ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

LABORATORIO DE GENÓMICA y BIOINFORMÁTICA

Dr. TOMÁS POKLÉPOVICH CARIDE
tjcaride@anlis.gob.ar
@Poklepovich

Unidad Operativa Centro Nacional de Genómica y Bioinformática

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

PORQUÉ SECUENCIAR A NIVEL GLOBAL

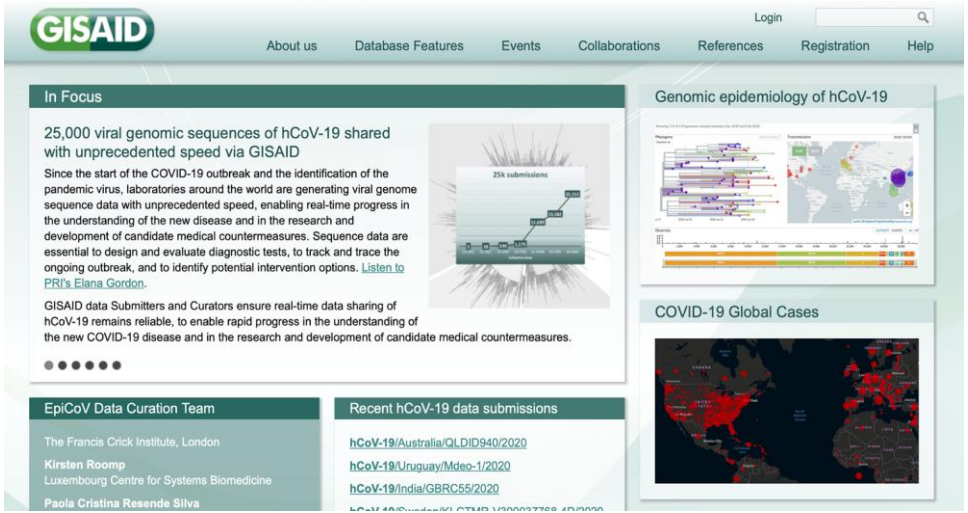
INVESTIGAR EL **ORIGEN** DEL VIRUS

ESTUDIAR LA **EXPANSIÓN** GLOBAL DE LA EPIDEMIA

ESTUDIAR LA **VARIABILIDAD** PARA ASEGURAR DIAGNÓSTICO Y DESARROLLO DE VACUNAS

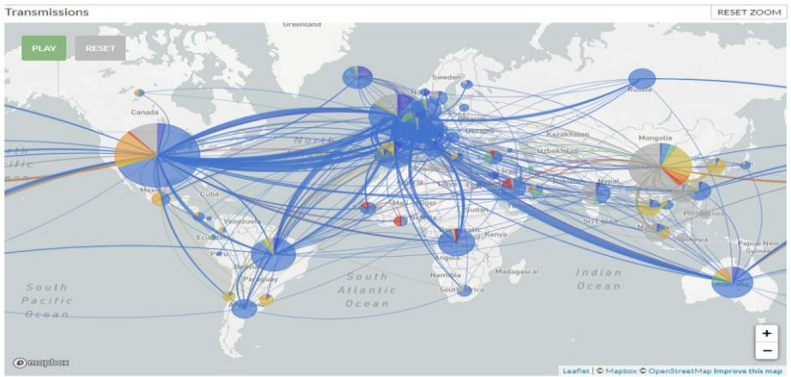
ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

SARS-CoV-2
ESTUDIO GLOBAL



>3.403.170 genomas en GISAID

VIGILANCIA GENÓMICA A TIEMPO REAL



PORQUÉ SECUENCIAR A NIVEL NACIONAL

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

ESTUDIAR LAS **INTRODUCCIONES**

VER LA **TRANSMISIÓN** COMUNITARIA LOCAL Y NACIONAL

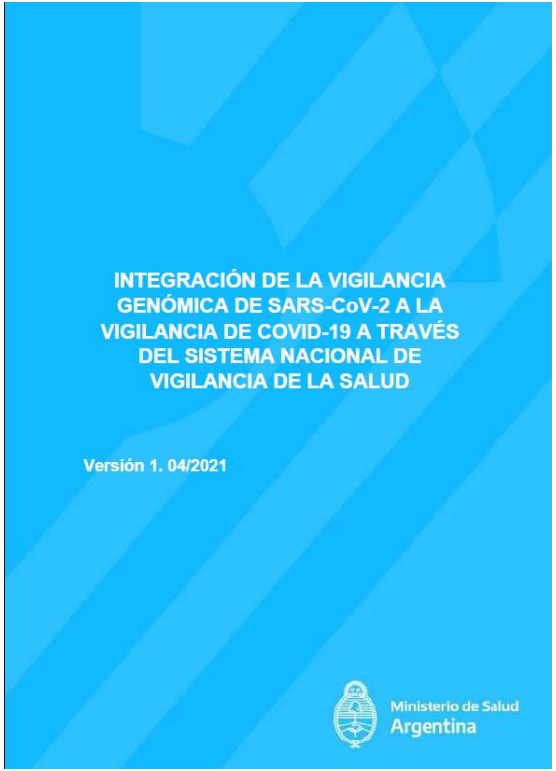
ESTUDIAR LA **VARIABILIDAD** LOCAL PARA ASEGURAR DIAGNÓSTICO Y DESARROLLO DE VACUNAS

HACER EL SEGUIMIENTO DE LOS SUCESIVOS **BROTES**

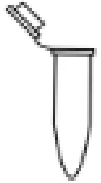
Criterios para selección de muestras para vigilancia genómica

- **Vigilancia general** de las variantes circulantes en la comunidad
- Investigación de nuevas variantes en sospechas de **reinfecciones**
- Investigación de nuevas variantes en pacientes **vacunados**
- Investigación de nuevas variantes en **escenarios de alta transmisibilidad o virulencia**
- Investigación de nuevas variantes en **viajeros** internacionales

- Muestras de HNF o Saliva con Ct menor o igual a 30. Conservadas en frio y enviadas correctamente.



Laboratorio con Miseq - Illumina



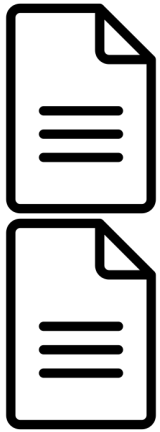
DNA input

Preparo bibliotecas

Library

Las cargo en un cartucho

Archivos



Reactivos necesarios: Kit para Library Prep + Cartucho de secuenciación + Flow Cell

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

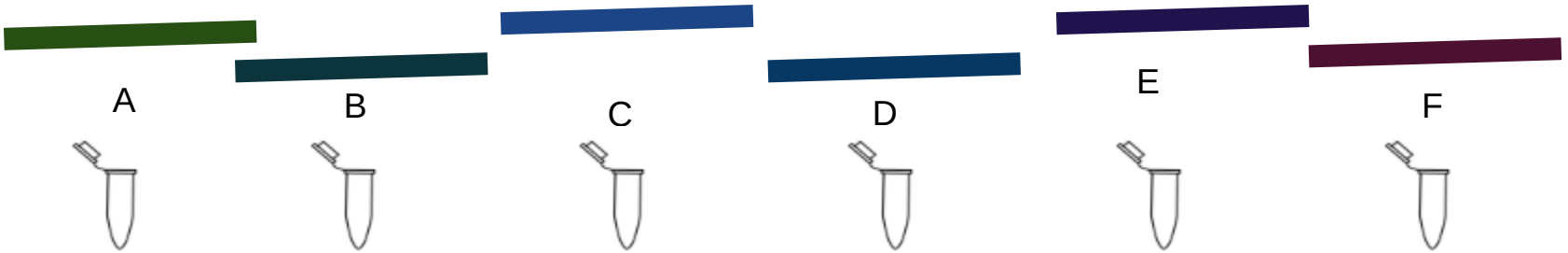
Protocolo Chile

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

ssRNA(+) ~30.000bp

One-Step RT-PCR

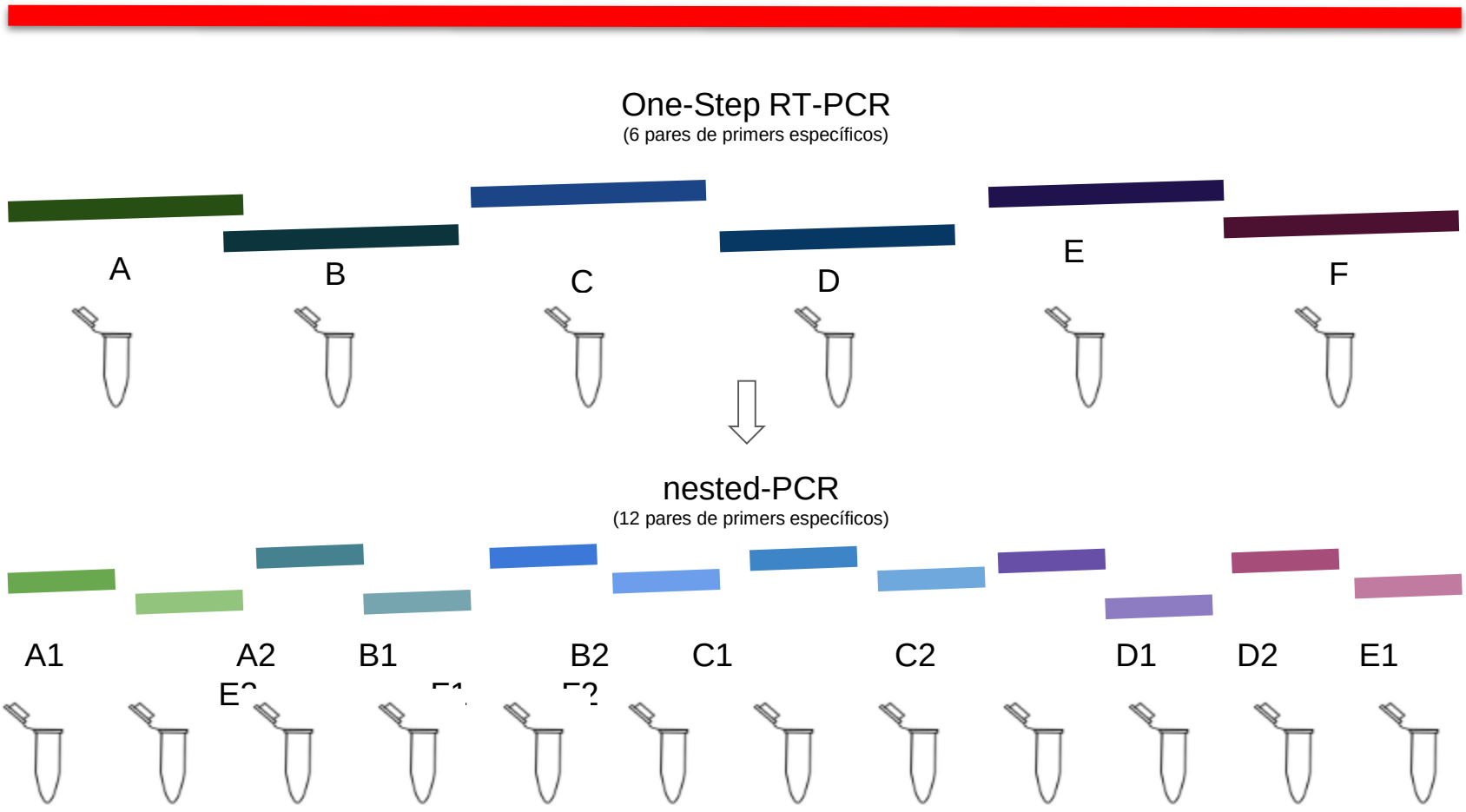
(6 pares de primers específicos)



Protocolo Chile

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

ssRNA(+) ~30.000bp



Protocolo ARTIC

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

ssRNA(+) ~30.000bp

RT
(random primers)

cDNA

1 h

PCR con primers de ARTIC
Ultraplex específica

5 h

Pool 1

Pool 2

Pool 1 + Pool 2 = DNA input

Evolución de multiplexado

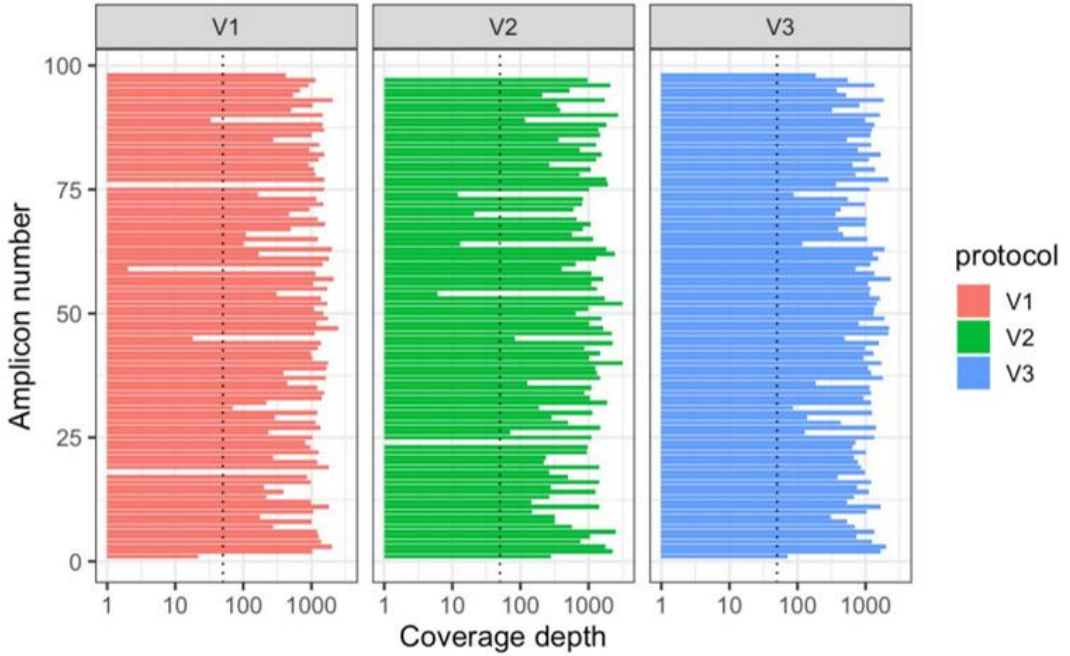


Figure 1: Comparison of V1, V2 and V3 protocols run on the same clinical sample with a barcoded library. For comparison, 100,000 randomly subsampled reads were used. As a guide to sufficient coverage, the vertical dotted line indicates 50x coverage.

Kits de **Library** prep para illumina®

Nextera XT

El kit que veníamos utilizando

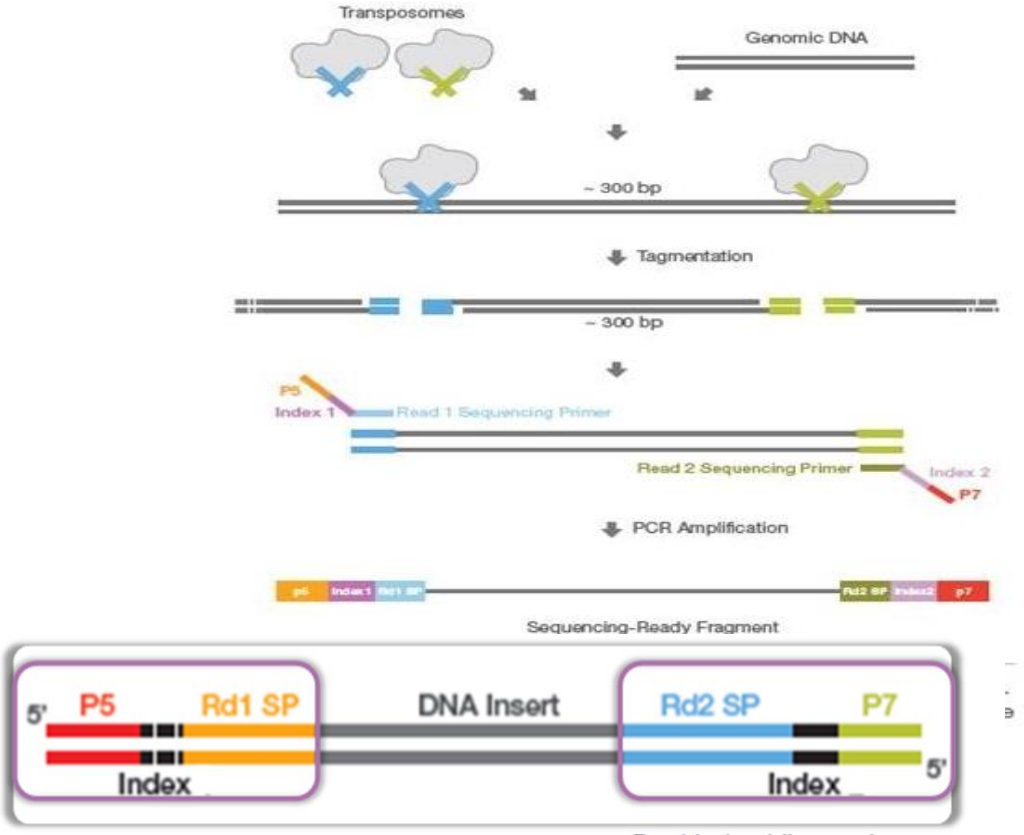
Nextera DNA Flex

El kit que se planeaba utilizar en el 2020

CovidSeq

Solución de Illumina optimizada para amplicones de ARTIC

SEGUNDA GENERACIÓN ILLUMINA



SEGUNDA GENERACIÓN ILLUMINA

Pueden quedar las secuencias de los primers Artic en los extremos de las lecturas

Es importante recortarlos para evitar errores

Library 1

Library 2

…

Library n

Pool de libraries

5 h

26 h

FastQ de muestra 1

FastQ de muestra 2

…

FastQ de muestra n

- En un archivo **FastQ** se almacenan las miles de lecturas o **reads** correspondientes a una library junto con su calidad.
- Las reads de Illumina son secuencias cortas.
- Para armar el genoma de ~30.000bp se requiere de **bioinformática**.

ANLIS MALBRÁN

ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

Control de Calidad

Reporte de FastQC

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content
- Kmer Content

Basic Statistics

Measure	Value
Filename	MI.M00833_0484.001.Index_16.D2_R2.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1396235
Sequences flagged as poor quality	0
Sequence length	250
%GC	34

Per base sequence quality

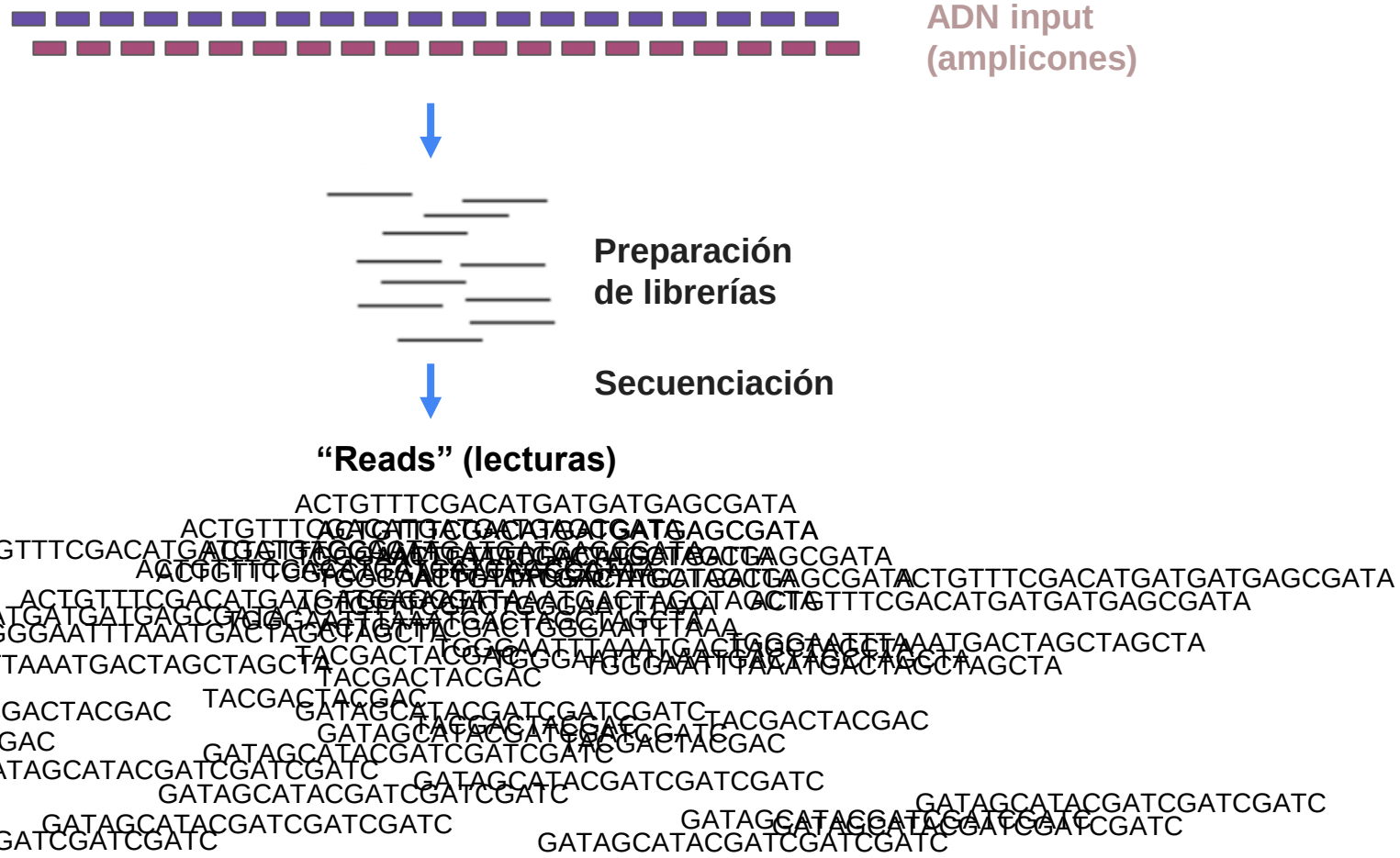
Control de Calidad

Reporte de **Kraken**

Permite la clasificación de **reads** de manera rápida a través de un algoritmo que las asocia al grupo taxonómico del ancestro común más bajo (LCA, Lowest Common Ancestor) usando un alineamiento exacto de **kmers.**

Categorías taxonómicas principales:

Domain D
Kingdom K
Phylum P
Class C
Order O
Family F
Genus G
Species S


ADN input
(amplicones)
Preparación
de librerías
Secuenciación
“Reads” (lecturas)
ACTGTTTCGACATGATGATGAGCGATA
TACGACTACGAC
GATAGCATACGATCGATCGATC
TGGGAATTTAAATGACTAGCTAGCTA

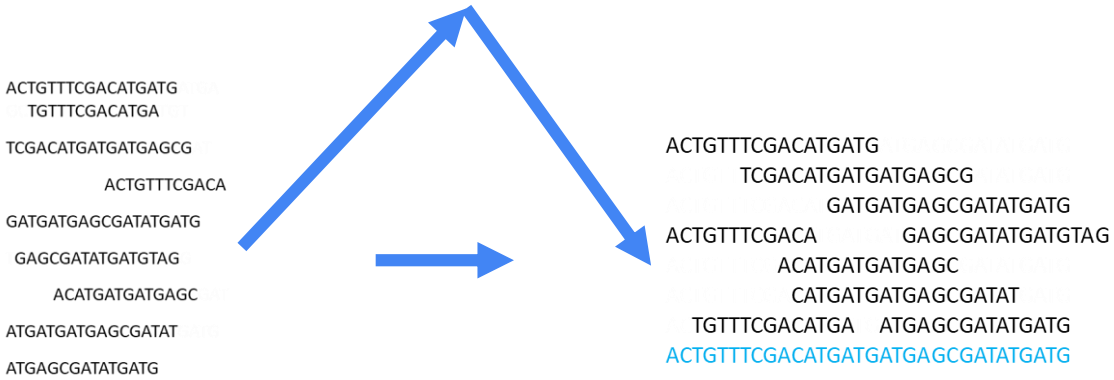
Como armar la secuencia del virus?

Ensamblado *de novo*

Mapeo contra referencia

¿EN QUÉ CONSISTE EL ENSAMBLE *DE NOVO*?

Remover secuencias de primers (cuando sea necesario)



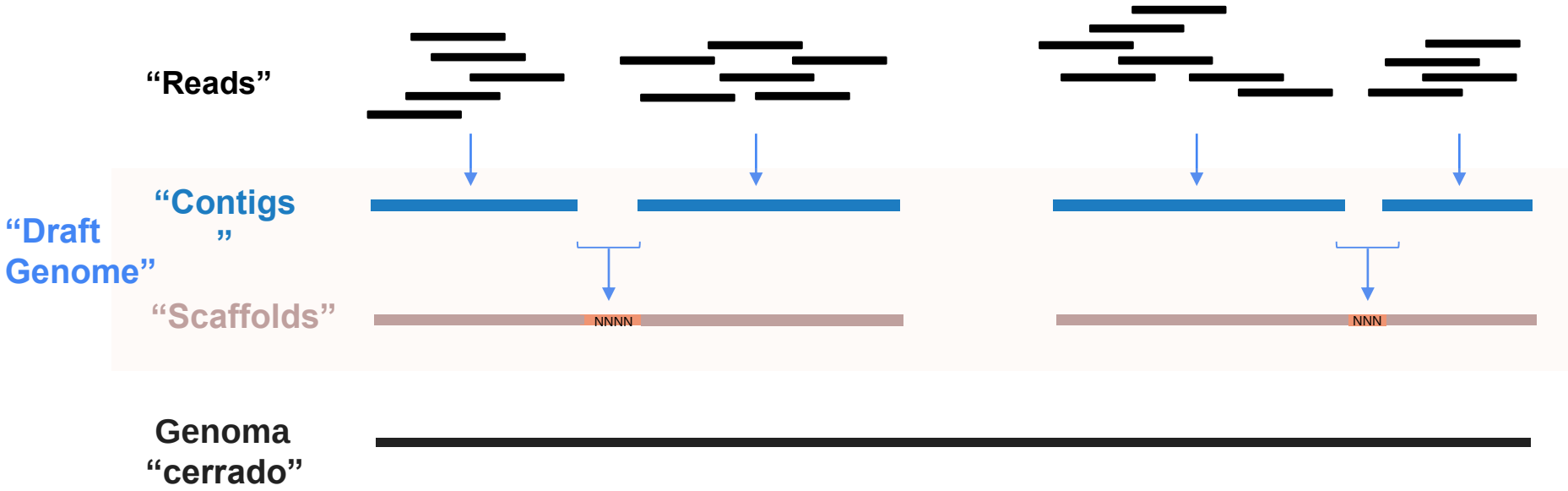
Lecturas (“Reads”)

≥x10⁶

“Contigs”

Muchos menos que x10⁶

¿EN QUÉ CONSISTE EL ENSAMBLE *DE NOVO*?



Parece hace siglos...

Nature. 2020; 579(7798): 265–269. PMCID: PMC7094943
Published online 2020 Feb 3. doi: 10.1038/s41586-020-2008-3 PMID: 32015508

A new coronavirus associated with human respiratory disease in China

Fan Wu,[#1] Su Zhao,[#2] Bin Yu,[#3] Yan-Mei Chen,[#1] Wen Wang,[#4] Zhi-Gang Song,[#1] Yi Hu,[#2] Zhao-Wu Tao,[2] Jun-Hua Tian,[3] Yuan-Yuan Pei,[1] Ming-Li Yuan,[2] Yu-Ling Zhang,[1] Fa-Hui Dai,[1] Yi Liu,[1] Qi-Min Wang,[1] Jiao-Jiao Zheng,[1] Lin Xu,[1] Edward C. Holmes,[1,5] and Yong-Zhen Zhang[1,4,6]

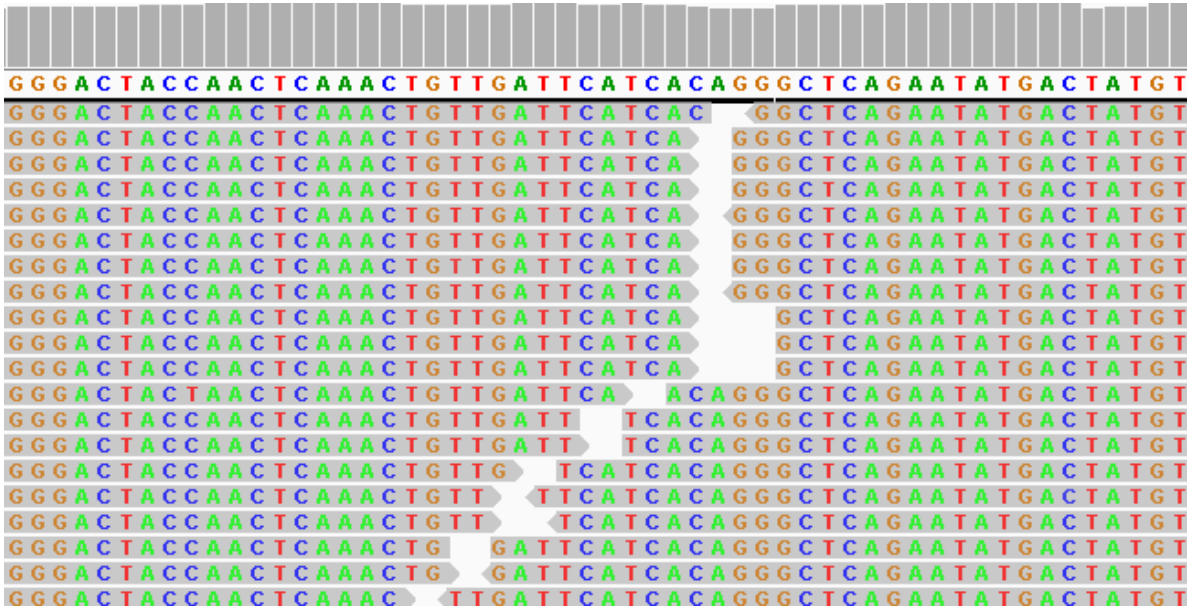
Genoma de referencia

Mapeo contra referencia

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

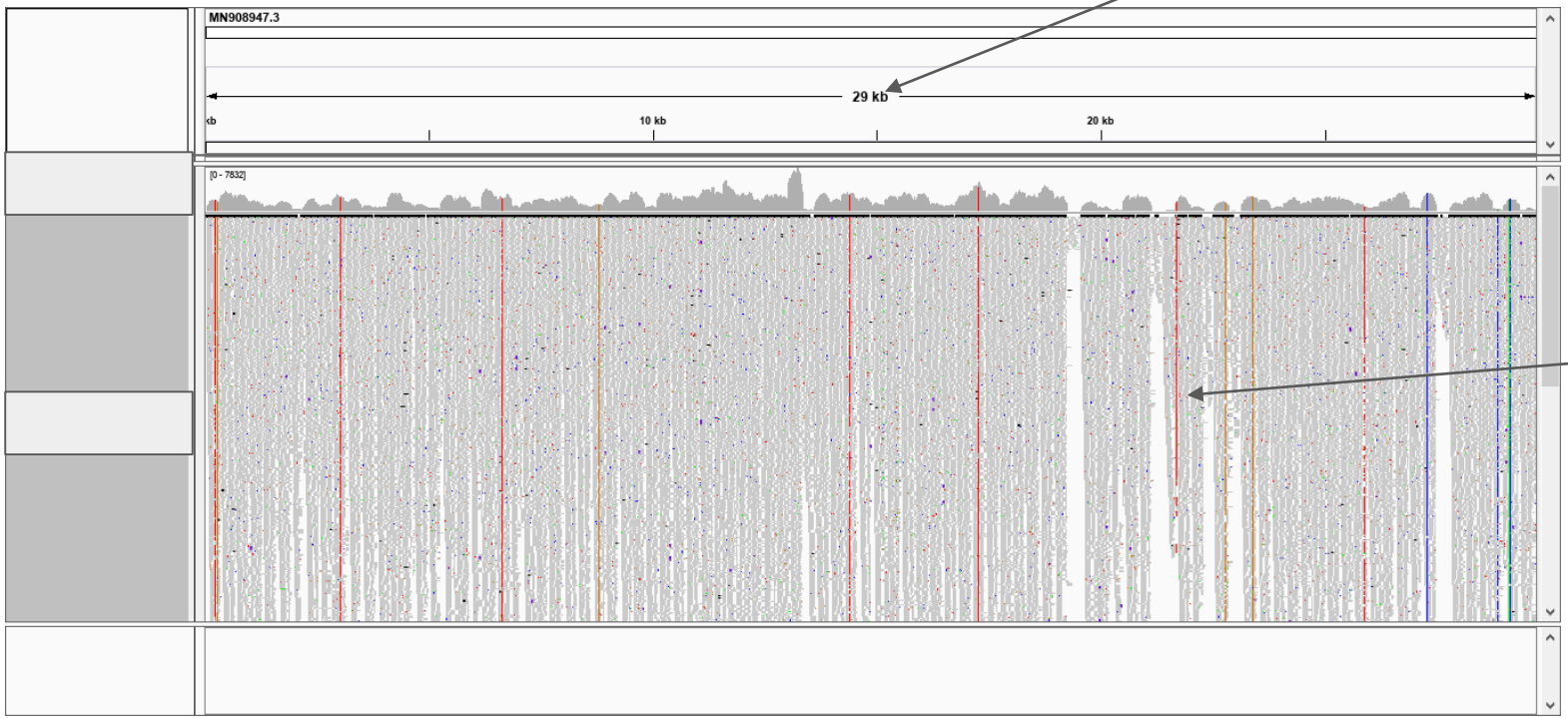
Secuencia de referencia Wuhan-Hu-1

Reads mapeadas

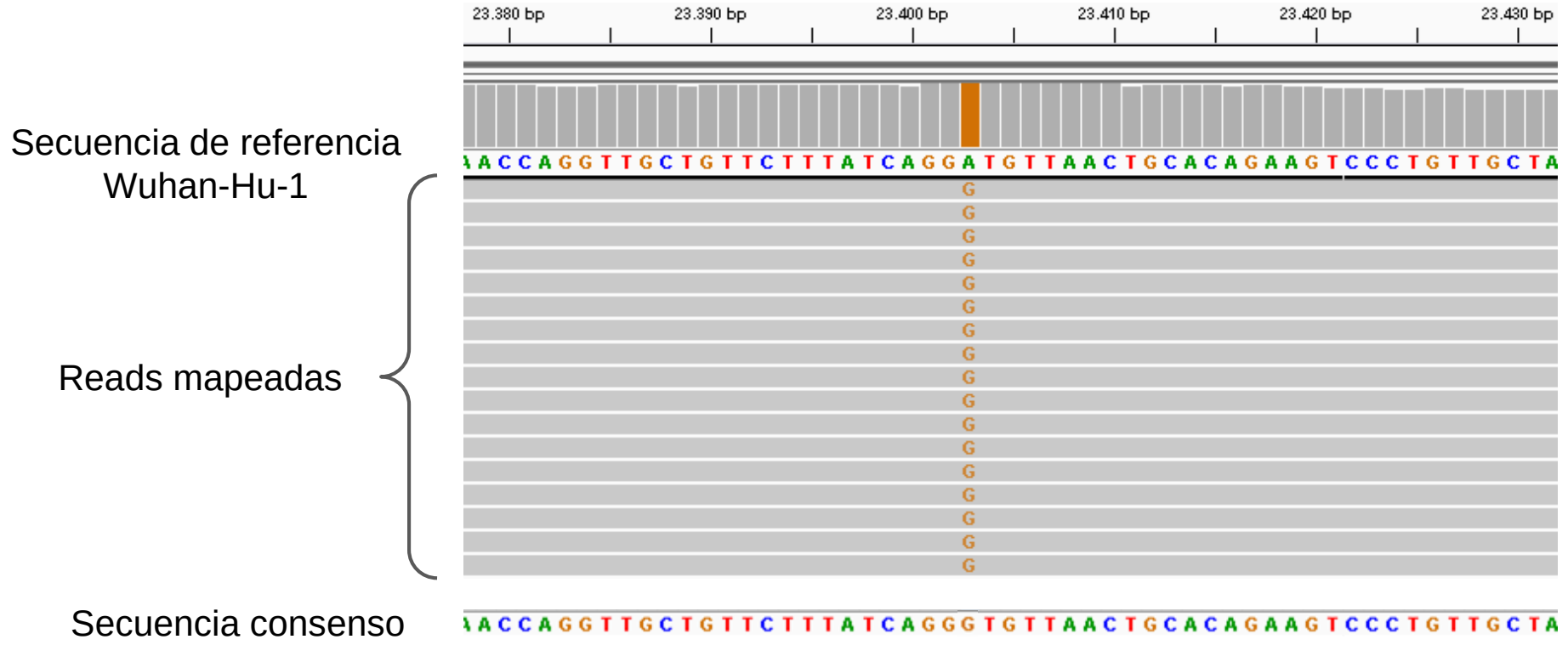


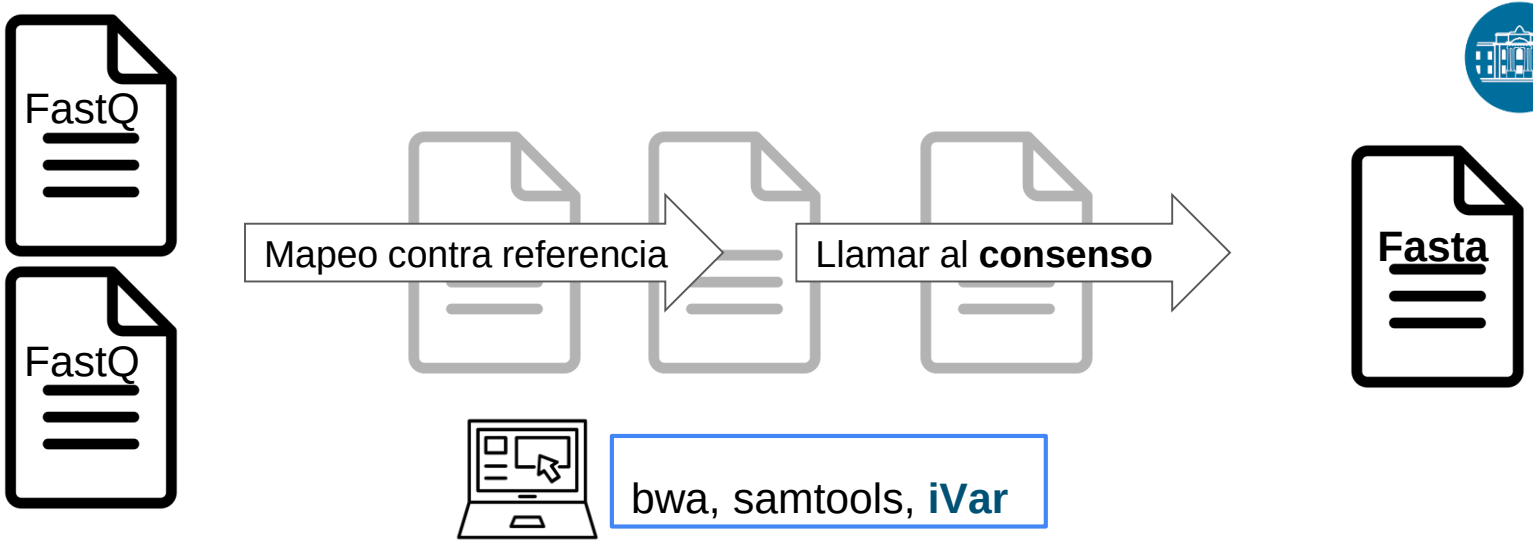
Mapeo contra referencia

Archivo BAM: guarda la información del mapeo



¿Cómo se ve una mutación?





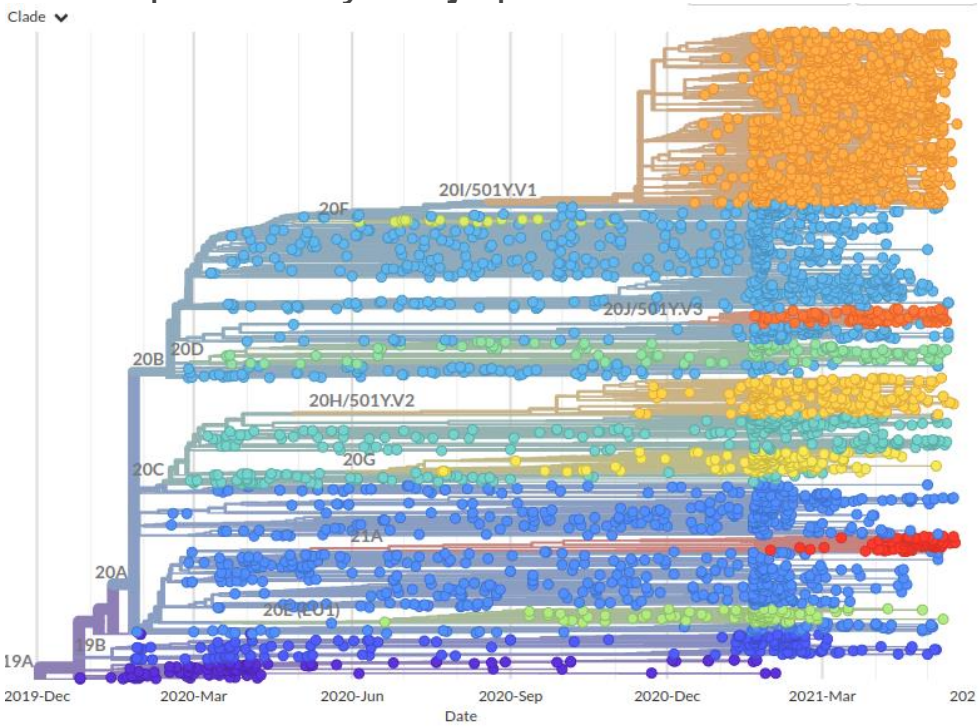
>hCoV-19/Argentina/INEI104556/2021

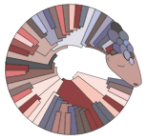
ACAGATAGATCAAGATACAGATACAACAGATAGATCAACAGATAGATCAAGATACAGAT
AC
AGCAGATCTCAACAGATAGATCAAGATACAGATACAACAGATAGATCAAGATACAGATA
CAAAGATAGATCAAGATACAGATACAACAGATAGATCAAGATACAGATACCCAACAGAT
AGATCCACAGATAGATCAAGATACAGGGGATACAAGATACAGATACNNNNNNNNNNNN
NTACAGATACAACAGATAGATCAAGATACAGATACAGATAGATCAAGATACTTTAGATAC
AACAACAGACCCTAGATCAAGATACAGATACAACAGATAGATCAAGATACAGATACAAC
AGATAGATAGTACATCATTAG.....

~30.000bp

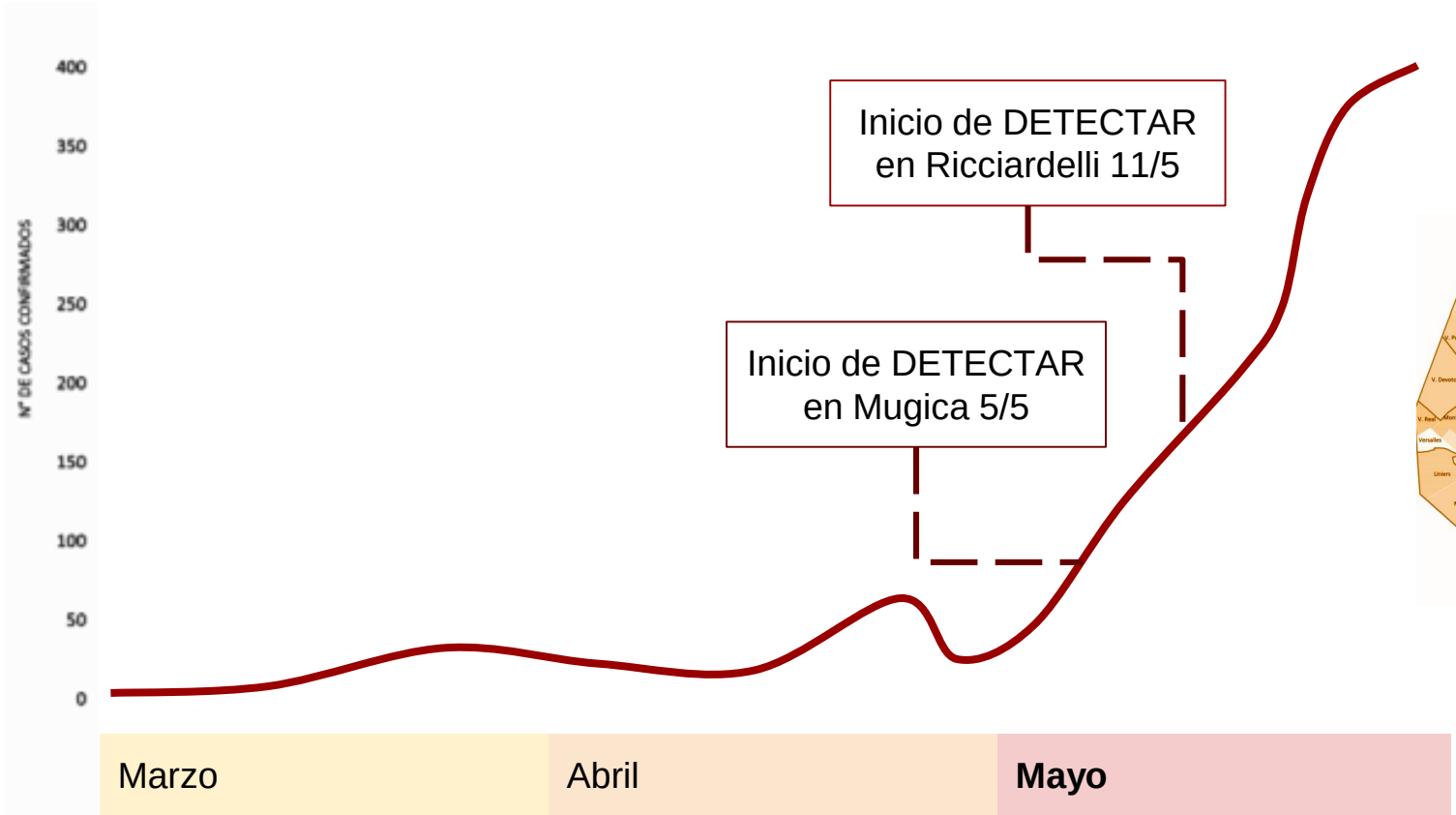
¿Qué hago con los FASTAs?

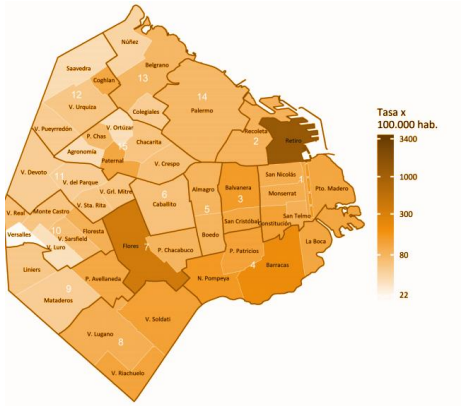
- Evaluar calidad y subirlos a **GISAID** junto a la metadata
- Ver **mutaciones** en regiones de interés
- Definir a qué **clado** y **linaje** pertenecen



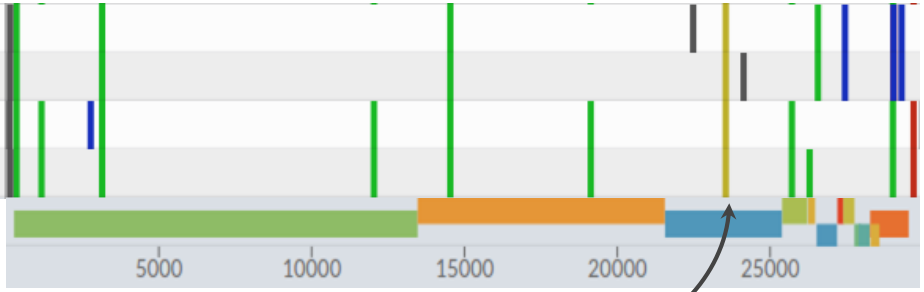
Coexistencia de Nomenclaturas				
Nextstrain	**PANGO lineages**	**GISAID**	**Medios de comunicación y redes sociales**	**WHO**
Clados (más macro)	**Linajes** (más micro y dinámico)	Clados	Supuesto origen geográfico	Nomenclatura reciente de VOC y VOI
19A, 19B 20A, 20B, …, 20G, 20I 21A	B.1.1.7 P.1 C.37 B.1.499 N.5 B.1.617	S L V O G GV GH GR ...	“La de UK” “La de Manaos” “La andina” “La de Sudáfrica” “La de India”	Alpha Gamma Delta Lambda Mu
Nextclade 	**Pangolin** 	GISAID		

Marzo-Mayo 2020 en CABA. Casos confirmados.

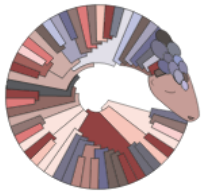




	QC	Clade	Mut.	non-ACGTN	Ns	Gaps	Sequence view
a	N M P C	20B	10	0	197	0	
b	N M P C	20B	10	0	43	0	
c	N M P C	20C	11	0	7	0	
d	N M P C	20C	11	0	48	0	



Mutación en Spike **D614G**

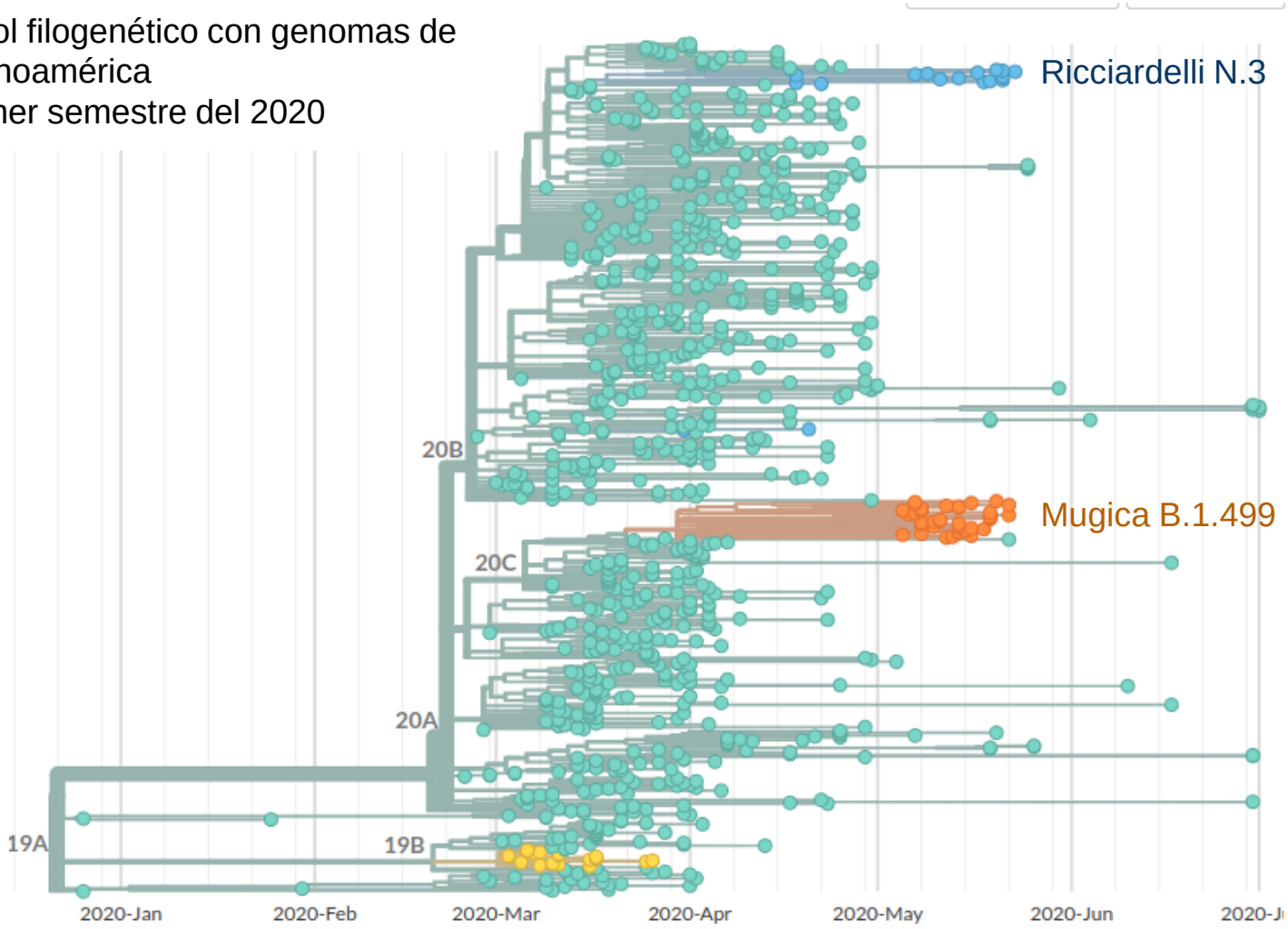


	Lineage	Assignment Conflict
a	N.3	0.0
b	N.3	0.0
c	B.1.499	0.0
d	B.1.499	0.0

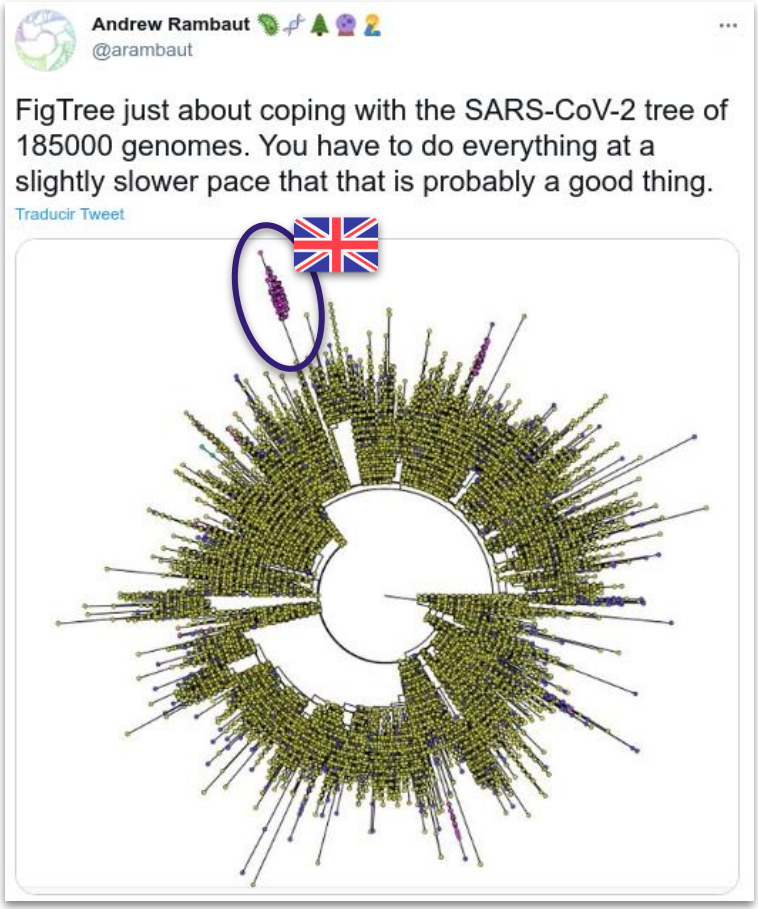
Ricciardelli	Mugica
>90% Clado 20B Linaje N.3	>90% Clado 20C Linaje B.1.499

ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

Árbol filogenético con genomas de Latinoamérica
Primer semestre del 2020



Diciembre 2020. Variantes.



Virological.org

Clado 20I

Linaje B.1.1.7

Gran número de mutaciones, principalmente en la Spike

Variants of Concern (VOC)

Working definition:

A SARS-CoV-2 variant that meets the definition of a VOI (see below) and, through a comparative assessment, has been demonstrated to be associated with one or more of the following changes at a degree of global public health significance:

- Increase in transmissibility or detrimental change in COVID-19 epidemiology; OR
- Increase in virulence or change in clinical disease presentation; OR
- Decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics

Variants of Interest (VOI)

Working definition

A SARS-CoV-2 variant :

- with genetic changes that are predicted or known to affect virus characteristics such as transmissibility, disease severity, immune escape, diagnostic or therapeutic escape; AND
- Identified to cause significant community transmission or multiple COVID-19 clusters, in multiple countries with increasing relative prevalence alongside increasing number of cases over time, or other apparent epidemiological impacts to suggest an emerging risk to global public health.

Reclassifying VOIs/ VOCs

A previously designated Variant of Interest (VOI) or Variant of Concern (VOC) which has conclusively demonstrated to no longer pose a major added risk to global public health compared to other circulating SARS-CoV-2 variants, can be reclassified.

This is undertaken through a critical expert assessment, in collaboration with Technical Advisory Group on Virus Evolution, of several criteria, such as the observed incidence/relative prevalence of variant detections among sequenced samples over time and between geographical locations, the presence/absence of other risk factors, and any ongoing impact on control measures.

Alerts for Further Monitoring

Working definition

A SARS-CoV-2 variant with genetic changes that are suspected to affect virus characteristics with some indication that it may pose a future risk, but evidence of phenotypic or epidemiological impact is currently unclear, requiring enhanced monitoring and repeat assessment pending new evidence.

Note: It is expected that our understanding of the impacts of these variants may fast evolve, and designated Alerts for Further Monitoring may be readily added/removed; therefore, WHO labels will not be assigned at this time. Former VOIs/VOCs may, however, be monitored for an extended period under this category, and will maintain their assigned WHO label until further notice.

Currently designated Variants of Concern:

WHO label	Pango lineage*	GISAID clade	Nextstrain clade	Additional amino acid changes monitored°	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		
Delta	B.1.617.2§	G/478K.V1	21A	+S:417N		

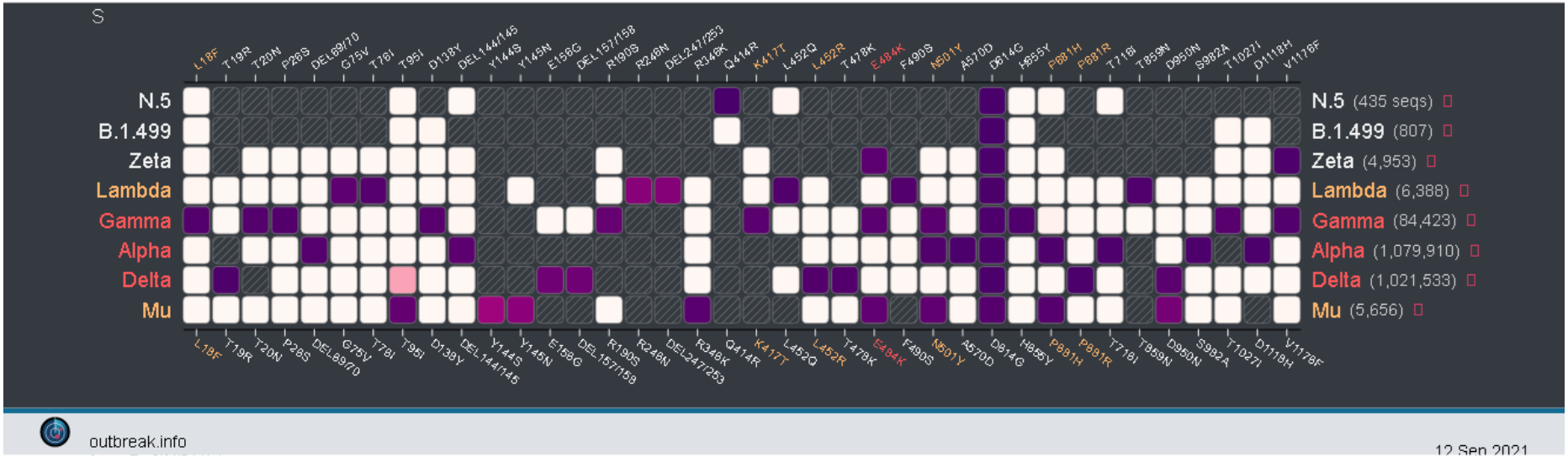
Currently designated Variants of Interest:

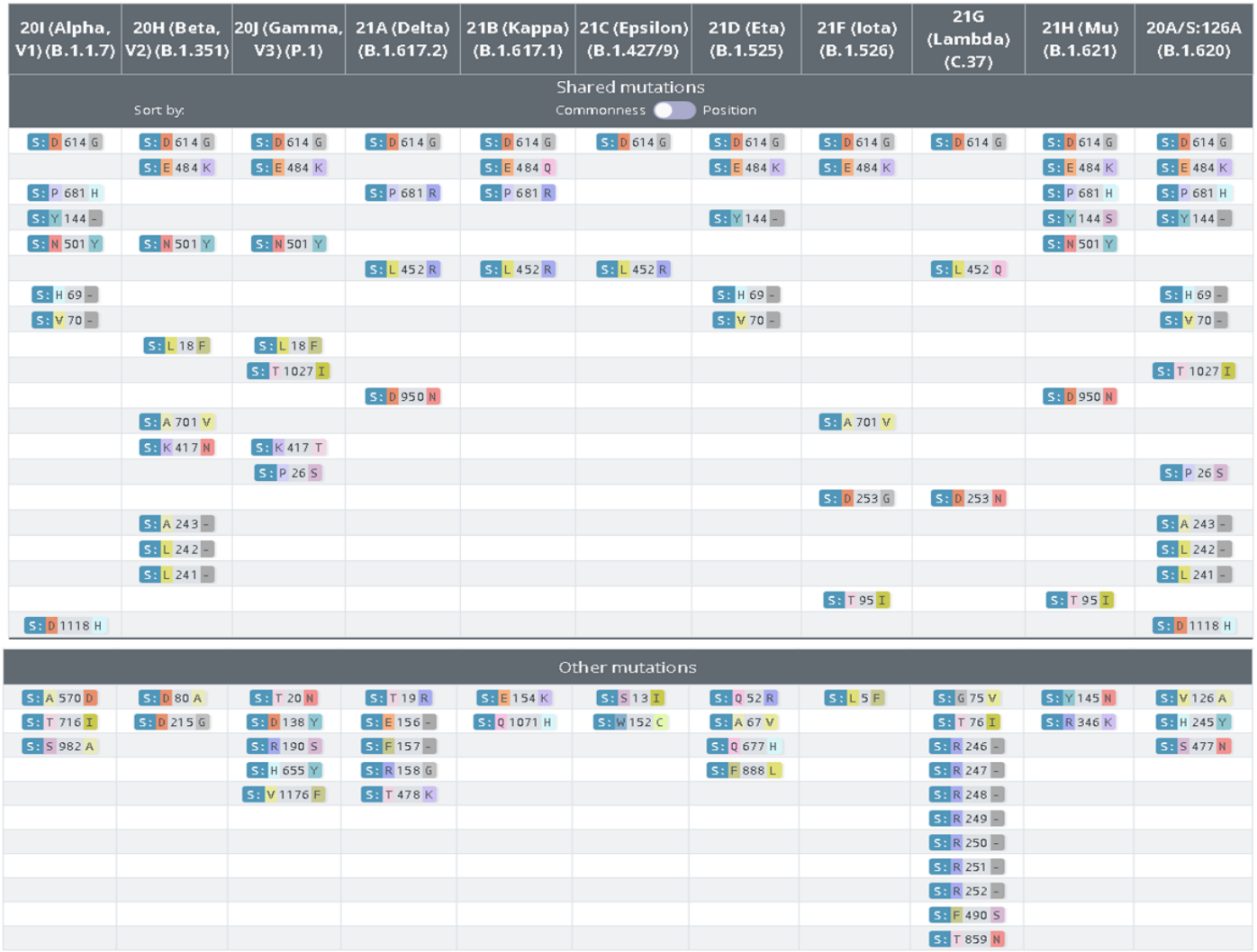
WHO label	Pango lineage*	GISAID clade	Nextstrain clade	Earliest documented samples	Date of designation
Eta	B.1.525	G/484K.V3	21D	Multiple countries, Dec-2020	17-Mar-2021
Iota	B.1.526	GH/253G.V1	21F	United States of America, Nov-2020	24-Mar-2021
Kappa	B.1.617.1	G/452R.V3	21B	India, Oct-2020	4-Apr-2021
Lambda	C.37	GR/452Q.V1	21G	Peru, Dec-2020	14-Jun-2021
Mu	B.1.621	GH	21H	Colombia, Jan-2021	30-Aug-2021

https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

Constelaciones y variantes (VOC, VOI….)

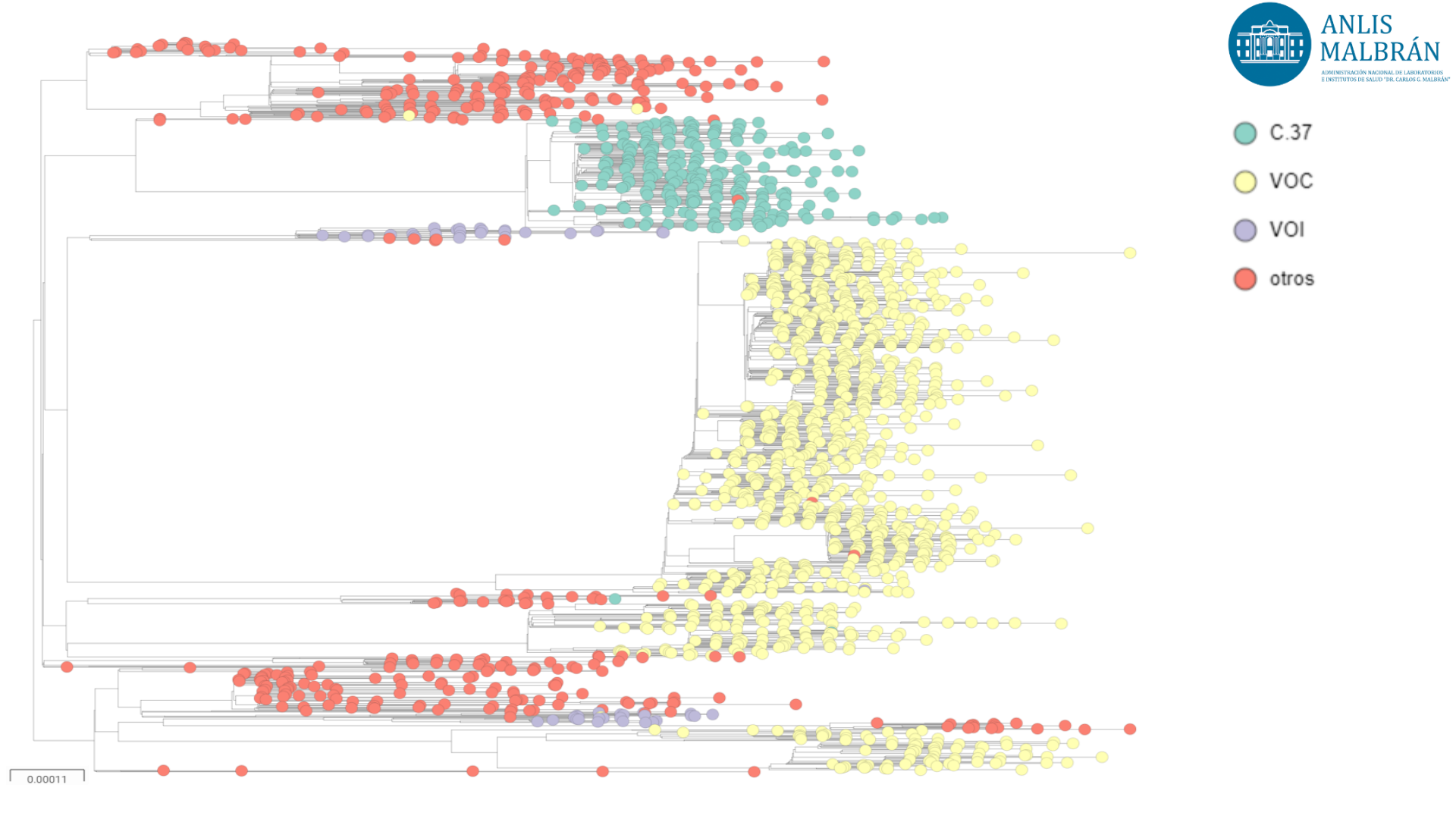
Mutations by lineage

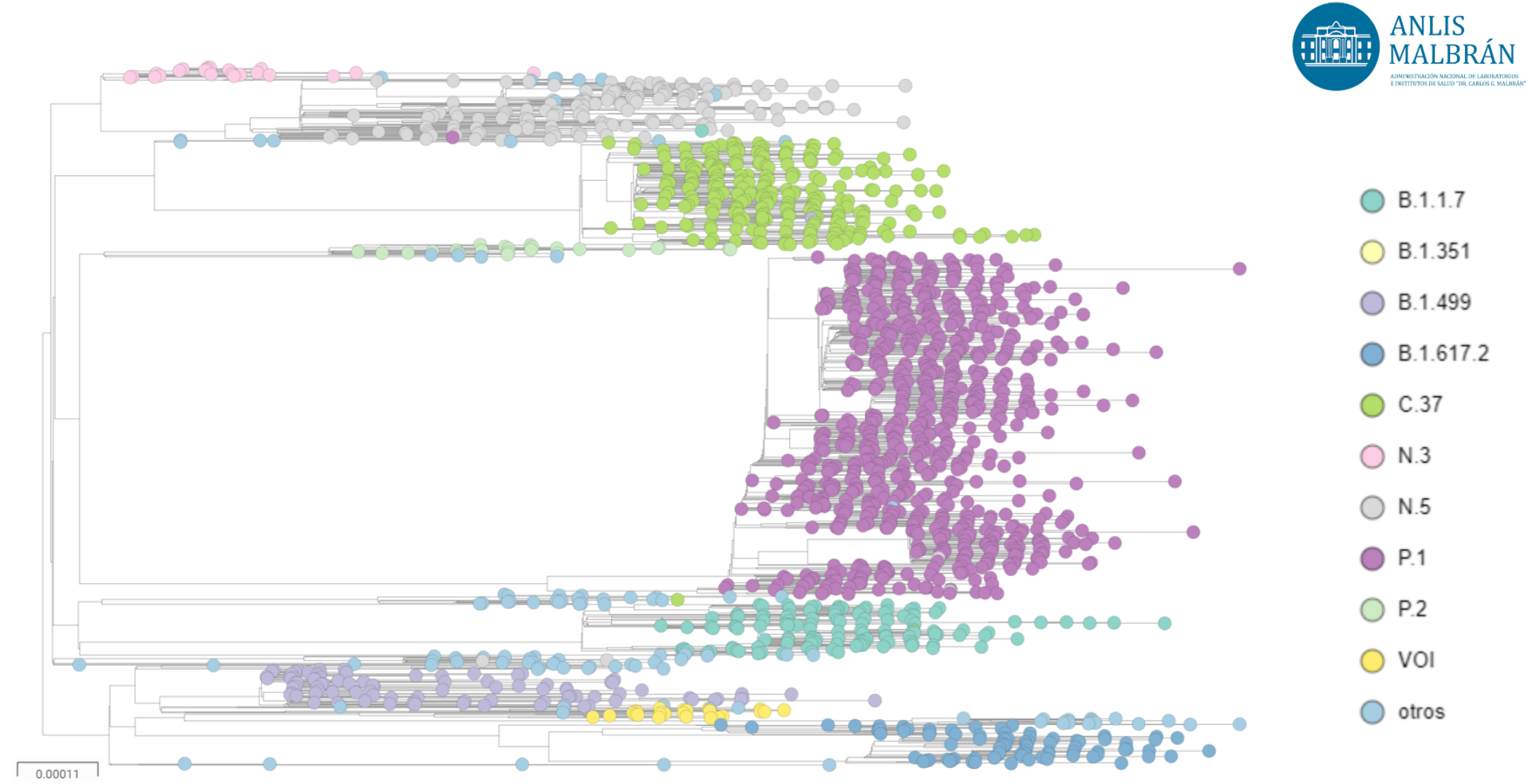


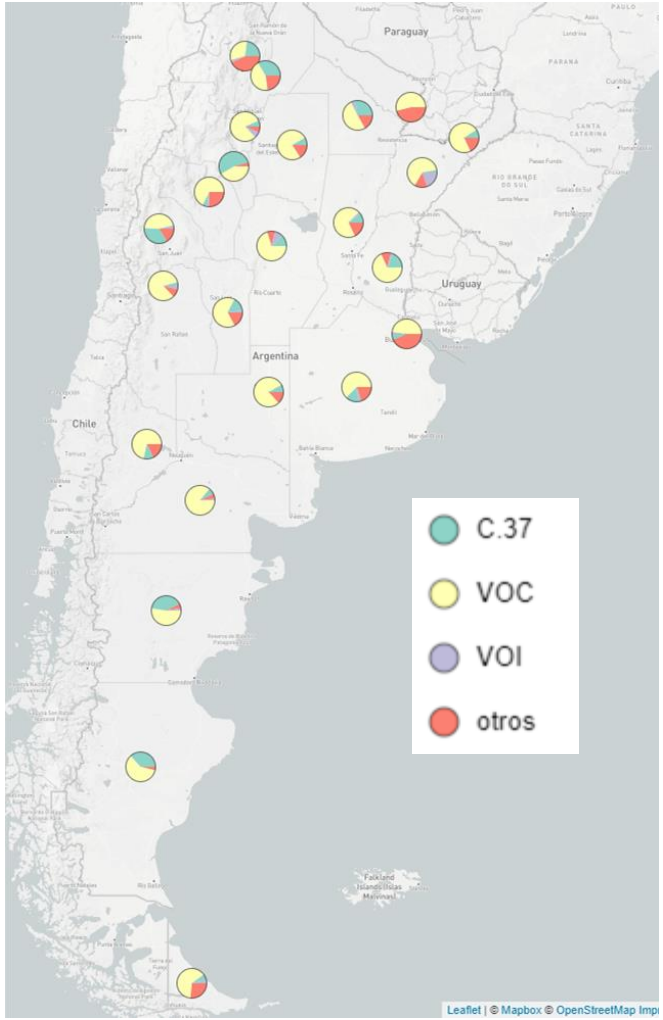


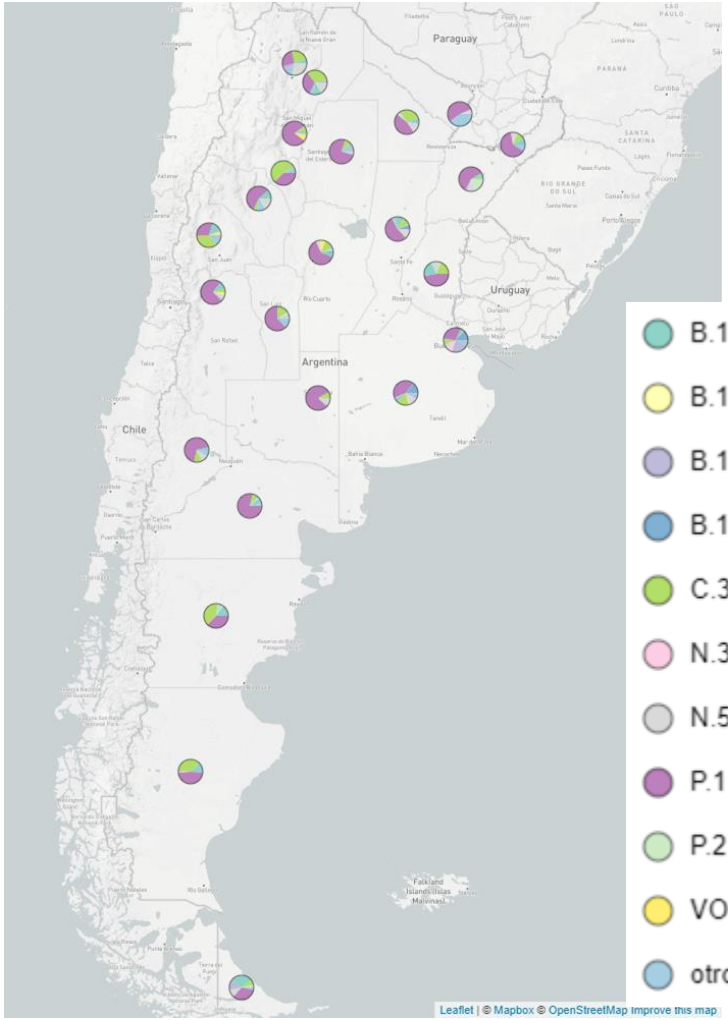
ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

https://covariants.org/


ANLIS
MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS
E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"
C.37
VOC
VOI
otros
0.00011


ANLIS
MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS
E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"
B.1.1.7
B.1.351
B.1.499
B.1.617.2
C.37
N.3
N.5
P.1
P.2
VOI
otros
0.00011

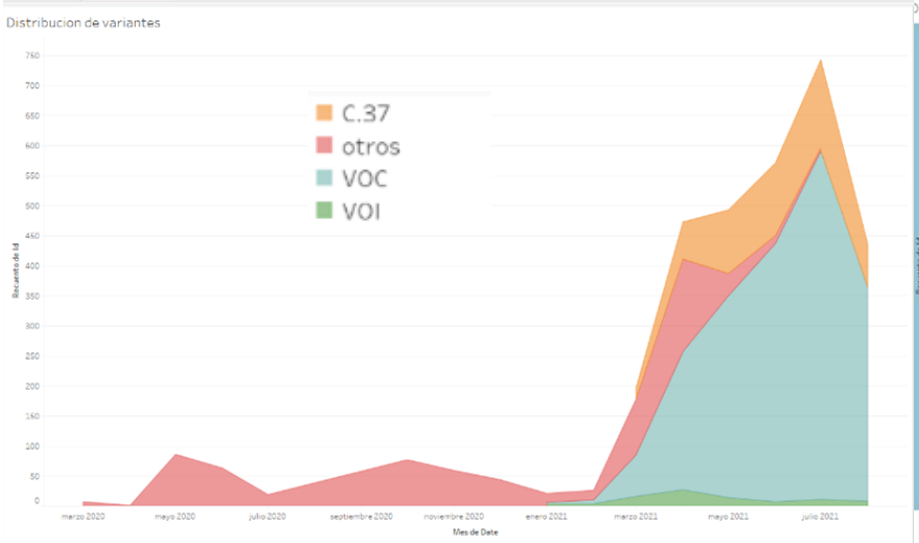




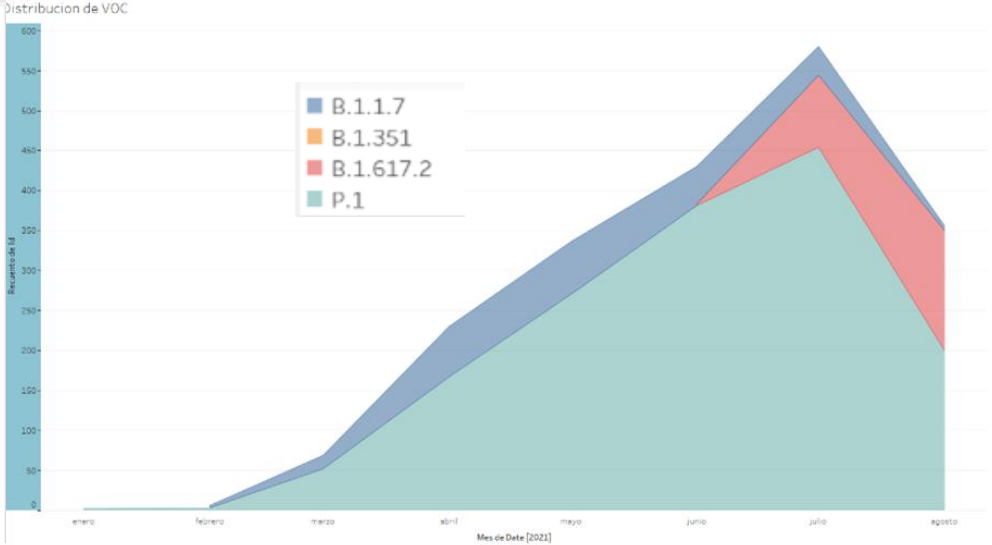
ANLIS MALBRÁN
ADMINISTRACIÓN NACIONAL DE LABORATORIOS E INSTITUTOS DE SALUD "DR. CARLOS G. MALBRÁN"

Distribución de variantes en el tiempo

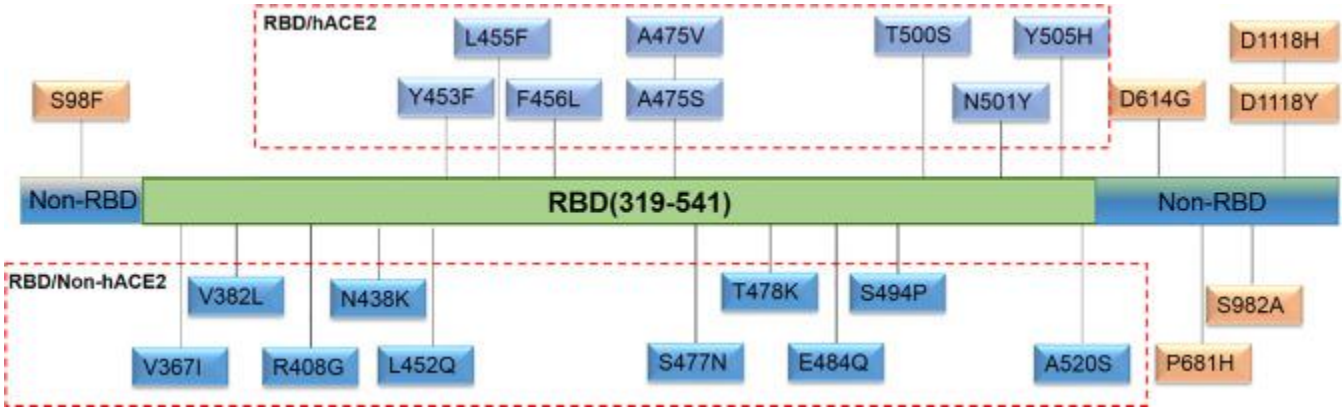
Distribucion de VOC, VOI y Lambda (C.37)



Distribucion de VOCs



Constelaciones y variantes (VOC, VOI….)



Natural and high-frequency amino acid mutations of the spike protein selected in this study. Twenty-four single amino acid mutations were located in three regions. Eight mutations (Y453F, L455F, F456L, A475V, A475S, T500S, N501Y, and Y505H) were in the RBD and hACE2 interaction region (**RBD/hACE2**); 10 mutations (V367I, V382L, R408G, N438K, L452Q, S477N, T478K, E484Q, S494P, and A520S) were in the RBD region but no interaction with hACE2 (**RBD/Non-hACE2**); 6 mutations (S98F, D614G, P681H, S982A, D1118H, and D1118Y) were not in the RBD region (**Non-RBD**).

Ding C, He J, Zhang X, Jiang C, Sun Y, Zhang Y, Chen Q, He H, Li W, Xie J, Liu Z, Gao Y. Crucial Mutations of Spike Protein on SARS-CoV-2 Evolved to Variant Strains Escaping Neutralization of Convalescent Plasmas and RBD-Specific Monoclonal Antibodies. Front Immunol. 2021 Aug 17;12:693775. doi: 10.3389/fimmu.2021.693775. PMID: 34484190; PMCID: PMC8416052.

Constelaciones y variantes (VOC, VOI….)

S:N501

This mutation is in the receptor binding domain (RDB), important to ACE2 binding and antibody recognition.

- May be associated with adaptation to rodents and mustelids: S:N501T in ferrets (Richard et al., Nature Comm.), in mink (Welkers et al., Virus Evolution), S:N501Y and in mice (Gu et al. Science).
    - Some have speculated on the risk of a possible persistent reservoir in wild rodents/mustelids
- May increase ACE2 binding (Bloom Lab ACE2 binding website , Nelson et al., bioRxiv) - in particular it is predicted to do this by increasing the time spent in the 'open' conformation (Teruel et al., bioRxiv)
- ACE2 binding may be further increased by the presence of S:E484K, and stabilized by the presence of S:K417N (Nelson et al., bioRxiv)
- S:N501Y was found in longitudinally-collected samples from an immunocompromised patient (Choi et al., NEJM)
- In one study, sera from previously infected patients neutralised viruses with S:501N and S:501Y equally (Xie et al., bioRxiv)
- Tests in people vaccinated with the Moderna and Pfizer-BioNTech vaccines suggest S:N501Y and S:E484K individually, and both together in combination with S:K417N, cause a small but significant reduction in neutralization (Wang et al., Nature)
- *In vitro* evolution to select for greater ACE2 binding resulted in mutations S:N501Y, S:E484K and S:S477N to be among the first selected (Zahradnik et al., bioRxiv).

Constelaciones y variantes (VOC, VOI….)

S:E484

This mutation is in the receptor binding domain (RBD), important to ACE2 binding and antibody recognition.

- Mutations at S:E484 may significantly reduce convalescent serum neutralization (Greaney et al., Cell Host & Microbe)
- There has been a case of reinfection associated with S:E484K: a woman previously infected with a non-S:E484K variant of SARS-CoV-2 was later reinfected with a virus carrying the S:E484K mutation (Nonaka et al., EID)
- In another study co-incubating pseudotyped virus with SARS-CoV-2 spike proteins and monocolonal antibodies, neutralization both by monoclonal antibodies and to convalescent sera was significantly reduced in viruses with S:E484 mutations (Liu et al., Cell Host & Microbe)
- May increase ACE2 binding, which may be further increased by the presence of S:N501Y, and stabilized by the presence of S:K417N (Nelson et al., bioRxiv)
- Tests in people vaccinated with the Moderna and Pfizer-BioNTech vaccines suggest S:E484K and S:N501Y individually, and both together in combination with S:K417N, cause a small but significant reduction in neutralization (Wang et al., Nature)

S:L452

- It was reported that the L452R mutation is associated with immune escape and could result in a stronger cell attachment of the virus, with both factors likely increasing viral transmissibility, infectivity and pathogenicity. (Tchesnokova, et al. Journal of clinical microbiology)
- L452Q increases spike protein affinity for ACE2, The neutralization resistance was attributed to the L452Q mutation. (Takuya Tada, et al. bioRxiv)

Constelaciones y variantes (VOC, VOI….)

S:H69-

S:H69- has arisen 3 times in association with other mutations and variants: MutationS:Y453F, Variant20A/S:439K, Variant20I(Alpha, V1) (Kemp et al. bioRxiv); and has additionally arisen multiple times outside of these variants.

- S:H69- **may alter the recognition by antibodies**, possibly impacting some antibody-therapy treatments, or immunity (Kemp et al. - Nature).
- In particular, the deletion is predicted structurally to 'tuck in' the Spike N-terminal domain)(Kemp et al. bioRxiv)
- In one study, was identified as a 'recurrent deletion region' (found multiple times in public sequences), but did not impact the 2 monoclonal antibodies tested (McCarthey et al., Science)
- Appeared in a chronically infected immunosuppressed patient treated with rituximab monoclonal antibodies (along with VariantS:Y453F) (Bazykin et al., Virological)

S:P681H

S:P681H is present in Variant20I(Alpha, V1), Variant21B(Kappa), and Variant21A(Delta) as well as some other circulating variants.

- S:P681H may reduce class 3 antibody recognition (Haynes et al., medRxiv)
- **This mutation is near the furin cleavage site, which may be important for immune recognition** (Johnson et al., bioRxiv

S:K417N

S:K417N is present in Variant20H(Beta, V2) and Variant20J(Gamma, V3).

- Mutations in S:K417 **may escape antibody binding** (Starr et al., bioRxiv , Wang et al., bioRxiv)
- May decrease binding to ACE2 receptor (Starr et al., Cell], Thomson et al, Cell , Tian et al., bioRxiv)

Bibliografía y webs utilizadas

https://nextstrain.org/
https://clades.nextstrain.org/

https://pangolin.cog-uk.io/

https://microreact.org/

https://outbreak.info/

https://covariants.org/

Boletín epidemiológico semanal N° 197 |gerencia operativa de epidemiología | ministerio de salud de la CABA
https://www.buenosaires.gob.ar/sites/gcaba/files/bes_197_se_20_vf_0.pdf

Integración de la vigilancia genómica de sars-cov-2 a la vigilancia de covid-19 a través del sistema nacional de vigilancia de la salud. versión 1. 04/2021
https://bancos.salud.gob.ar/sites/default/files/2021-04/SNVS_integracion-de-la-vigilancia-genomica_de_SARS-CoV-2.pdf

Informes vigilancia genómica. Ministerio de Salud. SE 34
https://www.argentina.gob.ar/salud/coronavirus-COVID-19/informacion-epidemiologica/agosto-2021

Situación actual nuevas variantes sars-cov-2. informe técnico. mayo 2021
https://www.argentina.gob.ar/sites/default/files/informe-vigilancia-genomica-mayo-direpi-anlis.pdf

Reporte epidemiologico semanal de WHO
https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---7-september-2021

Referencias a los protocolos utilizados

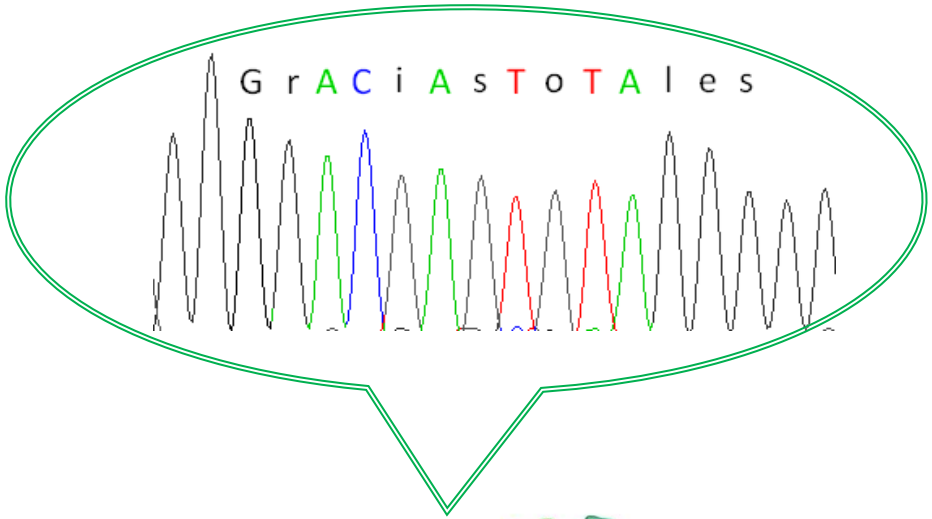
Castillo, AE, Parra, B, Tapia, P, et al. Phylogenetic analysis of the first four SARS-CoV-2 cases in Chile. *J Med Virol*. 2020; 92: 1562– 1566. https://doi.org/10.1002/jmv.25797

From total RNA extraction we performed the first amplification round using SuperScript III One-Step RT-PCR Platinum Taq DNA Polymerase (Invitrogen) and six pair of specific primers to obtain six complementary DNA fragments around 5 Kbp each, followed by a second amplification round with 24 specific primers (Table S1) to generate two fragments from each first round products, each subfragment (a total of 12) are around 2.3 to 2.7 Kbp.

https://artic.network/ncov-2019

Why do you sequence short fragments?

We have standardised on short amplicons. We do this because short amplicons are more likely to amplify with degraded RNA. Additionally, if regions fail to amplify in the scheme then less genome coverage is lost than it would be with a long amplicon. In an ideal world, amplicons would be as long as possible and it is often possible to sequence products longer than the 400 bp we typically use for intact RNA (we have tested up to 2kb with Ebola in the past). Aside from RNA fragment quality, PCR is less likely to work at very high fragment lengths.


GrACiAsToTAles