

NGS DES VIRUS RESPIRATOIRES ET SURVEILLANCE ÉPIDÉMIOLOGIQUE

SÉMINAIRE BIOINFODIAG2025



HCL

HOSPICES CIVILS
DE LYON

01/04/2025

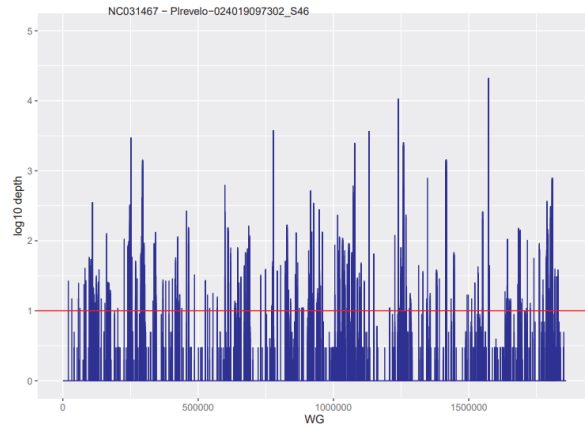
HADRIEN REGUE – CENTRE NATIONAL DE RÉFÉRENCE DES VIRUS RESPIRATOIRES

www.chu-lyon.fr

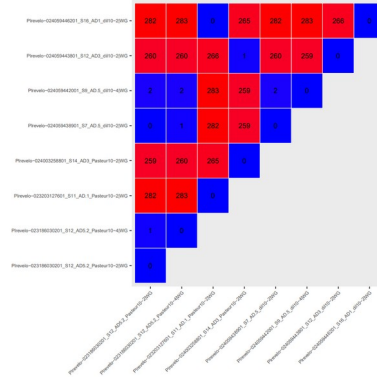
INTRODUCTION: DÉCOUVRIR LE NGS EN VIROLOGIE POUR L'ÉPIDÉMIOLOGIE

POURQUOI FAIRE DU NGS EN VIROLOGIE?

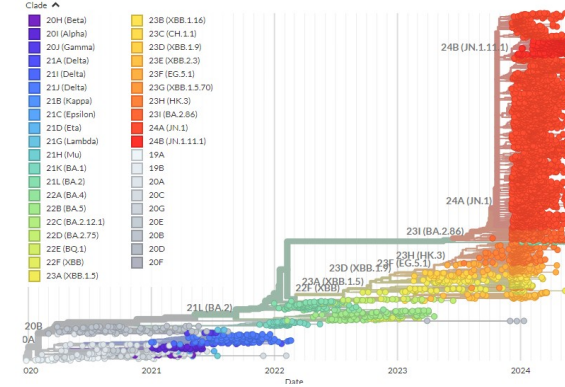
Détection de pathogènes



Analyses de Clusters

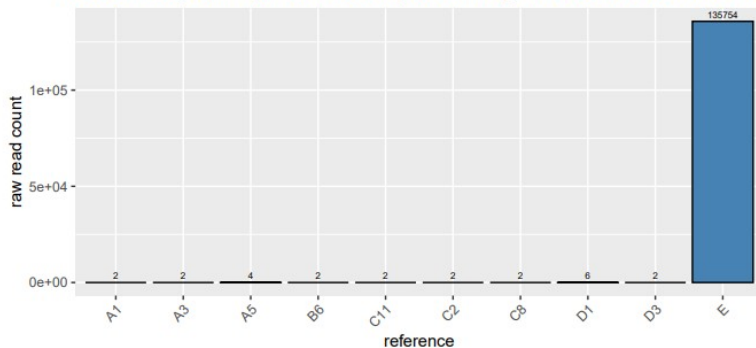


Epidémiologie



Sous-typage

Reference repartition per read for sample:barcode_PI07VHB-024064844501-81



Recherche de variants minoritaires

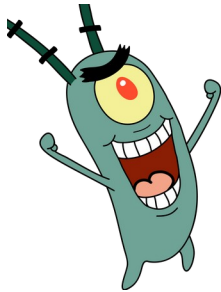
REFERENCE	REF_POS	REF	total_count	base_status	base	count	NUM_CODON	REF_AA	ALT_AA	Mutation	freq
GTA	2032	T	22324	majo	T	19222	180	L	L	L180L	86,1
GTA	2032	T	22324	mino_1st	A	2961	180	L	M	L180M	13,26
GTA	2032	T	22324	mino_2nd	C	107	180	L	L	L180L	0,48
GTA	2032	T	22324	mino_3d	G	34	180	L	V	L180V	0,15
GTA	2033	T	23596	majo	T	23455	180	L	L	L180L	99,4
GTA	2033	T	23596	mino_1st	C	74	180	L	S	L180S	0,31
GTA	2033	T	23596	mino_2nd	G	52	180	L	W	L180W	0,22
GTA	2033	T	23596	mino_3d	A	15	180	L	*	L180*	0,06
GTA	2034	G	23503	majo	G	23380	180	L	L	L180L	99,48
GTA	2034	G	23503	mino_1st	C	52	180	L	F	L180F	0,22
GTA	2034	G	23503	mino_2nd	A	51	180	L	L	L180L	0,22
GTA	2034	G	23503	mino_3d	T	20	180	L	F	L180F	0,09

CARACTÉRISTIQUES DES VIRUS



X1000

Virus:
2k-2.5Mpb



X1000

Bactérie:
0.2-13Mpb



Humain:
2.3Gpb

Taille

ADN simple brin
ADN double brin
ARN simple brin
ARN double brin
ARN simple brin circulaire
ARN double brin circulaire
Génome segmenté
...



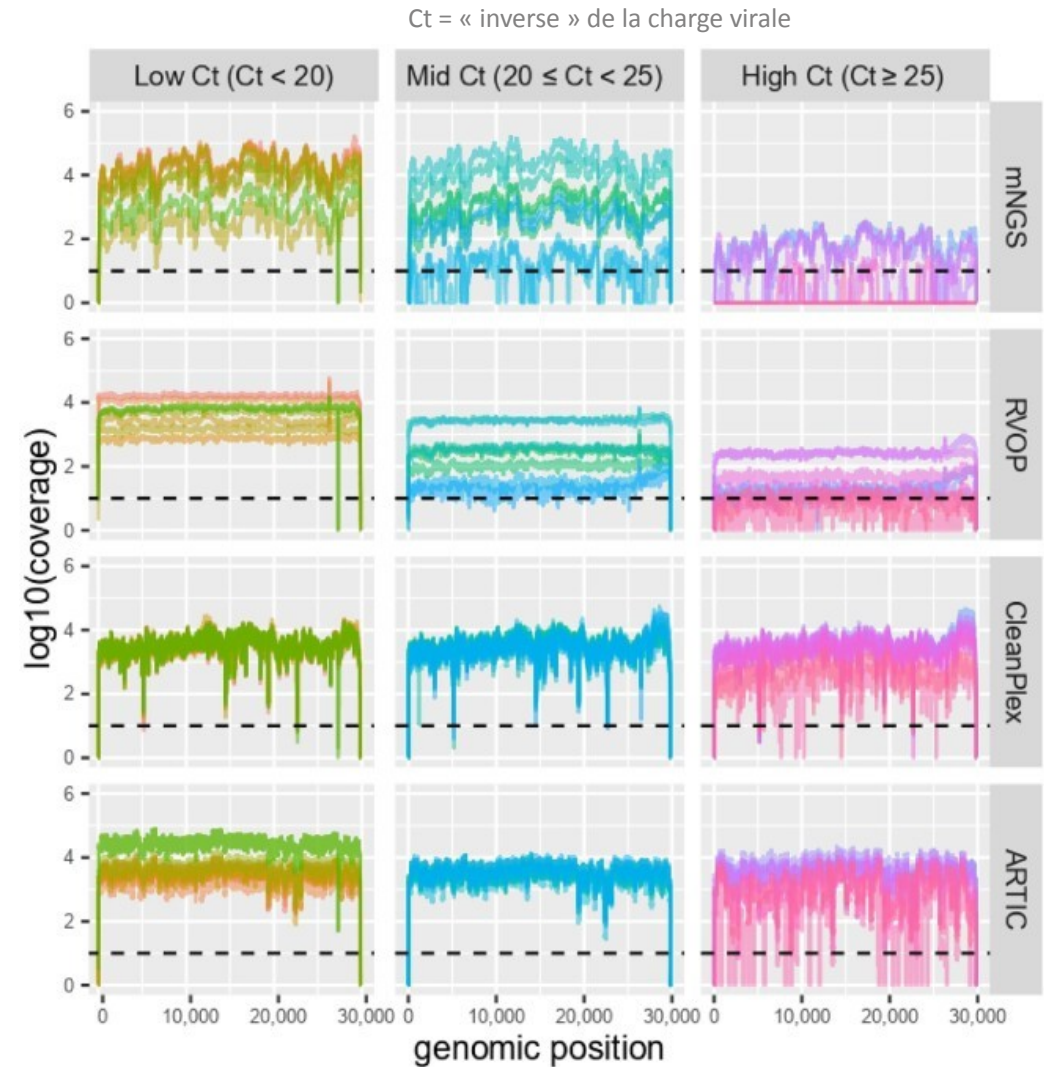
Réplication et évolution



Structures génomiques

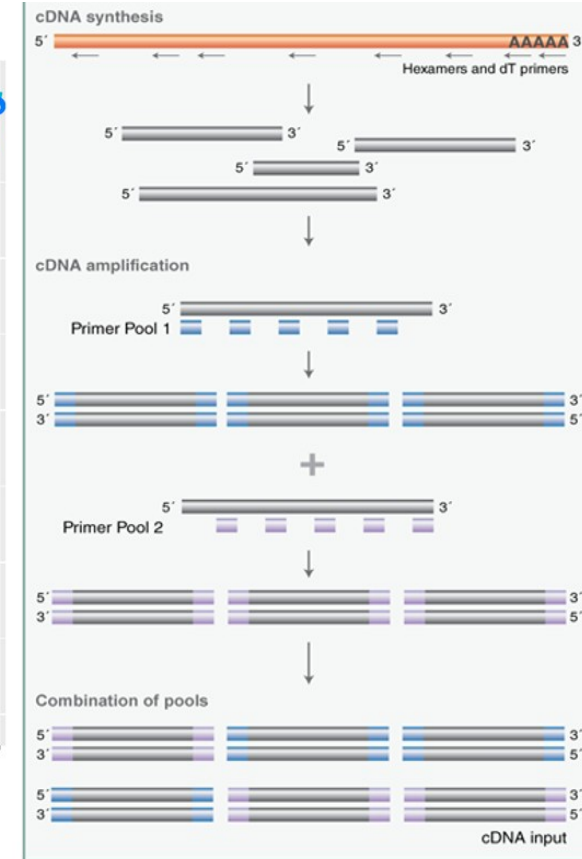
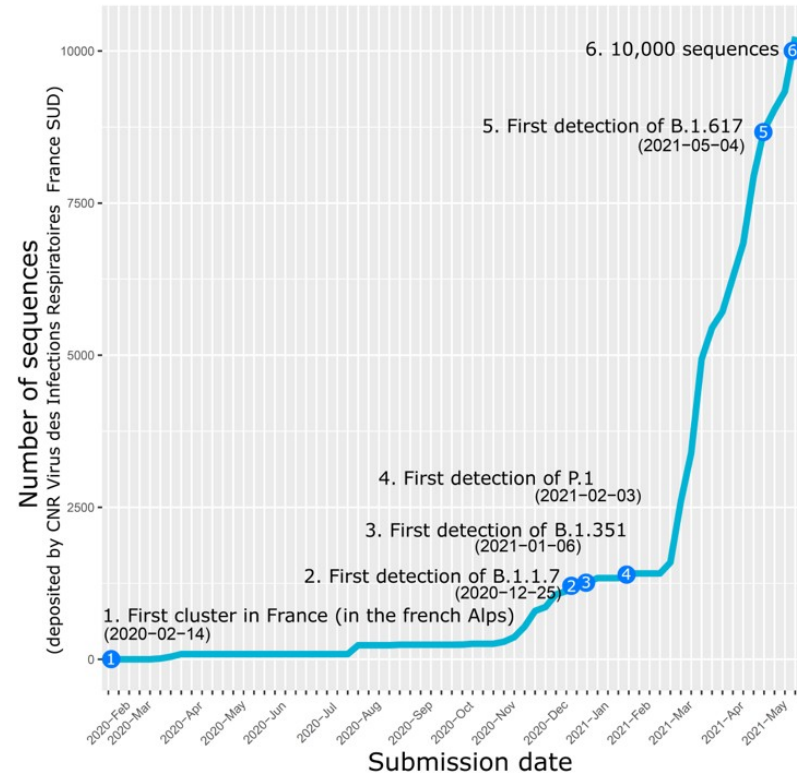
APPROCHES DE SÉQUENÇAGE : METAGENOMIQUE

- Méthode de référence (mNGS)
- Détection de pathogènes émergents
- Non ciblée et non biaisée
- Cout/échantillon élevé
- Manips longues
- sensible à la charge virale



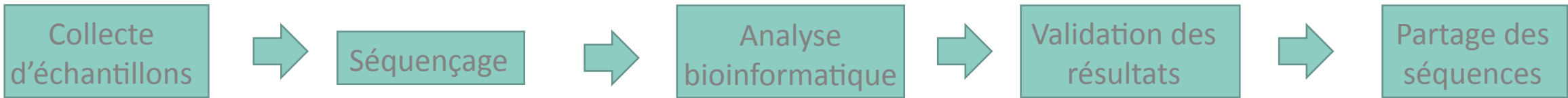
APPROCHES DE SÉQUENÇAGE : AMPLICON

- Méthode de routine
- Séquençage ciblé
- Cout/échantillon faible
- Profondeurs de séquençage élevées
- Fortement impacté par l'évolution du pathogène



LE PROCESSUS DE SÉQUENÇAGE EN ÉPIDÉMIOLOGIE

CAS DU CNR VIRUS RÉSPI' LYON – FRANCE SUD POUR LE SARS-COV-2



Récupération des échantillons provenant des régions France sud

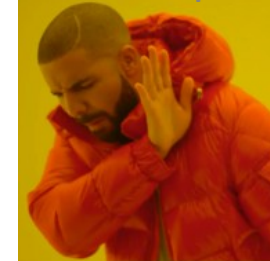


Séquençage Illumina en amplicon



Génération d'une séquence consensus

Mauvaise séquence



Bonne séquence



Validation et interprétation de la séquence

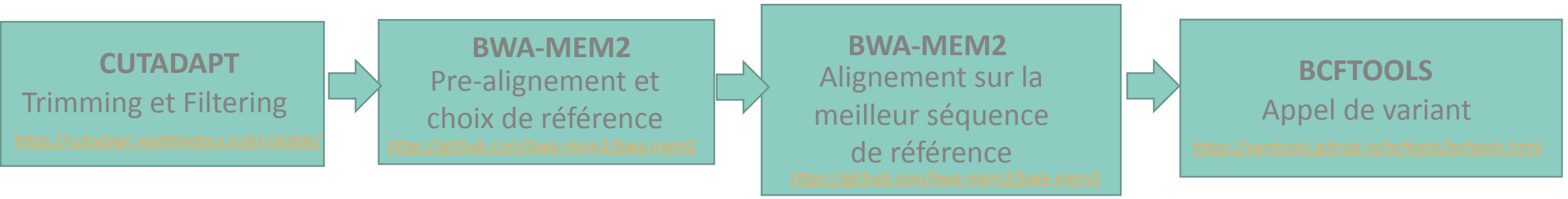


Soumission des séquences sur les bases de données nationales et internationales

L'ANALYSE BIOINFORMATIQUE

DES READS BRUTS À LA CONSENSUS

PRODUIRE UN GÉNOME CONSENSUS, C'EST SIMPLE...



-rognage des reads
-suppression des
reads trop courts

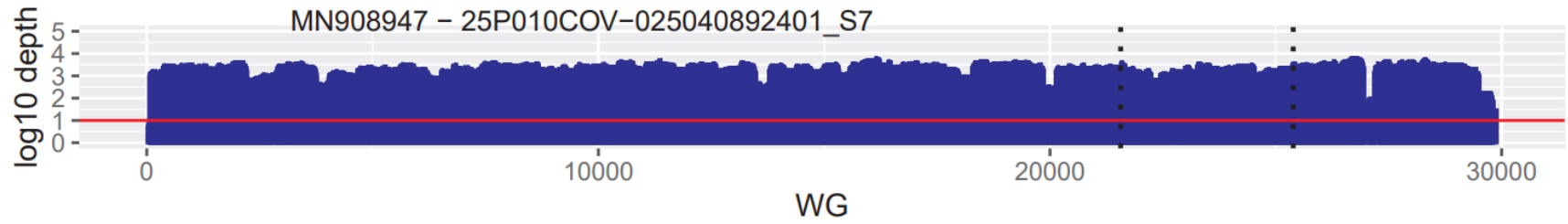
Alignement rapide sur
des fractions de
génomme et sur des
contrôles

Alignement sur le
génomme complet
sélectionné

Recherche des variants
majoritaires avec une
fréquence >50%

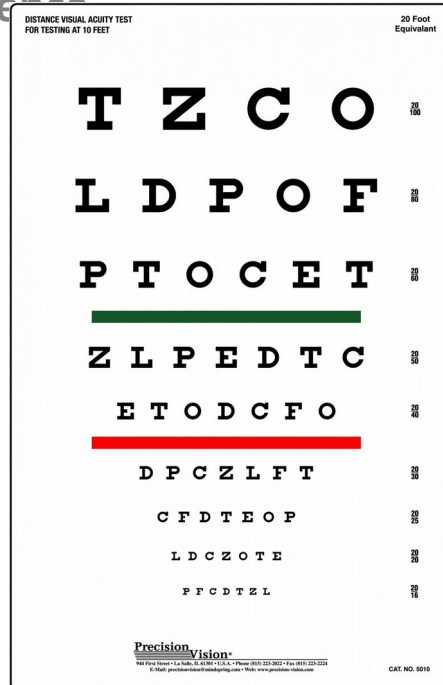
SCRIPT MAISON
Construction de
la consensus

Utilisation de la référence initiale
avec prise en compte la profondeur
de séquençage et les variants
retenus



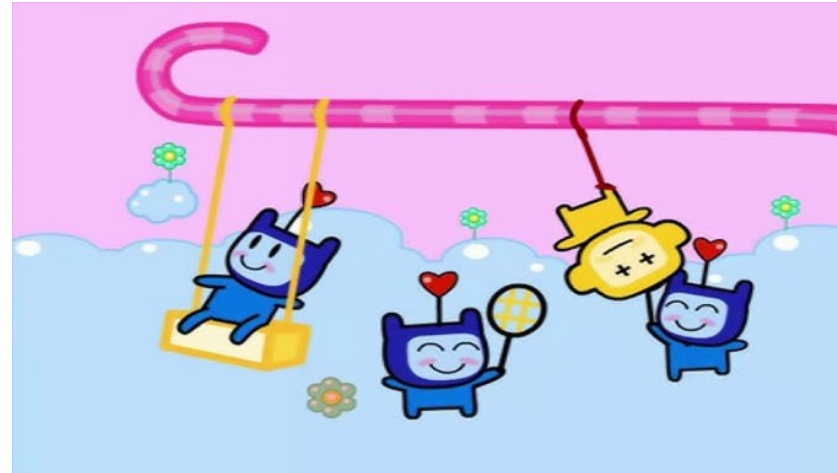
...MAIS ATTENTION AUX PIÈGES!

Choix de la séquence de référence



Utiliser une référence éloignée peut entraîner:
-erreurs d'alignements
-erreurs de variant calling
-le masquage d'évènements génomiques

Populations et coinfections



Plusieurs populations d'un même virus peuvent coexister au sein d'un même hôte:
-des populations de lignages différents
-des population d'un même lignage

Réassortants et recombinants

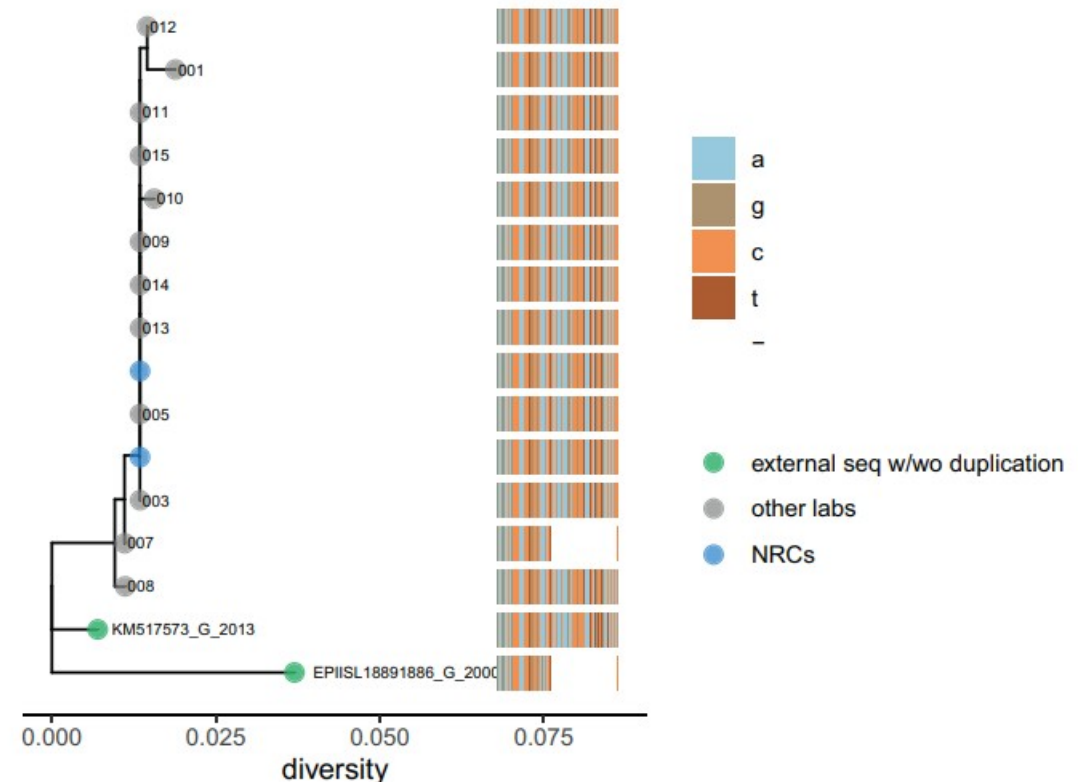


Dans des cas de coinfections et doubles populations:
-les génomes peuvent se recombiner (SARS-CoV-2)
-il peut y avoir des échanges de segments pour les génomes segmentés (Influenza)

L'IMPORTANCE DU CHOIX DE LA RÉFÉRENCE

DIRECTION PAR LES CNR VIRUS RÉSPI
D'UN EQA CONCERNANT LE
SÉQUENÇAGE DU VRS.
APPARITION D'UNE DUPLICATION
DANS UNE RÉGION DU GÉNOME DES
VRS DE TYPE B EN 2013.
SI UNE RÉFÉRENCE ANTÉRIEURE EST
UTILISÉE POUR L'ALIGNEMENT, IL Y A
DES RISQUES DE RATER CETTE
DUPLICATION DANS LA CONSENSUS
FINALE.

Moisan, CMI, 2024



<https://doi.org/10.1016/j.cmi.2024.12.018>

VARIANT CALLING, DOUBLE POP', COINFECTION

UTILISATION DE VARIANT
CALLER À MODÈLE
SOMATIQUE (==ANALOGUE
AUX TUMEURS?)
AF =]0;1]
NOMBRE DE VARIANTS
MINORITAIRES ÉLEVÉES?
PRÉSENCES DE VARIANTS
SPÉCIFIQUES À DEUX
LIGNAGES DIFFÉRENTS?
CONTAMINATION?
BIAS DE PRIMERS?

#CHROM	POS	ID	REF	ALT	QUAL	FILTER	INFO	FORMAT	22P1CoInf-V41xEMAG50x50_S306.bam
MN908947	WG	21618	.	C	G	9.03101	PASS	TYPE=snp;AF=0.731634;DP=4669;ADF=653,1599;ADR=599,1817;	
MN908947	WG	21762	.	C	T	1.34905	PASS	TYPE=snp;AF=0.281066;DP=2889;ADF=928,355;ADR=1147,457;A	
MN908947	WG	21764	.	ATACATG	A	2.60846	PASS	TYPE=del;AF=0.2771;IDV=795;IMF=0.2771;DP=2869;ADF=9	
MN908947	WG	21846	.	C	T	258.013	PASS	TYPE=snp;AF=0.999654;DP=2894;ADF=0,889;ADR=0,2004;AD=0,	
MN908947	WG	21986	.	GGTGTATT	G	3.21633	PASS	TYPE=del;AF=0.322026;IDV=734;IMF=0.323348;DP=22	
MN908947	WG	21987	.	G	A	261.022	PASS	TYPE=snp;AF=0.997883;DP=1417;ADF=1,903;ADR=1,511;AD=2,1	
MN908947	WG	22028	.	GAGTTCA	G	15.647	PASS	TYPE=del;AF=0.646465;IDV=1344;IMF=0.646465;DP=2079;	
MN908947	WG	22193	.	AATT	A	3.47603	PASS	TYPE=del;AF=0.282156;IDV=623;IMF=0.282156;DP=2208;A	
MN908947	WG	22204	.	T	TGAGCCAGAA	3.78821	PASS	TYPE=ins;AF=0.260625;IDV=435;IMF=0.271875;DP=16	
MN908947	WG	22578	.	G	A	0.579927	PASS	TYPE=snp;AF=0.357772;DP=2837;ADF=904,535;ADR=916,48	
MN908947	WG	22673	.	T	C	1.2494	PASS	TYPE=snp;AF=0.34342;DP=1421;ADF=385,189;ADR=548,299;AD=	
MN908947	WG	22674	.	C	T	1.2494	PASS	TYPE=snp;AF=0.34342;DP=1421;ADF=385,189;ADR=548,299;AD=	
MN908947	WG	22679	.	T	C	9.03101	PASS	TYPE=snp;AF=0.315079;DP=2593;ADF=721,267;ADR=1054,550;A	
MN908947	WG	22686	.	C	T	9.03101	PASS	TYPE=snp;AF=0.297996;DP=2745;ADF=728,262;ADR=1197,556;A	
MN908947	WG	22813	.	G	T	3.45311	PASS	TYPE=snp;AF=0.178132;DP=1381;ADF=786,181;ADR=349,65;AD=	
MN908947	WG	22882	.	T	G	3.45138	PASS	TYPE=snp;AF=0.179538;DP=1515;ADF=653,162;ADR=590,110;AD	
MN908947	WG	22898	.	G	A	0.725514	PASS	TYPE=snp;AF=0.168823;DP=1546;ADF=619,152;ADR=664,10	
MN908947	WG	22917	.	T	G	1.2494	PASS	TYPE=snp;AF=0.826087;DP=1564;ADF=155,583;ADR=117,709;AD	
MN908947	WG	22995	.	C	A	258.013	PASS	TYPE=snp;AF=0.999609;DP=2559;ADF=0,1286;ADR=1,1272;AD=1	
MN908947	WG	23040	.	A	G	1.2494	PASS	TYPE=snp;AF=0.328217;DP=1694;ADF=710,334;ADR=428,222;AD	
MN908947	WG	23048	.	G	A	0.321851	PASS	TYPE=snp;AF=0.327627;DP=1694;ADF=710,333;ADR=426,22	
MN908947	WG	23055	.	A	G	0.579927	PASS	TYPE=snp;AF=0.362456;DP=1694;ADF=699,344;ADR=379,27	
MN908947	WG	23063	.	A	T	9.03101	PASS	TYPE=snp;AF=0.333799;DP=2864;ADF=1213,528;ADR=693,428;A	

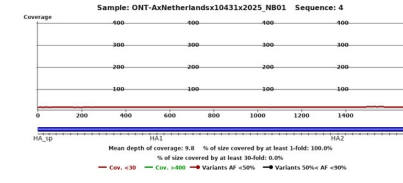
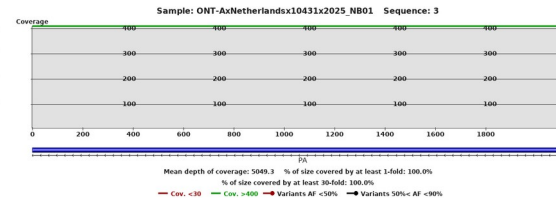
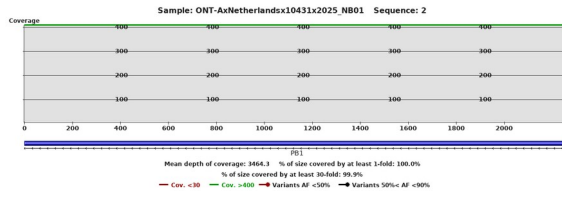
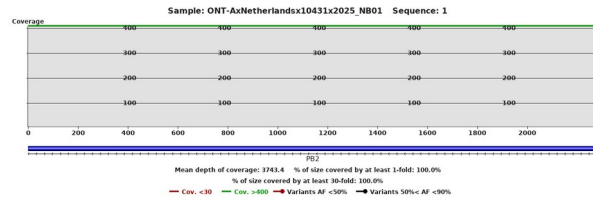
LE CAS DE LA GRIPPE : LES RÉASSORTIMENTS

S1/PB2

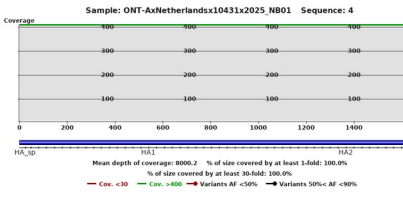
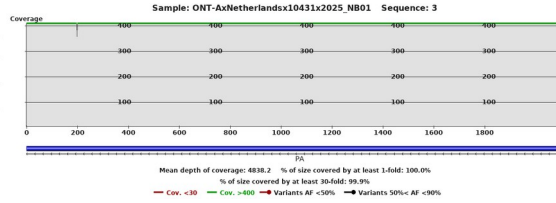
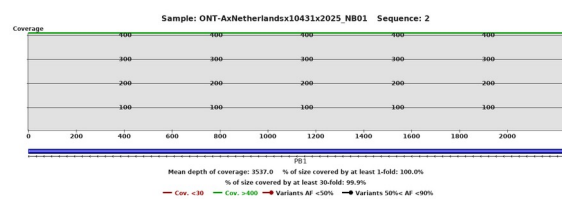
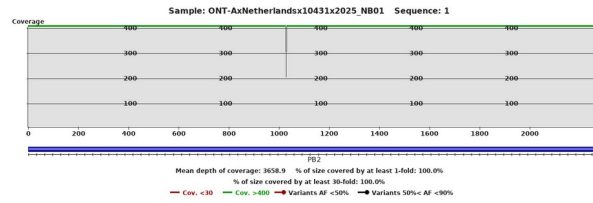
S2/PB1

S3/PA

S4/HA



Aln
H1N1



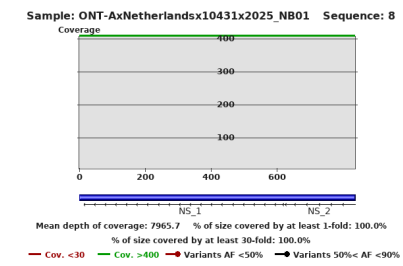
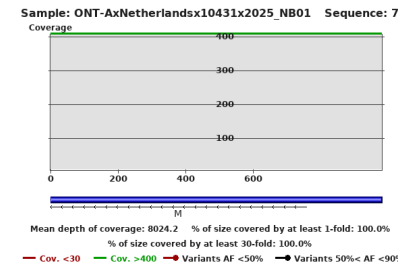
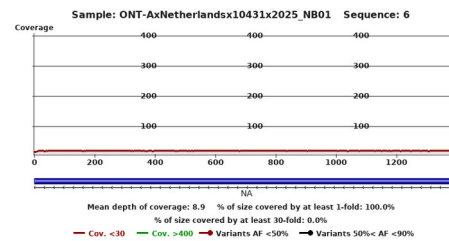
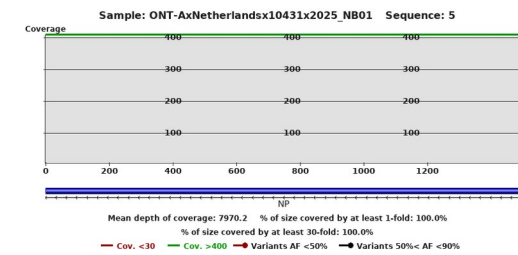
Aln
H3N2

S5/NP

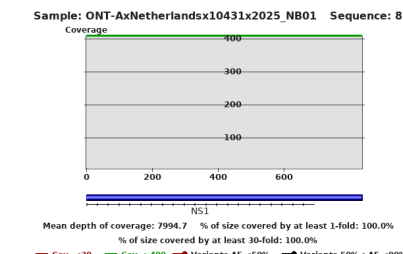
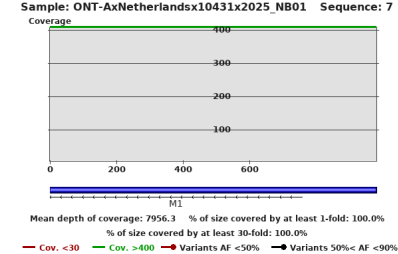
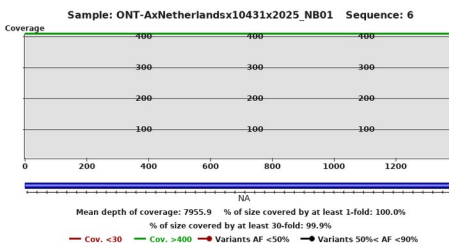
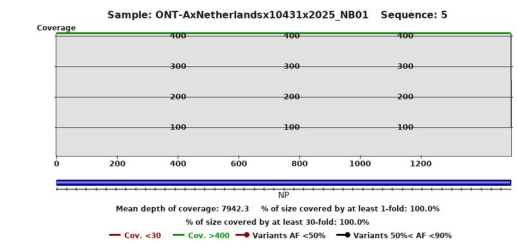
S6/NA

S7/MP

S8/NS



Aln
H1N1



Aln
H3N2



L'ANALYSE BIOINFORMATIQUE

PHYLOGENIE

CLASSIFICATION DES VIRUS RESPIRATOIRES

15



Nextstrain

Real-time tracking of pathogen evolution

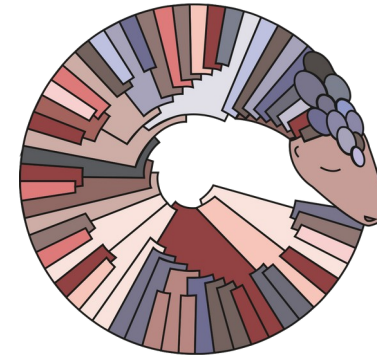
<https://nextstrain.org/>

Outil dédié à la classification des virus respiratoires, mais aussi à l'interprétation des séquences.

Peut être utilisé pour plusieurs pathogènes dont:

- le SARS-CoV-2
- la grippe saisonnière et aviaire
- le virus respiratoire syncycial
- les metapneumovirus
- le monkeypox

NB: disponibles pour d'autres pathogènes



Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN)

<https://cov-lineages.org/>

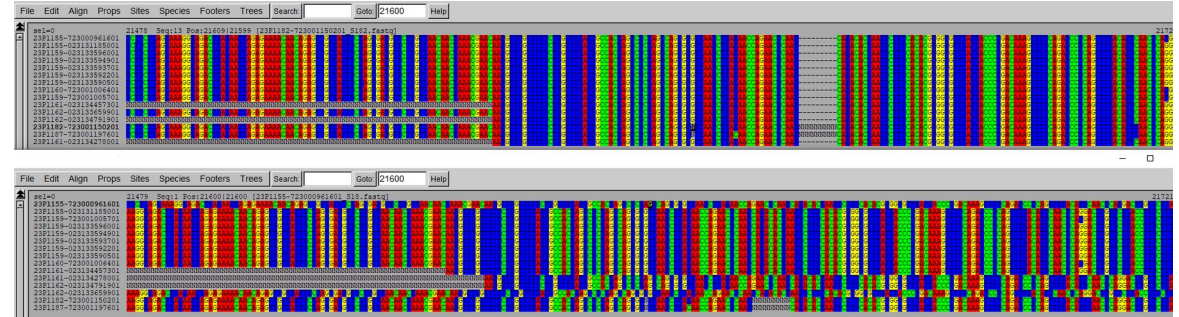
Outil dédié à la classification des séquences SARS-CoV-2.

CONSTRUCTION D'UN ARBRE PHYLOGENETIQUE

ANNOTATION

Inférence des métadonnées
Assignation des séquences
temporalité

MAFFT
Alignement multiple des séquences
<https://mafft.cbrc.jp/alignment/software/about.html>



*Nettoyage des séquences?
Gap penalties?*



IQTREE
Construction de l'arbre phylogénétique
<http://www.iqtree.org/doc/>

*Bootstrap?
Modèle de substitution?
Biologie du virus?*

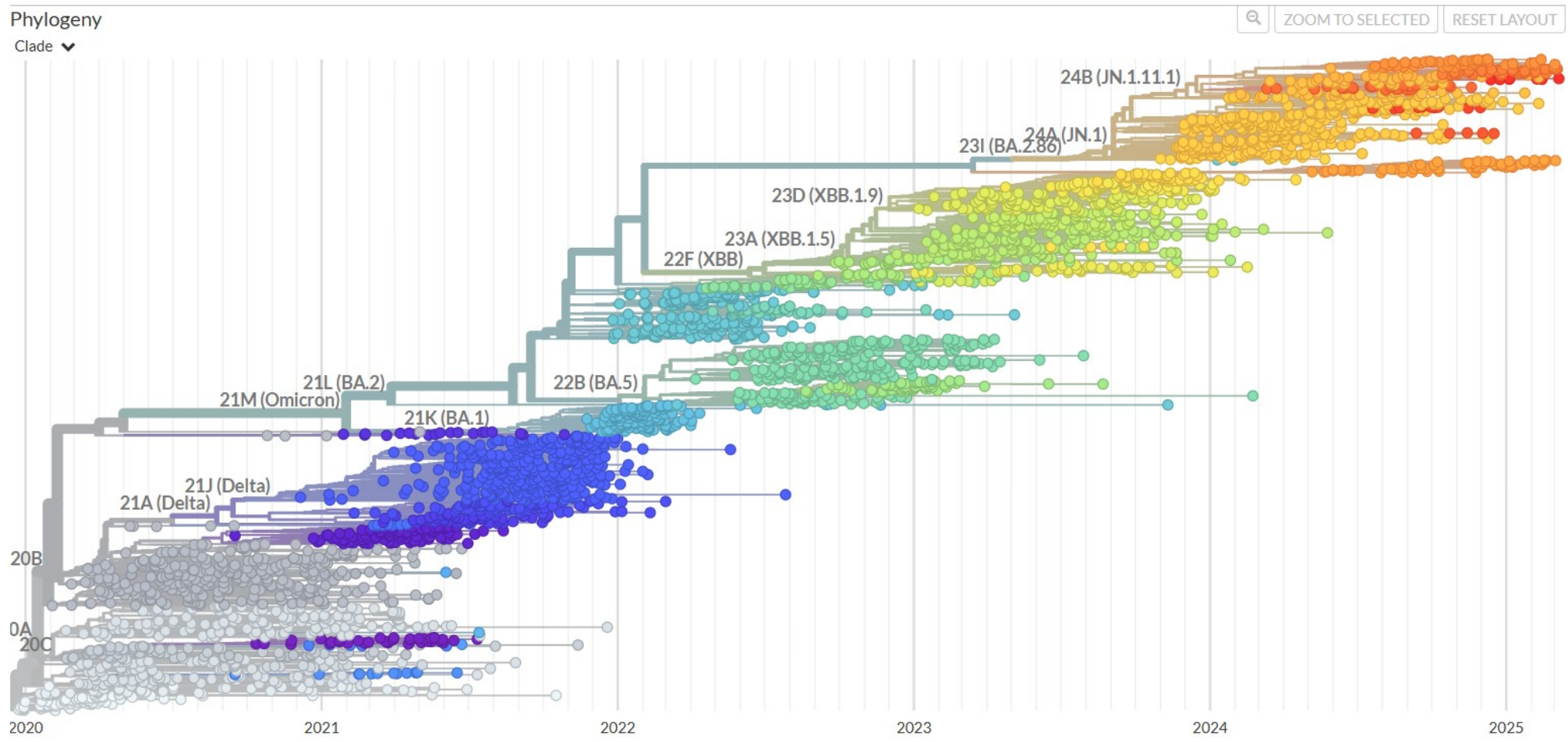


RÉTROSPECTIVE DE LA PANDÉMIE SARS-COV-2

Genomic epidemiology of SARS-CoV-2 with subsampling focused globally since pandemic start

Built with [nextstrain/ncov](#). Maintained by the [Nextstrain team](#). Data updated 2025-03-20. Enabled by data from [GISAID](#).

Showing 3439 of 3439 genomes sampled between Dec 2019 and Mar 2025.



<https://nextstrain.org/ncov/gisaid/global/all-time>



MERCI

www.chu-lyon.fr



HCL
HOSPICES CIVILS
DE LYON