



Robustez mutacional de virus ARN

Lic. Bioq. Rodrigo Arce Rama

Tutor: Dr. Gonzalo Moratorio

Co-tutora: Dra. Marianoel Pereira-Gómez

Tesis de maestría en Ciencias Biológicas
Opción Biología Celular y Molecular, PEDECIBA

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

Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo

Agosto 2022

Montevideo, Uruguay

Robustez mutacional de virus ARN

Lic. Rodrigo Arce Rama

Tesis presentada con el objetivo de obtener el título de Magíster en Ciencias Biológicas, opción Biología Celular y Molecular, en el marco del Programa PEDECIBA.

Tutor: Dr. Gonzalo Moratorio

Co-tutora: Dra. Marianoel Pereira-Gómez

Tribunal: Dr. Nicolás Sarute (presidente)

Dra. Natalia Olivero

Dr. Rodrigo Puentes

Tesis realizada 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 agosto de 2020 y agosto de 2022.

Agradecimientos

Luego de dos años, se cierra un ciclo formativo duro pero lleno de aprendizajes. Esta pequeña sección se la dedico a quienes me guiaron en este camino lleno de mutaciones y a la gente bella que en estos dos años me ayudó a bufferear el efecto de las mismas.

A Gonza y Noel, por la orientación y formación a lo largo de estos dos años.

A Pili y Gonza (nuevamente), por el aguante y consejos en estos dos años y por todos los escaloncitos que fueron poniendo en mi camino para que siga creciendo como profesional.

A los miembros del tribunal, Dres. Natalia Olivero, Rodrigo Puentes y Nicolás Sarute, por la buena voluntad de haber aceptado evaluar esta tesis en un período de tiempo tan corto.

A la familia LVM/LEEV, que sin ustedes hoy no habría llegado a destino. Fabi, Ali, Meche, Lompi, Naty, Varo, Diegao, Ire, Pau, Joaco y Lucas. Gracias por las risas, el aliento, las tardes de plena y trap en el laboratorio y el empuje de ustedes siempre que lo necesité. A Juan, por las buenas historias y siempre la buena onda de intentarnos enseñar cosas nuevas. Volvé Varo!!!

A la gente linda que conocí en estos dos años en el Institut: Valen, Marco (el reencuentro), los vecinos del LGM y muchas otras personas bellas del Institut.

A la gente de Zurgén, en especial a mi Team Finde. Richard, Dai, Facu, Vale, Gonza, Anita y Vicky por hacer que trabajar los fines de semana sea algo incluso lindo de vivir.

A los pibes, que espero que ya hayan asumido que año a año voy perdiendo mi estabilidad mental. En especial al Chevi por introducirme en el mundo de la ciencia. A los tortus, Gonza, Tincho y Garra, a Ana y James.

A mi familia de allá, Ana, Marc, Juanjo, Ana yaya, Susan, Ruth, María y papá.

A mi familia de acá por todos los momentos lindos, el cariño y el aguante en estos dos años duros. A Gra, Cami, Pipi, Flo, Rafa, Mili, Vale, Nacho, Pela, tío José, el primo Massi, tía Ramona Gus, Didier y a todos los peques y más seres queridos.

Gracias abuelos, por todo. Finalmente, a mis dos pilares en estos dos años, Lucía y mamá. Millones de gracias, soy quién soy, gracias a ustedes dos que son la base que me permite crecer en todos los buenos sentidos de la vida.

Resumen

Los virus son causantes de un gran número de enfermedades. Frecuentemente el desarrollo de tratamientos efectivos falla debido a la gran capacidad evolutiva de estos patógenos de adquirir resistencias, en gran parte, como consecuencia de sus altas tasas de mutación. Dada la ausencia de actividad correctora de errores, los virus de ARN presentan tasas de error de hasta un millón de veces mayor que otros organismos con genomas de ADN. Este fenómeno puede considerarse como su talón de Aquiles, ya que hasta el 40% de las mutaciones generadas son letales en estos virus. Por consiguiente, se cree que los virus evolucionaron para limitar el efecto deletéreo de las mutaciones, seleccionándose codones que, cuando mutan, lo hacen mayormente de forma silenciosa, es decir, sin cambiar la secuencia proteica. Este mecanismo de tolerancia a la mutación, también conocido como robustez mutacional o genética, contrarresta sus tasas de mutación extremas. El equilibrio entre estas dos propiedades determina la base del éxito evolutivo de los virus de ARN.

En este proyecto, nos propusimos estudiar la relación de la volatilidad genética con la robustez mutacional y su efecto en la evolución viral. La volatilidad genética es la probabilidad de un codón a cambiar frente a una mutación aleatoria. Con este fin trabajamos con 3 genotipos diferentes del enterovirus humano Coxsackievirus B3 que difieren en la volatilidad genética de su genoma. Comprobamos *in vitro* que estos tres genotipos virales tenían el mismo comportamiento a una baja multiplicidad de infección. Asimismo realizamos una evolución experimental con el fin de que la mayor fuerza evolutiva actuante sea la deriva génica y así, las mutaciones que se fijaran en la población sean producto del azar y no de la selección. Se cuantificó la capacidad evolutiva de estas poblaciones evolucionadas mediante el estudio relativo del fitness viral a las poblaciones ancestrales. Finalmente se analizaron las mutaciones fijadas durante la evolución y se relacionó la capacidad de tolerar mutaciones de los virus con sus características genómicas.

Palabras clave: evolución viral, robustez mutacional, volatilidad genómica, Coxsackie virus B3, virología molecular

Abstract

Viruses are the cause of many diseases, and treatment development frequently fails due to the evolutionary capacity of these pathogens to acquire resistance, largely as a consequence of their high mutation rates. Given the absence of error-correcting activity, RNA viruses have error rates up to one million times higher than other organisms with DNA genomes. Thus, each possible mutation is generated in a few replication cycles. However, this phenomenon can also be considered their Achilles heel since up to 40% of the mutations generated are lethal in RNA viruses. Therefore, it is believed that viruses evolved to limit the deleterious effect of mutations by selecting codons that, when they mutate, do so mostly silently, i.e. without changing the protein sequence. This mutation tolerance mechanism, also known as mutational or genetic robustness, counteracts their extreme mutation rates. The balance between these two properties determines the basis for the evolutionary success of RNA viruses.

In this project, we propose to study the relationship of genetic volatility with mutational robustness and its effect on viral evolution. Genetic volatility is the probability of a codon changing in the face of a random mutation. While mutational robustness refers to the ability of a genotype to maintain its phenotype despite the effect of the mutation. In other words, the ability to buffer the effect so as not to change phenotypically. In this work, 3 different genotypes of the human enterovirus Coxsackievirus B3 were used in which the genotypes differed in the genetic volatility of their genome. It was found that in vitro the viruses had the same behavior at low MOI. In addition, experimental evolution was carried out by making passages from plaque to plaque so that the major evolutionary force acting is genetic drift and thus the mutations that will be fixed in the population are random and not a product of selection. The evolutionary capacity of these evolved populations was quantified by studying their relative viral fitness to the ancestral populations. Finally, the mutations fixed in the evolved populations were analyzed, and the ability of the viruses to tolerate mutations was related to their genomic characteristics.

Keywords: viral evolution, mutational robustness, genomic volatility, Coxsackievirus B3, molecular virology

Tabla de contenido

1. Introducción	1
1.1 Fuerzas evolutivas	2
1.1.1 Mutación	2
1.1.2 Deriva genética	3
1.2 Robustez mutacional	4
1.2.1 Volatilidad genómica	5
1.2.2 Robustez mutacional de virus ARN	7
1.3 Coxsackie virus B3, nuestro modelo de estudio	9
1.3.1 Clasificación y patogenicia	9
1.3.2 Organización genómica y ciclo replicativo	9
2. Hipótesis	12
3. Objetivos	13
3.1 Objetivo General	13
3.2 Objetivos específicos	13
4. Materiales y métodos	14
4.1 Líneas celulares	14
4.2 Mutantes de CVB3	14
4.3 Generación de stock virales CVB3	15
4.4 Titulación viral	16
4.4.1 Plaque assay	17
4.4.2 TCID50	18
4.5 Extracción de ARN	19
4.5.1 Extracción con TRIzol	19
4.5.2 Extracción con kit Qiagen QIAamp® Viral RNA	19
4.5.3 Extracción de ARN con robot PurePrep 96 (Molgen)	19
4.6 Retrotranscripción (RT)	19
4.6.1 SuperScript III	19
4.6.2 LunaScript RT SuperMix Kit (New England BioLabs)	20
4.7 Amplificación del genoma de CVB3 por PCR a tiempo final	20
4.7.1 Amplificación de la región P1	20

4.7.2 Amplificación del genoma completo de CVB3	21
4.8 Ensayos de cinética replicativa	21
4.9 Pasajes de calva en calva	22
4.10 Ensayos de competencia	23
4.11 RT-qPCR para ensayos de competencia y cuantificación de genomas	24
4.11.1 RT-qPCR múltiplex para ensayos de competencia	24
4.11.2 RT-qPCR para cuantificar genomas	27
4.12 Medición del tamaño de calvas de las poblaciones evolucionadas	28
4.13 Next Generation Sequencing	29
5. Resultados	31
5.1 Generación y titulación de los Virus CVB3 WT, MoreV y LessV	31
5.2 PCR a tiempo final para secuenciación	33
5.3 Ensayos de cinética de replicación de los ancestros de CVB3	34
5.3.1 Cinética de replicación a MOI=3	35
5.3.2 Cinética de replicación viral a MOI=0.01	37
5.4 Evolución experimental por deriva génica	40
5.4.1 Tamaño poblacional durante la evolución calva a calva.	41
5.4.2 Fenotipo de calvas	43
5.4.3 Fitness relativo al CVB3 de Referencia	45
5.5 Análisis de las bases moleculares de las poblaciones generadas en la evolución experimental	46
6. Discusión	50
6.1 Con respecto al desarrollo de protocolos y herramientas moleculares para CVB3:	50
6.2 Con respecto al estudio de la cinética de replicación de las variantes de volatilidad de CVB3:	51
6.3 En relación al estudio por evolución experimental del impacto de la deriva génica en las variantes de volatilidad de CVB3:	53
6.4 Estudio de las mutaciones fijadas a lo largo de la evolución	57
6.5 En suma:	58
7. Conclusiones	60
8. Perspectivas	61

9. Bibliografía	62
Anexos	73
Anexo 1 - Cantidades utilizadas en cultivo celular	73
Anexo 2 - Generación de la progenie 0 y 1	74
Anexo 3 - Extracción de ARN con TRIzol	76
Anexo 4 - Extracción de ARN con QIAamp® Viral RNA	77
Anexo 5 - Protocolo de Extracción de ARN viral con robot Pureprep_Molgen	78
Anexo 6 - Generación de librerías para secuenciación en plataforma GridION	79
Anexo 7 - Tablas de mutaciones fijadas a lo largo de la evolución experimental por deriva génica	81
Anexo 8 - Calidad de la secuenciación de las poblaciones evolucionadas en plataforma GridION	84
Anexo 9 - Publicaciones realizadas durante el desarrollo de la tesis	87

1. Introducción

En la actualidad los virus ARN representan una gran problemática a nivel de salud, ya sea salud humana, animal o vegetal. Generar conocimientos y estrategias para combatirlos es uno de los grandes objetivos que enfrenta la ciencia. Para ello, es importante entender los principios evolutivos que subyacen el gran potencial adaptativo de estos patógenos.

La evolución se define como los cambios en los rasgos heredables de una población a medida que las sucesivas generaciones se van sustituyendo, siendo las poblaciones las que evolucionan y no los organismos individuales (National Academies of Sciences Engineering and Medicine (US)). Cuando hablamos de evolución pensamos inmediatamente en Charles Darwin quien en 1859 publicó “El origen de las especies” (Darwin, 1859). En este libro, Darwin definió la evolución como “descendencia con modificación”, un mecanismo por el cual las poblaciones cambian a lo largo del tiempo, dando origen a nuevas especies. También postulaba que toda la diversidad actual de especies compartiría un ancestro en común.

Otro fundamental aporte de Darwin, y de Alfred R. Wallace (Darwin & Wallace, 1858), es el de la selección natural como la principal fuerza por la que la evolución actúa. Asimismo, Darwin propuso que los individuos más adaptados al medio ambiente (más aptos) de una población tienden en promedio a sobrevivir o dejar más descendencia que aquellos menos adaptados (la 'supervivencia del más apto'). Cuando nos referimos a los individuos más adaptados, nos referimos a aquellos individuos con más fitness, siendo el fitness la capacidad de un organismo a sobrevivir y reproducirse en el ambiente en el que se encuentran (Orr, 2009).

Por otra parte, ya en el siglo XX, más precisamente en la década de 1960, Motoo Kimura (Kimura, 1968) desarrolló la teoría neutralista de la evolución molecular que establece que la mayoría de los cambios evolutivos que se producen son selectivamente neutros a nivel molecular, por lo que las

mutaciones beneficiosas no se verán seleccionadas y propone que estos cambios se deben a otras dos fuerzas: la variación, que se debe en gran parte a la mutación, y la deriva génica (Kern & Hahn, 2018), siendo la deriva el cambio genotípico ocurrido por el efecto del azar (Rye et al., 2016).

Existen además otras dos fuerzas que tienen roles evolutivos claves, como el flujo genético, que hace referencia a cambios en las frecuencias alélicas entre diferentes poblaciones, y la recombinación, que da lugar a una reestructuración de una parte del genoma. Estas dos fuerzas no serán abordadas en esta tesis.

1.1 Fuerzas evolutivas

1.1.1 Mutación

La mutación genética se refiere a cualquier cambio al azar que modifica la secuencia de nucleótidos de un organismo. Estos cambios pueden ser mutaciones puntuales, deleciones o inserciones. Las mutaciones puntuales son cambios de un nucleótido que es reemplazado por otro. Otros posibles errores de la replicación pueden ser las deleciones e inserciones. En las deleciones se elimina al menos un nucleótido de la secuencia. Mientras que en las inserciones se inserta al menos un nucleótido. Estos tipos de cambios pueden llevar a cambiar el marco abierto de lectura si se dan en una región codificante (Brown, 2002). La mutación provee la materia prima indispensable para la adaptación ya que “un organismo que no mutase sería incapaz de adaptarse a los cambios ambientales y por lo tanto estaría condenado a la extinción” (Sanjuán Verdeguer, 2005). Por otra parte, las mutaciones suelen tener impactos negativos sobre los organismos (Sanjuán Verdeguer, 2005). Cuando las mutaciones suceden en un marco abierto de lectura (ORF) estas pueden ser sinónimas o no sinónimas; neutras o deletéreas. Las mutaciones sinónimas son aquellas que el cambio nucleotídico no causará una sustitución aminoacídica. En las no sinónimas, la mutación dará lugar a un cambio aminoacídico. Con respecto a las mutaciones deletéreas, las mismas comprometen el fitness del organismo; un caso particular de mutación deletérea es aquella donde se genera un codón *stop*.

Si bien las mutaciones sinónimas deberían ser neutras a nivel fenotípico (Kimura, 1968, 1977), se ha visto que en virus ARN pueden tener efectos deletéreos (Cuevas et al., 2012) debido a cambios en regiones con estructuras secundarias que son fundamentales para la replicación, cambios que comprometan la expresión de genes o en señales de encapsidación (Mayrose et al., 2013; Pereira-Gómez, 2016).

1.1.2 Deriva genética

La deriva genética se define como el muestreo aleatorio de los genotipos entre generaciones y es uno de los modos en que la evolución progresa en ausencia de selección natural (Kimura, 1983; Sanjuán Verdeguer, 2005). En otras palabras, es la evolución de una población provocada por eventos aleatorios, impredecibles.

La influencia de la deriva en una población es dependiente de su tamaño efectivo (N_e). El tamaño efectivo de una población está determinado por factores que pueden reducir la variabilidad genética y de esta manera, hacer que N_e sea sustancialmente inferior al tamaño de la población. Algunos factores que afectan el N_e son el modo de reproducción, los cuellos de botella de la población (reducciones drásticas en el N_e), el enlace entre genes o la selección natural (Elena & Sanjuán, 2007). La deriva depende también del coeficiente de selección (s), que se define como la diferencia del fitness relativo entre un genotipo mutante y el genotipo wild type. En los casos en los que el $N_e \times s > 1$ la selección natural prevalecerá sobre la deriva, mientras que en el caso contrario será la deriva la que prevalecerá (Ohta, 1992).

A nivel experimental, una de las maneras de fomentar la deriva génica por sobre la selección natural se realiza con pasajes de calva en calva (**Figura 1**). Esta estrategia permite generar cuellos de botella en donde las generaciones se mantienen en pequeños tamaños y así las chances de que se fijen mutaciones beneficiosas o compensatorias sea menor. Además, en este caso, las mutaciones deletéreas se comportan como casi neutras por lo que se pueden acumular en la población y como producto de esto el fitness puede verse afectado de manera negativa, siempre y cuando las poblaciones sigan

manteniéndose en pequeños tamaños (Elena & Sanjuán, 2007). En poblaciones con estas características, es esperable que la acumulación de mutaciones lleve a las poblaciones a su extinción por colapso mutacional (Lynch et al., 1993). El mecanismo que actúa para llevar a una población a su fin es el trinquete de Müller que es un proceso, que en poblaciones asexuales y sin recombinación, da como resultado la acumulación de mutaciones deletéreas de manera irreversible (Muller, 1932, 1964).

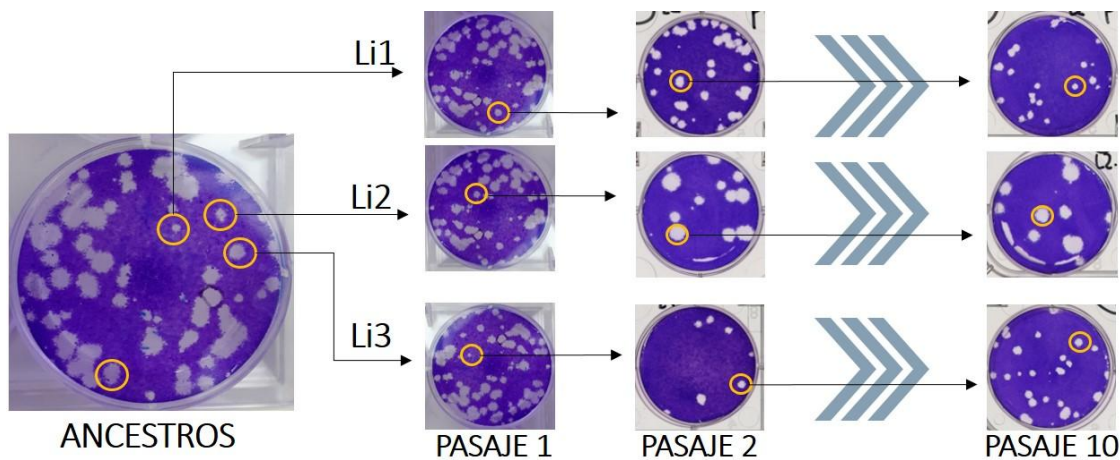


Figura 1. Pasajes de calva en calva. En estos experimentos se plaquean los ancestros de las poblaciones virales a estudiar. A partir de este plaqueo se muestrean aleatoriamente 3 calvas que van a dar origen a 3 linajes distintos (Li1-Li3). Estas calvas son sucesivamente pasadas de la misma manera.

1.2 Robustez mutacional

Los virus ARN poseen las tasas de mutación más altas conocidas hasta el momento. Estas se definen como el “número de cambios genéticos (mutaciones puntuales, inserciones, deleciones) introducidas en cada ronda de replicación” (Drake, 1993; Duffy et al., 2008). Para enfrentarse a la mutación, los organismos multicelulares con genomas de ADN grandes y complejos utilizan mecanismos como la ploidía, vías de señalización superpuestas o varias proteínas con la misma función, conocido como redundancia genética (Figura 2).

Estas características vuelven a este tipo de organismos más robustos mutacionalmente. El término de robustez mutacional, puede aplicarse a todos los organismos vivos y puede definirse como la capacidad del genotipo a mantener el fenotipo frente a la ocurrencia de mutaciones (Visser et al., 2003).

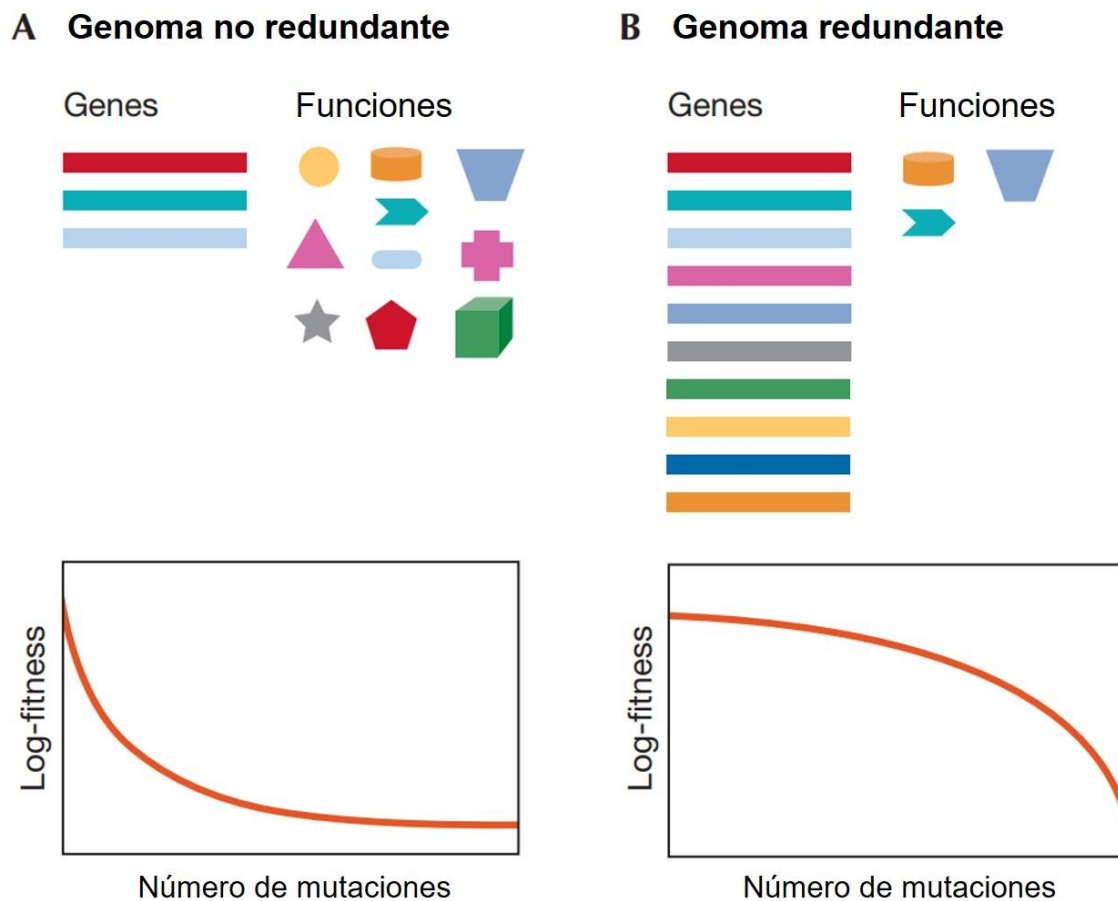


Figura 2. Redundancia genética. A) se ejemplifica la redundancia genética en organismos como virus y bacterias, quienes tienen pocos genes para muchas funciones por lo que a medida que aumenta el número de mutaciones, el fitness biológico disminuye rápidamente. B) Se observa el caso contrario, en la que hay muchos genes que tienen pocas funciones por lo que la cantidad de mutaciones que pueden acumular es mayor previo a comenzar a perder fitness. (Imagen tomada y traducida de Elena et al., 2006).

1.2.1 Volatilidad genómica

Dada la incapacidad de los virus a generar redundancia genética debido a que sus genomas se caracterizan por ser pequeños, compactos y no redundantes, se cree que los virus evolucionaron mediante redes de neutralidad en la que las progenies generadas tienen secuencias genéticas que permite que el fenotipo

no se vea afectado a pesar de las mutaciones. En otras palabras, evolucionaron manteniendo en sus genomas codones poco volátiles por lo que las mutaciones que sufren son más fáciles de amortiguar (**Figura 3**).

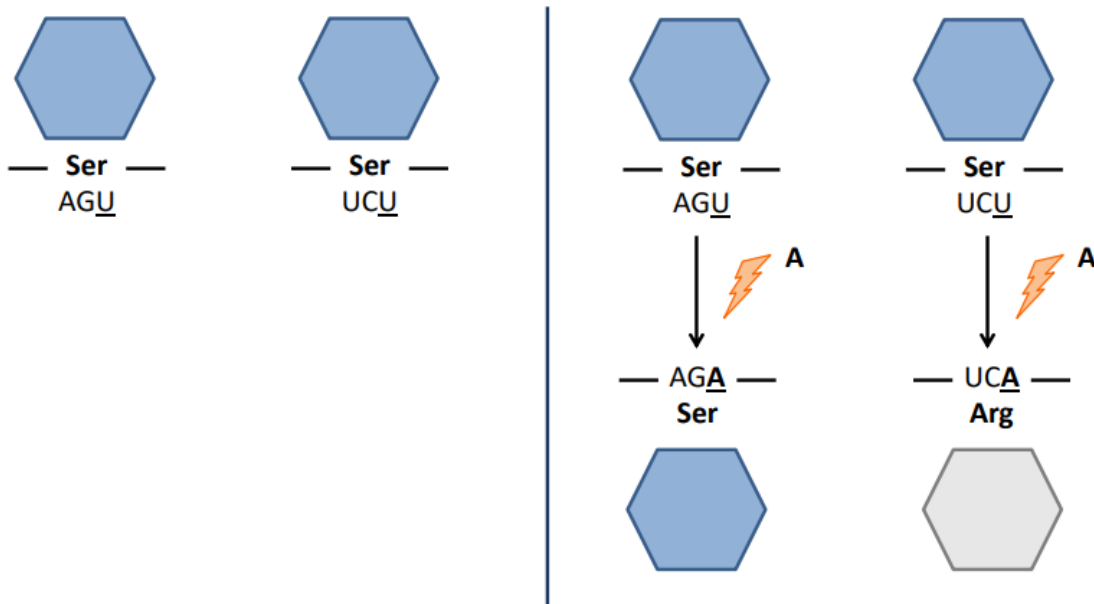


Figura 3. Impacto de las mutaciones en genomas con distinta volatilidad genética. En la parte izquierda se muestran dos codones sinónimos de serina. En la parte derecha se muestra el impacto de la misma mutación puntual en ambos codones. En uno la mutación es neutra mientras que en el otro codón la mutación conlleva un cambio aminoacídico por lo que el fenotipo cambiará. (Tomado de comunicación personal Moratorio&Carrau 2016).

La volatilidad genómica, o genética, se define como la suma de la volatilidad de los codones en la región codificante del genoma o del gen (Plotkin et al., 2004). La volatilidad codónica se define como la probabilidad de que una mutación puntual aleatoria en un codón genere un cambio no sinónimo (Zhang, 2005). La volatilidad de los codones se calcula a partir de la siguiente ecuación que fue extraída de Plotkin et al., 2004:

$$v(c) = \frac{1}{\text{no. de mut. puntuales vecinas}} \sum_{\text{mut. puntuales vecinas } c_i} D[\text{acid}(c), \text{acid}(c_i)],$$

en donde se suman los codones no stop, c_i , que pueden mutar en codones c por una mutación puntual. Ellos utilizan la medida más simple para medir D , la distancia de Hamming, que es cero si los dos aminoácidos son iguales y vale 1 si son distintos. En esta ecuación se asume que la tasa de mutación es igual

entre los codones, en otras palabras no se tiene en cuenta el sesgo transición/transversión.

Por lo que antecede, de los 20 aminoácidos solamente 4 son codificados por codones sinónimos con diferentes grados de volatilidad (Tabla 1). Estos aminoácidos son: serina, leucina, arginina y glicina.

Tabla 1. Volatilidad de cada uno de los codones del código genético. Tabla extraída y editada de Zhang, 2005. Recuadro azul: codones que codifican para leucina. Recuadros amarillos: codones que codifican para serina. Recuadros amarillos: codones que codifican para arginina. Recuadros rojos: codones que codifican para glicina.

TTT	Phe	0.889	TCT	Ser	0.667	TAT	Tyr	0.857	TGT	Cys	0.875
TTC	Phe	0.889	TCC	Ser	0.667	TAC	Tyr	0.857	TGC	Cys	0.875
TTA	Leu	0.714	TCA	Ser	0.571	TAA	Stop		TGA	Stop	
TTG	Leu	0.750	TCG	Ser	0.625	TAG	Stop		TGG	Trp	1.000
CTT	Leu	0.667	CCT	Pro	0.667	CAT	His	0.889	CGT	Arg	0.667
CTC	Leu	0.667	CCC	Pro	0.667	CAC	His	0.889	CGC	Arg	0.667
CTA	Leu	0.556	CCA	Pro	0.667	CAA	Gln	0.875	CGA	Arg	0.500
CTG	Leu	0.556	CCG	Pro	0.667	CAG	Gln	0.875	CGG	Arg	0.556
ATT	Ile	0.778	ACT	Thr	0.667	AAT	Asn	0.889	AGT	Ser	0.889
ATC	Ile	0.778	ACC	Thr	0.667	AAC	Asn	0.889	AGC	Ser	0.889
ATA	Ile	0.778	ACA	Thr	0.667	AAA	Lys	0.875	AGA	Arg	0.750
ATG	Met	1.000	ACG	Thr	0.667	AAG	Lys	0.875	AGG	Arg	0.778
GTT	Val	0.667	GCT	Ala	0.667	GAT	Asp	0.889	GGT	Gly	0.667
GTC	Val	0.667	GCC	Ala	0.667	GAC	Asp	0.889	GGC	Gly	0.667
GTA	Val	0.667	GCA	Ala	0.667	GAA	Glu	0.875	GGA	Gly	0.625
GTG	Val	0.667	GCG	Ala	0.667	GAG	Glu	0.875	GGG	Gly	0.667

1.2.2 Robustez mutacional de virus ARN

Para comprender las estrategias que utilizan los virus para ser más robustos es necesario primero introducir los conceptos de espacio de secuencia y de paisaje adaptativo. En biología evolutiva, el espacio de secuencia es una representación matemática de todas las combinaciones posibles para una secuencia de nucleótidos determinada (Baltscheffsky et al., 1986). Asimismo, los paisajes adaptativos (Wright, 1932) se utilizan para visualizar la relación entre genotipos y el fitness, donde los genotipos se representan en una o dos dimensiones y en otra dimensión se representa la eficacia biológica de cada genotipo como altura del paisaje. Los distintos picos, separados por valles, presentan diferentes alturas dependiendo de la eficacia de cada genotipo, siendo los picos más altos los de mayor eficacia (**Figura 4**). Esta teoría se pudo

comprobar mediante la utilización de organismos digitales (simulaciones bioinformáticas) (Wilke et al., 2001) pero aún no se han visto muchas aproximaciones experimentales precisas dada la alta dimensionalidad que presentan estos conceptos y a la dificultad de realizar evolución experimental en los laboratorios. Los virus ARN se presentan como buenos candidatos para este fin, tornándose más sencillo de estudiar que otros sistemas biológicos debido a que tienen elevadas tasas de mutación, cortos tiempos generacionales, elevados tamaños poblacionales y escasa complejidad genómica (Sanjuán Verdeguer, 2005).

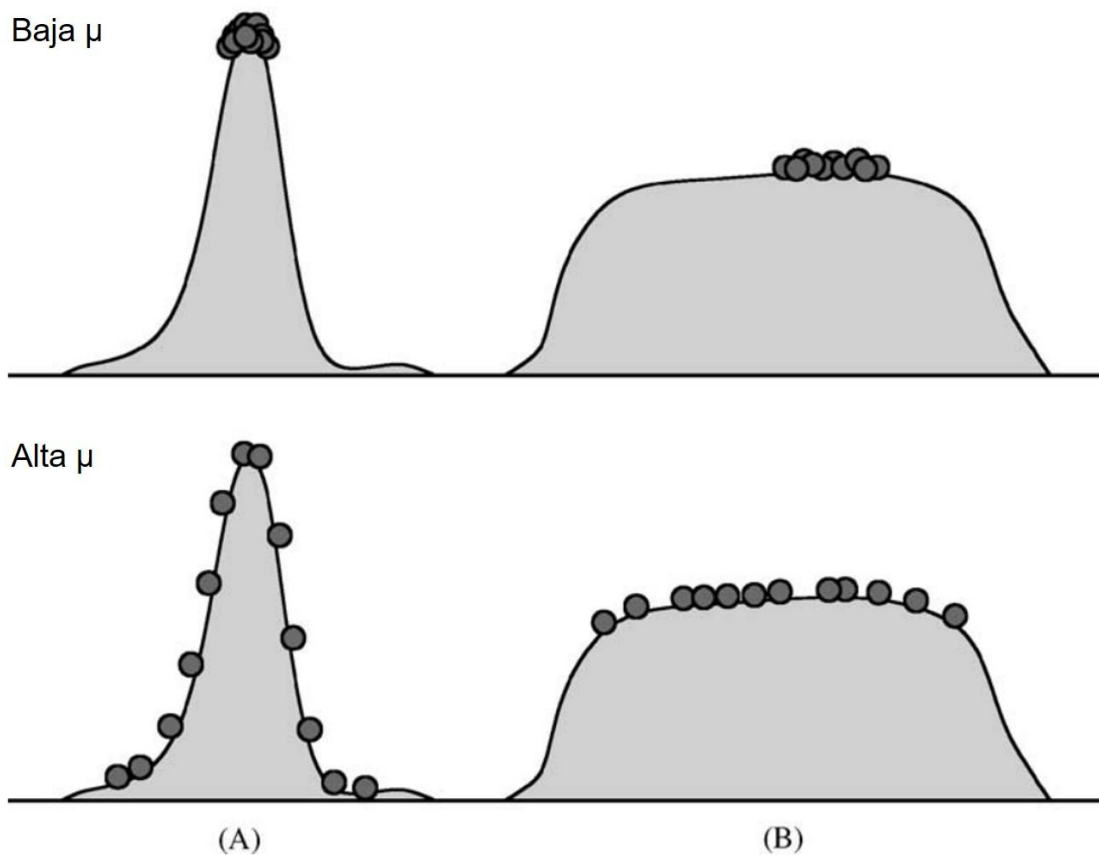


Figura 4. Supervivencia del más plano. En esta figura la altura de los picos representa el fitness viral. Cuando la tasa de mutación (μ) es pequeña, las poblaciones menos robustas (A) tienden a mantener su eficacia. Lo mismo sucede con las poblaciones más robustas (B). Sin embargo, cuando la tasa de mutación es alta, la población menos robusta pierde su fitness mientras que la población más robusta tiende a mantener su fitness. Tomado y adaptado de Wilke & Adami, 2003.

1.3 Coxsackie virus B3, nuestro modelo de estudio

1.3.1 Clasificación y patogenia

Los Coxsackievirus pertenecen al género Enterovirus dentro de la familia *Picornaviridae*. Estos virus se aislaron por primera vez durante un brote de poliomielitis en la ciudad de Coxsackie en el estado de Nueva York en 1947 al buscar virus en heces de niños infectados (Dalldorf & Sickles, 1948). El virus Coxsackie B3 (CVB3), al igual que muchos enterovirus, se caracteriza por replicarse en el tracto gástrico (Martino et al., 1994). La transmisión del mismo se produce por la vía fecal-oral y aunque la mayoría de las infecciones por CVB3 suelen ser leves, en algunos casos se han asociado con enfermedad febril, parálisis, pancreatitis, erupción cutánea, meningitis y miocarditis (Knipe & Howley, 2013).

1.3.2 Organización genómica y ciclo replicativo

Los miembros de la familia *Picornaviridae* se caracterizan por ser virus de ARN monocatenario de sentido positivo (6,7 a 10,1 kb), de tamaño pequeño (30-32 nm de diámetro), icosaédricos y sin envoltura (Zell et al., 2017). Particularmente, los Enterovirus presentan genoma de 7,5 kb y presentan la proteína Vpg (Viral Protein of the Genome) que está unida covalentemente al extremo 5' del ARN genómico viral (Flanegan et al., 1977). En la región 5'UTR contiene una región altamente estructurada denominada IRES (Internal Ribosome Entry Site), la cual permite el reclutamiento de las subunidades ribosomales promoviendo una traducción CAP independiente (Fitzgerald & Semler, 2009). A continuación del IRES se encuentra un único marco abierto de lectura (ORF) seguido de una cola poli-A en el extremo 3'UTR (**Figura 5**). La región codificante se caracteriza por tener 3 regiones, P1, P2 y P3. La región P1 codifica para las proteínas de la cápside (vinculadas a la estructura viral y la vehiculización de la partícula), mientras que P2 y P3 codifican polipéptidos precursores de las proteínas con funciones no estructurales que son necesarias para la morfogénesis y replicación viral.

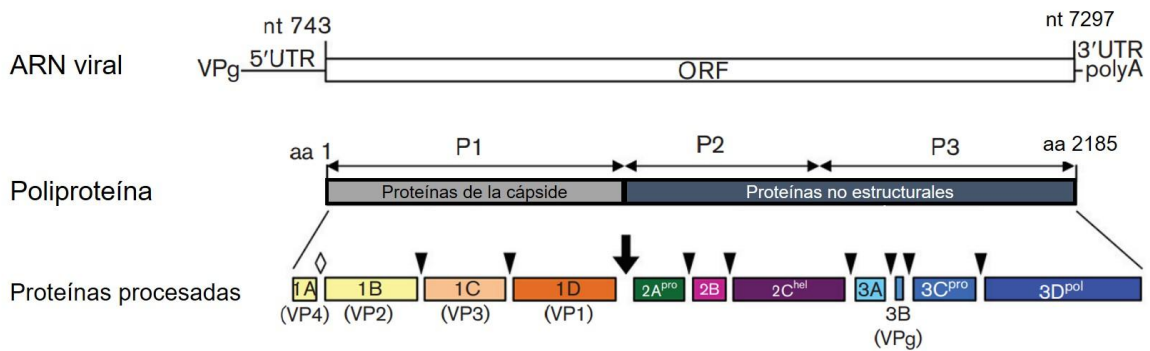


Figura 5. Estructura genómica de los Enterovirus. Este género es el más estudiado de la familia Picornaviridae debido a que el poliovirus (PV) pertenece a él. La similitud entre estos dos permite que lo aprendido PV se aplique en CVB3, teniendo en cuenta ciertas excepciones como son los receptores de entrada. Tomada, traducida y adaptada a CVB3 de Zell et al., 2017.

El ciclo replicativo de CVB3 comienza por la unión del virus al receptor CAR (receptor de coxsackievirus y adenovirus, por sus siglas en inglés) y correceptor DAF (factor acelerador de la descomposición, por sus siglas en inglés, que también es conocido como CD55) (Selinka et al., 2004). Algunos receptores celulares como el heparán sulfato podrían estar implicados en la adhesión del virus a las células huésped (Zautner et al., 2003). Debido a que el receptor CAR se encuentra en las uniones estrechas de las células epiteliales del intestino, CVB3 debe interactuar primero con DAF. La interacción con DAF produce una cascada de señalización que deriva en el movimiento de DAF a las uniones estrechas (Coyne & Bergelson, 2006; Shieh & Bergelson, 2002) y provoca la interacción de CVB3 con CAR para que sea internalizado por endocitosis. Tras su entrada al citoplasma el ARN viral es traducido como un gran polipéptido que luego se escinde en las proteínas estructurales y no estructurales individuales. Estos clivajes son realizados por las proteasas virales 2A y 3C (Badorff et al., 1999; Ehrenfeld, 1982). Mientras tanto, la traducción celular se corta, (Ehrenfeld, 1982). Luego de esa traducción inicial, la ARN polimerasa dependiente del ARN (proteína 3D), comienza la replicación del genoma, sintetizando en una primera instancia la hebra de ARN negativo la cual será utilizada como molde, también por la proteína 3D, para la generación de múltiples genomas progenies de cadena positiva (Garmaroudi et al., 2015).

Una vez producidos los genomas de ARN de cadena positiva y las proteínas estructurales, los mismos se ensamblarán para formar el provirión. La maduración del provirión a virión maduro ocurre cuando por autoclivaje de la VP0 se forman VP4 y VP2. Finalmente, el ciclo vital del virus se completa con la liberación de la progenie viral para la posterior infección de las células vecinas. Si bien se creía que esta liberación se producía de manera lítica, en la actualidad se demostró que también hay liberación viral asociada a membranas (Bou et al., 2019; Bou & Sanjuán, 2021).

2. Hipótesis

La volatilidad genómica y la robustez mutacional son dos caras de una misma moneda. Por lo que un virus que presente mayor grado de volatilidad genómica incorporará un mayor número de mutaciones bajo el efecto de pasajes en condiciones de cuello de botella lo cual impactará disminuyendo su fitness. Contrariamente, se espera que un virus que presente mayor robustez mutacional incorpore menos mutaciones y amortigüe su efecto deletéreo bajo estas mismas condiciones experimentales.

3. Objetivos

3.1 Objetivo General

El objetivo general de esta tesis es estudiar la relación entre la volatilidad genómica y la robustez mutacional de un enterovirus humano, CVB3, utilizando 3 poblaciones virales con diferente grado de volatilidad genómica teórica y misma secuencia aminoacídica. Se evaluará la capacidad de amortiguar las mutaciones de estas distintas poblaciones aplicando sucesivos cuellos de botella para luego evaluar el fitness de las poblaciones ancestrales e intentar determinar qué población tiene mayor robustez mutacional.

3.2 Objetivos específicos

- Objetivo específico 1:
Caracterización de las cepas virales a estudiar: virus salvaje (WT), virus de mayor volatilidad (MoreV) y virus de menor volatilidad (LessV).

- Objetivo específico 2:
Estudio del impacto de la deriva génica sobre las 3 variantes de volatilidad de CVB3 mediante la implementación de sucesivos cuellos de botella en su tamaño poblacional de modo de favorecer la fijación de mutaciones deletéreas. Se evaluará este impacto a través del estudio de subrogados del fitness viral: i) tamaño poblacional recuperado en cada pasaje, ii) fenotipo de calva y iii) cálculo del fitness relativo por ensayos de competencia.

- Objetivo específico 3:
Análisis de las bases moleculares de la evolución experimental mediante secuenciación de las poblaciones evolucionadas.

4. Materiales y métodos

4.1 Líneas celulares

Se trabajó con las líneas celulares Vero (riñón de *Cercopithecus aethiops*) y HeLa (cervix de *Homo sapiens*). Se mantuvieron en Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher) suplementado con 10% de suero fetal bovino (FBS; Thermo Fisher) y 1% de penicilina/estreptomicina (P/S; Thermo Fisher). Las células se mantuvieron a 37°C con 5% de CO₂. Para los pasajes celulares se retiró el medio, se lavó con PBS 1x (Sigma Aldrich), se agregó tripsina (Thermo Fisher) para despegar las células del frasco y tras 5 minutos a 37°C se neutralizó la tripsina con la misma cantidad de medio complementado. Finalmente se agregó medio para llevar la cantidad de células al volumen deseado. Las cantidades utilizadas para cada tamaño de frasco y de placas está en el Anexo 1.

4.2 Mutantes de CVB3

En este trabajo se utilizó como modelo viral el virus Coxsackie B3 (CVB3) Wild-type (Nancy strain) para el cual el laboratorio cuenta con el clon infeccioso. Este clon infeccioso fue modificado por ingeniería genética para obtener dos mutantes distintos con diferentes grados de volatilidad genética teórica que fueron generados por el Dr. Gonzalo Moratorio. Estos dos mutantes se caracterizan por tener su región P1 con mayor o menor grado de volatilidad y se los llamó MoreV y LessV respectivamente. Para obtener estos mutantes se reemplazaron los 117 codones que codifican para serina y leucina de la región P1 de CVB3 para que, en esos dos mutantes, estos aminoácidos se codifiquen únicamente a partir de los codones que se muestran en la **Figura 6**. Para estos reemplazos se basó en la matriz de similitud química de McLachlan (McLachlan, 1971) y en el marco matemático diseñado por Archetti (Archetti, 2009a, 2009b). Además de estos 3 virus hay un cuarto genotipo que se llama CVB3 de Referencia. Este se caracteriza por tener 4 mutaciones sinónimas en su polimerasa que permite diferenciarlo del resto de genotipos por RT-qPCR con sondas Taqman.

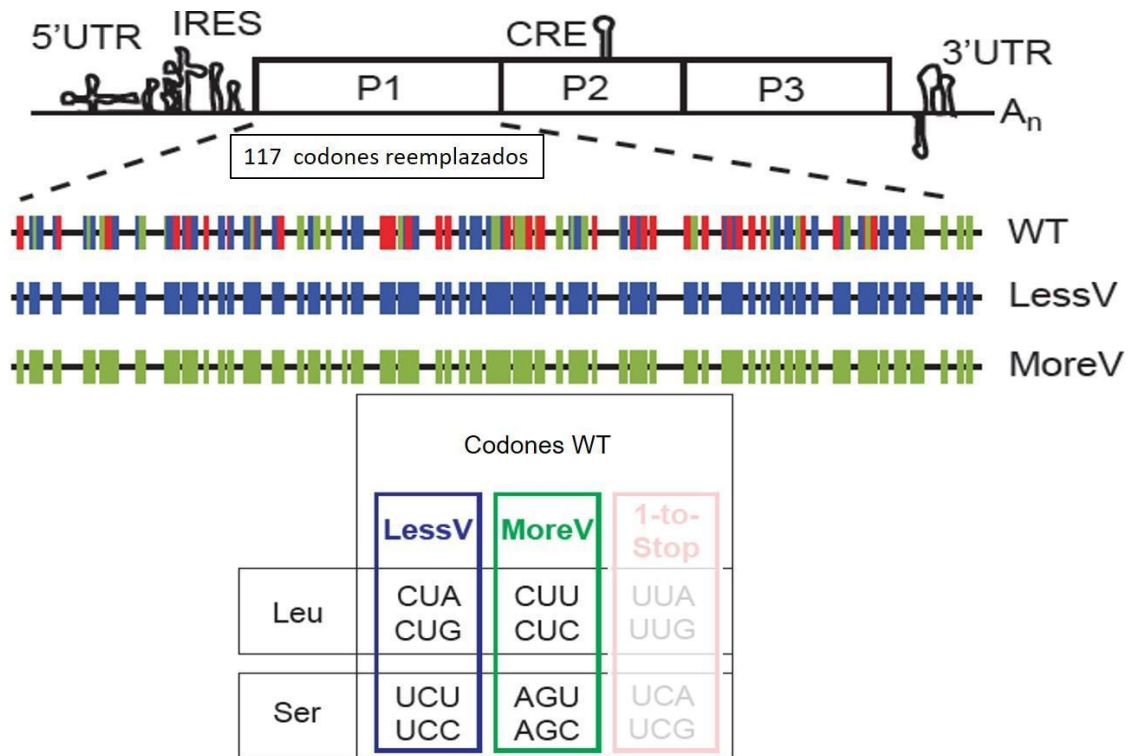


Figura 6. Mutantes de CVB3 más y menos volátiles (MoreV y LessV). En la parte superior de la imagen se muestra la estructura genómica de CVB3 y se muestra la región P1 amplificada con los codones sustituidos para generar los mutantes LessV y MoreV. En el recuadro de la parte inferior de la imagen se ven los codones que codifican para cada uno de estos aminoácidos en los mutantes. Figura realizada por el Dr. Moratorio y traducida para esta tesis.

4.3 Generación de stock virales CVB3

Las progenies 0 y 1 (p0 y p1, respectivamente) de CVB3 (WT, MoreV y LessV) fueron generadas por la Dra. Marianoel Pereira-Gómez previo al comienzo de esta tesis como se describe en el Anexo 2 siguiendo el esquema que se describe en la **Figura 7**. En este trabajo se procedió a generar, a partir de estas p1, la progenie 2 (p2) que fue el stock viral con el cual se realizaron todos los ensayos de esta tesis que a partir de ahora, en este documento, se llamarán ancestros. Para esto se infectó con la p1 un frasco T-175 de células HeLa con una confluencia de un 70%. Previo a la infección se retiró el medio de los frascos, se lavó con PBS 1x y se infectó con el volumen del criotubo (600 µL de p1). Se dejó incubar por una hora a 37°C y 5% de CO2 y se agregó DMEM con 2% de FBS y 1% de P/S. 48 horas luego de la infección se verificó la presencia

de efecto citopático al comparar estas células con un control que no había sido infectado.

Los ancestros virales fueron secuenciados por metodología Sanger para confirmar las mutaciones en la región P1 de los 3 mutantes. Posteriormente se realizó la secuenciación del genoma completo por la tecnología de Nanopore a fin de confirmar que el resto del genoma no tuviera mutaciones y descartar la existencia de contaminación cruzada entre mutantes y WT.

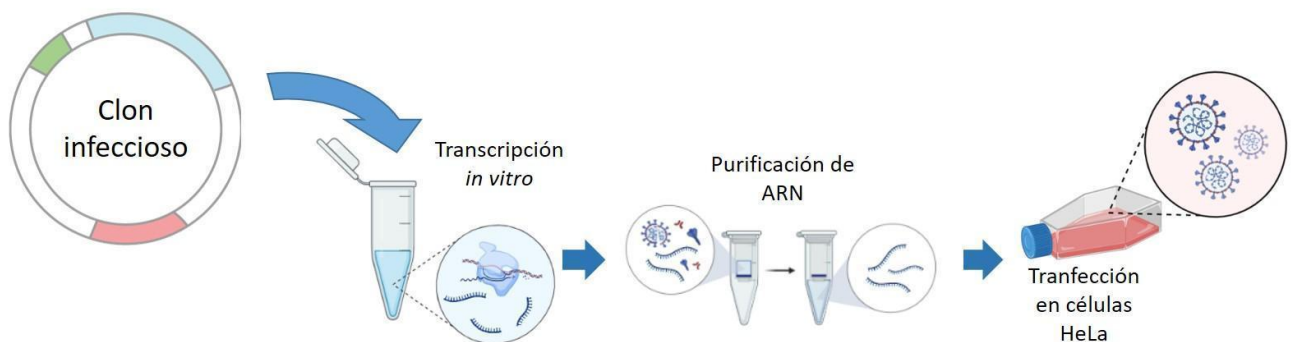


Figura 7. Generación de los ancestros virales. A partir de los clones infecciosos de cada virus se realizó una transcripción *in vitro* que se purificó y luego se transfectó en células HeLa. El éxito de la transfección se verificó evaluando el efecto citopático en la monocapa celular.

4.4 Titulación viral

Con el fin de estimar el título viral se realizó la titulación mediante dos abordajes: plaque assay y median Tissue Culture Infectious Dose (TCID₅₀) (Figura 8).

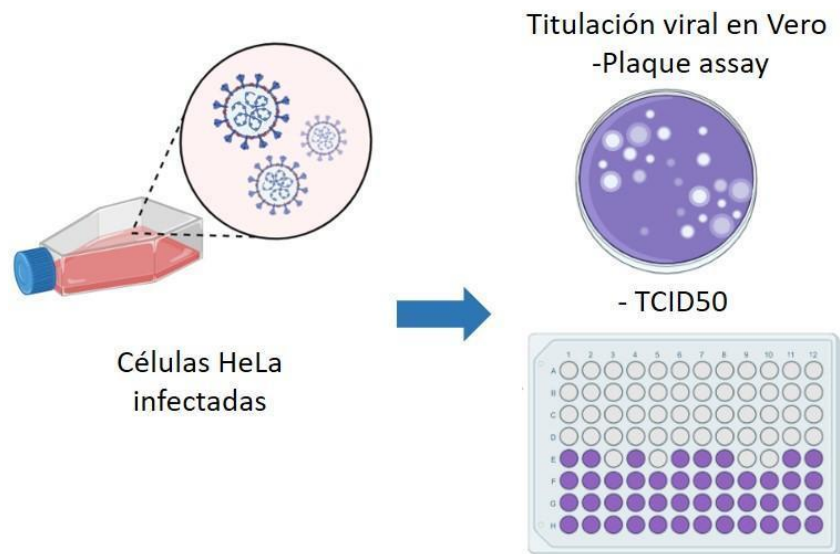


Figura 8. Metodología de titulación viral utilizadas. Durante esta maestría se realizaron titulaciones mediante dos metodologías, *plaque assay* y TCID₅₀.

4.4.1 Plaque assay

Los ensayos de placa son el método estándar que se utiliza para determinar el título viral, o lo que se conoce como dosis infecciosa. Con el fin de realizar este tipo de ensayos previamente se sembraron células Vero en placas de 12 pocillos (3×10^6 por placa) (**Figura 8**). Estas células fueron inoculadas con diluciones seriadas en base 10 de los diferentes virus. Para esto se utilizaron 100 μ L de cada dilución para infectar la monocapa celular y posteriormente se incubó por una hora a 37°C. Luego de la incubación se agregó una mezcla de 1 parte de agarosa al 4%, 1 parte de DMEM con 10% FBS y 1% P/S y 3 partes de DMEM solo. Dos días post infección se fijaron las células con formaldehído al 4% por 30 minutos, se retiraron los botones de agarosa y se realizó la tinción con cristal violeta. Luego de la tinción se enjuagaron con agua y se dejaron secar para luego contar las unidades formadoras de placas (PFU). Para el cálculo se utilizó la siguiente fórmula:

$$PFU/mL = \frac{N^{\circ} \text{ promedio de calvas de lisis en una dilución}}{\text{Dilución} \times \text{Volumen de dilución sembrado}}$$

4.4.2 TCID₅₀

Para el TCID₅₀, el día previo a la titulación, se sembraron células Vero en placas de 96 pocillos (1x10⁶ células por placa) (100 µL de DMEM con 10% de SFB que contenían 1 x 10⁴ células en cada pocillo) (**Figura 8**). Estas células fueron inoculadas con diluciones seriadas en base 10 de los diferentes virus (entre la dilución -1 y -8). Posteriormente se inocularon en cada pocillo 100 µL de cada una de las 8 diluciones (12 réplicas). Cinco días post infección las células fueron fijadas con formaldehído al 4% por 30 minutos y las monocapas se tiñeron con cristal violeta. Se enjuagaron en agua, se dejaron secar y se calcularon los valores de TCID₅₀.

El título viral por esta metodología se calculó por el método de Reed-Muench y que consiste en lo siguiente:

- 1) Calcular la proporción de las 12 réplicas que están infectados para cada dilución.
- 2) Calcular la distancia proporcional (DP) que se calcula a partir de la siguiente ecuación:

$$DP = \frac{(\% \text{ de positivos mayor al } 50\%) - 50\%}{(\% \text{ de positivos mayor al } 50\%) - (\% \text{ de positivos menor al } 50\%)}$$

- 3) Calcular el log TCID₅₀ con la siguiente fórmula:

$$\log \log TCID_{50} = (\log \log \text{ de la dilución mayor al } 50\%) - (PD \times \text{factor de dilución})$$

- 4) El valor de TCID₅₀ se calcula de la siguiente manera:

$$TCID_{50} = 10^{\log TCID_{50}}$$

- 5) Finalmente calcular el TCID₅₀/mL. Dado que se utilizan 0.1 mL para la siembra entonces el cálculo final será:

$$TCID_{50}/mL = \frac{TCID_{50}}{0.1 mL}$$

4.5 Extracción de ARN

Con el fin de realizar la extracción de ARN se utilizaron 3 metodologías distintas dependiendo del objetivo para el cual fuera utilizado el ARN posteriormente.

4.5.1 Extracción con TRIzol

Para la confirmación de las mutaciones en la región P1 de CVB3 se extrajo el ARN viral con el protocolo de TRIzol que se encuentra en el Anexo 3.

4.5.2 Extracción con kit Qiagen QIAamp® Viral RNA

Con el fin de secuenciar el genoma completo de CVB3 por metodología Oxford Nanopore, la extracción de ARN se realizó utilizando el kit Qiagen QIAamp® Viral RNA. El protocolo seguido para la extracción de estas muestras se encuentra en el Anexo 4.

4.5.3 Extracción de ARN con robot PurePrep 96 (Molgen)

El robot de extracción fue utilizado para extraer las muestras provenientes de los ensayos de competencia dada su uniformidad en la extracción de las mismas. El protocolo de extracción utilizado se puede ver en el Anexo 5.

4.6 Retrotranscripción (RT)

Se realizaron dos estrategias alternativas para la síntesis de ADNc:

4.6.1 SuperScript III

Con el fin de secuenciar los ARN virales por Sanger, se realizó la transcripción reversa de los mismos utilizando la enzima SuperScript III (Invitrogen), siguiendo las instrucciones del proveedor. Esta transcripción reversa fue realizada con un primer específico, el primer 4018Rev, que hibrida en la región P2 de CVB3, a una concentración final de 0.1 μ M.

4.6.2 LunaScript RT SuperMix Kit (New England BioLabs)

Por otra parte, con el fin de obtener el genoma completo de CVB3 para luego secuenciar por Nanopore se utilizó la LunaScript RT SuperMix Kit (New England BioLabs). Se utilizaron 2 µL de SuperMix para cada 8 µL de ARN. La reacción se llevó a cabo en termociclador y consistía en 3 ciclos, el primero a 25°C por 2 minutos, el segundo paso a 55°C por 10 minutos y finalmente se inactivó la enzima a 95°C por 1 minuto.

4.7 Amplificación del genoma de CVB3 por PCR a tiempo final

4.7.1 Amplificación de la región P1

La amplificación de esta región se utilizó para verificar que cada uno de los virus utilizados tenían las mutaciones correspondientes en la región estructural. La PCR se realizó bajo las siguientes condiciones:

	1x	<u>CICLADO</u>
5x Buffer Phusion	10 µL	98°C - 30 segundos - 1 Ciclo
dNTP's (10 mM)	0.75 µL	98°C - 10 segundos
Seq2For (20 µM)	1 µL	65°C - 30 segundos - 25 Ciclos
3573Rev (20 µM)	1 µL	72°C - 1.5 minutos
H2O DEPC	33.5 µL	72°C - 10 minutos - 1 ciclo
Phusion	0.25 µL	
ADNc	4 µL	

La secuencias de los primers utilizados fueron las siguientes:

Seq2For 5'-TGGCCATCCGGTGACTAATAGA-3',

3573Rev 5'-CTACTAGACCTGGTCCTTCAAACGA-3'. El primer Seq2For hibrida en el 5'UTR mientras que el 3573Rev hibrida en el comienzo de la región P2.

4.7.2 Amplificación del genoma completo de CVB3

El genoma completo de CVB3 se amplificó por PCR bajo las siguientes condiciones:

	1x	<u>CICLADO</u>
5x Buffer Phusion	5 µL	98°C - 30 segundos - 1 Ciclo
dNTP's (10 mM)	0.5 µL	98°C - 10 segundos
120For (10 µM)	0.5 µL	55°C - 30 segundos - 35 Ciclos
7430Rev (10 µM)	0.5 µL	72°C - 6 minutos
H2O DEPC	16.25 µL	72°C - 10 minutos - 1 Ciclo
Phusion	0.25 µL	
ADNc	2 µL	

Las secuencias de los primers utilizados fueron las siguientes:

120For 5'-ACCCCCTCCCCCACCTGTAA -3',

7430Rev 5'-CCGCACCGAATGCGGAGAATTTACC-3'.

El primer 120For hibrida en la región 5'UTR de CVB3 mientras que el primer 7430Rev hibrida en la región 3'UTR justo previo a la cola poli-A.

4.8 Ensayos de cinética replicativa

Las curvas de crecimiento fueron realizadas en la línea celular HeLa a una alta y baja multiplicidad de infección (MOI) (3 y 0.01). Para la MOI=3 se sembraron 5×10^4 células por pocillo en placa de 24 pocillos. El día de la infección se prepararon diluciones en DMEM solo para infectar con 1.5×10^5 PFU por pocillo. Para la MOI=0.01 se sembraron 3.5×10^5 células por pocillo para infectar con 3.5×10^3 PFU. El día del ensayo se retiró el medio, se lavó con PBS 1x y se infectaron los pocillos por 1 hora a 37°C. Luego de la incubación se agregó DMEM con 2% de FBS y 1% de P/S. Para cada tiempo de recolección se realizaron 3 ciclos de congelado-descongelado y se clarificaron los sobrenadantes. En ambos casos los tiempos evaluados fueron 4, 8, 12 y 24 horas post infección. Además para la curva de MOI=3 se evaluó el tiempo 30

horas mientras que para la curva de MOI=0.01 se evaluó el crecimiento a las 36, 48, 72 y 96 horas. Las curvas se realizaron por triplicados independientes.

En todos los ensayos se titularon las diluciones virales de partida para confirmar que hubiera una cantidad de virus similar a la esperada.

Los títulos obtenidos de la curva fueron graficados en GraphPad Prism 9 y a partir de los mismos se realizó una regresión no lineal para crecimientos logísticos. El modelo que utiliza el software se comporta acorde a la siguiente ecuación:

$$y = \frac{Y_M x Y_0}{(Y_M - Y_0) x \exp(-k*x) + Y_0}, \text{ (Ec. 1)}$$

donde Y_0 es la población inicial, Y_M es la población máxima, k es la constante de crecimiento y $1/k$ es el punto de inflexión de la curva de crecimiento. A partir de la regresión no lineal se obtuvo la tasa de crecimiento para cada una de las réplicas de las curvas de crecimiento (9 en total). Con estas tasas se realizó un t-test para comparar si habían diferencias significativas (p -valor<0.05) entre las distintas tasas de crecimiento.

4.9 Pasajes de calva en calva

Para realizar la evolución experimental en la que la deriva génica fuera la mayor fuerza evolutiva se realizaron 10 pasajes seriados de calva en calva por triplicado para los virus WT, MoreV y LessV (**Figura 1**). Los pasajes se realizaron en células Vero y se sembraron, el día previo, 3.5×10^5 células por pocillo en placas de 6 pocillos. Para el pasaje 1 se prepararon diluciones de los ancestros virales como para tener 20 calvas por pocillo. La infección se realizó con 100 μ L de esas diluciones durante 1 hora. Luego de esa hora se agregaron 2.5 mL medio semisólido (DMEM con 0.8% de agarosa, 2% de FBS y 1% de P/S) por 48 horas. Posteriormente se observaron las calvas a trasluz y se seleccionó aleatoriamente una calva. Para tomar la población viral generada de esa calva se picó el medio semisólido con un tip de 20 μ L, se retiró el medio semisólido y luego se procedió a tomar 5 μ L de DMEM (de un tubo previamente

preparado que tenía un volumen total de 50 μL) y se resuspendió con la micropipeta la calva en esos 5 μL en 3 ocasiones.

En cada pasaje realizado se tituló el pasaje anterior. Los títulos por calva se graficaron utilizando el GraphPad Prism 9.

4.10 Ensayos de competencia

Los ensayos de competencia relativa se realizaron en la línea celular HeLa realizando una competencia directa de los virus de interés con el CVB3 Referencia (**Figura 9**). Se sembraron en una placa de 96 pocillos 1×10^6 células por placa el día previo. El día del ensayo se retiró el medio de cada pocillo y se lavó con PBS 1x. Luego se infectó con 10 μL de dilución viral en los que se esperó tener 100 PFU de virus de interés y con 10 μL de virus de referencia (en proporción 1:1 de los virus). La adsorción duró una hora, se retiró todo el medio con virus y se agregaron 100 μL de DMEM con 2% FBS y 1 % de P/S. Se evaluó la competencia 24 horas post infección, se recogieron los sobrenadantes y se extrajo el ARN por las metodologías previamente descritas. La proporción de cada virus se determinó por RT-pPCR utilizando sondas Taqman con distintos fluoróforos para los virus de interés y el virus de referencia (ver siguiente sección). El fitness relativo se determinó por el método descrito en Carrasco et al, 2007 y por lo que se calculó de la siguiente manera:

$$W = \frac{R(t)}{R(0)},$$

donde W es el fitness relativo,

$$R(0) = \frac{\text{Genomas de virus de interés al tiempo } 0}{\text{Genomas de virus Ref al tiempo } 0},$$

$$R(t) = \frac{\text{Genomas de virus de interés al tiempo } t}{\text{Genomas de virus Ref al tiempo } t}.$$

En los casos en el que el W es mayor a 1, el virus de interés será más eficiente que el virus de Ref. En caso de que W sea menor a 1, ocurrirá lo contrario. Si es igual a 1 tendrán el mismo fitness.

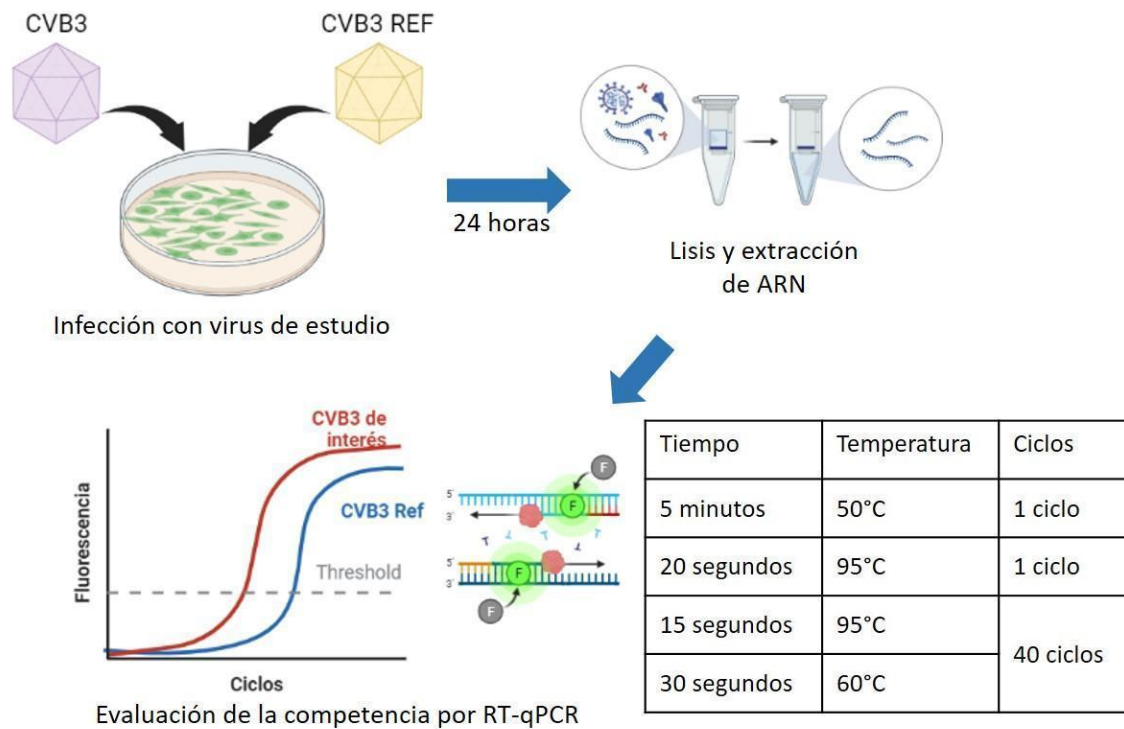


Figura 9. Ensayos de competencia y RT-qPCR. Para estos ensayos se siembra el virus a estudiar y el virus de Referencia en proporción de 1:1. Se infecta por 24 horas y luego de la infección se realizan 2 ciclos de congelado-descongelado para luego extraer el ARN con robot. El ARN extraído se utiliza para ser cuantificado por RT-qPCR y así poder calcular el fitness relativo para los virus de estudio.

4.11 RT-qPCR para ensayos de competencia y cuantificación de genomas

Se realizó la puesta a punto de una técnica de RT-qPCR con el fin de poder cuantificar genomas virales. En particular se utilizó para cuantificar los genomas producidos durante los ensayos de competencia y durante las curvas de crecimiento.

Tanto la curva estándar para los ensayos de competencia como para cuantificar genomas se realizaron por triplicados en un equipo QuantStudio 7.

4.11.1 RT-qPCR múltiple para ensayos de competencia

Se desarrolló una RT-qPCR que utiliza los primers, forward 5'-GATCGCATATGGTGATGATGTGA-3' y el reverso

5'-AGCTTCAGCGAGTAAAGATGCA-3', que amplifican un fragmento de la región P3 del genoma viral. Además se utilizaron dos tipos de sondas:

- La sonda MGB_CVB3_WT (5'FAM-CGCATCGTACCCATGG-3', la cual sirve para detectar los genomas del virus WT, MoreV y LessV,)
- La sonda MGB_CVB3_REF (5'VIC-CGCTAGCTACCCATGG-3') que tiene 4 mutaciones silenciosas que permiten diferenciar al virus de referencia del resto de los virus.

Las mezclas de reacción para los ensayos de competencia se hicieron a un volumen final de 20 μ L conteniendo 5 μ L de ARN, 1.8 μ L de cada uno de los primers (10 μ M), 1.5 μ L de la sonda concentración 2 μ M y 5 μ L de Fast 1-Step Mix 4x (Thermo Fischer). Este protocolo se basó en el descrito previamente en Moratorio et al., 2017.

Previo a realizar la curva estándar se procedió a verificar que no hubiera detección cruzada entre sondas y genomas WT y Referencia (**Figura 10**). Luego se realizaron las curvas estándar para poner a punto los ensayos de competencia, utilizando diluciones de concentración conocida de ARNs generados mediante IVT de los virus WT y de Referencia que había sido previamente generados por la Dra. Marianoel Pereira-Gómez. Con este fin se prepararon diluciones para tener una concentración de 1×10^9 genomas de cada virus en 5 μ L y a partir de esta se realizaron 8 diluciones seriadas de uno en diez, por lo que la concentración de ARN de cada virus de interés se pudo calcular con alta especificidad.

A partir de los resultados obtenidos se realizó una gráfica del Ct en función del logaritmo del número de copias (**Figura 11**). A partir de esta gráfica se realizó una regresión lineal (cuya función es $y=ax+b$).

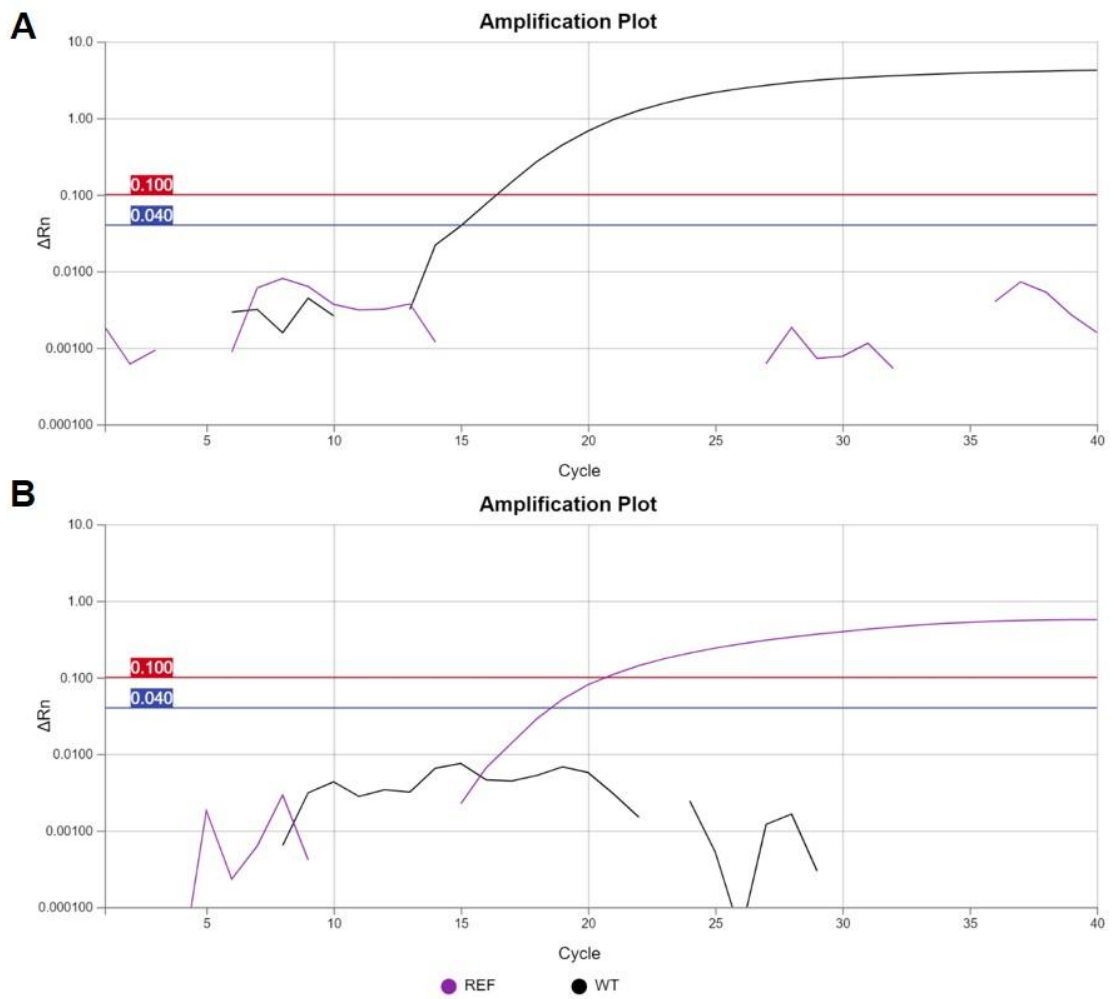


Figura 10. Verificación de detección cruzada entre las sondas específicas para los genomas WT y Referencia. A) se muestra que el mix con sonda WT detecta al genoma WT (en negro) pero no al Referencia (en violeta). B) se muestra que la sonda REF es específica para Referencia y no detecta al genoma WT.

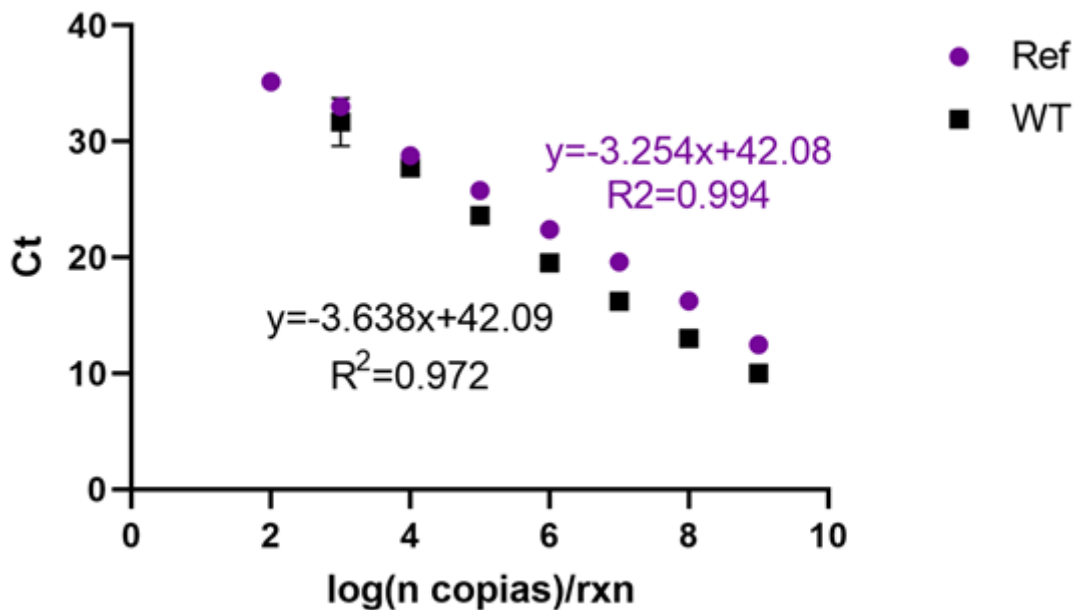


Figura 11. Curva estándar para los ensayos de competencia. Se graficaron los Ct en función del logaritmo de número de copias por reacción. En negro se muestran los valores obtenidos para los genomas WT en multiplex mientras que en violeta se muestran los resultados para los genomas de Referencia.

A partir de la regresión lineal para los genomas WT se obtuvo la siguiente ecuación de la recta: $y = -3.638x + 42.09$ con un $R^2 = 0.972$. El límite de detección es de 1000 genomas por reacción dado que en la última dilución (que no se muestra en la **Figura 11**) se detectó señal solamente en 2 de las 3 réplicas.

Para los genomas de Referencia se obtuvo la siguiente ecuación: $y = -3.254x + 42.08$ con un $R^2 = 0.994$. Se llegaron a detectar como mínimo 100 copias del genoma de Referencia por reacción.

4.11.2 RT-qPCR para cuantificar genomas

Para la cuantificación de genomas de WT, MoreV y LessV se utilizó el mismo protocolo que el anterior con la excepción que la sonda que detecta al CVB3 de Referencia fue sustituida con H₂O DEPC. Para la curva estándar se utilizaron 8

diluciones seriadas a partir de un stock que contenía 2×10^9 genomas cada 5 μL . Los ciclados para ambos ensayos aparecen en la **Figura 9**.

A partir de los resultados obtenidos se realizó una gráfica del Ct en función del logaritmo del número de copias (**Figura 12**). A partir de esta gráfica se realizó una regresión lineal (cuya función es $y=ax+b$).

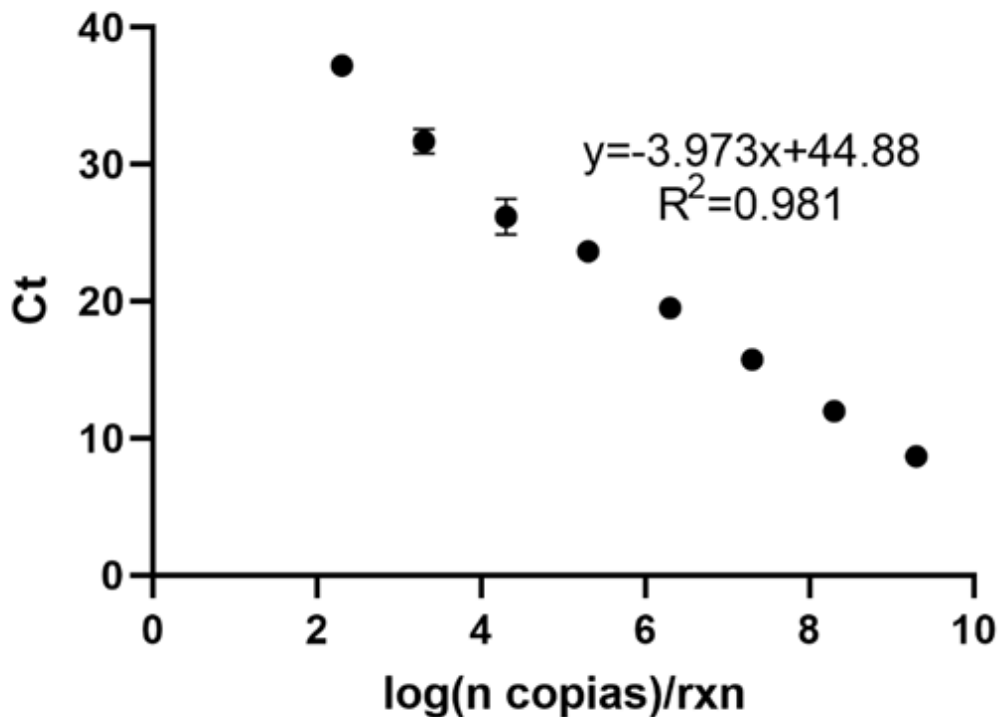


Figura 12. Curva estándar para los genotipos WT, MoreV y LessV. Se muestran los valores promedio de Ct con sus respectivos SEM.

Tras realizar la regresión lineal se obtuvo que la recta que ajustaba mejor tenía la siguiente ecuación: $y=-3.973x+44.88$ con un $R^2=0.981$. Se aprecia que la linealidad se conserva hasta la última dilución utilizada por lo que podemos detectar hasta 200 copias por reacción.

4.12 Medición del tamaño de calvas de las poblaciones evolucionadas

Los ensayos para comparar el tamaño de las calvas generadas por cada población evolucionada se realizaron en placas de 6 pocillos. El día previo al ensayo se sembraron placas con 3×10^6 células por placa (5×10^5 células por

pocillo). El día del ensayo se prepararon diluciones virales como para obtener unas 30 PFU de estas poblaciones en 100 μ L y se realizó la infección por una hora a 37°C. Luego de la incubación se agregó medio semisólido, DMEM con 2% de FBS, 0.8% de agarosa y 1% P/S. 48 horas después de la infección se fijaron las células con formalina, se retiró el botón de medio semisólido y se tiñó con cristal violeta. Las placas de 6 pocillos fueron fotografiadas en el equipo Amersham ImageQuant 800 (General Electric). El tamaño de las calvas se cuantificó utilizando el software ImageJ (<http://rsbweb.nih.gov/ij>) y se utilizó un macro (serie de comandos para automatizar el análisis en el ImageJ) generado por el Lic. Tabaré de los Campos de la Unidad de Bioimagenología Avanzada (UBA) del Institut Pasteur de Montevideo.

4.13 Next Generation Sequencing

Los genomas completos de los ancestros de CVB3 se secuenciaron en una plataforma GridION (Oxford Nanopore Technologies, United Kingdom) (**Figura 13**). La extracción de ARN, retrotranscripción y PCR del genoma completo fueron realizadas como se describe en los protocolos anteriores. Con los productos de PCR se generaron las librerías siguiendo los pasos mostrados en el Anexo 6.

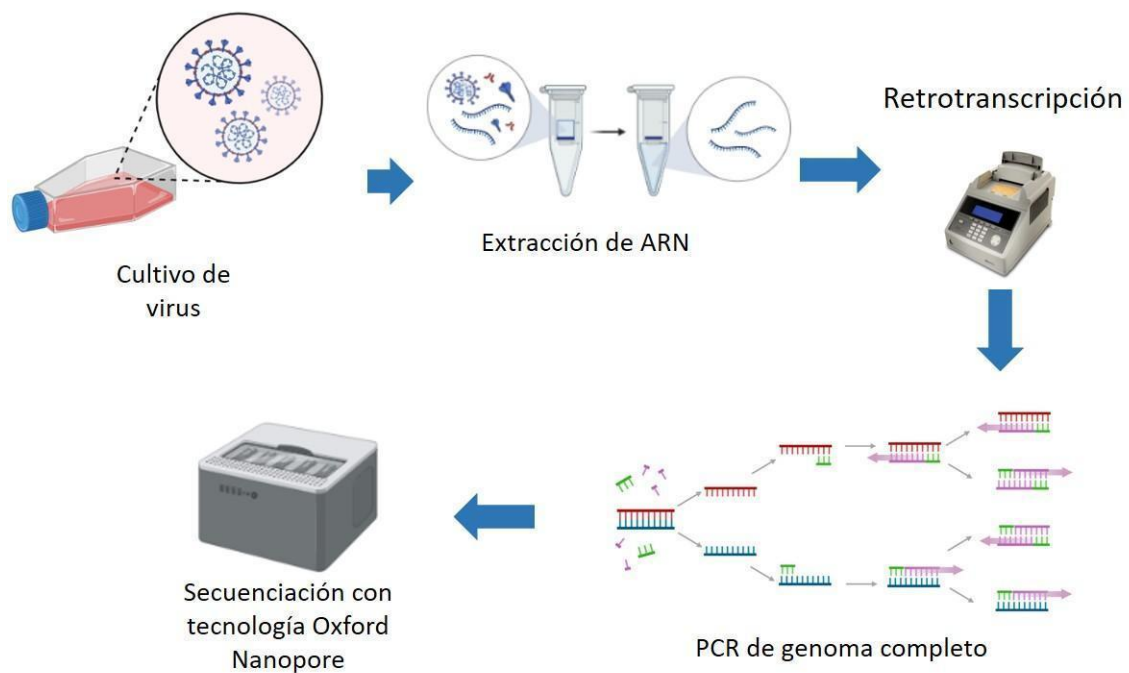


Figura 13. Secuenciación de los genomas. Para la secuenciación del genoma completo de CVB3 se realizó una extracción de ARN a partir de los cultivos virales. Esta extracción fue retrotranscrita y se realizó una PCR para amplificar el genoma completo en un solo fragmento. A partir de este producto de PCR se generaron las librerías para la secuenciación por tecnología Oxford Nanopore.

5. Resultados

5.1 Generación y titulación de los Virus CVB3 WT, MoreV y LessV

Con el fin de obtener los stocks virales (ancestros) de CVB3: i) WT; ii) el mutante de mayor volatilidad (MoreV) y iii) el mutante de menor volatilidad (LessV) se partió de progenies virales de primera generación (p1) de cada uno de estos virus, las cuales se encontraban previamente en el laboratorio (ver Sección 4.3). Es importante resaltar que los dos mutantes, MoreV y LessV, se caracterizan por tener su región P1 con mayor o menor grado de volatilidad respectivamente. Para obtener estos mutantes se reemplazaron los 117 codones que codifican para serina y leucina por codones sinónimos de ambos aminoácidos que difieren en su grado de volatilidad (ver **Figura 6**). De esta forma se obtuvieron las progenies virales de segunda generación (p2) con las cuales trabajamos a lo largo de toda la tesis. Recordemos que estas progenies serán llamadas ancestros. Estos ancestros fueron evaluados en una primera instancia en su capacidad de generar efecto citopático (CPE) en la línea celular HeLa. En la **Figura 14** se observa el efecto citopático generado por el virus wt (panel B), virus LessV (panel C) y el virus MoreV (panel d) en relación al mock (panel A) donde las células se evidencian totalmente adheridas a la placa y con morfología normal.

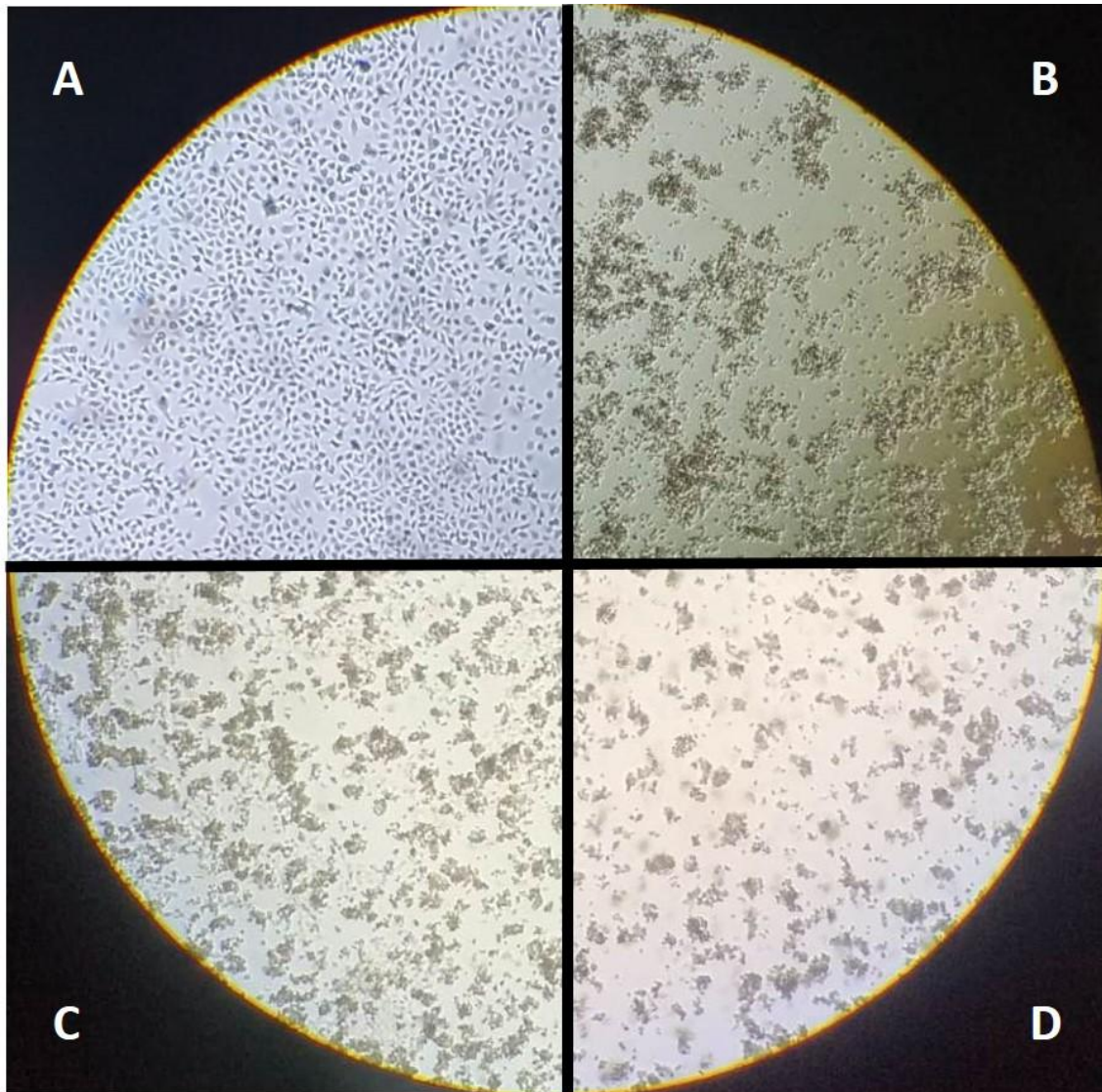


Figura 14. Efecto citopático observado en la línea celular HeLa post 48 horas de la infección con el p1 durante la generación de los ancestros virales (p2). A) MOCK, control negativo de infección, B) CPE generado por el virus WT; C) CPE generado por el virus LessV y D) CPE generado por el virus MoreV.

Una vez obtenido los ancestros virales los mismos fueron titulados utilizando la técnica de plaque assay. La titulación de estos ancestros resulta fundamental dado que son la base de partida de todos los experimentos realizados a lo largo de esta tesis.

Los resultados de la titulación de las diferentes variantes de volatilidad de CVB3 se realizó en triplicados biológicos y se muestran en la Tabla 2.

Tabla 2. Cálculo de títulos de los ancestros virales. Las unidades del título viral son unidades formadoras de placas/mililitros (PFU/mL).						
Genotipo	Réplicas	Dilución	Calvas	Título viral	Título viral	SEM
WT	4-A	-4	73	7.30E+06	7.80E+06	3.21E+05
	4-B		77	7.70E+06		
	4-C		84	8.40E+06		
MoreV	4-A	-4	74	7.40E+06	7.83E+06	3.38E+05
	4-B		85	8.50E+06		
	4-C		76	7.60E+06		
LessV	4-A	-4	43	4.30E+06	4.60E+06	5.13E+05
	4-B		56	5.60E+06		
	4-C		39	3.90E+06		
Ref	4-A	-4	89	8.90E+06	8.90E+06	5.77E+04
	4-B		88	8.80E+06		
	4-C		90	9.00E+06		

Para todos los virus la dilución que se utilizó para cuantificar el título viral fue la dilución -4, por lo que los títulos de los 4 genotipos estaban dentro del mismo orden.

Por último se realizó la secuenciación de los genomas completos de cada uno de mis ancestros virales logró obtener las secuencias completas de los genomas virales mediante la tecnología GridION, verificando que las secuencias eran idénticas a los clones infecciosos de partida.

5.2 PCR a tiempo final para secuenciación

Por un lado se logró la optimización de una PCR a tiempo final para amplificar la región P1 del genoma viral de forma completa (ver **Figura 6**) (única región que difiere en secuencia entre los tres virus). El producto de esta amplificación fue de unas 2.9 kb.

Por otra parte, a fin de poder secuenciar los genomas virales por la plataforma GridION, se puso a punto la amplificación del genoma completo de CVB3 en un solo fragmento de 7,3 Kb.

En la **Figura 15** se puede observar el producto de amplificación de unos 7300 pb, aproximadamente, correspondientes a los genomas completos de los virus CVB3 WT, MoreV y LessV.

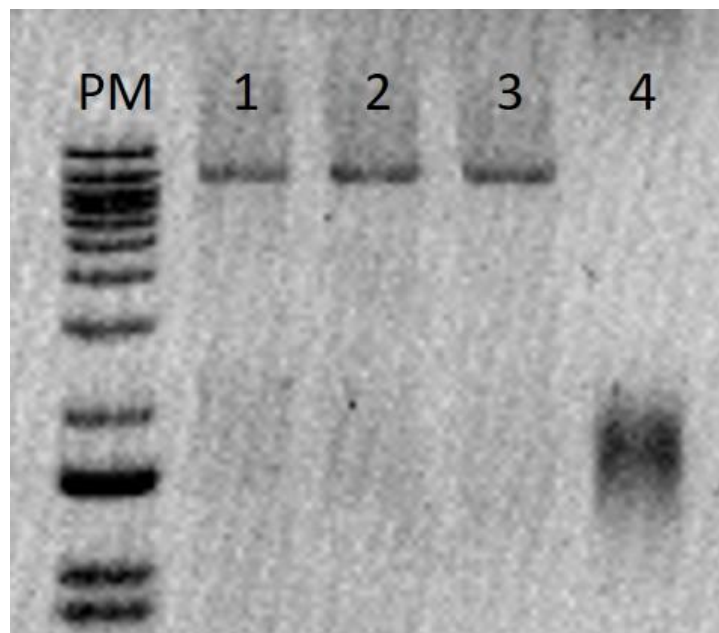


Figura 15. Electroforesis en gel de agarosa de los productos de amplificación del genoma completo de CVB3. Para confirmar la correcta amplificación de esta PCR a tiempo final se realizó una electroforesis en gel de agarosa al 1%. Carril 1: virus WT, carril 2 virus MoreV, carril 3 virus LessV, carril 4 es el control negativo de la PCR. No se utilizó control positivo. El peso molecular (PM) utilizado es 1kb Plus de ThermoFischer.

5.3 Ensayos de cinética de replicación de los ancestros de CVB3

Se realizaron ensayos de cinética de replicación en células HeLa para demostrar que los cambios en los codones de los dos genotipos generados (MoreV y LessV) no afectaban el fitness de los mismos en relación al genotipo WT.

Para esto se realizaron dos curvas de crecimiento, una curva de un solo paso con una multiplicidad de infección (MOI) de 3 y otra curva de múltiples pasos con una MOI de 0.01.

5.3.1 Cinética de replicación a MOI=3

Mediante estos ensayos se estudió el comportamiento de los virus a una alta MOI. En este caso, la MOI=3 implica que hay 3 virus por cada célula a infectar. Tras haber titulado por *plaque assay* cada uno de los tiempos y las réplicas, se graficaron los datos (**Figura 16A**) y se realizó una regresión no lineal como se describe en la Sección 4.8. Asimismo se cuantificaron los genomas virales producidos a los diferentes tiempos post infección (ver figura 16B).

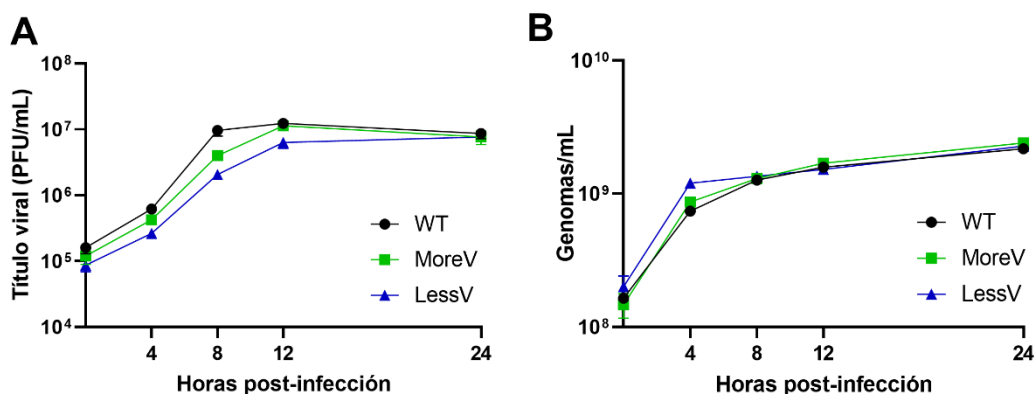


Figura 16. Curva de crecimiento de un solo paso en células HeLa. A) Se graficó el título viral en PFU/mL en función del tiempo tras el inicio de la infección en horas. B) Se cuantificaron los genomas por mL producidos durante la infección. En ambas gráficas se muestran los valores promedios con el error estándar de la media (SEM).

A partir de estas regresiones se obtuvieron las constantes de crecimiento (k) y se vió que, tras un One-way ANOVA con posterior test de múltiples comparaciones de Tukey para los triplicados de las curvas, hay diferencias significativas entre el WT y el genotipo LessV (**Figura 17A**).

Cabe destacar que los R^2 para los ajustes de las curvas a la Ec. 1 fueron, en promedio, de 0.785 para el WT, 0.728 para el MoreV y 0.928 para el LessV y las curvas ajustadas se muestran graficados en la **Figura 17 B-D**.

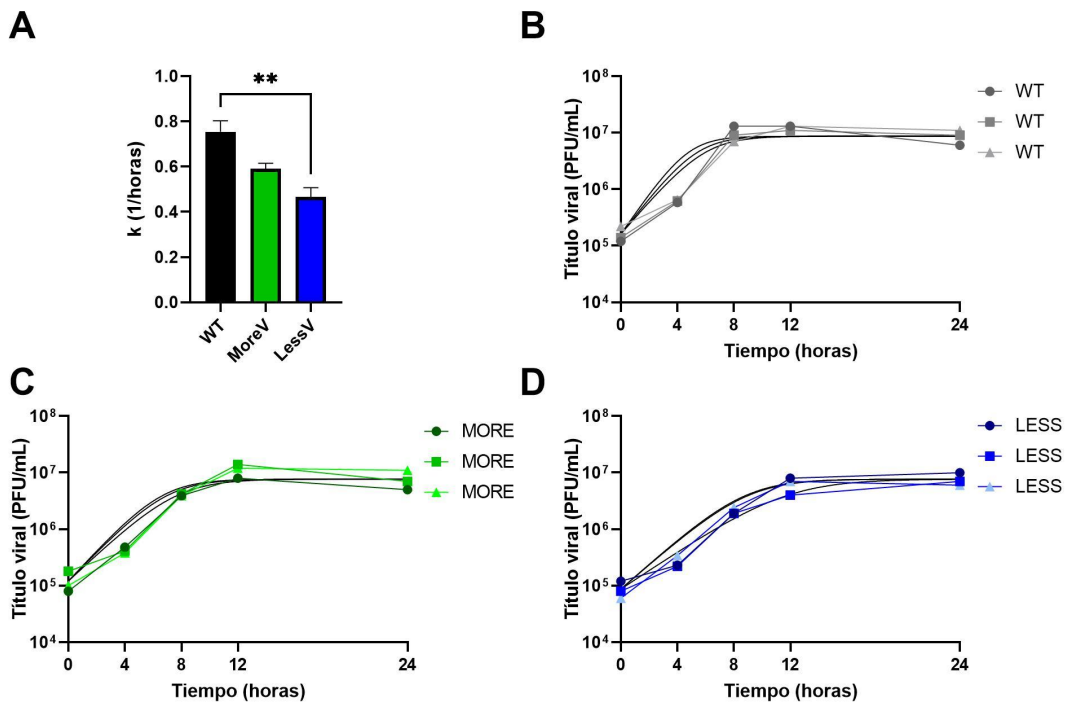


Figura 17. Regresiones no lineales para curva de crecimiento a MOI=3 y constantes de crecimiento. A) se comparan constantes de crecimiento para cada virus. Se puede observar que entre el virus WT y el genotipo LessV hay diferencias significativas en su crecimiento con un p-valor= 0.0049. El genotipo MoreV no tiene diferencias significativas en su crecimiento con los otros dos genotipos. En esta gráfica se muestran los valores promedios con el SEM. B) WT, C) MoreV y D) LessV; se muestran como se ajustan las regresiones no lineales a cada una de las curvas (curvas en negro).

A partir de los datos de genomas generados y de las partículas virales generadas en estas curvas de crecimiento se calculó la infectividad específica (IE) a las 24 horas post infección. (**Figura 18**). La infectividad específica es la relación entre PFU y genomas. Cuanto mayor es la IE, mayor es la cantidad de genomas viables encapsidados en relación a los genomas generados y viceversa.

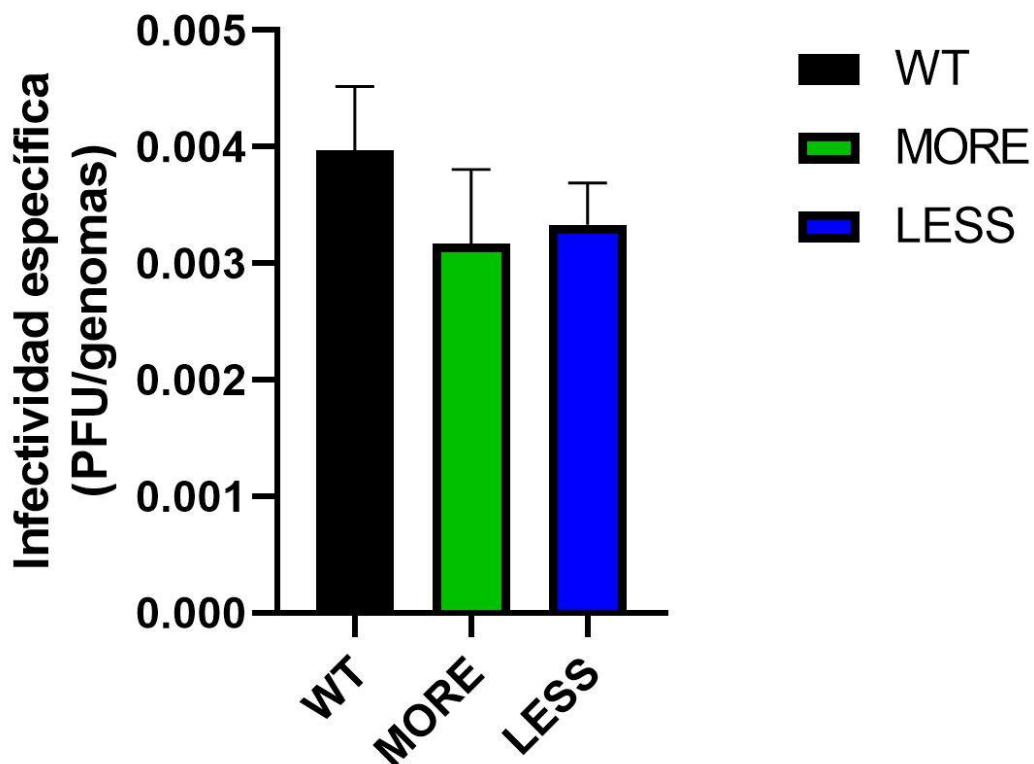


Figura 18. Infectividad específica de cada virus a las 24 horas en la curva de crecimiento de alta MOI. A las 24 horas los 3 genotipos tienen la misma infectividad específica tras realizar un One-way ANOVA con posterior test de Tukey. En esta gráfica se muestran los valores promedio de infectividad específica para cada genotipo en cada tiempo con el SEM.

La relación entre partículas virales infectivas y genomas no presenta diferencias significativas a las 24 horas de infección.

5.3.2 Cinética de replicación viral a MOI=0.01

Al contrario de la curva de crecimiento anterior, en esta se busca ver el comportamiento poblacional cuando se parte de una MOI baja. Para conseguir una MOI=0.01 se busca que una relación de una partícula viral infectiva por cada 100 células.

Esta curva de crecimiento se tituló por dos estrategias distintas, plaque assay y TCID₅₀. El principal objetivo de esta doble titulación fue familiarizarse con el TCID₅₀ y también poder comparar los resultados obtenidos entre las titulaciones realizadas con estas técnicas.

Al igual que para la curva anterior, para ver si hay diferencias estadísticas entre los 3 genotipos se realizó una regresión no lineal como se describe en la Sección 4.8. Se obtuvo la constante de crecimiento para cada uno de los virus, tanto para la curva titulada por plaque assay (**Figura 19.A**) como para la titulada por TCID₅₀ (**Figura 19.B**).

Los R² para los ajustes de las curvas a la Ec.1 fueron, en promedio, de 0.683 para el WT, 0.829 para el MoreV y 0.698 para el LessV..

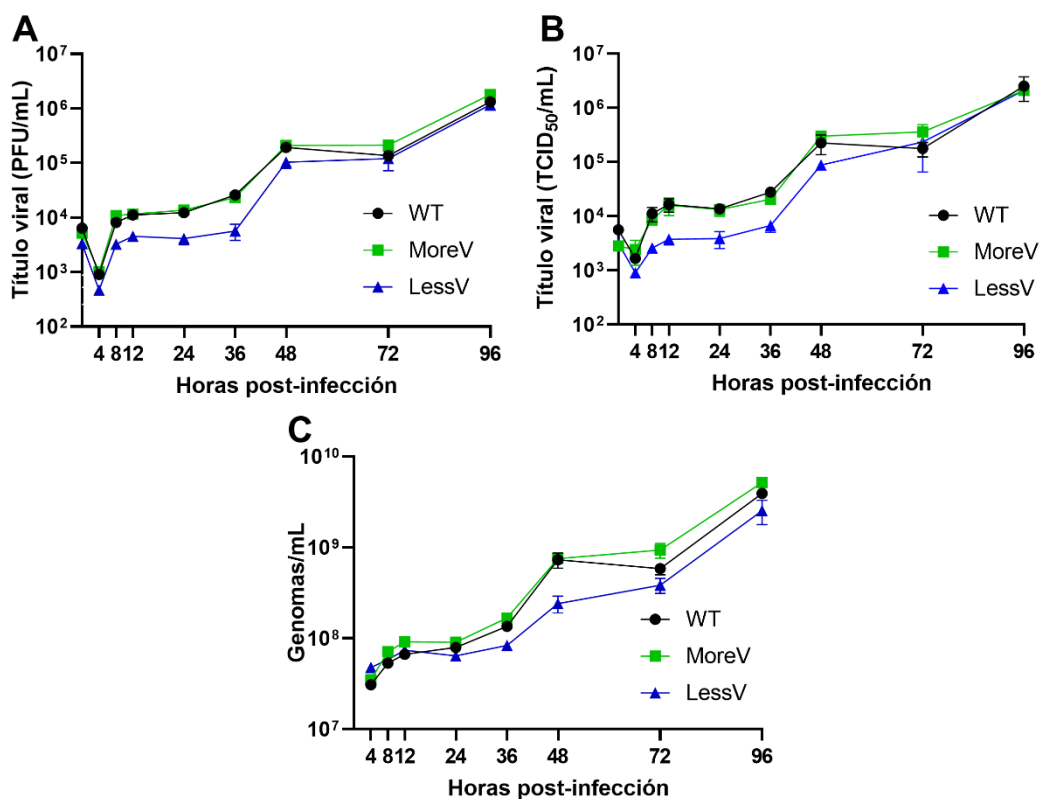


Figura 19. Curva de crecimiento en células HeLa a baja MOI. Para esta curva se utilizó una MOI=0.01. A) Curva de crecimiento titulada por plaque assay. B) Curva de crecimiento titulada por TCID₅₀. C) Curva de cuantificación de genomas por mL. En todas las gráficas se muestran los valores promedio con el SEM.

Las constantes de crecimiento no mostraron diferencias significativas (One-way ANOVA con test de múltiples comparaciones de Tukey) (**Figura 20**) para ninguna de las dos curvas por lo que podemos decir que cuando la MOI es baja, estos 3 virus se comportan similar.

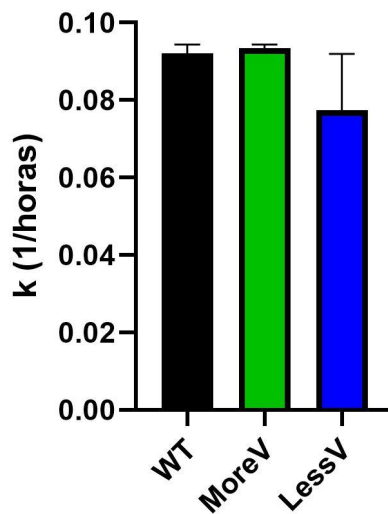


Figura 20. Comparación de las constantes de crecimiento para cada virus a MOI=0.01. Se puede ver que a baja MOI los virus se comportan igual. En esta gráfica se muestran los valores promedios con el SEM.

Respecto a la cantidad de genomas producidos durante la curva de crecimiento (**Figura 19.C**), son estadísticamente iguales en todos los tiempos menos a las 96 horas, tiempo en el que el genotipo MoreV tiene mayor cantidad de genomas que los otros dos genotipos (p -valor=0.0004 para el WT y <0.0001 para el LessV). Asimismo, la cantidad de genomas a las 96 horas es significativamente mayor para el WT que para el LessV (p -valor <0.0001). Para este análisis se realizó un Two-way ANOVA con posterior test de múltiples comparaciones de Tukey.

Otra manera por la que se estimó el fitness de estos 3 genotipos a baja MOI fue mediante ensayos de competencia en los que se obtiene un fitness relativo a un virus de referencia (CVB3 de Referencia) (**Figura 21**).

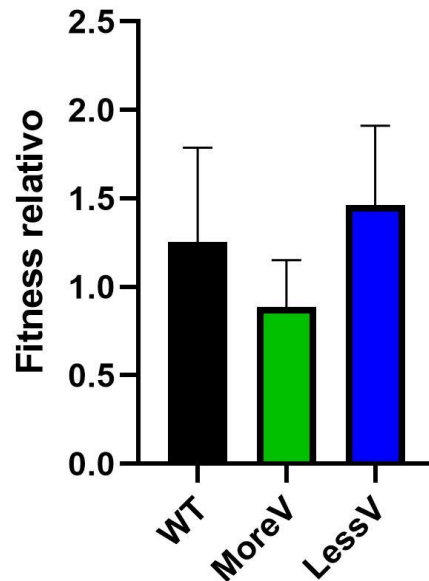


Figura 21. Fitness relativo de los ancestros virales. Se calculó el fitness relativo de las poblaciones ancestrales mediante ensayos de competencia. A los datos obtenidos se les realizó un One-way ANOVA con posterior test de Tukey para múltiples comparaciones y no se vieron diferencias significativas entre los 3 genotipos. La altura de las barras muestran el promedio de los triplicados y la barra de error el SEM.

A partir de estos resultados también podemos confirmar que estos 3 genotipos tienen un comportamiento similar a baja MOI.

5.4 Evolución experimental por deriva génica

El objetivo de aplicar este mecanismo evolutivo se debe a que mediante la deriva génica existe la posibilidad de que se fijen mutaciones deletéreas para la población y de esta manera estudiar qué tan robustos mutacionalmente son los 3 genotipos virales con los que se trabajaron. Por lo tanto, la evolución experimental mediante pasajes de calva en calva resulta ser un abordaje apropiado para generar grandes reducciones en el tamaño de la población y en el que el efecto de la deriva génica predomine por sobre la selección natural.

Como resultado de esta evolución por pasajes de calva en calva se analizaron 3 características de las poblaciones: i) el tamaño poblacional generado en cada

calva que se pasó; ii) el fenotipo de las calvas al final de los pasajes (pasaje 10) y iii) el fitness relativo de los virus en el pasaje 10.

5.4.1 Tamaño poblacional durante la evolución calva a calva.

A medida que se realizaron los pasajes de calva en calva se titularon las poblaciones que se plaquearon para saber el tamaño de la población de la calva que se había picado en el pasaje anterior.

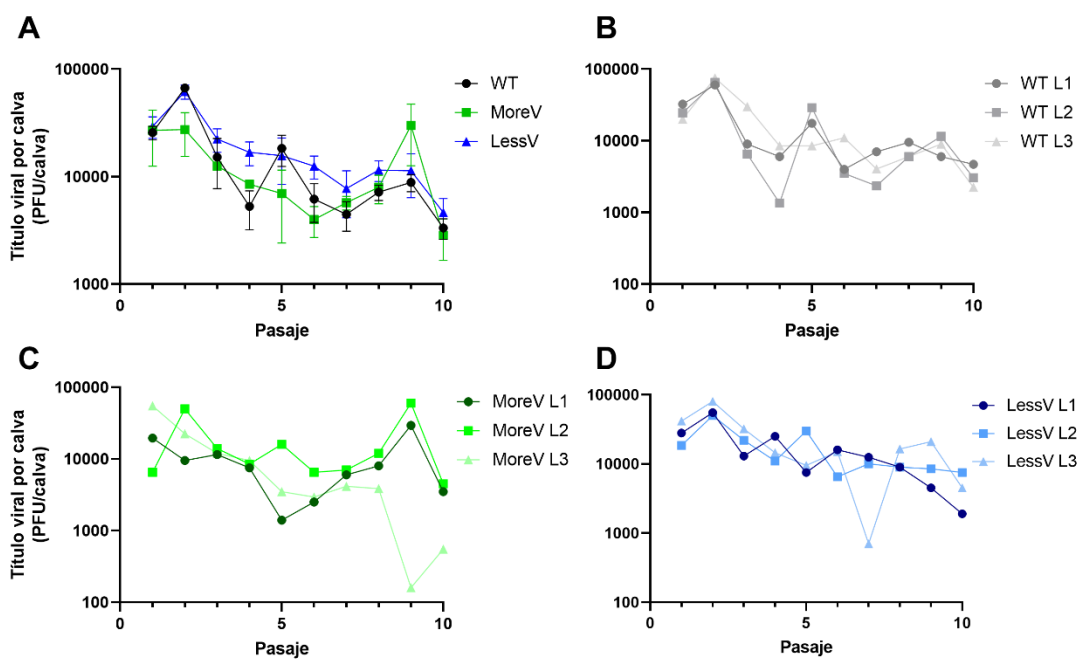


Figura 22. Tamaños poblacionales de cada calva durante la evolución experimental por deriva génica. A) Se muestra el promedio de los 3 linajes para los 3 genotipos con el SEM. B) Tamaños poblacionales de los 3 linajes independientes del genotipo WT. C) Tamaños poblacionales de los 3 linajes independientes del genotipo MoreV. D) Tamaños poblacionales de los 3 linajes independientes del genotipo LessV. La escala del eje de ordenadas es distinta en el panel A al compararla con los otros 3.

En la **Figura 22.A** vemos que en promedio, los 3 genotipos presentan una disminución en la población recuperada de cada pasaje. En general, esta disminución se mantiene pasaje tras pasaje con excepción del genotipo MoreV que luego del pasaje 6 tiene un aumento, siendo más importante tras el pasaje 9 para que luego, en el pasaje 10, continúe el descenso en el tamaño poblacional.

Si se desglosan los linajes independientes (Li) de cada genotipo se aprecia un comportamiento más estocástico del genotipo MoreV con respecto a los genotipos WT y LessV. Esto se observa en la variación del SEM a lo largo de los pasajes (generaciones) en la **Figura 23**.

En el genotipo WT (**Figura 22.B**) se ve que el Li3 se despega de los otros dos en el pasaje 3. En el pasaje 4 vuelve a igualarse con el Li1 pero el Li2 continúa con el descenso de su población. En el pasaje 5 las poblaciones de los 3 linajes aumentan para luego ir en un descenso no constante hasta el pasaje 10.

El genotipo MoreV (**Figura 22.C**) es el que presenta mayor variabilidad en su comportamiento. El Li3 es el que genera una mayor población en el pasaje 1 y va en descenso sostenido hasta el pasaje 5, en el que mantiene su población por 3 pasajes más. Posteriormente presenta un descenso abrupto en el pasaje 9 y recupera parcialmente el tamaño de su población en el pasaje 10. El Li1 presenta un descenso en su tamaño poblacional en los primeros 5 pasajes para luego tener un ascenso hasta el pasaje 9 donde alcanza su máximo tamaño poblacional. Finalmente en el pasaje 10 disminuye su población. El Li2 de este genotipo resulta ser el más caótico de todos. De los 9 linajes es el que comienza con menor tamaño poblacional. Al siguiente pasaje es el que mayor tamaño título viral tiene de los MoreV. En los dos siguientes pasajes se comporta similar a los otros dos linajes para luego tener un pequeño pico en el pasaje 5 y finalmente un gran pico en el pasaje 9. Luego del pasaje 7, los Li1 y Li2 se comportan muy similar en cuanto al tamaño poblacional.

El último genotipo, LessV, (**Figura 22.D**) presenta una disminución constante de su tamaño poblacional con excepción del Li3 que en el pasaje 7 tiene un cambio brusco en su tamaño poblacional. Si bien hay ciertos altibajos, se ve claramente que a lo largo de la evolución, la población tiende a ir en constante descenso. Esto es similar a lo que ocurre con el WT. Sin embargo, a pesar de la disminución constante en su población, este genotipo es el que mantiene los tamaños poblacionales más altos en todos los pasajes (con excepción del pasaje 9).

Los porcentajes de la reducción de la población entre el pasaje 1 y el pasaje 10 pueden verse en la Tabla 3 (página 46).

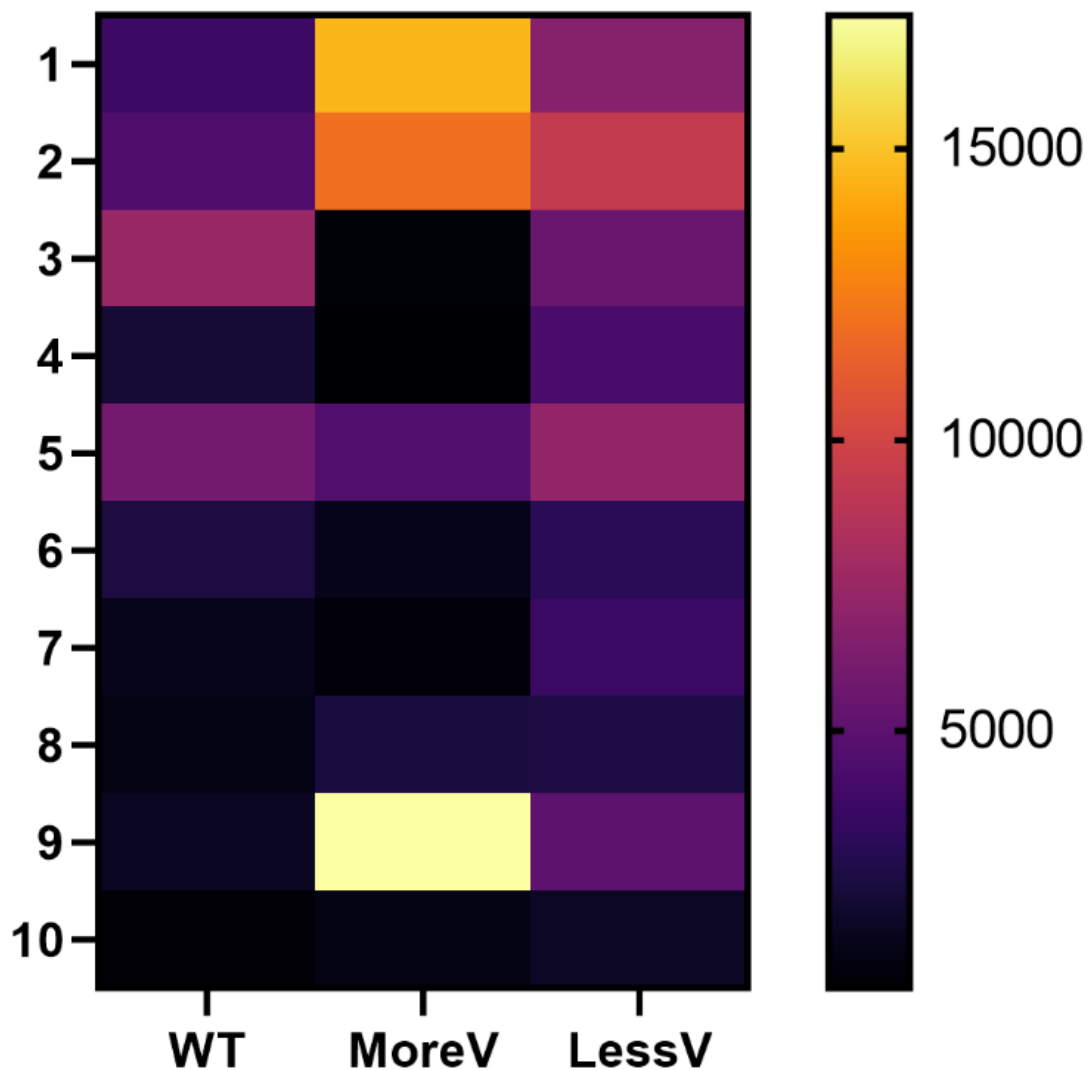


Figura 23. Mapa de calor del SEM entre los tamaños poblaciones de los 3 linajes durante la evolución por deriva génica. En el eje de las y se muestra el número de pasajes, mientras que en el de las x el genotipo viral.

5.4.2 Fenotipo de calvas

Tras finalizar la evolución experimental se comparó el tamaño de calvas del pasaje 10 de cada linaje con sus respectivos ancestros para ver si el fenotipo de las calvas era afectado por el impacto de los sucesivos cuellos de botella. Los resultados se ven en la **Figura 24**.

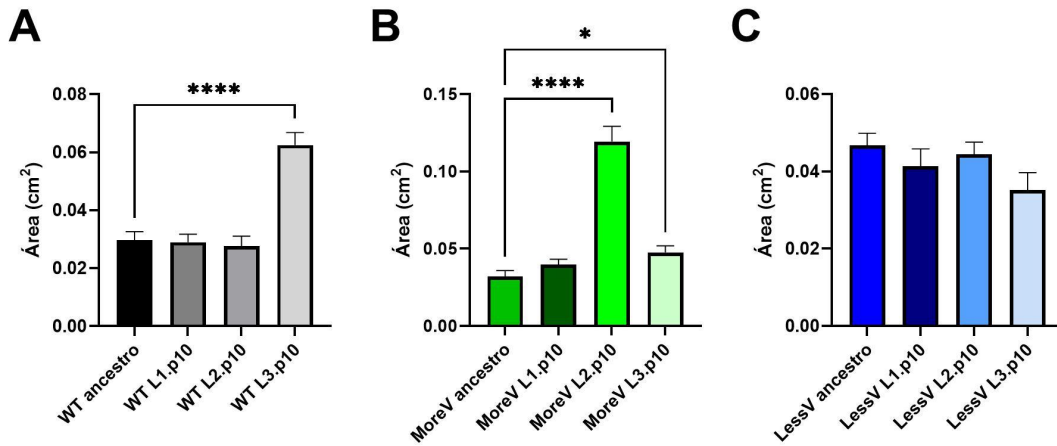


Figura 24. Fenotipo de calvas. Se calcularon las áreas de las calvas del pasaje 10 de cada linaje de cada genotipo. A) se muestran los fenotipos del genotipo WT. B) se muestra el genotipo MoreV. C) se muestra el genotipo LessV. En los 3 paneles se muestran los promedios de los tamaños de calvas en cm^2 con el SEM. Se realizó un One-way ANOVA con posterior test de múltiples comparaciones de Tukey y respecto a los p-valor, * indican que es menor a 0.05, y **** que es menor a 0.0001. Cabe destacar que los ejes no respetan los tamaños entre paneles.

A partir del panel A de la **Figura 24** se puede ver que el tamaño de las calvas para el WT para dos de los linajes se mantiene. Mientras que para el Li3, las calvas son 2.1 veces más grandes que las de la población ancestral.

Los tamaños de las calvas de los linajes evolucionados del genotipo MoreV son las más variables. El Li1 conserva el mismo tamaño de calva que el ancestro MoreV. Por otra parte, los linajes 2 y 3 tienen tamaños de calva mayores al de la población ancestral, con un p-valor menor a 0.0001 y de 0.044 respectivamente. Cabe destacar que las calvas generadas por el Li2 son 3,7 veces más grandes que las de las iniciales. Mientras que las del Li3 es 1.48 veces más grande. de los Li1 y Li3 del genotipo MoreV parecen mantenerse similares a la población ancestral. Sin embargo, durante la evolución del Li2 de MoreV se asume que se fija una mutación que genera un fenotipo de calvas más grandes y se mantiene a lo largo de la misma. Incluso en el pasaje 10, estas calvas son 3.7 veces más grandes que las ancestrales.

Finalmente, el genotipo LessV parece ser el que menos impacto sufre sobre el tamaño de calvas dado que mantuvo el mismo en los 3 linajes cuando se compara con la población ancestral.

La manera en la que el fenotipo de calva cambió desde los ancestros hasta que se finalizó la evolución puede verse en la Tabla 3 (página 46).

5.4.3 Fitness relativo al CVB3 de Referencia

La tercera característica de las poblaciones evolucionadas que se evaluó fue el fitness relativo para todos los linajes de cada genotipo del pasaje 10 con el fin de ver el impacto de los cuellos de botella sobre el fitness. Los resultados de los ensayos de competencia por los cuales se evaluó el fitness relativo se muestran en la **Figura 25** y se calcularon según lo descrito en la Sección 4.10. A su vez, esos fitness fueron relativizados a los del ancestro para poder comparar la variabilidad durante la evolución.

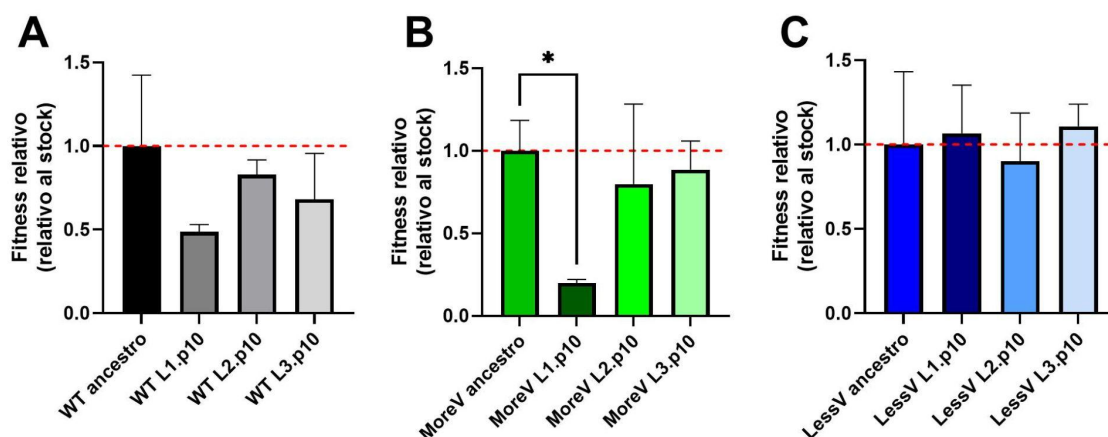


Figura 25. Fitness relativo de las poblaciones evolucionadas. Se cuantificó el fitness relativo mediante ensayos de competencia. En todos los paneles se muestran el fitness relativo al virus de Referencia y a su vez se utiliza al respectivo ancestro como normalizador de las poblaciones evolucionadas. A) se ve el fitness relativo de los linajes de WT evolucionados. B) están los que provienen del genotipo MoreV y C) los del genotipo LessV. Cabe aclarar que los colores de cada barra representan cada uno de los linajes. También es importante remarcar que los ejes de ordenadas respetan los tamaños entre paneles. Se realizaron t-test para cada muestra comparada al ancestro y se vio que el MoreV Li1 presenta un fitness relativo significativamente menor (p-valor=0.012).

Los datos utilizados para generar la **Figura 25** fueron sometidos individualmente a un t-test no pareado en los que se vio que solamente el Li1 del MoreV presentaba un fitness significativamente menor a su ancestro al finalizar la evolución experimental (p-valor=0.012). A partir de esto, los 3 linajes

del genotipo LessV son los que mejor mantienen el fitness tras los 10 pasajes. Dentro de las poblaciones evolucionadas del genotipo WT, parecería que uno de los linajes pierde su fitness a un 50% de la población ancestral. Mientras que los otros dos linajes del MoreV mantienen su fitness.

La variación entre el fitness relativo inicial (de las poblaciones ancestrales) y el de los linajes finales se puede ver en la Tabla 3.

Tabla 3. Comparación entre los valores analizados al inicio de la evolución y al final. Y es el tamaño poblacional. T. de calva es el tamaño de calva. W rel es el fitness relativo.				
Genotipo	Linaje	$\frac{Y \text{ en pasaje } 10}{Y \text{ en pasaje } 1} \times 100$	$\frac{T. \text{ de calva pasaje } 10}{T. \text{ de calva en ancestros}} \times 100$	$\frac{W \text{ rel en pasaje } 10}{W \text{ rel en ancestros}} \times 100$
WT	Li1	14.5	97.8	48.6
	Li2	12.5	93.3	82.9
	Li3	11.3	210.8	68.1
MoreV	Li1	17.9	123.5	19.9
	Li2	69.2	372.0	79.8
	Li3	1	148.1	88.6
LessV	Li1	6.8	89.7	106.7
	Li2	40.5	96.4	90.2
	Li3	10.8	76.3	110.8

5.5 Análisis de las bases moleculares de las poblaciones generadas en la evolución experimental

Mediante el uso de la plataforma GridION se logró obtener exitosamente la secuencia consenso de los 9 linajes del pasaje 10 de la evolución experimental por deriva génica. Las mutaciones fijadas a lo largo de la evolución se muestran en la **Figura 27**. Además, se desglosan con más detalles en las Tablas Anexo 2 y 3 (en el Anexo 7). Asimismo, el resumen general de la secuenciación se muestra en la Tabla Anexo 4 (en el Anexo 8). En general se obtuvieron buenos resultados para todas las secuencias con excepción de la secuencia del WT Li3.p10 en la que se obtuvo pocas lecturas. A pesar de esto, se pudo obtener una secuencia consenso para este genotipo.

Durante esta evolución se fijaron un total de 57 mutaciones: 19 se fijaron en los linajes WT, 21 en los linajes MoreV y 17 en los linajes LessV (**Figura 26**). De estas mutaciones, solamente 10 fueron no sinónimas y las mismas se fijaron: 3 en los linajes WT y 7 en los linajes MoreV. Los linajes provenientes del LessV no sufrieron ninguna mutación no sinónima tras los 10 pasajes.

Dentro de los linajes WT, en el Li1 se fijaron solamente dos mutaciones sinónimas luego de los 10 pasajes. El Li2 fue el que más mutaciones sufrió con la fijación de un total de 10 mutaciones, 9 sinónimas y 1 no sinónima. La mutación no sinónima se fijó en el residuo 323 de la proteína 3D (ARN polimerasa) e implicó un cambio de una lisina en donde había una arginina. En el Li3 las mutaciones fijadas fueron 7, de las cuales 2 fueron no sinónimas. Una de ellas se fijó en el residuo 63 de la VP3, en la que se cambió una asparagina por una histidina, y la otra fue en el residuo 13 de la proteína 3A, donde se cambió una treonina en lugar de una alanina.

Al desglosar los linajes del MoreV, se aprecia que fueron los que más impacto sufrieron sobre sus genomas. En el Li1 se fijaron 7 mutaciones, 2 de las cuales son no sinónimas. Una se fijó en el residuo 234 de la VP3 y el cambio aminoacídico fue de un ácido glutámico en lugar de una glutamina. La otra mutación fue el cambio de una cisteína en el lugar donde había una tirosina. Esto sucedió en el residuo 22 de la 2B. En el Li2, 4 mutaciones fueron sinónimas, 4 no sinónimas y 1 mutación fue en el IRES (5'UTR). De las 4 mutaciones no sinónimas, una fue en el residuo 63 de la VP3, donde una tirosina sustituyó una asparagina. Este residuo también sufrió una mutación en el Li3 del WT. La segunda mutación no sinónima fue en la 2A, en el residuo 54 donde una glutamina fue sustituida por una arginina. Las otras dos mutaciones no sinónimas caen en las proteínas 2C y 3A. En los dos casos, el nuevo aminoácido es una serina en lugar de una treonina y una prolina, respectivamente. Finalmente, en el Li3 se fijan un total de 5 mutaciones, 3 sinónimas, 1 no sinónima y 1 en el 5'UTR. La mutación no sinónima se da en el aminoácido 186 de la VP2 y una valina es el cambio que se introduce donde había una isoleucina.

Finalmente, en los linajes del genotipo LessV, si bien se fijaron 12 mutaciones, ninguna de ellas fue no sinónima. En el Li1 se fijó una mutación en el 5'UTR.

En el Li2, se fijaron 9 mutaciones, 2 de las cuales caen en el 5'UTR. Finalmente en el Li3, se fijaron 2 mutaciones y 1 es en el 5'UTR.

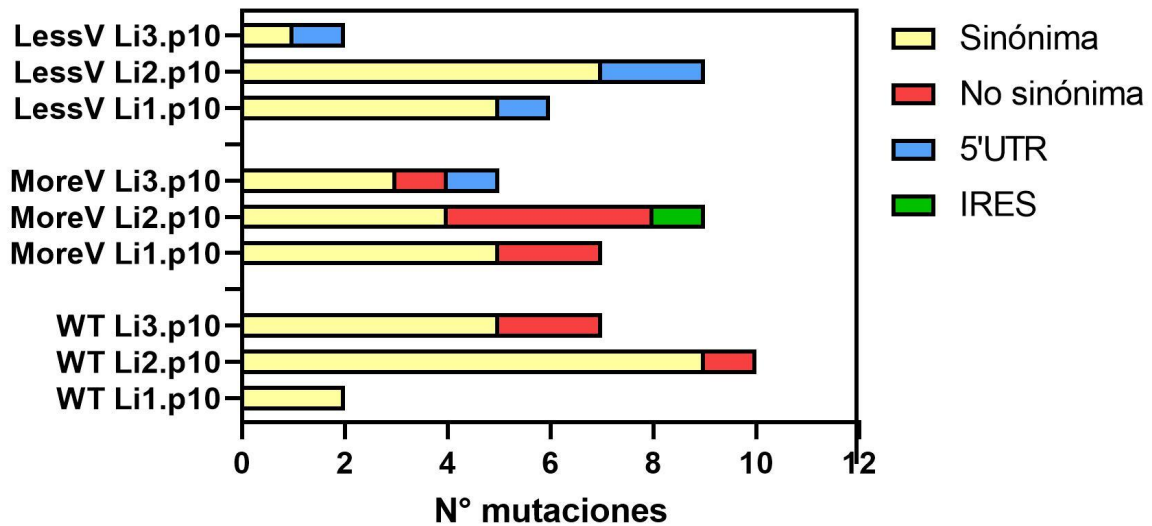


Figura 26. Mutaciones fijadas a lo largo de la evolución experimental. Se muestran las mutaciones fijadas acumuladas en función del linaje en el que ocurren. En amarillo se muestran la cantidad de mutaciones sinónimas, en rojo las no sinónimas, en azul las del 5'UTR y en verde las del IRES.

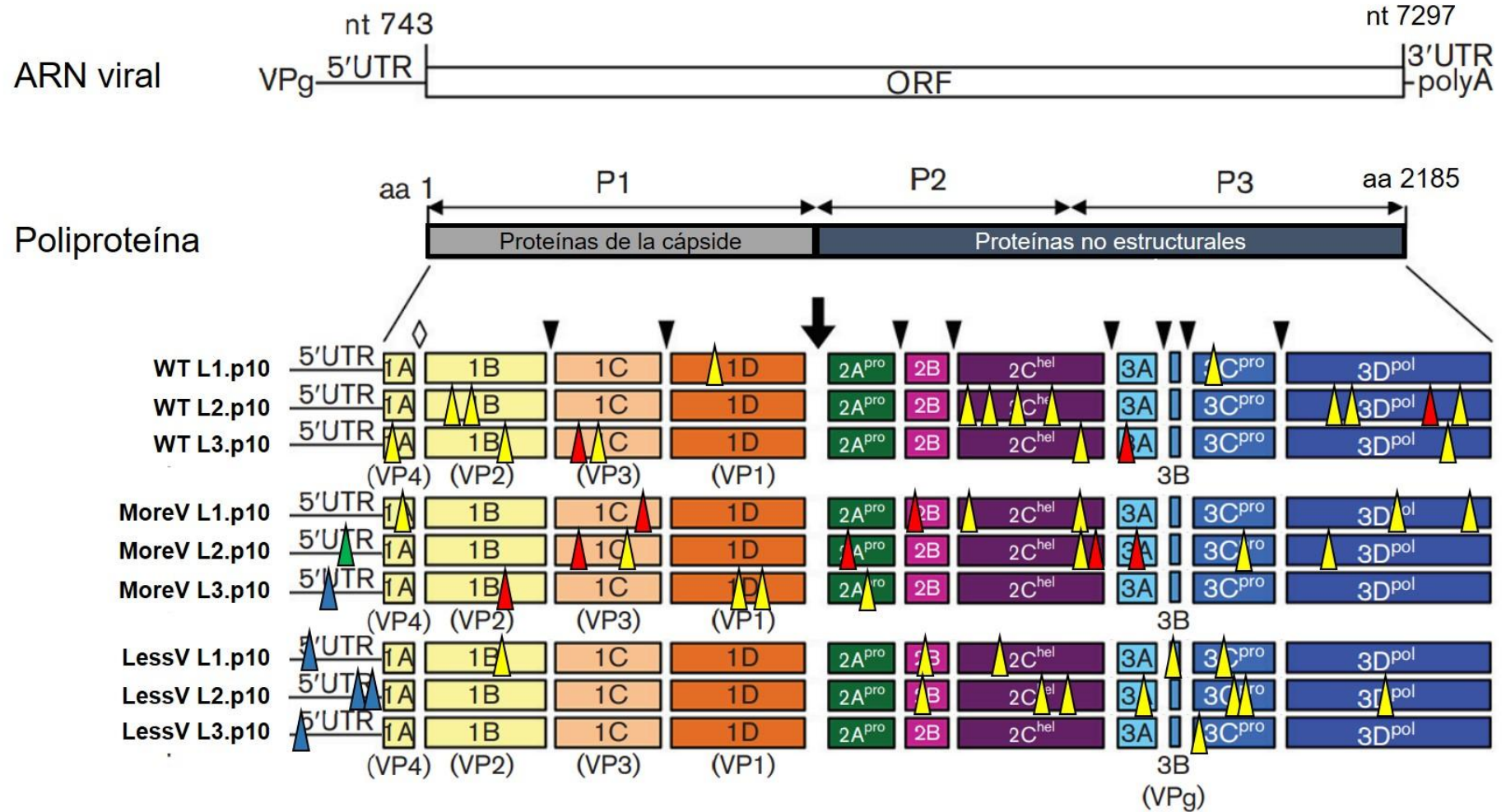


Figura 27. Mutaciones fijadas a lo largo de la evolución experimental. Los triángulos amarillos indican las mutaciones sinónimas fijadas; los triángulos rojos: mutaciones no sinónimas; triángulos verdes: mutaciones en el IRES, triángulos azules: mutación en el 5'UTR.

6. Discusión

6.1 Con respecto al desarrollo de protocolos y herramientas moleculares para CVB3:

i) Se logró amplificar el genoma completo de CVB3 en un segmento con dos protocolos distintos, el descrito en la Sección 4.7.2 y mediante el uso de la enzima Q5. Este último protocolo fue finalmente descartado para amplificar los genomas ancestrales dado que generaba amplificación inespecífica.

El uso de un solo amplicón representa una gran ventaja para la secuenciación dado que asegura la representación equimolar de todas las regiones del genoma al momento de generación de librerías. De esta manera se obtiene una profundidad y cobertura del genoma viral constante a lo largo del mismo. Por otra parte, posee la desventaja de que si se parte de poco material, como en el caso de la evolución por deriva génica, la amplificación de genomas completos en una única PCR resulta ser más compleja.

Una ventaja de los genomas amplificados en esta tesis es que los mismos tienen 220 nucleótidos más en su secuencia que los amplificados previamente (Moratorio et al., 2017). Dicha región de 220 nucleótidos se encuentran en la región 5'UTR del genoma de CVB3. Por lo que si bien, no implican cambios aminoacídicos ni afectan la región del IRES, podrían presentar mutaciones que afecten la traducción o replicación viral.

ii) Se desarrollaron dos estrategias para la cuantificación de genomas virales. En ambas, se utilizó información de la literatura (Gnädig et al., 2012) para seleccionar las sondas y primers. Esto resultó en protocolos optimizados realizados por primera vez en nuestro laboratorio y que estarán a disposición de otros integrantes. Además, me permitió familiarizarme con las técnicas de RT-qPCR multiplex.

6.2 Con respecto al estudio de la cinética de replicación de las variantes de volatilidad de CVB3:

Para discutir las curvas de crecimiento de las 3 variantes de volatilidad de CVB3, es importante recordar que todas las poblaciones tienen un crecimiento logístico. Este tipo de crecimiento se caracteriza por partir de un tamaño poblacional inicial (Y_0), de modo de ir incrementando dicho tamaño de manera exponencial hasta alcanzar la capacidad de carga (Y_M) que se define como el tamaño poblacional máximo que puede soportar una población en un ambiente determinado. Matemáticamente este tipo de crecimiento se representa con la Ec. 1, entre otras posibles. La principal diferencia entre los dos ensayos que realicé se basa en que para llegar al Y_M es necesario de un mayor o menor número de ciclos de replicación viral, MOI alta o baja, respectivamente.

i) Cinética de replicación a alta multiplicidad de infección (MOI=3) favoreciendo un bajo número de ciclos virales.

Como se observa en la **Figura 16A**, los genotipos WT y LessV tienen comportamientos diferentes en su crecimiento. Esto puede explicarse, por un lado, por factores que impactan cómo el Y_0 . En este caso la MOI de partida fue para algunos de los genotipos menor a la deseada: 1.73 para el LessV y 2.4 para el MoreV mientras que para el WT es de 3.2. Asimismo, el impacto de estas diferencias es fuerte en la pendiente de la curva que es muy importante a la hora de hacer la regresión no lineal.

A pesar de esto, a las 24 horas los títulos de los 3 genotipos son semejantes y al calcular la infectividad específica (IE) a ese tiempo, la misma no presenta diferencias significativas entre las 3 variantes de volatilidad de CVB3. Este resultado es llamativo, dado que teóricamente el genotipo más volátil (MoreV) podría presentar una IE diferente con respecto a los otros virus. Esto se debe a que la acumulación de mutaciones en este genotipo más propenso al cambio podría conducir a un mayor número de genomas no viables y por lo tanto impactar negativamente en el cociente de IE. A nivel teórico se esperaría que la generación de partículas infectivas se viera afectada debido a los cambios en las secuencias nucleotídicas, afectando el fenotipo y disminuyendo su robustez

mutacional. Posiblemente bajo condiciones de evolución acelerada, por ejemplo bajo el uso de mutágenos, los efectos de los genomas más volátiles y su hipotética caída en IE podrían evidenciarse más claramente. Lamentablemente, esto no fue realizado en esta tesis, sino que se emplearon eventos de cuello de botella sobre el tamaño poblacional para estudiar estos fenómenos que más adelante son discutidos.

ii) Cinética de replicación a baja multiplicidad de infección (MOI=0,01) favoreciendo un alto número de ciclos virales.

Frente a este ensayo la primera duda que me surge es: ¿Cuánto tiempo se tiene que evaluar el crecimiento viral? Entiendo que sería hasta que la población alcance su capacidad de carga (Y_M), a nivel práctico esto se puede ver cuando es evidenciada una meseta en la curva de crecimiento. En nuestro caso, se ve que el título viral alcanza su primera meseta entre las 12 y 24 horas para luego seguir subiendo en título hasta las 48-72 horas en el que se ve otra meseta. A nivel experimental, ambos tiempos serían correctos pero al observar las placas de 12 pocillos con células infectadas en el microscopio se apreciaba que aún había células vivas. Por lo tanto asumí, que el tamaño de la población viral seguiría creciendo y fue por eso que también se titularon las 96 horas luego de iniciada la infección. Sin duda alguna, falta al menos un tiempo más para evaluar si el título se mantiene en una meseta o si sigue creciendo. Sin embargo a las 96 horas las monocapas celulares estaban completamente arrasadas, por lo que el “sustrato” del virus ya se había terminado.

La constante de crecimiento se calcula realizando una regresión no lineal con los datos obtenidos hasta las 72 horas de crecimiento. La razón de realizar esto es que la ecuación utilizada para la regresión no lineal se ajusta a un crecimiento logístico y si consideramos que el crecimiento va hasta las 96 horas, este no se comporta de la misma manera que lo hace la regresión. A partir de estos datos, vemos que a baja MOI los virus crecen de la misma manera hasta llegar a su Y_M . Esto se confirma con los ensayos de competencia, que se realizaron con una MOI de 0.01 por 24 horas.

Es importante resaltar que existen diferentes aproximaciones para estudiar la cinética de replicación a baja MOI en otros virus. Algunos autores estudian el crecimiento viral hasta llegar a una meseta (Abdoli et al., 2013; Moser et al., 2018), mientras que otros dejan crecer el virus por un tiempo determinado (Liu et al., 2019; Takeuchi et al., 2002).

Finalmente, respecto a los ensayos de cinética de replicación, puede observarse que a alta MOI la variante de CVB3 LessV presenta una constante de crecimiento (k) menor al genotipo WT. Esto podría significar que su menor plasticidad genómica retrase su crecimiento con respecto a las otras variantes de CVB3 en estudio. Sin embargo, a baja MOI este genotipo menos volátil (LessV) no muestra diferencias significativas en su k . Esto podría explicarse ya que al favorecer el número de ciclos replicativos virales este genotipo es capaz de explorar mayores vecindarios mutacionales y por lo tanto replicar de forma similar a las otras variantes de CVB3.

6.3 En relación al estudio por evolución experimental del impacto de la deriva génica en las variantes de volatilidad de CVB3:

Una de las preguntas iniciales de esta maestría era: “¿cómo afecta la volatilidad genética a la robustez mutacional?”. Una forma de evaluar la tolerancia de los virus a la mutación es realizando pasajes de calva en calva. Este abordaje fomenta la deriva génica por sobre la selección natural. En cada pasaje, dado que la mayor fuerza evolutiva que se aplica es la deriva génica, es posible que una mutación deletérea se fije en la población. De esta manera se puede estudiar la capacidad de nuestras variantes de volatilidad de amortiguar la mutación. En nuestro caso pudimos observar mutaciones tras 10 pasajes realizados. Sin embargo estos cambios pueden evidenciarse más claramente luego de realizar entre 20 a 30 pasajes (Elena & Sanjuán, 2007).

Un factor que entra en discusión al momento de planificar este experimento es determinar el tiempo de desarrollo de las calvas. En nuestro caso elegimos 48

horas. Escarmís y colaboradores observaron que seleccionar distintas horas post infección impacta directamente sobre el descenso del tamaño de la población, a más horas menor es el descenso, por lo que este factor pudo jugar un rol fundamental para el estudio de las distintas poblaciones.

A lo largo de este abordaje de evolución experimental se evaluaron 3 características durante los pasajes: i) el título recuperado por calva (también llamado tamaño de la población); ii) el área o fenotipo de las calvas y iii) el fitness relativo. La razón de estudiar estas 3 características es que las mismas sirven como subrogado del fitness viral. El fitness se puede definir como la capacidad de un organismo a sobrevivir y reproducirse en el ambiente que se encuentra. Por lo que antecede, es que asumo que el título de virus recuperado de las calvas sirve como indicador de la descendencia de la partícula fundadora de la misma (Escarmís et al., 2002; Lázaro et al., 2003). Por otra parte, el tamaño (o área) de las calvas es un indicador de la capacidad de diseminación e infección de células vecinas por un virus en un medio semi sólido (Goh et al., 2016; Mandary et al., 2020). Finalmente, el fitness relativo por competencia es una manera de cuantificar el fitness utilizando otro virus, al cual llamamos virus de Referencia y cuantificamos la síntesis de genomas de ambos mientras compiten por los mismos recursos en un tiempo dado.

i) Título recuperado por calva de cada variante de volatilidad de CVB3 a lo largo de los pasajes en cuello de botella.

Durante este experimento se vio, en el pasaje 10, una reducción de los tamaños de las poblaciones virales de entre el 68,9% al 1% con respecto al tamaño poblacional presente en el pasaje 1. Este fenómeno ha sido previamente descrito por diferentes autores (Chao, 1990; Escarmís et al., 2002; Yuste et al., 2000) y es esperable dado la acción del trinquete de Muller. Este fenómeno consiste en la pérdida del fitness causada por la introducción de mutaciones a medida que se reducen los tamaños poblacionales. Sería interesante continuar con este abordaje para ver si el proceso evolutivo ejercido lleva a la extinción de alguno de los linajes. De todas formas, sería desafiar lo

previamente descrito en la literatura y además entra en juego la importancia de la volatilidad genética en este tipo de evolución, algo que sería novedoso.

A lo largo de los 10 pasajes, el genotipo LessV es el único que desciende en su tamaño poblacional de manera constante, lo que no significa que sea el que menos población tiene. De hecho, a pesar del descenso constante de su tamaño poblacional, mantiene en todos los pasajes un mayor tamaño poblacional que los otros dos genotipos. A partir de esto, podríamos afirmar que el genotipo LessV es el que menos sufre los sucesivos cuellos de botella. Si tomamos el tamaño poblacional como una medida de fitness, en vista de estos resultados, el genotipo LessV sería el más robusto mutacionalmente. Todo lo opuesto ocurre con el genotipo MoreV que en la mayoría de los pasajes es el que tiene los menores tamaños poblacionales. Cada cuello de botella parece impactar más en el tamaño poblacional del MoreV que en el del genotipo LessV. Esto coincide con la característica de que el MoreV es el genotipo más volátil, por lo que sería el que peor amortigua las mutaciones. El genotipo WT se mantiene entre los dos genotipos en casi todos los pasajes.

ii) Tamaño de calvas de cada variante de volatilidad de CVB3 a lo largo de los pasajes en cuello de botella.

Dado los tamaños incrementados de calvas de los linajes: WT Li3.p10 y MoreV Li2.p10, decidimos secuenciar el ARN extraído de cada una de estas calvas. Ambos linajes tienen una mutación en el residuo 63 de la proteína VP3, siendo N63H la mutación del linaje WT y N63Y la observada en la variante MoreV. Interesantemente, ambas mutaciones fueron previamente reportadas y estudiadas. La primera de ellas surge de un trabajo en el que se estudia la transmisión de CVB3 mediante vesículas (Bou & Sanjuán, 2021). Esta mutación aparece dentro de las partículas que se comportan de manera individual. En este mismo trabajo la caracterizan y ven que también tiene fenotipo de calva de mayor tamaño, aproximadamente 2 veces más grande que el WT, por lo que coincide con los resultados de esta tesis. Además, se realizaron ensayos de competencia, utilizando una estrategia distinta a la de

este trabajo, en la que demuestran que este mutante tiene más fitness que el WT, algo que no coincide con lo reportado en este trabajo. Sin embargo, hay que tener en cuenta que en nuestro caso la población viral presenta varias mutaciones que podrían generar una epistasis negativa sobre esta mutación. La otra mutación, N63Y, se reporta en un trabajo en el que se infectaron ratones y 72 horas después se recuperan partículas virales que generan calvas de gran tamaño (Wang & Pfeiffer, 2016). Asimismo, los autores demuestran que tiene defectos para crecer en cultivos celulares al ver que crece más lento que el WT y lo asocian a una reducción en la unión mediada por glicanos.

Además, estas dos mutaciones se describen en otros dos artículos. En Mattenberger et al., 2021, la mutación N63H surge como variante de alta frecuencia en varias réplicas en las que se somete al CVB3 WT a sucesivos cuellos de botella basados en el aumento de la temperatura (mayores a 37°C). Por otra parte, en Bordería et al., 2015 se reportan tanto la mutación N63H como la N63Y. La primera de ellas se asocia su aparición en ambas líneas celular, HeLa y A549, sin embargo la segunda mutación mencionada se fija solamente en A549.

En estudios con el Enterovirus-A71 se lograron aislar poblaciones virales con fenotipos de calva grande pero en este trabajo las mutaciones que se fijaban para lograrlo estaban todas en la VP1 y ninguna de las mutaciones fijadas en estos dos mutantes se encuentran esa región (Mandary et al., 2020).

Por otra parte, en CVB4 se asocian mutaciones en la VP4 y en el 5'UTR a los fenotipos de calvas grandes (Ramsingh et al., 1995), siendo esto interesante dado que muchas de las mutaciones fijadas fueron en el 5'UTR.

iii) Fitness relativo de cada variante de volatilidad de CVB3 a lo largo de los pasajes en cuello de botella.

Luego de haber realizado 10 pasajes de calva en calva a modo de disminuir significativamente el tamaño poblacional y favorecer la fijación de mutaciones, se evaluó el fitness relativo de los linajes evolucionados (ver **Figura 25**). El genotipo LessV fue el que menor variación presentó en los tres linajes en el

pasaje 10 (linajes evolucionados) con respecto a su fitness ancestral. Importantemente, este genotipo era el de menor volatilidad por lo que teóricamente sufriría menos la acción del trinquete de Muller. Por otro lado, uno de los linajes del genotipo MoreV es el linaje que más fitness relativo pierde en comparación con su ancestro. Con respecto al genotipo WT, el mismo pierde fitness pero en una menor fracción que la disminución observada para los linajes del genotipo MoreV. La pérdida de fitness a la medida que se realizan cuellos de botella es algo previamente descrito (Clarke et al., 1993; Escarmís et al., 2002; Lázaro et al., 2003; Peña et al., 2000; Yuste et al., 2000). Además, hay que recordar que el tiempo elegido para la generación de las calvas es un factor de impacto sobre la conservación del fitness (Escarmís et al., 2002; Manrubia et al., 2005). Por lo que podría observarse un mayor impacto de la evolución por deriva génica al reducir el tiempo de generación de calvas, y/o, aumentar la cantidad de pasajes en la evolución.

6.4 Estudio de las mutaciones fijadas a lo largo de la evolución

Dos de las diez mutaciones fijadas fueron discutidas en párrafos anteriores. De las otras ocho mutaciones sólo dos aparecen en la bibliografía. La primera de ellas es la mutación Q234E en VP3 que se fijó en el MoreV Li1.p10. Esta sustitución específica no está descrita en ningún artículo. Sin embargo, existe evidencia de que este residuo interacciona con el residuo 63 de la VP3, discutido previamente, y parecen ser seleccionados debido a que están involucrados en la unión viral a los receptores celulares, en particular a DAF (Bordería et al., 2015). Por otra parte, se ha descrito en un preprint que la mutación Q234R potencia la asociación de CVB3 al receptor CAR (Hulsebosch et al., 2021). La segunda mutación que aparece en la literatura es C22Y en la proteína 2B que aparece únicamente en (Bou & Sanjuán, 2021). Esta mutación no tiene ninguna función asociada, se encuentra en baja frecuencia en poblaciones de CVB3 (Bou Prados, 2021). La misma es interesante en vista de que ambos aminoácidos tienen propiedades fisicoquímicas muy diferentes. La tirosina es un aminoácido aromático y la cisteína es nucleofílico que tiene un grupo tiol que puede generar enlaces de disulfuro.

El resto de sustituciones, tanto sinónimas como no sinónimas, no parecen tener relevancia biológica descrita previamente, tanto en CVB3 como en otros picornavirus (Carson et al., 2011; Dial et al., 2019; McDonald et al., 2016; Wessels et al., 2006). A pesar de esto, dos mutaciones resultan interesantes. La primera de ellas es R323K que se sitúa en la polimerasa viral (3D). Si bien es un cambio entre aminoácidos con características fisicoquímicas similares, no deja de ser una mutación en una enzima clave para la replicación viral y altamente conservada. La otra mutación es una sustitución nucleotídica en el IRES (G499A) de la población MoreV Li2.p10, que recordemos que es el genotipo con calvas más grandes.

6.5 En suma:

El estudio de diferentes variantes de volatibilidad de CVB3 desde el punto de vista evolutivo nos ha permitido observar diferentes características de las poblaciones virales. Respecto al tamaño de calvas se aprecia que tanto el genotipo WT como el LessV son aquellos que menos varían el área de las mismas en relación al área de las calvas de sus poblaciones ancestrales. Exceptuando al linaje WT Li3 en el que se fijó la mutación N63H, lo cual explica su mayor tamaño (Bou & Sanjuán, 2021).

Si se utiliza el fitness relativo como guía, el genotipo LessV en sus 3 linajes evolucionados es en el que menos variación se observa en relación a su ancestro. Este resultado es coherente ya que es la variante de volatibilidad que menos mutaciones (n=17) adquirió luego de los 10 pasajes, no afectando ninguna de ellas la secuencia proteica original. Es importante destacar que 4 de estas mutaciones se encontraban en el 5'UTR aunque no parecen haber afectado los fitness de los linajes.

Con respecto a los 3 linajes WT, los mismos sufrieron 19 mutaciones en total, de las cuales solo 3 fueron no sinónimas. Esto contrasta con las 21 mutaciones encontradas en los 3 linajes MoreV donde 7 fueron no sinónimas.

En vista de la cantidad de mutaciones y de la conservación del fitness podríamos asumir que el más robusto es el menos volátil y a su vez que el más volátil es el menos robusto, teniendo al WT en una posición intermedia.

Sin embargo, al analizar las mutaciones fijadas en todos los linajes se ve que ninguna de ellas ocurre en los codones modificados que otorgan diferentes grados de volatilidad. Por lo que creemos que este fenómeno puede darse a nivel poblacional y de variantes minoritarias. Por lo que antecede, es necesario realizar este abordaje basado en evolución experimental que conlleve un mayor número de pasajes y utilizar técnicas de secuenciado ultra-profundo para examinar las frecuencias alélicas y determinar el rol de las variantes minoritarias en lo que observamos en las secuencias consenso

Asimismo, sería clave ver las secuencias de los pasajes intermedios pero al momento de la escritura de esta tesis carecemos de ellas. Teniendo estas secuencias, se puede hacer un análisis más exhaustivo.

Es interesante plantear que el genoma menos volátil LessV, de alguna manera, fue un sustrato que sirvió para no adquirir mutaciones no sinónimas. Según los conceptos de espacio de secuencia y paisaje adaptativo, la capacidad de amortiguar las mutaciones del genotipo LessV, generaría un enjambre de mutantes que se mantendrían dentro de vecindarios mutacionales más conectados y similares. Asumiendo el comportamiento intermedio del genotipo WT, continuaré discutiendo sobre el MoreV. Para este genotipo ocurriría lo contrario, la acumulación de mutaciones impacta más sobre el fenotipo al tener una mayor capacidad de cambiar de aminoácido frente a mutaciones. En vistas de que la evolución por deriva génica favorecerá la fijación de mutaciones aleatorias y por lo tanto, mayores chances de fijar mutaciones deletéreas dado que cada mutación es más probable que sea deletérea que beneficiosa, se esperaría que el genotipo MoreV sea el primero que sufra una extinción. Esto podría ocurrir en una evolución de calva en calva con un número de pasajes más elevado.

7. Conclusiones

- Con el fin de realizar los ensayos de competencia entre los virus de interés y el virus CVB3 de Referencia y la cuantificación de genomas virales, se pusieron a punto con éxito los protocolos descritos en la Sección 4.11.
- Se concluyó que a baja MOI, los 3 genotipos de estudio presentan un comportamiento similar en cuanto a su cinética de replicación. Los resultados obtenidos para los ensayos con alta MOI , debe volver a realizarse ensayos para confirmar los resultados.
- Respecto al fitness durante los 10 pasajes de calva en calva: i) el tamaño poblacional descendió en las variantes de volatilidad de CVB3 de acorde a lo descrito en la bibliografía, ii) surgieron 2 linajes con fenotipos de calvas grandes y iii) el fitness relativo se mantuvo o descendió respecto a la población ancestral.
- Tras la evolución por deriva génica, se determinó que el genotipo menos volátil, LessV es el más robusto de los 3 genotipos.
- Se logró la secuenciación de 7310 nucleótidos de los genomas de CVB3 por secuenciación NGS, confirmándose que todos los genotipos utilizados en esta tesis mantenían la secuencia nucleotídica del clon infeccioso

8. Perspectivas

Como perspectivas de esta tesis nos planteamos las siguientes actividades:

- Realizar nuevamente la curva de crecimiento a alta MOI para confirmar el comportamiento de los genotipos.
- Continuar con la evolución experimental realizando un mayor número de pasajes de calva a calva para comprobar si lo observado en esta tesis se mantiene en un abordaje de evolución experimental más extenso en el tiempo y bajo las mismas condiciones de cuello de botella.
- Secuenciar pasajes intermedios de la evolución experimental realizada para conocer más en detalle cómo sufrió cada genotipo en los diferentes pasajes, en particular sería interesante los pasajes 3 y 6. Además, utilizar tecnologías NGS, como Illumina, para poder estudiar las variantes minoritarias de la población.
- Realizar mutaciones de sitio dirigidas en los clones infecciosos con el fin de estudiar más a fondo las mutaciones fijadas a lo largo de la evolución experimental en los diferentes contextos de genomas. En particular las 4 mutaciones que se destacan en la discusión.

9. Bibliografía

- Abdoli, A., Soleimanjahi, H., Tavassoti Kheiri, M., Jamali, A., & Jamaati, A. (2013). Determining influenza virus shedding at different time points in madin-darby canine kidney cell line. *Cell Journal*, *15*(2), 130–135.
- Archetti, M. (2009a). Survival of the steepest: Hypersensitivity to mutations as an adaptation to soft selection. *Journal of Evolutionary Biology*, *22*(4), 740–750. <https://doi.org/10.1111/j.1420-9101.2009.01697.x>
- Archetti, M. (2009b). Genetic robustness at the codon level as a measure of selection. *Gene*, *443*(1–2), 64–69. <https://doi.org/10.1016/j.gene.2009.05.009>
- Badorff, C., Lee, G.-H., Lamphear, B. J., Martone, M. E., Campbell, K. P., Rhoads, R. E., & Knowlton, K. U. (1999). Enteroviral protease 2A cleaves dystrophin: Evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nature Medicine*, *5*(3), 320–326. <https://doi.org/10.1038/6543>
- Baltscheffsky, H., Jörnvall, H., & Rigler, R. (1986). *Molecular evolution of life: Proceedings of a conference held at Södergarn, Lidingö, Sweden, 8-12 september 1985*. Cambridge university press.
- Bordería, A. V., Isakov, O., Moratorio, G., Henningsson, R., Agüera-González, S., Organtini, L., Gnädig, N. F., Blanc, H., Alcover, A., Hafenstein, S., Fontes, M., Shomron, N., & Vignuzzi, M. (2015). Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLOS Pathogens*, *11*(5), e1004838.

<https://doi.org/10.1371/journal.ppat.1004838>

- Bou, J.-V., Geller, R., & Sanjuán, R. (2019). Membrane-Associated Enteroviruses Undergo Intercellular Transmission as Pools of Sibling Viral Genomes. *Cell Reports*, 29(3), 714-723.e4. <https://doi.org/10.1016/j.celrep.2019.09.014>
- Bou, J.-V., & Sanjuán, R. (2021). Experimental Evolution Reveals a Genetic Basis for Membrane-Associated Virus Release. *Molecular Biology and Evolution*, 38(2), 358–367. <https://doi.org/10.1093/molbev/msaa208>
- Bou Prados, J. V. (2021). *Transmisión colectiva de enterovirus asociada a membranas: Bases genéticas e implicaciones evolutivas*. Universitat de Valencia.
- Brown, T. A. (2002). *Genomes* (2nd ed.). Wiley-Liss. <http://www.ncbi.nlm.nih.gov/books/NBK21128/>
- Carson, S. D., Chapman, N. M., Hafenstein, S., & Tracy, S. (2011). Variations of Coxsackievirus B3 Capsid Primary Structure, Ligands, and Stability Are Selected for in a Coxsackievirus and Adenovirus Receptor-Limited Environment. *Journal of Virology*, 85(7), 3306–3314. <https://doi.org/10.1128/JVI.01827-10>
- Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348(6300), 454–455. <https://doi.org/10.1038/348454a0>
- Clarke, D. K., Duarte, E. A., Moya, A., Elena, S. F., Domingo, E., & Holland, J. (1993). Genetic bottlenecks and population passages cause profound fitness differences in RNA viruses. *Journal of Virology*, 67(1), 222–228. <https://doi.org/10.1128/jvi.67.1.222-228.1993>
- Coyne, C. B., & Bergelson, J. M. (2006). Virus-Induced Abl and Fyn Kinase

- Signals Permit Coxsackievirus Entry through Epithelial Tight Junctions. *Cell*, 124(1), 119–131. <https://doi.org/10.1016/j.cell.2005.10.035>
- Cuevas, J. M., Domingo-Calap, P., & Sanjuán, R. (2012). The Fitness Effects of Synonymous Mutations in DNA and RNA Viruses. *Molecular Biology and Evolution*, 29(1), 17–20. <https://doi.org/10.1093/molbev/msr179>
- Dalldorf, G., & Sickles, G. M. (1948). An Unidentified, Filtrable Agent Isolated From the Feces of Children With Paralysis. *Science*, 108(2794), 61–62. <https://doi.org/10.1126/science.108.2794.61>
- Darwin, C., 1809-1882. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London: John Murray, 1859. <https://search.library.wisc.edu/catalog/9934839413602122>
- Darwin, C., & Wallace, A. (1858). On the Tendency of Species to form Varieties; and on the Perpetuation of Varieties and Species by Natural Means of Selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, 3(9), 45–62. <https://doi.org/10.1111/j.1096-3642.1858.tb02500.x>
- Dial, C. N., Tate, P. M., Kicmal, T. M., & Mounce, B. C. (2019). Coxsackievirus B3 Responds to Polyamine Depletion via Enhancement of 2A and 3C Protease Activity. *Viruses*, 11(5), 403. <https://doi.org/10.3390/v11050403>
- Drake, J. W. (1993). Rates of spontaneous mutation among RNA viruses. *Proceedings of the National Academy of Sciences*, 90(9), 4171–4175. <https://doi.org/10.1073/pnas.90.9.4171>
- Duffy, S., Shackelton, L. A., & Holmes, E. C. (2008). Rates of evolutionary change in viruses: Patterns and determinants. *Nature Reviews Genetics*, 9(4), 267–276. <https://doi.org/10.1038/nrg2323>

- Ehrenfeld, E. (1982). Poliovirus-induced inhibition of host-cell protein synthesis. *Cell*, 28(3), 435–436. [https://doi.org/10.1016/0092-8674\(82\)90195-7](https://doi.org/10.1016/0092-8674(82)90195-7)
- Elena, S. F., Carrasco, P., Daròs, J., & Sanjuán, R. (2006). Mechanisms of genetic robustness in RNA viruses. *EMBO Reports*, 7(2), 168–173. <https://doi.org/10.1038/sj.embor.7400636>
- Elena, S. F., & Sanjuán, R. (2007). Virus Evolution: Insights from an Experimental Approach. *Annual Review of Ecology, Evolution, and Systematics*, 38(1), 27–52. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095637>
- Escarmís, C., Gómez-Mariano, G., Dávila, M., Lázaro, E., & Domingo, E. (2002). Resistance to extinction of low fitness virus subjected to plaque-to-plaque transfers: Diversification by mutation clustering. *Journal of Molecular Biology*, 315(4), 647–661. <https://doi.org/10.1006/jmbi.2001.5259>
- Fitzgerald, K. D., & Semler, B. L. (2009). Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1789(9–10), 518–528. <https://doi.org/10.1016/j.bbagr.2009.07.004>
- Flanegan, J. B., Petterson, R. F., Ambros, V., Hewlett, N. J., & Baltimore, D. (1977). Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *Proceedings of the National Academy of Sciences*, 74(3), 961–965. <https://doi.org/10.1073/pnas.74.3.961>
- Garmaroudi, F. S., Marchant, D., Hendry, R., Luo, H., Yang, D., Ye, X., Shi, J., & McManus, B. M. (2015). Coxsackievirus B3 replication and

- pathogenesis. *Future Microbiology*, 10(4), 629–653.
<https://doi.org/10.2217/fmb.15.5>
- Gnädig, N. F., Beaucourt, S., Campagnola, G., Bordería, A. V., Sanz-Ramos, M., Gong, P., Blanc, H., Peersen, O. B., & Vignuzzi, M. (2012). Coxsackievirus B3 mutator strains are attenuated in vivo. *Proceedings of the National Academy of Sciences*, 109(34).
<https://doi.org/10.1073/pnas.1204022109>
- Goh, K. C. M., Tang, C. K., Norton, D. C., Gan, E. S., Tan, H. C., Sun, B., Syenina, A., Yousuf, A., Ong, X. M., Kamaraj, U. S., Cheung, Y. B., Gubler, D. J., Davidson, A., St John, A. L., Sessions, O. M., & Ooi, E. E. (2016). Molecular determinants of plaque size as an indicator of dengue virus attenuation. *Scientific Reports*, 6(1), 26100.
<https://doi.org/10.1038/srep26100>
- Hulsebosch, B. M., Omoba, O. S., LoMascolo, N. J., & Mounce, B. C. (2021). *Polyamines mediate enterovirus attachment directly and indirectly through cellular heparan sulfate synthesis* [Preprint]. *Microbiology*.
<https://doi.org/10.1101/2021.10.31.466121>
- Kern, A. D., & Hahn, M. W. (2018). The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution*, 35(6), 1366–1371.
<https://doi.org/10.1093/molbev/msy092>
- Kimura, M. (1968). Evolutionary Rate at the Molecular Level. *Nature*, 217(5129), 624–626. <https://doi.org/10.1038/217624a0>
- Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature*, 267(5608), 275–276.
<https://doi.org/10.1038/267275a0>

- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution* (1st ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511623486>
- Knipe, D. M., & Howley, P. (2013). *Fields Virology*. Wolters Kluwer Health.
<https://books.google.com.uy/books?id=dxlrrMrot3gC>
- Lázaro, E., Escarmís, C., Pérez-Mercader, J., Manrubia, S. C., & Domingo, E. (2003). Resistance of virus to extinction on bottleneck passages: Study of a decaying and fluctuating pattern of fitness loss. *Proceedings of the National Academy of Sciences*, *100*(19), 10830–10835.
<https://doi.org/10.1073/pnas.1332668100>
- Liu, X., Yang, C., Sun, X., Lin, X., Zhao, L., Chen, H., & Jin, M. (2019). Evidence for a novel mechanism of influenza A virus host adaptation modulated by PB 2-627. *The FEBS Journal*, *286*(17), 3389–3400.
<https://doi.org/10.1111/febs.14867>
- Lynch, M., Bürger, R., Butcher, D., & Gabriel, W. (1993). The Mutational Meltdown in Asexual Populations. *Journal of Heredity*, *84*(5), 339–344.
<https://doi.org/10.1093/oxfordjournals.jhered.a111354>
- Mandary, M. B., Masomian, M., Ong, S.-K., & Poh, C. L. (2020). Characterization of Plaque Variants and the Involvement of Quasi-Species in a Population of EV-A71. *Viruses*, *12*(6), 651.
<https://doi.org/10.3390/v12060651>
- Manrubia, S. C., Escarmís, C., Domingo, E., & Lázaro, E. (2005). High mutation rates, bottlenecks, and robustness of RNA viral quasispecies. *Gene*, *347*(2), 273–282. <https://doi.org/10.1016/j.gene.2004.12.033>
- Martino, T. A., Liu, P., & Sole, M. J. (1994). Viral infection and the pathogenesis

- of dilated cardiomyopathy. *Circulation Research*, 74(2), 182–188.
<https://doi.org/10.1161/01.RES.74.2.182>
- Mattenberger, F., Vila-Nistal, M., & Geller, R. (2021). Increased RNA virus population diversity improves adaptability. *Scientific Reports*, 11(1), 6824.
<https://doi.org/10.1038/s41598-021-86375-z>
- Mayrose, I., Stern, A., Burdelova, E. O., Sabo, Y., Laham-Karam, N., Zamostiano, R., Bacharach, E., & Pupko, T. (2013). Synonymous site conservation in the HIV-1 genome. *BMC Evolutionary Biology*, 13(1), 164. <https://doi.org/10.1186/1471-2148-13-164>
- McDonald, S., Block, A., Beaucourt, S., Moratorio, G., Vignuzzi, M., & Peersen, O. B. (2016). Design of a Genetically Stable High Fidelity Coxsackievirus B3 Polymerase That Attenuates Virus Growth in Vivo. *Journal of Biological Chemistry*, 291(27), 13999–14011.
<https://doi.org/10.1074/jbc.M116.726596>
- McLachlan, A. D. (1971). Tests for comparing related amino-acid sequences. Cytochrome c and cytochrome c551. *Journal of Molecular Biology*, 61(2), 409–424. [https://doi.org/10.1016/0022-2836\(71\)90390-1](https://doi.org/10.1016/0022-2836(71)90390-1)
- Moratorio, G., Henningsson, R., Barbezange, C., Carrau, L., Bordería, A. V., Blanc, H., Beaucourt, S., Poirier, E. Z., Vallet, T., Boussier, J., Mounce, B. C., Fontes, M., & Vignuzzi, M. (2017). Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nature Microbiology*, 2(8), 17088. <https://doi.org/10.1038/nmicrobiol.2017.88>
- Moser, L. A., Boylan, B. T., Moreira, F. R., Myers, L. J., Svenson, E. L., Fedorova, N. B., Pickett, B. E., & Bernard, K. A. (2018). Growth and adaptation of Zika virus in mammalian and mosquito cells. *PLOS*

- Neglected Tropical Diseases*, 12(11), e0006880.
<https://doi.org/10.1371/journal.pntd.0006880>
- Muller, H. J. (1932). Some Genetic Aspects of Sex. *The American Naturalist*, 66(703), 118–138. <https://doi.org/10.1086/280418>
- Muller, H. J. (1964). THE RELATION OF RECOMBINATION TO MUTATIONAL ADVANCE. *Mutation Research*, 106, 2–9.
[https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8)
- National Academies of Sciences Engineering and Medicine (US). (n.d.). *Evolution Resources at the National Academies: Definitions*. Retrieved July 18, 2022, from <https://www.nationalacademies.org/evolution/definitions>
- Ohta, T. (1992). The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*, 23(1), 263–286.
<https://doi.org/10.1146/annurev.es.23.110192.001403>
- Orr, H. A. (2009). Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10(8), 531–539. <https://doi.org/10.1038/nrg2603>
- Peña, M. de la, Elena, S. F., & Moya, A. (2000). EFFECT OF DELETERIOUS MUTATION-ACCUMULATION ON THE FITNESS OF RNA BACTERIOPHAGE MS2. *Evolution*, 54(2), 686.
[https://doi.org/10.1554/0014-3820\(2000\)054\[0686:EODMAO\]2.0.CO;2](https://doi.org/10.1554/0014-3820(2000)054[0686:EODMAO]2.0.CO;2)
- Pereira-Gómez, M. (2016). *Genetic and Evolutionary Determinants of Mutation Rate in Bacteriophage ϕ X174*.
- Plotkin, J. B., Dushoff, J., & Fraser, H. B. (2004). Detecting selection using a single genome sequence of *M. tuberculosis* and *P. falciparum*. *Nature*, 428(6986), 942–945. <https://doi.org/10.1038/nature02458>

- Ramsingh, A. I., Caggana, M., & Ronstrom, S. (1995). Genetic mapping of the determinants of plaque morphology of coxsackievirus B4. *Archives of Virology*, 140(12), 2215–2226. <https://doi.org/10.1007/BF01323241>
- Rye, C., Wise, R., Jurukovski, V., DeSaix, J., Choi, J., & Avissar, Y. (2016). *Biology*. OpenStax. <https://openstax.org/books/biology/pages/19-2-population-genetics>
- Sanjuán Verdeguer, R. (2005). *ROBUSTEZ GENÉTICA, EPISTASIA Y EVOLUCIÓN DE LOS VIRUS DE RNA*.
- Selinka, H.-C., Wolde, A., Sauter, M., Kandolf, R., & Klingel, K. (2004). Virus-receptor interactions of coxsackie B viruses and their putative influence on cardiotropism. *Medical Microbiology and Immunology*, 193(2–3), 127–131. <https://doi.org/10.1007/s00430-003-0193-y>
- Shieh, J. T. C., & Bergelson, J. M. (2002). Interaction with Decay-Accelerating Factor Facilitates Coxsackievirus B Infection of Polarized Epithelial Cells. *Journal of Virology*, 76(18), 9474–9480. <https://doi.org/10.1128/JVI.76.18.9474-9480.2002>
- Takeuchi, K., Takeda, M., Miyajima, N., Kobune, F., Tanabayashi, K., & Tashiro, M. (2002). Recombinant Wild-Type and Edmonston Strain Measles Viruses Bearing Heterologous H Proteins: Role of H Protein in Cell Fusion and Host Cell Specificity. *Journal of Virology*, 76(10), 4891–4900. <https://doi.org/10.1128/JVI.76.10.4891-4900.2002>
- Visser, J. A. G. M., Hermisson, J., Wagner, G. P., Meyers, L. A., Bagheri-Chaichian, H., Blanchard, J. L., Chao, L., Cheverud, J. M., Elena, S. F., Fontana, W., Gibson, G., Hansen, T. F., Krakauer, D., Lewontin, R. C., Ofria, C., Rice, S. H., Dassow, G. von, Wagner, A., &

- Whitlock, M. C. (2003). PERSPECTIVE: EVOLUTION AND DETECTION OF GENETIC ROBUSTNESS. *Evolution*, 57(9), 1959–1972. <https://doi.org/10.1111/j.0014-3820.2003.tb00377.x>
- Wang, Y., & Pfeiffer, J. K. (2016). Emergence of a Large-Plaque Variant in Mice Infected with Coxsackievirus B3. *MBio*, 7(2), e00119-16. <https://doi.org/10.1128/mBio.00119-16>
- Wessels, E., Notebaart, R. A., Duijsings, D., Lanke, K., Vergeer, B., Melchers, W. J. G., & van Kuppeveld, F. J. M. (2006). Structure-Function Analysis of the Coxsackievirus Protein 3A. *Journal of Biological Chemistry*, 281(38), 28232–28243. <https://doi.org/10.1074/jbc.M601122200>
- Wilke, C. O., & Adami, C. (2003). Evolution of mutational robustness. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 522(1–2), 3–11. [https://doi.org/10.1016/S0027-5107\(02\)00307-X](https://doi.org/10.1016/S0027-5107(02)00307-X)
- Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E., & Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412(6844), 331–333. <https://doi.org/10.1038/35085569>
- Wright, S. (1932). The Roles of Mutation, Inbreeding, crossbreeding and Selection in Evolution. *Proceedings of the XI International Congress of Genetics*, 8, 209–222.
- Yuste, E., López-Galíndez, C., & Domingo, E. (2000). Unusual Distribution of Mutations Associated with Serial Bottleneck Passages of Human Immunodeficiency Virus Type 1. *Journal of Virology*, 74(20), 9546–9552. <https://doi.org/10.1128/JVI.74.20.9546-9552.2000>
- Zautner, A. E., Körner, U., Henke, A., Badorff, C., & Schmidtke, M. (2003). Heparan Sulfates and Coxsackievirus-Adenovirus Receptor: Each One

Mediates Coxsackievirus B3 PD Infection. *Journal of Virology*, 77(18), 10071–10077. <https://doi.org/10.1128/JVI.77.18.10071-10077.2003>

Zell, R., Delwart, E., Gorbalenya, A. E., Hovi, T., King, A. M. Q., Knowles, N. J., Lindberg, A. M., Pallansch, M. A., Palmenberg, A. C., Reuter, G., Simmonds, P., Skern, T., Stanway, G., Yamashita, T., & ICTV Report Consortium. (2017). ICTV Virus Taxonomy Profile: Picornaviridae. *Journal of General Virology*, 98(10), 2421–2422. <https://doi.org/10.1099/jgv.0.000911>

Zhang, J. (2005). On the Evolution of Codon Volatility. *Genetics*, 169(1), 495–501. <https://doi.org/10.1534/genetics.104.034884>

Anexos

Anexo 1 - Cantidades utilizadas en cultivo celular

Tabla Anexo 1. Cantidades de reactivos utilizados para cada etapa del pasaje de líneas celulares.				
	Área superficial (cm²)	Volumen de medio de cultivo (mL)	Volumen de PBS para lavado (mL)	Volumen de tripsina (mL)
Frasco T175	175	25	25	5
Frasco T75	75	10	10	2
Frasco T25	25	5	5	0,5

Anexo 2 - Generación de la progenie 0 y 1

Para la generación de las progenie 0 los plásmidos fueron linealizados con la enzima Sall (Invitrogen). A partir del plásmido linealizado, se realizó la transcripción *in vitro* (IVT) en un volumen final de 25 μ L y para el mix se utilizó 5 μ L de buffer T7 5x (Invitrogen), 1 μ L de T7 RNA pol (20 U/ μ L), 2 μ L de rNTPs (25nM), 1 μ L de Inhibidor de RNAsas (40 U/ μ L, NZTech), 1.25 μ M de DTT (0,1 M, Invitrogen) y 500 ng de ADN lineal purificado. Se complementó con H₂O con dietil pirocarbonato (DEPC) hasta llegar al volumen final. Se incubó 1 hora y 30 minutos a 37°C. Luego de la transcripción se trató la muestra con 1 μ L de Turbo DNase I (Invitrogen) por 15 minutos a 37°C con el fin de degradar el molde plasmídico lineal.

La IVT se purificó utilizando el Kit Quick RNA Miniprep Plus (Zymo Research) – Reaction Clean Up, según el protocolo recomendado por el fabricante.

A fin de evaluar los ARN generados se realizó una electroforesis en gel de agarosa para lo cual, se calentaron 5 μ L del ARN generado a 70°C con 1 μ L de Loading Buffer 6x (Thermo Fischer) por un minuto y posteriormente se pusieron en hielo hasta el momento de la carga del gel. Se preparó un gel de agarosa no desnaturante al 1% y se corrió la electroforesis a 100 V por 25 minutos con el fin de ver la integridad del transcripto. Como marcador de peso molecular se utilizó 1 kb Plus (Invitrogen).

El ARN purificado fue transfectado en placas de 6 pocillos con células HeLa. En cada pocillo se sembraron 250000 células el día antes para tener una confluencia del 70% al momento de la transfección y se siguió, por pocillo, el siguiente protocolo:

- 1) Diluir 6 μ L de Lipofectamine 3000 (Invitrogen) en 150 μ L de medio Opti-MEM® (Sol A).
- 2) Diluir ARN (400 ng) en medio Opti-MEM® (Vf=150 μ L) y agregar 6 μ L de P3000 reagent (Sol B).
- 3) Mezclar 150 μ L de Sol A y 150 μ L de Sol B en ratio 1:1.
- 4) Incubar 5 minutos a temperatura ambiente.

- 5) Durante la incubación, cambiar el medio a medio DMEM + 1,5% SFB sin antibiótico (1 mL por pocillo), previo lavado con 1-2 mL de PBS 1X.
- 6) Añadir la mezcla Opti-MEM + lipofectamina + ARN (Sol A + Sol B) a las células de a gotas, agregar 1 mL más de medio DMEM + 1,5% SFB sin antibiótico e incubar a 37°C 5% CO₂.

Como control negativo de transfección (MOCK) se realizó el mismo procedimiento con la excepción de que en el paso 2) de la transfección no se agregó ARN.

48 horas después de la transfección se evaluó el efecto citopático (CPE). Si en las células transfectadas se veía CPE y en el MOCK no, se asumía que se tenía la progenie 0 (p0) de CVB3. A partir de la p0, se generó la progenie 1 (p1) infectando células HeLa con una confluencia del 70% en un frasco T-75 con 500 µL de p0. El MOCK de la transfección se utilizó como control negativo de infección.

Luego de la generación de cada progenie se realizan 3 ciclos de congelado-descongelado y en todas las infecciones el medio utilizado era DMEM con 2% de FBS y 1% de P/S. Además finalizada la infección se clarificaron los sobrenadantes. Este paso consiste en tomar el sobrenadante de la infección, centrifugar 5 minutos a 6000 g, tomar el sobrenadante y descartar el pellet celular.

Anexo 3 - Extracción de ARN con TRIzol

- 1) Se agrega la muestra de superficies a 0,5 mL de TRIzol, mezclar por inversión y dejar incubar 10 minutos a temperatura ambiente.
- 2) Agregar 200 μ L de cloroformo. Asegurarse de que el tubo esté bien tapado y mezclar vigorosamente. Luego incubar por 3 minutos a temperatura ambiente.
- 3) Centrifugar la muestra por 10 minutos a 12000 g a 4°C. Como producto de la centrifugación se obtendrán tres fases, una fase inferior orgánica roja que contiene fenol y cloroformo, una interfase y una fase superior acuosa transparente.
- 4) Transferir 400 μ L de la fase acuosa que contiene el ARN a un nuevo tubo.
- 5) Agregar 500 μ L de isopropanol con Glycoblue (1 μ L) y dejar incubar 10 minutos a temperatura ambiente.
- 6) Centrifugar por 10 minutos a 12000 g y a 4°C. Terminado el centrifugado, el ARN se encontrará en el fondo y se verá como un precipitado azul parecido a un gel.
- 7) Descartar el sobrenadante por volcado.
- 8) Resuspender el pellet en 500 μ L de etanol al 70% y centrifugar a 12000 g por 5 minutos a 4°C.
- 9) Descartar el sobrenadante por volcado. Dejar secar el pellet al aire.
- 10) Resuspender el pellet en 40 μ L de agua libre de ARNasa e incubar a 60-65°C por 5 minutos para que el ARN se solubilice bien en el agua.

Anexo 4 - Extracción de ARN con QIAamp® Viral RNA

Procedimiento

- 1) Agregar 500 µL de Buffer AVL con carrier RNA (10 µL de carrier por cada mL de Buffer AVL) en un tubo de 1.5 ml.
- 2) Agregar 300 µl de muestra. Mezclar bien.
- 3) Incubar a 10 minutos. (LISIS)
- 4) Agregar 500 µl EtOH (96–100%) y mezclar bien.
- 5) Pasar 700 µl a la columna QIAamp Mini (en tubo colector). Centrifugar a 6000 x g por 1 minuto. Cambiar a tubo colector nuevo.
- 6) Repetir hasta pasar todo el volumen de muestra por la columna. Cambiar a tubo colector nuevo.
- 7) Agregar 500 µl de Buffer AW1. Centrifugar a 6000 x g por 1 minuto. Descartar Flow Through.
- 8) Agregar 500 µl de Buffer AW2. Centrifugar a 20,000 x g por 3 minutos. Descartar Flow through (paso opcional) y volver a Centrifugar a 20,000 x g por 1 minuto.
- 9) Pasar la columna a un tubo de 1.5 mL. Agregar 60 µl Buffer AVE a temperatura ambiente. Incubar a temperatura ambiente por 1 minuto.
- 10) Centrifugar a 6000 x g por 1 minuto.

Anexo 5 - Protocolo de Extracción de ARN viral con robot Pureprep_Molgen

1) Preparar Soluciones:

- Buffer de Lisis:

- Preparar: 200 μ L de buffer de lisis + 10 μ L proteinasa K + 1 μ L poly-A RNA para cada muestra
- Cargar en deepwell 211 μ L por pocillo

-Buffer U1 + BEADS:

- Preparar en Falcon: 400 μ L de Buffer U1 + 20 μ L de BEADS para cada muestra

- Buffers de Lavado:

- Cargar 2 Deepwells con 800 μ L de Wash Buffer I /pocillo
- Cargar 1 Deepwell con 800 μ L de Wash Buffer II /pocillo

-Buffer de Elución:

- Cargar placa de Elución con 100 μ L de Buffer de Elución/pocillo

2) Cargar 300 μ L de la muestra en deepwell con Buffer de Lisis

3) Agitar y dejar 10' a TA

4) Agregar a cada pocillo 420 μ L de BufferU1 + BEADS

5) Cargar en Robot:

Muestra: posición 2

Wash Buffer I: posición 3 y 4

Wash Buffer II: posición 5

Buffer de Elución: posición 8

* En posición 1 dejar una placa vacía y colocar las fundas dentro de la placa.

Anexo 6 - Generación de librerías para secuenciación en plataforma GridION

Paso 1: purificación de los amplicones

- 1) Mezclar el mismo volumen de *beads* y de producto de PCR por 10 minutos y dejar incubando en el Hula mixer.
- 2) Centrifugar brevemente, poner los tubos en el rack magnético por 5 minutos para concentrar las *beads* y retirar el sobrenadante.
- 3) Agregar 200 μL de etanol 80% con las *beads* unidas al imán y retirarlo.
- 4) Repetir paso anterior.
- 5) Agregar 15 μL de H₂O libre de ARN/ADNsa y resuspender el pellet de *beads*. Incubar por 2 minutos a temperatura ambiente.
- 6) Concentrar las *beads* en el fondo, tomar el eluido y ponerlo en un tubo LoBind.

Paso 2: Endprep

- 1) Mezclar 12,5 μL de los productos purificados con 1,75 μL de buffer de Ultra II End-prep reaction y 0,75 μL de Ultra II End-prep enzyme mix. En estos 12,5 μL debe haber un total de 75 femtomoles para cada muestra.
- 2) Este mix se debe incubar a 20 por 5 minutos y 65 por 5 minutos.

Paso 3: Barcoding

- 1) Para el barcoding se mezclan 3,5 μL de H₂O, 4 μL del Endprep del paso anterior, 2,5 μL de Barcode y 10 μL de Blunt TA ligase MasterMix (New England Biolabs).
- 2) Poolear todas las muestras con barcodes en un tubo de 1,5 mL Eppendorf DNA LoBind.
- 3) Agregar 0,4 volúmenes de *beads* y mezclar por pipeteo. Incubar 10 minutos a temperatura ambiente en Hula mixer.
- 4) Peletear las *beads* con imán por 5 minutos y luego pipetear el sobrenadante.
- 5) Lavar agregando 700 μL de Long Fragment Buffer (LFB). Resuspender las *beads* agitando suavemente y volver a peletear las *beads* con imán. Remover el sobrenadante mediante pipeteo y descartar.
- 6) Repetir paso anterior.
- 7) Lavar las *beads* con etanol al 80% evitando desintegrar el pellet. Remover el etanol con pipeta y descartar.
- 8) Centrifugar brevemente y poner el tubo sobre el imán. Retirar el etanol residual en caso de que haya.
- 9) Agregar 25 μL de H₂O libre de ARN/ADNsa e incubar a temperatura ambiente por 2 minutos.

- 10) Peletear las beads hasta que el eluido esté transparente. Retirar el sobrenadante y poner en un tubo Eppendorf DNA LoBind. Cuantificar con Qubit.

Paso 4: Ligación de adaptadores

- 1) Preparar mix que tenga 5 μL de Quick T4 DNA Ligase (New England BioLabs), 10 μL de NEBNext Quick Ligation reaction Buffer 5x (New England BioLabs), 5 μL de Adapter Mix II (New England BioLabs) , 20 μL de las muestras pooleadas con barcode y 10 μL de H₂O libre de ARN/ADNsas.
- 2) Incubar por 30 minutos a temperatura ambiente.
- 3) Agregar 0,4 volúmenes de beads y mezclar por pipeteo. Incubar 10 minutos a temperatura ambiente en Hula mixer.
- 4) Peletear las beads con imán por 5 minutos y luego pipetear el sobrenadante.
- 5) Lavar agregando 700 μL de Long Fragment Buffer (LFB). Resuspender las beads agitando suavemente y volver a peletear las beads con imán. Remover el sobrenadante mediante pipeteo y descartar
- 6) Repetir paso anterior.
- 7) Centrifugar brevemente y poner el tubo sobre el imán. Retirar el sobrenadante residual en caso de que haya.
- 8) Mover el tubo del imán, resuspender las beads en 7 μL de Elution Buffer e incubar 10 minutos a temperatura ambiente.
- 9) Colocar tubo en el imán nuevamente hasta que el eluido esté transparente. Recuperar los 7 μL de eluido y poner en un tubo Eppendorf DNA LoBind. Cuantificar con Qubit.

Paso 5: Carga de la muestra en Flow Cell

- 1) Mezclar 5 μL de librerías, 10 μL de Loading Beads y 15 μL de Sequencing Buffer.
- 2) Primear la Flow Cell con 117 μL de Flush Buffer y 3 μL de Flush Tether.
- 3) Cargar los 30 μL de librería en el Sample Port.
- 4) Dejar secuenciar por 24 horas y en caso de que la calidad de las lecturas y la velocidad de translocación bajen, se debe recargar la Flow Cell con 100 μL de Flush Buffer.

Paso 6: Análisis de las lecturas

- 1) Analizar con el protocolo artic (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). La única excepción es que se cambiaron el mínimo y máximo de largo de las lecturas a analizar a 6000 y 8000 respectivamente. Como consenso se utilizó la secuencia JX312064_1.

Anexo 7 - Tablas de mutaciones fijadas a lo largo de la evolución experimental por deriva génica

Tabla Anexo 2. Mutaciones sinónimas fijadas durante la evolución experimental. Nuc.=nucleótido; Pos.=posición; CDS=región de codificación; AA=aminoácido.								
Genotipo	Mutación nucleotídica			Posición CDS	Posición AA	AA	Proteína	Región
	Nuc. orig	Pos. Nuc.	Nuc. nuevo					
WT Li1.P10	G	2803	A	687	117	T	VP1	P1
WT Li1.P10	T	5504	C	1588	48	L	3C	P3
WT Li2.p10	A	1159	G	139	70	Q	VP2	P1
WT Li2.p10	T	1267	C	175	106	Y	VP2	P1
WT Li2.p10	T	4108	C	1123	23	V	2C	P2
WT Li2.p10	A	4270	G	1176	76	L	2C	P2
WT Li2.p10	G	4516	A	1258	158	P	2C	P2
WT Li2.p10	T	4711	C	1323	223	N	2C	P2
WT Li2.p10	T	6307	C	1855	132	S	3D	P3
WT Li2.p10	A	6385	G	1881	158	V	3D	P3
WT Li2.p10	T	7132	C	2130	407	D	3D	P3
WT Li3.p10	A	826	G	28	28	T	VP4	P1
WT Li3.p10	T	2023	C	427	95	F	VP3	P1
WT Li3.p10	T	4921	C	1393	293	I	2C	P2
WT Li3.p10	A	7060	G	2106	383	E	3D	P3
MoreV Li1.p10	T	877	C	45	45	D	VP4	P1
MoreV Li1.p10	T	4108	A	1122	22	A	2C	P2
MoreV Li1.p10	T	4924	C	1394	294	D	2C	P2
MoreV Li1.p10	T	6670	C	1976	253	L	3D	P3
MoreV Li1.p10	T	7172	C	2144	421	L	3D	P3

MoreV Li2.p10	G	499	A				IRES	5'UTR
MoreV Li2.p10	G	2215	C	491	159	G	VP3	P1
MoreV Li2.p10	A	4927	G	1395	295	R	2C	P2
MoreV Li2.p10	T	5746	C	1668	128	G	3C	P3
MoreV Li2.p10	T	6271	C	1843	120	Y	3D	P3
MoreV Li3.p10	A	317	G				5'UTR	5'UTR
MoreV Li3.p10	T	2908	C	722	152	D	VP1	P1
MoreV Li3.p10	C	3115	T	791	221	V	VP1	P1
MoreV Li3.p10	C	3574	G	944	90	R	2A	P2
LessV Li1.p10	A	253	G				5'UTR	5'UTR
LessV Li1.p10	T	1505	A	255	186	I	VP2	P1
LessV Li1.p10	T	3884	C	1048	47	V	2B	P2
LessV Li1.p10	T	4691	C	1317	99	F	2C	P2
LessV Li1.p10	A	5309	G	1523	5	G	3B	P3
LessV Li1.p10	A	5610	C	1623	83	F	3C	P3
LessV Li2.p10	A	598	C				5'UTR	5'UTR
LessV Li2.p10	A	706	G				5'UTR	5'UTR
LessV Li2.p10	A	3872	G	1044	43	L	2B	P2
LessV Li2.p10	T	4622	C	1294	194	V	2C	P2
LessV Li2.p10	T	4790	C	1350	250	V	2C	P2
LessV Li2.p10	G	5243	T	1501	72	V	3A	P3
LessV Li2.p10	C	5729	T	1663	123	G	3C	P3
LessV Li2.p10	C	5777	T	1679	139	N	3C	P3
LessV Li2.p10	T	6623	C	1961	238	D	3D	P3
LessV Li3.p10	G	114	A				5'UTR	5'UTR
LessV Li3.p10	G	5411	A	1557	17	T	3C	P3

Tabla Anexo 3. Mutaciones no sinónimas fijadas durante la evolución experimental. Nuc.=nucleótido; Pos.=posición; CDS=región de codificación; AA=aminoácido.

Genotipo	Mutación nucleotídica (N)			Pos. CDS	Modificación aminoacídica (AA)			Proteína	Región
	Nuc. original	Pos. N	Nuc. mutado		AA original	Pos. AA	AA mutado		
WT L2.p10	G	6879	A	2046	R	323	K	3D	P3
WT L3.p10	A	1925	C	395	N	63	H	VP3	P1
WT L3.p10	G	5066	A	1442	A	13	T	3A	P3
MoreV L1.p10	C	2438	G	566	Q	234	E	VP3	P1
MoreV L1.p10	G	3810	A	1023	C	22	Y	2B	P2
MoreV L2.p10	A	1925	T	395	N	63	Y	VP3	P1
MoreV L2.p10	A	3465	G	908	Q	54	R	2A	P2
MoreV L2.p10	A	4964	T	1408	T	308	S	2C	P2
MoreV L2.p10	C	5165	T	1475	P	46	S	3A	P3
MoreV L3.p10	A	1505	G	255	I	186	V	VP2	P1

Anexo 8 - Calidad de la secuenciación de las poblaciones evolucionadas en plataforma GridION

Tabla Anexo 4. Resumen general de la calidad de la secuenciación realizada en la plataforma GridION de las poblaciones virales evolucionadas.										
Resumen general	WT Li1.p10	WT Li2.p10	WT Li3.p10	MoreV Li1.p10	MoreV Li2.p10	MoreV Li3.p10	LessV Li1.p10	LessV Li2.p10	LessV Li3.p10	Control positivo SARS
Largo promedio de las lecturas	6,362.3	6,937.8	5,798.3	7,200.8	7,325.1	7,221.5	7,224.0	7,221.5	7,220.3	1,314.2
Calidad promedio de las lecturas	12	12.2	12.1	11.7	11.7	11.7	11.8	11.7	11.8	11.8
Largo mediano de las lecturas	7,227.0	7,234.0	7,239.0	7,248.0	7,244.0	7,253.0	7,254.0	7,253.0	7,254.0	1,127.0
Calidad mediana de las lecturas	12.1	12.2	12	11.7	11.8	11.8	11.8	11.8	11.9	11.9
Número de lecturas	649	523	88	309	651	1106	649	1106	753	40
Largo de lectura del N50	7,243.0	7,239.0	7,252.0	7,249.0	7,247.0	7,254.0	7,255.0	7,254.0	7,255.0	1,149.0
Largo de lectura STDEV	1711	1149.6	2653.1	951.3	1169.9	688.6	640.6	688.6	788.7	416.5
Total bases	4,129,110.00	3,628,445.00	510,249.00	2,225,048.00	4,768,665.00	7,986,925.00	4,688,394.00	7,986,925.00	5,436,922.00	52,566.00

Número, porcentaje y megabases de los reads en relación a la calidad										
>Q5	649 (100.0%) 4.1Mb	523 (100.0%) 3.6Mb	88 (100.0%) 0.5Mb	309 (100.0%) 2.2Mb	651 (100.0%) 4.8Mb	1106 (100.0%) 8.0Mb	649 (100.0%) 4.7Mb	1106 (100.0%) 8.0Mb	753 (100.0%) 5.4Mb	40 (100.0%) 0.1Mb
>Q7	649 (100.0%) 4.1Mb	523 (100.0%) 3.6Mb	88 (100.0%) 0.5Mb	309 (100.0%) 2.2Mb	651 (100.0%) 4.8Mb	1106 (100.0%) 8.0Mb	649 (100.0%) 4.7Mb	1106 (100.0%) 8.0Mb	753 (100.0%) 5.4Mb	40 (100.0%) 0.1Mb
>Q10	563 (86.7%) 3.5Mb	446 (85.3%) 3.1Mb	73 (83.0%) 0.4Mb	254 (82.2%) 1.8Mb	548 (84.2%) 4.0Mb	943 (85.3%) 6.8Mb	566 (87.2%) 4.1Mb	943 (85.3%) 6.8Mb	629 (83.5%) 4.5Mb	28 (70.0%) 0.0Mb
>Q12	335 (51.6%) 2.1Mb	289 (55.3%) 2.0Mb	44 (50.0%) 0.2Mb	132 (42.7%) 0.9Mb	280 (43.0%) 2.0Mb	487 (44.0%) 3.5Mb	297 (45.8%) 2.1Mb	487 (44.0%) 3.5Mb	351 (46.6%) 2.5Mb	19 (47.5%) 0.0Mb
>Q15	22 (3.4%) 0.1Mb	46 (8.8%) 0.3Mb	4 (4.5%) 0.0Mb	2 (0.6%) 0.0Mb	13 (2.0%) 0.1Mb	12 (1.1%) 0.1Mb	14 (2.2%) 0.1Mb	12 (1.1%) 0.1Mb	14 (1.9%) 0.1Mb	3 (7.5%) 0.0Mb
Mejores 5 scores de calidad y el largo de esa secuencia										
1	16.9 (4089)	18.6 (6751)	16.4 (453)	15.7 (6038)	16.4 (7266)	16.8 (7296)	16.6 (7299)	16.8 (7296)	15.9 (7305)	16.0 (1157)
2	16.8 (7292)	18.5 (6767)	16.1 (356)	15.0 (7275)	15.7 (7294)	16.4 (7292)	16.1 (7287)	16.4 (7292)	15.8 (7293)	15.7 (1147)

3	16.6 (7293)	18.1 (6725)	15.4 (6375)	15.0 (7252)	15.6 (7292)	16.1 (7274)	16.1 (7291)	16.1 (7274)	15.8 (7305)	15.1 (1128)
4	16.4 (4111)	17.8 (6750)	15.2 (7262)	14.7 (7284)	15.4 (7262)	15.7 (7276)	15.8 (5874)	15.7 (7276)	15.7 (7275)	14.6 (1083)
5	16.0 (7304)	17.7 (6713)	15.0 (6815)	14.6 (7272)	15.2 (7240)	15.7 (7250)	15.7 (7268)	15.7 (7250)	15.7 (7264)	14.6 (1111)
Las 5 lecturas más largas y su score de calidad promedio										
1	15319 (8.4)	13492 (8.3)	7517 (11.0)	14479 (10.4)	21496 (8.0)	14536 (10.9)	14683 (10.5)	14536 (10.9)	14558 (13.1)	2373 (9.7)
2	14397 (9.4)	13204 (8.1)	7407 (9.4)	14476 (12.4)	17067 (9.3)	14449 (11.6)	14571 (11.6)	14449 (11.6)	13443 (8.9)	2271 (9.4)
3	13379 (8.4)	13198 (8.1)	7362 (12.4)	12959 (8.4)	14479 (11.4)	14268 (8.4)	13115 (8.8)	14268 (8.4)	13330 (8.2)	2172 (13.6)
4	13376 (8.6)	8412 (9.4)	7355 (10.2)	9596 (11.6)	14446 (10.9)	13491 (8.9)	7564 (10.8)	13491 (8.9)	13134 (8.0)	2121 (9.5)
5	13003 (8.4)	7952 (8.9)	7326 (13.3)	8516 (14.1)	13535 (8.2)	13052 (8.2)	7515 (10.0)	13052 (8.2)	13124 (8.0)	2108 (10.7)

Anexo 9 - Publicaciones realizadas durante el desarrollo de la tesis

Durante el transcurso de esta tesis participé en varias líneas de trabajo que tenía el laboratorio debido al contexto pandémico. A partir del trabajo en estas se publicaron 4 trabajos que se muestran a continuación:

About the Author

Dr. Jayasinghearachchi is a senior lecturer at General Sir John Kotelawala Defense University, Ratmalana. Her research interests include molecular evolution of human pathogens and their epidemiology.

References

1. Corea EM, de Silva AD, Thevanesam V. Melioidosis in Sri Lanka. *Trop Med Infect Dis*. 2018;3:22. <https://doi.org/10.3390/tropicalmed3010022>
2. Ko WC, Cheung BM, Tang HJ, Shih HI, Lau YJ, Wang LR, et al. Melioidosis outbreak after typhoon, southern Taiwan. *Emerg Infect Dis*. 2007;13:896–8. <https://doi.org/10.3201/eid1306.060646>
3. Merritt A, Inglis TJ, Chidlow G, Harnett G. PCR-based identification of *Burkholderia pseudomallei*. *Rev Inst Med Trop São Paulo*. 2006;48:239–44. <https://doi.org/10.1590/S0036-46652006000500001>
4. Jayasinghearachchi HS, Corea EM, Krishnananthasivam S, Sathkumara HD, Francis VR, Abeysekere TR, et al. Whole-genome sequences of eight clinical isolates of *Burkholderia pseudomallei* from melioidosis patients in eastern Sri Lanka. *Microbiol Resour Announc*. 2019;8:e00645–19. <https://doi.org/10.1128/MRA.00645-19>
5. Okonechnikov K, Golosova O, Fursov M; UGENE team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28:1166–7. <https://doi.org/10.1093/bioinformatics/bts091>
6. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al.; Simon Fraser University Research Computing Group. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. 2017;45(W1):W30–5. <https://doi.org/10.1093/nar/gkx343>
7. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44(W1):W16–21. <https://doi.org/10.1093/nar/gkw387>
8. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;15;28(12):1647–9. <https://doi.org/10.1093/bioinformatics/bts199>
9. Sitthidet C, Stevens JM, Chantratita N, Currie BJ, Peacock SJ, Korbsrisate S, et al. Prevalence and sequence diversity of a factor required for actin-based motility in natural populations of *Burkholderia* species. *J Clin Microbiol*. 2008;46:2418–22. <https://doi.org/10.1128/JCM.00368-08>
10. Holden MT, Titball RW, Peacock SJ, Cerdeño-Tárraga AM, Atkins T, Crossman LC, et al. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. *Proc Natl Acad Sci U S A*. 2004;101:14240–5. <https://doi.org/10.1073/pnas.0403302101>

Address for correspondence: Enoka M. Corea, Department of Microbiology, Faculty of Medicine, University of Colombo, Colombo, Sri Lanka; email: enokac@micro.cmb.ac.lk; Aruna D. De Silva, Department of Paraclinical Sciences, Faculty of Medicine, General Sir John Kotelawala Defense University, Ratmalana, Sri Lanka; email: dharshan_fom@kdu.ac.lk

Real-Time Genomic Surveillance for SARS-CoV-2 Variants of Concern, Uruguay

Natalia Rego,¹ Alicia Costáble,¹ Mercedes Paz,¹ Cecilia Salazar,¹ Paula Perbolianachis,¹ Lucía Spangenberg, Ignacio Ferrés, Rodrigo Arce, Alvaro Fajardo, Mailen Arleo, Tania Possi, Natalia Reyes, Ma Noel Bentancor, Andrés Lizasoain, María José Benítez, Viviana Bortagaray, Ana Moller, Gonzalo Bello, Ighor Arantes, Mariana Brandes, Pablo Smircich, Odhille Chappos, Melissa Duquía, Belén González, Luciana Griffero, Mauricio Méndez, Ma Pía Techera, Juan Zanetti, Bernardina Rivera, Matías Maidana, Martina Alonso, Cecilia Alonso, Julio Medina, Henry Albornoz, Rodney Colina, Veronica Noya, Gregorio Iraola, Tamara Fernández-Calero, Gonzalo Moratorio, Pilar Moreno

Author affiliations: Institut Pasteur, Montevideo, Uruguay (N. Rego, A. Costáble, M. Paz, C. Salazar, P. Perbolianachis, L. Spangenberg, I. Ferrés, R. Arce, A. Fajardo, M. Brandes, B. Rivera, M. Maidana, M. Alonso, G. Iraola, T. Fernandez-Calero, G. Moratorio, P. Moreno); Universidad de la República, Montevideo, Uruguay (A. Costáble, P. Perbolianachis, R. Arce, A. Fajardo, A. Lizasoain, M.J. Benítez, V. Bortagaray, A. Moller, O. Chappos, M. Duquía, B. González, L. Griffero, M. Méndez, M.P. Techera, J. Zanetti, C. Alonso, R. Colina, G. Moratorio, P. Moreno); Sanatorio Americano, Montevideo (M. Arleo, T. Possi, N. Reyes, M.N. Bentancor, V. Noya); Instituto Oswaldo Cruz–Fiocruz, Rio de Janeiro, Brazil (G. Bello, I. Arantes); Instituto de Investigaciones Biológicas Clemente Estable, Montevideo (P. Smircich); Ministerio de Salud Pública, Montevideo (J. Medina, H. Albornoz)

DOI: <https://doi.org/10.3201/eid2711.211198>

We developed a genomic surveillance program for real-time monitoring of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern (VOCs) in Uruguay. We report on a PCR method for SARS-CoV-2 VOCs, the surveillance workflow, and multiple independent introductions and community transmission of the SARS-CoV-2 P.1 VOC in Uruguay.

By late 2020, because of natural viral evolution, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genetic variants emerged, some of which show increased transmissibility and cause more severe coronavirus disease (COVID-19) (1). In

¹These authors contributed equally to this article.

RESEARCH LETTERS

addition, these variants show reduced neutralization by antibodies generated during previous infection or vaccination, which can reduce effectiveness of treatments, vaccines, or diagnostic tests (1). By July 2021, a total of 4 variants of concern (VOCs) had been identified: B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.617 (Delta) (2). Nonetheless, a robust surveillance workflow for early VOC identification is key to accelerating the pandemic response.

Brazil demonstrated a sharp increase in SARS-CoV-2 cases, hospitalizations, and deaths after the emergence of the P.1 VOC in Amazonas State in November 2020 (3; F. Naveca et al., unpub. data, <https://doi.org/10.21203/rs.3.rs-275494/v1>). P.1 displays higher transmissibility than previous local SARS-CoV-2 lineages and rapidly became the predominant strain in most states of Brazil during February–March 2021 (3,4; F. Naveca et al., unpub. data, <https://doi.org/10.21203/rs.3.rs-275494/v1>). P.1 also has spread worldwide; by July 2021, P.1 had been detected in ≥ 41 countries (5), where it might replicate the epidemic trajectory observed in Brazil. Uruguay, which shares 600 miles of dry border with Brazil, has experienced an exponential increase in COVID-19 cases since February 2021; by June 2021, Uruguay was among countries with the highest number of daily cases and deaths per million persons (6). Despite closing the Brazil–Uruguay border to tourism on March 13, 2020, evidence suggests a high viral flux between the countries (7,8). Therefore, P.1 could be introduced into Uruguay and the country needs an organized strategy to monitor VOC emergence.

In response to concerns over VOCs, the Ministry of Public Health, the Pasteur Institut of Montevideo (Uruguay), University of the Republic, and Zurgén-Sanatorio Americano formed a multidisciplinary workgroup to develop a genomic surveillance program for real-time monitoring for VOC emergence in Uruguay. The workgroup aimed to provide expertise and resources for large-scale sequencing, genomic analysis, and an affordable and decentralized inhouse PCR to detect known VOCs, including B.1.1.7, B.1.351, and P.1. Within a few weeks, the working group developed a PCR VOC detection method and a national sample processing workflow (Appendix 1 Figure 1, <https://wwwnc.cdc.gov/EID/article/27/11/21-1198-App1.pdf>). In addition, we identified multiple independent introductions of P.1 and community transmission in Uruguay.

The workgroup processes $>3,000$ nasopharyngeal samples daily, and around 100–300 SARS-CoV-2-positive samples are sent weekly for PCR

VOC analysis and sequencing. During January 11–March 26, 2021, the working group collected and processed a total of 251 SARS-CoV-2-positive RNA samples from 15/19 departments in Uruguay (Appendix 1 Table 1). Cycle thresholds for initial diagnostic PCR were 9–34.7. Among patients with positive samples, 95 were male, 95 were female, and 61 were of unknown sex; ages ranged from 1–85 years. Results from PCR VOC assay showed that 67/251 (27%) samples corresponded to putative P.1/B.1.351 (Appendix 1 Figure 2).

To validate PCR VOC classification, we sequenced all VOC-positive samples, plus 31 additional samples, by applying the ARTIC Network protocol (J.R. Tyson et al., unpub. data, <https://doi.org/10.1101/2020.09.04.283077>) for the MinION sequencing platform (Oxford Nanopore Technologies, <https://nanoporetech.com>) (Appendix 1). For the final 74 high-quality consensus sequences that were assigned to SARS-CoV-2 lineages following Pango nomenclature (9), we achieved a 100% agreement between PCR VOC and genome sequencing results. Predictably, given Uruguay's proximity to Brazil, samples classified as P.1/B.1.351 by PCR VOC were assigned to lineage P.1 after genome sequencing. The PCR we developed is a feasible, precise, and scalable method for real-time surveillance of known VOCs and verified circulation of P.1 lineage in 15/19 departments in Uruguay (Appendix 1 Figures 2, 3).

To estimate geographic sources and the number of independent P.1 introductions into Uruguay, we combined P.1 sequences from Uruguay and 691 P.1 sequences from South America available from EpiCoV in GISAID (<https://www.epicov.org/epi3>; Appendix 1 Tables 2, 3; Appendix 2, <https://wwwnc.cdc.gov/EID/article/27/11/21-1198-App2.xlsx>). The maximum-likelihood phylogeographic analysis identified ≥ 12 independent P.1 introductions into Uruguay from Brazil and ≥ 6 local transmission clusters of 3–24 sequences (Figure; Appendix 1 Figure 4). We used Bayesian analysis to estimate the median time of most recent common ancestor of P.1 clades in Uruguay to be mid-February to early March 2021 (Appendix 1 Table 4, Figure 5), which coincides with increasing mobility and the beginning of the exponential surge in COVID-19 cases in the country (Appendix 1 Figure 6).

The rapid emergence of the SARS-CoV-2 lineage P.1 in South America justifies the need for increased screening for this highly transmissible virus. We elaborated a comprehensive genomic surveillance program and provide a clear example of how multidisciplinary



Figure. Time-scaled maximum likelihood Bayesian phylogeographic maximum clade credibility tree of 59 severe acute respiratory syndrome coronavirus 2 lineage P.1 whole-genome sequences from Uruguay and 691 reference sequences from South America. The tree was rooted with the EPI_ISL_833137 sequence from GISAID (<https://www.gisaid.org>), collected December 4, 2020. Branches are colored according to the most probable location state of their descendant nodes as indicated at the legend. Sequences from Uruguay are shown with dots at the end of the branch. Red shading indicates clades from Uruguay and their distribution along the P.1 tree demonstrates ≥ 12 independent introductions and locally transmitted clusters of 3–24 sequences. The tree suggests Brazil has been the source of P.1 dissemination to Uruguay and other countries in South America.

RESEARCH LETTERS

teams are key mechanisms for helping public health authorities manage the COVID-19 crisis. Our findings revealed that the P.1 VOC was introduced into Uruguay multiple times over a period of increasing mobility in binational cities along the Brazil–Uruguay border and in Uruguay between mid-February and early March 2021. The introduction of the highly transmissible P.1 VOC coupled with the increasing human mobility probably contributed to the rapid local spread of this variant and the worsening COVID-19 epidemic in Uruguay during January–July 2021.

This article was preprinted at <https://www.medrxiv.org/content/10.1101/2021.05.20.21256969v1>.

Acknowledgments

We thank the groups that contributed SARS-CoV-2 genomes to the EpiCoV GISAID initiative (Appendix 2). We thank Marcelo Fiori and María Inés Fariello for sharing mobility data from Uruguay and Christian Brandt from Institute for Infectious Diseases and Infection Control (Jena University Hospital), for adapting their poreCov Nextflow pipeline to our requirements and for clarifying doubts about its implementation. We thank Nicolas Nin and Javier Hurtado from Hospital Español for their work as medical doctors caring for patients included in this study.

This work was supported by FOCEM-Fondo para la Convergencia Estructural del Mercosur (COF03/11).

About the Author

Ms. Rego is a research biologist and trained in bioinformatics at Institut Pasteur de Montevideo, Uruguay. Her research interests include evolutionary biology, transcriptomics, host-pathogen interactions, and the epidemiology and phylogeny of coronaviruses.

References

1. Abdool Karim SS, de Oliveira T. New SARS-CoV-2 variants—clinical, public health, and vaccine implications. *N Engl J Med*. 2021;384:1866–8. <https://doi.org/10.1056/NEJMc2100362>
2. Centers for Disease Control and Prevention. SARS-CoV-2 variant classifications and definitions [cited 2021 Apr 10]. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html>
3. Faria NR, Mellan TA, Whittaker C, Claro IM, Candido DDS, Mishra S, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021;372:815–21. <https://doi.org/10.1126/science.abh2644>
4. Fiocruz Genomic Network. Frequency of the main SARS-CoV-2 lineages per month sampling [in Portuguese] [cited 2021 Apr 30]. <http://www.genomahcov.fiocruz.br/frequencia-das-principais-linhagens-do-sars-cov-2-por-mes-de-amostragem>
5. Outbreak.info. P.1 lineage report [cited 2021 Apr 30]. https://outbreak.info/situation-reports?pango=P.1&selected=BRA&loc=BRA&loc=USA&loc=USA_US-CA
6. Our World in Data. Daily new and confirmed COVID-19 cases per million people [cited 2021 Jun 30]. <https://ourworldindata.org/explorers/coronavirus-data-explorer>
7. Elizondo V, Harkins GW, Mabvakure B, Smidt S, Zappile P, Marier C, et al. SARS-CoV-2 genomic characterization and clinical manifestation of the COVID-19 outbreak in Uruguay. *Emerg Microbes Infect*. 2021;10:51–65. <https://doi.org/10.1080/22221751.2020.1863747>
8. Mir D, Rego N, Resende PC, Tort F, Fernández-Calero T, Noya V, et al. Recurrent dissemination of SARS-CoV-2 through the Uruguayan–Brazilian border. *Front Microbiol*. 2021;12:653986. <https://doi.org/10.3389/fmicb.2021.653986>
9. Rambaut A, Holmes EC, O’Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;5:1403–7. <https://doi.org/10.1038/s41564-020-0770-5>

Address for correspondence: Pilar Moreno, Laboratorio de Virología Molecular, Facultad de Ciencias, Universidad de la República, Mataojo 2055, Montevideo, Uruguay; email: pmoreno@pasteur.edu.uy

Highly Pathogenic Avian Influenza A(H5N1) Virus in Wild Red Foxes, the Netherlands, 2021








Jolianne M. Rijks, Hanna Hesselink, Pim Lollinga, Renee Wesselman, Pier Prins, Eefke Weesendorp, Marc Engelsma, Rene Heutink, Frank Harders, Marja Kik, Harry Rozendaal, Hans van den Kerkhof, Nancy Beerens

Author affiliations: Dutch Wildlife Health Centre, Utrecht University, Utrecht, the Netherlands (J.M. Rijks, H. Hesselink, M. Kik); Stichting Faunavisie Wildlife Care, Westernieland, the Netherlands (P. Lollinga); Stichting Faunavisie Wildlife Care, Blijham, the Netherlands (R. Wesselman); Dierenkliniek Winsum, Winsum, the Netherlands (P. Prins); Wageningen Bioveterinary Research, Lelystad, the Netherlands (E. Weesendorp, M. Engelsma, R. Heutink, F. Harders, N. Beerens); Dutch Food and Consumer Products Safety Authority, Utrecht (H. Rozendaal); Coordination Centre for Communicable Disease Control, National Institute for Public Health and the Environment, Bilthoven, the Netherlands (H. van den Kerkhof)

DOI: <https://doi.org/10.3201/eid2711.211281>

Article

Emergence and Spread of a B.1.1.28-Derived P.6 Lineage with Q675H and Q677H Spike Mutations in Uruguay

Natalia Rego ^{1,†} , Cecilia Salazar ^{2,3,†}, Mercedes Paz ^{3,†}, Alicia Costáble ^{3,4,5,6,†}, Alvaro Fajardo ^{4,5}, Ignacio Ferrés ^{2,3} , Paula Perbolianachis ^{4,5}, Tamara Fernández-Calero ^{1,7}, Veronica Noya ⁸, Matias R. Machado ⁹ , Mariana Brandes ¹, Rodrigo Arce ^{4,5,8}, Mailen Arleo ⁸, Tania Possi ⁸, Natalia Reyes ⁸, María Noel Bentancor ⁸, Andrés Lizasoain ¹⁰ , Viviana Bortagaray ¹⁰, Ana Moller ¹⁰, Odhille Chappos ¹¹, Nicolas Nin ¹², Javier Hurtado ¹², Melissa Duquía ¹¹, Maria Belén González ¹¹, Luciana Griffero ¹¹, Mauricio Méndez ¹¹, Maria Pía Techera ¹¹, Juan Zanetti ¹¹, Emiliano Pereira ¹¹ , Bernardina Rivera ¹³, Matias Maidana ¹³, Martina Alonso ¹³, Pablo Smircich ^{14,15}, Ighor Arantes ¹⁶, Daiana Mir ¹⁷, Cecilia Alonso ¹¹, Julio Medina ^{18,19}, Henry Albornoz ^{18,19}, Rodney Colina ¹⁰, Gonzalo Bello ¹⁶ , Pilar Moreno ^{3,4,5,*} , Gonzalo Moratorio ^{3,4,5,*}, Gregorio Iraola ^{2,3,20,21,*} and Lucía Spangenberg ^{1,22,*}



Citation: Rego, N.; Salazar, C.; Paz, M.; Costáble, A.; Fajardo, A.; Ferrés, I.; Perbolianachis, P.; Fernández-Calero, T.; Noya, V.; Machado, M.R.; 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*, 1801. <https://doi.org/10.3390/v13091801>

Academic Editors: Eurico Arruda, Helena Lage Ferreira, José Luiz Proença Módena and William Marciel de Souza

Received: 28 July 2021

Accepted: 31 August 2021

Published: 10 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

- 1 Bioinformatics Unit, Institut Pasteur de Montevideo, Montevideo 11400, Uruguay; natalia@pasteur.edu.uy (N.R.); tamfer@pasteur.edu.uy (T.F.-C.); mbrandes@pasteur.edu.uy (M.B.)
- 2 Laboratorio de Genómica Microbiana, Institut Pasteur Montevideo, Montevideo 11400, Uruguay; csalazar@pasteur.edu.uy (C.S.); iferres@pasteur.edu.uy (I.F.)
- 3 Centro de Innovación en Vigilancia Epidemiológica, Institut Pasteur Montevideo, Montevideo 11400, Uruguay; mpaz@pasteur.edu.uy (M.P.); acostabile@fcien.edu.uy (A.C.)
- 4 Laboratorio de Virología Molecular, Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay; afajardo@fcien.edu.uy (A.F.); paulaperbolianachis@fcien.edu.uy (P.P.); arcerama@gmail.com (R.A.)
- 5 Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo, Montevideo 11400, Uruguay
- 6 Sección Bioquímica, Facultad de Ciencias, Universidad de la República, Montevideo 11400, Uruguay
- 7 Department of Exact and Natural Sciences, Universidad Católica del Uruguay, Montevideo 11600, Uruguay
- 8 Laboratorio de Biología Molecular, Sanatorio Americano, Montevideo 11600, Uruguay; veronicanoyaro@gmail.com (V.N.); mailenarleo@gmail.com (M.A.); tpossi@pasteur.edu.uy (T.P.); reyesntl@gmail.com (N.R.); manoben2011@gmail.com (M.N.B.)
- 9 Protein Engineering, Institut Pasteur de Montevideo, Montevideo 11400, Uruguay; mmachado@pasteur.edu.uy
- 10 Laboratorio de Virología Molecular, Departamento de Ciencias Biológicas, CENUR Litoral Norte, Universidad de la República, Salto 50000, Uruguay; andres.lizasoain.cuelho@gmail.com (A.L.); viviborta@gmail.com (V.B.); anaclaramoller2@gmail.com (A.M.); rodneycolina1@gmail.com (R.C.)
- 11 Centro Universitario Regional Este, Universidad de la República, Rocha 27000, Uruguay; izisodhille@gmail.com (O.C.); meliduquia@gmail.com (M.D.); mabel2406@gmail.com (M.B.G.); luciana.griffero@cure.edu.uy (L.G.); mauriciomendezpeyre@gmail.com (M.M.); mariapiatechera@gmail.com (M.P.T.); juanzanetti04@hotmail.com (J.Z.); pereiramemo@gmail.com (E.P.); ceci.babilonia@gmail.com (C.A.)
- 12 Unidad de Cuidados Intensivos, Hospital Español “Juan J. Crottogini”, Montevideo 11800, Uruguay; niconin@hotmail.com (N.N.); javierhurtado2005@gmail.com (J.H.)
- 13 Laboratorio de Diagnóstico Molecular, Institut Pasteur de Montevideo, Montevideo 11400, Uruguay; bernardinariverasoto@gmail.com (B.R.); maidana@pasteur.edu.uy (M.M.); martina.alonso60@gmail.com (M.A.)
- 14 Bioinformatics Laboratory, Department of Genomics, Instituto de Investigaciones Biológicas Clemente Estable, MEC, Montevideo 11600, Uruguay; psmircich@gmail.com
- 15 Laboratory of Molecular Interactions, Facultad de Ciencias, UdelAR, Montevideo 11400, Uruguay
- 16 Laboratorio de AIDS e Imunología Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro 21040-900, Brazil; ighorarantes@gmail.com (I.A.); gbellobr@gmail.com (G.B.)
- 17 Unidad de Genómica y Bioinformática, Departamento de Ciencias Biológicas, Centro Universitario Regional Litoral Norte, Universidad de la República, Salto 50000, Uruguay; daianamir@gmail.com
- 18 Cátedra de Enfermedades Infecciosas, Facultad de Medicina, Universidad de la República, Montevideo 11300, Uruguay; jcmolina1@gmail.com (J.M.); henry_albornoz@hotmail.com (H.A.)
- 19 Dirección General de Salud, Ministerio de Salud Pública, Montevideo 11200, Uruguay
- 20 Center for Integrative Biology, Universidad Mayor, Santiago de Chile 8580745, Chile
- 21 Host-Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, Cambridge CB10 1SA, UK

²² Department of Informatics and Computer Science, Universidad Católica del Uruguay, Montevideo 11600, Uruguay

* Correspondence: pmoreno@pasteur.edu.uy (P.M.); moratorio@pasteur.edu.uy (G.M.); giraola@pasteur.edu.uy (G.I.); lucia@pasteur.edu.uy (L.S.)

† These authors equally contributed to the study.

Abstract: Uruguay controlled the viral dissemination during the first nine months of the SARS-CoV-2 pandemic. Unfortunately, towards the end of 2020, the number of daily new cases exponentially increased. Herein, we analyzed the country-wide genetic diversity of SARS-CoV-2 between November 2020 and April 2021. We identified that the most prevalent viral variant during the first epidemic wave in Uruguay (December 2020–February 2021) was a B.1.1.28 sublineage carrying Spike mutations Q675H + Q677H, now designated as P.6, followed by lineages P.2 and P.7. P.6 probably arose around November 2020, in Montevideo, Uruguay’s capital department, and rapidly spread to other departments, with evidence of further local transmission clusters; it also spread sporadically to the USA and Spain. The more efficient dissemination of lineage P.6 with respect to P.2 and P.7 and the presence of mutations (Q675H and Q677H) in the proximity of the key cleavage site at the S1/S2 boundary suggest that P.6 may be more transmissible than other lineages co-circulating in Uruguay. Although P.6 was replaced by the variant of concern (VOC) P.1 as the predominant lineage in Uruguay since April 2021, the monitoring of the concurrent emergence of Q675H + Q677H in VOCs should be of worldwide interest.

Keywords: SARS-CoV-2; B.1.1.28; Spike; Q675H; Q677H; Uruguay; phylogeography; phylogenetics

1. Introduction

By the end of 2020 and the beginning of 2021, several studies reported the emergence of novel SARS-CoV-2 variants of interest (VOIs) and concern (VOCs) with different missense mutations and deletions in the Spike (S) protein that impact viral transmissibility and escape from previous host’s immune responses, among other features [1,2]. In Brazil, the SARS-CoV-2 lineages B.1.1.28 and B.1.1.33 dominated the first epidemic wave [3,4], but were replaced by VOC P.1 (WHO name: Gamma) and P.2 (former VOI Zeta), both descendants of lineage B.1.1.28, by the end of 2020 and beginning of 2021 [5]. So far, five B.1.1.28 descendant sublineages carrying mutations of concern have emerged. The VOC P.1, which harbors the mutations of concern S:K417T/E484K/N501Y among its lineage defining mutations [6], originated in the Amazonas state in mid-November [7,8] and rapidly spread across Brazil and to over 50 countries globally [9]. The lineages P.2, P.4, and P.5, carrying the concerning amino acid changes S:E484K, S:L452R, and S:E484Q/N501T, respectively, were also initially detected in samples from Brazil [10–15]. The lineage P.3 (former VOI Theta) emerged in the Philippines, and it includes substitutions S:E484K/N501Y/P681H among the lineage-defining mutations; the first sample was collected on 8 January 2021, and later it further spread to the USA, Germany, and Malaysia, among other countries [16]. One additional B.1.1.28 descendant clade that emerged in southern Brazil carrying mutation N:P13L was recently defined as a new Pango P.7 lineage [17,18].

Uruguay was able to control the early viral dissemination during the first nine months of the SARS-CoV-2 pandemic by implementing a successful Test, Trace, and Isolation strategy (TETRIS). The low number of total cases, contained outbreaks, and few deaths were characteristic for this first period [19,20]. At the beginning, viral diversity was high, with cocirculation of strains A.2, A.5, B.1, B.1.195, and B.31, introduced mostly through Montevideo, Uruguay’s capital city and connection hub through its international airport and harbor [21]. Later, multiple introductions of SARS-CoV-2 lineages B.1.1.28 and B.1.1.33 of Brazilian origin were detected in Uruguay, mainly along the 1068 km long Uruguayan–Brazilian dry border, and these lineages became predominant between May and July 2020 [22]. Towards the end of 2020, the number of active cases exponentially increased, from an average of 60 cases per day during October and November to more than 400 during

December [19], concomitant with the loss of the TETRIS safety zone [23,24]. SARS-CoV-2 positive daily new cases decreased around mid-February after more stringent mobility measures were taken by the government [20,25]; but the total number of cases stayed outside the TETRIS zone and a second exponential growth period started in March 2021, coinciding with the introduction and dissemination of VOC P.1 [26]. Summer-related social gatherings and relaxed social distancing are some of the proposed reasons to explain the first epidemic wave in Uruguay, but there is currently a gap in knowledge concerning the potential influence of virological factors.

To understand the SARS-CoV-2 diversity associated with the first COVID-19 epidemic wave in Uruguay, we conducted a retrospective epidemiological and genomic analysis of SARS-CoV-2 complete genomes from COVID-19 patients diagnosed between November 2020 and April 2021. Our study revealed that a novel B.1.1.28 clade harboring two non-synonymous changes in the Spike protein: Q675H and Q677H, now designated as lineage P.6 [27], was the most prevalent SARS-CoV-2 variant by the end of 2020 and beginning of 2021. Lineages P.2 and P.7 were also detected at lower prevalence during the first epidemic wave in Uruguay. The Q675H and Q677H mutations are in the proximity of the polybasic cleavage site at the S1/S2 boundary, a region of biological relevance for virus replication [28], and also arose independently in many other SARS-CoV-2 VOIs circulating worldwide. These findings suggest that local emergence and spread of a more transmissible P.6 variant might have had a non-negligible role in the first epidemic wave of COVID-19 in Uruguay.

2. Materials and Methods

2.1. Ethics Statement

This work was done by the Inter-Institutional Working Group (IiWG) for SARS-CoV-2 genomic surveillance in Uruguay, which involves a diagnostic network, expertise and resources to handle large-scale sequencing, computational scientists for genomic analysis, and an affordable and decentralized “in-house” qPCR test designed to detect known VOCs [26]. Residual deidentified RNA samples from SARS-CoV-2 positive patients were remitted to the Institut Pasteur de Montevideo (IPMon). IPMon was validated by the Ministry of Health of Uruguay as an approved center providing diagnostic testing for COVID-19. All samples were deidentified before receipt by the study investigators. All relevant ethical guidelines were appropriately followed. Additionally, the project was approved by the Ethics Committee of the Sanatorio Americano SASA (Uruguay) on the 29 April 2021. Ethical approval was given, and signed informed consent was obtained from the participants.

2.2. SARS-CoV-2 Samples

In total, 260 SARS-CoV-2 RNA samples (Table S1) were recovered from nasopharyngeal–throat combined swabs collected from clinically ill or asymptomatic individuals that resided in different Uruguayan departments and were diagnosed from November 2020 to April 2021 in Uruguay. As the IiWG began working in March 2021, the availability of earlier samples was conditional to what was kept at any laboratory of the IiWG diagnostic network. Positive RNA samples were reverse transcribed using SuperScript™ II Reverse Transcriptase (Thermo Fisher Scientific Inc., Waltham, MA, USA) or the LunaScript® RT SuperMix Kit (New England Biolabs, Ipswich, MA, USA). A negative control was included at this point and carried throughout the protocol.

2.3. Genome Sequencing

Sequencing libraries were prepared according to the classic ARTIC protocol described by Quick J. [29,30], the 2000 bp long amplicon version described by Resende P.C. [31], or the sequencing protocol using a 1200 bp amplicon “midnight” primer set, with the Nanopore Rapid kit as described by Freed N. and Silander O. [32,33] (Table S1). The final library was eluted in EB buffer (ONT) and quantified using a fluorometric assay.

Recommended amounts of library were loaded into a FLO-MIN106D R9.4.1 flowcell and sequenced on the MinION Mk1C or GridION X5 sequencing platforms (ONT). Basecalling and demultiplexing were performed with Guppy 4.3.2 or higher [34] using the high or super accuracy mode. Consensus genomes were generated using the poreCov pipeline 0.7.0 or higher [35–44], and Nanopolish was used for consensus generation. Complete sequences with up to 15% of Ns were kept for further analysis. All genomes obtained in this study were uploaded to the EpiCoV database in the GISAID initiative under the accession numbers specified in Table S1.

2.4. SARS-CoV-2 Lineage Assignment

SARS-CoV-2 full-length consensus sequences were manually curated in specific genome positions, such as clade-defining mutations. Genotyping was performed according to Rambaut et al. [45] using the Pangolin application [46,47], and later confirmed using maximum likelihood (ML) analysis.

2.5. Phylogenetic and Phylogeographic Analysis of B.1.1.28

Uruguayan B.1.1.28 sequences ($n = 212$; samples assigned as lineage B.1.1.28, P.6, or P.7 in Table S1) were next analyzed in the context of additional B.1.1.28 sequences from Uruguay and Brazil, downloaded from the EpiCoV database of the GISAID initiative [48] (Table S3 showing EpiCoV/GISAID acknowledgments). Downloaded B.1.1.28 sequences from Uruguay ($n = 143$) were complete, with full collection date information and sampled before 31 May 2021. Sequences from Brazil ($n = 1428$) were complete and high quality with full collection date information, and were sampled before 31 May 2021 (Table S2). Additionally, we downloaded four B.1.1.28 from the USA ($n = 2$), Spain ($n = 1$), and Belgium ($n = 1$) that also harbored both S:Q675H and S:Q677H mutations (Tables S2 and S3). Alignment was performed with MAFFT v7.471 [49]. Maximum likelihood phylogenetic analysis of the 1787 B.1.1.28 sequences was performed with IQ-TREE version 1.6.12 under the model GTR + F + R3 of nucleotide substitution selected by the built-in ModelFinder option [50]. Branch support was assessed by the approximate likelihood-ratio test based on a Shimodaira–Hasegawa-like procedure (SH-aLRT) with 1000 replicates [51]. The tree root was established with the sequence EPI_ISL_416036 with the earliest collection date of 5 March 2020. This tree was time-scaled using TreeTime 0.8.3.1 [52], applying a fixed clock rate of 8×10^{-4} substitutions/site/year [53,54], and keeping polytomies. The time-scaled tree was then employed for the ancestral character state reconstruction (ACR) of epidemic locations with PastML v1.9.15 [55], using the marginal posterior probabilities approximation (MPPA) method with an F81-like model. Brazilian sequences were grouped according to the region: South, Southeast, Central West, North, and Northeast. A time-scaled Bayesian phylogeographic analysis was next performed to infer the geographical source and dissemination pattern of the Uruguayan B.1.1.28 + Q675H + Q677H (now P.6) samples, and to estimate the time of their most recent common ancestors (T_{MRCA}). Phylogenetic trees were estimated in BEAST v1.10 [56] using the GTR + F + I nucleotide substitution model, the non-parametric Bayesian skyline model as the coalescent tree prior [57], a strict molecular clock model with a uniform substitution rate prior ($8\text{--}10 \times 10^{-4}$ substitutions/site/year), and a reversible discrete phylogeographic model (using Uruguayan departments as epidemic locations) [58] with a continuous-time Markov chain (CTMC) rate reference prior [59]. MCMC chains were run for 100 million generations, and convergence (effective sample size > 200) in parameter estimates was assessed using Tracer v1.7 [60]. The maximum clade credibility (MCC) tree was summarized with TreeAnnotator v1.10 [61] and visualized using FigTree v1.4.4 [62]. Additional visualizations were implemented in the R environment with treeio 1.16.2 and ggtree 3.0.4 Bioconductor packages [63].

2.6. Phylogenetic and Phylogeographic Analysis of P.2

Uruguayan P.2 sequences ($n = 79$) were analyzed in the context of 1272 additional Brazilian P.2 sequences downloaded from EpiCoV/GISAID; sequences were complete and high quality with full collection date information, and were sampled before 31 May 2021 (Tables S2 and S3). Five Brazilian sequences with a collection date before September 2020 were discarded (the collection date might have been incorrect). Alignment and ML phylogenetic analysis were performed as above. The tree root was established with the sequence EPI_ISL_2344425 with the earliest collection date of 1 September 2020. The tree was time-scaled using TreeTime [52] and the ACR of epidemic locations inferred with PastML [55], as before. Brazilian locations were grouped according to the five Brazilian regions mentioned above. The resulting tree with inferred locations was visualized using FigTree [62].

2.7. Lineage Prevalence of Available Uruguayan Samples

To assess the prevalence of B.1.1.28 + Q675H + Q677H clade in the context of the different SARS-CoV-2 lineages circulating in Uruguay, we used the lineage assignments obtained for the 260 samples presented in this study, 335 additional samples already available at EpiCoV/GISAID, and 342 records available at the IiWG domain. As only one B.1.1.28 (not carrying Q675H + Q677H mutations) was identified in May 2021, we kept our analysis from November 2020 to April 2021. The dataset comprised a total of 937 SARS-CoV-2 genotypes, and prevalence (relative frequency) of each lineage was calculated monthly, from November to April. The “Others” category included B.1, B.1.1, B.1.1.1, B.1.1.34, B.1.177, B.1.177.12, and B.1.238.

In the case of geographic distribution of P.6 and P.2 variants (for the Uruguayan maps), the number of cases per department were obtained using samples from this study and also available at the IiWG domain, from November 2020 to April 2021 ($n = 174$ for P.6 and $n = 70$ for P.2).

2.8. Determination of Prevalence of Q675H + Q677H

To assess the prevalence of co-occurring S:Q675H and S:Q677H in worldwide SARS-CoV-2 genomes, we downloaded from EpiCoV/GISAID (accessed on 7 July 2021) 129 complete genomes, with high quality and full collection date information. We removed redundant samples with the ones included in this study, obtaining a final dataset of 259 sequences (Tables S2 and S3).

2.9. Structural Representation of the Spike Protein

The molecular model of SARS-CoV-2 Spike glycoprotein was taken from the D. E. Shaw Research database (DESRES-ANTON-11021566) [64]. The visual rendering was done with VMD 1.9.3 [65].

3. Results

At the IiWG, we sequenced 663 SARS-CoV-2 positive samples detected in Uruguay between November 2020 and April 2021 (Figure S1A and Table S1), which were classified in the following lineages: 333 (50%) P.1, 180 (27%) B.1.1.28, 70 (11%) P.2, 32 (5%) P.7, and 39 (6%) other B.1-derived lineages. The mutational profile of B.1.1.28 sequences identified 174 (26%) genomes carrying amino acid changes S:Q675H and S:Q677H that compose a new Pango lineage designated as P.6 [27,45]. Lineage P.6 was widely spread throughout the country, being detected in 12 out of 19 Uruguayan departments from 2 December 2020 to 26 April 2021; but was particularly prevalent in Montevideo, the capital department of Uruguay, and the surrounding metropolitan area (Canelones) (Figure 1A,B, right panel). Lineages P.2 and P.7, by contrast, were more frequently detected outside the metropolitan region (Figure 1B, right panel). Analysis of these 663 SARS-CoV-2 Uruguayan sequences determined by the IiWG, plus additional Uruguayan sequences obtained at the EpiCoV database in GISAID [48], revealed a changing molecular epidemiological pattern over time (Figure 1B, to the left). Lineage P.7 was the most prevalent variant (55%) detected

from late November to late December 2020, but was rapidly replaced by lineage P.6, the relative frequency of which increased from 26% in December 2020 to 76% in February 2021. Lineage P.2 was first detected in Uruguay in December 2020 and reached a relatively high prevalence during the summer season, with a maximum frequency of 36% in January 2021. As previously described [26], the VOC P.1 was first detected in Uruguay in February 2021 and then quickly outcompeted the other SARS-CoV-2 variants circulating in the country, becoming the predominant lineage in April 2021.

To identify the probable geographic source of B.1.1.28, P.6, and P.7 lineages detected in Uruguay, all Uruguayan sequences here obtained ($n = 212$, Table S1) were combined with complete genome sequences of those lineages available at the EpiCoV database in GISAID sampled in Uruguay ($n = 143$) and Brazil ($n = 1428$), and with all B.1.1.28 sequences sampled worldwide that carried mutations Q675H and Q677H (USA = 2, Spain = 1 and Belgium = 1; Table S2). The ML phylogeographic analysis supported at least 20 independent introductions of this lineage B.1.1.28 ($n = 14$) and P.7 ($n = 6$) from Brazil into Uruguay that mostly resulted in singletons, dyads, or small clades ($n < 5$ sequences), with no evidence of extensive dissemination in Uruguay (Figure 1C and Figure S2). One B.1.1.28 introduction, however, was successfully established and originated the lineage P.6 (SH-aLRT = 98) that comprised all Uruguayan sequences ($n = 314$), as well as three sequences collected in the USA and Spain carrying mutations S:Q675H + Q677H. According to our analysis, the lineage P.6 was most likely introduced from the southeastern Brazilian region (ACR-location marginal probability (LMP) = 0.99) and was disseminated from Uruguay to the USA (two independent times, ACR-LMP ≥ 0.99) and Spain (ACR-LMP = 0.99) (Figures 1C and S3). Of note, the B.1.1.28 + Q675H + Q677H collected in Belgium did not belong to clade P.6, indicating an independent and recurrent appearance of both mutations (see below; Figure 1C). We also identified one successful introduction of lineage P.7, probably from Southern Brazil (ACR-LMP = 0.99), that originated the highly supported (SH-aLRT = 100) Uruguayan clade designated as UY_{P7} composed by 27 Uruguayan sequences plus eight sequences from Southern Brazil (Figure 1C and Figure S3). We also performed a ML phylogeographic analysis of lineage P.2 sequences sampled in Uruguay ($n = 79$) and Brazil ($n = 1267$), and identified three successful introductions, most likely from Southern Brazil (ACR-LMP > 0.85), that originated the highly supported (SH-aLRT > 0.92) Uruguayan clades UY-I_{P2} ($n = 35$), UY-II_{P2} ($n = 30$), and UY-III_{P2} ($n = 12$) (Figure S4). It was noteworthy that most sequences branching within Uruguayan clades UY_{P7} (89%), UY-I_{P2} (100%) and UY-II_{P2} (83%) were sampled outside the metropolitan region.



Figure 1. Geographic distribution, prevalence, and maximum likelihood analysis of P.6. **(A)** Map of Uruguay showing the number of sequences classified as P.6 (B.1.1.28 + Q675H + Q677H) in every department ($n = 165$ for samples with known geographical source in Uruguay). Department labels are as follows: Sa (Salto), Ta (Tacuarembó), Ro (Rocha), Ca (Canelones), Mo (Montevideo). The metropolitan area corresponds to Mo and Ca. The border with Brazil is shown. RS stands for Rio Grande do Sul, the southernmost Brazilian state. **(B)** To the left, Pango lineage proportions of all available Uruguayan

samples calculated monthly, from November 2020 to April 2021. The “Others” category includes B.1, B.1.1, B.1.1.1, B.1.1.34, B.1.177, B.1.177.12, and B.1.238. To the right, pie charts show Pango lineage prevalence in the metropolitan area (top) versus the rest of the country (bottom). (C) Maximum likelihood phylogeographic analysis of lineage B.1.1.28 samples ($n = 1787$) from Uruguay ($n = 355$); Brazil ($n = 1428$); and the USA, Spain, and Belgium ($n = 4$) inferred by an ancestral character reconstruction method implemented in PastML. Tips and branches are colored according to sampling location and the most probable location state of their descendent nodes, respectively, as indicated in the legend. Shaded boxes highlight the major B.1.1.28 clades in Uruguay. P.6 is the assigned Pango name for the clade B.1.1.28 + Q675H + Q677H discussed here, while UY-P.7 (corresponding to UY_{P7} in the main text) is a B.1.1.28 clade carrying mutation N:P13L widely distributed in southern Brazil (recently assigned as P.7). Brazilian P.7 is shown with a gray shadow. Asterisks (*) indicate the sequences dispersed from Uruguay to the USA and Spain. The hash indicates the B.1.1.28 + Q675H + Q677H sample from Belgium. The time-scaled tree was rooted with the earliest sequence (collection date 5 March 2020). Branch lengths are drawn to scale indicating nucleotide substitutions per site per year.

To better understand the origin and spread of lineage P.6, we performed a Bayesian phylogeographic analysis of all P.6 Uruguayan sequences produced in this work that had a geographic source available ($n = 165$), and six Brazilian basal sequences. The spatiotemporal reconstruction suggested that an ancestral B.1.1.28 virus was probably introduced into Uruguay from Brazil around October 2020, and after a short period of local evolution, the lineage P.6 ancestor arose in Montevideo (posterior state probability (PSP) = 0.94) around 9 November 2020 (95% HPD: 20 October–26 November) (Figure 2A). Lineage P.6 was next disseminated from Montevideo to the surrounding metropolitan area and also to more distant Uruguayan departments. The T_{MRCA} of major transmission clusters outside the metropolitan region was traced to 23 December 2020 (95% HPD: 12 December 2020 to 30 December 2020) in Rocha and 30 December 2020 (95% HPD: 19 December 2020 to 9 January 2021) in Salto (Figure 2A,B). The introduction and dispersion of lineage P.6 in each department coincided with the increase in new COVID-19 cases reported daily (Figure 2A). Given that VOI P.2 was also cocirculating, we reinforce that there was a low spatiotemporal overlap of P.6 and P.2 in Montevideo and Salto, though this was not the case for Rocha (Figure S1C). The lineage P.6 was characterized by eight lineage-defining genetic changes in addition to S:Q675H and S:Q677H, including a total of five nonsynonymous mutations (Figure 2B). Of note, eight out of 10 lineage P.6-defining mutations (including S:Q675H) were also identified in a basal B.1.1.28 sequence sampled in Rio de Janeiro in January 2021, and were thus probably present in the ancestral virus that arrived from Brazil; while the remaining two mutations (ORF1ab: C8980T and S:Q677H) were fixed during the early local transmission in Uruguay. We also identified one additional mutation (ORF3a:M260I) that was fixed at a later step during evolution of P.6 in Uruguay (Figure 2B).

Structural analysis of the SARS-CoV-2 Spike glycoprotein showed that residues Q675H and Q677H were within the subdomain SD2 of each protomer constituting the homotrimer (Figure 2C). Particularly, they were located at the beginning of a very flexible loop (residues 675–690) [66], which embraced the solvent-accessible polybasic furin cleavage site [66]. These mutations were close to two experimentally observed *O*-glycosylation sites at T676 and T678 [67,68], and at the same domain of the D614G mutation [69] (Figure 2C, inset). Substitution S:Q677H has been reported as a recurrent mutation arising independently in many SARS-CoV-2 lineages, including several VOIs, circulating worldwide by the end of 2020 [70,71]. A search in the EpiCoV database (accessed on 7 July 2021) for high-quality SARS-CoV-2 genomes carrying both mutations S:Q675H and S:Q677H recovered 85 non-Uruguayan sequences. Overall, the pair S:Q675H + Q677H appeared to be distributed in 12 different countries (in decreasing frequency order: Uruguay, England, USA, Belgium, India, Australia, Switzerland, Spain, Netherlands, Japan, Germany, and France) and in 13 different Pango lineages (in decreasing frequency order: P.6, B.1.36, B.1.2, C.36, B.1.538, B.1.1.316, B.1.526 (VOI Iota), B.1.525 (VOI Eta), B.1.243, B.1.1.70, B.1.1.7 (VOC Alpha), B.1.1.63, and B.1) (Figure S5).

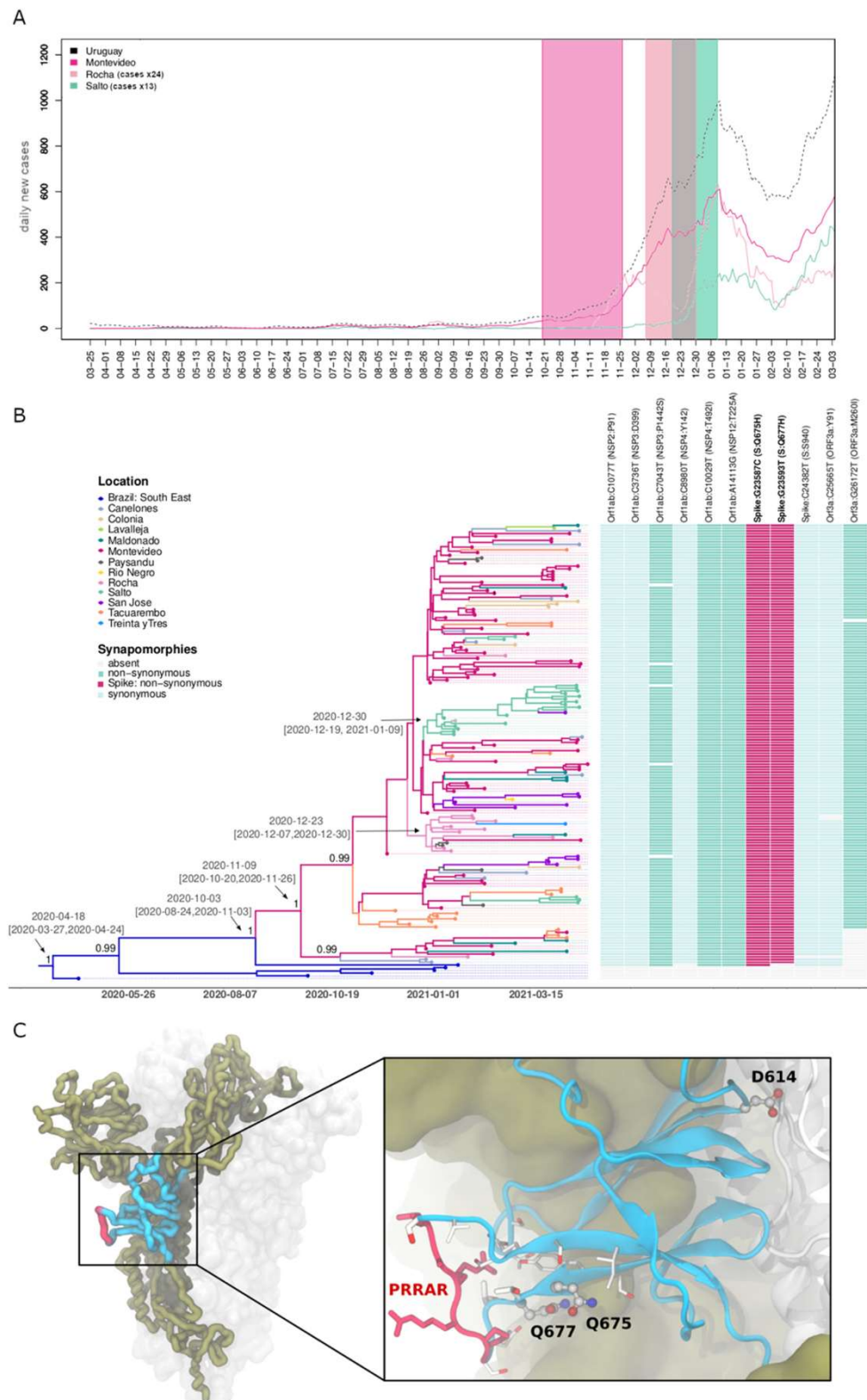


Figure 2. Bayesian phylogeographic analysis and description of synapomorphies of P.6. **(A)** Number of daily new cases from March 2020 to March 2021 in the country (black), in Montevideo (fuchsia), Rocha (pink), and Salto (green). Daily new cases for Rocha and Salto were multiplied by a factor proportional to the population of that department in comparison to Montevideo (times 24 and 13 for Rocha and Salto, respectively) for visualization purposes. Confidence intervals of TMRCA of Montevideo (fuchsia), Rocha (pink), and Salto (green) P.6 clades are shown as shaded areas. **(B)** Bayesian phylogeographic analysis of the P.6 clade in Uruguay, implemented in BEAST. Uruguayan sequences

generated by the iWiG with known geographic source ($n = 165$) were combined with six additional basal sequences from southeastern Brazil. Tips and branches of the time-scaled Bayesian tree are colored according to sampling location and the most probable location state of their descendent nodes, respectively, as indicated in the legend. Posterior probability support values and estimated TMRCAs are indicated at key nodes. Additionally, a heatmap represents the presence or absence of 10 synapomorphic sites and an additional change (ORF3a:M260I) shared by most sequences. The color scheme indicates the different mutations, as indicated in the legend. In each case, genomic position, nucleotide substitution, viral protein, and amino acid are shown. (C) One Spike protomer of the homotrimer is shown in thick ribbons, while the others are represented as transparent solvent accessible surfaces. The subdomain SD2 (residues 590–700) is indicated in blue, and the polybasic furin cleavage site (PRRAR, residues 681–685) is in red. The inset shows a zoom into the structural context of SD2, representing residues D614, Q675, and Q677 in balls and sticks, and nearby residues in sticks.

4. Discussion

Since March 2020, Uruguay had been successful at keeping the COVID-19 pandemic in check. Closed international borders and an aggressive contact-tracing system, among other government measures, were able to avoid virus transmission growing exponentially [19,20]. Brazil has been a COVID-19 hotspot in South America, and the 1068 km long Uruguayan–Brazilian dry border allowed the rapid local establishment of SARS-CoV-2 Brazilian lineages B.1.1.28 and B.1.1.33, initially associated with a few outbreaks that occurred in the departments bordering Brazil [22]. However, by the end of 2020, the pandemics worsened, with a clear increase of daily cases in December that extended until mid-February 2021. Summer-related social gatherings and relaxed social-distancing measures were some of the proposed reasons to explain the epidemic growth [20]. In this study, we described a new B.1.1.28 sublineage, designated P.6, that probably arose in Montevideo by November 2020 and spread throughout the country. Lineage P.6 comprised most (60%) Uruguayan virus genomes recovered between November 2020 and February 2021, and its spread coincided with national and local (as shown for Montevideo, Rocha, and Salto departments) increases in daily SARS-CoV-2 cases during the first epidemic wave.

Dispersion of lineage P.6 in Uruguay could have been fueled by changes in human behavior coinciding with the end of the austral spring and the ambiance of relaxed restrictions, as was demonstrated for variant 20E(EU1), which emerged in Spain and spread through Europe in the boreal summer of 2020 [72]. This hypothesis, however, failed to explain why lineage P.6 outcompeted both lineage P.7, which was the most prevalent variant in Uruguay in late 2020, and lineage P.2, which became the dominant SARS-CoV-2 strain in many Brazilian states by the end of 2020 [5,13]. An alternative hypothesis is that lineage P.6 was successfully spread in Uruguay because it was initially established in the capital city of Montevideo, which comprises nearly half of Uruguayan inhabitants and is strongly connected with all other departments. By contrast, lineages P.2 and P.7 were probably initially established outside the metropolitan region, and this may have reduced their chance of spreading at the country level. Finally, we hypothesized that the combined presence of amino acid changes S:Q675H + Q677H might have also produced a more transmissible P.6 variant, contributing to the rapid increase in the lineage dominance observed between December 2020 and February 2021.

We are not aware of any experimental assay that assessed the effect of mutations S:Q675H + Q677H on the viral fitness, but independent data indicated that these amino acid changes (either one or both of them) might facilitate viral transmissibility. These mutations are in close proximity to the polybasic cleavage site at the S1/S2 boundary that can be processed by furin and other proteases like TMPRSS2, mediating efficient entry into cells and increasing human-to-human transmission [73–75]. Mutations Q675H and Q677H might alter the properties of this nearby protease-cleavage site through changes in the structure conformation, glycosylation, and/or phosphorylation processes already

known to have a role in cleavage regulation [73]. Recently, the promotion of syncytium formation and virus infectivity has been shown for the D614G mutation, which is known to impact structural and thermodynamic aspects of the Spike [69,76–80] and to enhance the protease cleavage, likely by allosterically increasing the binding to furin [81–83]. Knowing that histidine residues function as pH sensors in other viruses [84], Q675H and Q677H mutations might also provide some synergic structural changes in the dynamics of the subdomain SD2, enhancing the effects of mutation D614G.

Convergent evolution is a hallmark of positive selection, and we identified the independent appearance of both S:Q675H and S:Q677H in 12 additional SARS-CoV-2 lineages. Moreover, mutations close to or at the polybasic cleavage site at the S1/S2 boundary have been reported in several VOCs and VOIs, including: Alpha (S:P681H), Beta (A701V), Delta (P681R), Eta (Q677H), Iota (A701V), Kappa (P681R), and Theta (P.3, P681H). These findings suggest that the S1/S2 boundary is a region particularly relevant for selection of mutations that resulted in more transmissible SARS-CoV-2 variants. Consistent with this notion, a recent study that used a reverse genetic system and primary human airway cultures identified mutation S:P681R as a significant determinant for enhanced viral replication fitness of the VOC Delta, and supported that Spike mutations that potentially affect furin cleavage efficiency must be closely monitored for future variant surveillance [85]. Notably, another study that developed an innovative model on epidemiological variables integrating the effect of Spike amino acid changes in viral fitness forecasted that mutations Q675H and Q677H could appear in emerging SARS-CoV-2 VOCs in the following months [86]. These observations underscore the importance of future experimental studies to assess the functional impact of Spike mutations Q675H and Q677H on virus infectivity and transmissibility.

In summary, this study described the emergence and local spread of lineage P.6, a new B.1.1.28-derived lineage carrying Spike mutations Q675H + Q677H, in Uruguay that coincided with the first exponential growth phase of the country's COVID-19 epidemic, which started by November 2020 and lasted until mid-February 2021. P.6 was the second recognized B.1.1.28-descendent lineage, together with lineage P.3, that emerged outside Brazil. An ancestral B.1.1.28 virus carrying mutation Q675H was probably introduced from southeastern Brazil into Montevideo, Uruguay's capital city, and by November 2020 the virus already fixed mutation Q677H and spread across the entire country, originating lineage P.6. We propose that simultaneous presence of Spike mutations Q675H and Q677H might confer to lineage P.6 a higher infectivity and increased transmissibility, which, combined with the establishment in the populated metropolitan region, contributed to its swift dissemination in Uruguay. Although the lineage P.6 was substituted by the VOC P.1 as the most prevalent lineage in Uruguay since April 2021, the concurrent emergence of Spike mutations Q675H and Q677H in VOIs and/or VOCs circulating worldwide should be closely monitored.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v13091801/s1>. Figure S1. (A) Monthly COVID-19 cases reported in Uruguay from November 2020 to April 2021 (*y*-axis to the left). In pink, total number of sequences generated in this study. In blue, total number of sequences generated by our group, including several not yet publicly available (*y*-axis to the right). (B) Uruguayan map with the number of P.2 sequences generated by the IiGW as indicated in the legend. Some department names are written accordingly. (C) Uruguayan map showing departments with simultaneous incidence of P.6 and P.2 based on data generated by the IiGW. The barplots show the absolute frequency of P.6 and P.2 monthly, from December 2020 to April 2021, in a few selected departments (Salto, Montevideo, and Rocha). Figure S2. Maximum likelihood tree of 1783 B.1.1.28 sequences found in Brazil and Uruguay. Additionally, four sequences from the USA, Spain, and Belgium were included. The root of the tree was established using the earliest sequence from Brazil, with a collection date of 5 March 2020. The Uruguayan P.6 clade and Brazilian basal sequences are highlighted in purple and pink, respectively. In addition, taxon names for samples collected in the USA and Spain are shown in blue. Uruguayan P.7 main clade and additional sequences are shown in green; Brazilian P.7 clade is shown in light green. The remaining

Uruguayan sequences through the ML tree are indicated by a blue taxon name, and supported clades are highlighted in blue. Figure S3. Schematic representation of migration events during dissemination of SARS-CoV-2 lineage B.1.1.28 between Uruguay and Brazil (including recently assigned P.6 and P.7 samples). The migration events were inferred by ancestral character reconstruction obtained through a maximum likelihood method implemented in PastML. Each node in the network is identified by location and number of sequences within different phylogenetic subclusters. Arrows indicate migration events deduced from location state changes across the B.1.1.28 ML tree. The shade of gray identifies marginal probabilities, and the numbers quantify the migration events connecting respective locations (no numbers represent one single event). Nodes are colored according to their location. BR: Brazil. Figure S4. Maximum likelihood phylogeographic analysis of lineage P.2 samples ($n = 1346$) from Uruguay ($n = 79$) and Brazil ($n = 1267$) inferred by ancestral character reconstruction method implemented in PastML. Tips and branches are colored according to sampling location and the most probable location state of their descendent nodes, respectively, as indicated in the legend. Shaded boxes highlight the major P.2 clades in Uruguay (labels UY-I, UY-II, and UY-III in the tree correspond to clades UY-I_{P2}, UY-II_{P2}, and UY-III_{P2} mentioned in the main manuscript, respectively). The time-scaled tree was rooted with the earliest P.2 sequence (collection date: 1 September 2020). Branch lengths are drawn to scale indicating nucleotide substitutions per site per year. Figure S5. Geographical and Pango lineage distribution of worldwide SARS-CoV-2 genome sequences carrying mutations Q675H + Q677H. Barplots indicate the number of observed cases per country and lineage. The analysis is based on the 174 B.1.1.28 + Q675H + Q677H sequences from this study and an additional 85 complete and high quality genomes, collected worldwide and carrying Q675H + Q677H independently of the assigned lineage. These sequences were obtained from GISAID on 7 July 2021. B.1.1.28 corresponds to the newly designated P.6 Pango lineage plus one independent arisen sequence collected in Belgium. Table S1. Information on SARS-CoV-2 samples from Uruguay used in this study. GISAID accession identifier, liWG identifier, collection date, and geographic origin in Uruguay (department) and Pango lineage are indicated. Additionally, # indicates sequences obtained with the standard ARTIC protocol, #c indicates those obtained with the standard ARTIC protocol with long amplicons, and * indicates the samples processed using the “midnight” primer set, using the Nanopore Rapid kit. Table S2. Information about Uruguayan, Brazilian, and worldwide samples obtained from the EpiCoV database in GISAID and used in this study (accessed on 7 July 2021). Table S3. GISAID (EpiCoV database) acknowledgments table.

Author Contributions: Conceptualization: N.R. (Natalia Rego), G.B., L.S.; methodology, C.S., M.P., A.C., A.F., P.P., V.N.; software, I.F.; formal analysis, N.R. (Natalia Rego), C.S., M.P., A.C., A.F., I.A., D.M., P.P., M.R.M., G.B.; investigation, N.R. (Natalia Rego), C.S., M.P., A.C., A.F., P.P., T.F.-C., V.N., M.B., R.A., M.A. (Mailen Arleo), T.P., N.R. (Natalia Reyes), M.N.B., A.L., V.B., A.M., O.C., N.N., J.H., M.D., M.B.G., L.G., M.M. (Mauricio Mendez), M.P.T., J.Z., B.R., M.M. (Matias Maidana), M.A. (Martina Alonso), P.S., I.A., E.P., D.M., C.A., J.M., H.A., R.C., G.B., P.M., G.M., G.I., L.S.; resources, V.N., M.B., R.A., M.A. (Mailen Arleo), T.P., N.R. (Natalia Reyes), M.N.B., A.L., V.B., A.M., O.C., N.N., J.H., M.D., M.B.G., L.G., M.M. (Mauricio Mendez), M.P.T., J.Z., B.R., M.M. (Matias Maidana), M.A. (Martina Alonso), P.S., C.A., J.M., H.A., R.C.; data curation, N.R. (Natalia Rego), C.S., I.F., T.F.-C.; writing—original draft preparation, N.R. (Natalia Rego), L.S.; writing—review and editing, C.S., T.F.-C., M.R.M., D.M., G.B., P.M., G.M., G.I.; visualization, M.R.M.; supervision, P.M., G.M., G.I., L.S.; funding acquisition, P.M., G.M., G.I. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by FOCEM-Fondo para la Convergencia Estructural del Mercosur (COF03/11).

Institutional Review Board Statement: The National Ministry of Health (Uruguay) is the main health Institution in our country. It is a dedicated ethics oversight body, and granted us the ethical approval for this work. All necessary patient/participant consent was obtained, and the appropriate institutional forms were archived.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All SARS-CoV-2 genome sequences have been submitted to the EpiCoV/GISAID database with accession numbers indicated in Table S1.

Acknowledgments: The authors wish to thank all the health care workers and scientists who have worked hard to deal with this pandemic threat, the GISAID team, and all the EpiCoV database's submitters (the GISAID acknowledgments table containing the sequences used in this study is shown in Table S3). We thank Christian Brandt from the Institute for Infectious Diseases and Infection Control (Jena University Hospital) for adapting their poreCov Nextflow pipeline to our requirements and quickly fixing reported bugs. We thank Maria Jose Benitez-Galeano for fruitful discussions. We thank the Zurich-Santander Foundation for their support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CDC Coronavirus Disease 2019 (COVID-19). Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html> (accessed on 17 August 2021).
2. Janik, E.; Niemcewicz, M.; Podogrocki, M.; Majsterek, I.; Bijak, M. The Emerging Concern and Interest SARS-CoV-2 Variants. *Pathogens* **2021**, *10*, 633. [CrossRef]
3. Candido, D.S.; Claro, I.M.; de Jesus, J.G.; Souza, W.M.; Moreira, F.R.R.; Dellicour, S.; Mellan, T.A.; du Plessis, L.; Pereira, R.H.M.; Sales, F.C.S.; et al. Evolution and Epidemic Spread of SARS-CoV-2 in Brazil. *Science* **2020**, *369*, 1255–1260. [CrossRef]
4. Resende, P.C.; Delatorre, E.; Gräf, T.; Mir, D.; Motta, F.C.; Appolinario, L.R.; da Paixão, A.C.D.; da Fonseca Mendonça, A.C.; Ogrzewalska, M.; Caetano, B.; et al. Evolutionary Dynamics and Dissemination Pattern of the SARS-CoV-2 Lineage B.1.1.33 During the Early Pandemic Phase in Brazil. *Front. Microbiol.* **2021**, *11*, 3565. [CrossRef] [PubMed]
5. Rede Genomica Fiocruz. Available online: <http://www.genomahcov.fiocruz.br/> (accessed on 17 August 2021).
6. Fujino, T.; Nomoto, H.; Kutsuna, S.; Ujiie, M.; Suzuki, T.; Sato, R.; Fujimoto, T.; Kuroda, M.; Wakita, T.; Ohmagari, N. Novel SARS-CoV-2 Variant in Travelers from Brazil to Japan. *Emerg. Infect. Dis.* **2021**, *27*, 1243. [CrossRef] [PubMed]
7. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.D.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.S.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil. *Science* **2021**, *372*, 815–821. [CrossRef] [PubMed]
8. Naveca, F.G.; Nascimento, V.; de Souza, V.C.; de Lima Corado, A.; Nascimento, F.; Silva, G.; Costa, Á.; Duarte, D.; Pessoa, K.; Mejía, M.; et al. COVID-19 in Amazonas, Brazil, Was Driven by the Persistence of Endemic Lineages and P.1 Emergence. *Nat. Med.* **2021**, *27*, 1230–1238. [CrossRef]
9. Cov-Lineages. Available online: <https://cov-lineages.org/lineage.html?lineage=P.1> (accessed on 17 August 2021).
10. Cov-Lineages. Available online: <https://cov-lineages.org/lineage.html?lineage=P.2> (accessed on 17 August 2021).
11. Voloch, C.M.; da Silva Francisco, R.; de Almeida, L.G.P.; Cardoso, C.C.; Brustolini, O.J.; Gerber, A.L.; de C Guimarães, A.P.; Mariani, D.; da Costa, R.M.; Ferreira, O.C.; et al. Genomic Characterization of a Novel SARS-CoV-2 Lineage from Rio de Janeiro, Brazil. *J. Virol.* **2021**, *95*. [CrossRef]
12. Lamarca, A.P.; de Almeida, L.G.P.; da S. Francisco, R.; Lima, L.F.A.; Scortecchi, K.C.; Perez, V.P.; Brustolini, O.J.; Sousa, E.S.S.; Secco, D.A.; Santos, A.M.G.; et al. Genomic Surveillance of SARS-CoV-2 Tracks Early Interstate Transmission of P.1 Lineage and Diversification within P.2 Clade in Brazil. *medRxiv* **2021**. Available online: <https://www.medrxiv.org/content/10.1101/2021.03.21.21253418v2> (accessed on 17 August 2021). [CrossRef]
13. Franceschi, V.B.; Ferrareze, P.A.G.; Zimmerman, R.A.; Cybis, G.B.; Thompson, C.E. Mutation hotspots and spatiotemporal distribution of SARS-CoV-2 lineages in Brazil, February 2020–2021. *Virus Res.* **2021**, *304*, 198532. [CrossRef] [PubMed]
14. da Silva Francisco, R., Jr.; Lamarca, A.P.; de Almeida, L.G.P.; Cavalcante, L.; Machado, D.T.; Martins, Y.; Brustolini, O.; Gerber, A.L.; de C Guimarães, A.P.; Gonçalves, R.B.; et al. Turnover of SARS-CoV-2 Lineages Shaped the Pandemic and Enabled the Emergence of New Variants in the State of Rio de Janeiro, Brazil. *medRxiv* **2021**. Available online: <https://www.medrxiv.org/content/10.1101/2021.07.20.21260890v1> (accessed on 17 August 2021). [CrossRef]
15. Potential New B.1.1.28 Sub-Lineage with L452R in Brazil Issue #68-Cov-Lineages/Pango-Designation. Available online: <https://github.com/cov-lineages/pango-designation/issues/68> (accessed on 17 August 2021).
16. Tablizo, F.A.; Kim, K.M.; Lapid, C.M.; Castro, M.J.R.; Yangzon, M.S.L.; Maralit, B.A.; Ayes, M.E.C.; Cutiongco-de la Paz, E.M.; De Guzman, A.R.; Yap, J.M.C.; et al. Genome Sequencing and Analysis of an Emergent SARS-CoV-2 Variant Characterized by Multiple Spike Protein Mutations Detected from the Central Visayas Region of the Philippines. *medRxiv* **2021**. Available online: <https://www.medrxiv.org/content/10.1101/2021.03.03.21252812v2> (accessed on 17 August 2021). [CrossRef]
17. da Silva Francisco, R., Jr.; Benites, L.F.; Lamarca, A.P.; de Almeida, L.G.P.; Hansen, A.W.; Gularte, J.S.; Demoliner, M.; Gerber, A.L.; de C Guimarães, A.P.; Antunes, A.K.E.; et al. Pervasive Transmission of E484K and Emergence of VUI-NP13L with Evidence of SARS-CoV-2 Co-Infection Events by Two Different Lineages in Rio Grande Do Sul, Brazil. *Virus Res.* **2021**, *296*, 198345. [CrossRef] [PubMed]
18. Sant'Anna, F.H.; Varela, A.P.M.; Prichula, J.; Comerlato, J.; Comerlato, C.B.; Roglio, V.S.; Pereira, G.F.M.; Moreno, F.; Seixas, A.; Wendland, E.M. Emergence of the Novel SARS-CoV-2 Lineage VUI-NP13L and Massive Spread of P.2 in South Brazil. *Emerg. Microbes Infect.* **2021**, *10*, 1431–1440. [CrossRef]
19. Estadisticasuy. Available online: <https://guiad-covid.github.io/estadisticasuy.html> (accessed on 26 July 2021).
20. Taylor, L. Why Uruguay Lost Control of COVID. *Nature* **2021**, *595*, 21. [CrossRef] [PubMed]

21. Elizondo, V.; Harkins, G.W.; Mabvakure, B.; Smidt, S.; Zappile, P.; Marier, C.; Maurano, M.T.; Perez, V.; Mazza, N.; Beloso, C.; et al. SARS-CoV-2 Genomic Characterization and Clinical Manifestation of the COVID-19 Outbreak in Uruguay. *Emerg. Microbes Infect.* **2021**, *10*, 51–65. [[CrossRef](#)] [[PubMed](#)]
22. Mir, D.; Rego, N.; Resende, P.C.; Tort, F.; Fernández-Calero, T.; Noya, V.; Brandes, M.; Possi, T.; Arleo, M.; Reyes, N.; et al. Recurrent Dissemination of SARS-CoV-2 Through the Uruguayan–Brazilian Border. *Front. Microbiol.* **2021**, *12*, 653986. [[CrossRef](#)]
23. Fraser, C.; Riley, S.; Anderson, R.M.; Ferguson, N.M. Factors That Make an Infectious Disease Outbreak Controllable. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6146–6151. [[CrossRef](#)]
24. Grantz, K.H.; Lee, E.C.; McGowan, L.D.; Lee, K.H.; Metcalf, C.J.E.; Gurley, E.S.; Lessler, J. Maximizing and Evaluating the Impact of Test-Trace-Isolate Programs: A Modeling Study. *PLoS Med.* **2021**, *18*, e1003585. [[CrossRef](#)]
25. Uruguay Medidas. Available online: https://medios.presidencia.gub.uy/tav_portal/2020/noticias/AH_204/Medidas%2016.12.2020.pdf (accessed on 16 December 2020).
26. Rego, N.; Costabile, A.; Paz, M.; Salazar, C.; Perbolianachis, P.; Spangenberg, L.; Ferrés, I.; Arce, R.; Fajardo, A.; Arleo, M.; et al. Real-Time Genomic Surveillance for SARS-CoV-2 Variants of Concern, Uruguay. *Emerg. Infect. Dis.* **2021**, *27*. [[CrossRef](#)]
27. Cov-Lineages. Available online: <https://cov-lineages.org/lineage.html?lineage=P.6> (accessed on 17 August 2021).
28. Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N.G.; Decroly, E. The Spike Glycoprotein of the New Coronavirus 2019-nCoV Contains a Furin-like Cleavage Site Absent in CoV of the Same Clade. *Antiviral Res.* **2020**, *176*, 104742. [[CrossRef](#)]
29. Quick, J. NCoV-2019 Sequencing Protocol v3 (LoCost). 2020. Available online: <https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bh42j8ye/metrics> (accessed on 31 August 2021).
30. Tyson, J.R.; James, P.; Stoddart, D.; Sparks, N.; Wickenhagen, A.; Hall, G.; Choi, J.H.; Lapointe, H.; Kamelian, K.; Smith, A.D.; et al. Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore. *BioRxiv* **2020**. [[CrossRef](#)]
31. Resende, P. Long Reads Nanopore Sequencing to Recover SARS-CoV-2 Whole Genome. *Protocols* **2020**. [[CrossRef](#)]
32. Freed, N.; Silander, O. SARS-CoV2 Genome Sequencing Protocol (1200bp Amplicon “Midnight” Primer Set, Using Nanopore Rapid Kit) V5. Available online: https://www.protocols.io/view/sars-cov2-genome-sequencing-protocol-1200bp-amplicon-midnight-6?version_warning=no (accessed on 31 August 2021).
33. 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.* **2020**, *5*, bpaa014. [[CrossRef](#)] [[PubMed](#)]
34. Oxford Nanopore Technologies. Available online: <https://nanoporetech.com/> (accessed on 2 June 2021).
35. GitHub-Replikation/PoreCov: SARS-CoV-2 Workflow for Nanopore Sequence Data. Available online: <https://github.com/replikation/poreCov> (accessed on 23 July 2021).
36. Hufsky, F.; Lamkiewicz, K.; Almeida, A.; Aouacheria, A.; Arighi, C.; Bateman, A.; Baumbach, J.; Beerenwinkel, N.; Brandt, C.; Cacciabue, M.; et al. Computational Strategies to Combat COVID-19: Useful Tools to Accelerate SARS-CoV-2 and Coronavirus Research. *Brief. Bioinform.* **2021**, *22*, 642–663. [[CrossRef](#)]
37. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [[CrossRef](#)]
38. Loman, N.J.; Quick, J.; Simpson, J.T. A Complete Bacterial Genome Assembled de Novo Using Only Nanopore Sequencing Data. *Nat. Methods* **2015**, *12*, 733–735. [[CrossRef](#)] [[PubMed](#)]
39. Di Tommaso, P.; Chatzou, M.; Floden, E.W.; Barja, P.P.; Palumbo, E.; Notredame, C. Nextflow Enables Reproducible Computational Workflows. *Nat. Biotechnol.* **2017**, *35*, 316–319. [[CrossRef](#)] [[PubMed](#)]
40. Kurtzer, G.M.; Sochat, V.; Bauer, M.W. Singularity: Scientific Containers for Mobility of Compute. *PLoS ONE* **2017**, *12*, e0177459. [[CrossRef](#)] [[PubMed](#)]
41. Ferguson, J.M.; Gamaarachchi, H.; Nguyen, T.; Gollon, A.; Tong, S.; Aquilina-Reid, C.; Bowen-James, R.; Deveson, I.W. InterARTIC: An Interactive Web Application for Whole-Genome Nanopore Sequencing Analysis of SARS-CoV-2 and Other Viruses. *bioRxiv* **2021**. [[CrossRef](#)]
42. Li, H. Minimap2: Pairwise Alignment for Nucleotide Sequences. *Bioinformatics* **2018**, *34*, 3094–3100. [[CrossRef](#)]
43. Ondov, B.D.; Bergman, N.H.; Phillippy, A.M. Interactive Metagenomic Visualization in a Web Browser. *BMC Bioinform.* **2011**, *12*, 385. [[CrossRef](#)] [[PubMed](#)]
44. Brandt, C.; Krautwurst, S.; Spott, R.; Lohde, M.; Jundzill, M.; Marquet, M.; Hölzer, M. PoreCov—An Easy to Use, Fast, and Robust Workflow for SARS-CoV-2 Genome Reconstruction via Nanopore Sequencing. *Front. Genet.* **2021**, *12*, 711437. [[CrossRef](#)] [[PubMed](#)]
45. Rambaut, A.; Holmes, E.C.; O’Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A Dynamic Nomenclature Proposal for SARS-CoV-2 Lineages to Assist Genomic Epidemiology. *Nat. Microbiol.* **2020**, *5*, 1403–1407. [[CrossRef](#)]
46. COG-UK. Available online: <https://pangolin.cog-uk.io/> (accessed on 26 July 2021).
47. O’Toole, Á.; Scher, E.; Underwood, A.; Jackson, B.; Hill, V.; McCrone, J.T.; Colquhoun, R.; Ruis, C.; Abu-Dahab, K.; Taylor, B.; et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evol.* **2021**, veab064. [[CrossRef](#)]
48. Shu, Y.; McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data—from Vision to Reality. *Eurosurveillance* **2017**, *22*, 30494. [[CrossRef](#)] [[PubMed](#)]
49. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]

50. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)] [[PubMed](#)]
51. Anisimova, M.; Gascuel, O. Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Syst. Biol.* **2006**, *55*, 539–552. [[CrossRef](#)]
52. Sagulenko, P.; Puller, V.; Neher, R.A. TreeTime: Maximum-Likelihood Phylodynamic Analysis. *Virus Evol.* **2018**, *4*. [[CrossRef](#)]
53. Duchene, S.; Featherstone, L.; Haritopoulou-Sinanidou, M.; Rambaut, A.; Lemey, P.; Baele, G. Temporal Signal and the Phylodynamic Threshold of SARS-CoV-2. *Virus Evol.* **2020**, *6*. [[CrossRef](#)]
54. Time Dependence of SARS-CoV-2 Substitution Rates—SARS-CoV-2 Coronavirus/NCoV-2019 Evolutionary History. Available online: <https://virological.org/t/time-dependence-of-sars-cov-2-substitution-rates/542> (accessed on 23 July 2021).
55. Ishikawa, S.A.; Zhukova, A.; Iwasaki, W.; Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Mol. Biol. Evol.* **2019**, *36*, 2069–2085. [[CrossRef](#)]
56. Suchard, M.A.; Lemey, P.; Baele, G.; Ayres, D.L.; Drummond, A.J.; Rambaut, A. Bayesian Phylogenetic and Phylodynamic Data Integration Using BEAST 1.10. *Virus Evol.* **2018**, *4*. [[CrossRef](#)]
57. Drummond, A.J.; Rambaut, A.; Shapiro, B.; Pybus, O.G. Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol. Biol. Evol.* **2005**, *22*, 1185–1192. [[CrossRef](#)]
58. Lemey, P.; Rambaut, A.; Drummond, A.J.; Suchard, M.A. Bayesian Phylogeography Finds Its Roots. *PLoS Comput. Biol.* **2009**, *5*, e1000520. [[CrossRef](#)] [[PubMed](#)]
59. Ferreira, M.A.R.; Suchard, M.A. Bayesian Analysis of Elapsed Times in Continuous-Time Markov Chains. *Can. J. Stat.* **2008**, *36*, 355–368. [[CrossRef](#)]
60. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [[CrossRef](#)] [[PubMed](#)]
61. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; Maio, N.D.; et al. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2019**, *15*, e1006650. [[CrossRef](#)]
62. FigTree. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 26 July 2021).
63. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T.-Y. Ggtree: An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol. Evol.* **2017**, *8*, 28–36. [[CrossRef](#)]
64. D. E. Shaw Research. Technical Data Molecular Dynamics Simulations Related to SARS-CoV-2. Available online: https://www.deshawresearch.com/downloads/download_trajectory_sarscov2.cgi/ (accessed on 26 July 2021).
65. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
66. Lemmin, T.; Kalbermatter, D.; Harder, D.; Plattet, P.; Fotiadis, D. Structures and Dynamics of the Novel S1/S2 Protease Cleavage Site Loop of the SARS-CoV-2 Spike Glycoprotein. *J. Struct. Biol. X* **2020**, *4*, 100038. [[CrossRef](#)]
67. Bagdonaite, I.; Thompson, A.J.; Wang, X.; Søgaard, M.; Fougereux, C.; Frank, M.; Diedrich, J.K.; Yates, J.R.; Salanti, A.; Vakhrushev, S.Y.; et al. Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses* **2021**, *13*, 551. [[CrossRef](#)] [[PubMed](#)]
68. Sanda, M.; Morrison, L.; Goldman, R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* **2021**, *93*, 2003–2009. [[CrossRef](#)]
69. Gobeil, S.M.-C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M.F.; Manne, K.; Li, D.; Wiehe, K.; et al. Effect of Natural Mutations of SARS-CoV-2 on Spike Structure, Conformation, and Antigenicity. *Science* **2021**. [[CrossRef](#)] [[PubMed](#)]
70. Hodcroft, E.B.; Domman, D.B.; Snyder, D.J.; Oguntuyo, K.Y.; Diest, M.V.; Densmore, K.H.; Schwalm, K.C.; Femling, J.; Carroll, J.L.; Scott, R.S.; et al. Emergence in Late 2020 of Multiple Lineages of SARS-CoV-2 Spike Protein Variants Affecting Amino Acid Position 677. *medRxiv* **2021**. [[CrossRef](#)]
71. Detection of the Recurrent Substitution Q677H in the Spike Protein of SARS-CoV-2 in Cases Descended from the Lineage B.1.429—SARS-CoV-2 Coronavirus/NCoV-2019 Genomic Epidemiology. Available online: <https://virological.org/t/detection-of-the-recurrent-substitution-q677h-in-the-spike-protein-of-sars-cov-2-in-cases-descended-from-the-lineage-b-1-429/660> (accessed on 22 July 2021).
72. Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Vaughan, T.G.; Crawford, K.H.D.; Althaus, C.L.; Reichmuth, M.L.; Bowen, J.E.; Walls, A.C.; Corti, D.; et al. Spread of a SARS-CoV-2 Variant through Europe in the Summer of 2020. *Nature* **2021**, 1–6. [[CrossRef](#)]
73. Örd, M.; Faustova, I.; Loog, M. The Sequence at Spike S1/S2 Site Enables Cleavage by Furin and Phospho-Regulation in SARS-CoV2 but Not in SARS-CoV1 or MERS-CoV. *Sci. Rep.* **2020**, *10*, 16944. [[CrossRef](#)]
74. Tang, T.; Jaimes, J.A.; Bidon, M.K.; Straus, M.R.; Daniel, S.; Whittaker, G.R. Proteolytic Activation of SARS-CoV-2 Spike at the S1/S2 Boundary: Potential Role of Proteases beyond Furin. *ACS Infect. Dis.* **2021**, *7*, 264–272. [[CrossRef](#)] [[PubMed](#)]
75. Peacock, T.P.; Goldhill, D.H.; Zhou, J.; Baillon, L.; Frise, R.; Swann, O.C.; Kugathasan, R.; Penn, R.; Brown, J.C.; Sanchez-David, R.Y.; et al. The Furin Cleavage Site in the SARS-CoV-2 Spike Protein Is Required for Transmission in Ferrets. *Nat. Microbiol.* **2021**, *6*, 899–909. [[CrossRef](#)] [[PubMed](#)]
76. Zhang, J.; Cai, Y.; Xiao, T.; Lu, J.; Peng, H.; Sterling, S.M.; Walsh, R.M.; Rits-Volloch, S.; Zhu, H.; Woosley, A.N.; et al. Structural Impact on SARS-CoV-2 Spike Protein by D614G Substitution. *Science* **2021**, *372*, 525–530. [[CrossRef](#)]

77. Mansbach, R.A.; Chakraborty, S.; Nguyen, K.; Montefiori, D.C.; Korber, B.; Gnanakaran, S. The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational State. *Sci. Adv.* **2021**, *7*, eabf3671. [[CrossRef](#)]
78. Yurkovetskiy, L.; Wang, X.; Pascal, K.E.; Tomkins-Tinch, C.; Nyalile, T.P.; Wang, Y.; Baum, A.; Diehl, W.E.; Dauphin, A.; Carbone, C.; et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020**, *183*, 739–751.e8. [[CrossRef](#)]
79. Benton, D.J.; Wrobel, A.G.; Roustan, C.; Borg, A.; Xu, P.; Martin, S.R.; Rosenthal, P.B.; Skehel, J.J.; Gamblin, S.J. The Effect of the D614G Substitution on the Structure of the Spike Glycoprotein of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* **2021**, *118*. [[CrossRef](#)]
80. Zhang, L.; Jackson, C.B.; Mou, H.; Ojha, A.; Peng, H.; Quinlan, B.D.; Rangarajan, E.S.; Pan, A.; Vanderheiden, A.; Suthar, M.S.; et al. SARS-CoV-2 Spike-Protein D614G Mutation Increases Virion Spike Density and Infectivity. *Nat. Commun.* **2020**, *11*, 6013. [[CrossRef](#)]
81. Gobeil, S.M.-C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Manne, K.; Stalls, V.; Kopp, M.F.; Henderson, R.; Edwards, R.J.; et al. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep.* **2021**, *34*, 108630. [[CrossRef](#)]
82. Mohammad, A.; Alshawaf, E.; Marafie, S.K.; Abu-Farha, M.; Abubaker, J.; Al-Mulla, F. Higher Binding Affinity of Furin for SARS-CoV-2 Spike (S) Protein D614G Mutant Could Be Associated with Higher SARS-CoV-2 Infectivity. *Int. J. Infect. Dis.* **2021**, *103*, 611–616. [[CrossRef](#)] [[PubMed](#)]
83. Cheng, Y.-W.; Chao, T.-L.; Li, C.-L.; Wang, S.-H.; Kao, H.-C.; Tsai, Y.-M.; Wang, H.-Y.; Hsieh, C.-L.; Lin, Y.-Y.; Chen, P.-J.; et al. D614G Substitution of SARS-CoV-2 Spike Protein Increases Syncytium Formation and Virus Titer via Enhanced Furin-Mediated Spike Cleavage. *mBio* **2021**, *12*, e00587-21. [[CrossRef](#)] [[PubMed](#)]
84. Kampmann, T.; Mueller, D.S.; Mark, A.E.; Young, P.R.; Kobe, B. The Role of Histidine Residues in Low-PH-Mediated Viral Membrane Fusion. *Structure* **2006**, *14*, 1481–1487. [[CrossRef](#)] [[PubMed](#)]
85. Liu, Y.; Liu, J.; Johnson, B.A.; Xia, H.; Ku, Z.; Schindewolf, C.; Widen, S.G.; An, Z.; Weaver, S.C.; Menachery, V.D.; et al. Delta Spike P681R Mutation Enhances SARS-CoV-2 Fitness over Alpha Variant. *bioRxiv* **2021**. Available online: <https://www.biorxiv.org/content/10.1101/2021.08.12.456173v1> (accessed on 26 June 2021). [[CrossRef](#)]
86. Maher, M.C.; Bartha, I.; Weaver, S.; di Iulio, J.; Ferri, E.; Soriaga, L.; Lempp, F.A.; Hie, B.L.; Bryson, B.; Berger, B.; et al. Predicting the Mutational Drivers of Future SARS-CoV-2 Variants of Concern. *medRxiv* **2021**. [[CrossRef](#)]



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Phylogenetic analysis of SARS-CoV-2 viruses circulating in the South American region: Genetic relations and vaccine strain match

Paula Perbolianachis, Diego Ferla, Rodrigo Arce, Irene Ferreiro, Alicia Costáble, Mercedes Paz, Diego Simón, Pilar Moreno, Juan Cristina*

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

ARTICLE INFO

Keywords:

Coronavirus
Evolution
SARS-CoV-2
South America
COVID-19

ABSTRACT

The pandemic of coronavirus disease 2019 (COVID-19) is caused by a novel member of the family *Coronaviridae*, now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Recent studies revealed the emergence of virus variants with substitutions in the spike and/or nucleocapsid and RNA-dependent RNA polymerase proteins that are partly responsible for enhanced transmission and reduced or escaped anti-SARS-CoV-2 antibodies that may reduce the efficacy of antibodies and vaccines against the first identified SARS-CoV-2 strains. In order to gain insight into the emergence and evolution of SARS-CoV-2 variants circulating in the South American region, a comprehensive phylogenetic study of SARS-CoV-2 variants circulating in this region was performed. The results of these studies revealed sharp increase in virus effective population size from March to April of 2020. At least 62 different genotypes were found to circulate in this region. Variants of concern (VOCs) Alpha, Beta, Gamma and Delta co-circulate in the region, together with variants of interest (VOIs) Lambda, Mu and Zeta. Most of SARS-CoV-2 variants circulating in the South American region belongs to B.1 genotypes and have substitutions in the spike and/or nucleocapsid and polymerase proteins that confer high transmissibility and/or immune resistance. 148 amino acid positions of the spike protein and 70 positions of the nucleocapsid were found to have substitutions in different variants isolated in the region by comparison with reference strain Wuhan-Hu-1. Significant differences in codon usage among spike genes of SARS-CoV-2 strains circulating in South America was found, which can be linked to SARS-CoV-2 genotypes.

1. Introduction

The pandemic of coronavirus disease 2019 (COVID-19) started in China in December of 2019 (Li, Guan and Wu, 2020) (Li et al., 2020). This severe respiratory pneumonia is caused by a novel member of the family *Coronaviridae*, now known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Gorbalenya, Baker and Baric, 2020; Gorbalenya et al., 2020). The World Health Organization declare this SARS-CoV-2 pandemic as a public health emergency of international concern on January 30th, 2020 (World Health Organization 2020a; WHO, 2020a). 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 2020b) (WHO, 2020b).

As all members of the family *Coronaviridae*, SARS-CoV-2 possess a single stranded, positive-sense RNA genome of approximately 30 kb

bases 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, Liu and Guo, 2020; Chen et al., 2020).

The spike (S) glycoprotein of SARS-CoV-2 facilitates coronavirus entry into host cells. The S protein forms a homotrimeric complex protruding from the viral surface and consists of two functional subunits, S1 and S2, which are responsible for host cell receptor binding and the viral fusion to the host cellular membranes (Xia, 2021) (Xia et al., 2021). The smaller S1 subunit consists of an N-terminal domain (NTD) and three C-terminal domains (CTD1–3), of which CTD1 forms the receptor-binding domain (RBD) and contributes to the stabilization of the membrane-anchored S2 subunit. The larger S2 subunit contains the machinery for viral fusion and comprises a hydrophobic fusion peptide (FP), heptad repeat 1 (HR1), central helix (CH), connector domain (CD), heptad repeat 2 (HR2), transmembrane domain (TM) and cytoplasmic

* Correspondence author.

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

<https://doi.org/10.1016/j.virusres.2022.198688>

Received 27 October 2021; Received in revised form 19 January 2022; Accepted 20 January 2022

Available online 21 January 2022

0168-1702/© 2022 Elsevier B.V. All rights reserved.

tail (CT) (Walls et al., 2020) (Walls et al., 2020). The HR1 is vital for the S protein stability, maintaining the correct protein fold in the closed pre-fusion conformation (Juraszek et al., 2020). As in all members of the family *Coronaviridae*, the S protein is cleaved by host proteases at the S1/S2 junction; cleavage has been suggested to activate the protein for host membrane fusion through irreversible conformational changes. There is a second cleavage site, S2', located 130 residues from the N terminus of the S2 subunit, which is highly conserved among coronaviruses. Cleavage at the S2' site by host cell proteases is vital for successful infection by coronaviruses (Belouzard, Chu and Whittaker, 2009; Belouzard et al., 2009). SARS-CoV-2 S protein interacts directly with the host cell receptor angiotensin-converting enzyme 2 (ACE2) (Hoffmann et al., 2020). ACE2 is a protease responsible for blood pressure and volume regulation and it is widely expressed on the cell membranes of the lung, heart, kidneys and gastrointestinal tract (Samavati and Uhal, 2020) (Samavati and Uhal, 2020).

The RBD of S protein is immunodominant and the target of 90% of the neutralizing activity present in SARS-CoV-2 immune sera (Piccoli, Park and Tortorici, 2020; Piccoli et al., 2020). Antibodies directed to S protein NTD and to RBD can neutralize with high potency (less than 0.01 $\mu\text{g}/\text{mL}$ IC50) (Cerutti et al., 2021). While RBD shows many non-overlapping antigenic sites (Barnes et al., 2020), NTD appears to contain only a single site of vulnerability to neutralization, identified as an antigenic supersite (McCallum, De Marco and Lempp, 2021a; McCallum et al., 2021).

Clinically applied monoclonal antibodies (Weinreich et al., 2021; Weinreich et al., 2021) and vaccinations (Wang and Cheng, 2021) (Wang et al., 2021) have shown a significant success in virus neutralization. However, recent studies on SARS-CoV-2 evolution revealed the emergence of variants with substitutions in the S protein that appears to be more transmissible, increasing the affinity with ACE-2 or are partly responsible for reduced or escaping to anti-SARS-CoV-2 antibodies (Davies et al., 2021; Davies et al., 2021). These variants are known as variants of concern (VOC), i.e.: 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). Very recently, by November 24th, 2021, a new VOC named Omicron (B.1.1.529) identified in South Africa (Wang and Cheng, 2021; Wang and Chen, 2021). All VOCs have substitutions in the S NTD and/or RBD domains (Chakraborty, Bhattacharya and Sharma, 2021; Chakraborty et al., 2021) (see Supplementary Material Table 1). Besides VOCs, there are variants that are being closely followed because of its prevalence or other phenotypic characteristics. These variants are known as variants of interest (VOI): VOI Eta (B.1.52, first identified in Nigeria); VOI Iota (B.1.53, first identified in the United States); VOI Kappa (B.1.617.1, first identified in India); VOI Lambda (C.37, first identified in Peru), VOI Mu (B.1.621, first identified in Colombia) and VOI Zeta (P.2, first identified in Brazil). These VOIs also have substitutions in S NTD and/or RBD domains (see Supplementary Material Table 1).

Therefore, there is a concern that VOCs and/or VOIs may reduce the efficacy of the anti-SARS-CoV-2 induced-antibodies or eventually evade them completely (Leach et al., 2021; Leach et al., 2021; Salleh, Derrick and Deris, 2021; Salleh et al., 2021).

In order to better understand the emergence, spread and evolution of SARS-CoV-2 variants circulating in the South American region, a comprehensive phylogenetic study of SARS-CoV-2 strains circulating in this region was performed.

2. Material and methods

2.1. Sequences

Available and comparable complete genome sequences of 933 SARS-CoV-2 variants isolated in South America from March 12th, 2020 to May 28th, 2021, were used throughout these studies (including Argentina,

Bolivia, Brazil, Chile, Colombia, Ecuador, Paraguay, Peru, Suriname, Uruguay and Venezuela). In the selection of the sequences to be included in the dataset, we carefully selected only those ones with a minimum of N in the full-length genome sequence. 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 2.

2.2. Sequence alignment

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

2.3. Data analysis

Nucleotide frequencies, codon and amino acid usage and relative synonymous codon usage (RSCU) (Sharp and Li, 1986; Sharp and Lee, 1986) of S proteins from SARS-CoV-2 variants isolated in South America were calculated using the program CodonW (written by John Peden) as implemented in the Galaxy server version 1.4.4 (Afgan, Baker and Batut, 2018; 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; Metsalu and Vilo, 2015).

Correspondence analysis (COA) is another multivariate statistical analysis. This method was used to analyze the RSCU of the S genes of SARS-CoV-2 variants enrolled in these studies. COA allows a geometrical representation of the sets of rows and columns in a dataset. Each ORF is represented as a 59-dimensional vector and each dimension corresponds to the RSCU value of one codon (excluding AUG, UGG, and stop codons). Major trends within a dataset can be determined using measures of relative inertia and genes ordered according to their position along the different axes (Greenacre, 1994; Greenacre, 1994). COA was performed on the RSCU values using the CodonW program (Afgan et al., 2018).

2.4. Bayesian Markov chain Monte Carlo analysis

To investigate the evolutionary patterns of SARS-CoV-2 variants circulating in the South American region, a Bayesian Markov Chain Monte Carlo (MCMC) approach was used as implemented in the BEAST package v2.5.2 (Bouckaert et al., 2019; Bouckaert et al., 2019). First, the evolutionary model that best fit the sequence dataset was determined using software from the IQ-TREE program (Trifinopoulos, Nguyen, von Haeseler and Minh, 2016; Trifinopoulos et al., 2016). Bayesian information criterion (BIC), Akaike information criterion (AIC), and the log of the likelihood (LnL), indicated that the GTR+ Γ +I model was the most suitable model (BIC = 16,824.83; AIC = 13,848.63; LnL = -6346.31). Both strict and relaxed molecular clock models were used to test different dynamic models (constant population size, exponential population growth, expansion population growth, logistic population growth and Bayesian Skyline). Statistical uncertainty in the data was reflected by the 95% highest probability density (HPD) values. Results were examined using the TRACER v1.6 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 using TRACER v1.6 program. Lower AICM values indicate better model fit. The Bayesian Skyline model was the best model to analyze the data. 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 v1.4.2 (available at: <http://tree.bio.ed.ac.uk>). A Bayesian Skyline was constructed using TRACER 1.6 software.

2.5. Epidemiology data

Daily number of cases and deaths due to COVID-19 in the South American region from Mar 12th, 2020 through May 28th, 2021, was obtained from Our World in Data (Dong, Du and Gardner, 2020; Dong et al., 2020).

2.6. 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 and nomenclature (Rambaut et al., 2020; Rambaut et al., 2020). In an ongoing and rapidly changing pandemic such as the one caused by SARS-CoV-2, this nomenclature system can facilitate real-time epidemiology by providing commonly agreed labels to refer to viruses circulating in different parts of the world (Rambaut et al., 2020). According to this, two main genetic lineages of SARS-CoV-2, named A and B, are still circulating in many countries around the world, reflecting the exportation of viruses from China to elsewhere before strict travel restrictions and quarantine measures were imposed there. Each descending lineage from either A or B is assigned by a numerical value (for example, lineage A.1 or lineage B.2). This iterative procedure refers to a maximum of three sublevels (for example, A.1.1.1) after which new descendant lineages are given a letter (i.e. A.1.1.1.1 would become C.1; A.1.1.1.2 would become C.2).

2.7. Prediction of N- and O-linked glycosylation sites in Spike protein

Potential N-linked glycosylation sites in S protein were predicted using the NetNGlyc 1.0 Server (Gupta and Brunak, 2002; Gupta and Brunak, 2002). The NetNGlyc server predicts N-Glycosylation sites in proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequons. A threshold value of >0.5 average potential score was set to predict glycosylated sites. By the same approach, potential O-linked glycosylation sites were predicted using the NetOGlyc-4.0 (Steenfot et al., 2013; Steenfot et al., 2013).

2.8. Amino acid sequence profiles

In order to construct and visualize amino acid sequence profiles in the S protein of SARS-CoV-2 variants isolated in South America we used SeqLogo-2.0 (Thomsen and Nielsen, 2012; Thomsen and Nielsen, 2012). This method allows a graphical representation of the information content stored in a multiple sequence alignment (MSA) and provide a highly intuitive representation of the position-specific amino acid composition. Moreover, this method includes sequence weighting to correct for data redundancy and pseudo counts to correct for low number of observations. We employed SeqLogo-2.0 using Kullback-Leibler logo type and 200 pseudo counts (Thomsen and Nielsen, 2012) (Thomsen and Nielsen, 2012).

2.9. Mapping of amino acid substitutions in a 3D structure of the receptor binding domain of Spike protein

Amino acid substitutions found in the receptor binding domain of S proteins from SARS-CoV-2 isolated in South America were mapped in the 3D structure of receptor binding domain complexed with its acceptor ACE2, available at the Protein Data Bank (PDB) under accession number 6LZG. Visualization was done using Jmol-14.0.4 software (available at: <http://www.jmol.org/>).

3. Results

3.1. Diversification and sharp increase of virus population size

To determine which SARS-CoV-2 genotypes circulate in the South American region, 933 genomes from strains circulating in this region were genotyped. 62 different genotypes were found to circulate in this region (see Fig. 1). Four VOCs circulate in the South American region. VOC Gamma (P.1) was found to circulate in almost all South American countries enrolled in these studies (Fig. 1). VOI Lambda (C.37) was found to circulate in Argentina, Chile and Peru. VOI Zeta (P.2) was found to circulate in Brazil, Ecuador, Paraguay, Suriname and Uruguay. VOI Mu (B.1.621) was found to circulate in Colombia.

Then, in order to gain insight into the mode of evolution of SARS-CoV-2 variants isolated in South America, a first analysis using 145 available and comparable complete genomes sequences from major SARS-CoV-2 genotypes isolated in South America was performed using a Bayesian MCMC approach (Bouckaert et al., 2019; Bouckaert et al., 2019) (for variants included in these analyses see Supplementary Material Table 2). The results shown in Table 1 are the outcome of 40 million steps of the MCMC, using the GTR+ Γ +I model, a relaxed molecular clock and the Bayesian Skyline model.

The results of these studies suggest that the SARS-CoV-2 variants isolated in the South American region evolved from ancestors that existed around November 26th, 2019. This is in agreement with recent estimations that point to all sequences sharing a common ancestor towards the end of 2019, supporting this as the period when SARS-CoV-2 jumped into its human host (van Dorp, Acman and Richard, 2020; Leung et al., 2021; Leung et al., 2021).

A skyline plot is a graphical representation of historical effective population sizes as a function of time. Past population sizes for these plots are estimated from genetic data, without *a priori* assumptions on the mathematical function defining the shape of the demographic trajectory. Due to these facts, skyline plots can provide realistic descriptions of the complex demographic scenarios that occur in natural populations (Navascués, Leblois and Burgarella, 2017; Navascues et al., 2017). At present, most of the methods to estimate demography from genetic data are based on the coalescent. The coalescent is a mathematical model that describes the rate at which genetic lineages coalesce (i.e., join in a common ancestor) towards the past, forming the genealogy of the sample (Kuhner, 2009; Kuhner, 2009). In order to reconstruct the demographic history of SARS-CoV-2 variants detected in South America enrolled in these studies, a Bayesian skyline plot was constructed (Drummond, Rambaut, Shapiro and Pybus, 2005). The results of these studies are shown in Fig. 2.

A sharp increase in effective population size from March to April of 2020 was observed. Then, a small descent is observed towards the end of 2020 and later recovery in the first months of 2021 to remained constant to the end of the period covered by these studies (see Fig. 2A). This is in agreement with the epidemiology observed in the region, where a sharp increase in the number of cases and deaths was observed in the region at the beginning of the pandemic period in South America, a small descent towards the end of 2020 and a new increase in the months of 2021 covered by these studies (Dong et al., 2020) (see Fig. 2B).

To study the phylogenetic relations among SARS-CoV-2 strains isolated in South America, maximum clade credibility trees were generated using software from the BEAST package (Rambaut et al., 2020). The results of these studies are shown in Supplementary Material Fig. 1.

The results of these studies revealed that both main SARS-CoV-2 genetic lineages (A and B) have been circulating in the South American region (see Supplementary Material Fig. 1). Lineage A was found circulating in Chile, Peru and Uruguay at the beginning of the pandemic period. This lineage is considered at the root of the pandemic, like Wuhan/WH04/2020 (EPI_ISL_406801), and share two nucleotide positions in SARS-CoV-2 genome (positions 8782 in ORF1ab and 28,144 in ORF8) with the closest known relative being a bat virus (RaTG13)

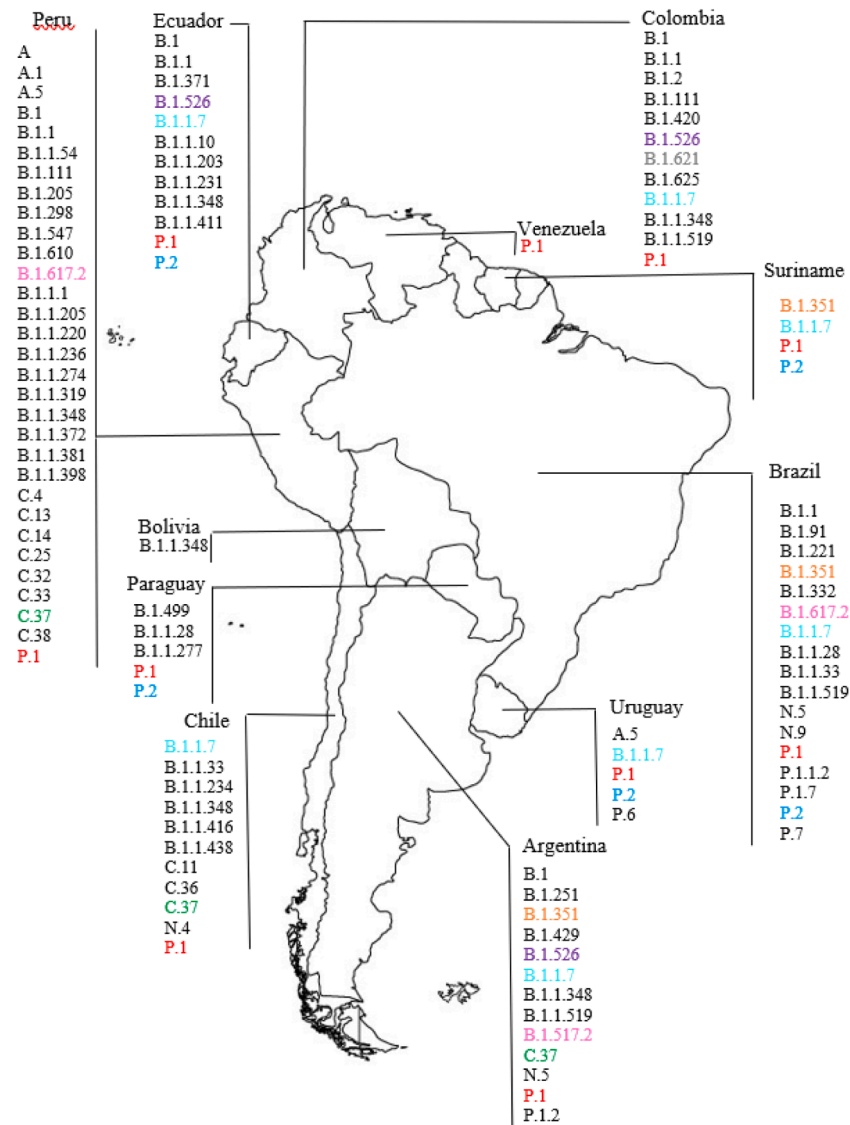


Fig. 1. Map of South America. Genotypes isolated in each country are shown under country names. VOCs Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1) and Delta (B.1.617.2) are shown in cyan, orange, red and fuchsia, respectively. VOIs Lambda (C.37), Mu (B.1.621) and Zeta (P.2) are shown in green, gray and blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
 Bayesian coalescent inference of SARS-CoV-2 strains isolated in South America.

Group ^a	Parameter	Value ^b	HPD ^c	ESS ^d
SARS-CoV-2 full-length sequences	Tree Likelihood	-52,353.03	-53,279.00 to -53,228.42	1189.15
	tMRCA ^e	1.564 11/26/ 2019	1.305 to 1.943	273.36

^a See Supplementary Material Table 2 for strains included in this analysis.
^b In all cases, the mean values are shown.
^c HPD, high probability density values.
^d ESS, effective sample size.
^e tMRCA, time of the most common recent ancestor is shown in years. The date estimated for the tMRCA is indicated in bold.

(Chan, Kok and Zhu, 2020). This is in agreement with ongoing studies showing that the lineage B (particularly B.1) have spread and replaced the lineage A in several different countries (Korber, Fischer and Gnanakaran, 2020; Korber et al., 2020). An extensive co-circulation of

different genetic lineages is observed in several countries in the region (Supplementary Material Fig. 1). Most of SARS-CoV-2 variants circulating in the South American region belongs to B.1 genotypes. These genotypes have diversified from its entry in the region. VOCs Alpha, Beta, Gamma and Delta were found to co-circulate in the region during the period covered by these studies (March 12th, 2020 to May 28th, 2021) (see Supplementary Material Fig. 1). Since most of the lineages found to circulate in the South American region were B.1 lineages, they carry an amino acid substitution at position 614 of the S protein (D614G). Variants having this substitution have shown to be more transmissible (Hou et al., 2020; Volz, Hill and McCrone, 2021; Volz et al., 2021).

3.2. Substitutions were found in all domains of the S protein

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).

In order to contribute to a better understanding of the results found in these phylogenetic analyses and to understand the relation among SARS-CoV-2 circulating in the region and vaccines, the complete amino

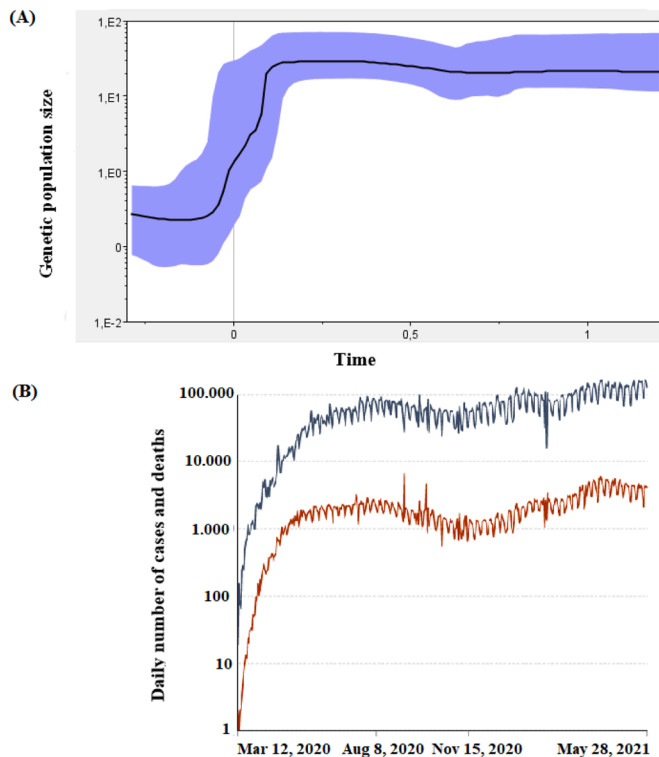


Fig. 2. Bayesian Skyline plot depicting the population history of SARS-CoV-2 strains isolated in South America. In (A) a Bayesian Skyline is shown. The thick solid black line represents the median estimate, and the blue area shows the 95% highest probability density (HPD) values (Bouckaert et al., 2019). Time is shown in the x-axis in years. In (B) the daily number of cases and deaths due to COVID-19 in South America are shown in blue and red, respectively. Time is shown in the x-axis as dates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

acid sequences of the S protein of 933 SARS-CoV-2 variants circulating in the South America region were aligned with corresponding sequences of reference strain Wuhan-Hu-1 (GenBank: NC_045512) and the S protein substitutions observed (see Fig. 3) (for sequences enrolled in these studies, see Supplementary Material Table 2).

From the 933 South American isolates enrolled in these studies 98.82% has a D614G substitution. This substitution is present in all the B.1 lineages established in South America (Martin, VanInsberghe and Koelle, 2021; Martin et al., 2021).

On the other hand, substitutions were found in all S protein domains. 148 amino acid positions of the S protein were found to have substitutions in different variants isolated in South America by comparison with reference strain Wuhan-Hu-1 (see Fig. 3).

In order to establish which of them were found significantly represented in the population studied, we employed SeqLogo (Thomsen and Nielsen, 2012) (see Fig. 3 and Supplementary Material Figure 2). Although S amino acid sequence resulted to be roughly conserved at population level in strains isolated in South America, we detected 11 sites in S protein with significant polymorphisms. Five of these polymorphisms were found in the N-terminal domain (substitutions L18F, T20N, P26S, D138I, R190S); three in the receptor binding domain (K417T, E484K and N501Y) and substitution H653Y in the S1 sub-unit of S protein. Substitution T1027I and V1176F were found in the S2 sub-unit, being the last one in the heptad repeat 2. Substitution L18F have been detected in SARS-CoV-2 clinical isolates, mainly from VOCs Alpha (B.1.1.7), Beta (B.1.351) and Gamma (P.1) and it is known to affect loop 1 of the antigenic supersite in the NTD domain (McCallum, De Marco and Lempp, 2021a; McCallum et al., 2021a). The finding that multiple circulating SARS-CoV-2 variants map to the NTD, several of

them in the antigenic supersite (site I), suggests that the NTD is subject to a strong selective pressure from the host humoral immune response (McCallum, Bassi and De, 2021b; McCallum et al., 2021b).

Due to the fact that the receptor-binding domain (RBD) of S protein is vital for virus attaching to the host receptor and triggering a conformational change in the protein that results in fusion with the host cell membrane, we mapped the substitutions sites where significantly polymorphisms was found in the RBD of S protein in the 3D structure of RBD complexed with its acceptor ACE2. The results of these studies are shown in Fig. 4. Substitutions K417T, E484K and N501Y map in the RBD region that interacts with ACE2 protein, in agreement with recent results (Winger and Caspari, 2021) (Winger and Caspari, 2021).

3.3. Nucleocapsid substitutions that increase infectivity and fitness were found in strains isolated in South America

In addition to the adaptive S protein substitutions, there are adaptive substitutions in other viral components that also contribute to the spread, fitness and infectivity of the virus. That is the case of the Nucleocapsid protein (N), where recent studies revealed that variants having substitutions R203K and G204R in N protein have replication advantages over the original virus, as well as show increased infectivity in human lung cells, contributing to an increased transmission and virulence of these variants (Wu et al., 2021). For these reasons, the same studies outlined above for S protein were performed for the N proteins of the same 933 strains. The results of these studies are shown in Fig. 5.

The crystal structure of N-protein revealed two distinct domains (Kang, Yang and Hong, 2020; Kang et al., 2020). One domain is present towards the N terminus of the protein and is also known as the RNA-binding domain (RBD). The C terminal side of the protein harbors a dimerization domain, which interacts with other N protein to make a dimer. Apart from these two domains there are three intrinsically disordered regions (IDRs) at N- and C-terminal ends as well as between the RBD and dimerization domain (Kang et al., 2020) (see Fig. 5). We have observed that R203K and G204R are the most frequently mutated residues of the N-protein. 89% of the SARS-CoV-2 variants isolated in South America and enrolled in these studies have these substitutions. Moreover, 70 substitutions were found by comparison with reference strain Wuhan-Hu-1 (see Fig. 5). 61% (43 out of 70) of these substitutions reside in the IDR regions. Recent studies have shown that positions S197, S202, R203 and G204 are important sites of phosphorylation by Aurora kinase A/B, GSK-3 as well as for its interactions with 14–3–3 protein (Tung and Limtung, 2020) (Tung and Limtung, 2020). Substitutions were found in two of these four positions. Interestingly, substitution S194L was observed in strains isolated in Peru and Colombia (see Fig. 5). Recent studies revealed an association of strains having this substitution and symptomatic patients (Barona-Gomez, Delaye and Diaz-Valenzuela, 2021; Barona-Gomez et al., 2021). Another recent study revealed that one of the most important B and T cell epitope of the N protein lies between residues 305–340 (Ahmed, Quadeer and McKay, 2020). Our study identified four substitutions in those positions (Fig. 5).

3.4. RNA-dependent RNA polymerase substitutions conferring epidemiological advance over Wuhan strains were found in strains circulating in South America

RNA-dependent RNA polymerase (RdRp or nsp12) is a key player in the synthesis of viral RNA (Ilmjärv, Abdul and Acosta-Gutiérrez, 2021; Ilmjärv et al., 2021). The structure of the SARS-CoV-2 nsp12 contains a cupped, right-handed RdRp domain linked to a nidovirus RdRp-associated nucleotidyl-transferase domain (NiRAN) via an interface domain (see Fig. 6). The RdRp domain adopts the conserved architecture of the viral polymerase family 4 and is composed of three domains: a fingers domain, a palm domain and a thumb domain (Fig. 6). The crystal structure of SARS-CoV-2 nsp12 in complex with its non-structural protein 7 and 8 (nsp7 and nsp8) co-factors underlines the

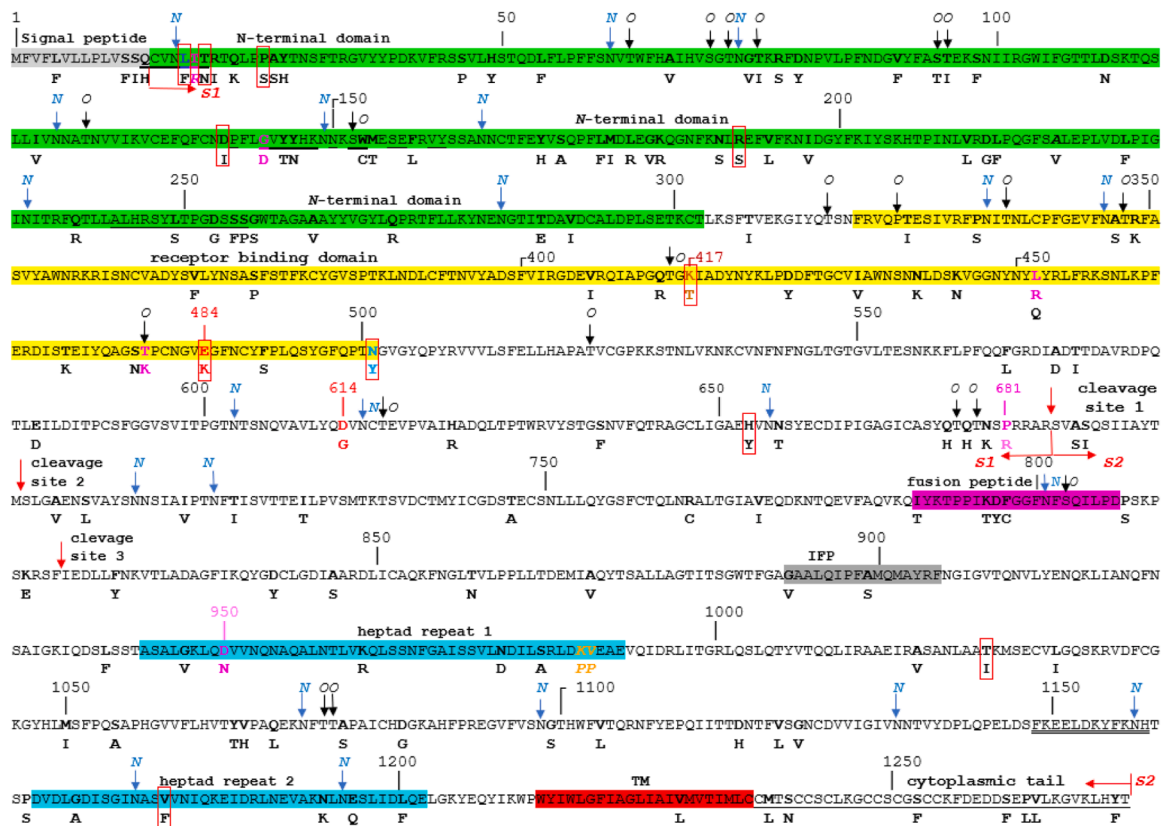


Fig. 3. Substitutions found in the Spike protein of SARS-CoV-2 viruses circulating in South America. The amino acid sequence of reference strain Wuhan-Hu-1 strain (GenBank: NC_045512) is shown. Domains of the Spike protein are shown at the top of the amino acid sequence. IFP, internal fusion peptide; TM, transmembrane domain (Xia, 2021). The cytoplasmic tail is shown underlined. Boundaries of S1 and S2 sub-units are shown in red and italics. Numbers at the top of the sequence denote amino acid position (relative to the reference strain). Substitution D614G, present in all strains B.1 strains and substitution E484K, present in all VOI Zeta (P.2) as well of VOCs Beta and Gamma are shown in red. Substitution N501Y, found in all VOC Gama (P.1) strains, is shown in bold blue. Substitution K417T, characteristic of VOC Beta (B.1.351) is shown in brown. Substitutions L452R, T478K, P681R and D950N characteristic of VOC Delta (B.1.1617.2) are shown in fuchsia. Positions where substitutions were found in strains isolated in South America are highlighted in bold and the corresponding substitution is shown below each position. Proline substitutions in Pfizer/BioTech and Moderna vaccine to stabilize the Spike protein at the prefusion conformation are shown in orange (Xia, 2021; Xia, 2021). Positions in the antigenic supersite of the N-terminal domain (NTD) are underlined. The N-glycosylation sites are shown by a blue arrow while the O-glycosylation sites are shown by a black arrow and they are indicated N or O next to the arrow. Sites where significant polymorphic sites were found are shown in red squares. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

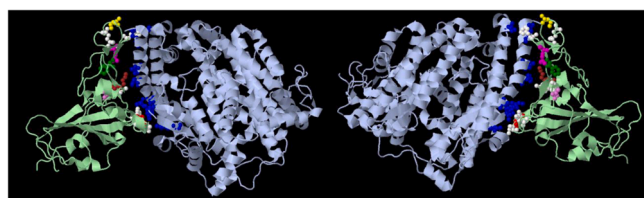


Fig. 4. 3D structure of Spike protein receptor-binding domain complexed with its receptor ACE2. Receptor-binding domain of Spike protein is shown in light green, whereas ACE2 protein is shown in gray. ACE2 amino acids interacting with Spike receptor-binding domain (Gln24, Asp30, His34, Tyr41, Gln42, Met82, Lys353 and Arg357) (according to Yi et al., 2020) are shown in blue. Substitutions K417N, L452R, T478K, E484K, F490S and N501Y are shown in brown, rose, yellow, fuchsia, green and red, respectively. Positions of other amino acids in the Spike receptor-binding domain also shown to interact with ACE2 (Gln498, Thr500, Tyr453, Gln474 and Phe 486) (also according to Yan et al., 2020) are shown in white. The 3D structure was obtained from the PDB Database under accession number 6LZG. Two views rotated 180° in the X-axis are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

central role these nsp in the replication and transcription of the virus (Gao, Yan and Huang, 2020; Gao et al., 2020). Recent studies have revealed that substitution P323L in RdRp together with D614G in the Spike protein have shown that G614/L323 variants are epidemiological highly successful and replaced the original D614/P323 variants (Ilmjärvi, Abdul and Acosta-Gutiérrez, 2021; Ilmjärvi et al., 2021). For these reasons, the same studies performed for S and N proteins were performed for the RdRp proteins for the same strains enrolled in S and N studies. The results of these studies are shown in Fig. 6. We have found that 98% of SARS-CoV-2 strains circulating in South America and enrolled in these studies carried the P323L in the RdRp protein (see Fig. 6). This substitution is located on the interface domain. We have observed 14 substitutions in the NiRAN domain, 7 substitutions in the interface domain, and 12, 9, and 5 substitutions in finger, palm and thumb domains of the RdRp proteins in strains isolated in South America (Fig. 6). The RdRp protein were found to be significantly conserved among variants and not polymorphic sites were observed among the RdRp of strains isolated in South America and enrolled in these studies.

3.5. A significant bias in nucleotide frequencies, codon and amino acid usage among S proteins

PCA is a statistic technique used to describe a collection of data in terms of variables (components) no correlated among themselves. The

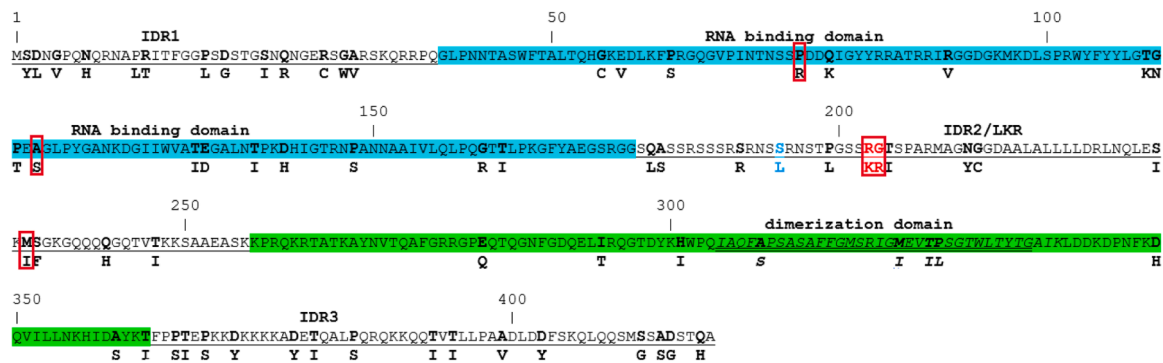


Fig. 5. Substitutions found in the Nucleocapsid (N) protein of SARS-CoV-2 viruses circulating in South America. The amino acid sequence of reference strain Wuhan-Hu-1 strain (GenBank: NC_045512) is shown. Domains of the N protein are shown at the top of the amino acid sequence. Intrinsically disordered regions (IDR) are shown underlined. LKR, liker region. Residues conforming B/T cell epitope are shown in italics and double underlined. Substitutions R203K and G204R, found in 89% of the strains enrolled in these studies are shown in red. Substitution S194S is shown in blue. Numbers at the top of the sequence denote amino acid position (relative to the reference strain). Sites where significant polymorphic sites were found are shown in red squares.

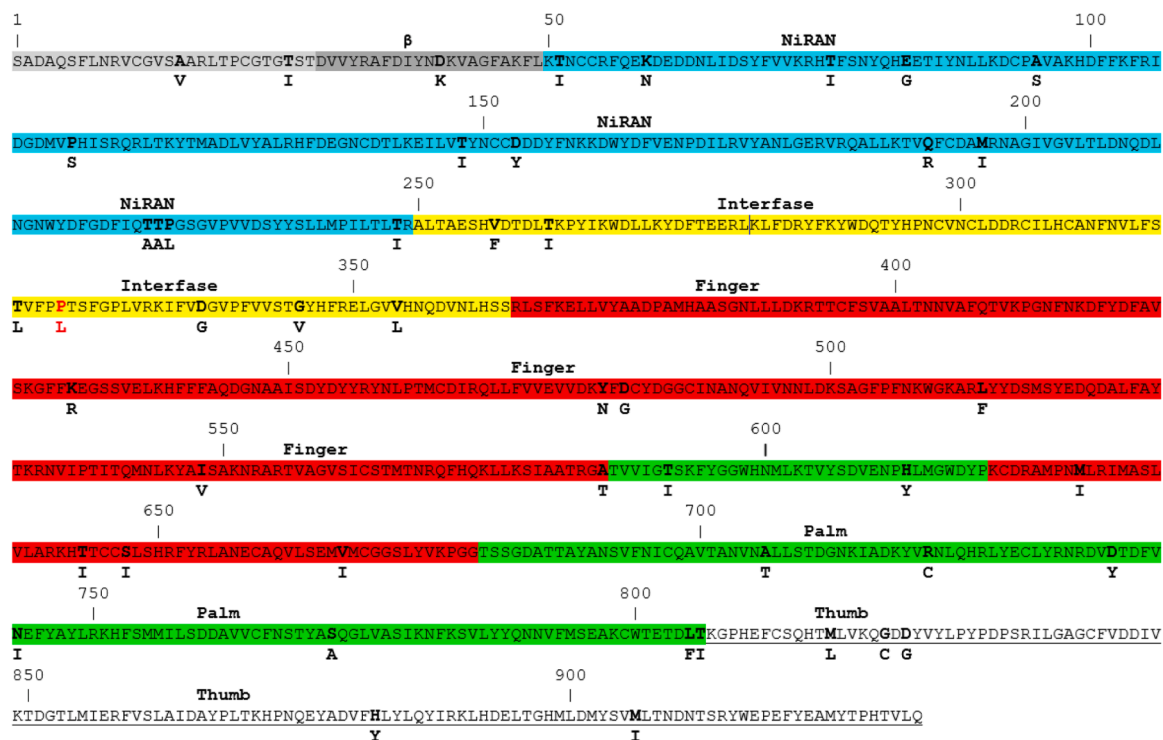


Fig. 6. Substitutions found in the RNA-dependent RNA polymerase (RdRp) protein of SARS-CoV-2 viruses circulating in South America. The amino acid sequence of reference strain Wuhan-Hu-1 strain (GenBank: NC_045512) is shown. Domains of the RdRp protein are shown at the top of the amino acid sequence. Numbers at the top of the sequence denote amino acid position (relative to the reference strain). Substitution P323L, found in 98% of the strains enrolled in these studies is shown in red. NiRAN domain is shown in cyan. Interfase domain is shown in yellow. RdRp domains Finger, Palm and Thumb are shown in red, blue and underlined, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

components are ordered according to the original variation that they describe, and for this reason it is a very useful technique to reduce the dimensionality of the collection of data while preserving as much of the data's variation as possible.

In order to gain insight into the trends of evolution of the S protein, a composition analysis among the 933 S proteins enrolled in these studies was performed. For this purpose, the nucleotide frequencies for first, second and third codon positions were established for S genes from 933 SARS-CoV-2 variants circulating in South America and PCA was performed (see Supplementary Material Fig. 3).

A significant bias in nucleotide frequencies was found in the S gene sequences from SARS-CoV-2 variants enrolled in these studies. In fact, PC1 component (that accounts for the 85% of the total variation

observed) has a strong positive correlation with A and U frequencies at the third codon position and a strong negative correlation with C and G frequencies at that position (see Supplementary Material Fig. 3). This bias in preferences for A and U ended codons accounts for highly preferred CCU, CUU, GAU, GGU, GCU, UUU and CGU codons, and an underrepresentation use of UCG, CUG, CCG, AGC, CGC and GCG, AGC and GGG codons (see Table 2).

To gain insight into the trends of the variation observed among VOCs circulating in the South American region, codon usage frequencies of 86 S genes from SARS-CoV-2 strains belonging to VOCs Alpha, Beta, Gamma and Delta were determined and PCA and Heatmap analysis was performed (for strains included in these analyses, see Supplementary Material Table 2) (see Fig. 7). This analysis revealed significant

Table 2
Codon usage in SARS-CoV-2 S genes from strains isolated in South America^a.

AA	Cod	Frequency	AA	Cod	Frequency	AA	Cod	Frequency	AA	Cod	Frequency
Phe	UUU	1.54	Ser	UCU	2.25	Tyr	UAU	1.49	Cys	UGU	1.40
	UUC	0.46		UCC	0.72		UAC	0.51		UGC	0.60
Leu	UUA	1.56		UCA	1.57	TER	UAA	***	TER	UGA	***
	UUG	1.12		UCG	0.12		UAG	***	Trp	UGG	1.00
	CUU	1.98	Pro	CCU	1.99	His	CAU	1.51	Arg	CGU	1.28
	CUC	0.67		CCC	0.28		CAC	0.49		CGC	0.15
	CUA	0.50		CCA	1.73	Gln	CAA	1.49		CGA	0.00
	CUG	0.17		CCG	0.00		CAG	0.51		CGG	0.29
Ile	AUU	1.74	Thr	ACU	1.81	Asn	AAU	1.22	Ser	AGU	1.04
	AUC	0.55		ACC	0.40		AAC	0.78		AGC	0.30
	AUA	0.71		ACA	1.65	Lys	AAA	1.26	Arg	AGA	2.87
Met	AUG	1.00		ACG	0.13		AAG	0.74		AGG	1.41
Val	GUU	1.97	Ala	GCU	2.12	Asp	GAU	1.37	Gly	GGU	2.31
	GUC	0.87		GCC	0.41		GAC	0.63		GGC	0.72
	GUA	0.62		GCA	1.37	Glu	GAA	1.41		GGA	0.82
	GUG	0.54		GCG	0.10		GAG	0.59		GGG	0.15

^a Average frequencies in 1:187,463 codons. AA, amino acid; Cod, codons; TER, termination codons. Preferred codons ($\Delta \geq 0.30$) are shown in bold. Underrepresented codons are shown in italics.

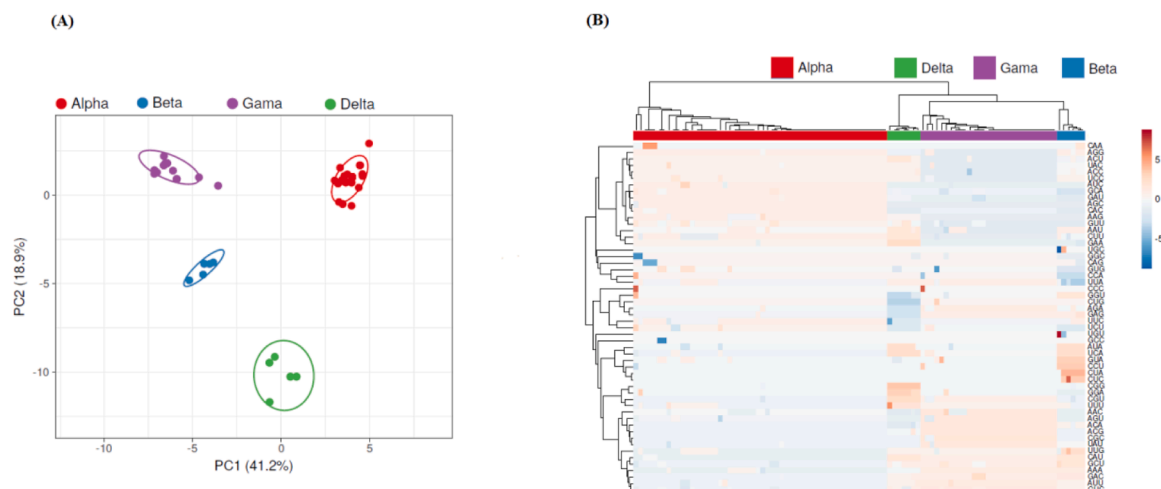


Fig. 7. PCA of codon usage in Spike proteins from VOC's SARS-CoV-2 strains. In (A) the position of the Spike proteins in the plane conformed by the first two major components of PCA is shown. SVD was used to calculate principal components and unit variance was applied. The proportion of variance explained 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 top of the figure. $N = 86$ 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 at the top of the figure. Both rows and columns are clustered using correlation distance and average linkage.

differences in codon usage among the S genes of SARS-CoV-2 VOCs circulating in South America, which can be linked to SARS-CoV-2 genotypes (see Fig. 7A and B). When the same analysis is done for amino acid usage frequencies, the same results are also found (see Supplementary Material Figure 4).

Since codon usage by its very nature is multivariate, it is necessary to analyze the data using different and complementary approaches. To confirm the results outlined above, we performed a COA analysis on the RSCU values for the same 86 S genes from VOCs strains and examined the distribution of the genes along the plane determined by the first two principal axes of COA. The results of these studies are shown in Supplementary Material Fig. 5. The distribution of the genes in the plane defined by the first two major axes of COA showed the same results, revealing that different S genes are located at different places, which again are linked to VOCs genotypes (see Supplementary Material Fig. 5). These results suggest codon usage plays an important role in shaping the evolution of Spike proteins from SARS-CoV-2 VOC's. Moreover, when codon usage frequencies are analyzed using 494 S genes from 10 different genotypes found to circulate in South America, a relation among genotype and codon usage is also found (see Supplementary

Material Fig. 6).

4. Discussion

SARS-CoV-2 has spread across the world, causing a health threat of international concern. As the virus circulation becomes widespread, phylodynamic analyses can give insight into how the virus spreads both spatially and temporally. Moreover, viruses from a given region can be placed in the context of those circulating globally, allowing for the number of independent virus introductions into a region to be estimated in these analyses (Rambaut et al., 2020; Rambaut et al., 2020). Phylodynamic analyses can also be extremely useful to study viral adaptation, a particular concern since SARS-CoV-2 has recently spilled to humans.

The results of these studies suggest that SARS-CoV-2 variants circulating in South America evolved from ancestors that existed around November 26th, 2019 (Table 2). Similar studies carried out at the beginning of the pandemic on SARS-CoV-2 strains circulating in the Hubei province of China trace the case index to November 9th, 2019 (Pekar et al., 2021; Pekar et al., 2020). These results revealed a rapid transmission of SARS-CoV-2 strains from China to South America.

Emerging SARS-CoV-2 variants with an amino acid substitution at position 614 of the S protein (D614G) have shown to be more transmissible (Hou, Chiba and Halfmann, 2020; Hou et al., 2020; Faria et al., 2021; Volz, Hill and McCrone, 2021; Volz et al., 2021). Although G614 resides at a fair distance from the RBD, it affects the ACE2 binding site through an allosteric link with T500 (Zhang et al., 2021; Zhang et al., 2021; Omotuyi et al., 2021) Omotuyi et al., 2021). This is in agreement with the results of this work, since most of the lineages found to circulate in the South American region where B.1 lineages who carry this substitution (see Figs. 1 and Supplementary Material Fig. 1). By the same token, most of the variants enrolled in these studies also carried substitutions R203K and G204R in the N protein (see Fig. 5), providing an enhanced replication advance by comparison with strains isolated early in the pandemic, as well as enhanced infectivity and disease severity in hamster model (Wu et al., 2021). Moreover, 98% of the strains enrolled in these studies carried the P232L substitution in the RdRp. Taking these results together, this may help to explain, at least in part, the sharp increase in population size observed at the beginning of the pandemic in the South American region (see Fig.2).

An emergence of Zeta (P.2) variants from a parental cluster of B.1.28 strains was observed in these studies (see Table 1). Besides the D614G, these strains also carry the substitution E484K in the receptor binding domain (RBD) of S protein (see also Fig. 3). Recent studies suggest that the substitution E484K permits virus variants to be selected as an escape mutation in the presence of neutralizing antibodies or plasma from immune humans in vitro (Weisblum et al., 2020; Weisblum et al., 2020). Interestingly, studies carried out in England and Wales suggest the accumulation of substitution E484K in the Alpha (B.1.1.7) background is the result of the vaccination program (Collier et al., 2021; Collier et al., 2021). This substitution, which is situated at the RBD interface with the receptor ACE2 (see Figs. 3 and 4), may also play an important role in neutralization escape of VOI Zeta as well of VOCs Beta and Gamma, that share this substitution (Salleh et al., 2021).

This same parental cluster B.1.28 give rise to the emergence of VOC Gamma (P.1) (see Supplementary Material Fig. 1). This variant was first observed in Manaus, Brazil, in December of 2020, where a sharp increase in the total COVID-19 infections, followed by an increase in the number of hospital admissions was observed (Faria et al., 2021; Faria et al., 2021; Candido, Claro and de, 2020; Candido et al., 2020). This variant shares with VOC Beta (B.1.351) three important substitutions in the RBD of the S protein: K417T, E484K and N501Y (see Fig. 3). These substitutions have been shown to increase the binding affinity of the S protein to its receptor ACE2, particularly substitution N501Y significantly contribute to this increase in binding affinity (Luan, Wang and Huynh, 2021; Luan et al., 2021; Ali, Kasry and Amin, 2021; Ali et al., 2021). Moreover, recent studies revealed that N439K viruses have similar in vitro replication fitness as compared to wild type, while at the same time N439K substitutions confers resistance against several neutralizing monoclonal antibodies (Thomson et al., 2021; Thomson et al., 2021). This highlights the importance of for molecular surveillance in all regions of the world to guide development and usage of vaccines and therapeutics.

VOI Mu was detected in Colombia (see Fig. 1). This VOI was isolated for the first time in January 11th, 2021. VOI Mu have substitutions in the S protein, as T95I, Y144S, Y145N in the N-terminal domain; R346K, E484K and N501Y in the receptor-binding domain and D614G, P681H and D950N in other regions (Uriu, Kimura and Shirakawam, 2021; Uriu et al., 2021). In these studies, Colombian VOI Mu variants having substitution Y144T were also observed (see Fig. 3). Several of these substitutions have been identified in other VOCs: e.g., E484K in Beta and Gamma, N501Y in Alpha and Beta, P681H in Alpha, and D950N in Delta. Virus neutralization studies, performed with the use of serum samples obtained from persons who had recovered from Covid-19 and who were infected early in the pandemic (April through September 2020), showed that the VOI Mu was 10.6 times as resistant to neutralization as the B.1 lineage parental virus (Uriu, Kimura and Shirakawam, 2021; Uriu et al.,

2021). Similar studies using sera from persons who had received the BNT162b2 vaccine showed that the VOI Mu variant was 9.1 as resistant as the parental virus (Uriu, Kimura and Shirakawam, 2021; Uriu et al., 2021). These studies highlight the importance of further studies regarding this variant.

Taking these results together it is possible to observe that most of the SARS-CoV-2 viruses circulating in South America have substitutions in the S and N proteins that confer high transmissibility and/or immune resistance (see Figs. 3 and 5). Therefore, it is extremely important to aim for the best possible vaccination campaign in South American in order to enhance protection against these and newly emerging SARS-CoV-2 variants. Vaccination campaigns in South American countries began in early 2021 (February-March) and most of the countries enrolled in these studies had less than 10% coverage of their population by the period covered by these studies (Dong et al., 2020).

Moreover, 148 amino acid positions in the S protein were found to have substitutions in variants from South America in the period covered by these studies (March 10th, 2020 to May 28th, 2021) (see Fig. 3). Continuous molecular surveillance of SARS-CoV-2 will be necessary to detect new variants of the virus with clinical relevance. This is extremely important to improve programs to control the virus (Flores-Alanis, Cruz-Rangel and Rodríguez-Gómez, 2021; Flores et al., 2021).

The SARS-CoV-2 S protein is covered by a shield of N-linked and O-linked glycans (Lo Presti, Rezza and Stefanelli, 2020; Lo Presti et al., 2020; Shajahan, Supekar, Gleinich and Azadi, 2020; Shajahan et al., 2020). An important fact in the development of effective subunit vaccines is the characterization of the glycosylation of key viral proteins, since they play a crucial role in immune recognition affecting vaccine designs. Glycosylation of viral surface proteins is extremely important for immune shielding and altering positions where glycosylation sites occur is a well-known immune evasion mechanism in viruses (Walls et al., 2016; Walls et al., 2016). From the predicted putative N- or O-glycosylated sites found in these studies (21 N- and 18 O- glycosylated sites), no substitutions were found in N- glycosylation sites and 4 O-glycosylation sites were found to have substitutions (Fig. 3). The results of these studies suggest that glycosylation sites are roughly conserved among S protein from SARS-CoV-2 viruses. More studies will be needed in order to address the effect of substitutions in glycosylation sites of this protein.

In these studies, a biased nucleotide composition was found in the S proteins of SARS-CoV-2 variants circulating in South America (Fig. 5). While S genes composition analysis revealed a positive correlation of U and A at the third codon positions, a negative correlation was found for C and G at these positions (see Fig. 5). This is in agreement with previous work carried out in other coronaviruses, where A/G bias is a relatively stable property shared in the family, while the C/U bias differs significantly per virus type (Berkhout and van Hemert, 2015; Berkout and Hemert, 2015). These biases also have a major influence on derived parameters as codon usage (see Table 2). PCA and Heatmap analysis revealed correlation among codon usage and genotypes in the S protein from VOC's strains (see Fig. 5). Interestingly, these results also demonstrate that S genes have suitable genetic information for clear assignment of emerging VOCs to its specific genotypes (see Fig. 6).

5. Conclusions

The results of these studies revealed that at least 62 different genotypes of SARS-CoV-2 were found to circulate in South America. Most of the variants circulating in this region belongs to B.1 genotypes. From the 933 South American isolates enrolled in these studies 98.82% has a D614G substitution in the spike protein, while 89% of them have substitutions R203K and G204R in the nucleocapsid protein and 98% have the substitution P323L. All VOCs (Alpha, Beta, Gamma and Delta) co-circulate in the region, together with VOI Lambda, VOI Mu and Zeta. Substitutions were found in all S protein domains in different variants from South America by comparison with reference strain Wuhan-Hu-1.

11 polymorphic sites were detected in the S protein of SARS-CoV-2 sequences from strains circulating in South America. Particularly, polymorphisms were found in the receptor binding domain with substitutions in positions that interact with ACE2 cellular receptor. Significant trends of variation in codon and amino acid usage were observed among VOCs circulating in the South American region. This variation can be linked to VOCs genotypes. These results of these studies suggest that codon usage plays an important role in shaping the evolution of S proteins from SARS-CoV-2 variants.

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

Paula Perbolianachis: Data curation, Visualization, Investigation. **Diego Ferla:** . **Rodrigo Arce:** Data curation, Visualization, Investigation. **Irene Ferreira:** Data curation, Visualization, Investigation. **Alicia Costáble:** Data curation, Visualization, Investigation. **Mercedes Paz:** Data curation, Visualization, Investigation. **Diego Simón:** . **Pilar Moreno:** Writing – review & editing. **Juan Cristina:** Conceptualization, Methodology, 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 GISRID. We thank Dr. Gonzalo Moratorio for critical reading of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2022.198688.

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>.
- Ali, F., Kasry, A., Amin, M., 2021. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. *Med. Drug Discov.* 10, 100086 <https://doi.org/10.1016/j.medidd.2021.100086>.
- Ahmed, S.F., Quadeer, A.A., McKay, M.R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* 12, 254. <https://doi.org/10.3390/v12030254>.
- Barona-Gomez, F., Delaye, L., Diaz-Valenzuela, E., et al., 2021. Phylogenomics and population genomics of SARS-CoV-2 in Mexico during the pre-vaccination stage reveals variants of interest B.1.1.28.4 and B.1.1.222 or B.1.1.519 and the nucleocapsid mutation S194L associated with symptoms. *Microb Genom* 7, 000684. <https://doi.org/10.1099/mgen.0.000684>.
- Belouzard, S., Chu, V.C., Whittaker, G.R., 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proc. Natl. Acad. Sci. USA* 106, 5871–5876. <https://doi.org/10.1073/pnas.0809524106>.
- Berkhout, B., van Hemert, F., 2015. On the biased nucleotide composition of the human coronavirus RNA genome. *Virus Res* 202, 41–47. <https://doi.org/10.1016/j.virusres.2014.11.031>.
- 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>.
- Candido, D.S., Claro, I.M., de Jesus, J.G., et al., 2020. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science* 369, 1255–1260. <https://doi.org/10.1126/science.abd2161>.
- Chakraborty, C., Bhattacharya, M., Sharma, A.R., 2021 Jun 27. Present variants of concern and variants of interest of severe acute respiratory syndrome coronavirus 2: their significant mutations in S-glycoprotein, infectivity, re-infectivity, immune escape and vaccines activity. *Rev Med Virol* e2270. <https://doi.org/10.1002/rmv.2270>.
- Chan, J.F., Kok, K.H., Zhu, Z., et al., 2020. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes Infectec.* 9, 21–236. <https://doi.org/10.1080/22221751.2020.1719902>.
- Cerutti, G., Guo, Y., Zhou, T., et al., 2021. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* 29 (5), 819–833.e7. <https://doi.org/10.1016/j.chom.2021.03.005>.
- 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>.
- Collier, D.A., De Marco, A., Ferreira, I.A., Meng, B., Datt, R.P., Walls, A.C., Kemp, S.A., et al., 2021. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* 593, 136–141. <https://doi.org/10.1038/s41586-021-03412-7>.
- 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.abg3055>.
- Dong, E., Du, H., Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* 20 (5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22, 1185–1192. <https://doi.org/10.1093/molbev/msi103>.
- Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D., Mishra, S., Crispim, M. A., Sales, F.C., Hawryluk, I., McCrone, J.T., et al., 2021. Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372, 815–821. <https://doi.org/10.1126/science.abh2644>.
- Flores-Alanis, A., Cruz-Rangel, A., Rodríguez-Gómez, F., et al., 2021. Molecular epidemiology surveillance of SARS-CoV-2: mutations and genetic diversity one year after emerging. *Pathogens* 10, 184. <https://doi.org/10.3390/pathogens10020184>.
- Gao, Y., Yan, L., Huang, Y., et al., 2020. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368, 779–782. <https://doi.org/10.1126/science.abb7498>.
- Gorbalenya, A.E., Baker, S.C., Baric, R.S., et al., 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. <https://doi.org/10.1038/s41564-020-0695-z>.
- Greenacre, M., 1994. *Theory and Applications of Correspondence Analysis*. Academic Press, London.
- Gupta, R., Brunak, S., 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* 310–322. <https://service.s.healthtech.dtu.dk/service.php?NetNGlyc-1.0>.
- Hou, Y.X., Chiba, S., Halfmann, P., et al., 2020. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* 370, 1464–1468. <https://doi.org/10.1126/science.abe8499>.
- Ilmjärvi, S., Abdul, F., Acosta-Gutiérrez, S., et al., 2021. Concurrent mutations in RNA-dependent RNA polymerase and spike protein emerged as the epidemiologically most successful SARS-CoV-2 variant. *Sci Rep.* 11, 13705. <https://doi.org/10.1038/s41598-021-91662-w>.
- Kang, S., Yang, M., Hong, Z., et al., 2020. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B* 10, 1228–1238. <https://doi.org/10.1016/j.apsb.2020.04.009>.
- Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* 4, 1160–1166. <https://doi.org/10.1093/bib/bbx108>.
- Korber, B., Fischer, W.M., Gnanakaran, S., et al., 2020. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182, 812–827. <https://doi.org/10.1016/j.cell.2020.06.043>.
- Kuhner, M.K., 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol* 24, 86–93. <https://doi.org/10.1016/j.tree.2008.09.007>.
- Leach, A., Ilca, F.T., Akbar, Z., Ferrari, M., Bentley, E.M., Mattiuzzo, G., Onuoha, S., Miller, A., Ali, H., Rabbitts, T.H., 2021. A tetrameric ACE2 protein broadly neutralizes SARS-CoV-2 spike variants of concern with elevated potency. *Antiviral Res* 194, 105147. <https://doi.org/10.1016/j.antiviral.2021.105147>.
- Leung, K.S., Ng, T.T., Wu, A.K., Yau, M.C., Lao, H.Y., et al., 2021. Territory wide study of early coronavirus disease outbreak, Hong Kong, China. *Emerg. Infect. Dis.* 27, 196–204. <https://doi.org/10.3201/eid2701.201543>.
- Li, Q., Guan, X., Wu, P., et al., 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* 26, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>.
- Lo Presti, A., Rezza, G., Stefanelli, P., 2020. Selective pressure on SARS-CoV-2 protein coding genes and glycosylation site prediction. *Heliyon*. 6, e05001. <https://doi.org/10.1016/j.heliyon.2020.e05001>.
- Luan, B., Wang, H., Huynh, T., 2021. Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. *FEBS Lett.* <https://doi.org/10.1002/1873-3468.14076>.

- Martin, M.A., VanInsberghe, D., Koelle, K., 2021. Insights from SARS-CoV-2 sequences. *Science* 371, 466–467. <https://doi.org/10.1126/science.abf3995>.
- McCallum, M., De Marco, A., Lempp, F.A., et al., 2021a. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* 184, 2332–2347.e16. <https://doi.org/10.1016/j.cell.2021.03.028>.
- McCallum, M., Bassi, J., De A., Marco, et al., 2021b. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* 373, 648–654. <https://doi.org/10.1126/science.abi7994>.
- Metsalu, T., Vilo, J., 2015. Clustvis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* 43 (W1), W566–W570. <https://doi.org/10.1093/nar/gkv468>.
- Navascués, M., Leblois, R., Burgarella, C., 2017. Demographic inference through approximate-Bayesian-computation skyline plots. *Peer J* 5, e3530. <https://doi.org/10.7717/peerj.3530>.
- Omotuyi, I.O., Nash, O., Ajiboye, O.B., Iwegbulam, C.G., Oyinloye, E.B., Oyediji, O.A., Kashim, Z.A., Okaiyeto, K., 2021. Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-COV-2. *J. Comput. Chem.* 41, 2158–2161. <https://doi.org/10.1002/jcc.26383>, 2020.
- Pekar, J., Worobey, M., Moshiri, N., Scheffler, K., Wertheim, J.O., 2021. Timing the SARS-CoV-2 index case in Hubei province. *Science* 372, 412–417. <https://doi.org/10.1126/science.abf8003>.
- Piccoli, L., Park, Y.J., Tortorici, M.A., et al., 2020. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* 183, 1024–1042.e21. <https://doi.org/10.1016/j.cell.2020.09.037>.
- Rambaut, A., Holmes, E.C., O’Toole, Á., Hill, V., McCrone, J.T., Ruis, C., du Plessis, L., Pybus, O.G., 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>.
- Samavati, L., Uhal, B.D., 2020. ACE2, much more than just a receptor for SARS-COV-2. *Front. Cell. Infect. Microbiol.* 10, 317. <https://doi.org/10.3389/fcimb.2020.00317>.
- Salleh, M.Z., Derrick, J.P., Deris, Z.Z., 2021. Structural evaluation of the spike glycoprotein variants on SARS-CoV-2 transmission and immune evasion. *Int. J. Mol. Sci.* 22, 7425. <https://doi.org/10.3390/ijms22147425>.
- Shajahan, A., Supekar, N.T., Gleinich, A.S., Azadi, P., 2020. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* 30, 981–988. <https://doi.org/10.1093/glycob/cwaa042>.
- Sharp, P.M., Li, W.H., 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24, 28–38. <https://doi.org/10.1007/BF02099948>.
- Steenfot, C., Vakhrushev, S.Y., Joshi, H.J., Kong, Y., Vester-Christensen, M.B., et al., 2013. Precision mapping of the human O-GalNAc glycoproteome through Simple Cell technology. *EMBO J* 32, 1478–1488. <https://doi.org/10.1038/emboj.2013.79>.
- Trifinopoulos, J., Nguyen, L.T., von Haeseler, A., Minh, B.Q., 2016. *Nucl. Acids Res.* 44 (W1), W232–W235. <https://doi.org/10.1093/nar/gkw256>.
- Thomsen, M.C.F., Nielsen, M., 2012. Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* 40, W281–W287. <https://doi.org/10.1093/nar/gks469>.
- Thomson, E.C., Rosen, L.E., Shepherd, J.G., Spreafico, R., da Silva, F., Wojcechowskyj, J. A., Davis, C., et al., 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* 184, 1171–1187.e20. <https://doi.org/10.1016/j.cell.2021.01.037>.
- Tung, H.Y.L., Limtung, P., 2020. Mutations in the phosphorylation sites of SARS-CoV-2 encoded nucleocapsid protein and structure model of sequestration by protein 14-3-3. *Biochem. Biophys. Res. Commun.* 532, 134–138. <https://doi.org/10.1016/j.bbrc.2020.08.024>.
- Uriu, K., Kimura, I., Shirakawa, K., et al., 2021. Neutralization of the SARS-CoV-2 Mu Variant by Convalescent and Vaccine Serum. *N. Engl. J. Med.* 385, 2397–2399. <https://doi.org/10.1056/NEJMc2114706>.
- van Dorp, L., Acman, M., Richard, D., et al., 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* 283, 104351. <https://doi.org/10.1016/j.meegid.2020.104351>.
- Volz, E., Hill, V., McCrone, J.T., et al., 2021. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* 1, 64–75. <https://doi.org/10.1016/j.cell.2020.11.020>.
- Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 Spike glycoprotein. *Cell* 181, 281–292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>.
- Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., DiMaio, F., Bosch, B. J., Veesler, D., 2016. Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* 23, 899–905. <https://doi.org/10.1038/nsmb.3293>.
- Wang, L., Cheng, G., 2021. Sequence analysis of the Emerging Sars-CoV-2 Variant Omicron in South Africa. *J. Med. Virol.* <https://doi.org/10.1002/jmv.27516>, 2021 Dec 12.
- Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finklin, S., Nussenzweig, M.C., 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature*. <https://doi.org/10.1038/s41586-021-03324-6>.
- Weinreich, D.M., Sivapalasingam, S., Norton, T., Ali, S., Gao, H., Bhore, R., Trial, I., 2021. REGN-COV2, a neutralizing antibody cocktail, in outpatients with covid-19. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2035002>.
- Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J.C., Muecksch, F., et al., 2020. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 9, e61312. <https://doi.org/10.7554/eLife.61312>.
- Winger, A., Caspari, T., 2021. The Spike of Concern-The Novel Variants of SARS-CoV-2. *Viruses* 13, 1002. <https://doi.org/10.3390/v13061002>.
- World Health Organization. 2020a. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). <https://www.who.int> (accessed August 12th, 2021).
- World Health Organization. 2020b. Coronavirus disease 2019 (COVID-19) Weekly epidemiological update on COVID-19 - 14 December 2021). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports> (accessed December 14th, 2021).
- Wu, H., Xing, N., Meng, K., et al., 2021. Nucleocapsid mutation R203K/G294R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host Microbe* 29, 1788–1801. <https://doi.org/10.1016/j.chom.2021.11.005>.
- Xia, X., 2021. Domains and functions of Spike protein in SARS-Cov-2 in the context of vaccine design. *viruses* 13, 109. <https://doi.org/10.3390/v13010109/v13010109>.
- 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.e5. <https://doi.org/10.1016/j.molcel.2020.04.022>.
- Yi, C., Sun, X., Jye, J.I., Ding, L., Liu, M., et al., 2020. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell. Mol. Immunol.* 17, 621–630. <https://doi.org/10.1038/s41423-020-0458-z>.
- Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S.M., Walsh, R.M., Rits-Volloch, S., Zhu, H., Woosley, A.N., et al., 2021. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 372, 525–530. <https://doi.org/10.1126/science.abf2303>, 2021.



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



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).

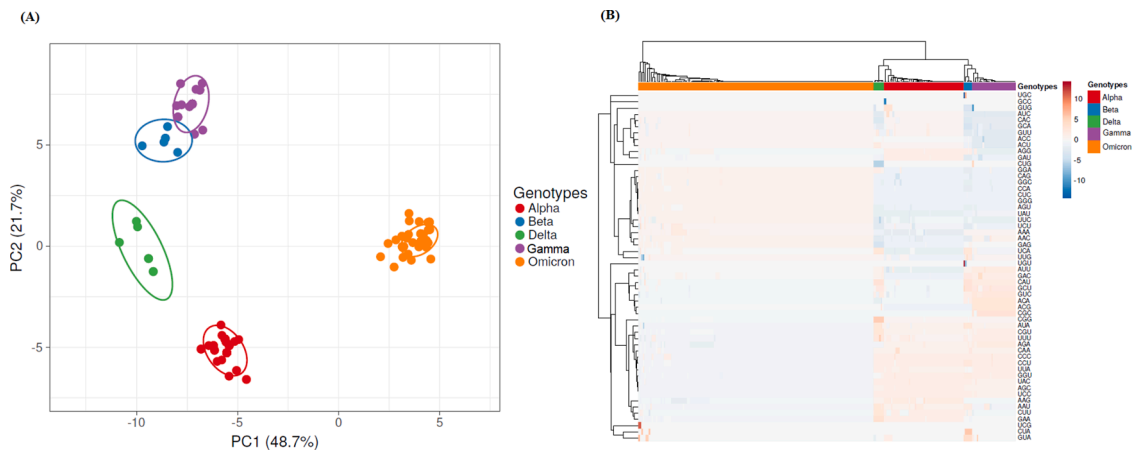


Fig. 1. PCA of codon usage in Spike proteins from SARS-CoV-2 VOCs strains. In (A) the position of the Spike proteins in the plane conformed by the first two major components of PCA is shown. SVD was used to calculate principal components and unit variance was applied. The proportion of variance explained 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.

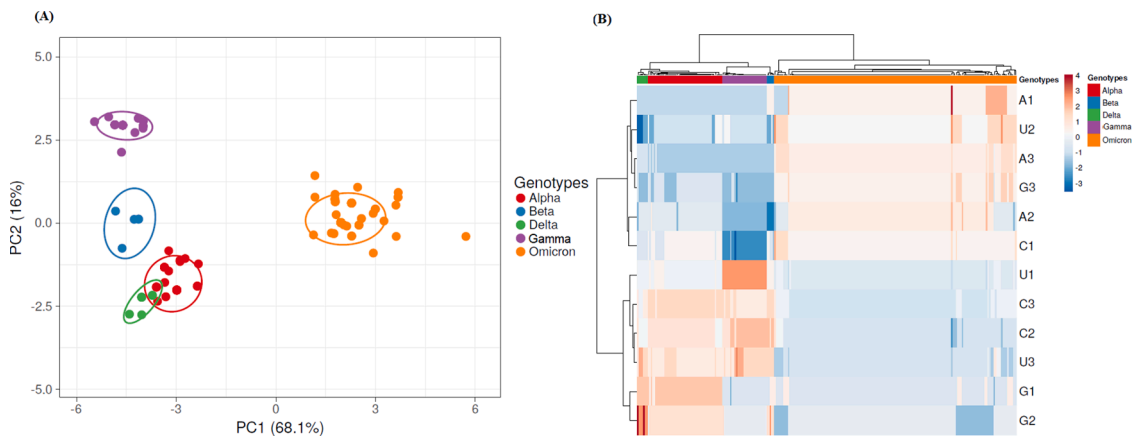


Fig. 2. PCA of nucleotide composition in Spike proteins from SARS-CoV-2 VOCs. In (A) a PCA analysis of nucleotide frequencies for first, second and third codon positions in S protein from SARS-CoV-2 strains is shown. SVD was used to calculate principal components and unit variance was applied. The proportion 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.

Phylogenetic 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 ^b	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.
^eIn all cases, the mean values are shown.
^fHPD, high probability density values.
^gESS, 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 sequencing capacities is currently unknown (Kupferschmidk, 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 Denmark (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.

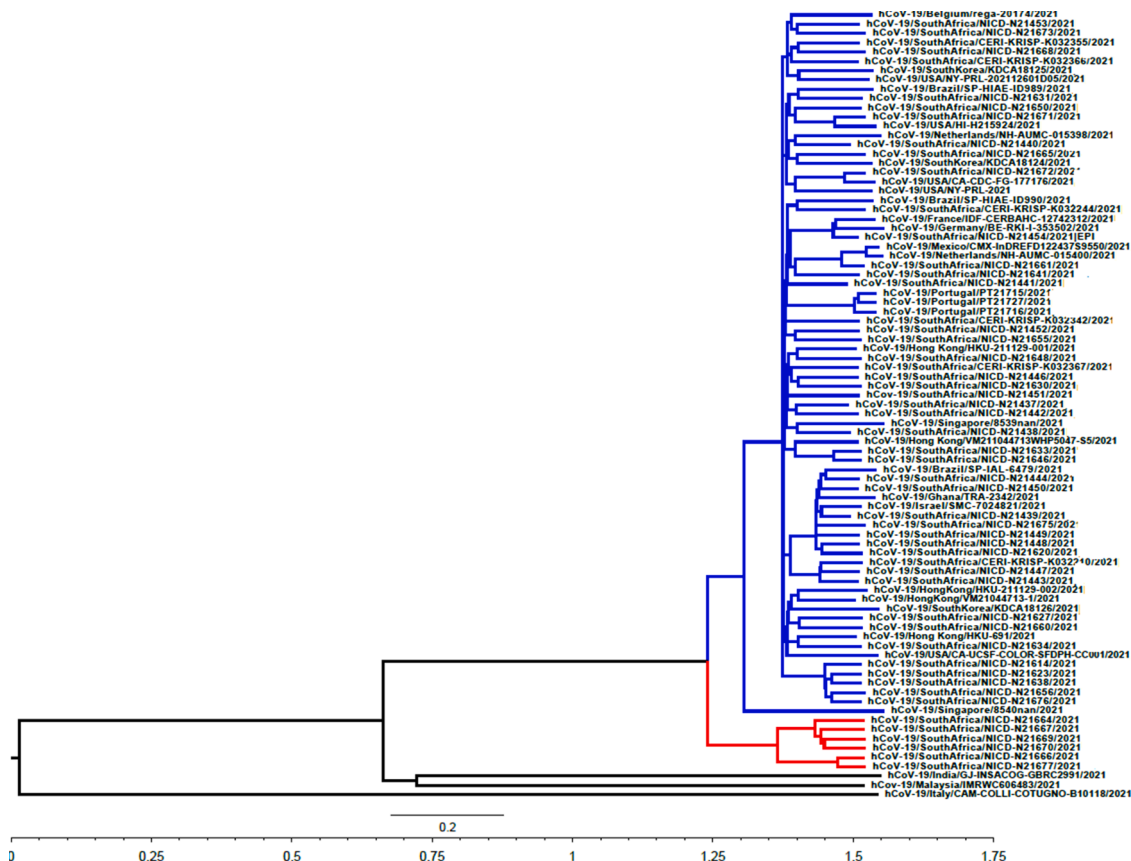


Fig. 3. Bayesian MCMC phylogenetic tree analysis of Omicron SARS-CoV-2 strains. A maximum clade credibility tree obtained using the HKY+I model, a strict molecular clock and the Birth-Death Skyline Serial population model is shown. The tree is rooted to the Most Recent Common Ancestor (MRCA). The two 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 Ferreiro 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 Ferreiro:** 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.
- Kandee, 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., Bhone, R., Trial, I., 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).

Xia, X., 2021. Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design. *Viruses* 13, 109. <https://doi.org/10.3390/v13010109/v13010109>.
Xia, X., 2021. Domains and Functions of Spike Protein in SARS-Cov-2 in the Context of Vaccine Design. *Viruses* 13, 109. <https://doi.org/10.3390/v13010109/v13010109>.

Zhang, L, Li, Q, Liang, Z, Li, T, Liu, S, Cui, Q, Nie, J, Wu, Q, Qu, X, Huang, W, Wang, Y., 2021. The significant immune escape of pseudotyped SARS-CoV-2 Variant Omicron. *Emerg Microbes Infect* 10, 1–11. <https://doi.org/10.1080/22221751.2021.2017757.41423-020-0458-z>.