Poster Session xxx
Abstract #xxx
Conference on Retroviruses and Opportunistic Infections
Feb 2023
Seattle, WA USA

ACCURACY OF MOLECULAR EPIDEMIOLOGICAL INFERENCES IN A SIMULATED EPIDEMIC

Fabricia F. Nascimento[1], **Sanjay R. Mehta**[2,3], Susan J. Little[2], Erik M. Volz[1]
[1]Imperial College London, UK, [2]University of California, San Diego, La Jolla, CA, [3]Veterans Affairs San Diego Healthcare System, San Diego, CA

Sanjay Mehta, MD, DTM&H, D(ABMM)
San Diego Veteran Affairs Medical Center
Department of Medicine
San Diego, CA 92161, USA
E-mail: srmehta@ucsd.edu

Background

- Molecular Epidemiology is the 4th pillar of the Ending the HIV Epidemic (EHE) plan, and it is hoped that the use of these analyses will lead to focused prevention and treatment interventions in at-risk populations
- There is concern that inferences derived from molecular epidemiological(ME) analyses could be used in criminal and civil litigation [1].
- The level of concern is impacted by the accuracy of these analyses.

Objective

- To assess the accuracy of ME analyses to **correctly** identify transmission pairs and the direction of transmission in a simulated HIV epidemic based upon the San Diego HIV epidemic using both consensus and next generation sequencing (NGS) technologies.

Methods

- An agent-based simulation based on the EpiModel software package was parameterized to fit the San Diego HIV epidemic among MSM [2].
- Our model utilized five stages of HIV infection and three stages of HIV diagnosis along with a module which allowed individuals to migrate in and out of the study population.
- Latin hypercube sampling (LHS) was used to sample the parameters we could not obtain for the MSM HIV surveillance data in San Diego.
- The simulated data, which included information of who transmitted to whom and timing of transmissions, was used to simulate phylogenetic trees which in turn were used to simulate HIV genetic sequences for the corresponding individuals, incorporating time from infection to sampling.
- Both consensus and next generation sequences were generated.
- These sequences were then analyzed to measure the ability of consensus and NGS to predict correct linkage and directionality [3].
- Analyses were then repeated varying sample fraction, portion of genome sequenced, human migration in and out of the region of interest, and the inclusions of assays for testing recency of infection when inferring who infected whom [4,5].

Figure 1: Simulation is a Reasonable Approximation of San Diego Epidemic

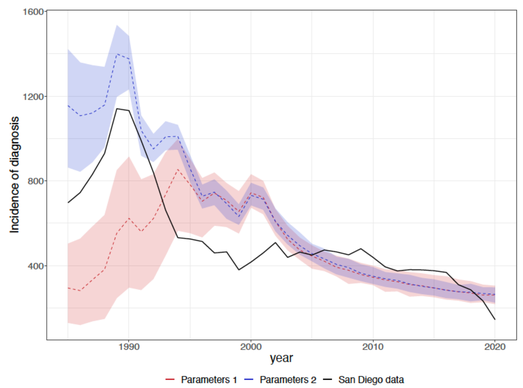


Figure 1: Incidence of diagnosis for the two best combinations of parameter values. Medians and 95% credible interval are shown as dashed line and shaded area, respectively. Black solid line represent the incidence of diagnosis from San Diego surveillance data.

Figure 2: Consensus Sequences do not Reliably Predict Correct Directionality and Linkage, Regardless of Sampling Depth or Sequence Length.

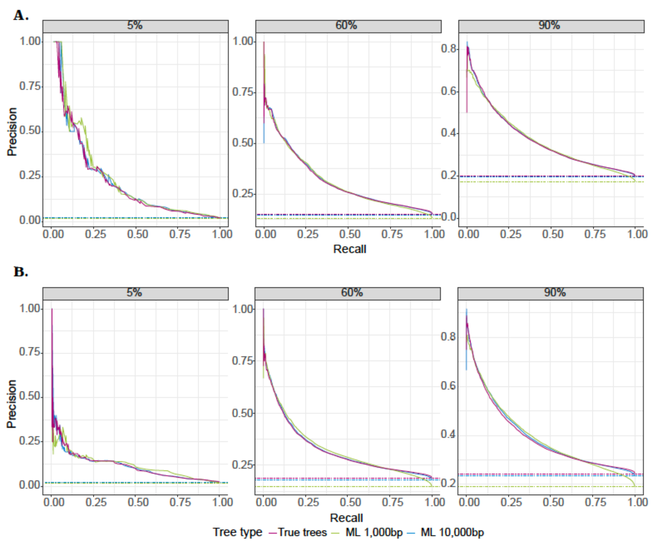


Figure 2: Precision-recall curves (PRC) for prediction of correct directionality and linkage for sampling depths of 5%, 60% and 90% of the diagnosed individuals not on ART for adjacent tips in the phylogenetic tree. (A) parameter combination 1, and (B) parameter combination 2. Migration rate set at 500 persons a year. Dotted line represents a random classifier, while predictions using maximum likelihood trees generated from 1000 bp (green) or 10000 bp (blue) showed similar AUC (values varying from 20% to 40%) as the true trees.

Figure 3: NGS Markedly Improves Predictions of Correct Linkage and Directionality

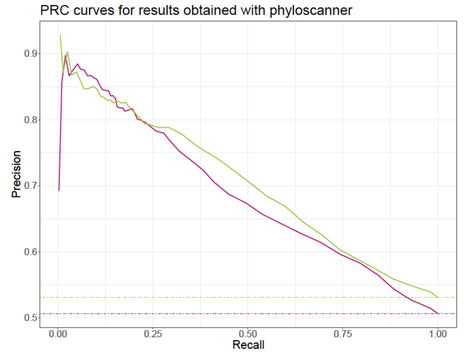


Figure 3: Precision-recall curves for deep-sequencing based on phyloscanner results filtered by infector probability ≥ 1% of the diagnosed individuals not on ART for parameter combinations 1 (red) and 2 (green) (see also Table 1), and 500 persons migrating to and from region annually.

Table 1: Sensitivity and Specificity with NGS

	W ≧80% / W ≧ 1%		
Parameter	**Sensitivity**	**Specificity**	**Precision**
Combination 1	0.06/0.19	0.99/0.96	0.86/0.82
Combination 2	0.07/0.21	0.99/0.95	0.90/0.80

Sensitivity is low and specificity is high for predicting correct linkage and direction with NGS focusing on individuals with an infector probability of either ≥ 80% or ≥ 1%.

Conclusions

- The accuracy of inference of directionality and linkage in a large epidemic using molecular epidemiology is low when consensus sequences are used.
- Accuracy increases significantly if additional epidemiologic data leading to a high pre-test probability are available and NGS data are used, however, even then sensitivity remains low

References

1. Mehta, S. R., Schairer, C., and Little, S. (2019). Ethical issues in HIV phylogenetics and molecular epidemiology. Current Opinion in HIV and AIDS, 14, 221–226.
2. Jenness, S. M., Goodreau, S. M., and Morris, M. (2018). EpiModel: An R package for mathematical modeling of infectious disease over networks. Journal of Statistical Software, 84, 8.
3. Wymant, C., Hall, M., Ratmann, O., Bonsall, D., Golubchik, T., de Cesare, M., Gall, A., Cornelissen, M., Fraser, C., STOP-HCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration (2018b). PHYLOSCANNER: Inferring transmission from within- and between-host pathogen genetic diversity. Molecular Biology and Evolution, 35(3),719–733.
4. Le Vu, S., Ratmann, O., Delpech, V., Brown, A. E., Gill, O. N., Tostevin, A., Fraser, C., and Volz, E. M. (2018). Comparison of cluster-based and source-attribution methods for estimating transmission risk using large HIV sequence databases. Epidemics, 23, 1–10.
5. Zhang, Y., Wymant, C., Laeyendecker, O., Grabowski, M. K., Hall, M., Hudelson, S., Piwowar-Manning, E., McCauley, M., Gamble, T., Hosseinipour, M. C., Kumarasamy, N., Hakim, J. G.,Kumwenda, J., Mills, L. A., Santos, B. R., Grinsztejn, B., Pilotto, J. H., Chariyalertsak, S., Makhema, J., Chen, Y. Q., Cohen, M. S., Fraser, C., and Eshleman, S. H. (2021). Evaluation of phylogenetic methods for inferring the direction of Human Immunodeficiency Virus (HIV) transmission: HIV Prevention Trials Network (HPTN) 052. Clinical Infectious Disease, 72, 30–37.

Acknowledgements: NIH Grants: R01MH124590