QUASIM: SIMulating Viral QUAsispecies evolution under immune response

Alexander Artyomenko[1], Igor Mandric[1], Pavel Skums[2], Yury Khudyakov[2], and Alex Zelikovsky[1]

[1]Georgia State University, Computer Science Department, Atlanta GA, 30303, USA
[2]Centers fo Desease Control and Prevention, Atlanta GA, 30333, USA

Abstract. Viral infection outbreak detection and prevention are very important tasks in epidemiology. Outbreak simulation can help to develop a reliable method for detection. Existing outbreak simulators do not take in account viral quasispecies. For realistic quasispecies simulation it is important to incorporate immune response. We developed a tool called QUASIM for simultaneous simulation of transmissions and intra-host quasispecies evolution under immune response. We applied existing outbreak simulation and transmission detection tool Outbreaker on our simulated datasets.

Background

Outbreak detection is a well known problem in epidemiology. Modern sequencing technologies provide big amounts of data and manual approach became inapplicable. Several computational methods were developed: threshold based methods [1], *SeqTrack* [2], *Outbraker* [3], etc. Most of them are not designed to work with quasispecies. We applied the most recent tool *Outbreaker* using consesuses of intra-host populations and transmission times.

Method Overview

We simulate evolution of viral infection inside a single host. Our model of infected host consists of 3 main components: the host, the immune system, and the virus. The host is represented as liver and blood, the immune system is represented as B-cells, which can differentiate into plasma cells, antibodies, and T-cells, and the virus is represented as virions in blood and infected cells in liver (see Fig. 1). In each iteration each virion in blood can either bind to a healthy liver cell or bind to a B-cell or stay in blood. In case if a B-cell catches a virion, it differentiates into plasma cell and produces antibodies. Each antibody is targeted at a specific epitope and binds to an antigen (virion) with this epitope. If sufficient amount of antibodies is bound to a virion, it is deactivated and removed from blood.

Inter-host simulation is based on contact graph, which is randomly generated scale-free graph. Transmissions can happen only along edges of a contact graph. When a transmission happens, a randomly chosen virion is transmitted

to another host and new intra-host simulation starts for this host. In our simulation we can store all transmissions and times, however, in real datasets this information is usually not available.

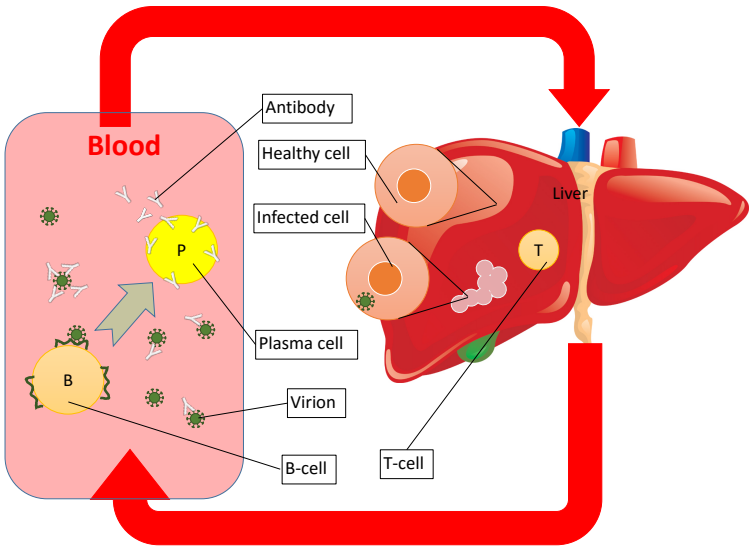

Fig. 1. Overview of a simulation model. Infection blood interacts with liver and host's immune system. Immune system is represented as B-cells, T-cells, plasma cells, and antibodies. Viral infection is represented as virions and infected liver cells.

Results

Diversity definition and plots We measure the *index of nucleotide diversity* or simply *nucleotide diversity* and *nucleotide divergence* [4] to measure evolution during simulations (see Fig. 2).

$$\pi = \frac{N}{N-1} * \sum_{i,j<n} x_i x_j \pi_{ij} = \sum_{i,j<n} \frac{c_i c_j d_{ij}}{N(N-1)L}$$

where c_i is number of virions of i-th variant, d_{ij} is distance between sequences of i-th and j-th variants, L is length of region and N is total number of virions in host's blood.

$$D = \sum_i x_i * \pi_i = \sum_i \frac{c_i d_i}{NL}$$

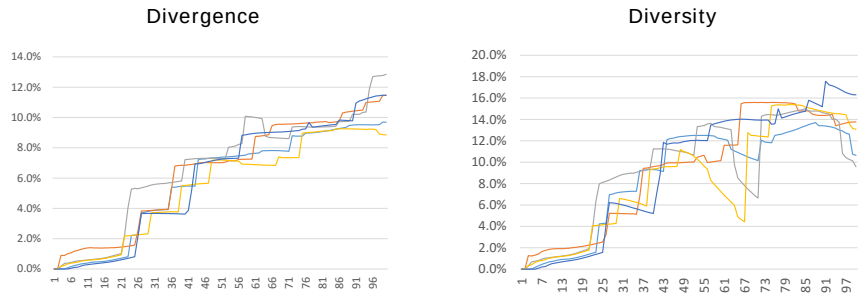

Fig. 2. Nucleotide diversity and divergence of 5 simulations 50 iterations each.

where d_i is the distance between sequences of i-th and initial variants.

In order to validate the simulation model of QUASIM, we generated random outbreaks and ran the Outbreaker tool on them. Our intuitive assumption was that Outbreaker results on QUASIM datasets should correlate well with the results of Outbreaker being run on datasets generated by its native simulator. The advantage of QUASIM is that it takes quasispecies as input, whereas Outbreaker is able to process only a single DNA sequence per sample. To run the experiments, we, therefore, were forced to downgrade the input of QUASIM by computing the consensus sequence for each sample's quasispecies.

To evaluate the relatedness of both simulating tools, we measured the percentage of correctly predicted ancestries by Outbreaker in the QUASIM outbreaks. As our experiments show, QUASIM is a competing tool which is able to simulate outbreaks close to what other state-of-the-art tools produce. Indeed, Outbreaker correctly identified in average 65% of ancestries in our outbreaks. We believe that this proves our initial assumptions regarding the simulation model of our tool.

References

1. Rytsareva, I., Sims, S., Chockalingam, S.P., Aluru, S., Campo, D.S., Tetik, C., Thankachan, S.V., Khudyakov, Y., Zheng, Y., Chirag, J., *et al.*: Efficient detection of viral transmission with threshold-based methods. In: Computational Advances in Bio and Medical Sciences (ICCABS), 2015 IEEE 5th International Conference On, pp. 1–6 (2015). IEEE
2. Jombart, T., Eggo, R., Dodd, P., Balloux, F.: Reconstructing disease outbreaks from genetic data: a graph approach. Heredity **106**(2), 383–390 (2011)
3. Jombart, T., Cori, A., Hellewell, J., Jombart, M.T., EpiEstim, S.: Package 'outbreaker' (2015)
4. Nei, M., Li, W.-H.: Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences **76**(10), 5269–5273 (1979)