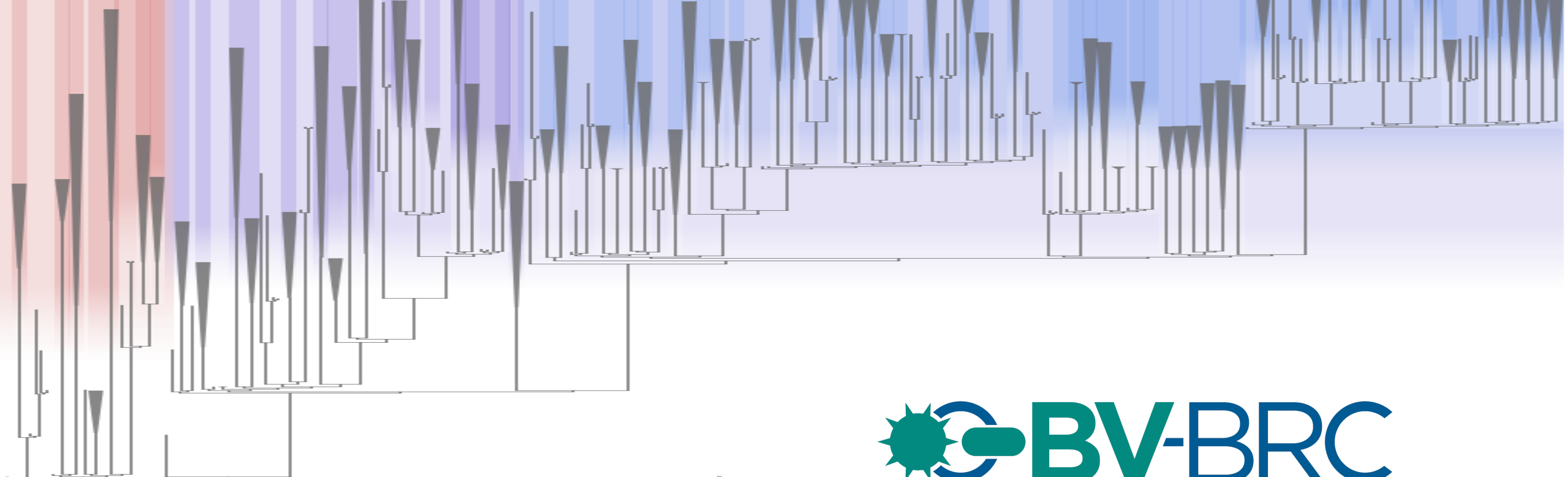


# Perspectives on the development of the Pango system and software

Áine O'Toole

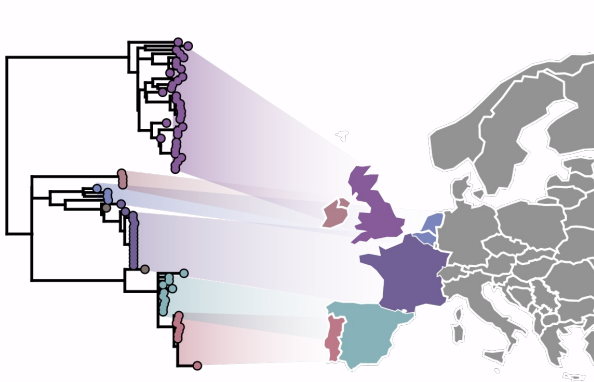
*Viral Sub-species Classification Workshop, BV-BRC*

2024-04-09

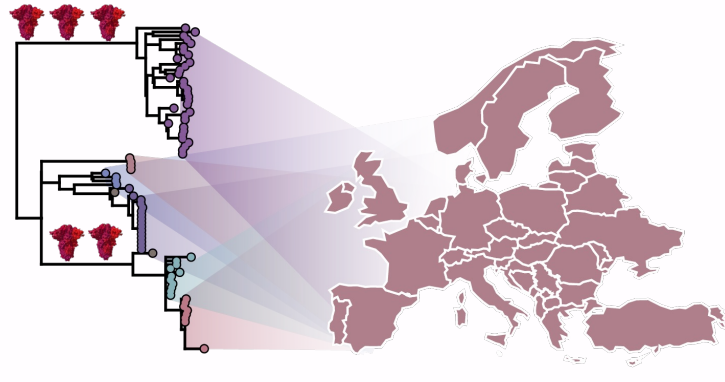


Thank you!

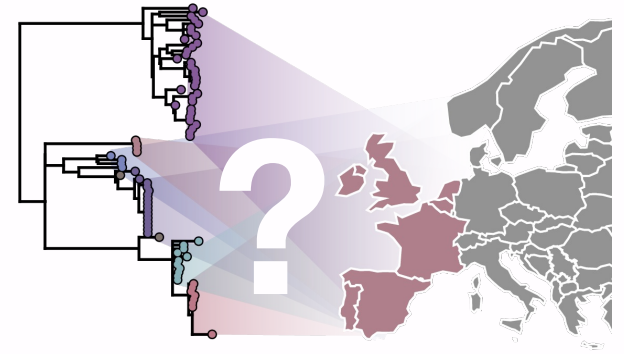




2020



2021-2024



2024



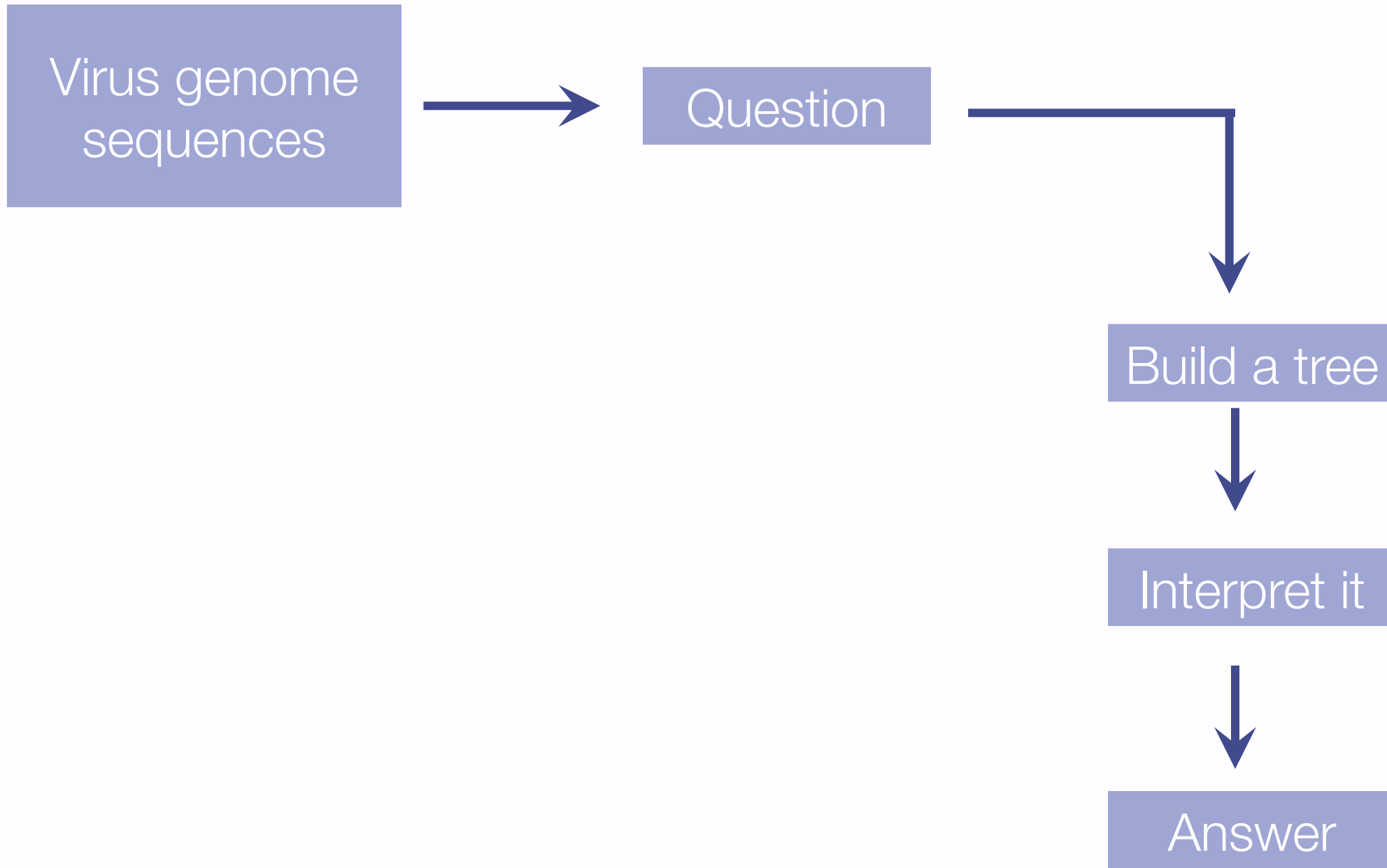
Rachel Colquhoun

Virus genome  
sequences



Question





Virus genome sequences



Question



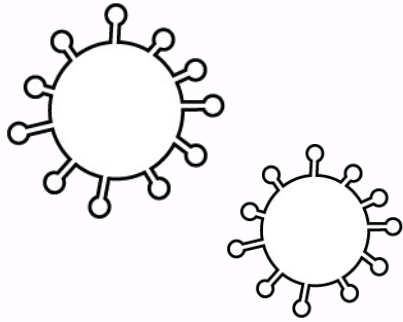
Build a tree



Interpret it



Answer



Virus genome sequences



Question



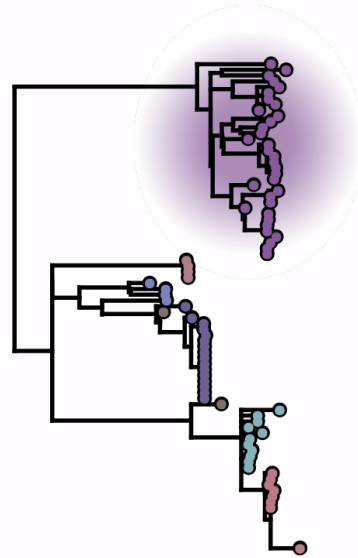
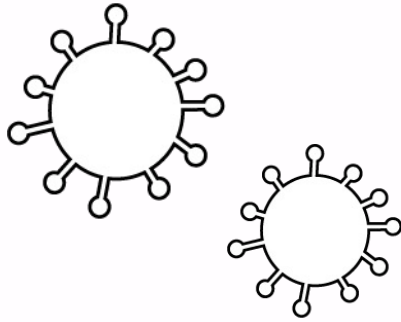
Build a tree



Interpret it

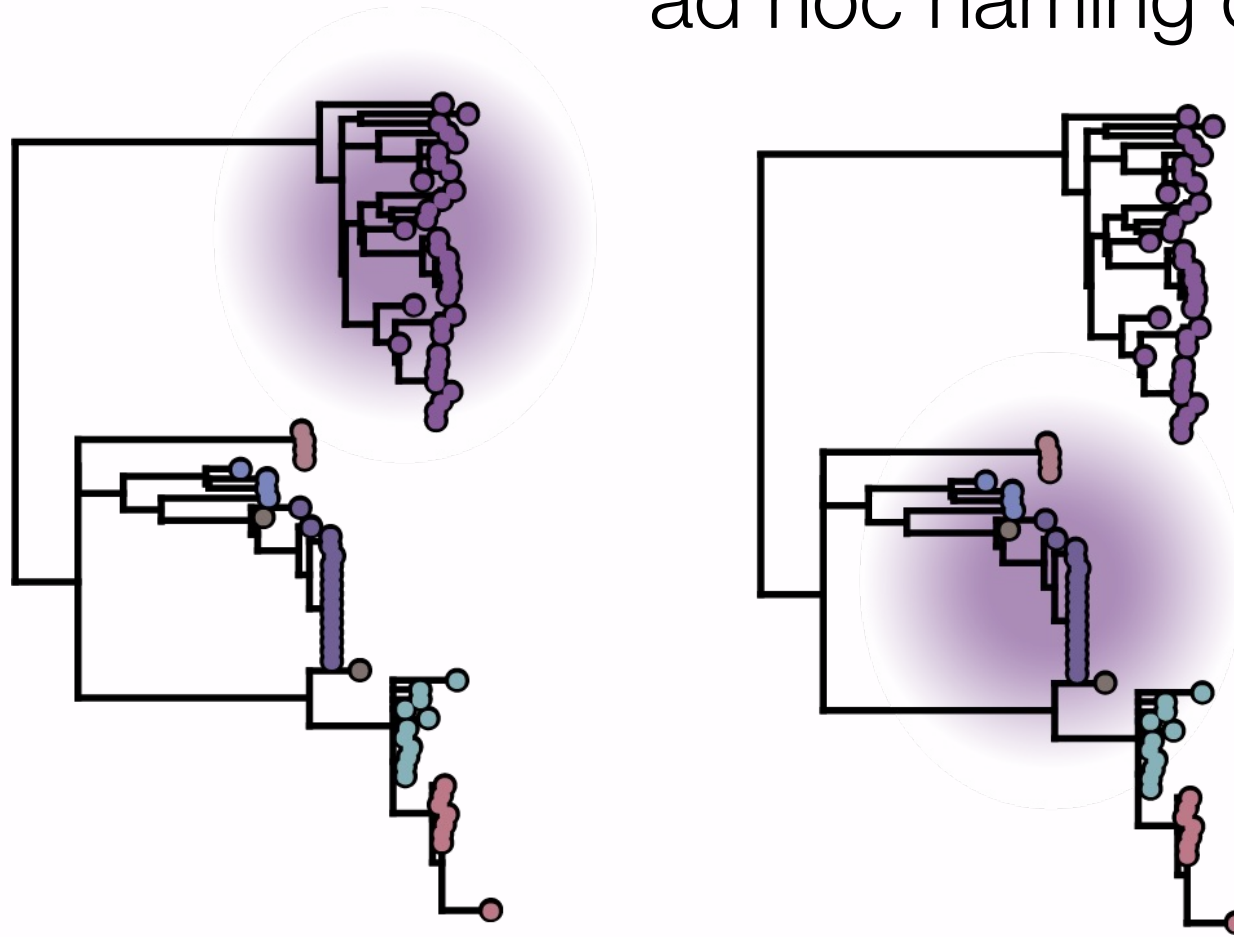


Answer

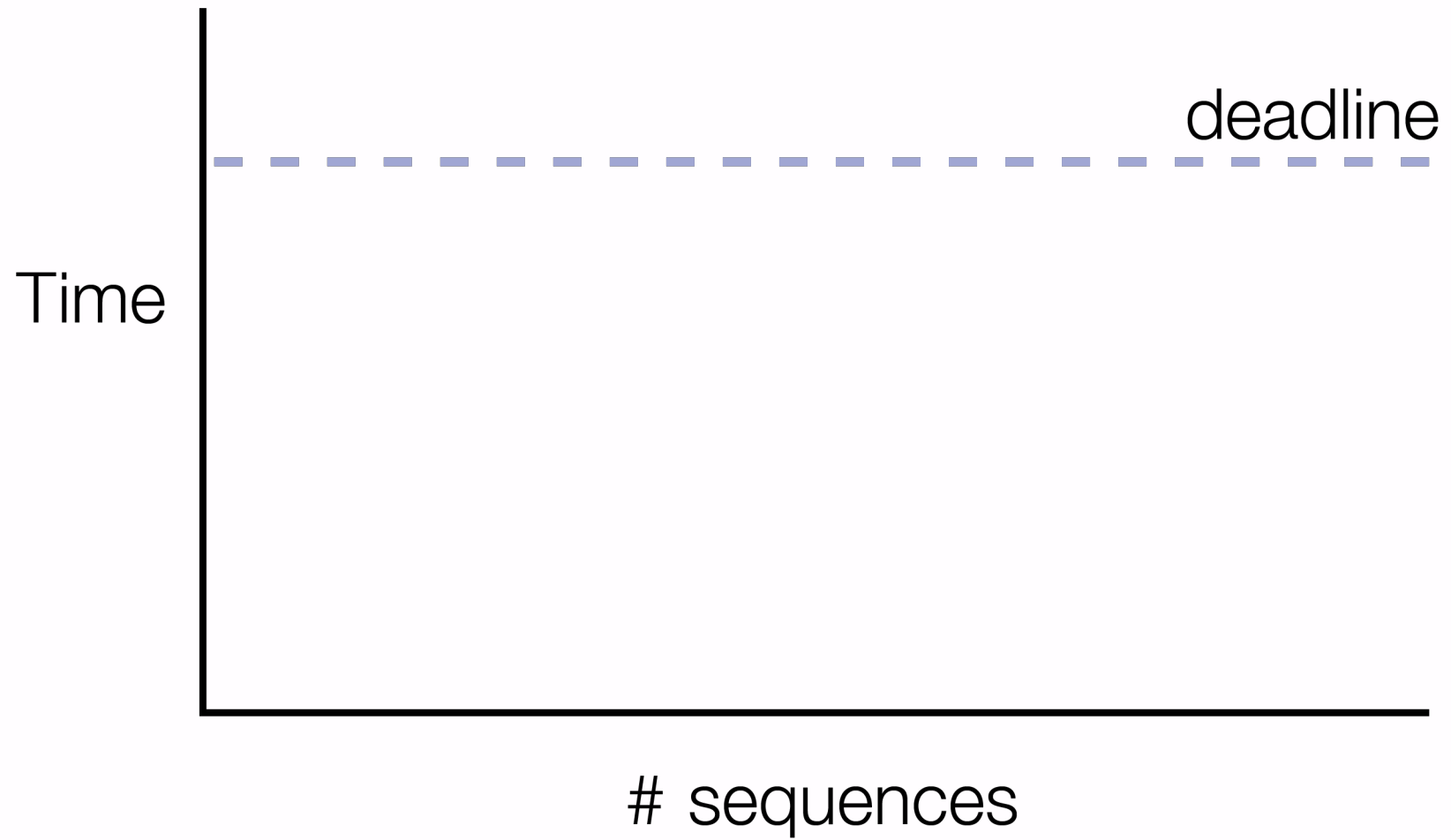


2020

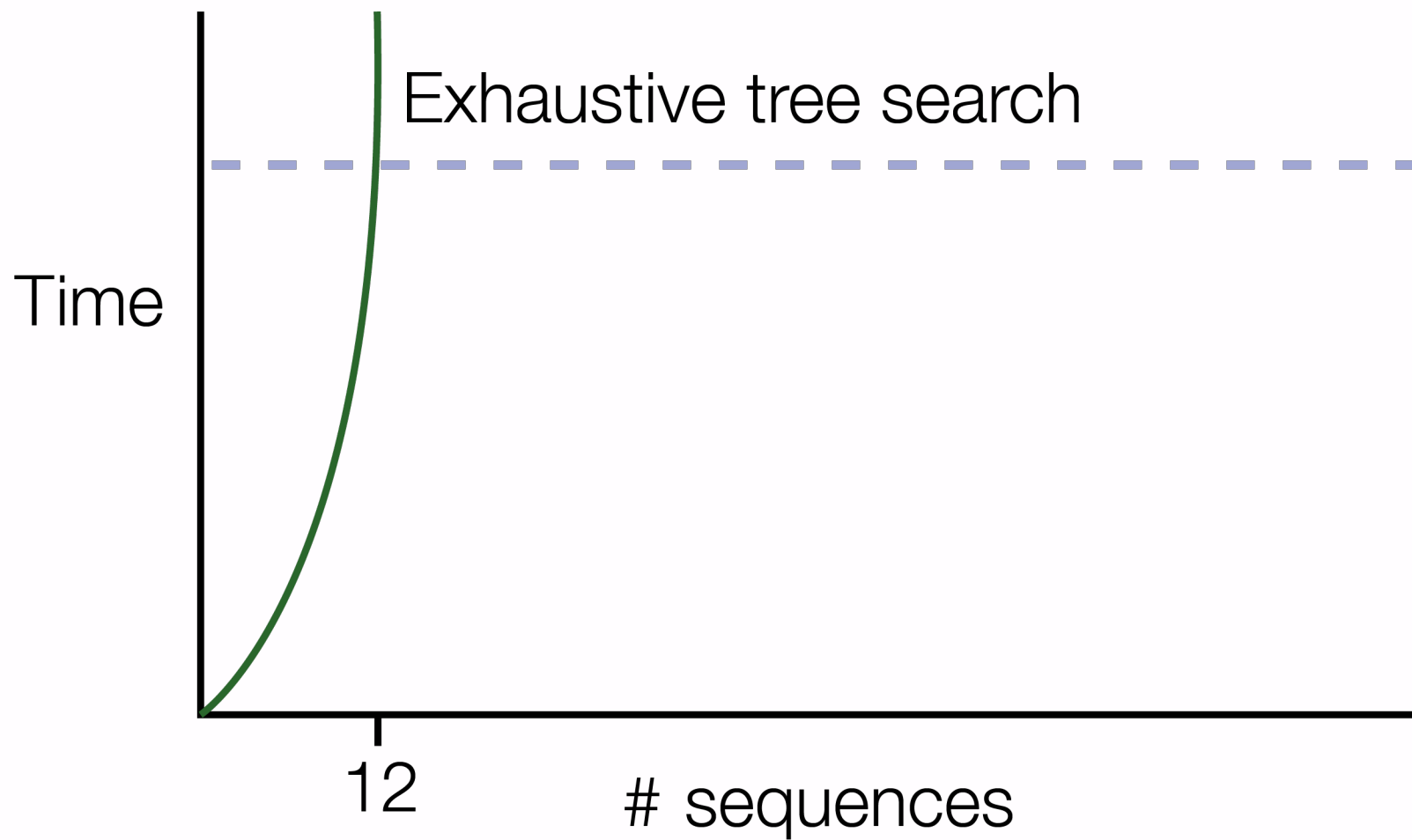
Early studies/ reports had  
ad hoc naming of clusters



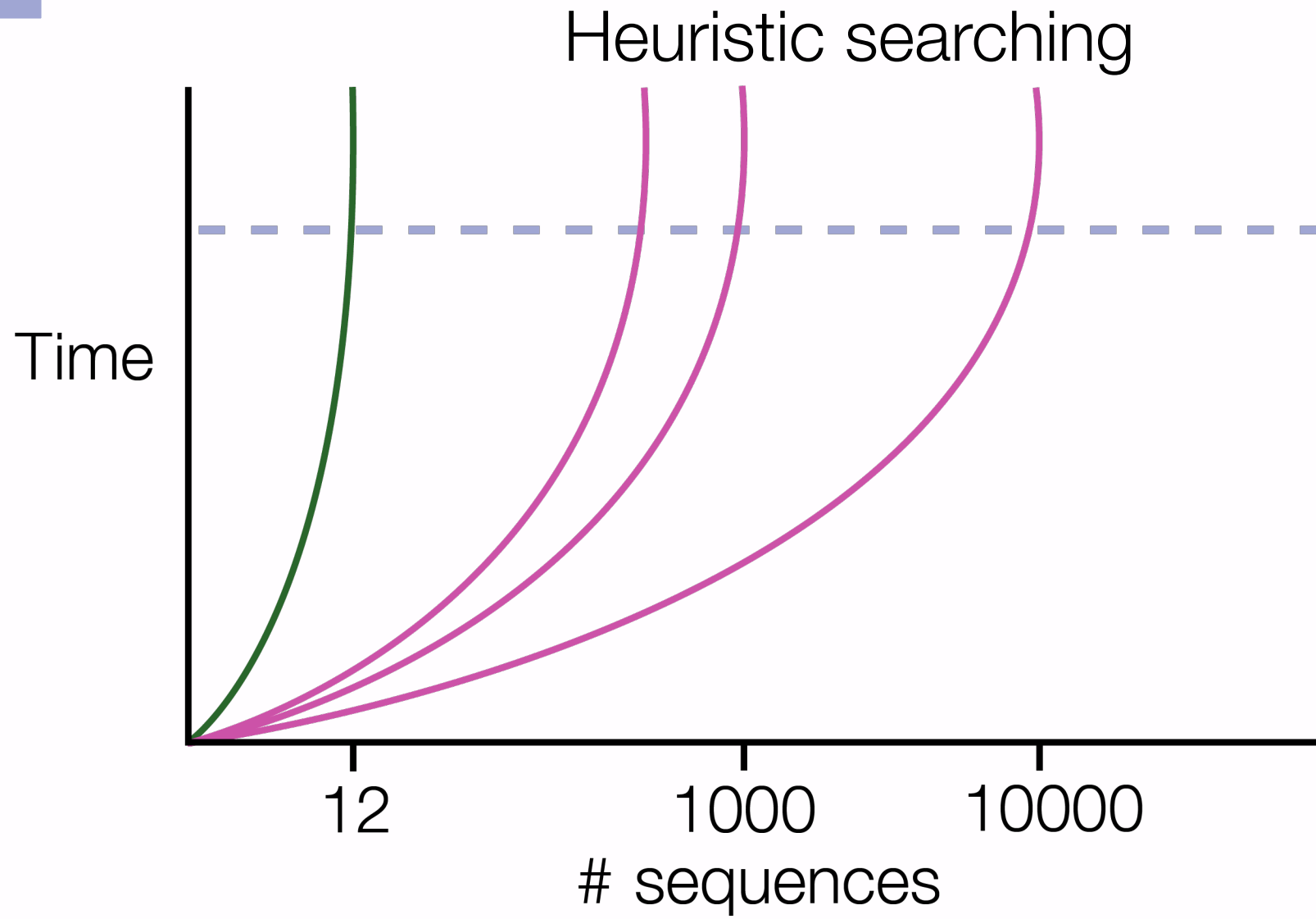
# Question



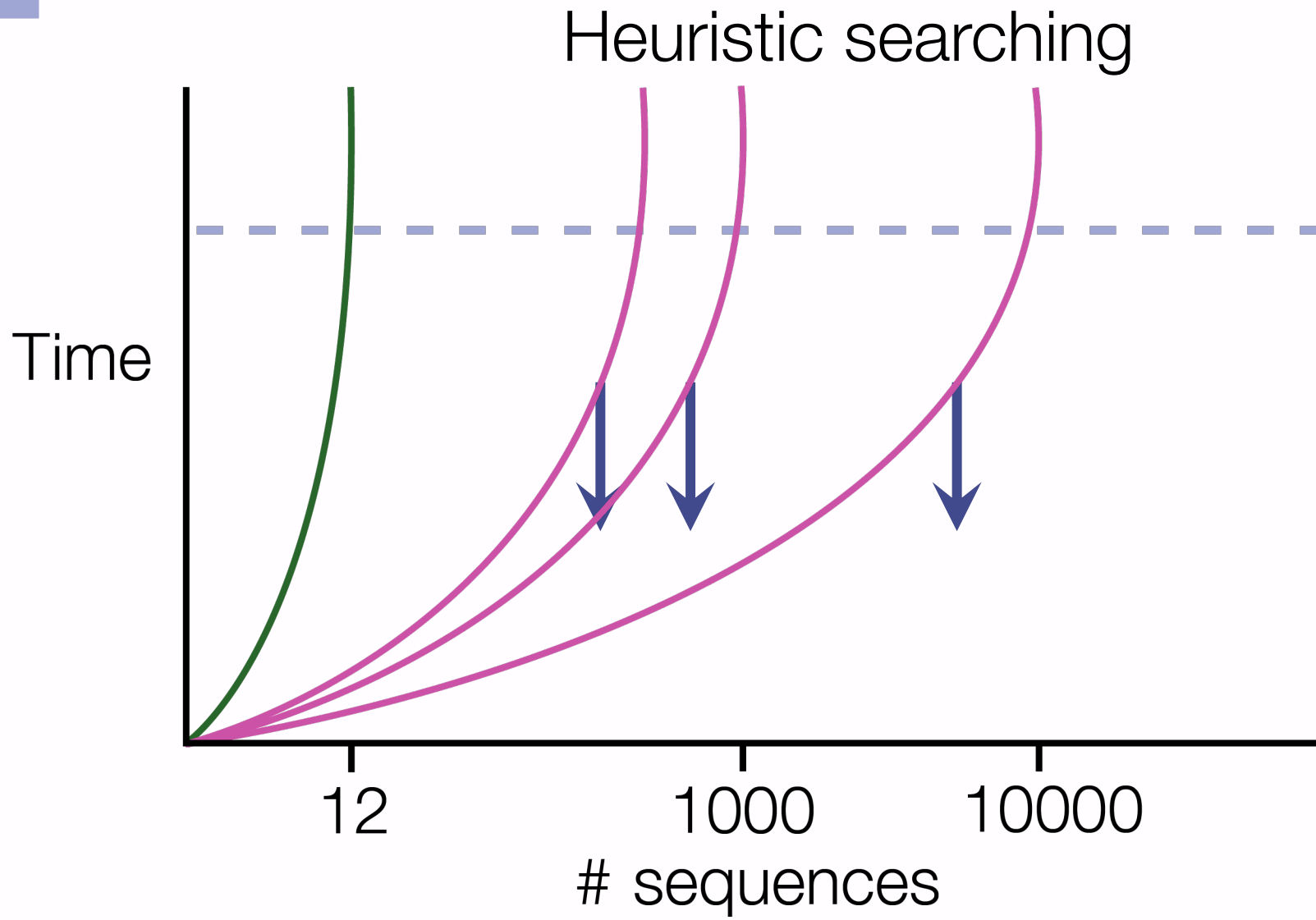
# Question



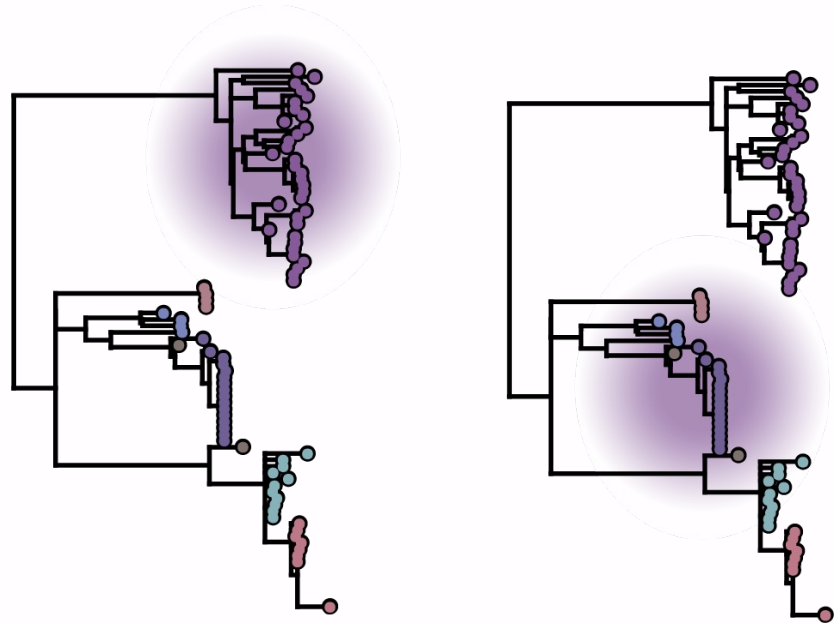
# Question



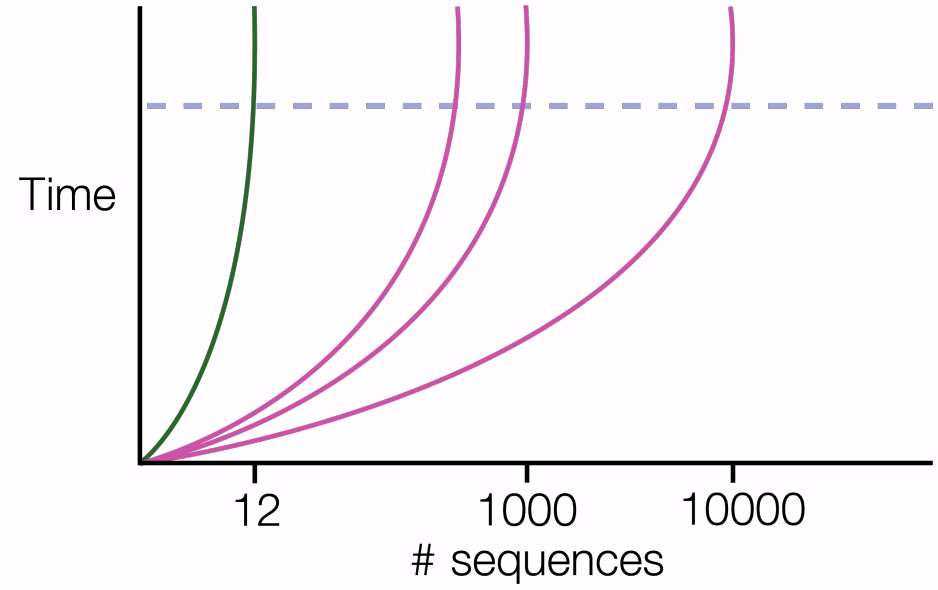
# Question







Consistent communication



Alleviate analytical bottleneck

2020



Oli Pybus



Andrew  
Rambaut

**nature microbiology**

# A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology

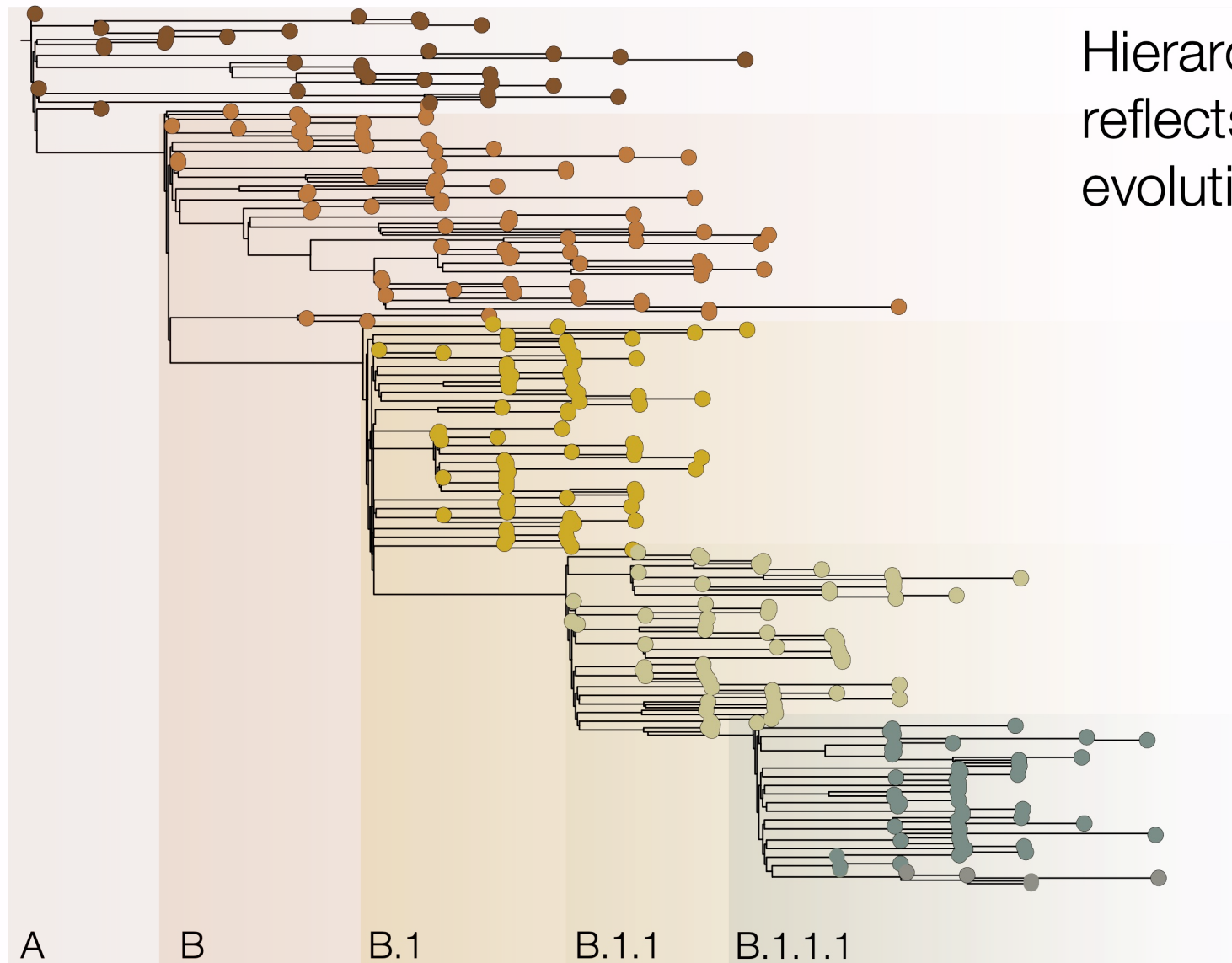
Andrew Rambaut , Edward C. Holmes , Áine O'Toole, Verity Hill, John T. McCrone, Christopher Ruis, Louis du Plessis & Oliver G. Pybus 

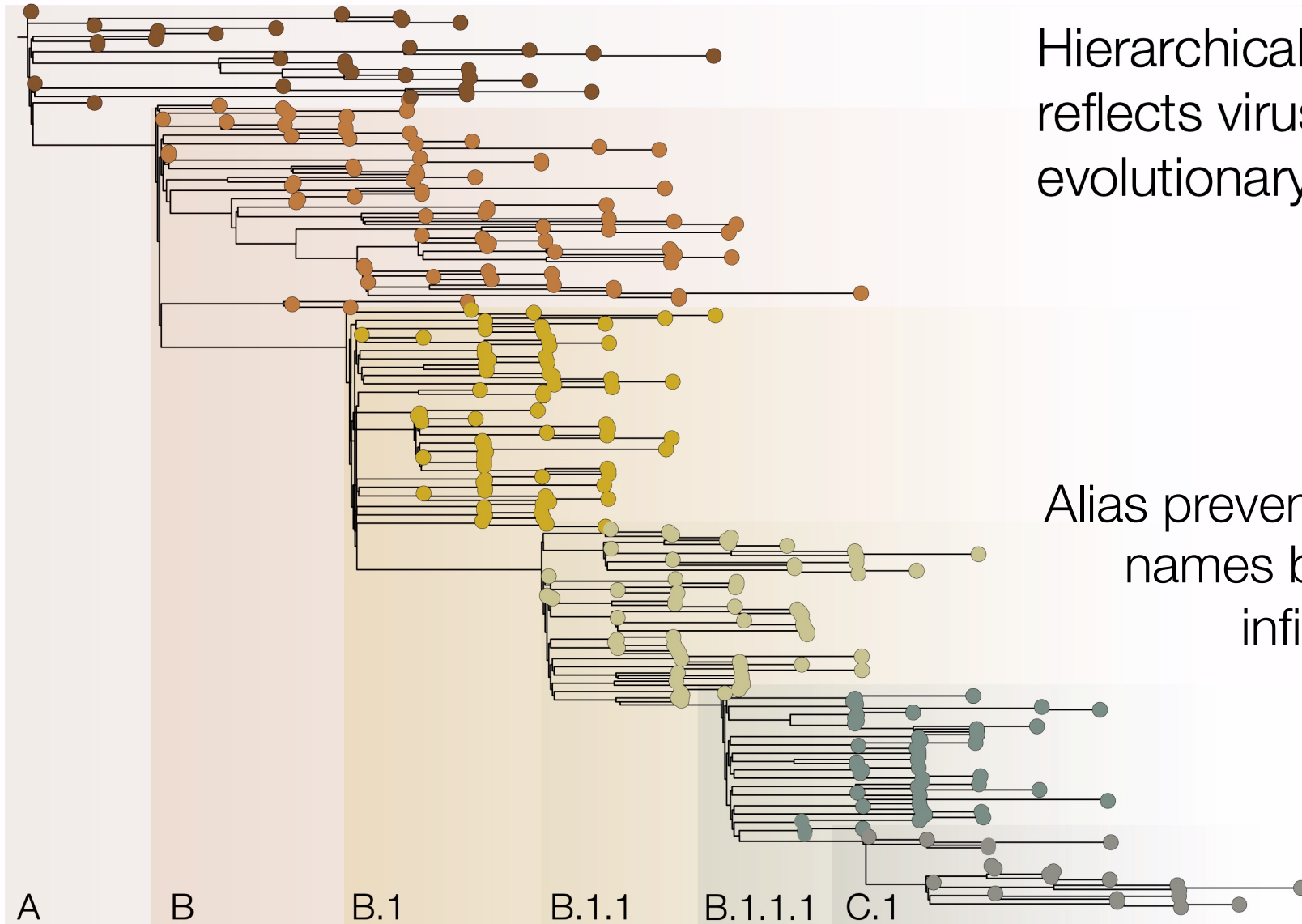
*Nature Microbiology* **5**, 1403–1407(2020) | [Cite this article](#)

**42k** Accesses | **216** Citations | **1071** Altmetric | [Metrics](#)



Hierarchical system  
reflects virus  
evolutionary history

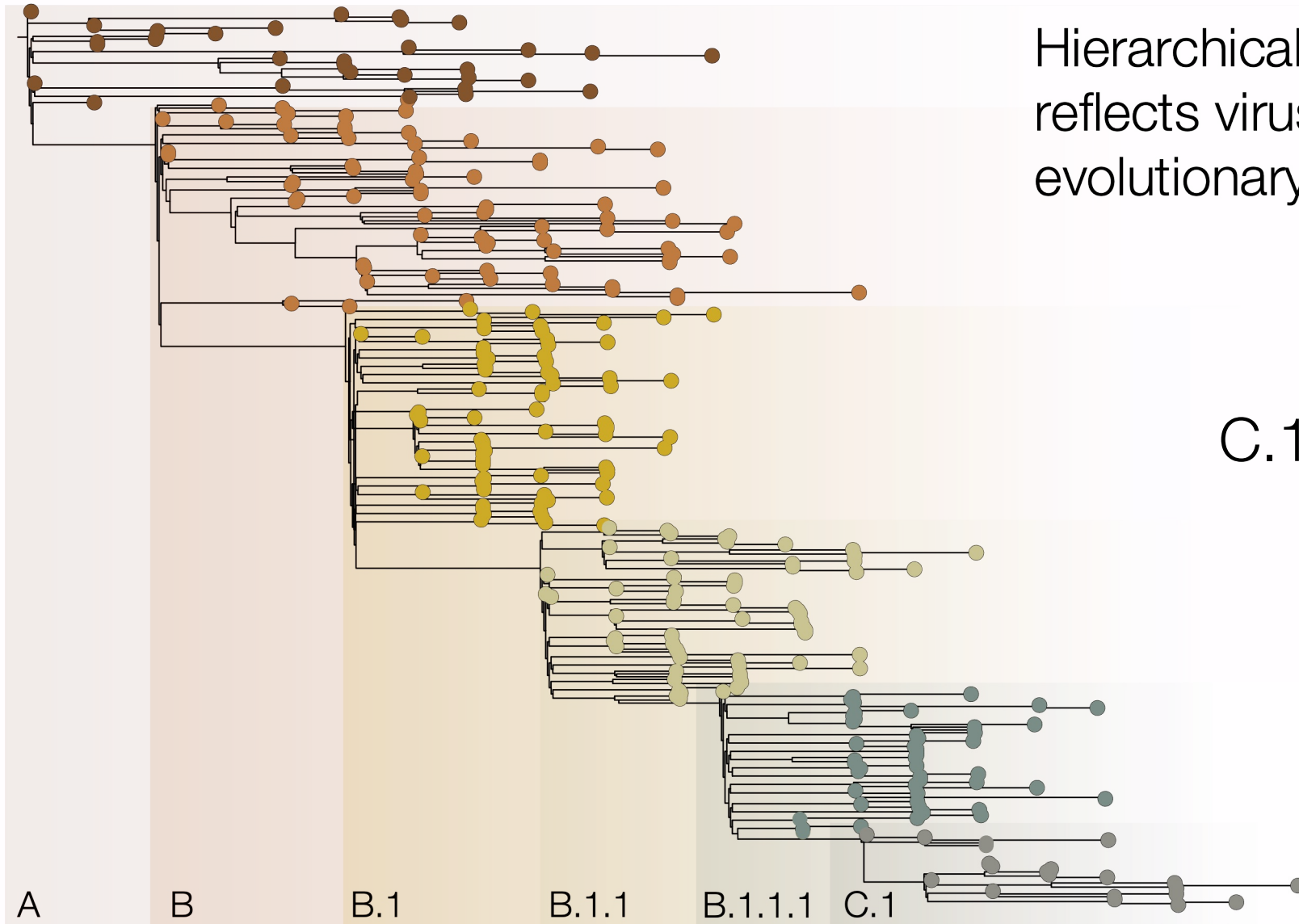




Hierarchical system  
reflects virus  
evolutionary history

Alias prevents lineage  
names becoming  
infinitely long

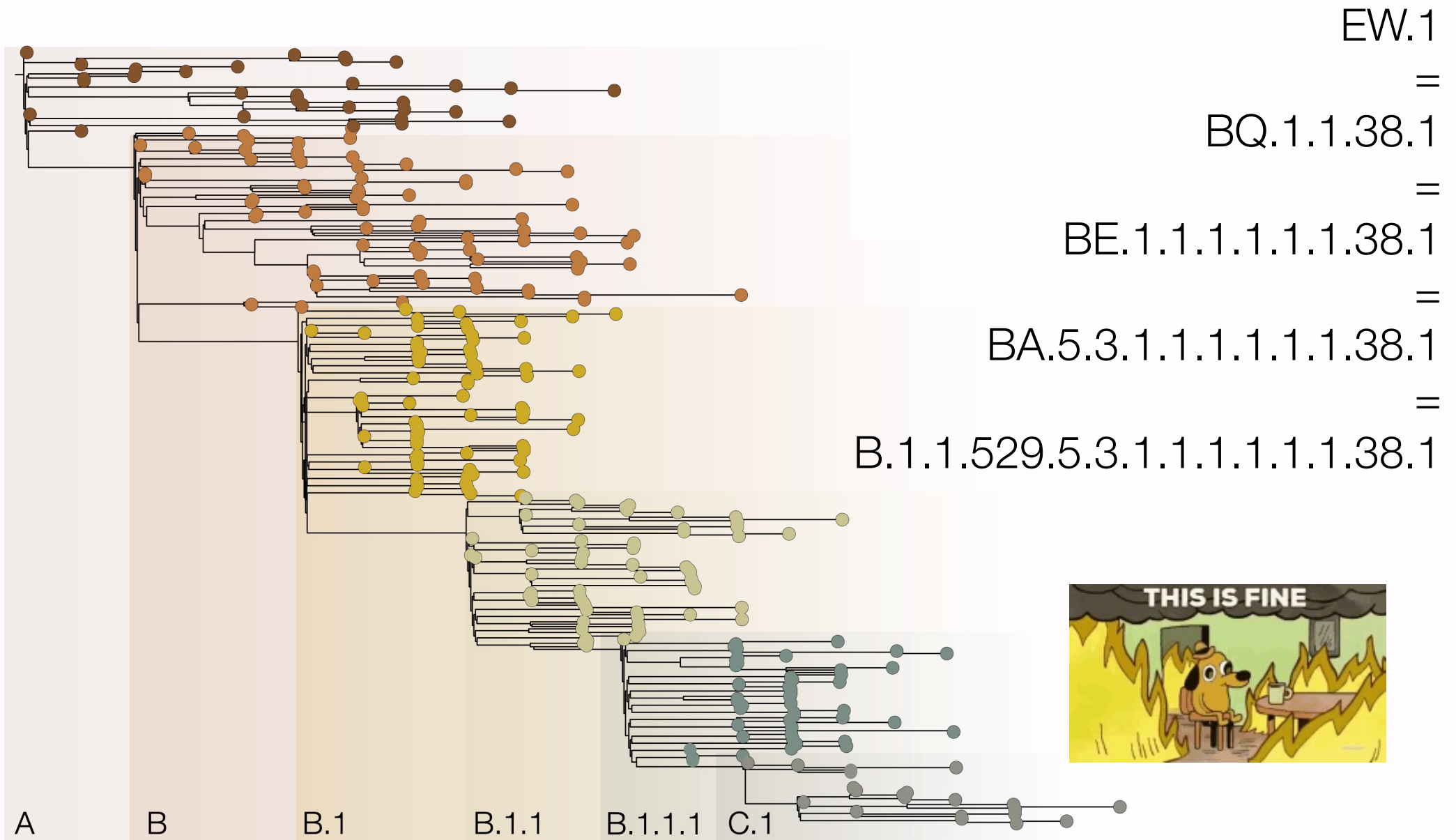




Hierarchical system  
reflects virus  
evolutionary history

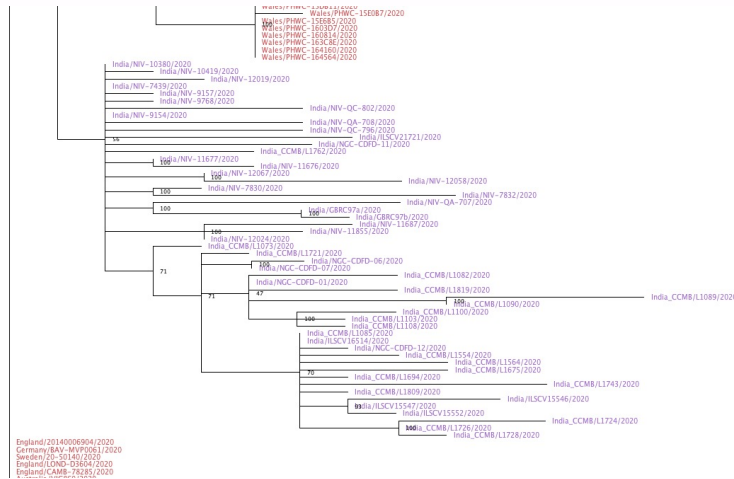
C.1 = B.1.1.1.1



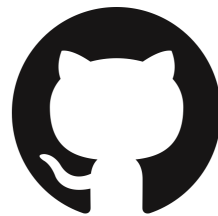


April 2020

19 lineages  
2,683 sequences



5<sup>th</sup> April 2020 initial lineages on GitHub

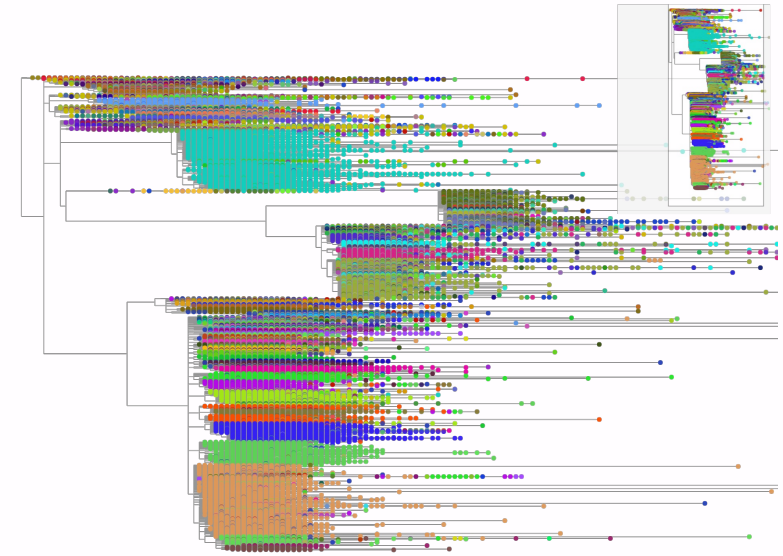
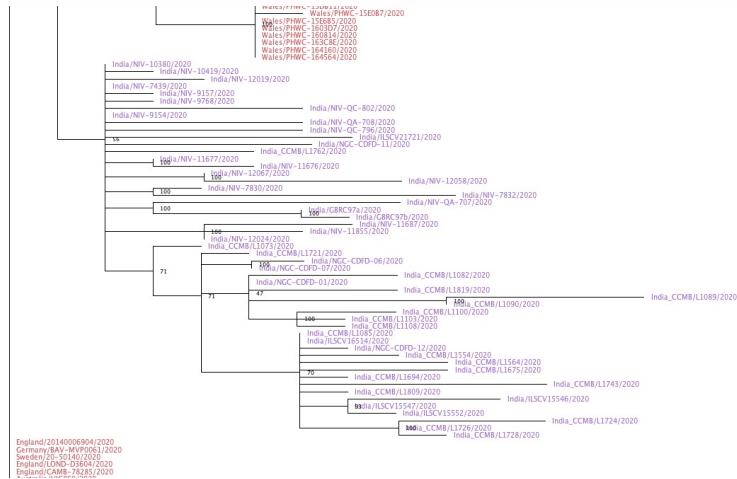


- |       |      |
|-------|------|
| A     | B    |
| A.1   | B.1  |
| A.1.1 | B.2  |
| A.1.2 | B.3  |
| A.2   | B.4  |
| A.3   | B.5  |
| A.4   | B.6  |
| A.5   | B.7  |
|       | B.8  |
|       | B.9  |
|       | B.10 |

April 2020



April 2024



19 lineages  
2,683 sequences

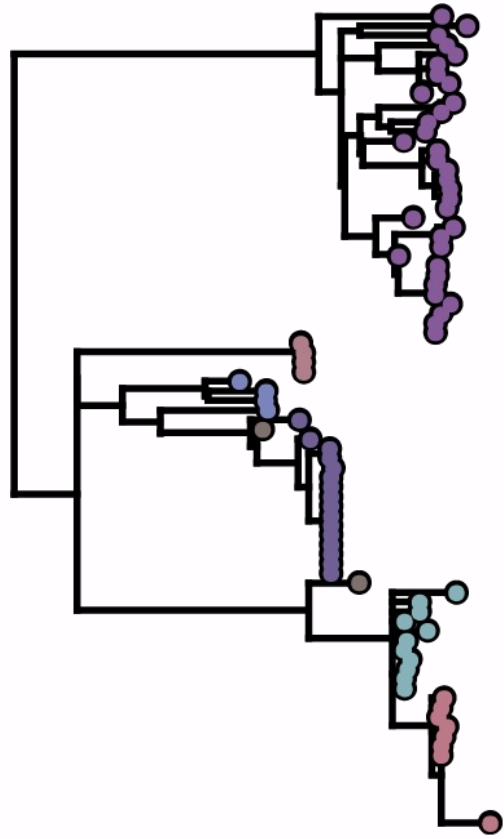
Annotated on a tree by hand

>4K lineages  
>12 million sequences

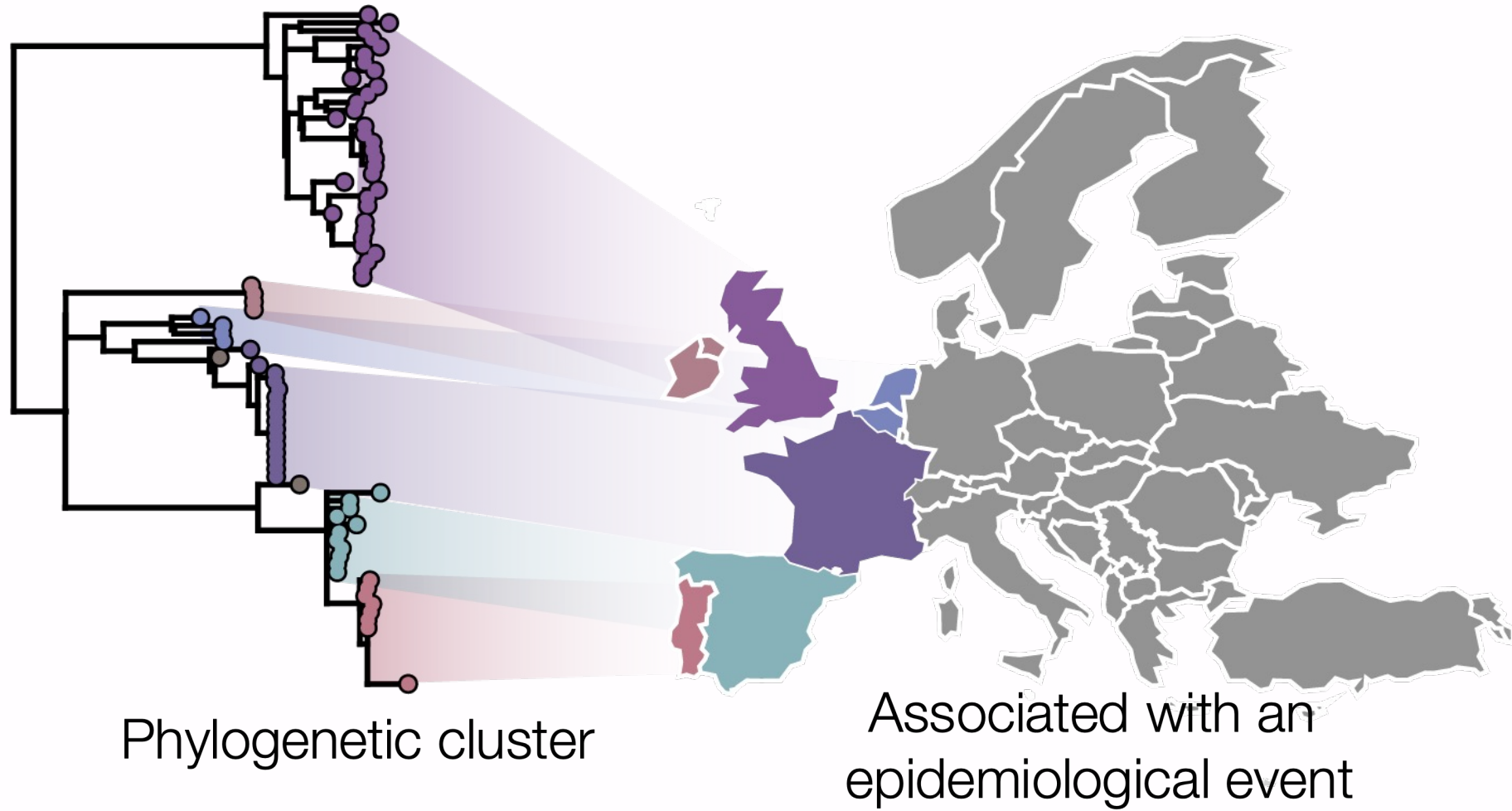
Still a largely manual process



2020



2020



2020

2021

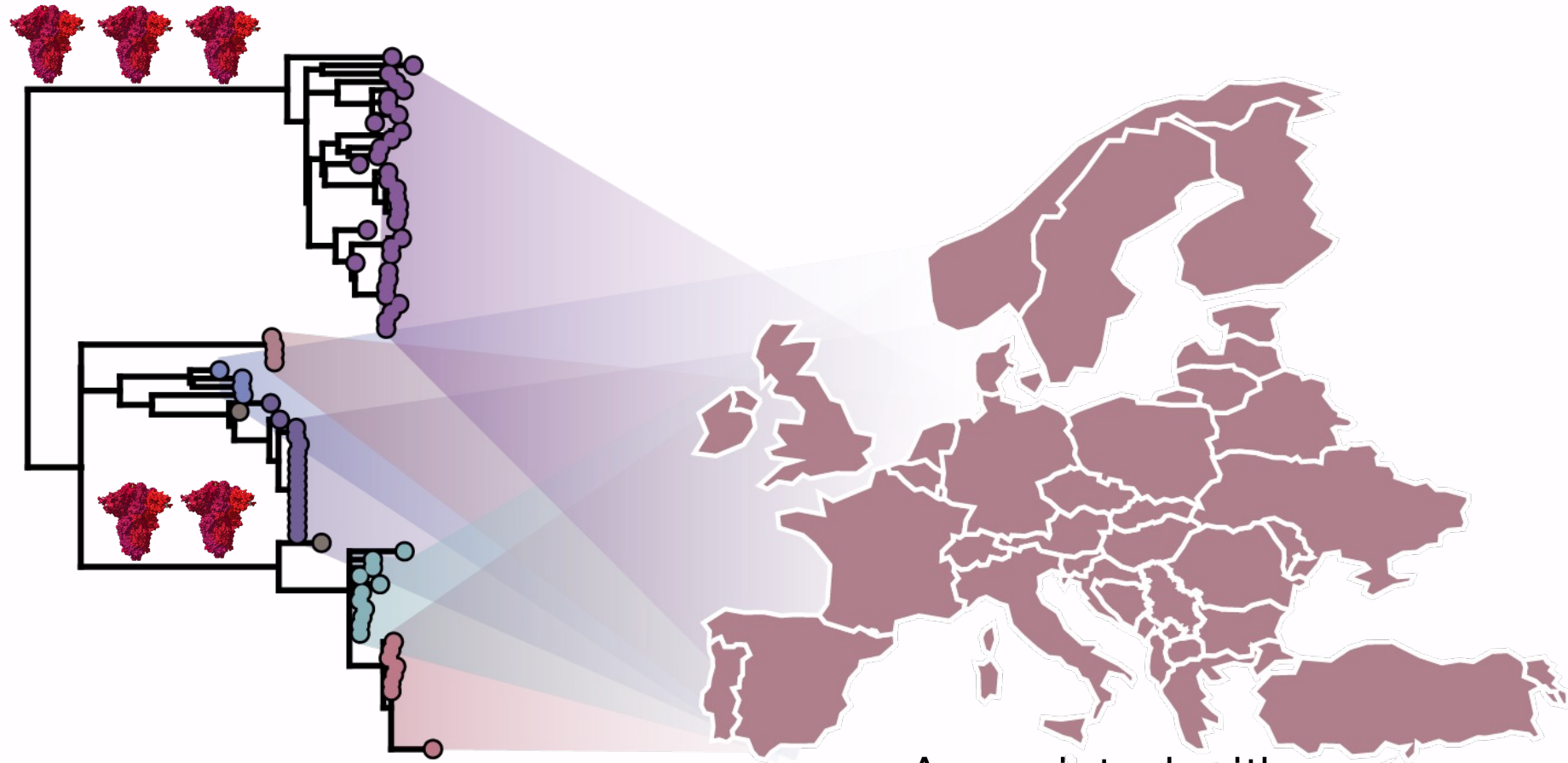


Phylogenetic cluster

Associated with an  
epidemiological event

2020

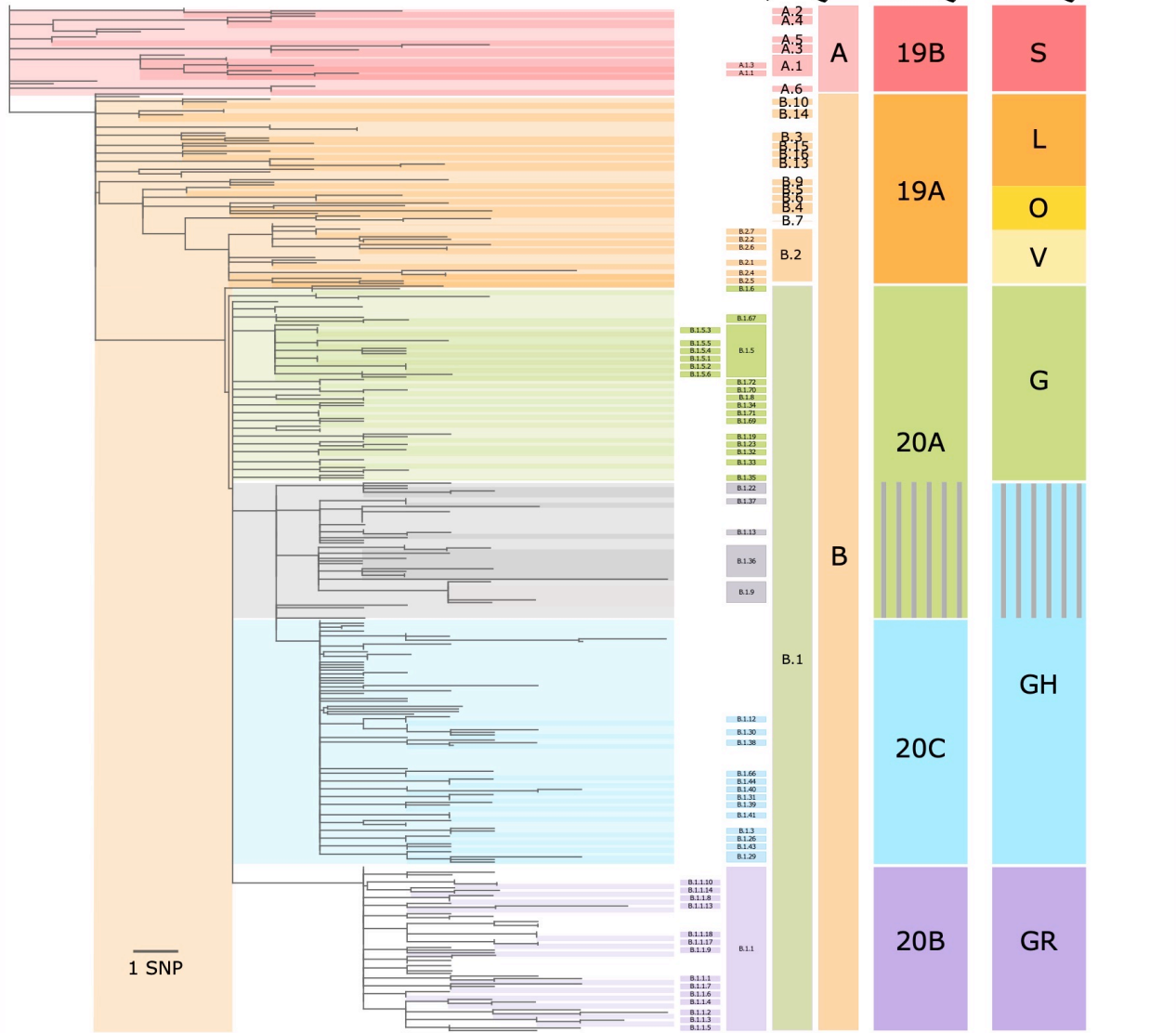
2021-2024



Phylogenetic cluster

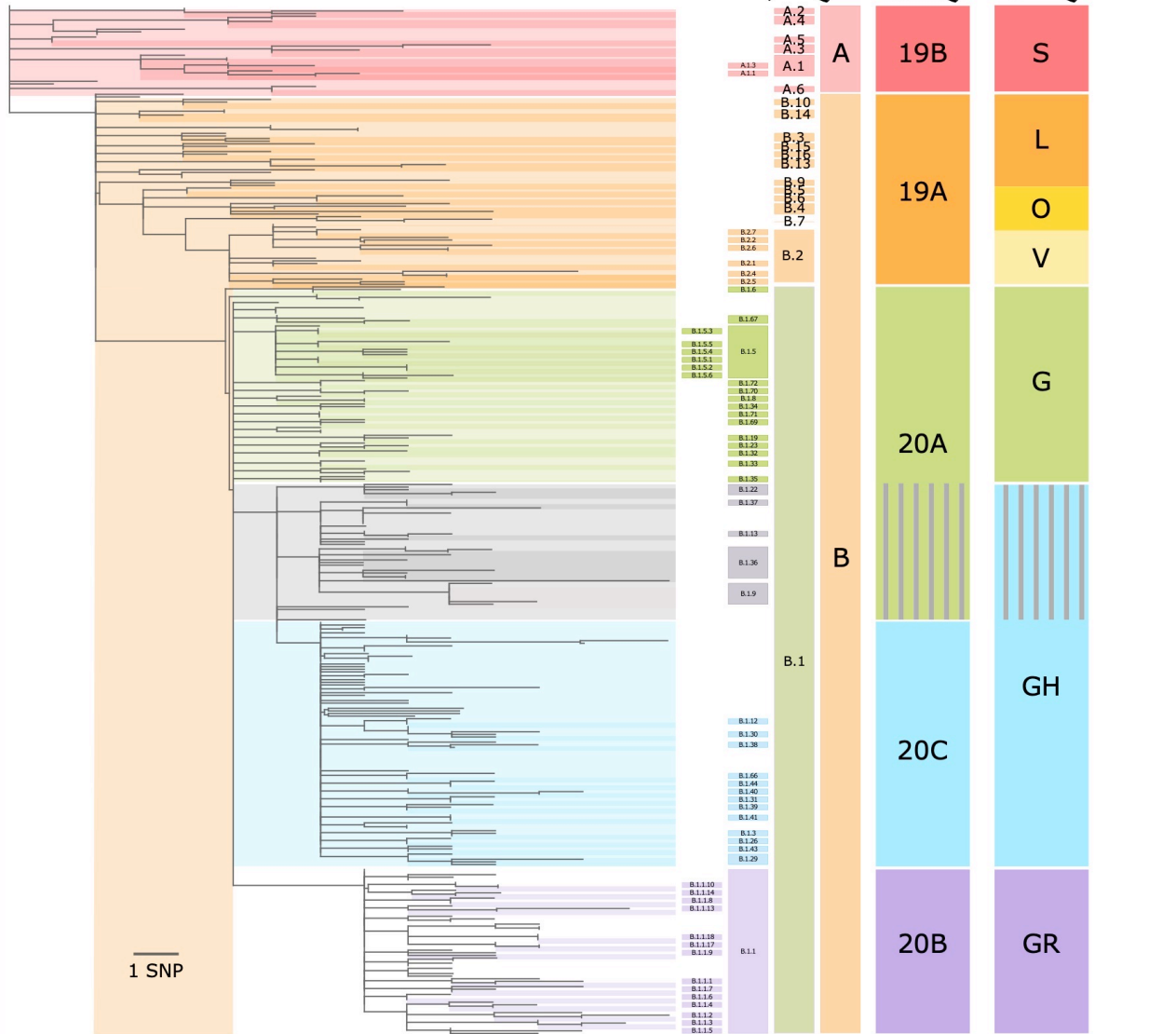
Associated with an  
epidemiological event

2020



Created with Emma Hodcroft, (Alm *et al* Eurosurveillance 2020)

2020



2021-2024

B.1.1.7



Pango Lineage

20I (501Y.V1)

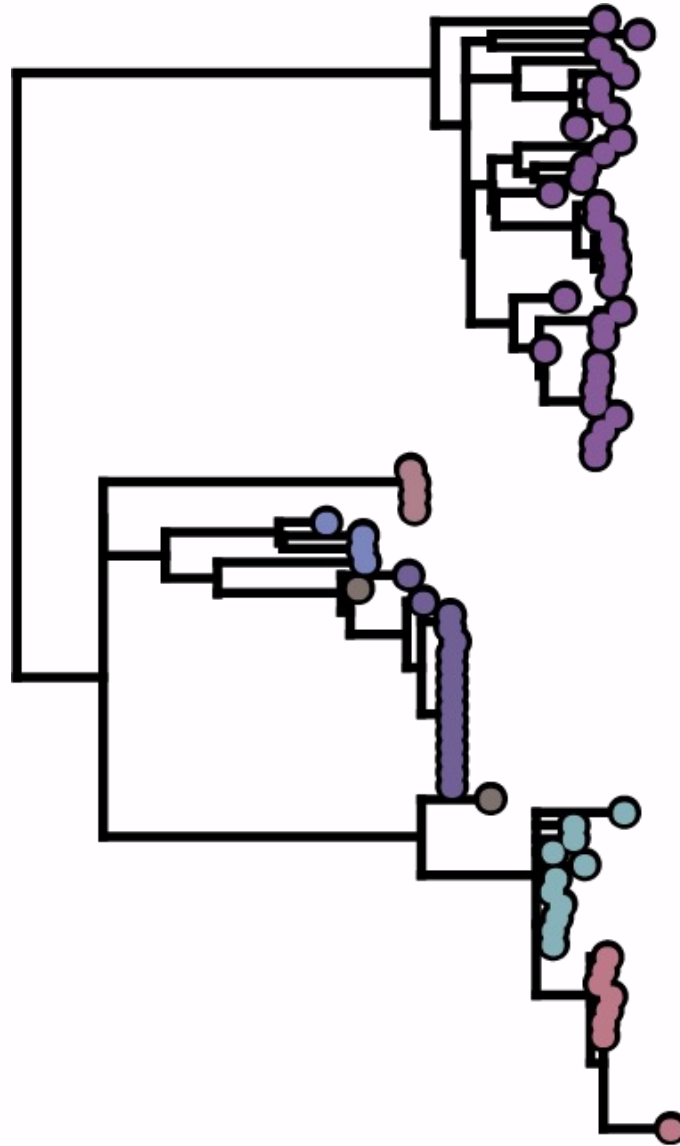


Nextstrain

Alpha



World Health Organization

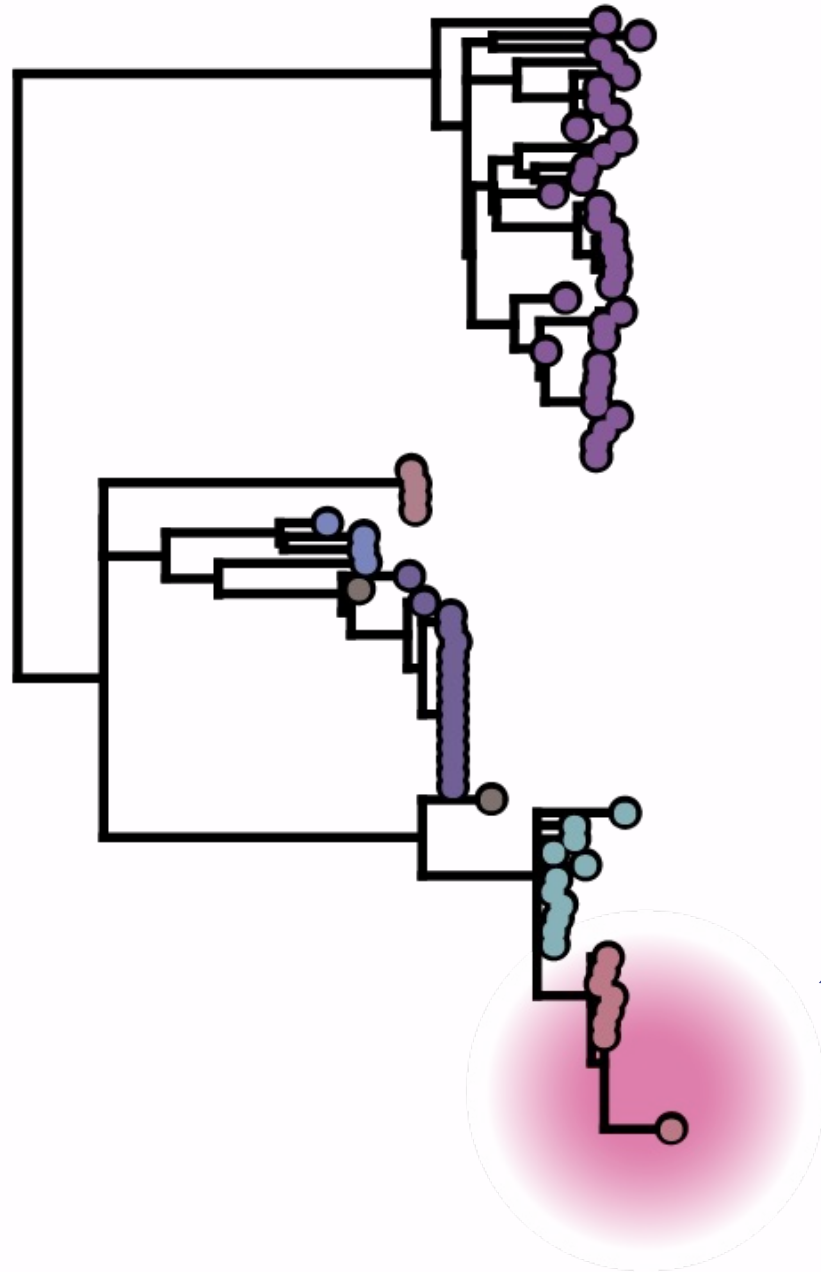


Fine scaled tracking

Unlimited names

No judgement on  
biology of the virus  
per se





Fine scaled tracking

Unlimited names

Something of concern comes along...





# Pango Network Formalisation

## Pango Committee

Creates rules

Reviews rule changes



Oli Pybus  
Pango chair

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop  
software

Update [cov-lineages.org](https://cov-lineages.org)



I acknowledge that the word “grinch” is protected by trademarks owned by Seuss Enterprises

# Pango Network Formalisation



2 postdocs

pango.network

## Pango Committee

Creates rules

Reviews rule changes



Oli Pybus  
Pango chair

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop

software

Update [cov-lineages.org](https://cov-lineages.org)

2 summer  
students

grinch server



**FAST — GRANTS**

# Pango Network Formalisation

## Pango Committee

Creates rules

Reviews rule changes

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop  
software

Update [cov-lineages.org](https://cov-lineages.org)



Lineage  
assignment

## Pango Committee

Creates rules

Reviews rule changes

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop  
software

Update [cov-lineages.org](https://cov-lineages.org)



Lineage  
assignment

## Pango Committee

Creates rules

Reviews rule changes

Generalised rules



Rachel Colquhoun

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop software

Update [cov-lineages.org](https://cov-lineages.org)



Lineage assignment

## Lineage Designation Committee

Horizon scan lineages

Analyse lineage requests

Maintain & develop  
software

Update [cov-lineages.org](https://cov-lineages.org)

Evolution of the lineage  
designation system

## Lineage Designation Committee

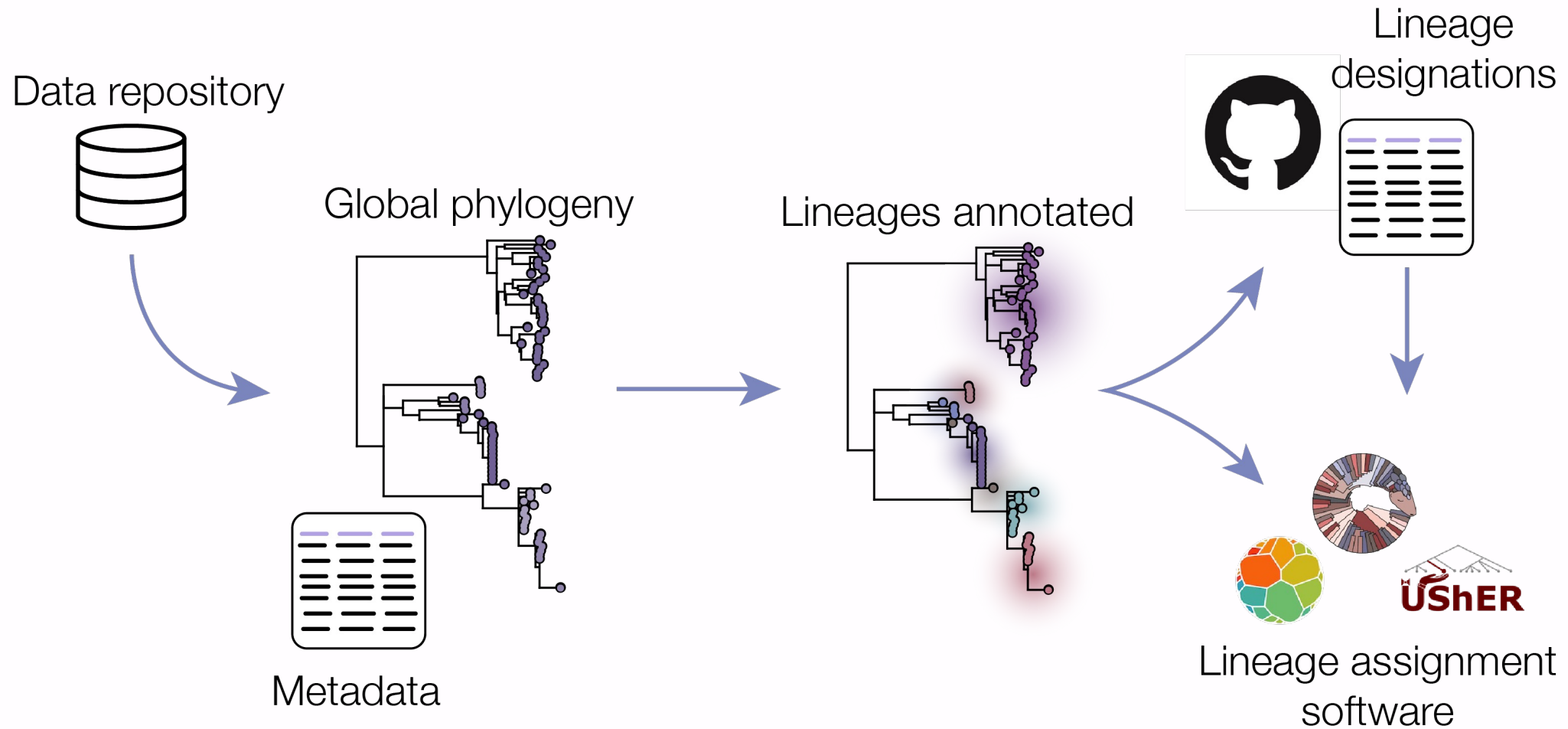
Horizon scan lineages

Analyse lineage requests

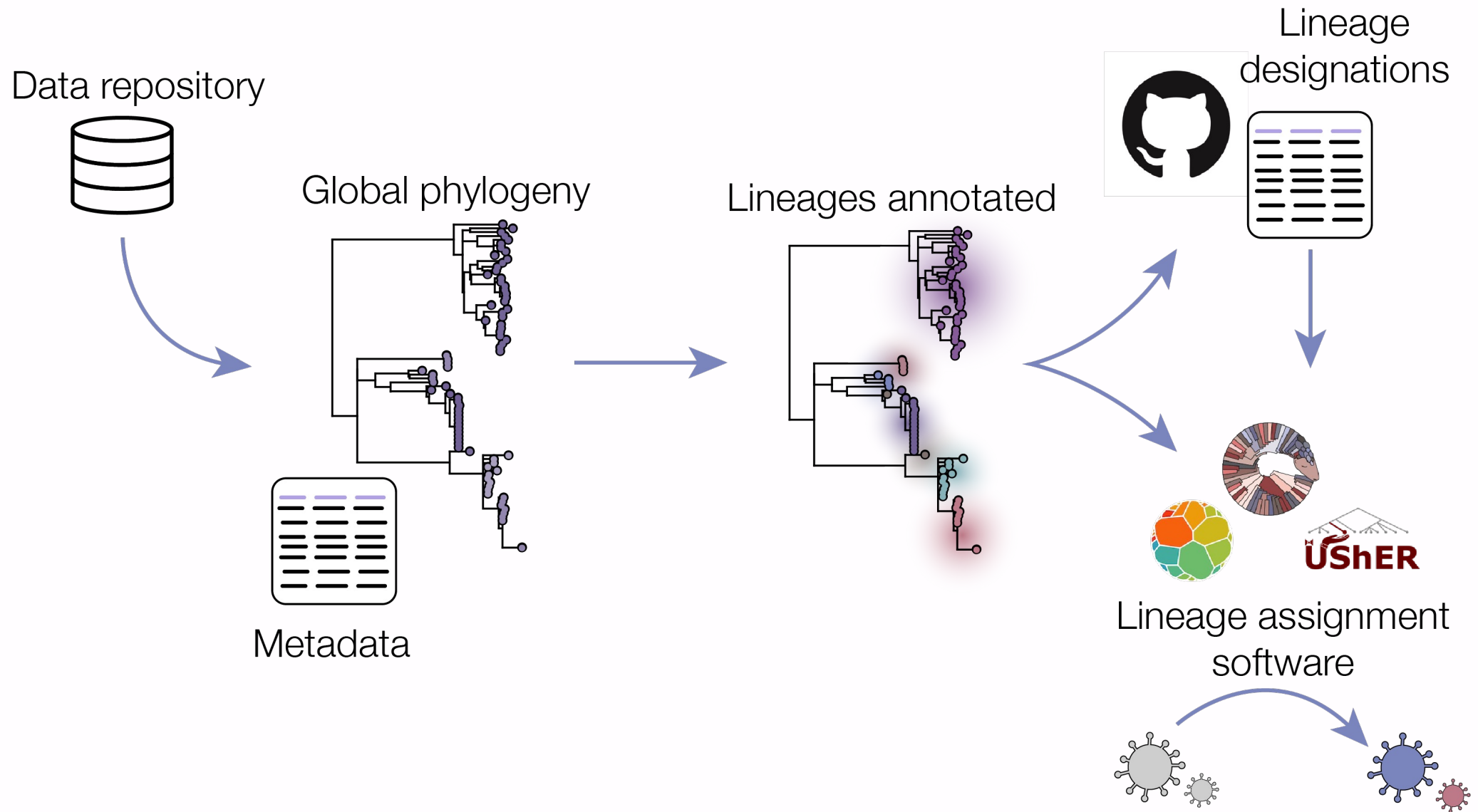
Maintain & develop  
software

Update [cov-lineages.org](https://cov-lineages.org)

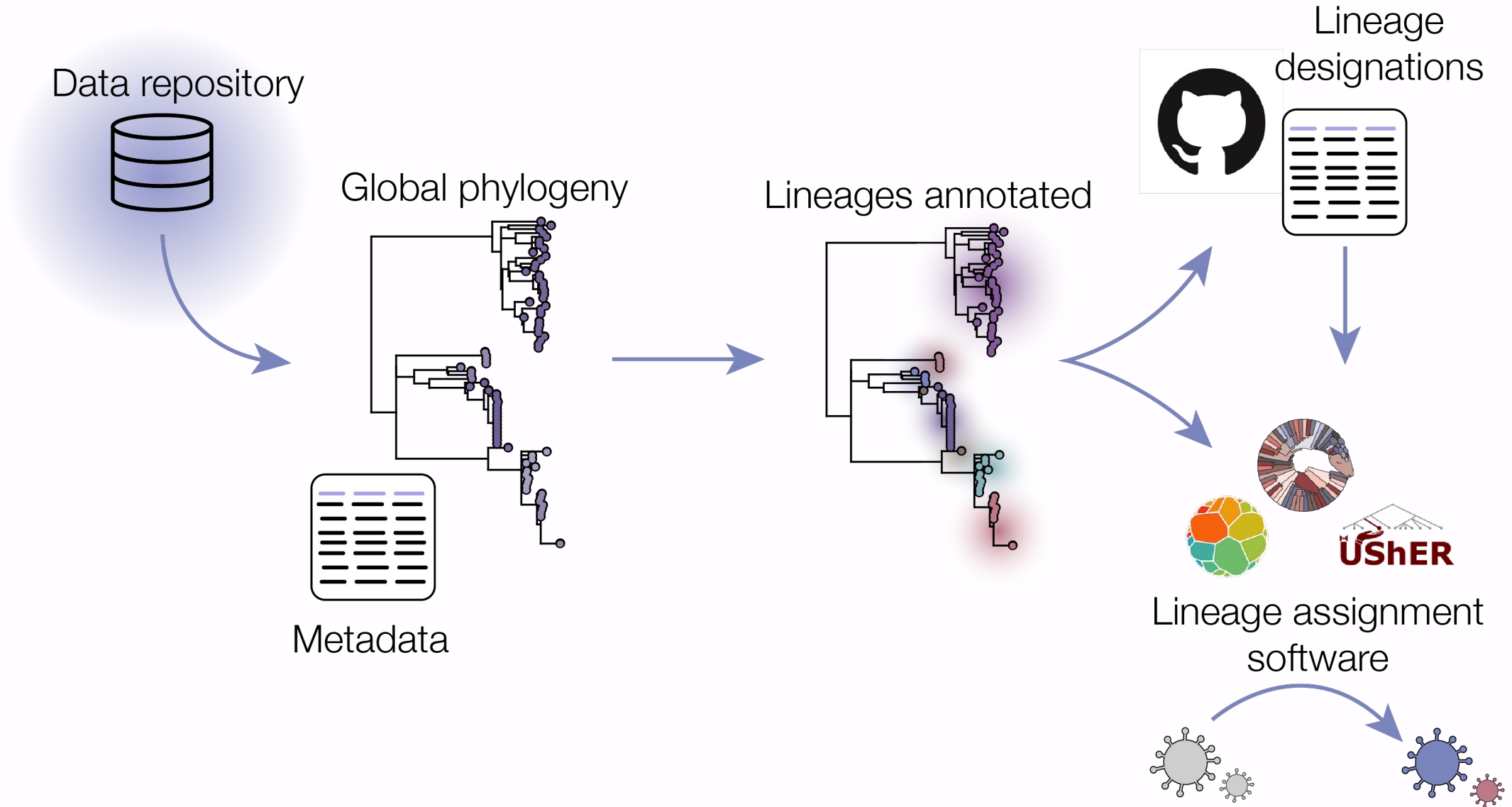
- 
- 5<sup>th</sup> March 1<sup>st</sup> lineages
  - 7<sup>th</sup> March tool to assign them (pangolin)
  - 5<sup>th</sup> July [cov-lineages.org](https://cov-lineages.org)

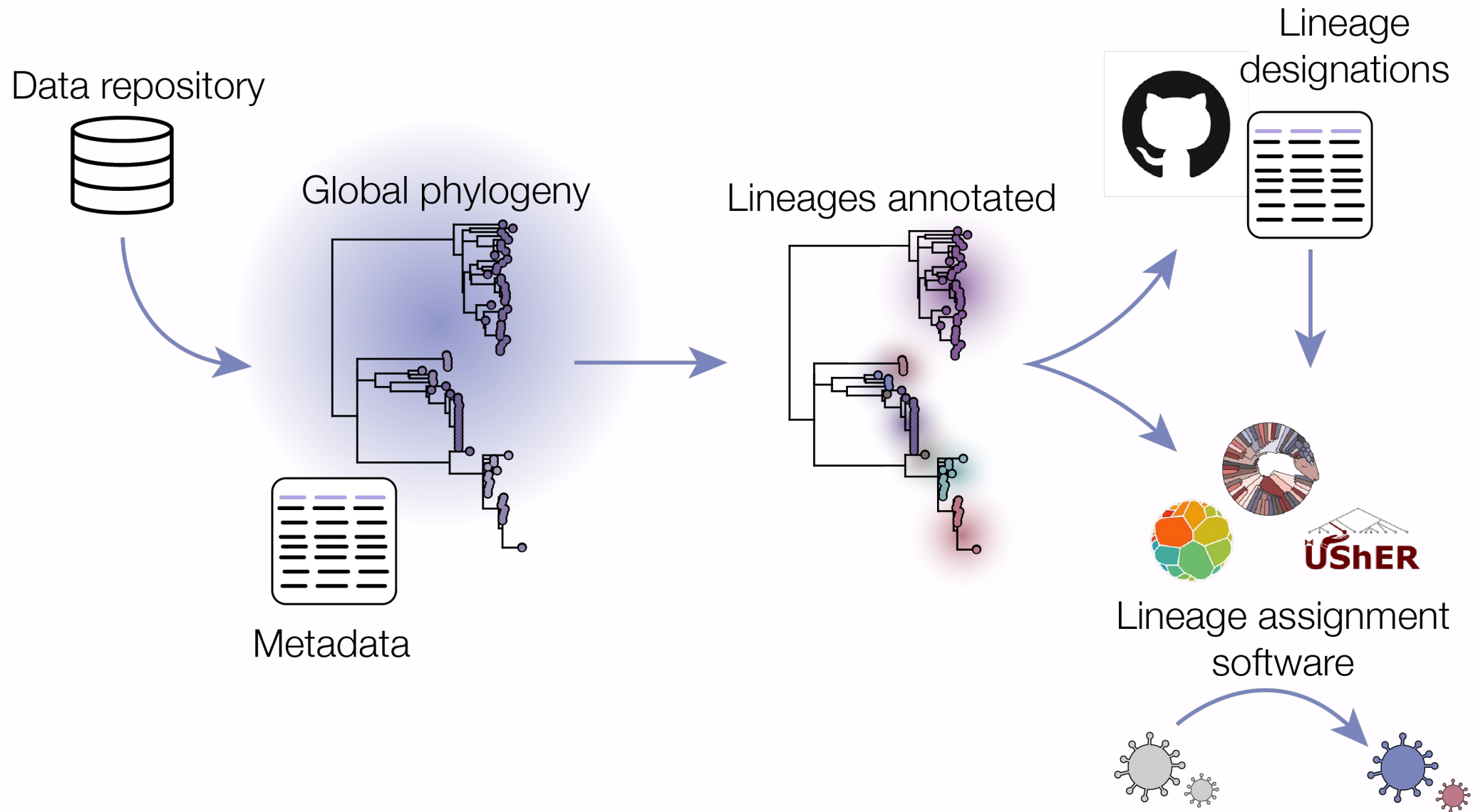


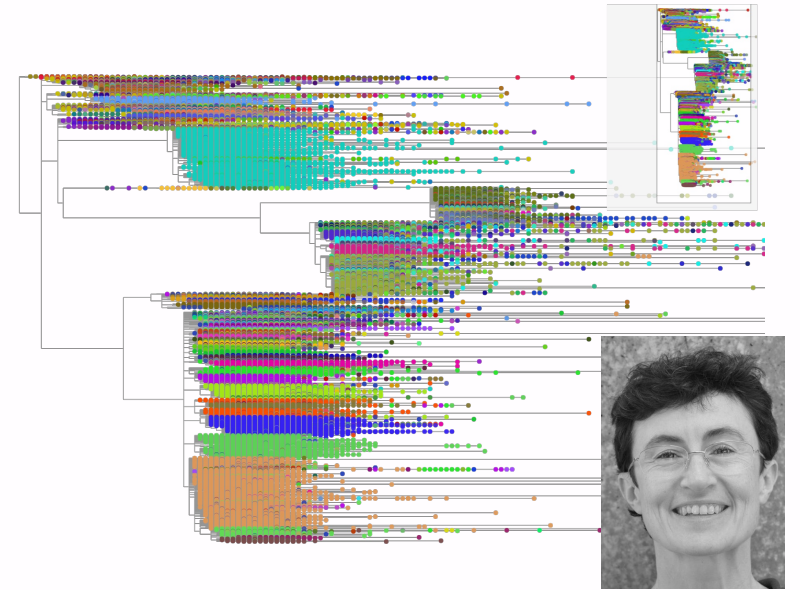
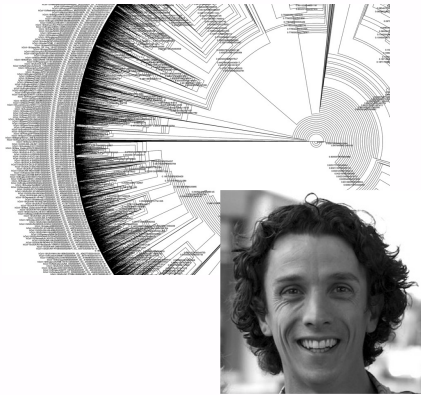
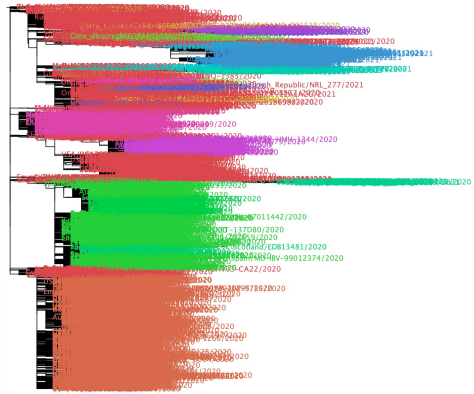
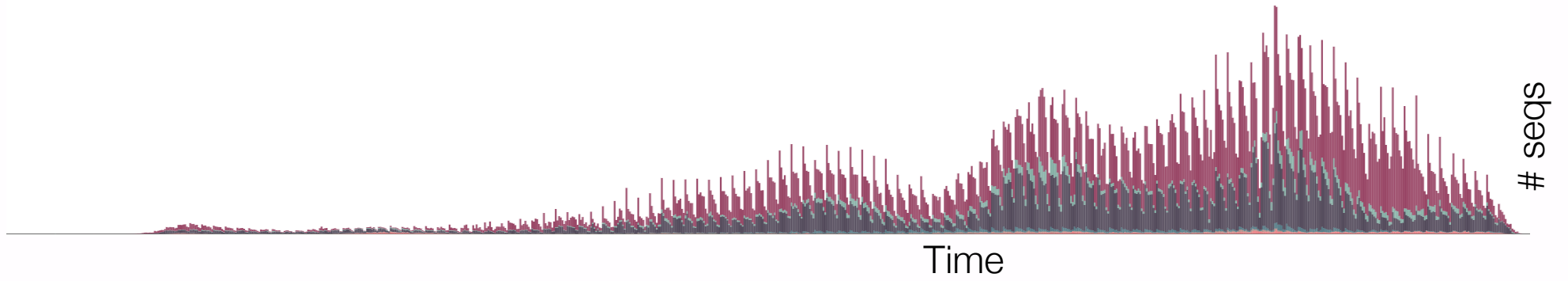
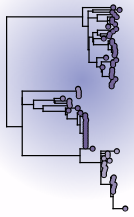




GISAID + COG-UK + DDBJ + ENA + NCBI







1 in-house  
tree



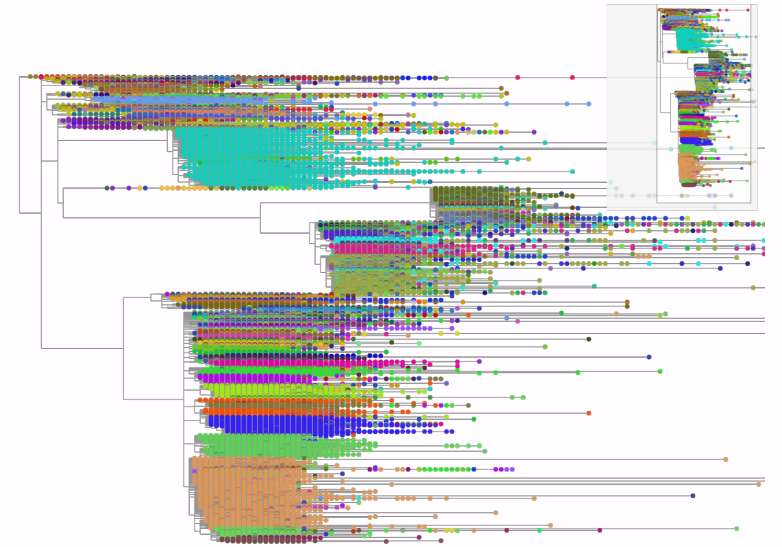
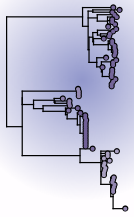
Rob Lanfear  
(sarscov2phylo)



The in-house  
chunk approach  
(phylopipe)



Angie  
Hinrichs  
(USHER)

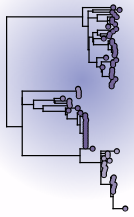


Yatish  
Turakhia

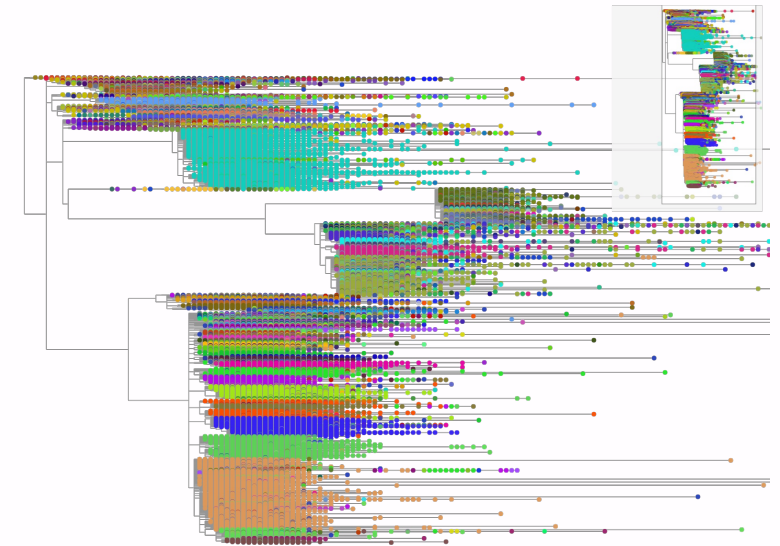
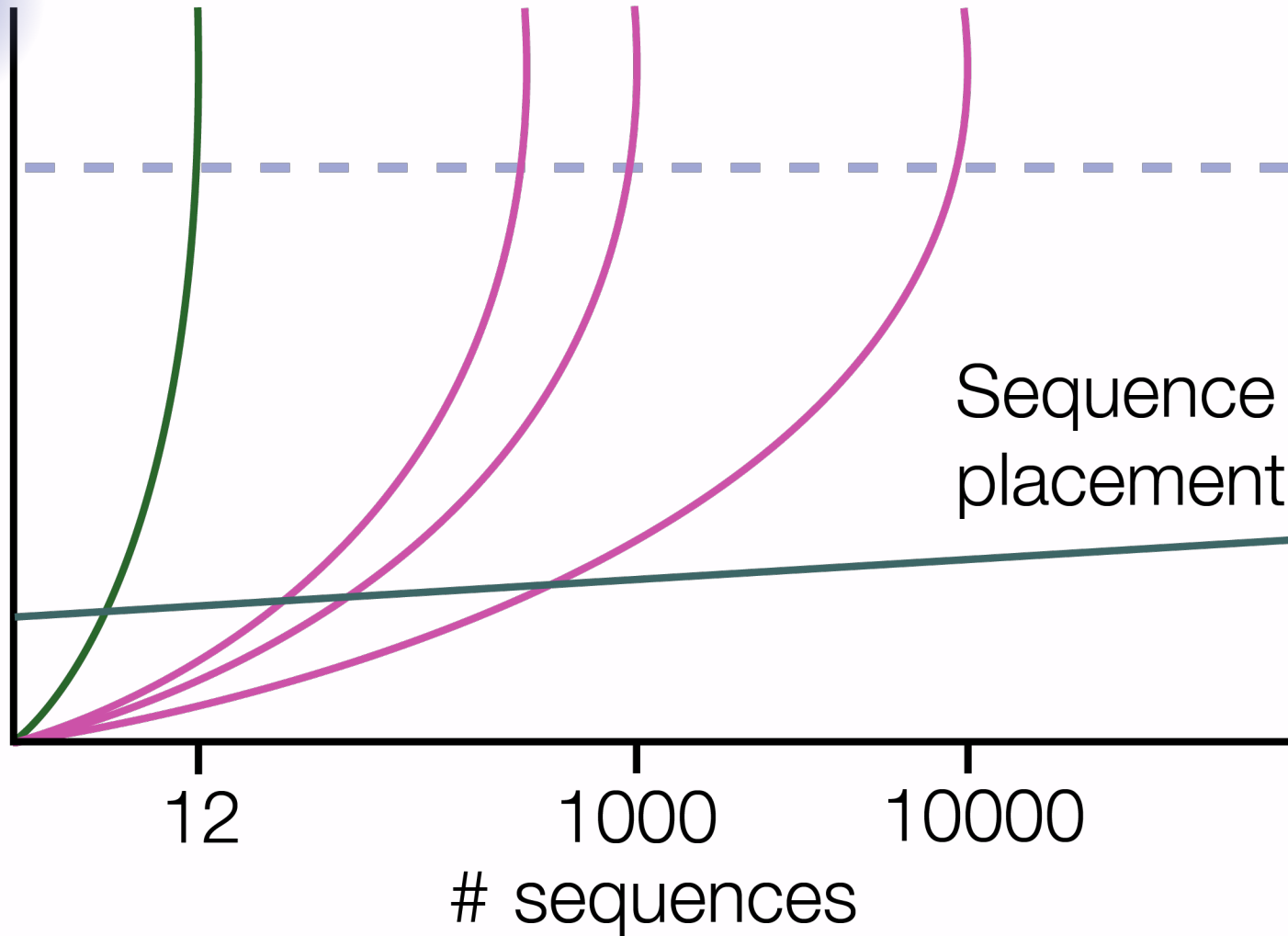


Angie  
Hinrichs

One tree to rule them all...



Time



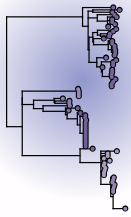
Yatish Turakhia




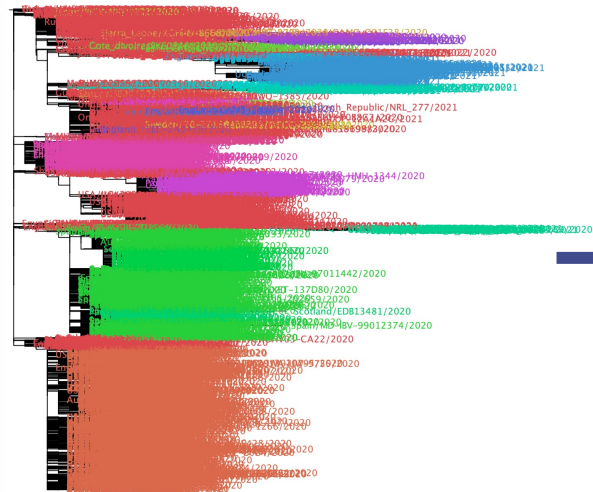
Angie Hinrichs

One tree to rule them all...

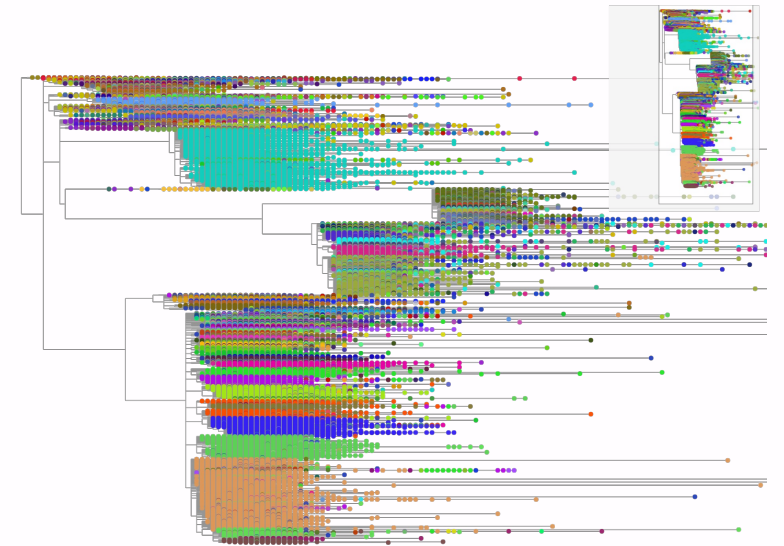




Cov2Tree  powered by Taxonium



Theo Sanderson

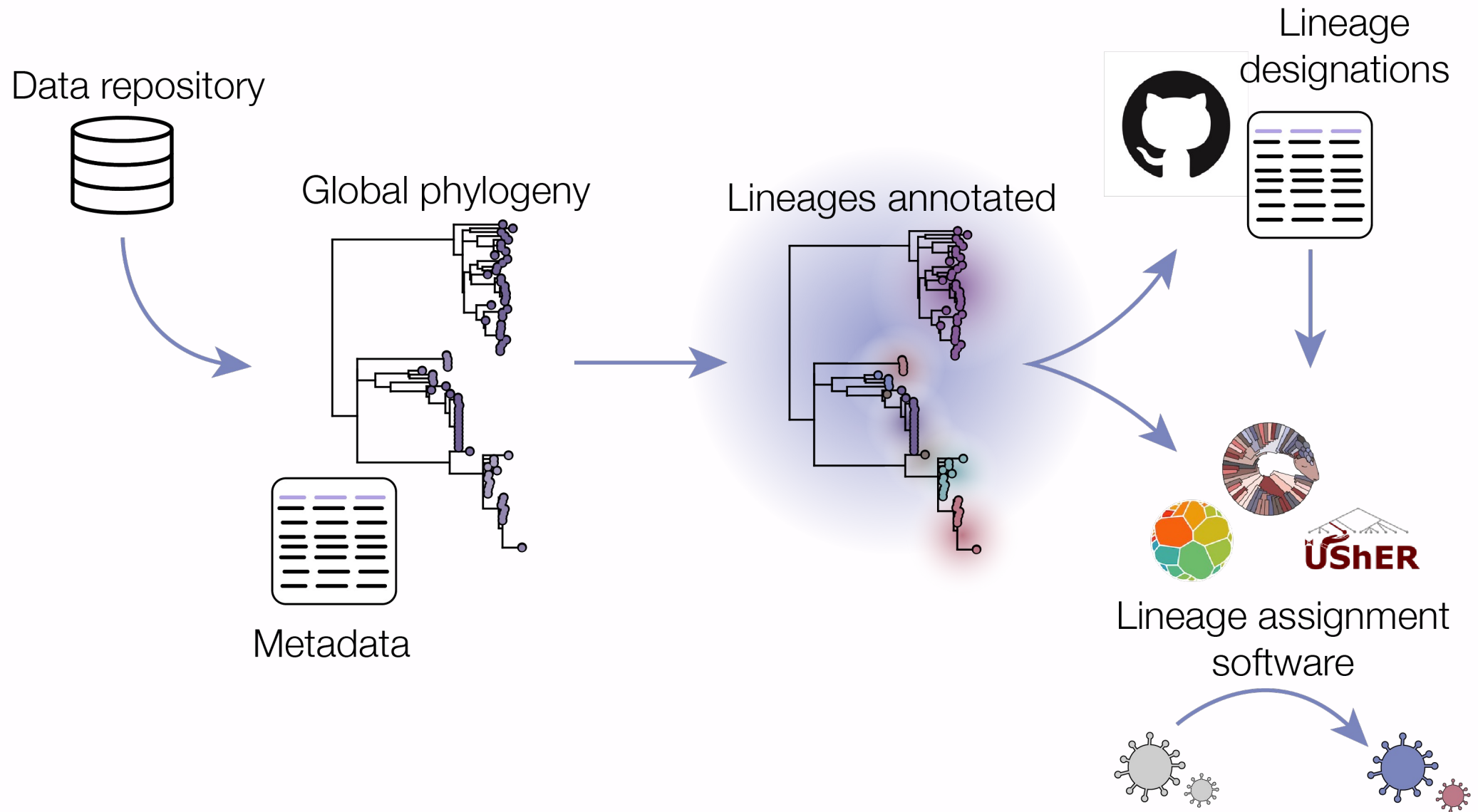


Yatish Turakhia

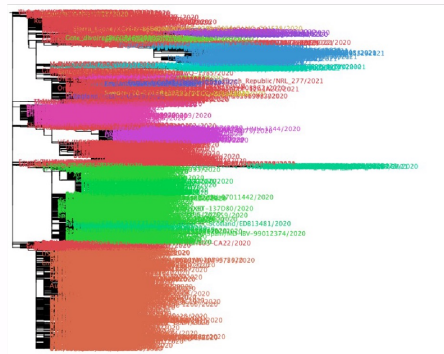
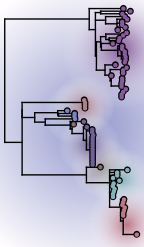


Angie Hinrichs

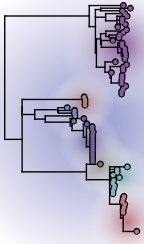
FigTree  
up to ~10K tips





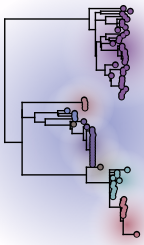


Full tree curation

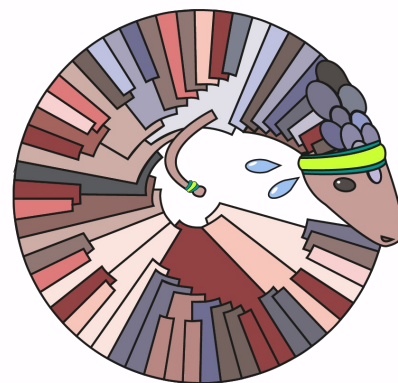


Full tree curation

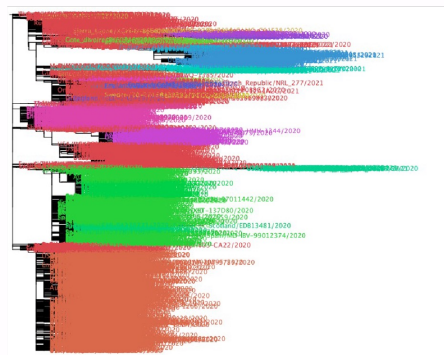
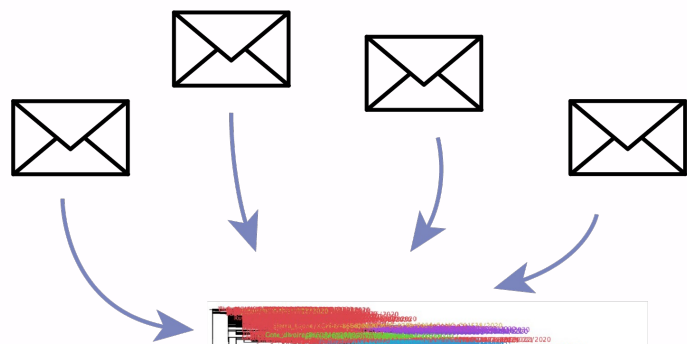
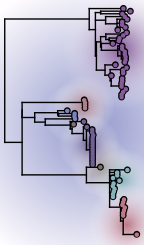




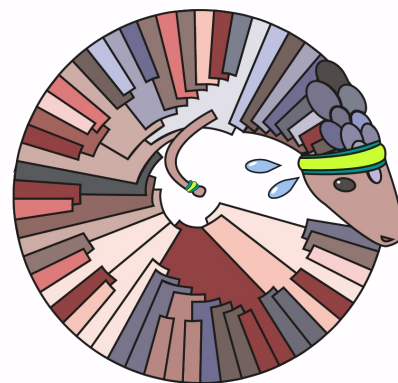
Full tree curation



“pangothon”  
750K sequences

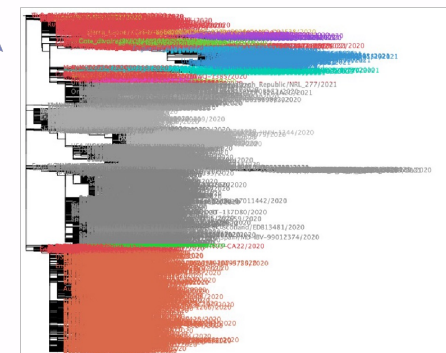
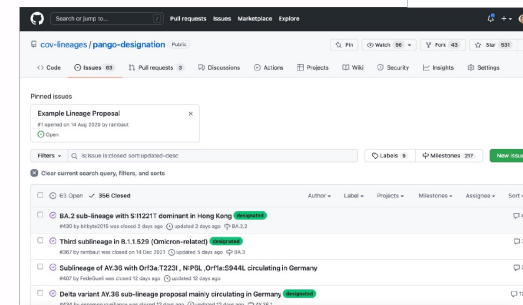


Full tree curation



“pangothon”  
750K sequences

## Crowd-sourcing lineages



Sub-tree curation



Chris Ruis

# Crowd-sourcing lineages: a global sentinel network?



Search or jump to... [7] Pull requests Issues Marketplace Explore

cov-lineages / pango-designation Public

Code Issues 63 Pull requests 3 Discussions Actions Projects Wiki Security Insights Settings

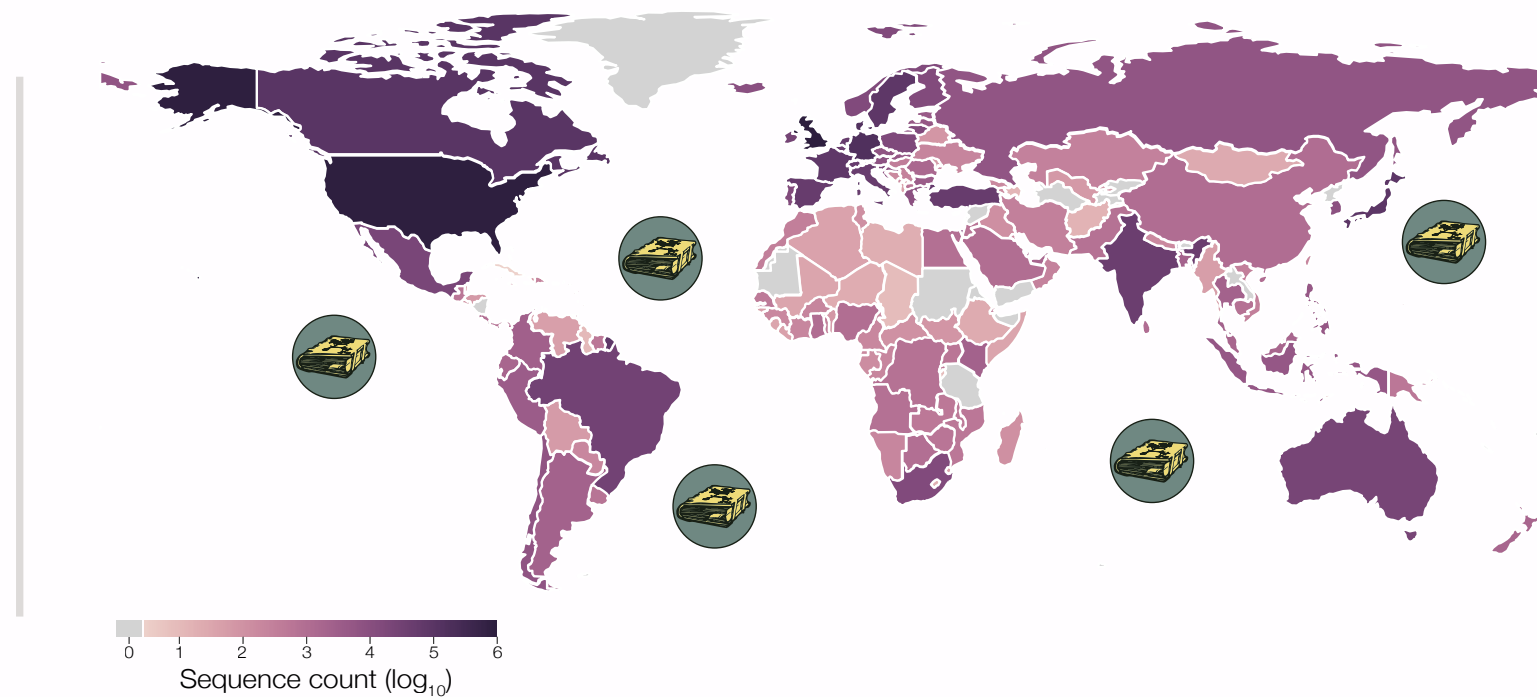
Pinned issues

Example Lineage Proposal  
#1 opened on 14 Aug 2020 by rambaut  
Open

Filters is:issue is:closed sort:updated-desc Labels 9 Milestones 217 New issue

Clear current search query, filters, and sorts

| <input type="checkbox"/> | <input type="radio"/>            | 63 Open  | <input checked="" type="radio"/> | 356 Closed | Author | Label | Projects | Milestones | Assignee | Sort |
|--------------------------|----------------------------------|--|----------------------------------|------------|--------|-------|----------|------------|----------|------|
| <input type="checkbox"/> | <input checked="" type="radio"/> | BA.2 sub-lineage with S:11221T dominant in Hong Kong <b>designated</b>                   |                                  |            |        |       |          |            |          | 4    |
|                          |                                  | #430 by bitbyte2015 was closed 2 days ago updated 2 days ago BA.2.2                      |                                  |            |        |       |          |            |          |      |
| <input type="checkbox"/> | <input checked="" type="radio"/> | Third sublineage in B.1.1.529 (Omicron-related) <b>designated</b>                        |                                  |            |        |       |          |            |          | 7    |
|                          |                                  | #367 by rambaut was closed on 14 Dec 2021 updated 5 days ago BA.3                        |                                  |            |        |       |          |            |          |      |
| <input type="checkbox"/> | <input checked="" type="radio"/> | Sublineage of AY.36 with Orf3a:T223I, N:P6L_Orf1a:S944L circulating in Germany           |                                  |            |        |       |          |            |          | 3    |
|                          |                                  | #407 by FedeGuelli was closed 12 days ago updated 12 days ago                            |                                  |            |        |       |          |            |          |      |
| <input type="checkbox"/> | <input checked="" type="radio"/> | Delta variant AY.36 sub-lineage proposal mainly circulating in Germany <b>designated</b> |                                  |            |        |       |          |            |          | 12   |
|                          |                                  | #434 by genomesurveillance was closed 12 days ago updated 12 days ago AY.36.1            |                                  |            |        |       |          |            |          |      |



# Crowd-sourcing lineages: a global sentinel network?



Tom Peacock



Cornelius  
Roemer



Memorablea Memorable



aviczhl2 Xu Zou · he/him



FedeGueli Federico Gueli



yanjin1234



AleksanderRing



RaphaelRaphael Raphael



BorisUitham



RajLABN Rajendram Rajnarayanan



Tom Peacock  
thomaspeacock



InfrPopGen



JosetteSchoenma



OrsonMM Orson



ryhisner Ryan Hisner



Mike-Honey Mike Honey



BorisUitham



AngieHinrichs Angie Hinrichs



oobb45729



larboh-ioc LARBOH IOC



NkRMnZr



Outpfmance



Over-There-Is Yat-sing Ng



arodzh-sudo Arnold Rodriguez-Hilario



Sinickle



HynnSpylor

Variant hunters A.K.A.

Soltation sentinels

Lineage lookers

Sequence surfers

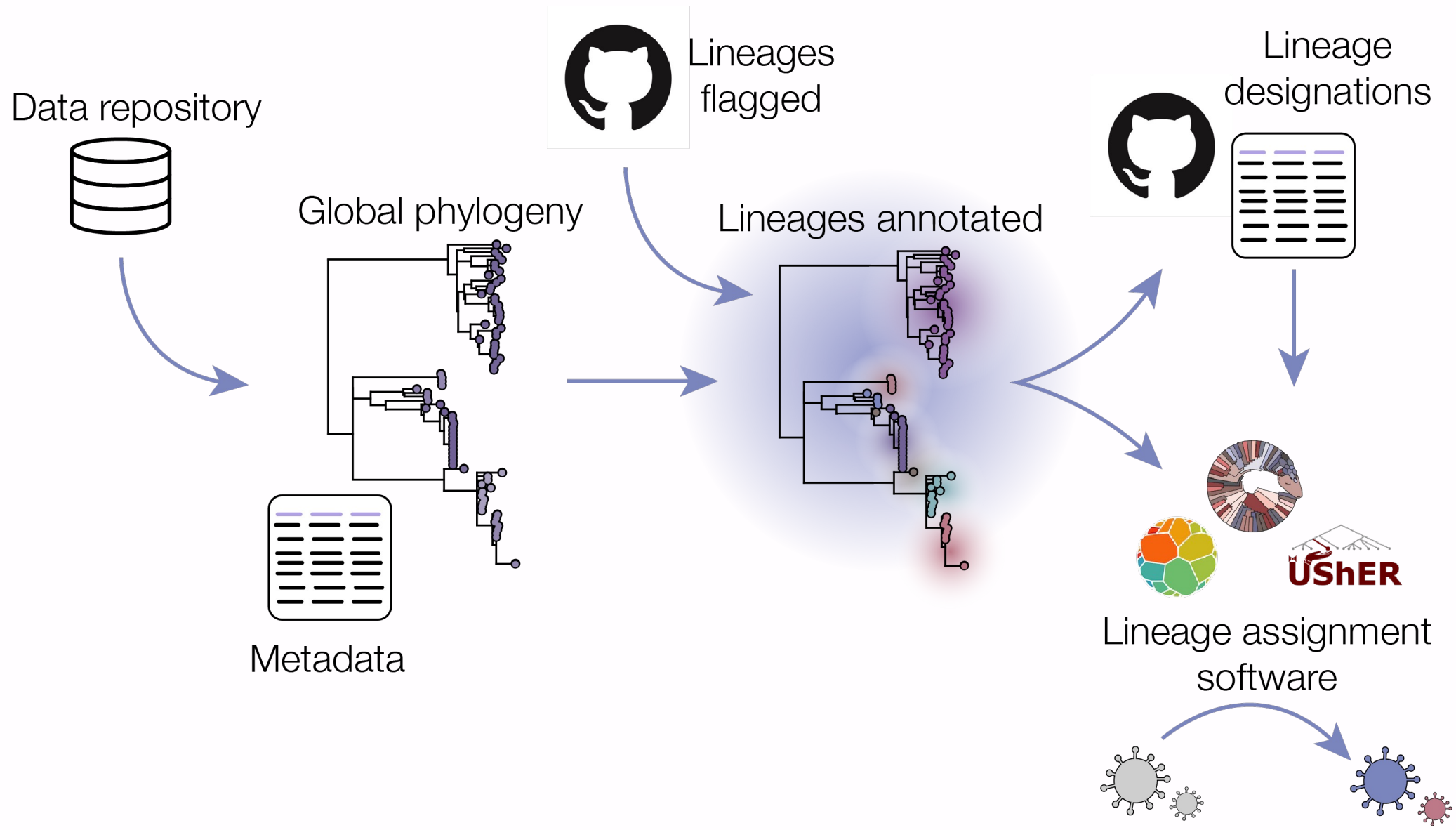
Phylo-philes

Variant viewers

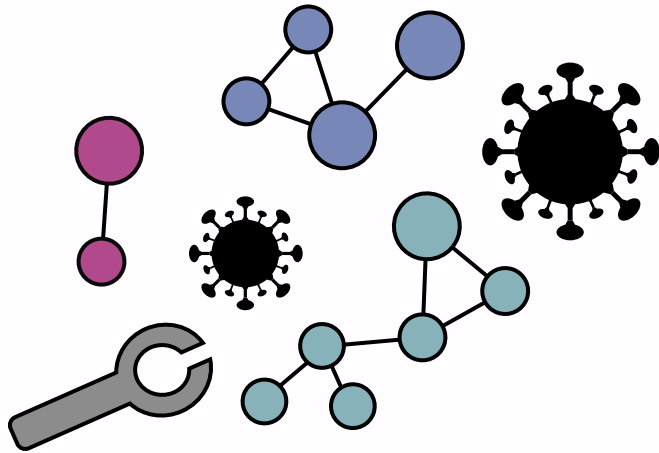
Fast-evo flaggers

Transmission trawlers

Spike seekers



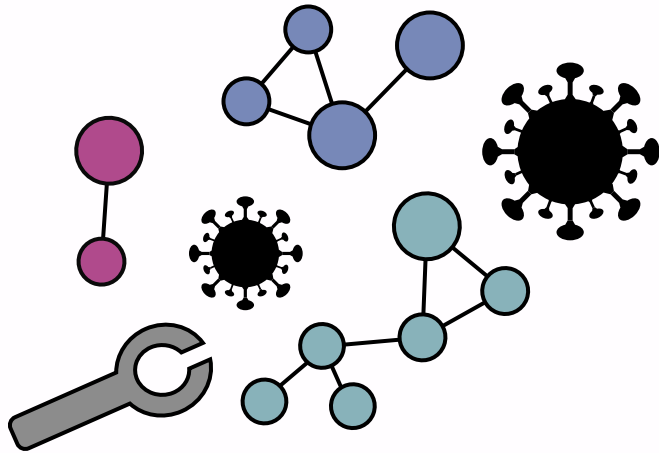
2024



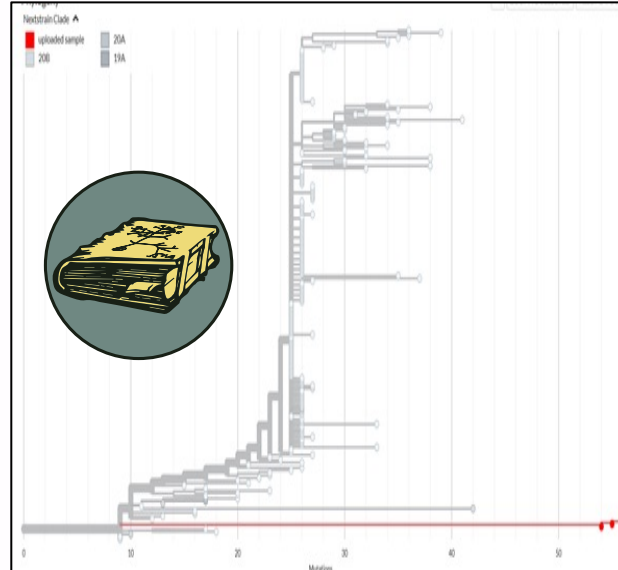
Adaptable system



2024

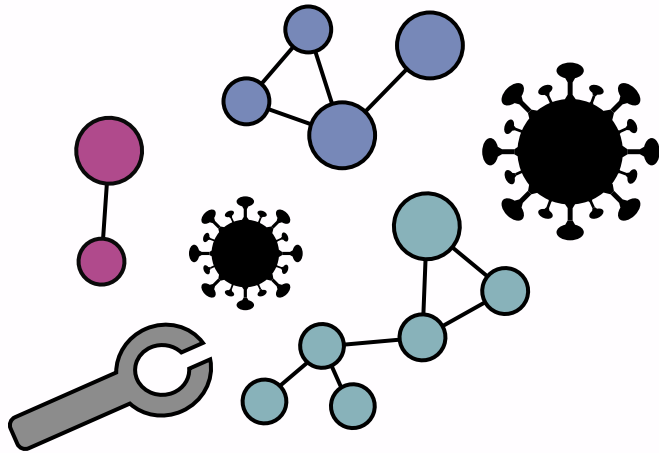


Adaptable tools

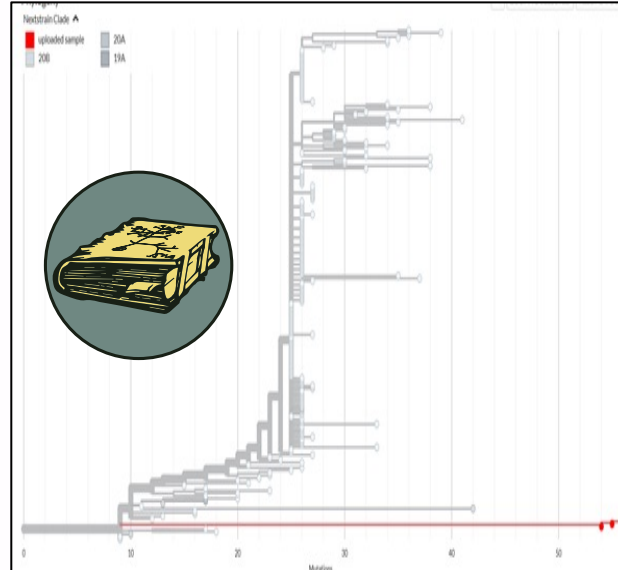


Horizon scanning capacity

2024



Adaptable tools



Horizon scanning  
capacity



Sustainable  
surveillance

Virus genome sequences



Question



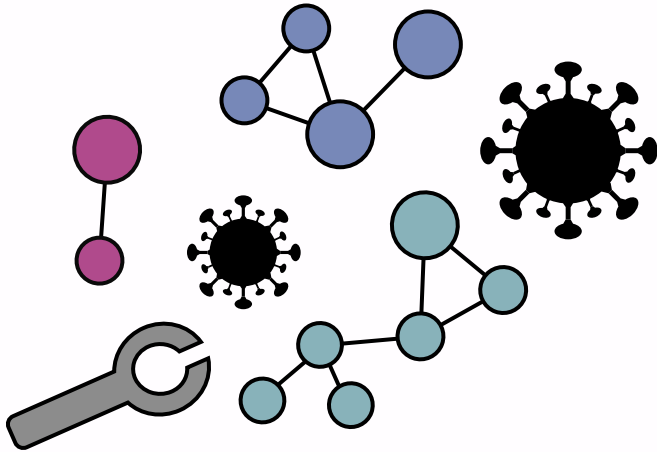
Build a tree



Interpret it



Answer



Virus genome sequences



Question



Build a tree



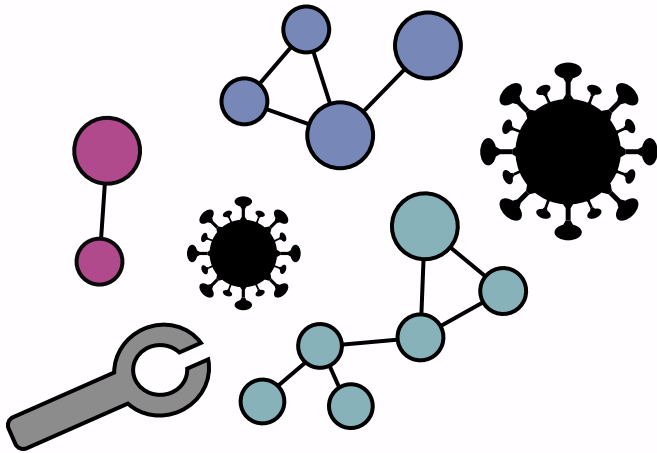
Interpret it



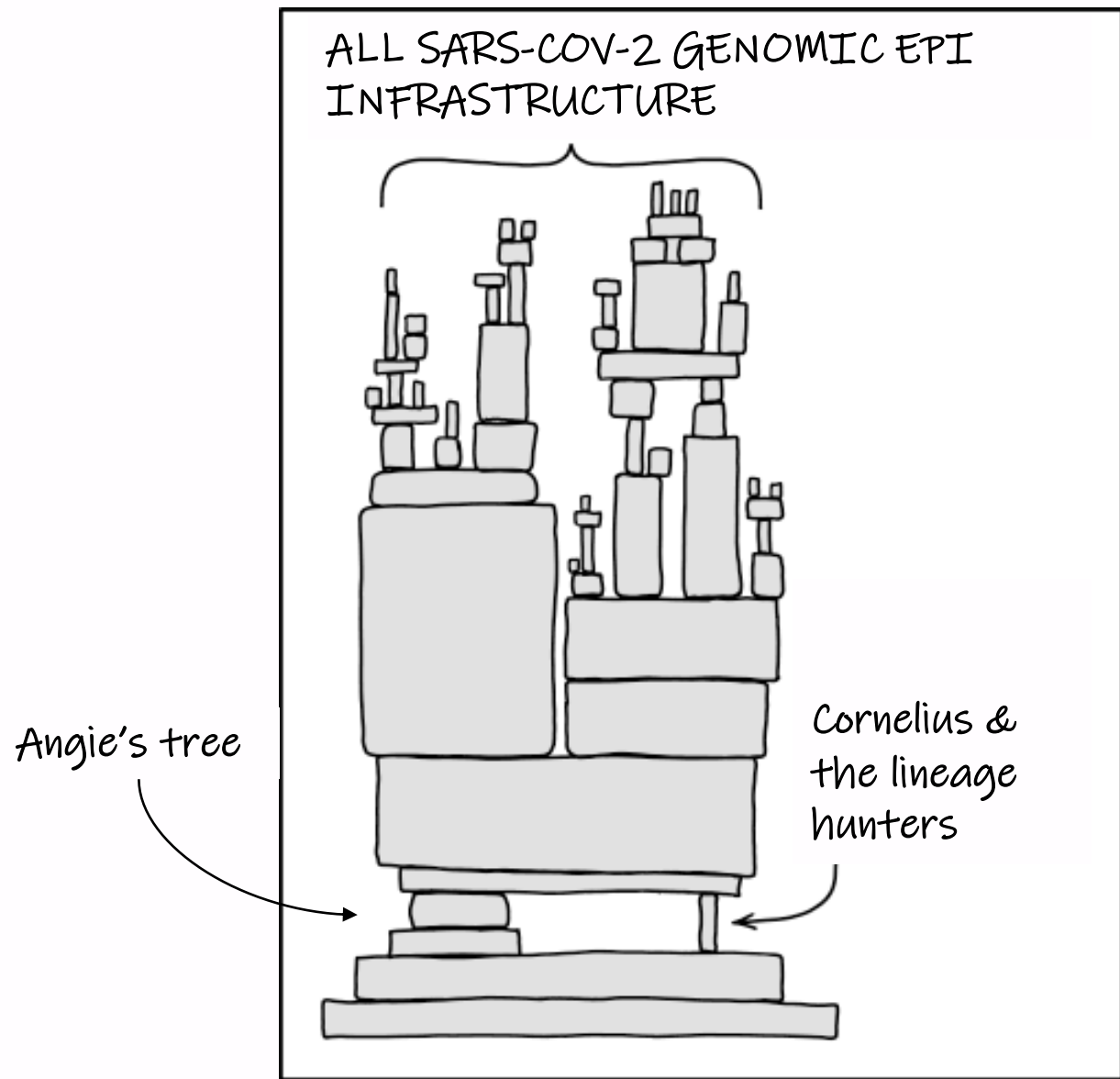
Answer



SARS-CoV-2 infrastructure



An alternative representation:



<https://xkcd.com/2347>

# The Pango Network

Oli Pybus  
Andrew Rambaut  
Chris Ruis  
Stephen Attwood  
Cornelius Roemer  
Many (!! ) volunteers



Summer students  
Isobel Guthrie  
Angela Sun

# The ARTIC Network

Andrew Rambaut  
Rachel Colquhoun  
Verity Hill  
  
JT McCrone  
Ben Jackson  
Emily Scher

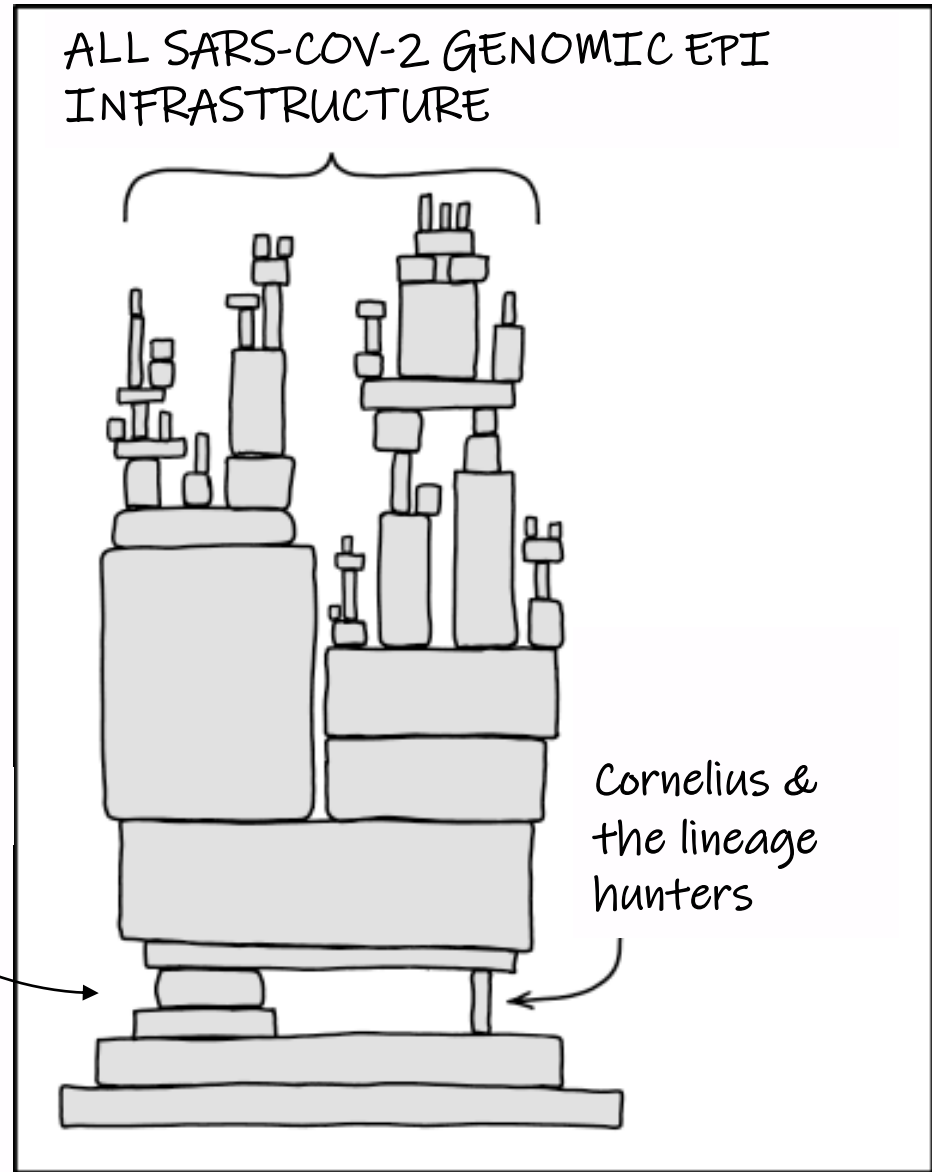
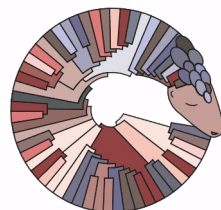
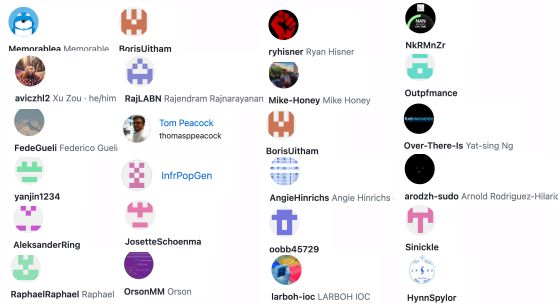
Angie Hinrichs  
Jakob McBroome  
Russ Corbett



THE UNIVERSITY  
of EDINBURGH



FAST — GRANTS



<https://xkcd.com/2347>