

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE MEDICINA
ESCUELA DE MEDICINA HUMANA



**Variaciones de aminoácidos de la proteína spike de SARS-CoV-2
secuenciados en América del Sur**

**TESIS PARA OPTAR EL TÍTULO DE
MÉDICO CIRUJANO**

AUTOR

Karoline Janet Olazabal Mancilla

ASESOR

Erick Giancarlo Suclupe Farro

<https://orcid.org/0000-0002-0334-2191>

Chiclayo, 2023

**Variaciones de aminoácidos de la proteína spike de SARS-CoV-2
secuenciados en América del Sur**

PRESENTADA POR

Karoline Janet Olazabal Mancilla

A la Facultad de Medicina de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

MÉDICO CIRUJANO

APROBADA POR

Cesar Armando Ñique Carbajal

PRESIDENTE

Victor Daniel Linares Baca
SECRETARIO

Erick Giancarlo Suclupe Farro
VOCAL

TESIS KAROLINE OLAZABAL MANCILLA

INFORME DE ORIGINALIDAD

17%

INDICE DE SIMILITUD

15%

FUENTES DE INTERNET

14%

PUBLICACIONES

%

TRABAJOS DEL
ESTUDIANTE

FUENTES PRIMARIAS

1	Rosas Moreno América Denisse. "Aspectos virológicos, clínicos, diagnósticos y terapéuticos de los coronavirus causantes de enfermedades graves en humanos; SARS-COV-1, MERS-COV, SARS-COV-2", TESIUNAM, 2022 Publicación	3%
2	www.flickr.com Fuente de Internet	2%
3	www.scielo.org.pe Fuente de Internet	2%
4	tesis.usat.edu.pe Fuente de Internet	1%
5	repositorio.unal.edu.co Fuente de Internet	1%
6	Lalitha Guruprasad. "HUMAN SARS CoV-2 SPIKE PROTEIN MUTATIONS", American Chemical Society (ACS), 2020 Publicación	1%

Índice

Resumen	5
Abstract	6
Introducción.....	7
Revisión de literatura	10
Materiales y métodos	16
Resultados y discusión	18
Conclusiones	28
Recomendaciones	29
Referencias.....	30
Anexos	36

Resumen

Introducción: La proteína spike está implicada en el tropismo celular, la patogenia y la antigenicidad cambiante de los coronavirus; son conocidas las secuencias de nucleótidos y aminoácidos del SARS-CoV-2, así como sus variaciones que este nuevo coronavirus presenta en relación con otros tipos de coronavirus, pero no se han descrito las variaciones de los aminoácidos de la proteína spike en los diferentes países de América del Sur donde ha sido secuenciado el genoma de SARS-CoV-2. **Objetivo:** Describir las variaciones de aminoácidos de la proteína spike en genomas del SARS-CoV-2 secuenciados en América del Sur. **Materiales y métodos:** Estudio observacional, descriptivo, retrospectivo, transversal. **Población:** Genes de la proteína spike de secuenciamientos completos en países de América del Sur publicados en la base de datos GISAID hasta el 20 de septiembre del 2021. **Procedimiento:** Se realizó un análisis bibliográfico científico para obtener una visión general de las mutaciones en la proteína spike del SARS-CoV-2 de las cepas circulantes en América del Sur; para lo cual se descargó de la base de datos de GISAID, la secuencia de aminoácidos de la proteína spike natural de Wuhan-China y todas las secuencias completas de proteína spike del genoma de SARS-Cov-2 secuenciados en América del Sur; luego se realizó un filtro de secuencias con el uso de un código escrito en Python y un alineamiento múltiple de las secuencias de aminoácidos de la proteína spike con el uso de un software de acceso libre MAFFT; posteriormente se descargaron y visualizaron las secuencias con ayuda del programa Jalview, se realizó un análisis e interpretación de datos y finalmente se crearon tablas y gráficos que permiten observar las variaciones de los aminoácidos de la proteína spike. Los resultados describen la mayor cantidad de variaciones en los aminoácidos que se encuentran en la región bisagra llamada RBD de la proteína spike.

Palabras clave: proteína spike, SARS-CoV-2, América del Sur (Fuente: DeCS BIREME)

Abstract

Introduction: The spike protein is involved in cell tropism, pathogenesis, and changing antigenicity of coronaviruses; The nucleotide and amino acid sequences of SARS-CoV-2 are known, as well as its variations that this new coronavirus presents in relation to other types of coronaviruses, but have not been described the amino acids variations of the spike protein in different countries of South America where the SARS-CoV-2 genome has been sequenced.

Objective: To describe the amino acid variations of the spike protein in SARS-CoV-2 genomes sequenced in South America. **Materials and methods:** Observational, descriptive, retrospective, cross-sectional study. The population will be constituted by spike protein genes from complete sequencings in South American countries published in the GISAID database until September 20, 2021. **Procedure:** A scientific literature review was performed to obtain an overview of spike protein mutations SARS-CoV-2 circulating strains in South America; for which the amino acid sequence of the natural spike protein from Wuhan-China and all the complete spike protein sequences of the SARS-Cov-2 genome sequenced in South America were downloaded from the GISAID database; then a sequence filter was performed with the use of a code written in Python and a multiple alignment of the amino acid sequences of the spike protein with the use of free access software MAFFT; Subsequently, the sequences were downloaded and visualized with the help of the Jalview program, an analysis and interpretation of the data was carried out and finally, tables and graphs were created that allow observing the variations of the spike protein amino acids. Results describe the largest number of variations in the amino acids found in the hinge region called RBD of the spike protein.

Keywords: protein spike, SARS-CoV-2, South American

Introducción

Los coronavirus atentan potencialmente contra la vida de los animales, incluyendo la de los seres humanos. En los años 2002 y 2003 el coronavirus responsable del conocido síndrome respiratorio agudo severo o SARS-CoV, obtuvo una tasa de mortalidad del 10% al lograr infectar a más de ocho mil personas; a finales del 2019 en Wuhan surgió un nuevo coronavirus, el virus SARS-CoV-2 que ocasionó la llegada de una nueva enfermedad con una alta capacidad infectiva que se extendió con mucha rapidez en todo el mundo, la COVID-19, la cual varía desde una enfermedad leve y autolimitante del sistema respiratorio hasta una neumonía grave, pudiendo llevar al paciente a una falla multiorgánica o la muerte.⁽¹⁻⁵⁾

A inicios de febrero del 2020 se publicaron reportes de la Organización Mundial de la Salud (OMS) en donde informaban sobre la gran velocidad de transmisión de este nuevo virus de persona a persona con más de treinta mil infecciones humanas confirmadas e incluso más de seiscientos cuarenta muertes, por lo que la OMS lo declararon como una emergencia de salud pública de preocupación internacional y fue el 10 de abril del mismo año en el que se notificaron 1.600.427 casos confirmados por laboratorio, con una tasa de mortalidad del 5,98% en 184 países de todo el mundo; fue gracias a esto que el 11 de marzo de 2020, la OMS lo declararon como una pandemia, la primera causada por los coronavirus a lo largo de la historia.⁽¹⁻⁵⁾

En efecto, los coronavirus además de destacar por sus genomas que presentan más de 30 000 pares de bases, también destacan por la frecuencia con la que se recombinan, intercambiando fragmentos de su ARN con otros coronavirus, lo cual es un “truco especial” que les otorga un dinamismo mortal; lo usual es un intercambio sin sentido de partes similares entre virus similares; pero cuando dos parientes lejanos del coronavirus terminan en la misma célula, la recombinación puede conducir a versiones formidables que infectan nuevos tipos de células y saltan a otras especies; este dinamismo se evidenció ya que en los primeros tres meses de la pandemia, la base de datos *GISAID* ya disponía de tres grandes grupos de mutaciones en las proteínas vitales del SARS-CoV-2, lo cual sugiere que este virus es muy venerable por tener cambios rápidos y mutar incluso durante la transmisión de una persona a otra, las posteriores mutaciones del SARS-CoV-2 o polimorfismo puntual llamado también polimorfismo de un solo nucleótido o en sus siglas SNP, muestran la complejidad y rápida evolución de este virus.⁽⁸⁾

Los SNP ocurrieron principalmente en la proteína spike, en la proteína no estructural (Nsp1) y en la ARN polimerasa dependiente de ARN (RdRp); las mutaciones de la proteína spike ocurrieron en la superficie y en las regiones ACE3 y CD26, las cuales son responsables de la entrada del virus al huésped al mediar la unión del receptor del huésped con el sitio de escisión de Furina del virus, estas mutaciones podrían hacer que el virus sea más resistente a los inhibidores de unión al receptor y a algunos de los anticuerpos. En sí, la mayoría de las mutaciones fueron expuestas a la proteína spike y muestran la alta posibilidad de impedir el efecto de los medicamentos antivirales que aunque hasta el momento no se dispone de un tratamiento adecuado o específico para el SARS-CoV-2, las mutaciones encontradas son de gran importancia tanto para la virulencia como para la resistencia y respuesta de este virus a los diferentes antivirales, por lo cual el estudio de las mutaciones en estas proteínas virales claves son de gran importancia.⁽⁸⁾

La proteína spike es una proteína estructural del SARS-CoV-2 que desempeña un papel esencial al mediar la entrada de este virus al huésped ya que posee una unidad llamada dominio S de unión al receptor que hace que estos receptores se unan a las células diana del huésped lo que es fundamental para el ingreso del virus a las células humanas, este dominio de unión del SARS-CoV-2 es particularmente eficiente y diferente de otros coronavirus, los cuales no son capaces de infectar a los seres humanos; la presencia de un motivo distinto en el S1 / S2 de la región de unión sugiere que la posible adquisición de sitios de escisión en la proteína spike fue lo que promovió la transmisión entre especies y es por ello que es importante analizar las variaciones de aminoácidos de la proteína spike ya que esta es un punto clave en la patogenia del virus, determina el tropismo y cambios conformacionales, entonces las mutaciones en esta proteína condujo con gran probabilidad a la antigenicidad cambiante del virus, sin embargo hasta la fecha no se dispone de un estudio sobre las mutaciones en la secuencia de aminoácidos de la proteína spike, a pesar que, la identificación de estas es importante y por ello debe ser investigada por otros estudios para una mayor comprensión de su árbol filogenético y su evolución.^(7,11,13)

Se realizaron estudios sobre las mutaciones del SARS-CoV-2 y se menciona que entre los 12 países diferentes (una secuencia de cada país) con datos completos de secuenciación del genoma, notaron 47 mutaciones de puntos clave o SNP ubicados a lo largo de todo el genoma que podrían tener un impacto en la virulencia y la respuesta a diferentes antivirales contra el SARS-CoV-2 . En este sentido, las proteínas virales clave como la proteína spike, Nsp1,

RdRp y la región ORF8 se mutaron fuertemente en los primeros tres meses de pandemia a través del paso de persona a persona ^[11]

Otro estudio en el que se alinearon secuencias de nucleótidos reveló que de 93 mutaciones en genomas completos de SARS-CoV-2, las principales son ocho de la proteína spike y tres mutaciones (D354, Y364 y F367) que se encuentran en el dominio de unión al receptor de la proteína spike.⁽⁹⁾

Los resultados del presente trabajo serán de ayuda para comprender qué tanto han variado las mutaciones de las secuencias de aminoácidos de la proteína spike del SARS-CoV-2, puesto que esta proteína es una máquina molecular multifuncional que media la entrada del virus al huésped y se cree que las mutaciones de esta, es la que le confiere al SARS-CoV-2 la capacidad de poder infectar a las personas a diferencia de los otros coronavirus; además, esta proteína spike contribuye al tropismo celular, a la patogenia y la antigenicidad cambiante de este virus; al mismo tiempo nos permitirán comprender mejor la evolución de este virus en cada país de América del Sur en los cuales se haya secuenciado el genoma completo del SARS-CoV-2, nos permitirá conocer las secuencias y establecer consensos en cuanto a los aminoácidos, comparar la forma que estos se mantienen o varían y en qué porcentaje lo hacen ya que por lo que se conoce hasta el momento lo que se sabe de su evolución es que a nivel vírico varía mucho y su virulencia aumenta debido a su gran dinamismo y constante intercambio de ARN con otros coronavirus, pero no se conoce qué tanto varían los aminoácidos de la proteína spike del SARS-CoV-2 en poblaciones con diferentes áreas geográficas.

Por ello, ¿Cuáles son las variaciones de los aminoácidos de la proteína spike en genomas de SARS-CoV-2 secuenciados en América del Sur? Se planteó como objetivo general describir las variaciones de aminoácidos de la proteína spike en genomas del SARS-CoV-2 secuenciados en América del Sur.

Revisión de literatura

El SARS-CoV-2, es un virus ARN monocatenario envuelto en sentido positivo; se conocían dos tipos: el Síndrome respiratorio agudo severo por Coronavirus tipo 1 (SARS-CoV) y en los humanos, el síndrome respiratorio del Medio Oriente (MERS-CoV) que en el 2012 fue el que causó una epidemia en la Península Arábiga infectando a más de 1700 personas, con una tasa de mortalidad de 36%.⁽⁸⁻¹⁰⁾

El SARS-CoV-2 tienen en común un 79 % de identidad de secuencia genómica con el SARS-CoV y un 50 % con el MERS-CoV, además está comprobado que filogenéticamente comparte genoma con los coronavirus vinculados con el SARS (SARSr-CoV) que se hallan en los murciélagos, lo que lo ubica en el subgénero Sarbecovirus del género Betacoronavirus, los cuales tienen el genoma más grande (entre 27 y 32 kb) de todos los virus ARN y este genoma está formado por una proteína de la nucleocápside y rodeada por una envoltura de tres proteínas estructurales: la proteína de envoltura involucrada en el ensamblaje del virus, la proteína de membrana y la proteína spike responsable de la entrada del virus en las células huésped. El SARS-CoV-2 y el SARS-CoV tienen en común los mismos genes que codifican sus proteínas estructurales y comparten más del 90 % de identidad de aminoácidos, excepto la proteína spike, la cual forma protuberancias en la superficie del virus, dándole el aspecto de corona y se le acuñó el nombre original "coronavirus".⁽¹⁰⁻¹³⁾.

La proteína spike del SARS-CoV-2 tiene 1273 aminoácidos siendo más larga que la proteína spike del SARS-CoV que cuenta con 1255 aminoácidos y la del SARSr-CoV de murciélago que tiene de 1245 aminoácidos; la secuencia de aminoácidos de la proteína spike tienen similitudes de hasta un 76,7 a 77,0 % con SARS-CoV de civetas y humanos, 75 a 97,7 % con coronavirus de murciélago y 90,7 a 92,6 % con coronavirus de pangolín; Esta proteína es un miembro de las proteínas de fusión de membrana viral de clase I que ayuda a la unión del receptor a la célula, el recorrido del virus en los tejidos y la nosogénesis; es una máquina molecular multifuncional que interviene en la entrada del virus al huésped.⁽¹⁴⁾

La estructura de la proteína spike, contiene tres segmentos: un ectodominio, un ancla transmembrana de un solo paso y una cola intracelular corta; el ectodominio se divide en subunidades: S1 N-terminal, la cual se une a un receptor de la célula y en S2 C-terminal que se encarga de la entrada del virus.⁽¹⁷⁻²⁰⁾ La S1 se une al receptor por la enzima convertidora de

angiotensina 2 o ACE2, que se encuentra sobre todo en las vías respiratorias del ser humano y esto es lo que le permite al virus poder invadir la célula.⁽⁶⁻⁸⁾

En el dominio de unión al receptor (RBD) de la proteína spike, la igualdad de aminoácidos entre el SARS-CoV-2 y el SARS-CoV es solo del 73 % por lo que se cree que las mutaciones de la proteína spike son responsables de la evolución del virus que pasó de infectar a murciélagos y otras especies, a infectar a seres humanos; Otra característica genómica específica del SARS-CoV-2 es la inserción de cuatro residuos de aminoácidos (PRRA) en la unión de las subunidades S1 y S2 de la proteína spike, las cuales son procesadas por las proteasas de la célula huésped posterior a la unión de la subunidad S1 al receptor de la célula del huésped y la fusión de membrana del virus a través de la subunidad S2, haciendo que la proteína spike medie no sólo la entrada del virus al huésped, sino también intervenga en el tropismo a nivel celular e influya en la respuesta inmune de este.⁽¹⁴⁻²⁰⁾

El Centro Nacional de Recursos de Bioinformación alineó 77 801 secuencias genómicas de SARS-CoV-2 detectadas a nivel mundial para valorar qué tanto variaba genéticamente las diferentes cepas del nuevo coronavirus e identificó un total de 15 018 mutaciones.⁽²³⁾ En la proteína spike e identificaron cuatro mutaciones de aminoácidos: V483A, L455I, F456V y G476S, de las cuales se conoce su ubicación, cerca de la unión del RBD, pero se desconocen sus efectos sobre este en el huésped; La alteración en la subunidad S1 de D614G se encontró con mucha más frecuencia que otros respecto a las variantes spike, y es el marcador de un subclado principal de SARS-CoV-2 (clado G).⁽²⁴⁾ La variante dominante de la proteína spike del SARS-CoV-2 que predomina a nivel mundial desde marzo del 2020 es la G614, la cual reemplazó a la variante original D614, en comparación con esta última, las cargas virales de la G614 son mucho más altas en los pacientes infectados, pero la clínica no sugiere una relación significativa entre esta variante predominante y la gravedad de la enfermedad⁽²²⁻²⁵⁾

Debido a estas variaciones, la interacción del SARS-CoV-2 con su receptor estabiliza los dos puntos críticos de unión al virus en la superficie de ACE2; Además, un motivo de cuatro variaciones en los aminoácidos 482–485: GVEG de la membrana de unión al receptor del SARS-CoV-2 da como resultado una conformación más compacta de su unión a la ACE2. La literatura confirma que las propiedades estructurales del dominio de unión del receptor del SARS-CoV-2 han consolidado su afinidad de unión a la ACE2 a diferencia de la del SARS-CoV; Además en los secuenciamientos del ARN se evidenció que TMPRSS2 se expresa con frecuencia en varios tejidos del cuerpo y se expresa con ACE2 en células del epitelio nasal,

pulmonar y bronquial, lo que explica el dinamismo celular de la proteína spike del SARS-CoV-2 para la evasión inmune.⁽²⁵⁻²⁶⁾

Pokhrel et al. (2023), en una investigación que tuvo como metodología la recolección de información bibliográfica científica a partir de la base de datos GISAID, determinó que la proteína SARS-CoV-2 S tiene 1273 aminoácidos de largo; contiene un péptido señal (aminoácidos 1 a 13), la subunidad S1 (14 a 685 residuos) que interviene en la unión al receptor y finalmente la subunidad S2 (686 a 1 273 residuos) que tiene que ver con la fusión de membranas. A diferencia de la secuencia índice WIV04 (MN996528.1, también conocida como variante de Wuhan o virus índice) de febrero de 2020, la proteína spike de 1273 aminoácidos tenía 3540 variantes. Este número de variantes solo incluye secuencias filtradas (441.168) que están completas y no contienen un número anormal de mutaciones. Debido que existen unas 3540 variantes, en promedio, cada posición en la secuencia de proteína de 1273 aminoácidos tiene aproximadamente tres variantes. Sin embargo, algunas regiones albergan 9 variantes en una sola posición de aminoácido, mientras que otras no tienen variantes, habiendo un total de 123 posiciones que son totalmente invariables⁽²⁷⁾.

Gómez (2022), en un estudio para el que empleó la recolección de artículos científicos, encontró hasta 13 variaciones de interés en lo que respecta al virus SARS-Cov-2, siendo las cepas más relevantes las de Brasil, Inglaterra y la Sudafricana, encontrando además que las cepas más influyentes en Colombia fueron las B.1.1.529 B.1.1.7, B.1.3.5.1, B.1.617.2, B.1.621, y P.1.⁽²⁸⁾

Núñez y Zabala (2023) mediante revisión bibliográfica, encontró que la variante Gamma, linaje P.1, contienen mutaciones en la proteína spike, generando resistencia en anticuerpos neutralizantes, lo que se relaciona a una propagación rápida de la enfermedad; la variante Delta, linaje B.1.617.2 también presenta mutaciones en la proteína spike (S): D111D, D614G, E484Q, G142D, L452R y P681R; en relación a la variante Ómicron, linaje B.1.1.529 en la proteína S se encuentran 37 mutaciones, de las que 26 son específicas a la variante en las demás guardan relación con las variantes alfa, beta y gamma, las que presentan evasión inmune y eficaz transmisión; la variante Mu, linaje B.1.621 es identificable a partir de la proteína S y tiene importancia fenotípica por su evasión inmune e infectividad (E484K, N501Y y P681H); la variable Eta linaje B.1.52 presenta las siguientes mutaciones en la

proteína S: D614G, E484K y Q677H. Dentro de los países más infectados con COVID en Latinoamérica (año 2022), tenemos, por orden de casos: Brasil con 34.477.539, Argentina con 9.689.861; México con 7.046.220, Colombia con 6.304.317, Chile con 4.549.927 y Perú con 4.121.036. ⁽²⁹⁾

De Groot et al. (2020), realizaron estudios sobre las mutaciones del SARS-CoV-2 llegándose a determinar luego de una exhaustiva recolección de material bibliográfico especializado que entre los 12 países diferentes (una secuencia de cada país) con datos completos de secuenciación del genoma, existían 47 mutaciones de puntos clave o SNP ubicados a lo largo de todo el genoma, que podrían tener un impacto en la virulencia y la respuesta a diferentes antivirales contra el SARS-CoV-2. En este sentido, las proteínas virales clave como la proteína spike, Nsp1, RdRp y la región ORF8 se mutaron fuertemente en los primeros tres meses de pandemia a través del paso de persona a persona ^[11]

Wuz y McGoogan (2019), en un estudio en el que emplearon la metodología de recopilación bibliográfica, donde se alinearon secuencias de nucleótidos revelaron que de 93 mutaciones en genomas completos de SARS-CoV-2, las principales son ocho de la proteína spike y tres mutaciones (D354, Y364 y F367) que se encuentran en el dominio de unión al receptor de la proteína spike. ⁽⁷⁾

Aguilar (2021), haciendo uso de metodología enfocada en el empleo de material bibliográfico para su posterior análisis, por la que seleccionó hasta 74 fuentes extraída de diversas bases de datos, entre ellas Medline y Pub Med., extrajo como conclusiones que la proteína spike y los cambios funcionales y estructurales que ella sufre son responsables de la elevada diversidad que presenta el genoma SARS-CoV-2, incidiendo directamente en los cambios epidemiológicos de dicha enfermedad originando sus diversas variantes, clados y linajes. Además, encontró que las variantes Lambda y Mu eran las que revestían mayor interés dentro del territorio peruano. Pudo identificar además que, dentro de la Variante Alfa, Linaje B.1.1.7., 3 mutaciones encontradas en la proteína spike fueron los responsables de efectos potenciales biológicos: N501Y, P681H y la delación 69/70; en la Variante Beta, Linaje B.1.351, las mutaciones en la proteína spike están en N501Y y las E464K, K417N que son las que más preocupan por la incidencia que pueden tener en la actividad neutralizante propia de los anticuerpos. ⁽³⁰⁾

Expósito (2021), cuya investigación estuvo basada en el método de recolección bibliográfica científica a partir de la secuenciación del genoma, encontró que, en América del Sur,

concretamente en Perú, se presentó la variante Lambda o C.37, encontrándose delación D246-252 y las mutaciones no sinónimas D614G, F490S, G75V, L452Q T76I y T859N dentro de la proteína spike. Asimismo, dio cuenta de la Variante Gamma o brasileña B.1.1.28.1 o P.1 o 20J/501Y.V3 siendo detectada posteriormente hasta en 52 países, siendo el 10% de casos de las infecciones habidas en los Estados Unidos en junio del 2021. Además, se dio cuenta de hasta 17 mutaciones no sinónimas en la proteína spike: L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I, and V1176F, observándose así que esta variante es la que más mutaciones tiene en relación a la proteína S.⁽³¹⁾

Loayza (2021), para cuya investigación empleó revisión bibliográfica, encontró que dentro de 25000 secuenciamientos genómicas analizados, pudo ser identificada la mutación D614G, la misma que desplaza el ácido aspártico con glicina dentro de la posición 614 en la proteína spike vírica, la misma a la que se asocia carga viral elevada en pacientes más jóvenes.⁽³²⁾

Vargas (2021), haciendo empleo de material bibliográfico, encontró que las mutaciones Lambda T76I y L452Q encontradas en la proteína spike, podrían ser responsables de mayores tasas de infectividad; mientras que F490S, RSYLTSGD246-253N, y particularmente L452Q, muy similar a la mutación de la variante Delta L452R, le darían a Lambda la propiedad de evadir el sistema inmunológico. Esto podría explicar en parte la expansión y predominio de la variable Lambda en Perú durante la segunda ola de la enfermedad por coronavirus (COVID-19), pese a la presencia de las otras variantes altamente transmisibles. Por otro lado, Lambda ha mostrado menor variabilidad en sólo un sub-linaje (C.37.1 encontrado en Europa), en tanto que Gamma presenta 10 sub-linajes portadores de mutaciones de la proteína spike, 3 de los cuales circulan en Perú (P.1.1, P.1.4 y P.1.7).⁽³³⁾

Romero et al. (2021), empleando revisión bibliográfica, encontró que a fines del 2020 surgió un nuevo sublinaje dentro de B.1.1.1, expandiéndose velozmente en Perú y Chile, estando luego presente en varios países de América, Europa y Oceanía. Está definido por el ORF1a: Δ 3675-3677. Además, se observó una nueva combinación de mutaciones no sinónimas en spike: G75V, T76I, L452Q, F490S, T859N. A las mutaciones L452Q y F490S se les asignó el dominio de unión al receptor de la proteína spike. Mientras que L452Q es casi exclusivo de C.37, L452R ha surgido de forma independiente en los VOI B.1.427 / B.1.429 (difundido en California)⁽³⁴⁾

Variaciones de aminoácidos: Mutaciones en las moléculas que conforman las proteínas.

Proteínas: Moléculas complejas que desempeñan determinadas funciones específicas y críticas en el cuerpo.

Proteína Spike: Proteína que se ubica en la superficie del virus del SARS-CoV-2 y es la que se encarga de hacer que el virus entre en el huésped.

SARS-CoV-2: Virus que ocasiona la enfermedad respiratoria denominada enfermedad por coronavirus de 2019 o COVID-19.

Secuenciación de proteínas: Proceso en el que se da a conocer la secuencia de los aminoácidos en una proteína, realizada para detectar alguna modificación en la misma.

Aminoácidos: Moléculas esenciales que forman parte y estructuran una proteína.

Materiales y métodos

El presente trabajo de tesis pertenece a la línea de investigación en enfermedades transmisibles de la Universidad Católica Santo Toribio de Mogrovejo, al área de ciencias médicas y de salud, sub área de Biotecnología en salud y disciplina en tecnología para la identificación y funcionamiento del ADN, proteínas y enzimas y como influyen la enfermedad; cuenta con un diseño de investigación de tipo observacional, descriptivo, retrospectivo y transversal, para lo que se llevó a cabo una recopilación sistemática de bibliografía científica.

Se realizó un procedimiento secuencial en el que primero se obtuvieron a partir de la base de datos GISAID, la secuencia de aminoácidos de la proteína spike natural de Wuhan - China la cuál presenta un total 1273 aminoácidos, luego se descargaron todas las secuencias de spike depositadas pertenecientes a América del Sur hasta el 20 de septiembre del 2021, fecha en la que según el cronograma fue el último día para la recolección de datos; la población diana de este estudio fueron todos los genes de la proteína spike del genoma del SARS-CoV-2 secuenciados en países de América del Sur, la población accesible fueron todos los genes de la proteína spike de secuenciamientos completos en países de América del sur que se encontraban publicadas en la base de datos GISAID hasta el 20 de septiembre del 2021, la población elegible fue la población accesible que cumplía con los criterios de inclusión y exclusión; los criterios de inclusión fueron los genes de la proteína spike del SARS-CoV-2 de la población accesible secuenciados en países de América del Sur y que tenían un secuenciamiento completo; los criterios de eliminación fueron todos los genes de la proteína spike del SARS-COV-2 que no fueron secuenciados en América del Sur y los genes de la proteína spike del SARS COV 2 que no tenían secuenciamiento completo.

Luego se realizó un filtro de secuencias y alineamiento múltiple ya que de la totalidad de secuencias descargadas pasaron por dos filtros: *i)* Se separaron las secuencias pertenecientes a los países de América del Sur, siendo considerados los siguientes: Perú, Chile, Brasil, Uruguay, Paraguay, Bolivia, Argentina, Venezuela, Colombia, Surinam, Ecuador y Guyana Francesa; *ii)* Fueron eliminadas todas aquellas secuencias depositadas de forma incompleta y las que solo presentan segmentos secuenciados de la proteína spike, solo se aceptaron por conveniencia una variación de ± 15 aminoácidos. Para ambos filtros se usó un código escrito en Python. El alineamiento de las secuencias de los aminoácidos fue realizado en el software de acceso libre MAFFT (<https://mafft.cbrc.jp/alignment/server/>). Los alineamientos fueron descargados y visualizados en el programa Jalview (<https://www.jalview.org/>), los índices de

similaridad evaluados fueron $\cong 95 - 99\%$, los espacios de aminoácidos faltantes fueron completados con líneas (-).

Finalmente se realizó el Análisis e interpretación de datos, a partir de los datos obtenidos, se crearon tablas y gráficos que permiten observar las variaciones de los aminoácidos de la proteína spike; no fue necesaria la aplicación de instrumentos ya que se contó con información de dominio público depositada en las bases de datos de GISAID y se realizaron análisis de alineamiento bioinformático con el uso del software MAFFT.

Se realizó una operacionalización de variables la cual se puede ver en anexo 1 y una matriz de consistencia la cuál se aprecia en anexo 2. En cuanto a los aspectos éticos, el presente trabajo de investigación fue aprobado por el Comité de Ética en Investigación de la Facultad de Medicina Humana de la Universidad Católica Santo Toribio de Mogrovejo. No incluye el trabajo con seres vivos o muestras biológicas ya que la información que se utilizó fue recolectada de la base de datos GISAID, la cual es una información de dominio público, por lo cual no se necesitó solicitar permisos; tampoco atenta contra el principio de beneficencia ni el principio de justicia al no ser necesaria la aplicación de un consentimiento informado y se cumple con el criterio de respeto hacia las personas, según lo establecido en el reporte de Belmont. Este trabajo otorga beneficios directos ya que nos permitirá conocer el grado de variaciones que se han dado en las secuencias de aminoácidos de la proteína spike, se demostrará en porcentajes el grado de homología con la identidad de nuestro genoma base del virus aislado de Wuhan, las variaciones serán representados en una tabla; asimismo nos brindará beneficios de tipo indirecto al ser todos los resultados de dominio público por lo cual esta información podrá ser utilizada para investigaciones futuras relacionadas con la COVID-19.

Resultados y discusión

1. Resultados de la búsqueda de la base de datos GISAID.

Luego de realizar la búsqueda en la base de datos GISAID, se obtuvieron más de un millón de secuencias de aminoácidos de la proteína spike que fueron secuencias a nivel mundial, la gran mayoría por técnica de Next Sequence Generation (NGS). A partir del resultado, se hizo un filtro utilizando un código escrito en Python (ver anexo 3), donde, primero se obtuvieron sólo las secuencias de América y usando el mismo scripts pero con ligeras modificaciones se obtuvieron las secuencias de América del sur, que estaban completas, comparadas con la secuencia originaria de Wuhan, China.

Del filtro se logró obtener un total de 45458 secuencias de aminoácidos de la proteína spike del SARS-CoV-2 para los 12 países de América del Sur donde se incluyen (ver tabla 01): Argentina, Bolivia, Brasil, Chile, Colombia, Ecuador, Guyana Francesa, Paraguay, Perú, Surinam, Uruguay y Venezuela.

Lo encontrado acá se condice con lo hallado por Núñez y Zabala, quienes identificaron que Brasil, Argentina, México, Colombia, Chile y Perú son los países con mayor incidencia de la enfermedad en Latinoamérica coincidiendo la totalidad de casos detectados por aquellos autores con las secuencias de spike analizadas en este estudio, lo que afirma la relación estrecha que hay entre las mutaciones de la denominada proteína S con la viralidad de la enfermedad.

Tabla 01: Resultado de secuencias de aminoácidos de América del Sur

País	# Total de secuencias Spike descargadas de GISAID	# Secuencias completas analizadas	# Secuencias incompletas excluidas
Brasil	29639	29502	137
Chile	5215	5206	9
Argentina	3969	2678	1291
Perú	3666	3646	20

Ecuador	1049	1022	27
Colombia	576	553	23
Suriname	462	462	0
Guyana Francesa	354	354	0
Uruguay	246	246	0
Venezuela	153	49	104
Paraguay	88	87	1
Bolivia	41	41	0
TOTAL	45458	43846	1612

Fuente: Elaboración propia

2. Análisis por alineamiento múltiple local de secuencias de aminoácidos

Las 43846 secuencias que se obtuvieron como resultado del filtro anterior, fueron llevadas a un análisis de alineamiento múltiple local (este tipo de alineamiento identifica regiones similares dentro de largas secuencias que normalmente son muy divergentes entre sí), para este objetivo se usó la herramienta bioinformática escrita en código abierto, MAFFT, tanto en su versión online, como en su versión de escritorio. y la herramienta JALVIEW, versión de escritorio.

El alineamiento múltiple, se realizó para el total de cada uno de los países hallados de forma independiente, el software fue ejecutado siguiendo los parámetros dados por defecto por MAFFT y JALVIEW donde obtuvieron los siguientes resultados. (Tabla 1)

A partir de los alineamientos con MAFFT, los resultados permitieron realizar un filtro donde se observan las variaciones en las diferentes posiciones dentro de los 1274 aminoácidos de la secuencia spike (anexo 4). Con estos datos fue posible la obtención de las posiciones que contienen a los aminoácidos mutados, para este corte se tuvo en consideración clasificación de la frecuencia de rareza de mutación, donde se selecciono como mejor valor de corte el 10% del total de los aminoácidos en spike.

3. Análisis de las mutaciones en la proteína spike

Para este análisis se utilizó la fórmula (**Anexo 5**). El resultado del análisis de las mutaciones de las 45458 secuencias depositadas de la proteína spike de los países de Sudamérica se muestran en las (Tablas 02 y 03). Para una mejor interpretación de los resultados, los análisis de aquí en adelante están basados en países con por lo menos 400 secuencias depositadas en GISAID (Argentina, Brasil, Chile, Colombia, Ecuador, Perú, Uruguay y Suriname).

Según lo investigado por Guruprasad (2021) y Mansbach (2023), la proteína spike está dividida en 21 regiones. La tabla 02. muestra la distribución de mutaciones en diferentes regiones de la proteína spike. Las mutaciones son frecuentes en la gran mayoría de las regiones de la proteína, a excepción de siete regiones (*S1A-S1B linker, S1B - S1C linker, S1C - S1D linker, S2' cleavage site, Fusion peptide, Transmembrane region y Cytoplasmic region*) que no presentaron ningún tipo de mutación. El dominio S1^A (residuos del 1 al 302) que contenía la mutación **XXX** se observaron 93 mutaciones. Nuestro análisis difiere de los resultados ya reportados por, donde se reportaba que el residuo D614G era el residuo con más frecuencia de mutaciones, sin embargo, nuestros datos muestran que sólo una secuencia procedente de Uruguay presenta esta mutación, el cual está relacionada con el alto grado de inafectabilidad de SARS-CoV-2. ⁽³⁵⁻³⁷⁾

Tabla 02: Distribución de mutaciones en las diferentes regiones de la proteína spike del SARS-CoV-2 humano.

Regiones de la proteína Spike	Número total de mutaciones	Número de tipos distintos de mutaciones
S1 ^A dominio (1-302)	93	38
S1 ^A -S1 ^B linker (303-332)	0	0
S1 ^B dominio (333-527)	59	55
S1 ^B - S1 ^C linker (528-533)	0	0
S1 ^C dominio (534-589)	3	3
S1 ^C - S1 ^D linker (590-593)	0	0
S1 ^D dominio (594-674)	16	13
Protease cleavage site (675-692)	13	11

S1-S2 subunits linker (693-710)	6	6
Central β -strand (711-737)	2	2
Downward helix (738-782)	1	1
S2' cleavage site (783-815)	0	0
Fusion peptide (816-828)	0	0
Connecting region (829-911)	5	5
Heptad repeat region (912-983)	6	6
Central helix (984-1034)	7	6
β -hairpin (1035-1068)	1	2
β -sheet dominio (1069-1133)	3	2
Heptad repeat region (1134-1213)	10	8
Transmembrane region (1214-1236)	0	0
Cytoplasmic region (1237-1273)	0	0

Fuente: Elaboración propia

La Figura 1 muestra la densidad de mutación estimada en función del número de mutaciones observadas a lo largo de la secuencia correspondiente a las diferentes regiones de la proteína espiga. El sitio de escisión de la proteasa (entre los residuos 675 y 692) se asoció con la densidad de mutación más alta. Además, el dominio S1^D (residuos del 594 al 674) y el dominio S1^B (residuos del 333 al 527) son las regiones que acumulan relativamente más mutaciones que el resto de otras regiones de la proteína spike.

Según Jaimes, et al. (2020) y Wang, et al. (2020), el sitio de escisión de la proteasa parece ser una característica distintiva entre las secuencias relacionadas con el SARS e introduce un sitio de escisión potencial para la proteasa, las mutaciones en este sitio en la proteína Spike le es ventajosa porque le permite al virus, experimentar una escisión proteolítica por un gran

número de enzimas del hospedero en donde se encuentra, Por ejemplo, una variedad de proteasas humanas, como como tripsina, triptasa Clara, proteasa similar a la tripsina de las vías respiratorias humanas (HAT) y proteasa transmembrana serina 2 (TMPRSS2), se sabe que escinden y activan la proteína S del SARS-CoV (Jaimes, et al., 2020; Wang, et al., 2020). Por lo que, la mutación en esta región es un factor clave para determinar las características epidemiológicas y patológicas del virus, incluidos el rango de huéspedes, el tropismo tisular, la transmisibilidad y la mortalidad. ⁽³⁸⁻³⁹⁾

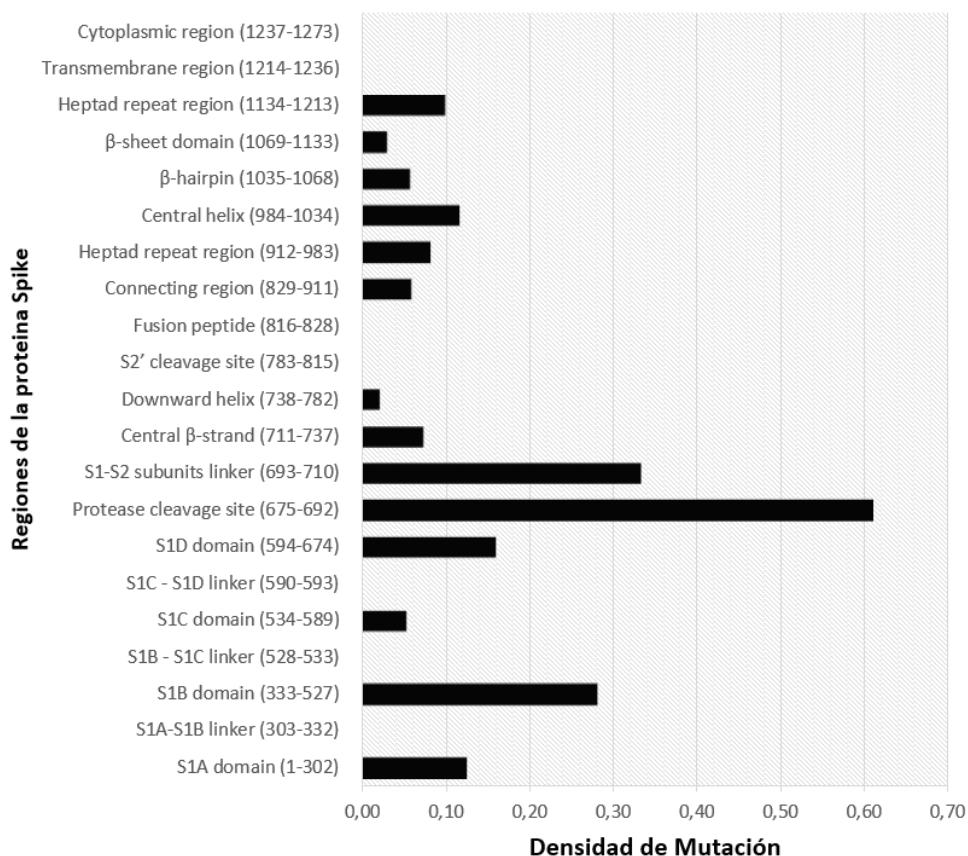


Figura 1. Densidad de mutaciones en las diferentes regiones de la proteína spike del SARS-CoV-2 humano.

Fuente: Elaboración propia

Sitios de mutación y tipos de mutaciones observados en la proteína spike de SARS-CoV-2 humano según ubicaciones geográficas

Se pueden encontrar múltiples tipos de mutaciones en la misma posición en la secuencia de la proteína spike. Por ejemplo, se puede ver que en la posición 18 el aminoácido L muta a S, F, R o K. La distribución geográfica basada en la ubicación de los sitios de mutación y los tipos de mutación se muestra en el Anexo 06. Nuestros resultados muestran que, de los 8 países analizados, hubo un total de 102 mutaciones de los 1274 aminoácidos de la secuencia spike,

siendo Leu18 (en 6 países), Thr20 (en 7 países), Pro26 (en 5 países) y Asp138 (en 7 países) los aminoácidos con más veces mutadas por país, Glu488, Asn504, Asn505, Asp617, His658, Pro684 y Glu1030 son los aminoácidos mutados en por lo menos en 2 países. Un dato sorprendente es que el país de Ecuador presentó un total de 32 aminoácidos mutados, lo que lo posiciona como el país con mayor número de mutaciones, superando a otros países con mayor densidad poblacional, como lo es Brasil. En este sentido, Vargas (2021), dio cuenta de que la variante Lambda T76I y L452Q ubicadas en la proteína spike, podrían ser responsables de mayores tasas de infectividad; mientras que F490S, RSYLTPGD246-253N, y particularmente L452Q, que tenía bastante similitud a la mutación de la variante Delta L452R, le darían a Lambda la propiedad de evadir el sistema inmunológico, por lo que esta variable fue culpable de una gran virulencia en el Perú. A su vez Gomez (2022), encontró que en Colombia algunas de las cepas que más virulencia causaron fueron las B.1.1.529 B.1.1.7, B.1.3.5.1, B.1.617.2, B.1.621, y P.1.

Según lo encontrado por Hady (2022), el número total de sitios de mutación observados es: Argentina (21), Brasil y Perú (18), Colombia y Uruguay (15), Suriname (14) y Chile (13); el número total de tipos de mutación observados fue: Brasil (51), Ecuador (47), Argentina y Perú (28), Chile (24), Colombia y Suriname (16), Uruguay (15). A partir de nuestro estudio, está claro que la proteína espiga del SARS-CoV-2 humano está mutada en muchos sitios y que puede haber múltiples tipos de mutaciones asociadas con el sitio de mutación. Otro dato interesante es que solo Uruguay presenta dos mutaciones el cual puede dar origen a otras variedades de SARS-CoV-2, entre ellas la variedad Alfa, Beta, Gamma y Delta, estas mutaciones son Asn501 y Asp614.⁽⁴⁰⁾

Por su parte, Núñez y Zabala (2023) encontraron que las variantes Alfa, Beta y Gamma son las que presentan evasión inmune y eficaz transmisión; la variante Mu, linaje B.1.621 es identificable a partir de la proteína S y tiene importancia fenotípica por su evasión inmune e infectividad (E484K, N501Y y P681H); la variable Eta linaje B.1.52 presenta las siguientes mutaciones en la proteína S: D614G, E484K y Q677H, lo que en conjunto explica la difusión que tuvo el virus en Latinoamérica.

Mapeo de mutaciones en el dominio de unión al receptor de la proteína spike de SARS-CoV-2

Según lo investigado por Robson (2020), todas nuestras secuencias conservan el motivo **RSFIEDLLFNKV** característico de SARS-CoV-2 ⁽⁴¹⁾, el cual no presenta ninguna mutación, sin embargo, según lo encontrado por Banerjee et al. (2020) como es de esperarse, el mayor número de mutaciones se encuentra en el subdominio S1 que alberga el dominio de unión al receptor ACE 2 o RBD de la proteína spike, el cual comprende desde el aminoácido 13 hasta el 685, (13aa – 685aa), nuestras secuencias analizadas presentan un mayor número de mutaciones, esta región estaría relacionado al grado de transmisibilidad, mostrando la gran versatilidad que posee para poder unirse al receptor. El subdominio S2, que contiene péptido de fusión, el cual cumple la función de una entrada exitosa de la proteína spike, se observan mutaciones características para cada país, el cual podría estar relacionado a la especificidad a cada país. ⁽⁴²⁾

La proteína spike juega un papel importante en la unión a receptores específicos en la superficie de la célula huésped y luego cataliza la fusión del virus con la membrana de la célula huésped, lo cual es esencial para que ocurra el proceso de infección. El RBD en la proteína spike interactúa con el receptor ACE2 del huésped, lo que provoca infección por coronavirus que conduce a la enfermedad del COVID-19. La estructura tridimensional del SARS-CoV-2 RBD humano (entre los residuos 333 y 527) junto con el complejo receptor ACE 2 (código PDB: 6LZG) fue utilizado para mapear los sitios de mutación, como se muestra en la Figura 2. El RBD de la proteína Spike de SARS-COV -2 presenta 28 sitios diferentes de mutación. Estas mutaciones se hallan en las posiciones; 351, 417, 418, 420, 421, 422, **429**, 453, 455, 456, 457, **464**, 484, 485, 487, **488**, 489, 490, 491, 493, 495, **496**, 501, 504, 505, 506 y 513. Los números en negrita indican las posiciones con cuatro mutaciones y los números subrayados indican las posiciones con tres apariciones de mutaciones observadas en un sitio en particular. Por su parte, Romero (2021) encontró que había una nueva combinación de mutaciones no sinónimas en spike: G75V, T76I, L452Q, F490S, T859N. A las mutaciones L452Q y F490S se les asignó el RBD de la proteína spike; además encontró que la delección S:Δ246-252 debería ser tomada en consideración para la validación de tests y vacunas. Toda esta data confirma lo encontrado por Robson (2023) en el sentido de que las propiedades estructurales del dominio de unión del receptor del SARS-CoV-2 han consolidado su afinidad de unión al receptor ACE2 del huésped.

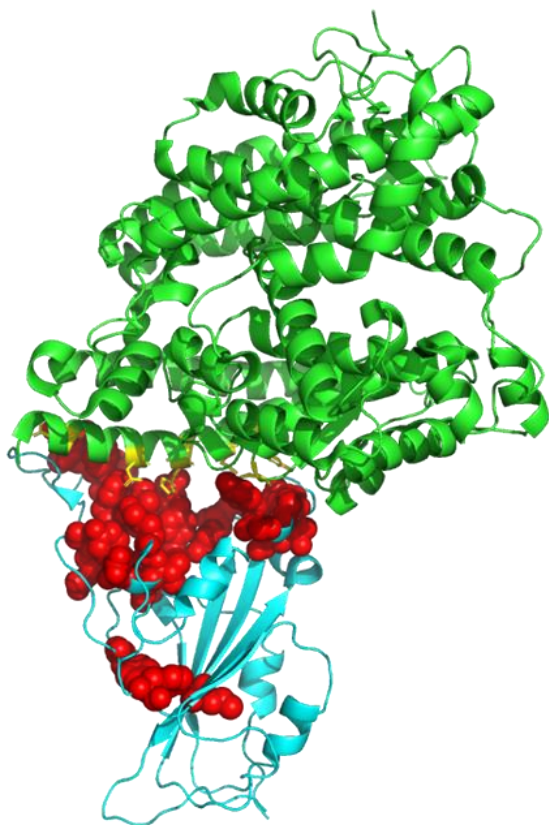


Figura 2. Mutaciones en la región RBD de la proteína spike. Las 44 mutaciones (esferas rojas) mapeadas en la estructura tridimensional de la proteína spike RBD (cian) complejada con el receptor ACE-2 (verde) (código PDB: 6LZG).

Fuente: Elaboración propia

La distribución del número de mutaciones RBD observadas en los 28 sitios diferentes se muestra en la Figura 3. Los sitios de mutación y los tipos de mutación de la proteína espiga humana del SARS-CoV-2 en el RBD según la distribución geográfica por ubicación se presentan en la Tabla Y. Por lo tanto, se observó que el número total de sitios de mutación únicos en el RBD es; Argentina y Ecuador (6), Brasil y Perú (5), Chile (4), Colombia y Uruguay (3), Suriname (2), así como el número total de diferentes tipos de mutaciones observadas; Brasil (16), Chile, Ecuador y Perú (9), Argentina (7), Colombia y Uruguay (3), Suriname (2).

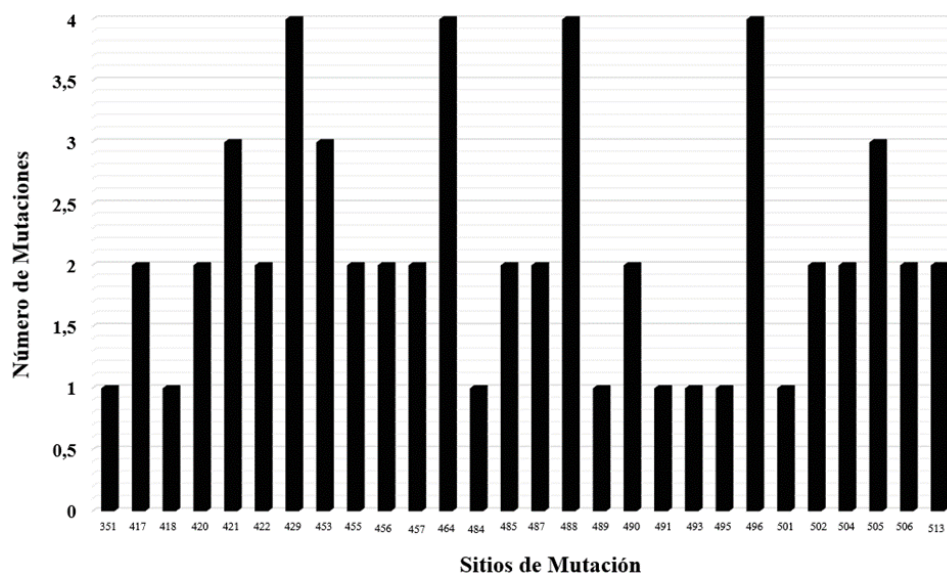


Figura 3. Número de mutaciones observadas en el sitio RBD de la proteína spike del SARS-CoV-2.

Fuente: Elaboración propia

En la Figura 4 se muestran varios residuos mutados de la región RBD de la proteína spike implicados en la interacción con el receptor ACE-2. Los residuos de la región RBD de la proteína spike de SARS-CoV-2 complejados con el receptor ACE-2 a una distancia interatómica de $\leq 3,2$ Å (código PDB: 6LZG) son; K417, Y449, Y453, A475, N487, T500, N501 y G502.

Para Guruprasad (2020), los residuos de la proteína spike con mayor proximidad al receptor ACE 2 son Y453 que está cerca de His34 del receptor ACE-2, y los residuos T500 y N501 en la proteína spike que están cerca de Tyr41 en ACE-2. Los residuos de la proteína spike cercanos a la base del receptor ACE-2, que están involucrados en las mutaciones son; K417 (próximo a I418), Y453, C488 (próximo a N487) y G502 (próximo a N501).⁽⁴³⁾

Por ende, siempre según Guruprasad, et al. (2021), los residuos en las posiciones; 417, 453, 487, 500, 501 y 502 que están asociados con las mutaciones RBD y que están cerca del receptor ACE-2 afectarían la forma y la carga de la proteína cerca de la interfaz de interacción proteína-receptor. Por lo tanto, estas regiones podrían ser consideradas sitios de interés, para el desarrollo de anticuerpos, vacunas y fármacos.

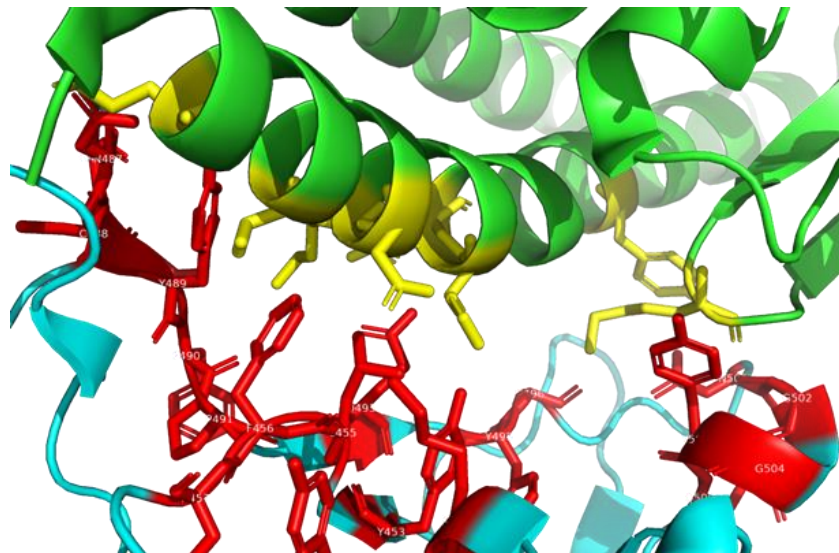


Figura 4. Interacciones de la proteína spike (cian) con ACE-2 (verde) y los residuos de la cadena lateral (amarillo) que están dentro de $3,2 \text{ \AA}$ en la estructura cristalina de la región RBD de la proteína spike RBD humana del SARS-CoV-2 complejada con el receptor ACE-2 (código PDB: 6LZG). Los residuos mutados de la proteína spike se muestran en (rojo).

Fuente: Elaboración propia

Conclusiones

Las variaciones de los aminoácidos en las proteínas forman parte de la evolución de las especies. En el análisis de los aminoácidos de la proteína spike del SARS-CoV-2 de los países de América del Sur, se logró describir la alta diversidad genómica, esto se debe a cambios en la secuencia de aminoácidos del ARN viral (mutaciones), las que favorecen de manera evolutiva debido a modificaciones estructurales en sus proteínas funcionales ocurre en la región RBD de la proteína spike.

La acumulación de mutaciones genera cambios en la epidemiología de la enfermedad, da lugar a variantes o linajes, y estos a su vez en clados, como se puede evidenciar en la figura 1, su alta variación en la región de corte.

Las variaciones en los aminoácidos de SARS-CoV-2 son de interés e importancia ya que presentan probable afectación de la transmisión, interferencia diagnóstica, susceptibilidad reducida a la terapéutica, gravedad de la enfermedad y respuesta inmunológica.

El conocimiento de las variaciones aminoacídicas de la proteína Spike, permite tener un mejor conocimiento del modo de acción de la epidemiología genómica del SARS-CoV-2 es imprescindible para explorar la transmisión, la evolución y también la patogenicidad de los virus.

Recomendaciones

Se recomienda realizar más estudios en los cuales se aumente la población, no solo en países de América del Sur, y realizar un análisis de las mutaciones tanto de la proteína spike como de las otras tres proteínas estructurales del SARS-CoV-2 para así tener una noción más clara de la diversidad genómica en SARS-CoV-2, entender mejor la patogenia de la infección por este virus y su mecanismo de transmisión, ya que esto ayudará a tomar medidas de contención adecuadas frente al virus.

Se recomienda a su vez una vigilancia genómica continua y suficiente para poder guiar el desarrollo y el uso de vacunas, terapias, diagnóstico y políticas en salud.

Referencias

- Tsang KW, Ho PL, Ooi GC, Yee WK, Wang T, Chan-Yeung M, et al. Un grupo de casos de síndrome respiratorio agudo severo en Hong Kong [Internet]. *N Engl J Med*; 2003 mayo [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/12671062/>
2. Lee N, Hui D, Wu A, Chan P, Cameron P, Joynt GM, et al. Un brote importante de síndrome respiratorio agudo severo en Hong Kong [Internet]. *N Engl J Med*; 2003 mayo [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/12682352/>
 3. Poutanen SM, Low DE, Henry B, Finkelstein S, Rose D, Green K, et al. Identificación del síndrome respiratorio agudo severo en Canadá [Internet]. *N Engl J Med*. 2003 Mayo [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/12671061/>
 4. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Características clínicas de los pacientes infectados con el nuevo coronavirus de 2019 en Wuhan, China [Internet]. *Lancet*; 2020 febrero [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31986264/>
 5. Mahase E. Covid-19: OMS declara pandemia por «niveles alarmantes» de propagación, gravedad e inacción [Internet]. *BMJ*. 2020 marzo [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32165426/>
 6. Zhu N, et al. Un nuevo coronavirus de pacientes con neumonía en China, 2019. [Internet] *N. Engl. J. Med.* 2020 [citado 12 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/31978945>
 7. WuZ, McGoogan JM. Características y lecciones importantes del brote de la enfermedad por coronavirus 2019 (COVID-19) en China: resumen de un informe de 72314 casos del Centro Chino para el Control y la Prevención de Enfermedades. [Internet]. *JAMA*. 2020 [citado 12 de Junio de 2020]. Disponible en <https://pubmed.ncbi.nlm.nih.gov/32091533>
- Vankadari N. Mutaciones abrumadoras o SNP del SARS-CoV-2: un punto de precaución [Internet]. *Gene*. 2020 Mayo [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32445924/>

8. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus A, Fouchier RAM. Aislamiento de un nuevo coronavirus de un hombre con neumonía en Arabia Saudita [Internet]. *N Engl J Med*. 2012 [citado 12 de Junio de 2020] Disponible en: <https://pubmed.ncbi.nlm.nih.gov/23075143/>
9. De Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, et al. Coronavirus del síndrome respiratorio de Oriente Medio (MERS-CoV): anuncio del Grupo de Estudio de Coronavirus [Internet]. *J Virol*. 2013 [citado 12 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3700179/>
10. Woo PC, Lau SK, Chu CM y col. Caracterización y secuencia completa del genoma de un nuevo coronavirus, el coronavirus HKU1, de pacientes con neumonía [Internet]. *J Virol*. 2005 [citado 12 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC538593/>
11. Perlman S, Netland J. Coronavirus post-SARS: actualización sobre replicación y patogénesis [Internet]. *Nat Rev Microbiol*. 2009 [citado 12 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2830095/>
12. Skehel JJ, Wiley DC. Unión de receptores y fusión de membranas en la entrada del virus: la hemaglutinina de la influenza [Internet]. *Annu Rev Biochem*. 2000 [citado 13 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/10966468/> Phan T. Diversidad genética y evolución del SARS-CoV-2 [Internet]. *Infect Genet Evol*. 2020 julio [citado 10 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32092483/>
13. Li F. Estructura, función y evolución de las proteínas de pico de coronavirus [Internet]. *Rev. Virol*. 2016 [citado 13 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5457962/>
14. Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, et al. Estructura previa a la fusión de una proteína de pico de coronavirus humano [Internet]. *Naturaleza*. 2016 [citado 13 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4860016/>
15. Walls AC, Tortorici MA, Bosch BJ, Frenz B, Rottier PJ, et al. Estructura de microscopía crioelectrónica de un trímero de glicoproteína de pico de coronavirus

- [Internet]. *Naturaleza*. 2016 [citado 13 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5018210/>
16. Beniac DR, Andonov A, Grudeski E, Stand TF. Arquitectura del pico de prefusión del coronavirus del SARS [Internet]. *Nat Struct Mol Biol*. 2006 [citado 13 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7097490/>
 17. Li F, Berardi M, Li WH, Farzan M, Dormitzer PR, Harrison SC. Estados conformacionales del ectodominio proteico de pico de coronavirus del síndrome respiratorio agudo severo [Internet]. *J Virol*. 2006 [citado 15 de Junio de 2020]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1489032/>
 18. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. La glicoproteína Spike del nuevo coronavirus 2019-nCoV contiene un sitio de escisión similar a furina ausente en el CoV del mismo lado [Internet]. *Antiviral Res*. 2020 [citado 15 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32057769/>
 19. Zhou P, Yang X - L, Wang X - G, et al. Un brote de neumonía asociado con un nuevo coronavirus de probable origen en murciélagos [Internet]. *Nature*. 2020 febrero [citado 15 de Junio de 2020]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32015507/>
 20. Zhao WM, et al. El nuevo recurso de coronavirus de 2019 [Internet]. *Yi Chuan*. 2020 [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32102777/>
 21. Korber B, et al. Seguimiento de los cambios en el pico de SARS-CoV-2: evidencia de que D614G aumenta la infectividad del virus COVID-19 [Internet]. *Celúla*. 2020 [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32697968/>
 22. Tang X, et al. Sobre el origen y continua evolución del SARS-CoV-2. ciencia nacional [Internet]. *Rev*. 2020 [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/34676127/>
 23. Forster P, Forster L, Renfrew C, Forster M. Análisis de redes filogenéticas de los genomas del SARS-CoV-2 [Internet]. *Academia Nacional de ciencia de EE.UU*. 2020

- [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32269081/>
24. Shang J, et al. Base estructural del reconocimiento de receptores por SARS-CoV-2 [Internet]. *Naturaleza*. 2020 [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32225175>
 25. Paredes AC, et al. Estructura, función y antigenicidad de la glicoproteína espiga del SARS-CoV-2 [Internet]. *Celúla*. 2020 [citado 10 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32155444>
 26. Robson, B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance [Internet]. *Computers in biology and medicine*. 2020 [citado 25 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32568687/>
 27. Pokhrel S., Kraemer B., Burkholz S., Mochly D. Natural variants in SARS-CoV-2 Spike protein pinpoint structural and functional hotspots with implications for prophylaxis and therapeutic strategies. [Internet]. *Scientific Reports*. 2021 [citado 1 de Marzo de 2023]. Disponible en: <https://www.nature.com/articles/s41598-021-92641-x>
 28. Gómez D. Principales variaciones genéticas del SARS-CoV2 en Latinoamérica ¿Cuáles son las variaciones genéticas del SARSCoV2 que se han manifestado en Colombia? [Internet]. *Ciencia Unisalle*. 2021 [citado 20 de Febrero de 2023]. Disponible en: <https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=2915&context=optometria>
 29. Núñez J., Zabala A. Covid-19: principales variantes genéticas. *Ciencia Latina*. [Internet]. 2022 [citado 2 de Marzo de 2023]. Disponible en: <https://ciencialatina.org/index.php/cienciala/article/view/4750/7197>
 30. Aguilar F., Suclupe D., Vega J., Silva H. Diversidad genómica en SARS-CoV-2: mutaciones y variantes. [Internet]. *Rev. Cuerpo Med. HNAAA*. 2021 [citado 1 de Marzo de 2023]. Disponible en: http://www.scielo.org.pe/scielo.php?pid=S2227-47312021000500020&script=sci_arttext

31. Expósito A., Feria González S., Miguel P. Variantes genéticas del SARS-CoV-2 y sus implicaciones clínicas. [Internet]. *Medisan*. (2021) [citado 1 de Marzo de 2023]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192021000601424
32. Loayza M., De la Cruz J. Effect of SARS-CoV-2 variants on the transmission of COVID-19 in Peru. [Internet]. *Revista de la Facultad de Medicina Humana*. (2021) [citado 1 de Marzo de 2023]. Disponible en: <https://inicib.urp.edu.pe/rfmh/vol21/iss1/2/>
33. Vargas N., Araujo R., Mestanza O., Galarza M., Rojas N., Solari L. SARS-CoV-2 Lambda and Gamma variants competition in Peru, a country with high seroprevalence. [Internet]. *The Lancet Regional Health*. 2021 [citado 1 de Marzo de 2023]. Disponible en: [https://www.thelancet.com/journals/lanam/article/PIIS2667-193X\(21\)00108-3/fulltext](https://www.thelancet.com/journals/lanam/article/PIIS2667-193X(21)00108-3/fulltext)
34. Romero P., Dávila A., González L., Salvatierra G., Cuicapuza D., Solis L, Marcos P., Huancachoque J., Carhuaricra D., Rosadio R., Luna L., Maturrano L. Tsukayama P. Novel sublineage within B.1.1.1 currently expanding in Peru and Chile, with a convergent deletion in the ORF1a gene ($\Delta 3675-3677$) and a novel deletion in the spike gene ($\Delta 246-252$, G75V, T76I, L452Q, F490S, T859N). [Internet]. *Virological*. 2021 [citado 1 de Marzo de 2023]. Disponible en: <https://virological.org/t/novel-sublineage-within-b-1-1-1-currently-expanding-in-peru-and-chile-with-a-convergent-deletion-in-the-orf1a-gene-3675-3677-and-a-novel-deletion-in-the-spike-gene-246-252-g75v-t76i-l452q-f490s-t859n/685>
35. Guruprasad L. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 spike proteins for host receptor recognition [Internet]. *Proteins: Structure, Function, and Bioinformatics*. 2020 [citado 20 de Febrero de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7323375/>
36. Guruprasad L , Human . SARS CoV-2 spike protein mutations. [Internet]. Wiley. 2021 [citado 20 de Febrero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/33423311/>

37. Mansbach R , Chakraborty S , Nguyen K , Montefiori D, Korber B , Gnanakaran S. The SARS-CoV-2 spike variant D614G favors an open conformational state [Internet]. *Science advances*. 2021 [citado 20 de Febrero de 2023]. Disponible en: <https://www.science.org/doi/full/10.1126/sciadv.abf3671?%ef%bf%bdParamMap}%ef%bf%bd=>
38. Jaimes J , Millet J , Whittaker G. Proteolytic cleavage of the SARS-CoV-2 spike protein and the role of the novel S1/S2 site [Internet]. *IScience*. 2020 [citado 20 de Febrero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32512386/>
39. Wang Q , Qiu Y , Li J , Zhou Z , Liao C , Ge X . A unique protease cleavage site predicted in the spike protein of the novel pneumonia coronavirus (2019-nCoV) potentially related to viral transmissibility. [Internet]. *Virologica Sinica*. 2020 [citado 20 de Febrero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32198713/>
40. Hady Hassine I. Covid-19 vaccines and variants of concern: A review. [Internet]. *Reviews in medical virology*. 2022 [citado 25 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/34755408/>
41. Robson, B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance [Internet]. *Computers in biology and medicine*. 2020 [citado 25 de Enero de 2023]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/32568687/>
42. Banerjee A, Begum F , Rayo U . Mutation hot spots in spike protein of COVID-19 [Internet]. *Preprints*. 2020 [citado 25 de Enero de 2023]. Disponible en: <https://www.preprints.org/manuscript/202004.0281/v1>
43. Guruprasad L. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 Spike proteins for host receptor recognition [Internet]. *Proteins: Structure, Function, and Bioinformatics*. 2020 [citado 20 de Febrero de 2023]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7323375/>

Anexos

<https://docs.google.com/document/d/1GspuWQ9O0huAWVSuPCdFahgIecKovplyFvAmfGG1DBI/edit?usp=sharing>