

sequence database
(NCBI, GISAID)

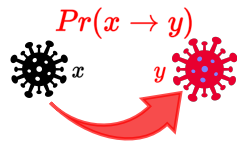


a.

Infer dependencies
across mutations
and genomic
variations



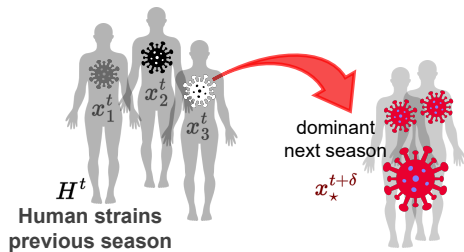
- Quantify probability of specific mutations in the wild
- Implicitly factor in functional constraints and time-varying selection pressures



c.

Application 1:

Forecast dominant strain(s) in upcoming season

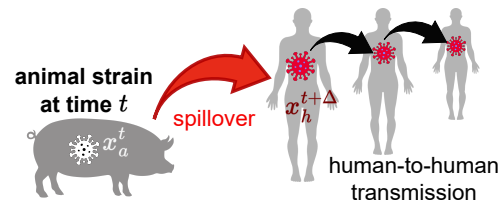


$$x_*^{t+\delta} = \underset{x}{\operatorname{argmax}} \prod_{x^t \in H^t} \Pr(x^t \rightarrow x)$$

d.

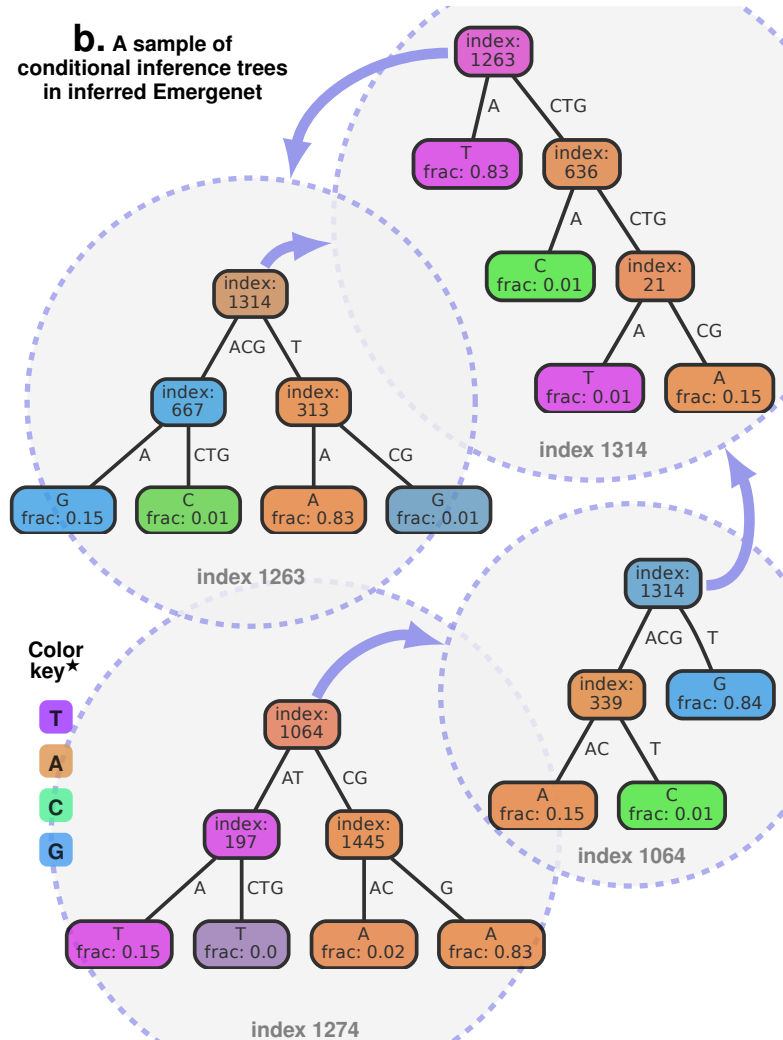
Application 2:

Estimate pandemic potential of animal strains



$$\text{risk}(x_a^t) \propto \langle \log \Pr(x_a^t \rightarrow x_h^{t+\delta}) \rangle$$

b. A sample of
conditional inference trees
in inferred Emergenet



H1N1 2018

Haemagglutinin Sequences

GGAAACAAAAGCAACAAAA...TGAA^AAAAAAGA...
AGCAAAAGCAGGGGAAACAA...GTT^CAACCAC...
AGCAAAAGCAGGGGAAACAA...GTT^TAACCAC...
ATGAAGACTATCATTGCTTT...ACC^TTGAGAA...

residue
at index 1274

★ mixed colors
represent distributions

A/Italy/7366/2018
A/Baltimore/P0264/2018
A/Baltimore/P0278/2018
A/Florida/61/2018