

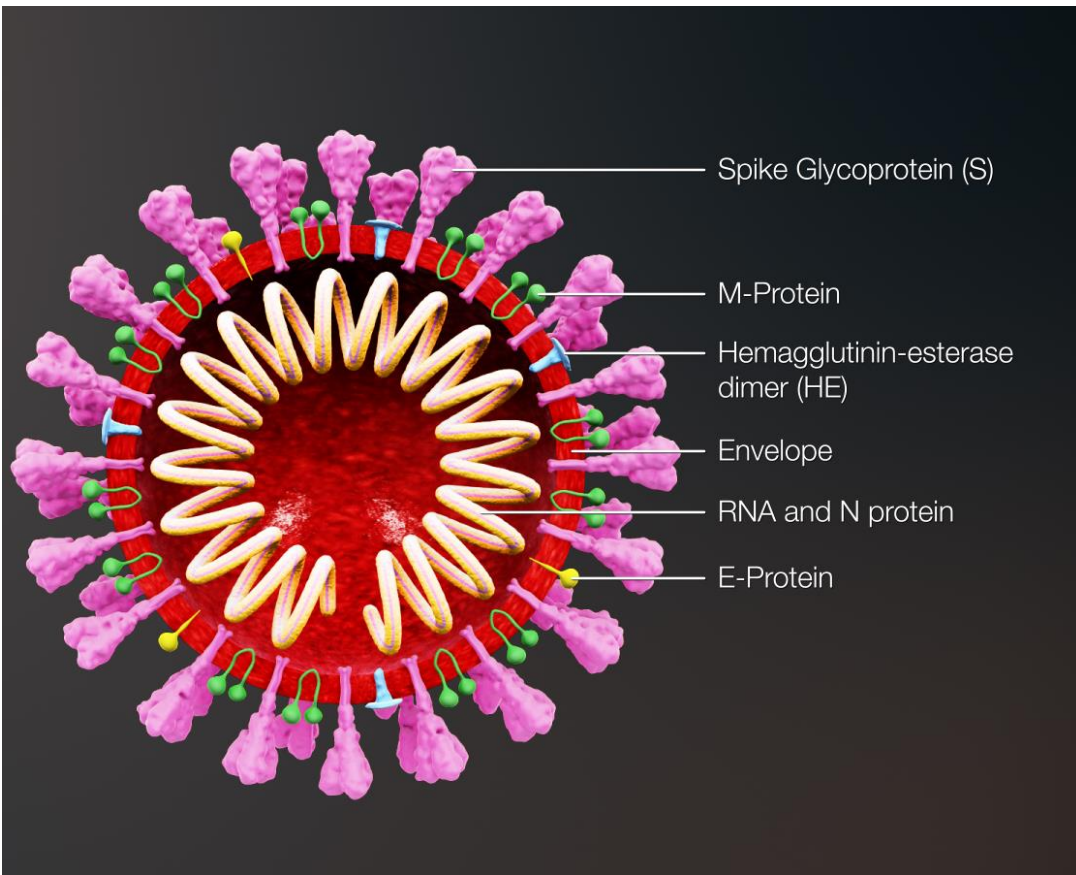
**PREPAREDNESS E
SORVEGLIANZA GENOMICA
NELLE PANDEMIE
IMPORTANZA DELLE RETI
DI COMPETENZE**

Sorveglianza genomica ai tempi del COVID

Gianguglielmo Zehender
Cattedra di Igiene del
Dipartimento di Scienze
Biomediche e Cliniche-Università
di Milano

Genoma e sequenziamento

2



Il **genoma** è il **materiale genetico** di un organismo. Nel caso di virus a RNA, come il Coronavirus, è composto da RNA.

Il sequenziamento dell'intero genoma è il **processo usato per ricavare la sequenza di un genoma**.

Una sequenza genomica è quindi una successione ordinata delle basi azotate presenti nel genoma.

Nel caso del Coronavirus, **la dimensione genomica è pari a circa 30k basi**

(credit: <https://www.scientificanimations.com/wiki-images/>).

I virus «mutano»

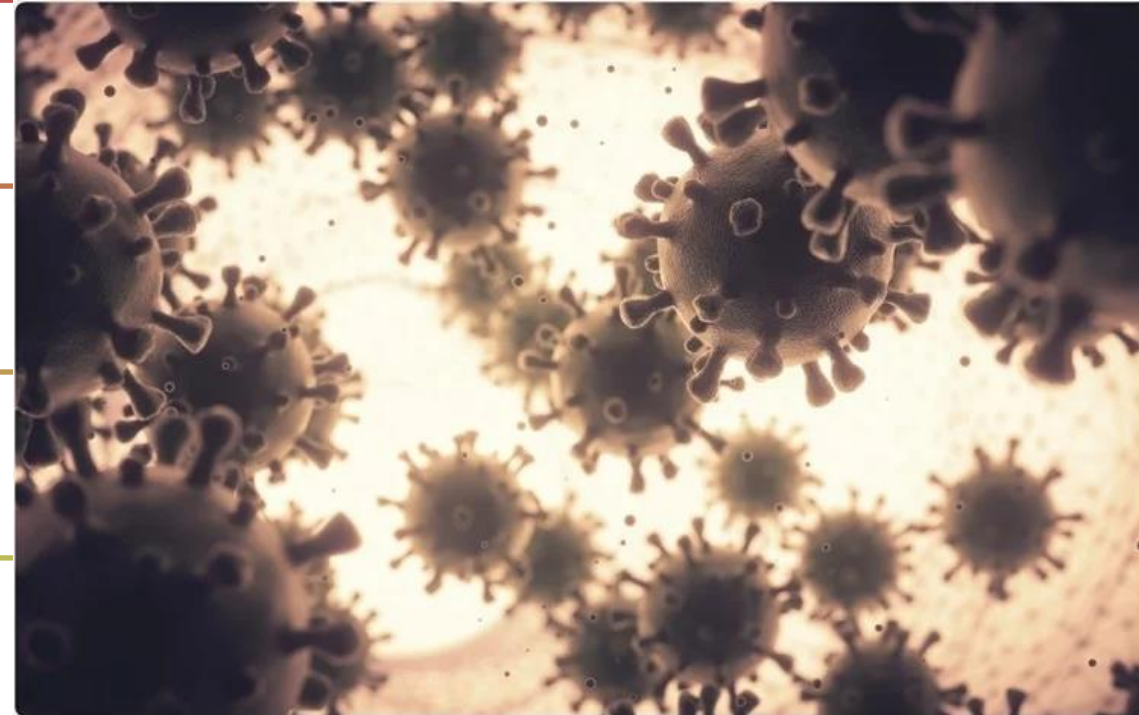
Tutti i virus accumulano mutazioni nei geni durante la replicazione a causa di meccanismi di duplicazione degli acidi nucleici che commettono errori di copiatura

La velocità della replicazione virale (in 24 ore si può avere il cambiamento del 50-90% della popolazione virale in un singolo ospite)

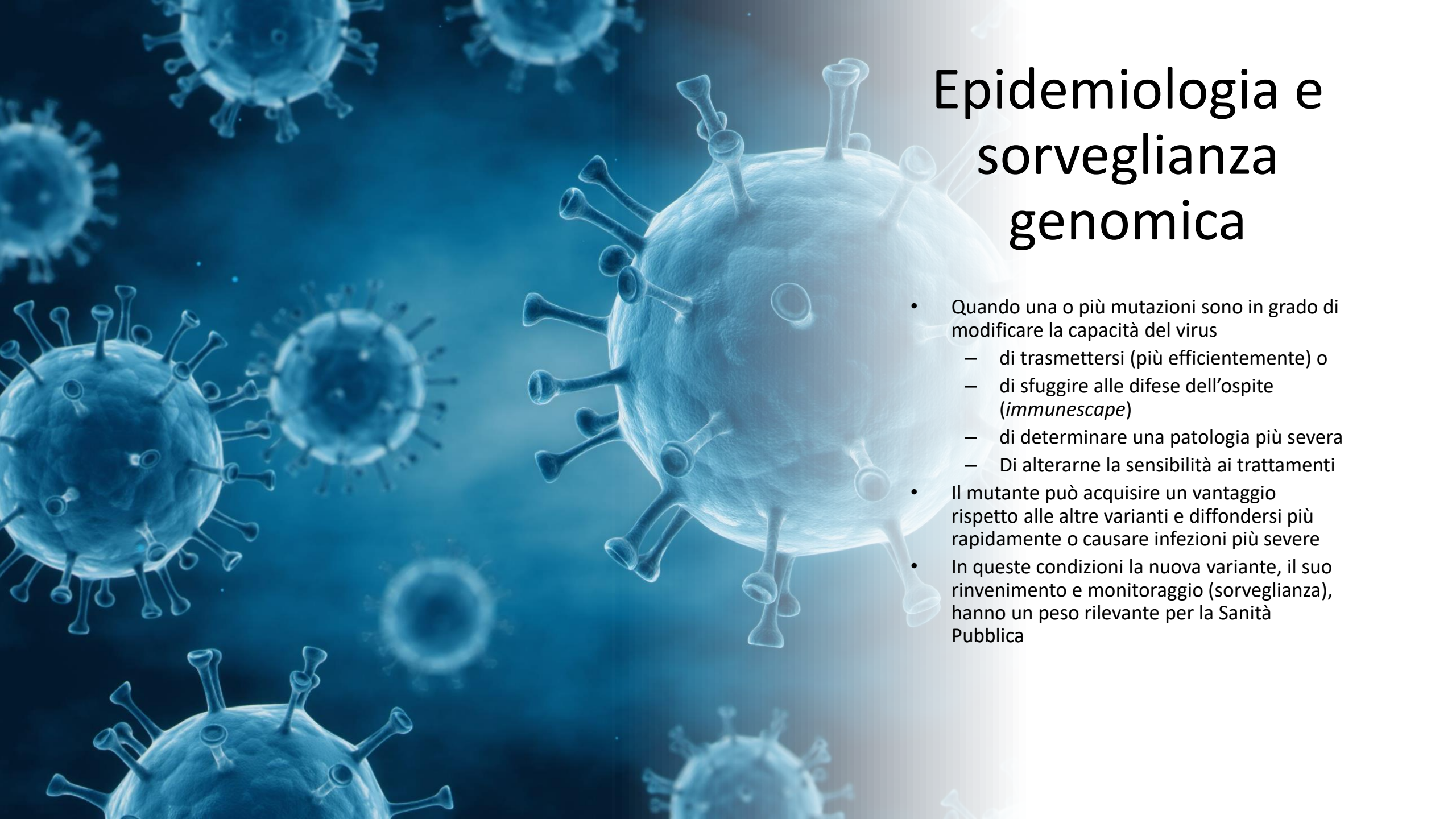
La intensità della replicazione (entità della progenie >100,000 volte per RNA virus)

Giustificano una variabilità che è di almeno 4 o 5 logaritmi superiore a quella dell'ospite umano

Molte mutazioni non hanno alcun impatto sulle caratteristiche fenotipiche del virus



<https://www.news-medical.net/news/20200812/SARS-CoV-2-a-swarm-of-mutant-viruses.aspx>

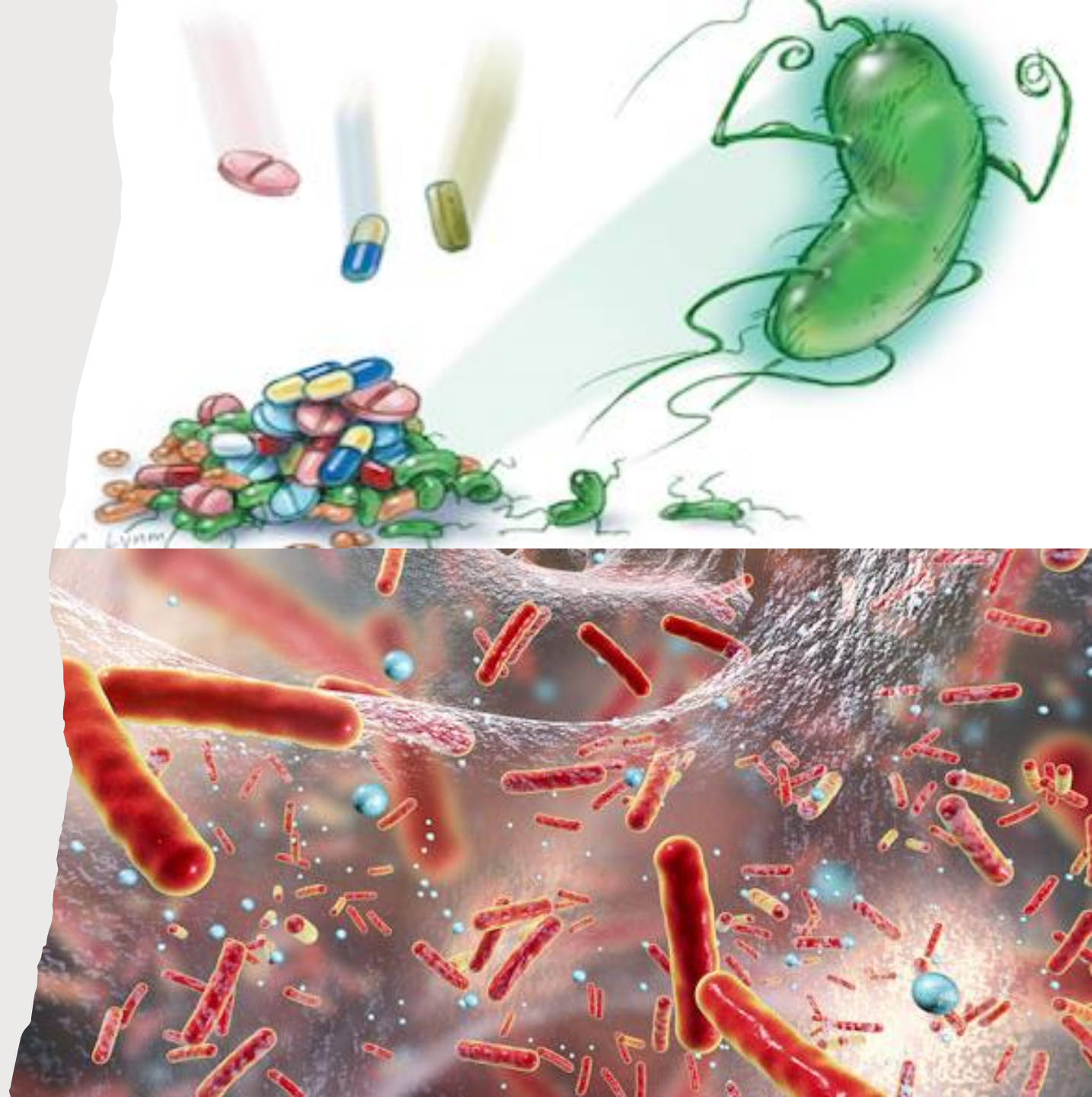


Epidemiologia e sorveglianza genomica

- Quando una o più mutazioni sono in grado di modificare la capacità del virus
 - di trasmettersi (più efficientemente) o
 - di sfuggire alle difese dell'ospite (*immunescape*)
 - di determinare una patologia più severa
 - Di alterarne la sensibilità ai trattamenti
- Il mutante può acquisire un vantaggio rispetto alle altre varianti e diffondersi più rapidamente o causare infezioni più severe
- In queste condizioni la nuova variante, il suo rinvenimento e monitoraggio (sorveglianza), hanno un peso rilevante per la Sanità Pubblica

Anche i batteri «mutano»

- Anche i batteri, pur avendo un tasso di mutazione più basso della maggior parte dei virus, sono soggetti a variazione anche attraverso meccanismi genetici più complessi
- La caratterizzazione del genoma batterico permette:
 - Individuare e delimitare i focolai epidemici in comunità o in ambito nosocomiale (assistenziale)
 - Identificare i ceppi resistenti ai farmaci, in particolare i ceppi multi-resistenti (MDR)
 - Identificare i fattori di virulenza e i ceppi dotati di aumentata patogenicità



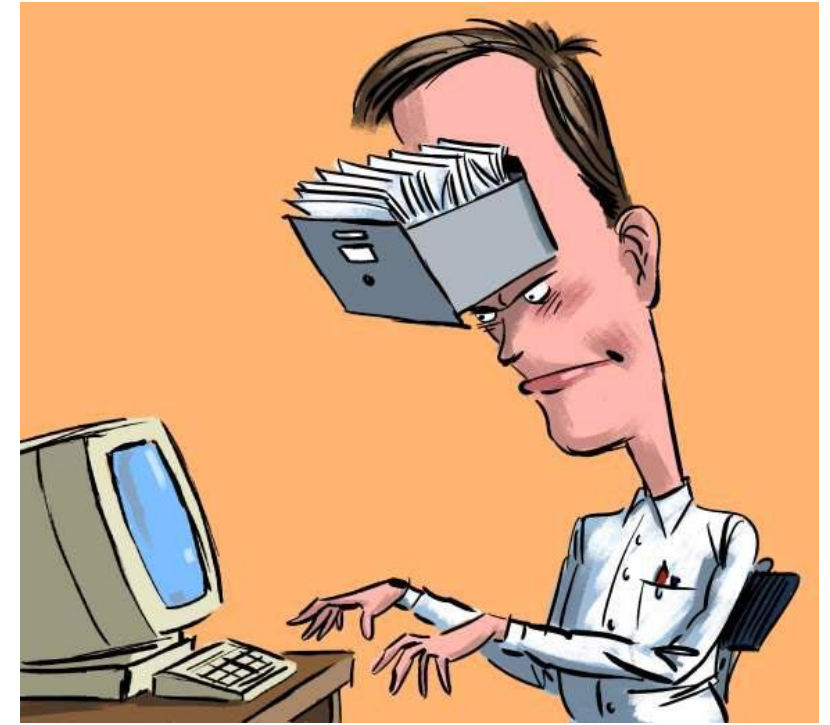
Next generation sequencing (NGS)



- Fino a 10 anni fa il sequenziamento completo di 3-4 milioni di nucleotidi di un genoma batterico richiedeva mesi e migliaia di euro per essere completato
- Le piattaforme di *Next Generation Sequencing* (NGS) oggi consentono il sequenziamento di decine di genomi batterici (centinaia di genomi virali) al giorno a un costo ridotto

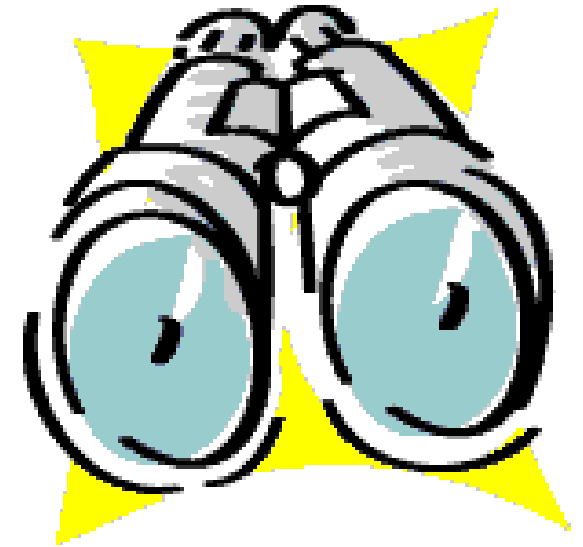
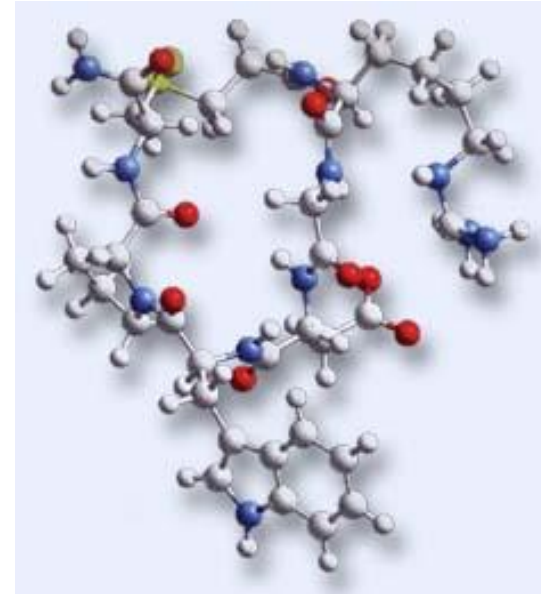
Big-data

- Questo approccio produce una gran quantità di dati complessi, difficili da analizzare (*Big data*)
 - Molto spesso occorrono una serie di analisi in cui il prodotto di una analisi rappresenta l'input dell'analisi successiva, fino all'ottenimento del dato necessario (*pipelines*)
- Da qui i laboratori di microbiologia e di Sanità Pubblica, devono dotarsi di competenze bioinformatiche e di strumenti informatici avanzati
 - Per l'archiviazione dei dati e l'interrogazione (intelligente) dei database (Bioinformatica) e per l'analisi degli stessi dati



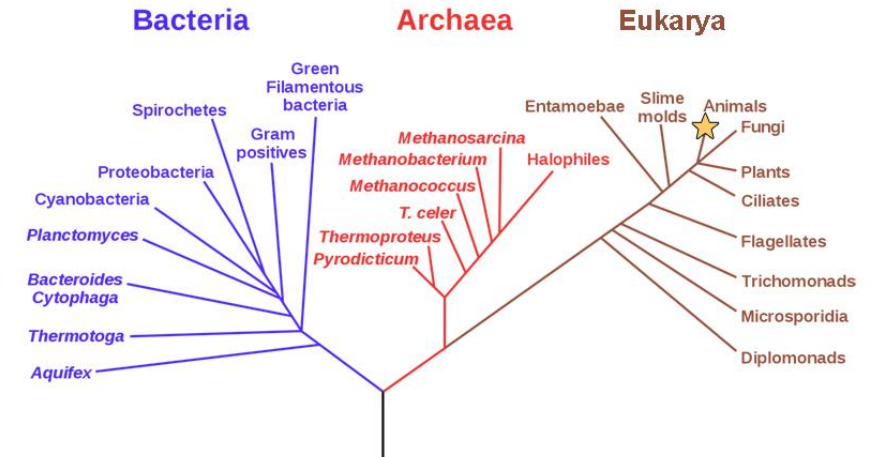
BIOINFORMATICA

- Uso dell'approccio informatico (computer) per risolvere problemi biologici
- Nata alla fine degli anni 70 per l'interrogazione delle banche di sequenze (proteine e acidi nucleici)
- Successivamente ha sviluppato competenze nel campo dell'analisi delle stesse:
- Per esempio:
 - Identificazione della funzione di geni
 - Studio della struttura 3D (molecular modeling)
 - Confronto tra organismi simili (filogenesi)



Phylogenetic Tree of Life

★ = You are here



Advanced Molecular detection

- Il riconoscimento di queste problematiche e dei vantaggi dell'uso dell'NGS nell'identificazione microbica ha spinto il CDC ad avviare l'iniziativa detta Advanced Molecular Detection (AMD)
- Finalizzata a integrare i nuovi strumenti di NGS e bioinformatici con le metodiche epidemiologiche nel campo delle malattie infettive



U.S. Department of
Health and Human Services
Centers for Disease
Control and Prevention

For more information visit www.cdc.gov/amd



IDENTIFYING EMERGING PATHOGENS **Zika. Ebola. Elizabethkingia.**

Studying the genetic makeup of microorganisms helps us identify and track rare but deadly pathogens, such as Ebola and Zika viruses. AMD technologies have uncovered previously unknown threats, including Bourbon virus.



IMPROVING VACCINES **Influenza. Pertussis. Pneumococcal disease.**

Applying AMD to vaccine-preventable diseases, such as pertussis and influenza, helps us monitor genetic changes and understand why vaccine effectiveness may decrease. Targeting evolving pathogens with AMD, we can develop more effective vaccines.



MAKING FOOD SAFER **Listeria. Salmonella. E. coli.**

State-of-the-art AMD methods help solve foodborne outbreaks faster by linking food sources to clusters of illness. By expanding these methods nationwide, we can get contaminated products off store shelves and out of people's homes sooner to save more lives.



DEVELOPING FASTER TESTS **Meningitis. AR Gonorrhea. EV-D68.**

AMD leads to more rapid and accurate tests to diagnose respiratory infections and to detect antimicrobial resistance (AR). Reference genomic databases built through AMD are speeding up diagnosis and public health response.



CONNECTING THE DOTS **HIV. Hepatitis. TB.**

AMD connects information from public health investigations with genomic data from pathogens to help us better understand how an infection is spreading. By mapping the transmission sooner, we can target resources where they will have the most impact.

AMD
ADVANCED
MOLECULAR
DETECTION

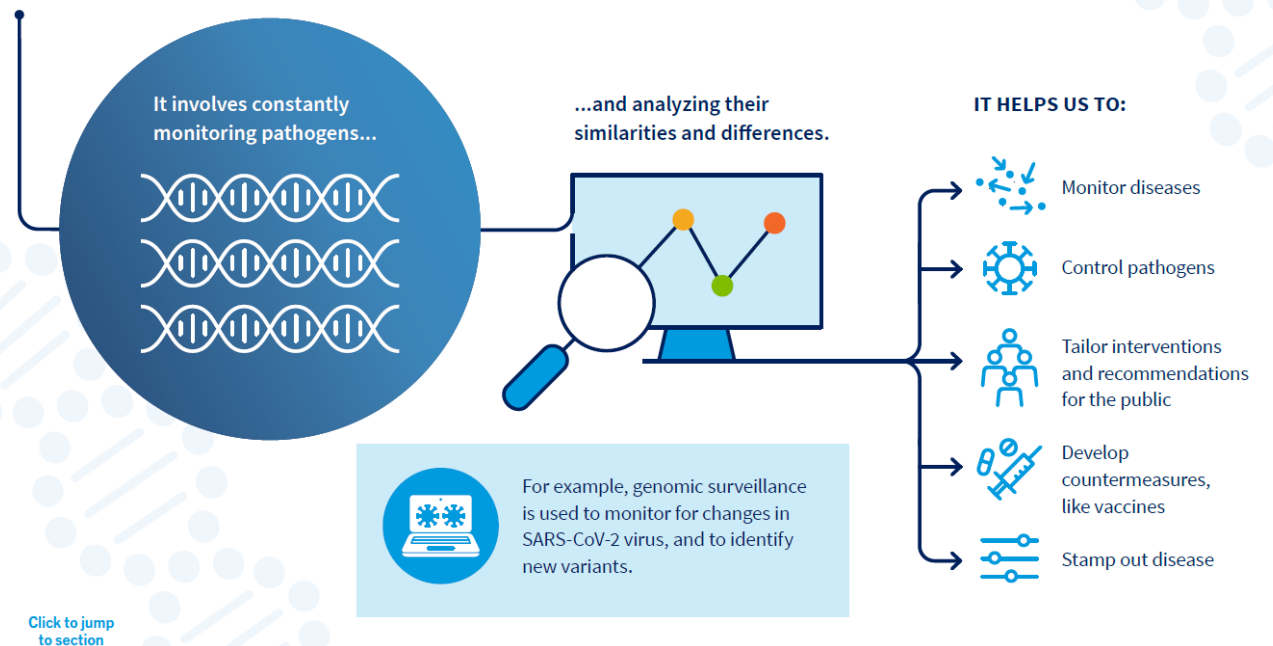
AMD harnesses the power of next-generation genomic sequencing, high-performance computing, and epidemiology to study pathogens.

La Sorveglianza Genomica

- Sorveglianza Epidemiologica:
 - consiste in una attività di continuo monitoraggio e vigilanza sulla distribuzione e diffusione nella popolazione di malattie/infezioni, che consenta la prevenzione, il controllo e la rapida individuazione delle modificazioni nelle caratteristiche clinico-epidemiologiche delle stesse
- Genomica:
 - Caratterizzazione e analisi di geni/genomi interi di patogeni attraverso approcci di sequenziamento ad alta resa

What is genomic surveillance?

Genomic surveillance is monitoring the genetic make-up of pathogens, such as viruses and bacteria, to understand how they behave and how we can control them.



«Chi deve fare cosa...» in caso di emergenza

Ha lo scopo di ridurre i rischi dovuti alle malattie trasmissibili e mitigare l'impatto in emergenze

E' un processo continuo che coinvolge molteplici settori sociali (non solo la sanità pubblica)

Richiede **coordinamento e comunicazione**.

Pianificazione di misure comuni

Determinazione dei **livelli di rischio** e delle **priorità** di intervento

Valutazione post-emergenza

Supporto a **RICERCA, FORMAZIONE e INFORMAZIONE**



La *preparedness*

What are the Strategy's goals and objectives?



WHO-Global genomic surveillance strategy

L'Organizzazione Mondiale della Sanità ha inaugurato nel giugno 2021 un piano per promuovere una **strategia comune globale** per la sorveglianza genomica

Coinvolgendo esperti nel settore non solo di COVID-19 ma anche di Polio, Ebola, HIV, ceppi microbici resistenti

Obiettivo generale: potenziare la sorveglianza genomica a livello locale e globale

Sorveglianza genomica e COVID-19

- La pandemia da COVID-19 ha mostrato fin dall'inizio l'importanza della sorveglianza genomica
- Nel monitoraggio delle varianti (di interesse o di preoccupazione) di SARS-CoV-2 in grado di influenzare la diffusione nella popolazione e l'impatto sulla clinica nonché la capacità di fuga dalle difese indotte da vaccino o da anticorpi monoclonali

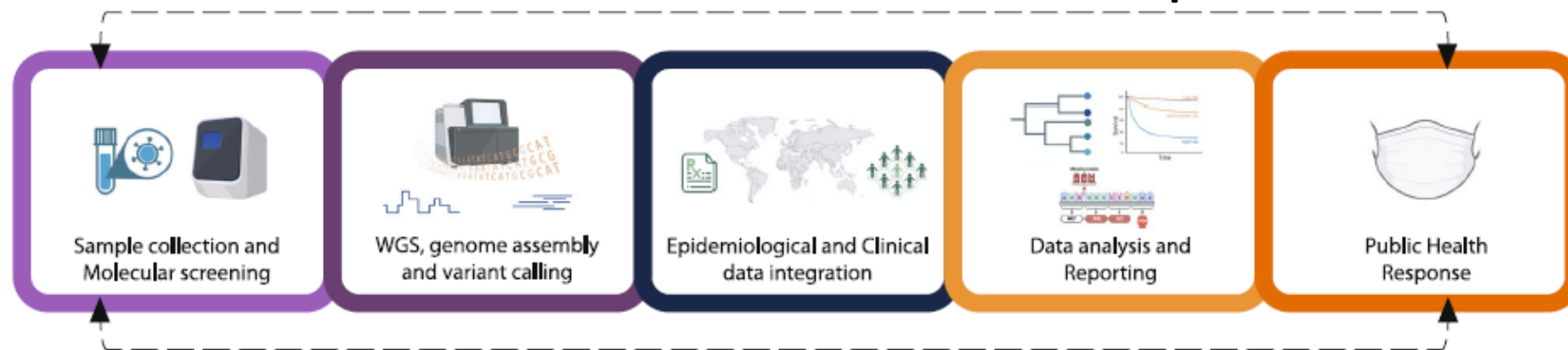
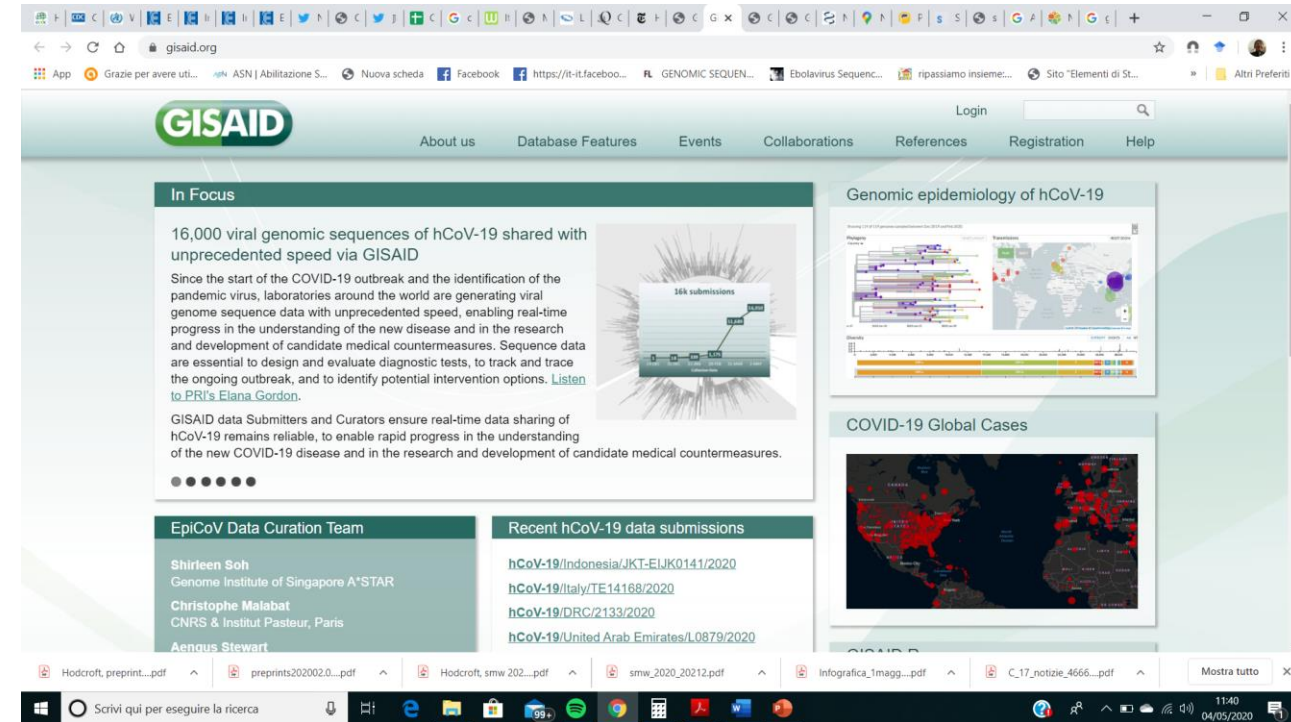


FIGURE 2 Genomic surveillance framework for the real-time monitoring of viral pathogens. The framework consists of five essential steps.

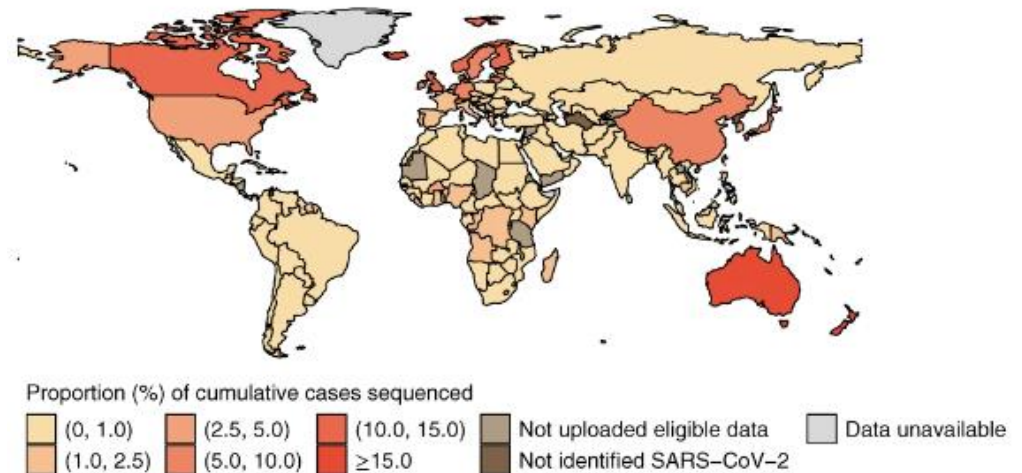
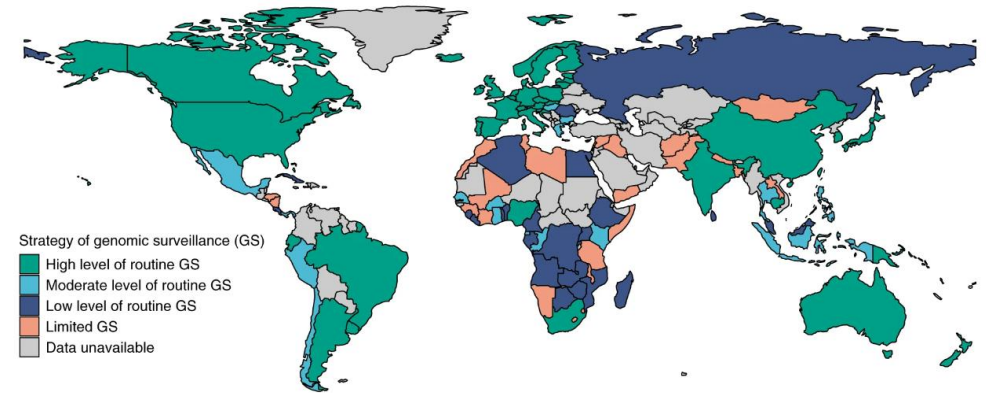
Banche dati di SARS-CoV-2: GISAID

- La prima sequenza di genoma completo di SARS-CoV-2 depositata sulla piattaforma **Global Initiative on Sharing All Influenza Data (GISAID)** risale al 10 gennaio 2020
- Attualmente (febbraio 2024) sono più di 16 milioni e mezzo i genomi completi di SARS-COV-2 disponibili
- **Il database garantisce l'accesso libero ai dati in modo gratuito, sia per il download che per l'upload**
- L'elevata disponibilità di sequenze genomiche complete collezionate in tutto il mondo permette
 - lo studio dell'epidemiologia molecolare/genomica del virus nel mondo
 - L'identificazione di nuove varianti (sorveglianza molecolare)
 - Lo studio delle proprietà cliniche e di fuga dalle difese immuni o resistenza ai farmaci dei mutanti
 - Lo sviluppo di farmaci e/o vaccini



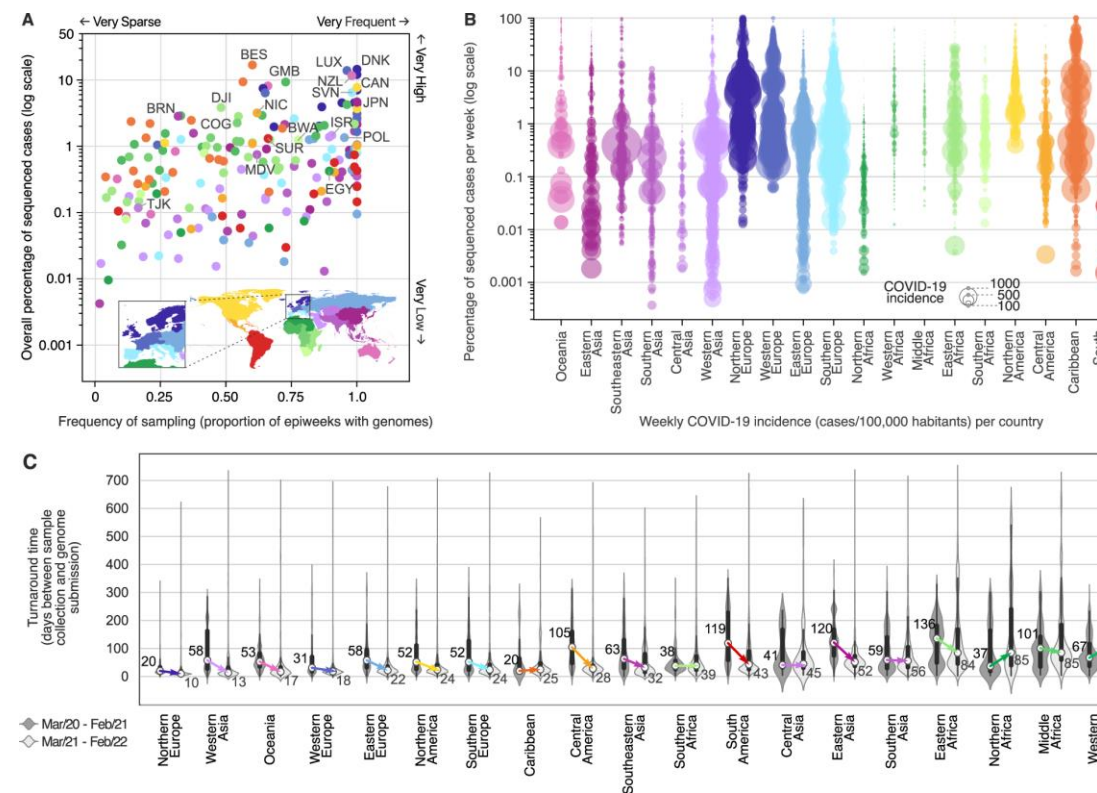
Capacità di sequenziamento globale al 12/2021

- Marcata eterogeneità geografica
- Globalmente il livello di sorveglianza genomica è
 - Alto nel 38.1% dei paesi (45)
 - Sconosciuto nel 64% (76)
- In particolare i paesi *high-income* (HIC) sequenziano molto di più (mediamente 4.36% dei casi) dei paesi *non-high-income* (lower-middle-income-LMIC) (0.27%)
- Ma l'*income* level non è l'unico fattore che limita il sequenziamento (diversa politica di sorveglianza, non condivisione dei dati, alta incidenza COVID-19)
- **Diversa capacità di sequenziamento può influenzare la pronta individuazione di una nuova variante**
 - *Zhiyuan Chen, Nature Genetics, 2022*



Diverso tempismo a livello globale

- Durante i primi 2 anni di pandemia: da Marzo 2020 a Febbraio 2022 (*Brito, Nat Comm 2022*)
- Il 25% dei genomi ottenuti da HICs sono stati pubblicati in banche dati <21 giorni dopo il campionamento (*Turnaround Time*)
- Solo il 5% dei genomi caratterizzati da Paesi a basso-medio reddito vengono pubblicati con la stessa rapidità
- La rapidità con cui vengono sequenziate e pubblicati i genomi influenza la capacità di individuare per tempo le varianti (settimane o mesi per diffondersi e divenire prevalenti)



Capacità di sequenziamento in Europa

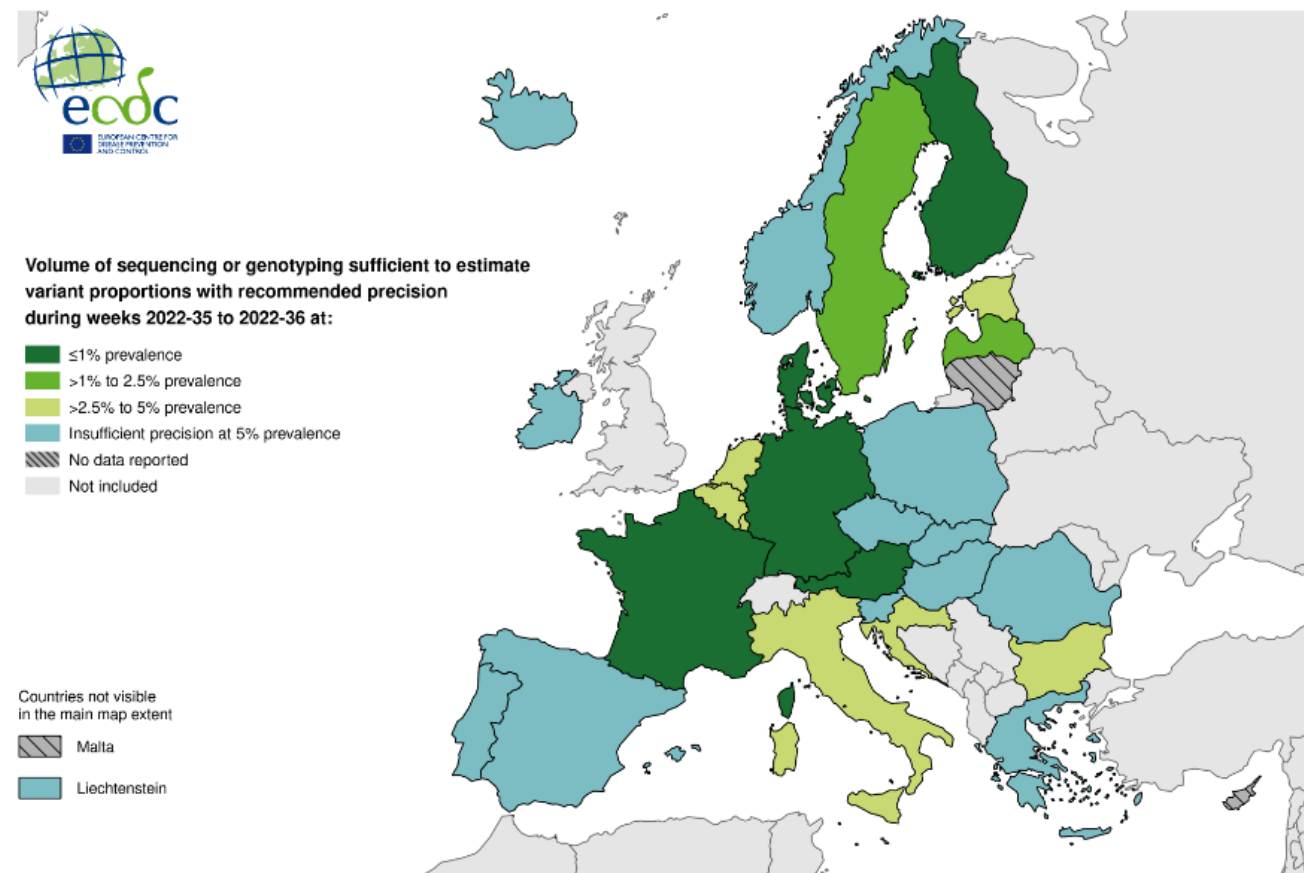
Dati ECDC delle settimane 35 e 36 del 2022

Proporzione di campioni SARS-CoV-2 sequenziati in 27 Paesi Europei

Media 10.0%

Minimo 0.4% (Grecia Italia)

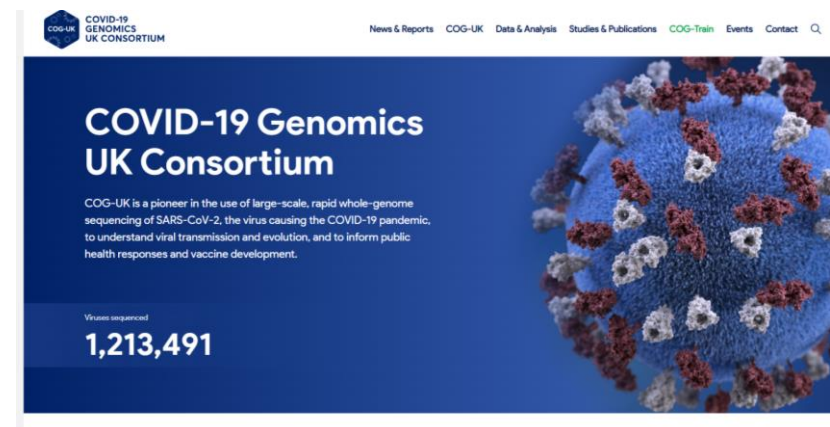
Max 62.4% (Danimarca)



Source: GISAID EpiCoV™ and ECDC TESSy data. Administration boundaries: © Eurographics
The boundaries and names shown on this map do not imply official endorsement or acceptance by the European Union. ECDC. Map produced on 22 September 2022

Capacità di sequenziamento in UK

- The COG-UK Consortium is an innovative partnership of NHS organisations, the four Public Health Agencies of the UK, the Wellcome Sanger Institute and more than 12 academic institutions providing sequencing and analysis capacity.
- It is led by Professor Sharon Peacock, Director of the COVID-19 Genomics UK (COG-UK) Consortium.
- It is supported by £20 million funding from the UK Department of Health and Social Care (DHSC), UK Research and Innovation (UKRI) and the Wellcome Sanger Institute.
- Il Consorzio è stato chiuso il 31 marzo 2023
- A marzo 2023 più di 3 milioni di sequenze su 15 milioni (1:3) depositate su GISAID, provenivano dal Regno Unito



Numero campioni sequenziati in Italia tra gennaio 2022 e gennaio 2023

Tra gennaio 2022 e gennaio 2023

Mediana c.ca 1,000 seq a settimana

Min 476

Max 3223

Percentuale positivi sequenziati

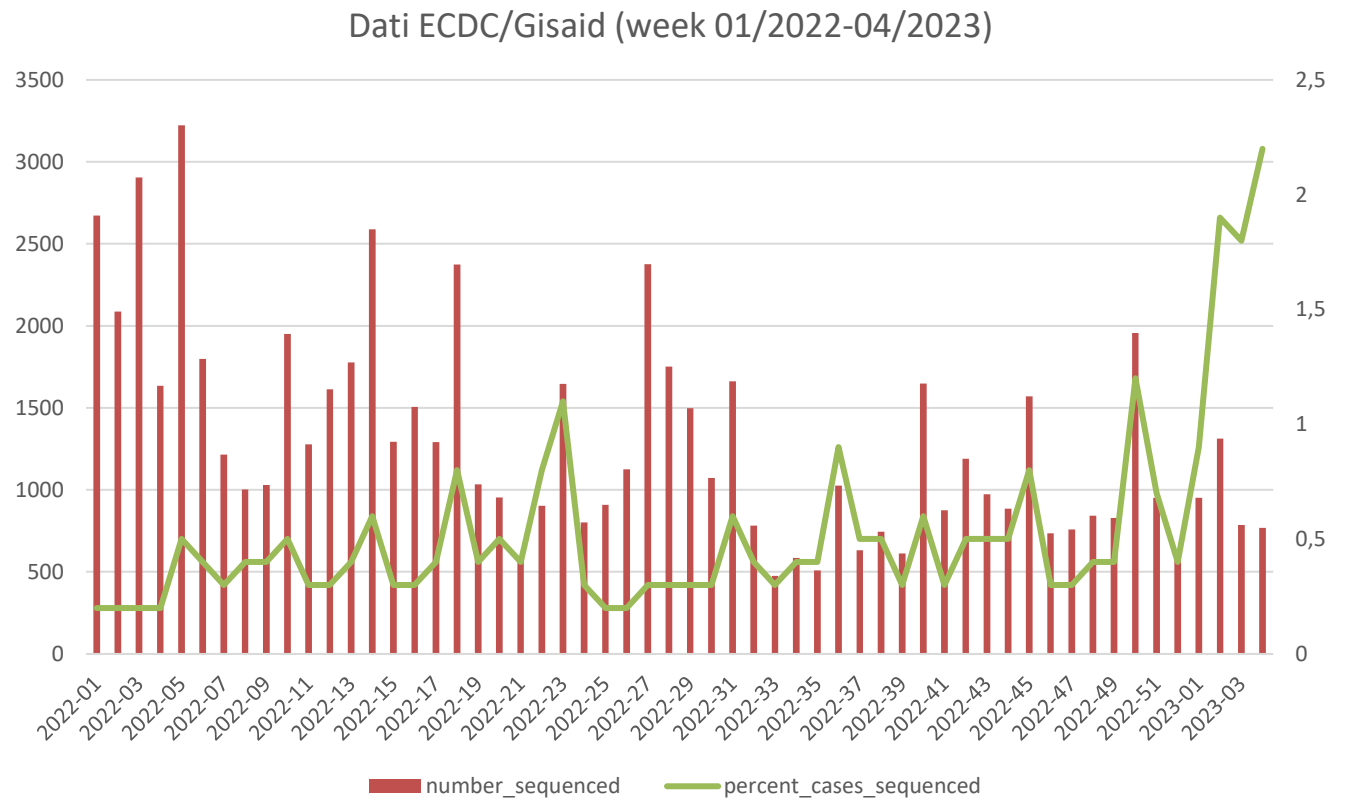
Mediana 0.4%

Min 0.2%

Max 2.2%

Tempi relativamente lunghi per il
consolidamento del dato?

Dopo gennaio 2023 il dato diventa irrealistico
a causa della riduzione del numero di tamponi



Intelligenza artificiale

- L'Intelligenza Artificiale rappresenta un'opportunità fondamentale per la gestione di una gran numero di dati (Big-data) genomici (ma anche epidemiologici) e l'uso di modelli (bioinformatici e statistici) sempre più complessi, in tempi brevi.
- Fondamentale nella sorveglianza genomica
 - Identificazione pronta di nuove varianti o emergenza di mutazioni di impatto
 - Comprensione dell'effetto delle mutazioni sulla clinica e sulla capacità di fuga dalle difese
 - Valutazione dei rischi di emergenza
- Disegno di nuovi farmaci e/o di nuovi vaccini
- Determinare la resistenza agli antibiotici nei batteri



Importanza delle reti di competenze

- Per fare tutto questo è fondamentale la costituzione di reti di competenze e l'interazione pubblico/privato
 - Biomedicina umana, animale e ambiente (approccio One Health)
 - Competenze bioinformatiche
 - Reti informatiche
 - Competenze nell'intelligenza artificiale
 - Infrastruttura informatica e gestione Dati

Gruppo di Lavoro



Research Team

- G. Zehender
- A. Lai

Artificial Intelligence Lab

- L. Ruggerone
- L. Li Puma
- S. Ronchiadin
- A. Miola

Ospedale Luigi Sacco

- P. Carrer
- M. Galli

Intesa Sanpaolo

- F. Rastrelli
- D. Russignaga
- A. Izzo

Intesa Sanpaolo Innovation Center

- M. Montagnese
- G. de Vecchi

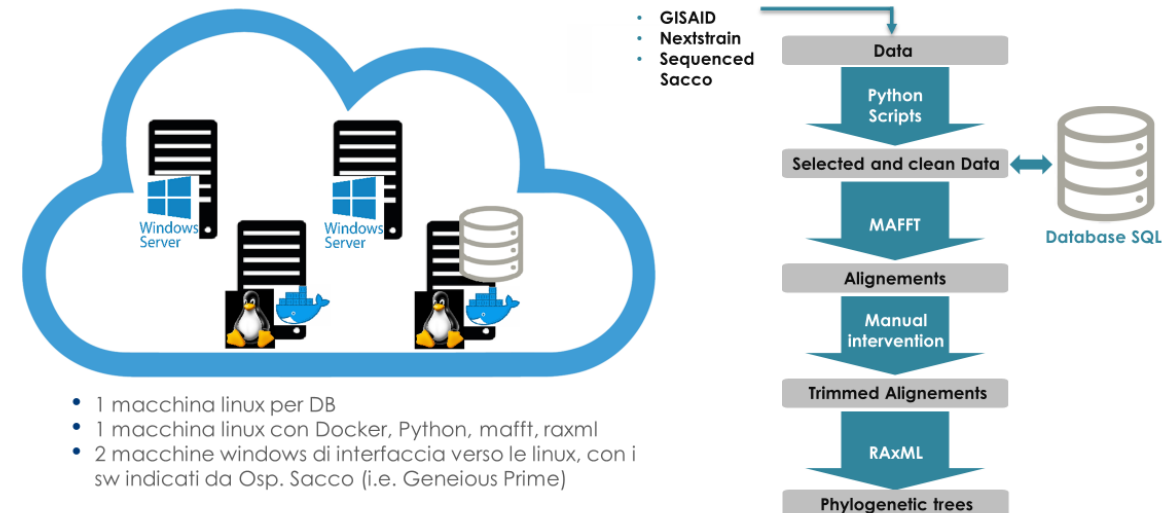
Tech Meeting
Every week

Progress Meeting
Every 2 weeks



Infrastruttura e software

Esigenza: Ospedale Sacco vorrebbe avere un'infrastruttura dedicata e più efficiente rispetto al loro attuale ambiente di analisi (laptop e servizi da remoto «best effort»)



FINE

Gracie

**PREPAREDNESS E
SORVEGLIANZA GENOMICA
NELLE PANDEMIE**
IMPORTANZA DELLE RETI
DI COMPETENZE

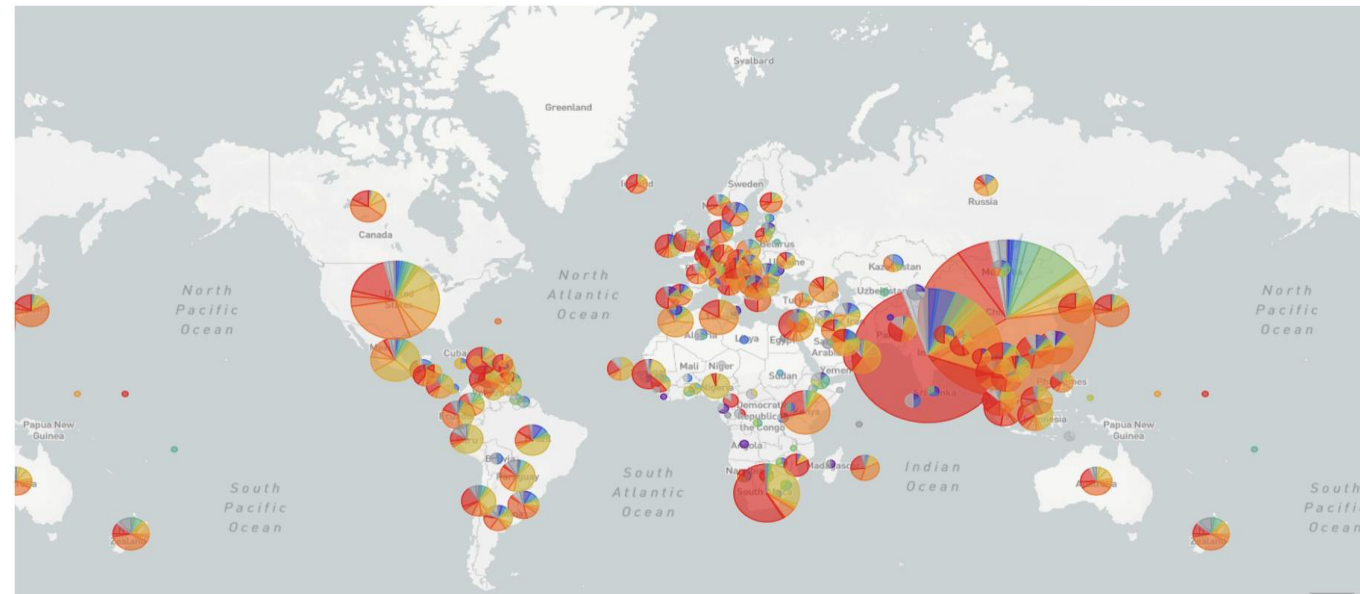
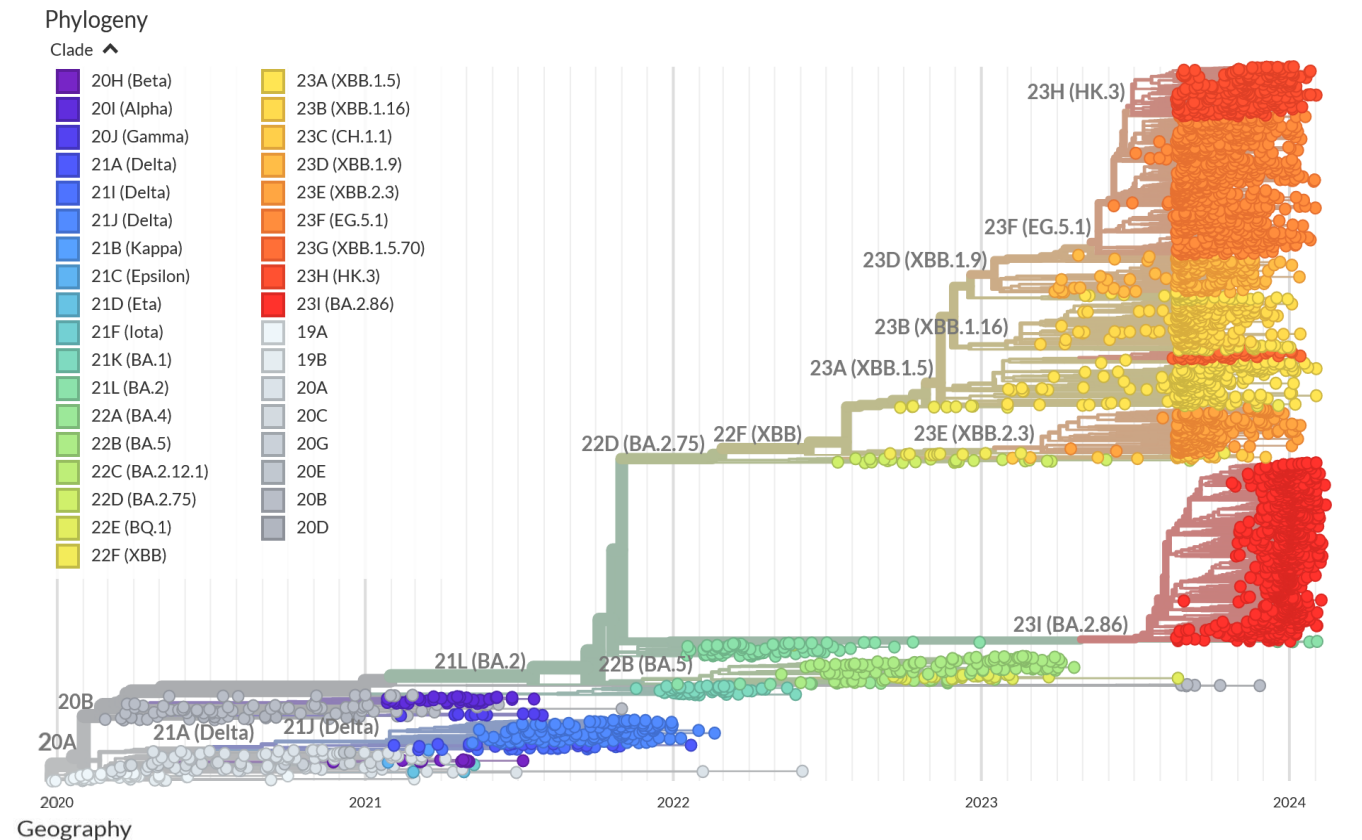
Il ruolo delle banche dati nello studio dell'epidemiologia genomica

Alessia Lai

Dipartimento di Scienze Biomediche e Cliniche

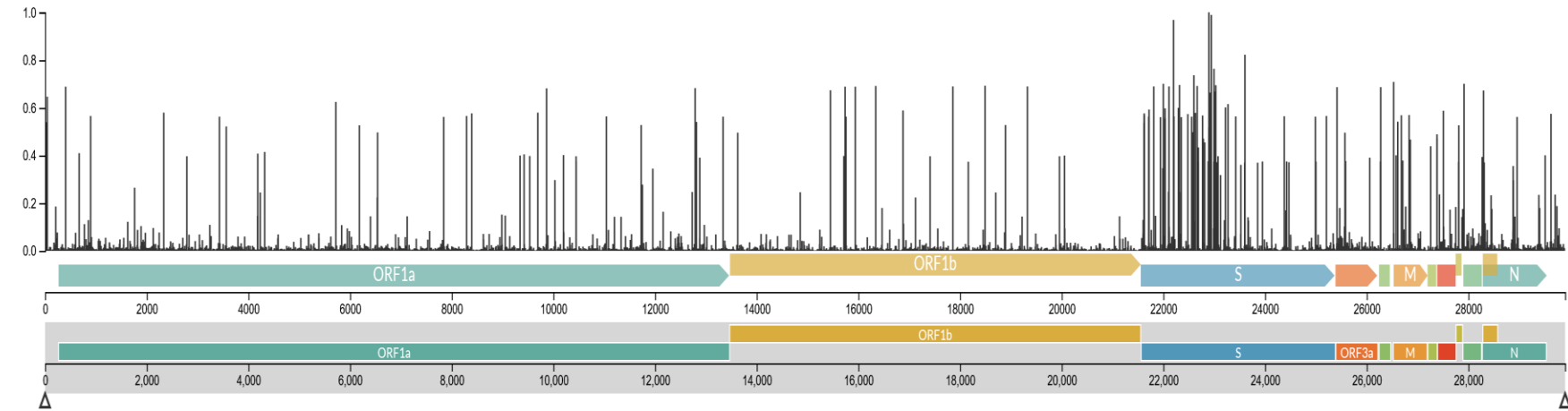
Università degli Studi di Milano

- ✓ Nextstrain è un progetto *open source* per sfruttare la disponibilità di genomi dei patogeni a beneficio della ricerca scientifica e della sanità pubblica
- ✓ Attinge i dati genomici da database pubblici, tra cui principalmente GISAID
- ✓ Costruisce un albero filogenetico datato
- ✓ Combina dati relativi al tempo, allo spazio (mappa) e alla classificazione dei genotipi
- ✓ Fornisce una visione costantemente aggiornata dei lignaggi di SARS-CoV-2 circolanti in tutto il mondo

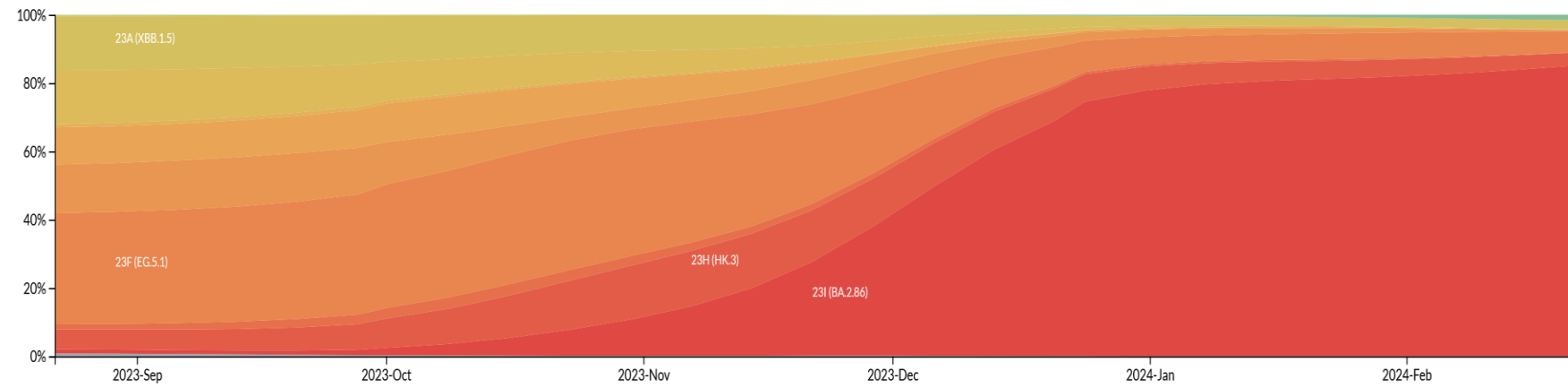


- ✓ Fornisce anche una visione completa delle mutazioni osservate e della loro frequenza nel genoma del virus
- ✓ Nonché un quadro aggiornato delle varianti (*clades*) circolanti negli ultimi mesi
- ✓ L'obiettivo è fornire uno strumento potente per la sorveglianza genomica
- ✓ Aiutare la comprensione epidemiologica/evolutiva
- ✓ Contribuire alla risposta contro le epidemie/pandemie

Nucleotide diversity of genome



Frequencies (colored by Clade)



Nextstrain SARS-CoV-2 resources



Clade assignment, mutation calling, and sequence quality checks

The image shows two side-by-side panels of the Nextclade web interface. The left panel, titled "Provide sequence data", has tabs for "File", "Link", "Text", and "Example". The "File" tab is active, showing a dashed box for dragging files, a FASTA file icon, and a "Select files" button. The right panel, titled "Select reference dataset", has a "Suggest automatically" toggle (checked), "Reset", and "Suggest" buttons. A "Select reference dataset" button is also visible at the bottom.

- [Latest global SARS-CoV-2 analysis \(GISAID data\)](#)

Jump to our globally-subsampled SARS-CoV-2 dataset which is updated daily using data from GISAID. We also maintain additional analyses to focus subsampling on different geographic regions and different time periods. These include analyses that focus on the previous 6 months for [Africa](#), [Asia](#), [Europe](#), [North America](#), [Oceania](#) and [South America](#), as well as analyses that focus on the entire pandemic for [Africa](#), [Asia](#), [Europe](#), [North America](#), [Oceania](#) and [South America](#).

- [Latest global SARS-CoV-2 analysis \(open data\)](#)

Jump to our globally-subsampled SARS-CoV-2 dataset which is updated daily using open data from GenBank. Additional analyses that focus subsampling on different geographic regions and different time periods include analyses that focus on the previous 6 months for [Africa](#), [Asia](#), [Europe](#), [North America](#), [Oceania](#) and [South America](#), as well as analyses that focus on the entire pandemic for [Africa](#), [Asia](#), [Europe](#), [North America](#), [Oceania](#) and [South America](#).

- [Nextclade \(sequence analysis webapp\)](#)

Drag and drop your sequences to assign them to clades, report potential sequence quality issues and view samples on a phylogenetic tree

- [CoVariants \(mutations and variants of interest\)](#)

An overview of SARS-CoV-2 variants, their prevalence around the world and constituent mutations

- [SARS-CoV-2 variant growth rates and frequency forecasts](#)

Historical changes in frequency of SARS-CoV-2 variants are used to estimate variant growth rates and forecast changes in frequency

- [How to run your own phylogenetic analysis of SARS-CoV-2](#)

A tutorial walking through running your own analysis using Nextstrain tools

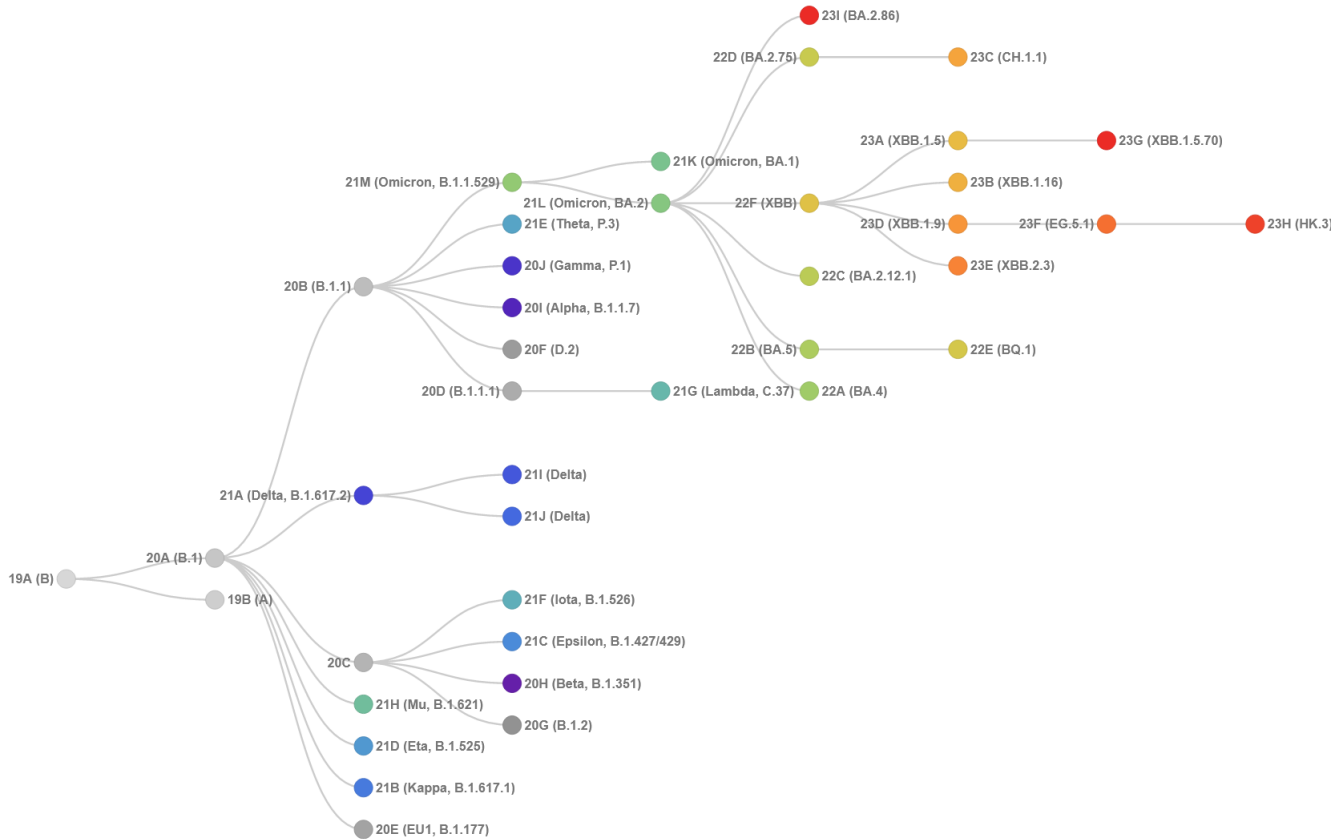
- [Scroll down to all available datasets](#)

- [Scroll down to all available interactive situation reports](#)

SARS-CoV-2 evoluzione

- SARS-CoV-2 ha un grande genoma (29 kb) ed evolve rapidamente, ma meno degli altri RNA virus
- Variabilità limitata dal possesso di una **ESONUCLEASI ad attività *proofreading***
- All'inizio (spillover) non aveva praticamente diversità genetica
- Durante l'espansione nell'uomo si è diversificato producendo numerosi lignaggi virali distinti
- **Ad una velocità evolutiva stimata tra a 0.4 e 2 sostituzioni ogni 1000 siti all'anno**
 - una frequenza di circa **1-2 mutazioni mese**
- **Molti lignaggi sono ora estinti e sostituiti da altri con caratteristiche diverse**
- Primo caso la variante **D614G**
- **Il tasso di trasmissione di SARS-CoV-2 è più elevato del suo tasso evolutivo**, quindi molti genomi identici si diffondono rapidamente e possono essere presenti ovunque!

SARS-CoV-2 clades (Nextstrain)

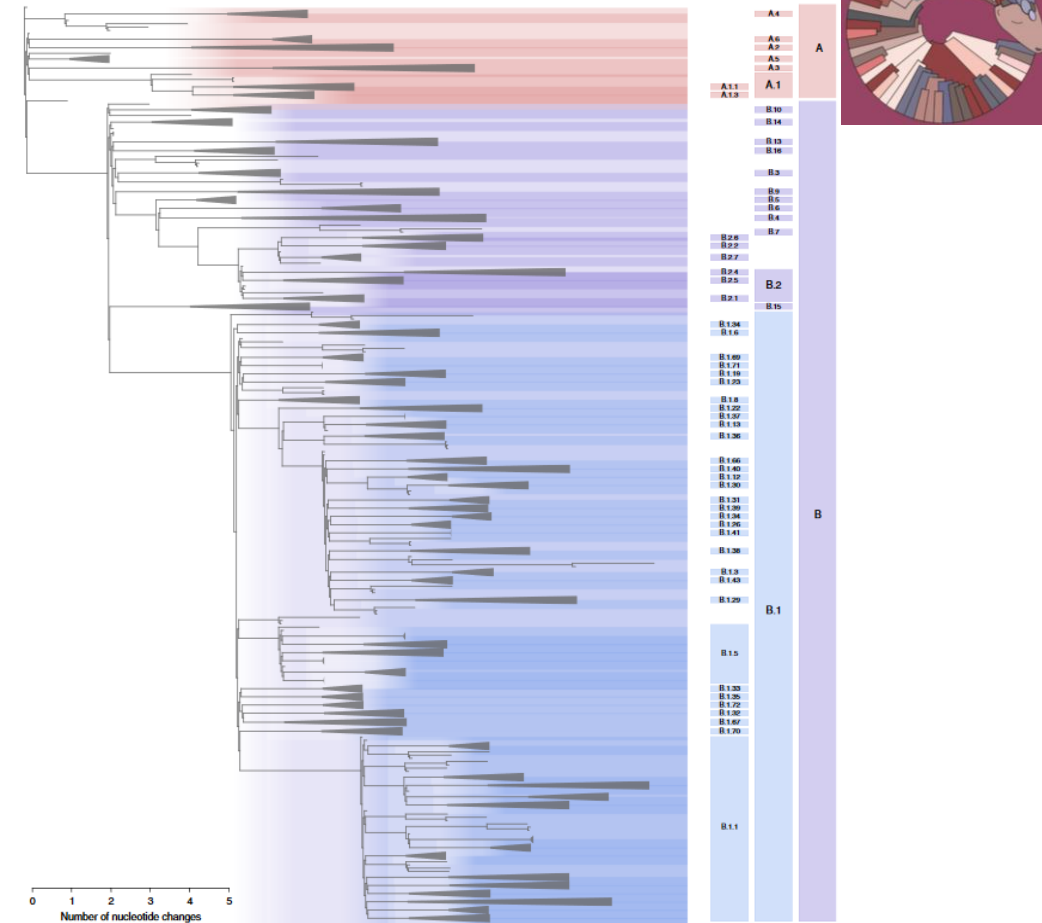


Assegna ad ogni clade una lettera sulla base della diversificazione dal progenitore e un numero corrispondente all'anno di comparsa.

- ✓ **19A e 19B** gli originari di Wuhan
- ✓ **20A** emerso in Europa in Marzo da 19A
- ✓ **20B e 20C** emersi successivamente nel 2020 dal 20A
- ✓ **20D-20I** emersi dopo l'estate 2020

<https://clades.nextstrain.org/>

SARS-CoV-2 lineages (Pangolin)



Radice dell'albero: **lineage A** (EPI_ISL_406801-il più vicino al pipistrello) e **B** (Wuhan-Hu-1 AN MN908947)

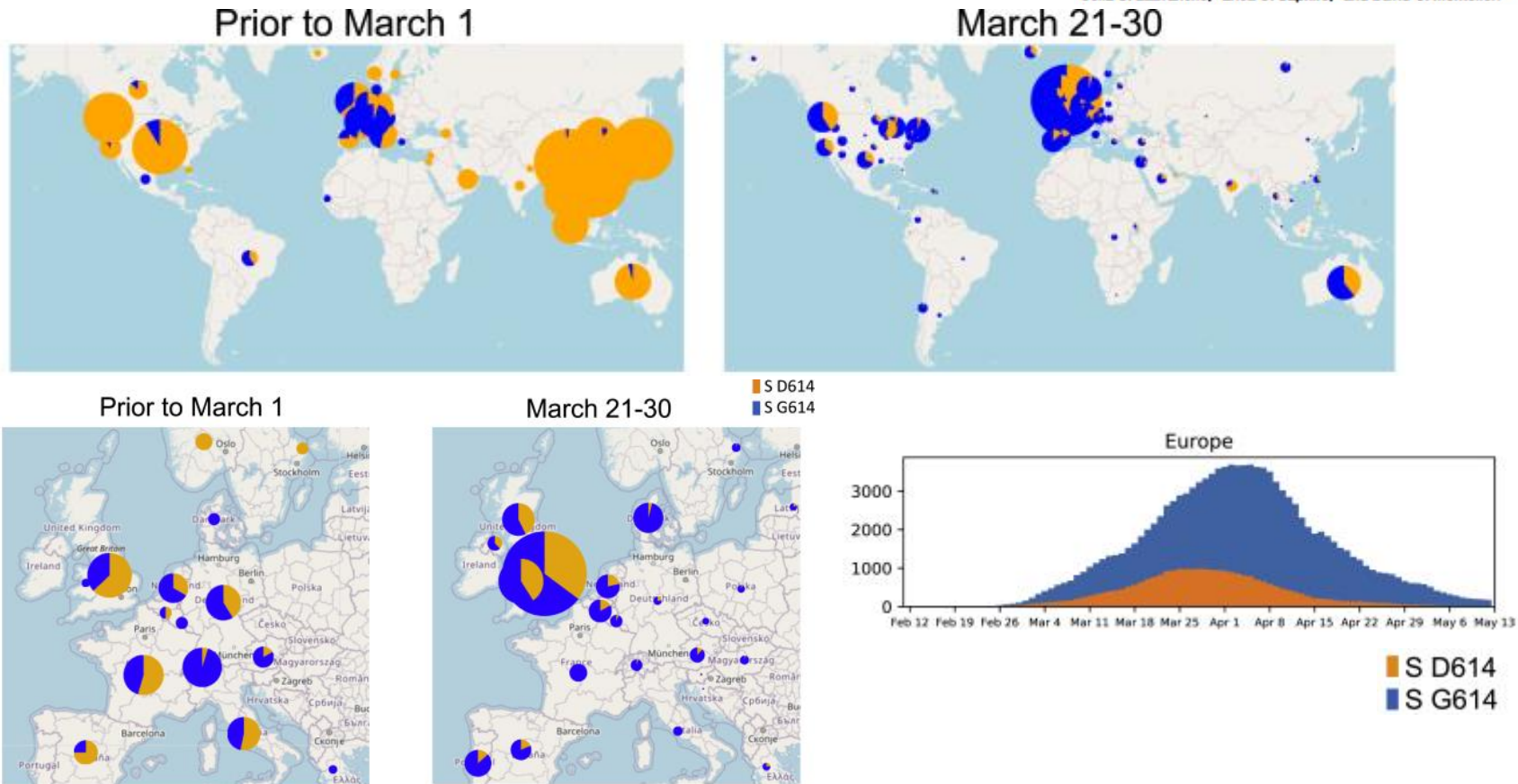
- ✓ **Lineages discendenti** dai primi sono numerati (**A.1, B.1, B.2**, ecc con almeno 1 nucleotide di differenza) e possono essere **progenitori** di altri *lineage* che vengono nominati aggiungendo una numerazione (**B.1.1, A.1.1....**). Questo **fino al 3° livello** (es. **B.1.1.1**), quindi cambia lettera

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

MUTAZIONE D614G

Evidence that D614G Increases Infectivity of the COVID-19 Virus

Bette Korber,^{1,2,10,*} Will M. Fischer,¹ Sandrasegaram Gnanakaran,¹ Hyejin Yoon,¹ James Theiler,¹ Werner Abfalterer,¹ Nick Hengartner,¹ Elena E. Giorgi,¹ Tanmoy Bhattacharya,¹ Brian Foley,¹ Kathryn M. Hastio,³ Matthew D. Parker,⁴ David G. Partridge,⁵ Ciarán M. Evans,⁵ Timothy M. Freeman,⁴ Thushan I. de Silva,^{5,6} on behalf of the Sheffield COVID-19 Genomics Group, Charlene McDaniel,⁷ Lautaro G. Perez,⁷ Haili Tang,⁷ Alex Moon-Walker,^{3,8,9} Sean P. Whelan,⁹ Celia C. LaBranche,² Erica O. Saphiro,³ and David C. Montefiori⁷



- D614G: sostituzione nucleotidica da A a G in posizione 23403
- La sostituzione potrebbe aver indotto un cambiamento conformazionale nella proteina S
- E' diventata la forma dominante nella pandemia
- La variante G614 è associata a una maggiore infettività e a cariche virali più elevate, ma non ad un aumento della severità

La storia delle varianti

- ✓ Alla fine del 2020 iniziano a emergere le varianti «di preoccupazione» (VOC)
- ✓ Evidente dinamica di sostituzione di un lignaggio con quello successivo
- ✓ Dovuta ad una maggiore trasmissibilità o capacità di *immune escape* (reinfezioni e infezione vaccinati)

Alpha (lineage B.1.1.7, clade 20B):

Prima identificazione: in Europa (GB), settembre-ottobre 2020

Prevalente da aprile 2021 (6m dopo)

6 mutazioni e 2 delezioni nella proteina spike; aumento della trasmissibilità del 50% rispetto ai ceppi circolanti precedentemente

Delta (lineage B.1.617.2, clade 21A, 21I, 21J):

Prima identificazione: India, ottobre 2020

Prevalente da agosto 2021 (8m dopo)

Numerosi sotto-lignaggi (AY.x)

10 mutazioni nella proteina spike; aumento della trasmissibilità del 40-60% rispetto alla variante Alpha

Omicron (lineage B.1.529, clade 21K, 21L e 21M, 22B):

Africa, novembre 2021

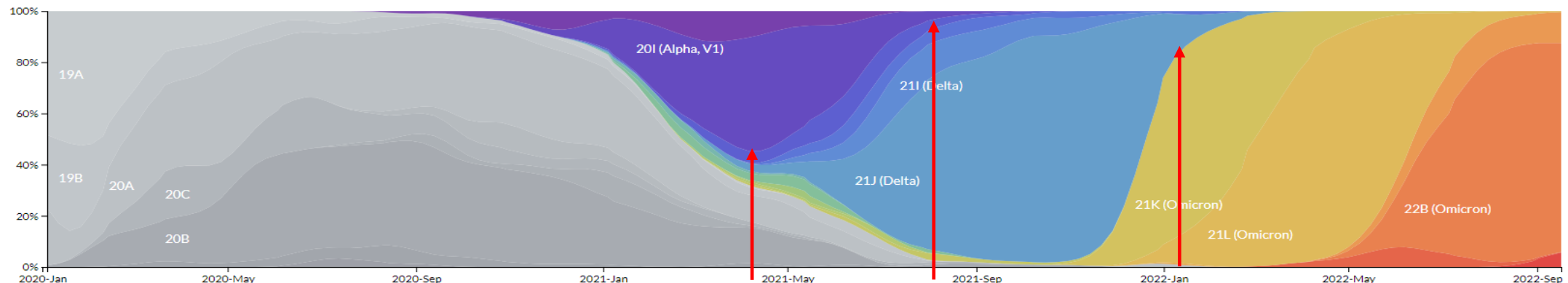
Prevalente da gennaio 2022 (3m dopo)

Numerosi sotto-lignaggi (BA.1-5)

Oltre 30 mutazioni nella proteina spike; aumento della trasmissibilità di circa 3 volte rispetto alla variante Delta

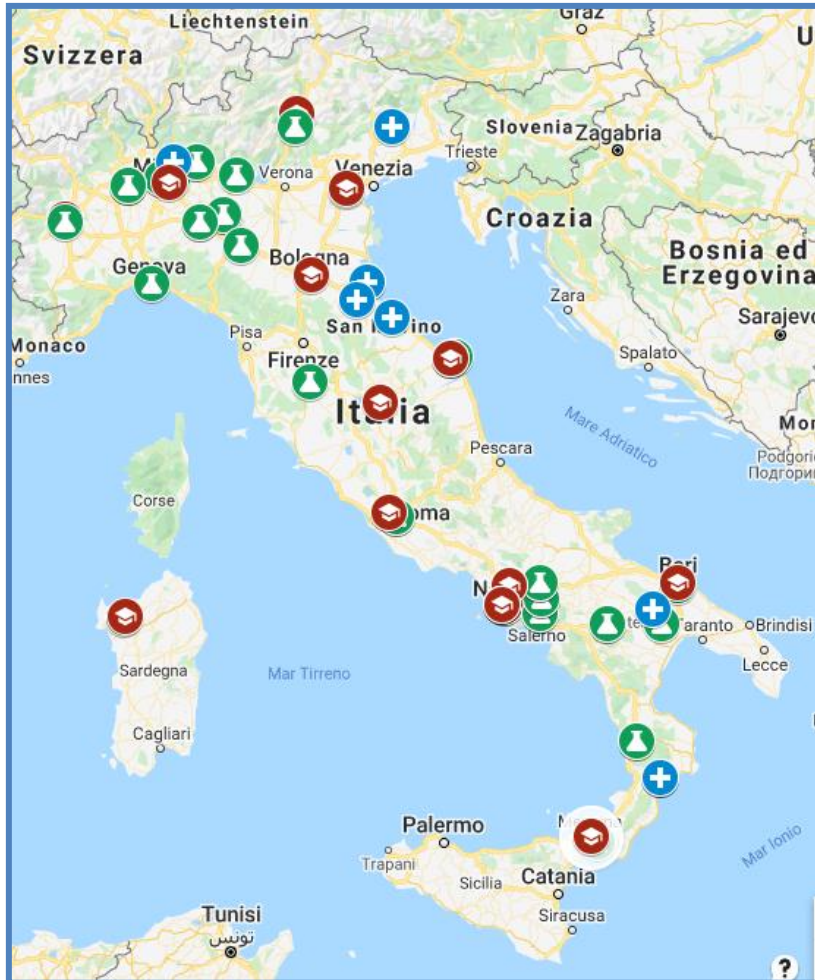
A
STAR
IS
BORN

Frequencies (colored by Clade)



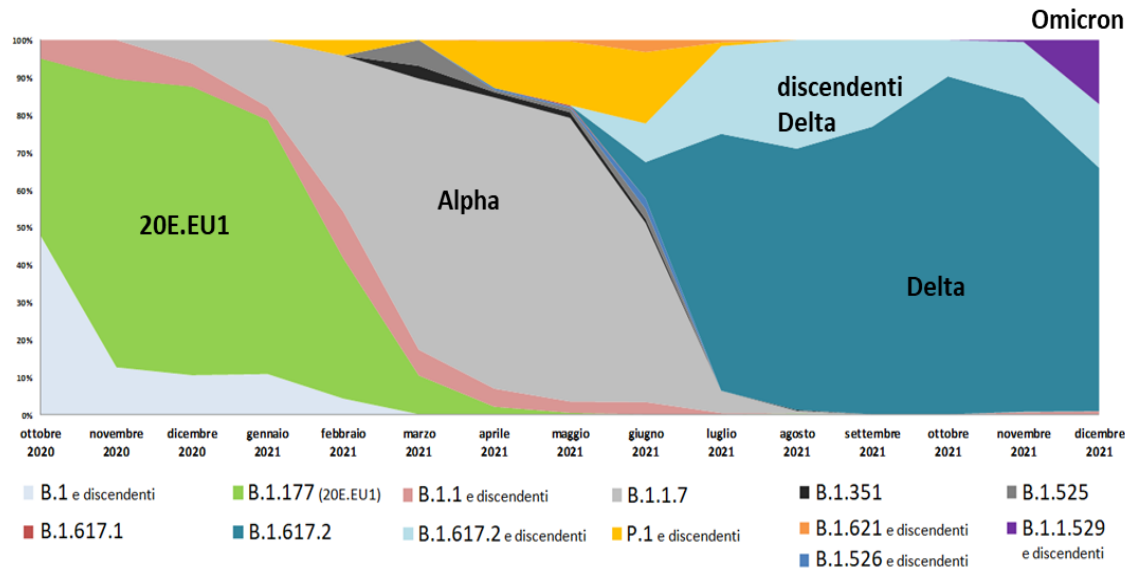
Gruppo collaborativo SCIRE

(SARS-CoV-2 Italian Research Enterprise)- 74 centri



- Department of Biomedical and Clinical Sciences L.Sacco, University of Milan, Milan
- Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano
- IRCCS Ospedale San Raffaele, Milan
- ASST Spedali Civili, University of Brescia, Brescia
- ASST Papa Giovanni XXIII, Bergamo
- ASST Cremona, Cremona
- University of Padova, Padua
- AUSL Piacenza, Piacenza
- IZLER Parma, Parma
- University of Turin, Turin
- Polytechnic University of Marche, Ancona
- University of Siena, Siena
- IRCCS AOU San Martino-IST, Genoa,
- University of Bari, Bari
- AORN S. Anna e S. Sebastiano Caserta, Caserta
- University of Campania "Luigi Vanvitelli", Naples
- Tor Vergata University, Rome
- University Hospital Policlinico Umberto I, Sapienza University of Rome, Rome,
- University of Messina, Messina,
- University of Sassari, Sassari
- University of Perugia, Perugia
- ASST-Lecco, Lecco
- AOU'SS Giovanni di Dio e Ruggi, University of Salerno, Salerno
- A.O.S.G. Moscati, Avellino
- AORN 'San Pio', G.Rummo, Benevento
- University of Naples, Federico II, Naples
- AORN - Ospedali dei ColliAORN
- University "MAGNA GRÆCIA" , Catanzaro
- AOP S.S. Annunziata di Cosenza, Cosenza
- Grande Ospedale Metropolitano "Bianchi Melacrino Morelli", Reggio Calabria
- Azienda Ospedaliera Regionale San Carlo, Contrada Macchia Romana, Potenza
- Presidio Ospedaliero "Madonna delle Grazie", Matera
- AUSL della Romagna- Ravenna, Rimini, Forlì
- Edmund Mach Foundation, Trento
- ASL Bari, Bari
- APSS Trento, Trento
- Università del Piemonte Orientale, Novara
- Fondazione IRCCS Istituto Nazionale dei Tumori, Milan
- University Hospital Major of the Novara Charity, Novara
- Intesa San Paolo Innovation Center-AI LAB, Turin

BACKGROUND



- **Gennaio 2022:** circolazione Omicron **BA.1**, e primi casi **BA.2** e **BA.5**
- **Febbraio-luglio 2022:** circolazione Omicron **BA.2**
- **Luglio- novembre 2022:** prevalenza Omicron **BA.5**

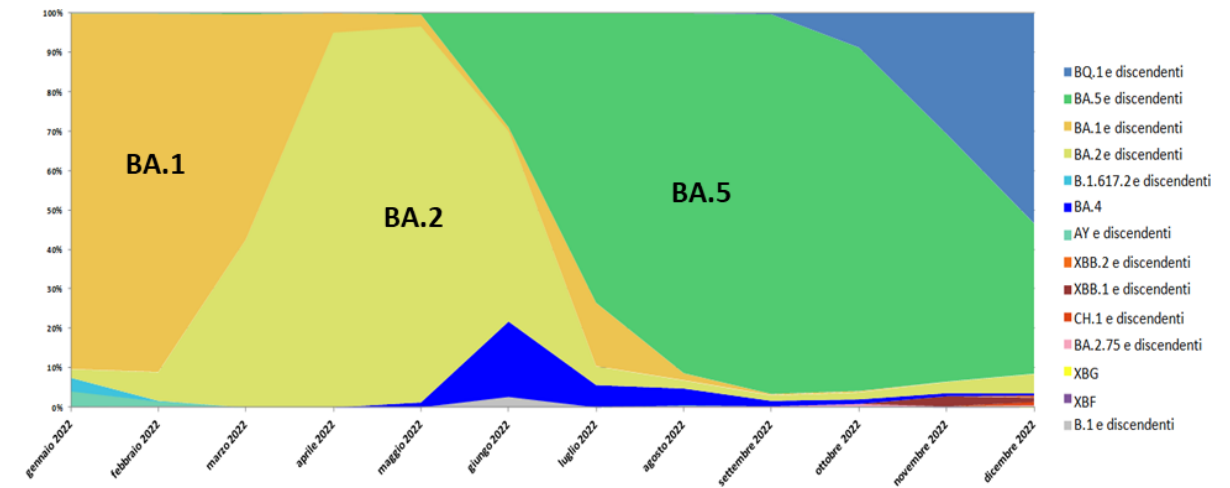
Circulating SARS-CoV-2 variants in Italy, October 2020–March 2021

Alessia Lai^{1,2}, Annalisa Bergna^{1,3}, Stefano Menzo⁴, Gianguglielmo Zehender¹, Sara Caucci², Valeria Ghisetti¹, Francesca Rizzo⁵, Fabrizio Maggi⁶, Francesco Cerutti¹, Giorgio Giurato⁴, Alessandro Weiss^{4,7}, Chiara Turchi¹, Bianca Bruzzone⁸, Francesca Ceccherini-Silberstein⁹, Nicola Clementi¹⁰, Annapaola Callegaro¹¹, Fabio Sagradi¹², Daniela Francisci¹³, Emanuele Venanzi Rullo¹⁴, Ilaria Vicenti¹⁵, Massimo Clementi¹⁶ and Massimo Galli¹ collaborative group SCIRE SARS-CoV-2 Italian Research Enterprise

Epidemiological and Clinical Features of SARS-CoV-2 Variants Circulating between April–December 2021 in Italy

Alessia Lai^{1,2}, Annalisa Bergna^{1,3}, Carla Della Ventura⁴, Stefano Menzo⁴, Bianca Bruzzone⁵, Fabio Sagradi⁴, Francesca Ceccherini-Silberstein⁵, Alessandro Weiss⁶, Nicola Clementi⁷, Gaetano Brindici⁸, Ilaria Vicenti⁹, Lolita Sasset¹⁰, Sara Caucci¹¹, Benedetta Corvaro¹², Silvia Ippoliti¹³, Carla Acciarri¹⁴, Vanessa De Pace¹⁵, Leonardo Lanfranchi¹⁶, Maria C. Bellocchi¹⁷, Giorgio Giurato¹⁸, Roberto Ferraresse¹⁹, Antonella Lagiola²⁰, Daniela Francisci²¹, Martina L. Colombo²², Samuel Lazzarin²³, Matilde Ogliastro²⁴, Maria R. Cappelletti²⁵, Marco Iannetta²⁶, Francesca Rizzo²⁷, Carlo Torti²⁸, Maurizio Fumi²⁹, Morena d'Avenia³⁰, Stefano Brusa³¹, Francesca Greco³², Angela Menchise³³, Vittoria Letizia³⁴, Emilia Vaccaro³⁵, Carmen R. Santoro³⁶, Chiara Fracalvieri³⁷, Sophie Testa³⁸, Luca Carloti³⁹, Teresa Rocco⁴⁰, Annalisa Saracino⁴¹, Annamaria Cattelan⁴², Massimo Clementi⁴³, Loredana Sarmati⁴⁴, Agostino Riva⁴⁵, Massimo Galli⁴⁶, Spinello Antinori⁴⁷, Gianguglielmo Zehender⁴⁸ and SARS-CoV-2 ITALIAN RESEARCH ENTERPRISE (SCIRE) Collaborative Group^{1,4}

- **Ottobre 2020:** ingresso del *clade* **20E.EU1**;
- **Dicembre 2020:** ingresso della **variante Alpha (B.1.1.7)**
- **Maggio 2021:** primi casi di variante **Delta (B.1.617.2)**;
- **Novembre 2021:** primi casi di variante **Omicron (B.1.1.529)**



Nextstrain-maintained pathogen analyses



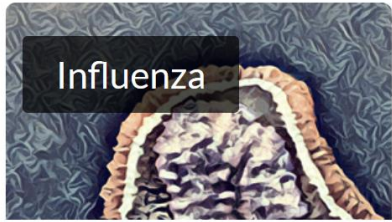
Dengue



West African Ebola
2013-16



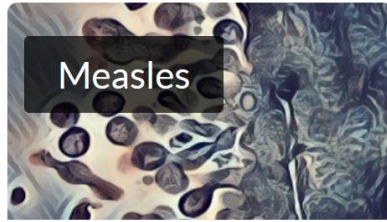
Enterovirus D68



Influenza



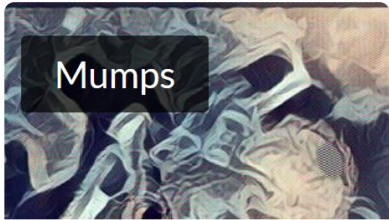
Lassa



Measles



Mpox



Mumps



SARS-CoV-2



RSV



Tuberculosis



West Nile virus



Zika

- ✓ Data la rilevanza epidemiologica di RSV (ogni anno 64 milioni di infezioni e 160.000 decessi), gli attuali sistemi di sorveglianza integrata lo includono



RespiVirNet - Sorveglianza integrata dei virus respiratori
Istituto Superiore di Sanità

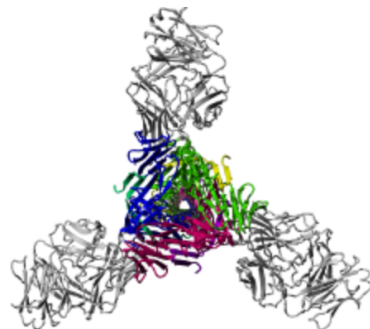
- ✓ Estesa diffusione del sequenziamento dell'intero genoma e integrazione dei dati su RSV nella banca dati GISAID e nella piattaforma Nextstrain

- ✓ RSV è diffuso a livello globale
- ✓ Virus a RNA (circa 15.2Kb) con un tasso di mutazione di $10^{-3}/10^{-4}$ sostituzioni per sito/anno
- ✓ Due sottotipi (**A** e **B**), suddivisi in ulteriori genotipi
- ✓ I sottotipi co-circolano, anche se normalmente prevale un solo sottotipo

Surveillance of respiratory syncytial virus (RSV)

RSV is one of the main causes of lower respiratory tract infections (LRTI), primarily affecting pediatric populations, older adults and those with chronic co-morbidities. RSV-associated respiratory infections are a major cause of hospital admission and hospital deaths in infants and young children. Many countermeasures against RSV, including vaccines, therapeutics and antibodies are in various stages of development. There is a pressing need to monitor genetic evolution, epidemiology and disease characteristics of RSV, particularly in low and middle-income countries (LMICs).

Global surveillance for RSV, currently built on the Global Influenza Surveillance and Response System (GISRS), will make it possible to further investigate and elucidate knowledge gaps for understanding the transmission of the virus in the world, risk factors for RSV infection and severe disease, but above all, will further describe the genetic background of the virus. Genomic surveillance for RSV will enable predictions on the effectiveness of prophylactic treatments and support countermeasure developments against all circulating viral variants, and is essential for optimal implementation of vaccines to analyze in real-time the impact of interventions on circulating viruses.



by A*STAR Singapore



BLAST



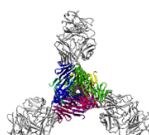
Emerging Variants



EpiCharts



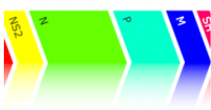
PrimerChecker



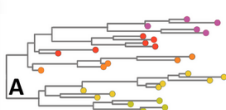
RSVsurver



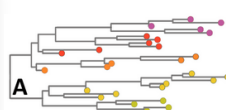
Search & Browse



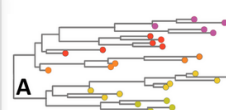
Reference Sequence
Subtype A



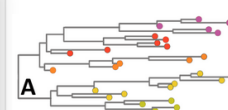
Phylodynamics
Global



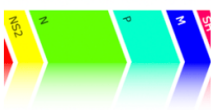
Phylodynamics
Africa



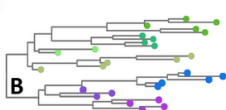
Phylodynamics
Asia



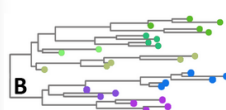
Phylodynamics
South America



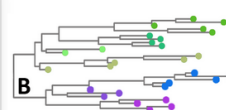
Reference Sequence
Subtype B



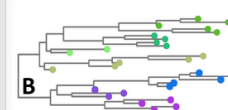
Phylodynamics
Global



Phylodynamics
Africa



Phylodynamics
Asia



Phylodynamics
South America

Important note: In the [GISAID EpiFlu™ Database Access Agreement](#), you have accepted certain terms and conditions for viewing and using data regarding influenza viruses. To the extent the Database contains data relating to non-influenza viruses, the viewing and use of these data is subject to the same terms and conditions, and by viewing or using such data you agree to be bound by the terms of the [GISAID EpiFlu™ Database Access Agreement](#) in respect of such data in the same manner as if they were data relating to influenza viruses.

63 countries and territories shared 16.825 viral sequences from human cases of RSV via GISAID

Country	Continent	WHO region	RSV sequences submitted
United Kingdom	Europe	WHO European Region	1,806
Spain	Europe	WHO European Region	629
France	Europe	WHO European Region	445
Netherlands	Europe	WHO European Region	387
Germany	Europe	WHO European Region	343
Portugal	Europe	WHO European Region	205
Italy	Europe	WHO European Region	197
Slovenia	Europe	WHO European Region	194
Ireland	Europe	WHO European Region	151
Croatia	Europe	WHO European Region	71
Finland	Europe	WHO European Region	69
Russia	Europe	WHO European Region	57
Slovakia	Europe	WHO European Region	49
Austria	Europe	WHO European Region	19
Bulgaria	Europe	WHO European Region	17
Greece	Europe	WHO European Region	13
North Macedonia	Europe	WHO European Region	10
Belgium	Europe	WHO European Region	4
Bosnia and Herzegovina	Europe	WHO European Region	2
Switzerland	Europe	WHO European Region	1

128 genomi interi, 88
sequenze depositate della
stagione 2021-2022 dal
nostro Laboratorio

***APPLICAZIONI DELLA
GENOMICA AL
MONITORAGGIO DI
VARIANTI VIRALI
EMERGENTI***

AI ARTIFICIAL
INTELLIGENCE
LAB Iniziativa di  INTESA SANPAOLO
INNOVATION CENTER

**PREPAREDNESS E
SORVEGLIANZA GENOMICA
NELLE PANDEMIE**
IMPORTANZA DELLE RETI
DI COMPETENZE

Annalisa Bergna

Dipartimento di Scienze Biomediche e Cliniche
Università degli Studi di Milano

Article

Molecular Tracing of SARS-CoV-2 in Italy in the First Three Months of the Epidemic

Alessia Lai ^{1,*}, Annalisa Bergna ¹, Sara Caucci ², Nicola Clementi ³, Ilaria Vicenti ⁴ , Filippo Dragoni ⁴, Anna Maria Cattelan ⁵, Stefano Menzo ², Angelo Pan ⁶, Annapaola Callegaro ⁷, Adriano Tagliabracci ⁸ , Arnaldo Caruso ⁹, Francesca Caccuri ⁹, Silvia Ronchiadin ¹⁰ , Claudia Balotta ¹, Maurizio Zazzi ⁴ , Emanuela Vaccher ¹¹ , Massimo Clementi ³, Massimo Galli ¹ and Gianguglielmo Zehender ¹ on behalf of SARS-CoV-2 ITALIAN RESEARCH ENTERPRISE—(SCIRE) Collaborative Group (Appendix A)

¹ Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, 20157 Milan, Italy; bergna.anna@gmail.com (A.B.); claudia.balotta@unimi.it (C.B.); massimo.galli@unimi.it (M.G.); gianguglielmo.zehender@unimi.it (G.Z.)

² Department of Biomedical Sciences and Public Health, Virology Unit, Polytechnic University of Marche, 60131 Ancona, Italy; s.caucci@univpm.it (S.C.); menzo@univpm.it (S.M.)

³ Microbiology and Virology Unit, “Vita-Salute” San Raffaele University, 20132 Milan, Italy; clementi.nicola@hsr.it (N.C.); clementi.massimo@hsr.it (M.C.)

⁴ Department of Medical Biotechnologies, University of Siena, 53100 Siena, Italy; ilariavicenti@gmail.com (I.V.); dragonifilippo@gmail.com (F.D.); maurizio.zazzi@gmail.com (M.Z.)

⁵ Infectious Diseases Unit, Department of Internal Medicine, Azienda Ospedaliera-Universitaria di Padova, 35128 Padua, Italy; annamaria.cattelan@aopd.veneto.it

⁶ Infectious Diseases, ASST Cremona, 26100 Cremona, Italy; a.pan@asst-cremona.it

⁷ Microbiology and Virology Laboratory, ASST Papa Giovanni XXIII, 24127 Bergamo, Italy; acallegaro@hpg23.it

⁸ Section of Legal Medicine, Università Politecnica delle Marche, 60126 Ancona, Italy; a.tagliabracci@staff.univpm.it

⁹ Microbiology Unit, Department of Molecular and Translational Medicine, University of Brescia and ASST Spedali Civili Hospital, 25123 Brescia, Italy; arnaldo.caruso@unibs.it (A.C.); francesca.caccuri@unibs.it (F.C.)

¹⁰ Intesa Sanpaolo Innovation Center—AI LAB, 10138 Turin, Italy; silvia.ronchiadin@intesasnpaolo.com

¹¹ Medical Oncology and Immune-related Tumors, Centro di Riferimento Oncologico di Aviano (CRO), IRCCS, 33081 Aviano, Italy; evaccher@cro.it

* Correspondence: alessia.lai@unimi.it; Tel.: +39-0250319775

Received: 3 July 2020; Accepted: 21 July 2020; Published: 24 July 2020



PRIMI 3 MESI DI EPIDEMIA IN ITALIA (FEBBRAIO-APRILE 2020)



Caratterizzazione e tracciamento
genomico mediante analisi filogenetica di
SARS-CoV-2 in Italia durante la prima fase
dell’epidemia

DATASET

dicembre 2019-febbraio 2020

gennaio-febbraio 2020



GISAID

sequenze SCIRE

sequenze italiane

sequenze europee e asiatiche

Totale

59

77

452

588



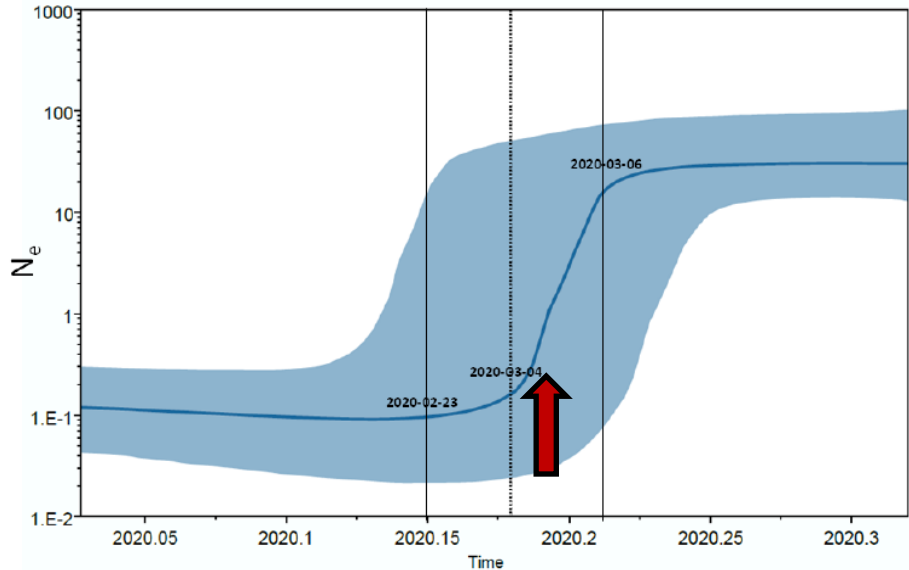
- Criteri di inclusione: 10 genomi italiani e 5 genomi completi europei e asiatici per mese di campionamento
- Criteri di esclusione: genomi con > 5% gap

Sequenze genomiche allineate mediante BioEdit



RISULTATI

- Tutte le sequenze italiane ed europee erano di **lineage B.1**, **clades 20A** e **20B**
- Il **lineage B.1.1** era frequente tra gli isolati più recenti
- Stima dell'**evolutionary rate** di 8×10^{-4} sostituzioni/sito/anno
- La stima del tMRCA (*time of the most recent common ancestor*) conferma la probabile origine dell'epidemia tra la fine di gennaio e l'inizio di febbraio, con un incremento del numero di infezioni tra la fine di febbraio e metà marzo 2020



SITUAZIONE IN ITALIA (FEBBRAIO-GIUGNO 2020)



Conduzione di analisi filodinamiche e
filogeografiche per comprendere la storia
epidemiologica del virus in Italia, il suo
ingresso nel nostro Paese e le dinamiche
spazio-temporali della sua dispersione a
livello nazionale e internazionale

OPEN

Phylogeography and genomic epidemiology of SARS-CoV-2 in Italy and Europe with newly characterized Italian genomes between February-June 2020

Alessia Lai^{1,2}, Annalisa Bergna¹, Stefano Toppo^{3,4}, Marina Morganti⁵, Stefano Menzo⁶, Valeria Ghisetti⁷, Bianca Bruzzone⁸, Mauro Codeluppi⁹, Vito Fiore¹⁰, Emmanuele Venanzi Rullo¹¹, Guido Antonelli¹², Loredana Sarmati¹³, Gaetano Brindicci¹⁴, Annapaola Callegaro¹⁵, Caterina Sagnelli¹⁶, Daniela Francisci¹⁷, Ilaria Vicenti¹⁸, Arianna Miola¹⁹, Giovanni Tonon^{20,21}, Daniela Cirillo²², Ilaria Menozzi⁵, Sara Caucci⁶, Francesco Cerutti⁷, Andrea Orsi²³, Roberta Schiavo²⁴, Sergio Babudieri¹⁰, Giuseppe Nunnari¹¹, Claudio M. Mastroianni²⁵, Massimo Andreoni¹³, Laura Monno¹⁴, Davide Guarneri¹⁵, Nicola Coppola¹⁶, Andrea Crisanti^{26,27}, Massimo Galli¹, Gianguglielmo Zehender^{1,2,28} & SCIRE-SARS-CoV-2 Italian Research Enterprise-Collaborative Group*

The aims of this study were to characterize new SARS-CoV-2 genomes sampled all over Italy and to reconstruct the origin and the evolutionary dynamics in Italy and Europe between February and June 2020. The cluster analysis showed only small clusters including <80 Italian isolates, while most of the Italian strains were intermixed in the whole tree. Pure Italian clusters were observed mainly after the lockdown and distancing measures were adopted. Lineage B and B.1 spread between late

¹Department of Biomedical and Clinical Sciences Luigi Sacco, University of Milan, Milan, Italy. ²Pediatric Clinical Research Center Fondazione Romeo ed Enrica Invernizzi, University of Milan, Milan, Italy. ³Department of Molecular Medicine, University of Padova, Padua, Italy. ⁴CRIBI Biotech Center, University of Padova, Padua, Italy. ⁵Risk Analyses and Genomic Epidemiology Unit, Istituto Zooprofilattico Sperimentale della Lombardia e dell'Emilia Romagna, Parma, Italy. ⁶Department of Biomedical Sciences and Public Health, Virology Unit, Polytechnic University of Marche, Ancona, Italy. ⁷Laboratory of Microbiology and Virology, Amedeo di Savoia, ASL Città di Torino, Torino, Italy. ⁸Hygiene Unit, IRCCS AOU San Martino-IST, Genoa, Italy. ⁹UOC of Infectious Diseases, Department of Oncology and Hematology, Guglielmo da Saliceto Hospital, AUSL Piacenza, Piacenza, Italy. ¹⁰Infectious and Tropical Disease Clinic, Department of Medical, Surgical and Experimental Sciences, University of Sassari, Sassari, Italy. ¹¹Unit of Infectious Diseases, Department of Experimental and Clinical Medicine, University of Messina, Messina, Italy. ¹²Department of Molecular Medicine, University Hospital Policlinico Umberto I, Sapienza University of Rome, Rome, Italy. ¹³Infectious Diseases, Tor Vergata University, Rome, Italy. ¹⁴Infectious Diseases Unit, University of Bari, Bari, Italy. ¹⁵Microbiology and Virology Laboratory, ASST Papa Giovanni XXIII, Bergamo, Italy. ¹⁶Department of Mental Health and Public Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy. ¹⁷Department of Medicine and Surgery, Clinic of Infectious Diseases, "Santa Maria della Misericordia" Hospital, University of Perugia, Perugia, Italy. ¹⁸Department of Medical Biotechnologies, University of Siena, Siena, Italy. ¹⁹Intesa San Paolo Innovation Center-AI LAB, Turin, Italy. ²⁰Center for Omics Sciences, IRCCS Ospedale San Raffaele, Milan, Italy. ²¹Division of Experimental Oncology, IRCCS Ospedale San Raffaele, Milan, Italy. ²²Division of Immunology, Transplantation and Infectious Disease, IRCCS Ospedale San Raffaele, Milan, Italy. ²³Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy. ²⁴UOC of Microbiology, Department of Clinical Pathology, Guglielmo da Saliceto Hospital, AUSL Piacenza, Piacenza, Italy. ²⁵Department of Public Health and Infectious Diseases, University Hospital Policlinico Umberto I, Sapienza University of Rome, Rome, Italy. ²⁶Microbiology and Virology Diagnostic Unit, Padua University Hospital, Padua, Italy. ²⁷Department of Life Science, Imperial College London, South Kensington Campus Imperial College Road, London SW7 2AZ, UK. ²⁸CRC- Coordinated Research Center "EpiSoMI", University of Milan, Milan, Italy. *A list of authors and their affiliations appears at the end of the paper. [✉]email: gianguglielmo.zehender@unimi.it

DATASET

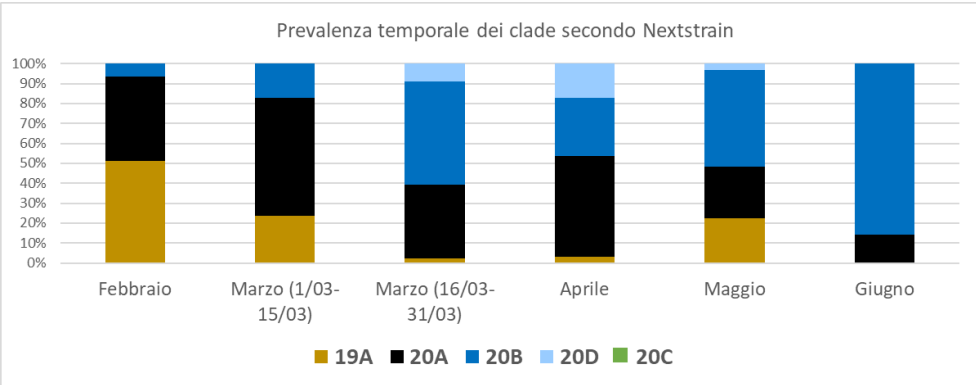
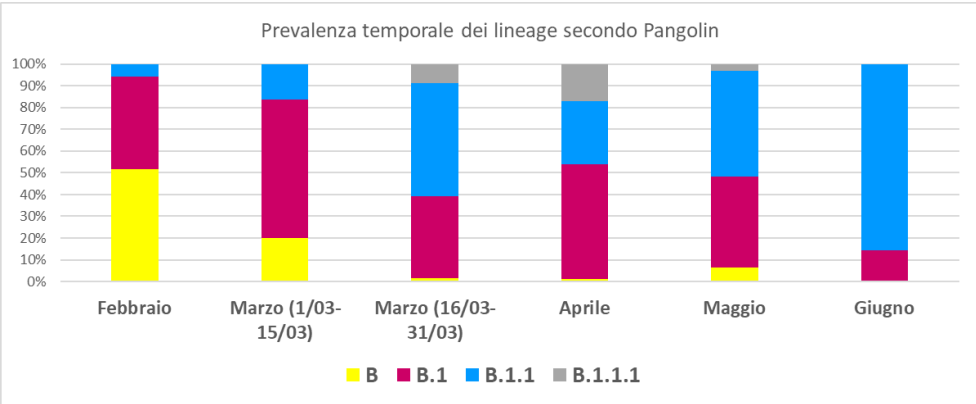
192 nuovi genomi: 24 febbraio 2020 al 18 giugno 2020 (Gruppo collaborativo SARS-CoV- 2 ITALIAN RESEARCH ENTERPRISE - SCIRE)

DATASET NAZIONALE			DATASET INTERNAZIONALE		
sequenze SCIRE	sequenze GISAID	Totale	sequenze GISAID europee	sequenze GISAID asiatiche	Totale
254	211	465	858	52	1,375

PREVALENZA SPAZIALE E TEMPORALE DEI LINEAGES E CLADES

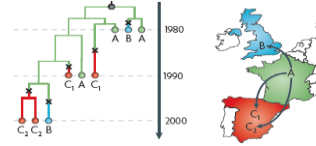


- Presenza nel nostro Paese di diversi *lineage* e *clade* già a inizio della pandemia con differente distribuzione spazio-temporale
- Origine precoce dei ceppi B e B.1 a partire dalla seconda metà di gennaio 2020, seguita dalla comparsa dei lineage B.1.1 e B.1.1.1, rispettivamente tra la fine di febbraio 2020 e la prima settimana di marzo 2020

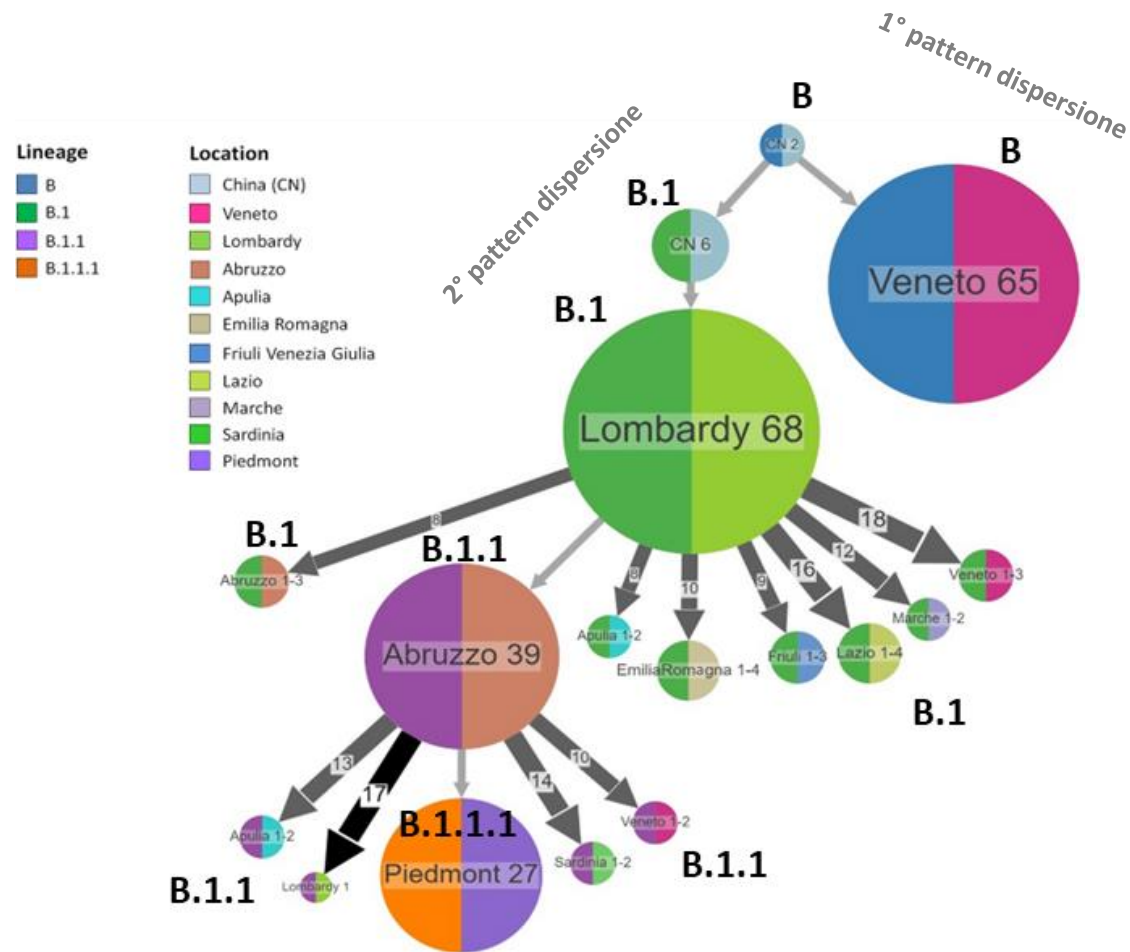


ANALISI FILOGEOGRAFICA

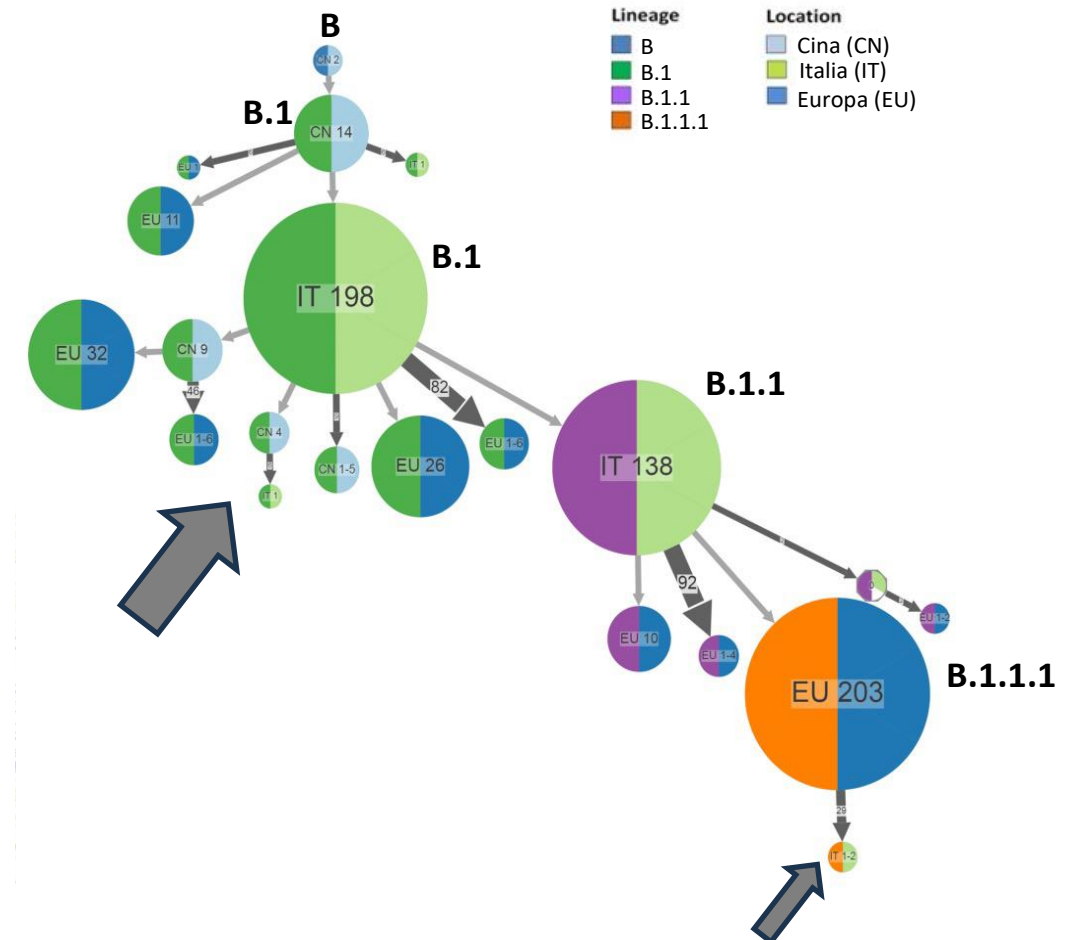
Relazione tra genomi (con località di campionamento nota) e distribuzione spaziale



DATASET ITALIANO



DATASET INTERNAZIONALE

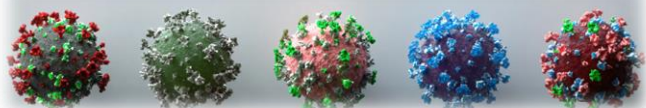


Ruolo centrale dell'Italia nell'**esportazione di alcuni lineages** all'inizio dell'epidemia in Europa, mentre successivamente, ha rappresentato un **centro di importazione** da altri Paesi europei

PERIODI PRE-OMICRON E OMICRON (2020-2021/2022)



Studio della diversità genomica e analisi
filodinamica delle principali varianti di SARS-CoV-2
circolanti nel nostro Paese prima (20E.EU1, Alpha e
Delta) e durante l'era Omicron (BA.1, BA.2 e BA.5)



Received: 13 July 2023 | Accepted: 12 October 2023

DOI: 10.1002/jmv.29193

RESEARCH ARTICLE

JOURNAL OF
MEDICAL VIROLOGY WILEY

Genomic epidemiology of the main SARS-CoV-2 variants in Italy between summer 2020 and winter 2021

Annalisa Bergna¹ | Alessia Lai¹ | Carla Della Ventura¹ | Bianca Bruzzone² |
Alessandro Weisz³ | Morena d'Avenia⁴ | Sophie Testa⁵ | Carlo Torti⁶ |
Caterina Sagnelli⁷ | Angela Menchise⁸ | Gaetano Brindicci⁹ | Daniela Francisci¹⁰ |
Ilaria Vicenti¹¹ | Nicola Clementi^{12,13} | Annapaola Callegaro¹⁴ |
Emmanuele Venanzi Rullo¹⁵ | Sara Caucci¹⁶ | Vanessa De Pace² | Andrea Orsi^{2,17} |
Stefano Brusa¹⁸ | Francesca Greco¹⁹ | Vittoria Letizia²⁰ | Emilia Vaccaro²¹ |
Gianluigi Franci³ | Francesca Rizzo³ | Fabio Sagradi⁵ | Leonardo Lanfranchi⁵ |
Nicola Coppola⁷ | Annalisa Saracino⁹ | Michela Sampaolo¹³ | Silvia Ronchiadin²² |
Massimo Galli¹ | Agostino Riva¹ | Gianguglielmo Zehender¹ | SARS-CoV-2
ITALIAN RESEARCH ENTERPRISE – (SCIRE) collaborative Group

¹Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy

²Hygiene Unit, IRCCS AOU San Martino-IST, Genoa, Italy

³Laboratory of Molecular Medicine and Genomics, Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana", University of Salerno and Genome Research Center for Health, Baronissi, Italy

⁴UOSVD of Cytopathology and Screening, Department of Laboratory Medicines, Ospedale di Veneri, Asl Bari, Bari, Italy

⁵Unit of Infectious Diseases, Azienda Socio Sanitaria Territoriale Cremona, Cremona, Italy

⁶Infectious and Tropical Disease Unit, Department of Medical and Surgical Sciences, Magna Graecia University of Catanzaro, Catanzaro, Italy

⁷Department of Mental Health and Public Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy

⁸Microbiology and Virology Laboratory, A.O.R. San Carlo Potenza, Potenza, Italy

⁹Infectious Diseases Unit, University of Bari, Bari, Italy

¹⁰Department of Medicine and Surgery, Clinic of Infectious Diseases, "Santa Maria della Misericordia" Hospital, University of Perugia, Perugia, Italy

¹¹Department of Medical Biotechnologies, University of Siena, Siena, Italy

¹²Laboratory of Microbiology and Virology, Università "Vita-Salute" San Raffaele, Milan, Italy

¹³Laboratory of Microbiology and Virology, IRCCS San Raffaele Scientific Institute, Milan, Italy

¹⁴Ospedali Riuniti, Department of Infectious Diseases, Bergamo, Italy

¹⁵Unit of Infectious Diseases, Department of Experimental and Clinical Medicine, University of Messina, Messina, Italy

¹⁶Department of Biomedical Sciences and Public Health, Virology Unit, Polytechnic University of Marche, Ancona, Italy

¹⁷Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy

¹⁸Department of Translational Medical Sciences, Federico II University, Naples, Italy

¹⁹UOC Microbiology and Virology, PO Cosenza, Cosenza, Italy

²⁰UOSD Genetic and Molecular Biology, AORN Sant'Anna and San Sebastiano di Caserta, Caserta, Italy

²¹Molecular Biology Units, AOU "S. Giovanni di Dio e Ruggi d'Aragona" Università di Salerno, Salerno, Italy

²²Artificial Intelligence Laboratory, Intesa Sanpaolo Innovation Center, Turin, Italy

DATASET

PERIODO PRE-OMICRON: 20E.EU1, ALPHA E DELTA

609 genomi completi collezionati da agosto 2020 a gennaio 2022
(SARS-CoV- 2 ITALIAN RESEARCH ENTERPRISE - SCIRE)

DATASET NAZIONALI				DATASET INTERNAZIONALI		
	sequenze nostre	sequenze GISAID	Totale	sequenze GISAID Europee	sequenze GISAID non Europee	Totale
20E.EU1	269	802	1,071	1,144	210	2,495
Alpha	148	1,175	1,323	1,470	381	3,174
Delta	176	1,319	1,495	1,451	302	3,248

20E.EU1 → agosto 2020-ottobre 2021
Alpha → novembre 2020-agosto 2021
Delta → maggio 2021-gennaio 2022

- Criteri inclusione: 10 genomi completi per ogni regione italiana e mese di campionamento
- Criteri esclusione: genomi > 5% gap

- Criteri inclusione: 5 genomi completi Europei e non europei per mese di campionamento
- Criteri esclusione: genomi > 5% gap

PERIODO OMICRON: BA.1, BA.2 E BA.5

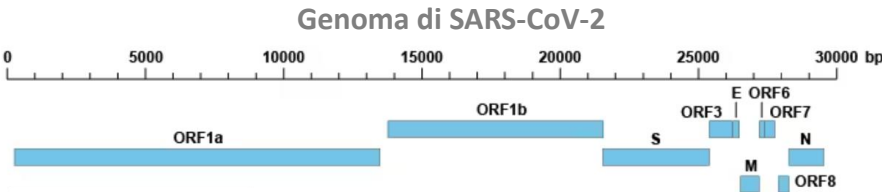
1,658 genomi completi collezionati tra dicembre 2021 e dicembre 2022
(SARS-CoV- 2 ITALIAN RESEARCH ENTERPRISE - SCIRE)

DATASET NAZIONALI				
		sequenze SCIRE	sequenze GISAID	Totale
Omicron	BA.1	268	612	880
	BA.2	677	950	1,627
	BA.5	713	1,048	1,761

BA.1 → dicembre 2021 - aprile 2022
BA.2 → gennaio-dicembre 2022
BA.5 → maggio-dicembre 2022

- Criteri inclusione: 10 genomi completi per ogni regione italiana e mese di campionamento
- Criteri esclusione: genomi > 5% gap

ANALISI DELLE MUTAZIONI DEI DATASEI



Clade 20E.EU1

- **10 sostituzioni** aminoacidiche (5 mutazioni in gene S)

Gene	Mutazione	Totale n=1,071 (%)
ORF 1a	A3623S	214 (20%)
ORF 1b	P314L	1,053 (98.3%)
	A222V	1,048 (97.8%)
	A262S	352 (32.9%)
S	P272L	371 (34.6%)
	D614G	1,061 (99.1%)
	Q675H	149 (13.9%)
N	A220V	1,043 (97.4%)
ORF 10	V30L	1,064 (99.3%)
ORF 14	L67F	1,064 (99.3%)

Variante Alpha

- **17 sostituzioni** aminoacidiche; (7 mutazioni in gene S)
- **3 delezioni** (2 in gene S)

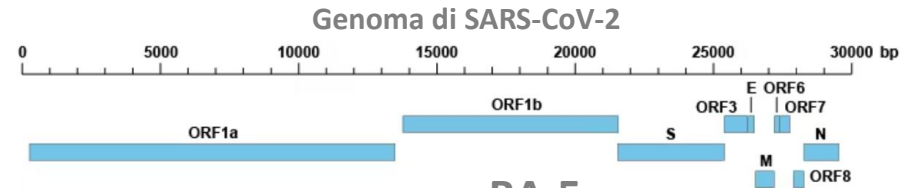
Gene	Mutazione	Totale n=1,323 (%)
ORF 1a	T1001I	1,305 (98.6%)
	A1708D	1,305 (98.6%)
	I2230T	1,314 (99.3%)
	SGF 3675-3677del	1,136 (85.9%)
	N501Y	1,270 (95.9%)
	A570D	1,309 (98.9%)
	D614G	1,321 (99.8%)
	P681H	1,280 (96.7%)
S	T716I	1,272 (96.1%)
	S982A	1,323 (94.6%)
	D1118H	1,306 (98.7%)
	HV 69-70del	1,233 (93.2%)
	Y144del	1,248 (94.3%)
ORF 8	Q27 [#]	1,320 (99.8%)
	R52I	1,197 (90.5%)
	Y73C	1,316 (99.5%)
	D3L	1,316 (99.4%)
N	R203K	1,312 (99.1%)
	G204R	1,281 (96.8%)
	S235F	1,280 (96.7%)

Variante Delta

- **34 sostituzioni** aminoacidiche (10 mutazioni in gene S)
- **2 delezioni** (2 in gene S)

Gene	Mutazione	Totale n=1,495 (%)		
ORF 1a	K621N	204 (13.6%)		T19R 1,441 (96.4%)
	A1306S	1,293 (86.5%)		T95I 635 (42.5%)
	P1640L	195 (13%)		G142D 948 (63.4%)
	P2046L	1,292 (86.4%)		R158G 1,433 (95.9%)
	P2287S	1,287 (86.1%)		A222V 191 (12.8%)
	A2529V	167 (11.2%)		L452R 1,458 (97.5%)
	V2930L	1,197 (80.1%)		T478K 1,431 (95.7%)
	A3209V	188 (12.6%)		D614G 1,384 (96.2%)
	T3255I	1,287 (86.1%)		P681R 138 (9.2%)
	T3646A	1,293 (86.5%)		D950N 1,303 (87.2%)
	V3718A	187 (12.5%)		E156del 1,429 (95.6%)
	T3750I	153 (10.2%)		F157del 1,451 (97.1%)
ORF 1b	P314L	1,442 (96.5%)		ORF 3a S26L 1,486 (99.4%)
	G662S	1,484 (99.3%)		ORF 7a V82A 1,355 (90.6%)
	P1000L	1,452 (97.2%)		T120I 1,425 (95.3%)
	A1918V	1,225 (81.9%)		
				Q9L 325 (21.7%)
				D63G 1,424 (95.3%)
				D377Y 1,486 (99.4%)
				R203M 1,488 (99.5%)
				G215C 1,286 (86%)

ANALISI DELLE MUTAZIONI DEI DATASEI



BA.1

- 49 sostituzioni aminoacidiche (32 in gene S)
- 7 delezioni in più del 10% degli isolati (3 in gene S)

Gene	Mutazione	n=880 (%)			
ORF 1a	K856R*	830 (94.3)		T478K*	495 (56.3)
	V1887I	285 (32.4)	BA.1.17 (91%)	E484A*	511 (58.1)
	S2083del*	863 (98.1)	BA.1.17.2 (98%)	Q493R*	492 (55.9)
	L2084I*	862 (97.9)		G496S*	475 (54)
	A2710T*	813 (92.4)		Q498R*	446 (50.7)
	T3255I*	850 (96.6)		N501Y*	448 (50.9)
	P3395H*	851 (96.7)		Y505H*	465 (52.8)
	L3674del*	849 (96.5)		T547K*	839 (95.3)
	S3675del*	852 (96.8)		D614G*	812 (92.3)
	G3676del*	851 (96.7)		H655Y*	832 (94.5)
ORF1b	I3758V*	811 (92.2)		N679K*	818 (92.9)
	P314L*	879 (99.9)		P681H*	807 (91.7)
	I1566V*	832 (94.5)	BA.1.17.2 (86%)	A701V	180 (20.5)
	A67V*	843 (95.8)		N764K*	839 (95.3)
	H69del*	843 (96.5)		D796Y*	799 (90.8)
	V70del*	849 (96.1)		N856K*	851 (96.7)
	T95I*	777 (88.3)		Q954H*	851 (96.7)
	G142del*	814 (92.5)		N969K*	812 (92.3)
	VY 143-144del*	815 (92.6)		L981F*	809 (91.9)
	Y145D*	874 (99.3)		T9I*	822 (93.4)
S	N211del	787 (89.4)	BA.1.1 (95%)	D3G*	757 (86)
	L212I*	772 (87.7)	BA.1.1.1 (88%)	Q19E*	810 (92)
	G339D*	771 (87.6)		A63T*	848 (93.5)
	R346K	345 (39.2)		P10S*	877 (99.7)
	S371L*	722 (82)		ENA 27-29del*	880 (100)
	S373P*	754 (85.7)		P13L*	880 (100)
	S375F*	727 (82.6)		ER 31-32del*	880 (100)
	K417N*	668 (75.9)		S33del*	872 (99.1)
	N440K*	628 (71.4)	BA.1.1 (74%)	R203K*	845 (96)
	G446S	665 (75.6)		G204R*	845 (96)
	S477N*	495 (56.3)			

BA.2

- 51 sostituzioni aminoacidiche (28 in gene S)
- 4 delezioni in più del 10% degli isolati (1 in gene S)

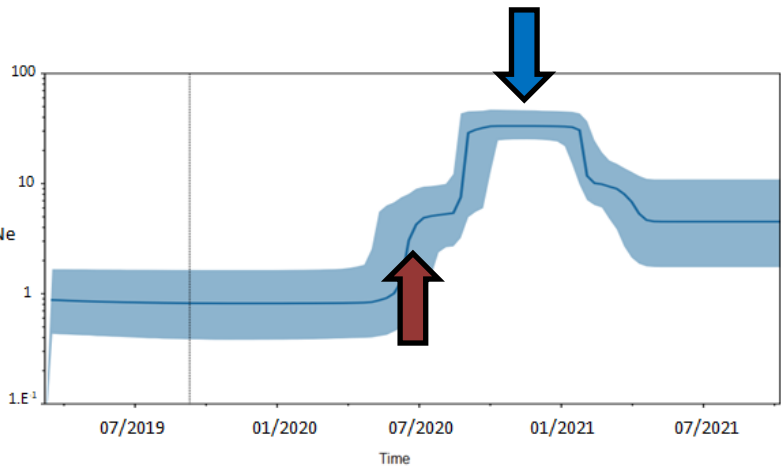
Gene	Mutazione	Totale n=1,627 (%)				
ORF 1a	S135R*	1,603 (98.5)	S	E484A*	1506 (92.6)	
	T842I*	1,465 (90)		Q493R*	1417 (87.1)	
	G1307S*	1,626 (99.9)		Q498R*	1240 (76.2)	
	L3027F*	1,618 (99.4)		N501Y*	1277 (78.5)	
	T3090I*	1,516 (93.2)		Y505H*	1296 (79.7)	
	L3201F*	1470 (90.3)		D614G*	1512 (92.9)	
	T3255I*	1621 (99.6)		H655Y*	1618 (99.4)	
	P3395H*	1624 (99.8)		N679K*	1567 (96.3)	
	SG 3675-3676del*	1627 (100)		P681H*	1594 (98)	
	F3677del*	1626 (99.9)		N764K*	1607 (98.8)	
ORF1b	P314L*	1626 (99.9)	BA.2.9 (88.5%)	D796Y*	1549 (95.2)	
	R1315C*	1619 (99.5)		Q954H*	1621 (99.6)	
	I1566V*	1553 (95.4)		N969K*	1566 (96.3)	
	T2163I*	1534 (94.3)		H78Y	268 (16.5)	
	T19I*	1614 (99.2)		ORF3a	T223I*	1618 (99.4)
	LPP 24-26del	1616 (99.3)		E	T9I*	1537 (94.5)
	A27S	1614 (99.2)		M	Q19E*	1562 (96)
	G142D*	1547 (95)		A63T*	1622 (99.7)	
	V213G*	1565 (96.2)		ORF6	D61L*	1523 (93.6)
	G339D*	1333 (81.9)		ORF9b	P10S	1620 (99.6)
S	S371F*	1506 (92.5)	N	EN 27-28del	1627 (100)	
	S373P*	1518 (93.3)		A29del	1626 (99.9)	
	S375F*	1491 (91.6)		P13L*	1625 (99.9)	
	T376A*	1482 (91.1)		E31del	1626 (99.9)	
	D405N*	1572 (96.6)		R32del	1627 (100)	
	R408S*	1562 (96)		S33del	1610 (98.9)	
	K417N*	1564 (96.1)		R203K*	1625 (99.9)	
	N440K*	1391 (85.5)		G204R*	1624 (99.8)	
	S477N*	1397 (85.9)		S413R*	1549 (95.2)	
	T478K*	1426 (87.6)				

BA.5

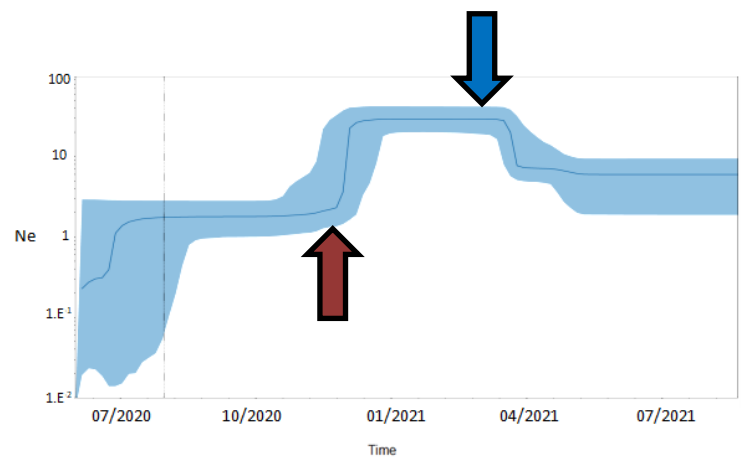
- 52 sostituzioni aminoacidiche (29 in gene S)
- 5 delezioni in più del 10% degli isolati (4 in gene S)

Gene	Mutazione	Totale n=1,761 (%)			
ORF 1a	S135R*	1753 (99.5)		S477N*	1589 (90.2)
	T842I*	1652 (93.8)		T478K*	1657 (94.1)
	G1307S*	1758 (99.8)		E484A*	1735 (98.5)
	L3027F*	1743 (99)		F486V*	1735 (98.5)
	T3090I*	1654 (93.9)		Q498R*	1622 (92.1)
	T3255I*	1749 (99.3)		N501Y*	1647 (93.5)
	P3395H*	1753 (99.5)		Y505H*	1661 (94.3)
	SGF 3675-3677del*			S	
	1760 (99.9)		D614G*	1654 (93.9)	
P314L*	1756 (99.7)	BA.5.2 (90%)	H655Y*	1759 (99.9)	
T1050N	367 (20.8)		BA.5.2.2 (99%)		
ORF1b	R1315C*	1754 (99.6)		N679K*	1704 (96.7)
	P681H*			P681H*	1709 (97)
	I1566V*	1732 (98.4)		N764K*	1736 (98.6)
	T2163I*	1665 (94.5)		D796Y*	1725 (98)
S	T19I*	1753 (99.5)		Q954H*	1758 (99.8)
	L24del*	1753 (99.5)		N969K*	1732 (98.3)
	PP 25-26del*	1752 (99.5)	ORF3a	T223I*	1747 (99.2)
	A27S*	1749 (99.3)	E	T9I*	1716 (97.4)
	HV 69-70del*	1749 (99.3)		D3N*	1732 (98.4)
	F140-141del*	1755 (99.6)	M	Q19E*	1685 (95.7)
	G142Y*	1728 (98.1)	BA.5.2 (100%)	A63T*	1756 (99.7)
	V213G*	1742 (98.9)		BA.5.2.1 (98%)	P10S
	G339D*	1495 (84.9)	ORF9b		D16G
	S371F*	1622 (92.1)		EN 27-28del*	1745 (99.1)
S373P*	1640 (93.1)		A29del*	1628 (92.4)	
S375F*	1603 (91.0)		P13L*	1740 (98.8)	
T376A*	1604 (91.1)		E31del*	1745 (99.1)	
D405N*	1724 (97.9)		R32del*	1744 (99)	
R408S*	1686 (95.7)		S33del*	1607 (91.2)	
K417N*	1729 (98.2)		N	R203K*	1760 (99.9)
N440K*	1590 (90.3)			G204R*	1759 (99.9)
L452R*	1728 (98.1)			S413R*	1731 (98.3)

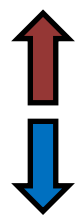
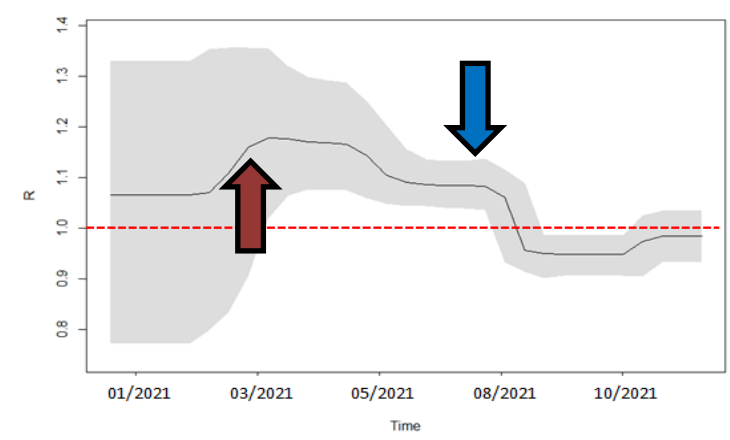
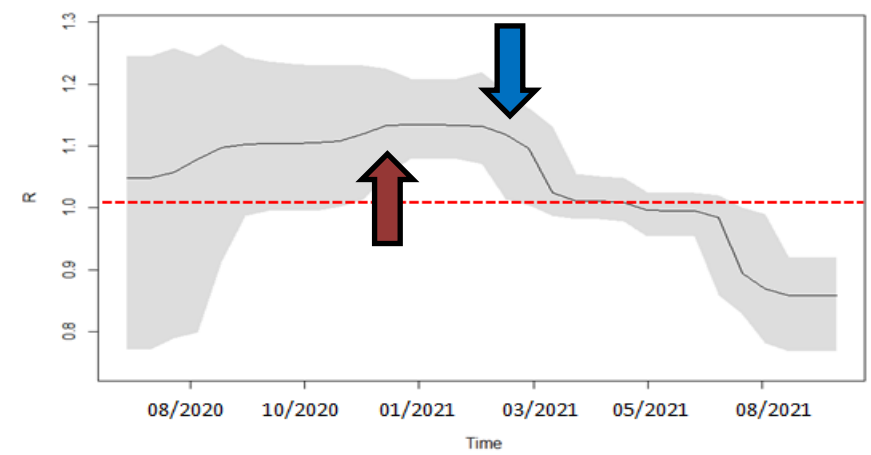
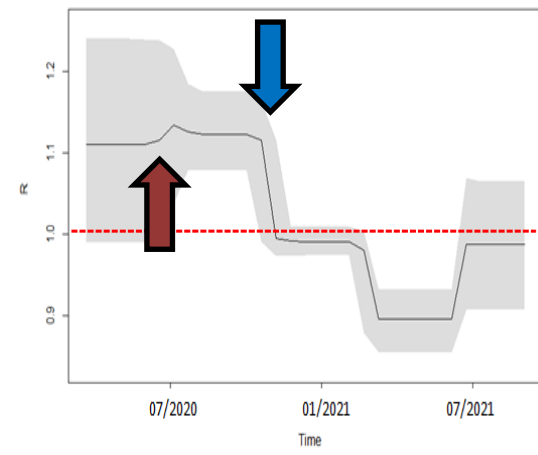
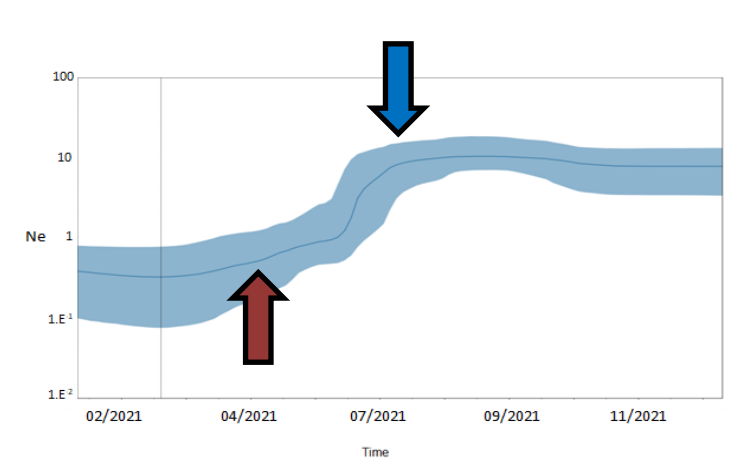
20E.EU1



Alpha

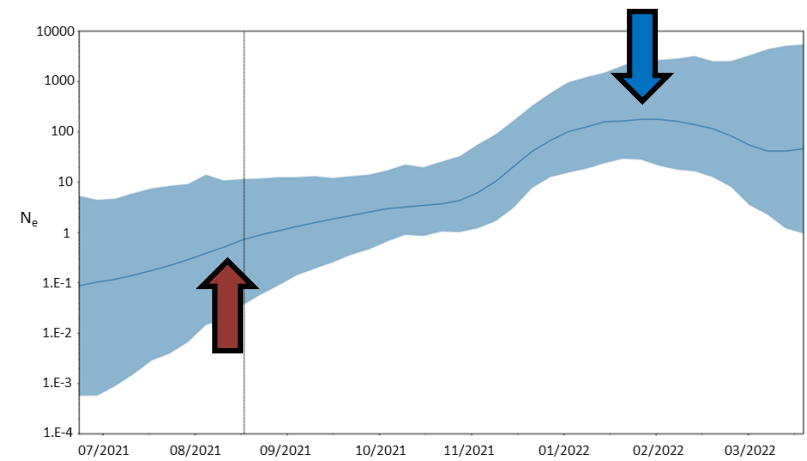


Delta

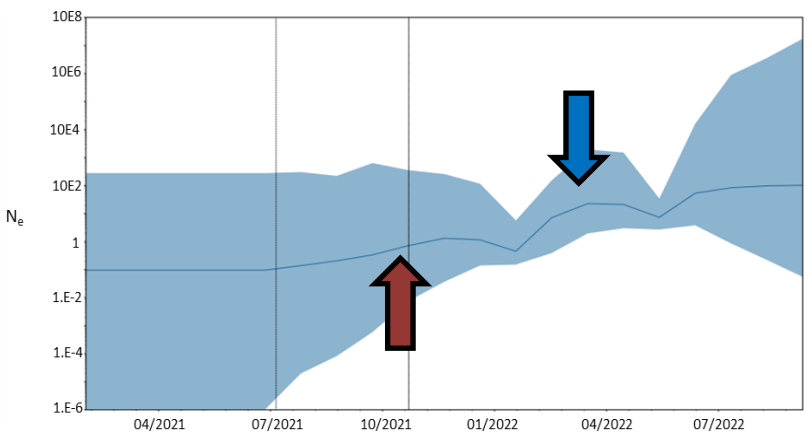


Aumento numero infezioni e parametro R_e
Fase stazionaria numero infezioni e decremento R_e

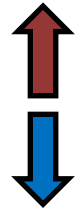
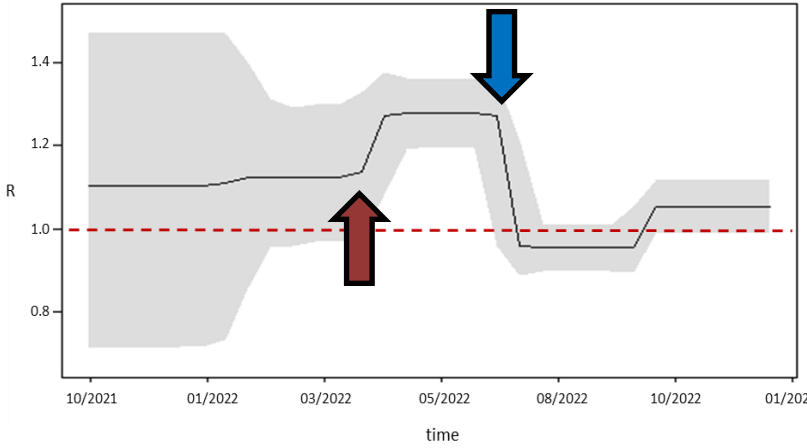
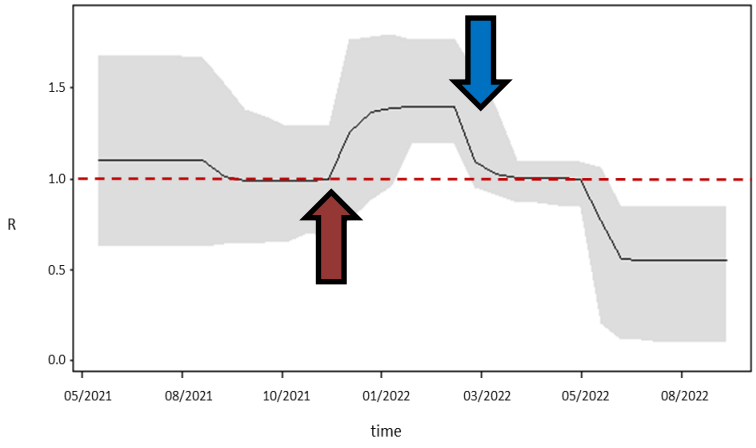
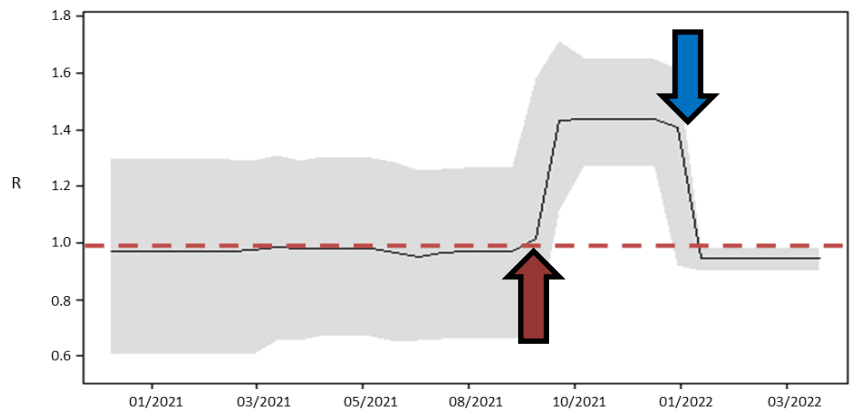
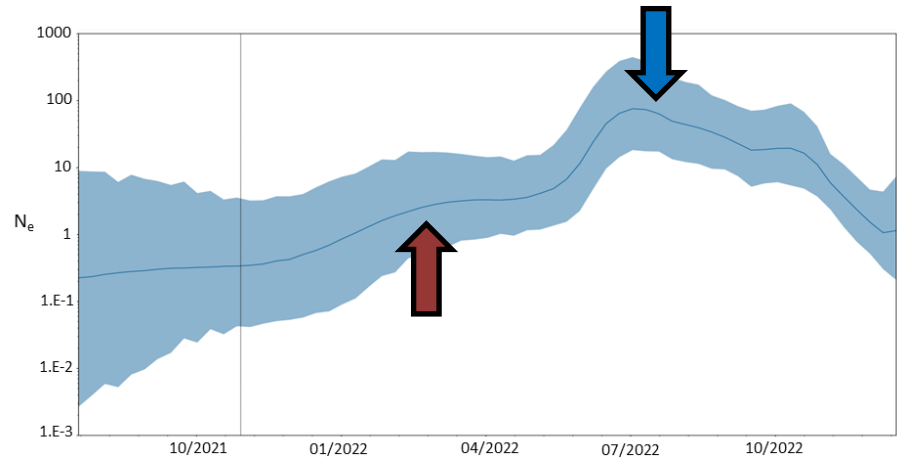
BA.1



BA.2



BA.5



Aumento numero infezioni e parametro R_e
Fase stazionaria numero infezioni e decremento R_e

VIRUS RESPIRATORIO SINCIZIALE (RSV)



 **frontiers** | Frontiers in **Microbiology**

TYPE Original Research
PUBLISHED 04 January 2024
DOI 10.3389/fmicb.2023.1327239

 Check for updates

OPEN ACCESS

EDITED BY
Kazuya Shirato,
National Institute of Infectious Diseases
(NIID), Japan

REVIEWED BY
Koo Nagasawa,
Chiba Cancer Center, Japan
Oliver Schildgen,
Kliniken der Stadt Köln, Germany

*CORRESPONDENCE
Alessia Lai
✉ alessia.lai@unimi.it

RECEIVED 24 October 2023
ACCEPTED 28 November 2023
PUBLISHED 04 January 2024

Epidemiology and molecular analyses of respiratory syncytial virus in the 2021–2022 season in northern Italy

Alessia Lai^{✉*}, Annalisa Bergna¹, Valentina Fabiano^{1,2}, Carla della Ventura¹, Giulia Fumagalli¹, Alessandra Mari², Martina Liodice², Gian Vincenzo Zuccotti^{1,2} and Gianguglielmo Zehender¹

¹Department of Biomedical and Clinical Sciences, University of Milan, Milan, Italy, ²Pediatric Department, "Vittore Buzzi" Children's Hospital, Milan, Italy



- **Identificazione e discriminazione** dei **sottotipi** di RSV nella popolazione pediatrica afferente all'Ospedale dei Bambini "Vittore Buzzi" nelle stagioni 2021/2022 e 2022/2023
- Studio dell'**evoluzione molecolare** e analisi della **filogenesi** dei sottotipi di HRSV

DATASET

104 tamponi nasofaringei: stagione 2021-2022 (novembre 2021 -gennaio 2022)

105 tamponi nasofaringei: stagione 2022-2023 (dicembre 2022 - gennaio 2023)

DATASET INTERNAZIONALE

DATASET NAZIONALE

sequenze italiane

sequenze europee/internazionali

Totale

sottotipo A

89

752

841

sottotipo B

133

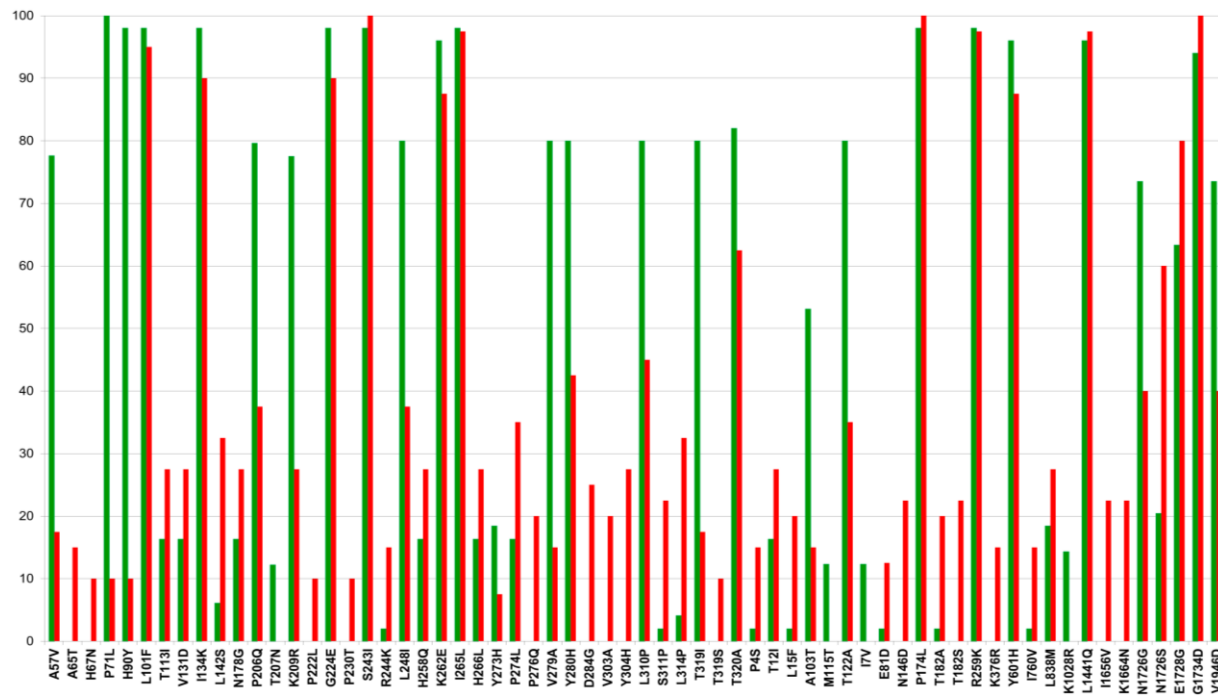
646

779

MUTAZIONI e VARIABILITA'

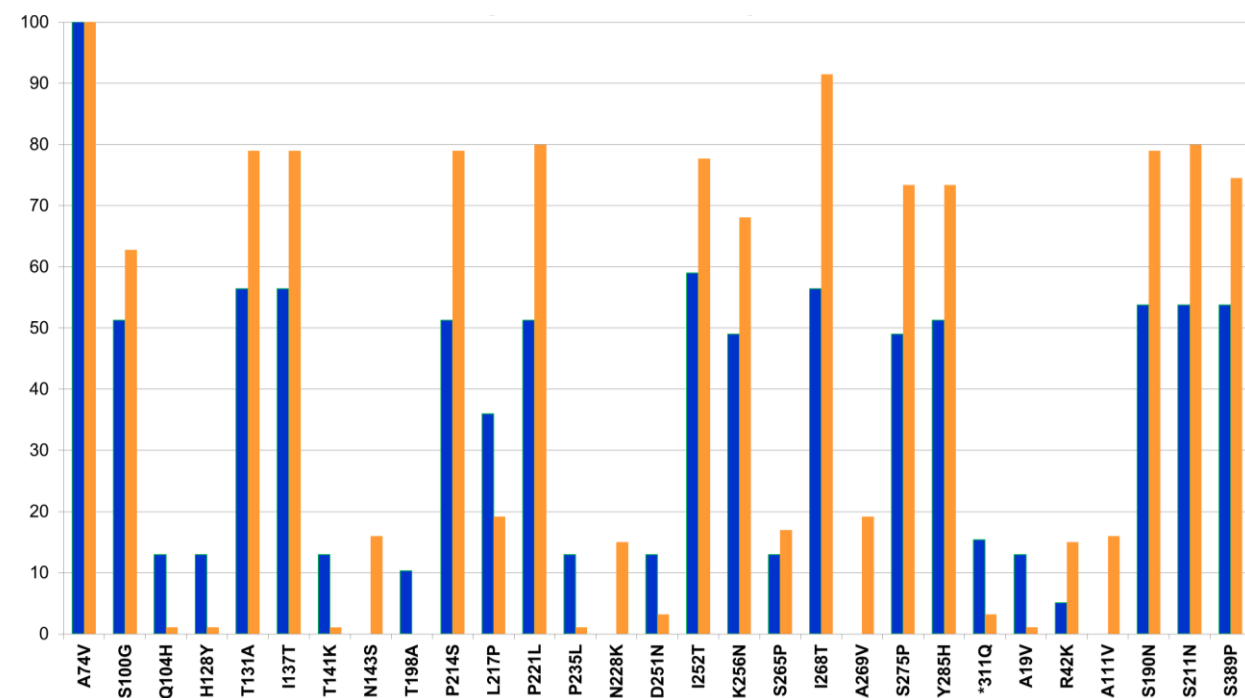
RSV-B

- Stagione 2021-2022 (n=49)
- Stagione 2022-2023 (n=40)



RSV-A

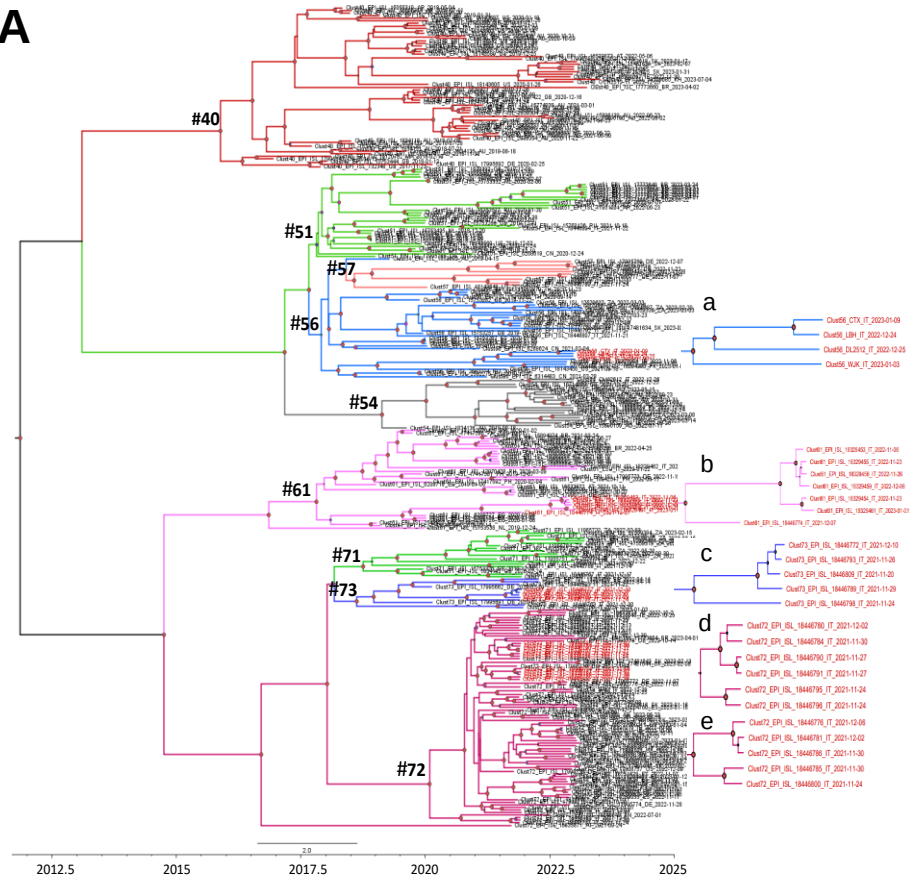
- Stagione 2021-2022 (n=39)
- Stagione 2022-2023 (n=94)



Elevata variabilità soprattutto nei campioni più recenti

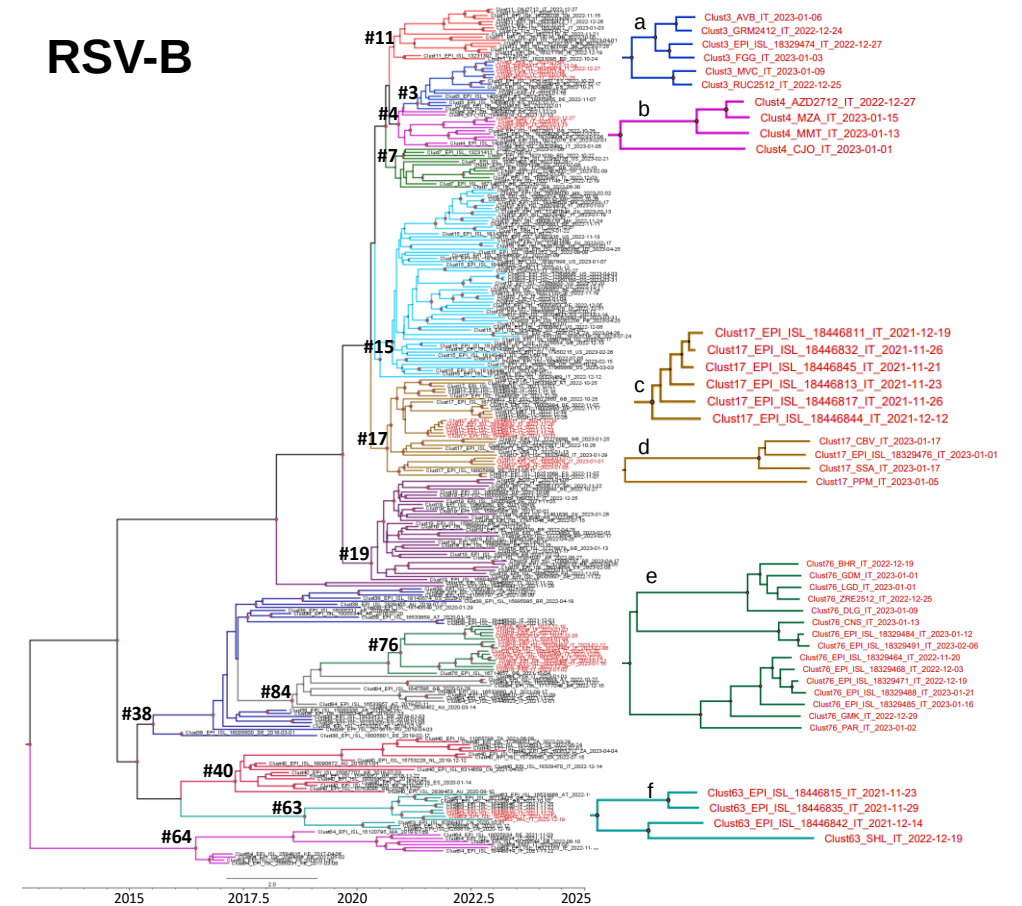
Cluster di trasmissione

RSV-A



Cluster misti: datazione novembre 2015-gennaio 2020
5 clade puri italiani: 1 della stagione 2022-2023,
 3 della stagione 2021-2022
 1 misto

RSV-B

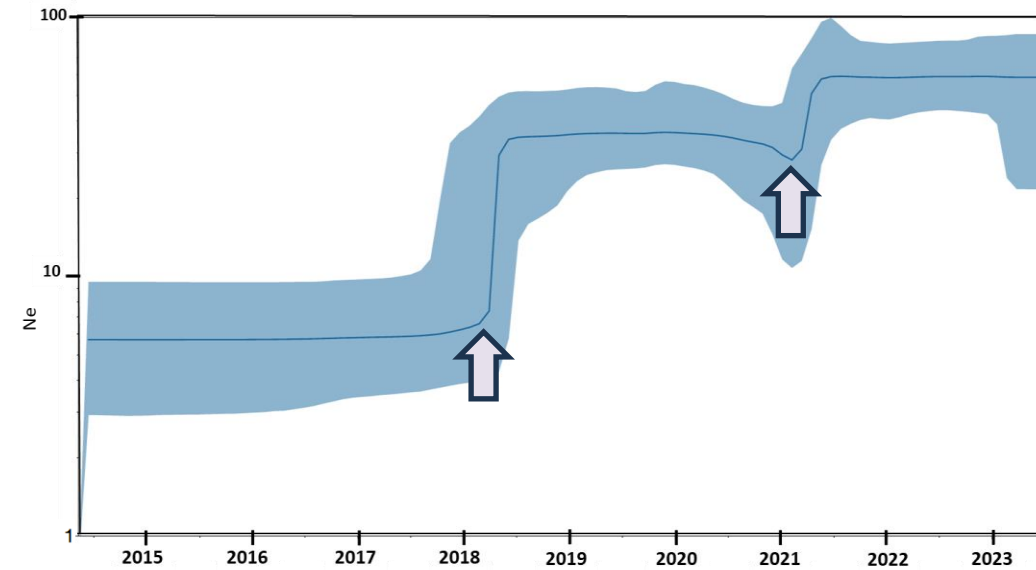


Cluster misti: datazione luglio 2015- aprile 2021
6 clade puri italiani: 4 della stagione 2022-2023,
 1 della stagione 2021-2022
 1 misto

Le due epidemie di RSV sono state determinate dall'ingresso e dall'amplificazione di cluster che circolavano da almeno 8 anni in Europa, ma anche da varianti di più recente formazione

ANALISI FILODINAMICA

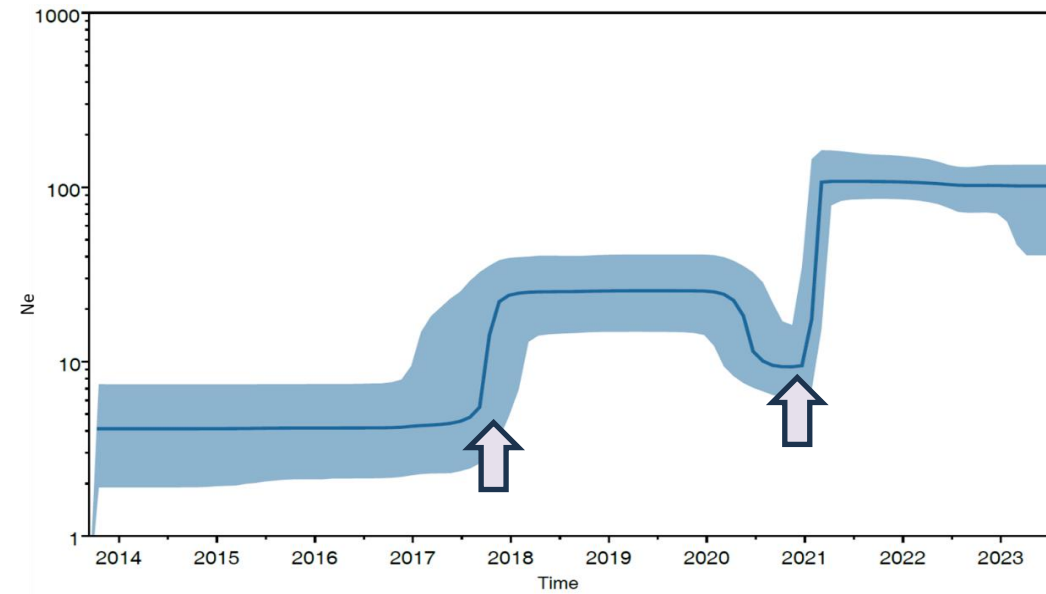
RSV-A



2 fasi di crescita:

- inizio 2018
- inizio 2021

RSV-B



Progetto SMART IT

- La **collaborazione** tra UNIMI, Intesa Sanpaolo e Innovation Center ha permesso la **rapida costruzione dei dataset** utilizzati **selezionando** le sequenze sulla base di specifici criteri e garantendo un **continuo update** delle sequenze campionate disponibili
- Intesa Sanpaolo e Innovation Center continueranno a supportare l'attività di ricerca di UniMI, attraverso la **messa a disposizione di un ambiente informatico protetto e gestito** ("Smart IT") da dedicare alle analisi svolte per lo studio della filogenesi dei virus, in particolare quelle su SARS-CoV-2
- Il sistema ha l'obiettivo di permettere ai ricercatori dell'Ospedale Sacco e del gruppo SCIRE di **conservare le sequenze genomiche di SARS-CoV-2 sequenziate, integrandole con i relativi metadati**. I ricercatori potranno effettuare analisi su tali sequenze, eventualmente integrandole con dataset temporanei caricati *ad hoc*



GRAZIE!

**AI ARTIFICIAL
LAB INTELLIGENCE**
Iniziativa di  INTESA SANPAOLO
INNOVATION CENTER

PREPAREDNESS E

SORVEGLIANZA GENOMICA

NELLE PANDEMIE

IMPORTANZA DELLE RETI

DI COMPETENZE