

Age-structured Stochastic Model of Inpatient HIV Evolution

Elena E Giorgi^{1,2}, Alan S. Perelson¹ and Tanmoy Bhattacharya^{1,3}

Short Abstract — We study a stochastic linear birth and death process where the birth and death rates depend on age. The novelty of this work is in keeping track of the distribution of generation numbers, *i.e.*, the number of HIV replication cycles that have occurred since the initial infecting genome entered the host. We derive the master equation of the process and compare to a Monte Carlo simulation. The summary statistics obtained here provide a more robust alternative to coalescent theory for estimating times to the last common ancestor for an exponentially growing population.

I. BACKGROUND

THE occurrence of a genetic bottleneck in HIV sexual or mother to infant transmissions has been well documented [1-4]. This results in a majority of new infections being homogeneous, *i.e.*, initiated by a single genetic strain [5]. Furthermore, the viral population grows exponentially during the early phases of infection prior to the onset of the host immune response. In this simple setting, an approach based on comparison of summary statistics is a feasible alternative to the existing Bayesian methods based on simulation of genealogies (e.g. BEAST [6]) for estimating evolutionary and demographic parameters.

II. METHODS

A. Stochastic Model

We implement an age-structured stochastic model that follows the total population of infected cells. At any given time t , the state of the system is described by a function $I(a,g)$ which represents the number of infected cells of age a (where here by age we mean time since the viral particle has entered the cell) and generation g (*i.e.*, the number of replication cycles since the initial infecting strain). Given arbitrary birth and death rates $\alpha(a,g)$ and $\mu(a,g)$ respectively, we derive the master equation of the system for the probability density functional $P(I,t)$ and the evolution equation for the generating functional $F(\theta,t)$.

B. Monte Carlo Simulation

We generate a realization of $I(a,g)$ at all times t as follows: at each time step, we allow for a random number of cell deaths and let the survivors produce a random number of newly infected cells. The simulations are carried out in discrete time, and the convergence as the time step is reduced is studied. The probability of a cell dying and the distribution of the

number of newly infected cells depend on the cell's age and generation and can be chosen to reflect the range of values presented in the literature for the average generation time τ , the basic reproductive ratio R_0 , and the time from cell entry until viral production.

C. Hamming Distance Distribution

Assuming a random base substitution model, we use $I(a,g)$ to derive at all times the Hamming Distance (HD, number of base pairs any two sequences differ at) probability distribution and obtain the time evolution of summary statistics such as the mean, variance and maximum HD. We then compare them to the sequence diversity data in acute HIV infections obtained in recent studies as a robust alternative to coalescent approaches for studying early viral evolution. These simulations can be used to understand early longitudinal within-patient data: by comparing the observed diversity to our model predictions we provide accurate estimates of the time elapsed since the Most Recent Common Ancestor of the sampled sequences with reliable errors bars, even when the sequence data is unsuitable for a phylogenetic analysis.

III. CONCLUSION

This method is useful when the data to be analyzed are not suitable for a complete phylogenetic analysis, for example, due to in vitro recombination, and yet the number of generations since the genetic bottleneck is still small so that exponential population growth can be *assumed*. A modification of our simulations can easily allow for the inclusion of other in vivo processes, such as recombination and selection, which are difficult to handle in standard coalescent theory.

REFERENCES

- [1] Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, et al. (1992) Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. *Science* 255: 1134-1137.
- [2] Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, et al. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* 303: 2019-2022.
- [3] Delwart EL, Magierowska M, Royz M, Foley B, Peddada L, et al. (2001) Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. *AIDS* 15: 1-7.
- [4] Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, et al. (1993) Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* 67: 3345-3356.
- [5] Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, et al. (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A*.
- [6] Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.

¹Los Alamos National Laboratory, Los Alamos NM 87544. E-mail: eggiorgi@lanl.gov (EEG), asp@lanl.gov (ASP)

²University of Massachusetts, Amherst MA 01002

³Santa Fe Institute, Santa Fe 87505. E-mail: tanmoy@lanl.gov