STATS 700-002 Class 1.
Background on phylodynamics

Edward Ionides and Aaron King

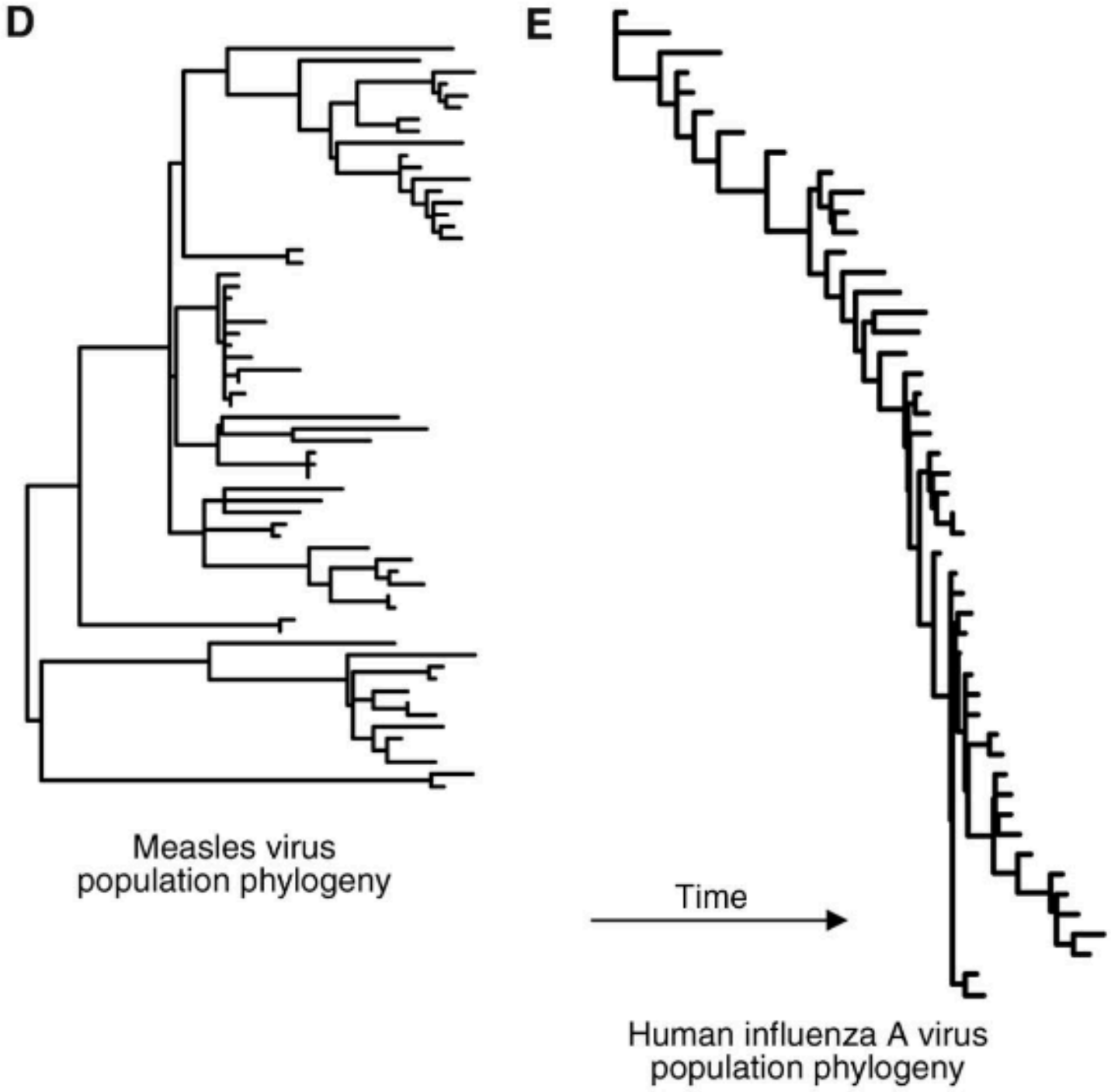
August 28, 2025

Outline

- The pre-history of phylodynamics and its initial synthesis by Grenfell et al. (2004).
  - Increases in the availability and length of genetic sequence data
  - Challenges for interpretation of the resulting trees of evolutionary relationships.
  - A need for model-based statistical inference
- A brief overview of advances 2005-2025 motivating this course.
  - Continuing growth in data collection and developments in inference methodology
  - Lessons from the COVID-19 pandemic
  - Current challenges, scientific and statistical

Glossary

- Pathogen. Virus, bacterium or other micro-organism which multiply in a host, causing disease.
- Phylogeny/genealogy. Ancestral tree for a group of individuals. Here, the individuals are pathogens.
- Epidemiology. The study of the cause and spread of disease, defaulting to infectious disease of humans.
- Immunity. Partial or complete protection from infection.


D
E
Measles virus
population phylogeny
Time
Human influenza A virus
population phylogeny

What is phylodynamics?

The forward problem: How do observed phylogenies arise from "pathogen genetic variation, modulated by host immunity, transmission bottlenecks, and epidemic dynamics."

The inverse problem: Inferring disease dynamics from observed phylogenies. What scientific & public health questions can phylogenies address?

Note: raw data are genetic sequences (or possibly phenotype measurements) not trees. We may analyze derived data, "observed trees," or the raw data.

The need for models and formal statistical inference

- Good descriptive statistics and graphics are important for data analysis.
- Formal quantitative inference also has a role.
  - Models about which we can make statistical statements to compare them to data.
  - What are the tools we need for quantitative phylodynamics?

Reading Grenfell et al. (2004) as a statistical scientist in 2025.

- There are many words to look up on Google/Wikipedia.
- The paper focuses on the forward problem (how biology affects the observed tree).
- This raises the possibility of the inferential inverse problem, but does not tell you how to do the statistics.
- An example of a forward-looking review paper.

Statistical/probabilistic abstraction

- Hosts are susceptible (S), infected/infectious (I), or recovered/removed (R).
- Chance interactions between S and I hosts lead to a transmission event ($S \to I$) and a new infection.
- Supposing each new infection has a unique source, called its ancestor, this sequence of random infection events implies a random ancestral tree connecting any collection of observed individuals.
- Probability models of this kind, built up of random variables that determine how the process progresses through time, are called **stochastic processes**.
- The corresponding statistical inverse problem is **inference for stochastic processes**.

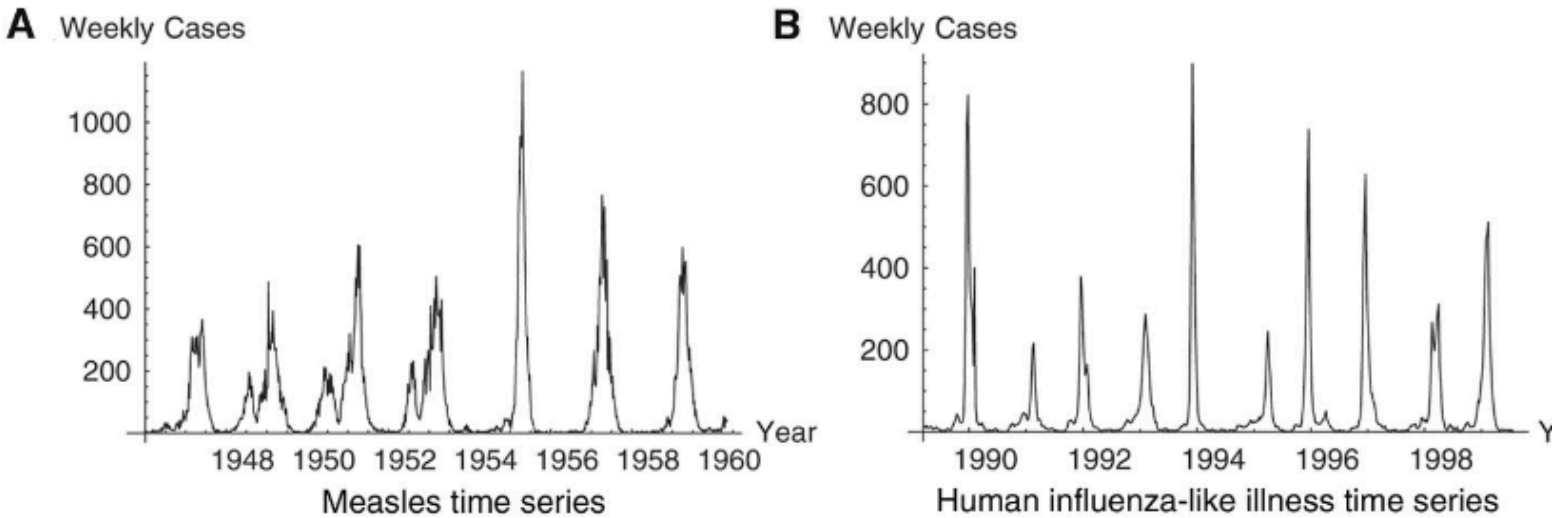
Knowledge pre-test

Which of the following are at least somewhat familiar to you? If you can, write a few words to say what they mean to you.

Epidemiology. DNA. RNA. Darwinian evolution. Virus. Vaccine. R_0.

Stochastic processes. Markov chain. Kolmogorov's forward equation. Branching process. Birth-death process. Brownian motion. Poisson process. Stochastic differential equation.

Statistics. Efficient estimate. Maximum likelihood estimate. Akaike's information criterion. Importance sampling. Markov chain Monte Carlo. Sequential Monte Carlo. Fisher information.


A
Weekly Cases
1000
800
600
400
200
Year
1948 1950 1952 1954 1956 1958 1960
Measles time series
B
Weekly Cases
800
600
400
200
1990 1992 1994 1996 1998
Human influenza-like illness time series

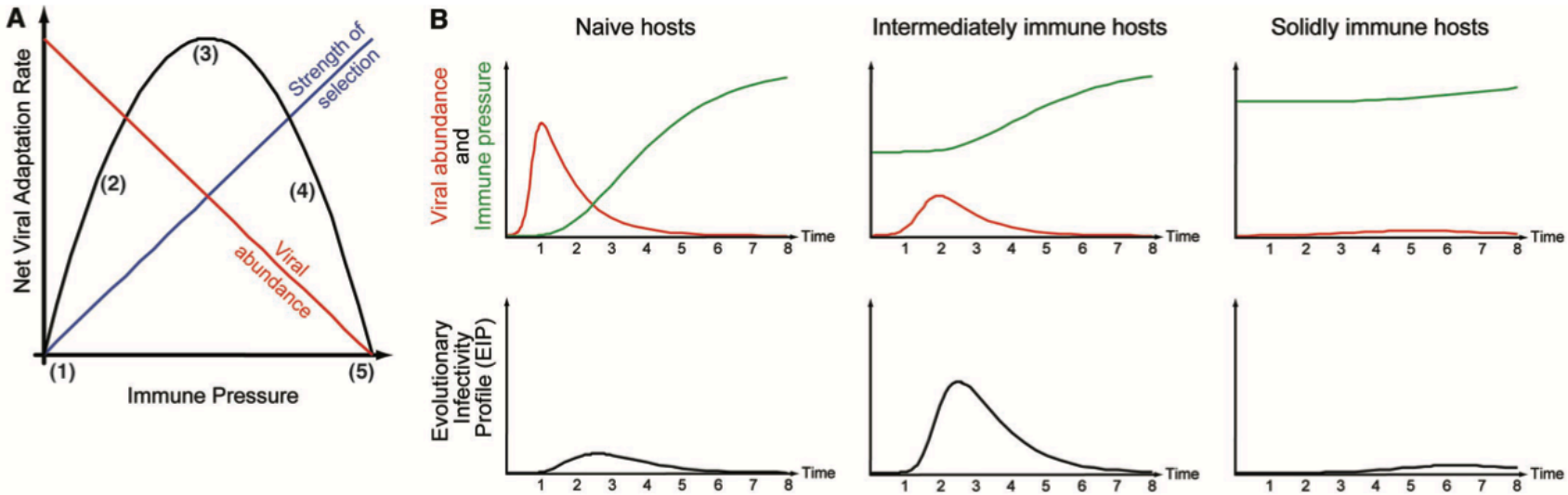


Fig. 2. (**A**) Schematic diagram of a static phylodynamic model for virus adaptation as a function of average immune pressure. Numbers correspond to phylodynamic patterns: 1) no effective response and no adaptation; 2) low immune pressure and low adaptation; 3) medium immune pressure and high adaptation; 4) high immune pressure and low adaptation; and 5) overwhelming immune pressure and no adaptation. (**B**) A phylodynamic framework allowing for within-host viral and immune kinetics. Time is measured in days after infection. Top: Schematic viral (red) and immunological (green) trajectories in individual hosts, based on experimental infection of horses with equine influenza virus (*28*, *35*). Bottom: The corresponding EIPs (*34*). Left, center, and right columns respectively reflect infection in naïve, intermediately, and solidly immune individuals. In naïve hosts, virus shedding generally peaks ~2 days after infection, declining to negligible levels by day 5. The humoral response rises by ~day 6, underlining the idea that innate immunity, loss of susceptible cells, or other mechanisms play the major role in initially limiting infection (*11*). The EIP for naïve hosts is relatively low, because little viral replication coincides with selective immunity, so these hosts are unlikely to be a major source of host-selected variants. The EIP for highly immune hosts is also very low, because adaptive immunity generally prevents substantial virus excretion, other than rare immune escape variants. For intermediately immune hosts, existing immunity limits viral excretion compared to the naïve case, also increasing earlier and more rapidly. The EIP shows a high potential for the transmission of selected viral variants, as substantial viral replication occurs during a time of substantial immune selection.

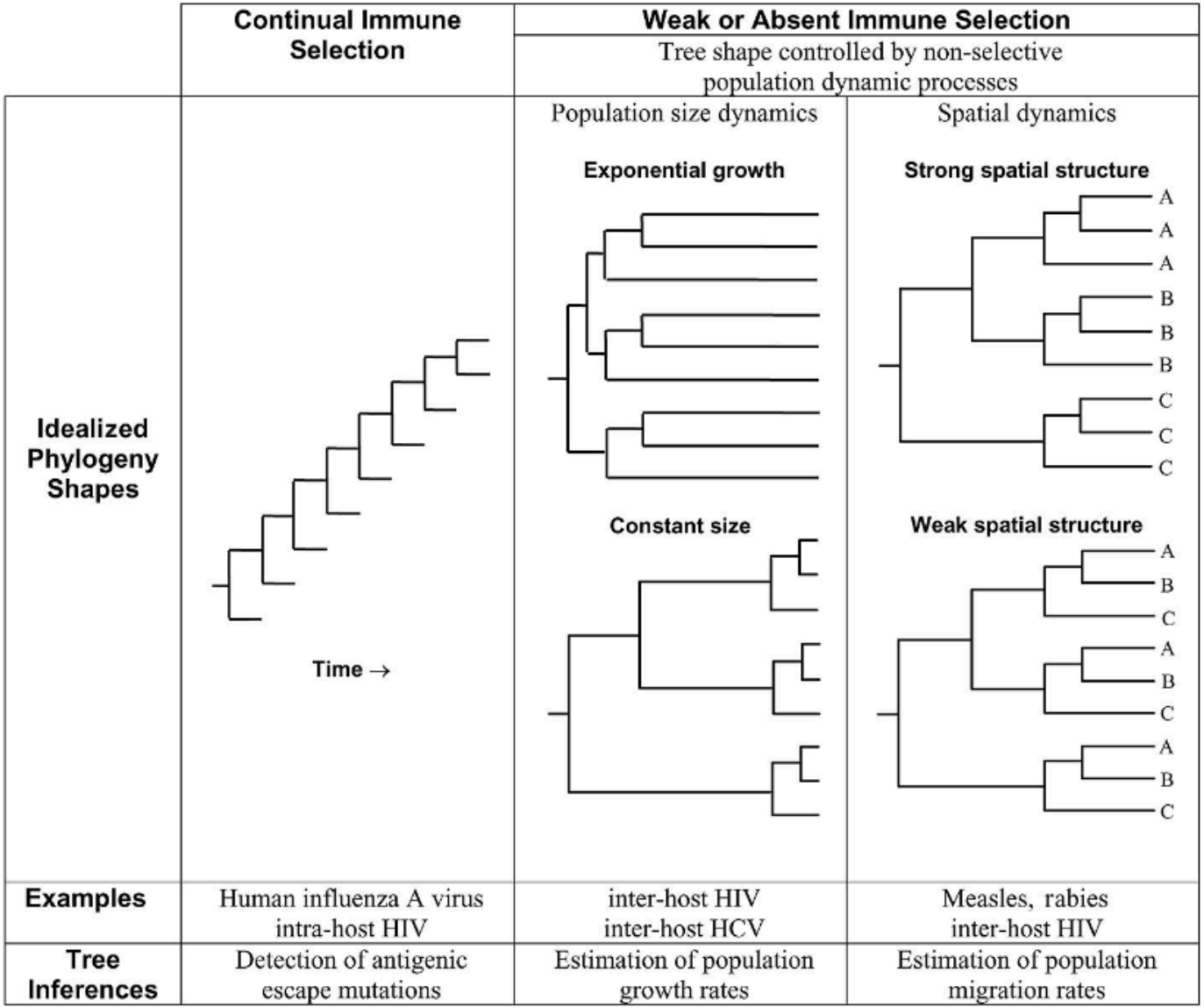


Fig. 3. Idealized tree shapes under different phylodynamic processes. The main division is between those viruses subject to continual immune-driven selection (such as human influenza A virus and intra-host HIV), in which trees have a strong temporal structure, and viruses where immune selection is absent or weak (such as many RNA viruses), in which the trees depict population size and spatial dynamics. The types of evolutionary inference that can be made from the various phylogenies are also indicated. (A, B, and C represent three subpopulations from which viruses have been sampled.)

References I

Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303:327–332.