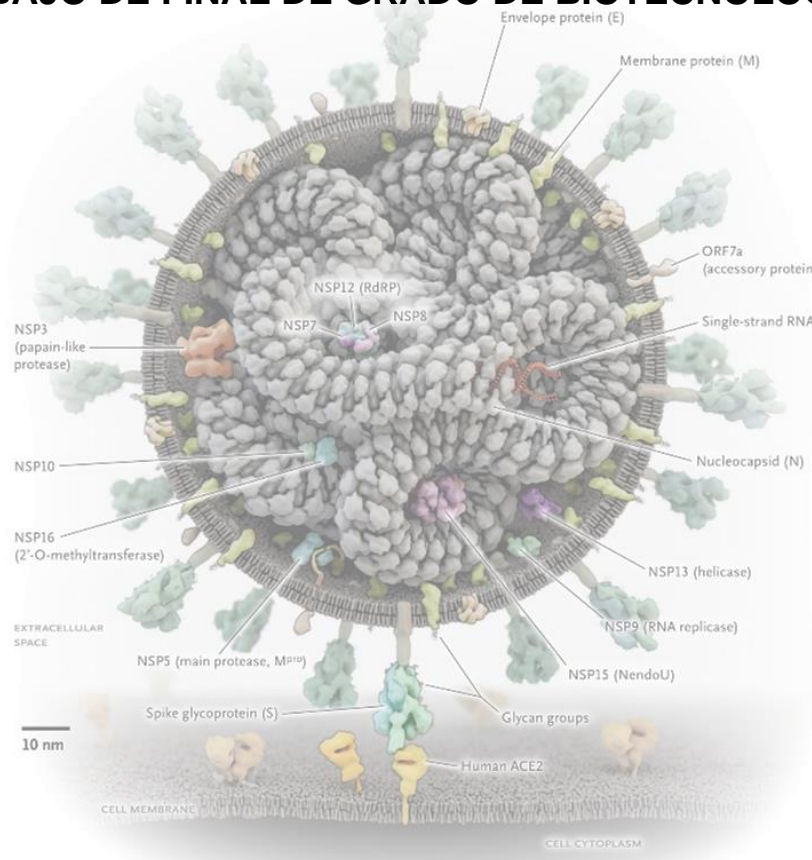


## CARACTERÍSTICAS GENÉTICAS Y VARIABILIDAD DEL SARS-CoV-2, AGENTE CAUSAL DE LA COVID-19

CAROLINA CAMPOS MARTÍNEZ

### TRABAJO DE FINAL DE GRADO DE BIOTECNOLOGÍA



**Tutor académico:** Dra. Katihuska Paredes

Grado de Biotecnología

Departamento de Bioquímica y Biotecnología  
URV

**Email:** [kathiuska.paredes@urv.cat](mailto:kathiuska.paredes@urv.cat)

**En cooperación con:** Fundación Vall d'Hebrón  
Institut de Recerca (VHIR)

**Supervisor:** Dr. Josep Quer Sivilia

Enfermedades Hepáticas, Hepatitis Virales

**Email:** [josep.quer@vhir.org](mailto:josep.quer@vhir.org)

10 de julio de 2020

*Dedicado a todas las víctimas de la COVID-19 y a aquellos que han perdido a alguien durante la pandemia, pero especialmente dedicado a los científicos/as que se están volcando en esta lucha.*

Yo, Carolina Campos Martínez , con DNI 48278460-b, soy conocedor de la guía de prevención de plagio de la URV *Prevención, detección y tratamiento del plagio en la docencia: guía para estudiantes* (aprobada en julio de 2017) (<http://www.urv.cat/ca/vidacampus/serveis/crai/que-us-oferim/formacio-competencies-nuclears/plagi/>) y afirmo que este TFG no constituye ninguna de las conductas consideradas como plagio por la URV.

Tarragona, 10 de julio de 2020

A handwritten signature in black ink, appearing to read 'Carolina', enclosed within a circular scribble.

# ÍNDICE

RESUMEN .....	5
INTRODUCCIÓN .....	6
1. Filogenia de los coronavirus, familia y relacionados, características e infecciones que producen .....	7
1.1. Análisis filogenético y clasificación taxonómica del SARS-CoV-2.....	8
1.2. SARS-CoV, características, huéspedes y tratamientos .....	10
1.3. MERS- CoV, características, huéspedes y tratamientos.....	12
2. Epidemiología del SARS-CoV-2, agente causal de la COVID-19.....	13
3. Clínica de la COVID-19 .....	14
4. Ciclo viral .....	14
5. Caracterización genómica del SARS-CoV-2 .....	16
5.1. Estructura del genoma. Genes y proteínas .....	16
5.2. Diferencias genómicas del SARS-CoV-2 en comparación con otros virus de la familia <i>Coronaviridae</i> .....	18
5.3. Proteína <i>spike</i> .....	20
5.4. Reconocimiento del receptor ECA2 .....	23
OBJETIVOS.....	26
1. Objetivo principal .....	27
2. Objetivo secundario .....	27
METODOLOGÍA.....	28
1. Secuencias de estudio .....	28
2. Alineamiento de las secuencias .....	28
2.1. Análisis de las mutaciones.....	28
RESULTADOS .....	32
DISCUSIÓN.....	34
CONCLUSIONES .....	36
AUTOEVALUACIÓN.....	37

AGRADECIMIENTOS.....	38
BIBLIOGRAFÍA.....	39
ANEXOS .....	44
1. Tabla Excel de mutaciones nucleotídicas.....	44
2. Tabla Excel de mutaciones peptídicas.....	45
3. Tabla Excel de mutaciones por muestra .....	46

## RESUMEN

El anterior brote del SARS-CoV presenta características muy similares a las del nuevo brote provocado por el SARS-CoV-2. Ambos se iniciaron en dos mercados de comida probablemente sin control sanitario a partir de la transferencia zoonótica entre especies. Epidemiológicamente han empezado de forma similar, aunque han evolucionado de forma distinta: en el segundo caso, la dispersión del virus ha sido indiscutiblemente de mayor extensión, probablemente por los cambios en las características intrínsecas del virus. En cuanto a su estructura genómica, su organización es muy parecida y muchas de las regiones que determinan la capacidad patogénica están conservadas, pero con algunas diferencias en la proteína *spike*, que podría ser la razón por la cual el virus ha optimizado su capacidad de infección. En cuanto al tratamiento, ambos son responsables de generar el síndrome respiratorio agudo grave (SARS) por lo que los tratamientos paliativos usados para el SARS-CoV podrían contemplarse para el SARS-CoV-2. Se presenta un pequeño estudio de las mutaciones nucleotídicas presentes en 28 secuencias de SARS-CoV-2 de pacientes del hospital Vall d'Hebrón de Barcelona, los cambios aminoacídicos que producen y el tipo de mutación generada en la secuencia peptídica. Los resultados muestran que la región más variable del genoma afecta a la poliproteína 1a y a la proteína de la nucleocápside. Las mutaciones halladas corresponden a los resultados de estudios previos, pero varían en su frecuencia.

**Palabras clave:** SARS-CoV-2, transferencia zoonótica, proteína *spike*, mutaciones, secuencia peptídica.

# INTRODUCCIÓN

Des del pasado diciembre de 2019, una nueva cepa vírica de coronavirus ha azotado el planeta provocando una pandemia a nivel mundial. Lo primeros casos se detectaron en la ciudad china de Wuhan y fueron relacionados con el mercado de Huanan. La nueva cepa fue inicialmente denominada como *2019 novel Coronavirus (2019-nCov)*, pero tras su análisis genético ha sido relacionado filogenéticamente con los virus SARS, responsables del síndrome agudo respiratorio grave. Dadas sus características genéticas y su estructura genómica y molecular, la nueva cepa de coronavirus ha sido denominada como SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus 2*).

La capacidad patogénica y de transmisión del virus ha permitido su propagación a escala mundial. Las prioridades actuales de la comunidad científica son el desarrollo de una vacuna efectiva y un tratamiento para los pacientes, así como detener su transmisión y minimizar los contagios para descongestionar la red sanitaria y disminuir la tasa de mortalidad. Para ello, concentra sus esfuerzos en dos vertientes para el estudio del virus: la epidemiología y la caracterización genómica. En primer lugar, el estudio de la epidemiología permite evaluar la dinámica de la infección, cómo varía su comportamiento en función del tiempo, los diferentes factores de riesgo y la población. En segundo lugar, el estudio genético del virus permitirá la identificación del origen, la variabilidad de aislados e identificar zonas del virus esenciales, por ejemplo, para el diseño de un fármaco o una vacuna. La comunidad científica se ha centrado en dos regiones para conseguir ambos objetivos: la región que codifica para la proteína *spike* y las secuencias conservadas del genoma para encontrar una diana terapéutica.

El descubrimiento del origen del virus y el huésped en el que se adaptó respondería muchas de las incógnitas<sup>1</sup>. El estudio de la variabilidad del virus abriría una puerta al conocimiento sobre su comportamiento actual y futuro, considerando la posibilidad de que tras la pandemia que ha generado su expansión aparecerán nuevos brotes e incluso se podría contemplar la posibilidad de que se convierta en un virus estacional como la gripe.

# 1. Filogenia de los coronavirus, familia y relacionados, características e infecciones que producen

La familia *Coronaviridae* pertenece al orden de los *Nidovirales* y engloba a los coronavirus en la subfamilia *Coronavirinae*. Éstos se clasifican en cuatro géneros: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus*. Los siete coronavirus capaces de infectar a la especie humana pertenecen a los géneros *Alphacoronavirus* y *Betacoronavirus*: son el HCoV-NL63, HCoV-229E, HCoV-OC43, HKU1, SARS-CoV, MERS-CoV y SARS-CoV-2. En el género *Betacoronavirus*, los virus se clasifican en tres linajes o subgéneros distintos: linaje A (subgénero *Embecovirus*), linaje B (subgénero *Sarbecovirus*), linaje C (subgénero *Marbecovirus*) y linaje D (subgénero *Hibecovirus*) (Ilustración 1)<sup>2,3</sup>.

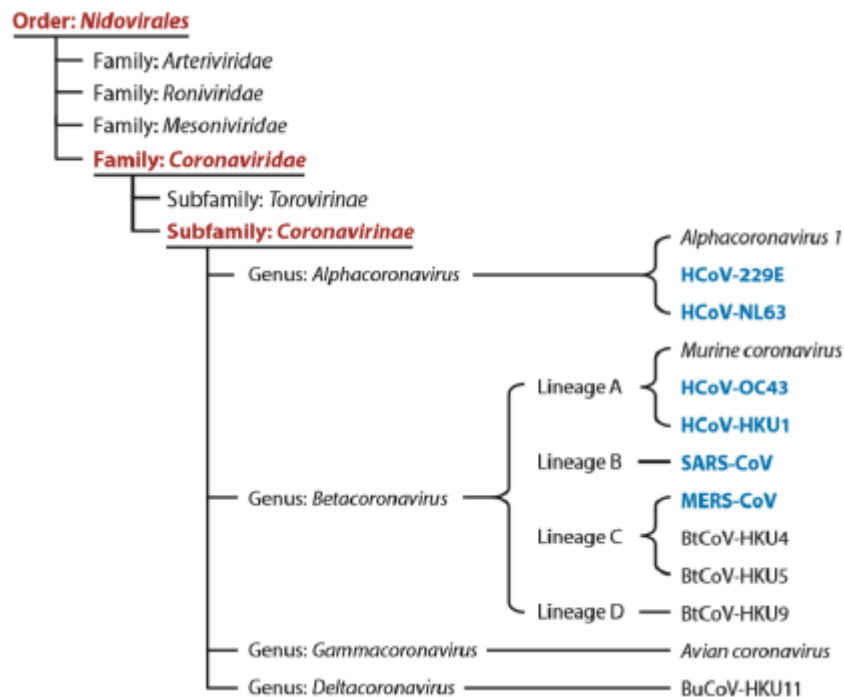
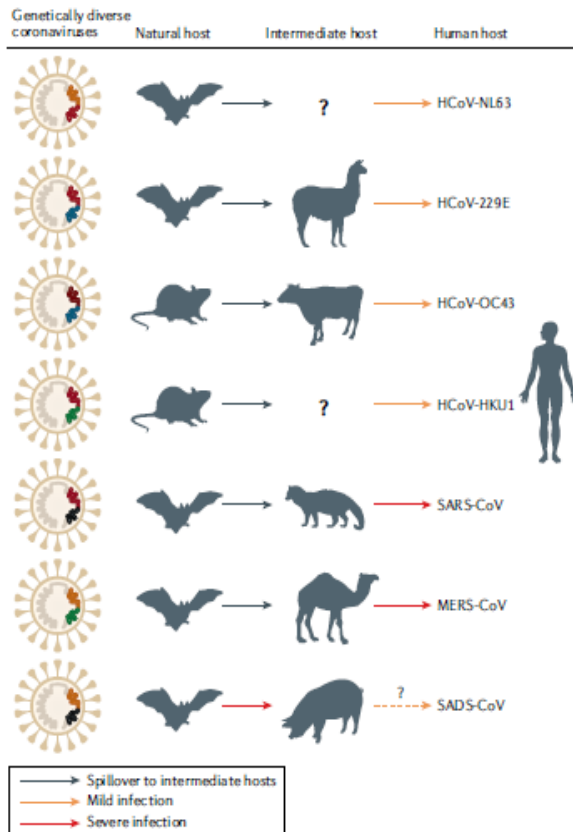


Ilustración 1. Taxonomía de la familia *Coronaviridae*. Fuente: Yan et al., 2020.

La organización genómica de los siete coronavirus que infectan a humanos (HCoVs) es muy similar: todos contienen dos grandes marcos de lectura abiertos (*open reading frames*, ORFs en inglés). Ambos codifican para dos poliproteínas que generalmente se dividen en 16 proteínas no estructurales involucradas en el ciclo viral. El resto de ORFs codifican para proteínas estructurales y asociadas. Las proteínas estructurales son la proteína S (*spike*), proteína M (de membrana), proteína E (de envoltura) y proteína N (de la nucleocápside)<sup>4</sup>.

Son virus de ARN monocadena positiva (ARNmc+). Se sospecha que algunos de ellos tienen su origen en murciélagos o en roedores antes de realizar el salto zoonótico a la especie humana. Para ello, se han tenido que adaptar en algún huésped intermediario, algunos de los cuales todavía son desconocidos (Ilustración 2) <sup>2</sup>.



**Ilustración 2. Huéspedes originales y huéspedes intermediarios antes. Flechas negras: paso directo de un huésped a otro. Flechas amarillas: ligera infección. Flechas rojas: infección severa. Fuente: Cui *et al.*, 2020**

Las siete cepas de coronavirus que infectan a la especie humana se caracterizan por generar infecciones en el tracto respiratorio superior, sobre todo en individuos inmunocomprometidos <sup>5</sup>. Las cepas HCoV-NL63, HCoV-229, HCoV-OC43 y HCoV-HKU1 se caracterizan por generar infecciones fácilmente confundibles con un simple resfriado o la gripe<sup>5,6</sup>. En cambio, otros coronavirus como el SARS-CoV, el MERS-CoV o el nuevo SARS-CoV-2, se caracterizan por generar síndromes respiratorios severos además de afectar a otros tejidos como el tracto gastrointestinal, los riñones, el hígado o el tejido neuronal <sup>6</sup>.

### 1.1. Análisis filogenético y clasificación taxonómica del SARS-CoV-2

El estudio de Lu *et al.*, 2020 sobre la caracterización genómica del SARS-CoV-2 concluye que su secuencia genética se encuentra lo suficientemente alejada de la del SARS-CoV como para considerarse un nuevo componente del subgénero *Sarbecovirus* (Ilustración 3)<sup>1</sup>.

En el caso del SARS-CoV-2, se ha tenido en cuenta la región ORF 1a y ORF1ab del genoma del virus para determinar que pertenece al clado 2 de los *Sarbecovirus*<sup>7</sup>. Estas regiones codifican para varias proteínas no estructurales. Algunas de ellas forman el complejo RpRd. Las regiones conservadas de ambos ORFs, con un elevado porcentaje de identidad en comparación a las cepas de otros *Betacoronavirus*, son la prueba definitiva de que el SARS-Cov-2 pertenece al mismo género (con un 90,7%, 90,3% y un 88,9% y 88,7% de similitud con los ORF1a y 1ab de las cepas bat-SL-ZC45 y bat-SL-ZXC21 respectivamente) <sup>8</sup>.

Category	Coronaviruses	Humans	Divergence
Realm	Riboviria		●
Order	Nidovirales	Primates	●
Suborder	Cornidovirineae		●
Family	Coronaviridae	Hominidae	●
Subfamily	Orthocoronavirinae	Homininae	●
Genus	Betacoronavirus	Homo	●
Subgenus	Sarbecovirus		●
Species	Severe acute respiratory syndrome-related coronavirus	Homo sapiens	●

Ilustración 3. Tabla comparativa de la clasificación taxonómica del SARS-CoV-2 y del Homo Sapiens.  
Fuente: Gorbalenya *et al.*, 2020.

El estudio filogenético de Lu *et. al*, 2020 ha comparado las semejanzas genéticas y estructurales del SARS-CoV-2 y de otros virus del SARS y del MERS a través de la secuenciación masiva, con el objetivo de aclarar el origen filogenético del nuevo coronavirus, concluyendo que respecto a su organización y contenido genético presenta una mayor similitud con cepas de SARS de los murciélagos del subgénero *Sarbecovirus* (88% de identidad con la cepa bat-SL-CoV-ZC45 y 98% con la cepa bat-SL-CoV-ZXC21S) que con las cepas de SARS-CoV y del MERS-CoV (del 79% y el 50% de identidad respectivamente). Sin embargo, el estudio de la proteína S reveló que ésta es más similar a las cepas de SARS-CoV a pesar de presentar unas ligeras diferencias en ciertos aminoácidos <sup>8,9</sup>. Estos resultados ratifican la idea de que el huésped original del virus sea el murciélago y que éste se haya adaptado de una forma similar al SARS-CoV, pero con las diferencias suficientes como para formar dos clados distintos dentro del subgénero al que pertenecen (Ilustración 4)<sup>7-9</sup>.

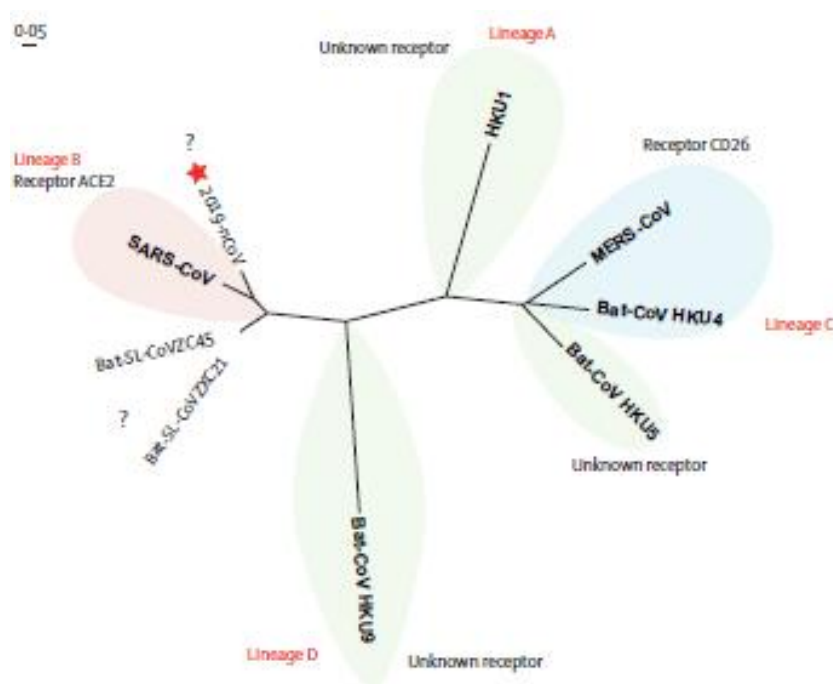


Ilustración 4. Árbol filogenético de los coronavirus del género *Betacoronavirus* que infectan a humanos. Hay 5 subgéneros dentro del género *Betacoronaviurs*, uno de ellos no está representado. Fuente: Lu *et al.*, 2020.

## 1.2. SARS-CoV, características, huéspedes y tratamientos

El antecedente más cercano es el brote de SARS-CoV entre los años 2002 y 2003. A finales de noviembre del año 2002, se detectó un incremento de pacientes con problemas respiratorios en la provincia china de Guangdong. Fue durante los siguientes meses que se reportaron un gran número de casos de una neumonía atípica <sup>6</sup>. Un nuevo patógeno había sido identificado con el nombre SARS-CoV: un nuevo coronavirus (CoV) responsable de causar el Síndrome Respiratorio Agudo Severo (SARS en inglés). El huésped original del SARS-CoV es el murciélago (*Rhinolophus sinicus*), que actúa a su vez como huésped de otras especies de coronavirus<sup>10</sup>. Probablemente el SARS-CoV se adaptó en un huésped intermediario, posiblemente las civetas de las palmeras (*Paguma larvata*) <sup>11</sup>, las cuales, fueron encontradas en uno de los mercados de comida de la provincia de Guangdong.. Este brote del SARS llegó a extenderse hasta a 29 países incluyendo China (su país de origen), Taiwán, Singapur y Canadá<sup>10,11</sup>. El paso del SARS por el mundo dejó más de 8.000 infectados <sup>10,12</sup> con un porcentaje de mortalidad del 10% <sup>10</sup>.

El genoma del SARS-CoV contiene casi 30.000 nucleótidos, que se dividen en 11 ORFs, dos de los cuáles (ORF 1a y ORF 1b o 1b) codifican para dos enzimas importantes del ciclo replicativo del virus (la proteasa y la polimerasa) (Ilustración 5)<sup>10,11</sup>. Estas pautas de lectura son un buen objeto de estudio para los coronavirus porque se tratan de regiones bastante conservadas en el genoma. El resto de genes codifican para proteínas estructurales y no estructurales que determinan la capacidad patogénica del virus: la proteína S (formada por la subunidad 1 y la subunidad 2) que reconoce el receptor de las células que infecta el virus y la proteína de envoltura que permite la entrada del virus a la célula infectada, entre otras <sup>10</sup>.

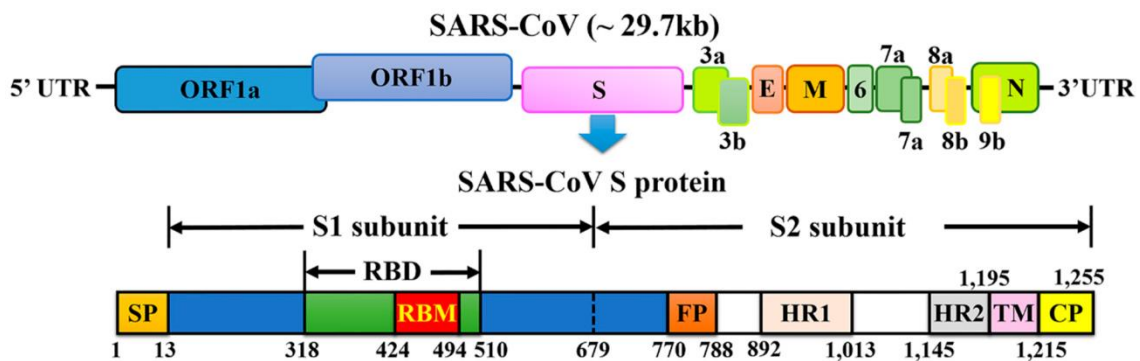


Ilustración 5. Estructura genómica del virus SARS-CoV. Fuente: Song *et al.*, 2019.

Los síntomas del SARS se pueden confundir con los de una gripe común: tos y fiebre y en algunos casos se han reportado disnea y diarrea <sup>12</sup>. La transmisión entre humanos se produce fácilmente a través de fómites respiratorios, tanto que hasta el 22% de los sanitarios llegaron a infectarse en China y alrededor del 40% en Canadá, por lo que la mayoría de las infecciones se produjeron vía nosocomial<sup>11,12</sup>. La técnica de diagnóstico por excelencia es la RT-PCR o real time RT-PCR, aunque existen otras como la detección de anticuerpos, no tan efectivas como la RT-PCR porque necesita un periodo mínimo de incubación para que los anticuerpos sean detectados <sup>10,11,13</sup>.

El tratamiento de elección son los antivirales. Los que han mostrado mejores resultados son el ribavirin, el lopinavir y el ritonavir <sup>12,13</sup>. El primero inhibe la síntesis de ARN mientras que los dos últimos actúan como inhibidores de la proteasa del virus, el lopinavir por ejemplo, es utilizado en pacientes infectador por VIH<sup>11-13</sup>. Se ha observado que la combinación algunos de estos antivirales con el interferón  $\beta$  (INF- $\beta$ ) reduce la respuesta inflamatoria<sup>13</sup>. Otras sustancias también se han utilizado como tratamiento para modular la respuesta inflamatoria como los corticoides o el ácido micofenólico (MPA) <sup>11</sup>.

La proteína S es una de las estructuras clave para que el SARS-CoV pueda producir la infección. Por ello, algunos tratamientos están dirigidos a bloquear su función ya sea impidiendo el ensamblaje de sus estructuras, inhibiendo el reconocimiento ligando-receptor o la fusión con la membrana celular a infectar <sup>12,13</sup>. El problema de esta clase de tratamientos es que dentro de una misma infección pueden aparecer mutaciones, seleccionando indirectamente la supervivencia de los mutantes no susceptibles al fármaco <sup>11</sup>. Finalmente, el uso de plasma de pacientes recuperados fue otra opción terapéutica utilizada en el momento del brote <sup>11</sup> resultando efectiva en los pacientes que se encontraban en las fases iniciales de la infección <sup>12,13</sup>.

### 1.3. MERS-CoV, características, huéspedes y tratamientos

En el año 2012 en la península arábiga surgió un brote de un nuevo virus denominado *Middle East Respiratory Syndrome Coronavirus* (MERS-CoV). En total, se infectaron 2254 personas alrededor de 27 países diferentes con un total de 800 muertes (35% de mortalidad)<sup>10</sup>. El huésped original del MERS-CoV es el murciélago y se demostró que los dromedarios (*Camelus dromedarius*) actuaron como huésped intermediario permitiendo la adaptación del virus y la transferencia al ser humano <sup>10</sup>.

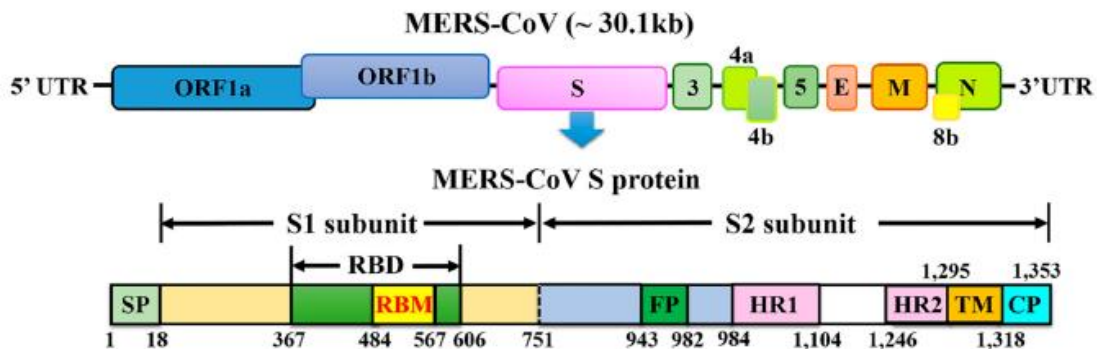


Ilustración 6. Organización genómica del MERS-CoV. Fuente: Song *et al.*, 2019.

A diferencia del SARS-CoV, la proteína S del MERS-CoV reconoce el receptor peptidasa dipeptidil 4 (DPP4 o receptor CD26) <sup>10</sup>. El genoma viral es una cadena simple de ARN positiva, de una longitud de 30,1 kb formado por 10 ORFs: el primer ORF codifica para 16 proteínas no estructurales al igual que en el SARS-CoV y el resto del genoma codifica para las 4 proteínas estructurales principales: S, M, E y N, y otras proteínas accesorias (Ilustración 6) <sup>14</sup>.

El síndrome respiratorio del medio este (MERS) se caracteriza por generar complicaciones en el tracto respiratorio superior, acompañado de síntomas como fiebre, escalofríos, tos y dolor de cabeza. Los casos más graves pueden derivar en diarrea o vómitos. Así como, también se reportaron pacientes asintomáticos<sup>15</sup>. La mayoría de los pacientes hospitalizados en el primer brote, presentaban otras patologías previas como la obesidad (75%), diabetes o hipertensión<sup>15</sup>. El diagnóstico se realiza mediante una RT-PCR que detecta el genoma viral o mediante detección de anticuerpos<sup>15</sup>.

No hay un tratamiento específico para los infectados de MERS, pero a fecha de hoy la medicación más utilizada es el antiviral rivabirin combinado con INF- $\beta$ <sup>13</sup>. Los tratamientos en desarrollo para el MERS-CoV se dividen en vacunas y anticuerpos. Las vacunas están basadas en partículas víricas neutralizadas o en proteínas recombinantes, mientras que los anticuerpos están dirigidos a neutralizar la funcionalidad de la proteína S. Son anticuerpos monoclonales (mAbs) que han mostrado su eficacia *in vitro* y en modelos animales<sup>16</sup>. Los mAbs desarrollados tienen capacidad para identificar diferentes regiones del dominio *receptor binding domain* (RBD) o de la interfaz de conexión entre el dominio RBD y el receptor DDP4. Algunos ejemplos en desarrollo son el MERS-4, CD-27, CDC2-CD o el MCA1. Finalmente, otra estrategia son los nanoanticuerpos, pequeñas porciones del anticuerpo: están formados únicamente por la estructura de reconocimiento del anticuerpo: cuatro cadenas ligeras, lo que los hace accesibles a regiones del dominio RBD que no pueden ser reconocidas por los mAbs. Algunos de estos nanoanticuerpos en estudio son: VHH-1, VHH-1, VHH-83 y VHH-101<sup>17</sup>.

## 2. Epidemiología del SARS-CoV-2, agente causal de la COVID-19

El 21 de enero de 2020, la Organización Mundial de la Salud (OMS) publicó el primer informe sobre la situación del nuevo brote de coronavirus<sup>16</sup>. Dos días después, con un total de 581 de pacientes infectados en 4 países del continente asiático y EEUU, la OMS consideró que el riesgo propagación del virus era elevado a nivel regional y mundial<sup>1,18</sup>. Finalmente el 12 de marzo la OMS declaró la pandemia<sup>16</sup>.

Los estudios realizados apuntan que el SARS-CoV-2 presenta regiones similares a las cepas de SARS obtenidas de murciélago (*bat-SARS-CoV-like*) y que probablemente hay un huésped intermedio, que permitió su adaptación antes del salto zoonótico<sup>19</sup>. Tras el salto entre especies, el virus se ha transmitido de humano a humano, a través de los fómites respiratorios, sobre todo durante el periodo sintomático, permitiendo su dispersión a gran escala<sup>20,21,22</sup>.

A principios de junio, el total de infectados en España fue de 123.974, con un total de 11.413 de pacientes ingresados en las UCI, por un cuadro clínico respiratorio severo con complicaciones respiratorias (ARDS) y 27.127 fallecidos <sup>22</sup>. Como en los brotes anteriores, se han reportado casos asintomáticos <sup>21,22</sup>.

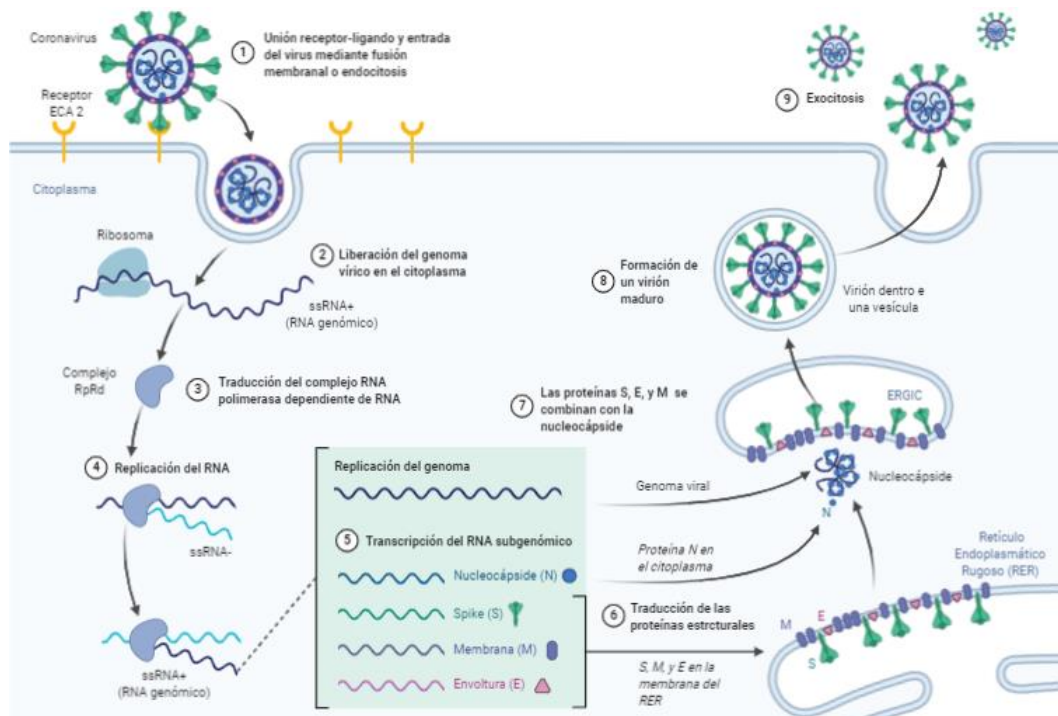
### **3. Clínica de la COVID-19**

La enfermedad causada por la infección del SARS-CoV-2 se denomina COVID-19 (*Coronavirus Disease-2019*). Los síntomas que la caracterizan son la fiebre y la tos. La mayoría de los pacientes corresponden a varones de mediana edad <sup>23</sup> y entre los factores de riesgo encontramos patologías previas como la linfocitopenia, la diabetes, problemas cerebrovasculares, la hipertensión y la obesidad <sup>20,24</sup>.

La diferencia principal entre los síntomas clínicos de la COVID-19 y los anteriores coronavirus, es que los pacientes casi no presentan diarrea (entre el 2 y el 3%) <sup>7,10,11,23</sup>. En el estudio de Ramathan *et al.*, se hizo una comparación sistemática entre el SARS-CoV y el SARS-CoV-2. Los resultados revelaron que ambos coronavirus podían infectar células gastrointestinales y pulmonares, pero que la replicación del SARS-CoV en células del tracto intestinal era mucho más significativa que en el células del tejido pulmonar <sup>23</sup>. Esto podría explicar porque existía una mayor incidencia de diarrea en pacientes de SARS que en pacientes de COVID-19.

### **4. Ciclo viral**

El ciclo viral se realiza gracias al reconocimiento del receptor ECA2 que expresan las células diana, replicándose en el interior de todas aquellas células que lo expresen. El receptor ECA2 se encuentra en células de muchos tejidos: células endoteliales, tejido neuronal, gastrointestinal, renal e incluso en órganos inmunológicos, pero en mayor medida en el tejido pulmonar <sup>9,24</sup>. Esta es la razón por la cual la COVID-19 puede derivar en complicaciones respiratorias severas <sup>7,9</sup>.



**Ilustración 7. Ciclo viral del SARS-CoV-2.**

En función de las características de membrana de la célula diana, el virus puede entrar a través de dos vías: la vía de fusión de membranas o la vía endocítica (Ilustración 7)<sup>24</sup>. La fusión entre membranas se lleva a cabo cuando la célula diana presenta una serie de proteasas ancladas en la membrana plasmática, como las serín-proteasas transmembranales TMPRSS2<sup>24</sup>. Tras la interacción ligando-receptor, las TMPRSS2 desplazan al receptor ECA2 e interactúan con la proteína S. Esto activa a la proteína S y genera un cambio conformacional de la nucleocápside que permite la liberación del genoma en el citoplasma celular<sup>7,9,10</sup>. Por otra parte, en la vía endocítica, el virus entra en la célula a través de la formación de un endosoma tras el reconocimiento proteína *spike*-receptor ECA2. El pH endosómico disminuye y produce la activación de la catepsina L, que permite la liberación del ARN viral en el citoplasma celular<sup>7,10</sup>.

El genoma del SARS-CoV y del SARS-CoV-2 son cadenas ARN positiva (ARNmc<sup>+</sup>), por lo tanto, puede actuar directamente como ARNm para la traducción de proteínas víricas o como cadena molde para la producción de más ARNs genómicos<sup>10</sup>. El ARN viral codifica para proteínas estructurales y no estructurales. Las primeras proteínas en ser traducidas son las proteasas y las proteínas no estructurales que formarán el complejo proteico RnRd (ARN polimerasa ARN dependiente), encargado de formar los genomas virales para los nuevos viriones.

El ORF 1a y el ORF 1ab se traducen y forman dos grandes poliproteínas (pp1a y pp1ab) que son divididas por una proteasa en 16 proteínas no estructurales involucradas en la replicación<sup>9</sup>. Algunas de ellas formarán el complejo RpRd<sup>7</sup>. El complejo RpRd replicará el ARN<sup>+</sup> en cadenas de ARN<sup>-</sup> que se usarán como molde para la producción de más genomas víricos (cadenas de ARN<sup>+</sup>)<sup>24</sup>. Además de replicar el genoma, el complejo RpRd también se encarga de transcribir las cadenas de ARN<sup>+</sup> en otras pequeñas moléculas de ARN subgenómico a través de un cambio en la pauta de lectura ribosomal<sup>7</sup>. Se producen entre 7 y 9 ARNs subgenómicos que codifican para las proteínas estructurales como la proteína S y la proteína E<sup>24</sup>. Éstos se traducen a proteínas y se ensamblan en el retículo endoplasmático rugoso (RER) y el aparato de Golgi (AG)<sup>7,9,10,24</sup>. Simultáneamente, la proteína N se ensambla para la formación de la cápside viral y se une al complejo ERGIC, donde se encuentran las demás proteínas estructurales<sup>10</sup>. A su vez la proteína furina completa la estructura tridimensional de la proteína S<sup>9</sup>. Finalmente se encapsula el ARN viral en su interior y se forma una vesícula que será secretada mediante exocitosis al exterior celular<sup>25</sup>.

El ciclo viral del SARS-CoV y el SARS-CoV-2 son prácticamente iguales. La única diferencia es la eficiencia de reconocimiento del receptor ECA2. Para que se dé el reconocimiento virus-célula huésped, la proteína S interacciona con el receptor ECA2 principalmente, aunque también puede interactuar con otros receptores como el CD209L<sup>1</sup>. Concretamente la interacción del receptor ECA2 se da con el dominio RBD de la proteína S. Los cambios en residuos clave del dominio RBD de la proteína S del SARS-CoV-2 pueden ser los responsables de una mejora en el reconocimiento del receptor ECA2 y la razón por la que el virus haya optimizado su capacidad de transmisión<sup>1</sup>.

## 5. Caracterización genómica del SARS-CoV-2

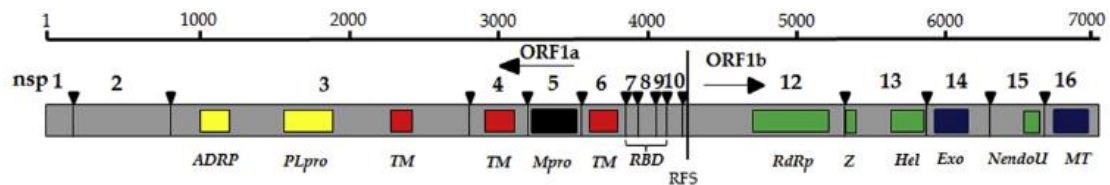
### 5.1. Estructura del genoma. Genes y proteínas



Ilustración 8. Estructura genómica del SARS-CoV-2. Fuente: Wu *et al.*, 2020.

El genoma del SARS-CoV-2 está formado por 29.900 nucleótidos organizados en 14 ORFs que codifican para un total de 27 proteínas (Ilustración 8)<sup>7,9,26</sup>. Los dos primeros (ORF1a y ORF 1ab) comprenden el 67% del genoma y codifican para dos grandes poliproteínas (pp1a y

pp1ab) que se subdividen en 16 proteínas no estructurales (nsps) (Ilustración 9)<sup>26</sup>. Los otros ORFs codifican para proteínas estructurales como la proteína S, proteína E, proteína M, la proteína N y otras proteínas estructurales accesorias entre las cuáles se encuentran: 3a, 3b, p6, 7a, 7b, 8b, 9b y el ORF14<sup>7,9</sup>. Además, en sus extremos 5' y 3' contienen una región UTR que no se transcribe y que flanquea todo el genoma<sup>7,9</sup>.



**Ilustración 9. Representación de los ORFs 1a y 1ab. Fuente: Kumar et al., 2020.**

Las proteínas no estructurales están involucradas en la transcripción y replicación durante el ciclo viral infeccioso. Entre las 16 proteínas no estructurales (nsps) se encuentran: la nsp 3 (*papain-like protease* o PLP) y nsp 5 (3CLpro, *chymotrypsin-like protease* o Mpro) que dividen las poliproteínas e inhiben la respuesta inmunitaria innata del huésped<sup>27,28</sup>, la nsp 6 que inhibe la formación de autofagosomas que puedan eliminar al virus<sup>27,28</sup>, el complejo que forman la nsp 7 y 8 que permiten la unión de la nsp 12 (ARN polimerasa dependiente de ARN) al genoma viral para la replicación, la nsp 13 que actúa como helicasa y la nsp 15, una endoribonucleasa, todas necesarias en la replicación<sup>27</sup>.

Las proteínas estructurales más importantes del SARS-CoV-2 determinan la estructura vírica y la interacción con la célula huésped (Ilustración 10). La glicoproteína transmembranal S se encuentra por toda la superficie viral y es la encargada de mediar la interacción con el receptor ECA2 de la célula diana. La proteína E es la proteína estructural más pequeña de todas y junto con la proteína M se incrustan en la bicapa lipídica que deriva de la membrana plasmática de la célula huésped cuando se liberan los nuevos viriones. Además, la proteína M es la responsable en mantener la estructura vírica estable y promueve la fusión del virus cuando se ha dado la interacción ligando-receptor. Finalmente la proteína N envuelve todo el genoma viral y media el ensamblaje vírico en el complejo ERGIC al final del ciclo vírico en la formación de los nuevos viriones<sup>27</sup>.

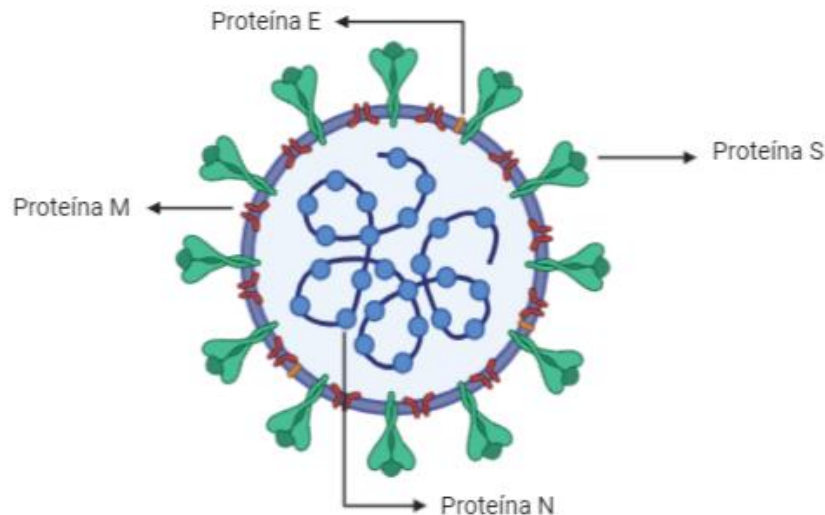


Ilustración 10. Estructura proteica del SARS-CoV-2.

## 5.2. Diferencias genómicas del SARS-CoV-2 en comparación con otros virus de la familia *Coronaviridae*

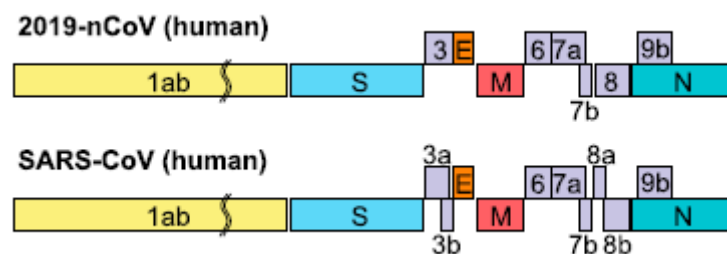
El SARS-CoV-2 presenta modificaciones en su estructura genómica que se traducen en cambios en la estructura tridimensional de las proteínas para las que codifica. La región menos conservada es la que codifica para la proteína S, pero hay otras en las que también se han observado modificaciones que pueden afectar al comportamiento de las proteínas que codifican, como el ORF1ab, el ORF 3b y el ORF8 (Ilustración 11).

El ORF1ab del SARS-CoV-2 presenta cambios que afectan a las proteínas no estructurales, en la región que codifica para la nsp 6, nsp 5, nsp 2 y 3. La proteína no estructural 6 (nsp 6) del SARS-CoV-2 presenta una sustitución en la posición 37 de su cadena polipeptídica. En las cepas que pertenecen al subgénero *Sarbecovirus* contienen una valina mientras que la secuencia polipeptídica de la nsp 6 del SARS-CoV-2 contiene una fenilalanina o una leucina <sup>8</sup>. La nsp 6 forma un dominio transmembranal con los autofagosomas que forma la célula como respuesta contra el virus. La sustitución en la nsp 6 podría modificar sus interacciones con el autofagosoma incrementando su efectividad contra la autofagia <sup>7</sup>. Además, en el estudio de Cárdenas-Conejo *et al.*, 2020, se ha detectado la presencia de una secuencia de 126 nucleótidos en la región del gen de la proteasa parecida a la papaína (nsp 5), que se traduce en 42 aminoácidos parcialmente conservados en otros coronavirus. Se analizó la identidad de esta nueva secuencia del SARS-CoV-2 en otros *Sarbecovirus* del mismo clado: bat-SARS-CoV-RATG13 (76%), bat-SL-CoV-ZC45 (40,5%) y bat-SL-CoV-ZXC21S (16%) <sup>29</sup>. Otros *Betacoronavirus* se incluyeron en el estudio como las cepas de Pangolin-BetaCoV-Guangdong-2019 y Pangolin-BetaCoV-Guangxi-2017 (con un 35% de similitud en la misma región y un 0%

respectivamente)<sup>30</sup>. Finalmente el estudio de Wu *et al.*, 2020, también ha hallado sustituciones en las cadenas polipeptídicas de la nsp 3 (102 aminoácidos diferentes) y de la nsp 2 (61 aminoácidos)<sup>8</sup>.

El ORF 3b codifica para una proteína que inhibe la señal del INF que permite a la célula inhibir la replicación viral<sup>30</sup>. En el SARS-CoV-2 el ORF 3b codifica para una proteína nueva que contiene un codón STOP en la posición 23<sup>29</sup> (afecta al extremo C-terminal de la proteína)<sup>8,29</sup> que inhibe su actividad. No obstante, en el SARS-CoV-2 la pérdida de la actividad de la proteína 3b se sustituye por la proteína del ORF6 y la proteína N<sup>31</sup>.

El ORF8 de las cepas de coronavirus de murciélagos y de civetas de las palmeras codifica para una única proteína<sup>8,31</sup>, mientras que el ORF8 del SARS-CoV sufrió una delección de 29 nucleótidos que generó su división en dos ORFs funcionales<sup>8,29</sup>. Ambos codifican para dos proteínas independientes: la proteína 8a y 8b (de 39 y 89 aminoácidos respectivamente)<sup>1</sup>, implicadas en el desarrollo del estrés intracelular<sup>14</sup>. La proteína 8a del SARS-CoV está ausente en el SARS-CoV-2, por lo que su ORF8 tan solo codifica para la proteína 8b, de 121 aminoácidos que además ha perdido su motivo funcional<sup>32</sup>. Por lo tanto, el ORF8 del SARS-CoV-2 codifica para una proteína diferente a la del ORF8 del SARS-CoV. Ambas secuencias presentan una identidad del 32%<sup>33</sup>.



**Ilustración 11. Comparación de la estructura genómica del SARS-CoV-2 y del SARS-CoV. Los ORF3 y ORF8 forman estructuras nuevas respecto al SARS-CoV. Fuente: Chan *et al.*, 2020.**

Finalmente, la proteína S es la región más variable del genoma. El gen S está dividido en dos subunidades (S1 y S2). La región S1 del SARS-CoV-2 tiene una similitud del 68% con las cepas bat-SL-CoV-ZC45 y bat-SL-CoV-ZXC21 (50 aminoácidos conservados) mientras que la similitud en la región S2 asciende hasta el 93%<sup>32</sup>.

## 5.3. Proteína *spike*

### 5.3.1. Estructura de la proteína *spike* del SARS-CoV-2

La proteína S reconoce la célula huésped y media la entrada del virus al interior citoplasmático<sup>14,32</sup>. En el extremo N-terminal de la proteína está la subunidad S1 que contiene el dominio RBD (reconoce específicamente el receptor), y el extremo C-terminal presenta la subunidad S2 (media la fusión). Las subunidad S1 y S2 contienen las regiones de reconocimiento para las proteasas transmembranales y lisosomales que permiten la entrada del virus a la célula y la liberación del genoma vírico en el citoplasma<sup>14,34</sup>. Ambas subunidades se organizan formando tres monómeros idénticos que generan la estructura trimérica de la proteína S<sup>14</sup>. El dominio RBD de la S1 reconoce específicamente el receptor ECA2. Está formado por un subdominio central y un subdominio externo conocido como RBM<sup>32</sup>.

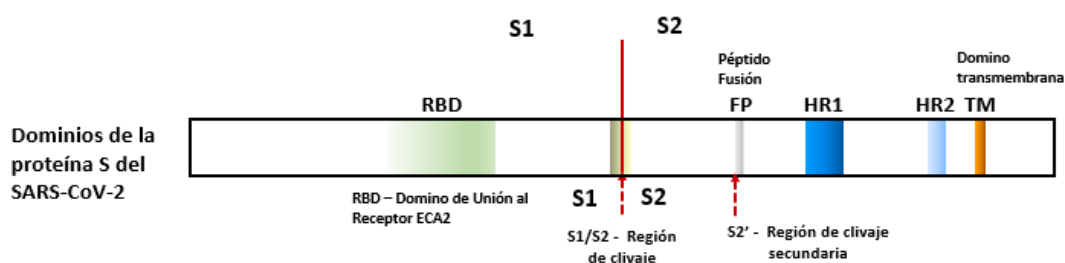


Ilustración 12. Gen de la proteína S con los dominios representados.

La subunidad S2 contiene las regiones HR1 y HR2, que son los principales dominios implicados en la fusión<sup>32,33</sup>. Están basadas en hélices- $\alpha$  que se ensamblan formando una conformación típica que permite el acercamiento de la membrana vírica y la membrana citoplasmática para la fusión<sup>35</sup>. Si observamos la Ilustración 12, se halla la región S2' (S2 *cleavage region*) que permite la exposición de la proteína S a las proteasas. Después hay un péptido de fusión (*fusion peptide*) conectado al dominio HR1 a través de tres hélices- $\alpha$ . Tras el dominio HR1 se encuentra una hélice central (*central helix*) y una horquilla- $\beta$  situada en la base del trímero<sup>34</sup>.

Su secuencia aminoacídica es un 76% idéntica a la del SARS-CoV y un 80% similar a la de las cepas bat-SL-CoV-ZX21 y bat-SL-CoV-ZC45<sup>36</sup>. Las variaciones en la secuencia se deben a las diferencias en el RBD, el dominio de reconocimiento del receptor y que determina el tropismo del virus. Por lo tanto, estos cambios en el dominio RBD pueden ser los responsables del aumento de la capacidad patogénica del SARS-CoV-2.

La proteína S cambia la conformación en función del estado del virus. Existen dos estados tridimensionales para la proteína S: el estado de prefusión y el de postfusión (Ilustración 13). El primero concibe la glicoproteína como una estructura más cerrada de los dominios que forma la S1. Cuando la proteína interacciona con el receptor, se abre y deja al descubierto otras estructuras<sup>18</sup>. Ambos estados dependen de la interacción del virus con las proteasas celulares. Una de las novedades que presenta el dominio RBD del SARS-CoV-2, es que contiene una nueva región de reconocimiento para la proteasa furina, que activa a la proteína S cuando se ha ensamblado el nuevo virión al final del ciclo vírico<sup>34</sup> y por lo tanto presenta una estructura tridimensional diferente a la proteína S del SARS-CoV y del MERS-CoV antes de su activación.

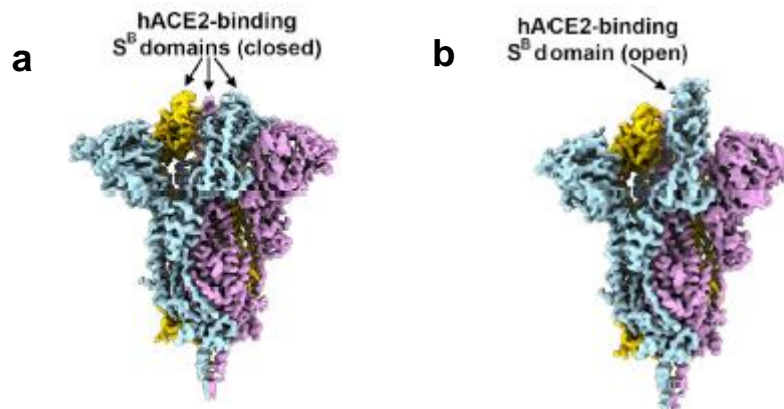


Ilustración 13. (A) Proteína *spike* del SARS-CoV-2 conformacionalmente abierta (B) Conformación cerrada. Los dominios SB hacen referencia a los dominios de la subunidad 1. Fuente: Walls *et al.*, 2020.

### 5.3.1.1. Mutaciones en el dominio RBD

El dominio RBD ocupa la región aminoacídica entre los residuos 270 y el 510 de la proteína S. La identidad entre dominios del SARS-CoV y del SARS-CoV-2 es del 73,3%<sup>35</sup>. La superficie de del RBD está formada por 14 aminoácidos<sup>34</sup>, 6 de los cuales son clave para la interacción con el receptor ECA2. En el SARS-CoV-2 han mutado 5 de los 6 residuos:



Ilustración 14. Alineamiento de secuencias del dominio RBD de la proteína *spike* de diferentes cepas de SARS. Fuente: Andersen *et al.*, 2020.

En la ilustración 14 se muestran alineadas 6 secuencias de diferentes cepas de SARS <sup>18</sup>: SARS-CoV-2, bat-SL-CoV-RATG13, Pangolin, SARS-CoV, dos cepas bat-SARS-CoV relacionadas. Los recuadros señalan los 6 residuos clave del dominio RBD para la interacción con el receptor ECA2. Los residuos que han mutado en el SARS-CoV-2 en comparación con la cepa de SARS-CoV son los siguientes: Y455 > L455, L463 > F463, N493 > Q493, D494 > S494 y T501 > N501.

### 5.3.1.2. Polybasic cleavage site and O linked glycans

La segunda característica única en la secuencia del dominio RBD del SARS-CoV-2 es la una nueva región de reconocimiento para la furina (Ilustración 15). En la región de reconocimiento S1/S1 para las proteasas celulares, se ha hallado una nueva región de reconocimiento para la proteasa furina<sup>18,35</sup>. La secuencia de la proteína S del SARS-CoV-2 contiene una inserción de 12 nucleótidos *upstream* de la región S1/S2 <sup>34,35</sup>, que codifica para dos residuos de arginina, una alanina y una prolina (región polibásica de reconocimiento).

La proteína S de otros SARS (como el SARS-CoV) presenta oligosacáridos que sobresalen de su superficie. Estas estructuras participan en el plegado proteico y en la interacción con otras proteínas <sup>37</sup>. Los nuevos residuos básicos en la región S1/S2 del SARS-CoV-2 interaccionan y dan como resultado la formación de tres enlaces glucosídicos <sup>38</sup>. Éstos permitirían la exposición al exterior de residuos clave para la interacción con la proteasa furina<sup>39</sup>.



Ilustración 15. Alineamiento de la secuencia de la región de reconocimiento de proteasas S1/S2. Se encuentran encuadrados los nuevos residuos que el SAR-CoV-2 ha incorporado en esta región. Fuente: Andersen *et al.*, 2020.

La nueva región hace que la proteína S del SARS-CoV-2 adopte más estados conformacionales que el SARS-CoV o el MERS-CoV, en el ciclo viral durante el ensamblaje del virión, la proteína S permanece inactiva hasta que la proteasa furina reconoce esta nueva región. Su actividad proteolítica genera un cambio en la estructura tridimensional de la proteína S y permite su activación <sup>39</sup>.

## 5.4.Reconocimiento del receptor ECA2

### 5.4.1. Perfil de expresión del receptor ECA2

La enzima convertidora de angiotensina es una proteína con múltiples funciones: regula negativamente el sistema renina-angiotensina, actúa como transportador de aminoácidos y es el receptor de los virus SARS-CoV y SARS-CoV-2<sup>40</sup>. ECA2 se expresa en diferentes tejidos del sistema vascular, corazón, hígado, intestino delgado, riñones, sistema nervioso central y la retina, pero es más abundante en las vías respiratorias superiores y pulmones. Esto explicaría la diversidad de síntomas que se observan en los pacientes de COVID-19. Además, también se ha propuesto que el perfil de expresión del receptor ECA2, en los diferentes tejidos u otros factores de riesgo como el género o la edad, influyen en el desarrollo de la infección.

Las células AT2 (células alveolares de tipo 2)<sup>39</sup> expresan el receptor ECA2 en mayor proporción en el tejido pulmonar<sup>39</sup>. En las células AT2 se ha observado que la expresión del ECA2 está relacionada con la expresión de las proteasas TMPRSS2 involucradas en el ciclo viral<sup>39</sup>. Así mismo también se ha demostrado que la expresión de la proteína furina en el tejido pulmonar se incrementa en células ECA2<sup>+</sup>/TMPRSS2<sup>-</sup> y ECA2<sup>+</sup>/TMPRSS2<sup>+</sup><sup>38</sup>. Las células epiteliales bronquiales humanas primarias (HBEC) también presentan niveles intermedios de expresión del receptor ECA2. En concreto se ha observado que son las *transient secretory cells* HBEC las que presentan una mayor expresión del receptor<sup>37,39</sup>. El estudio de Lukassen *et al.*, 2020 ha demostrado que las *transient secretory cells* presenta un incremento en ciertas enzimas involucradas en el ciclo replicativo viral, por lo que podrían ser más susceptibles a la infección por SARS-CoV-2 añadiendo los niveles significativos de expresión de ECA2 que presentan <sup>41</sup>.

La cavidad bucal es la primera barrera física que se encuentra el virus. Se ha observado que varios tejidos de la lengua y de la mucosa oral contienen elevados niveles de expresión del receptor ECA2: células epiteliales, fibroblastos, linfocitos T y B, macrófagos, mastocitos y células endoteliales convirtiéndola en una región vulnerable, que puede facilitar la entrada del virus<sup>42</sup>.

Otros tejidos expresan el receptor ECA2. En el estudio de Li *et al.*, 2020 los órganos del tracto gastrointestinal, además del corazón, las tiroides y riñones presentaron los niveles de expresión más elevados de ECA2, incluso más que en los pulmones. La razón por la que los pacientes de COVID-19 no suelen presentar síntomas relacionados con estos órganos y tejidos (como la diarrea) es la vía de transmisión. Al ser la cavidad bucal una zona tan susceptible a la

infección, al virus le resulta más fácil llegar al tejido pulmonar que contiene niveles de expresión intermedios y localizar la infección en esa zona que en otras como el intestino <sup>41-43</sup>. También se han tratado de asociar la gravedad de la infección ciertos factores de riesgo como el sexo, la edad o el historial fumador del paciente. Los resultados indican que las diferencias de los niveles de expresión de ECA2 entre hombres y mujeres, jóvenes y adultos no son significativos. Aun así, el riesgo de muerte sí que podría estar asociado con la edad o el sexo debido a la respuesta inmune del huésped, que parece ser mayor en hombres y en adultos <sup>42</sup>.

#### 5.4.2. Optimización del reconocimiento

La interfaz generada entre el RBM del SARS-COV-2 y el receptor ECA2, es muy similar a la interfaz del complejo RBM del SARS-CoV-ECA2. La cristalización de las estructuras ha permitido observar que el solapamiento es prácticamente idéntico (Ilustración 16). Sin embargo, por los cambios en la secuencia del dominio RBD, el RBM del SARS-CoV-2 tiene una superficie de interacción mayor que la del RBM del SARS-CoV, por lo que es capaz de generar un mayor número de uniones y ligeramente distintas, que permiten incrementar la capacidad de interacción con el receptor ECA2. 21 aminoácidos en el RBD del SARS-CoV-2 interactúan con 24 aminoácidos del receptor ECA2, 15 más que con el SARS-CoV. Por lo tanto, se generan más fuerzas de Van der Waals, puentes de hidrógeno e interacciones iónicas <sup>41</sup>.

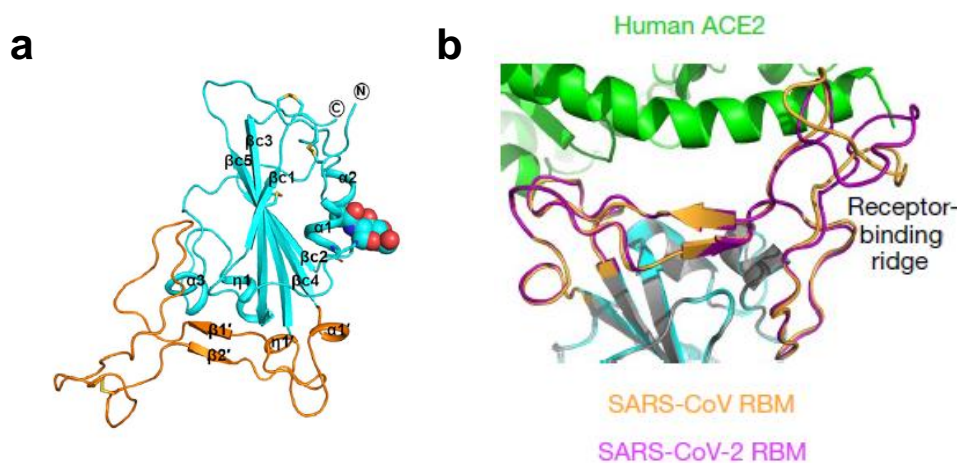


Ilustración 16. (A) Representación tridimensional del *receptor binding motif* (RBM) del SARS-CoV-2. (B) Representación de los RBM del SARS-CoV y SARS-CoV-2 superpuestos interaccionando con el receptor ECA2 humano. RBM del SARS-CoV en naranja, RBM del SARS-CoV-2 en morado, receptor ECA2 en verde. El *receptor-binding bridge* corresponde a la cresta de unión con el receptor. Fuente: Wang *et al.*, 2020, Shang *et al.*, 2020.

Las diferencias entre ambos RBMs que optimizan la interacción con el receptor ECA2 son tres: las variaciones del *loop*  $\beta 1/\beta 2$  en la cresta de unión con el ECA2, la formación de una región de interacciones hidrofóbicas y la optimización de la interacción con los *hotspots* del receptor ECA2.

El *loop*  $\beta 1/\beta 2$  es la región estructural más variable y se encuentra en el RBM del SARS-CoV y del SARS-CoV-2. Su función consiste en generar interacciones con el receptor ECA2 mediante fuerzas de Van der Waals o interacciones iónicas. Los aminoácidos que forman este *loop* en el SARS-CoV y el SARS-CoV-2 son distintos: el SARS-CoV contiene un motivo de 3 residuos distinto al motivo de 4 residuos del SARS-CoV-2. El motivo del SARS-CoV-2 está formado por aminoácidos más flexibles que permiten al *loop* tomar una conformación distinta y más compacta que genera el acercamiento de la alanina 475 al receptor ECA2. Por lo tanto, el cambio conformacional del *loop* permite la formación de más interacciones con la hélice N-terminal del receptor ECA2 <sup>42</sup>.

El extremo C-terminal de la hélice- $\alpha 1$  del RBM del SARS-CoV-2 contiene una leucina en la posición 472, que en el RBM del SARS-CoV-2 es sustituida por una fenilalanina en la posición 486. La fenilalanina adopta una direccionalidad diferente en el espacio y genera una fuerte interacción hidrofóbica con la tirosina de la posición 83 del receptor ECA2. La sustitución de L472 por F486 favorece la unión con el receptor ECA2 (Ilustración 17)<sup>41-43</sup>.

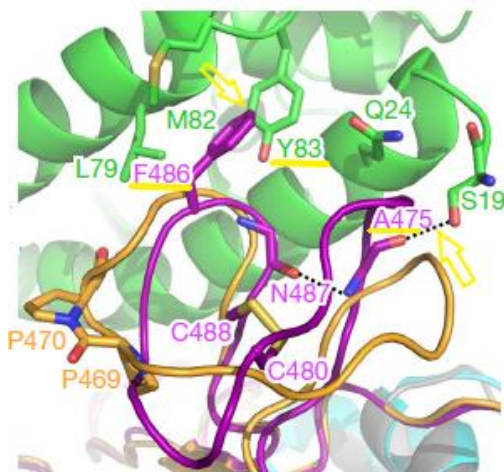


Ilustración 17. Representación de la interacción entre la fenilalanina de la posición 486 del RBM del SARS-CoV-2 con la arginina de la posición 83 del receptor ECA2. Fuente: Shang *et al.*, 2020.

El receptor ECA2 contiene dos *hotspots* o puntos calientes de interacción: son el residuo lisina en posición 31 (denominado *hotspot* 31) y los residuos lisina en posición 353 y ácido aspártico en posición 38 (denominado *hotspot* 353).

Los aminoácidos que interactúan con el *hotspot* 31 en el SARS-CoV-2 y el SARS-CoV son diferentes (Ilustración 18). En el SARS-CoV-2 es una leucina en posición 455 con una cadena lateral poco voluminosa. La conformación de la cadena ha provocado un reajuste del *hotspot* 31 que ha generado la formación de un nuevo puente de hidrógeno. Con el *hotspot* 353 ocurre algo parecido formando un nuevo puente de hidrógeno. Ambos *hotspot* reajustan su estructura conformacional y generan nuevas interacciones que estabilizan la unión del RBM del SARS-CoV-2 y del receptor ECA2 de una forma más eficaz que el SARS-CoV <sup>42</sup>.

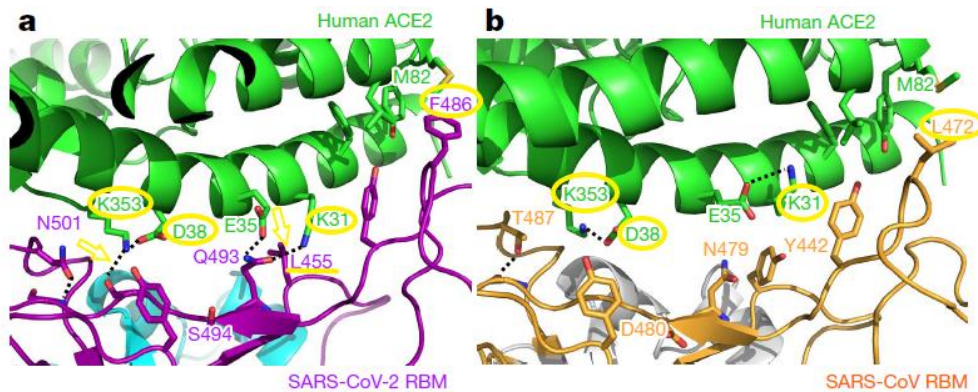


Ilustración 18. (A) Interacciones entre el RBM del SARS-CoV-2 y el receptor ECA2. (B) Interacciones entre el RBM del SARS-CoV y el receptor ECA2. Se señalan en amarillo los cambios de interacciones entre un RBM y otro.

Fuente: Shang *et al.*, 2020.

# **OBJETIVOS**

## **1. Objetivo principal**

Identificar las mutaciones nucleotídicas y aminoacídicas de 28 secuencias de SARS-CoV-2 de pacientes del hospital Vall d'Hebrón de Barcelona. A partir de dicho análisis, identificar los codones mutados, las proteínas afectadas y qué tipo de mutación producen.

## **2. Objetivo secundario**

Revisar si las mutaciones halladas corresponden con las de otros análisis de variabilidad genómica del SARS-CoV-2 y si presentan una frecuencia similar.

# METODOLOGÍA

## 1. Secuencias de estudio

Los datos utilizados son 28 secuencias nucleotídicas de SARS-CoV-2 que se obtuvieron en la base de datos de la GISAID. A través de la cuenta personal de GISAID se buscaron las secuencias del Hospital Universitario Vall d'Hebrón de Barcelona (HUVH). Mediante el buscador *virus name* "hCoV-19/Spain/Barcelona" y *host* "Human" se obtuvieron 26 de las 28 secuencias. Las dos secuencias restantes se obtuvieron a través de su *Accession ID* "EPI\_ISL\_418860" y "EPI\_ISL\_418861".

## 2. Alineamiento de las secuencias

Para el análisis de las mutaciones de las secuencias se utilizó el programa bioinformático BioEdit y a través de un alineamiento múltiple se analizaron las secuencias. Para ello, se ha abierto en Bioedit una ventana con las 28 secuencias de la GISAID. A través de la opción *Accessory Application > ClustalW Multiple Alignment* se obtuvo el alineamiento de las secuencias y una secuencia consenso originada a partir de las 28 muestras.

### 2.1. Análisis de las mutaciones

Para el análisis de las diferencias en los nucleótidos se seleccionó la opción *View conservation by plotting identity to a standard as a dot*. Esta opción permite observar tan sólo los nucleótidos de la primera muestra (estándar). El resto de las secuencias se observan como un punto en los nucleótidos que son iguales y muestra la letra de las bases que son distintas a la estándar.



**Ilustración 19. Programa BioEdit con el alineamiento múltiple. La flecha señala un cambio de nucleótido respecto la secuencia estándar. La última secuencia representa la secuencia consenso. Los asteriscos son los nucleótidos que coinciden con la secuencia estándar.**

Conociendo la región exacta de nts que ocupa cada ORF en el genoma, mediante la opción *Edit > Search > Find Next ORF*, se hallaron y anotaron todos los cambios nucleotídicos, su posición exacta en el genoma y la muestra a la que pertenecen en un cuadro Excel que se muestra en el Anexo 1. Posteriormente, se hizo lo mismo pero con la secuencia aminoacídica. Mediante la opción *Sequence > Toggle Translation* se obtuvieron las secuencias traducidas. Para identificar los aminoácidos afectados por las mutaciones en la secuencia nucleotídica, se siguió el siguiente proceso:

1. Se obtuvieron las secuencias aminoacídicas (mediante la ejecución descrita anteriormente).
2. Para la comparación de los aminoácidos se utilizó como secuencia de referencia la primera muestra (al igual que en la comparación de los nucleótidos). Para ello, la secuencia estándar fue traducida o a través del programa BioEdit mediante la ejecución *Sequence > Nucleic acid > Translate > Frame 1*. De esta manera se obtuvo un archivo con las secuencia genómica de la primera muestra con los nucleótidos divididos en codones. Debajo de ésta se encuentran los codones traducidos al aminoácido correspondiente (Ilustración 21).
3. Con el archivo obtenido en el paso 2, se marcaron todas las mutaciones anotadas en la tabla excel del Anexo 1.
4. Una vez halladas todas las mutaciones en el archivo del paso 2, se procedió a identificar los codones con mutaciones nucleotídicas. Por ejemplo:
  - 4.1. Se identificó la posición de la mutación (posición 50nts, ilustración 20).



4.5. Finalmente, se anotó el cambio de aminoácido en la tabla Excel del Anexo 2 (en este caso, la mutación de la posición 50 ha producido un cambio en el residuo 17 de serina > tirosina, ilustración 23)).

			MUESTRAS	
REGIÓN	NTS	AAS	101226372	9461
5' UTR	50	17	S	Y

**Ilustración 23. Ejemplo. Región afectada. Posición nts de la mutación y aminoácido afectado. Se muestra el aminoácido de la secuencia estándar (S) y el de la secuencia mutada (Y).**

Una vez creada la tabla Excel de los cambios de aminoácidos, se identificó qué clase de mutaciones provocan los cambios nucleotídicos de la secuencia genómica (Anexo 3).

## RESULTADOS

En la tabla Excel del Anexo 3 están anotadas las posiciones de cada mutación, a qué residuo afectan, en qué región se hallan, el cambio de codón y de aminoácido que producen y finalmente de qué tipo de mutación se trata (por muestra).

En total, se hallaron 53 posiciones nts mutadas entre las 28 muestras. De las 53:

- 34 son mutaciones no sinónimas. La mutación en la secuencia genómica genera un codón que codifica para un aminoácido distinto.
- 18 son mutaciones sinónimas. La mutación en la secuencia genómica genera un codón que codifica para el mismo aminoácido.
- 4 mutaciones son sin sentido (*non sense*). La mutación en la secuencia genómica genera un codón de terminación que provoca una parada prematura de la traducción.

Los resultados de las mutaciones nucleotídicas se hallan en la tabla Excel del Anexo 1 y los cambios en los aminoácidos se muestran en la tabla Excel del Anexo 2. A continuación, se muestra una tabla resumen de todas las mutaciones nucleotídicas obtenidas y el tipo de cambio que producen en la secuencia peptídica (tabla 1).

Tabla 1. Resumen de las mutaciones nucleotídicas, el cambio de aminoácido que generan y el tipo de mutación.

Posición nts	Cambio de aminoácido	Tipo de mutación	20239		
				Serina→ Cisteína	No sinónima
50	Serina→ Tirosina	No sinónima	20268	Leucina→ Leucina	Sinónima
241	Cisteína→ Arginina	No sinónima	20762	Treonina→ Isoleucina	No sinónima
313	Glutamina→ Codón de terminación	<i>Non sense</i>	20844	Prolina→ Prolina	Sinónima
922	Glicina→ Arginina	No sinónima	21625	Asparagina→ Aspartato	No sinónima
3036	Fenilalanina→ Fenilalanina	Sinónima	21694	Isoleucina→ Valina	No sinónima
5176	Asparagina→ Aspartato	No sinónima	22295	Treonina→ Isoleucina	No sinónima
5230	Valina→ Leucina	No sinónima	22893	Codón de terminación→ Codón de terminación	Sinónima
6354	Valina→ Valina	Sinónima	23403	Glicina→ Glicina	Sinónima
6894	Leucina→ Fenilalanina	No sinónima	23430	Cisteína→ Cisteína	Sinónima
7735	Arginina→ Cisteína	No sinónima	24793	Cisteína→ Serina	No sinónima
8739	Tirosina→ Tirosina	Sinónima	25485	Alanina→ Alanina	Sinónima
8782	Prolina→ Serina	No sinónima	25979	Glicina→ Valina	No sinónima
9477	Fenilalanina→ Leucina	No sinónima	27215	Valina→ Alanina	No sinónima
9988	Glutamina→ Codón de terminación	<i>Non sense</i>	27302	Asparagina→ Serina	No sinónima
11083	Valina→ Leucina	No sinónima	28144	Tirosina→ Histidina	No sinónima
13426	Arginina→ Codón de terminación	<i>Non sense</i>	28325	Triptófano→ Leucina	No sinónima
14408	Leucina→ Prolina	Sinónima	28863	Fenilalanina→ Fenilalanina	Sinónima
14481	Histidina→ Histidina	Sinónima	28881	Codón de terminación→ Codón de terminación	Sinónima
14487	Arginina→ Arginina	Sinónima	28882	Codón de terminación→ Codón de terminación	Sinónima
14805	Tirosina→ Tirosina	Sinónima	28948	Glutamina→ Codón de temrinación	<i>Non sense</i>
15738	Fenilalanina→ Fenilalanina	Sinónima	29688	Glutamina→ Histidina	No sinónima
16393	Prolina→ Serina	No sinónima	29734	Glutamato→ Glutamina	No sinónima
17211	Leucina→ Fenilalanina	No sinónima	29866	Metionina→ Cualquier aminoácido	No sinónima
18462	Prolina→ Prolina	Sinónima	29867	Metionina→ Cualquier aminoácido	No sinónima
18744	Tirosina→ Tirosina	Sinónima	29868	Metionina→ Cualquier aminoácido	No sinónima

## DISCUSIÓN

Los resultados indican que la región con una mayor tasa de mutación es el ORF1a. De todas las proteínas no estructurales que se hallan codificadas en el ORF1a, la proteína nsp3 (PLP o *papain-like protease*) es la que acumula más cambios (tabla 1). 4 de las 6 mutaciones de esta región son no sinónimas. Además, la mutación 5230nts de la PLP está presente en el 14,28% de las muestras analizadas (4/28). La PLP juega un rol importante durante el ciclo replicativo. Es una proteína transmembrana que se sitúa en el retículo endoplasmático y está formada por varios dominios con acción proteolítica. Ésta, como la Mpro, se encarga de hidrolizar las poliproteínas (pp1a y pp1ab) y de inhibir la respuesta inmune del huésped<sup>44</sup>.

La siguiente región que muestra más variabilidad es la proteína S. Presenta 7 mutaciones (tabla 1), cuatro de las cuáles son no sinónimas. La proteína S es la proteína de reconocimiento del receptor ECA2 humano. Como se ha comentado anteriormente, los cambios en la secuencia peptídica de la proteína S han permitido el aumento de su capacidad patogénica. Si estas mutaciones no sinónimas favorecen el reconocimiento del receptor y son persistentes en otros análisis de variabilidad del SARS-CoV-2, podrían ser un indicativo de que el virus continúa en un proceso de adaptación y mejora del reconocimiento del receptor ECA2.

Finalmente, la proteína N es la última región que presenta más variabilidad en las muestras analizadas (tabla 1). Presenta 5 mutaciones, 3 de las cuáles son sinónimas, una no sinónima y por última una sin sentido (*non sense*). Las mutaciones sin sentido provocan que un codón que codificaba para un residuo sea un codón de terminación, pudiendo provocar una parada prematura de la traducción. De las 53 mutaciones, solo el 14,28% (4/28) son mutaciones sin sentido. Estas se hallan en: la región 5' UTR (no codifica para ninguna proteína), la proteína N y en el ORF1a, concretamente en la nsp 4 y en la nsp 10.

Cabe destacar que, de las 28 secuencias analizadas, hay dos que presentan un 100% de homología entre ellas y con la secuencia consenso, es decir, no presentan ninguna mutación (101211438 y 101216837).

Los genomas víricos de ARN tienen una tasa de mutación de  $10^{-4}$  pares de bases cada año en cada ciclo de replicación<sup>45,46</sup>. El SARS-CoV-2 presenta una homología del 99,99% entre las secuencias de los virus aislados de diversos pacientes de COVID-19<sup>1,45,46</sup>, algo que indica que la variación de la secuencia genómica de este coronavirus es muy baja. La mayoría de las regiones presentan una elevada homología entre las diferentes secuencias de SARS-CoV-2, incluso algunas de ellas como las regiones E, 6 y 7b presentan un 100% de homología<sup>45</sup>. Aun

así, se pueden observar en los resultados de la tabla 1 que existen mutaciones puntuales en su genoma que generan cambios en la secuencia peptídica.

Según otros estudios como los de Dorp *et al.*, y Wang *et al.*, las regiones más hipervariables del SARS-CoV-2 son el ORF 1a y el ORF8 con una tasa mutagénica del 29,47% y del 30,53% respectivamente<sup>45,46</sup>. Otras regiones variables menos frecuentes que las anteriormente mencionadas son la región 3a, S y N<sup>45,46</sup>. A pesar de que las mutaciones obtenidas de la tabla 1 también se hallan en los estudios anteriormente mencionados, en éstos también se encuentran más mutaciones que no se han obtenido aquí. La cantidad de muestras estudiadas influye en la cantidad de mutaciones obtenidas y en la frecuencia de éstas. Por ejemplo, los estudios de Dorp *et al.*, y Wang *et al.*, sí que coinciden con éste en que la región con una mayor tasa de mutación es el ORF1a, sin embargo, Dorp *et al.*, y Wang *et al.*, muestran que el ORF8 es otra región hipervariable, mientras que en este estudio tan solo se observa una mutación puntual del ORF8 en el 3,57% de las muestras (1/28).

## CONCLUSIONES

La prioridad de la comunidad científica ante la pandemia provocada por el SARS-CoV-2 se ha basado en la búsqueda de un tratamiento que permita reducir el tiempo de recuperación de los pacientes de COVID-19, y en el desarrollo de una vacuna eficiente para la prevención de la enfermedad. Recientemente ha sido aprobado el remdesivir para el tratamiento de pacientes severos y graves de COVID-19<sup>47</sup>. Además, algunos proyectos para el desarrollo de vacunas están en estado muy avanzado<sup>48</sup>. Esto significa que la comunidad científica deberá centrar sus esfuerzos en el estudio del genoma del virus y su evolución. Esto permitirá asociar a mutaciones en la secuencia, con cambios en la patogenicidad y transmisibilidad del virus.

El SARS-CoV-2 tiene una tasa mutagénica muy por debajo del resto de virus de ARN. Las secuencias genómicas presentan entre el 99-100% de homología entre ellas. Aun así, existen algunas posiciones que son propensas a sufrir mutaciones, algo que indicaría que, a pesar de la baja tasa mutagénica del virus, este se encuentra en un proceso de adaptación. Por ello, es importante vigilar estas regiones que en función del tiempo podrían determinar la evolución patogénica del virus.

Las secuencias analizadas en este trabajo son una pequeña representación de los cambios que podría estar sufriendo el virus. Aun así, los análisis deben realizarse con una mayor cantidad de muestras para poder demostrar qué cambios son significativos y saber a qué mutaciones se les debe prestar mayor atención.

En conclusión, el análisis de la secuencia genómica del SARS-CoV-2 es un terreno bastante prematuro en el contexto de pandemia. Hasta hoy en día, los esfuerzos se han centrado en la búsqueda de tratamientos para la COVID-19, por lo que los pocos estudios disponibles sobre la variabilidad del genoma del SARS-CoV-2, no presentan unos datos lo suficientemente sólidos como para deducir si el virus ha incrementado su capacidad patogénica o su transmisibilidad.

## **AUTOEVALUACIÓN**

Mi trabajo de final de grado no iba a ser sobre el SARS-CoV-2. Inicialmente realizaría un proyecto experimental en las instalaciones de la Fundación Vall d'Hebrón Institut de Recerca (VHIR). Debido a la declaración del estado de alarma y al confinamiento, tuve que suspender mi estancia y cambiar la tipología de trabajo experimental a uno bibliográfico. El grupo de investigación en el que iba a trabajar inició un proyecto para la investigación del SARS-CoV-2 y aquello fue lo que me hizo cambiar la temática.

Realizar el trabajo sobre un tema tan novedoso y que actualmente está en curso, ha sido bueno y a su vez muy dificultoso. Des del punto de vista bibliográfico, al ser un tema de investigación puntero, he podido obtener una gran cantidad de información muy reciente, casi toda publicada este año 2020. A su vez, me ha resultado muy complicado enlazar la información y tratar de encontrar un consenso entre los datos disponibles sobre ciertos puntos del trabajo. Esto revela que a pesar de que casi todas las entidades científicas están volcadas en la investigación del SARS-CoV-2, queda mucho por hacer para que la información sea más entendible para las futuras generaciones científicas, que vayan a dedicarse a este sector. A pesar de las dificultades que la bibliografía me haya podido generar, ha resultado muy satisfactorio informarme tan extensivamente sobre el virus que ha provocado una pandemia mundial en pleno siglo XXI. Además, he aprendido a realizar una búsqueda bibliográfica y a utilizar herramientas *in silico*, esenciales para la bioinformática que todo estudiante de ciencias naturales debe saber utilizar.

Finalmente, gracias a la realización del trabajo en este contexto, donde hay mucha información que no siempre es transmitida al público adecuadamente, he aprendido a dudar y a contrastar yo misma los datos disponibles para crear una idea sólida de los conceptos consultados.

## **AGRADECIMIENTOS**

En primer lugar, a mi tutora la Dra. Katy, que me ha guiado en este trayecto con sus correcciones y consejos, y que me ha ayudado a hacer mi trabajo mucho más atractivo.

A mi madre y amigos, por apoyarme durante el confinamiento y por confiar en mí mucho más que yo misma.

Finalmente, quiero destacar la inmensa ayuda que he recibido por parte del Dr. Quer y el grupo de investigación de Hepatitis Virales del Vall d'Hebrón Institut de Recerca (VHIR). Gracias Damir por enseñarme todo lo que sé hacer en el laboratorio y por ser siempre tan atento conmigo. Por último, gracias Josep por permitirme realizar las prácticas curriculares con vosotros el verano pasado, y a pesar de que no haya podido volver a hacer el TFG presencialmente con vosotros, gracias por darme la oportunidad de realizar este trabajo junto con toda la información y ayuda que me has dado. Para finalizar, no hay suficientes palabras para agradecer todas las oportunidades que me estás dando ahora, que sin duda marcarán mi futuro profesional como científica. Eres mi ejemplo a seguir. Espero algún día ser la mitad de buena viróloga y persona que tú.

## BIBLIOGRAFÍA

1. Lu R, Zhao X, Li J, *et al.* (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 395(10224):565-574.
2. Cui J, Li F, Shi ZL. (2019). Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*. 17(3):181-192.
3. Yan Y, Shin WI, Pang YX, *et al.* (2020). The first 75 days of novel coronavirus (SARS-CoV-2) outbreak: Recent advances, prevention, and treatment. *Int J Environ Res Public Health*. 2020;17(7).
4. Forni D, Cagliani R, Clerici M, Sironi M. (2017). Molecular evolution of human coronavirus genomes. *Trends Microbiol*. 25(1):35-48.
5. Bouvier M, Chen WJ, Arnold JC, *et al.* (2018). Species-specific clinical characteristics of human coronavirus infection among otherwise healthy adolescents and adults. *Influenza Other Respi Viruses*. 12(2):299-303.
6. Xu L, Liu J, Lu M, Yang D, Zheng X. (2020). Liver injury during highly pathogenic human coronavirus infections. *Liver Int*. 1-7.
7. Astuti I, Ysrafil. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): an overview of viral structure and host response. *Diabetes Metab Syndr Clin Res Rev*. 14(4):407-412.
8. Wu A, Peng Y, Huang B, *et al.* (2020) Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. *Cell Host Microbe*. 27(3):325-328.
9. Kumar S, Nyodu R, Maurya VK, Saxena SK. (2019). Morphology, genome organization, replication and pathogenesis of Severe Acute Respiratory Syndrome Coronavirus 2. 2:23-31.
10. Song Z, Xu Y, Bao L, *et al.* (2019). From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses*. 11(1).
11. Xu J, Zhao S, Teng T, *et al.* (2014). Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV.

12. Hui D, Zumla A. (2020). Severe Acute Respiratory Syndrome. Historical, epidemiologic, and clinical features. *Infectious Disease Clinics of North America*.
13. Yin Y, Wunderink RG. (2018). MERS, SARS and other coronaviruses as causes of pneumonia. *Respirology*. 23(2):130-137.
14. Daniel E Shumer, Natalie J Nokoff NPS. (2017). MERS-CoV spike protein: a key target for antivirals. *Physiol Behav*. 176(12):139-148.
15. Chafekar A, Fielding BC. (2018). MERS-CoV: Understanding the latest human coronavirus threat. *Viruses*. 10(2).
16. Sohrabi C, Alsafi Z, O'Neill N, et al. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int J Surg*. 71-76.
17. Zhou Y, Yang Y, Huang J, Jiang S, Du L. (2019) Advances in MERS-CoV vaccines and therapeutics based on the receptor-binding domain. *Viruses*. 11(1):1-18.
18. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. (2020). The proximal origin of SARS-CoV-2. *Nat Med*. 26:450-452.
19. Li Q, Guan X, Wu P, et al. (2020) Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*. 382(13):1199-1207.
20. Guan WJ, Ni ZY, Hu Y, et al. (2020). Clinical characteristics of Coronavirus Disease 2019 in China. *N Engl J Med*. 1-13.
21. Huang C, Wang Y, Li X, et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 395(10223):497-506.
22. Chen N, Zhou M, Dong X, et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet*. 395(10223):507-513.
23. Ramanathan K, Antognini D, Combes A, et al. (2020). Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. 19-21.

24. Tang T, Bidon M, Jaimes JA, Whittaker GR, Daniel S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* 178:104792.
25. Gorbalenya AE, Baker SC, Baric RS, *et al.* (2020). The species Severe Acute Respiratory Syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 5(4):536-544.
26. Chen Y, Liu Q, Guo D. (2020). Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol.* 92(4):418-423.
27. Cárdenas-Conejo Y, Liñan-Rico A, García-Rodríguez DA, Centeno-Leija S, Serrano-Posada H. (2020). An exclusive 42 amino acid signature in pp1ab protein provides insights into the evolutive history of the 2019 novel human-pathogenic coronavirus (SARS-CoV-2). *J Med Virol.* 92(6):688-692.
28. Benvenuto D, Angeletti S, Giovanetti M, *et al.* (2020). Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *Journal of Infection.*
29. Cagliani R, Forni D, Clerici M, Sironi M. (2020). Coding potential and sequence conservation of SARS-CoV-2 and related animal viruses. *Infect Genet Evol.* 83:104353.
30. Chan JFW, Kok KH, Zhu Z, *et al.* (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect.* 9(1):221-236.
31. Khailany RA, Safdar M, Ozaslan M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Reports.*
32. Yuan Y, Cao D, Zhang Y, *et al.* (2017). Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun.* 8(China CDC):1-9.
33. Song W, Gui M, Wang X, Xiang Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog.* 14(8):1-19.
34. Alexandra C. Walls, Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire DV. (2020). Structure, Function, and Antigenicity of the SARSCoV- 2 Spike Glycoprotein. *Cell.*

35. Coutard B, Valle C, Lamballerie X De, Canard B, Seidah NG, Decroly E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*.
36. Kumar S, Maurya VK, Prasad AK, Bhatt MLB, Saxena SK. (2020). Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). *VirusDisease*. 31(1):13-21.
37. Li MY, Li L, Zhang Y, Wang XS. (2020). Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect Dis Poverty*. 9(1):1-7.
38. Xu H, Zhong L, Deng J, *et al.* (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *Int J Oral Sci*. 12(1):1-5.
39. Lukassen S, Chua RL, Trefzer T, *et al.* (2020). SARS -CoV-2 receptor ACE 2 and TMPRSS 2 are primarily expressed in bronchial transient secretory cells. *EMBO J*. 1-15.
40. Gheblawi M, Wang K, Viveiros A, *et al.* (2020). Angiotensin-Converting Enzyme 2: SARS-CoV-2 receptor and regulator of the renin-angiotensin system: celebrating the 20th anniversary of the discovery of ACE2. *Circ Res*. 1456-1474.
41. Wang Q, Zhang Y, Wu L, *et al.* (2020). Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. 181(4):894-904.e9.
42. Shang J, Ye G, Shi K, *et al.* (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 581(7807):221-224.
43. 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. *Science*. 367(6485):1444-1448.
44. Báez-santos YM, John SES, Mesecar AD. (2015). The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds COVID-19. *Antiviral Res*. 115:21-38.
45. Dorp L Van, Acman M, Richard D, *et al.* (2020). Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection Genetics and Evolution*.
46. Wang C, Liu Z, Chen Z, *et al.* (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J Med Virol*. 92(6):667-674.

47. First COVID-19 treatment recommended for EU authorisation | European Medicines Agency. <https://www.ema.europa.eu/en/news/first-covid-19-treatment-recommended-eu-authorisation>.
48. World Health Organization. (2020). Draft landscape of COVID-19 candidate vaccines. <https://www.who.int/who-documents-detail/draft-landscape-of-covid-19-candidate-vaccines>.

# ANEXOS

## 1. Tabla Excel de mutaciones nucleotídicas

Tabla 2. Mutaciones nucleotídicas de las secuencias analizadas en BioEdit.

GEN		NTS	MUESTRAS																													
			101226372	101226819	101227280	101227458	101231407	101227773	101211438	101214141	101214488	101214964	101215081	101215325	101215994	101216552	101216837	101218938	101219434	101220851	101221484	101221713	101211449	101217663	101217924	8002	9461	9758	1133	198152683		
ORF 1a	5' UTR	50	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
	241	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
	nsp 1	315	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
	322	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
	3029	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
	5176	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
	5230	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
	6354	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	6894	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
	7735	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	8739	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	8782	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	9477	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
	9888	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	nsp 6	11283	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
	nsp 10	13425	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
ORF 1ab	14408	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
	14481	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
	nsp 12	14487	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
	14805	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
	15738	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
	nsp 13	16393	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	17211	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
	nsp 14	18462	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
	18744	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	nsp 15	20239	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
	20269	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
	nsp 16	20762	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	20844	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
	SPIKE	21625	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
		21694	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
		22295	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C		
22893		A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A			
23405		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
23430		C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
24793		T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
ORF 3b		25483	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
25979		G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
ORF 6a		27215	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T		
		27302	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A		
ORF 8		28144	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T			
		28325	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G		
N		28657	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
		28863	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C			
		28881	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
	28882	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G				
	28883	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G				
	28884	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				
	28948	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C				
	29688	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			
3' UTR	29734	C	G	C	G	C	G	C	G	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A				
	29866	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A				
	29867	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T				
	29868	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G			



### 3. Tabla Excel de mutaciones por muestra

Tabla 4. Resumen de las mutaciones nucleotídicas, peptídicas y el tipo de mutación por muestra.

PROTEÍNA	NTS	AAS	CODÓN ORIGINAL	CODÓN MUTADO	AAS ORIGINAL	CAMBIO DE AAS	TIPO DE MUTACIÓN	MUESTRA
<b>SPIKE</b>	21625	7209	TCT	TAT	Asparagina	Aspartato	No sinónima	101226372
<b>nsp 6</b>	11083	3695	GTA	TTA	Valina	Leucina	No sinónima	101226819
<b>SPIKE</b>	23430	7810	TGC	TGT	Cisteína	Cisteína	Sinónima	
<b>nsp 15</b>	20239	6747	AGT	TGT	Serina	Cisteína	No sinónima	101227280
<b>3' UTR</b>	29734	9912	GAG	CAG	Glutamato	Glutamina	No sinónima	
<b>nsp 10</b>	13426	4476	CGA	TGA	Arginina	Codón de terminación	Non sense	101227458
<b>nsp 3</b>	5176	1726	AAC	GAC	Asparagina	Aspartato	No sinónima	
	6354	2118	GTC	GTT	Valina	Valina	Sinónima	
<b>N</b>	28883	9628	GGG	GAG	Glicina	Glutamato	No sinónima	101231407
<b>3' UTR</b>	29688	9896	CAG	CAC	Glutamina	Histidina	No sinónima	
<b>nsp 4</b>	8739	2913	TAC	TAT	Tirosina	Tirosina	Sinónima	
<b>nsp 14</b>	18744	6248	TAC	TAT	Tirosina	Tirosina	Sinónima	101227773
<b>SPIKE</b>	22893	7631	TAA	TAG	Codón de terminación	Codón de terminación	Sinónima	
<b>N</b>	<b>O</b>		<b>MU</b>	<b>TA</b>	<b>CIO</b>	<b>NES</b>	<b>*</b>	101211438
<b>nsp 2</b>	922	308	GGA	AGA	Glicina	Arginina	No sinónima	101214141
<b>nsp 14</b>	18462	6154	CCG	CCT	Prolina	Prolina	Sinónima	
<b>ORF 3a</b>	25485	8495	GCT	GCC	Alanina	Alanina	Sinónima	101214488
<b>N</b>	28948	9650	CAG	TAG	Glutamina	Codón de terminación	Non sense	
<b>nsp 3</b>	7735	2579	CGT	TGT	Arginina	Cisteína	No sinónima	101214964
<b>nsp 13</b>	16393	5465	CCA	TCA	Prolina	Serina	No sinónima	
<b>SPIKE</b>	22295	7432	ACA	ATA	Treonina	Isoleucina	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101215081
<b>nsp 12</b>	14487	4829	AGA	AGG	Arginina	Arginina	Sinónima	101215325
<b>N</b>	28325	9442	TGG	TTG	Triptófano	Leucina	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101215994
<b>nsp 3</b>	5230	1744	GTG	TTG	Valina	Leucina	No sinónima	101216552
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>N</b>	<b>O</b>		<b>MU</b>	<b>TA</b>	<b>CIO</b>	<b>NES</b>	<b>*</b>	101216837
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101218938
<b>nsp 3</b>	5230	1744	GTG	TTG	Valina	Leucina	No sinónima	101219434
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>nsp 3</b>	5230	1744	GTG	TTG	Valina	Leucina	No sinónima	
<b>nsp 13</b>	17211	5737	TTG	TTT	Leucina	Fenilalanina	No sinónima	101220851
<b>nsp 16</b>	20762	6821	ACA	ATA	Treonina	Isoleucina	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>nsp 3</b>	6894	2298	TTG	TTT	Leucina	Fenilalanina	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101221484
	29866	9956	ATG	NNN	Metionina	Cualquier aas	No sinónima	
	29867	9956	ATG	NNN	Metionina	Cualquier aas	No sinónima	
	29868	9956	ATG	NNN	Metionina	Cualquier aas	No sinónima	
<b>5' UTR</b>	241	81	TGT	CGT	Cisteína	Arginina	No sinónima	
<b>nsp 3</b>	3036	1012	TTT	TTC	Fenilalanina	Fenilalanina	Sinónima	
<b>nsp 4</b>	8782	2928	CCA	TCA	Serina	Prolina	No sinónima	
	9477	3159	TTT	TTA	Fenilalanina	Leucina	No sinónima	
<b>nsp 12</b>	14408	4803	CIT	CCT	Leucina	Prolina	No sinónima	
	14805	4935	TAC	TAT	Tirosina	Tirosina	Sinónima	
<b>nsp 15</b>	20268	6756	TTG	TTA	Leucina	Leucina	Sinónima	101221713
<b>SPIKE</b>	21694	7232	ATC	GTC	Isoleucina	Valina	No sinónima	
<b>ORF 3a</b>	23403	7801	GGG	GGA	Glicina	Glicina	Sinónima	
	25979	8660	GGA	GTA	Glicina	Valina	No sinónima	
<b>ORF 8</b>	28144	9382	TAC	CAC	Tirosina	Histidina	No sinónima	
<b>N</b>	28657	9553	CGG	TGG	Arginina	Triptófano	No sinónima	
<b>3' UTR</b>	28863	9621	TTT	TTT	Fenilalanina	Fenilalanina	Sinónima	
	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101211449
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	101217663
<b>ORF 6a</b>	27215	9072	GTT	GCT	Valina	Alanina	No sinónima	101217924
<b>nsp 3</b>	5230	1744	GTG	TTG	Valina	Leucina	No sinónima	
<b>nsp 12</b>	14481	4827	CAC	CAT	Histidina	Histidina	Sinónima	8002
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>5' UTR</b>	50	17	TCT	TAT	Serina	Tirosina	No sinónima	
<b>nsp 12</b>	15738	5246	TTT	TTT	Fenilalanina	Fenilalanina	Sinónima	9461
<b>nsp 16</b>	20844	6948	CCC	CCT	Prolina	Prolina	Sinónima	
<b>SPIKE</b>	24793	8265	TGC	AGC	Cisteína	Serina	No sinónima	
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>nsp 4</b>	9988	3329	CAG	TAG	Glutamina	Codón de terminación	Non sense	9758
<b>nsp 15</b>	20268	6756	TTG	TTA	Leucina	Leucina	Sinónima	1133
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	
<b>nsp 1</b>	313	105	CAG	TAG	Glutamina	Codón de terminación	No sinónima	
<b>ORF 6a</b>	27302	9101	AAC	AGC	Asparagina	Serina	No sinónima	198152683
	28881	9627	TAG	TA	Codón de terminación	Codón de terminación	Sinónima	
	28882	9628	GGG	ACG	Glicina	Treonina	No sinónima	
	28883	9628						
<b>3' UTR</b>	29734	9912	CAG	GAG	Glutamina	Glutamato	No sinónima	