

Análisis de genómica comparativa: del virus SARS al SARS-CoV-2. Similitudes y diferencias

Comparative genomic analysis: from SARS to SARS-CoV-2 virus-types. Similarities and differences

Irma Berenice Mulato-Briones¹, Rosa María Ribas-Aparicio², Abraham Reyes-Gastellou², Ismael Olan Rodríguez-Ildelfonso² y Mauricio Salcedo-Vargas¹

Resumen

Introducción: en este momento somos testigos de un evento de magnitud mundial provocado por el brote pandémico derivado del nuevo virus SARS-CoV-2, lo cual requiere la generación de conocimiento. Por lo novedoso que resulta, muchas hipótesis y teorías son discutidas a diario respecto al origen de este nuevo virus. Varios estudios están enfocados en demostrar la similitud que el SARS-CoV-2 tiene con otros virus.

Objetivo: resaltar las diferencias del SARS-CoV-2 con otros virus SARS, a partir de un análisis de genómica comparativa, y determinar si se pueden atribuir a eventos de manipulación.

Material y métodos: se descargaron dos genomas completos de virus SARS, seis genomas completos de coronavirus humanos y 16 de coronavirus tipo SARS; fueron analizados en un estudio de genómica comparativa mediante la herramienta BLAST Ring Image Generator, y a continuación se examinaron las diferencias evidentes mediante el uso de los programas MAFFT y BLAST.

Resultados: se observó una alta identidad en fragmentos de los genomas tipo SARS de mamíferos con los genomas SARS-CoV-1 y SARS-CoV-2, y se identificaron tres diferencias nucleotídicas principales: en el gen ORF1ab región nsp3, en el gen S de reconocimiento al receptor y en el gen

Abstract

Background: We are currently witnessing a worldwide event caused by the pandemic outbreak derived from the new SARS-CoV-2 virus, which requires the generation of knowledge. Due to its novelty, many hypotheses and theories are discussed daily regarding the origin of this new virus. Several studies are focused on demonstrating how similar it is to other viruses.

Objective: To highlight the differences of SARS-CoV-2 with other SARS viruses, from a comparative genomics analysis, and determine if these can be attributed to manipulation events.

Material and methods: Complete genomes of two SARS viruses were downloaded, along with other six of human coronaviruses, and 16 of SARS-type coronaviruses. These were analyzed using the BLAST Ring Image Generator tool; afterwards, the evident differences were examined by MAFFT and BLAST programs.

Results: High identity was observed in fragments of the mammalian SARS-like genomes with the SARS-CoV-1 and SARS-CoV-2 genomes, identifying three main nucleotide differences, in the ORF1ab nsp3 region gene, in the receptor recognition S gene, and in the ORF8 gene, with which the

¹Instituto Politécnico Nacional, Escuela Nacional de Ciencias Biológicas Campus Casco de Santo Tomás, Departamento de Microbiología, Laboratorio de Producción y Control de Biológicos "Dr. Mario González Pacheco"; ²Instituto Mexicano del Seguro Social, Centro Médico Nacional Siglo XXI, Hospital de Oncología, Unidad de Investigación Médica en Enfermedades Oncológicas. Ciudad de México, México

Correspondencia:

*Mauricio Salcedo Vargas
E-mail: masava89@gmail.com

Fecha de recepción: 13/04/2020

Fecha de aceptación: 29/05/2020
DOI: 10.24875/RMIMSS.M20000124

Rev Med Inst Mex Seguro Soc. 2020;58 Supl 2:S121-133
<http://revistamedica.imss.gob.mx/>

2448-5667 / © 2020 Instituto Mexicano del Seguro Social. Publicado por Permayer. Éste es un artículo open access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ORF8, con el cual se pueden separar las cepas tipo SARS de mamíferos en tipo SARS-CoV-1 y SARS-CoV-2.

Conclusión: el genoma completo de SARS-CoV-2 posee una alta identidad con cepas tipo SARS de mamíferos, por lo que su aparición más probable podría ser el resultado de la evolución natural.

Palabras clave: Evolución Biológica; Pandemias; Coronavirus; Recombinación Genética; SARS-CoV-2

La actual pandemia que padecemos en el mundo, derivada de la aparición del virus SARS-CoV-2, que inició hace cuatro meses en Wuhan, provincia de Hubei, China, ha causado inquietud por su origen tan abrupto y sobre su posible existencia; en especial porque se encontró un caso inexplicable de agregación de casos de neumonía,¹ similar al producido por los coronavirus del síndrome respiratorio agudo severo (SARS-CoV) y al síndrome respiratorio del Medio Oriente (MERS-CoV).² Cuando el caso se presentó al mundo, la Organización Mundial de la Salud (OMS) anunció el brote de una nueva enfermedad de neumonía, causada por un coronavirus, que se nombró COVID-19; rápidamente se convirtió en una pandemia por su alta contagiosidad y el desenlace fatal en alrededor del 10% de los afectados. Al corte —inicio de abril del 2020—, más de un millón de casos de COVID-19, distribuidos en al menos 100 países, han sido diagnosticados en el mundo.³ Esta abrupta aparición ha generado diversas hipótesis, algunas derivadas de las suposiciones que circulan en las redes sociales y constituyen noticias falsas (*fake news*). Sin embargo, se han publicado artículos que, a pesar de que se orientan al ámbito científico, sugieren una probable manipulación.⁴ Esto ha sucedido a pesar de que se cuenta con datos que, por un lado, muestran el análisis de un alineamiento de secuencias altamente conservadas de los betacoronavirus; con estos datos se descarta la manipulación humana y se concluye que hay un origen derivado de un proceso de selección natural.⁵ Por otro lado, hay trabajos que sugieren una plasticidad genómica alta en virus de ARN, como el SARS-CoV-2, por lo que sería factible especular sobre la posible adquisición de material genético por parte del virus, mediante mecanismos de recombinación homóloga u otros mecanismos con otros coronavirus o con sus hospederos.^{6,7,8} Todos los anteriores son estudios bien fundamentados en evidencia robusta, mientras que las suposiciones de la aparición del virus debido a la intervención científica son menos categóricas. Por lo

SARS-type strains of mammals can be separated into the SARS-CoV-1 and SARS-CoV-2 type.

Conclusion: The complete SARS-CoV-2 genome has high identity with mammalian SARS-type strains, which is why its most probable appearance could be the result of natural evolution.

Keywords: Biological Evolution; Pandemics; Coronavirus; Recombination, Genetic; SARS-CoV-2

anterior, este trabajo se propuso para determinar las similitudes y diferencias en el genoma completo de SARS-CoV-2 con otros virus de ARN, por medio de análisis *in silico*, así como un análisis genético de las diferencias nucleotídicas encontradas.

Material y métodos

Descarga de librerías genómicas

Para llevar a cabo el análisis genómico se descargaron los genomas completos, parciales y las secuencias de aminoácidos del sitio de NCBI (National Center of Biotechnology Information, www.ncbi.nlm.nih.gov), con fecha del 20 de marzo de 2020, en una máquina Laptop DELL Intel Core i7-7700 HQ, CPU @2.80 GHz 7th Gen, sistema operativo de 64 bits, con procesador x64 y una de tarjeta de video NVIDIA 1050 Ti GeForce GTX en ambiente Windows. A continuación, se indican las librerías de genomas completos descargadas:

SARS-CoV-2 *Homo sapiens*:

- NC_045512.2 SARS-CoV-2 isolate Wuhan-Hu-1 (genoma de referencia)
- MT123290.1 SARS-CoV-2 isolate SARS-CoV-2/IQTC01/human/2020/CHN
- MT126808.1 SARS-CoV-2 isolate SARS-CoV-2/SP02/human/2020/BRA
- MT072688 SARS-CoV-2 SARS-CoV-2/61-TW/human/2020/NPL
- MT188340.1 SARS-CoV-2 isolate USA/MN2-MDH2/2020
- SARS-CoV-1
- NC_004718.3 SARS coronavirus
- MERS-CoV
- NC_019843.3 Middle East respiratory syndrome coronavirus

Betacoronavirus humanos:

- NC_006577.2 Human coronavirus HKU1
- NC_006213 Human coronavirus OC43 strain ATCC VR-759

Alfacoronavirus humanos:

- NC_005831 Human Coronavirus NL63
- NC_002645.1 Human coronavirus 229E

Coronavirus tipo SARS en otros mamíferos:

Quirópteros

- MN996532.1 Bat coronavirus RaTG13
- KY770859.1 Bat coronavirus isolate Anlong-112
- KY770860.1 Bat coronavirus isolate Jiyuan-84
- DQ022305.2 Bat SARS coronavirus HKU3-1
- DQ084199.1 bat SARS coronavirus HKU3-2
- DQ084200.1 bat SARS coronavirus HKU3-3
- MK211378.1 Coronavirus BtRs BetaCoV/YN2018D
- KJ473814.1 BtRs-BetaCoV/HuB2013
- JX993987.1 Bat coronavirus Rp/Shaanxi2011

Pangolín

- MT084071.1 Pangolin coronavirus isolate MP789
- Mus musculus
- HQ890526.1 SARS coronavirus MA15 ExoN1 isolate d2ym1

Recombinante

- FJ211859.1 Recombinant coronavirus clone Bat SARS-CoV

Secuencia de aminoácidos de *ORF-8*

- YP_009724396.1 ORF8 protein [SARS-CoV-2]

Secuencia de aminoácidos de *S*, para alineamiento con MAFFT

- YP_009724390.1 surface glycoprotein [SARS-CoV2]
- QHR63300.2 spike glycoprotein [Bat coronavirus RaTG13]
- QIA48632.1 spike protein [Pangolin coronavirus]
- QIA48641.1 spike protein [Pangolin coronavirus]
- QIA48614.1 spike protein [Pangolin coronavirus]
- QIA48623.1 spike protein [Pangolin coronavirus]
- QIQ54048.1 spike protein [Pangolin coronavirus]
- AVP78042.1 spike protein [Bat SARS-like coronavirus]
- AVP78031.1 spike protein [Bat SARS-like coronavirus]
- ADE34766.1 spike glycoprotein [Bat SARS coronavirus HKU3-8]
- ADE34755.1 spike glycoprotein [Bat SARS coronavirus HKU3-7]
- AAZ41329.1 spike glycoprotein [Bat SARS coronavirus HKU3-2]
- ADE34722.1 spike glycoprotein [Bat SARS coronavirus HKU3-4]
- ADE34823.1 spike glycoprotein [Bat SARS coronavirus HKU3-13]
- ADE34812.1 spike glycoprotein [Bat SARS coronavirus HKU3-12]

- AID16716.1 spike glycoprotein [Bat SARS-like coronavirus]
- ACU31051.1 spike protein [Bat SARS Cov Rs806/2006]
- AIA62320.1 spike glycoprotein [BtRs-BetaCoV/GX2013]
- ATO98181.1 spike protein [Bat SARS-like coronavirus]
- QDF43820.1 spike glycoprotein [BtRs-BetaCoV/YN2018A]
- ATO98108.1 spike protein [Bat SARS-like coronavirus]
- QDF43830.1 spike glycoprotein [BtRs-BetaCoV/YN2018C]
- ATO98169.1 spike protein [Bat SARS-like coronavirus]
- ABD75332.1 spike protein [Bat SARS CoV Rm1/2004]
- ATO98193.1 spike protein [Bat SARS-like coronavirus]
- QDF43835.1 spike glycoprotein [BtRs-BetaCoV/YN2018D]
- ATO98120.1 spike protein [Bat SARS-like coronavirus]
- AGC74176.1 spike protein [Bat coronavirus Cp/Yunnan2011]
- AIA62339.1 spike glycoprotein [BtRf-BetaCoV/HeN2013]
- AIA62290.1 spike glycoprotein [BtRf-BetaCoV/HeB2013]
- ARI44809.1 spike glycoprotein [Bat coronavirus]
- AIA62300.1 spike glycoprotein [BtRf-BetaCoV/SX2013]
- ABD75323.1 spike protein [Bat SARS CoV Rf1/2004]
- ABG47060.1 spike protein [Bat CoV 273/2005]
- AIA62340.1 spike glycoprotein [BtRf-BetaCoV/HuB2013]
- AKZ19087.1 spike glycoprotein [Bat SARS-like coronavirus YNLF_34C]
- AKZ19076.1 spike glycoprotein [Bat SARS-like coronavirus YNLF_31C]
- AIA62310.1 spike glycoprotein [BtRs-BetaCoV/HuB2013]
- AGC74165.1 spike protein [Bat coronavirus Rp/Shaanxi2011]
- QDF43815.1 spike glycoprotein [BtRI-BetaCoV/SC2018]
- ARI44799.1 spike glycoprotein [Bat coronavirus]
- AIA62330.1 spike glycoprotein [BtRs-BetaCoV/YN2013]
- ATO98145.1 spike protein [Bat SARS-like coronavirus]
- ANA96027.1 spike protein [Bat coronavirus]

- AIA62277.1 spike glycoprotein [BtRf-BetaCoV/JL2012]
- ANH10613.1 spike protein [Bat coronavirus]
- ASO66810.1 spike glycoprotein [Bat coronavirus]
- ATO98205.1 spike protein [Bat SARS-like coronavirus]
- ALK02457.1 spike protein [SARS-like coronavirus WIV16]
- AAS00003.1 spike glycoprotein [SARS coronavirus GZ02]
- AAP51227.1 spike glycoprotein S [SARS coronavirus GD01]
- ATO98157.1 spike protein [Bat SARS-like coronavirus]
- AGZ48806.1 spike protein [Bat SARS-like coronavirus RsSHC014]
- ATO98132.1 spike protein [Bat SARS-like coronavirus]
- ATO98218.1 spike protein [Bat SARS-like coronavirus]
- ATO98231.1 spike protein [Bat SARS-like coronavirus]
- QDF43825.1 spike glycoprotein [BtRs-BetaCoV/YN2018B]
- AGZ48828.1 spike protein [Bat SARS-like coronavirus WIV1]
- AGZ48818.1 spike protein [Bat SARS-like coronavirus Rs3367]
- AHX37558.1 spike protein [Rhinolophus affinis coronavirus]
- AHX37569.1 spike protein [Rhinolophus affinis coronavirus]
- YP_003858584.1 spike protein [Bat coronavirus BM48-31/BGR/2008]
- ALJ94036.1 spike glycoprotein [BatCoV/BB9904/BGR/2008]

Análisis in silico de genómica comparativa

Todas las librerías se descargaron y almacenaron en formato FASTA. Después, para realizar la comparación de secuencias completas, se utilizó la herramienta de *BLAST BRIG-095*, la cual permitió límites de identidad de entre 50 y 70%. *BLAST Ring Image Generator* (BRIG) es una plataforma de libre acceso que puede ser utilizada en diferentes ambientes, como Windows/Mac/Unix, y que permite hacer comparaciones de un gran número de genomas largos y visualizarlas en forma circular con el ensamblado de los datos. Esta herramienta está disponible en el sitio <http://sourceforge.net/projects/brig> de manera gratuita y accesible.⁹ Para alinear secuencias de interés, se utilizaron los

programas MAFFT, versión 7 para Windows¹⁰ y *Basic Local Alignment Search Tool* (BLAST)(NCBI-NIH).¹¹

Resultados

Estabilidad del genoma del SARS-CoV-2

En general, la comparación de los genomas completos se hizo con la herramienta de *BLAST BRIG-095*. Los genomas de virus SARS-CoV-2 analizados de varias regiones del mundo presentan valores de identidad en más del 99%, virtualmente en el 100% de los casos, con no más de cuatro pares de bases de diferencia con el genoma de referencia, lo que podría indicar que el virus se ha mantenido relativamente estable en su genoma. En otras palabras, independientemente del sitio geográfico de donde provenga el genoma viral, la identidad es alta, y a este nivel, y con estas herramientas, no parece que existan variantes geográfico-específicas (Fig. 1).

Genómica comparativa entre coronavirus humanos

Para determinar si el SARS-CoV-2 posee zonas de identidad con otros coronavirus humanos (alfa y beta), se comparó el genoma de la cepa de referencia de SARS-CoV-2 (NC_045512.2) con otros genomas de coronavirus humanos (ver las librerías). Los resultados de la comparación genómica muestran claramente que el genoma de SARS-CoV-2 conserva mayor identidad, casi en su totalidad con SARS-CoV-1, con excepción del gen *ORF8* y la región de *nsp3* en el *ORF1ab*, por lo que se distinguen tres fragmentos con identidades en cerca del 80% y es evidente que MERS, cuya relevancia médica es similar a la de los SARS, se separa de ambos; se comparten cuatro fragmentos con identidad de alrededor del 70% con los demás genomas de coronavirus, pero sí se puede sugerir una zona conservada para coronavirus en la región del *ORF1ab*, perteneciente a la ARN-polimerasa dependiente del ARN del virus, importante para su replicación y transcripción (Fig. 2), lo que sugiere que todos estos genomas regulan su actividad de replicación de manera semejante.

Comparación con coronavirus no mamíferos

Posteriormente se realizó el análisis comparativo para conocer qué proporción del genoma del SARS-CoV-2

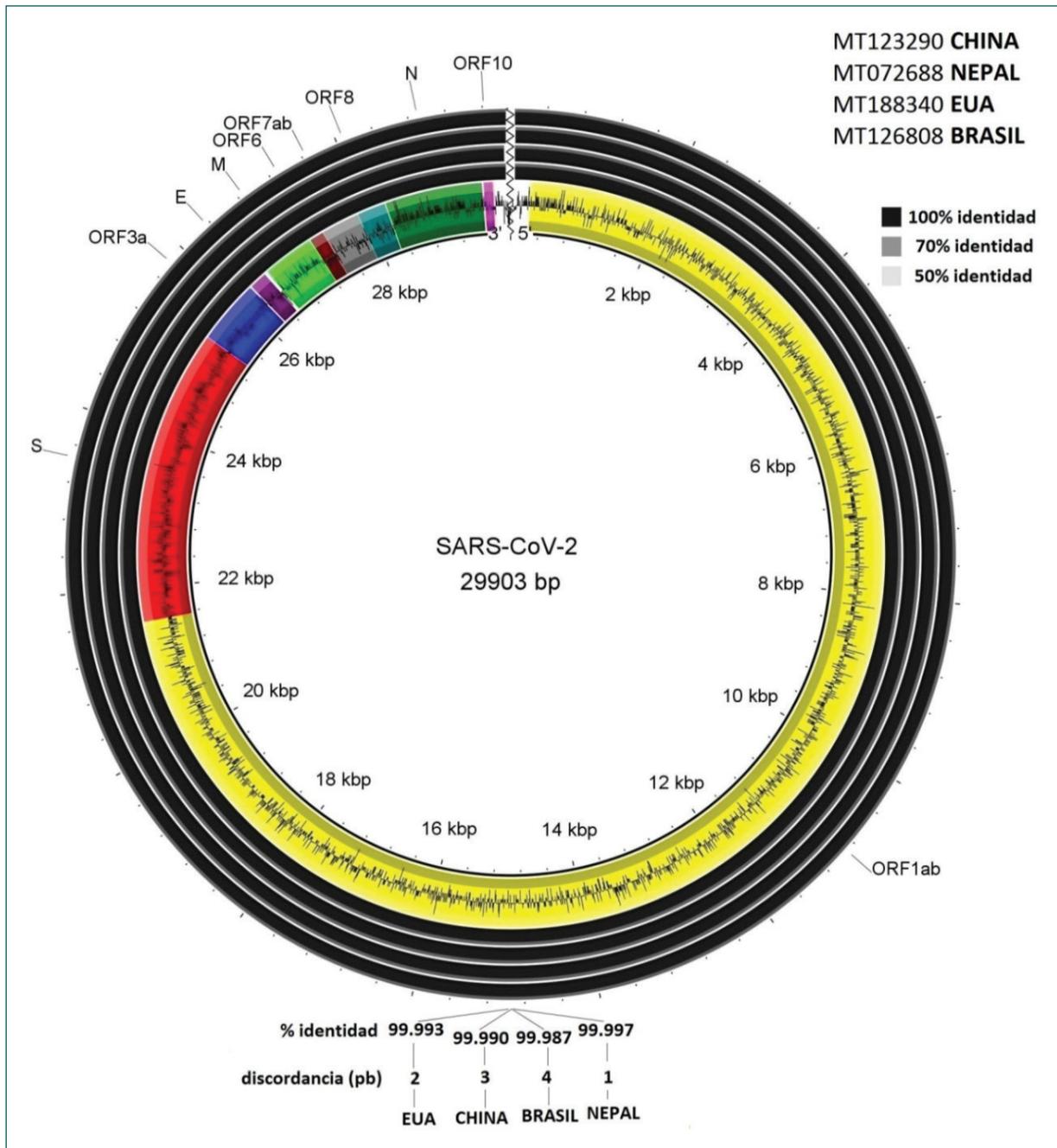


Figura 1. Comparación entre los genomas completos de SARS-CoV-2.

Representación gráfica circular que favorece la visualización de las zonas de similitudes y diferencias entre los genomas estudiados. El círculo central corresponde al genoma de referencia lineal (5'-3') de 29 903 pb NC_04512.2 (SARS-CoV-2 Wuhan-Hu-1), el cual se comparó contra otros cuatro genomas de SARS-CoV-2 (círculos siguientes) de China, Nepal, EUA y Brasil. Se agregaron etiquetas y colores específicos para cada región del virus, señaladas, que corresponden a: ORF1ab (amarillo), S (rojo), ORF3a (azul), E (morado), M (limón), ORF6 (guinda), ORF7ab (gris), ORF8 (celeste), N (verde), ORF10 (rosa). La identidad entre los genomas se visualiza en la intensidad del color: blanco: sin identidad; gris claro: identidades cercanas al 50%; gris: identidades cercanas al 70%; negras: identidades cercanas al 100%; "e-value" < 0.0001.

es similar a genomas SARS en otros mamíferos. La [figura 3](#) muestra que existe más del 80% de identidad (tonalidad en gris) y distintos fragmentos que cubren

casi la totalidad entre los genomas analizados. Aquí se hace evidente que, en específico, las cepas de murciélago (MN996532.1) y pangolín (MT084071.1) de reciente

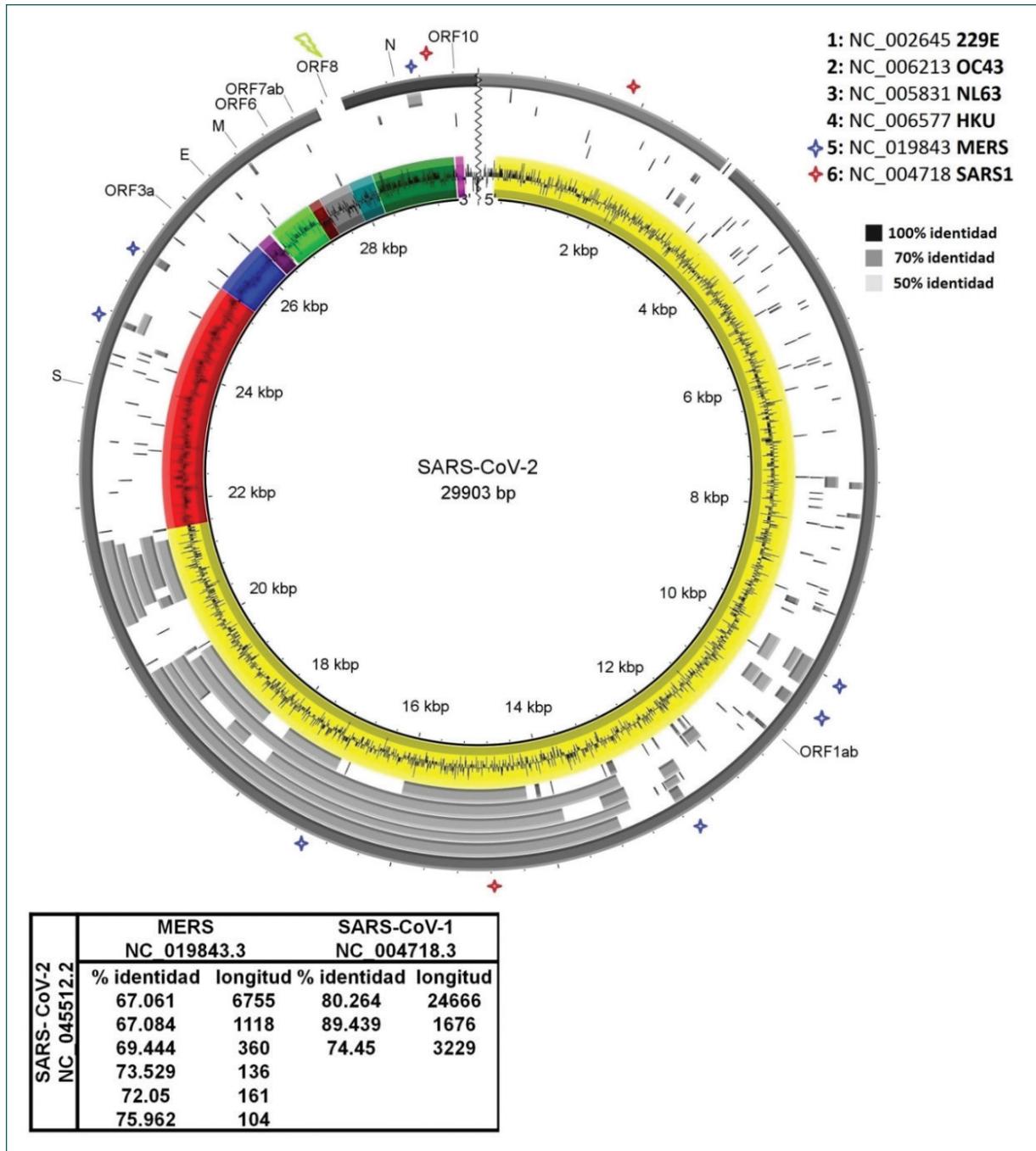


Figura 2. Comparación del genoma completo de SARS-CoV-2 con otros genomas de coronavirus humanos. Representación gráfica circular que favorece la visualización de las zonas de similitudes y diferencias entre los genomas estudiados. El círculo central corresponde al genoma de referencia lineal (5'-3') de 29 903 pb NC_045512.2 (SARS-CoV-2 Wuhan-Hu-1), que se comparó contra otros seis genomas de coronavirus humanos (círculos siguientes) (NC_006577.2 HKU1; NC_004718.3 SARS1; NC_019843.3 MERS; NC_002645.1 229E; NC_006213.3 OC43; NC_005831.1 NL63). Se agregaron etiquetas y colores específicos para cada región del virus, las cuales corresponden a: ORF1ab (amarillo), S (rojo), ORF3a (azul), E (morado), M (limón), ORF6 (guinda), ORF7ab (gris), ORF8 (celeste), N (verde), ORF10 (rosa). La identidad entre los genomas se visualiza en la intensidad del color: blanco: sin identidad; gris claro: identidades cercanas al 50%; gris: identidades cercanas al 70%; negras: identidades cercanas al 100%. Se colocó una tabla de porcentaje de identidades para los genomas de SARS-COV-1 y MERS para cada fragmento con "e-value" < 0.0001, en la que resaltaron los fragmentos anotados y se evidenció la diferencia del ORF8 entre SARS-CoV-1 y SARS-CoV-2.

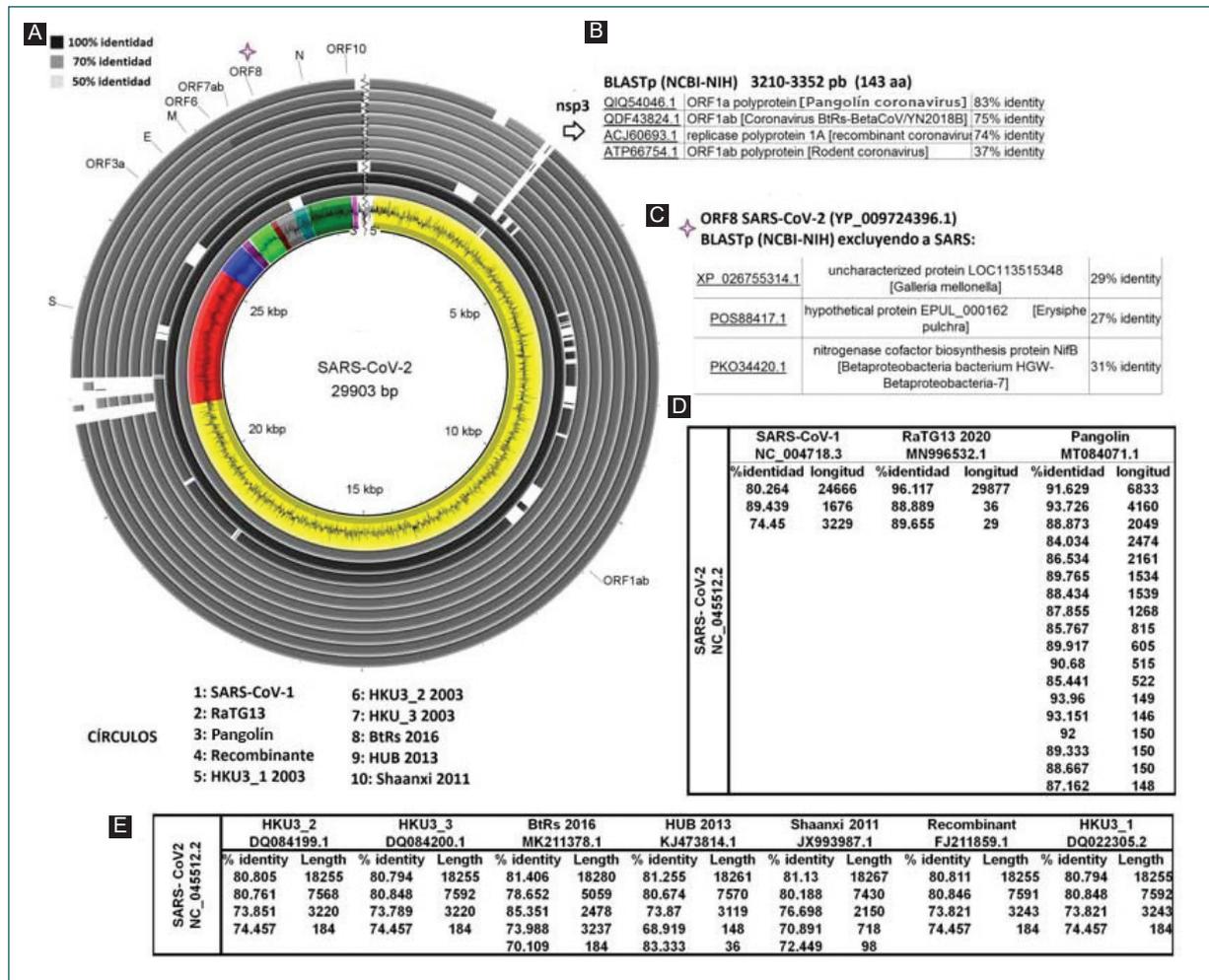


Figura 3. Cepas no humanas tipo SARS-CoV-2 de acuerdo con *ORF8*.

A: gráfico circular que favorece la visualización de las zonas de similitudes y diferencias entre los genomas estudiados. El círculo central corresponde al genoma (29 903 pb) lineal (5' - 3') de referencia NC_045512.2 (SARS-CoV-2 Wuhan-Hu-1), el cual se comparó con otros 10 genomas de virus tipo SARS (círculos siguientes): siete aislados de murciélagos (MN996532.1 RaTG13, DQ022305.2 HKU3-1, DQ084199.1 HKU3-2, DQ084200.1 HKU3-3, MK211378.1 BtRs 2016, KJ473814.1 HuB 2013 y JX993987.1 Shaanxi 2011), uno recombinante de laboratorio consenso de murciélago (FJ211859.1), un aislado de pangolín (MT084071.1) y SARS-CoV-1 (NC_004718.3). Se agregaron etiquetas y colores específicos para cada región del virus, señaladas con etiquetas, las cuales corresponden a: ORF1ab (amarillo), S (rojo), ORF3a (azul), E (morado), M (limón), ORF6 (guinda), ORF7ab (gris), ORF8 (celeste), N (verde), ORF10 (rosa). La identidad entre los genomas se visualiza en la intensidad del color: blanco: sin identidad; gris claro: identidades cercanas al 50%; gris: identidades cercanas al 70%; negras: identidades cercanas al 100% (con el genoma completo de coronavirus del pangolín guarda varios fragmentos de alta identidad, pero dispersos en todo el genoma). Son observables las diferencias en varias regiones de la espiga (S). **B:** se muestran las diferencias con los genomas de murciélagos en ORF1ab (nsp3) región 3210-3352, seguido de una tabla de similitud en aminoácidos con otros coronavirus derivada de un análisis de blastp. **C:** se muestra también la diferencia del ORF8 únicamente con SARS-CoV-1, seguida de la tabla de similitud al realizar un blastp de la secuencia de aminoácidos YP_009724396.1, correspondiente a ORF8 de SARS-CoV-2, análisis realizado excluyendo virus tipo SARS. **D-E:** finalmente se colocaron tablas de porcentaje de identidades de todos los genomas para cada fragmento con "e-value" < 0.0001.

recolección son las que conservan una mayor identidad. En tanto estas, como las cepas cuyos genomas tienen más de cinco años de recolección, como DQ022305.2 HKU3-1, DQ084199.1 HKU3-2,

DQ084200.1 HKU3-3, MK211378.1 BtRs 2016, KJ473814.1 HuB 2013 y JX993987.1 Shaanxi 2011, de murciélagos; así como FJ211859.1, que es una cepa recombinante de laboratorio, muestran también

identidad en el marco de lectura *ORF8* a diferencia de NC_004718.3 SARS1, lo que de alguna manera podría indicar cierta ascendencia con estas cepas, de acuerdo con este gen.

Para inferir el origen y la función posible de este gen, hasta ahora desconocido pero asociado con la patogenicidad y virulencia, se realizó un análisis de BLAST de la región *ORF8* de SARS-CoV-2 (27894-28259 pares de bases [pb]), excluyendo del análisis los virus tipos SARS, ya que sabíamos que contienen un gen con identidad mayor del 80%. Los resultados indican que no se encontraron secuencias similares en otros virus u organismos, por lo que decidimos hacer el análisis con la secuencia de aminoácidos YP_009724396.1 de esta misma región. De manera interesante, se encontraron identidades bajas cercanas al 30% entre unas proteínas no caracterizadas del lepidóptero *Galleria mellonella*, una proteína bacteriana asociada a la biosíntesis del cofactor de nitrogenasas NifB y una enzima tipo RNasa H del hongo *Erysiphe pulchra*. De forma adicional se observaron regiones claras de eliminación en el extremo del gen *S* y en la región 3210-3352 pb *nsp3* del *ORF1ab*, conservada para todas las cepas, razón por la cual también se realizó el análisis de BLAST bajo las mismas condiciones en las que se realizó con *ORF8*, y no se encontró similitud.

Al analizar los 143 aminoácidos (aa) correspondientes de *nsp3*, se encontró que tiene identidad con las cepas ya analizadas MK211378.1 BtRs 2016, y FJ211859.1 recombinante, así como con otros SARS de pangolín y roedores; por lo que se sugiere que esta región de eliminación en realidad representa una mutación de cambio de bases que deriva en la misma proteína.

Finalmente, al observar que existían virus similares al SARS que contenían el gen *ORF8* con identidad al de SARS-CoV-2, se llevó a cabo la comparación genómica y se tomó como referencia el genoma de SARS-CoV-1 (NC_004718.3) (Fig. 4). Los resultados indicaron que nuevamente la cepa de murciélago (MN996532.1), de pangolín (MT084071.1) y la recombinante FJ211859.1 conservan casi el 80% de identidad con SARS-CoV-2 y SARS-CoV-1 y mantienen la eliminación en la totalidad del gen *ORF8*, a diferencia de los virus de murciélago (KY770859.1), (KY770860.1) y el SARS de ratón (HQ890526.1), con los cuales SARS-CoV-1 sí comparte identidad en esta región, por lo que podría ser factible llevar a cabo una clasificación de los virus SARS a partir de la secuencia del *ORF8*.

Variabilidad en la proteína S

Como se observó en el análisis de genómica comparativa, una de las zonas en donde se encuentra mayor variabilidad entre los genomas SARS es la proteína S, o espícula, lo cual es lógico, debido a que su composición dependerá especialmente del hospedero. Por tal motivo, se realizó un alineamiento múltiple de secuencias de aminoácidos para la proteína S. Los resultados muestran que hay diferencias en las secuencias a pesar de presentarse aisladas de los mismos hospederos, como los murciélagos del género *Rhinolophus*, y se corrobora que los SARS de pangolín (RaTG13) son los que guardan una mayor identidad con SARS-CoV-2; también se aprecia que existen cepas que guardan más similitud con SARS-CoV-1. Con el resto de las cepas, ambos SARS se pueden distinguir por la presencia de dos insertos de 5 y 14 aa. Al analizar el péptido de fusión se indica que existen mutaciones puntuales en el primer y sexto aminoácido, pero se observa que la secuencia está conservada en los diferentes coronavirus provenientes de distintos hospederos. Adicionalmente se presenta un inserto de 6 aa NSPRRA (683-688 aa) de S2, justo a un lado del péptido de fusión, en el que se puede apreciar la diferencia de los SARS-CoV-1 con SARS-CoV-2, mientras que el SARS RaTG13 se diferencia solo por 4 aa PRRA, y el SARS de pangolín por 5 aa NSPRS, por lo que, nuevamente, de acuerdo con este inserto, las cepas de murciélago y pangolín conservan una mayor identidad con SARS-CoV-2 (Fig. 5).

Discusión

La manifestación abrupta de la COVID-19 ha generado diversas hipótesis, entre las que sobresale una probable conspiración sobre la creación de un arma biológica, derivada de una “descabellada manipulación de la ingeniería genética”, situación que genera desconfianza en los científicos y su conocimiento en el campo de la biología molecular. Sin embargo, la biología molecular va más allá, pues si la combinamos con la nueva área de la genómica, abre campos para poder elucidar mecanismos ante este tipo de situaciones de índole mundial y regional.¹² Así, una manera de aplicación de los conocimientos y conceptos moleculares es la genómica comparativa, la cual se basa en la búsqueda de las semejanzas y diferencias entre genomas de diferentes organismos, e intenta de alguna manera dar una explicación ante un problema en este caso de salud.⁹

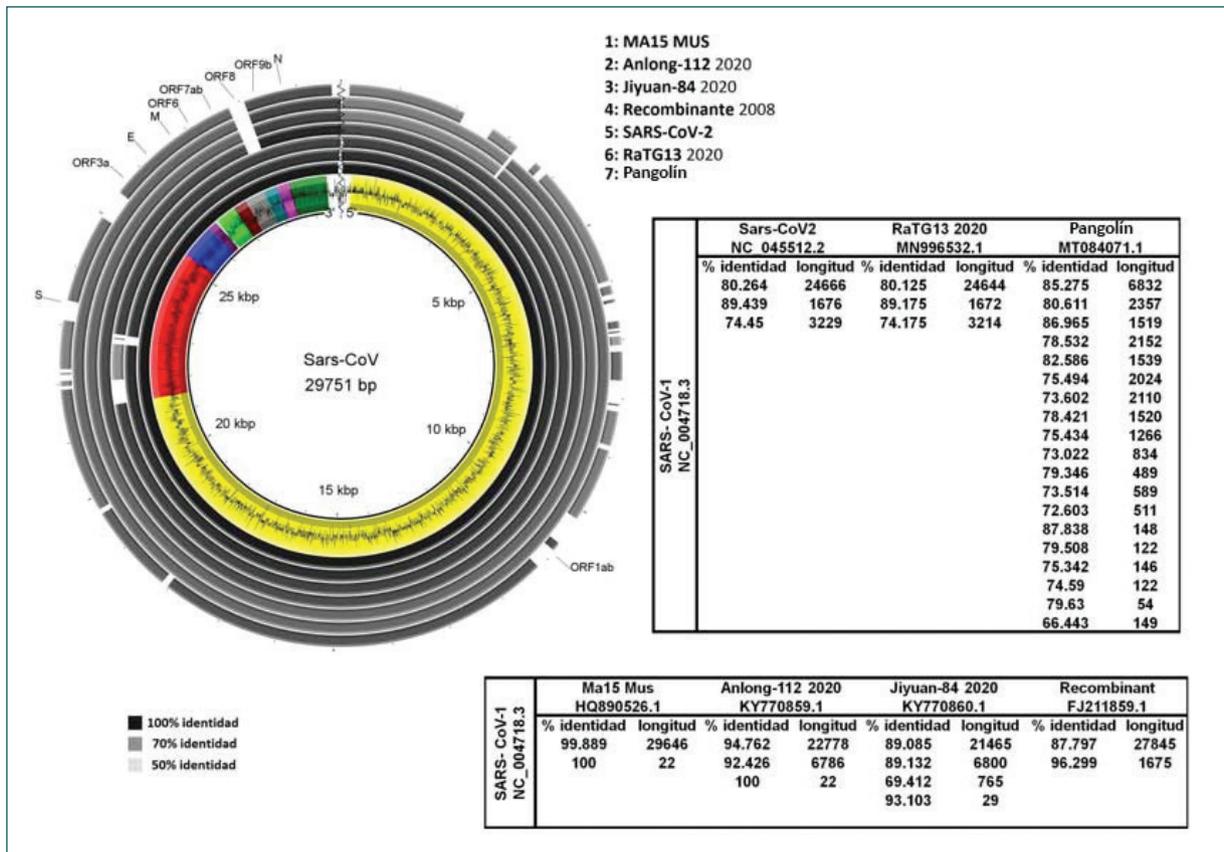


Figura 4. Cepas no humanas tipo SARS-CoV-1 de acuerdo con *ORF8*.

El círculo central corresponde al genoma de referencia NC_004718.3 de SARS-CoV-1, el cual se comparó con otros siete genomas de virus tipo SARS (círculos siguientes): tres aislados de murciélagos (KY770859.1 Anlong-112, KY770860.1 Jiyuan-84 y MN996532.1 RaTG13), uno aislado de pangolín (MT084071.1), el SARS recombinante (FJ211859.1), el SARS adecuado a ratón (HQ890526.1 MA15) y SARS-CoV-2 (NC_045512.2). Se agregaron etiquetas y colores específicos para cada región del virus, señaladas con etiquetas, que corresponden a: ORF1ab (amarillo), S (rojo), ORF3a (azul), E (morado), M (limón), ORF6 (guinda), ORF7ab (gris), ORF8 (celeste), N (verde) y ORF10 (rosa). La identidad entre los genomas se visualiza en la intensidad del color: blanco: sin identidad; gris claro: identidades cercanas al 50%; gris: identidades cercanas al 70% y negras: identidades cercanas al 100%. Se colocó una tabla de porcentaje de identidades de todos los genomas para cada fragmento con "e-value" < 0.0001, lo cual evidenció la disminución del porcentaje de identidad con RaTG y pangolín, y las zonas de no concordancia con ORF8, ORF1ab (nsp3) y S, con SARS-COV-2 y virus similares a este de murciélago, como RaTG, la recombinante y la de pangolín.

Distintas áreas de estudio se están enfocando en buscar soluciones para la COVID-19, como la fisiopatogenia, desde la que se busca explicar los desenlaces de la enfermedad, el pronóstico de esta, la búsqueda inmediata de medicamentos antivirales, la posibilidad de un producto vacunal, así como la determinación de los orígenes del virus.¹³ En este contexto, la idea es aportar evidencias que respalden el origen natural del virus mediante la genómica comparativa, gracias a la rápida aplicación de la secuencia masiva y a la liberación de los datos obtenidos.

Debido a las similitudes del SARS-CoV-2 con los SARS presentes en murciélagos y otros mamíferos,

las cuales han sido propuestas por distintos investigadores^{14,15} y descritas en el presente estudio, proyectamos dos hipótesis. La primera es que todos estos genomas derivaron de un mismo virus ancestral, pero debido a diversas mutaciones, cada uno evolucionó y se adaptó de acuerdo con las condiciones de los hospederos que han colonizado, lo que demuestra e implica que estos virus son altamente mutables; por lo tanto, es probable que el SARS-CoV-2 pudiera cambiar pronto y que algunas de estas mutaciones derivarán en virus más o menos agresivos.^{16,17} Esto apoya parcialmente la información, ya reportada, que refiere que los virus cuyo material genético está compuesto

La segunda hipótesis tiene un gran reto, pues excluye la participación de otros hospederos y vagamente se podría argumentar con el grado de conservación observado en las cepas analizadas en el presente estudio de SARS-CoV-2, las cuales provienen de pacientes de distintos puntos geográficos. Adicionalmente, si revisamos los resultados derivados del estudio de más 326 genomas de pacientes, realizado en 2020 por Zhang *et al.*, en el que sugieren que el genoma del virus conserva una gran estabilidad y que, si bien pueden existir variantes, este no guarda relación con el mercado húmedo de Wuhan, por lo que los genomas se pudieron haber desarrollado de manera independiente desde el surgimiento del brote.¹⁹ De alguna manera, ambos resultados podrían generar una legítima incógnita bajo la lupa de cualquier observador, pero la idea de otros mamíferos hospederos implicaría una cantidad de mutaciones que debieron suceder en las distintas cepas tipo SARS, lo cual representa un “esfuerzo”, por así llamarlo, de cuyas dimensiones no estamos conscientes,²⁰ desde la velocidad y temporalidad de ocurrencia de estas mutaciones, hasta la posibilidad de cambios y adaptación a una infinidad de hospederos, virus, e inclusive otros organismos presentes en los hospederos, con los cuales podría haber intercambio de material genético con mecanismos que aún no se describen para este tipo de virus y que necesariamente deben ser estudiados. Sin embargo, este acertijo no basta para desestimar la primera hipótesis, que contempla la participación de distintos hospederos mamíferos.

De modo que la primera hipótesis reúne más argumentos a favor, pues otros análisis han determinado al margen la posible presencia de tres variantes derivadas de procesos de mutaciones rápidas en SARS-CoV-2. La primera, extendida por América y Australia, posiblemente deriva de virus encontrados en otros mamíferos; la segunda deriva de la primera mediante dos mutaciones que se mantuvieron en Asia, y la tercera deriva de la segunda por una mutación que se esparció por Europa.²¹ De manera complementaria, diversos estudios apoyan la idea de que los virus tipos SARS de otros mamíferos se pueden considerar como sucesores del SARS-CoV-2 y son los pangolines y los murciélagos los hospederos intermediarios más probables, propuestos gracias al análisis filogenético de diferentes genes, como *ORF1ab*, pero principalmente por el estudio de secuencias del gen *S*.^{22,23} En nuestro caso, al analizar todo el genoma no encontramos secuencias o genes del virus SARS-CoV-2, “nuevos/artificiales” o que no pudieran haber descendido evolutivamente (con el esfuerzo que implique) de los virus tipo SARS ya existentes o analizados. Con este

virus en particular, estas características pudieran sugerir una capacidad de mutación rápida entre hospederos para su posible adaptación, creencia que puede ser apoyada con la capacidad de que el SARS-CoV-2 ha demostrado en laboratorio que se puede adecuar rápidamente a otros animales domésticos.²⁴ En este caso analizamos también los genomas de otros coronavirus humanos para ver si las secuencias eran “familiares de acuerdo con el hospedero humano”. Por lo que analizamos de 0C43, HKU, MERS, y SARS, los primeros distan demasiado del SARS-CoV-2; pero el SARS-CoV-1, al igual que los virus tipo SARS de los mamíferos, demostró gran identidad, aunque no completa, de modo que pudimos separarlos, de acuerdo con sus diferencias, en virus tipo SARS-CoV-1 y SARS-CoV-2. Es de resaltar que identificamos tres cambios importantes de nucleótidos. El primero, en el ORF1ab, que a pesar de ser evidente a nivel genoma, al momento de realizar la comparación de ambos fragmentos a nivel secuencia de aminoácidos, resulta en una proteína cuya identidad es concordante con la cepa recombinante FJ211859.1 y otros diversos coronavirus, prueba tal vez burda pero que refiere la capacidad de conservación de su naturaleza. El segundo hueco que denominaremos “cambio estrella,” por ser el sitio ideal en el que se esperarían cambios bajo la teoría de adecuación a un hospedero, lo encontramos en la proteína S, cuyos cambios son directamente importantes en el reconocimiento de los receptores en los probables hospederos.^{15,24} El tercer cambio es el evidente entre estos dos virus y corresponde a la totalidad de la región ORF8, cuyo origen y función aún no están descritos, pues a nivel nucleotídico solo conserva identidad con cepas de coronavirus parecidos a SARS de murciélagos, y a nivel de aminoácidos posee una baja identidad con tres proteínas no virales, una proteína del lepidóptero, en distintas bacterias, y con un hongo; esto último llama la atención, pues en el caso de las RNAsas H generalmente suelen estar involucradas en distintos procesos fundamentales en eucariotas, como replicación y reparación de ADN, recombinación homóloga y transposición o interferencia de ARN.^{25,26} Sin embargo, por los porcentajes de identidad y cobertura, son datos que podrían caer en lo artificioso al querer encontrar el origen o función de la proteína; por lo tanto, solo se queda como un dato observado. Lo que sí se sabe de ORF8 es que su papel está relacionado con la patogenicidad de los virus tipo SARS, ya que distintas mutaciones realizadas en esta región durante la epidemia del virus SARS-CoV-1 derivaron en la atenuación de la cepa,²⁷ lo que nos lleva a las incógnitas de si esta diferencia podría ser importante en la eficiencia de la transmisión y la letalidad de SARS-CoV-2

sobre su contraparte, y si se realizarán mutaciones también en ORF8 de SARS-CoV-2.

Cabe aclarar que no se realizó un análisis filogenético del gen S, como en otros estudios,^{22,23} razón por la cual no sería factible contrastar nuestros datos. Sin embargo, al analizar el inserto (683-688 aa) de S2 en las cepas tipo SARS de murciélago y pangolín, y utilizarlo de manera didáctica como evento evolutivo, bajo la regla de cambios menores a mayores^{26,28} que pudieran haber llevado a la adecuación de la proteína S de SARS-CoV-2 al receptor ACE humano, sería lógico pensar que tanto el murciélago RatG y el pangolín pudieran ser de los últimos hospederos, y que probablemente otros animales pudieran estar involucrados, pues la inserción de cuatro aminoácidos involucraría la inserción de al menos 12 nucleótidos a nivel genómico. Al tratar de buscar el origen del cuarto inserto mediante el alineamiento con otros coronavirus de mamíferos, no se encontró ninguna similitud. Esto pudiera deberse a que es posible que aún no se haya muestreado al animal o a los animales involucrados. Este inserto llamó nuestra atención, debido a que en la fecha de corte de nuestros análisis no se encuentra representado en las estructuras cristalográficas disponibles en las bases de datos como PDB.²⁹ Por esta razón, de momento no es posible inferir si tiene participación en la interacción del virus con el receptor, para lo cual se tendrían que modelar las proteínas, incluido el inserto y estudios adicionales de dinámica, para inferir si tiene una función importante de extensión o conferir alguna otra característica fisicoquímica importante o no para la interacción con el receptor, además de que debe probarse a nivel experimental.³⁰

Otro dato reciente indica que la región genómica entre las bases 21600-22350 bp, que involucra a las regiones ORF1ab y S, es completamente nueva, lo cual sugiere una secuencia “extranjera” a la secuencia de la familia de los coronavirus.³¹ Para ello, Perez³² construyó un genoma hipotético que contenía esta secuencia de nucleótidos y demostró que, tal como sucede en la naturaleza, los eventos biológicos siguen distribuciones tipo ondas de Fibonacci 5 8 13 21 y ahora 34 bp, así como regiones estructurales tipo fractal. Algo por demás interesante es que este tipo de nivel de organización matemática es similar a la secuencia del cromosoma 4 de los humanos, y el mismo autor cuestiona si esta similitud pudiera sugerir también que, debido a la afinidad natural, algún sitio de este cromosoma sería punto de integración para este tipo de virus y, de esta manera, permanece como una infección latente. Mayor cantidad de estudios deben ser realizados para dar respuesta a esta posibilidad.

Respecto a los eventos de recombinación molecular, es importante mencionar que presentan una gran adaptabilidad en diversos hospederos, entre ellos muy cerca de los murciélagos del género *Rhinolophus* (reservorio mamífero); así, no es difícil pensar en esta recombinación genómica, pero lo más complicado es que un nuevo virus sea capaz de infectar a nuevas especies, es decir, para que ocurra la recombinación, los virus divergentes tienen que infectar al mismo organismo simultáneamente y lo más drástico es que, debido a la velocidad de replicación, la recombinación es un evento frecuente que sucede a diario en lugares donde conviven o cohabitan diversos géneros y especies.³³

Con el análisis genómico realizado y con los reportes previos, podemos sugerir que se descarta la presencia de secuencias que pudieran haber sido introducidas en el laboratorio experimental, y con el análisis realizado se corrobora la presencia y la similitud entre diversos virus de ARN de SARS, como resultado de eventos de evolución natural en más de un solo hospedero.

Agradecimientos

Durante la elaboración de este manuscrito, IBMB, ARG e IORI fueron becarios de doctorado por el CONACyT, México, y estudiantes del Programa de Doctorado en Biomedicina y Biotecnología Molecular, ENCB-IPN, Ciudad de México, México. Los autores agradecen al Instituto Politécnico Nacional y al Instituto Mexicano del Seguro Social por su apoyo.

Conflicto de intereses

Los autores han completado y enviado la forma traducida al español de la declaración de conflictos potencial de intereses del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno relacionado con este artículo.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Referencias

1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*. 2020;382(8):727-33.
3. World Health Organization. Coronavirus disease (COVID-19) Pandemic [Internet]. Geneva, Switzerland: WHO; 2020. Disponible en: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>
4. Pradhan P, Pandey A, Mishra A, Gupta P, Tripathi P, Menon M, et al. Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag. *bioRxiv*. 2020;2020.01.30.927871
5. Andersen K, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS CoV 2. *Nat Med*. 2020. Disponible en: <https://www.nature.com/articles/s41591-020-0820-9>
6. Simon-Lorieri E, Holmes E. Why do RNA viruses recombine? *Nat Rev Microbiol*. 2011;9(8):617-26.
7. Bentley K, Evans D. Mechanisms and consequences of positive-strand RNA virus recombination. *J Gen Virol*. 2018;99(10):1345-56.
8. Fhang Z, Li W, Xu H, He Z. Identification Sus scrofa and Mus musculus as Potential Parasitifers of SARS-CoV-2 via Phylogenetic and Homologous Recombination Analysis. *Journal of Cell Signaling*. 2020.
9. Alikhan N, Petty N, Ben-Zakour N, Beatson S. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*. 2011;12:402.
10. Katoh K, Misawa K, Kuma Ki, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30(14):3059-66.
11. Altschul S, Gish W, Miller W, Myers E, Lipman D. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-10.
12. Hadfield J, Megill C, Bell S, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34(23):4121-3.
13. Ahn D, Shin H, Kim M, Lee S, Kim H, Myoung J, et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). *J Microbiol Biotechnol*. 2020;30(3):313-24.
14. Anthony S, Johnson C, Greig D, Kramer S, Che X, Wells H, et al. Global patterns in coronavirus diversity. *Virus Evol*. 2017;3(1):vex012.
15. Ge X, Li J, Yang X, Chmura A, Zhu G, Epstein J, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 2013;503(7477):535-8.
16. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-3.
17. Shereen M, Khan S, Kazmi A, Bashir N, Siddique R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res*. 2020;24:91-8.
18. Grandi N, Tramontano E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front Immunol*. 2018;9:2039.
19. Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. *Nature*. 2020;583(7816):437-440.
20. Baltrus DA. Exploring the costs of horizontal gene transfer. *Trends Ecol Evol*. 2013;28(8):489-95.
21. Forster P, Forster L, Renfrew C, Forster M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc Natl Acad Sci*. 2020;117(17):9241-9243. Disponible en: <https://www.pnas.org/content/117/17/9241>
22. Zhang T, Wu Q, Zhang Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. *Curr Biol*. 2020;30(7):1346-1351.e2
23. Lam T, Shum M, Zhu H, Tong Y, Ni X, Liao Y, et al. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature*. 2020;10.1038/s41586-020-2169-0
24. Shi J, Wen Z, Zhong G, Yang H, Wang C, Huang B, et al. Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*. 2020;368(6494):1016-1020. doi: 10.1126/science.abb7015
25. Majorek K, Dunin-Horkawicz S, Steczkiewicz K, Muszewska A, Nowotny M, Ginalski K, et al. The RNase H-like superfamily: new members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res*. 2014;42(7):4160-79.
26. Moelling K, Broecker F, Russo G, Sunagawa S. RNase H As Gene Modifier, Driver of Evolution and Antiviral Defense. *Front Microbiol*. 2017;8:1745.
27. Muth D, Corman V, Roth H, Binger T, Dijkman R, Gottula L, et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep*. 2018;8(1):15177.
28. Broecker F, Moelling K. What viruses tell us about evolution and immunity: beyond Darwin? *Ann N Y Acad Sci*. 2019;1447(1):53-68.
29. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42.
30. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature*. 2020;581(7807):221-224. Disponible en: <https://www.nature.com/articles/s41586-020-2179-y>
31. Hao P, Zhong W, Song S, Fan S, Li X. Is SARS-CoV-2 originated from laboratory? A rebuttal to the claim of formation via laboratory recombination. *Emerg Microbes Infect*. 2020;9(1):545-7.
32. Perez J. Wuhan nCoV-2019 SARS Coronaviruses Genomics Fractal Metastructures Evolution and Origins. 2020; doi: 10.20944/preprints202002.0025.v2
33. Zheng J. SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. *Int J Biol Sci*. 2020;16(10):1678-85.

Cómo citar este artículo:

Mulato-Briones IB, Ribas-Aparicio RM, Reyes-Gastellou A, Rodríguez-Ildelfonso IO, Salcedo Vargas M. Análisis de genómica comparativa: del virus SARS al SARS CoV 2 Similitudes y diferencias. *Rev Med Inst Mex Seguro Soc*. 2020;58 Supl 2:S121-133.