COVID-19: descifrando el origen

Antoni Romeu[1] y Enric Ollé[2]

Departamento de Bioquímica y Biotecnología, Universidad Rovira i Virgili, E-43007 Tarragona, España.

[1]: Catedrático de Bioquímica y Biología Molecular. E-mail antonioramon.romeu@iubilo.urv.cat
[2]: Profesor asociado. E-mail enric.olle@urv.cat

El origen del coronavirus del síndrome respiratorio severo agudo 2 (SARS-CoV-2), el agente de la enfermedad por coronavirus 2019 (COVID-19), es controvertido. La pandemia de la COVID-19 no solo tiene efectos dramáticos sobre las personas, sino que también produce efectos irreversibles en la sociedad, por ejemplo, en la economía, en la política, en las relaciones personales. De repente, el coronavirus llegó y nadie sabe de dónde. Lo que es verdad que ha llegado para quedarse. Establecer el origen del SARS-CoV-2 es un reto, el cual está ligado al origen de otros coronavirus relacionados. En el presente trabajo se ha abordado dicho reto mediante una aproximación bioinformática. El objetivo es intentar encajar las piezas del rompecabezas. La fuente de información ha sido las bases de datos del Centro Nacional de Información Biotecnológica (NCBI) (https://www.ncbi.nlm.nih.gov/) y la metodología se ha basado en los recursos bioinformáticos del propio NCBI y del Laboratorio Europeo de Biología Molecular (EMBL) (https://www.embl.de/). En el espíritu del trabajo, ha estado siempre la reproducibilidad de los resultados, y mantener un debate sobre el origen del SARS-CoV-2.

¿Qué es lo aportado de nuevo con respecto al conocimiento previo?

Los resultados revelan la presencia de unos marcadores genéticos (*fingerprints*) en las secuencias de los genomas de los coronavirus relacionados con la COVID-19 y establecen una relaciones filogenéticas, que no se han sido descritas hasta la fecha:

- Primero, utilizando la herramienta de búsqueda de alineación local básica (*Basic Local Alignment Search Tool*, BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) del NCBI, se identificaron claramente tres marcadores genéticos en las secuencias de los genomas de coronavirus relacionados con la COVID-19 (Tabla 1): (*i*) al inicio del genoma, también en el inicio del gen *Orf1ab* de la replicasa de ARN, donde se codifica el Macrodominio N-terminal (módulo de unión de ATP); (*ii*) en el gen S, donde se codifica el dominio N-terminal y en el dominio de unión al receptor (RBD) de la glicoproteína de pico (en adelante en inglés *spike glycoprotein*) (la *spike glycoprotein* es una proteína de membrana, en la parte exterior del coronavirus, representada mediante triángulos azules sobre la superficie (tan familiar en nuestros días), que al sobresalir, da un aspecto de corona; de aquí el nombre de "coronavirus"); y (*iii*) al final del genoma, el propio gen *NS8*. Las secuencias de los coronavirus *Bat SARS like* (Bat-SL-CoV) (muestras ZXC21 y ZC45, números de acceso de la base de datos GenBank MG772934.1 y MG772933.1, respectivamente) fueron las únicas secuencias naturales de ADN que coincidieron perfectamente (*match* perfecto) con los marcadores de los genes *Orf1ab* y *NS8*. Por otro lado, el marcador del gen *S* no presentó ninguna coincidencia (*match*) con ninguna otra secuencia natural de ADN.

- Segundo, el análisis filogenético basado en las secuencias de genomas completos mostró que los coronavirus Bat-SL-CoV (ZXC21 y ZC45), Pangolin-CoV (muestra MP789, GenBank MT121216.1),

BatCoV-RaTG13 (GenBank MN996532.1) y SARS -CoV-2 eran ortólogos. Es decir, se originaron a partir de un ancestro común, y se separaron entre sí mediante posteriores eventos de especiación. Es importante remarcar que no todos los coronavirus de pangolín se comportaron igual. Solo el coronavirus Panglolin-CoV (MP789) pertenece a este grupo de ortólogos. Otros coronavirus de pangolín estaban más distanciados filogenéticamente, consistentemente (soporte *bootstrap* 1000), se agruparon en otra rama del árbol filogenético (Figura 1).

- Tercero, el análisis filogenético basado en secuencias de la *spike glycoprotein* (concretamente, desde el dominio N-terminal hasta el RBD, inclusive) mostró un patrón diferente. Las secuencias de BatCoV-RaTG13 y SARS-CoV-2 divergieron significativamente (soporte *bootstrap* 1000) de las secuencias de Bat-SL-CoV (ZXC21 y ZC45) y Pangolin-CoV (MP789). Además, las dos últimas se agruparon consistentemente (soporte *bootstrap* 1000) con otras secuencias de murciélagos. En este análisis filogenético, los otros coronavirus de pangolín también se agruparon consistentemente (soporte *bootstrap* 1000) en otra rama del árbol (*cluster*) (Figura 2).

¿Cómo y cuándo apareció el coronavirus de murciélago BatCoV-RaTG13?

El coronavirus BatCoV-RaTG13 se considera, de forma muy probable, que sea el progenitor directo del SARS-CoV-2 (1). Nuestros resultados también mostraron la estrecha relación filogenética entre ambas especies, sin embargo, ¿qué es lo que sabemos de este coronavirus de murciélago? El genoma de BatCoV-RaTG13 fue identificado y secuenciado por Zheng-Li Shi y colaboradores, 2020 (1), en el marco del estudio de un "nuevo coronavirus", que provocó la epidemia de síndrome respiratorio agudo en humanos en Wuhan, China (diciembre de 2019). Literalmente de (1): *Descubrimos que una región corta de ARN polimerasa dependiente de ARN (RdRp) de un coronavirus de murciélago (BatCoV RaTG13), que se detectó previamente en Rhinolophus affinis de la provincia de Yunnan, mostró una alta identidad de secuencia con 2019-nCoV (SARS-CoV-2)*. Entonces, Zheng-Li Shi y colaboradores, 2020 (1), llevaron a cabo la secuenciación completa de los dos genomas de BatCoV-RaTG13 y SARS-CoV-2, encontrando una identidad de secuencia genómica entre ambos de 96,2%. La secuencia completa del genoma BatCoV-RaTG13 fue sometida e introducida a GenBank por Z.-L. Shi y colaboradores el 27 de enero de 2020. En la ficha de BatCoV-RaTG13 GenBank (MN996532.1), aparece como fuente, muestra fecal; como huesped, *Rhinolophus affinis*; como país, China; y como fecha de recogida de la muestra, 24-jul-2013. Es decir, el coronavirus BatCoV RaTG13 fue aislado de una muestra fecal de murciélago en Yunnan en 2013, años antes de que se identificara por primera vez el SARS-CoV-2 y del inicio de la pandemia.

"Spike glycoprotein" de BatCoV-RaTG13: una incongruencia filogenética

En relación a la divergencia de BatCoV-RaTG13 y SARS-CoV-2 en la filogenia basada en las *spike glicoproteins* (Figura 1 y Figura 2), cabe mencionar que la región RBD se considera como la región más variable de los genomas del coronavirus (2). Sin embargo, como se muestra en el multialineamiento (Figura 3), la región del dominio N-terminal y del RBD está altamente conservada y muchos residuos de aminoácidos están estrictamente conservados. De modo que, de acuerdo con la teoría de Motoo Kimura (3) sobre evolución molecular, en esta región tan crítica de los genomas del coronavirus, el equilibrio entre la deriva aleatoria y las constricciones funcionales debe haber estado muy comprometido, a lo largo de la evolución. No obstante, la *sipke glicoproteína* de SARS-CoV-2 se ha optimizado mucho para unirse al receptor humano ACE2 (2,4). ¡Por supuesto que sí! Es precisamente a través del RBD que el coronavirus se adhiere a la membrana celular e interactúa con el receptor del huésped, iniciando la infección (4,5).

No obstante, ¿qué evento evolutivo podría explicar esta extraordinaria optimización de la *spike glycoprotein* del SARS-CoV-2 para la infección humana, y también la incongruencia filogenética? Desde una perspectiva de evolución molecular, asumiendo que BatCoV-RaTG13 sea el progenitor directo del SARS-CoV-2, debemos situar este evento evolutivo en el genoma de BatCoV-RaTG13. Se podría considerar un evento de recombinación entre BatCoV-RaTG13 y otros coronavirus, o una transferencia horizontal génica que

reemplaza regiones variables en el gen *S* de BatCoV-RaTG13 (6,7). Pero, ya sea tanto en una recombinación o en una transferencia horizontal, siempre hay un "donador" del ADN que se recombina o se transfiere. En el presente estudio, no hemos podido identificar a ningún "donador" analizando la colección completa de nucleótidos del NCBI. Este razonamiento está de acuerdo con Zheng-Li Shi y colaboradores, 2020 (1), literalmente de (1): *Usando multialineamientos de secuencias del genoma de 2019-nCoV, RaTG13, SARS-CoV y SARSr-CoV de murciélago reportados anteriormente, no se detectaron evidencias de eventos de recombinación en el genoma de 2019-nCoV (SARS-CoV-2)*.

El misterio del sitio de escisión de furina en la "spike glycoprotein" de SARS-CoV-2

En la secuencia de la *spike glycoprotein* de SARS-CoV-2 hay una pequeña inserción de cuatro aminoácidos, que no está en la de BatCoV-RaTG13. Esta pequeña inserción es responsable de la alta patogenicidad de la COVID-19. Esta inserción se conoce como "sitio polibásico de escisión de furina" (en adelante en inglés *furin site*). Es polibásico porque contiene aminoácidos básicos y es una pequeña región de la proteína que interacciona con la "furina", otra proteína de membrana de las células humana (una proteasa), que favorece enormemente la infección del SARS-CoV-2.

Por otro lado, BatCoV-RaTG13 y SARS-CoV-2 también comparten otras tres pequeñas inserciones en el dominio N-terminal de la proteína (Figura 3), en comparación con la misma proteína de los coronavirus de su grupo taxonómico. Dado que coronavirus de murciélago BatCoV-RaTG13 se aisló en 2013, es poco probable que ambos coronavirus hayan adquirido idénticas inserciones en tres sitios distintos de la proteína (8). El hecho de compartir estas tres inserciones, apoya la hipótesis de que BatCoV-RaTG13 sea precursor de SARS-CoV-2. Se considera que el SARS-CoV-2 surgió directamente del murciélago al ser humano (1). Entonces, ¿cuándo ocurrió la inserción del *furin site*? Debe haber ocurrido en el huésped o durante la transmisión por recombinación, pero, no se han detectado eventos de recombinación en el SARS-CoV-2 (1).

Concretamente, el *furin site,* está constituido por los cuatro aminoácidos "PRRA", codificados por la inserción de 12 nucleótidos (CCT CGG CGG GCA) en la región conservada y bisagra del centro de escisión de las cadenas S1/S2 de la *spike glycoprotein*. A nivel de secuencia, el *furin site* se localiza 144 posiciones de aminoácidos aguas abajo del RBD (Figura 3). Por otra parte, el *furin site* tambiés se encuentra en otros beta-coronavirus de otros linajes, pero no del linaje B, en el cual se ha clasificado el SARS-CoV-2 (9). A modo de ejemplo, a continuación se muestra un fragmento de un múltiple alineamiento de secuencias de *spike glycoprotein*, de beta-coronavirus de otros linajes, que cubre la región del *furin site* (indicados en amarillo). La presencia de un doblete de arginina es una característica estructural del sitio de unión a la furina (10):

```
Lineage A, Human coronavirus HKU1 (1495877059)                          YALPSSRRKRRGI    758
Lineage A, Murine hepatitis virus (ADI59790.1)                          YST--AHRARTSV    759
Lineage A, Human coronavirus OC43 (998640295)                           YSK--TRRSRRAI    766
Lineage B, SARS-CoV-2 (QHR63260.2)                                      YQTQTN-SPRRAR    685
Lineage B, SARS-CoV-2 (QII57208.1)                                      YQTQTN-SPRRAR    685
Lineage C, Middle East respiratory syndrome-related CoV (1453282535)    PDTPSTLTPRSVR    751
Lineage C, Pipistrellus bat coronavirus HKU5 (1386872228)               PPSPSARLARSAR    749
                                                                                 *
```

Como idea fundamental, el *furin site* es responsable de la alta infectividad y transmisibilidad de la COVID-19 (11). La interacción con la furina mejora la fusión célula-célula y media la fusión de la membrana. También está involucrada en otras enfermedades infecciosas y el cáncer (12), ahora, es la clave de la COVID-19. Es a través de este sitio que el SARS-CoV-2 se ha optimizado realmente para unirse al receptor humano ACE2 e ingresar a las células humanas (13). El origen del *furin site* en el SARS-CoV-2 es desconcertante.

¿Existe algún animal huested intermedio?

Inicialmente, se consideró que los pangolines eran el huésped intermedio (14). Sin embargo, en base a los presentes resultados y los descritos en la literatura, los análisis moleculares y filogenéticos no apoyan la hipótesis que el SARS-CoV-2 surgiera directamente de coronavirus de pangolín (15). Recientemente, C.M. Freuling et al., 2020 (16) también plantean la hipótesis de que los perros mapache (*Nyctereutes procyonoides*) podrían haber sido huéspedes intermediarios del SARS-CoV-2, pero solo muestran que estos animales son susceptibles al coronavirus (16). Es un caso de zoonosis, donde probablemente el SARS-CoV-2 saltó entre humanos y animales no humanos. Otro caso de zoonosis ha sido el de Dinamarca que ha de matar a 17 millones de visones para frenar una nueva variante del coronavirus que ha saltado a los humanos.

Entonces, a pesar de que el contacto directo entre humanos y murciélagos es limitado, y generalmente una especie intermedia a menudo está implicada en la transmisión de un virus emergentes de murciélagos a humanos (17), hasta la fecha, no es el caso en el SARS-CoV-2. O al menos, no se ha descubierto ningún huésped intermedio.

Interfaz entre biología y lógica

Los principios básicos de la biología también se cumplen en el mundo de los virus. De la Teoría Celular de Rudolf Virchow (1858), *Omnis cellula ex cellula* (cada célula deriva de otra célula preexistente), podría inferirse como "cada virus deriva de un virus preexistente". La selección natural de Charles Darwin (1859) sobre mutaciones y la lucha por la existencia es totalmente aplicable a los virus. El principio de Theodosius Dobzhansky (1973) "Nada en biología tiene sentido excepto a la luz de la evolución" (18) adquiere relevancia en el caso del origen del SARS-CoV-2. A partir de la información actual disponible en las bases de datos, es difícil encajar dicho origen en un modelo evolutivo racional.

La clasificación taxonómica del SARS2-CoV-2 también es confusa. Para abordar este punto, hemos considerado un criterio de genética forense, que se basa en una relación de probabilidades (*likelihood ratio*, LR), y tiene validez en un juicio. El LR es un parámetro que compara dos probabilidades de encontrar un mismo genotipo o marcadores genéticos de una evidencia, en dos personas distintas. El "LR" se utiliza habitualmente para expresar los resultados de una prueba de ADN. Cuando un juez dicta sentencia sobre la culpabilidad de un sospechoso, basada en una prueba de ADN, requiere que la probabilidad de que el ADN de la evidencia sea del sospechoso, debe ser mucho mayor (billones superior, $> 10^{12}$) que la probabilidad de que el mismo ADN sea de una persona aleatorio de la población a la que pertenece el sospechoso. Así, en genética forense, para que el LR sirva de testigo de cargo, debe ser enormemente alto.

En el caso de la clasificación taxonómica del SARS2-CoV-2, hay una duda razonable de que pertenezca al grupo taxonómico, al cual se ha clasificado (linaje B del beta-coronavirus). Debido a la presencia del *furin site*, un LR es "infinito". Esto es, la relación entre la probabilidad (P1) de que dada la secuencia actual del genoma del SARS-CoV-2, pertenezca a SARS-CoV-2, obviamente, es 1; y la probabilidad (P2) de que dada la misma secuencia del genoma del SARS-CoV-2, pertenezca a otro coronavirus de su grupo taxonómico, por el momento es 0 (porque el *furin site* solo se ha encontrado en el SARS-CoV-2). Entonces ese LR que compara las dos probabilidades P1 y P2 es "infinito" (LR = P1 / P2 = 1/0 = infinito). Esto es una forma probabilística de mostrar la dificultad de asociar el SARS-CoV-2 con el linaje B de los beta-coronavirus. En este sentido, Z.-L. Shi et al., 2020 (1), (sin tener en cuenta el *furin site*), también apuntaron esta duda. Literalmente de (1): *El análisis filogenético del genoma completo y las secuencias de genes de RdRp y spike (S) mostró que, para todas las secuencias, RaTG13 es el pariente más cercano de 2019-nCoV (SARS-CoV-2) y forman un linaje distinto de otros SARSr-CoVs*.

Dado que BatCoV-RaTG13 se considera el origen probable del SARS-CoV-2 (1), es sorprendente que después de casi un año de la pandemia de la COVID-19, no se hayan aislado más coronavirus de murciélago de la especie BatCoV-RaTG13. El murciélago es el reservorio natural de virus más estudiado. Por otro lado, los coronavirus de murciélago podrían aislarse de muestras fecales (¡no sería necesario un hisopo nasal de

murciélago!). Una vez aislados, sus genomas completos se introducen en la base de datos GenBank. Actualmente solo hay un genoma disponible de BatCoV-RaTG13 en GenBank (el referenciado, MN996532.1). Se requieren más genomas de BatCoV-RaTG13 para validar estadísticamente el 96,2% de identidad con el genoma de SARS-CoV-2. todavía queda un 3,8% de diferencia entre ambos, lo que puede ser clave para explicar la presencia del *furin site* en la *spike glycoprotein* del SARS-CoV-2. Por el contrario, actualmente hay miles de genomas completos disponibles de SARS-CoV-2 en GenBank.

En la evolución biológica siempre hay "eslabones perdidos", en algunos casos, partir de descubrimientos de nuevos restos fósiles se pueden resolver. En el mundo de los virus no hay restos fósiles, pero los “eslabones perdidos” se pueden solucionar mediante el descubrimiento de nuevos virus. La presencia del *furin site* en el SARS-CoV-2 (en la parte más conservada de la proteína, y lejos de la región más variable) es señal de un "eslabón perdido" en nuestro conocimiento de su proceso evolutivo, y en definitiva de su origen. Este eslabón perdido también pone en duda que el BatCoV-RaTG13 sea el progenitor directo del SARS-CoV-2. Como sucede en genética forense, que también se aplica en el mundo de los virus, cuando un marcador genético falla en una prueba de ADN, es suficiente para descartar una culpabilidad. Por lo tanto, existe una duda razonable sobre el origen del SARS-CoV-2, lo cual puede ser "un nuevo paradigma para la virología". La disponibilidad de nuevos genomas BatCoV-RaTG13 es esencial para poder abrir perspectivas. De lo contrario, el origen del SARS-CoV-2 es un lado oscuro de la COVID-19. La duda puede hacer pensar que el SARS-CoV-2 sea un producto de laboratorio o un virus manipulado a propósito. La tecnología necesaria para ello está disponible. Para fines científicos y médicos, existen varios constructos sintéticos del genoma del SARS-CoV-2 en GenBank (MT108784.1, MT461669.1, MT461671.1, MT461670.1). Con fines terapéuticos, en 2008 se creó un coronavirus recombinante sintético similar al SARS de murciélago, que resultó infeccioso en células cultivadas y en ratones (19). En este sentido, el debate debe continuar (20).

Como motivo de esperanza, por razones de probabilidad, y debido al mecanismo de "ensayo-error" de la evolución biológica, hay que pensar que un virus humano como el SARS-CoV-2, no volverá a surgir de la Madre Naturaleza por mucho, mucho tiempo (a nadie le toca el Gordo de la Lotería dos veces). Finalmente, hay que esperar una próxima vacuna y/o medicamentos efectivos para la COVID-19.

Referencias

1. Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao , Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, Zheng-Li Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273, 2020. PMID: 32015507. doi: 10.1038/s41586-020-2012-7.

2 . Kristian G Andersen, Andrew Rambaut, W Ian Lipkin, Edward C Holmes, Robert F Garry. The proximal origin of SARS-CoV-2. Nat. Med. 26:450-452, 2020. PMID: 32284615. doi: 10.1038/s41591-020-0820-9.

3. Kimura, M. The neutral theory of molecular evolution. Cambridge University, Cambridge. UK (1983).

4. Yushun Wan, Jian Shang, Rachel Graham, Ralph S Baric, Fang Li. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J. Virol. 94, e00127-20, 2020.PMID: 31996437. doi.org/10.1128/JVI.00127-20.

5. UNIPROT. P59594, Spike glycoprotein. Accessed October 09, 2020. https://www.uniprot.org/uniprot/P59594.

6 Dong-Sheng Chen, Yi-Quan Wu, Wei Zhang, San-Jie Jiang, Shan-Ze Chen. Horizontal gene transfer events

reshape the global landscape of arm race between viruses and homo sapiens. Sci. Rep. 6:26934, 2016. PMID: 27270140. doi: 10.1038/srep26934.

7. Shahana S Malik, Syeda Azem-E-Zahra, Kyung Mo Kim, Gustavo Caetano-Anollés, Arshan Nasir. Do Viruses Exchange Genes across Superkingdoms of Life? Front. Microbiol. 8, 2110, 2017. PMID: 29163404. doi.org/10.3389/fmicb.2017.02110.

8. Xiaojun Li, Elena E Giorgi, Manukumar Honnayakanahalli Marichannegowda, Brian Foley, Chuan Xiao, Xiang-Peng Kong, Yue Chen, S Gnanakaran, Bette Korber, Feng Gao. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. 6(27):eabb9153. PMID: 32937441. doi: 10.1126/sciadv.abb9153.

9. Javier A Jaimes, Nicole M André, Joshua S Chappie, Jean K Millet, Gary R Whittaker. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. J. Mol. Biol. 432:3309-3325, 2020. PMID: 32320687. doi: 10.1016/j.jmb.2020.04.009.

10. Imène Kara, Marjorie Poggi, Bernadette Bonardo, Roland Govers, Jean-François Landrier, Sun Tian, Ingo Leibiger, Robert Day, John W M Creemers, Franck Peiretti. The Paired Basic Amino Acid-cleaving Enzyme 4 (PACE4) Is Involved in the Maturation of Insulin Receptor Isoform B. J. Biol. Chem. 290:2812–2821. PMID: 25527501. doi: 10.1074/jbc.M114.592543.

11. Shuai Xia, Qiaoshuai Lan, Shan Su, Xinling Wang, Wei Xu, Zezhong Liu, Yun Zhu, Qian Wang, Lu Lu, Shibo Jiang. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. Signal Transduct. Target Ther. 5:92, 2020. PMID: 32532959. doi.org/10.1038/s41392-020-0184-0.

12. Elisabeth Braun, Daniel Sauter. Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunol. E1073, 2019. PMID: 31406574. doi.org/10.1002/cti2.1073.

13. Markus Hoffmann, Hannah Kleine-Weber, Stefan Pöhlmann. Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol. Cell 78:779–784, 2020. PMID: 32362314. doi: 10.1016/j.molcel.2020.04.022.

14. Tao Zhang, Qunfu Wu, Zhigang Zhang. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr. Biol. 30:1346–1351.e2, 2020. PMID: 32315626. doi: 10.1016/j.cub.2020.03.022.

15. Ping Liu, Jing-Zhe Jiang, Xiu-Feng Wan, Yan Hua, Linmiao Li, Jiabin Zhou, Xiaohu Wang, Fanghui Hou, Jing Chen, Jiejian Zou, Jinping Chen. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog. 16(5):e1008421, 2020. PMID: 32407364. doi: 10.1371/journal.ppat.1008421.

16. Conrad M. Freuling, Angele Breithaupt, Thomas Müller, Julia Sehl, Anne Balkema-Buschmann, Melanie Rissmann, Antonia Klein, Claudia Wylezich, Dirk Höper, Kerstin Wernike, Andrea Aebischer, Donata Hoffmann, Virginia Friedrichs, Anca Dorhoi, Martin H. Groschup, Martin Beer, Thomas C. Mettenleiter. Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. Emerg Infect Dis. 2020 Oct 22;26(12), 2020. PMID: 33089771. doi: 10.3201/eid2612.203733.

17. Shauna Milne-Price, Kerri L Miazgowicz, Vincent J Munster. The emergence of the Middle East Respiratory Syndrome coronavirus. Pathog. Dis. 71:21-176, 2014. PMID: 24585737. doi: 10.1111/2049-632X.12166.

18. Theodosius Dobzhansky. Nothing in Biology Makes Sense except in the Light of Evolution. The American

Biology Teacher 35:125-129, 1973.

19. Michelle M Becker , Rachel L Graham, Eric F Donaldson, Barry Rockx, Amy C Sims, Timothy Sheahan, Raymond J Pickles, Davide Corti, Robert E Johnston, Ralph S Baric, Mark R Denison. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc. Natl. Acad. Sci. USA 105:19944–19949, 2008. PMID: 19036930. doi: 10.1073/pnas.0808116105.

20. Heidi J Larson. A lack of information can become misinformation. Nature 580:306, 2020. PMID: 32231320. doi: 10.1038/d41586-020-00920-w.

21. Muhamad Fahmi, Yukihiko Kubota, Masahiro Ito. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. Infect. Genet. Evol. 81:104272, 2020. PMID: 32142938. doi.org/10.1016/j.meegid.2020.104272.

22. Fábio Madeira, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R N Tivey, Simon C Potter, Robert D Finn, Rodrigo Lopez. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucl. Acids Res. 47(W1):W636-W641, 2019. PMID: 30976793. doi: 10.1093/nar/gkz268.

23. Ivica Letunic, Peer Bork. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucl. Acids Res. 2011, Vol. 39(W475–W478), 2011. PMID: 21470960. doi:10.1093/nar/gkr201.

24. Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, Xinquan Wang. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215-220, 2020. PMID: 32225176. doi:10.1038/s41586-020-2180-5.

Agradecimientos

Este trabajo no ha recibido subvenciones de ninguna institución de apoyo a la investigación.

Declaración de conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Tabla 1. Marcadores genómicos de SARS-CoV-2, BatCoV-RaTG13, Pangolin-CoVs y Bat-SL-CoV

Especie	Referencia genoma	Gen	Coordenadas del genoma	Proteina. GenBank id	Posición proteína
SARS-CoV-2	GenBank MN996528.1	*orf1ab*	1940 - 3955	Orf1ab polyprotein. QHR63259.1	559 - 1230
	isolate WIV04	*S*	21563 - 22963	Spike glycoprotein. QHR63260.2 (RBD)	1 - 467
		NS8	27912 - 28256	Nonstructural protein NS8. QHR63267.1	7 - 121
BatCoV-RaTG13	GenBank MN996532.1	*orf1ab*	1925 - 3937	Orf1ab polyprotein. QHR63299.1	559 - 1229
		S	21545 - 22945	Spike glycoprotein. QHR63300.2 (RBD)	1- 467
		NS8	27872 - 28222	Nonstructural protein NS8. QHR63307.1	5 - 121
Pangolin-CoV	GenBank MT040335.1	*orf1ab*	2220 - 3884	Orf1ab polyprotein. QIA48631.1	652 - 1206
	isolate PCoV_GX-P5L	*S*	21540 - 22940	Spike glycoprotein. QIA48632.1 (RBD)	1 - 467
		orf8	27875 - 28210	Orf8 protein. QIA48638.1	9 - 121
Pangolin-CoV	GenBank MT121216.1	*orf1ab*	2102 - 3814	Orf1ab polyprotein. QIG55944.1	653 - 1223
	isolate MP789	*S*	21421 -22821	Spike glycoprotein. QIG55945.1 (RBD)	1 - 467
		orf8	27728 - 28042	Orf8 protein. QIG55952.1	1 - 105
Bat-SL-CoV	GenBank MG772933.1	*1ab*	2218 - 3948	non-structural polyprotein 1ab.AVP78030.1	652 - 1228
	isolate Bat-SL-CoVZC45	*S*	21549 - 22949	Spike protein. AVP78031.1 (RBD)	1 - 467
		10b	27799 - 28161	Hypothetical protein. AVP78037.1	1 - 121

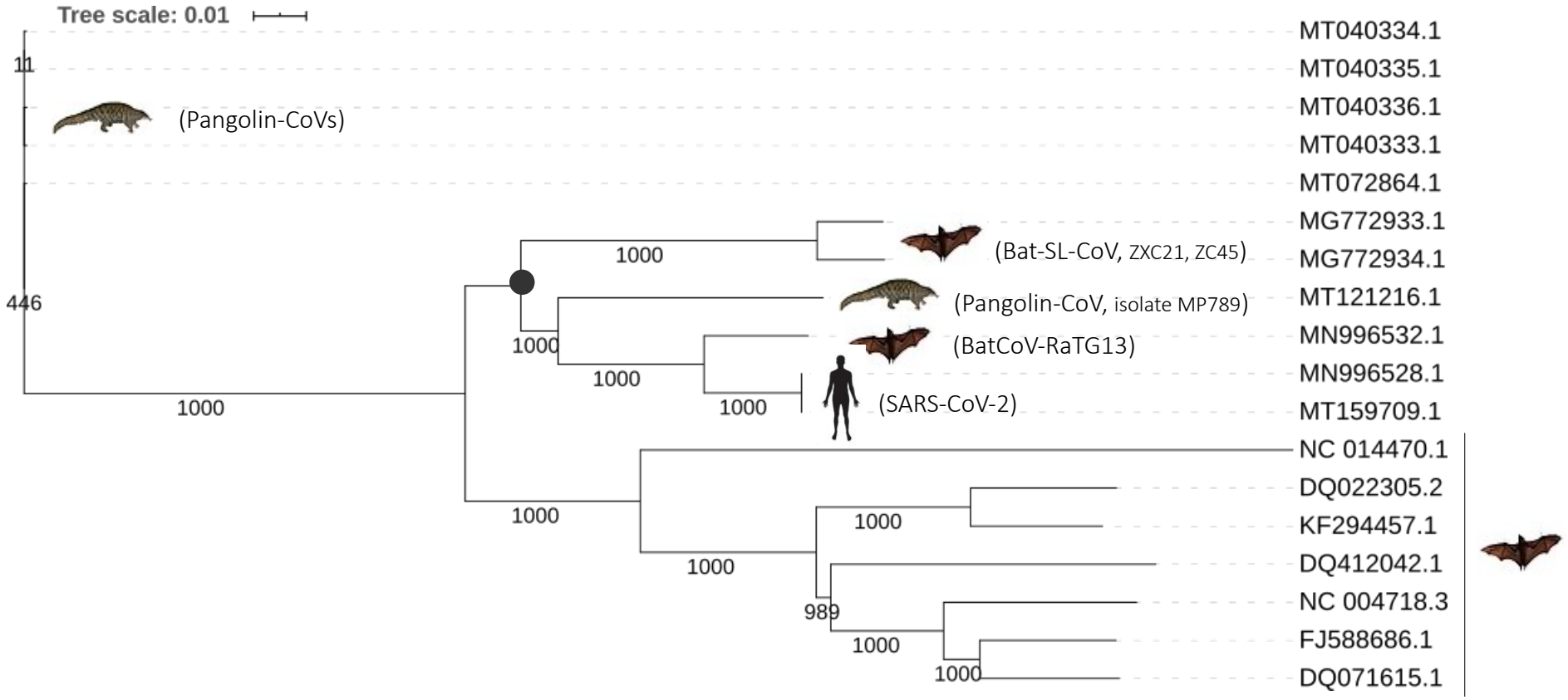
Tree scale: 0.01
11
(Pangolin-CoVs)
446
1000
1000
1000
1000
(Bat-SL-CoV, ZXC21, ZC45)
(Pangolin-CoV, isolate MP789)
(BatCoV-RaTG13)
(SARS-CoV-2)
1000
1000
1000
989
1000
1000
MT040334.1
MT040335.1
MT040336.1
MT040333.1
MT072864.1
MG772933.1
MG772934.1
MT121216.1
MN996532.1
MN996528.1
MT159709.1
NC 014470.1
DQ022305.2
KF294457.1
DQ412042.1
NC 004718.3
FJ588686.1
DQ071615.1

Figure 1

Figura 1. Árbol filogenético de una muestra de coronavirus basado en secuencias de genomas completos

Árbol filogenético basado en un alineamiento múltiple de una selección de genomas completos de coronavirus. La muestra incluye el BatCoV-RaTG13, los coronavirus de pangolín disponibles, una selección de coronavirus de murciélago extraídos de la literatura (1,14,21) y dos genomas del SARS-CoV-2 como representantes del grupo taxonómico: NCBI, *Severe acute respiratory syndrome coronavirus 2 (taxid:2697049)*. El árbol filogenético se construyó utilizando el método *Neighbor Joining* del software Clustal Omega (v.1.2.4) con parámetros por defecto (22) y la herramienta *iTol Interactive Tree Of Life* (23). La consistencia de las agrupaciones se calculó mediante *bootstrap* utilizando 1000 repeticiones. La barra de escala de árbol representa la distancia evolutiva. El punto negro indica el ancestro común del grupo de coronavirus relacionados con la COVID-19. El número de acceso de GenBank de los genomas completos y el coronavirus son los siguientes (en la misma disposición que en el árbol filogenético): MT040333.1 a MT040336.1, coronavirus pangolín; MG772933.1 y MG772934.1, coronavirus de tipo murciélago SARS; MT121216.1 (aislado MP789) Coronavirus de pangolín; MN996532.1, BatCoV-RaTG13; MN996528.1 y MT159709.1, SARS-CoV-2; NC_014470.1, Coronavirus de murciélago BM48-31/BGR/2008; DQ022305.2, coronavirus HKU3-1 del SARS de murciélago; KF294457.1, coronavirus de tipo murciélago SARS; DQ412042.1, Bat SARS CoV Rf1/2004; NC_004718.3, coronavirus Tor2 del SARS; FJ588686.1, coronavirus del SARS Rs_672/2006; DQ071615.1, Bat SARS CoV Rp3/2004.

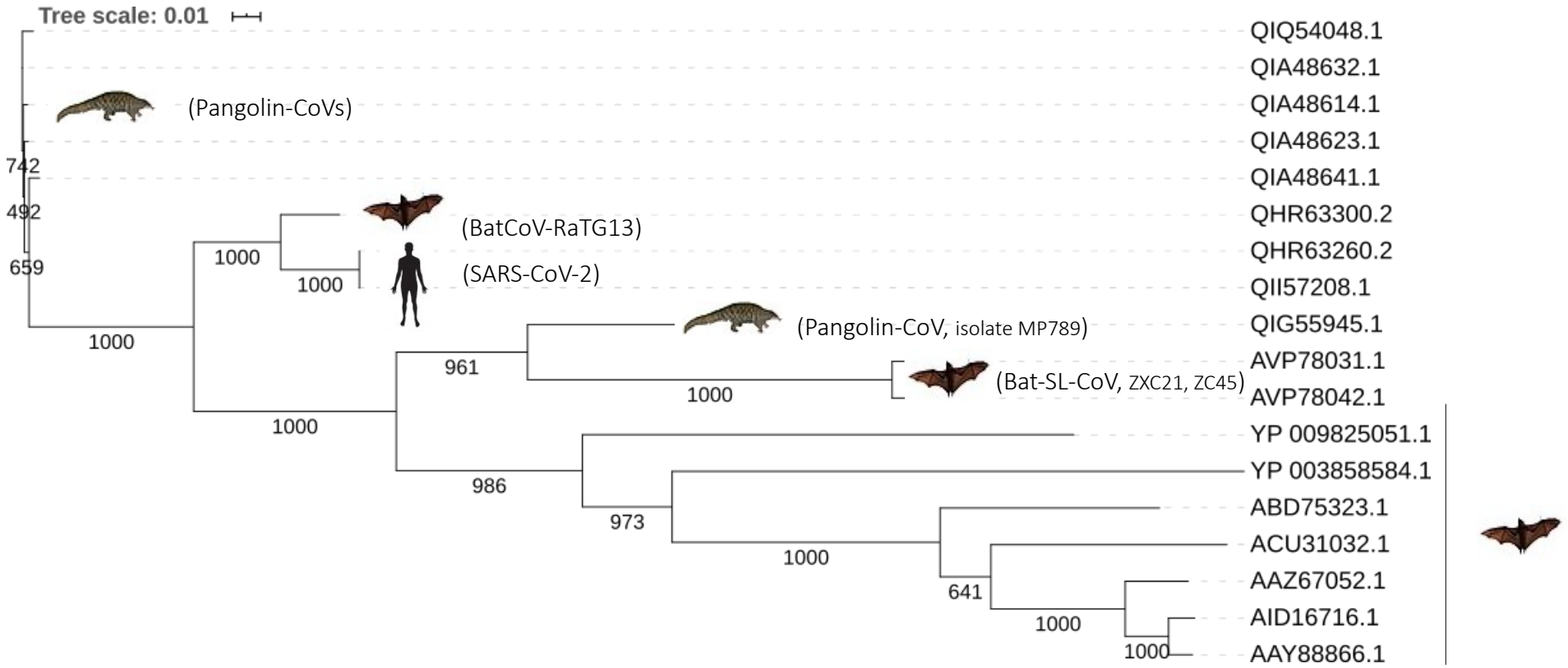
Tree scale: 0.01
(Pangolin-CoVs)
(BatCoV-RaTG13)
(SARS-CoV-2)
(Pangolin-CoV, isolate MP789)
(Bat-SL-CoV, ZXC21, ZC45)
QIQ54048.1
QIA48632.1
QIA48614.1
QIA48623.1
QIA48641.1
QHR63300.2
QHR63260.2
QII57208.1
QIG55945.1
AVP78031.1
AVP78042.1
YP 009825051.1
YP 003858584.1
ABD75323.1
ACU31032.1
AAZ67052.1
AID16716.1
AAY88866.1
742
492
659
1000
1000
1000
961
1000
1000
986
973
1000
641
1000
1000

Figure 2

Figura 2. Árbol filogenético de una muestra coronavirus basado en secuencias de *spike glycoprotein*

Árbol filogenético basado en el multialineamiento de la región más variable de la *spike glycoprotein* (desde el dominio N-terminal hasta el *RB*D). La muestra incluye las secuencias de los mismos coronavirus de la Figura 1, con el objetivo de comparar el comportamiento filogenético, en función del tipo de seciencia. La posición de RBD se basó en (24). El árbol filogenético se construyó utilizando el método *Neighbor Joining* del software Clustal Omega (v.1.2.4), con parámetros por defecto (22) y la herramienta *iTol Interactive Tree Of Life* (23). La consistencia de las agrupaciones se calculó mediante *bootstrap* utilizando 1000 repeticiones. La barra de escala de árbol representa la distancia evolutiva. El número de acceso de GenBank de la *spike glycoprotein* y el correspondiente coronavirus con los siguientes (en la misma disposición que en el árbol filogenético): QIA48641.1, QIA48632.1, QIA48614.1, QIA48623.1, QIQ54048.1, coronavirus pangolín; QHR63300.2, coronavirus de murciélago RaTG13; QHR63260.2, QII57208.1, SARS-CoV-2; QIG55945.1, coronavirus de pangolín; AVP78031.1, AVP78042.1, coronavirus de tipo murciélago SARS; YP_009825051.1, coronavirus Tor2 del SARS; YP_003858584.1, Coronavirus de murciélago BM48-31/BGR/2008; ABD75323.1, Bat SARS CoV Rf1/2004; ACU31032.1, coronavirus del SARS Rs_672/2006; AAZ67052.1, Bat SARS CoV Rp3/2004; AID16716.1, coronavirus de tipo murciélago SARS; AAY88866.1, coronavirus HKU3-1 del SARS de murciélago.

```
ADC35483.1                -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
sp|P59594|SPIKE_CVHSA     -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
AAR07630.1                -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
AVP78031.1                MLFFLFLQFALVNS----QCVNLTGRTPLNPNYTNSSQRGVYYPDTIYRSDTLVLSQGYF  56
AVP78042.1                MLFFLFLQFALVNS----QC-DLTGRTPLNPNYTNSSQRGVYYPDTIYRSDTLVLSQGYF  55
QIG55945.1                MLFFFFLHFALVNS----QCVNLTGRAAIQPSFTNSSQRGVYYPDTIFRSNTLVLSQGYF  56
QIQ54048.1                -MFVFLFVLPLVSS----QCVNLTTRTGIPPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QIA48614.1                -MFVFLFVLPLVSS----QCVNLTTRTGIPPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QIA48623.1                -MFVFLFVLPLVSS----QCVNLTTRTGIQPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QHR63260.2                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QII57208.1                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QIA98554.1                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QHR63300.2                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSSTRGVYYPDKVFRSSVLHLTQDLF  55
                           :*.::. : *...    :*  :           .*  ******* ::**. *  :*. *

ADC35483.1                LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
sp|P59594|SPIKE_CVHSA     LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
AAR07630.1                LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
AVP78031.1                LPFYSNVSWYYSLTTN-NAATKRTDNPILDFKDGIYFAATEHSNIIRGWIFGTTLDNTSQ  115
AVP78042.1                LPFYSNVSWYYSLTTN-NAATKRTDNPILDFKDGIYFAATEHSNIVRGWIFGTTLDNTSQ  114
QIG55945.1                LPFYSNVSWYYALTKT-NSAEKRVDNPVLDFKDGIYFAATEKSNIVRGWIFGTTLDNTSQ  115
QIQ54048.1                LPFFSNVTWFNTIHLNYQGGFKKFDNPVLPXNDGVYFASTEKSNIIRGWIFGTTLDARTQ  115
QIA48614.1                LPFFSNVTWFNTI--NYQGGFKKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDARTQ  113
QIA48623.1                LPFFSNVTWFNTI--NYQGGFKKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDARTQ  113
QHR63260.2                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QII57208.1                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QIA98554.1                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QHR63300.2                LPFFSNVTWFHAIHVSGTNGIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
                          ***:***: : ::        :  .**::  :**:***:**:**::***:**:*::  :*

ADC35483.1                SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
sp|P59594|SPIKE_CVHSA     SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
AAR07630.1                SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
AVP78031.1                SLLIVNNATNVIIKVCNFDFCYDPYLSGYYHN-NKTWSIREFAVYSSYANCTFEYVSKSF  174
AVP78042.1                SLLIVNNATNVIIKVCNFDFCYDPYLSGYYHN-NKTWSIREFAVYSFYANCTFEYVSKSF  173
QIG55945.1                SLLIVNNATNVIIKVCNFQFCYDPYLSGYYHN-NKTWSTREFAVYSSYANCTFEYVSKSF  174
QIQ54048.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  175
QIA48614.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  173
QIA48623.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  173
QHR63260.2                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QII57208.1                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QIA98554.1                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QHR63300.2                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
                          *::*:**:***:*:.*:*::* :*::.      .      .  ::.   ******:*. *

ADC35483.1                SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
sp|P59594|SPIKE_CVHSA     SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
AAR07630.1                SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
AVP78031.1                MLNISGNGGLFNTLREFVFRNVDGHFKIYSKFTPVNLNRGLPTGLSVLQPLVELPVSINI  234
AVP78042.1                MLNISGNGGLFNTLREFVFRNVDGHFKIYSKFTPVNLNRGLPTGLSVLQPLVELPVSINI  233
QIG55945.1                MLDIAGKSGLFDTLREFVFRNVDGYFKIYSKYTPVNVNSNLPIGFSALEPLVEIPAGINI  234
QIQ54048.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  235
QIA48614.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  233
QIA48623.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  233
QHR63260.2                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QII57208.1                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QIA98554.1                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QHR63300.2                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPGFSALEPLVDLPIGINI  235
                           :::  : * *. ******:* **.: :*. . *:::  .** *: .*:*:..:* .***

ADC35483.1                TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
sp|P59594|SPIKE_CVHSA     TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
AAR07630.1                TNFRAILTAFS------PAQDTWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
AVP78031.1                TKFRTLLTIHRGDPMP---NNGWTAFSAAYFVGYLKPRTFMLKYNENGTITDAVDCALDP  291
AVP78042.1                TKFRTLLTIHRGDPMS---NNGWTAFSAAYFVGYLKPRTFMLKYNENGTITDAVDCALDP  290
QIG55945.1                TKFRTLLTIHRGDPMP---NNGWTVFSAAYYVGYLAPRTFMLNYNENGTITDAVDCALDP  291
QIQ54048.1                TRFQTLLALHRSYLTPGKLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  295
QIA48614.1                TRFQTLLALHRSYLTPGNLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  293
QIA48623.1                TRFQTLLALHRSYLTPGKLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  293
QHR63260.2                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QII57208.1                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QIA98554.1                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QHR63300.2                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
                          *.*:::*: .         .. * . :***:****   **:*.*::**********: :*

ADC35483.1                LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
sp|P59594|SPIKE_CVHSA     LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
AAR07630.1                LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
AVP78031.1                LSETKCTLKSLTVQKGIYQTSNFRVQPTQSVVRFPNITNVCPFHKVFNATRFPSVYAWER  351
AVP78042.1                LSETKCTLKSLSVQKGIYQTSNFRVQPTQSIVRFPNITNVCPFHKVFNATRFPSVYAWER  350
QIG55945.1                LSEAKCTLKSLTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATTFASVYAWNR  351
QIQ54048.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  355
QIA48614.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  353
QIA48623.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  353
QHR63260.2                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QII57208.1                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QIA98554.1                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QHR63300.2                LSETKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNITNLCPFGEVFNATTFASVYAWNR  355
                          *:* **::**: ::*********** *: .:********:*** :****: * *****:*
```

Figure 3

```
ADC35483.1                KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
sp|P59594|SPIKE_CVHSA     KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
AAR07630.1                KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
AVP78031.1                TKISDCIADYTVFYNSTSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGQT  411
AVP78042.1                TKISDCIADYTVFYNSTSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGQT  410
QIG55945.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVRQIAPGQT  411
QIQ54048.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  415
QIA48614.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  413
QIA48623.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  413
QHR63260.2                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QII57208.1                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QIA98554.1                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QHR63300.2                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVRQIAPGQT  415
                          .:**:*:***:*:***: ********** :** ****:.****:*::  .:***:*****

ADC35483.1                GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
sp|P59594|SPIKE_CVHSA     GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
AAR07630.1                GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
AVP78031.1                GVIADYNYKLPDDFTGCVIAWNTAKQDV-----GNYFYRSHRSTKLKPFERDLSSDEN--  464
AVP78042.1                GVIADYNYKLPDDFTGCVIAWNTAKQDT-----GHYFYRSHRSTKLKPFERDLSSDEN--  463
QIG55945.1                GRIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  471
QIQ54048.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNYGYLYRLFRKSKLKPFERDISTEIYQA  475
QIA48614.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNYGYLYRLFRKSKLKPFERDISTEIYQA  473
QIA48623.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNY--LYRLFRKSKLKPFERDISTEIYQA  471
QHR63260.2                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QII57208.1                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QIA98554.1                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QHR63300.2                GKIADYNYKLPDDFTGCVIAWNSKHIDAKEGGNFNYLYRLFRKANLKPFERDISTEIYQA  475
                          * ************ ***:***: : *          **  *  :*:*****:*.

ADC35483.1                DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
sp|P59594|SPIKE_CVHSA     DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
AAR07630.1                DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
AVP78031.1                ------G------VRTLSTYDFNPNVPLEYQATRVVVLSFELLNAPATVCGPKLSTQLVK  512
AVP78042.1                ------G------VRTLSTYDFNPNVPLEYQATRVVVLSFELLNAPATVCGPKLSTQLVK  511
QIG55945.1                GSTPCNGVEGFNCYFPLQSYGFHPTNGVGYQPYRVVVLSFELLKAPATVCGPKQSTNLVK  531
QIQ54048.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSXELLNGPATVCGPKLSTTLVK  535
QIA48614.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSFELLNGPATVCGPKLSTTLVK  533
QIA48623.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSFELLNGPATVCGPKLSTTLVK  531
QHR63260.2                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QII57208.1                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QIA98554.1                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QHR63300.2                GSKPCNGQTGLNCYYPLYRYGFYPTDGVGHQPYRVVVLSFELLNAPATVCGPKKSTNLVK  535
                                     *  *.*  .  : :*  ****** ***:.******** ** *:*

ADC35483.1                NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
sp|P59594|SPIKE_CVHSA     NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
AAR07630.1                NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
AVP78031.1                NQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGKDASDFIDSVRDPQTLEILDITPCSFGGV  572
AVP78042.1                NQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGKDASDFIDSVRDPQTLEILDITPCSFGGV  571
QIG55945.1                NKCVNFNFNGLTGTGVLTESSKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  591
QIQ54048.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  595
QIA48614.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  593
QIA48623.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  591
QHR63260.2                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QII57208.1                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QIA98554.1                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QHR63300.2                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
                          ::**********.****** *.*:*  *****:* :*  *:****:* *****:*******

ADC35483.1                SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
sp|P59594|SPIKE_CVHSA     SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
AAR07630.1                SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
AVP78031.1                SVITPGTNTSLEVAVLYQDVNCTDVPTTIHADQLTPAWRIYATGTNVFQTQAGCLIGAEH  632
AVP78042.1                SVITPGTNTSSEVAVLYQDVNCTDVPTTIHADQLTPAWRIYAIGTSVFQTQAGCLIGAEH  631
QIG55945.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  651
QIQ54048.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  655
QIA48614.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  653
QIA48623.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  651
QHR63260.2                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QII57208.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QIA98554.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QHR63300.2                SVITPGTNASNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
                          ********:* :***********:*  :***:****:**:*: * .****:****:****

ADC35483.1                VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
sp|P59594|SPIKE_CVHSA     VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
AAR07630.1                VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
AVP78031.1                VNASYECDIPIGAGICASYHTASI----LRSTSQKAIVAYTMSLGAENSIAYANNSIAIP  688
AVP78042.1                VNASYECDIPIGAGICASYHTASI----LRSTGQKAIVAYTMSLGAENSIAYANNSIAIP  687
QIG55945.1                VNNTYECDIPIGAGICASYQTQTN----SRSVSSQAIIAYTMSLGAENSVAYANNSIAIP  707
QIQ54048.1                VNNSYECDIPVGAGICASYHSMS----SFRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  711
QIA48614.1                VNNSYECDIPVGAGICASYHSMS----SLRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  709
QIA48623.1                VNNSYECDIPVGAGICASYHSMS----SLRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  707
QHR63260.2                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QII57208.1                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QIA98554.1                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QHR63300.2                VNNSYECDIPIGAGICASYQTQTNS----RSVASQSIIAYTMSLGAENSVAYSNNSIAIP  711
                          *: :******:********:: :      **. .::*:********:.*:**:**:****
```

Figure 3 (continued)

Figura 3. Multialneamiento de *spike glycoprotein* de una selección de coronavirus.

Alineamiento múltiple de secuencia *spike glycoprotein*. Los grupos de las muestras seleccionadas son: (*i*) coronavirus del SARS humano (SARS-CoV) (longitud de 1255 amoniacidos); (*ii*) Bat-SL-CoV (1245); (*iii*) Pangolin-CoVs (1265-1269); (*iv*) SARS-CoV-2 (1273); y (*v*) BatCoV-RaTG13 (1269). Para visualizar mejor las características de cada grupo, solo hay tres secuencias representativas de cada uno de ellos. El multialineamiento se construyó mediante el software Clustal Omega (v.1.2.4), con los parámetros por defecto (22). Los aminoácidos estrictamente conservados se indican con *, los espacios con -. La posición de los aminoácidos en cada secuencia está indicada por los números a la derecha. Los espacios correspondientes a la delección y las tres inserciones características de las secuencias relacionadas con COVID-19, el RBD y el sitio de escisión de furina (*furin site*) del SARS-CoV-2 están resaltados en amarillo. La banda de colores (gris-blanco) tiene por objeto resaltar las características de secuencia de: SARS-CoV; Bat-SL-CoV y Pangolin-CoV (MP789); otros Pangolin-CoV; SARS-CoV-2; y BatCoV-RaTG13. La posición de RBD se basó en (24). La figura solo se muestra hasta el *furin site*. Hasta el extremo C-terminal de la proteína, la mayoría de las posiciones eran estrictamente conservadas. El número de acceso de GenBank de las secuencias de *spike glycoprotein* y los respectivos coronavirus son los siguientes: ADC35483.1, coronavirus del SARS HKU-39849; sp | P59594 | (UNIPROT SPIKE_CVHSA); AAR07630.1, coronavirus BJ302 del SARS; AVP78031.1, Bat-SL-CoV ZC45; AVP78042.1, Bat-SL-CoV ZXC21; QIG55945.1, PangolinCoV, MP789; QIQ54048.1, Pang-CoV, GX-P2V; QIA48614.1, Pang-CoV, GX-P4L; QIA48623.1, Pang-CoV, GX-P1E; QHR63260.2, SARS-CoV-2; QII57208.1 SARS-CoV-2; QIA98554.1, SARS-CoV-2; QHR63300.2, BatCoV-RaTG13.

COVID-19: deciphering the origin

Antoni Romeu[1] and Enric Ollé[2]

Department of Biochemistry and Biotechnology, Rovira i Virgili University, E-43007 Tarragona, Spain.

[1]: Professor of Biochemistry and Molecular Biology. E-mail antonioramon.romeu@iubilo.urv.cat
[2]: Associated Professor. E-mail enric.olle@urv.cat

The origin of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the coronavirus disease 2019 (COVID-19), is controversial. The COVID-19 pandemic not only has dramatic effects on people, but also produces irreversible effects on society, *e.g.*, economy, politics, personal relationships. Suddenly, the coronavirus arose and nobody knows where it came from. The simple truth is that it is here to stay. Establishing the origin of SARS-CoV-2 is a challenge, which is linked to that of the most related coronaviruses. We addressed the issue through a bioinformatic approach. The aim of this work is to try to fit together all the pieces of the puzzle. The source of information was the National Center for Biotechnological Information (NCBI) databases (https://www.ncbi.nlm.nih.gov/) and the methodology was based on the bioinformatic resources of NCBI and the European Laboratory of Molecular Biology (EMBL) (https://www.embl.de/). In was in the spirit of the work the reproducibility of the results, as well as maintain a debate on the origin of the SARS-CoV-2.

What is new to previous knowledge?

The main result reveals genomic sequence fingerprints and kinship relationship among COVID-19 related coronaviruses that has not been described so far:

- First, using the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) of the NCBI, there were clearly identified three sequence genomic fingerprints of the COVID-19 related coronaviruses (Table 1): (*i*) at the beginning of the of the genome, also in the beginning of the *Orf1ab* RNA replicase gene, encoding the N-terminal Macro domain (ATP binding module); (*ii*) at the beginning of the S gene, encoding the N-terminal domain and receptor binding domain (RBD) of the spike glycoprotein; and (*iii*) at the end of the genome, the *NS8* gene itself. The sequences of the Bat SARS-like coronaviruses (Bat-SL-CoV) (isolates ZXC21 and ZC45, GenBank accession numbers MG772934.1 and MG772933.1, respectively) were the only natural DNA sequences that gave a perfect match with the *Orf1ab*- and *NS8* -gene fingerprints. On the other hand, the spike glycoprotein fingerprint was unmatched with any natural DNA sequence.

- Second, the phylogenetic analysis based on complete genomes showed that Bat-SL-CoV (ZXC21 and ZC45), Pangolin-CoV (isolate MP789, GenBank MT121216.1), BatCoV-RaTG13 (GenBank MN996532.1) and SARS-CoV-2 were orthologous. That is, they originated from a common ancestor coronavirus species, which were separated from each other after a speciation events. It is worth noting that not all pangolin coronaviruses were the same. Only the Panglolin-CoV (MP789) belonged to that group of orthologous coronavirus. Other pangolin coronaviruses were phylogenetically more distant, they consistently (bootstrap support 1000) grouped in a different cluster (Figure 1).

- Third, the phylogenetic analysis was based on spike glycoprotein sequences (specifically, from the N-terminal domain to the RBD, inclusive) showed a different pattern. The sequences of BatCoV-RaTG13 and SARS-CoV-2 diverged significantly (bootstrap support 1000) from Bat-SL-CoV (ZXC21 and ZC45) and Pangolin-CoV (MP789). Furthermore, the last two, were consistently grouped (bootstrap

support 1000) with other bat sequences. In this phylogenetic analysis, oOther pangolin coronaviruses were also consistently (bootstrap support 1000) grouped in another branch of the tree or cluster (Figure 2).

How and when did the BatCoV-RaTG13 appear?

BatCoV-RaTG13 has been considered likely to be the direct progenitor of SARS-CoV-2 (1), our results also showed the close phylogenetic relationship between both species, however, what is known about this bat coronavirus? The BatCoV-RaTG13 genome was identified and sequenced by Zheng-Li Shi and coworkers, 2020 (1), within the framework of the study of a "new coronavirus", which caused the epidemic of acute respiratory syndrome in humans in Wuhan, China (December 2019). Literally from (1): "We found that a short region of RNA-dependent RNA polymerase (RdRp) from a bat coronavirus (BatCoV RaTG13) —which was previously detected in *Rhinolophus affinis* from Yunnan province— showed high sequence identity to 2019-nCoV (SARS-CoV-2)". Then, Zheng-Li Shi and coworkers, 2020 (1), carried out full-length BatCoV-RaTG13 and SARS-CoV-2 genome sequencing, with an overall genome sequence identity of 96.2%. BatCoV-RaTG13 complete genome sequence was submitted to GenBank by Z.-L. Shi and coworkers on 27-JAN-2020. In the BatCoV-RaTG13 GenBank record, it appears as isolation source, fecal swab; as host, *Rhinolophus affinis*; as country, China; and as collection date, 24-Jul-2013. That is, the BatCoV RaTG13 sequence was sampled from a bat in Yunnan in 2013, years before SARS-CoV-2 was first identified, and the start of the pandemic.

BatCoV-RaTG13 spike glycoprotein: an phylogenetic incongruence

Regarding the divergence of BatCoV-RaTG13 and SARS-CoV-2 in the spike glycoprotein phylogeny (Figure 1 and Figure 2), it is worth mentioning that the RBD region is described as the most variable region of the coronavirus genomes (2). However, as it is shown in the multialignment (Figure 3), the N-terminal domain and RBD region is highly conserved, and many amino acid residues were strictly conserved. So that, according to Motoo Kimura's neutral theory of molecular evolution (3), in this so critical region of the coronavirus genomes the balance between the random drift and the functional constraints must be enormously stressed. Nonetheless, the SARS-CoV-2 sipke glycoprotein appears to be optimized for binding to the human receptor ACE2 (2,4). Of course it has! Through The RBD of the spike glycoprotein attaches the virion to the cell membrane by interacting with host receptor, initiating the infection (4,5).

However, What evolutionary event could explain the extraordinary optimization of the SARS-CoV-2 spike glycoprotein for human infection, as well as, the phylogenetic incongruence? From a molecular evolution perspective, considering that BatCoV-RaTG13 is the direct progenitor of SARS-CoV-2, we must place this evolutionary event in BatCoV-RaTG13. A possible recombination event between BatCoV-RaTG13 and other coronaviruses, or a horizontal gene transfer that replacing variable regions in BatCoV-RaTG13 *S* gene, could be considered (6,7). But, either in recombination or a horizontal transfer, there always had to be a donor, and we could not identify any donor by analysing the entire NCBI nucleotide collection. This reasoning agrees with Zheng-Li Shi and coworkers, 2020 (1), literally from (1): "Using the aligned genome sequences of 2019-nCoV, RaTG13, SARS-CoV and previously reported bat SARSr-CoVs, no evidence for recombination events was detected in the genome of 2019-nCoV (SARS-CoV-2)".

The mystery of the polybasic furin cleavage site in the SARS-CoV-2 spike glycoprotein

One small insertion of four aminoacids is present in the SARS-CoV-2 spike glycoprotein, but not found in that of BatCoV-RaTG13 protein. It is responsible of the high COVID-19 pathogenesis. This insertion is the known as polybasic furin cleavage site. Polybasic because there are basic amino acids, and It is a small region (site) of the spike glycoprotein that interacts with the furin, another membrane protein of human cellsl (a protease), which is the key in the infection process of SARS-CoV-2.

On the other hand, BatCoV-RaTG13 and SARS-CoV-2 also share other three small insertions in the N-terminal domain of the protein (Figure 3), in comparison with the same protein from the coronaviruses of its taxonomic group. Since the BatCoV-RaTG13 sequence was sampled in 2013, it is unlikely that both coronaviruses independently acquired identical insertions at three different locations in the protein (8). This, agrees that BatCoV-RaTG13 was the progenitor of SARS-CoV-2. It is considered that SARS-CoV-2 directly emerged from bat to human (1). Then, when did the insertion of the furin site occur? It must have occurred either in the host or during the transmission of emerging coronavirus, through recombination. However, recombination events were not detected in SARS-CoV-2 (1).

The polybasic furin cleavage site is a four aminoacid insertion “PRRA”, encoded by the inserted of 12 nucleotides (CCT CGG CGG GCA) in the conserved region of the S1/S2 cleavage site, At sequence level, the furin site is located at144 amino acid residues downstream the RBD (Figure 3). This site is typical of beta-coronaviruses of other lineages, but not of the lineage B, to which SARS-CoV-2 has been classified (9). As an example, below is a fragment of an amino acid multiple alignment of several spike glycoprotein sequences, from several beta-coronavirus lineages, and covering the respective polybasic furin sites (denoted in yellow). The presence of a doublet of Arginine is a distinctive structural feature (10):

```
Lineage A, Human coronavirus HKU1 (1495877059)                          YALPSSRRKRRGI    758
Lineage A, Murine hepatitis virus (ADI59790.1)                          YST--AHRARTSV    759
Lineage A, Human coronavirus OC43 (998640295)                           YSK--TRRSRRAI    766
Lineage B, SARS-CoV-2 (QHR63260.2)                                      YQTQTN-SPRRAR    685
Lineage B, SARS-CoV-2 (QII57208.1)                                      YQTQTN-SPRRAR    685
Lineage C, Middle East respiratory syndrome-related CoV (1453282535)    PDTPSTLTPRSVR    751
Lineage C, Pipistrellus bat coronavirus HKU5 (1386872228)               PPSPSARLARSAR    749
                                                                                  *
```

As important poin, the polybasic furin clevage site is responsible for the high SARS-CoV-2 infectivity and transmissibility (11). The furin site enhances cell-cell fusion and mediates membrane fusion. It is involved in infectious diseases and cancer (12), and now, also in COVID-19. It is through this site that the coronavirus has optimized for binding to the human receptor ACE2 to entry into human cells (13). The origin of the furin cleavage site it SARS-CoV-2 is puzzling.

Is there an intermediate host?

Initially, pangolins were considered to be the intermediate host (14). However, based on the present results and those of the literature, the molecular and phylogenetic analyses do not support that SARS-CoV-2 emerged directly from the pangolin coronaviruses (15). Recently, C.M. Freuling et al., 2020 (16) also make the hypothesis that Raccoon dogs (*Nyctereutes procyonoides*) might have been intermediate hosts for SARS-CoV-2, but they only show that these animals are susceptible to the coronavirus (16). It is a case of zoonosis, where probably SARS-CoV-2 jumped between human a non-human animal. Another case of zoonosis was that of Denmark that will kill 17 million minks to stop a new variant of the coronavirus that has jumped to humans.

So despite direct contact between humans and bats is limited, and an intermediate species often plays a role in the transmission of emerging viruses from bats to humans (17), however, it is not the case in SARS-CoV-2. Or at least, no intermediate host has been discovered to date.

Interface between biology and logic

The basic principles of biology are also fulfilled in the world of viruses. From the Cell Theory of Rudolf Virchow (1858), *Omnis cellula ex cellula* (each cell derived from another pre-existing cell), it could be inferred as "each virus derives from a pre-existing virus". The Natural Selection of Charles Darwin (1859) of change mutations in the struggle for existence is fully applicable to the viruses. The principle of Theodosius Dobzhansky (1973) "Nothing in Biology Makes Sense Except in the Light of Evolution" (18) becomes relevant

in the case of SARS-CoV-2 origin. From the current information available of the public genomic data bases, it is hard to fit it into a rational evolutionary model.

The taxonomic classification of the SARS2-CoV-2 is also confusing. To address this point, we considered a criterion of forensic genetics, which is based on the likelihood ratio (LR), and is valid in a trial. The LR compares two probabilities to find the same genotype or genetic markers from a sample, in two different persons or systems. This parameter is used to express the result of a DNA test. When passes judgement on convicting a suspect, it requires the probability that DNA of the evidence was from the suspect, must be much greater (more than billions, $> 10^{12}$) than the probability that would be from a random member of the population to which the suspect belongs. Thus, in forensic genetics, for the LR to serve as a prosecution witness, it must be enormously high.

In the taxonomic classification of the SARS2-CoV-2, there are reasonable doubts that it belongs to the taxonomic group, to which it is classified (lineage B of beta-coronavirus). Due the presence of the polybasic furin cleavage site, an LR is “infinite”. That is, the relationship between the probability (P1) that given the present SARS-CoV-2 genome sequence, it belongs to SARS-CoV-2, obviously, it is 1; and the probability (P2) that given the same SARS-CoV-2 genome sequence, it belongs to another coronavirus of the taxonomic group, at the moment it is 0 (because the polybasic furin site has only been found in SARS-CoV-2). So that LR that compares the two probabilities P1 and P2 is "infinity" (LR = P1/P2 = 1/0 = infinite). It is a probabilistic way of showing the difficulty of associating SARS-CoV-2 with lineage B of the beta-coronaviruses. In this sense, Z.-L. Shi and coworkers, 2020 (1), (not taking into account the furin site), also pointed out this doubt. Literally from (1): "Phylogenetic analysis of the full-length genome and the gene sequences of RdRp and spike (S) showed that—for all sequences—RaTG13 is the closest relative of 2019-nCoV (SARS-CoV-2) and they form a distinct lineage from other SARSr-CoVs".

Since BatCoV-RaTG13 is the probable origin of SARS-CoV-2 (1), it is surprising that after almost a year of the COVID-19 pandemic, no more BatCoV-RaTG13 species have been isolated. Bat is the most studied natural reservoir of viruses. On the other hand, bat coronavirus strains could be isolated from fecal swabs (no bat nasal swab is needed!), and their complete genomes could be submitted to GenBank. Currently there is only one available genome of BatCoV-RaTG13. More complete genomes of BatCoV-RaTG13 are required to validate statistically the 96.2% of genome identity with SARS-CoV-2. there is still 3.8% of difference between them, which may hold the key to explain the presence of the polybasic furin site in the SARS-CoV-2 spike glycoprotein. In contrast, there are now thousands of available complete genomes of SARS-CoV-2.

In biological evolution there are always "missing links", however, based on the discovery of new fossil remains, they could be resolved. In the world of viruses there are no fossil remains, but “missing links” could be solved with the discovery of new viruses. The presence of the furin site in SARS-CoV-2 (in the most conserved part of the protein, far from the most variable region) indicates one “missing link” in our understanding of its evolutionary process, and its origin. This calls into question whether BatCoV-RaTG13 is the direct progenitor of SARS-CoV-2. As in forensic genetics, which also applies in the world of viruses, when a genetic marker fails a DNA test, it is enough to rule out guilt. So, there is a reasonable doubt about SARS-CoV-2 origin, which could be "A New Paradigm for Virology". The availability of new BatCoV-RaTG13 genomes is essential to provide insight. Otherwise, SARS-CoV-2 origin is a dark side of COVID-19. The doubt may lead to concern that it could be a laboratory construct or a purposefully manipulated virus. The right technology for this is available. For scientific and medical purposes, several synthetic constructs of SARS-CoV-2 genome exist in GenBank (MT108784.1, MT461669.1, MT461671.1, MT461670.1). For therapeutic purposes, in 2008 synthetic recombinant bat SARS-like coronavirus was created, and was infectious in cultured cells and in mice (19). In this sense, the debate must go on (20).

As a reason for hope, for probability, and because the "trial-error" mechanism of the biological evolution, a human virus like SARS-CoV-2, will not appear again from Mother Nature, for a long, long time (nobody gets the lottery jackpot twice). Finally, there is no other alternative but to wait for an upcoming vaccine and/or specific therapeutic drugs available for COVID-19.

References

1. Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, Hui-Dong Chen, Jing Chen, Yun Luo, Hua Guo, Ren-Di Jiang, Mei-Qin Liu, Ying Chen, Xu-Rui Shen, Xi Wang, Xiao-Shuang Zheng, Kai Zhao , Quan-Jiao Chen, Fei Deng, Lin-Lin Liu, Bing Yan, Fa-Xian Zhan, Yan-Yi Wang, Geng-Fu Xiao, Zheng-Li Shi. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270–273, 2020. PMID: 32015507. doi: 10.1038/s41586-020-2012-7.

2 . Kristian G Andersen, Andrew Rambaut, W Ian Lipkin, Edward C Holmes, Robert F Garry. The proximal origin of SARS-CoV-2. Nat. Med. 26:450-452, 2020. PMID: 32284615. doi: 10.1038/s41591-020-0820-9.

3. Kimura, M. The neutral theory of molecular evolution. Cambridge University, Cambridge. UK (1983).

4. Yushun Wan, Jian Shang, Rachel Graham, Ralph S Baric, Fang Li. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. J. Virol. 94, e00127-20, 2020.PMID: 31996437. doi.org/10.1128/JVI.00127-20.

5. UNIPROT. P59594, Spike glycoprotein. Accessed October 09, 2020. https://www.uniprot.org/uniprot/P59594.

6 Dong-Sheng Chen, Yi-Quan Wu, Wei Zhang, San-Jie Jiang, Shan-Ze Chen. Horizontal gene transfer events reshape the global landscape of arm race between viruses and homo sapiens. Sci. Rep. 6:26934, 2016. PMID: 27270140. doi: 10.1038/srep26934.

7. Shahana S Malik, Syeda Azem-E-Zahra, Kyung Mo Kim, Gustavo Caetano-Anollés, Arshan Nasir. Do Viruses Exchange Genes across Superkingdoms of Life? Front. Microbiol. 8, 2110, 2017. PMID: 29163404. doi.org/10.3389/fmicb.2017.02110.

8. Xiaojun Li, Elena E Giorgi, Manukumar Honnayakanahalli Marichannegowda, Brian Foley, Chuan Xiao, Xiang-Peng Kong, Yue Chen, S Gnanakaran, Bette Korber, Feng Gao. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. 6(27):eabb9153. PMID: 32937441. doi: 10.1126/sciadv.abb9153.

9. Javier A Jaimes, Nicole M André, Joshua S Chappie, Jean K Millet, Gary R Whittaker. Phylogenetic Analysis and Structural Modeling of SARS-CoV-2 Spike Protein Reveals an Evolutionary Distinct and Proteolytically Sensitive Activation Loop. J. Mol. Biol. 432:3309-3325, 2020. PMID: 32320687. doi: 10.1016/j.jmb.2020.04.009.

10. Imène Kara, Marjorie Poggi, Bernadette Bonardo, Roland Govers, Jean-François Landrier, Sun Tian, Ingo Leibiger, Robert Day, John W M Creemers, Franck Peiretti. The Paired Basic Amino Acid-cleaving Enzyme 4 (PACE4) Is Involved in the Maturation of Insulin Receptor Isoform B. J. Biol. Chem. 290:2812–2821. PMID: 25527501. doi: 10.1074/jbc.M114.592543.

11. Shuai Xia, Qiaoshuai Lan, Shan Su, Xinling Wang, Wei Xu, Zezhong Liu, Yun Zhu, Qian Wang, Lu Lu, Shibo Jiang. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. Signal Transduct. Target Ther. 5:92, 2020. PMID: 32532959. doi.org/10.1038/s41392-020-0184-0.

12. Elisabeth Braun, Daniel Sauter. Furin-mediated protein processing in infectious diseases and cancer. Clin. Transl. Immunol. E1073, 2019. PMID: 31406574. doi.org/10.1002/cti2.1073.

13. Markus Hoffmann, Hannah Kleine-Weber, Stefan Pöhlmann. Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. Mol. Cell 78:779–784, 2020. PMID: 32362314. doi: 10.1016/j.molcel.2020.04.022.

14. Tao Zhang, Qunfu Wu, Zhigang Zhang. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. Curr. Biol. 30:1346–1351.e2, 2020. PMID: 32315626. doi: 10.1016/j.cub.2020.03.022.

15. Ping Liu, Jing-Zhe Jiang, Xiu-Feng Wan, Yan Hua, Linmiao Li, Jiabin Zhou, Xiaohu Wang, Fanghui Hou, Jing Chen, Jiejian Zou, Jinping Chen. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog. 16(5):e1008421, 2020. PMID: 32407364. doi: 10.1371/journal.ppat.1008421.

16. Conrad M. Freuling, Angele Breithaupt, Thomas Müller, Julia Sehl, Anne Balkema-Buschmann, Melanie Rissmann, Antonia Klein, Claudia Wylezich, Dirk Höper, Kerstin Wernike, Andrea Aebischer, Donata Hoffmann, Virginia Friedrichs, Anca Dorhoi, Martin H. Groschup, Martin Beer, Thomas C. Mettenleiter. Susceptibility of Raccoon Dogs for Experimental SARS-CoV-2 Infection. Emerg Infect Dis. 2020 Oct 22;26(12), 2020. PMID: 33089771. doi: 10.3201/eid2612.203733.

17. Shauna Milne-Price, Kerri L Miazgowicz, Vincent J Munster. The emergence of the Middle East Respiratory Syndrome coronavirus. Pathog. Dis. 71:21-176, 2014. PMID: 24585737. doi: 10.1111/2049-632X.12166.

18. Theodosius Dobzhansky. Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher 35:125-129, 1973.

19. Michelle M Becker , Rachel L Graham, Eric F Donaldson, Barry Rockx, Amy C Sims, Timothy Sheahan, Raymond J Pickles, Davide Corti, Robert E Johnston, Ralph S Baric, Mark R Denison. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc. Natl. Acad. Sci. USA 105:19944–19949, 2008. PMID: 19036930. doi: 10.1073/pnas.0808116105.

20. Heidi J Larson. A lack of information can become misinformation. Nature 580:306, 2020. PMID: 32231320. doi: 10.1038/d41586-020-00920-w.

21. Muhamad Fahmi, Yukihiko Kubota, Masahiro Ito. Nonstructural proteins NS7b and NS8 are likely to be phylogenetically associated with evolution of 2019-nCoV. Infect. Genet. Evol. 81:104272, 2020. PMID: 32142938. doi.org/10.1016/j.meegid.2020.104272.

22. Fábio Madeira, Young Mi Park, Joon Lee, Nicola Buso, Tamer Gur, Nandana Madhusoodanan, Prasad Basutkar, Adrian R N Tivey, Simon C Potter, Robert D Finn, Rodrigo Lopez. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucl. Acids Res. 47(W1):W636-W641, 2019. PMID: 30976793. doi: 10.1093/nar/gkz268.

23. Ivica Letunic, Peer Bork. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucl. Acids Res. 2011, Vol. 39(W475–W478), 2011. PMID: 21470960. doi:10.1093/nar/gkr201.

24. Jun Lan, Jiwan Ge, Jinfang Yu, Sisi Shan, Huan Zhou, Shilong Fan, Qi Zhang, Xuanling Shi, Qisheng Wang, Linqi Zhang, Xinquan Wang. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581:215-220, 2020. PMID: 32225176. doi:10.1038/s41586-020-2180-5.

Acknowledgements

This work has not been awarded grants by any research-supporting institution

Competing interest declaration

All authors declare that they have no conflicts of interest.

Table 1. SARS-CoV-2, BatCoV-RaTG13, Pangolin-CoVs and Bat-SL-CoV genomic fingerprints

Virus species	Genome reference	Gene	Genome coordinates	Protein gene product. GenBank id	Protein position
SARS-CoV-2	GenBank MN996528.1	*orf1ab*	1940 - 3955	Orf1ab polyprotein. QHR63259.1	559 - 1230
	isolate WIV04	*S*	21563 - 22963	Spike glycoprotein. QHR63260.2 (RBD)	1 - 467
		NS8	27912 - 28256	Nonstructural protein NS8. QHR63267.1	7 - 121
BatCoV-RaTG13	GenBank MN996532.1	*orf1ab*	1925 - 3937	Orf1ab polyprotein. QHR63299.1	559 - 1229
		S	21545 - 22945	Spike glycoprotein. QHR63300.2 (RBD)	1- 467
		NS8	27872 - 28222	Nonstructural protein NS8. QHR63307.1	5 - 121
Pangolin-CoV	GenBank MT040335.1	*orf1ab*	2220 - 3884	Orf1ab polyprotein. QIA48631.1	652 - 1206
	isolate PCoV_GX-P5L	*S*	21540 - 22940	Spike glycoprotein. QIA48632.1 (RBD)	1 - 467
		orf8	27875 - 28210	Orf8 protein. QIA48638.1	9 - 121
Pangolin-CoV	GenBank MT121216.1	*orf1ab*	2102 - 3814	Orf1ab polyprotein. QIG55944.1	653 - 1223
	isolate MP789	*S*	21421 -22821	Spike glycoprotein. QIG55945.1 (RBD)	1 - 467
		orf8	27728 - 28042	Orf8 protein. QIG55952.1	1 - 105
Bat-SL-CoV	GenBank MG772933.1	*1ab*	2218 - 3948	non-structural polyprotein 1ab.AVP78030.1	652 - 1228
	isolate Bat-SL-CoVZC45	*S*	21549 - 22949	Spike protein. AVP78031.1 (RBD)	1 - 467
		10b	27799 - 28161	Hypothetical protein. AVP78037.1	1 - 121

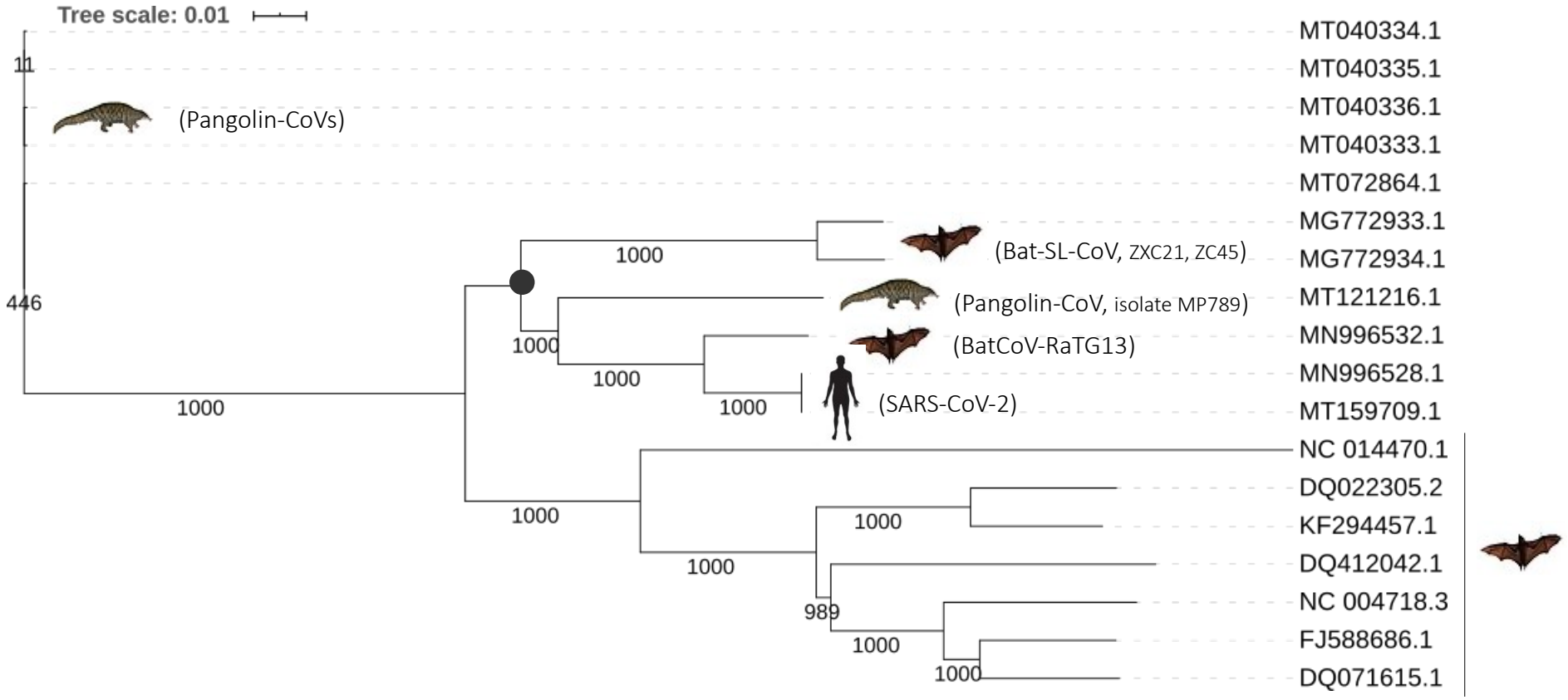
Tree scale: 0.01
11
(Pangolin-CoVs)
446
1000
1000
1000
1000
(Bat-SL-CoV, ZXC21, ZC45)
(Pangolin-CoV, isolate MP789)
(BatCoV-RaTG13)
(SARS-CoV-2)
1000
1000
1000
989
1000
1000
MT040334.1
MT040335.1
MT040336.1
MT040333.1
MT072864.1
MG772933.1
MG772934.1
MT121216.1
MN996532.1
MN996528.1
MT159709.1
NC 014470.1
DQ022305.2
KF294457.1
DQ412042.1
NC 004718.3
FJ588686.1
DQ071615.1

Figure 1

Figure 1. Phylogenetic tree based on coronavirus complete genome

Phylogenetic tree based on a multialignment of complete genomes of selected coronaviruses. The sample includes the BatCoV-RaTG13, the available pangolin coronaviruses, a selection of bat coronaviruses extracted from the literature (1,14,21), and two SARS-CoV-2 genomes as representatives of the NCBI "Severe acute respiratory syndrome coronavirus 2 (taxid:2697049)" taxonomic group. The phylogenetic tree was constructed with the Neighbor Joining method of the Clustal Omega (v.1.2.4) software package using default parameters (22), and iTol Interactive Tree Of Life tool (23). Assessed clustering strength was calculated by bootstrap using 1000 replicates. Tree scale bar stands for the evolutionary distance, based on genome multialigment. The black point depicts the ancestral coronavirus species from which the group COVID-19 related coronaviruses separated from each other after a speciation event. GenBank accession number of the complete genomes and the coronavirus were the following (in the same arrangement as in the phylogenetic tree): MT040333.1 to MT040336.1, Pangolin coronavirus; MG772933.1 and MG772934.1, Bat SARS-like coronavirus; MT121216.1 (isolate MP789) Pangolin coronavirus; MN996532.1, BatCoV-RaTG13; MN996528.1 and MT159709.1, SARS-CoV-2; NC_014470.1, Bat coronavirus BM48-31/BGR/2008; DQ022305.2, Bat SARS coronavirus HKU3-1; KF294457.1,Bat SARS-like coronavirus; DQ412042.1, Bat SARS CoV Rf1/2004; NC_004718.3, SARS coronavirus Tor2; FJ588686.1, SARS coronavirus Rs_672/2006; DQ071615.1, Bat SARS CoV Rp3/2004.

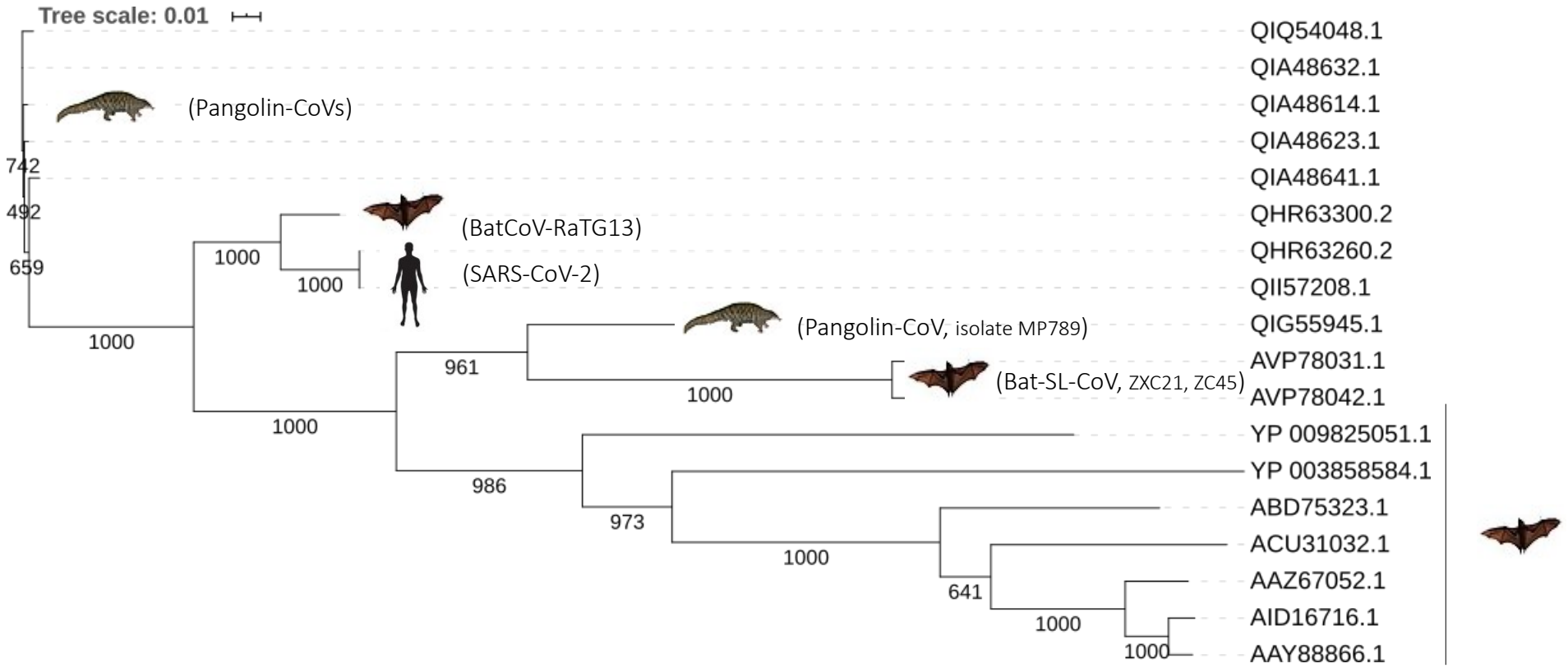
Tree scale: 0.01
(Pangolin-CoVs)
(BatCoV-RaTG13)
(SARS-CoV-2)
(Pangolin-CoV, isolate MP789)
(Bat-SL-CoV, ZXC21, ZC45)
QIQ54048.1
QIA48632.1
QIA48614.1
QIA48623.1
QIA48641.1
QHR63300.2
QHR63260.2
QII57208.1
QIG55945.1
AVP78031.1
AVP78042.1
YP 009825051.1
YP 003858584.1
ABD75323.1
ACU31032.1
AAZ67052.1
AID16716.1
AAY88866.1
742
492
659
1000
1000
1000
961
1000
1000
986
973
1000
641
1000
1000

Figure 2

Figure 2. Phylogenetic tree based on coronavirus spike glycoprotein

Phylogenetic tree based on the multialignment of the most variable region of the coronavirus spike glycoprotein: from the N-terminal domain to the RBD, inclusive. In this phylogenetic analysis, we used the sequences of the same coronaviruses of Figure 1, with the particular intention of comparing the phylogenetic behaviour, when it is based on the complete genome and when it is based on this strategic region of the genome. RBD position was based on (24). The phylogenetic tree was constructed with the Neighbor Joining method of the Clustal Omega (v.1.2.4) software package, using default parameters (22), and iTol Interactive Tree Of Life tool (23). Assesed clustering strength was calculated by bootstrap using 1000 replicates. Tree scale bar stands for the evolutionary distance, based on sequence multialigmnet. GenBank accession number of the spike glycoprotein and the coronavirus were the following (in the same arrangement as in the phylogenetic tree): QIA48641.1, QIA48632.1, QIA48614.1, QIA48623.1, QIQ54048.1, Pangolin coronavirus; QHR63300.2, Bat coronavirus RaTG13; QHR63260.2, QII57208.1, SARS-CoV-2; QIG55945.1, Pangolin coronavirus; AVP78031.1, AVP78042.1, Bat SARS-like coronavirus; YP_009825051.1, SARS coronavirus Tor2; YP_003858584.1, Bat coronavirus BM48-31/BGR/2008; ABD75323.1, Bat SARS CoV Rf1/2004; ACU31032.1, SARS coronavirus Rs_672/2006; AAZ67052.1, Bat SARS CoV Rp3/2004; AID16716.1, Bat SARS-like coronavirus; AAY88866.1, Bat SARS coronavirus HKU3-1.

```
ADC35483.1                -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
sp|P59594|SPIKE_CVHSA     -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
AAR07630.1                -MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLF  59
AVP78031.1                MLFFLFLQFALVNS----QCVNLTGRTPLNPNYTNSSQRGVYYPDTIYRSDTLVLSQGYF  56
AVP78042.1                MLFFLFLQFALVNS----QC-DLTGRTPLNPNYTNSSQRGVYYPDTIYRSDTLVLSQGYF  55
QIG55945.1                MLFFFFLHFALVNS----QCVNLTGRAAIQPSFTNSSQRGVYYPDTIFRSNTLVLSQGYF  56
QIQ54048.1                -MFVFLFVLPLVSS----QCVNLTTRTGIPPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QIA48614.1                -MFVFLFVLPLVSS----QCVNLTTRTGIPPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QIA48623.1                -MFVFLFVLPLVSS----QCVNLTTRTGIQPGYTNSSTRGVYYPDKVFRSSILHLTQDLF  55
QHR63260.2                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QII57208.1                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QIA98554.1                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLF  55
QHR63300.2                -MFVFLVLLPLVSS----QCVNLTTRTQLPPAYTNSSTRGVYYPDKVFRSSVLHLTQDLF  55
                           :*.::. : *...    :*  :           .*  ******* ::**. *  :*. *

ADC35483.1                LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
sp|P59594|SPIKE_CVHSA     LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
AAR07630.1                LPFYSNVTGFHTIN-------HTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQ  112
AVP78031.1                LPFYSNVSWYYSLTTN-NAATKRTDNPILDFKDGIYFAATEHSNIIRGWIFGTTLDNTSQ  115
AVP78042.1                LPFYSNVSWYYSLTTN-NAATKRTDNPILDFKDGIYFAATEHSNIVRGWIFGTTLDNTSQ  114
QIG55945.1                LPFYSNVSWYYALTKT-NSAEKRVDNPVLDFKDGIYFAATEKSNIVRGWIFGTTLDNTSQ  115
QIQ54048.1                LPFFSNVTWFNTIHLNYQGGFKKFDNPVLPXNDGVYFASTEKSNIIRGWIFGTTLDARTQ  115
QIA48614.1                LPFFSNVTWFNTI--NYQGGFKKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDARTQ  113
QIA48623.1                LPFFSNVTWFNTI--NYQGGFKKFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDARTQ  113
QHR63260.2                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QII57208.1                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QIA98554.1                LPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
QHR63300.2                LPFFSNVTWFHAIHVSGTNGIKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQ  115
                          ***:***: : ::        :  .**::  :**:***:**:**::***:**:*::  :*

ADC35483.1                SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
sp|P59594|SPIKE_CVHSA     SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
AAR07630.1                SVIIINNSTNVVIRACNFELCDNPFFAVSKPMGT----QTHTMIFDNAFNCTFEYISDAF  168
AVP78031.1                SLLIVNNATNVIIKVCNFDFCYDPYLSGYYHN-NKTWSIREFAVYSSYANCTFEYVSKSF  174
AVP78042.1                SLLIVNNATNVIIKVCNFDFCYDPYLSGYYHN-NKTWSIREFAVYSFYANCTFEYVSKSF  173
QIG55945.1                SLLIVNNATNVIIKVCNFQFCYDPYLSGYYHN-NKTWSTREFAVYSSYANCTFEYVSKSF  174
QIQ54048.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  175
QIA48614.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  173
QIA48623.1                SLLIVNNATNVVIKVCEFQFCTDPFLGVYYHNNNKTWVENEFRVYSSANNCTFEYISQPF  173
QHR63260.2                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QII57208.1                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QIA98554.1                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
QHR63300.2                SLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPF  175
                          *::*:**:***:*:.*:*::* :*::.      .      .  ::.   ******:*. *

ADC35483.1                SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
sp|P59594|SPIKE_CVHSA     SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
AAR07630.1                SLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINI  228
AVP78031.1                MLNISGNGGLFNTLREFVFRNVDGHFKIYSKFTPVNLNRGLPTGLSVLQPLVELPVSINI  234
AVP78042.1                MLNISGNGGLFNTLREFVFRNVDGHFKIYSKFTPVNLNRGLPTGLSVLQPLVELPVSINI  233
QIG55945.1                MLDIAGKSGLFDTLREFVFRNVDGYFKIYSKYTPVNVNSNLPIGFSALEPLVEIPAGINI  234
QIQ54048.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  235
QIA48614.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  233
QIA48623.1                LMDLEGKQGNFKNLREFVFKNVDGYFKIYSKHTPIDLVRDLPRGFAALEPLVDLPIGINI  233
QHR63260.2                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QII57208.1                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QIA98554.1                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINI  235
QHR63300.2                LMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPPGFSALEPLVDLPIGINI  235
                           :::  : * *. ******:* **.: :*. . *:::  .** *: .*:*:..:* .***

ADC35483.1                TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
sp|P59594|SPIKE_CVHSA     TNFRAILTAFS------PAQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
AAR07630.1                TNFRAILTAFS------PAQDTWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNP  282
AVP78031.1                TKFRTLLTIHRGDPMP---NNGWTAFSAAYFVGYLKPRTFMLKYNENGTITDAVDCALDP  291
AVP78042.1                TKFRTLLTIHRGDPMS---NNGWTAFSAAYFVGYLKPRTFMLKYNENGTITDAVDCALDP  290
QIG55945.1                TKFRTLLTIHRGDPMP---NNGWTVFSAAYYVGYLAPRTFMLNYNENGTITDAVDCALDP  291
QIQ54048.1                TRFQTLLALHRSYLTPGKLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  295
QIA48614.1                TRFQTLLALHRSYLTPGNLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  293
QIA48623.1                TRFQTLLALHRSYLTPGKLESGWTTGAAAYYVGYLQQRTFLLSYNQNGTITDAVDCSLDP  293
QHR63260.2                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QII57208.1                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QIA98554.1                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
QHR63300.2                TRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDP  295
                          *.*:::*: .         .. * . :***:****   **:*.*::**********: :*

ADC35483.1                LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
sp|P59594|SPIKE_CVHSA     LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
AAR07630.1                LAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWER  342
AVP78031.1                LSETKCTLKSLTVQKGIYQTSNFRVQPTQSVVRFPNITNVCPFHKVFNATRFPSVYAWER  351
AVP78042.1                LSETKCTLKSLSVQKGIYQTSNFRVQPTQSIVRFPNITNVCPFHKVFNATRFPSVYAWER  350
QIG55945.1                LSEAKCTLKSLTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATTFASVYAWNR  351
QIQ54048.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  355
QIA48614.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  353
QIA48623.1                LSETKCTLKSLTVEKGIYQTSNFRVQPTISIVRFPNITNLCPFGEVFNASKFASVYAWNR  353
QHR63260.2                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QII57208.1                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QIA98554.1                LSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNR  355
QHR63300.2                LSETKCTLKSFTVEKGIYQTSNFRVQPTDSIVRFPNITNLCPFGEVFNATTFASVYAWNR  355
                          *:* **::**: ::*********** *: .:********:*** :****: * *****:*
```

Figure 3

```
ADC35483.1                KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
sp|P59594|SPIKE_CVHSA     KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
AAR07630.1                KKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQT  402
AVP78031.1                TKISDCIADYTVFYNSTSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGQT  411
AVP78042.1                TKISDCIADYTVFYNSTSFSTFKCYGVSPSKLIDLCFTSVYADTFLIRFSEVRQVAPGQT  410
QIG55945.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVRGDEVRQIAPGQT  411
QIQ54048.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  415
QIA48614.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  413
QIA48623.1                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVVKGDEVRQIAPGQT  413
QHR63260.2                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QII57208.1                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QIA98554.1                KRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT  415
QHR63300.2                KRISNCVADYSVLYNSTSFSTFKCYGVSPTKLNDLCFTNVYADSFVITGDEVRQIAPGQT  415
                          .:**:*:***:*:***: ********** :** ****:.****:*::  .:***:*****

ADC35483.1                GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
sp|P59594|SPIKE_CVHSA     GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
AAR07630.1                GVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSP  462
AVP78031.1                GVIADYNYKLPDDFTGCVIAWNTAKQDV-----GNYFYRSHRSTKLKPFERDLSSDEN--  464
AVP78042.1                GVIADYNYKLPDDFTGCVIAWNTAKQDT-----GHYFYRSHRSTKLKPFERDLSSDEN--  463
QIG55945.1                GRIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  471
QIQ54048.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNYGYLYRLFRKSKLKPFERDISTEIYQA  475
QIA48614.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNYGYLYRLFRKSKLKPFERDISTEIYQA  473
QIA48623.1                GVIADYNYKLPDDFTGCVIAWNSVKQDALTGGNY--LYRLFRKSKLKPFERDISTEIYQA  471
QHR63260.2                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QII57208.1                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QIA98554.1                GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQA  475
QHR63300.2                GKIADYNYKLPDDFTGCVIAWNSKHIDAKEGGNFNYLYRLFRKANLKPFERDISTEIYQA  475
                          * ************ ***:***: : *          **  *  :*:*****:*.

ADC35483.1                DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
sp|P59594|SPIKE_CVHSA     DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
AAR07630.1                DGKPCT-PPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIK  521
AVP78031.1                ------G------VRTLSTYDFNPNVPLEYQATRVVVLSFELLNAPATVCGPKLSTQLVK  512
AVP78042.1                ------G------VRTLSTYDFNPNVPLEYQATRVVVLSFELLNAPATVCGPKLSTQLVK  511
QIG55945.1                GSTPCNGVEGFNCYFPLQSYGFHPTNGVGYQPYRVVVLSFELLKAPATVCGPKQSTNLVK  531
QIQ54048.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSXELLNGPATVCGPKLSTTLVK  535
QIA48614.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSFELLNGPATVCGPKLSTTLVK  533
QIA48623.1                GSTPCNGQVGLNCYYPLERYGFHPTTGVNYQPFRVVVLSFELLNGPATVCGPKLSTTLVK  531
QHR63260.2                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QII57208.1                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QIA98554.1                GSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVK  535
QHR63300.2                GSKPCNGQTGLNCYYPLYRYGFYPTDGVGHQPYRVVVLSFELLNAPATVCGPKKSTNLVK  535
                                     *  *.*  .  : :*  ****** ***:.******** ** *:*

ADC35483.1                NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
sp|P59594|SPIKE_CVHSA     NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
AAR07630.1                NQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGV  581
AVP78031.1                NQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGKDASDFIDSVRDPQTLEILDITPCSFGGV  572
AVP78042.1                NQCVNFNFNGLKGTGVLTDSSKRFQSFQQFGKDASDFIDSVRDPQTLEILDITPCSFGGV  571
QIG55945.1                NKCVNFNFNGLTGTGVLTESSKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  591
QIQ54048.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  595
QIA48614.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  593
QIA48623.1                DKCVNFNFNGLTGTGVLTTSKKQFLPFQQFGRDISDTTDAVRDPQTLEILDITPCSFGGV  591
QHR63260.2                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QII57208.1                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QIA98554.1                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
QHR63300.2                NKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGV  595
                          ::**********.****** *.*:*  *****:* :*  *:****:* *****:*******

ADC35483.1                SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
sp|P59594|SPIKE_CVHSA     SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
AAR07630.1                SVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEH  641
AVP78031.1                SVITPGTNTSLEVAVLYQDVNCTDVPTTIHADQLTPAWRIYATGTNVFQTQAGCLIGAEH  632
AVP78042.1                SVITPGTNTSSEVAVLYQDVNCTDVPTTIHADQLTPAWRIYAIGTSVFQTQAGCLIGAEH  631
QIG55945.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  651
QIQ54048.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  655
QIA48614.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  653
QIA48623.1                SVITPGTNTSNQVAVLYQDVNCTEVPMAIHAEQLTPAWRVYSAGANVFQTRAGCLVGAEH  651
QHR63260.2                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QII57208.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QIA98554.1                SVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
QHR63300.2                SVITPGTNASNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEH  655
                          ********:* :***********:*  :***:****:**:*: * .****:****:****

ADC35483.1                VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
sp|P59594|SPIKE_CVHSA     VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
AAR07630.1                VDTSYECDIPIGAGICASYHTVSL----LRSTSQKSIVAYTMSLGADSSIAYSNNTIAIP  697
AVP78031.1                VNASYECDIPIGAGICASYHTASI----LRSTSQKAIVAYTMSLGAENSIAYANNSIAIP  688
AVP78042.1                VNASYECDIPIGAGICASYHTASI----LRSTGQKAIVAYTMSLGAENSIAYANNSIAIP  687
QIG55945.1                VNNTYECDIPIGAGICASYQTQTN----SRSVSSQAIIAYTMSLGAENSVAYANNSIAIP  707
QIQ54048.1                VNNSYECDIPVGAGICASYHSMS----SFRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  711
QIA48614.1                VNNSYECDIPVGAGICASYHSMS----SLRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  709
QIA48623.1                VNNSYECDIPVGAGICASYHSMS----SLRSVNQRSIIAYTMSLGAENSVAYSNNSIAIP  707
QHR63260.2                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QII57208.1                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QIA98554.1                VNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIP  715
QHR63300.2                VNNSYECDIPIGAGICASYQTQTNS----RSVASQSIIAYTMSLGAENSVAYSNNSIAIP  711
                          *: :******:********:: :      **. .::*:********:.*:**:**:****
```

Figure 3 (continued)

Figure 3. Coronavirus spike glycoprotein multiple sequence alignment.

Spike glycoprotein multiple sequence alignment. Groups of sequence sample: (*i*) human SARS coronavirus (SARS-CoV) (1255 amoniacids length); (*ii*) Bat-SL-CoV (1245); (*iii*) Pangolin-CoVs (1265-1269); (*iv*) SARS-CoV-2 (1273); and (*v*) BatCoV-RaTG13 (1269). To better visualize the characteristics of each group, there were three representative sequences of each. The protein alignments were created by Clustal Omega (v.1.2.4) using default parameters (22). Strictly conserved amino acids are denoted by *, gaps are denoted by -. The position of the amino acids in each sequence is indicated by the numbers to the right. The characteristic deletion and insertions of the COVID-19 related sequences, the RBD, and the SARS-CoV-2 polybasic furin cleavage site are highlighted in yellow. The color of the bands (gray-white) are intended to highlight the sequence characteristics of: SARS-CoV; Bat-SL-CoV and Pangolin-CoV (MP789); other Pangolin-CoVs; SARS-CoV-2; and BatCoV-RaTG13. The RBD position was based on (24). The figure only show up to the furin cleavage site. Up to the C-terminal end, most positions were strictly conserved. GenBank accession number of the spike glycoprotein sequences and the respective coronavirus are the following: ADC35483.1, SARS coronavirus HKU-39849; sp|P59594| (UNIPROT SPIKE_CVHSA); AAR07630.1, SARS coronavirus BJ302; AVP78031.1, Bat-SL-CoV ZC45; AVP78042.1, Bat-SL-CoV ZXC21; QIG55945.1, PangolinCoV, MP789; QIQ54048.1, Pang-CoV,GX-P2V; QIA48614.1, Pang-CoV,GX-P4L; QIA48623.1, Pang-CoV,GX-P1E; QHR63260.2, SARS-CoV-2; QII57208.1 SARS-CoV-2; QIA98554.1, SARS-CoV-2; QHR63300.2, BatCoV-RaTG13.