

BASE Research

Evaluating AI agents for pathogen genomic surveillance

First meeting
12 October 2026

Bryan Tegomoh, MD, MPH

Introductions and today's agenda

Conversation	Time
Our backgrounds, skills and learning goals	15 min
Public health and genomic epidemiology	15 min
Research question, data and agent evaluation	15 min
Working agreement and first deliverables	10 min
Questions and next meeting	5 min

Genomics in public health

Detect and investigate clusters

Sequence similarity gains meaning from exposure and sampling data.

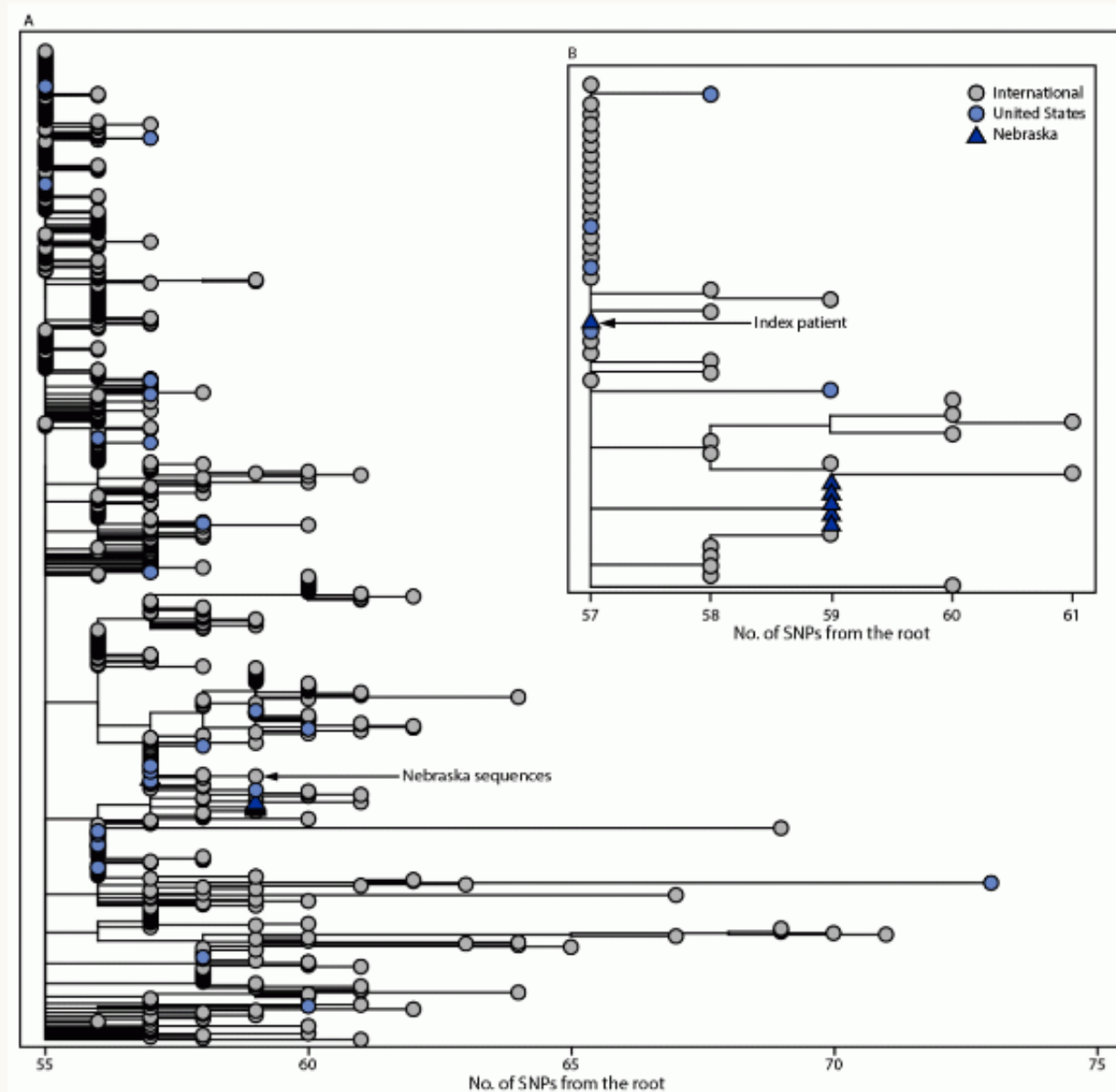
Follow changes over time

Lineage and clade patterns depend on who and what gets sampled.

Connect evidence to a response

Laboratory findings inform an investigation alongside epidemiology.

Nebraska's Omicron household cluster



Six household cases

Genomic relatedness

Exposure and travel history

Timing of illness

What could an agent conclude from the tree alone?

Clinical, wastewater and environmental sampling

Source	Observation	Interpretation
Clinical	Specimen or isolate linked to a sampled person	Case context and sampling selection
Wastewater	Mixed signal from a catchment and time	Community trends, with mixture uncertainty
Other environmental	Sample from water, air or a surface	Collection setting and background context

Wastewater is environmental surveillance.
Targeted and metagenomic describe sequencing approaches.

The research question

How accurately and reproducibly can supervised AI agents analyze sequencing data to support pathogen cluster detection and variant surveillance?

How does orchestration affect performance against a defined baseline?

Clinical and environmental data, including wastewater and metagenomics.
One primary setting first, with an agreed comparison.

Agent orchestration and human review

Responsibility	What it produces
Analyst / planner	A bounded question and proposed analysis
Agent with allowed tools	Outputs with a record of its actions
Independent checks	Reference comparisons and flagged uncertainty
Human reviewer	An accepted interpretation or a request for more evidence

Proposed comparison: baseline, one agent, then justified role separation.

A fictional surveillance summary

What proportion of eligible records belong to category A?

Supplied summary	Records
Total specimen records	60
Meet the stated eligibility rules	50
Category A among eligible records	10

What would you check?

The denominator
The missing context
The limits of the claim

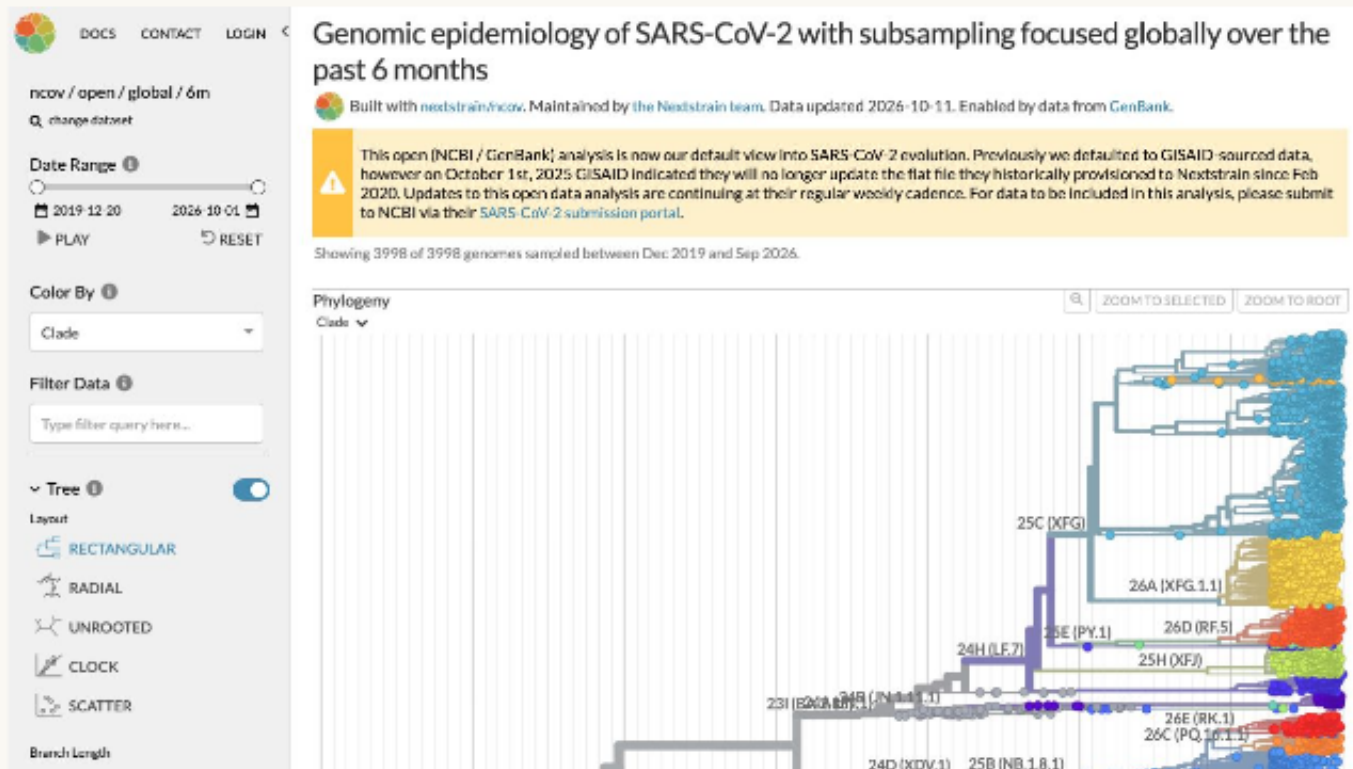
Illustrative data. Category A has no biological meaning.

Databases and the evidence they hold

Resource	What it provides	Scope to check
NCBI GenBank	Nucleotide sequences and annotations	Record quality and linked metadata
NCBI SRA	Raw sequencing data and alignments	Study design, assay and access terms
GISAID	Influenza and SARS-CoV-2 data	Access agreement, attribution and reuse
NCBI Pathogen Detection	Comparative information for supported organisms	Bacterial / fungal coverage and version

Database coverage and sampling determine what an agent can infer.

Tools and the questions they answer



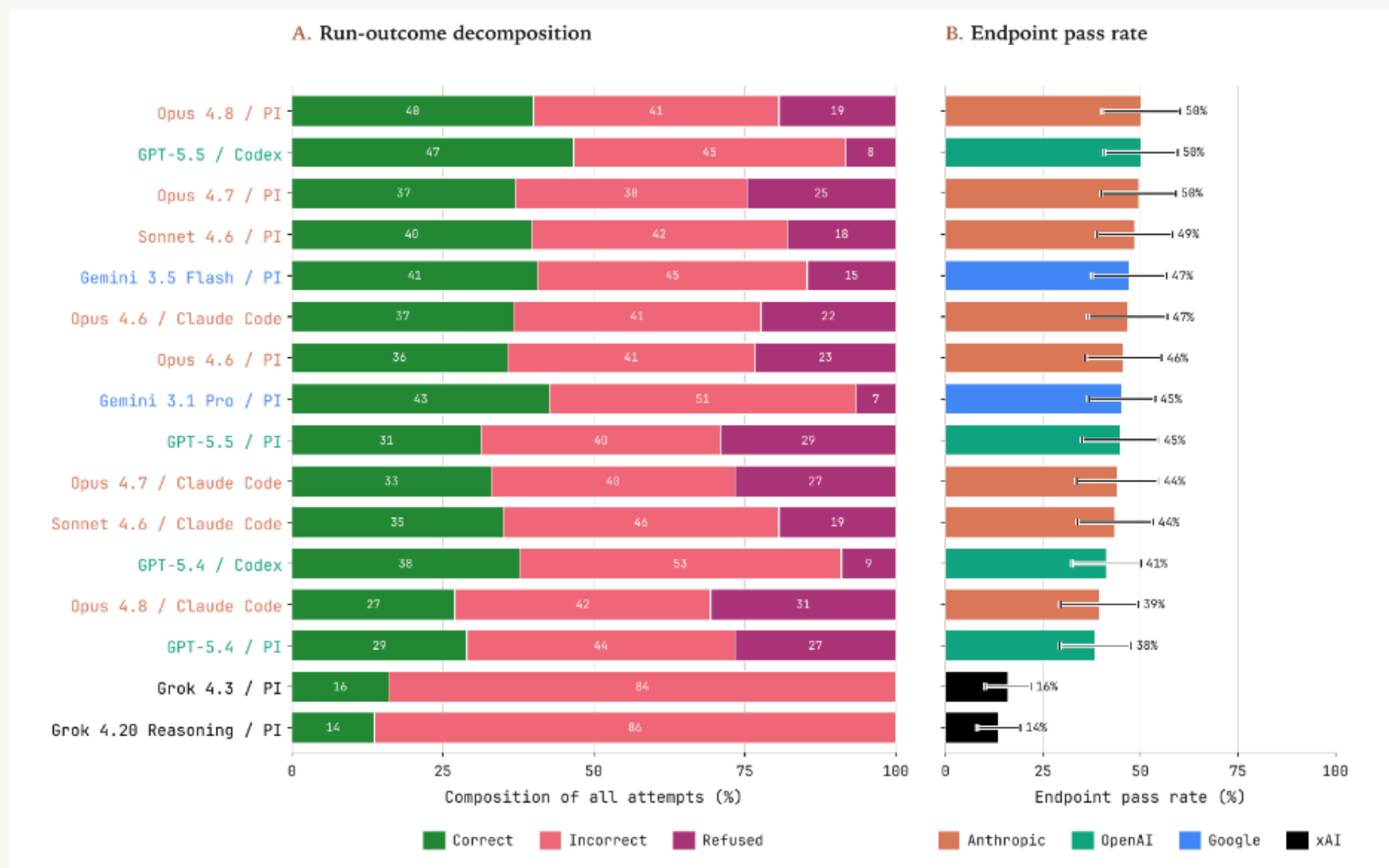
Nextstrain public view

A tree still needs sampling context.

Image: nextstrain.org

Tool / system	Role
Pango / Pangolin	Lineage names / lineage assignment
Nextclade	QC and clade outputs
Nextstrain	Trees and metadata
Python / R	Reference calculations
Inspect AI	Agent evaluation and logs

BioSecBench-Surveillance: published results



A: all attempts. B: mean pass rate among gradable attempts, excluding refusals.

Evidence for a credible evaluation

01 A checkable reference

Expected fields, calculation, and source evidence

02 Separate scoring rules

Deterministic checks plus a review of interpretation

03 Controls for shortcuts

Neutral filenames and reordered records

04 An auditable run record

Baseline, allowed tools, versions, errors, and all attempts

A manageable first month

Proposed checkpoint	Artifact for discussion
19 October	Reading notes, candidate questions and constraints
26 October	Joint two-page scope note and one task sketch
2 November	One worked example with a reference answer and rubric
16 November	A small example set and a draft failure taxonomy

Agree the pace together. Each checkpoint produces something reviewable.

Working agreement and decisions

Question and scope	One primary setting and a clear comparison
Ownership	A lead for each artifact, with cross-review
Support	Learning goals, feedback preferences and constraints
Cadence	Weekly meeting, brief progress note and decision log
Next checkpoint	A feasible date and an artifact each person can explain

Reading and resources

Pathogen Genomics in Public Health

Armstrong et al., NEJM 2019. Includes Duncan MacCannell.

Nebraska Omicron cluster investigation

Jansen, Tegomoh et al., MMWR 2021.

BioSecBench-Surveillance

Bhasin et al., arXiv 2026. Preprint.

WHO genomic surveillance strategy