Insights at your fingertips - GISAID's Enhanced Tool Ecosystem

Krista Queen[1], Raphael Tze Chuen Lee [2], Yani Xu [2], Lucas Freitas [3,4,5], Ondina Palmeira [4], Felipe Andre Silva [3], Agnello Picorelli [4], Gunter Bach [4], Yuri Rodrigues [4], Gabriel Luz Wallau [1,5], Suma Tiruvayipati [2,3], Sandy Tze-Minn Mak [2], Sebastian Maurer-Stroh [2]

[1] GISAID Database Technical Committee @ LSU Health Shreveport
[2] GISAID Data Science Centre @ Bioinformatics Institute, A*STAR, Singapore
[3] GISAID Global Curation Team
[4] GISAID Technical Development Team
[5] Oswaldo Cruz Foundation (FIOCRUZ), Brazil

INTRODUCTION

Timely access to genomic and associated metadata is critical for responding to infectious threats (1). Although numerous data repositories containing sequence data scatter the landscape, the ability to integrate vast and complex datasets and extract meaningful data remains a challenge. User-friendly tools to identify the emergence and spread of new variants, evaluate their potential risk and define populations most vulnerable to these threats, are essential for an effective "sequence-first" approach to surveillance. Created in 2008 and maintained by scientists for scientists, the GISAID data science platform accelerates the extraction of actionable insights from its comprehensive pathogen specific databases through a seamlessly integrated suite of tools.

METHOD

Built on the foundation of trust, GISAID respects the rights of data submitters through its unique data governance (Fig. 1) (2).

Our commitment to respond to evolving public health needs guides the continuous development of GISAID's database to support its growth (Fig. 2).
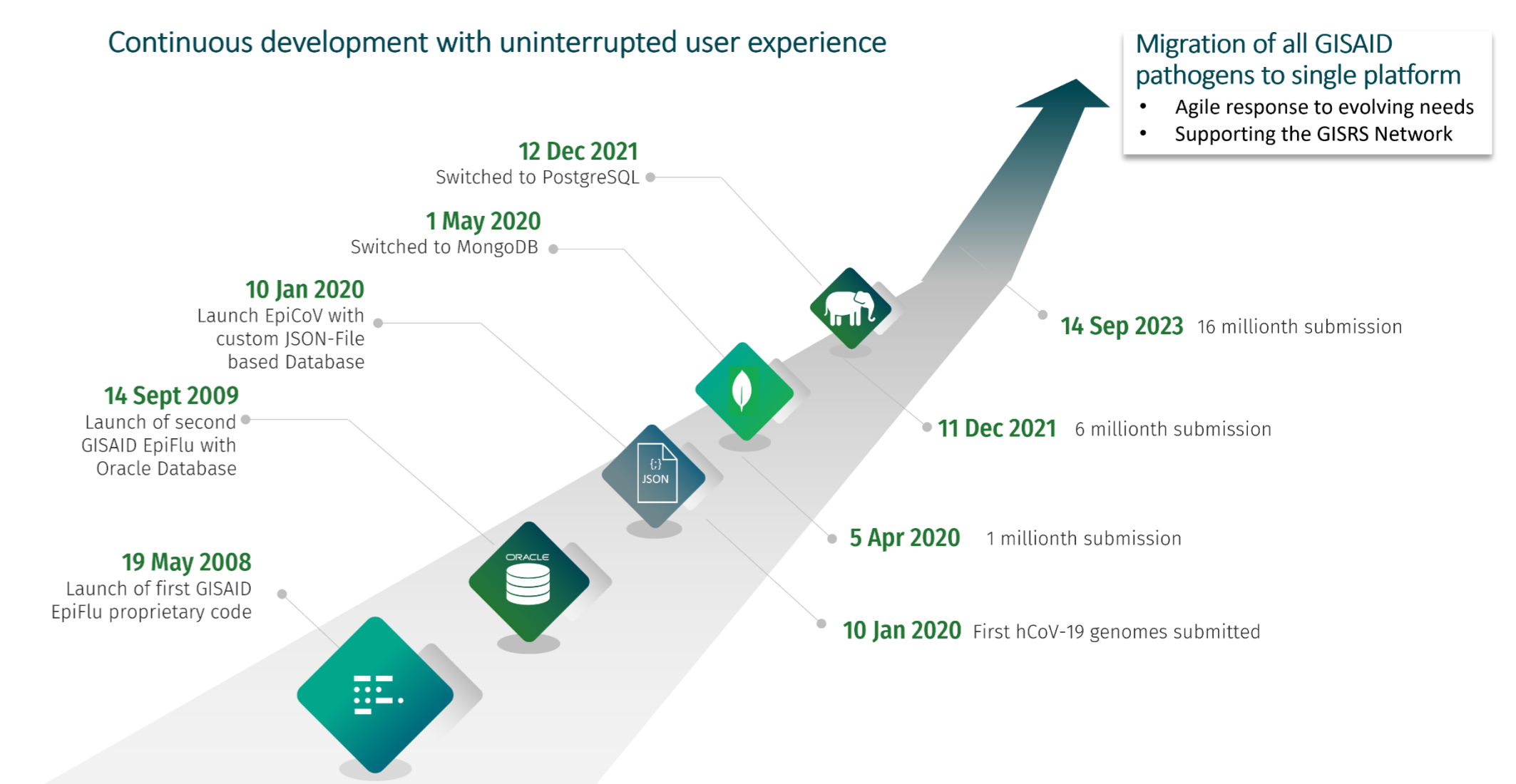

Fig. 1: GISAID's data sharing philosophy

Fig. 2: Seamless evolution of GISAID Database Architecture

INTEGRATED ECOSYSTEM OF TOOLS

GISAID's enhanced tools ecosystem allows users to inspect data from single entries to user-selected datasets with its intuitive search and browse tool. As all sequences are quality-checked and annotated during curation, they can be readily tracked across subtypes, clades, and defined mutation constellations (emerging variant), right down to an individual mutation (Fig. 3). Each tool generates an interactive output, with customizable graphs and reports that can be readily downloaded from the database, thus empowering data-driven conclusions at your fingertips.

An ecosystem of integrated tools supporting scale-free view of virus evolution

Fig. 3: An ecosystem of integrated tools to facilitate comprehensive analyses

Question	Tool	Question	Tool
Which sequences in the database are most similar to my own sequence?	AudacityInstant	Are there any mutations affecting my primers?	PrimerChecker
What are the emerging variants with unique constellation of mutations?	Emerging Variants	Can I browse, search and filter entries in the database?	Search & Browse
How can I create summary graphs for hits from a custom database search?	EpiCharts	Summary of phylogenetic evolution in subsampled tree?	Phylodynamics Global
Are there subclusters in my dataset and is there correlation with clinical or epidemiological phenotypes, based on input metadata?	FluCluster-AI	Are there phenotypically important mutations in my sequences (host specificity, antibody binding, drug resistance, glycosylation, virulence)?	FluSurver

Fig. 4: Tools to empower actionable data- Answering your questions.

FROM EQUITABLE DATA ACCESS TO EQUITABLE DATA SCIENCE – READY FOR DISEASE X

GISAID, the Global Data Science Initiative provides transparent access to the world's largest collection of genomic and associated metadata of high-priority pathogens. With about 20 million curated sequences, shared by 215 countries and territories, GISAID remains an essential asset for the WHO GISRS-network enabling genomic surveillance for Influenza, RSV, SARS-CoV-2, in addition to Mpox and Arboviruses e.g. Dengue (2, 3).

GISAID continually evolves to meet the needs of its users by offering:

- Options for high-throughput data submissions
- Assurance of high-quality data through sequence review and annotation performed by a global curation team
- Tools and data packages that empower equitable data utilization

With its long-standing collaboration with public health and research laboratories worldwide, GISAID is poised to respond to Disease X (Fig 5).
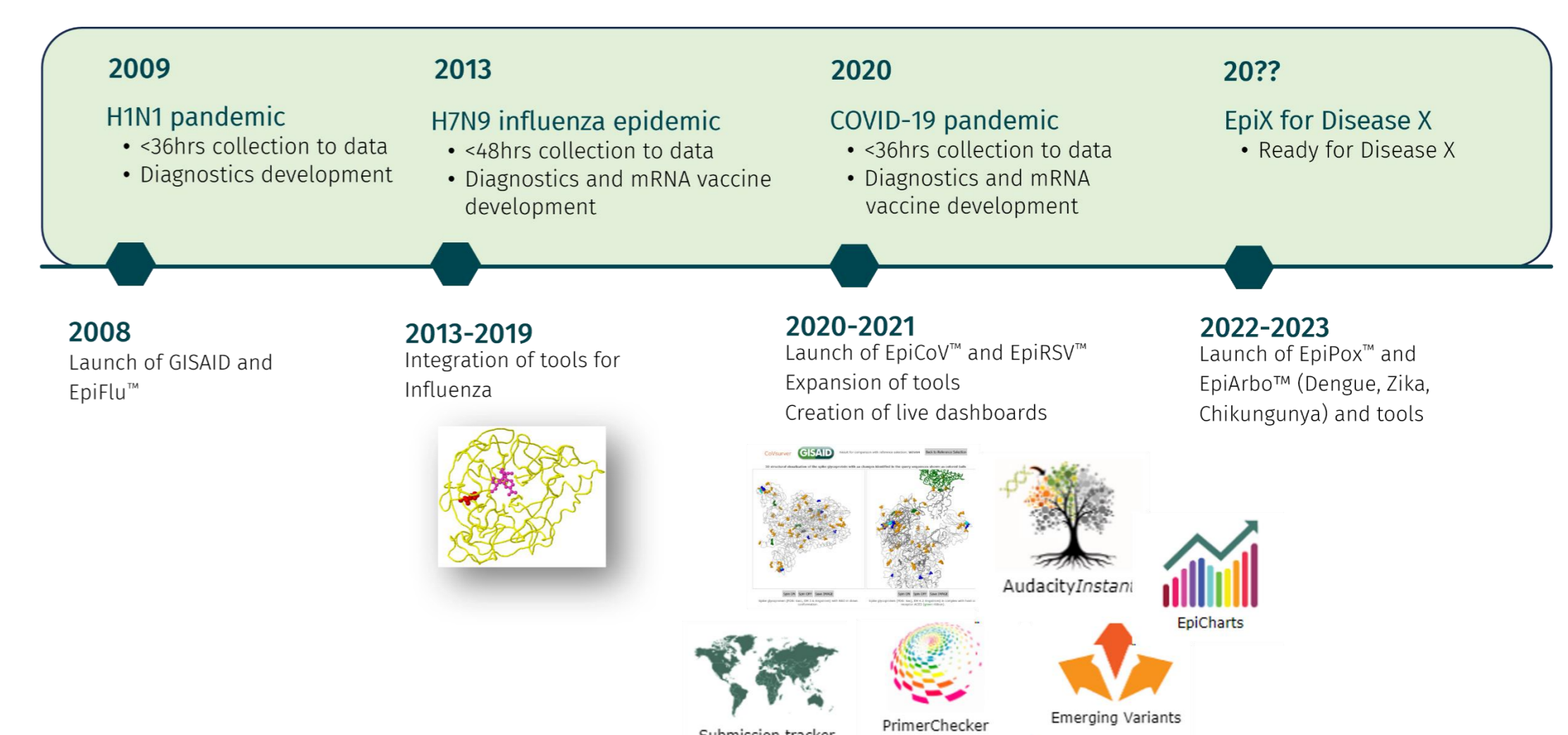

Fig. 5: Key milestones- Building resilience for future pandemics

ACKNOWLEDGEMENTS

REFERENCES

1. Carter LL, Yu MA, Sacks JA, Barnadas C, Pereyaslov D, Cognat S, et al. Bull World Health Organ 2022 Apr 1;100(4):239-239A
2. Elbe S, Buckland-Merrett G. Global Challenges 2017 Jan;1(1):33
3. Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. China CDC Wkly 2021 Dec 3;3(49):1049–51.