

ETIENNE DECROLY
JACQUES VAN HELDEN
MARTINE QUINIO

COVID-19 : APPROCHES SCIENTIFIQUES & ENJEUX SOCIÉTAUX



Covid-19 : approches scientifiques et enjeux sociétaux

par Etienne Decroly, Jacques van Helden & Martine Quinio

Campus Saint-Charles, Bâtiment 2, grand amphi

Inscription : contact@universite-populaire-de-marseille.net

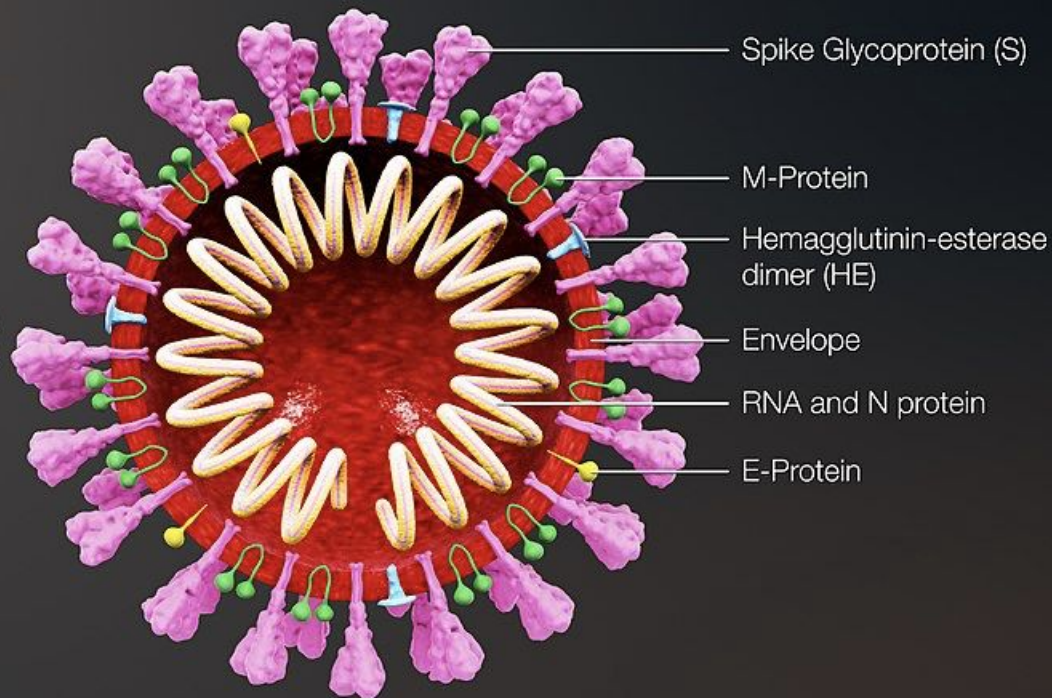
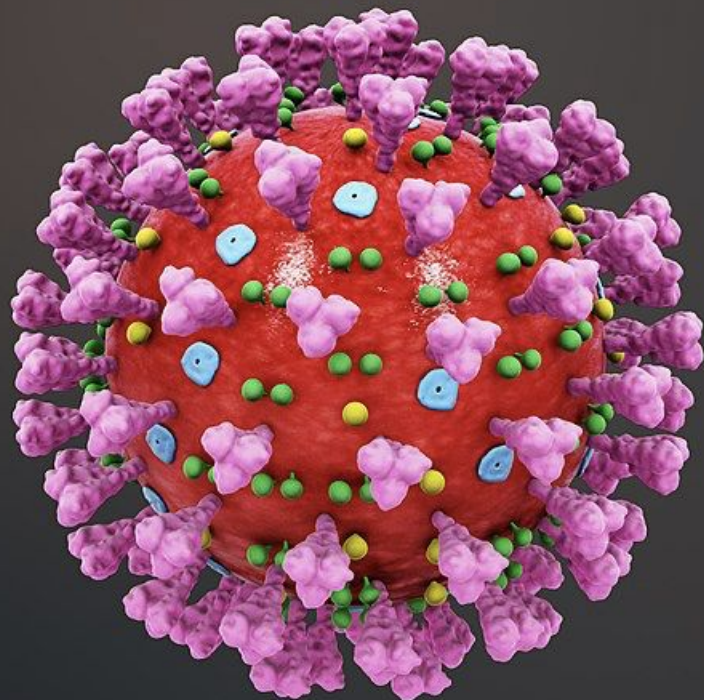
Info et documents : http://universite-populaire-de-marseille.net/covid-19_2020/

Vidéos : <https://www.youtube.com/c/UniversitePopulairedeMarseille/>

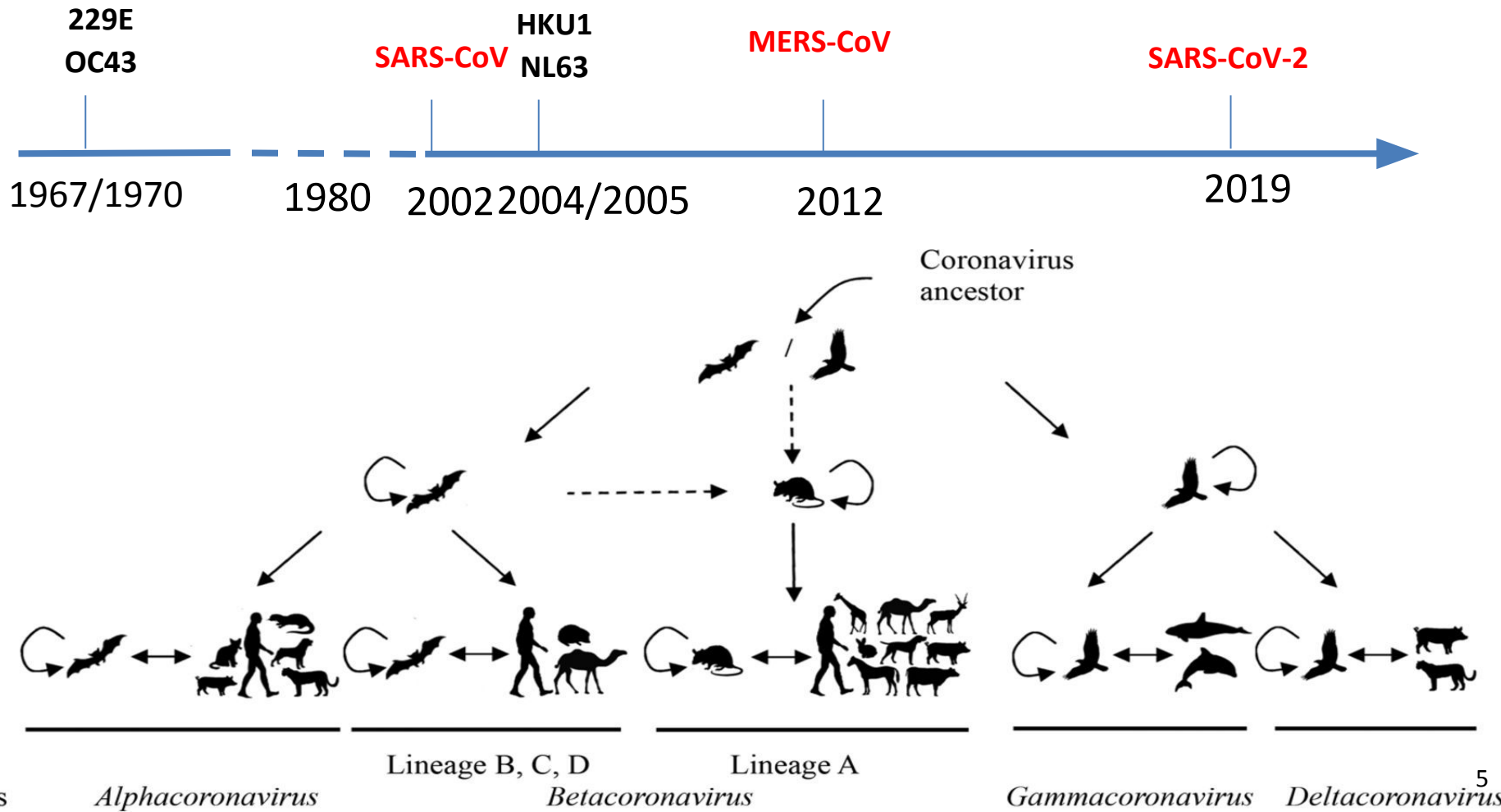
Lundi 28/09	Etienne Decroly Virologue, Directeur de Recherches au CNRS	Approches de virologie moléculaire
Jeudi 01/10	Jacques van Helden Professeur de bioinformatique à Aix-Marseille Université	Retracer dans les génomes l'origine et la propagation du virus
Jeudi 08/10	Martine Quinio Professeure de mathématiques à Aix-Marseille Université	Raisonnement statistique et modélisation
Mercredi 14/10	Annick Stevens, Jacques van Helden, Martine Quinio et Etienne Decroly	Covid-19: et maintenant ? a. Table-ronde-débat b. Compléments d'information

Rappel du premier épisode (Etienne Decroly)

Résumé de la Séance 1 : Approches de virologie moléculaire

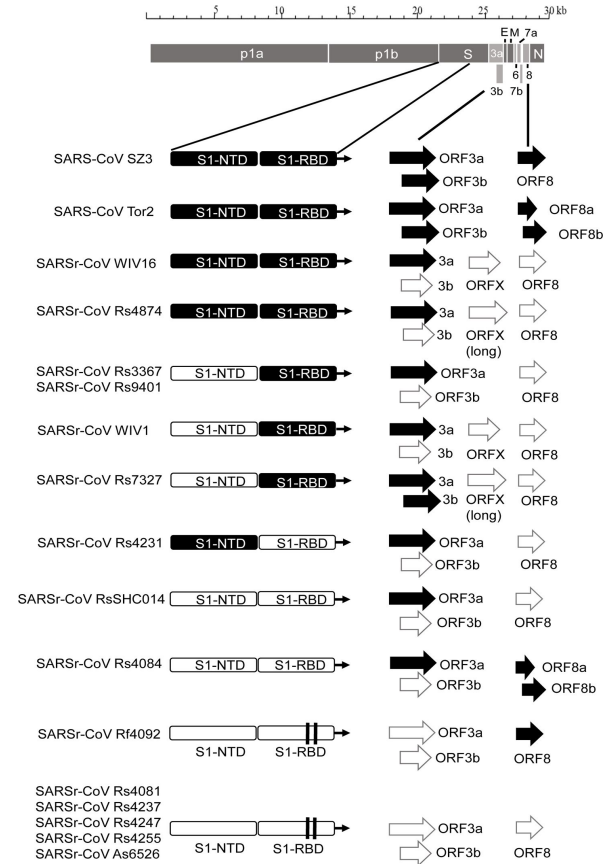
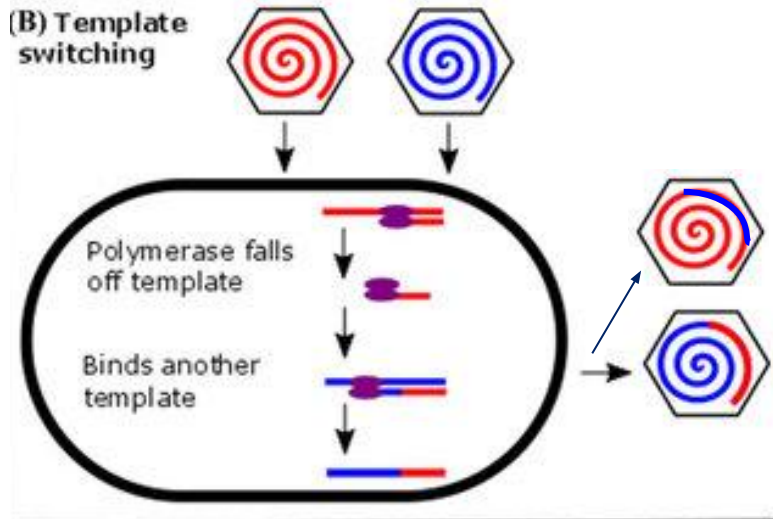


Cycle zoonotique et émergence des coronavirus



La recombinaison est un facteur d'évolution des virus chez les chauves-souris

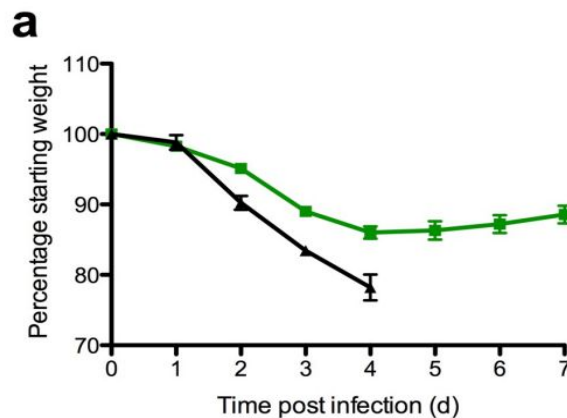
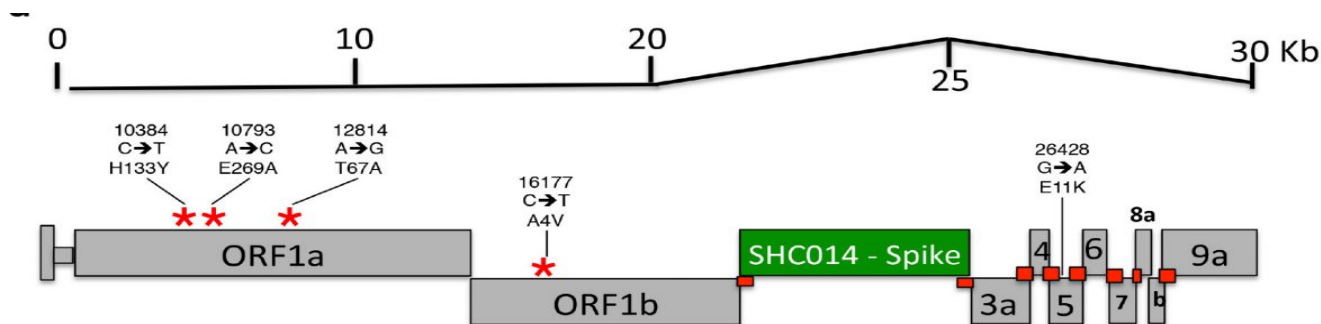
- Les chauve souris sont infectées par de nombreuses espèces des CoV.
- Des évènements de **recombinaisons** sont détectables dans les études de phylogénie (génomés chimériques).



1. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus
2. Ben Hu ,et al Plos path : November 30, 2017
3. J. [Dennehy](#), Evolutionary ecology of virus emergence: Virus emergence, 2016, Annals of the New York Academy of Science

Comment construire des virus chimériques qui franchissent la barrière d'espèce?

Expériences de Gain de fonction:



Expériences de gain de fonction

- 2011
 - résultats de Ron Fouchier
 - Gros débat au sein de la communauté
- 2021 publication
- 2014
 - Embargo sur les financements d'expériences de les gains de fonction
- 2017
 - Levée de l'embargo

Comment un virus passe il la barrière d'espèce?

La protéine spicule (Spike, S) joue un rôle crucial dans l'infection

Protéine initialement inactive

→ Clivage par une enzyme de l'hôte

→ Protéine active

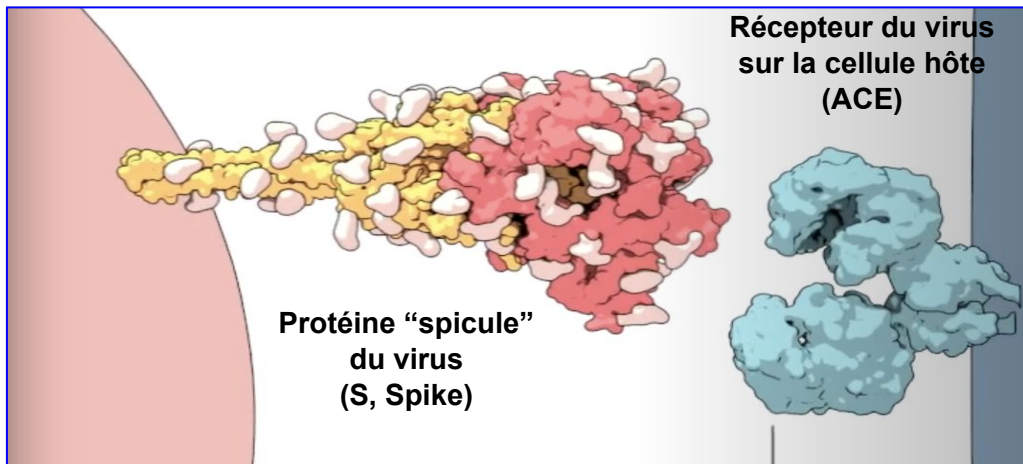
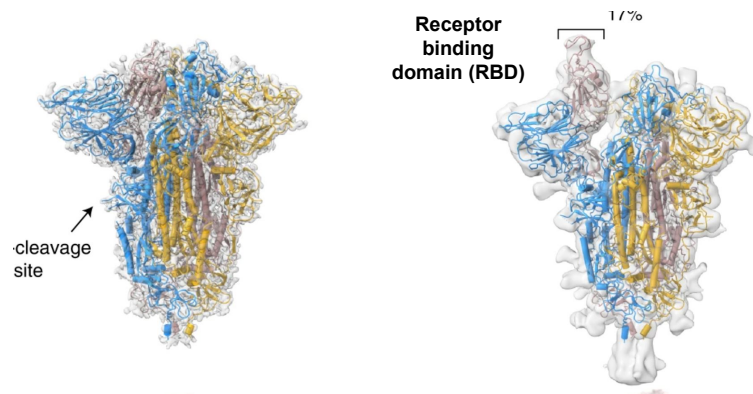
→ interaction avec protéine de membrane de l'hôte

(**ACE2** : angiotensin-converting enzyme)

→ entrée dans la cellule

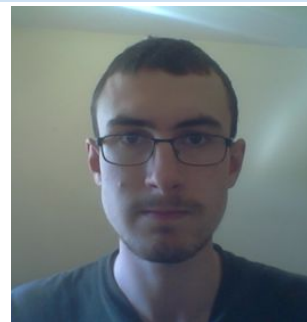
- ACE: angiotensin-converting enzyme
- RBD: receptor binding domain

Activation par clivage protéolytique



A l'origine de ce cours

- Ce cours repose en grande partie sur un travail d'équipe mené durant le confinement.
- 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](https://doi.org/10.1007/s12250-020-00145-5)

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
 - ❑ 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
- Comment distinguer le vrai du faux ?
 - ❑ Approche: ignorer délibérément les débats médiatiques et analyser les séquences des virus

Hypothèses sur l'origine du SARS-CoV-2

- Zoonose naturelle ... c'est le processus usuel des émergences
 - modèle spillover (débordement)
 - modèle des “quasi espèces”
 - Note: la collecte des échantillons sur les sites naturels génère des risques de zoonose

- Origine de laboratoire (accidentelle) ?
 - Il y a eu plusieurs précédents de virus échappés de laboratoires, par exemple grippe, ebola, SARS-CoV (de 2002), ...
 - Virus “naturel” cultivé en laboratoire
 - culture en cellules
 - expérimentation animale
 - Virus génétiquement modifié (et cultivé en laboratoire)
 - expérience de gain de fonction

- Caractéristiques des coronavirus
 - Virus à ARN
 - Rôle de la protéine S
 - RBD - Liaison au récepteur de l'hôte (ACE)
 - Recombinaison
- Zoonoses
 - Transferts entre espèces
 - Les chauve-souris constituent le principal réservoir animal des coronavirus
 - Hôte intermédiaire
- Expérimentation sur les génomes de virus
 - Cultures sur cellules de différents animaux
 - Passage d'espèce (sur culture, sur animaux de labo)
 - Expériences de gain de fonction
 -

Le génome de SARS-CoV-2

Le génome des coronavirus est constitué d'ARN

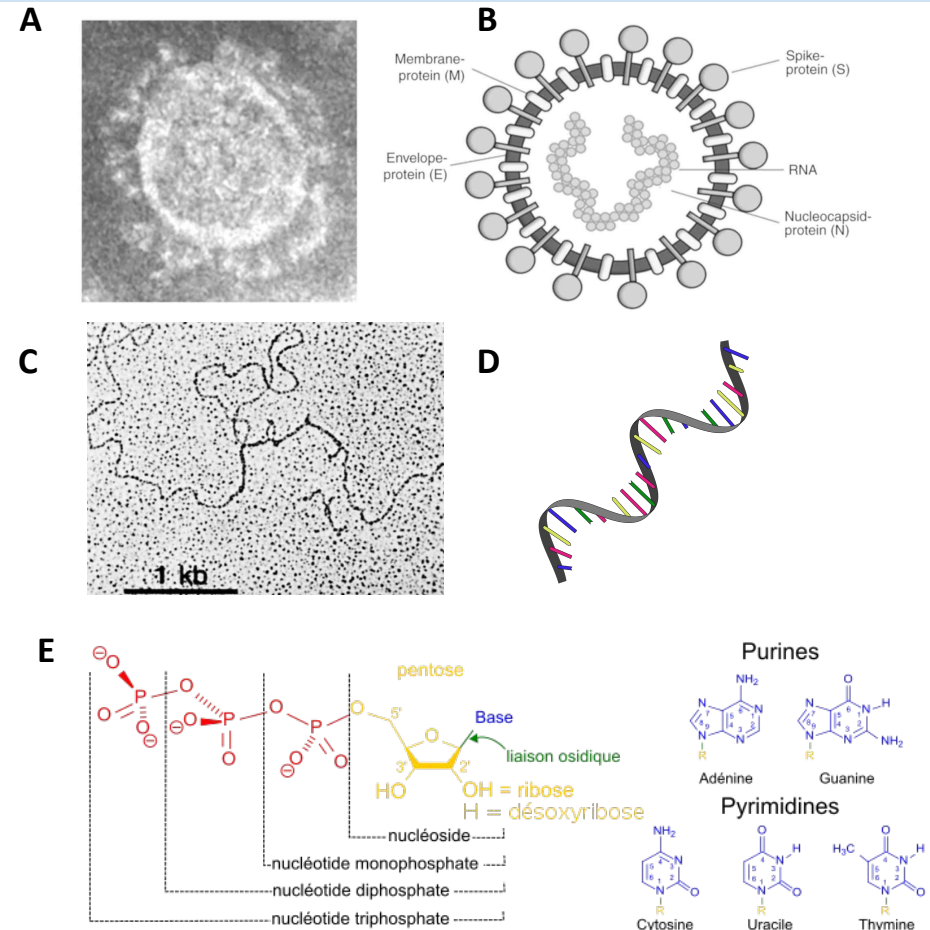
- Le génome est l'ensemble du matériel génétique d'un organisme
- Le génome des coronavirus est constitué d'acide ribonucléique (ARN).

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

```
ATTAAAGGTTTATACCTTCCAGGTAACAAACCAACCACTTTCGATCTCTGTAGATCTGTTCTCTAAA
CGAACTTTAAAATCTGTGTGGCTGTCACCTCGGCTGCATGCTTAGTGCACCTCACGCAGTATAATTAATAAC
TAATTAAGTGTGCTGTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCTGT
TTGCAGCCGATCATCAGCACATCTAGGTTTCGTCCGGGTGTGACCGAAAGGTAAGATGGAGAGCCTTGTCTC
CCTGGTTTCAACGAGAAAAACACAGTCCAACTCAGTTTGCCTGTTTTACAGGTTTCGCGACGTGCTCGTAC
GTGGCTTTGGAGACTCCGTGGAGGAGGTTTATCAGAGGCACGTCAACATCTTAAAGATGGCACTTGTGG
CTTAGTAGAAGTTGAAAAGGCGTTTTCCTCAACTTGAACAGCCCTATGTGTTTCATCAAACGTTCCGGAT
GCTCGAAGTGCACCTCATGGTTCATGTTATGGTTGAGCTGGTAGCAGAACTCGAAGGCATTTCAGTACGGTC
GTAGTGGTGGAGACTTGGTGTCTTGTCCCTCATGTGGGCGAAATACAGTGGCTTACCGCAAGGTTCT
TCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGATCTAAAGTCAATTTGACTTA
GGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAAGTGGAACTAAACATAGCAGTGGTG
TTACCCGTGAAGTTCATGCGTGAGCTTAAACGGAGGGGCATACACTCGCTATGTCGATAACAACCTTCTGTGG
CCCTGATGGCTACCCTCTTGAGTGCATTAAGACCTTCTAGCACGTGCTGGTAAAGCTTTCATGCACTTTG
TCCGAACAAGTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTGAACATGAGCATGAAATTTG
CTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGAATTAATTTGGCAAAGAA
ATTTGACACCTTCAATGGGGAATGTCCAAATTTTGTATTTCCCTTAAATTCATAATCAAGACTATTCAA
CCAAGGTTGAAAAGAAAAGCTTGTATGGCTTTTATGGGTAGAATTCGATCTGTCTATCCAGTTGCGTCAC
CAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTGTGGTGAAGCTTTCATGGCA
GACGGGCGATTTTGTAAAGCCACTTGCGAATTTTGTGGCACTGAGAATTTGACTAAAGAAGGTGCCACT
ACTGTGGTTACTTACCCCAAATGCTGTGTTAAAATTTATTGTCAGCATGTCACAATTCAGAAGTAG
GACCTGAGCATAGTCTTGCCGAATACATAATGAATCTGGCTTGAACCACTTCTTCGTAAGGGTGGTCCG
CACTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAGTGTGCCATTATGGGTTCCA
CGTGCTAGCGCTAACATAGGTTGTAACCATAACAGGTGTTGTTGGAGAAGGTTCCGAAGGCTTAAATGACA
ACCTTCTTGAAATACTCCAAAAGAGAAAGTCAACATCAATATGTTGGTGACTTTAACTTAAATGAAGA
GATCGCCATTATTTTGGCATCTTTTCTGCTTCCACAAGTGTCTTTGTGGAAACTGTGAAAGGTTTGGAT
TATAAAGCATTCAAAACAAATGTTGAATCCTGTGGTAATTTTAAAGTTACAAAAGGAAAAGCTAAAAAG
GTGCTTGAATATTTGGTGAACAGAAATCAATACTGAGTCTCTTATGCAATTTGCATCAGAGGCTGCTCG
```


- 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

Localisation des gènes

- Le génome de SARS-CoV-2 comporte 10 gènes
- Le schéma ci-dessous indique la position de ces gènes (flèches épaisses) le long du génome (ligne supérieure)
- Ces gènes présentent des tailles très variables.
- Le premier (ORF1ab) couvre $\frac{2}{3}$ du génome !
 - Cas tout-à-fait particulier: ce gène code pour plusieurs protéines



Localisation des gènes

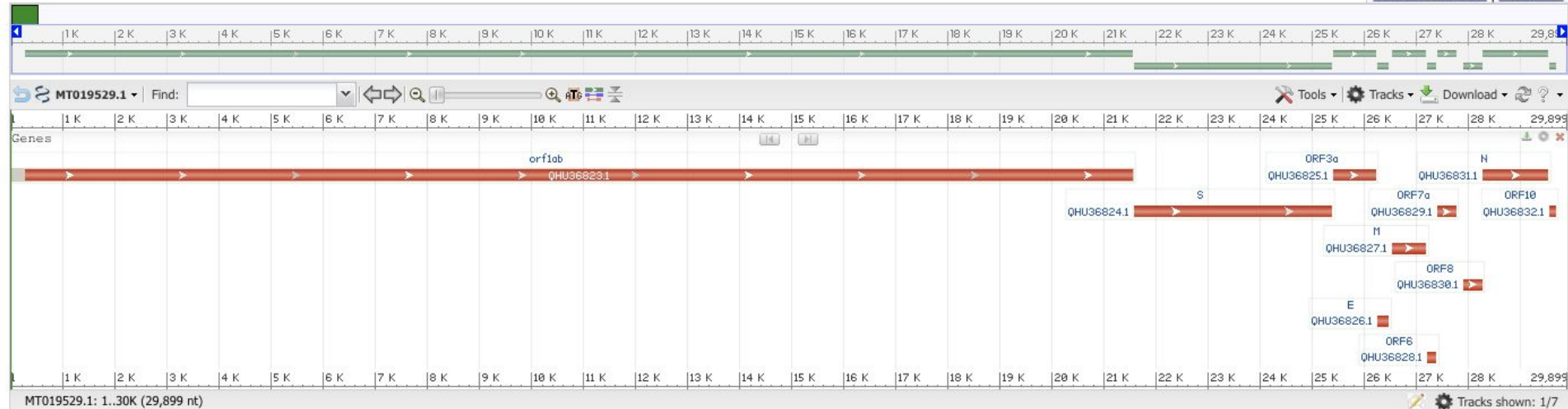
- Le génome de SARS-CoV-2 comporte 10 gènes
- Le schéma ci-dessous indique la position de ces gènes (flèches épaisses) le long du génome (ligne supérieure)
- Ces gènes présentent des tailles très variables.
- Le premier (ORF1ab) couvre $\frac{2}{3}$ du génome !
 - Cas tout-à-fait particulier: ce gène code pour plusieurs protéines

Severe acute respiratory syndrome coronavirus 2 isolate BetaCoV/Wuhan/IPBCAMS-WH-01/2019, complete genome

GenBank: MT019529.1

[GenBank](#) [FASTA](#)

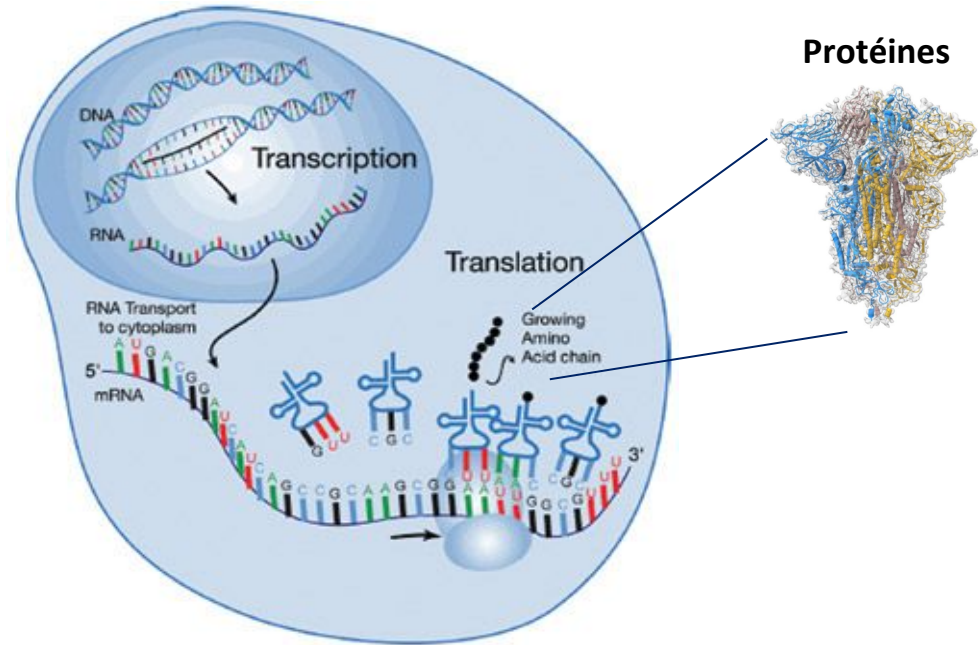
[Link To This View](#) | [Feedback](#)



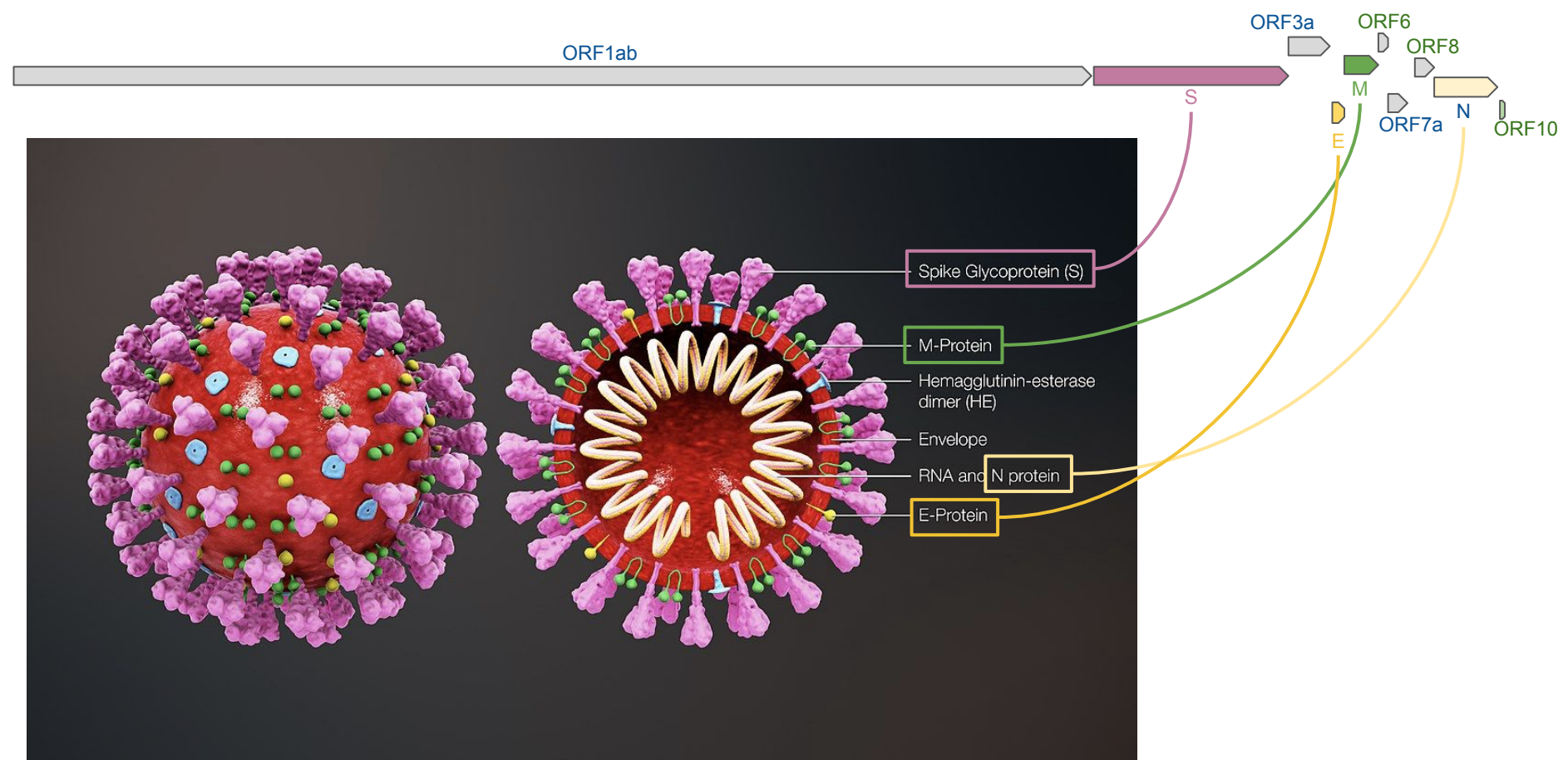
RNA makes protein makes a lot of things *

* Paraphrase de Francis Crick : "DNA makes RNA makes protein"

- L'ARN sert de modèle à la synthèse des protéines
- Les protéines sont les principaux acteurs moléculaires des organismes vivants
 - Enzymes
 - Transporteurs
 - Régulateurs
 - Cycle cellulaire
 - Différenciation cellulaire
 - ... un tas d'autres fonctions



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

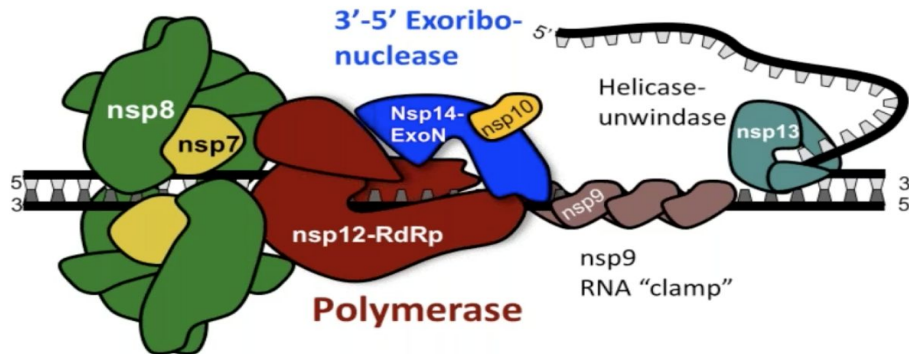
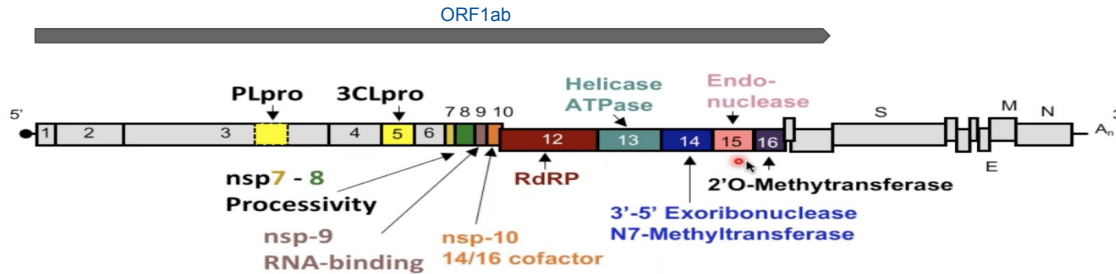


Le complexe répllication/transcription

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.

Des chauves-souris et des hommes

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>

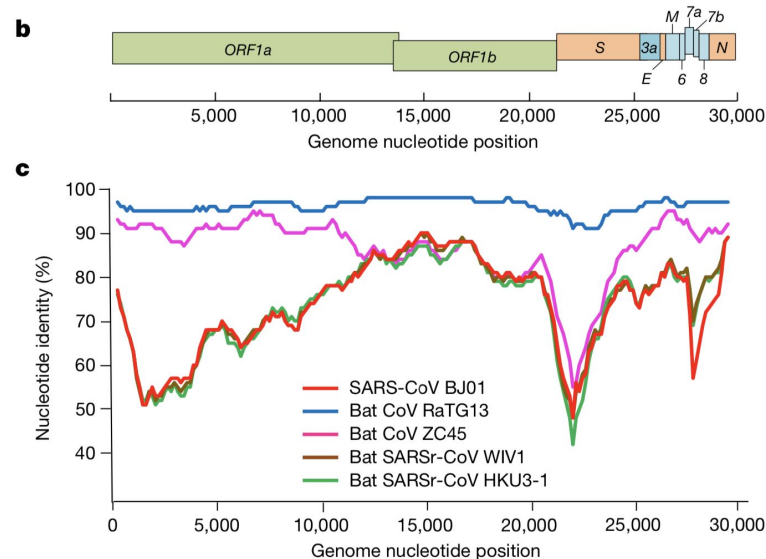
Received: 20 January 2020

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^{1,5}, 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^{1,5}, Ying Chen^{1,5}, Xu-Rui Shen^{1,3}, Xi Wang^{1,3}, Xiao-Shuang Zheng^{1,3}, Kai Zhao^{1,3}, Qian-Jiao Chen¹, Fei Deng¹, Lin-Lin Liu⁴, Bing Yan¹, Fa-Xian Zhan⁴, Yan-Yi Wang¹, Geng-Fu Xiao¹ & Zheng-Li Shi^{1,5*}



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



Alignement d'une paire de séquences

Typologie des mutations

- Substitution

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

Avant répllication

ATGACCATGA



Après répllication

ATGACCA**G**GA

- Délétion

- Perte d'un fragment de la molécule

ATG**ACCAT**GA



↓
ATGGA

- Insertion

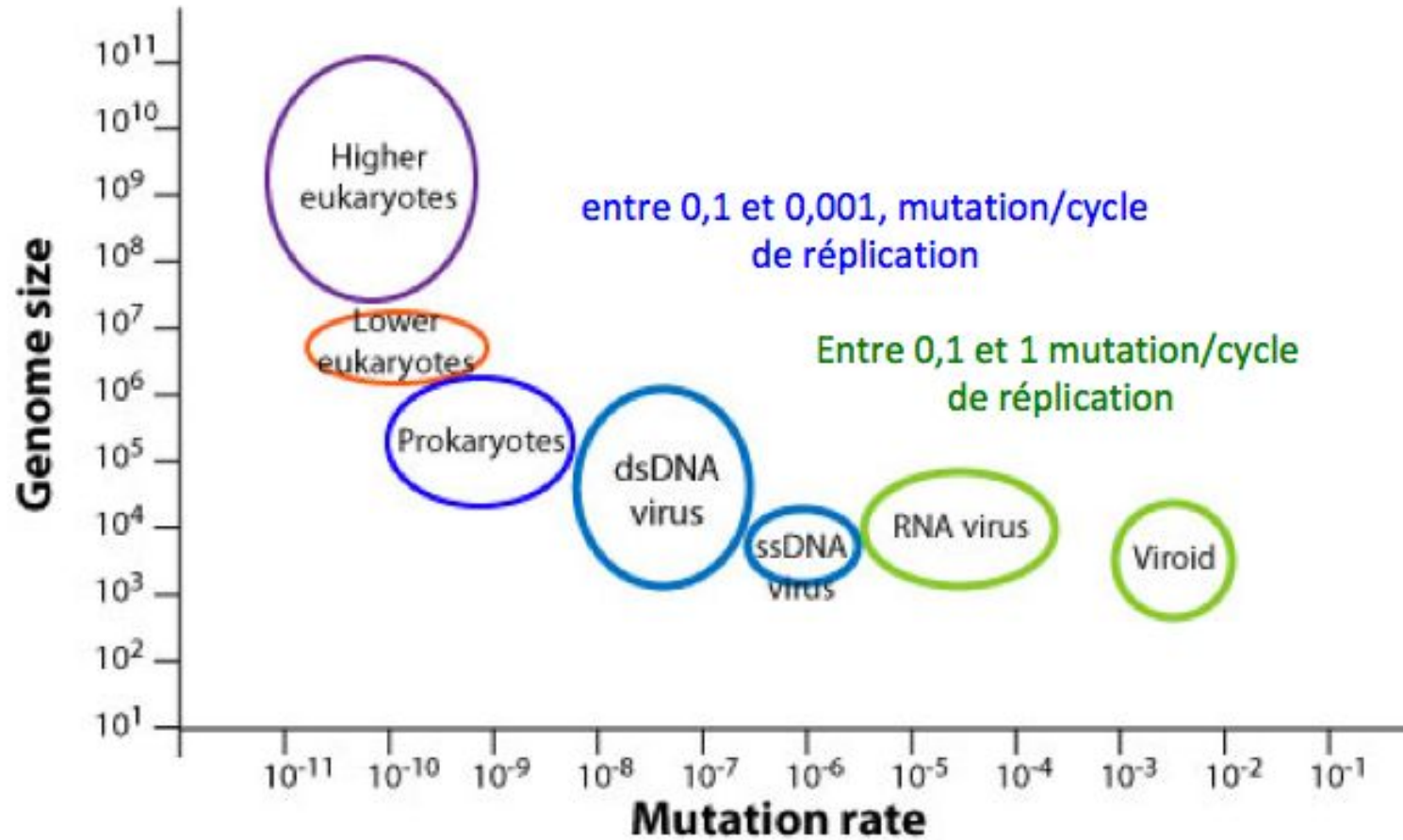
- Ajout d'un fragment de molécule

ATGACCATGA



ATGA**AAA**CATGA

Taux de mutation 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
```

Substitution

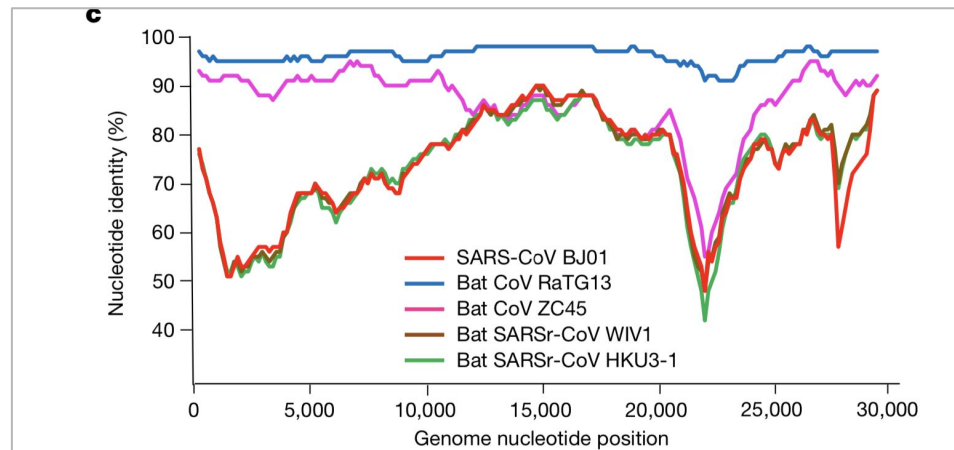
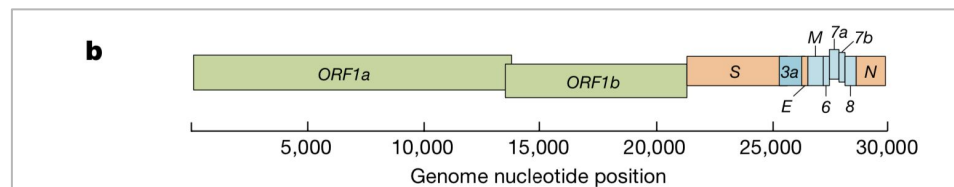
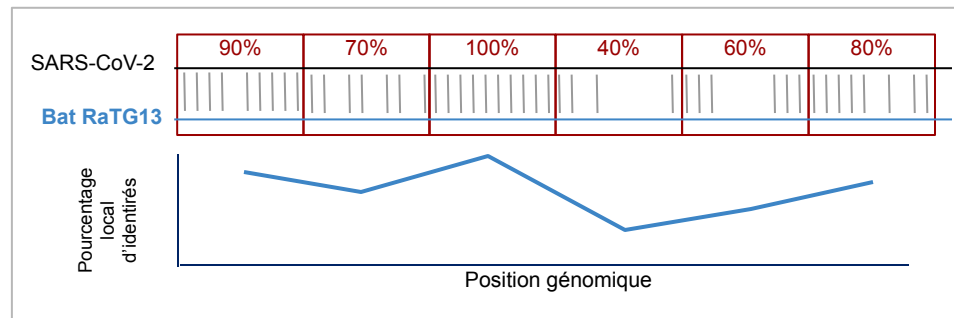
		Identités			Identités	
Human_SARS-CoV-2	1	ATGTTTGTTCCTTCTGTTTTATGCCACTAGTCTCTAGTCAGTGTGTAA				50
Bat_RaTG13	21545	ATGTTTGTTCCTTCTGTTTTATGCCACTAGTTTCTAGTCAGTGTGTAA				21594
...						
Human_SARS-CoV-2	2001	TGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCCTCCTCGGCGGG	2050			
Bat_RaTG13	23545	TGCAGGAATATGCGCCAGTTATCAGACTCAACTAATTC-----	23583		Indel	
...						
Human_SARS-CoV-2	2051	CACGTAGTGTAGCTAGTCAATCCATCATTGCCTACACTATGTCACCTGGT	2100			
Bat_RaTG13	23584	-ACGTAGTGTGGCCAGTCAATCTATTATTGCCTACACTATGTCACCTGGT	23632		Indel	

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

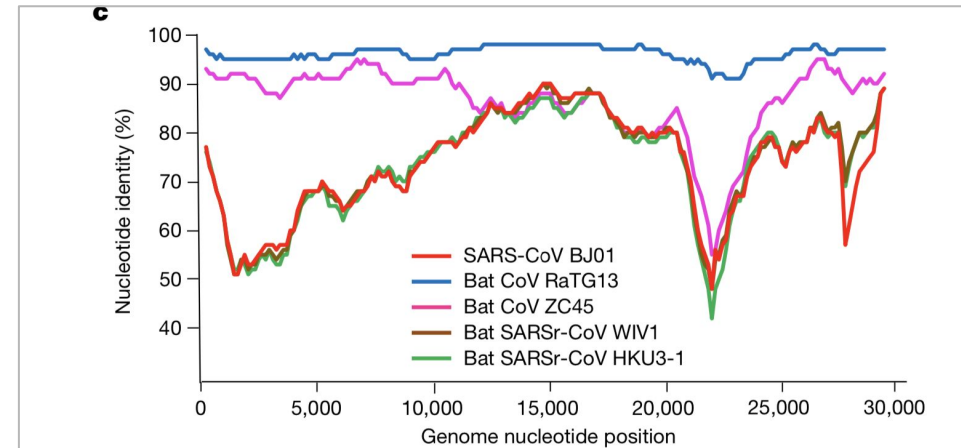
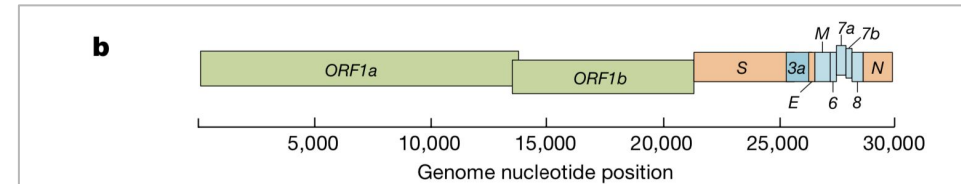
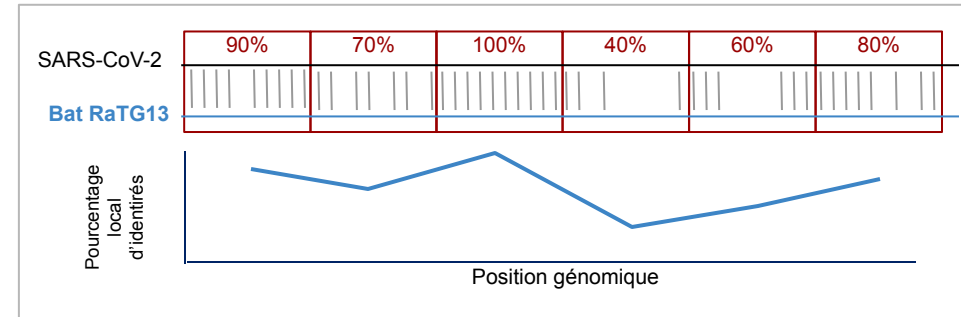
Profils de positions identiques (PPI)

- 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



Profils de positions identiques (PPI) de génomes de coronavirus

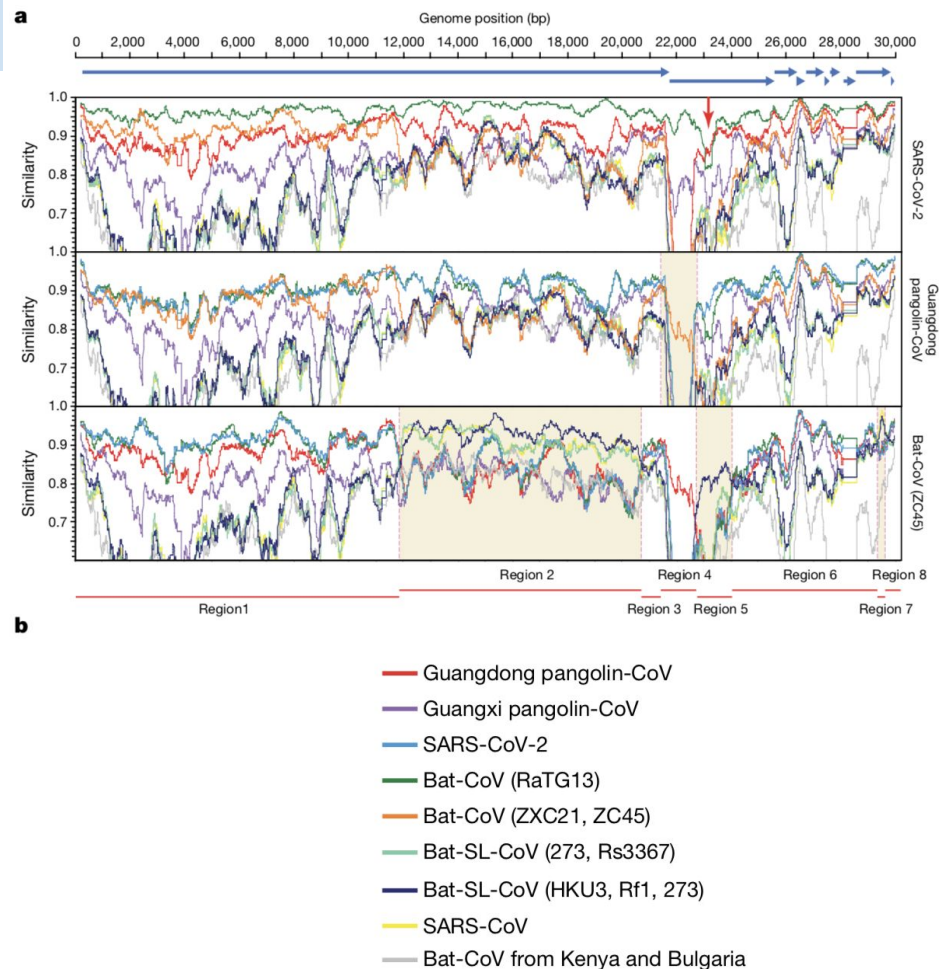
- RaTG13 (virus de chauve-souris) est le génome le plus proche de SARS-CoV-2
- On a identifié d'autres virus de chauve-souris relativement proches de SARS-CoV-2 (Cov ZC45)
- Les virus SARS-CoV humains (pandémie 2002-2003) sont moins proches
- Pour chaque espèce, on observe des fluctuations le long du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



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

... 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.



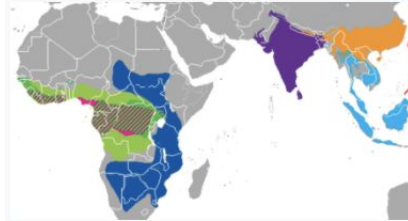
Pangolins

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

Répartition géographique de différentes espèces de pangolin

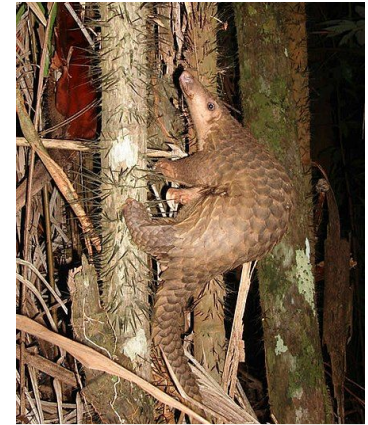
Genera

- *Manis*
- *Phataginus*
- *Smutsia*



Species ranges

-  *Manis crassicaudata*
-  *Manis pentadactyla*
-  *Manis javanica*
-  *Manis culionensis*
-  *Phataginus tricuspis*
-  *Phataginus tetradactyla*
-  *Smutsia gigantea*
-  *Smutsia temminckii*



Manis javanica
(Pangolin malais)

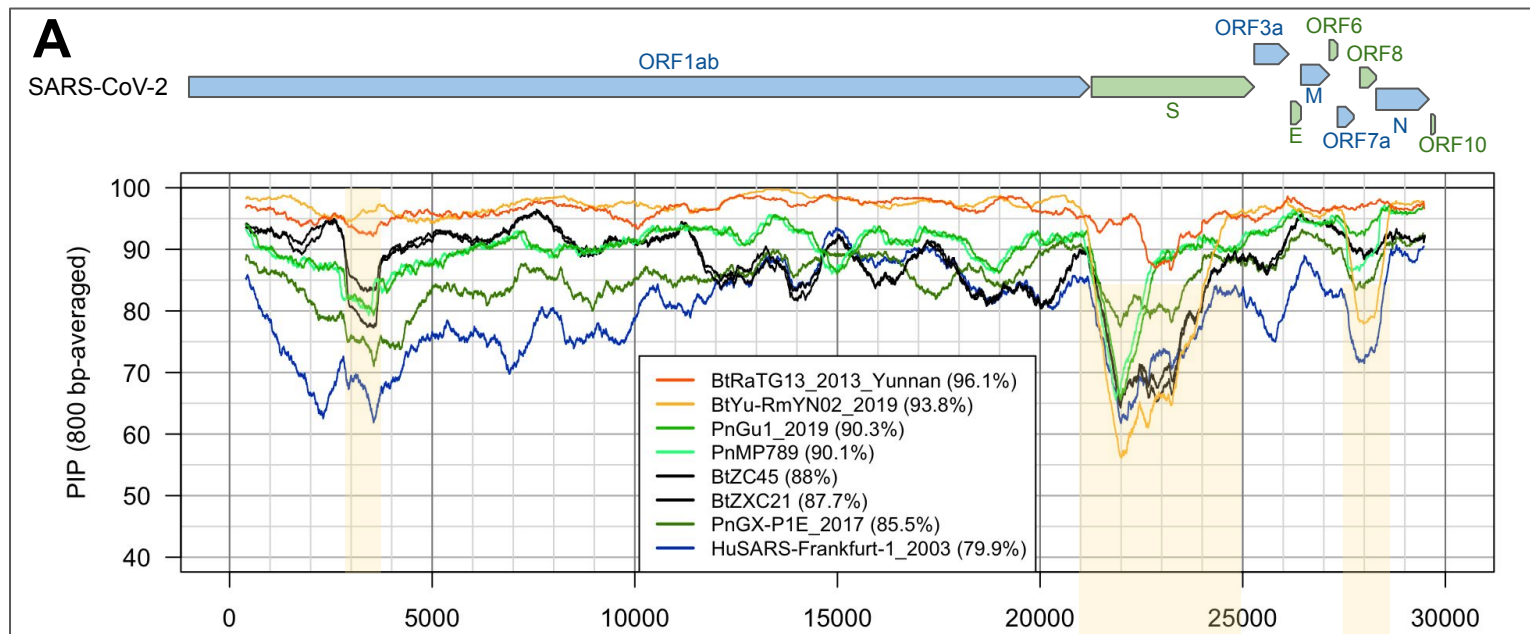
Aire de répartition de *Manis javanica*



Détecter les recombinaisons par comparaison de génomes

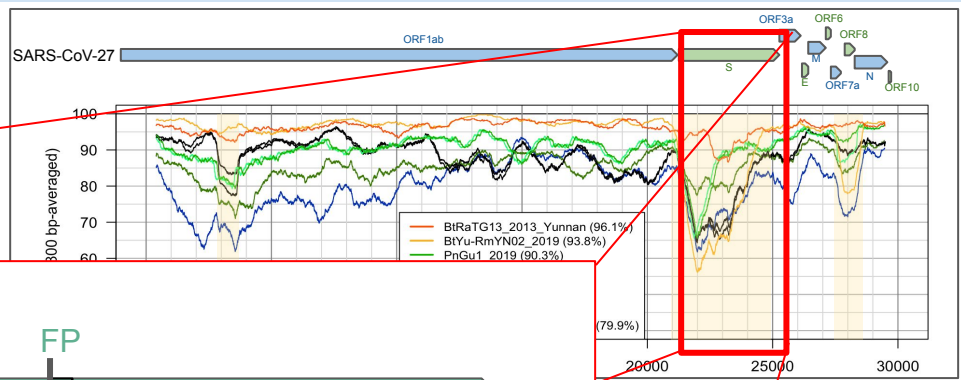
Détection des recombinaisons génomiques

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

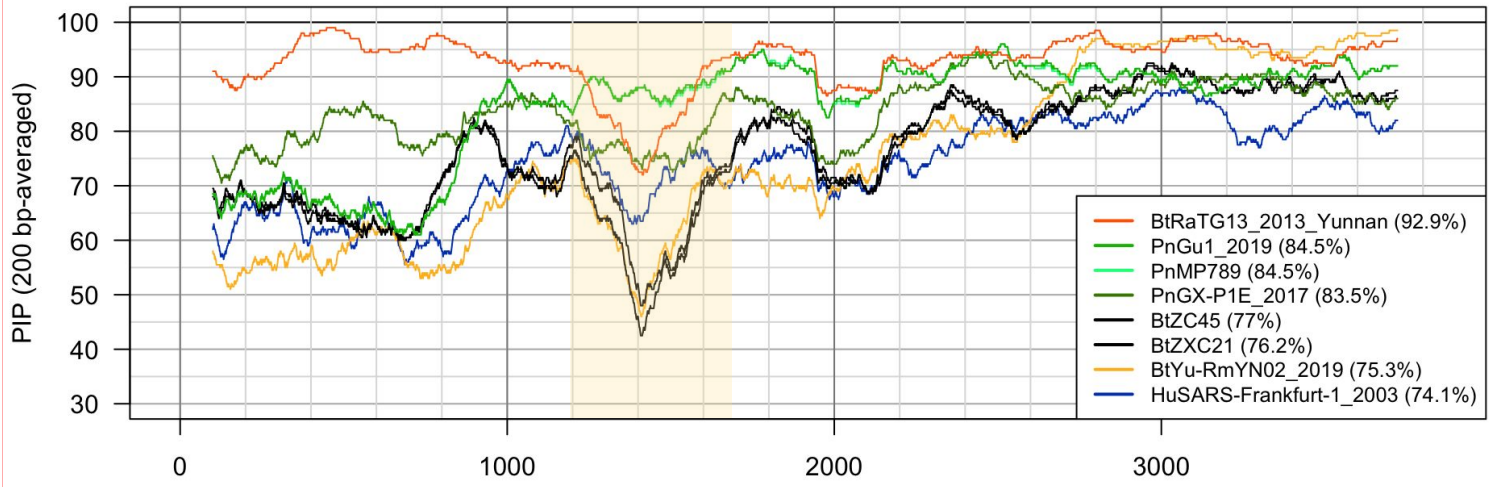
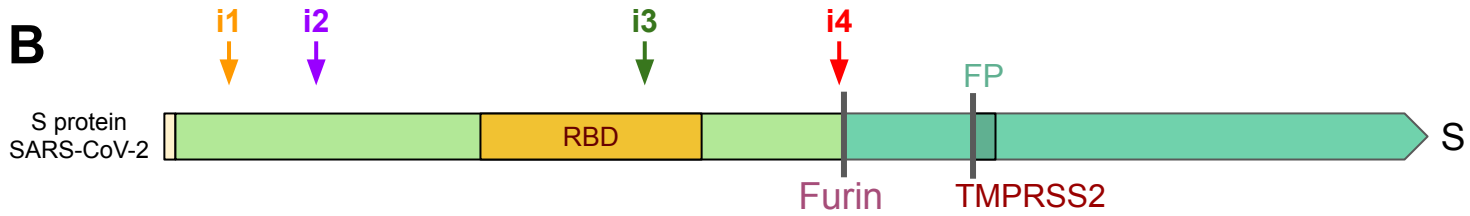


Comparaison entre coronavirus - gène S

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



B



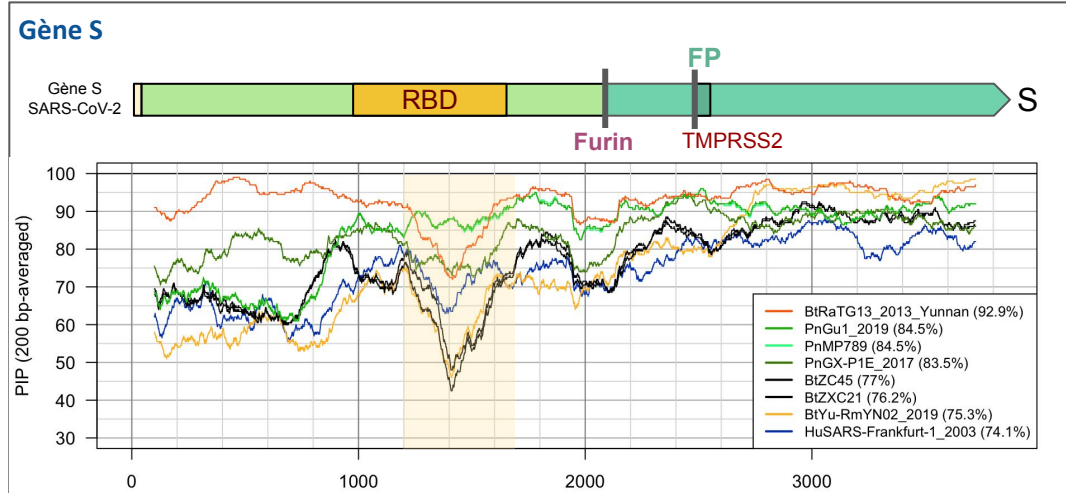
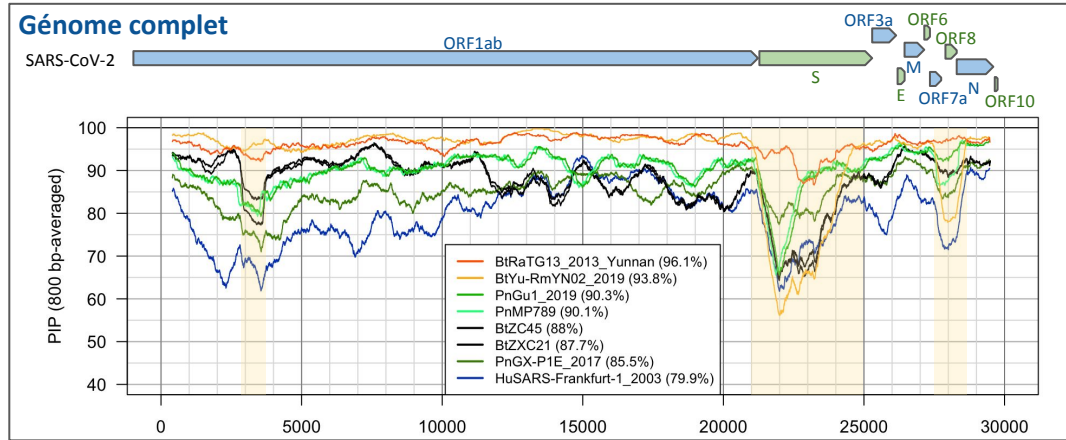
Recombinaisons génomiques dans les génomes de coronavirus

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



Un métagénome de virus de chauve-souris (*RmYN02*)

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

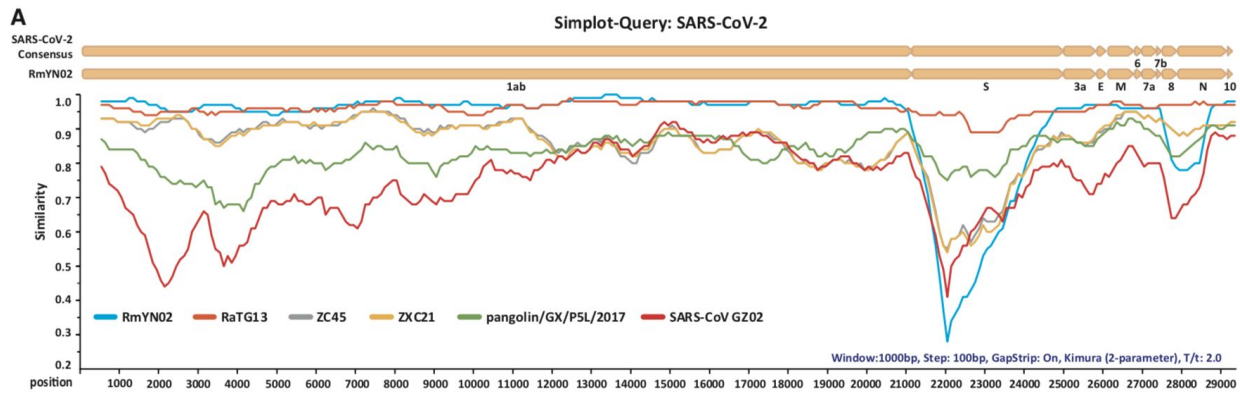
- Hong Zhou et coll. (2019)
- Séquence métagénomique d'un coronavirus de chauve-souris assemblée à partir de 11 échantillons de chauves-souris de Yunnan
- Génome très proche de SARS-CoV-2 **sauf** dans la région du gène spicule (S).
 - ce virus serait donc un recombinant de RaTG13
 - **oui mais** ce métagénome est reconstruit à partir de 11 échantillons différents. La recombinaison pourrait donc être un artéfact
- 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

	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	S	V	A	S	Q	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	L	-	-	-	-	R	S	T	S	Q	K	S	I	V
RmYN01	G	A	G	I	C	A	S	Y	H	T	A	S	L	L	-	-	-	-	R	N	T	G	Q	K	S	I	V

O-linked glycan residues

Origine 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)
- Leurs conclusions sont très généralement adoptées par les scientifiques et un bon nombre de publications ultérieures s'y réfèrent.

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.

- 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** la sélection n'est pas pour autant naturelle : le fait de maintenir des virus sur des cultures cellulaires (en labo) revient à réaliser une sélection artificielle, et peut créer des possibilités inconnues.
- 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.

- Sirotkin & Sirotkin 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

BioEssays

www.bioessays-journal.com

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 Sirotkin* and Dan Sirotkin

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,

Des fragments de HIV dans le génome de SARS-CoV-2 ?

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 :

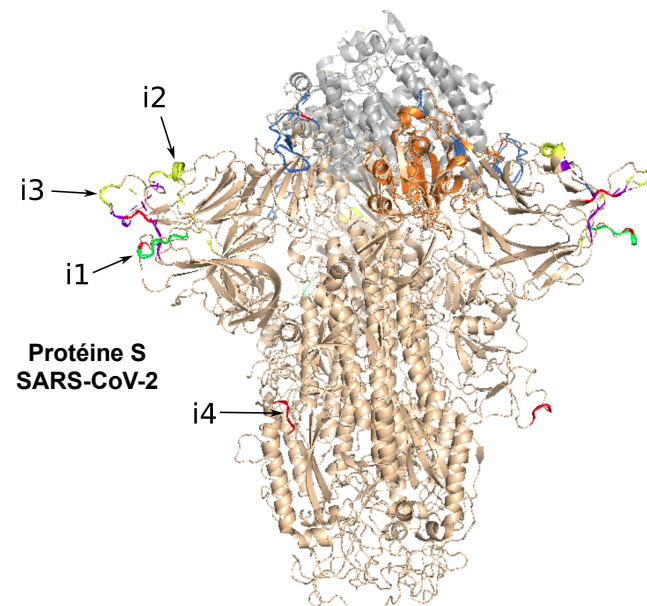
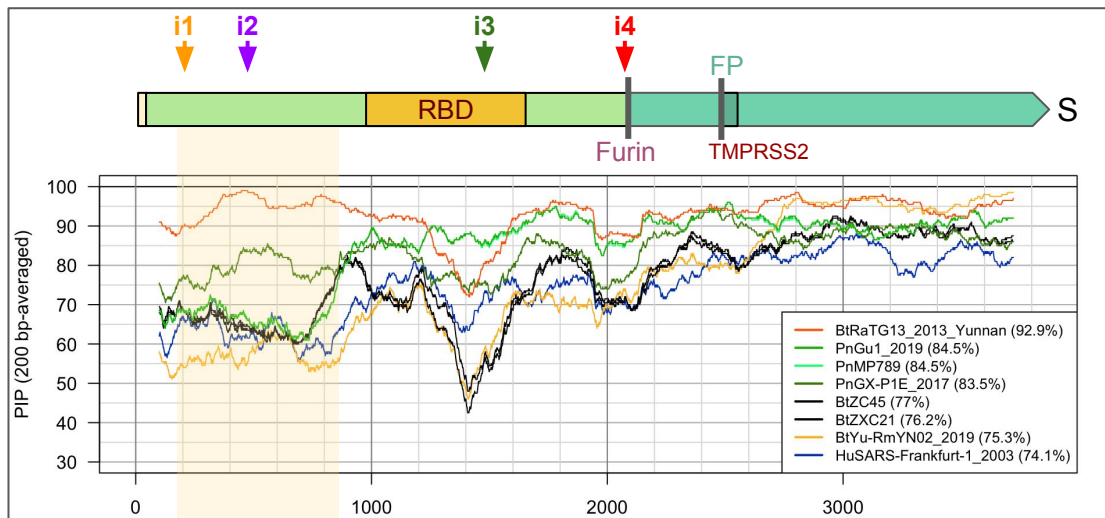
- le travail d'un collègue mathématicien, Jean-Claude Perez, qui "a fouillé les moindres détails de la séquence",
- une analyse des séquences génomiques et protéiques des coronavirus préalablement publiée par une équipe indienne, qui a, selon lui, "été forcée de rétracter" sa publication.

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'étouffement, 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". "

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".



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 19828.PPH11 from Netherlands envelope glycoprotein (env) gene, partial cds				
Sequence ID: HQ644953.1		Length: 1143	Number of Matches: 1	Range 1: 967 to 994
Score	Expect	Identities	Gaps	Strand
38.3 bits(41)	7.5	25/28(89%)	0/28(0%)	Plus/Plus
Query	86	AATGGTACTAAGAGGTTTGATAACCCTG	113	
Sbjct	967	AATGGTACTAAAAGGTTAGATAACACTG	994	

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	CCTAAAAGTTCTTTGTAATAACTGTATTATT	381	
Sbjct	2523	CCTAAAAGTTCTTTGTAATATTTCTATAATT	2493	

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.

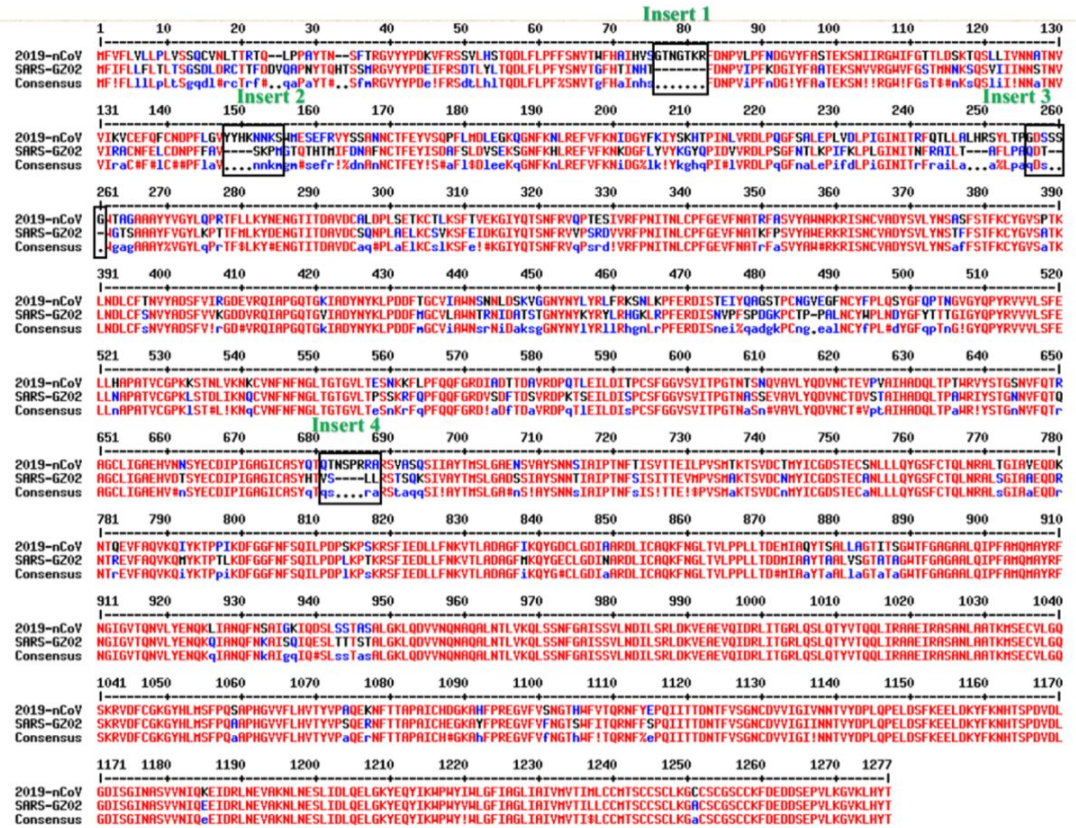


Figure 2: Multiple sequence alignment between spike proteins of 2019-nCoV and SARS. The sequences of spike proteins of 2019-nCoV (Wuhan-HU-1, Accession NC_045512) and of SARS CoV (GZ02, Accession AY390556) were aligned using MultiAlin software. The sites of difference are highlighted in boxes.

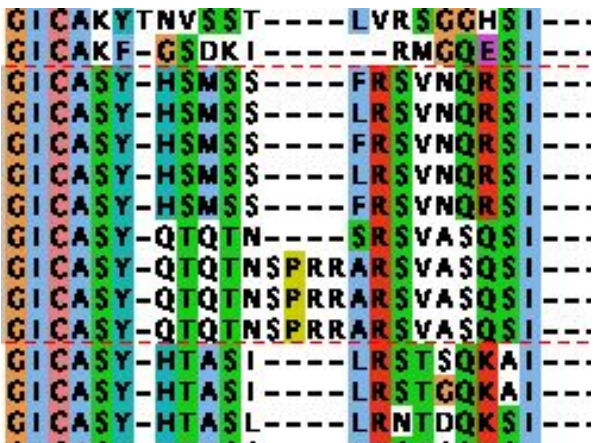
Insertion partagée entre tous les virus du groupe CoV-2

- 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.

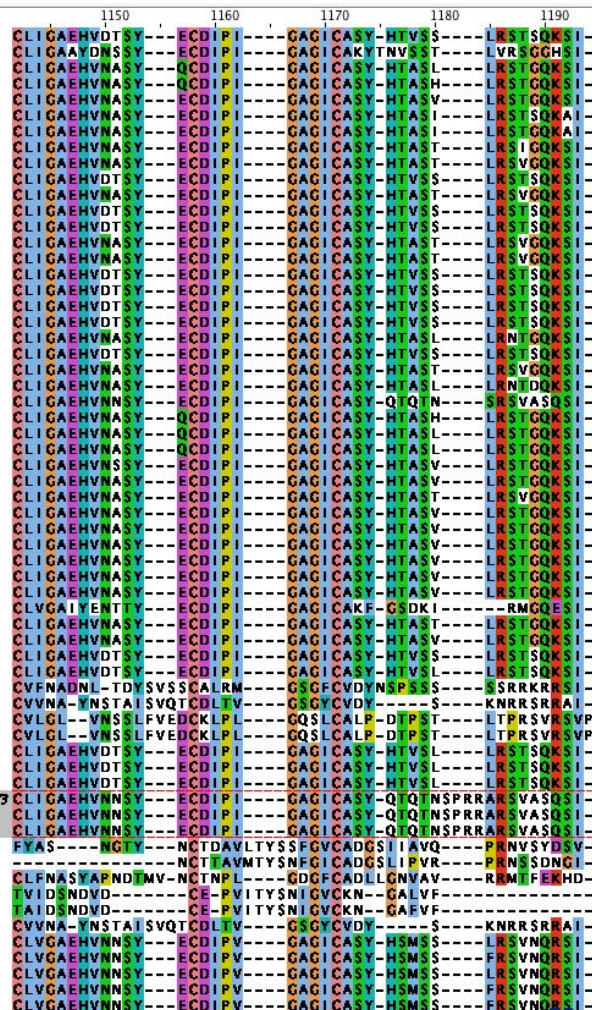
	420	430	440	450	460	470	480
Pig_SADS_AY441569.1_ref/1-1130	N S V V T V R L C R	--- W W ---	- Q F M S F N S T S H A A D A	--- G P T N A	--- F	--- E	--- C L T
Pig_PRCV_AKV62755.1_ref/1-1232	S G K L V T K Q F L	--- L V N C L W P V P F S E E A A S	--- T F	--- C F E G A D F	--- H C N C A V L N N T V D V I R F I		
Human_TGVB_CA891145.1_ref/1-1447	S G K L V T K Q F L	--- L V N C L W P V P F S E E A A S	--- T F	--- C F E G A D F	--- Q D N C A V L N N T V D V I R F I		
Human_229E_AAG48592.1_ref/1-1173	S F Q P L L L N C L	--- W ---	- S V S G L R F T T G F	--- Y F Y N G	--- T G R G	--- D	--- C K C F S S D V L S D V I R F I
Human_ML63_AA558177.1_ref/1-1356	L Y Q P L R L T C L	--- W ---	- P V P G L K S S T G F	--- Y F Y N A	--- T G S D V Y	--- N	--- C N G Y Q H N S V D V I R F I
Human_h-HKU1_N23_ABD96197.1_ref/1-1356	N G V L E I T A C Q	--- Y T M C E Y P H T I C K S K G S S R N E	--- S W	--- H F D K S	--- E	--- P L C L	--- F K --- K I
Human_HK04-02_ABN19366.1_ref/1-1361	Q G L L E M S V C Q	--- N M M C E Y P H T I C H P K L G N H F K E L W	--- H L D T G V	--- V	--- S	--- C L	--- Y K --- R I
Human_OC43_ADX10763.1_ref/1-1358	Q G L L E M S V C Q	--- N M M C E Y P H T I C H P N L G N H F K E L W	--- H L D T G V	--- V	--- S	--- C L	--- Y K --- R I
Human_h-MERS_KOR/CNMH_SNU/172_06_2015_ALK80311.1_ref/1-1353	T L V L L P D G C C T L L R A	--- F Y C I L E P R S G N H C P A G N S Y T	--- S F	--- A	--- Y H T P	--- A	--- T C S D G I
Camel_MERS_AHE78097.1_ref/1-1353	T L V L L P D G C C T L L R A	--- F Y C I L E P R S G N H C P A G N S Y T	--- S F	--- A	--- Y H T P	--- A	--- T C S D G I
Rat_RM48-31_ADK66841.1/1-1259	G T H I V L D Y C N	--- F N F C A D P M F A W N S ---	- G Q Y H K	--- T W	--- I	--- T S A	--- A --- H C
Rat_RoKY72_APO40579/1-1257	G T H I V L D Y C N	--- F Y F C Q D P M L A A N ---	- G S H F K	--- S W	--- Y	--- F L N A	--- I --- N R Y H
Pangolin_PCoV_GX-PSE_QIA48641_ref/1-1267	A N N V V I K V C E	--- F Q F C T D P F L G Y Y H N N N K	--- T W N E F E R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Pangolin_PCoV_GX-P4L_QIA48614/1-1267	A N N V V I K V C E	--- F Q F C T D P F L G Y Y H N N N K	--- T W N E F E R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Pangolin_PCoV_GX-P2V_QHQ54048/1-1269	A N N V V I K V C E	--- F Q F C T D P F L G Y Y H N N N K	--- T W N E F E R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Pangolin_PCoV_GX-P1E_QIA48623_ref/1-1265	A N N V V I K V C E	--- F Q F C T D P F L G Y Y H N N N K	--- T W N E F E R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Pangolin_PCoV_GX-P5L_QIA48632/1-1267	A N N V V I K V C E	--- F Q F C T D P F L G Y Y H N N N K	--- T W N E F E R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Rat_RaTG13_QHR63300_ref/1-1269	A N N V V I K V C E	--- F Q F C D P F L G Y Y H K N N K	--- S W M E S E F R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Human_SARS-CoV-2_BetaCoV/Wuhan/PPRCAMS-WH-01/2019_QHU36824/1-1273	A N N V V I K V C E	--- F Q F C D P F L G Y Y H K N N K	--- S W M E S E F R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Human_SARS-CoV-2_WIV04_QHR63260/1-1273	A N N V V I K V C E	--- F Q F C D P F L G Y Y H K N N K	--- S W M E S E F R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Human_SARS-CoV-2_Wuhan-Hu-1_YP_009724390_ref/1-1273	A N N V V I K V C E	--- F Q F C D P F L G Y Y H K N N K	--- S W M E S E F R Y Y S S A	--- N	--- N	--- C	--- F E Y I S Q I
Rat_Bat-SI-CoVZC45_AVP7803.1_ref/1-1246	A N N V I I K V C N	--- F D F Y D P L F G Y Y H N N K	--- H K N N K	--- T W S I R E F A N Y S S F	--- A	--- N	--- C
Rat_Bat-SI-CoVZC12_AVP78042_ref/1-1245	A N N V I I K V C N	--- F D F Y D P L F G Y Y H N N K	--- H K N N K	--- T W S I R E F A N Y S S F	--- A	--- N	--- C
Rat_Yiba11_AHK37558/1-1259	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_Ro4084_ATO98132.1/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_RoSHC014_AG248806.1/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_WIV1_AG248828.1/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_Ro3367_AG248818/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_Ro7327_ATO98218/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L K S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_Ro8r-BetaCoV/YN20108_QDF43825/1-1256	S T N V L V R V C N	--- F E L C D N P F F V W L R S N N T I P	--- S Y	--- F	--- F N N A	--- F	--- N
Rat_Ro4231_ATO98157/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P T G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Rat_S23_ATO98205.1/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P T G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Rat_WIV16_AIK02457.1/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P T G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Civet_SARS-CoV_007/2004_AAU04646/1-1255	S T N V V I R V C N	--- F E L C D N P F F V W S K P M G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Civet_S23_P59594.1/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P M G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Human_SARS-CoV_Tor2/FPI-10895_APR8742/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P M G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Human_SARS-CoV_TW11_AA87512/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P M G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Human_SARS-CoV_Frankfurt_LAAP33697.1_ref/1-1255	S T N V V I R V C N	--- F E L C D N P F F A M S K P M G T Q T H	--- T M	--- I	--- F D N A	--- F	--- N
Ra1.2012_AIA62277.1_ref/1-1236	G S A I T I E V C Y	--- F Q F C D N P F A I I R D ---	- G A Q I N	--- T A	--- I	--- Y I N I	--- R --- V Y D I L R I
Rat_TMC15_ANA96027/1-1236	G S A I T I E V C Y	--- F Q F C D N P F A I I S G ---	- G A Q T N	--- T A	--- I	--- Y I N I	--- R --- V Y D I L R I
Rat_Ro8r-BetaCoV/YN2013_AIA62330/1-1233	S T N L I V R V C N	--- F E L C K V P L F V W F K S N N S Q L	--- S H	--- L	--- F S D S	--- F	--- N
Rat_Ro8r-BetaCoV/SC2018_QDF43815/1-1256	S T H I L V K V C N	--- F I I C K E P M F T V S Q	--- N R H F K	--- S W	--- V Y Q D A	--- R	--- N
Rat_Ro/Shaanx2011_AGC74163/1-1240	S T H I L V K V C N	--- F V L C T E P M F T V S R	--- N Q H Y K	--- S W	--- V Y Q H A	--- R	--- N
Rat_Ro8r-BetaCoV/Hu2013_AIA62310/1-1241	S T H I L V K V C N	--- F V L C T E P M F T V S R	--- N Q H Y K	--- S W	--- V Y Q H A	--- R	--- N
Rat_YNF_31C_AK219076/1-1241	S T H I I I R V C Y	--- F N L C K D P M Y T V S A	--- G T Q I S	--- S W	--- V Y Q N A	--- F	--- N
Rat_Ro8r-BetaCoV/He2013_AIA62290/1-1241	S T H I I I R V C Y	--- F N L C K D P M Y T V S A	--- G T Q V S	--- S W	--- V Y Q S A	--- F	--- N
Rat_RY1/2004_ABD75323/1-1241	S T H I I I R V C Y	--- F N L C K D P M Y T V S A	--- G T Q K S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Bat-CoV-273/2005_ABG47060/1-1241	S T H I I I R V C Y	--- F N L C K D P M Y T V S A	--- G T Q K S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Co/Yunnan2011_AGC74176/1-1241	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G V H F S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro_672/2006_ACU31032/1-1241	S T H I I I R V C Y	--- F N L C K E P M Y A I S N	--- E Q H Y K	--- S W	--- V Y Q N A	--- V	--- N
Rat_Ro4081_ATO98120/1-1241	S T H I I I R V C Y	--- F N L C K E P M Y A I S N	--- E Q H Y K	--- S W	--- V Y Q N A	--- V	--- N
Rat-recombinant_Bat-SRoBQ_AJ60703.1_ref/1-1259	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q N	--- A W	--- V Y Q S A	--- F	--- N
Rat_Bat-SARS-CoV-Rm1/2004_ABD75332/1-1241	S T H I I I R V C N	--- F N L C K E P M Y T V S K	--- G T Q Q S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro3/2004_AA267052/1-1241	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G A Q Q S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro8r-BetaCoV/YN2010C_QDF43830/1-1241	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro4247_ATO98181.1/1-1242	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro4237_ATO98169/1-1241	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q S	--- S W	--- V Y Q S A	--- F	--- N
Rat_Ro8r-BetaCoV/GZ0213_AIA62320/1-1242	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q N	--- S W	--- V Y Q S A	--- F	--- N
Rat_HKU3-8_ADE34766/1-1242	S T H I I I R V C N	--- F N L C K E P M Y T V S M	--- G T Q Q N	--- S W	--- V Y Q S A	--- F	--- N
Rat_Longquan-14Q_AID16716/1-1242	S T H I I I R V C N	--- F N L C R E P M Y T V S R	--- G T Q Q N	--- S W	--- V Y Q S A	--- F	--- N
Rat_HKU3-12_ADE34812/1-1242	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q N	--- A W	--- V Y Q S A	--- F	--- N
Rat_HKU3-6_ADE34744/1-1242	S T H I I I R V C N	--- F N L C K E P M Y T V S R	--- G T Q Q N	--- A W	--- V Y Q S A	--- F	--- N

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).



RaC_WV1_AG248828.1/1-1256
 RaC_BM48-31_ADK66841.1/1-1259
 RaC_H.2012_AIA62277.1_ref/1-1236
 RaC_RF1/2004_ABD75323/1-1241
 RaC_Longquan-140_AID16716/1-1242
 RaC_Bat-SL-CoVZC45_AVP78031_ref/1-1246
 RaC_Bat-SL-CoVZC21_AVP78042_ref/1-1245
 RaC_BtRs-BetaCoV/YN2013_AIA62330/1-1233
 RaC_Rs_672/2006_ACU31032/1-1241
 RaC_Rs3967_AG248818/1-1256
 RaC_Rs4081_ATO98120/1-1241
 RaC_Rs4084_ATO98132.1/1-1256
 RaC_Rs4231_ATO98157/1-1255
 RaC_Rs4237_ATO98169/1-1241
 RaC_Rs4247_ATO98181.1/1-1242
 RaC_Rs7327_ATO98218/1-1256
 RaC_RsSHC014_AG248806.1/1-1256
 RaC_S23_ATO98205.1/1-1255
 RaC_WV16_ALK02457.1/1-1255
 RaC_Cp/Yunnan2011_AGC74176/1-1241
 RaC_BtRs-BetaCoV/YN2018_QDF43825/1-1256
 RaC_BtRs-BetaCoV/YN2018C_QDF43830/1-1241
 RaC_LYRa11_AHR37558/1-1259
 RaC_TG13_QHR63300_ref/1-1269
 RaC_Bat-CoV-273/2005_ABG47060/1-1241
 RaC_BtRf-BetaCoV/HeB2013_AIA62290/1-1241
 RaC_TMCI5_ANA96027/1-1236
 RaC_YNF_31C_AK219076/1-1241
 RaC_Bat-SARS-CoV-Rm1/2004_ABD75332/1-1241
 RaC_Rp3/2004_AA267052/1-1241
 RaC_Rp/Shaanx2011_AGC74165/1-1240
 RaC_BtRs-BetaCoV/GX2013_AIA62320/1-1242
 RaC_BtRs-BetaCoV/Hud2013_AIA62310/1-1241
 RaC_HKU3-12_ADE34812/1-1242
 RaC_HKU3-6_ADE34744/1-1242
 RaC_HKU3-8_ADE34766/1-1242
 RaC_BtKY72_APO40579/1-1257
 RaC_BtRf-BetaCoV/SC2018_QDF43815/1-1256
 Bat-recombinant_Bat-SRR0_AQ60703.1_ref/1-1259
 ClveC_SARS-CoV_007/2004_AAU04646/1-1255
 ClveC_S23_P59594.1/1-1255
 Human_h-HKU1_N23_ABD96197.1_ref/1-1356
 Human_HK04-02_ABN19366.1_ref/1-1361
 Human_h-MERS_KOR/CNMU_SNU/172_06_2015_ALK80311.1_ref/1-1353
 Camel_LMERS_AHE78097.1_ref/1-1353
 Human_SARS-CoV_Frankfurt-1_AAP33697.1_ref/1-1255
 Human_SARS-CoV_Tor2/FP1-10895_AFR58742/1-1255
 Human_SARS-CoV_TW11_AAR87512/1-1255
 Human_SARS-CoV-2_BetaCoV/Wuhan/IPRCAMS-WH-01/2019_QHU36824/1-1273
 Human_SARS-CoV-2_WV04_QHR63260/1-1273
 Human_SARS-CoV-2_Wuhan-Hu-1_YP_009724390_ref/1-1273
 Human_229E_AAG48592.1_ref/1-1173
 Human_NL63_AASS8177.1_ref/1-1356
 Pig_SADS_AVN41569.1_ref/1-1130
 Pig_PRCV_AKV62755.1_ref/1-1232
 Human_TGCV_CAB91145.1_ref/1-1447
 Human_OC43_ADX10763.1_ref/1-1358
 Pangolin_PCoV_GX-P1E_QIA48623_ref/1-1265
 Pangolin_PCoV_GX-P2V_QIQ54048/1-1269
 Pangolin_PCoV_GX-P4L_QIA48614/1-1267
 Pangolin_PCoV_GX-P5E_QIA48641_ref/1-1267
 Pangolin_PCoV_GX-PS1_QIA48632/1-1267



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.

```

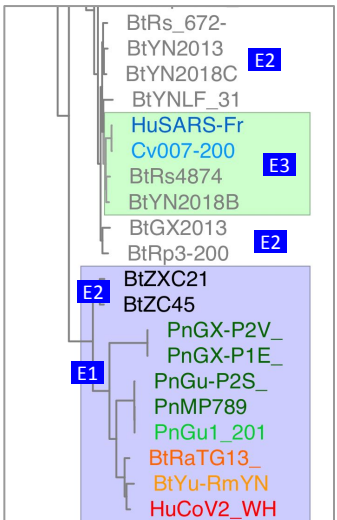
HuCoV2_WH01_2019  KSNLKPFERDISTEITYOAGSTPCNGVEGFNCYFPLOS YGFOPTNGVGYOPYRVVVL
BtRaTG13_2013_Yunnan  KANLKPFERDISTEITYOAGSKPCNGQTGLNCYYPLRYRGFYPTDGVGHOPYRVVVL
PnGX-P1E_2017        KSKLKPFERDISTEITYOAGSTPCNGQVGLNCYYPLERYGFHPTTG VNYOPFRVVVL
PnGX-P2V_2018        KSKLKPFERDISTEITYOAGSTPCNGQVGLNCYYPLERYGFHPTTG VNYOPFRVVVL
PnGu1_2019           KSNLKPFERDISTEITYOAGSTPCNGVEGFNCYFPLOS YGFOPTNGVGYOPYRVVVL
BtZC45                STKLNKPFERDLSSDE-----NGVR-----TLSTYDFNPNVPLEYQATRVVVL
BtZXC21              STKLNKPFERDLSSDE-----NGVR-----TLSTYDFNPNVPLEYQATRVVVL
BtYu-RmYN02_2019    AVKLNKPFERDLSSDE-----NGVR-----TLSTYDFNPNVPLDYQATRVVVL
BtYN2013             STKLNKPFERDLSSDE-----NGVR-----TLSTYDFNPNVPLEYQATRVVVL
BtHKU3-12            KTKLNKPFERDLSSDD-----GNGVY-----TLSTYDFNPNVPLVAYQATRVVVL
BtGX2013             KTKLNKPFERDLSSDD-----GNGVY-----TLSTYDFNPNVPLVAYQATRVVVL
BtYN2018B           HGKLRPFERDISNVFFSPDGKPCTPPA-FNCYWPLNDYGFYITNGIGYOPYRVVVL
BtRs4874             HGKLRPFERDISNVFFSPDGKPCTPPA-FNCYWPLNDYGFYITNGIGYOPYRVVVL
Cv007-2004          HGKLRPFERDISNVFFSDGKPCTPPA-PNCYWPLRGYGYTTSGIGYOPYRVVVL
HuSARS-Frankfurt-1_2003  HGKLRPFERDISNVFFSPDGKPCTPPA-LNCYWPLNDYGFYITNGIGYOPYRVVVL
0.....480.....490.....500.....510.....520.....

```

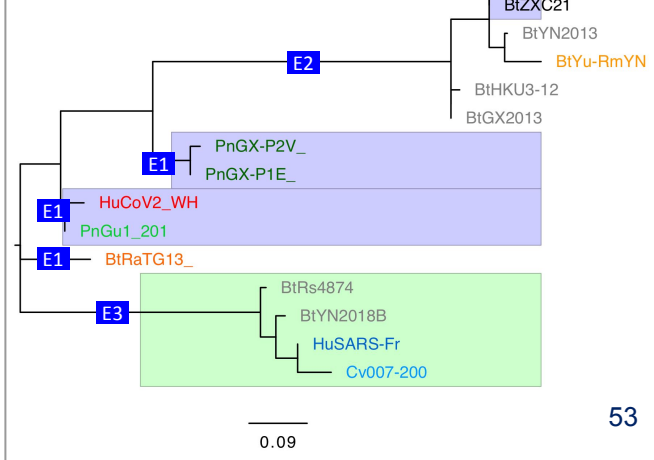
E1
E2
E3



Arbre des génomes



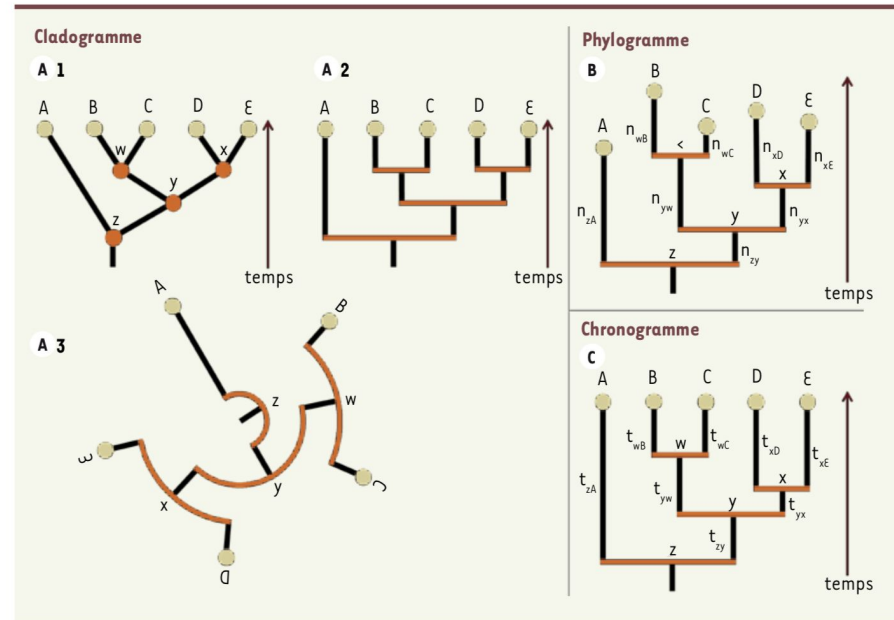
Arbre de la région protéique

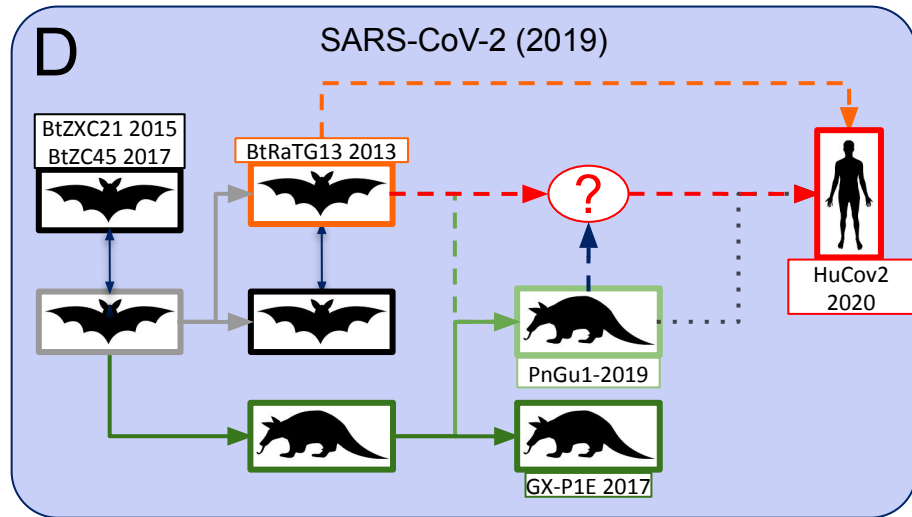
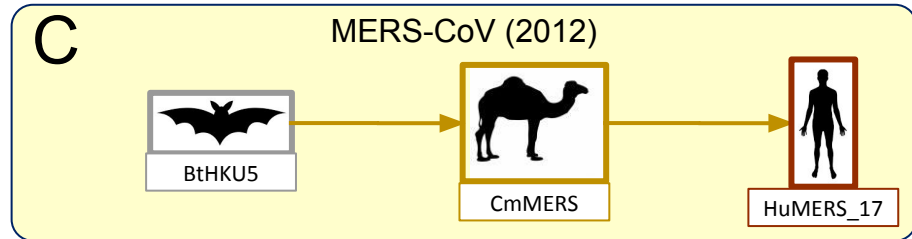
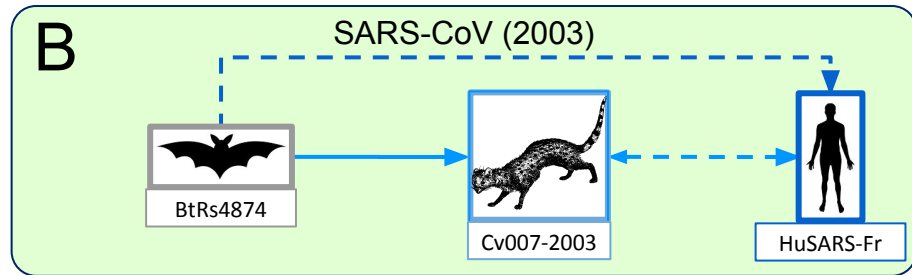
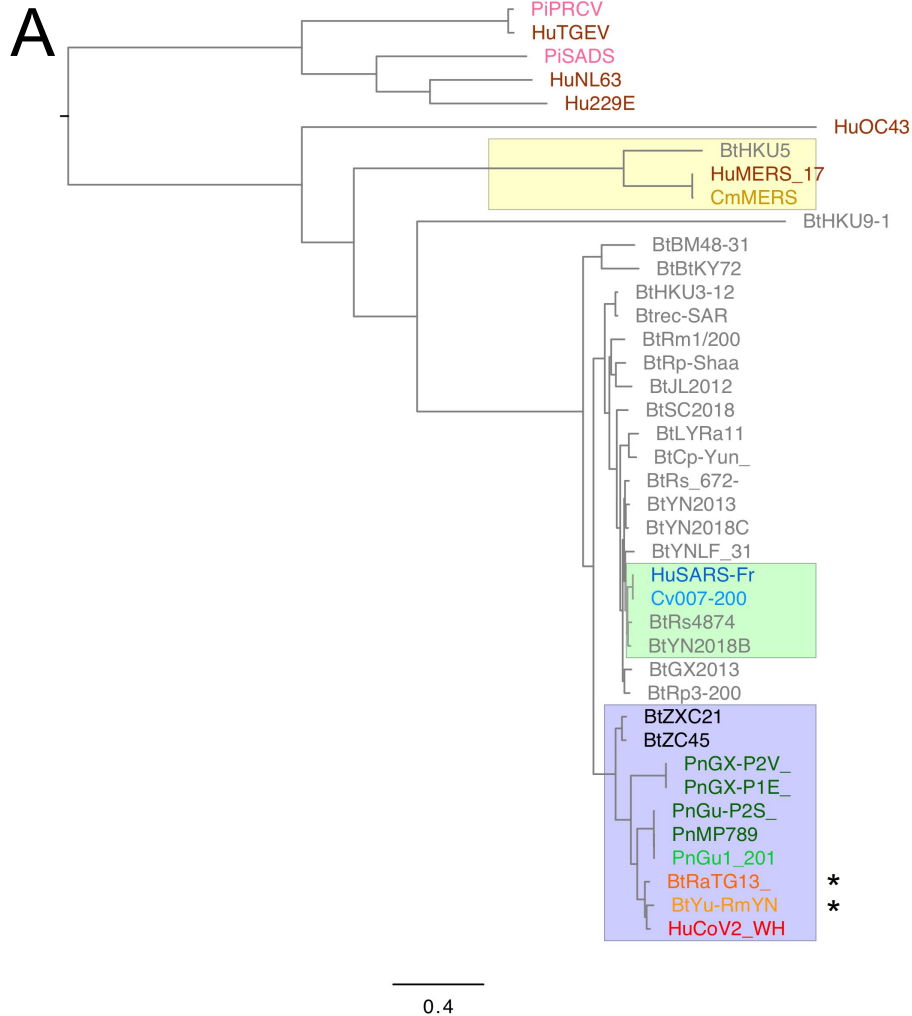


Inférer l'histoire évolutive à partir des séquences génomiques

Représentations arborescentes des histoires évolutives

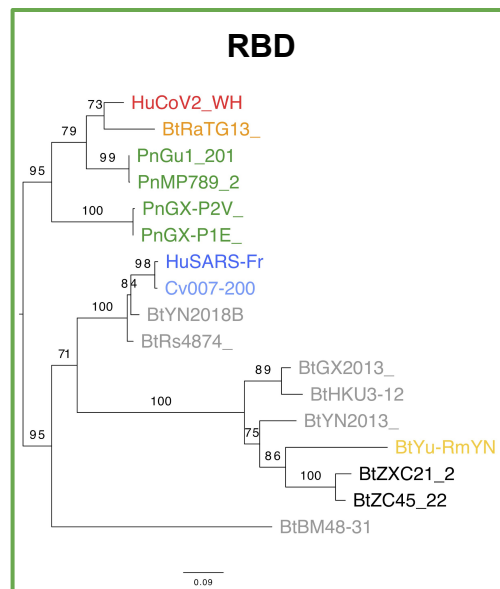
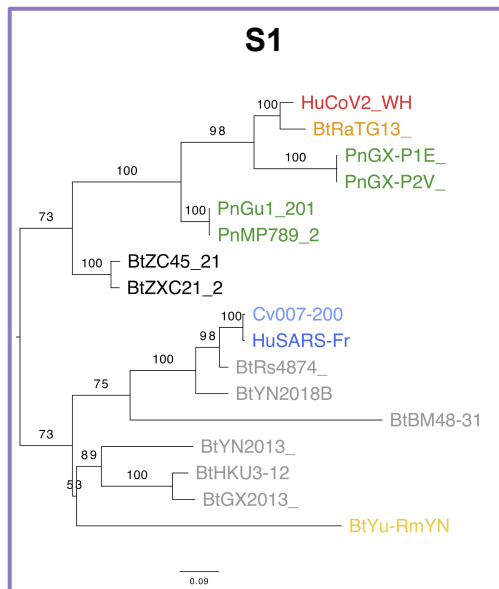
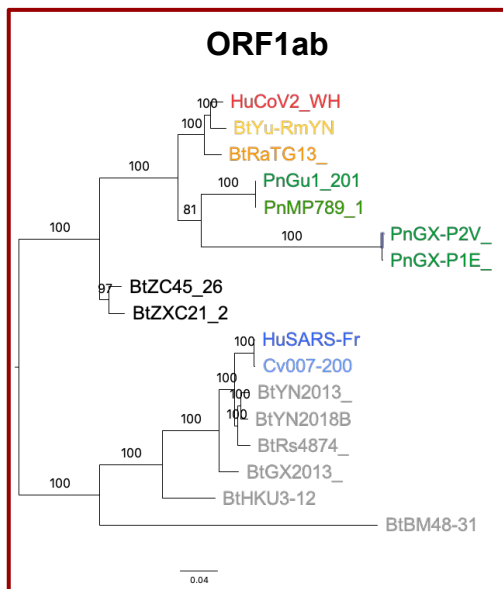
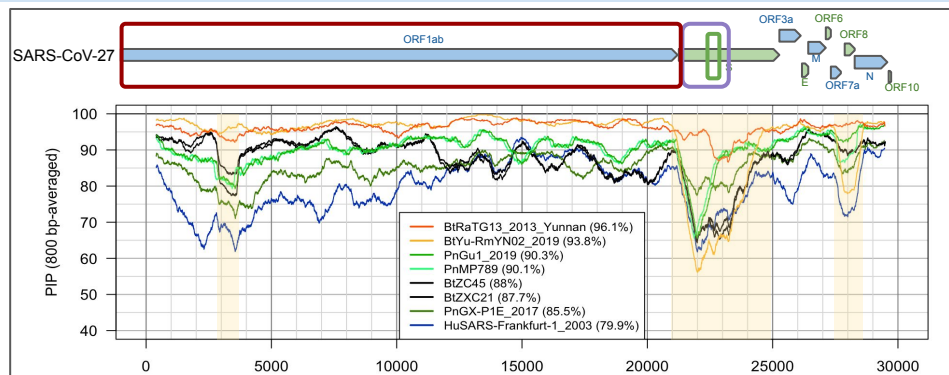
- On représente les histoires évolutives sous forme d'arbre
- Différents types de représentation peuvent être utilisés selon les cas.
 - Bifurcations triangulaires ou rectangulaires
 - Disposition radiale
- Selon les cas, les longueurs des branches représentent
 - Le nombre de divergences (cladogramme)
 - le nombre de différences entre deux espèces (phylogramme)
 - Le temps de divergence (chronogramme)





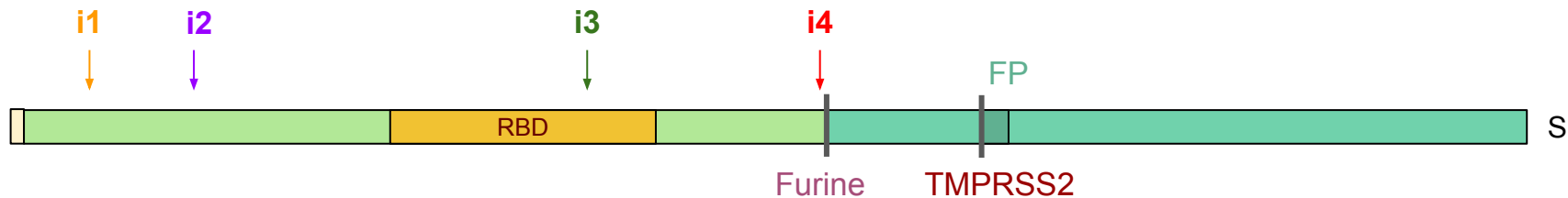
Comparaison entre coronavirus - gène S

Profils de pourcentages de positions identiques (PPI) entre différents génomes de coronavirus (d'humain, de chauve-souris et de pangolin) et SARS-CoV-2 (la référence qui correspond à 100% sur toute la largeur).



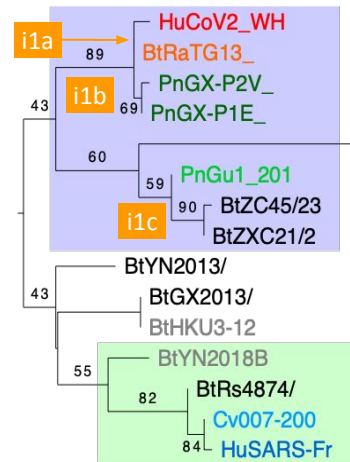
- Alignements multiples

Insertion 1



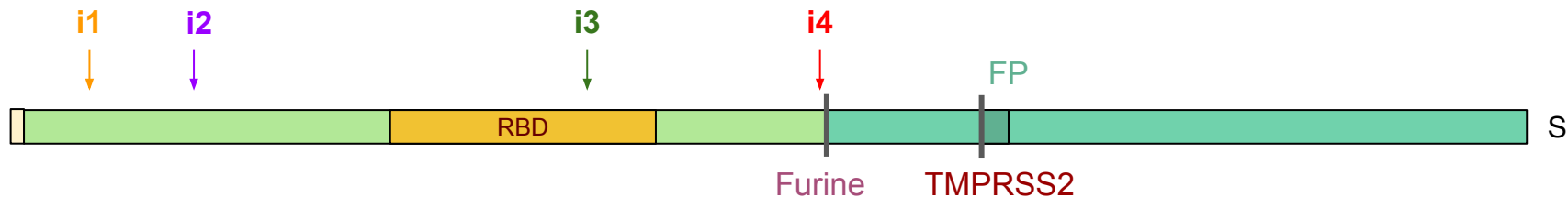
Strain	Sequence
HuCoV2_WH01_2019	WFHAIHVSGTNGTKRFDNP
BtRaTG13_2013_Yunnan	WFHAIHVSGTNGIKRFDNP
PnGX-P1E_2017	WFNTI--NYQGGFKKFDNP
PnGX-P2V_2018	WFNTIHLNYQGGFKKFDNP
PnGu1_2019	WYYAL-TKTNSAEKRVDNP
BtZC45	WYYSL-TTNNAATKRFDNP
BtZXC21	WYYSL-TTNNAATKRFDNP
BtYu-RmYN02_2019	WYNFW-----NQAYTSR
BtYN2013	QFFTQ-----GTNIDNP
BtHKU3-12	QYFSL-NVSDRYTYFDNP
BtGX2013	QYFSL-NVSDRYTYFDNP
BtYN2018B	RFITF-----GLNFDNP
BtRs4874	GFHTI-----NHRFDNP
Cv007-2004	GFHTI-----NHTFDNP
HuSARS-Frankfurt-1_2003	GFHTI-----NHTFGNP

.....80.....90

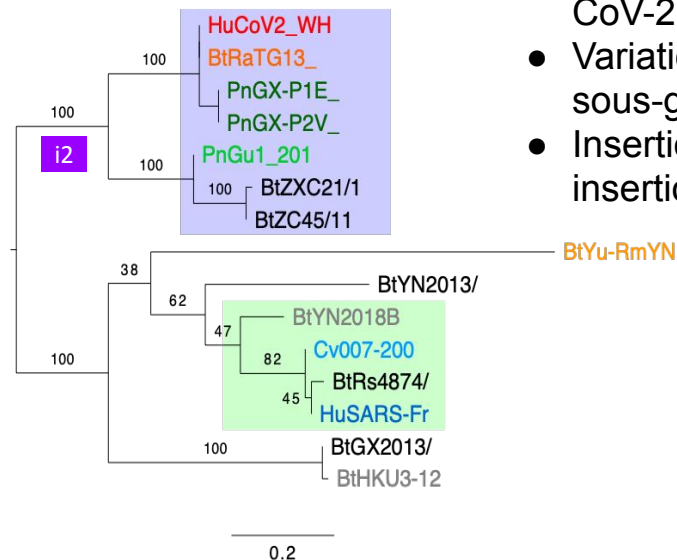


- Insertion retrouvée dans tous les génomes de la branche CoV-2.
- Variations selon les sous-groupes.
- Insertions indépendantes ou insertion suivie de mutations ?

Insertion 2

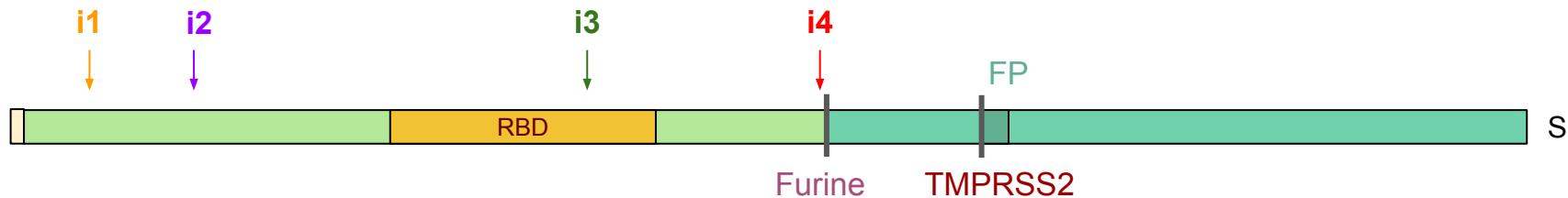


HuCoV2_WH01_2019	YYHKNNKSWMESEFRVYSS	.	::
BtRaTG13_2013_Yunnan	YYHKNNKSWMESEFRVYSS		
PnGX-P1E_2017	YYHNNNKTWVENEFRVYSS		
PnGX-P2V_2018	YYHNNNKTWVENEFRVYSS		
PnGu1_2019	YYH-NNKTWSTREFAVYSS	i2	
BtZC45	YYH-NNKTWSIREFAVYSS		
BtZXC21	YYH-NNKTWSIREFAVYSF		
BtYu-RmYN02_2019	AGGQOTSAA-----VYIS		
BtYN2013	FKSNNSQLSH-----LFS		
BtHKU3-12	SRGTQONAW-----VYQS		
BtGX2013	SRGTQONS-----VYQS		
BtYN2018B	LRSNNTQIPSY----IFNN		
BtRs4874	SKPTGTQTHM----IFDN		
Cv007-2004	SKPMGTQTHM----IFDN		
HuSARS-Frankfurt-1_2003	SKPMGTQTHM----IFDN		
160.....170		



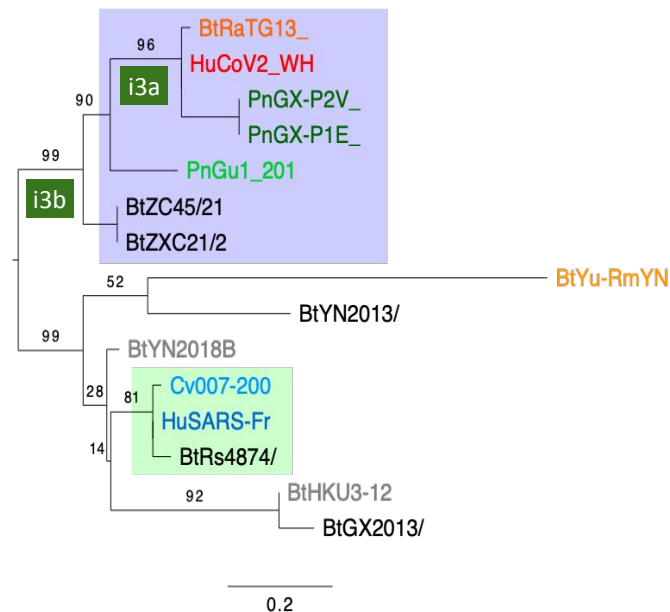
- Insertion retrouvée dans tous les génomes de la branche CoV-2.
- Variations selon les sous-groupes.
- Insertions indépendantes ou insertion suivie de mutations ?

Insertion 3

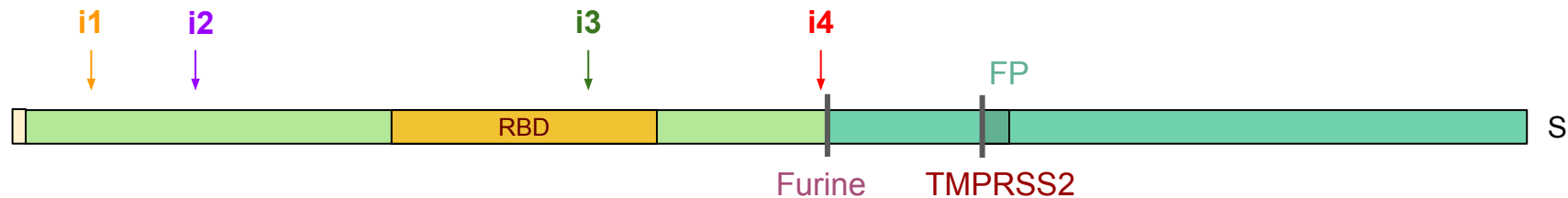


```

HuCoV2_WH01_2019      ...:
BtRaTG13_2013_Yunnan i3a LLALHRSYLTTPGDSSSGWTA
PnGX-P1E_2017         LLALHRSYLTTPGKLESGWTT
PnGX-P2V_2018        LLALHRSYLTTPGKLESGWTT
PnGu1_2019           LLTIHRGDPMP---NNGWTV
BtZC45                LLTIHRGDPMP---NNGWTA
BtZXC21              LLTIHRGDPMS---NNGWTA
BtYu-RmYN02_2019     VLTF-----RSNSQP
BtYN2013              FLAVYRVA-----AGSISV
BtHKU3-12             VMAMFSQT-----TSNFLP
BtGX2013              VMAMFSQS-----TSNFLP
BtYN2018B            LLTAFPPN-----PGYWGT
BtRs4874              ILTAFSPA-----QDTWGT
Cv007-2004           ILTAFSPA-----QGTWGT
HuSARS-Frankfurt-1_2003 ILTAFSPA-----QDIWGT
.....260.....270..
    
```

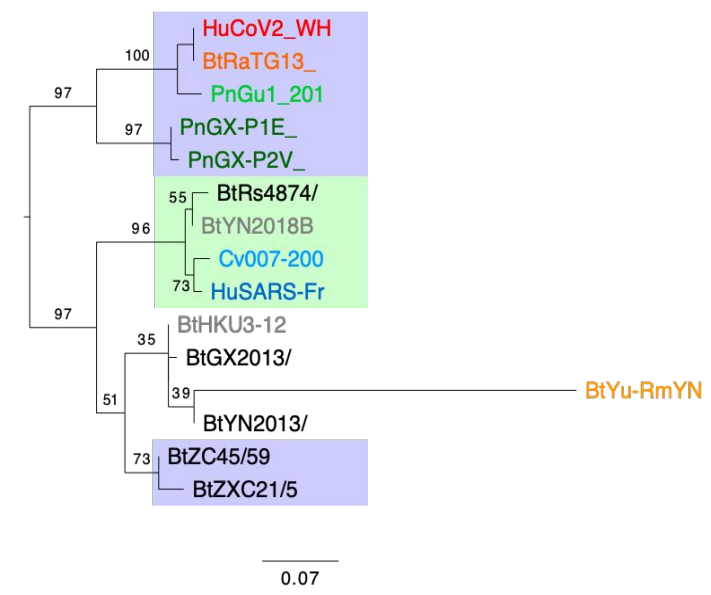


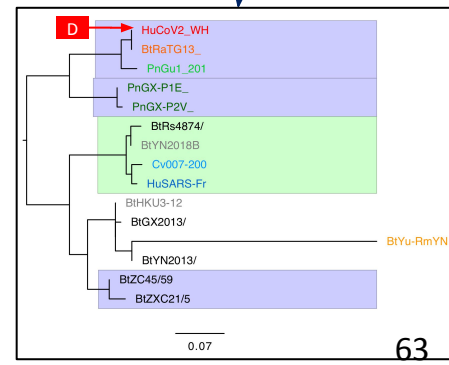
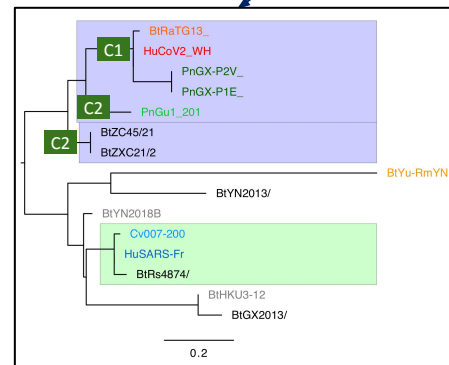
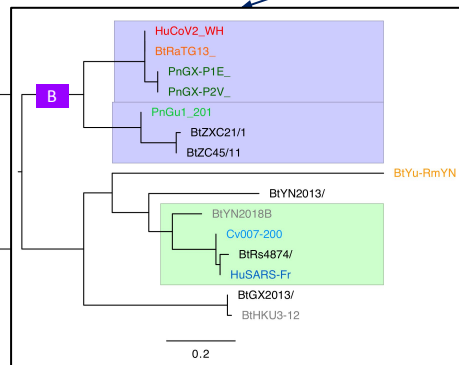
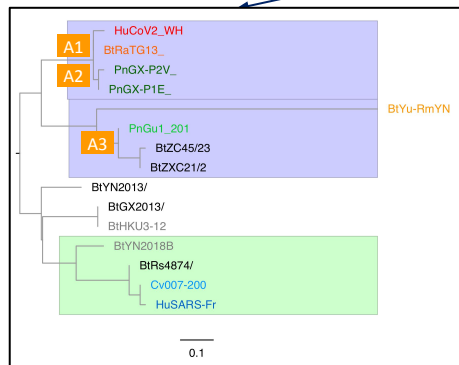
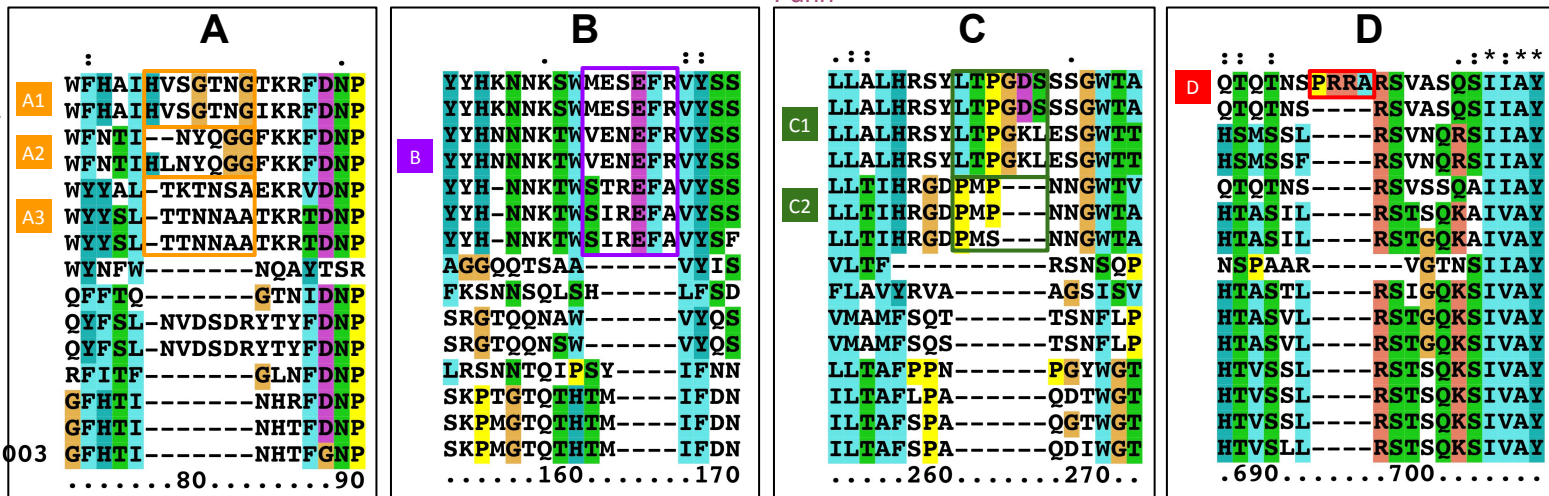
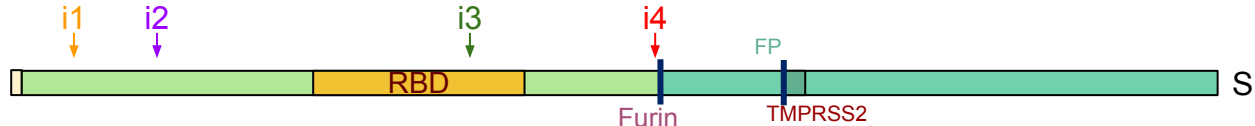
Insertion 4



Sequence	Conservation	Insertion Site
HuCoV2_WH01_2019	QTQTNSPRRARSVASQSIIAY	i4
BtRaTG13_2013_Yunnan	QTQTNS----RSVASQSIIAY	
PnGX-P1E_2017	HSMSSL----RSVNORSIIAY	
PnGX-P2V_2018	HSMSSF----RSVNORSIIAY	
PnGu1_2019	QTQTNS----RSVSSQAI IAY	
BtZC45	HTASIL----RSTSOKAIVAY	
BtZXC21	HTASIL----RSTGOKAIVAY	
BtYu-RmYN02_2019	NSPAAR-----VGTNSIIAY	
BtYN2013	HTASTL----RSIGOKSIVAY	
BtHKU3-12	HTASVL----RSTGOKSIVAY	
BtGX2013	HTASVL----RSTGOKSIVAY	
BtYN2018B	HTVSLL----RSTSOKSIVAY	
BtRs4874	HTVSLL----RSTSOKSIVAY	
Cv007-2004	HTVSLL----RSTSOKSIVAY	
HuSARS-Frankfurt-1_2003	HTVSLL----RSTSOKSIVAY	

.690.....700.....



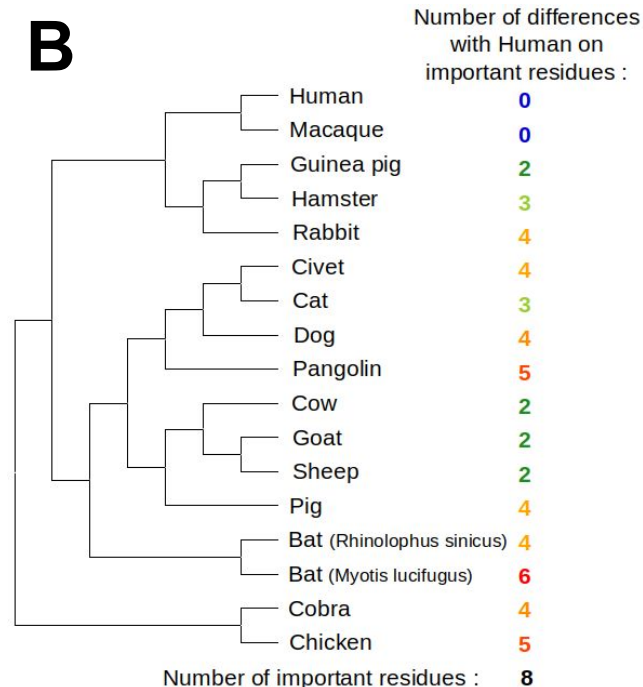


A

ACE2 receptor



S protein

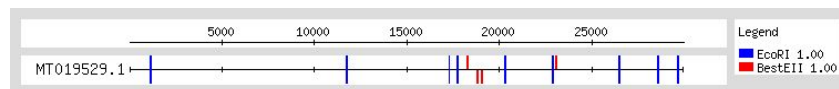
B

Un virus construit par ingénierie moléculaire ?

Un virus construit par ingénierie moléculaire ?

- Sept 2020: un preprint fait du bruit
- Li-Meng Yan
 - chercheuse chinoise
 - travaillait dans le laboratoire de référence de l'OMS pour la Chine
 - réfugiée aux Etats-Unis
- 600.000 téléchargements en 10 jours
- Arguments
 - Présence de sites de restriction dans le génome de SARS-Cov-2

The screenshot shows the Zenodo interface for a preprint. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The preprint title is 'Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route' by Yan, Li-Meng; Kang, Shu; Guan, Jie; Hu, Shanchang. It is dated September 14, 2020, and is marked as a 'Working paper' and 'Open Access'. The page shows 776,634 views and 597,172 downloads. The abstract discusses the COVID-19 pandemic and the origin of SARS-CoV-2, arguing for a laboratory origin based on genomic, structural, and medical evidence. It mentions that the natural origin theory lacks substantial support and that the alternative theory of laboratory origin is more consistent with the data. The paper is indexed in OpenAIRE and has a DOI of 10.5281/zenodo.4028830. It is associated with the 'Coronavirus Disease Research Community - COVID-19' and has a license for files.



Des sites de restriction créés dans le génome de SARS-CoV-2 ?

- Li-Meng Yan détecte dans la séquence de SARS-CoV-2 des séquences correspondant aux sites de restriction de EcoRI et BstEII, souvent utilisés en biologie moléculaire pour créer de l'ADN recombinant au moyen d'enzymes de restriction ("ciseaux" moléculaires).
- Elle souligne que ces sites auraient pu être facilement créés à partir de sites très similaires présents dans le virus de chauve-souris Bat ZC45.
- Des questions se posent sur la signification de ce résultat
 - Probabilité de trouver ces sites aléatoirement ?
 - Les sites sont-ils uniques à SARS-CoV-2 dans le lignage des SARS-CoV ?

A SARS-CoV-2		EcoRI						
		W	N	S				
tataattata	aattaccaga	tgattttaca	ggctgcgta	tagcttg	gaa	ttg	taacaat	1320
cttgattcta	aggttggtgg	taattataat	tacctgtata	gattgtttag	gaagtcta	aat		1380
ctcaaacctt	ttagagagaga	tatttcaact	gaaatctatc	aggccggtag	cacaccttgt			1440
aatggtgttg	aaggttttaa	ttgttacttt	cctttacaat	catatggttt	ccaaccact			1500
aatggtgttg	gttacc	atacagagta	gtagtacttt	cttttgaact	tctacatgca			1560
	G	Y	Q					
		BstEII						
B ZC45		W		N	T			
ttacctgatg	atthtacagg	ttgtgtcata	gcttg	gaaca	ct	tgccaaaca	ggatgtaggt	1320
aattatttct	acaggtctca	tcgttctacc	aaattgaaac	catttgaaag	agatctttcc			1380
tcagacgaga	atggtgtccg	tacacttagt	acttatgact	tcaaccctaa	tgtaccactt			1440
gaatacca	aaag	ctacaagggt	tgttgttttg	tcatttgagc	ttctaaatgc	accagctaca		1500
	E	Y	Q					

- An EcoRI site is found at the 5'-end of the RBM and a BstEII site at the 3'-end.*
- Although these two restriction sites do not exist in the original spike gene of ZC45, they can be conveniently introduced given that the sequence discrepancy is small (2 nucleotides) in either case.*

Un virus construit par ingénierie moléculaire ?

- Le site de restriction **EcoRI** (séquence **GAATTC**) se trouve dans le virus humain, mais également dans les virus les plus proches de chauve-souris (**RatG13**) et de pangolin (**MP789**).
- La majorité de ce site, ainsi que les séquences avoisinantes (**TTGGAAT*CT**) est conservée dans l'ensemble de la lignée SARS.
- Les séquences des autres SARS sont encore plus proche du site SARS-CoV-2 que celui de BtZ45 mis en avant par Li-Meng Yan.
- Il est donc plus vraisemblable que le site de SARS-CoV-2 soit apparu par une substitution d'un seul nucléotide à partir de n'importe quel SARS que par modification de 2 nucléotides du virus de chauve-souris BtZC45.

	EcoRI
HuCoV2_WH01_2019_215	ACAGGCTGCGTTATAGCTTGAATTCGAACAA
BtRaTG13_2013_Yunnan_215	ACTGGTTGTGTTATAGCTTGAATTCGAACCA
PnMP789_214	ACAGGTTGTGTAATAGCTTGAATTCGAACAA
PnGX-P1E_2017_215	ACTGGTTGTGTTATTGCTTGAATTCAGTTAA
PnGX-P2V_2018_215	ACTGGTTGTGTTATTGCTTGAATTCAGTTAA
BtZC45_215	ACAGGTTGTGCATAGCTTGAACACAGCCAA
BtZXC21_214	ACAGGTTGTGCATAGCTTGAACACAGCCAA
BtLYRa11_215	ATGGGTTGTGCTTCTGCTTGAACACAGGAA
BtRs4874_214	ACGGGTTGTGCTTCTGCTTGAATACAGGAA
BtYN2018B_215	ATGGGTTGTGCTTCTGCTTGAATACAGGAA
huSARS-Frankfurt-1_2003_215	ATGGGTTGTGCTTCTGCTTGAATACAGGAA
Cv007-2004_214	ATGGGTTGTGCTTCTGCTTGAATACAGGAA
BtHKU3-12_214	ACTGGCTGTGTAATTGCTTGAATACAGCTAA
Btrec-SARsg_2008_215	ACTGGCTGTGTAATTGCTTGAATACAGCTAA
BtYN2013_212	ACAGGCTGTGCATAGCTTGAATACAGCTAA
BtRm1/2004_215	ACAGGCTGTGTAATAGCTTGAATACAGCTAA
BtRp3-2004_214	ACTGGTTGGGTAATAGCTTGAATACAGCCAA
BtGX2013_212	ACTGGCTGTGTAATAGCTTGAATACAGCTAA
BtRp-Shaanxi2011_213	ACAGGCTGTGTAATAGCTTGAACACAGCCAA
BtSC2018_214	ACTGGCTGTGTAATAGCTTGAATACAGCTAA
BtCp-Yun_2011_214	ACAGGTTGGGTAATTGCTTGAATACAGCTAA
BtRs_672-2006_208	ACAGGCTGTGCATAGCTTGAACACAGCTAA
BtYN2018C_215	ACAGGCTGTGTTATTGCTTGAATACAGCTAA
BtYNLF_31C_215	ACAGGCTGTGTTATAGCTTGAACACAGCCAA
BtJL2012_212	ATAGGTTGTGTTATAGCTTGAACACAGCCAA
BtBtKY72_214	ACTGGCTGTGTTTATAGCTTGAATACAGCTAA
BtBM48-31_213	ACAGGTTGTGTAATAGCTTGAATACAGCTAA
BtHKU5_219	TCT---TCACCAATTGCAATTACAACCTAACCA
HuOC43_241	ACAAGTTGTGAGTTGTAATTAATTTACCTGTC
CmMERS_218	CCCACATGTTTGTATTTAGCGACTGTTCCCTCA
HuMERS_172-06_2015_218	CCCACATGTTTGTATTTAGCGACTGTTCCCTCA
HuTGEV_210	TATTGATTTGATATCTTTTAAATTTGACCAGT
BtHKU9-1_211	TACGGTTGTTTGCATGCAATCTAATTTGAATTC
Hu229E_205	-----GCTAGTATTAAACACGGGAAA
HuNL63_210	---TCATGGCACATTTATTTAAAGAGTGGCAC
PiSADS_205	CCTTATGAATGTTTGGGTTGTCATGGAATGA
PiPRCV_205	-----GCCACCGCTGTTATAAAAAGTGGTAC

Quasi-identique
à EcoRI

Un virus construit par ingénierie moléculaire ?

- Le site de restriction BstEII (**séquence GGTTACC**) se retrouve dans le virus humain, mais également dans les virus les plus proches de chauve-souris (**RatG13**) et de pangolin (**MP789**).
- Ce site se trouve également dans d'autres virus de la lignée, notamment ceux du SRAS humain (HuSARS-Frankfurt-1_2003), et de la civette (Cv007_2004).
- L'hypothèse de Li-Meng Yan (création à partir du génome de chauve-souris BtZC45) n'est donc pas convaincante.

BstEII

HuCoV2_WH01_2019_215	GGTGTGGTTACCAACCATACAGAA
BtRaTG13_2013_Yunnan_215	GGTGTGGTCCACCAACCTTATAGGA
PnMP789_214	GGTGTGGTTACCAACCTTATAGAA
PnGX-P1E_2017_215	GGTGTAACTACCAACCTTTTAGAA
PnGX-P2V_2018_215	GGTGTAACTACCAACCTTTTAGAA
BtZC45_215	CCACTTGAATACCAAGCTACAAGGA
BtZXC21_214	CCGGTTGAATATCAAGCTACAAGGA
BtLYRa11_215	GGCATGGTTACCAACCTTATAGAA
BtRs4874_214	GGCATGGCTACCAACCTTATAGAA
BtYN2018B_215	GGCATGGCTATCAACCTTATAGAA
HuSARS-Frankfurt-1_2003_215	GGCATGGCTACCAACCTTACAGAA
Cv007-2004_214	GGCATGGCTACCAACCTTACAGAA
BtHKU3-12_214	CCAGTAGCATATCAGGCTACTAGGA
Btrec-SARSg_2008_215	CCAGTAGCATATCAGGCTACTAGGA
BtYN2013_212	CCCTCTGATTATCAAGCCACCAGAA
BtRm1/2004_215	CCAGTTGAATACCAAGCCACTAGGA
BtRp3-2004_214	CCGGTTGCCTATCAGGCTACTAGGA
BtGX2013_212	CCAGTGGCATATCAGGCTACTAGGA
BtRp-Shaanxi2011_213	CCACTTGAATATCAGGCTACTAGGA
BtSC2018_214	CCGGTGGCATATCAGGCTACTAGAA
BtCp-Yun_2011_214	CCACTTGAATACCAAGCTACTAGAA
BtRs_672-2006_208	CCATATTGAATATCAGGCTACTAGGA
BtYN2018C_215	CCATATTGAATATCAGGCTACTAGGA
BtYNLF_31C_215	CCCCTTGAATATCAAGCCACTAGAA
BtJL2012_212	CCCTCTTGAGTACCAAGCCACTAGAA
BtBtKY72_214	GGTGTGGTTACCAACCATATAGAA
BtBM48-31_213	GGAAATTGGCTTTCAACCATACAGAA
BtHKU5_219	GTTACTCTTCTCTTACAGTGGACTT
HuOC43_241	AAGTGCCCCCAAACCTAAATCTTTAA
CmMERS_218	CCACTTGAAGGTGGTGGCTGGCTT
HuMERS_172-06_2015_218	CCACTTGAAGGTGGTGGCTGGCTT
HuTGEV_210	AAACACAGCTATTACAAAGGTGAC
BtHKU9-1_211	CCTTTTCTTAT---GTTTATGGT
Hu229E_205	ATACCCGGTGGTTGCGCAATGCC
HuNL63_210	GTGCCTGGTAGTTGTAATTTTCCG
PiSADS_205	-----GGTGG
PiPRCV_205	GTTGGTGTAAATGTAAAGTTTGA

Que peut-on conclure à ce stade ?

Que peut-on conclure à ce stade ?

- Les éléments disponibles ne permettent pas de démontrer si le génome est d'origine naturelle ou artificielle
- Comment progresser sur cette question, selon les différentes hypothèses ?
 - Zoonose récente d'origine naturelle : intensifier la collecte d'échantillons animaux (sites naturels, élevages, marchés, laboratoires).
 - Circulation à bas bruit dans les populations humaines: caractériser les virus dans les échantillons prélevés depuis 2013 chez des patients atteints de pneumonies atypiques (en particulier les mineurs de 2013).
 - Virus développé en laboratoire : contrôle du laboratoire par une commission indépendante.

Reporté à la 4ème séance

Retracer par analyse des génomes la propagation du virus
dans les populations humaines

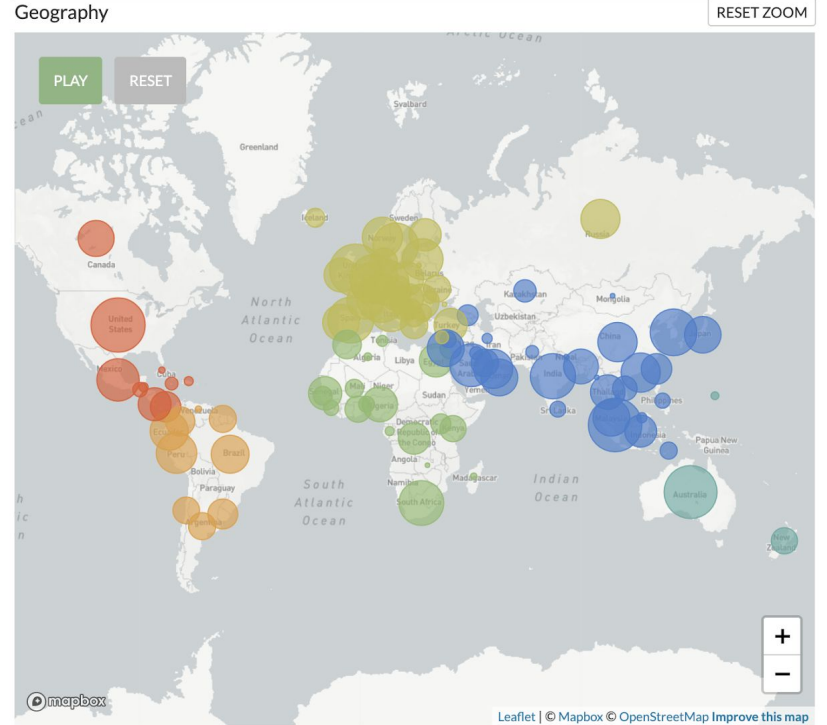
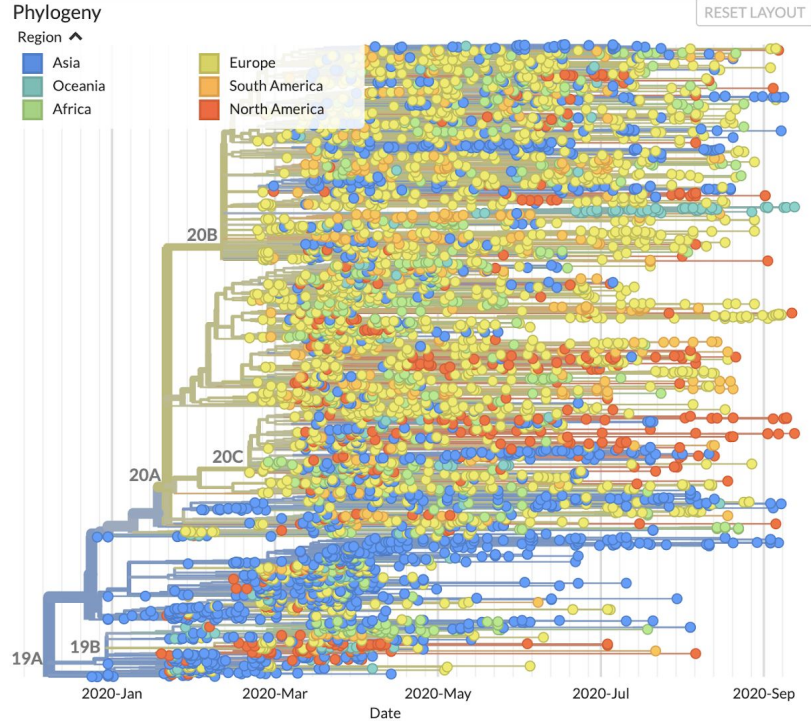
Jacques van Helden

Un outil informatique pour retracer la progression de l'épidémie

Genomic epidemiology of novel coronavirus - Global subsampling

Maintained by the Nextstrain team. Enabled by data from **GISAID**

Showing 4678 of 4678 genomes sampled between Dec 2019 and Sep 2020.



Evolution du virus pendant sa propagation

Review

SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective



Manuela Sironi^a, Seyed E. Hasnain^b, Benjamin Rosenthal^j, Tung Phan^c, Fabio Luciani^d, Marie-Anne Shaw^e, M. Anice Sallum^f, Marzieh Ezzaty Mirhashemi^g, Serge Morand^h, Fernando González-Candelas^{i,*,} on behalf of the Editors of *Infection, Genetics and Evolution*

M. Sironi, et al.

Infection, Genetics and Evolution 84 (2020) 104384

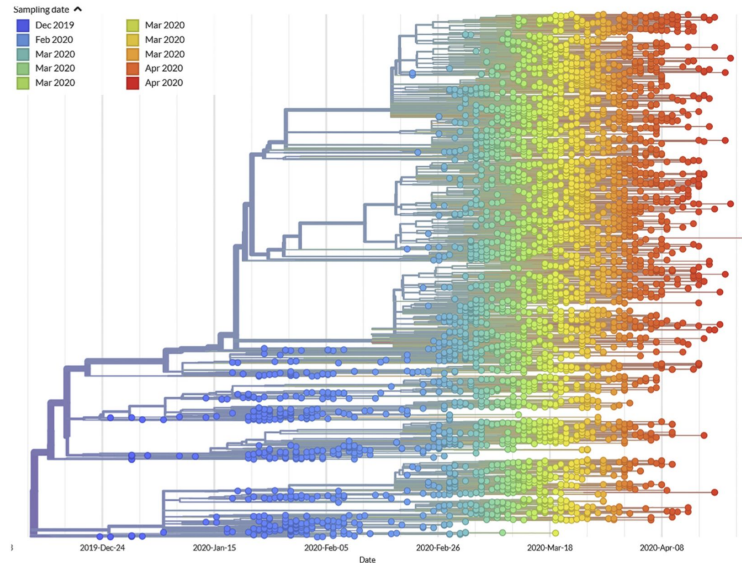


Fig. 3. Time-stamped maximum likelihood phylogenetic reconstruction of SARS-CoV-2 isolates deposited to GISAID.org and rendered by <https://nextstrain.org/ncov>. Isolates are represented by colored circles with the color code corresponding to time of sampling as detailed in the legend.

M. Sironi, et al.

Infection, Genetics and Evolution 84 (2020) 104384

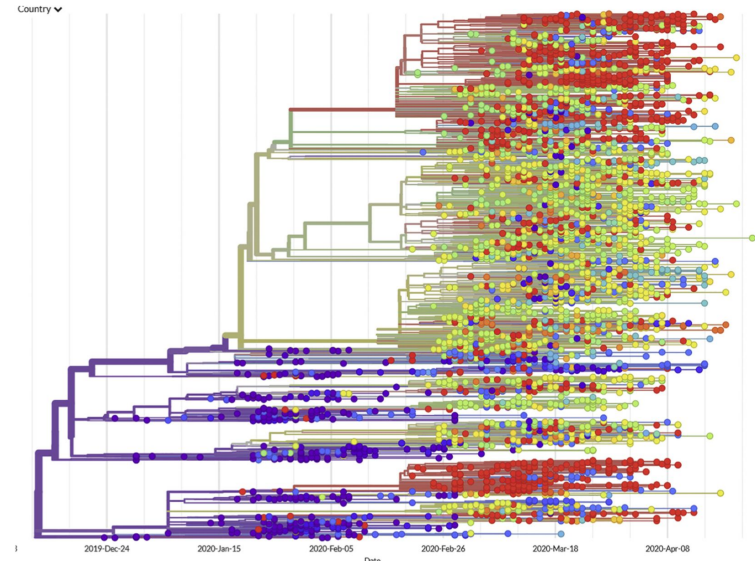


Fig. 4. The same reconstruction of SARS-CoV-2 phylogeny, now denoted by geography. Isolates originated and initially diversified in China (purple), followed by multiple and independent introductions to Oceania (blue), Europe (green and yellow), and North America (red). Less information is known about Africa, India, South America, and other populations of major concern in the Global South. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Special Section: SARS-CoV-2

On the origin and continuing evolution of SARS-CoV-2

Xiaolu Tang^{1,†}, Changcheng Wu^{1,†}, Xiang Li^{2,3,4,†}, Yuhe Song^{2,5,†}, Xinmin Yao¹,
Xinkai Wu¹, Yuange Duan¹, Hong Zhang¹, Yirong Wang¹, Zhaohui Qian⁶,
Jie Cui^{ID 2,3,*} and Jian Lu^{ID 1,*}

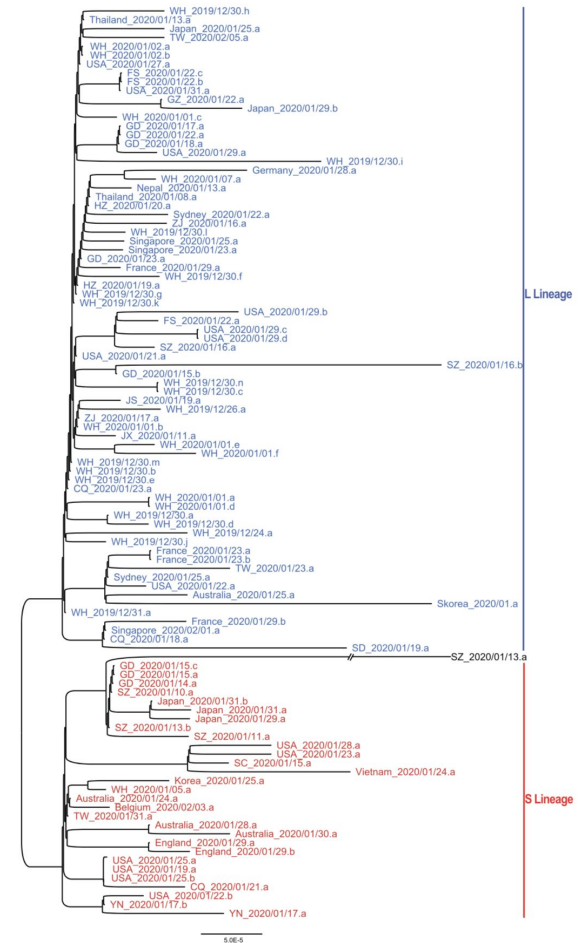
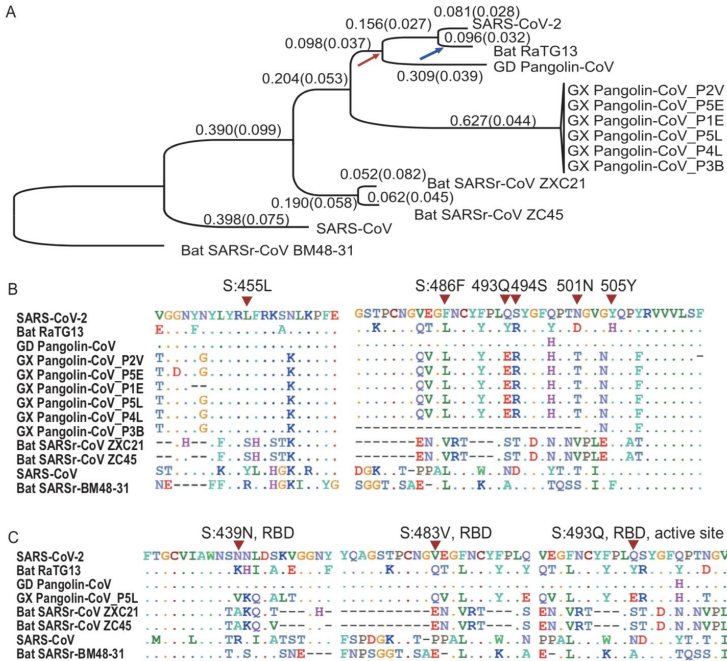


Figure 5. The unrooted phylogenetic tree of the 103 SARS-CoV-2 genomes. The ID of each sample is the same as in Fig. 4A. Note WH/2019/12/31.a represents the reference genome (NC_045512). Note SZ.2020/01/13.a had C at both positions 8,782 and 28,144 in the genome, belonging to neither L nor S lineage.

Matériel supplémentaire

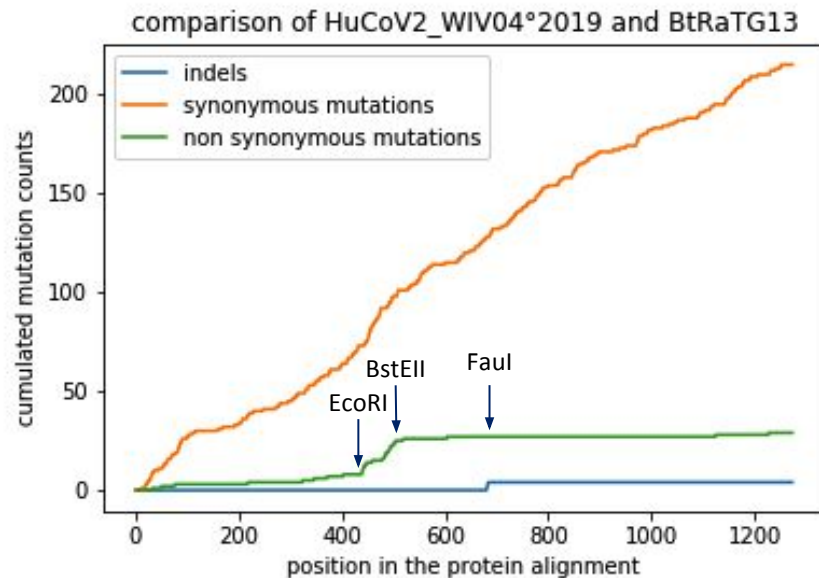
L'hypothèse de la reine rouge



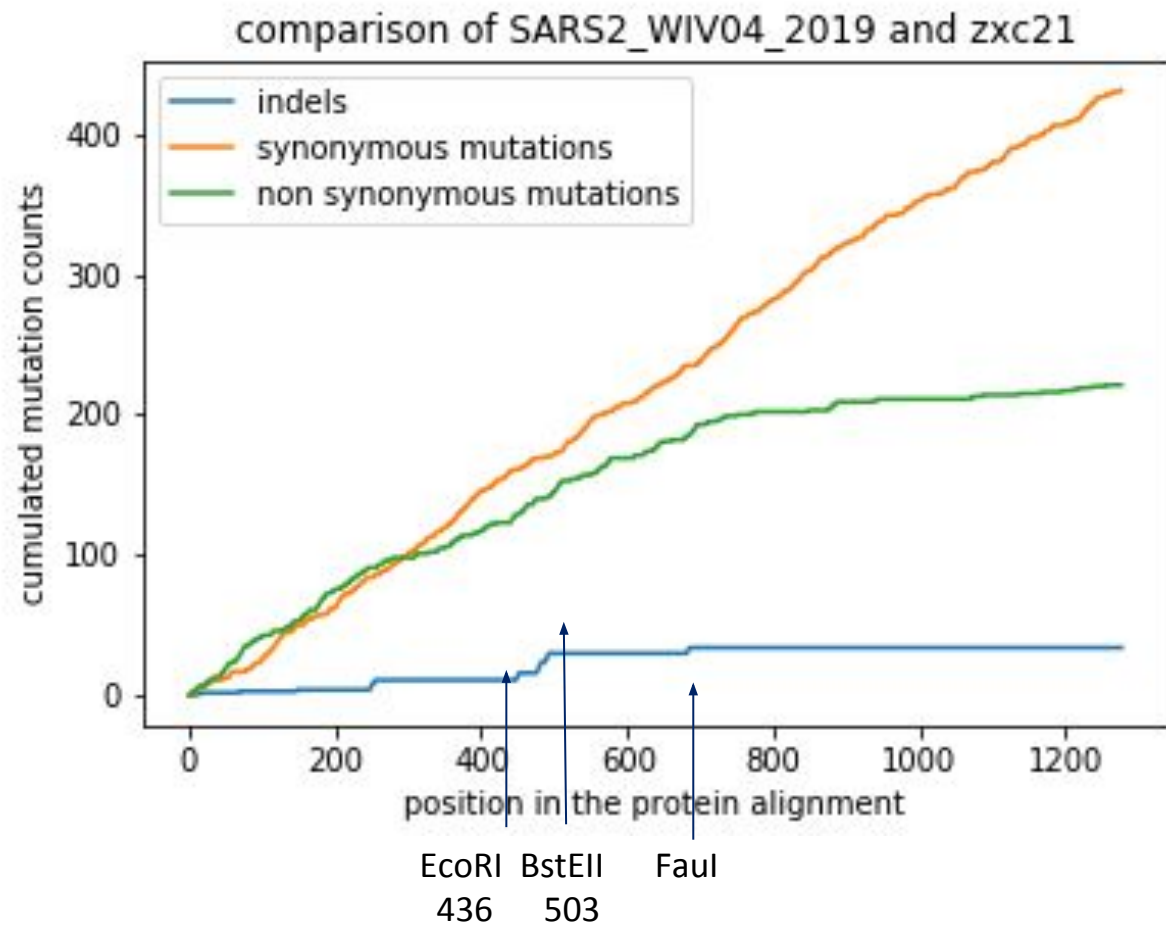
Dégénérescence du code et mutations synonymes

Profils de mutations synonymes sur la protéine S

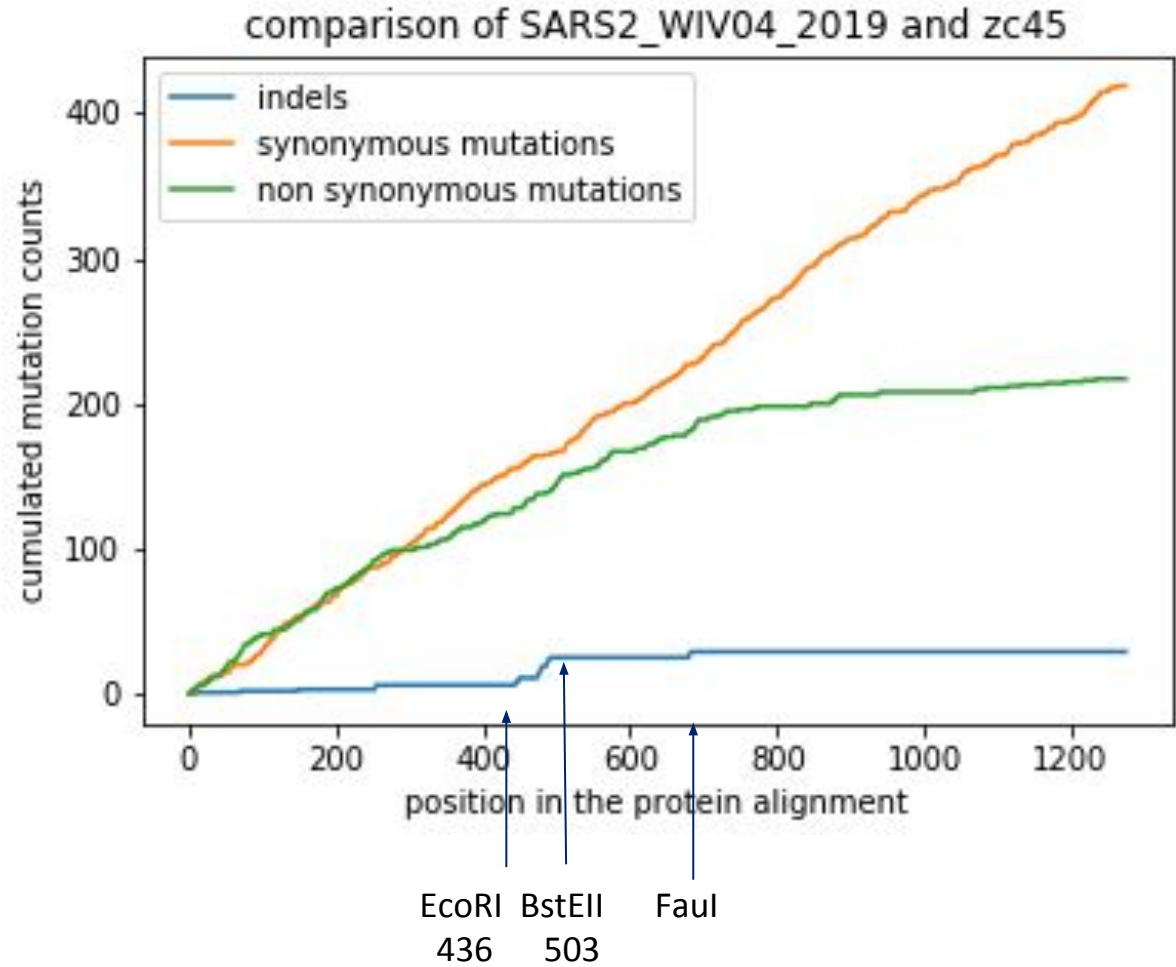
- Les plateaux horizontaux de mutations non-synonymes indiquent des régions soumises à forte pression sélective
-



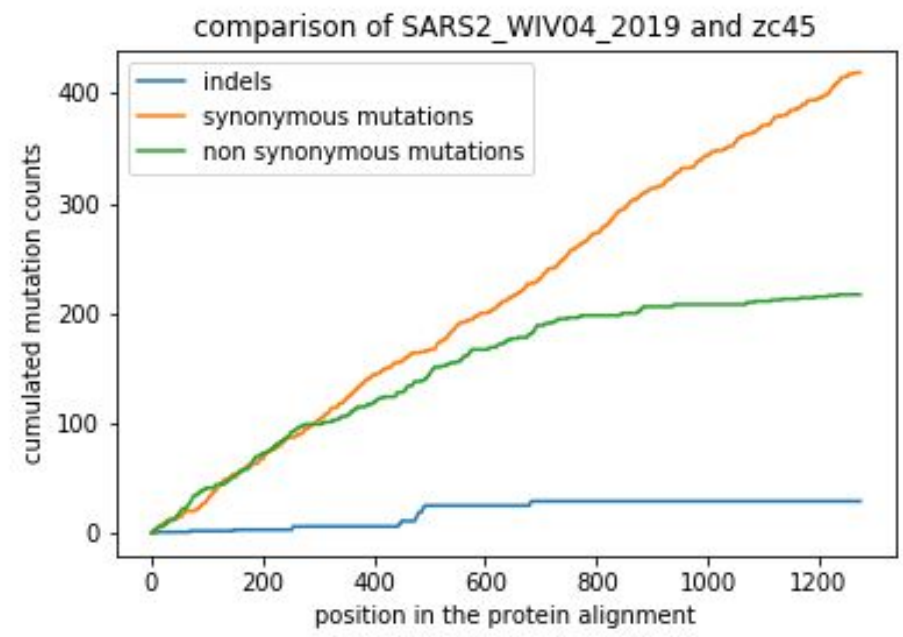
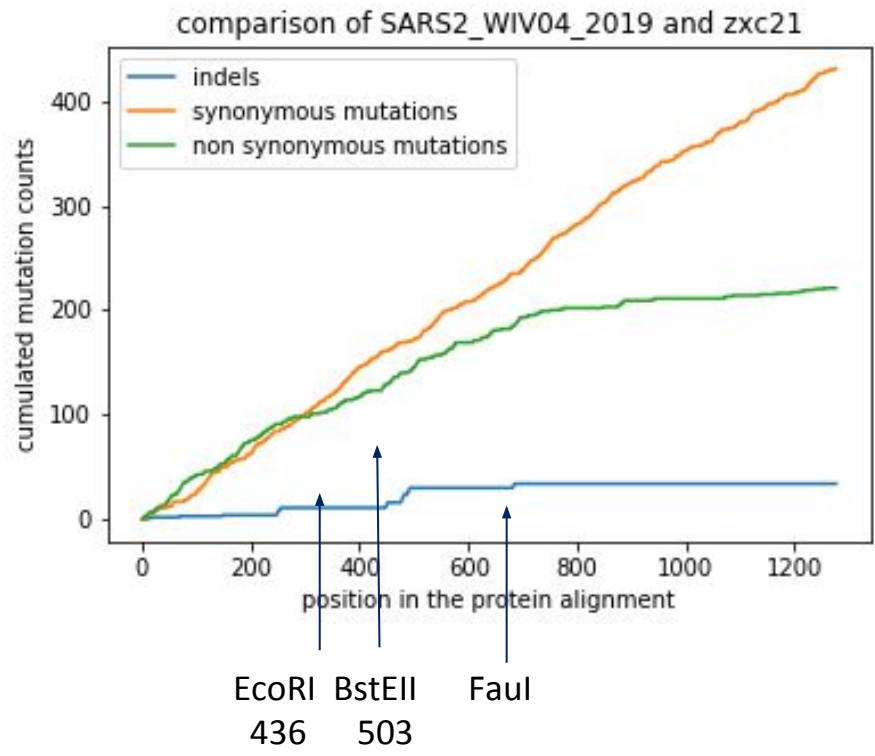
Profils de mutations synonymes sur la protéine S



Profils de mutations synonymes sur la protéine S

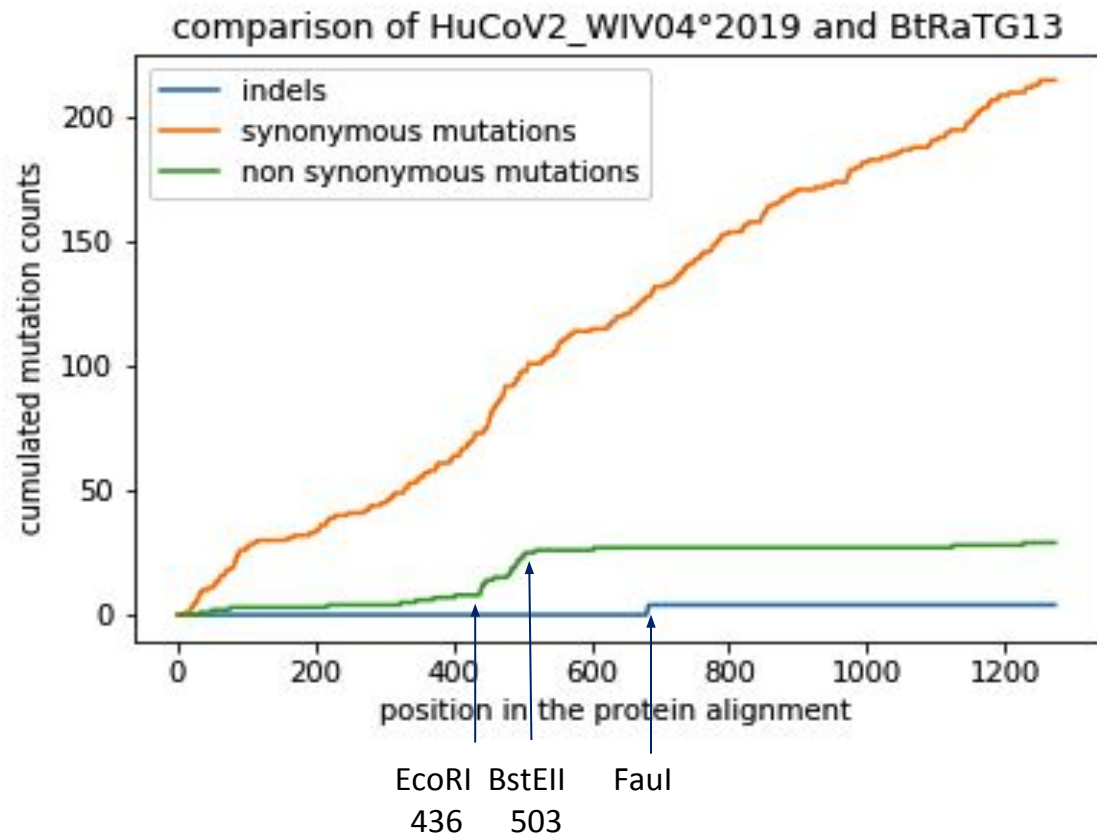


Profils de mutations synonymes sur la protéine S

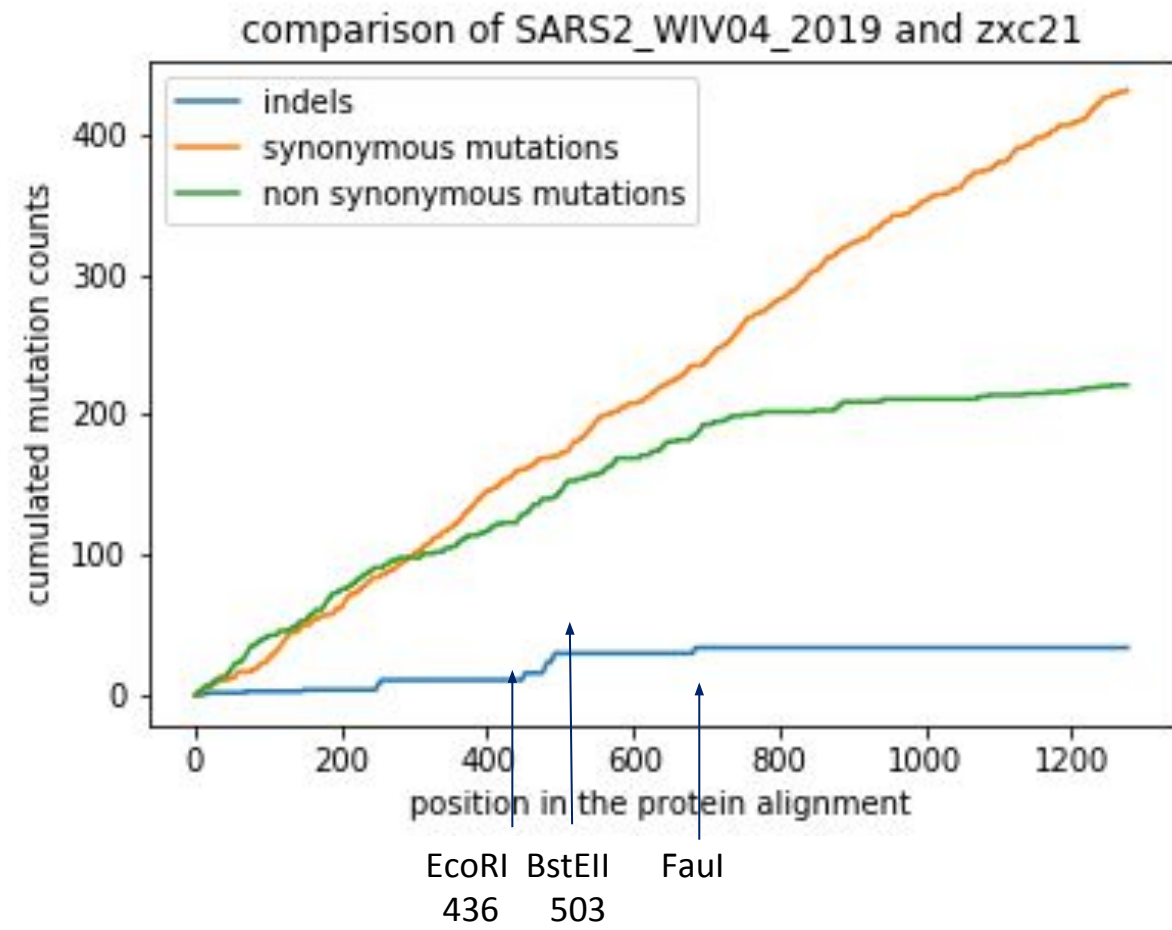


Profils PPI génomiques versus protéiques

Mutations synonymes



Mutations synonymes



Mutations synonymes

