INTRODUCTION

- Reconstructing transmission histories from molecular data may be able to enhance intervention
- Some viruses (e.g. HIV) evolve quickly, and phylogenetic relationships between viruses reflect transmission histories
- With clinical and epidemiological data, these relationships can provide critical information about drug resistance, associations between sociodemographic characteristics, and epidemic time scales
- Not always clear which method/setting combination performs best for specific epidemiological conditions (e.g. HIV-TRACE vs. TreeCluster)

Simulation provides an inexpensive method to investigate questions about epidemics.

Simulation of transmission networks needs to combine models of social networks, transmission, evolution, and ideally sampling biases and errors.

PANGEA-HIV simulation (developed by the PANGEA-HIV consortium) can simulate realistic HIV transmission dynamics, phylogenetic trees, and sequence data, but the statistical models at each step of the workflow are fixed.

We introduce FAVITES (F:Framework for ViReal Transmission and Evolution Simulation), which can simulate numerous models of contact networks, viral transmission, phylogenetic and sequence evolution, data (sub)sampling, and real-world data perturbations, and which was built to be flexible such that users can seamlessly plug in statistical models and model parameters at every step of the simulation process.

HIV SIMULATION EXPERIMENTAL SETUP

- Emulate HIV transmission in San Diego from 2005 to 2014
- Contact network includes 100,000 individuals with an average of 4 relations per individual under the Barabási–Albert (BA) model
- 15,000 total infected seed individuals
- HIV transmission modeled as a Markov chain epidemic model (see figure below), with states Susceptible (S), Acute Untreated (AU), Acute Treated (AT), Chronically Untreated (CU), and Chronic Treated (CT)
- Transmission model parameters learned from real HIV epidemiological data
- Viral phylogeny in unit of time (years) sampled under coalescent model with logistic growth

Resulting trees merged by simulating a seed tree under pure neutral Kingman coalescent model with an expected height of ~40 years

Evolutionary rate sampled from a log-normal random variable

Root sequence sampled from profile Hidden Markov Model (HMM) generated from HIV-1 subtype B pol sequences from San Diego

Sequences evolved under GTR+I model

RESULTS

DISCUSSION

- FAVITES can simulate under numerous different models and produce realistic data
- TreeCluster and HIV-TRACE, when paired with temporal monitoring, can identify infectious individuals
- HIV-TRACE outperforms TreeCluster under most tested conditions
- Both methods use a threshold to define clusters, and the choice of threshold defines a trade-off between cluster sensitivity and specificity
- FAVITES enables public health officials to simulate conditions similar to their own epidemic and pick the best method/threshold tailored to their situation.

FUNDING

- NIH subaward S00AR027767-28 to SM and NM
- NIH-NIAID K01 Career Development Award K01AI101818, NIH-NIAID R01 AI135992, and CHRP IDEA Award ID15-SD-052 to JOW

REFERENCES