

Mutations et variants

Jacques van Helden

Mutations et recombinaisons

Typologie des mutations

- Substitution

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

Avant réPLICATION

ATGACCATGA



Après réPLICATION

ATGACCAGGA



- Délétion

- Perte d'un fragment de la molécule

ATGACCATGA



ATGGA

- Insertion

- Ajout d'un fragment de molécule

ATGACCATGA

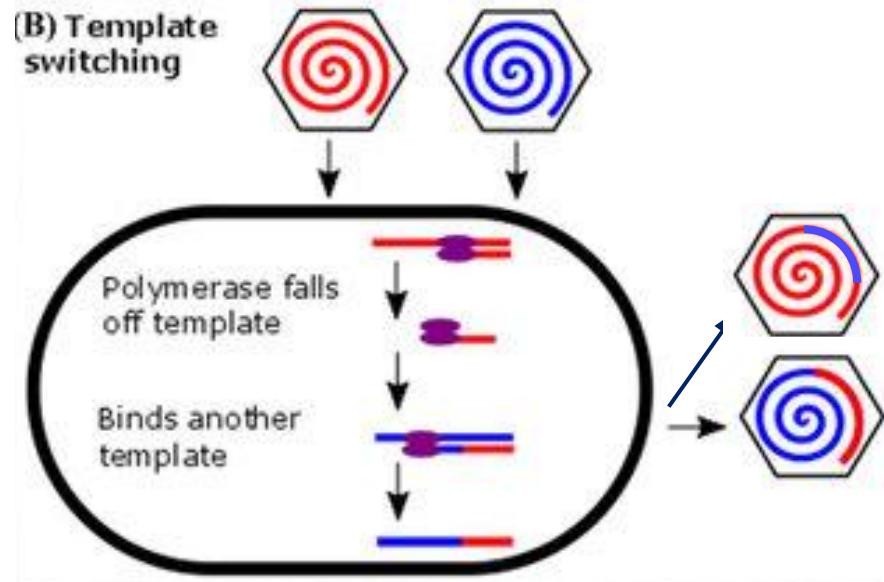


ATGACAAACATGA



Recombinations 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.



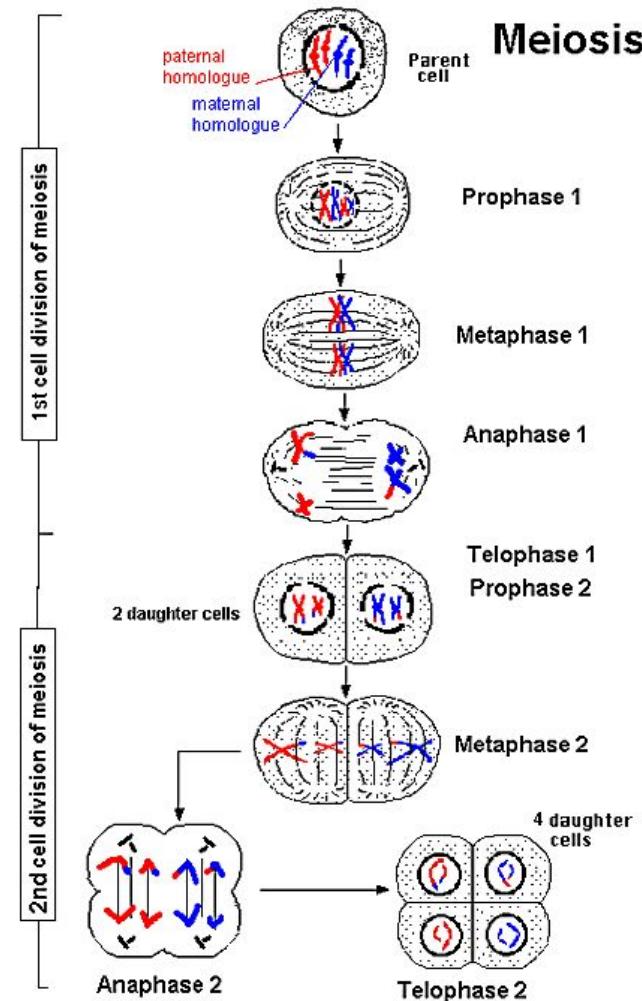
Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus

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

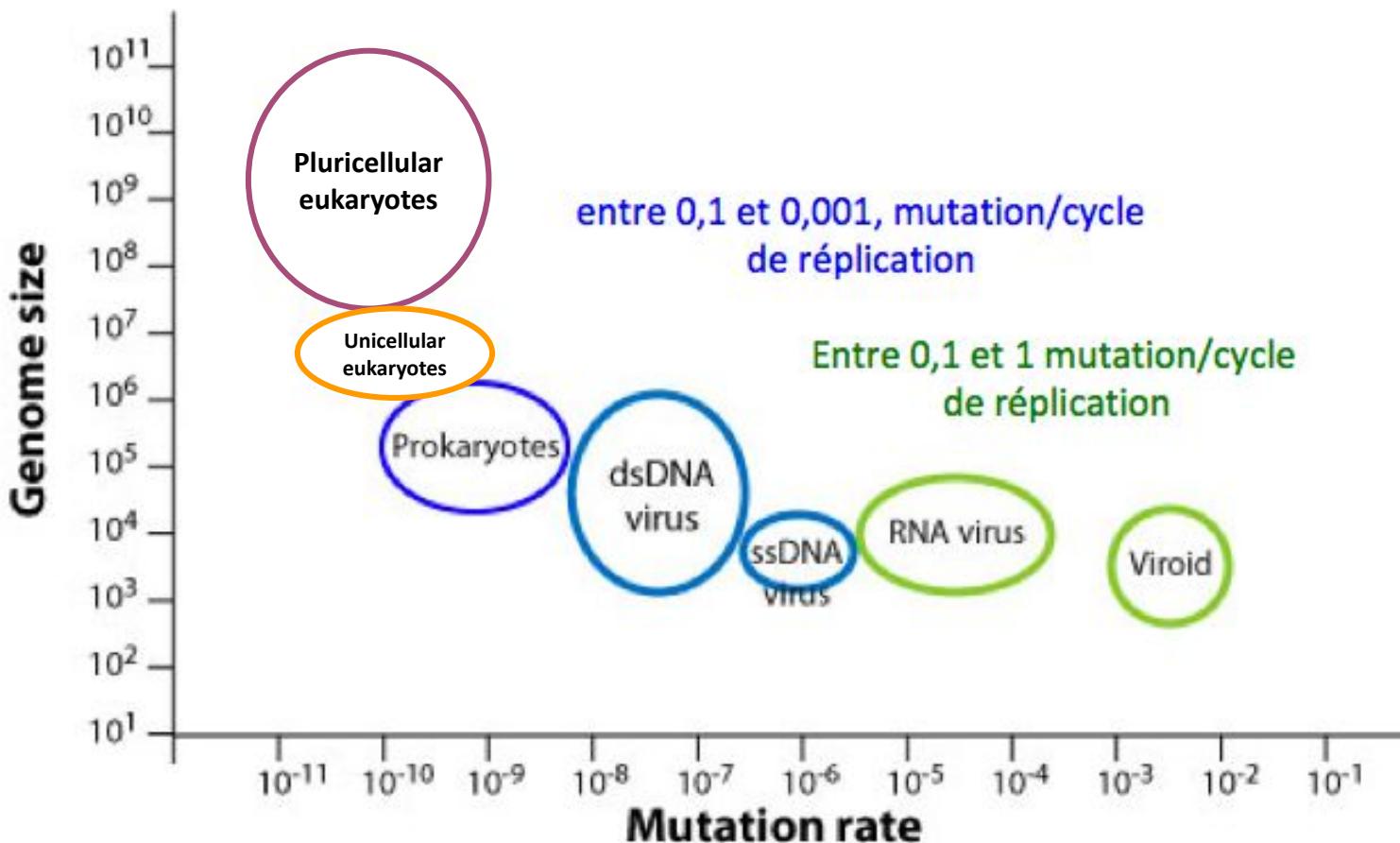
J. Dennehy, Evolutionary ecology of virus emergence: Virus emergence, 2016,
Annals of the New York Academy of Science

Recombinaison chez les eucaryotes

- Chez les organismes cellulaires eucaryotes (cellules à noyau), 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.



Taux de substitutions représentatifs chez différents groupes taxonomiques

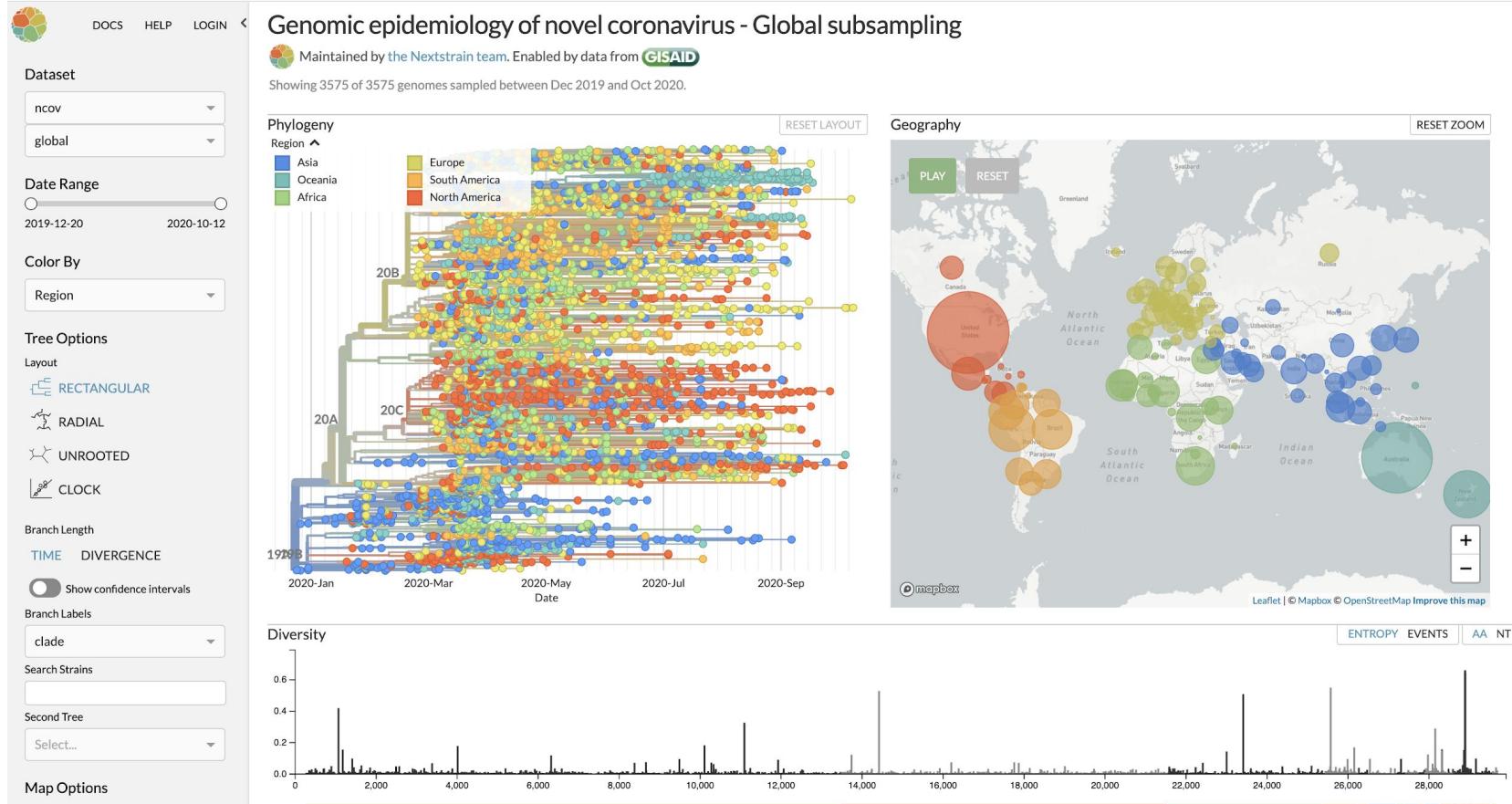


Mutations et variations dans les génomes de SARS-CoV-2

La collecte des génomes de SARS-CoV-2

- Les technologies de séquençage massivement parallèle (“Next-Generation Sequencing) sont utilisées pour séquencer le génome complet de virus prélevés sur des patients Covid-19.
- Ces séquences sont déposées dans la base de données GISAID
 - 144 127 génomes complets à ce jour (14 octobre 2020)
 - Note : données accessibles gratuitement pour les chercheurs *mais* on n'a pas le droit de les diffuser.
- Les bioinformaticiens ont développé des outils pour extraire des informations pertinentes à partir de ces génomes.

The screenshot shows the GISAID website homepage. At the top, there's a navigation bar with 'Login' and a search icon. The main header 'GISAID' is in a large, bold, white font on a dark green background. Below it is a sidebar with links: 'About us', 'Database Features', 'Events', 'Collaborations', 'References', 'Registration', and 'Help', each preceded by a '+' sign. To the right of the sidebar is a section titled 'In Focus' with the text: 'Over 143,000 viral genomic sequences of hCoV-19 shared with unprecedented speed via GISAID'. It includes a small chart showing the cumulative number of submissions from January to October 2020, reaching 143k. Below this is a paragraph about the importance of real-time data sharing for understanding the disease and developing countermeasures. Further down, there are sections for 'Enabled by hCoV-19 data from GISAID' featuring the 'Geographic Mutation Tracker' (KAUST, Saudi Arabia) and the 'Interactive Real-Time Mutation Tracker' (Michigan State University, USA). At the bottom, there's a section titled 'Recent hCoV-19 data submissions' listing several entries: hCoV-19/Japan/NGY-NNH-021/2020, hCoV-19/Peru/LIM-UPCH-0127/2020, hCoV-19/Gambia/NPHL-2892/2020, hCoV-19/Australia/VIC13964/2020, and hCoV-19/Bangladesh/JUST-GC46-003a/2020. A footer at the very bottom says 'Number of entries: 144,127'.



Propagation d'une souche mutante

- Analyse de l'ensemble des génomes SARS-CoV-2 dans GISAID
- Variation de séquence à la position 614 de la protéine spicule, avec selon les cas
 - D acide aspartique
 - G glycine
- Souche originale: D614 (en orange)
- La souche "G614" (en bleu sur le graphe)
 - apparaît en janvier.
 - se propage plus rapidement que la D614.
 - remplace progressivement la souche originale (D614).

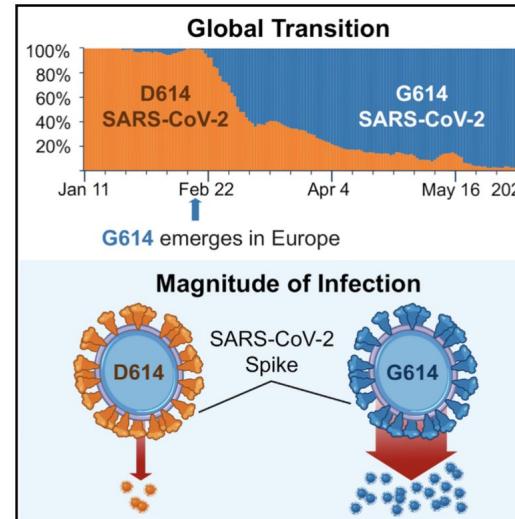
Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* S0092867420308205.

Grubaugh, N.D., Hanage, W.P. & Rasmussen, A.L. 2020. Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. *Cell* S0092867420308175.

Cell

Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus

Graphical Abstract



Authors

Bette Korber, Will M. Fischer, Sandrasegaram Gnanakaran, ..., Celia C. LaBranche, Erica O. Saphire, David C. Montefiori

Correspondence

btk@lanl.gov

In Brief

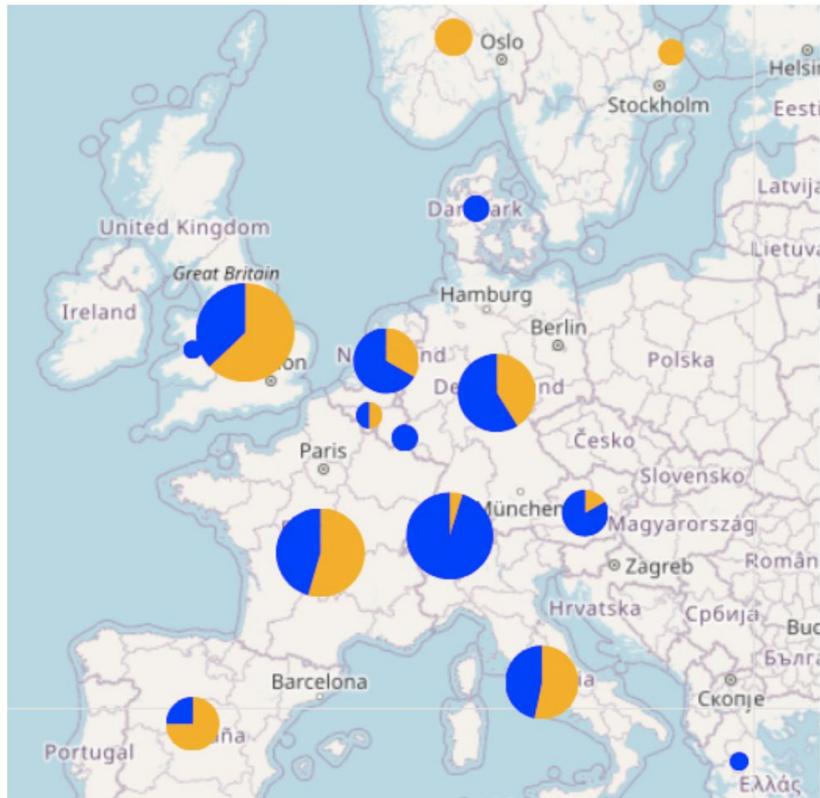
Korber et al. present evidence that there are now more SARS-CoV-2 viruses circulating in the human population globally that have the G614 form of the Spike protein versus the D614 form that was originally identified from the first human cases in Wuhan, China. Follow-up studies show that patients infected with G614 shed more viral nucleic acid compared with those with D614, and G614-bearing viruses show significantly higher infectious titers *in vitro* than their D614 counterparts.

Highlights

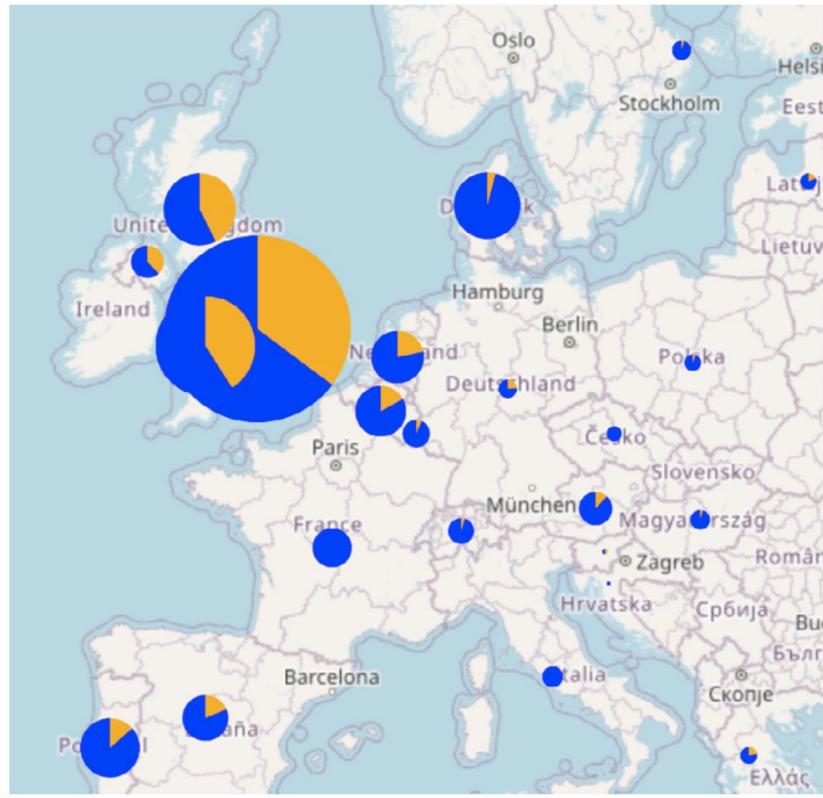
- A SARS-CoV-2 variant with Spike G614 has replaced D614 as the dominant pandemic form
- The consistent increase of G614 at regional levels may indicate a fitness advantage
- G614 is associated with lower RT PCR Cts, suggestive of higher viral loads in patients
- The G614 variant grows to higher titers as pseudotyped virions

Fréquences des variants D614 et G614 dans les populations européennes

Avant le 1er mars 2020

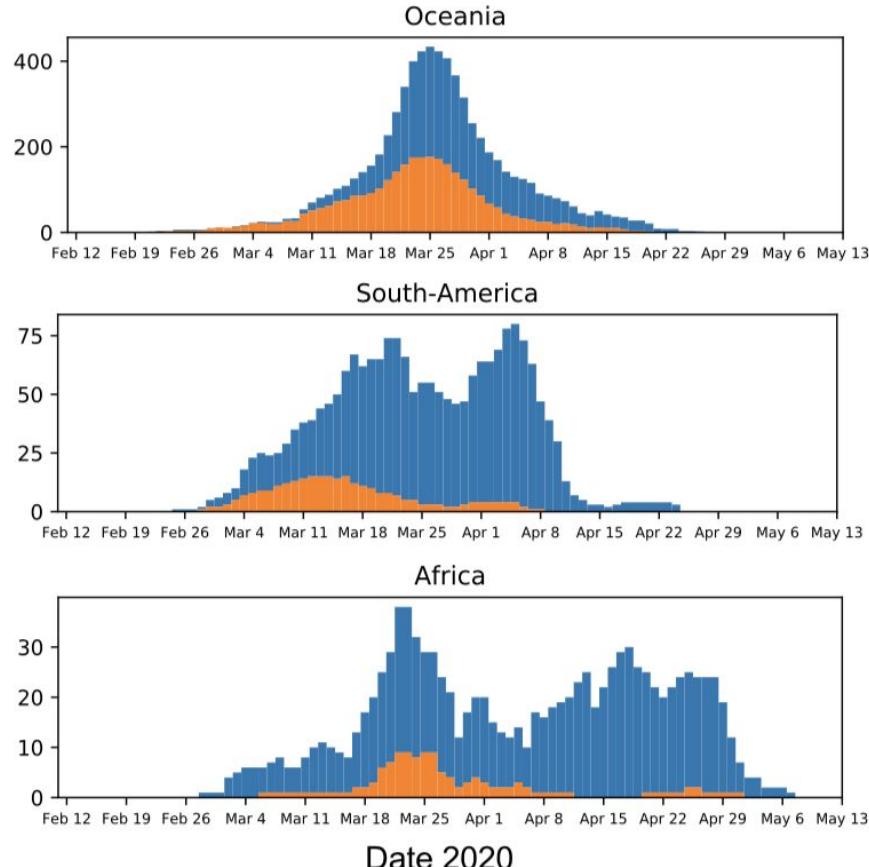
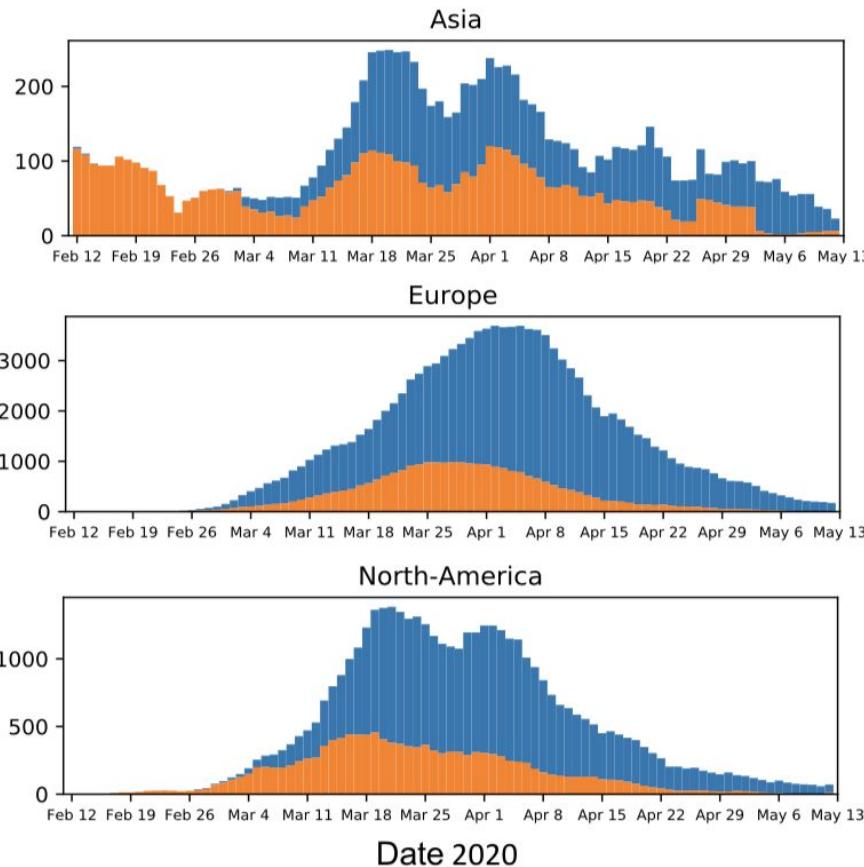


Du 21 au 30 mars

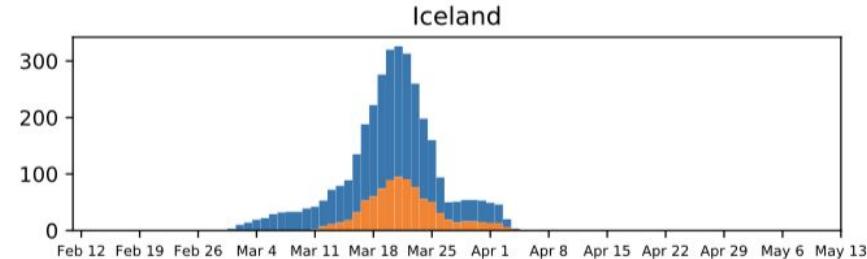
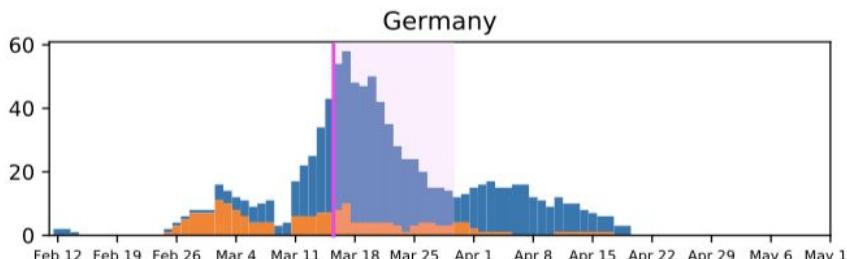
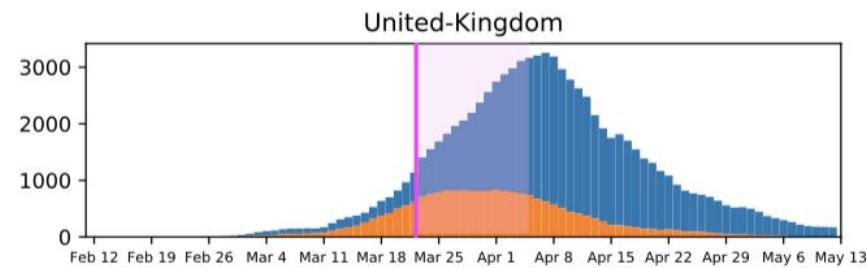
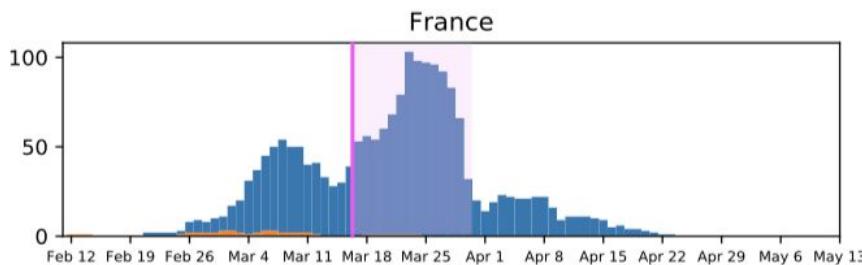
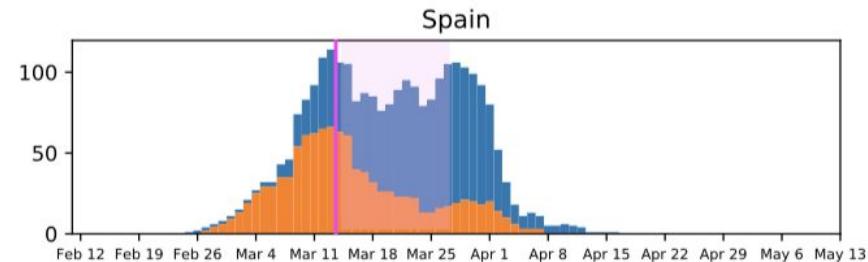
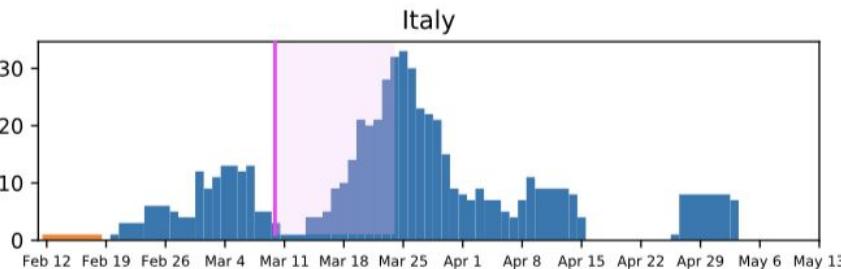


Evolution temporelle par continent

Weekly running counts



Evolution temporelle par pays



Une souche plus envahissante mais pas plus sévère

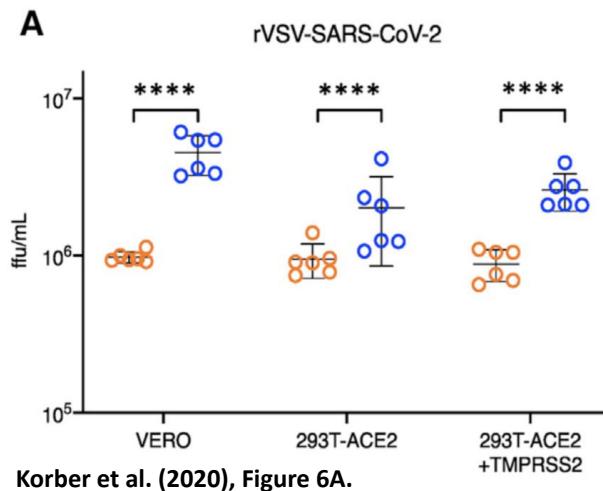
Plus envahissante ...

- Figure 6A: Sur cultures cellulaires, la charge virale est plus importante pour la souche **G614** que pour la **D614**. La mutation D614G pourrait faciliter l'entrée du virus dans la cellule.

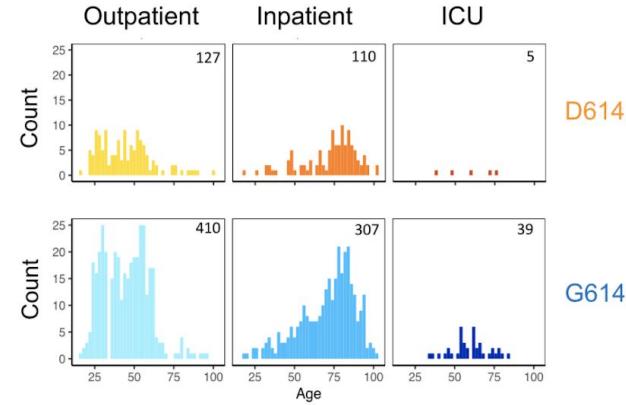
... mais pas plus sévère

- Figure 5B: les nombres d'hospitalisations par tranches d'âge ne montrent pas de différences entre souches **G614** et **D614**. La mutation n'aggrave donc pas les symptômes cliniques.

Note: une étude détaillée des patients anglais (novembre 2020) confirme ces observations: la souche **G614** est associée à une charge virale plus élevée, mais n'augmente pas la sévérité clinique ou la mortalité.



B Hospitalization **D614** vs **G614**
Fisher' p = 0.66



	OP	IP	ICU
D614	127	110	5
G614	410	307	39

Fisher's exact, 2x2: (OP+IP) x ICU = 0.047

Fisher's exact, 2x2, OP x (IP+ICU): p = 0.66

Korber et al. (2020), Figure 5B.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalsterer, W., et al. 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell S0092867420308205.

Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., et al. 2020. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell, doi: 10.1016/j.cell.2020.11.020.

Matrices de substitution

Matrices de substitutions

- Une **matrice de substitution** associe un score à chaque paire de résidus qu'on peut trouver dans un alignement.
 - Chaque ligne et chaque colonne représente l'un des résidus (4 nucléotides, 20 acide aminés).
 - La diagonale correspond aux identités.
 - Le triangle inférieur correspond à des substitutions.
 - Le triangle supérieur est symétrique au triangle inférieur, il n'est pas nécessaire d'indiquer les nombres.
 - Les **scores négatifs** sont considérés comme des pénalités associées à certaines substitutions qu'on n'observe que rarement dans les alignements. Les algorithmes d'alignements tenteront donc d'éviter ces substitutions.
 - Les **scores positifs** correspondent à des substitutions qu'on observe plus souvent que prévu, dans les alignements d'un grand nombre de séquences. Ceci suggère que ces substitutions particulières sont moins dommageable que d'autres, et on les qualifie donc de « **substitutions conservatives** » ou encore de « **mutations ponctuelles acceptées** » (**PAM**).
 - Au sein d'un alignement, le terme **similarité** désigne les positions où se superposent des résidus ayant un score positif dans la matrice de substitution (identité ou substitution conservative).

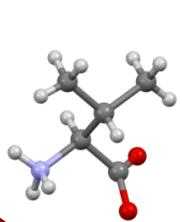
Matrice de substitutions entre nucléotides

	A	C	G	T
A	2			
C	-2	2		
G	-2	-2	2	
T	-1	-2	-2	2

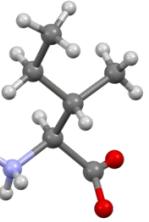
Matrice de substitutions entre acides aminés

Similarités chimiques entre acides aminés

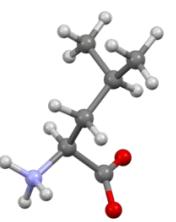
Valine



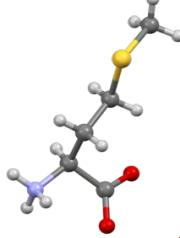
Isoleucine



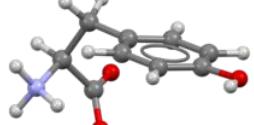
Leucine



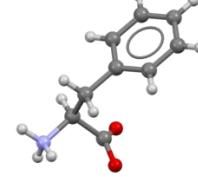
Méthionine



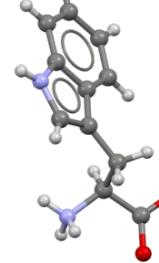
Tyrosine



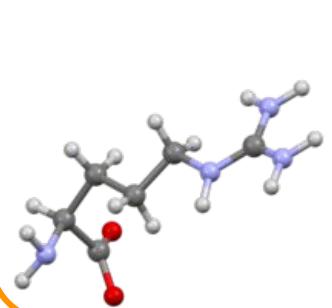
Phénylalanine



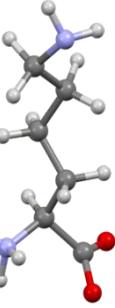
Tryptophane



Lysine



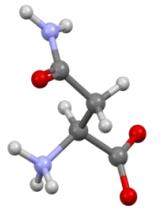
Arginine



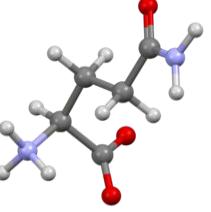
Matrice de substitutions entre acides aminés

	Ala	A	R	Z	Asn	D	Asp	C	Cys	Q	Gln	E	Glu	G	Gly	H	His	Ile	Lys	M	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																										
Arg	-1	5																									
Asn	-2	0	6																								
Asp	-2	-2	1	6																							
Cys	0	-3	-3	-3	9																						
Gln	-1	1	0	0	-3	5																					
Glu	-1	0	0	2	-4	2	5																				
Gly	0	-2	0	-1	-3	-2	-2	6																			
His	-2	0	1	-1	-3	0	0	-2	8																		
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	-4	-3	-4	-3	-4	-3	4											
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	-4	-3	-4	-3	-4	-3	2	4										
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	-1	-3	-2	-5												
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	-3	-2	5												
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6													
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7												
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4											
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-1	-1	5											
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-2	-3	-1	11											
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	-3	-2	7											
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-2	-2	0	-3	-1	4								
	Ala	A	Arg	Z	Asn	D	Asp	C	Cys	Q	Gln	E	Glu	G	Gly	H	His	Ile	Lys	M	Phe	Pro	Ser	Thr	Trp	Tyr	Val

Asparagine



Glutamine



Des mutations aux “variants”

Variants de SARS-CoV-2

En épidémiologie, un “variant” est une souche qui présente une combinaison particulière de mutations.

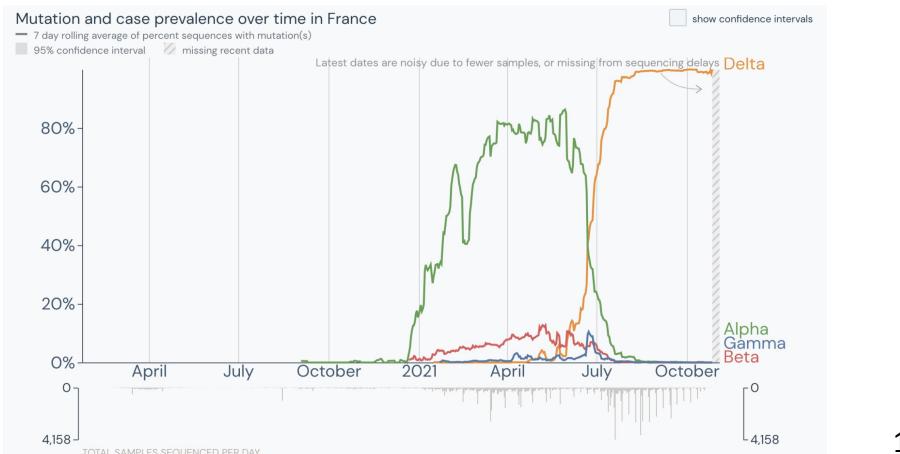
Dès le mois de mars 2020, le Royaume-Uni met en place un ambitieux programme de séquençage des génomes de SARS-CoV-2 pour

- surveiller l'évolution du virus
- détecter l'apparition de nouveaux variants
- étudier l'impact de ces variants sur
 - la transmissibilité
 - la pathogénicité
 - l'échappement immunitaire
- préparer des stratégies de lutte adaptées à l'évolution des variants (vaccins, médicaments)

- Variants OMS: <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>
- Evolution variants en France:
<https://outbreak.info/location-reports?loc=FRA&selected=Delta&selected=Alpha&selected=Beta&selected=Gamma>

Currently designated Variants of Concern⁺:

WHO label	Pango lineage*	GISAID clade	Nextstrain clade	Additional amino acid changes monitored ^a	Earliest documented samples	Date of designation
Alpha	B.1.1.7 #	GRY	20I (V1)	+S:484K +S:452R	United Kingdom, Sep-2020	18-Dec-2020
Beta	B.1.351	GH/501Y.V2	20H (V2)	+S:L18F	South Africa, May-2020	18-Dec-2020
Gamma	P.1	GR/501Y.V3	20J (V3)	+S:681H	Brazil, Nov-2020	11-Jan-2021
Delta	B.1.617.2§	G/478K.V1	21A, 21I, 21J	+S:417N	India, Oct-2020	VOI: 4-Apr-2021 VOC: 11-May-2021

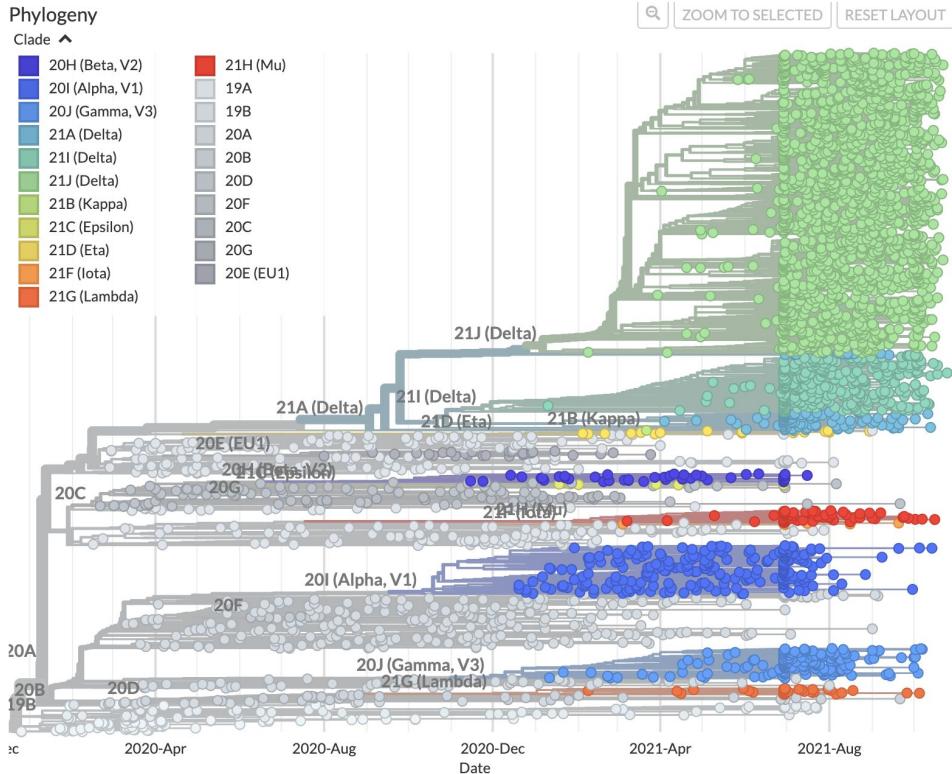


Cartographie mondiale des variants

- nextstrain.org/ncov/gisaid/global

L'outil Nextstrain permet d'afficher

- La pylogénie des variants, établie sur base de milliers de génomes complètement séquencés

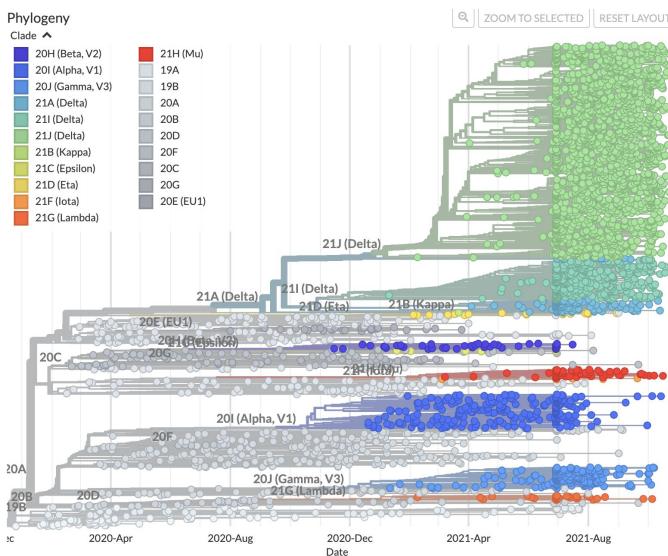


Cartographie mondiale des variants

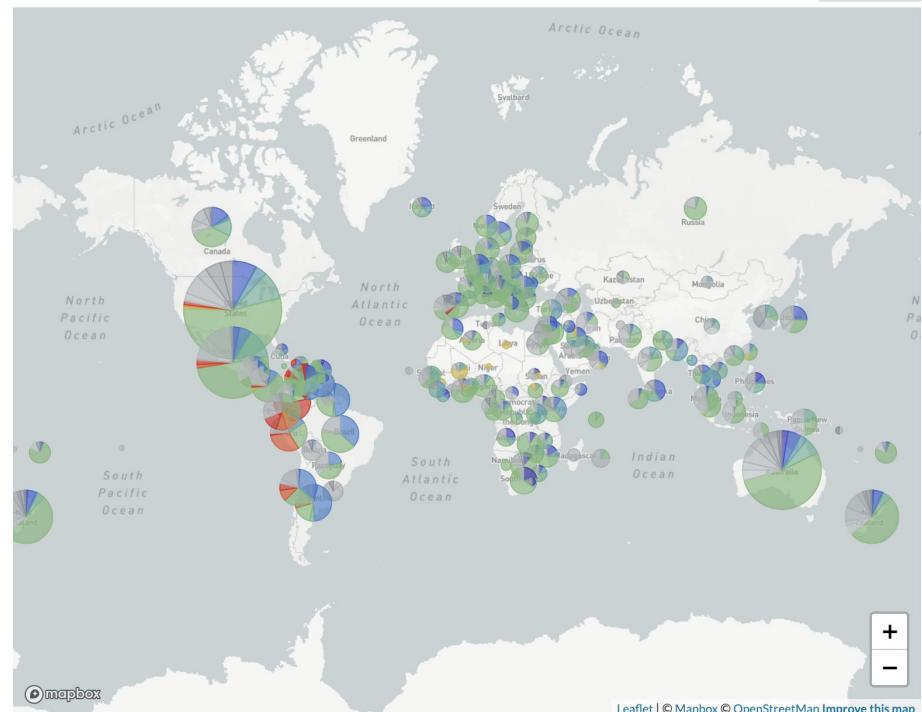
- nextstrain.org/ncov/gisaid/global

L'outil Nextstrain permet d'afficher

- La phylogénie des variants, établie sur base de milliers de génomes complètement séquencés
- la distribution géographique des variants



Geography



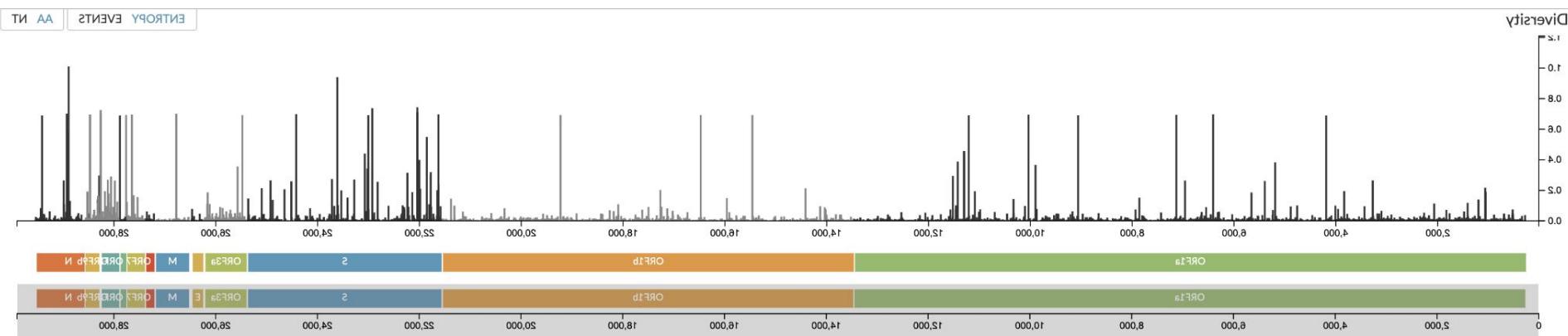
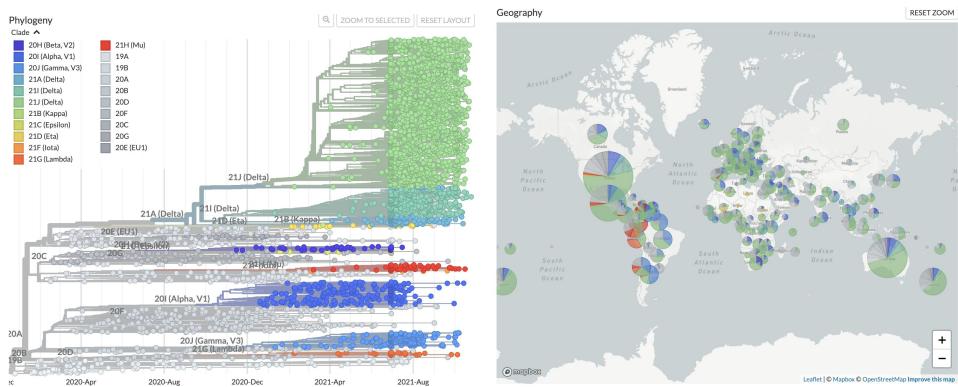
Leaflet | © Mapbox © OpenStreetMap Improve this map

Cartographie mondiale des variants

▪ nextstrain.org/ncov/gisaid/global

L'outil Nextstrain permet d'afficher

- La phylogénie des variants, établie sur base de milliers de génomes complètement séquencés
- la distribution géographique des variants
- le nombre de mutations à chaque position du génome



Cartographie mondiale des variants

- nextstrain.org/ncov/gisaid/global

L'outil Nextstrain permet d'afficher

- La phylogénie des variants, établie sur base de milliers de génomes complètement séquencés
- la distribution géographique des variants
- le nombre de mutations à chaque position du génome
- l'évolution temporelle des fréquences de variants

