la activación de estos sensores, sí observamos que la variante FG+ mostró una menor inducción de los transcritos de las quimioquinas inflamatorias CCL5 y CXCL10, así como una tendencia hacia niveles más bajos de IFN β , en comparación con el virus WT, lo que es consistente con una activación más débil de la respuesta mediada por PRRs. (**Fig. 2.19**). En este contexto, el pretratamiento de células A549 con IFN α permitió disociar el efecto del contenido de CpG de los niveles de expresión de ZAP. La estimulación con IFN α promovió una fuerte inducción de ZAPS y, en menor medida de ZAPL en células infectadas con ambos virus (**Fig. 2.17**), y bajo estas condiciones, se observó una inhibición significativamente mayor de la replicación de MAYV FG+ en comparación con el virus WT (**Fig. 2.18**). Por el contrario, este efecto se perdió completamente en células ZAP KO, donde el pretratamiento con IFN α no logró inhibir la variante FG+, demostrando el rol crucial de ZAP en la restricción mediada por IFN α (**Fig. 2.20**).

Estos resultados sugieren que la sensibilidad de MAYV FG+ no solo depende de su contenido de CpG, sino que está fuertemente influenciada por el nivel de expresión de ZAP y la activación de la respuesta antiviral. En línea con esto, estudios realizados en HIV-1 han demostrado que la introducción sinónima de CpG en las regiones *gag* o *pol* del genoma viral no genera una atenuación significativa bajo niveles endógenos de ZAP. Sin embargo, tras aumentar la expresión de ZAP, tanto con tratamiento con IFN-I o mediante una sobreexpresión de ZAP de forma heteróloga, se logró convertir estas

variantes originalmente resistentes a ZAP, en variantes sensibles¹⁶⁰. En resumen, los resultados sugieren que la restricción viral mediada por ZAP resulta de una interacción dinámica entre la composición del genoma viral y el nivel de activación de los mecanismos antivirales del hospedero.

En este contexto, cofactores como TRIM25, KHNYN y Riplet también podría contribuir a la mayor inhibición observada de la variante FG+ tras el tratamiento con IFN α . No obstante, el hecho de que el pretratamiento con IFN α no haya logrado inhibir la replicación de FG+ en células ZAP KO (**Fig. 2.20**) sugiere que la presencia de estos cofactores, en ausencia de ZAP, no es suficiente para ejercer un efecto antiviral independiente. Este resultado refuerza el papel indispensable de ZAP como mediador principal de la restricción viral de la variante FG+ inducida por IFN α .

Además de su acción antiviral, la isoforma ZAPS modula la respuesta inmune innata con funciones duales sobre el IFN-I. Mientras que algunos estudios la describen como un potenciador de la activación de PRRs (como RIG-I) y de la inducción de IFN β , investigaciones recientes sugieren que actúa como un mecanismo de retroalimentación negativa, suprimiendo la producción de interferón tras la detección de ARN viral¹¹⁵. En este sentido, nuestros resultados con MAYV son consistentes con un rol de ZAP como regulador negativo de la respuesta inmune innata. En células ZAP KO, la infección indujo una activación más intensa y sostenida de IFN β , así como de CXCL10 y CCL5, ambas implicadas en el reclutamiento de células inmunes y asociadas a la inmunopatogénesis de alfavirus (**Fig. 2.19**). En particular, en modelos murinos que desarrollan artritis tras la infección viral con CHIKV y ONNV, CXCL10 se ha asociado con una mayor persistencia viral y daño articular, sugiriendo un papel en la cronificación de la respuesta inflamatoria¹⁶⁴. Asimismo, se ha reportado que la infección de células neuronales humanas con MAYV induce niveles elevados de CXCL10, lo que apunta a una posible implicancia de esta quimioquina en la neuropatogénesis¹⁶⁵. En el caso de CCL5, se ha observado que su expresión se ve aumentada considerablemente en tejidos musculares de ratones jóvenes y en ratones adultos IFNAR KO, ambos modelos altamente susceptibles a la infección por MAYV. Este aumento se correlacionó con inflamación muscular severa y necrosis tisular, sugiriendo que CCL5 facilita el reclutamiento de células inmunes que agravan el daño tisular durante la fase aguda de la infección⁹⁹. En contraste, en modelos murinos adultos e inmunocompetentes, que no desarrollan síntomas clínicos graves, no se observaron cambios significativos en la expresión de CCL5. Estos resultados indican que la expresión exacerbada de esta quimioquina está asociada con un fenotipo clínico más severo y al establecimiento del daño muscular. En conjunto, nuestros

resultados donde se observó aumento de CXCL10 y CCL5 en células ZAP KO (**Fig. 2.19**) sugieren que ZAP podría tener un rol importante en el control de la inmunopatología inducida por MAYV al modular la expresión de estas quimioquinas.

En contraste, la expresión de OAS1, un ISGs dependiente de la señalización a través de IFNAR, no mostró diferencias significativas en su expresión entre células A549 y A549 ZAP KO, a pesar de observarse mayores niveles de IFN β (**Fig. 2.19**). Este desacople entre la producción de IFN-I y la inducción de ISGs sugiere que ZAP no modularía directamente la señalización *downstream* de IFNAR, sino que podría estar actuando principalmente limitando la disponibilidad de ARN viral citoplasmático, restringiendo la activación prolongada de PRRs y, en consecuencia, regulando la magnitud y duración de la respuesta antiviral *upstream* de la activación por IFNAR.

No obstante, serían necesarios estudios adicionales para profundizar en estos resultados. Un abordaje transcriptómico global permitiría definir con mayor precisión el impacto de ZAP sobre la regulación coordinada de genes de la respuesta inmune innata y evaluar la posible existencia de umbrales o mecanismos de saturación en la señalización *downstream* de IFNAR.

Nuestros resultados son consistentes con reportes recientes en modelos murinos ZAP KO, donde se observa una expresión prolongada de IFN-I tras la activación de PRRs con análogos de ácidos nucleicos, sin observarse una inducción proporcional de ISGs canónicos ¹²⁶. Sería interesante extender estos estudios en un contexto de infección por MAYV en modelos murinos ZAP KO, para caracterizar si ZAP (o específicamente la isoforma ZAPS) puede actuar como un modulador clave de la homeostasis inmune previniendo un mayor daño tisular.

En conjunto, nuestros hallazgos apoyan un modelo en el cual ZAP cumple una doble función durante la infección por MAYV: por un lado, actúa como un factor antiviral intrínseco que restringe la replicación viral mediante el reconocimiento de ARN enriquecido en CpG; particularmente en etapas tempranas del ciclo infeccioso. Por otro lado, modula la respuesta inmune innata al limitar la activación excesiva de PRRs y la producción de citoquinas proinflamatorias. En este sentido, ZAP no solo contribuiría al control de la replicación viral, sino también a la regulación de la magnitud y duración de la respuesta inmune, lo que podría impactar en la prevención de la inmunopatología asociada a la infección (**Fig. 2.23**).

Además, los resultados sugieren que potenciar la acción de ZAP podría ser una estrategia efectiva para controlar la infección por MAYV y posiblemente por otros alfavirus. Este estudio ofrece nuevas perspectivas sobre la interacción entre los factores de restricción del hospedero, la composición genómica viral y la respuesta

inmune, con implicancias relevantes para la patogénesis viral y el diseño racional de vacunas atenuadas.

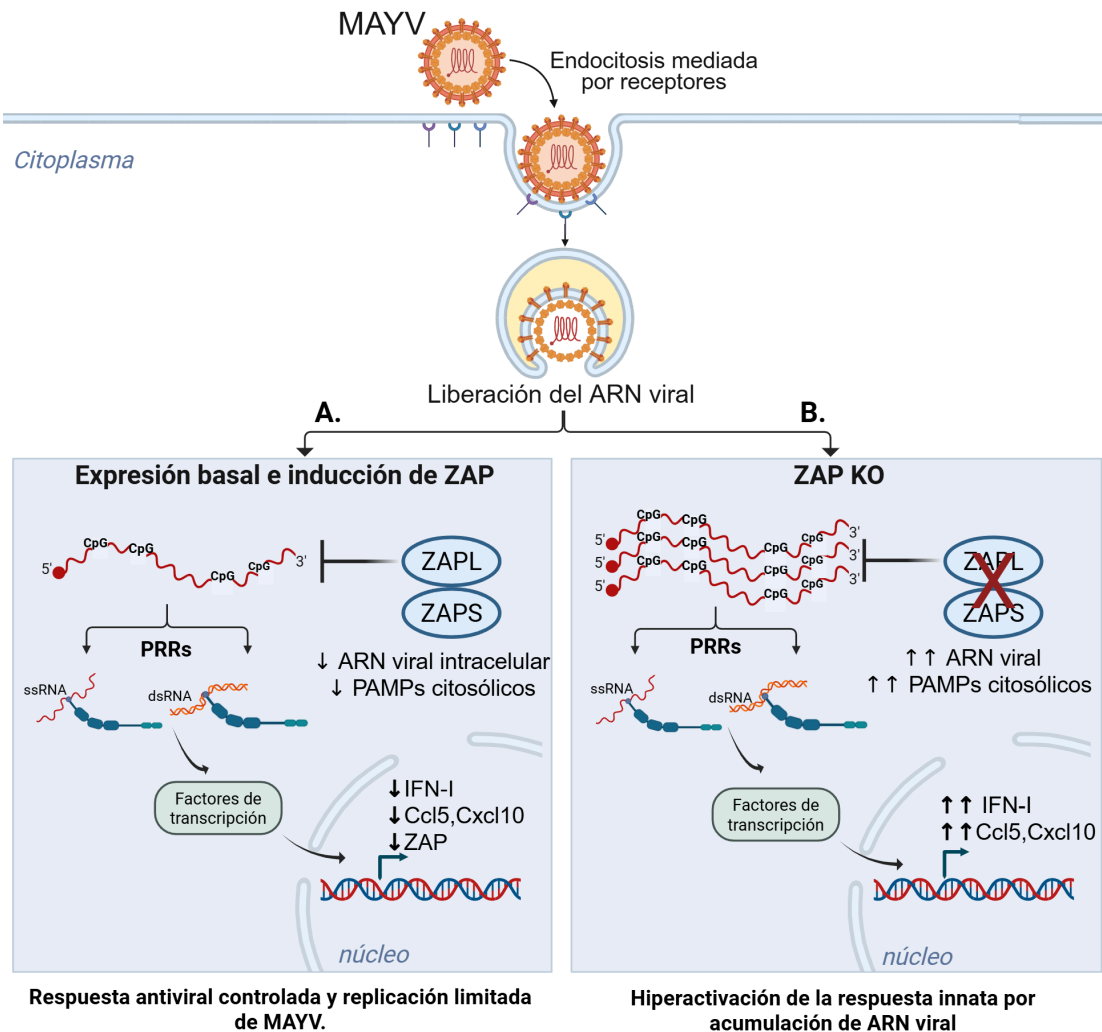


Figura 2.23: Modelo propuesto del papel de ZAP en el control de la infección por MAYV. A partir de nuestros resultados, proponemos un modelo hipotético que podría explicar nuestras observaciones. En condiciones normales (panel A), ZAP restringe la replicación de MAYV, particularmente de la variante enriquecida en CpG, disminuyendo los niveles de ARN viral intracelular y contribuyendo a mantener regulada la respuesta inmune innata. En este contexto, la expresión basal de ZAPL actúa como un factor antiviral intrínseco clave. ZAPS se induce moderadamente tras la infección por MAYV y requiere el tratamiento con IFN α para potenciar su expresión. Además, MAYV podría estar retrasando la inducción de ambas isoformas para evadir su restricción. En ausencia de ZAP (panel B), la acumulación de ARN viral intracelular e intermediarios de replicación incrementa la activación de PRRs, promoviendo una mayor producción de interferón y quimioquinas. Esta respuesta exacerbada podría contribuir a un mayor daño tisular; por lo tanto, nuestros resultados sugieren que ZAP podría representar un posible blanco para estrategias antivirales dirigidas a limitar la replicación viral y la diseminación del virus.

Conclusiones

- Se generó una línea A549 ZAP KO, que permitió demostrar que ZAP es indispensable para la restricción de MAYV FG+ en células humanas.
- El enriquecimiento de CpG aumenta la sensibilidad de MAYV a ZAP, lo que sugiere que la supresión de CpG podría contribuir a modular la interacción del virus con este factor antiviral.
- La isoforma ZAPL constituye el principal factor de restricción celular frente a la variante MAYV FG+, actuando como un factor intrínseco que limita la replicación viral en células humanas.
- La restricción viral inducida por IFN α depende directamente de la inducción de ZAP, donde posiblemente otros ISGs no compensan su función antiviral en este contexto.
- MAYV FG+ desencadena una respuesta inmune innata más débil que el virus WT, reflejada en menor activación de ZAP. Para ambos virus la inducción de ZAP fue tardía, sugiriendo que MAYV podría evadir parcialmente la respuesta antiviral mediante modulación de la cinética de activación de ZAP.
- ZAP limita la magnitud y duración de la inducción de citoquinas durante la infección por MAYV posiblemente evitando una activación excesiva de PRRs, lo que sugiere un rol adicional en la regulación de la inmunopatología asociada a la infección.

Perspectivas

Los resultados obtenidos en el marco de esta tesis abren nuevas líneas de trabajo a futuro, orientadas a profundizar en los mecanismos que subyacen a esta interacción en sistemas más complejos. Una limitación de este trabajo es que los modelos celulares utilizados no capturan completamente la complejidad de la respuesta inmunológica sistémica durante una infección viral en un organismo completo. En este sentido, la transición hacia modelos murinos deficientes en ZAP (ZAP KO) permitirá evaluar el impacto de esta proteína sobre la replicación viral, la diseminación sistémica y la respuesta inflamatoria en un contexto fisiológico.

La utilización de estos modelos también permitirá estudiar el papel de esta proteína en la regulación de la producción de IFN-I y de quimioquinas dependientes de RLRs durante la infección. Dado que la infección por MAYV induce quimioquinas proinflamatorias como CXCL10 y CCL5, asociadas al daño articular y muscular característico de la enfermedad, estos modelos permitirían evaluar si la ausencia de ZAP exacerba la respuesta inflamatoria *in vivo* y agrava la sintomatología clínica o favorece la cronificación de la enfermedad.

Además, el hallazgo de que MAYV es susceptible a la restricción mediada por ZAP, y que potenciar su expresión limita la replicación viral y la inducción de citoquinas, abre una línea de investigación orientada a explorar el potencial antiviral de esta vía. Futuros estudios podrían evaluar estrategias destinadas a modular la actividad o los niveles de expresión de ZAP, por ejemplo mediante agonistas de la respuesta de interferón u otros compuestos capaces de potenciar su expresión.

Finalmente, será relevante profundizar en los determinantes moleculares que regulan la interacción entre ZAP y el ARN viral. Estudios basados en técnicas de inmunoprecipitación de ARN (RIP/CLIP-seq) permitirán identificar las regiones específicas del genoma de MAYV reconocidas por ZAP y evaluar si la restricción depende de la distribución local de CpG o de su contenido global. Estos enfoques contribuirán a comprender con mayor precisión los principios que gobiernan la sensibilidad viral a ZAP y a optimizar estrategias de diseño viral basadas en este mecanismo.

CAPÍTULO 3

Desarrollo de variantes de MAYV como vectores de ARN guías (sgRNAs) para la edición génica *in vitro*, hacia la identificación de factores del hospedero que afecten el ciclo replicativo viral

Introducción

CRISPR/Cas9 para el estudio funcional de interacciones virus–hospedero

La herramienta de edición génica dirigida CRISPR/Cas9 ha sido clave para el desarrollo de estudios funcionales del genoma, tanto en cultivos celulares como en modelos *in vivo*, consolidándose como una de las metodologías más utilizadas en biología molecular para este fin ¹⁶⁶. En el contexto de las infecciones virales, su aplicación ha permitido la identificación sistemática de factores del hospedero involucrados en la replicación viral o en mecanismos de defensa antiviral.

En particular, los *CRISPR screens*, basados en la edición dirigida de genes celulares mediante bibliotecas de sgRNAs clonados en vectores lentivirales, han sido fundamentales para comprender procesos como la entrada viral, replicación, ensamblaje y evasión inmune, así como para identificar posibles blancos terapéuticos ^{78,167,168}.

En este tipo de abordajes, se selecciona una línea celular permisiva y susceptible que permita una replicación viral eficiente y, en el caso de estrategias basadas en viabilidad, presente efectos citopáticos notorios. En la **figura 3.1** se muestra el flujo experimental de este abordaje, en donde los sgRNAs se diseñan *in silico* o se utilizan bibliotecas comerciales preexistentes, seguidas de su clonación y validación para generar una biblioteca agrupada (*pooled*). Posteriormente, las células se transducen con la biblioteca de sgRNAs diseñada a escala genómica mediante sistemas lentivirales, generando una población celular mutada, en la cual cada célula porta la inactivación de un gen distinto.

Posteriormente, la población celular se somete a infección viral y posiblemente a una presión selectiva de interés. Dependiendo del diseño experimental, la selección puede basarse en supervivencia/muerte celular o en separación mediante citometría de flujo, por ejemplo, utilizando virus reporteros fluorescentes. Tras la selección, se extrae el ADN genómico de las poblaciones celulares y se amplifican las secuencias de sgRNA integradas al genoma para su análisis mediante secuenciación de nueva generación (NGS). El análisis bioinformático posterior permite cuantificar el enriquecimiento o la depleción de cada sgRNA respecto a la población control, generando una lista jerarquizada de genes candidatos. Posteriormente, estos factores del hospedero se deben validar experimentalmente y caracterizarse en estudios mecanísticos que investiguen su papel en etapas específicas del ciclo replicativo viral.

Además del enfoque clásico de pérdida de función mediado por la nucleasa Cas9 (CRISPR KO), existen variantes como CRISPR de interferencia (CRISPRi) o de activación (CRISPRa), que utilizan una Cas9 catalíticamente inactiva (dCas9) fusionada a dominios represores o activadores para silenciar o activar la expresión génica respectivamente. Estas herramientas permiten modular la expresión génica sin generar cortes en el ADN, ampliando el espectro de aproximaciones funcionales disponibles, aunque no serán desarrolladas en detalle en esta tesis.

Análisis funcional de factores del hospedero implicados en la infección viral mediante CRISPR screening a escala genómica

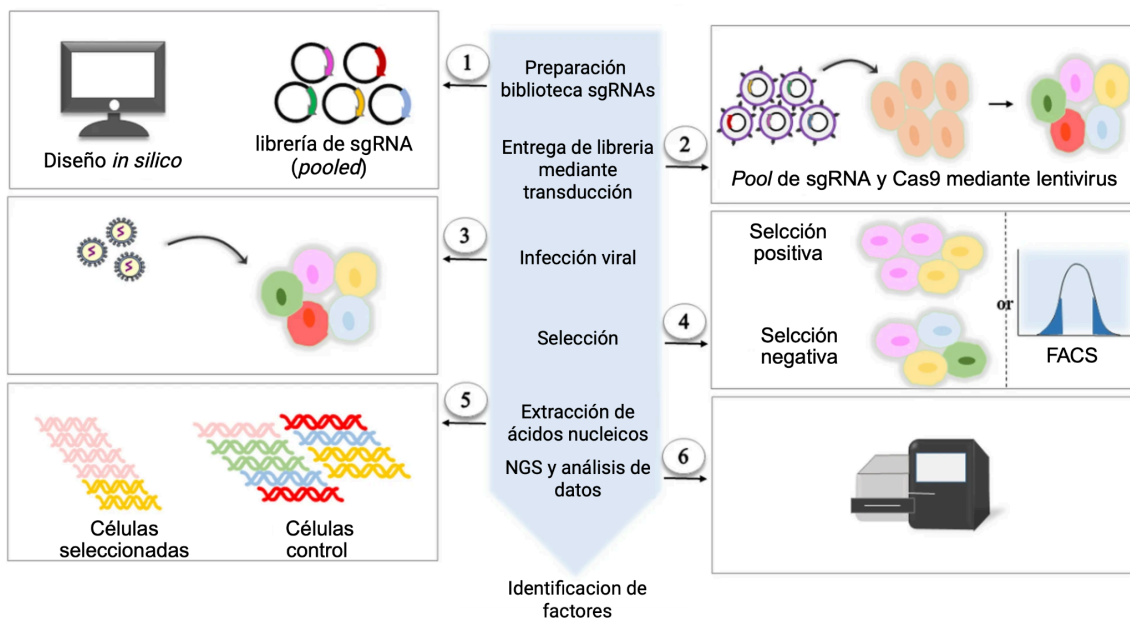


Figura 3.1: Esquema general del flujo de trabajo de un cribado funcional basado en CRISPR screens. (1) Los sgRNAs se diseñan *in silico* o se utilizan bibliotecas comerciales preexistentes, seguidas de su clonación y validación para generar una biblioteca agrupada (*pooled*). (2) Las sgRNA se empaquetan en lentivirus y se transducen en células diana, induciendo inactivación génica (knockout) o modulación de la expresión génica según la plataforma empleada. (3) La población celular editada se somete a infección con el virus en estudio. (4) Tras la infección, se aplica una selección basada en viabilidad. En infecciones con virus citolíticos, los ensayos de supervivencia permiten una selección positiva de células resistentes, en las que la pérdida de factores provirales favorece la supervivencia. Los cribados negativos identifican genes cuya inactivación compromete la viabilidad bajo presión selectiva, siendo útiles para detectar genes esenciales. Alternativamente, puede emplearse clasificación celular activada por fluorescencia (FACS) en infecciones persistentes o con virus no citolíticos marcados con reporteros fluorescentes. (5) Se extrae ADN genómico de células seleccionadas y controles, y se amplifican por PCR las secuencias guía integradas. (6) El análisis mediante NGS compara la representación relativa de sgRNA entre poblaciones, permitiendo identificar genes enriquecidos o depletados asociados al fenotipo de interés. Adaptada de ¹⁶⁹

Limitaciones de los CRISPR *screens* convencionales

En su formato más comúnmente utilizado, los CRISPR *screens* en virología se basan en la transducción previa de células con bibliotecas lentivirales que codifican sgRNAs, seguida de selección celular editada mediante FACS o selección con antibióticos, y posterior infección con el virus de interés. Este esquema ha demostrado ser extremadamente útil, pero introduce limitaciones a la hora de capturar interacciones dinámicas o dependientes del contexto replicativo.

En este sentido, los pasos de transducción y selección pueden generar heterogeneidad en la población celular editada. Por otro lado, la edición génica ocurre antes de la infección, lo que dificulta el análisis de procesos que dependen estrictamente del estado infeccioso o que se manifiestan en etapas tardías del ciclo viral. Además, muchos de estos estudios utilizan la viabilidad celular como principal criterio de selección e indicador de la edición génica, donde se analizan las células resistentes a la infección. Si bien este enfoque ha permitido identificar genes cuya alteración confiere resistencia celular a la infección y se trata de una estrategia extremadamente valiosa, tiende a favorecer la detección de factores celulares que intervienen en etapas tempranas de la replicación viral, en particular durante la entrada del virus a la célula, lo que limita el espectro de factores funcionales identificables.

Estas limitaciones han impulsado el desarrollo de estrategias que integren directamente la edición génica con la replicación viral.

Virus replicativos como plataformas para CRISPR *screens* funcionales

Con el fin de superar algunas de estas limitaciones, diversos grupos han desarrollado una alternativa a los CRISPR *screens* convencionales basados en lentivirus. Esta alternativa se basa en un sistema que utiliza virus replicativos con genomas modificados para expresar sgRNAs ¹⁷⁰⁻¹⁷³. En estos sistemas, las bibliotecas de sgRNA se expresan directamente desde el genoma viral, lo que permite monitorear de forma directa la propagación del virus a lo largo de su ciclo de replicación. Tras la infección de células que expresan la proteína Cas9 de manera constitutiva, los sgRNA codificados por el virus se producen en el interior celular tras la infección y dirigen la edición específica de genes celulares. Esta modificación genética afecta la cinética de replicación viral, en donde los niveles de sgRNA funcionan como una medida directa de la propagación del virus (**Fig. 3.2**) ¹⁷⁴. De esta forma, en un experimento de pérdida de función génica (KO), los sgRNA dirigidos contra genes esenciales para la replicación viral se verán subrepresentados en las partículas virales producidas,

mientras que aquellos que apuntan contra genes con efecto restrictivo de la replicación viral se verán enriquecidos.

CRISPR *screens* basados en virus codificantes para sgRNAs

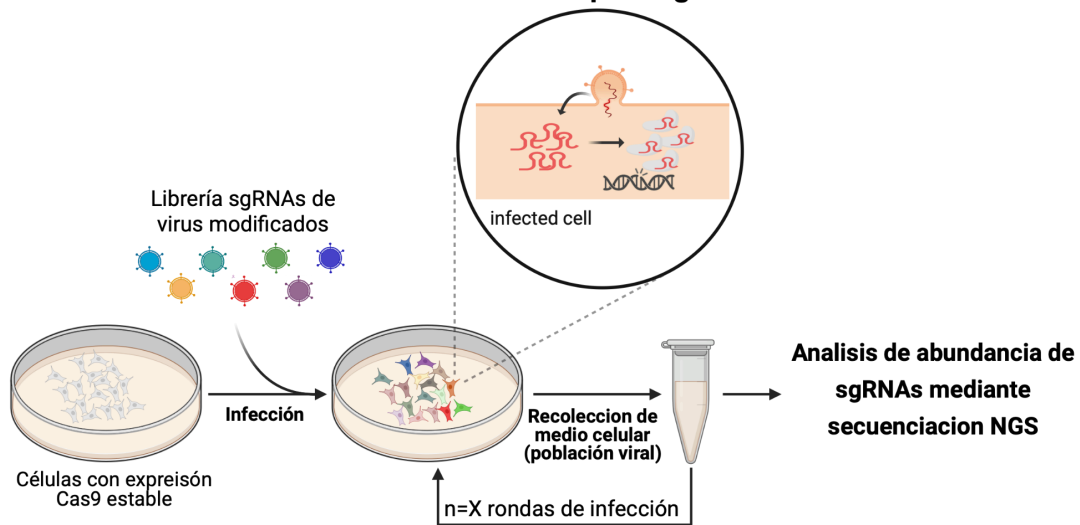


Figura 3.2: Esquema del funcionamiento de un *CRISPR screen* mediado por una librería de virus codificantes para sgRNA. Las bibliotecas de sgRNAs se expresan directamente desde el genoma viral del virus en estudio, lo que permite monitorizar de forma directa la propagación del virus a lo largo de todo el ciclo de infección. Tras la infección de una célula que expresa la proteína Cas9, la sgRNA codificada por el virus es liberada y modula la expresión de genes celulares. Esta alteración impacta en la cinética replicativa viral, de modo que los niveles de cada sgRNA constituyen una lectura directa de la propagación del virus. En consecuencia, las sgRNA dirigidas contra genes esenciales para la replicación viral tenderán a estar subrepresentadas, mientras que aquellas que inactivan factores restrictivos aparecerán enriquecidas en las partículas virales producidas por células que expresan Cas9.¹⁷⁴

Este enfoque ha sido implementado en distintos sistemas virales. En el caso de VIH-1, se utilizaron genomas lentivirales capaces de transportar sgRNAs empaquetados en viriones infecciosos, lo que permitió evaluar la eficiencia replicativa tras la inactivación sistemática de genes celulares. Mediante el análisis de la representación diferencial de sgRNAs tras múltiples ciclos de replicación, estos estudios identificaron factores antivirales (como MxB, TRIM5 α , IFITM1 y Tetherin) así como factores de dependencia que favorecen la replicación viral¹⁷². De manera similar, un sistema denominado VECOS (del inglés, *virus-encoded CRISPR-based direct readout system*) desarrollado para el citomegalovirus humano (HCMV) permitió expresar bibliotecas de sgRNAs desde el genoma viral a partir de la expresión de sgRNAs desde un promotor eucariota y cuantificar su abundancia a lo largo de pasajes seriados, identificando factores críticos en distintas etapas del ciclo infeccioso¹⁷⁰. Asimismo, en influenza A, se desarrollaron construcciones capaces de expresar sgRNAs codificados en una secuencia intrónica del gen no estructural (NS) que es procesado durante la

replicación viral, sin comprometer la producción de proteínas virales esenciales. Este sistema se basó en la activación de genes en células que expresan dCas9 de forma constitutiva y demostró que la propagación viral puede emplearse como una métrica directa de selección funcional, permitiendo identificar factores celulares provirales ¹⁷¹.

Una ventaja clave de estos sistemas frente a los CRISPR *screens* convencionales (**Fig. 3.1**) es que, dado que durante cada ciclo de replicación se sintetizan miles de copias del genoma viral, la señal asociada al impacto funcional de cada sgRNA se amplifica de manera natural. Esto aumenta significativamente la sensibilidad del análisis, permitiendo detectar efectos tanto positivos como negativos sobre la replicación viral ¹⁷⁰.

En conjunto, esta metodología ofrece mayor resolución funcional que los CRISPR *screens* tradicionales y permite analizar etapas avanzadas del ciclo infeccioso que suelen quedar subrepresentadas en los enfoques basados en viabilidad celular. Además, al acoplar la edición génica a la dinámica replicativa del virus, amplía el espectro de factores identificables en las interacciones virus–hospedero y abre nuevas posibilidades para el diseño racional de estrategias antivirales.

Alfavirus como plataforma de expresión de secuencias heterólogas.

Dentro de los virus ARN, los alfavirus ofrecen características particularmente atractivas para la expresión de genes o ARN de forma heteróloga. Como se describió anteriormente (en la introducción del capítulo 2, **Fig. 2.2** y **Fig. 2.3**), el genoma de +ssRNA de alfavirus está organizado en dos ORF funcionales: una región no estructural traducida directamente a partir del genoma, y una región estructural cuya expresión depende de la síntesis de un ARN subgenómico transcrito desde un promotor interno, denominado promotor subgenómico.

Este sistema, de expresión de ARN subgenómicos, ha sido ampliamente explotado para la expresión de genes heterólogos en clones infecciosos recombinantes, incluyendo proteínas fluorescentes, enzimas reporteras como la nano luciferasa (Nluc) o también para la expresión de ARNs pequeños como miRNA a partir de la incorporación de un locus pri-miR (**Fig. 3.3**) ^{175–178}. En este contexto, la introducción de un segundo promotor subgenómico permite la expresión de secuencias heterólogas sin comprometer necesariamente la replicación viral. Dado que la transcripción del ARN subgenómico depende estrictamente del complejo de replicación viral, se asegura de que la expresión de secuencias heterólogas ocurra únicamente en células donde el virus se está replicando activamente.

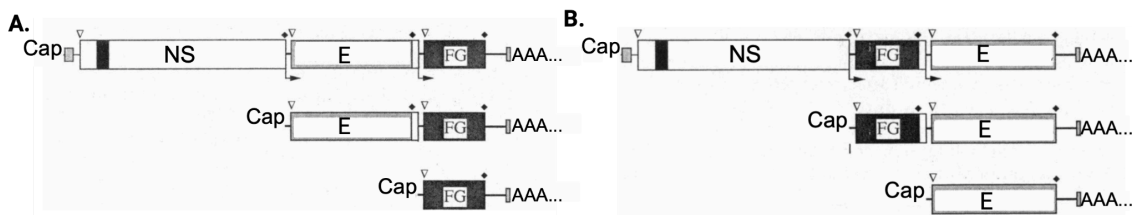


Figura 3.3 Genoma de alfavirus con doble promotor subgenómico doble para la expresión de genes heterólogos. Clones infecciosos de alfavirus que contienen tanto la maquinaria de replicación como las proteínas estructurales (se muestran como NS y E respectivamente). Los genes heterólogos (FG, del inglés *foreign gene*) se expresan mediante la duplicación del promotor subgenómico entre los ORF NS o E (panel A) o hacia el extremo 3' (panel B). Las señales de inicio de traducción (AUG) y de terminación (TRM) se indican mediante triángulos abiertos y rombos sólidos, respectivamente. Adaptada de ¹⁷⁶.

Origen del proyecto presentado en este capítulo y justificación del uso de MAYV como vector de sgRNAs.

Por lo explicado en el apartado anterior, MAYV constituye un sistema adecuado para evaluar experimentalmente la viabilidad de incorporar y expresar sgRNAs del sistema CRISPR/Cas9 desde su genoma viral con el objetivo de inducir edición génica en células infectadas.

Esta línea de investigación se originó en el marco de una propuesta colaborativa entre el Laboratorio de Evolución Experimental de Virus del Institut Pasteur de Montevideo y el laboratorio Stern-Ginossar del Weizmann Institute of Science (Israel), pionero en el desarrollo de sistemas virales para la expresión de sgRNAs en el contexto de cribados funcionales utilizando al HCMV como modelo ¹⁷⁰. El proyecto, seleccionado para su financiamiento, tuvo como objetivo integrar la experiencia del grupo israelí en CRISPR *screens* acoplados a infección viral, con la trayectoria de nuestro laboratorio en virología molecular y en el desarrollo de clones infecciosos de MAYV.

El plan experimental contemplaba una primera etapa a desarrollarse en Montevideo, centrada en el diseño y desarrollo de clones infecciosos de MAYV capaces de expresar sgRNAs funcionales, seguida de una segunda fase experimental en el Weizmann Institute orientada a optimizar el sistema y avanzar hacia la implementación de un CRISPR *screen* acoplado a la replicación viral, aprovechando la infraestructura y experiencia del laboratorio anfitrión en este tipo de abordajes.

Sin embargo, debido al contexto geopolítico internacional y al conflicto armado en curso en la región, esta segunda etapa prevista en Israel no pudo concretarse en los

plazos originalmente establecidos. A pesar de ello, el proyecto continuó en Montevideo, donde se desarrollaron las primeras pruebas de concepto destinadas a evaluar la viabilidad técnica de expresar sgRNAs desde el genoma de MAYV, analizar su estabilidad durante la replicación viral y determinar su capacidad para inducir edición génica en células que expresan Cas9 de manera constitutiva.

Hipótesis

Planteamos que es posible modificar el genoma de MAYV para expresar ARN guía (sgRNA) funcionales a partir del promotor subgenómico, de modo que tras la infección en células que expresan Cas9, estas variantes sean capaces de inducir edición génica específica, permitiendo evaluar la funcionalidad de los sgRNA en el contexto de una infección activa. La validación de este sistema establecería a MAYV como una plataforma viral innovadora para edición génica dirigida, con potencial de aplicación en el estudio mecanístico de interacciones virus-hospedero y que podría ser adaptable a otros alfavirus.

Objetivo general

Desarrollar variantes sintéticas de MAYV que expresen ARN guías del sistema CRISPR/Cas9 y evaluar su capacidad para inducir edición génica *in vitro* como prueba de concepto para el estudio de factores del hospedero que modulan la replicación viral.

Objetivos específicos

- 1- Diseñar *in silico* clones infecciosos de MAYV que codifiquen ARN guías (sgRNAs) del sistema CRISPR/Cas9.
- 2- Generar clones infecciosos codificantes para sgRNAs mediante herramientas de biología molecular.
- 3- Evaluar la viabilidad viral y la estabilidad del inserto de sgRNA en el genoma de MAYV.
- 4- Determinar la funcionalidad del sistema evaluando la capacidad de los virus recombinantes para expresar sgRNAs e inducir edición génica específica en células que expresan Cas9.

Materiales y Métodos

Células

La línea celular A549 con expresión constitutiva de Cas9 (*Streptococcus pyogenes* Cas9) fue proporcionada gentilmente por el Dr. Benjamin tenOever (New York University, USA). Se utilizaron además las líneas A549 (carcinoma pulmonar humano, ATCC CCL-185) y Vero (riñón de mono verde africano, ATCC CCL-81) en las condiciones de cultivo descritas en el capítulo 2, sección materiales y métodos.

Generación de clones infecciosos

La generación del clon infeccioso MAYV_sgRNA2_miR se realizó mediante la clonación de un inserto de ADN sintético de 351 pb (gBlock™, Integrated DNA Technologies, USA) que abarca desde el extremo 5' del promotor subgenómico (SGP) hasta el final 3' de la secuencia del mmu-miR-124-2 (chromosome 3 [chr3]: 17,795,418–17,796,022) (**Fig. 3.4**). La clonación se llevó a cabo mediante el kit *In-Fusion® HD EcoDry™ Cloning* (Takara Bio, USA). Para ello, tanto el inserto como el clon infeccioso de MAYV WT (*backbone*) se amplificaron por PCR en tiempo final utilizando cebadores con extremos 5' y 3' con 20 pb de homología, permitiendo la ligación entre el inserto y el vector mediante el kit *In-Fusion® HD EcoDry™ Cloning* (**Fig. 3.4**).

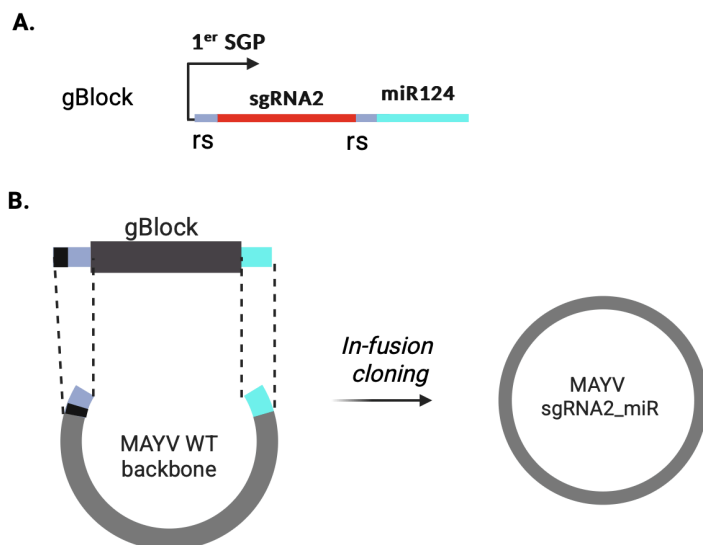


Figura 3.4: Generación de MAYV_sgRNA2_miR por el sistema de clonado *In-Fusion*. **A)** estructura del gBlock conteniendo las secuencias de interés: 1er SGP, 1er promotor subgenómico; rs, sitio de corte de restricción; sgRNA2; sgRNA contra el gen ZC3HAV1 codificante para ZAP; miR124, pri mmu-miR-124-2 **B)** El inserto (gBlock) fue amplificado mediante PCR utilizando cebadores diseñados para incorporar extremos solapantes complementarios a MAYV WT. El backbone de MAYV WT fue linealizado por PCR. Ambos fragmentos fueron ligados mediante la estrategia de clonación sin restricción *In-Fusion*, y el producto resultante fue transformado en bacterias *E. coli* TURBO quimiocompetentes para la obtención del constructo MAYV_sgRNA2_miR.

La PCR del inserto de 351 pb (vol. final 50 μ L) contuvo 2 ng de ADN molde, 25 μ L Q5® High-Fidelity DNA Polymerase Master Mix (NEB, USA), 2.5 μ L (10 μ M) de los cebadores Fragment_for y Fragment_rev y agua libre de nucleasas. La PCR del backbone (vol. final 50 μ L) contuvo 22 ng del clon infeccioso MAYV WT, 25 μ L del Master Mix Platinum™ SuperFi™ PCR (Invitrogen, USA) y 2.5 μ L (10 μ M) de los cebadores Vector_for y Vector_rev (**tabla S3.1**).

Los amplicones obtenidos se analizaron por electroforesis en gel de agarosa al 1 % p/v en TAE 1X, teñidos con SYBR® Safe (Invitrogen, USA), utilizando como referencia el marcador de peso molecular 1 Kb Plus DNA Ladder (Thermo Scientific, USA). Los productos de PCR fueron purificados a partir de geles de agarosa utilizando los kits Zymoclean Gel DNA Recovery para el inserto y Zymoclean Large Fragment DNA Recovery para el vector (Zymo Research, USA).

La reacción de *In-Fusion Cloning* (vol. final 10 μ L) se realizó siguiendo las indicaciones del fabricante, conteniendo 2 ng de inserto y 75 ng del vector, a una relación molar inserto:vector de 1.5:1, en agua libre de nucleasas. El producto de reacción se utilizó para transformar bacterias quimiocompetentes *E. coli* Turbo (NEB®, USA) mediante

choque térmico. Las células se recuperaron en medio SOC y se sembraron en placas de LB-agar suplementadas con ampicilina (100 µg/mL), incubándose a 37 °C O/N.

El clon infeccioso MAYV_sgRNA2_HDV se generó reemplazando la secuencia del mmu-miR-124-2 en el constructo MAYV_sgRNA2_miR por la secuencia codificante de la ribozima del virus de la hepatitis delta (HDV). Esto se realizó mediante la clonación direccional de un producto de PCR de fusión (2746 pb), utilizando las enzimas de restricción *Ascl* y *BsiWI-HF*® (NEB) (**Fig. 3.5**). La PCR de fusión se realizó a partir de dos productos solapantes (PCR 1 de 103 pb y PCR 2 de 2658 pb). Para la PCR 1 se emplearon los cebadores For_fusionHDV y Rev_fusionHDV, y para la PCR 2, los cebadores For_restofusion y Rev_restofusion (**tabla S3.1**, información suplementaria). En cada PCR (vol. final 25 µL), se utilizó la enzima Q5® High-Fidelity DNA Polymerase Master Mix (New England Biolabs, MA, USA), 1,25 µL de los cebadores (10 µM), 5 ng de ADN (clon infeccioso MAYV WT) y agua libre de nucleasas. Los amplicones obtenidos fueron purificados desde el gel de agarosa con el kit PureLink™ Quick Gel Extraction Kit (ThermoFisher, USA). La PCR de fusión (vol. final 25 µL) se realizó con la enzima Q5® High-Fidelity DNA Polymerase Master Mix, 1,25 µL del cebador For_fusionHDV (10 µM), 1,25 µL del cebador reverso Rev_restofusion (10 µM), 5 ng de la PCR 1 y 133 ng de la PCR 2. Los amplicones fueron purificados con el PureLink™ Quick Gel Extraction Kit (ThermoFisher, USA) y secuenciados por ONT.

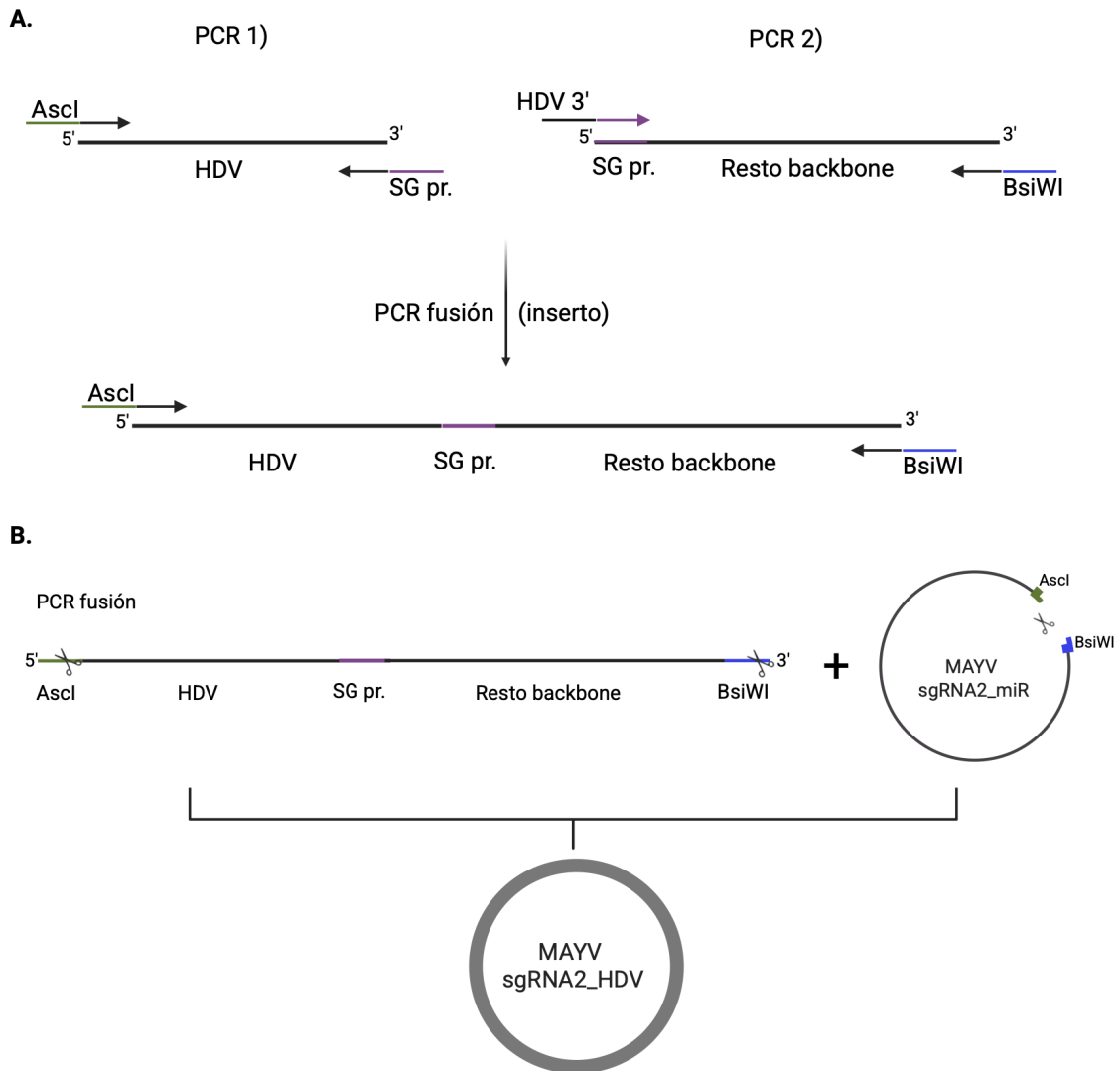


Figura 3.5: Generación de MAYV sgRNA2_HDV por PCR de fusión y clonación direccional en el *backbone* de MAYV sgRNA2_miR. A) Representación esquemática de las PCRs realizadas para generar el inserto a clonar (PCR fusión). **B)** La PCR de fusión mostrada en A. y el constructo MAYV sgRNA2_miR fueron digeridos con las enzimas de restricción AsclI y BsiWI y posteriormente ligados con una enzima T4 ligasa para la obtención del constructo MAYV sgRNA2_HDV.

A continuación, el clon infeccioso MAYV_sgRNA2_miR y el amplicón de 2746 pb fueron digeridos con las enzimas AsclI y BsiWI-HF® por separado en 50 µl finales de reacción por 1 hora a 37°C. Los mismos fueron purificados y concentrados con el Kit DNA Clean & Concentrator-5 (Zymo Research) siguiendo las instrucciones del fabricante, y posteriormente ligados con la enzima T4 DNA ligasa (Invitrogen) a 16 °C durante toda la noche (O/N), seguido de una inactivación de la enzima por 15 minutos a 65 °C. Bacterias NEB® Turbo Competent *E. coli* (New England Biolabs, USA) se transformaron con el producto de ligación mediante choque térmico siguiendo las instrucciones del fabricante. Las bacterias se recuperaron en medio SOC y se

sembraron en placas de LB-agar suplementadas con ampicilina (100 µg/mL), incubándose a 37 °C O/N.

Secuenciación de clones infecciosos

Se seleccionaron al menos cuatro colonias por placa de LB agar conteniendo las bacterias transformadas con los constructos MAYV_sgRNA2_miR o MAYV_sgRNA2_HDV, y se procedió con la extracción de los plásmidos utilizando el kit Zippy Plasmid Miniprep (Zymo Research, USA) siguiendo las instrucciones del fabricante. La secuencia consenso de ambos constructos fue verificada mediante secuenciación con Oxford Nanopore a través del servicio CiveTerus ofrecido por el CiVE en el Institut Pasteur de Montevideo.

Generación de stocks virales

Con el fin de generar la progenie viral inicial (p0) a partir de los clones infecciosos MAYV WT, MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV, los mismos fueron transfectados en células Vero en placas de 6 pocillos con Lipofectamine™ 3000 Transfection Reagent (Invitrogen, USA), siguiendo las especificaciones del fabricante. 48 horas post-transfección, se continuó con el protocolo descrito bajo el apartado “Generación de stocks virales” del capítulo 2, sección materiales y métodos.

Cinética de replicación viral

Para evaluar la cinética de replicación viral de las variantes MAYV sgRNA2_miR y MAYV sgRNA2_HDV en células A549 y A549-Cas9, las mismas fueron infectadas con estos virus y el virus MAYV WT a modo de control. En el caso de MAYV sgRNA2_miR, las infecciones se realizaron a multiplicidades de infección (MOI) de 0,0001; 0,1; 1 y 3, mientras que para MAYV sgRNA2_HDV se utilizaron MOIs de 0,1; 1 y 3. A las 8, 12, 24, 48 y 72 HPI se recogieron 150 µL de sobrenadante del medio de cultivo para estudiar la cinética de replicación y realizar la extracción de ARN viral. A las 24 y 48 HPI se recogieron también los pellets celulares para analizar la expresión de la proteína ZAP por WB y para la extracción de ARN viral y ADN celular. El ARN extraído de ambas fracciones se utilizó para analizar la estabilidad del inserto conteniendo el sgRNA en los genomas virales de los virus sintéticos. El ADN celular se utilizó para analizar la presencia de *indels* en el sitio blanco del sgRNA2.

Evaluación de la estabilidad genómica del inserto en los virus sintéticos MAYV sgRNA2_miR y MAYV sgRNA2_HDV

Para evaluar la estabilidad genómica del inserto en los virus MAYV sgRNA2_miR y MAYV sgRNA2_HDV, se extrajo ARN total a partir de 100 μ L de medio de cultivo celular (fracción extracelular), como a partir de células A549-Cas9 (fracción intracelular), a distintas HPI con el kit QIAamp Viral RNA Mini (Qiagen, USA), siguiendo las indicaciones del fabricante. El ARN extraído (0,3 μ g) fue retrotranscrito con el master mix LunaScript™ RT SuperMix (2 μ L, NEB, USA) y agua libre de nucleasas (vol. final 10 μ L). El ADNc fue amplificado por PCR (vol. final 25 μ L) con los cebadores For_checkinsKO820 y Rev_checkinsKO820 (1,25 μ L, 10 μ M) detallados en la **tabla S3.1** (información suplementaria). Se utilizó el master mix Q5® High-Fidelity DNA Polymerase (12,5 μ L, NEB, USA). Los amplicones se analizaron mediante electroforesis en geles de agarosa al 1%, junto con amplicones obtenidos a partir del clon infeccioso MAYV WT utilizando los mismos cebadores, con el fin de comparar los tamaños de los fragmentos obtenidos. De acuerdo con el diseño de los constructos, se espera un amplicón de 820 pb para MAYV_sgRNA2_miR, de 697 pb para MAYV_sgRNA2_HDV y de 468 pb para el clon infeccioso MAYV WT.

De manera complementaria, los amplicones fueron analizados mediante secuenciación masiva en la plataforma Oxford Nanopore Technologies (ONT).

En el caso de MAYV_sgRNA2_miR, se secuenciaron los amplicones de 820 pb obtenidos a MOI 1 y 3, a las 24 y 48 HPI. Para MAYV_sgRNA2_HDV, se analizaron los amplicones de 697 pb correspondientes a MOI 0,1; 1 y 3, en las mismas HPI.

Los amplicones fueron purificados con perlas magnéticas AMPure XP (0,6X, Beckman Coulter, USA) y cuantificados mediante un método fluorimétrico con el kit Qubit dsDNA High Sensitivity (ThermoFisher, USA). Las bibliotecas de secuenciación se prepararon utilizando el kit Rapid Barcoding SQK-RBK110.96 (ONT), siguiendo un protocolo optimizado basado en Freed y Silander ¹⁵⁷. Para cada preparación se emplearon entre 50 y 100 ng de ADN. Las bibliotecas se cargaron (20 fmol) en una flow cell FLO-FLG001 R9.4.1 y se secuenciaron en un equipo GridION X5 durante aproximadamente 12 horas. El basecalling y la demultiplexación se realizaron con Guppy v6.3.9 en modo de super alta precisión, y los genomas consenso se generaron mediante el flujo de trabajo Epi2Me utilizando los parámetros predeterminados.

Cuantificación relativa del micro-RNA 124 (miR-124) maduro mediante stem-loop RT-qPCR

La extracción de ARN de células A549-Cas9 infectadas con el virus MAYV sgRNA2_miR (MOI=3) a las 24 y 48 HPI, así como de células no infectadas (mock), se realizó con el kit Zymo RNA Clean & Concentrator (Zymo Research, USA). Este kit permite separar los fragmentos de ARN menores a 200 nt de los mayores a 200 nt, recuperando miRNA maduros de 20 nt en la fracción de menor tamaño. Se cuantificó el ARN total de todas las muestras de manera estándar mediante absorbancia a 260 nm por NanoDrop. La síntesis de ADNc se realizó a partir de 0,1 µg de ARN total con la enzima SuperScript II Reverse Transcriptase (Invitrogen, USA) siguiendo las instrucciones del fabricante. Se emplearon cebadores específicos de stem-loop para el miR124 y el control endógeno miR423-5p. La cuantificación relativa del mir124 se llevó a cabo mediante qPCR utilizando el kit TaqMan Fast Virus 1-Step Master Mix (Invitrogen, USA) con los cebadores y sondas del TaqMan Assay 5X siguiendo las instrucciones del fabricante. Los datos obtenidos fueron analizados utilizando el método comparativo del Cq ($\Delta\Delta Cq$), donde se normalizó la expresión del miRNA miR124 con el control endógeno miR423-5p.

Eficiencia de edición génica en células infectadas con MAYV sgRNA2_miR y MAYV sgRNA2_HDV

El análisis de la eficiencia de edición génica en el gen de ZAP se realizó en células A549-Cas9 infectadas con los virus MAYV_sgRNA2_miR (MOI=0,1 y MOI=1) o MAYV_sgRNA2_HDV (MOI=1 y MOI=3). A partir de ARN extraído mediante el kit RNAeasy Mini (QIAGEN), se llevó a cabo una PCR utilizando los cebadores For_sgRNA2 y Rev_sgRNA2, detallada en materiales y métodos del capítulo 2, obteniéndose un amplicón de 566 pb flanqueante a la región blanco del sgRNA2 en el gen codificante para ZAP. Los amplicones obtenidos se purificaron con el kit PureLink PCR Purification (ThermoFisher, USA) y se cuantificaron por nanodrop.

A continuación, se emplearon dos tipos de análisis para cuantificar la eficiencia de edición génica:

1) *Secuenciación por Sanger y análisis mediante ICE*

Para el análisis inicial de indels, los amplicones de 566 pb se secuenciaron por Sanger (Macrogen Inc., Seúl, Corea del Sur), empleando el cebador For_sgRNA2. Los cromatogramas obtenidos fueron analizados utilizando la herramienta en línea ICE (Synthego, USA).

2) *Secuenciación masiva por Illumina y análisis mediante CRISPResso2*

Para aumentar la sensibilidad y resolución en la detección de eventos de edición, los amplicones de 566 pb previamente purificados fueron utilizados para la generación de librerías compatibles con secuenciación Illumina. Para ello, los amplicones fueron fragmentados para la construcción de librerías y posteriormente secuenciados en la plataforma NovaSeq X Plus en modalidad *paired-end* 150 pb (PE150). Tanto la fragmentación como la construcción de librerías y secuenciación fueron llevadas a cabo por la empresa Haplox GeneTech Limited (Hong Kong).

La cuantificación de edición se realizó con el software en línea CRISPResso2¹⁷⁹ utilizando una secuencia de referencia definida como un fragmento de 150 pb interno al amplicón de 566 pb, que contiene el sitio blanco del sgRNA2 (detallado en materiales y métodos del capítulo 2). El fragmento de referencia de 150 pb empleado para el alineamiento fue: “5’ – gttataagggagaggggtcggcagcagatttgaaccagcagccaccgtgtcaagactccacatctgtgaccacttcaccgaggggaactgctggtttccaactgcctccgggtccataacctgatggacagaaaggtgctggcca –3”.

Dado que las librerías fueron generadas a partir de un amplicón de 566 pb que sufrió un proceso de fragmentación para la construcción de librerías de 150 pb, únicamente las lecturas que abarcan completamente la región de referencia de 150 pb pudieron alinearse al amplicón definido en CRISPResso2.

La eficiencia de edición se calculó como el porcentaje de lecturas alineadas clasificadas como modificadas dentro de la ventana de cuantificación. Además, las métricas completas de análisis de control de calidad de la secuenciación (reporte QC) para todas las muestras analizadas se presentan en la **tabla S3.3**.

Western blot para ZAP y Cas9 en células A549-Cas9

Se realizó una preparación de lisados celulares para electroforesis desnaturizante SDS-PAGE y western blotting para el análisis de la expresión de ZAP en células A549 y A549-Cas9 infectadas con MAYV sgRNA2_miR o MAYV sgRNA2_HDV según el procedimiento detallado en el capítulo 2, sección materiales y métodos. Se utilizó el anticuerpo primario anti-spCas9 a una dilución de 1:2000 (7A9-3A3, Abcam, USA), como control en células A549-Cas9.

Resultados

Diseño in silico de clones infecciosos codificantes para ARN guías.

Con el objetivo de desarrollar clones infecciosos de MAYV que codifiquen un sgRNA del sistema CRISPR/Cas9, se realizó el diseño in silico utilizando SnapGene®, basado en la recodificación del clon MAYV WT. Se generaron dos construcciones, MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV. La **figura 3.6A** muestra los elementos incorporados y su ubicación en el genoma, mientras que la **figura 3.6B** esquematiza el procesamiento y la expresión del sgRNA durante la replicación viral.

Estos virus serán evaluados en células A549-Cas9 (que expresan constitutivamente la nucleasa Cas9 de *Streptococcus pyogenes*), donde Cas9, guiada por el sgRNA producido desde el genoma viral, se espera que genere *indels* en su sitio blanco.

Se seleccionó el sgRNA2 dirigido contra ZAP, previamente validado en células A549 (ver capítulo 2), donde mostró alta eficiencia de edición (~80%). La inserción se realizó downstream de los genes no estructurales y del primer promotor subgenómico (1er SGP), con el objetivo de minimizar interferencias con procesos esenciales como la replicación y el ensamblaje, en concordancia con reportes previos en alfavirus.^{177,180}

Con el fin de facilitar la clonación de otros sgRNAs o la generación de una biblioteca de clones infecciosos, se añadieron sitios de restricción PacI (5') y AscI (3') flanqueando el sgRNA2. Para asegurar el procesamiento adecuado del sgRNA y su interacción funcional con Cas9, se incorporaron dos estrategias alternativas. Por un lado, utilizó un pri-miRNA mmu-miR-124-2, generando el clon MAYV_sgRNA2_miR. Por otro lado, se utilizó una ribozima del virus de hepatitis delta (HDV), dando lugar al clon MAYV_sgRNA2_HDV. Ambas estrategias han sido utilizadas previamente en sistemas virales (alfavirus y otros virus ARN) como herramientas para el procesamiento preciso del ARN.^{171,178,181} El miR-124 es específico de células neuronales, por lo que no se espera su expresión en A549-Cas9. Además, el pri-miR-124-2 ha sido incorporado previamente en el genoma de SINV, un alfavirus similar a MAYV, demostrando su procesamiento citoplasmático y su viabilidad en alfavirus recombinantes.¹⁷⁸

Finalmente, se incorporó un segundo promotor subgenómico duplicado (2do SGP), downstream del cassette que contiene el sgRNA, seguido de los genes estructurales, para permitir la expresión de un ARN subgenómico intacto necesario para la producción de las proteínas estructurales (**Fig.3.6**).

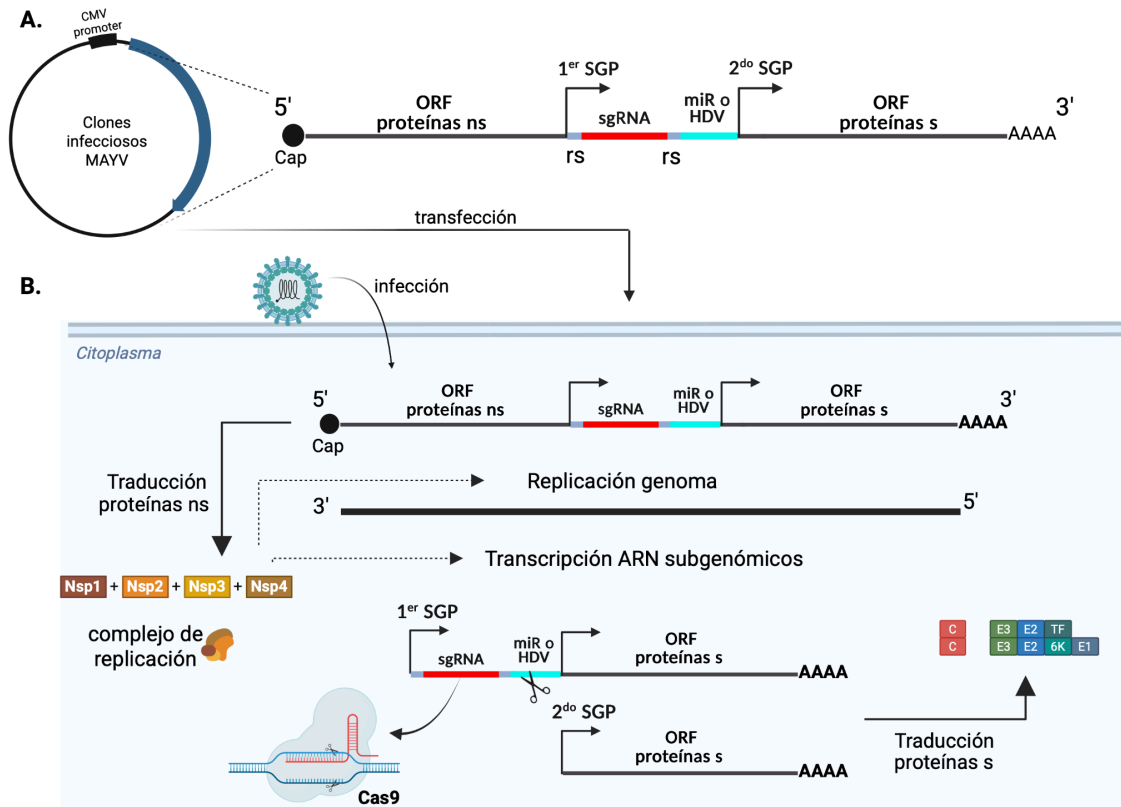


Figura 3.6: Representación esquemática del diseño *in silico* de clones infecciosos MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV y procesamiento intracelular hipotético. Se recodificó el genoma de MAYV WT para generar dos clones infecciosos: MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV, para expresar un ARN guía en células infectadas A549-Cas9. CMV, promotor citomegalovirus; ORF ns y ORF s, marco abierto de lectura para proteínas no estructurales y estructurales respectivamente; rs, sitio de corte con enzimas de restricción; sgRNA, single-guided RNA; miR, ARN primario largo; HDV, secuencia de ribozima del virus de hepatitis delta; SGP, promotor subgenómico.

Generación de clones infecciosos de MAYV codificantes para un sgRNA contra el gen *ZC3HAV1*.

Se generaron los clones infecciosos recombinantes MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV mediante técnicas de biología molecular. Para MAYV_sgRNA2_miR, se amplificó un fragmento sintético (gBlock™) junto con el backbone del clon MAYV WT, los cuales fueron ensamblados mediante In-Fusion. La optimización de las PCRs y el análisis de los amplicones se muestran en la **figura S3.1**. Tras la transformación en *E. coli*, se seleccionaron colonias, se realizó la extracción de plásmidos (miniprep) y se confirmó la correcta construcción y ausencia de mutaciones mediante secuenciación por ONT.

A partir de este constructo, se generó MAYV_sgRNA2_HDV mediante el reemplazo del pri-miRNA mmu-miR-124-2 por la ribozima HDV, utilizando una estrategia de PCR de fusión seguida de digestión y ligación. La optimización de la PCR de fusión y el

análisis en gel se muestran en la **figura S3.2**. El producto fue transformado en bacterias, y los plásmidos obtenidos fueron validados por secuenciación, confirmando la correcta construcción del clon.

Obtención de stock virales (P2) a partir de los clones infecciosos desarrollados.

Para producir stocks virales de MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV, se transfectaron los clones infecciosos en células Vero, junto con el clon WT como control (**Fig. 3.7A**). A las 48 h post-transfección se observó efecto citopático en los tres virus (**Fig. 3.7B**), confirmando la producción de partículas infectivas y la viabilidad de los clones.

El sobrenadante obtenido (P0) se utilizó para realizar dos pasajes en células Vero, generando la progenie P2 (stock viral). Los títulos virales, determinados por ensayo de placa, fueron comparables entre ambos virus, en el rango de 10^7 – 10^8 PFU/mL.

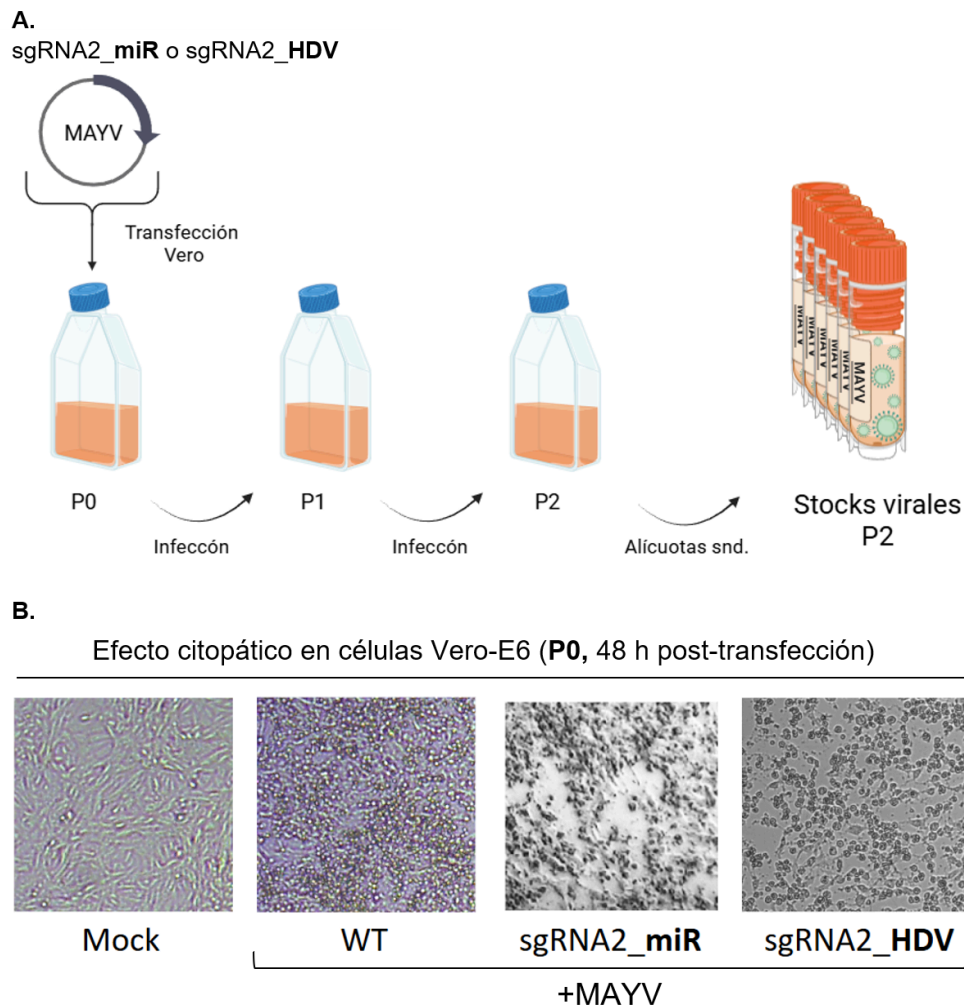


Figura 3.7: Efecto citopático en células Vero-E6 transfectadas con distintos clones infecciosos de MAYV. Las células fueron transfectadas con los clones infecciosos de MAYV (clon WT, MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV) utilizando el reactivo Lipofectamine™ 3000 (Invitrogen, USA), según las indicaciones del fabricante. A las 48 horas post-transfección, se evidenció un efecto citopático característico, donde se recolectó la progenie viral correspondiente al pasaje 0 (P0). Como control negativo (Mock), se incluyeron células tratadas únicamente con el reactivo de transfección, sin adición de ADN.

Para confirmar la presencia del inserto en MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV en las progenies P0 y P2, se realizó una RT-PCR utilizando cebadores flanqueantes al inserto (**Fig. 3.8A**). Esta estrategia permitió diferenciar los genomas según el tamaño del amplicón: 468 pb para MAYV WT, 820 pb para MAYV_sgRNA2_miR y 697 pb para MAYV_sgRNA2_HDV (**Fig. 3.8B**). En conjunto, estos resultados, junto con los de la **figura 3.7**, confirman la viabilidad de los virus sintéticos y sugieren estabilidad del inserto durante la producción de stocks virales.

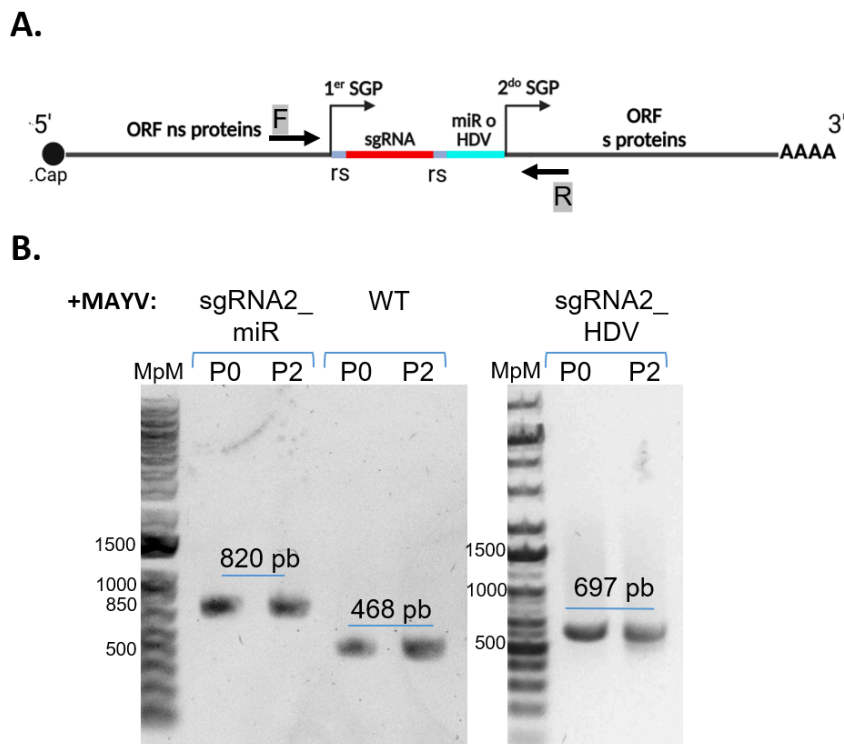


Figura 3.8. Análisis de la estabilidad del inserto en MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV durante pasajes seriados (P0–P2) mediante RT-PCR. A) Esquema de la región amplificada. Tamaños esperados: 468 pb (WT), 820 pb (MAYV_sgRNA2_miR) y 697 pb (MAYV_sgRNA2_HDV). **B)** Gel de agarosa al 1% de los productos de RT-PCR a partir de ARN viral de P0 y P2. MpM: marcador Invitrogen 1 kb Plus DNA Ladder. La línea azul indica los tamaños esperados.

Caracterización funcional de los virus desarrollados para dirigir la edición génica mediante CRISPR/Cas9 en células infectadas.

Para evaluar la funcionalidad del sgRNA2 codificado en el genoma viral, se realizaron tres tipos de análisis presentados a continuación:

1- Análisis de cinética de replicación en células A549-Cas9.

Se espera que la infección con MAYV_sgRNA2_miR o MAYV_sgRNA2_HDV genere mutaciones en ZAP y reduzca su expresión, favoreciendo la replicación viral. Para evaluar este efecto, se comparó la cinética de replicación de los virus desarrollados con MAYV WT en células A549 y A549-Cas9.

MAYV sgRNA2_miR:

Para MAYV_sgRNA2_miR, a MOI bajas (1×10^{-4} y 0.1), se observó una atenuación significativa respecto al WT (Fig. 3.9). A MOI = 1×10^{-4} , esta atenuación fue mayor en presencia de Cas9, mientras que a MOI = 0.1, el virus alcanzó títulos comparables al

WT a las 48 HPI en ambas condiciones, indicando que este efecto no se asocia al silenciamiento de ZAP.

A MOI altas (1 y 3; **Fig. 3.10**), MAYV_sgRNA2_miR continuó atenuado en células A549. Sin embargo, en A549-Cas9 su replicación fue comparable o superior al WT, con mayores títulos a las 48 HPI.

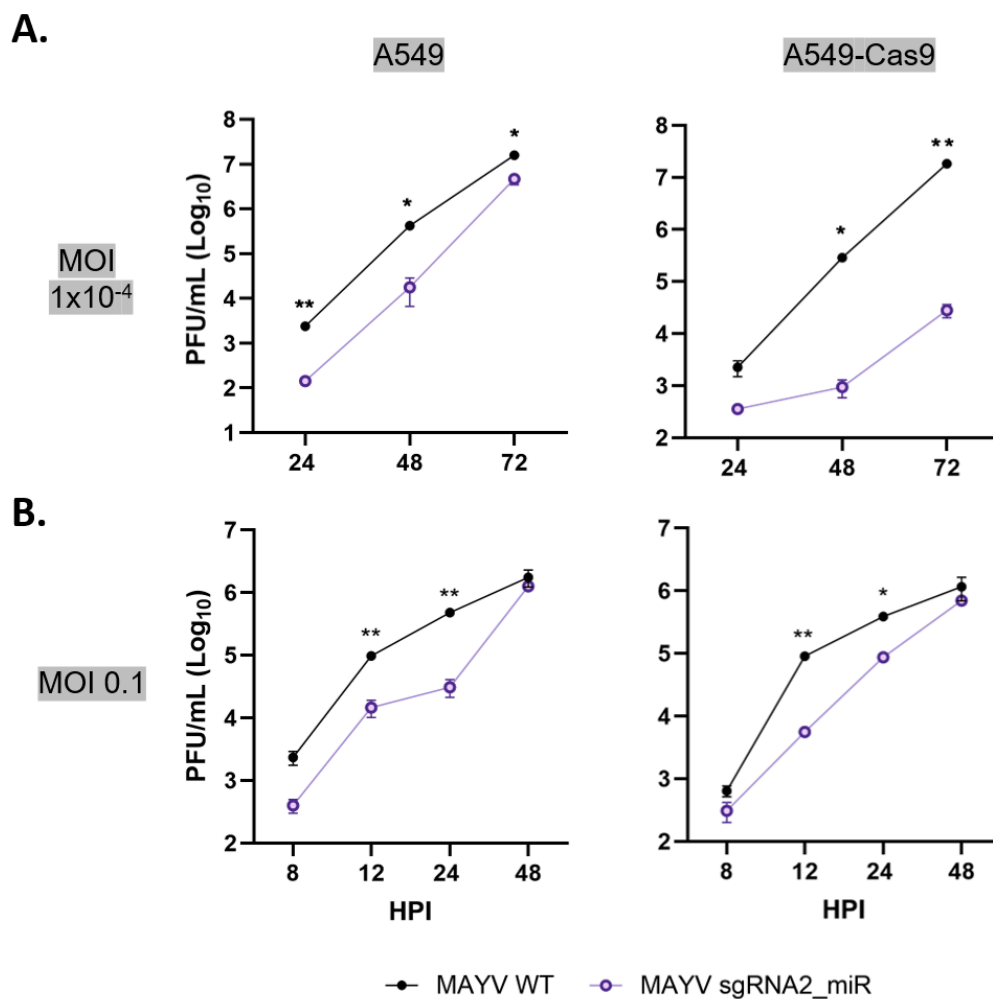


Figura 3.9: Cinética de replicación viral a baja MOI muestra un fenotipo atenuado de MAYV_sgRNA2_miR en células A549 y A549-Cas9 respecto a MAYV WT. Células A549 y A549-Cas9 fueron infectadas con MAYV WT o MAYV_sgRNA2_miR a MOI = 1×10^{-4} (A) y 0.01 (B). Los títulos virales (PFU/mL) se determinaron entre 8–72 HPI mediante plaque assay en células Vero. Datos expresados como media \pm error estándar (n=3). *P < 0.05, **P < 0.01, ***P < 0.001 (two-way ANOVA, post-test Šidák).

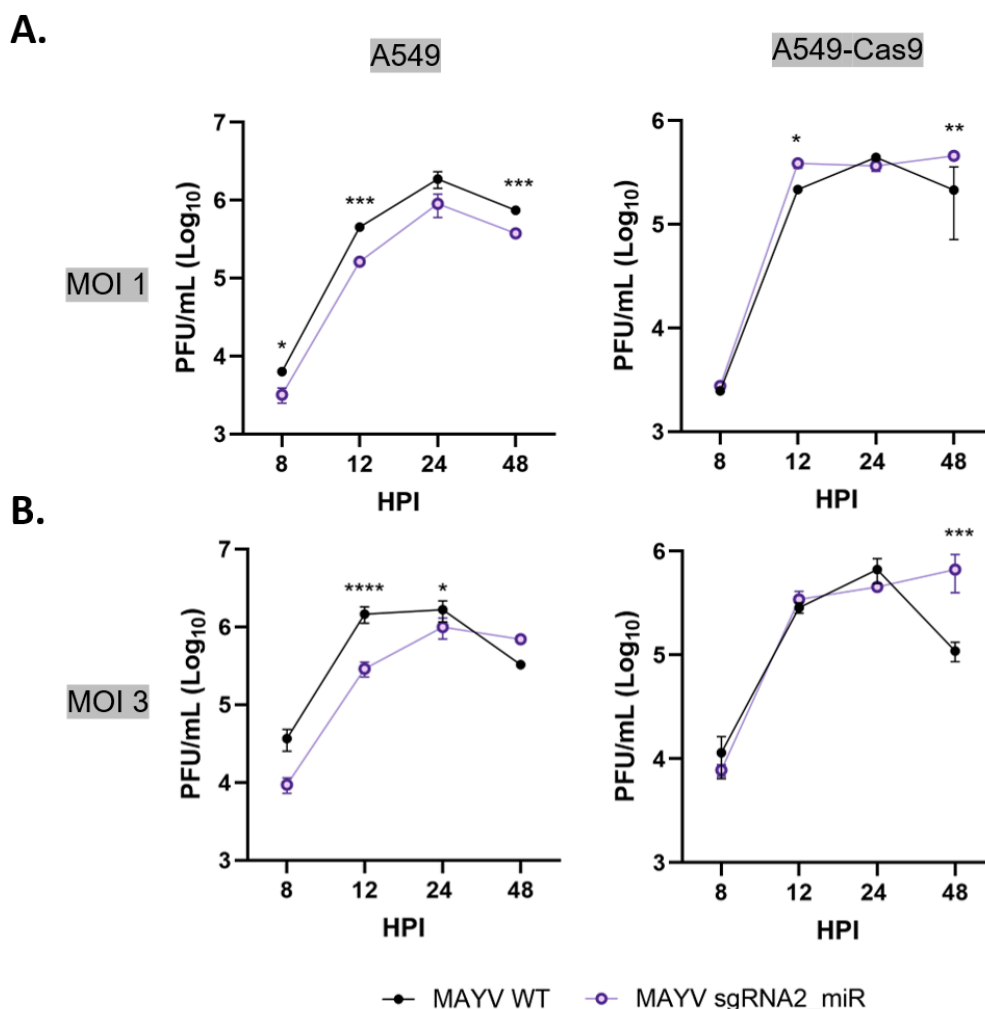


Figura 3.10: Cinética de replicación viral de MAYV_sgRNA2_miR a alta MOI. Células A549 y A549-Cas9 fueron infectadas con MAYV WT o MAYV_sgRNA2_miR a MOI = 1 (A) y 3 (B). Los títulos virales (PFU/mL) se determinaron entre 8–72 HPI. Datos expresados como media \pm error estándar ($n=3$). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$ (two-way ANOVA, post-test Šidák).

Como control, se evaluó la estabilidad del inserto mediante RT-PCR (según **Fig. 3.8A**). A MOI = 1×10^{-4} (**Fig. 3.11A**), el inserto se mantuvo estable en la fracción extracelular hasta 72 HPI, mientras que en la fracción intracelular se detectaron amplicones heterogéneos de menor tamaño, sugiriendo genomas defectivos, lo que podría explicar la atenuación observada en la **figura 3.9A**.

A MOI = 0.1, 1 y 3 (**Fig. 3.11B**), el inserto se mantuvo mayoritariamente estable en la fracción extracelular, aunque a 48 HPI se detectó una banda de menor tamaño compatible con pérdida parcial del inserto. Este patrón también se observó en la fracción intracelular desde 24 HPI.

MAYV sgRNA2 miR

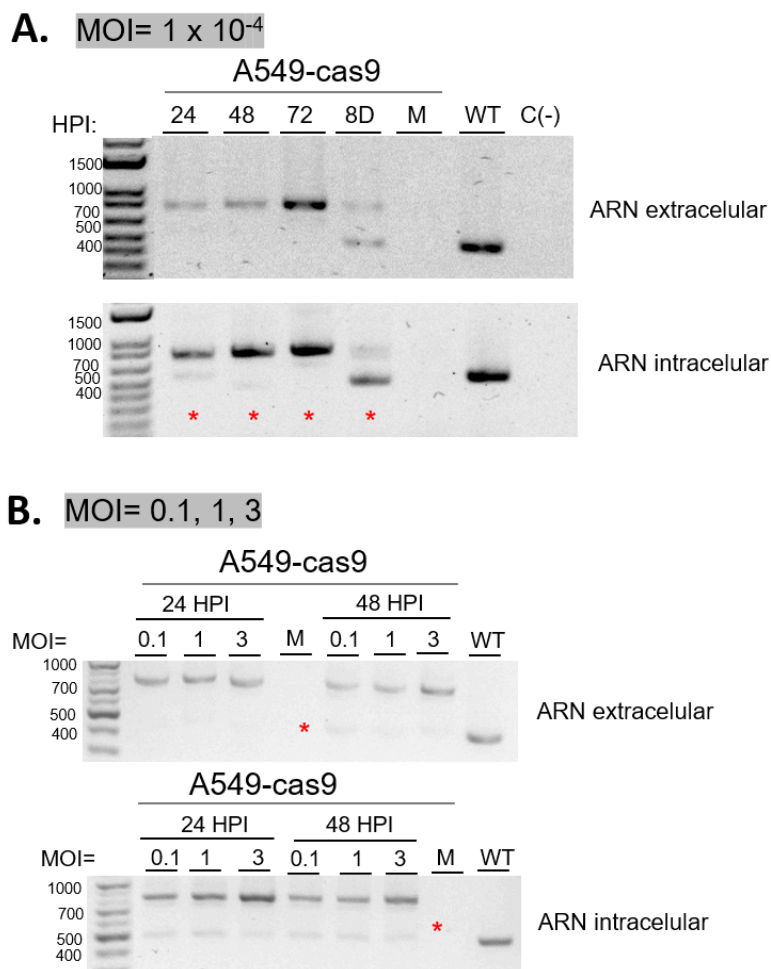


Figura 3.11: Estabilidad del inserto en MAYV_sgRNA2_miR a lo largo del tiempo.

Células A549 y A549-Cas9 fueron infectadas a MOI = 1×10⁻⁴ (**A**) y 0.1, 1 y 3 (**B**). Se extrajo ARN de fracciones extracelular e intracelular a distintos tiempos postinfección (8D: 8 días) y se analizaron los amplicones por RT-PCR en gel de agarosa al 1%. M: mock; WT: MAYV WT. Los asteriscos rojos indican amplicones de menor tamaño.

Los amplicones correspondientes a MOI = 3 (24 y 48 HPI) fueron analizados por secuenciación masiva (ONT) (**Fig. 3.12**). En la fracción intracelular se observó una disminución de cobertura hacia el extremo 3' y deleciones en la región del miRNA, consistentes con los amplicones truncados (**Fig. 3.11B**). En contraste, la fracción extracelular mostró perfiles homogéneos sin deleciones, indicando predominio de genomas completos, aunque con menor profundidad a 48 HPI. Los valores de cobertura se detallan en la **tabla S3.2** (información suplementaria).

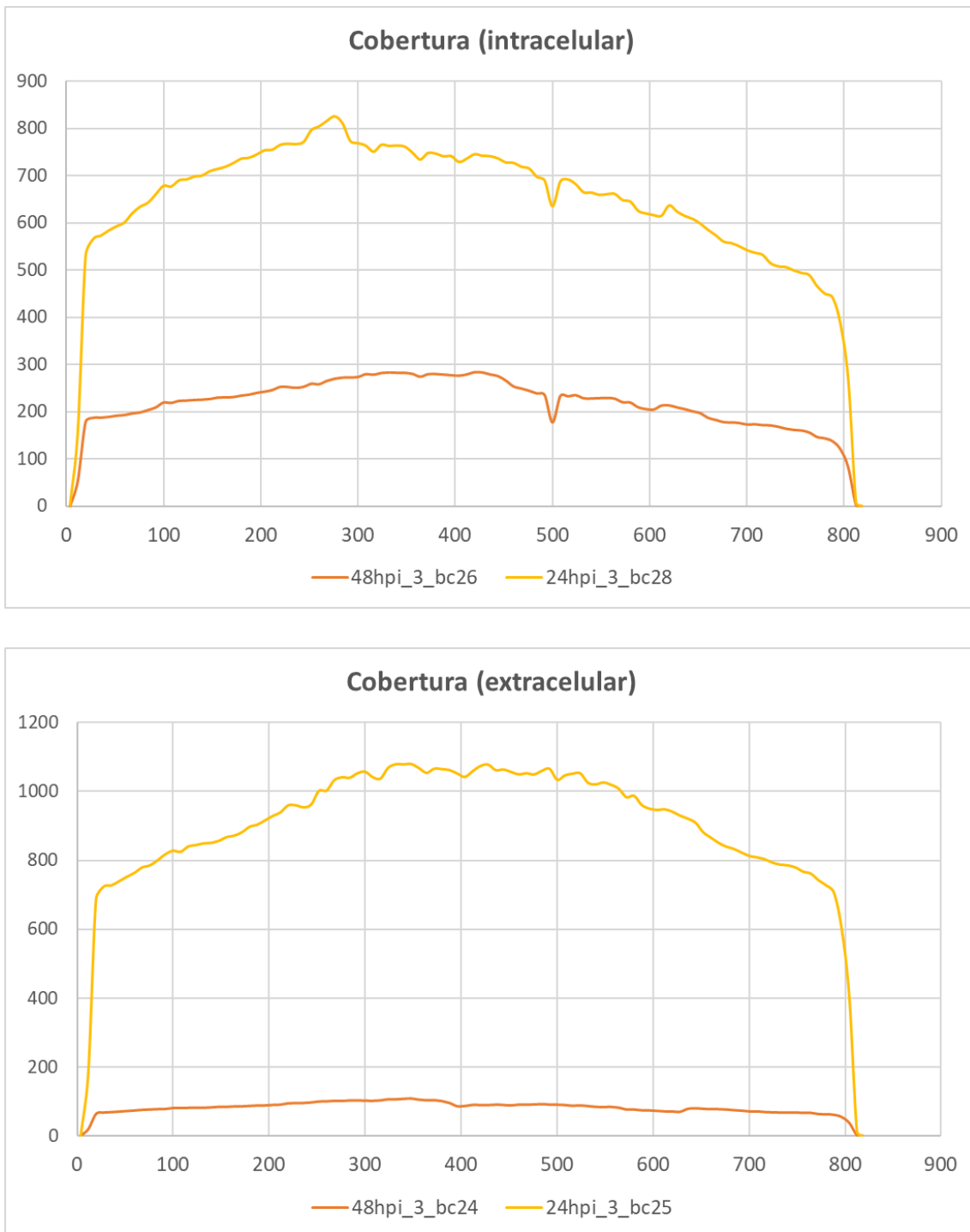


Figura 3.12. Perfiles de cobertura de amplicones de MAYV_sgRNA2_miR obtenidos por secuenciación ONT. El panel superior corresponde a ARN intracelular y el inferior a ARN extracelular. Se analizaron muestras a MOI = 3, a 24 HPI (24hpi_3_bc28 y bc25) y 48 HPI (48hpi_3_bc26 y bc24). Los datos de cobertura se detallan en la tabla S3.2.

MAYV sgRNA2 HDV:

La cinética de replicación para MAYV_sgRNA2_HDV se evaluó en células A549 y A549-Cas9 a MOI 0.1, 1 y 3. Este virus mostró una atenuación significativa respecto al WT entre 12 y 24 HPI en ambas líneas celulares (**Fig. 3.13**), independientemente de Cas9. A las 48 HPI, ambos virus alcanzaron títulos comparables, lo que sugiere que la recuperación no se asocia al silenciamiento de ZAP.

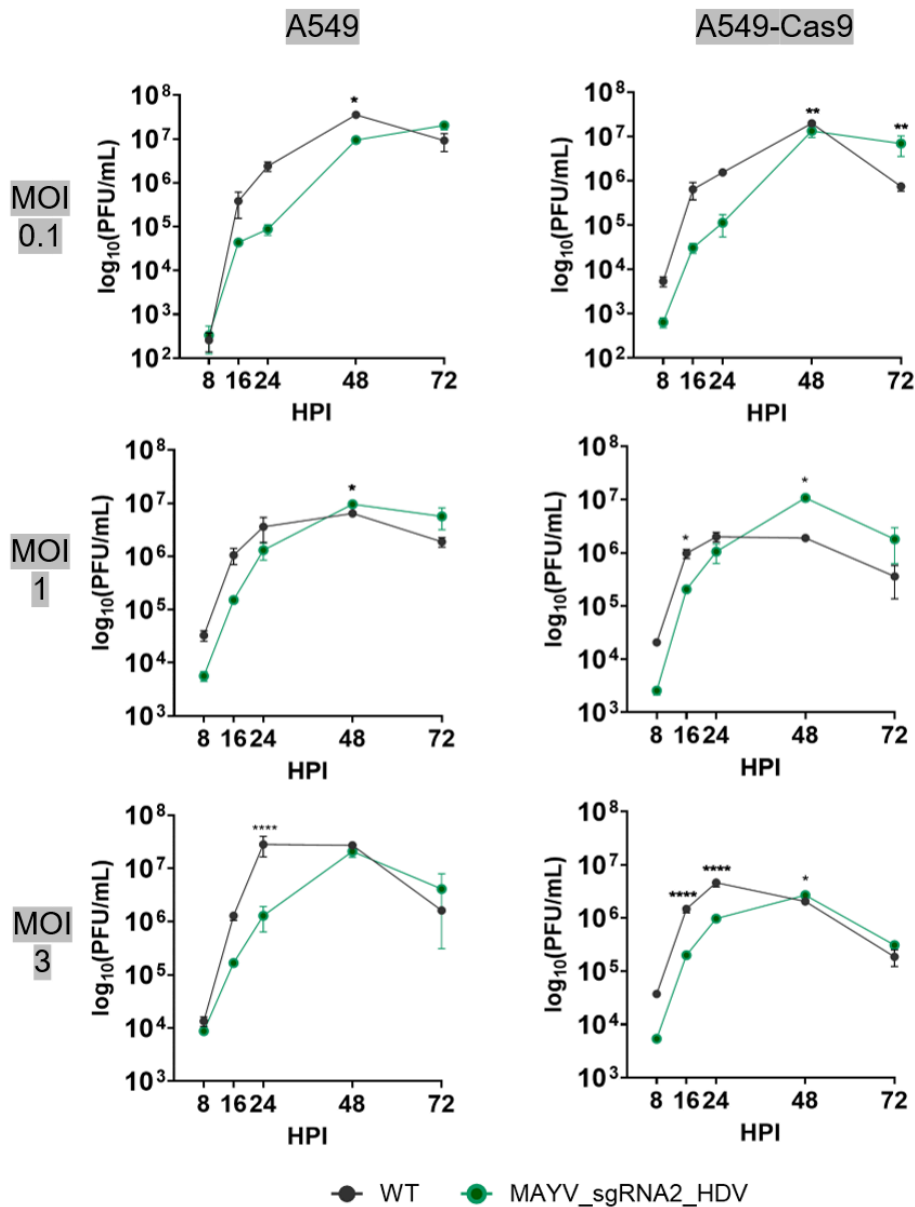


Figura 3.13: Cinética de replicación de MAYV_sgRNA2_HDV en células A549 y A549-Cas9. Células A549 y A549-Cas9 fueron infectadas con MAYV WT o MAYV_sgRNA2_HDV a MOI = 0.1, 1 y 3. Los títulos virales (PFU/mL) se determinaron entre 8–72 HPI mediante plaque assay en células Vero-E6. Datos expresados como media ± error estándar (n=3). *P < 0.05, **P < 0.01, ***P < 0.001 (two-way ANOVA, post-test Šídák).

La estabilidad del inserto se evaluó mediante RT-PCR (**Fig. 3.14**), siguiendo el mismo enfoque utilizado previamente para MAYV_sgRNA2_miR (mostrado en las **Figs. 3.11 y 3.12**). En la fracción extracelular se observó una banda predominante de menor tamaño al esperado (697 pb), mientras que en la fracción intracelular se detectaron dos bandas cercanas en tamaño, posiblemente correspondientes a genomas con y sin el inserto, aunque la sensibilidad de este análisis no permite confirmarlo.

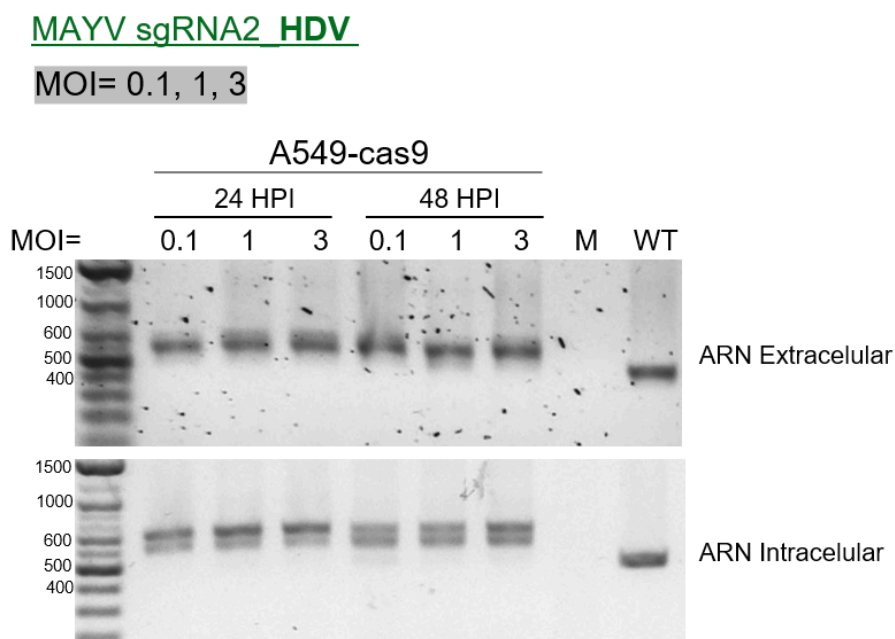


Figura 3.14: Estabilidad del inserto en MAYV_sgRNA2_HDV. Células A549 y A549-Cas9 fueron infectadas a MOI = 0.1, 1 y 3. Se extrajo ARN de fracciones extracelular e intracelular a distintos tiempos postinfección y se analizaron los amplicones por RT-PCR en gel de agarosa al 1%. M: mock; WT: MAYV WT.

Posteriormente los amplicones fueron analizados por secuenciación masiva ONT (**Fig. 3.15**). En la fracción intracelular solo se obtuvieron lecturas para las MOI 1 y 3 (24 HPI), mientras que en la fracción extracelular se obtuvieron lecturas en todas las condiciones excepto para MOI 3 (48 HPI). Los perfiles de cobertura mostraron una disminución en la región del inserto en la mayoría de las condiciones, indicando deleciones frecuentes. Solo en MOI 3 a 24 HPI (24hpi_3_bc30) se detectaron lecturas con cobertura en dicha región, aunque de baja abundancia, sugiriendo que sólo una fracción minoritaria mantiene el inserto intacto.

Como control, se analizaron los amplicones de las progenies P0 y P2. Mientras que en P0 se observó cobertura completa, en P2 se detectó una caída en la región del inserto, aunque aproximadamente el 70% de las lecturas conservaron dicha región, indicando

una población mixta (**Fig. 3.15**). En conjunto, estos resultados evidencian una baja estabilidad del inserto en MAYV_sgRNA2_HDV.

Cabe señalar que la profundidad de secuenciación fue variable entre los distintos amplicones, lo cual puede contribuir a las diferencias observadas en los perfiles de cobertura. En la **tabla S3.2** se detalla la cantidad de lecturas por posición del amplicón utilizados para la generación de los gráficos de cobertura.

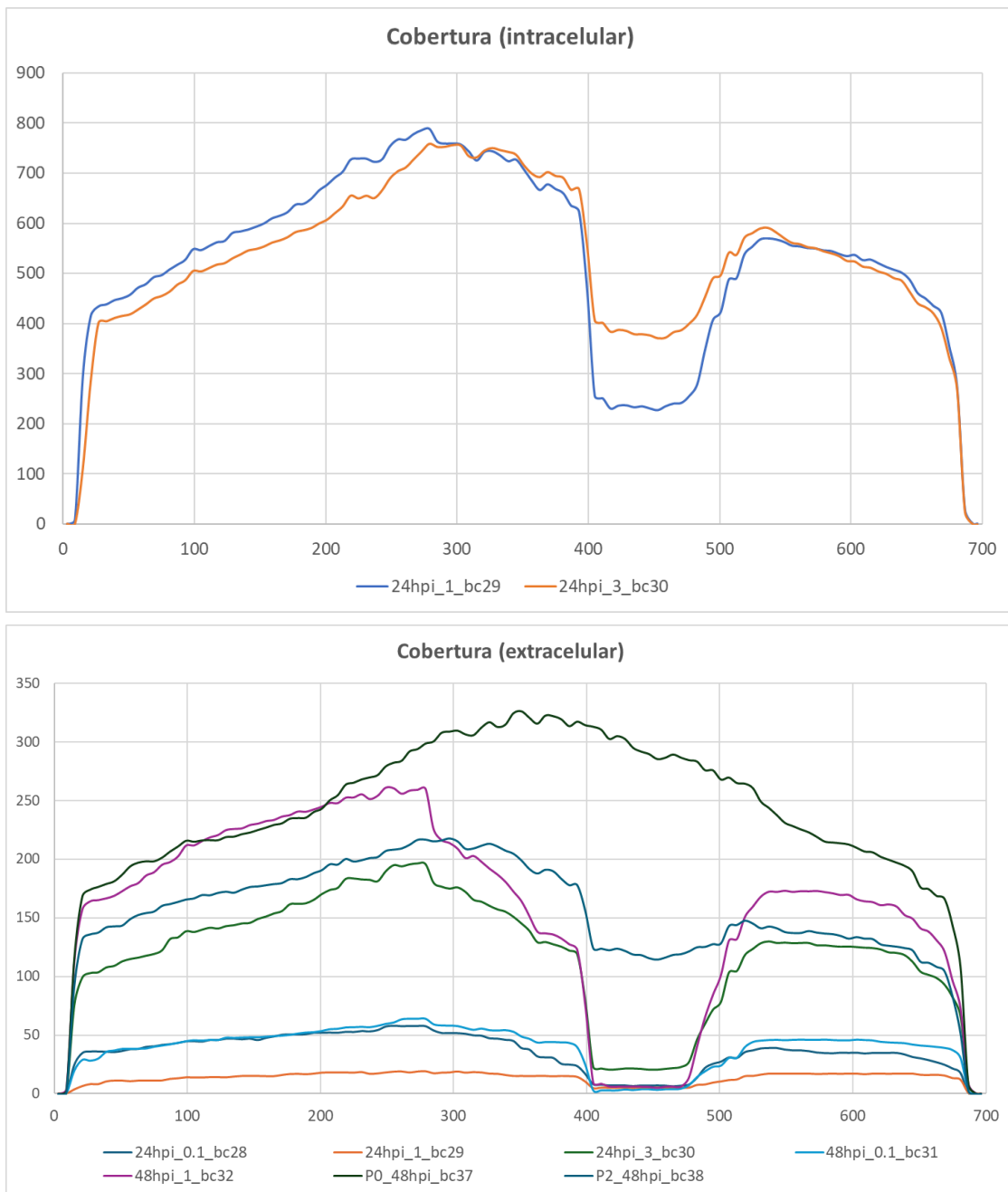


Figura 3.15. Perfiles de cobertura de amplicones de MAYV_sgRNA2_HDV obtenidos por secuenciación ONT. El panel superior corresponde a ARN intracelular (24 hpi MOI 1 y 3; bc29 y bc30), sin lecturas en otras condiciones. El panel inferior muestra ARN extracelular (24 hpi MOI 0,1; 1; 3 y 48 hpi MOI 0,1; 1), sin lecturas a 48 hpi MOI 3 en A549-Cas9. Se incluyen además amplicones de las progenies P0 (bc37) y P2 (bc38). Los datos se detallan en la **tabla S3.2**.

2- Análisis de expresión de ZAP en células A549-Cas9.

Como segundo análisis de funcionalidad del sgRNA2 codificado en el genoma viral, se analizó la expresión de ZAP por WB en células A549-Cas9 infectadas con MAYV_sgRNA2_miR o MAYV_sgRNA2_HDV para evaluar su capacidad de silenciar ZAP. Las células se infectaron con MAYV WT, MAYV_sgRNA2_miR o MAYV_sgRNA2_HDV (MOI = 0,1; 1 y 3) y se recolectaron a las 24 y 48 HPI. Los resultados mostraron que la expresión de la isoforma corta ZAPS fue menor en células infectadas con ambos virus sintéticos en comparación con el virus WT, de forma consistente en todas las MOI y tiempos analizados. En contraste, no se observaron diferencias significativas en la expresión de ZAPL bajo ninguna condición evaluada (Fig. 3.16 y 3.17).

En conjunto, estos resultados sugieren un efecto diferencial sobre las isoformas de ZAP, aunque no permiten concluir de forma directa que el sgRNA2 inducido por los virus recombinantes genere una inhibición eficiente de la expresión de ZAP.

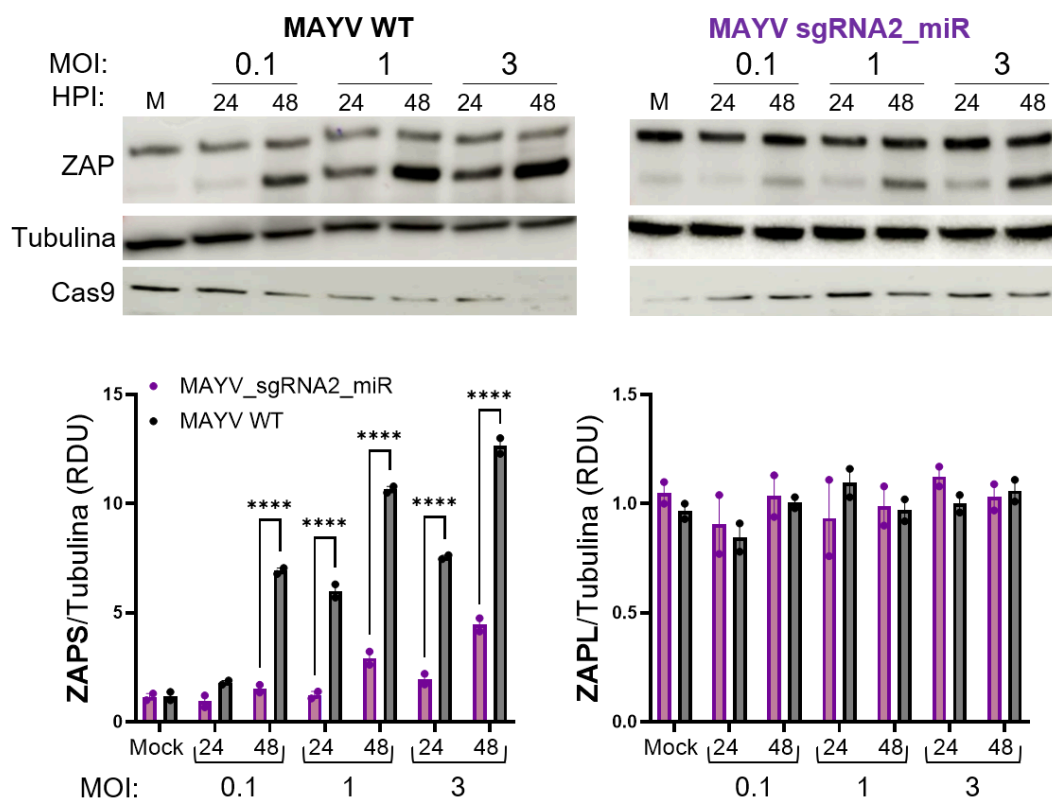


Figura 3.16: Expresión de ZAP en células A549-Cas9 infectadas con MAYV_sgRNA2_miR. Análisis por WB de ZAP. Se muestra WB para tubulina y Cas9 como control. Células A549-Cas9 fueron infectadas con MAYV WT o MAYV_sgRNA2_miR a MOI = 0,1, 1 y 3, o sin infectar (M). Los pellets se recolectaron a los tiempos indicados (HPI). Datos expresados como media \pm error estándar (n=2). Prueba t de Student vs. mock: ns, no significativo; *p \leq 0.05; **p \leq 0.01; ***p \leq 0.001.

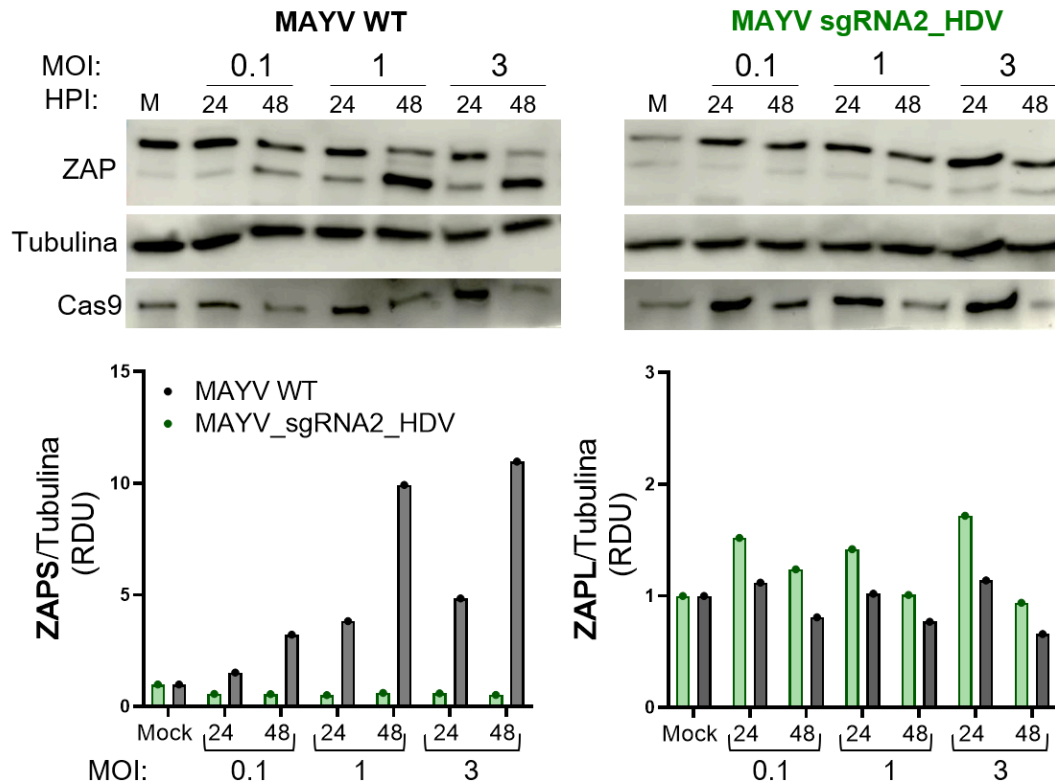


Figura 3.17: Expresión de ZAP en células A549-Cas9 infectadas con MAYV_sgRNA2_HDV. Análisis por WB de ZAP. Se muestra WB para tubulina y Cas9 como control. Células A549-Cas9 fueron infectadas con MAYV WT o MAYV_sgRNA2_HDV a MOI = 0.1, 1 y 3, o sin infectar (M). Los pellets se recolectaron a los tiempos indicados (HPI). n=1.

3- Eficiencia de edición génica en células infectadas con MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV.

Como tercer análisis, se evaluó la eficiencia de edición génica en células infectadas mediante secuenciación del locus blanco de ZAP (ZC3HAV1). Para ello, se extrajo ADN genómico de células infectadas con MAYV_sgRNA2_miR (MOI = 0.1 y 1; 24 y 48 HPI) y MAYV_sgRNA2_HDV (MOI = 0.1 y 3; 24 y 48 HPI), incluyendo células mock sin infectar como control. Una región flanqueante al sitio blanco del sgRNA2 fue amplificada por PCR (produciendo un amplicón de 566 pb), purificada (**Fig. S3.2A**) y analizada mediante dos estrategias de secuenciación.

El análisis por secuenciación Sanger, junto con la deconvolución mediante ICE, no evidenció *indels* detectables, obteniéndose secuencias que fueron indistinguibles de la condición mock (**Fig. S3.2B**), lo que indica ausencia de edición detectable bajo estas condiciones.

Para aumentar la sensibilidad, los amplicones se sometieron a secuenciación masiva por Illumina (PE150) y se analizaron con CRISPResso2. En todas las condiciones, la gran mayoría de las lecturas correspondieron a secuencias no modificadas (>98%).

Sin embargo, se observó un aumento de la frecuencia de lecturas modificadas en función de la MOI y del tiempo postinfección. Los valores máximos (indicados en rojo) se observaron a las 48 HPI: 1,5% (738 lecturas) para MAYV_sgRNA2_HDV (MOI = 3) y 1,3% (1080 lecturas) para MAYV_sgRNA2_miR (MOI = 1), mientras que el resto de las condiciones mostró niveles bajos comparables al control mock (0,09%).

En conjunto, aunque la secuenciación masiva por Illumina permitió detectar eventos de edición en condiciones de mayor MOI y tiempo, la baja frecuencia de estos (<1,5%) sugiere que no se produjo una edición génica robusta en el locus de ZAP tras la infección viral en las condiciones evaluadas.

Tabla 3.1 Frecuencia de edición génica en el locus *ZC3HAV1* luego de la infección con virus MAYV sintéticos. Resultado del análisis realizado con CRISPResso versión: 2.3.3¹⁷⁹

Virus MAYV	MOI	HPI	Lecturas alineadas ^a (n)	Lecturas sin modificaciones (n, %)	Lecturas modificadas (n, %) ^b
HDV	0.1	24	36.357	36.311(99,87%)	46 (0,1%)
HDV	0.1	48	19.392	19.279 (99,42%)	113 (0,6%)
HDV	3	24	34.555	34.333 (99,36%)	222 (0,7%)
HDV	3	48	49.910	49.172(98,52%)	738 (1,5%)
miR	0.1	24	82.060	82.021 (99,95%)	39 (0,05%)
miR	0.1	48	19.600	19.574(99,87%)	26 (0,13%)
miR	1	24	48.115	48.055 (99,88%)	60 (0,12%)
miR	1	48	84.653	83.573 (98,72%)	1080 (1,3%)
Mock	—	—	58.794	58.742 (99,91%)	51 (0,09%)

^atotal de lecturas alineadas a un amplicón de referencia de 150 pb conteniendo el sitio blanco del sgRNA2. ^bLas lecturas modificadas incluyen inserciones y/o deleciones dentro de la ventana de cuantificación definida alrededor del sitio de corte.

Expresión del miR-124 en células infectadas con MAYV_sgRNA2_miR.

El constructo MAYV_sgRNA2_miR incorpora el pri-miR-124-2 (208 nt), el cual se procesa intracelularmente a través de la maquinaria endógena de la biogénesis de mirRNAs para generar miR-124 maduro (~20 nt) y liberar el sgRNA2 en el extremo 5' del ARN subgenómico (**Fig. 3.18**).

Con el objetivo de confirmar la producción de miR-124, se utilizó *stem-loop* RT-PCR, una técnica específica y sensible para detectar miARNs maduros de tamaño pequeño. Células A549-Cas9 fueron infectadas (MOI = 3) y el ARN se extrajo a 24 y 48 HPI, separándose en fracciones <200 nt y >200 nt. Se empleó miR-423-5p como control endógeno, considerando que miR-124 no se expresa en A549. Como se muestra en la **figura 3.18**, la fracción <200 nt presentó niveles detectables de miR-124 maduro a ambos tiempos postinfección, lo que confirma el correcto procesamiento del pri-miR-124-2 en células infectadas.

En conjunto, estos resultados demuestran que el sistema viral permite la expresión y maduración eficiente del miRNA exógeno en el contexto de la infección.

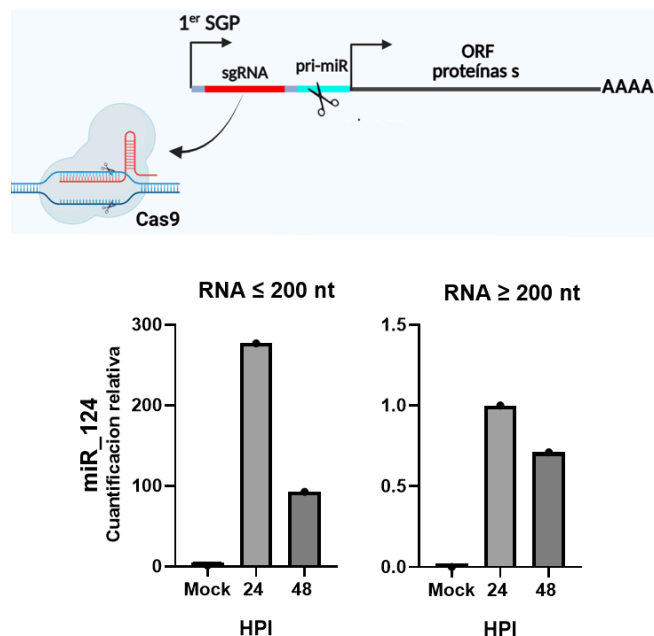


Figura 3.18: Cuantificación relativa de miR-124 maduro en células A549-Cas9 infectadas con MAYV sgRNA2_miR. Células A549-Cas9 fueron infectadas con MAYV sgRNA2_miR (MOI = 3). El ARN intracelular se extrajo a las 24 y 48 HPI y se fraccionó en ARN de tamaño menor a 200 nt (gráfico izquierdo) y mayor a 200 nt (gráfico derecho). La cuantificación relativa de miR-124 maduro se realizó mediante *stem-loop* RT-PCR utilizando miR-423-5p como control endógeno. Cada punto en los gráficos representa una muestra biológica individual, por lo que no se muestran barras de error ni test estadísticos.

Discusión

En este capítulo se evaluó la viabilidad de utilizar el virus MAYV, un alfavirus de replicación citoplasmática eficiente, como plataforma viral para la expresión intracelular de ARN guías (sgRNA) del sistema CRISPR/Cas9 durante una infección viral activa. La hipótesis planteada proponía que la incorporación de un sgRNA en el genoma de MAYV permitiría inducir edición génica específica en células que expresen Cas9 de manera constitutiva, facilitando el estudio funcional de factores celulares relevantes en la interacción virus–hospedero. En este contexto, los resultados obtenidos constituyen una prueba de concepto que demuestra que un alfavirus puede ser reprogramado para expresar ARN pequeños funcionales en células infectadas, aunque también revelan limitaciones importantes en términos de eficiencia de edición génica.

Desde el punto de vista experimental, la generación exitosa de dos clones infecciosos sintéticos de MAYV: MAYV_sgRNA2_miR y MAYV_sgRNA2_HDV, que incorporan en su genoma un sgRNA dirigido contra ZAP, demuestra que el genoma de MAYV tolera la inserción de secuencias exógenas asociadas a la expresión de sgRNAs sin comprometer de forma crítica su infectividad. Este hallazgo es relevante, ya que confirma la plasticidad del genoma alfaviral y su potencial como vector para aplicaciones biotecnológicas más allá de la expresión proteica convencional.

Sin embargo, los resultados también evidencian que la estrategia de procesamiento del sgRNA es un determinante crítico del sistema. Para el procesamiento del sgRNA durante la replicación viral se utilizaron dos estrategias distintas, un precursor de microARN (pri-miR-124-2) y una ribozima autocatalítica del virus de la hepatitis delta (HDV). En ambos casos se obtuvieron virus viables, capaces de producir stocks virales con títulos comparables al virus WT, demostrando que el genoma de MAYV tolera la incorporación de secuencias exógenas asociadas a la expresión de ARN pequeños sin comprometer completamente su viabilidad (**Fig. 3.7**).

Los estudios de cinética de replicación revelaron diferencias claras entre ambos virus, en términos de estabilidad genómica del inserto y compatibilidad con la biología viral. El clon MAYV_sgRNA2_miR mostró mayor estabilidad del inserto, el cual se mantuvo predominantemente en las partículas infectivas presentes en el medio celular, a pesar de la aparición de genomas defectivos en la fracción intracelular, particularmente a muy baja MOI (**Fig. 3.11**). Además, el pri-miR-124-2 incorporado en el genoma viral fue correctamente procesado en células infectadas, generando miR-124 maduro (**Fig. 3.16**). Dado que este microARN no se expresa de forma basal en células A549, su detección constituye una prueba funcional sólida de que MAYV puede actuar como

plataforma para la expresión y procesamiento intracelular de ARN pequeños durante la infección.

En contraste, MAYV_sgRNA2_HDV presentó una inestabilidad del inserto mucho mayor, con pérdida frecuente de la secuencia correspondiente a la ribozima HDV en los genomas virales de las partículas infectivas. Aunque se detectaron genomas con el inserto intacto, estos representaron una fracción minoritaria de la población viral (**Fig. 3.13** y **Fig. 3.14**). Estos resultados indican que, en las condiciones evaluadas, la estrategia basada en ribozimas autocatalíticas es menos compatible con la dinámica replicativa de MAYV y se pierde rápidamente.

Desde el punto de vista funcional, la evaluación de la eficiencia de edición génica mostró resultados consistentes entre ambas plataformas de secuenciación. El análisis mediante ICE-Sanger no detectó *indels* en el locus de ZAP (**Fig. S3.2A**). De forma acorde, la secuenciación masiva por Illumina permitió corroborar este resultado en la mayoría de las muestras. No obstante, se llegó a evidenciar *indels* de muy baja frecuencia (~1,5%) en condiciones de alta MOI y a las 48 HPI, hasta 738 lecturas modificadas (1,5%) para MAYV_sgRNA2_HDV (MOI = 3, 48 HPI) y 1080 lecturas modificadas (1,3%) para MAYV_sgRNA2_miR (MOI = 1, 48 HPI) (**Tabla 3.1**).

En conjunto, estos resultados indican que el sistema desarrollado puede generar eventos detectables de edición génica, pero a una frecuencia muy baja (<1,5%), insuficiente para evidenciar una edición robusta y específica del locus de ZAP en comparación con el control. La ausencia de una reducción clara y consistente de las isoformas ZAPS y ZAPL respalda esta interpretación. Dado que el sgRNA2 está diseñado contra todas las isoformas de ZAP, la disminución observada específicamente en ZAPS probablemente refleja un efecto indirecto asociado al fenotipo atenuado de los virus desarrollados, más que un silenciamiento específico de esta proteína.

Diversos factores biológicos podrían estar contribuyendo a la baja eficiencia de edición observada. Dado que el sgRNA2 mostró ser muy eficiente en células A549 (>80%, **Fig. 2.13**, Capítulo 2), sugiere que su incorporación en el genoma viral podría no favorecer una expresión o procesamiento adecuados. Por ejemplo, la expresión y disponibilidad intracelular del sgRNA derivado del virus podrían no alcanzar niveles suficientes o sostenidos para permitir una interacción eficiente con Cas9 en el contexto de una infección aguda. Además, el efecto citopático marcado inducido por MAYV en células A549 reduce la ventana temporal disponible para que se acumulen eventos de edición detectables luego de la infección, limitando la expansión de células potencialmente editadas.

En conjunto, estos hallazgos indican que, si bien MAYV puede ser utilizado como vector para la expresión de ARN pequeños, su aplicación como plataforma para edición génica eficiente presenta limitaciones importantes en su configuración actual impuestas por la biología de MAYV y el contexto celular. La identificación de estas limitaciones constituye un paso esencial para orientar futuras optimizaciones del diseño viral y del sistema experimental. No obstante, este sistema abre nuevas posibilidades para el desarrollo de vectores virales basados en ARN capaces de modular la expresión génica de manera transitoria durante la infección.

Conclusiones

- Se construyeron exitosamente dos clones infecciosos sintéticos de MAYV, diseñados para portar un sgRNA contra ZAP. El uso de secuencias de procesamiento basadas en pri-miRNA y la ribozima HDV permitió la recuperación de virus con capacidad infectiva, demostrando que el genoma de MAYV tolera la inserción de secuencias exógenas asociadas a la expresión de ARN pequeños.
- El constructo **MAYV sgRNA2_miR** exhibió una mayor estabilidad genómica frente a la variante con ribozima HDV y permitió validar el procesamiento funcional del ARN heterólogo mediante la detección de miR-124 maduro en células infectadas, confirmando la competencia del sistema para ser procesado.
- La variante **MAYV_sgRNA2_HDV** presentó una rápida delección del inserto en la progenie viral. Esta inestabilidad estructural restringe su potencial como vector para la expresión sostenida de secuencias de sgRNA en ensayos de cinética de replicación o en múltiples pasajes.
- La secuenciación masiva por Illumina permitió detectar eventos de edición de muy baja frecuencia ($\leq 1,5\%$) en el locus de ZAP a MOI altas, sugiriendo que los virus desarrollados no fueron suficientes para provocar una edición génica eficiente.
- La reducción en los niveles de la isoforma ZAPS durante la infección no pudo atribuirse a un silenciamiento mediado por CRISPR/Cas9. Este fenómeno sugiere un efecto indirecto vinculado a la atenuación en la replicación de los virus sintéticos, en lugar de una edición génica específica.
- En conjunto, los resultados indican que el sistema permitió la generación de virus sintéticos viables, pero no fue funcional como plataforma eficiente de edición génica en las condiciones evaluadas.

Perspectivas

Los resultados obtenidos permitieron identificar diversas limitaciones del sistema y delinear posibles estrategias para su optimización. La detección de eventos de edición génica de muy baja frecuencia sugiere que el sgRNA2 codificado desde el genoma viral podría ser funcional sólo en un subconjunto reducido de células, lo que indica que el rendimiento del sistema estaría condicionado por múltiples factores. Entre ellos se incluyen: (i) una expresión insuficiente o una estabilidad limitada del sgRNA maduro generado a partir del genoma viral; (ii) una eficiencia intrínseca subóptima del sgRNA2 en el contexto de expresión viral; y (iii) el efecto citopático de la infección por MAYV, que produce la lisis celular tras la infección y dificulta el análisis de poblaciones celulares infectadas y potencialmente editadas.

Para abordar estas limitaciones, una primera estrategia sería avanzar en el desarrollo de una biblioteca que incluya al menos tres sgRNAs validados dirigidos contra un mismo gen. Este enfoque es común en los CRISPR *screens*, donde se emplean múltiples sgRNAs por gen blanco para contrarrestar posibles efectos derivados de estructuras secundarias del ARN o variaciones en la eficiencia de edición intrínseca de cada sgRNA. En este trabajo se utilizó únicamente un sgRNA dirigido contra ZAP, debido a que su eficiencia había sido previamente validada en células A549 y a que disponíamos de herramientas experimentales e insumos para medir su funcionalidad, como ensayos de cinética de replicación y anticuerpos específicos contra ZAP. Sin embargo, es posible que este sgRNA2 no mantenga la misma eficiencia en el contexto del inserto diseñado, el cual además incluye secuencias de enzimas de restricción. Asimismo, el sgRNA2 comercial utilizado para generar la línea celular ZAP KO presenta una eficiencia particularmente elevada, ya que incorpora modificaciones químicas que protegen al ARN de la degradación por exonucleasas y reducen la activación de la respuesta inmune innata intracelular, lo que favorece una mayor estabilidad del sgRNA en el interior celular.

A su vez, resulta fundamental implementar estrategias experimentales que permitan una lectura directa, sensible y cuantificable de la actividad del sistema CRISPR, como la incorporación de controles positivos fácilmente medibles, como reporteros fluorescentes (GFP) o lumínicos (luciferasa). De esta forma, la pérdida de señal de un reportero podría utilizarse como una métrica más sensible para detectar eventos de edición génica.

Otro aspecto relevante a considerar es la biología replicativa de MAYV. Como se discutió anteriormente, este virus presenta una replicación rápida en células A549, con una progresión acelerada de la infección y una marcada lisis celular, un rasgo bien

documentado de su interacción con células susceptibles. En nuestro caso utilizamos células A549 que expresan constitutivamente Cas9 (A549-Cas9), una línea celular previamente disponible en el laboratorio y generada para estudios de edición génica. Mientras que este sistema permitió evaluar de forma directa la funcionalidad del sgRNA expresado por el virus en un contexto celular permisivo para la replicación de MAYV, la dinámica replicativa de MAYV restringe la ventana temporal disponible para la acumulación y detección de eventos de edición génica y reduce la proporción de células infectadas viables en las que puedan analizarse indels de manera confiable.

Para abordar esta limitación, el sistema podría evaluarse en líneas celulares alternativas que presenten una cinética de infección más lenta o una mayor tolerancia a la infección por MAYV, lo que facilitaría la recuperación y el análisis de células potencialmente editadas. Dado que actualmente no se dispone de un conocimiento exhaustivo sobre qué líneas celulares presentan un efecto citopático reducido frente a MAYV manteniendo una replicación eficiente, este aspecto deberá ser evaluado empíricamente. En estos casos, las líneas celulares seleccionadas deberían además modificarse para expresar Cas9 de manera constitutiva.

A largo plazo, los resultados de este capítulo establecen una base conceptual sólida para trabajar en el desarrollo de plataformas virales destinadas a la expresión intracelular de ARN pequeños en un contexto de infección activa. La optimización del sistema podría permitir su aplicación en estudios funcionales de interacción virus–hospedero e incluso en el desarrollo de estrategias experimentales de *screening* basadas en alfavirus. Además, el conocimiento generado podría orientar el diseño de vectores virales capaces de modular la expresión de factores antivirales específicos, como ZAP, con posibles aplicaciones terapéuticas o biotecnológicas.

CONCLUSIONES FINALES

Esta tesis investigó cómo la plasticidad genómica de los virus ARN y la composición de sus genomas contribuyen a su capacidad de adaptación y diversificación en distintos contextos biológicos. A través de diferentes modelos virales y enfoques experimentales, se exploró cómo variaciones en la arquitectura genómica impactan tanto en procesos evolutivos como en mecanismos funcionales durante la infección

En el primer capítulo demostramos que la composición nucleotídica y los patrones de uso de codones del gen S de SARS-CoV-2 reflejan la estructura evolutiva de los linajes virales y permiten distinguir variantes emergentes. En particular, la variante Ómicron mostró un perfil composicional claramente diferenciado del resto de las VOCs, consistente con una trayectoria evolutiva independiente y con un período prolongado de circulación no detectada previo a su identificación.

En el segundo capítulo analizamos cómo la composición de dinucleótidos CpG en el genoma de MAYV está estrechamente vinculada a mecanismos de inmunidad innata del hospedador. El enriquecimiento experimental de CpG en la variante FG+, que incorpora sitios CpG presentes únicamente en secuencias naturales aisladas del virus, resultó en una marcada atenuación dependiente de la proteína antiviral ZAP, identificando a la isoforma ZAPL como el principal mediador restrictivo en etapas tempranas de la infección. El hecho de que el virus silvestre mantenga sensibilidad a ZAP, a pesar de presentar una subrepresentación de CpG, sugiere que la composición genómica de MAYV refleja un equilibrio evolutivo asociado a su ciclo entre hospedadores vertebrados e invertebrados.

Finalmente, en el tercer capítulo exploramos la posibilidad de reprogramar el genoma de MAYV como plataforma experimental para la expresión de ARN pequeños, mediante la incorporación de un sgRNA del sistema CRISPR/Cas9 dirigido contra ZAP. Si bien no se detectó edición génica eficiente en las condiciones evaluadas, los resultados constituyen un paso esencial para orientar futuras optimizaciones del diseño viral y el sistema experimental.

En conjunto, los resultados de esta tesis demuestran que la composición genómica no constituye únicamente una característica pasiva de los virus ARN, sino un determinante activo de su biología, capaz de modular tanto su evolución como su interacción con el hospedador. Estos hallazgos integran evidencia evolutiva, mecanística y experimental que posiciona a la arquitectura del genoma viral como un eje central en la comprensión de la dinámica virus–hospedero y como una herramienta potencial para el diseño racional de virus con aplicaciones biotecnológicas.

REFERENCIAS BIBLIOGRÁFICAS

1. Salyer, S. J., Silver, R., Simone, K. & Barton Behravesh, C. Prioritizing Zoonoses for Global Health Capacity Building—Themes from One Health Zoonotic Disease Workshops in 7 Countries, 2014–2016. *Emerg. Infect. Dis.* **23**, S55–S64 (2017).
2. Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
3. Kreuder Johnson, C. *et al.* Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* **5**, 14830 (2015).
4. Dolan, P. T., Whitfield, Z. J. & Andino, R. Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. *Annu. Rev. Virol.* **5**, 69–92 (2018).
5. Domingo, E., García-Crespo, C., Lobo-Vega, R. & Perales, C. Mutation Rates, Mutation Frequencies, and Proofreading-Repair Activities in RNA Virus Genetics. *Viruses* **13**, 1882 (2021).
6. Domingo, E. *et al.* The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance — a review. *Gene* **40**, 1–8 (1985).
7. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626 (2011).
8. Chen, R. *et al.* Pattern recognition receptors: function, regulation and therapeutic potential. *Signal Transduct. Target. Ther.* **10**, 216 (2025).
9. Muhammad, I., Contes, K., Bility, M. T. & Tang, Q. Chasing Virus Replication and Infection: PAMP-PRR Interaction Drives Type I Interferon Production, Which in Turn Activates ISG Expression and ISGylation. *Viruses* **17**, (2025).
10. Ji, L. *et al.* The crucial regulatory role of type I interferon in inflammatory diseases. *Cell Biosci.* **13**, 230 (2023).
11. Gaunt, E. R. & Digard, P. Compositional biases in RNA viruses: Causes, consequences and applications. *Wiley Interdiscip. Rev. RNA* **13**, e1679 (2022).
12. Tsetsarkin, K. A., Vanlandingham, D. L., McGee, C. E. & Higgs, S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. *PLoS Pathog.* **3**, e201 (2007).
13. Urbanowicz, R. A. *et al.* Human Adaptation of Ebola Virus during the West African Outbreak. *Cell* **167**, 1079-1087.e5 (2016).
14. Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
15. Sánchez, A., Contreras, A., Corrales, J. C. & de la Fe, C. En el principio fue la zoonosis: One Health para combatir esta y futuras pandemias. Informe SESPAS 2022. *Gac. Sanit.* **36**, S61–S67 (2022).
16. Novel 2019 coronavirus genome - SARS-CoV-2 coronavirus. *Virological* <https://virological.org/t/novel-2019-coronavirus-genome/319> (2020).
17. Naming the coronavirus disease (COVID-19) and the virus that causes it. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it).
18. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **19**, 141–154 (2021).
19. Pagani, I., Ghezzi, S., Alberti, S., Poli, G. & Vicenzi, E. Origin and evolution of SARS-CoV-2. *Eur. Phys. J. Plus* **138**, 157 (2023).
20. Temmam, S. *et al.* Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **604**, 330–336 (2022).
21. Xu, Z. *et al.* CX1/BtSY2 and BANAL-20-52 exhibit broader receptor binding and higher affinities to multiple animal ACE2 orthologs than SARS-CoV-2 prototype. *J. Virol.* **99**, e00283-25 (2025).
22. Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).
23. Zhang, T., Wu, Q. & Zhang, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr. Biol. CB* **30**, 1346-1351.e2 (2020).
24. El Grupo Consultivo Científico de la OMS publica un informe sobre los orígenes de la COVID-19. <https://www.who.int/es/news/item/27-06-2025-who-scientific-advisory-group-issues-report-on-origins-of-covid-19>.
25. Gorkhali, R. *et al.* Structure and Function of Major SARS-CoV-2 and SARS-CoV Proteins. *Bioinforma. Biol. Insights* **15**, 11779322211025876 (2021).

26. Hartenian, E. *et al.* The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
27. Poland, G. A., Ovsyannikova, I. G. & Kennedy, R. B. SARS-CoV-2 immunity: review and applications to phase 3 vaccine candidates. *Lancet* **396**, 1595–1606 (2020).
28. Markov, P. V. *et al.* The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* **21**, 361–379 (2023).
29. Biebricher, C. K. & Eigen, M. What is a quasispecies? *Curr. Top. Microbiol. Immunol.* **299**, 1–31 (2006).
30. Domingo, E., Sheldon, J. & Perales, C. Viral quasispecies evolution. *Microbiol. Mol. Biol. Rev. MMBR* **76**, 159–216 (2012).
31. Forni, D., Cagliani, R., Pontremoli, C., Clerici, M. & Sironi, M. The substitution spectra of coronavirus genomes. *Brief. Bioinform.* **23**, bbab382 (2022).
32. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).
33. Wang, D. *et al.* The SARS-CoV-2 subgenome landscape and its novel regulatory features. *Mol. Cell* **81**, 2135-2147.e5 (2021).
34. Colson, P. *et al.* “Fertile” Mutations in SARS-CoV-2 RNA More Frequently Occurred in Hairpin Loops That Determine Virus Evolution. *Apmis* **133**, e70113 (2025).
35. Rambaut, A. *et al.* A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
36. Hadfield, J. *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).
37. GISAID - Initiative. <https://www.gisaid.org/>.
38. Parums, D. Editorial: Revised World Health Organization (WHO) Terminology for Variants of Concern and Variants of Interest of SARS-CoV-2. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.* **27**, e933622 (2021).
39. Marjanovic, S. *et al.* COVID-19 Genomics UK (COG-UK) Consortium: Final Report. *Rand Health Q.* **9**, 24 (2022).
40. Panzera, Y. *et al.* Transmission cluster of COVID-19 cases from Uruguay: emergence and spreading of a novel SARS-CoV-2 ORF6 deletion. *Mem. Inst. Oswaldo Cruz* **116**, e210275 (2022).
41. Rego, N. *et al.* Real-Time Genomic Surveillance for SARS-CoV-2 Variants of Concern, Uruguay. *Emerg. Infect. Dis.* **27**, 2957–2960 (2021).
42. Tosta, S. *et al.* Global SARS-CoV-2 genomic surveillance: What we have learned (so far). *Infect. Genet. Evol.* **108**, 105405 (2023).
43. Sarkar, M. M. H. *et al.* M gene targeted qRT-PCR approach for SARS-CoV-2 virus detection. *Sci. Rep.* **13**, 16659 (2023).
44. Vogels, C. B. F. *et al.* Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLOS Biol.* **19**, e3001236 (2021).
45. Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
46. Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
47. Faria, N. R. *et al.* Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).
48. Tegally, H. *et al.* Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021).
49. Dhar, M. S. *et al.* Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science* **374**, 995–999 (2021).
50. Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **603**, 679–686 (2022).
51. Kudriavtsev, A. V. *et al.* Immune Escape Associated with RBD Omicron Mutations and SARS-CoV-2 Evolution Dynamics. *Viruses* **14**, (2022).
52. Otto, S. P., MacPherson, A. & Colijn, C. Endemic does not mean constant as SARS-CoV-2 continues to evolve. *Evol. Int. J. Org. Evol.* **78**, 1092–1108 (2024).
53. Gorbalenya, A. E. & Lauber, C. Phylogeny of Viruses. *Ref. Module Biomed. Sci.* B978-0-12-801238-3.95723-4 (2017) doi:10.1016/B978-0-12-801238-3.95723-4.
54. Molecular Clocks | BEAST Documentation. <https://beast.community/clocks#fixed-local-clock>.

55. Ventoso, I. Codon Usage Bias in Human RNA Viruses and Its Impact on Viral Translation, Fitness, and Evolution. *Viruses* **17**, 1218 (2025).
56. Kustin, T. & Stern, A. Biased Mutation and Selection in RNA Viruses. *Mol. Biol. Evol.* **38**, 575–588 (2021).
57. Latrille, T. & Lartillot, N. An Improved Codon Modeling Approach for Accurate Estimation of the Mutation Bias. *Mol. Biol. Evol.* **39**, msac005 (2022).
58. Simón, D., Cristina, J. & Musto, H. Nucleotide Composition and Codon Usage Across Viruses and Their Respective Hosts. *Front. Microbiol.* **12**, 646300 (2021).
59. Davidson, A. *et al.* Over time analysis of the codon usage of SARS-CoV-2 and its variants. *Comput. Struct. Biotechnol. J.* **27**, 2034–2050 (2025).
60. Tort, F. L., Castells, M. & Cristina, J. A comprehensive analysis of genome composition and codon usage patterns of emerging coronaviruses. *Virus Res.* **283**, 197976 (2020).
61. Błażej, P., Mackiewicz, D. & Mackiewicz, P. Temporal and gene-specific dynamics of codon usage evolution in SARS-CoV-2 genomes. *BMC Genomics* <https://doi.org/10.1186/s12864-026-12672-4> (2026) doi:10.1186/s12864-026-12672-4.
62. Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570 (2015).
63. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **15**, e1006650 (2019).
64. Molecular Evolution, Phylogenetics and Epidemiology. <https://tree.bio.ed.ac.uk/>.
65. Fumagalli, S. E. *et al.* Analysis of 3.5 million SARS-CoV-2 sequences reveals unique mutational trends with consistent nucleotide and codon frequencies. *Virology* **20**, 31 (2023).
66. Fischer, C. *et al.* Emergence and spread of the SARS-CoV-2 omicron (BA.1) variant across Africa: an observational study. *Lancet Glob. Health* **13**, e256–e267 (2025).
67. Paz, M., Aldunate, F., Arce, R., Ferreira, I. & Cristina, J. An evolutionary insight into Severe Acute Respiratory Syndrome Coronavirus 2 Omicron variant of concern. *Virus Res.* **314**, 198753 (2022).
68. Kafai, N. M., Diamond, M. S. & Fox, J. M. Distinct Cellular Tropism and Immune Responses to Alphavirus Infection. *Annu. Rev. Immunol.* **40**, 615–649 (2022).
69. Wei, L. L. L., Tom, R. & Kim, Y. C. Mayaro Virus: An Emerging Alphavirus in the Americas. *Viruses* **16**, 1297 (2024).
70. Long, K. C. *et al.* Experimental transmission of Mayaro virus by *Aedes aegypti*. *Am. J. Trop. Med. Hyg.* **85**, 750–757 (2011).
71. Jose, J., Snyder, J. E. & Kuhn, R. J. A structural and functional perspective of alphavirus replication and assembly. *Future Microbiol.* **4**, 837–856 (2009).
72. Ribeiro-Filho, H. V. *et al.* Cryo-EM structure of the mature and infective Mayaro virus at 4.4 Å resolution reveals features of arthritogenic alphaviruses. *Nat. Commun.* **12**, 3038 (2021).
73. Marinho, M. dos S. *et al.* Evolutionary Profile of Mayaro Virus in the Americas: An Update into Genome Variability. *Viruses* **16**, 809 (2024).
74. Hean Tang, P. C., Hann Ng, W. & Liu, X. Host Immune Responses to Arthritogenic Alphavirus Infection, with Emphasis on Type I IFN Responses. *Zoonoses* **2**, (2022).
75. Wu, S.-R., Haag, L., Sjöberg, M., Garoff, H. & Hammar, L. The dynamic envelope of a fusion class II virus. E3 domain of glycoprotein E2 precursor in Semliki Forest virus provides a unique contact with the fusion protein E1. *J. Biol. Chem.* **283**, 26452–26460 (2008).
76. Mendonça, D. C., Reis, Erik. V. S., Arias, Nidia. E. C., Valencia, H. J. & Bonjardim, C. A. A study of the MAYV replication cycle: Correlation between the kinetics of viral multiplication and viral morphogenesis. *Virus Res.* **323**, 199002 (2023).
77. Holmes, A. C., Basore, K., Fremont, D. H. & Diamond, M. S. A molecular understanding of alphavirus entry. *PLOS Pathog.* **16**, e1008876 (2020).
78. Zhang, R. *et al.* Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* **557**, 570–574 (2018).
79. Andreolla, A. P., Borges, A. A., Bordignon, J. & Duarte Dos Santos, C. N. Mayaro Virus: The State-of-the-Art for Antiviral Drug Development. *Viruses* **14**, 1787 (2022).
80. Spuul, P., Balistreri, G., Kääriäinen, L. & Ahola, T. Phosphatidylinositol 3-kinase-, actin-, and microtubule-dependent transport of Semliki Forest Virus replication complexes from the plasma membrane to modified lysosomes. *J. Virol.* **84**, 7543–7557 (2010).

81. Ahola, T. & Kääriäinen, L. Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsP1 with 7-methyl-GMP. *Proc. Natl. Acad. Sci.* **92**, 507–511 (1995).
82. Ishida, R. *et al.* Mayaro Virus Non-Structural Protein 2 Circumvents the Induction of Interferon in Part by Depleting Host Transcription Initiation Factor IIE Subunit 2. *Cells* **10**, 3510 (2021).
83. Rupp, J. C., Sokoloski, K. J., Gebhart, N. N. & Hardy, R. W. Alphavirus RNA synthesis and non-structural protein functions. *J. Gen. Virol.* **96**, 2483–2500 (2015).
84. Tomar, S., Hardy, R. W., Smith, J. L. & Kuhn, R. J. Catalytic core of alphavirus nonstructural protein nsP4 possesses terminal adenylyltransferase activity. *J. Virol.* **80**, 9962–9969 (2006).
85. Paschoalino, M. *et al.* An update on the development of antiviral against Mayaro virus: from molecules to potential viral targets. *Arch. Microbiol.* **205**, 106 (2023).
86. Anderson, C. R., Downs, W. G., Wattley, G. H., Ahin, N. W. & Reese, A. A. Mayaro virus: a new human disease agent. II. Isolation from blood of patients in Trinidad, B.W.I. *Am. J. Trop. Med. Hyg.* **6**, 1012–1016 (1957).
87. Caicedo, E.-Y. *et al.* The epidemiology of Mayaro virus in the Americas: A systematic review and key parameter estimates for outbreak modelling. *PLoS Negl. Trop. Dis.* **15**, e0009418 (2021).
88. Navarrete-Espinosa, J. & Gómez-Dantés, H. [Arbovirus causing hemorrhagic fever at IMSS]. *Rev. Medica Inst. Mex. Seguro Soc.* **44**, 347–353 (2006).
89. Lorenz, C., Freitas Ribeiro, A. & Chiaravalloti-Neto, F. Mayaro virus distribution in South America. *Acta Trop.* **198**, 105093 (2019).
90. Diagne, C. T. *et al.* Mayaro Virus Pathogenesis and Transmission Mechanisms. *Pathogens* **9**, 738 (2020).
91. Ganjian, N. & Riviere-Cinamond, A. Mayaro virus in Latin America and the Caribbean. *Rev. Panam. Salud Pública* **44**, 1 (2020).
92. da Silva Neves, N. A. *et al.* Chikungunya, Zika, Mayaro, and Equine Encephalitis virus detection in adult Culicinae from South Central Mato Grosso, Brazil, during the rainy season of 2018. *Braz. J. Microbiol. Publ. Braz. Soc. Microbiol.* **53**, 63–70 (2022).
93. de Curcio, J. S. *et al.* Detection of Mayaro virus in *Aedes aegypti* mosquitoes circulating in Goiânia-Goiás-Brazil. *Microbes Infect.* **24**, 104948 (2022).
94. Dudley-Moore, D. RNA viruses: all bases covered? *Nat. Rev. Immunol.* **6**, 341–341 (2006).
95. Song, J. *et al.* Friend or foe: RIG- I like receptors and diseases. *Autoimmun. Rev.* **21**, 103161 (2022).
96. Wu, B. & Hur, S. How RIG-I like receptors activate MAVS. *Curr. Opin. Virol.* **12**, 91–98 (2015).
97. de Castro-Jorge, L. A. *et al.* The NLRP3 inflammasome is involved with the pathogenesis of Mayaro virus. *PLoS Pathog.* **15**, e1007934 (2019).
98. de Oliveira Souza, R. *et al.* Unraveling the complex interplay: immunopathology and immune evasion strategies of alphaviruses with emphasis on neurological implications. *Front. Cell. Infect. Microbiol.* **14**, (2024).
99. Figueiredo, C. M. *et al.* Mayaro Virus Replication Restriction and Induction of Muscular Inflammation in Mice Are Dependent on Age, Type-I Interferon Response, and Adaptive Immunity. *Front. Microbiol.* **10**, 2246 (2019).
100. Hernández-Sarmiento, L. J., Tamayo-Molina, Y. S., Valdés-López, J. F. & Urcuqui-Inchima, S. Mayaro virus infection elicits a robust pro-inflammatory and antiviral response in human macrophages. *Acta Trop.* **252**, 107146 (2024).
101. Rosa, R. B. *et al.* A comprehensive murine clinical model for development of countermeasures and studying Mayaro virus infection. *PLoS Negl. Trop. Dis.* **19**, e0013333 (2025).
102. Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**, 1499–1504 (1980).
103. Schorderet, D. F. & Gartler, S. M. Analysis of CpG suppression in methylated and nonmethylated species. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 957–961 (1992).
104. Gao, G., Guo, X. & Goff, S. P. Inhibition of retroviral RNA production by ZAP, a CCCH-type zinc finger protein. *Science* **297**, 1703–1706 (2002).

105. Charron, G., Li, M. M. H., MacDonald, M. R. & Hang, H. C. Prenylome profiling reveals S-farnesylation is crucial for membrane targeting and antiviral activity of ZAP long-isoform. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11085–11090 (2013).
106. Hayakawa, S. *et al.* ZAPS is a potent stimulator of signaling mediated by the RNA helicase RIG-I during antiviral responses. *Nat. Immunol.* **12**, 37–44 (2011).
107. Li, M. M. H. *et al.* Characterization of Novel Splice Variants of Zinc Finger Antiviral Protein (ZAP). *J. Virol.* **93**, e00715-19 (2019).
108. Wang, N. *et al.* Viral Induction of the Zinc Finger Antiviral Protein Is IRF3-dependent but NF- κ B-independent*. *J. Biol. Chem.* **285**, 6080–6090 (2010).
109. Shao, R., Visser, I., Fros, J. J. & Yin, X. Versatility of the Zinc-Finger Antiviral Protein (ZAP) As a Modulator of Viral Infections. *Int. J. Biol. Sci.* **20**, 4585–4600 (2024).
110. Shao, R. *et al.* ZAP inhibits double-stranded RNA virus infection by degrading negative-strand RNA and blocking the elongation phase of viral protein synthesis. *Cell Rep.* **45**, 116742 (2026).
111. Chiu, H.-P. *et al.* Inhibition of Japanese encephalitis virus infection by the host zinc-finger antiviral protein. *PLoS Pathog.* **14**, e1007166 (2018).
112. Bick, M. J. *et al.* Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562 (2003).
113. Lista, M. J. *et al.* Strain-Dependent Restriction of Human Cytomegalovirus by Zinc Finger Antiviral Proteins. *J. Virol.* **97**, e0184622 (2023).
114. Su, C., Zhang, J. & Zheng, C. Herpes simplex virus 1 UL41 protein abrogates the antiviral activity of hZAP by degrading its mRNA. *Virol. J.* **12**, 203 (2015).
115. Schwerk, J. *et al.* RNA-binding protein isoforms ZAP-S and ZAP-L have distinct antiviral and immune resolution functions. *Nat. Immunol.* **20**, 1610–1620 (2019).
116. Ly, P. T., Xu, S., Wirawan, M., Luo, D. & Roca, X. ZAP isoforms regulate unfolded protein response and epithelial- mesenchymal transition. *Proc. Natl. Acad. Sci.* **119**, e2121453119 (2022).
117. Lin, Y.-T. *et al.* Human cytomegalovirus evades ZAP detection by suppressing CpG dinucleotides in the major immediate early 1 gene. *PLOS Pathog.* **16**, e1008844 (2020).
118. Guo, X., Carroll, J.-W. N., Macdonald, M. R., Goff, S. P. & Gao, G. The zinc finger antiviral protein directly binds to specific viral mRNAs through the CCCH zinc finger motifs. *J. Virol.* **78**, 12781–12787 (2004).
119. Luo, X. *et al.* Molecular Mechanism of RNA Recognition by Zinc-Finger Antiviral Protein. *Cell Rep.* **30**, 46-52.e4 (2020).
120. Meagher, J. L. *et al.* Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24303–24309 (2019).
121. Guo, X., Ma, J., Sun, J. & Gao, G. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proc. Natl. Acad. Sci.* **104**, 151–156 (2007).
122. Mao, R. *et al.* Inhibition of Hepatitis B Virus Replication by the Host Zinc Finger Antiviral Protein. *PLoS Pathog.* **9**, e1003494 (2013).
123. Zhao, Y. *et al.* ZAP, a CCCH-Type Zinc Finger Protein, Inhibits Porcine Reproductive and Respiratory Syndrome Virus Replication and Interacts with Viral Nsp9. *J. Virol.* **93**, e00001-19 (2019).
124. Kerns, J. A., Emerman, M. & Malik, H. S. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. *PLoS Genet.* **4**, e21 (2008).
125. Karlberg, T. *et al.* Structural basis for lack of ADP-ribosyltransferase activity in poly(ADP-ribose) polymerase-13/zinc finger antiviral protein. *J. Biol. Chem.* **290**, 7336–7344 (2015).
126. Gonçalves-Carneiro, D., Mastrocola, E., Lei, X. & Bieniasz, P. D. Modulation of host gene expression by the zinc finger antiviral protein. *Proc. Natl. Acad. Sci. U. S. A.* **122**, e2420819122 (2025).
127. Leung, A. K. L. *et al.* Poly(ADP-ribose) regulates stress responses and microRNA activity in the cytoplasm. *Mol. Cell* **42**, 489–499 (2011).
128. Xue, G. *et al.* Poly(ADP-ribose) potentiates ZAP antiviral activity. *PLoS Pathog.* **18**, e1009202 (2022).

129. Goonawardane, N., Nguyen, D. & Simmonds, P. Association of Zinc Finger Antiviral Protein Binding to Viral Genomic RNA with Attenuation of Replication of Echovirus 7. *mSphere* **6**, e01138-20 (2021).
130. Law, L. M. J. *et al.* ZAP's stress granule localization is correlated with its antiviral activity and induced by virus replication. *PLoS Pathog.* **15**, e1007798 (2019).
131. Zhu, Y., Wang, X., Goff, S. P. & Gao, G. Translational repression precedes and is required for ZAP-mediated mRNA decay. *EMBO J.* **31**, 4236–4246 (2012).
132. Li, M. M. H. *et al.* TRIM25 Enhances the Antiviral Action of Zinc-Finger Antiviral Protein (ZAP). *PLOS Pathog.* **13**, e1006145 (2017).
133. Zhu, Y. *et al.* Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15834–15839 (2011).
134. Chen, G., Guo, X., Lv, F., Xu, Y. & Gao, G. p72 DEAD box RNA helicase is required for optimal function of the zinc-finger antiviral protein. *Proc. Natl. Acad. Sci.* **105**, 4352–4357 (2008).
135. Ficarelli, M. *et al.* KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife* **8**, e46767 (2019).
136. Youle, R. L. *et al.* KHNYN is a manganese-dependent endoribonuclease required for ZAP-mediated antiviral restriction. *Nucleic Acids Res.* **53**, gkaf1360 (2025).
137. Atkinson, N. J., Witteveltdt, J., Evans, D. J. & Simmonds, P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.* **42**, 4527–4545 (2014).
138. Fros, J. J. *et al.* CpG and UpA dinucleotides in both coding and non-coding regions of echovirus 7 inhibit replication initiation post-entry. *eLife* **6**, e29112 (2017).
139. Fros, J. J. *et al.* The dinucleotide composition of the Zika virus genome is shaped by conflicting evolutionary pressures in mammalian hosts and mosquito vectors. *PLOS Biol.* **19**, e3001201 (2021).
140. Gonçalves-Carneiro, D. *et al.* Rational attenuation of RNA viruses with zinc finger antiviral protein. *Nat. Microbiol.* **7**, 1558–1567 (2022).
141. Takata, M. A. *et al.* CG-dinucleotide suppression enables antiviral defense targeting non-self RNA. *Nature* **550**, 124–127 (2017).
142. Kmiec, D. *et al.* CpG Frequency in the 5' Third of the env Gene Determines Sensitivity of Primary HIV-1 Strains to the Zinc-Finger Antiviral Protein. *mBio* **11**, e02903-19 (2020).
143. Odon, V. *et al.* The role of ZAP and OAS3/RNaseL pathways in the attenuation of an RNA virus with elevated frequencies of CpG and UpA dinucleotides. *Nucleic Acids Res.* **47**, 8061–8083 (2019).
144. Gonçalves-Carneiro, D., Takata, M. A., Ong, H., Shilton, A. & Bieniasz, P. D. Origin and evolution of the zinc finger antiviral protein. *PLoS Pathog.* **17**, e1009545 (2021).
145. Simón, D. *et al.* Impact of the Zinc Antiviral Protein on the Genomic Composition of RNA Viruses Infecting Vertebrates. *Mol. Biol. Evol.* **42**, msaf135 (2025).
146. Falckenhayn, C. *et al.* Comprehensive DNA methylation analysis of the *Aedes aegypti* genome. *Sci. Rep.* **6**, 36444 (2016).
147. Klughammer, J. *et al.* Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nat. Commun.* **14**, 232 (2023).
148. Colmant, A. M. G. *et al.* Insect-Specific Flavivirus Replication in Mammalian Cells Is Inhibited by Physiological Temperature and the Zinc-Finger Antiviral Protein. *Viruses* **13**, 573 (2021).
149. Junglen, S. *et al.* Host Range Restriction of Insect-Specific Flaviviruses Occurs at Several Levels of the Viral Life Cycle. *mSphere* **2**, e00375-16 (2017).
150. Provataris, P., Meusemann, K., Niehuis, O., Grath, S. & Misof, B. Signatures of DNA Methylation across Insects Suggest Reduced DNA Methylation Levels in Holometabola. *Genome Biol. Evol.* **10**, 1185–1197 (2018).
151. Piyasena, T. B. H. *et al.* Infectious DNAs derived from insect-specific flavivirus genomes enable identification of pre- and post-entry host restrictions in vertebrate cells. *Sci. Rep.* **7**, 2940 (2017).
152. Simmonds, P., Xia, W., Baillie, J. K. & McKinnon, K. Modelling mutational and selection pressures on dinucleotides in eukaryotic phyla--selection against CpG and UpA in cytoplasmically expressed RNA and in RNA viruses. *BMC Genomics* **14**, 610 (2013).

153. Forni, D., Pozzoli, U., Cagliani, R., Clerici, M. & Sironi, M. Dinucleotide biases in RNA viruses that infect vertebrates or invertebrates. *Microbiol. Spectr.* **11**, e0252923 (2023).
154. Di Giallonardo, F., Schlub, T. E., Shi, M. & Holmes, E. C. Dinucleotide Composition in Animal RNA Viruses Is Shaped More by Virus Family than by Host Species. *J. Virol.* **91**, e02381-16 (2017).
155. Odon, V., Fiddaman, S. R., Smith, A. L. & Simmonds, P. Comparison of CpG- and UpA-mediated restriction of RNA virus replication in mammalian and avian cells and investigation of potential ZAP-mediated shaping of host transcriptome compositions. *RNA* **28**, 1089–1109 (2022).
156. Nguyen, L. P., Aldana, K. S., Yang, E., Yao, Z. & Li, M. M. H. Alphavirus Evasion of Zinc Finger Antiviral Protein (ZAP) Correlates with CpG Suppression in a Specific Viral nsP2 Gene Sequence. *Viruses* **15**, 830 (2023).
157. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore Rapid Barcoding. *Biol. Methods Protoc.* **5**, bpaa014 (2020).
158. Franz, S. *et al.* Human IFITM3 restricts chikungunya virus and Mayaro virus infection and is susceptible to virus-mediated counteraction. *Life Sci. Alliance* **4**, e202000909 (2021).
159. Trus, I. *et al.* CpG-Recoding in Zika Virus Genome Causes Host-Age-Dependent Attenuation of Infection With Protection Against Lethal Heterologous Challenge in Mice. *Front. Immunol.* **10**, 3077 (2019).
160. Ficarelli, M. *et al.* CpG Dinucleotides Inhibit HIV-1 Replication through Zinc Finger Antiviral Protein (ZAP)-Dependent and -Independent Mechanisms. *J. Virol.* **94**, e01337-19 (2020).
161. Sharp, C. P. *et al.* CpG dinucleotide enrichment in the influenza A virus genome as a live attenuated vaccine development strategy. *PLOS Pathog.* **19**, e1011357 (2023).
162. Gonzalez-Perez, A. C. *et al.* The Zinc Finger Antiviral Protein ZAP Restricts Human Cytomegalovirus and Selectively Binds and Destabilizes Viral UL4/UL5 Transcripts. *mBio* **12**, 10.1128/mbio.02683-20 (2021).
163. Fros, J. J. & Pijlman, G. P. Alphavirus Infection: Host Cell Shut-Off and Inhibition of Antiviral Responses. *Viruses* **8**, 166 (2016).
164. Lin, T. *et al.* CXCL10 Signaling Contributes to the Pathogenesis of Arthritogenic Alphaviruses. *Viruses* **12**, (2020).
165. Bengue, M. *et al.* Mayaro Virus Infects Human Brain Cells and Induces a Potent Antiviral Response in Human Astrocytes. *Viruses* **13**, 465 (2021).
166. Pacesa, M., Pelea, O. & Jinek, M. Past, present, and future of CRISPR genome editing technologies. *Cell* **187**, 1076–1100 (2024).
167. Karakus, U. *et al.* MHC class II proteins mediate cross-species entry of bat influenza viruses. *Nature* **567**, 109–112 (2019).
168. Ma, H. *et al.* A CRISPR-Based Screen Identifies Genes Essential for West-Nile-Virus-Induced Cell Death. *Cell Rep.* **12**, 673–683 (2015).
169. Ramezannia, Z., Shamekh, A. & Bannazadeh Baghi, H. CRISPR–Cas system to discover host-virus interactions in Flaviviridae. *Virol. J.* **20**, 247 (2023).
170. Finkel, Y. *et al.* A virally encoded high-resolution screen of cytomegalovirus dependencies. *Nature* **630**, 712–719 (2024).
171. King, C. R. *et al.* Pathogen-driven CRISPR screens identify TREX1 as a regulator of DNA self-sensing during influenza virus infection. *Cell Host Microbe* **31**, 1552-1567.e8 (2023).
172. OhAinle, M. *et al.* A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. *eLife* **7**, e39823 (2018).
173. Prelli Bozzo, C. *et al.* Replication competent HIV-guided CRISPR screen identifies antiviral factors including targets of the accessory protein Nef. *Nat. Commun.* **15**, 3813 (2024).
174. Paz, M. & Moratorio, G. Deep mutational scanning and CRISPR-engineered viruses: tools for evolutionary and functional genomics studies. *mSphere* **10**, e00508-24 (2025).
175. Boorsma, M., Koller, D., Renner, W. A. & Bachmann, M. F. New applications of alphavirus-based expression vectors. *Cytotechnology* **35**, 203–212 (2001).
176. Frolov, I. *et al.* Alphavirus-based expression vectors: strategies and applications. *Proc. Natl. Acad. Sci.* **93**, 11371–11377 (1996).
177. Marinho, M. dos S. *et al.* Development and validation of Mayaro virus with luciferase reporter genes as a tool for antiviral assays. *Heliyon* **10**, e33885 (2024).
178. Shapiro, J. S., Varble, A., Pham, A. M. & Tenoever, B. R. Noncanonical cytoplasmic processing of viral microRNAs. *RNA N. Y. N* **16**, 2068–2074 (2010).

179. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
180. Sun, C., Gardner, C. L., Watson, A. M., Ryman, K. D. & Klimstra, W. B. Stable, High-Level Expression of Reporter Proteins from Improved Alphavirus Expression Vectors To Track Replication and Dissemination during Encephalitic and Arthritogenic Disease. *J. Virol.* **88**, 2035–2046 (2014).
181. Steel, J. J., Henderson, B. R., Lama, S. B., Olson, K. E. & Geiss, B. J. Infectious alphavirus production from a simple plasmid transfection. *Virology* **8**, 356 (2011).

INFORMACIÓN SUPLEMENTARIA

Tablas y figuras suplementarias:

Capítulo 1

Tabla S1.1: Secuencias de Ómicron (BA.2 y BA.3) incluidas en el análisis filodinámico bayesiano complementario de la figura S1.1. Las secuencias fueron obtenidas de la base GISAID.

Nombre de la secuencia	Número de acceso (GISAID)	Fecha de aislamiento	Genotipo
hCoV-19/Poland/Kce-2877/2022	EPI_ISL_10703609	2022-02-28	Omicron BA.2
hCoV-19/Poland/Kce-2874/2022	EPI_ISL_10703606	2022-02-28	Omicron BA.2
hCoV-19/Slovakia/BA_22_00008002/2022	EPI_ISL_10702197	2022-02-02	Omicron BA.2
hCoV-19/Slovakia/BA_22_00005176/2022	EPI_ISL_10702196	2022-01-22	Omicron BA.2
hCoV-19/Netherlands/FR-RIVM-89749/2022	EPI_ISL_10696417	2022-02-17	Omicron BA.2
hCoV-19/Belgium/ULG-25154/2022	EPI_ISL_10695957	2022-02-23	Omicron BA.2
hCoV-19/Switzerland/BE-IFIK-6878-2052/2022	EPI_ISL_10695813	2022-02-20	Omicron BA.2
hCoV-19/Switzerland/BE-IFIK-6875-2044/2022	EPI_ISL_10695803	2022-02-21	Omicron BA.2
hCoV-19/Austria/LB-R00063-S371/2022	EPI_ISL_10693259	2022-02-23	Omicron BA.2
hCoV-19/Austria/LB-R00063-S361/2022	EPI_ISL_10693251	2022-02-22	Omicron BA.2
hCoV-19/Norway/OUS-21440/2022	EPI_ISL_10681732	2022-02-16	Omicron BA.2
hCoV-19/Norway/OUS-21437/2022	EPI_ISL_10681731	2022-02-16	Omicron BA.2
hCoV-19/Italy/EMR_AUSLRomagna_C022-22-04/2022	EPI_ISL_10681468	2022-02-23	Omicron BA.2
hCoV-19/Italy/EMR_AUSLRomagna_C022-22-02/2022	EPI_ISL_10681465	2022-02-22	Omicron BA.2
hCoV-19/Denmark/DCGC-401971/2022	EPI_ISL_10680071	2022-02-26	Omicron BA.2
hCoV-19/Denmark/DCGC-401970/2022	EPI_ISL_10680070	2022-02-27	Omicron BA.2
hCoV-19/France/BRE-CHUP-8600827992/2022	EPI_ISL_10694901	2022-02-18	Omicron BA.2
hCoV-19/Denmark/DCGC-392185/2022	EPI_ISL_10506665	2022-02-14	Omicron BA.3
hCoV-19/England/PHEC-5U0BCZD1/2022	EPI_ISL_10327378	2022-01-22	Omicron BA.3
hCoV-19/Denmark/DCGC-382270/2022	EPI_ISL_10216492	2022-02-15	Omicron BA.3
hCoV-19/Denmark/DCGC-362114/2022	EPI_ISL_9854919	2022-02-06	Omicron BA.3
hCoV-19/Poland/PZH-GUM-10091/2022	EPI_ISL_9817248	2022-01-20	Omicron BA.3

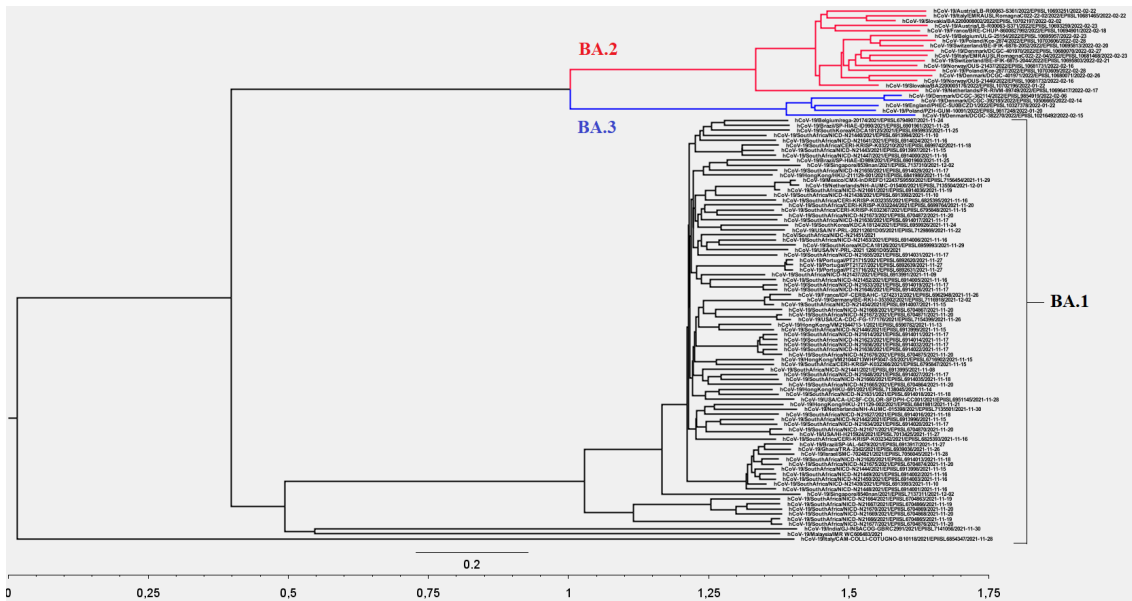


Figura S1.1. Análisis Bayesiano MCMC de las cepas BA.1, BA.2 y BA.3 de SARS-CoV-2 (Ómicron). Se muestra un árbol de máxima credibilidad de clados obtenido mediante el modelo GTR+I, un reloj molecular estricto y un modelo de población coalescente estructurado. El árbol está enraizado en el ancestro común más reciente (MRCA). Los clados de las cepas de Ómicron BA.1, BA.2 y BA.3 se muestran en negro, rojo y azul, respectivamente. El tiempo hasta el MRCA se indica en años en la parte inferior de la figura. La barra en la base del árbol denota el tiempo en años. Para una descripción completa del análisis y la figura original, consultar ⁶⁷.

Capítulo 2

Tabla S2.1: Lista de cebadores y su temperatura de fusión (T_m) en grados centígrados (°C).

Cebador (nombre)	Secuencia 5'-3'	T _m (°C)
Primers para RT-qPCR		
MAYV_qPCR_For	AGGACCCAGAGGAACACTAATA	60
MAYV_qPCR_Rev	CGATACTCTTTGCCCCACTTAC	60
IFN β _for	CAGCAATTTTCAGTGTCAGAAGC	60
IFN β _rev	TCATCCTGTCCTTGAGGCAGT	60
CCL5 (RANTES)_for	CTGTCATCCTCATTGCTACTGCC	60
CCL5 (RANTES)_rev	AGCACTTGCCACTGGTGTAGA	60
CXCL10 (IP-10)_for	TCAGCACCATGAATCAAA	60
CXCL10 (IP-10)_rev	GCAGGTACAGCGTACAGT	60
OAS1_for	GATCTCAGAAATACCCCAGCCA	60
OAS1_rev	AGCTACCTCGGAAGCACCTT	60
GAPDH_for	GCAGGTACAGCGTACAGT	60
GAPDH_rev	CGCTCCTGGAGGATGGTGAT	60
Primers para PCR en tiempo final		
JW55-MAYV 1-21F	ATGGCGGGCAAGTGACACTTG	57
NotI_Rv	CGCTAATACATTCTGCAACG	57
MAYV 5856-5875F	CTAAGCCTCGGTACTCTGTG	57
MAYV 11382-11401_Rv	TTCTATGGTTAGTAGGTGCC	57
For_sgRNA2	CACACGTGAGCTTGCCTTAA	60
Rev_sgRNA2	GCGAAACCCCGTCTCTAATTTAT	60

Figura S2.1

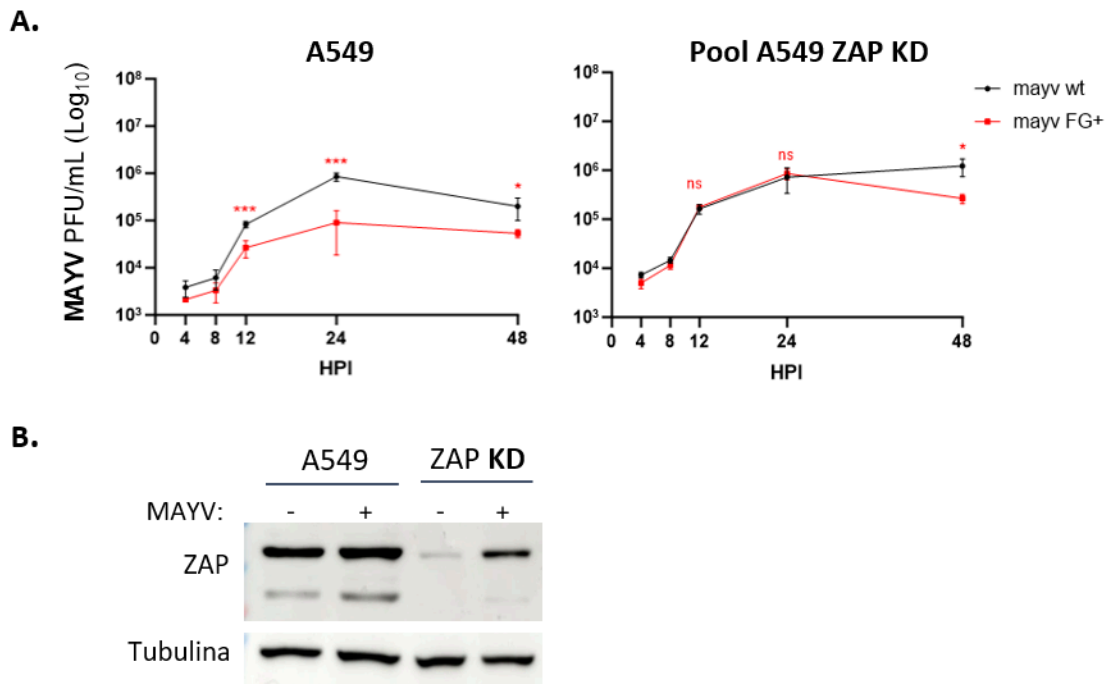


Figura S2.1: Cinética de replicación de MAYV WT y FG+ en el pool celular A549 ZAP KD **A)** Titulación viral por ensayo de placa a las 4, 8, 12, 24 y 48 HPI de MAYV WT y FG+ en células A549 (gráfico izquierdo) o pool A549 ZAP knock down (KD) (gráfico derecho) infectadas a MOI = 3. Los datos a cada tiempo se presentan como media \pm SEM (n = 3 experimentos independientes). ns, non significant; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; two-way ANOVA (comparando cepa FG+ versus WT MAYV en cada tiempo). **B)** WB de homogeneizados de células A549 o pool A549 ZAP KD infectadas con MAYV WT (+) o sin infectar (-); se utilizaron anticuerpos contra ZAP y tubulina.

Figura S2.2

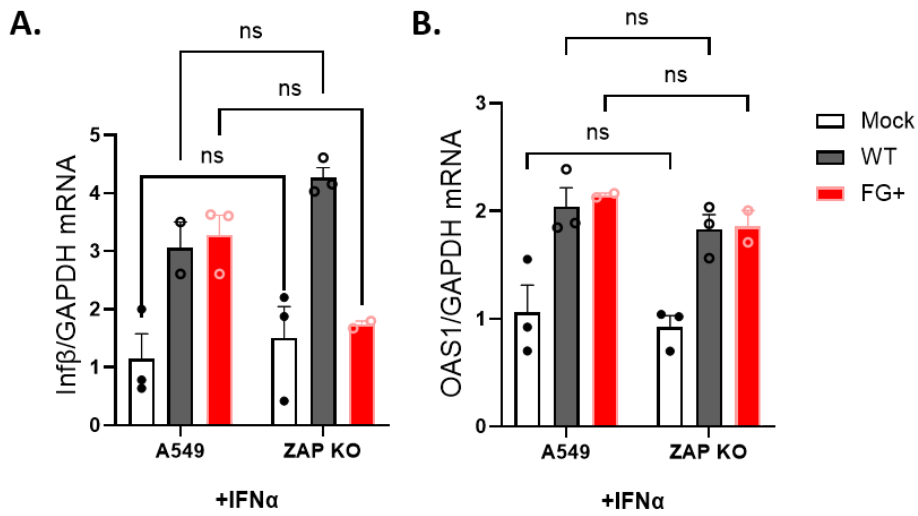


Figura S2.2: inducción de IFNβ y OAS1 en células pretratadas con IFNα. RT-qPCR para la cuantificación transcripcional relativa de la expresión de IFNβ (panel A) y OAS1 (panel B) en células A549 y A549 ZAP KO pretratadas con IFNα e infectadas con MAYV WT o FG+ (MOI=3, 24 HPI). Se graficó la media ± SEM (n =3 experimentos independientes). ns, non significant; *p ≤ 0.05; **p ≤ 0.01; ***p ≤ 0.001; two-way ANOVA (comparando cepa FG+ versus WT MAYV en cada tiempo).

Capítulo 3

Tabla S3.1: Lista de cebadores y su temperatura de fusión (T_m) en grados centígrados (°C).

Primers para PCR en tiempo final		
Cebador (nombre)	Secuencia 5'-3'	T_m (°C)
Fragment_for	ggggaccgctacatacACTTGTACGGC GGTCCTAAATAGG	72
Fragment_rev	gaccgccgtacaagtCATTCTAGAATT GTTTCTCCCCACCCT	72
Vector_for	actgtacggcggctcctaaatagg	66
Vector_rev	gtatgacgggtcccctcaatg	66
For_fusionHDV	AAGGCGCGCCGGCCGGCATG GTCCCAGC	57
Rev_fusionHDV	gtacaagtCGTCCCATTGCGCC	57
For_restofusion	TGGGACGactgtacggcg	52
Rev_restofusion	actttacgtacgggtgatgggacct	52
For_checkinsKO820	GTGTGGAATAGCGGACCCAT	60
Rev_checkinsKO820	TTTGCCTAAGGGCAAGCGTA	60

Tabla S3.2 Cobertura secuenciación por ONT de amplicones realizados para evaluar la estabilidad genómica del inserto en los virus MAYV_sgRNA2_miR (820 nt) y MAYV_sgRNA2_HDV (697 nt).

MAYV sgRNA2_miR				
	ARN Extracelular		ARN Intracelular	
Posición (nt)	48hpi_3_bc24	24hpi_3_bc25	48hpi_3_bc26	24hpi_3_bc28
4	0	0	0	1.12
12	19.25	187.25	54.5	158.88
20	63.62	686.75	178.25	530.25
28	68	723.75	186.88	565.62
36	69.12	727.5	187.25	572.88
44	70.75	739.5	188.88	583.62
52	72.38	752.25	191.38	592.38
60	74	763.75	192.75	600.5
68	76	779.5	196.12	620
76	76.88	785.38	198.12	633.62
84	78	800	203.12	642.12
92	78.38	817.38	208.62	659.75
100	81.38	827.88	219	677.88
108	81.12	824.75	218.25	676.5
116	81.75	840.25	222.5	689.5
124	81.88	844.38	223.25	691.75
132	82	849.25	224.5	697.75
140	83.25	850.75	225.12	700
148	84.88	856.88	226.5	709.38
156	84.75	867.5	229.5	713.88
164	86	871.88	229.75	718.62
172	86.12	881.75	230.62	726.25
180	87.5	897.62	233.62	735.38
188	88.62	903.62	235.62	737.5
196	88.88	916	239.38	744.5

Información suplementaria - capítulo 3

204	90.38	928.75	242	753.25
212	91.25	939.25	245.38	754.75
220	95.12	959.38	252	764.75
228	95.38	959.62	251.75	767.12
236	96.12	953.88	250.12	766.5
244	97.75	962.12	252.12	770.88
252	100.5	1001.5	258.5	795.88
260	100.88	1002.38	257.62	804
268	102.62	1032.38	264.62	815.5
276	102	1040.88	269.25	825.12
284	103.5	1039.5	271.5	810.62
292	103.88	1052.38	271.88	773.38
300	102.75	1057.12	272.75	768.38
308	102.38	1041.38	278.75	763.25
316	103.75	1037.12	277.62	750.38
324	107	1068.25	281.25	764.88
332	106.88	1078.75	282.38	762.5
340	108.12	1078.38	281.62	763.25
348	109.25	1079.5	281.5	761
356	106	1067.75	279.38	748.38
364	104	1053.75	273.38	733.62
372	104.38	1065.88	278.75	747.62
380	101.75	1064.5	279	746.25
388	96	1061.62	277.88	740.5
396	86.5	1051.88	276.5	740.88
404	87.38	1041.88	275.62	728.38
412	90.5	1058.62	278.25	735.88
420	89.88	1073.38	282.88	745
428	89.88	1077.88	282.38	741.75
436	91.12	1061.62	278	740.75
444	90.25	1063.38	274.62	736
452	88.88	1056.38	265.25	727.5

Información suplementaria - capítulo 3

460	91.25	1049.38	253	726.5
468	91.25	1052.62	248.5	718.38
476	91.62	1049.12	243.75	714.62
484	92.75	1059.5	238.12	696.88
492	91.38	1065.5	234.5	688.75
500	91	1033.75	177.25	634.75
508	90.12	1046.38	232.75	687.25
516	88	1051.25	232.25	691.38
524	89	1052.38	234.38	681.25
532	87	1025.12	228	664.25
540	84.75	1020.38	227.5	663.62
548	84	1025.88	228.25	658.88
556	84.62	1019.25	228.75	660.25
564	82.5	1007.88	227.12	660.88
572	77	983.25	219.75	647.62
580	76.75	987.25	219	644.62
588	74.62	960.75	209	625.12
596	74.25	949.75	205.25	619.5
604	72.88	946.38	204.25	615.88
612	71.5	947.88	212.62	614.62
620	71	940.88	213	636.38
628	70.38	929.5	208.75	623.38
636	79	921	205.12	614.12
644	80.38	909.12	200.62	607.88
652	79	881.62	196.5	597.88
660	78	865.75	186.75	584.38
668	78	850.88	182.38	572.88
676	77	839.75	177.75	559.62
684	75	833.12	177.12	556.5
692	73.5	822.5	176	550
700	71.38	812.75	172.75	541.5
708	71.38	808.75	173.25	536.12

716	69.62	803.38	171.5	531.75
724	68.5	793.38	171	514.25
732	68	787.88	168	507.38
740	68.12	785.5	163.88	505.75
748	67.75	779.5	161.38	498.75
756	67.38	767.25	160.12	493.38
764	67	761.5	155.88	488.75
772	63.62	742.5	146.5	465.25
780	62.88	727.38	143.5	449.38
788	61.38	706.75	137.38	441
796	55	602.75	120.38	385.25
804	37.62	404.62	82.38	268.12
812	1.38	13.75	3.38	6.38
818	0	0.25	0.25	0

MAYV sgRNA2_HDV							
Posición (nt)	ARN Extracelular					ARN Intracelular	
	24hpi_0_1_bc28	24hpi_1_bc29	24hpi_3_bc30	48hpi_0_1_bc31	48hpi_1_bc32	24hpi_3_bc29	24hpi_1_bc30
3	0	0	0	0	0	0.33	1
9	0.5	0	1.17	0.67	3.83	8.5	1
15	24.67	3.5	75.67	19.83	112.83	293.83	111.5
21	34.67	6.33	99.17	28.83	157	414.83	285.33
27	36	8	103	28.17	164	434.33	402.17
33	36	8	103.67	30	165	438.33	404.83
39	35.83	10.33	107.83	35.67	166.5	446.5	411.33
45	35.5	11	109.17	36.83	168.67	450.67	415.67
51	36.5	11	113	38.17	172.67	457.5	418.83
57	38	10.5	115	38.17	176.83	471.5	428.17
63	38.17	11	116.17	38.33	179.5	478.67	438.5
69	40	11	117.83	38.5	186	492.5	450.17

Información suplementaria - capítulo 3

75	40.5	11	119.33	40	188.83	496.5	455.33
81	41.67	11.17	122.5	41	195.17	507.67	463.83
87	42.33	12.5	132	42	197.5	517.33	478
93	43.33	13	133.5	43.33	202.33	527.17	486.67
99	44.67	13.83	138.67	44.67	211.5	548.17	505.17
105	44.83	13.67	138	45.5	211.67	546.5	504.33
111	44.5	13.67	140	45.17	215.33	554.33	510.83
117	45.83	14	141.67	45.83	218.5	562	517.83
123	45.83	14	141	45.67	220.5	564.83	520.83
129	47	13.83	143	47.67	224.67	580.67	530.5
135	47	14.5	143.83	47.5	225.67	583.67	538
141	46.5	15	145.33	47.83	226.17	587.5	546
147	47	15	145.83	48	229	593.17	549
153	45.83	15	149.17	48	230.17	599.67	554
159	47.67	14.83	151.17	48.5	232.33	609.67	561.67
165	48.83	14.83	153.83	49	233.33	615	566.17
171	50.33	15.83	155.83	49.33	236.17	622.33	572.67
177	50.67	16.67	161.67	50	237.5	637.17	582.67
183	50.5	16.83	162	51.17	240.5	638.67	586.33
189	50.83	16.5	162.5	52	240.33	648.83	590.83
195	52	17	165.5	52.33	242.33	665.67	600
201	52	17.83	170.33	53.33	244.67	676.67	607.5
207	52.17	18	173.83	54.83	247.83	691	620.5
213	52	18	175.67	55.17	247.67	703	634.17
219	52.83	18	183.33	56.33	252.33	726.83	655.33
225	52.67	17.83	183.67	56.5	252.5	729	650
231	53.5	18.33	182.83	56.83	255.17	728.83	655.33
237	53.17	16.83	182.5	56.5	251.17	722.5	650.5
243	54.33	17.17	181.33	57.67	254	727.67	666.17
249	57.33	18	189.67	59.33	260.83	753.67	689.83
255	58	18.67	195	60.5	260.33	767.17	704.17
261	57.67	18.83	194	63.17	255.67	766.67	711.67

Información suplementaria - capítulo 3

267	57.83	18	195.67	63.67	258.33	778	727.83
273	57.83	18.83	196.5	63.67	259	785.5	743.33
279	57.67	19	195.83	63.67	259.83	788.33	758.83
285	54	17.33	180	59	225.5	763.17	752.5
291	51.83	17.83	176.67	58	216.17	759	753
297	51.83	18	175	57.83	213.67	759.33	756.33
303	51.67	18.83	176	57.5	208.83	757	755.17
309	51.17	17.83	172.17	55.83	200.83	743.33	734.17
315	49.83	18.33	165.5	54.33	202.67	725.33	731.83
321	49.5	18	163.83	55.33	197.67	742.5	745.33
327	47.33	17	160.33	54	191.83	743.67	750.33
333	47	16.83	157.33	53.67	186.83	735.17	746.17
339	46	15.83	155.33	54	180.67	723.83	742.67
345	45	15	151	52.67	172.33	727	737.17
351	39	15	145.33	48.67	164.17	707.17	715.83
357	37.67	14.83	139	46.67	151.33	685	699.5
363	31.83	15	129.33	43.5	138.17	666.5	692.5
369	31	15	129.5	44	136.67	677.67	702.33
375	30.5	14.67	127.5	44	135.67	668.5	694.83
381	25.5	15	125.5	43.67	132.33	659.83	690.83
387	24.83	14.83	122.17	43.17	127.33	635	667
393	23.5	14.17	119	39.5	122.5	623.17	668.33
399	16.83	10.17	80	24.83	75.5	472.17	558
405	8	4.5	23	3	9.33	255.5	406.67
411	7.67	5	21.67	3	8.5	251	401.83
417	7	4.83	20.67	3	6.17	230.5	384
423	7	4.83	21	3	6	236.33	387.83
429	7	5	21.67	3.83	6	237	385.67
435	6.5	4.67	21.67	3.33	6	233	379.17
441	7	5	21.33	4	6	234.83	379
447	6.83	5	20.67	4	5.83	230.67	376.83
453	7	5	20.67	3.5	5	227.33	371.17

Información suplementaria - capítulo 3

459	6.83	5	21.33	4	6	235.17	372.33
465	6.33	5	22	4	6	240.5	383.17
471	6.83	5	23	4.5	6.17	241.83	387.33
477	7.83	5	27.33	7.5	14	255.5	400.33
483	13.67	7.33	46.33	14.33	40.17	278.67	418
489	22	7.67	59.67	19.17	65	347.67	453.5
495	25.5	9.33	71.83	23	84.33	407	490.67
501	27.33	10.33	78.5	24	101.17	423.33	497
507	31	11.5	103.5	30.5	130.5	487.33	541.33
513	30.5	12	104.83	30.67	131.83	490.5	537.5
519	35.5	14.83	118.33	39.67	150.67	538	571.5
525	37	15	123.67	44	159.5	553.33	580.67
531	38.67	16.17	128.5	45.17	168	567.83	590
537	39	17	130	45.67	172.17	569.83	591.17
543	39	17	128.67	45.83	172	567.83	582.5
549	37.67	17	129	45.5	173	563	571.17
555	37	17	128.5	45.83	172	555.5	561.5
561	37	17	128.67	46	172.67	554	558.83
567	36.33	17	128.67	45.83	172.5	550.67	553
573	35.17	16.67	126.67	45.83	172.83	549.83	550.67
579	34.67	16.83	126.67	46	171.83	546	544.67
585	34.5	17	126	45.67	170.83	544.67	540.5
591	35	16.83	125.5	45.33	169.33	539	535
597	34.83	17	125.67	45.33	169.5	534.5	525.33
603	35	17	125.17	46	165.67	536.83	524.17
609	34.33	16.5	124.67	45.67	164	526.67	514.17
615	34.83	17	124.33	45.33	163.33	527.67	511.67
621	34.83	17	123.17	44.17	160.67	520.33	504.33
627	34.83	16.83	120.67	43.67	161.33	513	499.5
633	34.83	17	120.17	43.33	159.5	507.17	490.83
639	33.5	17	118.33	43	152.33	501.17	485.33
645	31.33	17	112.67	41.83	149.17	486.83	463.17

651	30.17	16	104.17	41	141	460.67	440.67
657	28.67	15.83	101.5	40.5	138.17	449.83	432.83
663	26.83	16	98.83	39.83	130.83	435.17	420.17
669	24.67	15.33	93.17	39	121.17	421.17	391.67
675	21.5	13.17	82.5	37.17	96.5	354.83	332.17
681	17.67	11.83	66.5	30.67	73	274.33	269.5
687	0.67	0.5	7.17	2.83	8.33	28.5	24.67
693	0	0	0	0	0	1	0.33
696.5	0	0	0	0	0	1	0

Tabla S3.3: Métricas de control de calidad de la secuenciación Illumina para las bibliotecas de amplicones secuenciadas en la plataforma NovaSeq X Plus (PE150). **Rendimiento (Gb):** cantidad total de bases generadas en gigabases. **Número de lecturas (M):** total de fragmentos secuenciados en millones. **Q20 y Q30:** porcentaje de bases con una precisión de secuenciación del 99% y 99,9%, respectivamente. **Contenido GC:** proporción de bases guanina y citosina. **Datos efectivos:** porcentaje de datos de alta calidad remanentes tras el filtrado de adaptadores y lecturas erróneas. **Tasa de duplicación:** porcentaje de secuencias idénticas detectadas. Los términos Bruto y Limpio/Filtrado indican el estado de los datos antes y después del procesamiento de calidad, respectivamente.

Muestra	Rendimiento bruto (Gb)	Lecturas Brutas (M)	Q30 Bruto (%)	Q20 Bruto (%)	Contenido GC Bruto (%)
HDV 0.1_24 HPI	1.0155	6.77	92.48	97.77	49.23
HDV 0.1_48HPI	1.0274	6.85	92.33	97.77	49.36
HDV 3_24HPI	1.045	6.97	92.49	97.73	49.37
HDV 3_48HPI	1.7597	11.73	92.38	97.71	49.56
miR 0.1_24HPI	2.6565	17.71	92.76	97.87	49.35
miR 0.1_48HPI	1.1098	7.4	92.47	97.85	49.5
miR 1_24HPI	1.7769	11.85	92.51	97.77	49.49
miR 1_48HPI	1.8209	12.14	92.42	97.67	49.37
Mock	1.8988	12.66	92.43	97.74	49.26

Tabla S3.3. Continuación

Muestra	Rendimiento Limpio/Filtrado (Gb)	Lecturas Limpias (M)	Q30 Limpio (%)	Q20 Limpio (%)	Contenido GC Limpio (%)	Porcentaje de Datos Efectivos (%)
HDV 0.1_24HPI	0.993	6.65	93.23	98.17	49.06	97.79
HDV 0.1_48HPI	1.0	6.69	93.28	98.25	49.07	97.38
HDV 3_24HPI	1.015	6.81	93.45	98.22	49.08	97.17
HDV 3_48HPI	1.710	11.45	93.41	98.26	49.3	97.2
miR 0.1_24HPI	2.600	17.4	93.52	98.29	49.19	97.88
miR 0.1_48HPI	1.082	7.24	93.36	98.3	49.22	97.56
miR 1_24HPI	1.729	11.58	93.49	98.28	49.25	97.32
miR 1_48HPI	1.771	11.89	93.37	98.19	49.17	97.31
Mock	1.855	12.42	93.27	98.2	49.08	97.71

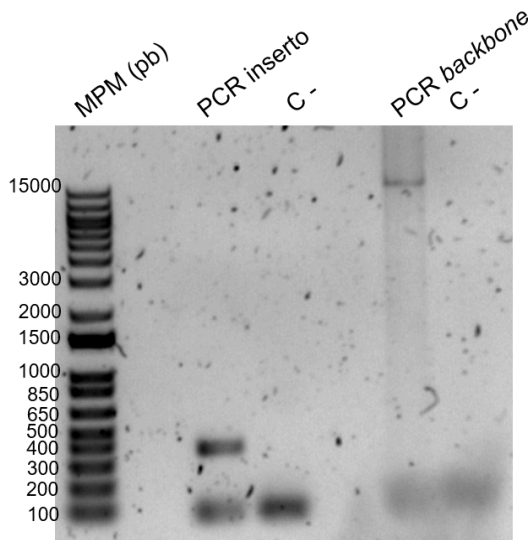


Figura S3.1. PCR para la amplificación del gBlock de 351 pb y el backbone MAYV WT. Análisis en gel de agarosa al 1% de los amplicones solapantes, obtenidos en las PCRs del gBlock de 351 pb (PCR inserto) y el backbone del clon infeccioso MAYV WT (PCR *backbone*) de 16417 pb. MPM, marcador de peso molecular; C -, control negativo sin ADN.

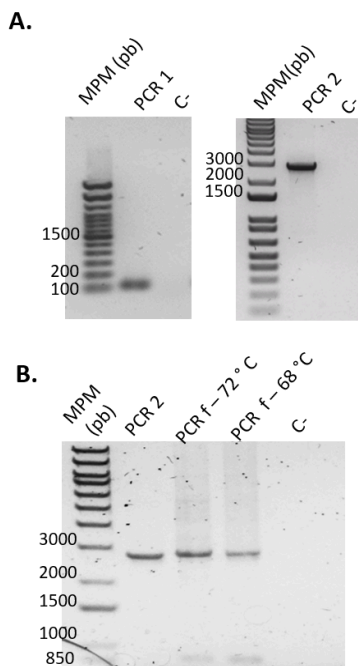


Figura S3.2. PCR de fusión para la construcción del clon infeccioso MAYV_sgRNA2_HDV. Análisis en gel de agarosa 1% de los productos de PCR en **A)** Amplicones obtenidos de las PCR 1 (103 pb) y PCR 2 (2658 pb) luego de la purificación. **B)** La PCR de fusión (PCR f) se realizó a partir de los amplicones mostrados en A) Se probaron dos temperaturas de fusión, 72 y 68 grados. MPM, marcador de peso molecular 1Kb plus (ThermoFisher™, USA). C-, control negativo sin ADN.

Producción científica asociada directamente a esta tesis

Publicaciones

- 3) **Paz M**, Aldunate F, Arce R, Ferreiro I, Cristina J. An evolutionary insight into Severe Acute Respiratory Syndrome Coronavirus 2 Omicron variant of concern. *Virus Res.* 2022;314:198753. [doi:10.1016/j.virusres.2022.198753](https://doi.org/10.1016/j.virusres.2022.198753)
- 4) **Paz M**, Moratorio G. Deep mutational scanning and CRISPR-engineered viruses: tools for evolutionary and functional genomics studies. *mSphere.* 2025;9(3):e00508-24. [doi:10.1128/msphere.00508-24](https://doi.org/10.1128/msphere.00508-24). (Review)
- 5) **Paz M**, Ferreiro I, Fajardo Á, Simón D, Moreno P, Moratorio G. CpG Recoding of Mayaro Virus Reveals Mechanistic Basis of Alphavirus Host Adaptation [manuscrito en preparación]. Montevideo: Institut Pasteur de Montevideo; 2026.

Presentaciones en congresos

1. "Role of the Zinc finger antiviral protein in the innate immune antiviral response against MAYV virus" – Poster presentation (2025), III Encuentro Uruguayo de Virología, Salto, Uruguay
2. "Delving into the Mechanism of Attenuation in CpG Dinucleotide-enriched Mayaro Virus" — Poster Presentation (2024), EMBL Symposium - Innate Immunity in Host-Pathogen Interactions, Germany
3. *Investigating the Mechanisms of CpG Dinucleotide-enriched Mayaro Virus Attenuation* — Poster Presentation (2023), Scientific Days at the Institut Pasteur Montevideo, Uruguay

Producción científica relacionada a la temática de la tesis

Publicaciones

1. Salazar C, Trovão NS, Ferrés I, **Paz M**, et al. Impact of timely vaccination and genomic surveillance on controlling consecutive waves of SARS-CoV-2 variants. *Sci Rep.* 2026. [doi:10.1038/s41598-026-48131-z](https://doi.org/10.1038/s41598-026-48131-z)
2. Salazar C, Ferrés I, **Paz M**, et al. Fast and cost-effective SARS-CoV-2 variant detection using Oxford Nanopore full-length spike gene sequencing. *Microb Genom.* 2023;9(5):mgen001013. [doi:10.1099/mgen.0.001013](https://doi.org/10.1099/mgen.0.001013)
3. **Paz M**, Rego N, Costábile A, et al. Real-Time Genomic Surveillance for SARS-CoV-2 Variants of Concern, Uruguay. *Emerg Infect Dis.* 2021;27(11):2957-2960. [doi:10.3201/eid2711.211198](https://doi.org/10.3201/eid2711.211198)
4. **Paz M**, Rego N, Salazar C, Costábile A, Fajardo A, Ferrés I, Perbolianachis P, Fernández-Calero T, Noya V, Machado MR, et al. Emergence and Spread of a B.1.1.28-Derived P.6 Lineage with Q675H and Q677H Spike Mutations in Uruguay. *Viruses.* 2021; 13(9):1801. doi.org/10.3390/v13091801

5. **Paz M**, Franco-Trecu V, Szteren D, et al. Understanding the emergence of highly pathogenic avian influenza A virus H5N1 in pinnipeds: An evolutionary approach. *Virus Res.* 2024;350:199472. [doi:10.1016/j.virusres.2024.199472](https://doi.org/10.1016/j.virusres.2024.199472)

Capítulos de libros:

1. **Paz, M.**, Moreno, P., & Moratorio, G. (2023). Perspective Chapter: Real-Time Genomic Surveillance for SARS-CoV-2 on Center Stage. IntechOpen. [doi: 10.5772/intechopen.107842](https://doi.org/10.5772/intechopen.107842). Escrito por invitación

Presentaciones en congresos

1. *“First tools for a Bio-smart City: Geolocalized Surveillance of Emerging Arboviruses in Uruguay”* Oral Presentation (2023), Workshop: Development of an Epidemiological Surveillance Network in South America, Uruguay
2. *“Genomic Surveillance of SARS-CoV-2 in Uruguay: An Interinstitutional Approach”* — Poster Presentation (2022), Biosciences Society Congress (SMB), Uruguay
3. *“Molecular Biology Applications for Diagnosis and Monitoring of SARS-CoV-2 Variants”* — Oral Presentation (2022), Interdisciplinary Congress: COVID-19, Pandemic and Post-pandemic, Uruguay



Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

An evolutionary insight into Severe Acute Respiratory Syndrome Coronavirus 2 Omicron variant of concern.

Mercedes Paz^{a,b}, Fabián Aldunate^{a,c}, Rodrigo Arce^{a,c}, Irene Ferreiro^{a,c}, Juan Cristina^{a,*}

^a Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la República, Iguá 4225, Montevideo 11400, Uruguay

^b Centro de Innovación en Vigilancia Epidemiológica, Institut Pasteur Montevideo, Mataojo 2020, Montevideo 11400, Uruguay

^c Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo, Mataojo 2020, Montevideo 11400, Uruguay

ARTICLE INFO

Keywords:
Coronavirus
evolution
SARS-CoV-2
Omicron
COVID-19

ABSTRACT

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel virus that belongs to the family *Coronaviridae*. This virus produces a respiratory illness known as coronavirus disease 2019 (COVID-19) and is to blame for the pandemic of COVID-19. Due to its massive circulation around the world and the capacity of mutation of this virus, genomic studies are much needed in order to reveal new variants of concern (VOCs). On November 26th, 2021, the WHO announced that a new SARS-CoV-2 VOC, named Omicron, had emerged. In order to get insight into the emergence, spread and evolution of Omicron SARS-CoV-2 variants, a comprehensive phylogenetic study was performed. The results of these studies revealed significant differences in codon usage among the S genes of SARS-CoV-2 VOCs Alfa, Beta, Gamma, Delta and Omicron, which can be linked to SARS-CoV-2 genotypes. Omicron variant did not evolve out of one of the early VOCs, but instead it belongs to a complete different genetic lineage from previous ones. Strains classified as Omicron variants evolved from ancestors that existed around May 15th, 2020, suggesting that this VOC may have been circulating undetected for a period of time until its emergence was observed in South Africa. A rate of evolution of 5.61×10^{-4} substitutions/site/year was found for Omicron strains enrolled in these analyses. The results of these studies demonstrate that S genes have suitable genetic information for clear assignment of emerging VOCs to its specific genotypes.

1. Introduction

In December, 2019, a pandemic of coronavirus disease (COVID-19) started in Wuhan, China (Li, Zai and Zhao, 2020a). This pandemic is caused by a virus known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the infection by this virus leads to a severe respiratory pneumonia (Gorbalenya et al., 2020). As December 14th, 2021, there have been more than 270 million confirmed cases worldwide and the global deaths of SARS-CoV-2 disease surpasses 5 million people (World Health Organization 2021a).

SARS-CoV-2 possess a single stranded, positive-sense RNA genome of approximately 30 kilobases in length, which encodes for multiple structural and non-structural proteins. The structural proteins include the spike (S) protein, the envelope (E) protein, the membrane (M) protein, and the nucleocapsid (N) protein (Chen et al., 2020).

The replication cycle of SARS-CoV-2 starts when it infects epithelial cells using the viral S protein to bind host angiotensin-converting

enzyme 2 (ACE2) and by this means fusing with cell membrane to gain cell entry (Hoffmann et al., 2020; Xia, 2021).

From the beginning of this pandemic several therapies and preventive health cares were developed, such as clinically applied monoclonal antibodies (Weinreich et al., 2021) or vaccinations (Wang et al., 2021), and both were successfully used to neutralize the virus. However, the emergence of variants of concern (VOCs) with substitutions in the S protein may make these therapies and vaccines to reduce its efficiency by escaping to anti-SARS-CoV-2 antibodies (Davies et al., 2021).

Previous studies have identified four VOCs currently circulating in the human population: VOC Alpha (B.1.1.7, first identified in the United Kingdom); VOC Beta (B.1.351, first identified in South Africa); VOC Gamma (P.1, first identified in Brazil) and VOC Delta (B.1.617.2, first isolated in India). By November 24th, 2021, a new SARS-CoV-2 VOC, now known as VOC Omicron (B.1.1.529) was first identified in South Africa (Wang & Chen, 2021).

The emergence of Omicron VOC raised a concern that this variant

* Correspondence: Dr. Juan Cristina. Laboratorio de Virología Molecular, Centro de Investigaciones Nucleares, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

E-mail address: cristina@cin.edu.uy (J. Cristina).

<https://doi.org/10.1016/j.virusres.2022.198753>

Received 5 January 2022; Received in revised form 18 March 2022; Accepted 20 March 2022

Available online 22 March 2022

0168-1702/© 2022 Published by Elsevier B.V.

may reduce the efficacy of the anti-SARS-CoV-2 induced-antibodies or be more transmissible (Callaway, 2021; Zhang et al., 2021).

In order to better understand the emergence, spread and evolution of Omicron SARS-CoV-2 variants, a comprehensive phylogenetic study was performed.

2. Material and Methods

2.1. Sequences

Available and comparable complete S gene sequences of 159 Omicron SARS-CoV-2 strains isolated from November 13th to December 2nd, 2021, in South Africa, Ghana, Singapore, Portugal, Netherlands, Mexico, Malaysia, Japan, Hong Kong, Germany, Canada, Botswana, Israel, Ireland, Belgium, Austria, Australia, South Korea, USA, Sri Lanka and Switzerland, were used throughout these studies. These sequences were aligned with corresponding sequences from 97 Alfa, Beta, Gamma and Delta SARS-CoV-2 strains isolated elsewhere. Sequences were obtained from the Global Initiative on Sharing Avian Influenza Data (GISAID) database. For accession numbers, country of origin and date of isolation, see Supplementary Material Table 1.

2.2. Sequence alignment

Sequences were aligned using MAFFT version 7 program (Katoh et al., 2019).

2.3. SARS-CoV-2 genotype assignment

In order to capture local and global patterns of virus genetic diversity in a timely and coherent manner, we employed Pangolin COVID-19 genetic lineage strain assignment (Rambaut, Holmes, O'Toole et al., 2020).

2.4. Data analysis

Nucleotide frequencies and codon usage of S proteins from SARS-CoV-2 variants were calculated using the program CodonW (written by John Peden) as implemented in the Galaxy server version 1.4.4 (Afgan et al., 2018). The relationship between compositional variables and samples was obtained using Principal Component Analysis (PCA). Singular value decomposition (SVD) method was used to calculate the PCA method. The unit variance was used as the scaling method. This means that all variables are scaled so that they will be equally important (variance = 1) when finding the components. By the same approach, Heatmaps were also constructed, which is a data matrix for visualizing values in the dataset by the use of a color gradient. Rows and/or columns of the matrix are clustered so that sets of rows or columns rather than individual ones can be interpreted. PCA and Heatmaps analysis were done using the ClustVis program (Metsalu and Vilo, 2015).

To reconstruct the evolutionary history of Omicron SARS-CoV-2 strains, a Bayesian Markov Chain Monte Carlo (MCMC) approach was used as implemented in the BEAST package v2.5.2 (Bouckaert et al., 2019). First, the evolutionary model that best fit the sequence dataset was determined using the IQ-TREE program (Trifinopoulos et al., 2016). Bayesian information criterion (BIC), Akaike information criterion (AIC), and the log of the likelihood (LnL) were used to identify the best model. Both strict and relaxed molecular clock models were used to test different dynamic models (constant population size, exponential population growth, Bayesian Skyline and Birth-Death Skyline Serial). Statistical uncertainty in the data was reflected by the 95 % highest probability density (HPD) values. Results were examined using the TRACER v1.7.2 program (available from <http://beast.bio.ed.ac.uk/Tracer>). Convergence was assessed by effective sample sizes (ESS) above 200. Models were compared by AICM from the likelihood output of each of the models. Maximum clade credibility trees were generated

by means of the use of the Tree Annotator program from the BEAST package. Visualization of the annotated trees was done using the FigTree program v1.4.4 (available at: <http://tree.bio.ed.ac.uk>).

3. Results

3.1. Trends in evolution across S proteins from SARS-CoV-2 VOCs variants

In order to gain insight into the trends of evolution of the S protein, codon usage frequencies of 256 S genes from SARS-CoV-2 strains belonging to VOCs Alpha, Beta, Gamma, Delta and Omicron were determined and PCA and Heatmap analysis were performed (for strains included in these analyses, see Supplementary Material Table 1). This analysis revealed significant differences in codon usage among the S genes of SARS-CoV-2 VOCs, which can be linked to SARS-CoV-2 genotypes (see Fig. 1A and B). Average linkage suggest that VOCs Alfa, Beta, Gamma and Delta have a closer genetic relation among themselves and a more distant genetic relation with Omicron variants (see Fig. 1B, upper part). Moreover, average linkage also shows that Omicron isolates are not identical and heterogeneity can be observed (see Fig. 1B).

In order to study if these trends of evolution of the VOC's S genes are due to differences in nucleotide composition, an analysis of nucleotide frequencies for first, second and third codon positions were established for the complete S genes from VOCs SARS-CoV-2 variants included in these analyses and PCA and Heatmap analysis was performed (see Fig. 2). A significant bias in nucleotide composition frequencies was found among VOCs S proteins.

3.2. Bayesian coalescent analysis of Omicron SARS-CoV-2 strains

In order to reconstruct the evolutionary history of Omicron SARS-CoV-2 population, a Bayesian MCMC approach was employed (Bouckaert et al., 2019) using 89 available and comparable full-length genomes by December 2nd, 2021, (for isolates included in these analyses see Supplementary Material Table 1). The results shown in Table 1 are the outcome of 20 million steps of the MCMC, using the HKY+I nucleotide model, a strict molecular clock and the Birth-Death Skyline Serial population model.

The results of these studies suggest that strains classified as Omicron variants evolved from ancestors that existed around May 15th, 2020. A mean rate of evolution of 5.61×10^{-4} substitutions/site/year (s/s/y) was found for Omicron strains enrolled in these analyses (95 % high probability density values of 3.051×10^{-4} to 9.014×10^{-4} s/s/y). This is in agreement with previous estimations of SARS-CoV-2 rate of evolution of SARS-CoV-2 populations (6.57×10^{-4} s/s/y, Castells et al., 2020; 7.80×10^{-4} s/s/y, Lai et al., 2020; 9.90×10^{-4} s/s/y, Nie et al., 2020; 3.0×10^{-4} s/s/y, Simmonds, 2020; 1.60×10^{-3} s/s/y, Bai et al., 2020; $1.19-1.31 \times 10^{-3}$ s/s/y, Li, Zai and Zhao, 2020b).

To study the phylogenetic relations among Omicron variants, maximum clade credibility trees were generated using software from the BEAST package (Bouckaert et al., 2019). The results of these studies are shown in Fig. 3.

Strains isolated in South Africa belong to two different genetic clades, revealing that Omicron variants have diversified into two distinct genetic groups in that country. Strains isolated in Malaysia and India conform another genetic group (see Fig. 3). Inside main cluster of South African strains, strains isolated in Singapore, Hong Kong, Europe, Brazil and the USA are observed, suggesting a rapid spread of this lineage in different continents.

4. Discussion

SARS-CoV-2 S protein plays a key role in virus biology, epidemiology and adaptation of virus to its human host. Moreover, almost all vaccine candidates against SARS-CoV-2 are based on the S protein (Xia, 2021).

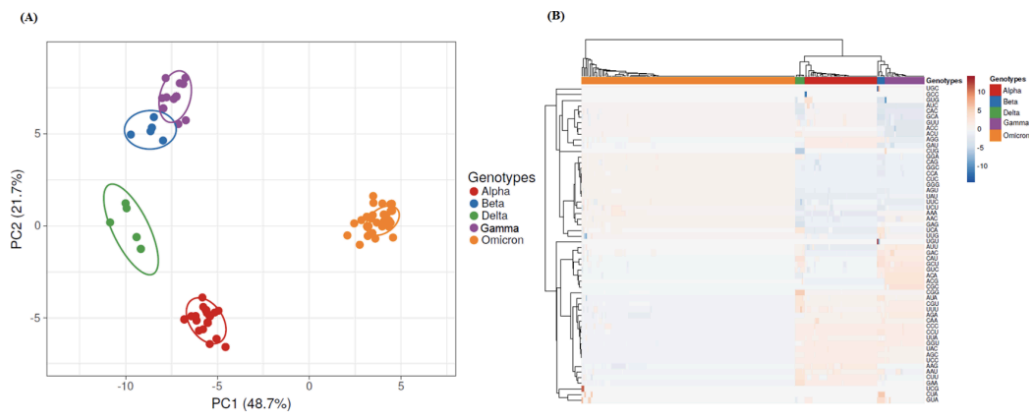


Fig. 1. PCA of codon usage in Spike proteins from SARS-CoV-2 VOCs strains. In (A) the position of the Spike proteins in the plane conformed by the first two major components of PCA is shown. SVD was used to calculate principal components and unit variance was applied. The proportion of variance explained by each axis is shown between parentheses. Prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse. Genotypes are indicated at the right of the figure. $N = 256$ data points. In (B) Heatmaps of codon usage in Spike proteins are shown. Unit variance scaling was applied. Each column corresponds to a different Spike protein from SARS-CoV-2 VOCs strains, who's genotype is shown in the upper part of the figure. Both rows and columns are clustered using correlation distance and average linkage.

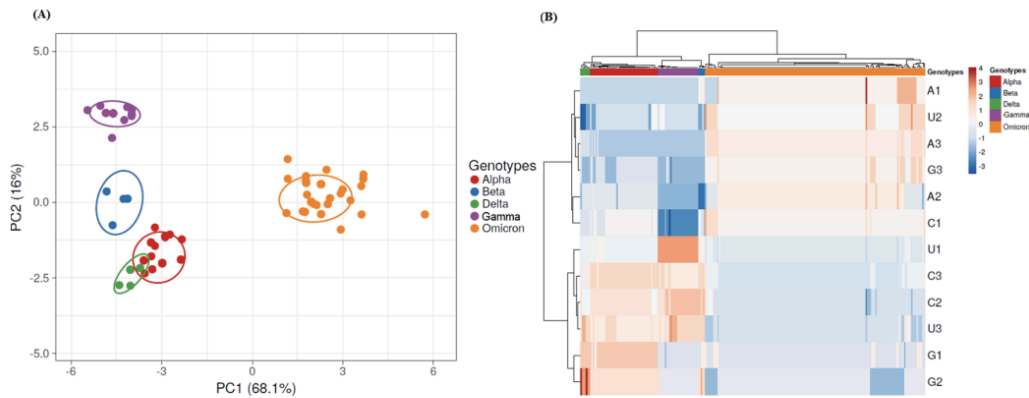


Fig. 2. PCA of nucleotide composition in Spike proteins from SARS-CoV-2 VOCs. In (A) a PCA analysis of nucleotide frequencies for first, second and third codon positions in S protein from SARS-CoV-2 strains is shown. SVD was used to calculate principal components and unit variance was applied. The proportion of variance explained by each axis shown between parentheses. Prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse. Genotypes are indicated at the right of the figure. $N = 256$ data points. In (B) Heatmaps of nucleotide frequencies in Spike proteins are shown. Frequencies for A, C, U and G at first, second and third codon positions are indicated 1 through 3. Unit variance scaling was applied. Each column corresponds to a different Spike protein from SARS-CoV-2 VOCs strains, who's genotype is shown in the upper part of the figure. Both rows and columns are clustered using correlation distance and average linkage.

Phyldynamic analyses can be extremely useful to study viral adaptation, particularly since new SARS-CoV-2 VOCs have emerged.

On November 26th, 2021, the WHO announced that a new SARS-CoV-2 VOC, named Omicron (initially named B.1.1.529), appeared to be increasing in almost all of South Africa's provinces, particularly Gauteng (World Health Organization, 2021b). Just three days after the announcement, cases of VOC Omicron have been detected in several other countries, including Austria, Australia, Belgium, Brazil, Canada, Czech Republic, Denmark, France, Germany, Italy, the Netherlands and the United Kingdom.

Omicron has some deletions and more than 30 substitutions in S protein, several of which (e.g., 69–70del, T95I, G142D/143–145del, K417N, T478K, N501Y, N655Y, N679K, and P681H) overlap with those in alpha, beta, gamma, or delta VOCs (GISAID, 2021). Nevertheless, Omicron also has specific S protein substitutions and insertions, like A67V, L212I, ins214EPE, G339D, S371P, S375F, N440K, N466S, S477N, Q493R, Y505H, T547K, N679K, N764K, D796Y, N856K, Q954H, N969K and L981F (GISAID, 2021).

In these studies, PCA and Heatmap analysis revealed correlation among codon usage and genotypes in the S protein from VOC's strains (see Fig. 1). These results demonstrate that S genes have suitable genetic information for clear assignment of emerging VOCs to its specific genotypes (Fig. 1). These differences are also related to biases in nucleotide composition among VOC's strains (Fig. 2). These findings highlight the latent diversity of SARS-CoV-2 that has yet to be fully explored. This is also in agreement with similar results found in other members of the family *Coronaviridae* (Kumar et al., 2021).

Bayesian coalescent analysis revealed that Omicron strains evolved from ancestors that existed by May 15th, 2020. This result revealed that this VOC may have been circulating undetected for a period of time until its emergence was observed in South Africa (Petersen et al., 2021). This is in agreement with recent results suggesting that Omicron variants have been circulating much longer than previously anticipated (Kandee et al., 2021). The earliest known case of Omicron in South Africa was a patient diagnosed with COVID-19 on Nov 9, 2021 (Karim & Karim, 2021). The results of these studies revealed that South African's

Table 1
Bayesian coalescent inference of Omicron SARS-CoV-2 strains.

Value ^a	HPD ^c	ESS ^d
-41527.05	-41604.75 to -41432.15	214.50
318.78	255.70 to 397.14	209.40
-41845.83	-41863.00 to -41830.62	314.10
5.613×10^{-4}	3.051×10^{-4} to 9.014×10^{-4}	208.10
1.539	0.806 to 2.251	210.70
May 15th, 2020		
0.310	0.120 to 0.570	
October 10th, 2021		

^aSee Supplementary Material Table 1 for strains included in this analysis.

^bThe rate of evolution is indicated in substitutions/site/year.

^ctMRCA, time of the most common recent ancestor is shown in years.

^dtMRCA South Africa, time to the most recent common ancestor for all strains isolated in South Africa. The date estimated for the tMRCAs are indicated in bold.

^b In all cases, the mean values are shown.

^c HPD, high probability density values.

^d ESS, effective sample size.

Omicron strains evolved from ancestors that circulated around October 10th, 2021, just before it emerged in the South African population (see Table 1). Nevertheless, strains isolated at Singapore and Hong Kong can be traced to one of the main genetic lineages of Omicron, suggesting that although Omicron strains were first detected in South Africa, they may have also been circulating in other regions of the world undetected (see Fig. 3). Besides, Omicron strains isolated in South Africa can be assigned to two different genetic groups (see Fig. 3). This is in agreement with recent results suggesting that Omicron strains diversified in two

different sub-lineages in that country (Wang & Chen, 2021).

The results of these studies revealed that Omicron variant did not evolve out of one of the early VOC's. Instead, it belongs to a complete different genetic lineage from previous VOC's. These results highlight the possibility of Omicron (or other VOCs currently unknown) may change the course of the pandemic. Whether the virus could have been circulating undetected in countries with little surveillance and sequences capacities is currently unknown (Kupferschmid, 2021). More studies will be needed to address these important questions.

While this manuscript was in the review process, very recent studies revealed that Omicron SARS-CoV-2 viruses have diversified in three different genetic lineages: BA.1, composed by the first emerging viruses isolated in South Africa, Botswana, and elsewhere and being enrolled in the analyses shown in this work; and two more genetic lineages, BA.2, which have been remained minoritarian in several countries although its incidence became predominant in Demark (Houhamdi et al., 2022) and BA.3 (Colson et al., 2022). In order to gain insight into this recent evolutionary process, the same strains enrolled in these studies were aligned with comparable BA.2 and BA.3 strains recently isolated in different countries (for strains, accession numbers, date and country of isolation, see Supplementary Material Table 2). Then, the same Bayesian MCMC approach depicted in these studies were performed and their results are shown in Supplementary Material Fig. 1. As it can be seen in the figure, the same conclusions about the origin and phylogenetic relations of BA.1 strains were obtained. Moreover, BA.2 and BA.3 strains belong to different genetic lineages in agreement with recent results (Colson et al., 2022). More studies will be needed in order to establish the origin and evolution of BA.2 and BA.3 strains.

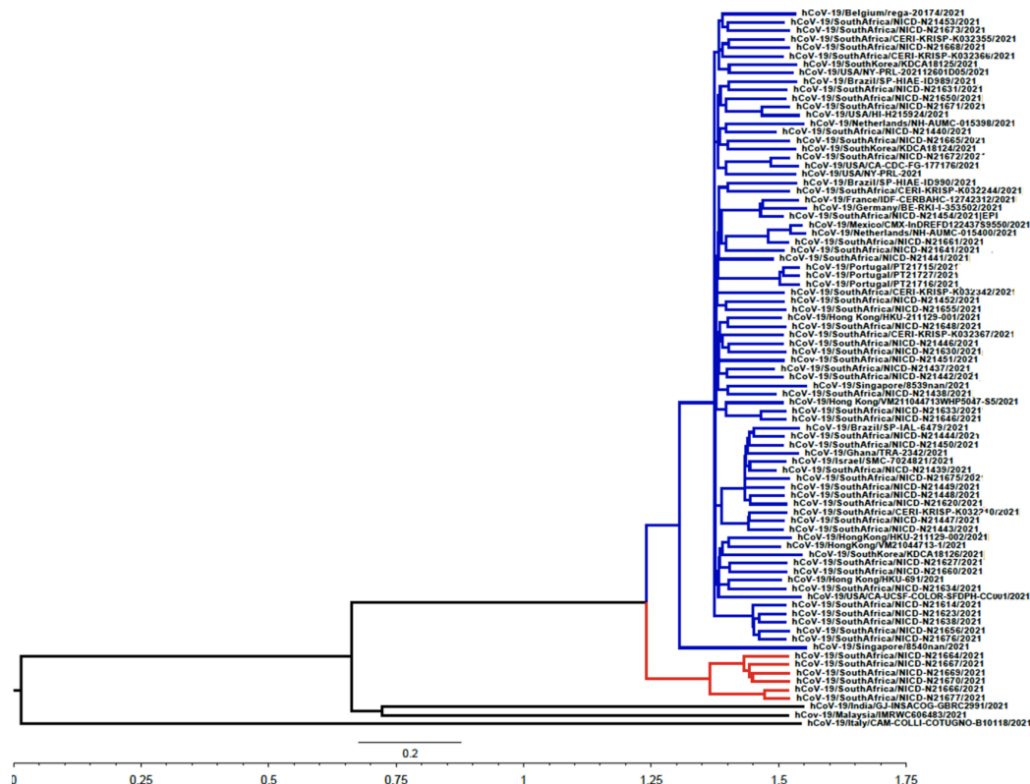


Fig. 3. Bayesian MCMC phylogenetic tree analysis of Omicron SARS-CoV-2 strains. A maximum clade credibility tree obtained using the HKY+I model, a strict molecular clock and the Birth-Death Skyline Serial population model is shown. The tree is rooted to the Most Recent Common Ancestor (MRCA). The two main clades containing strains isolated in South Africa are shown in blue and red, respectively. Time to the MRCA is shown in years at the bottom of the figure. Bar at the bottom of the tree denotes time in years.

5. Conclusions

The results of these studies revealed significant differences in codon usage among the S genes of SARS-CoV-2 VOCs, which can be linked to SARS-CoV-2 genotypes. These results demonstrate that S genes have suitable genetic information for clear assignment of emerging VOCs to its specific genotypes. VOCs Alfa, Beta, Gamma and Delta have a closer genetic relation among themselves and a more distant genetic relation with Omicron variants. Omicron variants did not evolve out of one of the early VOCs, but instead it belongs to a complete different genetic lineage from previous VOCs. A significant bias in nucleotide composition frequencies was found among VOCs S proteins. Omicron variants evolved from ancestors that existed around May 15th, 2020 suggesting that this VOC may have been circulating undetected for a period of time until its emergence was observed in South Africa. Strains isolated in South Africa belong to two different genetic clades, suggesting that Omicron variants have diversified into two distinct genetic groups in that country.

Author individual contributions: Mercedes Paz, Fabián Aldunate, Rodrigo Arce, Irene Ferreira were involved in the research, data curation and visualization of the results obtained. Juan Cristina was involved in conceptualization, methodology, data curation and wrote the original draft of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by Agencia Nacional de Investigación e Innovación, PEDECIBA and Comisión Sectorial de Investigación Científica (Grupos I+D grant), Universidad de la República, Uruguay.

CRedit authorship contribution statement

Mercedes Paz: Data curation, Visualization, Investigation. **Fabián Aldunate:** Data curation, Visualization, Investigation. **Rodrigo Arce:** Data curation, Visualization, Investigation. **Irene Ferreira:** Data curation, Visualization, Investigation. **Juan Cristina:** Conceptualization, Methodology, Data curation, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was funded by Agencia Nacional de Investigación e Innovación and PEDECIBA, Uruguay. We acknowledge Comisión Sectorial de Investigación Científica, Universidad de la República, Uruguay, for support through Grupos I + D grant. We gratefully acknowledge the Originating and Submitting Laboratories for sharing newly identified coronavirus sequences through GISAID. We thank Drs. Pilar Moreno and Gonzalo Moratorio for critical reading and support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.virusres.2022.198753](https://doi.org/10.1016/j.virusres.2022.198753).

References

- Afgan, E., Baker, D., Batut, B., et al., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46, W537–W544. <https://doi.org/10.1093/nar/gky379>.
- Bai, Y., Jiang, D., Lon, J.R., et al., 2020. Comprehensive evolution and molecular characteristics of a large number of SARS-CoV-2 genomes reveal its epidemic trends. *Int. J. Infect. Dis.* 100, 164–173. <https://doi.org/10.1016/j.ijid.2020.08.066>.

- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al., 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15, e1006650. <https://doi.org/10.1371/journal.pcbi.1006650>.
- Callaway, E., 2021. Beyond Omicron: what's next for COVID's viral evolution. *Nature* 600, 204–207. <https://doi.org/10.1038/d41586-021-03619-8>.
- Castells, M., Lopez-Tort, F., Colina, R., Cristina, J., 2020. Evidence of increasing diversification of emerging Severe Acute Respiratory Syndrome Coronavirus 2 strains. *J. Med. Virol.* 92, 2165–2172. <https://doi.org/10.1002/jmv.26018>.
- Chen, Y., Liu, Q., Guo, D., 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. *J. Med. Virol.* 92, 418–423. <https://doi.org/10.1002/jmv.25681>.
- Colson, P., Delerue, J., Beye, M., et al., 2022. First cases of infection with the 21L/BA.2 Omicron variant in Marseille, France. *J. Med. Virol.* <https://doi.org/10.1002/jmv.27695>, 2022 doiEpub ahead of print.
- Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372. <https://doi.org/10.1126/science.abc3055>.
- GISAID, 2021. Tracking of variants. <https://www.gisaid.org/hcov19-variants/> (accessed December 12th, 2021).
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., et al., 2020. Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the Coronavirus Study Group. *bioRxiv*. <https://doi.org/10.1101/2020.02.07.937862>.
- Hoffmann, M., Kleine-Weber, H., Pöhlmann, S., 2020. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol. Cell.* 78, 779–784. <https://doi.org/10.1016/j.molcel.2020.04.022> e5.
- Houhamdi, L., Gautret, P., Hoang, V.T., Fournier, P.E., Colson, P., Raoult, D., 2022. Characteristics of the first 1119 SARS-CoV-2 Omicron variant cases, in Marseille, France, November-December 2021. *J. Med. Virol.* <https://doi.org/10.1002/jmv.27613>. Epub ahead of print.
- Kandeel, M., Mohamed, MEM, Abd El-Lateef, HM, Venugopala, KN, El-Beltagi, HS, 2021. Omicron variant genome evolution and phylogenetics. *J. Med. Virol.* 2021 Dec 10. <https://doi.org/10.1002/jmv.27515>.
- Karim, S., Karim, QA, 2021. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet* 398, 2126–2128. [https://doi.org/10.1016/S0140-6736\(21\)02758-6](https://doi.org/10.1016/S0140-6736(21)02758-6).
- Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* 4, 1160. <https://doi.org/10.1093/bib/bbx108>. –116.
- Kupferschmidt, K., 2021. Where did “weird” Omicron come from? *Science* 374, 1179. <https://doi.org/10.1126/science.acx9738>.
- Kumar, N., Kaushik, R., Tennakoon, C., Uversky, V.N., Mishra, A., Sood, R., Srivastava, P., Tripathi, M., Zhang, K.Y.J., Bhatia, S., 2021. Evolutionary Signatures Governing the Codon Usage Bias in Coronaviruses and Their Implications for Viruses Infecting Various Bat Species. *Viruses* 13, 1847. <https://doi.org/10.3390/v13091847>.
- Li, Q., Guan, X., Wu, P., et al., 2020a. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 26, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>.
- Li, X., Zai, J., Zhao, Q., et al., 2020b. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J. Med. Virol.* 92, 602–611. <https://doi.org/10.1002/jmv.25731>.
- Metsalu, T., Vilo, J., 2015. Clustvis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research* 43 (W1), W566–W570. <https://doi.org/10.1093/nar/gkv468>.
- Nie, Q., Li, X., Chen, W., Liu, D., Chen, Y., Li, H., Li, D., Tian, M., Tan, W., Zai, J., 2020. Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Res* 287, 198098. <https://doi.org/10.1016/j.virusres.2020.198098>, 2020.
- Petersen, E., Ntoumi, F., Hui, DS, Abubakar, A., Kramer, LD, et al., 2021. Emergence of new SARS-CoV-2 Variant of Concern Omicron (B.1.1.529) - highlights Africa's research capabilities, but exposes major knowledge gaps, inequities of vaccine distribution, inadequacies in global COVID-19 response and control efforts. *Int J Infect Dis* 114, 268–272. <https://doi.org/10.1016/j.ijid.2021.11.040>.
- Rambaut, A., Holmes, E.C., O'Toole, Á., et al., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 5, 1403–1407. <https://doi.org/10.1038/s41564-020-0770-5>.
- Simmonds, P., 2020. Rampant C→U Hypermutation in the Genomes of SARS-CoV-2 and Other Coronaviruses: Causes and Consequences for Their Short- and Long-Term Evolutionary Trajectories. *mSphere* 5, e00408–e00420. <https://doi.org/10.1128/mSphere.00408-20>.
- Wang, L., Cheng, G., 2021 Dec 12. Sequence analysis of the Emerging Sars-CoV-2 Variant Omicron in South Africa. *J. Med. Virol.* <https://doi.org/10.1002/jmv.27516>.
- Weinreich, D.M., Sivapalasingam, S., Norton, T., Ali, S., Gao, H., Bhore, R., Trial, L., 2021. REGN-COV2, a neutralizing antibody cocktail, in outpatients with covid-19. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2035002>.
- World Health Organization, 2021a. Coronavirus disease 2019 (COVID-19) Weekly epidemiological update on COVID-19 –14 December 2021 (accessed 12.12.2021). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed December 14th, 2021).
- World Health Organization, 2021b. WHO 2nd Global consultation on assessing the impact of SARS-CoV-2 variants of concern on public health interventions. <https://www.who.int/publications/m/item/2nd-global-consultation-on-assessing-the-impact-of-sars-cov-2-variants-of-concern-on-public-health-interventions> (accessed 26.11.2021).

Deep mutational scanning and CRISPR-engineered viruses: tools for evolutionary and functional genomics studies

Mercedes Paz,^{1,2} Gonzalo Moratorio^{1,2,3}

AUTHOR AFFILIATIONS See affiliation list on p. 11.

ABSTRACT Recent advancements in synthetic biology and sequencing technologies have revolutionized the ability to manipulate viral genomes with unparalleled precision. This review focuses on two powerful methodologies: deep mutational scanning and CRISPR-based genome editing, that enable comprehensive mutagenesis and detailed functional characterization of viral proteins. These approaches have significantly deepened our understanding of the molecular determinants driving viral evolution and adaptation. Furthermore, we discuss how these advances provide transformative insights for future vaccine development and therapeutic strategies.

KEYWORDS deep mutational scanning, CRISPR-engineered viruses, virus evolution

In our 2020 mSphere of Influence article (1), we unfold groundbreaking work in synthetic biology that revolutionized how we study viral genomes, introducing an approach to decode how genetic information drives function, fitness, and evolutionary trajectories. Among these, the seminal study by Coleman et al. (2) demonstrated how engineering codon pair bias in poliovirus genes created a synthetic virus with unchanged protein composition but altered nucleotide sequences. By incorporating host-unfavorable codon pairs, this virus showed reduced replication efficiency and attenuated virulence, while retaining robust immunogenicity, a milestone in designing rationally attenuated viruses.

Building on this foundation, subsequent studies expanded the scope (3–6), applying recoding strategies across diverse viral families. These efforts revealed the intricate links between genome architecture and viral fitness, paving the way for innovative approaches to studying viral evolution and developing vaccine candidates. More recently, researchers have reframed these concepts to engineer RNA viruses using genome-wide recoding strategies aimed at promoting the accumulation of the most detrimental mutations, stop codons, thereby steering viral progenies toward evolutionary “dead ends.” This approach, tested across a spectrum of RNA viruses, including human enteroviruses, influenza, chikungunya, and coronaviruses (7–9), highlights the potential of this strategy to develop live attenuated vaccines and investigate the viral capacity to buffer mutations, known as genetic robustness.

Technological advancements in computational and experimental virology have further catalyzed this progress. Cutting-edge tools now enable precise manipulation of viral genomes, encompassing genome-wide mutagenesis and targeted insertions or deletions to dissect the functional roles of proteins and sequence motifs. This Full Circle review builds upon the existing body of work on viral fitness, focusing specifically on protein-focused methods and their application in deep mutational scanning (DMS) and CRISPR-based engineered viruses. These methodologies are pivotal for dissecting viral mechanisms, developing novel therapeutics, and understanding viral evolution. Here, we highlight recent advances in these two transformative techniques, which represent the next chapter in our understanding of viral fitness. Taken together, these techniques

Editor Michael J. Imperiale, University of Michigan, Ann Arbor, Michigan, USA

Address correspondence to Gonzalo Moratorio, moratorio@pasteur.edu.uy.

The authors declare no conflict of interest.

See the companion article at <https://doi.org/10.1128/msphere.00179-20>.

Published 24 April 2025

Copyright © 2025 Paz and Moratorio. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

integrate viral genomes with protein function, showcasing the power of these tools in engineering viral genomes. The review will explore the principles behind DMS and CRISPR-based genome editing, discuss their applications in studying viral fitness, and highlight key findings from recent studies. We will also examine the challenges and future directions of these techniques in the context of viral research.

DMS IN RNA VIRUSES

Overview and applications

DMS is a powerful high-throughput technique that systematically investigates how genetic variation translates into phenotypic variation. It enables the creation of fitness maps that detail the effects of nearly all possible nucleotide substitutions within a viral genome or protein sequence. This approach has helped to uncover key molecular mechanisms underlying viral evolution, the roles of non-coding viral sequences, protein function, and resistance to antiviral therapies, among many other applications as summarized in Fig. 1.

Methodologically, a DMS experiment follows at least five key steps. It starts with the design and construction of a large library of sequence variants, which defines the target sequence under study. This sequence can range from an entire viral protein to a particular protein domain, such as active sites or regions involved in important molecular interactions, or even a full viral proteome. The mutational scope is also defined at this stage, including single-nucleotide or amino acid substitutions covering all 19 non-native residues per position, and codon insertions and deletions (known as indels). To generate the mutant libraries, various strategies have been employed. Error-prone PCR has been widely used due to its cost-effectiveness and scalability in high-throughput experiments for introducing random nucleotide substitutions (10–13). However, it is less suitable for experiments in which specific mutations are required and can also introduce biases in the mutation frequency depending on the polymerase used in the PCR (14). Over the past two decades, rapid advances in *de novo* DNA and RNA synthesis have made controlled mutagenesis approaches more accessible, for example, the use of a pool of oligos containing the desired mutations, ensuring precise amino acid or nucleotide substitutions (15, 16) or indels (17).

Following library construction, the mutant proteins must be expressed in a relevant experimental system which is chosen based on the biological question. Infectious virus systems, for example, are used to assess viral fitness by rescuing mutant viral genomes into live viruses and measuring replication efficiency in cells permissive to infections (18–21). Library screening tools used for high-throughput selection of desired traits based on the display of proteins or peptides on the cell surface, like yeast (22–24) or mammalian (16, 25, 26) display systems, and more recently lentiviral or vesicular stomatitis virus (VSV) based pseudoviruses, have been highly beneficial for studying envelope proteins. For example, it was used in studies of the spike (S) protein of SARS-CoV-2 and the hemagglutinin (HA) protein of H5 influenza, to characterize the effects of mutations in phenotypes critical to viral fitness, like viral cell entry, antibody neutralization, and receptor-binding assays (27, 28).

Once expressed, mutants must be filtered to identify beneficial, neutral, and deleterious mutations. For example, in experiments using live viruses, assessing the replicative fitness of mutants after a certain time post-infection or after several rounds of serial passages has been used for evaluating amino acid substitutions in the RNA-dependent RNA polymerase (RdRp) of influenza A virus (IAV) and dengue virus (DENV) (18, 19). Binding affinity assays evaluate how mutations in receptor-binding domains (RBDs) affect interactions with host receptors; for instance, SARS-CoV-2 RBD mutant S proteins are tested for binding strength to human angiotensin-converting enzyme 2 (ACE2) using labeled sorting methods (22, 29). Additionally, neutralization and immune-escape selection experiments can test how mutant envelope proteins influence viral evasion from antibodies, revealing key antigenic sites (23, 29, 30).

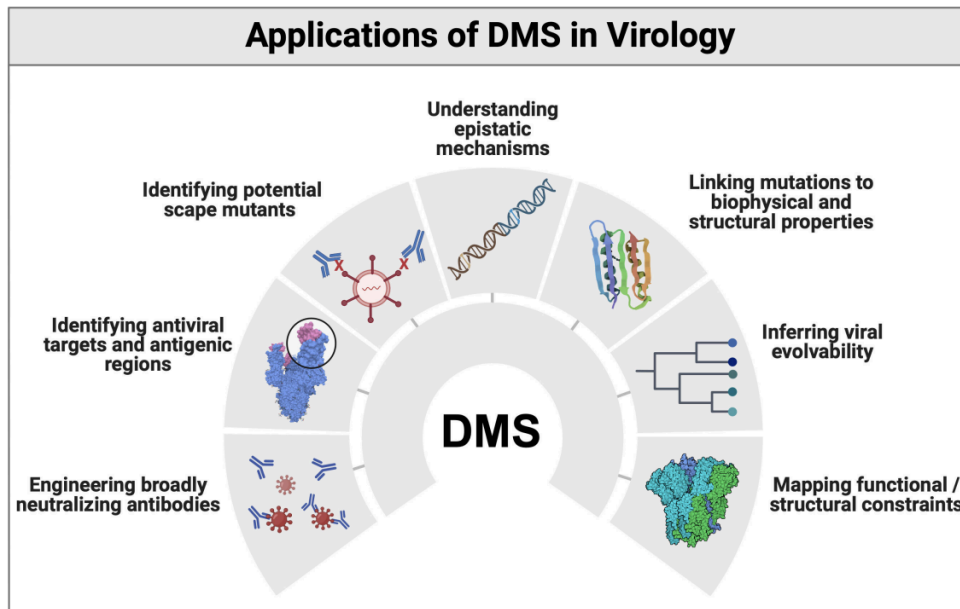


FIG 1 Summary of the main applications of deep mutational scanning in virology. Created in <https://BioRender.com>.

Following selection, next-generation sequencing is employed to quantify mutation frequencies before and after functional screening. DNA or RNA from the selected mutant populations is extracted and sequenced using high-throughput platforms. Computational analyses are then used to calculate fitness scores for each variant, categorizing them as deleterious, neutral, or beneficial based on their enrichment or depletion in the selected population (31).

The final step involves interpreting the mutational effects within a structural and evolutionary context. Mapping these changes onto known protein structures using molecular modeling, cryo-electron microscopy, or computational tools like AlphaFold helps predict how mutations alter protein stability and function (15, 16, 32).

Seminal papers published on DMS studies around 2010, originally focused on understanding protein sequence-function relationships. For example, the work by Fowler et al. (33) focused on WW domains, stable, triple-stranded beta-sheet motifs found in signaling and structural proteins that are characterized by their ability to bind proline-rich sequences and evaluated the ligand-binding performance of over 600,000 sequence variants in a single experiment. In another foundational study by Hietpas et al. (34), they developed an approach to systematically generate mutant libraries and measure fitness in a high-throughput manner. Precisely, they mapped the fitness effects of point mutations on a nine amino acid (AA) region of the yeast Hsp90 protein, a chaperone that mediates, directly or indirectly, the folding and activation of 10–15% of all proteins in yeast and humans. Another example of the applicability of this technique is the work of Starita et al. (35), where they used DMS to create a sequence-function map of the U-box domain of the murine E3 ligase Ube4b, characterizing mutations that enhance ubiquitin transfer activity. These works provided the first insights into the distribution and evolutionary weight of different types of mutations, established a tool for identifying mutations that enhance enzymatic activity for biotechnology applications, and demonstrated the potential of DMS in precision medicine, by elucidating the influence of human genome variations on disease susceptibility. As discussed in the following sections, DMS has been used extensively in virology to study viral protein function and

immune evasion mechanisms, to contribute to the development of vaccines and antiviral strategies, and to study viral evolution.

How does DMS shape our understanding of viral proteins?

DMS has greatly been used to map fitness landscapes and constraints on RNA viral proteins. It has helped to identify which AA sites are essential for molecular interactions or enzymatic activity and which ones are more tolerant to mutation, providing a detailed picture of protein function (10, 18, 19, 36–38). Historically, this analysis has been addressed by exploring the grade of sequence conservation of available deposited sequences (39–41). Although sequencing viral genomes from naturally isolated strains has become technically straightforward, only a limited number of viruses have sufficient sequence data available for comprehensive analysis. Furthermore, this approach has been proven to obscure critical residues within highly conserved regions (10, 42). Coupled with corroborative approaches combining experimental validation and computational modeling, DMS has advanced in understanding conserved functional and structural areas, as well as epistatic mechanisms in viral proteins.

Multiple studies have applied DMS to examine viral polymerases and structural proteins. For example, different groups have focused on uncovering functional constraints on the IAV polymerase subunits. Wu et al. (10) combined experimental fitness profiling across 94% of the polymerase A (PA) AA sequence with computational protein structure stability predictions, revealing high inter-type diversity, as numerous functional residues were non-conserved among each influenza virus (A or B) type. Additionally, this integrative methodology successfully distinguished residues critical for viral replication from those essential for protein stability, outperforming conventional approaches like dN/dS analysis (nonsynonymous/synonymous substitution ratio, a method in evolutionary biology to assess selection pressure) and highlighting the limitations of relying solely on sequence conservation or phylogenetic analyses.

Li et al. (19) used DMS to study the mutational landscape of the polymerase basic 1 (PB1) protein. They generated a library covering more than 90% of all possible AA substitutions and quantified their effects on replicative fitness, measured by a viral reconstitution system in mammalian cells. The study revealed that functional and structural constraints primarily mapped to residues involved in molecular interactions with RNA or proteins rather than broader conserved subdomains. In addition to capturing mutational effects on fitness, pairwise competition assays experimentally validated more than 10 mutations, highlighting regions critical for stability and polymerase function. This comprehensive approach demonstrated how DMS can delineate essential regions of RNA virus polymerases while correlating mutational tolerance with evolutionary diversity across natural influenza populations. Similarly, Suphatrakul et al. (18) conducted a DMS study of the nonstructural protein 5 (NS5) protein of DENV, focusing on its roles as an RNA-dependent RNA polymerase and methyltransferase. They performed a saturation mutagenesis experiment targeting nearly all single AA substitutions and assessed their effects on viral replication in mammalian cells. The resulting fitness maps demonstrated high mutational constraint in the conserved RdRp active site but unexpectedly revealed tolerance for select substitutions in universally conserved residues lining the priming loop. These findings suggest potential alternative roles for these residues in regulating polymerase activity. Further integration with structural and biochemical data highlighted the enzymatic versatility of NS5, as certain constrained positions influenced both RNA synthesis and host immune antagonism. This study underscored the importance of combining DMS data with structural modeling to identify residues critical for viral replication and therapeutic targeting. Moreover, recent research on the hepatitis B virus (HBV) polymerase unveiled the mechanism behind cis-preferential reverse transcription (38). They identified conserved prolines near the polymerase's termination codon that stall ribosomes, effectively tethering the nascent polymerase to its template RNA. Conse-

quently, this interaction ensures cis-preferential RNA packaging and facilitates reverse transcription of the HBV genome.

Structural proteins also play a crucial role in many aspects of viral fitness. They dictate the assembly, infectivity, host-tropism, and survival of viruses. DMS studies on the HA and neuraminidase (NA) proteins of IAV have revealed a clear distinction between antigenic and functional/structural regions (32, 43, 44). Antigenic regions, such as the HA globular head or surface-exposed regions of NA, show high mutational tolerance, enabling immune escape by allowing the virus to evade host immune responses. In contrast, functional domains, including the HA and NA catalytic sites, exhibit low mutational tolerance due to their critical roles in maintaining essential viral functions such as receptor binding and enzymatic activity. These findings highlight the evolutionary constraints imposed by the need to balance immune evasion with functional integrity.

The M1 matrix protein is the most conserved in IAV, leading to the historical assumption that it has a low tolerance for mutations. However, a recent study demonstrated that mutational tolerance varies across the protein, despite the low sequence diversity observed in natural isolates (42). The N-terminal third of M1 exhibited low mutational tolerance, likely due to its critical role in oligomerization and interactions with other viral components. In contrast, the C-terminal two-thirds were more tolerant of AA variations, which structural analysis suggests may result from its disordered configuration, reducing constraints on AA preferences. During the COVID-19 pandemic, multiple studies focused on performing DMS to explore the mutational fitness landscape of the SARS-CoV-2 S protein, shedding light on structural and functional constraints (16, 22, 45, 46). For example, Starr et al. (22) conducted a comprehensive DMS of the RBD of the S protein, measuring the effects of all possible AA substitutions on folding stability and ACE2 binding affinity, which is the main entry point into cells for some coronaviruses including SARS-CoV-2. Their findings showed that whereas most mutations reduced binding or compromised protein folding, some substitutions improved ACE2 binding, most notably, aligning with changes observed in circulating variants. Similarly, Ouyang et al. (16) examined the spike N-terminal domain (NTD) and quantified the impact of thousands of single mutations on protein expression levels, revealing that mutational tolerance was inversely correlated with proximity to interdomain interfaces. Their findings highlighted mutations that enhanced expression without altering antigenicity, offering insights for immunogen design.

How does DMS advance our understanding of viral evolution?

Over the past decade, DMS has also been instrumental in experimental evolution studies, identifying mutations linked to adaptation and shedding light on evolutionary mechanisms.

Several studies demonstrate the importance of epistasis, which is defined as the interactions between mutations in which one mutation's effect depends on others' presence. For example, this was done for studying SARS-CoV-2 evolution (29, 47–53), where compensatory mutations (e.g., at the interaction interface of the S protein RBD with the NTD) were demonstrated to reconfigure viral fitness. Taylor and Starr (48) showed that mutations at the RBD of the Omicron SARS-CoV-2 variants BQ.1.1 and XBB.1.5 can induce substantial epistatic effects, where earlier substitutions like N501Y reshaped the mutational fitness landscape, making previously deleterious mutations either neutral or beneficial. Using an infectious system of the whole S protein, Taminihi et al. (51) also revealed epistasis between RBD and NTD mutations, demonstrating compensatory relationships that stabilize overall S protein fitness while enabling immune evasion, such as structural compensation in the NTD for destabilizing RBD mutations observed in Omicron variants BA.1 and BA.2. Recently, epistatic interactions were observed in the Omicron BA.2.86 spike variant, where specific combinations of mutations (e.g., Q493E coupled with L455S and F456L) enhanced ACE2 binding, reversing the deleterious effect of Q493E in previous variants (49). These studies

consistently emphasize that epistasis plays an important role in modulating mutational constraints, either relaxing or tightening them based on the genetic and structural context, enabling viruses to navigate evolutionary trade-offs. Moreover, these findings also underscore the importance of investigating epistatic networks across RNA virus families to uncover conserved principles of viral evolution.

DMS has also been applied to investigate how changes in RNA viral proteins influence host tropism, the adaptation of a virus to infect different host species. The work by Thyagarajan and Bloom on the IAV HA protein (32) identified mutations that enable shifts in receptor-binding specificity, highlighting antigenic sites that enhance binding to human-like α 2-6-linked sialic acids, a critical determinant for allowing IAV to infect humans. More recently, a comprehensive proteome-wide DMS analysis of coxsackievirus B3 evaluated how over 40,000 nonsynonymous mutations and almost all possible codon deletions in the nonstructural proteins affected its fitness (15). The study revealed considerable variation in mutation tolerance both within and across individual proteins, which is associated with their structural and functional roles. It was found that while many viral proteins face strict functional constraints, nonstructural proteins especially display unique mutational tolerance patterns that may relate to mechanisms of adaptation specific to the host. In another proteome-wide DMS analysis of enterovirus A71, the virus was modified with around 51,000 indels and 41,000 AA substitutions. Studies experimentally examining the effects of indels on viral fitness are scarce. Still, most have indicated that indels are lethal to the virus and can only be tolerated at certain hotspots, which coincide with regions crucial for host recognition and immune interaction. Conversely, the viral proteome shows a higher tolerance for AA substitutions compared to indels, underscoring the evolutionary pressures exerted by this type of mutation and their possible influence on phenotypic diversity (17).

Regarding arboviruses, their dual-host cycle, which involves replication in both invertebrate and vertebrate hosts, imposes a strong purifying selection on the viral genomic determinants required for replication in evolutionarily distant hosts (54). This selective pressure makes it challenging to resolve which determinants are optimal for viral fitness in each host. In a work by Setoh et al. on Zika virus (ZIKV) (55), this was addressed by coupling DMS with fitness measurements across cell types from different organisms (e.g., mosquito cells vs human cells). They characterized conserved residues in the envelope protein's RBD, two substitutions in the ZIKV E protein, K316Q and S461G, that combined, strongly favored virus replication in mosquito cells. The mutant virus 316Q/461G containing these substitutions was found to be highly attenuated in mammalian cells, organoids, and mice, demonstrating host-specific trade-offs. Similarly, constrained mutational tolerance in critical catalytic sites of nonstructural proteins often links structural stability to host-specific replication efficiency (18). These studies shed light on evolutionary selection, where host-specific evolutionary pressures, whether structural, biochemical, or immune-related, tend to exploit residues or domains that balance adaptability with the preservation of essential functions. They also uncover conserved mutational constraints and hotspots that govern cross-species adaptability in RNA viruses.

How can DMS identify therapeutic targets?

As a powerful tool for systematically assessing the impact of mutations across viral proteins, several DMS studies have contributed insights that could be relevant to the development of different therapeutic approaches. For instance, mapping viral protein regions related to immune evasion and receptor binding may be possible to enhance the development of neutralizing antibodies (56–58). Moreover, highly conserved regions in viral proteins could be ideal targets for therapeutic and preventive strategies against RNA viruses. For example, these regions have the potential to be used for the development of broadly neutralizing antibodies that focus on conserved epitopes in envelope glycoproteins, to ensure cross-strain efficacy, such as those reported to target a discrete membrane-proximal anchor epitope of the HA stalk domain in IAV (59). In addition,

antivirals targeting critical enzymatic or functional sites in these conserved regions found by DMS experiments could have the potential to effectively inhibit viral replication, offering a promising alternative to the limited number of effective vaccines available to combat diverse RNA viruses (44).

A promising alternative approach to monoclonal antibodies and antivirals, which has capitalized on DMS studies identifying mutational rigid and conserved receptor-binding regions, involves the designing of soluble decoy receptors that mimic the virus-binding regions of host receptors. In a series of studies, a research group used deep mutagenesis of the ACE2 receptor to first engineer a decoy receptor sACE22.v2.4, with enhanced affinity for the SARS-CoV-2 S protein (60). The decoy receptor, containing three mutations (T27Y, L79T, and N330Y), showed a 35-fold increase in binding affinity and neutralized SARS-CoV-2 at levels comparable to high-affinity monoclonal antibodies. In the first study, the engineered receptor was demonstrated to be highly soluble, stable, dimeric, and catalytically active, with broad neutralization capabilities against both SARS-CoV-1 and SARS-CoV-2. In a second study (61), the researchers explored the potential for viral resistance to sACE22.v2.4 and found that no mutations within the RBD of SARS-CoV-2 redirected specificity toward the wild-type ACE2 receptor, suggesting that resistance would be rare. In another collaborative work, the *in vivo* efficacy of a derivative of the sACE22.v2.4 decoy fused with an IgG1 constant region (sACE22.v2.4-IgG1) was evaluated in K18-hACE2 mice (62). This work proved that sACE22.v2.4-IgG1 exhibited tight and persistent binding to SARS-CoV-2 and several variants of concern, and notably, that the decoy significantly reduced viral entry, lung vascular hyperpermeability, acute respiratory distress syndrome, and mortality in mice infected with both the WA-1/2020 and P.1 VOCs. These works highlight the promising therapeutic potential of an engineered decoy receptor *in vivo*. However, its clinical application is still being studied due to the possible short serum half-life, which may compromise its practical use. A recent study on the engineered decoy receptor sACE22.v2.4-IgG1 points out that sialylation is a vital factor in enhancing the *in vivo* efficacy of sACE2 decoy receptors and stresses the importance of glycosylation modifications to improve their pharmacokinetics (63).

Virus-driven crispr screens to identify host factors modulating viral replication

Overview

The pioneering work of Mojica et al. (64) and Barrangou et al. (65) demonstrated CRISPR-Cas9 as the adaptive immune mechanism of bacteria and archaea to fight foreign genetic material such as plasmids and phages. In prokaryotes, this system specifically recognizes and binds DNA through CRISPR RNA (crRNA), guiding Cas proteins to identify and cleave exogenous DNA with the assistance of trans-activating CRISPR RNA (tracrRNA). Subsequently, the groups of Charpentier and Doudna (66) combined crRNA and tracrRNA into a single-guide RNA (sgRNA), which more efficiently facilitated Cas9's genome-editing role *in vitro*. This groundbreaking discovery, which earned them the Nobel Prize in Chemistry in 2020, transformed our ability to investigate gene functions. In virology, the CRISPR-Cas9 system has widely been used to study virus-host interactions through whole-genome screening, using vector-based sgRNA libraries targeting the whole coding host genome, identifying host factors essential for virus replication (67–72).

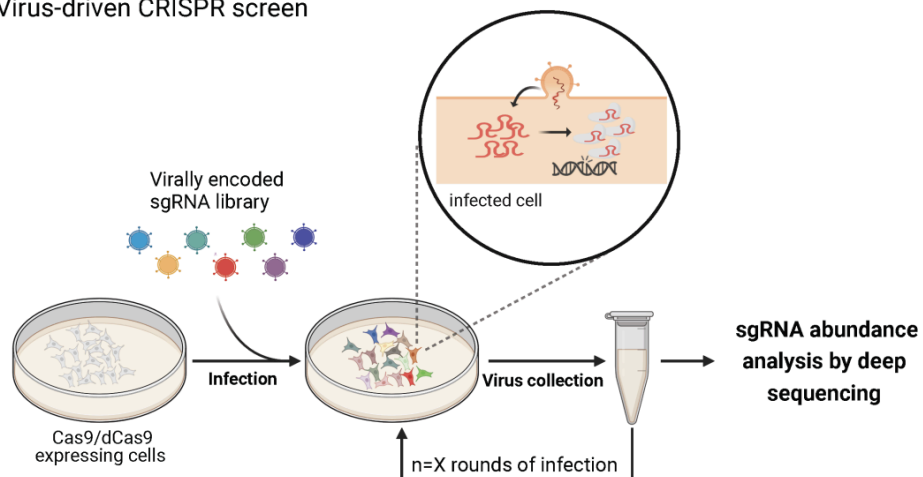
Until just a couple of years ago, most multiplexed pooled CRISPR screens in the context of infection were based on cell viability as their readout, leading to the discovery of cellular genes that allow cells to resist infection. Although extremely powerful, cell viability approaches predominantly reveal hits that block the early steps of viral replication, specifically entry, limiting the breadth of captured host factors. As reviewed in the following section, various groups have developed a novel screening system based on CRISPR-engineered replicating viral systems to expand the breadth of discoverable functional factors and increase sensitivity. In this platform, sgRNA libraries are expressed directly from the viral genome, enabling the direct follow-up of virus

propagation throughout the infection cycle. Upon infection of a cell expressing the Cas9 protein, the sgRNA encoded by the virus is released and alters the expression of cellular genes. This gene alteration affects the replication kinetics of the virus, where the sgRNA levels serve as a direct readout for viral propagation (Fig. 2A). In this way, sgRNAs that target genes that are essential or restrict virus propagation are expected to be underrepresented or enriched, respectively, in viruses budding from Cas9-expressing cells. Additionally, since thousands of viral genome copies are synthesized during each replication cycle, the signal is naturally amplified, making this screening approach highly sensitive and, potentially enabling the detection of even small positive and negative effects on viral propagation. This methodology surpasses conventional CRISPR screens by offering greater specificity, resolution, and mostly, the ability to profile later stages of the infection cycle, paving the way for novel antiviral strategies.

How can viruses be engineered to conduct CRISPR screens?

This novel approach was first applied in a study to identify interferon-stimulated genes against HIV (73). In this study, the researchers used lentiviral genomes encoding sgRNAs packaged into HIV virions, enabling the delivery and expression of Cas9 and sgRNA in target cells. Knockout cells were pooled, and deep sequencing of CRISPR RNAs in newly budded virions was used to assess HIV replication efficiency across thousands of genes in a heterogeneous population. The study identified antiviral factors like MxB, TRIM5α, IFITM1, and Tetherin, with sensitivity varying by viral strain. Additionally,

A. Virus-driven CRISPR screen



B. Modified alphavirus genome encoding sgRNAs

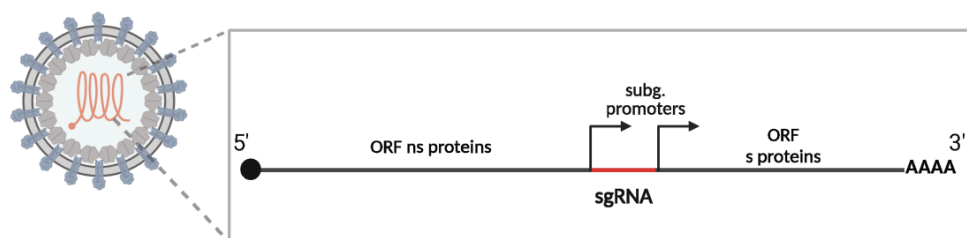


FIG 2 Virus-driven CRISPR screens. (A) Overview of a virus-driven CRISPR screen. (B) Potential strategy for the development of alphaviruses encoding sgRNAs. Open reading frame (ORF), nonstructural (ns) proteins, subgenomic promoters (subg. promoters), structural (S) proteins. Created in <https://BioRender.com>.

host dependency factors such as CD169, SEC62, and TLR2 were shown to support HIV replication in THP-1 cells. In another study on HIV (74), the researchers employed a library of over 1,500 replication-competent HIV-1 constructs, each expressing a single sgRNA targeting more than 500 cellular genes, to identify antiviral restriction factors across the entire HIV-1 replication cycle. Passaging in CD4⁺T cells enriched sgRNAs targeting several key genes, including GRN, CIITA, EHMT2, CEACAM3, CC2D1B, and RHOA, by over 50-fold. Additionally, the technique identified IFI16 as a restriction factor specifically targeted by the HIV-1 accessory gene *nef*. Functional validation in cell lines and primary CD4⁺T cells confirmed that these factors impact various stages of the HIV life cycle, including entry, transcription, release, and infectivity.

Recently, Finkel et al. (75) introduced virus-encoded CRISPR-based direct readout screening (VECOS), a system leveraging human cytomegalovirus (HCMV) engineered to express sgRNAs targeting the human coding genome. This was achieved by cloning the HCMV genome into a bacterial artificial chromosome system coupled with a Gateway cloning system to streamline the process of modifying the viral genome to encode a robust sgRNA library. Primary human foreskin fibroblasts (HFFs) were employed as the experimental model due to their permissiveness to viral replication. Sequential infections were performed by harvesting viral supernatants from infected HFFs and transferring them to fresh cultures to propagate the virus over multiple passages. To quantify changes in sgRNA abundance across passages and identify host factors critical for distinct stages of the HCMV infection cycle, the researchers extracted viral DNA after each sequential passage from cell lysates and culture supernatants. The abundance of sgRNAs encoded within the viral genome was then analyzed by PCR and high-throughput sequencing. This experimental approach demonstrated that RRM2, a key enzyme in dNTP production, is a critical host factor for HCMV replication, as its depletion severely impaired viral propagation and DNA synthesis. This finding opens up a new therapeutic potential for combining RRM2 inhibitors with existing anti-HCMV treatments, such as ganciclovir, to enhance antiviral efficacy. Importantly, the study demonstrated the unique strengths of VECOS in identifying a broader range of host factors, including those critical for non-lethal stages of infection, which are often missed by traditional survival-based CRISPR screens. They also provide evidence of the poorly understood functions of ARL6IP6 and its importance in viral propagation, demonstrating that ARL6IP6 plays a critical role in the late stages of the HCMV infection cycle, likely contributing to virus assembly through its association with ER-derived structures.

Similarly to the group above, King et al. (76) developed a novel fitness-based screen named transcriptional regulation by pathogen-programmed Cas9 (TRPPC), which uses IAV to deliver single guide RNAs (sgRNAs) that program CRISPR activation or CRISPR inhibition of cellular genes exploiting the virus's ability to express non-coding RNAs. They modified the influenza virus nonstructural (NS) gene, which normally encodes two proteins, nonstructural protein 1 (NS1) and the nuclear export protein, by altering its overlapping reading frames into non-overlapping ones, separated by an artificial intron. This created a "split NS" system that allowed the expression of non-coding RNAs without disrupting viral protein production or replication. They incorporated a microRNA (miRNA) encoded within this artificial intron, which would be processed by the host's Drosha nuclease in the nucleus, along with a sgRNA downstream of the miRNA-124, releasing the sgRNA from the viral genome. After several rounds of selection of the virus-encoded sgRNA library in A549 cells, the screen identified the protein TREX1, a cytoplasmic DNA exonuclease, as a key pro-viral factor, as it degrades cytoplasmic DNA to prevent immune activation, thus aiding influenza replication. This methodology also allowed the creation of a ranking of host factors that affect IAV viral replication, providing valuable insights into host-pathogen interactions.

Challenges and future directions

Despite their potential, the mentioned methodologies face challenges that create opportunities for refinement to address research gaps and limitations. A technique like

DMS produces a big volume of data, and the main challenge remains in analyzing and making sense of it to predict the effects of mutations. Machine learning offers a powerful approach for these methods that require the analysis of vast amounts of data and exhaustive experimental testing. For example, automated viral infectivity assays utilizing convolutional neural networks have been used to predict infection phenotypes from microscopy data, improving speed and precision in virus characterization, which could be applied for a more efficient analysis of viral variants (77–79). Moreover, machine learning has also been used to predict the functional outcomes of millions of untested mutations from DMS data sets, helping researchers to identify and filter promising variants for further study and reduce the need for exhaustive experimental testing (80–82). This approach could fill gaps in DMS by enabling real-time predictions of mutation effects in complex protein landscapes and expand the ability to explore larger mutational spaces with high accuracy and fewer laboratory resources. Another major challenge, shared by both CRISPR screening and DMS, is that they rely on cell culture as their experimental system, which can influence the interpretation of mutational effects. While these systems are used primarily for their ability to facilitate high-volume infections and vigorous viral replication, they may not fully capture the complexities of *in vivo* scenarios, where factors such as tissue-specific interactions and immune responses may play a significant role. For instance, research teams employing DMS on the same viral protein have reported inconsistent results across diverse cellular environments (42, 83). The authors discuss that discrepancies can largely be attributed to differences in viral strains, cell types, and experimental methodologies, but cannot rule out factors like population bottlenecks and epistatic interactions (42). Therefore, any conclusions drawn should be contextualized within the specifics of the viral strain and experimental setting.

Regarding DMS, although several studies have identified conserved functional regions in RNA virus proteins, such as the flavivirus NS5 or influenza PB1 polymerases (18, 19), systematic cross-family comparisons of mutational constraints are scarce. Current research mostly focuses on specific viral targets or homology-based analyses, lacking comprehensive data sets that consolidate DMS mappings across diverse RNA virus families to identify conserved functional or structural “hotspots.” This gap may largely arise from inconsistencies in experimental models and the lack of standardized fitness metrics across studies. Future research could expand DMS efforts to homologous proteins across major RNA viral families, targeting universal antiviral sites such as RdRp enzymatic cores and conserved glycoprotein domains. Such integration could inform the design of broad-spectrum inhibitors targeting residues less prone to mutational escape, paving the way for cross-family antiviral strategies, like the development of pan-viral vaccines to combat diverse RNA viruses. To support this idea, initiatives such as MaveDB (84), an open-source repository for sharing and analyzing DMS data sets, and the Atlas of Variant Effects Alliance (<https://www.varianteffect.org/>) have been established to foster collaboration and enhance the utility of mutational scanning research (85).

As part of this section discussing future directions, we hypothesize that alphaviruses, with their positive-sense, single-stranded RNA genomes that replicate in the cytoplasm, offer a distinct opportunity for exploration as a possible advancement in CRISPR/Cas9 technology. These viruses exhibit unique characteristics compared to previously employed viral systems. Specifically, the potential development of replication-competent Mayaro virus constructs encoding sgRNAs represents an innovative approach worth investigating. This idea is grounded in the capacity that alphaviruses have to tolerate the insertion of exogenous coding and non-coding sequences downstream of their nonstructural genes, expressed via a subgenomic promoter (Fig. 2B). To facilitate the release of non-coding RNAs like sgRNAs during replication, these sequences can be positioned downstream of a duplicated subgenomic promoter, followed by a self-excising ribozyme (e.g., HDV) or a miRNA sequence, as demonstrated in prior studies (76, 86). Nevertheless, RNA viruses face additional complexities, including high mutation rates and genomic instability, which increase the risk of evolutionary escape and may compromise embedded sgRNA functionality. Optimizing experimental designs and

methodologies will be essential to make CRISPR-based tools broadly applicable across multiple RNA viral families. In addition, given the recent development of this technology, establishing robust reproducibility standards is critical to ensure consistent and reliable outcomes across viral strains. Variability in experimental protocols, such as delivery methods, cell lines, and infection conditions, affects sgRNA targeting efficiency and the functional relevance of identified host factors. Addressing these issues will require standardized sgRNA libraries and multi-laboratory validation studies to benchmark findings and enhance reproducibility.

CONCLUDING REMARKS

In this Full Circle review, we highlight how DMS and CRISPR-engineered viral systems serve as essential tools in virology, enabling researchers to map viral fitness landscapes, identify drivers of evolution, and explore host-virus dynamics. DMS has revolutionized the study of viral proteins, allowing us to create fitness maps that detail the effects of nearly all possible nucleotide substitutions within a viral genome. Combining experimental techniques with computational tools such as modeling and machine learning has provided deeper insights into mutations, protein functions, and epistatic interactions to unveil viral evolution. Global collaboration is crucial, with open-access repositories like MakeDB enabling data sharing to address broader challenges like assessing mutational landscapes across viral families or identifying pan-antiviral targets. The newly developed strategy, virus-driven CRISPR screenings, focuses on the role of host factors in viral replication, uncovering novel host-viral mechanisms of interaction and revealing new therapeutic targets. It represents a significant advancement over earlier conventional CRISPR screening, addressing limitations such as profiling later stages of the infection cycle. As this technology develops, new experimental designs and methods will broaden the application across multiple viruses, revealing new insights into viral-host dynamics and offering promising leads for antiviral development.

Looking ahead, the convergence of high-throughput technologies, artificial intelligence, and viral genome engineering will unlock an era of unprecedented precision in virology. This multidisciplinary approach will not only accelerate our understanding of viral biology but will also lead to personalized, adaptive treatment regimens, representing a new frontier in the fight against viral infections.

ACKNOWLEDGMENTS

The authors thank Dr. Pilar Moreno for her thoughtful review of the manuscript and for providing invaluable feedback.

This work was supported by FOCEM COF 03/11, the G4 Program, the Institut Pasteur de Montevideo, and the Agencia Nacional de Investigación e Innovación.

AUTHOR AFFILIATIONS

¹Laboratory of Experimental Virus Evolution, Institut Pasteur de Montevideo, Montevideo, Uruguay

²Molecular Virology Laboratory, Faculty of Sciences, University of the Republic, Montevideo, Uruguay

³Center for Innovation in Epidemiological Surveillance, Institut Pasteur de Montevideo, Montevideo, Uruguay

AUTHOR ORCIDS

Mercedes Paz  <http://orcid.org/0000-0002-9700-0327>

Gonzalo Moratorio  <http://orcid.org/0000-0002-4555-4926>

AUTHOR CONTRIBUTIONS

Mercedes Paz, Writing – original draft, Writing – review and editing | Gonzalo Moratorio, Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review and editing

REFERENCES

- Moratorio G. 2020. mSphere of influence: experimental evolution of RNA viruses. *mSphere* 5:e00179–20. <https://doi.org/10.1128/mSphere.00179-20>
- Coleman JR, Papamichail D, Skiena S, Futcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 320:1784–1787. <https://doi.org/10.1126/science.1155761>
- 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. <https://doi.org/10.1016/j.chom.2012.10.008>
- Mueller S, Coleman JR, Papamichail D, Ward CB, Nimmual A, Futcher B, Skiena S, Wimmer E. 2010. Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* 28:723–726. <https://doi.org/10.1038/nbt.1636>
- Mueller S, Stauff CB, Kalkeri R, Koidei F, Kushnir A, Tasker S, Coleman JR. 2020. A codon-pair deoptimized live-attenuated vaccine against respiratory syncytial virus is immunogenic and efficacious in non-human primates. *Vaccine (Auckl)* 38:2943–2948. <https://doi.org/10.1016/j.vaccine.2020.02.056>
- Stauff CB, Yang C, Coleman JR, Boltz D, Chin C, Kushnir A, Mueller S. 2019. Live-attenuated H1N1 influenza vaccine candidate displays potent efficacy in mice and ferrets. *PLoS One* 14:e0223784. <https://doi.org/10.1371/journal.pone.0223784>
- Moratorio G, Henningson R, Barbezange C, Carrau L, Bordería AV, Blanc H, Beaucourt S, Poirier EZ, Vallet T, Boussier J, Mounce BC, Fontes M, Vignuzzi M. 2017. Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat Microbiol* 2:17088. <https://doi.org/10.1038/nmicrobiol.2017.88>
- Carrau L, Rezelj VV, Noval MG, Levi LI, Megrian D, Blanc H, Weger-Lucarelli J, Moratorio G, Stapleford KA, Vignuzzi M. 2019. Chikungunya virus vaccine candidates with decreased mutational robustness are attenuated *in vivo* and have compromised transmissibility. *J Virol* 93:00775–19. <https://doi.org/10.1128/JVI.00775-19>
- Schön J, Barut GT, Trüeb BS, Halwe NJ, Berenguer Veiga I, Kratzel A, Ulrich L, Kelly JN, Brügger M, Wylezich C, et al. 2024. A safe, effective and adaptable live-attenuated SARS-CoV-2 vaccine to reduce disease and transmission using one-to-stop genome modifications. *Nat Microbiol* 9:2099–2112. <https://doi.org/10.1038/s41564-024-01755-1>
- Wu NC, Olson CA, Du Y, Le S, Tran K, Remenyi R, Gong D, Al-Mawsawi LQ, Qi H, Wu T-T, Sun R. 2015. Functional constraint profiling of a viral protein reveals discordance of evolutionary conservation and functionality. *PLoS Genet* 11:e1005310. <https://doi.org/10.1371/journal.pgen.1005310>
- Gallo G, Conceicao C, Tsigotgi C, Willett B, Graham SC, Bailey D. 2021. Application of error-prone PCR to functionally probe the morbillivirus Haemagglutinin protein. *J Gen Virol* 102:001580. <https://doi.org/10.1099/jgv.0.001580>
- Gong D, Zhang T-H, Zhao D, Du Y, Chapa TJ, Shi Y, Wang L, Contreras D, Zeng G, Shi P-Y, Wu T-T, Arumugaswami V, Sun R. 2018. High-throughput fitness profiling of zika virus E protein reveals different roles for glycosylation during infection of mammalian and mosquito cells. *iScience* 1:97–111. <https://doi.org/10.1016/j.isci.2018.02.005>
- Ikegami S, Hashiguchi T, Hung C-T, Dobrindt K, Brennand KJ, Takeda M, Lee B. 2021. Fitness selection of hyperfusogenic measles virus F proteins associated with neuropathogenic phenotypes. *Proc Natl Acad Sci U S A* 118:e2026027118. <https://doi.org/10.1073/pnas.2026027118>
- Vanhercke T, Ampe C, Tirry L, Denolf P. 2005. Reducing mutational bias in random protein libraries. *Anal Biochem* 339:9–14. <https://doi.org/10.1016/j.ab.2004.11.032>
- Álvarez-Rodríguez B, Velandia-Álvarez S, Toft C, Geller R. 2024. Mapping mutational fitness effects across the coxsackievirus B3 proteome reveals distinct profiles of mutation tolerability. *PLoS Biol* 22:e3002709. <https://doi.org/10.1371/journal.pbio.3002709>
- Ouyang WO, Tan TJC, Lei R, Song G, Kieffer C, Andrabi R, Matreyek KA, Wu NC. 2022. Probing the biophysical constraints of SARS-CoV-2 spike N-terminal domain using deep mutational scanning. *Sci Adv* 8:eadd7221. <https://doi.org/10.1126/sciadv.add7221>
- Bakhache W, Symonds-Orr W, McCormick L, Dolan PT. 2025. Deep mutation, insertion and deletion scanning across the Enterovirus A proteome reveals constraints shaping viral evolution. *Nat Microbiol* 10:158–168. <https://doi.org/10.1038/s41564-024-01871-y>
- Suphatrakul A, Posiri P, Srisuk N, Nantachokchawapan R, Onnong S, Mongkolsapaya J, Siridechadilok B. 2023. Functional analysis of flavivirus replicase by deep mutational scanning of dengue NS5. *Microbiology*. <https://doi.org/10.1101/2023.03.07.531617>
- Li Y, Arcos S, Sabsay KR, Te Velthuis AJW, Lauring AS. 2023. Deep mutational scanning reveals the functional constraints and evolutionary potential of the influenza A virus PB1 protein. *J Virol* 97:e01329-23. <https://doi.org/10.1128/jvi.01329-23>
- Haddox HK, Diggins AS, Bloom JD. 2016. Experimental estimation of the effects of all amino-acid mutations to HIV's envelope protein on viral replication in cell culture. *PLoS Pathog* 12:e1006114. <https://doi.org/10.1371/journal.ppat.1006114>
- Wu NC, Young AP, Al-Mawsawi LQ, Olson CA, Feng J, Qi H, Chen S-H, Lu I-H, Lin C-Y, Chin RG, Luan HH, Nguyen N, Nelson SF, Li X, Wu T-T, Sun R. 2014. High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Sci Rep* 4:4942. <https://doi.org/10.1038/srep04942>
- Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Diggins AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veelsler D, Bloom JD. 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182:1295–1310. <https://doi.org/10.1016/j.cell.2020.08.012>
- Gaiotto T, Hufton SE. 2016. Cross-neutralising nanobodies bind to a conserved pocket in the hemagglutinin stem region identified using yeast display and deep mutational scanning. *PLoS One* 11:e0164296. <https://doi.org/10.1371/journal.pone.0164296>
- Gai SA, Wittrup KD. 2007. Yeast surface display for protein engineering and characterization. *Curr Opin Struct Biol* 17:467–473. <https://doi.org/10.1016/j.sbi.2007.08.012>
- Javanmardi K, Chou C-W, Terrace CI, Annapareddy A, Kaoud TS, Guo Q, Lutgens J, Zorkic H, Horton AP, Gardner EC, Nguyen G, Boutz DR, Goike J, Voss WN, Kuo H-C, Dalby KN, Gollihar JD, Finkelstein J. 2021. Rapid characterization of spike variants via mammalian cell surface display. *Mol Cell* 81:5099–5111. <https://doi.org/10.1016/j.molcel.2021.11.024>
- Frank F, Keen MM, Rao A, Bassit L, Liu X, Bowers HB, Patel AB, Cato ML, Sullivan JA, Greenleaf M, Piantadosi A, Lam WA, Hudson WH, Ortlund EA. 2022. Deep mutational scanning identifies SARS-CoV-2 Nucleocapsid escape mutations of currently available rapid antigen tests. *Cell* 185:3603–3616. <https://doi.org/10.1016/j.cell.2022.08.010>
- Dadonaite B, Ahn JJ, Ort JT, Yu J, Furey C, Dosey A, Hannon WW, Baker ALV, Webby RJ, King NP, Liu Y, Hensley SE, Peacock TP, Moncla LH, Bloom JD. 2024. Deep mutational scanning of H5 hemagglutinin to inform influenza virus surveillance. *bioRxiv:2024.05.23.595634*. <https://doi.org/10.1101/2024.05.23.595634>
- Dadonaite B, Crawford KHD, Radford CE, Farrell AG, Yu TC, Hannon WW, Zhou P, Andrabi R, Burton DR, Liu L, Ho DD, Chu HY, Neher RA, Bloom JD. 2023. A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* 186:1263–1278. <https://doi.org/10.1016/j.cell.2023.02.001>
- Starr TN, Greaney AJ, Hannon WW, Loes AN, Hauser K, Dillen JR, Ferri E, Farrell AG, Dadonaite B, McCallum M, Matreyek KA, Corti D, Veelsler D, Snell G, Bloom JD. 2022. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* 377:420–424. <https://doi.org/10.1126/science.abo7896>
- Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, Hilton SK, Huddleston J, Eguia R, Crawford KHD, Diggins AS, Nargi RS, Sutton RE, Suryadevara N, Rothlauf PW, Liu Z, Whelan SPJ, Carnahan RH, Crowe JE,

- Bloom JD. 2021. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* 29:44–57. <https://doi.org/10.1016/j.chom.2020.11.007>
31. Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. *Nat Methods* 11:801–807. <https://doi.org/10.1038/nmeth.143027>
 32. Thyagarajan B, Bloom JD. 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* 3:e03300. <https://doi.org/10.7554/eLife.03300>
 33. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, Fields S. 2010. High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7:741–746. <https://doi.org/10.1038/nmeth.1492>
 34. Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. *Proc Natl Acad Sci U S A* 108:7896–7901. <https://doi.org/10.1073/pnas.1016024108>
 35. Starita LM, Pruneda JN, Lo RS, Fowler DM, Kim HJ, Hiatt JB, Shendure J, Brzovic PS, Fields S, Kleit RE. 2013. Activity-enhancing mutations in an E3 ubiquitin ligase identified by high-throughput mutagenesis. *Proc Natl Acad Sci USA* 110. <https://doi.org/10.1073/pnas.1303309110>
 36. Carr CR, Crawford KHD, Murphy M, Galloway JG, Haddock HK, Matsen FA, Andersen KG, King NP, Bloom JD. 2024. Deep mutational scanning reveals functional constraints and antibody-escape potential of Lassa virus glycoprotein complex. *Immunity* 57:2061–2076. <https://doi.org/10.1016/j.immuni.2024.06.013>
 37. Teo QW, Wang Y, Lv H, Oade MS, Mao KJ, Tan TJC, Huan YW, Rivera-Cardona J, Shao EK, Choi D, Wang C, Tavakoli Dargani Z, Brooke CB, te Velthuis AJW, Wu NC. 2025. Probing the functional constraints of influenza A virus NEP by deep mutational scanning. *Cell Rep* 44:115196. <https://doi.org/10.1016/j.celrep.2024.115196>
 38. Yu Y, Kass MA, Zhang M, Youssef N, Freije CA, Brock KP, Aguado LC, Seifert LL, Venkittu S, Hong X, Shlomai A, de Jong YP, Marks DS, Rice CM, Schneider WM. 2024. Deep mutational scanning of hepatitis B virus reveals a mechanism for cis-preferential reverse transcription. *Cell* 187:2735–2745. <https://doi.org/10.1016/j.cell.2024.04.008>
 39. ElHefnawi M, Alaidi O, Mohamed N, Kamar M, El-Azab I, Zada S, Siam R. 2011. Identification of novel conserved functional motifs across most Influenza A viral strains. *Virology* 418:44–49. <https://doi.org/10.1016/j.virol.2011.07.011>
 40. Marsh GA, Rabadán R, Levine AJ, Palese P. 2008. Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. *J Virol* 82:2295–2304. <https://doi.org/10.1128/JVI.02267-07>
 41. Ngo V-N, Winkler DP, Aho B, Kamath PL, King BL, Waters H, Zimmerberg J, Sodt A, Hess ST. 2024. Conserved sequence features in intracellular domains of viral spike proteins. *Virology (Auckl)* 599:110198. <https://doi.org/10.1016/j.virol.2024.110198>
 42. Hom N, Gentles L, Bloom JD, Lee KK. 2019. Deep mutational scan of the highly conserved influenza A virus M1 matrix protein reveals substantial intrinsic mutational tolerance. *J Virol* 93:e00161-19. <https://doi.org/10.1128/JVI.00161-19>
 43. Doud MB, Bloom JD. 2016. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* 8:155. <https://doi.org/10.3390/v8060155>
 44. Lei R, Hernandez García A, Tan TJC, Teo QW, Wang Y, Zhang X, Luo S, Nair SK, Peng J, Wu NC. 2023. Mutational fitness landscape of human influenza H3N2 neuraminidase. *Cell Rep* 42:111951. <https://doi.org/10.1016/j.celrep.2022.111951>
 45. Dadonaite B, Brown J, McMahon TE, Farrell AG, Figgins MD, Asarnow D, Stewart C, Lee J, Logue J, Bedford T, Murrell B, Chu HY, Veelsler D, Bloom JD. 2024. Spike deep mutational scanning helps predict success of SARS-CoV-2 clades. *Nature* 631:617–626. <https://doi.org/10.1038/s41586-024-07636-1>
 46. Bdeir N, Lüddecke T, Maaß H, Schmelz S, Rand U, Jacobsen H, Metzdorf K, Kulkarni U, Cossmann A, Stankov MV, Hoffmann M, Pöhlmann S, Blankenfeldt W, Dopfer-Jablonka A, Behrens GMN, Čičin-Šain L. 2025. Reverse mutational scanning of SARS-CoV-2 spike BA.2.86 identifies epitopes contributing to immune escape from polyclonal sera. *Nat Commun* 16:809. <https://doi.org/10.1038/s41467-025-55871-5>
 47. Jian F, Feng L, Yang S, Yu Y, Wang L, Song W, Yisimayi A, Chen X, Xu Y, Wang P, Yu L, Wang J, Liu L, Niu X, Wang J, Xiao T, An R, Wang Y, Gu Q, Shao F, Jin R, Shen Z, Wang Y, Wang X, Cao Y. 2023. Convergent evolution of SARS-CoV-2 XBB lineages on receptor-binding domain 455-456 synergistically enhances antibody evasion and ACE2 binding. *PLOS Pathog* 19:e1011868. <https://doi.org/10.1371/journal.ppat.1011868>
 48. Taylor AL, Starr TN. 2023. Deep mutational scans of XBB.1.5 and BQ.1.1 reveal ongoing epistatic drift during SARS-CoV-2 evolution. *bioRxiv:2023.09.11.557279*. <https://doi.org/10.1101/2023.09.11.557279>
 49. Taylor AL, Starr TN. 2024. Deep mutational scanning of SARS-CoV-2 Omicron BA.2.86 and epistatic emergence of the KP.3 variant. *Virus Evol* 10:veae067. <https://doi.org/10.1093/ve/veae067>
 50. Haddock HK, Galloway JG, Dadonaite B, Bloom JD, Matsen Iv FA, DeWitt WS. 2023. Jointly modeling deep mutational scans identifies shifted mutational effects among SARS-CoV-2 spike homologs. *bioRxiv:2023.07.31.551037*. <https://doi.org/10.1101/2023.07.31.551037>
 51. Taminishi S, Li S, Higuchi Y, Kirita Y, Motooka D, Ozaki Y, Arimori T, Ikemura N, Itoh Y, Matoba S, Okamoto T, Takagi J, Standley DM, Hoshino A. 2023. Deep mutational scanning of whole SARS-CoV-2 spike in an inverted infection system. *bioRxiv*. <https://doi.org/10.1101/2023.07.17.549430>
 52. Raisingshani N, Alshahrani M, Gupta G, Verkhivker G. 2024. AlphaFold2 modeling and molecular dynamics simulations of the conformational ensembles for the SARS-CoV-2 spike Omicron JN.1, KP.2 and KP.3 variants: mutational profiling of binding energetics reveals epistatic drivers of the ACE2 affinity and escape hotspots of antibody resistance. *Viruses* 16:1458. <https://doi.org/10.3390/v16091458>
 53. Raisingshani N, Alshahrani M, Gupta G, Verkhivker G. 2024. Ensemble-based mutational profiling and network analysis of the SARS-CoV-2 spike Omicron XBB lineages for interactions with the ACE2 receptor and antibodies: cooperation of binding hotspots in mediating epistatic couplings underlies binding mechanism and immune escape. *Int J Mol Sci* 25:4281. <https://doi.org/10.3390/ijms25084281>
 54. Weaver SC. 2006. Evolutionary influences in arboviral disease. *Curr Top Microbiol Immunol* 299:285–314. https://doi.org/10.1007/3-540-26397-7_10
 55. Setoh YX, Amarilla AA, Peng NYG, Griffiths RE, Carrera J, Freney ME, Nakayama E, Ogawa S, Watterson D, Modhiran N, et al. 2019. Determinants of Zika virus host tropism uncovered by deep mutational scanning. *Nat Microbiol* 4:876–887. <https://doi.org/10.1038/s41564-019-0399-4>
 56. Stumpf MM, Brunetti T, Davenport BJ, McCarthy MK, Morrison TE. 2024. Deep mutationally scanned (DMS) CHIKV E3/E2 virus library maps viral amino acid preferences and predicts viral escape mutants of neutralizing CHIKV antibodies. *bioRxiv:2024.12.04.626854*. <https://doi.org/10.1101/2024.12.04.626854>
 57. Garrett ME, Itell HL, Crawford KHD, Basom R, Bloom JD, Overbaugh J. 2020. Phage-DMS: a comprehensive method for fine mapping of antibody epitopes. *iScience* 23:101622. <https://doi.org/10.1016/j.isci.2020.101622>
 58. Hanning KR, Minot M, Warrander AK, Kelton W, Reddy ST. 2022. Deep mutational scanning for therapeutic antibody engineering. *Trends Pharmacol Sci* 43:123–135. <https://doi.org/10.1016/j.tips.2021.11.010>
 59. Guthmiller JJ, Han J, Utset HA, Li L, Lan LY-L, Henry C, Stamper CT, McMahon M, O'Dell G, Fernández-Quintero ML, et al. 2022. Broadly neutralizing antibodies target a haemagglutinin anchor epitope. *Nature* 602:314–320. <https://doi.org/10.1038/s41586-021-04356-8>
 60. Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, Kranz DM, Herbert AS, Procko E. 2020. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* 369:1261–1265. <https://doi.org/10.1126/science.abc0870>
 61. Chan KK, Tan TJC, Narayanan KK, Procko E. 2021. An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci Adv* 7:eabf1738. <https://doi.org/10.1126/sciadv.abf1738>
 62. Zhang L, Dutta S, Xiong S, Chan M, Chan KK, Fan TM, Bailey KL, Lindeblad M, Cooper LM, Rong L, Gugliuzza AF, Shukla D, Procko E, Rehman J, Malik AB. 2022. Engineered ACE2 decoy mitigates lung injury and death induced by SARS-CoV-2 variants. *Nat Chem Biol* 18:342–351. <https://doi.org/10.1038/s41589-021-00965-6>
 63. Skeeters S, Bagale K, Stepanyuk G, Thieker D, Aguhob A, Chan KK, Dutzar B, Shalygin S, Shajahan A, Yang X, DaRosa PA, Frazier E, Sauer MM, Bogatzki L, Byrnes-Blake KA, Song Y, Azadi P, Tarcha E, Zhang L, Procko E. 2024. Modulation of the pharmacokinetics of soluble ACE2 decoy receptors through glycosylation. *Mol Ther Methods Clin Dev* 32:101301. <https://doi.org/10.1016/j.omtm.2024.101301>
 64. Mojica FJM, Diez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive

- from foreign genetic elements. *J Mol Evol* 60:174–182. <https://doi.org/10.1007/s00239-004-0046-3>
65. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315:1709–1712. <https://doi.org/10.1126/science.1138140>
 66. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821. <https://doi.org/10.1126/science.1225829>
 67. Karakus U, Thamamongood T, Ciminski K, Ran W, Günther SC, Pohl MO, Eiletto D, Jeney C, Hoffmann D, Reiche S, et al. 2019. MHC class II proteins mediate cross-species entry of bat influenza viruses. *Nature* 567:109–112. <https://doi.org/10.1038/s41586-019-0955-3>
 68. Zhu Y, Feng F, Hu G, Wang Y, Yu Y, Xu W, Cai X, Sun Z, Han W, Ye R, Qu D, Ding Q, Huang X, Chen H, Xu W, Xie Y, Cai Q, Yuan Z, Zhang R. 2021. A genome-wide CRISPR screen identifies host factors that regulate SARS-CoV-2 entry. *Nat Commun* 12:961. <https://doi.org/10.1038/s41467-021-21213-4>
 69. Zhang R, Kim AS, Fox JM, Nair S, Basore K, Klimstra WB, Rimkunas R, Fong RH, Lin H, Poddar S, Crowe JE, Doranz BJ, Fremont DH, Diamond MS. 2018. Mxra8 is a receptor for multiple arthritogenic alphaviruses. *Nature* 557:570–574. <https://doi.org/10.1038/s41586-018-0121-3>
 70. Richardson RB, Ohlson MB, Eitson JL, Kumar A, McDougal MB, Boys IN, Mar KB, De La Cruz-Rivera PC, Douglas C, Konopka G, Xing C, Schoggins JW. 2018. A CRISPR screen identifies IFI6 as an ER-resident interferon effector that blocks flavivirus replication. *Nat Microbiol* 3:1214–1223. <https://doi.org/10.1038/s41564-018-0244-1>
 71. Marceau CD, Puschnik AS, Majzoub K, Ooi YS, Brewer SM, Fuchs G, Swaminathan K, Mata MA, Elias JE, Sarnow P, Carette JE. 2016. Genetic dissection of *Flaviviridae* host factors through genome-scale CRISPR screens. *Nature* 535:159–163. <https://doi.org/10.1038/nature18631>
 72. Ma Y, Walsh MJ, Bernhardt K, Ashbaugh CW, Trudeau SJ, Ashbaugh IY, Jiang S, Jiang C, Zhao B, Root DE, Doench JG, Gewurz BE. 2017. CRISPR/Cas9 screens reveal Epstein-Barr virus-transformed B cell host dependency factors. *Cell Host Microbe* 21:580–591. <https://doi.org/10.1016/j.chom.2017.04.005>
 73. OhAinle M, Helms L, Vermeire J, Roesch F, Humes D, Basom R, Delrow JJ, Overbaugh J, Emerman M. 2018. A virus-packageable CRISPR screen identifies host factors mediating interferon inhibition of HIV. *Elife* 7:e39823. <https://doi.org/10.7554/eLife.39823>
 74. Prelli Bozzo C, Laliberté A, De Luna A, Pastorio C, Regensburger K, Krebs S, Graf A, Blum H, Volcic M, Sparrer KMJ, Kirchhoff F. 2024. Replication competent HIV-guided CRISPR screen identifies antiviral factors including targets of the accessory protein Nef. *Nat Commun* 15:3813. <https://doi.org/10.1038/s41467-024-48228-x>
 75. Finkel Y, Nachshon A, Aharon E, Arazi T, Simonovsky E, Dobešová M, Saud Z, Gluck A, Fisher T, Stanton RJ, Schwartz M, Stern-Ginossar N. 2024. A virally encoded high-resolution screen of cytomegalovirus dependencies. *Nature* 630:712–719. <https://doi.org/10.1038/s41586-024-07503-z>
 76. King CR, Liu Y, Amato KA, Schaack GA, Mickelson C, Sanders AE, Hu T, Gupta S, Langlois RA, Smith JA, Mehle A. 2023. Pathogen-driven CRISPR screens identify TREX1 as a regulator of DNA self-sensing during influenza virus infection. *Cell Host Microbe* 31:1552–1567. <https://doi.org/10.1016/j.chom.2023.08.001>
 77. Dodkins R, Delaney JR, Overton T, Scholle F, Frias-De-Diego A, Crisci E, Huq N, Jordan I, Kimata JT, Findley T, Goldberg IG. 2023. A rapid, high-throughput, viral infectivity assay using automated brightfield microscopy with machine learning. *SLAS Technol* 28:324–333. <https://doi.org/10.1016/j.slast.2023.07.003>
 78. Wang T-E, Chao T-L, Tsai H-T, Lin P-H, Tsai Y-L, Chang S-Y. 2020. Differentiation of cytopathic effects (CPE) induced by influenza virus infection using deep convolutional neural networks (CNN). *PLOS Comput Biol* 16:e1007883. <https://doi.org/10.1371/journal.pcbi.1007883>
 79. Werner J, Kronberg RM, Stachura P, Ostermann PN, Müller L, Schaal H, Bhatia S, Kather JN, Borkhardt A, Pandya AA, Lang KS, Lang PA. 2021. Deep transfer learning approach for automatic recognition of drug toxicity and inhibition of SARS-CoV-2. *Viruses* 13:610. <https://doi.org/10.3390/v13040610>
 80. Sarfati H, Naftaly S, Papo N, Keasar C. 2022. Predicting mutant outcome by combining deep mutational scanning and machine learning. *Proteins* 90:45–57. <https://doi.org/10.1002/prot.26184>
 81. Zhou Y, Pan Q, Pires DEV, Rodrigues CHM, Ascher DB. 2023. DDMut: predicting effects of mutations on protein stability using deep learning. *Nucleic Acids Res* 51:W122–W128. <https://doi.org/10.1093/nar/gkad472>
 82. Gelman S, Fahlberg SA, Heinzelman P, Romero PA, Gitter A. 2021. Neural networks to learn protein sequence–function relationships from deep mutational scanning data. *Proc Natl Acad Sci USA* 118:e2104878118. <https://doi.org/10.1073/pnas.2104878118>
 83. Wu NC, Du Y, Le S, Young AP, Zhang T-H, Wang Y, Zhou J, Yoshizawa JM, Dong L, Li X, Wu T-T, Sun R. 2016. Coupling high-throughput genetics with phylogenetic information reveals an epistatic interaction on the influenza A virus M segment. *BMC Genomics* 17:46. <https://doi.org/10.1186/s12864-015-2358-7>
 84. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, Fowler DM, Rubin AF. 2019. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome Biol* 20:223. <https://doi.org/10.1186/s13059-019-1845-6>
 85. Atlas of Variant Effects Alliance. Atlas Var Eff Alliance. Available from: <https://www.varianteffect.org>. Retrieved 23 Dec 2024.
 86. Shapiro JS, Langlois RA, Pham AM, Tenover BR. 2012. Evidence for a cytoplasmic microprocessor of pri-miRNAs. *RNA* 18:1338–1346. <https://doi.org/10.1261/ma.032268.112>