

Estructura i genoma del SARS-CoV-2

Pau Clavell Revelles

Aquest article és una edició d'un fil de Twitter publicat originalment a @cienciaoberta.

Ciència Oberta és un projecte de divulgació científica en català iniciat per 5 estudiants de la Facultat de Biologia de la UB, en el qual actualment hi participen més de 15 persones. Al web, cienciaoberta.cat, hi podràs trobar un **reportatge** nou setmanalment, a més d'altres seccions com els **contes científics** o les **experiències**. També pots seguir el projecte a les xarxes socials, tant a [Twitter](https://twitter.com/cienciaoberta) com a [Instagram](https://www.instagram.com/cienciaoberta) (@cienciaoberta), on també es crea contingut divers i atractiu diàriament!



Introducció

Quan els historiadors estudiïn la dècada que encetem, forçosament hauran de parar atenció a un virus que no només ha marcat la nostra vida quotidiana, sinó que també ens haurà afectat socialment, políticament, econòmicament, etc. Per aquest motiu, a nosaltres, que ens ha tocat viure aquesta realitat, hem d'entendre amb relativa profunditat alguns trets generals sobre l'estructura i genoma del SARS-CoV-2.

Més de cent mil publicacions a febrer del 2021 són el resultat d'un esforç titànic al qual s'ha abocat tota la comunitat científica, un esforç per identificar i caracteritzar un virus en temps rècord.

Estructura del coronavirus SARS-CoV-2

El SARS-CoV-2 és un dels molts coronavirus que existeixen. El seu material genètic és l'RNA i està empaquetat al voltant de les proteïnes N (nucleocàpside). Tot això s'envolta d'una doble capa de lípids -obtinguda de la cèl·lula on s'han generat- que integra 3 proteïnes estructurals diferents: l'espícula (S), l'embolcall (E) i la membrana (M), figura 1.

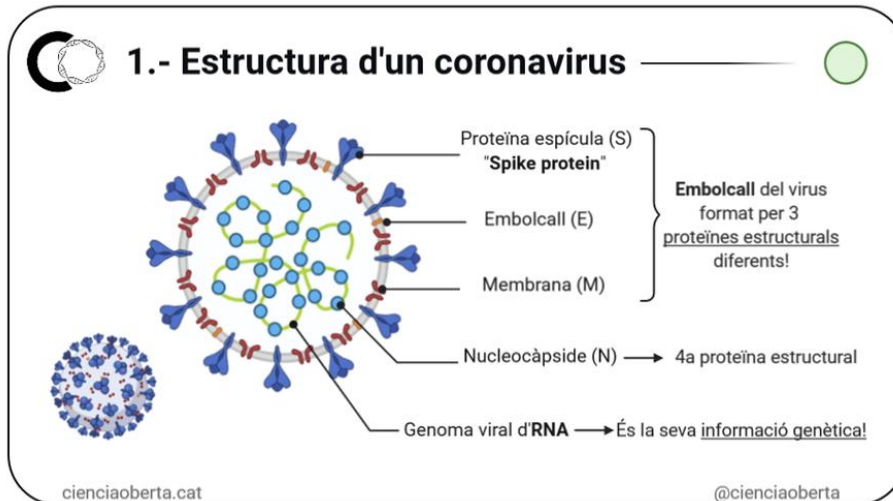


Figura 1. Esquema de l'estructura d'un coronavirus. Imatge generada a <https://biorender.com/>

La proteïna espícula

La proteïna de l'espícula és coneguda com a *spike protein* i, en realitat és un conjunt de 3 proteïnes S acompanyades de sucres, és a dir, altament glicosilades, fet molt important en la immunitat. Aquesta proteïna homotrimerica és importantíssima perquè encaixa amb el receptor ACE2 (enzim conversiu d'angiotensina 2) de les cèl·lules pulmonars -que també podem trobar a altres teixits com el ronyó-. Aquest reconeixement o encaix fa que ACE2 esdevingui la porta d'entrada per infectar-nos.

Cadascuna de les tres unitats que formen la proteïna espícula presenta dues regions importants: la S2 que ancora la proteïna al virus i la S1 que conté el Domini d'Unió al Receptor (RBD). L'RBD és concretament la part de la proteïna que permet el reconeixement entre el receptor ACE2 humà i el virus, figura 2.

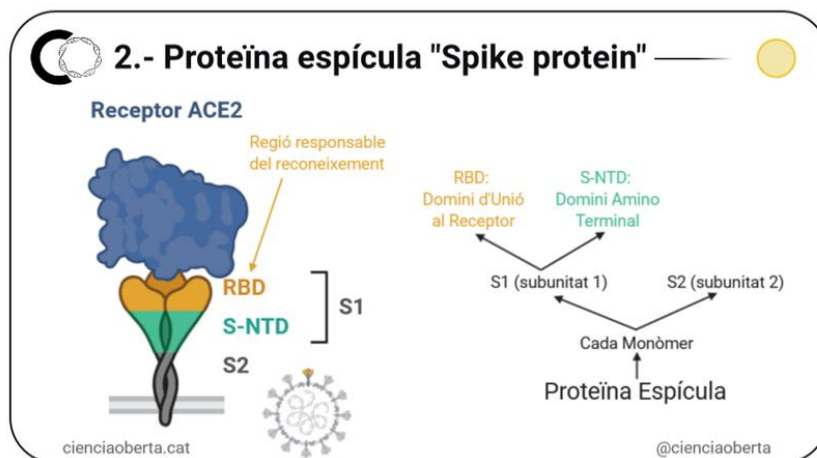


Figura 2. Esquema de l'estructura de la proteïna espícula generada amb <https://biorender.com/>

Després de la unió entre RBD (del virus) i ACE2 (de la cèl·lula), s'activa una proteïna que tallarà l'espícula: la proteasa TMPRSS2. Es produeixen dos talls, el primer al lloc de tall S1/S2 i després, el segon al lloc S2'. Així s'exposa el pèptid de fusió (verd a la figura 3) que provoca la fusió entre la membrana viral i la membrana de la cèl·lula.

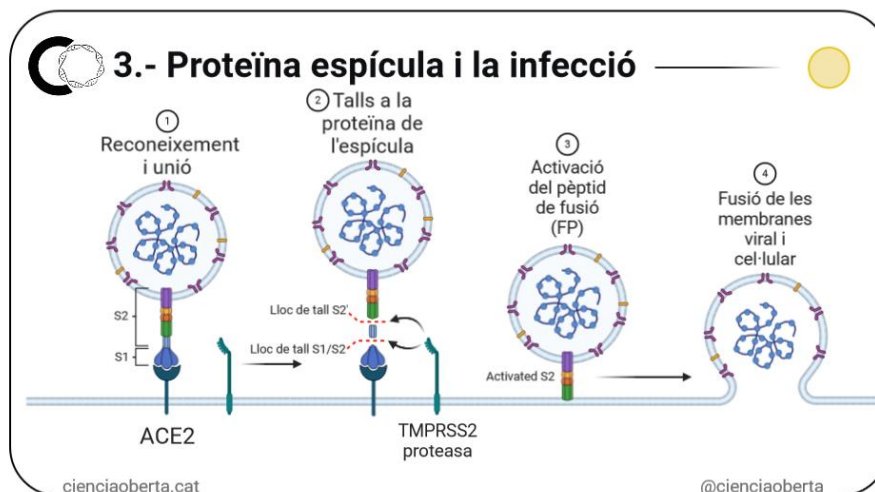


Figura 3. Model de la infecció de la cèl·lula pel coronavirus des del reconeixement RBD-ACE2 a la fusió virus-cèl·lula. Generat amb <https://biorender.com/>

D'aquesta manera, el virus és capaç d'entrar a les nostres cèl·lules, és a dir, d'infectar-les.

Tanmateix, aquesta unió per acabar infectant la cèl·lula no es produeix sempre, cal que la proteïna espícula tingui una conformació molt concreta. La interacció RBD-ACE2 es produeix quan una de les tres proteïnes S està "oberta" (RBD-up) i les altres dues "tancades" (RBD-down). Dit d'una altra manera, cal que la proteïna espícula tingui una subunitat S1 amb conformació RBD-up, figura 4.

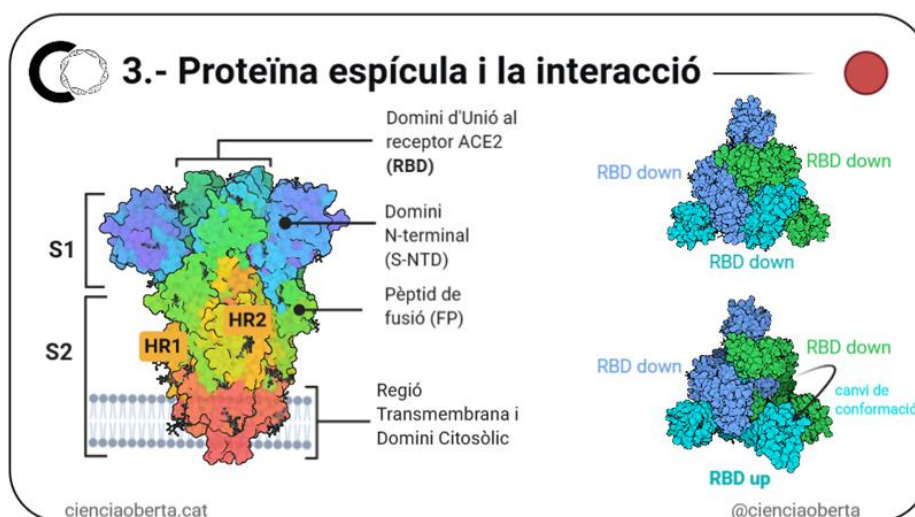


Figura 4. Estructura de la proteïna espícula on s'aprecien les regions S1, S2 i l'RBD. A la dreta observem el canvi conformacional de down-down-down a down-down-up. Generat amb <https://biorender.com/>

La proteïna N

A més de la proteïna espícula, hi ha altres proteïnes estructurals com la N, M i E.

La proteïna N (nucleocàpside) empaqueta el genoma i participa en l'encapsidació del virus, és a dir, és important per construir nous virus correctament. Es creu que les proteïnes N fan fibres al voltant de les quals s'enrotlla el RNA viral, com si fos fil de cosir, figura 5.

Aquesta fibra es forma per l'associació de parelles o dímers de proteïnes N que tenen zones amb potencials positius que atrauen la càrrega negativa del RNA.

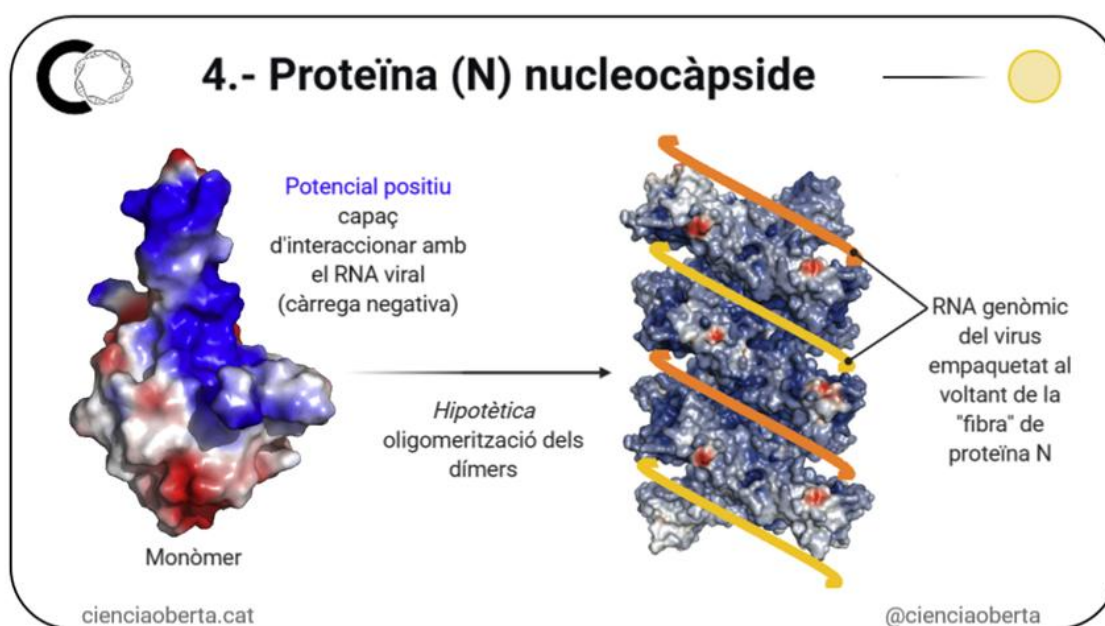


Figura 5. Estructura d'un monòmer de proteïna N amb els potencials de superfície i hipotètica estructura del complex proteïnes N i RNA viral. Generat amb <https://app.biorender.com/>

Passem a veure el material genètic del SARS-CoV-2

A diferència dels humans, que tenim una doble cadena de DNA, el coronavirus emmagatzema la seva informació genètica en forma d'una sola cadena d'RNA.

L'RNA és un manual d'instruccions que serveix per ensenyar a les cèl·lules infectades com fer més virus. Concretament, farà de motlle per fer més còpies d'RNA igual i per fer proteïnes virals, tot plegat per acabar muntant nous virus.

A l'RNA del virus li diem RNA genòmic (gRNA) i gràcies a ell, les cèl·lules poden crear dos tipus de molècules. Per una banda creen més RNA genòmic idèntic, que serà la informació genètica dels nous virions i per l'altra banda crearan RNA subgenòmic que només servirà de manual d'instruccions per fer proteïnes virals -també per formar o ajudar a formar nous virions-.

El coronavirus té un genoma molt gran per ser un virus d'RNA, té gairebé 30.000 nucleòtids (lletres), tots en una sola cadena "desxifrada, que es pot llegir directament" (gRNA+). M'explico: l'RNA pot emmagatzemar informació genètica "xifrada o desxifrada". Està

desxifrada si es pot llegir per fer proteïnes directament, però si està xifrada cal fer el pas de desxifrat. El coronavirus té RNA desxifrat (gRNA+).

Què és realment aquest xifrat? Seria com tenir un llibre on les A són U, les U són A, les C són G i les G són C. No ho podríem llegir, sinó que hauríem de transcriure-ho utilitzant aquesta complementarietat entre lletres. Per exemple: un RNA xifrat (RNA-) diria GAGU, però si fem la transcripció per desxifrar-ho (RNA+) llegim CUCA, que sí que té un sentit, codifica per un significat (proteïna). Per tant l'RNA amb sentit li diem RNA+, figura 6.

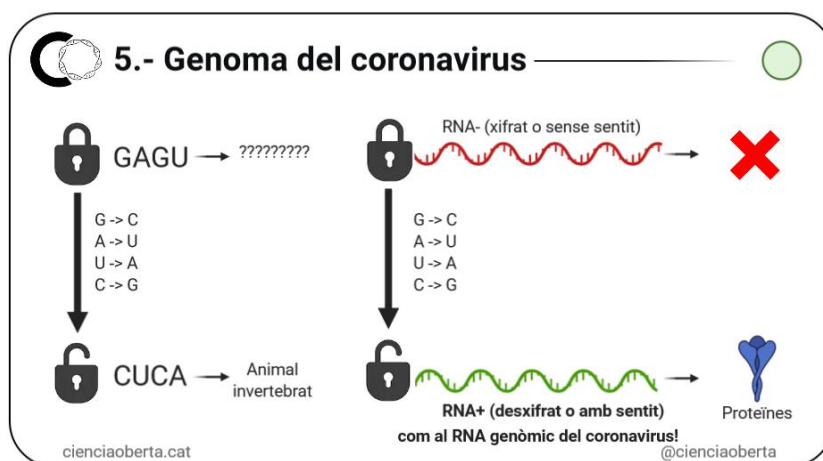


Figura 6. Esquema metafòric per explicar les cadenes d'àcids nucleics amb sentit i sense sentit. Generat amb <https://app.biorender.com/>

De què serveix el RNA- si no té sentit? Serveix com a motlle per fer RNA+. Per tant, amb el RNA genòmic amb sentit del coronavirus (gRNA+) podem fer RNA genòmic antisentit que servirà de motlle (gRNA-). Llavors aquest motlle de RNA genòmic antisentit (gRNA-) és immediatament utilitzat per fer tant RNA genòmic amb sentit (gRNA+) per fer nous virus, com RNAs subgenòmics (sgRNA+), que són trossets de gRNA+, és a dir, les instruccions per fer proteïnes per acabar formant nous virus, figura 7.

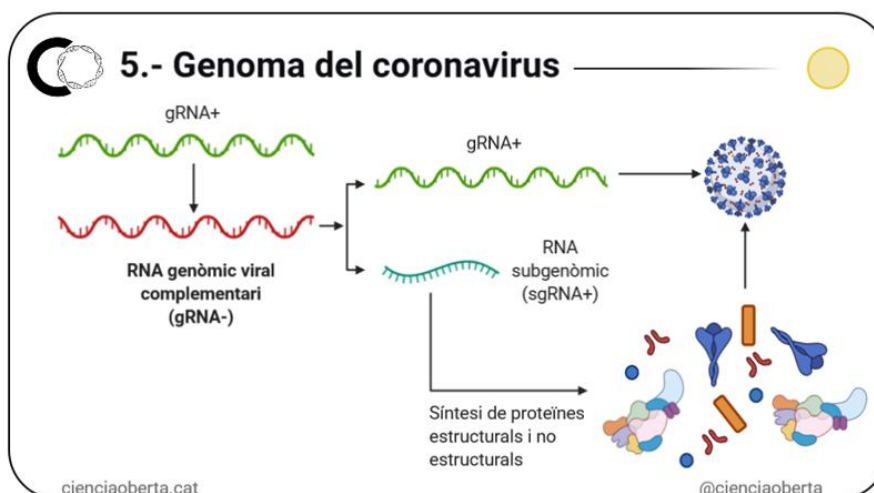


Figura 7. Flux de la informació genètica del coronavirus per acabar formant nous virions. Generat amb <https://app.biorender.com/>

L'encarregat de fer aquests processos de síntesi de RNA tant genòmic com subgenòmic és el complex de replicació-transcripció (RTC). Aquesta màquina està formada per moltes proteïnes no estructurals (Nsp) virals que s'envolten per una vesícula de doble membrana (DMV). L'eix del RTC és la nsp12, una RNA polimerasa dependent de RNA, és a dir, un enzim que crea una còpia d'RNA a partir d'RNA -pensa que normalment el RNA es fa a partir de DNA, no de més RNA-. Seguidament trobem la nsp14, un enzim que detecta i corregeix errors en la replicació gràcies a la seva activitat exonucleasa -és capaç de retirar el nucleòtids erronis-.

També trobem la nsp16, una 2'-O MTasa que protegeix l'RNA viral acabat de sintetitzar i l'ajuda a evitar la resposta immunitària. Finalment trobem altres proteïnes no estructurals que participen en el manteniment de l'estabilitat del complex i al correcte desenvolupament del procés, com la nsp9 i la nsp13, entre d'altres, figura 8.

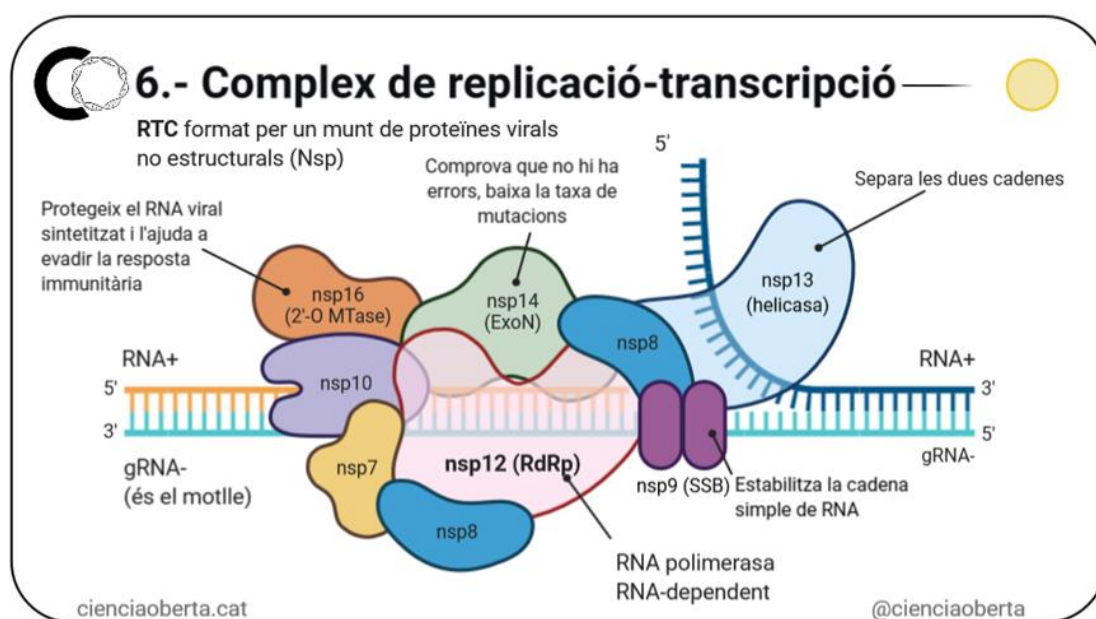


Figura 8. Complex de replicació-transcripció (RTC) format per proteïnes no estructurals (Nsp). Generat amb <https://app.biorender.com/>

Com s'organitza el seu genoma?

Al genoma del SARS-CoV-2 hi ha molts gens, que s'organitzen en dues parts. En la primera tenim els gens de proteïnes no estructurals (Nsp) que s'utilitzaran directament com a instruccions pels ribosomes, les fàbriques de proteïnes. Seran traduïts directament. Aquests ribosomes faran dos productes diferents anomenats poliproteïnes, cadascuna de les quals són realment moltes proteïnes diferents fusionades. Per activar-les i obtenir la seva forma funcional cal separar-les utilitzant proteïnes tisores: les proteases.

A la segona part del genoma, tenim les proteïnes estructurals i accessòries. Aquestes no es podran traduir directament, sinó que s'haurà de passar per la fase de RNA- per fer trossets de RNA+ (els RNA subgenòmics sgRNA+ prèviament explicats), figura 9.

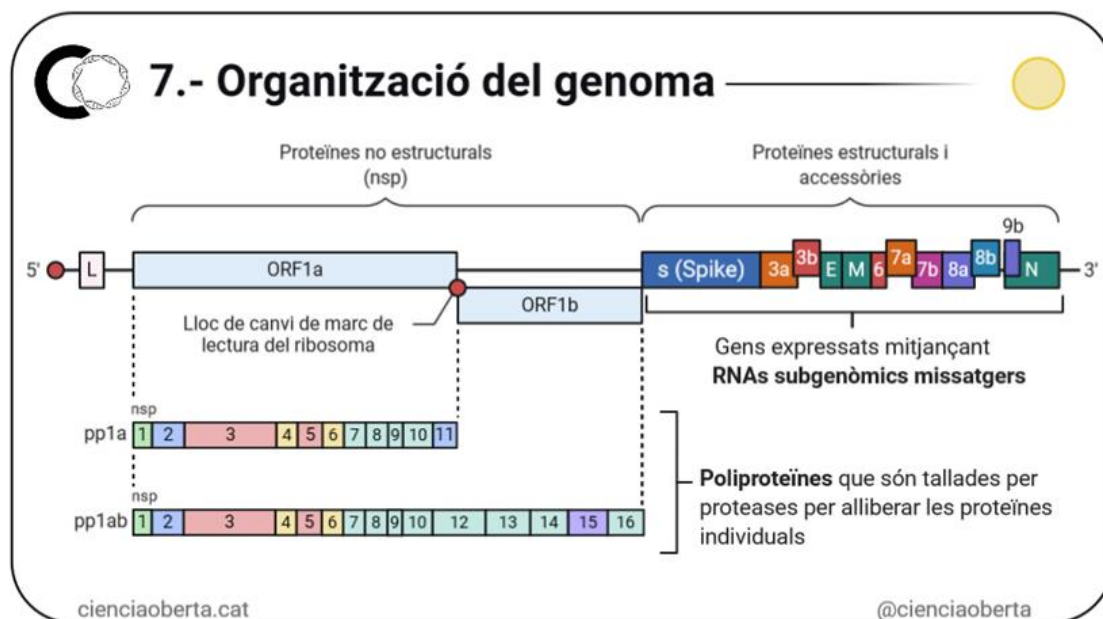


Figura 9. Organització del genoma del coronavirus en dues regions amb diferent funcionament. A l'esquerra les proteïnes no estructurals i a la dreta les estructurals i accessorïes. Generat amb <https://app.biorender.com/>

Ja per acabar...

Podríem parlar de molts aspectes més, però és evident que el funcionament del coronavirus és fascinant. A hores d'ara en tenim molta informació que s'ha obtingut tant de l'estudi d'altres coronavirus similars com el SARS-CoV com del propi SARS-CoV-2. Precisament gràcies a aquest coneixement, concretament al de la seva seqüència, s'ha pogut observar la seva proximitat evolutiva amb altres coronavirus. De fet, diferents estudis estimen que el SARS-CoV-2 i el SARS-CoV van divergir fa entre 40 i 70 anys!

Bibliografia:

Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., & Glaunsinger, B. A. (2020). The molecular virology of coronaviruses. *The Journal of biological chemistry*, 295(37), 12910–12934. <https://doi.org/10.1074/jbc.REV120.013930>

Sisi Kang, Mei Yang, Zhongsi Hong, Liping Zhang, Zhaoxia Huang, Xiaoxue Chen, Suhua He, Ziliang Zhou, Zhechong Zhou, Qiuyue Chen, Yan Yan, Changsheng Zhang, Hong Shan, Shoudeng Chen. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, *Acta Pharmaceutica Sinica B*, Volume 10, Issue 7, Pages 1228-1238, ISSN 2211-3835. <https://doi.org/10.1016/j.apsb.2020.04.009>.

Walls, A., Tortorici, M., Bosch, B.J. *et al.* Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **531**, 114–117 (2016). <https://doi.org/10.1038/nature16988>