Tracking and predicting the spread and evolution of RNA viruses
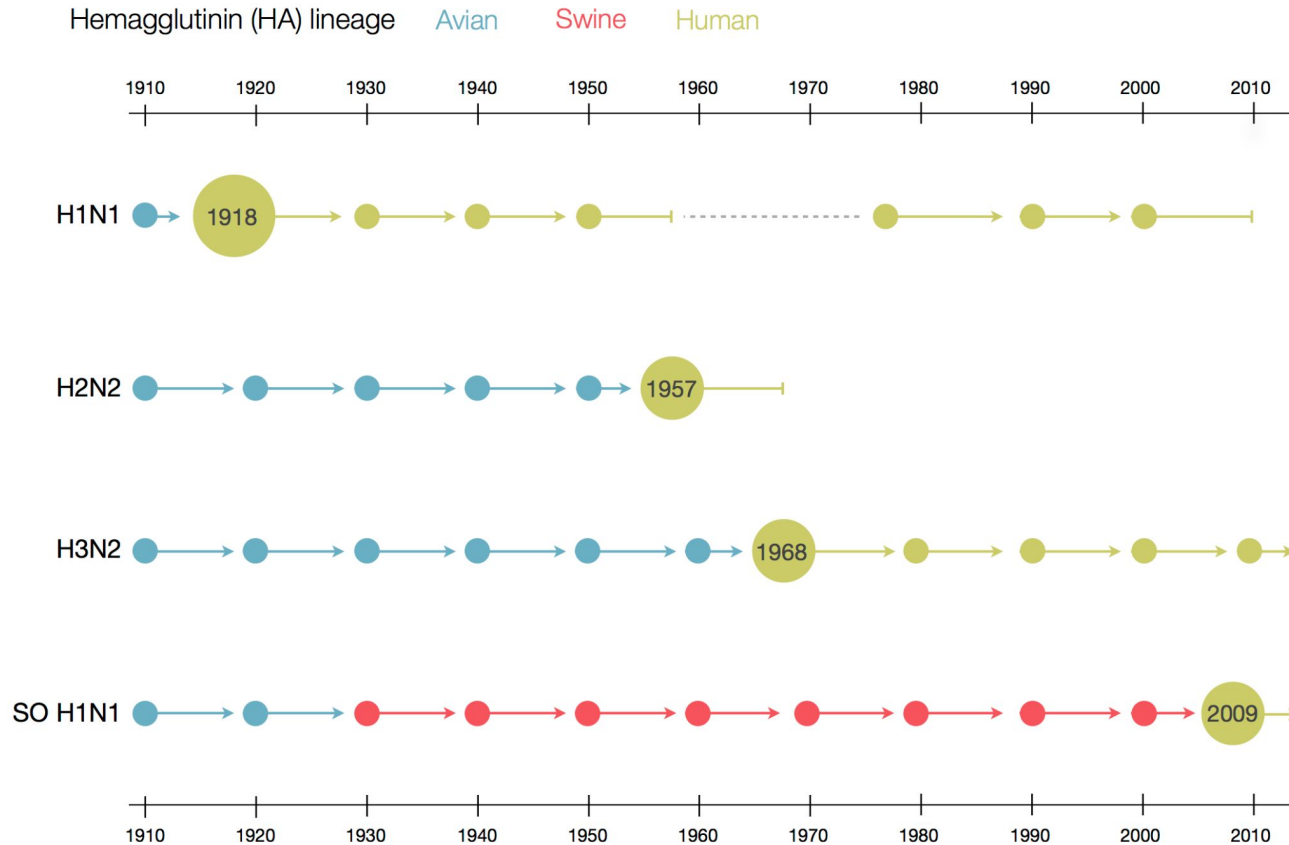
Richard Neher, University of Basel

BV-BRC Viral Sub-species Classification Workshop
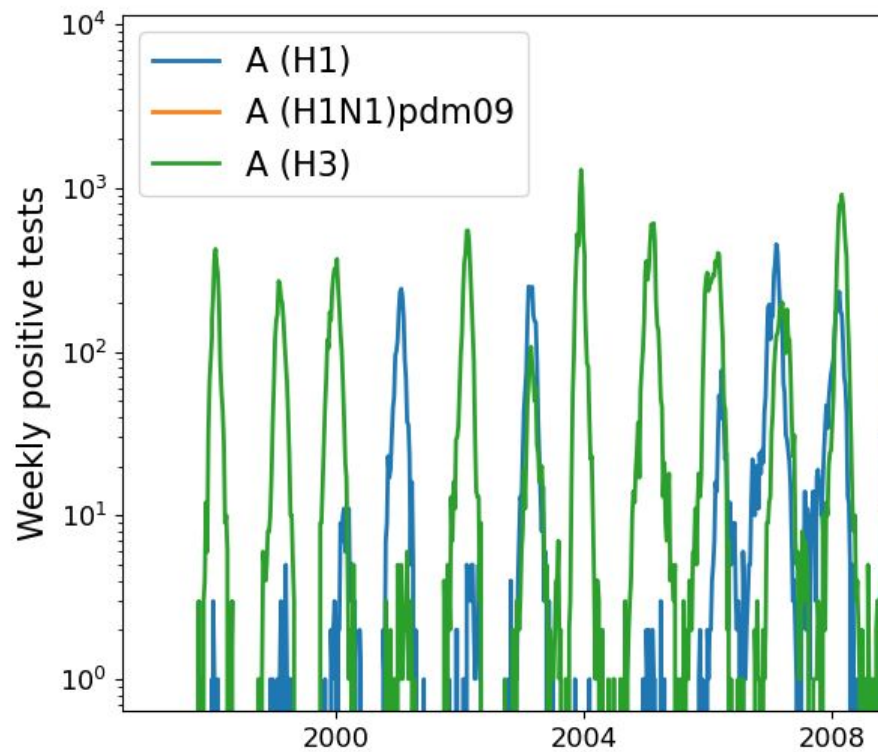April 2024, Rockville, MA

Disclosures:

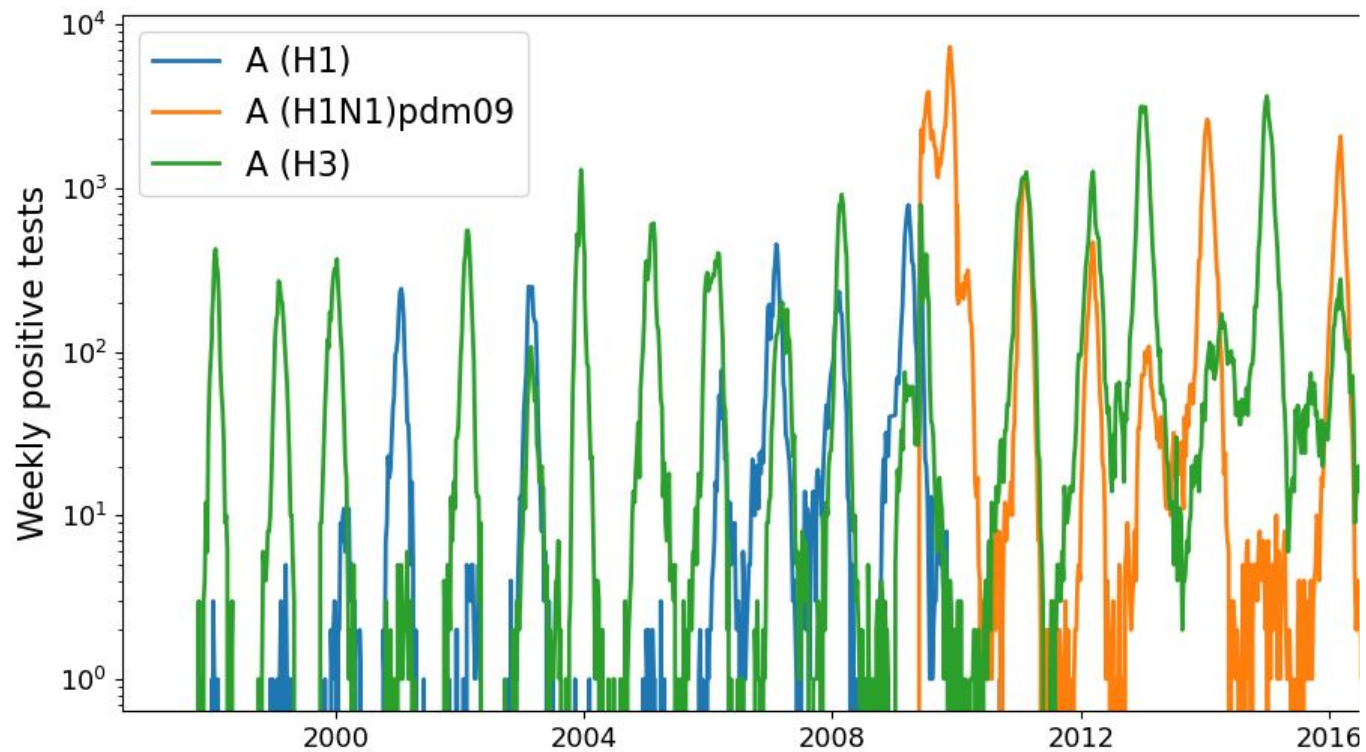I consult for Moderna TX on questions of influenza virus evolution.
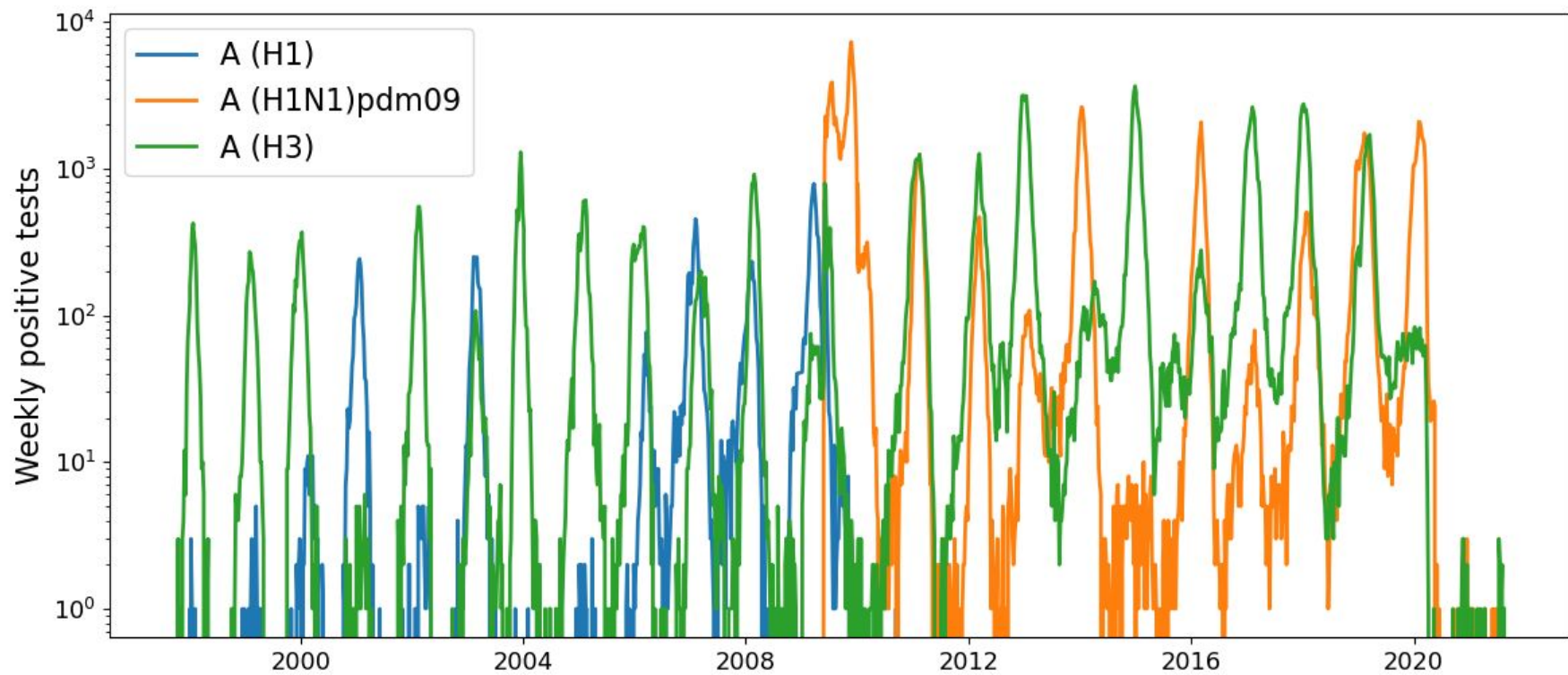
Human Seasonal Influenza Virus A

Seasonal dynamics of IAV in the Northern Hemisphere

Seasonal dynamics of IAV in the Northern Hemisphere

Seasonal dynamics of IAV in the Northern Hemisphere

What sustains endemic circulation?

- Birth of immunologically naive individuals
- Waning of pre-existing immunity
- Antigenic change of the virus
- Non-sterilizing immunity → persistence via asymptomatic acute infections

What sustains endemic circulation?

- Birth of immunologically naive individuals
- Waning of pre-existing immunity
- Antigenic change of the virus
- Non-sterilizing immunity → persistence via asymptomatic acute infections

Mostly population turnover:

Measles Virus (pre vaccine)

- Extremely rapid transmission among naive children
- Long-lived immunity → rare re-infection

Many animal viruses

- Rapid population turnover

What sustains endemic circulation?

- Birth of immunologically naive individuals
- Waning of pre-existing immunity
- Antigenic change of the virus
- Non-sterilizing immunity → persistence via asymptomatic acute infections

Mostly population turnover:

Measles Virus (pre vaccine)

- Extremely rapid transmission among naive children
- Long-lived immunity → rare re-infection

Many animal viruses

- Rapid population turnover

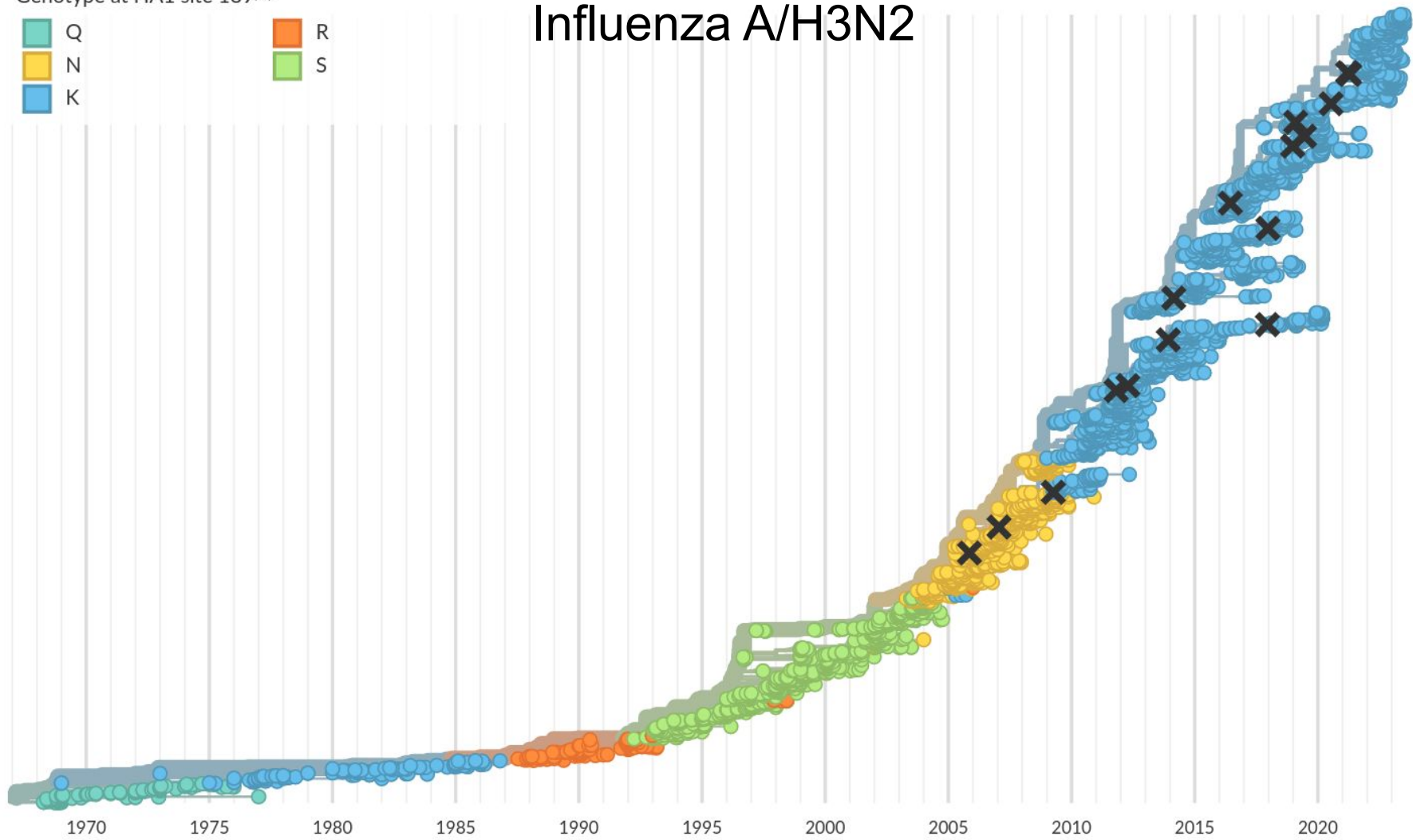## Antigenic evolution and waning

Influenza Virus, SARS-CoV-2, ...

- Frequent reinfection of adults
- Combination of waning and antigenic evolution
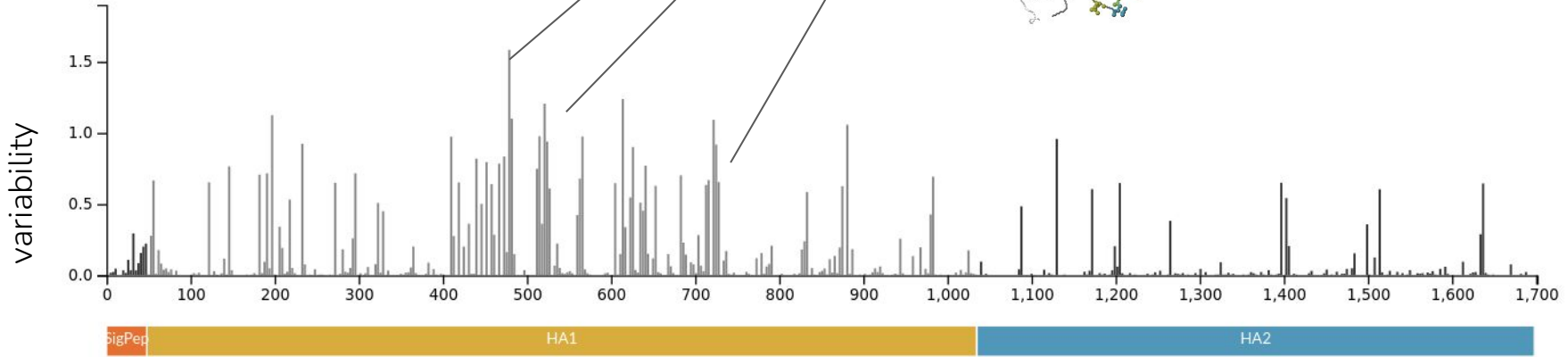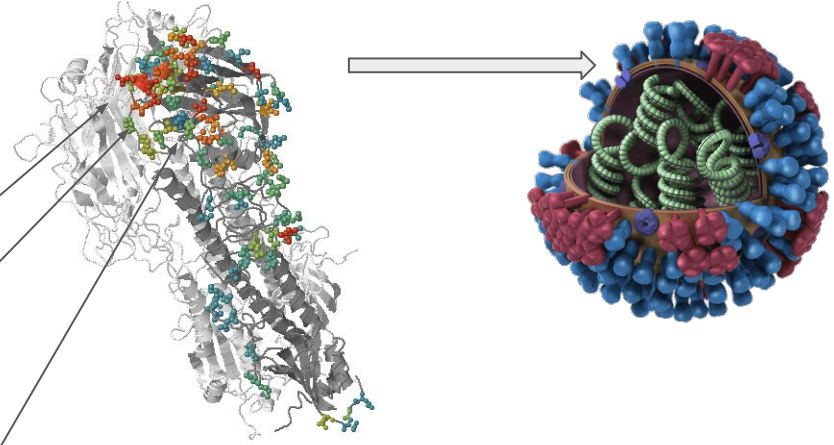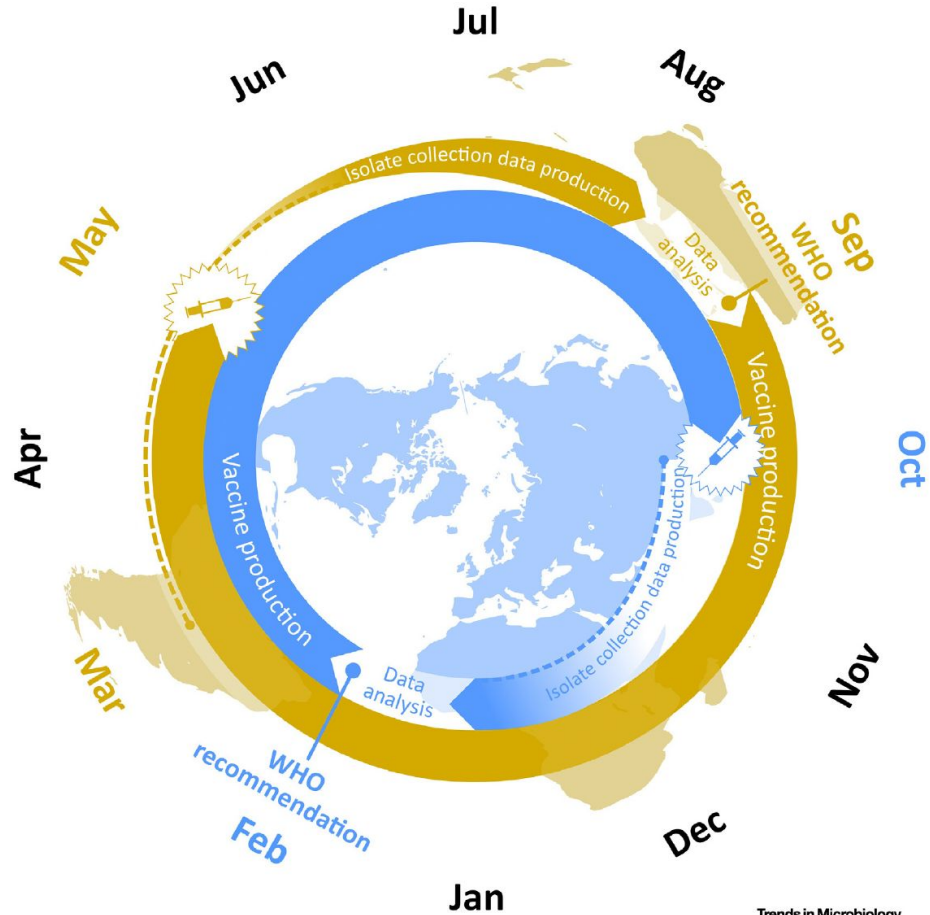- Approximate equilibrium between immunity and infections

Influenza A/H3N2

Genotype at HA1 site 189
- Q
- N
- K
- R
- S

Influenza surface proteins change rapidly
→ Vaccines need updating

Vaccine selection process
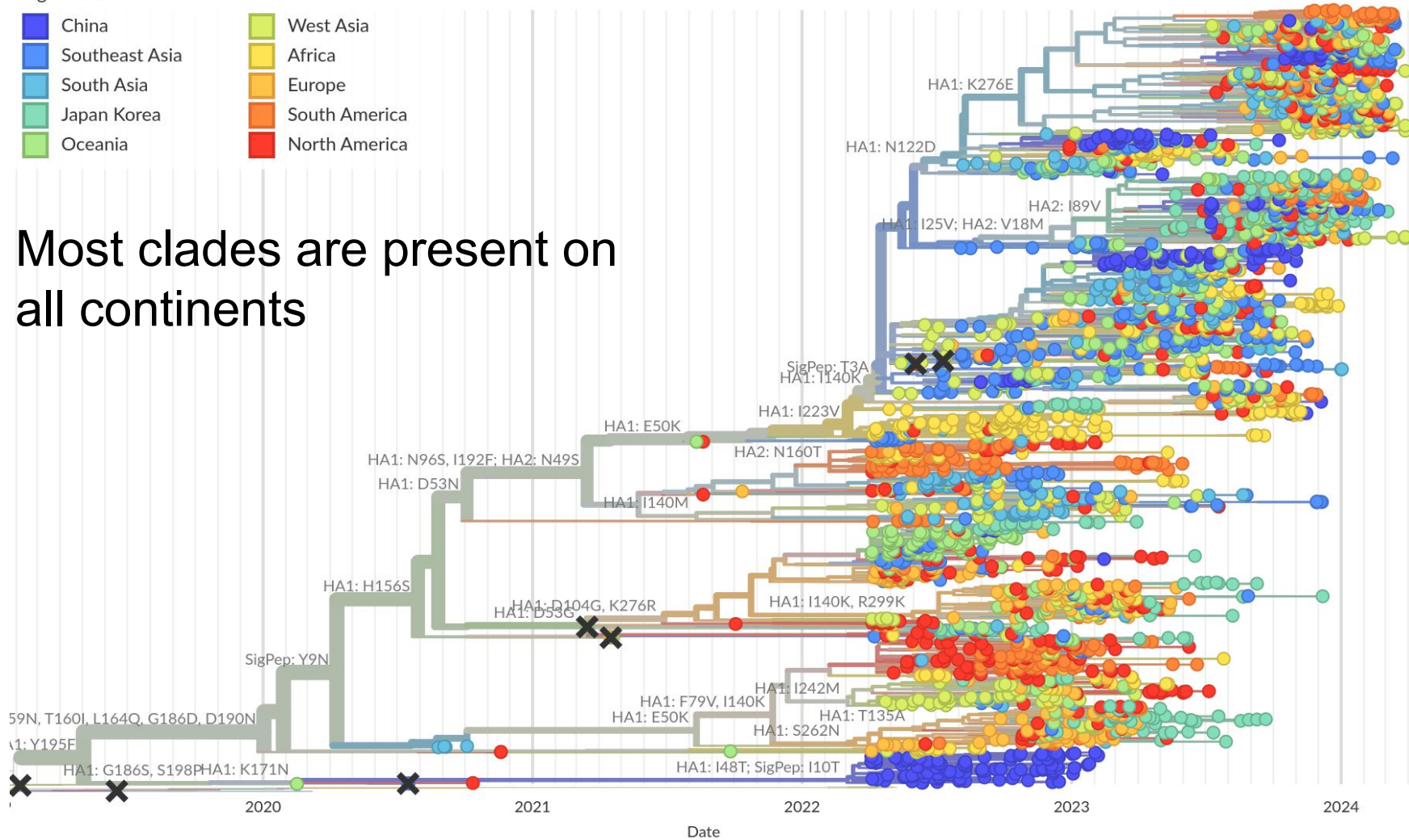

Trends in Microbiology

- Two WHO convened meetings per year
- Input data:
    - Epidemiology
    - Genetic diversity and dynamics
    - Antigenic characterization of many viruses using ferret serology
    - Human serology (pre/post vaccination titers)
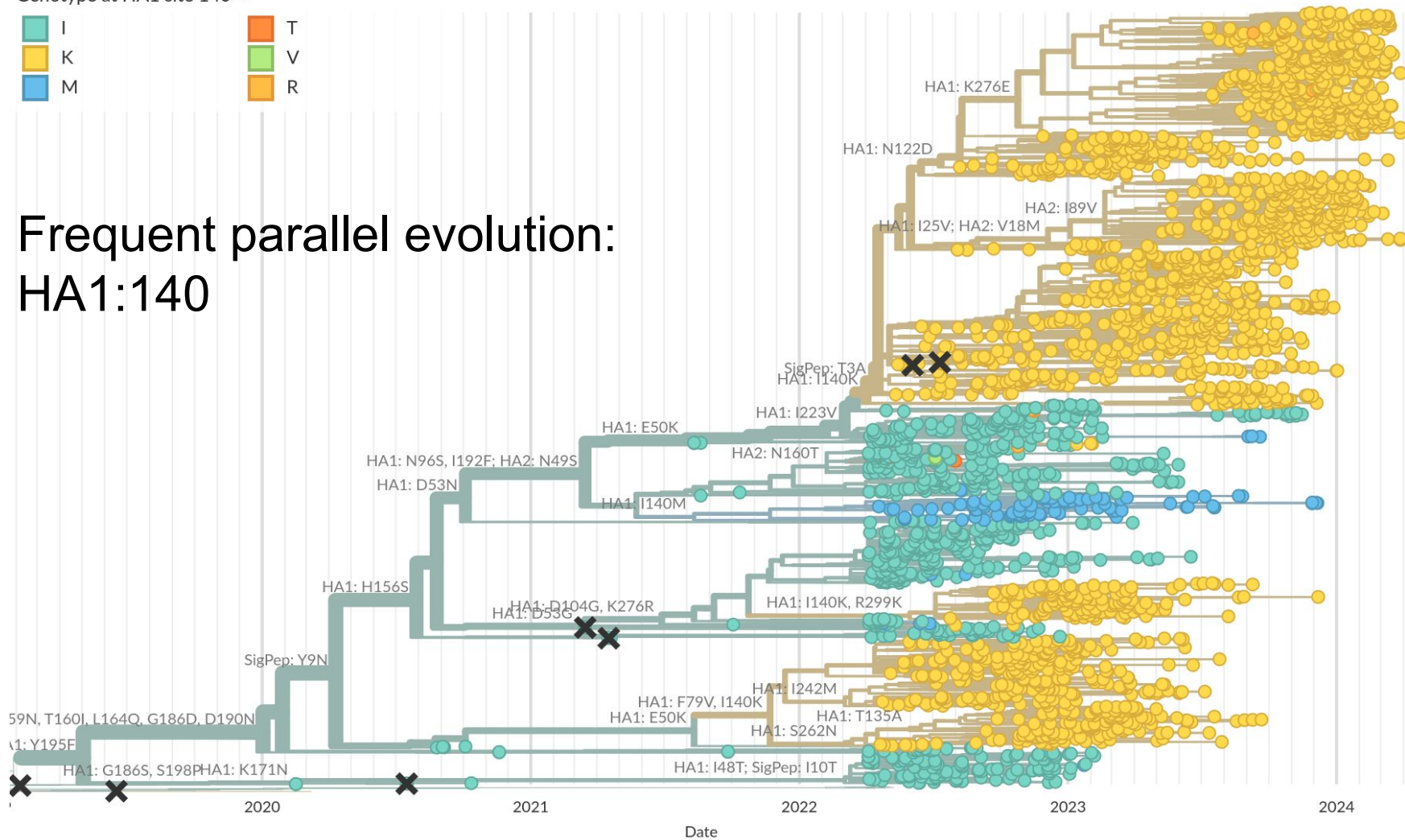   - Modeling and prediction

→ Vaccine composition recommendation
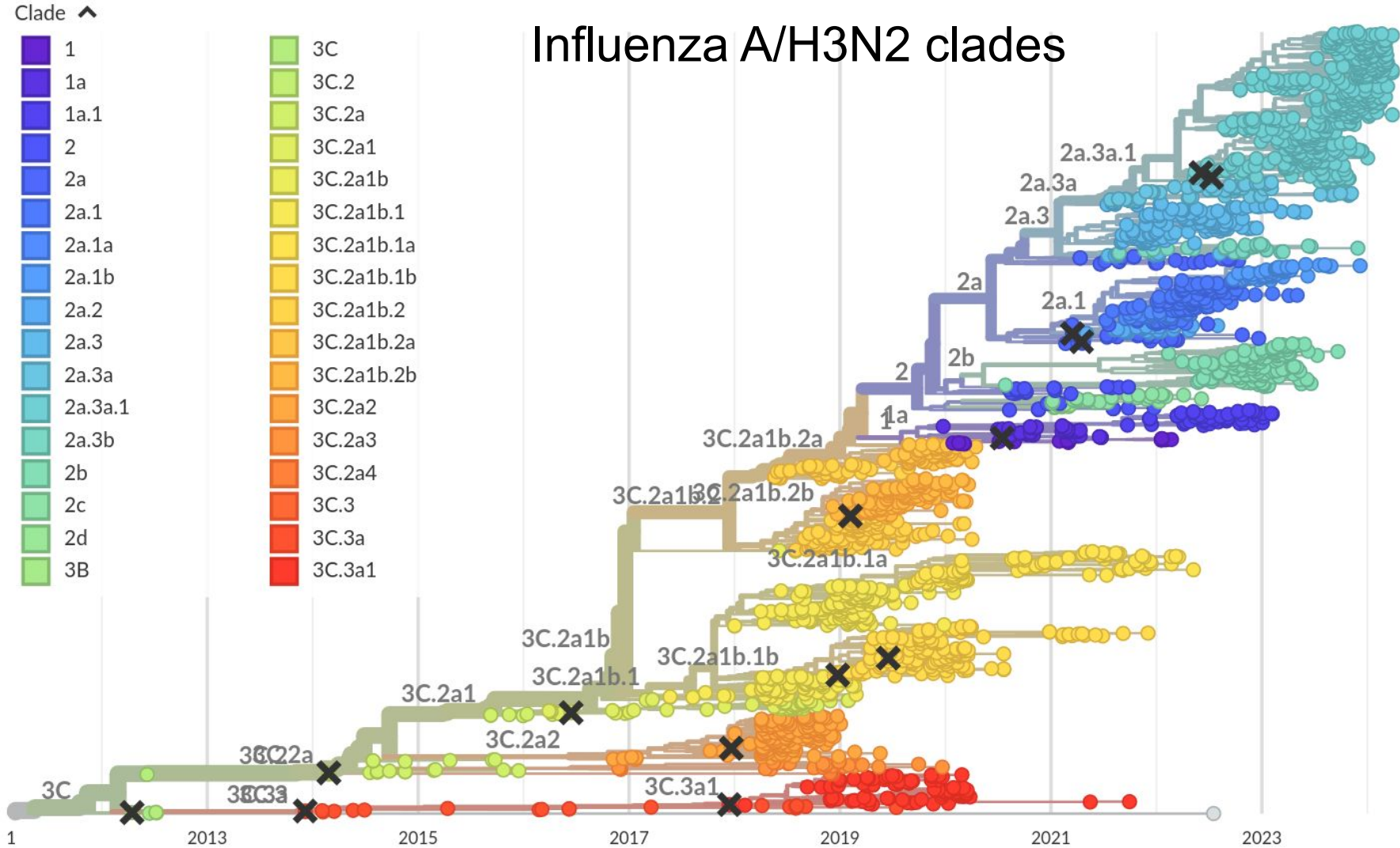
Most clades are present on all continents

Frequent parallel evolution:
HA1:140

Influenza A/H3N2 clades

Purpose of clade nomenclature

- Units of tracking and prediction
  - Frequency trajectories and geographic frequency differences
  - Frequency predictions/extrapolations
- Aggregation of phenotypic data
  - Neutralization titers
  - Resistance phenotypes
  - Virulence
- Reassortment analysis
  - Tracking genotypic constellations of clades across segments
- Communication
  - Signal that viruses within a clade are similar (no need for vaccine updates etc)

Tracking frequency trajectories of clades by region
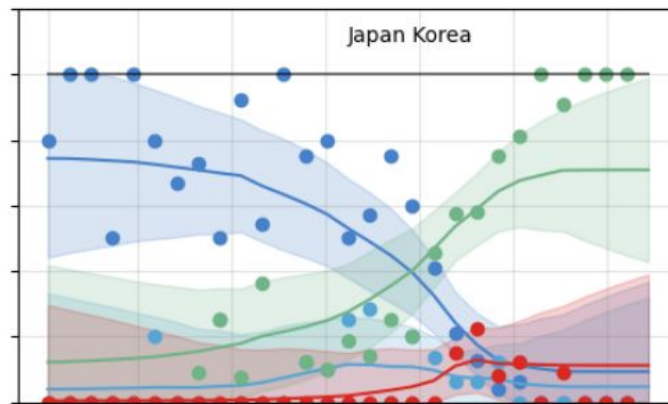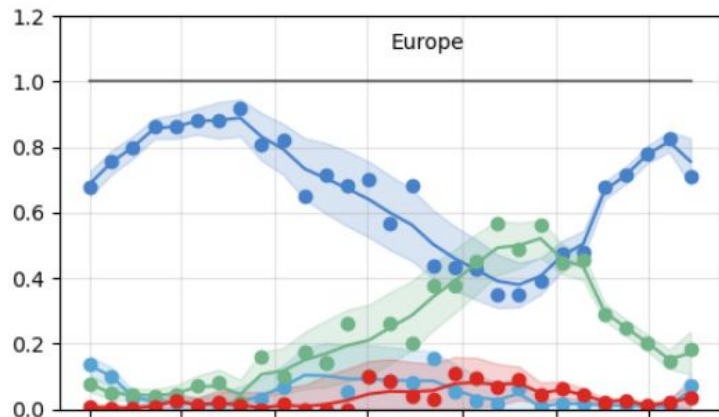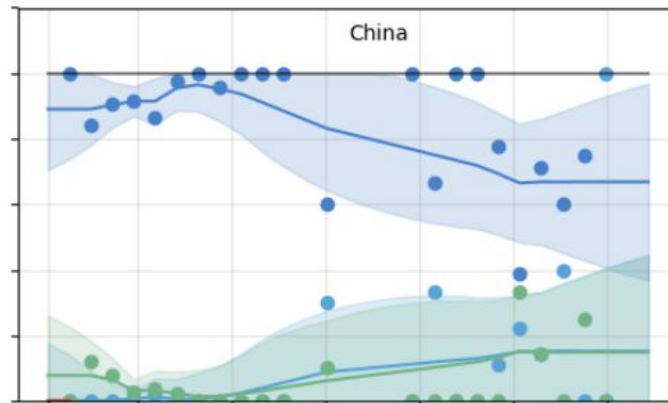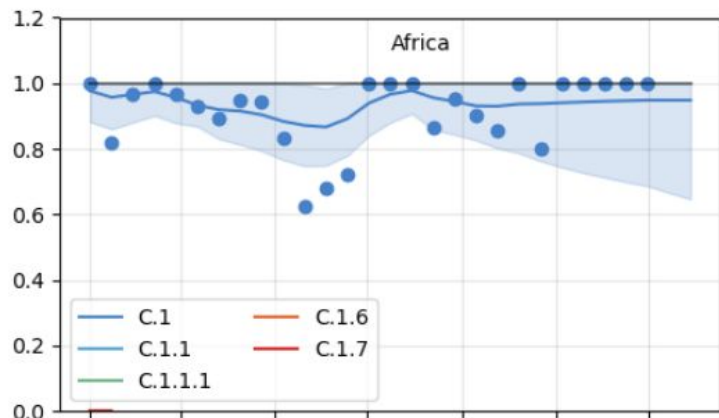
Aggregating serological data
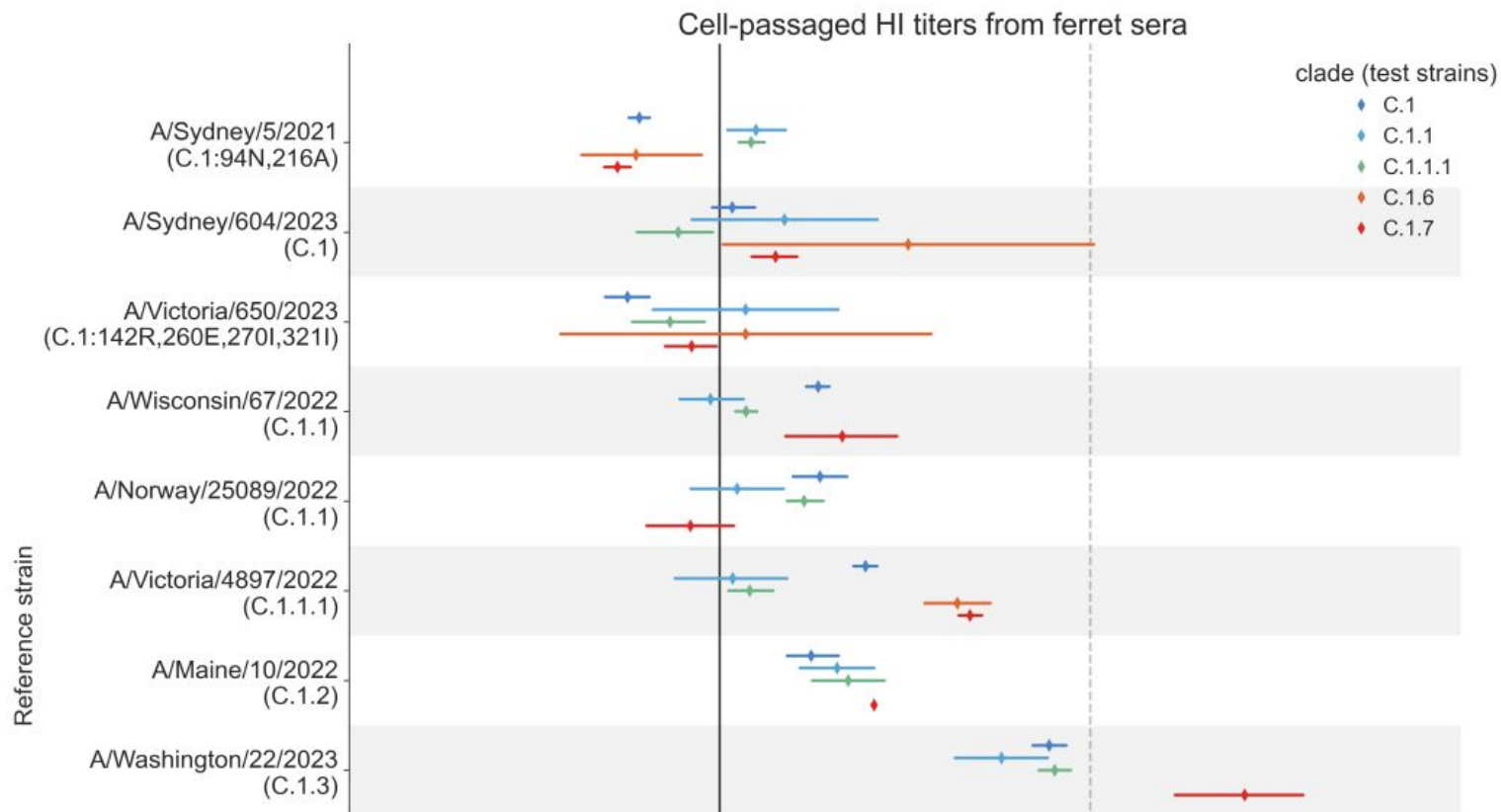
Reassortment analysis

Table 3.1 - Global A(H1N1)pdm09 Collected Since September 1, 2023
Number of H1N1pdm09 Sequences (Percent of Sequences in Each HA Subclade)

HA Clade_Subclade			Number of Seq	Percent in Total	NA Subclade									
					B.3.1.1	C	C.2	C.3	C.4	C.5	C.5.1	C.5.1.1	C.5.2	C.5.3
	6B.1A.5a.2a	C.1	3,389	52.42%	36 (1.06%)	13 (0.38%)				1,935 (57.10%)			3 (0.09%)	1,402 (41.37%)
		C.1.2	9	0.14%			8 (88.89%)							1 (11.11%)
		C.1.5	2	0.03%				1 (50.00%)		1 (50.00%)				
		C.1.7	367	5.68%				1 (0.27%)	345 (94.01%)	1 (0.27%)				20 (5.45%)
	6B.1A.5a.2a.1	C.1.1	140	2.17%	7 (5.00%)			73 (52.14%)					16 (11.43%)	44 (31.43%)
		C.1.1.1	2,558	39.57%					4 (0.16%)	36 (1.41%)	1 (0.04%)	302 (11.81%)	1,208 (47.22%)	1,008 (39.41%)

Number of Sequences
1 ▭▭▭▭ 1,935

US CDC, Becky Kondor and team.

Challenges with existing nomenclature

- Slow to update

  - New clades are often only defined after a previous clade has fixed

  - Different groups "invent" their own nomenclature (e.g. 2a.1b/53G)

  - Clades lose their value as units of tracking, prediction, or phenotype aggregation

- Names become longer and longer

  - Ad-hoc shortening of names results in ambiguities

- No central reference for updates to the nomenclature

New clades are often designated too late

In Influenza B/Vic, V1A.3a.2 almost fixed for 2 years
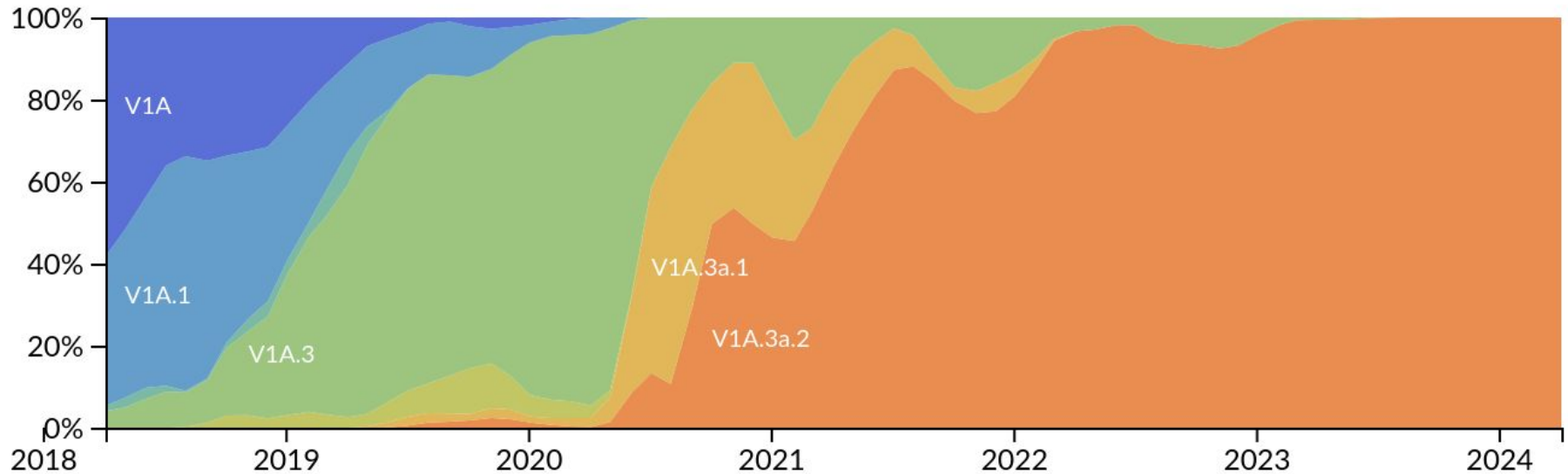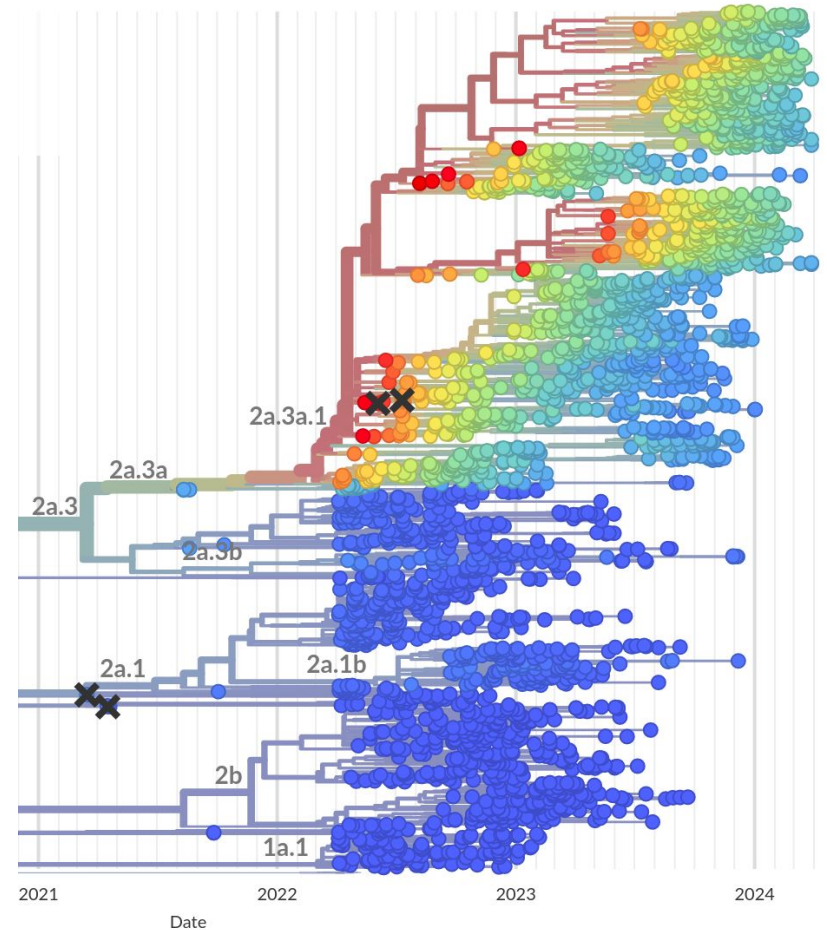
Prediction of IAV evolution

Several groups including:
- PreVir (Lukza, Ruchnewitz, Laessig)
- Nextstrain

Features:
- Phylogenetic signatures of expansion/growth
- Mutations in epitopes and around the receptor binding site
- Mutational load
- Parallel evolution
- Genetic distinctiveness
- Past circulation

Work in Progress: more dynamic classification

- Adopt a Pango-style `letter.number.number…` pattern with letters as aliases
- New clades: Suggested by an algorithm, picked manually
  - Importantly: addition to a labeled tree, not de-novo classification
- Criteria (inspired by predictive features):
  - Phylogenetic expansion
  - Branches with amino acid mutations, differentiated by importance
  - Divergence from parent clade
  - Minimal size
- Combined into a score
- Updates aligned with WHO vaccine composition meeting schedules
- Aliasing not necessarily automatic
- GitHub repo with machine and human readable definitions
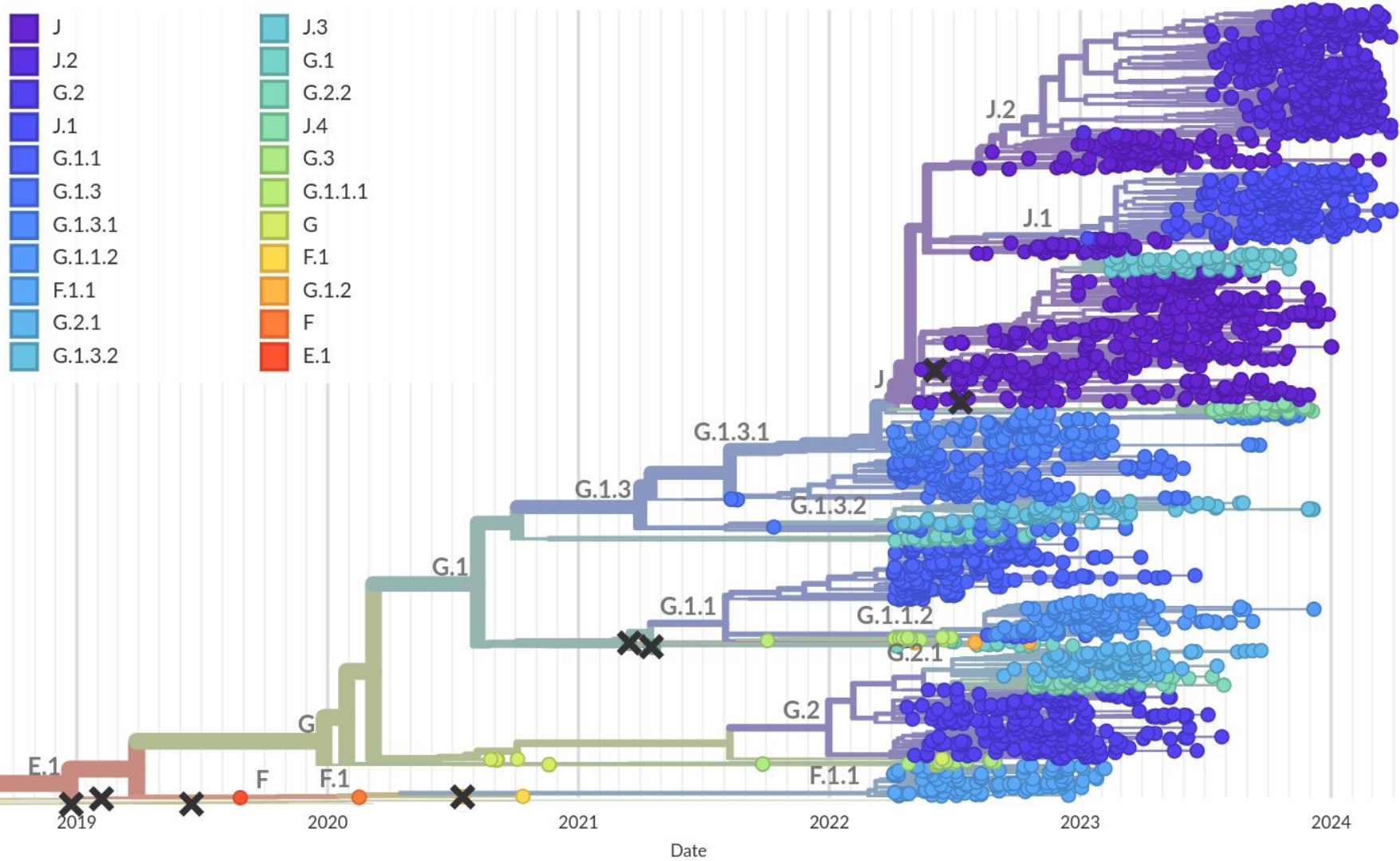
Multiple previous proposals

- Ramaekers et al. *Towards a unified classification for human respiratory syncytial virus genotypes.* 2020
- Goya et al. *Toward unified molecular surveillance of RSV: A proposal for genotype definition*. 2020
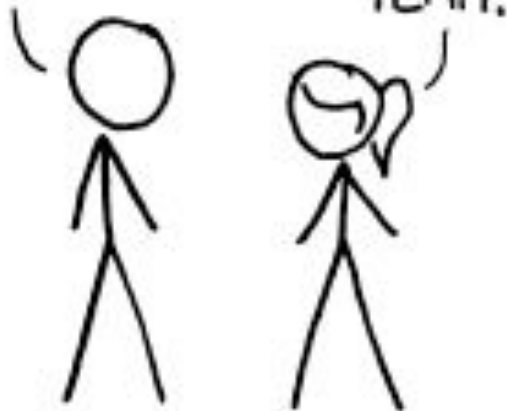- …

→ Lack mechanism to update
→ Reduced information content by the time they were published
   (only 1-2 clades still in circulation)

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION: THERE ARE 14 COMPETING STANDARDS.

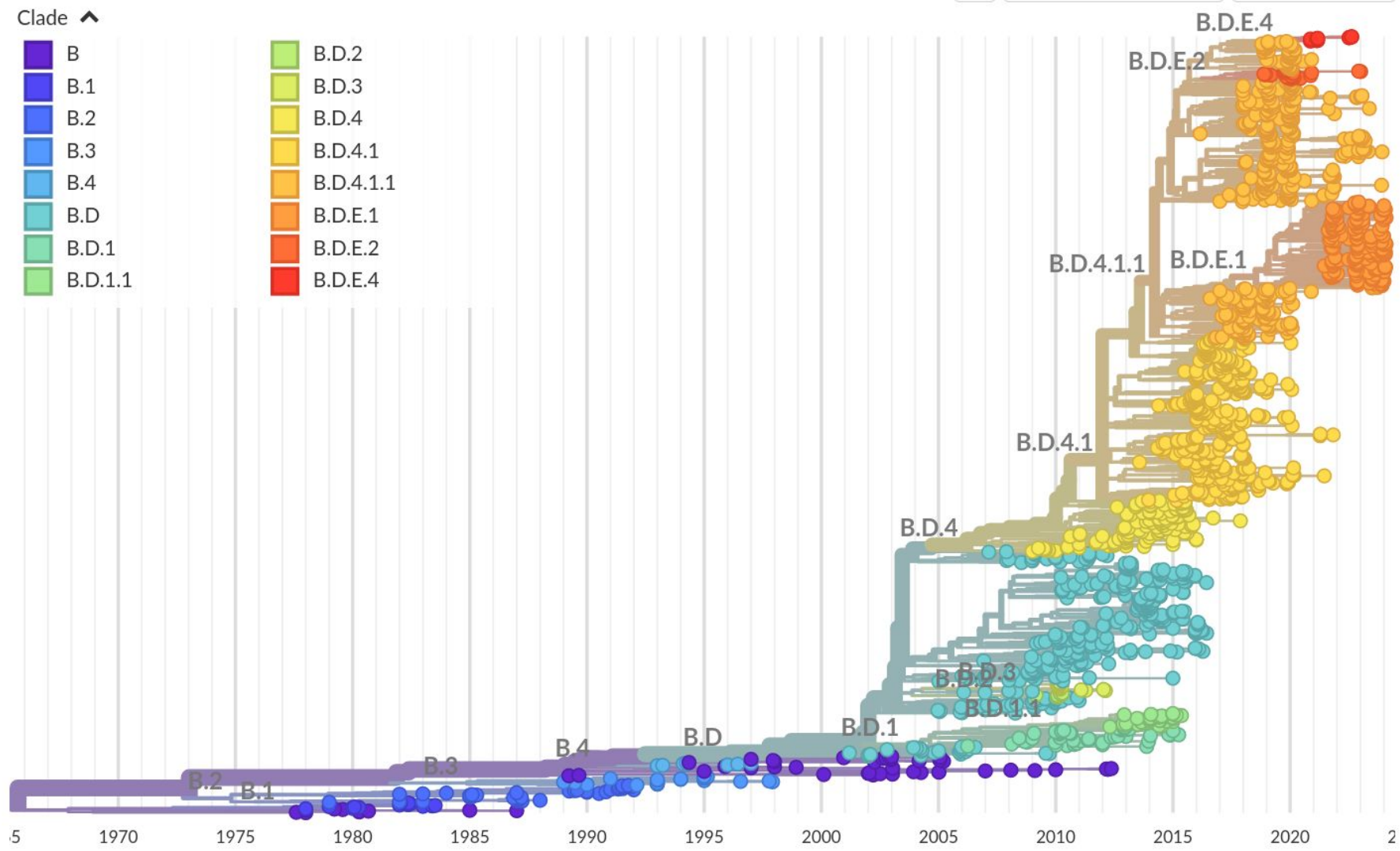14?! RIDICULOUS! WE NEED TO DEVELOP ONE UNIVERSAL STANDARD THAT COVERS EVERYONE'S USE CASES.

YEAH!

SOON:

SITUATION: THERE ARE 15 COMPETING STANDARDS.

https://xkcd.com/927/

RSV consortium proposal

- Led by Stephanie Goya, preprint on medrxiv
- Broad consortium with many relevant stakeholders
- Pango-style
- "Cumulative" aliasing: B.D.4.1.1 → B.D.E

The unified proposal for classification of human respiratory syncytial virus below the subgroup level

Stephanie Goya, Christopher Ruis, Richard A. Neher, Adam Meijer, Ammar Aziz, Angie S. Hinrichs, Anne von Gottberg, Cornelius Roemer, Daniel G. Amoako, Dolores Acuña, Jakob McBroome, James R. Otieno, Jinal N. Bhiman, Josie Everatt, Juan C. Muñoz-Escalante, Kaat Ramaekers, Kate Duggan, Lance D. Presser, Laura Urbanska, Marietjie Venter, Nicole Wolter, Teresa C. T. Peret, Vahid Salimi, Varsha Potdar, Vítor Borges, Mariana Viegas

General thoughts

- Publications are fine to announce criteria and systems, but are not suitable for frequent updates
- Dynamic updates need to be built into the system and have a low barrier
- Consensus is more important than "perfect"
- Requires a dynamic repository as a source of truth, ideally run by the community → Github has worked well for SARS-CoV-2, mpox
- Definitions and representative sequences should be openly accessible
- Challenges:
  - Balancing proliferation of names and sufficient granularity; depends on the use case
  - Balancing aliasing with preserving hierarchical relationships
  - Competing objectives: tracking emerging variants vs labeling phenotypically different groups
  - Pronouncability (3C.3b.2b) and memorability
- Easy to use tools to assign sequences to clades/lineages/cluster are critical

Acknowledgements

- WHO Collaborating Centers for Influenza

- Scientists and laboratories that share viral sequence data via INSDC or GISAID

- Stephanie Goya and the RSV nomenclature consortium

- Nextstrain team