

Enquête bioinformatique sur les origines de SARS-CoV-2

*Cours donné par Jacques van Helden, Emese Mégalecz et Gabriel Neve
Sur base d'une enquête menée par Erwan Salard, José Haloy, Didier Casane,
Etienne Decroly et Jacques van Helden*

03/11	10:00	12:00	CM1	(1) La question des origines de SARS-CoV-2. (2) Biologie de SARS-CoV-2. (3) Événements évolutifs. (4) Bases de données biologiques. (5) Alignement de séquences par paires. (6) Recherche de séquences par similarité.
10/11	8:30	11:30	TP1 groupe 1	(1) Bases de données de séquences biologiques (Uniprot, NCBI). (2) Alignement par paires (needle). (3) Recherche de similarités (BLAST).
	12:30	15:30	TP1 groupe 2	cf groupe 1
	15:45	18:00	CM2	(1) Interprétation des résultats du TP 1. (2) Profils de pourcents de positions identiques. (3) Alignements multiples. (4) Inférence phylogénétique
17/11	8:30	11:30	TP2 groupe 1	(1) Alignements 1 à N et profils de PPI (PIPprofileR) (2) Alignements multiples (clustal). (3) Inférence phylogénétique (phylogeny.fr)
	12:30	15:30	TP2 groupe 2	cf groupe 1
	16:00	17:00	CM3	Interprétation, résumé et conclusion du cours

Pourquoi l'origine de SARS-CoV-2 suscite-t-elle des débat ?

Pourquoi la question des origines se pose-t-elle ?

- Dès le début de la pandémie, la question des origines du virus a suscité de fortes controverses : origine naturelle (transfert d'animal à humain) ou artificielle (produit dans un laboratoire) ?
- Les éléments de la controverse
 - Le virus a émergé dans la ville de Wuhan, où est situé le laboratoire de référence pour les recherches sur les coronavirus
 - Ce laboratoire réalise des expériences sur les coronavirus, notamment de gain de fonction
 - L'hypothèse d'une contamination dans le marché de Wuhan est fortement remise en cause
 - Certains chercheurs affirment que le génome de SARS-CoV-2 résulte de manipulations génétiques
 - Certains chercheurs pensent que le virus pourrait s'être échappé accidentellement d'un laboratoire. D'autres affirment carrément qu'il a été conçu pour servir d'arme biologique
 - Ces affirmations sont fortement médiatisées par les réseaux sociaux, mais ne sont pas publiées dans les revues scientifiques avec comité de lecture
 - Le débat scientifique est fortement perturbé par le contexte politique et géopolitique (relations Chine - USA, rôle de l'OMS, ...)
- Comment distinguer le vrai du faux ?
 - **Approche: ignorer délibérément les débats médiatiques et analyser les séquences des virus**

A la recherche de l'hôte manquant

- Les coronavirus responsables des dernières émergences épidémiques chez l'humain trouvaient leur origine dans un "réservoir naturel" constitué par les chauves-souris.
- La transmission à l'homme passe généralement par un hôte animal intermédiaire: la civette pour le SRAS (2002) et le dromadaire pour le MERS (2012).
- Pour la pandémie COVID-19 on a invoqué le pangolin comme hôte intermédiaire, en suggérant que le passage à l'homme provenait d'animaux vendus sur le marché de Wuhan.
- Cette hypothèse est cependant remise en cause pour différentes raisons.
 - Les premiers patients ne fréquentaient pas le marché
 - Les génomes des coronavirus de pangolin dont on dispose sont beaucoup plus éloignés de SARS-CoV-2 que ceux des chauves-souris

Expériences de gain de fonction

- Les coronavirus responsables des dernières émergences épidémiques chez l'humain trouvaient leur origine dans un "réservoir naturel" constitué par les chauves-souris.
- La transmission à l'homme passe généralement par un hôte animal intermédiaire: la civette pour le SRAS (2002) et le dromadaire pour le MERS (2012).
- Pour la pandémie COVID-19 on a invoqué le pangolin comme hôte intermédiaire, en suggérant que le passage à l'homme provenait d'animaux vendus sur le marché de Wuhan.
- Certains laboratoires de virologie pratiquent des expériences dites "de gain de fonction" qui consistent à modifier un virus pour le rendre plus virulent ou plus contagieux, afin d'étudier les mécanismes moléculaires de l'infection virale.
- Ceci suscite de fortes réticences au sein même de leur communauté.
- 2011: Ron Fouchier annonce qu'il a produit un virus H5N1 modifié (9 mutations ponctuelles) pour augmenter sa contagiosité chez le furet.
- Moratoire:
 - En 2014 le NIH annonce qu'il ne financera plus les expériences de gain de fonction
 - En 2017 il lève cette mesure, en invoquant l'intérêt de ces expériences pour comprendre les mécanismes de l'infection virale.
- Plusieurs incidents ont déjà été rapportés concernant des fuites accidentnelles de virus de laboratoires.

Arguments en faveur d'une zoonose naturelle

- Andersen et collègues publient dès janvier 2020 un article affirmant que SARS-CoV-2 est sans aucun doute d'origine naturelle.
- Argument principal : la séquence du domaine de liaison au récepteur (RBD) est optimale pour se lier au récepteur ACE2, mais d'une façon différente de celles qu'on connaissait jusqu'alors. D'après les auteurs, si on avait conçu un virus dans le but de le rendre infectieux pour l'homme, on n'aurait pas pu concevoir cette séquence.
- Autre argument: on ne trouve pas dans ce génome de traces d'ingénierie moléculaire (par exemple des sites de restriction)

correspondence



The proximal origin of SARS-CoV-2

NATURE MEDICINE | VOL 26 | APRIL 2020 | 450–455 | www.nature.com/naturemedicine

While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding. Thus, the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of purposeful manipulation.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine 26, 450–452.

Arguments en faveur d'une zoonose naturelle

- Andersen et collègues publient dès janvier 2020 un article affirmant que SARS-CoV-2 est sans aucun doute d'origine naturelle.
- Argument principal : la séquence du domaine de liaison au récepteur (RBD) est optimale pour se lier au récepteur ACE2, mais d'une façon différente de celles qu'on connaissait jusqu'alors. D'après les auteurs, si on avait conçu un virus dans le but de le rendre infectieux pour l'homme, on n'aurait pas pu concevoir cette séquence.
 - *Oui mais les auteurs n'envisagent pas une autre voie possible : en cultivant des virus sur des cellules (en labo) on peut réaliser une sélection artificielle, qui peut déboucher sur des souches adaptées sans nécessiter de connaissance a priori des séquences.*
- Autre argument: on ne trouve pas dans ce génome de traces d'ingénierie moléculaire (par exemple des sites de restriction)
 - *Oui mais les techniques de biologie synthétique permettent depuis 15 ans de générer une molécule d'ADN de novo, sans recourir à des enzymes de restriction (et donc sans trace).*

correspondence



The proximal origin of SARS-CoV-2

NATURE MEDICINE | VOL 26 | APRIL 2020 | 450–455 | www.nature.com/naturemedicine

While the analyses above suggest that SARS-CoV-2 may bind human ACE2 with high affinity, computational analyses predict that the interaction is not ideal and that the RBD sequence is different from those shown in SARS-CoV to be optimal for receptor binding. Thus, the high-affinity binding of the SARS-CoV-2 spike protein to human ACE2 is most likely the result of natural selection on a human or human-like ACE2 that permits another optimal binding solution to arise. This is strong evidence that SARS-CoV-2 is not the product of purposeful manipulation.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nature Medicine 26, 450–452.

- Sirokin & Sirokin discutent des faiblesses des arguments d'Andersen
- Ils soulignent également l'absence générale d'évaluation sérieuse des hypothèses alternatives concernant les possibilité d'échappement de laboratoire.
- Ils développent l'historique des accidents de laboratoire, et évaluent les scénarios qui permettraient également de comprendre la nature des données en notre possession.
- Des passages successifs d'une souche virale d'une espèce à l'autre (en culture cellulaire ou sur animaux) donneraient le même effet : une divergence accélérée par rapport aux taux de mutations en milieu naturel.

PROBLEMS & PARADIGMS

Prospects & Overviews

Might SARS-CoV-2 Have Arisen via Serial Passage through an Animal Host or Cell Culture?

A potential explanation for much of the novel coronavirus' distinctive genome

Karl Sirokin* and Dan Sirokin

Despite claims from prominent scientists that SARS-CoV-2 indubitably emerged naturally, the etiology of this novel coronavirus remains a pressing and open question. Without knowing the true nature of a disease, it is impossible for clinicians to appropriately shape their care, for policy-makers to correctly gauge the nature and extent of the threat, and for the public to appropriately modify their behavior. Unless the intermediate host necessary for completing a natural zoonotic jump is identified, the dual-use gain-of-function research practice of viral serial passage should be considered a viable route by which the novel coronavirus arose. The practice of serial passage mimics a natural zoonotic jump, and offers explanations for SARS-CoV-2's distinctive spike-protein region and its unexpectedly high affinity for angiotensin converting enzyme (ACE2), as well as the notable polybasic furin cleavage site within it. Additional molecular clues raise further questions, all of which warrant full investigation into the novel coronavirus's origins and a re-examination of the risks and rewards of dual-use gain-of-function research.

same genetic signatures behind as a natural jump but occurring in a much shorter period of time.

The genetic signatures in question includes two distinctive features possessed by SARS-CoV-2's spike-protein: the unique sequence in the receptor binding domain (RBD), a region known to be critical for SARS-CoV-2's utilization of human angiotensin converting enzyme (ACE2), which is the cell surface receptor used by both SARS-CoV and SARS-CoV-2 for fusion with target cells and subsequent cell entry. The second feature is the presence of a polybasic furin cleavage site, which is also known as a multibasic cleavage site (MBS)—a four amino acid insertion with limited sequence flexibility—within the coronavirus's novel spike-protein, that is not found in SARS-CoV or other lineage B coronaviruses. This furin cleavage site,

Un virus synthétique avec des bouts de HIV ?

Le 17 avril 2020, le Professeur Luc Montagnier, Prix Nobel de médecine pour sa contribution à la découverte du HIV (le virus responsable du SIDA), défraie la chronique en annonçant sur plusieurs médias (Pourquoi Docteur, CNEWS) que le génome du coronavirus SARS-CoV-2, agent de la pandémie COVID-19, comporte quatre fragments de séquences provenant du HIV. De plus, il affirme que la présence de ces séquences ne résulte pas d'une recombinaison naturelle (fréquente chez les virus) ou d'un accident, mais d'un vrai travail d'ingénieur, effectué intentionnellement, vraisemblablement dans le cadre de recherches visant à développer des vaccins contre le HIV.

Pour appuyer sa théorie, Luc Montagnier cite deux études :

- le travail d'un collègue mathématicien, Jean-Claude Perez, qui "a fouillé les moindres détails de la séquence",
- une analyse des séquences génomiques et protéiques des coronavirus préalablement publiée par une équipe indienne, qui a, selon lui, "été forcée de rétracter" sa publication.

Professeur Luc Montagnier : Le virus covid19 est une manipulation humaine

(<https://www.youtube.com/watch?v=qSWCLHIOiMo>).

"Je suis arrivé à la conclusion qu'il y avait eu une manipulation de ce virus. [...] Il y a un modèle qui est évidemment le virus classique, et là c'était un modèle venant de la chauve-souris, et là, à ce modèle on a par-dessus ajouté les séquences du VIH, du SIDA. ... Non, ce n'est pas naturel, c'était un travail de professionnel, de biologiste moléculaire, très minutieux, on peut dire d'horloger, au niveau des séquences. Dans quel but ce n'est pas clair. Mon travail c'est d'exposer les faits, c'est tout. Je n'accuse personne, je ne sais pas qui a fait ça et pourquoi. La possibilité c'est qu'on a voulu faire un vaccin contre le SIDA. Donc on a pris des petites séquences du virus [HIV] et on les a installées dans la séquence plus grande du coronavirus. [...] Il y a quand même une volonté d'étoffement, nous ne sommes pas les premiers. Un groupe de chercheurs indiens très renommés avaient publié la même chose, on les a forcés à rétracter. Si vous regardez leur publication vous voyez une grande bande "annulé". "

Une arme biologique ?

- Li-Meng Yan
 - chercheuse chinoise réfugiée aux Etats-Unis
 - travaillait dans le laboratoire de référence de l'OMS pour la Chine
- En septembre et octobre 2020, elle publie sur zenodo (dépôt d'oeuvres électroniques, sans révision par des experts) deux articles, où elle affirme que
 - le virus résulte de manipulations génétiques reposant sur les méthodes classiques de biologie moléculaire (recombinaison d'ADN au moyen d'enzymes de restriction) ;
 - ce virus est une arme biologique.
- Ses articles n'ont pas encore été publiés, mais ils sont téléchargés et fortement médiatisés.
- Les arguments sous-jacents font cependant l'objet de critiques par les spécialistes.

Yan, Li-Meng, Kang, Shu, Guan, Jie & Hu, Shanchang. 2020a. SARS-CoV-2 Is an Unrestricted Bioweapon: A Truth Revealed through Uncovering a Large-Scale, Organized Scientific Fraud. , doi: [10.5281/zendo.4073131](https://doi.org/10.5281/zendo.4073131). Zenodo.

Yan, Li-Meng, Kang, Shu, Guan, Jie & Hu, Shanchang. 2020b. Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route. , doi: [10.5281/zendo.4028830](https://doi.org/10.5281/zendo.4028830). Zenodo.

The screenshot shows two research papers on the Zenodo platform. The top paper, "Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route" (September 14, 2020), has 776,634 views and 597,172 downloads. The bottom paper, "SARS-CoV-2 Is an Unrestricted Bioweapon: A Truth Revealed through Uncovering a Large-Scale, Organized Scientific Fraud" (October 8, 2020), has 161,914 views and 90,966 downloads. Both papers are indexed in OpenAIRE. The Zenodo interface includes a search bar, upload and communities buttons, and user login/sign-up options.

zenodo

Search Upload Communities

Log in Sign up

September 14, 2020

Working paper Open Access

776,634 597,172

views downloads

See more details...

Indexed in

OpenAIRE

Publication date: September 14, 2020

DOI: DOI 10.5281/zenodo.4028830

Communities: Coronavirus Disease Research Community - COVID-19

License (for files):

Jacques.van-Helden@univ-amu.fr

zenodo

Search Upload Communities

October 8, 2020

Working paper Open Access

161,914 90,966

views downloads

See more details...

Indexed in

OpenAIRE

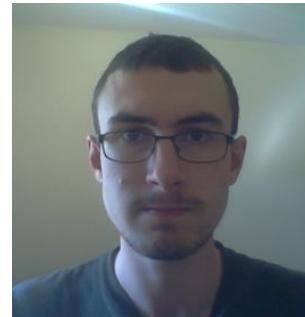
Publication date: October 8, 2020

DOI: DOI 10.5281/zenodo.4073131

License (for files): Creative Commons Attribution 4.0 International

A l'origine de ce cours

- Ce cours repose en grande partie sur un travail d'équipe mené durant le confinement d'avril-mai 2020.
- Nous remercions les nombreux collègues qui nous ont suggéré des améliorations sur les premières versions du manuscrit.
- Afin d'assurer la traçabilité et la reproductibilité de nos analyses, les données utilisées et les logiciels développés sont en libre accès
 - https://jvanheld.github.io/SARS-CoV-2_origins/



Erwan Sallard

Étudiant à l'Institut de Biologie
Ecole Normale Supérieure, Paris



Didier Casane

Biologie évolutive
Professeur, Université de Paris



José Halloy

Modélisation des systèmes biologiques
Professeur, Université de Paris



Etienne Decroly

Virologie
Dir. de Recherches, CNRS



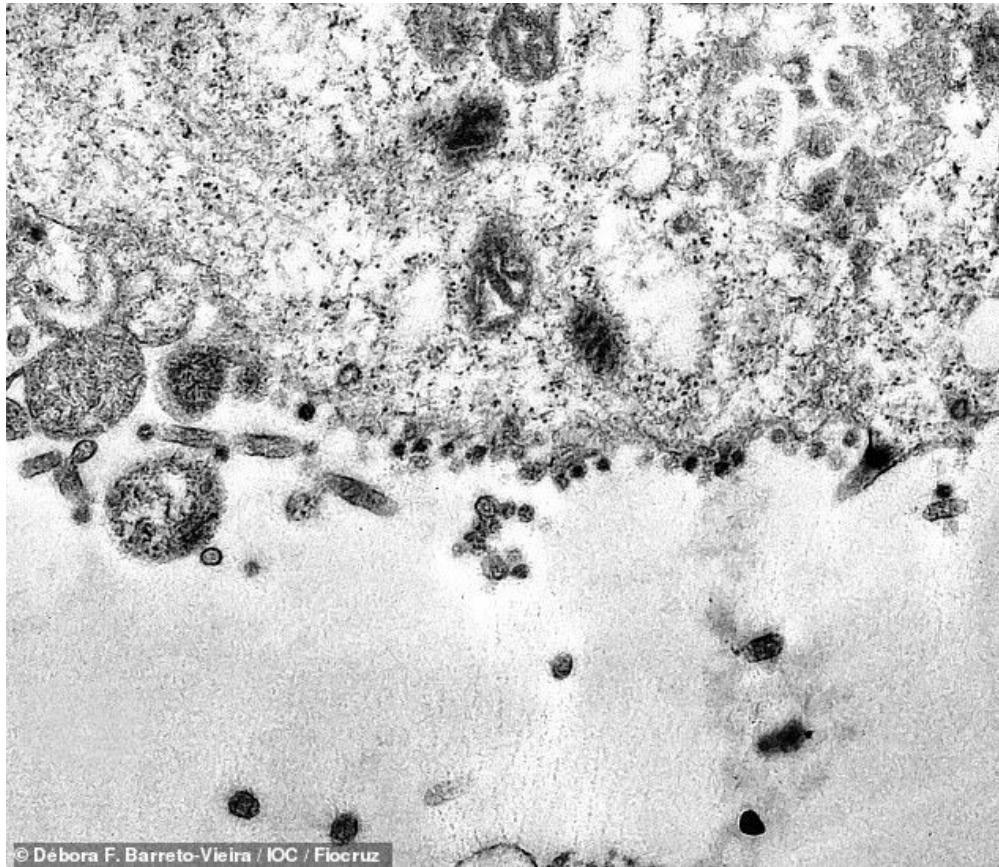
Jacques van Helden

Bioinformatique
Professeur, Aix-Marseille Université

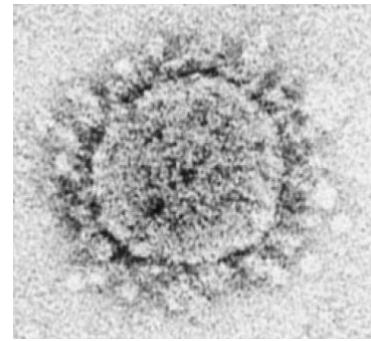
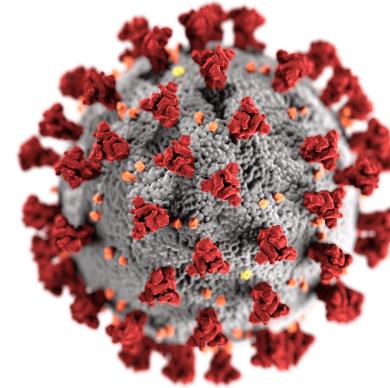
- Sallard, E., Halloy, J., Casane, D., van Helden, J. & Decroly, É. 2020. Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *Med Sci (Paris)* 36: 783–796.
English version : Erwan Sallard, José Halloy, Didier Casane, Etienne Decroly, Jacques van Helden. Tracing the origins of SARS-CoV-2 in coronavirus phylogenies. hal-02891455

Biologie de SARS-CoV-2

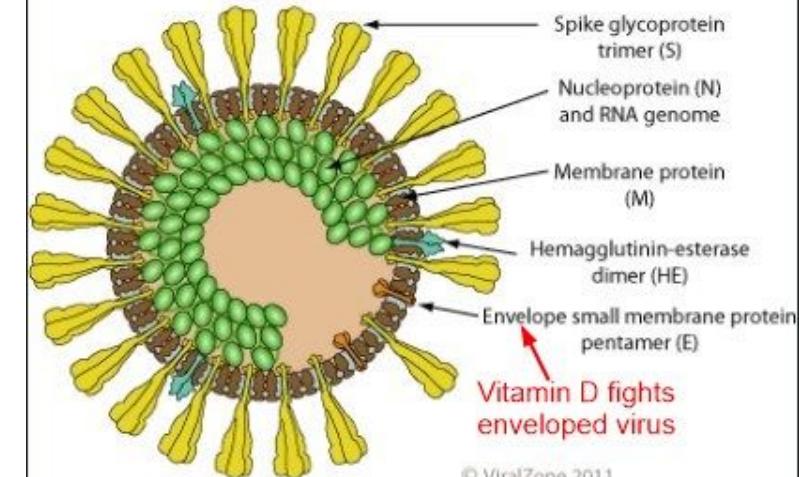
Coronavirus



© Débora F. Barreto-Vieira / IOC / Fiocruz



Murine Hepatitis Virus (MHV)



Le génome des coronavirus est constitué d'ARN

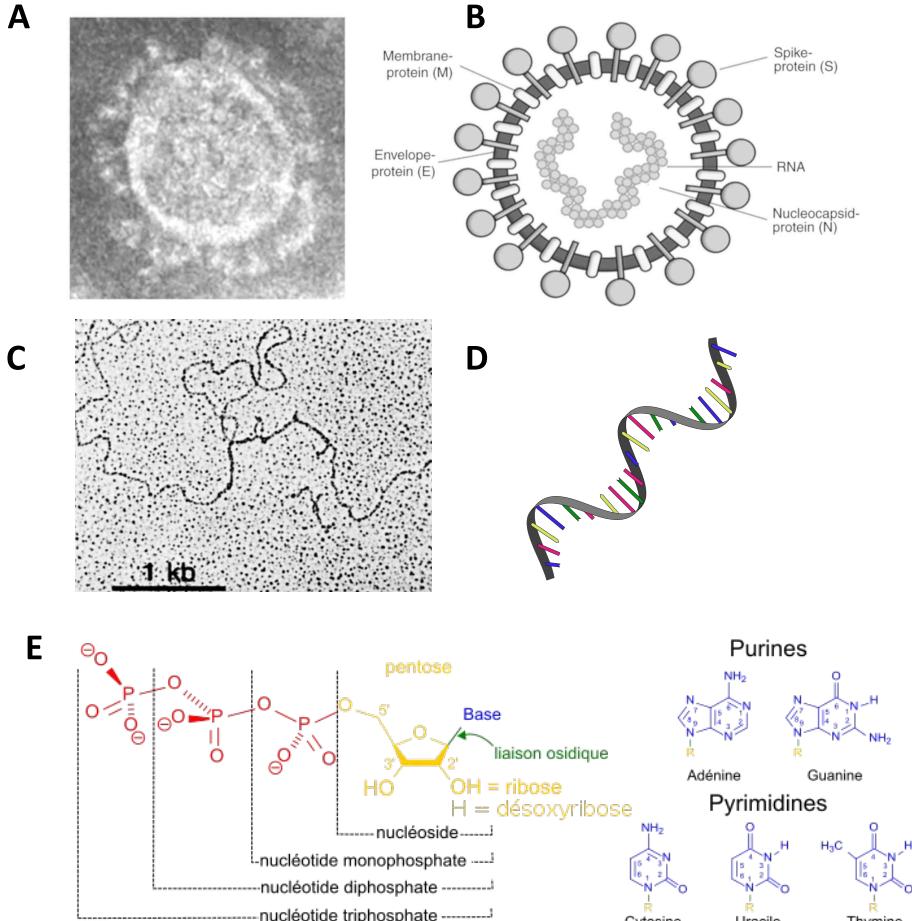
- Le génome est l'ensemble du matériel génétique d'un organisme
- Le génome des coronavirus est constitué d'acide ribonucléique (ARN).

Légende de la figure

- A. Micrographie électronique d'un virion de coronavirus.
- B. Schéma de la structure du virion. Le chapelet à l'intérieur symbolise l'ARN.
- C. Micrographie électronique de l'ARN viral extrait de son enveloppe
- D. Schéma de la structure de l'ARN. Les bâtonnets de couleur représentent les nucléotides

Adénine
Cytosine
Guanine
Uracile

- E. Structure chimique des nucléotides



Génome de SARS-CoV-2 de référence

- Début de la séquence génomique de SARS-CoV-2
- La taille totale fait 29 899 nucléotides
- Un des plus grands génomes parmi les virus à ARN
 - A Adénine
 - C Cytosine
 - G Guanine
 - T Uracile (on remplace le U par un T)

>MT019529.1 Severe acute respiratory syndrome coronavirus 2 isolate BetaCoV/Wuhan/IPBCAMS-WH-01/2019, complete genome
ATTAAGGTTACCTCCAGGTACAAACCAACTTCGATCTTGAGATCTGTTCTAAA
CGAACCTTAAATCTGTGGCTGACTCGGCTGCATGCTTAGTGCACTCACGCAAGTATAATTAAAC
TAATTACTGCTTGACAGGACACGAGTAACCTGCTATCTTCTGCAGGCTGCTACGGTTCGTCCGTG
TTGCAGCCGATCATCAGCACATCTAGGTTCTCGCCGGTGTGACCAGAAGGTAAAGATGGAGAGCCTTGTG
CCTGGTTCAACGAGAAAACACGTCCAACTCAGTTGCCCTGTTACAGGTTCGGACGTGCTCGTAC
GTGGCTTGGAGACTCCGTGGAGGGTCTTACAGAGGCACGTCAACATCTAAAGATGGCACTGTGG
CTTAGTAGAAGTTGAAAAAGGCGTTTGCTCAACTTGAACAGCCATATGTGTTCATCAAACGTTCCGAT
GCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACCTCGAAGGGCATTCAGTACGGTC
GTAGTGGTAGACACTTGGTCCTTGTCCCTCATGTGGCGAAATACCAAGTGGCTACCGCAAGGGTCT
TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCGATCTAAAGTCATTGACTTA
GGCGACGAGCTTGCACTGATCTTATGAAGATTTCAAGAAAACGGAAACTAAACATAGCAGTGGTG
TTACCCGTAACTCATGCGTGAGCTTAAACGGAGGGCATACACTCGCTATGTCGATAACAACCTCTGTGG
CCCTGATGGCTACCTCTTGAGTGCATTAAGACCTCTAGCACGTGGTAAAGCTTCATGCACCTTG
TCCGAACAACTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTG
CTTGGTACCGAACGTTCTGAAAAGAGCTATGAATTGAGACACCTTTGAAATTAAATTGGCAAAGAA
ATTGACACCTCAATGGGAATGTCCAAATTGTTATTTCCCTTAAATTCCATAATCAAGACTATTCAA
CCAAGGGTTGAAAAGAAAAAGCTTGTGGCTTATGGTAGAATTGAGCTGTCTATCCAGTGGCTCAC
CAAATGAATGCAACCAAATGTGCCCTCAACTCTCATGAAGTGTGATCATTGTTGAAACTTCATGGCA
GACGGCGATTGTTAAAGCCACTTGCATGGCAATTGTTGACTGAGAATTGACTAAAGAAGGTGCCACT
ACTTGTGGTACTTACCCCAAATGCTGTTAAATTATTGTCTCAGCATGTCACAATTCAAAGTAG
GACCTGAGCATAGTCTTGCATGGCAATTGACTGAGAATTGACTAAAGAAGGTGCCACT
CACTATTGCCCTGGAGGCTGTGTTCTTATGTTGGTGCATAACAAGTGTGCCATTGGTTCCA
CGTGCTAGCGCTAACATAGGTTGAACCATAACAGGTTGGAGAAGGTTCCGAAGGTCTTAAATGACA
ACCTTCTGAAATACTCCAAAAGAGAAAGCTAACATCAATATTGTTGGTAGCTTAAACTTAATGAAGA
GATGCCATTATTGGCATCTTCTGCTTCCACAAGTGCTTGTGGAAACTGTGAAAGGTTGGAT
TATAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTAAAGTACAAAGGAAAGCTAAAAAG
GTGCCATTGGAAATTGGTAGACAGAAATCAAACTGAGTCCTTATGCATTTGCAGAGGCTGCTCG

Appports de la génomique

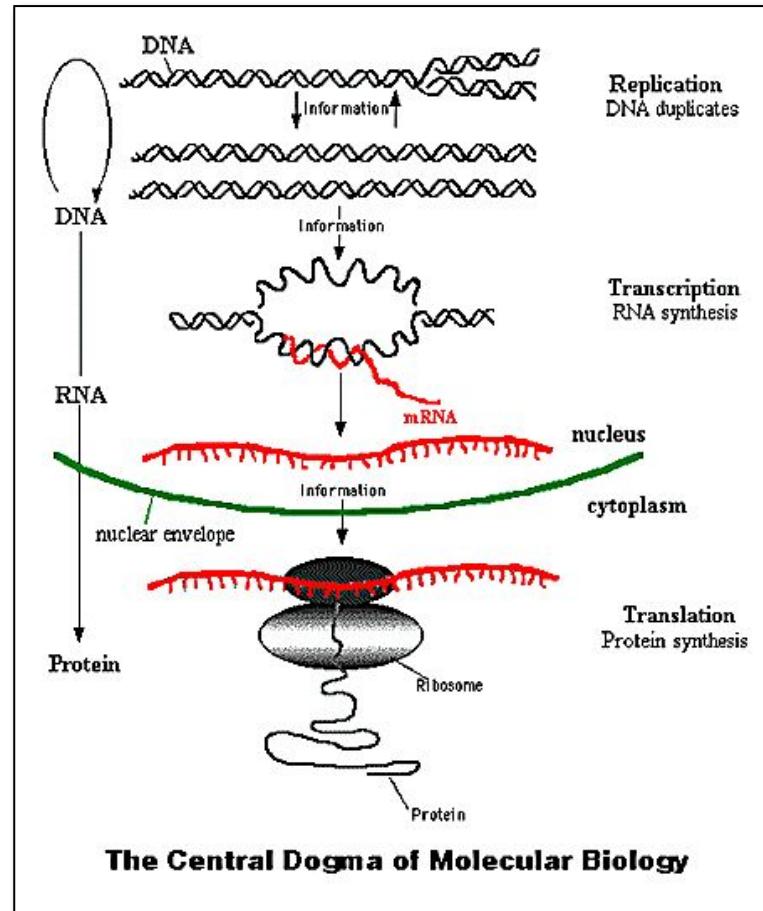
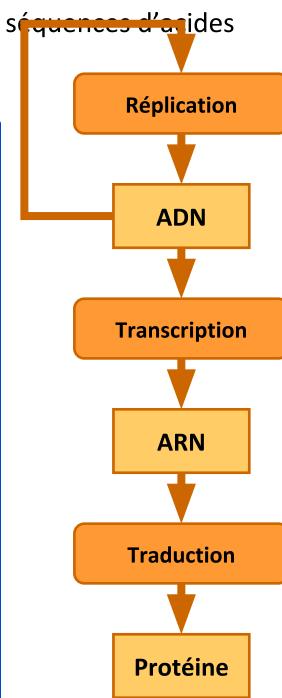
- Le séquençage d'un génome viral représente un coût modique et se fait très rapidement.
- On dispose de
 - plusieurs dizaines de génomes de coronavirus non-humains
 - plusieurs dizaines de milliers de génomes de SARS-CoV-2 (humain)
- Ces données permettent de poser des questions concernant
 - l'origine animale de SARS-CoV-2
 - sa propagation dans la population humaine

Le “dogme central” de la biologie

- Formulé en 1958 par Francis Crick
 - Crick, F. H. (1958). On protein synthesis. Symp Soc Exp Biol 12, 138-63.
 - Je recommande également de lire cette discussion ultérieure : Crick, F. (1970). Central dogma of molecular biology. Nature 227, 561-3.
- On le résume souvent de la façon suivante
 - **DNA makes RNA makes protein**
(l'ADN fait l'ARN qui fait les protéines)
- Cependant le dogme ne dit pas exactement cela. Il énonce les transferts d'information qui sont possibles ou impossibles entre les séquences d'acides nucléiques et celles des protéines.

Le dogme central stipule que, une fois que l' « information » est passée dans la protéine elle ne peut pas en ressortir. Plus précisément, le transfert d'information serait possible d'acide nucléique à acide nucléique, ou d'acide nucléique à protéine, mais le transfert de protéine à protéine, ou de protéine à acide nucléique est impossible. Information signifie ici la détermination précise de la séquence, soit des bases dans l'acide nucléique, soit des résidus aminoacides dans la protéine.

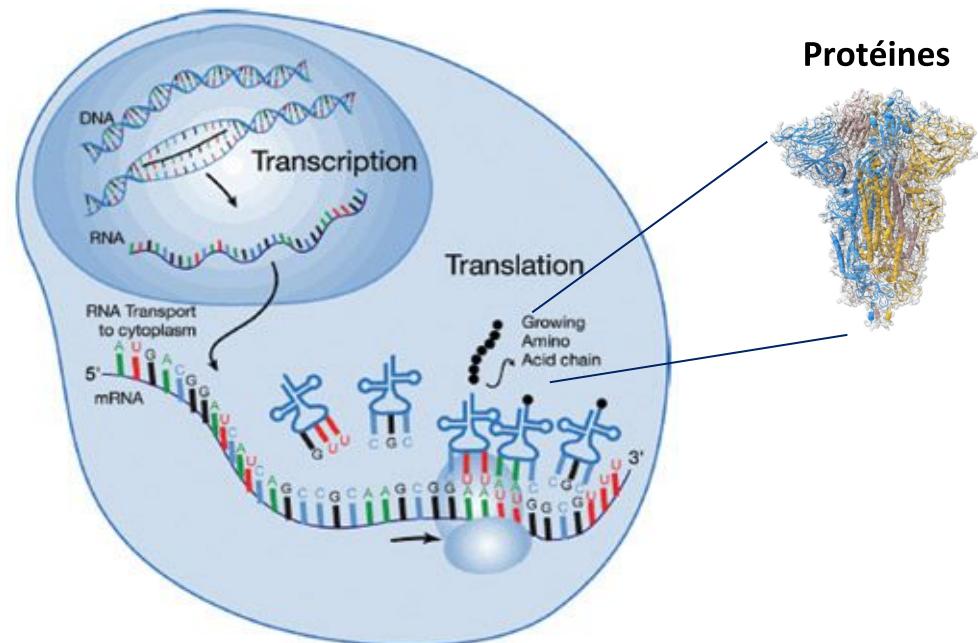
Crick, F. H. (1958). On protein synthesis. Symp Soc Exp Biol 12, 138-63.



DNA makes RNA makes protein*

* Formulation compacte des flux de l'information moléculaire, par Francis Crick

- Dans tous les organismes cellulaires
 - L'ADN sert de modèle à la synthèse d'ARN (*transcription*)
 - L'ARN sert de modèle à la synthèse des protéines (*traduction*)
- Les protéines sont les principaux acteurs moléculaires des organismes vivants
 - Enzymes
 - Transporteurs
 - Régulateurs
 - Cycle cellulaire
 - Différenciation cellulaire
 - ... un tas d'autres fonctions



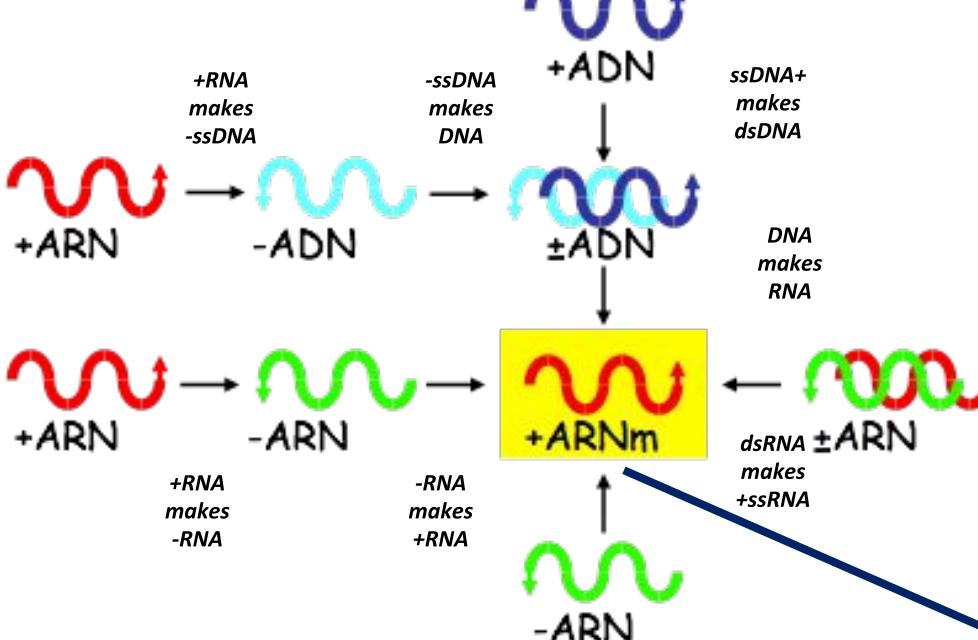
Protéines

Le matériel génétique des virus

rétrovirus

Poliomyélite
Hépatite A, C
Fièvre jaune
CORONAVIRUS

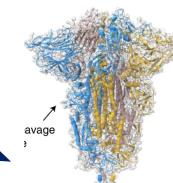
parvovirus



Papilloma
Herpès
Variole

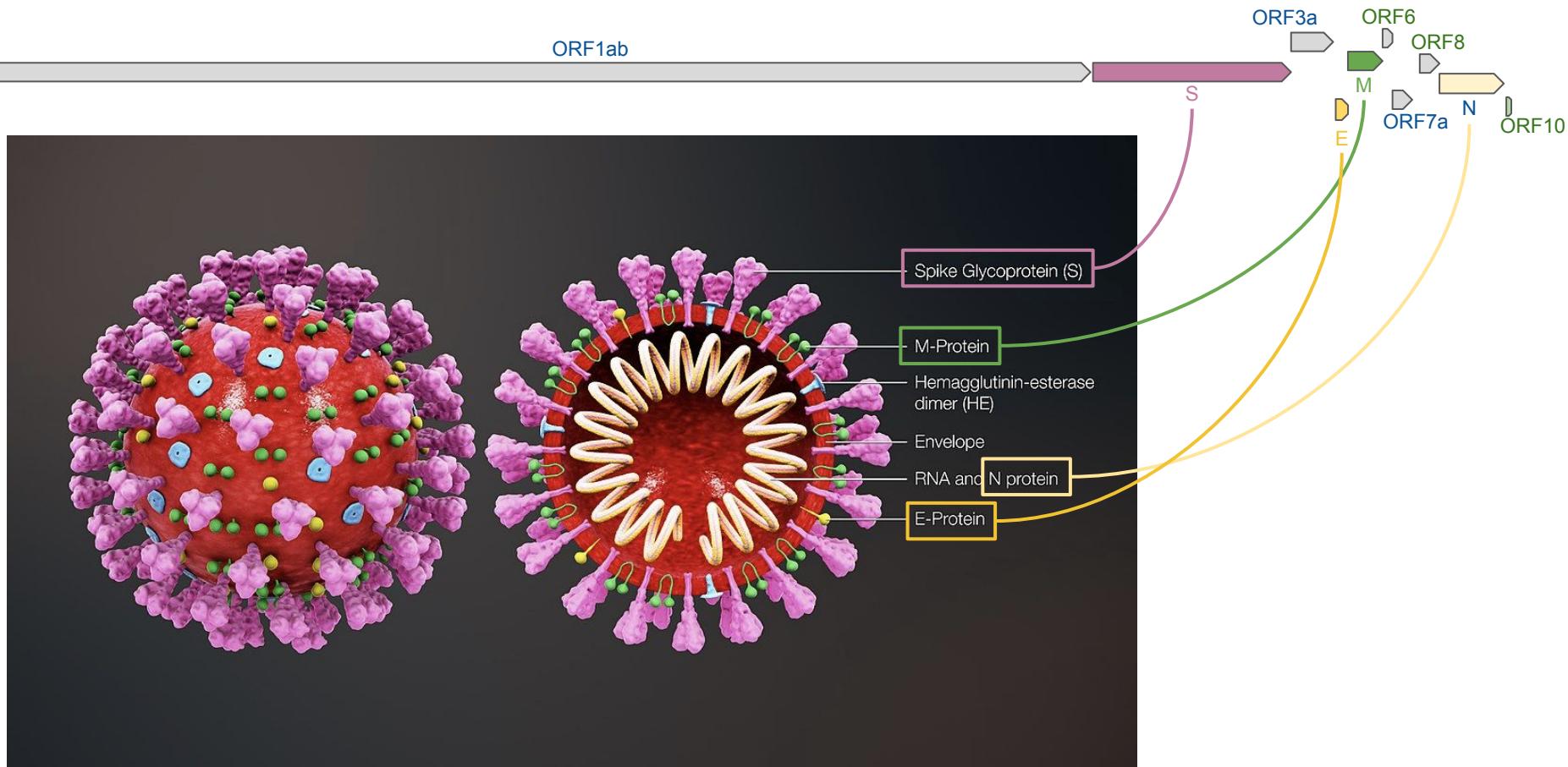
rotavirus

Grippe Rage
Rougeole
Oreillons
Ebola



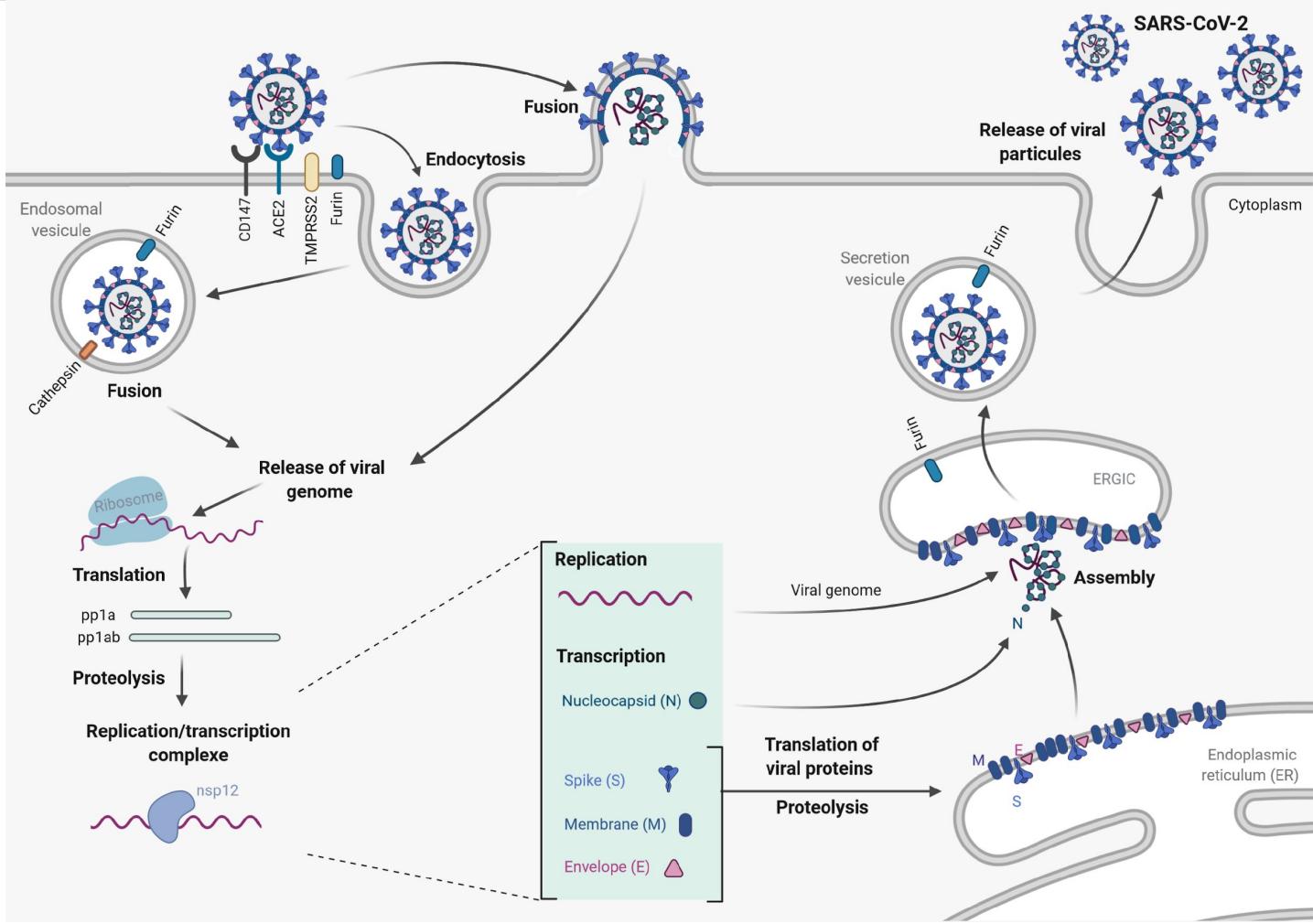
Protéines

Fonction des gènes de SARS-CoV-2 – Gènes structuraux

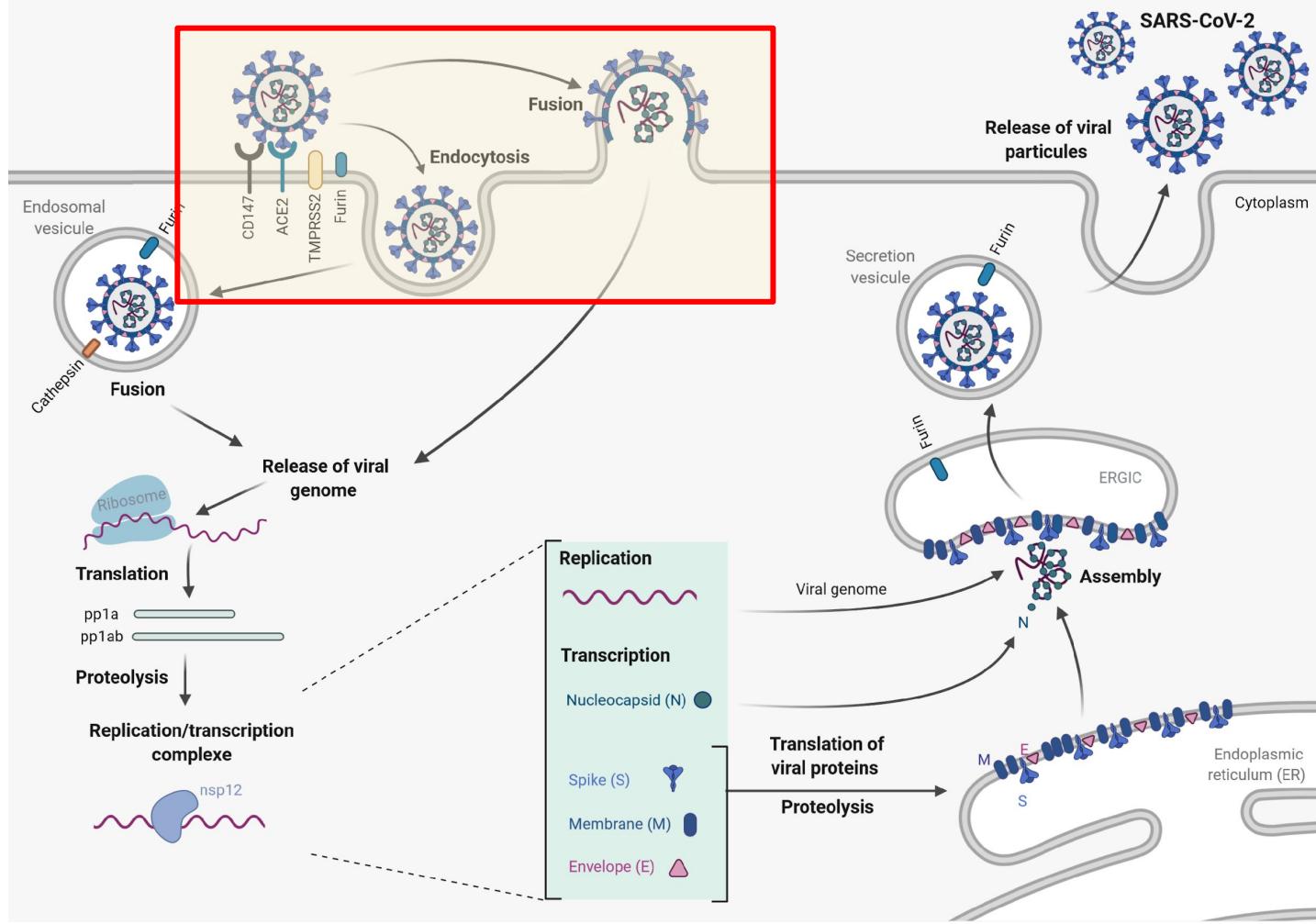


Mécanismes d'infection

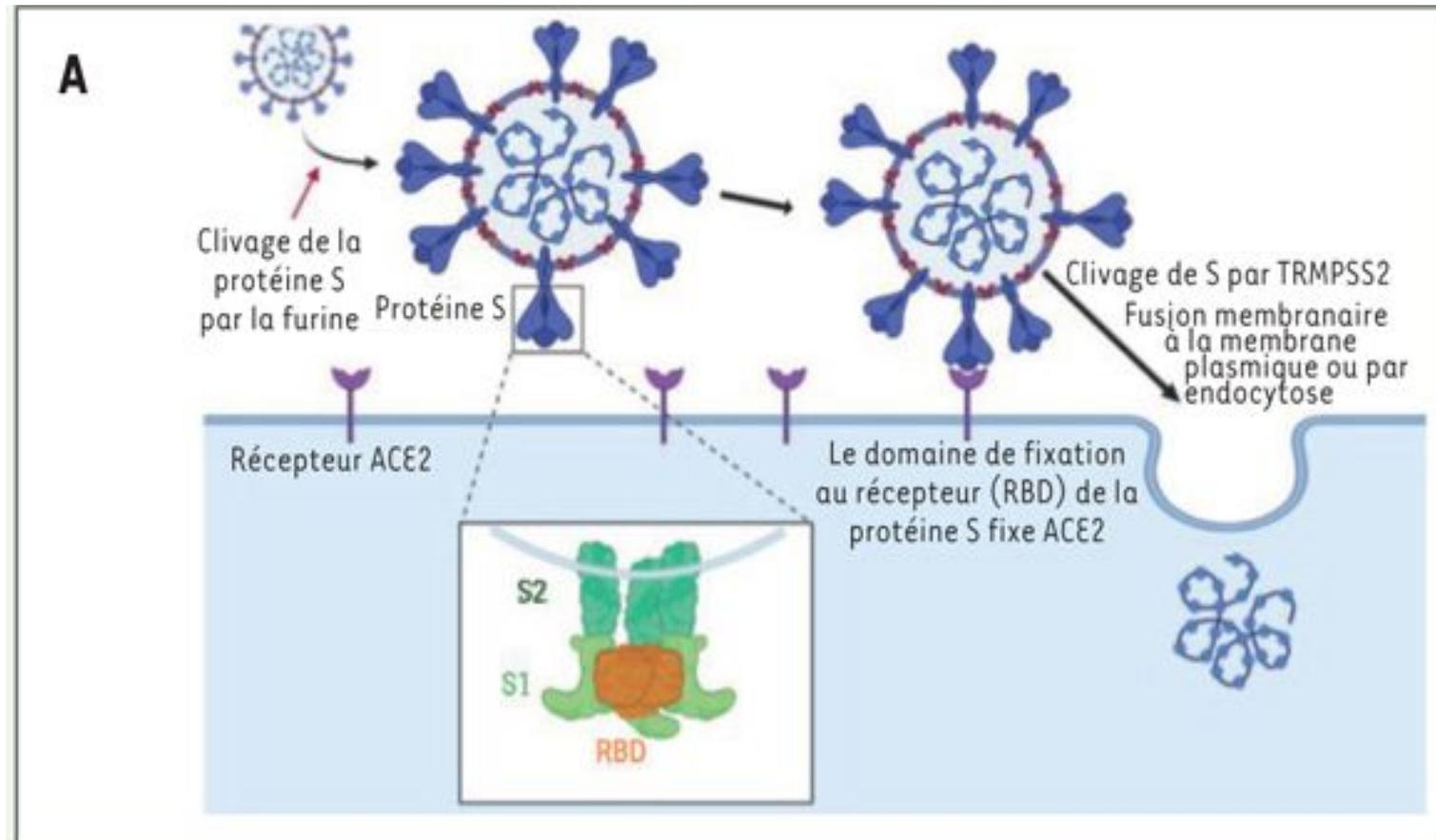
Cycle de réPLICATION des coronavirus



Cycle de réplication des coronavirus - Entrée dans la cellule hôte

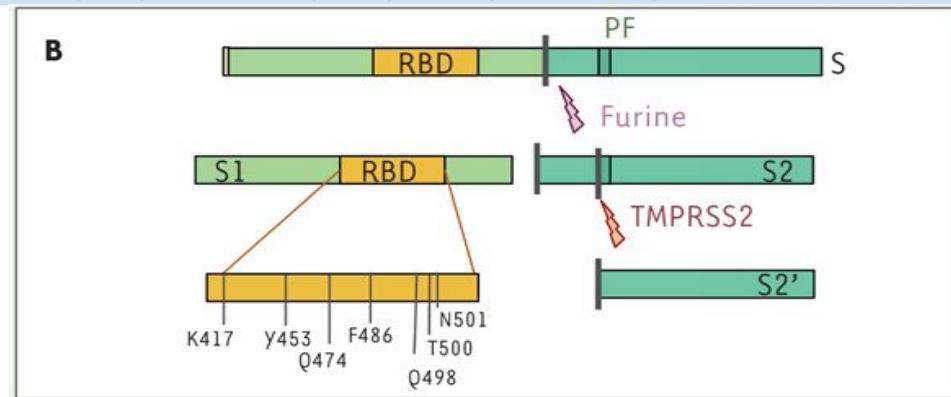


La protéine S reconnaît un récepteur cellulaire (SARS likes ACE2)

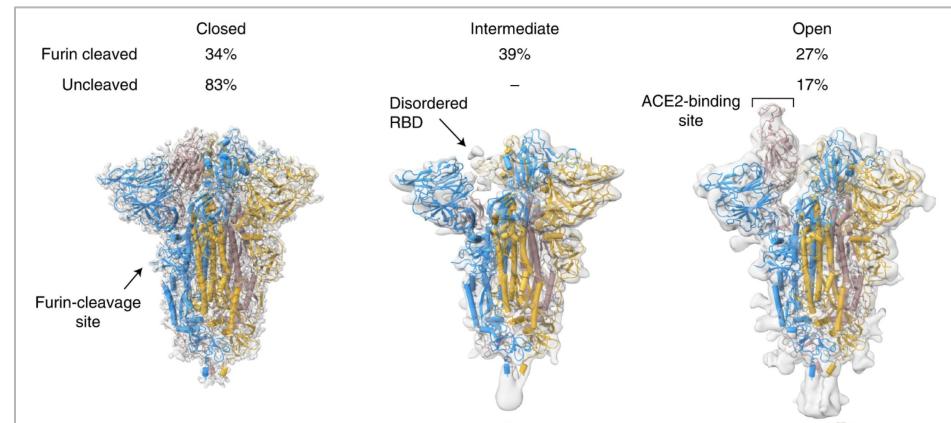


La protéine S doit être activée par un clivage protéolytique (priming)

- Juste après la traduction, le produit du gène S est une protéine inactive.
- L'activation requiert le clivage (coupe) de la protéine en deux parties (S1, S2).
- Chez SARS-CoV-2, ce clivage est réalisé par une enzyme appelée furine.
- Ceci a une grave conséquence, car cette enzyme est ubiquitaire dans les cellules humaines, ce qui explique en partie que les symptômes de la Covid-19 ne se limitent pas aux voies respiratoires.

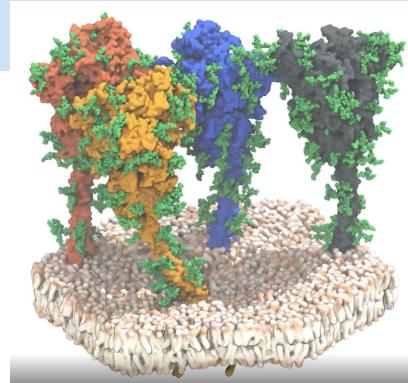
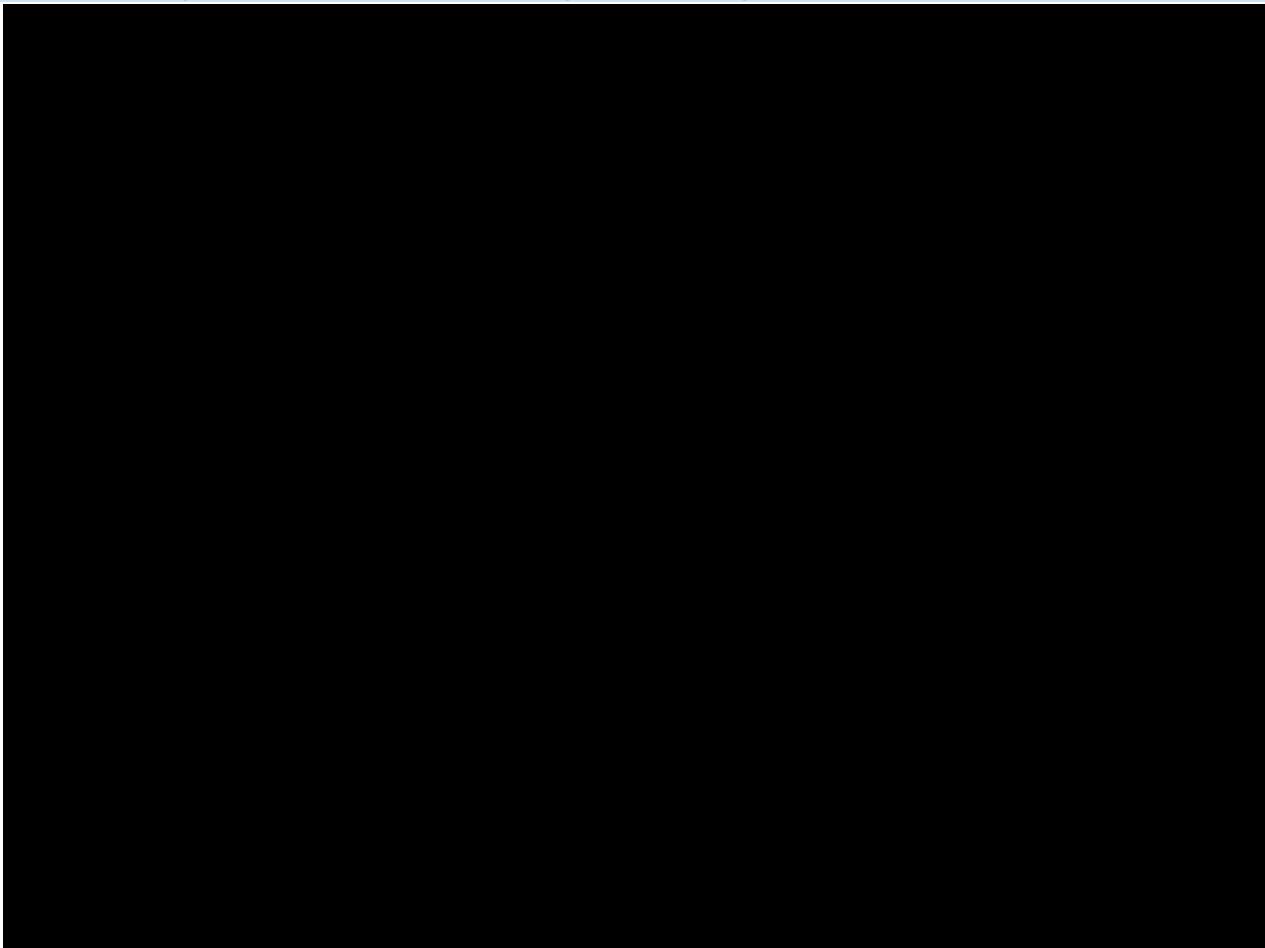


Sallard, E., Halloy, J., Casane, D., van Helden, J. & Decroly, É. 2020. Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *Med Sci (Paris)* 36: 783–796



Wrobel, A.G., Benton, D.J., Xu, P., Roustan, C., Martin, S.R., Rosenthal, P.B., et al. 2020. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* 27: 763–767.

Structure de la protéine S & dynamique moléculaire



Simulation moléculaire de l'interaction spicule - récepteur

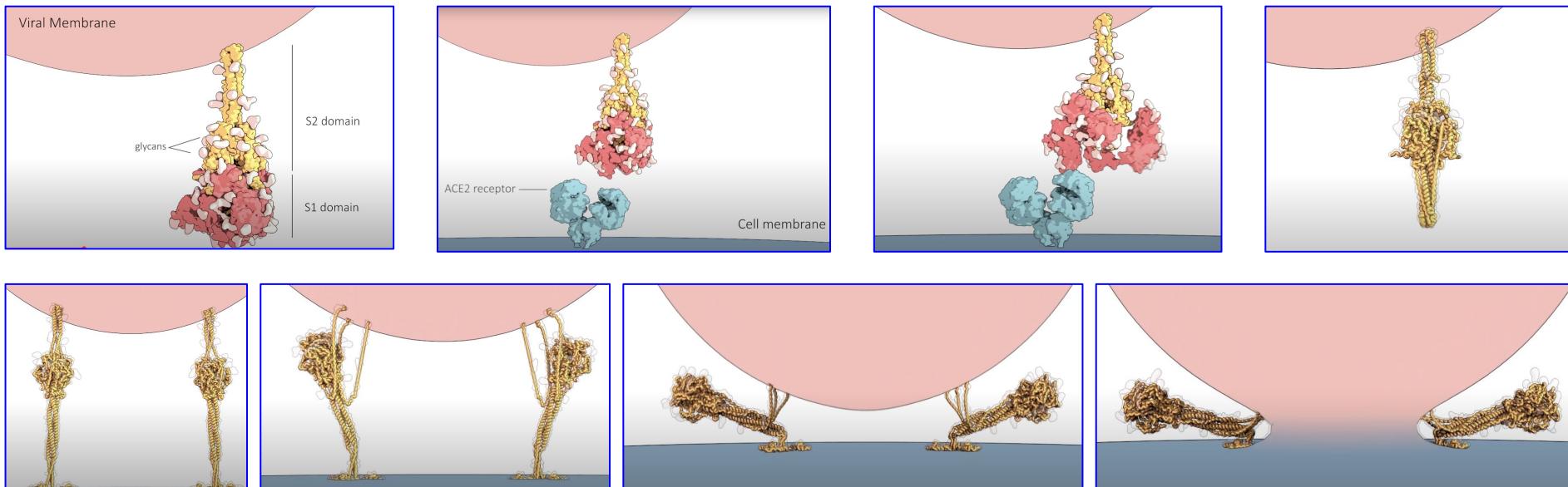
<https://youtu.be/e2Qi-hAXdJo?t=12>

Un modèle visuel de simulation dynamique illustrant la façon dont la protéine de spicule (spike protein) assure la fusion des membranes virale et cellulaire.

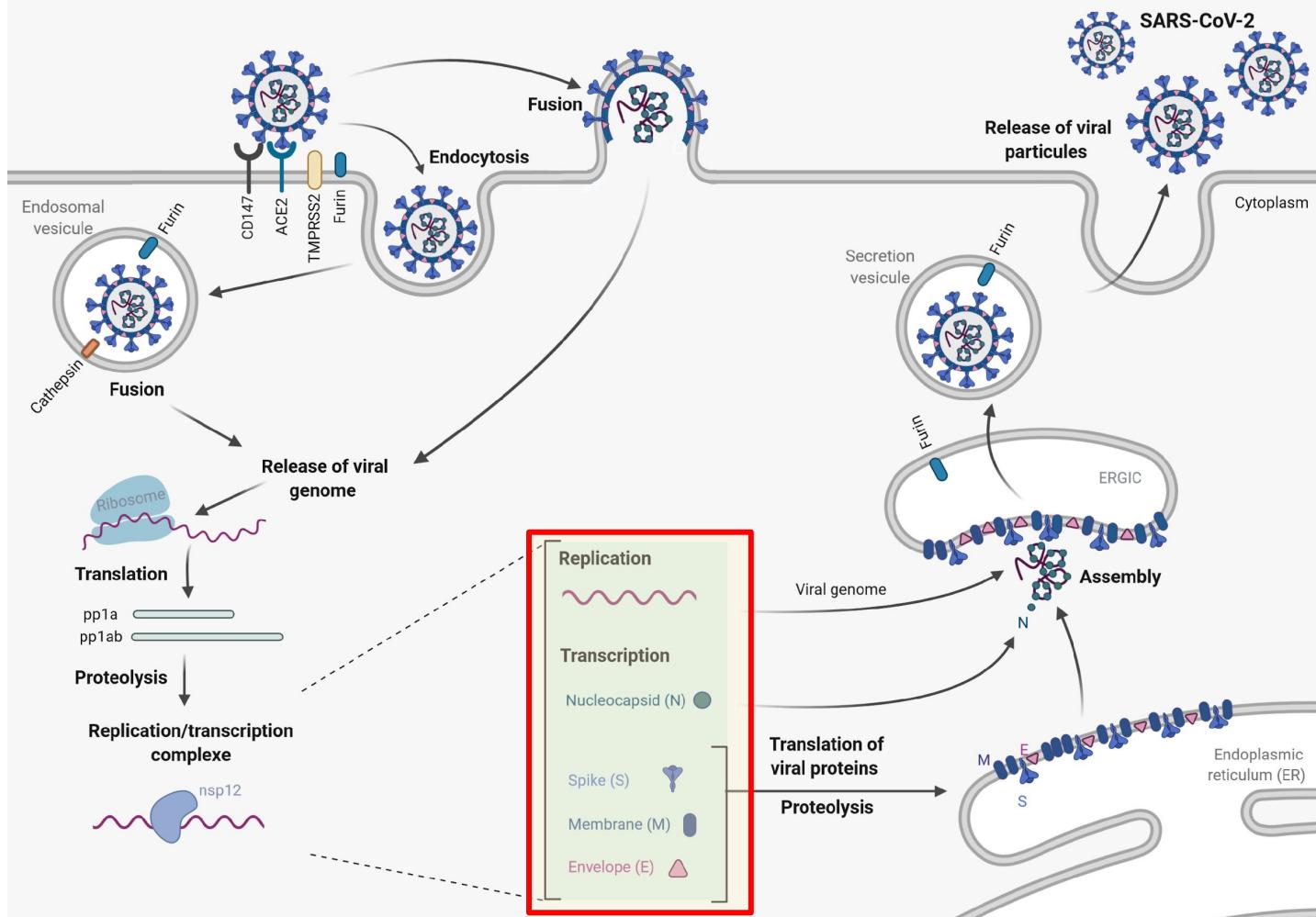
Created by Jonathan Khao, Ph.D. & Gaël McGill, Ph.D.
Digizyme Inc.
www.digizyme.com

Modeled & Simulated with Molecular Maya (Modeling & Rigging kits)
www.clarafi.com/tools/mmaya

We wish to thank Bing Chen, Ph.D. and Stephen Harrison, Ph.D. for their guidance and sharing data prior to publication.



Cycle de réplication des coronavirus - Réplication et transcription

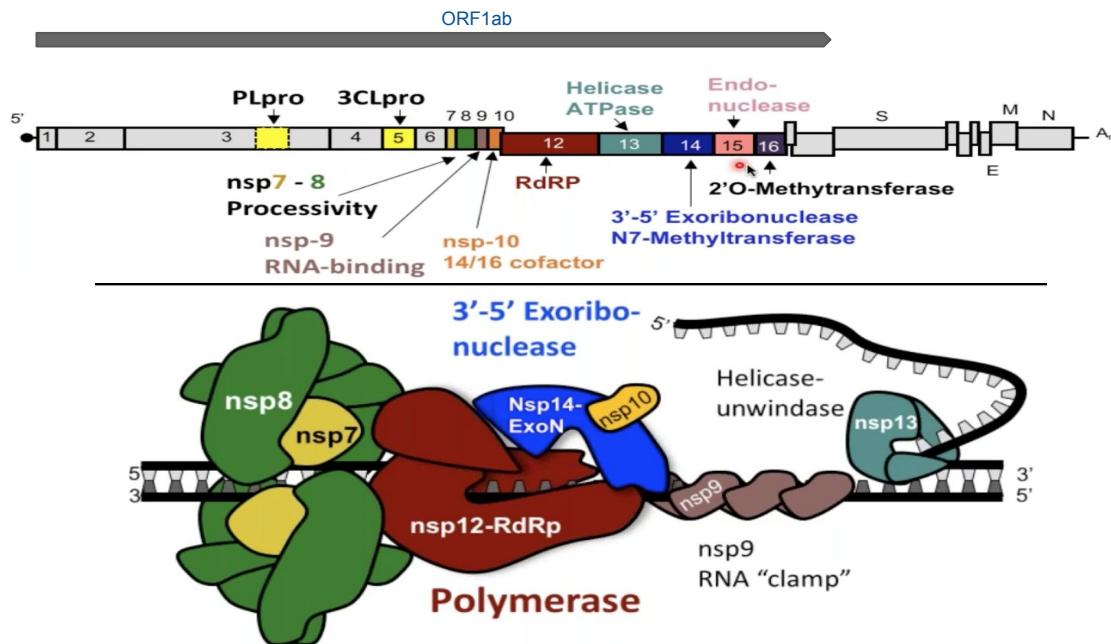


Le complexe réplication/transcription

SARS-CoV-2

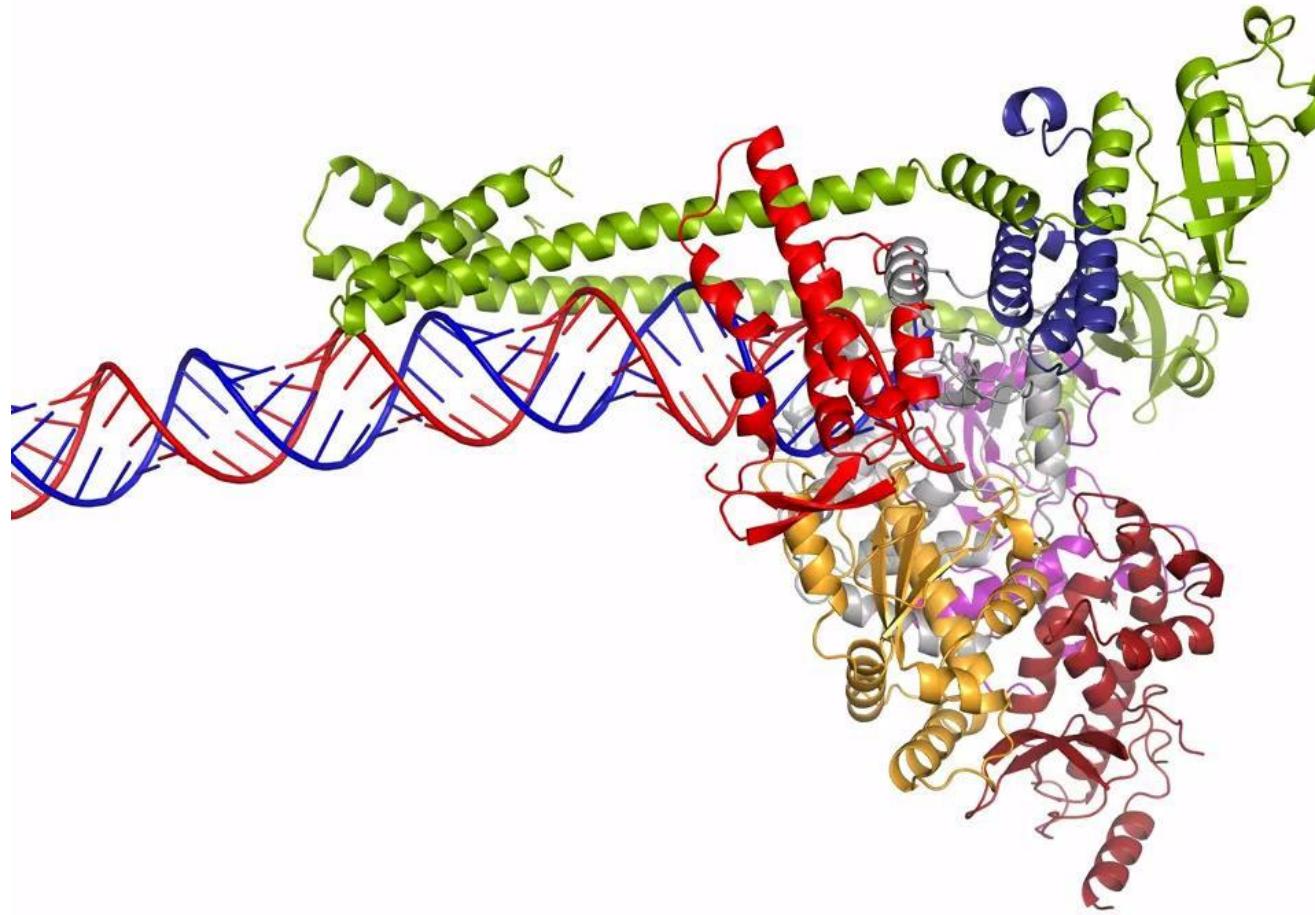


SARS-CoV

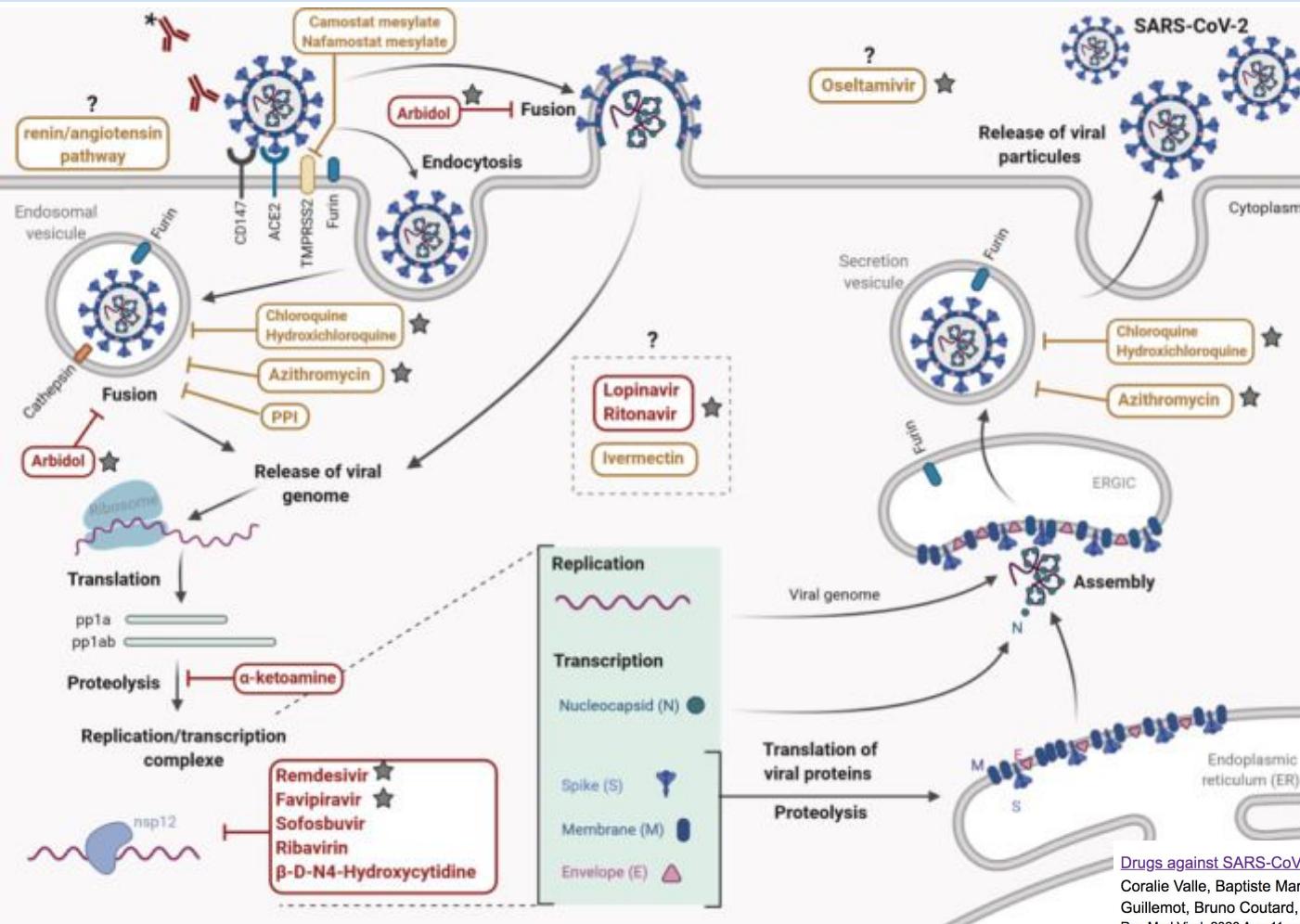


- Le gène ORF1ab code pour une “polyprotéine”, qui contient 16 protéines distinctes
- Une dizaine de ces protéines forment un complexe (figure du bas) qui assure la réplication et la transcription de l’ARN.

SARS-CoV : le complexe réplication/transcription



Mécanismes d'action des médicaments contre le SARS-CoV-2



- Des études sont menées dans un grand nombre de laboratoires pour identifier des molécules qui pourraient bloquer l'infection par SARS-CoV-2.
- Ces études partent de médicaments antiviraux connus, qui interagissent à différents niveau du cycle infectieux des virus.

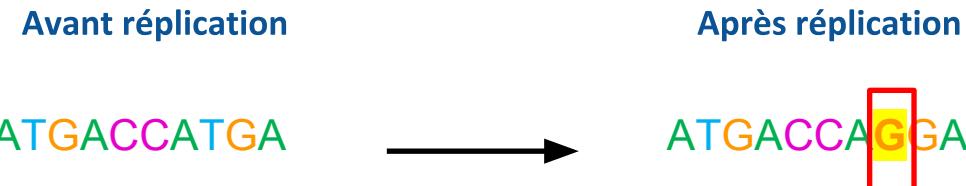
Drugs against SARS-CoV-2: What do we know about their mode of action?

Coralie Valle, Baptiste Martin, Franck Touret, Ashleigh Shannon, Bruno Canard, Jean-Claude Guillemot, Bruno Coutard, Etienne Decroly
 Rev Med Virol. 2020 Aug 11 : e2143. doi: 10.1002/rmv.2143 [Epub ahead of print]

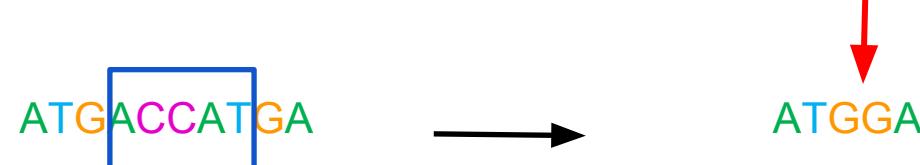
Événements évolutifs

Typologie des mutations

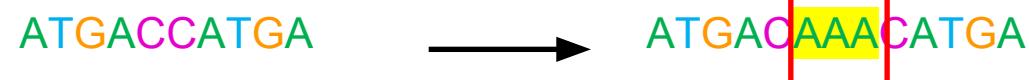
- Substitution
 - Remplacement d'un résidu (une lettre) par un autre



- Délétion
 - Perte d'un fragment de la molécule

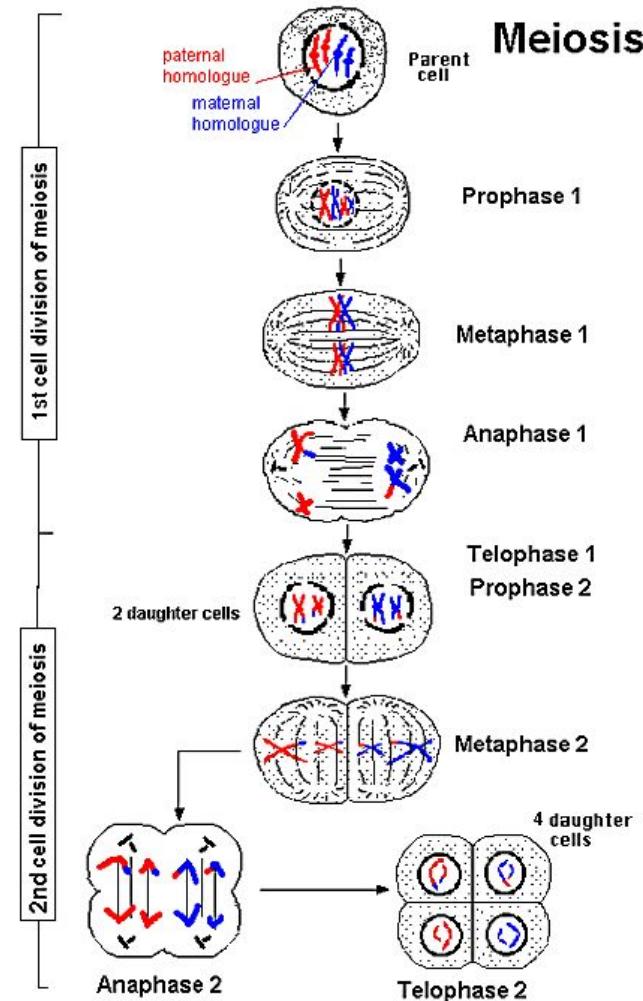


- Insertion
 - Ajout d'un fragment de molécule



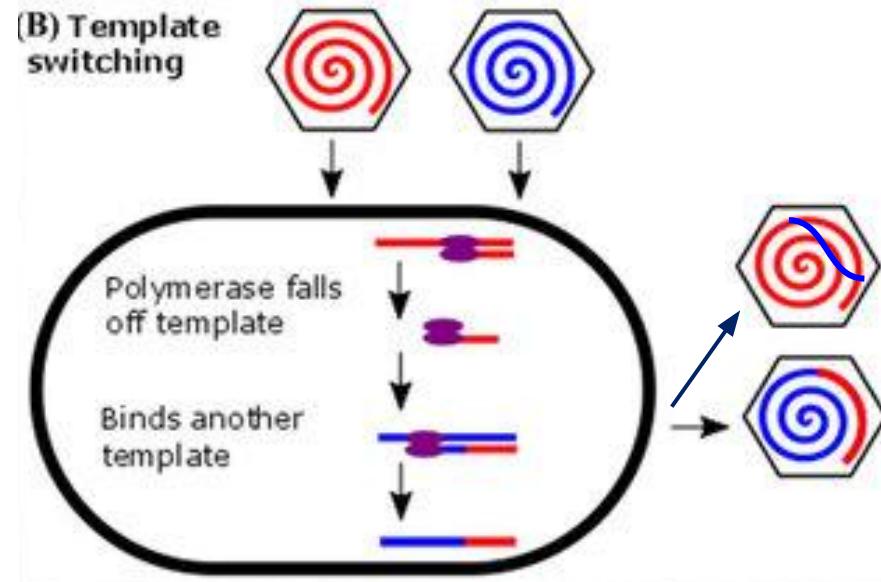
Recombinaison

- Chez les organismes cellulaires, lors de la méiose, une cellule mère diploïde forme 4 cellules-filles haploïdes.
- Les chromosomes parentaux sont distribués aléatoirement et de façon indépendante entre les 4 cellules-filles.
- Des événements de “crossing-over” (croisements) provoquent une recombinaison de fragments de chromosomes.
- La liaison génétique entre les gènes d'un même chromosome n'est pas complète.



Recombinaisons chez les coronavirus

- Une chauve-souris peut se retrouver infectée par plusieurs coronavirus en même temps.
- Pendant la réPLICATION des coronavirus, il arrive que la polymérase de l'ARN "saute" d'un virus à l'autre.
- Ceci donne naissance à un virus "chimérique", dont le génome est composé de fragments d'origines différentes.
- Ceci complique l'analyse de la phylogénie des virus, car différents fragments génomiques résultent d'histoires évolutives différentes.

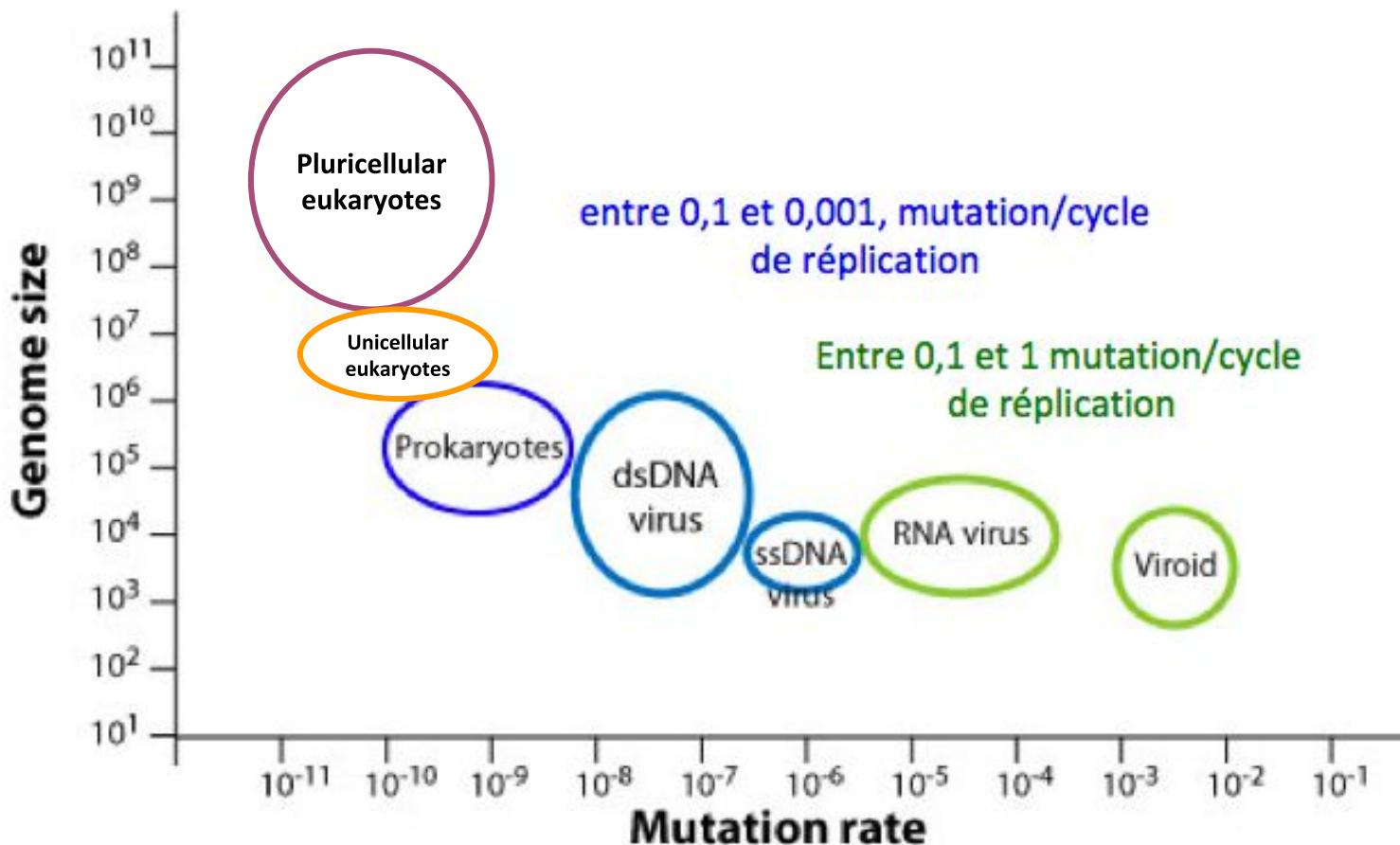


Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus

Ben Hu ,et al Plos path : November 30, 2017

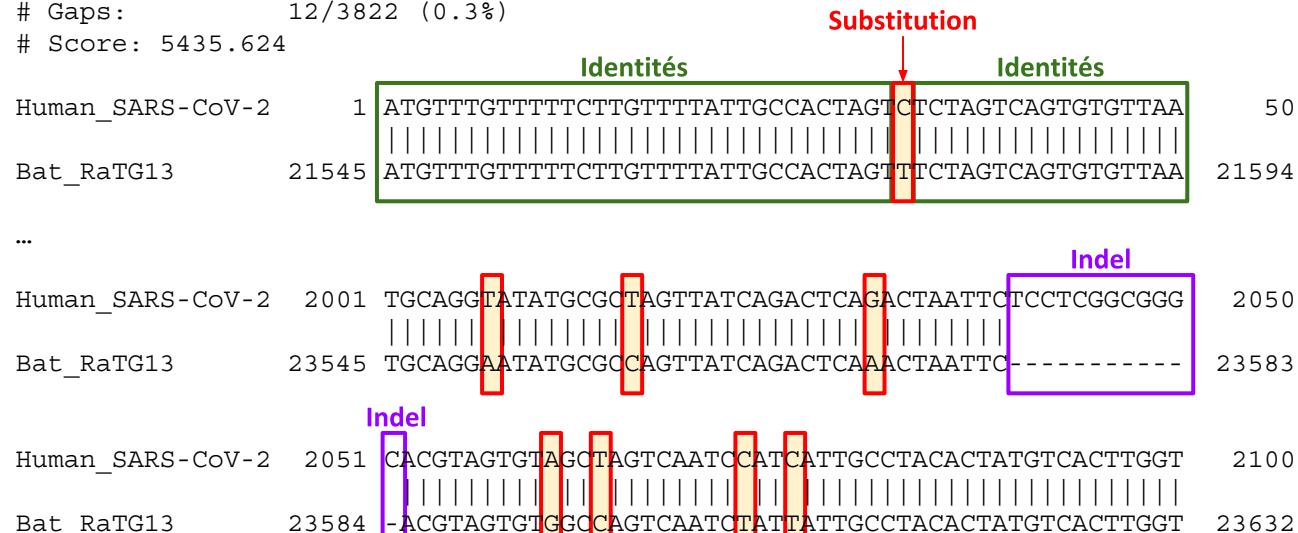
J. Dennehy, Evolutionary ecology of virus emergence: Virus emergence, 2016,
Annals of the New York Academy of Science

Taux de substitutions représentatifs chez différents groupes taxonomiques



Alignement de séquences – Gènes S de SARS-CoV-2 et RaTG13

```
# Aligned_sequences: 2
# 1: Human_SARS-CoV-2_BetaCoV/Wuhan/IPBCAMS-WH-01/2019
# 2: Bat_RaTG13
#
# Length: 3822
# Identity: 3549/3822 (92.9%)
# Similarity: NA/3822 (NA%)
# Gaps: 12/3822 (0.3%)
# Score: 5435.624
```

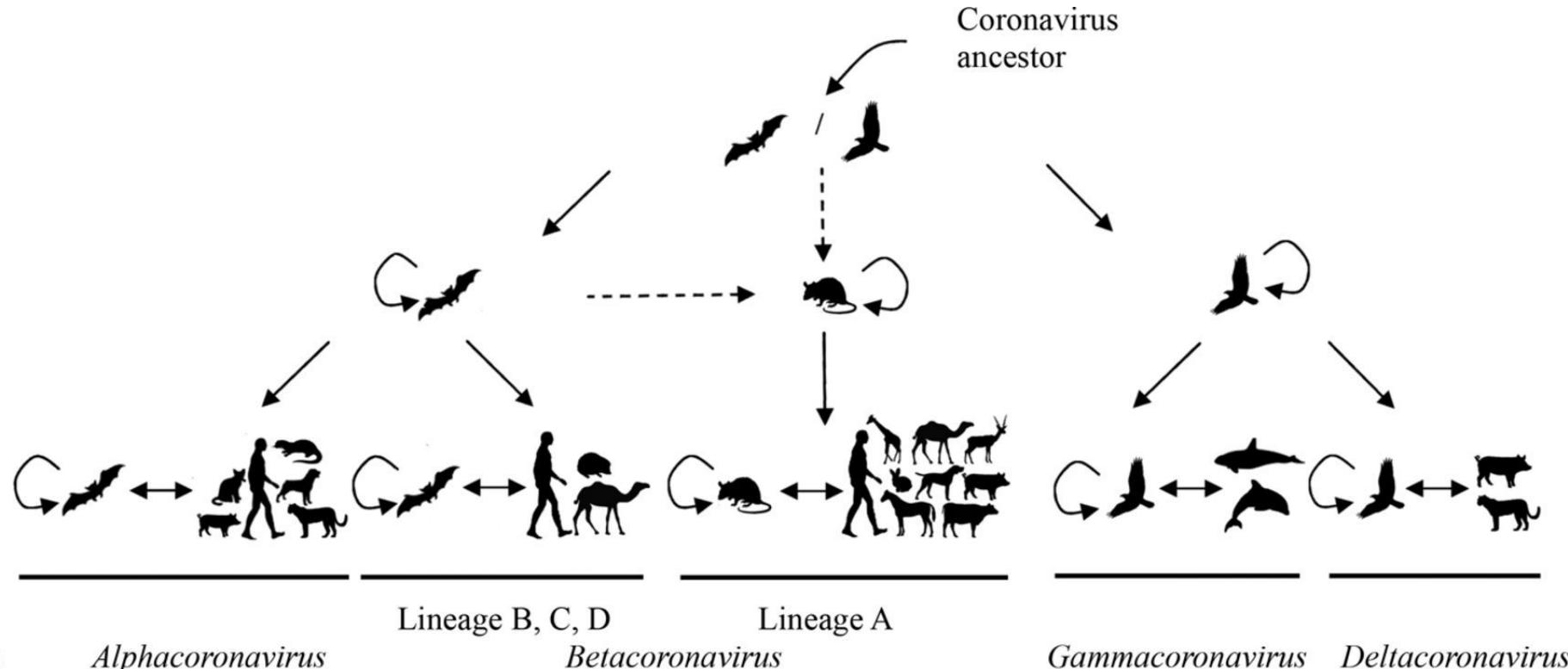


Note

- “Indel” signifie “Insertion ou délétion”
- Sur base de ce résultat, la différence observée peut provenir soit d'une insertion chez un ancêtre de SARS-CoV-2, soit d'une délétion chez un ancêtre de RaTG13

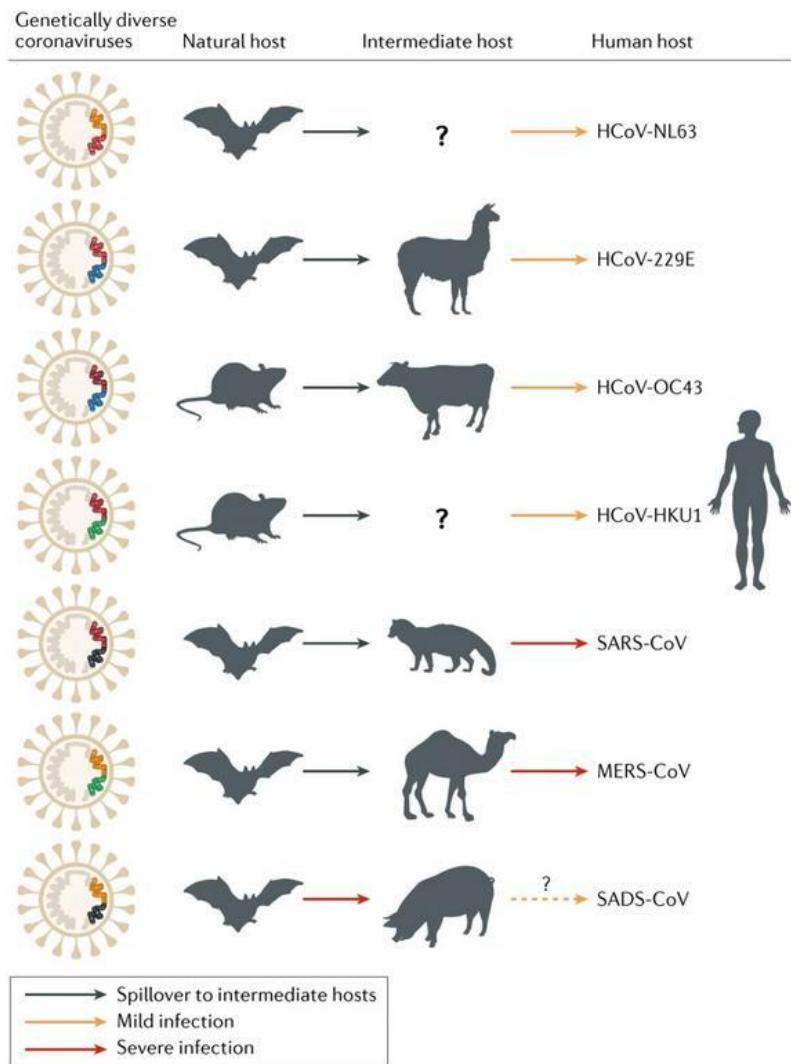
Des chauves-souris et des hommes

Cycle zoonotique des coronavirus



- Modèle du “débordement”
- Modèle de la circulation de « quasi espèces »

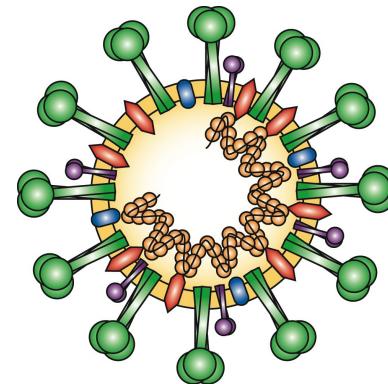
Mécanismes d'émergence des CoV humains



Origin and evolution of pathogenic coronaviruses. Jie Cui, Fang Li & Zheng-Li Shi.

Nature Reviews Microbiology volume 17, pages 181–192(2019)

Emergences de coronavirus humains



Common cold
OC43 can infect
lower respiratory

track
229E

OC43

HKU1; Pneumonia
NL63; Bronchiolitis

SARS-CoV
HKU1
NL63

MERS-CoV

2019-nCoV

1967/1970

1980

2002 2004/2005

2012

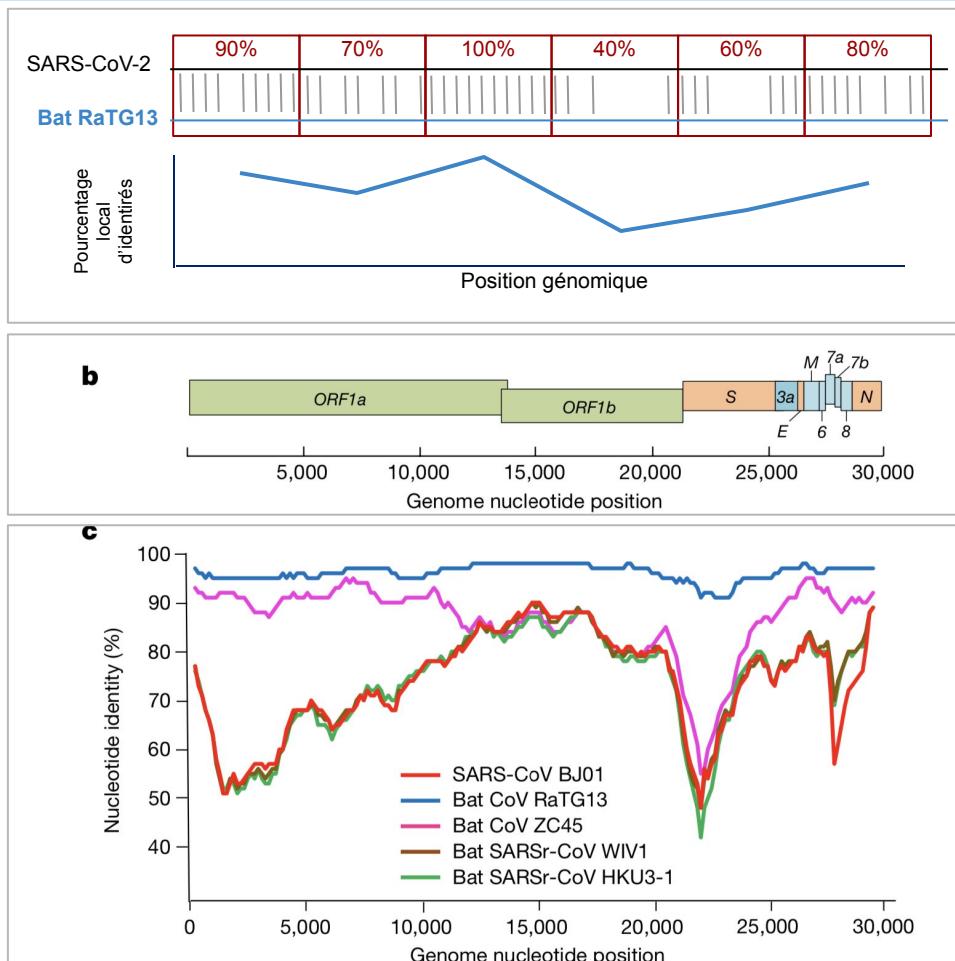
2019

OC43 genome similar to
Bovine coronavirus

SARS-CoV. MERS-CoV, 2019-nCoV
Severe respiratory disease

Profils de positions identiques (PPI)

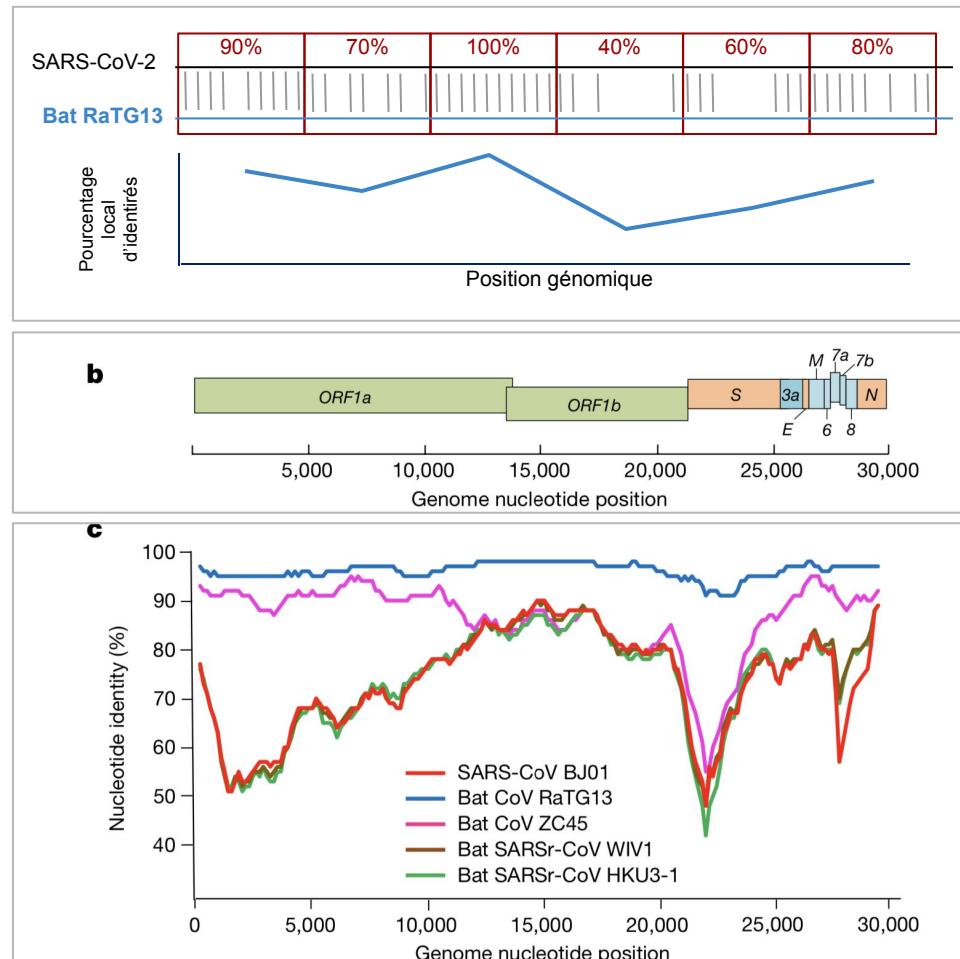
- Haut: principe de calcul du PPI
 - alignement d'une paire de séquences
 - découpage de la séquence en "fenêtres"
 - calcul du pourcentage local d'identité de chaque fenêtre
 - dessin du profil de positions identiques (PPI)
- Milieu : positions des gènes de SARS-CoV-2 sur le génome
- Bas: PPI de quelques génomes de coronavirus sur celui de SARS-CoV-2
- Commentaires dans la diapo suivante



Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.

Profils de positions identiques (PPI) de génomes de coronavirus

- RaTG13 (virus de chauve-souris) est le génome le plus proche de SARS-CoV-2
- On a identifié d'autres virus de chauve-souris relativement proches de SARS-CoV-2 (Cov ZC45)
- Les virus SARS-CoV humains (pandémie 2002-2003) sont moins proches
- Pour chaque espèce, on observe des fluctuations le long du profil de PPI
- Entre 22.000 et 25.000 : chute brutale des PPI



Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.

Une origine probable: la chauve-souris

- 3 février 2020: publication du génome complet de SARS-CoV-2
- Recherche de virus similaires dans les bases de données de séquence
 - Les virus les plus proches sont des virus de chauves-souris (Bat CoV ZC45)
- Dans le même article, les auteurs décrivent un nouveau génome de chauve-souris:
RaTG13
 - A ce jour la souche virale la plus proche de SARS-CoV-2 connue
- Figures du bas: profil de positions identiques (PPI), expliqué ci-après.

Article

A pneumonia outbreak associated with a new coronavirus of probable bat origin

<https://doi.org/10.1038/s41586-020-2012-7>

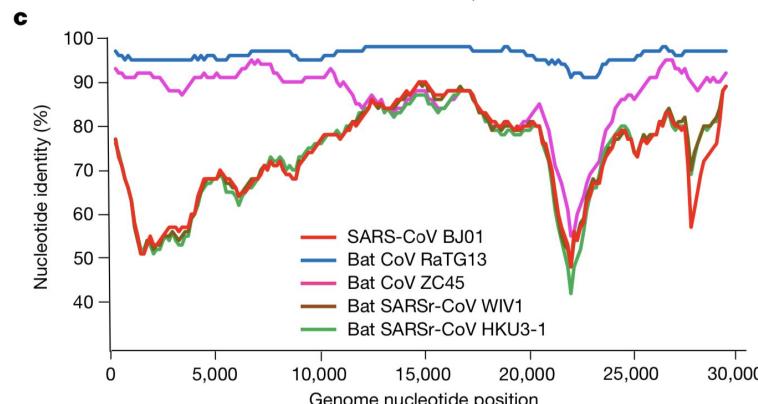
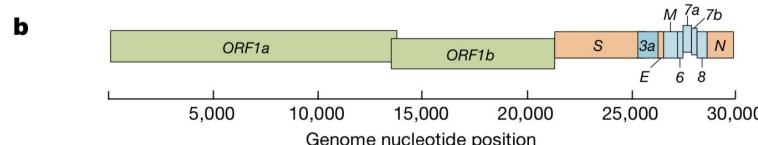
Received: 20 January 2020

Accepted: 29 January 2020

Published online: 3 February 2020

Open access

Peng Zhou^{1,5}, Xing-Lou Yang^{1,5}, Xian-Guang Wang^{2,5}, Ben Hu¹, Lei Zhang¹, Wei Zhang¹, Hao-Rui Si^{1,3}, Yan Zhu¹, Bei Li¹, Chao-Lin Huang², Hui-Dong Chen², Jing Chen^{1,3}, Yun Luo^{1,3}, Hua Guo^{1,3}, Ren-Di Jiang^{1,3}, Mei-Qin Liu^{1,3}, Ying Chen^{1,3}, Xu-Rui Shen^{1,3}, Xi Wang^{1,3}, Xiao-Shuang Zheng^{1,3}, Kai Zhao^{1,3}, Quan-Jiao Chen¹, Fei Deng¹, Lin-Lin Liu⁴, Bing Yan¹, Fa-Xian Zhan¹, Yan-Yi Wang¹, Geng-Fu Xiao¹ & Zheng-Li Shi^{1,3*}



Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.

De Yunnan à Wuhan

- 2013
 - Pneumonie atypique chez 6 mineurs dans la province de Yunnan, 3 décès
 - Plusieurs pistes sont évoquées (levures, virus) dont un coronavirus
 - Collecte d'échantillons de chauves-souris dans la mine
- 2016 : publication d'un fragment de séquence (360 nucléotides, 1% du génome) de virus de *Rhinolophus affinis*
- 2018 : dépôt des fragments de séquençage (reads) dans une base de données, pour ~90% du génome
- 2020 :
 - publication de la séquence complète du génome viral, sous l'identifiant “**BatCoV RaTG13**”.
 - Ce génome est le plus proche connu de celui de SARS-CoV-2 (96,2% nucléotides identiques).
- Note: l'implication de coronavirus dans le décès des mineurs fait encore l'objet de débats

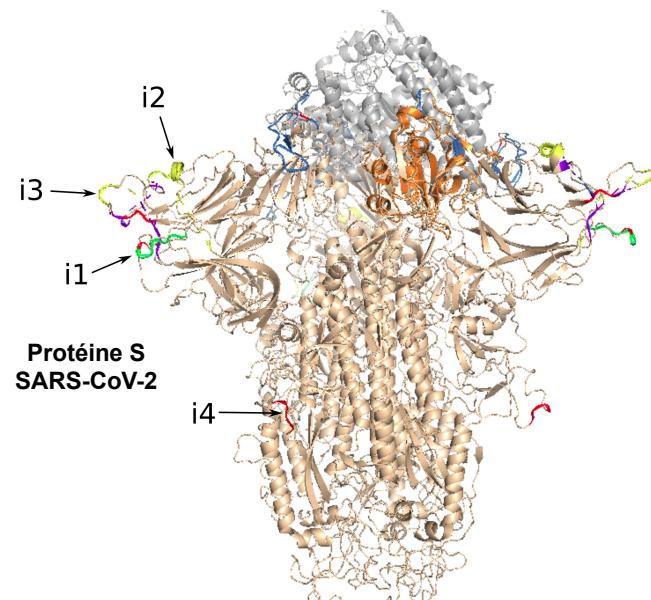
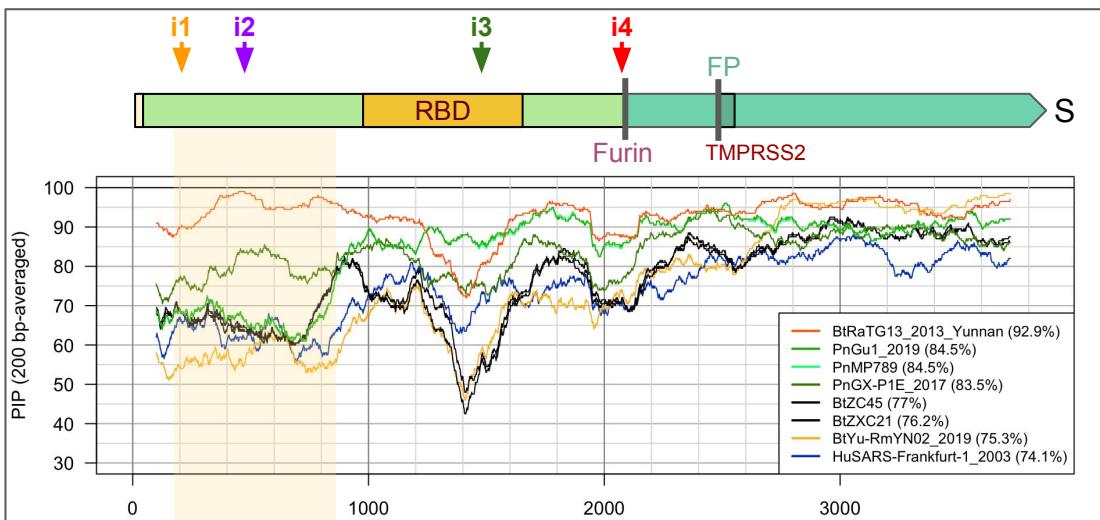


Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., et al. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579: 270–273.

Insertions dans les séquences du gène S

Quatre insertions dans le gène S de SARS-CoV-2

- Les flèches indiquent la position des 4 insertions sur le gène S (gauche) et sur la protéine spicule (droite).
- Les 3 premières sont situées à l'extérieur de la protéine, dans des régions “exposées”.



- Le site de clivage par la furine qu'on observe dans la protéine spicule de SARS-CoV-2 ne se trouve dans aucun autre coronavirus.
- Il résulte de l'insertion de 12 nucléotide à un endroit particulier du le gène S.

≡ EL PAÍS

CORONAVIRUS

ccu cgg cgg gca The 12 letters that changed the world

The genome of the new coronavirus harbors a short sequence suspected of being the main culprit of its uniquely infectious and aggressive nature



MANUEL ANSEDE | ARTUR GALOCHA | MARIANO ZAFRA

19 MAY 2020 - 18:25 CEST

Des insertions bizarres?

- Figure from Pradhan et al (2020), initially published on bioRxiv and retracted.
- The “multiple alignment” is actually a pairwise alignment + a consensus.
- The gaps obtained from a multiple alignment overlap with these ones, but they start and end at different positions.
- It is precisely because they did not do a multiple alignment that they did not realize that 3 of these insertions were not unique to SARS-CoV-2.

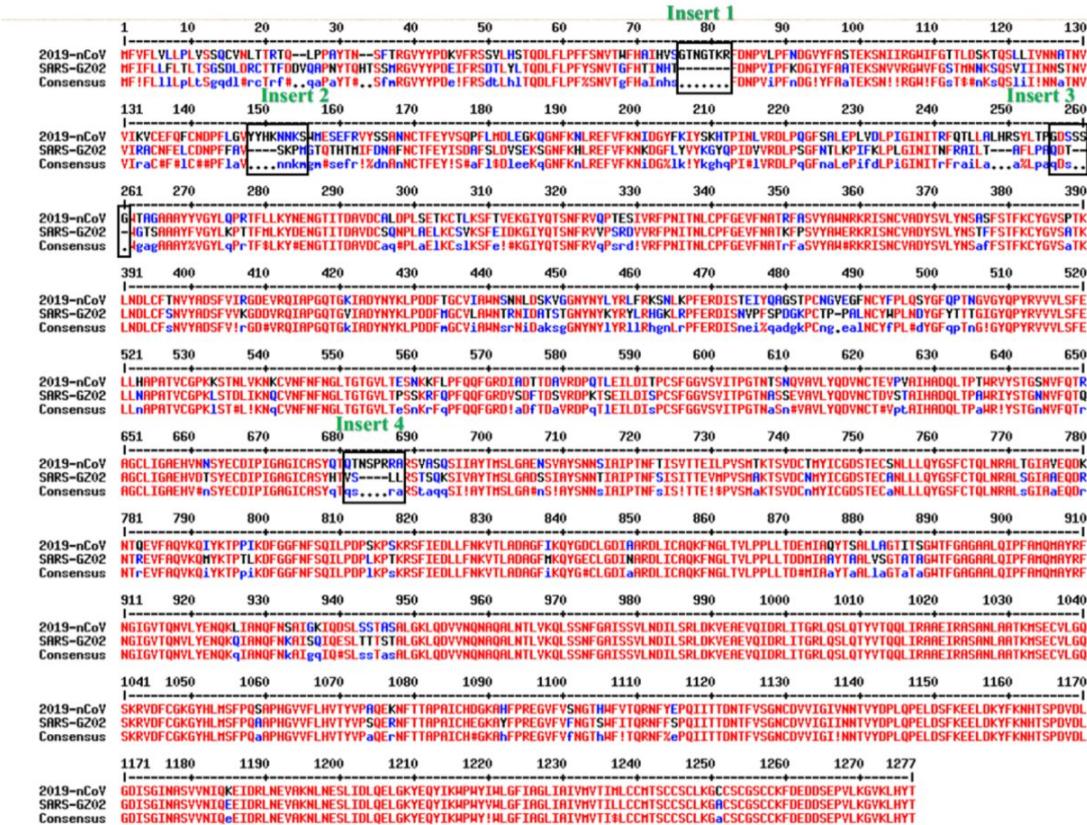
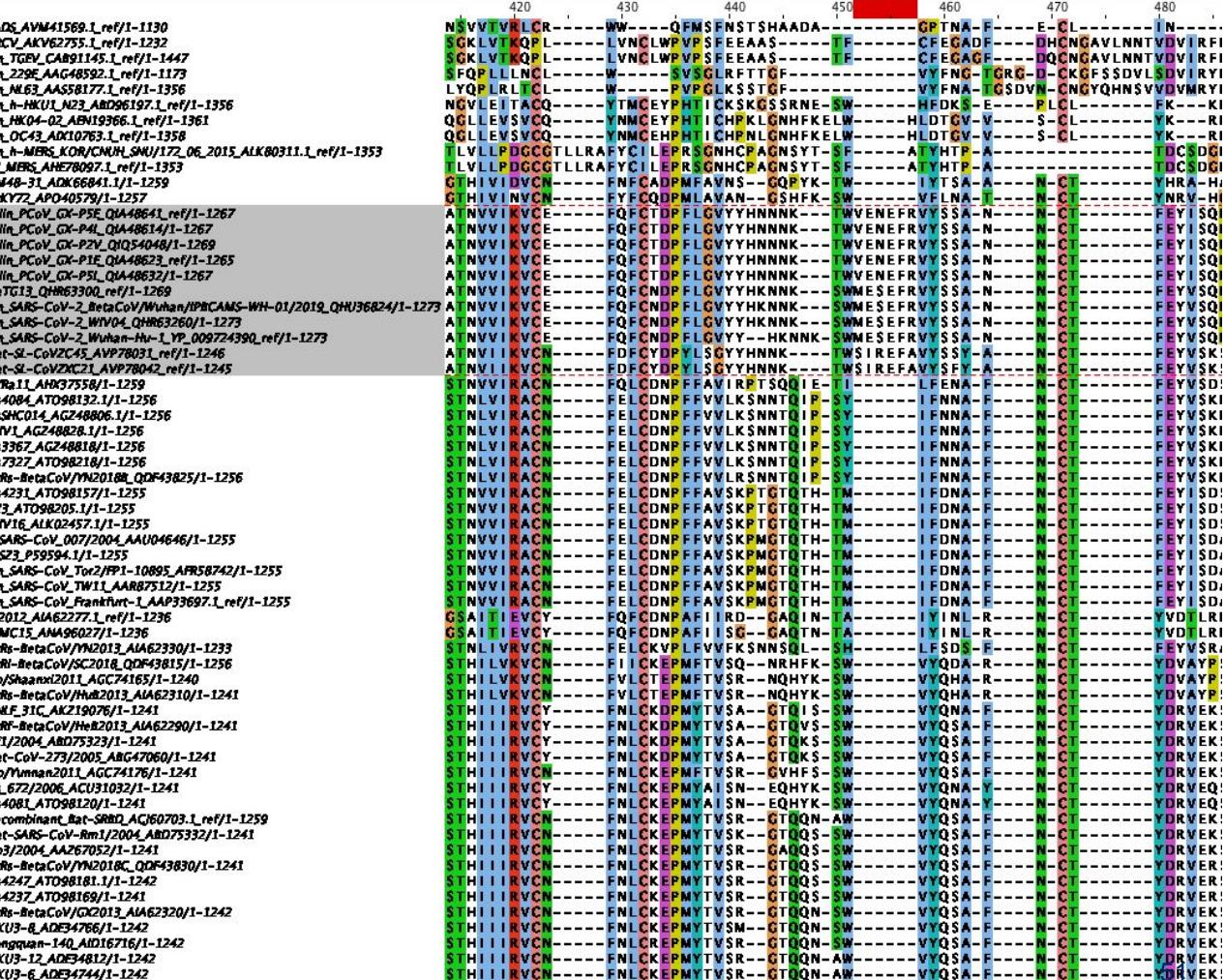


Figure 2: Multiple sequence alignment between spike proteins of 2019-nCoV and SARS. The sequences of spike proteins of 2019-nCoV (Wuhan-HU-1, Accession NC_045512) and of SARS CoV (GZ02, Accession AY390556) were aligned using MultiAlin software. The sites of difference are highlighted in boxes.

Insertion partagée entre tous les virus du groupe CoV-2

- Position: 153-158 de SARS-CoV-2
- Cette insertion se trouve chez les virus de pangolin + plusieurs chauve-souris
- Les résidus sont identiques entre SARS-CoV-2 et la souche RaTG13 de chauve-souris (la plus proche de SARS-CoV-2)
- Par contre elle présente 3 substitutions entre les souches de pangolin et SARS-CoV-2.



Insertion partagée par la majorité des virus du groupe CoV-2

- Position: 245-251 de SARS-CoV-2
 - Cette insertion se trouve chez les virus de pangolin + la souche RaTG13 de chauve-souris
 - Elle est cependant absente de 2 souches de chauves-souris appartenant au groupe CoV-2 : CoVZC45 et CoVZXC21
 - Au-delà de l'insertion on trouve un bloc conservé (jusqu'à la position 595 de l'alignement).
 - Au sein de ce bloc, une paire de résidus distingue les pangolins du groupe SARS2 + Bat RaTG13 .

M41_NL_1_ref/1-1130	QLDGQLQWERTV	-QFNNSPVNAGCHAIRFNVVVKDITSTLVETNSI
V62755_1_ref/1-1232	GTALKYGLTL-PPSVKEI	-AIKWKWGHFYINGYNNSFTPDCISFLNT--
CA891145_1_ref/1-1447	GTALKYGLL-PPSVKEI	-AISKWGHFYINGYNNSFTPDCISFLNT--
AAC485922_1_ref/1-1173	ETTSAFFVGAL-PKTVREF-VI	-BRTCHFYVINGYVFTLGNVNEAVNFVNNTA
AA558177_1_ref/1-1356	TTHYSFTFGV-LPTTVREIVARTGTY	-INGFKYVFDLGFLFVEAVNFVNNTA
U_123_AB06197_L_ref/1-1356	GTLLSHYYVPLTC	--NAISSNNETLQYWVTPLSKRQYLLKFDNF
U_2_ABW19366_L_ref/1-1361	GMALSHYYVMP LTC	--SRLDLCFTLE--YWVTPLSRQYLLAFNG
ADX10763_L_ref/1-1358	GMALSHYYVMP LTC	--SRRDLGFTLE--YWVTPLSRQYLLAFNG
CVN_CPNH_SAU/172_06_2015_ALK00311_1_ref/1-1353	YDTIKYSS1--PHS	-RSIQSDRKAW---AAFFYVYKLQPLTFLLDFSVI
AHE78097_1_ref/1-1353	YDTIKYSS1--PHS	-RSIQSDRKAW---AAFFYVYKLQPLTFLLDFSVI
ADK664841_1/-1-1259	GLNITQFKV1-MTL	--F-SSTQSNFDADASAYFVGHLKPLTMVLDFDE
APO40579/-1-1257	GLNITQFKV1-MTL	--F-SPTTSSFNADASAYFVGHLKPLTMVLDFDE
V_GX_PSE_QIA48641_ref/1-1267	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPKLESGWTTGAAAYYYGYLQRTFLLSYNQ
V_GX_PAE_QIA48614/-1-1267	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGNLESGWTTGAAAYYYGYLQRTFLLSYNQ
V_GX_P2V_QIA50448/-1-1269	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGKLESGWTTGAAAYYYGYLQRTFLLSYNQ
V_GX_PIE_QIA48623_ref/1-1265	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGKLESGWTTGAAAYYYGYLQRTFLLSYNQ
V_GX_P51_QIA48632/-1-1267	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGKLESGWTTGAAAYYYGYLQRTFLLSYNQ
QH063300_ref/1-1269	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGDSSSGWWTAGAAAAAAGGYLQRTFLLKYNE
CoV-2_Wuhan-Hu-1_YP_009724390_ref/1-1273	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGDSSSGWWTAGAAAAAAGGYLQRTFLLKYNE
CoV-2_WV04_QRH63260/-1-1273	GINITRFQTL-LALHRSYLT	-LALHRSYLTCPGDSSSGWWTAGAAAAAAGGYLQRTFLLKYNE
CoV-2_Wuhan-Hu-1_YP_009724390_ref/1-1246	SINITKFRTL-LTI	-HRGDPMNNNGWTAFSAAAYFVGYKLPRTFMLKYNE
v2ZC45_AVP78031_ref/1-1245	SINITKFRTL-LTI	-HRGDPMNNNGWTAFSAAAYFVGYKLPRTFMLKYNE
AM037558/-1-1259	GLNITNFRVL-LTA	--F-IPNITGWTGSVAYFVGYKLPTTFMLKYDY
AT098132/-1/-1256	GLNITNFRVL-LTA	--F-PPRDPDWGTSAAAAYFVGYKLPTTFMLKYDE
4_AG24880/-1/-1256	GLNITNFRVL-LTA	--F-PPRDPDWGTSAAAAYFVGYKLPTTFMLKYDE
Z48823_1/-1-1256	GLNITNFRVL-LTA	--F-PPRDPDWGTSAAAAYFVGYKLPTTFMLKYDE
AG248816/-1-1256	GLNITNFRVL-LTA	--F-PPRDPDWGTSAAAAYFVGYKLPTTFMLKYDE
AT098218/-1-1256	GLNITNFRVL-LTA	--F-PPRDPDWGTSAAAAYFVGYKLPTTFMLKYDE
aCoV/YN/2016/QD43825/-1-1256	GLNITNFRVL-LTA	--F-PPNPGWGTSAAAAYFVGYKLPTTFMLKYDE
AT098157/-1-1255	GLNITNFRVL-LTA	--F-LPAQDTWGTSSAAAAYFVGYKLPATFMLKYDE
98205_1/-1-1255	GLNITNFRVL-LTA	--F-LPAQDTWGTSSAAAAYFVGYKLPATFMLKYDE
LK20457_1/-1-1255	GLNITNFRVL-LTA	--F-LPAQDTWGTSSAAAAYFVGYKLPATFMLKYDE
CoV_007/2004_AA004646/-1-1255	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
9594_1/-1-1255	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
CoV_Tor2/PPI-10895_AFR58742/-1-1255	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
CoV_TW11_AAR87512/-1-1255	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
CoV_Frankfurt_1_AAP33697_1_ref/1-1255	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
IA62227_1_ref/1-1236	GIKINTNRAI-LTA	--F-SPAQDIWGTSAAAAYFVGYKLPTTFMLKYDE
AMA96027/-1-1236	GLNITNKKV-TTL	--KPTNQAF--QAAAYIVGNLKHMTMFLSFNE
aCoV/YN/2013_AA62330/-1-1233	GLNITNKKV-TTL	--KPTNQAF--QAAAYIVGNLKHMTMFLSFNE
aCoV/SC2018_ODF43815/-1-1256	GLNITNSFKTFL-LAV	-Y-RVAAGSISVASSAAYYYVGKYLPLTFMLSYDL
xx2011_AGC74165/-1-1240	GLNITGMRVV-MTM	-SNTQQNFLTENAAAYYYVGKYLPRTFMLQFNT
aCoV/Hu2013_AA62310/-1-1241	GLNITGMRVV-MTM	-SNTQQNFLTENAAAYYYVGKYLPRTFMLQFNT
_AK219076/-1-1241	GLNITSFRVV-MAM	-SKTTSNVPESAAAAYVGKYLPRTFMLQFNT
aCoV/HeB2013_AA62290/-1-1241	GLNITSFRVV-MAM	-SKTTSNVPESAAAAYVGKYLQSTFMLSFNO
_ABD75223/-1-1241	GLNITSFRVV-MAM	-SKTTSNVPESAAAAYVGKYLQSTFMLSFNO
273/2005_ABG47060/-1-1241	GLNITSFRVV-MAM	-SKTTSNVPESAAAAYVGKYLQSTFMLSFNO
2011_AGC74176/-1-1241	GLNITSFKVV-MTM	-Y-QTTTSNLFPESESAAAAYVGKLYVTTFMFQFNE
2006_ACU31032/-1-1241	GLNITSFKVV-MSM	-RRTTSNLFPEVAAAYVGKLYSTFMLNFE
AT098120/-1-1241	GLNITSFKVV-MSM	-RRTTSNLFPEVAAAYVGKLYSTFMLNFE
Want_Bat-SMBI_AC60703_1_ref/1-1259	GLNITSYRVV-MAM	-QTTTSNLFPEIAYVGKLYSTFMLRFE
5-CoV-Bm1/2004_ABD75332/-1-1241	GLNITSYRVV-MTM	-SQFNSNLFPEESAAYYYVGKLYTTFMLRFE
4_AA07052/-1-1241	GLNITSYRVV-MAM	-QTTTSNLFPEESAAYYYVGKLYTTFMLRFE
aCoV/YN/2018/QD49830/-1-1241	GLNITSYRVV-MAM	-QTTTSNLFPEESAAYYYVGKLYTTFMLRFE
AT098181_1/-1-1242	GLNITSYRVV-MAM	-QTTTSNLFPEESAAYYYVGKLYTTFMLRFE
AT098169/-1-1241	GLNITSYRVV-MAM	-QTTTSNLFPEESAAYYYVGKLYTTFMLRFE
aCoV/GX/2013_AIA62320/-1-1242	GLNITSYRVV-MAM	-SQSTSNSNLFPEESAAYYYVGKLYSTFMLRFE
AD05476/-1-1242	GLNITSYRVV-MAM	-SQSTSNSNLFPEESAAYYYVGKLYSTFMLRFE
n-140_A016716/-1-1242	GLNITSYRVV-MAM	-SQSTSNSNLFPEESAAYYYVGKLYSTFMLRFE
_AD04812/-1-1242	GLNITSYRVV-MAM	-SQSTSNSNLFPEESAAYYYVGKLYSTFMLRFE
AD054744/-1-1242	GLNITSYRVV-MAM	-SQSTSNSNLFPEESAAYYYVGKLYSTFMLRFE

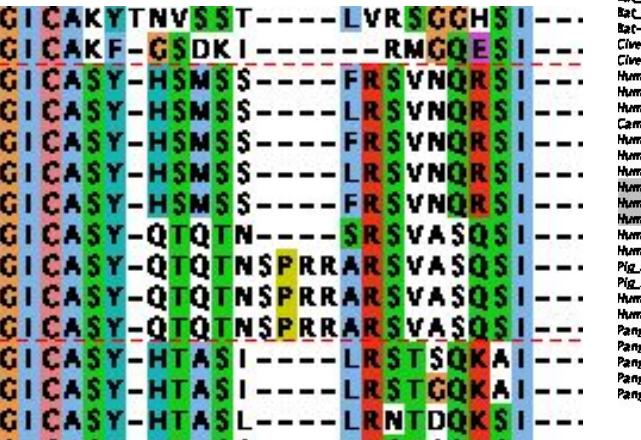
Insertion i3

- Position
 - 470-486 de SARS-CoV-2
 - 855-872 sur l'alignement
- Commune au groupe pangolin + Bat_RaTG13 + SARS-CoV-2
- 2 substitutions uniques à Bat_RaTG13



Insertion d'un site Furine (i4)

- Positions : 1181-1184 de l'alignement
 - On trouve chez SARS-CoV-2 un site unique SPRRAR, qui résulte d'une insertion SRR et d'une substitution L -> A
 - Ce site correspond au motif reconnu par la furine (protéase).

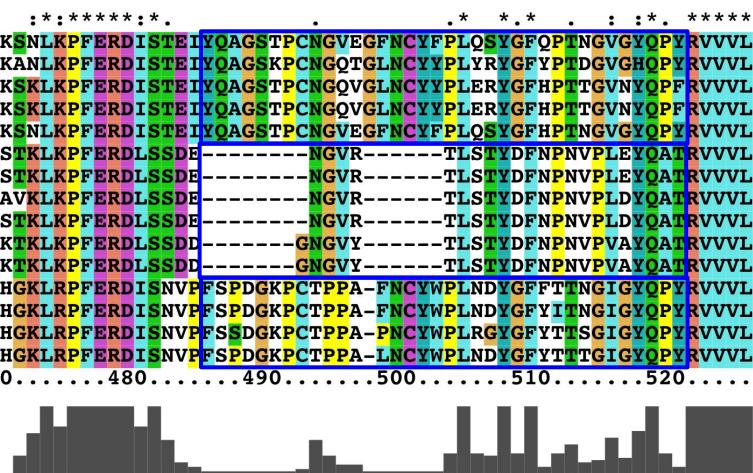


V_AUG48828/1-1256
448_31_ATG656841/1-1259
2012_AIA622771.ref/1-1296
1/2004_AB075323/1-1241
Rn-BetaCoV/MN2013_AA622330/1-1233
672/2006.ACUS1032/1-1241
3367_AZG248818/1-1256
4081_ATD98120/1-1241
4084_ATD98132/1-1256
4231_ATD98157/1-1255
4232_ATD98169/1-1241
4247_ATD98181/1-1242
7322_ATD98216/1-1256
SHC014_AZG248806/1-1256
3_AT098205.1/1-1255
V16_ALK02457.1/1-1255
V/Funn2011_AGC74176/1-1241
Rn-BetaCoV/VN2018C_QDF43825/1-1256
Rn-BetaCoV/VN2018C_QDF43830/1-1241
Ra11_ANH37558/1-1259
TG12_OHR633300.ref/1-1269
tr-CoV-27/2005_ABG47060/1-1241
Rn-BetaCoV/HeB2013_AA62290/1-1241
MC15_AMA96027/1-1236
Mf_3IC_AK219076/1-1241
r-SARS-CoV-Rm1/2004_AB075332/1-1241
3/2004_AA267052/1-1241
/Shaanxi2011_AGC74165/1-1240
Rn-BetaCoV/CQ2013_AA62320/1-1242
Rn-BetaCoV/HuB2013_AA62310/1-1241
(U3-12_ADE34812/1-1242
(U3-6_ADE34744/1-1242
(U3-8_ADE34766/1-1242
KY72_AP040579/1-1257
Rn-BetaCoV/SC2018_QDF43815/1-1256
combinant_Rat-SRD₅AG67073.L.ref/1-1259
CoV_CoV_007/2004_AA040646/1-1255
S23_P59594/1-1235
h_HKU1_N23_AB096197.L.ref/1-1356
h_NK04_02_AB193966.L.ref/1-1361
h_MERS_KOR_CMGH_SNU/172/2016_ALIK80311.I.ref/1-1353
MERS_AB070097/1-1353
Sars-CoV_Frankfurt-1_AA3P3697_L.ref/1-1255
Sars-CoV_Tur2/PP1-10895_ARF58742/1-1255
Sars-CoV_TW1L_AAR87512/1-1255
Sars-CoV-2_betaCoV_Wuhan/PRCAMS-WH-01/2019_QHU936824/1-1279
Sars-CoV-2_WIV10A_QHM63260/1-1273
Sars-CoV-2_Wuhan/PRCAMS-WH-0072490.ref/1-1273
_229_AAG65922_L.ref/1-1173
_ME6_AAS58177.L.ref/1-1356
DS_AV4M41569.L.ref/1-1130
CV_AKV622755.L.ref/1-1232
1_TGPV_CAB911455.L.ref/1-1447
1_OC43_A0010763.L.ref/1-1358
In_PCoV_GX-P16_QAA48623.ref/1-1265
In_PCoV_GX-P2V_Q0354048/L-1269
In_PCoV_GX-P4L_QAA48614.ref/1-1267
In_PCoV_GX-P5E_QAA48641.ref/1-1267
In_PCoV_GX-P5L_QAA48632/1-1267

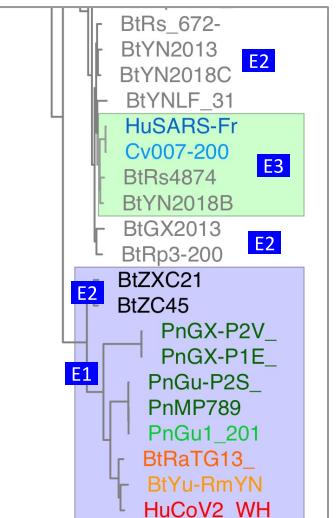
Un site recombinant?

- L'interprétation de ce site est plus complexe.
- Il existe clairement trois groupes de séquences.
- Ceux-ci s'étendent au-delà des deux indels.
- L'arbre construit à partir de cette région est peu robuste, et incohérent avec celui des génomes.
- La répartition des sous-blocs de séquences est plus cohérente avec l'arbre des espèces.
- Cette région a échappé à Pradhan et al. parce qu'ils ont réalisé un alignement par paire SARS-CoV-2 vs SARS plutôt qu'un alignement multiple.

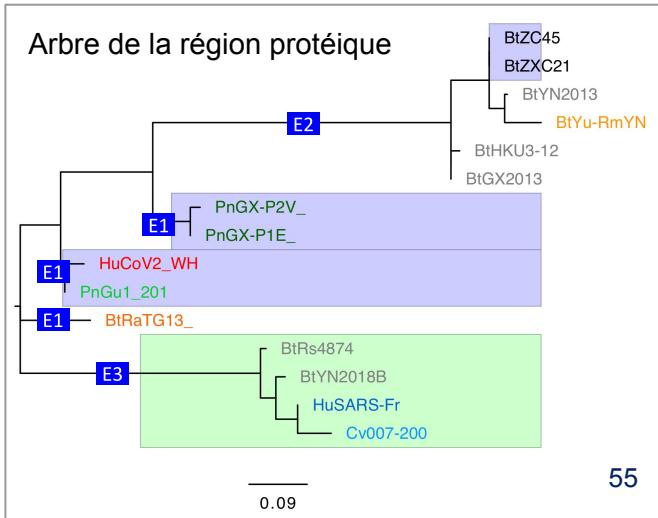
HuCoV2_WH01_2019
 BtRaTG13_2013_Yunnan
 PnGX-P1E_2017
 PnGX-P2V_2018
 PnGu1_2019
 BtZC45
 BtZXC21
 BtYu-RmYN02_2019
 BtYN2013
 BtHKU3-12
 BtGX2013
 BtYN2018B
 BtRs4874
 Cv007_2004
 HuSARS-Frankfurt-1_2003



Arbre des génomes



Arbre de la région protéique

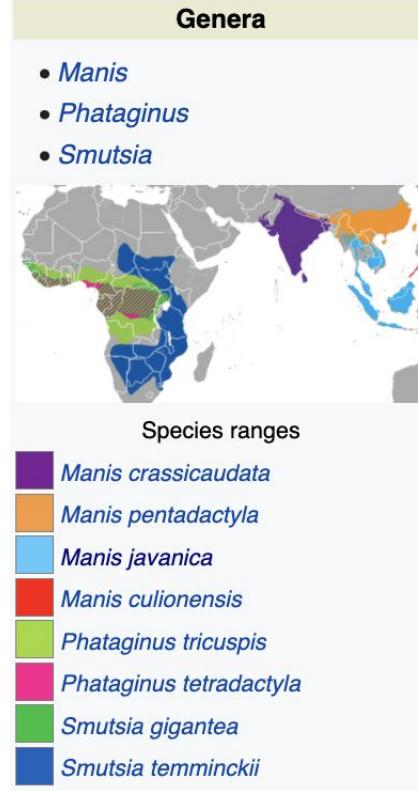


Des chauves-souris et des hommes ... et des pangolins ?

Pangolins

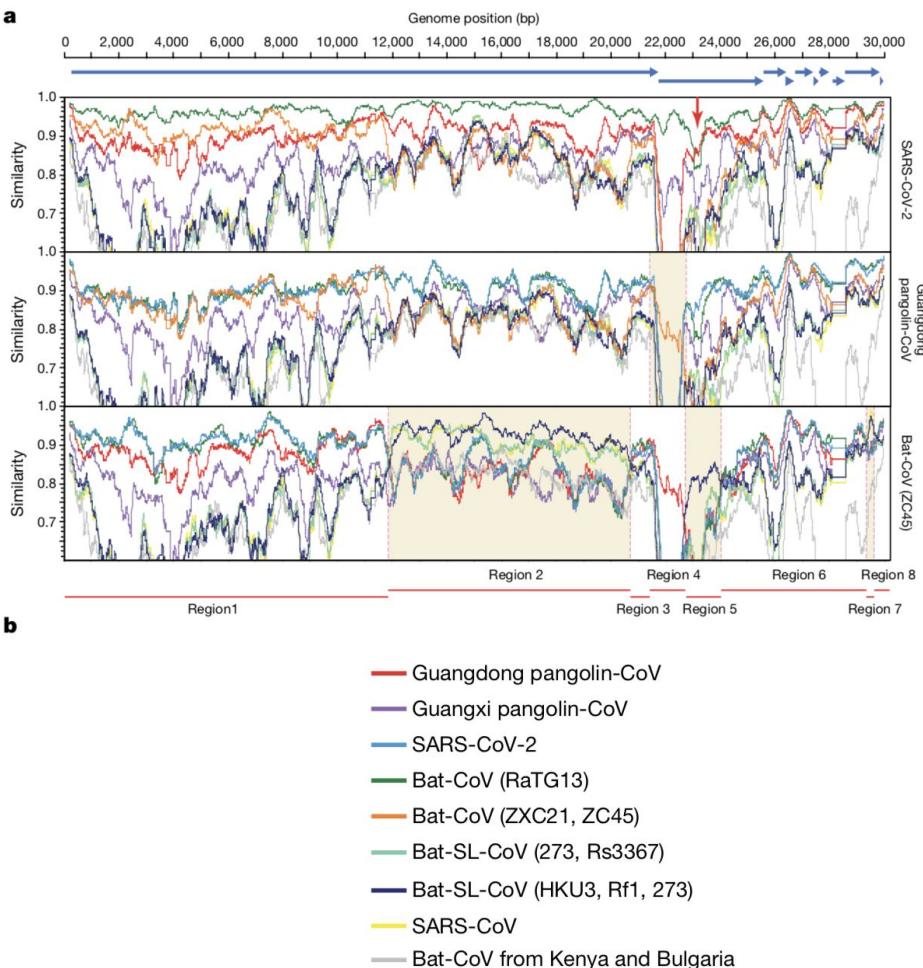
- Mode de vie
- Aire géographique
- Migration
- Contacts avec les chauves-souris
- Contacts avec l'humain

Répartition géographique de différentes espèces de pangolin



... et des pangolins ?

- Lam et collègues comparent le génome de SARS-CoV-2 à des génomes de virus isolés à partir de pangolins.
- Ces virus sont globalement plus éloignés de SARS-CoV-2 que ceux de chauve-souris.
- Cependant, on observe une identité plus élevée dans la région particulière où les PPI des autres coronavirus s'affaissent.
- Ceci suggère la possibilité d'une recombinaison entre des virus de chauve-souris et de pangolin.



Lam, T.T.-Y., Shum, M.H.-H., Zhu, H.-C., Tong, Y.-G., Ni, X.-B., Liao, Y.-S., et al. 2020.
Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature,
doi: 10.1038/s41586-020-2169-0.

Et si les pangolins n'y étaient pour rien ?

L'hypothèse du pangolin est fortement remise en cause pour plusieurs raisons.

- Les génomes des virus de pangolin les plus proches dont on dispose sont plus éloignés de SARS-CoV-2 que ceux des génomes de chauves-souris.
- La région où la similarité est la plus forte correspondent au gène S, qui code pour la protéine spicule. Or la protéine spicule de ces virus de pangolin n'est pas capable d'adhérer au récepteur de cellules humaines.

franceinfo:

vidéos | radio | jt | magazines

DIRECT TV

DIRECT RADIO



★ / replay radio / Le billet sciences

Covid-19 : et si le pangolin n'y était pour rien ?

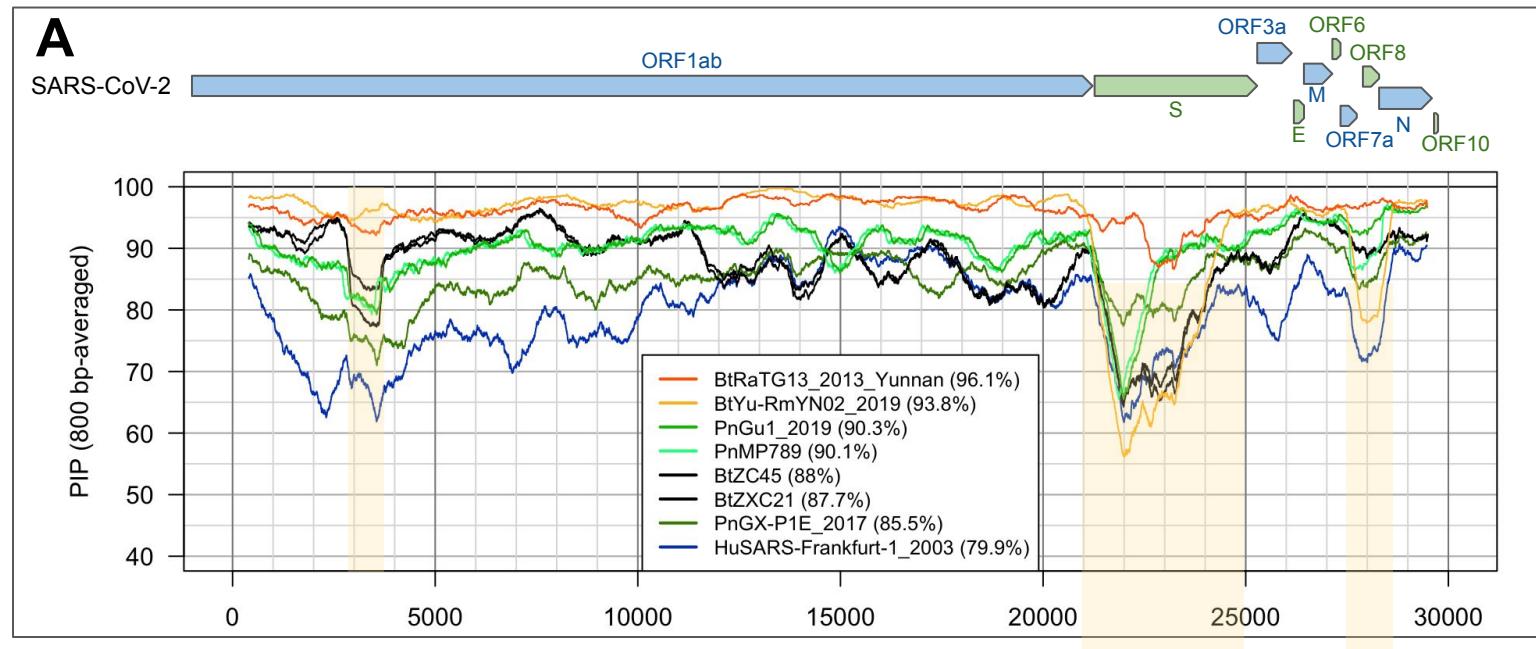
Les chercheurs ont peut-être accusé un peu trop vite le pangolin d'être à l'origine de la pandémie qui a débuté à Wuhan en Chine fin décembre 2019. Ce petit mammifère à écailles est de plus en plus disculpé par des études scientifiques.



Déetecter les recombinaisons par comparaison de génomes

Détection des recombinaisons génomiques

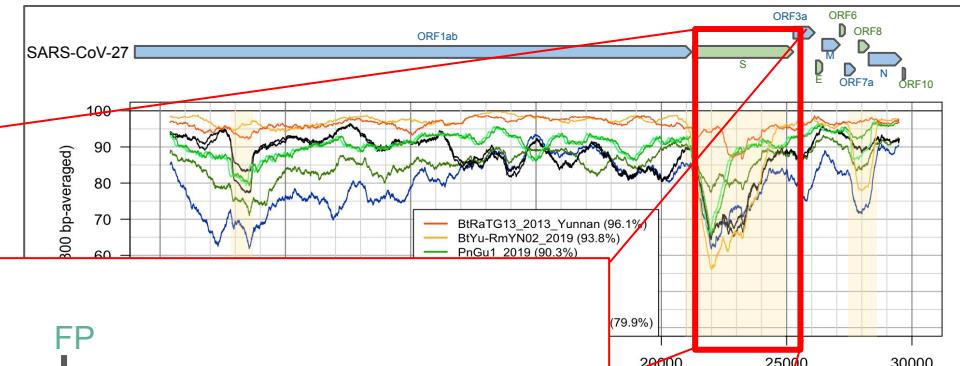
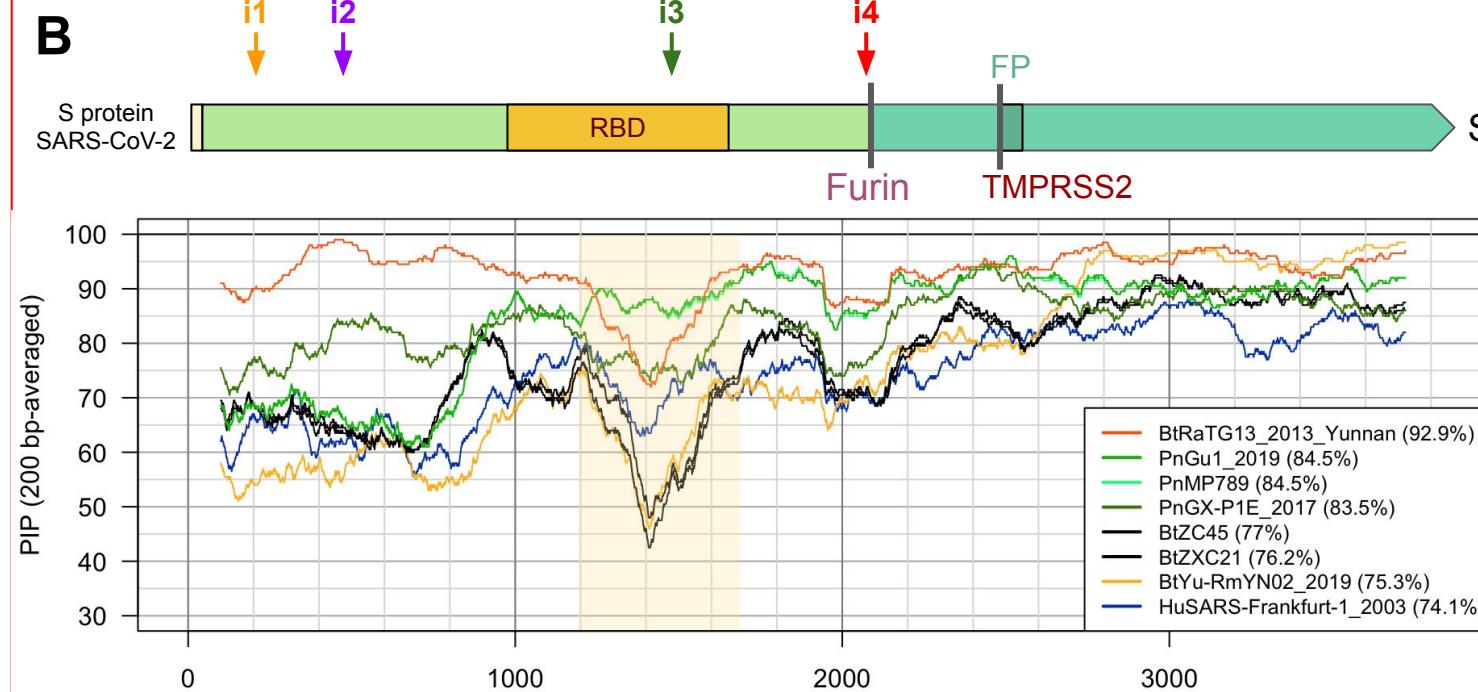
Les chutes brutales d'identité sur les profils PPI (fond jaune) dénotent des régions résultant vraisemblablement de recombinaisons.



- Sallard, E., Halloy, J., Casane, D., van Helden, J. & Decroly, É. 2020. Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *Med Sci (Paris)* 36: 783–796.
- English version : Erwan Sallard, José Halloy, Didier Casane, Etienne Decroly, Jacques van Helden. Tracing the origins of SARS-CoV-2 in coronavirus phylogenies. hal-02891455

Comparaison entre coronavirus - gène S

Profils de pourcentages de positions identiques (PPI) entre régions génomiques du gène S de différents coronavirus et SARS-CoV-2 (la référence à 100%).



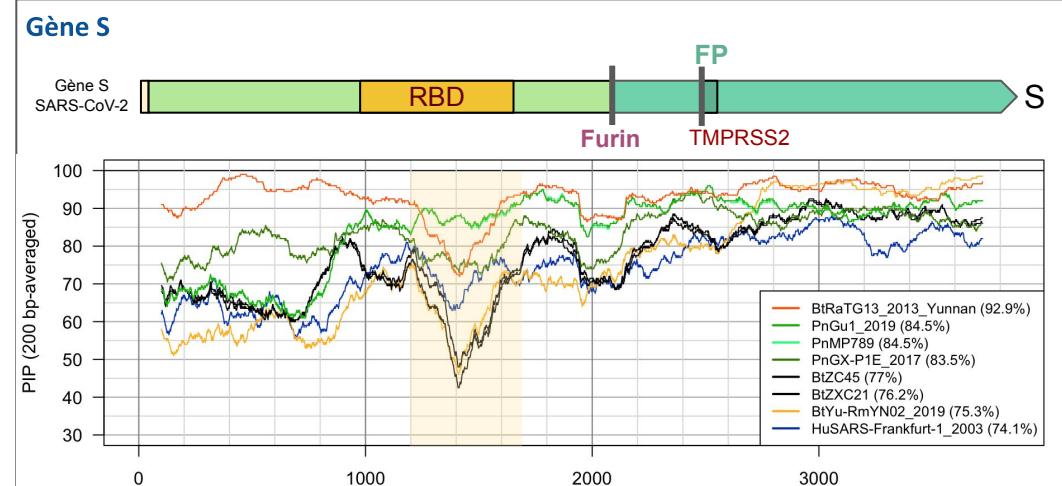
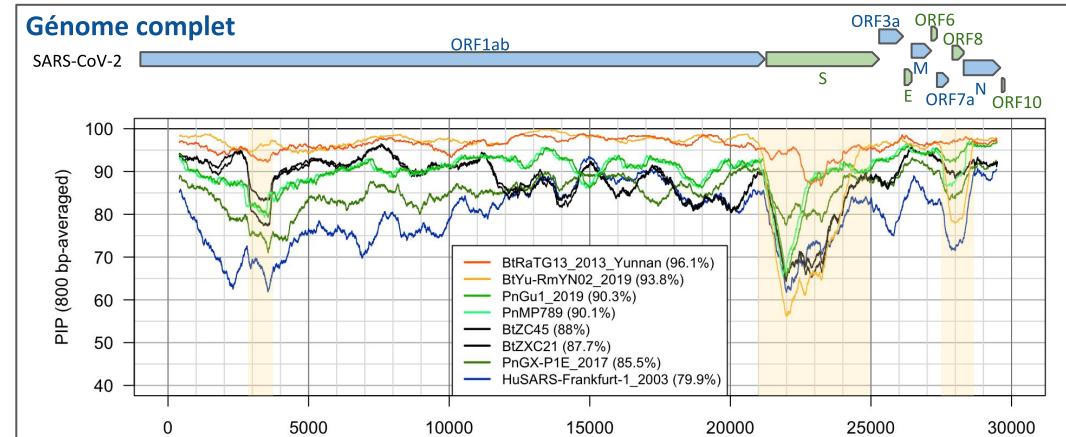
Recombinaisons génomiques dans les génomes de coronavirus

Le profil génomique

- régions ayant vraisemblablement fait l'objet de recombinaisons (fond jaune).

Profil PPI du gène spicule (S)

- **S:** spicule
- **RBD:** receptor binding domain
- Dans la région du RBD forte baisse des PPI
- Le RBD est en évolution rapide, pourquoi ?
 - Immunogène → forte pression sélective en faveur de variations qui permettent d'échapper à l'immunité
 - Spécificité d'espèce → modifications permettent de changer d'hôte



■ Sallard, E., Halloy, J., Casane, D., van Helden, J. & Decroly, É. 2020. Retrouver les origines du SARS-CoV-2 dans les phylogénies de coronavirus. *Med Sci (Paris)* 36: 783–796.

Un métagénom de virus de chauve-souris (RmYN02)

Report

Current Biology

A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein

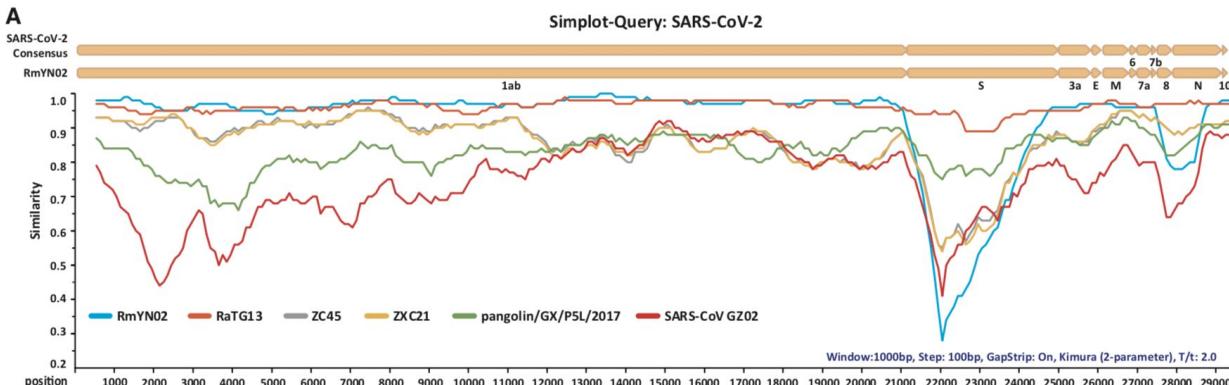
- Hong Zhou et coll. (2019)
- Séquence métagénomique d'un coronavirus de chauve-souris assemblée à partir de 11 échantillons de chauves-souris de Yunnan
- Génome très proche de SARS-CoV-2 **sauf** dans la région du gène spicule (S).
 - ce virus serait donc un recombinant de RaTG13
 - **oui mais** ce métagénom est reconstruit à partir de 11 échantillons différents. La recombinaison pourrait donc être un artefact
- Les auteurs soulignent aussi la présence d'une insertion similaire au site furine de SARS-CoV-2.
 - **Oui mais** ce scénario est peu vraisemblable, car il aurait nécessité quatre événements évolutifs (une insertion et trois délétions).

Highlights

- Metagenomic analysis identified a novel coronavirus, RmYN02, from *R. malayanus*

Authors

Hong Zhou, Xing Chen, Tao Hu, ..., Alice C. Hughes, Yuhai Bi, Weifeng Shi



H

		Polybasic cleavage site																										
		667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693																										
SARS-CoV-2 Numbering		G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	P	R	R	A	R	S	V	A	S	Q	S	I	I
Consensus SARS-CoV-2		G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	P	R	R	A	R	-	V	G	T	N	S	I	I
RmYN02		G	A	G	V	C	A	S	Y	-	-	-	-	N	S	P	-	A	A	R	-	V	G	T	N	S	I	I
RaTG13		G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	-	-	-	-	R	S	V	A	S	Q	S	I	I
ZC45		G	A	G	I	C	A	S	Y	H	T	A	S	I	L	-	-	-	-	R	S	T	S	Q	K	A	I	V
ZXC21		G	A	G	I	C	A	S	Y	H	T	A	S	I	L	-	-	-	-	R	S	T	G	Q	K	A	I	V
pangolin/MP789/2019		G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	-	-	-	-	R	S	V	S	S	X	A	I	I
pangolin/GX/P5L/2017		G	A	G	I	C	A	S	Y	H	S	M	S	S	F	-	-	-	-	R	S	V	N	Q	R	S	I	I
SARS-CoV GZ02		G	A	G	I	C	A	S	Y	H	T	V	S	L	-	-	-	-	R	S	T	S	Q	K	S	I	V	
RmYN01		G	A	G	I	C	A	S	Y	H	T	A	S	L	-	-	-	-	R	N	T	G	Q	K	S	I	V	

O-linked glycan residues

Bases de données de séquences biologiques

- KB: knowledge base
- Swiss-prot contient des séquences protéiques “annotées” par des biologistes.
L’annotation consiste à associer à une séquence les connaissances résultant d’expérimentation.
- Deux limitations
 - Le nombre de publications augmente tellement qu’il n’est pas possible à l’équipe de Swiss-prot de tout annoter
 - Le nombre de séquences augmente de façon tellement rapide qu’il est impossible de toutes les caractériser expérimentalement
- TREMBL: annotation automatique des séquences traduites de la base de données EMBL.
- Uniprot = Swiss-prot + TREMBL

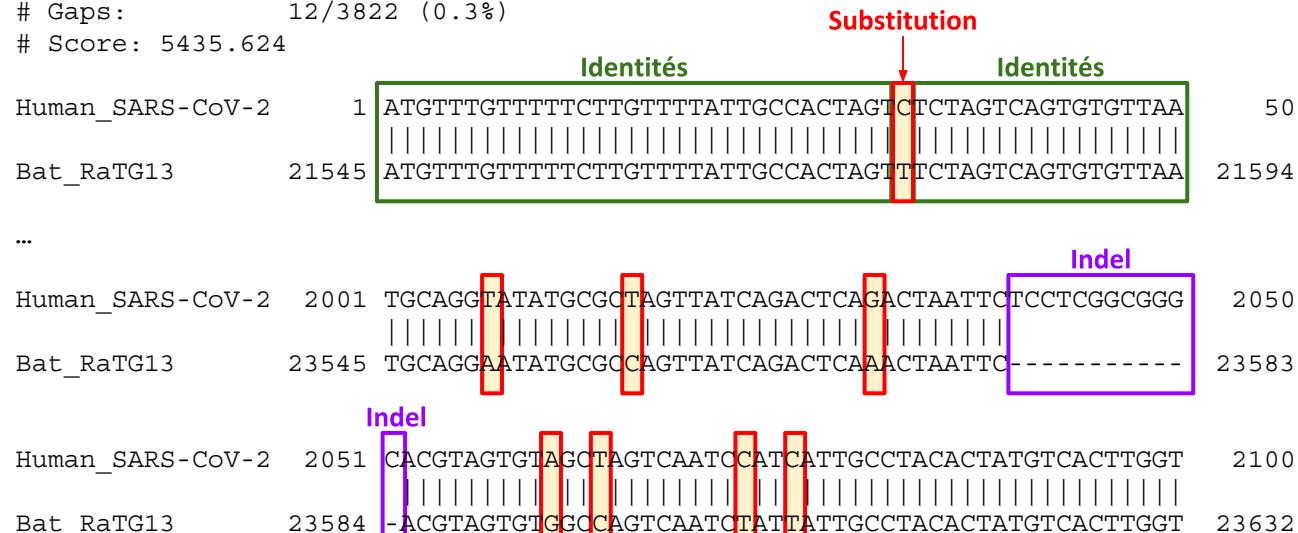
Le NCBI (USA) propose une série de bases de données couvrant les différents domaines de la biologie : taxonomie, séquences protéiques, séquences nucléiques, publications biologiques, ...

Nous utiliserons le site Web “Entrez” pour consulter les séquences génomiques de coronavirus et leurs annotations.

Alignement d'une paire de séquences

Alignement de séquences – Gènes S de SARS-CoV-2 et RaTG13

```
# Aligned_sequences: 2
# 1: Human_SARS-CoV-2_BetaCoV/Wuhan/IPBCAMS-WH-01/2019
# 2: Bat_RaTG13
#
# Length: 3822
# Identity: 3549/3822 (92.9%)
# Similarity: NA/3822 (NA%)
# Gaps: 12/3822 (0.3%)
# Score: 5435.624
```



Note

- “Indel” signifie “Insertion ou délétion”
- Sur base de ce résultat, la différence observée peut provenir soit d'une insertion chez un ancêtre de SARS-CoV-2, soit d'une délétion chez un ancêtre de RaTG13

Recherche de séquences par similarité

Alignement de séquences de SARS-CoV-2 sur le génome du HIV

Haut: fragment le plus significatif de l'alignement de la séquence du gène S sur le génome du VIH. Noter le score Expect = 7.5. Ce score n'est significatif que s'il est nettement inférieur à 1.

Bas: fragment le plus significatif de l'alignement d'une séquence aléatoire sur le génome du VIH. Noter le score Expect = 2.1, supérieur à 1 et donc non-significatif (comme on s'y attend, puisque la séquence est aléatoire).

Conclusion: l'alignement sur lequel s'appuient Perez et Luc Montagnier correspond à ce qu'on s'attend à trouver par hasard en alignant des séquences de cette taille.

HIV-1 isolate patient B clone 16.3 from Netherlands envelope glycoprotein (env) gene, complete cds				
Sequence ID: HQ386166.1		Length: 2580	Number of Matches: 1	Range 1: 2493 to 2523
Score	Expect	Identities	Gaps	Strand
39.2 bits(42)	2.1	27/31(87%)	0/31(0%)	Plus/Minus
Query 351	CCTAAAAGTTCTTGTAATAACTGTATTATT		381	
Sbjct 2523	CCTAAAAGTTCTTGTAATATTCTATAATT		2493	

Travaux pratiques

Travaux pratiques

Nous mobiliserons une série d'outils bioinformatiques accessibles en ligne pour analyser les séquences de coronavirus et pour tenter de trouver des éléments informatifs concernant l'origine de SARS-CoV-2.

Supports de ce cours	Diapos, tuto, données	https://ivanheld.github.io/shnc-origines-sars-cov-2/
Uniprot	Base de donnée de séquences protéiques	https://www.uniprot.org/
NCBI Entrez	Bases de données biologiques	https://www.ncbi.nlm.nih.gov/
EMBOSS needle	Alignement de paires de séquences	https://www.ebi.ac.uk/Tools/psa/emboss_needle/
NCBI BLAST	Recherche de séquences par similarité	https://blast.ncbi.nlm.nih.gov/Blast.cgi
PIPprofileR	Profils de pourcentages de positions identiques	https://pipprofiler.france-bioinformatique.fr/
Clustal	Alignement de séquences multiples	https://www.ebi.ac.uk/Tools/msa/clustalo/
phylogeny.fr	Phylogénie moléculaire	https://www.phylogeny.fr/
AMU	page AMETICE de N&C3	https://ametice.univ-amu.fr/course/view.php?id=62928