

Basic Coronavirus biology and vaccines for COVID-19

Biología básica del Coronavirus y vacunas para COVID-19

Hernan Garcia-Ruiz^{1,2*}, Katherine LaTourrette^{1,2,3}, Mayra Teresa Garcia-Ruiz⁴, ¹Nebraska Center for Virology, ²Department of Plant Pathology, ³Complex Biosystems Interdisciplinary Life Sciences Program, University of Nebraska-Lincoln, Lincoln, NE, USA, 68503. ⁴Universidad Autónoma Chapingo, México, CP 56230. *Corresponding author: hgarciarui2@unl.edu

Received: December 21, 2020.

Accepted: March 01, 2021.

Garcia-Ruiz H, LaTourrette K and Garcia-Ruiz MT. 2021. Basic Coronavirus biology and vaccines for COVID-19. Mexican Journal of Phytopathology 39(4): 69-87.

DOI: <https://doi.org/10.18781/R.MEX.FIT.2021-1>

Abstract. *Severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) is the causal agent of the COVID-19 pandemic. Two mRNA vaccines based on the spike protein S have been authorized by the Food and Drug Administration. Antibody-based diagnostic test detect antibodies developed against protein S. Mutations in the genome of SARS-CoV-2 might compromise the precision of diagnostic tests and the efficacy of vaccines and antiviral drugs. We recently profiled genomic variation in human coronaviruses SARS-CoV, SARS-CoV-2, and *Middle East respiratory syndrome coronavirus* (MERS-CoV). As in all species of the genus *Betacoronavirus*, the genome is hyper variable, and mutations are not random. The most variable cistron codes for the spike S protein. Hyper variation in protein S has the potential to affect the efficacy of vaccines, the reliability

Resumen. El *coronavirus de tipo 2 causante del síndrome respiratorio agudo severo* (SARS-CoV-2) es el agente causal de la pandemia de COVID-19. Dos vacunas de ARNm basadas en la proteína espicular S han sido autorizadas por la Administración de Alimentos y Fármacos de Estados Unidos (FDA, por sus siglas en inglés). La prueba de diagnóstico basada en anticuerpos detecta anticuerpos desarrollados contra la proteína S. Las mutaciones en el genoma de SARS-CoV-2 podrían poner en riesgo la precisión de las pruebas de diagnóstico y la eficacia de las vacunas y los fármacos antivirales. Recientemente, realizamos un perfil de la variación genómica en los coronavirus humanos SARS-CoV, SARS-CoV-2 y el *Coronavirus del síndrome respiratorio de Oriente Medio* (MERS-CoV). Al igual que en todas las especies del género *Betacoronavirus*, el genoma es hipervariable y las mutaciones no son aleatorias. El cistron más variable codifica la proteína espicular S. La hipervariación en la proteína S tiene el potencial de afectar la eficacia de las vacunas, la confiabilidad de una prueba de diagnóstico basada en anticuerpos y predice el potencial de infecciones recurrentes de SARS-CoV-2. En este trabajo revisamos lo básico de la biología y

of antibody-based diagnostic test, and predicts potential for repeated SARS-CoV-2 infections. Here we review the basics of coronavirus biology and genomic variation, and link them to diagnostic tests, vaccines, and antiviral drugs.

Key words: Antiviral, Coronavirus, COVID-19, MERS-CoV, mRNA vaccine, protein S, spike protein.

The virus

The causal agent of the COVID-19 pandemic is the *Severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) first described in Wuhan, China in December of 2019 (Lu *et al.*, 2020; Zhu *et al.*, 2020). Two other coronaviruses are highly pathogenic to humans. The *Severe acute respiratory syndrome coronavirus* (SARS-CoV) was described in China in 2002, and the *Middle East respiratory syndrome coronavirus* (MERS-CoV) was described in South Arabia in 2012 (Cui *et al.*, 2019). Both SARS-CoV and SARS-CoV-2 originated in bats, in China, and adapted to infect humans (Cui *et al.*, 2019; Cagliani *et al.*, 2020; Lu *et al.*, 2020).

Coronaviruses form spherical virions with a membrane envelope. The genome is single-stranded RNA (Cui *et al.*, 2019). As in all RNA viruses, in coronaviruses sources of genetic variation include nucleotide insertions, deletions, substitutions and include RNA recombination. These events occur naturally during RNA replication (Sanjuán and Domingo-Calap, 2016). Genetic variation and selection favor accumulation of mutations in parts of the genome responsible for critical processes, such as host adaptation, vector transmission, entry into the cell, and suppression of antiviral defense (Obenauer *et al.*, 2006; Nigam and Garcia-Ruiz, 2020).

variación genómica del coronavirus y los vinculamos a pruebas de diagnóstico, vacunas y fármacos antivirales.

Palabras clave: Antivirales, COVID-19, MERS-CoV, vacuna ARNm, proteína S, proteína espicular.

El virus

El agente causal de la pandemia de COVID-19 es el *coronavirus de tipo 2 causante del síndrome respiratorio agudo severo* (SARS-CoV-2, por sus siglas en inglés), descrito por primera vez en Wuhan, China en diciembre de 2019 (Lu *et al.*, 2020; Zhu *et al.*, 2020). Existen otros dos coronavirus que son altamente patogénicos para los humanos. El *coronavirus del síndrome respiratorio agudo severo* (SARS-CoV, por sus siglas en inglés) fue descrito en China en el año 2002 y el *Coronavirus del síndrome respiratorio de Oriente Medio* (MERS-CoV, por sus siglas en inglés) fue descrito en el sur de Arabia Saudita en el 2012 (Cui *et al.*, 2019). Tanto el SARS-CoV como el SARS-CoV-2 se originaron en murciélagos en China y se adaptaron para infectar a los humanos (Cui *et al.*, 2019; Cagliani *et al.*, 2020; Lu *et al.*, 2020).

Los coronavirus forman viriones esféricos y están cubiertos por una membrana. El genoma es un ARN monocatenario (Cui *et al.*, 2019). Al igual que en todos los virus de ARN, en los coronavirus, fuentes de variación genética incluyen inserciones, eliminaciones, sustituciones e incluyen la recombinación de ARN. Estos ocurren de manera natural durante la replicación del ARN (Sanjuán y Domingo-Calap, 2016). La variación y selección genética favorecen la acumulación de mutaciones en partes del genoma encargadas de procesos críticos, tales como la adaptación del hospedante, la transmisión de vectores, entrada a la célula y la supresión de la defensa antiviral (Obenauer *et al.*, 2006; Nigam y Garcia-Ruiz, 2020).

At the population level, genetic variation and selection drive the formation of new strains and species (Lauring and Andino, 2010). This model supports the emergence of SARS-CoV and SARS-CoV-2 in bats followed by adaptation to humans (Cui *et al.*, 2019; Cagliani *et al.*, 2020; Lu *et al.*, 2020). SARS-CoV never reached pandemic level. One of the differences is that SARS-CoV-2 is more readily transmissible than SARS-CoV. The genetic difference in transmissibility and pathogenicity maps to the spike protein S (Zhou *et al.*, 2020).

The spike protein decorates the coronavirus virion and mediates entry into the cell to initiate infection (Li, 2016; Wrapp *et al.*, 2020a). SARS-CoV-2 entry is mediated by the specific interaction between the spike protein S and cellular receptor angiotensin-converting enzyme 2 (ACE2) (Cai *et al.*, 2020). Infected people develop neutralizing antibodies against the entire protein S and non-neutralizing antibodies against fractions or a subunit of protein S (Brochot *et al.*, 2020; Cai *et al.*, 2020). Accordingly, antibodies against the S protein are used as markers in diagnostic assays (Zhu *et al.*, 2007; Li *et al.*, 2008; Brochot *et al.*, 2020). Other, coronavirus diagnostic protocols are based on the detection of viral RNA by RT-PCR, viral proteins, or antibodies developed against viral proteins (Brochot *et al.*, 2020; Phan, 2020b; Zhu *et al.*, 2020).

Vaccines against SARS-CoV-2 are being developed using multiple approaches, including attenuated or inactivated viruses, DNA, adenovirus-based and mRNA vaccines (Amanat and Krammer, 2020; Dearlove *et al.*, 2020). In the United States of America, two mRNA vaccines based on protein S have been authorized for use by the Food and Drug Administration. The end goal of these vaccines is to block virus entry into the cell by activating the formation of antibodies against protein S. We recently showed that the cistron coding for protein S is the most variable in the genome of SARS-CoV-2

A nivel poblacional, la variación y la selección genética promueven la formación de nuevas cepas y especies (Lauring y Andino, 2010). Este modelo apoya la emergencia del SARS-CoV y el SARS-CoV-2 en murciélagos, seguido de su adaptación en humanos (Cui *et al.*, 2019; Cagliani *et al.*, 2020; Lu *et al.*, 2020). El SARS-CoV nunca alcanzó un nivel de pandemia. Una de las diferencias es que SARS-CoV-2 es más fácilmente transmisible que el SARS-CoV. La diferencia genética en transmisibilidad y patogenicidad se encuentra en la proteína espicular S (Zhou *et al.*, 2020).

La proteína espicular adorna de virión del coronavirus y media la entrada a la célula para iniciar la infección (Li, 2016; Wrapp *et al.*, 2020a). La entrada del SARS-CoV-2 es mediada por la interacción específica entre la proteína espicular S y el receptor celular ACE2 (por sus siglas en inglés), o enzima convertidora de angiotensina (Cai *et al.*, 2020). Las personas infectadas desarrollan anticuerpos neutralizantes contra toda la proteína S y anticuerpos no neutralizantes contra fracciones o una subunidad de la proteína S (Brochot *et al.*, 2020; Cai *et al.*, 2020). Por consiguiente, los anticuerpos contra la proteína S son usados como marcadores en ensayos de diagnóstico (Zhu *et al.*, 2007; Li *et al.*, 2008; Brochot *et al.*, 2020). Otros protocolos de diagnóstico del coronavirus se basan en la detección de ARN viral por RT-PCR, proteínas virales o anticuerpos desarrollados contra las proteínas virales (Brochot *et al.*, 2020; Phan, 2020b; Zhu *et al.*, 2020).

Varias vacunas contra SARS-CoV-2 se encuentran en desarrollo, usando varios enfoques, incluyendo virus atenuados o desactivados, ADN, con base en adenovirus y vacunas de ARNm (Amanat y Krammer, 2020; Dearlove *et al.*, 2020). En Estados Unidos, dos vacunas de ARNm basadas en la proteína S han sido autorizadas para su uso por la Administración de Alimentos y Fármacos de Estados Unidos (FDA, por sus siglas en inglés). La finalidad de estas vacunas es bloquear la entrada del

and in all species in the genus *Betacoronavirus* (LaTourrette *et al.*, 2021), which includes SARS-CoV and MERS-CoV. The wide genetic diversity of their host has selected *Betacoronavirus* for hyper variation in protein S (LaTourrette *et al.*, 2021).

The hyper variable nature of protein S has several biological functions. One is to maintain functionality and recognize a genetically diverse group of potential hosts, such as humans or bats (Zhai *et al.*, 2020). Another is to trigger the formation of non-neutralizing antibodies that serve as decoys. Protein S variation may also escape neutralizing antibodies formed by natural infection or triggered by vaccines (Long *et al.*, 2020; Walls *et al.*, 2020). Accordingly, efficacy of vaccines, and reliability of antibody-based diagnostic test has potential to be affected by variation in protein S. Protein S variation also explains the occurrence of repeated SARS-CoV-2 infections (Tillett *et al.*, 2020). Fundamental understanding of the coronavirus biology and genomic variation establish the basis for designing and deploying diagnostic tests, vaccines, and antiviral drugs.

Coronavirus taxonomy

Coronaviruses belong to the order *Nidovirales*, the family *Coronaviridae*, the sub-family *Orthcoronavirinae*, and four genera (*Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*) (Lu *et al.*, 2020; Zhu *et al.*, 2020) (Figure 1). Alphacoronaviruses infect mammals. Betacoronaviruses mainly infect bats and humans. Gammacoronaviruses and Deltacoronaviruses infect birds, and some species infect mammals. The genus *Betacoronavirus* is divided into five sub-genera (Figure 1): *Embecovirus*, *Merbecovirus*, *Nobecovirus*, *Hibecovirus*, and *Sarbecovirus* (Lu *et al.*, 2020; Zhu *et al.*, 2020). The sub-genus

virus a la célula al activar la formación de anticuerpos contra la proteína S. Hace poco demostramos que el cistron que codifica la proteína S es el más variable en el genoma del SARS-CoV-2 y en toda la especie del género *Betacoronavirus* (LaTourrette *et al.*, 2021), que incluye a SARS-CoV y a MERS-CoV. La amplia diversidad genética de su huésped ha seleccionado a *Betacoronavirus* para la hipervariación en la proteína S (LaTourrette *et al.*, 2021). La naturaleza hipervariable de la proteína S tiene varias funciones biológicas. Una es mantener la funcionalidad y reconocer a un grupo genéticamente diverso de hospedantes en potencia, tales como los humanos o los murciélagos (Zhai *et al.*, 2020). Otra es desencadenar la formación de anticuerpos no neutralizantes que sirven de señuelos. La variación de la proteína S también podrían escapar a anticuerpos neutralizantes formados por una infección natural o desencadenados por vacunas (Long *et al.*, 2020; Walls *et al.*, 2020). Por ello, la eficacia de las vacunas y la confiabilidad de la prueba de diagnóstico tiene el potencial de ser afectado por la variación en la proteína S. La variación en la proteína S también explica la ocurrencia de infecciones recurrentes de SARS-CoV-2 (Tillett *et al.*, 2020). Una comprensión de los fundamentos de la biología y la variación genómica del coronavirus sientan las bases para el diseño y la implementación de pruebas de diagnóstico, vacunas y fármacos antivirales.

Taxonomía del Coronavirus

Los coronavirus pertenecen al orden *Nidovirales*, la familia *Coronaviridae*, la subfamilia *Orthcoronavirinae* y cuatro géneros (*Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* y *Deltacoronavirus*) (Lu *et al.*, 2020; Zhu *et al.*, 2020) (Figura 1). Los alfacoronavirus infectan a los mamíferos. Los betacoronavirus infectan principalmente a

Family	Sub-Family	Genus	Sub-genus	Clade	Representative species	Host	Tropism	
Coronaviridae	Letovirinae	<i>Alphaletovirus</i>	<i>Milecovirus</i>		<i>Microhylla letovirus 1</i>	Amphibians	Not determined	
					<i>Alphacoronavirus 1</i>	Cat	Enteric	
	Orthocoronavirinae	<i>Betacoronavirus</i>				<i>Human coronavirus HKU1</i>	Human	Respiratory
						<i>Middle East respiratory syndrome-related coronavirus (MERS-CoV)</i>	Human, bat, camel	Respiratory
						<i>Rousettus bat coronavirus HKU9</i>	Bat	Not determined
						<i>Bat Hp-betacoronavirus</i>	Bat	Not determined
						<i>Sarbecovirus</i>		
			I	<i>Bat coronavirus</i>	Bat	Respiratory		
			II	<i>SARS-CoV-2</i> <i>Bat SARS coronavirus RaTG13</i>	Human Bat	Respiratory Respiratory		
			III	<i>SARS-CoV</i> <i>Bat SARS-like coronavirus</i>	Human Bat	Respiratory Respiratory		
			<i>Deltacoronavirus</i>	<i>Bulbul coronavirus HKU11</i>	Bird	Not determined		
			<i>Gammacoronavirus</i>	<i>Avian infectious bronchitis virus</i>	Commercial pheasants	Respiratory, enteric, and neurological		

Figure 1. Taxonomic organization of coronaviruses. Representative species are indicated for every genus and sub-genus. Their hosts and tropism are indicated. Important human pathogens are highlighted in blue.

Figura 1. Organización taxonómica de los coronavirus. Las especies representativas están señaladas para cada género y subgénero. Sus hospedantes y tropismo están señalados. Los patógenos humanos importantes aparecen resaltados en color azul.

Sarbecovirus contains species that infect only bats or humans and includes SARS-CoV and SARS-CoV-2. Another human coronavirus, MERS-CoV, belongs to the sub-genus *Merbecovirus*. The sub-genera *Nobecovirus*, *Hibecovirus* are integrated by species that infect bats (Figure 1).

Coronavirus genome organization

In betacoronaviruses the genome consists of a single RNA, linear, of positive polarity and is approximately 30,000 nt long. The virion is spherical, enveloped, and is 120 nm in diameter

murciélagos y humanos. Los gammacoronavirus y deltacoronavirus infectan a aves y algunas especies, a mamíferos. El género *Betacoronavirus* se divide en cinco subgéneros (Figura 1): *Embecovirus*, *Merbecovirus*, *Nobecovirus*, *Hibecovirus* y *Sarbecovirus* (Lu *et al.*, 2020; Zhu *et al.*, 2020). El subgénero *Sarbecovirus* contiene especies que infectan sólo a murciélagos y humanos e incluye a SARS-CoV y a SARS-CoV-2. Otro coronavirus humano, el MERS-CoV, pertenece al subgénero *Merbecovirus*. Los subgéneros *Nobecovirus* y *Hibecovirus* están integrados por especies que infectan a murciélagos (Figura 1).

(Brian and Baric, 2005; Cui *et al.*, 2019). The genomic RNA is protected by nucleoprotein N in a nucleocapsid. The envelope is formed by the membrane (M) protein and the small membrane protein E. A distinctive feature of the coronavirus virion is the presence of spikes formed by the glycoprotein S (protein S) (Figure 2) (Lan *et al.*, 2020; Walls *et al.*, 2020; Wrapp *et al.*, 2020a).

The coronavirus genomic RNA (Figure 3) is capped, polyadenylated and encodes multiple cistrons in open reading frames 1 (ORF1a) and 1b (ORF1b) joined by a ribosomal frameshift. Polyproteins 1a and 1ab are processed by papain-like proteinase NSP3 and 3C-like proteinase NSP5 to form the viral RNA-dependent RNA polymerase and several non-structural proteins necessary for RNA replication. M, E, S and other structural proteins are expressed from subgenomic RNAs co-terminal with the 3' end, and contain a 5' leader that is 65 to 89 nt long (Brian and Baric, 2005).

Entry into the cell

Entry into the cell is mediated by protein S spikes on the virion surface that interact with cellular receptors. The process is facilitated by entry cofactors (Gallagher and Buchmeier, 2001; Li, 2013; Cantuti-Castelvetri *et al.*, 2020; Yan *et al.*, 2020). Protein S is divided into S1 and S2 subunits cleaved by cellular proteases and cofactors (Millet and Whittaker, 2015; Cantuti-Castelvetri *et al.*, 2020; Coutard *et al.*, 2020; Xia *et al.*, 2020). The receptor binding domain is located at the tip of the S1 head, and mediates recognition and binding to the surface of the receptor ACE2 (Cai *et al.*, 2020). Interactions between protein S and de cellular receptor are critical for cell entry, and highly specific (Cai *et al.*, 2020). Thus, protein S is a determinant of coronavirus host range (Gallagher and Buchmeier, 2001; Zhai *et al.*, 2020).

Organización del genoma del Coronavirus

En los betacoronavirus, el genoma consiste en un único ARN, lineal, de polaridad positiva y con una longitud de aproximadamente 30,000 nt. El virión es esférico, está envuelto y tiene un diámetro de 120 nm (Brian y Baric, 2005; Cui *et al.*, 2019). El ARN genómico está protegido por la nucleoproteína N en una nucleocápside. La envoltura está formada por la proteína de membrana (M) y la proteína E de la membrana pequeña. Una característica distintiva del virión del coronavirus es la presencia de puntas formadas por la glucoproteína S (proteína S) (Figura 2) (Lan *et al.*, 2020; Walls *et al.*, 2020; Wrapp *et al.*, 2020a).

El ARN genómico del coronavirus (Figura 3) está protegido, poliandenilado y codifica múltiples cistrones en los marcos abiertos de lectura 1 (ORF1a) y 1b (ORF1b), unidos por un desplazamiento de marco ribosómico. Las poliproteínas 1a y 1ab son procesadas por proteinasa similar al NSP3 y la proteinasa NSP5 similar al 3C para formar la polimerasa ARN dependiente del ARN viral y varias proteínas no estructurales necesarias para la replicación del ARN. M, E, S y otras proteínas estructurales se expresan desde ARNs subgenómicos coterminales con la punta de 3' y contienen un líder de 5' con una longitud de entre 65 y 89 nt (Brian y Baric, 2005).

Entrada a la célula

La entrada a la célula está mediada por puntas de proteína S en la superficie del virión que interactúan con receptores celulares. El proceso es facilitado por cofactores de entrada (Gallagher y Buchmeier, 2001; Li, 2013; Cantuti-Castelvetri *et al.*, 2020; Yan *et al.*, 2020). La proteína S se divide en las subunidades S1 y S2, divididas por proteasas

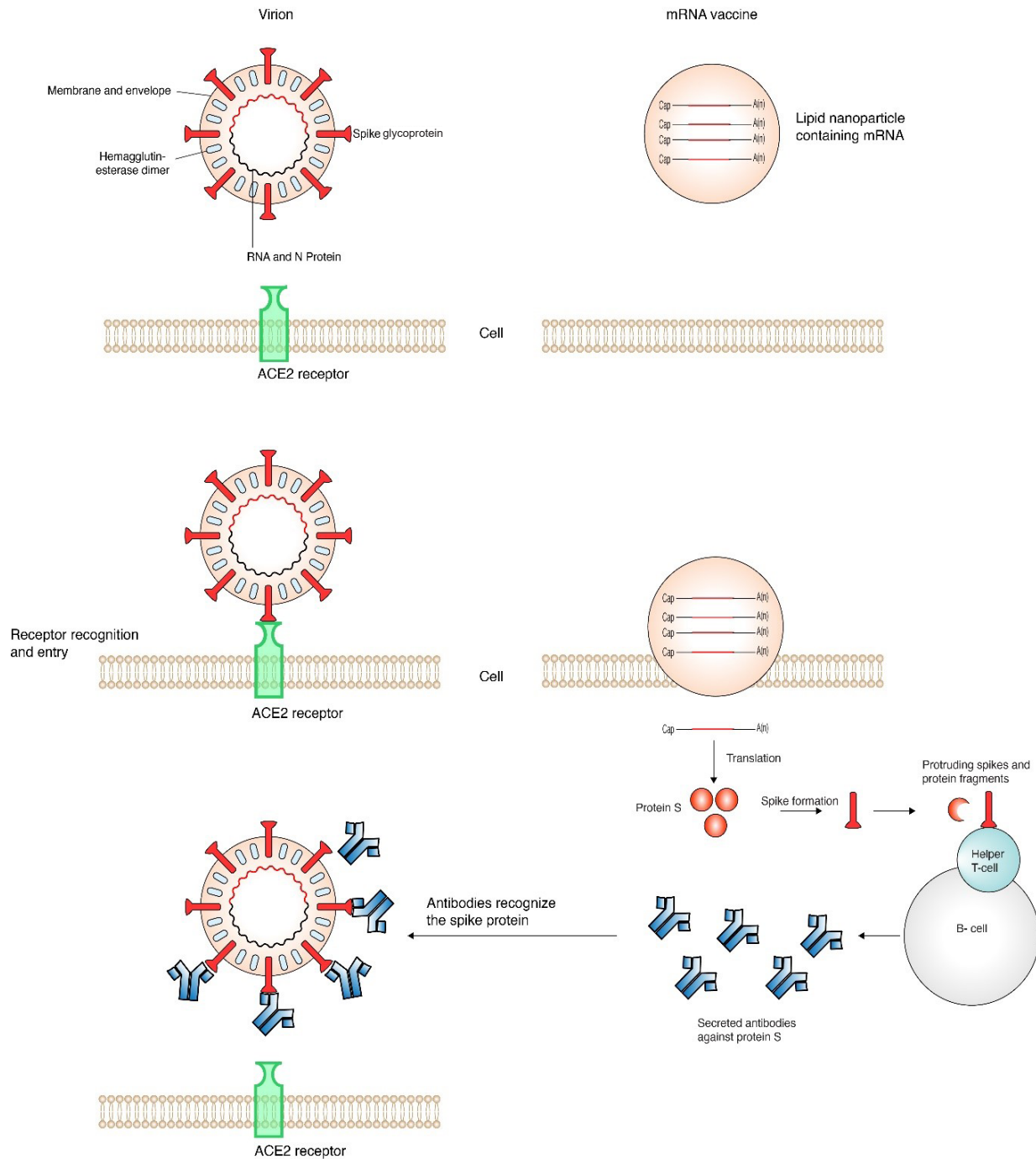


Figure 2. Schematic representation of coronavirus virions, entry into cells, and generation and activity of neutralizing antibodies against protein S triggered by natural infection or an mRNA vaccine.

Figura 2. Representación esquemática de viriones de coronavirus, la entrada a las células y generación y actividad de anticuerpos neutralizantes contra la proteína S desencadenada por infección natural o una vacuna de ARNm.

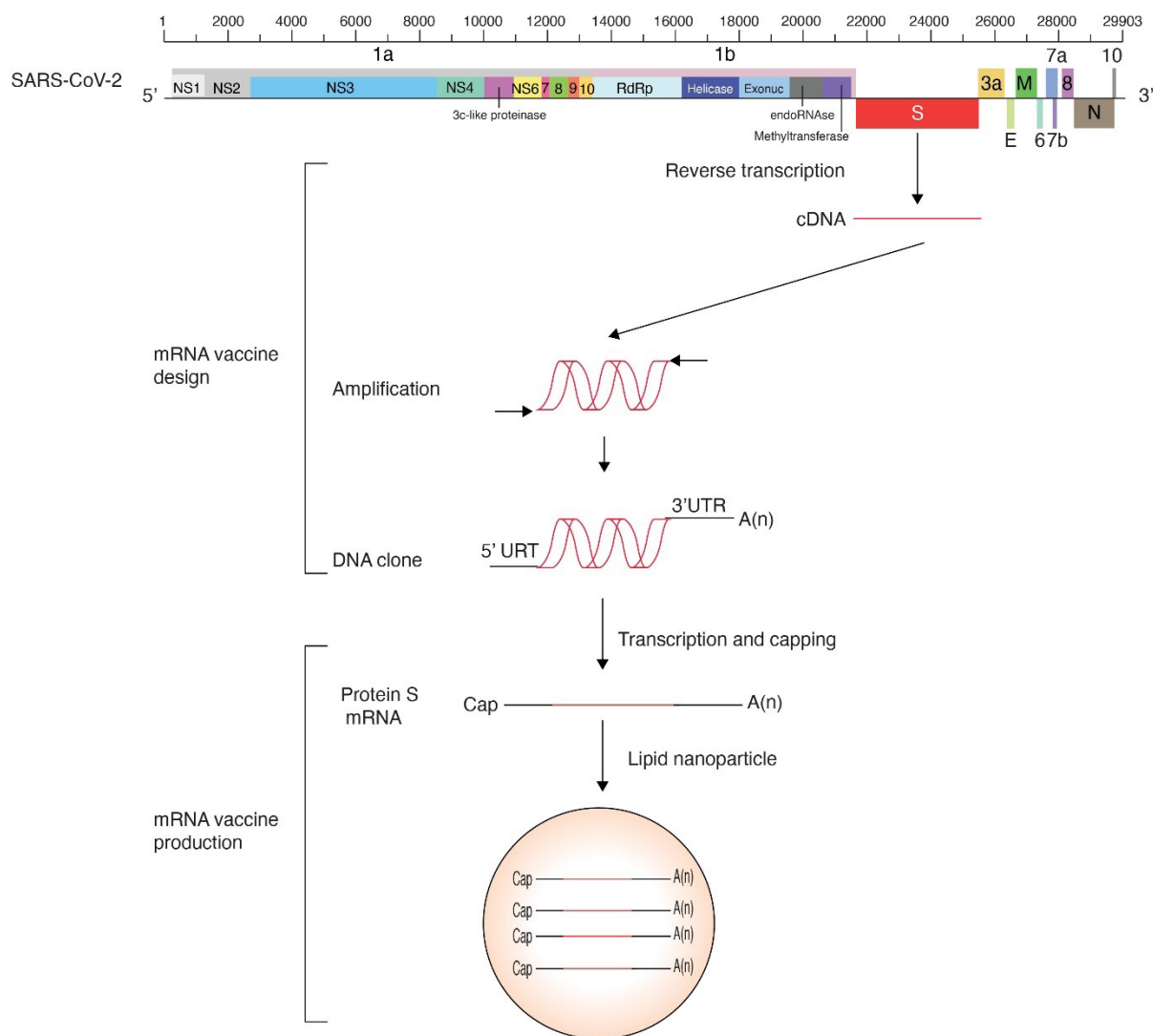


Figure 3. SARS-CoV-2 genome organization and design of a mRNA vaccine. Coordinates are based on the reference isolate Wuhan-Hu-1 (NC_045512.2).

Figura 3. Organización de genomas de SARS-CoV-2 y el diseño de una vacuna de ARNm. Las coordenadas se basan en el aislamiento de referencia Wuhan-Hu-1 (NC_045512.2).

The mRNA vaccine

For its critical role in cell entry, the spike S protein is the common target for neutralizing antibodies and vaccines (Brochot *et al.*, 2020; Cai *et al.*, 2020). In people infected with coronavirus,

celulares y cofactores (Millet y Whittaker, 2015; Cantuti-Castelvetri *et al.*, 2020; Coutard *et al.*, 2020; Xia *et al.*, 2020). El dominio de unión del receptor está ubicado en la punta del cabezal S1 y media en el reconocimiento y unión con la superficie del receptor ACE2 (Cai *et al.*, 2020). Las

neutralizing antibodies are formed against the entire protein S. However, non-neutralizing antibodies are also developed against the S2 subunit (Brochot *et al.*, 2020; Cai *et al.*, 2020).

Vaccines trigger the formation of neutralizing antibodies against protein S, in the absence of infection. Two mRNA vaccines based on protein S have been authorized. Their design is similar and are based on the genome organization and gene expression of coronaviruses (Figure 3). The cistron coding for protein S was cloned using the sequence from the reference isolate Wuhan-Hu-1 (NC_045512.2). A 5' UTR, a 3' UTR, and a poly A tail were added to provide stability and enhance translation efficiency. To account for variation in protein S (Becerra-Flores and Cardozo, 2020) prevalent mutations were introduced, and to further enhance translation efficiency, nucleoside modification were introduced (Pardi *et al.*, 2018). The basic design was made by integrating all basic information previously accumulated from SARS-CoV and MERS (Corbett *et al.*, 2020). For delivery purposes, and to avoid degradation, the mRNA is enclosed in a lipid nanodrop that releases the mRNA into the cell. Ribosomes translate the mRNA into protein S that triggers the formation of neutralizing antibodies (Pardi *et al.*, 2018).

Genome variation

In SARS-CoV-2, mutations have been detected and are being tracked using on-line tools (Hadfield *et al.*, 2018; Fernandes *et al.*, 2020). We recently profiled the genomic variation in all species in the genus *Betacoronavirus* (LaTourrette *et al.*, 2021). Results showed betacoronaviruses are hyper variable (Figure 4). The most diversity was observed in *Rousettus bat coronavirus HKU9*, other species infecting bats, and MERS-CoV. In these species, more than 25% of the nucleotides in the

interacciones entre la proteína S y el receptor celular son fundamentales para la entrada a la célula y altamente específicas (Cai *et al.*, 2020). Por ello, la proteína S es un factor decisivo del rango del hospedante del coronavirus (Gallagher y Buchmeier, 2001; Zhai *et al.*, 2020).

La vacuna de ARNm

Por su papel fundamental en la entrada a la célula, la proteína S es el objetivo común de anticuerpos neutralizantes y vacunas (Brochot *et al.*, 2020; Cai *et al.*, 2020). En personas infectadas con el coronavirus, los anticuerpos neutralizantes son formados en contra de toda la proteína S. Sin embargo, también se desarrollan anticuerpos no neutralizantes en contra de la subunidad S2 (Brochot *et al.*, 2020; Cai *et al.*, 2020).

Las vacunas desencadenan la formación de anticuerpos neutralizantes contra la proteína S en la ausencia de una infección. Dos vacunas de ARNm basadas en la proteína S han sido autorizadas. Sus diseños son similares y están basados en la organización genómica y la expresión de los genes de los coronavirus (Figura 3). La codificación del cistron para la proteína S fue clonada usando la secuencia del aislamiento de referencia Wuhan-Hu-1 (NC_045512.2). Un UTR de 5', un UTR de 3' y una cola de poli A se adicionaron para brindar estabilidad y aumentar la eficiencia de traducción. Para dar cuenta de la variación en la proteína S (Becerra-Flores y Cardozo, 2020), se introdujeron mutaciones prevalentes, y para aumentar la eficiencia de traducción aún más, se introdujeron modificaciones de nucleósidos (Pardi *et al.*, 2018). El diseño básico se hizo mediante la integración de toda la información previamente acumulada acerca de SARS-CoV y MERS (Corbett *et al.*, 2020). Para su entrega, y para evitar su degradación, el ARNm está cubierta por una nanogota lipídica que descarga el ARNm dentro de la célula. Los ribosomas tradu-

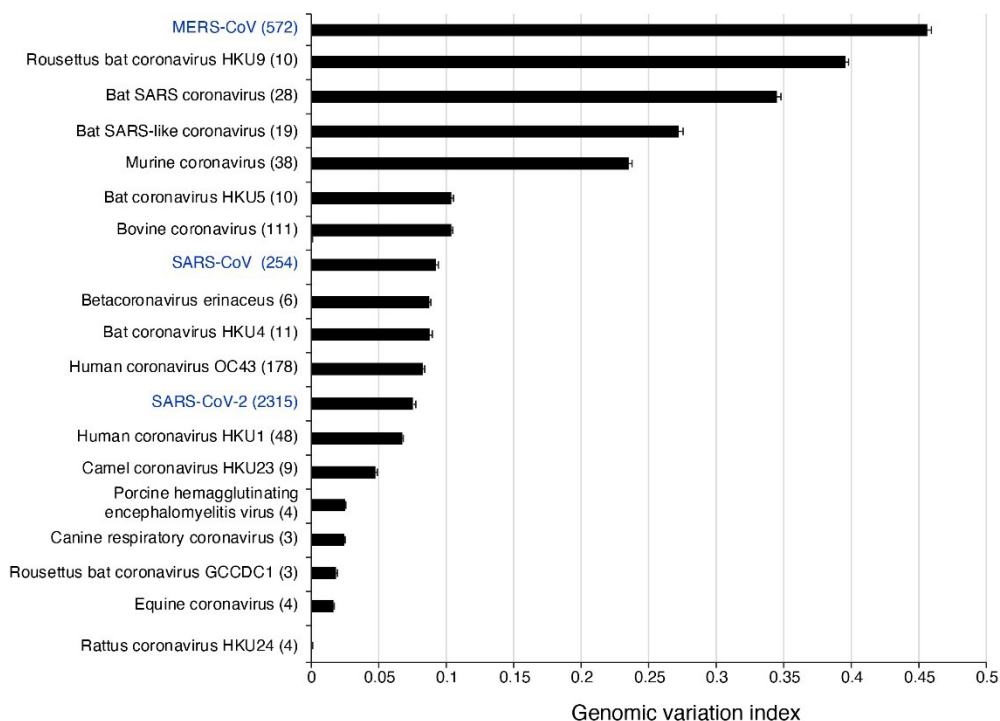


Figure 4. Genomic variation in betacoronaviruses. A genomic variation index indicates the proportion of single nucleotide polymorphisms normalized to the length of the genome. Modified from LaTourrette *et al.* (2021).

Figura 4. Variación genómica en los betacoronavirus. Un índice de variación genómica indica la proporción de polimorfismos de un solo nucleótido normalizados al largo del genoma. Modificado de LaTourrette *et al.* (2021).

genome are polymorphic (Figure 4). The genome of betacoronaviruses consists of 11 to 14 cistrons. The most variable cistron codes for the spike protein S. The lowest variation was detected the cistrons that code for proteins that mediate virus replication: RNA-dependent RNA polymerase, RNA helicase, exonuclease, endo RNase and methyltransferase, and that are located in open reading frame 1b (LaTourrette *et al.*, 2021).

Protein S variation

Mutations in the genome of SARS-CoV-2 have the potential to affect the precision of diagnostic tests and the efficacy of vaccines. In a recent genome-wide analysis, we showed that hyper variation in

cen el ARNm en una proteína S que desencadena la formación de anticuerpos neutralizantes (Pardi *et al.*, 2018).

Variación genómica

En el SARS-CoV-2 se han detectado mutaciones que están siendo rastreadas con el uso de herramientas en línea (Hadfield *et al.*, 2018; Fernandes *et al.*, 2020). Recientemente trazamos un perfil de la variación genómica en todas las especies del género *Betacoronavirus* (LaTourrette *et al.*, 2021). Los resultados muestran que los betacoronavirus son hipervariables (Figura 4). La mayor diversidad se observó en *Roussettus bat coronavirus* HKU9, otras especies que infectan a murciélagos,

protein S is a general feature of betacoronaviruses (LaTourrette *et al.*, 2021). Hyper variation in protein S is evident in betacoronavirus highly pathogenic to humans: MERS-CoV (Figure 5A), SARS-CoV (Figure 5B), and SARS-CoV-2 (Figure 5C). The pattern is also clear in species that infect bats (Figure 5B). Specifically, in SARS-CoV-2 several regions in protein S are hyper variable, including the ACE2 receptor binding domain and the fusion peptide proximal region (Figure 5D).

Betacoronaviruses mainly infect bats and humans (Figure 1). Given the large genetic diversity of bats, and possibly humans, the cellular receptors, proteases, and entry cofactors are likely diverse (Kuo *et al.*, 2000; Cantuti-Castelvetri *et al.*, 2020). Accordingly, protein S hyper variation may provide an evolutionary advantage. Mechanisms driving diversifying selection in protein S may include diversity in cellular receptors, cellular proteases that process the S1/S2 cleavage site, entry cellular cofactors, and antibodies.

Several domains in protein S are intrinsically disordered (LaTourrette *et al.*, 2021): the receptor binding domain and the C-terminal domain 2 in S1, and the fusion peptide proximal region in S2 (Figure 5D). This observation is important because intrinsically disordered proteins mediate functional diversity and interactions with genetically diverse partners such as cellular receptors and entry cofactor in bats and humans (Hebrard *et al.*, 2009; Rantalainen *et al.*, 2011; Charon *et al.*, 2018). Selection for hyper variation and disorder in protein S are consistent with the bat origin of SARS-CoV and SARS-CoV-2 followed by adaptation to humans (Cui *et al.*, 2019; Lu *et al.*, 2020).

In betacoronaviruses, protein S is hyper variable, disordered, mutationally robust (LaTourrette *et al.*, 2021), and is a determinant of host adaptation and host range (Kuo *et al.*, 2000; Muth *et al.*, 2018; Zhai *et al.*, 2020). The emerging model is that in protein S

y el MERS-CoV. En estas especies, más de 25% de los nucleótidos en el genoma son polimórficos (Figura 4). El genoma de los betacoronavirus consiste en entre 11 y 14 cistrones. El cistrón más variable codifica la proteína espicular S. La menor variación fue detectada en los cistrones que codifican las proteínas que median la replicación del virus: polimerasa de RNA dependiente de ARN, helicasa de ARN, exonucleasa, endo-RNasa y metiltransferasa, y que se encuentran en el marco de lectura abierto 1b (LaTourrette *et al.*, 2021).

Variación de la proteína S

Las mutaciones en el genoma de SARS-CoV-2 tienen el potencial de afectar la precisión de las pruebas de diagnóstico y la eficacia de las vacunas. Mediante un reciente análisis del genoma entero comprobamos que la hipervariación de la proteína S es una característica general de los betacoronavirus (LaTourrette *et al.*, 2021). La hipervariación en la proteína S es evidente en los betacoronavirus altamente patogénicos para los humanos: MERS-CoV (Figura 5A), SARS-CoV (Figura 5B) y SARS-CoV-2 (Figura 5C). El patrón también es claro en las especies que infectan a los murciélagos (Figura 5B). Específicamente en SARS-CoV-2, varias regiones en la proteína S son hipervariables, incluyendo el dominio de unión al receptor ACE2 y la región proximal del péptido de fusión (Figura 5D).

Los betacoronavirus infectan principalmente a murciélagos y humanos (Figura 1). Dada la gran diversidad genética de los murciélagos, y posiblemente de humanos, los receptores celulares, proteasas y cofactores de entrada posiblemente sean diversos (Kuo *et al.*, 2000; Cantuti-Castelvetri *et al.*, 2020). Por consiguiente, la hipervariación de la proteína S puede brindar una ventaja evolutiva. Los mecanismos que impulsan la selección diversificadora en la proteína S podrían incluir la diversidad en

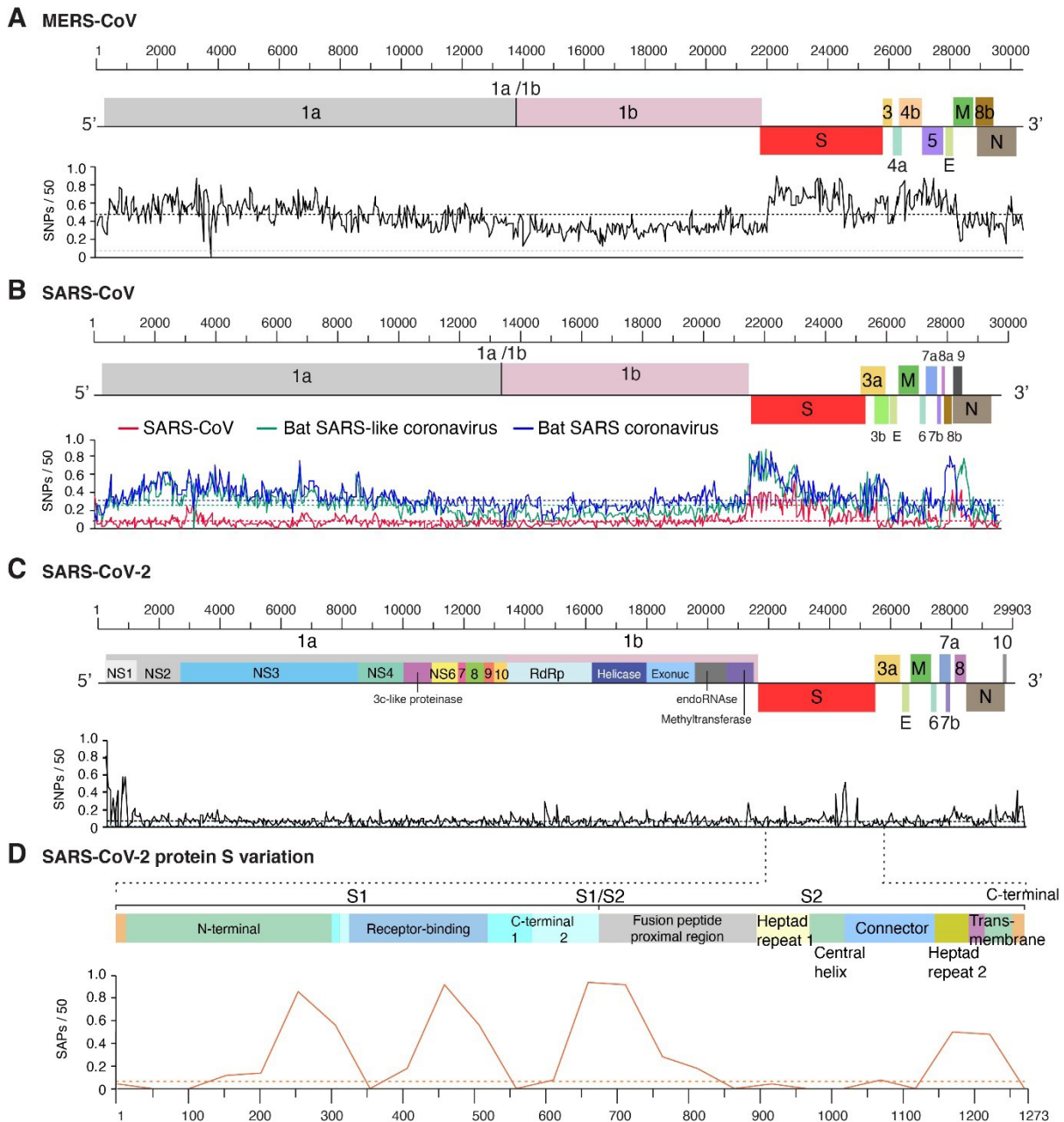


Figure 5. Distribution of genomic variation in human betacoronaviruses. Single nucleotide polymorphisms were plotted with respect to the genome. A 99% confidence interval is indicated as a horizontal line. A) MERS-CoV. B) SARS-CoV. Two species infecting bats are included for comparison. C) SARS-CoV-2. D) Single amino acid polymorphism in SARS-CoV-2 protein S. Domains are annotated and color coded. Modified from LaTourrette *et al.* (2021).

Figura 5. Distribución de variación genómica en betacoronavirus humanos. Polimorfismos de un solo nucleótido fueron graficados con respecto al genoma. Un intervalo de confianza de 99% está señalado como una línea horizontal. A) MERS-CoV. B) SARS-CoV. Dos especies que infectan a murciélagos están incluidas para comparar. C) SARS-CoV-2. D) Polimorfismo de un solo aminoácido en la proteína S de SARS-CoV-2. Los dominios están anotados y con códigos de color. Modificado de LaTourrette *et al.* (2021).

hyper variation provides an evolutionary advantage and is an intrinsic property of betacoronaviruses (LaTourrette *et al.*, 2021).

Antibodies against protein S

In infected cells, neutralizing antibodies are developed against protein S (Brochot *et al.*, 2020; Cai *et al.*, 2020). The receptor-binding domain is a critical antigen (Noy-Porat *et al.*, 2020). Additionally, non-neutralizing antibodies against protein S fragments of subunit two are also present (Brochot *et al.*, 2020; Cai *et al.*, 2020). Non-neutralizing antibodies may serve as decoys to reduce biogenesis and targeting efficiency of neutralizing antibodies (Cai *et al.*, 2020). Thus, the hyper variation of protein S might be a mechanism for betacoronaviruses to escape the immune system.

Variation in protein S and implications for vaccine use

Vaccines against SARS-CoV-2 induce neutralizing antibodies against protein S (Figure 2) (Cai *et al.*, 2020; Wrapp *et al.*; Yuan *et al.*, 2020). Hyper variation in protein S has potential to reduce efficacy of vaccines by multiple mechanisms. In an infected individual, new virus variants might be generated and have been detected (Jary *et al.*, 2020) with potential to escape neutralizing antibodies. Furthermore, antibodies developed after vaccination are a selection agents with potential to favor virus variants that can escape neutralizing antibodies (Baum *et al.*, 2020).

Under this scenario, if SARS-CoV-2 remains genetically stable, vaccines will be efficient (Dearlove *et al.*, 2020), antibody-based diagnostic test highly reliable, and infected people who develop antibodies will likely acquire immunity to SARS-CoV-2. However, if SARS-CoV-2 differentiates

receptores celulares, proteasas celulares que procesan el sitio de escisión S1/S2, cofactores de entrada celular y anticuerpos.

Varios dominios en la proteína S están intrínsecamente desordenados (LaTourrette *et al.*, 2021): el dominio de unión del receptor y el dominio 2 de la terminal C en S1, así como la región proximal del péptido de fusión en S2 (Figura 5D). Esta observación es importante, ya que las proteínas intrínsecamente desordenadas median en la diversidad funcional y las interacciones con parejas genéticamente diversas, tales como los receptores celulares y el cofactor de entrada en murciélagos y humanos (Hebrard *et al.*, 2009; Rantalainen *et al.*, 2011; Charon *et al.*, 2018). La selección para la hipervariación y el desorden en la proteína S son consistentes con el origen en el murciélago de SARS-CoV y SARS-CoV-2, seguido de su adaptación en los humanos (Cui *et al.*, 2019; Lu *et al.*, 2020).

En los betacoronavirus, la proteína S es hipervariable, desordenada, mutacionalmente robusta (LaTourrette *et al.*, 2021) y es determinante en la adaptación y el rango de los hospedantes (Kuo *et al.*, 2000; Muth *et al.*, 2018; Zhai *et al.*, 2020). Estas son características intrínsecas de todas las especies de betacoronavirus que les da una ventaja evolutiva (LaTourrette *et al.*, 2021).

Anticuerpos contra la proteína S

En células infectadas se desarrollan anticuerpos neutralizantes contra la proteína S (Brochot *et al.*, 2020; Cai *et al.*, 2020). El dominio de unión del receptor es un antígeno crítico (Noy-Porat *et al.*, 2020). Además, se encuentran presentes anticuerpos no neutralizantes contra fragmentos de proteína S de la subunidad dos (Brochot *et al.*, 2020; Cai *et al.*, 2020). Anticuerpos que no neutralizan proporcionan varias ventajas al virus. Reducen la síntesis de anticuerpos neutralizantes y mantienen ocupados

into strains, vaccines will be efficient only against closely related strains, ineffective against diverse strains, and people might be repeatedly infected by SARS-CoV-2.

Re-infection in humans has been confirmed (Tillett *et al.*, 2020), and hyper variation in protein S is a general feature of betacoronaviruses (LaTourrette *et al.*, 2021). These observations predict that adjustments to vaccine design and antibody-based diagnostic tests will be needed. Vaccines administered to people may consist of a cocktail of protein S variants (Baum *et al.*, 2020; Cai *et al.*, 2020). Alternatively, or in addition, the vaccines may need to be re-designed based on SARS-CoV-2 population dynamics, structure and their geographic distribution (Korber *et al.*, 2020; Taboada *et al.*, 2020; LaTourrette *et al.*, 2021).

It is likely that SARS-CoV-2 will accumulate mutations for efficient replication and differentiate into biological strains as the virus faces selection pressure from genetically different human populations (LaTourrette *et al.*, 2021). Multiple lines of evidence support this model. Despite not reaching pandemic levels, protein S accumulated large numbers of mutations in MERS-CoV and SARS-CoV (Figure 5). SARS-CoV-2 variants infecting the same individual have been detected (Tillett *et al.*, 2020; Jary *et al.*, 2020), and recurrent mutations in open reading frame 1ab, and cistrons coding for NSP6 and protein S have been identified (van Dorp *et al.*, 2020). Additionally, contrasting mutations have also be described. In Mexico, SARS-CoV-2 population were grouped into clades (Taboada *et al.*, 2020), and in Arizona, a 27-amino acid deletion was detected in protein 7 (Holland *et al.*, 2020).

The first genome of the SARS-CoV-2 came from the strain initially described in China, (Wuhan-Hu-1 NC_045512.2). The cistron coding for protein S contains residues that are compatible, but not

al sistema inmunológico (Cai *et al.*, 2020). Por ello, la hipervariación de la proteína S podría ser un mecanismo para que los betacoronavirus escapen del sistema inmune.

Variación en la proteína S e implicaciones para el uso en vacunas

Las vacunas contra el SARS-CoV-2 inducen anticuerpos neutralizantes contra la proteína S (Figura 2) (Cai *et al.*, 2020; Wrapp *et al.*; Yuan *et al.*, 2020). La hipervariación en la proteína S tiene el potencial de reducir la eficacia de las vacunas mediante varios mecanismos. En un individuo infectado se podrían generar, y se han detectado, varias variantes del virus (Jary *et al.*, 2020) con el potencial de escapar a los anticuerpos neutralizantes. Además, los anticuerpos desarrollados después de recibir la vacuna son agentes de selección con el potencial de favorecer a variantes del virus que pueden escapar a anticuerpos neutralizantes (Baum *et al.*, 2020).

En este escenario, si el SARS-CoV-2 permanece genéticamente estable, las vacunas serán eficientes (Dearlove *et al.*, 2020), las pruebas de diagnósticos basadas en anticuerpos serán muy confiables y la gente infectada que desarrollen anticuerpos probablemente se vuelvan inmunes al SARS-CoV-2. Sin embargo, si el SARS-CoV-2 se separa en cepas, las vacunas serán eficientes sólo contra cepas estrechamente relacionadas e ineficiente contra cepas diversificadas, con lo que la gente se infectará con SARS-CoV-2 de manera recurrente.

La reinfección en los humanos ha sido confirmada (Tillett *et al.*, 2020) y la hipervariación en la proteína S es una característica general de los betacoronavirus (LaTourrette *et al.*, 2021). Estas observaciones predicen que se requerirán ajustes al diseño de la vacuna y pruebas de diagnóstico con base en anticuerpos. Las vacunas administradas a las personas podrían consistir en un cóctel

optimal, for binding human receptor ACE2 (Wan *et al.*, 2020). Thus, there is potential for protein S to accumulate mutations for more efficient entry into human cells. Consistent with this model, a D614G mutation makes the virus more transmissible, more pathogenic to humans (Becerra-Flores and Cardozo, 2020), and has replaced the initial strain (Long *et al.*, 2020; Volz *et al.*, 2020). The D614G mutation and others in the receptor binding domain reduce affinity to monoclonal antibody CR3022 (Long *et al.*, 2020). This is consistent with a role for protein S variation in escaping from neutralizing antibodies.

Future challenges

Multiple lines of evidence support the model that SARS-CoV-2 is mutating (Forster *et al.*, 2020; Korber *et al.*, 2020; Phan, 2020a), and that as a group betacoronaviruses are hypervariable and variation mainly accumulates in protein S (LaTourrette *et al.*, 2021). This variation has the potential to affect the both the efficacy of vaccines and the reliability of antibody-based diagnostic test. Collectively this information predicts that vaccine design and deployment will be based on fundamental understanding and characterization of proteins S, and other genes, in the SARS-CoV-2 genome in a combination of factors such as human population genetics, age groups, health underlying conditions, geographical and regional boundaries. Characterizing the genetic structure of SARS-CoV-2 at fine scale, and translating this variation into the design and deployment of SARS-CoV-2 vaccines is one of the main challenges. To answer this challenge, it will be essential profile the genetic structure of the virus in different parts of the world, in human populations of different genetic backgrounds, and before and after administration of the SARS-CoV-2 vaccines

de variantes de la proteína S (Baum *et al.*, 2020; Cai *et al.*, 2020). Además, o como alternativa, las vacunas quizá necesiten ser rediseñadas con base en las dinámicas de la población del SARS-CoV-2, su estructura y distribución geográfica (Korber *et al.*, 2020; Taboada *et al.*, 2020; LaTourrette *et al.*, 2021).

Es posible que el SARS-CoV-2 acumule mutaciones para su replicación eficiente y se diferencie en cepas biológicas a medida que el virus enfrenta presión de selección de poblaciones humanas genéticamente diferentes (LaTourrette *et al.*, 2021). Son muchas las líneas de evidencia que apoyan este modelo. A pesar de no haber alcanzado niveles pandémicos, la proteína S acumuló grandes cantidades de mutaciones en MERS-CoV y SARS-CoV (Figura 5). Se han detectado variantes del SARS-CoV-2 que han infectado al mismo individuo (Tillett *et al.*, 2020; Jary *et al.*, 2020) y se han identificado mutaciones recurrentes en el marco abierto de lectura 1ab y cistrones que codifican NSP6 y proteína S (van Dorp *et al.*, 2020). Además, también se han descrito mutaciones contrastantes. En México, la población de SARS-CoV-2 se agrupó en clados (Taboada *et al.*, 2020) y en Arizona se detectó la eliminación de 27 aminoácidos en la proteína 7 (Holland *et al.*, 2020).

El primer genoma del SARS-CoV-2 surgió de la cepa inicialmente descrita en China, (Wuhan-Hu-1 NC_045512.2). El cistron que codifica la proteína S contiene residuos que son compatibles, aunque no óptimos, para el receptor humano de unión ACE2 (Wan *et al.*, 2020). Por ello, la proteína S tiene el potencial de acumular mutaciones para una entrada más eficiente a las células humanas. Consistente con este modelo, una mutación de D614G hace al virus más transmisible, más patogénico para humanos (Becerra-Flores y Cardozo, 2020) y ha reemplazado a la cepa inicial (Long *et al.*, 2020;

ACKNOWLEDGMENTS

This research was supported by NIH grant R01GM120108 to HG-R and by the Nebraska Agricultural Experiment Station with funding from the Hatch Act (Accession Number 1007272) through the USDA National Institute of Food and Agriculture.

LITERATURE CITED

- Amanat F and Krammer F. 2020. SARS-CoV-2 Vaccines: Status Report. *Immunity* 52(4): 583-589. <https://doi.org/10.1016/j.immuni.2020.03.007>
- Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, Wei Y, Atwal GS, Murphy AJ, Stahl N, Yancopoulos GD and Kyratsous CA. 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 369(6506): 1014-1018. <https://doi.org/10.1126/science.abd0831>
- Becerra-Flores M and Cardozo T. 2020. SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice* 74(8): e13525. <https://doi.org/10.1111/ijcp.13525>
- Brian DA and Baric RS. 2005. Coronavirus genome structure and replication. *Current Topics and Microbiology Immunology* 287: 1-30. https://doi.org/10.1007/3-540-26765-4_1
- Brochot E, Demey B, Touzé A, Belouzard S, Dubuisson J, Schmit J-L, Duverlie G, Francois C, Castelain S and Helle F. 2020. Anti-spike, Anti-nucleocapsid and Neutralizing Antibodies in SARS-CoV-2 Inpatients and Asymptomatic carriers. *Frontiers in Microbiology*. 24 p. <https://doi.org/10.3389/fmicb.2020.584251>
- Cagliani R, Forni D, Clerici M and Sironi M. 2020. Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Virology* 94: e00411-00420. <https://doi.org/10.1128/JVI.00411-20>
- Cai Y, Zhang J, Xiao T, Peng H, Sterling SM, Walsh RM, Rawson S, Rits-Volloc, S and Chen B. 2020. Distinct conformational states of SARS-CoV-2 spike protein. *Science* 369(6511) 1586-1592. <https://doi.org/10.1126/science.abd4251>
- Cantuti-Castelvetri L, Ojha R, Pedro LD, Djannatian M, Franz J, Kuivanen S, van der Meer F, Kallio K, Kaya T, Anastasina M, Smura T, Levanov L, Szivoczka L, Tobi A, Kallio-Kokko H, Österlund P, Joensuu M, Meunier FA, Butcher SJ, Winkler MS, Mollenhauer B, Helenius A, Gokce O, Teesalu T, Hepojoki J, Vapalahti O, Stadelmann C, Balistreri G and Simons M. 2020. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370(6518): 856-860. <https://doi.org/10.1126/science.abd2985>
- Charon J, Barra A, Walter J, Millot P, Hebrard E, Moury B and Michon T. 2018. First Experimental Assessment of Protein

Volz *et al.*, 2020). La mutación de D614G y otros en el dominio de unión del receptor reducen la afinidad al anticuerpo monoclonal CR3022 (Long *et al.*, 2020). Esto es consistente con un rol de la variación de la proteína S para escapar de anticuerpos neutralizantes.

Retos a futuro

El SARS-CoV-2 está acumulando mutaciones y cambiando (Forster *et al.*, 2020; Korber *et al.*, 2020; Phan, 2020a) y que, como grupo, los betacoronavirus son hipervariables y la variación se acumula principalmente en la proteína S (LaTourrette *et al.*, 2021). Esta variación tiene el potencial de afectar tanto la eficacia de las vacunas como la confiabilidad de la prueba de diagnóstico basada en anticuerpos. De forma colectiva, esta información predice que el diseño y el uso de las vacunas se basará en la comprensión fundamental y la caracterización de las proteínas S y otros genes en el genoma del SARS-CoV-2 en una combinación de factores tales como la genética de las poblaciones humanas, grupos de edades, condiciones de salud subyacentes y fronteras geográficas y regionales. Caracterizar la estructura genética de SARS-CoV-2 a una escala fina y traducir esta variación al diseño y la implementación de las vacunas contra SARS-CoV-2 es uno de los desafíos principales. Para responder a este desafío será esencial realizar un perfil de la estructura genética de este virus en diferentes partes del mundo, en poblaciones humanas de diversos orígenes genéticos y antes y después de administrar las vacunas contra el SARS-CoV-2.

AGRADECIMIENTOS

Esta investigación contó con el apoyo de la beca NIH R01GM120108 a HG-R y por la Estación Experimental Agrícola de Nebraska con la financiación del Acta de Hatch (Número de Acceso 1007272) a través del Instituto Nacional de Alimentos y Agricultura del USDA.

- Intrinsic Disorder Involvement in an RNA Virus Natural Adaptive Process. *Molecular Biology and Evolution* 35(1): 38-49. <https://doi.org/10.1093/molbev/msx249>.
- Corbett KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, Himansu S, Schäfer A, Ziwawo CT, DiPiazza AT, Dinnon KH, Elbashir SM, Shaw CA, Woods A, Fritch EJ, Martinez DR, Bock KW, Minai M, Nagata BM, Hutchinson GB, Wu K, Henry C, Bahl K, Garcia-Dominguez D, Ma L, Renzi I, Kong W-P, Schmidt SD, Wang L, Zhang Y, Phung E, Chang LA, Loomis RJ, Altaras NE, Narayanan E, Metkar M, Presnyak V, Liu C, Louder MK, Shi W, Leung K, Yang ES, West A, Gully KL, Stevens LJ, Wang N, Wrapp D, Doria-Rose NA, Stewart-Jones G, Bennett H, Alvarado GS, Nason MC, Ruckwardt TJ, McLellan JS, Denison MR, Chappell JD, Moore IN, Morabito KM, Mascola JR, Baric RS, Carfi A and Graham BS. 2020. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* 586: 567-571. <https://doi.org/10.1038/s41586-020-2622-0>
- Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG and 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* 176: 104742. <https://doi.org/10.1016/j.antiviral.2020.104742>
- Cui J, Li F and Shi ZL. 2019. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology* 17: 181-192. <https://www.nature.com/articles/s41579-018-0118-9>
- Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, Scott PT, Amare MF, Vasan S, Michael NL, Modjarrad K and Rolland M. 2020. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proceedings of the National Academy of Sciences* 117(38): 23652-23662. <https://doi.org/10.1073/pnas.2008281117>
- Fernandes JD, Hinrichs AS, Clawson H, Gonzalez JN, Lee BT, Nassar LR, Raney BJ, Rosenbloom KR, Nerli S, Rao AA, Schmelter D, Fyfe A, Maulding N, Zweig AS, Lowe TM, Ares M, Corbet-Detig R, Kent WJ, Haussler D and Haussler M. 2020. The UCSC SARS-CoV-2 Genome Browser. *Nature Genetics* 52: 991-998. <https://www.nature.com/articles/s41588-020-0700-8>
- Forster P, Forster L, Renfrew C and Forster M. 2020. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences* 117(17): 9241-9243. <https://doi.org/10.1073/pnas.2004999117>
- Gallagher TM and Buchmeier MJ. 2001. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* 279(2): 371-374. <https://doi.org/10.1006/viro.2000.0757>
- Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T and Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34(23): 4121-4123. <https://doi.org/10.1093/bioinformatics/bty407>
- Hebrard E, Bessin Y, Michon T, Longhi S, Uversky VN, Delalande F, Van Dorsselaer A, Romero P, Walter J, Declerck N and Fargette D. 2009. Intrinsic disorder in Viral Proteins Genome-Linked: experimental and predictive analyses. *Virology Journal* 6: 23. <https://virologyj.biomedcentral.com/articles/10.1186/1743-422X-6-23>
- Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, Halden RU, Hogue BG, Scotch M, and Lim ES. 2020. An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020). *Journal Virology* 94(14): JVI.00711-00720. <https://doi.org/10.1128/JVI.00711-20>
- Jary A, Leducq V, Malet I, Marot S, Klement-Frutos E, Teysou E, Soulié C, Abdi B, Wirden M, Pourcher V, Caumes E, Calvez V, Burrel S, Marcelin A-G and Boutolleau D. 2020. Evolution of viral quasispecies during SARS-CoV-2 infection. *Clinical Microbiology and Infection* 26(11):1560.e1-1560.e4. <https://doi.org/10.1016/j.cmi.2020.07.032>
- Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, Angyal A, Brown RL, Carrilero L, Green LR, Groves DC, Johnson KJ, Keeley AJ, Lindsey BB, Parsons PJ, Raza M, Rowland-Jones S, Smith N, Tucker RM, Wang D, Wyles MD, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO and Montefiori DC. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182(4): 812-827. e819. <https://doi.org/10.1016/j.cell.2020.06.043>
- Kuo L, Godeke GJ, Raamsman MJ, Masters PS and Rottier PJ. 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *Journal Virology* 74(3): 1393-1406. <https://doi.org/10.1128/jvi.74.3.1393-1406.2000>
- Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L and Wang X. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581: 215-220. <https://www.nature.com/articles/s41586-020-2180-5>
- LaTourrette K, Holste NM, Rodriguez-Peña R, Arruda-Leme R and García-Ruiz H. 2021. Genome-wide variation in betacoronaviruses. *Journal Virology* Submitted.
- Lauring AS and Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 6: e1001005. <https://doi.org/10.1371/journal.ppat.1001005>
- Li CK-f, Wu H, Yan H, Ma S, Wang L, Zhang M, Tang X, Temperton NJ, Weiss RA, Brenchley JM, Douek DC, Mongkolsapaya J, Tran B-H, Lin C-IS, Screaton GR, Hou J-I, McMichael AJ and Xu X-N. 2008. T cell responses to whole SARS coronavirus in humans. *The Journal of Immunology* 181(8): 5490-5500. <https://doi.org/10.4049/jimmunol.181.8.5490>
- Li F. 2013. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral Research* 100(1): 246-254. <https://doi.org/10.1016/j.antiviral.2013.08.014>
- Li F. 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology* 3: 237-261. <https://doi.org/10.1146/annurev-virology-110615-042301>
- Long SW, Olsen RJ, Christensen PA, Bernard DW, Davis JJ, Shukla M, Nguyen M, Saavedra MO, Yerramilli P, Pruitt L, Subedi S, Kuo H-C, Hendrickson H, Eskandari G, Nguyen HAT, Long JH, Kumaraswami M, Goike J, Boutz D, Gollihar J, McLellan JS, Chou C-W, Javanmardi K,

- Finkelstein IJ and Musser J. 2020. Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. *mBio* 2020.2009.2022.20199125. <https://doi.org/10.1128/mBio.02707-20>
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W and Tan W. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* 395(10224): 565-574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Millet JK, and Whittaker GR. 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Research* 202: 120-134. <https://doi.org/10.1016/j.virusres.2014.11.021>
- Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, Gloza-Rausch F, Balboni A, Battilani M, Rihtaric D, Toplak I, Ameneiros RS, Pfeifer A, Thiel V, Drexler JF, Muller MA and Drosten C. 2018. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Scientific Report* 8: 15177. <https://doi.org/10.1038/s41598-018-33487-8>
- Nigam D and Garcia-Ruiz H. 2020. Variation Profile of the Orthospovirus Genome. *Pathogens* 9(7): 521. <https://doi.org/10.3390/pathogens9070521>
- Noy-Porat T, Makdasi E, Alcalay R, Mechaly A, Levy Y, Bercovich-Kinori A, Zauberman A, Tamir H, Yahalom-Ronen Y, Israeli Ma, Epstein E, Achdout H, Melamed S, Chitlaru T, Weiss S, Peretz E, Rosen O, Paran N, Yitzhaki S, Shapira SC, Israely T, Mazor O and Rosenfeld R. 2020. A panel of human neutralizing mAbs targeting SARS-CoV-2 spike at multiple epitopes. *Nature Communications* 11: 4303. <https://www.nature.com/articles/s41467-020-18159-4>
- Obenauer JC, Denson J, Mehta PK, Su X, Mukatira S, Finkelstein DB, Xu X, Wang J, Ma J, Fan Y, Rakestraw KM, Webster RG, Hoffmann E, Krauss S, Zheng J, Zhang Z and Naeve CW. 2006. Large-Scale Sequence Analysis of Avian Influenza Isolates. *Science* 311(5767): 1576-1580. <https://doi.org/10.1126/science.1121586>
- Pardi N, Hogan MJ, Porter FW and Weissman D. 2018. mRNA vaccines — a new era in vaccinology. *Nature Reviews Drug Discovery* 17: 261-279. <https://www.nature.com/articles/nrd.2017.243>
- Phan T. 2020a. Genetic diversity and evolution of SARS-CoV-2. *Infection, Genetics and Evolution* 81: 104260. <https://doi.org/10.1016/j.meegid.2020.104260>
- Phan T. 2020b. Novel coronavirus: From discovery to clinical diagnostics. *Infection, Genetics and Evolution* 79: 104211. <https://doi.org/10.1016/j.meegid.2020.104211>
- Rantalainen KI, Eskelin K, Tompa P, and Mäkinen K. 2011. Structural flexibility allows the functional diversity of potyvirus genome-linked protein VPg. *Journal of virology* 85: 2449-2457. <https://jvi.asm.org/content/85/5/2449>
- Sanjuán R and Domingo-Calap P. 2016. Mechanisms of viral mutation. *Cellular and Molecular Life Sciences* 73(23): 4433-4448. <https://doi.org/10.1007/s00018-016-2299-6>
- Taboada B, Vazquez-Perez JA, Muñoz-Medina JE, Ramos-Cervantes P, Escalera-Zamudio M, Boukadida C, Sanchez-Flores A, Isa P, Mendieta-Condado E, Martínez-Orozco JA, Becerril-Vargas E, Salas-Hernández J, Grande R, González-Torres C, Gaytán-Cervantes FJ, Vazquez G, Pulido F, Araiza-Rodríguez A, Garcés-Ayala F, González-Bonilla CR, Grajales-Muñoz C, Borja-Aburto VH, Barrera-Badillo G, López S, Hernández-Rivas L, Perez-Padilla R, López-Martínez I, Ávila-Ríos S, Ruiz-Palacios G, Ramírez-González JE and Arias CF. 2020. Genomic Analysis of Early SARS-CoV-2 Variants Introduced in Mexico. *Journal of Virology* 94(1): e01056-01020. <https://doi.org/10.1128/JVI.01056-20>
- Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, Laverdure C, Verma SC, Rossetto CC, Jackson D, Farrell MJ, Van Hooser S and Pandori M. 2020. Genomic evidence for reinfection with SARS-CoV-2: a case study. *The Lancet Infectious Diseases* 21(1): 52-58. [https://doi.org/10.1016/S1473-3099\(20\)30764-7](https://doi.org/10.1016/S1473-3099(20)30764-7)
- van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, Owen CJ, Pang J, Tan CCS, Boshier FAT, Ortiz AT and Balloux F. 2020. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution* 83: 104351. <https://doi.org/10.1016/j.meegid.2020.104351>
- Volz EM, Hill V, McCrone JT, Price A, Jorgensen D, Toole A, Southgate JA, Johnson R, Jackson B, Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, da Silva Filipe A, Shepherd JG, Pascall DJ, Shah R, Jesudason N, Li K, Jarrett R, Pacchiarini N, Bull M, Geidelberg L, Siveroni I, Goodfellow IG, Loman NJ, Pybus O, Robertson DL, Thomson EC, Rambaut A and Connor TR. 2020. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity 184(1): 64-75. <https://doi.org/10.1016/j.cell.2020.11.020>
- Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT and Veerler D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181(2): 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>
- Wan Y, Shang J, Graham R, Baric RS and Li F. 2020. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *Journal of Virology* 94(7): e00127-00120. <https://doi.org/10.1128/JVI.00127-20>
- Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS and McLellan JS. 2020a. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367(6483): 1260-1263. <https://doi.org/10.1126/science.abb2507>
- Wrapp D, De Vlioger D, Corbett KS, Torres GM, Wang N, Van Breedam W, Roose K, van Schie L, Team V-CC-R, Hoffmann M, Pohlmann S, Graham BS, Callewaert N, Schepens B, Saelens X and McLellan JS. 2020b. Structural Basis for Potent Neutralization of Betacoronaviruses by Single-Domain Camelid Antibodies. *Cell* 181(5): 1004-1015.e15. <https://doi.org/10.1016/j.cell.2020.04.031>
- Xia S, Liu M, Wang C, Xu W, Lan Q, Feng S, Qi F, Bao L, Du L, Liu S, Qin C, Sun F, Shi Z, Zhu Y, Jiang S and Lu L. 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high

- capacity to mediate membrane fusion. *Cell Research* 30: 343-355. <https://www.nature.com/articles/s41422-020-0305-x>
- Yan R, Zhang Y, Li Y, Xia L, Guo Y and Zhou Q. 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 367(6485): 1444-1448. <https://doi.org/10.1126/science.abb2762>
- Yuan M, Wu NC, Zhu X, Lee C-CD, So RTY, Lv H, Mok CKP, and Wilson IA. 2020. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* 368(6491): 630-633. <https://doi.org/10.1126/science.abb7269>
- Zhai X, Sun J, Yan Z, Zhang J, Zhao J, Zhao Z, Gao Q, He W-T, Veit M and Su S. 2020. Comparison of Severe Acute Respiratory Syndrome Coronavirus 2 Spike Protein Binding to ACE2 Receptors from Human, Pets, Farm Animals, and Putative Intermediate Hosts. *Journal of Virology* 94: e00831-00820. <https://doi.org/10.1128/JVI.00831-20>
- Zhou H, Chen X, Hu T, Li J, Song H, Liu Y, Wang P, Liu D, Yang J, Holmes EC, Hughes AC, Bi Y and Shi W. 2020. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Current Biology* 30(11): 2196-2203 e2193. <https://doi.org/10.1016/j.cub.2020.05.023>
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus I and Research T. 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England Journal of Medicine* 382: 727-733. <https://doi.org/10.1056/NEJMoa2001017>
- Zhu Z, Chakraborti S, He Y, Roberts A, Sheahan T, Xiao X, Hensley LE, Prabaharan P, Rockx B, Sidorov IA, Corti D, Vogel L, Feng Y, Kim JO, Wang LF, Baric R, Lanzavecchia A, Curtis KM, Nabel GJ, Subbarao K, Jiang S and Dimitrov DS. 2007. Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. *Proceeding of the National Academy of Sciences of United States of America* 104(29): 12123-12128. <https://doi.org/10.1073/pnas.0701000104>