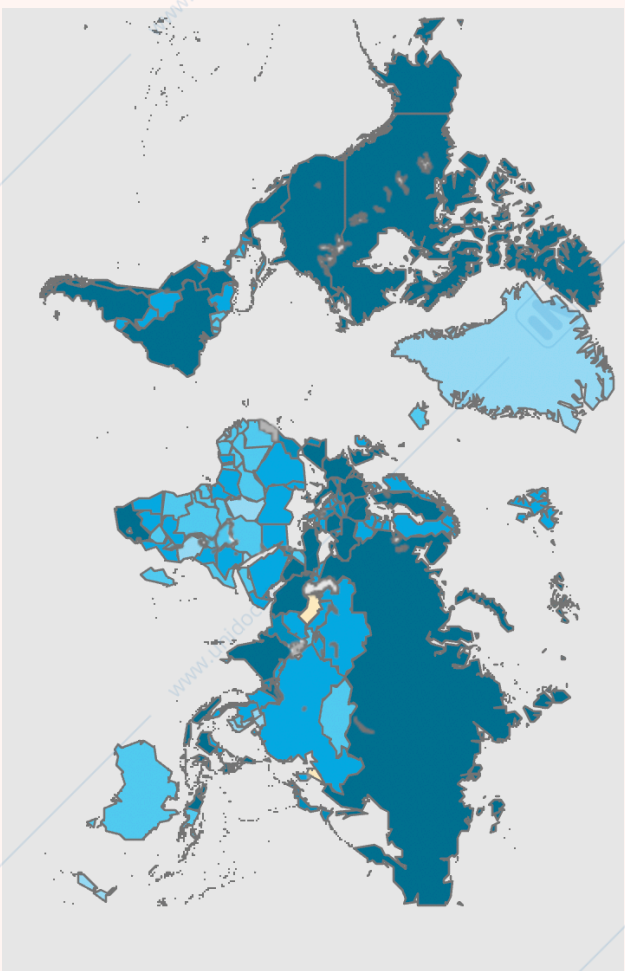


# NEXT GENERATION SEQUENCING OF SARS-COV-2 GENOMES

**Challenges, application and opportunities**

**Francesca Roselli**



**Gennaio 2020**  **nuovo betacoronavirus (SARS-CoV-2)**

**11 marzo 2020**  **la WHO dichiara lo stato di pandemia**

**160 813 869**

**Confirmed cases**

Last update: 14 May 2021, 06:18 pm CEST

**3 339 002**

**Confirmed deaths**

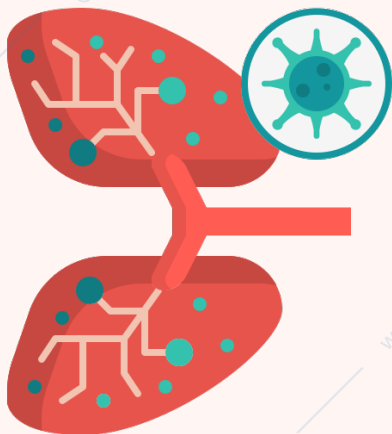
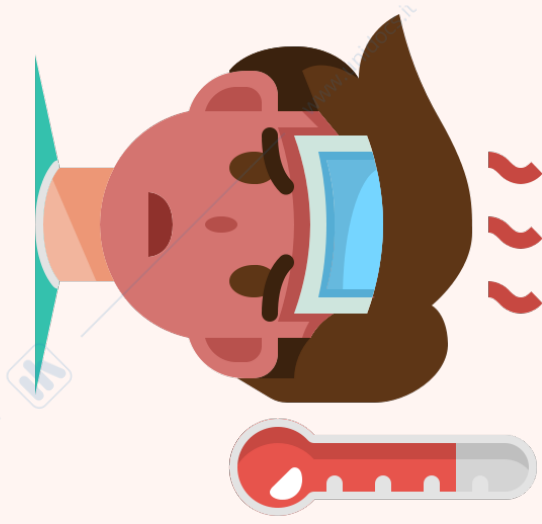
Last update: 14 May 2021, 06:18 pm CEST

**223**

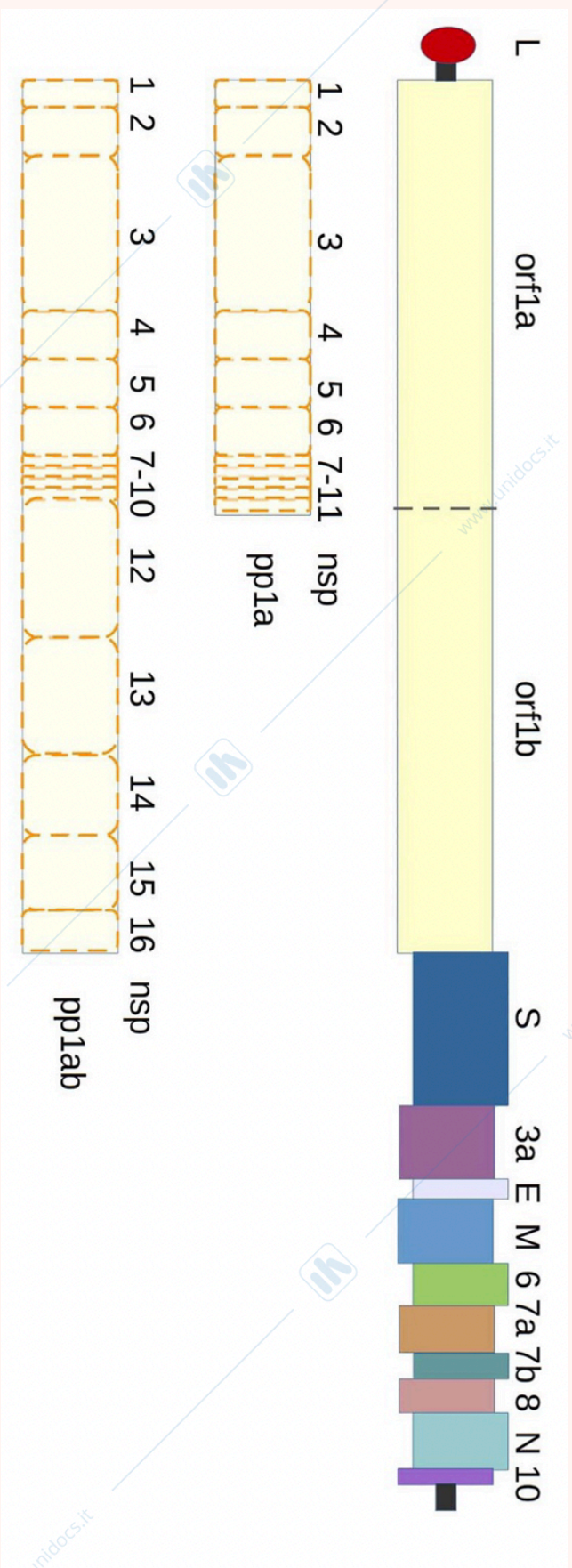
**Countries, areas or territories with cases**

Last update: 14 May 2021, 06:18 pm CEST

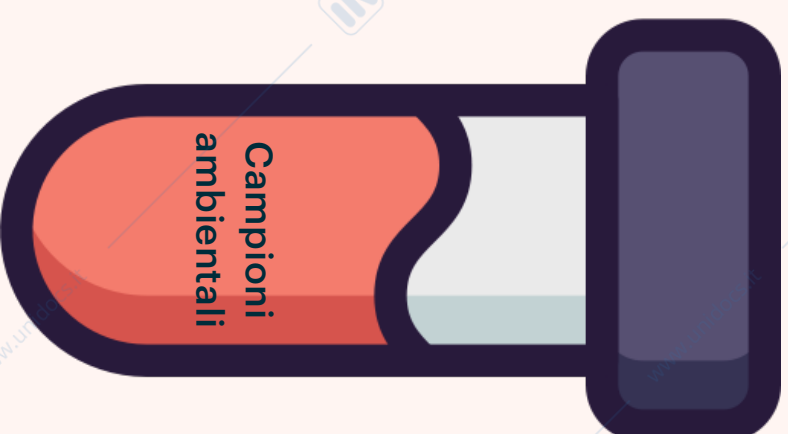
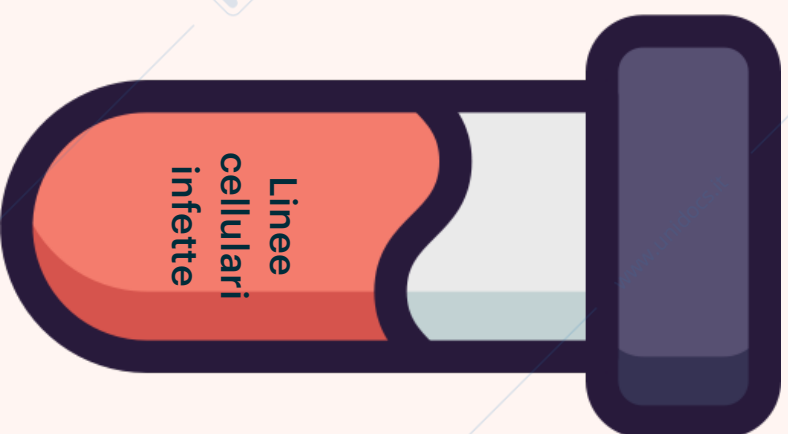
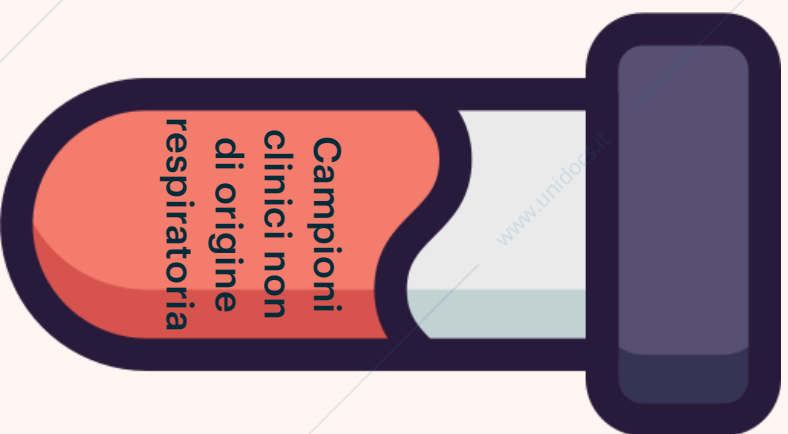
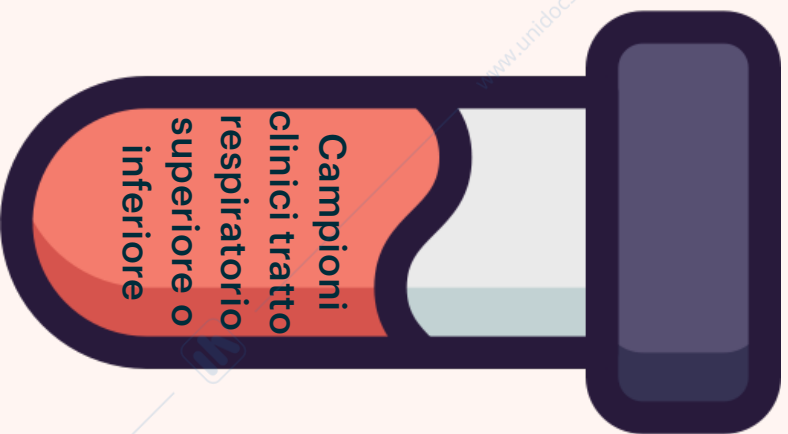
# SINTOMI



**Dicembre 2019**  **Prima sequenza genomica di SARS-CoV-2**



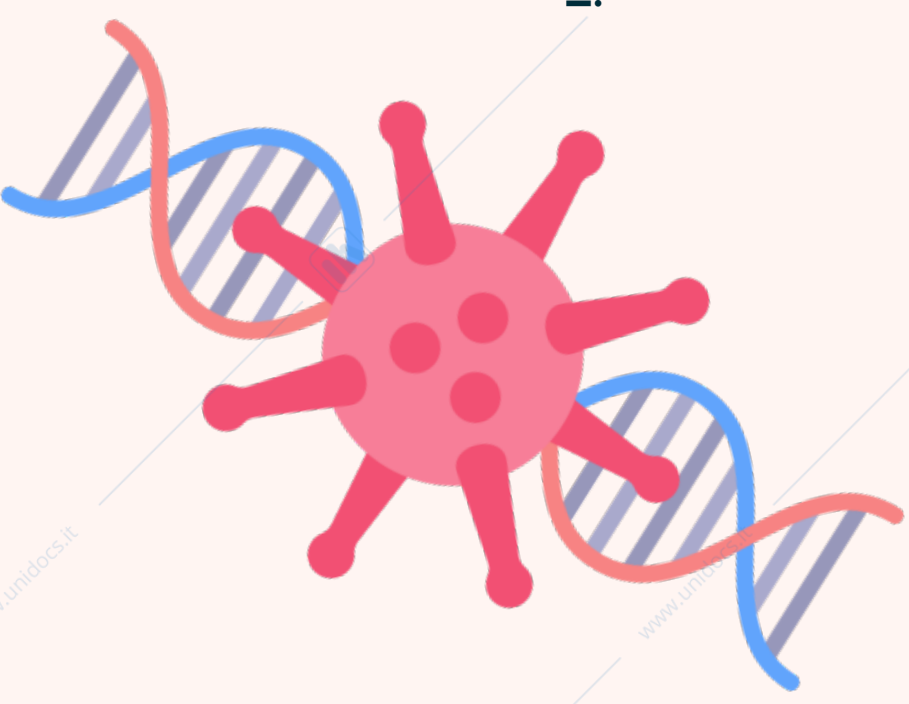
# SAMPLE COLLECTION



# ESTRAZIONE RNA

Da campioni clinici, colture isolate o campioni ambientali

- sale di guanidina
- fenolo



# STRATEGIE DI SEQUENZIAMENTO

## Next Generation Sequencing



quattro possibili approcci



	Shotgun metatranscriptomics
Goals	SARS-CoV-2, host microbiota, and host response to infection
Co-infection detection	Yes
Minimum number of reads	20-50 M
Genome Coverage	$\geq 99\%$
Accuracy in SNV identification	High
Sample viral load (Ct) requested (ref Xiao)	<24-28
Sample RNA input (ng)	10-200
Sample type	Patient specimens
Cost	High
NGS sequencing platforms	High- or ultra high-throughput platforms

# SHOTGUN METATRASCRIPTOMICS

## VANTAGGI:

- Sequenziamento dell'intero DNA di un campione
- Non è richiesta nessuna conoscenza a priori

## SVANTAGGI:

- Campioni con carico virale alto
- Elevata profondità di sequenziamento
- Tecnica lenta, che richiede molto lavoro e l'utilizzo di laboratori con requisiti specifici

# AMPLICON-BASED SEQUENCING

## VANTAGGI:

- Approccio altamente specifico
- Richiede poche quantità di RNA
- Poco costoso
- Campioni con basse cariche virali

## SVANTAGGI:

- Significativa conoscenza a priori della sequenza
- Possibile inefficienza dei primer nell'amplificazione

	Amplicon-based
Goals	SARS-CoV-2 genome
Co-infection detection	No
Minimum number of reads	5-20 M
Genome Coverage	≥95-99%
Accuracy in SNV identification	High
Sample viral load (Ct) requested (ref Xiao)	≥24-28
Sample RNA input (ng)	1-50
Sample type	Patient specimens, environmental samples
Cost	Low
NGS sequencing platforms	Mid-throughput platforms

# HYBRID CAPTURE ENRICHMENT SEQUENCING

	Hybrid capture-enrichment
Goals	SARS-CoV-2 genome
Co-infection detection	No/yes (depending on gene panel)
Minimum number of reads	5–20 M
Genome Coverage	≥95–99%
Accuracy in SNV identification	Moderate
Sample viral load (Ct) requested (ref Xiao)	≥24–28
Sample RNA input (ng) Sample type	10–50 Patient specimens, environmental samples
Cost	Moderate
NGS sequencing platforms	Mid- or high-throughput platforms

- VANTAGGI:**
- Approccio altamente specifico
  - Bassa profondità di sequenziamento
  - Approccio robusto rispetto alla variabilità genomica

- SVANTAGGI:**
- Usato prevalentemente con campioni con alta carica virale

# DIRECT RNA SEQUENCING

## VANTAGGI:

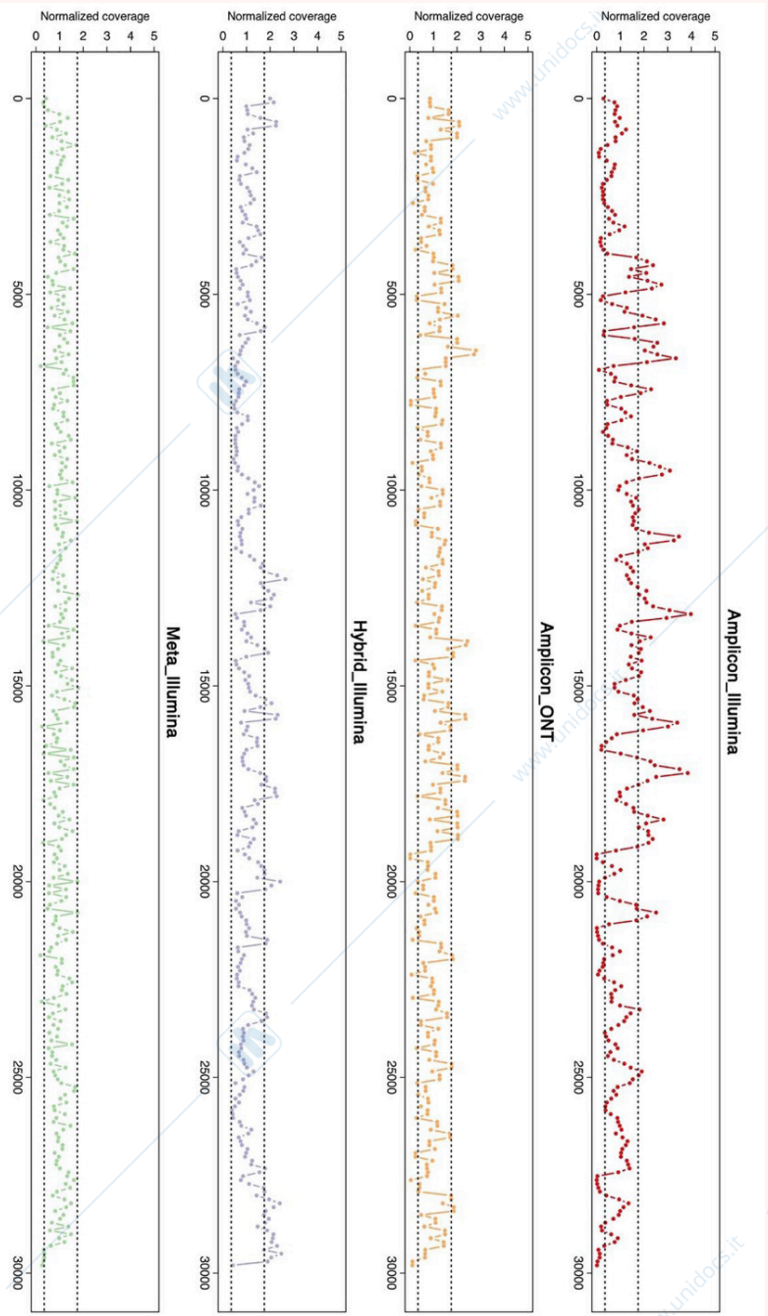
- Non richiede la retrotrascrizione ne l'amplificazione

## SVANTAGGI:

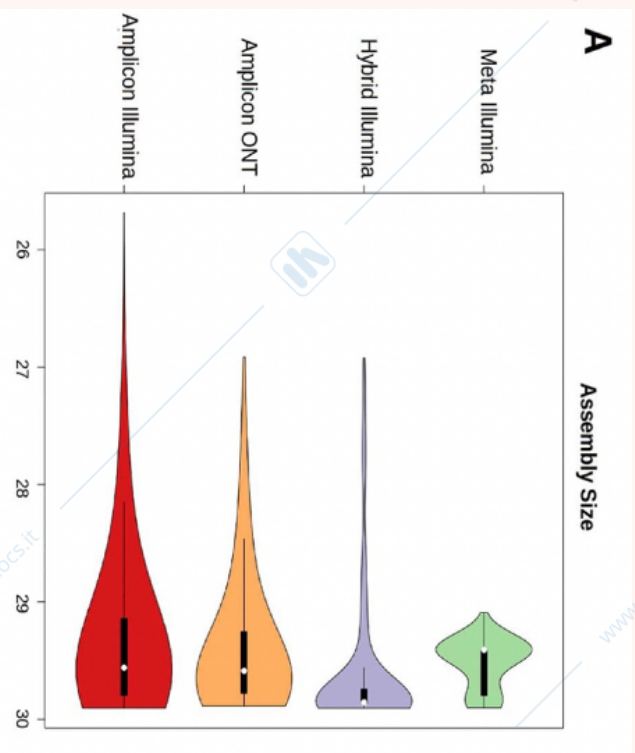
- Forniscono reads con un tasso di errore maggiore

	Direct RNA sequencing <sup>a</sup>
Goals	SARS-CoV-2 and host transcriptome and epitranscriptome
Co-infection detection	Yes
Minimum number of reads	0.5 M
Genome Coverage	≥99%
Accuracy in SNV identification	Low
Sample viral load (Ct) requested (ref Xiao)	<24-28
Sample RNA input (ng)	≥1000
Sample type	Viral cell cultures
Cost	High
NGS sequencing platforms	ONT

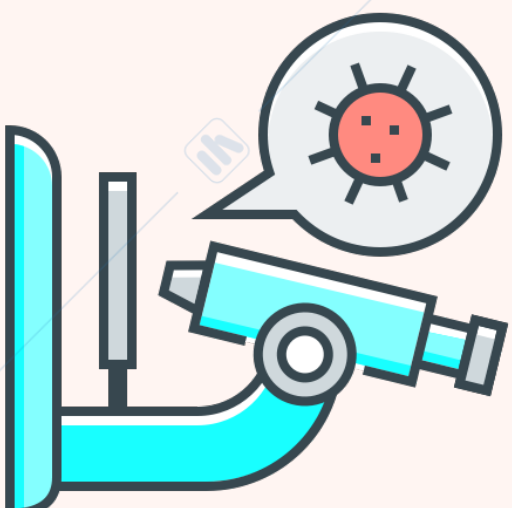
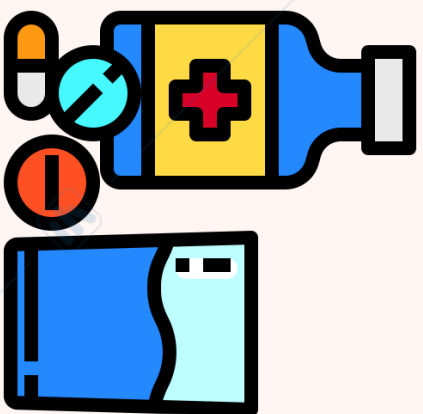
# ASSEMBLAGGIO GENOMICO



Library preparation	Sequencing technology
Amplicon	Illumina Oxford Nanopore
Hybrid capture	Illumina
Metatranscriptomics	Illumina



# DEPOSIZIONE E ACCESSO AI DATI



# GISAID EPICOV

## Contiene oltre 100 000 genomi di SARS-CoV-2

## Contiene pochi metadati

The screenshot displays the GISAID website interface. At the top, the GISAID logo is visible. The main navigation menu includes: About us, Database Features, Events, Collaborations, References, Registration, and Help. A search bar is located on the right side of the navigation menu.

The main content area is divided into several sections:

- In Focus:** A section titled "Additional closely related bat CoVs identified in Cambodia" featuring a phylogenetic tree and a map of Cambodia. The text states: "The Virology Unit of the Institut Pasteur du Cambodge has recently submitted the full genome sequences for 2 bat coronaviruses from samples taken from bats in 2010 in the Steung Treng province of Cambodia; they appear to join the list of coronaviruses that are closely related to the current pandemic virus." Below this, it mentions: "Other potential precursors known so far include bat coronaviruses from caves in Yunnan (Southern China), sampled in 2013 and 2019, and samples derived from pangolins in Southern China from 2017 and 2019." A citation is provided: "see: Vibol Hul et al (2021) A novel SARS-CoV-2 related coronavirus in bats from Cambodia".
- hCoV-19 Submission Tracking:** A section with two world maps showing submission locations.
- hCoV-19 Tracking of Variants:** A world map with red dots indicating the global distribution of hCoV-19 variants.
- Genomic epidemiology of hCoV-19:** A section showing a detailed phylogenetic tree and a map of the world with colored dots representing different genetic clusters.
- hCoV-19 data sharing via GISAID:** A section stating "1,563,988 submissions".
- Enabled by data shared via GISAID:** A section listing "SARS-CoV-2 variants and mutations of interest" and "Covid-Miner Regina Elena National Cancer Institute (Italy) Detect the emergence and reemerging GISAID Resources".
- Public-Private Partnerships of the GISAID Initiative:** A section with the text: "The GISAID Initiative involves public-private partnerships between the Initiative's administrative..."
- Free Access Credentials:** A section with the text: "Register here and join thousands of researchers around the globe."

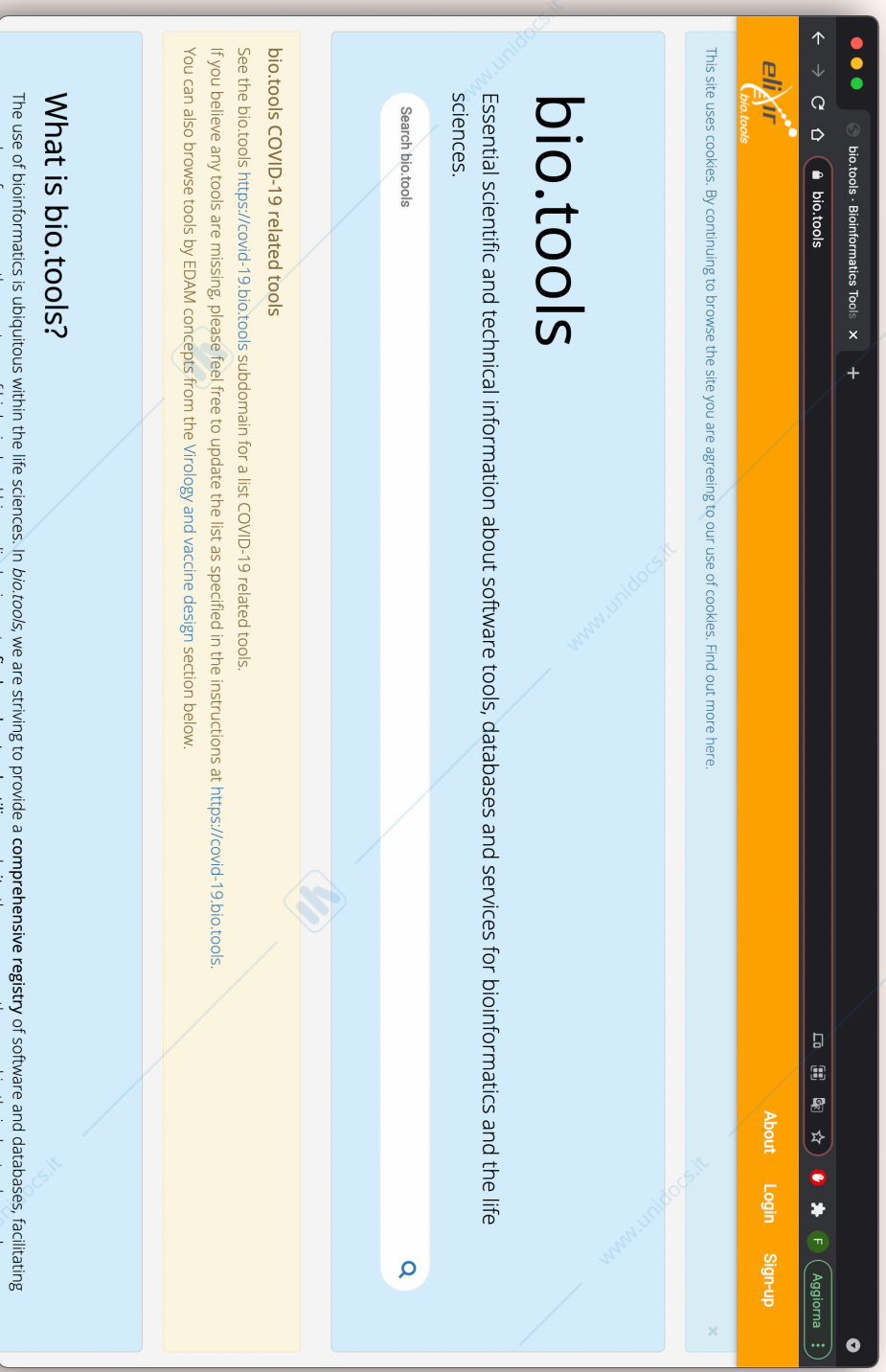
# RESEARCH DATA ALLIANCE

Contiene dati più conformi  
allo standard Findability  
Accessibility Interoperability  
Reuse (FAIR)

The screenshot shows the RDA Research Data Alliance website. The main navigation bar includes: ABOUT RDA, GET INVOLVED, GROUPS, RECOMMENDATIONS & OUTPUTS, RDA FOR DISCIPLINES, PLENARIES & EVENTS, NEWS & MEDIA. The header features the RDA logo and the tagline "Building the social and technical bridges to enable open sharing and re-use of data". Key statistics are displayed: O&A Members (61), MEMBERSHIP (11795), and RDA Groups (96). A central banner for the "Community Review: Request for Comments Open!" is overlaid on a background image of penguins. The banner text reads: "Data Versioning Interest Group Charter - Deadline June 7" and "CURE-FAIR WG: Challenges of Curating for Reproducible and FAIR Research Output - Deadline May 21". A sidebar on the left contains "NEWS & EVENTS" with a link to "10/05/2021 - Register for Info" and "REQUEST FOR COMMENTS ONGOING!". A "FOLLOW US" section at the bottom right shows "Tweets by @resdatail".

# BIO.TOOLS

Contiene dati riguardo  
l'analisi bioinformatica e i  
workflow usati

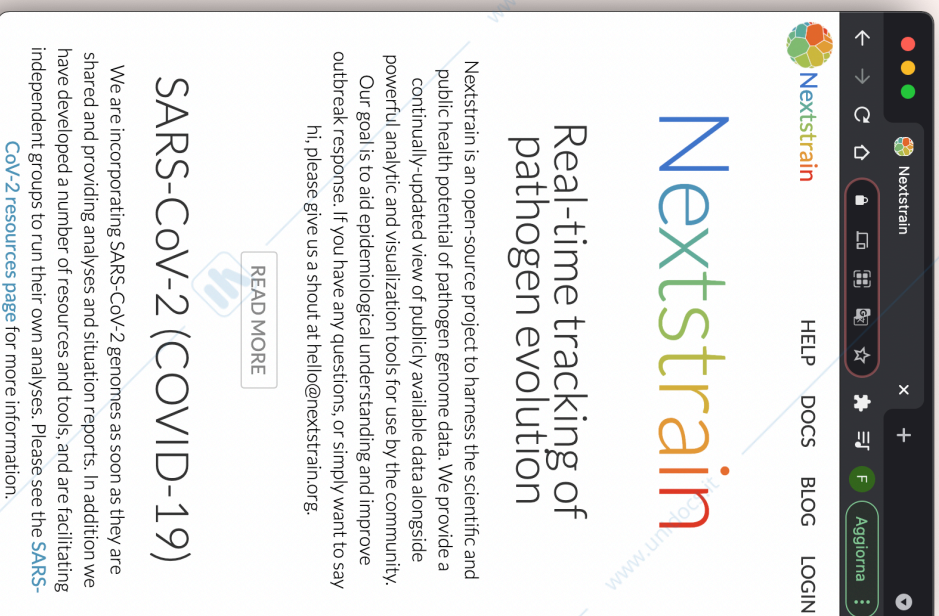


# NEXTSTRAIN

# E

# HYPHY

**Permettono il monitoraggio in tempo reale dell'evoluzione di SARS-CoV-2**



The screenshot shows the Nextstrain website homepage. The browser's address bar displays "Nextstrain". The navigation menu includes "HELP", "DOCS", "BLOG", and "LOGIN". The main heading reads "Real-time tracking of pathogen evolution". Below this, a paragraph explains that Nextstrain is an open-source project for harnessing scientific and public health data. A "READ MORE" button is visible at the bottom of the text block.

**Nextstrain**

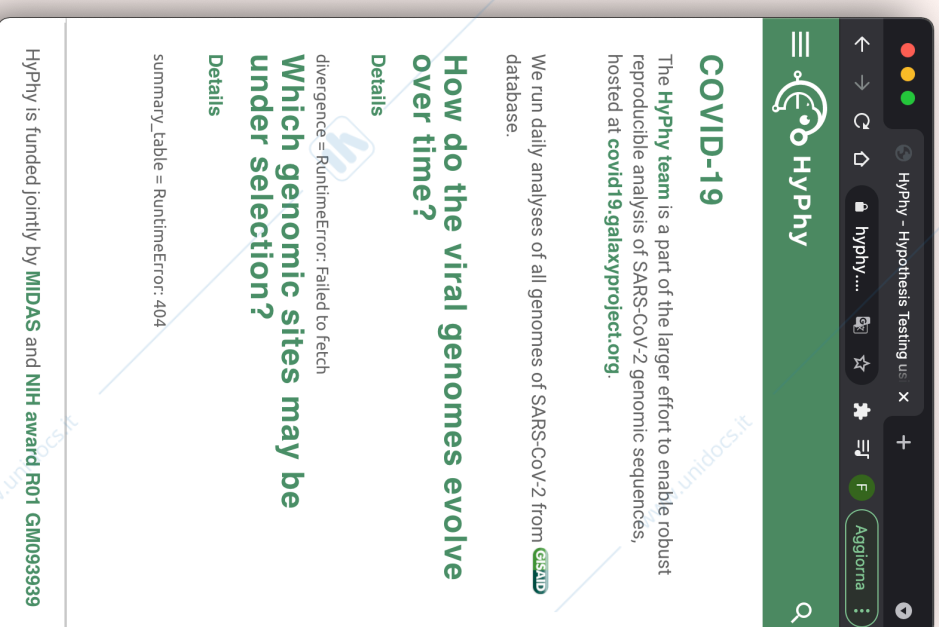
## Real-time tracking of pathogen evolution

Nextstrain is an open-source project to harness the scientific and public health potential of pathogen genome data. We provide a continually-updated view of publicly available data alongside powerful analytic and visualization tools for use by the community. Our goal is to aid epidemiological understanding and improve outbreak response. If you have any questions, or simply want to say hi, please give us a shout at [hello@nextstrain.org](mailto:hello@nextstrain.org).

[READ MORE](#)

### SARS-CoV-2 (COVID-19)

We are incorporating SARS-CoV-2 genomes as soon as they are shared and providing analyses and situation reports. In addition we have developed a number of resources and tools, and are facilitating independent groups to run their own analyses. Please see the [SARS-CoV-2 resources page](#) for more information.



The screenshot shows the HyPhy website. The browser's address bar displays "HyPhy - Hypothesis Testing". The navigation menu includes "HOME", "ABOUT", "FAQ", "CONTACT", and "DONATE". The main heading reads "COVID-19". Below this, a paragraph explains that the HyPhy team is a part of the larger effort to enable robust reproducible analysis of SARS-CoV-2 genomic sequences. A "READ MORE" button is visible at the bottom of the text block.

**HyPhy**

## COVID-19

The **HyPhy team** is a part of the larger effort to enable robust reproducible analysis of SARS-CoV-2 genomic sequences, hosted at [covid19.galaxyproject.org](https://covid19.galaxyproject.org).

We run daily analyses of all genomes of SARS-CoV-2 from [GISAID](https://gisaid.org) database.

### How do the viral genomes evolve over time?

**Details**

divergence = RuntimeError: Failed to fetch

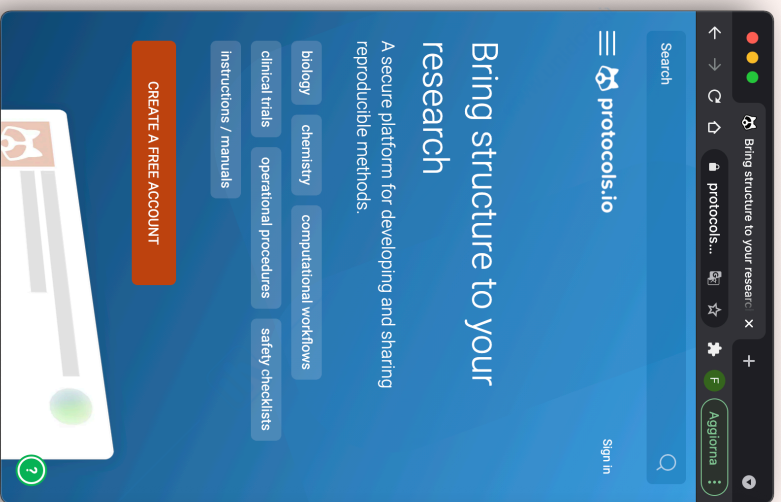
### Which genomic sites may be under selection?

**Details**

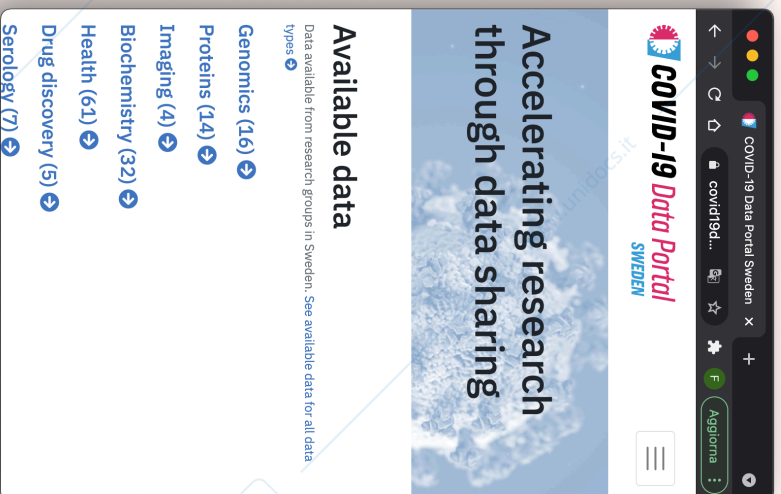
summary\_table = RuntimeError: 404

HyPhy is funded jointly by **MIDAS** and **NIH award R01 GM093939**

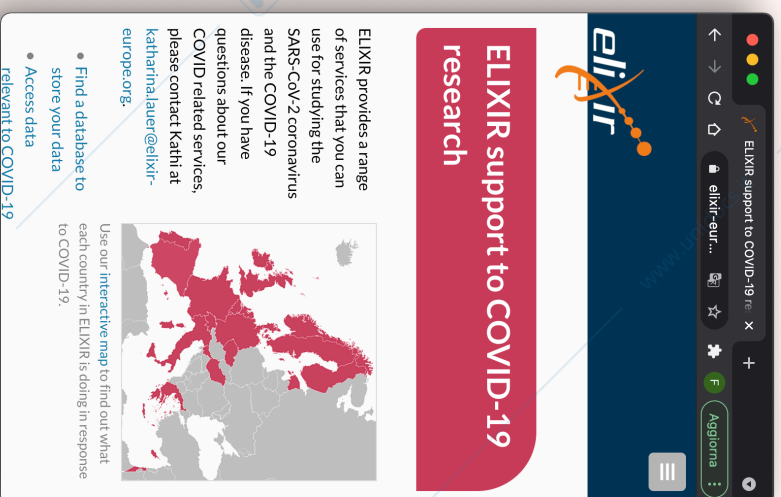
# PROTOCOLS.IO



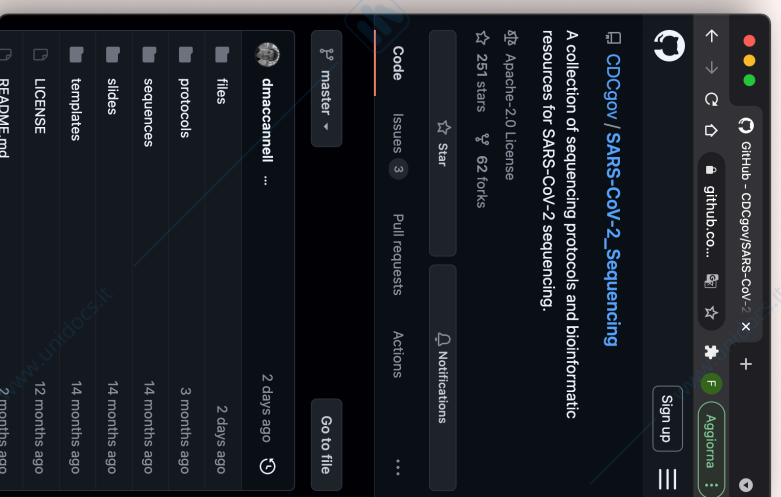
# COVID-19 DATA PORTAL



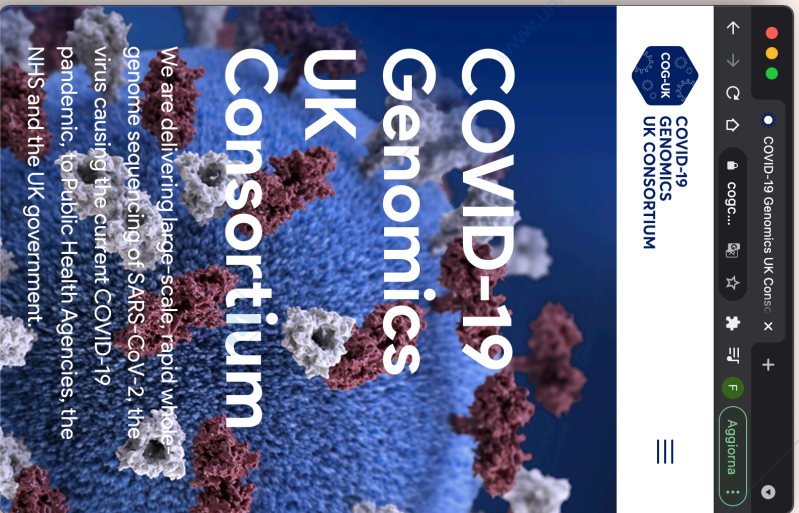
# ELIXIR



# GITHUB



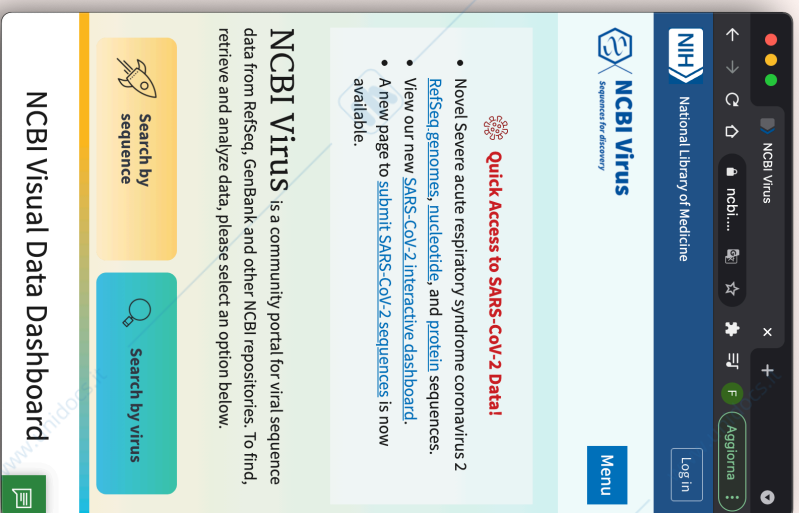
# COG-UK

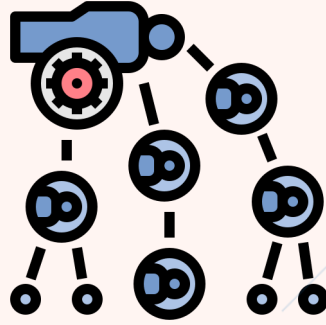
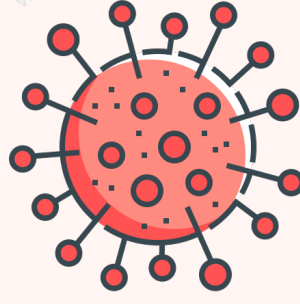
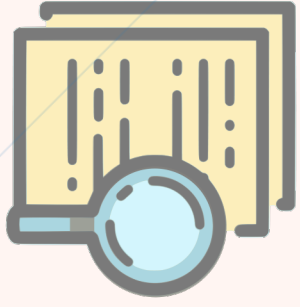


# GISAID EPICOV



# NCBI VIRUS





# CONCLUSIONI

