



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Análisis de la variabilidad genética en siete exones del gen ACE2 humano y su relación con el linaje de SARS-CoV-2 en un grupo de individuos diagnosticados con COVID-19

Natalia Andrea Hernández Bocanegra

Universidad Nacional de Colombia
Facultad de ciencias
Maestría en ciencias- Microbiología
Bogotá, Colombia
2024

Análisis de la variabilidad genética en siete exones del gen ACE2 humano y su relación con el linaje de SARS-CoV-2 en un grupo de individuos diagnosticados con COVID-19

Natalia Andrea Hernández Bocanegra

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título
de:

Magister en Ciencias-Microbiología

Director (a):

Ph.D. Andrés M. Pinzón V.

Línea de Investigación:

Biología molecular de agentes infecciosos

Grupo de Investigación:

Bioinformática y Biología de sistemas

Universidad Nacional de Colombia

Facultad de ciencias

Maestría en ciencias- Microbiología

Bogotá, Colombia

2024

Agradecimientos

Al culminar una labor tan exigente y llena de desafíos como la elaboración de mi tesis resulta inevitable expresar mi agradecimiento a las personas e instituciones que contribuyeron a este proceso. Sin su colaboración, alcanzar la conclusión de este trabajo habría sido prácticamente imposible. Por esta razón, me complace sinceramente aprovechar este espacio para reconocer y expresar mi gratitud hacia ellos.

Al profesor Andrés Mauricio Pinzón Velasco, por brindarme su apoyo y orientación para llevar a cabo este trabajo de maestría y al profesor Carlos Eduardo Arboleda Bustos por sus enseñanzas y aportes en la elaboración del proyecto.

Al laboratorio de biotecnología a cargo del profesor Emiliano Barreto, al laboratorio de microbiología de la facultad de medicina a cargo de la docente Martha Murcia y al servicio de secuenciación y análisis molecular (SSiGMol) por su participación en diferentes actividades que contribuyeron a la realización del proyecto.

A la Clínica Colsanitas y al Biorepositorio de la fundación Universitaria Unisanitas por su aporte para la obtención de las muestras para el desarrollo de la presente investigación.

Resumen

Análisis de la variabilidad genética en siete exones del gen *ACE2* humano y su relación con el linaje de SARS-CoV-2 en un grupo de individuos diagnosticados con COVID-19

La interacción entre la proteína Spike del SARS-CoV-2 y el receptor ACE2 humano es crucial en el desarrollo de la enfermedad COVID-19. Este estudio buscó evaluar la variabilidad genética del linaje viral y de siete exones del gen ACE2 humano, así como analizar antecedentes clínicos y condiciones sociodemográficas relacionadas con la susceptibilidad y severidad de la enfermedad. Utilizando tecnología de secuenciación, se examinaron muestras de hisopado nasofaríngeo para el genoma viral y muestras sanguíneas de 50 individuos diagnosticados con infección o neumonía viral por SARS-CoV-2. Estos individuos fueron ingresados a la Clínica Universitaria Colombia, ubicada en la ciudad de Bogotá, desde octubre del año 2020 a junio del año 2021, y se clasificaron en tres grupos de severidad según los síntomas, el diagnóstico, la insuficiencia respiratoria y la necesidad de ventilación como: leve, moderado y severo.

A pesar de no encontrar asociaciones estadísticamente significativas entre los linajes virales y la gravedad de la enfermedad, se observó que el linaje B.1.621 predomina en casos leves, mientras que el linaje B.1.420 está presente solo en casos moderados y el linaje P.1 se encontró en individuos moderados y severos. El análisis de las secuencias de los siete exones del gen *ACE2* reveló una única variante homocigota en el exón 17; que corresponde a la sustitución de una guanina (G) por una adenina (A), causando un cambio sinónimo en la posición 748 de la proteína ACE2 que resulta en la codificación de una valina. No se encontraron correlaciones significativamente estadísticas entre datos demográficos, antecedentes clínicos y gravedad de la enfermedad. Se requieren estudios futuros con muestras representativas para comprender la fisiopatología del virus y su relación con el fenotipo desarrollado.

Palabras clave: Receptor ACE2 humano, región Spike, variante, linaje viral.

Abstract

Analysis of genetic variability in seven exons of the human ACE2 gene and its relationship with the SARS-CoV-2 lineage in a group of individuals diagnosed with COVID-19

The interaction between the Spike protein of SARS-CoV-2 and the human ACE2 receptor plays a pivotal role in the development of COVID-19. This study aimed to assess the genetic variability of the viral lineage and seven exons of the human ACE2 gene, along with analyzing clinical backgrounds and sociodemographic conditions related to susceptibility and disease severity. Using sequencing technology, samples from nasopharyngeal swabs were examined for the viral genome, and blood samples from 50 individuals diagnosed with SARS-CoV-2 infection or viral pneumonia were sequenced.

Although no statistically significant associations were found between viral lineages and disease severity, it was observed that the B.1.621 lineage predominated in mild cases, while the B.1.420 lineage was present in moderate cases, and the P.1 lineage was found in both moderate and severe individuals. Analysis of exonic sequences of the *ACE2* gene revealed a homozygous variant in exon 17, involving a synonymous substitution at position 133, resulting in the encoding of valine. No significant associations were found between demographic data, clinical history, and disease severity.

These results indicate genetic variability in the studied regions of the ACE2 gene, underscoring the importance of future large-scale studies with representative samples to comprehend the virus's pathophysiology and its relationship with the developed phenotype.

Keywords: Human ACE2 receptor, Spike región, variant, viral lineage.

Tabla de contenido

Resumen	V
Abstract	VI
Introducción	11
1. Marco Teórico	14
1.1 Origen y aislamiento del SARS-CoV-2	14
1.2 Linajes de SARS-CoV-2	15
1.2.1 Sistema pangolín	17
1.2.2 Nextclade.....	18
1.2.3 Variantes de Interés (VOI) y variantes de preocupación (VOC)	19
1.3 Genoma y proteínas del SARS-CoV-2	20
1.4 Interacción de la proteína de superficie viral (Spike) con el receptor ACE2 humano.....	22
1.4.1 El receptor ACE2, exones de importancia.....	24
1.5 variabilidad genética.....	25
2. Justificación y planteamiento de la problemática	26
3. Pregunta de investigación	27
4. Hipótesis	27
5. Objetivos	28
5.1 Objetivo general	28
5.2 objetivos específicos	28
5.2.1 Obtener la secuencia genómica completa del SARS-CoV-2 para cada uno de los cincuenta individuos infectados y establecer su linaje.....	28
5.2.2 Secuenciar los exones 2, 3, 9, 10, 16, 17 y 18 del gen ACE2 humano y determinar su variabilidad genética para un grupo de cincuenta individuos diagnosticados con COVID-19.	28
5.2.3 Establecer la relación entre la variabilidad genética del gen ACE2, el linaje viral SARS-CoV-2, variables sociodemográficas y los antecedentes clínicos existentes en tres estados de severidad de la enfermedad.	28
6. Metodología	29
6.1 Individuos, tipo de muestras y grupos de gravedad.....	29
6.1.1 Individuos	29
6.1.2 Obtención de las muestras	31
6.1.3 Determinación del grupo de gravedad	31
6.1.4 Extracción de material genético viral y RT-PCR.....	32
6.1.5 Preparación de librerías.....	33
6.1.6 Secuenciación del genoma de SARS-CoV-2	33
6.1.7 Ensamblaje del genoma de SARS-CoV-2.....	34

6.1.8	Determinación del linaje viral.....	34
6.1.9	Extracción de ADN de individuos con COVID-19.....	34
6.1.10	Diseño de primers de amplificación y secuenciación para exones del gen ACE2	34
6.1.11	Amplificación y purificación de los exones del gen ACE2.....	35
6.1.12	Secuenciación de exones del gen ACE2.....	37
6.1.13	Alineación y análisis de secuencias.....	37
6.2	Análisis estadístico	38
7.	Resultados	39
7.1	Linajes de SARS-CoV-2 identificados	39
7.2	Variabilidad genética en el gen ACE2.....	44
7.3	Correlación entre la variabilidad genética del gen ACE2, el linaje viral SARS-CoV-2, variables sociodemográficas y los antecedentes clínicos	47
8.	Discusión.....	49
9.	Conclusiones	57
10.	Recomendaciones	58
	Anexos.....	59
	Referencias	60

Lista de figuras

Figura 1. Árbol filogenético que abarca las secuencias completas del genoma del SARS-CoV, SARS-CoV-2 y otros betacoronavirus.....	17
Figura 2. Clados del SARS-CoV-2 en la clasificación con Nextclade.....	19
Figura 3. Anotación del genoma del SARS-CoV-2..	21
Figura 4. Rutas empleadas por el SARS para entrar a la célula huésped.....	23
Figura 5. Flujo de trabajo para el tratamiento de las muestras pertenecientes al biorepositorio.....	31
Figura 6. Diagrama de flujo para la estandarización in silico de la PCR.	36
Figura 8. Porcentaje de distribución de los distintos linajes y clados virales presentes en la muestra analizada.	41
Figura 9. Distribución de Clados según la herramienta en línea Nextclade.	42
Figura 10. Alineamiento exón 17 del gen ACE2.	46
Figura 11. Gráfico de la matriz de correlación de múltiples variables..	48

Lista de tablas

	Pág.
Tabla 1. Información de los individuos estudiados.	30
Tabla 2. Diseño de primers para los exones del gen ACE2.....	35
Tabla 3. Diseño del programa de PCR.....	36
Tabla 4. Cálculos para la Mastermix.	36
Tabla 5. Linaje y clado viral determinado para cada individuo y su estado de gravedad..	40
Tabla 6. Exones secuenciados por individuo.	46
Tabla 1. Mutaciones de linajes de frecuencia relativa baja.....	61

Introducción

La circulación del síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2), agente causal del COVID-19, se inició en Wuhan, China, en diciembre de 2019. Este brote fue identificado después de la notificación de varios centros de salud locales mediante un mecanismo de vigilancia para "neumonía de etiología desconocida", implementado posterior al brote de SARS en el 2003 (Li et al., 2020). En marzo de 2020, se declaró oficialmente una pandemia mundial, marcando uno de los eventos más notorios en la historia de la humanidad y señalando la posibilidad de eventos similares en el futuro (Singh & Yi, 2021). El Comité Internacional sobre Taxonomía de Virus estableció el nombre SARS-CoV-2 para este nuevo patógeno humano basándose en su filogenia y taxonomía (Gorbalenya, Baker, Baric, de Groot, Drosten, Gulyaeva, Haagmans, Lauber, Leontovich, Neuman, Penzar, Perlman, Poon, et al., n.d.).

(Vadgama et al., 2022)

Según datos de la Organización Panamericana de la Salud, desde enero de 2020 hasta el año 2023 se reportan 193.21 millones de casos y alrededor de 2.99 millones de muertes por COVID-19 en las Américas (*Situación Del SARS CoV2 - Región de Las Américas - OPS/OMS | Organización Panamericana de La Salud*, n.d.). Para Colombia se registran 142.780 muertes según datos del Instituto Nacional de Salud (Instituto Nacional de Salud, n.d.). Este microorganismo y la enfermedad que causa han sido objeto de un extenso estudio; sin embargo, existen puntos clave en la patogenia del virus y la interacción con ACE2 que aún no están resueltos.

Identificado como el noveno coronavirus que afecta a los humanos, comparte un 79% de similitud con el SARS-CoV, lo que lo clasifica en otro clado (Holmes et al., n.d.). Este nuevo *betacoronavirus* se clasifica como parte del subgénero *Sarbecovirus* de la familia *Coronaviridae* (N. Zhu et al., 2020). Su genoma es de ARN monocatenario de sentido

positivo, con una longitud aproximada de 30 Kb, envuelto en una nucleocápside helicoidal. Su genoma consta de 11 genes con 11 marcos abiertos de lectura (ORF), que en total codifican 29 proteínas (F. Wu et al., 2008). La proteína Spike (S), envoltura (E), membrana (M) y nucleocápside (N) son las proteínas estructurales del virus. Mientras que las proteínas S, M y E se encuentran en la membrana del virión, la proteína N juega un papel crucial en la unión y empaquetamiento del genoma de ARN (C. rong Wu et al., 2022).

Al inicio de la pandemia se identificaron dos tipos de linajes virales 'A' y 'B'. Sin embargo, los avances en la secuenciación han permitido que a la fecha se hayan reportado 16,288,764 secuencias del genoma de SARS-CoV-2 en la base de datos de GISAID. Esto llevó a la creación de un sistema de clasificación de linajes denominado Pangolin, donde los linajes son evolutivamente descendientes de un linaje "padre" o linaje de origen (Rambaut et al., n.d.) y la herramienta Nextclade usada para clasificar las secuencias del SARS-CoV-2 en función de su relación genética (Hadfield et al., 2018). Así mismo El Departamento de Salud y Servicios Humanos (HHS) de los Estados Unidos designo El Grupo Interagencial por el SARS-CoV-2 (SIG), encargado de monitorear variantes emergentes, con su sistema de clasificación de Variante de gran consecuencia (VOHC), Variante de preocupación (VOC), Variante de interés (VOI), y Variantes bajo monitoreo (VBM) (Centro para el control y la prevención de enfermedades, n.d.).

Se ha descrito que las interacciones entre la proteína de superficie Spike del virus y el receptor ACE2 humano son cruciales para la entrada del virus en la célula. Este proceso complejo requiere la acción coordinada del receptor y la activación proteolítica de la glicoproteína S en la superficie de la célula huésped para lograr la fusión del virus y la membrana celular (Yang et al., 2020).

El gen *ACE2*, con una longitud de 39.98 Kb, ubicado en el cromosoma X (Xp22.2), tiene 18 exones que codifican una glicoproteína de 805 aminoácidos (Senapati et al., 2021). Los exones 2, 3, 9 y 10 del ACE2 codifican 24 aminoácidos claves en la participación del reconocimiento viral y pertenecientes al dominio peptidasa (Hoffmann et al., 2020; Lan et al., 2020; Yan et al., 2020). El exón 16 codifica para la metaloproteasa ADAM-17, que participa en el procesamiento de ACE2 (Heurich et al., 2014; Shovlin & Vizcaychipi, 2020), por último, los exones 17 y 18 codifican para la proteasa transmembrana TMPRSS2 y la

serina TMPRSS11D, encargadas del reconocimiento y unión con el sitio de escisión S1/S2 de la proteína Spike del virus (Senapati et al., 2021).

Se ha propuesto que la variabilidad genética del receptor ACE2 puede modular la susceptibilidad y conducir a una gravedad variable de la enfermedad (Zeberg & Pääbo, 2020). Esta teoría ha dado origen a la Iniciativa de Genética del Huésped COVID-19, que busca reunir datos genómicos de investigaciones en diferentes poblaciones a nivel mundial (“The COVID-19 Host Genetics Initiative, a Global Initiative to Elucidate the Role of Host Genetic Factors in Susceptibility and Severity of the SARS-CoV-2 Virus Pandemic,” 2020).

A pesar de la extensa literatura e investigación activa sobre la interacción ACE2-virus y la variabilidad genética viral, persisten vacíos en el conocimiento sobre el papel de las variaciones del gen ACE2 en esta interacción. Esto es particularmente relevante en el contexto de las variaciones específicas de ACE2 en diferentes poblaciones, incluida la colombiana.

Esta investigación se centra en la evaluación de la variabilidad genética en siete exones del gen ACE2, así como la presente en SARS-CoV-2. La premisa fundamental es que estas variaciones genéticas pueden incidir en el proceso de entrada del virus a la célula huésped, lo que podría afectar la susceptibilidad y, consecuentemente, la gravedad de la infección por SARS-CoV-2. Para llevar a cabo este análisis, se examinarán individuos residentes en Bogotá positivos para SARS-CoV-2 y clasificados por el estado de gravedad de la enfermedad.

1. Marco Teórico

En el siguiente marco teórico, se exponen los elementos conceptuales fundamentales para el desarrollo de este trabajo. Se aborda en primer lugar el origen del SARS-CoV-2, sus características principales, los linajes virales y sus sistemas de clasificación, el genoma del virus y las principales proteínas codificadas. Seguido por la interacción con el receptor ACE2 y la relevancia de este gen. Por último, se expone la importancia de la variabilidad genética.

1.1 Origen y aislamiento del SARS-CoV-2

La enfermedad del coronavirus 2019 (COVID-19) surgió tras la notificación de varios centros de salud con casos de neumonía de causa desconocida, 27 de estos, vinculados epidemiológicamente con un mercado mayorista de la provincia de Hubei, en la ciudad de Wuhan, China (N. Zhu et al., 2020). El 31 de diciembre del año 2019 el Centro Chino para el Control y la Prevención de Enfermedades (CDC) junto con las autoridades chinas anunciaron un nuevo coronavirus como el agente causal del brote respiratorio, inicialmente denominado 2019-nCoV. Su propagación se extendió masivamente a todas las provincias chinas y a nivel internacional. Posteriormente, el 11 de marzo de 2020, la Organización Mundial de la Salud declararía el brote global de COVID-19 como una pandemia (Deng & Peng, 2020; Hu et al., 2020). (Shang et al., 2020a)

La recolección de cuatro muestras de lavado broncoalveolar del Hospital Wuhan Jin Yin-Tan fue esencial para la detección y aislamiento del patógeno viral. La extracción de ARN se empleó para la clonación y secuenciación del genoma, revelando una coincidencia del 85% de identidad con un CoV similar al SARS de murciélago previamente reportado (bat-SL-CoVZC45, MG772933.1). Las tres secuencias obtenidas se enviaron a GISAID (Global Initiative on Sharing All Influenza Data) una herramienta implementada desde el 2008 que permite la disponibilidad de genomas de virus gripales (ID de acceso: EPI_ISL_402119, EPI_ISL_402120, EPI_ISL_402121) (Hernández et al., n.d.; N. Zhu et al., 2020). Además, el ensayo de RT-PCR dirigido a la región de consenso RdRp de los betacoronavirus fue exitoso. Se llevó a cabo el cultivo en células epiteliales de vías respiratorias humanas y en líneas celulares Vero E6 y Huh-7, seguido de evaluación en microscopía electrónica de

transmisión. Los resultados mostraron un efecto citopático 96 horas después de la inoculación y partículas virales esféricas con diámetros entre 60 y 140 nm, con picos distintivos que daban la apariencia de una corona solar (N. Zhu et al., 2020).

Alrededor del 55% de los casos reportados hasta diciembre de 2019 habían tenido exposición a Huanan y a otros mercados en Wuhan. Después de su cierre, se detectó SARS-CoV-2 en muestras ambientales; Sin embargo, los restos de animales dieron negativos para el virus, lo que no proporcionó resultados concluyentes sobre su origen (Organization, 2021). Un estudio retrospectivo mostró que desde el año 2017 al 2019 se comercializaron 38 especies de animales en los mercados de Wuhan, dentro de los cuales no se incluyeron pangolines, lo que respalda la idea de que estos no son su huésped indirecto. Por otro lado, se evidenciaron malas condiciones de higiene y se detallaron otras infecciones zoonóticas (Xiao et al., 123 C.E.).

La gran mayoría de los virus humanos tienen un origen zoonótico, y los coronavirus anteriormente reportados no son la excepción (Holmes et al., n.d.). Se han reportado algunos SARS-CoV de murciélagos que pueden infectar humanos (Menachery et al., 2015; Wang et al., 2018). El reporte de la secuenciación completa de una muestra de SARS-CoV-2 obtenida a inicios de la epidemia en Wuhan (número de acceso de GISAID: EPI_ISL_402131) mostró un alto porcentaje de identidad con el genoma de RaTG13 un coronavirus de murciélago aislado previamente de *Rhinolophus affinis* (Wang et al., 2018).

Sin embargo, el brote se inició cuando los murciélagos estaban hibernando y no se registró comercialización en el mercado Huanan. Además, el porcentaje de identidad con otros coronavirus de murciélagos fue inferior al 90% lo que indica que otro animal estaría actuando como huésped intermediario entre murciélagos y humanos (Lu et al., 2020).

1.2 Linajes de SARS-CoV-2

La percepción histórica de los virus como agentes causales de enfermedades ha llevado al establecimiento de un sistema de nomenclatura formal que facilita la catalogación de un virus nuevo. Si un virus emergente se encuentra dentro de un grupo conocido de aislados, se considera una variante de este. En cambio, si está fuera de un grupo y muestra distancias intergrupales comparables, el virus se clasifica como distinto y puede denominarse como nuevo (Gorbalenya, Baker, Baric, de Groot, Drosten, Gulyaeva, Haagmans, Lauber, Leontovich, Neuman, Penzar, Perlman, M Poon, et al., n.d.). La

clasificación actual de los coronavirus comprende 39 especies en 27 subgéneros cinco géneros y dos subfamilias que pertenecen a la familia *Coronaviridae*, suborden *Cornidovirineae*, orden *Nidovirales* y reino *Riboviria* (*Propuestas Aprobadas | ICTV*, n.d.; Siddell et al., 2019). Los coronavirus (CoV) pertenecen al orden *Nidovirales*, que incluye las familias *Coronaviridae*, *Arteriviridae*, *Mesoniviridae* y *Roniviridae*, y a su vez los *Coronaviridae* se subdividen en cuatro géneros, los *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus* (Gorbalenya et al., 2020).

Los grupos de estudio del ICTV (Comité Internacional de Taxonomía de Virus) cuantifican y dividen la variación en las proteínas replicativas más conservadas, codificadas en los marcos de lectura abiertos 1a y 1b (ORF1a/1b) del genoma del coronavirus. La estimación de las distancias entre el SARS-CoV-2 y los coronavirus estrechamente relacionados varía entre estudios, dependiendo de la elección de la medida, ya sea nucleotídica o aminoacídica, y de la región del genoma. Por lo tanto, aún no se ha alcanzado un acuerdo sobre la posición taxonómica exacta del virus dentro del subgénero Sarbecovirus (Gorbalenya, Baker, Baric, de Groot, Drosten, Gulyaeva, Haagmans, Lauber, Leontovich, Neuman, Penzar, Perlman, M Poon, et al., n.d.).

Según un análisis filogenético, se reveló que el SARS-CoV-2 comparte un 79% de identidad con el SARS-CoV y un 50% con el MERS-CoV (Lu et al., 2020), Por lo tanto, pertenece a un clado diferente, ya que la identidad de secuencia de los dominios conservados de la replicasa (ORF1a) es inferior al 90%. Este nuevo betacoronavirus se clasifica como parte del subgénero Sarbecovirus de la familia *Coronaviridae* (N. Zhu et al., 2020).

Asimismo, el Grupo de Estudio *Coronaviridae* (CSG) del ICTV propone un formato de nomenclatura estandarizado que incluye una referencia el organismo huésped del cual se aisló el virus, el lugar de aislamiento (ubicación geográfica), un número de aislamiento o cepa, y el tiempo de aislamiento (año o de manera más detallada) en el formato virus/hospedador/ubicación/aislamiento/fecha. (Gorbalenya, Baker, Baric, de Groot, Drosten, Gulyaeva, Haagmans, Lauber, Leontovich, Neuman, Penzar, Perlman, M Poon, et al., n.d.)

Los virus del linaje 'A' exhiben un patrón "T/C", mientras que los virus del linaje 'B' presentan un patrón "C/T" en estos sitios clave (Pekar et al., n.d.).

El sistema de clasificación de linajes denominado Pangolin se basa en evidencia filogenética y características genómicas. Cada linaje muestra diferencias nucleotídicas con respecto a un linaje ancestral, se compone de al menos cinco genomas secuenciados con >95% del genoma, exhibe cambios de nucleótidos compartidos y tiene un valor de >70% para el nodo que lo define. Los linajes pueden ser ancestros de otros en diferentes áreas geográficas, dando lugar a nuevas designaciones de linaje (por ejemplo, A.1.1). El proceso iterativo puede continuar por tres subniveles, después de los cuales se asigna una letra a los nuevos linajes descendientes (por ejemplo, A.1.1.1 se convierte en C.1). Cada secuencia se asigna automáticamente al nivel más bajo que cumple con los criterios de inclusión, como "A" o "B" (Rambaut et al., n.d.).

1.2.2 Nextclade

Nextclade es una herramienta que implementa una serie de pasos para obtener sus resultados. En primer lugar, realiza un alineamiento con una secuencia de referencia, en este caso, la Wuhan-Hu-1/2019 (MN908947). Utiliza pequeños fragmentos con suficiente similitud en la secuencia consultada y de referencia mediante una variante del algoritmo Smith-Waterman. Luego, lleva a cabo la traducción de nucleótidos a aminoácidos y realiza un llamado de variantes.

Cada secuencia se coloca en un árbol filogenético de referencia para determinar su ubicación filogenética. Esto se logra al comparar las mutaciones de la secuencia de consulta con las mutaciones de cada nodo y punta del árbol de referencia. Se selecciona el nodo con el conjunto de mutaciones más similares. Finalmente, sitúa la secuencia en un clado, definido por combinaciones específicas de variantes (clades.nextstrain.org) (Nextclade, n.d.).

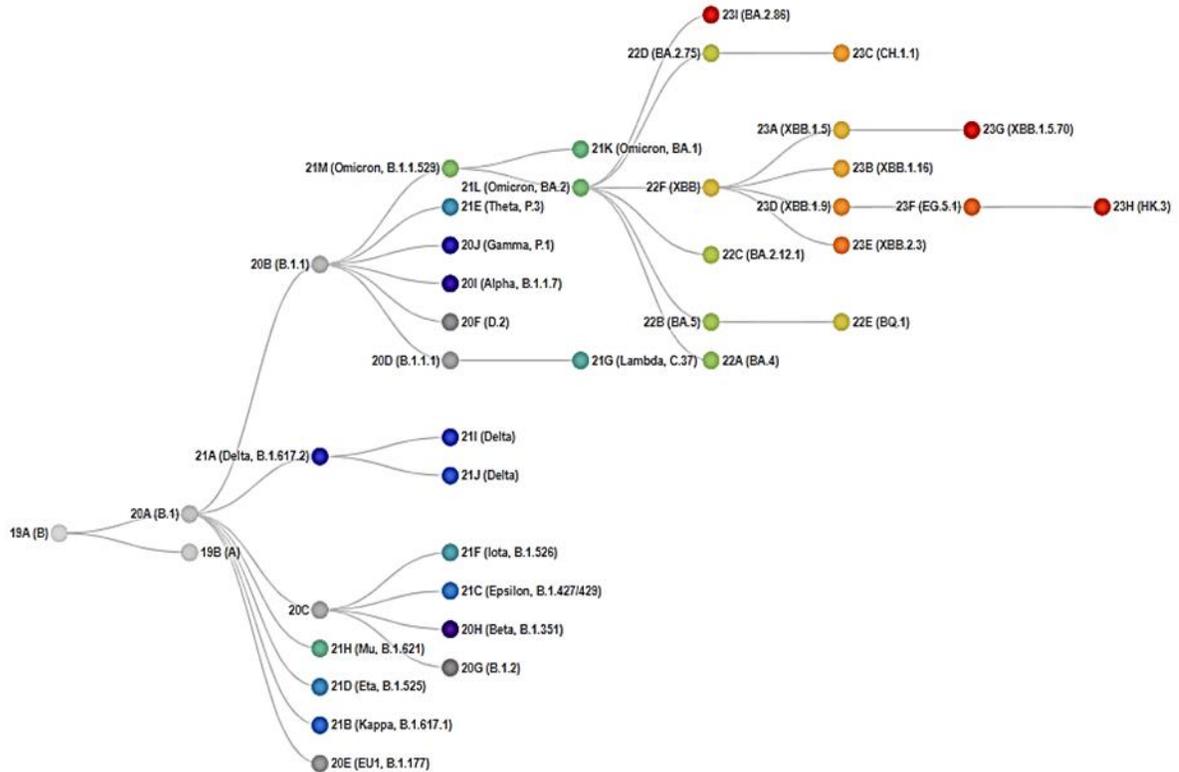


Figura 2. Clados del SARS-CoV-2 en la clasificación con Nextclade. Ilustración de las relaciones filogenéticas de los clados del SARS-CoV-2, según la definición de Nextstrain. Tomado y disponible en (clades.nextstrain.org) (*Nextclade*, n.d.).

1.2.3 Variantes de Interés (VOI) y variantes de preocupación (VOC)

Se establecieron dos grupos de clasificación para las variantes a finales de 2020, conocidas como variantes de interés (VOI) y variantes de preocupación (VOC), con el propósito de priorizar el seguimiento e investigación a nivel global (*WHO COVID-19 Weekly Epidemiological Update Vol. 74...* - *Google Académico*, n.d.). El primer grupo presenta marcadores genéticos específicos que afectan procesos como la transmisibilidad viral, la virulencia, la capacidad de evadir la detección, la reducción de la neutralización de anticuerpos producidos por infecciones naturales o vacunaciones, y la eficacia de la terapéutica o la vacunación. Dentro de este grupo se encuentran Épsilon (B.1.427 y B.1.429); Zeta (P.2); Eta (B.1.525); Theta (P.3); Iota (B.1.526); Kappa (B.1.617.1), Lambda (C.37) y Mu (B.1.621) (Ravi et al., 2022).

Por otro lado, las variantes de preocupación (VOC) poseen un potencial de transmisibilidad mejorada. El linaje Alfa (B.1.1.7), identificado por primera vez en el Reino Unido, presenta 17 mutaciones, de las cuales 8 afectan la proteína Spike (Volz et al., 2021). El linaje Beta (B.1.351 o GH501Y.V2), detectado inicialmente en Sudáfrica, exhibe nueve mutaciones en la proteína S, incluyendo tres en el dominio RBD (Tegally et al., 2021). Gamma (P.1 o GR/501Y.V3), originario de Brasil, cuenta con 10 mutaciones en la proteína S, de las cuales tres se localizan en el RBD. Delta (B.1.617.2), procedente de la India, presenta 10 mutaciones en la proteína Spike, y su sublinaje (AY.1) posee una variante adicional en el RBD (Ravi et al., 2022). La variante Omicron (B.1.1.529) se ha convertido en la más dominante a nivel mundial, con al menos 30 mutaciones en el gen S (Tegally et al., 2021).

1.3 Genoma y proteínas del SARS-CoV-2

El SARS-CoV-2 está compuesto por un ARN monocatenario de sentido positivo con una longitud aproximada de 30 Kb, envuelto en una nucleocápside helicoidal. Su genoma consta de 11 genes con 11 marcos abiertos de lectura (ORF), que en total codifican 29 proteínas (F. Wu et al., 2008). En sentido 5' a 3', comienza con dos marcos de lectura abiertos, ORF1a y b, los cuales se escinden proteolíticamente para formar 16 proteínas no estructurales que componen el complejo replicasa/transcriptasa (RTC) (Gordon et al., 2020; Yoshimoto, 1234). Los ORFs ubicados en el extremo 3' codifican para proteínas estructurales, se encuentran intercalados con varios genes accesorios (ORFs 3a, 3b, 6, 7a, 7b, 8, 9b, 9c y 10) (Thomas, 2021).

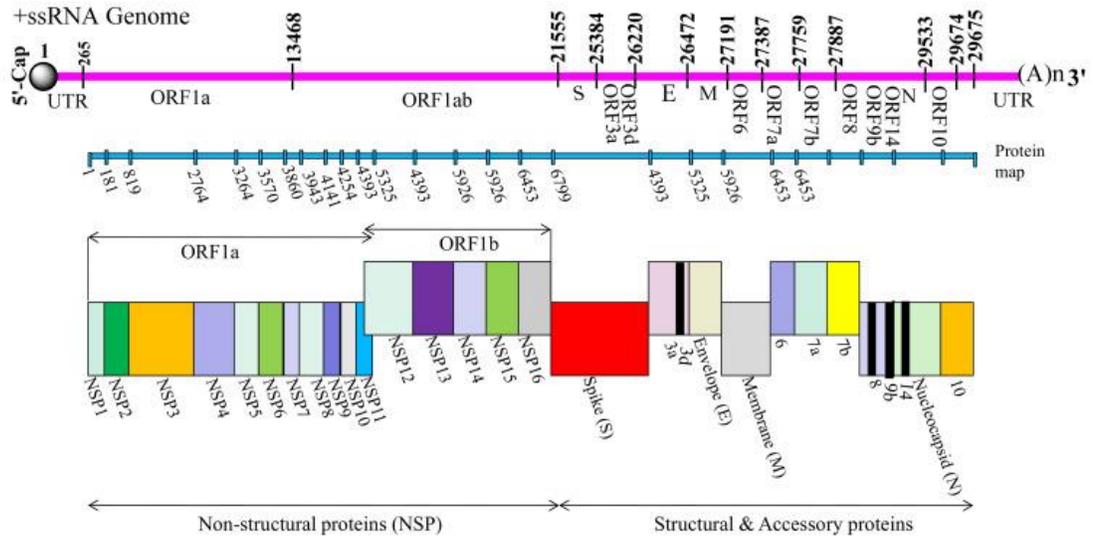


Figura 3. Anotación del genoma del SARS-CoV-2. Múltiples enzimas componen el complejo replicasa/transcriptasa (RTC), como la proteasa similar a la papaína (Nsp3), la proteasa principal (Nsp5), el complejo primasa Nsp7-Nsp8, la ARN polimerasa primaria dependiente de ARN (Nsp12), la helicasa/trifosfatasa (Nsp13), la exoribonucleasa (Nsp14), la endonucleasa (Nsp15) y las N7- y 2'O-metiltransferasas (Nsp10/Nsp16). Tomado y adaptado de: (Gordon et al., 2020; Yadav et al., 2021).

Las proteínas accesorias desempeñan diversas funciones que respaldan la infección, supervivencia y transmisión viral en las células huésped. En contraste, las proteínas no estructurales (NSP) son esenciales para la replicación del ARN viral y la evasión inmunológica (Yoshimoto, 2020). Las partículas virales maduras se forman gracias a las proteínas estructurales, las cuales tienen la responsabilidad del ensamblaje viral (C. rong Wu et al., 2022).

Dentro de las proteínas estructurales se incluyen la proteína de pico (S), la proteína de envoltura (E), la glicoproteína de membrana (M) y la proteína de nucleocápside (N). Mientras que S, M y E se encuentran en la membrana del virión, la proteína N participa en la unión y empaquetamiento del genoma de ARN (C. rong Wu et al., 2022).

La proteína S cumple un papel crucial al facilitar la unión, fusión y entrada del virus en la célula huésped. Una proteasa similar a la furina, presente en la célula huésped, escinde la

proteína S en dos subunidades: S1 y S2. S1 constituye el gran dominio de unión al receptor, mientras que S2 forma el tallo de la espiga (Glowacka et al., 2011a).

La retención de la proteína S en el compartimento intermedio ER-Golgi (ERGIC)/complejo de Golgi y su incorporación en nuevos viriones dependen de la interacción entre las proteínas S y M (Fehr & Perlman, 2015). En cambio, la unión entre la proteína M y la proteína N estabiliza el núcleo interno y la nucleocápside (complejo proteína N-ARN) de los viriones, completando así el proceso de ensamblaje viral. La envoltura viral está compuesta por las proteínas M y E, que desempeñan un papel fundamental en la producción y liberación de partículas similares a virus (VLP) (Yadav et al., 2021). La proteína E se expresa ampliamente dentro de la célula infectada su función es incorporarse a la envoltura del virus y participa activamente en el ensamblaje y la gemación viral (Nieto-Torres et al., 2014).

1.4 Interacción de la proteína de superficie viral (Spike) con el receptor ACE2 humano

La interacción inicial entre el virión y la célula huésped está mediada por la proteína S que forma homotrímeros sobresalientes de la superficie viral (Tortorici & Veessler, 2019). Al igual que en el SARS-CoV, la glicoproteína S en la superficie del virión facilita la entrada viral (Simmons et al., 2013). Este proceso es complejo y requiere la acción coordinada del receptor y la activación proteolítica de la glicoproteína S en la superficie de la célula huésped para lograr la fusión del virus con la membrana celular (Yang et al., 2020).

La proteína S se compone de una región N-terminal denominada S1, que a su vez contiene el dominio de unión al receptor, y una unidad transmembrana C-terminal designada como S2 (Harrison, 2008).

Se ha confirmado que un fragmento de 193 aminoácidos del dominio S1 se une al receptor enzima convertidora de angiotensina (ACE2) de manera más eficiente que todo el dominio completo, y este fragmento se denomina dominio de unión al receptor (RBD) (Wong et al., 2004). Su estructura tridimensional presenta una conformación compacta, generando una estabilidad de unión y una afinidad mejorada (Shang et al., 2020b). La unión al receptor desencadena cambios conformacionales en la proteína S, requiriendo la expresión de la

serina proteasa transmembrana tipo II (TMPRSS2) para la escisión y activación de S, facilitando así la fusión entre el virus y la célula, promoviendo la patogénesis y la propagación viral (Glowacka et al., 2011b) También es crucial la actividad de la cisteína proteasa endosomal, la cathepsina L, que prepara al SARS para la fusión de membranas después de la activación con ACE2 (Simmons et al., 2005).

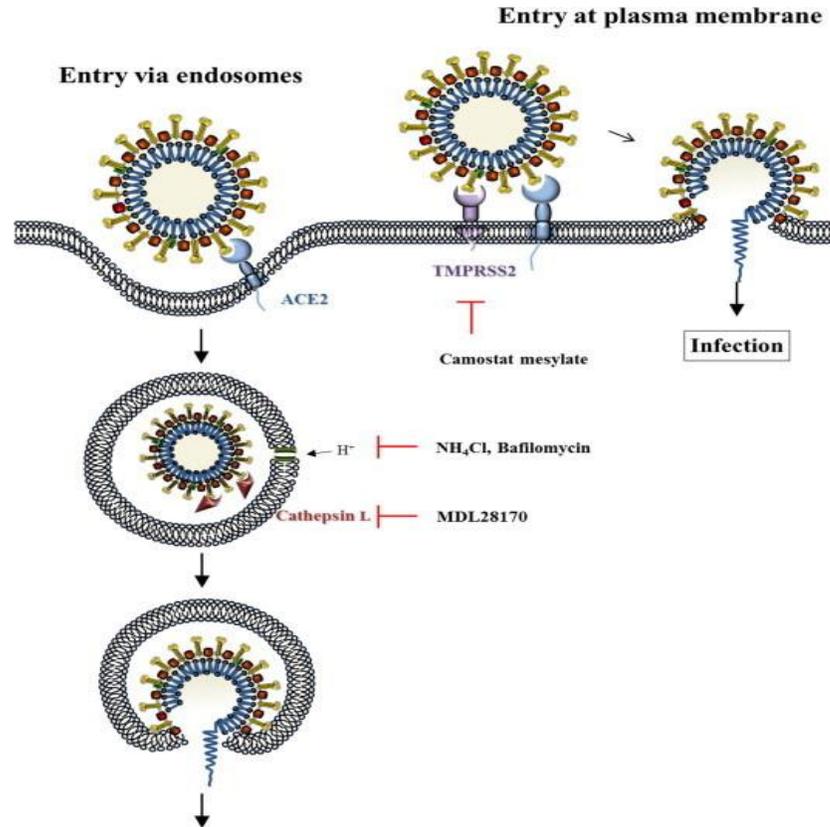


Figura 4. Rutas empleadas por el SARS para entrar a la célula huésped. El virus puede ingresar a la célula huésped de dos maneras diferentes, dependiendo de la ubicación de las proteasas necesarias para activar la proteína S del SARS. La unión al receptor celular ACE2 puede desencadenar la absorción de viriones en los endosomas. En estos endosomas, la cisteína proteasa cathepsina L, activa la proteína S. Como alternativa, TMPRSS2 puede activar la proteína S en la superficie celular o en sus proximidades, provocando la fusión de la membrana viral con la membrana plasmática. Tomado de: (Simmons et al., 2013).

Una característica distintiva de la proteína S de SARS-CoV-2 es la adopción de un sitio de escisión de furina entre S1 y S2 (S1/S2), mediado por la proteasa TMPRSS2. Además, se produce una segunda escisión proteolítica en el sitio S2', liberando el péptido de fusión

(FP) que penetra la membrana de la célula huésped, preparándola para la fusión (Casalino et al., 2020).

1.4.1 El receptor ACE2, exones de importancia

La enzima convertidora de angiotensina 2 (ACE2) es una peptidasa que se expresa en la superficie de células pulmonares epiteliales alveolares tipo I y II, así como en otros tejidos. Desempeña un papel crucial en la regulación del sistema renina-angiotensina-aldosterona (SRAA) y actúa como un receptor funcional para la glicoproteína S del SARS-CoV-2 (Devaux et al., 2020; Sun et al., 2020).

El gen *ACE2* abarca 39,98 Kb, está ubicado en el cromosoma X (Xp22.2) y consta de 18 exones que codifican una glicoproteína de 805 aminoácidos, perteneciente a la familia de enzimas convertidoras de angiotensina (Senapati et al., 2021). Este gen es polimórfico, y sus variantes genéticas están asociadas a diversas enfermedades como hipertensión, fibrilación auricular, diabetes mellitus, miocardiopatía dilatada, niveles de lipoproteína de alta densidad (HDL) alterados, miocardiopatía hipertrófica, proteína C reactiva de alta sensibilidad, entre otras (Medina-Enríquez et al., 2020).

Los análisis estructurales han identificado residuos de aminoácidos clave en el dominio peptidasa del gen *ACE2*, así como otros residuos que entran en contacto con la región RBD del virus. Estos incluyen residuos codificados por los exones 2, 3, 9 y 10 del gen *ACE2*, como Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L79, M82, Y83, N330, K353, G354, D355, R357 y R393 (Hoffmann et al., 2020; Lan et al., 2020; Yan et al., 2020).

Además, se ha identificado la presencia de serinas proteasas como la TMPRSS11D y la TMPRSS2, codificadas por los exones 17 y 18. Estas serinas proteasas interactúan tanto con ACE2 como con la proteína Spike. En su forma catalíticamente activa, escinden el dominio C-terminal transmembrana, induciendo cambios conformacionales en el complejo S1-ACE2 que facilitan la entrada viral. La metaloproteasa ADAM17, codificada por el exón 16, también participa en el procesamiento de ACE2, eliminándolo del espacio extracelular y promoviendo la entrada del virus a la célula (Heurich et al., 2014; Senapati et al., 2021; Shovlin & Vizcaychipi, 2020).

1.5 variabilidad genética

La variabilidad genética se define como la capacidad de los organismos para heredar, modificar y transmitir información genética ya que esto permite la generación de diversidad genética lo que a su vez contribuye a la variabilidad fenotípica observada en poblaciones naturales ((US) & Study, 2007). A pesar de que la especie *Homo sapiens* es relativamente joven en términos evolutivos, la población humana exhibe un significativo grado de variación genética, lo que significa que ningún individuo comparte una identidad genética idéntica con otro ser humano ((US) & Study, 2007).

Las tecnologías de secuenciación de nueva generación (NGS) han proporcionado información de secuenciación de alto rendimiento para abordar diversas preguntas de investigación, marcando así el inicio de una nueva era en la investigación genómica. Técnicas como el ensamblaje de genomas de novo y la resecuenciación a gran escala del ADN han servido para comprender las variaciones genómicas en la investigación genómica, permitiendo procesar eficientemente volúmenes masivos de datos a un costo reducido (Park & Kim, 2016). Estos avances han sido fundamentales para investigaciones genómicas, y estudios notables como el HapMap y el Proyecto de los 1,000 Genomas han sentado las bases para la planificación y ejecución de investigaciones que buscan establecer asociaciones entre numerosas variantes genómicas y una amplia variedad de enfermedades y fenotipos complejos (Caicedo et al., 2019) .

En la última década, la identificación de polimorfismos de un solo nucleótido (SNP) en todo el genoma y el acceso a conjuntos de datos de secuencias del genoma completo han posibilitado la evaluación de diversas hipótesis sobre la influencia de la selección natural, en sus diversas formas e intensidades, en la variabilidad del genoma humano (Quintana-Murci, 2016). Además, la variabilidad genética también se ha convertido en un tema relevante en el contexto de la pandemia de COVID-19, donde la interacción entre la variabilidad genética del huésped y la del virus SARS-CoV-2 puede desempeñar un papel crucial en la susceptibilidad y gravedad de la infección.

2. Justificación y planteamiento de la problemática

Los coronavirus han demostrado ser capaces de causar brotes mortales en los humanos, como se observó con el SARS-CoV y el MERS, la pandemia de COVID-19 ha resaltado la capacidad de una enfermedad contagiosa para propagarse rápidamente a nivel mundial. La COVID-19, causada por el SARS-CoV-2, representa la tercera enfermedad por coronavirus humano altamente patógena hasta la fecha. La rápida propagación de esta enfermedad altamente contagiosa ha planteado la amenaza más grave para la salud mundial en este siglo. Dada la duración del brote y la probabilidad de que el virus continúe circulando en la población humana durante mucho tiempo, es esencial realizar un seguimiento genómico continuo del SARS-CoV-2 en casos nuevos en todo el mundo.

La investigación genética se presenta como una herramienta crucial para identificar rápidamente cualquier mutación que pueda provocar cambios en el genoma viral. La variabilidad genética en pacientes con COVID-19, podría ofrecer información valiosa sobre la susceptibilidad y la gravedad de la enfermedad. La relación entre ciertas variantes genéticas y el riesgo de desarrollar síntomas graves de COVID-19 está siendo investigada, y este estudio contribuirá al conocimiento de la biología de la infección y la enfermedad por SARS-CoV-2.

La esperanza es que estos descubrimientos puedan no solo ayudar a identificar a las personas con un riesgo inusualmente alto o bajo al desarrollo de la enfermedad, sino también generar hipótesis para la reutilización de medicamentos y contribuir a estrategias de prevención, diagnóstico y tratamiento más efectivas. Dada la incertidumbre en torno a la variabilidad genética de ACE2 en pacientes con COVID-19 en población colombiana, este estudio en pacientes de Bogotá aborda una brecha en el conocimiento y podría tener implicaciones significativas para la gestión de la COVID-19 y futuras amenazas virales.

3.Pregunta de investigación

¿Existe una relación entre la variabilidad genética del linaje de SARS-CoV-2 y de siete exones del receptor ACE2 humano, con susceptibilidad a la infección viral o gravedad de los síntomas en un grupo de individuos diagnosticados con COVID-19?

4.Hipótesis

La diversidad genética del genoma de SARS-CoV-2 y del receptor ACE2 afectan su interacción y puede ser relacionado con la gravedad de los síntomas en la enfermedad COVID-19.

5. Objetivos

5.1 Objetivo general

Identificar la relación que tiene la variabilidad genética del linaje del SARS-CoV-2 y de siete exones del receptor ACE2 humano, con susceptibilidad a la infección viral o severidad de los síntomas en un grupo de individuos diagnosticados con COVID-19 en la ciudad de Bogotá.

5.2 objetivos específicos

- 5.2.1 Obtener la secuencia genómica completa del SARS-CoV-2 para cada uno de los cincuenta individuos infectados y establecer su linaje.
- 5.2.2 Secuenciar los exones 2, 3, 9, 10, 16, 17 y 18 del gen ACE2 humano y determinar su variabilidad genética para un grupo de cincuenta individuos diagnosticados con COVID-19.
- 5.2.3 Establecer la relación entre la variabilidad genética del gen ACE2, el linaje viral SARS-CoV-2, variables sociodemográficas y los antecedentes clínicos existentes en tres estados de severidad de la enfermedad.

6. Metodología

El estudio fue consistente con las pautas institucionales y fue aprobado por el Comité de Ética de la Fundación Universitaria Sanitas (CEIFUS 1610-20) y de la Universidad Nacional de Colombia. De todos los individuos participantes en el estudio se obtuvo consentimiento informado firmado por escrito para el procesamiento de las muestras, análisis de secuenciación, almacenamiento y uso de datos clínicos y moleculares.

6.1 Individuos, tipo de muestras y grupos de gravedad

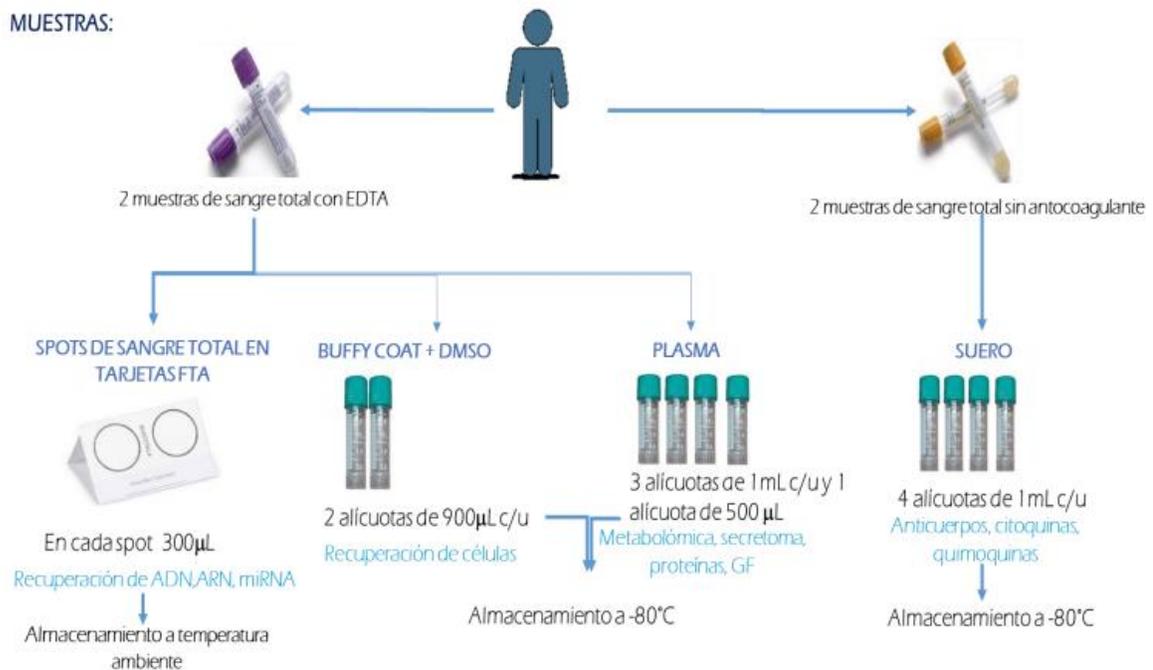
6.1.1 Individuos

La muestra comprendió 50 individuos distribuidos en cuatro cohortes de pacientes: ambulatorios, hospitalizados, hospitalizados que posteriormente ingresaron a UCI e ingresados directamente a UCI. Estos pacientes fueron admitidos en el servicio de urgencias de la Clínica Colsanitas en la ciudad de Bogotá, entre octubre de 2020 y junio de 2021.

Todos los participantes residían en la ciudad de Bogotá y tenían edades comprendidas entre los 24 y 87 años, distribuidos en 26 mujeres y 24 hombres. Fueron clasificados en tres grupos de gravedad, siendo 17 individuos leves, 18 moderados y 15 severos. En cuanto a sus antecedentes personales, se identificaron condiciones como diabetes, hipertensión, enfermedades cardiovasculares, enfermedad renal, asma, rinitis, hiperlipidemia, hipotiroidismo, y diferentes tipos de carcinomas, incluyendo gástrico, de próstata, de ovario, mielofibrosis, osteoma y de colon. Los antecedentes familiares incluyen diabetes, hipertensión, problemas cardiovasculares, y varios tipos de carcinomas, como seno, próstata, hígado, vesícula biliar, gástrico y tiroides (tabla 1).

6.1.2 Obtención de las muestras

Las muestras fueron recolectadas en dos momentos: al día cero de ingreso al servicio de urgencias y siete días después. Se obtuvo una muestra de hisopado nasofaríngeo y cuatro muestras de sangre total de cada individuo, distribuidas entre dos con anticoagulante EDTA y dos sin anticoagulante. Estas muestras fueron procesadas para su almacenamiento en las tarjetas FTA. Para el presente estudio se emplearon los hisopados nasofaríngeos y dos cortes de las tarjetas FTA que contenían sangre total tomados en el tiempo 1 es decir en el día cero de ingreso al servicio de urgencias y su posterior clasificación en un grupo de gravedad de la enfermedad.



6.1.3 Determinación del grupo de gravedad

Los 50 individuos se agruparon en tres grupos de gravedad cualitativa dependiendo de sus síntomas, diagnóstico, la insuficiencia respiratoria y la necesidad de ventilación (*Clinical Spectrum | COVID-19 Treatment Guidelines*, n.d.):

Figura 5. Flujo de trabajo para el tratamiento de las muestras pertenecientes al biorepositorio.

El biorepositorio se encuentra bajo custodia de la Fundación Universitaria Sanitas en asocio con el laboratorio clínico de la Clínica Colsanitas S.A.

- Grupo de gravedad leve: Estos pacientes que no manifestaron imágenes anormales de tórax, dificultad para respirar o disnea, a pesar de presentar signos y síntomas de COVID-19. Además, estos individuos demostraron una saturación de oxígeno medida por oximetría de pulso (SpO₂) superior al 94 % y no requirieron ventilación. Los diagnósticos en este grupo incluyeron infección por el SARS-CoV-2 no especificada e infección debida a coronavirus sin otra especificación.
- Grupo de gravedad moderado: incluyó pacientes que presentaron signos de enfermedad en las vías respiratorias inferiores o infiltrados pulmonares superiores durante la evaluación clínica o mediante imágenes. Estos individuos también demostraron una saturación de oxígeno medida por oximetría de pulso (SpO₂) inferior al 94 %, una relación entre la presión parcial arterial de oxígeno y la fracción de oxígeno inspirado (PaO₂/FiO₂) inferior a 300 mm Hg y una frecuencia respiratoria superior a 30 respiraciones por minuto. Requirieron ventilación no invasiva, y sus diagnósticos incluyeron enfermedad respiratoria aguda debida al SARS-CoV-2 (COVID-19), infección debida a coronavirus sin otra especificación, e infección por SARS-CoV-2 confirmada (hipoxemia + SDR) con factores de riesgo.
- Grupo de gravedad severo: comprendió a pacientes que experimentaron insuficiencia respiratoria, shock séptico y/o disfunción orgánica múltiple, lo que llevó a la necesidad de ventilación invasiva. Los diagnósticos para este grupo incluyeron neumonía viral por SARS-CoV-2 y neumonía adquirida en la comunidad por SARS-CoV-2.

6.1.4 Extracción de material genético viral y RT-PCR

La extracción de RNA viral se realizó a partir de 200 µl del hisopados nasofaríngeos, con el sistema automatizado MagNA Pure 96 (Roche, Penzberg, Alemania), de acuerdo con las instrucciones del fabricante. Para detectar la presencia del SARS-CoV-2 y determinar el umbral de ciclos (ct) de la muestra, se realizó una RT-qPCR utilizando el protocolo cDNA con el kit LunaScript® RT SuperMix de BioLabs® y un equipo de amplificación en tiempo real CFX96 Real de Bio-rad.

6.1.5 Preparación de librerías para secuenciación por Oxford nanopore

El proceso de preparación de librerías inicia con una PCR multiplex, utilizando los reactivos 5X Q5 Reaction Buffer, 10 mM dNTPs (NEB N0447), Q5 Hot Start DNA Polymerase (NEB M0493), Primer Pools 1 y 2 (10 μ M), agua libre de nucleasas y cDNA. Posteriormente, se cuantifican los amplicones con el Kit de ensayo Qubit dsDNA HS (Thermo Fisher Scientific) y se realiza la reparación de extremos utilizando Ultra II End Prep Reaction Buffer, Ultra II End Prep Enzyme Mix y agua libre de nucleasas.

Para la ligación, se prepara una máster mix con los reactivos Ultra II Ligation Master Mix, Ligation Enhancer, agua libre de nucleasas y el barcode asignado a la muestra. Luego, se lleva a cabo una purificación utilizando AMPure XP beads (Beckman A63880-A63881), seguida de un punto de control mediante la cuantificación del amplicon con el barcode. La última PCR tiene como objetivo ligar los adaptadores, utilizando los reactivos NEBNext Quick Ligation Reaction Buffer (5X), Mezcla de adaptadores (AMII), Quick T4 DNA Ligase y la mezcla de amplicones con barcodes.

Finalmente, se realiza una purificación con las perlas ya mencionadas y se cuantifica la librería, asegurándose de tener un mínimo de 15 a 20 ng en 15 μ l (1 ng/ μ l) (*Secuenciacion-Genoma-Completo-SARS-COV-2-Secuenciador-MINION*, n.d.). Protocolo disponible en: (https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3?version_warning=no&step=2)

6.1.6 Secuenciación del genoma de SARS-CoV-2

La secuenciación se llevó a cabo en el laboratorio de Biotecnología de la Universidad Nacional de Colombia, en el MinION de Oxford Nanopore y con una celda de flujo con los pozos disponibles para la secuenciación de las muestras (ONT FLO-MIN106). Para programar y monitorear la corrida de secuenciación se empleó el software MinKNOW. El control de calidad analítico se realizó a partir de la profundidad y cobertura de las secuencias (*Secuenciacion-Genoma-Completo-SARS-COV-2-Secuenciador-MINION*, n.d.).

6.1.7 Ensamblaje del genoma de SARS-CoV-2

Los análisis bioinformáticos se utilizaron para procesar los datos crudos obtenidos después de la secuenciación, lo que permite ensamblar el genoma completo mediante el mapeo o alineamiento de las lecturas de los fragmentos contra el genoma de referencia de SARS-CoV-2, lo que permite ensamblar por similitud de secuencia siguiendo el protocolo bioinformático (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>), así mismo se extrajo computacionalmente la región Spike para cada uno de los genomas ensamblados.

6.1.8 Determinación del linaje viral

La identificación del linaje viral se realizó a partir de las secuencias obtenidas en formato FASTA usando la plataforma Nextclade (<https://clades.nextstrain.org/>) (Nextclade, n.d.).

6.1.9 Extracción de ADN de individuos con COVID-19

La extracción de DNA genómico se realizó a partir de sangre total preservada en tarjetas Whatman FTA™, con dos cortes de 3mm preservados en tubos de eppendorf hasta su uso. EL kit para la extracción fue Invisorb® Spin Universal kit de invitak, que utiliza columnas para la extracción, siguiendo el protocolo de aislamiento de ADN genómico de sangre total con la modificación de un tratamiento previo a los spots con 500 µL buffer FTA y 20 µL proteinasa K en agitación durante 24 horas e incubación a 37°C (Sahajpal et al., 2019). Para controlar la extracción se realizó medición en el NanoDrop™ 2000 de Thermo Scientific™ utilizando como blanco el buffer de elución del Kit de extracción.

6.1.10 Diseño de primers de amplificación y secuenciación para exones del gen ACE2

El diseño de primers de amplificación para cada exón se realizó usando en el programa Primer3web versión 4.1.0 (<https://primer3.ut.ee/>) (Primer3, n.d.)y utilizando la secuencia nucleotídica de referencia descargadas de CCDS Database (CCDS ID: 14169 V.1) (<https://ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi?REQUEST=CCDS&DATA=CCDS14169>) Estos primers de amplificación también fueron empleados para la reacción de secuenciación.

DISEÑO DE PRIMERS PARA LOS EXONES DEL GEN ACE2	
Exón 2	
Left primer	5'AGCAATGCCATTCCAACCTTC3'
Right primer	5'AGCAATGCCATTCCAACCTTC3'
Exón 3	
Left primer	5'TCTTCAGCAAATTTCCATTGTT3'
Right primer	5'TGTTGAAACACACATATCTGCAA3'
Exón 9	
Left primer	5'TTCATCTTGTCCATTTTCATGC3'
Right primer	5'CCTCCAAAACAAGCCATGAG3'
Exón 10	
Left primer	5'CATCCACTGTCATCTTCATCG3'
Right primer	5'CACAGCAAACCTCATTAGAAAACA3'
Exón 16	
Left primer	5'CCTCTGTGCCACAAGTGAAG3'
Right primer	5'AACCCCCTCACCCCTTAGATG3'
Exón 17	
Left primer	5'GTGTCCCCTTTGCTGTTTTT3'
Right primer	5'GTTAAGTCTAGGAAAGGCCACTT3'
Exón 18	
Left primer	5'CCCAGCCTCCTAACACAGA3'
Right primer	5'CAAATCACCTCAAGAGGAAAAA3'

Tabla 3. Diseño de primers para los exones del gen ACE2. Secuencia de primers utilizados para amplificar y secuenciar los siete exones del gen ACE2.

6.1.11 Amplificación y purificación de los exones del gen ACE2

La PCR de amplificación se empezó estandarizando de manera in silico empleando la herramienta Biotools (<http://biotools.nubic.northwestern.edu/OligoCalc.html>) como se muestra en el diagrama.

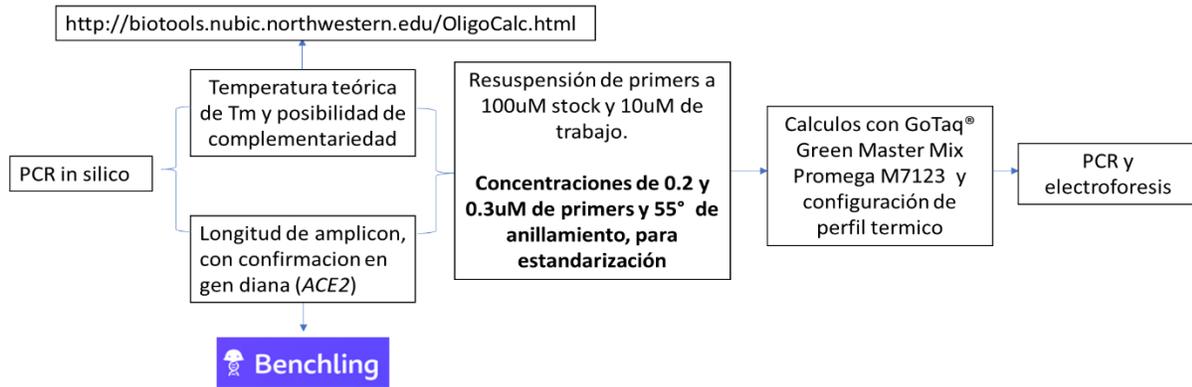


Figura 6. Diagrama de flujo para la estandarización in silico de la PCR. Se inició verificando la temperatura T_m teórica de los primers reverse (Rw) y forward (Fw) con la longitud en pares de bases Pb del exón descargando la secuencia para realizar el alineamiento con los primers.

De esta manera se determinó el siguiente programa de PCR de punto final:

Número de ciclos	Temperatura en °C	Tiempo
1X	95°C	2:00
30X	95°C	0:30
	55°C	0:40
	72°C	0:30
1X	72°C	5:00
1X	4°C	∞

Tabla 4. Diseño del programa de PCR. Programa utilizado para la PCR de punto final de los siete exones seleccionados.

Usando los siguientes cálculos de la enzima GoTaq® Green Master Mix de Promega (M7123), primers, agua libre de nucleasas y DNA:

Reactivo	Volumen en μL para 1 Rx
GoTaq® Green Master	15
Primer Rw	1
Primer Fw	1
DNA	5
Agua libre de nucleasas	8

Tabla 5. Cálculos para la Mastermix. Cantidad en μL usada de cada reactivo para preparar la Mastermix de amplificación.

Se llevaron a cabo dos reacciones de PCR idénticas por muestra, cada una con un volumen final de 30 μL . Posteriormente, los dos productos amplificados se combinaron para obtener un volumen total de 60 μL . Se confirmó el tamaño del fragmento en los productos de PCR mediante una electroforesis en gel de agarosa (Figura. 7).

Después de la amplificación, los productos fueron tratados antes de la secuenciación. Este tratamiento incluyó una purificación mediante columnas, utilizando el kit DNA Clean and Concentrator -5 de Zymo Research, siguiendo las instrucciones del fabricante. La resuspensión final se llevó a cabo con 10 μL de agua libre de nucleasas y se procedió a un secado completo de la muestra utilizando el concentrador Savant™ DNA SpeedVac™ de Thermo Scientific, seguido de una nueva resuspensión en agua libre de nucleasas con 5 μL . La cuantificación de las muestras se realizó fluorométricamente mediante el kit QuantiFluor® dsDNA System en el equipo Quantus de Promega.

6.1.12 Secuenciación de exones del gen *ACE2*

El proceso comienza realizando una mezcla con 7 μL de Terminator Ready Reaction Mix*, se estandarizó un volumen de 2,5 μL de ADN y 0,5 μL de cebador, y el volumen restante se completó con agua libre de nucleasas. Se utilizó el programa de PCR descrito en la página 5 del capítulo 3 del protocolo (BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol, 2002). El buffer de secuenciación BigDye® Terminator v1.1/3.1 se usó a una concentración de 1X, siguiendo las instrucciones del protocolo del fabricante. Posteriormente, se llevó a cabo una purificación de placas y columna de centrifugación con el kit de filtración en gel de Edge Biosystems, tal como lo indica el fabricante (BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol, 2002).

La secuenciación capilar se llevó a cabo en el servicio de secuenciación SiGMOL del instituto de genética de la Universidad Nacional de Colombia, con el equipo Abby 3500 y el kit BigDye™ Terminator v3.1 Cycle Sequencing con programación en el software data collection.

6.1.13 Alineación y análisis de secuencias

Los datos obtenidos por secuenciación capilar de los siete exones del gen *ACE2* fueron alineados con el programa Geneious versión 2023.2.1 y las secuencias de referencia

empleadas fueron CCDS Database (CCDS ID: 14169 V.1)
(<https://ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi?REQUEST=CCDS&DATA=CCDS14169>)

6.2 Análisis estadístico

Todos los datos de los pacientes fueron recolectados en un archivo de Excel. El análisis estadístico se realizó con el software del programa GraphPad Prims 10.0.2(232). Utilizando el paquete estadístico de matriz de correlación de Spearman para datos no paramétricos.

7.Resultados

7.1 Linajes de SARS-CoV-2 identificados

Las secuencias obtenidas fueron reportadas en GISAID, y el informe MultiQC del ensamblaje de los genomas reveló la presencia de lecturas en un rango de 5000 a 200 mil. Sin embargo, de las 50 muestras secuenciadas, 32 presentaron coberturas entre el 70% y el 100%, mientras que 18 muestras mostraron una cobertura inferior al 70%. Por esta razón, las muestras con cobertura inferior al 70% fueron excluidas del proceso de clasificación del linaje. Las secuencias de datos en formato FASTA se utilizan para que la herramienta Nextclade Web determine el linaje viral. La información sobre el linaje y clado viral identificado para cada individuo, junto con su correspondiente estado de gravedad, se muestra en la Tabla 5.

LINAJES VIRALES IDENTIFICADOS			
INDIVIDUO	ESTADO DE GRAVEDAD	LINAJE VIRAL	CLADO
RVD-001-011	Leve	B.1.1	20B
FRBE-002-046	Leve	B.1.625	20A
AGMP-002-050	Leve	B.1.1.348	20B
PBAM-002-051	Leve	B.1.621	21H (Mu)
CPNE-002-055	Leve	B.1.621	21H (Mu)
SMA-002-056	Leve	B.1.621	21H (Mu)
BGJF-002-087	Leve	B.1.621	21H (Mu)
GGVA-002-101	Leve	B.1.625	20A
SRLM-002-125	Leve	B.1.621	21H (Mu)
NRA-002-130	Leve	B.1.621	21H (Mu)
CRLDP-002-133	Leve	B.1.621	21H (Mu)
BJM-001-024	Moderado	B.1.111	20A
GOAG-001-025	Moderado	B.1.420	20A
RPJC-001-028	Moderado	B.1.420	20A
MTS-001-029	Moderado	B.1.420	20A
NTJP-001-034	Moderado	B.1	20A
SBLM-001-003	Moderado	B.1.420	20A
LMCA-001-010	Moderado	B.1	20A
CSJA-001-013	Moderado	B.1	20A
PNR-001-015	Moderado	B.1.621	21H (Mu)
GBJM-001-016	Moderado	P.1	20J (Gamma)

VPA-002-170	Severo	B.1.621	21H (Mu)
RSMDC-002-169	Severo	P.1	20J (Gamma)
RPMJ-002-150	Severo	P.1	20J (Gamma)
TLYI-002-044	Severo	B.1.621	21H (Mu)
TPMA-002-090	Severo	A.2.5	19B
ABF-002-092	Severo	P.1	20J (Gamma)
MALH-002-097	Severo	B.1.621	21H (Mu)
AAML-002-127	Severo	B.1.111	20A
LPFA-002-131	Severo	B.1.111	20A
MLNR-001-014	Severo	B.1.243	20A
EAJH-001-021	Severo	B.1.111	20A

Tabla 6. Linaje y clado viral determinado para cada individuo y su estado de gravedad.

Clasificación detallada del genoma viral, así como su correspondiente clado, para cada individuo incluido en el estudio. Además, se proporciona información sobre el estado de gravedad de cada paciente.

Los resultados obtenidos de la secuenciación genómica revelaron una diversidad de linajes del SARS-CoV-2 entre los pacientes, siendo el linaje B.1.621 el más predominante en la muestra, alcanzando un 34%. Este linaje se encontró distribuido en todos los niveles de severidad, aunque se observa con mayor frecuencia en individuos con síntomas leves. Le sigue el linaje B.1.111, identificado en un caso moderado y en tres casos severos. El linaje P.1 se detectó en un caso moderado y en dos casos severos, mientras que el linaje B.1.420, exclusivo de pacientes moderados, se identificó en cuatro individuos, representando estos tres linajes un 13% cada uno en la muestra. En menor proporción, se observaron los linajes B.1.625, A.2.5, B.1.243, B.1.1 y B.1.1.348.

El análisis de las frecuencias de los clados virales en la muestra revela una distribución significativa de las variantes genéticas del SARS-CoV-2. El clado 20A se destaca como el más predominante, abarcando el 44% de la población estudiada, seguido por el clado 21H, que constituye el 34%. En menor proporción, se registra la presencia del clado 20J, con un 13%, mientras que los clados 20B y 19B presentan frecuencias más bajas, con el 6% y el 3%, respectivamente. La figura 8 representa el porcentaje de distribución de los diferentes linajes y clados virales presentes en la muestra analizada.

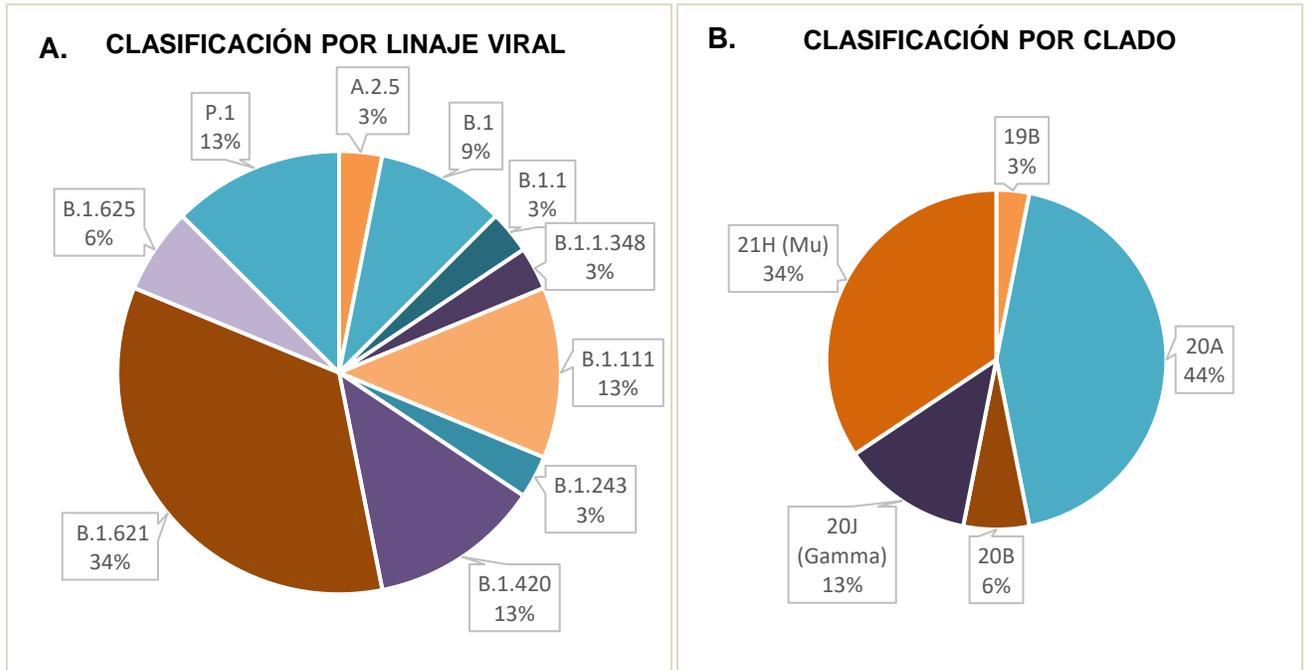


Figura 7. Porcentaje de distribución de los distintos linajes y clados virales presentes en la muestra analizada. Visualización gráfica de la frecuencia relativa de cada: **A.** variante genética **B.** Clado, del virus dentro de la población estudiada.

Es importante destacar que la interpretación precisa de cada clado puede variar según las actualizaciones en la nomenclatura y clasificación de las variantes del SARS-CoV-2 realizadas por la comunidad científica (ver figura 9).

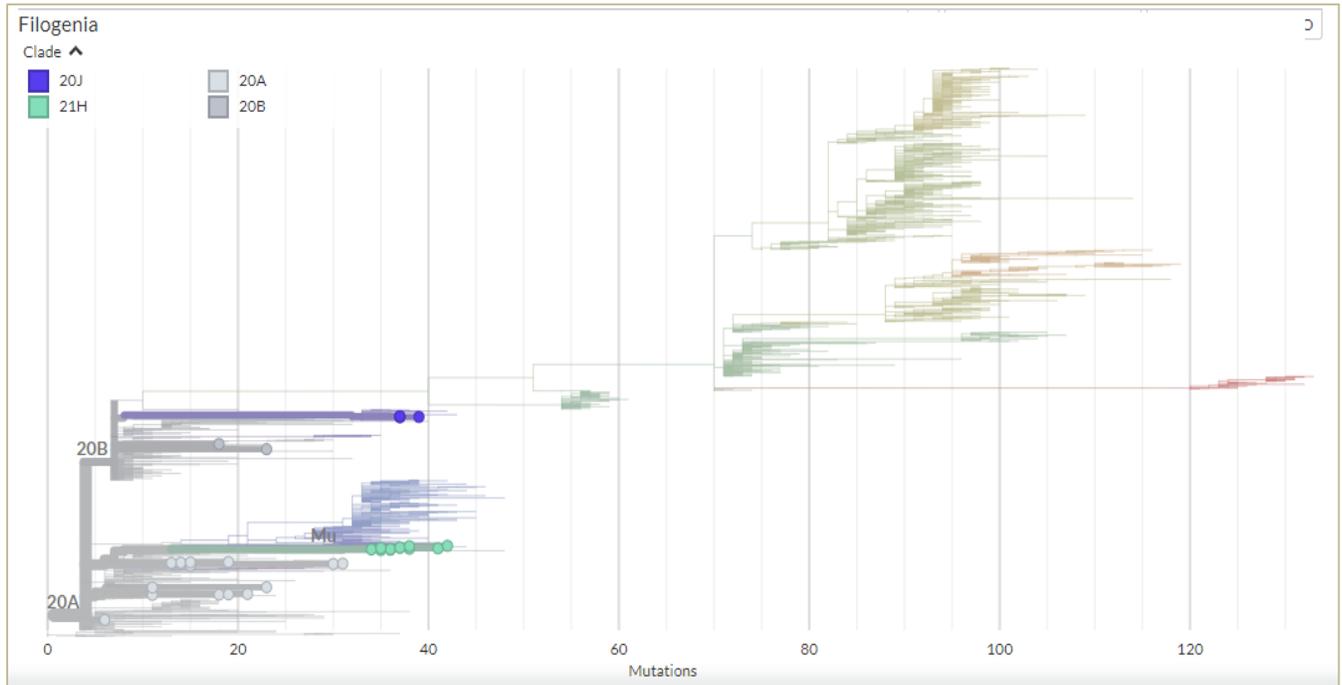


Figura 8. Distribución de Clados según la herramienta en línea Nextclade. Nuestras secuencias se representan como puntos grandes en azul clado 20J, verde clado 21H y gris claro 20A y gris oscuro 20B.

A continuación, se mencionan las sustituciones, deleciones, inserciones, y cambios frameShifts de los linajes con mayor frecuencia relativa en nuestra muestra. Las sustituciones representan cambios en la secuencia de nucleótidos, las deleciones indican la pérdida de ciertos fragmentos de la secuencia, las inserciones son la adición de nuevos fragmentos en la secuencia y los cambios frameShifts afectan el marco de lectura de la secuencia, lo que puede tener consecuencias en la función de las proteínas.

En nuestros resultados el linaje B.1 presenta de 10 a 23 sustituciones en su genoma: G219T, C241T, T1576C, C1684T, C2675A, C3037T, G3483A, C5310T, C9438T, G11670A, C12400T, C14408T, C14599T, G16181A, C17104T, A18141G, C19152T, A20268G, A23403G, T23443C, A25974G, C26028T, C28854T, T11790C, C18452T, C4921T, G6032A, C9118T, A13189T, T13804C, C18877T, G20410A, C25452T, G25563T, C28869T, y las siguientes sustituciones de aminoácidos N:S194L, ORF1a:P804T, ORF1a:G1073E, ORF1a:T1682I, ORF1a:T3058I, ORF1a:R3802H, ORF1b:P314L, ORF1b:R905K, ORF1b:H1213Y, S:D614G, ORF1a:I3842T, ORF1b:A1662V, ORF3a:Q57H, ORF1a:A1923T, ORF1b:Y113H, ORF1b:E2315K.

El linaje B.1.111 presenta entre 13 y 19 sustituciones: C241T, C3037T, C3743T, C10507T, C14408T, C18877T, A23403G, G24356T, G25440T, G25563T, G26951T, C27603T, C28854T, G2782T, A8148C, G16301A, C26029A, C26058T, C27389T, G28975T, T606C, C1060T, C2334T, C4582T, C11653T, C12120T, C25854T, C26060T, T27995C, C28887T, las cuales generan las siguientes sustituciones de aminoácidos N:S194L, ORF1a:H1160Y, ORF1b:P314L, ORF3a:K16N, ORF3a:Q57H, S:D614G, S:G932C, N:M234I, ORF1a:N2628T, ORF1b:R945K, ORF3a:Q213K, N:T205I, ORF1a:I114T, ORF1a:A690V, ORF1a:P3952L, ORF1b:M2414I, ORF3a:T223I.

Así mismo, el linaje B.1.420 presenta entre 10 a 20 sustituciones incluidas: C2334T, C3037T, G3483A, C4113T, G8368T, C14408T, C15194T, A23403G, T23443C, C23893T, C26013T, C26570T, C27143T, A27277G, T27483C, G28001C, C28708T, C1342T, A2031G, T4330C, G8353A, A8374G, G19684T, T27760A, C28923T, A29221T, C29421T, A29723T, cuyos cambios en proteína son: ORF1a:A690V, ORF1a:G1073E, ORF1a:A1283V, ORF1b:P314L, ORF1b:A576V, ORF6:I26V, S:D614G, N:A217V, N:P383L, ORF1a:D589G, ORF1b:V2073L, ORF7b:I2N, ORF1b:G575E, ORF1b:Y1200C, ORF3a:W45C.

Mientras que la variante Mu (B.1.621), exhibe un rango de sustituciones de 27 a 35: C3037T, A3428G, C4878T, C6037T, T7060C, C10029T, A11451G, A12889G, A13057T, C14408T, C14768T, C17491T, C18877T, T19035C, C20148T, C21846T, T21995A, G22599A, A23403G, C23604A, G24410A, G25563T, A26492T, C27925A, C28005T, C28093T, A28272T, C28887T, C29108T, G29427T. Con los siguientes cambios de aminoácidos: N:T205I, N:P279S, N:R385I, ORF1a:T1055A, ORF1a:T1538I, ORF1a:T3255I, ORF1a:Q3729R, ORF1b:P314L, ORF1b:A434V, ORF1b:P1342S, ORF3a:Q57H, ORF8:T11K, ORF8:P38S, ORF8:S67F, ORF8:D113Y, S:T95I, S:Y145N, S:R346K, S:D614G, S:P681H, S:D950N, ORF1b:S2430I, ORF3a:D222Y, ORF3a:L106F, ORF1b:G2005V, ORF1b:T2488M, ORF3a:P104S, ORF7a:A8T, ORF1b:L1701F, ORF1b:Y512C, ORF1b:E1184D, ORF3a:I20M, ORF7a:I10T, ORF1b:Y1200C, ORF1b:P1095H, ORF1b:V1649F, ORF9b:R32L, ORF1a:G894S, ORF1a:Q970K, ORF1b:T2537I, ORF1b:L2560F. A lo largo del genoma también se observa una inserción 21991:ACT (S:143:T), una delección del nucleótido 26158 al 26161 (ORF3a:V256-, ORF3a:N257) y una mutación con cambio en el marco de lectura (frameshift) ORF3a:258-276.

Por último el linaje gamma (P.1) presento de 30 a 38 sustituciones: C504T, T733C, C3037T, C3828T, A5648C, G6476T, A6613G, C6627T, C10519T, C12778T, C13860T, C14408T, G17259T, C21614T, C21621A, C21638T, G21974T, A22812C, A23403G, C23525T, C24642T, G25088T, T26149C, G28167A, C28512G, A28877T, G28878C, G28881A, G28882A, G28883C, C29769T, G237T, C475T, C2749T, C1663T, T2899G, C3902T, G20464A, G22132T, A23063T, G29543T, A1547G, C5784T, C10519T, A29517G, con las siguientes sustituciones de aminoácidos: N:P80R, N:R203K, N:G204R, ORF1a:P80L, ORF1a:S1188L, ORF1a:K1795Q, ORF1a:V2071F, ORF1a:T2121I, ORF1b:P314L, ORF1b:E1264D, ORF1b:D2333N, ORF3a:S253P, ORF8:E92K, ORF9b:Q77E, S:L18F, S:T20N, S:P26S, S:D138Y, S:K417T, S:D614G, S:H655Y, S:T1027I, S:V1176F, S:E484K, S:N501Y, S:R190S. Además, una delección del nucleótido 11288 al 11296 (ORF1a:S3675-, ORF1a:G3676-, ORF1a:F3677-).

Las mutaciones presentes en linajes con frecuencia relativa baja en nuestros resultados (A.2.5, B.1.1, B.1.1348, B.1.243 y B.1.625), se encuentran descritos en la tabla anexa 1.

7.2 Variabilidad genética en el gen ACE2

El número inicial de muestras destinadas a la secuenciación por exón fue de 50; sin embargo, algunas de ellas fueron excluidas debido a la insuficiencia de material genético extraído o a la mala calidad de la secuencia obtenida. Por consiguiente, se obtuvieron secuencias para 46 muestras en el exón 2, 47 en el exón 3, 48 en el exón 9, 48 en el exón 10, 43 en el exón 16, 49 en el exón 17 y, finalmente, 46 en el exón 18.

EXONES SECUENCIADOS POR INDIVIDUO								
INDIVIDUO	ESTADO DE GRAVEDAD	EXONES						
		2	3	9	10	16	17	18
RVD-001-011	Leve							
POMR-001-127	Leve							
FRBE-002-046	Leve							
FPLC-002-048	Leve							
GOOL-002-049	Leve							
AGMP-002-050	Leve							
PBAM-002-051	Leve							
RRAM-002-052	Leve							

OAMI-001-033	Severo								
--------------	--------	--	--	--	--	--	--	--	--

Tabla 7. Exones secuenciados por individuo. Se muestra el total de exones que se secuenciaron para cada individuo, los recuadros en blanco indican ausencia de la secuencia.

El análisis de los alineamientos de todos los exones, solo nos permitió identificar una única variante ubicada en el exón 17 que corresponde a el cambio de una guanina (G) por una adenina (A), causando un cambio sinónimo en la proteína en la posición 748 (Figura 10). Este cambio se presentó en dos individuos leves, uno moderado y uno severo, siendo homocigoto para un individuo leve y uno severo (NRA-002-130 y EAJH-001-021) y heterocigoto para un individuo leve y uno moderado (RRAM-002-052 y ADCM-001-020).



Figura 9. Alineamiento exón 17 del gen ACE2. Cromatograma obtenido mediante el programa Geneious versión 2023.2.1. Alineamiento de las secuencias obtenidas para el exón 17 en donde se evidencia la variante sinónima en cuatro individuos.

Figura 10. Gráfico de la matriz de correlación de múltiples variables. Las correlaciones pueden variar de -1 a 1, donde -1 indica una relación negativa perfecta (rojo oscuro), 1 indica una relación positiva perfecta (azul oscuro) y 0 indica ausencia de correlación.

Por lo tanto, se procedió a realizar una matriz de correlación de Spearman no paramétrica entre las variables: Estado de severidad, linaje establecido, edad, antecedentes clínicos personales (diabetes, hipertensión, enfermedades cardiovasculares, enfermedad renal, asma, rinitis, hiperlipidemia, hipotiroidismo y carcinomas) y antecedentes clínicos familiares (diabetes familiar, hipertensión familiar, enfermedades cardiovasculares familiares y carcinomas familiares), (Figura 11).

Es importante destacar que la correlación no implica causalidad. Es decir, aunque dos variables estén correlacionadas, no significa necesariamente que una variable cause la otra. La matriz de correlación es una herramienta descriptiva que proporciona información sobre las relaciones estadísticas, pero se necesita un diseño experimental cuidadoso para establecer causalidad.

Los resultados de la matriz de correlación indican que, en general, no se encontraron correlaciones estadísticamente significativas entre las variables analizadas. No obstante, existen algunas correlaciones que podrían requerir atención adicional:

La edad y estado de gravedad ($p = 0,07$, $r = 0,32$), mostró una correlación positiva moderada ($r = 0,32$), aunque el valor p ($0,07$) sugiere que esta asociación no alcanza significancia estadística convencional ($p < 0,05$). El antecedente personal de enfermedades cardiovasculares con el linaje identificado ($p = 0,04$, $r = 0,364$) muestra una correlación positiva moderada ($r = 0,364$) y este resultado alcanza significación estadística ($p = 0,04$). Entre las enfermedades renales con la severidad y el Linaje (Severidad: $p = 0,394$, $r = 0,155$) y (Linaje: $p = 0,289$, $r = 0,193$) existen correlaciones débiles ($r = 0,155$ y $r = 0,193$), aunque ninguno de los resultados alcanza significancia estadística. Por último, el hipotiroidismo y la diabetes familiar con la severidad y el Linaje (Hipotiroidismo: $p = 0,57$, $r = 0,1$) y (Diabetes Familiar: $p = 0,48$, $r = 0,129$) presentan correlaciones débiles ($r = 0,1$ y $r = 0,129$) ninguna de estas correlaciones es estadísticamente significativa.

8. Discusión

El SARS-CoV-2 ha experimentado mutaciones, dando lugar a diversos linajes a lo largo del tiempo. Estudiar la variabilidad genética de estos linajes es crucial para rastrear la evolución del virus, entender su propagación y evaluar cualquier cambio en su virulencia o capacidad de transmisión. Además, comprender cómo las diferencias genéticas del huésped pueden afectar la respuesta al virus, es esencial para identificar poblaciones en riesgo y desarrollar estrategias de prevención y tratamiento personalizadas. En este contexto, el presente estudio se propone identificar la relación que tiene la variabilidad genética del linaje del SARS-CoV-2 y de siete exones del receptor ACE2 humano con la susceptibilidad a la infección viral o la severidad de los síntomas en un grupo de individuos diagnosticados con COVID-19 en la ciudad de Bogotá, abriendo así la puerta a una discusión sobre la interacción de las principales mutaciones, sus implicaciones clínicas y la importancia de evaluar variabilidad genética del gen *ACE2*.

Entre diciembre de 2019 y junio de 2020, se identificaron 11 clados, de los cuales el 19B, 20A y 20B están presentes en nuestros resultados y de julio a diciembre de 2020, se detectaron cuatro clados, siendo el 20J uno de los presentes en nuestros resultados (Wolf et al., 2023). En nuestra investigación, encontramos dos variantes de interés (Mu y Gamma), cuyos datos filogenéticos sugieren su origen en Colombia y Brasil, respectivamente. La OMS ha clasificado estas dos variantes como de preocupación (VOC) debido a su mayor transmisibilidad, gravedad de la enfermedad, reducción en la neutralización por los anticuerpos del suero convaleciente y posibles fallos en los tratamientos o diagnósticos (Organización mundial de la salud, n.d.).

Un estudio sobre el origen y la evolución del SARS-CoV-2 al inicio de la pandemia en Colombia caracterizó la diversidad de los linajes en muestras de pacientes residentes en diferentes departamentos del país. Se determinó que los linajes B.1, B.1.1 y B.1.5 estaban epidemiológicamente vinculados en las primeras cadenas de transmisión, siendo el linaje B.1 el observado con mayor frecuencia. Este hallazgo coincide con informes de otros países de América Latina, como Chile y Brasil (Candido et al., 2020; Castillo et al., 2020; Laiton-Donato, Franco-Muñoz, et al., 2021), y podría explicar la presencia del linaje B.1 y B.1.1, así como sus sublinajes, en nuestra muestra.

La evolución del virus propició la aparición de ciertos linajes, como el B.1.111 y B.1.420, identificados como linajes colombianos debido a una mayor presencia de secuencias provenientes de Colombia (37,4% y 85,4%, respectivamente) en la base de datos GISAID hasta el 28 de febrero de 2021 (Laiton-Donato, Usme-Ciro, et al., 2021a). Desde enero de 2021, la vigilancia genómica en Colombia se intensificó con el programa de caracterización genómica a cargo del Instituto Nacional de Salud para un monitoreo más sensible de las variantes de SARS-CoV-2 de importación. En ese momento, la variante B.1.111 era la más prevalente, seguida por B.1 y B.1.420. Sin embargo, esta situación experimentó un cambio significativo en julio de 2021, con una marcada reducción en la presencia de estos linajes, siendo reemplazados por las variantes P.1 y B.1.621 (Laiton-Donato, Franco-Muñoz, et al., 2021).

El linaje B.1.621 se detectó por primera vez en el país el 11 de enero de 2021, en una muestra recolectada en el Magdalena. Desde febrero hasta diciembre del mismo año, mostró una frecuencia elevada en comparación con otros países de América del Sur (Álvarez-Díaz et al., 2022; Laiton-Donato, Franco-Muñoz, et al., 2021). Posteriormente, en una provincia italiana, fue aislado de un grupo de individuos de origen sudamericano (Bolivia, Argentina y Colombia) de los cuales solo un paciente fue hospitalizado y diagnosticado con neumonía leve (Mio et al., 2022). El linaje B.1.625 se informó en enero de 2021 y representa el 5,8% de las secuencias genómicas virales del primer trimestre de este mismo año en el país (Rivero et al., 2022). La variante B.1.1.348 se registra en la semana del 9 de abril del año 2021 con una frecuencia del 6,2% en Colombia según la distribución geográfica de linajes de SARS-CoV-2 del Instituto Nacional de Salud (*Instituto Nacional de Salud | Colombia Documentos Tecnicos Genomica*, n.d.). En nuestros linajes también se identificó el A.2.5 del cual no se tiene un aislamiento específico, pero se cree que puede originarse en Costa Rica (Kuja et al., 2022).

Durante el tercer pico epidémico en Colombia entre abril y julio de 2021, período en el cual también se llevó a cabo la recolección de muestras para el presente estudio, se registró el mayor número de contagios y muertes. Aproximadamente el 50% de los casos se atribuyeron al linaje B.1.621 (Mu) (Álvarez-Díaz et al., 2022). Estos hallazgos destacan la relevancia y la prevalencia significativa de este linaje específico durante un período crítico

de la pandemia en Colombia y con lo que se podría explicar la sobrerrepresentación de este linaje en nuestra muestra.

Así mismo la secuenciación de genomas virales muestreados en Manaus, Brasil, permitieron aislar un nuevo linaje, que contenía 17 cambios de aminoácidos, 10 de estos en la proteína Spike, tres deleciones, cuatro mutaciones sinónimas, y una inserción de cuatro nucleótidos al cual designaron como P.1, descendiente del linaje B.1.1.28 detectado por primera vez en Brasil (Faria et al., n.d.).

Faria et al, analizaron el número de ciclos en el que la señal de fluorescencia cruza el umbral Cycle threshold (Ct) de RT-PCR cuantitativo en muestras positivas para SARS-CoV-2, encontrando una asociación estadísticamente significativa entre la variante P.1 y valores más bajos de Ct, indicando una posible relación con mayores cargas virales (Faria et al., n.d.). Estimaron que las infecciones por la variante P.1 en Manaus tienen entre 1,2 y 1,9 veces más probabilidades de provocar mortalidad en comparación con linajes que circulaban anteriormente. Sin embargo, se destaca la necesidad de investigaciones clínicas detalladas para determinar si el aumento en el riesgo relativo se debe a la infección por P.1, al estrés en el sistema de atención de salud o a ambos (Faria et al., n.d.)

Un estudio realizado en la población colombiana (Álvarez-Díaz et al., 2022), que incluyó a 535 casos positivos para SARS-CoV-2 durante el tercer pico de pandemia, sugiere que los hombres tienen más probabilidades de experimentar resultados más graves y fallecer, aunque la gravedad de la enfermedad no difiere significativamente entre hombres y mujeres. La evaluación por edad revela tasas más altas de casos graves y fallecimientos en individuos mayores de 60 años. Entre las comorbilidades más comunes de los fallecidos y los casos graves se encuentran la EPOC, la diabetes y las enfermedades cardíacas, siendo la hipertensión más asociada con casos graves que con la mortalidad. Así mismo identificaron 39 linajes de SARS-CoV-2, pero solo nueve representan el 88,5% de los casos, con los linajes Mu (B.1.621) y Gamma (P.1) predominando entre los fallecidos, abarcando el 69% y el 18% de las muertes, respectivamente (Álvarez-Díaz et al., 2022).

Un estudio de casos y controles de infección grave por COVID-19, definiendo como grave a individuos que necesiten oxígeno suplementario, ventilación mecánica o muerte determinaron que la edad avanzada, un índice de masa corporal (IMC) elevado, historial

de insuficiencia cardíaca crónica, fibrilación auricular o diabetes, y la infección con cepas de SARS-CoV-2 del clado 20A pueden ser indicativos de un curso grave de la enfermedad (L. Zhu et al., 2022).

Con respecto a estos hallazgos, en los resultados de este estudio, se observó que las enfermedades cardiovasculares muestran correlación estadísticamente significativa con el linaje identificado ($p = 0,04$) ($r = 0,364$), al igual que la edad y la severidad de la enfermedad ($p = 0,07$, $r = 0,32$), aunque no se logró establecer una correlación estadísticamente significativa. Además, se observó que el linaje B.1.621 está presente en el 63,3% de los individuos leves, mientras que el linaje P.1 está presente en un 27,2% de los individuos severos. En resumen, aunque se estableció correlaciones moderadas se observan en los datos, la falta de significación estadística podría atribuirse al tamaño de la muestra. Estos resultados subrayan la necesidad de investigaciones adicionales con muestras más grandes y un diseño experimental más detallado para validar y comprender mejor estas posibles asociaciones. Además, el estado de gravedad de los individuos puede estar influenciado por factores propios del huésped, como la respuesta inmunológica o las interacciones específicas del virus con el huésped, incluso factores como el sistema de salud o las medidas de control implementadas.

Cabe mencionar que en el linaje B.1, se destacó la presencia de la mutación D614G en la proteína Spike, originada por la sustitución A23403G. Esta mutación suele encontrarse en conjunto con tres mutaciones adicionales: la sustitución C241T en 5'UTR, el cambio sinónimo C3037T en Nsp3, y la mutación de aminoácido P314L en ORF1b. Además, se observa una mutación de C a T en la posición 14,408, que induce un cambio específico en el aminoácido de la ARN polimerasa dependiente de ARN (RdRp P323L) (Korber et al., 2020). Este conjunto de cuatro mutaciones fue identificado inicialmente en aislamientos de genomas virales en China y Alemania a finales de enero de 2020 (Abdullaev et al., 2022).

Desde el punto de vista estructural, la mutación D614G podría potencialmente eliminar un enlace de hidrógeno en la cadena lateral de la proteína, lo que a su vez podría aumentar la flexibilidad de la cadena principal y tener un impacto en las interacciones entre los protómeros. Además, esta sustitución podría afectar la glicosilación en el sitio N616

cercano y modificar la dinámica del péptido de fusión del protómero vecino, entre otros posibles efectos (Hwang et al., 2020; Korber et al., 2020; Walls et al., 2020).

La mutación D614G se relaciona con cargas virales elevadas, pero no con la gravedad de la enfermedad, según dos estudios (Korber et al., 2020; Lorenzo-Redondo et al., n.d.). El primero incluyó datos clínicos como edad, sexo, fecha de muestreo, estado de hospitalización y el número de ciclos en el que la señal de fluorescencia cruza el umbral Cycle threshold (Ct) de la RT-PCR cuantitativo basada en el gen E, en 999 muestras positivas para SARS-CoV-2. El análisis estadístico, reveló que la presencia de la variante D614G se encuentra asociado a valores bajos de Cycle threshold (Ct) en la RT-PCR, lo que indica mayor carga viral (Korber et al., 2020). Esta teoría fue respaldada por otra investigación realizada con individuos infectados con el linaje B.1, los cuales exhiben niveles de carga viral considerablemente superiores en las vías respiratorias superiores en comparación con aquellos infectados con otra variante sin importar el lapso transcurrido desde la manifestación de los síntomas o la gravedad de la enfermedad (Lorenzo-Redondo et al., n.d.).

El linaje B.1.621 fue reportado con seis sustituciones de aminoácidos (Y144T, Y145S, R346K, E484K, N501Y y P681H) (Laiton-Donato, Franco-Muñoz, et al., 2021), sin embargo, en nuestras secuencias no encontramos las sustituciones Y144T ni Y145S. La sustitución E484K junto con la N501Y ubicadas en el dominio RBD de la proteína Spike, han sido asociadas con la inhibición de la actividad neutralizante del plasma convaleciente (Greaney et al., 2021; Xie et al., 2021). Además, se identificó la sustitución P681H, ubicada en el sitio de escisión de furina S1/S2, la cual, según estudios previos, no afecta la entrada viral, la tasa de infección ni la prevalencia (Örd et al., 2020). Por otro lado, sustituciones en los aminoácidos como L249S y E484K, localizadas en los dominios CTD y RBD de la proteína Spike respectivamente, podrían ser de interés particular debido a su potencial papel biológico en la interacción entre el virus y el huésped (Laiton-Donato, Usme-Ciro, et al., 2021b).

Estas últimas dos sustituciones (L249S y E484K) fueron inicialmente asignadas al linaje B.1.111, posteriormente fueron reasignados al linaje B.1 (Laiton-Donato, Usme-Ciro, et al., 2021b). Sin embargo, en nuestros resultados la mutación E484K está presente en orden de frecuencia en los linajes B.1.621, B.1.625 y P.1, mientras que la mutación L249S no se

encontró en ninguno de los genomas virales secuenciados. A medida en que se realizan investigaciones y se obtiene más datos de secuenciación genómica viral, las sustituciones pueden ser reclasificadas para reflejar mejor la filogenia y la evolución del virus, y entender mejor cómo ciertas sustituciones se agrupan en linajes específicos, e incluso las diferencias en las metodologías de secuenciación y análisis genómico entre laboratorios o investigadores pueden influir en la identificación y clasificación de las sustituciones (*Secuenciación Genómica y Caracterización Genética Del Virus de La Influenza | CDC*, n.d.).

El linaje P.1 se caracteriza por tener 10 mutaciones de aminoácidos en la proteína de Spike, las cuales son L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, H655Y y T1027I (Faria et al., n.d.), nueve de las 10 sustituciones fueron encontradas en nuestras secuencias genómicas virales, excepto H655Y. Las sustituciones N501Y y K417T se encuentran ubicadas en el dominio de unión al receptor (RBD) de la proteína spike, facilita la unión directa al receptor de la enzima convertidora de angiotensina 2 (ACE2) en la célula huésped, lo que resulta en una mayor capacidad de transmisión (Hirotsu & Omata, 2021).

Por otro lado, se ha planteado la hipótesis de que la variabilidad genética del receptor ACE2 puede influir en la susceptibilidad y la gravedad de la enfermedad (Zeberg & Pääbo, 2020). Como se ha observado en otras enfermedades, como el VIH, los factores genéticos del huésped podrían contribuir (Biti et al., 1997; Lieb et al., 2006). Por este motivo, se estableció la Iniciativa de Genética del Huésped COVID-19, que tiene como objetivo recopilar datos genómicos generados por investigaciones en diversas poblaciones a nivel mundial. Uno de los resultados destacados de esta iniciativa señala la implicación de una región específica del cromosoma 3 (3p21.31) asociada con el fenotipo grave de la enfermedad por COVID-19 ("The COVID-19 Host Genetics Initiative, a Global Initiative to Elucidate the Role of Host Genetic Factors in Susceptibility and Severity of the SARS-CoV-2 Virus Pandemic," 2020). Este hallazgo coincide con otras dos investigaciones que también relacionan genes en los cromosomas 3 y 9 con formas graves de COVID-19 (Ellinghaus et al., 2020; Zeberg & Pääbo, 2020)

Un estudio realizado en una población italiana evaluó la variabilidad del gen *ACE2* utilizando 6930 exomas completos (WES) y descubrió tres variantes missense: p.(Asn720Asp), p.(Lys26Arg) y p.(Gly211Arg). Estas variantes interfieren en la estructura

y estabilización de las proteínas. Además, se identificaron variantes poco comunes, como p.(Leu351Val) y p.(Pro389His), que podrían obstaculizar la internalización y se espera que impidan la unión de la proteína de pico del SARS-CoV-2. El estudio también determinó la variabilidad de ACE2 en una cohorte de 131 individuos afectados con COVID-19 y 258 controles sanos, encontrando una mayor variabilidad del ACE2 en el grupo de control. Estos hallazgos respaldan la idea de que las variantes en ACE2 podrían explicar algunas diferencias en la susceptibilidad y gravedad de la enfermedad entre individuos (Benetti et al., 2020).

En la población española, se llevaron a cabo dos estudios sobre la variabilidad en ACE2. Uno de ellos, que incluyó 204 casos de COVID-19 en estados moderados y graves y 536 controles, encontró asociación de la variante rs2285666 con hipertensión en sujetos de edad avanzada, pero no se observaron diferencias significativas entre pacientes leves y graves (Gómez et al., 2021). Por otro lado, Torre et al. utilizaron datos de secuenciación de exoma completo para analizar variantes en ACE2 y TMPRSS2, y encontraron un bajo nivel de polimorfismo. Ninguna variante se asoció significativamente con la infección por SARS-CoV-2, excepto en el gen TMPRSS2, donde la variante rs75603675 (c.23G>T) podría estar relacionada con el desarrollo de la enfermedad (Torre-Fuentes et al., 2021).

Estos dos genes también fueron evaluados en sujetos del Reino Unido, pero no se encontraron secuencias de ADN con frecuencias notablemente diferentes entre casos positivos para SARS-CoV-2 y controles (Curtis, 2021). Además, Shovlin et al. propusieron varias variantes que afectan la región de unión del SARS-CoV-2 de ACE2, como rs73635825, rs4646116 y rs766996587 (Shovlin & Vizcaychipi, 2020).

Se han identificado polimorfismos en genes diferentes al ACE2, que logran asociación con la susceptibilidad y gravedad de la enfermedad (Dieter et al., 2022). Incluso en población colombiana se identificó el polimorfismo en el gen *LZTFL1* rs11385942 como un factor importante de hospitalización (Angulo-Aguado et al., 2022). El genotipo T/T del SNP ACE2 rs2285666 mostró una contribución significativa a resultados críticos y necesidad de suplementación de oxígeno, con una fuerte asociación del alelo T en individuos masculinos (Martínez-Gómez et al., 2022). En otra investigación se identificaron dos polimorfismos en ACE2 asociados con la enfermedad rs2106809 y rs2285666 se vincularon con mayor riesgo de hospitalización y gravedad. Factores como la edad avanzada, sexo masculino y

comorbilidades aumentaron el riesgo de ingreso en UCI o muerte (Sabater Molina et al., 2022).

En nuestros resultados no se detectó variabilidad genética en los siete exones del gen ACE2 evaluados, por lo tanto, no es posible realizar un análisis de asociación entre los estados de gravedad de la enfermedad y la variabilidad en los exones de ACE2. El hecho de no encontrar variabilidad genética puede ser debido al pequeño tamaño de la muestra analizada en el presente estudio, sin embargo, esta observación también se ve reflejada en la nula detección de variabilidad en estos exones y a lo largo del gen ACE2 en los 94 individuos de la población colombiana proveniente de Medellín (CML) en el proyecto 1000 genomas, datos disponibles en la base de datos ensembl (Auton et al., 2015).

9. Conclusiones

La presente investigación realizó secuenciación genómica viral de pacientes diagnosticados con COVID-19, identificando múltiples linajes del SARS-CoV-2. El linaje B.1.621 fue el más prevalente (34%), siendo común en casos leves. Otros linajes, como B.1.111, P.1 y B.1.420, también estuvieron presentes, en diferente frecuencia relativa en la muestra analizada.

Aunque se observaron cambios en la prevalencia de linajes a lo largo del tiempo, como el aumento del B.1.621 durante el tercer pico epidémico, y la incorporación del linaje P.1, no se estableció una correlación estadísticamente significativa entre los linajes y la gravedad de la enfermedad debido al tamaño limitado de la muestra y la falta de una distribución normal en los datos. La presencia de ciertos linajes colombianos, como B.1.111 y B.1.420, se atribuyen a la evolución del virus en la región.

Este estudio destaca la diversidad de linajes del SARS-CoV-2 en la ciudad de Bogotá en un periodo de ocho meses. Además, se destaca la importancia de monitorear la evolución genética del virus para comprender mejor su impacto clínico y la necesidad de seguir estudiando el restante de los exones del gen *ACE2* como de otros genes implicados en la interacción huésped- virus.

El análisis de los exones del receptor *ACE2* reveló una única variante sinónima en el exón 17 en cuatro individuos, sin asociación estadísticamente significativa con la gravedad de la enfermedad. Este estudio observó una mayor representación de hipertensión en individuos severos e hiperlipidemia en moderados, y encontró correlación de las enfermedades cardiovasculares con el linaje identificado.

10. Recomendaciones

La actual muestra puede presentar limitaciones para establecer correlaciones significativas. Se sugiere expandir el tamaño de la muestra, lo que mejoraría la representatividad y robustez estadística de los resultados. Además, se aconseja recopilar información demográfica más detallada, incluyendo el historial médico completo y otros factores relevantes. Un seguimiento a largo plazo de los pacientes permitiría evaluar la persistencia de los efectos de los linajes del virus y su evolución temporal. Para obtener una comprensión más completa, se propone analizar otros marcadores genéticos adicionales, además de los exones de ACE2, que puedan estar vinculados con la susceptibilidad o gravedad de la infección por SARS-CoV-2. La inclusión de factores de riesgo adicionales, como exposición ambiental y estado socioeconómico, sería esenciales para entender mejor la gravedad de la enfermedad. Finalmente, se destaca la importancia de compartir los resultados de manera transparente y accesible en la comunidad científica, facilitando así la revisión por pares y contribuyendo al conocimiento global sobre la relación genético-viral en la COVID-19.

Anexos

Tabla 8. Mutaciones de linajes de frecuencia relativa baja. Se describe el clado, el linaje, y 1. Sustituciones, 2.Delecciones, 3. Inserciones, 4. Cambios frameShifts, 5. aaSustituciones, 6. aaDelecciones, 7. aaInserciones, 8. Score de alineamiento, 9. Inicio del alineamiento, 10. Fin del alineamiento.

clade	Next clade pango	clade display	1	2	3	4	5	6	7	8	9	10	coverage	substitutions	deletions	frameShifts	aaSubstitutions	aaDeletions
20B	B.1.1	20B	17	0	0	0	12	0	0	89641	1	29903	0,94	A2653G,C3037T,C6539T,C6990T,G11417T,A11983G,C14408T,G20274T,C21614T,G21974C,A23403G,C23996T,A25315G,C26060T,G28881A,G28882A,G28883C	N:R203K,N:G204R,ORF1a:H2092Y,ORF1a:S2242F,ORF1a:V3718F,ORF1b:P314L,ORF1b:M2269I,ORF3a:T2231,S:L18F,S:D138H,S:D614G,S:P812S			
20A	B.1.625	20A	31	23	0	1	23	8	0	89487	1	29903	0,99	C241T,G521T,C2306T,C3037T,A4220G,A4495G,C5730T,C6401T,A10323G,C10775T,T12730G,A13712G,C14408T,C16694T,C18877T,C21846T,C22033A,T22882G,C22993T,G23012A,A23403G,G24410A,G25244T,G25563T,C25916T,T26767C,A27721G,A28272G,G28975T,G28979A,T29619A	11288-11296,21765-21770,21992-21994,27897-27901	ORF8:4-122	M:I82T,N:M234I,N:G236S,ORF1a:V86F,ORF1a:L681F,ORF1a:K1319E,ORF1a:T1822I,ORF1a:P2046S,ORF1a:K3353R,ORF1a:P3504S,ORF1b:K82R,ORF1b:P314L,ORF1b:T1076I,ORF3a:Q57H,ORF3a:T175I,ORF7a:I110V,S:T95I,S:F157L,S:N440K,S:E484K,S:D614G,S:D950N,S:V1228L	ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,ORF8:K2-,ORF8:F3-,S:H69-,S:V70-,S:Y144-
20B	B.1.1.348	20B	23	13	0	0	13	0	0	89572	1	29903	0,98	C241T,C3037T,A3790T,T6386A,G9802T,C10726T,G11417T,C11824T,C14408T,T15642C,G16647T,C17550T,A20288G,A23403G,G25062C,C25521T,G25913A,G27852A,C28278A,G28881A,G28882A,G28883C,C29370T	29761-29773		N:S2Y,N:R203K,N:G204R,N:T366I,ORF1a:L1175F,ORF1a:C2041S,ORF1a:V3718F,ORF1b:P314L,ORF1b:E2274G,ORF3a:G174D,ORF7b:E33K,S:D614G,S:G1167A	
20A	B.1.625	20A	30	23	0	1	25	8	0	89491	1	29903	0,99	C241T,G521T,C2306T,C3037T,A4220G,C5730T,C6401T,A10323G,C10775T,T12730G,A13712G,C14408T,C16468T,C16694T,C18877T,C21846T,A22008C,C22033A,A22407G,T22882G,G23012A,A23403G,G24410A,G25244T,G25563T,T26767C,A27721G,A28272G,G28975T,G29527T	11288-11296,21765-21770,21992-21994,27897-27901	ORF8:4-122	M:I82T,N:M234I,N:Q418H,ORF1a:V86F,ORF1a:L681F,ORF1a:K1319E,ORF1a:T1822I,ORF1a:P2046S,ORF1a:K3353R,ORF1a:P3504S,ORF1b:K82R,ORF1b:P314L,ORF1b:P1001S,ORF1b:T1076I,ORF3a:Q57H,ORF7a:I110V,S:T95I,S:N149T,S:F157L,S:N282S,S:N440K,S:E484K,S:D614G,S:D950N,S:V1228L	ORF1a:S3675-,ORF1a:G3676-,ORF1a:F3677-,ORF8:K2-,ORF8:F3-,S:H69-,S:V70-,S:Y144-
20A	B.1.243	20A	3	1	0	1	2	2	0	89687	1	29903	0,97	C14408T,T15642C,C28278A	11075	ORF1a:3605-4401	N:S2Y,ORF1b:P314L	ORF1a:F3603-,ORF1a:F3604-
19B	A.2.5	19B	36	9	0	0	21	3	0	89531	1	29903	0,98	T224C,C275T,C344T,C1567T,C3165T,C4543T,A5234G,A5488G,C8407T,C8782T,T9477A,C10029T,C11005A,C12815T,C13115T,C14805T,C16466T,T16857C,G17019T,T18417C,A22525G,T22917G,A23403G,G24697A,C25613T,G25979T,C26681T,T28144C,A28272T,C28291T,A28295T,C28657T,C28863T,G28975T,C29366T,C29421T	21983-21991		N:N8Y,N:S197L,N:M234I,N:P365S,N:P383L,ORF1a:L4F,ORF1a:L27F,ORF1a:A967V,ORF1a:K1657E,ORF1a:F3071Y,ORF1a:T325S,ORF1a:H3580Q,ORF1b:P1000L,ORF1b:E1184D,ORF3a:S74F,ORF3a:G196V,ORF8:L84S,ORF9b:P3L,ORF9b:K4N,S:L452R,S:D614G	S:L141-,S:G142-,S:V143-

11. Referencias

- Abdullaev, A., Abdurakhimov, A., Mirakbarova, Z., Ibragimova, S., Tsoy, V., Nuriddinov, S., Dalimova, D., Turdikulova, S., & Abdurakhmonov, I. (2022). Genome sequence diversity of SARS-CoV-2 obtained from clinical samples in Uzbekistan. *PLoS ONE*, *17*(6). <https://doi.org/10.1371/JOURNAL.PONE.0270314>
- Álvarez-Díaz, D. A., Ruiz-Moreno, H. A., Zapata-Bedoya, S., Franco-Muñoz, C., Laiton-Donato, K., Ferro, C., Sepulveda, M. T. H., Pacheco-Montealegre, M., Walteros, D. M., Carrero-Galindo, L. C., & Mercado-Reyes, M. (2022). Clinical outcomes associated with Mu variant infection during the third epidemic peak of COVID-19 in Colombia. *International Journal of Infectious Diseases*, *125*, 149–152. <https://doi.org/10.1016/j.ijid.2022.10.028>
- Angulo-Aguado, M., Corredor-Orlandelli, D., Carrillo-Martínez, J. C., Gonzalez-Cornejo, M., Pineda-Mateus, E., Rojas, C., Triana-Fonseca, P., Contreras Bravo, N. C., Morel, A., Parra Abaunza, K., Restrepo, C. M., Fonseca-Mendoza, D. J., & Ortega-Recalde, O. (2022). Association Between the LZTFL1 rs11385942 Polymorphism and COVID-19 Severity in Colombian Population. *Frontiers in Medicine*, *9*, 910098. <https://doi.org/10.3389/FMED.2022.910098/FULL>
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., ... Schloss, J. A. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68. <https://doi.org/10.1038/NATURE15393>
- Benetti, E., Tita, R., Spiga, O., Ciolfi, A., Birolo, G., Bruselles, A., Doddato, G., Giliberti, A., Marconi, C., Musacchia, F., Pippucci, T., Torella, A., Trezza, A., Valentino, F., Baldassarri, M., Brusco, A., Asselta, R., Bruttini, M., Furini, S., ... Pinto, A. M. (2020). ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. *European Journal of Human Genetics*, *28*(11), 1602–1614. <https://doi.org/10.1038/s41431-020-0691-z>
- BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol*. (2002).
- Biti, R., French, R. F., Young, J., Bennetts, B., Stewart, G., & Liang, T. (1997). HIV-1 infection in an individual homozygous for the CCR5 deletion allele. *Nature Medicine*, *3*(3), 252–253. <https://doi.org/10.1038/NM0397-252>

- Caicedo, J. D., Cáceres, A., Arboleda-Bustos, C. E., Mahecha, M. F., Ortega, J., Arboleda, G., & Arboleda, H. (2019). Genetic variability analysis in a population from Bogota: Towards a haplotype map. *Biomedica*, 39(3).
<https://doi.org/10.7705/BIOMEDICA.4753>
- 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., Manuli, E. R., Thézé, J., Almeida, L., Menezes, M. T., Voloch, C. M., Fumagalli, M. J., Coletti, T. M., da Silva, C. A. M., Ramundo, M. S., ... Faria, N. R. (2020). Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science (New York, N.y.)*, 369(6508), 1255–1260. <https://doi.org/10.1126/SCIENCE.ABD2161>
- Casalino, L., Gaieb, Z., Goldsmith, J. A., Hjorth, C. K., Dommer, A. C., Harbison, A. M., Fogarty, C. A., Barros, E. P., Taylor, B. C., McLellan, J. S., Fadda, E., & Amaro, R. E. (2020). Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein. *ACS Central Science*, 6(10), 1722–1734.
https://doi.org/10.1021/ACSCENTSCI.0C01056/SUPPL_FILE/OC0C01056_SI_006.ZIP
- Castillo, A. E., Parra, B., Tapia, P., Acevedo, A., Lagos, J., Andrade, W., Arata, L., Leal, G., Barra, G., Tambley, C., Tognarelli, J., Bustos, P., Ulloa, S., Fasce, R., & Fernández, J. (2020). Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *Journal of Medical Virology*, 92(9), 1562. <https://doi.org/10.1002/JMV.25797>
- Centro para el control y la prevención de enfermedades. (n.d.). *Clasificaciones y definiciones de las variantes del SARS-CoV-2*.
- Clinical Spectrum | COVID-19 Treatment Guidelines*. (n.d.). Retrieved January 5, 2024, from <https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/>
- Curtis, D. (2021). Variants in ACE2 and TMPRSS2 Genes Are Not Major Determinants of COVID-19 Severity in UK Biobank Subjects. *Human Heredity*, 85(2), 66–68.
<https://doi.org/10.1159/000515200>
- Deng, S. Q., & Peng, H. J. (2020). Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. In *Journal of Clinical Medicine* (Vol. 9, Issue 2). MDPI. <https://doi.org/10.3390/jcm9020575>
- Devaux, C. A., Rolain, J. M., & Raoult, D. (2020). ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. In *Journal of Microbiology, Immunology and Infection* (Vol. 53, Issue 3, pp. 425–435). Elsevier Ltd. <https://doi.org/10.1016/j.jmii.2020.04.015>
- Dieter, C., Brondani, L. de A., Leitão, C. B., Gerchman, F., Lemos, N. E., & Crispim, D. (2022). Genetic polymorphisms associated with susceptibility to COVID-19 disease and severity: A systematic review and meta-analysis. *PLOS ONE*, 17(7), e0270627.
<https://doi.org/10.1371/JOURNAL.PONE.0270627>

- Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., Fernández, J., Prati, D., Baselli, G., Asselta, R., Grimsrud, M. M., Milani, C., Aziz, F., Kässens, J., May, S., Wendorff, M., Wienbrandt, L., Uellendahl-Werth, F., Zheng, T., ... Karlsen, T. H. (2020). Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *New England Journal of Medicine*, 383(16), 1522–1534. <https://doi.org/10.1056/NEJMoa2020283>
- Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., da Candido, D. S., Mishra, S., E Crispim, M. A., S Sales, F. C., Hawryluk, I., McCrone, J. T., G Hulswit, R. J., M Franco, L. A., Ramundo, M. S., de Jesus, J. G., Andrade, P. S., Coletti, T. M., Ferreira, G. M., M Silva, C. A., Manuli, E. R., ... Sabino, E. C. (n.d.). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. In *Mélodie Monod* (Vol. 25). <http://pangolin.cog-uk.io>
- Fehr, A. R., & Perlman, S. (2015). Coronaviruses: An overview of their replication and pathogenesis. *Coronaviruses: Methods and Protocols*, 1282, 1–23. https://doi.org/10.1007/978-1-4939-2438-7_1/TABLES/2
- GISAID - gisaid.org. (n.d.). Retrieved December 7, 2023, from <https://gisaid.org/>
- Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pöhlmann, S. (2011a). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *Journal of Virology*, 85(9), 4122–4134. <https://doi.org/10.1128/jvi.02232-10>
- Glowacka, I., Bertram, S., Müller, M. A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T. S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., & Pöhlmann, S. (2011b). Evidence that TMPRSS2 Activates the Severe Acute Respiratory Syndrome Coronavirus Spike Protein for Membrane Fusion and Reduces Viral Control by the Humoral Immune Response. *Journal of Virology*, 85(9), 4122–4134. <https://doi.org/10.1128/JVI.02232-10>
- Gómez, S. A., Rojas-Valencia, N., Gómez, S., Cappelli, C., & Restrepo, A. (2021). The Role of Spike Protein Mutations in the Infectious Power of SARS-COV-2 Variants: A Molecular Interaction Perspective. *Chembiochem : A European Journal of Chemical Biology*. <https://doi.org/10.1002/cbic.202100393>
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., M Poon, L. L., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (n.d.). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*. <https://doi.org/10.1038/s41564-020-0695-z>

- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I., & Ziebuhr, J. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology*, *5*(4), 536–544. <https://doi.org/10.1038/S41564-020-0695-Z>
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L., Samborskiy, D., Sidorov, I. A., Sola, I., & Ziebuhr, J. (n.d.). *Severe acute respiratory syndrome-related coronavirus: The species and its viruses—a statement of the Coronavirus Study Group*. <https://doi.org/10.1101/2020.02.07.937862>
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O’Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., ... Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), 459–468. <https://doi.org/10.1038/s41586-020-2286-9>
- Greaney, A. J., Loes, A. N., Crawford, K. H. D., Starr, T. N., Malone, K. D., Chu, H. Y., & Bloom, J. D. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe*, *29*(3), 463. <https://doi.org/10.1016/J.CHOM.2021.02.003>
- Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics*, *34*(23), 4121–4123. <https://doi.org/10.1093/BIOINFORMATICS/BTY407>
- Harrison, S. C. (2008). Viral membrane fusion. *Nature Structural & Molecular Biology*, *15*(7), 690. <https://doi.org/10.1038/NSMB.1456>
- Hernández, M., García-Morán, E., Abad, D., & Eiros, J. M. (n.d.). *GISAID: INICIATIVA INTERNACIONAL PARA COMPARTIR DATOS GENÓMICOS DEL VIRUS DE LA GRIPE Y DEL SARS-CoV-2*. Retrieved December 3, 2023, from www.msccbs.es/resp
- Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., & Pöhlmann, S. (2014). TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. *Journal of Virology*, *88*(2), 1293–1307. <https://doi.org/10.1128/JVI.02202-13>

- Hirotsu, Y., & Omata, M. (2021). Discovery of a SARS-CoV-2 variant from the P.1 lineage harboring K417T/E484K/N501Y mutations in Kofu, Japan. *The Journal of Infection*, 82(6), 276. <https://doi.org/10.1016/J.JINF.2021.03.013>
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2), 271-280.e8. <https://doi.org/10.1016/j.cell.2020.02.052>
- Holmes, E. C., Goldstein, S. A., Rasmussen, A. L., Robertson, D. L., Crits-Christoph, A., Wertheim, J. O., Anthony, S. J., Barclay, W. S., Boni, M. F., Doherty, P. C., Farrar, J., Geoghegan, J. L., Jiang, X., Leibowitz, J. L., Neil, S. J. D., Skern, T., Weiss, S. R., Worobey, M., Andersen, K. G., ... Rambaut, A. (n.d.). *Leading Edge The origins of SARS-CoV-2: A critical review*. <https://doi.org/10.1016/j.cell.2021.08.017>
- Hu, B., Guo, H., Zhou, P., & Shi, Z.-L. (n.d.). *Characteristics of SARS-CoV-2 and COVID-19*. <https://doi.org/10.1038/s41579-020-00459-7>
- Hwang, S. S., Lim, J., Yu, Z., Kong, P., Sefik, E., Xu, H., Harman, C. C. D., Kim, L. K., Lee, G. R., Li, H. B., & Flavell, R. A. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.y.)*, 367(6483), 1260. <https://doi.org/10.1126/SCIENCE.ABB2507>
- Instituto Nacional de Salud. (n.d.). *Coronavirus Colombia*. Coronavirus Colombia. Retrieved January 26, 2024, from <https://www.ins.gov.co/Noticias/Paginas/Coronavirus.aspx>
- Instituto Nacional de Salud | Colombia Documentos tecnicos genomica. (n.d.). Retrieved January 23, 2024, from <https://www.ins.gov.co/Paginas/Documentos-tecnicos-genomica.aspx>
- Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E. E., Bhattacharya, T., Foley, B., Hastie, K. M., Parker, M. D., Partridge, D. G., Evans, C. M., Freeman, T. M., de Silva, T. I., Angyal, A., Brown, R. L., Carrilero, L., ... Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell*, 182(4), 812. <https://doi.org/10.1016/J.CELL.2020.06.043>
- Kuja, J. O., Kanoi, B. N., Balboa, R. F., Shiluli, C., Maina, M., Waweru, H., Gathii, K., Mungai, M., Masika, M., Anzala, O., Mwau, M., Clark, T. G., Waitumbi, J., & Gitaka, J. (2022). Genomic surveillance of SARS-COV-2 reveals diverse circulating variant lineages in Nairobi and Kiambu Counties, Kenya. *BMC Genomics*, 23(1). <https://doi.org/10.1186/S12864-022-08853-6>

- Laiton-Donato, K., Franco-Muñoz, C., Álvarez-Díaz, D. A., Ruiz-Moreno, H. A., Usme-Ciro, J. A., Prada, D. A., Reales-González, J., Corchuelo, S., Herrera-Sepúlveda, M. T., Naizaque, J., Santamaría, G., Rivera, J., Rojas, P., Ortiz, J. H., Cardona, A., Malo, D., Prieto-Alvarado, F., Gómez, F. R., Wiesner, M., ... Mercado-Reyes, M. (2021). Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. *Infection, Genetics and Evolution*, 95, 105038. <https://doi.org/10.1016/J.MEEGID.2021.105038>
- Laiton-Donato, K., Usme-Ciro, J. A., Franco-Muñoz, C., Álvarez-Díaz, D. A., Ruiz-Moreno, H. A., Reales-González, J., Prada, D. A., Corchuelo, S., Herrera-Sepúlveda, M. T., Naizaque, J., Santamaría, G., Wiesner, M., Walteros, D. M., Ospina Martínez, M. L., & Mercado-Reyes, M. (2021a). Novel Highly Divergent SARS-CoV-2 Lineage With the Spike Substitutions L249S and E484K. *Frontiers in Medicine*, 8, 697605. <https://doi.org/10.3389/FMED.2021.697605/FULL>
- Laiton-Donato, K., Usme-Ciro, J. A., Franco-Muñoz, C., Álvarez-Díaz, D. A., Ruiz-Moreno, H. A., Reales-González, J., Prada, D. A., Corchuelo, S., Herrera-Sepúlveda, M. T., Naizaque, J., Santamaría, G., Wiesner, M., Walteros, D. M., Ospina Martínez, M. L., & Mercado-Reyes, M. (2021b). Novel Highly Divergent SARS-CoV-2 Lineage With the Spike Substitutions L249S and E484K. *Frontiers in Medicine*, 8, 697605. <https://doi.org/10.3389/FMED.2021.697605/FULL>
- Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., & Wang, X. (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, 581(7807), 215–220. <https://doi.org/10.1038/S41586-020-2180-5>
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., ... Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *The New England Journal of Medicine*, 382(13), 1199. <https://doi.org/10.1056/NEJMOA2001316>
- Lieb, W., Graf, J., Götz, A., König, I. R., Mayer, B., Fischer, M., Stritzke, J., Hengstenberg, C., Holmer, S. R., Döring, A., Löwel, H., Schunkert, H., & Erdmann, J. (2006). Association of angiotensin-converting enzyme 2 (ACE2) gene polymorphisms with parameters of left ventricular hypertrophy in men. Results of the MONICA Augsburg echocardiographic substudy. *Journal of Molecular Medicine (Berlin, Germany)*, 84(1), 88–96. <https://doi.org/10.1007/S00109-005-0718-5>
- Lorenzo-Redondo, R., Nam, H. H., Roberts, S. C., Simons, L. M., Jennings, L. J., Qi, C., Achenbach, C. J., Hauser, A. R., Ison, M. G., Hultquist, J. F., & Ozer, E. A. (n.d.). A Unique Clade of SARS-CoV-2 Viruses is Associated with Lower Viral Loads in Patient Upper Airways. <https://doi.org/10.1101/2020.05.19.20107144>

- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Www.TheLancet.Com*, 395, 565. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Martínez-Gómez, L. E., Herrera-López, B., Martínez-Armenta, C., Ortega-Peña, S., Camacho-Rea, M. del C., Suarez-Ahedo, C., Vázquez-Cárdenas, P., Vargas-Alarcón, G., Rojas-Velasco, G., Fragoso, J. M., Vidal-Vázquez, P., Ramírez-Hinojosa, J. P., Rodríguez-Sánchez, Y., Barrón-Díaz, D., Moreno, M. L., Martínez-Ruiz, F. de J., Zayago-Angeles, D. M., Mata-Miranda, M. M., Vázquez-Zapién, G. J., ... López-Reyes, A. (2022). ACE and ACE2 Gene Variants Are Associated With Severe Outcomes of COVID-19 in Men. *Frontiers in Immunology*, 13, 1. <https://doi.org/10.3389/FIMMU.2022.812940/FULL>
- Medina-Enríquez, M. M., Lopez-León, S., Carlos-Escalante, J. A., Aponte-Torres, Z., Cuapio, A., & Wegman-Ostrosky, T. (2020). ACE2: the molecular doorway to SARS-CoV-2. In *Cell and Bioscience* (Vol. 10, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s13578-020-00519-8>
- Menachery, V. D., Yount, B. L., Debbink, K., Agnihothram, S., Gralinski, L. E., Plante, J. A., Graham, R. L., Scobey, T., Ge, X. Y., Donaldson, E. F., Randell, S. H., Lanzavecchia, A., Marasco, W. A., Shi, Z. L., & Baric, R. S. (2015). A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nature Medicine*, 21(12), 1508–1513. <https://doi.org/10.1038/nm.3985>
- Mio, C., Secco, C. D., Marzinotto, S., Pipan, C., Sozio, E., Tascini, C., Damante, G., & Curcio, F. (2022). Correspondence: Monitoring the SPREAD of the SARS-CoV-2 lineage B.1.621 in Udine, Italy. *Journal of Clinical Pathology*, 75(10), 712. <https://doi.org/10.1136/JCLINPATH-2021-207810>
- Nextclade*. (n.d.). Retrieved January 6, 2024, from <https://clades.nextstrain.org/>
- Nieto-Torres, J. L., DeDiego, M. L., Verdiá-Báguena, C., Jimenez-Guardeño, J. M., Regla-Nava, J. A., Fernandez-Delgado, R., Castaño-Rodríguez, C., Alcaraz, A., Torres, J., Aguilera, V. M., & Enjuanes, L. (2014). Severe Acute Respiratory Syndrome Coronavirus Envelope Protein Ion Channel Activity Promotes Virus Fitness and Pathogenesis. *PLoS Pathogens*, 10(5). <https://doi.org/10.1371/journal.ppat.1004077>
- Örd, M., Faustova, I., & Loog, M. (2020). 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. *Scientific Reports 2020 10:1*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-74101-0>

- Organización mundial de la salud. (n.d.). *Panel de control de la OMS sobre el coronavirus (COVID-19)*.
- Organization, W. H. (2021). *WHO-convened global study of origins of SARS-CoV-2: China part*.
- Park, S. T., & Kim, J. (2016). Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. *International Neurology Journal*, 20(Suppl 2), S76. <https://doi.org/10.5213/INJ.1632742.371>
- Pekar, J. E., Magee, A., Parker, E., Moshiri, N., Izhikevich, K., Havens, J. L., Gangavarapu, K., Mariana, L., Serrano, M., Crits-Christoph, A., Matteson, N. L., Zeller, M., Levy, J. I., Wang, J. C., Hughes, S., Lee, J., Ching, K., Yan, Z., Tzer, R., ... Wertheim, J. O. (n.d.). *The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2*. <https://www.science.org>
- Primer3. (n.d.). Retrieved January 5, 2024, from <https://primer3.org/>
- Propuestas aprobadas | ICTV. (n.d.). Retrieved December 8, 2023, from <https://ictv.global/files/proposals/approved?fid=4929>
- Quintana-Murci, L. (2016). Genetic and epigenetic variation of human populations: An adaptive tale. *Comptes Rendus Biologies*, 339(7–8), 278–283. <https://doi.org/10.1016/J.CRVI.2016.04.005>
- Rambaut, A., Holmes, E. C., OToole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (n.d.). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. <https://doi.org/10.1038/s41564-020-0770-5>
- Ravi, V., Saxena, S., & Panda, P. S. (2022). Basic virology of SARS-CoV 2. *Indian Journal of Medical Microbiology*, 40(2), 182. <https://doi.org/10.1016/J.IJMMB.2022.02.005>
- Rivero, R., Garay, E., Botero, Y., Serrano-Coll, H., Gastelbondo, B., Muñoz, M., Ballesteros, N., Castañeda, S., Patiño, L. H., Ramirez, J. D., Calderon, A., Guzmán, C., Martinez-Bravo, C., Aleman, A., Arrieta, G., & Mattar, S. (2022). Human-to-dog transmission of SARS-CoV-2, Colombia. *Scientific Reports*, 12(1). <https://doi.org/10.1038/S41598-022-11847-9>
- Sabater Molina, M., Nicolás Rocamora, E., Bendicho, A. I., Vázquez, E. G., Zorio, E., Rodríguez, F. D., Gil Ortuño, C., Rodríguez, A. I., Sánchez-López, A. J., Jara Rubio, R., Moreno-Docón, A., Marcos, P. J., García Pavía, P., Villa, R. B., & Gimeno Blanes, J. R. (2022). Polymorphisms in ACE, ACE2, AGTR1 genes and severity of COVID-19 disease. *PloS One*, 17(2), e0263140. <https://doi.org/10.1371/JOURNAL.PONE.0263140>

- Sahajpal, V., Rajput, S., Sharma, T., Sharma, A., & Thakar, M. K. (2019). Development and evaluation of a novel DNA purification buffer and protocol for blood samples on FTA cards. *Forensic Science International: Reports*, 1, 100014. <https://doi.org/10.1016/J.FSIR.2019.100014>
- Secuenciación genómica y caracterización genética del virus de la influenza | CDC.* (n.d.). Retrieved January 27, 2024, from <https://espanol.cdc.gov/flu/about/professionals/genetic-characterization.htm>
- Secuenciacion-genoma-completo-SARS-COV-2-secuenciador-MINION.* (n.d.).
- Senapati, S., Banerjee, P., Bhagavatula, S., Kushwaha, P. P., & Kumar, S. (2021). Contributions of human ACE2 and TMPRSS2 in determining host–pathogen interaction of COVID-19. In *Journal of Genetics* (Vol. 100, Issue 1). Springer. <https://doi.org/10.1007/s12041-021-01262-w>
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020a). Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020 581:7807, 581(7807), 221–224. <https://doi.org/10.1038/s41586-020-2179-y>
- Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., & Li, F. (2020b). Structural basis of receptor recognition by SARS-CoV-2. *Nature*, 581(7807), 221. <https://doi.org/10.1038/S41586-020-2179-Y>
- Shovlin, C. L., & Vizcaychipi, M. P. (2020). COVID-19 genomic susceptibility: Definition of ACE2 variants relevant to human infection with SARS-CoV-2 in the context of ACMG/AMP Guidance. *MedRxiv*, 2020.05.12.20098160. <https://doi.org/10.1101/2020.05.12.20098160>
- Siddell, S. G., Walker, P. J., Lefkowitz, E. J., Mushegian, A. R., Adams, M. J., Dutilh, B. E., Gorbalenya, A. E., Harrach, B., Harrison, R. L., Junglen, S., Knowles, N. J., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Nibert, M., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., ... Davison, A. J. (2019). Additional changes to taxonomy ratified in a special vote by the International Committee on Taxonomy of Viruses (October 2018). *Archives of Virology*, 164(3), 943–946. <https://doi.org/10.1007/S00705-018-04136-2>
- Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., & Bates, P. (2005). Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), 11876. <https://doi.org/10.1073/PNAS.0505577102>
- Simmons, G., Zmora, P., Gierer, S., Heurich, A., & Pöhlmann, S. (2013). Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting

- edge of antiviral research. *Antiviral Research*, 100(3), 605.
<https://doi.org/10.1016/J.ANTIVIRAL.2013.09.028>
- Singh, D., & Yi, S. V. (2021). On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, 53, 537–547. <https://doi.org/10.1038/s12276-021-00604-z>
- Situación del SARS CoV2 - Región de las Américas - OPS/OMS | Organización Panamericana de la Salud.* (n.d.). Retrieved January 13, 2024, from <https://www.paho.org/en/covid-19-weekly-updates-region-americas>
- Sun, P., Lu, X., Xu, C., Sun, W., & Pan, B. (2020). Understanding of COVID-19 based on current evidence. *Journal of Medical Virology*, 92(6), 548–551.
<https://doi.org/10.1002/JMV.25722>
- Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E. J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A. J., Engelbrecht, S., Van Zyl, G., ... de Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*, 592(7854), 438–443. <https://doi.org/10.1038/S41586-021-03402-9>
- The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. (2020). *European Journal of Human Genetics*, 28(6), 715–718.
<https://doi.org/10.1038/s41431-020-0636-6>
- Thomas, S. (2021). Mapping the Nonstructural Transmembrane Proteins of Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Computational Biology*, 28(9), 909–921. <https://doi.org/10.1089/cmb.2020.0627>
- Torre-Fuentes, L., Matías-Guiu, J., Hernández-Lorenzo, L., Montero-Escribano, P., Pytel, V., Porta-Etessam, J., Gómez-Pinedo, U., & Matías-Guiu, J. A. (2021). ACE2, TMPRSS2, and Furin variants and SARS-CoV-2 infection in Madrid, Spain. *Journal of Medical Virology*, 93(2), 863–869. <https://doi.org/10.1002/jmv.26319>
- Tortorici, M. A., & Vesler, D. (2019). Structural insights into coronavirus entry. In *Advances in Virus Research* (Vol. 105, pp. 93–116). Academic Press Inc.
<https://doi.org/10.1016/bs.aivir.2019.08.002>
- (US), N. I. of H., & Study, B. S. C. (2007). *Understanding Human Genetic Variation*. <https://www.ncbi.nlm.nih.gov/books/NBK20363/>
- Vadgama, N., Kreymerman, A., Campbell, J., Shamardina, O., Brugger, C., Research Consortium, G. E., Deaconescu, A. M., Lee, R. T., Penkett, C. J., Gifford, C. A., Mercola, M., Nasir, J., & Karakikes, I. (2022). SARS-CoV-2 Susceptibility and ACE2 Gene Variations Within Diverse Ethnic Backgrounds. *Frontiers in Genetics*, 13, 888025. <https://doi.org/10.3389/FGENE.2022.888025/BIBTEX>

- Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., ... Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, *593*(7858), 266–269. <https://doi.org/10.1038/S41586-021-03470-X>
- Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Velesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, *181*(2), 281. <https://doi.org/10.1016/J.CELL.2020.02.058>
- Wang, N., Li, S. Y., Yang, X. Lou, Huang, H. M., Zhang, Y. J., Guo, H., Luo, C. M., Miller, M., Zhu, G., Chmura, A. A., Hagan, E., Zhou, J. H., Zhang, Y. Z., Wang, L. F., Daszak, P., & Shi, Z. L. (2018). Serological Evidence of Bat SARS-Related Coronavirus Infection in Humans, China. *Virologica Sinica*, *33*(1), 104. <https://doi.org/10.1007/S12250-018-0012-7>
- WHO COVID-19 weekly epidemiological update vol. 74... - Google Académico. (n.d.). Retrieved January 2, 2024, from <https://scholar.google.com/scholar?q=WHO+COVID-19+weekly+epidemiological+update+vol.+74+January+2022+World+Health+organization>
- Wolf, J. M., Wolf, L. M., Bello, G. L., Maccari, J. G., & Nasi, L. A. (2023). Molecular evolution of SARS-CoV-2 from December 2019 to August 2022. *Journal of Medical Virology*, *95*(1). <https://doi.org/10.1002/JMV.28366>
- Wong, S. K., Li, W., Moore, M. J., Choe, H., & Farzan, M. (2004). A 193-Amino Acid Fragment of the SARS Coronavirus S Protein Efficiently Binds Angiotensin-converting Enzyme 2. *Journal of Biological Chemistry*, *279*(5), 3197–3201. <https://doi.org/10.1074/jbc.C300520200>
- Wu, C. rong, Yin, W. chao, Jiang, Y., & Xu, H. E. (2022). Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19. In *Acta Pharmacologica Sinica* (Vol. 43, Issue 12, pp. 3021–3033). Springer Nature. <https://doi.org/10.1038/s41401-021-00851-w>
- Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2008). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*. <https://doi.org/10.1038/s41586-020-2008-3>

- Xiao, X., Newman, C., Buesching, C. D., Macdonald, D. W., & Zhou, Z.-M. (2020). Animal sales from Wuhan wet markets immediately prior to the COVID-19 pandemic. *Scientific Reports*, *11*, 11898. <https://doi.org/10.1038/s41598-021-91470-2>
- Xie, X., Liu, Y., Liu, J., Zhang, X., Zou, J., Fontes-Garfias, C. R., Xia, H., Swanson, K. A., Cutler, M., Cooper, D., Menachery, V. D., Weaver, S. C., Dormitzer, P. R., & Shi, P. Y. (2021). Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nature Medicine*, *27*(4), 620–621. <https://doi.org/10.1038/S41591-021-01270-4>
- Yadav, R., Chaudhary, J. K., Jain, N., Chaudhary, P. K., Khanra, S., Dhamija, P., Sharma, A., Kumar, A., & Handu, S. (2021). *cells Role of Structural and Non-Structural Proteins and Therapeutic Targets of SARS-CoV-2 for COVID-19*. *10*, 821. <https://doi.org/10.3390/cells>
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. In *Science* (Vol. 367).
- Yang, J., Petitjean, S. J. L., Koehler, M., Zhang, Q., Dumitru, A. C., Chen, W., Derclaye, S., Vincent, S. P., Soumillion, P., & Alsteens, D. (2020). Molecular interaction and inhibition of SARS-CoV-2 binding to the ACE2 receptor. *Nature Communications*, *11*(1). <https://doi.org/10.1038/s41467-020-18319-6>
- Yoshimoto, F. K. (2020). The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *The Protein Journal*, *39*, 198–216. <https://doi.org/10.1007/s10930-020-09901-4>
- Zeberg, H., & Pääbo, S. (2020). The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* *2020* *587*:7835, *587*(7835), 610–612. <https://doi.org/10.1038/s41586-020-2818-3>
- Zhu, L., Marsh, J. W., Griffith, M. P., Collins, K., Srinivasa, V., Waggle, K., Van Tyne, D., Snyder, G. M., Phan, T., Wells, A., Marroquin, O. C., & Harrison, L. H. (2022). Predictive model for severe COVID-19 using SARS-CoV-2 whole-genome sequencing and electronic health record data, March 2020-May 2021. *PLoS ONE*, *17*(7). <https://doi.org/10.1371/JOURNAL.PONE.0271381>
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W., & China Novel Coronavirus Investigating and Research Team. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England Journal of Medicine*, *382*(8), 727–733. <https://doi.org/10.1056/NEJMoa2001017>